The Actuary's Free Study Guide for Exam 4 / Exam C

Second Edition

G. Stolyarov II,

ASA, ACAS, MAAA, CPCU, ARe, ARC, API, AIS, AIE, AIAF

First Edition Published in July-October 2009
Second Edition Published in July 2014

Table of Contents

Section 1

Raw Moments, Central Moments, Excess Loss Variables, and Left-Censored and Shifted Variables

The **kth raw moment** of a random variable X is the expected value of X^k and can be denoted as $E(X^k)$ or μ'_k.

The following formulas hold:

For a continuous distribution with probability density function f(x):

$\boldsymbol{\mu'_k = E(X^k) = {}_{-\infty}{}^{\infty}\int x^k * f(x)dx}$

For a discrete distribution with probability density function $p(x_j)$ for all relevant index values j:

$\boldsymbol{\mu'_k = E(X^k) = {}_j\Sigma x_j{}^k * p(x_j)}$

The **kth central moment** of a random variable X with mean μ is the expected value of the kth power of the variable's deviation from its mean and can be denoted as $E((X-\mu)^k)$ or μ_k.

For a continuous distribution with probability density function f(x):

$\boldsymbol{\mu_k = E((X-\mu)^k) = {}_{-\infty}{}^{\infty}\int(x-\mu)^k * f(x)dx}$

For a discrete distribution with probability density function $p(x_j)$ for all relevant index values j:

$\boldsymbol{\mu_k = E((X-\mu)^k) = {}_j\Sigma(x_j-\mu)^k * p(x_j)}$

An **excess loss variable** can be defined in terms of a random variable X, some given value d such that Pr(X > d) > 0. The excess loss variable is denoted Y^P and is equal to X - d.

The expected value of the excess loss variable is called the **mean excess loss function**, the **mean residual life function**, and the **complete expectation of life.** The following are all acceptable notations for the mean excess loss function:

$e_X(d) = e(d) = E(Y^P) = E(X - d \mid X > d) = \dot{e}_d$.

We can find the kth moment of the excess loss variable as follows.

For a continuous distribution with probability density function f(x) and cumulative distribution function F(x):

$$\mathbf{e_X{}^k(d) = {}_d{}^{\infty}\int(x - d)^k * f(x)dx/(1 - F(d))}$$

For a discrete distribution with probability density function $p(x_j)$ and a cumulative distribution function $F(x_j)$ for all relevant index values j:

$$\mathbf{e_X{}^k(d) = {}_{x\_j>d}\Sigma(x_j - d)^k * p(x_j)/(1 - F(d))}$$

An **left-censored and shifted variable** can be defined in terms of a random variable X, some given value d such that Pr(X > d) > 0. The excess loss variable is denoted Y^L or $(X-d)_+$ and is equal to 0 when X ≤ d and X - d when X > d.

We can find the kth moment of the left-censored and shifted variable as follows.

For a continuous distribution with probability density function f(x) and cumulative distribution function F(x):

$$\mathbf{E((X-d)_+{}^k) = {}_d{}^{\infty}\int(x - d)^k * f(x)dx}$$

For a discrete distribution with probability density function $p(x_j)$ and all relevant index values j:

$$\mathbf{E((X-d)_+{}^k) = {}_{x\_j>d}\Sigma(x_j - d)^k * p(x_j)}$$

A useful shortcut formula to remember in relating $e_X{}^k(d)$ and $E((X-d)_+{}^k)$ is the following:

$$\mathbf{E((X-d)_+{}^k) = e_X{}^k(d)(1 - F(d))}$$

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 3, pp. 21-27.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C1-1. The fraction of a whole pie that Mr. Zolatrax eats during a meal is a random variable X with the following probability density function: $f(x) = 3x^2/8$ for $0 \le x \le 2$ and 0 otherwise. Find the first raw moment of X.

Solution S4C1-1. We have a continuous distribution, so we use the formula $\mu'_k = E(X^k) = {}_{-\infty}{}^{\infty}\int x^k * f(x)dx$ for k =1. Here, we will integrate from 0 to 2, because the random variable is 0 for all other values.

$\mu'_1 = E(X^1) = {}_0{}^2\int x*(3x^2/8)dx = {}_0{}^2\int(3x^3/8)dx = 3x^4/32 \,\Big|\, {}_0{}^2 = 3*16/32 =$ **$E(X) = 3/2 = 1.5$**.

(I know that I just had you find the mean of a fairly simple continuous distribution. This will be useful for further problems in this section. The problems will also gradually get harder from here.)

Problem S4C1-2. The fraction of a whole pie that Mr. Zolatrax eats during a meal is a random variable X with the following probability density function: $f(x) = 3x^2/8$ for $0 \le x \le 2$ and 0 otherwise. Find the second central moment of X.

Solution S4C1-2. We have a continuous distribution, so we use the formula

$\mu_k = E((X-\mu)^k) = {}_{-\infty}{}^{\infty}\int(x - \mu)^k * f(x)dx$ for k = 2. Here, we will integrate from 0 to 2, because the random variable is 0 for all other values. From Solution S4C1-1, we know that μ, the mean of X, is 1.5.

$\mu_2 = E((X-1.5)^2) = {}_0{}^2\int(x - 1.5)^2*(3x^2/8)dx = {}_0{}^2\int(x^2 - 3x + 2.25)*(3x^2/8)dx =$
${}_0{}^2\int(3x^4/8 - 9x^3/8 + 0.84375x^2)dx = (3x^5/40 - 9x^4/32 + 0.28125x^3)\,\vert_0{}^2 = \boldsymbol{\mu_2 = 0.15}$.

(Note that this is the variance of X.)

Problem S4C1-3. Which of the following is the *only* difference between an excess loss variable and a left-censored and shifted variable?You can answer the question by examining the definitions of these two concepts.

(a) A left-censored and shifted variable is censored on the left, while an excess loss variable is censored on the right.

(b) A left-censored and shifted variable assumes values of zero for X < d, while an excess loss variable is undefined for X < d.

(c) An excess loss variable assumes values of zero for X < d, while a left-censored and shifted variable is undefined for X < d.

(d) An excess loss variable applies to discrete functions only, while a left-censored and shifted variable applies to continuous functions only.

(e) An excess loss variable assumes values of zero for X > d, while a left-censored and shifted variable assumes values of zero for X < d.

(f) A left-censored and shifted variable assumes values of zero for X > d, while an excess loss variable assumes values of zero for X < d.

(g) A left-censored and shifted variable is friendly and tame, while an excess loss variable will make you scream in frustration.

Solution S4C1-3.

The correct answer is **(b)** A left-censored and shifted variable assumes values of zero for X < d, while an excess loss variable is undefined for X < d. An excess loss variable is only defined for values of X greater than the threshold d, whereas a left-censored and shifted variable is defined to be zero for all values of X less than d.

Problem S4C1-4. The fraction of a whole pie that Mr. Zolatrax eats during a meal is a random variable X with the following probability density function: $f(x) = 3x^2/8$ for $0 \le x \le 2$ and 0 otherwise. Mr. Zolatrax is concerned that health complications will ensue if he eats more than one whole pie per meal. If he eats more than at least one pie per meal, how much of the second pie can he be expected to eat?

Solution S4C1-4. This question is asking us to determine a mean excess loss function. In other words, we want to find $e_X(1)$. We use the formula

$e_X^k(d) = {}_d^\infty\int(x - d)^k * f(x)dx/(1 - F(d))$ for k = 1 and d = 1. The upper bound of the integral in this case is 2. We also note that $F(x) = \int f(x)dx = x^3/8$ for $0 \le x \le 2$, 0 for x < 0, and 1 for x > 2. Therefore $F(1) = 1^3/8 = 1/8$.

Thus, $e_X(1) = {}_1^2\int(x - 1)*(3x^2/8)dx/(1 - 1/8) = (8/7)\ {}_1^2\int(3x^3/8 - 3x^2/8)dx =$

$(8/7)(3x^4/32 - 3x^3/24)\Big|_1^2 = (8/7)(3*16/32 - 3*8/24 - 3/32 + 3/24) =$

$e_X(1)$ = 0.6071428571 = 17/28 of the second pie.

Problem S4C1-5. The fraction of a whole pie that Mr. Zolatrax eats during a meal is a random variable X with the following probability density function: $f(x) = 3x^2/8$ for $0 \le x \le 2$ and 0 otherwise. Mr. Zolatrax is concerned that health complications will ensue if he eats more than one whole pie per meal. Therefore, he makes an agreement with Mr. Finericlog such that Mr. Finericlog will receive anything in excess of one pie per meal that Mr. Zolatrax would have ordinarily eaten during that meal. What is the expected fraction of a pie that Mr. Finericlog will be getting per meal?

Solution S4C1-5. This question is asking us to determine the expected value of a left-censored and shifted random variable. Fortunately, most of the work has already been done in the previous solutions. We recall the shortcut formula $E((X\text{-}d)_+^k) = e_X^k(d)(1 - F(d))$. We know from Solution S4C1-4 that, for d = 1, F(1) = 1/8 and $e_X(1)$ = 17/28. Therefore, our desired value, $E((X\text{-}1)_+)$ = (17/28)(1 - 1/8) = **0.53125 = 17/32 of a pie**.

Note: All names of hypothetical persons used in this section have been randomly generated.

Section 2

The Empirical Model and Limited Loss Variables

We now introduce another kind of discrete distribution, the **empirical model**. If there is a sample size of n values, the empirical model assigns a probability of 1/n to each individual value in the sample.

A **limited loss variable** can be defined in terms of a random variable X and some given value u at which we censor the distribution of X on the right. If a limited loss variable is denoted as Y, then it can be described as follows: $Y = X \wedge u = X$ when $X < u$ and $Y = X \wedge u = u$ when $X \geq u$. The expected value of the limited loss value is $E(X \wedge u)$ and is called the **limited expected value.**

We can find the kth moment of the limited loss variable as follows.

For a continuous distribution with probability density function f(x) and cumulative distribution function F(x):

$$\mathbf{E((X \wedge u)^k) = {}_{-\infty}{}^{u}\!\int x^k * f(x)dx + u^k*(1 - F(u))}$$

For a discrete distribution with probability density function $p(x_j)$ and a cumulative distribution function $F(x_j)$ for all relevant index values j:

$$\mathbf{E((X \wedge u)^k) = {}_{x\_j \leq u}\Sigma(x_j{}^k * p(x_j)) + u^k*(1 - F(u))}$$

For a continuous distribution with cumulative distribution function F(x) and survival function S(x), the kth moment of the limited loss variable can be expressed as follows:

$$\mathbf{E((X \wedge u)^k) = -({}_{-\infty}{}^{0}\!\int kx^{k-1}*F(x)dx) + {}_{0}{}^{u}\!\int kx^{k-1}*S(x)dx}$$

In particular, $\mathbf{E((X \wedge u)) = -({}_{-\infty}{}^{0}\!\int F(x)dx) + {}_{0}{}^{u}\!\int S(x)dx.}$

Source:

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 3, pp. 22-27.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C2-1. Bombast Corporation provides Internet services. During each of the past 12 months, it has provided poor customer service the following numbers of times per month:

January: 120 times
February: 122 times
March: 124 times
April: 124 times
May: 120 times
June: 120 times
July: 167 times
August: 245 times
September: 245 times
October: 245 times
November: 245 times
December: 150 times

Use an empirical model to construct a discrete probability mass function for X, the number of times per month that Bombast Corporation provides poor customer service.

Solution S4C2-1. Each month is a single data point in our sample. Thus, we have 12 data points, each assigned probability of 1/12. The data point 120 occurs 3 times; 122 occurs 1 time; 124 occurs 2 times; 150 occurs 1 time, 167 occurs 1 time, and 245 occurs 4 times. Thus, we have the following probability mass function:

Pr(X = 120) = 3/12 = 1/4
Pr(X = 122) = 1/12
Pr(X = 124) = 2/12 = 1/6
Pr(X = 150) = 1/12
Pr(X = 167) = 1/12
Pr(X = 245) = 4/12 = 1/3

Problem S4C2-2. Refer to the empirical model developed in Solution S4C2-1 for the number of times Bombast Corporation provides poor customer service in each month. Bob Bombastic, the CEO of Bombast, has decided that his company will only resolve a maximum of 150 customer complaints per month. Assuming that every customer who complains has been mistreated, what is the expected value of customer complaints per month that Bombast will resolve?

Solution S4C2-2. We want to find E(X Λ 150). To do this, we use the formula

$E((X \wedge u)^k) = {}_{x\_j\leq u}\Sigma(x_j{}^k*p(x_j)) + u^k*(1 - F(u))$ for u = 150 and k = 1. Thus:
$E(X \wedge 150) = {}_{x\_j\leq 150}\Sigma(x_j *p(x_j)) + 150*(1 - F(150))$.

We find F(150), the probability that customer complaints in a given month will be less than 150. This is 1 - Pr(X = 167) - Pr(X = 245) = 1 - 1/12 - 4/12 = 7/12. Therefore, 150*(1 - F(150)) = 150(1 - 7/12) = 5*150/12 = 62.5. Now we find ${}_{x\_j\leq 150}\Sigma(x_j *p(x_j))$. The four values of x_j that are less than or equal to 150 are 120, 122, 124, and 150. These have associated probabilities of 3/12, 1/12, 2/12, and 1/12, respectively.

Thus, ${}_{x\_j\leq 150}\Sigma(x_j *p(x_j))$ = 120(3/12) + 122(1/12) + 124(2/12) + 150(1/12) = 73.333333333.

Hence, $E(X \Lambda 150) = 73.333333333 + 62.5 =$ **$E(X \Lambda 150) = 135.8333333333$ complaints per month resolved**.

Problem S4C2-3. The fraction of a whole pie that Mr. Zolatrax eats during a meal is a random variable X with the following probability density function: $f(x) = 3x^2/8$ for $0 \leq x \leq 2$ and 0 otherwise. His health insurance company is willing to pay him for any medical costs that result from his consuming at most 1 pie per meal. Given this limitation, for what fraction of one pie per meal can the health insurance company be expected to pay the associated medical costs?

Solution S4C2-3. We want to find $E(X \Lambda 1)$ for a continuous random variable. Thus, we use the formula $E((X \Lambda u)^k) = {}_{-\infty}{}^{u}\int x^k * f(x)dx + u^k*(1 - F(u))$. Here, $k = 1$, and we integrate from 0 to 1. Moreover, $F(x) = \int f(x)dx = x^3/8$ for $0 \leq x \leq 2$, 0 for $x < 0$, and 1 for $x > 2$. Thus, $F(1) = 1^3/8 = 1/8$. Hence, $E(X \Lambda 1) = {}_0{}^1\int x*(3x^2/8)dx + 1*(1 - 1/8) = {}_0{}^1\int(3x^3/8)dx + 7/8 = (3x^4/32)\big|_0{}^1 + 7/8 = 3/32 + 7/8 =$ **$0.96875 = 31/32$ of a pie**.

Problem S4C2-4. The lifetime of a seven-legged snail in years is exponentially distributed with the following survival function: $S(x) = e^{-0.034x}$ for $x \geq 0$ and 1 for $x < 0$. The Young Seven-Legged Snails Association (YSLSA) accepts all seven-legged snails younger than age 40. Members who reach 40 are allowed to remain in the association and pretend to be 40 in order to fulfill the formal age-based membership requirement. Assuming the snails in the YSLSA are representative of the general seven-legged snail population, what is the expected age of a snail in the YSLSA, as would be reported by that snail (assuming that all members want to stay members)?

Solution S4C2-4. We want to find $E(X \Lambda 40)$. Here, it will be fastest to use the formula $E((X \Lambda u)) = -({}_{-\infty}{}^0\int F(x)dx) + {}_0{}^u\int S(x)dx$. We are already given $S(x) = e^{-0.034x}$, so it follows that $F(x) = 1 - e^{-0.034x}$ for $x \geq 0$. However, since $S(x) = 1$ for $x < 0$, $F(x) = 1 - 1 = 0$ for $x < 0$. Thus, $-({}_{-\infty}{}^0\int F(x)dx)$ becomes $-({}_{-\infty}{}^0\int 0*dx) = 0$. Moreover, $u = 40$. Thus, $E(X \Lambda 40) = {}_0{}^{40}\int e^{-0.034x}dx = -(1/0.034)e^{-0.034x}\big|_0{}^{40} = -(1/0.034)(e^{-0.034*40} - 1) =$ **21.86291832 years**.

Problem S4C2-5. The number of dollars Podcar gains or loses per round of poker follows a uniform distribution and ranges from -\$5 to \$15. The Gamblers' Charitable Fund has made the following deal with Podcar: it will pay for all of his losses but will also receive his profits up to \$10 per round. How many dollars per round can the Gamblers' Charitable Fund be expected to gain?

Solution S4C2-5. Since the range of outcomes for Podcar is 20 dollars, the uniform distribution for X, the number of dollars Podcar gains or loses per round, has a probability density function of $f(x) = 1/20$ and therefore a cumulative distribution function of $F(x) = (x+5)/20$ for $-5 \leq x \leq 15$. Hence, $S(x) = 1 - x/20$ for $-5 \leq x \leq 15$. Since we are dealing with a continuous distribution, we can use the formula $E((X \Lambda u)) = -({}_{-\infty}{}^0\int F(x)dx) + {}_0{}^u\int S(x)dx$. Here, $u = 10$, and the lower bound of integration is -5. Thus, $E((X \Lambda 10)) = -({}_{-5}{}^0\int((x+5)/20)dx) + {}_0{}^{10}\int(1 - (x+5)/20)dx = -((x+5)^2/40)\big|_{-5}{}^0 + (x - (x+5)^2/40)\big|_0{}^{10} = -(25/40) + (10 - 225/40) + 25/40 = 10 - 225/40 =$ **\$4.375**.

Section 3

Coefficient of Variation, Skewness, Kurtosis, and a Shortcut Formula for Mean Excess Loss Functions

The **coefficient of variation** of a random variable is the ratio of its standard deviation σ to its mean μ. We shall abbreviate the coefficient of variation for some random variable X as CV(X). Thus, $\mathbf{CV(X) = \sigma/\mu}$.

The **skewness** of a random variable is the ratio of its third central moment μ_3 to the cube of its standard deviation σ. Skewness is denoted as γ_1. Thus, $\boldsymbol{\gamma_1 = \mu_3/\sigma^3}$.

The **kurtosis** of a random variable is the ratio of its fourth central moment μ_4 to the fourth power of its standard deviation σ. Kurtosis is denoted as γ_2. Thus, $\boldsymbol{\gamma_2 = \mu_4/\sigma^4}$.

To find the mean excess loss function $e_X(d)$ of a continuous random variable X at some given value d, there is a useful shortcut formula: $\mathbf{e_X(d) = {}_d{}^{\infty}\!\int S(x)dx/S(d)}$, where S(x) is the survival function of X.

Source:

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 3, pp. 23-27.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C3-1. The fraction of a whole pie that Mr. Zolatrax eats during a meal is a random variable X with the following probability density function: $f(x) = 3x^2/8$ for $0 \le x \le 2$ and 0 otherwise. Find the coefficient of variation of X.

Solution S4C3-1. We seek $CV(X) = \sigma/\mu$. We first find μ, which is the first raw moment of X:

$\mu'_1 = E(X^1) = {}_0{}^2\!\int x*(3x^2/8)dx = {}_0{}^2\!\int(3x^3/8)dx = 3x^4/32 \,|\, {}_0{}^2 = 3*16/32 = E(X) = 3/2 = 1.5.$

We then find Var(X), which is the second central moment of X. We use the formula

$\mu_k = E((X- \mu)^k) = {}_{-\infty}{}^{\infty}\!\int(x - \mu)^k*f(x)dx$. Thus,
$\mu_2 = E((X- 1.5)^2) = {}_0{}^2\!\int(x - 1.5)^2*(3x^2/8)dx = {}_0{}^2\!\int(x^2 - 3x + 2.25)*(3x^2/8)dx =$

${}_0{}^2\!\int(3x^4/8 - 9x^3/8 + 0.84375x^2)dx = (3x^5/40 - 9x^4/32 + 0.28125x^3) \,|\, {}_0{}^2 = \mu_2 = 0.15.$

Since $\sigma = \sqrt{(Var(X))}$, it follows that $\sigma = \sqrt{(0.15)} = 0.3872983346$. Therefore, CV(X) = 0.3872983346/1.5 = **CV(X) = 0.2581988897**.

Problem S4C3-2. Recall the empirical model we developed in Section 2 for the number of customer complaints Bombast Corporation receives per month:

Pr(X = 120) = 3/12 = 1/4
Pr(X = 122) = 1/12
Pr(X = 124) = 2/12 = 1/6
Pr(X = 150) = 1/12
Pr(X = 167) = 1/12
Pr(X = 245) = 4/12 = 1/3

Find the coefficient of variation of X.

Solution S4C3-2. We seek CV(X) = σ/μ. First, we find the mean μ = 120(1/4) + 122(1/12) + 124(1/6) + 150(1/12) + 167(1/12) + 245(1/3) = μ = 168.9166667.

Then we find Var(X), the second central moment of X.

We use the formula $\mu_k = E((X- \mu)^k) = {}_j\Sigma(x_j - \mu)^k *p(x_j)$ for k = 2.

$E((X- 168.9166667)^2) = (120 - 168.9166667)^2(1/4) + (122 - 168.9166667)^2(1/12) + (124 - 168.9166667)^2(1/6) + (150 - 168.9166667)^2(1/12) + (167 - 168.9166667)^2 (1/12) + (245 - 168.9166667)^2 (1/3)$ = Var(X) = 3077.576389.

Since $\sigma = \sqrt{(Var(X))}$, it follows that $\sigma = \sqrt{(3077.576389)} = 55.47590818$.
Therefore, CV(X) = 55.47590818/168.9166667 = **CV(X) = 0.3284217754**.

Problem S4C3-3. Recall the empirical model we developed in Section 2 for the number of customer complaints Bombast Corporation receives per month:

Pr(X = 120) = 3/12 = 1/4
Pr(X = 122) = 1/12
Pr(X = 124) = 2/12 = 1/6
Pr(X = 150) = 1/12
Pr(X = 167) = 1/12
Pr(X = 245) = 4/12 = 1/3

Find the skewness of X.

Solution S4C3-3. We seek $\gamma_1 = \mu_3/\sigma^3$. We know from Solution S4C3-2 that σ = 55.47590818. Now we use the formula $\mu_k = E((X- \mu)^k) = {}_j\Sigma(x_j - \mu)^k *p(x_j)$ to find μ_3.

$\mu_3 = E((X- 168.9166667)^3) = (120 - 168.9166667)^3(1/4) + (122 - 168.9166667)^3(1/12) + (124 - 168.9166667)^3(1/6) + (150 - 168.9166667)^3(1/12) + (167 - 168.9166667)^3(1/12) + (245 - 168.9166667)^3(1/3) = 93270.81134$.

Thus, $\gamma_1 = 93270.81134/55.47590818^3 = \boldsymbol{\gamma_1 = 0.5463016252}$.

Problem S4C3-4. Recall the empirical model we developed in Section 2 for the number of customer complaints Bombast Corporation receives per month:

$\Pr(X = 120) = 3/12 = 1/4$
$\Pr(X = 122) = 1/12$
$\Pr(X = 124) = 2/12 = 1/6$
$\Pr(X = 150) = 1/12$
$\Pr(X = 167) = 1/12$
$\Pr(X = 245) = 4/12 = 1/3$

Find the kurtosis of X.

Solution S4C3-4. We seek $\gamma_2 = \mu_4/\sigma^4$. We know from Solution S4C3-2 that $\sigma = 55.47590818$. Now we use the formula $\mu_k = E((X- \mu)^k) = {}_j\Sigma(x_j - \mu)^k * p(x_j)$ to find μ_4.

$\mu_4 = E((X- 168.9166667)^4) = (120 - 168.9166667)^4(1/4) + (122 - 168.9166667)^4(1/12) + (124 - 168.9166667)^4(1/6) + (150 - 168.9166667)^4(1/12) + (167 - 168.9166667)^4(1/12) + (245 - 168.9166667)^4(1/3) = 13693826.62$.

Thus, $\gamma_2 = 13693826.62/55.47590818^4 = \boldsymbol{\gamma_2 = 1.44579641}$.

Problem S4C3-5. A Pareto distribution has the following survival function: $S\boldsymbol{(x) = \theta^\alpha/(x + \theta)^\alpha}$for some specified parameters α and θ with $\alpha \geq 1$ and $\theta \geq 0$. Find the mean excess loss function for a Pareto distribution with some given values of α and θ at $x = 3\theta$. Express your answer in terms of α and θ if necessary.

Solution S4C3-5. We seek $e_X(3\theta)$. We can now use our shortcut formula $e_X(d) = {}_d{}^\infty\int S(x)dx/S(d)$ for $d = 3\theta$. We find $S(3\theta) = \boldsymbol{\theta^\alpha/(3\theta + \theta)^\alpha} = \boldsymbol{\theta^\alpha/(4\theta)^\alpha} = 1/4^{\boldsymbol{\alpha}}$. Therefore, $1/S(3\theta) = 4^{\boldsymbol{\alpha}}$.

Hence, $e_X(3\theta) = 4^{\boldsymbol{\alpha}} * {}_{3\theta}{}^\infty\int(\boldsymbol{\theta^\alpha/(x + \theta)^\alpha})dx = 4^{\boldsymbol{\alpha}} * {}_d{}^\infty\int S(x)dx = 4^{\boldsymbol{\alpha}} * -(1/(\alpha - 1))\theta^\alpha/(x + \theta)^{\alpha-1} \Big|_{3\theta}^{\infty} =$

$4^{\boldsymbol{\alpha}} * (1/(\alpha - 1))\theta^\alpha/(3\theta + \theta)^{\alpha-1} = 4^{\boldsymbol{\alpha}} * \theta^\alpha/((4\theta)^{\alpha-1}(\alpha - 1)) = \boldsymbol{e_X(3\theta) = 4\theta/(\alpha - 1)}$.

Section 4

Percentiles, Value-at-Risk, and Tail-Value-at-Risk

For a random variable, the **100pth percentile** is denoted $\boldsymbol{\pi_p}$. It can also be called **Value-at-Risk** at the 100p% level,denoted $\mathbf{VaR_p(X)}$ for some random variable X. $VaR_p(X) = \pi_p$ can most conveniently be found using the following formulas:

$\mathbf{1 - p = S(\pi_p)}$, where S(x) is the survival function for the random variable X.

$\mathbf{p = F(\pi_p)}$, where F(x) is the cumulative distribution function for the random variable X.

The **Tail-Value-at-Risk** is related to the Value-at-Risk. For some random variable X, the Tail-Value-at-Risk at the 100p% level, denoted $\mathbf{TVaR_p(X)}$, is "the expected loss given that the loss exceeds the 100p percentile of the distribution of X" (Klugman, Panjer, and Wilmot 2008, p. 45). In other words, $\mathbf{TVaR_p(X) = E(X \mid X > \pi_p)}$**.** For continuous distributions, We can calculate $TVaR_p(X)$ using the following formula:

$$\mathbf{TVaR_p(X) = \pi_p + (E(X) - E(X \wedge \pi_p))/(1 - p)}$$

Source: *Loss Models: From Data to Decisions*, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 3, pp. 29, 44-45.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C4-1. Marder the Mugger occasionally mugs people in dark alleys and takes their money. Lexthor takes daily walks near the alleys where Marder mugs people. The number of golden hexagons (GH) that Lexthor can expect to surrender to Marder on any given day follows a distribution with a cumulative distribution function of $F(x) = 1 - 0.5/(x + 1)$ for $x \geq 0$ and 0 otherwise. Find the 40^{th} percentile of the amount Lexthor can be expected to surrender.

Solution S4C4-1. We note that if we follow the formula $p = F(\pi_p)$ and set $0.4 = 1 - 0.5/(\pi_{0.4} + 1)$, we get $\pi_{0.4} = -0.2$, and Marder does not give money away to the people he mugs! Indeed, the distribution is defined such that $F(x) = 1 - 0.5/(x + 1)$ only applies for $x \geq 0$, which means that any amount less than 0 has a probability of 0 associated with it. We find $F(0) = 1 - 0.5/(0 + 1) = 0.5$, meaning that 0.5 of the time, Marder will fail to mug Lexthor and will not take any money. Hence, since $0.4 < 0.5$, $\boldsymbol{\pi_{0.4} = 0}$.

Problem S4C4-2. X follows a Pareto distribution with $\theta = 6000$ and $\alpha = 5$. Find $VaR_{0.7}(X)$.

Relevant properties for Pareto distributions: $S(x) = \theta^\alpha/(x + \theta)^\alpha$.

$E(X) = \theta/(\alpha - 1)$

$E(X \wedge K) = (\theta/(\alpha - 1))(1 - (\theta/(K + \theta))^{(\alpha - 1)})$

Note: You will be able to find information about a wide variety of continuous distributions from the Exam 4 / C Tables, which will be provided to you during the exam. The relevant information for each distribution will, however, also be listed here to save you the time required to reference it.

Solution S4C4-2. We use the formula $1 - p = S(\pi_p)$ to find $\pi_{0.7}$. Here, $S(x) = 6000^5/(x + 6000)^5$.

Thus, for p = 0.7, $1 - 0.7 = 6000^5/(\pi_{0.7} + 6000)^5 \rightarrow 0.3^{1/5} = (6000)/(\pi_{0.7} + 6000) \rightarrow 0.7860030856 = (6000)/(\pi_{0.7} + 6000) \rightarrow (\pi_{0.7} + 6000) = 7633.557819 \rightarrow \pi_{0.7}$ = **$VaR_{0.7}(X)$** = 1633.557819**.**

Problem S4C4-3. X follows a Pareto distribution with $\theta = 6000$ and $\alpha = 5$. Find $TVaR_{0.7}(X)$.

Relevant properties for Pareto distributions: $S(x) = \theta^\alpha/(x + \theta)^\alpha$.

$E(X) = \theta/(\alpha - 1)$

$E(X \wedge K) = (\theta/(\alpha - 1))(1 - (\theta/(K + \theta))^{(\alpha - 1)})$

Solution S4C4-3. We use the formula $TVaR_p(X) = \pi_p + (E(X) - E(X \wedge \pi_p))/(1 - p)$. From Solution S4C4-2, we know that $\pi_{0.7} = 1633.557819$. Moreover p = 0.7, so 1 - p = 0.3.

We find $E(X) = \theta/(\alpha - 1) = 6000/(5 - 1) = E(X) = 1500$.

Now we find $E(X \wedge \pi_{0.7}) = E(X \wedge 1633.557819) =$

$(6000/(5 - 1))(1 - (6000/(6000 + 1633.557819))^{(5 - 1)}) = 1500(1 - 0.7860030856^4) = 1500*0.618322109 = E(X \wedge \pi_{0.7}) = 927.4831635$.

Now we find $TVaR_{0.7}(X) = 1633.557819 + (1500 - 927.4831635)/0.3 =$ **$TVaR_p(X)$ = 3541.947274**.

Problem S4C4-4. X follows a Weibull distribution with $\theta = 100$ and $\tau = 0.25$. Find $VaR_{0.9}(X)$.

Relevant properties for Weibull distributions:

$S(x) = \exp(-(x/\theta)^\tau)$

$E(X) = \theta*\Gamma(1 + 1/\tau)$.

$E(X \wedge K) = \theta*\Gamma(1 + 1/\tau)*\Gamma(1 + 1/\tau; (K/\theta)^\tau) + K\exp(-(x/\theta)^\tau)$.

$\Gamma(r)$ is the Gamma function for r, whereas $\Gamma(r; K)$ is the incomplete Gamma function for r and K.

Note that $\Gamma(r) = (r - 1)!$ whenever r is an integer.

The incomplete Gamma function can be found as follows: $\Gamma(r; K) = (1/\Gamma(r))({}_0{}^K\int t^{r-1}*e^{-t}dt)$.

Solution S4C4-4. We use the formula $1 - p = S(\pi_p)$ to find $\pi_{0.9}$. Here, $S(x) = \exp(-(x/100)^{0.25})$, and p = 0.9. Thus, $1 - 0.9 = \exp(-(\pi_{0.9}/100)^{0.25}) \rightarrow \ln(0.1) = -(\pi_{0.9}/100)^{0.25} \rightarrow$

$\pi_{0.9} = 100*(-\ln(0.1))^4 = \pi_{0.9} =$ $\mathbf{VaR_{0.9}(X) = 2811.012357}$.

Problem S4C4-5. X follows a Weibull distribution with $\theta = 100$ and $\tau = 0.25$. Find $TVaR_{0.9}(X)$. *Use a calculator or computer to perform any necessary integration.*

Relevant properties for Weibull distributions:

$S(x) = \exp(-(x/\theta)^{\tau})$
$E(X) = \theta*\Gamma(1 + 1/\tau)$.
$E(X \wedge K) = \theta*\Gamma(1 + 1/\tau)*\Gamma(1 + 1/\tau; (K/\theta)^{\tau}) + K\exp(-(K/\theta)^{\tau})$.

$\Gamma(r)$ is the Gamma function for r, whereas $\Gamma(r; K)$ is the incomplete Gamma function for r and K.

Note that $\Gamma(r) = (r - 1)!$ whenever r is an integer.

The incomplete Gamma function can be found as follows: $\Gamma(r; K) = (1/\Gamma(r))({}_0{}^K\int t^{r-1}*e^{-t}dt)$.

Solution S4C4-5. We use the formula $TVaR_p(X) = \pi_p + (E(X) - E(X \wedge \pi_p))/(1 - p)$. From Solution S4C4-4, we know that $\pi_{0.9} = 2811.012357$. Moreover p = 0.9, so 1 - p = 0.1.

We find $E(X) = \theta*\Gamma(1 + 1/\tau) = 100*\Gamma(1 + 1/0.25) = 100*\Gamma(5) = 100*4! = E(X) = 2400$.

Now we find $E(X \wedge \pi_{0.9}) = E(X \wedge 2811.012357) =$
$\theta*\Gamma(1 + 1/\tau)*\Gamma(1 + 1/\tau; (2811.012357/\theta)^{\tau}) + 2811.012357\exp(-(2811.012357/\theta)^{\tau})$.
We know that $\theta*\Gamma(1 + 1/\tau) = E(X) = 2400$ and that $\exp(-(2811.012357/\theta)^{\tau}) = 0.1$.
Thus, $2811.012357\exp(-(2811.012357/\theta)^{\tau}) = 281.1012357$.
Hence, $E(X \wedge \pi_{0.9}) = 2400*\Gamma(1 + 1/\tau; (2811.012357/\theta)^{\tau}) + 281.1012357$.
We want to find $\Gamma(1 + 1/\tau; (2811.012357/\theta)^{\tau}) = \Gamma(1 + 1/0.25; (2811.012357/100)^{0.25}) =$ $\Gamma(5; 2.302585093)$. We now use the formula $\Gamma(r; K) = (1/\Gamma(r))({}_0{}^K\int t^{r-1}*e^{-t}dt)$.
$\Gamma(5; 2.302585093) = (1/\Gamma(5))({}_0{}^{2.302585093}\int t^4*e^{-t}dt) = (1/24)({}_0{}^{2.302585093}\int t^4*e^{-t}dt) = \Gamma(5; 2.302585093) = 0.0840532111$. Hence, $E(X \wedge \pi_{0.9}) = 2400*0.0840532111 + 281.1012357 =$ 482.8289422.
Therefore, $TVaR_{0.9}(X) = 2811.012357 + (2400 - 482.8289422)/0.1 =$ $\mathbf{TVaR_{0.9}(X) = 21982.72294}$.

Note: All names of hypothetical persons used in this section have been randomly generated.

Section 5

The Central Limit Theorem and Applications to Special Distributions

The **central limit theorem** can be phrased as follows (paraphrasing Klugman, Panjer, and Wilmot, p. 30).

Let the random variables $X_1, ..., X_k$ be independent.
Let S_k be a random variable such that $E(S_k) = E(X_1) + ... + E(X_k)$.
Since $X_1, ..., X_k$ are independent, $Var(S_k) = Var(X_1) + ... + Var(X_k)$.

Then, if the first and second moments of these random variables meet certain conditions, it is the case that $\mathbf{\lim_{k\to\infty}(S_k - E(S_k))/\sqrt{(Var(S_k))}}$ **follows a normal distribution with mean 0 and variance 1**.

This means that for continuous distributions, the following formulas apply:

$\mathbf{Pr(S_k > x) \approx 1 - \Phi((x - E(S_k))/\sqrt{(Var(S_k))})}$

$\mathbf{Pr(S_k < x) \approx \Phi((x - E(S_k))/\sqrt{(Var(S_k))})}$

$\mathbf{Pr(a < S_k < b) \approx \Phi((b - E(S_k))/\sqrt{(Var(S_k))}) - \Phi((a - E(S_k))/\sqrt{(Var(S_k))})}$

Here, $\mathbf{\Phi(x)}$ is the standard normal cumulative distribution function at x. It can be found in Microsoft Excel using the formula input "=NORMSDIST(x)" with a numerical value substituted for x.

Note: The strict inequalities above can be replaced with "≤" or "≥" signs, where appropriate, without affecting the formulas, since no continuity correction is necessary for already continuous distributions.

This section will focus on using the central limit theorem to evaluate sums of random variables that follow special distributions that are likely to appear on the exam.

Source: *Loss Models: From Data to Decisions*, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 3, p. 30.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C5-1. Each of the 20 independent random variables $X_1, ..., X_{20}$ follows a Pareto distribution with $\theta = 6000$ and $\alpha = 5$. Use the central limit theorem to find the probability that $S_{20} = X_1 + ... + X_{20}$ is greater than 36000.

Relevant properties for Pareto distributions: $S(x) = \theta^\alpha/(x + \theta)^\alpha$.

$E(X) = \theta/(\alpha - 1)$

$E(X^2) = 2\theta^2/((\alpha - 1)(\alpha - 2))$

Note: You will be able to find information about a wide variety of continuous distributions from the Exam 4 / C Tables, which will be provided to you during the exam. The relevant information for each distribution will, however, also be listed here to save you the time required to reference it.

Solution S4C5-1. We first find $E(X_j) = \theta/(\alpha - 1) = 6000/(5 - 1) = E(X_j) = 1500$ for each j.

Thus, $E(S_{20}) = 20*E(X_j) = 30000$.

Then we find $Var(X_j)$ for each j.

$Var(X_j) = E(X_j^2) - E(X_j)^2 = 2\theta^2/((\alpha - 1)(\alpha - 2)) - 1500^2 = 2*6000^2/((5 - 1)(5 - 2)) - 1500^2 =$ $Var(X_j) = 3750000$. Thus, $Var(S_{20}) = 20*Var(X_j) = 75000000$, implying that $\sqrt{(Var(S_{20}))} =$ 8660.254038.

We are asked to find $Pr(S_{20} > 36000)$, which we can do using the formula

$Pr(S_k > x) \approx 1 - \Phi((x - E(S_k))/\sqrt{(Var(S_k))})$:

$Pr(S_{20} > 36000) = 1 - \Phi((36000 - 30000)/8660.254038) = 1 - \Phi(0.692820323)$, which we can find in MS Excel using the input "1-NORMSDIST(0.692820323)". +The desired probability is therefore **0.244211158**.

Problem S4C5-2. Each of the 36 independent random variables $X_1, ..., X_{36}$ follows an exponential distribution with $\theta = 20$. Use the central limit theorem to find the probability that $S_{36} = X_1 + ... + X_{36}$ is less than 690.

Relevant properties for exponential distributions: $S(x) = e^{-x/\theta}$
$E(X) = \theta$

$Var(X) = \theta^2$.

Solution S4C5-2. We are given that $E(X_j) = \theta = 20$ for each j. Thus, $E(S_{36}) = 36*E(X_j) = 720$. We can find $Var(X_j) = \theta^2 = 400$ for each j. Thus**,** $Var(S_{36}) = 36*400 = 14400$. This means that $\sqrt{(Var(S_{36}))} = 120$. We are asked to find $Pr(S_{36} < 690)$, which we can do using the formula

$Pr(S_k < x) \approx \Phi((x - E(S_k))/\sqrt{(Var(S_k))})$:

$Pr(S_{36} < 690) = \Phi((690 - 720)/120) = \Phi(-0.25)$, which we can find in MS Excel using the input "=NORMSDIST(-0.25)". The desired probability is therefore **0.401293674**.

Problem S4C5-3. Each of the 10 independent random variables X_1, ..., X_{10} follows a Weibull distribution with $\theta = 100$ and $\tau = 0.25$. Use the central limit theorem to find the probability that $S_{10} = X_1 + ... + X_{10}$ is between 23000 and 26000.

Relevant properties for Weibull distributions:

$S(x) = \exp(-(x/\theta)^\tau)$

$E(X) = \theta*\Gamma(1 + 1/\tau)$

$E(X^2) = \theta^2*\Gamma(1 + 2/\tau)$

Note that $\Gamma(r) = (r - 1)!$ whenever r is an integer.

Solution S4C5-3. We find $E(X_j) = \theta*\Gamma(1 + 1/\tau) = 100*\Gamma(1 + 1/0.25) = 100*\Gamma(5) = 100*4! =$ $E(X_j) = 2400$ for each j. Thus, $E(S_{10}) = 10*E(X_j) = 24000$.

Then we find $Var(X_j)$ for each j.

$Var(X_j) = E(X_j^2) - E(X_j)^2 = \theta^2*\Gamma(1 + 2/\tau) - 2400^2 = 100^2*\Gamma(1 + 2/0.25) - 2400^2 = 10000*\Gamma(9) -$ $5760000 = 10000*8! - 5760000 = Var(X_j) = 397440000$.

Thus, $Var(S_{10}) = 10*397440000 = 3974400000$ and $\sqrt{(Var(S_{10}))} = 63042.84258$.

We are asked to find $Pr(23000 < S_k < 26000)$, which we can do using the formula

$Pr(a < S_k < b) \approx \Phi((b - E(S_k))/\sqrt{(Var(S_k))}) - \Phi((a - E(S_k))/\sqrt{(Var(S_k))})$:

$Pr(23000 < S_k < 26000) = \Phi((26000 - 24000)/63042.84258) - \Phi((23000 - 24000)/63042.84258)$ $= \Phi(0.0317244578) - \Phi(-0.0158622289)$, which we can find in MS Excel using the input "=NORMSDIST(0.0317244578)-NORMSDIST(-0.0158622289)". The desired probability is therefore **0.018981953**.

Problem S4C5-4. Each of the 50 independent random variables X_1, ..., X_{50} follows a Gamma distribution with $\theta = 900$ and $\alpha = 8$. Use the central limit theorem to find the probability that $S_{50} = X_1 + ... + X_{50}$ is greater than 364000.

Relevant properties for Gamma distributions:

$E(X) = \alpha\theta$

$Var(X) = \alpha\theta^2$

Solution S4C5-4.

We find $E(X_j) = \alpha\theta$ **= 900*8 = 7200 for each** X_j**.**

This means that $E(S_{50}) = 50^* E(X_j) = 360000$.

We find $Var(X_j) = \alpha\boldsymbol{\theta}^2$ **= 8*900^2 = 64800000 for each** X_j**.**

This means that $Var(S_{50}) = 50^*Var(X_j) = 324000000$ and $\sqrt{}(Var(S_{50})) = 18000$.

We are asked to find $Pr(S_{50} > 364000)$, which we can do using the formula

$Pr(S_k > x) = 1 - \Phi((x - E(S_k))/\sqrt{}(Var(S_k)))$:

$Pr(S_{50} > 364000) = 1 - \Phi((364000 - 360000)/18000) = 1 - \Phi(0.2222222222)$, which we can find in MS Excel using the input "=1-NORMSDIST(0.2222222222)". The desired probability is therefore **0.412070448**.

Problem S4C5-5. Of the 900 corporate bailouts in the United States of Bailoutland (USB) in 2010, the mean amount taxpayers lost per bailout was $2 billion, with a standard deviation of $50 million. It is expected that there will be 10000 corporate bailouts in 2011. Use the central limit theorem to find the probability that taxpayers will lose more than $19.98 trillion to corporate bailouts in 2011.

Solution S4C5-5. In billions, we are given $E(X_j) = 2$, so $E(S_{10000}) = 10000^*E(X_j) = 20000$.

We are given $\sqrt{}(Var(X_j)) = 0.05$, so $\sqrt{}(Var(S_{10000})) = \sqrt{}(10000)\sqrt{}(Var(X_j)) = 100^*0.05 = 5$.

We are asked to find $Pr(S_{10000} > 19980)$, which we can do using the formula

$Pr(S_k > x) \approx 1 - \Phi((x - E(S_k))/\sqrt{}(Var(S_k)))$:

$Pr(S_{10000} > 19980) = 1 - \Phi((19980 - 20000)/5) = 1 - \Phi(-4)$, which we can find in MS Excel using the input "1-NORMSDIST(-4)". The desired probability is therefore **0.999968329**, a virtual certainty!

Section 6

Moment Generating Functions and Probability Generating Functions

Let X be a random variable.

The **moment generating function** (mgf) of X is $\mathbf{M_X(t)}$ and is equal to $\mathbf{E(e^{tX})}$ for all t for which the expected value exists (Klugman, Panjer, and Willmot, p. 30).

The **probability generating function** (pgf) of X is $\mathbf{P_X(z)}$ and is equal to $\mathbf{E(z^X)}$ for all z for which the expectation exists (Klugman, Panjer, and Willmot, p. 30).

It is useful to note that $\mathbf{M_X(t) = P_X(e^t)}$ and $\mathbf{P_X(z)} = \mathbf{M_X(ln(z))}$. Two random variables with different distribution functions also cannot have the same mgf and pgf. Thus, an mgf or pgf provides a unique indicator of what a random variable's distribution is.

The following theorem is useful for working with sums of random variables:
Let $S_k = X_1 + ... + X_k$, where the random variables X_1, ..., X_k are independent. Then $\mathbf{M_{S\_k}(t) = {}_{j=1}{}^{k}\Pi(M_{X\_j}(t))}$ and $\mathbf{P_{S\_k}(z) = {}_{j=1}{}^{k}\Pi(P_{X\_j}(z))}$ given that all the component mgfs and pgfs exist (Klugman, Panjer, and Willmot, p. 30).

For the special distributions described in the Exam 4 / C Tables, a few of the moment generating functions are given, but most are not. Therefore, it is useful to be able to derive them oneself, starting with the probability density functions. Here, we will focus on deriving the mgfs and pdfs of exponential and gamma distributions. These mgfs *are* provided in the exam tables, but their derivations will enable us to have some practice with the formulas and properties above.

Source: *Loss Models: From Data to Decisions*, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 3, pp. 30-34.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C6-1. Using the formulas discussed above, find the moment generating function $M_X(t)$ for an exponential distribution.

Relevant properties for exponential distributions: $f(x) = e^{-x/\theta}/\theta$
$E(X) = \theta$
$Var(X) = \theta^2$.

Solution S4C6-1. We use the formula $E(X) = {}_{-\infty}{}^{\infty}\int x*f(x)dx$. Since an exponential distribution's lower bound is 0, we find $E(e^{tX})$ as follows: $E(e^{tX}) = {}_0{}^{\infty}\int e^{tx}*f(x)dx = {}_0{}^{\infty}\int(e^{tx}*e^{-x/\theta}/\theta)dx =$ ${}_0{}^{\infty}\int(e^{(t-1/\theta)x}/\theta)dx = (1/((t-1/\theta)*\theta))e^{(t-1/\theta)x}\Big|_0{}^{\infty} = (1/(\theta t-1))e^{(t-1/\theta)x}\Big|_0{}^{\infty} = 0 - (1/(\theta t-1)) =$

$M_X(t) = 1/(1 - \theta t)$.

Problem S4C6-2. Find the probability generating function $P_X(z)$ for an exponential distribution.
Relevant properties for exponential distributions: $f(x) = e^{-x/\theta}/\theta$; $E(X) = \theta$; $Var(X) = \theta^2$.

Solution S4C6-2. We use the formula $P_X(z) = M_X(\ln(z))$. From Solution S4C6-2, we know that $M_X(t) = 1/(1 - \theta t)$. Thus, $M_X(\ln(z)) =$ **$P_X(z) = 1/(1 - \theta\ln(z))$.**

Problem S4C6-3. The random variable X follows an exponential distribution with mean 45. The random variable Y follows an exponential distribution with mean 50. The random variable Z follows an exponential distribution with mean 99. Find the moment generating function $M_S(t)$ of $S = X + Y + Z$.

Solution S4C6-3. We use the formula $M_{S\_k}(t) = {}_{j=1}{}^{k}\Pi(M_{X\_j}(t))$. Here, $M_S(t) = M_X(t)*M_Y(t)*M_Z(t)$. We know that every exponential distribution with mean θ has an mgf of $1/(1 - \theta t)$. Therefore, $M_X(t) = 1/(1 - 45t)$, $M_Y(t) = 1/(1 - 50t)$, and $M_Z(t) = 1/(1 - 99t)$. Thus, $M_S(t) = (1/(1 - 45t))*(1/(1 - 50t))*(1/(1 - 99t)) = M_S(t) = 1/((1 - 45t)*(1 - 50t)*(1 - 99t))$.

Problem S4C6-4. The sum of α independent exponential random variables, each with mean θ, is a gamma random variable with parameters α and θ. What is the probability generating function $P_Y(z)$ of a gamma random variable Y with these parameters?

Solution S4C6-4. We use the formula $P_{S\_k}(z) = {}_{j=1}{}^{k}\Pi(P_{X\_j}(z))$. From Solution S4C6-2, we know that $P_X(z) = 1/(1 - \theta\ln(z))$ for any exponential random variable X with mean θ. Thus, for a random variable Y that is the sum of α exponential random variables X, $P_Y(z) = (1/(1 - \theta\ln(z)))^{\alpha}$ = **$P_Y(z) = (1 - \theta\ln(z))^{-\alpha}$**.

Problem S4C6-5. The fraction of a whole pie that Mr. Zolatrax eats during a meal is a random variable X with the following probability density function: $f(x) = 3x^2/8$ for $0 \le x \le 2$ and 0 otherwise. What is the moment generating function $M_X(t)$ of X?

Solution S4C6-5. We find $E(e^{tX})$ as follows: $E(e^{tX}) = {}_0{}^2\int e^{tx}*f(x)dx = {}_0{}^2\int e^{tx}*(3x^2/8)dx$.

We use the Tabular Method of Integration by Parts, setting $u = 3x^2/8$ and $dv = e^{tx}$.

Then we have the following table:

....u..............dv
$+3x^2/8$ .... e^{tx}
$-3x/4$ .... $(1/t)e^{tx}$
$+3/4$............$(1/t^2)e^{tx}$
$- 0$.............$(1/t^3)e^{tx}$

Thus, ${}_0{}^2\int e^{tx}*(3x^2/8) = ((3x^2/8t)e^{tx} - (3x/4t^2)e^{tx} + (3/4t^3)e^{tx})) \vert_0{}^2 =$
$((3*4/8t)e^{2t} - (3*3/4t^2)e^{2t} + (3/4t^3)e^{2t})) - (3/4t^3) =$ **$M_X(t) = (3/2t)e^{2t} - (9/4t^2)e^{2t} + (3/4t^3)e^{2t} - (3/4t^3)$**.

Section 7

Hazard Rates and Light and Heavy Tails of Distributions

The part of a distribution which corresponds to large values of a random variable is called the **right tail** (sometimes simply called the tail) of that distribution.

Generally, for a distribution with a **light right tail,** all positive moments exist. For a distribution with a **heavy right tail**, positive moments only exist up to a certain moment or do not exist at all.

For example, the gamma distribution has a light right tail because all positive moments of gamma random variables exist. The Pareto distribution with parameters α and θ has a heavy right tail, because the only kth moments that exist for this distribution are moments such that $k < \alpha$.

One can also make a *comparison* between two distributions to see which one has a heavy tail relative to the other. Let $S_1(x)$ and $S_2(x)$ be the survival functions of X for two different distributions, 1 and 2, and let $f_1(x)$ and $f_2(x)$ be the corresponding probability density functions. **If $\lim_{x\to\infty}(S_1(x)/S_2(x))$ increases without bound, then Distribution 1 has a heavy tail relative to Distribution 2**. Equivalently, **if $\lim_{x\to\infty}(-f_1(x)/-f_2(x))$ increases without bound, then Distribution 1 has a heavy tail relative to Distribution 2.**

A useful property to remember when making the comparisons above is **L'Hôpital's Rule**, which states the following: If **$g_1(x)/g_2(x)$ is of the form 0/0 or** ∞/∞**, then $\lim_{x\to\infty}(g_1(x)/g_2(x)) = \lim_{x\to\infty}(g_1'(x)/g_2'(x))$**.

The hazard rate of a distribution is **$h(x) = f(x)/S(x)$**. If a distribution has a decreasing hazard rate, then that distribution has a heavy tail. A decreasing function need not be *strictly* decreasing. If a distribution has an increasing hazard rate, then that distribution has a light tail. An increasing function need not be *strictly* increasing, and, under this definition, a constant hazard function can be called both "increasing" and "decreasing."

When comparing two increasing distributions, distribution A has a lighter tail than distribution B, if the hazard function of A is increasing faster than the hazard function of B.

When comparing two decreasing distributions, distribution C has a heavier tail than distribution D, if the hazard function of C is decreasing faster than the hazard function of D.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 3, pp. 34-37.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C7-1. Does the exponential distribution have a light right tail or a heavy right tail? Justify your answer. Refer to the Exam 4 / C Tables (p. 10 of the file).

Solution S4C7-1. The kth moment of an exponential distribution with mean θ is $E(X^k) = \theta^k * k!$. This moment exists for any integer value of k. This, the exponential distribution has a **light right tail**.

Problem S4C7-2. Consider two distributions:
A: An exponential distribution with mean θ and survival function $S_1(x) = e^{-x/\theta}$ for $x \geq 0$.

B: A distribution with survival function $S_2(x) = 1/x$ for $x \geq 1$ and 0 otherwise.

Assume that both distributions have the same mean. Does either of these distributions have a heavy right tail relative to the other? If so, which one has the heavy right tail?

Solution S4C7-2. We determine $\lim_{x\to\infty}(S_1(x)/S_2(x)) = \lim_{x\to\infty}(e^{-x/\theta}/(1/x)) = \lim_{x\to\infty}(xe^{-x/\theta}/1) = \lim_{x\to\infty}(xe^{-x/\theta}) = \lim_{x\to\infty}(x/e^{x/\theta})$. By L'Hôpital's Rule, $\lim_{x\to\infty}(g_1(x)/g_2(x)) = \lim_{x\to\infty}(g_1'(x)/g_2'(x))$, so $\lim_{x\to\infty}(x/e^{x/\theta}) = \lim_{x\to\infty}(1/(1/\theta)e^{x/\theta})$. Since $(1/\theta)e^{x/\theta}$ increases without bound as x increases, it means that $\lim_{x\to\infty}(1/(1/\theta)e^{x/\theta}) = \lim_{x\to\infty}(S_1(x)/S_2(x)) = 0$. Correspondingly, $\lim_{x\to\infty}(S_2(x)/S_1(x)) = \lim_{x\to\infty}((1/\theta)e^{x/\theta})$ increases without bound. Therefore, **Distribution 2 has a heavy right tail relative to Distribution 1**.

Problem S4C7-3. The survival function of a distribution of X is $S(x) = 1/x^2$ for $x \geq 1$ and 0 otherwise. Does the distribution have a light right tail or a heavy right tail? Answer the question by looking at the hazard function of the distribution.

Solution S4C7-3. Since $S(x) = 1/x^2$, it follows that $F(x) = 1 - 1/x^2$. Then $f(x) = F'(x) = 2/x^3$.

$h(x) = f(x)/S(x) = (2/x^3)/(1/x^2) = h(x) = 2/x$ for $x \geq 1$.

$h'(x) = -2/x^2$ for $x \geq 1$, which means that $h'(x) < 0$ for $x \geq 1$, and so the hazard function is decreasing, which means **the distribution has a heavy right tail**.

Problem S4C7-4. Consider two distributions:

C: An exponential distribution with mean 10.

D: An exponential distribution with mean 20.

Which of these has a lighter right tail relative to the other?

Solution S4C7-4. For an exponential distribution, $S(x) = e^{-x/\theta}$, and $f(x) = (e^{-x/\theta}/\theta)$, and so $h(x) = f(x)/S(x) = (e^{-x/\theta}/\theta)/e^{-x/\theta} = h(x) = 1/\theta$. Note that for any value of θ, 1/θ is constant, so neither distribution is increasing at a faster rate than the other. Therefore, **neither of these distributions**

has a lighter right tail relative to the other. They both have light tails, but neither distribution's tail is light*er*.

Problem S4C7-5. Consider two distributions:

E: A Pareto distribution with $\alpha = 4$ and $\theta = 340$.

F: A Pareto distribution with $\alpha = 6$ and $\theta = 340$.

Which of these has a heavier right tail relative to the other? Perform *both* a hazard rate analysis and a survival function analysis and compare the results.

Relevant properties for Pareto distributions: $S(x) = \theta^\alpha/(x + \theta)^\alpha$.

$f(x) = \alpha\theta^\alpha/(x + \theta)^{\alpha+1}$.

Solution S4C7-5.

We compare the two distributions' survival functions:
$\lim_{x\to\infty}(S_E(x)/S_F(x)) = (340^4/(x + 340)^4)/(340^6/(x + 340)^6)$
$\lim_{x\to\infty}(S_E(x)/S_F(x)) = ((x + 340)^2)/340^2$**.**
$\lim_{x\to\infty}(S_E(x)/S_F(x))$ increases without bound, so, **by the survival function analysis, E has a heavier right tail** compared to F.

Now we perform the hazard rate analysis:

For a Pareto distribution, $h(x) = f(x)/S(x) = (\alpha\theta^\alpha/(x + \theta)^{\alpha+1})/(\theta^\alpha/(x + \theta)^\alpha) =$

$\alpha/(x + \theta)$. Since $h'(x) = -\alpha/(x + \theta)^2 < 0$, the Pareto distribution has a heavy right tail. *How fast* h(x) decreases is a function of the value of α; the larger α is, the faster h(x) decreases. Thus, the distribution with the larger value of α has the heavier tail. Hence, **F has a heavier right tail by the hazard rate analysis**.

It appears that the two tests do not always give the same results!

Section 8

The Equilibrium Distribution, Mean Excess Loss Functions, and Tails of Distributions

Another method can be used to compare distributions with heavy tails to those with light tails.

A distribution has a heavy right tail if its mean excess loss function $e_X(d)$ is increasing in d. A distribution has a heavier right tail than another if its mean excess loss function is increasing at a *lower* rate.

A distribution has a light right tail if its mean excess loss function $e_X(d)$ is decreasing in d.

If a distribution's hazard rate function h(x) is increasing, then its mean excess loss function $e_X(d)$ will be decreasing, but the converse is not necessarily the case. Likewise, if a distribution's hazard rate function h(x) is decreasing, then its mean excess loss function $e_X(d)$ will be increasing, but the converse is not necessarily the case.

Moreover, the following formula is useful: $\mathbf{\lim_{d\to\infty}(e_X(d)) = \lim_{d\to\infty}(1/h(d)) = \lim_{d\to\infty}(S(d)/f(d))}$.

The **equilibrium distribution** or **integrated tail distribution** associated with some distribution of the random variable X has the following properties:

Probability density function: $\mathbf{f_e(x) = S(x)/E(X), x \geq 0}$.

Survival function: $\mathbf{S_e(x) = {}_x{}^{\infty}\!\int S(t)dt/E(X), x \geq 0}$.

Hazard rate function: $\mathbf{h_e(x) = f_e(x)/S_e(x) = 1/e_X(x)}$.

Source:

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 3, pp. 37-39.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C8-1. The survival function of a distribution of X is $S(x) = 1/(x+1)^2$ for $x \geq 0$ and 0 otherwise. What is the probability density function of the associated equilibrium distribution?

Solution S4C8-1. We use the formula $f_e(x) = S(x)/E(X)$. We want to find E(X). Since $S(x) = 1/(x+1)^2$, it follows that $F(x) = 1 - 1/(x+1)^2$. Then $f(x) = F'(x) = 2/(x+1)^3$. Then $E(X) = {}_0{}^{\infty}\!\int x*(2/(x+1)^3)dx = {}_1$. Hence, $f_e(x) = S(x)/E(X) = (1/(x+1)^2)/1 =$ $\mathbf{f_e(x) = 1/(x+1)^2}$**, for** $\mathbf{x \geq 0}$.

Problem S4C8-2. The survival function of a distribution of X is $S(x) = 1/(x+1)^2$ for $x \geq 0$ and 0 otherwise. What is the survival function of the associated equilibrium distribution?

Solution S4C8-2. We use the formula $S_e(x) = {}_x{}^{\infty}\int S(t)dt/E(X)$.We know that $S(t) = 1/(t+1)^2$ for $t \geq 1$ and, from Solution S4C8-1, that $E(X) = 1$. Thus, ${}_x{}^{\infty}\int S(t)dt = {}_x{}^{\infty}\int(1/(t+1)^2)dt = (-1/(t+1))\Big|_x{}^{\infty} = 1/(x+1)$. Hence, $\mathbf{S_e(x) = 1/(x+1)}$**, for** $\mathbf{x \geq 0}$.

Problem S4C8-3. The survival function of a distribution of X is $S(x) = 1/(x+1)^2$ for $x \geq 0$ and 0 otherwise. What is the hazard rate function of the associated equilibrium distribution?

Solution S4C8-3. We use the formula $h_e(x) = f_e(x)/S_e(x)$. From Solution S4C8-1, $f_e(x) = 1/(x+1)^2$. From Solution S4C8-2, $S_e(x) = 1/(x+1)$. Thus, $h_e(x) = (1/(x+1)^2)/(1/(x+1)) =$ $\mathbf{h_e(x) = 1/(x+1)}$**, for** $\mathbf{x \geq 0}$**.**

Problem S4C8-4. The survival function of a distribution of X is $S(x) = 1/(x+1)^2$ for $x \geq 0$ and 0 otherwise. What is the mean excess loss function $e_X(x)$ of this distribution? What can this function tell us with regard to whether the distribution has a heavy right tail or a light right tail?

Solution S4C8-4. We use the formula $h_e(x) = 1/e_X(d)$. Thus, $e_X(x) = 1/h_e(x)$. From Solution S4C8-3, $h_e(x) = 1/(x+1)$, so $\mathbf{e_X(x) = x+1}$ **for** $\mathbf{x \geq 1}$**.** A distribution has a heavy right tail if its mean excess loss function $e_X(d)$ is increasing in d. This **mean excess loss function is increasing** for all values of x, which means that **the distribution has a heavy right tail**. **Problem S4C8-5.** Random variable X follows a Pareto distribution with $\alpha = 4$ and $\theta = 340$. What is the limit of the mean excess loss function of X, $\lim_{d\to\infty}(e_X(d))$, as d increases without bound? What does this tell us about the lightness or heaviness of the tail of this distribution?

Relevant properties for Pareto distributions: $S(x) = \theta^\alpha/(x + \theta)^\alpha$.

$f(x) = \alpha\theta^\alpha/(x + \theta)^{\alpha+1}$.

Solution S4C8-5. We use the formula $\lim_{d\to\infty}(e_X(d)) = \lim_{d\to\infty}(1/h(d)) = \lim_{d\to\infty}(S(d)/f(d)) =$

$\lim_{d\to\infty}((\theta^\alpha/(d + \theta)^\alpha)/(\alpha\theta^\alpha/(d + \theta)^{\alpha+1})) = \lim_{d\to\infty}(1/(\alpha/(d + \theta)) = \lim_{d\to\infty}((d + \theta)/\alpha) = \infty$. This means that this **distribution has a heavy right tail, since** $\mathbf{e_X(d)}$ **increases** without bound in d.

Section 9

Risk Measures and Coherence

For random loss variables X and Y, a **coherent risk measure** $\rho(X)$ has the following properties:
1. Subadditivity: $\rho(X+Y) \le \rho(X) + \rho(Y)$.
2. Monotonicity: If $X \le Y$ for all possible outcomes, then $\rho(X) \le \rho(Y)$.
3. Positive homogeneity: For any positive constant c, $\rho(cX) = c\rho(X)$.
4. Translation invariance: For any positive constant c, $\rho(X + c) = \rho(X) + c$.

Tail-Value-at-Risk is an example of a coherent risk measure. Value-at-Risk, however, is not coherent because it fails to meet the subadditivity criterion.

Source: *Loss Models: From Data to Decisions*, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 3, pp. 42-48.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C9-1. The following risk measure is proposed for the random loss variables X and Y, which can each only be positive: $q(X) = x^2$ and $q(Y) = y^2$. Is q a coherent risk measure? If it is, show that it verifies all the properties of a coherent risk measure. If it is not, which properties does the measure not fulfill?

Solution S4C9-1. We check to see whether all four properties of a coherent risk measure hold:
1. Subadditivity: $q(X+Y) = (x+y)^2 = x^2 + 2xy + y^2$. $q(X) + q(Y) = x^2 + y^2$. If x and y are both positive, then $x^2 + 2xy + y^2 > x^2 + y^2$, and so **the subadditivity condition is not met.**
2. Monotonicity: Let $X \le Y$ for all possible outcomes. Since X and Y can only assume positive values, it *is* indeed the case that $q(X) \le q(Y)$, since for $X \le Y$, $x^2 \le y^2$. So the Monotonicity condition is met.
3. Positive homogeneity: For any positive constant c, $q(cX) = (cx)^2 = c^2x^2 \neq cx^2$ for any $c \neq 1$. Thus, for most c, $q(cX) \neq cq(X)$ and **the positive homogeneity condition is not met.**
4. Translation invariance: For any positive constant c, $q(X + c) = (x + c)^2 = x^2 + 2cx + c^2 \neq x^2 + c$ for most x and c. Thus, for most x and c, $q(X + c) \neq q(X) + c$, and so **the translation invariance condition is not met.** Clearly, failing to meet three of the four required conditions, **$q(X) = x^2$ is not a coherent risk measure**.

Problem S4C9-2. Which of the following is an implication of the subadditivity requirement for a coherent risk measure? More than one answer may be correct.

(a) If the subadditivity requirement is met, then diversification of risks reduces one's overall exposure.
(b) If the subadditivity requirement is met, then diversification of risks increases one's overall exposure.
(c) If the subadditivity requirement is met, then diversification of risks does not affect one's

overall exposure.
(d) If the subadditivity requirement is met, then a company holding both risk X and risk Y should split up into two separate companies holding each risk individually.
(e) If the subadditivity requirement is met, then a company holding both risk X and risk Y should remain an integrated entity, as doing so is safer than holding each risk individually would be.
(f) If the subadditivity requirement is met, then the executives of a company holding both risk X and risk Y should shrug and give up in disgust, because the situation is hopeless and utterly beyond the possibility of rescue.

Solution S4C9-2. If the subadditivity requirement is met, then $\rho(X+Y) \leq \rho(X) + \rho(Y)$. This means that the risk measure of the sum of two random variables is less than the sum of the risk measures of the two variables. Thus, holding the two or more risks together is safer than holding them separately. Hence, the following answers are correct:
(a) If the subadditivity requirement is met, then diversification of risks reduces one's overall exposure.
(e) If the subadditivity requirement is met, then a company holding both risk X and risk Y should remain an integrated entity, as doing so is safer than holding each risk individually would be.

Problem S4C9-3. Which of the following is an implication of the monotonicity requirement for a coherent risk measure? More than one answer may be correct.

(a) Even if one risk always has greater losses than another, it can sometimes result in a lower risk measure.
(b) If one risk always has fewer losses than another, its risk measure will always be greater.
(c) If one risk always has fewer losses than another, its risk measure will always be smaller.
(d) The risk measures of different risks are entirely independent of whether one risk has greater losses than another.
(e) As the random variable representing the risk increases in magnitude, the risk measure must increase.
(f) As the random variable representing the risk increases in magnitude, the risk measure must decrease.

Solution S4C9-3. The purpose of the monotonicity requirement is to ensure that greater risks always have larger associated risk measures. Thus, the following answers are true:
(c) If one risk has fewer losses than another, its risk measure will always be smaller.
(e) As the random variable representing the risk increases in magnitude, the risk measure must increase.

Problem S4C9-4. Which of the following is an implication of the positive homogeneity requirement for a coherent risk measure? More than one answer may be correct.

(a) The risk measure of a risk in yen will be its risk measure in dollars multiplied by the exchange ratio of yen to dollars.
(b) The risk measure of a risk in yen may be different its risk measure in dollars multiplied by the exchange ratio of yen to dollars.
(c) One can reduce the riskiness of an endeavor by converting the funds used to a different

currency.
(d) One cannot reduce the riskiness of an endeavor by converting the funds used to a different currency.
(e) If one assumes twice the amount of risk formerly assumed, one will need twice the capital.
(f) If one assumes twice the amount of risk formerly assumed, one will need four times the capital.
(g) If one assumes twice the amount of risk formerly assumed, one will need one-half the capital.
(h) If one assumes twice the amount of risk formerly assumed, one will need $\sqrt{(2)}$ times the capital, because of the risk-reducing effects of diversification.
(i) If one assumes twice the amount of risk formerly assumed, one will need π times the capital.

Solution S4C9-4. The positive homogeneity requirement states that $\rho(cX) = c\rho(X)$. If this holds, then converting to a different currency should simply scale the risk measure by the exchange ratio and should not affect overall riskiness. Likewise, the risk measure is directly proportional to the amount of risk undertaken, so twice the risk necessitates twice the capital. Therefore, the following answers are correct:
(a) The risk measure of a risk in yen will be its risk measure in dollars multiplied by the exchange ratio of yen to dollars.
(d) One cannot reduce the riskiness of an endeavor by converting the funds used to a different currency.
(e) If one assumes twice the amount of risk formerly assumed, one will need twice the capital.

Problem S4C9-5. Which of the following is an implication of the translation invariance requirement for a coherent risk measure? More than one answer may be correct.

(a) Getting additional capital, even if it is from a risk-free source, can fundamentally alter the riskiness of an endeavor.
(b) Getting additional capital, if it is from a risk-free source, cannot fundamentally alter the riskiness of an endeavor.
(c) If there is no uncertainty associated with a particular source of funds, then one does not need to devote additional capital to addressing the risk.
(d) If there is no uncertainty associated with a particular source of funds, then one still needs to devote to it an amount of additional capital proportional to the risk measure.
(e) The riskiness of an endeavor can be reduced by introducing a different, certain quantity.
(f) The riskiness of an endeavor can be increased by introducing a different, certain quantity.
(g) The riskiness of an endeavor is unaffected by introducing a different, certain quantity.

Solution S4C9-5. The translation invariance requirement states that $\rho(X + c) = \rho(X) + c$. This means that if c were an additional amount of funds (or of anything else) of which one could be certain, then the risk measure would not change it at all; it stays as c. Therefore, the following answers are correct:
(b) Getting additional capital, if it is from a risk-free source, cannot fundamentally alter the riskiness of an endeavor.
(c) If there is no uncertainty associated with a particular source of funds, then one does not need to devote additional capital to addressing the risk.
(g) The riskiness of an endeavor is unaffected by introducing a different, certain quantity.

Section 10

Parametric Distributions and Scale Parameters

A **parametric distribution** "is a set of distribution functions, each member of which is determined by specifying one or more values called **parameters.** The number of parameters is fixed and finite" (Klugman, Panjer, and Wilmot 2008, p. 54).

Examples of parametric distributions include the normal distribution with parameters μ and σ^2, the Pareto distribution with parameters α and θ, and the exponential distribution with parameter θ.

A parametric distribution is a **scale distribution** "if, when a random variable from that set of distributions is multiplied by a positive constant, the resulting random variable is also in that set of distributions" (Klugman, Panjer, and Wilmot 2008, p. 54).

A **scale parameter** is a parameter of a scale distribution that meets the following two conditions:

1. If a random variable X following a scale distribution is multiplied by a constant c, then the scale parameter of the distribution of cX is also multiplied by c.

2. If a random variable X following a scale distribution is multiplied by a constant c, then *only* the scale parameter changes.

For instance, the exponential distribution has a scale parameter of θ, while the Gamma distribution with parameters α and θ has a scale parameter of θ. Likewise, the Pareto distribution with parameters α and θ has a scale parameter of θ.

Source:

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 4, pp. 53-55.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C10-1. The amount of money in dollars that Laticfer received in 2010 from his investment in strontium futures follows a Pareto distribution with $\alpha = 3$ and $\theta = 1900$. Annual inflation in the United States from 2010 to 2011 is 140%. If Laticfer's investment income keeps up with inflation but is otherwise unaffected, what is the probability that Laticfer will receive more than \$2000 in 2011?

Relevant properties for Pareto distributions: $S(x) = \theta^\alpha/(x + \theta)^\alpha$.

Solution S4C10-1. The Pareto distribution is a scale distribution with a scale parameter of θ. Thus, inflation of 140% implies that θ in the Pareto distribution for 2010 gets multiplied by 2.4 to get the corresponding parameter for the Pareto distribution in 2011. Hence, Laticfer's earnings in 2011 follow a Pareto distribution with α = 3 and θ = 1900*2.4 = 4560. We want to find $S(2000) = \theta^\alpha/(2000 + \theta)^\alpha = 4560^3/(2000 + 4560)^3$ = **S(2000) = 0.3358791225.**

Problem S4C10-2. Rodres has two investments – one in tulips and the other in daffodils. The annual income X from the tulip investment follows an exponential distribution with θ = 500. The annual income Y from the daffodil investment is equal to 3X, but what happens to the tulip investment is independent from what happens to the daffodil investment. Find the probability that Rodres will get more than 700 from his investment in tulips and more than 3400 from his investment in daffodils this year.

Relevant properties for exponential distributions: $S(x) = e^{-x/\theta}$

Solution S4C10-2. We want to find Pr(X > 700 and Y > 3400) = Pr(X > 700)*Pr(Y > 3400) = $S_X(700)*S_Y(3400)$, since X and Y are independent. We know that $\theta_X = 500$, and since Y = 3X and the exponential distribution is a scale distribution, $\theta_Y = 1500$.

$S_X(700) = e^{-700/500} = 0.2465969639$.

$S_Y(3400) = e^{-3400/1500} = 0.1036571286$

Thus, the desired probability is 0.2465969639*0.1036571286 = **0.0255615332**.

Problem S4C10-3. The number of years a yellow-spotted snail lives (X) follows a Weibull distribution with parameters τ = 5 and θ = 120. Green-spotted snail lifespans (Y) are typically twice as long as yellow-spotted snail lifespans. The lifespans of the snails are independent. Find the probability that a green-spotted snail will live to be at least 300 years old.

Relevant properties for Weibull distributions:
$S(x) = \exp(-(x/\theta)^\tau)$

Solution S4C10-3. A Weibull distribution is a scale distribution with parameter θ. We can see this as follows: Let X follow a Weibull distribution such that $S_X(x) = \exp(-(x/\theta)^\tau)$. Let Y = 2X. Then $S_Y(y) = \Pr(Y > y) = \Pr(2X > y) = \Pr(X > y/2) = S_X(y/2) = \exp(-(y/(2\theta))^\tau)$. Thus, the Weibull distribution for Y has parameters τ_X and $2\theta_X$. We want to find

$S_Y(300) = \exp(-(300/(2*120)^5) = S_Y(300) = \exp(-(5/4)^5)$ = **0.0472757494**.

Problem S4C10-4. The number of years a red-spotted snail lives (X) follows a Weibull distribution with parameters τ = 0.5 and θ = 120. Blue-spotted snail lifespans (Y) are typically twice as long as yellow-spotted snail lifespans. Ten red-spotted snails are born on the same day as a blue-spotted snail. The lifespans of the snails are independent. Find the probability that at least one of the ten red-spotted snails will live more than three times the expected lifespan of the blue-spotted snail.

Relevant properties for Weibull distributions:
$S(x) = \exp(-(x/\theta)^{\tau})$

$E(X) = \theta*\Gamma(1 + 1/\tau)$
Note that $\Gamma(r) = (r - 1)!$ whenever r is an integer.

Solution S4C10-4. Since the Weibull distribution is a scale distribution with scale parameter θ and Y = 2X, we know that Y follows a Weibull distribution with parameters τ = 0.5 and θ = 240. Thus, E(Y) = 240*Γ(1 + 1/0.5) = 240*Γ(3) = 240*2! = E(Y) = 480.

We want to find Pr(at least one red-spotted snail lives longer than 3*480 = 1440) =

1 - Pr(No red-spotted snail lives to be 1440.) = $1 - F_X(1440)^{10}$

$F_X(1440) = 1 - S_X(1440) = 1 - \exp(-(1440/120)^{0.5}) = 0.9686988868$.

Thus, the desired probability is $1 - 0.9686988868^{10}$ = **0.2724078474**.

Problem S4C10-5. Random variable X follows the distribution with survival function

$S_X(x) = \alpha/x$ for $x \geq \alpha$ and 0 otherwise. Is α a scale parameter of the distribution? Why or why not?

Solution S4C10-5. Since the distribution has only one parameter, we do not need to worry about the behavior of any others. We only need to check whether the distribution of Y = cX for some constant c has the parameter cα.

$S_Y(y) = \Pr(Y > y) = \Pr(cX > y) = \Pr(X > y/c) = S_X(y/c) = \alpha/(y/c) = c\alpha/y$. Thus, **the distribution of Y *does* indeed have the parameter cα, and so α is a scale parameter of the distribution**.

Note: All names of hypothetical persons used in this section have been randomly generated.

Section 11

K-Point Mixtures, Variable-Component Mixture Distributions, and Data-Dependent Distributions

A random variable Y is a **k-point mixture** of the random variables X_1, X_2, ..., X_k if its cumulative distribution function (cdf) is given by

$\mathbf{F_Y(y) = a_1*F_{X\_1}(y) + a_2*F_{X\_2}(y) + ... + a_k*F_{X\_k}(y)}$, where each of the $a_j > 0$ and

$a_1 + a_2 + ... + a_k = 1$. This formula works for when k is known.

If we do not know how many distributions should be in a mixture, then the value of k itself becomes a parameter in a **variable-component mixture distribution**, whose cdf can be written as follows: $\mathbf{F(x) = {}_{j=1}{}^{K}\Sigma(a_jF_j(x))}$, where ${}_{j=1}{}^{K}\Sigma(a_j) = 1$, each of the $a_j > 0$, K can be any positive integer, and j ranges from 1 to K.

In a variable-component mixture distribution, each of the parameters associated with each individual $F_j(x)$ is a parameter. Also, there are (K-1) parameters corresponding to the weights a_1 through a_{K-1}. The weight a_K is not itself a parameter, since the value of a_K is determined by the value of the constants a_1 through a_{K-1}.

A **data-dependent distribution** "is at least as complex as the data or knowledge that produced it, and the number of 'parameters' increases as the number of data points or the amount of knowledge increases" (Klugman, Panjer, and Willmot 2008, p. 59).

We will develop one data-dependent distribution in this section.

Source:

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 4, pp. 56-59.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C11-1. The number of kilograms corn stored in a silo (Y) is a 2-point mixture of two random variables, each of which follows a Pareto distribution. The first Pareto distribution has $\alpha = 3$ and $\theta = 900$ and accounts for 0.6 of the mixture. The second Pareto distribution has $\alpha = 5$ and $\theta = 1500$. What is the cdf of Y?

Relevant properties for Pareto distributions: $S(x) = \theta^\alpha/(x + \theta)^\alpha$.

Solution S4C11-1. We use the formula $F_Y(y) = a_1*F_{X\_1}(y) + a_2*F_{X\_2}(y) + ... + a_k*F_{X\_k}(y)$ for k =2.

Since for a Pareto distribution, $S(x) = \theta^\alpha/(x + \theta)^\alpha$, it follows that $F(x) = 1 - \theta^\alpha/(x + \theta)^\alpha$.

The first Pareto distribution in the mixture has $F_1(x) = 1 - 900^3/(x + 900)^3$. It accounts for 0.6 of the mixture, so it will appear as one of the components of the mixture in the following form: $0.6(1 - 900^3/(x + 900)^3)$.

The second Pareto distribution in the mixture has $F_2(x) = 1 - 1500^5/(x + 1500)^5$. It accounts for the remaining 0.4 of the mixture, so it will appear as one of the components of the mixture in the following form: $0.4(1 - 1500^5/(x + 1500)^5)$.

The 2-point mixture's cdf therefore is as follows:
$\mathbf{F_Y(y) = 0.6(1 - 900^3/(x + 900)^3) + 0.4(1 - 1500^5/(x + 1500)^5)}$.

$\mathbf{F_Y(y) = 1 - 0.6(900^3/(x + 900)^3) - 0.4(1500^5/(x + 1500)^5)}$.

Problem S4C11-2. The number of kilograms corn stored in a silo (Y) is a 2-point mixture of two random variables, each of which follows a Pareto distribution. The first Pareto distribution has $\alpha = 3$ and $\theta = 900$ and accounts for 0.6 of the mixture. The second Pareto distribution has $\alpha = 5$ and $\theta = 1500$. What is the probability that there are more than 1000 kilograms of corn stored?

Solution S4C11-2. We want to find $S_Y(1000)$.

In general, it is the case that $F_Y(y) = a_1*F_{X\_1}(y) + a_2*F_{X\_2}(y) + ... + a_k*F_{X\_k}(y)$ implies that $S_Y(y) = a_1*S_{X\_1}(y) + a_2*S_{X\_2}(y) + ... + a_k*S_{X\_k}(y)$:

Here, in particular, $S_Y(y) = 1 - 0.6(1 - 900^3/(x + 900)^3) + 0.4(1 - 1500^5/(x + 1500)^5) =$

$1 - 0.6*1 - 0.4*1 + 0.6(900^3/(x + 900)^3) + 0.4(1500^5/(x + 1500)^5) = 0.6(900^3/(x + 900)^3) + 0.4(1500^5/(x + 1500)^5)$.

Thus, $S_Y(1000) = 0.6(900^3/(1000 + 900)^3) + 0.4(1500^5/(1000 + 1500)^5) =$ $\mathbf{S_Y(1000) = 0.0948742289}$.

Problem S4C11-3. A Weibull distribution has two parameters: θ and τ. An actuary is creating variable-component mixture distribution consisting of K Weibull distributions. If the actuary chooses to use 17 Weibull distributions instead of 12, how many more parameters will the variable-component mixture distribution have as a result?

Solution S4C11-3. In a variable-component mixture distribution, each of the parameters associated with each individual $F_j(x)$ is a parameter. Also, there are (K-1) parameters corresponding to the weights a_1 through a_{K-1}.

Each individual Weibull distribution will have 2 parameters: θ_j and τ_j.

If K = 12, then there will also be 11 parameters corresponding to the weights associated with each Weibull distribution. Thus, if K = 12, there will be 12*2 + 11 = 35 parameters.

If K = 17, then there will be 16 parameters corresponding to the weights associated with each Weibull distribution. Thus, if K = 17, there will be 17*2 + 16 = 50 parameters.

The difference in the number of parameters between K = 17 and K = 12 is thus 50-35 = **15 additional parameters.**

Problem S4C11-4. The numbers of cockroach sightings have been reported for 8 different geographic areas this year: 60, 42, 50, 135, 340, 53, 100, 20. To model Y, the average number of cockroach sightings per area, a data-dependent distribution is being created using a sum of exponential distributions to accommodate each of these data points. The number of cockroach sightings for an area will be the mean of the corresponding exponential distribution. No single exponential distribution in the model will be weighted more heavily than any other. Give the survival function $S_Y(y)$ associated with this data-dependent distribution.

Relevant properties for exponential distributions: $S(x) = e^{-x/\theta}$

Solution S4C11-4. There are 8 distributions that will be incorporated into the model. Since each has an equal weight, that weight will be (1/8). For each distribution, the θ is the corresponding data point. Thus, we have

$\mathbf{S_Y(y) = (1/8)e^{-x/60} + (1/8)e^{-x/42} + (1/8)e^{-x/50} + (1/8)e^{-x/135} + (1/8)e^{-x/340} + (1/8)e^{-x/53} + (1/8)e^{-x/100} + (1/8)e^{-x/20}}$.

Problem S4C11-5. The numbers of cockroach sightings have been reported for 8 different geographic areas this year: 60, 42, 50, 135, 340, 53, 100, 20. To model Y, the average number of cockroach sightings per area, a data-dependent distribution is being created using a sum of exponential distributions, one such distribution corresponding to each data point. The number of cockroach sightings for an area will be the mean of the corresponding exponential distribution. No single exponential distribution in the model will be weighted more heavily than any other. What is the probability that the average number of cockroach sightings next year will be greater than the highest number of cockroaches observed in a single area this year?

Solution S4C11-5. The highest number of cockroaches observed in a single area this year is 340. Thus, we want to find $S_Y(340)$. From Solution S4C11-4, we know that $S_Y(y) = (1/8)e^{-x/60} + (1/8)e^{-x/42} + (1/8)e^{-x/50} + (1/8)e^{-x/135} + (1/8)e^{-x/340} + (1/8)e^{-x/53} + (1/8)e^{-x/100} + (1/8)e^{-x/20}$. Thus,

$S_Y(340) = (1/8)e^{-340/60} + (1/8)e^{-340/42} + (1/8)e^{-340/50} + (1/8)e^{-340/135} + (1/8)e^{-340/340} + (1/8)e^{-340/53} + (1/8)e^{-340/100} + (1/8)e^{-340/20}$ = $\mathbf{S_Y(340) = 0.0610433066}$.

Section 12

Exam-Style Questions on the Central Limit Theorem, Skewness, Percentile Matching, and Ogives

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C12-1. Similar to Question 255 of the Exam C Sample Questions from the Society of Actuaries. The population standard deviation of the amount of grams of corn stored in silos is 0.5 times the population mean. Use the central limit theorem to find the minimum number of silos you would have to investigate in order to be 99.6% confident that the mean amount of grams of corn observed is within 7% of the population mean. Refer to the Exam 4 / C Tables as necessary.

Solution S4C12-1. We recall the central limit theorem from Section 5. In particular, the following formula will be useful:
$\Pr(a < S_k < b) \approx \Phi((b - E(S_k))/\sqrt{(Var(S_k))}) - \Phi((a - E(S_k))/\sqrt{(Var(S_k))})$
Let x be the population mean; thus, the population standard deviation is 0.5x. We want to find n such that the following holds:
$\Pr((0.93nx - nx)/(\sqrt{(n)}*(0.5x)) < Z < (1.07nx - nx)/(\sqrt{(n)}*(0.5x))) = 0.996$
$\Pr(-0.07\sqrt{(n)})/0.5 < Z < 0.07\sqrt{(n)})/0.5) = 0.996$
$\Pr(-0.14\sqrt{(n)} < Z < 0.14\sqrt{(n)}) = 0.996$
$\Phi(0.14\sqrt{(n)}) - \Phi(-0.14\sqrt{(n)}) = 0.996$
$\Phi(0.14\sqrt{(n)}) = 1 - (1 - 0.996)/2$, since what the area to the right of $0.14\sqrt{(n)}$ will be the same as the area to the left of $-0.14\sqrt{(n)}$.
$\Phi(0.14\sqrt{(n)}) = 0.998$

Using the Normal Distribution Table, we find that $0.14\sqrt{(n)} = 2.88 \rightarrow \sqrt{(n)} = 20.57142857 \rightarrow n^2 = 423.1836735$. We are looking for the smallest possible sample size, which is **n = 424.**

Problem S4C12-2. Similar to Question 248 of the Exam C Sample Questions from the Society of Actuaries. The following is true about the masses of elephants from a sample of 12:

5 elephants weigh 800 kilograms; 4 elephants weigh 1000 kilograms; 3 elephants weigh 1400 kilograms. What is the empirical skewness coefficient for this sample?

Solution S4C12-2. We use the formula $\gamma_1 = \mu_3/\sigma^3$.

Let X be the random variable representing the mass of an elephant.

We know that Pr(X = 800) = 5/12, Pr(X = 1000) = 4/12 = 1/3, and Pr(X = 1400) = 3/12 = ¼.

First, we find E(X) = 800(5/12) + 1000(4/12) + 1400(3/12) = E(X) = 1016.666666667.

Next, we find Var(X) = μ_2 = $(800- 1016.666666)^2(5/12) + (1000-1016.666666)^2(4/12) + (1400-1016.666666)^2(3/12)$ = Var(X) = 56388.888889. We note that $\sigma^3 = \text{Var}(X)^{1.5}$.

Next, we find $\mu_3 = (800- 1016.666666)^3(5/12) + (1000-1016.666666)^3(4/12) + (1400-1016.666666)^3(3/12) = \mu_3 = 9842592.593$.

Thus, $\gamma_1 = \mu_3/\sigma^3 = 9842592.593/56388.888889^{1.5}$ = $\boldsymbol{\gamma_1}$ **= 0.735053929**.

Problem S4C12-3. Similar to Question 216 of the Exam C Sample Questions from the Society of Actuaries. A certain Burr distribution has parameters α = 1 and θ and γ unknown. You know the following about a random variable that follows a Burr distribution.

60% of values are greater than 50.

10% of values are greater than 9200.

Find the value of θ for this Burr distribution using this data. The method you will be using is known as **percentile matching.**

Relevant properties regarding the Burr distribution: $S(x) = 1/(1 + (x/\theta)^\gamma)^\alpha$

Solution S4C12-3. We need to set up a system of two equations and solve it for the unknown parameters θ and γ. Since α = 1, $S(x) = 1/(1 + (x/\theta)^\gamma)$.
We know that $S(50) = 1/(1 + (50/\theta)^\gamma) = 0.6$
Moreover, $S(9200) = 1/(1 + (9200/\theta)^\gamma) = 0.1$
Thus, $5/3 = (1 + (50/\theta)^\gamma) \rightarrow 2/3 = (50/\theta)^\gamma$.
Moreover, $10 = (1 + (9200/\theta)^\gamma) \rightarrow 9 = (9200/\theta)^\gamma$
Hence, $9/(2/3) = (9200/\theta)^\gamma/(50/\theta)^\gamma \rightarrow 13.5 = 184^\gamma \rightarrow \gamma = \ln(13.5)/\ln(184) = \gamma = 0.4990837484$.
Then $2/3 = (50/\theta)^{0.4990837484} \rightarrow (2/3)^{1/0.4990837484} = 50/\theta \rightarrow \theta = 50/(2/3)^{1/0.4990837484} \rightarrow$
θ = 112.6676103.

Problem S4C12-4. Similar to Question 211 of the Exam C Sample Questions from the Society of Actuaries. There are two probability distributions:
A is an exponential distribution with mean θ = 7.

B is a distribution that is uniform on the interval from 0 to 8, and thereafter proportional to A. The distribution's probability density function (pdf) is everywhere continuous.

Relevant properties for exponential distributions: $S(x) = e^{-x/\theta}$; $f(x) = e^{-x/\theta}/\theta$.

Find the probability that a random variable that follows distribution B will be between 0 and 8.

Solution S4C12-4. We can define the probability density function (pdf) of B as a system of two equations with two unknowns:
$f(x) = c$ for $0 \leq x \leq 8$;

$f(x) = ke^{-x/7}/7$ for $x > 8$.

At $x = 8$, because the pdf is continuous, $f(8) = ke^{-8/7}/7 = c$.

We also know that the area underneath a pdf must equal 1. Thus, $1 = 8c + {}_8{}^{\infty}\int(ke^{-x/7}/7)dx \rightarrow$

$1 = 8c + -ke^{-x/7} \mid_8^{\infty} \rightarrow 1 = 8c + ke^{-8/7}$. But $c = ke^{-8/7}/7$, so $ke^{-8/7} = 7c$, and therefore $1 = 8c + 7c = 15c$, and so $c = 1/15$. Over the interval $0 \leq x \leq 8$, the area underneath the pdf is $8c =$ **8/15 = 0.5333333333**.

Problem S4C12-5. Similar to Question 195 of the [Exam C Sample Questions] from the Society of Actuaries. An **ogive** is a cumulative frequency curve that can be drawn using data from a table of frequencies and/or cumulative frequencies. For instance, we can start with the following frequency table:

Range of X........Number of observations in this range
0 - 10.....................12
10 - 20.....................30
20 - 40.....................15
40 - 90.....................18

The following cumulative frequencies can be ascertained from this data:

Value of X........Cumulative frequency for observations at or below this value
10.........................12
20.........................42
40........................57
90........................75

When creating an ogive, draw the value of the variable on the horizontal axis and the cumulative frequency for observations at or below this value on the vertical axis. Then plot the points from a cumulative frequency chart like the one above and connect them with straight lines. You can use linear interpolation between any two known points of an ogive.

Now you can answer a question regarding the frequency table and cumulative frequency table above. Using the ogive, what is $\Pr(15 \leq X \leq 84)$?

Solution S4C12-5. You can plot an ogive, which will visually illustrate the following reasoning. 15 is halfway between 10 and 20, and there are 30 observations for the interval [10, 20]. The ogive would thus show 15 observations for the interval [15, 20].

84 is (84 - 40)/(90 - 40) = 0.88 of the way between 40 and 90, and there are 18 observations for the interval [40, 90]. The ogive would thus show 0.88*18 = 15.84 observations for the interval [40, 84]. There are 75 observations in all.

Thus, $\Pr(15 \leq X \leq 84)$ = (Number of observations for [15, 20] + Number of observations for [20, 40] + Number of observations for [40, 84])/75 = (15 + 15 + 15.84)/75 = **$\Pr(15 \leq X \leq 84)$ = 0.6112**.

Section 13

Distributions of Multiples and Powers of Random Variables

What happens to a random variable's distribution when that variable is multiplied by a constant?

The following theorem is true (Klugman, Panjer, and Willmot 2008, p. 62):

Theorem 13.1. "Let X be a continuous random variable with pdf $f_X(x)$ and cdf $F_X(x)$. Let $Y = \theta X$ with $\theta > 0$. Then $F_Y(y) = F_X(y/\theta)$ and $f_Y(y) = (1/\theta)f_X(y/\theta)$." This implies that θ is a scale parameter for the random variable Y.

What happens to a random variable's distribution when that variable is raised to a power?

The following theorem is true (Klugman, Panjer, and Willmot 2008, p. 62):

Theorem 13.2. "Let X be a continuous random variable with pdf $f_X(x)$ and cdf $F_X(x)$ with $F_X(0) = 0$. Let $Y = X^{1/\tau}$. Then, if $\tau > 0$, $F_Y(y) = F_X(y^\tau)$, $f_Y(y) = \tau y^{\tau-1}f_X(y^\tau)$, $y > 0$.

If $\tau < 0$, $F_Y(y) = 1 - F_X(y^\tau)$, $f_Y(y) = -\tau y^{\tau-1}f_X(y^\tau)$."

If a distribution is raised to a power and $\tau > 0$, the resulting distribution is called **transformed.** If $\tau = -1$, the resulting distribution is called **inverse.** If $\tau < 0$ and $\tau \neq 1$, the resulting distribution is called **inverse transformed.**

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 5, pp. 61-63.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C13-1. The fraction of a whole pie that Mr. Zolatrax eats during a meal is a random variable X with the following probability density function: $f_X(x) = 3x^2/8$ for $0 \leq x \leq 2$ and 0 otherwise. Mr. Zolatrax decides to go on a diet and eat only half as many pies per meal as he does presently. Let Y be the new fraction of a whole pie that Mr. Zolatrax eats during a meal. Find $f_Y(y)$.

Solution S4C13-1. We are given that Y = (1/2)X. Thus, by Theorem 13.1, for θ = ½, $f_Y(y) = (1/\theta)f_X(y/\theta) = 2f_X(2y) = 2*3(2y)^2/8 = 24y^2/8 =$ **$f_Y(y) = 3y^2$ for $0 \leq y \leq 1$ and 0 otherwise**.

Problem S4C13-2. The lifetime of a seven-legged snail in years is exponentially distributed with the following survival function: $S_X(x) = e^{-0.034x}$ for $x \geq 0$. A sapient seven-legged snail scientist

(SSSSS) discovers a way to increase all seven-legged snail lifespans threefold. The new random variable for seven-legged snail lifespans is Y. Find $S_Y(y)$.

Solution S4C13-2. From Theorem 13.1, $F_Y(y) = F_X(y/\theta)$ for $\theta = 3$. (Note: This θ is *not* the mean of the exponential distribution!) Since $S_Y(y) = 1 - F_Y(y)$, $S_Y(y) = 1 - F_X(y/\theta) = S_X(y/\theta) = e^{-0.034y/3}$ = $\mathbf{S_Y(y) = e^{-0.01133333333333y}}$ **for $\mathbf{y \geq 0}$**.

Problem S4C13-3. The fraction of a whole pie that Mr. Zolatrax eats during a meal is a random variable X with the following probability density function: $f_X(x) = 3x^2/8$ for $0 \leq x \leq 2$ and 0 otherwise. Mr. Zolatrax finds out, to his regret, that his former diet was insufficient for his purposes. Therefore, he decides on a more intense diet, in which he will raise the number of pies he originally ate (X) to the power of (1/4). Let Z be the new fraction of a whole pie that Mr. Zolatrax eats during a meal. Find $f_Z(z)$.

Solution S4C13-3. We are given that $Z = X^{1/4}$. Thus, by Theorem 13.2, for $\tau = 4$, $f_Z(z)$ =

$\tau z^{\tau-1} f_X(z^\tau) = 4z^3 f_X(z^4) = 4z^3 * 3(z^4)^2/8$ = $\mathbf{f_Z(z) = 1.5z^{11}}$**, for $\mathbf{0 \leq z \leq 2^{1/4}}$ and 0 otherwise**.

Problem S4C13-4. After the successes of a heroic SSSSS, the lifetime of a seven-legged snail in years is exponentially distributed with the following survival function: $S_Y(y) = e^{-0.01133333333333y}$ for $y \geq 0$. Now a second SSSSS comes along and discovers a way to raise all seven-legged snail lifespans to the power of 2. The new random variable for seven-legged snail lifespans is X. Find $S_Z(z)$.

Solution S4C13-4. From Theorem 13.2, $F_Z(z) = F_Y(z^\tau)$, so $S_Z(z) = 1 - F_Z(z) = 1 - F_Y(z^\tau) = S_Y(z^\tau)$, for $\tau = ½$. Thus, $\mathbf{S_Z(z) = \exp(-0.0113333333333z^{1/2})}$ **for $\mathbf{z \geq 0}$**.

Problem S4C13-5. Let the random variable X follow a probability distribution for which $F_X(0) = 0$. Indicate whether each of the following random variables follows a transformed, inverse, or inverse transformed distribution of X:

$Y = X^2$
$Z = X^{-1}$
$A = X^{-0.25}$
$B = X^{-4}$
$C = X^{0.25}$

Solution S4C13-5.
For each variable, the τ is the reciprocal of the power of X.
For Y, $\tau = 0.5 > 0$, so **the distribution of Y is transformed**.
For Z, $\tau = -1$, so **the distribution of Z is inverse**.
For A, $\tau = -4 < 0$ and is not -1, so **the distribution of A is inverse transformed**.
For B, $\tau = -0.25 < 0$ and is not -1, so **the distribution of B is inverse transformed**.
For C, $\tau = 4 > 0$, so **the distribution of C is transformed**.

Section 14

The Gamma Function, Incomplete Gamma Function, and Exponentiation of Random Variables

The **gamma function** with parameter α > 0 is defined as follows:

$\mathbf{\Gamma(\alpha) = {}_0^\infty\!\int t^{\alpha-1} * e^{-t} * dt}$.

There are two useful properties of the gamma function:
$\mathbf{\Gamma(\alpha) = (\alpha-1)*\Gamma(\alpha-1)}$

$\mathbf{\Gamma(n) = (n-1)!}$ if n is a positive integer.

In MS Excel, you can get Γ(α) for any α you specify via the following input:
"=EXP(GAMMALN(α))"

The **incomplete gamma function** with parameter α > 0 is defined as follows:
$\mathbf{\Gamma(\alpha; x) = (1/\Gamma(\alpha))({}_0^x\!\int t^{\alpha-1} * e^{-t} * dt)}$.

In MS Excel, you can get Γ(α; x)for any α and x you specify via the following input:
"=GAMMADIST(x, α, 1, TRUE)"

What happens to a random variable's distribution when that variable is exponentiated?

The following theorem is true (Klugman, Panjer, and Willmot 2008, p. 64):

Theorem 14.1. "Let X be a continuous random variable with pdf $f_X(x)$ and cdf $F_X(x)$ with $f_X(x)$ 0 for all real x. Let Y = exp(X). Them, for y > 0, $F_Y(y) = F_X(\ln(y))$, $f_Y(y) = (1/y)f_X(\ln(y))$."

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 5, pp. 63-64.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C14-1. Find Γ(1.4). Set up the integral and use any calculator or program to solve it.

Solution S4C14-1. We use the formula $\Gamma(\alpha) = {}_0^\infty\!\int t^{\alpha-1} * e^{-t} * dt = \mathbf{{}_0^\infty\!\int t^{0.4} * e^{-t} * dt = \Gamma(1.4) =}$ **0.887263817.**

Problem S4C14-2. Find Γ(3.4) without doing any additional integration. Refer to Solution S4C14-1 for necessary information.

Solution S4C14-2. We use the formula $\Gamma(\alpha) = (\alpha-1)*\Gamma(\alpha-1)$. Thus, $\Gamma(3.4) = (2.4)*\Gamma(2.4) = (2.4)*(1.4)*\Gamma(1.4) = (2.4)*(1.4)*0.887263817 =$ **$\Gamma(3.4) = 2.981206425$**.

Problem S4C14-3. Find Γ(5; 10).

Solution S4C14-3. We use the following formula: $\Gamma(\alpha; x) = (1/\Gamma(\alpha))({}_0{}^x\int t^{\alpha-1}*e^{-t}*dt)$.

Since $\alpha = 5$, $\Gamma(\alpha) = \Gamma(5) = 4! = 24$. Thus, $\Gamma(5; 10) = (1/24)({}_0{}^{10}\int t^4*e^{-t}*dt) =$ **$\Gamma(5; 10) = 0.9707473119$**.

Problem S4C14-4. The lifetime of a seven-legged snail in years is exponentially distributed with the following survival function: $S_Z(z) = \exp(-0.0113333333333z^{1/2})$ for $z \geq 0$. Yet another sapient seven-legged snail scientist (SSSSS) discovers a way to increase all seven-legged snail lifespans by raising e to the power of each current lifespan. That is, the new lifespan $A = e^Z$. Find $S_A(a)$.

Solution S4C14-4. $S_A(a) = 1 - F_A(a) = 1 - F_Z(\ln(a)) = S_Z(\ln(a)) =$

$S_A(a) = \exp(-0.0113333333333*\ln(a)^{1/2})$.

Problem S4C14-5. The fraction of a whole pie that Mr. Zolatrax eats during a meal is a random variable X with the following probability density function: $f_X(x) = 3x^2/8$ for $0 \leq x \leq 2$ and 0 otherwise. After his previous pie-related diets failed, Mr. Zolatrax decides, in desperation, to give up and increase his per-meal pie consumption by raising e to the power of the current number of pies he eats per meal. That is, the new amount he eats per meal is $A = e^X$. Find $f_A(a)$.

Solution S4C14-5. We use the formula $f_A(a) = (1/a)f_X(\ln(a)) = (1/a)*3(\ln(a))^2/8 =$ **$f_A(a) = 3(\ln(a))^2/(8a)$, for $0 \leq x \leq e^2$ and 0 otherwise**.

Section 15

Mixture Distributions and Some Review Questions

Theorem 15.1. "Let X have pdf $f_{X|\Lambda}(x|\lambda)$ and cdf $F_{X|\Lambda}(x|\lambda)$, where λ is a parameter of X... Let λ be a realization of the variable Λ with pdf $f_\Lambda(\lambda)$. Then the unconditional pdf of X is $f_X(x) = \int f_{X|\Lambda}(x|\lambda)*f_\Lambda(\lambda)d\lambda$, where the integral is taken over all values of λ with positive probability. The resulting distribution is a **mixture distribution**, for which $F_X(x) = \int F_{X|\Lambda}(x|\lambda)*f_\Lambda(\lambda)d\lambda$."

The following is true about moments of the mixture distribution:

$\mathbf{E(X^k) = E(E(X^k|\Lambda))}$ and $\mathbf{Var(X) = E(Var(X|\Lambda)) + Var(E(X|\Lambda))}$.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 5, pp. 64-65.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C15-1. Let the random variable $X|\Lambda$ have a uniform distribution with parameter 1/Λ. (That is, $f_{X|\Lambda}(x|\lambda) = 1/(1/\lambda) = \lambda$ for $0 \leq x \leq 1/\lambda$ and 0 otherwise). Λ follows an exponential distribution with mean θ. Determine the pdf of the unconditional distribution of X.

Solution S4C15-1.

We are given that $f_{X|\Lambda}(x|\lambda) = \lambda$ and that $f_\Lambda(\lambda) = e^{-\lambda/\theta}/\theta$. Thus, we use the formula $f_X(x) = \int f_{X|\Lambda}(x|\lambda)*f_\Lambda(\lambda)d\lambda = {}_0^{1/x}\int(\lambda/\theta)e^{-\lambda/\theta}d\lambda$. The upper bound of the integral is 1/x, because if $0 \leq x \leq 1/\lambda$, then $0 \leq \lambda \leq 1/x$, and so $f_{X|\Lambda}(x|\lambda)$ is equal to λ only on the interval $0 \leq \lambda \leq 1/x$. To solve this integral, we use the tabular method of integration by parts:

......u...........dv
+.... λ........... $(1/\theta)e^{-\lambda/\theta}$
-.....1...........$-e^{-\lambda/\theta}$
+....0...........$\theta e^{-\lambda/\theta}$

Thus, ${}_0^{1/x}\int(\lambda/\theta)e^{-\lambda/\theta}d\lambda = (-\lambda e^{-\lambda/\theta} - \theta e^{-\lambda/\theta})|_0^{1/x} = \mathbf{f_X(x) = -(1/x)e^{-(1/(x\theta))} - \theta e^{-(1/(x\theta))} + \theta}$.

Problem S4C15-2. Review of Sections 13 and 14: Let $Y = e^{2X} + 5$. Find $F_Y(y)$, expressed in terms of $F_X(g(y))$, where g(y) is some function solely of the variable y.

Solution S4C15-2. $Y = e^{2X} + 5$ implies that $F_Y(y) = \Pr(Y \leq y) = \Pr(e^{2X} + 5 \leq y) = \Pr(e^{2X} \leq y-5) =$

Pr(2X ≤ ln(y-5)) = Pr(X ≤ ln(y-5)/2) = **F_Y(y) = F_X(ln(y-5)/2)**.

Problem S4C15-3. Let the random variable Y│Λ have a Pareto distribution with parameters α = 3 and θ = Λ. Λ follows an exponential distribution with mean 100. Determine E(Y).

Relevant properties for exponential distributions: S(x) = $e^{-x/\theta}$; f(x) = $e^{-x/\theta}/\theta$; E(X) = θ when X follows an exponential distribution.

Relevant properties for Pareto distributions: S(x) = $\theta^\alpha/(x + \theta)^\alpha$; E(X) = θ/(α - 1)

Solution S4C15-3. We use the formula E(Y^k) = E(E(Y^k│Λ)) for k = 1. We find E(Y│Λ), which is the mean of a Pareto distribution with parameters α = 3 and θ = Λ. That is, E(Y│Λ) = Λ/(3-1) = Λ/2.

Now E(Λ/2) = (1/2)E(Λ) = (1/2)*100 = **E(Y) = 50**.

Problem S4C15-4. Let the random variable Y│Λ have a gamma distribution with parameters α = 4 and θ_{gamma} = Λ. Λ follows a Weibull distribution with $\theta_{Weibull}$ = 500 and τ = 0.125. Determine E(Y). *Hint*: Think about a possible scale parameter!

Relevant properties for Gamma distributions:
E(X) = αθ
Var(X) = $\alpha\theta^2$

Relevant properties for Weibull distributions:
S(x) = $\exp(-(x/\theta)^\tau)$
E(X) = θ*Γ(1 + 1/τ)
E(X^2) = θ^2*Γ(1 + 2/τ)

Solution S4C15-4. We use the formula E(Y^k) = E(E(Y^k│Λ)) for k = 1. We find E(Y│Λ), which is the mean of a gamma distribution with parameters α = 4 and θ_{gamma} = Λ. That is, E(Y│Λ) = 4Λ. Then E(Y) = E(E(Y│Λ)) = E(4Λ). For the Weibull distribution, $\theta_{Weibull}$ is a scale parameter, so if Λ follows a Weibull distribution with $\theta_{Weibull}$ = 500 and τ = 0.125, then 4Λ follows a Weibull distribution with $\theta_{Weibull}$ = 2000 and τ = 0.125. Thus, E(4Λ) = 2000*Γ(1 + 1/0.125) = 2000*Γ(9) = 2000*8! = **E(Y) = 80,640,000**.

Problem S4C15-5. Let the random variable Y│Λ have a gamma distribution with parameters α = 7 and θ_{gamma} = Λ. Λ follows a Weibull distribution with $\theta_{Weibull}$ = 24 and τ = 1/3. Determine Var(Y). *Hint*: Think about a possible scale parameter!

Relevant properties for Gamma distributions:
E(X) = αθ
Var(X) = $\alpha\theta^2$

Relevant properties for Weibull distributions:
S(x) = $\exp(-(x/\theta)^\tau)$

$E(X) = \theta*\Gamma(1 + 1/\tau)$
$E(X^2) = \theta^2*\Gamma(1 + 2/\tau)$

Solution S4C15-5. We use the formula $Var(Y) = E(Var(Y | \Lambda)) + Var(E(Y | \Lambda))$.

We find $E(Y | \Lambda) = 7\Lambda$.

We find $Var(E(Y | \Lambda)) = Var(7\Lambda)$ = the variance of a Weibull distribution with parameters $\theta_{Weibull} = 7*24 = 168$ and $\tau = 1/3$. This is so because $\theta_{Weibull}$ is the scale parameter in a Weibull distribution. $E((7\Lambda)^2) = 168^2*\Gamma(1 + 2/(1/3)) = 28224*\Gamma(1 + 2/(1/3)) = 28224*\Gamma(7) = 28224*6! = 20321280$. Also, $E((7\Lambda))^2 = (168*\Gamma(1 + 1/(1/3)))^2 = (168*\Gamma(4))^2 = (168*3!)^2 = 1016064$.

Thus, $Var(E(Y | \Lambda)) = Var(7\Lambda) = E((7\Lambda)^2) - E((7\Lambda))^2 = 20321280 - 1016064 = Var(E(Y | \Lambda)) = 19305216$.

Now we find $Var(Y | \Lambda) = 7\Lambda^2$. We find $E(Var(Y | \Lambda)) = E(7\Lambda^2) = 7E(\Lambda^2) = 7*24^2*\Gamma(1 + 2/(1/3)) = 4032*6! = E(Var(Y | \Lambda)) = 2903040$.

Hence, $Var(Y) = E(Var(Y | \Lambda)) + Var(E(Y | \Lambda)) = 2903040 + 19305216 =$ **Var(Y) = 22,208,256**.

Section 16

Mixture Distributions Involving Discrete Random Variables

In this section, we will practice further with mixture distributions, this time involving some discrete random variables. The following theorem is useful:

Theorem 16.1. "Let X have pdf $f_{X|\Lambda}(x|\lambda)$ and cdf $F_{X|\Lambda}(x|\lambda)$, where λ is a parameter of X... Let λ be a realization of the variable Λ with a *discrete probability mass function* $f_\Lambda(\lambda)$. Then the unconditional pdf of X is $f_X(x) = \Sigma f_{X|\Lambda}(x|\lambda)*f_\Lambda(\lambda)$, where the *sum* is taken over all values of λ with positive probability. The resulting distribution is a **mixture distribution**, for which $F_X(x) = \Sigma F_{X|\Lambda}(x|\lambda)*f_\Lambda(\lambda)d\lambda$."

Note that Theorem 16.1 only applies when Λ is the discrete random variable, not when X|Λ is discrete but Λ itself is continuous. If X|Λ is discrete but Λ itself is continuous, you should still integrate.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 5, pp. 64-67.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C16-1. Our first practice with mixture distributions involving discrete distributions will concern a **two-point mixture distribution.** This distribution of a random variable N has a probability of 0.34 of being a Poisson distribution with parameter λ = 0.3, and a probability of 0.66 of being a binomial distribution with parameters m = 5 and q = 0.3. Find the probability that N = 3. **Relevant properties for Poisson distributions:** $Pr(N = k) = e^{-\lambda}*\lambda^k/k!$

Relevant properties for binomial distributions: $Pr(N = k) = C(m, k)*q^k*(1-q)^{m-k}$

Solution S4C16-1. Pr(N = 3) = 0.34*(Pr(N = 3 and the distribution is Poisson)) + 0.66* Pr(N = 3 and the distribution is binomial)) = $0.34*e^{-0.3}*0.3^3/3! + 0.66*C(5, 3)*0.3^3*(0.7)^{5-3}$ = $0.0011334519 + 0.66*10*0.3^3*0.7^2$ = 0.0011334519 + 0.087318 = **Pr(N = 3) = 0.0884514519**.

Problem S4C16-2. Another two-point mixture distribution of a random variable X has a probability p of being a geometric distribution with parameter β = 0.44 and a probability (1-p) of being a Poisson distribution with parameter λ = 5. It is known that Pr(X = 2) = 0.0823. Find p.

Relevant properties for Poisson distributions: $Pr(X = k) = e^{-\lambda}*\lambda^k/k!$

Relevant properties for geometric distributions: $Pr(X = k) = \beta^k/(1+\beta)^{k+1}$

Solution S4C16-2. Pr(X = 2) = p*(Pr(X = 2 and the distribution is geometric)) + (1-p)*(Pr(X = 2 and the distribution is Poisson)) = $p*0.44^2/(1.44)^3 + (1-p)*e^{-5}*5^2/2!$ = 0.0648362483p + 0.0842243375(1-p) = 0.0842243375 - 0.0193880892p. We are given that Pr(X = 2) = 0.0823. Therefore, 0.0823 = 0.0842243375 - 0.0193880892p → -0.0019243375 = -0.0193880892p → **p = 0.0992535916**.

Problem S4C16-3. (X | Λ) follows a Poisson distribution with parameter Λ. Λ follows a uniform distribution over the interval (1, 5). Find Pr(X = 3). After you set up any integral, you may use a calculator to evaluate it if this is possible.

Relevant properties for Poisson distributions: Pr(X = k) = $e^{-\lambda}*\lambda^k/k!$

Solution S4C16-3. Since Λ follows a continuous distribution, we will need to use an integral here. In particular, we refer to Theorem 15.1: $f_X(x) = \int f_{X|\Lambda}(x|\lambda)*f_\Lambda(\lambda)d\lambda$.

We know that $f_\Lambda(\lambda)$ = ¼ over the interval (1, 5).

Moreover, $f_{X|\Lambda}(x|\lambda) = e^{-\lambda}*\lambda^k/k!$ for x = k.

Thus, $f_X(3) = {}_1{}^5\int(e^{-\lambda}*\lambda^3/6)*(1/4)d\lambda = {}_1{}^5\int(e^{-\lambda}*\lambda^3/24)d\lambda$ = $\mathbf{f_X(3) = 0.178996482}$.

Problem S4C16-4. (X | Λ) follows a geometric distribution with parameter β = Λ. The distribution of Λ can be determined by creating an empirical model from the following observed values of Λ: 0.56, 0.56, 0.54, 0.46, 0.36, 0.36, 0.36. Find Pr(X = 1).

Relevant properties for geometric distributions: Pr(X = k) = $\beta^k/(1+\beta)^{k+1}$

Solution S4C16-4. Here, Λ follows a discrete distribution, so we will need to refer to the following equation from Theorem 16.1: $f_X(x) = \Sigma f_{X|\Lambda}(x|\lambda)*f_\Lambda(\lambda)$. Fortunately (and by design), we will need to sum only over a finite number of values. We can develop our empirical model for $f_\Lambda(\lambda)$ to see why this is so:

$f_\Lambda(0.36)$ = 3/7, since 3 of the 7 data points are 0.36.
$f_\Lambda(0.46)$ = 1/7, since 1 of the 7 data points is 0.46.
$f_\Lambda(0.54)$ = 1/7, since 1 of the 7 data points is 0.54.
$f_\Lambda(0.56)$ = 2/7, since 2 of the 7 data points are 0.56.

Thus, $f_X(x) = (3/7)*f_{X|\Lambda}(x|0.36) + (1/7)*f_{X|\Lambda}(x|0.46) + (1/7)*f_{X|\Lambda}(x|0.54) + (2/7)*f_{X|\Lambda}(x|0.56)$.

For geometric distributions with parameter β, Pr(X = 1) = $\beta/(1+\beta)^2$, so

$f_X(1) = (3/7)*0.36/(1.36)^2 + (1/7)*0.46/(1.46)^2 + (1/7)*0.54/(1.54)^2 + (2/7)*0.56/(1.56)^2$ =

Pr(X = 1) = 0.2125183276.

Problem S4C16-5. $(X \mid \Lambda)$ follows a geometric distribution with parameter $\beta = \Lambda$. The distribution of Λ can be determined by creating an empirical model from the following observed values of Λ: 0.56, 0.56, 0.54, 0.46, 0.36, 0.36, 0.36. Find $\Pr(X > 2)$.

Relevant properties for geometric distributions: $\Pr(X = k) = \beta^k/(1+\beta)^{k+1}$

Solution S4C16-5. X follows a discrete distribution, so

$\Pr(X > 2) = 1 - \Pr(X = 2) - \Pr(X = 1) - \Pr(X = 0)$.

From Solution S4C16-4, we know that $f_X(x) = (3/7)*f_{X\mid\Lambda}(x \mid 0.36) + (1/7)*f_{X\mid\Lambda}(x \mid 0.46) + (1/7)*f_{X\mid\Lambda}(x \mid 0.54) + (2/7)*f_{X\mid\Lambda}(x \mid 0.56)$. We also know that $\Pr(X = 1) = 0.2125183276$.

We need to find $\Pr(X = 0)$ and $\Pr(X = 2)$.

For geometric distributions with parameter β, $\Pr(X = 0) = 1/(1+\beta)$, so

$\Pr(X = 0) = (3/7)*(1/1.36) + (1/7)*(1/1.46) + (1/7)*(1/1.54) + (2/7)*(1/1.56) = \Pr(X = 0) = 0.6888879702$.

For geometric distributions with parameter β, $\Pr(X = 2) = \beta^2/(1+\beta)^3$, so

$\Pr(X = 2) = (3/7)*0.36^2/(1.36)^3 + (1/7)*0.46^2/(1.46)^3 + (1/7)*0.54^2/(1.54)^3 + (2/7)*0.56^2/(1.56)^3 = \Pr(X = 2) = 0.0668008064$.

Thus, $\Pr(X > 2) = 1 - 0.0668008064 - 0.2125183276 - 0.6888879702 =$ **$\Pr(X > 2) =$ 0.0317928958**.

Section 17

Frailty Models

In this section, we will discuss the components of a **frailty model.**

A frailty model consists some primary random variable X and a **frailty random variable** $\Lambda > 0$. The **conditional hazard rate** of X given that $\Lambda = \lambda$ is denoted $h_{X|\Lambda}(x \mid \lambda)$ and is defined as follows: $\mathbf{h_{X|\Lambda}(x \mid \lambda) = \lambda * a(x)}$, where a(x) is some function of x that will be specified in any particular instance. If you are using an exponential mixture in the frailty model, a(x) will be 1, and A(x) will be x. If you are using a Weibull mixture in the frailty model, a(x) will be $\gamma x^{\gamma-1}$, and A(x) will be x^γ.

The conditional survival function of X, $S_{X|\Lambda}(x \mid \lambda)$,can be found as follows:

$\mathbf{S_{X|\Lambda}(x \mid \lambda) = exp(-\lambda A(x))}$**, where** $\mathbf{A(x) = {}_0^x\int a(t)dt}$.

Alternatively, $\mathbf{S_{X|\Lambda}(x \mid \lambda) = exp(-{}_0^x\int h_{X|\Lambda}(t \mid \lambda)dt)}$.

The survival function of X, $S_X(X)$, can be found as follows:
$\mathbf{S_X(x) = E(exp(-A(x))) = M_\Lambda(-A(x)))}$.

The value of the $M_\Lambda(t)$function depends on the kind of frailty random variable selected. Some common frailty random variables include gamma and inverse Gaussian random variables. Useful properties of these random variables are described in the Exam 4 / C Tables.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 5, p. 68.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C17-1. A Weibull mixture is used in a frailty model with frailty random variable Λ, such that $a(x) = \gamma x^{\gamma-1}$, and, correspondingly, $A(x) = x^\gamma$. Find $S_{X|\Lambda}(x \mid \lambda)$ for this model.

Solution S4C17-1. We use the formula $S_{X|\Lambda}(x \mid \lambda) = exp(-\lambda A(x)) =$ $\mathbf{S_{X|\Lambda}(x \mid \lambda) = exp(-\lambda x^\gamma)}$.

Problem S4C17-2. A Weibull mixture is used in a frailty model with frailty random variable Λ, such that $a(x) = \gamma x^{\gamma-1}$, and, correspondingly, $A(x) = x^\gamma$. Here, $\gamma = 3$. The frailty random variable itself is a gamma random variable with parameters $\alpha = 5$ and $\theta = 90$. Find $S_X(x)$ for this model.

Useful property for gamma distributions: $M_X(t) = (1-\theta t)^{-\alpha}$.

Solution S4C17-2. We use the formula $S_X(x) = M_\Lambda(-A(x))) = M_\Lambda(-x^\gamma) = M_\Lambda(-x^3)$.

Since $M_\Lambda(t) = (1-\theta t)^{-\alpha} = (1-90t)^{-5}$, it follows that $M_\Lambda(-x^3) =$ $\mathbf{S_X(x) = (1+90x^3)^{-5}}$.

Problem S4C17-3. Now we have another frailty model with $S_X(x) = (1-90x^3)^{-5}$. However, this time, an exponential mixture is used, so a(x) = 1 and A(x) = x. What must $M_\Lambda(t)$ be in order for the above survival function to be the case?

Solution S4C17-3. We use the formula $S_X(x) = M_\Lambda(-A(x)))$. We know that $M_\Lambda(-A(x))) = M_\Lambda(-x) = (1-90x^3)^{-5}$, so if x is substituted for -x, all that will change is the sign within the parentheses: $M_\Lambda(x) = (1+90x^3)^{-5}$. Therefore, $\mathbf{M_\Lambda(x) = (1+90t^3)^{-5}}$.

Problem S4C17-4. A frailty model exists such that $S_X(x) = 2*(M_\Lambda(x))$. The frailty random variable is a gamma random variable with $\alpha = 2$ and $\theta = 12$. What must a(x) be for this model?

Useful property for gamma distributions: $M_X(t) = (1-\theta t)^{-\alpha}$.

Solution S4C17-4. We know that $S_X(x) = M_\Lambda(-A(x)))$. We are also given that $S_X(x) = 2*(M_\Lambda(x))$. Thus, $M_\Lambda(-A(x))) = 2*(M_\Lambda(x))$.

We can find $M_\Lambda(x) = (1-\theta x)^{-\alpha} = (1-12x)^{-2}$, so $M_\Lambda(-A(x))) = 2(1-12x)^{-2}$.

Moreover, $M_\Lambda(-A(x))) = (1+\theta A(x))^{-\alpha} = (1+12A(x))^{-2}$.
Thus, $(1+12A(x))^{-2} = 2(1-12x)^{-2} \rightarrow$
$(1+12A(x))^{-2} = (1/\sqrt{(2)})^{-2}(1-12x)^{-2} \rightarrow$
$1+12A(x) = (1/\sqrt{(2)})(1-12x) \rightarrow$
$\sqrt{(2)} + 12\sqrt{(2)}A(x) = (1-12x) \rightarrow$
$12\sqrt{(2)}A(x) = 1- \sqrt{(2)} - 12x \rightarrow$
$A(x) = (1- \sqrt{(2)} - 12x)/(12\sqrt{(2)}) \rightarrow$
$A(x) = (\sqrt{(2)} - 2 - 12\sqrt{(2)}x)/24$.
$a(x) = A'(x) = -12\sqrt{(2)}/24 =$ $\mathbf{a(x) = -\sqrt{(2)}/2}$.

Problem S4C17-5. A frailty model exists such that $S_X(x) = \ln((M_\Lambda(x)))$, and $M_\Lambda(x) = (1-100x)^{-1}$. Find a(x) for this model.

Solution S4C17-5. We know that $S_X(x) = M_\Lambda(-A(x))) = \ln((M_\Lambda(x))) = \ln((1-100x)^{-1})$.
Thus, $(1+100A(x))^{-1} = \ln((1-100x)^{-1}) \rightarrow$
$1+100A(x) = (\ln((1-100x)^{-1}))^{-1} \rightarrow$
$1+100A(x) = (-\ln(1-100x))^{-1} \rightarrow$
$100A(x) = (-\ln(1-100x))^{-1} - 1 \rightarrow$
$A(x) = (-\ln(1-100x))^{-1}/100 - 1/100 \rightarrow$
$A'(x) = a(x) = -(-\ln(1-100x))^{-2}/100*(-\ln(1-100x))' \rightarrow$
$a(x) = -(\ln(1-100x))^{-2}/100*(100/(1-100x))'$
$\mathbf{a(x) = -(\ln(1-100x))^{-2}/(1-100x)}$.

Section 18

Spliced Distributions

"A **k-component spliced distribution** has a density function that can be expressed as follows:

$f_X(x) = a_1 f_1(x),\ c_0 < x < c_1;$
$f_X(x) = a_2 f_2(x),\ c_1 < x < c_2;$
...
$f_X(x) = a_k f_k(x),\ c_{k-1} < x < c_k.$

For j = 1, ..., k, each $a_j > 0$ and each $f_j(x)$ must be a legitimate density function with all probability on the interval (c_{j-1}, c_j). Also, $a_1 + ... + a_k = 1$" (Klugman , Panjer, and Willmot 2008, p. 69).

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 5, pp. 69-70.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C18-1. A spliced distribution consists of a uniform distribution on the interval $0 \leq x < 10$ and another uniform distribution on the interval $10 \leq x < 40$. It is known that F(10) = 0.5. Find $f_X(x)$, the pdf of this spliced distribution.

Solution S4C18-1. We use the formula

$f_X(x) = a_1 f_1(x),\ c_0 < x < c_1;$
$f_X(x) = a_2 f_2(x),\ c_1 < x < c_2.$

The fact that F(10) = 0.5 indicates to us that the first uniform distribution's pdf needs to be multiplied by 0.5, as it applies to only the first 50% of possible values. Thus, $a_1 = 0.5$.

Since $a_1 + a_2 = 1$, it follows that $a_2 = 0.5$. Now we simply need to find the pdfs of the original specified uniform distributes and multiply them by the appropriate factors a_1 and a_2. Since the first uniform distribution is over the interval $0 \leq x < 10$, its pdf, $f_1(x)$, is 1/10. Since the first uniform distribution is over the interval $10 \leq x < 40$, its pdf, $f_2(x)$, is 1/30. Thus, the following is our formula for $f_X(x)$:

$\mathbf{f_X(x) = 1/20,\ 0 \leq x < 10;}$
$\mathbf{f_X(x) = 1/60,\ 10 \leq x < 40.}$

Problem S4C18-2. The survival function of an exponential distribution, $S_X(x) = e^{-x/\theta}$, and the corresponding probability density function (pdf), is $f_X(x) = e^{-x/\theta}/\theta$. Derive the formula for the pdf of an exponential distribution truncated on the right at c. (That is, an exponential distribution for which the entire domain of possible values is the interval from 0 to c.)

Solution S4C18-2. For a pdf, it must always be the case that ${}_a{}^b\int f_X(x)dx = 1$, for domain endpoints a and b. Since the regular exponential pdf is $e^{-x/\theta}/\theta$, we posit that, for the truncated distribution, the pdf is scaled by some factor k. We need to find k. We set up the following integral and equation:

${}_0{}^c\int k(e^{-x/\theta}/\theta)dx = 1 \rightarrow$
$-ke^{-x/\theta} \Big|_0^c = 1 \rightarrow$
$k-ke^{-c/\theta} = 1 \rightarrow$
$1 - e^{-c/\theta} = 1/k \rightarrow$
$k = 1/(1 - e^{-c/\theta})$.

Thus, for an exponential distribution truncated on the right at c, $\mathbf{f_X(x) = e^{-x/\theta}/((1 - e^{-c/\theta})*\theta)}$.

Formulas for pdfs of truncated distributions are not given in the Exam 4/C Tables, and so you may need to be able to derive them yourself on the exam in order to use certain truncated distributions as components of spliced distributions. This problem was intended to give you some practice with this skill.

Problem S4C18-3. The formula for the pdf of an exponential distribution truncated at some value c is $f_Y(y) = e^{-y/\theta}/((1 - e^{-c/\theta})*\theta)$. (We derived this formula in Problem S4C18-2.) A certain spliced distribution contains two components:

1. An exponential distribution on the interval $0 \leq x < 15$. This exponential distribution has mean $\theta = 5$.

2. A uniform distribution on the interval $15 \leq x \leq 95$.

It is known that $S_X(15) = 0.7$.

Find the pdf, $f_X(x)$, of this distribution.

Solution S4C18-3. We use the formula

$f_X(x) = a_1 f_1(x),\ c_0 < x < c_1$;

$f_X(x) = a_2 f_2(x),\ c_1 < x < c_2$.

From the given information, we can determine $f_1(x) = e^{-x/\theta}/((1 - e^{-c/\theta})*\theta) = e^{-x/5}/((1 - e^{-15/5})*5) =$

$f_1(x) = e^{-x/5}/((1 - e^{-3})*5)$.

We can also determine $f_2(x) = 1/(95-15) = f_2(x) = 1/80$.

Since S(15) = 0.7, 70% of all possible values occur over the domain of the uniform distribution. Thus, $a_2 = 0.7$, and, by implication, $a_1 = 1 - 0.7 = 0.3$.

Hence, we have

$\mathbf{f_X(x) = 3e^{-x/5}/(50(1 - e^{-3})), 0 \leq x < 15;}$

$\mathbf{f_X(x) = 7/800, 15 \leq x \leq 95}$.

Problem S4C18-4. The survival function of an exponential distribution, $S_X(x) = e^{-x/\theta}$, and the corresponding probability density function (pdf), is $f_X(x) = e^{-x/\theta}/\theta$. Derive the formula for the pdf of an exponential distribution truncated on the left at c. (That is, an exponential distribution for which the entire domain of possible values is the interval from c to ∞.)

Solution S4C18-4. For a pdf, it must always be the case that ${}_a{}^b\int f_X(x)dx = 1$, for domain endpoints a and b. Since the regular exponential pdf is $e^{-x/\theta}/\theta$, we posit that, for the truncated distribution, the pdf is scaled by some factor k. We need to find k. We set up the following integral and equation:

${}_c{}^\infty\int k(e^{-x/\theta}/\theta)dx = 1 \rightarrow$
$-ke^{-x/\theta}\Big|_c^\infty = 1 \rightarrow$
$ke^{-c/\theta} = 1 \rightarrow$
$k = e^{c/\theta}$.

Thus, for an exponential distribution truncated on the left at c, $f_X(x) = e^{c/\theta}e^{-x/\theta}/\theta =$

$\mathbf{f_X(x) = e^{(c-x)/\theta}/\theta}$.

Problem S4C18-5. A certain spliced distribution contains two components:
1. An exponential distribution on the interval $0 \leq x < 100$.

This exponential distribution has mean $\theta_1 = 50$.

2. An exponential distribution on the interval $100 \leq x$.

This exponential distribution has mean $\theta_2 = 200$.

It is known that $F_X(100) = 0.4$.

Find the pdf, $f_X(x)$, of this distribution.

Solution S4C18-5. We use the formula

$f_X(x) = a_1f_1(x), c_0 < x < c_1;$

$f_X(x) = a_2f_2(x)$, $c_1 < x < c_2$.

We use the formulas obtained from Solution S4C18-2 and Solution S4C18-4 regarding the pdfs of exponential distributions truncated on the right and on the left, respectively:

From the given information, we can determine $f_1(x) = e^{-x/\theta\_1}/((1 - e^{-c/\theta\_1})*\theta_1) =$

$f_1(x) = e^{-x/50}/((1 - e^{-100/50})*50) = f_1(x) = e^{-x/50}/((1 - e^{-2})*50)$.

From the given information, we can determine $f_2(x) = e^{(c-x)/\theta\_2}/\theta_2 = f_2(x) = e^{(100-x)/200}/200$.

Since $F_X(100) = 0.4$, we know that 0.4 of the possible values of x apply to the first distribution, so $a_1 = 0.4$, and, by implication, $a_2 = 1 - 0.4 = 0.6$. Hence, we have

$\mathbf{f_X(x) = e^{-x/50}/(125(1 - e^{-2})), 0 \leq x < 100;}$

$\mathbf{f_X(x) = 3e^{(100-x)/200}/1000, 100 \leq x}$.

Section 19

The Linear Exponential Family of Distributions

There exist special families of distributions, wherein the distributions are related to one another via their parameters, survival functions, pdfs, and cdfs. One such family is the **linear exponential family** of distributions, whose pdf is defined for a random variable X in terms of a parameter θ as follows:

$f_X(x; \theta) = p(x)e^{r(\theta)x}/q(\theta)$, where p(x) depends solely on x and not on θ, q(θ) is a constant, and r(θ) is called the **canonical parameter** of the distribution. Other parameters of the distribution (parameters that are not θ) may occur in the expressions p(x), q(θ), and r(θ). But they have no role in determining whether the distribution belongs to the linear exponential family.

If X follows a distribution of the linear exponential family, then **$E(X) = \mu(\theta) = q'(\theta)/(r'(\theta)*q(\theta))$** and **$Var(X) = v(\theta) = \mu'(\theta)/r'(\theta)$**.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 5, pp. 78-79.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C19-1. Does the Pareto distribution belong to the linear exponential family? Show why or why not. For a Pareto distribution with parameters α and θ, $f_X(x) = \alpha\theta^\alpha/(x+\theta)^{\alpha+1}$.

Solution S4C19-1. We need to show whether $f_X(x) = \alpha\theta^\alpha/(x+\theta)^{\alpha+1}$ can be transformed into an expression of the form $p(x)e^{r(\theta)x}/q(\theta)$. The problem arises with the $e^{r(\theta)x}$ component of the previous expression. The only occurrence of x in $f_X(x)$ is in the $(x+\theta)^{-(\alpha+1)}$ term. Creating some expression of the form $e^{r(\theta)x}$ out of that term would require us to do the following to it: $(x+\theta)^{-(\alpha+1)} = (\exp(\ln(x+\theta)^{-(\alpha+1)})) = (\exp((-\alpha+1)\ln(x)*\ln(\theta))) = (\exp((-\alpha+1)\ln(\theta)\ln(x)))$. But this still does not enable us to get an expression of the form $e^{r(\theta)x}$; it only enables us to get an expression of the form $e^{r(\theta)*\ln(x)}$.

Therefore, it is impossible to express $f_X(x)$ as a function of the form $p(x)e^{r(\theta)x}/q(\theta)$. Therefore, **the Pareto distribution does not belong to the linear exponential family.**

Problem S4C19-2. Does the Poisson distribution belong to the linear exponential family? Show why or why not. For a Poisson distribution with parameter λ, $Pr(X = k) = e^{-\lambda}*\lambda^k/k!$. Let that be the value of $f_X(k)$, a discrete probability mass function in this case.

Solution S4C19-2. We need to show whether $f_X(x) = e^{-\lambda}*\lambda^x/x!$ can be transformed into an expression of the form $p(x)e^{r(\theta)x}/q(\theta)$. x occurs in the expressions λ^x and x!. Transforming a factorial into an exponentiated expression in terms of x is difficult, so we will transform λ^x:

$\lambda^x = \exp(\ln(\lambda^x)) = \exp(x*\ln(\lambda))$, which is precisely an expression of the form $e^{r(\theta)x}$, where $\lambda = \theta$ and $r(\theta) = \ln(\theta)$. We can also let $p(x) = 1/x!$ and let $q(\theta) = e^\theta$, in which case $f_X(x) = (1/x!)e^{\ln(\theta)x}/e^\theta$ is the pdf of a Poisson distribution with c. This means that **the Poisson distribution belongs to the linear exponential family.**

Problem S4C19-3. A Poisson distribution has parameter $\lambda = 3$. For a Poisson distribution with parameter λ, $\Pr(X = k) = e^{-\lambda}*\lambda^k/k!$. What is the canonical parameter of this Poisson distribution?

Solution S4C19-3. In Solution S4C19-2, we determined that the Poisson distribution belongs to the linear exponential family, with probability mass function (pmf) $f_X(x) = (1/x!)e^{\ln(\theta)x}/e^\theta$ for $\lambda = \theta$, $p(x) = 1/x!$, $q(\theta) = e^\theta$, and $r(\theta) = \ln(\theta)$. The canonical parameter is $r(\theta) = \ln(\theta) = \ln(\lambda) =$ **ln(3) = 1.098612289**.

Problem S4C19-4. Let X follow a Poisson distribution with parameter λ, $\Pr(X = k) = e^{-\lambda}*\lambda^k/k!$. Derive the value of E(X) using solely the fact that the Poisson distribution belongs to the linear exponential family of distributions and the implications of this fact (including any formulas that apply).

Solution S4C19-4. We let $\lambda = \theta$. We refer to the formula $E(X) = \mu(\theta) = q'(\theta)/(r'(\theta)*q(\theta))$. We established in Solution S4C19-2 that for a Poisson distribution, $r(\theta) = \ln(\theta)$ and $q(\theta) = e^\theta$. Therefore, $q'(\theta)$ (differentiated with respect to θ) is e^θ, and $r'(\theta) = 1/\theta$. Thus, $q'(\theta)/(r'(\theta)*q(\theta)) =$

$e^\theta/(e^\theta*(1/\theta)) = \theta = \lambda$. We have just derived the fact that **E(X) = λ.**

Problem S4C19-5. Let X follow a Poisson distribution with parameter λ, $\Pr(X = k) = e^{-\lambda}*\lambda^k/k!$. Derive the value of Var(X) using solely the fact that the Poisson distribution belongs to the linear exponential family of distributions and the implications of this fact (including any formulas that apply).

Solution S4C19-5. We let $\lambda = \theta$. We refer to the formula $Var(X) = v(\theta) = \mu'(\theta)/r'(\theta)$. We know from Solution S4C19-4 that $\mu(\theta) = \theta$, so $\mu'(\theta) = 1$. We also know that $r'(\theta) = 1/\theta$. Therefore, $Var(X) = 1/(1/\theta) = \theta = \lambda$. We have just derived the fact that **Var(X) = λ.**

The results for both the mean and variance of the Poisson distribution were probably well known to you before these exercises. However, the purpose of these problems was to enable you to practice with another way of deriving these values, which can be applied to more complicated distributions of the linear exponential family as well.

Section 20

Properties of Poisson, Negative Binomial, and Geometric Distributions

This section addresses several extremely common discrete distributions, many of which we have already encountered in applications of other topics we have covered so far.

For the **Poisson distribution**, random variable N, and integer values of k:

The probability function or probability mass function (pf or pmf): $p_k = Pr(N = k) = e^{-\lambda}*\lambda^k/k!$

The probability generating function (pgf): $P(z) = e^{\lambda(z-1)}$, $\lambda > 0$.

$E(N) = \lambda$; $Var(N) = \lambda$.

Theorem 20.1. The sumof N independent Poisson random variables is a Poisson distribution whose parameter is the sum of the λ parameters of each individual Poisson random variable.

If $N_1, ..., N_n$ are n independent Poisson random variables with parameters $\lambda_1, ..., \lambda_n$, then

$N_1 + ... + N_n$ follows a Poisson distribution with parameter $\lambda_1 + ... + \lambda_n$.

The following theorem is also true (Klugman, Panjer, and Willmot 2008, p. 103):

Theorem 20.2. "Suppose that the number of events N is a Poisson random variable with mean λ. Further suppose that each event can be classified into one of m types with probabilities $p_1, ..., p_m$ independent of all other events. Then the number of events $N_1, ..., N_m$ corresponding to event types 1, ..., m, respectively, are mutually independent Poisson random variables with means λ_{p_1}, ..., λ_{p_m}, respectively."

For the **negative binomial distribution**, random variable N, and integer values of k:

The probability function or probability mass function (pf or pmf):

$p_k = Pr(N = k) = (r(r+1)...(r+k-1)/k!)*(\beta^k/(1+\beta)^{r+k})$, $k = 0, 1, 2, ..., r > 0, \beta > 0$.

The probability generating function (pgf): $P(z) = (1- \beta(z-1)^{-r})$.

$E(N) = r\beta$; $Var(N) = r\beta(1+\beta)$.

When r = 1, the negative binomial distribution becomes the **geometric distribution**. The geometric distribution, like the exponential distribution, has the **memoryless property**: the distribution of a variable that is known to be in excess of some value d does not depend on d.

Based on the pgfs of the Poisson and negative binomial distributions, it is possible to prove (Klugman, Panjer, and Willmot 2008, p. 107) that the Poisson distribution is a limiting case of the negative binomial distribution if $r\beta$ is taken to equal λ (i.e., the means of the two distributions are assumed to be the same). Then, $\lim_{r\to\infty}$(pgf of negative binomial distribution) = (pgf of Poisson distribution).

Because most of the pfs, pgfs, expected values, and variances for the special discrete distributions are provided in the Exam 4 / C Tables, we will focus in this section on the special properties of the distributions presented.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 6, pp. 101-107.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C20-1. Poisson random variable X has $\lambda_X = 4$.
Poisson random variable Y has $\lambda_Y = 5$.
Poisson random variable Z has $\lambda_Z = 0.6$.
Find the probability function (pf) p_k of A = X + Y + Z.

Solution S4C20-1. By Theorem 20.1, X + Y + Z follows a Poisson distribution with $\lambda = \lambda_X + \lambda_Y + \lambda_Z = 4 + 5 + 0.6 = 9.6$. Therefore, $p_k = \Pr(A = k) = e^{-\lambda}*\lambda^k/k! = \mathbf{p_k = e^{-9.6}*9.6^k/k!}$

Problem S4C20-2. The number of hovermobile crashes in a particular region per year follows a Poisson distribution with mean 40. There are three possible mutually exclusive causes of hovermobile crashes. Given that a crash has occurred, the following are the probabilities of each cause being the case:

Operator error: 0.15
Broken equipment: 0.45
Giant dead bug hitting windshield: 0.40

What is the probability that exactly 16 hovermobiles will crash this year due to giant dead bugs hitting their windshields?

Solution S4C20-2. We apply Theorem 20.2, which implies that the event of a hovermobile crashing due to a giant dead bug follows a Poisson distribution with mean $\lambda = 40*0.4 = 16$.

Thus, $p_{16} = e^{-16}*16^{16}/16! = \mathbf{p_{16} = 0.0992175316}$.

Problem S4C20-3. The number of hovermobile crashes in a particular region per year follows a Poisson distribution with mean 40. There are three possible mutually exclusive causes of hovermobile crashes. Given that a crash has occurred, the following are the probabilities of each cause being the case:

Operator error: 0.15
Broken equipment: 0.45
Giant dead bug hitting windshield: 0.40

What is the probability that exactly 35 hovermobiles will crash this year due to causes *other than* operator error?

Solution S4C20-3. We apply Theorem 20.2, which implies that the event of a hovermobile crashing due to an event other than operator error follows a Poisson distribution with mean λ = 40*(1-0.15) = 34. Thus, $p_{35} = e^{-34}*34^{35}/35! =$ $\mathbf{p_{35} = 0.0663005397}$.

Problem S4C20-4. You are modeling the frequency of events, and you need to select a distribution to use. You observe that the variance of the number of events is less than the mean number of events. Which of the following distributions should you use?

(a) Poisson
(b) Negative binomial
(c) Geometric
(d) None of the above are correct.

Solution S4C20-4. For the Poisson distribution, the variance is equal to the mean, so (a) is incorrect. For both the negative binomial distribution and its special r = 1 case, the geometric distribution, the variance is greater than the mean, since $E(N) = r\beta$, $Var(N) = r\beta(1+\beta)$, and $\beta > 0$. Thus, (b) and (c) do not fit the observation and are incorrect. Therefore, we are left with answer **(d) None of the above are correct**.

Problem S4C20-5. The number of anvils dropping from the sky in a particular region follows a geometric distribution with $\beta = 0.4$. What is the difference between the following two values:
A. The expected number of anvils falling from the sky in excess of 6 if it is known that the number of anvils falling from the sky is greater than 6.

B. The expected number of anvils falling from the sky in excess of 2 if it is known that the number of anvils falling from the sky is greater than 2.

Solution S4C20-5. Because the geometric distribution has the memoryless property, the expected number of anvils falling from the sky in excess of x does not depend on x, and so A and B are equal to one another, and **A - B = 0**.

Section 21

Properties of the Binomial Distribution and the (a, B, 0) Class of Distributions

For the **binomial distribution**, random variable N, integer values of k, and parameters m and q, where $0 < q < 1$.

The probability function or probability mass function (pf or pmf):

$p_k = Pr(N = k) = C(m, k)*q^k*(1-q)^{m-k}$, where k is a nonnegative integer less than or equal to m.

The probability generating function (pgf): $P(z) = (1 + q(z-1))^m$, $0 < q < 1$.

$E(N) = mq$; $Var(N) = mq(1-q)$.

If p_k is the pf of a discrete random variable, p_k is a member of the **(a, b, 0) class of distributions** if and only if there exist constants a and b such that $p_k/p_{k-1} = a + b/k$ for positive integer values of k. There are only four possible kinds of distributions belonging to the (a, b, 0) class: Poisson, binomial, negative binomial, and geometric, which is a special (r = 1) case of the negative binomial distribution. Here are their parameters (a and b) as pertain to the (a, b, 0) class:

Distribution	**a**	**b**	**p_0**
Poisson	0	λ	$e^{-\lambda}$
Binomial	$-q/(1-q)$	$(m+1)(q/(1-q))$	$(1-q)^m$
Negative Binomial	$\beta/(1+\beta)$	$(r-1)\beta/(1+\beta)$	$(1+\beta)^{-r}$
Geometric	$\beta/(1+\beta)$	0	$(1+\beta)^{-1}$

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 6, pp. 107-111.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C21-1. You are again modeling the frequency of events, and you need to select a distribution to use. You observe that the variance of the number of events is less than the mean number of events. Which of the following distributions should you use?

(a) Binomial
(b) Poisson
(c) Negative binomial
(d) Geometric
(e) None of the above are correct.

Solution S4C21-1. For the Poisson distribution, the variance is equal to the mean, so (b) is incorrect. For both the negative binomial distribution and its special r = 1 case, the geometric distribution, the variance is greater than the mean, since E(N) = rβ, Var(N) = rβ(1+β), and β > 0. Thus, (c) and (d) do not fit the observation and are incorrect. However, for the binomial distribution, 0 < q < 1, E(N) = mq, and Var(N) = mq(1-q), so Var(N) = (1-q)E(N) < E(N), and so the binomial distribution has a variance less than its mean. Thus, the correct answer is **(a) Binomial.**

Problem S4C21-2. A particular binomial distribution has b = -14a, in terms of the parameters describing the distribution as a member of the (a, b, 0) class. Find the value of m for this distribution.

Solution S4C21-2. We note that, for a binomial distribution, a = -q/(1-q) and b = (m+1)(q/(1-q)), so b = -(m+1)a. Thus, we have -(m+1) = 14 → m+1 = 14 → **m = 13**.

Problem S4C21-3. For a binomial distribution with m = 345 and q = 0.55, what is p_{232}/p_{231}?

Solution S4C21-3. We use the formula p_k/p_{k-1} = a + b/k, where a = -q/(1-q), b = -(m+1)a, and k = 232. a = -0.55/0.45 = -1.222222222. b = -(232+1)(-1.222222222) = 284.777777777778.

Thus, p_{232}/p_{231} = -1.222222222 + 284.777777777778/232 = **p_{232}/p_{231} = 0.0052681992**.

Problem S4C21-4. For a Poisson distribution with λ = 5, what is p_3/p_0?

Solution S4C21-4. We use multiple iterations of the formula p_k/p_{k-1} = a + b/k. For a Poisson distribution, $p_0 = e^{-\lambda}$, a = 0, and b = λ, so p_1/p_0 = 0 + λ/1 = λ.

$p_2/p_1 = 0 + \lambda/2 = \lambda/2$.

$p_3/p_2 = 0 + \lambda/3 = \lambda/3$.

Thus, $p_3/p_0 = (\lambda/3)(\lambda/2)(\lambda) = \lambda^3/6 = 5^3/6$ = **125/6 = p_3/p_0 = 20.833333333333**.

Problem S4C21-5. For a negative binomial distribution with r = 40 and β = 0.5, what is p_{23}/p_{21}?

Solution S4C21-5. We use multiple iterations of the formula p_k/p_{k-1} = a + b/k. For a negative binomial distribution, a = β/(1+β), and b = (r-1)a. here, a = 0.5/1.5 = (1/3). b = 39(1/3) = 13.

Thus, p_{22}/p_{21} = 1/3 + 13/22 = 61/66.

p_{23}/p_{22} = 1/3 + 13/23 = 62/69.

$p_{23}/p_{21} = (p_{23}/p_{22})(p_{22}/p_{21})$ = (61/66)(62/69) = **1891/2277 = p_{23}/p_{21} = 0.8304787**.

Section 22

The (a, B, 1) Class of Distributions

If p_k is the probability function (pf) of a discrete random variable, p_k is a member of the **(a, b, 1) class of distributions** if and only if there exist constants a and b such that $p_k/p_{k-1} = a + b/k$ for integer values of k *that are at least 2*. The only difference between the (a, b, 0) class of distributions and the (a, b, 1) class of distributions is that $p_k/p_{k-1} = a + b/k$ does not necessarily hold for k = 1.

For the (a, b, 1) class of distributions, if $p_0 = 0$, the distribution is called **zero-truncated.** The pf of a zero-truncated distribution is denoted p_k^T.

For the (a, b, 1) class of distributions, if $p_0 \geq 0$, the distribution is called **zero-modified.** A zero-truncated distribution is a type of zero-modified distribution. The pf of a zero-modified distribution is denoted p_k^M.

Theorem 22.1. If P(z) is the probability generating function (pgf) of a distribution of the (a, b, 0) class, then the pgf of the corresponding zero-modified distribution of the (a, b, 1) class has pgf $P^M(z) = (1-p_0^M)P(z)/(1- p_0)$.

If p_k is the pf of a distribution of the (a, b, 0) class, then the following are true:

Equation 22.2. $p_k^M = (1-p_0^M)p_k/(1- p_0)$.

Theorem 22.3. If P(z) is the probability generating function (pgf) of a distribution of the (a, b, 0) class, then the pgf of the corresponding zero-truncated distribution of the (a, b, 1) class has pgf $P^T(z) = (P(z) - p_0)/(1 - p_0)$.

Equation 22.4. $P^M(z) = p_0^M + (1- p_0^M)P^T(z)$.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 6, pp. 121-123.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C22-1. Recall that the pgf of an ordinary Poisson distribution is $P(z) = e^{\lambda(z-1)}$ and $p_0 = e^{-\lambda}$ if the distribution is not zero-modified. What is the pgf $P^M(z)$ of a zero-modified Poisson distribution with $\lambda = 4$ and $p_0^M = 0.3$?

Solution S4C22-1. We use the formula $P^M(z) = (1-p_0^M)P(z)/(1- p_0)$ from Theorem 22.1. We are given that $P(z) = e^{\lambda(z-1)} = e^{4(z-1)}$. Moreover, $p_0 = e^{-4}$, and $p_0^M = 0.3$. Thus, $P^M(z) = (1-0.3)e^{4(z-1)}/(1-e^{-4})$ = $\mathbf{P^M(z) = 0.7e^{4(z-1)}/(1-e^{-4}) = 0.7130601523e^{4(z-1)}}$.

Problem S4C22-2. A negative binomial distribution with r = 12 and β = 0.2 is zero-modified such that $p_0^M = 0.5$. Find p_2^M for this distribution.

Relevant properties of negative binomial distributions:

$p_k = \Pr(N = k) = (r(r+1)...(r+k-1)/k!)*(\beta^k/(1+\beta)^{r+k})$, k = 0, 1, 2, ..., r > 0, β > 0.

$p_0 = (1+\beta)^{-r}$.

Solution S4C22-2. We use the formula $p_k^M = (1-p_0^M)p_k/(1- p_0)$.
Here, $p_0 = 1/(1+0.2)^{12} = p_0 = 0.1121566548$.
$p_2 = (12(12+2-1)/2!)*(0.2^2/(1.2)^{14}) = 78(0.0031154626) = 0.2430060854$.
Thus, $p_2^M = (1-p_0^M)p_2/(1- p_0) = (1 - 0.5)0.2430060854/(1 - 0.1121566548) =$ $\mathbf{p_2^M =}$ **0.1368518932**.

Problem S4C22-3. Recall that the pgf of an ordinary geometric distribution is

$P(z) = (1-\beta(z-1))^{-1}$, and $p_0 = 1/(1+\beta)$. What is the pgf $P^T(z)$ of a zero-truncated geometric distribution with β = 2?

Solution S4C22-3. We use the formula $P^T(z) = (P(z) - p_0)/(1 - p_0)$. Here, $p_0 = 1/(1+2) = 1/3$. Moreover, $P(z) = (1-2(z-1))^{-1} = (3-2z)^{-1}$. Thus, $P^T(z) = ((3-2z)^{-1}- 1/3)/(1-1/3) =$ $\mathbf{P^T(z) = (3/2)(3-2z)^{-1} - (1/2)}$.

Problem S4C22-4. In Solution S4C22-3, we found that the pgf $P^T(z)$ of a zero-truncated geometric distribution with β = 2 is $P^T(z) = (3/2)(3-2z)^{-1} - (1/2)$. Using only this information, find the pgf $P^M(z)$ of a zero-modified geometric distribution with $p_0^M = 0.1$.

Solution S4C22-4. We use the formula $P^M(z) = p_0^M + (1- p_0^M)P^T(z)$.
$P^M(z) = 0.1 + (1-0.1)((3/2)(3-2z)^{-1} - (1/2))$.
$P^M(z) = 0.9 + 1.35(3-2z)^{-1} - 0.45$
$\mathbf{P^M(z) = 0.45 + 1.35(3-2z)^{-1} = (9/20) + (27/20)(3-2z)^{-1}}$.

Problem S4C22-5. A binomial distribution has m = 5 and q = 0.2. The distribution has been zero-truncated. Find p_4^T.

Relevant properties of binomial distributions:

$p_k = \Pr(N = k) = C(m, k)*q^k*(1-q)^{m-k}$, where k is a nonnegative integer less than or equal to m.

Solution S4C22-5. A zero-truncated distribution is a special instance of a zero-modified distribution, so the following formula still applies: $p_k^M = (1-p_0^M)p_k/(1- p_0)$, where $p_0^M = 0$. P
We find $p_0 = (1-q)^m = 0.8^5 = 0.32768$.
We find $p_4 = C(5, 4)*0.2^4*0.8 = 0.0064$.
Thus, $p_4^T = (1-0)0.0064/(1-0.32768) =$ $\mathbf{p_4^T = 0.0095192765}$.

Section 23

The Logarithmic Distribution

If a negative binomial distribution is zero-truncated and the parameter $r \to 0$, the **logarithmic distribution** arises.

This distribution belongs to the (a, b, 1) class (for $a = \beta/(1+\beta)$ and $b = -\beta/(1+\beta)$), has a single parameter, $\beta > 0$, and has the following properties:

$p_1^T = \beta/(\ln(1+\beta)(1+\beta))$

$p_k^T = \beta^k/(k*\ln(1+\beta)*(1+\beta)^k)$

$E(N) = \beta/(\ln(1+\beta))$; $Var(N) = \beta(1+\beta - \beta/(\ln(1+\beta)))/(\ln(1+\beta))$

$P(z) = 1 - \ln(1-\beta(z-1))/(\ln(1+\beta))$

Note that the probability function (pf) is expressed as a zero-truncated pf, because the logarithmic distribution is a limiting case of a zero-truncated negative binomial distribution.

Source: *Loss Models: From Data to Decisions*, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 6, p. 124.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C23-1. Random variable X follows a logarithmic distribution with $\beta = 3$. Find p_2^T.

Solution S4C23-1. We use the formula $p_k^T = \beta^k/(k*\ln(1+\beta)*(1+\beta)^k)$. Here, k = 2, so

$p_2^T = 3^2/(2*\ln(1+3)*(1+3)^2) = 9/(2*\ln(4)*16) =$ $\mathbf{p_2^T = 0.2028789901}$.

Problem S4C23-2. Random variable X follows a logarithmic distribution with $\beta = 3$. Find E(X).

Solution S4C23-2. We use the formula $E(X) = \beta/(\ln(1+\beta)) = 3/\ln(4) =$ **E(X) = 2.164042561**.

Problem S4C23-3. Random variable X follows a logarithmic distribution with $\beta = 3$. Find Var(X).

Solution S4C23-3. We use the formula $Var(X) = \beta(1+ \beta - \beta/(\ln(1+\beta)))/(\ln(1+\beta)) =$

$3(1+3-3/\ln(4))/\ln(4) = 3(4-3/\ln(4))/\ln(4) =$ **Var(X) = 3.973090038**.

Problem S4C23-4. Is it ever possible for the variance of a logarithmic distribution to be less than the mean? If so, what are the conditions that β must meet for this to happen? If not, why not?

Solution S4C23-4. For a logarithmic distribution, E(X) = β/(ln(1+β)) and

Var(X) = β(1+ β - β/(ln(1+β)))/(ln(1+β)). In order for Var(X) < E(X), it would need to be the case that

(1+ β - β/(ln(1+β))) < 1 or, equivalently, β - β/(ln(1+β)) < 0 or, equivalently, β < β/(ln(1+β)).

We can divide both sides of the inequality by β to get 1 < 1/ ln(1+β) → ln(1+β) < 1→ 1+β < e → β < e-1.

We note that it is still the case that β must be greater than 0.

Thus, **the variance of a logarithmic distribution can be less than its mean, if and only if**

0 < β < e-1.

Problem S4C23-5. Random variable X follows a logarithmic distribution with β = 0.2. Find P(2).

Solution S4C23-5. We use the formula P(z) = 1 - ln(1-β(z-1))/(ln(1+β)) for z = 2:

P(2) = 1 - ln(1-0.2(2-1))/ln(1+0.2) = 1 - ln(0.8)/ln(1.2) = **P(2) = 2.223901086**.

Section 24

Assorted Exam-Style Questions for Exam 4 / Exam C – Part 1

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

This section will also serve to give some practice with the **Chi-square goodness of fit test.**

The **Chi-square distribution (χ^2 distribution)** can be defined as follows:

The p.d.f. of U = ${}_{j=1}{}^{m}\sum[Z_j{}^2]$, where each Z_j is a standard normal random variable independent of all the other Z_j's, is called the **Chi-square distribution with m degrees of freedom.** (Larsen and Marx 2006, p. 474).

It is easy to remember the Chi-square distribution as the p.d.f. that results from adding the *squares* of m *standard normal random variables*, where m is the number of degrees of freedom for the Chi-square distribution.

The chi-square distribution has **critical values** associated with a given **significance level** and a number of degrees of freedom. You can look up these critical values in a Table of Critical Values for the Chi-Square Distribution.

Larsen and Marx define a **goodness of fit test** as "any procedure that seeks to determine whether a set of data could reasonably have originated from some given probability distribution, or *class* of probability distributions" (599).

The **Chi-square goodness of fit test** can be phrased as follows.

"Let r_1, r_2, ..., r_t be the set of possible outcomes (or ranges of outcomes) associated with each of n independent trials, where $P(r_i) = p_i$, for i = 1, 2, ..., t. Let X_i = the number of times r_i occurs. Then the following holds.

a. The random variable D = ${}_{i=1}{}^{t}\sum[(X_i - np_i)^2/np_i]$ has approximately a χ^2 distribution with t-1 degrees of freedom. For the approximation to be adequate, the t classes should be defined so that $np_i \geq 5$ for all i.

b. Let $k_1, k_2, ..., k_t$ be the observed frequencies for the outcomes $r_1, r_2, ..., r_t$, respectively, and let $np_{1o}, np_{2o}, ..., np_{to}$ be the corresponding expected frequencies based on the null hypothesis. At the α level of significance, H_0: $f_Y(y) = f_o(y)$ (or H_0: $p_X(k) = p_o(k)$) is rejected if

$d = {}_{i=1}{}^{t}\sum[(k_i - np_{io})^2/(np_{io})] \geq \chi^2_{1-\alpha,t-1}$." (Larsen and Marx 2006, p. 616)

For instance, if we have a distribution of observations, and we know both the expected numbers of observations of each kind and the actual numbers of observations of this kind, then we can let k_i be the actual number of observations of the type i and np_{io} be the expected number of observations of the type i. Then we can find $d = {}_{i=1}{}^{t}\sum[(k_i - np_{io})^2/(np_{io})]$ and compare it to

$\chi^2_{1-\alpha,t-1}$, the chi-square statistic with t-1 degrees of freedom and significance level of 1-α.

If $d \geq \chi^2_{1-\alpha,t-1}$, then we reject the null hypothesis H_0.

Sources:

Broverman, Sam. Actuarial Exam Solutions - CAS Exam 3 - Fall 2006.

Exam C Sample Questions and Solutions from the Society of Actuaries.

Larsen, Richard J. and Morris L. Marx. *An Introduction to Mathematical Statistics and Its Applications.* Fourth Edition. Pearson Prentice Hall: 2006. pp. 474, 599, 616.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C24-1. Similar to Question 11 of the Exam C Sample Questions from the Society of Actuaries. A Pareto distribution has pdf $f(x \mid \theta) = 2\theta^2/(x+\theta)^3$. There is a 1/3 probability that θ = 2 and a 2/3 probability that θ = 6. Variable X follows this distribution and denotes the number of blue slugs that will land on the roof of Mr. Zolatrax's house in a given month. In April, it is known that 4 blue slugs landed on the roof of Mr. Zolatrax's house. Given this information, find the probability that more than 10 slugs will land on the roof of Mr. Zolatrax's house in May. (This is called a **posterior probability**.)

Solution S4C24-1. First, we find what the probability of 4 slugs having landed on Mr. Zolatrax's roof is, given each of the possible values of θ:

$f(4 \mid 2) = 2*2^2/(4+2)^3 = 1/27$

$f(4 \mid 6) = 2*6^2/(4+6)^3 = 9/125$

Now we want to find the probability that θ = 2, *given* our past observation, X = 4:
$\Pr(\theta = 2 \mid X = 4) = \Pr(\theta = 2)*\Pr(X = 4 \mid \theta = 2)/\Pr(X = 4) =$

$\Pr(\theta = 2)*\Pr(X = 4 \mid \theta = 2)/(\Pr(\theta = 2)*\Pr(X = 4 \mid \theta = 2) + \Pr(\theta = 6)*\Pr(X = 4 \mid \theta = 6))$.

We know that $\Pr(\theta = 2) = 1/3$ and $\Pr(X = 4 \mid \theta = 2) = f(4 \mid 2) = 1/27$.

We also know that $\Pr(\theta = 6) = 2/3$ and $\Pr(X = 4 \mid \theta = 6) = f(4 \mid 6) = 9/125$.

Thus, $\Pr(\theta = 2 \mid X = 4) = (1/3)(1/27)/((1/3)(1/27) + (2/3)(9/125)) = 125/611 = 0.2045826514$.

From this it follows that $\Pr(\theta = 6 \mid X = 4) = 1 - \Pr(\theta = 2 \mid X = 4) = 486/611 = 0.7954173486$.

In calculating $S_X(10)$, we will thus assume that there is a 125/611 probability that $\theta = 2$ and a 486/611 probability that $\theta = 6$.

Now we need to find $S_X(x) = 1 - {}_0{}^{\infty}\int(2\theta^2/(x+\theta)^3)dx = 1 - (-\theta^2/(x+\theta)^2) = \theta^2/(x+\theta)^2$.

Hence, $S_X(10) = (125/611)*2^2/(10+2)^2 + (486/611)*6^2/(10+6)^2 =$ **$S_X(10) = 0.1175384161$**.

Problem S4C24-2. Similar to Question 13 of the Exam C Sample Questions from the Society of Actuaries. There are three colors of cat-dogs: yorange, grue, and rurple. The following are their historical probabilities of occurring: yorange: 0.43; grue: 0.24; rurple: 0.33. In a sample of 1000 cat-dogs, it is observed that 465 are yorange, 198 are grue, and 337 are rurple. You are testing the hypothesis that the observed probabilities of each of the three colors are the same as the historical probabilities. Find the Chi-square goodness of fit statistic for this test.

Solution S4C24-2. Our test statistic is calculated as follows: ${}_{i=1}{}^{t}\sum[(X_i - np_i)^2/np_i]$. Here, the expected numbers of each color of cat-dog, based on the historical probabilities, are as follows:

Yorange: 0.43*1000 = 430; Grue: 0.24*1000 = 240; Rurple: 0.33*1000 = 330.

These are our values of np_i in each term of the formula calculating the Chi-square statistic:
$\chi^2 = (465-430)^2/430 + (198-240)^2/240 + (337-330)^2/330 =$ **$\chi^2 = 10.34732206$**.

Problem S4C24-3. Similar to Question 19 of the Exam C Sample Questions from the Society of Actuaries. A random variable X follows a probability distribution with the following pdf: $f_X(x) = (1/4)(1/\theta)e^{-x/\theta} + (3/4)(1/\sigma)e^{-x/\sigma}$. A sample of 200 values is obtained, with the following sample data: ${}_{i=1}{}^{200}\Sigma x_i = 96000$; ${}_{i=1}{}^{200}\Sigma x_i{}^2 = 80000000$. It is known that $\theta > \sigma$. Assuming that the sample is representative of the population, estimate the value of θ by using the moments of the sample distribution. Use the Exam 4 / C Tables where necessary.

Solution S4C24-3. We note that the distribution in question is a mixture of two exponential distributions. The exponential distribution with mean θ has weight (1/4), and the exponential distribution with mean σ has weight (3/4). Thus, $E(X) = 0.25\theta + 0.75\sigma$, which, using the first moment of the sample, we know to be 96000/200 = 480.

The second moment $E(X^2)$ of an exponential distribution is $2\theta^2$, so in this case, $E(X^2) = 0.25*2\theta^2 + 0.75*2\sigma^2 = 0.5\theta^2 + 1.5\sigma^2$. Using the second moment of the sample, we know this to be equal to 80000000/200 = 40000. Thus, we must solve the following system of equations:
$0.25\theta + 0.75\sigma = 480$

$0.5\theta^2 + 1.5\sigma^2 = 400000$.

$0.25\theta + 0.75\sigma = 480 \rightarrow \sigma = (480 - 0.25\theta)/0.75 = 640 - \theta/3$.

This means that $400000 = 0.5\theta^2 + 1.5(640 - \theta/3)^2 \rightarrow$

$400000 = 0.5\theta^2 + 1.5(409600 - 6400 + \theta^2/9) \rightarrow$

$0 = (2/3)\theta^2 - 960\theta + 214400$. Using the Quadratic Formula, we find that

$\theta = 1163.62146$ or $\theta = 276.3785397$.

Of these values, if $\theta = 276.3785397$, it is clearly the case that $\sigma > \theta$, or else any weighted average of these two values would not be equal to 480. Thus, in order for $\theta > \sigma$ to be the case, it must be that **θ = 1163.62146**.

Problem S4C24-4. Similar to Question 25 of the Exam C Sample Questions from the Society of Actuaries. You are modeling the number of earthquakes in a variety of counties per year. A sample of 100 counties produced the following results:

53 counties had 0 earthquakes.

23 counties had 1 earthquake.

13 counties had 2 earthquakes.

4 counties had 3 earthquakes.

3 counties had 4 earthquakes.

3 counties had 5 earthquakes.

1 county had 6 earthquakes.

Which of the following distributions would be best for modeling this data?

(a) Poisson

(b) Negative binomial

(c) Binomial

(d) Discrete uniform

(e) Either Poisson or binomial.

Solution S4C24-4. We compare the variance to the mean. We find E(X) = 0*53/100 + 1*23/100 + 2*13/100 + 3*4/100 + 4*3/100 + 5*3/100 + 6*1/100 = E(X) = 0.94.

$E(X^2) = 0^2*53/100 + 1^2*23/100 + 2^2*13/100 + 3^2*4/100 + 4^2*3/100 + 5^2*3/100 + 6^2*1/100 =$
$E(X^2) = 2.7$.

$Var(X) = E(X^2) - E(X)^2 = 2.7 - 0.94^2 = 1.8164$.

Here, the variance is almost twice as large as the mean. Earthquakes in the sample are clearly not uniformly distributed between 0 and 6 and, of the distributions listed above, only the negative binomial has a variance larger than the mean. Thus, the correct answer is **(b) Negative binomial**.

Problem S4C24-5. Similar to Question 28 of the Exam C Sample Questions from the Society of Actuaries. 100 giraffes are measured for their height (X) in centimeters. The following data are observed:

34 giraffes are between 300 and 400 centimeters in height.

54 giraffes are between 400 and 500 centimeters in height.

12 giraffes are between 500 and 600 centimeters in height.

Assume that the heights are uniformly distributed within each interval.

Find $E(X^2) - E((X \wedge 570)^2)$.

Solution S4C24-5.

We use the formulas $E(X^k) = {}_{-\infty}{}^{\infty}\int x^k*f(x)dx$ and $E((X \wedge u)^k) = {}_{-\infty}{}^{u}\int x^k*f(x)dx + u^k*(1 - F(u))$.

Thus, $E(X^2) - E((X \wedge 570)^2) = {}_{300}{}^{600}\int x^2*f(x)dx - {}_{300}{}^{570}\int x^2*f(x)dx - 570^2(1 - F(570)) \rightarrow$

$E(X^2) - E((X \wedge 570)^2) = {}_{570}{}^{600}\int x^2*f(x)dx - 570^2(1 - F(570))$.

The uniform distribution on the interval from 500 to 600 has a pdf of (1/100), multiplied by the weight of (12/100) - as only 12 of the 100 observed giraffes are within the interval in question. Thus, the relevant value of f(x) is 12/10000 = 0.0012.

1 - F(570) is the same as S(570) = the probability that a giraffe is taller than 570 centimeters. Because we assume a uniform distribution within each interval, 12(600-570)/(600-500) = 3.6 of 100 giraffes can be expected to be taller than 570 centimeters. Therefore, 1 - F(570) = 0.036, and so

$E(X^2) - E((X \wedge 570)^2) = {}_{570}{}^{600}\int 0.0012x^2dx - 570^2(0.036) =$

$\mathbf{E(X^2) - E((X \wedge 570)^2) = 626.4}$.

Section 25

Applications of Ordinary Deductibles Per Payment and Per Loss

An **ordinary deductible**, d, modifies the amount X that an insurance company needs to pay for a loss by transforming X into X - d. Deductibles can be considered *per payment* (the corresponding payment variable is Y^P) or *per loss* (the corresponding payment variable is Y^L).

The per-payment payment variable is as follows:

Y^P = undefined when $X \leq d$. (This is so because the insurer does not need to make a payment.)

$Y^P = X - d$ when $X > d$.

This corresponds to an excess loss random variable.

The per-loss payment variable is as follows:

$Y^L = 0$ when $X \leq d$.

$Y^L = X - d$ when $X > d$.

This corresponds to a left-censored and shifted random variable.

Here are the properties of these two random variables:

$f_{Y^\wedge P}(y) = f_X(y+d)/S_X(d)$, $y > 0$

$S_{Y^\wedge P}(y) = S_X(y+d)/S_X(d)$

$F_{Y^\wedge P}(y) = (F_X(y+d) - F_X(d))/(1 - F_X(d))$

$h_{Y^\wedge P}(y) = f_X(y+d)/S_X(y+d) = h_X(y+d)$

$f_{Y^\wedge L}(y) = f_X(y+d)$, $y > 0$. (Note: The probability $f_{Y^\wedge L}(0)$ is a discrete probability.)

$S_{Y^\wedge L}(y) = S_X(y+d)$, $y \geq 0$

$F_{Y^\wedge L}(y) = F_X(y+d)$, $y \geq 0$

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 8, pp. 179-180.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C25-1. Losses from falling cows follow an exponential distribution with mean θ = 340. Falling Cow Insurance Company requires policyholders to have an ordinary deductible of 200. Find $f_{Y^\wedge P}(y)$, the pdf of the per-payment payment variable under this policy.

Solution S4C25-1. We use the formula $f_{Y^\wedge P}(y) = f_X(y+d)/S_X(d)$, y > 0.

For an exponential distribution, $f_X(x) = e^{-x/\theta}/\theta$, and $S_X(x) = e^{-x/\theta}$. Here, d = 200.

Thus, $f_{Y^\wedge P}(y) = e^{-(y+200)/340}/340/(e^{-200/340})$ = $\mathbf{f_{Y^\wedge P}(y) = e^{-y/340}/340, y > 0}$.

Why did we get as our answer the pdf of the given exponential distribution, except as applied to Y^P instead of X? This is so because the exponential distribution has the memoryless property, and so any random variable that is undefined when X ≤ d and is equal to X- d when X > d will have the same pdf, irrespective of the value of d.

Problem S4C25-2. Losses from falling cows follow an exponential distribution with mean θ = 340. Falling Cow Insurance Company requires policyholders to have an ordinary deductible of 200. Find $f_{Y^\wedge L}(y)$, the pdf of the per-loss payment variable under this policy. Do not forget to consider what happens when y = 0!

Solution S4C25-2. We use the formula $f_{Y^\wedge L}(y) = f_X(y+d)$, y > 0.

For an exponential distribution, $f_X(x) = e^{-x/\theta}/\theta$. Here, d = 200.

Thus, $f_{Y^\wedge L}(y) = e^{-(y+200)/340}/340$, y > 0.

We also want to find $f_{Y^\wedge L}(0)$, which is the probability that the loss will be less than or equal to 200, which is the same as $F_X(200) = 1 - e^{-200/340} = 0.444693627$.

Our answer is as follows:

$\mathbf{f_{Y^\wedge L}(y) = 0.444693627}$ **when y = 0;**

$\mathbf{f_{Y^\wedge L}(y) = e^{-(y+200)/340}/340}$ **when y > 0**.

Problem S4C25-3. Losses from house-smashing snakes (HSSs) follow a Pareto distribution with α = 3 and θ = 1000. HSS Mutual Insurance Company requires policyholders to have an ordinary deductible of 500. Find $S_{Y^\wedge P}(y)$, the survival function of the per-payment payment variable under this policy.

Relevant properties for Pareto distributions: $S(x) = \theta^\alpha/(x + \theta)^\alpha$.

Solution S4C25-3. We use the formula $S_{Y^\wedge P}(y) = S_X(y+d)/S_X(d)$, for d = 500.

Thus, $S_X(y+d) = 1000^3/(y + 500 + 1000)^3 = 1000^3/(y + 1500)^3$ and

$S_X(d) = 1000^3/(500 + 1000)^3 = 8/27$.

Hence, $S_{Y^\wedge P}(y) = 27*1000^3/(8(y + 1500)^3)$ = $\mathbf{S_{Y^\wedge P}(y) = 3.375*10^9/(y + 1500)^3= 3000^3/(2y + 3000)^3}$.

Problem S4C25-4. Losses from house-smashing snakes (HSSs) follow a Pareto distribution with $\alpha = 3$ and $\theta = 1000$. HSS Mutual Insurance Company requires policyholders to have an ordinary deductible of 500. Find $F_{Y^\wedge L}(y)$, the cumulative distribution function of the per-loss payment variable under this policy.

Relevant properties for Pareto distributions: $S(x) = \theta^\alpha/(x + \theta)^\alpha$.

Solution S4C25-4. We use the formula $F_{Y^\wedge L}(y) = F_X(y+d)$, $y \geq 0$. Here, d = 500 and

$F_X(x) = 1 - S_X(x) = 1 - \theta^\alpha/(x + \theta)^\alpha = 1 - 1000^3/(x + 1000)^3$.

Thus, $F_X(y+d) = 1 - 1000^3/(y + 1500)^3$, so

$\mathbf{F_{Y^\wedge L}(y) = 1 - 1000^3/(y + 1500)^3, y \geq 0}$.

Problem S4C25-5. Global warming/cooling/no-temperature-change is a serious environmental problem. Annual dollar losses from this peril follow a binomial distribution with m = 4 and q = 0.23. The Untied Nations has decided to implement an insurance policy against this peril, with an ordinary deductible of 1. Find $F_{Y^\wedge P}(y)$, the cdf of the per-payment payment variable under this policy.

Solution S4C25-5. We use the formula $F_{Y^\wedge P}(y) = (F_X(y+d) - F_X(d))/(1 - F_X(d))$.

$F_X(d) = F_X(1) = f_X(0) + f_X(1)$, since the binomial distribution is discrete.

$f_X(0) = (1-q)^m = (1-0.23)^4 = 0.35153041$.

$f_X(1) = C(m,1)*q*(1-q)^{m-1} = m*q*(1-q)^{m-1} = 4*0.23*0.77^3 = 0.42001036$.

Thus, $F_X(1) = f_X(0) + f_X(1) = 0.35153041 + 0.42001036 = 0.77154077$.

We now need to find $F_X(y+d)$. We note that the only possible losses occur in whole dollar amounts, and the value of $F_{Y^\wedge P}(y)$ will depend on the amount of loss that has occurred.

We already know that $F_X(0+d) = F_X(0+1) = F_X(1) = 0.77154077$.

$f_X(2) = C(4,2)*0.23^2*0.77^2 = 0.18818646$.

Thus, $F_X(1+d) = F_X(2) = 0.77154077 + 0.18818646 = 0.95972723$.

$f_X(3) = C(4,3)*0.23^3*0.77= 0.03747436$.

Thus, $F_X(2+d) = F_X(3) = 0.95972723 + 0.03747436 = 0.99720159$.

Since no loss can be greater than 4, $F_X(4) = F_X(3+d) = 1$.

Now we can apply our formula:

$F_{Y^\wedge P}(0) = (F_X(0+d) - F_X(d))/(1 - F_X(d)) = F_{Y^\wedge P}(0) = 0$.

$F_{Y^\wedge P}(1) = (F_X(1+d) - F_X(d))/(1 - F_X(d)) = (0.95972723 - 0.77154077)/(1 - 0.77154077) = F_{Y^\wedge P}(1) = 0.823720101$.

$F_{Y^\wedge P}(2) = (F_X(2+d) - F_X(d))/(1 - F_X(d)) = (0.99720159 - 0.77154077)/(1 - 0.77154077) = F_{Y^\wedge P}(2) = 0.9877509436$.

$F_{Y^\wedge P}(3) = (F_X(3+d) - F_X(d))/(1 - F_X(d)) = (1 - 0.77154077)/(1 - 0.77154077) = F_{Y^\wedge P}(3) = 1$.

Thus, we have the following answer:

$\mathbf{F_{Y^\wedge P}(0) = 0;}$

$\mathbf{F_{Y^\wedge P}(1) = 0.823720101;}$

$\mathbf{F_{Y^\wedge P}(2) = 0.9877509436;}$

$\mathbf{F_{Y^\wedge P}(3) = 1}$.

Section 26

Applications of Franchise Deductibles Per Payment and Per Loss

A **franchise deductible** differs from an ordinary deductible in that the loss is paid in full if the loss is in excess of the deductible. We can again apply franchise deductibles to per-payment and per-loss payment random variables.

The per-payment payment variable is as follows:

Y^P = undefined when $X \le d$. (This is so because the insurer does not need to make a payment.)

$Y^P = X$ when $X > d$.

The per-loss payment variable is as follows:

$Y^L = 0$ when $X \le d$.
$Y^L = X$ when $X > d$.
$f_{Y^\wedge P}(y) = f_X(y)/S_X(d)$, $y > d$
$S_{Y^\wedge P}(y) = 1$ if $0 \le y \le d$;
$S_{Y^\wedge P}(y) = S_X(y)/S_X(d)$ if $y > d$.
$F_{Y^\wedge P}(y) = 0$ if $0 \le y \le d$;
$F_{Y^\wedge P}(y) = (F_X(y) - F_X(d))/(1 - F_X(d))$ if $y > d$.
$h_{Y^\wedge P}(y) = 0$ if $0 < y < d$;
$h_{Y^\wedge P}(y) = h_X(y)$ if $y > d$.
$f_{Y^\wedge L}(y) = F_X(d)$, $y = 0$;
$f_{Y^\wedge L}(y) = f_X(y)$, $y > d$.
$S_{Y^\wedge L}(y) = S_X(d)$, $0 \le y \le d$;
$S_{Y^\wedge L}(y) = S_X(y)$, $y > d$.
$F_{Y^\wedge L}(y) = F_X(d)$, $0 \le y \le d$;
$F_{Y^\wedge L}(y) = F_X(y)$, $y > d$.
$h_{Y^\wedge L}(y) = 0$ if $0 < y < d$;
$h_{Y^\wedge L}(y) = h_X(y)$ if $y > d$.

For an ordinary deductible, the expected cost per payment is $(E(X) - E(X \wedge d))/(1-F(d))$.

For a franchise deductible, the expected cost per payment is $(E(X) - E(X \wedge d))/(1-F(d)) + d$.

For an ordinary deductible, the expected cost per loss is $E(X) - E(X \wedge d)$.

For a franchise deductible, the expected cost per loss is $E(X) - E(X \wedge d) + d(1 - F(d))$.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 8, pp. 182-183.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C26-1. Losses from flying cows follow an exponential distribution with mean θ = 340. Flying Cow Insurance Company requires policyholders to have a franchise deductible of 200. Find $f_{Y^P}(y)$, the pdf of the per-payment payment variable under this policy.

Solution S4C26-1. We use the formula $f_{Y^P}(y) = f_X(y)/S_X(d)$, y > d.

For an exponential distribution, $f_X(x) = e^{-x/\theta}/\theta$, and $S_X(x) = e^{-x/\theta}$. Here, d = 200.

Thus, $f_{Y^P}(y) = f_X(y)/S_X(d) = (e^{-y/340}/340)/(e^{-200/340})$ = $\mathbf{f_{Y^P}(y) = e^{(-y+200)/340}/340, y > 200}$.

Problem S4C26-2. Losses from flying cows follow an exponential distribution with mean θ = 340. Flying Cow Insurance Company requires policyholders to have a franchise deductible of 200. Find $f_{Y^L}(y)$, the pdf of the per-loss payment variable under this policy.

Solution S4C26-2. We use the formula $f_{Y^L}(y) = F_X(d)$, y = 0; $f_{Y^L}(y) = f_X(y)$, y > d.
For an exponential distribution, $f_X(x) = e^{-x/\theta}/\theta$, and $S_X(x) = e^{-x/\theta}$. Here, d = 200.
$F_X(200) = 1 - S_X(200) = 1 - e^{-200/340} = 1 - e^{-10/17} = 0.444693627$.
Thus,
$\mathbf{f_{Y^L}(y) = 1 - e^{-10/17} = 0.444693627}$ **if y = 0;**
$\mathbf{f_{Y^L}(y) = e^{-y/340}/340, y > 200}$.

Problem S4C26-3. Losses from house-smashing snakes (HSSs) are denoted by a random variable X that follows a Pareto distribution with α = 3 and θ = 1000. HSS Mutual Insurance Company requires policyholders to have an ordinary deductible of 500. Find the expected cost per loss under this policy.

Relevant properties for Pareto distributions: $S(x) = \theta^\alpha/(x + \theta)^\alpha$
$E(X) = \theta/(\alpha - 1)$
$E(X \wedge K) =(\theta/(\alpha - 1))(1 - (\theta/(K + \theta))^{(\alpha - 1)})$

Solution S4C26-3. The expected cost per loss for a policy with an ordinary deductible d is

$E(X) - E(X \wedge d) = E(X) - E(X \wedge 500)$.

Here, $E(X) = \theta/(\alpha - 1) = 1000/(3-1) = 500$.

$E(X \wedge 500) = (\theta/(\alpha - 1))(1 - (\theta/(K + \theta))^{(\alpha - 1)}) = 500(1 - (1000/(500 + 1000))^2) = 500(1 - (2/3)^2) =$ 277.777777778.

Thus, $E(X) - E(X \wedge d) = 500 - 277.777777778 =$ **222.22222222222**.

Problem S4C26-4. Losses from house-smashing snakes (HSSs) are denoted by a random variable X that follows a Pareto distribution with $\alpha = 3$ and $\theta = 1000$. The House-Smashing Snake Special (HSSS) Mutual Insurance Company requires policyholders to have a *franchise* deductible of 500. Find the expected cost per loss under this policy.

Relevant properties for Pareto distributions: $S(x) = \theta^\alpha/(x + \theta)^\alpha$
$E(X) = \theta/(\alpha - 1)$
$E(X \Lambda K) =(\theta/(\alpha - 1))(1 - (\theta/(K + \theta))^{(\alpha - 1)})$

Solution S4C26-4.

For a franchise deductible, the expected cost per loss is $E(X) - E(X \Lambda d) + d(1 - F(d))$.

From Solution S4C26-4, we already know that $E(X) - E(X \Lambda d) = 222.22222222$.

We need to find $d(1 - F(d))$. We know that $d = 500$. We need to find $1 - F(500) = S(500) =$

$1000^3/(500 + 1000)^3 = 8/27$. Thus, $d(1 - F(d)) = 500*8/27 = 148.1481481$.

Hence, $E(X) - E(X \Lambda d) + d(1 - F(d)) = 222.22222222 + 148.1481481 =$ **370.3703704.**

Problem S4C26-5. Losses from flying cows follow an exponential distribution with mean θ = 340. Flying Cow Insurance Company requires policyholders to have a franchise deductible of 200. Find the expected cost per payment under this policy.

Relevant properties for exponential distributions: $S_X(x) = e^{-x/\theta}$;

$E(X) = \theta$; $E(X \Lambda K) = \theta(1 - e^{-K/\theta})$.

Solution S4C26-5.

For a franchise deductible, the expected cost per payment is $(E(X) - E(X \Lambda d))/(1-F(d)) + d$.

We know that $E(X) = \theta = 340$ and that $d = 200$.
We calculate $E(X \Lambda 200) = 340(1 - e^{-200/340}) = 151.1958332$.
We find $1-F(d) = S(d) = S(200) = e^{-200/340} = 0.555306373$.

Thus, the expected cost per payment is $(340 - 151.1958332)/0.555306373 + 200 =$ **540**.

(Note that the memoryless property of an exponential distribution implies that the expected cost per paymentwith an *ordinary* deductible is the same as E(X), the expected cost per payment without any deductible at all.)

Section 27

Loss Elimination Ratios and Inflation Effects for Ordinary Deductibles

The **loss elimination ratio** is the decrease in the expected payment with an ordinary deductible (d), divided by the expected payment without the deductible (E(X)). It is computed via the following formula: **Loss Elimination Ratio = E(X Λ d)/E(X)**.

If there is uniform inflation at a rate r and an ordinary deductible d, then the expected cost per loss is **(1+r)(E(X) - E(X Λ d/(1+r))**

Furthermore, F(d/(1+r)) < 1, then the expected cost per payment is

(1+r)(E(X) - E(X Λ d/(1+r))/(1 - F(d/(1+r)))

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 8, p. 185.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C27-1. Losses from falling cows follow an exponential distribution with mean θ = 340. Falling Cow Insurance Company requires policyholders to have an ordinary deductible of 200. Find the loss elimination ratio for this policy.

Relevant properties for exponential distributions: $S_X(x) = e^{-x/\theta}$;

$E(X) = \theta$; $E(X \wedge K) = \theta(1 - e^{-K/\theta})$.

Solution S4C27-1. We use the formula Loss Elimination Ratio = E(X Λ d)/E(X). For an exponential distribution, E(X) = θ, which, in this case, is 340. Here, d = 200, so E(X Λ d) = $340(1 - e^{-200/340}) = 151.1958332$. Thus, our loss elimination ratio is 151.1958332/340 = **0.4446936271**.

Problem S4C27-2. Losses from falling cows follow an exponential distribution with mean θ = 340. Falling Cow Insurance Company requires policyholders to have an ordinary deductible of 200. Now the Fed begins to print vast amounts of money, and the annual rate of inflation is 30%. What is the expected cost per loss at this inflation rate after one year?

Solution S4C27-2.
We know from Solution S4C27-1 that E(X) = 340 and E(X Λ d) = 151.1958332.
For one year, r = 0.3, so we can use the formula

$(1+r)(E(X) - E(X \wedge d/(1+r)) = 1.3(E(X) - E(X \wedge d/1.3)) =$
$1.3(E(X) - E(X \wedge 153.8461538)) =$
$1.3(340 - 340(1 - e^{-153.8461538/340})) = 1.3(340e^{-153.8461538/340}) =$ **281.1311241.**

Problem S4C27-3. Losses from falling cows follow an exponential distribution with mean θ = 340. Falling Cow Insurance Company requires policyholders to have an ordinary deductible of 200. Now the Fed begins to print vast amounts of money, and the annual rate of inflation is 30%. What is the expected cost *per payment* at this inflation rate after one year?

Solution S4C27-3. The formula to use in this instance is
$(1+r)(E(X) - E(X \wedge d/(1+r))/(1 - F(d/(1+r)))$.
However, we first need to verify that $F(d/(1+r)) < 1$.
$d/(1+r) = 200/1.3 = 153.8461538$.
$F(153.8461538) = 1 - e^{-153.8461538/340} = 0.3639567328 < 1$, so the formula applies.
We know from Solution S4C27-2 that $(1+r)(E(X) - E(X \wedge d/(1+r)) = 281.1311241$.
Thus, the expected cost per payment is 281.1311241/(1-0.3639567328) = **442.**

Problem S4C27-4. Losses from house-smashing snakes (HSSs) follow a Pareto distribution with α = 3 and θ = 1000. HSS Mutual Insurance Company requires policyholders to have an ordinary deductible of 500. Find the loss elimination ratio for this policy.

Relevant properties for Pareto distributions: $S(x) = \theta^\alpha/(x + \theta)^\alpha$.

$E(X) = \theta/(\alpha - 1)$
$E(X \wedge K) = (\theta/(\alpha - 1))(1 - (\theta/(K + \theta))^{(\alpha - 1)})$

Solution S4C27-4. We use the formula Loss Elimination Ratio = $E(X \wedge d)/E(X)$, which in this case is $(\theta/(\alpha - 1))(1 - (\theta/(d + \theta))^{(\alpha - 1)})/(\theta/(\alpha - 1)) = (1 - (\theta/(d + \theta))^{(\alpha - 1)}) =$
$(1 - (1000/1500)^2) =$ **5/9 = 0.5555555556**.

Problem S4C27-5. Losses from house-smashing snakes (HSSs) follow a Pareto distribution with α = 3 and θ = 1000. HSS Mutual Insurance Company requires policyholders to have an ordinary deductible of 500. The Fed has gone mad and doubled the money supply within one year, so that the inflation rate is 100%. What is the expected cost *per payment* at this inflation rate after one year?

Solution S4C27-5. We use the formula $(1+r)(E(X) - E(X \wedge d/(1+r))/(1 - F(d/(1+r)))$.
However, we first need to verify that $F(d/(1+r)) < 1$.
Here, $d/(1+r) = 500/(1+1) = 250$.
$F(d/(1+r)) = 1 - S(d/(1+r)) = 1 - S(250) = 1 - 1000^3/(250 + 1000)^3 = 0.488 < 1$, so the formula applies.
We find $E(X) = \theta/(\alpha - 1) = 1000/(3-1) = 500$.
We find $E(X \wedge d/(1+r)) = (\theta/(\alpha - 1))(1 - (\theta/(d/(1+r) + \theta))^{(\alpha - 1)}) = 500*(1 - 1000^2/(250 + 1000)^2) =$ 180.
Thus, our desired answer is 2(500 - 180)/(1-0.488) = **1250.**

Section 28

Policy Limits and Associated Effects of Inflation

If an insurance policy has a **policy limit,** then the policy only pays the full loss up to magnitude u and then pays only u for losses greater than u. A policy limit corresponds to a right-censored payment random variable. Let X be the loss random variable. Let Y be the random variable after a policy limit is applied. Then the following is true:

$F_Y(y) = F_X(y)$ if $y < u$;
$F_Y(y) = 1$ if $y \geq u$.
$f_Y(y) = f_X(y)$ if $y < u$;
$f_Y(y) = 1 - F_X(y)$ if $y = u$.

The expected cost of a policy with a limit of u is $E(Y) = E(X \wedge u)$. If there is a uniform rate of inflation r, then the expected cost is $E(Y) = (1+r)E(X \wedge u/(1+r))$.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 8, pp. 187.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C28-1. Losses from falling cows follow an exponential distribution with mean θ = 340. Inverse Falling Cow Insurance Company requires policyholders to have a policy limit of 200. Let Y be the payment random variable after the policy limit has been applied. Find $f_Y(200)$.

Relevant properties for exponential distributions: $S_X(x) = e^{-x/\theta}$;

$E(X) = \theta$; $E(X \wedge K) = \theta(1 - e^{-K/\theta})$.

Solution S4C28-1. We use the formula $f_Y(y) = 1 - F_X(y)$ if $y = u$. Here, u = 200.

$1 - F_X(y) = S_X(y) = S_X(200) = e^{-200/340}$ = $\mathbf{f_Y(200) = 0.555306373}$.

Problem S4C28-2. Losses from falling cows follow an exponential distribution with mean θ = 340. Inverse Falling Cow Insurance Company requires policyholders to have a policy limit of 200. Let Y be the payment random variable after the policy limit has been applied. Find E(Y).

Relevant properties for exponential distributions: $S_X(x) = e^{-x/\theta}$;
$E(X) = \theta$; $E(X \wedge K) = \theta(1 - e^{-K/\theta})$.

Solution S4C28-2. The expected cost of a policy with a limit of u is E(Y) = E(X Λ u). Here, E(X Λ 200) = $340(1 - e^{-200/340})$ = **E(Y) = 151.1958332**.

Problem S4C28-3. Losses from falling cows follow an exponential distribution with mean θ = 340. Inverse Falling Cow Insurance Company requires policyholders to have a policy limit of 200. Now the Fed begins to print vast amounts of money, and the annual uniform rate of inflation is 30%. Let Y be the payment random variable after the policy limit has been applied. What is the expected cost of the policy after one year at this inflation rate?

Relevant properties for exponential distributions: $S_X(x) = e^{-x/\theta}$;
E(X) = θ; E(X Λ K) = $\theta(1 - e^{-K/\theta})$.

Solution S4C28-3. We use the formula E(Y) = (1+r)E(X Λ u/(1+r)) =

1.3*E(X Λ 200/1.3) = 1.3*E(X Λ 153.8461538) = $1.3*340(1 - e^{-153.8461538/340})$ = **E(Y) = 160.8688759**.

Problem S4C28-4. Losses from house-smashing snakes (HSSs) follow a Pareto distribution with α = 3 and θ = 1000. The House-Smashing Snake Safety Service (HSSSS) provides insurance against this peril with a limit of 680. Let Y be the payment random variable after the policy limit has been applied. Find E(Y).

Relevant properties for Pareto distributions: $S(x) = \theta^\alpha/(x + \theta)^\alpha$.
E(X) = θ/(α - 1)
E(X Λ K) = $(\theta/(\alpha - 1))(1 - (\theta/(K + \theta))^{(\alpha - 1)})$

Solution S4C28-4. The expected cost of a policy with a limit of u is E(Y) = E(X Λ u). Here, u = 680, so E(Y) = $(1000/(3 - 1))(1 - (1000/(680 + 1000))^{(3 - 1)})$ = **E(Y) = 322.845805**.

Problem S4C28-5. Losses from house-smashing snakes (HSSs) follow a Pareto distribution with α = 3 and θ = 1000. The House-Smashing Snake Safety Service (HSSSS) provides insurance against this peril with a limit of 680. The Fed has gone mad and quintupled the money supply within one year, so that the inflation rate is 400%. Let Y be the payment random variable after the policy limit has been applied. What is the expected cost of the policy after one year at this inflation rate?

Relevant properties for Pareto distributions: $S(x) = \theta^\alpha/(x + \theta)^\alpha$.

E(X) = θ/(α - 1)
E(X Λ K) = $(\theta/(\alpha - 1))(1 - (\theta/(K + \theta))^{(\alpha - 1)})$

Solution S4C28-5. The formula to use here is E(Y) = (1+r)E(X Λ u/(1+r)), where r = 4, and so u/(1+r) = 680/5 = 136.

Thus, our answer will be 5*E(X Λ 136) = $5*(1000/(3 - 1))(1 - (1000/(136 + 1000))^{(3 - 1)})$ = **E(Y) = 562.760365**.

Section 29

Coinsurance and Treatment Thereof in Combination with Ordinary Deductibles, Policy Limits, and Inflation

In an arrangement of **coinsurance,** an insurance company pays a proportion of the loss, denoted α, and the policyholder pays $(1-\alpha)$ of the loss, i.e., the remainder. For a policy with coinsurance and loss random variable X, the random variable for the coinsurance payment becomes $Y = \alpha X$.

This random variable is treated as any other multiple of a random variable, and this treatment is discussed in depth in Section 13.

A useful set of formulas applies to a case where a policy has all of the following: coinsurance with proportion α, an ordinary deductible of amount d, and a limit of amount u. There is also inflation at a uniform rate r. Note that for the equations below to apply, the coinsurance must apply *last*, after the deductible and limit have already been considered. In that case, the per-loss payment random variable is denoted Y^L and defined as follows:

$\mathbf{Y^L = 0}$ **when** $\mathbf{X < d/(1+r)}$.

$\mathbf{Y^L = \alpha((1+r)X - d)}$ **when** $\mathbf{d/(1+r) \leq X < u/(1+r)}$.

$\mathbf{Y^L = \alpha(u - d)}$ **when** $\mathbf{u/(1+r) \leq X}$.

$\mathbf{E(Y^L) = \alpha(1+r)(E(X \wedge u/(1+r)) - E(X \wedge d/(1+r)))}$.

$\mathbf{E((Y^L)^2) = \alpha^2(1+r)^2(E((X \wedge u/(1+r))^2) - E((X \wedge d/(1+r))^2) - 2(d/(1+r))E(X \wedge u/(1+r)) + 2(d/(1+r))E(X \wedge d/(1+r)))}$

For the per-payment payment random variable, Y^P, the following is true.

$\mathbf{E(Y^P) = E(Y^L)/(1-F_X(d/(1+r)))}$.

$\mathbf{E((Y^P)^2) = E((Y^L)^2)/(1-F_X(d/(1+r)))}$.

Source: *Loss Models: From Data to Decisions,* (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 8, pp. 189-190.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C29-1. Losses from falling cows follow an exponential distribution with mean θ = 340. Sophisticated Cow Insurance Company has a policy in place with the following features:

- A deductible of 100;
- A policy limit of 400;
- Coinsurance of 60%, applied after the deductible and limit have been considered.

Now the Fed begins to print vast amounts of money, and the annual rate of inflation is 30%.

After one year, an insured under this policy suffers a falling cow loss of 230. What is the value of Y^L, the per-loss payment random variable associated with this loss?

Solution S4C29-1. Here, u = 400, d = 100, r = 0.3, and α = 0.6. We find u/(1+r) = 400/1.3 = 307.6923077 and d/(1+r) = 100/1.3 = 76.92307692.

The loss X = 230, which is between d/(1+r) and u/(1+r). Thus, the following formula applies:

$Y^L = \alpha((1+r)X - d) = 0.6((1.3)230-100)$ = $\mathbf{Y^L = 119.4}$.

Problem S4C29-2. Losses from falling cows follow an exponential distribution with mean θ = 340. Sophisticated Cow Insurance Company has a policy in place with the following features:

- A deductible of 100;
- A policy limit of 400;
- Coinsurance of 60%, applied after the deductible and limit have been considered.

Now the Fed begins to print vast amounts of money, and the annual rate of inflation is 30%.

Find $E(Y^L)$ for this policy after one year.

Relevant properties for exponential distributions: $S_X(x) = e^{-x/\theta}$;

$E(X) = \theta$; $E(X \wedge K) = \theta(1 - e^{-K/\theta})$.

Solution S4C29-2. Here, u = 400, d = 100, r = 0.3, and α = 0.6.

We use the formula $E(Y^L) = \alpha(1+r)(E(X \wedge u/(1+r)) - E(X \wedge d/(1+r)))$.

From Solution S4C29-1, we know that d/(1+r) = 76.92307692 and u/(1+r) = 307.6923077.

$E(X \wedge u/(1+r)) = E(X \wedge 307.6923077) = 340(1 - e^{-307.6923077/340}) = 202.4526472$.

$E(X \wedge d/(1+r)) = E(X \wedge 76.92307692) = 340(1 - e^{-76.92307692/340}) = 68.8421093$.

Thus, $E(Y^L) = 0.6*1.3(202.4526472-68.8421093)$ = $\mathbf{E(Y^L) = 104.2162196}$.

Problem S4C29-3. Losses from falling cows follow an exponential distribution with mean θ = 340. Sophisticated Cow Insurance Company has a policy in place with the following features:

- A deductible of 100;
- A policy limit of 400;
- Coinsurance of 60%, applied after the deductible and limit have been considered.

Now the Fed begins to print vast amounts of money, and the annual rate of inflation is 30%.

Find $E(Y^P)$ for this policy after one year.

Relevant properties for exponential distributions: $S_X(x) = e^{-x/\theta}$;

$E(X) = \theta$; $E(X \wedge K) = \theta(1 - e^{-K/\theta})$.

Solution S4C29-3. We use the formula $E(Y^P) = E(Y^L)/(1-F_X(d/(1+r)))$.

We know from Solution S4C29-2 that $E(Y^L) = 104.2162196$.

We also know from Solution S4C29-1 that $d/(1+r) = 76.92307692$.

$1-F_X(d/(1+r)) = S_X(d/(1+r)) = e^{-76.92307692/340} = 0.7975232079$.

Thus, $E(Y^P) = 104.2162196/0.7975232079$ = $\mathbf{E(Y^P) = 130.6748425}$.

Problem S4C29-4. Losses from falling cows follow an exponential distribution with mean θ = 340. Sophisticated Cow Insurance Company has a policy in place with the following features:

- A deductible of 100;
- A policy limit of 400;
- Coinsurance of 60%, applied after the deductible and limit have been considered.

Now the Fed begins to print vast amounts of money, and the annual rate of inflation is 30%.

Find $Var(Y^L)$ for this policy after one year. You may use the following Excel input format for any incomplete Gamma function computations required in this problem:

"=GAMMADIST(x, α, 1, TRUE)"

Relevant properties for exponential distributions: $S_X(x) = e^{-x/\theta}$;

$E(X) = \theta$; $E(X \wedge K) = \theta(1 - e^{-K/\theta})$; $E((X \wedge K)^2) = 2\theta^2*\Gamma(3; K/\theta) + K^2e^{-K/\theta}$.

Solution S4C29-4. $Var(Y^L) = E((Y^L)^2)- E((Y^L))^2$.

We know from Solution S4C29-2 that $E(Y^L) = 104.2162196$.

Thus, $E((Y^L))^2 = 10861.02042$.

To find $E((Y^L)^2)$, we use the formula $E((Y^L)^2) = \alpha^2(1+r)^2(E((X \wedge u/(1+r))^2)- E((X \wedge d/(1+r))^2) - 2(d/(1+r))E(X \wedge u/(1+r)) + 2(d/(1+r))E(X \wedge d/(1+r)))$

From Solution S4C29-2, we know that $E(X \wedge u/(1+r)) = 202.4526472$ and $E(X \wedge d/(1+r)) = 68.8421093$.

From Solution S4C29-1, we know that $d/(1+r) = 76.92307692$ and $u/(1+r) = 307.6923077$.

We need to find $E((X \wedge u/(1+r))^2) = (E(X \wedge 307.6923077)^2) =$

$2*340^2*\Gamma(3; 307.6923077/340) + 307.6923077^2e^{-307.6923077/340}$.

To find $\Gamma(3; 307.6923077/340)$, we use the Excel input

"=GAMMADIST(307.6923077/340, 3, 1, TRUE)", which gives the result of 0.063679002. Thus,

$(E(X \wedge 307.6923077)^2) = 2*340^2*0.063679002 + 38300.68996 = E((X \wedge u/(1+r))^2) = 53023.27523$.

We need to find $E((X \wedge d/(1+r))^2) = (E(X \wedge 76.92307692)^2) =$

$2*340^2*\Gamma(3; 76.92307692/340) + 76.92307692^2e^{-76.92307692/340}$.

To find $\Gamma(3; 76.92307692/340)$, we use the Excel input

"=GAMMADIST(76.92307692/340, 3, 1, TRUE)", which gives the result of 0.001630465.

Thus, $(E(X \wedge 76.92307692)^2) = 2*340^2*0.001630465 + 4719.072236 = E((X \wedge d/(1+r))^2) = 5096.035744$.

Therefore, $E((Y^L)^2) = 0.6^2(1.3)^2(53023.27523 - 5096.035744 - 2*76.92307692*202.4526472 + 2*76.92307692*68.8421093) = E((Y^L)^2) = 16652.98616$.

Thus, $Var(Y^L) = 16652.98616 - 10861.02042 =$ **$Var(Y^L) = 5719.965736$**.

Problem S4C29-5. Losses from falling cows follow an exponential distribution with mean $\theta = 340$. Sophisticated Cow Insurance Company has a policy in place with the following features:

- A deductible of 100;
- A policy limit of 400;
- Coinsurance of 60%, applied after the deductible and limit have been considered.

Now the Fed begins to print vast amounts of money, and the annual rate of inflation is 30%.

Find $\mathrm{Var}(Y^P)$ for this policy after one year.

Solution S4C29-5. $\mathrm{Var}(Y^P) = E((Y^P)^2) - E((Y^P))^2$.

We know from Solution S4C29-3 that $E(Y^P) = 130.6748425$.

Thus, $E((Y^P))^2 = 17075.91446$.

We now find $E((Y^P)^2) = E((Y^L)^2)/(1-F_X(d/(1+r)))$.

We know from Solution S4C29-4 that $E((Y^L)^2) = 16652.98616$.

We know from Solution S4C29-3 that $1-F_X(d/(1+r)) = 0.7975232079$

Thus, $E((Y^P)^2) = 16652.98616/0.7975232079 = 20880.8797$.

Hence, $\mathrm{Var}(Y^P) = 20880.8797 - 17075.91446 =$ **$\mathrm{Var}(Y^P) = 3804.965237$**.

Section 30

Impact of a Deductible on the Number of Payments

The existence of a deductible in an insurance policy will affect the frequency of payments under that policy. Let X be the loss amount random variable. Let d be a deductible, and let v be the probability that a loss results in a payment. That is, $v = \Pr(X > d)$. Let N^L be the random variable denoting the *number* of losses, and let N^P be the random variable denoting the *number* of payments. We can compare the probability generating functions (pgfs) of these two random variables as follows:

$\mathbf{P_{N^\wedge P}(z) = P_{N^\wedge L}(1 + v(z-1))}$

The above is true generally. There is also an important special case in which $P_{N^\wedge L}(z)$ is a function of some parameter θ, such that $P_{N^\wedge L}(z; \theta) = B(\theta(z-1))$, where the function B(z) is itself entirely independent of θ. In that case, the following is true: $\mathbf{P_{N^\wedge P}(z) = P_{N^\wedge L}(z; v\theta) = B(v\theta(z-1))}$.

It is also possible to obtain the pgf of N^L by knowing the pdf of N^P. This can be done as follows:

$\mathbf{P_{N^\wedge L}(z) = P_{N^\wedge P}(1 - v^{-1} + zv^{-1})}$.

If $P_{N^\wedge P}(z)$ is a function of some parameter θ, then $\mathbf{P_{N^\wedge L}(z) = P_{N^\wedge P}(z; \theta/v)}$.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 8, pp. 192-196.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C30-1. Losses from falling cows follow an exponential distribution with mean θ = 340. The *number* of losses for an insured per year follows a Poisson distribution with mean λ = 0.5. The Simple Cow Insurance Company initially does not require any deductible from policyholders insured from this peril. However, the imposition of an ordinary deductible of 200 is being contemplated. What is v, the probability that a loss will exceed this deductible?

Relevant properties for exponential distributions: $S_X(x) = e^{-x/\theta}$.

Solution S4C30-1. We find $v = \Pr(X > d)$, where d = 200. Essentially, $v = S_X(200) = e^{-200/340} =$ **v = 0.555306373**.

Problem S4C30-2. Losses from falling cows follow an exponential distribution with mean θ = 340. The *number* of losses for an insured per year follows a Poisson distribution with mean λ =

0.5. The Simple Cow Insurance Company initially does not require any deductible from policyholders insured from this peril. However, the imposition of an ordinary deductible of 200 is being contemplated. What would the pgf of the number of payments be if this deductible were imposed?

Relevant properties for Poisson distributions: $P(z) = e^{\lambda(z-1)}$.

Solution S4C30-2. We want to find $P_{N^{\wedge}P}(z)$, which we will do using the formula

$P_{N^{\wedge}P}(z) = P_{N^{\wedge}L}(1 + v(z-1))$.

Here, $\lambda = 0.5$, so $P_{N^{\wedge}L}(z) = e^{0.5(z-1)}$.

Thus, $P_{N^{\wedge}L}(1 + v(z-1)) = e^{0.5(1+v(z-1)-1)} = e^{0.5(v(z-1))}$, where v, from Solution S4C30-1, is 0.555306373.

Thus, $P_{N^{\wedge}P}(z) = e^{0.5(0.555306373(z-1))} =$ $\mathbf{P_{N^{\wedge}P}(z) = e^{0.2776531865(z-1)}}$.

Problem S4C30-3. Losses from falling cows follow an exponential distribution with mean θ = 340. The *number* of losses for an insured per year follows a Poisson distribution with mean λ = 0.5. The Simple Cow Insurance Company initially does not require any deductible from policyholders insured from this peril. However, the imposition of an ordinary deductible of 200 is being contemplated. If this deductible were imposed, what fraction of policyholders would be expected to have at least one claim paid to them?

Relevant properties for Poisson distributions: $P(z) = e^{\lambda(z-1)}$.

$p_k = e^{-\lambda} * \lambda^k / k!$

Solution S4C30-3. We want to find $\Pr(N_P \geq 1)$. From Solution S4C30-2, we know that

$P_{N^{\wedge}P}(z) = e^{0.2776531865(z-1)}$. This is the pgf of a Poisson distribution with a mean $\lambda = 0.2776531865$.

Thus, $\Pr(N_P \geq 1) = 1 - \Pr(N_P = 0) = 1 - e^{-0.2776531865} =$ $\mathbf{\Pr(N_P \geq 1) = 0.2424404922}$.

Problem S4C30-4. Losses from house-smashing snakes (HSSs) follow a Pareto distribution with $\alpha = 3$ and $\theta = 1000$. HSS Mutual Insurance Company requires policyholders to have an ordinary deductible of 500. The *number* of losses for an insured per year follows a geometric distribution with $\beta = 0.4$. Find $P_{N^{\wedge}P}(z)$ for this policy.

Relevant properties for Pareto distributions: $S(x) = \theta^\alpha/(x + \theta)^\alpha$.

Relevant properties for geometric distributions: $P(z) = (1 - \beta(z-1))^{-1}$.

Solution S4C30-4. First, we find $v = \Pr(X > d) = S_X(500) = 1000^3/(500 + 1000)^3 = v = 8/27$.

Here, $P_{N^\wedge L}(z) = (1 - \beta(z-1))^{-1} = (1 - 0.4(z-1))^{-1}$

We note that this is a case where the formula $P_{N^\wedge L}(z; \theta) = B(\theta(z-1))$ applies, with β serving the role of the parameter θ (not to be confused with the θ from the Pareto distribution above!).

Thus, $P_{N^\wedge L}(z; \beta) = P_{N^\wedge L}(z; 0.4)$, so $P_{N^\wedge P}(z) = P_{N^\wedge L}(z; v\beta) = P_{N^\wedge L}(z; (8/27)*0.4) = P_{N^\wedge L}(z; 16/135)$ =

$P_{N^\wedge P}(z) = (1 - 16(z-1)/135)^{-1}$.

Problem S4C30-5. Losses from house-smashing snakes (HSSs) follow a Pareto distribution with α = 3 and θ = 1000. HSS Mutual Insurance Company requires policyholders to have an ordinary deductible of 500. The *number* of losses for an insured per year follows a geometric distribution with β = 0.4. What fraction of policyholders will incur losses but be unable to file claims under this policy, because the losses of these policyholders are not in excess of the deductible?

Relevant properties for geometric distributions: $p_k = \beta^k/(1+\beta)^{k+1}$.

Solution S4C30-5. We want to find $\Pr(N_P = 0 \text{ and } N_L \neq 0)$.

We can think of this value in the following way:
$\Pr(N_L \neq 0) = \Pr(N_P = 0 \text{ and } N_L \neq 0) + \Pr(N_P \neq 0 \text{ and } N_L \neq 0)$.

If $N_P \neq 0$ and $N_L \neq 0$, then this is the same as simply saying that $N_P \neq 0$, since a payment cannot occur without a loss. Thus,

$\Pr(N_L \neq 0) = \Pr(N_P = 0 \text{ and } N_L \neq 0) + \Pr(N_P \neq 0) \rightarrow$

$\Pr(N_P = 0 \text{ and } N_L \neq 0) = \Pr(N_L \neq 0) - \Pr(N_P \neq 0) \rightarrow$

$\Pr(N_P = 0 \text{ and } N_L \neq 0) = 1 - \Pr(N_L = 0) - (1 - \Pr(N_P = 0)) \rightarrow$

$\Pr(N_P = 0 \text{ and } N_L \neq 0) = \Pr(N_P = 0) - \Pr(N_L = 0)$.

We know from Solution S4C30-4 that $P_{N^\wedge P}(z) = (1 - 16(z-1)/135)^{-1}$. Thus, N_P follows a geometric distribution with β = 16/135.

We know from Solution S4C30-4 that $P_{N^\wedge L}(z) = (1 - 0.4(z-1))^{-1}$. Thus, N_L follows a geometric distribution with β = 0.4.

Therefore, $\Pr(N_P = 0) = 1/(1+16/135) = 0.8940397351$.

$\Pr(N_L = 0) = 1/(1+0.4) = 0.7142857143$.

Hence, our desired answer is 0.8940397351 - 0.7142857143 = **0.1797540208**.

Section 31

Assorted Exam-Style Questions for Exam 4 / Exam C – Part 2

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C31-1. Similar to Question 54 of the Exam C Sample Questions from the Society of Actuaries. The following data is obtained regarding a sample of 1000 gambling losses (X) at a casino:

Range of Losses..........Number of Losses
$0 \le x \le 25$...........................124
$25 < x \le 50$.........................174
$50 < x \le 100$.......................195
$100 < x \le 320$.....................185
$320 < x \le 3200$...................322

You are using an exponential distribution with mean θ to model the data in this sample. Estimate θ by using percentile matching at the 49.3rd percentile. Assume that losses are uniformly distributed within each range specified.

Solution S4C31-1. We note that 124+174+195 = 493, so the 493rd loss of 1000 occurs at x = 100. Thus, we want to find θ such that $S_X(100) = 1\text{-}0.493 = 0.507 = e^{-100/\theta} \rightarrow \theta = -100/\ln(0.507)$ = **θ = 147.2224406**.

Problem S4C31-2. Similar to Question 57 of the Exam C Sample Questions from the Society of Actuaries. You have collected the following data on a sample of 100 giraffes:

Height of Giraffe (cm.).............Number of Giraffes
0-300...34
300-400...14
400-500...36
500-600...16

Assume that height is uniformly distributed within each specified range. Estimate the third raw moment of this distribution of giraffe heights.

Solution S4C31-2. For each specified range, we use the following formula from Section 1:

$\mu'_k = E(X^k) = \int_{-\infty}^{\infty} x^k * f(x)dx$, with the bounds modified according to each range. Here, k = 3.

The third raw moment will be the probability-weighted average of four integrals, with one integral pertaining to each of the four ranges.

For the 0-300 range the integral is $\int_0^{300} x^3*(1/300)dx = x^4/1200 \Big|_0^{300} = 6750000$.

For the 300-400 range the integral is $\int_{300}^{400} x^3*(1/100)dx = x^4/400 \Big|_{300}^{400} = 43750000$.

For the 400-500 range the integral is $\int_{400}^{500} x^3*(1/100)dx = x^4/400 \Big|_{400}^{500} = 92250000$.

For the 500-600 range the integral is $\int_{500}^{600} x^3*(1/100)dx = x^4/400 \Big|_{500}^{600} = 167750000$.

Thus, the third raw moment is (34/100)*6750000 + (14/100)*43750000 + (36/100)*92250000 + (16/100)*167750000 = $\boldsymbol{\mu'_3 = 68{,}470{,}000}$.

Problem S4C31-3. Similar to Question 71 of the Exam C Sample Questions from the Society of Actuaries. Every year, some people come to disbelieve some popular superstitions. You are performing a test of the null hypothesis that the observed frequencies of the numbers of popular superstitions disbelieved this year accord with the historical frequencies. This is the data you gathered.

Number of Superstitions Disbelieved
0: 240 people
1: 23 people
2: 5 people
3+: 2 people

Historically, in any given year, 0.93 of the population will not abandon any of their superstitions in a given year; 0.04 of the population will abandon 1 superstition, 0.02 of the population will abandon 2 superstitions, and 0.01 of the population will abandon 3 or more superstitions.

What is the result of the hypothesis test? Use the table of values for the Chi-square distribution associated with the Exam 4 / C Tables (p. 4).

(a) Reject at the 0.005 significance level.

(b) Reject at the 0.010 significance level, but not at the 0.005 level.

(c) Reject at the 0.025 significance level, but not at the 0.010 level.

(d) Reject at the 0.050 significance level, but not at the 0.025 level.

(e) Do not reject at the 0.050 significance level.

Solution S4C31-3.

The sample contains a total of 240 + 23 + 5 + 2 = 270 members. Since there are four categories, the Chi-square statistic will have 4-1 = 3 degrees of freedom.

First, we find the Chi-square statistic: $\chi^2 = {}_{i=1}{}^{t}\sum((X_i - np_i)^2/np_i)$.

$(240 - 0.93*270)^2/(0.93*270) + (23 - 0.04*270)^2/(0.04*270) + (5 - 0.02*270)^2/(0.02*270) + (2 - 0.01*270)^2/(0.01*270) = 14.4832736$.

We examine p. 4 of the Exam 4 / C Tables in the row corresponding to 3 degrees of freedom, and we find that the Chi-square statistic value for P = 0.995 (significance level 0.005) is 12.838. Since 14.4832736 > 12.838, **we reject the null hypothesis at the 0.005 significance level**. This means that the correct answer is **(a)**.

Problem S4C31-4. Similar to Question 81 of the Exam C Sample Questions from the Society of Actuaries. There is a 0.36 probability that X follows a Pareto distribution with $\alpha = 4$ and $\theta = 99$. There is a 0.64 probability that X follows an exponential distribution with $\theta = 150$. You estimate X using the following two-step procedure:

1. A random number from 0 to 1 is generated, indicating to which distribution in this two-point mixture X belongs. The lower random numbers get assigned to the exponential distributions.

2. Another random number from 0 to 1 is generated, where lower numbers correspond to lower values of X, and the percentile of the random number within the [0, 1] interval corresponds to the percentile of X within the distribution.

One such application of this procedure yields the following randomly generated numbers, in order: 0.67, 0.35. What is the estimate x of X corresponding to these random numbers?

Relevant properties for exponential distributions: $S_X(x) = e^{-x/\theta}$.

Relevant properties for Pareto distributions: $S(x) = \theta^\alpha/(x + \theta)^\alpha$.

Solution S4C31-4. The lower random numbers get assigned to the exponential distributions, and the exponential distribution has a probability of 0.64 of occurring. Since the first randomly generated number is 0.67 > 0.64, X here follows the Pareto distribution.

The second randomly generated number of 0.35 indicates that x is at the 35^{th} percentile of this Pareto distribution. Hence, $S_X(x) = 1 - 0.35 = 0.65 = 99^4/(x + 99)^4 \rightarrow (x + 99)^4 = 99^4/0.65 \rightarrow x = 99/0.65^{0.25} - 99 =$ **x = 11.25717363**.

Problem S4C31-5. Similar to Question 84 of the Exam C Sample Questions from the Society of Actuaries. In an attempt to discourage gambling, the Gamblers' Charity Fund pays gamblers an incentive payment I, consisting of certain fraction p of the amount by which their annual gambling losses are less than $1000. It is known that E(I) = 200. Gambling losses (X) follow an exponential distribution with θ = 1400. Find the value of p.

Relevant properties for exponential distributions: $E(X \wedge K) = \theta(1 - e^{-K/\theta})$.

Solution S4C31-5.

$I = p(1000 - x)$ if $0 < x < 1000$;

$I = 0$ if $x \geq 1000$.

Thus, $E(I) = p(1000 - E(X \wedge 1000))$.

$E(I) = p(1000 - 1400(1 - e^{-1000/1400}))$.

$E(I) = 285.3583234p$.

$200 = 285.3583234p$

Thus, $p = 0.7008731956$.

Section 32

The Collective Risk Model and the Individual Risk Model

The following description of two types of risk models paraphrases Klugman, Panjer, and Willmot 2008, p. 200.

The **collective risk model** represents the aggregate losses S as a sum of a random number N of the individual payment amounts X_1, X_2, ..., X_N. Thus, in the collective risk model,

$S = X_1 + X_2 + ... + X_N$ for nonnegative integer values of N, and $S = 0$ when $N = 0$.

N is called the **claim count random variable.**

Each of the subscripted Xs are called the **individual** or **single-loss random variables**

S is called the **aggregate loss random variable** or **total loss random variable.**

All of the subscripted Xs are independent and identically distributed (i.i.d.) random variables. The following assumptions are true in the collective risk model:

1. Conditional on $N = n$, the variables X_1, X_2, ..., X_n are i.i.d. random variables.
2. Conditional on $N = n$, the common distribution of the random variables X_1, X_2, ..., X_n does not depend on n.
3. The distribution of N does not depend in any way on the values of X_1, X_2, ...

The **individual risk model** represents the aggregate loss as a sum, **$S = X_1 + X_2 + ... + X_n$**of a fixed number of insurance contracts, n. The random variables X_1, X_2, ..., X_n are assumed to be independent but are not necessarily identically distributed. The distribution of each of these subscripted Xs usually has a probability mass at zero, corresponding to the probability of no loss or payment.

It is best for the severity distribution (the distribution of the subscripted Xs) to be a scale distribution in order to model the effects of inflation and currency conversions.

For the frequency distribution (the distribution of N), it is best to have a distribution with the following probability generating function: $P_N(z; \alpha) = Q(z)^\alpha$, where Q(z) is some function of z independent of α. This implies that if the volume of business of the insurance company increases by 100r%, expected claims will increase in a manner proportional to (1+r)α. Zero-modified distributions are not of this form, but it may still be desirable to use them in circumstances where some insurance policies never result in a claim.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 9, p. 200-203.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C32-1. An insurance company insures 6000 miners against catastrophic injuries. The number of miners insured is fixed at 6000. Which of the following risk models is better suited for evaluating the risk to the insurance company?

(a) A collective risk model
(b) An individual risk model

Solution S4C32-1. A collective risk model does not specify the number n of the insurance contracts to be issued, whereas an individual risk model does. Here, n is fixed at 6000. Therefore, an individual risk model is better suited in this case. Thus, the correct answer is **(b) An individual risk model**.

Problem S4C32-2. An insurance company insures rabbits against being eaten by wolves. Wolves do not discriminate among rabbits, and no rabbit is better than any other at evading wolves. Thus, the probability of being eaten by a wolf is the same for every rabbit. Moreover, the insurance company will pay the same amount for a loss, irrespective of which rabbit gets eaten. Which of the following risk models is better suited for evaluating the risk to the insurance company?

(a) A collective risk model
(b) An individual risk model

Solution S4C32-2. The random variable corresponding to losses is the same for each rabbit. Thus, the loss random variables are i.i.d., which is the case for a collective risk model, but not necessarily for an individual risk model. Moreover, the number of rabbits to be insured is not specified, as it must be for an individual risk model. Thus, a collective risk model is preferable in this case, and the correct answer is **(a) A collective risk model**.

Problem S4C32-3. Which of the following statements are true? More than one answer may be correct.

(a) In the individual risk model, all of the random variables X_1, X_2, ..., X_n must have a positive probability at zero.
(b) In the collective risk model, the frequency and severity of payments are modeled separately.
(c) The individual risk model is always a special case of the collective risk model.
(d) The individual risk model is a special case of the collective risk model, but only if the random variables X_1, X_2, ..., X_n in the individual risk model are i.i.d..
(e) In the collective risk model, the number of payments affects the size of each individual payment.

(f) The individual risk model only applies when there is only a single insured; the collective risk model applies when there are more than one insureds.

Solution S4C32-3.

(a) is false: In the individual risk model, the random variables do not have to be i.i.d. While most variables will have a probability mass at zero, not all of them need to.

(b) is true: In the collective risk model, N (the frequency random variable) is determined independently of each of the subscripted Xs, the severity random variables.

(c) is false: The individual risk model is not a special case of the collective risk model when the random variables in the individual risk model are not i.i.d.

(d) is true: If the random variables in the individual risk model are i.i.d, then the individual risk model is a special case of the collective risk model, where the probability distribution of N is such that Pr(N = n) = 1.

(e) is false: If frequency is independent of severity, as it is in the collective risk model, then this implies that the number of payments *does not* affect the size of each individual payment.

(f) is false: The individual risk model can apply to multiple insureds, provided there is a fixed number of them.

Thus, **(b) and (d) are true.**

Problem S4C32-4. Let N, the claim count random variable, follow a Poisson distribution with mean λ. Does the Poisson distribution meet the criterion that $P_N(z; \alpha) = Q(z)^\alpha$?

Relevant properties for Poisson distributions: $P(z) = e^{\lambda(z-1)}$.

Solution S4C32-4. We consider $P(z) = e^{\lambda(z-1)}$ in terms of the parameter λ being α.
$e^{\lambda(z-1)} = (e^{(z-1)})$^λ, where $e^{(z-1)} = Q(z)$, and λ = α. Thus, **the Poisson distribution satisfies the criterion that $P_N(z; \alpha) = Q(z)^\alpha$.**

Problem S4C32-5. Let N, the claim count random variable, follow a negative binomial distribution with parameters r and β. Does the negative binomial distribution meet the criterion that $P_N(z; \alpha) = Q(z)^\alpha$?

Relevant properties for negative binomial distributions: $P(z) = (1 - \beta(z-1))^{-r}$.

Solution S4C32-5. We can consider $P(z) = (1 - \beta(z-1))^{-r}$ in terms of the parameter r being α.

Then $(1 - \beta(z-1))^{-r} = ((1 - \beta(z-1))^{-1})$^r, where $(1 - \beta(z-1))^{-1} = Q(z)$ and r = α. Thus, **the negative binomial distribution satisfies the criterion that $P_N(z; \alpha) = Q(z)^\alpha$.**

Section 33

Mean, Variance, Third Central Moment, and Probability Calculations Pertaining to Aggregate Loss Random Variables

Let N be the frequency random variable for losses.

Let X be the severity random variable for losses.

Let S be the aggregate loss random variable.

Then the following equations are true:
$E(S) = \mu'_{S1} = \mu'_{N1}*\mu'_{X1} = E(N)E(X)$.

$Var(S) = \mu_{S2} = \mu'_{N1}*\mu_{X2} + \mu_{N2}*(\mu'_{X1})^2 = E(N)*Var(X) + Var(N)*E(X)^2$.

$E((S-E(S))^3) = \mu_{S3} = \mu'_{N1}*\mu_{X3} + 3\mu_{N2}*\mu'_{N1}*\mu_{X2} + \mu_{N3}*(\mu'_{X1})^3$.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 9, pp. 203-205.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C33-1. Loss frequency N follows a Poisson distribution with $\lambda = 55$. Loss severity X follows an exponential distribution with $\theta = 200$. Find E(S), the expected value of the aggregate loss random variable.

Solution S4C33-1. We use the formula E(S) = E(N)E(X). For a Poisson distribution, the expected value is λ; for an exponential distribution, the expected value is θ. Thus, $E(S) = \lambda\theta = 55*200 =$ **E(S) = 11000**.

Problem S4C33-2. Loss frequency N follows a Poisson distribution with $\lambda = 55$. Loss severity X follows an exponential distribution with $\theta = 200$. Find Var(S), the expected value of the aggregate loss random variable.

Relevant properties of Poisson distributions: $E(N) = \lambda$; $Var(N) = \lambda$.

Relevant properties of exponential distributions: $E(X) = \theta$; $Var(X) = \theta^2$.

Solution S4C33-2. We use the formula $Var(S) = E(N)*Var(X) + Var(N)*E(X)^2$.

Here, E(N) and Var(N) are both 55; $E(X)^2$ and Var(X) are both $200^2 = 40000$.

Thus, Var(S) = 55*40000 + 55*40000 = **Var(S) = 4,400,000**.

Problem S4C33-3. Loss frequency N follows a Poisson distribution with λ = 55. Loss severity X follows an exponential distribution with θ = 200. Find $E((S-E(S))^3)$ for the aggregate loss random variable S.

Relevant properties of Poisson distributions: E(N) = λ; Var(N) = λ; $E((N-E(N))^3) = \lambda$.

Relevant properties of exponential distributions: E(X) = θ; Var(X) = θ^2; $E(X^k) = k!*\theta^k$ if k is an integer.

Solution S4C33-3. We use the formula

$E((S-E(S))^3) = \mu_{S3} = \mu'_{N1}*\mu_{X3} + 3\mu_{N2}*\mu'_{N1}*\mu_{X2} + \mu_{N3}*(\mu'_{X1})^3$.

We know the following values:

$\mu'_{N1} = E(N) = \lambda = 55$;

$\mu_{N2} = Var(N) = \lambda = 55$;

$\mu_{X2} = Var(X) = \theta^2 = 40000$;

$\mu_{N3} = E((N-E(N))^3) = \lambda = 55$;

$\mu'_{X1} = E(X) = \theta = 200$.
We need to find $\mu_{X3} = E((X-E(X))^3) = E((X- \theta)^3) = E(X^3) - 3\theta E(X^2) + 3\theta^2 E(X) - \theta^3 =$

$6\theta^3 - 3\theta*2\theta^2 + 3\theta^2*\theta - \theta^3 = \mu_{X3} = 2\theta^3 =16000000$.

Thus, $\mu_{S3} = 55*16000000 + 3*55*55*40000 + 55*200^3$ = $\boldsymbol{\mu_{S3}}$ **= 1,683,000,000**.

Problem S4C33-4. Loss frequency N follows a Poisson distribution with λ = 55. Loss severity X follows an exponential distribution with θ = 200. Use a normal approximation to find the probability that S, the aggregate loss, will exceed 11500. To calculate Φ(x) for any value of x, use the Excel input "=NORMSDIST(x)".

Solution S4C33-4. Using a normal approximation,

Pr(S ≥ 11500) ≈ 1 - Φ((11500 - E(S))/√(Var(S))).
From Solution S4C33-1, we know that E(S) = 11000, and Var(S) = 4,400,000. Thus, √(Var(S)) = 2097.617696. Hence, Pr(S ≥ 11500) ≈ 1 - Φ((11500 - 11000)/2097.617696) =

1 - Φ(0.2383656474). To find this value, we use the Excel input "=1-NORMSDIST(0.2383656474)". The resulting approximation of Pr(S ≥ 11500) is **0.405798755**.

Problem S4C33-5. Loss frequency N follows a Poisson distribution with λ = 55. Loss severity X follows an exponential distribution with θ = 200. Use a *lognormal* approximation to find the probability that S, the aggregate loss, will exceed 11500. To calculate Φ(x) for any value of x, use the Excel input "=NORMSDIST(x)".

Note: For a lognormal distribution with mean μ and standard deviation σ,

$E(S) = \exp(\mu + 0.5\sigma^2)$ and $E(S^2) = \exp(2\mu + 2\sigma^2)$.

In doing the approximation, you will need to take the natural logarithm of the value of S that you are considering.

Solution S4C33-5. We first need to find μ and σ by setting up and solving a system of equations.

From Solution S4C33-1, we know that E(S) = 11000, and Var(S) = 4,400,000.

Thus, $E(S^2) = Var(S) + E(S)^2 = 4400000 + 11000^2 = 125400000$.

Thus, we have the following system of equations:

$11000 = \exp(\mu + 0.5\sigma^2) \rightarrow 9.305650552 = \mu + 0.5\sigma^2$;

$125400000 = \exp(2\mu + 2\sigma^2) \rightarrow 18.64701919 = 2\mu + 2\sigma^2$;

From this, it follows that $\sigma^2 = 2(9.305650552 - \mu)$ and so

$18.64701919 = 2\mu + 4(9.305650552 - \mu) \rightarrow$

$-18.57558302 = -2\mu \rightarrow$

$\mu = 9.28779151 \rightarrow$

$\sigma^2 = 2(9.305650552 - 9.28779151) = \sigma^2 = 0.0357180826 \rightarrow \sigma = 0.1889922819$.

Now we can conduct the lognormal approximation:

$\Pr(S \geq 11500) \approx 1 - \Phi((\ln(11500) - 9.28779151)/0.1889922819) =$

1 - Φ(0.3297002569). To find this value, we use the Excel input

"=1-NORMSDIST(0.3297002569)". The resulting approximation of Pr(S ≥ 11500) is **0.37081323**. This result differs somewhat from the result of using the normal approximation.

Section 34

Stop-Loss Insurance and the Net Stop-Loss Premium

Stop-loss insurance is "insurance on the aggregate losses, subject to a deductible" (Klugman, Panjer, and Willmot 2008, p. 208).

The **net stop-loss premium** is the expected cost of stop-loss insurance. It is equal to $E((S-d)_+)$, where S is the aggregate loss random variable, d is the deductible, and the $()_+$notation implies that the value S-d is only used if S-d > 0. Otherwise, 0 is used in the parentheses.

The following equation is true with respect to the net stop-loss premium for any aggregate distribution:

Equation 34.1. $E((S-d)_+) = \int_d^{\infty}(1-F_S(x))dx$.

The following equation is true for any *continuous* aggregate distribution:

Equation 34.2. $E((S-d)_+) = \int_d^{\infty}(x-d)f_S(x)dx$.

The following equation is true for any *discrete* aggregate distribution:

Equation 34.3. $E((S-d)_+) = \sum_{x>d}((x-d)f_S(x))$.

The following theorems are also true:

Theorem 34.4. Suppose $\Pr(a < S < b) = 0$. Then, for $a \le d \le b$,

$E((S-d)_+) = ((b-d)/(b-a))E((S-a)_+) + ((d-a)/(b-a))E((S-b)_+)$.

Theorem 34.5. Assume $\Pr(S = kh) = f_k \ge 0$ for some fixed $h > 0$, k is a nonnegative integer, and $\Pr(S = x) = 0$ for all other x. Then, provided $d = jh$, with j a nonnegative integer,

$E((S-d)_+) = h*\sum_{m=0}^{\infty}(1-F_S((m+j)h))$.

Corollary 34.6 to Theorem 34.5. Under the conditions of Theorem 9.5,

$E((S-(j+1)h)_+) = E((S-jh)_+) - h(1- F_S(jh))$.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 9, pp. 208-210.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C34-1. Aggregate losses covered under a policy of stop-loss insurance have a cumulative distribution function $F_S(x) = 1 - 1/x^2$ for $x > 1$, where S is the aggregate loss random variable. There is a deductible of 2 on the policy. Find $E((S-d)_+)$, the net stop-loss premium for the policy.

Solution S4C34-1. We use the formula $E((S-d)_+) = {}_d{}^{\infty}\int(1-F_S(x))dx$, where d = 2:

$E((S-d)_+) = {}_2{}^{\infty}\int(1 - (1 - 1/x^2))dx = {}_2{}^{\infty}\int(1/x^2)dx = (-1/x) \,\Big|\, {}_2{}^{\infty}$ = $\mathbf{E((S\text{-}d)_+) = ½ = 0.5}$.

Problem S4C34-2. Aggregate losses covered under a policy of stop-loss insurance have a probability density function $f_S(x)$ of 1/40 for $200 \leq x < 220$ and 1/500 for $220 \leq x \leq 470$. A policy deductible of 210 applies. Find $E((S-d)_+)$, the net stop-loss premium for the policy.

Solution S4C34-2. Since we have a continuous aggregate loss distribution, we use the formula $E((S-d)_+) = {}_d{}^{\infty}\int(x-d)f_S(x)dx$ =

${}_{210}{}^{470}\int(x-210)f_S(x)dx = {}_{210}{}^{220}\int(x-210)(1/40)dx + {}_{220}{}^{470}\int(x-210)(1/500)dx$ =

$(x-210)^2/80 \,\Big|\, {}_{210}{}^{220} + (x-210)^2/1000 \,\Big|\, {}_{220}{}^{470}$ =

1.25 + 67.6 - 0.1 = $\mathbf{E((S\text{-}d)_+) = 68.75}$.

Problem S4C34-3. The following is the distribution of aggregate losses covered under a policy of stop-loss insurance:

Pr(S = 230) = ¼

Pr(S = 340) = 1/5

Pr(S = 345) = 1/6

Pr(S = 370) = 3/60

Pr(S = 500) = 1/3.

A policy deductible of 300 applies. Find $E((S-d)_+)$, the net stop-loss premium for the policy.

Solution S4C34-3. Since we have a discrete aggregate loss distribution, we use the formula

$E((S-d)_+) = {}_{x>d}\Sigma((x-d)f_S(x))$, where d = 300.

$E((S-d)_+) = ((340-300)f_S(340)) + ((345-300)f_S(345)) + ((370-300)f_S(370)) + ((500-300)f_S(500))$ =

40*(1/5) + 45*(1/6) + 70*(3/60) + 200*(1/3) = **$E((S-d)_+)$ = 85.6666666667**.

Problem S4C34-4. The following facts are known about a stop loss policy:

If the deductible were 100, the net stop-loss premium would be 500.

If the deductible were 50, the net stop-loss premium would be 530.

It is the case that Pr(50 < S < 100) = 0. That is, it is not possible that aggregate losses will be between 50 and 100.

Find the net stop-loss premium if the deductible were 85.

Solution S4C34-4. This is a case where Theorem 34.4 applies and

$E((S-d)_+) = ((b-d)/(b-a))E((S-a)_+) + ((d-a)/(b-a))E((S-b)_+)$, where d = 85, a = 50, and b = 100.

We also know that $E((S-a)_+) = 530$ and $E((S-b)_+) = 500$. Hence,

$E((S-85)_+) = ((100-85)/(100-50))530 + ((85-50)/(100-50))500 =$

(15/50)530 + (35/50)500 = **$E((S-85)_+)$ = 509**.

Problem S4C34-5. For a policy of stop-loss insurance, it is known with respect to the aggregate loss random variable S that S occurs only in multiples of 300, and $\Pr(S = 300k) = (1/2)^k$ for each k except 0. The policy has a deductible of 900. Find $E((S-d)_+))$, the net stop-loss premium for the policy.

Solution S4C34-5. We apply Theorem 34.5:Assume $\Pr(S = kh) = f_k \geq 0$ for some fixed h > 0, k is a nonnegative integer, and Pr(S = x) = 0 for all other x. Then, provided d = jh, with j a nonnegative integer, $E((S-d)_+) = h*_{m=0}{}^{\infty}\Sigma(1-F_S((m+j)h))$.

This situation meets the criteria for Theorem 34.5 to apply.

Here, h = 300, and d = 900 = 3*h, so j = 3.

Thus, $E((S-d)_+) = 300*_{m=0}{}^{\infty}\Sigma(1-F_S((m+3)300))$.

For every k, $F_S(300k) = (1/2)*(1-(1/2)^k)/(1-(1/2)) = (1-(1/2)^k)$.

Thus, $1 - F_S(300k) = (1/2)^k$.

Hence, $E((S-d)_+) = 300*_{m=0}{}^{\infty}\Sigma((1/2)^{m+3}) = 300*(1/2)^3/(1-(1/2)) = 300(1/2)^2$ = **$E((S-d)_+)$ = 75**.

Section 35

Moment Generating Functions of Aggregate Random Variables and Probability Calculations for Aggregate Random Variables with Exponential Severity

For a compound distribution where the aggregate random variable is S, the frequency random variable is N, and the severity random variable is X, the following is true:

Moment generating function of S:

$\mathbf{M_S(z) = P_N(M_X(z))}$.

If the severity random variable X is an exponential random variable with parameter θ, the following is true:

$\mathbf{F_S(x) = 1 - (e^{-x/\theta}) * {}_{j=0}{}^{\infty}\Sigma((x/\theta)^j/j!) * {}_{n=j+1}{}^{\infty}\Sigma(p_n)}$.

Note that the summation ${}_{n=j+1}{}^{\infty}\Sigma(p_n)$ is encompassed within the other summation.

A **compound Poisson distribution** is an aggregate distribution S where the frequency random variable is Poisson. The following theorem is true (Klugman, Panjer, and Willmot 2008, p. 221):

Theorem 35.1. "Suppose that S_j has a compound Poisson distribution with Poisson parameter λ_j and severity distribution with cdf $F_j(x)$ for j = 1, 2, ..., n. Suppose also that $S_1, S_2, ..., S_n$ are independent. Then $S = S_1 + S_2 + ... + S_n$ has a compound Poisson distribution with Poisson parameter $\lambda = \lambda_1 + \lambda_2 + ... + \lambda_n$ and severity distribution with cdf $F(x) = {}_{j=1}{}^{n}\Sigma((\lambda_j/\lambda)F_j(x))$."

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 9, pp. 217-222.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C35-1. The frequency random variable is a Poisson random variable with parameter λ. The severity random variable is exponential with parameter θ.

Find the moment generating function of S, the aggregate random variable.

Relevant properties for Poisson distributions: $P(z) = e^{\lambda(z-1)}$.

Relevant properties for exponential distributions: $M(t) = (1-\theta t)^{-1}$.

Solution S4C35-1. We use the formula $M_S(z) = P_N(M_X(z)) =$ $\mathbf{M_S(z) = exp(\lambda((1-\theta z)^{-1}-1))}$.

Problem S4C35-2. The frequency random variable is a geometric random variable with parameter $\beta = 1.2$. The severity random variable is a gamma random variable with parameters $\alpha = 3$ and $\theta = 3$. Find $M_S(2)$ for the aggregate random variable S.

Relevant properties for geometric distributions: $P(z) = (1-\beta(z-1))^{-1}$.

Relevant properties for gamma distributions: $M(t) = (1-\theta t)^{-\alpha}$.

Solution S4C35-2. We use the formula $M_S(z) = P_N(M_X(z)) =$

$P_N((1-\theta z)^{-\alpha}) = (1-\beta((1-\theta z)^{-\alpha}-1))^{-1}$ for $z = 2$, $\beta = 1.2$, $\alpha = 3$, and $\theta = 3$.

Thus, $M_S(2) = (1-1.2((1-3*2)^{-3}-1))^{-1} = (1-(-1.2096))^{-1} =$ $\mathbf{M_S(2) = 0.452570601}$.

Problem S4C35-3. The severity random variable X is exponential with parameter $\theta = 100$. The frequency random variable N has the following distribution:

$Pr(N = 0) = 1/3$;

$Pr(N = 1) = ½$;

$Pr(N = 2) = 1/6$;

Find $F_S(x)$ for the aggregate random variable S.

Solution S4C35-3. Since the severity random variable is exponential, we use the formula $F_S(x) = 1 - (e^{-x/\theta})*{}_{j=0}{}^{\infty}\Sigma((x/\theta)^j/j!)*{}_{n=j+1}{}^{\infty}\Sigma(p_n))$.

Since N only takes on positive probabilities for n = 0, 1, and 2, and in the inner summation, the lower bound on n is j + 1, the only values of j we would need to consider are 0 and 1. Thus,

$F_S(x) = 1 - (e^{-x/100})*(((x/100)^0/0!)*{}_{n=1}{}^{2}\Sigma(p_n)) + ((x/100)^1/1!)*{}_{n=2}{}^{2}\Sigma(p_n))) \rightarrow$
$F_S(x) = 1 - (e^{-x/100})*((p_1 + p_2) + (x/100)(p_2)) \rightarrow$
$F_S(x) = 1 - (e^{-x/100})*((1/2 + 1/6) + (x/100)(1/6)) \rightarrow$
$\mathbf{F_S(x) = 1 - (e^{-x/100})*(2/3 + x/600)}$.

Problem S4C35-4. Let $S_1, S_2, ..., S_5$ be i.i.d. compound Poisson random variables, each with parameter $\lambda_j = 7$. The pdf of the severity distribution for each S_j is $F_j(x) = x^2$ for $0 < x < 1$. For $S = S_1 + S_2 + ... + S_5$, find $M_S(2)$. Use a calculator to compute any definite integrals.

Relevant properties for Poisson distributions: $P(z) = e^{\lambda(z-1)}$.

Solution S4C35-4. We invoke Theorem 35.1:

"Suppose that S_j has a compound Poisson distribution with Poisson parameter λ_j and severity distribution with cdf $F_j(x)$ for j = 1, 2, ..., n. Suppose also that $S_1, S_2, ..., S_n$ are independent. Then $S = S_1 + S_2 + ... + S_n$ has a compound Poisson distribution with Poisson parameter $\lambda = \lambda_1 + \lambda_2 + ... + \lambda_n$ and severity distribution with cdf $F(x) = {}_{j=1}{}^{n}\Sigma((\lambda_j/\lambda)F_j(x))$."

Thus, S has Poisson parameter $\lambda = \lambda_1 + \lambda_2 + ... + \lambda_5 = 5*7 = 35$.

Now we can use the formula $F(x) = {}_{j=1}{}^{5}\Sigma((7/35)F_j(x)) = {}_{j=1}{}^{5}\Sigma((1/5)F_j(x)) = F_j(x) = 1 - x^2$, so the severity distribution is unchanged.

To find $M_S(z)$, we use the formula $M_S(z) = P_N(M_X(z))$.

We know that $P(z) = e^{35(z-1)}$ for the Poisson frequency distribution associated with S.

Now we need to find $M_X(t) = E(e^{Xt})$.

We note that $f(x) = F'(x) = 2x$.

Thus, $M_X(t) = {}_0{}^1\int(e^{xt}*(2x))dx$ and

$M_X(2) = {}_0{}^1\int(e^{2x}*2x)dx = 4.194528049$.

Hence, $M_S(2) = P_N(M_X(2)) = P_N(4.194528049) = e^{35(4.194528049-1)} =$ **exp(111.8084817) = approximately 3.612489936*10^{48}**.

Problem S4C35-5. The frequency random variable is a Poisson random variable with parameter 2. The severity random variable is exponential with parameter 50. For the aggregate random variable S, estimate $F_S(120)$ to 2 decimal places.

Relevant properties for Poisson distributions: $p_k = e^{-\lambda}\lambda^k/k!$

Solution S4C35-5. Since severity is exponentially distributed, we use the formula

$F_S(x) = 1 - (e^{-x/\theta})*{}_{j=0}{}^{\infty}\Sigma((x/\theta)^j/j!)*{}_{n=j+1}{}^{\infty}\Sigma(p_n))$, which contains an infinite sum. However, we only need to develop several terms of the formula to achieve the desired level of specificity.

We note that the frequency distribution is Poisson, so a useful shortcut can apply. For instance, ${}_{n=1}{}^{\infty}\Sigma(p_n) = 1 - p_0 = 1 - e^{-2} = 0.8646647168$.

${}_{n=2}{}^{\infty}\Sigma(p_n) = (1 - p_0) - p_1 = 0.8646647168 - 2e^{-2} = 0.5939941503$.

First Estimate:

We apply the formula above but only consider j = 0:

$F_S(120) \approx 1 - (e^{-120/50})*((120/50)^0/0!)*{}_{n=1}{}^{\infty}\Sigma(p_n)) = 1 - (e^{-120/50})*(0.8646647168) =$

0.9215593866. This is just our first estimate.

Second Estimate:

In our second estimate, we also consider j = 1:

$F_S(120) \approx 1 - ((e^{-120/50})*((120/50)^0/0!)*{}_{n=1}{}^{\infty}\Sigma(p_n)) + (e^{-120/50})*((120/50)^1/1!)*{}_{n=2}{}^{\infty}\Sigma(p_n)))$

$= 1 - ((e^{-120/50})*(0.8646647168) + (12/5)(e^{-120/50})*0.5939941503) =$

$0.9215593866 - (12/5)(e^{-120/50})*0.5939941503 = 0.792233146$, our second estimate.

Third estimate:

${}_{n=3}{}^{\infty}\Sigma(p_n) = (1 - p_0 - p_1) - p_2 = 0.5939941503 - 2^2e^{-2}/2 = 0.3233235838.$

$F_S(120) \approx$ Second estimate $- (e^{-120/50})*((120/50)^2/2!)*{}_{n=3}{}^{\infty}\Sigma(p_n))) =$

$0.792233146 - (e^{-120/50})2.88*0.3233235838 = 0.7077591351$, our third estimate.

Fourth estimate:

${}_{n=4}{}^{\infty}\Sigma(p_n) = (1 - p_0 - p_1 - p_2) - p_3 = 0.3233235838 - 2^3e^{-2}/3! = 0.1428765395.$

$F_S(120) \approx$ Third estimate $- (e^{-120/50})*((120/50)^3/3!)*{}_{n=4}{}^{\infty}\Sigma(p_n))) =$

$0.7077591351 - (e^{-120/50})*2.304*0.1428765395 = 0.6778959146$, our fourth estimate.

Fifth estimate:

${}_{n=5}{}^{\infty}\Sigma(p_n) = (1 - p_0 - p_1 - p_2 - p_3) - p_4 = 0.1428765395 - 2^4e^{-2}/4! = 0.0526530173.$

$F_S(120) \approx$ Fourth estimate $- (e^{-120/50})*((120/50)^4/4!)*{}_{n=5}{}^{\infty}\Sigma(p_n))) =$

$0.6778959146 - (e^{-120/50})*1.3824*0.0526530173 = 0.6712927787.$

We note that the estimates are getting quite close to one another, and the fourth and fifth estimates are less than 0.01 apart. It is highly unlikely that any of the subsequent estimates will decline below 0.665, as the impact of each subsequent term of the summation in the original equation steadily approaches zero. Thus our answer, rounded to the nearest hundredth, is **0.67**.

Section 36

Using Convolutions to Determine the Probability Distribution of Aggregate Random Variables

For a compound distribution where the aggregate random variable is S, the frequency random variable is N, and the severity random variable is X, the following is true:

$F_S(x) = {}_{n=0}{}^{\infty}\Sigma(p_n{*}F_X{}^{*n}(x))$, where

$F_X{}^{*n}(x)$ is called the **n-fold convolution** of the cdf of X. The formula for calculating $F_X{}^{*k}(x)$ for any positive integer k is as follows:
$F_X{}^{*k}(x) = {}_{-\infty}{}^{\infty}\int(F_X{}^{*(k-1)}(x-y){*}d(F_X(y)))$.

Moreover, $F_X{}^{*0}(x) = 0$ if $x < 0$ and 1 if $x \geq 0$.

If losses are always nonnegative, the integral becomes as follows:

$F_X{}^{*k}(x) = {}_0{}^{x}\int(F_X{}^{*(k-1)}(x-y){*}d(F_X(y)))$.

The n-fold convolutions are often extremely difficult to compute in practice, so various approximation techniques are used.

The **recursive method** of approximation works when the frequency random variable belongs to either the (a, b, 0) or the (a, b, 1) class of distributions.

For the (a, b, 0) class, the following equation holds:

Equation 36.1. $f_S(x) = {}_{y=1}{}^{x\Lambda m}\Sigma((a+by/x)f_X(y)f_S(x-y))/(1-af_X(0))$.

For the (a, b, 1) class, the following equation holds:

Equation 36.2. $f_S(x) = ((p_1-(a+b)p_0)f_X(x) + {}_{y=1}{}^{x\Lambda m}\Sigma((a+by/x)f_X(y)f_S(x-y))/(1-af_X(0))$.

In both equations above, $x \Lambda m = \min(x, m)$, where m is some specified amount that values of the severity random variable cannot exceed. It is possible for m to be infinite. In both of the equations above, the severity distribution of x is defined on the nonnegative integers up to and including m.

For the Poisson distribution, the following equation holds:

Equation 36.3. $f_S(x) = (\lambda/x)_{y=1}{}^{x\Lambda m}\Sigma(y*f_X(y)f_S(x-y))$ for positive integers x and

$f_S(0) = \exp(-\lambda(1-f_X(0)))$.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 9, pp. 204, 225-227.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C36-1. Let X, the severity random variable, follow a uniform distribution with pdf $f_X(x) = 1/100$ for $0 \leq x \leq 100$ and $f_X(x) = 0$ otherwise. The frequency random variable N has the following distribution:

Pr(N = 0) = 1/3;
Pr(N = 1) = ½;
Pr(N = 2) = 1/6.

Find the 1-fold convolution of the cdf of X.

Solution S4C36-1. We note that $F_X{}^{*0}(x) = 0$ if $x < 0$ and 1 if $x \geq 0$. For this distribution of X, X is always at least 0, so $F_X{}^{*0}(x) = 1$. Moreover, this is a uniform distribution, so $F_X(x) = x/100$.

We use the formula $F_X{}^{*k}(x) = {}_0{}^x\int(F_X{}^{*(k-1)}(x-y)*d(F_X(y))$.

Thus, $F_X{}^{*1}(x) = {}_0{}^x\int(F_X{}^{*0}(x-y)*d(y/100)$. $F_X{}^{*0}(x-y)$ is always 1, so
$F_X{}^{*1}(x) = {}_0{}^x\int 1*dy/100 \rightarrow$
$F_X{}^{*1}(x) = {}_0{}^x\int(0.01)dy \rightarrow$
$F_X{}^{*1}(x) = 0.01y \mid_0{}^x \rightarrow$
$\mathbf{F_X{}^{*1}(x) = 0.01x = x/100}$.

Note that $F_X{}^{*1}(x) = F_X(x)$ if x can only assume nonnegative values.

Problem S4C36-2. Let X, the severity random variable, follow a uniform distribution with pdf $f_X(x) = 1/100$ for $0 \leq x \leq 100$ and $f_X(x) = 0$ otherwise. The frequency random variable N has the following distribution:

Pr(N = 0) = 1/3;
Pr(N = 1) = ½;
Pr(N = 2) = 1/6.

Find the 2-fold convolution of the cdf of X.

Solution S4C36-2. We know from Solution S4C36-1 that $F_X{}^{*1}(x) = x/100$ and that $F_X(x) = x/100$. We use the following formula: $F_X{}^{*k}(x) = {}_0{}^x\int(F_X{}^{*(k-1)}(x-y)*d(F_X(y))$.

Thus, $F_X^{*2}(x) = {}_0{}^x\int(F_X^{*1}(x-y)*d(y/100) \rightarrow$
$F_X^{*2}(x) = {}_0{}^x\int 0.01(x-y)*dy/100 \rightarrow$
$F_X^{*2}(x) = {}_0{}^x\int(0.0001x-0.0001y)dy \rightarrow$
$F_X^{*2}(x) = (0.0001xy - 0.00005y^2) \mid_0{}^x \rightarrow$
$F_X^{*2}(x) = (x^2/10000) - (x^2/20000) =$
$\mathbf{F_X^{*2}(x) = x^2/20000 = 5*10^{-5}*x^2}$.

Problem S4C36-3. Let X, the severity random variable, follow a uniform distribution with pdf $f_X(x) = 1/100$ for $0 \leq x \leq 100$ and $f_X(x) = 0$ otherwise. The frequency random variable N has the following distribution:
Pr(N = 0) = 1/3;
Pr(N = 1) = ½;
Pr(N = 2) = 1/6.

Find $F_S(x)$ for the aggregate random variable S.

Solution S4C36-3. We use the formula $F_S(x) = {}_{n=0}{}^{\infty}\Sigma(p_n*F_X^{*n}(x))$, where
$p_n = 1/3$, $p_1 = ½$, and $p_2 = 1/6$. Moreover, from Solutions S4C36-1 and S4C36-2,
$F_X^{*0}(x) = 1$, $F_X^{*1}(x) = x/100$, and $F_X^{*2}(x) = x^2/20000$.
Here, n can only assume the values 0, 1, and 2.
Thus, $F_S(x) = (1/3)*1 + (1/2)x/100 + (1/6)x^2/20000 \rightarrow$
$\mathbf{F_S(x) = 1/3 + x/200 + x^2/120000}$.

Problem S4C36-4. Let X, the severity random variable, follow a distribution with pf $f_X(x) = 1/100$ for every integer x from 0 to 99 and $f_X(x) = 0$ otherwise. The frequency random variable N has a Poisson distribution with λ = 3. Find $f_S(0)$ for the aggregate random variable S.

Solution S4C36-4. We use the formula $f_S(0) = \exp(-\lambda(1-f_X(0)) = \exp(-3(1-f_X(0)))$.

Here, $f_X(0) = 1/100$, so $f_S(0) = \exp(-3(1-1/100)) = \exp(-2.97) =$ $\mathbf{f_S(0) = 0.0513033103}$.

Problem S4C36-5. Let X, the severity random variable, follow a distribution with pf $f_X(x) = 1/100$ for every integer x from 0 to 99 and $f_X(x) = 0$ otherwise. The frequency random variable N has a Poisson distribution with λ = 3. Find $f_S(1)$ for the aggregate random variable S.

Solution S4C36-5. We use the formula $f_S(x) = (\lambda/x)_{y=1}{}^{x\Lambda m}\Sigma(y*f_X(y)f_S(x-y))$, where, in this case, x = 1 and m is infinite. Thus, there is only one term in the summation (pertaining to y = 1) and

$f_S(1) = (\lambda/1)*1*f_X(1)f_S(1-1) = 3*1*(1/100)*f_S(0)$. From Solution S4C36-4, we know that $f_S(0) = 0.0513033103$. Therefore, $f_S(1) = (3/100)* 0.0513033103 =$ $\mathbf{f_S(1) = 0.0015390993}$.

Section 37

The Method of Rounding or Mass Dispersal

If the frequency distribution is of the (a, b, 1) class and the severity distribution is continuous with positive probabilities over the positive real numbers, then the following equation can be used to apply the recursive method of estimating the pdf of the aggregate distribution:

$f_S(x) = p_1 * f_X(x) + {}_0{}^x\int(a + by/x)f_X(y)f_S(x-y)dy$.

However, this equation is difficult to solve for any x, as we would need to know all the values of $f_S(x-y)$ for $0 \leq y \leq x$. Klugman, Panjer, and Willmot address this problem by approximating the continuous severity distribution with a discrete severity distributions. They present several methods of doing so. Here, we will focus on one method in particular.

Method of Rounding (Mass Dispersal) (Klugman, Panjer, and Willmot 2008, p. 232):

Here, h is defined as some unit of measurement such that the behavior of the original continuous severity distribution is still reasonably taken into account.

"Let fj denote the probability placed at jh, for nonnegative integer values of j. Then set

$f_0 = \Pr(X < h/2) = F_X(h/2)$,

$f_j = \Pr(jh - h/2 \leq X < jh + h/2) = F_X(jh + h/2) - F_X(jh - h/2)$."

At some point, it may be reasonable to halt the discretization process at some point for which j = m. In that case, f_m will be $1-F_X((m-0.5)h)$.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 9, pp. 230-232.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C37-1. The frequency random variable N follows a Poisson distribution with $\lambda = 5$. Let X, the severity random variable, follow a uniform distribution with pdf $f_X(x) = 1/100$ for $0 \leq x \leq 100$ and $f_X(x) = 0$ otherwise. Use the Method of Rounding to approximate f_0 in the discretized distribution. For your unit of measurement, use h = 20.

Solution S4C37-1. We use the formula $f_0 = F_X(h/2)$. For this uniform distribution, $F_X(x) = x/100$, so $F_X(h/2) = F_X(10) = 10/100 = 0.1$. Thus, $\mathbf{f_0 = 0.1}$.

Problem S4C37-2. The frequency random variable N follows a Poisson distribution with $\lambda = 5$. Let X, the severity random variable, follow a Pareto distribution with $S_X(x) = (300/(x+300))^4$ for $x \geq 0$. Use the Method of Rounding to approximate f_0 in the discretized distribution. For your unit of measurement, use h = 60.

Solution S4C37-2. We use the formula $f_0 = F_X(h/2)$. $F_X(h/2) = F_X(30) = 1 - S_X(30) =$

$1 - (300/(30+300))^4 =$ $\mathbf{f_0 = 0.3169865446}$.

Problem S4C37-3. The frequency random variable N follows a Poisson distribution with $\lambda = 5$. Let X, the severity random variable, follow a Pareto distribution with $S_X(x) = (300/(x+300))^4$ for $x \geq 0$. Use the Method of Rounding to approximate f_1 in the discretized distribution. For your unit of measurement, use h = 60.

Solution S4C37-3. We use the formula $f_j = F_X(jh + h/2) - F_X(jh - h/2)$.

Here, $f_1 = F_X(1*60 + 60/2) - F_X(1*60 - 60/2) = F_X(90) - F_X(30) = S_X(30) - S_X(90) =$

$(300/(30+300))^4 - (300/(90+300))^4 =$ $\mathbf{f_1 = 0.3328856587}$.

Problem S4C37-4. The frequency random variable N follows a Poisson distribution with $\lambda = 5$. Let X, the severity random variable, follow a Pareto distribution with $S_X(x) = (300/(x+300))^4$ for $x \geq 0$. Use the Method of Rounding to approximate f_2 and f_3 in the discretized distribution such that f_3 is the *last* positive probability (that is, all of the probability in the discretized distribution occurs over the values 0, 1, 2, and 3). For your unit of measurement, use h = 60.

Solution S4C37-4. For f_2 we use the formula $f_j = F_X(jh + h/2) - F_X(jh - h/2)$.

$f_2 = F_X(2*60 + 60/2) - F_X(2*60 - 60/2) = F_X(150) - F_X(90) = S_X(90) - S_X(150) =$

$(300/(90+300))^4 - (300/(150+300))^4 =$ $\mathbf{f_2 = 0.1525969324}$.

For f_3 we use the formula $f_m = 1-F_X((m-0.5)h)$, since 3 is our value of m at which the discretized distribution terminates. $F_3 = 1-F_X((3-0.5)h) = S_X(2.5*60) = S_X(150) = (300/(150+300))^4 =$ $\mathbf{f_3 = 0.1975308642}$.

Problem S4C37-5.

The frequency random variable N follows a Poisson distribution with $\lambda = 5$. Let X, the severity random variable, follow a Pareto distribution with $S_X(x) = (300/(x+300))^4$ for $x \geq 0$. Find $f_S(120)$. Note that $f_S(120) = f_A(2)$, where A = N*J, where J from the discretized distribution takes on the values j = 0, 1, 2, or 3, with S = hJ for h = 60. Use the discretized distribution previously obtained via the Method of Rounding in Problems S4C37-2 through S4C37-4 and the formula $f_A(x) = (\lambda/x)_{y=1}{}^{x\Lambda m}\Sigma(y*f_y*f_A(x-y))$ for positive integers x and $f_A(0) = \exp(-\lambda(1-f_0))$.

Solution S4C37-5. First, we recall our discretized distribution as found in Problems S4C37-2 through S4C37-4:

f_0 = 0.3169865446;

f_1 = 0.3328856587;

f_2 = 0.1525969324;

f_3 = 0.1975308642.

Now we find $f_A(0) = \exp(-\lambda(1-f_0)) = \exp(-5(1-0.3169865446)) = \exp(-3.415067277) = f_A(0) = 0.0328741949$.

Now we can find $f_A(1) = (\lambda/1)\sum_{y=1}^{1\wedge m}(y^* f_y{*}f_A(1-y)) = (\lambda/1)*1*f_1*f_A(1-1) = 5*0.3328856587*0.0328741949 = f_A(1) = 0.0547161402$.

Now we can find $f_A(2) = (\lambda/2)\sum_{y=1}^{2\wedge m}(y*f_y*f_A(2-y))$ =

$(\lambda/2)(1*f_1*f_A(2-1)) + (\lambda/2)(2*f_2*f_A(2-2))$ =

$(\lambda/2)(1*f_1*f_A(1)) + (\lambda/2)(2*f_2*f_A(0))$ =

2.5*0.3328856587*0.0547161402 + 2.5*2*0.1525969324*0.0328741949 =

$f_A(2)$ =$\mathbf{f_S(120) = 0.0706180524}$.

Section 38

The Method of Local Moment Matching

This section presents another method for discretizing a continuous severity distribution, the Method of Local Moment Matching. We let ph be the length of one of the intervals over which the distribution will be discretized. This interval begins at some value x_k and ends at x_k + ph.

The point masses m_0^k, m_1^k, ... , m_p^k will be placed at points x_k, x_k + h, ..., x_k + ph, respectively.

This results in a system of p+1 equations, each of the following form:

${}_{j=0}{}^{p}\Sigma(x_k+jh)^r*m_j^k = {}_{x\_k-0}{}^{x\_k+ph-0}\int x^r*dF_X(x)$, for each r = 0, 1, 2, ..., p.

This system of p+1 equations can (at least in theory) be solved to find each of the m_j^k.

The "-0" notation at the limits of the integral means that one should include the discrete probability at x_k but exclude the discrete probability at $x_{k+1} = x_k$ + ph.

If we assume that $x_0 = 0$, then the following are true:

$f_0 = m_0^0$; $f_1 = m_1^0$; $f_2 = m_2^0$;

$f_p = m_p^0 + m_0^1$; $f_{p+1} = m_1^1$; $f_{p+2} = m_2^1$

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 9, p. 232.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C38-1. The frequency random variable N follows a Poisson distribution with λ = 5. Let X, the severity random variable, follow a Pareto distribution with $S_X(x) = (300/(x+300))^4$ for x ≥ 0. Use the Method of Local Moment Matching with h = 60 and p = 1 to find the equation corresponding to r = 0 in the resulting system of equations. Assume that $x_0 = 0$.

Solution S4C38-1. The formula for setting up the system of equations is

${}_{j=0}{}^{p}\Sigma(x_k+jh)^r*m_j^k = {}_{x\_k-0}{}^{x\_k+ph-0}\int x^r*dF_X(x)$, for each r = 0, 1, 2, ..., p.

Here, p = 1. Moreover, since $x_0 = 0$ and $x_{k+1} = x_k + ph = x_k + h$, it follows that

x_k = hk = 60k, and x_{k+1} = 60k + 60. Thus, 60k and 60k+60 are the bounds of our integral.

r takes on values from 0 to p, i.e., from 0 to 1. This implies that we have a system of two equations.

Moreover, $dF_X(x) = f_X(x)dx$, which, for a Pareto distribution, is $\alpha\theta^\alpha/(x+\theta)^{\alpha+1}$. Here,

$f_X(x) = 4*300^4/(x+300)^5$.

For r = 0, the corresponding equation is as follows:

${}_{j=0}{}^{1}\Sigma(x_k+jh)^0*m_j{}^k = {}_{60k}{}^{60k+60}\int x^0*f_X(x)dx \rightarrow$

${}_{j=0}{}^{1}\Sigma m_j{}^k = {}_{60k}{}^{60k+60}\int f_X(x)dx \rightarrow$

$m_0{}^k + m_1{}^k = F_X(x) \mid_{60k}^{60k+60} \rightarrow$

$m_0{}^k + m_1{}^k = F_X(60k+60) - F_X(60k) \rightarrow$

$m_0{}^k + m_1{}^k = S_X(60k) - S_X(60k+60) \rightarrow$

Our desired equation is as follows:

$\mathbf{m_0{}^k + m_1{}^k = (300/(60k +300))^4 - (300/(60k +360))^4}$.

Problem S4C38-2. The frequency random variable N follows a Poisson distribution with $\lambda = 5$. Let X, the severity random variable, follow a Pareto distribution with $S_X(x) = (300/(x+300))^4$ for $x \geq 0$. Use the Method of Local Moment Matching with h = 60 and p = 1 to find the equation corresponding to r = 1 in the resulting system of equations. Assume that $x_0 = 0$.

Solution S4C38-2. For r = 1, the equation is as follows:
${}_{j=0}{}^{1}\Sigma(x_k+jh)^1*m_j{}^k = {}_{60k}{}^{60k+60}\int x^1*f_X(x)dx \rightarrow$
$(x_k+0h)^1*m_0{}^k + (x_k+ h)^1*m_1{}^k = {}_{60k}{}^{60k+60}\int x*f_X(x)dx \rightarrow$
$x_k*m_0{}^k + (x_k+ 60)*m_1{}^k = {}_{60k}{}^{60k+60}\int x*f_X(x)dx$.

From Solution S4C38-1, $f_X(x) = 4*300^4/(x+300)^5$.
Thus, $x_k*m_0{}^k + (x_k+ 60)*m_1{}^k = {}_{60k}{}^{60k+60}\int 4x*300^4/(x+300)^5$.

To integrate $4x*300^4/(x+300)^5$, we perform integration by parts using the tabular method:

Sign.........u............dv
+...........x...........$4*300^4/(x+300)^5$
- ....1............$-300^4/(x+300)^4$
+............0...........$300^4/(3*(x+300)^3)$

Thus, we have

${}_{60k}{}^{60k+60}\int 4x*300^4/(x+300)^5 = (-x*300^4/(x+300)^4 - 300^4/(3*(x+300)^3)) \mid_{60k}^{60k+60} =$

$((60k)300^4/(60k+300)^4 + 300^4/(3*(60k+300)^3)) - (60k+60)300^4/(60k+360)^4 - 300^4/(3*(60k+360)^3))$.

Hence, our desired equation is as follows:

$\mathbf{x_k*m_0^k + (x_k+ 60)*m_1^k = ((60k)300^4/(60k+300)^4 + 300^4/(3*(60k+300)^3)) - (60k+60)300^4/(60k+360)^4 - 300^4/(3*(60k+360)^3))}$.

Problem S4C38-3. The frequency random variable N follows a Poisson distribution with $\lambda = 5$. Let X, the severity random variable, follow a Pareto distribution with $S_X(x) = (300/(x+300))^4$ for $x \geq 0$. Using the system of equations you found for Problems S4C38-1 and S4C38-2, solve for m_0^0 and m_1^0. Assume that $x_0 = 0$.

Solution S4C38-3. We use the following system of equations:

(i) $m_0^k + m_1^k = (300/(60k +300))^4 - (300/(60k +360))^4$

(ii) $x_k*m_0^k + (x_k+ 60)*m_1^k = ((60k)300^4/(60k+300)^4 + 300^4/(3*(60k+300)^3)) - (60k+60)300^4/(60k+360)^4 - 300^4/(3*(60k+360)^3))$.

Here, k = 0, implying that $x_k = 0$, which considerably simplifies our task.

The two equations become as follows:
(i) $m_0^0 + m_1^0 = (300/300)^4 - (300/360)^4 \rightarrow$

(i) $m_0^0 + m_1^0 = 0.5177469136$.

(ii) $0*m_0^0 + (0+ 60)*m_1^0 = ((60*0)300^4/(60*0+300)^4 + 300^4/(3*(60*0+300)^3)) - (60*0+60)300^4/(60*0+360)^4 - 300^4/(3*(60*0+360)^3)) \rightarrow$

(ii) $60*m_1^0 = 300^4/(3*(300)^3) - 60*300^4/(360)^4 - 300^4/(3*(360)^3) \rightarrow$

(ii) $60*m_1^0 = 100 - 28.93518519 - 57.87037037 \rightarrow$

(ii) $60*m_1^0 = 13.194444444444 \rightarrow$

$\mathbf{m_1^0 = 0.2199074073} \rightarrow$

$m_0^0 = 0.5177469136 - m_1^0 = \mathbf{m_0^0 = 0.2978395063}$.

Problem S4C38-4. The frequency random variable N follows a Poisson distribution with $\lambda = 5$. Let X, the severity random variable, follow a Pareto distribution with $S_X(x) = (300/(x+300))^4$ for $x \geq 0$. Using the system of equations you found for Problems S4C38-1 and S4C38-2, solve for m_0^1 and m_1^1. Assume that $x_0 = 0$.

Solution S4C38-4. We use the following system of equations:

(i) $m_0^k + m_1^k = (300/(60k+300))^4 - (300/(60k+360))^4$

(ii) $x_k*m_0^k + (x_k+60)*m_1^k = ((60k)300^4/(60k+300)^4 + 300^4/(3*(60k+300)^3)) - (60k+60)300^4/(60k+360)^4 - 300^4/(3*(60k+360)^3))$.

Here, k = 1, implying that $x_k = x_1 = x_0 + ph = 0 + 1*60 = 60$.

The two equations become as follows:
(i) $m_0^1 + m_1^1 = (300/(360))^4 - (300/(420))^4 \rightarrow$
(i) $m_0^1 + m_1^1 = 0.2219448815$.
(ii) $60*m_0^1 + (60+60)*m_1^1 = ((60*1)300^4/(60*1+300)^4 + 300^4/(3*(60*1+300)^3)) - (60*1+60)300^4/(60*1+360)^4 - 300^4/(3*(60*1+360)^3)) \rightarrow$
(ii) $60*m_0^1 + 120*m_1^1 = 28.93518519 + 57.87037037 - 31.23698450 - 36.44314869 \rightarrow$
(ii) $60*m_0^1 + 120*m_1^1 = 19.12542237$.

We can take 60*(i) and subtract it from (ii), leaving the following:
(iii) $60*m_1^1 = 19.12542237 - 60*0.2219448815 \rightarrow$
(iii) $60*m_1^1 = 5.808729482 \rightarrow$
$\mathbf{m_1^1 = 0.096812158} \rightarrow$
$m_0^1 = 0.2219448815 - 0.096812158 = \mathbf{m_0^1 = 0.1251327235}$.

Problem S4C38-5. The frequency random variable N follows a Poisson distribution with λ = 5. Let X, the severity random variable, follow a Pareto distribution with $S_X(x) = (300/(x+300))^4$ for $x \geq 0$. Use your work for problems Problems S4C38-1 through S4C38-4 to define f_0 and f_1 for the discretized distribution.

Solution S4C38-5. Two formulas here are useful: $f_0 = m_0^0$ and $f_p = m_p^0 + m_0^1$. Here, p = 1.
Thus, $f_0 = m_0^0 = \mathbf{f_0 = 0.2978395063}$.
$f_1 = m_1^0 + m_0^1 = 0.2199074073 + 0.1251327235 = \mathbf{f_1 = 0.3450401308}$.

So much work for only two probability masses of the discretized distribution! The Method of Rounding is certainly more efficient.

Section 39

Effects of Individual Policy Modifications on the Aggregate Payment Random Variable

When an individual insurance policy is modified by a deductible, limit, and/or coinsurance, this affects one's treatment of the aggregate payment random variable S.

We define the following random variables:

Let N^L be the number of losses. Let N^P be the number of payments.

Let Y^L be the amount of a payment on a per-loss basis. Let Y^P be the amount of a payment on a per-payment basis.

We also let v be the probability that a loss results in a payment.

The following equations hold:

$F_{Y^\wedge L}(y) = (1\text{-}v) + v*F_{Y^\wedge P}(y)$, $y \geq 0$.

$M_{Y^\wedge L}(t) = (1\text{-}v) + v*M_{Y^\wedge P}(t)$.

$P_{N^\wedge P}(z) = P_{N^\wedge L}(1\text{-}v\text{+}vz)$.

$M_S(t) = P_{N^\wedge L}(M_{Y^\wedge L}(t)) = P_{N^\wedge P}(M_{Y^\wedge P}(t))$.

Source: *Loss Models: From Data to Decisions*, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 9, pp. 238-240.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C39-1. An insurance company has a policy where the amount of each payment for losses on a per-payment basis follows a Pareto distribution with $\alpha = 5$ and $\theta = 1000$. The probability that a payment will be made under this policy if a loss occurs is 0.55. Find the cumulative distribution function $F_{Y^\wedge L}(y)$ of the payment amount of a per-loss basis.

Relevant properties for Pareto distributions: $S(x) = \theta^\alpha/(x + \theta)^\alpha$.

Solution S4C39-1. We use the formula $F_{Y^\wedge L}(y) = (1\text{-}v) + v*F_{Y^\wedge P}(y)$, $y \geq 0$.

The value v is the probability that the loss will result in a payment, i.e, 0.55.

Thus, $F_{Y^L}(y) = (1 - 0.55) + 0.55*F_{Y^P}(y) \rightarrow$

$F_{Y^L}(y) = 0.45 + 0.55(1 - S_{Y^P}(y)) \rightarrow$

$F_{Y^L}(y) = 0.45 + 0.55(1 - 1000^5/(y + 1000)^5) \rightarrow$

$F_{Y^L}(y) = 1 - 0.55*1000^5/(y + 1000)^5$.

Problem S4C39-2. An insurance company has a policy where the amount of each payment for losses on a per-payment basis follows an exponential distribution with θ = 100. The probability that a loss will result in a payment is 0.76. Find $M_{Y^L}(3)$, the value for t = 3 of the moment-generating function of the per-loss payment amount random variable.

Relevant properties for exponential distributions: $M(t) = (1-\theta t)^{-1}$.

Solution S4C39-2. We are given that $M_{Y^P}(t) = (1-100t)^{-1}$.

We use the formula $M_{Y^L}(t) = (1-v) + v*M_{Y^P}(t)$ for v = 0.76.

Then $M_{Y^L}(t) = 0.24 + 0.76(1-100t)^{-1}$.
So $M_{Y^L}(3) = 0.24 + 0.76(1-100t)^{-1} = 0.24 + 0.76(1-100*3)^{-1} = 0.24 + 0.76(-299)^{-1} =$
$0.24 + 0.76(-0.0033444816) =$ **$M_{Y^L}(3) = 0.237458194$**.

Problem S4C39-3. The number of losses for a given peril follows a Poisson distribution with λ = 7. The probability that a loss will result in a payment is 0.46. Find $P_{N^P}(4)$, the value for t = 4 of the probability generating function of the number-of-payments random variable.

Relevant properties for Poisson distributions: $P(z) = e^{\lambda(z-1)}$.

Solution S4C39-3. We use the formula $P_{N^P}(z) = P_{N^L}(1-v+vz)$, for v = 0.46. Thus,

$P_{N^P}(z) = P_{N^L}(0.54+0.46z) = \exp(\lambda(0.54+0.46z-1)) = \exp(7(-0.46+0.46z)) =$
$\exp(3.22z - 3.22) = \exp(3.22z-3.22)$.

Thus, $P_{N^P}(4) = \exp(3.22*4-3.22) =$ **$P_{N^P}(4) = 15677.78467$.**

Problem S4C39-4. An insurance company has a policy where the amount of each payment for losses on a per-payment basis follows an exponential distribution with θ = 100. The number of losses for a given peril follows a Poisson distribution with λ = 7. The probability that a loss will result in a payment is 0.46. Find $M_S(t)$, the moment generating function of the aggregate payment random variable.

Relevant properties for Poisson distributions: $P(z) = e^{\lambda(z-1)}$.

Relevant properties for exponential distributions: $M(t) = (1-\theta t)^{-1}$.

Solution S4C39-4. We use the formula $M_S(t) = P_{N\wedge P}(M_{Y\wedge P}(t))$.

We know that $M_{Y\wedge P}(t) = (1\text{-}100t)^{-1}$.

Moreover, from Solution S4C39-3, we know that $P_{N\wedge P}(z) = \exp(3.22z\text{-}3.22)$.

Thus, $\mathbf{M_S(t) = exp(3.22(1\text{-}100t)^{-1}\text{-}3.22)}$.

Problem S4C39-5. An insurance company has a policy where the amount of each payment for losses follows an exponential distribution with θ = 100. The number of losses for a given peril follows a Poisson distribution with λ = 7. There is an ordinary deductible of 30, a policy limit of 340 - applied before the policy limit and the coinsurance, and coinsurance of 53%. Find the expected value of S, the aggregate payments on a per-loss basis.

Relevant properties for exponential distributions: $E(X \wedge k) = \theta(1\text{-}e^{-k/\theta})$.

Solution S4C39-5. $E(S) = E(N^L)*E(Y^L)$.

Since N^L is Poisson, $E(N^L) = \lambda = 7$.

$E(Y^L) = \alpha(E(X \wedge u) - E(X \wedge d))$, where u is the limit, d is the deductible, and α is the coinsurance amount. Here,

$E(Y^L) = 0.53(E(X \wedge 340) - E(X \wedge 30)) =$

$0.53(100(1\text{-}e^{-340/100}) - 100(1\text{-}e^{-30/100})) = E(Y^L) = 37.49458239$.

Thus, E(S) = 7*37.49458239 = **E(S) = 262.4620767**.

Section 40

Properties of Aggregate Losses in the Individual Risk Model

We recall the individual risk model from Section 32:

The **individual risk model** represents the aggregate loss as a sum, $\mathbf{S = X_1 + X_2 + ... + X_n}$ of a fixed number of insurance contracts, n. The random variables $X_1, X_2, ..., X_n$ are assumed to be independent but are not necessarily identically distributed.

This model was originally developed in a life insurance context, where the probability of death for the jth individual within the given year is q_j, and the benefit paid for the death of the jth individual is b_j. In such a situation, the following is true about the distribution of loss to the insurer (random variable X_j):

$f_{X\_j}(x) = 1 - q_j$ if $x = 0$;

$f_{X\_j}(x) = q_j$ if $x = b_j$.

For aggregate losses S, the following are the case:

$E(S) = {}_{j=1}{}^{n}\Sigma(b_j*q_j)$;

$Var(S) = {}_{j=1}{}^{n}\Sigma(b_j{}^2*q_j*(1-q_j))$;

$P_S(z) = {}_{j=1}{}^{n}\Pi(1-q_j + q_j*z^{b\_j})$.

If all risks are identical, $q_j = q$, and $b_j = 1$, then $P_S(z) = (1+q(z-1))^n$, which is the probability generating function of a binomial distribution.

Let B_j be the random variable corresponding to the distribution of benefits for the jth individual. Now B_j can assume a single value with probability 1, or it can vary among any number of possible values.

Then the following are the case:
$M_S(t) = {}_{j=1}{}^{n}\Pi(1 - q_j + q_j*M_{B\_j}(t))$.

We define $\mu_j = E(B_j)$ and $\sigma_j{}^2 = Var(B_j)$.

Then $E(S) = {}_{j=1}{}^{n}\Sigma(\mu_j*q_j)$;

$Var(S) = {}_{j=1}{}^{n}\Sigma(\sigma_j{}^2*q_j + \mu_j{}^2*q_j*(1-q_j))$.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 9, pp. 253-254.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C40-1. Life or Death Insurance Company insures 5 individuals with the following characteristics:

Individual 1: Probability of death in the next year is 0.34. Benefit receivable upon death is 500.

Individual 2: Probability of death in the next year is 0.034. Benefit receivable upon death is 230.

Individual 3: Probability of death in the next year is 0.54. Benefit receivable upon death is 90.

Individual 4: Probability of death in the next year is 0.0001. Benefit receivable upon death is 12000.

Individual 5: Probability of death in the next year is 0.053. Benefit receivable upon death is 1000.

Find E(S), the expected value of aggregate losses next year for this group of five life insurance policies.

Solution S4C40-1. We use the formula $E(S) = {}_{j=1}{}^{n}\Sigma(b_j * q_j) =$

$500*0.34 + 230*0.034 + 90*0.54 + 12000*0.0001 + 1000*0.053 =$ **$E(S) = 280.62$**.

Problem S4C40-2. Life or Death Insurance Company insures 5 individuals with the following characteristics:

Individual 1: Probability of death in the next year is 0.34. Benefit receivable upon death is 500.

Individual 2: Probability of death in the next year is 0.034. Benefit receivable upon death is 230.

Individual 3: Probability of death in the next year is 0.54. Benefit receivable upon death is 90.

Individual 4: Probability of death in the next year is 0.0001. Benefit receivable upon death is 12000.

Individual 5: Probability of death in the next year is 0.053. Benefit receivable upon death is 1000.

Find Var(S), the variance of aggregate losses next year for this group of five life insurance policies.

Solution S4C40-2. We use the formula Var(S) = $_{j=1}{}^{n}\Sigma(b_j{}^2*q_j*(1-q_j))$ =

500^2*0.34*(1-0.34) + 230^2*0.034*(1-0.034) + 90^2*0.54*(1-0.54) + 12000^2*0.0001*(1-0.0001) + 1000^2*0.053*(1-0.053) = **Var(S) = 124439.0476**.

Problem S4C40-3. Life or Death Insurance Company insures 5 individuals with the following characteristics:

Individual 1: Probability of death in the next year is 0.34. Benefit receivable upon death is 500.

Individual 2: Probability of death in the next year is 0.034. Benefit receivable upon death is 230.

Individual 3: Probability of death in the next year is 0.54. Benefit receivable upon death is 90.

Individual 4: Probability of death in the next year is 0.0001. Benefit receivable upon death is 12000.

Individual 5: Probability of death in the next year is 0.053. Benefit receivable upon death is 1000.

Find $P_S(z)$, the probability generating function of aggregate losses next year for this group of five life insurance policies. You do not need to expand the expressions in the parentheses.

Solution S4C40-3. We use the formula $P_S(z) = {}_{j=1}{}^{n}\Pi(1-q_j + q_j*z^{b\_j})$ =

$(1-0.34+ 0.34*z^{500})(1-0.034+ 0.034*z^{230})(1-0.54+ 0.54*z^{90})(1-0.0001+ 0.0001*z^{12000})(1-0.053+ 0.053*z^{1000})$ = $\mathbf{P_S(z) = (0.66 + 0.34*z^{500})(0.966+ 0.034*z^{230})(0.46+ 0.54*z^{90})(0.9999+ 0.0001*z^{12000})(0.947+ 0.053*z^{1000})}$.

Problem S4C40-4. Standard Life Insurance Company insures 12 individuals in Category A, who each have probability 0.01 of dying next year. It also insures 18 individuals in Category B, who each have probability 0.005 of dying next year. The benefit payable to a dead individual in Category A is 50 with probability 0.42 and 130 with probability 0.58. The benefit payable to a dead individual in Category B is 500 with probability 0.22 and 520 with probability 0.78.

Find E(S), the expected value of aggregate losses next year for this group of 30 life insurance policies.

Solution S4C40-4. We use the formula E(S) = $_{j=1}{}^{n}\Sigma(\mu_j*q_j)$.
For Category A, μ_A = E(B_j) = 0.42*50 + 0.58*130 = μ_A = 96.4.

For Category A, μ_A*q_A = 96.4*0.01 = 0.964, and so 12*μ_A*q_A = 11.568.

For Category B, μ_B = E(B_j) = 0.22*500 + 0.78*520 = μ_B = 515.6.

For Category B, μ_B*q_B = 515.6*0.005 = 2.578 and so 18*μ_B*q_B = 46.404.

Thus, $E(S) = 12*\mu_A*q_A + 18*\mu_B*q_B = 11.568 + 46.404 =$ **E(S) = 57.972**.

Problem S4C40-5. Standard Life Insurance Company insures 12 individuals in Category A, who each have probability 0.01 of dying next year. It also insures 18 individuals in Category B, who each have probability 0.005 of dying next year. The benefit payable to a dead individual in Category A is 50 with probability 0.42 and 130 with probability 0.58. The benefit payable to a dead individual in Category B is 500 with probability 0.22 and 520 with probability 0.78.

Find Var(S), the variance of aggregate losses next year for this group of 30 life insurance policies.

Solution S4C40-5. We use the formula $Var(S) = {}_{j=1}{}^{n}\Sigma(\sigma_j{}^2*q_j + \mu_j{}^2*q_j*(1-q_j))$.

For Category A, $\mu_A = 96.4$, $q_A = 0.01$ (from Solution S4C40-4).

We find $\sigma_A{}^2 = 0.42*(50-96.4)^2 + 0.58*(130-96.4)^2 = 1559.04$.

Thus, $\sigma_A{}^2*q_A + \mu_A{}^2*q_A*(1-q_A) = 1559.04*0.01 + 96.4^2*0.01*0.99 = 107.590704$.

It follows that $12(\sigma_A{}^2*q_A + \mu_A{}^2*q_A*(1-q_A)) = 1291.088448$.

For Category B, $\mu_B = 515.6$, $q_B = 0.005$ (from Solution S4C40-4).

We find $\sigma_B{}^2 = 0.22*(500-515.6)^2 + 0.78*(520-515.6)^2 = 68.64$.

Thus, $\sigma_B{}^2*q_B + \mu_B{}^2*q_B*(1-q_B) = 68.64^2*0.005 + 515.6^2*0.005*0.995 = 1346.127964$.

It follows that $18(\sigma_B{}^2*q_B + \mu_B{}^2*q_B*(1-q_B)) = 24230.30335$.

Thus, $Var(S) = 12(\sigma_A{}^2*q_A + \mu_A{}^2*q_A*(1-q_A)) + 18(\sigma_B{}^2*q_B + \mu_B{}^2*q_B*(1-q_B)) =$

$1291.088448 + 24230.30335 =$ **Var(S) = 25521.3918**.

Section 41

Parametric Approximations of Probability Distributions of Aggregate Losses in the Individual Risk Model

The probability distribution of aggregate losses in the individual risk model can be approximated using a parametric approximation, such as a normal or lognormal approximation.

The normal approximation uses the central limit theorem.

For a lognormal distribution with mean μ and standard deviation σ,
$E(S) = \exp(\mu + 0.5\sigma^2)$ and $E(S^2) = \exp(2\mu + 2\sigma^2)$.

The lognormal approximation replaces $E(S)$ with μ and $Var(S)$ with σ^2. Moreover, x, the value at which the distribution is being evaluated, is replaced by ln(x).

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 9, pp. 255-256.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C41-1. Life or Death Insurance Company insures 5 individuals with the following characteristics:

Individual 1: Probability of death in the next year is 0.34. Benefit receivable upon death is 500.

Individual 2: Probability of death in the next year is 0.034. Benefit receivable upon death is 230.

Individual 3: Probability of death in the next year is 0.54. Benefit receivable upon death is 90.

Individual 4: Probability of death in the next year is 0.0001. Benefit receivable upon death is 12000.

Individual 5: Probability of death in the next year is 0.053. Benefit receivable upon death is 1000. From Section 40, we know that for the aggregate loss random variable S, E(S) = 280.62, and Var(S) = 124439.0476. The company adds a 50% relative loading to the net pure premium (that is, the company charges a premium so as to collect 50% more than its expected losses). Use a normal approximation to find the probability that the company will lose money next year. To calculate $\Phi(x)$ for any value of x, use the Excel input "=NORMSDIST(x)".

Solution S4C41-1. The company will take in 1.5*E(S) = 420.93.

We want to find $\Pr(S > 420.93) \approx 1 - \Phi((420.93 - E(S))/\sqrt{(Var(S))}) =$

$1 - \Phi((420.93 - 280.62)/\sqrt{(124439.076)}) =$

$1 - \Phi(0.3977500435)$. To find this answer in MS Excel, we use the input

"=1-NORMSDIST(0.3977500435)". Our result is **0.345407222**.

Problem S4C41-2. Life or Death Insurance Company insures 5 individuals with the following characteristics:

Individual 1: Probability of death in the next year is 0.34. Benefit receivable upon death is 500.

Individual 2: Probability of death in the next year is 0.034. Benefit receivable upon death is 230.

Individual 3: Probability of death in the next year is 0.54. Benefit receivable upon death is 90.

Individual 4: Probability of death in the next year is 0.0001. Benefit receivable upon death is 12000.

Individual 5: Probability of death in the next year is 0.053. Benefit receivable upon death is 1000. From Section 40, we know that for the aggregate loss random variable S, E(S) = 280.62, and Var(S) = 124439.0476. The company adds a 50% relative loading to the net pure premium (that is, the company charges a premium so as to collect 50% more than its expected losses). Use a *lognormal* approximation to find the probability that the company will lose money next year. To calculate $\Phi(x)$ for any value of x, use the Excel input "=NORMSDIST(x)".

Solution S4C41-2. The company will take in 1.5*E(S) = 420.93.

We first find μ and σ by using the equations.
$E(S) = \exp(\mu + 0.5\sigma^2)$ and $E(S^2) = \exp(2\mu + 2\sigma^2)$.

Thus, $280.62 = \exp(\mu + 0.5\sigma^2) \rightarrow$ (i) $5.637001441 = \mu + 0.5\sigma^2$.

$E(S^2) = Var(S) + E(S)^2 = 124439.0476 + 280.62^2 = E(S^2) = 203186.632$.

Thus, $203186.632 = \exp(2\mu + 2\sigma^2) \rightarrow$ (ii) $12.22188021 = 2\mu + 2\sigma^2$.

We take (ii) - 2*(i), giving us the equation $\sigma^2 = 0.9478773232$.

Thus, $\mu = 5.637001441 - 0.5*0.9478773232 = \mu = 5.163062779$.

Now we can use the following approximation:

Pr(S > 420.93) ≈ 1 - Φ((ln(420.93) - 5.163062779)/√(0.9478773232)) =

1 - Φ(0.9032589141). To find this answer in MS Excel, we use the input

"=1-NORMSDIST(0.9032589141)". Our result is **0.183194248**.

Problem S4C41-3. Standard Life Insurance Company insures 12 individuals in Category A, who each have probability 0.01 of dying next year. It also insures 16 individuals in Category B, who each have probability 0.005 of dying next year. The benefit payable to a dead individual in Category A is 50 with probability 0.42 and 130 with probability 0.58. The benefit payable to a dead individual in Category B is 500 with probability 0.22 and 520 with probability 0.78. From Section 40, we know that for the aggregate loss random variable S, E(S) = 57.972 and Var(S) = 25521.3918.The company adds a 75% relative loading to the net pure premium (that is, the company charges a premium so as to collect 75% more than its expected losses). Use a normal approximation to find the probability that the company will lose money next year. To calculate Φ(x) for any value of x, use the Excel input "=NORMSDIST(x)".

Solution S4C41-3. The company will take in 1.75*E(S) = 101.451.

We want to find Pr(S > 101.451) ≈ 1 - Φ((101.451- E(S))/√(Var(S))) =

1 - Φ((101.451 - 57.972)/√(25521.3918)) =

1 - Φ(0.2721619259). To find this answer in MS Excel, we use the input

"=1-NORMSDIST(0.2721619259)". Our result is **0.392748758**.

Problem S4C41-4. Standard Life Insurance Company insures 12 individuals in Category A, who each have probability 0.01 of dying next year. It also insures 16 individuals in Category B, who each have probability 0.005 of dying next year. The benefit payable to a dead individual in Category A is 50 with probability 0.42 and 130 with probability 0.58. The benefit payable to a dead individual in Category B is 500 with probability 0.22 and 520 with probability 0.78. From Section 40, we know that for the aggregate loss random variable S, E(S) = 57.972 and Var(S) = 25521.3918.The company adds a 75% relative loading to the net pure premium (that is, the company charges a premium so as to collect 75% more than its expected losses). Use a *lognormal* approximation to find the probability that the company will lose money next year. To calculate Φ(x) for any value of x, use the Excel input "=NORMSDIST(x)".

Solution S4C41-4. The company will take in 1.75*E(S) = 101.451.
We first find μ and σ by using the equations.
$E(S) = \exp(\mu + 0.5\sigma^2)$ and $E(S^2) = \exp(2\mu + 2\sigma^2)$.
Thus, $57.972 = \exp(\mu + 0.5\sigma^2) \to$ (i) $4.059960135 = \mu + 0.5\sigma^2$.
$E(S^2) = Var(S) + E(S)^2 = 25521.3918 + 57.972^2 = 28882.14458$
Thus, $28882.14458 = \exp(2\mu + 2\sigma^2) \to$ (ii) $10.27097885 = 2\mu + 2\sigma^2$.
We take (ii) - 2*(i), giving us the equation $\sigma^2 = 2.151058578$.
Thus, μ = 4.059960135 - 0.5*2.151058578 = μ = 2.984430846.

Now we can use the following approximation:
$\Pr(S > 101.451) \approx 1 - \Phi((\ln(101.451) - 2.984430846)/\sqrt{(2.151058578)}) =$
$1 - \Phi(1.114885257)$. To find this answer in MS Excel, we use the input "=1-NORMSDIST(1.114885257)". Our result is **0.132449798**.

Problem S4C41-5. Review of Section 40. Standard Life Insurance Company insures 12 individuals in Category A, who each have probability 0.01 of dying next year. It also insures 16 individuals in Category B, who each have probability 0.005 of dying next year. The benefit payable to a dead individual in Category A is 50 with probability 0.42 and 130 with probability 0.58. The benefit payable to a dead individual in Category B is 500 with probability 0.22 and 520 with probability 0.78. From Section 40, we know that for the aggregate loss random variable S, E(S) = 57.972 and Var(S) = 25521.3918. Find $M_S(t)$, the moment generating function of the aggregate loss random variable. You do not need to expand the expression in the parentheses.

Solution S4C41-5. We use the formula $M_S(t) = {}_{j=1}{}^{n}\Pi(1 - q_j + q_j*M_{B\_j}(t))$.

For individuals in Category A, the distribution of benefit random variable B_A is

$\Pr(B_A = 50) = 0.42$;
$\Pr(B_A = 130) = 0.58$.

The corresponding $M_{B\_A}(t) = E(\exp(B_A*t)) = 0.42e^{50t} + 0.58e^{130t}$.

Moreover, for individuals in Category A, $q_A = 0.01$. Thus, for *every* individual in Category A, the function $M_S(t)$ contains a factor of $(1 - 0.01+ 0.01(0.42e^{50t} + 0.58e^{130t})) =$

$(0.99+ 0.0042e^{50t} + 0.0058e^{130t})$. Since there are 12 individuals in Category A, their contribution to the function $M_S(t)$ is the factor $(0.99+ 0.0042e^{50t} + 0.0058e^{130t})^{12}$.

For individuals in Category B, the distribution of benefit random variable B_B is

$\Pr(B_B = 500) = 0.22$;
$\Pr(B_B = 520) = 0.78$.

The corresponding $M_{B\_B}(t) = E(\exp(B_B*t)) = 0.22e^{500t} + 0.78e^{520t}$.

Moreover, for individuals in Category B, $q_B = 0.005$. Thus, for *every* individual in Category B, the function $M_S(t)$ contains a factor of $(1 - 0.005+ 0.005(0.22e^{500t} + 0.78e^{520t})) =$

$(0.995+ 0.0011e^{500t} + 0.0039e^{520t})$. Since there are 18 individuals in Category B, their contribution to the function $M_S(t)$ is the factor $(0.995+ 0.0011e^{500t} + 0.0039e^{520t})^{18}$.

Therefore, $\mathbf{M_S(t) = (0.99+ 0.0042e^{50t} + 0.0058e^{130t})^{12}(0.995+ 0.0011e^{500t} + 0.0039e^{520t})^{18}}$.

Section 42

The Empirical Distribution Function, the Cumulative Hazard Rate Function, and the Nelson-Åalen Estimate

The **empirical distribution function** is $F_n(x)$ = (number of observations $\leq$ x)/n.

The **cumulative hazard rate function** is $H(x) = -\ln(S(x))$. It is noteworthy that $H'(x) = h(x)$, the hazard rate function.

Let n be the size of a sample, and let k be the number of *unique* values (where the jth value is denoted as y_j) that appear in the sample such that $y_1 < y_2 < ... < y_k$. Let s_j be the number of times y_j appears in the sample.

The **risk set** consists of the observations greater than or equal to some given value.

Let $r_j = {}_{i=j}{}^{k}\Sigma s_i$ be the number of observations greater than some given value y_j. Then the empirical distribution is found as follows:

$F_n(x) = 0$ if $x < y_1$;

$F_n(x) = 1 - r_j/n$ if $y_{j-1} \leq x < y_j$, for j = 2, ..., k.

$F_n(x) = 1$ if $x \geq y_k$.

The **Nelson-Åalen estimate** of the cumulative hazard rate function in this situation is as follows:

$\hat{H}(x) = 0$ if $x < y_1$;

$\hat{H}(x) = {}_{i=1}{}^{j-1}\Sigma(s_i/r_i)$ if $y_{j-1} \leq x < y_j$, for j = 2, ..., k.

$\hat{H}(x) = {}_{i=j}{}^{k}\Sigma(s_i/r_i)$ if $x \geq y_k$.

Problems involving the Nelson-Åalen estimate are extremely likely to appear multiple times on Exam 4/C. Subsequent sections of the study guide will also include such problems.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 13, pp. 335-337.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C42-1. You are considering the following observations in a data set:
124, 234, 234, 234, 324, 324, 500, 500, 500, 500, 900, 1020, 1600. For the observation 500, find the number of elements in the associated risk set.

Solution S4C42-1. We wish to find $r_j = {}_{i=j}{}^{k}\Sigma s_i$. There are 7 unique observations in the data set: 124, 234, 324, 500, 900, 1020, 1600. Thus, k = 7, and 500 corresponds to j = 4.

For j = 4 (value 500), $s_4 = 4$.
For j = 5 (value 900), $s_5 = 1$.
For j = 6 (value 1020), $s_6 = 1$.
For j = 7 (value 1600), $s_7 = 1$.

Thus, the number of elements in the risk set associated with the value 500 is 4 + 1 + 1 + 1 = **7**.

Problem S4C42-2. You are considering the following observations in a data set:
124, 234, 234, 234, 324, 324, 500, 500, 500, 500, 900, 1020, 1600. Find the empirical distribution function $F_n(x)$ for this set.

Solution S4C42-2. There are 13 elements altogether in the set, so n = 13.

We use the formula $F_n(x)$ = (number of observations ≤ x)/n to get the following:

$F_{13}(x) = 0$ if $x < 124$;
$F_{13}(x) = 1/13$ if $124 \leq x < 234$;
$F_{13}(x) = 4/13$ if $234 \leq x < 324$;
$F_{13}(x) = 6/13$ if $324 \leq x < 500$;
$F_{13}(x) = 10/13$ if $500 \leq x < 900$;
$F_{13}(x) = 11/13$ if $900 \leq x < 1020$;
$F_{13}(x) = 12/13$ if $1020 \leq x < 1600$;
$F_{13}(x) = 1$if $1600 \leq x$.

Problem S4C42-3. You are considering the following observations in a data set:
124, 234, 234, 234, 324, 324, 500, 500, 500, 500, 900, 1020, 1600. Find the Nelson-Åalen estimate of the cumulative hazard rate function for this data set.

Solution S4C42-3. We use the following formula:

$\hat{H}(x) = 0$ if $x < y_1$;
$\hat{H}(x) = {}_{i=1}{}^{j-1}\Sigma(s_i/r_i)$ if $y_{j-1} \leq x < y_j$, for j = 2, ..., k.
$\hat{H}(x) = {}_{i=j}{}^{k}\Sigma(s_i/r_i)$ if $x \geq y_k$.
Naturally, $\hat{H}(x) = 0$ if $x < 124$.
If $124 \leq x < 234$, then j = 2, so $\hat{H}(x) = {}_{i=1}{}^{2-1}\Sigma(s_i/r_i) = s_1/r_1 = 1/13$, so $\hat{H}(x) = 1/13$.
If $234 \leq x < 324$, then j = 3, so $\hat{H}(x) = {}_{i=1}{}^{3-1}\Sigma(s_i/r_i) = s_1/r_1 + s_2/r_2 = 1/13 + 3/12 = \hat{H}(x) = 17/52$.

If $324 \leq x < 500$, then j = 4, so $\hat{H}(x) = {}_{i=1}{}^{4-1}\Sigma(s_i/r_i) = s_1/r_1 + s_2/r_2 + s_3/r_3 = 17/52 + 2/9 = \hat{H}(x) =$ 257/468.
If $500 \leq x < 900$, then j = 5 so $\hat{H}(x) = {}_{i=1}{}^{4}\Sigma(s_i/r_i) = s_1/r_1 + s_2/r_2 + s_3/r_3 + s_4/r_4 = 257/468 + 4/7 =$ 3671/3276.
If $900 \leq x < 1020$, then j = 6 so $\hat{H}(x) = {}_{i=1}{}^{5}\Sigma(s_i/r_i) = 3671/3276 + s_5/r_5 = 3671/3276 + 1/3 =$ 4763/3276.
If $1020 \leq x < 1600$, then j = 7, so $\hat{H}(x) = {}_{i=1}{}^{6}\Sigma(s_i/r_i) = 4763/3276 + ½ = 6401/3276$.
If $1600 \leq x$, then $\hat{H}(x) = {}_{i=j}{}^{7}\Sigma s_i = 6401/3276 + 1/1 = 9677/3276$.

Thus, the Nelson-Åalen estimate of the cumulative hazard rate function for this data set is as follows:

$\hat{H}(x) = 0$ if $x < 124$;
$\hat{H}(x) = 1/13 = 0.0769230769$ if $124 \leq x < 234$;
$\hat{H}(x) = 17/52 = 0.3269230769$ if $234 \leq x < 324$;
$\hat{H}(x) = 257/468 = 0.5491452991$ if $324 \leq x < 500$;
$\hat{H}(x) = 3671/3276 = 1.120563871$ if $500 \leq x < 900$;
$\hat{H}(x) = 4763/3276 = 1.453907204$ if $900 \leq x < 1020$;
$\hat{H}(x) = 6401/3276 = 1.953907204$ if $1020 \leq x < 1600$;
$\hat{H}(x) = 9677/3276 = 2.953907204$ if $1600 \leq x$.

Problem S4C42-4. The survival function for an exponential distribution with parameter θ is $S(x) = e^{-x/\theta}$. Find the cumulative hazard rate function for an exponential distribution.

Solution S4C42-4. We use the formula $H(x) = -\ln(S(x)) = -\ln(e^{-x/\theta}) = -(-x/\theta) =$ **$H(x) = x/\theta$**.

Problem S4C42-5. You are considering the following observations in a data set:
124, 234, 234, 234, 324, 324, 500, 500, 500, 500, 900, 1020, 1600. Using the Nelson-Åalen estimate of the cumulative hazard rate function for this data set, estimate the empirical distribution function $F^{\wedge}(x)$.

Solution S4C42-5. $H(x) = -\ln(S(x))$, so $\hat{H}(x) = -\ln(\hat{S}(x))$, and thus $\hat{S}(x) = \exp(-\hat{H}(x))$, which implies that

$F^{\wedge}(x) = 1 - \hat{S}(x) = 1 - \exp(-\hat{H}(x))$. The values of $\hat{H}(x)$ are known from Solution S4C42-3. Thus,
$F^{\wedge}(x) = 1 - \exp(-0) = 1 - 1 =$ **$F^{\wedge}(x) = 0$ if $x < 124$;**
$F^{\wedge}(x) = 1 - \exp(-1/13) =$ **$F^{\wedge}(x) = 0.0740389214$ if $124 \leq x < 234$;**
$F^{\wedge}(x) = 1 - \exp(-17/52) =$ **$F^{\wedge}(x) = 0.2788607869$ if $234 \leq x < 324$;**
$F^{\wedge}(x) = 1 - \exp(-257/468) =$ **$F^{\wedge}(x) = 0.4225568593$ if $324 \leq x < 500$;**
$F^{\wedge}(x) = 1 - \exp(-3671/3276) =$ **$F^{\wedge}(x) = 0.673907394$ if $500 \leq x < 900$;**
$F^{\wedge}(x) = 1 - \exp(-4763/3276) =$ **$F^{\wedge}(x) = 0.7663444377$ if $900 \leq x < 1020$;**
$F^{\wedge}(x) = 1 - \exp(-6401/3276) =$ **$F^{\wedge}(x) = 0.8582807376$ if $1020 \leq x < 1600$;**
$F^{\wedge}(x) = 1 - \exp(-9677/3276) =$ **$F^{\wedge}(x) = 0.947864397$ if $1600 \leq x$.**

Section 43

Kernel Smoothed Distributions, Ogives, and Histograms

Sections 13.1 and 13.3 of *Loss Models* (cited below) provide definitions of some relevant kinds of distributions. Some of these may be review, while others introduce new ideas.

"A **data-dependent distribution** is at least as complex as the data or knowledge that produced it, and the number of 'parameters' increases as the number of data points or amount of knowledge increases."

"A **parameteric distribution** is a set of distribution functions, each member of which is determined by specifying one or more values called **parameters**. The number of parameters is fixed and finite."

"A **kernel smoothed distribution** is obtained by replacing each data point with a continuous random variable and then assigning probability 1/n to each such random variable. The random variables used must be identical except for a location or scale change that is related to its associated data point."

The **empirical distribution** is a type of kernel smoothed distribution where the random variable gives probability 1 to each data point.

Sometimes data is grouped in categories or ranges of values such that creating an empirical distribution is not possible. In that situation, we can denote the group boundaries as $c_0 < c_1 < ... < c_k$, where c_k is either the maximum possible value or ∞. We can denote the number of observations whose values are between c_{j-1} and c_j as n_j. The sum of all the n_j is n, the total number of observations. We define $F_n(c_j) = (1/n)_{i=1}{}^j\Sigma n_i$. That is, $F_n(c_j)$ is the empirical distribution function evaluated at c_j.Using this notation, the following definitions can be established:

"For grouped data, the distribution function obtained by connecting the values of the empirical distribution function at the group boundaries with straight lines is called the **ogive.** The formula is $F_n(x) = (c_j-x)*F_n(c_{j-1})/(c_j-c_{j-1}) + (x-c_{j-1})F_n(c_j)/(c_j-c_{j-1})$."

"For grouped data, the empirical density function can be obtained by differentiating the ogive. The resulting function is called a **histogram**. The formula is

$f_n(x) = (F_n(c_j) - F_n(c_{j-1}))/(c_j-c_{j-1}) = n_j/(n(c_j-c_{j-1}))$, for $c_{j-1} \leq x < c_j$.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 13, pp. 331-332, 339-341.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C43-1. You are considering the following observations in a data set:
124, 234, 234, 234, 324, 324, 500, 500, 500, 500, 900, 1020, 1600. Develop a kernel-smoothed distribution associated with this set, such that each point x is associated with a uniform distribution that has positive probability over the interval from x-10 to x + 10. As your answer, write the probability density function (pdf) of the kernel smoothed distribution, denoting it as $f_X(x)$.

Solution S4C43-1. A uniform distribution that has positive probability over the interval from x-10 to x + 10 has a pdf of 1/20. Thus, for each value x, we take 1/20 and multiply it by the empirical probability of that value occurring to get $f_X(x)$ for the interval from x-10 to x + 10.

Thus, the following are true:

If $114 \leq x \leq 134$, then $f_X(x)$ = (1/20)*(empirical probability of 124) = (1/20)*(1/13) = 1/260.

Likewise, if $224 \leq x \leq 244$, $f_X(x)$ = (1/20)*(3/13) = 3/260.

If $314 \leq x \leq 334$, $f_X(x)$ = (1/20)*(2/13) = 2/260 = 1/130.

If $490 \leq x \leq 510$, $f_X(x)$ = (1/20)*(4/13) = 4/260 = 1/65.

If $890 \leq x \leq 910$, $f_X(x)$ = (1/20)*(1/13) = 1/260.

If $1010 \leq x \leq 1030$, $f_X(x)$ = (1/20)*(1/13) = 1/260.

If $1590 \leq x \leq 1610$, $f_X(x)$ = (1/20)*(1/13) = 1/260.

Thus, our answer is as follows:

$f_X(x)$ = 1/260 if $114 \leq x \leq 134$;
$f_X(x)$ = 3/260 if $224 \leq x \leq 244$;
$f_X(x)$ = 1/130 if $314 \leq x \leq 334$;
$f_X(x)$ = 1/65 if $490 \leq x \leq 510$;
$f_X(x)$ = 1/260 if $890 \leq x \leq 910$;
$f_X(x)$ = 1/260 if $1010 \leq x \leq 1030$;
$f_X(x)$ = 1/260 if $1590 \leq x \leq 1610$.

Problem S4C43-2. You have the following data set describing the number of slugs eaten by 1000 purple slug-eating monsters yesterday.

Number of Slugs Eaten.............Number of Monsters in this Range
0-35..18
35-95...230
95-145...320
145-300...120
300-500...312

Find the distribution function of the *ogive* corresponding to this data set.

Solution S4C43-2. We use the formula $F_n(x) = (c_j-x)*F_n(c_{j-1})/(c_j-c_{j-1}) + (x-c_{j-1})*F_n(c_j)/(c_j-c_{j-1})$.

First, we find all of the $F_n(c_j)$:

$F_n(0) = 0$.

$F_n(35) = 18/1000 = 0.018$.

$F_n(95) = (18+230)/1000 = 0.248$.

$F_n(145) = (18+230+320)/1000 = 0.568$.

$F_n(300) = (18+230+320+120)/1000 = 0.688$.

$F_n(500) = 1$.

Now, we can develop the ogive:

$F_n(x) = (35-x)*F_n(0)/(35-0) + (x-0)F_n(35)/(35-0) = 0.018x/35 =$

$\mathbf{F_{1000}(x) = 0.0005142857143x \text{ if } 0 \leq x \leq 35}$.

$F_n(x) = (95-x)*F_n(35)/(95-35) + (x-35)F_n(95)/(95-35) = (95-x)(0.0003) + (x-35)(0.00413333333) =$ $\mathbf{F_{1000}(x) = -0.1161666667 + 0.00383333333x \text{ if} 35 \leq x \leq 95}$.

$F_n(x) = (145-x)*F_n(95)/(145-95) + (x-95)F_n(145)/(145-95) =$

$(145-x)(0.00496) + (x-95)(0.01136) =$ $\mathbf{F_{1000}(x) = -0.36 + 0.0064x \text{if} 95 \leq x \leq 145}$.

$F_n(x) = (300-x)*F_n(145)/(300-145) + (x-145)F_n(300)/(300-145) =$

$(300-x)*0.0036645161 + (x-145)*0.0044387097 =$

$\mathbf{F_{1000}(x) = 0.4557419235 + 0.0007741935774x \text{ if} 145 \leq x \leq 300}$.

$F_n(x) = (500-x)*F_n(300)/(500-300) + (x-300)F_n(500)/(500-300) =$

(500-x)*0.00344 + (x-300)*0.005 = **F_{1000}(x) = 0.22 + 0.00156x if 300 ≤ x ≤ 500**.

Problem S4C43-3. You have the following data set describing the number of slugs eaten by 1000 purple slug-eating monsters yesterday.

Number of Slugs Eaten	Number of Monsters in this Range
0-35	18
35-95	230
95-145	320
145-300	120
300-500	312

Find the probability density function of the *histogram* corresponding to this data set.

Solution S4C43-3. The histogram pdf is just the derivative of the ogive distribution function, which we obtained in Solution S4C43-2. Thus, the histogram pdf is as follows:

f_{1000}(x) = 0.0005142857143 if 0 ≤ x ≤ 35;
f_{1000}(x) = 0.00383333333 if 35 ≤ x ≤ 95;
f_{1000}(x) = 0.0064 if95 ≤ x ≤ 145;
f_{1000}(x) = 0.0007741935774 if145 ≤ x ≤ 300;
f_{1000}(x) = 0.00156 if 300 ≤ x ≤ 500.

Problem S4C43-4. You have the following data set describing the number of slugs eaten by 1000 purple slug-eating monsters yesterday.

Number of Slugs Eaten	Number of Monsters in this Range
0-35	18
35-95	230
95-145	320
145-300	120
300-500	312

Using an ogive and/or histogram, approximate the probability that a given purple slug-eating monster has eaten between 234 and 315 slugs, inclusive.

Solution S4C43-4. We will use the histogram from Solution S4C43-3.

For x between 234 and 300, we will use pdf f_{1000}(x) = 0.0007741935774.

For x between 300 and 315 we will use pdf f_{1000}(x) = 0.00156.

These are uniform distribution pdfs, so to get the probability for an interval to which they apply, we simply multiply them by the width of the interval.

Thus, Pr(234 ≤ X ≤ 315) = (300-234)(0.0007741935774) + (315-300)(0.00156) =

Pr(234 ≤ X ≤ 315) = 0.0744967761.

Problem S4C43-5. You have the following data set describing the number of slugs eaten by 1000 purple slug-eating monsters yesterday.

Number of Slugs Eaten	Number of Monsters in this Range
0-35	18
35-95	230
95-145	320
145-300	120
300-500	312

Using an ogive and/or histogram, determine E(X Λ 200), where X is the number of slugs eaten by a monster. You may use a calculator or a computer to solve any integrals, although the integrals are also doable by hand.

Solution S4C43-5. We use the formula $E(X \wedge u) = {}_0^u\int x*f(x)dx + u*(1 - F(u))$ for u = 200.

First, we find $F_{1000}(200)$ using our ogive from Solution S4C43-2, where we found that

$F_{1000}(x) = 0.4557419235 + 0.0007741935774x$ if $145 \leq x \leq 300$.

Thus, $F_{1000}(200) = 0.4557419235 + 0.0007741935774*200 = F_{1000}(200) = 0.610580639$.

Thus, $u*(1 - F(u)) = 200(1-0.610580639) = 77.8838722$.

The ${}_0^{200}\int x*f(x)dx$ component of the formula will need to be split into separate integrals based on differences in the values of $f_{1000}(x)$, which we obtain from our histogram from Solution S4C43-2.

Thus, ${}_0^{200}\int x*f(x)dx = {}_0^{35}\int 0.0005142857143x*dx + {}_{35}^{95}\int 0.00383333333x*dx + {}_{95}^{145}\int 0.0064x*dx + {}_{145}^{200}\int 0.0007741935774x*dx = 61.01016144$.

Thus, E(X Λ 200) = 61.01016144 + 77.8838722 = **E(X Λ 200) = 138.8940336**.

Section 44

Exam-Style Questions on the Smoothed Empirical Estimate, the Nelson-Åalen Estimate, and Aggregate Random Variables

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

For this section, it will be useful to employ two definitions from Klugman, Panjer, and Willmot 2008, p. 377:

"A **percentile matching estimate** of θ is any solution of the p equations

$\pi_{g\_k}(\theta) = \hat{\pi}_{g\_k}$ for k = 1, 2, ..., p where $g_1, g_2, ..., g_p$ are p arbitrarily chosen percentiles. From the definition of percentile, the equations can also be written as $F(\hat{\pi}_{g\_k} | \theta) = g_k$,for k = 1, 2, ..., p."

"The **smoothed empirical estimate** of a percentile is found by

$\hat{\pi}_g = (1-h)x_{(j)} + hx_{(j+1)}$, where $j = \lfloor (n+1)g \rfloor$ and $h = (n+1)g-j$, where $\lfloor . \rfloor$ is the greatest integer function and $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$ are the order statistics from the sample." In this definition, n is the number of values in the sample.

(In this situation, it is easier to think of $x_{(i)}$ as the ith value in the sample, counting up from the sample's smallest value.)

Source:

Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 15, p. 377.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C44-1. Similar to Question 1 of the Exam C Sample Questions from the Society of Actuaries. A loglogistic distribution has the following cumulative distribution function (cdf): $F(x) = (x/\theta)^{\gamma}/(1 + (x/\theta)^{\gamma})$.

You have the following sample of 14 values:
3, 4, 6, 7, 18, 29, 40, 43, 55, 67, 190, 290, 339, 435.

Find the parameter γ of this distribution by using percentile matching at the 20th and 50th smoothed empirical percentile estimates.

Solution S4C44-1.

We use the definition $\hat{\pi}_g = (1-h)x_{(j)} + hx_{(j+1)}$, where $j = \lfloor (n+1)g \rfloor$ and $h = (n+1)g-j$.

Here, our values of g are 0.2 and 0.5.

There are 14 values in the sample, so n = 14, but the empirically smoothed percentile estimate uses (n+1) as the number by which g is multiplied, so the 20th percentile will occur at the 15*0.2 = the 3rd value, or 6. The 50th percentile will occur at the 15*0.5 = 7.5th value. This holds because we are using smoothed empirical percentile estimates. The 7.5th value is halfway between the 7th value, 40, and the 8th value, 43. Thus, the 50th percentile will occur at 41.5.

We set up our system of equations:

(i) $0.2 = (6/\theta)^\gamma/(1 + (6/\theta)^\gamma)$.

(ii) $0.5 = (41.5/\theta)^\gamma/(1 + (41.5/\theta)^\gamma)$.

These equations will require some creative manipulation to be tractable. We can solve (i) for $(6/\theta)^\gamma$: $0.2(1 + (6/\theta)^\gamma) = (6/\theta)^\gamma \rightarrow 0.2 = 0.8(6/\theta)^\gamma \rightarrow (1/4) =(6/\theta)^\gamma$.

We can solve (ii) for $(41.5/\theta)^\gamma$: $0.5(1 + (41.5/\theta)^\gamma) = (41.5/\theta)^\gamma \rightarrow 0.5 = 0.5(41.5/\theta)^\gamma \rightarrow$

$(41.5/\theta)^\gamma = 1$.

Now we can perform the following division: $(41.5/\theta)^\gamma /(6/\theta)^\gamma = 1/(1/4) = 4 \rightarrow (41.5/6)^\gamma = 4 \rightarrow 6.9166666667^\gamma = 4 \rightarrow \gamma = \ln(4)/\ln(6.9166666667) = \boldsymbol{\gamma = 0.7168261126}$.

Problem S4C44-2. Similar to Question 30 of the Exam C Sample Questions from the Society of Actuaries. There are 500 green rabbits in a population, and every time a rabbit emigrates, another rabbit moves in to take its place. Thus, there are always exactly 500 rabbits in the population at any given time. Each month, the following numbers of rabbits emigrate:

Month 1: 34
Month 2: 12
Month 3: 24
Month 4: 36
....
Month q: 12(q-1)

The cumulative distribution function of the number of rabbits that have emigrated has been approximated using a Nelson-Åalen estimate such that F^(q) = 0.3481885852 at the end of month q. Find q.

Solution S4C44-2. The Nelson-Åalen estimateof the cumulative hazard rate function in this situation is as follows:
$\hat{H}(x) = 0$ if $x < y_1$;
$\hat{H}(x) = {}_{i=1}{}^{j-1}\Sigma(s_i/r_i)$ if $y_{j-1} \leq x < y_j$, for j = 2, ..., k.
$\hat{H}(x) = {}_{i=j}{}^{k}\Sigma s_i$ if $x \geq y_k$.

To find q, we first want to find $\hat{H}(q)$.
$F\char`^(x) = 1-\exp(-\hat{H}(x))$, so $0.3481885852 = 1-\exp(-\hat{H}(q)) \rightarrow \exp(-\hat{H}(q)) = 0.6518114148$, and
$\hat{H}(q) = -\ln(0.6518114148) = \hat{H}(q) = 0.428$.

Since the number of rabbits is always the same, our risk set is always the same, and $r_i = 500$ for each i. This means that for month q, the Nelson-Åalen estimateof the cumulative hazard rate function is 34/500 + 12/500 + 24/500 + ... + 12(q-1)/500 =
(34 + 12(1 + 2 + ... (q-1)))/500 = (34 + 12(q-1)(q)/2)/500 = (34 + 6(q-1)(q))/500 = 0.428.
Thus, 214 = (34 + 6(q-1)(q)) → 180 = 6(q-1)(q) → 30 = (q-1)(q).
We know that 30 = 6*5, so q-1= 5 and **q = 6**.

Problem S4C44-3. Similar to Question 31 of the Exam C Sample Questions from the Society of Actuaries. You are given the following sample of values:
234, 235, 335, 369, 456, 476, 549, 560, 578, 591, 610, 634, 655, 678, 701, 1270

Calculate the 78th empirically smoothed percentile of this sample.

Solution S4C44-3.

We use the definition $\hat{\pi}_g = (1-h)x_{(j)} + hx_{(j+1)}$, where $j = \lfloor (n+1)g \rfloor$ and $h = (n+1)g-j$.

Here, g = 0.78 and n = 16, so $j = \lfloor (17)*0.78 \rfloor = \lfloor 13.26 \rfloor = j = 13$. Our 13th value is 655. Thus, $\hat{\pi}_{0.78}$ is 0.26 of the way between 655 and 678 or 655 + (678-655)*0.26 = $\boldsymbol{\hat{\pi}_{0.78} = 660.98}$.

Problem S4C44-4. Similar to Question 33 of the Exam C Sample Questions from the Society of Actuaries. There is only a certain number of cows living in the sky. Once a cow falls from the sky, no other cow ascends to take its place. Exactly one cow falls from the sky every day. By the end of the second day, the Nelson-Åalen estimateof the cumulative hazard rate function for falling cows was $\hat{H}(2) = 59/870$. By the end of the sixth day, what will be the value of $\hat{H}(6)$?

Solution S4C44-4. We use the formula $\hat{H}(x) = {}_{i=1}{}^{j-1}\Sigma(s_i/r_i)$ if $y_{j-1} \leq x < y_j$, for j = 2, ..., k.

We know that $s_i = 1$ for all i. Because the number of cows in the sky is limited, r_i will decrease by 1 every day. Let x be the initial number of cows in the sky. Then $\hat{H}(2) = 1/x + 1/(x-1) =$
$(x + x - 1)/(x(x-1)) = (2x-1)/(x(x-1)) = 59/870$.

We thus need to find x such that (2x-1) = 59 and x(x-1) = 870.

(2x-1) = 59 → 2x = 60 → x = 30. It so happens that 30*29 = 870, so there were 30 cows in the sky initially. By the end of day 6, only 24 cows will remain.

Therefore, $\hat{H}(6)$ = 1/30 + 1/29 + 1/28 + 1/27 + 1/26 + 1/25 = **$\hat{H}(6)$ = 0.2190289532**.

Problem S4C44-5. Similar to Question 53 of the Exam C Sample Questions from the Society of Actuaries.

A gambler uses a slot machine where he can win either 0, 1, or 2 times. *Each* payoff he receives is independent from every other; thus, even if he wins twice, his payoff can be different each time. The following are the probabilities associated with the gambler's winnings:

Number of Wins (N)......Probability(Number of wins)
0...2/3
1...1/4
2.......................................1/12

If the Gambler Wins Once:

Amount (X)...........Probability of Amount
30.....................................1/4
50.....................................3/4

If the Gambler Wins Twice:

Amount (X) for *One Win*...........Probability of Amount
60...2/5
100...3/5

Find the variance of the gambler's aggregate winnings S.

Solution S4C44-5. First, we want to find the probability of each amount of aggregate winnings.
Pr(S = 0) = 2/3, since the gambler wins no money if he does not win any games.
Pr(S = 30) = Pr(N = 1 and X = 30) = (1/4)(1/4) → Pr(S = 30) = 1/16.
Pr(S = 50) = Pr(N = 1 and X = 50) = (1/4)(3/4) → Pr(S = 50) = 3/16.
If the gambler wins twice, he can win either 60*2 = 120, 100*2 = 200, or 60 + 100 = 160.
Pr(S = 120) = (1/12)(2/5)(2/5) = Pr(S = 120) = 1/75.
Pr(S = 160) = 2(1/12)(2/5)(3/5) = Pr(S = 160) = 1/25.
Pr(S = 200) = (1/12)(3/5)(3/5) = Pr(S = 200) = 3/100.
Thus, E(S) = (2/3)(0) + (1/16)(30) + (3/16)(50) + (1/75)(120) + (1/25)(160) + (3/100)(200) = E(S) = 25.25.
Hence, Var(S) = $(2/3)(0-25.25)^2 + (1/16)(30-25.25)^2 + (3/16)(50-25.25)^2 + (1/75)(120-25.25)^2 + (1/25)(160-25.25)^2 + (3/100)(200-25.25)^2$ = **Var(S) = 2303.4375**.

Section 45

Properties of Estimators

When we are given a sample and are trying to estimate a population parameter, some estimators may have more desirable properties than others. The following are definitions (given in Klugman, Panjer, and Willmot, Chapter 12, cited below) of some concepts useful in making such determinations.

An estimator $\theta^{\wedge}$ of θ is **unbiased** if $E(\theta^{\wedge} | \theta) = \theta$ for all θ. The **bias** is represented as $bias_{\theta^{\wedge}}(\theta)= E(\theta^{\wedge} | \theta) - \theta$.

Let $\theta^{\wedge}{}_n$ be an estimator of θ based on a sample size of n. The estimator is **asymptotically unbiased** if $\lim_{n\to\infty}(E(\theta^{\wedge}{}_n | \theta)) = \theta$ for all θ.

An estimator is **consistent** or **weakly consistent** if, for all $\delta > 0$ and any θ,

$\lim_{n\to\infty}(\Pr(|\theta^{\wedge}{}_n - \theta| > \delta)$. To show that an estimator is weakly consistent, it is sufficient to show that the estimator is asymptotically unbiased and that $\lim_{n\to\infty}(Var(\theta^{\wedge}{}_n)) = 0$.

The **mean squared error (MSE)** of an estimator is $MSE_{\theta^{\wedge}}(\theta) = E((\theta^{\wedge} - \theta)^2 | \theta)$.

The following equation is true: $MSE_{\theta^{\wedge}}(\theta) = Var(\theta^{\wedge} | \theta) + (bias_{\theta^{\wedge}}(\theta))^2$.

An estimator $\theta^{\wedge}$ is called a **uniform minimum variance unbiased estimator (UMVUE)** if it is unbiased and for any true value of θ there is no other unbiased estimator that has a smaller variance.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source:

Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 12, pp. 315-321.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C45-1. A population has three entities with the following heights in meters: 3, 5, 6.

A random sample of two entities from this population is taken *without replacement*, and the mean of the sample $\mu^{\wedge}$ is used as an estimator of the population mean μ. Is $\mu^{\wedge}$ an unbiased estimator? If not, find the bias.

Solution S4C45-1. Here, μ = (3 + 5 + 6)/3 = 14/3.

The possible samples of two values that can be taken without replacement are 3,5; 3,6; and 5;6 - each of which has a probability of 1/3 of occurring. The mean of 3,5 is 4. The mean of 3,6 is 4.5. The mean of 5;6 is 5.5. Thus, $E(\mu^{\wedge} \mid \mu) = (4+4.5+5.5)/3 = 14/3 = \mu$. Therefore, **$\mu^{\wedge}$ is an unbiased estimator of the population mean**.

Problem S4C45-2. A population has three entities with the following heights in meters: 3, 5, 6.

A random sample of two entities from this population is taken *with replacement*, and the mean of the sample $m^{\wedge}$ is used as an estimator of the population mean μ. Is $m^{\wedge}$ an unbiased estimator? If not, find the bias.

Solution S4C45-2. Here, μ = (3 + 5 + 6)/3 = 14/3.

The possible samples of two values that can be taken without replacement are

3,3; 3,5; 3,6; 5,3; 5,5; 5,6; 6,3; 6,5; 6,6. Each of these has probability of (1/9) of occurring.

3,3 has a mean of 3; 3,5 and 5,3 have means of 4; 3,6 and 6,3 have means of 4.5; 5,5 has a mean of 5; 5,6 and 6,5 have means of 5.5; 6,6 has a mean of 6.

Thus, $E(m^{\wedge} \mid \mu) = (3 + 2*4 + 2*4.5 + 5 + 2*5.5 + 6)/9 = 42/9 = 14/3 = \mu$. Therefore, **$m^{\wedge}$ is an unbiased estimator of the population mean**.

Problem S4C45-3. Similar to Question 161 of the Exam C Sample Questions from the Society of Actuaries. Which of the following statements is true? More than one answer may be correct.

(a) An unbiased estimator always has a mean square error of zero.
(b) A uniform minimum variance unbiased estimator is an unbiased estimator such that for the true value of the parameter, no other estimator has a smaller variance.
(c) Any asymptotically unbiased estimator is consistent.
(d) Mean square error is always a function of the true value of the parameter being estimated.
(e) For a consistent estimator, the variance of the estimator increases without bound as the sample size increases without bound.

(f) It is possible for an estimator to exist such that the variance of this estimator is smaller than the variance of the uniform minimum variance unbiased estimator.

Solution S4C45-3.

(a) is false: $MSE_{\theta^\wedge}(\theta) = Var(\theta^\wedge \mid \theta) + (bias_{\theta^\wedge}(\theta))^2$ implies that if the estimator is unbiased, $bias_{\theta^\wedge}(\theta) = 0$, but $MSE_{\theta^\wedge}(\theta) = Var(\theta^\wedge \mid \theta)$, which can still be positive.

(b) is true: An estimator $\theta^\wedge$ is called a uniform minimum variance unbiased estimator (UMVUE)if it is unbiased and for any true value of θ there is no other unbiased estimator that has a smaller variance.

(c) is false: An asymptotically unbiased estimator is guaranteed to be consistent only if $\lim_{n\to\infty}(Var(\theta^\wedge{}_n)) = 0$.

(d) is true: $MSE_{\theta^\wedge}(\theta) = E((\theta^\wedge - \theta)^2 \mid \theta)$, so the mean square error depends on the true value of θ.

(e) is false: For many consistent estimators, the variance of the estimator approaches zero as the sample size increases without bound.

(f) is true: Such an estimator can exist, provided that it is not unbiased. The UMVUE only has the smallest variance among *unbiased* estimators.

Thus, **(b), (d), and (f) are true.**

Problem S4C45-4. A population parameter θ is 9340. An estimator $\theta^\wedge$ of θgives one of four different values, with equal probability: 9300, 9324, 9450, 9460. Find the bias of $\theta^\wedge$.

Solution S4C45-4. We use the formula $bias_{\theta^\wedge}(\theta) = E(\theta^\wedge \mid \theta) - \theta$.

$E(\theta^\wedge \mid \theta) = (9300 + 9324 + 9450 + 9460)/4 = 9383.5$. Thus, $bias_{\theta^\wedge}(\theta) = 9383.5 - 9340 =$ $\mathbf{bias_{\theta^\wedge}(\theta) = 43.5}$.

Problem S4C45-5. A population parameter θ is 9340. An estimator $\theta^\wedge$ of θgives one of four different values, with equal probability: 9300, 9324, 9450, 9460. Find the mean square error of $\theta^\wedge$.

Solution S4C45-5. We use the formula $MSE_{\theta^\wedge}(\theta) = E((\theta^\wedge - \theta)^2 \mid \theta) =$

$(1/4)(9300-9340)^2 + (1/4)(9324-9340)^2 + (1/4)(9450-9340)^2 + (1/4)(9460-9340)^2 =$

$\mathbf{MSE_{\theta^\wedge}(\theta) = 7089}$.

Section 46

Interval Estimators and Hypothesis Testing

This section will review some of the basic ideas of mathematical statistics, including confidence intervals, Type I and Type II errors, and hypothesis testing. All definitions are found in Section 12.3 of *Loss Models* (cited below).

"A 100(1-α)% **confidence interval** for a parameter θ is a pair of random values, L and U, computed from a random sample such that $\Pr(L \le \theta \le U) \ge 1 - \alpha$ for all θ." The value 1-α is also called the **level of confidence.**

A **hypothesis test** involves a null hypothesis (H_0) and an alternative hypothesis (H_1 or H_A). The null hypothesis typically includes a statement of equality such as "=", "≤", or "≥". The alternative hypothesis typically includes a statement of strict inequality. The values for which the null hypothesis is rejected in favor of the alternative hypothesis are called the **rejection region.** A **test statistic** is calculated for a hypothesis test; it is a function of the observations and is treated as a random variable. If the test statistic is beyond the boundary of the rejection region - or the **critical value** - on the side containing the rejection region, then the null hypothesis is rejected in favor of the alternative hypothesis. Otherwise, we *fail to reject* the null hypothesis, but we never conclusively accept it.

Type I error is the probability of rejecting a null hypothesis, given that the null hypothesis is true. "The **significance level** of a hypothesis test is the probability of making a Type I error given that the null hypothesis is true. If it can be true in more than one way, the level of significance is the maximum of such probabilities. The significance level is usually denoted by α."

Type II error is the probability of not rejecting the null hypothesis when the alternative hypothesis is true. Typically, attempting to reduce a test's Type I error increases its Type II error, and vice versa, but not necessarily by the same amount.

"A hypothesis test is **uniformly most powerful** if no other test exists that has the same or lower significance level and, for a particular value within the alternative hypothesis, has a smaller probability of making a Type II error."

"For a hypothesis test, the **p-value** is the probability that the test statistic takes on a value that is less in agreement with the null hypothesis than the value obtained from the sample. Tests conducted at a significance level that is greater than the p-value will lead to a rejection of the null hypothesis, while tests conducted at a significance level that is smaller than the p-value will lead to a failure to reject the null hypothesis."

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries.

They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source:
Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 12, pp. 324-330.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C46-1. Which of these tests is the uniformly most powerful, based on the associated probabilities of Type I and Type II errors? The Type II errors are associated with a particular value within the alternative hypothesis. Assume that these are all the possible hypothesis tests in this situation.

(a) Type I error probability: 0.10; Type II error probability: 0.15
(b) Type I error probability: 0.05; Type II error probability: 0.15
(c) Type I error probability: 0.05; Type II error probability: 0.05
(d) Type I error probability: 0.10; Type II error probability: 0.05

Solution S4C46-1. The most powerful test has both the lowest Type I error probability and the lowest Type II error probability. Thus, the correct answer among the four choices is

(c) Type I error probability: 0.05; Type II error probability: 0.05.

Problem S4C46-2. A hypothesis test has a p-value of 0.042. At which of these significance levels would you reject the null hypothesis?

(a) 0.20
(b) 0.15
(c) 0.10
(d) 0.05
(e) 0.025
(f) 0.01
(g) 0.001

Solution S4C46-2. The null hypothesis is rejected whenever p-value < significance level. Thus, all the significance levels greater than 0.042 will result in rejection of the null hypothesis. These values are **(a) 0.20; (b) 0.15; (c) 0.10; (d) 0.05**.

Problem S4C46-3. Similar to Question 71 of the Exam C Sample Questions from the Society of Actuaries. You are examining the distribution of revolutions in countries throughout the world during the past 10 years. Here is how observed revolutions correspond with historical probabilities.

Number of Revolutions in Region...Historical Probability....Countries With This Number
0..0.665....................................60
1..0.315....................................35
2+..0.020.....................................5

Conduct a Chi-square goodness of fit test to determine whether the observed revolutions correspond to the historical frequency of revolutions. At which level of confidence would you reject the null hypothesis that there has been no significant change in the observed frequency of revolutions? Use the Exam 4 / C Tables where necessary.

(a) Reject at the 0.005 significance level.
(b) Reject at the 0.010 significance level, but not at the 0.005 level.
(c) Reject at the 0.025 significance level, but not at the 0.010 level.
(d) Reject at the 0.050 significance level, but not at the 0.025 level.
(e) Do not reject at the 0.050 significance level.

Solution S4C46-3. There are 100 countries in the sample, so 66.5 can be expected to have no revolutions, 31.5 can be expected to have 1 revolution, and 2 can be expected to have 2 or more revolutions. Our chi-square statistic is thus $(60\text{-}66.5)^2/66.5 + (35\text{-}31.5)^2/31.5 + (5\text{-}2)^2/2 =$ 5.524227235. We have three categories of data and therefore 3-1 = 2 degrees of freedom.

We look at our table (p. 4 of the Exam 4 / C Tables) to find that a significance level of 0.950 has a critical value of 5.991 for 2 degrees of freedom. Our test statistic is 5.524227235 < 5.991, so we fail to reject the null hypothesis at the most generous of the significance levels. The correct answer is therefore **(e) Do not reject at the 0.050 significance level**.

Problem S4C46-4. You are performing a hypothesis test as follows:

H_0: X follows a uniform distribution from 3 to 10.

H_1: X follows a uniform distribution from 8 to 15.

You pick a random value of X. This is x, your test statistic. Your test statistic in this case is 9. What is the p-value of this test?

Solution S4C46-4. In this case, the p-value is the probability that, if the null hypothesis is true, a higher value than the test statistic is observed.

The pdf of X if H_0 is true is f(x) = 1/(10-3) = 1/7, which means that S(x) = 1 - (x-3)/7. For x = 9, our p-value is S(9) = 1 - (9-3)/7 = **p = 1/7**.

Problem S4C46-5. You are performing a hypothesis test as follows:
H_0: X follows a uniform distribution from 3 to 10.

H_1: X follows a uniform distribution from 8 to 15.

You pick a random value of X. This is x, your test statistic. The significance level of the test is set to 0.05. What is the probability of Type II error for this test?

Solution S4C46-5. The rejection region for the null hypothesis is such that Pr(Type I error) = 0.05. The pdf of X if H_0 is true is f(x) = 1/(10-3) = 1/7, which means that S(x) = 1 - (x-3)/7. For Pr(Type I error) to be 0.05, we would want S(x) if H_0 is true to be 0.05. Thus, 0.05 = 1 - (x-3)/7 → 0.95 = (x-3)/7 → 6.65 = x-3 → 9.65 = x. Thus, any x greater than 9.65 would result in rejecting H_0. If H_1 is true, then the pdf of X is 1/(15-8) = 1/7, where x ranges from 8 to 15. Thus, if H_1 is true, any observed x between 8 and 9.65 would result in a false failure to reject the null hypothesis. The probability of this happening is (1/7)*(9.65-8) = **Pr(Type II error) = 0.2357142857**.

Section 47

Modifications of Observations and Calculations Pertaining to the Risk Set for Modified Observations

Observations in a sample can be modified by truncation and censoring, as discussed in Section 14.1 of *Loss Models* (cited below):

"An observation is **truncated from below** or **left truncated** at d if when it is below d it is not recorded, but when it is above d it is recorded at its observed value."

"An observation is **truncated from above** or **right truncated** at u if when it is above u it is not recorded, but when it is below u it is recorded at its observed value."

"An observation is **censored from below** or **left censored** at d if when it is below d it is recorded as being equal to d, but when it is above d it is recorded at its observed value."

"An observation is **censored from above** or **right censored** at u if when it is above u it is recorded as being equal to u, but when it is below u it is recorded at its observed value."

Left truncation and right censoring are the most common of the above modifications. Left truncation applies, for instance, when there is an ordinary deductible of d. Right censoring applies when there is a policy limit of u.

Now we introduce some notation: If the jth observation was left truncated, we call the value at which this truncation occurred d_j. If no truncation occurred, then $d_j = 0$. If the jth observation was right censored, we call the value at which this censoring occurred u_j. We let the jth observation be denoted x_j if the observed value does not have truncation or censoring applied to it. If censoring was applied to the observed value, we call the resulting observation u_j. We also let k be the number of *unique* observations that appear in a sample, with the value of the jth unique observation denoted y_j and $y_1 < y_2 < ... < y_k$. We then let s_j be the number of times that an *uncensored* observation appears in the sample. Now can present some formulas for determining the size of the **risk set**, r_j, associated with each unique observation y_j:

Formula 47.1: r_j = (number of $x_i \geq y_j$) + (number of $u_i \geq y_j$) - (number of $d_i \geq y_j$)

Formula 47.2: r_j = (number of $d_i < y_j$) - (number of $x_i < y_j$) - (number of $u_i < y_j$)

Formula 47.3: $r_j = r_{j-1}$ + (number of d_i such that $y_{j-1} \leq d_i < y_j$) - (number of x_i equal to y_{j-1}) - (number of u_i such that $y_{j-1} \leq u_i < y_j$)

Note: r_0 is assumed to be zero.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 14, pp. 343-345.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C47-1. You have the following modified observations:

A: Observed value of 4, left truncated at 5, right censored at 7.

B: Observed value of 2, left truncated at 1, right censored at 7.

C: Observed value of 5, left truncated at 3, right censored at 6.

D: Observed value of 5, left truncated at 2, right censored at 4.

E: Observed value of 3, left truncated at 2, right censored at 5.

Find the size of the risk set associated with the value 4.

Solution S4C47-1.

A was left truncated at 5, so the observed value of 4 becomes 5.

B's observed value was 2, outside the boundaries for left truncation or right censoring.

C's observed value was 5, outside the boundaries for left truncation or right censoring.

D was right censored at 4, so the observed value of 5 becomes 4.

E's observed value was 3, outside the boundaries for left truncation or right censoring.

Thus, for observation 4, we have observations A, C, and D that have values greater or equal to 4. Thus, our risk set has **3 elements.**

Problem S4C47-2.

There are 15 unique observations in a sample, encompassing all the integers from 1 to 15.

(Note: There may be more than 15 overall observations in the sample.)

You know that the size of the risk set associated with the second unique observation (2) is 30.

There are 4 observations that are left truncated at 2. There are 7 observations that are left truncated at 3.

There are 6 observations that are right censored at 2. There are 2 observations that are right censored at 3.

There are also 7 observations that are equal to 2 and 8 observations that are equal to 3 without any truncation or censoring.
Find the size of the risk set associated with the third unique observation (3).

Solution S4C47-2. We use the formula $r_j = r_{j-1}$ + (number of d_i such that $y_{j-1} \leq d_i < y_j$) - (number of x_i equal to y_{j-1}) - (number of u_i such that $y_{j-1} \leq u_i < y_j$) for j = 3.

Thus, (number of d_i such that $y_2 \leq d_i < y_3$) is the number of observations that are left truncated at 2, i.e., 4.

The value (number of x_i equal to y_2) is the number of observations that are equal to 2 and were not truncated or censored in order to obtain this value. This number is given as 7.

The value (number of u_i such that $y_2 \leq u_i < y_3$) is the number of observations that are right censored at 2, i.e., 6.

Thus, $r_3 = r_2 + 4 - 7 - 6 = 30 + 4 - 7 - 6 =$ $\mathbf{r_3 = 21}$.

Problem S4C47-3. There are 15 unique observations in a sample, encompassing all the integers from 1 to 15.

(Note: There may be more than 15 overall observations in the sample.)

You know that the size of the risk set associated with the second unique observation (2) is 30.

There are 4 observations that are left truncated at 2. There are 7 observations that are left truncated at 3.

There are 6 observations that are right censored at 2. There are 2 observations that are right censored at 3.

There are also 7 observations that are equal to 2 and 8 observations that are equal to 3 without any truncation or censoring.
Find the size of the risk set associated with the fourth unique observation (4).

Solution S4C47-3. We use the formula $r_j = r_{j-1}$ + (number of d_i such that $y_{j-1} \leq d_i < y_j$) - (number of x_i equal to y_{j-1}) - (number of u_i such that $y_{j-1} \leq u_i < y_j$) for j = 4. From Solution S4C47-2, we know that $r_3 = 21$.

Thus, (number of d_i such that $y_3 \leq d_i < y_4$) is the number of observations that are left truncated at 3, i.e., 7.

The value (number of x_i equal to y_3) is the number of observations that are equal to 3 and were not truncated or censored in order to obtain this value. This number is given as 8.

The value (number of u_i such that $y_3 \leq u_i < y_4$) is the number of observations that are right censored at 3, i.e., 2.

Thus, $r_4 = r_3 + 7 - 8 - 2 = 21 + 7 - 8 - 2 =$ $\mathbf{r_4 = 18}$.

Problem S4C47-4. You are considering the following unmodified observations in a data set: 124, 234, 234, 234, 324, 324, 500, 500, 500, 500, 900, 1020, 1600.

Each jth observation is now left truncated at an amount equal to

((j-1)/j + (j/2))*original observation. This is done for every observation except the seventh observation (one of the values of 500). For the observation 500, find the number of elements in the associated risk set.

Solution S4C47-4.

The modified observations are as follows:

124 is left truncated at 124*((1-1)/1 + 1/2) = 62 → result is 124.

234 is left truncated at 234*((2-1)/2 + 2/2) = 351 → result is 351.

234 is left truncated at 234*((3-1)/3 + 3/2) = 507 → result is 507.

234 is left truncated at 234*((4-1)/4 + 4/2) = 643.5 → result is 643.5.

We note that for j ≥ 3, the observations are left truncated by a number that is at least 2.166666667 times the original observation, as ((j-1)/j + (j/2)) increases for larger values of j. Thus, the third modified observation and all subsequent modified observations will be 500 or greater. There are 13 total modified observations, of which only 2 are *less than* 500. For the observation 500, find the number of elements in the associated risk set is 13 - 2 = **11** for the modified data set.

Problem S4C47-5. There are 3 unique observations in a sample: 15, 65, and 67.

In this sample, the following is true:

19 observations have been left truncated at 4.

30 observations have been left truncated at 34.

26 observations have been left truncated at 66.

21 observations have been left truncated at 84.

4 observations have been right censored at 4.

3 observations have been right censored at 34.

4 observations have been right censored at 66.

11 observations have been right censored at 84.

No observations were both truncated and censored, and all truncated and censored observations have been affected by the truncation or censoring (i.e., the original value of the observation was not preserved).

There is exactly one observation for each of the following values: 15, 65, and 67. None of these observations has been modified at all.

Find the size of the risk set associated with the observation 65.

Solution S4C47-5. We use the formula

r_j = (number of $d_i < y_j$) - (number of $x_i < y_j$) - (number of $u_i < y_j$), where $y_j = 65$.

We are given that 19 observations have been left truncated at 4 and 30 observations have been left truncated at 34. Both 4 and 34 are less than 65.

Moreover, all the censored observations, and the three non-modified observations, were not truncated, so for them $d_i = 0 < 65$.

The sum of these observations is 4 + 3 + 4 + 11 + 3 = 25.

Thus, (number of $d_i < y_j$) = 19 + 30 + 25 = 74.

The value (number of $x_i < y_j$) is in this case the number of non-modified observations less than 65, which is equal to 1.

The value (number of $u_i < y_j$) is the number of observations that have been right-censored below 65. 4 observations have been right censored at 4. 3 observations have been right censored at 34. Thus,

(number of $u_i < y_j$) = 3 + 4 = 7.

Our answer is thus r_j = 74 - 1 - 7 = **66**.

Section 48

The Kaplan-Meier Product-Limit Estimator

Using the data in a sample, it is possible to estimate the survival function for the relevant population by using the **Kaplan-Meier product-limit estimator.**

Let a sample have size n. Let there be k unique observations in the sample, with the value of the jth unique observation denoted y_j and $y_1 < y_2 < ... < y_k$. We then let s_j be the number of times that an *uncensored* observation appears in the sample and let r_j be the size of the risk set associated with that observation. Then the Kaplan-Meier product-limit estimator approximates the survival function as follows:

$S_n(t) = 1$ if $0 \le t < y_1$;

$S_n(t) = {}_{i=1}{}^{j-1}\Pi((r_i - s_i)/r_i)$ if $y_{j-1} \le t < y_j$ for j = 2, ..., k;

$S_n(t) = {}_{i=1}{}^{k}\Pi((r_i - s_i)/r_i)$ or 0 for $y_k \le t$.

Note that the variable t refers to the value of an *unmodified* observation - one that has not been truncated or censored.

Alternatively, if $y_k \le t$ and if right-censoring has occurred in the sample, we can use an exponential curve to estimate the survival distribution past the point for which sample data is available. Let $w = \max\{x_1, ..., x_n, u_1, ..., u_n\}$, where $u_1, ..., u_n$ are the values, if any, at which observations $x_1, ..., x_n$ are right-censored. Using this method of estimation,

$S_n(t) = (s^*)^{t/w}$ for $w \le t$. Here, $s^* = {}_{i=1}{}^{k}\Pi((r_i - s_i)/r_i)$.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Sources:
Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 14, pp. 345-346.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C48-1. You are considering the following observations in a data set: 124, 234, 234, 234, 324, 324, 500, 500, 500, 500, 900, 1020, 1600. The observations are not truncated or censored. Use a Kaplan-Meier product-limit estimator to approximate the survival function for this data.

Solution S4C48-1. We use the definition of the estimator:

$S_n(t) = 1$ if $0 \leq t < y_1$;

$S_n(t) = {}_{i=1}{}^{j-1}\Pi((r_i - s_i)/r_i)$ if $y_{j-1} \leq t < y_j$ for j = 2, ..., k;

$S_n(t) = {}_{i=1}{}^{k}\Pi((r_i - s_i)/r_i)$ or 0 for $y_k \leq t$. Here, n = 13 and k = 7

Thus, $\mathbf{S_{13}(t) = 1}$ **if** $\mathbf{0 \leq t < 124}$;

$S_n(124) = (13 - 1)/13 =$ $\mathbf{S_{13}(t) = 12/13}$ **if** $\mathbf{124 \leq t < 234}$;

$S_n(234) = (12/13)((12 - 3)/12) =$ $\mathbf{S_{13}(t) = 9/13}$ **if** $\mathbf{234 \leq t < 324}$;

$S_n(324) = (9/13)((9-2)/9) =$ $\mathbf{S_{13}(t) = 7/13}$ **if** $\mathbf{324 \leq t < 500}$;

$S_n(500) = (7/13)((7-4)/7) =$ $\mathbf{S_{13}(t) = 3/13}$ **if** $\mathbf{500 \leq t < 900}$;

$S_n(900) = (3/13)((3-1)/3) =$ $\mathbf{S_{13}(t) = 2/13}$ **if** $\mathbf{900 \leq t < 1020}$;

$S_n(1020) = (2/13)((2-1)/2) =$ $\mathbf{S_{13}(t) = 1/13}$ **if** $\mathbf{1020 \leq t < 1600}$;

$S_n(1600) = (1/13)((1-1)/1) =$ $\mathbf{S_{13}(t) = 0}$ **if** $\mathbf{1600 \leq t}$.

Note that this is the same as the empirical survival function for this data set.

In general, for unmodified data, the Kaplan-Meier product-limit estimate of the survival function is the same as the empirical survival function. Remembering this may prove highly useful on the exam.

Problem S4C48-2. Similar to Question 135 of the Exam C Sample Questions from the Society of Actuaries.

You are examining a data set. The notation used for each observation i is as follows:
d_i is the left truncation point;

x_i is the observed value if not right censored;

u_i is the observed value if right censored.

There are 6 data points:

A: $d_i = 0$; $x_i = 5$.
B: $d_i = 0$; $u_i = 7$.
C: $d_i = 0$; $x_i = 7$.
D: $d_i = 0$; $x_i = 9$.
E: $d_i = 6$; $x_i = 19$.
F: $d_i = 8$; $u_i = 25$.

Use the Kaplan-Meier product-limit estimator to estimate the survival function $S_6(9)$.

Solution S4C48-2.

We recall that the value 9 refers to the unmodified observation. For the observations that have been right-censored at values less than 9, we cannot say whether the unmodified observations are greater than 9 or not. Thus, we only know of three unmodified observations that are less than or equal to 9: A, C, and D.

For A, the risk set size is (number of $d_i < 5$) - (number of $x_i < 5$) - (number of $u_i < 5$) = 4 - 0 - 0 = 4.

For A, $s_i = 1$, since $x_i = 5$ only appears once.

Thus, for A, $((r_i - s_i)/r_i) = (4-1)/4 = ¾$.

For C, the risk set size is (number of $d_i < 7$) - (number of $x_i < 7$) - (number of $u_i < 7$) = 5 - 1 - 0 = 4.

For C, $s_i = 1$, since $x_i = 7$ only appears once.

Thus, for C, $((r_i - s_i)/r_i) = (4-1)/4 = ¾$.

For D, the risk set size is (number of $d_i < 9$) - (number of $x_i < 9$) - (number of $u_i < 9$) = 6 - 2 - 1 = 3.

For D, $s_i = 1$, since $x_i = 9$ only appears once.

Thus, for D, $((r_i - s_i)/r_i) = (3-1)/3 = 2/3$.

Our Kaplan-Meier product-limit estimate is the product (3/4)(3/4)(2/3) = $\mathbf{S_6(9) = 3/8 = 0.375}$.

Problem S4C48-3. Similar to Question 174 of the Exam C Sample Questions from the Society of Actuaries. There is a population of n rhinoceroses. Mars has just been terraformed, and the rhinoceroses want to go there. Once a rhinoceros emigrates to Mars, it never comes back. Only one rhinoceros will emigrate to Mars at any given point in time. Let t_k be the time of the kth rhinoceros emigration. No data pertaining to rhinoceros emigration is censored or truncated. At t_2, the Nelson-Åalen estimate of the cumulative hazard rate function is $\hat{H}(t_2) = 69/1190$. Find the Kaplan-Meier product-limit estimate of the survival function at time t_{17}.

Solution S4C48-3. Since only one rhinoceros emigrates at a time, $s_i = 1$ for all i. This means that

$\hat{H}(t_2) = 1/n + 1/(n-1) = (2n-1)(n(n-1))$. If 69 = 2n-1, then n = 35. Also, (35(35-1)) = 1190, so both the numerator and the denominator of the fraction indicate that n = 35. Since the data is unmodified, the Kaplan-Meier product-limit estimate of the survival function is the same as the empirical survival function. Thus, at t_{17}, 17 rhinoceroses of 35 will have emigrated and 18 will remain. Thus, $\mathbf{S_{35}(t_{17}) = 18/35 = 0.5142857143}$, the empirical survival function at t_{17}.

Problem S4C48-4. Similar to Question 135 of the Exam C Sample Questions from the Society of Actuaries.

You are examining a data set. The notation used for each observation i is as follows:
d_i is the left truncation point;

x_i is the observed value if not right censored;

u_i is the observed value if right censored.

There are 8 data points:

A: $d_i = 0$; $x_i = 2$.
B: $d_i = 0$; $u_i = 4$.
C: $d_i = 0$; $x_i = 5$.
D: $d_i = 0$; $x_i = 5$.
E: $d_i = 5$; $x_i = 7$.
F: $d_i = 6$; $u_i = 10$.
G: $d_i = 0$; $x_i = 14$.
H: $d_i = 12$; $u_i = 15$.

Use the Kaplan-Meier product-limit estimator to estimate the survival function $S_8(14)$.

Solution S4C48-4. We recall that the value 14 refers to the unmodified observation. For the observations that have been right-censored at values less than 14, we cannot say whether the unmodified observations are greater than 14 or not. Thus, we only know of the following five observations that are less than or equal to 14: A, C, D, E, G.

For A, the risk set size is (number of $d_i < 2$) - (number of $x_i < 2$) - (number of $u_i < 2$) = 5 - 0 - 0 = 5.

For A, $s_i = 1$, since $x_i = 2$ only appears once.

Thus, for A, $((r_i - s_i)/r_i) = (5-1)/5 = 4/5$.

For C and D, the risk set size is (number of $d_i < 5$) - (number of $x_i < 5$) - (number of $u_i < 5$) = 5 - 1 - 1 = 3. For C and D, $s_i = 2$, since $x_i = 5$ appears twice.

Thus, for C and D, $((r_i - s_i)/r_i) = (3-2)/3 = 1/3$. *This single factor accounts for both C and D, since it takes each observation into consideration in the value $s_i = 2$.*

For E, the risk set size is (number of $d_i < 7$) - (number of $x_i < 7$) - (number of $u_i < 7$) = 7 - 3 - 1 = 3.

For E, $s_i = 1$, since $x_i = 7$ only appears once.

Thus, for E, $((r_i - s_i)/r_i) = (3-1)/3 = 2/3$.

For G, the risk set size is (number of $d_i < 14$) - (number of $x_i < 14$) - (number of $u_i < 14$) = 8 - 4 - 2 = 2.

For G, $s_i = 1$, since $x_i = 14$ only appears once.

Thus, for G, $((r_i - s_i)/r_i) = (2-1)/2 = 1/2$.

Our Kaplan-Meier product-limit estimate is the product (4/5)(1/3)(2/3)(1/2) = $\mathbf{S_8(14) = 4/45 = 0.08888888889}$.

Problem S4C48-5. There are 23 unique values in a sample. A Kaplan-Meier product-limit estimate of the survival function at the largest of these values, 75, is $_{i=1}{}^{23}\Pi((r_i - s_i)/r_i) = 0.15$. You also know that the maximum *uncensored* observation in the sample is 75. You use an exponential curve to estimate the survival distribution past the point for which sample data is available. Estimate the survival function at 90.

Solution S4C48-5. We use the formula $S_n(t) = (s^*)^{t/w}$ for $w \leq t$. Here, $s^* = {}_{i=1}{}^{k}\Pi((r_i - s_i)/r_i)$.

Here, t = 90 and w is given as 75. Moreover, $s^* = 0.15$. Thus, $S_{23}(90) = 0.15^{90/75}$ = $\mathbf{S_{23}(90) = 0.1026383143}$.

Section 49

Full Credibility of Data in Limited Fluctuation Credibility Theory

This section will discuss the concept of full credibility of data within the framework of limited fluctuation credibility theory.

Let X_1, ..., X_n be claims or losses (or magnitudes of something else) experienced by a policyholder (or some other entity), with X_j being experienced at time j.

Meaning of variables used in this section:
ξ: The mean across the members of a group or class of data - presumed to be stable over time. It is the case that $\xi = E(X_j)$ for the values of j under consideration.

X$^-$(X with a bar over it): $X^- = (X_1 + ... + X_n)/n$. That is, $X^-$is the average of past experience for the given policyholder or other entity. It is the case that $E(X^-) = \xi$, but similar values are possible.

Remember that X^- is a *random variable*, while ξ is a *specificnumerical value*.

r: A value greater than 0 but close to 0, selected judgmentally. A usual selection for r is 0.05.

p: A value such that $0 < p < 1$ and such that p is close to 1. This value is selected judgmentally. A usual selection for p is 0.9, but similar values are possible.

σ: The standard deviation of each X_j. That is, $Var(X_j) = \sigma^2$ for all X_j.

The conditions for data from a sample to be **fully credible** are as follows:

Inequality 49.1. $Pr(-r\xi \leq X^- - \xi \leq r\xi) \geq p$

Inequality 49.2. $Pr(|(X^- - \xi)/(\sigma/\sqrt{n})| \leq r\xi\sqrt{n}/\sigma) \geq p$

Now we define y_p as the smallest value for which the inequality

$Pr(|(X^- - \xi)/(\sigma/\sqrt{n})| \leq y) \geq p$ holds. In other words,

$Pr(|(X^- - \xi)/(\sigma/\sqrt{n})| \leq y_p) = p$.

We also define $\boldsymbol{\lambda_0 = (y_p/r)^2}$.

Now we can express several other inequalities which are satisfied if data from a sample is fully credible:

Inequality 49.3. $r\xi\sqrt{(n)}/\sigma \geq y_p$.

Inequality 49.4. $\sigma/\xi \leq r\sqrt{(n)}/y_p = \sqrt{(n)}/\sqrt{(\lambda_0)}$.

Inequality 49.5. $\mathrm{Var}(X^{-}) = \sigma^2/n \leq \xi^2/\lambda_0$.

Inequality 49.6. $n \geq \lambda_0(\sigma/\xi)^2$.

It is often useful to approximate the distribution of X^{-} by using a normal distribution with mean ξ and variance σ^2/n. If n is large, you can use the central limit theorem and assume that

$(X^{-} - \xi)/(\sigma/\sqrt{(n)})$ has a standard normal distribution. In that case, $p = 2\Phi(y_p)-1$ and $\Phi(y_p) = (1+p)/2$, where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 20, pp. 555-559.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C49-1. Data for a policyholder is known for the past 30 years. During that time, the policyholder's average losses per year were 100. To determine full credibility, you select the values r = 0.05 and p = 0.9. The standard deviation of losses in each year is 30. Use a standard normal distribution to approximate the distribution of X^{-} and find the value of y_p.

Solution S4C49-1. Since we are using a standard normal approximation, we use the formula $\Phi(y_p) = (1+p)/2 = (1+0.9)/2 = \Phi(y_p) = 0.95$. To find y_p, we use the Excel input "=NORMSINV(0.95)", obtaining as our answer $\mathbf{y_p = 1.644853627}$.

Problem S4C49-2. Data for a policyholder is known for the past 30 years. During that time, the policyholder's average losses per year were 100. To determine full credibility, you select the values r = 0.05 and p = 0.9. The standard deviation of losses in each year is 30. Use a standard normal distribution to approximate the distribution of X^{-} and find the value of λ_0.

Solution S4C49-2. We use the formula $\lambda_0 = (y_p/r)^2$. We know from Solution S4C49-1 that $y_p = 1.644853627$, so $\lambda_0 = (1.644853627/0.05)^2 = \boldsymbol{\lambda_0} \mathbf{= 1082.217382}$.

Problem S4C49-3. Data for a policyholder is known for the past 30 years. During that time, the policyholder's average losses per year were 100. To determine full credibility, you select the values r = 0.05 and p = 0.9. The standard deviation of losses in each year is 30. Use a standard normal distribution to approximate the distribution of X^{-} and use limited fluctuation credibility theory to determine whether these data are fully credible.

Solution S4C49-3. Here, n = 30. We use the inequality $n \geq \lambda_0(\sigma/\xi)^2$ to verify whether the condition for full credibility is met. We know that $\sigma = 30$, $\xi = 100$, and $\lambda_0 = 1082.217382$ (from

Solution S4C49-2). Thus, $\lambda_0(\sigma/\xi)^2 = 1082.217382(30/100)^2 = 97.39956435$. Since 30 < 97.39956435, the inequality above is not fulfilled, and thus **the data are not fully credible.**

Problem S4C49-4. You are trying to determine whether data for a sample size of n elephants is fully credible. Each elephant's average number of peanuts eaten is 250. The standard deviation of peanuts eaten for each elephant is 10. To determine full credibility, you select the values r = 0.025 and p = 0.97. You use a standard normal distribution to approximate the distribution of X^- .Determine the minimum number of elephants necessary for the sample data to be fully credible, according to limited fluctuation credibility theory.

Solution S4C49-4.

First, we find y_p.

Since we are using a standard normal approximation, we use the formula $\Phi(y_p) = (1+p)/2 = (1+0.97)/2 = \Phi(y_p) = 0.985$. To find y_p, we use the Excel input "=NORMSINV(0.985)", obtaining as our answer $y_p = 2.170090378$. Now we find $\lambda_0 = (y_p/r)^2 = (2.170090378/0.025)^2 = \lambda_0 = 7534.867598$. Now we analyze the inequality $n \geq \lambda_0(\sigma/\xi)^2$. $\lambda_0(\sigma/\xi)^2 = 7534.867598(10/250)^2 = 12.05578816$. Thus, $n \geq 12.05578816$, and the minimum number of elephants needed for full credibility of data is **n = 13**.

Problem S4C49-5. You know that the minimum number of data points in a sample required for the data to be credible is 120. You also know that the mean value of each data point is 1000, and the standard deviation is 800. To determine full credibility, you use r = 0.09 and p = 0.94. Find the value of y_p using limited fluctuation credibility theory.

Solution S4C49-5. The most useful inequality in this case is $\sigma/\xi \leq r\sqrt{(n)}/y_p$. Since the minimum n is 120, we know that $\sigma/\xi = r\sqrt{(120)}/y_p$. Thus, $y_p = r\sqrt{(120)}\xi/\sigma$. Here, $\sigma = 800$ and $\xi = 1000$. Thus,

$y_p = 0.09\sqrt{(120)}*1000/800 =$ $\mathbf{y_p = 1.232375754}$.

Section 50

Estimating the Variances of Empirical Survival Functions and Empirical Probability Estimates

The empirical estimate of the survival function S(x) is $S_n(x) = Y/n$, where n is the sample size, and Y is the number of sample observations greater than x. Y has a binomial distribution with parameters n and S(x). The following formulas are true:

$E(S_n(x)) = S(x)$

$Var(S_n(x)) = S(x)(1-S(x))/n$

It is also possible to estimate $Var(S_n(x))$ using $S_n(x)$ where S(x) is unknown:

$Var\text{^}(S_n(x)) = S_n(x)(1-S_n(x))/n$

Note that in standard uses of this notation, Var^ would have the carrot (^) placed over the entirety of "Var".

We can also produce empirical estimates of probabilities. Let p = Pr(a < X < b). Then p can be estimated by $p\text{^} = S_n(a) - S_n(b)$. Moreover, Var(p^) = p(1-p)/n. In the estimate of variance when p is unknown, p^ is used as a substitute. Var^(p^) = p^(1-p^)/n.

This estimate of variance can be used to create confidence intervals for the true probability. The formula for the endpoints of such a 1-α confidence interval is as follows:

$p_n \pm z_{1-\alpha/2}*\sqrt{(p_n(1-p_n)/n)}$, where $z_{1-\alpha/2}$ is the value for which a standard normal cumulative distribution function $\Phi(z_{1-\alpha/2})$ would give an output of 1-α/2.

Now we consider the situation where data are censored or truncated. The size of the risk set for the jth unique observation, r_j, is assumed to be fixed, as are each of the unique observations y_j. The *number of occurrences of each unique observation*, s_j, can vary and is represented by the variable S_j. Then the following formulas are true:

$E((r_j - S_j)/r_j) = S(y_j)/S(y_{j-1})$.

$Var((r_j - S_j)/r_j) = (S(y_{j-1})- S(y_j))S(y_j)/(r_jS(y_{j-1})^2)$.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 14, pp. 351-353.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C50-1. You are considering the following observations in a data set:
124, 234, 234, 234, 324, 324, 500, 500, 500, 500, 900, 1020, 1600. Estimate the variance of the empirical survival function when x = 500.

Solution S4C50-1. We want to find Var^(S_{13}(500)).Here, n = 13. We find S_{13}(500) = Y/13, where Y is 3, the number of observations greater than 500. Thus, S_{13}(500) = 3/13. Now we use the approximation Var^(S_n(x)) = S_n(x)(1-S_n(x))/n = (3/13)(1-3/13)/13 = **Var^(S_{13}(500)) = 30/169 = 0.1775147929**.

Problem S4C50-2. You are considering the following observations in a data set:
124, 234, 234, 234, 324, 324, 500, 500, 500, 500, 900, 1020, 1600. Empirically estimate the probability that x is between 300 and 940.

Solution S4C50-2. The empirical estimate formula is p^ = S_n(a) - S_n(b) = S_{13}(300) - S_{13}(940).

There are 13-4 = 9 observations greater than 300 and 2 observations greater than 940. Thus,

S_{13}(300) - S_{13}(940) = 9/13 - 2/13 = **p^ = 7/13**.

Problem S4C50-3. You are considering the following observations in a data set:
124, 234, 234, 234, 324, 324, 500, 500, 500, 500, 900, 1020, 1600. Empirically estimate the variance of the probability that x is between 300 and 940.

Solution S4C50-3. We use the formula Var^(p^) = p^(1-p^)/n, where we know from Solution S4C50-2 that p^ = 7/13. Thus, Var^(p^) = (7/13)(1-7/13)/13 = **Var^(p^) = 42/169 = 0.2485207101**.

Problem S4C50-4. You are considering the following observations in a data set:
124, 234, 234, 234, 324, 324, 500, 500, 500, 500, 900, 1020, 1600. Let p(324) be the true probability of getting a value of 324. Let p^(324) be the empirical estimate of that probability. Construct a 90% confidence interval for p(324) using p^(324). In solving the problem, estimate the variance of p(324) empirically.

Solution S4C50-4. Here, sample size n is 13, and 324 occurs twice in the sample, so p^(324) = 2/13. To construct a confidence interval, we need to estimate a variance. We use the formula Var^(p^) = p^(1-p^)/n = Var^(p^(234)) = (2/13)(1-2/13)/13 = 22/169.

A 90% confidence interval is constructed by excluding the rightmost 5% and the leftmost 5% of the area underneath a standard normal curve. Thus, we want to find z such that Φ(z) = 0.95 and Φ(-z) = 0.05. We use the Excel input "=NORMSINV(0.95)" to get z = 1.644853627.

Thus, we know that 0.90 = Pr(-1.644853627 ≤ (p^(234) - p(234))/√(Var^(p^(234))) ≤ 1.644853627) →

0.90 = Pr(-1.644853627 ≤ (2/13 - p(234))/√(22/169) ≤ 1.644853627) →

0.90 = Pr(-0.5934651827 ≤ (2/13 - p(234)) ≤ 0.5934651827) →

0.90 = Pr(-0.7473113365 ≤ (- p(234)) ≤ 0.4396190289) →

0.90 = Pr(-0.4396190289 ≤ p(234) ≤ 0.7473113365).

However, probabilities cannot be less than zero, so

0.90 = Pr(0 ≤ p(234) ≤ 0.7473113365)

and our 90% confidence interval for p(234) is **[0, 0.7473113365]**.

Problem S4C50-5. Assume that the number of rhinoceroses left to emigrate to Mars in the year 2130 is fixed, and that rhinoceroses only emigrate to Mars once a year. It is known that there were 560 rhinoceroses who have not emigrated to Mars in the year 2129, and there are 543 rhinoceroses who have not emigrated to Mars in 2130. Let r_{2130} be the number of rhinoceroses left to emigrate in 2130 and let S_{2130} be the number that emigrate in 2130. Using this information, find $E(((r_{2130} - S_{2130})/r_{2130})^2)$.

Solution S4C50-5. $E(((r_{2130} - S_{2130})/r_{2130})^2) = Var((r_{2130} - S_{2130})/r_{2130})) + E((r_{2130} - S_{2130})/r_{2130}))^2$.

First, we find $E((r_{2130} - S_{2130})/r_{2130}))$ using the formula $E((r_j - S_j)/r_j) = S(y_j)/S(y_{j-1})$.

Here, $y_j = 2130$ and $y_{j-1} = 2129$. Thus, S(2130) = 543/n, where n was the original number of rhinoceroses, and S(2129) = 560/n. Thus, $E((r_{2130} - S_{2130})/r_{2130}))$ = (543/n)/(560/n) = 543/560 and

$E((r_{2130} - S_{2130})/r_{2130}))^2 = (543/560)^2 = 0.9402072704$.

Now we find $Var((r_{2130} - S_{2130})/r_{2130}))$ using the formula

$Var((r_j - S_j)/r_j) = (S(y_{j-1})- S(y_j))S(y_j)/(r_jS(y_{j-1})^2)$.

We know that S(2129) = 560/n, S(2130) = 543/n, and $r_{2130} = 543$.

Thus, $Var((r_{2130} - S_{2130})/r_{2130})) = (560/n - 543/n)(543/n)/(543*(560/n)^2) =$

$(9231/n^2)/(170284800/n^2) = 9231/170284800 = Var((r_{2130} - S_{2130})/r_{2130})) = 0.00005420918367$.

Thus, $E(((r_{2130} - S_{2130})/r_{2130})^2) = Var((r_{2130} - S_{2130})/r_{2130})) + E((r_{2130} - S_{2130})/r_{2130}))^2 =$

0.00005420918367 + 0.9402072704 = $\mathbf{E(((r_{2130} - S_{2130})/r_{2130})^2) = 0.9402614796}$.

Section 51

Greenwood's Approximation and Variances of Products of Random Variables

Let $X_1, ..., X_n$ be n independent random variables. Each X_j has mean μ_j and standard deviation σ_j. Then $Var(X_1*...*X_n) = (\mu_1+\sigma_1)^2*....*(\mu_n+\sigma_n)^2 - \mu_1^2*....*\mu_n^2$.

Now we consider the situation where data are censored or truncated. The size of the risk set for the jth unique observation, r_j, is assumed to be fixed, as are each of the unique observations y_j. The *number of occurrences of each unique observation*, s_j, can vary and is represented by the variable S_j. In Section 50, the following formulas were presented:

$E((r_j - S_j)/r_j) = S(y_j)/S(y_{j-1})$.

$Var((r_j - S_j)/r_j) = (S(y_{j-1}) - S(y_j))S(y_j)/(r_jS(y_{j-1})^2)$.

Using the Kaplan-Meier product-limit estimator, it is possible to estimate the survival function $\hat{S}(y_j)$ for each y_j. The following formulas are true:

$E(\hat{S}(y_j)) = S(y_j)/S(y_0)$

$Var(S_n(y_j)) = (S(y_j)^2/S(y_0)^2)(_{i=1}{}^{j}\Pi(1 + (S(y_{j-1}) - S(y_j))/(r_j*S(y_j)) - 1)$

The latter formula is quite cumbersome, so **Greenwood's approximation** is used to estimate it as follows:

$Var^\wedge(S_n(y_j)) = S(y_j)^2*\ _{i=1}{}^{j}\Sigma(s_i/(r_i(r_i-s_i)))$.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Sources:

Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 14, pp. 356-357.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C51-1. An archaeologist has collected some rare artifacts and has retired from his profession, so he will not collect any more artifacts. He then gives away some number of them once every fixed time period. After his initial giveaway (time period t=0), he had not yet given away 43 artifacts. At the end of his giveaway in time period t=6, he had not yet given away 25 artifacts. A Kaplan-Meier product-limit estimator is used to estimate the survival function of the artifacts in the archaeologist's possession. Find $E(\hat{S}(6))$, the expected value of the survival function at t=6.

Solution S4C51-1. We use the formula $E(\hat{S}(y_j)) = S(y_j)/S(y_0)$. Here, $y_0 = 0$ and $y_j = 6$. Let n be the original number of artifacts. Then 43 of them were given away at some time later than t = 0, and 25 would be given away at some time later than t = 6. Thus, the values of the relevant survival functions are as follows: S(6) = 25/n and S(0) = 43/n. Thus, $E(\hat{S}(6)) = (25/n)/(43/n) =$ **$E(\hat{S}(6))$ = 25/43 = 0.5813953488**.

Problem S4C51-2. You have the following random variables:

X_1 follows an exponential distribution with $\theta_1 = 300$.
X_2 follows a Pareto distribution with $\alpha_2 = 3$ and $\theta_2 = 500$.
X_3 follows a Gamma distribution with $\alpha_3 = 4$ and $\theta_3 = 200$.

Find $Var(X_1*X_2*X_3)$. Round the answer to four decimal places.

Relevant properties for exponential distributions:

$E(X) = \theta$
$Var(X) = \theta^2$

Relevant properties for Pareto distributions:

$E(X) = \theta/(\alpha-1)$
$Var(X) = 2\theta^2/((\alpha-1)(\alpha-2)) - \theta^2/(\alpha-1)^2$

Relevant properties for Gamma distributions:

$E(X) = \alpha\theta$
$Var(X) = \alpha\theta^2$

Solution S4C51-2. We use the formula $Var(X_1*...*X_n) = (\mu_1+\sigma_1)^2*....*(\mu_n+\sigma_n)^2 - \mu_1^2*....*\mu_n^2$.

This formula requires us to find the mean and standard deviation of each of the above random variables.

For X_1: $\mu_1 = \theta_1 = 300$; $\sigma_1 = \sqrt{(\theta_1^2)} = \theta = 300$.

For X_2: $\mu_2 = \theta_2/(\alpha_2-1) = 500/(3-1) = 250$; $\sigma_2 = \sqrt{(2\theta_2^2/((\alpha_2-1)(\alpha_2-2)) - \theta_2^2/(\alpha_2-1)^2)} =$

$\sqrt{(2*500^2/((3-1)(3-2)) - 500^2/(3-1)^2)} = \sqrt{(187500)} = \sigma_2 = 433.0127019$.

For X_3: $\mu_3 = \alpha_3\theta_3 = 4*200 = \mu_3 = 800$; $\sigma_3 = \sqrt{(\alpha_3\theta_3^2)} = \sqrt{(4*200^2)} = \sigma_3 = 400$.

$\mathrm{Var}(X_1*X_2*X_3) = (\mu_1+\sigma_1)^2*(\mu_2+\sigma_2)^2*(\mu_3+\sigma_3)^2 - \mu_1^2*\mu_2^2*\mu_3^2 =$

$(300+300)^2(250+433.0127019)^2(800+400)^2 - 300^2*250^2*800^2$ = approximately **$2.3824*10^{17}$**.

Problem S4C51-3. You are analyzing the following censored and truncated data for numbers of homes struck by meteorites.

Time t=1: There are 39 homes at risk, of which 9 are struck by meteorites.
Time t=2: There are 29 homes at risk, of which 3 are struck by meteorites.
Time t=3: There are 32 homes at risk, of which 5 are struck by meteorites.

Use Greenwood's approximation to estimate the variance of the empirical survival function at time t = 3. That is, find $\mathrm{Var}^\wedge(S_n(3))$. Use the Kaplan-Meier product-limit estimator to estimate values of the survival function.

Solution S4C51-3. The formula for Greenwood's approximation is $S(y_j)^2 * {}_{i=1}{}^{j}\Sigma(s_i/(r_i(r_i-s_i)))$.

We will first find an estimate of S(3), the probability that a home will not be struck by a meteorite at or before time t = 3. We can approximate S(3) using the Kaplan-Meier product-limit estimator: $\Pi((r_i - s_i)/r_i)$ for i = 1, 2, and 3, where r_i is the number of homes at risk at time i and s_i is the number of homes struck. Thus, $S(3) = ((39-9)/39)((29-3)/29)((32-5)/32) = \hat{S}(3) = 135/232$.

Then $\hat{S}(3)^2 = 0.3386035969$.

Now we can find ${}_{i=1}{}^{3}\Sigma(s_i/(r_i(r_i-s_i))) = 9/(39*30) + 3/(29*26) + 5/(32*27) = 0.0174581246$.

Hence, $\mathrm{Var}^\wedge(S_n(3)) = 0.3386035969*0.0174581246$ = **$\mathrm{Var}^\wedge(S_n(3)) = 0.0059113838$**.

Problem S4C51-4. Similar to Question 16 of the Exam C Sample Questions from the Society of Actuaries. You are exploring censored and truncated data for failures of electrical equipment at various times. Here is a segment of this data:

Time t = 30: There are 50 pieces of equipment at risk, and 4 of them fail at t = 30.
Time t = 31: There are 53 pieces of equipment at risk, and 5 of them fail at t = 31.
Time t = 32: There are 32 pieces of equipment at risk, and 9 of them fail at t = 32.
Time t = 33: There are 45 pieces of equipment at risk, and 11 of them fail at t = 33.
Time t = 34: There are 20 pieces of equipment at risk, and 2 of them fail at t = 34.

Let ${}_3q_{30}$ be the probability that a piece of equipment which has survived past time t = 30 will fail at or before time t = 33. Use Greenwood's approximation to estimate the variance of ${}_3q\text{^}_{30}$, the empirical estimate of ${}_3q_{30}$.

Solution S4C51-4. First we note that ${}_3q_{30} = 1 - {}_3p_{30}$, where ${}_3p_{30}$ is the probability that a piece of equipment which has survived past time t = 30 will also survive past time t = 33. Thus, $\text{Var}({}_3q\text{^}_{30}) = \text{Var}(1 - {}_3p\text{^}_{30}) = (-1)^2\text{Var}({}_3p\text{^}_{30}) = \text{Var}({}_3p\text{^}_{30})$.

We can thus use Greenwood's approximation for $\text{Var}({}_3p\text{^}_{30})$ to find our answer. The formula for this approximation is $\text{Var}({}_3p\text{^}_{30}) = ({}_3p\text{^}_{30})^2 * {}_{i=31}{}^{33}\Sigma(s_i/(r_i(r_i-s_i)))$.

Now we find ${}_3p\text{^}_{30} = ((53-5)/53)((32-9)/32)((45-11)/45) = {}_3p\text{^}_{30} = 391/795$.

Next we find ${}_{i=31}{}^{33}\Sigma(s_i/(r_i(r_i-s_i))) = 5/(53*48) + 9/(32*23) + 11/(45*34) = 0.0213832122$.

Thus, $\text{Var}({}_3p\text{^}_{30}) = \text{Var}({}_3q\text{^}_{30}) = (391/795)^2*0.0213832122 =$ **$\text{Var}({}_3q\text{^}_{30}) = 0.0051724012$**.

Problem S4C51-5. Similar to Question 252 of the Exam C Sample Questions from the Society of Actuaries. For a sample of pieces of electronic equipment, either failure times were recorded, or we know that failure occurred later than a specified time. For the latter cases, the specified time at which monitoring ceased is given, along with a "+" superscript to indicate that actual failure occurred later than the specified time. Here are the sample data:

$34, 34, 34, 45^+, 56^+, 77, 77, 89, 97^+, 100, 123, 123, 220.$

Use Greenwood's approximation to find the variance of Ŝ(95), the Kaplan-Meier estimate of the survival function for this data at time 95.

Solution S4C51-5. The formula for Greenwood's approximation is $S(y_j)^2 * {}_{i=1}{}^{j}\Sigma(s_i/(r_i(r_i-s_i)))$.

First, we find Ŝ(95). At time 34, 13 pieces of equipment are at risk, and 3 fail. Thus, 10 do not fail. At time 77, 8 pieces of equipment are at risk; 2 fail and 6 do not fail. At time 89, 6 pieces of equipment are at risk; 1 fails and 5 do not fail. These are all the data we need to consider to find Ŝ(95) = (10/13)(6/8)(5/6) = Ŝ(95) = 0.4807692308.

Now we find ${}_{i=1}{}^{95}\Sigma(s_i/(r_i(r_i-s_i)))$, where the values of i for which s_i is nonzero are 34, 77, and 89.

Then ${}_{i=1}{}^{95}\Sigma(s_i/(r_i(r_i-s_i))) = 3/(13*10) + 2/(8*6) + 1/(6*5) = 0.0980769231$.

Greenwood's approximation gives Var^(Ŝ(95)) = $0.4807692308^2*0.0980769231$ = **Var^(Ŝ(95)) = 0.0226694071**.

Section 52

Log-Transformed Confidence Intervals for Survival Functions and Cumulative Hazard Rate Functions

In this section, we develop **log-transformed confidence intervals** for survival functions using empirical survival functions and estimates of their variances. We also develop linear and log-transformed confidence intervals for cumulative hazard rate functions using the Nelson-Åalen estimator.

Log-transformed confidence intervals using empirical survival functions:

Let $S_n(t)$ be the empirical survival function.

Let v^ = Var^($S_n(t)$).

Let U = exp($z_{1-\alpha/2}$√(v^)/($S_n(t)$*ln($S_n(t)$))) for confidence level 1-α.

Here, $z_{1-\alpha/2}$ is the value of z such that Φ(z) = 1-α/2, where Φ(z) is the standard normal cumulative distribution function.

Then the log-transformed confidence interval using empirical survival functions has the following upper and lower limits:

Lower limit: $S_n(t)^{1/U}$
Upper limit: $S_n(t)^U$

We note that these log-transformed confidence intervals *never* result in probabilities outside the range from 0 to 1, which is extremely desirable for confidence intervals involving probabilities and survival functions.

Linear and log-transformed confidence intervals using the Nelson-Åalen estimator:

An estimate of the variance of the Nelson-Åalen estimator is Var^($\hat{H}(y_j)$) = $_{i=1}{}^{j}\Sigma(s_i/r_i^2)$, where s_i is the number of times the ith unique observation occurs and r_i is the size of the risk set associated with the ith unique observation.

The *linear* 1-α confidence interval for the cumulative hazard rate function is
$\hat{H}(t) \pm z_{1-\alpha/2}$*√(Var^($\hat{H}(y_j)$)).

We let U = exp(±$z_{1-\alpha/2}$*√(Var^($\hat{H}(y_j)$))/$\hat{H}(t)$).

We can call the larger of these values U^+ and the smaller of these values U^-.

The *log-transformed* 1-α confidence interval for the cumulative hazard rate function has the following upper and lower limits:

Lower limit: $\hat{H}(t)*U^-$
Upper limit: $\hat{H}(t)*U^+$

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Sources: Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 14, pp. 358-359.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C52-1. You are exploring censored and truncated data for failures of electrical equipment at various times. Here is a segment of this data:

Time t = 1: There are 50 pieces of equipment at risk, and 4 of them fail.
Time t = 2: There are 53 pieces of equipment at risk, and 5 of them fail.
Time t = 3: There are 32 pieces of equipment at risk, and 9 of them fail.
Time t = 4: There are 45 pieces of equipment at risk, and 11 of them fail.
Time t = 5: There are 20 pieces of equipment at risk, and 2 of them fail.

Find Var^($\hat{H}(5)$), the variance of the Nelson-Åalen estimator at time t = 5.

Solution S4C52-1. We use the formula Var^($\hat{H}(y_j)$) = $_{i=1}{}^{j}\Sigma(s_i/r_i^2)$. Here, the s_i are the numbers of pieces of equipment that fail, and the r_i are the numbers of pieces of equipment that are at risk. Thus, Var^($\hat{H}(5)$) = $4/50^2 + 5/53^2 + 9/32^2 + 11/45^2 + 2/20^2$ = **Var^($\hat{H}(5)$) = 0.0226011541**.

Problem S4C52-2. You are exploring censored and truncated data for failures of electrical equipment at various times. Here is a segment of this data:

Time t = 1: There are 50 pieces of equipment at risk, and 4 of them fail.
Time t = 2: There are 53 pieces of equipment at risk, and 5 of them fail.
Time t = 3: There are 32 pieces of equipment at risk, and 9 of them fail.
Time t = 4: There are 45 pieces of equipment at risk, and 11 of them fail.
Time t = 5: There are 20 pieces of equipment at risk, and 2 of them fail.

Establish a 90% *linear* confidence interval for H(5), the cumulative hazard rate function at time t = 5.

Solution S4C52-2. First, we find the Nelson-Åalen estimate $\hat{H}(5) = {}_{i=1}{}^{5}\Sigma(s_i/r_i)$ = 4/50 + 5/53 + 9/32 + 11/45 + 2/20 = $\hat{H}(5)$ = 0.8000349671.

A 90% confidence interval is constructed by excluding the rightmost 5% and the leftmost 5% of the area underneath a standard normal curve. Thus, we want to find z such that Φ(z) = 0.95 and Φ(-z) = 0.05. We use the Excel input "=NORMSINV(0.95)" to get z = 1.644853627.

Now we use the formula $\hat{H}(t) \pm z_{1-\alpha/2}*\sqrt{}(\text{Var}\hat{}(\hat{H}(y_j)))$.

From Solution S4C52-1, we know that Var^($\hat{H}$(5)) = 0.0226011541.

Thus, $\sqrt{}(\text{Var}\hat{}(\hat{H}(y_j)))$ = 0.1503368024.

Hence, the boundaries of our confidence interval are

0.8000349671 ± 1.644853627*0.1503368024.

The confidence interval is therefore **(0.5527520324, 1.047316102)**.

Problem S4C52-3. Similar to Question 17 of the Exam C Sample Questions from the Society of Actuaries. You are exploring censored and truncated data for failures of electrical equipment at various times. Here is a segment of this data:

Time t = 1: There are 50 pieces of equipment at risk, and 4 of them fail.
Time t = 2: There are 53 pieces of equipment at risk, and 5 of them fail.
Time t = 3: There are 32 pieces of equipment at risk, and 9 of them fail.
Time t = 4: There are 45 pieces of equipment at risk, and 11 of them fail.
Time t = 5: There are 20 pieces of equipment at risk, and 2 of them fail.

Establish a 90% *log-transformed* confidence interval for H(5), the cumulative hazard rate function at time t = 5.

Solution S4C52-3. First, we find the Nelson-Åalen estimate $\hat{H}(5) = {}_{i=1}{}^{5}\Sigma(s_i/r_i)$ = 4/50 + 5/53 + 9/32 + 11/45 + 2/20 = $\hat{H}(5)$ = 0.8000349671.

A 90% confidence interval is constructed by excluding the rightmost 5% and the leftmost 5% of the area underneath a standard normal curve. Thus, we want to find z such that Φ(z) = 0.95 and Φ(-z) = 0.05. We use the Excel input "=NORMSINV(0.95)" to get z = 1.644853627.

Now we calculate $U = \exp(\pm z_{1-\alpha/2}*\sqrt{}(\text{Var}\hat{}(\hat{H}(y_j)))/\hat{H}(t))$.

From Solution S4C52-1, we know that Var^($\hat{H}$(5)) = 0.0226011541.

Thus, $\sqrt{(Var\hat{}(\hat{H}(y_j)))} = 0.1503368024$.

Thus, U = exp(±1.644853627*0.1503368024/0.8000349671).

Hence, $U^+ = 1.362184419$ and $U^- = 0.7341151511$.

Thus, the upper bound of the confidence interval is $\hat{H}(t)*U^+ = 0.8000349671*1.362184419 = 1.0897937$.

The upper bound of the confidence interval is $\hat{H}(t)*U^- = 0.8000349671*0.7341151511 = 0.58731713$.

Our confidence interval is therefore **(0.58731713, 1.0897937)**.

Problem S4C52-4. Similar to Question 36 of the Exam C Sample Questions from the Society of Actuaries. You know that the 95% log-transformed confidence interval for the survival function S(t) at time t = 2 is (0.246, 0.356). A product-limit estimate $\hat{S}(2)$ was used to create this confidence interval. Find $\hat{S}(2)$.

Hint: A confidence interval created using a product-limit estimate is calculated in the exact same manner as a confidence interval created using an empirical survival function. Just substitute the former for the latter. Also, you do not need to use the Kaplan-Meier product-limit estimator formula in this problem.

Hint II: Remember that $\sqrt{(x^2)} = \pm x$. Do not neglect to consider the negative value!

Solution S4C52-4. Our log-transformed confidence interval is as follows:

$(\hat{S}(2)^{1/U}, \hat{S}(2)^U)$. Based on the values of the limits of the interval, we can set up the following system of equations:

(i) $\hat{S}(2)^{1/U} = 0.246$
(ii) $\hat{S}(2)^U = 0.356$

We can take the natural logarithm of each side of each equation:

(i) $(1/U)*\ln(\hat{S}(2)) = \ln(0.246)$
(ii) $U*\ln(\hat{S}(2)) = \ln(0.356)$

Notice that U can be eliminated by multiplying the equations together:

$\ln(\hat{S}(2))^2 = \ln(0.246)*\ln(0.356) \rightarrow$
$\ln(\hat{S}(2))^2 = 1.448457669 \rightarrow$
$\ln(\hat{S}(2)) = \pm 1.203518869$.

Since we are estimating a survival function, values above 1 are nonsensical. Moreover, we know that e taken to a positive power gives an answer greater than 1. Thus, we pick $\ln(\hat{S}(2)) = -1.203518869 \rightarrow$ **$\hat{S}(2)$ = 0.3001362114**.

Problem S4C52-5. Similar to Question 77 of the Exam C Sample Questions from the Society of Actuaries. You are conducting a study, and you know that the 90% *linear* confidence interval for the cumulative hazard rate function at time t = 6 is (0.345, 0.934). Find the 90% *log-transformed* confidence interval for the cumulative hazard rate function at time t = 6.

Solution S4C52-5. It may come as a surprise, but the linear confidence interval is *all the information we need* to determine the log-transformed confidence interval. Consider the formula for the linear confidence interval: $\hat{H}(t) \pm z_{1-\alpha/2}*\sqrt{(\text{Var}^\wedge(\hat{H}(y_j)))}$.

We know that $\hat{H}(6)$ is exactly halfway between the limits of the interval. That is, $\hat{H}(6)$ = (0.345+0.934)/2 = $\hat{H}(6)$ = 0.6395.

A 90% confidence interval is constructed by excluding the rightmost 5% and the leftmost 5% of the area underneath a standard normal curve. Thus, we want to find z such that $\Phi(z) = 0.95$ and $\Phi(-z) = 0.05$. We use the Excel input "=NORMSINV(0.95)" to get z = 1.644853627. (But even finding the value of z is not crucial to solving this problem. You will see why below.

Thus, we know that $0.934 = 0.6395 + 1.644853627*\sqrt{(\text{Var}^\wedge(\hat{H}(6)))}$. From this, we get $1.644853627*\sqrt{(\text{Var}^\wedge(\hat{H}(6)))} = 0.2945$. We keep the square root of the variance estimate multiplied by the z-score, because this is also an expression that appears in the formula for U in the log-transformed confidence interval: $U = \exp(\pm z_{1-\alpha/2}*\sqrt{(\text{Var}^\wedge(\hat{H}(y_j)))}/\hat{H}(t))$.

Thus, here, $U = \exp(\pm 0.2945/0.6395)$.

Hence, $U^+ = 1.584891623$ and $U^- = 0.6309579694$.

Thus, the upper bound of the confidence interval is $\hat{H}(t)*U^+ = 0.6395*1.584891623 = 1.013538193$.

The upper bound of the confidence interval is $\hat{H}(t)*U^- = 0.6395*0.6309579694 = 0.4034976214$.

Thus, our log-transformed confidence interval is **(0.4034976214, 1.013538193)**.

Section 53

Kernel Density Estimators and Uniform, Triangular, and Gamma Kernels

A distribution function can be estimated by means of a **kernel density estimator**, which has the following formula for the cumulative distribution function (cdf):
$F\hat{}(x) = {}_{j=1}{}^{k}\Sigma(p(y_j)*K_{y\_j}(x))$.

The formula for the probability density function is as follows:
$f\hat{}(x) = {}_{j=1}{}^{k}\Sigma(p(y_j)*k_{y\_j}(x))$.

The function $k_y(x)$ is called the **kernel.** There are three common kernels: the uniform, the triangular, and the gamma.

For the uniform kernel:
$k_y(x) = 0$ if $x < y - b$;
$k_y(x) = 1/(2b)$ if $y-b \leq x \leq y + b$;
$k_y(x) = 0$ if $x > y + b$.
$K_y(x) = 0$ if $x < y - b$;
$K_y(x) = (x-y+b)/(2b)$ if $y-b \leq x \leq y + b$;
$K_y(x) = 1$ if $x > y + b$.

Essentially, the uniform kernel is uniform distribution over an interval of width 2b, centered at x.

The value b is called the **bandwidth** of the kernel.

For the triangular kernel:
$k_y(x) = 0$ if $x < y - b$;
$k_y(x) = (x-y+b)/(b^2)$ if $y-b \leq x \leq y$;
$k_y(x) = (y+b-x)/(b^2)$ if $y \leq x \leq y + b$;
$k_y(x) = 0$ if $x > y + b$.
$K_y(x) = 0$ if $x < y - b$;
$K_y(x) = (x-y+b)^2/(2b^2)$ if $y-b \leq x \leq y$;
$K_y(x) = 1-(y+b-x)^2/(2b^2)$ if $y \leq x \leq y + b$;
$K_y(x) = 1$ if $x > y + b$.

For the gamma kernel:

The gamma kernel has a gamma distribution with parameters α and y/α, where y/α is the scale parameter. Thus, $k_y(x) = (x^{\alpha-1}*e^{-x\alpha/y})/((y/\alpha)^{\alpha}*\Gamma(\alpha))$. The mean of the gamma kernel is y and the variance is y^2/α.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Sources: Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 14, pp. 362-365.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C53-1. A uniform kernel of bandwidth 5 is established around each observation x. Find the value of cumulative distribution function $K_{100}(x)$ for x = 104.

Solution S4C53-1. We first see how x compares to y and b. Here, y = 100 and b = 5.

Thus, y + b = 105 and y - b = 95. Since x = 104, y-b ≤ x ≤ y + b and so $K_y(x) = (x-y+b)/(2b) = (104-100+5)/(2*5)$ = **9/10 = $K_{100}(104)$ = 0.9**.

Problem S4C53-2. A triangular kernel of bandwidth 5 is established around each observation x. Find the value of cumulative distribution function $K_{100}(x)$ for x = 104.

Solution S4C53-2. We first see how x compares to y and b. Here, y = 100 and b = 5.

Thus, y + b = 105 and y - b = 95. Since x = 104, y ≤ x ≤ y + b, and so $K_y(x) = 1-(y+b-x)^2/(2b^2) = 1-(105-104)^2/(2*5^2) = 1 - 1/50$ = **$K_{100}(104)$ = 49/50 = 0.98.**

Problem S4C53-3. A gamma kernel with α = 2 is established around each observation x. Find the value of probability density function $k_{100}(x)$ for x = 104.

Solution S4C53-3. We are given that y = 100, so we can use the formula
$k_y(x) = (x^{\alpha-1}*e^{-x\alpha/y})/((y/\alpha)^{\alpha}*\Gamma(\alpha)) = (104^{2-1}*e^{-104*2/100})/((100/2)^2*\Gamma(2))$. Γ(2) = (2-1)! = 1, so
$k_{100}(104) = (104*e^{-2.08})/(2500)$ = **$k_{100}(104)$ = 0.0051970968**.

Problem S4C53-4. Similar to Question 3 of the Exam C Sample Questions from the Society of Actuaries. You are studying 10 pieces of data with the following observed values:
19, 19, 19, 23, 23, 24, 28, 30, 30, 31

You use a triangular kernel of bandwidth 4 to estimate the probability density function. Using this method, find f^(20).

Solution S4C53-4. We use the formula $f\hat{}(x) = {}_{j=1}{}^{k}\Sigma(p(y_j)*k_{y\_j}(x))$, which in this case becomes

$p(19)*k_{19}(20) + p(23)*k_{23}(20) + p(24)*k_{24}(20) + p(28)*k_{28}(20) + p(30)*k_{30}(20) + p(31)*k_{31}(20)$.

To make the calculations more concise, we can try right away to identify values of y for which $k_y(20)$ would be zero. These are cases for which 20 < y - b or 20 > y + b, where b = 4.

Since 20 < y - 4 for all y > 24, we can disregard the terms $p(28)*k_{28}(20)$, $p(30)*k_{30}(20)$, and $p(31)*k_{31}(20)$.

For y = 19, 20 is between y and y + b, and thus we can apply formula $k_y(x) = (y+b-x)/(b^2) = (19+4-20)/(4^2) = 3/16$. Moreover, empirically, p(19) = 3/10.

For y = 23, 20 is between y-b and y, and thus we can apply formula $k_y(x) = (x-y+b)/(b^2) = (20-23+4)/(4^2) = 1/16$. Moreover, empirically, p(23) = 2/10.

For y = 24, 20 is exactly y-b, which is non-strictly between y-b and y, and thus we can apply formula $k_y(x) = (x-y+b)/(b^2) = (20-24+4)/(4^2) = 0/16$. Hence, we can disregard the term $p(24)*k_{24}(20)$ as well.

We have as our answer (3/16)(3/10) + (1/16)(2/10) = **11/160 = f^(20) = 0.06875**.

Problem S4C53-5. Similar to Question 147 of the Exam C Sample Questions from the Society of Actuaries. You know the following ages (X) of a population of squirrel-tailed squirrels: 1, 3, 3, 4, 5, 5, 6, 6, 6, 6.

Use a uniform kernel of bandwidth 3 to estimate the cumulative distribution function at x = 5.

Solution S4C53-5. We use the formula $F\hat{}(x) = {}_{j=1}{}^{k}\Sigma(p(y_j)*K_{y\_j}(x))$. Thus,
$F\hat{}(x) = p(1)*K_1(5) + p(3)*K_3(5) + p(4)*K_4(5) + p(5)*K_5(5) + p(6)*K_6(5)$.

For each value of y, we compare x to y + b and y - b.

For y = 1, y + b = 4, so x > y + b and therefore $K_1(5) = 1$. Empirically, p(1) = 1/10.

For y = 3, y + b = 6, so y - b < x < y + b and therefore $K_3(5) = (x-y+b)/(2b) = (5-3+3)/(2*3) = 5/6$. Empirically, p(3) = 2/10.

For y = 4, y + b = 7, so y - b < x < y + b and therefore $K_4(5) = (x-y+b)/(2b) = (5-4+3)/(2*3) = 4/6$. Empirically, p(4) = 1/10.

For y = 5, it is evident that $K_5(5)$ = ½, since the uniform kernel is centered at x = 5. Empirically, p(5) = 2/10.

For y = 6, y-b = 3 and y + b = 9, so y - b < x < y + b and therefore $K_6(5) = (x-y+b)/(2b) = (5-6+3)/(2*3) = 2/6$. Empirically, p(6) = 4/10.

Thus, F^(5) = 1*1/10 + (5/6)(2/10) + (4/6)(1/10) + (1/2)(2/10) + (2/6)(4/10) = **F^(5) = 17/30 = 0.5666666666667**.

Section 54

Partial Credibility in Limited Fluctuation Credibility Theory

Limited fluctuation credibility theory can also be used to address data that is only *partially* credible and which must be weighted with data gathered from some external source. The following equation holds for partially credible data:

$P_c = Z\bar{X} + (1-Z)M$.

P_c is called the **credibility premium.**

Z is the credibility factor and is chosen to be between 0 and 1, inclusive.

ξ: The mean across the members of a group or class of data - presumed to be stable over time. It is the case that $\xi = E(X_j)$ for the values of j under consideration.

$\bar{X}$ (X with a bar over it): $\bar{X} = (X_1 + ... + X_n)/n$. That is, $\bar{X}$ is the average of past experience for the given policyholder or other entity. It is the case that $E(\bar{X}) = \xi$, but similar values are possible.

Remember that $\bar{X}$ is a *random variable*, while ξ is a *specific numerical value*.

M is the mean of data obtained independently of the data which is being characterized as partially credible.

One way in which Z can be selected is as follows:

$Z = \min((\xi/\sigma)\sqrt{(n/\lambda_0)}, 1)$.

That is, if the data is not fully credible, than $Z = (\xi/\sigma)\sqrt{(n/\lambda_0)}$, which is $\sqrt{(n/\lambda_0)}$ divided by (σ/ξ), the coefficient of variation for the data. This formula is also called the **square root rule** because Z can be conceived of as √(number of observations in data set)/√(number of observations required for full credibility).

Meaning of other variables used in this section:
r: A value greater than 0 but close to 0, selected judgmentally. A usual selection for r is 0.05.

p: A value such that $0 < p < 1$ and such that p is close to 1. This value is selected judgmentally. A usual selection for p is 0.9, but similar values are possible.

σ: The standard deviation of each X_j. That is, $Var(X_j) = \sigma^2$ for all X_j.

y_p: The smallest value for which the inequality

$Pr(|(X^- - \xi)/(\sigma/\sqrt{(n)})| \leq y) \geq p$ holds. In other words,

$Pr(|(X^- - \xi)/(\sigma/\sqrt{(n)})| \leq y_p) = p$.

λ_0 $= (y_p/r)^2$.

When one is working with an aggregate random variable, for which the frequency random variable is Poisson with mean λ and the severity random variable has mean θ_Y and variance $\sigma_Y{}^2$, then the expected number of events needed for the data to be fully credible is nλ and the following inequality is true: $n\lambda \geq \lambda_0(1+(\sigma_Y/\theta_Y)^2)$.

If the standard by which credibility is being judged is in terms of values of the aggregate (be it aggregate losses, aggregate claims, aggregate premiums, or something else), then the expected number of events needed for the data to be fully credible is $n\lambda\theta_Y$ and the following inequality is true: $n\lambda\theta_Y \geq \lambda_0(\theta_Y+(\sigma_Y{}^2/\theta_Y))$.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Sources:

Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 20, pp. 560-563.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C54-1. Data for a policyholder is known for the past 30 years. During that time, the policyholder's average losses per year were 100. To determine full credibility, you select the values r = 0.05 and p = 0.9. The standard deviation of losses in each year is 30. Use a standard normal distribution to approximate the distribution of X^- and use limited fluctuation credibility theory to determine the credibility factor Z that should be assigned to the data.

Solution S4C54-1. Here, n = 30. We use the inequality $n \geq \lambda_0(\sigma/\xi)^2$ to verify the threshold for full credibility. First, we need to determine y_p. Since we are using a standard normal approximation, we use the formula $\Phi(y_p) = (1+p)/2 = (1+0.9)/2 = \Phi(y_p) = 0.95$. To find y_p, we use the Excel input "=NORMSINV(0.95)", obtaining as our answer $y_p = 1.644853627$.

To find λ_0, we use the formula $\lambda_0 = (y_p/r)^2 = (1.644853627/0.05)^2 = \lambda_0 = 1082.217382$.

We know that $\sigma = 30$, $\xi = 100$, and $\lambda_0 = 1082.217382$. Thus, $\lambda_0(\sigma/\xi)^2 = 1082.217382(30/100)^2 = 97.39956435$.

Now we can apply the square root rule: Z = √(number of observations in data set)/√(number of observations required for full credibility) = √(30)/√(97.39956435) = **Z = 0.554986118**.

Problem S4C54-2. Data for a policyholder is known for the past 30 years. During that time, the policyholder's average losses per year were 100. To determine full credibility, you select the values r = 0.05 and p = 0.9. The standard deviation of losses in each year is 30. This data is not fully credible, and so in order for the credibility premium to be determined, it needs to be compared to external data. For policyholders in general average losses per year are 150. Find the credibility premium associated with the given policyholder.

Solution S4C54-2. We use the formula $P_c = ZX^{-} + (1-Z)M$. From Solution S4C54-1, we know that Z = 0.554986118. Moreover, we are given that $X^{-} = 100$ and M = 150. Hence,

$P_c = 0.554986118*100 + (1-0.554986118)*150 =$ **P_c = 122.2506941**.

Problem S4C54-3. Similar to Question 2 of the Exam C Sample Questions from the Society of Actuaries. Claim frequency follows a Poisson distribution, and claim severity follows a Pareto distribution with parameters $\theta = 10$ and $\alpha = 3$. The number of claims and claim sizes are independent, and it is specified that the observed pure premium should be within 3% of the expected pure premium 96% of the time, using a standard normal approximation. Find the minimum number of claims required for full credibility.

Relevant properties for Pareto distributions:

$E(X) = \theta/(\alpha-1)$

$Var(X) = 2\theta^2/((\alpha-1)(\alpha-2)) - \theta^2/(\alpha-1)^2$

Solution S4C54-3. Since we are concerned about the number of events, we use the inequality

$n\lambda \geq \lambda_0(1+(\sigma_Y/\theta_Y)^2)$. Note that we seek to find not n but $n\lambda$, so the value of λ itself is irrelevant.

We are given that r = 0.03 and p = 0.96.

First, we need to determine y_p. Since we are using a standard normal approximation, we use the formula $\Phi(y_p) = (1+p)/2 = (1+0.96)/2 = \Phi(y_p) = 0.98$. To find y_p, we use the Excel input "=NORMSINV(0.98)", obtaining as our answer $y_p = 2.053748911$.

Now we find $\lambda_0 = (y_p/r)^2 = (2.053748911/0.03)^2 = \lambda_0 = 4686.538433$.

We also need to find θ_Y, the mean of the severity distribution, which is equal to

$\theta/(\alpha-1) = 10/(3-1) = 5$.

Furthermore, we need to find σ_Y^2, the variance of the severity distribution:

$\sigma_Y^2 = 2\theta^2/((\alpha-1)(\alpha-2)) - \theta^2/(\alpha-1)^2 = 2*10^2/((3-1)(3-2)) - 10^2/(3-1)^2 = \sigma_Y^2 = 75$.

Thus, $n\lambda \geq \lambda_0(1+(\sigma_Y/\theta_Y)^2) = 4686.538433(1 + 75/5^2) \rightarrow n\lambda \geq 18746.15373$, implying that the minimum number of claims required for full credibility is **18747**.

Problem S4C54-4. Similar to Question 27 of the Exam C Sample Questions from the Society of Actuaries. It is known that, for fully credible data, $\Pr(-r\xi \leq X^- - \xi \leq r\xi) \geq p$. Now we are given partially credible data such that $P_c = ZX^- + (1-Z)M$. It is possible to find an expression of the form $\Pr(\xi - r\xi \leq A \leq \xi + r\xi)$. Find the value of A using the variables given here.

Solution S4C54-4. In partially credible data, X^- is credibility-weighted with M, and so X^- must be replaced by $ZX^- + (1-Z)M$. Likewise, the value ξ is only partially credible and so must be credibility-weighted with the mean M of the external data in the following manner: $Z\xi + (1-Z)M$.

Making this substitution, we get the following result:

$\Pr(-r\xi \leq (ZX^- + (1-Z)M) - (Z\xi + (1-Z)M) \leq r\xi) \geq p$.

We note that the two instances of (1-Z)M cancel each other, and we are left with

$\Pr(-r\xi \leq ZX^- - Z\xi \leq r\xi) \geq p$.

Now we add ξ to each side of the inequality:

$\Pr(\xi - r\xi \leq ZX^- - Z\xi + \xi \leq \xi + r\xi) \geq p \rightarrow$

$\Pr(\xi - r\xi \leq ZX^- + (1- Z)\xi \leq \xi + r\xi) \geq p$.

We have the inequality in its desired form, and so we conclude that $\mathbf{A = ZX^- + (1- Z)\xi}$.

Problem S4C54-5. Similar to Question 65 of the Exam C Sample Questions from the Society of Actuaries. There are 3400 policies of insurance. The frequency of claims per policy follows a negative binomial distribution with $r = 6$ and $\beta = 0.5$. The severity of each claim follows a Pareto distribution with parameters $\theta = 450$ and $\alpha = 3$. The values $r = 0.03$ and $p = 0.96$ are selected in calculations pertaining to credibility. Use limited fluctuation credibility theory (also known as *classical* credibility theory) to determine the credibility factor Z of the annual claims for this subset of insurance policyholders.

Relevant properties for negative binomial distributions:

$E(N) = r\beta$

$Var(N) = r\beta(1+\beta)$

Relevant properties for Pareto distributions:

$E(X) = \theta/(\alpha-1)$

$Var(X) = 2\theta^2/((\alpha-1)(\alpha-2)) - \theta^2/(\alpha-1)^2$

Solution S4C54-5. First, we need to determine the threshold for full credibility of data.

To do this, we refer to the inequality $n \geq \lambda_0(\sigma/\xi)^2$ from Section 49.

We need to find λ_0.

First, we need to determine y_p. Since we are using a standard normal approximation, we use the formula $\Phi(y_p) = (1+p)/2 = (1+0.96)/2 = \Phi(y_p) = 0.98$. To find y_p, we use the Excel input "=NORMSINV(0.98)", obtaining as our answer $y_p = 2.053748911$.

Now we find $\lambda_0 = (y_p/r)^2 = (2.053748911/0.03)^2 = \lambda_0 = 4686.538433$.

Now we need to find ξ and σ, the mean and standard deviation, respectively, of the *aggregate* claims.

To do this we refer to the following formulas from Section 33:

$E(S) = E(N)E(X)$.
$Var(S) = E(N)*Var(X) + Var(N)*E(X)^2$.

Here, $E(N) = r\beta = 6*0.5 = 3$ and $Var(N) = r\beta(1+\beta) = 3*1.5 = 4.5$.

Moreover, $E(X) = 450/(3-1) = 225$ and $Var(X) = 2*450^2/((3-1)(3-2)) - 450^2/(3-1)^2 = Var(X) = 151875$.

Thus, $Var(S) = 3*151875 + 4.5*225^2 = Var(S) = \sigma^2 = 683437.5$.

Moreover, $E(S) = 3*225 = \xi = 675$.

Thus, $(\sigma/\xi)^2 = 683437.5/675^2 = 1.5$.
Hence, $n \geq \lambda_0(\sigma/\xi)^2 = 4686.538433*1.5 = 7029.80765$, which is our threshold for full credibility.
To determine Z for our sample of 3400 policyholders, we apply the square root rule:

$Z = \sqrt{(3400/7029.80765)}$ = **$Z = 0.6954529242$**.

Section 55

The Kaplan-Meier Approximation for Large Data Sets and for Multiple Decrements

If there are large amounts of data, including censored and truncated data, it is convenient to modify the approach taken by the Kaplan-Meier product-limit estimator with regard to smaller data.

Let sample data be split into k intervals with boundary values $c_0 < c_1 < \ldots < c_k$.

Let d_j be the number of observations that are left truncated at some value within the half-closed interval $[c_j, c_{j+1})$.

Let u_j be the number of observations that are right censored at some value within the half-closed interval $(c_j, c_{j+1}]$.

Note that the intervals for d_j and u_j differ by which endpoints are included and which are omitted. This is because left truncation is not possible at the right end of a closed interval, while right censoring is not possible at the left end of a closed interval.

Let x_j be the number of uncensored observations within the interval $(c_j, c_{j+1}]$.

Then the sample size n can be determined as follows: $n = {}_{j=0}{}^{k-1}\Sigma(d_j) = {}_{j=0}{}^{k-1}\Sigma(u_j + x_j)$.

A **Kaplan-Meier approximation for large data sets** is presented here. In this approximation, we designate $\hat{S}(c_0) = 1$.

The risk set r_j at any value j is the number of available observations that could produce an uncensored observation at that value. Here, the following formulas are true:

$r_0 = d_0$;
$r_j = {}_{i=0}{}^{j}\Sigma(d_i) - {}_{i=0}{}^{j-1}\Sigma(u_i + x_i)$ for j = 1, 2, ..., k.
$\hat{S}(c_0) = 1$
$\hat{S}(c_j) = {}_{i=0}{}^{j-1}\Pi(1 - x_i/r_i)$ for j = 1, 2, ..., k.

We can also find an approximation for q_j, the probability that an observation with value greater than c_j will have a value of at most c_{j+1}. This approximation is $q\hat{}_j = x_j/r_j$. This corresponds to the life table formula q_j = (number of deaths in time period)/(number of lives considered during that time period).

In life table applications, it is possible to have multiple causes of decrement or multiple reasons for which individuals die or observations are otherwise observed to occur at certain values.

When there are multiple causes of decrement, we can express life table functions pertaining to *all* causes of decrement with the superscript $^{(\tau)}$. If assign numbers 1 through n to each of the causes of decrement, then we can express life table functions pertaining some cause of decrement i with the superscript $^{(i)}$, where i can be an integer from 1 through n. If you are given each ${}_mq_x^{(i)}$, then the following holds: $\mathbf{q_x^{(\tau)} = 1 - {}_{i=1}{}^{n}\prod(1- q_x^{(i)})}$.In this section, we will work with a Kaplan-Meier estimate involving multiple forces of decrement.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Sources: Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 14, pp. 366-369.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C55-1.

An insurance policy can have either a deductible of 20 or a deductible of 40. Here is data for losses for 200 insurance policies, with the loss intervals given below:

Loss interval......Policies with deductible = 20..........Policies with deductible = 40
(20, 40]........................10...N/A
(40, 100]......................34...34
(100, 400]....................36...46
(400, 9000]..................10...30

Use the Kaplan-Meier approximation for large data sets to estimate F(400).

Solution S4C55-1. Similar to Question 178 of the Exam C Sample Questions from the Society of Actuaries. To estimate the cumulative distribution function, we first estimate the survival function. We use the formula $\hat{S}(c_j) = {}_{i=0}{}^{j-1}\Pi(1- x_i/r_i)$ for j = 1, 2, ..., k.

There are 90 policies with a deductible of 20. Of those policies, 80 have losses over 40. This contributes a factor of 80/90 to our estimate.

There are a total of 190 policies with losses over 40. Of these, 122 policies have losses over 100. This contributes a factor of 122/190 to our estimate.

Of the 122 policies with losses over 100, 40 have losses over 400. This contributes a factor of 40/122 to our estimate.

Thus, $\hat{S}(400) = (80/90)(122/190)(40/122) = 32/171$.
Thus, $F\hat{}(400) = 1 - \hat{S}(400) = 1 - 32/171 =$ **$F\hat{}(400) = 139/171 = 0.8128654931$**.

Problem S4C55-2. Similar to Question 262 of the Exam C Sample Questions from the Society of Actuaries. There are two causes of decrement, falling and explosion. It is known that each decade, 0.03 of the population dies from falling. The total probability of death every decade j is denoted as $q_j^{(\tau)}$ and is given below:

Decade (j)..... $q_j^{(\tau)}$
1..................0.23
2..................0.04
3..................0.15
4..................0.36

Some members of this population are immune to falling. 10000 such individuals are monitored from the beginning of decade 1. For each decade, find $q_j^{(2)}$, the probability that death occurs due to explosion.

Solution S4C55-2. We use our formula $q_x^{(\tau)} = 1 - {}_{i=1}{}^{n}\prod(1 - q_x^{(i)})$, which in this case becomes $q_j^{(\tau)} = 1 - (1 - q_j^{(1)})(1 - q_j^{(2)}) \rightarrow (1 - q_j^{(2)}) = (1 - q_j^{(\tau)})/(1 - q_j^{(1)}) \rightarrow q_j^{(2)} = 1 - (1 - q_j^{(\tau)})/(1 - q_j^{(1)})$. We know that $q_j^{(1)} = 0.03$ for each j. Thus, $q_j^{(2)} = 1 - (1 - q_j^{(\tau)})/(1 - 0.03) = 1 - (1 - q_j^{(\tau)})/0.97 = (0.97 - (1 - q_j^{(\tau)}))/0.97 = (q_j^{(\tau)} - 0.03)/0.97$.

Thus, for j = 1, $q_1^{(2)} = (0.23 - 0.03)/0.97 =$ **$q_1^{(2)} = 0.206185567$**.
For j = 2, $q_2^{(2)} = (0.04 - 0.03)/0.97 =$ **$q_2^{(2)} = 0.0103092784$**.
For j = 3, $q_3^{(2)} = (0.15 - 0.03)/0.97 =$ **$q_3^{(2)} = 0.1237113402$**.
For j = 4, $q_4^{(2)} = (0.36 - 0.03)/0.97 =$ **$q_4^{(2)} = 0.3402061856$**.

Problem S4C55-3. Similar to Question 262 of the Exam C Sample Questions from the Society of Actuaries. There are two causes of decrement, falling and explosion. It is known that each decade, 0.03 of the population dies from falling. The total probability of death every decade j is denoted as $q_j^{(\tau)}$ and is given below:

Decade (j)..... $q_j^{(\tau)}$
1..................0.23
2..................0.04
3..................0.15
4..................0.36

Some members of this population are immune to falling. 10000 such individuals are monitored from the beginning of decade 1. Use a Kaplan-Meier approximation for multiple decrements to find how many of these individuals can be expected to survive to survive to past the end of the fourth decade.

Solution S4C55-3. In Solution S4C55-2, we identified the probability of death due to explosion alone during each decade. Since the data are not censored or truncated for the time period of observation, the Kaplan-Meier estimate is the same as the empirical survival function. Thus, the number of surviving individuals at the end of decade 4 is $10000(1-q_1^{(2)})(1-q_2^{(2)})(1-q_3^{(2)})(1-q_4^{(2)})$ =

10000(1-0.206185567)(1-0.0103092784)(1-0.1237113402)(1-0.3402061856) = 4542.280198 or approximately **4542 individuals.**

Problem S4C55-4. There are 3 unique unmodified observations in a sample: 15, 65, and 67.

In this sample, the following is true:

19 observations have been left truncated at 4.
30 observations have been left truncated at 34.
26 observations have been left truncated at 66.
21 observations have been left truncated at 84.
4 observations have been right censored at 4.
3 observations have been right censored at 34.
4 observations have been right censored at 66.
11 observations have been right censored at 84.

No observations were both truncated and censored, and all truncated and censored observations have been affected by the truncation or censoring (i.e., the original value of the observation was not preserved).

There is exactly one observation for each of the following values: 15, 65, and 67. None of these observations has been modified at all.

The data are split into four intervals: (0, 15], (15, 65], (65, 67], and (67, ∞). Find the size of the risk set associated with the interval (15, 65].

Solution S4C55-4. The boundaries of the intervals are $c_0 = 0$, $c_1 = 15$, $c_2 = 65$, and $c_3 = 67$.

We use the formula $r_0 = d_0$; $r_j = {}_{i=0}{}^{j}\Sigma(d_i) - {}_{i=0}{}^{j-1}\Sigma(u_i + x_i)$ for j = 1, 2, ..., k.

For the interval (0, 15], $r_0 = d_0$ = the number of observations left-truncated within this interval or not at all (i.e., $d_i = 0$). These observations consist of the following:

1. 19 observations that have been left truncated at 4.
2. *All* 22 of the right censored observations (which are indicated as not having been left truncated).
3. The three unique observations.

Thus, $d_0 = 19 + 22 + 3 = r_0 = 44$ for the interval (0, 15].

Now we can find $r_1 = {}_{i=0}{}^{1}\Sigma(d_i) - {}_{i=0}{}^{1-1}\Sigma(u_i + x_i) = d_0 + d_1 - (u_0 + x_0)$.

We find the value of x_0: This value includes the number of unmodified observations in the interval (0, 15]: 1 - and the number of left truncated but not right censored observations in the interval (0, 15]: 19.

Thus, $x_0 = 1 + 19 = 20$.

The value u_0 is equal to 4, the number of observations that have been right censored at 4, the only location within (0, 15] at which right censoring occurs.
Now we find d_1 = the number of observations left truncated within (15, 65]. The only value at which such truncation occurs is 34, at which 30 observations are truncated. Thus, $d_1 = 30$.

Hence, $r_1 = 44 + 30 - (4 + 20) = \mathbf{r_1 = 50}$. This is the size of the risk set associated with the interval (15, 65].

Problem S4C55-5. There are 3 unique unmodified observations in a sample: 15, 65, and 67.

In this sample, the following is true:

19 observations have been left truncated at 4.
30 observations have been left truncated at 34.
26 observations have been left truncated at 66.
21 observations have been left truncated at 84.
4 observations have been right censored at 4.
3 observations have been right censored at 34.
4 observations have been right censored at 66.
11 observations have been right censored at 84.

No observations were both truncated and censored, and all truncated and censored observations have been affected by the truncation or censoring (i.e., the original value of the observation was not preserved).

There is exactly one observation for each of the following values: 15, 65, and 67. None of these observations has been modified at all.

The data are split into four intervals: (0, 15], (15, 65], (65, 67], and (67, ∞). Use the Kaplan-Meier approximation for large data sets to find $\hat{S}(65)$, the estimate of the survival function at 65.

Solution S4C55-5. We use the formula $\hat{S}(c_0) = 1$; $\hat{S}(c_j) = {}_{i=0}{}^{j-1}\Pi(1- x_i/r_i)$ for j = 1, 2, ..., k.
Here, $c_2 = 65$, so $\hat{S}(65) = (1- x_0/r_0)(1- x_1/r_1)$.
From Solution S4C55-4, we know that $r_0 = 44$, $r_1 = 50$, and $x_0 = 20$.
Now we find x_1:
This value includes the number of unmodified observations in the interval (15, 65]: 1 - and the number of left truncated but not right censored observations in the interval (15, 65]: 30.
Thus, $x_1 = 1 + 30 = 31$.
Thus, $\hat{S}(65) = (1- 20/44)(1- 31/50) = \mathbf{\hat{S}(65) = 0.2072727273 = 57/275}$.

Section 56

The Method of Moments and Maximum Likelihood Estimation

The **method of moments** is one way of estimating the parameters of a distribution by reference to sample data.

In Section 15.1 of *Loss Models* (cited below), the method of moments is defined as follows:

"A **method-of-moments estimate** of θ is any solution of the p equations $\mu_k'(\theta) = \mu\hat{}_k'$, for k = 1, 2, ..., p."

Let M be the random variable representing the values in our *sample* and X be the random variable representing the values given by the *distribution* we are considering. According to the method of moments, the system of equations that can be set up to solve for our p unknown parameters can be conceived of as follows:

$E[X] = E[M]$.

$E[X^2] = E[M^2]$

...

$E[X^p] = E[M^p]$

This is a more convenient way to remember the method of moments and is especially useful when one has memorized the formulas for the moments of the distribution with which one is working.

The following definition of the **likelihood function** is offered by Larsen and Marx.

"Let $k_1, k_2, \ldots, k_n$ be a random sample of size n from the discrete p. d. f. $p_X(k; \theta)$, where θ is an unknown parameter. The *likelihood function,* L(θ), is the product of the p. d. f. evaluated at the n k_is. That is, **$L(\theta) = {}_{i=1}{}^{n}\Pi\, p_X(k_i \mid \theta)$**."

If the p. d. f. in question is equal to $f_Y(y_i; \theta)$ and is continuous for random sample $y_1, y_2, \ldots, y_n$ of size n, then **$L(\theta) = {}_{i=1}{}^{n}\Pi\, f_Y(y_i \mid \theta)$.**

The **loglikelihood function** is **$l(\theta) = \ln(L(\theta))$.** It is sometimes more convenient to convert a likelihood function into a loglikelihood function.

Now we can define the **maximum likelihood estimate,** following Larsen and Marx.

"Let $L(\theta) = {}_{i=1}{}^{n}\Pi\, p_X(k_i \mid \theta)$and $L(\theta) = {}_{i=1}{}^{n}\Pi\, f_Y(y_i \mid \theta)$ be the likelihood functions corresponding to random samples $k_1, k_2,..., k_n$ and $y_1, y_2,..., y_n$ drawn from the discrete p.d.f. $p_X(k_i \mid \theta)$ and the continuous p.d.f. $f_Y(f_i \mid \theta)$, respectively, where θ is an unknown parameter. In each case, let θ_e be a value of the parameter such that $L(\theta_e) \geq L(\theta)$ for all possible values of θ. Then θ_e is called a **maximum likelihood estimate** for θ."

In other words, *a maximum likelihood* estimate is a vector of parameters such that the likelihood function is maximized. Maximizing the loglikelihood function is equivalent to maximizing the likelihood function.

If some of the data in the sample has been censored, then each observation in the sample censored at u contributes a factor of $S(u) = 1 - F(u)$ to the likelihood function.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source:

An Introduction to Mathematical Statistics and Its Applications. (Fourth Edition), Larsen, Richard J. and Morris L. Marx. Pearson Prentice Hall: 2006. pp. 241-243, 347.

Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 15, pp. 376, 381-384.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C56-1. Similar to Question 4 of the Exam C Sample Questions from the Society of Actuaries. You know that ages in a certain population follow a single-parameter Pareto distribution with $f(x) = \alpha/x^{\alpha+1}$. You are analyzing a sample of six observations, of which four observations are 19, 23, 75, and 95, and the other two observations have been right censored at 100. Use the method of maximum likelihood estimation to estimate the parameter α.

Solution S4C56-1. We first create our likelihood function. The first four factors in the function each correspond to $\alpha/x^{\alpha+1}$, with the uncensored values substituted for x.

Thus, the product of the first four factors is $(\alpha/19^{\alpha+1})(\alpha/23^{\alpha+1})(\alpha/75^{\alpha+1})(\alpha/95^{\alpha+1})$.

The last two observations are censored at 100, so their contribution to the likelihood function will be a factor of $S(100)^2$. We first find the cdf $F(x) = \int\alpha/x^{\alpha+1} = 1\text{-}x^{-\alpha}$ (the constant of 1 is necessary to make the cdf equal 0 at x = 0). Then $S(x) = 1 - F(x) = x^{-\alpha}$, and so $S(100) = 100^{-\alpha}$, and thus the likelihood function is $L(\alpha) = (\alpha/19^{\alpha+1})(\alpha/23^{\alpha+1})(\alpha/75^{\alpha+1})(\alpha/95^{\alpha+1})(100^{-\alpha})(100^{-\alpha}) =$

$(\alpha^4/(3113625*31136250000^{\alpha}))$.

Now we find $L'(\alpha) = (4\alpha^3/(3113625*31136250000^{\alpha})) +$

$(-\ln(31136250000)\alpha^4/(3113625*31136250000^{\alpha}))$.

To maximize $L(\alpha)$, we set $L'(\alpha) = 0$, which implies that

$4\alpha^3/(3113625*31136250000^{\alpha}) = \ln(31136250000)\alpha^4/(3113625*31136250000^{\alpha})$ and

$\ln(31136250000)\alpha = 4$. Thus, $4/\ln(31136250000) =$ **α = 0.1655516859**.

Problem S4C56-2. Similar to Question 6 of the Exam C Sample Questions from the Society of Actuaries. For a sample of 100 pieces of data, you know the following: $\Sigma(x_i) = 97000$; $\Sigma(x_i^2) = 890340000$. A lognormal distribution with parameters μ and σ is used to fit the data. Estimate μ and σ using the method of moments.

Relevant properties for lognormal distributions:

$E(X^k) = \exp(k\mu + k^2\sigma^2/2)$.

Solution S4C56-2. We use two moments, since there are two parameters to estimate:

$E[X] = E[M]$
$E[X^2] = E[M^2]$

We are given that there are 100 pieces of data and that their sum is 97000. Thus, E(X) = 97000/100 = 970.

Likewise, $E(X^2) = 890340000/100 = 8903400$.

Thus, we set up the following system of equations:
(i) $E(X) = 970 = \exp(\mu + \sigma^2/2)$.

(ii) $E(X^2) = 89034000000 = \exp(2\mu + 2\sigma^2)$.

We take the natural logarithm of each side of each equation:

(i) $6.877296071 = \mu + \sigma^2/2$;

(ii) $16.00194378 = 2\mu + 2\sigma^2$;

To get σ^2, we can take (ii) - 2*(i): 16.00194378-2*6.877296071 = σ^2 = 2.247351641 → **σ = 1.499116954**. Moreover, μ = 6.877296071 - 2.247351641/2 = **μ = 5.753620251**.

Problem S4C56-3. Similar to Question 6 of the Exam C Sample Questions from the Society of Actuaries. For a sample of 100 pieces of data, you know the following: $\Sigma(x_i)$ = 97000; $\Sigma(x_i^2)$ = 890340000. A lognormal distribution with parameters μ and σ is used to fit the data. Find $E(X \Lambda 600)$ for the fitted distribution.

Relevant properties for lognormal distributions:

$E(X^k) = \exp(k\mu + k^2\sigma^2/2)$.

$E((X \Lambda x)^k) = \exp(k\mu + k^2\sigma^2/2)*\Phi((\ln(x) - \mu - k\sigma^2)/\sigma) + x^k(1-F(x))$, where $F(x) = \Phi((\ln(x)- \mu)/\sigma)$.

Solution S4C56-3. From Solution S4C56-2, we know that μ = 5.753620251 and σ = 1.499116954. We also know that $\exp(k\mu + k^2\sigma^2/2) = E(X) = 970$. Now we find

$\Phi((\ln(x) - \mu - k\sigma^2)/\sigma)$ for k = 1 and x = 600:

$\Phi((\ln(x) - \mu - k\sigma^2)/\sigma) = \Phi((\ln(600) - 5.753620251 - 2.247351641)/1.499116954) =$

$\Phi(-1.069991392)$, which we can find via the Excel input "=NORMSDIST(-1.069991392)". The result is 0.142311592.

Now we find $\Phi((\ln(x)- \mu)/\sigma) = \Phi((\ln(600)- 5.753620251)/1.499116954) = \Phi(0.4291255613)$, which we can find via the Excel input "=NORMSDIST(0.4291255613)". The result is 0.666084074. Thus, our answer is 970*0.142311592 + 600*(1-0.666084074) = **E(X Λ 600) = 338.3917998**.

Problem S4C56-4. Similar to Question 49 of the Exam C Sample Questions from the Society of Actuaries. You are analyzing population data with the following probabilities:

Pr(X = 2) = 0.23
Pr(X = 5) = 0.34
Pr(X = 7) = 0.43.

A sample of this population is taken. A method of moments procedure is used to estimate the mean μ via the sample mean X^- and the population variance σ^2 by the statistic $S_n^2 = \Sigma(X_i-X^-)^2/n$, where n is the sample size. Find the bias of S_n^2 when n = 13. **Hint:** $\Sigma(X_i-X^-)^2 = (n-1)\sigma^2$.

Solution S4C56-4. We first find μ = 0.23(2) + 0.34(5) + 0.43(7) = μ = 5.17.
Now we find $\sigma^2 = 0.23(2-5.17)^2 + 0.34(5-5.17)^2 + 0.43(7-5.17)^2 = \sigma^2 = 3.7611$.
Now $S_n^2 = \Sigma(X_i-X^-)^2/n = (n-1)\sigma^2/n$, so for n = 13, $S_{13}^2 = 12\sigma^2/13 = 3.471784615$.
Thus, the bias of S_n^2 is $S_n^2 - \sigma^2$ = 3.471784615 - 3.7611 = **-0.2893153846**.

Problem S4C56-5. Similar to Question 75 of the Exam C Sample Questions from the Society of Actuaries. A sample of 50 values is given, such that $\Sigma(x_i) = 8200$ and $\Sigma(x_i^2) = 2000000$. You are modeling the data via a shifted exponential distribution such that

$f(x) = (1/\theta)e^{-(x-\delta)/\theta}$ for $\delta < x$. Use a method of moments procedure to estimate the value of δ.

Solution S4C56-5. We use two moments, since there are two parameters to estimate:

$E[X] = E[M]$
$E[X^2] = E[M^2]$

We are given that there are 50 pieces of data and that their sum is 8200. Thus, $E(X) = 8200/50 = 164$. Likewise, $E(X^2) = \Sigma(x_i^2)/50 = 40000$.

What does the mean of a shifted exponential distribution equal? The distribution, other than the shift, is identical to a regular exponential distribution, so the mean must be θ increased by the size of the shift, or $\theta + \delta$. The variance of a shifted exponential distribution is unaffected by the shift and so should remain as θ^2. This means that $E(X^2) = Var(X) + E(X)^2 = \theta^2 + (\theta + \delta)^2$.

Thus, we set up the following system of equations:
(i) $164 = \theta + \delta$;

(ii) $40000 = \theta^2 + (\theta + \delta)^2$.

We know that $(\theta + \delta)^2 = 164^2 = 26896$, so $\theta^2 = 40000 - 26896 = 13104$, and thus $\theta = 114.4727042$. Hence, $\delta = 164 - \theta =$ **$\delta = 49.52729583$**.

Section 57

Exam-Style Questions on Maximum Likelihood Estimation

When the data is grouped into intervals of the form $(c_{j-1}, c_j]$, where $c_0 < c_1 < ... < c_k$ are the boundaries of the intervals, and there are n_j observations in each interval $(c_{j-1}, c_j]$, then the likelihood function is $\mathbf{L(\theta) = {}_{j=1}{}^{k}\Pi(F(c_j | \theta) - F(c_{j-1} | \theta))^{n\_j}}$. The corresponding loglikelihood function is $\mathbf{l(\theta) = {}_{j=1}{}^{k}\Sigma(n_j * \ln(F(c_j | \theta) - F(c_{j-1} | \theta)))}$.

If the data are truncated, they can be shifted by subtracting the point of truncation from each observation. Alternatively, one can simplify the data by assuming a lack of information for values below the truncation point. In that case, the denominator of each factor of the likelihood function becomes 1-F(d), where d is the truncation point for the observation in question.

The following properties are useful:

Theorem 57.1.

For *any* exponential distribution and *any* sample of *any* size n, both the maximum likelihood estimate and the estimate using the method of moments will be equal to the sample mean.

Theorem 57.2.

The maximum likelihood estimator is **asymptotically unbiased**: its bias approaches zero as the sample size increases without bound.

The variance of the maximum likelihood estimator is extremely difficult to calculate directly, but it can in some cases be approximated. The approximation is itself a difficult process, and so we will approach it in a piecewise fashion here.

The concept of **information** associated with a maximum likelihood estimator of a parameter θ is important in the approximation.

The following is the relationship between **information** and **asymptotic variance** of a maximum likelihood estimator. This relationship holds when there is a single parameter θ and when information is a function of the sample size.

Information = 1/Asymptotic variance = 1/Var(θ^)

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source:
Exam C Sample Questions and Solutions from the Society of Actuaries.

"Maximum Likelihood." Wikipedia, the Free Encyclopedia.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 15, pp. 384-388.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C57-1. Similar to Question 14 of the Exam C Sample Questions from the Society of Actuaries. The informationassociated with a maximum likelihood estimator of a parameter θ is 14n, where n is the number of observations. Find the asymptotic variance of the maximum likelihood estimator of 95θ.

Solution S4C57-1. We use the formula Information = 1/Var(θ^) → Var(θ^) = 1/Information → Var(θ^) = 1/(14n). We want to find Var(95*θ^) = 95^2*Var(θ^) = 9025/(14n) =

Var(95*θ^) = 644.6428571/n.

Problem S4C57-2. Similar to Question 26 of the Exam C Sample Questions from the Society of Actuaries. There are three types of slugs: green, purple, and striped. You are aware that the lifetimes of green slugs follow an exponential distribution with parameter θ. The lifetimes of purple slugs follow an exponential distribution with parameter 4θ. The lifetimes of striped slugs follow an exponential distribution with parameter 16θ. You observe the following sample:

1 green slug with a lifetime of 39.

2 purple slugs with lifetimes of 45 and 165.

1 striped slug with a lifetime of 900.

Use the method of maximum likelihood estimation to find θ.

Solution S4C57-2. We use Theorem 57.1: For *any* exponential distribution and *any* sample of *any* size n, both the maximum likelihood estimate and the estimate using the method of moments will be equal to the sample mean.

However, we must be careful here, because we are trying to find the sample mean as pertains to θ. For purple slugs, the parameter is 4θ, so we must divide each observation by 4 to find the observation relevant to calculating the sample mean for θ. For striped slugs, the parameter is 16θ, so we must divide each observation by 16 to find the observation relevant to calculating the sample mean for θ.

Thus, the sample of which we will be taking the mean is 39, 45/4 = 11.25, 165/4 = 41.25, and 900/16 = 56.25. Hence, θ = (39+11.25+41.25+56.25)/4 = **θ = 36.9375**.

Problem S4C57-3. Similar to Question 34 of the Exam C Sample Questions from the Society of Actuaries. A negative binomial distribution has a value of r = 3. For a certain sample size n, the sample mean is 15. Find the value of β for this distribution using maximum likelihood estimation.

Relevant properties of negative binomial distributions: $p(k) = (r(r-1)...(r-k+1)/k!)\beta^k/(1+\beta)^{r+k}$.

Solution S4C57-3. We first find $L(\theta) = {}_{j=1}{}^{n}\Pi((r(r-1)...(r-k_j+1)/k_j!)\beta^{k\_j}/(1+\beta)^{k\_j}) =$

${}_{j=1}{}^{n}\Pi((3(3-1)...(3-k_j+1)/k_j!)\beta^{k\_j}/(1+\beta)^{3+k\_j})$.

We notice that this function is proportional to another function, which we call $M(\theta) = {}_{j=1}{}^{n}\Pi(\beta^{k\_j}/(1+\beta)^{3+k\_j})$.

We can take the logarithm of M(θ):

${}_{j=1}{}^{n}\Sigma(\ln(\beta^{k\_j}/(1+\beta)^{3+k\_j})) =$

${}_{j=1}{}^{n}\Sigma(\ln(\beta^{k\_j}) - \ln((1+\beta)^{3+k\_j})) =$

${}_{j=1}{}^{n}\Sigma(k\_j*\ln(\beta) - (3+k\_j)\ln(1+\beta))$.

To maximize this function, we can take its derivative and set it equal to zero:
$(\ln(M(\theta)))' = {}_{j=1}{}^{n}\Sigma(k\_j/\beta- (3+k\_j)/(1+\beta)) = 0 \rightarrow$

${}_{j=1}{}^{n}\Sigma(k\_j/\beta) = {}_{j=1}{}^{n}\Sigma((3+k\_j)/(1+\beta)) \rightarrow$

${}_{j=1}{}^{n}\Sigma(k\_j*(1+\beta)) = {}_{j=1}{}^{n}\Sigma((3+k\_j)* \beta) \rightarrow$

${}_{j=1}{}^{n}\Sigma(k\_j*(1+\beta) - (3+k\_j)* \beta) = 0 \rightarrow$

${}_{j=1}{}^{n}\Sigma(k\_j* - 3* \beta) = 0 \rightarrow$

${}_{j=1}{}^{n}\Sigma(k\_j) - 3n\beta = 0 \rightarrow$

${}_{j=1}{}^{n}\Sigma(k\_j) = 3n\beta \rightarrow$

${}_{j=1}{}^{n}\Sigma(k\_j)/n = 3\beta$.

However, ${}_{j=1}{}^{n}\Sigma(k\_j)/n$ is the sample mean or 15, so 15 = 3β → **β = 5**.

Problem S4C57-4. Similar to Question 37 of the Exam C Sample Questions from the Society of Actuaries. For a Pareto distribution, it is known that the parameter $\theta = 30$. The parameter α is estimated via the method of maximum likelihood by analyzing data from the following sample: 13, 25, 36, 40. Find the maximum likelihood estimate of α.

Relevant properties of Pareto distributions: $f(x) = \alpha\theta^\alpha/(x+\theta)^{\alpha+1}$

Solution S4C57-4. We first find $L(\alpha) =$
$(\alpha\theta^\alpha/(13+\theta)^{\alpha+1})(\alpha\theta^\alpha/(25+\theta)^{\alpha+1})(\alpha\theta^\alpha/(36+\theta)^{\alpha+1})(\alpha\theta^\alpha/(40+\theta)^{\alpha+1}) =$

$\alpha^4*(30^4)^\alpha/(43*55*66*70)^{\alpha+1} = \alpha^4*(30^4)^\alpha/(10926300)^{\alpha+1} =$

$\alpha^4*0.0741330551^\alpha/10926300$.
Now we find $L'(\alpha)$ and set it equal to zero:

$L'(\alpha) = 4\alpha^3*0.0741330551^\alpha/10926300 + \ln(0.0741330551)*\alpha^4*0.0741330551^\alpha/10926300 = 0$

$-\ln(0.0741330551)*\alpha^4*0.0741330551^\alpha/10926300 = 4\alpha^3*0.0741330551^\alpha/10926300 \rightarrow$

$-\ln(0.0741330551)*\alpha = 4 \rightarrow \alpha = 4/-\ln(0.0741330551) =$ $\boldsymbol{\alpha = 1.537341787}$.

Problem S4C57-5. Similar to Question 44 of the Exam C Sample Questions from the Society of Actuaries. You are trying to estimate the parameter θ of an exponential distribution. You know the following: 10 values fall within the interval from 0 to 25, inclusive. 34 values fall within the interval (25,50]. 9 values are greater than 50. Find the maximum likelihood estimate of θ.

Solution S4C57-5. This is a case where data are grouped into intervals, so $L(\theta) = {}_{j=1}{}^{k}\Pi(F(c_j \mid \theta) - F(c_{j-1} \mid \theta))^{n\_j}$. Therefore, we consider the cumulative distribution function at each endpoint of each interval:
$F(0 \mid \theta) = 0$; $F(25 \mid \theta) = 1 - e^{-25/\theta}$; $F(50 \mid \theta) = 1 - e^{-50/\theta}$; $F(\infty \mid \theta) = 1$.

Thus, our likelihood function will have 10 factors of $(F(25 \mid \theta)- F(0 \mid \theta)) = (1 - e^{-25/\theta})$.
Our likelihood function will have 34 factors of $(F(50 \mid \theta)- F(25 \mid \theta)) = (e^{-25/\theta}- e^{-50/\theta})$.
Our likelihood function will have 9 factors of $(F(\infty \mid \theta)- F(50 \mid \theta)) = (e^{-50/\theta})$.

Hence, $L(\theta) = (1 - e^{-25/\theta})^{10}(e^{-25/\theta}- e^{-50/\theta})^{34}(e^{-50/\theta})^9$.

It is cumbersome to directly maximize this function in terms of θ. However, maximization will be easier if we let $e^{-25/\theta} = p$. Then, $L(p) = (1-p)^{10}(p-p^2)^{34}(p^2)^9 = (p^{52})(1-p)^{44}$.

Then $\ln(L(p)) = l(p) = \ln((p^{52})(1-p)^{44}) = \ln(p^{52}) + \ln((1-p)^{44}) = 52\ln(p) + 44\ln(1-p)$.

Thus, $l'(p) = 52/p - 44/(1-p) = 0$ at its maximum, so $52/p = 44/(1-p)$ and $52(1-p) = 44p \rightarrow 52 = 96p$ and $p = 52/96$. Thus, $e^{-25/\theta} = 52/96 \rightarrow -25/\ln(52/96) =$ $\boldsymbol{\theta = 40.77608484}$.

Section 58

Fisher Information for One Parameter

Fisher's information or simply **information** is defined as follows for independent and identically distributed random variables:

$I(\theta) = nE((\partial(\ln(f(X; \theta)/\partial\theta)^2)$

$I(\theta) = n*\int f(x; \theta)(\partial(\ln(f(x; \theta)/\partial\theta)^2)dx$

More generally, if only one parameter θ is being estimated via the method of maximum likelihood:

$I(\theta) = E((\partial(l(\theta)/\partial\theta)^2)$.

$I(\theta) = -E((\partial^2(l(\theta)/\partial\theta^2))$.

This means that Fisher's information is the expected value of the square of the partial derivative with respect to θ of the loglikelihood function as well as the negative of the expected value of the second partial derivative with respect to θ of the loglikelihood function.

The inverse of Fisher's information is a good estimate for the asymptotic variance of the maximum likelihood estimator: $1/I(\theta) = Var(\theta\hat{}_n)$ for sample size n.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 15, pp. 393-394.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C58-1. For a Pareto distribution, it is known that the parameter θ = 30. The parameter α is estimated via the method of maximum likelihood by analyzing data from the following sample: 13, 25, 36, 40. Find $(\partial(l(\alpha)/\partial\alpha)^2$, the square of the first partial derivative of the loglikelihood function.

Relevant properties of Pareto distributions: $f(x) = \alpha\theta^\alpha/(x+\theta)^{\alpha+1}$

Solution S4C58-1. We first find $L(\alpha) =$
$(\alpha\theta^\alpha/(13+\theta)^{\alpha+1})(\alpha\theta^\alpha/(25+\theta)^{\alpha+1})(\alpha\theta^\alpha/(36+\theta)^{\alpha+1})(\alpha\theta^\alpha/(40+\theta)^{\alpha+1}) =$
$\alpha^4*(30^4)^\alpha/(43*55*66*70)^{\alpha+1} = \alpha^4*(30^4)^\alpha/(10926300)^{\alpha+1} =$
$\alpha^4*0.0741330551^\alpha/10926300$.
Now we find $l(\alpha) = \ln(L(\alpha)) = \ln(\alpha^4*0.0741330551^\alpha/10926300) =$
$(\ln(\alpha^4)) + \ln(0.0741330551^\alpha) - \ln(10926300) =$
$l(\alpha) = 4*\ln(\alpha) + \alpha*\ln(0.0741330551) - \ln(10926300)$.
$\partial(l(\alpha)/\partial\alpha = 4/\alpha + \ln(0.0741330551)$
$(\partial(l(\alpha)/\partial\alpha)^2 = 16/\alpha^2 + (8/\alpha)\ln(0.0741330551) + \ln(0.0741330551)^2$.
$\boldsymbol{(\partial(l(\alpha)/\partial\alpha)^2 = 16/\alpha^2 - 20.851515007/\alpha + 6.76985113}$.

Problem S4C58-2. For a Pareto distribution, it is known that the parameter $\theta = 30$. The parameter α is estimated via the method of maximum likelihood by analyzing data from the following sample: 13, 25, 36, 40. Find $(\partial^2(l(\alpha)/\partial\alpha^2))$, the second partial derivative of the loglikelihood function.

Solution S4C58-2. We know from Solution S4C58-1 that $\partial(l(\alpha)/\partial\alpha = l'(\alpha) = 4/\alpha + \ln(0.0741330551)$.

Thus, $l''(\alpha) = \boldsymbol{\partial^2(l(\alpha)/\partial\alpha^2) = -4/\alpha^2}$.

Problem S4C58-3. For a Pareto distribution, it is known that the parameter $\theta = 30$. The parameter α is estimated via the method of maximum likelihood by analyzing data from the following sample: 13, 25, 36, 40. Find the Fisher information associated with the maximum likelihood estimator.

Relevant properties of Pareto distributions: $f(x) = \alpha\theta^\alpha/(x+\theta)^{\alpha+1}$

Solution S4C58-3. We can use the formula $I(\alpha) = E((\partial(l(\alpha)/\partial\alpha)^2)$ or $I(\alpha) = -E((\partial^2(l(\alpha)/\partial\alpha^2))$. In this case, the second formula is easier because $\partial^2(l(\alpha)/\partial\alpha^2) = -4/\alpha^2$. Thus, $I(\alpha) = -E(-4/\alpha^2) = E(4/\alpha^2)$.

We have enough information to find the maximum likelihood estimate of α, which we would be able to substitute in the expression $4/\alpha^2$. Since the maximum likelihood estimate of α is a single value α^*, we know that $-E((\partial^2(l(\alpha)/\partial\alpha^2))$ is equal to $-\partial^2(l(\alpha^*)/\partial\alpha^2 = 4/\alpha^{*2}$.

Fortunately we already know that $l'(\alpha) = \partial(l(\alpha)/\partial\alpha = 4/\alpha + \ln(0.0741330551)$.

To find α^*, we set $l'(\alpha) = 0 = 4/\alpha^* + \ln(0.0741330551) \rightarrow \alpha^* = 4/-\ln(0.0741330551) = \alpha^* = 1.537341787$.

Thus, $I(\alpha) = 4/1.537341787^2 = \boldsymbol{I(\alpha) = 1.692462782}$.

Problem S4C58-4. For a Pareto distribution, it is known that the parameter θ = 30. The parameter α is estimated via the method of maximum likelihood by analyzing data from the following sample: 13, 25, 36, 40. Estimate the asymptotic variance of the maximum likelihood estimator.

Solution S4C58-4. We use the formula $1/I(\alpha) = Var(\alpha\hat{}_n) = 1/1.692462782 =$ **$Var(\alpha\hat{}_4) =$ 0.5908549425**.

Problem S4C58-5. Review of Section 56: Similar to Question 239 of the Exam C Sample Questions from the Society of Actuaries. You have the following data for numbers of flies swatted by a sample of cows:

29 cows swatted 0 flies.
23 cows swatted 1 fly.
34 cows swatted 2 flies.
34 cows swatted 3 flies.

This data is modeled in two ways:
1. A geometric distribution is fitted to the data, where the parameter β is estimated by the method of maximum likelihood. Using this distribution, let A be the probability that a cow swats 0 flies.

2. A Poisson distribution is fitted to the data, where the parameter λ is estimated using the method of moments. Using this distribution, let B be the probability that a cow swats 0 flies. Find │B-A│.

Relevant properties of Poisson distributions: $Pr(n = 0) = e^{-\lambda}$.

Relevant properties of geometric distributions: $Pr(n = 0) = 1/(1+\beta)$

$Pr(n = k) = \beta^k/(1+\beta)^{k+1}$.

Solution S4C58-5. We first estimate β by finding L(β) =
$Pr(n = 0)^{29} * Pr(n = 1)^{23} * Pr(n = 2)^{34} * Pr(n = 3)^{34} =$
$(1/(1+\beta))^{29} * (\beta/(1+\beta)^2)^{23} * (\beta^2/(1+\beta)^3)^{34} * (\beta^3/(1+\beta)^4)^{34} =$
$L(\beta) = \beta^{193}/(1+\beta)^{313}$
$l(\beta) = \ln(\beta^{193}/(1+\beta)^{313}) = 193*\ln(\beta) - 313*\ln(1+\beta)$
$l'(\beta) = 193/\beta - 313/(1+\beta) = 0$ at the maximum.
Thus, $193/\beta = 313/(1+\beta)$ and $193(1+\beta) = 313\beta \rightarrow 193 = 120\beta \rightarrow \beta = 193/120 = 1.6083333333$.
Thus, $A = 1/(1+\beta) = 1/2.6083333333 = A = 0.3833865815$.

Now we estimate λ via the method of moments. Fortunately, for a Poisson distribution, λ = E(X), so we just need to find the empirical expected value:
$\lambda = (0*29 + 1*23 + 2*34 + 3*34)/(29+23+34+34) = \lambda = 1.6083333333$.

Thus $B = e^{-1.6083333333} = B = 0.2002210379$ and │B-A│=│0.2002210379-0.3833865815│= **0.1831655436**.

Section 59

Fisher Information for Multiple Parameters

If there is more than one parameter being estimated via the method of maximum likelihood (call the parameters $\theta_1, ..., \theta_n$), then Fisher information can be expressed as a matrix $I(\theta)$, where each entry $I(\theta)_{r,s}$ of the matrix in row r, column s is expressible as follows:

$I(\theta)_{r,s} = -E((\partial^2(l(\theta)/\partial\theta_r\partial\theta_s))$.

$I(\theta)_{r,s} = E((\partial(l(\theta)/\partial\theta_r)(\partial(l(\theta)/\partial\theta_s))$

The estimate of the **covariance matrix** of the Fisher information is the inverse of $I(\theta)$:
Covariance matrix = $I^{-1}(\theta)$.

The entry in row r, column s of the covariance matrix corresponds to the covariance of θ_r with θ_s.

Note that $Cov(\theta_r, \theta_r) = Var(\theta_r)$.

Sometimes it is not possible to take the expected value of either the second derivatives or the product of the first derivatives of the loglikelihood function. In that case, one can instead substitute observed data points where appropriate. This method will result in the **observed information.**

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 15, pp. 395-397.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C59-1. For a Pareto distribution, the parameters α and θ are both estimated via the method of maximum likelihood by analyzing data from the following sample: 13, 25, 36, 40. Find the loglikelihood function for this distribution.

Relevant properties of Pareto distributions: $f(x) = \alpha\theta^\alpha/(x+\theta)^{\alpha+1}$

Solution S4C59-1. We first find $L(\alpha, \theta) = (\alpha\theta^\alpha/(13+\theta)^{\alpha+1})(\alpha\theta^\alpha/(25+\theta)^{\alpha+1})(\alpha\theta^\alpha/(36+\theta)^{\alpha+1})(\alpha\theta^\alpha/(40+\theta)^{\alpha+1}) =$

$L(\alpha, \theta) = \alpha^4 * \theta^{4\alpha}/((13+\theta)(25+\theta)(36+\theta)(40+\theta))^{\alpha+1}$.

The loglikelihood function is $\ln(L(\alpha, \theta)) = l(\alpha, \theta) =$

$\ln(\alpha^4 * \theta^{4\alpha}/((13+\theta)(25+\theta)(36+\theta)(40+\theta))^{\alpha+1}) =$

$\mathbf{l(\alpha, \theta) = 4*\ln(\alpha) + 4\alpha*\ln(\theta) - (\alpha+1)*\ln(13+\theta) - (\alpha+1)*\ln(25+\theta) - (\alpha+1)*\ln(36+\theta) - (\alpha+1)*\ln(40+\theta)}$.

Problem S4C59-2. For a Pareto distribution, the parameters α and θ are both estimated via the method of maximum likelihood by analyzing data from the following sample: 13, 25, 36, 40. Find the following second partial derivatives of the loglikelihood function for this distribution:

$\partial^2(l(\alpha, \theta)/(\partial\theta\partial\alpha))$

$\partial^2(l(\alpha, \theta)/(\partial\theta^2))$

$\partial^2(l(\alpha, \theta)/(\partial\alpha^2))$

Relevant properties of Pareto distributions: $f(x) = \alpha\theta^\alpha/(x+\theta)^{\alpha+1}$

Solution S4C59-2. We begin with $l(\alpha, \theta) = 4*\ln(\alpha) + 4\alpha*\ln(\theta) - (\alpha+1)*\ln(13+\theta) - (\alpha+1)*\ln(25+\theta) - (\alpha+1)*\ln(36+\theta) - (\alpha+1)*\ln(40+\theta)$.

The first partial derivative with respect to α is

$\partial(l(\alpha, \theta)/\partial\alpha) = 4/\alpha + 4*\ln(\theta) - \ln(13+\theta) - \ln(25+\theta) - \ln(36+\theta) - \ln(40+\theta)$.

The second partial derivative with respect to α is thus

$\mathbf{\partial^2(l(\alpha, \theta)/(\partial\alpha^2)) = -4/\alpha^2}$.

The partial derivative $\partial^2(l(\alpha, \theta)/(\partial\theta\partial\alpha))$ is the partial derivative of $\partial(l(\alpha, \theta)/\partial\alpha)$ with respect to θ:

$\mathbf{\partial^2(l(\alpha, \theta)/(\partial\theta\partial\alpha)) = 4/\theta - 1/(13+\theta) - 1/(25+\theta) - 1/(36+\theta) - 1/(40+\theta)}$.

The first partial derivative with respect to θ is

$\partial(l(\alpha, \theta)/\partial\theta) = 4\alpha/\theta - (\alpha+1)/(13+\theta) - (\alpha+1)/(25+\theta) - (\alpha+1)/(36+\theta) - (\alpha+1)/(40+\theta)$.

The second partial derivative with respect to θ is thus

$\mathbf{\partial^2(l(\alpha, \theta)/(\partial\theta^2)) = -4\alpha/\theta^2 + (\alpha+1)/(13+\theta)^2 + (\alpha+1)/(25+\theta)^2 + (\alpha+1)/(36+\theta)^2 + (\alpha+1)/(40+\theta)^2}$.

Problem S4C59-3. For a Pareto distribution, the parameters α and θ are both estimated via the method of maximum likelihood by analyzing data from the following sample: 13, 25, 36, 40.

A possible combination of maximum likelihood estimates for α and θ is θ = 30 and α = 1.537341787. Use these values to find the *observed* information matrix I(α, θ).

Solution S4C59-3. The observed information matrix will be of the following form, where α* and θ* are the provided maximum likelihood estimates:

$[-\partial^2(l(\alpha^*, \theta^*)/(\partial\alpha^2)), -\partial^2(l(\alpha^*, \theta^*)/(\partial\theta\partial\alpha))]$

$[-\partial^2(l(\alpha^*, \theta^*)/(\partial\theta\partial\alpha)), -\partial^2(l(\alpha^*, \theta^*)/(\partial\theta^2))]$

We use the formulas for these partial derivatives that we found in Solution S4C59-2:

$-\partial^2(l(\alpha^*, \theta^*)/(\partial\alpha^2)) = -(-4/\alpha^{*2}) = -(-4/1.537341787^2) = 1.692462782.$

$-\partial^2(l(\alpha^*, \theta^*)/(\partial\theta\partial\alpha)) = -(4/\theta - 1/(13+\theta) - 1/(25+\theta) - 1/(36+\theta) - 1/(40+\theta)) =$

$-(4/30 - 1/43 - 1/55 - 1/66 - 1/70) = -0.0624584718.$

$-\partial^2(l(\alpha^*, \theta^*)/(\partial\theta^2)) =$

$-(-4\alpha/\theta^2 + (\alpha+1)/(13+\theta)^2 + (\alpha+1)/(25+\theta)^2 + (\alpha+1)/(36+\theta)^2 + (\alpha+1)/(40+\theta)^2)$

$= -(-4*1.537341787/(30)^2 + 2.537341787/(43)^2 + 2.537341787/(55)^2 + 2.537341787/(66)^2 + 2.537341787/(77)^2) = 0.0036111137.$

Thus, the observed information matrix is

[1.692462782, -0.0624584718]
[-0.0624584718, 0.0036111137].

Problem S4C59-4. For a Pareto distribution, the parameters α and θ are both estimated via the method of maximum likelihood by analyzing data from the following sample: 13, 25, 36, 40.

A possible combination of maximum likelihood estimates for α and θ is θ = 30 and α = 1.537341787. Use these values to find the estimate of the covariance matrix for α and θ.

Solution S4C59-4. The estimate of the covariance matrix is the inverse of the observed information matrix in this case. The observed information matrix (from Solution S4C59-3) is

[1.692462782, -0.0624584718]
[-0.0624584718, 0.0036111137].

What matrix,

(a, b)
(c, d)

multiplied by the matrix above, will give the identity matrix

[1, 0]
[0, 1]?

We consider that 1.692462782b + -0.0624584718d = 0, so
1.692462782b = 0.0624584718d and d = 27.09740942b.

Moreover, -0.0624584718a + 0.0036111137c = 0, so
0.0624584718a = 0.0036111137c and c = 17.29617979a.

Thus, we have
1.692462782a - 0.0624584718(17.29617979a) = 1 → 0.6121698243a = 1 → a = 1/0.6121698243 = a = 1.633533638.

From this it follows that c = 17.29617979a = 28.25389149.

Now -0.0624584718*b + 0.0036111137(27.09740942b) = 1 → 0.0353933546b = 1 → b = 28.25389149 and d = 27.09740942*28.25389149 = 765.6072653.

Therefore, the estimate of the covariance matrix is
[1.633533638, 28.25389149]
[28.25389149,765.6072653].

Problem S4C59-5. Review of Section 57: Similar to Question 179 of the Exam C Sample Questions from the Society of Actuaries. The mean of a sample of size 2 is 19. An exponential distribution is used to fit the data, and the parameter θ of the distribution is found via the method of maximum likelihood. Now a new sample of size 2 is taken. Find the probability that the mean of this new sample is greater than 23. Use the Exam 4 / C Tables where necessary.

Solution S4C59-5. The probability that the mean the new sample is greater than 23 is the same as the probability that the sum of the 2 values in the new sample exceeds 23*2 = 46. We can consider each value in the new sample as following an exponential distribution with parameter θ. Thus, the sum of the two values in the new sample follows a distribution that is the sum of these two exponential distributions - i.e. a gamma random variable with parameters α = 2 and θ. Since θ was estimated using the method of maximum likelihood, and the maximum likelihood estimate for an exponential distribution is just the sample mean, the estimate of θ is 19.

Thus, the sum of the two values in the new sample follows a gamma distribution with parameters α = 2 and θ = 19. We wish to find S(46) for this distribution. For a gamma distribution, F(x) = $\Gamma(\alpha; x/\theta) = (1/\Gamma(\alpha))_0{}^{x/\theta}\int t^{\alpha-1}e^{-t}dt$. Here, x/θ = 46/19, so F(46) = $\Gamma(2; 46/19) = (1/\Gamma(2))_0{}^{46/19}\int te^{-t}dt$ = $(1/(2-1)!)_0{}^{46/19}\int te^{-t}dt = {}_0{}^{46/19}\int te^{-t}dt$ = 0.6961145145, implying that S(46) = 1-0.6961145145 = **0.3038854855**.

Section 60

The Delta Method for Estimating Functions of Parameters

The **delta method** is used to estimate quantities that are the functions of parameters in a distribution, where the parameters themselves need to be estimated through methods such as maximum likelihood estimation.

Let θ be a parameter of a distribution, let θ^ be an estimate of θ (for instance, a maximum likelihood estimate), and let g(θ) be a function of θ. Here, we are ultimately concerned with estimating g(θ) by using g(θ^).

If there is only one parameter under consideration, the following is true:

"Let θ^ be an estimator of θ that has an asymptotic normal distribution with mean θ and variance σ^2/n. Then g(θ^) has an asymptotic normal distribution with mean g(θ) and asymptotic variance $g'(\theta)^2\sigma^2/n$." (Klugman, Panjer, and Willmot 2008, p. 398).

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Sources: Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 15, pp. 397-398.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C60-1. For an exponential distribution of random variable X, you are trying to find Pr(X > 50). However, the parameter θ is not known for the distribution and is estimated using the method of maximum likelihood. The sample used to estimate θ has size 34. Moreover, θ^, the maximum likelihood estimate of θ, follows an asymptotic normal distribution with mean θ and variance 9. Estimate the asymptotic variance of Pr(X > 50) using the delta method.

Solution S4C60-1. We are given that g(θ) = Pr(X > 50) = S(50) = $e^{-50/\theta}$ for an exponential distribution.

The asymptotic variance can be estimated via the formula Var^(g(θ)) = $g'(\theta)^2\sigma^2/n$.

First we find $g'(\theta) = (50/\theta^2)(e^{-50/\theta})$. Thus, $g'(\theta)^2 = (2500/\theta^4)(e^{-100/\theta})$.

We are given that the variance of θ^ is $\sigma^2/n = 9$. For an exponential distribution, the maximum likelihood estimate is the same as the sample mean, so Var(θ^) = Var(sample mean) = Var(X)/n. For an exponential distribution, Var(X) = θ^2, so Var(θ^) = $\theta^2/n = 9$. Since n = 34, $\theta^2 = 9*34 = 306 \rightarrow \theta = 17.49285568$. Thus, our estimate Var^(g(θ)) is $9(2500/\theta^4)(e^{-100/\theta})$ =

$9(2500/17.49285568^4)(e^{-100/17.49285568})$ = **Var^(g(θ)) = 0.0007907575714*10^{-4}**.

Problem S4C60-2. For an exponential distribution of random variable X, you are trying to find Pr(X > 50). However, the parameter θ is not known for the distribution and is estimated using the method of maximum likelihood. The sample used to estimate θ has size 34 and sample mean of 18. Moreover, θ^, the maximum likelihood estimate of θ, follows an asymptotic normal distribution with mean θ and variance 9. Use the delta method to develop a 90% confidence interval for Pr(X > 50).

Solution S4C60-2. Using our sample mean of 18 as the maximum likelihood estimate, we set as the center of the confidence interval the value S(50) ≈ $e^{-50/18}$ = 0.062176524.

We happen to know the true value of θ from Solution S4C60-1, and the variance of S(50) based on the true value of θ is 0.0007907575714*10^{-4}. Note that if we did not know the true value of θ, we would need to estimate Var(X) = θ^^{2}/n, where θ^ would be the maximum likelihood estimate of 18.

The z-scores associated with a 90% confidence interval are found via the Excel inputs "=NORMSINV(0.95)" and "=NORMSINV(0.05)". These values are 1.644853627 and -1.644853627, respectively.

Hence, Pr(X > 50) = 0.062176524 ± 1.644853627*√(0.0007907575714*10^{-4}) and our 90% confidence interval for Pr(X > 50) is **(0.0159225623, 0.1084304857)**.

Problem S4C60-3. Similar to Question 180 of the Exam C Sample Questions from the Society of Actuaries. The mean of a sample of size 2 is 19. An exponential distribution is used to fit the data, and the parameter θ of the distribution is found via the method of maximum likelihood. Now a new sample of size 2 is taken. The mean of this new sample is denoted by random variable Z. Use the delta method to estimate the variance of $F_Z(23)$, the probability that the mean of this new sample is 23 or less. Use the Exam 4 / C Tables where necessary.

Solution S4C60-3. The probability that the mean the new sample is less than or equal to 23 is the same as the probability that the sum of the 2 values in the new sample is less than or equal to 23*2 = 46. We can consider each value in the new sample as following an exponential distribution with parameter θ. Thus, the sum of the two values in the new sample follows a distribution that is the sum of these two exponential distributions - i.e. a gamma random variable with parameters α = 2 and θ. Since θ was estimated using the method of maximum likelihood, and the maximum likelihood estimate for an exponential distribution is just the sample mean, the estimate of θ is 19. We create a new random variable Q = 2Z, where Q is the sum of the values of the sample of size 2.

For a gamma distribution, $F(x) = \Gamma(\alpha; x/\theta) = (1/\Gamma(\alpha))_0{}^{x/\theta}\int t^{\alpha-1}e^{-t}dt$. Here, x/θ = 46/19, so $F_Q(46)$ =

$\Gamma(2; 46/\theta) = (1/\Gamma(2))_0{}^{46/\theta}\int te^{-t}dt = {}_0{}^{46/\theta}\int te^{-t}dt$.

To integrate te^{-t}, we use the Tabular Method:

Sign........u...........dv
+............t..............e^{-t}
-.............1.............$-e^{-t}$
+............0.............e^{-t}

Thus, $\int te^{-t}dt = C - te^{-t} - e^{-t}$, where C is a constant.

Hence, ${}_0{}^{46/\theta}\int te^{-t}dt = (-te^{-t} - (e^{-t})\big|{}_0{}^{46/\theta} = 1 - \exp(-46/\theta) - (46/\theta)\exp(-46/\theta) = F_Q(46) = F_Z(23) =$ g(θ).

Therefore, $g'(\theta) = (-46/\theta^2)\exp(-46/\theta) - (-46/\theta^2)\exp(-46/\theta) - (46^2/\theta^3)\exp(-46/\theta) =$

$g'(\theta) = -(46^2/\theta^3)\exp(-46/\theta)$. Thus, $g'(\theta)^2 = (46^4/\theta^6)\exp(-92/\theta)$. The maximum likelihood estimate for θ is 19, and the maximum likelihood estimate for the variance of the sample mean is the maximum likelihood estimate for the variance of θ (19^2), divided by the sample size of 2. Thus, the estimate for the variance of the sample mean is $19^2/2 = 180.5$.

Likewise, $g'(\theta)^2 \approx (46^4/\theta^6)\exp(-92/\theta) = (46^4/19^6)\exp(-92/19) = 0.0007509484486$ with the maximum likelihood estimate 19 of θ substituted for θ.

Thus, the estimate of the variance of $F_Z(23)$ is 0.0007509484486*180.5 = **0.135546195**.

Problem S4C60-4. Similar to Question 231 of the Exam C Sample Questions from the Society of Actuaries. It is known that the 90% linear confidence interval for the hazard rate function H(x) is (0.45, 0.99). Use the delta method to find the 90% linear confidence interval for the survival function S(x).

Solution S4C60-4. First, we find the center of the confidence interval for H(x): (0.45+0.99)/2 = 0.72.

The z-scores associated with a 90% confidence interval are found via the Excel inputs "=NORMSINV(0.95)" and "=NORMSINV(0.05)". These values are 1.644853627 and

-1.644853627, respectively.

We know that 0.99 = 0.72 + 1.644853627*√(Var(H(x))). Thus, 0.27 = 1.644853627*√(Var(H(x))) →

0.1641483446 = √(Var(H(x))) →

Var(H(x)) = 0.026944679.

We can also find the center of the confidence interval for S(x) by noting that S(x) = exp(-H(x)), so the center of the interval for S(x) is exp(-0.72) = 0.486752256.

To find Var(S(x)), we use the delta method. Let g(θ) = S(x) = exp(-H(x)).

Then g'(θ) is the first derivative of S(x) with respect to H(x):

g'(θ) = -exp(-H(x)) and so $g'(\theta)^2$ = exp(-2H(x)).

Hence,
Var^(S(x)) = exp(-2H(x))*Var(H(x)), where we use 0.72 as our estimate of H(x) in exp(-2H(x)).
Therefore, Var^(S(x)) = exp(-2*0.72)*0.026944679 = 0.0063839424.
Our 90% confidence interval for the survival function is thus
0.486752256 ± 1.644853627*√(0.0063839424) or **(0.3553291469, 0.6181753651)**.

Problem S4C60-5. Similar to Question 277 of the Exam C Sample Questions from the Society of Actuaries. You have collected a sample consisting of the following values:

23, 34, 46, 68, 89, 120, 120, 150, 250.

This data is known to follow an exponential distribution, with the parameter θ being estimated by the method of maximum likelihood. Use the delta method to estimate the variance of the survival function S(200).

Solution S4C60-5. First, we note that, since the distribution in question is exponential the maximum likelihood estimate of θ is the sample mean or
(23+34+46+68+89+120+120+150+250)/9 = 100.
Let S(200) = g(θ) = exp(-200/θ).
Then g'(θ) = $(200/\theta^2)$exp(-200/θ) and $g'(\theta)^2 = (200^2/\theta^4)$exp(-400/θ).
We substitute our maximum likelihood estimate of 100 for θ:
$g'(100)^2 = (200^2/100^4)$exp(-400/100) = 7.3262555555555*10^{-6}.
Now we find an estimate of Var(θ^) = Var(θ)/n = $100^2/9$ = 1111.11111111.
Thus, our estimate Var^(S(200)) = 1111.11111111*7.3262555555555*10^{-6} =
Var^(S(200)) = 0.008140284.

Section 61

Non-Normal Confidence Intervals for Parameters

It is sometimes necessary to develop confidence intervals for maximum likelihood estimators, where the maximum likelihood estimator is not normally distributed. The general formula for such a confidence interval or region is $\{\theta \mid l(\theta) \geq c\}$, i.e., the set of all values for which the loglikelihood function exceeds some value c. To produce a (100-α)% confidence interval, we set $c = l(\theta\hat{}) - 0.5\chi^2$. The number of degrees of freedom for the associated chi-square distribution is the number of parameters being estimated via the method of maximum likelihood, and the percentile of the chi-square distribution is the 1-α percentile.

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 15, pp. 402-404.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C61-1. The Weird distribution has a single parameter θ and a loglikelihood function in a particular instance of $l(\theta) = -\theta^2 + 11\theta - 24$. The maximum likelihood estimate of θ is θ^ = 3. Find the value of c to be used in constructing a 97.5% non-normal confidence interval for α. To look up relevant values of the chi-square distribution, see page 4 of the Exam 4 / C Tables.

Solution S4C61-1. First we find $l(\theta\hat{}) = -3^2 + 11*3 - 24 = l(\theta\hat{}) = 0$.

Now we find $0.5\chi^2$. Since one parameter is being estimated, there is one degree of freedom associated with the chi-square distribution. We find the value of the chi-square distribution for 1 degree of freedom and the 95th percentile: 3.841. Thus, $c = l(\theta\hat{}) - 0.5\chi^2 = 0 - 0.5*3.841 =$

c = -1.9205.

Problem S4C61-2. The Weird distribution has a single parameter θ and a loglikelihood function in a particular instance of $l(\theta) = -\theta^2 + 11\theta - 24$. The maximum likelihood estimate of θ is θ^ = 3. Develop a 95% non-normal confidence interval for θ. To look up relevant values of the chi-square distribution, see page 4 of the Exam 4 / C Tables.

Solution S4C61-2. The confidence interval contains all values for which $l(\theta) \geq c \rightarrow -\theta^2 + 11\theta - 24 \geq -1.9205 \rightarrow -\theta^2 + 11\theta - 22.0795 \geq 0$. We can find the endpoints of the confidence interval by solving the quadratic equation $-\theta^2 + 11\theta - 22.0795 = 0 \rightarrow \theta = 2.641591352$ and $\theta = 8.358408648$.

Thus, our 95% non-normal confidence interval for θ is **(2.641591352, 8.358408648)**.

Problem S4C61-3. For a Pareto distribution, it is known that the parameter θ = 30. The parameter α is estimated via the method of maximum likelihood by analyzing data from the following sample: 13, 25, 36, 40. It is known that the maximum likelihood estimate of α is α^ = 1.537341787 and that l(α) = 4*ln(α) + α*ln(0.0741330551) - ln(10926300). Find the value of c to be used in constructing a 97.5% non-normal confidence interval for α. To look up relevant values of the chi-square distribution, see page 4 of the Exam 4 / C Tables.

Solution S4C61-3. For α^ = 1.537341787, l(α^) = 4*ln(1.537341787) + 1.537341787*ln(0.0741330551) - ln(10926300) = l(α^) = -18.48646403.

Since the parameter θ is known, only one parameter (α) is being estimated, meaning that there is one degree of freedom associated with the chi-square distribution. We find the value of the chi-square distribution for 1 degree of freedom and the 97.5th percentile: 5.024.

Thus, c = -18.48646403 - $0.5\chi^2$ = -18.48646403 - 0.5*5.024 = **c = -20.99846403**.

Problem S4C61-4. You know that ages in a certain population follow a single-parameter Pareto distribution with $f(x) = \alpha/x^{\alpha+1}$. You are analyzing a sample of six observations, of which four observations are 19, 23, 75, and 95, and the other two observations have been right censored at 100. The method of maximum likelihood estimation estimates the parameter α to be 0.1655516859. Find the value of c to be used in constructing a 90% non-normal confidence interval for α. To look up relevant values of the chi-square distribution, see page 4 of the Exam 4 / C Tables.

Solution S4C61-4.

To develop the confidence interval, we need to find the loglikelihood function l(α):

We first create our likelihood function. The first four factors in the function each correspond to $\alpha/x^{\alpha+1}$, with the uncensored values substituted for x.

Thus, the product of the first four factors is $(\alpha/19^{\alpha+1})(\alpha/23^{\alpha+1})(\alpha/75^{\alpha+1})(\alpha/95^{\alpha+1})$.

The last two observations are censored at 100, so their contribution to the likelihood function will be a factor of $S(100)^2$. We first find the cdf $F(x) = \int\alpha/x^{\alpha+1} = 1-x^{-\alpha}$ (the constant of 1 is necessary to make the cdf equal 0 at x = 0). Then $S(x) = 1 - F(x) = x^{-\alpha}$, and so $S(100) = 100^{-\alpha}$, and thus the likelihood function is $L(\alpha) = (\alpha/19^{\alpha+1})(\alpha/23^{\alpha+1})(\alpha/75^{\alpha+1})(\alpha/95^{\alpha+1})(100^{-\alpha})(100^{-\alpha}) =$

$(\alpha^4/(3113625*31136250000^\alpha))$.

The loglikelihood function is $l(\alpha) = \ln(\alpha^4/(3113625*31136250000^\alpha))$ = 4*ln(α) - ln(3113625) - α*ln(31136250000). Thus, for α^ = 0.1655516859, l(α^) = 4*ln(0.1655516859) - ln(3113625) - 0.1655516859*ln(31136250000) = -26.14518553.

We also need to find χ^2. This is the value of the chi-square distribution at the 90th percentile for 1 degree of freedom. The Exam 4 / C Tables show this value as 2.706.

Thus, c = l(α^) - χ^2 = -26.14518553 - 0.5*2.706 = **c = -27.49818553**.

Problem S4C61-5. You know that ages in a certain population follow a single-parameter Pareto distribution with f(x) = $\alpha/x^{\alpha+1}$. You are analyzing a sample of six observations, of which four observations are 19, 23, 75, and 95, and the other two observations have been right censored at 100. The method of maximum likelihood estimation estimates the parameter α to be 0.1655516859. Develop a 90% non-normal confidence interval for α. To look up relevant values of the chi-square distribution, see page 4 of the Exam 4 / C Tables.

(a) (0.023, 0.6038)
(b) (0.0175, 0.6038)
(c) (0.0175, 0.359)
(d) (0.023, 0.359)
(d) (0.0638, 0.3415)
(e) (0.023, 0.3415)
(f) (0.0175, 0.3415)
(g) (0.0638, 0.6038)

Hint: Once you have reduced the inequality as far as possible, substitute the endpoints suggested in the answer choices above into the inequality to see whether the inequality holds. Otherwise, the inequality will not be possible to solve directly.

Solution S4C61-5. Thus, we know that, in constructing our confidence interval, l(α) ≥ c, and we know from Solution S4C61-1 that l(α) = 4*ln(α) - ln(3113625) - α*ln(31136250000) and c = -31.55718553. Thus, l(α) ≥ -27.49818553→
4*ln(α) - ln(3113625) - α*ln(31136250000) ≥ -27.49818553→
4*ln(α) - α*ln(31136250000) ≥ -12.54688733 →
ln(α) - 6.040409643α ≥ -3.136721832 →
exp(ln(α) - 6.040409643α) ≥ 0.043424919 →
exp(ln(α))*exp(-6.040409643α) ≥ 0.043424919 →
α*exp(-6.040409643α) ≥ 0.043424919.

To find the endpoints of the interval, we need to find the values of α such that α*exp(-6.040409643α) = 0.043424919.

We need to substitute the values 0.03925, 0.23, 0.359, and 0.439 for α in α*exp(-6.040409643α):
0.0175*exp(-6.040409643*0.03925) ≈ 0.01574;
0.023*exp(-6.040409643*0.023) ≈ 0.02002;
0.0638*exp(-6.040409643*0.0638) ≈ 0.0434;
0.3415*exp(-6.040409643*0.3415) ≈ 0.0434;
0.359*exp(-6.040409643*0.359) ≈ 0.04105;
0.6038*exp(-6.040409643*0.6038) ≈ 0.01574.

Thus, our best estimate of the desired confidence interval is **(d) (0.0638, 0.3415)**.

Section 62

Prior, Model, Marginal, Posterior, and Predictive Distributions in Bayesian Estimation

All information in the preface to the problems in this study guide is found in Section 15.5 of *Loss Models: From Data to Decisions* (cited below).

Bayesian estimation differs from conventional frequentist estimation in that, in Bayesian estimation, it is assumed that the population is variable and that only the actually observed data are relevant.

"The **prior distribution** is a probability distribution over the space of possible parameter values. It is denoted $\pi(\theta)$ and represents our opinion concerning the relative chances that various values of θ are the true value of the parameter." (404)

"An **improper prior distribution** is one for which the probabilities (or the probability density function) are nonnegative but their sum (or integral) is infinite." (405)

"The **model distribution** is the probability distribution for the data as collected given a particular value for the parameter. Its pdf is denoted $f_{X|\Theta}(x|\theta)$... [T]his is identical to the likelihood function..." (405)

"The **joint distribution** has pdf $f_{X,\Theta}(x, \theta) = f_{X|\Theta}(x|\theta)*\pi(\theta)$." (405)

"The **marginal distribution** has pdf $f_X(x) = \int f_{X|\Theta}(x|\theta)*\pi(\theta)*d\theta$." (405)

"The **posterior distribution** is the conditional probability distribution of the parameters, given the observed data. It is denoted as $\pi_{\Theta|X}(\theta|x)$." (406)

"The **predictive distribution** is the conditional probability distribution of a new observation y, given the data x. It is denoted $f_{Y|X}(y|x)$." (406)

The following formula is true for the posterior distribution:

$$\pi_{\Theta|X}(\theta|x) = f_{X|\Theta}(x|\theta)*\pi(\theta)/\int f_{X|\Theta}(x|\theta)*\pi(\theta)*d\theta$$

That is, (Posterior distribution) = (Joint distribution)/(Marginal distribution).

The following formula is true for the predictive distribution:

$f_{Y|X}(y|x) = \int f_{Y|\Theta}(y|\theta)*\pi_{\Theta|X}(\theta|x)*d\theta$, where $f_{Y|\Theta}(y|\theta)$ is the pdf of the new observation, given the parameter value. (406)

Source:
Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 15, pp. 404-406.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C62-1. The prior distribution for a particular population is assumed to be $\pi(\theta) = 1/\theta$, for $\theta > 0$. Two points of data are observed: 34 and 64. They are modeled to an exponential distribution with parameter θ. Find the probability density function for the joint distribution $f_{X,\Theta}(x, \theta)$.

Solution S4C62-1. We use the formula $f_{X,\Theta}(x, \theta) = f_{X|\Theta}(x|\theta)*\pi(\theta)$. Here, we know that $\pi(\theta) = 1/\theta$ and that $f_{X|\Theta}(x|\theta)$, the model distribution, is the exponential likelihood function

$(1/\theta)\exp(-34/\theta)*(1/\theta)\exp(-64/\theta) = (1/\theta^2)\exp(-98/\theta)$.

Thus, $f_{X,\Theta}(x, \theta) = (1/\theta^2)\exp(-98/\theta)*(1/\theta) =$ $\mathbf{f_{X,\Theta}(x, \theta) = (1/\theta^2)\exp(-98/\theta)}$.

Problem S4C62-2. The prior distribution for a particular population is assumed to be $\pi(\theta) = \theta^2$, for $\theta > 0$. Two points of data are observed: 0.1 and 0.12. They are modeled to an exponential distribution with parameter θ. Find the probability density function for the joint distribution $f_{X,\Theta}(x, \theta)$.

Solution S4C62-2. We use the formula $f_{X,\Theta}(x, \theta) = f_{X|\Theta}(x|\theta)*\pi(\theta)$. Here, we know that $\pi(\theta) = \theta^2$ and that $f_{X|\Theta}(x|\theta)$, the model distribution, is the exponential likelihood function

$(1/\theta)\exp(-0.1/\theta)*(1/\theta)\exp(-0.12/\theta) = (1/\theta^2)\exp(-0.22/\theta)$.

Thus, $f_{X,\Theta}(x, \theta) = (1/\theta^2)\exp(-0.22/\theta)*(\theta^2) =$ $\mathbf{f_{X,\Theta}(x, \theta) = \exp(-0.22/\theta)}$.

Problem S4C62-3. The prior distribution for a particular population is assumed to be $\pi(\theta) = \theta^2$, for $\theta > 0$. Two points of data are observed: 0.1 and 0.12. They are modeled to an exponential distribution with parameter θ. Find the probability density function for the marginal distribution $f_X(x)$. Based on your answer, determine whether the choice of an exponential model distribution is appropriate.

Solution S4C62-3. We use the formula $f_X(x) = \int f_{X|\Theta}(x|\theta)*\pi(\theta)*d\theta$.

That is, the marginal distribution is the integral of the joint distribution.

We know from Solution S4C62-2 that $f_{X|\Theta}(x|\theta)*\pi(\theta) = f_{X,\Theta}(x, \theta) = \exp(-0.22/\theta)$.

Thus, we take $_0{}^{\infty}\int \exp(-0.22/\theta)*d\theta = 83.33333333*\theta^2*\exp(-0.22/\theta)\,|\,_0{}^{\infty}$ = $\mathbf{f_X(x) = \infty}$**.** This is *not a valid probability density function*, so **the choice of an exponential model distribution is not appropriate**.

Problem S4C62-4. The value of the marginal distribution is $f_X(x) = 0.034$ for all x. For x = 5, the value of the joint distribution $f_{X,\Theta}(x, \theta)$ is 0.0019. Find the value of the posterior distribution $\pi_{\Theta|X}(\theta\,|\,x)$ at x = 5.

Solution S4C62-4. We use the formula (Posterior distribution) = (Joint distribution)/(Marginal distribution). Thus, $\pi_{\Theta|X}(\theta\,|\,5) = f_{X,\Theta}(5, \theta)/f_X(5) = 0.0019/0.034$ = $\mathbf{\pi_{\Theta|X}(\theta\,|\,5) = 0.055823529}$.

Problem S4C62-5. The value of the posterior distribution is $\pi_{\Theta|X}(\theta\,|\,6) = 0.53$ for all x = 6. For x = 6, the value of the joint distribution $f_{X,\Theta}(x, \theta)$ is 0.045. Find the value of the marginal distribution, given that it is constant for all x.

Solution S4C62-5. We use the formula (Posterior distribution) = (Joint distribution)/(Marginal distribution), which implies that (Marginal distribution) = (Joint distribution)/(Posterior distribution). For x = 6, (Marginal distribution) = 0.045/0.53 =$\mathbf{f_X(x) = 0.0849056604}$.

Section 63

Loss Functions and the Expected Value of the Predictive Distribution in Bayesian Estimation

All information in the preface to the problems in this study guide is found in Section 15.5 of *Loss Models: From Data to Decisions* (cited below).

"A **loss function** $l_j(\theta\hat{}_j, \theta_j)$ describes the penalty paid by the investigator when $\theta\hat{}_j$ is the estimate and θ_j is the true value of the jth parameter." (407)

"The **Bayes estimate** for a given loss function is the one that minimizes the expected loss given the posterior distribution of the parameter in question." (407)

"For **squared-error loss,** the loss function is $l(\theta\hat{}, \theta) = (\theta\hat{} - \theta)^2$. For **absolute loss,** the loss function is $l(\theta\hat{}, \theta) = |\theta\hat{} - \theta|$. For **zero-one loss** it is $l(\theta\hat{}, \theta) = 0$ if $\theta\hat{} = 0$ and is 1 otherwise." (407)

Theorem 63.1. "For squared-error loss, the Bayes estimate is the mean of the posterior distribution; for absolute loss, it is a median; for zero-one loss, it is a mode." (408)

The **expected value of the predictive distribution** provides an estimate of the (n+1)st observation if the first n observations and the prior distribution are known. The expected value of the predictive distribution is
$E(Y|X) = \int E(Y|\theta)*\pi_{\Theta|X}(\theta|X)*d\theta$.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Sources: Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 15, pp. 407-408.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C63-1. The mean of an exponential distribution is θ. The median of an exponential distribution is ln(2)*θ. The mode of an exponential distribution is 0. A posterior distribution is exponential with parameter θ = 90. Find the Bayes estimate of the mean for a loss function associated with this posterior distribution under the following circumstances:

(a) The loss function is a squared-error loss function.
(b) The loss function is an absolute loss function.
(c) The loss function is a zero-one loss function.

Solution S4C63-1. We use Theorem 63.1: "For squared-error loss, the Bayes estimate is the mean of the posterior distribution; for absolute loss, it is a median; for zero-one loss, it is a mode."

Thus, for (a), the Bayes estimate is the mean of the posterior distribution, or θ = **90**.

For (b), the Bayes estimate is the median of the posterior distribution or ln(2)*θ = ln(2)*90 = **62.38324625.**

For (c), the Bayes estimate is the mode of the posterior distribution or **0.**

Problem S4C63-2. A parameter is estimated as being 89, whereas the true value of the parameter is 85. Find the difference between the squared-error loss function and the absolute loss function in this case.

Solution S4C63-2. The squared-error loss function is $(θ\text{^}- θ)^2 = (89-85)^2 = 16$. The absolute loss function is │θ^- θ│ = │89- 85│= 4. Thus, our answer is 16-4 = **12.**

Problem S4C63-3. Similar to Question 5 of the Exam C Sample Questions from the Society of Actuaries. The number of elephants in a room in one particular day follows a binomial distribution with n = 3 and binomial probability q. (The associated probability function is p(x│q) = $C(3, x)*q^x*(1-q)^{3-x}$.) The following observations are made.

Day 1: There are 3 elephants in the room.
Day 2: There is 1 elephant in the room.
Day 3: There are 2 elephants in the room.

The prior distribution of elephants in the room is $8q^3$.

Develop a Bayesian estimate for the number of elephants in the room on Day 4.

Use a calculator or computer to evaluate any integrals.

Solution S4C63-3. We will ultimately be using the formula

$E(Y | X) = \int E(Y | q)*\pi_{Q|X}(q | X)*d\theta q$.

Let X be the previously observed data and let Y be the number of elephants in the room on Day 4. Then $E(Y | q) = nq$ (as this is a binomial distribution) = 3q.

We still need to find the posterior distribution $\pi_{Q|X}(q | X)$.

To do so, we need to find the joint distribution and the marginal distribution.

(Joint distribution) = (Prior distribution)*(Model distribution).

Thus, we need to find the model distribution, which is the probability, given q, that the numbers of elephants described in the problem will occur during the first three days. This probability is

Pr(3 elephants on Day 1)*Pr(1 elephant on Day 2)*Pr(2 elephants on Day 3) =

$(q^3)(3q*(1-q)^2)(3q^2*(1-q))$ = Model distribution = $9q^6(1-q)^3$.

Thus, (Joint distribution) = $(8q^3)(9q^6(1-q)^3) = (72q^9(1-q)^3)$.

(Marginal distribution) = ∫(Joint distribution) = $_0{}^1\int(72q^9(1-q)^3)*dq$, since q is a probability and can only range from 0 to 1. Thus, (Marginal distribution) = 0.0251748252 →

(Posterior distribution) = (Joint distribution)/(Marginal distribution) =

$(72q^9(1-q)^3)/0.0251748252 = 2860q^9(1-q)^3$.

Thus, $\int E(Y | q)*$ (Posterior distribution)*dq = $_0{}^1\int 3q*2860q^9(1-q)^3*dq = {}_0{}^1\int 8580q^{10}(1-q)^3*dq$ = $E(Y | X) = 15/7 = 2.142857143$. Thus, we can expect **15/7 elephants** to be in the room on Day 3.

Problem S4C63-4. Similar to Question 15 of the Exam C Sample Questions from the Society of Actuaries. The probability that a squirrel will eat a nut in a given hour is p. The prior distribution of p is uniform on the interval from 0 to 0.75. During every hour for the past 14 hours, the squirrel has eaten a nut. Estimate the posterior probability that the squirrel will also eat a nut during the 15th hour.

Solution S4C63-4. Let X be the previously observed data.

Our estimate will be $\int E(p | p)*\pi_{P|X}(p | X)*dp$.

Since p is both the parameter and the quantity being estimated, $E(p | p) = p$.

We still need to find the posterior distribution $\pi_{P|X}(p | X)$.

To do so, we need to find the joint distribution and the marginal distribution.

(Joint distribution) = (Prior distribution)*(Model distribution).

Thus, we need to find the model distribution, which is the probability, given p, that the squirrel ate a nut each of the 14 hours. Thus, (Model distribution) = p^{14}.

(Prior distribution) = 1/(0.75-0) = 4/3.

Thus, (Joint distribution) = $(4/3)p^{14}$ and (Marginal distribution) = ∫(Joint distribution) = $_0{}^{0.75}\int(4/3)p^{14}*dp = (4/45)p^{15} \mid_0{}^{0.75} = (4/45)*0.75^{15}$.

Thus, (Posterior distribution) = (Joint distribution)/(Marginal distribution) =

$(4/3)p^{14}/((4/45)*0.75^{15}) = 1122.463708p^{14}$.

Our estimate for p during the 15th hour is $_0{}^{0.75}\int p*1122.463708p^{14} = {}_0{}^{0.75}\int 1122.463708p^{15} = (1122.463708/16)p^{16} \mid_0{}^{0.75} = (1122.463708/16)(0.75)^{16}$ = **p = 45/64 = 0.703125**.

Problem S4C63-5. Similar to Question 24 of the Exam C Sample Questions from the Society of Actuaries. The probability that a person is named Quintus is q. The prior distribution of q is $p(q) = (19/7)q^{12/7}$, with 0 < q < 1. You observe one person named Quintus. Find the posterior probability that q is greater than 0.5.

Solution S4C63-5. We need to find the posterior distribution of q.

To do so, we need to find the joint distribution and the marginal distribution.

(Joint distribution) = (Prior distribution)*(Model distribution).

The model distribution is the probability, given q, that one person named Quintus was observed. This means that the model distribution is q.

Thus, (Joint distribution) = $q*(19/7)q^{12/7} = (19/7)q^{19/7}$.

Hence, (Marginal distribution) = ∫(Joint distribution) = $_0{}^{1}\int(19/7)q^{19/7}*dq = (19/7)(7/26)q^{26/7} \mid_0{}^{1}$ =19/26. Thus, (Posterior distribution) = (Joint distribution)/(Marginal distribution) =

$(19/7)q^{19/7}/(19/26) = (26/7)q^{19/7}$ and $\Pr(q > 0.5) = {}_{0.5}{}^{1}\int(26/7)q^{19/7}*dq = q^{26/7} \mid_{0.5}{}^{1} = 1 - 0.5^{26/7}$ = **Pr(q > 0.5) = 0.9238116466**.

Section 64

Credibility Intervals and the Bayesian Central Limit Theorem in Bayesian Estimation

All information in the preface to the problems in this study guide is found in Section 15.5 of *Loss Models: From Data to Decisions* (cited below).

In Bayesian estimation, a **credibility interval** can be constructed as follows:

"The points a < b define a 100(1-α)% **credibility interval** for θ_j, provided that $\Pr(a \leq \Theta_j \leq b \mid x) \geq (1-\alpha)$." (408)

Theorem 64.1. "If the posterior random variable $\theta_j \mid x$ is continuous and unimodal, then the 100(1-α)% credibility interval with smallest width b - a is the unique solution to the following system of equations:

$\int_a^b \pi_{\Theta\_j \mid X}(\theta_j \mid X)*d\theta_j = 1-\alpha$

$\pi_{\Theta \mid X}(a \mid X) = \pi_{\Theta \mid X}(b \mid X)$.

This interval is a special case of a highest posterior density (HPD) credibility set." (409)

Theorem 64.2. "For any posterior distribution, the 100(1-α)% **HPD credibility set** is the set of parameter values C such that $\Pr(\theta_j \in C) \geq (1-\alpha)$ and $C = \{\theta_j: \pi_{\Theta\_j \mid X}(\theta_j \mid X) \geq c\}$ for some c, where c is the largest value for which the first inequality holds." (410)

Theorem 64.3. "If $\pi(\theta)$ (the prior distribution) and $f_{X \mid \Theta}(x \mid \theta)$ (the model distribution) are both twice differentiable in the elements of θ and other commonly satisfied assumptions hold, then the posterior distribution of Θ given X = x is asymptotically normal." (410) This theorem is called the **Bayesian central limit theorem.**

Several other equations are useful in Bayesian estimation:

$E(Y) = E(E(Y \mid X)) \rightarrow$ For the predictive distribution, $E(Y \mid X) = E_{\Theta \mid X}(E(Y \mid \Theta))$.

$Var(Y) = E(Var(Y \mid X)) + Var(E(Y \mid X)) \rightarrow$ For the predictive distribution, $Var(Y \mid X) = E_{\Theta \mid X}(Var(Y \mid \Theta)) + Var_{\Theta \mid X}(E(Y \mid \Theta))$.

Source:

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 15, pp. 408-413.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C64-1. The probability that a squirrel will eat a nut in a given hour is p. The prior distribution of p is uniform on the interval from 0 to 0.75. During every hour for the past 14 hours, the squirrel has eaten a nut. Use Theorem 64.1 to develop a nonzero width for a 95% credibility interval for the posterior probability that the squirrel will also eat a nut during the 15th hour. You are aware that the Bayesian estimate of p is 0.703125.

Solution S4C64-1. We first need to find the posterior distribution $\pi_{P|X}(p|X)$.

To do so, we need to find the joint distribution and the marginal distribution.

(Joint distribution) = (Prior distribution)*(Model distribution).

Thus, we need to find the model distribution, which is the probability, given p, that the squirrel ate a nut each of the 14 hours. Thus, (Model distribution) = p^{14}.

(Prior distribution) = 1/(0.75-0) = 4/3.

Thus, (Joint distribution) = $(4/3)p^{14}$ and (Marginal distribution) = ∫(Joint distribution) = $_0{}^{0.75}\int(4/3)p^{14}*dp = (4/45)p^{15}|_0{}^{0.75} = (4/45)*0.75^{15}$.

Thus, (Posterior distribution) = (Joint distribution)/(Marginal distribution) =

$(4/3)p^{14}/((4/45)*0.75^{15}) = 1122.463708p^{14}$.

Thus, according to Theorem 64.1, one of the equations of the system that we will need to solve is

$_a{}^b\int\pi_{\Theta\_j|X}(\theta_j|X)*d\theta_j = 1-\alpha \rightarrow {}_a{}^b\int 1122.463708p^{14}*dp = 0.95$.

The other equation is

$\pi_{\Theta|X}(a|X) = \pi_{\Theta|X}(b|X) \rightarrow 1122.463708a^{14} = 1122.463708b^{14}$.

The second equation can only be solved either such that a = b or that a = -b.

The solution a = b would give a width of zero.

The solution a = -b implies the following:
$_{-b}{}^b\int 1122.463708p^{14}*dp = 74.83089805p^{15}|_{-b}{}^b = 149.6617961*b^{15} = 0.95$.

Thus, $b^{15} = 0.0063476453 \rightarrow b = 0.7136865466$.

The width of the credibility interval is (b-a) = (b - (-b)) = 2b = **1.427373093**, which is not particularly informative, because it encompasses all the possible values of p - since p is a probability that can only range from 0 to 1.

Problem S4C64-2. The probability that a squirrel will eat a nut in a given hour is p. The prior distribution of p is uniform on the interval from 0 to 0.75. During every hour for the past 14 hours, the squirrel has eaten a nut. Use Theorem 64.3 (the Bayesian central limit theorem) to develop a nonzero width for a 95% credibility interval for the posterior probability that the squirrel will also eat a nut during the 15th hour. You are aware that the Bayesian estimate of p (the mean of the posterior distribution) is 0.703125 and that the credibility interval is centered at that value.

Solution S4C64-2. From Solution S4C64-1, the model distribution is p^{14}, which is twice differentiable. The prior uniform distribution is also twice differentiable. Thus, Theorem 64.3 applies, and the posterior distribution can be thought as being asymptotically normal. For developing the confidence interval, we assume normality.

From Solution S4C64-1, the posterior distribution is $1122.463708p^{14}$.

The mean of the posterior distribution is given as 0.703125.

The variance is $E(p^2) - E(p)^2 = {}_0{}^{0.75}\int 1122.463708p^{14}*p^2*dp - 0.703125^2 =$
${}_0{}^{0.75}\int 1122.463708p^{16}*dp - 0.703125^2 = 0.4963235293 - 0.703125^2 = Var(p) = 0.0019387637$.

Now we find the z-scores associated with the endpoints of the interval.

These values can be found in MS Excel using the inputs "=NORMSINV(0.975)" and "=NORMSINV(0.025)". The desired z-scores are ±1.959963985. Thus, the credibility interval's endpoints are $0.703125 \pm 1.959963985\sqrt{(0.0019387637)}$ and the interval is **(0.6168250528, 0.7894249472)**. This shows that the assumptions of the Bayesian central limit theorem can generate much more precise and useful credibility intervals than would result if these assumptions were not invoked.

Problem S4C64-3. The probability that a squirrel will eat a nut in a given hour is p. The prior distribution of p is uniform on the interval from 0 to 0.75. During every hour for the past 14 hours, the squirrel has eaten a nut. Use Theorem 64.3 (the Bayesian central limit theorem) to estimate the probability that $p > 0.8$.

Solution S4C64-3. From Solution S4C64-2, we know that E(p) = 0.703125 and Var(p) = 0.0019387637. Assuming normality, we can use the central limit theorem:
$Pr(p > 0.8) = 1-\Phi((0.8-0.703125)/\sqrt{(0.0019387637)}) = 1-\Phi(2.200134729)$, which can be found in Excel via the input "=1-NORMSDIST(2.200134729)". Our desired answer is **0.013898669**.

Problem S4C64-4. This problem is meant to provide practice with the double-expectation formulas. It is known that $f_{Y|X}(y|x) = (1/x)e^{-y/x}$. X is uniformly distributed on the interval (3, 6). Find E(Y).

Solution S4C64-4. We use the formula $E(Y) = E(E(Y|X))$. First, we find $E(Y|X)$. Since the distribution of $Y|X$ is exponential with parameter X, and the expected value of an exponential distribution is its parameter, it follows that $E(Y|X) = X$. Thus, $E(Y) = E(X) = (6+3)/2 =$ **E(Y) = 4.5**.

Problem S4C64-5. This problem is meant to provide practice with the double-expectation formulas. It is known that $f_{Y|X}(y|x) = (1/x)e^{-y/x}$. X is uniformly distributed on the interval (3, 6). Find Var(Y).

Solution S4C64-5. We use the formula $\text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X))$.

From Solution S4C64-4, we know that $E(Y|X) = X$. Thus, $\text{Var}(E(Y|X)) = \text{Var}(X)$, where X is uniformly distributed on the interval (3, 6). Thus, $\text{Var}(X) = (6-3)^2/12 = 9/12 = ¾ = 0.75$.

Hence, $\text{Var}(E(Y|X)) = 0.75$.

Now we find $\text{Var}(Y|X)$. Since the distribution of $Y|X$ is exponential with parameter X, and the variance of an exponential distribution is the square of its parameter, $\text{Var}(Y|X) = X^2$.

$E(\text{Var}(Y|X)) = E(X^2)$, where X is uniformly distributed on the interval (3, 6).

$E(X^2) = \text{Var}(X) + E(X)^2 = 0.75 + 4.5^2 = 21$.

Thus, $E(\text{Var}(Y|X)) = 21$ and $\text{Var}(Y) = 21 + 0.75 =$ **Var(Y) = 21.75**.

Section 65

Conjugate Prior Distributions in Bayesian Estimation

"A prior distribution is said to be a **conjugate prior distribution** for a given model if the resulting posterior distribution is from the same family as the prior (but perhaps with different parameters." (414)

Theorem 65.1. "Suppose that given $\Theta = \theta$ the random variables $X_1, ..., X_n$ are independent and identically distributed (i.i.d.) with probability function (pf) $f_{X\_j|\Theta}(x_j | \theta) = p(x_j)*\exp(r(\theta)*x_j)/q(\theta)$, where Θ has pdf $\pi(\theta) = (q(\theta))^{-k}*\exp(\mu k r(\theta))*r'(\theta)/c(\mu,k)$, where k and μ are parameters of the distribution and $c(\mu,k)$ is the normalizing constant. Then the posterior pf $\pi_{\Theta|X}(\theta | x)$ is of the same form as $\pi(\theta)$." (414)

The above is a case where the model distribution belongs to the linear exponential family.

The posterior distribution would be of the same form as the prior distribution $\pi(\theta)$, but with parameters $k^* = k + n$ and $\mu^* = (k\mu + nx^-)/(k+n)$, where n is the sample size and x^- is the sample mean.

The following are useful cases of specific conjugate prior, model, and posterior distributions:

Case 1:
- Prior distribution of λ is gamma with parameters α and θ.
- Model distribution is Poisson with parameter λ.
- Posterior distribution is gamma with parameters $\alpha + \Sigma x_i$ and $\theta/(n\theta+1)$.

Case 2:
- Prior distribution of λ is inverse gamma with parameters α and θ.
- Model distribution is exponential with parameter λ.
- Posterior distribution is inverse gamma with parameters $\alpha + n$ and $\theta + \Sigma x_i$.

Case 3:
- Prior distribution of q is beta with parameters a, b, and 1.
- Model distribution is binomial with parameters m and q.
- Posterior distribution is beta with parameters $a + \Sigma x_i$, $b + km - \Sigma x_i$, and 1.

Case 4:
- Prior distribution of λ is gamma with parameters α and θ.
- Model distribution is inverse exponential with parameter λ.
- Posterior distribution is gamma with parameters $\alpha + n$ and $(1/\theta + \Sigma(1/x_i))^{-1}$.

Case 5:

●Prior distribution of λ is normal with parameters μ and a^2.
● Model distribution is normal with parameters λ and σ^2.
● Posterior distribution is normal with parameters $(x/\sigma^2 + \mu/a^2)/(1/\sigma^2 + 1/a^2)$ and $1/(1/\sigma^2 + 1/a^2)$.

Case 6:

● Prior distribution of λ is Pareto with parameters α and θ.
● Model distribution is uniform from 0 to λ.
● Posterior distribution is Pareto with parameters $\alpha + n$ and max(x) - i.e., the maximum value encountered in the sample.

Integrals to memorize

The following two integrals are useful to memorize for the exam:

${}_0{}^\infty\int x^a*e^{-cx}*dx = a!/(c^{a+1})$ if a is an integer.
${}_0{}^\infty\int(e^{-c/x}/x^k)*dx = (k-2)!/c^{k-1}$ if k is an integer ≥ 2.

Sources: *Loss Models: From Data to Decisions,* (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 15, pp. 414-416.

Loss Models Summary by JavaGeek. Section 12, pp. 37-38.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C65-1. Find ${}_0{}^\infty\int x^6*e^{-7x}*dx + {}_0{}^\infty\int(e^{-4/x}/x^7)*dx$.

Solution S4C65-1. We use the formulas ${}_0{}^\infty\int x^a*e^{-cx}*dx = a!/(c^{a+1})$ and ${}_0{}^\infty\int(e^{-c/x}/x^k)*dx = (k-2)!/c^{k-1}$.
Thus, ${}_0{}^\infty\int x^6*e^{-7x}*dx = 6!/(7^7) = 0.0008742712888$ and ${}_0{}^\infty\int(e^{-4/x}/x^7)*dx = (7-2)!/4^{7-1} = 120/4096 = 0.029296875$. Thus, our answer is 0.0008742712888 + 0.029296875= **0.0301711463.**

Problem S4C65-2. You are analyzing a data set where the model distribution is Poisson with parameter λ. The prior distribution of λ is gamma with parameters $\alpha = 6$ and $\theta = 24$. The sample size is 9, and the sample mean is 3. Find the mean of the posterior distribution. Refer to the Exam 4 / C Tables as needed.

Solution S4C65-2. This is an instance of Case 1, described above. The posterior distribution is gamma with parameters $\alpha + \Sigma x_i$ and $\theta/(n\theta+1)$. We are given that n = 9. We can find Σx_i as (sample mean)*(sample size) = 9*3 = 27. Thus, the posterior distribution has parameters 6 + 27 = 33 and 24/(9*24+1) = 24/217. The posterior distribution is gamma, and the mean of a gamma distribution is the product of its parameters, so the mean of the posterior distribution is 33*(24/217) = **792/217 = 3.649769585**.

Problem S4C65-3. You are analyzing a data set where the model distribution is exponential with parameter λ. The prior distribution of λ is inverse gamma with parameters $\alpha = 6$ and $\theta = 24$. The sample size is 9, and the sample mean is 3. Find the mean of the posterior distribution. Refer to the Exam 4 / C Tables as needed.

Solution S4C65-3. This is an instance of Case 2, described above. In this case, the posterior distribution is inverse gamma with parameters α + n and θ + Σx_i. We are given that n = 9. We can find Σx_i as (sample mean)*(sample size) = 9*3 = 27. Thus, the posterior distribution has parameters α* = 6 + 9 = 15 and θ* = 24 + 27 = 51.

The expected value of an inverse gamma distribution is θ*Γ(α-1)/Γ(α). Here, since α is an integer, Γ(α-1)/Γ(α) = (α-2)!/(α-1)! = 1/(α-1), so the mean is θ/(α-1) = **51/14 = 3.642857143**.

Problem S4C65-4. You are analyzing a data set where the model distribution is normal with mean λ and variance 16, and the prior distribution of λ is normal with mean 30 and variance 25. You observe a single observation: x = 40. Find the coefficient of variation of the posterior distribution.

Solution S4C65-4. This is an instance of Case 5, described above, so the posterior distribution is normal with parameters $(x/\sigma^2 + \mu/a^2)/(1/\sigma^2 + 1/a^2)$ and $1/(1/\sigma^2 + 1/a^2)$.
Here, $\sigma^2 = 16$ and $a^2 = 25$, so the variance is 1/(1/16 + 1/25) = 400/41.
Here, x = 40 and μ = 30, so the mean is (40/16 + 30/25)/(1/16 + 1/25) = 1480/41.
Coefficient of variation = (Standard deviation)/(Mean) = √(400/41)/(1480/41) = **0.0865287059.**

Problem S4C65-5. You are analyzing a data set where the model distribution is uniform from 0 to λ. The prior distribution is Pareto with parameters α = 3 and θ = 6000. You observe 18 values, of which the greatest is 5000. Find the mean of the posterior distribution. Refer to the Exam 4 / C Tables as needed.

Solution S4C65-5. This is an instance of Case 5, described above, so the posterior distribution is Pareto with parameters α + n and max(x). Here, n = 18, and max(x) = 5000, so for the posterior distribution, α* = 3 + 18 = 21, and θ* = 5000. Thus, the mean of the posterior distribution is θ*/(α*-1) = 5000/20 = **250.**

Section 66

Exam-Style Questions on Bayesian Estimation

This section provides some exam-style questions intended to help students review Bayesian estimation concepts and procedures.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source:

Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C66-1. Similar to Question 29 of the Exam C Sample Questions from the Society of Actuaries.

There are two types of porcupines - spiky and sharp. The prior probability of a random porcupine being spiky is 0.43, and the prior probability of a random porcupine being sharp is 0.57. A spiky porcupine has a 0.5 probability of confronting a predator within a year, whereas a sharp porcupine has a 0.3 probability of confronting a predator within a year. A porcupine can only confront one predator per year. Within 7 years, you observe that a randomly chosen porcupine has confronted 4 predators. Find the posterior probability that the porcupine will confront a predator during the 8th year.

Solution S4C66-1. First, we find the probabilities of confronting 4 predators in 7 years, given each type of porcupine: Pr(4 predators │ spiky porcupine) = C(7, 4)*0.5^4*0.5^3 = 0.2734375.

Pr(4 predators │ sharp porcupine) = C(7, 4)*0.3^4*0.7^3 = 0.0972405.

So the probability of a randomly chosen porcupine confronting 4 predators in 7 years is

Pr(spiky porcupine)*Pr(4 predators │ spiky porcupine) + Pr(sharp porcupine)*Pr(4 predators │ sharp porcupine) = 0.43*0.2734375 + 0.57*0.0972405 = 0.17300521.

Now we can find the posterior probabilities of a porcupine being spiky or sharp.

Pr(spiky porcupine | 4 predators) = Pr(spiky porcupine)*Pr(4 predators | spiky porcupine)/Pr(4 predators) = 0.43*0.2734375/0.17300521 = 0.6796218738.

Pr(sharp porcupine | 4 predators) = Pr(sharp porcupine)*Pr(4 predators | sharp porcupine)/Pr(4 predators) = 0.57*0.0972405/0.17300521 = 0.3203781262.

Thus, Pr(predator in 8^{th} year) =

Pr(predator in a given year | spiky porcupine)*Pr(spiky porcupine | 4 predators in prior years) + Pr(predator in a given year | sharp porcupine)*Pr(sharp porcupine | 4 predators in prior years) = 0.5*0.6796218738 + 0.3*0.3203781262 = **0.4359243748.**

Problem S4C66-2. Similar to Question 43 of the Exam C Sample Questions from the Society of Actuaries. The prior distribution of Θ has pdf $\pi(\theta) = 1/\theta^5$ for θ > 1. Given Θ = θ, observations follow a Pareto distribution with parameters α =5 and θ. A single observation of size 7 is made. Estimate the posterior probability that θ exceeds 3.

Useful properties of Pareto distributions: $f(x) = \alpha\theta^\alpha/(x+\theta)^{\alpha+1}$.

Solution S4C66-2. First, we find the model distribution, where we substitute the observed value of 7 for x: $\alpha\theta^\alpha/(7+\theta)^{\alpha+1} = 5\theta^5/(7+\theta)^6$.

(Joint distribution) = (Prior distribution)*(Model distribution) = $(1/\theta^5)(5\theta^5/(7+\theta)^6) = 5/(7+\theta)^6$.

(Marginal distribution) = ∫(Joint distribution) = $_1{}^\infty\int(5/(7+\theta)^6)d\theta = -1/(7+\theta)^5 \,|\, _1{}^\infty = 1/8^5 = 1/32768$.

Thus, (Posterior distribution) = (Joint distribution)/(Marginal distribution) =

$(5/(7+\theta)^6)/(1/32768) = 163840/(7+\theta)^6$.

$\Pr(\theta > 3) = {}_3{}^\infty\int(163840/(7+\theta)^6)d\theta = -32768/(7+\theta)^5 \,|\, _3{}^\infty = 32768/10^5 =$ **0.32768.**

Problem S4C66-3. Similar to Question 45 of the Exam C Sample Questions from the Society of Actuaries.

The amount of a claim Y is uniformly distributed on the interval (0, θ).

You observe two claims of magnitudes 20 and 40. The posterior distribution of θ given those claims is $\pi_{\Theta|X}(\theta \,|\, 20, 40) = 3*40^3/\theta^4$ for θ > 40.

Find the Bayesian premium E(Y | 20, 40).

Solution S4C66-3. We use the formula $E(Y | X) = \int E(Y | \theta)*\pi_{\Theta|X}(\theta | X)*d\theta$.

Here, X = {20, 40} and E(Y | θ) = θ/2, since Y is uniformly distributed. Thus,

$E(Y \mid 20, 40) = {}_{40}{}^{\infty}\int(\theta/2)(3*40^3/\theta^4)*d\theta = {}_{40}{}^{\infty}\int(1.5*40^3/\theta^3)*d\theta = -0.75*40^3/\theta^2 \mid_{40}^{\infty} =$

$0.75*40 =$ **E(Y | 20, 40) = 30**.

Problem S4C66-4. Similar to Question 64 of the Exam C Sample Questions from the Society of Actuaries. The amount of a claim Y is uniformly distributed on the interval (0, θ).

You observe two claims of magnitudes 20 and 40. The posterior distribution of θ given those claims is $\pi_{\Theta|X}(\theta \mid 20, 40) = 3*40^3/\theta^4$ for θ > 40. Find the probability that the next claim will exceed 35.

Solution S4C66-4. Let Y_3 be the next claim. Pr(next claim exceeds 35) = $\Pr(Y_3 > 35) =$

$\int \Pr(Y_3 > 35 \mid X)*\pi_{\Theta|X}(\theta \mid X)*d\theta.$

Here, X = {20, 40} and $\Pr(Y_3 > 35 \mid X) = (\theta-35)/\theta$, since Y is uniformly distributed on (0, θ).

Thus, $\Pr(Y_3 > 35) = {}_{40}{}^{\infty}\int((\theta-35)/\theta)(3*40^3/\theta^4)*d\theta = {}_{40}{}^{\infty}\int(3*40^3/\theta^4 - 105*40^3/\theta^5) =$

$(-40^3/\theta^3 + 26.25*40^3/\theta^4) \mid_{40}^{\infty} = 1 - 26.25/40 =$ **0.34375.**

Problem S4C66-5. Similar to Question 142 of the Exam C Sample Questions from the Society of Actuaries. The number of elephants that step on houses in a year follows a Poisson distribution with mean θ. The prior distribution of θ has pdf $\pi(\theta) = e^{-\theta}/(1-e^{-k})$, where $0 < \theta < k$. You know that the unconditional probability of *no* elephants stepping on houses in a given year is 0.93. Find k.

Solution S4C66-5. Let X be the number of elephants that step on houses in a year. Pr(X = 0) = ${}_0{}^{k}\int \Pr(X = 0 \mid \theta)*\pi(\theta)*d\theta$.

Since X follows a Poisson distribution, $\Pr(X = 0 \mid \theta) = e^{-\theta}$. Thus,

$\Pr(X = 0) = {}_0{}^{k}\int(e^{-\theta}*e^{-\theta}/(1-e^{-k}))*d\theta = {}_0{}^{k}\int(e^{-2\theta}/(1-e^{-k})) = -0.5e^{-2\theta}/(1-e^{-k}) \mid_0^{k} = (-0.5e^{-2k} + 0.5)/(1-e^{-k}) =$

$0.5(1-e^{-2k})/(1-e^{-k}) = 0.5(1-e^{-k})(1+e^{-k})/(1-e^{-k}) = 0.5(1+e^{-k}) = 0.93 \rightarrow (1+e^{-k}) = 1.86 \rightarrow e^{-k} = 0.86 \rightarrow$ $k = -\ln(0.86) =$ **k = 0.1508228897**.

Section 67

Exam-Style Questions on Bayesian Estimation – Part 2

This section provides some further exam-style questions intended to help students review Bayesian estimation concepts and procedures.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C67-1. Similar to Question 157 of the Exam C Sample Questions from the Society of Actuaries. An elephant can run away from at most one mouse per year. The probability that an elephant runs away from one mouse in a given year is q, which has prior density $\pi(q) = q^3/0.157625$, $0.4 < q < 0.9$. An elephant is observed to run away from no mice during Year 1 and then to run away from one mouse each in Year 2 and Year 3. Find the posterior probability that $0.6 < q < 0.7$.

Solution S4C67-1. The model distribution is $(1-q)*q*q = (1-q)*q^2$. We do not multiply this by C(3, 1) because the sequence of events is established. Thus, the joint distribution is

(Model distribution)*(Prior distribution) = $(1-q)*q^2*q^3/0.157625 = (1-q)*q^5/0.157625$.

(Marginal distribution) = ∫(Joint distribution) = $\int_{0.4}^{0.9}((1-q)*q^5/0.157625)dq = 0.1255940486$.

(Posterior distribution) = (Joint distribution)/(Marginal distribution) =

$(1-q)*q^5/(0.157625*0.1255940486) = 50.51331146(1-q)*q^5$.

Thus, $\Pr(0.6 < q < 0.7) = \int_{0.6}^{0.7}(50.51331146(1-q)*q^5)dq =$ **0.2054049243.**

Problem S4C67-2. Similar to Question 184 of the Exam C Sample Questions from the Society of Actuaries. The number of camels that fly in a given year follows a Poisson distribution with mean λ. The prior distribution of λ is a mixture of exponential distributions:

$\pi(\lambda) = 0.2(5)e^{-5\lambda} + 0.8(9)e^{-9\lambda}$ for $\lambda > 0$. In Year 1, you observe that 3 camels have flown. Find the Bayesian expectation of the number of camels that will fly in Year 2.

Solution S4C67-2. The model distribution is the Poisson probability that the number of camels that will fly is 3, i.e., $e^{-\lambda}*\lambda^3/(3!) = e^{-\lambda}*\lambda^3/6$.

Thus, the joint distribution is (Model distribution)*(Prior distribution) =

$(e^{-\lambda}*\lambda^3/6)(0.2(5)e^{-5\lambda} + 0.8(9)e^{-9\lambda}) = (1/6)\lambda^3e^{-6\lambda} + (6/5)\lambda^3e^{-10\lambda}$

(Marginal distribution) = ∫(Joint distribution) = $_0{}^{\infty}\int((1/6)\lambda^3e^{-6\lambda} + (6/5)\lambda^3e^{-10\lambda})d\lambda$.

We recall the following formula from Section 65: $_0{}^{\infty}\int x^a*e^{-cx}*dx = a!/(c^{a+1})$ if a is an integer.

Thus, $_0{}^{\infty}\int\lambda^3e^{-6\lambda} = 3!/(6^4) = 1/(6^3)$ and $_0{}^{\infty}\int \lambda^3e^{-10\lambda}d\lambda = 3!/(10^4)$.

$_0{}^{\infty}\int((1/6)\lambda^3e^{-6\lambda} + (6/5)\lambda^3e^{-10\lambda})d\lambda = (1/6)(1/(6^3)) + (6/5)(6/(10^4)) =$

(Marginal distribution) = 0.0014916049.

Thus, (Posterior distribution) = (Joint distribution)/(Marginal distribution) =

$((1/6)\lambda^3e^{-6\lambda} + (6/5)\lambda^3e^{-10\lambda})/0.0014916049=$

$(111.7364675)\lambda^3e^{-6\lambda} + (804.5025658)\lambda^3e^{-10\lambda}$.

Thus, $E(\lambda) = {}_0{}^{\infty}\int((\lambda)(111.7364675)\lambda^3e^{-6\lambda} + (804.5025658)\lambda^3e^{-10\lambda})d\lambda =$

$_0{}^{\infty}\int((111.7364675)\lambda^4e^{-6\lambda} + (804.5025658)\lambda^4e^{-10\lambda})d\lambda =$

$(111.7364675)*4!/6^5 + (804.5025658)*4!/10^5 =$ **E(λ) = 0.5379462562camels**.

Problem S4C67-3. Similar to Question 191 of the Exam C Sample Questions from the Society of Actuaries. The number of incidents of simultaneous global warming and cooling per year follow a Poisson distribution with mean λ. The distribution of λ is a gamma distribution with the following prior density function: $\pi(\lambda) = (5\lambda)^7*e^{-5\lambda}/(720\lambda)$. During the past 4 years, the following numbers of incidents of simultaneous global warming and cooling were observed:

3, 3, 2, 1. Find the Bayesian expectation for λ given this data.

Useful information about gamma distributions: $f(x) = (x/\theta)^{\alpha}*e^{-x/\theta}/(x\Gamma(\alpha))$; $E(X) = \alpha\theta$.

Solution S4C67-3. This situation is an instance of Case 1 from Section 65:

Prior distribution of λ is gamma with parameters α and θ.

Model distribution is Poisson with parameter λ.

Posterior distribution is gamma with parameters $\alpha + \Sigma x_i$ and $\theta/(n\theta+1)$.

Here, $\pi(\lambda) = (5\lambda)^7 * e^{-5\lambda}/(720\lambda)$, so it can be inferred that α = 7 and θ = 1/5.

The sample size n is 4, and Σx_i, the sum of sample values, is 9. Thus, in the posterior distribution, α* = 7 + 9 = 16 and θ* = (1/5)/(4(1/5) + 1) = (1/9).

The Bayesian expectation for λ is the mean of the posterior distribution, i.e., 16(1/9) = **16/9 = 1.77777778**.

Problem S4C67-4. Similar to Question 242 of the Exam C Sample Questions from the Society of Actuaries. A llama can wear 0, 1, or 2 green socks on a given day. The probability that the llama will wear 0 green socks is 0.4. The probability that the llama will wear 1 green sock is 0.6-q. The probability that the llama will wear 2 green socks is q. The probability q has prior density $\pi(q) = q^4/0.000484$ for 0.1 < q < 0.3. On three consecutive days, the llama was observed to wear 2 green socks each day. Find the Bayesian expectation for the number of green socks the llama wears per day.

Solution S4C67-4. The model distribution is q^3. Thus, the joint distribution is
(Model distribution)*(Prior distribution) = $q^3 * q^4/0.000484 = q^7/0.000484$.
(Marginal distribution) = ∫(Joint distribution) = $_{0.1}{}^{0.3}\int(q^7/0.000484)dq$ = 0.0169421488.
Posterior distribution) = (Joint distribution)/(Marginal distribution) =
$(q^7/0.000484)/0.0169421488 = 121951.2195q^7$.
E(q) = $_{0.1}{}^{0.3}\int q*121951.2195q^7*dq = {}_{0.1}{}^{0.3}\int 121951.2195q^8*dq$ = E(q) = 0.2666937669.
E(green socks per day) = 0*0.4 + 1*(0.6 - 0.2666937669) = 2*0.2666937669 = **0.8666937669 green socks per day.**

Problem S4C67-5. Similar to Question 253 of the Exam C Sample Questions from the Society of Actuaries. There are m binomial random variables, each with parameter q. The sum of the binomial random variables is S_m. The prior distribution of Q is beta with parameters a = 3, b = 9, and θ = 1. Find the smallest value of m such that $E(S_m)$, the mean of the marginal distribution of S_m, is at least 20.

Useful information about beta distributions: E(X) = θ*Γ(a+b)*Γ(a+1)/(Γ(a)*Γ(a+b+1)).

Solution S4C67-5. The expected value of each of the binomial random variables is E(q), so $E(S_m)$ = m*E(q), and E(q) can be found by analyzing the prior distribution of Q.
We find E(q) = θ*Γ(a+b)*Γ(a+1)/(Γ(a)*Γ(a+b+1)) = 1*Γ(12)*Γ(4)/(Γ(3)*Γ(13)) = 11!*3!/(2!*12!) = 3/12 = ¼. Thus, if we want $E(S_m)$ = 20, then 20 = m*(1/4) → **m = 80**.

Section 68

Exam-Style Questions on Bühlmann Credibility

Bühlmann credibility, also known as **least squares credibility**, differs from limited fluctuation credibility in the manner the credibility factor Z is calculated. Under Bühlmann credibility, Z = N/(N + K), where N is the number of observations, and K is calculated as follows:
K = EPV/VHM.

Let θ be the parameter that is being estimated.

EPV is the Expected Value of the Process Variance, which is $E_\theta(\mathrm{Var}(X|\theta)) = E(X^2) - E_\theta(E(X|\theta)^2)$. The variance X is first determined for each of the possible values of θ, after which the expected value this variance is taken over the distribution of θ to find EPV.

VHM is the Variance of Hypothetical Means, which is $\mathrm{Var}_\theta(E(X|\theta)) = E_\theta(E(X|\theta)^2) - E(X)^2$.

To find VHM, first find the mean of all observations, then find the second raw moment of the means of all observations for each of the possible values of θ, then subtract the square of the former result from the latter result.

It is useful to know that EPV + VHM = Total variance.

For Bühlmann credibility, as for classical (limited fluctuation) credibility, the credibility-weighted result is $ZX^- + (1\text{-}Z)M$, where X^- is the sample mean and M is the external mean.

Moreover, Bühlmann credibility offers the least squares approximation of the Bayesian estimate. This means that the Bühlmann credibility result is one such that the weighted average of the squared differences between the Bühlmann credibility result and the Bayesian estimate for each possible case is the smallest possible.

If the prior distribution of θ is gamma and the model distribution is Poisson, the Bühlmann credibility estimate is the same as the Bayesian estimate.

An excellent free source for learning about both limited fluctuation credibility and Bühlmann credibility is "Credibility" by Howard C. Mahler and Curtis Gary Dean. Actuarial students are strongly encouraged to read this entire study guide and to solve all of the examples and practice problems provided. Solutions are offered by Mr. Mahler and Mr. Dean in the study guide, so students have an opportunity to check their work and answers.

The current study guide will focus on exam-style questions pertaining to Bühlmann credibility. If more basic practice is desired, Sections 3-5 of Mr. Mahler's and Mr. Dean's study guide offer a variety of practice problems that will build up students' skills to the point of their being able to solve the problems here.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source:
Exam C Sample Questions and Solutions from the Society of Actuaries.

Mahler, H.C.; and Dean, C.G., "Credibility," *Foundations of Casualty Actuarial Science* (Fourth Edition), 2001, Casualty Actuarial Society, Chapter 8, Section 1 and Sections 2-5.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C68-1. Similar to Question 18 of the Exam C Sample Questions from the Society of Actuaries. Event A is four times more likely to occur than Event B, and one of the events must occur, but only once a year. The following are probabilities of monetary losses L in the event that each event occurs:

Event A: Pr(L = 34) = 0.3, Pr(L = 56) = 0.4, Pr(L = 89) = 0.3.

Event B: Pr(L = 34) = 0.2, Pr(L = 56) = 0.6, Pr(L = 89) = 0.2.

In Year 1, you observe a single loss of 56. Use Bühlmann credibility to find the expected loss in Year 2.

Solution S4C68-1. First, we find EPV and VHM.

EPV: E(A) = 0.3*34 + 0.4*56 + 0.3*89 = 59.3.

$E(A^2) = 0.3*34^2 + 0.4*56^2 + 0.3*89^2 = 3977.5$.

$Var(A) = E(A^2) - E(A)^2 = 3977.5 - 59.3^2 = Var(A) = 461.01$.

E(B) = 0.2*34 + 0.6*56 + 0.2*89 = 58.2.

$E(B^2) = 0.2*34^2 + 0.6*56^2 + 0.2*89^2 = 3697$.

$Var(B) = E(B^2) - E(B)^2 = 3697 - 58.2^2 = Var(B) = 309.76$.

Pr(A) = 4*Pr(B) and Pr(A) + Pr(B) = 1 → Pr(A) = 4/5 and Pr(B) = 1/5.

Thus, EPV = (4/5)Var(A) + (1/5)Var(B) = (4/5)461.01 + (1/5)309.76 = EPV = 430.76.

VHM: E(Total) = (4/5)E(A) + (1/5)E(B) = (4/5)59.3 + (1/5)58.2 = 59.08.

E(Total^2) = (4/5)E(A)^2 + (1/5)E(B)^2 = (4/5)59.3^2 + (1/5)58.2^2 = 3490.64.

VHM = E(Total^2) - E(Total)^2 = 3490.64 - 59.08^2 = VHM = 0.1936.

K = EPV/ VHM = K = 2225.

N = 1, so Z = 1/(1+K) = 1/2226 = 0.000492362983 →

Expected loss in Year 2 = 0.000492362983*56 + (1-0.000492362983)*59.08 = **59.07861635.**

Problem S4C68-2. Similar to Question 32 of the Exam C Sample Questions from the Society of Actuaries. The number of elephants that a snake steps on in a year follows a Poisson distribution with mean λ. The prior distribution of λ is a gamma distribution with α = 3 and θ = 1.9. During a three-year period, you observe that the snake has stepped on 2 elephants in Year 1, 4 elephants in Year 2, and 0 elephants in Year 3. Use *Bühlmann credibility* to estimate the number of elephants the snake will step on in Year 4. Use the Exam 4 / C Tables as needed.

Solution S4C68-2. First, we find the prior mean E(λ), which is also the EPV, since the distribution of the relevant variable is Poisson. Since λ is gamma-distributed, E(λ) = αθ = 3*1.9 = 5.7. Now we find VHM = Var(λ) = αθ^2 = 3*1.9^2 = 10.83.

Thus, K = EPV/ VHM = 10/19. Since N = 3, Z = 3/(3+10/19) = Z = 0.8507462687.

The mean number of observed stepped-on elephants per year is (2+4+0)/3 = 2.

Thus, the expected number of stepped-on elephants in Year 4 is

2*0.8507462687 + 5.7(1-0.8507462687) = **2.552238806 elephants.**

Problem S4C68-3. Similar to Question 32 of the Exam C Sample Questions from the Society of Actuaries. The number of elephants that a snake steps on in a year follows a Poisson distribution with mean λ. The prior distribution of λ is a gamma distribution with α = 3 and θ = 1.9. During a three-year period, you observe that the snake has stepped on 2 elephants in Year 1, 4 elephants in Year 2, and 0 elephants in Year 3. Use *Bayesian estimation* to find the expected number of elephants the snake will step on in Year 4. Use the Exam 4 / C Tables as needed.

Solution S4C68-3. This is an instance of Case 1 of conjugate prior distributions from Section 65:

Prior distribution of λ is gamma with parameters α and θ.

Model distribution is Poisson with parameter λ.

Posterior distribution is gamma with parameters $\alpha + \Sigma x_i$ and $\theta/(n\theta+1)$.

The prior distribution is gamma with $\alpha = 3$ and $\theta = 1.9$. Here, n = number of years observed = 3, and Σx_i = sum of observations = 2+4+0 = 6. Thus, $\alpha^* = 3 + 6 = 9$ and $\theta^* = 1.9/(3*1.9+1) = 19/67$. The expected number of stepped-on elephants in Year 4 is thus $\alpha^*\theta^* = 9(19/67) =$ **2.552238806 elephants.** Note the Bühlmann credibility and Bayesian estimation give the exact same result!

Problem S4C68-4. Similar to Question 35 of the Exam C Sample Questions from the Society of Actuaries. The unconditional probability of observing a value of 3 is ½, and the unconditional probability of observing a value of 5 is ½. If 3 is observed, the Bayesian estimate of the second observation is 3.4. If 5 is observed, the Bayesian estimate of the second observation is 4.8. Use Bühlmann credibility to estimate the second observation, given that the first observation is 3.

Solution S4C68-4. Bühlmann credibility is known as least squares credibility, because Bühlmann credibility gives a least squares approximation of the Bayesian estimate. This means that the weighted average of the squared differences between the Bühlmann credibility estimate and each of the Bayesian estimates must be minimized.

The credibility factor Z is associated with the value of the first observation, whereas the complement of credibility (1-Z) is associated with the external mean, which is the weighted average of the two Bayesian estimates: (1/2)3.4 + (1/2)4.8 = 4.1.

Hence, the quantity to be minimized in order to find Z is

$q(Z) = (1/2)(3Z + 4.1(1-Z) - 3.4)^2 + (1/2)(5Z + 4.1(1-Z) - 4.8)^2$

$q(Z) = (1/2)(-1.1Z + 0.7)^2 + (1/2)(0.9Z - 0.7)^2$

$q'(Z) = -1.1(-1.1Z + 0.7) + 0.9(0.9Z - 0.7) = 0$ at minimum.

Thus, $1.1(-1.1Z + 0.7) = 0.9(0.9Z - 0.7) \rightarrow$

$-1.21Z + 0.77 = 0.81Z - 0.63 \rightarrow 2.02Z = 1.4 \rightarrow Z = 70/101 = 0.6930693069$.

Hence, our Bühlmann credibility estimate is (70/101)3 + (31/101)4.1 = **3.337623762.**

Problem S4C68-5. Similar to Question 48 of the Exam C Sample Questions from the Society of Actuaries. There are 15 observations in a sample, and the sum of the observations is 78. The magnitude of each observation X depends on the parameter Θ via the following joint distribution:

$\Pr(X = 4, \Theta = 3) = 0.3$

$\Pr(X = 6, \Theta = 3) = 0.4$

$\Pr(X = 4, \Theta = 4) = 0.2$

$\Pr(X = 6, \Theta = 4) = 0.1$

Use this information to find the Bühlmann credibility premium, i.e., the expected value of the next observation.

Solution S4C68-5. First, we find EPV and VHM. We can separate the joint distribution by values of Θ. $\Pr(\Theta = 3) = 0.3+0.4 = 0.7$, so $\Pr(X = 4 \mid \Theta = 3) = 0.3/0.7 = 3/7$ and $\Pr(X = 6 \mid \Theta = 3) = 4/7$. Thus, $E(X \mid \Theta = 3) = 4(3/7) + 6(4/7) = 36/7$. $E(X^2 \mid \Theta = 3) = 4^2(3/7) + 6^2(4/7) = 192/7$. Hence, $\mathrm{Var}(X \mid \Theta = 3) = 192/7 - (36/7)^2 = 0.9795918367$.

$\Pr(\Theta = 4) = 1 - 0.7 = 0.3$, so $\Pr(X = 4 \mid \Theta = 4) = 0.2/0.3 = 2/3$, and $\Pr(X = 6 \mid \Theta = 4) = 1/3$.

Thus, $E(X \mid \Theta = 4) = 4(2/3) + 6(1/3) = 14/3$. $E(X^2 \mid \Theta = 4) = 4^2(2/3) + 6^2(1/3) = 68/3$.

Hence, $\mathrm{Var}(X \mid \Theta = 4) = 68/3 - (14/3)^2 = 8/9$.

$EPV = 0.7*\mathrm{Var}(X \mid \Theta = 3) + 0.3*\mathrm{Var}(X \mid \Theta = 4) = 0.7*0.9795918367 + 0.3(8/9) = 0.9523809524$.

$E(X) = 0.7*(36/7) + 0.3*(14/3) = 5$.

$E(X^2) = 0.7*(36/7)^2 + 0.3*(14/3)^2 = 25.04761905$

$VHM = E(X^2) - E(X)^2 = 25.04761905 - 5^2 = 0.0476190476$.

$K = EPV/VHM = 0.9523809524/0.0476190476 = 20$.

The number of observations, N, is 15 so $Z = 15/(15 + 20) = (3/7)$.

The observed mean is $78/15 = 5.2$. Hence, the Bühlmann credibility premium is $(3/7)5.2 + (4/7)5$ = **5.085714286 = 178/35**.

Section 69

Exam-Style Questions on Bühlmann Credibility – Part 2

This section gives additional exam-style practice with Bühlmann credibility.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Sources:
Exam C Sample Questions and Solutions from the Society of Actuaries.

Mahler, H.C.; and Dean, C.G., "Credibility," *Foundations of Casualty Actuarial Science* (Fourth Edition), 2001, Casualty Actuarial Society, Chapter 8, Section 1 and Sections 2-5.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C69-1. Similar to Question 62 of the Exam C Sample Questions from the Society of Actuaries. An individual can only suffer no losses or one loss in a year. Losses for each individual are denoted by random variable X. Any individual's probability of loss is constant over time and is independent of any other individual's probability of loss. A large number of individuals is analyzed, and the probability of a single loss occurring in a year for a randomly selected individual is 0.4. The variance of the individual loss probabilities is known to be 0.05. An individual is observed for 20 years, during which time he has suffered 5 losses. Find the Bühlmann credibility estimate of the number of losses expected for that individual during the subsequent 30 years.

Solution S4C69-1. We can find the Bühlmann credibility estimate of the number of losses expected for that individual during the subsequent year and multiply that answer by 30.

First, though, we need to find K = EPV/VHM.

Let θ be the probability of loss for each individual.

$\text{EPV} = E_\theta(\text{Var}(X \mid \theta)) = E(X^2) - E_\theta(E(X \mid \theta)^2)$

$\text{VHM} = \text{Var}_\theta(E(X \mid \theta)) = E_\theta(E(X \mid \theta)^2) - E(X)^2$

VHM is the variance of the individual loss probabilities, which is given as 0.05. VHM is known, and E(X) is known to be 0.4, so $E_\theta(E(X|\theta)^2)$ = VHM + $E(X)^2 = 0.05 + 0.4^2 = 0.21$.

Now we can find EPV = $E(X^2) - E_\theta(E(X|\theta)^2)$

Since X can be at most 1, $E(X^2) = E(X)$, since 1^2 is always 1. Thus,

EPV = $E(X) - E_\theta(E(X|\theta)^2) = 0.4 - 0.21 = 0.19$.

Hence, K = 0.19/0.05 = 3.8.

N is the number of years of observation, i.e., 20, so Z = N/(N+K) = 20/23.8 = Z = 0.840336134.

The prior mean of losses per year is 0.4, and the observed mean is 5/20 = 0.25.

Therefore, in a single year, the number of losses can be expected to be

0.25*0.840336134 + 0.4(1-0.840336134) = 0.2739495799.
In 30 years, the number of losses can be expected to be 30*0.2739495799 = **8.218487397 losses.**

Problem S4C69-2. Similar to Question 67 of the Exam C Sample Questions from the Society of Actuaries. The random variable $X_i = {}_{j=1}{}^{N\_i}\Sigma(Y_{ij})$ represents the annual amount spent on bouncy castles by the ith customer. $N_1, N_2, ..., N_{440}$ are independent random variables which follow a negative binomial distribution with parameters r and β = 0.7. The distribution of r is exponential with mean 4. The random variables $Y_{i1}, Y_{i2}, ..., Y_{iN\_i}$ are independent, and each follows a Pareto distribution with α = 4 and θ = 90. Find Z, the Bühlmann credibility factor for this group of customers.

Relevant properties for exponential distributions: E(X) = θ; Var(X) = θ^2.

Relevant properties for negative binomial distributions: E(X) = rβ; Var(X) = rβ(1+β).

Relevant properties for Pareto distributions: E(X) = θ/(α -1); Var(X) = $2\theta^2/((\alpha -1)(\alpha -2)) - \theta^2/(\alpha -1)^2$.

Solution S4C69-2. For each customer,

E(X|r) = E(N|r)*E(Y) = rβ*θ/(α -1) = r*0.7*90/(4-1) = E(X|r) = 21r.

The VHM is the variance of the mean amount spent for each consumer: Var(21r) = 21^2*Var(r) = $21^2 * 4^2$ = VHM = 7056. Var(r) = 4^2 since r follows an exponential distribution with mean 4, and the variance of an exponential distribution is the square of the mean.

The EPV is E(Var(X|r)) = E(E(N|r)*Var(Y) + Var(N|r)*$E(Y)^2$), by the formula for the variance of an aggregate random variable in Section 33.

$E(Y) = \theta/(\alpha - 1) = 90/(4-1) = E(Y) = 30$.

$Var(Y) = 2\theta^2/((\alpha - 1)(\alpha - 2)) - 30^2 = 2*90^2/((4-1)(4-2)) - 30^2 = 2700 - 900 = Var(Y) = 1800$.

$E(N \mid r) = r\beta = 0.7r$.

$Var(N \mid r) = r\beta(1+\beta) = 0.7*1.7r = 1.19r$.

Thus, $EPV = E(0.7r*1800 + 1.19r*30^2) = E(2331r) = 2331*E(r) = 2331*4 = EPV = 9324$.

Thus, $K = EPV/VHM = 9324/7056 = 1.321428571$.

Our sample size is 440, so $Z = 440/(440 + 1.321428571) =$ **$Z = 0.9970057457$**.

Problem S4C69-3. Similar to Question 70 of the Exam C Sample Questions from the Society of Actuaries. Elephants are either pink or orange, with equal probability. Each type of elephant is either spotted or striped. The following is data collected on the number of peanuts, X, eaten by each sub-type of elephant.

For pink and spotted elephants: $E(X) = 3$, $Var(X) = 0.3$
For pink and striped elephants: $E(X) = 5$, $Var(X) = 0.1$
For pink elephants in total: $E(X) = 3.6$, $Var(X) = 0.45$

For orange and spotted elephants: $E(X) = 6$, $Var(X) = 0.5$
For orange and striped elephants: $E(X) = 2$, $Var(X) = 0.6$
For orange elephants in total: $E(X) = 4.8$, $Var(X) = 0.8$.

Each elephant's number of peanuts eaten is independent of the data for every other elephant. Find the Bühlmann credibility factor Z for an individual elephant.

Solution S4C69-3. First, we need to find the probabilities of each sub-type of elephant.

$E(X \mid Pink) = 3.6 = Pr(Spotted \mid Pink)*3 + (1 - Pr(Spotted \mid Pink))*5 \rightarrow Pr(Spotted \mid Pink) = 0.7$.

Since $Pr(Pink) = 0.5$, $Pr(Spotted\ and\ Pink) = 0.7*0.5 = 0.35$ and thus $Pr(Striped\ and\ Pink) = 0.5 - 0.35 = 0.15$.

$E(X \mid Orange) = 4.8 = Pr(Spotted \mid Orange)*6 + (1 - Pr(Spotted \mid Orange))*2 \rightarrow$
$Pr(Spotted \mid Orange) = 0.7$, implying that $Pr(Spotted\ and\ Orange) = 0.35$ and $Pr(Striped\ and\ Orange) = 0.15$.

Thus, for all elephants, $E(X) = 0.35*3 + 0.15*5 + 0.35*6 + 0.15*2 = E(X) = 4.2$.

$E(X^2) = 0.35*3^2 + 0.15*5^2 + 0.35*6^2 + 0.15*2^2 = 20.1 \rightarrow VHM = 20.1 - 4.2^2 = 2.46$.

EPV is the weighted average of the variances for each sub-type:
EPV = 0.35*0.3 + 0.15*0.1 + 0.35*0.5 + 0.15*0.6 = EPV = 0.385.

Thus, K = EPV/VHM = 0.385/2.46 = 0.156504065. Since we are considering an individual elephant, N = 1, and Z = 1/(1+0.156504065) = **Z = 0.8646748682**.

Problem S4C69-4. Similar to Question 78 of the Exam C Sample Questions from the Society of Actuaries. The value X of a billionaire's bank account has mean μ and variance 400 billion. The random variable μ has mean 900 billion and variance 70 billion.

In 6 successive years, the following "snapshot" values of the bank account were observed at the end of each year: 540 billion, 450 billion, 230 billion, 900 billion, 800 billion, 800 billion.

Use Bühlmann credibility to estimate the value of the bank account during the 7^{th} year.

Solution S4C69-4. First, we need to find EPV and VHM. VHM is the variance of hypothetical means, i.e., Var(μ) = 70 billion. EPV is the expected value of the process variance, i.e., Var(X) = 400 billion. Thus, K = EPV/VHM = 400/70 = 40/7. We observe for 6 years, so N = 6 and Z = 6/(6+40/7) = Z = 0.512195122. We are given that the prior mean is E(μ) = 900 billion, while the observed mean is (540+450+230+900+800+800)/6 = 620 billion.

Hence, our answer, in billions, is 620*(0.512195122) + 900(1-0.512195122) = **756.5853659 billion = approximately $756,585,365,900.**

Problem S4C69-5. Similar to Question 133 of the Exam C Sample Questions from the Society of Actuaries. The annual number of new stores built in a town follows a binomial distribution with probability function $C(4, x)*q^x(1-q)^{4-x}$. The pdf of the prior distribution of q is π(q) = 1.5q for 0 < q < 1. Over the past 2 years, 5 stores have been built. Use Bühlmann credibility to estimate the number of stores that will be built during the next year.

Relevant properties for binomial distributions: E(X) = mq; Var(X) = mq(1-q).

Solution S4C69-5. First, we find E(X│q) = mq = 4q and Var(X│q) = 4q(1-q).
The prior mean is $E(X) = E_Q(E(X│q)) = {}_0{}^1\int 4q*1.5q*dq = 6q^3/3 │_0{}^1 = E(X) = 6/3 = 2$.
EPV is $E(Var(X│q)) = {}_0{}^1\int 4q(1-q)*1.5q*dq = {}_0{}^1\int 6q(1-q)dq = 1$, so EPV = 1.
VHM is $Var(E(X│q)) = Var(4q) = E((4q)^2) - E((4q))^2$.
We already know $E((4q)) = E_Q(E(X│q)) = 2$.
$E((4q)^2) = {}_0{}^1\int 16q^2*1.5q*dq = 6q^4 │_0{}^1 = 6$, so VHM = 6 - 2^2 = 2 and K = EPV/VHM = ½.
We have observations for 2 years, so N = 2 and Z = 2/(2+1/2) = 0.8.
The observed mean is 5/2 = 2.5. Thus, our estimate is 0.8*2.5 + 0.2*2 = **2.4 stores.**

Section 70

Exam-Style Questions on Bühlmann-Straub Credibility

Bühlmann-Straub credibility is a more general case of Bühlmann credibility. In Bühlmann-Straub credibility, N is calculated as the number of *exposures*, i.e., the number of observations of which values can potentially be ascertained - irrespective of which time period these observations occur. So, for instance, if there were X insureds in period 1 with average claims of A per insured and Y insureds in period 2 with average claims of B per insured, then N would be X + Y, and total claims would be AX + BY.

With Bühlmann-Straub credibility, first calculate the credibility-weighted estimate *per exposure* and then multiply by the desired number of exposures to get the total credibility-weighted estimate. The rest of the calculations are substantively identical to Bühlmann credibility.

Some special cases:

Property 70.1. If the prior distribution is gamma with parameters α and θ, and the model distribution is Poisson, EPV = $\alpha\theta$, VHM = $\alpha\theta^2$, and K = $1/\theta$.

Property 70.2. If the model distribution is Poisson, then EPV = mean of the prior distribution and VHM = variance of the prior distribution of the Poisson parameter λ.

A good formal introduction to Bühlmann-Straub credibility, with some additional practice problems and solutions, can be found in Curtis Gary Dean's study note, "Topics in Credibility Theory."

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Sources: Dean, Curtis Gary, "Topics in Credibility Theory," 2004 (SOA Study Note).

Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C70-1. Similar to Question 21 of the Exam C Sample Questions from the Society of Actuaries. The number of rabbits pulled out of a hat by an individual magician in a day follows a Poisson distribution with mean λ. The rabbit frequencies for each magician are independent. The prior distribution of λ is a gamma distribution such that $f(\lambda) = (20\lambda)^4 * e^{-20\lambda}/(6\lambda)$.

You observe the following:

On Day 1, 40 magicians pulled 23 rabbits out of hats.
On Day 2, 50 magicians pulled 34 rabbits out of hats.

There are 60 magicians on Day 3. Determine the Bühlmann-Straub credibility estimate of the number of rabbits pulled out of hats on Day 3.

Relevant properties of gamma distributions: $f(x) = (x/\theta)^\alpha * e^{-x/\theta}/(x*\Gamma(\alpha))$.

$E(X) = \alpha\theta$; $Var(X) = \alpha\theta^2$.

Solution S4C70-1. Examining the prior distribution of λ, we see that it has α = 4 and θ = 1/20.

If the model distribution is Poisson and the prior distribution is gamma, it is the case that $K = \alpha\theta/(\alpha\theta^2) = 1/\theta$. Here, K = 1/(1/20) = 20.

N is the number of exposures observed, which is 40+50 = 90.

The prior mean per magician is the mean of the gamma distribution or 4/20 = 1/5.

The observed mean per magician is (23+34)/90 = 19/30.

Z = N/(N+K) = 90/(90+20) = 9/11.

Thus, the Bühlmann-Straub credibility estimate *per magician* is (9/11)(19/30) + (2/11)(1/5) = 61/110 = 0.55454545454545.

Bühlmann-Straub credibility estimate for 60 magicians is 60*61/110 = **366/11 = 33.2727272727 rabbits.**

Problem S4C70-2. Similar to Question 50 of the Exam C Sample Questions from the Society of Actuaries. There are three classes of rhinoceroses: blue, striped, and invisible.

Each rhinoceros can eat at most one chocolate bar per year. The probability of a blue rhinoceros eating a chocolate bar in a year is 0.3. The probability of a striped rhinoceros eating a chocolate bar in a year is 0.4. The probability of an invisible rhinoceros eating a chocolate bar in a year is 0.2. A type of rhinoceros is selected at random (with probability 1/3), and 5 rhinoceroses are selected at random from that type. The total number of chocolate bars eaten for one year is 1. If 15 rhinoceroses are selected at random from the same type, use Bühlmann-Straub credibility to estimate the total number of chocolate bars eaten in a year.

Solution S4C70-2. Here, the observed mean is 1 chocolate bar/5 rhinoceroses = 0.2 chocolate bars/rhinoceros. The prior mean is 0.3*(1/3)+0.4*(1/3) + 0.2*(1/3) = 0.3. N, the number of exposures, is 5. For each individual rhinoceros,

E(Chocolate bars | Blue rhinoceros) = 0*0.7 + 1*0.3 = 0.3.

Var(Chocolate bars | Blue rhinoceros) = $0.7(0 - 0.3)^2 + 0.3(1 - 0.3)^2$ = 0.21.

E(Chocolate bars | Striped rhinoceros) = 0*0.6 + 1*0.4 = 0.4

Var(Chocolate bars | Striped rhinoceros) = $0.6(0-0.4)^2 + 0.4(1-0.4)^2$ = 0.24.

E(Chocolate bars | Invisible rhinoceros) = 0*0.8 + 1*0.2 = 0.2.

Var(Chocolate bars | Invisible rhinoceros) = $0.8(0-0.2)^2 + 0.2(1-0.2)^2$ = 0.16.

EPV = (0.21 + 0.24 + 0.16)/3 = EPV = 0.2033333333.

E(Total) = 0.3; E(Total^2) = 0.3^2*(1/3)+0.4^2*(1/3) + 0.2^2*(1/3) =0.09666666667.

VHM = 0.09666666667 - 0.3^2 = 0.00666666667.

K = EPV/VHM = 0.2033333333/0.00666666667 = K = 30.5.

Thus, Z = 5/(5+30.5) = 0.1408450704.

For each individual rhinoceros, estimated number of chocolate bars eaten is

0.2*0.1408450704 + 0.3(1-0.1408450704) = 0.285915493.

Thus, for 15 rhinoceroses, the estimated total number of chocolate bars eaten is 15*0.285915493 = **4.288732394 chocolate bars.**

Problem S4C70-3. Similar to Question 72 of the [Exam C Sample Questions] from the Society of Actuaries. Monetary losses for each gambler are independent and have a common mean and variance. For all gamblers, the overall average loss per gambler is 100. The variance of hypothetical means is 49, and the expected value of the process variance is 392. For a randomly selected group of gamblers, you know that there were 34 gamblers in year 1, whose average losses per gambler were 200. In year 2, there were 56 gamblers, and the average loss per gambler was 70. In year 3, there were 98 gamblers, and the average loss per gambler was 130. In year 4, there were 89 gamblers, and the average loss per gambler was 150. Use Bühlmann-Straub credibility to determine the expected loss per gambler in year 5.

Solution S4C70-3. The prior mean is 100, and the observed mean is (Total losses)/(Total gamblers) = (34*200 + 56*70 + 98*130 + 89*150)/(34+56+98+89) = 132.8880866.

K = EPV/VHM = 392/49 = 8.

N, the number of exposures, is 34 + 56 + 98 + 89 = 277.

Thus, Z = 277/(277+8) = 0.9719298246.

Hence, expected losses per gambler in year 5 are estimated to be

132.8880866*0.9719298246 + 100(1-0.9719298246) = **131.9649123.**

Problem S4C70-4. Similar to Question 139 of the Exam C Sample Questions from the Society of Actuaries. Members of three classes of birds, I, II, and III, can lay either 0, 1, or 2 eggs per year, with the following probabilities:

Class I: Pr(0 eggs) = 0.54; Pr(1 egg) = 0.34; Pr(2 eggs) = 0.12.
Class II: Pr(0 eggs) = 0; Pr(1 egg) = 0.5; Pr(2 eggs) = 0.5.
Class III: Pr(0 eggs) = 0.4; Pr(1 egg) = 0.16; Pr(2 eggs) = 0.44.

A class is chosen at random (with probability 1/3), and members of the class are observed for three years.

In year 1, 90 birds were observed to have laid 40 eggs.
In year 2, 50 birds were observed to have laid 80 eggs.
In year 3, 100 birds were observed to have laid 90 eggs.

Use Bühlmann-Straub credibility to estimate the number of eggs 40 birds can be expected to lay in year 4.

Solution S4C70-4.

For an individual bird:
E(Eggs | Class I) = 0.54*0 + 0.34*1 + 0.12*2 = 0.58 eggs.
Var(Eggs | Class I) = $0.54*(0-0.58)^2 + 0.34*(1-0.58)^2 + 0.12*(2-0.58)^2$ = 0.4836
E(Eggs | Class II) = 0.5*1 + 0.5*2 = 1.5 eggs.
Var(Eggs | Class II) = $0.5*(1-1.5)^2 + 0.5*(2-1.5)^2$ = 0.25.
E(Eggs | Class III) = 0.4*0 + 0.16*1 + 0.44*2 = 1.04 eggs.
Var(Eggs | Class III) = $0.4*(0-1.04)^2 + 0.16*(1-1.04)^2 + 0.44*(2-1.04)^2$ = 0.8384.
EPV = (0.4836 + 0.25 + 0.8384)/3 = 0.524.
E(Total) = Prior mean = (0.58 + 1.5 + 1.04)/3 = 1.04 eggs.
E(Total^2) = $(0.58^2 + 1.5^2 + 1.04^2)/3$ = 1.2226666667.
VHM = 1.2226666667 - 1.04^2 = 0.14106666667.
K = EPV/VHM = 3.71455766.
N is the total number of exposures, which is 90+50+100 = 240.
Thus, Z = 240/(240 + 3.71455766) = 0.9847585806.
Observed mean = (40 + 80 + 90)/240 = 0.875.

Expected eggs per bird in year 4 = 0.9847585806*0.875 + (1-0.9847585806)*1.04 = 0.8775148342.
Total expected eggs for 40 birds = 40*0.8775148342 = **35.10059337 eggs.**

Problem S4C70-5. Similar to Question 263 of the Exam C Sample Questions from the Society of Actuaries. The number of giant flies encountered per month in any given neighborhood follows a Poisson distribution with mean λ. The fly frequencies for each neighborhood are independent. The prior distribution of λ is Weibull with θ = 0.3 and τ = 2.

In month 1, you observe 20 neighborhoods and 4 total flies.
In month 2, you observe 90 neighborhoods and 7 total flies.
In month 3, you observe 800 neighborhoods and 80 total flies.
In month 4, you observe 200 neighborhoods and 10 total flies.

Use Bühlmann-Straub credibility to estimate the expected number of giant flies in 100 neighborhoods during the next 36 months.

Relevant information about Weibull distributions:

E(X) = θ*Γ(1 + 1/τ); E(X^2) = $θ^2$*Γ(1 + 2/τ).

We also know that Γ(0.5) = 1.77245; Γ(1) = 1; Γ(1.5) = 0.88623; Γ(2) = 1.

Solution S4C70-5. First, we find a per-neighborhood per-month estimate of the number of flies.
The prior mean is θ*Γ(1 + 1/τ) = 0.3*Γ(1 +1/2) = 0.3*0.88623 = 0.265869.
The observed mean is (4+7+80+10)/(20+90+800+200) = 0.090990991.
The number of exposures is 20+90+800+200 = N = 1110.
Since the model distribution is Poisson, EPV is the mean of the prior distribution or 0.265869.
VHM is the variance of the prior distribution, i.e., E(X^2) - E(X)^2 =
0.3^2*Γ(1 + 2/2) - 0.265869^2 = 0.3^2 - 0.265869^2 = VHM = 0.0193136748.
K = EPV/VHM = 0.265869/0.0193136748 = 13.76584219.
Z = 1110/(1110 + 13.76584219) = Z = 0.9877502575.
Thus, the per-neighborhood per-month estimate is
0.090990991*0.9877502575 + 0.265869*(1-0.9877502575) = 0.0931334588 flies.
We multiply this value by (number of neighborhoods)*(number of months) to get
0.0931334588*100*36 = **335.2804518 giant flies.**

Section 71

Distribution and Probability Density Functions for Modified Data, Difference Functions, and the Kolmogorov-Smirnov Test

Let us say that there is a distribution of random variable X with cumulative distribution function (cdf) F(x) and probability density function (pdf) f(x). We have some empirical data about X, and the observations are left truncated at some point t. Then we can say the following about the cdf and pdf of the modified data, which we shall call F*(x) and f*(x), respectively:

$F^*(x) = 0$ if $x < t$;

$F^*(x) = (F(x) - F(t))/(1-F(t))$ if $x \geq t$.

$f^*(x) = 0$ if $x < t$;

$f^*(x) = f(x)/(1-F(t))$ if $x \geq t$.

We recall that, for a sample size n, $F_n(x)$ is the empirical distribution function. Let F*(x) be the model distribution to which we are trying to fit the data. This distribution may or may nor involve truncation or censoring.We define the **difference function** D(x) as $D(x) = F_n(x) - F^*(x)$.

A **probability plot** or **p-p plot** plots the value of $F_n(x_j)$ on the horizontal axis against $F^*(x_j)$ on the vertical axis for each observed value x_j. The closer the data are to the 45-degree (slope = 1, intercept at the origin) line of the plot, the better the fit of the model distribution F*(x).

The **Kolmogorov-Smirnov test** measures the goodness of fit of the model to the data. The following are the general hypotheses being tested:
H_0: The data came from a population with the stated model.

H_1: The data did not come from a population with the stated model.

Let t be the point of left truncation (t = 0 if no truncation exists). Let u be the point of right censoring ($u = \infty$ if no censoring exists). The Kolmogorov-Smirnov test statistic is

$D = \max_{t\leq x\leq u} \left| F_n(x) - F^*(x) \right| = \max_{t\leq x\leq u} \left| D(x) \right|$.

One can think of the Kolmogorov-Smirnov test statistic as the maximum of the absolute value of the difference function, with a few caveats. The test assumes that F*(x) is continuous over the applicable range of data and that the empirical function $F_n(x)$ is well-defined (i.e., individual empirical data are actually used). For each value of x within the observed data set, one needs to calculate both $F_n(x)$ and $F_n(x^-)$, which is the empirical distribution function for all values *up to x but excluding x*. Then, for each x, the maximum of $|F_n(x) - F^*(x)|$ and the maximum $|F_n(x^-) - F^*(x)|$ are each considered to be possible candidates for D.

Let n be the sample size. Then the following critical values are used for the test:

For significance level $\alpha = 0.10$, a critical value of $1.22/\sqrt{(n)}$ is used.

For significance level $\alpha = 0.05$, a critical value of $1.36/\sqrt{(n)}$ is used.

For significance level $\alpha = 0.01$, a critical value of $1.63/\sqrt{(n)}$ is used.

Let c be the critical value for a Kolmogorov-Smirnov test. If $D < c$, then we do not reject the null hypothesis H_0. If $D \geq c$, then we reject the null hypothesis H_0.

Source: *Loss Models: From Data to Decisions*, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 16, pp. 441-450.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C71-1. You observe eight different data points: 2, 2, 3, 4, 4, 4, 4, 5.

You are trying to fit the data to an exponential distribution with mean 3.5.

What is the difference function for this model distribution?

Solution S4C71-1. We use the formula $D(x) = F_n(x) - F^*(x)$.

We find the empirical distribution function $F_n(x)$ by inspecting our sample.

We find that n = 8, Pr(2) = ¼, Pr(3) = 1/8, Pr(4) = ½, and Pr(5) = 1/8.

Thus, $F_8(x) = 0$ if $x < 2$;

$F_8(x) = 1/4$ if $2 \leq x < 3$;

$F_8(x) = ¼ + 1/8 = 3/8$ if $3 \leq x < 4$;
$F_8(x) = 3/8 + ½ = 7/8$ if $4 \leq x < 5$;

$F_8(x) = 1$ if $5 \leq x$.

Our model distribution is exponential with mean 3.5, so $F^*(x) = 1 - e^{-x/3.5}$ for all $x > 0$.

Thus, $D(x) = 0 - (1 - e^{-x/3.5})$ = **$D(x) = e^{-x/3.5} - 1$ if $0 < x < 2$**;

$D(x) = ¼ - (1 - e^{-x/3.5})$ = **$D(x) = e^{-x/3.5} - ¾$ if $2 \leq x < 3$**;

$D(x) = 3/8 - (1 - e^{-x/3.5}) = \mathbf{D(x) = e^{-x/3.5} - 5/8 \text{ if } 3 \leq x < 4}$;

$D(x) = 7/8 - (1 - e^{-x/3.5}) = \mathbf{D(x) = e^{-x/3.5} - 1/8 \text{ if } 4 \leq x < 5}$;

$D(x) = 1 - (1 - e^{-x/3.5}) = \mathbf{D(x) = e^{-x/3.5} \text{ if } 5 \leq x}$.

Problem S4C71-2. You observe eight different data points: 2, 2, 3, 4, 4, 4, 4, 5.

You are trying to fit the data to an exponential distribution with mean 3.5.

To test the goodness of fit for the model, you conduct a Kolmogorov-Smirnov test at the 0.05 significance level. Determine the Kolmogorov-Smirnov test statistic D.

Solution S4C71-2. We use the formula $D = \max_{t \leq x \leq u} |D(x)|$. Since there is no truncation or censoring, and since we are required to only consider actual empirical data, the only relevant values to compare are D(2), D(3), D(4), and D(5). We refer to our general solution for D(x) in Solution S4C71-1.

For x = 2, $D(x) = e^{-x/3.5} - ¾$, so $D(2) = e^{-2/3.5} - ¾ = D(2) = -0.185281878$.

For x = 3, $D(x) = e^{-x/3.5} - 5/8 = e^{-3/3.5} - 5/8 = D(3) = -0.20062716543$.

For x = 4, $D(x) = e^{-x/3.5} - 1/8 = e^{-4/3.5} - 1/8 = D(4) = 0.1939065573$.

For x = 5, $D(x) = e^{-x/3.5} = e^{-5/3.5} = D(5) = 0.2396510364$.

For each x, we also need to find $D(x^-) = F_n(x^-) - F^*(x)$.

For x = 2, $D(x^-) = 0 - (1 - e^{-2/3.5}) = -0.435281878$.

For x = 3, $D(x^-) = 1/4 - (1 - e^{-3/3.5}) = -0.3256271543$.

For x = 4, $D(x^-) = 3/8 - (1 - e^{-4/3.5}) = -0.3060934427$.

For x = 5, $D(x^-) = 7/8 - (1 - e^{-5/3.5}) = 0.1146510364$.

As can be seen from the above results, $|D(x)|$ is maximized by $|D(5)| = 0.2396510364$, and

$|D(x^-)|$ is maximized by $|D(2^-)| = 0.435281878$. Thus, **D = 0.435281878**.

Problem S4C71-3. You observe eight different data points: 2, 2, 3, 4, 4, 4, 4, 5.

You are trying to fit the data to an exponential distribution with mean 3.5.

To test the goodness of fit for the model, you conduct a Kolmogorov-Smirnov test at the 0.05 significance level. Do you reject or fail to reject the null hypothesis that the data came from a population having the exponential distribution above? Perform the test and justify your answer.

Solution S4C71-3. Our sample size is n = 8.

From Solution S4C71-2, we know that our test statistic is D = 0.435281878. We compare this to our critical value, which, for $\alpha = 0.05$, is $c = 1.36/\sqrt{(n)} = 1.36/\sqrt{(8)} = 0.4808326112$.

As 0.435281878 < 0.4808326112, we know that **D < c, sowe fail to reject the null hypothesis.**

Problem S4C71-4. You observe eight different data points: 2, 2, 3, 4, 4, 4, 4, 5.

You are trying to fit the data to an exponential distribution with mean 3.5. Now, however, the model distribution has been left-truncated at x = 2. Find the new model distribution function.

Solution S4C71-4. We use the formula

$F^*(x) = 0$ if $x < t$;

$F^*(x) = (F(x) - F(t))/(1-F(t))$ if $x \geq t$.

Here, t = 2, and $F(x) = 1 - e^{-x/3.5}$ for all x > 0.

Thus, for $x \geq 2$, $F^*(x) = ((1 - e^{-x/3.5}) - (1 - e^{-2/3.5}))/(1 - (1 - e^{-2/3.5})) =$

$F^*(x) = ((1 - e^{-x/3.5}) - (1 - e^{-2/3.5}))/(e^{-2/3.5}) = (e^{-2/3.5} - e^{-x/3.5})/(e^{-2/3.5}) = 1 - e^{-(x-2)/3.5}$.

Therefore $\mathbf{F^*(x) = 0}$ **if** $\mathbf{x < 2}$**;**

$\mathbf{F^*(x) = 1 - e^{-(x-2)/3.5}}$ **if** $\mathbf{x \geq 2}$**.**

Problem S4C71-5. You observe eight different data points: 2, 2, 3, 4, 4, 4, 4, 5.

You are trying to fit the data to an exponential distribution with mean 3.5. Now, however, the model distribution has been left-truncated at x = 2. To test the goodness of fit for the model, you conduct a Kolmogorov-Smirnov test at the 0.10 significance level. Do you reject or fail to reject the null hypothesis that the data came from a population having the left-truncated exponential distribution above? Perform the test and justify your answer.

Solution S4C71-5. We note that the empirical function is left unchanged, since the model distribution was all that was truncated, and for every point of available empirical data, the difference function has only changed from the one in Solution S4C71-1 in that $e^{-x/3.5}$is now replaced by $e^{-(x-2)/3.5}$. Therefore,

For x = 2, $D(x) = e^{-(x-2)/3.5}$- ¾, so $D(2) = e^{-0/3.5}$- ¾ = D(2) = 0.25.

For x = 3, $D(x) = e^{-(x-2)/3.5} - 5/8 = e^{-1/3.5} - 5/8 = D(3) = 0.1264772931$.

For x = 4, $D(x) = e^{-(x-2)/3.5} - 1/8 = e^{-2/3.5} - 1/8 = D(4) = 0.439718122$.

For x = 5, $D(x) = e^{-(x-2)/3.5} = e^{-3/3.5} = D(5) = 0.4243728457$.

For each x, we also need to find $D(x^-) = F_n(x^-) - F^*(x)$.

For x = 2, $D(x^-) = 0 - (1 - e^{-0/3.5}) = 0$

For x = 3, $D(x^-) = 1/4 - (1 - e^{-1/3.5}) = 0.0014772931$.

For x = 4, $D(x^-) = 3/8 - (1 - e^{-2/3.5}) = -0.060281878$.

For x = 5, $D(x^-) = 7/8 - (1 - e^{-3/3.5}) = 0.2993728457$.

Our Kolmogorov-Smirnov test is the maximum of the values of $|D(x)|$ and $|D(x^-)|$, which is D(4) = 0.439718122.

Our sample size is n = 8. Our significance level is α = 0.10, so the critical value c = 1.22/√(n) =

1.22/√(8) = 0.4313351365. We note that 0.439718122 > 0.4313351365, so **D > c, and we reject the null hypothesis**.

Section 72

Exam-Style Questions on the Kolmogorov-Smirnov Test and Bühlmann Credibility

This section offers additional exam-style practice on the Kolmogorov-Smirnov test and Bühlmann credibility.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C72-1. Similar to Question 40 of the Exam C Sample Questions from the Society of Actuaries.

A sample of 3 values is given: 111, 234, 555.

You attempt to fit the sample data to an exponential distribution, whose mean is obtained from the sample data via the method of moments.

Find the Kolmogorov-Smirnov test statistic for this model.

Solution S4C72-1. The method of moments estimate is the sample mean: (111 + 234 + 555)/3 = 300. Thus, the model distribution $F^*(x) = 1 - e^{-x/300}$. Each data point of the sample has empirical probability (1/3) of occurring, so $F_3(111) = (1/3)$, $F_3(234) = (2/3)$, $F_3(555) = 1$.

However, we also need to consider the values of $F_3(x)$ immediately prior to each of these data points. These values can be denoted as $F_3(x^-)$.

$F_3(111^-) = 0$; $F_3(234^-) = 1/3$; $F_3(555^-) = 2/3$.

Our test statistic D is the maximum of either of the following quantities: $|F_3(x) - F^*(x)|$ or $|F_3(x^-) - F^*(x)|$ for one of the three x values from the sample.

For x = 111: $|F_3(x) - F^*(x)| = |1/3 - (1 - e^{-111/300})| = 0.024067764$.

$|F_3(x^-) - F^*(x)| = |0 - (1 - e^{-111/300})| = 0.3092656694$.

For x = 234: $|F_3(x) - F^*(x)| = |2/3 - (1 - e^{-234/300})| = 0.125072678$.

$|F_3(x^-) - F^*(x)| = |1/3 - (1 - e^{-234/300})| = 0.2082606554$.

For x = 555: $|F_3(x) - F^*(x)| = |1 - (1 - e^{-555/300})| = 0.1572371663$.

$|F_3(x^-) - F^*(x)| = |2/3 - (1 - e^{-555/300})| = 0.176096167$.

The highest of the above values is our test statistic D. The answer is thus **D = 0.3092656694**.

Problem S4C72-2. Similar to Question 160 of the Exam C Sample Questions from the Society of Actuaries. The following is a sample of data: 0.3, 0.6, 0.9.

You attempt to fit the sample data to a distribution with the following probability density function: $f^*(x) = 6/(1+x)^7$ for x > 0.

Find the Kolmogorov-Smirnov test statistic for this model.

Solution S4C72-2. First, we find the model distribution function $F^*(x) = \int f^*(x)dx = \int(6/(1+x)^7)dx = F^*(x) = 1 - (1+x)^{-6}$ for x > 0.

Each data point of the sample has empirical probability (1/3) of occurring, so $F_3(0.3) = (1/3)$, $F_3(0.6) = (2/3)$, $F_3(0.9) = 1$.

However, we also need to consider the values of $F_3(x)$ immediately prior to each of these data points. These values can be denoted as $F_3(x^-)$.

$F_3(0.3^-) = 0$; $F_3(0.6^-) = 1/3$; $F_3(0.9^-) = 2/3$.

Our test statistic D is the maximum of either of the following quantities: $|F_3(x) - F^*(x)|$ or $|F_3(x^-) - F^*(x)|$ for one of the three x values from the sample.

For x = 0.3: $|F_3(x) - F^*(x)| = |1/3 - (1 - (1+0.3)^{-6})| = 0.4594904556$.

$|F_3(x^-) - F^*(x)| = |0 - (1 - (1+0.3)^{-6})| = 0.792823789$.

For x = 0.6: $|F_3(x) - F^*(x)| = |2/3 - (1 - (1+0.6)^{-6})| = 0.2737286886$.

$|F_3(x^-) - F^*(x)| = |1/3 - (1 - (1+0.6)^{-6})| = 0.6070620219$.

For x = 0.9: $|F_3(x) - F^*(x)| = |1 - (1 - (1+0.9)^{-6})| = 0.021255846$.

$|F_3(x^-) - F^*(x)| = |2/3 - (1 - (1+0.9)^{-6})| = 0.3120774874$.

The highest of the above values is our test statistic D. The answer is thus **D = 0.792823789.**

Problem S4C72-3. Similar to Question 172 of the Exam C Sample Questions from the Society of Actuaries. The following is a sample of data: 0.3, 0.6, 0.9.

You use the Kolmogorov-Smirnov test to evaluate the null hypothesis H_0 that the sample data comes from a population whose distribution has the following probability density function: $f^*(x) = 6/(1+x)^7$ for $x > 0$.

The following are the critical values c for various significance levels α:

$\alpha = 0.10 \rightarrow c = 1.22/\sqrt{(n)}$

$\alpha = 0.05 \rightarrow c = 1.36/\sqrt{(n)}$

$\alpha = 0.025 \rightarrow c = 1.48/\sqrt{(n)}$

$\alpha = 0.01 \rightarrow c = 1.63/\sqrt{(n)}$

Determine the result of the test.

(A) Do not reject H_0 at the 0.10 significance level.

(B) Reject H_0 at the 0.10 significance level, but not at the 0.05 significance level.

(C) Reject H_0 at the 0.05 significance level, but not at the 0.025 significance level.

(D) Reject H_0 at the 0.025 significance level, but not at the 0.01 significance level.

(E) Reject H_0 at the 0.01 significance level.

Solution S4C72-3. We know from Solution S4C72-2 that the test statistic D = 0.792823789.

Sample size n = 3. We find the critical value for each significance level.

$\alpha = 0.10 \rightarrow c = 1.22/\sqrt{(3)} = 0.7043673284$.

$\alpha = 0.05 \rightarrow c = 1.36/\sqrt{(3)} = 0.7851963661$.

$\alpha = 0.025 \rightarrow c = 1.48/\sqrt{(3)} = 0.8544783984$.

$\alpha = 0.01 \rightarrow c = 1.63/\sqrt{(3)} = 0.9410809388$.

Since $0.7851963661 < D < 0.8544783984$, we reject H_0 for $\alpha = 0.05$ and do not reject H_0 for $\alpha = 0.025$. Thus, the answer is **(C) Reject H_0 at the 0.05 significance level, but not at the 0.025 significance level**.

Problem S4C72-4. Similar to Question 8 of the Exam C Sample Questions from the Society of Actuaries. Frequency of losses follows a Poisson distribution with mean θ. Severity of losses follows an exponential distribution with mean 4θ. Given θ, frequency and severity are independent. The prior distribution of θ is $\pi(\theta) = 7\theta^{-8}$ for θ > 1. Find Bühlmann's K for aggregate losses.

Relevant properties for exponential distributions: $E(X^2) = 2\theta^2$.

Solution S4C72-4. K = EPV/VHM. Let S denote aggregate losses. Let N denote frequency, and let X denote severity.

$E(S \mid \theta) = E(N \mid \theta)*E(X \mid \theta) = \theta*4\theta = 4\theta^2$.

Since N is Poisson-distributed, $Var(S \mid \theta) = E(N \mid \theta)*E(X^2 \mid \theta) = \theta*(2*(4\theta)^2) = 32\theta^3$.

$EPV = E(Var(S \mid \theta)) = {}_1{}^{\infty}\int(32\theta^3)(7\theta^{-8})d\theta = {}_1{}^{\infty}\int(224\theta^{-5})d\theta = 224/4 = EPV = 56$.

$VHM = Var(E(S \mid \theta)) = Var(4\theta^2) = E(16\theta^4) - E(4\theta^2)^2 =$

${}_1{}^{\infty}\int(16\theta^4)(7\theta^{-8})d\theta - ({}_1{}^{\infty}\int(4\theta^2)(7\theta^{-8})d\theta)^2 = {}_1{}^{\infty}\int(112\theta^{-4})d\theta - ({}_1{}^{\infty}\int((28\theta^{-6})d\theta)^2 = 112/3 - (28/5)^2 = VHM$ = 5.9733333333. Thus, K = EPV/VHM = 56/5.9733333333 = **K = 9.375**.

Problem S4C72-5. Similar to Question 41 of the Exam C Sample Questions from the Society of Actuaries. Frequency of losses per year has mean λ and variance σ^2. The prior distribution of λ is uniform from 3 to 6. The prior distribution of σ^2 is exponential with mean 9. An individual was observed in year 1 to have 4 losses. Use Bühlmann credibility to estimate this individual's number of losses in year 2.

Relevant properties for uniform distributions: $Var(X) = (b-a)^2/12$.

Solution S4C72-5. We find K = EPV/VHM. Let N be the frequency of losses.

EPV = E(Var(N)). Since $Var(N) = \sigma^2$, $EPV = E(\sigma^2) = 9$.

VHM = Var(E(N)). Since E(N) = λ, $VHM = Var(\lambda) = (6-3)^2/12 = 9/12$ = VHM = ¾. K = 9/(3/4) = 12. Since one person was observed for one year, we have N = 1, so Z = 1/(1+12) = 1/13.

The prior mean is E(λ) = (6+3)/2 = 4.5, so our answer is

(1/13)*4 + (12/13)*4.5 = **4.461538462 losses.**

Section 73

Nonparametric and Semiparametric Empirical Bayes Estimation of Bühlmann And Bühlmann-Straub Credibility Factors

The most complex topic on the Exam 4/C syllabus pertaining to Bühlmann and Bühlmann-Straub credibility is empirical Bayes estimation of credibility factors. Empirical Bayes estimation is applied when there are multiple risks (e.g. different insureds or different groups of policyholders) and multiple exposures per risk (e.g. multiple years or other time periods of observations available for each risk). The situation is further complicated if some risks have different numbers of exposures from other risks.

There are two types of empirical Bayes estimation on the syllabus: nonparametric and semiparametric. Both are presented formally in Curtis Gary Dean's "Topics in Credibility Theory." This study guide will take a somewhat different approach in emphasizing easy remembering and application of each of these methods of estimation in the circumstances to which they might apply. The present section will be a relatively gentle introduction, while the subsequent section will focus on exam-style questions.

First, some notation that applies to each of the methods below:

R = number of risks;
N = number of exposures - which could be particular to a risk or could be overall, depending on the context in which it is used. If necessary, subscripts will clarify any distinctions.

EPV^ = estimate of the expected value of the process variance.

VHM^ = estimate of the variance of hypothetical means.

K = EPV^/VHM^ = Bühlmann's K, arising out of estimates of EPV and VHM.

Z = N/(N+K) = the credibility factor applying to the observed mean. N is typically the number of exposures of the risk whose data is being credibility-weighted with the overall mean.

Credibility-weighted estimate = Z*E(R_i) + (1-Z)*E(Total), where E(R_i) is the observed mean of the ith risk.

Nonparametric Bayes Estimation - Bühlmann Credibility Model

Let R_i denote the ith risk.

Let SampleVar(R_i) denote the sample variance for the ith risk.

Crucial note: The sample variance, unlike the population variance, is calculated by taking the sum of the squared deviations of each observation about the mean and dividing it by *the sample size minus one*, not the whole sample size. That is, SampleVar(R_i) = $_j\Sigma(X_j - E(R_i))^2/(N-1)$, where the X_j are the observations in the sample.

There are two possible formulas for EPV^:

EPV^ = ($_i\Sigma$(SampleVar(R_i))/R): this is the average of the sample variances for each risk.

EPV^ =$((1/(R(N-1))\ {}_{i=1}{}^{R}\Sigma({}_{t=1}{}^{N}\Sigma((X_{it} - E(R_i))^2))$: this is the sum of all the deviations of each observation about the mean of the risk to which it pertains, divided by R(N-1).

If N is the same for each risk:

VHM^ = $(1/(R-1))*{}_i\Sigma(E(R_i) - E(\text{Total}))^2$ -EPV^/N.

The quantity $(1/(R-1))*{}_i\Sigma(E(R_i) - E(\text{Total}))^2$ can be considered the total sample variance, so VHM^ = Total Sample Variance - EPV^/N.

It is possible for this method to produce VHM^ < 0 some data sets. In this case, we simply assume that VHM = 0 and that the data under consideration has no credibility.

Nonparametric Bayes Estimation - Bühlmann-Straub Credibility Model

In the Bühlmann-Straub model, it is possible for each risk to have a different number of exposures and, moreover, to be monitored over different time periods. This changes the calculation of EPV^ and VHM^.

First, we designate weights $w_i = (N_i-1)/{}_j\Sigma(N_j-1)$. Each weight w_i is the number of exposures for the ith risk minus one, divided by the sum of the (number of exposures for each risk minus one).

Then EPV^ = $_i\Sigma(w_i$*SampleVar(R_i)) = $_i\Sigma(N_{it}*(X_{it} - E(R_i))^2/{}_j\Sigma(N_j-1)$, where N_{it} is the number of times each observation X_{it} occurs.

The expression for VHM^ is cumbersome to express at once, so we can split it into two components:

VHM Numerator = $_i\Sigma(N_i*(E(R_i) - E(\text{Total}))^2) - (R-1)$EPV^.

VHM Denominator = $N_{total} - (1/N_{total})*{}_i\Sigma(N_i{}^2)$.

VHM^ = VHM Numerator/VHM Denominator.

K = EPV^/ VHM^.

For the ith risk, the credibility factor is $Z_i = N_i/(N_i + K)$.

What is the External Mean?

In empirical Bayes estimation, what is the external mean by which the complement of credibility (1-Z) is multiplied? There are two options.

1) Use E(Total), the mean of all observations. This is the most common approach.

2) Use a credibility-weighted average of the means of all the individual risks. This value is M = ${}_i\Sigma(Z_i*E(R_i))/{}_i\Sigma(Z_i)$. This is called the method that "preserves total losses" or any other total quantity under consideration.

Semiparametric Bayes Estimation

The procedure for semiparametric empirical Bayes estimation is the same as for nonparametric empirical Bayes estimation, except with the simplifying assumption that the value per exposure follows a common probability distribution. If the value per exposure follows a Poisson distribution, then EPV^ = E(Total). VHM^ is calculated in the same manner as for nonparametric estimation and depends on whether the Bühlmann or the Bühlmann-Straub model is being used.

Source:
Dean, Curtis Gary, "Topics in Credibility Theory," 2004 (SOA Study Note).

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C73-1. You are given the following data:

Risk 1 has 4 claims in year 1, 3 claims in year 2, 6 claims in year 3, and 4 claims in year 4.

Risk 2 has 4 claims in year 1, 4 claim in year 2, 4 claims in year 3, and 6 claims in year 4.

Use nonparametric Bayes estimation within the Bühlmann credibility model to find EPV^, the estimate of the expected value of the process variance.

Solution S4C73-1. We use the formula EPV^ = $({}_i\Sigma(\text{SampleVar}(R_i))/R)$. Here, R = 2, and N = 4 for each R.

For Risk 1: $E(R_1) = (4 + 3 + 6 + 4)/4 = 4.25$

$\text{SampleVar}(R_1) = (2*(4-4.25)^2 + (3-4.25)^2 + (6-4.25)^2)/(4-1) = 1.583333333$.

For Risk 2: $E(R_2) = (4 + 4 + 4 + 6)/4 = 4.5$

$\text{SampleVar}(R_2) = (3*(4-4.5)^2 + (6-4.5)^2)/(4-1) = 1.08333333333$.

EPV^ = (SampleVar(R_1) + SampleVar(R_2))/2 = 2.6666666667/2 = **EPV^ = 1.33333333333**.

Problem S4C73-2. You are given the following data:

Risk 1 has 4 claims in year 1, 3 claims in year 2, 6 claims in year 3, and 4 claims in year 4.

Risk 2 has 4 claims in year 1, 1 claim in year 2, 4 claims in year 3, and 9 claims in year 4.

Use nonparametric Bayes estimation within the Bühlmann credibility model to find VHM^, the estimate of the variance of hypothetical means.

Solution S4C73-2. We use the formula VHM^ = $(1/(R-1))*{}_i\Sigma(E(R_i) - E(Total))^2$ -EPV^/N.

We know from Solution S4C73-1 that $E(R_1)$ = 4.25 and $E(R_2)$ = 4.5, implying that E(Total) = (4.25+4.5)/2 = 4.375.

Thus, ${}_i\Sigma(E(R_i) - E(Total))^2 = (4.25-4.375)^2 + (4.5-4.375)^2$ = 0.03125. Since R = 2, (R-1) = 1. Also, from Solution S4C73-1, EPV = 1.333333333. As N = 4, VHM^ = 0.03125 - 1.333333333/4 = **VHM^ = -0.3177083333333**. In this case, we would assume that VHM = 0, and the data has no credibility.

Problem S4C73-3. You observe data for an insurance company for one year. There are 90 insureds with no claims, 6 insureds with one claim, and 4 insureds with two claims. The number of claims for each insured is Poisson-distributed, with an unknown mean, and each insured has the same expected claim frequency. Use semiparametric Bayes estimation within the Bühlmann model to calculate VHM^, the estimate of the variance of hypothetical means.

Solution S4C73-3.

Here, R = 100 and N = 1, since 100 insureds are observed for 1 year.

There are 100 insureds, so Pr(0 claims) = 0.9, Pr(1 claim) = 0.06, Pr(2 claims) = 0.04. Thus, E(Total) = 0*0.9 + 1*0.06 + 2*0.04 = EPV^ = 0.14.

VHM = $(1/(R-1))*{}_i\Sigma(E(R_i) - E(Total))^2$ -EPV^/N.

There are 100 risks, so (R-1) = 99.

${}_i\Sigma(E(R_i) - E(Total))^2 = 90(0 - 0.14)^2 + 6(1 - 0.14)^2 + 4(2 - 0.14)^2$ = 20.04.

Thus, VHM^ = 20.04/99 - 0.14/1 = **VHM^ = 0.062424242424**.

Problem S4C73-4. You observe data for an insurance company for one year. There are 90 insureds with no claims, 6 insureds with one claim, and 4 insureds with two claims. The number of claims for each insured is Poisson-distributed, with an unknown mean, and each insured has the same expected claim frequency. An individual insured has one claim this year. Use semiparametric Bayes estimation within the Bühlmann model to calculate this insured's expected claims next year.

Solution S4C73-4. Here, N = 1, and, from Solution S4C73-3, EPV^ = 0.14 and VHM^ = 0.062424242424. Thus, K = 0.14/0.062424242424 = 2.242718447 and Z = 1/(1+2.242718447) = 0.3083832335. The external mean is E(Total) = 0.14, and the observed mean is 1. Thus, our answer is 1*(0.3083832335) + 0.14(1-0.3083832335) = **0.4052095808 claims.**

Problem S4C73-5. There are two groups of elephants: A and B. There are 90 elephants in group A, who ate a mean number of 9 peanuts per elephant during the period of observation. There are 70 elephants in group B, who ate a mean number of 10 peanuts per elephant during the period of observation. The expected value of the process variance is known to be 30. Use nonparametric Bayes estimation within the Bühlmann-Straub credibility model to find VHM^, the estimate of the variance of hypothetical means.

Solution S4C73-5.

We use the following formulas:

VHM Numerator = $_i\Sigma(N_i*(E(R_i) - E(Total))^2) - (R-1)EPV^\wedge$.

VHM Denominator = $N_{total} - (1/N_{total})*_i\Sigma(N_i^2)$.

VHM^ = VHM Numerator/VHM Denominator.

We are given that R = 2, so (R - 1) = 1. Also, EPV^ = 30 and

E(Total) = (90*9 + 70*10)/(90 + 70) = 9.4375, and

$_i\Sigma(N_i*(E(R_i) - E(Total))^2) = 90(9-9.4375)^2 + 70(10-9.4375)^2 = 39.375$.

Thus, VHM Numerator = 39.375 - 30 = 9.375.

$N_{total} = 90 + 70 = 160$, so

VHM Denominator = $160 - (1/160)(90^2 + 70^2) = 78.75$.

Hence, VHM^ = 9.375/78.75 = **VHM^ = 0.119047619**.

Section 74

Exam-Style Questions on Nonparametric Empirical Bayes Estimation of Bühlmann and Bühlmann-Straub Credibility Factors

This section provides exam-style practice with nonparametric Bayes estimation of Bühlmann and Bühlmann-Straub credibility factors.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Sources:
Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C74-1. Similar to Question 38 of the Exam C Sample Questions from the Society of Actuaries. You have access to data on number of peanuts eaten by 12 elephants for 19 days. Here is what you know: $\sum_{i=1}^{12}(\sum_{j=1}^{19}(X_{ij}-\bar{X}_i)^2) = 900$.

$\sum_{i=1}^{12}(\bar{X}_i - \bar{X})^2 = 10$. Use nonparametric empirical Bayes estimation to find the Bühlmann credibility factor for an individual elephant.

Solution S4C74-1. Here, N = 19 - i.e., the number of days, and R = 12 - i.e., the number of elephants.

EPV^ = $\sum_{i=1}^{12}(\sum_{j=1}^{19}(X_{ij}-\bar{X}_i)^2)/(R*(N-1))$ = 900/(12*18) = EPV^ = 4.16666667.

VHM^ = $(1/(R-1))\sum_{i=1}^{12}(\bar{X}_i - \bar{X})^2$ - EPV^/N = 10/11 - 4.16666667/19 = 0.6897926635.

K = EPV^/VHM^ = 4.16666667/0.6897926635 = K = 6.040462428.

Since N = 19, Z = 19/(19+6.040462428) = **Z = 0.7587719298.**

Problem S4C74-2. Similar to Question 12 of the Exam C Sample Questions from the Society of Actuaries. You are given information for the number of peanuts eaten by 2 elephants,

A and B, over 3 years:
A ate 97 peanuts in year 1, 65 peanuts in year 2, and 60 peanuts in year 3.

B ate 45 peanuts in year 1, 70 peanuts in year 2, and 65 peanuts in year 3.

Use nonparametric empirical Bayes estimation to find the Bühlmann credibility estimate of the number of peanuts elephant B will eat in year 4.

Solution S4C74-2. Here, N = 3 - i.e., the number of years, and R = 2 - i.e., the number of elephants. Let P be the number of peanuts eaten.

For A, E(P│A) = (97 + 65 + 60)/3 = 74.

SampleVar(P│A) = $((97-74)^2 + (65-74)^2 + (60-74)^2)/(3-1) = 403$.

For B, E(P│B) = (45 + 70 + 65)/3 = 60.

SampleVar(P│B) = $((45-60)^2 + (70-60)^2 + (65-60)^2)/(3-1) = 175$.

EPV^ = (SampleVar(P│A) + SampleVar(P│B))/2 = EPV^ = 289.

E(Total) = (74 + 60)/2 = 67.

SampleVar(Total) = $(1/(1-1))((74-67)^2 + (60-67)^2) = 98$.

VHM^ = SampleVar(Total) - EPV^/N = 98 - 289/3 = VHM^ = 5/3.

K = EPV^/VHM^ = 289/(5/3) = 173.4.

Thus, Z = 3/(3+173.4) = 0.0170068027.

For elephant B, the Bühlmann credibility estimate is

0.0170068027*(60) + (1-0.0170068027)*67 = **66.88095238 peanuts.**

Problem S4C74-3. Similar to Question 145 of the Exam C Sample Questions from the Society of Actuaries. You are given information for car-related losses in three neighborhoods.

Neighborhood A: In year 1, there were 30 cars suffering a total of 400 in losses.
In year 2, there were 50 cars suffering a total of 500 in losses.

Neighborhood B: In year 1, there were 60 cars suffering a total of 200 in losses.
In year 3, there were 80 cars suffering a total of 800 in losses.

Neighborhood C: In year 2, there were 40 cars suffering a total of 200 in losses.
In year 3, there were 40 cars suffering a total of 100 in losses.

Use nonparametric empirical Bayes estimation to determine the Bühlmann-Straub credibility factor Z for Neighborhood C.

Solution S4C74-3.

Here, R = 3.

First, we can find the total number of exposures N for each neighborhood:

For A, $N_A = 30 + 50 = 80$.

For B, $N_B = 60 + 80 = 140$.

For C, $N_C = 40 + 40 = 80$.

In all, $N_{total} = 80 + 140 + 80 = 300$.

Let L denote losses per car.

Then E(L│A) = (400 + 500)/80 = 11.25.

SampleVar(L│A) = $(30(400/30 - 11.25)^2 + 50(500/50 - 11.25)^2)/(2\text{-}1)$ =

SampleVar(L│A) = 208.333333333.

E(L│B) = (200 + 800)/140 = 7.142857143.

SampleVar(L│B) = $(60(200/60 - 7.142857143)^2 + 80(800/80 - 7.142857143)^2)/(2\text{-}1)$ =

SampleVar(L│B) = 1523.809524.

E(L│C) = (200 + 100)/80 = 3.75.

SampleVar(L│C) = $(40(200/40 - 3.75)^2 + 40(100/40 - 3.75)^2)/(2\text{-}1)$ = SampleVar(L│C) = 25.

EPV^ = (SampleVar(L│A) + SampleVar(L│B) + SampleVar(L│C))/3 = EPV^ = 619.0476191.

E(Total) = (400 + 500 + 200 + 800 + 200 + 100)/300 = 7.3333333333.

VHM^ = (Sum(N_X_i*(E(L│X_i) - 7.3333333333))^2 - (R-1)EPV^)/(N_{total} - (1/N_{total})($N_A^2 + N_B^2 + N_C^2$)), where the X_i are either A, B, or C.

First, we find (N_{total} - (1/N_{total})($N_A^2 + N_B^2 + N_C^2$)) = 300 - (1/300)($80^2 + 140^2 + 80^2$) = 192.

Next, we find (Sum(N_X_i*(E(L│X_i) - 7.3333333333))^2 - (R-1)EPV^) =

$80*(11.25 - 7.3333333333)^2 + 140*(7.142857143 - 7.3333333333)^2 + 80*(3.75 - 7.3333333333)^2 - 2*619.0476191 = 1021.428571$.

VHM^ = 1021.428571/192 = 5.319940477.

K = EPV^/VHM^ = 619.0476191/5.319940477 = K = 116.3636364.

For C, N_C = 80, so Z = 80/(80 + 116.3636364) = **Z = 0.4074074074**.

Problem S4C74-4. Similar to Question 194 of the Exam C Sample Questions from the Society of Actuaries. There are two groups of tarantulas, A and B.

In year 2, group A had 30 tarantulas and an average of 30 webs spun per tarantula. In year 3, group A had 40 tarantulas and an average of 50 webs spun per tarantula. On average, 41.42857143 webs per tarantula were spun by group A.

In year 1, group B had 50 tarantulas and an average of 70 webs spun per tarantula. In year 2, group B had 20 tarantulas and an average of 40 webs spun per tarantula. On average, 61.42857143 webs per tarantula were spun by group B.

The Tarantula Regulation Authority has recently issued an order requiring all analysts who use nonparametric empirical Bayes estimation to determine Bühlmann-Straub credibility factors for tarantulas to use a VHM value of 90, irrespective of its mathematical correctness.

Given this information and constraint, estimate the Bühlmann-Straub credibility factor Z for group A using nonparametric Bayes estimation.

Solution S4C74-4. We are required to use VHM = 90, but we still need to find EPV from the data. SampleVar(A) = $(30*(30 - 41.42857143)^2 + 40*(50 - 41.42857143)^2)/(2-1) = 6857.142857$.

SampleVar(B) = $(50*(70 - 61.42857143)^2 + 20*(40-61.42857143)^2)/(2-1) = 12857.14286$.

EPV = (SampleVar(A) + SampleVar(B))/2 = (6857.142857+12857.14286)/2 = 9857.142857.

Thus, K = EPV^/VHM = 9857.142857/90 = 109.5238095.

For group A, N = 30+40 = 70, so Z = 70/(70+109.5238095) = **Z = 0.3899204244.**

Problem S4C74-5. Similar to Question 223 of the Exam C Sample Questions from the Society of Actuaries.

You are examining group C of tarantulas. In year 1, there were 80 tarantulas in this group, who spun a total of 8900 webs. In year 2, there were 40 tarantulas in this group, who spun a total of 3500 webs.

The Tarantula Regulation Authority has recently issued an order requiring all analysts who use nonparametric empirical Bayes estimation to determine Bühlmann-Straub credibility factors for tarantulas to use a VHM value of 90, irrespective of its mathematical correctness.

Given this information and constraint, estimate the Bühlmann-Straub credibility factor Z for group C in year 3 using nonparametric Bayes estimation.

Solution S4C74-5. We are required to use VHM = 90, but we still need to find EPV from the data. The average number of webs spun is (8900 + 3500)/(80+40) = 103.33333333.

EPV^ = SampleVar(C) = (80*(8900/80-103.33333333)^2 + 40*(3500/40-103.33333333)^2)/(2-1) = 15041.6666667. K = EPV^/VHM = 15041.6666667/90 = 167.1296296.

Here, N = 80 + 40 = 120, so Z = 120/(120 +167.1296296) = **Z = 0.4179297001.**

Section 75

Exam-Style Questions on Semiparametric Empirical Bayes Estimation of Bühlmann and Bühlmann-Straub Credibility Factors and the Chi-Square Goodness of Fit Test

This section provides exam-style practice with semiparametric empirical Bayes estimation of Bühlmann and Bühlmann-Straub credibility factors.

A tangentially related matter is maximum likelihood estimation of a Poisson parameter λ, which is also the sample mean - just like the estimate of the EPV in semiparametric empirical Bayes estimation. Problem S4C75-4 in this section involves this idea.

Regarding the Chi-square goodness of fit test, if any category of data needs to be estimated by reference to other categories of data, doing so takes away one additional degree of freedom.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source:
Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C75-1. Similar to Question 197 of the Exam C Sample Questions from the Society of Actuaries. You have the following data for peanuts eaten by 1000 elephants over 8 hours:

5 elephants ate 0 peanuts.

30 elephants ate 1 peanut.

300 elephants ate 2 peanuts.

340 elephants ate 3 peanuts.

80 elephants ate 4 peanuts.

80 elephants ate 5 peanuts.

95 elephants ate 6 peanuts.

370 elephants ate 7 peanuts.

The number of peanuts eaten per hour by any given elephant follows a Poisson distribution. A randomly chosen elephant was found to have eaten 2 peanuts over 8 hours. Use semiparametric empirical Bayes estimation of the Bühlmann credibility factor to find the number of peanuts the elephant can be expected to eat during hour 9.

Solution S4C75-1. For the purposes of calculating Z, the credibility factor, we can think of a single time period as being equal to 8 days, so N = 1.

EPV^ = E(Total) = (5*0 + 30*1 + 300*2 + 340*3 + 80*4 + 80*5 + 95*6 + 370*7)/1000 = 5.53.

To find VHM^, we use the formula

VHM^ = $(1/(R-1))*{}_i\Sigma(E(R_i) - E(Total))^2$ -EPV^/N.

Here, R = 1000, so $(1/(R-1))*{}_i\Sigma(E(R_i) - E(Total))^2$ =

$(5*(0-5.53)^2 + 30*(1-5.53)^2 + 300*(2-5.53)^2 + 340*(3-5.53)^2 + 80*(4-5.53)^2 + 80*(5-5.53)^2 + 95*(6-5.53)^2 + 370*(7-5.53)^2)/999 = 7.721091091$.

Thus, VHM^ = 7.721091091 - 5.53 = 2.191091091, and K = EPV^/VHM^ = 2.523856732.

Therefore, Z = 1/(1+2.523856732) = 0.283779982.

Over the next *eight* hours, the elephant in question can be expected to eat a total of

2*0.283779982 + 5.53(1-0.283779982) = 4.528256663 peanuts. To get the number of peanuts eaten over the next hour, we divide this result by 8 to get 4.528256663/8 = **0.5660320829 peanuts.**

Problem S4C75-2. Similar to Question 240 of the Exam C Sample Questions from the Society of Actuaries. During 2010, you observe the number of federally issued corporate bailouts for 500 firms in the United States of Bailoutland (USB).

220 firms received 0 bailouts.

30 firms received 1 bailout.

140 firms received 2 bailouts.

60 firms received 3 bailouts.

50 firms received 4 bailouts.

No firms received more than 4 bailouts.

The number of bailouts per year for each firm is Poisson-distributed.

A randomly chosen firm was found to not have been bailed out in 2010. How many bailouts can it be expected to receive in 2011? Use semiparametric empirical Bayes estimation of the Bühlmann credibility factor to find the answer.

Solution S4C75-2. Here, we let our time period be one year, so N = 1.

EPV^ = E(Total) = (220*0 + 30*1 + 140*2 + 60*3 + 50*4)/500 = 1.38.

To find VHM^, we use the formula

VHM^ = $(1/(R-1))*_i\Sigma(E(R_i) - E(Total))^2$ -EPV^/N.

Here, R = 500, so $(1/(R-1))*_i\Sigma(E(R_i) - E(Total))^2$ =

$(220*(0-1.38)^2 + 30*(1-1.38)^2 + 140*(2-1.38)^2 + 60*(3-1.38)^2 + 50*(4-1.38)^2)/499$ = 1.959519038.

Thus, VHM^ = 1.959519038 - 1.38 = 0.5895190381, and K = EPV^/VHM^ = 2.381285013.

Therefore, Z = 1/(1+2.381285013) = 0.2957455512.

Our answer is therefore 0*0.2957455512 + 1.38*(1-0.2957455512) = **0.9718711393 bailouts.**

Problem S4C75-3. Similar to Question 257 of the Exam C Sample Questions from the Society of Actuaries. You observe two neighborhoods, A and B.

Neighborhood A: In year 1, neighborhood A had 4 robberies for 30 houses. In year 2, neighborhood A had 3 robberies for 20 houses. In year 3, neighborhood A had 3 robberies for 40 houses.

Neighborhood B: In year 2, neighborhood B had 3 robberies for 50 houses. In year 3, neighborhood B had 2 robberies for 45 houses.

The number of robberies for each neighborhood follows a Poisson distribution.

Use semiparametric empirical Bayes estimation within the Bühlmann-Straub credibility model to find the number of robberies that can be expected per house in neighborhood A in year 4.

Solution S4C75-3. EPV^ = E(Total) = (Number of robberies observed)/(Number of houses observed) = (4 + 3 + 3 + 3 + 2)/(30 + 20 + 40 + 50 + 45) = EPV^ = 3/37.

VHM Numerator = $_i\Sigma(N_i*(E(R_i) - E(\text{Total}))^2) - (R-1)\text{EPV}$^. Here, R = 2.

E(A) = (4 + 3 + 3)/(30 + 20 + 40) = 1/9. N_A = 30 + 20 + 40 = 90.

E(B) = (3 + 2)/(50 + 45) = 1/19. N_B = 50 + 45 = 95.

VHM Numerator = $90(1/9-3/37)^2 + 95(1/19-3/37)^2 - 3/37=$

= 0.0769717086.

VHM Denominator = $N_{total} - (1/N_{total})*_i\Sigma(N_i^2)$, where N_{total} = 30 + 20 + 40 + 50 + 45 = 185,

N_A = 30 + 20 + 40 = 90, and N_B = 50 + 45 = 95.

Thus, VHM Denominator = $185 - (1/185)(90^2 + 95^2)$ = 92.4324324.

VHM^ = VHM Numerator/VHM Denominator = 0.0769717086/92.4324324 = VHM^ = 0.0008327348591.

Thus, K = EPV^/VHM^ = (3/37)/0.0008327348591 = 97.36722343.

For neighborhood A, N_A = 90, so Z = 90/(90+97.36722343) = 0.4803401489.

Our answer is Z*E(A) + (1-Z)*E(Total) = 0.4803401489*(1/9) + (1-0.4803401489)*(3/37) = **0.0955057102 robberies per house.**

Problem S4C75-4. Similar to Question 47 of the Exam C Sample Questions from the Society of Actuaries. You have the following data:
30 people have 0 dogs.

50 people have 1 dog.

20 people have 2 dogs.

0 people have 3 or more dogs.

Fit this data to a Poisson distribution, estimating the Poisson parameter λ via the method of maximum likelihood. Then regroup the data into the following categories:
People who have 0 dogs, people who have 1 dog, people who have 2 or more dogs.

Then perform a Chi square goodness of fit test for the null hypothesis that observed data fits the Poisson distribution obtained as described above. What is the result of the hypothesis test?

(A) Reject the null hypothesis at the 0.005 significance level.

(B) Reject the null hypothesis at the 0.010 significance level, but not at the 0.005 level.

(C) Reject the null hypothesis at the 0.025 significance level, but not at the 0.010 level.

(D) Reject the null hypothesis at the 0.050 significance level, but not at the 0.025 level.

(E) Do not reject the null hypothesis at the 0.050 significance level.

Use the fourth page of the Exam 4 / C Tables to find critical values of the Chi-square distribution.

Relevant property of Poisson distribution: $\Pr(X=k) = e^{-\lambda}*\lambda^k/k!$

Solution S4C75-4. In maximum likelihood estimation of a Poisson parameter, the estimate for λ is the sample mean. There are 100 observations, so λ = (30*0 + 50*1 + 20*2)/100 = 0.9.

Now we can regroup the data as follows:
People who own 0 dogs: Observed: 30. Expected: $100*e^{-0.9}$ = 40.65696597.

People who own 1 dog: Observed: 50. Expected: $100*0.9e^{-0.9}$ = 36.59126938.

People who own 2 or more dogs: Observed: 20. Expected: 100 - 40.65696597 - 36.59126938 = 22.75176465.

Our Chi-square test statistic is $\sum_{i=1}^{t}((X_i - np_i)^2/np_i)$ =

$(30-40.65696597)^2/40.65696597 + (50-36.59126938)^2/36.59126938 + (20-22.75176465)^2/22.75176465 = 8.039790516$.

The number of degrees of freedom is 3-1-1 = 1. We subtract one additional degree of freedom, because the expected value for "two or more dogs" category was inferred from the expected values for the "0 dogs" and "1 dog" categories.

Looking at the table of critical values, we see that the value associated with 1 degree of freedom and a significance level of 0.005 is 7.879. As 8.039790516 > 7.879, our answer is **(A) Reject the null hypothesis at the 0.005 significance level**.

Problem S4C75-5. Similar to Question 140 of the Exam C Sample Questions from the Society of Actuaries. You observe the following 15 data points:

20, 20, 30, 40, 40, 40, 50, 60, 60, 60, 60, 70, 70, 80, 95

You hypothesize that the data fits the following cumulative distribution function:
F(x) = 0.3 when x = 31;

F(x) = 0.5 when x = 51;

F(x) = 0.8 when x = 61;

F(x) = 0.95 when x = 81.

You arrange the data into the largest number of classes such that each class is expected to have at least 3 data points. Find the test statistic for a Chi-square goodness of fit test that would evaluate whether the data fits the hypothesized distribution.

Solution S4C75-5. We first develop our classes of data.

We expect 0.3 of the data to occur prior to x = 31. As 15*0.3 = 4.5, this expected value is sufficiently high for us to designate the 0-31 range of values as its own class.

We expect 0.2 of the data to occur within the range 31-51. As 15*0.2 = 3, expected value is sufficiently high for us to designate the 31-51 range of values as its own class.

The same can be said of the 51-61 range of values, which has a probability of 0.3 and thus an expected value of 4.5.

The 61-81 range of values only has an expected value of 15*0.15 = 2.25, so we would need to broaden this class to 61+, which would have an expected value of 15 - 4.5 - 3 - 4.5 = 3.

Now we compare expected to observed data:

0-31 range: Observed: 3; Expected: 4.5.

31-51 range: Observed: 4; Expected: 3.

51-61 range: Observed: 5; Expected: 4.5.

61+ range: Observed: 4; Expected: 3.

Our Chi-square test statistic is thus $(3-4.5)^2/4.5 + (4-3)^2/3 + (5-4.5)^2/4.5 + (4-3)^2/3 =$ **1.222222222.**

Section 76

The Anderson-Darling Test

The **Anderson-Darling test** compares the empirical distribution function $F_n(x)$ with the model distribution function $F^*(x)$ via the following test statistic:

$A^2 = n*_t{}^u\int((F_n(x) - F^*(x))^2/(F^*(x)(1- F^*(x))))(f^*(x))dx.$

Here, n is the sample size and t and u are the points of left truncation and right censoring, respectively. If there is no left truncation or right censoring, t = 0 and u = ∞.

The Anderson-Darling test places more emphasis on good fit in the tails rather than in the intermediate parts of the distribution.

Like the Kolmogorov-Smirnov test, the Anderson-Darling test does not work for grouped data. When individual data are available, the test statistic becomes

$A^2 = -n(F^*(u)) + n({}_{j=0}{}^k\Sigma((1- F_n(y_j))^2(\ln(1- F^*(y_j)) - \ln(1- F^*(y_{j+1}))) + n({}_{j=0}{}^k\Sigma(F_n(y_j)^2(\ln(F^*(y_{j+1})) - \ln(F^*(y_j)))).$

Here, the y_j are the unique observed noncensored data points, $t = y_j$, and $u = y_{k+1}$.

When u = ∞, the last term of the first sum is 0.

The critical values c for the Anderson-Darling test are as follows:

For α = 0.1, c = 1.933.

For α = 0.05, c = 2.492.

For α = 0.01, c = 3.857.

A note on the Anderson-Darling test and the Kolmogorov-Smirnov test: When the data are right censored, the critical value is smaller than when no right censoring occurs.

A note on the Chi-square goodness-of-fit test: The chi-square goodness-of-fit test works best when the expected number of observations is close to the same from interval to interval.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Sources: Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 16, pp. 450-454.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C76-1. Similar to Question 189 of the Exam C Sample Questions from the Society of Actuaries. Evaluate the truth or falsehood of each of the following statements:

(A) For a null hypothesis that the population follows a particular distribution, using sample data to estimate the parameters of the distribution tends to keep constant the probability of a Type II error.

(B) The Kolmogorov-Smirnov test can only be used on individual data.

(C) The Anderson-Darling test can only be used on grouped data.

(D) The Anderson-Darling test tends to place more emphasis on a good fit in the tails of the distribution, rather than in the middle of the distribution.

(E) For a given number of cells, the critical value for the chi-square goodness-of-fit test becomes smaller with increased sample size.

Solution S4C76-1.

(A) Type II error is failure to reject a false null hypothesis. If sample data are used, the fit of the data to the null hypothesis is better than expected, so the probability of Type II error should *increase* compared to a situation where sample data are not used. Thus, **(A) is false.**

(B) It is indeed true that the Kolmogorov-Smirnov test can only be used on individual data, so **(B) is true.**

(C) The Anderson-Darling test can only be used on individual data, so **(C) is false.**

(D) It is indeed true that the Anderson-Darling test tends to place more emphasis on a good fit in the tails of the distribution, rather than in the middle of the distribution, so **(D) is true.**

(E) The critical value of the chi-square goodness-of-fit test depends on the number of degrees of freedom, which depends on the "number of cells," not the sample size. Therefore, **(E) is false.**

Problem S4C76-2. Similar to Question 244 of the Exam C Sample Questions from the Society of Actuaries. Evaluate the truth or falsehood of each of the following statements:

(A) The chi-square goodness-of-fit test works best when the expected number of observations is close to the same from interval to interval.

(B) For the Kolmogorov-Smirnov test, when the parameters of the distribution in the null hypothesis are estimated from the data, the probability of rejecting the null hypothesis increases.

(C) For the Kolmogorov-Smirnov test, the critical value for right censored data should be the same as the critical value for uncensored data.

(D) The Anderson-Darling test works for grouped data, but integration instead of summation must be used in such cases.

Solution S4C76-2.

(A) is true: The chi-square goodness-of-fit test indeed works best when the expected number of observations is close to the same from interval to interval.

(B) is false: When sample data are used, the probability of Type II error increases, and so the probability of rejecting a null hypothesis that should be rejected decreases.

(C) is false: For the Kolmogorov-Smirnov test, the critical value for right censored data should be *smaller than* the critical value for uncensored data.

(D) is false: The Anderson-Darling test can only be used on individual data.

Problem S4C76-3. You are examining a sample of 10 values, but you have lost the data for all except two of them - 3 and 9. There is no censoring or truncation of data. You are also aware that $F_n(3) = 0.2$ and $F_n(9) = 0.8$. The model distribution being used is exponential with mean 6. What is the value $n({}_{j=0}{}^{k}\Sigma((1- F_n(y_j))^2(\ln(1- F^*(y_j)) - \ln(1- F^*(y_{j+1})))$ in the Anderson-Darling test statistic?

Solution S4C76-3. We can only use the data we have available, so we need to find F*(3) and F*(9), which, respectively, are $(1-e^{-3/6}) = 0.3934693403$ and $(1-e^{-9/6}) = 0.7768698399$. The sample size n is 10, so our answer is

$10({}_{j=0}{}^{k}\Sigma((1- F_n(y_j))^2(\ln(1- F^*(y_j)) - \ln(1- F^*(y_{j+1}))) =$

$10((1-F_n(3))^2(\ln(1- F^*(3)) - \ln(1-F^*(9))) + 10((1-F_n(9))^2(\ln(1- F^*(9)) - \ln(1-F^*(\infty)))$.

When u = ∞, the last term of the first sum is 0, so the answer is

$10(1-0.2)^2(\ln(1- 0.3934693403) - \ln(1- 0.7768698399))$

$+ 10(1-0.8)^2(\ln(1- 0.7768698399)) = 6.4 + (-0.6) =$ **5.8**.

Problem S4C76-4. You are examining a sample of 10 values, but you have lost the data for all except two of them - 3 and 9. There is no censoring or truncation of data. You are also aware that $F_n(3) = 0.2$ and $F_n(9) = 0.8$. The model distribution being used is exponential with mean 6.

What is the value $n({}_{j=0}{}^{k}\Sigma(F_n(y_j)^2(\ln(F^*(y_{j+1})) - \ln(F^*(y_j))))$ in the Anderson-Darling test statistic?

Solution S4C76-4. Since n = 10,

$n({}_{j=0}{}^{k}\Sigma(F_n(y_j)^2(\ln(F^*(y_{j+1})) - \ln(F^*(y_j)))) =$

$10(F_n(3)^2(\ln(F^*(9)) - \ln(F^*(3)))) + 10(F_n(9)^2(\ln(F^*(\infty)) - \ln(F^*(9)))) =$

$10*0.2^2*(\ln(0.7768698399) - \ln(0.3934693403)) + 10*0.8^2*(\ln(1) - \ln(0.7768698399)) =$ **1.887995605**.

Problem S4C76-5. You are examining a sample of 10 values, but you have lost the data for all except two of them - 3 and 9. There is no censoring or truncation of data. You are also aware that $F_n(3) = 0.2$ and $F_n(9) = 0.8$. The model distribution being used is exponential with mean 6. Perform an Anderson-Darling test at the 0.01 significance level. Do you reject the null hypothesis that the observed data fits the exponential model distribution?

Solution S4C76-5. We calculate our test statistic:

$A^2 = -n(F^*(u)) + n({}_{j=0}{}^{k}\Sigma((1- F_n(y_j))^2(\ln(1- F^*(y_j)) - \ln(1- F^*(y_{j+1}))) + n({}_{j=0}{}^{k}\Sigma(F_n(y_j)^2(\ln(F^*(y_{j+1})) - \ln(F^*(y_j))))$.

We know from Solution S4C76-3 that $n({}_{j=0}{}^{k}\Sigma((1- F_n(y_j))^2(\ln(1- F^*(y_j)) - \ln(1- F^*(y_{j+1}))) = 5.8$.

We know from Solution S4C76-4 that $n({}_{j=0}{}^{k}\Sigma(F_n(y_j)^2(\ln(F^*(y_{j+1})) - \ln(F^*(y_j)))) = 1.887995605$.

Moreover, as $u = \infty$, $F^*(u) = 1$, so $-n(F^*(u)) = -10$, since n = 10.

Thus, $A^2 = -10 + 5.8 + 1.887995605 = A^2 = -2.312004395$.

The critical value for $\alpha = 0.01$ is c = 3.857.

Since -2.312004395 < 3.857, we **do not reject the null hypothesis, as $A^2 < c$.**

Section 77

The Likelihood Ratio Test

The following discussion of the likelihood ratio test is based on Chapter 16 of *Loss Models* (cited below).

The **likelihood ratio test** evaluates the following two hypotheses:

H_0: The data came from a population with distribution A.
H_1: The data came from a population with distribution B.

For the hypothesis test to be performed, distribution A must be a special case of distribution B. For instance, B can be a gamma distribution, while A can be an exponential distribution.

Let $L(\theta)$ be the likelihood function. Let θ_0 be the value of parameters maximizing the likelihood function. Let $L_0 = L(\theta_0)$ and let θ_1 be "the maximum likelihood estimator where the parameters can vary over all possible values from the alternative hypothesis" and let $L_1 = L(\theta_1)$. (455)

The test statistic is $T = 2*\ln(L_1/L_0) = 2(\ln(L_1) - \ln(L_0))$.

We reject the null hypothesis if $T > c$, where c is the critical value and is determined thus: $\alpha = \Pr(T > c)$, where T has "degrees of freedom equal to the number of free parameters in the model from the alternative hypothesis less the number of free parameters in the model from the null hypothesis" (455).

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Sources: Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 16, pp. 455-456.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C77-1. You have collected the following data:
3, 3, 4, 6, 7, 8, 10, 25.

You are performing the following hypothesis test:

H_0: The data came from a population that follows an exponential distribution with mean equal to 8.
H_1: The data came from a population that follows an exponential distribution with mean not equal to 8.

What is the test statistic T for a likelihood ratio test of this hypothesis?

Solution S4C77-1. Our parameter θ_0 is 8, since this is the hypothesized parameter in the null hypothesis. Our parameter θ_1 is the maximum likelihood estimate of the exponential parameter for the data, i.e., the sample mean of the data: (3 + 3 + 4 + 6 + 7 + 8 + 10 + 25)/8 = 8.25

Now we find the likelihood function:

$L(\theta) = \exp(-3/\theta)*\exp(-3/\theta)*\exp(-4/\theta)*\exp(-6/\theta)*\exp(-7/\theta)*\exp(-8/\theta)*\exp(-10/\theta)*\exp(-25/\theta) = \exp(-61/\theta)$, and therefore, $\ln(L(\theta)) = -66/\theta$.

Our test statistic is $T = 2(\ln(L_1) - \ln(L_0)) = 2(-66/8.25 - (-66/8)) =$ **T = 0.5**.

Problem S4C77-2. You have collected the following data:
3, 3, 4, 6, 7, 8, 10, 25.

You are performing the following hypothesis test:
H_0: The data came from a population that follows an exponential distribution with mean equal to 8.

H_1: The data came from a population that follows an exponential distribution with mean not equal to 8.

Perform a likelihood ratio test at the 0.1 significance level and determine whether the null hypothesis ought to be rejected. Use the fourth page of the Exam 4 / C Tables to find critical values of the Chi-square distribution.

Solution S4C77-2. From Solution S4C77-1, we know that the test statistic is T = 0.5. The number of degrees of freedom is (Free parameters in alternative hypothesis) - (Free parameters in null hypothesis). In the null hypothesis, the parameter θ is specified to be 8, so there are no free parameters. In the alternative hypothesis, the parameter θ is free, with the exception that it must not be 8. Thus, there is 1-0 = 1 degree of freedom for this test. Hence, we look up the Chi-square distribution critical value associated with 1 degree of freedom and $\alpha = 0.1$, which is 2.706. Therefore, since $0.5 < 2.706$, **we do not reject the null hypothesis, as T < c.**

Problem S4C77-3. Similar to Question 22 of the Exam C Sample Questions from the Society of Actuaries. You are examining 1000 data points and testing the following hypotheses:

H_0: The data came from a Pareto distribution with $\alpha = 3$ and $\theta = 60$.

H_1: The data came from a Pareto distribution with $\alpha \neq 3$ and $\theta \neq 60$.

You are given that $_x\Sigma(\ln(x+60)) = 12000$.

The maximum likelihood estimates for α and θ from this data result in $\ln(L_1) = -34615.58$.

Perform a likelihood ratio test of these hypotheses. Use the fourth page of the Exam 4 / C Tables to find critical values of the Chi-square distribution. Which of the following is the outcome of the test?

(A) Reject the null hypothesis at the 0.005 significance level.

(B) Reject the null hypothesis at the 0.010 significance level, but not at the 0.005 level.

(C) Reject the null hypothesis at the 0.025 significance level, but not at the 0.010 level.

(D) Reject the null hypothesis at the 0.050 significance level, but not at the 0.025 level.

(E) Reject the null hypothesis at the 0.10 significance level, but not at the 0.050 level.

(F) Do not reject the null hypothesis at the 0.10 significance level.

Relevant properties for Pareto distributions: $f(x) = \alpha\theta^\alpha/(x+\theta)^{\alpha+1}$.

Solution S4C77-3. We need to find $\ln(L_0)$. To do so, we first find the formula for the likelihood function $L(\theta) = {}_x\Pi(\alpha\theta^\alpha/(x+\theta)^{\alpha+1})$, and, by implication the formula for $\ln(L(\theta))$ is $_x\Sigma(\ln(\alpha\theta^\alpha/(x+\theta)^{\alpha+1})) = {}_x\Sigma(\ln(\alpha\theta^\alpha)) - {}_x\Sigma(\ln(x+\theta)^{\alpha+1})) =$

$n*\ln(\alpha\theta^\alpha) - (\alpha+1){}_x\Sigma(\ln(x+\theta))$.

We are given that $n = 1000$, and, for L_0, $\alpha = 3$ and $\theta = 60$. Moreover, we are given that $_x\Sigma(\ln(x+60)) = {}_x\Sigma(\ln(x+\theta)) = 12000$.

Thus, $\ln(L_0) = 1000*\ln(3*60^3) - (3+1)*12000 = -34618.35402$.

Thus, we can find our test statistic $T = 2(\ln(L_1) - \ln(L_0)) = 2(-34615.58 - (-34618.35402)) = T = 5.548049332$.

We have 0 degrees of freedom in the null hypothesis, since both parameters are specified, and 2 degrees of freedom in the alternative hypothesis, since neither parameter is specified. We thus have 2 degrees of freedom overall. We examine the table of critical values to find that for significance level of 0.10, the critical value is 4.605, while for significance level of 0.050, the critical value is 5.991. Since $4.605 < 5.548049332 < 5.991$, the answer is **(E) Reject the null hypothesis at the 0.10 significance level, but not at the 0.050 level**.

Problem S4C77-4. Similar to Question 235 of the Exam C Sample Questions from the Society of Actuaries. You have collected the following sample data: 3, 4, 6. You make the following null and alternative hypotheses:

H_0: The data came from a Weibull distribution with $\tau = 3$.
H_1: The data came from a Weibull distribution with $\tau \neq 3$.

Find $\ln(L_0)$, the value of the loglikelihood function at $\tau = 3$.

Relevant properties for Weibull distributions: $f(x) = \tau(x/\theta)^\tau*\exp(-(x/\theta)^\tau)/x$.

Solution S4C77-4. We first find the formula for the likelihood function $L(\theta, \tau) =$

${}_x\Pi(\tau(x/\theta)^\tau*\exp(-(x/\theta)^\tau)/x)$. The loglikelihood function is $\ln(L(\theta, \tau)) =$

${}_x\Sigma(\ln(\tau(x/\theta)^\tau*\exp(-(x/\theta)^\tau)/x)) =$

${}_x\Sigma(\ln(\tau) + \tau*\ln(x) - \tau*\ln(\theta) - (x/\theta)^\tau - \ln(x)) =$

${}_x\Sigma(\ln(\tau) + (\tau-1)*\ln(x) - \tau*\ln(\theta) - (x/\theta)^\tau) =$

$n(\ln(\tau) - \tau*\ln(\theta)) + {}_x\Sigma((\tau-1)*\ln(x) - (x/\theta)^\tau)$.

For $\tau = 3$, $n = 3$, and the given sample data.

$\ln(L(\theta, 3)) = 3(\ln(3) - 3*\ln(\theta)) + (3-1)(\ln(3) + \ln(4) + \ln(6) - (3/\theta)^3*(4/\theta)^3*(6/\theta)^3)$

$-9*\ln(\theta) + \ln(27) + 2*\ln(74) - 746496/\theta^3$.

To find $\ln(L_0)$, we first need to maximize $\ln(L(\theta))$ to find the maximum likelihood estimate for θ when $\tau = 3$.

$d(\ln(L(\theta, 3)))/d\theta = -9/\theta + 2239488/\theta^4 = 0$ at maximum.

Thus, $2239488/\theta^4 = 9/\theta \rightarrow 2239488/\theta^3 = 9$ and $\theta = (2239488/9)^{1/3} = \theta = 62.89779346$.

Therefore, $\ln(L_0) = -9*\ln(62.89779346) + \ln(27) + 2*\ln(74) - 746496/62.89779346^3 =$

$\ln(L_0) = -28.36963269$.

Problem S4C77-5. Similar to Question 235 of the Exam C Sample Questions from the Society of Actuaries. You have collected the following sample data: 3, 4, 6. You make the following null and alternative hypotheses:

H_0: The data came from a Weibull distribution with $\tau = 3$.
H_1: The data came from a Weibull distribution with $\tau \neq 3$.

At the maximum likelihood estimates of θ and τ, you are instructed to use $\ln(L_1) = -26$.

Perform a likelihood ratio test of these hypotheses. Use the fourth page of the Exam 4 / C Tables to find critical values of the Chi-square distribution. Which of the following is the outcome of the test?

(A) Reject the null hypothesis at the 0.005 significance level.

(B) Reject the null hypothesis at the 0.010 significance level, but not at the 0.005 level.

(C) Reject the null hypothesis at the 0.025 significance level, but not at the 0.010 level.

(D) Reject the null hypothesis at the 0.050 significance level, but not at the 0.025 level.

(E) Reject the null hypothesis at the 0.10 significance level, but not at the 0.050 level.

(F) Do not reject the null hypothesis at the 0.10 significance level.

Solution S4C77-5.

In Solution S4C77-4, we found that $\ln(L_0) = -28.36963269$.

We find our test statistic $T = 2(\ln(L_1) - \ln(L_0)) = 2(-26 - (-28.36963269)) = 4.739265389$.

In the null hypothesis, there is one free parameter (θ), whereas both τ and θ are free in the alternative hypothesis. Therefore, this test has 2-1 = 1 degree of freedom. The critical value of the Chi-square distribution associated with the 0.050 significance level is 3.841, whereas the critical value associated with the 0.025 significance level is 5.024. We note that 3.841 < 4.739265389 < 5.024, and therefore the correct answer is **(D) Reject the null hypothesis at the 0.050 significance level, but not at the 0.025 level**.

Section 78

The Schwarz Bayesian Criterion and Exam-Style Questions on Hypothesis Testing of Models

The **Schwarz Bayesian criterion (SBC)** "recommends that, when ranking models, a deduction of (r/2)*ln(n) should be made from the loglikelihood value, where r is the number of estimated parameters and n is the sample size" (Klugman, Panjer, and Willmot 2008, p. 461).

The **Schwarz Bayesian adjustment** is the application of the SBC to a loglikelihood value. Multiple models' loglikelihood values can be compared after the Schwarz Bayesian adjustment is made. Then, the model that has the highest (which can be the *least negative*) loglikelihood value is favored over all the others.

Here, some practice problems regarding the SBC are offered. This section also gives some exam-style practice with the other hypothesis tests previously covered.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Sources: Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 16, pp. 459-461.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C78-1. You are given that a particular model has a maximum loglikelihood value of -340. You also know that the model uses 3 parameters and is used to interpret a sample of 89 data points. What is the adjusted loglikelihood value after the Schwarz Bayesian adjustment has been made?

Solution S4C78-1. We subtract (r/2)*ln(n) from the loglikelihood value of -340. Here, r = 3 and n = 89. Thus, our answer is -340 - (3/2)*ln(89) = **-346.7329546.**

Problem S4C78-2. Similar to Question 149 of the Exam C Sample Questions from the Society of Actuaries. What happens to each of the following quantities as the sample size n increases without bound? Assume that the model distribution fits the data.

(A) Kolmogorov-Smirnov test statistic
(B) Anderson-Darling test statistic
(C) Chi-square goodness-of-fit test statistic
(D) The effect of Schwarz Bayesian adjustment on the loglikelihood function.

Solution S4C78-2.

(A) The Kolmogorov-Smirnov test statistic is $D = \max_{t \leq x \leq u} | F_n(x) - F^*(x) |$. If the model fits the data, then, as the sample size n increases, the empirical distribution is able to take finer increments of probability into account, and so $F_n(x)$ becomes closer to $F^*(x)$ for each x. Therefore, the difference $F_n(x) - F^*(x)$ ought to decrease, and **D ought to approach 0 as n increases without bound**.

(B) The Anderson-Darling test statistic is $A^2 = -n(F^*(u)) + n({}_{j=0}{}^{k}\Sigma((1- F_n(y_j))^2(\ln(1- F^*(y_j)) - \ln(1- F^*(y_{j+1}))) + n({}_{j=0}{}^{k}\Sigma(F_n(y_j)^2(\ln(F^*(y_{j+1})) - \ln(F^*(y_j))))$. We note that n, the sample size, is present as a coefficient of all three factors. If A^2 is positive, higher n means higher A^2.

If A^2 is negative, higher n means lower (more negative) A^2. Thus, **as n increases without bound, $|A^2| \to \infty$**.

(C) The Chi-square test statistic has a factor of (Observed value - Expected value)^2/(Expected value) for each data point in the sample. This is always a positive or zero factor. Therefore, **as the number of sample data points increases without bound, the Chi-square test statistic increases without bound** as well.

(D) The Schwarz Bayesian adjustment subtracts (r/2)*ln(n) from the likelihood function, which is a quantity that increases without bound as n increases without bound. So **the effect of theSchwarz Bayesian adjustment increases without bound**, *but* the loglikelihood function itself is likely to increase as n increases, so the overall effect on the loglikelihood function is determined by a multiplicity of factors, and not the Schwarz Bayesian adjustment alone.

Problem S4C78-3. Similar to Question 266 of the Exam C Sample Questions from the Society of Actuaries. You are evaluating a sample of 300 values and comparing the following models, each with its own loglikelihood value and number of parameters:

I: 4 parameters, loglikelihood = -300.
II: 7 parameters, loglikelihood = -295.
III: 9 parameters, loglikelihood = -293.
IV: 35 parameters, loglikelihood = -208.
V: 40 parameters, loglikelihood = -200.

Which of these models is favored by the Schwarz Bayesian criterion?

Solution S4C78-3. The favored model has the highest or *least negative* adjusted loglikehood value. Let ln(L) be the loglikelihood value. Then the adjusted value is ln(L) - (r/2)*ln(n). For all models, n = 300.

Model I: r = 4, so adjusted value is -300 - (4/2)*ln(300) = -311.4075649.
Model II: r = 7, so adjusted value is -295 - (7/2)*ln(300) = -314.9632387.
Model III: r = 9, so adjusted value is -293 - (9/2)*ln(300) = -318.6670211.
Model IV: r = 35, so adjusted value is -208 - (35/2)*ln(300) = -307.8161933.
Model V: r = 40, so adjusted value is -200 - (40/2)*ln(300) = -314.0756495.

The highest adjusted loglikehood value among these is -307.8161933, so **Model IV** is favored by the Schwarz Bayesian criterion.

Problem S4C78-4. Similar to Question 201 of the Exam C Sample Questions from the Society of Actuaries. You are conducting a Chi-square goodness-of-fit test to determine whether observed data fits a certain distribution with cumulative distribution function F(x). There are 1000 observations in all. You know the following.

For $x \leq 10$, there are 90 observed values of x, and F(10) = 0.1.
For $10 < x \leq 50$, there are 304 observed values of x, and F(50) = 0.4.
For $50 < x \leq 200$, there are 455 observed values of x, and F(200) = 0.85.
For $x > 200$, there are 151 observed values of x.

What is the result of the hypothesis test? Use the fourth page of the Exam 4 / C Tables to find critical values of the Chi-square distribution.

(A) Reject the null hypothesis at the 0.005 significance level.
(B) Reject the null hypothesis at the 0.010 significance level, but not at the 0.005 level.
(C) Reject the null hypothesis at the 0.025 significance level, but not at the 0.010 level.
(D) Reject the null hypothesis at the 0.050 significance level, but not at the 0.025 level.
(E) Reject the null hypothesis at the 0.10 significance level, but not at the 0.050 level.
(F) Do not reject the null hypothesis at the 0.10 significance level.

Solution S4C78-4. First, we find the expected number of observations in each category.

For $x \leq 10$, there are F(10)*1000 = 100 expected observations.

For $10 < x \leq 50$, there are (F(50)-F(10))*1000 = (0.4-0.1)*1000 = 300 expected observations.

For $50 < x \leq 200$, there are (F(200)-F(50))*1000 = (0.85-0.4)*1000 = 450 expected observations.

It follows that for x > 200 there are 1000 - 450 - 300 - 100 = 150 expected observations.

The test statistic is therefore

$(90-100)^2/100 + (304-300)^2/300 + (455-450)^2/450 + (151-150)^2/150 = 1.1155555556.$

There are four categories used, so there are 4-1 = 3 degrees of freedom.

We note that the critical value for 3 degrees of freedom and $\alpha = 0.10$ is 7.779. As 1.1155555556 < 7.779, we are far from any threshold of rejection, and so the correct answer is **(F) Do not reject the null hypothesis at the 0.10 significance level**.

Problem S4C78-5. Similar to Question 222 of the Exam C Sample Questions from the Society of Actuaries. You observe 100 elephants, of which 50 ate 0 peanuts, 26 ate 1 peanut, 16 ate 2 peanuts, and 8 ate 3 or more peanuts. In all, 86 peanuts were eaten. You hypothesize that the number of peanuts eaten per elephant follows a Poisson distribution, whose parameter λ is estimated by the sample mean. You then conduct a Chi-square goodness-of-fit test to see whether actual data fits the hypothesized model. What is the result of the test? Use the fourth page of the Exam 4 / C Tables to find critical values of the Chi-square distribution.

(A) Reject the null hypothesis at the 0.005 significance level.
(B) Reject the null hypothesis at the 0.010 significance level, but not at the 0.005 level.
(C) Reject the null hypothesis at the 0.025 significance level, but not at the 0.010 level.
(D) Reject the null hypothesis at the 0.050 significance level, but not at the 0.025 level.
(E) Reject the null hypothesis at the 0.10 significance level, but not at the 0.050 level.
(F) Do not reject the null hypothesis at the 0.10 significance level.

Solution S4C78-5. We first estimate λ = 86 peanuts/100 elephants = 0.86.

Thus, $100*e^{-0.86}$ = 42.31620823 elephants can be expected to eat 0 peanuts.
$100*0.86*e^{-0.86}$ = 36.39193908 elephants can be expected to eat 1 peanut.
$100*0.86^2*e^{-0.86}/2$ = 15.6485338 elephants can be expected to eat 2 peanuts.
100 - 15.6485338 - 36.39193908 - 42.31620823 = 5.643318885 elephants can be expected to eat 3 or more peanuts.

Now we find our test statistic:
$(50-42.31620823)^2/42.31620823 + (26-36.39193908)^2/36.39193908 + (16-15.6485338)^2/15.6485338 + (8-5.643318885)^2/5.643318885 = 5.354763826$.

There are four categories of data, one of which is derived from the others. The parameter λ was also estimated from the data. Thus, we must subtract 2 from 4 to get our number of degrees of freedom: 2.

We examine the table of critical values to find that for significance level of 0.10, the critical value is 4.605, while for significance level of 0.050, the critical value is 5.991. Since 4.605 < 5.354763826 < 5.991, the answer is **(E) Reject the null hypothesis at the 0.10 significance level, but not at the 0.050 level**.

Section 79

Basics of Simulation and the Inversion Method

The following are the general steps to follow in the simulation approach:

"1. Build a model for S that depends on random variables X, Y, Z,..., where their distributions and any dependencies are known.

"2. For j = 1, ..., n generate pseudorandom variables x_j, y_j, z_j, ... and then compute s_j using the model from step 1.

"3. The cumulative distribution function (cdf) of S may be approximated by $F_n(s)$, the empirical cdf based on the pseudorandom sample s_1, ..., s_n.

"4. Compute quantities of interest, such as the mean, variance, percentiles, or probabilities, using the empirical cdf." (Klugman, Panjer, and Willmot 2008, p. 644)

Generating a pseudorandom variable:

Let X be the real random variable of interest. Let X* be another random variable with an identical distribution to that of X. Then, we can take a **pseudorandom sample** of values from X* that is generated by some known process which is based on X*. Although these values are not actual random values of X, an external observer would not be able to tell the difference between these values and actual observed random values of X.

The **inversion method** of simulation works as follows:

1. Let U have a uniform distribution on the interval from 0 to 1.

2. Generate pseudorandom numbers u_1**, ..., u_n**.

3. For each j from 1 to n, let x_j** = $F_X^{-1}(u_j**)$.

This gives us our pseudorandom sample of values x_j**.

The inversion method can be used for both discrete and continuous distributions.

Property 79.1: When the random variable of *occurrences per unit time* is Poisson with mean λ, the random variable of *time to the next occurrence* is exponential with mean 1/λ.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Sources: Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 21, pp. 643-645.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C79-1. You are hypothesizing that data fits an exponential distribution with mean 5. To simulate this random variable, you use the inversion method, where U has a uniform distribution on the interval from 0 to 1. The following sequence of 5 pseudorandom values of U was generated in Excel: 0.916656645, 0.464812622, 0.950301766, 0.895541938, 0.303598448. Determine the pseudorandom sample of values of X* from these values.

Solution S4C79-1. We use the formula $x_j^{**} = F_X^{-1}(u_j^{**})$. Here, because the hypothesized distribution is exponential with mean 5, $F_X(x_j^{**}) = 1 - \exp(-x_j^{**}/5) = u_j^{**}$, and so

$1 - u_j^{**} = \exp(-x_j^{**}/5) \rightarrow \ln(1 - u_j^{**}) = -x_j^{**}/5 \rightarrow x_j^{**} = -5*\ln(1 - u_j^{**})$.

Therefore, we get the following pseudorandom sample of values of X*:

-5*ln(1-0.916656645) = 12.42393199

-5*ln(1-0.464812622) = 3.125691771

-5*ln(1-0.950301766) = 15.0089294

-5*ln(1-0.895541938) = 11.29484804

-5*ln(1-0.303598448) = 1.809144212

Thus, our sample is as follows: **1.809144212, 3.125691771, 11.29484804, 12.42393199, 15.0089294**.

Problem S4C79-2. Similar to Part of Question 131 of the Exam C Sample Questions from the Society of Actuaries. The number of large losses for a gambler per year follows a Poisson distribution with mean 0.5. To simulate time to the next large loss, the inversion method is used, where U has a uniform distribution on the interval from 0 to 1. A pseudorandom sample of four values of U is obtained as follows: 0.15, 0.35, 0.65, 0.95. Each of these values corresponds to one large loss suffered by the gambler. Determine the times at which each of these simulated losses occurred.

Solution S4C79-2. We refer to Property 79.1: When the random variable of *occurrences per unit time* is Poisson with mean λ, the random variable of *time to the next occurrence* is exponential with mean 1/λ. We use the formula $x_j^{**} = F_X^{-1}(u_j^{**})$, where $F_X(x_j^{**}) = 1 - \exp(-x_j^{**}/2) = u_j^{**}$, and so $\exp(-x_j^{**}/2) = 1 - u_j^{**} \rightarrow x_j^{**} = -2*\ln(1 - u_j^{**})$.

We find each of the four relevant x_j^{**} values to be as follows:
Time to first loss: -2*ln(1 - 0.15) = 0.325037859 years.

Time between first and second loss: -2*ln(1 - 0.35) = 0.8615658322 years.

Time between second and third loss: -2*ln(1 - 0.65) = 2.099644249 years.

Time between third and fourth loss: -2*ln(1 - 0.95) = 5.991464547.

To find time to each kth loss, we add up all the intervals between the (j-1)st to the jth loss up to the desired kth loss.

Thus, **the first loss occurs after 0.325037859 years.**

The second loss occurs after 1.186603691 years.

The third loss occurs after 3.28624794 years.

The fourth loss occurs after 9.277712487 years.

Problem S4C79-3. Similar to Part of Question 131 of the Exam C Sample Questions from the Society of Actuaries. The number of large losses for a gambler per year follows a Poisson distribution with mean 0.5. To simulate time to the next large loss, the inversion method is used, where U has a uniform distribution on the interval from 0 to 1. A pseudorandom sample of four values of U is obtained as follows: 0.15, 0.35, 0.65, 0.95. Each of these values corresponds to one large loss suffered by the gambler.

The amount of each loss X for the gambler follows a discrete probability distribution as follows: Pr(X = 1000) = 0.34; Pr(X = 3000) = 0.35; Pr(X = 10000) = 0.31. To simulate the amount of each loss, the inversion method is used, where V has a uniform distribution on the interval from 0 to 1. A pseudorandom sample of four values of V is obtained as follows: 0.46, 0.78, 0.36, 0.15.

Find the sum of the losses the gambler sustained during the 10 years over which the simulation was performed.

Solution S4C79-3. We use the formula $x_j^{**} = F_X^{-1}(u_j^{**})$, where $F_X(1000) = 0.34$, $F_X(3000) = 0.34 + 0.35 = 0.69$, and $F_X(10000) = 1$. Thus, any values of V under 0.34 correspond to a loss of 1000. This means that the fourth loss is 1000. Any values of V between 0.34 and 0.69

correspond to losses of 3000. This means that the first and third losses are 3000. Any values of V between 0.69 and 1 correspond to losses of 10000. This means that the second loss is 10000. Thus, total losses are 1000 + 2*3000 + 10000 = **17000.**

Problem S4C79-4. Similar to Part of Question 131 of the Exam C Sample Questions from the Society of Actuaries. The number of large losses for a gambler per year follows a Poisson distribution with mean 0.5. To simulate time to the next large loss, the inversion method is used, where U has a uniform distribution on the interval from 0 to 1. A pseudorandom sample of four values of U is obtained as follows: 0.15, 0.35, 0.65, 0.95. Each of these values corresponds to one large loss suffered by the gambler.

The amount of each loss X for the gambler follows a discrete probability distribution as follows: Pr(X = 1000) = 0.34; Pr(X = 3000) = 0.35; Pr(X = 10000) = 0.31. To simulate the amount of each loss, the inversion method is used, where V has a uniform distribution on the interval from 0 to 1. A pseudorandom sample of four values of V is obtained as follows: 0.46, 0.78, 0.36, 0.15.

The gambler purchases a special insurance policy that pays his losses and charges him an annual premium equal to (Expected annual loss) + (Standard deviation of expected annual loss). Find the amount of money the insurance company has gained or lost during the 10 years over which the simulation was performed.

Solution S4C79-4. Let N = number of annual losses. Let X = size of each loss.

E(X) = 1000*0.34 + 3000*0.35 + 10000*0.31 = 4490.

Var(X) = $0.34*(1000-4490)^2 + 0.35*(3000-4490)^2 + 0.31*(10000-4490)^2$ = 14329900.

$E(X^2) = Var(X) + E(X)^2$ = 34490000.

E(N) = λ = 0.5, and Var(N) = λ = 0.5.

Annual expected loss is E(N)*E(X) = 0.5*4490 = 2245.

Standard deviation of expected annual loss is $\sqrt{(E(N)*E(X)^2)}$ = √(0.5*34490000) = 4152.709959.

Thus, annually, the gambler is charged 2245 + 4152.709959 = 6397.709959.

Over 10 years, the gambler will pay 63977.09959. The insurance company ends up paying 17000 to the gambler (as per Solution S4C79-3), so its gain is 63977.09959 - 17000 = **gain of 46977.09959.**

Problem S4C79-5. Similar to Part of Question 182 of the Exam C Sample Questions from the Society of Actuaries. There are 90 people in a group, of which each person has an independent probability of 0.05 of flying to the moon in a year. To simulate numbers of people flying to the moon in a year, the inversion method is used, where U has a uniform distribution on the interval from 0 to 1. A pseudorandom sample of four values of U, where each value corresponds to one year of observation, is obtained as follows: 0.02, 0.06, 0.13, 0.19. On the basis of the simulated data, find the per-year average number of people that fly to the moon.

Solution S4C79-5. Based on the given conditions, the number of people who fly to the moon in a year is binomially distributed, with m = 90 and q = 0.05.

We can find Pr(0) = 0.95^{90} = 0.0098883647.

Pr(1) = 100*0.05*0.95^{89} = 0.0520440248.

Pr(2) = C(100,2)*0.05^2*0.95^{88} = 0.1355883804.

As $x_j** = F_X^{-1}(u_j**)$, we note that any values of u less than 0.0098883647 correspond to 0 people flying to the moon; any values of u between 0.0098883647 and 0.0098883647+0.0520440248 = 0.0619323895 correspond to 1 person flying to the moon. Thus, the values 0.02 and 0.06 in the pseudorandom sample of values of U each correspond to 1 person flying to the moon.

Any values between 0.0619323895 and 0.0619323895+0.1355883804 = 0.1975207699 correspond to 2 people flying to the moon. Thus, the values 0.13 and 0.19 in the pseudorandom sample of values of U each correspond to 2 people flying to the moon. Over the four simulated years, a total of 1+1+2+2 = 6 people fly to the moon, so the average is 6/4 = **1.5 people per year.**

Section 80

Exam-Style Questions on the Inversion Method

This section affords some exam-style practice regarding the inversion method of simulation.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source:

Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C80-1. Similar to Question 202 of the Exam C Sample Questions from the Society of Actuaries. Claim sizes follow a lognormal distribution with parameters $\mu = 6.5$ and $\sigma = 3$. There is a policy limit of 2000 for each claim. Observed claims where simulated using the inversion method, where U has a uniform distribution on the interval from 0 to 1. The following sequence of 5 pseudorandom values of U was generated: 0.34, 0.12, 0.67, 0.94, 0.29. Using the simulated data, find the average value paid out by the insurer per claim.

Solution S4C80-1. We use the formula $x_j^{**} = F_X^{-1}(u_j^{**})$.

For each u_j^{**}, it is the case that $u_j^{**} = \Phi((\ln(x_j^{**}) - \mu)/\sigma) = \Phi((\ln(x_j^{**}) - 6.5)/3)$

Thus, $\Phi^{-1}(u_j^{**}) = (\ln(x_j^{**}) - 6.5)/3$ and $x_j^{**} = \exp(3*\Phi^{-1}(u_j^{**})+6.5)$. This can be found in MS Excel for each u_j^{**} using the input "=EXP(3*NORMSINV(u_j**) + 6.5)". We use the following inputs:

"=EXP(3*NORMSINV(0.34) + 6.5)" → 192.9846422

"=EXP(3*NORMSINV(0.12) + 6.5)" → 19.59039948

"=EXP(3*NORMSINV(0.67) + 6.5)" → 2489.256865

"=EXP(3*NORMSINV(0.94) + 6.5)" → 70567.20406

"=EXP(3*NORMSINV(0.29) + 6.5)" → 126.4498569

We note that two of these amounts, 2489.256865 and 70567.20406, are capped at 2000, so our average is (192.9846422+19.59039948+2000+2000+126.4498569)/5 = **867.8049797**.

Problem S4C80-2. Similar to Question 220 of the Exam C Sample Questions from the Society of Actuaries. The number of elephants (N) eating peanuts follows a Poisson distribution with mean λ = 1.4. The peanuts eaten per elephant (X) follows an exponential distribution with mean 300. You simulate N using the inversion method, where U has a uniform distribution on the interval from 0 to 1. You observe a value of U to be 0.87. You simulate X using the inversion method, where V has a uniform distribution on the interval from 0 to 1. The values of V, as needed, are v_1 = 0.56, v_2 = 0.34, v_3 = 0.23. Find the total number of peanuts eaten.

Solution S4C80-2. We use the formula $x_j^{**} = F_X^{-1}(u_j^{**})$.

For N, we can find the following:
$\Pr(N = 0) = e^{-1.4} = 0.2465969639$.

$\Pr(N = 1) = 1.4e^{-1.4} = 0.3452357495$.

$\Pr(N = 2) = 1.4^2e^{-1.4}/2 = 0.2416650247$.

We note that F(2) = Pr(N = 0) + Pr(N = 1) + Pr(N = 2) = 0.8334977381 < 0.87.

$\Pr(N = 3) = 1.4^3e^{-1.4}/6 = 0.1127770115$. Clearly, F(3) = F(2) + Pr(N = 3) > 0.87, so the simulated value of N is 3. For X, we must thus use all three of the simulated values of V.

$u_j^{**} = 1 - \exp(-x_j^{**}/300) \rightarrow x_j = -300*\ln(1- u_j^{**})$.

Thus, our three simulated values of numbers of peanuts eaten are as follows:

-300*ln(1- 0.56) = 246.2941656

-300*ln(1- 0.34) = 124.6546332

-300*ln(1- 0.23) = 78.40942924

Our answer is therefore 246.2941656 + 124.6546332 + 78.40942924 = **449.358228 peanuts**.

Problem S4C80-3. Similar to Question 237 of the Exam C Sample Questions from the Society of Actuaries. Claim sizes follow a lognormal distribution with parameters μ = 6.5 and σ = 3. There is a deductible 150 for each claim. Observed claims where simulated using the inversion method, where U has a uniform distribution on the interval from 0 to 1. The following sequence of 5 pseudorandom values of U was generated: 0.80, 0.50, 0.23, 0.32, 0.15. Using the simulated data, find the average value paid out by the insurer per claim.

Solution S4C80-3. We use the formula $x_j^{**} = F_X^{-1}(u_j^{**})$.

For each u_j**, it is the case that u_j** = $\Phi((\ln(x_j**) - \mu)/\sigma) = \Phi((\ln(x_j**) - 6.5)/3)$

Thus, $\Phi^{-1}(u_j**) = (\ln(x_j**) - 6.5)/3$ and $x_j** = \exp(3*\Phi^{-1}(u_j**)+6.5)$. This can be found in MS Excel for each u_j** using the input "=EXP(3*NORMSINV(u_j**) + 6.5)". We use the following inputs:
"=EXP(3*NORMSINV(0.80) + 6.5)" → 8307.082147

"=EXP(3*NORMSINV(0.50) + 6.5)" → 665.141633

"=EXP(3*NORMSINV(0.23) + 6.5)" → 72.49078515

"=EXP(3*NORMSINV(0.32) + 6.5)" → 163.5148158

"=EXP(3*NORMSINV(0.15) + 6.5)" → 29.68672071

We note that two of these claims, 72.49078515 and 29.68672071, are less than the deductible of 150 and so will not cost the insurer anything.

The average amount paid out is therefore

(8307.082147 + 665.141633 + 163.5148158 - 3*150)/5 = **1737.147719**.

Problem S4C80-4. Similar to Question 249 of the Exam C Sample Questions from the Society of Actuaries. The number of claims in a year (N) is known to follow a distribution with the following cumulative distribution functions.

$F_N(0) = 0.34$, $F_N(1) = 0.56$, $F_N(2) = 0.66$, $F_N(3) = 0.78$, $F_N(4) = 1$.

You simulate N using the inversion method, where U has a uniform distribution on the interval from 0 to 1. You observe a value of U to be 0.73.

The amount of each claim (X) follows a Weibull distribution with $\theta = 600$ and $\tau = 3$.

You simulate X using the inversion method, where V has a uniform distribution on the interval from 0 to 1. You observe the following values of V, in the order below:
0.34, 0.75, 0.35, 0.39.
Each claim has a deductible of 450, and the total claims are subject to a maximum out-of-pocket limit for the insured of 1000. Find the total amount paid out by the insurer on the basis of this simulated data.

Useful properties regarding Weibull distributions: $F(x) = 1 - \exp(-(x/\theta)^\tau)$.

Solution S4C80-4. Since the simulated value of U is between $F_N(2) = 0.66$ and $F_N(3) = 0.78$, we conclude that N = 3 and there were 3 claims.

Now we find the magnitude of each claim.

We use the formula $x_j^{**} = F_X^{-1}(u_j^{**})$.

For each u_j^{**}, it is the case that $u_j^{**} = 1 - \exp(-(x_j^{**}/600)^3) \rightarrow$

$\exp(-(x_j^{**}/600)^3) = 1 - u_j^{**} \rightarrow (x_j^{**}/600)^3 = -\ln(1 - u_j^{**}) \rightarrow x_j^{**} = 600(-\ln(1 - u_j^{**}))^{1/3}$.

Thus, we have the following claims:

$600(-\ln(1 - 0.34))^{1/3} = 447.7273664$
$600(-\ln(1 - 0.75))^{1/3} = 669.0158433$
$600(-\ln(1 - 0.35))^{1/3} = 453.1452246$

The first claim is fully within the insured's deductible, but contributes 447.7273664 toward the out-of-pocket limit. Of the second claim, only 219.0158433 is paid by the insurer, and 450 is contributed to the out-of-pocket limit, leaving 1000-450-447.7273664 = 102.2726336 as the deductible for the third claim. Hence, the insurer pays 453.1452246-102.2726336 = 350.872591 of the third claim, resulting in a total payment of 219.0158433+350.872591 = **569.8884343**.

Problem S4C80-5. Similar to Question 265 of the Exam C Sample Questions from the Society of Actuaries. Paid losses in a year follow a lognormal distribution with parameters $\mu = 10$ and $\sigma = 2$. The ratio of unpaid losses to paid losses is estimated to be $y = 0.35*x^{0.56}*e^{-0.34x}$, where x is the year. You simulate paid losses using the inversion method, where U has a uniform distribution on the interval from 0 to 1. You observe the following three values of U: 0.34, 0.12, 0.98. Use these values to calculate the empirical estimate of average unpaid losses in year 2.

Solution S4C80-5. First, we determine the estimated paid losses.

We use the formula $x_j^{**} = F_X^{-1}(u_j^{**})$.

For each u_j^{**}, it is the case that $u_j^{**} = \Phi((\ln(x_j^{**}) - \mu)/\sigma) = \Phi((\ln(x_j^{**}) - 10)/2)$

Thus, $\Phi^{-1}(u_j^{**}) = (\ln(x_j^{**}) - 10)/2$ and $x_j^{**} = \exp(2*\Phi^{-1}(u_j^{**})+10)$. This can be found in MS Excel for each u_j^{**} using the input "=EXP(2*NORMSINV(u_j^{**}) + 10)". We use the following inputs:
"=EXP(2*NORMSINV(0.34) + 10)" → 9653.479904
"=EXP(2*NORMSINV(0.12) + 10)" → 2100.701081
"=EXP(2*NORMSINV(0.98) + 10)" → 1339085.962

The average paid loss is thus (9653.479904 + 2100.701081 + 1339085.962)/3 = 450280.0477.

The value by which each paid loss is multiplied to get the unpaid loss is $y = 0.35*2^{0.56}*e^{-0.34*2} = 0.2614114429$. Since this value is a coefficient of every term in the average unpaid loss, it makes sense to simply multiply the average paid loss by this value to get our answer.

Average unpaid loss is 0.2614114429*450280.0477 = **117708.357**.

Section 81

Exam-Style Questions on Simulation

This section affords some exam-style practice regarding cases in which simulation may be applied. A few properties are useful here:

Property 81.1. Where one distribution is gamma and the other is Poisson, the aggregate distribution created from these two distributions is negative binomial.

Property 81.2. Where the population mean is not known directly in a simulation, it can be estimated using the mean of the simulated sample.

Property 81.3. Where the population variance is not known directly in a simulation, it can be estimated using the variance of the simulated sample. *However*, to get the sample variance, one must divide the sum of squares by the *sample size minus one*, not by the sample size.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source:
Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C81-1. Similar to Question 66 of the Exam C Sample Questions from the Society of Actuaries. You are examining the following sample of 6 simulated values of X: 2, 3, 5, 6, 8, 12. You are trying to estimate E(X) with a standard deviation of at most 0.1. Estimate the minimum number of simulations (n) needed to get a standard deviation that low.

Solution S4C81-1. We need to estimate population mean and standard deviation by the sample mean and standard deviation. We find sample mean x^{-} = (2+3+5+6+8+12)/6 = 6.

We find the sample variance as follows (remember that we must divide by sample size minus one!):

$s^2 = ((2-6)^2 + (3-6)^2 + (5-6)^2 + (6-6)^2 + (8-6)^2 + (12-6)^2)/5 = 13.2 \rightarrow s = 3.633180425.$

For n simulations, we estimate the standard deviation as $s/\sqrt{(n)}$, so we want to find the smallest integer n such that $s/\sqrt{(n)} \leq 0.1$. We set $3.633180425/\sqrt{(n)} = 0.1 \rightarrow n = (3.633180425/0.1)^2 = n =$ **1320 simulations**.

Problem S4C81-2. Similar to Question 82 of the Exam C Sample Questions from the Society of Actuaries. The number of claims N follows a probability distribution where Pr(N = n) = $0.8*0.3^{n-1}$ for positive integer values of n. You simulate N using the inversion method (a.k.a. the inverse transformation method), where U has a uniform distribution on the interval from 0 to 1. You observe a value of U to be 0.45. Each claim X has the following probability density function (pdf): $f(x) = (1/30)e^{-x/30}$. You simulate the size of each claim (X_i) using the inversion method, where V_i has a uniform distribution on the interval from 0 to 1. For the ith claims, you get the following values of V_i: $v_1 = 0.34$, $v_2 = 0.34$, $v_3 = 0.78$, $v_4 = 0.67$. Use the simulation data to estimate the total cost of all the observed claims.

Solution S4C81-2. For N, we find Pr(N = 1) = $0.8*0.3^{1-1} = 0.8$, so the observed value of U = 0.45 is clearly less than $F_N(1)$, and thus we have one claim to consider.

Now we estimate the size of the one claim, considering only the value of v_1.

Since $f(x) = (1/30)e^{-x/30}$ is an exponential pdf, the corresponding cumulative distribution function is $F(x) = 1 - e^{-x/30}$. Thus,
$0.34 = 1 - e^{-x/30} \rightarrow x = -30*\ln(1-0.34) =$ **12.46546332.**

Problem S4C81-3. Similar to Question 104 of the Exam C Sample Questions from the Society of Actuaries. You use a two-step simulation process. You simulate λ using a gamma distribution with α = 3 and θ = 4 (implying a mean of 12 and a variance of 48). Then you use λ as the mean of a Poisson distribution used to estimate values of the random variable X. You perform 10000 simulations. How many of these simulations can be expected to result in a value of X = 13?

Solution S4C81-3. We know that an aggregate random variable where one of the component random variables is gamma and the other is Poisson follows a negative binomial distribution.

To find the mean and variance of this distribution, we use the formulas

$E(N) = E_\Lambda(E(N | \Lambda))$ and $Var(N) = E_\Lambda(Var(N | \Lambda)) + Var_\Lambda(E(N | \Lambda))$. Since N follows a Poisson distribution, $Var(N | \Lambda) = E(N | \Lambda) = \Lambda$. Hence, $Var(N) = E_\Lambda(\Lambda) + Var_\Lambda(\Lambda) = 12 + 48 = 60$.

$E_\Lambda(E(N | \Lambda)) = E_\Lambda(\Lambda) = 12$.

For a negative binomial distribution, the mean is equal to rβ, and the variance is equal to rβ(1+β). We note that the variance is 5 times the mean, so 1 + β = 5 and β = 4. Therefore, 4r = 12, and r = 3.

For a negative binomial distribution, $Pr(N = k) = r(r+1)...(r+k-1)*\beta^k/(k!*(1+\beta)^{r+k})$, so

Pr(N = 13) = (3*4*5*6*7*8*9*10*11*12*13*14*15)*4^{13}/(13!*5^{16}) = (15!)*4^{13}/(2!*13!*5^{16}) =

15*14*4^{13}/(2*5^{16}) = 0.0461794884. Thus, for 10000 simulations, we can expect **461.7948837** to give a value of X = 13.

Problem S4C81-4. Similar to Question 122 of the Exam C Sample Questions from the Society of Actuaries. You simulate the number of claims N using a binomial distribution with m = 4 and a mean of 3. You simulate the size of each claim using a discrete distribution with equal probabilities attached to each of the following observations: 3, 4, 9.

You use the inversion method (a.k.a. the inverse transformation method), where U has a uniform distribution on the interval from 0 to 1. You generate the following values of U:

0.5, 0.3, 0.2, 0.7, 0.6, 0.2, 0.1, 0.9, 0.9

You use these values in the following order:

1. Use the first not-yet-used value of U to determine N for the insured. (If the same value of U repeats, you should use the repeated value.) If N = 0, go on to the next value of U to determine N for the next insured.

2. Use the next N not-yet-used values of U to determine the size of each claim for the insured.

3. Repeat the process for the next insured.

What are the total claims for the second insured using this method?

Solution S4C81-4. We first determine the properties of our binomial distribution. Its mean is mq = 3, so q = 3/m = ¾ = 0.75.

Pr(N = 0) = 0.25^4 = 0.00390625.

Pr(N = 1) = 4*0.75*0.25^3 = 0.046875 → F(1) = 0.05078125.

Pr(N = 2) = 6*0.75^2*0.25^2 = 0.2109375 → F(2) = 0.26171875.

Pr(N = 3) = 4*0.75^3*0.25 = 0.421875 → F(3) = 0.68359375.

Naturally, as m = 4, F(4) = 1.

For the first insured, U = 0.5, so N = 3, as F(2) < 0.5 < F(3).

Thus, we can skip the next three values of U, since the question is not asking about claims for the first insured.

For the second insured, U = 0.6, so N = 3, as F(2) < 0.6 < F(3).

The values of U determining claim amounts are 0.2, 0.1, and 0.9.

Since the claim amount values 3, 4, and 9 have equal probability, F(3) = 1/3, F(4) = 2/3, and F(9) = 1. It follows that the first two claims are of amount 3, and the last claim is of amount 9. The total claim amount is 3 + 3 + 9 = **15.**

Problem S4C81-5. Similar to Question 255 of the Exam C Sample Questions from the Society of Actuaries. You are simulating the mean of a nonnegative random variable, where the standard deviation for the population is known to be 0.5 times the population mean. You want to be 97% confident that the simulated mean is within 3% of the actual mean. How many trials of the simulation (n) would you need to achieve this level of confidence? Use the central limit theorem.

Solution S4C81-5. Let X^{-} be the random variable denoting the simulated sample mean. Let μ be the population mean, implying that population standard deviation is 0.5μ. We want to find n such that $\Pr(0.97\mu \leq X^{-} \leq 1.03\mu) = 0.97$.

Thus, $\Pr((0.97\mu - \mu)/(0.5\mu/\sqrt{(n)}) \leq (X^{-} - \mu)/(0.5\mu/\sqrt{(n)}) \leq (1.03\mu - \mu)/(0.5\mu/\sqrt{(n)})) = 0.97 \rightarrow$

$\Pr(-0.03\sqrt{(n)}/0.5 \leq (X^{-} - \mu)/(0.5\mu/\sqrt{(n)}) \leq 0.03\sqrt{(n)}/0.5) = 0.97 \rightarrow$

$\Pr(-0.06\sqrt{(n)} \leq (X^{-} - \mu)/(0.5\mu/\sqrt{(n)}) \leq 0.06\sqrt{(n)}) = 0.97$. By the central limit theorem, if n is large,

$(X^{-} - \mu)/(0.5\mu/\sqrt{(n)})$ approaches the standard normal random variable Z. Therefore,

$\Pr(-0.06\sqrt{(n)} \leq Z \leq 0.06\sqrt{(n)}) = 0.97$, which means that

$0.06\sqrt{(n)} = \Phi^{-1}(1 - (1-0.97)/2) = \Phi^{-1}(0.985)$, which we can find via the Excel input "=NORMSINV(0.985)", for which the result is 2.170090378.

Thus, $2.170090378 = 0.06\sqrt{(n)}$, and n = (2.170090378/0.06)^2 = 1308.136736, implying that the smallest number of trials is **1309**.

Section 82

Estimates for Value-at-Risk and Tail-Value-at-Risk for Simulated Samples

Let $y_1 \leq y_2 \leq ... \leq y_n$ be a simulated sample of size n of a random variable. Let p be the percentile being used. Let k = $\lfloor pn \rfloor$ + 1, where $\lfloor . \rfloor$ is the greatest integer function. Then the following are estimators for value-at-risk and tail value-at-risk.

$\text{VaR}_p\text{^}(X) = y_k$;

$\text{TVaR}_p\text{^}(X) = (1/(n-k+1))_{j=k}{}^n\Sigma(y_j)$.

Let the sample variance in this case be

$s_p{}^2 = (1/(n-k))_{j=k}{}^n\Sigma(y_j - \text{TVaR}_p\text{^}(X))^2$.

The following is an asymptotically unbiased estimate of the variance of $\text{TVaR}_p\text{^}(X)$:

$\text{Var}(\text{TVaR}_p\text{^}(X)) = (s_p{}^2 + p*(\text{TVaR}_p\text{^}(X) - \text{VaR}_p\text{^}(X))^2)/(n-k+1)$.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 21, pp. 656.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C82-1. You have the following sample of simulated values:

93, 109, 120, 123, 150, 153, 189, 190, 195, 200.

Find $\text{VaR}_{0.3}\text{^}(X)$, the estimate of the 30^{th} percentile of X (or value-at-risk at the 30% level).

Solution S4C82-1. We use the formula $\text{VaR}_p\text{^}(X) = y_k$, where p = 0.3, and n = sample size = 10.

We thus find k = $\lfloor pn \rfloor$ + 1 = $\lfloor 0.3*10 \rfloor$ + 1 = 3 + 1 = k = 4, so $VaR_{0.3}$^(X) = y_4 = **$VaR_{0.3}$^(X) = 123**.

Problem S4C82-2. You have the following sample of simulated values:

93, 109, 120, 123, 150, 153, 189, 190, 195, 200.

Find $TVaR_{0.3}$^(X), the estimate of tail value-at-risk at the 30% level.

Solution S4C82-2. We use the formula $TVaR_p$^(X) = $(1/(n-k+1))_{j=k}{}^{n}\Sigma(y_j)$. We found in Solution S4C82-1 that k = 4, so our answer is $(1/(10-4+1))_{j=4}{}^{10}\Sigma(y_j)$ = (1/7)(123 + 150 + 153 + 189 + 190 + 195 + 200) = **$TVaR_{0.3}$^(X) =171.4285714.**

Problem S4C82-3. You have the following sample of simulated values:

93, 109, 120, 123, 150, 153, 189, 190, 195, 200.

Given that the values under consideration exceed the 30th percentile, find the sample variance.

Solution S4C82-3. We seek to find $s_{0.3}{}^2$, which we do using the formula

$s_p{}^2 = (1/(n-k))_{j=k}{}^{n}\Sigma(y_j$ - $TVaR_p$^(X)$)^2$. We found in Solution S4C82-1 that k = 4,. From Solution S4C82-2, we know that $TVaR_{0.3}$^(X) = 171.4285714.

Thus, our answer is (1/(10-4))((123- 171.4285714)^2 + (150- 171.4285714)^2 + (153- 171.4285714)^2 + (189- 171.4285714)^2 + (190- 171.4285714)^2 + (195- 171.4285714)^2 + (200- 171.4285714)^2) = **$s_{0.3}{}^2$ = 861.6190476**.

Problem S4C82-4. You have the following sample of simulated values:

93, 109, 120, 123, 150, 153, 189, 190, 195, 200.

Find Var($TVaR_{0.3}$^(X)), the variance of the estimate of tail value-at-risk at the 30% level.

Solution S4C82-4. We use the formula

Var($TVaR_p$^(X)) = ($s_p{}^2$ + p*($TVaR_p$^(X) - VaR_p^(X)$)^2$)/(n-k+1).

We know the following:

From Solution S4C82-1, we know that k = 4 and $VaR_{0.3}$^(X) = 123.
From Solution S4C82-2, we know that $TVaR_{0.3}$^(X) = 171.4285714.
From Solution S4C82-3, we know that $s_{0.3}{}^2$ = 861.6190476.

Thus, our answer is Var($TVaR_{0.3}$^(X)) = (861.6190476 + (171.4285714-123)^2)/(10-4+1) = **Var($TVaR_{0.3}$^(X)) = 458.1350826**.

Problem S4C82-5. Review of Section 79. Similar to Question 83 of the Exam C Sample Questions from the Society of Actuaries. You are simulating a speculative venture, where the probability is 0.6 that there will be a loss of 600, and the probability of a gain is 0.4. Where there is a gain, the gain is uniformly distributed on the interval from 0 to 5000. You use the inverse transformation (inversion) method to simulate the outcome, where large random numbers from the uniform distribution on the interval from 0 to 1 correspond to large outcomes. For two trials, the observed values from the uniform distribution are 0.4 and 0.9. Find the average gain or loss for these two trials.

Solution S4C82-5. As it is given that F(-600) = 0.6, any trial value less than 0.6 will correspond to a loss of 600. So the trial value of 0.4 corresponds to a loss of 600. Any trial value greater than 0.6 corresponds to a gain that is uniformly distributed from 0 to 5000. The gain associated with a trial value of 0.9 is (0.9-0.6)(5000-0)/(1-0.6) = 3750. The average of these two trials is therefore (-600 + 3750)/2 = **1575.**

Section 83

The Bootstrap Method for Estimating Mean Square Error

In cases where the population distribution is not available to be used in a simulation, the **bootstrap method** can be used to estimate the mean square error of an estimator.

Here is how the bootstrap method works, given a sample of size n and an estimator that uses k of the sample values:

1. Assume a population in which each of the sample values occurs with probability (1/n).

2. For the given sample, find the value Ψ of the quantity being estimated (which can be the mean, variance, third central moment, or anything else that is specified).

3. Find all the possible values Ψ_i of the estimator whose mean square error is being estimated. To do so, determine all the possible samples of n values (with repetition allowed) that can be drawn from the original sample of size n. This value is n^k. Find Ψ_i for each of the n^k samples, and assign to each Ψ_i a probability of $1/n^k$.

4. The bootstrap estimate of mean squared error is $(1/n^k)_i\Sigma(\Psi_i - \Psi)^2$.

Note: The notation above was developed by Mr. Stolyarov to facilitate an easier conceptualization of the bootstrap method.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Sources: Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 21, pp. 657-658.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C83-1. You have a sample of two values: 3 and 5. You take a sample, with repetition possible, of two of these values and use the mean of that sample as the estimate of the given sample mean. That is, you estimate the given sample mean X^{-} as $g(X_1, X_2)$, where X_1 and

X_2 are the values you have picked. Use the bootstrap method to estimate the mean square error of g.

Solution S4C83-1. There are four possible samples to select from the given sample:

(3, 3), (3, 5), (5, 3), and (5, 5). Their means, respectively, are 3, 4, 4, and 5. The mean of the given sample is 4. We estimate the mean squared error of g thus:
$(1/4)((3-4)^2 + 2(4-4)^2 + (5-4)^2) = 2/4 =$ **MSE^(X¯) = ½ = 0.5**.

Problem S4C83-2. Similar to Question 20 of the Exam C Sample Questions from the Society of Actuaries. You are given a sample of 2 values of X: 4 and 18. To estimate Var(X), you are using an estimator $h(X_1, X_2) = (1/2)\Sigma(X_i - X^-)^2$, where X¯ is the sample mean.

Use the bootstrap method to estimate the mean square error of h.

Solution S4C83-2. First, we need to find each of the possible values of h.

The possible samples to be drawn from the 2 given values of X are (4, 4), (4, 18), (18, 4), and (18, 18). For (4, 4) and (18, 18), $h(X_1, X_2) = 0$, since the mean of those samples is equal to each of the sample values. For (4, 18) and (18, 4), $h(X_1, X_2) = (1/2)((4 - 11)^2 + (18 - 11)^2) = 49$.

The population variance is estimated by assuming a population where each of the values 4 and 18 is distributed with probability (1/2). The mean of such a population is 11, so the population variance is $(1/2)((4-11)^2 + (18 - 11)^2) = 49$.

Thus, the estimate of the MSE of h is $(1/4)(2*(0-49)^2 + 2*(49-49)^2) =$ **1200.5.**

Problem S4C83-3. Similar to Question 52 of the Exam C Sample Questions from the Society of Actuaries. In the bootstrapping method, which distribution function is used to estimate the underlying population?

(a) Lognormal
(b) Normal
(c) Empirical
(d) Uniform
(e) Exponential
(f) Whichever distribution the modeler selects.

Solution S4C83-3. The answer is **(c) Empirical**. The use of the empirical distribution function to estimate the underlying population is the basis for the modeler's assignment of a (1/n) probability to each of the given sample values in estimating the population distribution.

Problem S4C83-4. Similar to Question 144 of the Exam C Sample Questions from the Society of Actuaries. You are given the following sample of claim amounts: {43, 56, 98}.

You obtain the following three simulations from this sample:

1) 43, 56, 98
2) 43, 43, 43
3) 56, 98, 98

The policyholder has a deductible of 20, and you estimate that the loss elimination ratio per claim is 0.4. Find the bootstrap estimate of the mean square error of the loss elimination ratio per claim.

Solution S4C83-4. The formula for the loss elimination ratio is $E(X \Lambda d)/E(X)$. Since d = 20, and each of the possible claim amounts exceeds the deductible, it is always the case that our loss elimination ratio for each simulation will be 20/E(X), where E(X) is the mean of the simulated sample.

Thus, for simulation 1), loss elimination ratio is 20/((1/3)(43+56+98)) = 0.3045685279.

For simulation 2), loss elimination ratio is 20/43 = 0.4651162791.

For simulation 3), loss elimination ratio is 20/((1/3)(56+98+98)) = 0.2380952381.

To find the estimate of the MSE, we take the average of the squared differences between each of the loss elimination ratios for our simulations and the estimate of 0.4:
$MSE\text{^} = (1/3)((0.3045685279-0.4)^2 + (0.4651162791-0.4)^2 + (0.2380952381-0.4)^2)$ = **MSE^ =0.0131868159.**

Problem S4C83-5. Similar to Question 175 of the Exam C Sample Questions from the Society of Actuaries. You are given a sample of 3 values of X: 3, 3, and 10. To estimate the third central moment of X you are using an estimator $j(X_1, X_2, X_3) = (1/3)\Sigma(X_i - X^-)^3$, where X^- is the sample mean.

Use the bootstrap method to estimate the mean square error of j.

Solution S4C83-5. The mean of the given sample is (3+3+10)/3 = 16/3.

The third central moment of the given sample is $(1/3)(2(3-16/3)^3 + (10-16/3)^3) = 25.40740741$.

There are $3^3 = 27$ possible samples with repetition that can be drawn from the given sample.

We group them as follows:
Sample with 3 values of 3: Probability is $(2/3)^3 = 8/27$. Mean is 3, and j is 0, since the difference between each value and the mean is 0.

Sample with 3 values of 10: Probability is $(1/3)^3 = 1/27$. Mean is 10, and j is 0, since the difference between each value and the mean is 0.

Sample with 2 values of 3, 1 value of 10: Probability is $3*(2/3)^2*(1/3) = 4/9$. Mean is 16/3, and j = 25.40740741, as with the given sample.

Sample with 2 values of 10, 1 value of 3: Probability is $3*(1/3)^2*(2/3) = 6/27$. Mean is (10+10+3)/3 = 23/3, and $j = (1/3)(2(10-23/3)^3 + (3-23/3)^3) = -25.40740741$.

Thus, we have j = 0 with probability 9/27 = 1/3.

We have j = 25.40740741 with probability 4/9. This is the same as the third central moment of the given sample, so it results in a squared difference of zero.

We have j = -25.40740741 with probability 6/27.

Mean square error is thus $(1/3)(0-25.40740741)^2 + 6/27(-25.40740741-25.40740741)^2$**= MSE^ = 788.9888736.**

Section 84

Simulations of Cumulative Distribution Functions and Exam-Style Questions on Simulation

This section offers additional exam-style practice regarding simulation. Moreover, it introduces another situation which may occur in simulation.

Suppose you are estimating the value of the cumulative distribution function $F_X(k)$ for some specific value k. You are trying to get the estimate to have a maximum error of ± r (where r is a fraction of 1), within a confidence level of c. How many simulations (n) do you need to perform? The following formula sets a lower bound on n:

$n \geq (\Phi^{-1}((1+c)/2)/r)^2*(n - P_n)/P_n$

Meanings of variables:
c = the desired confidence level.
r = the desired maximum error level, as a fraction of 1.
n = total number of simulations.
P_n = number of simulations at or below k, where $F_X(k)$ is being estimated.
$\Phi(x)$ = cumulative standard normal distribution function.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Sources: Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 21, pp. 648.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C84-1. Similar to Question 121 of the Exam C Sample Questions from the Society of Actuaries. Failure time of electronic equipment follows a uniform distribution (DeMoivre's law of mortality) from 0 to Ω. The distribution of Ω has cumulative distribution function $F(\omega) = (\omega/900)^3$ for $0 \leq \omega \leq 900$. You simulate Ω and then the failure time by picking two random values of the random variable U, which is uniformly distributed on the interval from

0 to 1, where smaller values of U correspond to smaller values of Ω and the failure time, as applicable. You then apply the inversion (inverse transformation) method. Your two values of U are 0.3 and 0.9. Find the failure time for this simulation.

Solution S4C84-1. The first value of U, 0.3, corresponds to the distribution of Ω. Therefore, 0.3 = $(\omega/900)^3$ and so ω = $900*0.3^{1/3}$ = 602.4896551. So failure time is uniformly distributed between 0 and 602.4896551. The second value of U, 0.9, corresponds to the distribution of the failure time, which is uniformly distributed. Therefore, the simulated failure time is 0.9*(602.4896551-0) = **542.2406896.**

Problem S4C84-2. Similar to Question 132 of the Exam C Sample Questions from the Society of Actuaries. The number of claims follows a Poisson distribution with mean 2.6. The size of each claim follows a Pareto distribution with α = 3 and θ = 60. An insurance company pays 90% of the total claim amount up to 40, and then 100% of the total claim amount above 100. To simulate the number of claims and size of each claim, you use the inverse transformation (inversion) method, with random variable U, which is uniformly distributed on the interval from 0 to 1, where smaller values of U correspond to smaller values of the relevant variable.
When simulating number of claims, you get U = 0.6.

When simulating size of each claim, you get the following values, in order, of which you use as many values as there are claims: 0.3, 0.7, 0.2, 0.9, 0.1.
Run the simulation to find the total amount the insurance company has to pay.

Relevant property of Poisson distributions: $\Pr(N = k) = e^{-\lambda}*\lambda^k/k!$

Relevant property of Pareto distributions: $F(x) = 1 - \theta^\alpha/(x+\theta)^\alpha$.

Solution S4C84-2. First, we find the relevant probabilities for N, the number of claims.

$\Pr(N = 0) = e^{-2.6} = 0.0742735782$.

$\Pr(N = 1) = 2.6e^{-2.6} = 0.1931113034$, so F(1) = 0.2673848816.

$\Pr(N = 2) = 2.6^2e^{-2.6}/2 = 0.2510446944$, so F(2) = 0.5184295759.

$\Pr(N = 3) = 2.6^3e^{-2.6}/6 = 0.2175720684$, so F(3) = 0.7360016444.

This means that a simulated value of 0.6 corresponds to 3 claims.

The relevant simulated values pertaining to claim amounts are therefore 0.3, 0.7, and 0.2.

Let u be a simulated value. Then

$u = 1 - \theta^\alpha/(x+\theta)^\alpha = 1 - 60^3/(x+60)^3 \rightarrow 1 - u = 60^3/(x+60)^3 \rightarrow (1 - u)^{1/3} = 60/(x + 60) \rightarrow$

$x = 60/(1 - u)^{1/3} - 60$.

Thus, the three simulated claim amounts are

$60/(1 - 0.3)^{1/3} - 60 = 7.574872827$;
$60/(1 - 0.7)^{1/3} - 60 = 29.62809493$;
$60/(1 - 0.2)^{1/3} - 60 = 4.633040701$.

The total claim amount is thus 7.574872827 + 29.62809493 + 4.633040701 = 41.83600846, of which the insurer pays 1.83600846 and 90% of 40 or, altogether 0.9*40 + 1.83600846 = **37.83600846.**

Problem S4C84-3. You are attempting to simulate $F_X(50)$ so that you can be 93% confident that the error is at most ±2%. You know that P_n, the number of simulated observations less than or equal to 50, is 34. Find the minimum total number of simulations.

Solution S4C84-3. We use the formula

$n \geq (\Phi^{-1}((1+c)/2)/r)^2*(n - P_n)/P_n$.

Here, c = 0.93 and r = 0.02. Therefore,

$(\Phi^{-1}((1+c)/2)/r)^2 = (\Phi^{-1}((1.93)/2)/0.02)^2 = (\Phi^{-1}(0.965)/0.02)^2$, which we can find via the Excel input "=(NORMSINV(0.965)/0.02)^2", giving a result of 8207.550717. We also know that P_n = 34, and so at minimum n = 8207.550717*(n - 34)/34 →

34n = 8207.550717n - 279056.7244 →

279056.7244 = 8173.550717n → n = 279056.7244/8173.550717 = 34.1414318. We round this up to get the minimum number of total simulations: **35.**

Problem S4C84-4. Similar to Question 227 of the Exam C Sample Questions from the Society of Actuaries. You are attempting to simulate $F_X(278)$ so that you can be 98% confident that the error is at most ±3%. Let P_n denote the number of simulated values less than or equal to 278. Which of these combinations of n and P_n would result in the desired confidence level and limitation on error? More than one answer may be correct.

(a) n = 300, P_n = 200;
(b) n = 3000, P_n = 2000;
(c) n = 3400, P_n = 1200;
(d) n = 1080, P_n = 700;
(e) n = 14800, P_n = 5200;

Solution S4C84-4. We use the formula

$n \geq (\Phi^{-1}((1+c)/2)/r)^2*(n - P_n)/P_n$.

Here, c = 0.98 and r = 0.03. Therefore,

$(\Phi^{-1}((1+c)/2)/r)^2 = (\Phi^{-1}((1.98)/2)/0.03)^2 = (\Phi^{-1}(0.99)/0.03)^2$, which we can find via the Excel input "=(NORMSINV(0.99)/0.03)^2", giving a result of 6013.216035. Thus,

$n \geq 6013.216035(n - P_n)/P_n \rightarrow$

$n*P_n/(n - P_n) \geq 6013.216035$. We test each of the possible answer choices to see if this condition is met:
(a): $n*P_n/(n - P_n) = 300*200/(300 - 200) = 600 < 6013.216035$, so (a) does not give us the desired conditions.

(b): $n*P_n/(n - P_n) = 3000*2000/(3000 - 2000) = 6000 < 6013.216035$, so (b) does not give us the desired conditions.

(c): $n*P_n/(n - P_n) = 3400*1200/(3400 - 1200) = 1854.545455 < 6013.216035$, so (c) does not give us the desired conditions.

(d): $n*P_n/(n - P_n) = 1080*700/(1080 - 700) = 1989.473684 < 6013.216035$, so (d) does not give us the desired conditions.

(e): $n*P_n/(n - P_n) = 14800*5200/(14800-5200) = 8016.666666667 > 6013.216035$, so **(e) is the only correct answer.**

Problem S4C84-5. You are attempting to simulate $F_X(100)$ so that you can be 95% confident that the error is at most ±1%. You know that n, the total number of simulated observations, is 3000. Find the largest number of these observations that can be at or below 100 in order for the desired confidence level and limitations on error to hold.

Solution S4C84-5. We use the formula

$n \geq (\Phi^{-1}((1+c)/2)/r)^2*(n - P_n)/P_n$.

Here, c = 0.95 and r = 0.01. Therefore,

$(\Phi^{-1}((1+c)/2)/r)^2 = (\Phi^{-1}((1.95)/2)/0.01)^2 = (\Phi^{-1}(0.975)/0.01)^2$, which we can find via the Excel input "=(NORMSINV(0.975)/0.01)^2", giving a result of 38414.58821.

Thus, $n \geq 38414.58821(n - P_n)/P_n \rightarrow$
$3000 \geq 38414.58821(3000 - P_n)/P_n$.

At the largest value of P_n, $3000 = 38414.58821(3000 - P_n)/P_n \rightarrow$
$3000P_n = 38414.58821(3000 - P_n) \rightarrow$
$115243764.6 = 41414.58821*P_n \rightarrow$

$P_n = 115243764.6/41414.58821 = 2782.685271$, which we round down to the largest whole number. Therefore, our answer is **2782 observations.**

Section 85

Assorted Exam-Style Questions for Exam 4/C – Part 3

This section offers exam-style practice with a variety of items covered in the study guide thus far. Now that we have introduced virtually all the material that is likely to be on the exam, the next step in students' preparation should concentrate on extensive practice with each of the syllabus topics so that students might attain the ability to solve exam-level problems using internalized knowledge.

A few reminders are in order.

Note 85.1. When, in a Poisson distribution, the parameter λ follows a gamma distribution with parameters α and θ, the overall distribution is negative binomial with parameters $r = \alpha$ and $\beta = \theta$.

Note 85.2. When using classical credibility, the number of observations required for full credibility of aggregate data is $n = \lambda_0*(\sigma_f^2/\mu_f + \sigma_s^2/\mu_s^2)$, where $\lambda_0 = (\Phi^{-1}((1+p)/2)/k)^2$, p is desired confidence level, k is the desired maximum error, μ_f and μ_s are, respectively, the mean frequency and mean severity, and σ_f and σ_s are, respectively, the standard deviations of frequency and severity.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Sources: Exam C Sample Questions and Solutions from the Society of Actuaries.

Mahler, H.C.; and Dean, C.G., "Credibility," *Foundations of Casualty Actuarial Science* (Fourth Edition), 2001, Casualty Actuarial Society, Chapter 8, Sections 2.5 and 2.6.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C85-1. Similar to Question 23 of the Exam C Sample Questions from the Society of Actuaries. You have gathered 9 observations from three intervals, distributed as follows:

There are 3 observations in the interval from 0 to 14.
There are 3 observations in the interval from 14 to 30.
There are 3 observations greater than 30.

You know that observations follow a uniform distribution from 0 to θ.

You estimate θ as follows: you minimize the function q = $_{j=1}^{3}\Sigma((E_j-O_j)^2/O_j)$, where E_j is the expected number of observations, and O_j is the actual number of observations for the jth interval. Perform this minimization procedure to obtain the estimate of θ.

Solution S4C85-1. Naturally, since there exist observations greater than 30, θ must be greater than 30 as well. If we have nine total observations, here is what we would expect on each interval:

For the interval from 0 to 14, the expected number of observations is 9*(14/θ) = 126/θ.
For the interval from 14 to 30, the expected number of observations is 9*((30-14)/θ) = 144/θ.
For the interval from 30 to θ, the expected number of observations is 9*((θ-30)/θ) = 9 - 270/θ.
Thus, our function $q = ((126/\theta - 3)^2/3) + ((144/\theta - 3)^2/3) + ((9 - 270/\theta - 3)^2/3) \rightarrow$
$q = ((126/\theta - 3)^2/3) + ((144/\theta - 3)^2/3) + ((6 - 270/\theta)^2/3)$.
Thus, $q' = 2(-126/\theta^2)(126/\theta - 3)/3 + 2(-144/\theta^2)(144/\theta - 3)/3 + 2(270/\theta^2)(6 - 270/\theta)/3)$.

We set q' to 0 to find the minimum. We note that a denominator of $3\theta^2$ is common to each of the three expressions whose sum is q', so we can simply diving q' by $3\theta^2$, and the result will still equal zero.

Therefore, $2(-126)(126/\theta - 3) + 2(-144)(144/\theta - 3) + 2(270)(6 - 270/\theta) = 0 \rightarrow$
$(-126)(126/\theta - 3) + (-144)(144/\theta - 3) + (270)(6 - 270/\theta) = 0 \rightarrow$
$(270)(6 - 270/\theta) = (126)(126/\theta - 3) + (144)(144/\theta - 3) \rightarrow$
$1620 - 72900/\theta = 15876/\theta - 378 + 20736/\theta - 432 \rightarrow$
$2430 = 109512\theta \rightarrow$
109512/2430 = **θ = 45.066666667**.

Problem S4C85-2. Similar to Question 39 of the Exam C Sample Questions from the Society of Actuaries. Frequency of losses follows a Poisson distribution with mean Λ. The prior distribution of Λ is a gamma distribution with α = 3 and θ = 0.9. Severity follows an exponential distribution with mean 4000. You want a full credibility standard for aggregate losses to have a probability of 0.98 of being within 2% of the expected value. Using classical (limited fluctuation) credibility theory, find the minimum number of observations required for full credibility. Refer to the Exam 4 / C Tables as necessary.

Solution S4C85-2. The formula we will ultimately use is $n = \lambda_0*(\sigma_f^2/\mu_f + \sigma_s^2/\mu_s^2)$.

First, we find $\lambda_0 = (\Phi^{-1}((1+p)/2)/k)^2$, where p = 0.98 and k = 0.02, so

$\lambda_0 = (\Phi^{-1}((1.98)/2)/0.02)^2 = (\Phi^{-1}(0.99)/0.02)^2$, which we find in Excel using the input "=(NORMSINV(0.99)/0.02)^2", of which the result is 13529.73608.

We analyze characteristics of the severity distribution. In an exponential distribution, the variance is the square of the mean, so standard deviation is equal to the mean. Therefore, $\sigma_s^2/\mu_s^2 = 1^2 = 1$.

We analyze characteristics of the frequency distribution. By Note 85.1, the combined Poisson-gamma distribution is negative binomial with $r = \alpha = 3$ and $\beta = \theta = 0.9$. The mean of the negative binomial distribution is $r\beta = 3*0.9 = 2.7$. The variance is $r\beta(1+\beta) = 2.7*1.9$. Therefore, $\sigma_f^2/\mu_f = 2.7*1.9/2.7 = 1.9$. Hence, our standard for full credibility is n = 13529.73608(1.9 + 1) = 39236.23463, which we round up to **39237 observations**.

Problem S4C85-3. Similar to Question 46 of the Exam C Sample Questions from the Society of Actuaries. You are aware of the following loss data, where a "+" superscript indicates that the loss has exceeded the policy limit:
6, 7, 8, 8, 9, 9, 9+, 10, 20, 23, 23+, 25, 25, 30, 30+.

Use the Kaplan-Meier product-limit estimator to find the probability that a loss exceeds 24.

Solution S4C85-3. For the losses that exceed the policy limit, we only know that they exceed the limit, not what the actual amount of loss is. Thus, where the observation under consideration is *less than the policy limit* for such losses, we can consider the excess-loss observations as part of the risk set. But we cannot determine a risk set for the excess-loss observations themselves, since we do not know the magnitude of the actual losses associated with them.

We are trying to estimate S(24), for which we use the following formulas:

$S_n(t) = 1$ if $0 \leq t < y_1$;
$S_n(t) = {}_{i=1}{}^{j-1}\Pi((r_i - s_i)/r_i)$ if $y_{j-1} \leq t < y_j$ for j = 2, ..., k;
$S_n(t) = {}_{i=1}{}^{k}\Pi((r_i - s_i)/r_i)$ or 0 for $y_k \leq t$.

For 6, $s_i = 1$ and $r_i = 15$, so $S(6) \approx 14/15$.
For 7, $s_i = 1$ and $r_i = 14$, so $S(7) \approx (14/15)(13/14) = 13/15$.
For 8, $s_i = 2$ and $r_i = 13$, so $S(8) \approx (13/15)(11/13) = 11/15$.
For 9, $s_i = 2$ and $r_i = 11$, so $S(8) \approx (11/15)(9/11) = 9/15$.
For 10, $s_i = 1$ and $r_i = 8$, so $S(9) \approx (9/15)(7/8) = 21/40$.
For 20, $s_i = 1$ and $r_i = 7$, so $S(9) \approx (21/40)(6/7) = 9/20$.
For 23, $s_i = 1$ and $r_i = 6$, so $S(23) \approx (9/20)(5/6) = 3/8$.

Since no losses are *known* to occur between 23 and 24, we estimate S(4) to be **3/8 or 0.375**.

Problem S4C85-4. Similar to Question 55 of the Exam C Sample Questions from the Society of Actuaries. You have the following information about groups elephants:
Group A has 80 elephants. The probability of a group A elephant eating no peanuts is 0.23.

The probability of a group A elephant eating 1 peanut is 0.45. The probability of a group A elephant eating 2 peanut is 0.32.
Group B has 100 elephants. The probability of a group B elephant eating no peanuts is 0.33. The probability of a group B elephant eating 1 peanut is 0.12. The probability of a group B elephant eating 2 peanuts is 0.55.

You observe an elephant eating 1 peanut on day 1. What is the expected number of peanuts this elephant will eat on day 2?

Solution S4C85-4.

There are 180 elephants in all, so the probability of a randomly chosen elephant being in Group A is 80/180 = Pr(A) = 4/9, while Pr(B) = 5/9.

Given that the elephant has eaten 1 peanut, we want to find the probability that it came from each of the two groups. By Bayes's Theorem, Pr(A │1) = Pr(A and 1)/Pr(1).

We find Pr(1) = Pr(A)*Pr(1│A) + Pr(B)*Pr(1│B) = (4/9)*0.45 + (5/9)*0.12 = 4/15.
Thus, Pr(A │1) = (4/9)*0.45/(4/15) = ¾.
Pr(B │1) = Pr(B and 1)/Pr(1) = (5/9)*0.12/(4/15) = ¼.
E(A) = 0*0.23 + 1*0.45 + 2*0.32 = 1.09.
E(B) = 0*0.33 + 1*0.12 + 2*0.55 = 1.22.

Thus, for the elephant in question the expected number of peanuts eaten on day 2 is (3/4)*1.09 + (1/4)*1.22 = **1.1225.**

Problem S4C85-5. Similar to Question 56 of the Exam C Sample Questions from the Society of Actuaries. You are given the following information about claim payments and policy limits on the policies for which the respective claims were paid:

Claim Payment..........Policy Limits
2..............................3
3..............................3
3.............................120
3.............................500
120..........................120
120..........................220.

Let f(x) be the probability density function for loss amount; let F(x) be the cumulative distribution function for loss amount. Find the likelihood function for this data in terms of values of f(x) and F(x), where appropriate.

Solution S4C85-5. Every time a claim payment is equal to the loss x, a factor of f(x) is contributed to the likelihood function. Every time a claim payment is capped at the limit, we do not know the loss x, so our best estimate for the factor contributing to the likelihood function is (1-F(x)). We know that claim payments are capped at the limit when we observe a claim payment equaling the policy limit. Therefore, we have the following likelihood function:
f(2)*(1-F(3))*f(3)*f(3)*(1-F(120))*f(120).

Section 86

Assorted Exam-Style Questions for Exam 4/C – Part 4

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C86-1. Similar to Question 58 of the Exam C Sample Questions from the Society of Actuaries. The number of pet grooming liability claims per insured pet groomer follows a Poisson distribution with mean Λ. The prior distribution of Λ has the following probability density function: $f_\Lambda(\lambda) = (30\lambda)^{12} * e^{-30\lambda}/(\lambda * \Gamma(12))$.

In year 1, there were 23 insured pet groomers and 4 pet grooming liability claims.

In year 2, there were 67 insured pet groomers and 20 pet grooming liability claims.

It is expected that there will be 100 insured pet groomers in year 3. Find the expected number of pet grooming liability claims for this group of 100 insured pet groomers. Refer to the Exam 4 / C Tables as necessary.

Solution S4C86-1. From the exam tables, we note that $f_\Lambda(\lambda)$ has the format of a pdf of a gamma distribution with α = 12 and θ = (1/30). Since the prior distribution is gamma and the model distribution is Poisson, we are working with a common case of conjugate prior distributions (Case 1 from Section 65):

Prior distribution of λ is gamma with parameters α and θ.
Model distribution is Poisson with parameter λ.
Posterior distribution is gamma with parameters $\alpha + \Sigma x_i$ and $\theta/(n\theta+1)$.

We find n = number of exposures in the two years for which we have observations = 67 + 23 = 90.

We also find Σx_i = number of claims in the two years for which we have observations = 4 + 20 = 24.

Thus, the posterior distribution is gamma and has parameters $\alpha^* = \alpha + \Sigma x_i = 12 + 24 = 36$ and $\theta^* = \theta/(n\theta+1) = (1/30)/(90/30 + 1) = 1/120$.

The mean of the posterior distribution *per insured* is $(\alpha^*)(\theta^*) = 36/120 = 0.3$, and so for 100 insureds, the expected number of claims is 100*0.3 = **30 claims.**

Problem S4C86-2. Similar to Question 59 of the [Exam C Sample Questions] from the Society of Actuaries. Construct a p-p plot by adhering to the following instructions. Remember that you are plotting two cumulative distribution functions (cdfs) against one another:

- Draw the horizontal axis and label it "Sample distribution." The values on the axis should range from 0 to 1.

- Draw the vertical axis and label it "Fitted distribution." The values on the axis should range from 0 to 1.

- Draw a 45-degree (slope = 1) line from the origin to the point (1, 1).

- Draw two points on the 45-degree line at (0.2, 0.2) and (0.8, 0.8).

- Now you will draw the actual p-p plot. Draw an always increasing curve between the origin and (0.2, 0.2) such that the entire curve is below the 45-degree line.

- Draw an always increasing curve between (0.2, 0.2) and (0.8, 0.8) such that the entire curve is above the 45-degree line.

- Draw an always increasing curve between (0.8, 0.8) and (1, 1) such that the entire curve is below the 45-degree line.

Answer the following questions about the fitted distribution:

(a) Does the fitted distribution have a lighter or heavier left tail than the sample distribution in the region from the origin to (0.2, 0.2)?
(b) Does the fitted distribution place greater or smaller probabilities on the "middle range" between (0.2, 0.2) and (0.8, 0.8) than the sample distribution?
(c) Does the fitted distribution have a lighter or heavier right tail than the sample distribution in the region from (0.8, 0.8) to (1, 1)?

Solution S4C86-2. We answer the questions in sequence:
(a) In the region from the origin to (0.2, 0.2), the curve is below the 45-degree line, meaning that the fitted probabilities are smaller than the sample probabilities, which means the fitted distribution emphasizes the left tail less and has a **lighter left tail.**

(b) In the region between (0.2, 0.2) and (0.8, 0.8), the curve is above the 45-degree line, meaning that the fitted probabilities are larger than the sample probabilities, which means that the fitted distribution emphasizes the "middle range" more and **places greater probabilities on the "middle range"** than the sample distribution.

(c) In the region between (0.8, 0.8) and (1, 1), the fitted distribution assigns smaller probabilities than the sample distribution for being below a particular value (as the p-p plot is a plot of cumulative distribution functions). This means that the fitted distribution assigns *larger* probabilities for being *above* the value in question, meaning that the fitted distribution emphasizes the right tail more than the sample distribution and so has a **heavier right tail.**

Problem S4C86-3. Similar to Question 60 of the Exam C Sample Questions from the Society of Actuaries. You know the following about three dice:
Die A has 0.2 probability of value 0, 0.34 probability of value 1, 0.46 probability of value 2.

Die B has 0.5 probability of value 0, 0.4 probability of value 1, 0.1 probability of value 2.

Die C has 0.43 probability of value 0, 0.55 probability of value 1, 0.02 probability of value 2.
You pick one die at random and observe the following sequence of rolls S: S = {0, 2, 1, 1}.

Find the posterior expected value of the next roll of the die.

Solution S4C86-3. We want to find E(Next roll │S). To do this, we need to find the posterior probability of each possible outcome.

We note that Pr(S and Die A) = Pr(Die A)*Pr(S │ Die A) = (1/3)(0.2*0.46*0.34^2) = 0.0035450667.

Pr(S and Die B) = Pr(Die B)*Pr(S │ Die B) = (1/3)(0.5*0.1*0.4^2) = 0.00266666667.

Pr(S and Die C) = Pr(Die C)*Pr(S │ Die C) = (1/3)(0.43*0.02*0.55^2) = 0.00086716666667.

Posterior probability of A is Pr(S and Die A)/Pr(S) =

0.0035450667/(0.0035450667 + 0.00266666667 + 0.00086716666667) = 0.5007934378.

Posterior probability of B is Pr(S and Die B)/Pr(S) =

0.00266666667/(0.0035450667 + 0.00266666667 + 0.00086716666667) = 0.3767063621.

Posterior probability of C is Pr(S and Die C)/Pr(S) =

0.00086716666667/(0.0035450667 + 0.00266666667 + 0.00086716666667) = 0.1225002001.

Moreover, E(A) = 0.2*0 + 0.34*1 + 0.46*2 = 1.26.

E(B) = 0.5*0 + 0.4*1 + 0.1*2 = 0.6.

E(C) = 0.43*0 + 0.55*1 + 0.02*2 = 0.59.

Thus, our posterior expected value is

0.5007934378*1.26 + 0.3767063621*0.6 + 0.1225002001*0.59 = **0.9292986669.**

Problem S4C86-4. Similar to Question 61 of the Exam C Sample Questions from the Society of Actuaries. You observe the following twelve raw claim amounts:
20, 20, 25, 26, 27, 35, 37, 45, 67, 89, 100, 103

Each claim is subject to a deductible of 10. You also know that raw claim amounts follow an exponential distribution. Use maximum likelihood estimation to find the average amount *the insurer will have to pay out* per claim.

Solution S4C86-4. The exponential distribution has the memoryless property, which means that the excess over a certain value (say, our deductible of 10) follows an exponential distribution if the original values follow an exponential distribution. This means that we can present our sample data as data for claims minus deductible:
10, 10, 15, 16, 17, 25, 27, 35, 57, 79, 90, 93.

This sample is also presumed to follow an exponential distribution. The maximum likelihood estimate for the mean θ of an exponential distribution is equal to the sample mean. In this case, our desired estimate is (10 + 10 + 15 + 16 + 17 + 25 + 27 + 35 + 57 + 79 + 90 + 93)/12 = **39.5.**

Problem S4C86-5. Similar to Question 68 of the Exam C Sample Questions from the Society of Actuaries.

You have the following data for two groups of hippopotami, A and B:
For Group A:

In year 1, 20 hippopotami fly to the moon, out of 90 that could have.

In year 2, 40 hippopotami fly to the moon, out of 70 that could have.

In year 3, 10 hippopotami fly to the moon, out of 30 that could have.

For Group B:

In year 1, 230 hippopotami fly to the moon, out of 1000 that could have.

In year 2, 320 hippopotami fly to the moon, out of 770 that could have.

In year 3, 150 hippopotami fly to the moon, out of 450 that could have.

Let $\hat{S}_T(3)$ be the Kaplan-Meier product-limit estimate of the survival function (i.e., the proportion of hippopotami *from both groups* that did *not* fly to the moon after all three years have elapsed.

Let $\hat{S}_B(3)$ be the Kaplan-Meier product-limit estimate of the survival function (i.e., the proportion of hippopotami *from Group B* that did *not* fly to the moon after all three years have elapsed.

Find $\hat{S}_T(3)$ - $\hat{S}_B(3)$.

Solution S4C86-5. In each year, our risk set is the number of hippopotami that could have flown to the moon; our s_i is the number of hippopotami that actually flew to the moon.

We use the formula for the Kaplan-Meier product-limit estimator:

$S_n(t) = 1$ if $0 \leq t < y_1$;
$S_n(t) = {}_{i=1}{}^{j-1}\Pi((r_i - s_i)/r_i)$ if $y_{j-1} \leq t < y_j$ for j = 2, ..., k;
$S_n(t) = {}_{i=1}{}^{k}\Pi((r_i - s_i)/r_i)$ or 0 for $y_k \leq t$.

Thus, we find $\hat{S}_B(3)$ = ((1000-230)/1000)*((770-320)/770)*((450-150)/450) = 300/1000 = 0.3.

We find $\hat{S}_T(3)$ = (((1000+90)-(230+20))/(1000+90))*((770+70)-(40+320))/(770+70))*((450+30)-(150+10))/(450+30)) = ((450+30)-(150+10))/(1000+90) = 0.2935779817.

Thus, $\hat{S}_T(3)$ - $\hat{S}_B(3)$ = 0.2935779817 - 0.3 = **-0.0064220183.**

Section 87

The Life Table Methodology for Estimating Risk Sets and Assorted Exam-Style Questions for Exam 4/C

This section provides additional exam-style practice with a variety of syllabus topics.

Note 87.1. The **life table methodology** for estimating the size of the risk set when using the Kaplan-Meier approximation for large data sets assumes the following:
When there are right censored observations within a particular interval whose risk set is being calculated, *half* of the right censored observations within the interval go into the risk set. The rest of the risk set calculation follows the usual procedure for the Kaplan-Meier approximation.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Revelle, Dave. Discussion of SOA Question 73.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C87-1. Similar to Question 69 of the Exam C Sample Questions from the Society of Actuaries. You have the following sample of values:

12, 21, 24, 25, 53, 65, 67, 99, 100

You fit these values to an exponential distribution. Find the coefficient of variation for the maximum likelihood estimate of this distribution.

Solution S4C87-1. Coefficient of variation = (Standard deviation)/(Mean).
The maximum likelihood estimate θ^ of the mean θ for an exponential distribution is the sample mean. The variance we want to find is the *variance of the maximum likelihood estimate*, i.e., the variance of the sample mean, which is (Estimated variance of distribution)/(sample size), or, in this case, $\theta\text{^}^2/n$. Thus, the estimate of standard deviation is $\sqrt{}(\theta\text{^}^2/n) = \theta\text{^}/\sqrt{}(n)$, and
Coefficient of variation = (θ^/√(n))/(θ^) = 1/√(n). Here, the sample size is n = 9, so our answer is

$1/\sqrt{(9)}$ = **1/3.**

Problem S4C87-2. You have the following data:
In the interval (0, 4000], there are 89 uncensored observations and 94 observations that are right censored within the interval.

In the interval (4000, 10000], there are 23 uncensored observations and 83 observations that are right censored within the interval.

In the interval (10000, 90000], there are 4 uncensored observations and 50 observations that are right censored within the interval.

What is the size of risk set in the interval (0, 4000], according to the life table methodology?

Solution S4C87-2. By the life table methodology, the risk set includes all uncensored observations that are within the interval, all observations that are known to be greater than the values in the interval, *plus* one-half of all the censored observations in the interval.

The number of uncensored observations within (0, 4000] is 89.

The number of observations that are known to be greater than the values in the interval is 23+83+4+50 = 160.

One-half of all the censored observations in the interval is equal to 94/2 = 47.

Thus, according to the life table methodology, the size of the risk set for (0, 4000] is 89 + 160 + 47 = **296**.

Problem S4C87-3. Similar to Question 73 of the Exam C Sample Questions from the Society of Actuaries. You have the following data:

In the interval (0, 4000], there are 89 uncensored observations and 94 observations that are right censored within the interval.

In the interval (4000, 10000], there are 23 uncensored observations and 83 observations that are right censored within the interval.

In the interval (10000, 90000], there are 4 uncensored observations and 50 observations that are right censored within the interval.

Use the Kaplan-Meier approximation for large data sets and the life table methodology for determining the size of risk sets to estimate S(10000), the probability that a randomly selected value will exceed 10000.

Solution S4C87-3. For each interval, the value of s_i is the number of uncensored observations within the interval (as always in the Kaplan-Meier approximation). The value of r_i is the size of the risk set, determined using the life table methodology. Each interval containing values less than 10000 contributes a factor of $(r_i - s_i)/r_i$ for that interval. For the interval (0, 4000], $s_i = 89$, $r_i = 296$ (from Solution S4C87-2), and so the contributing factor is (296-89)/296 = 207/296.

For the interval (4000, 10000], the size of the risk set is as follows:

The number of uncensored observations within (4000, 10000] is 23.

The number of observations that are known to be greater than the values in the interval is 4+50 = 54.

One-half of all the censored observations in the interval is equal to 83/2 = 41.5.

Thus, the size of the risk set is 23 + 54 + 41.5 = 118.5, and the contributing factor for the interval is (118.5-23)/118.5 = 191/237. Thus, our estimate for S(10000) is (207/296)(191/237) = **0.5635904892.**

Problem S4C87-4. Similar to Question 76 of the Exam C Sample Questions from the Society of Actuaries. You know that losses follow a Poisson distribution with mean θ. The prior distribution of θ is $f(\theta) = \theta e^{-5\theta}$ for θ > 0. You also know that $_0{}^{\infty}\int \theta e^{-n\theta} * d\theta = 1/n^2$. You observe that a policyholder has sustained at least one loss this year. Find the posterior probability that the policyholder will sustain more than one loss next year.

Solution S4C87-4. We first need to find the posterior distribution.

We know that the prior distribution is $f(\theta) = \theta e^{-5\theta}$.

The model distribution is the Poisson probability that the policyholder has sustained at least one loss this year. This is 1 - Pr(0 losses) = $1 - e^{-\theta}$.

Joint distribution = (Prior distribution)*(Model distribution) = $(\theta e^{-5\theta})(1 - e^{-\theta}) = \theta e^{-5\theta} - \theta e^{-6\theta}$.

Marginal distribution = $_0{}^{\infty}\int$(Joint distribution) = $_0{}^{\infty}\int(\theta e^{-5\theta} - \theta e^{-6\theta}) * d\theta$ = (by our given information) $1/5^2 - 1/6^2 = 11/900$.
Posterior distribution = (Joint distribution)/(Marginal distribution) = $(\theta e^{-5\theta} - \theta e^{-6\theta})/(11/900) = (900/11)(\theta e^{-5\theta} - \theta e^{-6\theta})$.

Our answer is E(Pr(More than one loss)) =

$_0{}^{\infty}\int \Pr(\text{More than one loss}) * (900/11)(\theta e^{-5\theta} - \theta e^{-6\theta}) * d\theta =$

$(900/11)\,_0{}^{\infty}\int(1 - e^{-\theta})(\theta e^{-5\theta} - \theta e^{-6\theta}) * d\theta = (900/11)\,_0{}^{\infty}\int(\theta e^{-5\theta} - 2\theta e^{-6\theta} + \theta e^{-7\theta}) * d\theta =$

$(900/11)(1/5^2 - 2/6^2 + 1/7^2)$ = **214/539 = 0.3970315399**.

Problem S4C87-5. Similar to Question 79 of the Exam C Sample Questions from the Society of Actuaries. Losses come from a distribution that is a mixture of two Pareto distributions - distribution A with $\alpha = 4$ and $\theta = 30$ and distribution B with $\alpha = 6$ and $\theta = 50$.

Distribution A occurs with probability p, and distribution B occurs with probability (1-p).

Two losses of 10 and 20 were observed. You know that p = 0.3. Find the value of the likelihood function for this set of observations.

Relevant property of Pareto distributions: $f(x) = \alpha\theta^{\alpha}/(x+\theta)^{\alpha+1}$.

Solution S4C87-5. We know that the likelihood function is the product f(10)*f(20). What is f(x)? A mixture of two Pareto distributions is a probability-weighted average of these two distributions. In our case, $f(x) = p*4*30^4/(x+30)^5 + (1-p)*6*50^6/(x+50)^7 =$

$3240000p/(x+30)^5 + 93750000000(1-p)/(x+50)^7$. The likelihood function is

$f(10)*f(20) = (3240000p/(10+30)^5 + 93750000000(1-p)/(10+50)^7)*$

$(3240000p/(20+30)^5 + 93750000000(1-p)/(20+50)^7) =$

(0.031640625p + 0.0334897977(1-p))*(0.010368p + 0.0113837407(1-p)) =

(0.031640625*0.3 + 0.0334897977*0.7)*(0.010368*0.3 + 0.0113837407*0.7) = **0.000368879831.**

Section 88

Assorted Exam-Style Questions for Exam 4/C – Part 5

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C88-1. Similar to Question 85 of the Exam C Sample Questions from the Society of Actuaries. You know that the number of accidents an individual insuredsuffers in a year follows a Poisson distribution with mean 0.1. The magnitude of loss from each accident has mean 500 and standard deviation 1000. Frequency and severity of losses are independent. At least how many insureds must the company have in order that the probability of losses exceeding 1.5 times the expected amount is no greater than 0.05? Use a normal approximation and the central limit theorem to obtain your answer.

Solution S4C88-1. First, we find the mean loss per insured.
Let N = frequency of losses; Let X = severity of losses. Let S = total loss to individual insured.
E(S) = E(N)*E(X) = 0.1*500 = 50. For n insureds, the expected loss is therefore 50n, and 1.5 times the expected loss is 75n.
Now we find the variance of the loss per insured.
Since the frequency distribution is Poisson, Var(S) = E(N)*E(X^2).
E(X^2) = Var(X) + E(X)^2 = 1000^2 + 500^2 = 1250000.
Thus, Var(S) = 0.1*1250000 = 125000. The variance for n insureds is thus 125000n, and the standard deviation is 353.5533906√(n).
We want to find n such that 0.05 = Pr(Total loss ≥ 75n).
By the central limit theorem, 0.05 = 1 - Φ((75n - 50n)/(353.5533906√(n))) =
1 - Φ((25n)/(353.5533906√(n))).

Thus, 0.95 = Φ(0.0707106781√(n)) and 0.0707106781√(n) = Φ^{-1}(0.95), which we find in MS Excel via the input "=NORMSINV(0.95)". We get 0.0707106781√(n) = 1.644853627 and n = $(1.644853627/0.0707106781)^2$ = 541.1086911, which we round up to **542 insureds**.

Problem S4C88-2. Similar to Question 86 of the Exam C Sample Questions from the Society of Actuaries. Frequency of losses follows a Poisson distribution with mean 0.3. Severity of losses follows a uniform distribution on the interval from 0 to 900. The insurance company pursues the following plan: For every loss there is a deductible of 100, and only 80% of each loss above the deductible is paid by the insurer. Find the mean of the frequency of *losses for which the insurance company has to pay.*

Solution S4C88-2. If the insurer has to pay for some losses, what percentage of the losses the insurer has to pay has no reflection on the frequency of the paid losses. Thus, the 80% payment figure for paid losses is not material to the question at hand. We consider that losses from 0 to 100 - or 100/900 = 1/9 of the losses - are excluded from payment altogether because of the deductible, which means that 8/9 of the losses are eligible for payment. This means that the mean of paid losses is 8/9 of the mean of incurred losses or (8/9)*0.3 = **4/15 = 0.2666666667**.

Problem S4C88-3. Similar to Question 87 of the Exam C Sample Questions from the Society of Actuaries. Losses (X) follow a uniform distribution with f(x) = 0.06 on the interval from 0 to 6. After x = 6, the probability density function (pdf) for losses decreases linearly until reaching f(x) = 0 at x = 21.33333333. An insurance company implements a deductible of 3 for each loss. Find the loss elimination ratio for this deductible.

Solution S4C88-3. Loss elimination ratio = E(X Λ d)/E(X) = E(X Λ 3)/E(X) in this case.
Determining E(X Λ 3) is straightforward.
Pr(X < 3) = 0.06*3 = 0.18, so E(X Λ 3) = $_0{}^3\int$x*f(x)*dx + (1-0.18)*3 =
$_0{}^3\int$0.06x*dx + 2.46 = $0.03x^2 \Big|_0^3$ + 2.46 = 0.27 + 2.46 = E(X Λ 3) = 2.73.

To find E(X), we first need to find f(x) past x = 6.
We know that f(x) is a linear function. Its slope is -0.06/(21.33333333-6) = -9/2300

Its (extrapolated) y-intercept can be found by recalling the familiar formula y = mx + b → b = y - mx. We know a point on the line: (6, 0.06). Thus, b = 0.06 + (9/2300)*6 = 48/575, and so, between 6 and 21.3333333333, f(x) = (-9/2300)x + 48/575.

need to take $_0{}^{21.33333333}\int$x*f(x)*dx = $_0{}^6\int$0.06x*dx + $_6{}^{21.33333333}\int$((-9/2300)x + 48/575)x*dx =

$0.03x^2 \Big|_0^6 + ((-3/2300)x^3 + (24/575)x^2) \Big|_6^{21.33333333}$ = E(X) = 14.44592593.

Thus, the loss elimination ratio is 2.73/14.44592593 = **0.1889806174.**

Problem S4C88-4. Similar to Question 88 of the Exam C Sample Questions from the Society of Actuaries. Gamblers at a casino can lose in three ways.

A small loss occurs with probability 0.3 and has loss amount 900.

A medium loss occurs with probability 0.5 and has loss amount 4000.

A large loss occurs with probability 0.2 and has loss amount 15000.

Each gambler incurs one and only one of these three losses when visiting the casino.

The number of gamblers at the casino follows a Poisson distribution with mean 5600.

As an incentive to keep gamblers in the casino after they lose money, the casino reimburses them for 20% of their losses.

Use the central limit theorem to find the probability that total losses of all the gamblers (minus the total reimbursements) exceed 24,000,000.

Solution S4C88-4. Let N be the frequency of losses (i.e., the number of gamblers). Since N is Poisson, E(N) = Var(N) = 5600.

Let X be the severity of losses: E(X) = 900*0.3 + 4000*0.5 + 15000*0.2 = 5270.
$E(X^2) = 900^2*0.3 + 4000^2*0.5 + 15000^2*0.2 = 53243000$.
Let S be the total losses.
E(S) = E(N)*E(X) = 5600*5270 = 29512000.

Since the frequency distribution is Poisson, $Var(S) = E(N)*E(X^2) = 5600*53243000 = 298160800000$, and so the standard deviation of S is√(298160800000) = 546041.0241.

If the total retained losses of the gamblers are 24000000, the total losses prior to reimbursement must be 24000000/(1-0.2) = 30000000. Thus, we seek to find Pr(S ≥ 30000000) =

1-Φ((30000000-29512000)/546041.0241) = 1-Φ(0.8937057446), which we find via the Excel input "=1-NORMSDIST(0.8937057446)", getting as our answer **0.185739675**.

Problem S4C88-5. Similar to Question 89 of the Exam C Sample Questions from the Society of Actuaries. Losses follow an exponential distribution. An insurance company insuring against the losses requires policyholders to pay a deductible of d. This year, the loss elimination ratio was 40%. Next year, the deductible will increase to 2d. What will be the new loss elimination ratio, if the distribution of losses remains the same?

Relevant properties of exponential distributions: $E(X \wedge k) = \theta(1-e^{-k/\theta})$.

Solution S4C88-5. Loss elimination ratio = $E(X \wedge d)/E(X) = \theta(1-e^{-d/\theta})/\theta = 1-e^{-d/\theta}$ in this case. We are given that $1-e^{-d/\theta} = 0.4 \rightarrow e^{-d/\theta} = 0.6$. Next year, since the deductible will be 2d, the loss elimination ratio will be $1-e^{-2d/\theta}$. We note that $e^{-2d/\theta} = (e^{-d/\theta})^2 = 0.6^2 = 0.36$, so the loss elimination ratio will be 1 - 0.36 = **0.64 = 64%.**

Section 89

Assorted Exam-Style Questions for Exam 4/C – Part 6

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C89-1. Similar to Question 90 of the Exam C Sample Questions from the Society of Actuaries. The frequency distribution of insured losses is Poisson with mean Λ, where Λ follows a gamma distribution with mean 5 and variance 5. You select a random insured. Find the probability that the insured will suffer at most two losses next year. Refer to the Exam 4 / C Tables as necessary.

Solution S4C89-1. When a Poisson parameter follows a gamma distribution, the overall frequency distribution follows a negative binomial distribution with $r = \alpha$ and $\beta = \theta$.

For the given gamma distribution, the mean is $\alpha\theta = 5$, and the variance is $\alpha\theta^2 = 5$, which implies that $\beta = \theta$ could only be 1 and, therefore, that $\alpha = r = 5$.

Our desired answer is Pr(0) + Pr(1) + Pr(2) for the negative binomial distribution.

From the exam tables, for a negative binomial distribution, $\Pr(0) = (1+\beta)^{-r} = 2^{-5} = 0.03125$.

$\Pr(1) = r*\beta/(1+\beta)^{1+r} = 5*1/2^6 = 0.078125$.

$\Pr(2) = r(r+1)\beta^2/(2(1+\beta)^{2+r}) = 5*6*1^2/(2(2)^7) = 0.1171875$.

Our answer, therefore, is 0.03125 + 0.078125 + 0.1171875 = **0.2265625 = 29/128.**

Problem S4C89-2. Similar to Question 91 of the Exam C Sample Questions from the Society of Actuaries. The frequency of losses per year has mean 4059 and variance 3059. The severity of each loss has mean 120 and standard deviation 78. Use the central limit theorem to find the probability that aggregate losses this year will be less than 500000.

Solution S4C89-2. Let N = frequency, X = severity, and S = aggregate losses.

First, we find E(S) = E(N)*E(X) = 4059*120 = 487080.

Now we find Var(S) = E(N)*Var(X) + Var(N)*$E(X)^2$ = $4059*78^2 + 3059*120^2$ = 68744556.

Thus, the standard deviation of S is √(68744556) = 8921.233684.

Our desired probability is Pr(S < 500000) = Φ((500000-487080)/8921.233684) = Φ(1.558272326), which we can find via the Excel input "=NORMSDIST(1.558272326)", giving us our answer: **0.940415647**.

Problem S4C89-3. Review of Section 1. Similar to Question 92 of the Exam C Sample Questions from the Society of Actuaries. You know that frequency of losses follows a geometric distribution with mean 7. Each loss has the same severity of 5. Let S denote aggregate losses. Find $E((S-18.5)_+)$.

Relevant properties regarding geometric distributions: E(X) = β; p(0) = 1/(1+β).

$p(k) = \beta^k/(1+\beta)^{k+1}$.

Solution S4C89-3. From Section 1, for a discrete random variable S,

$E((S-d)_+) = {}_{s\_j>d}\Sigma(s_j - d)*p(s_j)$. Here, d = 18.5 and S = 5N, where N is the frequency random variable. Thus, $E((S-18.5)_+) = E((5N-18.5)_+) = 5*E((N-3.7)_+)$.

Thus, we want to find

$5({}_{n\_j>3.7}\Sigma(n_j - 3.7)*p(n_j))$ =

$5({}_{n\_j>3.7}\Sigma(n_j)*p(n_j) - {}_{n\_j>3.7}\Sigma(3.7*p(n_j)))$ =

$5({}_{n\_j>0}\Sigma(n_j)*p(n_j) - 0*p(0) - 1*p(1) - 2*p(2) - 3*p(3) - {}_{n\_j>0}\Sigma(3.7*p(n_j)) + 3.7*p(0) + 3.7*p(1) + 3.7*p(2) + 3.7*p(3))$.

By the definition of expected value, ${}_{n\_j>0}\Sigma(n_j)*p(n_j)$ = E(N) = 7, and ${}_{n\_j>0}\Sigma(3.7*p(n_j))$ =

$3.7*{}_{n\_j>0}\Sigma(p(n_j))$ = 3.7, as, by the definition of probability, ${}_{n\_j>0}\Sigma(p(n_j))$ = 1.

Thus, our answer is equal to

5(7 - 3.7 + 3.7*p(0) + 2.7*p(1) + 1.7*p(2) + 0.7*p(3)) =

5(3.3 + 3.7*p(0) + 2.7*p(1) + 1.7*p(2) + 0.7*p(3)).

As E(N) = 7 and N is geometric, we know that $\beta = 7$.

We also find

p(0) = 1/(1+7) = 1/8 = 0.125;

p(1) = $7/8^2$ = 7/64 = 0.109375;

p(2) = $7^2/8^3$ = 49/512 = 0.095703125.

p(3) = $7^3/8^4$ = 343/4096 = 0.0837402344.

$E((S-18.5)_+)$ = 5(3.3 + 3.7*p(0) + 2.7*p(1) + 1.7*p(2) + 0.7*p(3)) =

5(3.3 + 3.7*(1/8) + 2.7*(7/64) + 1.7*(49/512) + 0.7*(343/4096)) = **$E((S-18.5)_+)$ = 21.39562988**.

Problem S4C89-4. A single round of a game is played as follows:
You roll a fair 8-sided die with faces numbered 1 through 8, whose outcome determines the value of N.

Then you roll N fair 8-sided dice of the same sort and receive an amount of money equal to the sum of their faces.
Find the variance of your payoff in each individual round.

Solution S4C89-4. Let N be frequency (number of dice rolled) and let X be severity (total value of *one* die rolled). Both N and X are based on the roll of a single 8-sided die, so

E(X) = E(N) and Var(X) = Var(N).

First, we find E(N) = (1/8)(1+2+3+4+5+6+7+8) = 4.5.

Now we find Var(X) = (1/8)($(1-4.5)^2 + (2-4.5)^2 + (3-4.5)^2 + (4-4.5)^2 + (5-4.5)^2 + (6-4.5)^2 + (7-4.5)^2 + (8-4.5)^2$) = 5.25.

Let S be the payoff in each individual round. Then

Var(S) = E(N)*Var(X) + Var(N)*$E(X)^2$ = (since Var(X) = Var(N)) 5.25($4.5 + 4.5^2$) = **Var(S) = 129.9375.**

Problem S4C89-5. Similar to Question 93 of the [Exam C Sample Questions] from the Society of Actuaries. A single round of a game is played as follows:
You roll a fair 8-sided die with faces numbered 1 through 8, whose outcome determines the value of N.

Then you roll N fair 8-sided dice of the same sort and receive an amount of money equal to the sum of their faces.

You play the game for 30 rounds, and you must pay a fee of 15 to play in each round. Your starting balance is 500. Use the central limit theorem to find the probability that your starting balance will exceed 500 after 30 rounds.

Solution S4C89-5. If we start with 500 and pay 15*30 = 450 to play for 30 rounds, then we are left with a balance of 50 at the beginning. Thus, we want to find the probability that, after 30 rounds, we will win at least 450, i.e., Pr(30S > 450).

We find E(30S) = 30E(S) = 30*E(X)*E(N) = 30*4.5^2 (from Solution S4C89-4) = 607.5.

We find Var(30S) = 30Var(S) = 30*129.9375 = 3898.125 → SD(30S) = $3898.125^{1/2}$ = 62.43496616.

Thus, by the central limit theorem, we want to find 1-Φ((450-607.5)/62.43496616) =

1-Φ(-2.522624896), which we can find via the Excel input "=1-NORMSDIST(-2.522624896)", giving us **0.994175871** as our answer.

Section 90

Assorted Exam-Style Questions for Exam 4/C – Part 7

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C90-1. Similar to Question 94 of the Exam C Sample Questions from the Society of Actuaries. You know that a random variable X follows a distribution that belongs to the (a, b, 0) class. You also know that Pr(X = 0) = 0.2, Pr(X = 1) = 0.4, Pr(X = 2) = 0.3. Find Pr(X = 3).

Solution S4C90-1. The distinctive feature of distributions of the (a, b, 0) class is that the following formula holds: $p_k/p_{k-1} = a + b/k$ for positive integer values of k.

Thus, $p_1/p_0 = a + b/1 = a + b = 0.4/0.2 = 2$, so a + b = 2.
$p_2/p_1 = a + b/2 = 0.3/0.4 = 0.75$.
If a + b = 2 and a + b/2 = 0.75, then b/2 = 2 - 0.75 = 1.25, implying that b = 2.5.
Therefore, a = 2 - b = 2 - 2.5 = -0.5.
Thus $p_3/p_2 = a + b/3 = -0.5 + 2.5/3 = 1/3$, implying that $p_3 = (1/3)p_2 = 0.3/3 =$ **Pr(X = 3) = 0.1**.

Problem S4C90-2. Similar to Question 95 of the Exam C Sample Questions from the Society of Actuaries. The number of claims (N) follows a geometric distribution with mean 3. The size of each claim (X) has probability 0.2 at each of the values 4, 6, 10, 12, and 15. Let S be the aggregate claim amount. *If there are multiple claims, each claim could be of a difference size, but does not need to be.* Find $F_S(15)$.

Relevant properties regarding geometric distributions: E(X) = β; p(0) = 1/(1+β).

$p(k) = \beta^k/(1+\beta)^{k+1}$.

Solution S4C90-2. In how many ways can we get claim sizes less than or equal to 15?

If there is 1 claim, then any of the possible claim sizes will be less than or equal to 15.

If there are 2 claims, then the following sequences of claim sizes will have sums than or equal to 15:

4 and 4: probability $0.2^2 = 0.04$.

4 and 6, 6 and 4: probability $2*0.2^2 = 0.08$

6 and 6: probability $0.2^2 = 0.04$.

4 and 10, 10 and 4: probability $2*0.2^2 = 0.08$

Overall, $\Pr(S < 15 \mid N = 2) = 2*0.04 + 2*0.08 = 0.24$.

If there are 3 claims, then the following sequences of claim sizes will have sums than or equal to 15:

4, 4, and 4: probability $0.2^3 = 0.008$.

4, 4, and 6, in any order: probability $3*0.2^3 = 0.024$.

Overall, $\Pr(S < 15 \mid N = 3) = 0.008 + 0.024 = 0.032$.

If there are 4 claims, even if all claims were of the smallest size (4), the sum would always exceed 15. Thus, we only need to concern ourselves about values of N from 0 to 3, inclusive.

$F_S(15) = \Pr(N = 0) + \Pr(N = 1)*\Pr(S < 15 \mid N = 1) + \Pr(N = 2)*\Pr(S < 15 \mid N = 2) + \Pr(N = 3)*\Pr(S < 15 \mid N = 3)$. We are given $E(N) = 3 = \beta$. Therefore, we can find the following probabilities:

$\Pr(N = 0) = 1/(1+3) = ¼ = 0.25$.
$\Pr(N = 1) = 3/4^2 = 3/16$.
$\Pr(N = 2) = 3^2/4^3 = 9/64$.
$\Pr(N = 3) = 3^3/4^4 = 27/256$.

Thus, $F_S(15) = ¼ + (3/16)*1 + (9/64)*0.24 + (27/256)*0.032 =$ **$F_S(15) = 0.474625 = 3797/8000$**.

Problem S4C90-3. Similar to Question 96 of the Exam C Sample Questions from the Society of Actuaries. An insurance company pays an agent a bonus if losses are less than 30% of the premium collected by the agent - which is 1000. The magnitude of the bonus is 40% of the difference between the 30% threshold and the amount of losses. Losses (X) follow a Pareto distribution with $\alpha = 3$ and $\theta = 3000$. Find the expected value of the agent's bonus.

Relevant properties regarding Pareto distributions:
$E(X \wedge k) = (\theta/(\alpha-1))*(1 - (\theta/(x+\theta))^{\alpha-1})$.

Solution S4C90-3. We define B as a random variable in terms of X.
B = 0.4(0.3*1000 - X), but only if X < 0.3*1000 = 300; otherwise, B = 0. Thus,
B = 0.4(300 - X Λ 300) and E(B) = E(0.4(300 - X Λ 300)) =
0.4*300 - 0.4*E(X Λ 300).
We find E(X Λ 300) = (3000/(3-1))*(1 - (3000/(3300))$^{3-1}$) = 260.3305785.
Thus, E(B) = 0.4*300 - 0.4*260.3305785 = **E(B) = 15.8677686**.

Problem S4C90-4. Similar to Question 97 of the [Exam C Sample Questions] from the Society of Actuaries. Claim frequency (N) follows a negative binomial distribution with mean 5667 and variance 46090. Claim severity (X) has the following probabilities:

Pr(X = 350) = 0.4;
Pr(X = 460) = 0.3;
Pr(X = 900) = 0.2;
Pr(X = 3400) = 0.1.

Next year, severities are expected to decrease by 40%, and a deductible of 200 will be imposed for each claim. Find the insurance company's expected payout under next year's conditions.

Solution S4C90-4. Next year, the severities will change to 0.6 of their previous amounts; thus, the severities will become 0.6*350 = 210, 0.6*460 = 276, 0.6*900 = 540, 0.6*3400 = 2040.

With the 200 deductible, the payout for each magnitude of loss will be

10, 76, 340, and 1840, respectively.

Thus, the expected payout *per claim* is 10*0.4 + 76*0.3 + 340*0.2 + 1840*0.1 = 278.8.

Total expected payout is E(N)*(Expected payout per claim) = 5667*278.8 = **1,579,959.6**.

Problem S4C90-5. Similar to Question 98 of the [Exam C Sample Questions] from the Society of Actuaries. A cat scratches an average of 12 pieces of furniture a year, with a variance of 50. The monetary damage to each piece of furniture scratched has a mean of 600 and a variance of 1000. Frequency and severity of scratches are independent. If the cat's total monetary damage due to furniture scratches is under 5000 in a year, the cat gets a treat. Use the central limit theorem to find the probability that the cat will get a treat this year.

Solution S4C90-5. Let N be the frequency, X be the severity, and S be the aggregate loss due to scratches. E(S) = E(N)*E(X) = 12*600 = 7200. Var(S) = E(N)*Var(X) + Var(N)*E(X^2) = 12*1000 + 50*600^2 = 18012000, meaning that SD(S) = √(18012000) = 4244.054665.

Pr(Treat) = Pr(S < 5000) ≈ Φ((5000-7200)/4244.054665) = Φ(-0.5183722109), which we can find via the Excel input "=NORMSDIST(-0.5183722109)", giving us the answer of **0.3020993**.

Section 91

Assorted Exam-Style Questions for Exam 4/C – Part 8

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C91-1. Similar to Question 99 of the Exam C Sample Questions from the Society of Actuaries. The distribution of frequency of claims (N) is Poisson with mean 4. The distribution of severity of claims (X) is such that ¼ probability is assigned to each of the following values: 4, 7, 8, 10. An insurance policy has a deductible of 8 that is applied to aggregate losses. Find the insurer's expected payout under this policy.

Relevant properties of Poisson distributions: Pr(N = k) = $e^{-\lambda}*\lambda^k/k!$

Solution S4C91-1. Let S denote aggregate losses. The expected payout with a deductible d is E(S) - E(S Λ d). Here, E(S) = E(N)*E(X). E(X) = (4 + 7 + 8 + 10)/4 = 7.25, so E(S) = 4*7.25 = 29. Now we find E(S Λ 8).

What is the probability of aggregate claims less than 8?

The only ways this can happen are if there are no claims, a single claim of 4, or a single claim of 7.

Pr(N = 1 and X = 4) = e^{-4}(1/4) = 0.0045789097.

Pr(N = 1 and X = 7) = e^{-4}(1/4) = 0.0045789097.

Thus, E(S Λ 8) = 4*0.0045789097 + 7*0.0045789097 + 8(1 - 2*0.0045789097) = 7.977105451.

Therefore, the expected payout is 29 - 7.977105451 = **21.02289455.**

Problem S4C91-2. Similar to Question 100 of the Exam C Sample Questions from the Society of Actuaries. The distribution of claims is a mixture of two exponential distributions, such that the cumulative distribution function is as follows: $F(x) = 1 - 0.46e^{-0.06x} - 0.54e^{-0.05x}$ for all $x > 0$. An insurance company only pays claims up to a limit of 25. Find the expected payout for a single claim.

Relevant properties of exponential distributions: $E(X \wedge k) = \theta(1-e^{-k/\theta})$.

Solution S4C91-2. This is a mixture of two exponential values, for which we are trying to find $E(X \wedge 25)$. Naturally, the expected value of the mixture will be the appropriately weighted average of the expected values of the right-censored component distributions, whose means, respectively, are 1/0.06 = 16.66666667 and 1/0.05 = 20.

Thus, $E(X \wedge 25) = 0.46*16.66666667(1 - e^{-25/16.66666667}) + 0.54*20(1 - e^{-25/20})$ = **E(X Λ 25) =13.6617503.**

Problem S4C91-3. Similar to Question 101 of the Exam C Sample Questions from the Society of Actuaries. For a random variable X, you are given the following information:
F(340) = 0.3, E(X Λ 340) = 295;
F(430) = 0.5, E(X Λ 430) = 356;
F(550) = 1, E(X Λ 550) = 400.

An insurance policy where losses are represented by X has a deductible of 340. Find the mean excess loss for this deductible.

Solution S4C91-3. For a deductible d, the mean excess loss is

(Expected payout with deductible considered)/(1-F(d)) = (E(X) - E(X Λ d))/(1-F(d)).

We note here that, since F(550) = 1, all possible values of X occur at or below 550, and so E(X) = E(X Λ 550) = 400. We are given E(X Λ 340) = 295 and F(340) = 0.3, and so

(400 - 295)/(1 - 0.3) = **mean excess loss = 150**.

Problem S4C91-4. Similar to Question 102 of the Exam C Sample Questions from the Society of Actuaries. A company earns revenues of 1000 in a year. The company owns two offices, each of which has (1/3) annual probability of suffering each of the following magnitudes of losses: 0, 500, 1000. The company insures itself against these losses by purchasing a policy with a deductible of 250. The premium for the policy is 105% of the expected losses *that the insurer will have to pay*. The company also has annual expenses equaling 10% of revenues. What profit, if any, can the company be expected to have in a year?

Solution S4C91-4. First, we find the company's predictable expenses. The non-premium expenses are 10% of 1000, or 100.

The premium cost will be 1.05(Expected value of insurance payout).
Because of the deductible of 250, for each building, the possible insurance payouts are as follows, with each payout having probability 1/3: 0, 250, 750. Thus, for each building, the expected payout is (250 + 750)/3 = 1000/3, implying that for both buildings, the expected payout is 2000/3, and the premium cost is 1.05*2000/3 = 700.

Therefore, the company has 1000 - 100 - 700 = 200 left over.

If even one office suffers a loss of 500 or 1000, the company will have to pay a deductible of 250 and so will have a -50 profit. If no offices suffer a loss, then the company will have a profit of 200.

Pr(no offices suffer a loss) = $(1/3)^2$ = 1/9, and so the expected profit is 200*(1/9) + (-50)(8/9) =

-22.2222222222.

Problem S4C91-5. Similar to Question 103 of the Exam C Sample Questions from the Society of Actuaries. The lifetimes of green slugs follow a single-parameter Pareto distribution with θ = 12 and α > 1. You know that the mean of the distribution is 18 years. Find the probability that a green slug will live to be older than 15 years.

Relevant properties of single-parameter Pareto distributions:
E(X) = αθ/(α-1); F(x) = 1 - $(\theta/x)^\alpha$ for x > θ.

Solution S4C91-5. We want to find S(15), but first we need to find α. We are given that E(X) = αθ/(α-1) = 18 = 12α/(α-1), so α/(α-1) = 3/2, which is consistent with α = 3.

We know that S(x) = 1 - F(x) = $(\theta/x)^\alpha$, and so S(15) = $(12/15)^3$ = **0.512.**

Section 92

Assorted Exam-Style Questions for Exam 4/C – Part 9

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C92-1. Similar to Question 105 of the Exam C Sample Questions from the Society of Actuaries. The total number of peanuts eaten by an elephant follows a negative binomial distribution with mean 45 and variance 65. This random variable could also be said to follow a Poisson distribution with mean Λ, where Λ has a gamma distribution. Find the variance of that gamma distribution.

Relevant properties of negative binomial distributions: E(N) = rβ; Var(N) = rβ(1+β)

Relevant properties of gamma distributions: E(X) = αθ; Var(X) = $\alpha\theta^2$.

Solution S4C92-1. When a Poisson parameter follows a gamma distribution, the overall frequency distribution follows a negative binomial distribution with r = α and β = θ.

Here, we first need to find β. We are given that rβ = 45 and rβ(1+β) = 65, so (1+β) = 65/45 = 13/9, and thus β = 4/9. Thus, r = 45/(4/9) = 101.25. Thus, in the gamma distribution, α = 101.25 and θ = (4/9), meaning that the variance of the gamma distribution is $101.25*(4/9)^2$ = **20.**

Problem S4C92-2. Similar to Question 106 of the Exam C Sample Questions from the Society of Actuaries. Every day, the mean number of valuable mineral specimens that an explorer of another planet collects is 700, and the variance of this number is 350. For each specimen the explorer collects, he can make a mean income of 4000, with a variance of 3540. For at least how many whole days must the explorer gather mineral specimens to have a 0.9 probability that his total income will exceed 10,000,000? Use the central limit theorem to find the answer.

Solution S4C92-2. The explorer's mean income for *one* day of exploration is 700*4000 = 2,800,000.

The variance of the explorer's income per day is

Var(S) = E(N)*Var(X) + Var(N)*E(X)^2, where N is number of mineral specimens collected and X is income per specimen. Thus, Var(S) = 700*3540 + 350*4000^2 = 5602478000, implying that SD(S) = 5602478000$^{0.5}$ = 74849.70274.

Let n be the number of days the explorer spends. Then, over n days, the mean is 2800000n and the standard deviation is 74849.70274√(n).

The minimum n is such that

Pr(Total income < 10,000,000) = 0.1. Using the central limit theorem, we set

Φ((10000000-2800000n)/74849.70274√(n)) = 0.1.

To find the associated z value, we use the Excel input "=NORMSINV(0.1)", getting as our result

-1.281551566. Therefore, (10000000-2800000n)/74849.70274√(n) = -1.281551566 →

10000000-2800000n = -95923.75376√(n) →

2800000n - 95923.75376√(n) - 10000000 = 0.

We can think of the above equation as a quadratic equation in terms of √(n). The positive solution to this quadratic equation is √(n) = 1.907029234, implying that the minimum n is 3.636760499, which we round up to **4 days.**

Problem S4C92-3. Similar to Question 107 of the Exam C Sample Questions from the Society of Actuaries. There are 9 insureds, each of whom has the following probability distribution of losses (X):
Pr(X = 0) = 1/5;
Pr(X = 1) = 1/5;
Pr(X = 2) = 1/5;
Pr(X = 3) = 1/5;
Pr(X = 4) = 1/5.

A policy of stop-loss insurance covers all 9 insureds, with a deductible of 1 for *the entire group*. Find the net stop-loss premium for this policy.

Solution S4C92-3. The net stop-loss premium is the expected payout for the insurer, with the deductible taken into consideration. Let S be aggregate losses. In this case, the net stop-loss premium is E(S) - E(S Λ d), where d = 1. For each individual insured, the expected loss is (0+1+2+3+4)/5 = 2. Thus, for 9 insureds, the total expected loss, E(S), is 18.

Now we find E(S Λ 1). In this calculation, all losses are capped at 1, so there are only two possible loss amounts: 0 and 1. We find Pr(S Λ 1 = 0) = $(1/5)^9$, implying that

Pr(S Λ 1 = 1) = 1 - $(1/5)^9$.

Thus, E(S Λ 1) = 0*$(1/5)^9$ + 1*(1 - $(1/5)^9$) = 0.999999488.

Therefore, our net stop-loss premium is 18 - 0.999999488 = **17.00000051.**

Problem S4C92-4. Similar to Question 108 of the Exam C Sample Questions from the Society of Actuaries. The following recursion relation pertains to a probability distribution: Pr(X = k) = (8/k)*Pr(X = k-1). Find Pr(X = 13). Refer to the Exam 4 / C Tables as necessary.

Solution S4C92-4. We note that Pr(X = k)/Pr(X = k-1) = (8/k) = 0 + 8/k, which is of the format a + b/k, where a = 0 and b = 8. This means that the distribution belongs to the (a, b, 0) class of distributions. Since a = 0, the distribution is Poisson with b = λ = 8. Thus,

Pr(X = 13) = e^{-8}*8^{13}/13! = **Pr(X = 13) = 0.0296164949**.

Problem S4C92-5. Similar to Question 109 of the Exam C Sample Questions from the Society of Actuaries. Loss frequency per year has a mean of 89. Loss severity follows an exponential distribution with θ = 3540. Next year, loss frequency is expected to fall by 50%, and a deductible of 3521 will be applied to each loss. Find the total expected payout for the insurance company under these circumstances.

Relevant properties of exponential distributions: E(X Λ k) = θ(1-$e^{-k/\theta}$).

Solution S4C92-5. If frequency decreases by 50%, the new expected frequency will be 0.5*89 = 44.5. Let X denote loss severity. With the imposition of the deductible, expected payout per loss will be E(X) - E(X Λ d) = θ - θ(1-$e^{-d/\theta}$) = θ$e^{-d/\theta}$ = 3540$e^{-3521/3540}$ = 1309.301722.

Thus, total expected payout will be 44.5*1309.301722 = **58263.92665.**

Section 93

The Continuity Correction for Normal Approximations and Exam-Style Questions on Mixed and Aggregate Distributions

This section provides additional exam-style practice with a variety of syllabus topics.

When making approximations using the normal distribution, it is often necessary to use a **continuity correction.** This needs to happen when you use a continuous distribution such as the normal distribution to approximate a distribution where only discrete outcomes (such as integer outcomes) are possible. For instance, using the normal distribution in approximating a binomial distribution with a large number of possible discrete outcomes requires the continuity correction.

When you have an integer g, then, by the continuity correction,

Pr(More than g) = Pr(At least g + 1) = Pr(g + 1 or more) = 1 - Φ((g + 0.5 - μ)/σ).

When you have an integer g, then, by the continuity correction,

Pr(Less than g) = Pr(At most g - 1) = Pr(g - 1 or less) = Φ((g - 0.5 - μ)/σ).

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Larsen, Richard J. and Morris L. Marx. *An Introduction to Mathematical Statistics and Its Applications.* Fourth Edition. Pearson Prentice Hall: 2006. p. 302.

Mahler, Howard. "Sample Pages for Stochastic Models (CAS 3L / SOA MLC)."

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C93-1. Similar to Question 110 of the Exam C Sample Questions from the Society of Actuaries. There can be either 3 or 5 losses. The probability of 3 losses is 0.3, and the probability of 5 losses is 0.7. For each loss, the loss amount is 436 with probability 0.4, 3450 with probability 0.4, and 10000 with probability 0.2. The premium for an insurance policy is equal to the sum of the mean and the standard deviation of aggregate losses. Determine this premium.

Solution S4C93-1. Let N be frequency of losses, and let X be severity of losses.
We find E(N) = 0.3*3 + 0.7*5 = E(N) = 4.4.
Var(N) = $0.3*(3-4.4)^2 + 0.7*(5-4.4)^2$ = Var(N) = 0.84.
We find E(X) = 0.4*436 + 0.4*3450 + 0.2*10000 = 3554.4.
Var(X) = $0.4(436-3554.4)^2 + 0.4(3450-3554.4)^2 + 0.2(10000-3554.4)^2$ = Var(X) = 12203279.04.
Let S be aggregate losses. Then E(S) = E(N)*E(X) = 4.4*3554.4 = 15639.36.
We can now find Var(S) = E(N)*Var(X) + Var(N)*$E(X)^2$ =
4.4*12203279.04 + 0.84*3554.4^2 = Var(S) = 64306785.64.
Thus, SD(S) = $64306785.64^{1/2}$ = 8019.151179.
The premium is E(S) + SD(S) = 15639.36 + 8019.151179 = **23658.51118.**

Problem S4C93-2. Similar to Question 111 of the Exam C Sample Questions from the Society of Actuaries. Frequency of losses follows a Poisson distribution with λ = 50. Severity of each loss is 460 with probability 0.5, 600 with probability 0.3, and 1000 with probability 0.2. Find the variance of aggregate losses.

Solution S4C93-2. Let N be frequency of losses, and let X be severity of losses. Let S be aggregate losses. If frequency is Poisson, then
Var(S) = E(N)*E(X^2). We are given E(N) = λ = 50.
We find E(X^2) = 0.5*460^2 + 0.3*600^c + 0.2*1000^2 = 413800.
Thus, Var(S) = 50*413800 = **Var(S) = 20690000**.

Problem S4C93-3. Similar to Question 112 of the Exam C Sample Questions from the Society of Actuaries. Frequency of losses is uniformly distributed *on the integers* from 6 to 18. Severity of each loss is Poisson with λ = 90. Use the normal approximation *with the continuity correction* to find the probability that aggregate losses are 1350 or more. Express your answer as 1 - Φ(k), where k is a number, and Φ(k) is the standard normal cumulative distribution function at k.

Solution S4C93-3. Let N be frequency of losses, and let X be severity of losses. Let S be aggregate losses. E(N) = (6+7+8+9+10+11+12+13+14+15+16+17+18)/13= E(N) = 12.

E(N^2) = $(6^2+7^2+8^2+9^2+10^2+11^2+12^2+13^2+14^2+15^2+16^2+17^2+18^2)/13$ = E(N^2) = 158.

Var(N) = E(N^2) - $E(N)^2$ = 158 - 12^2 = Var(N) = 14.

Since X is Poisson-distributed, E(X) = Var(X) = λ = 90.

Thus, E(S) = E(N)*E(X) = 12*90 = 1080.

Var(S) = E(N)*Var(X) + Var(N)*$E(X)^2$ = 12*90 + 14*90^2 = 114480, implying that SD(S) = 338.3489323.

Pr(S ≥ 1350) = Pr(S is 1349 + 1 or more) ≈ 1 - Φ((1349.5 - 1080)/338.3489323) =

1 - Φ(0.7965150005).

Problem S4C93-4. Similar to Question 113 of the Exam C Sample Questions from the Society of Actuaries. The number of claims is 0 with probability 0.4, 1 with probability 0.45, and 2 with probability 0.05. The size of each claim is either 2 with probability 0.7 or 55 with probability 0.3. Find the probability that aggregate claim amount S, exceeds the mean aggregate claim amount E(S) plus two standard deviations of S.

Solution S4C93-4. Let N be frequency of claims, and let X be severity of claims.
We find E(N) = 0*0.4 + 1*0.45 + 2*0.05 = 0.55.
Var(N) = $(0-0.55)^2$*0.4 + $(1-0.55)^2$*0.45 + $(2-0.55)^2$*0.05 = 0.31725.
We find E(X) = 2*0.7 + 55*0.3 = 17.9.
Var(X) = $(2-17.9)^2$*0.7 + $(55-17.9)^2$*0.3 = 589.69.
Thus, E(S) = E(N)*E(X) = 0.55*17.9 = 9.845.

Var(S) = E(N)*Var(X) + Var(N)*$E(X)^2$ = 0.55*589.69 + 0.31725*17.9^2 = 425.9795725, implying that SD(S) = $425.9795725^{1/2}$ = 20.63927258.

We want to find Pr(S > 9.845 + 2*20.63927258) = Pr(S > 51.12354515).

We note that if there is even one claim of 55, S will exceed 51.12354515. The most efficient way to find our answer is to find the complement of the probability that no claim is 55.

Thus, Pr(S > 51.12354515) = 1 - Pr(N = 0) - Pr(N = 1 and X = 2) - Pr(N = 2 and both claims are 2) = 1 - 0.4 - 0.45*0.7 - 0.05*0.7^2 = **Pr(S > 51.12354515) = 0.2605**.

Problem S4C93-5. Similar to Question 114 of the Exam C Sample Questions from the Society of Actuaries. The number of asteroids hitting the surface of a planet follows a mixed Poisson distribution with mean Λ, where Λ is uniformly distributed on the interval from 5 to 10. Find the probability that fewer than 2 asteroids will hit the surface of the planet. Refer to the Exam 4 / C Tables as necessary.

Solution S4C93-5. For a Poisson distribution, Pr(N = 0) = $e^{-\lambda}$, and Pr(N = 1) = $\lambda e^{-\lambda}$, and so Pr(N < 2 | Λ) = $e^{-\lambda} + \lambda e^{-\lambda}$. We know that f(λ) = 1/5 for 5 < λ < 10, and so Pr(N < 2) =
$_5{}^{10}\int f_\Lambda(\lambda)$*Pr(N > 2 | Λ)*dλ = $_5{}^{10}\int(1/5)(e^{-\lambda} + \lambda e^{-\lambda})$*dλ =
$-(1/5)e^{-\lambda} - (1/5)e^{-\lambda} - (1/5)\lambda e^{-\lambda} \Big|_5^{10}$ =**0.009324166**.

Section 94

Assorted Exam-Style Questions for Exam 4/C – Part 10

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C94-1. Similar to Question 115 of the Exam C Sample Questions from the Society of Actuaries. Losses follow an exponential distribution with mean 4. An insurance company pays all the losses once a deductible of 1 has been applied. Find the variance of the insurance company's payout per loss. Consider in your answer all the possible losses for which the insurance company's payout may be zero.

Relevant properties of exponential distributions: $F(x) = 1 - e^{-x/\theta}$; $E(X^2) = 2\theta^2$.

Solution S4C94-1. Let L be the loss amount and P be the insurance company payout. Exponential distributions have the memoryless property, so the distribution of losses *in excess* of 1 will also be exponential with mean 4. The probability of a loss being less than 1 is $F(1) = 1 - e^{-1/4} = 0.2211992169$, which is the probability of the company paying out nothing.

Thus, $f(p) = 0.2211992169*0 + 0.7788007831*(1/4)e^{-p/4}$.

Therefore, E(P) = 0.2211992169*E(0) + 0.7788007831*E(Exponential distribution with mean 4) = 0.7788007831*4 = 3.115203132.

$E(P^2) = 0.2211992169*0^2 + 0.7788007831*E(X^2$, where X follows an exponential distribution with mean 4) = $0 + 0.7788007831*2*4^2 = 24.92162506$.

Thus, $Var(P) = E(P^2) - E(P)^2 = 24.92162506 - 3.115203132^2 =$ **Var(P) = 15.2171345**.

Problem S4C94-2. Similar to Question 116 of the Exam C Sample Questions from the Society of Actuaries. Initial aggregate losses for all claims follow a Pareto distribution with α =

2 and θ = 34. Claims adjusters are paid a bonus if the claims they adjust have total losses of less than 20, with the bonus being the amount by which the loss is less than 20. If the loss exceeds 20, no bonus is paid. As a result of the bonus, the distribution of aggregate losses changes to another Pareto distribution, where α = 2 and θ = R, where R + E(B) = 34, and B is the amount of the bonus. Find R.

Relevant properties regarding Pareto distributions: $E(X) = \theta/(\alpha-1)$;

$E(X \wedge k) = (\theta/(\alpha-1))(1 - (\theta/(k+\theta))^{\alpha-1})$

Solution S4C94-2. First, we find the expression for the amount of bonus: $B = 20 - X \wedge 20$, and so $E(B) = 20 - E(X \wedge 20) = 20 - (R/(2-1))(1-(R/(20+R))^{2-1}) =$

$20 - R(20/(20+R))$. Note that we calculate the expected value of the bonus using the *new* loss distribution, because the application of the bonus means that the new loss distribution is the distribution that describes losses and thus the distribution using which E(B) should be calculated.

Hence, we have $R + 20\ R(20/(20+R)) = 34 \rightarrow 14 = R - 20R/(20+R) \rightarrow$

$14(20+R) = R(20 + R) - 20R \rightarrow 280 + 14R = R^2 \rightarrow R^2 - 14R - 280 = 0$. Solving this quadratic equation for R, we get as our positive solution **R = 25.13835715**.

Problem S4C94-3. Similar to Question 117 of the Exam C Sample Questions from the Society of Actuaries. The following observations are made:

10000 peanuts have not been eaten at time 0.

3400 peanuts have not been eaten at time 5.

3150 peanuts have not been eaten at time 10.

2350 peanuts have not been eaten at time 15.

400 peanuts have not been eaten at time 20.

0 peanuts have not been eaten at time 25.

Assume that peanuts eaten follow a uniform distribution within each time interval of length 5 specified above. If a peanut is observed not to have been eaten at time 17, in how many units of time can it be expected to be eaten?

Solution S4C94-3. Because peanuts eaten are uniformly distributed within each interval, we expect to observe 2350 - (2/5)(2350 - 400) = 1570 peanuts to remain uneaten at time 17.

We also expect 1570-400 = 1170 of these peanuts to be eaten by time 20, so

S(20 | survival to 17) = 400/1570 = 0.2547770701, and S(25 | survival to 17) = 0.

To solve this problem, we can draw a graph with probability on the vertical axis and time (t) on the horizontal axis. We plot three points: (17, 1), (20, 0.2547770701), and (25, 0) and connect these points with straight lines. The area underneath the resulting graph is our expected value of time past t = 17. We find the area thus:
3*0.2547770701 + 3*(1-0.2547770701)/2 + 5*0.2547770701/2 = **2.51910828 units of time.**

Problem S4C94-4. Similar to Question 118 of the Exam C Sample Questions from the Society of Actuaries. The frequency of losses (N) follows a Poisson distribution with mean 46. Severity of losses (X) follows a uniform distribution on the interval from 15 to 145. Use a normal approximation to find the probability that aggregate losses (S) exceed 4000. You do not need to apply a continuity correction. Express your answer as 1-Φ(k), where k is a number and Φ(k) is the standard normal cumulative distribution function at k.

Solution S4C94-4. First, we find E(X) = (145+15)/2 = 80 and Var(X) = $(145-15)^2/12$ = 1408.333333, implying that $E(X^2)$ = 1408.333333 + 80^2 = 7808.3333333.

Thus, E(S) = E(N)*E(X) = 46*80 = 3680.
Var(S) = 46*7808.3333333 = 359183.333333, and so SD(S) = 599.319058.

Thus, our normal approximation for Pr(S > 4000) = 1-Φ((4000-3680)/599.319058) =

Pr(S > 4000) = 1-Φ(0.5339393028).

Problem S4C94-5. Similar to Question 124 of the Exam C Sample Questions from the Society of Actuaries. The number of losses follows a binomial distribution with m = 46 and q = 0.35. The size of each loss follows a uniform distribution from 1000 to 2000. Find the probability that at least 28 losses occur and exactly 28 of them are less than 1500. Hint: The answer is very small!

Solution S4C94-5. The probability that a loss is less than 1500 is 0.5, so the probability that a given possible loss occurs and is less than 1500 is 0.5*0.35 = 0.175. This means that the number of losses which occur and are less than 1500 follows a binomial distribution with m = 46 and q = 0.175. Thus, Pr(28 losses less than 1500) = C(46, 28)*0.175^{28}*0.825^{18} = **$5.64058681*10^{-11}$**.

Section 95

Assorted Exam-Style Questions for Exam 4/C – Part 11

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C95-1. Similar to Question 119 of the Exam C Sample Questions from the Society of Actuaries. A reinsurance policy pays the excess loss over 675 on each of an insurance company's individual insured losses. The reinsurance premium is equal to 120% of the expected total payout by the reinsurance company in a year.

In the year 3029, each individual loss follows a Pareto distribution with cumulative distribution function $F(x) = 1 - (500/(500+x))^2$. Severity of losses increases each year at an annual rate of 8%. The frequency of losses does not change from year to year.

Total loss amount in 3029 was observed to be 500000.

Find the reinsurance premium for the year 3029.

Relevant properties regarding Pareto distributions: $E(X) = \theta/(\alpha-1)$;

$E(X \wedge k) = (\theta/(\alpha-1))(1 - (\theta/(k+\theta))^{\alpha-1})$

Solution S4C95-1. For 3029, each individual loss (X) follows a Pareto distribution with $\theta = 500$ and $\alpha = 2$, so $E(X) = 500/(2-1) = 500$. Reinsurance pays $E(X) - E(X \wedge 675)$ on each loss, which is equal to $500 - 500(1 - (500/(675+500))^{2-1}) = 500 - 500(27/47) = 212.7659574$.

How many losses are there expected to be in 3029? With 500000 as the amount of total losses and 500 as the expected value per loss, we can expect there to be 500000/500 = 1000, so reinsurance will be expected to pay out 212765.9574. The reinsurance premium is 120% of this amount or 1.2*212765.9574 = **255319.1489.**

Problem S4C95-2. Similar to Question 120 of the Exam C Sample Questions from the Society of Actuaries. A reinsurance policy pays the excess loss over 675 on each of an insurance company's individual insured losses. The reinsurance premium is equal to 120% of the expected total payout by the reinsurance company in a year.

In the year 3029, each individual loss follows a Pareto distribution with cumulative distribution function $F(x) = 1 - (500/(500+x))^2$. Severity of losses increases each year at an annual rate of 8%. The frequency of losses does not change from year to year.

Total loss amount in 3029 was observed to be 500000.

Find the ratio of the reinsurance premium for the year 3030 to the reinsurance premium for the year 3029.

Relevant properties regarding Pareto distributions: $E(X) = \theta/(\alpha-1)$;

$E(X \Lambda k) = (\theta/(\alpha-1))(1 - (\theta/(k+\theta))^{\alpha-1})$

Solution S4C95-2. In the Pareto distribution, the scale parameter is θ, which is multiplied by 1 plus the rate of inflation to get a new θ for 3030: 500*1.08 = 540. Thus, the new expected reinsurance payout per loss is $E(X) - E(X \Lambda 675) = 540 - 540(1 - (540/(675+540))^{2-1}) =$

540(4/9) = 240. As there are still 1000 losses, total reinsurance payout is 240000, and the reinsurance premium is 1.2*240000 = 288000 for 3030. From Solution S4C95-1, we know that the premium for 3029 was 255319.1489 our answer is thus 288000/255319.1489 = **1.128.**

Problem S4C95-3. Similar to Question 123 of the Exam C Sample Questions from the Society of Actuaries. Insured losses for an individual insured follow a Pareto distribution with θ = 400 and α = 2.

Under a policy of coinsurance, the insured pays all the losses up to a deductible of 300.

The insured pays 46% of the losses between 300 and 500.

The insured pays all the losses above 500 until the insurer has paid 600 in total.

The insured pays 4% of the remaining costs.

Find the expected payout *by the insurance company* under this policy for an individual insured.

Relevant properties regarding Pareto distributions: $E(X) = \theta/(\alpha-1)$;

$E(X \Lambda k) = (\theta/(\alpha-1))(1 - (\theta/(k+\theta))^{\alpha-1})$

Solution S4C95-3. The insurance company will pay 54% of the losses between 300 and 500.

The insured will have paid 300 + 0.46*(500 - 300) = 392 prior to reaching 500 in losses, so the insured will need to pay all the losses between 500 and 500 + (600 - 392) = 708. Thus, the insurance company will also pay 96% of all losses above 708.

The insurance company's expected payout is thus

$0.54(E(X \wedge 500) - E(X \wedge 300)) + 0.96(E(X) - E(X \wedge 708))$.

We find E(X) = 400/(2-1) = 400.

We find $E(X \wedge 300)$ = 400(1 - 400/700) = 171.4285714.

We find $E(X \wedge 500)$ = 400(1 - 400/900) = 222.22222222.

We find $E(X \wedge 708)$ = 400(1 - 400/1108) = 255.5956679.

Thus, the insurer's expected payout is

0.54(222.22222222 - 171.4285714) + 0.96(400 - 255.5956679) = **166.0567303.**

Problem S4C95-4. Similar to Question 125 of the [Exam C Sample Questions] from the Society of Actuaries. There are two groups of elephants, A and B. The number of peanuts eaten by elephants in group A follows a Poisson distribution with $\lambda = 35$. The caloric content of each peanut eaten by elephants in group A follows a uniform distribution from 2 to 10. The number of peanuts eaten by elephants in group B follows a Poisson distribution with $\lambda = 20$. The caloric content of each peanut eaten by elephants in group B follows a uniform distribution from 4 to 20. Use a normal approximation (no continuity correction necessary) to find the probability that more than 470 calories in peanuts will be consumed by the elephants in both groups.

Solution S4C95-4. We can find the mean and variance of calories consumed for each elephant group and then add together the means and variances of the respective groups.

Let N be peanut frequency and let X be peanut calorie content (the analog of severity). Let S be aggregate calories per elephant group.
For group A, E(N) = Var(N) = 35, and E(X) = (2+10)/2 = 6, and Var(X) = $(10-2)^2/12$ = 5.33333333. Thus, $E(X^2) = Var(X) + E(X)^2$ = 41.33333333, and so, since N is Poisson-distributed, $Var(S) = E(N)*E(X^2)$ = 35*41.33333333 = 1446.666666667.

E(S) = E(N)*E(X) = 35*6 = 210.

For group B, E(N) = Var(N) = 20, and E(X) = (4+20)/2 = 12, and Var(X) = $(20-4)^2/12$ = 21.33333333. Thus, $E(X^2) = Var(X) + E(X)^2$ = 165.33333333, and so, since N is Poisson-distributed, $Var(S) = E(N)*E(X^2)$ = 20*165.33333333 = 3306.6666667.

E(S) = E(N)*E(X) = 20*12 = 240.

Thus, total expected value is 210 + 240 = 450, and total variance is 1446.666666667 + 3306.6666667 = 4753.33333333, implying that total standard deviation is the square root of that amount, i.e., 68.94442206.
Using the normal approximation, Pr(Total calories > 470) = 1 - Φ((470-450)/68.94442206) =

1 - Φ(0.2900887324), which we can find in Excel using the input

"=1-NORMSDIST(0.2900887324)". Our answer is thus **0.385874178**.

Problem S4C95-5. Similar to Question 126 of the Exam C Sample Questions from the Society of Actuaries. Claim frequency follows a Poisson distribution with $\lambda = 90$. Claim severity follows a Pareto distribution with $\theta = 3400$ and $\alpha = 4.3$. A deductible of 1000 is imposed per claim. Find the expected value of aggregate claim payments for this insurance. Assumethat claim frequency is not affected by the existence of the deductible.

Relevant properties regarding Pareto distributions: $E(X) = \theta/(\alpha-1)$;

$E(X \wedge k) = (\theta/(\alpha-1))(1 - (\theta/(k+\theta))^{\alpha-1})$

Solution S4C95-5. The expected value for an individual claim payment is $E(X) - E(X \wedge 1000)$ = 3400/(4.3-1) - (3400/(4.3-1))(1 - $(3400/4400)^{3.3}$) = 439.99888. Assuming that claim frequency is not affected by the existence of the deductible, expected aggregate payments are 90*439.99888 = **39599.8992.**

Section 96

Assorted Exam-Style Questions for Exam 4/C – Part 12

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C96-1. Similar to Question 127 of the Exam C Sample Questions from the Society of Actuaries. An insurance company applies a deductible of 50 to each policy. In the year 5460, insured losses follow a Pareto distribution with $\alpha = 4$ and $\theta = 300$. In the year 5461, losses are uniformly observed to be 0.6 of the previous amount. Calculate the loss elimination ratio during the year 5461.

Relevant properties regarding Pareto distributions: $E(X) = \theta/(\alpha-1)$;

$E(X \wedge k) = (\theta/(\alpha-1))(1 - (\theta/(k+\theta))^{\alpha-1})$

Solution S4C96-1. Loss elimination ratio = $E(X \wedge d)/E(X)$, which, for a Pareto distribution is

$((\theta/(\alpha-1))(1 - (\theta/(d+\theta))^{\alpha-1}))/(\theta/(\alpha-1)) = (1 - (\theta/(d+\theta))^{\alpha-1})$. Since θ is the scale parameter for a Pareto distribution, the year 5461 value of θ will be 0.6 times the year 5460 value of 300, i.e., θ = 180 in 5461, and α is unchanged at 4. Therefore, our loss elimination ratio is

$1 - (180/(50+180))^3$ = **0.5206706666.**

Problem S4C96-2. Similar to Question 128 of the Exam C Sample Questions from the Society of Actuaries. An insurance company charges a premium that is 1.2 times the expected loss in a year.

Losses are either 4 with probability 0.6, 6 with probability 0.2, or 10 with probability 0.2.

The insurance company starts out with a surplus of 4. It collects its premium at the beginning of each year. There are no interest payments or inflation. Find the probability that the insurance company will not have funds to pay losses within the next two years.

Solution S4C96-2. The expected loss amount is 4*0.6 + 6*0.2 + 10*0.2 = 5.6. Thus, the premium the insurance company collects each year is 1.2*5.6 = 6.72.

Thus, in year 1, the insurance company has a total of 6.72 + 4 = 10.72.

Irrespective of what loss occurs in year 1, the company will not run out of money.

At the beginning of year 2, the company will collect another premium of 6.72.

If a loss of 4 occurs in year 1, the company will have 10.72 - 4 + 6.72 > 10.72, which will suffice to withstand any loss that occurs in year 2.

If a loss of 6 occurs in year 1, the company will have 10.72 - 6 + 6.72 > 10.72, which will suffice to withstand any loss that occurs in year 2.

If a loss of 10 occurs in year 1, the company will have 10.72 - 10 + 6.72 = 7.44, which will only result in ruin for the company if the loss in year 2 is 10. Thus, we want to find Pr(losses in both years are 10) = 0.2^2 = **0.04.**

Problem S4C96-3. Similar to Question 130 of the Exam C Sample Questions from the Society of Actuaries. The random variable N follows a Poisson distribution with mean Λ, where Λ is uniformly distributed on the interval from 4 to 10. Let Q = 3^W. Find E(Q).

Relevant property of Poisson distributions: Pr(N) = $e^{-\lambda}*\lambda^n/n!$

Solution S4C96-3. We use the formula E(Q) = $E_\Lambda(E(Q \mid \Lambda))$. We note that N is discrete, so Q is discrete, while Λ is continuous. So in calculating E(Q│Λ), we use a summation, and in calculating $E_\Lambda(.)$, we use an integral.

First, we find E(Q│Λ) = ${}_{n=0}{}^{\infty}\Sigma(3^n*e^{-\lambda}*\lambda^n/n!)$ = ${}_{n=0}{}^{\infty}\Sigma(e^{-\lambda}*(3\lambda)^n/n!)$. We note that, since the sum of all possible probabilities for a random variable must equal 1, ${}_{n=0}{}^{\infty}\Sigma(e^{-3\lambda}*(3\lambda)^n/n!)$, which is the sum of all possible probabilities for a Poisson random variable with mean λ, must equal 1.

Therefore, ${}_{n=0}{}^{\infty}\Sigma(e^{-\lambda}*(3\lambda)^n/n!)$ = ${}_{n=0}{}^{\infty}\Sigma(e^{-3\lambda}*(3\lambda)^n/n!)/e^{-2\lambda}$ = $1/e^{-2\lambda}$ = $e^{2\lambda}$.

Thus, E(Q) = $E_\Lambda(e^{2\lambda})$ = ${}_4{}^{10}\int e^{2\lambda}*f(\lambda)*d\lambda$ = ${}_4{}^{10}\int e^{2\lambda}*(1/6)*d\lambda$ = $(1/12)e^{2\lambda}\Big|_4^{10}$ =

$(1/12)(e^{20}- e^8)$ = **E(Q) =40430184.54.**

Problem S4C96-4. Similar to Question 134 of the Exam C Sample Questions from the Society of Actuaries. You have the following sample of 21 values:

45, 47, 56, 59, 61, 64, 68, 76, 80, 81, 82, 83, 87, 89, 90, 94, 96, 100, 106, 110, 111

Find the smoothed empirical estimate of the 76th percentile.

Solution S4C96-4. Since the sample has 21 values, the 76th percentile can be approximated by the 0.76*(21+1)th = the 16.72nd sample value, with linear interpolation being used between the 16th and 17th values. The 16th value is 94; the 17th value is 96. The difference between them is 2. Thus, our estimate is 94 + 2*0.72 = **95.44.**

Problem S4C96-5. Similar to Question 136 of the [Exam C Sample Questions] from the Society of Actuaries. There are two possible groups of gamblers, A and B.

Gamblers in group A have a probability of 0.3 of losing 1000, a probability of 0.5 of losing 5000, and a probability of 0.2 of losing 10000.

Gamblers in group B have a probability of 0.1 of losing 1000, a probability of 0.3 of losing 5000, and a probability of 0.6 of losing 10000.

There are 3 times as many gamblers in group A as in group B.

A loss of 5000 was observed. Find the Bayesian estimate of the expected value of the next loss from the same gambler.

Solution S4C96-5. We first want to find Pr(A | Loss of 5000) and Pr(B | Loss of 5000).

We note that Pr(A | Loss of 5000) = Pr(A and Loss of 5000)/Pr(Loss of 5000).

Since there are 3 times as many gamblers in group A as in group B, Pr(A) = ¾ and Pr(B) = ¼. Thus, Pr(A and Loss of 5000) = (3/4)(0.5) = 3/8, and

Pr(B and Loss of 5000) = (1/4)(0.3) = 3/40.

Thus, Pr(Loss of 5000) = 3/8 + 3/40 = 9/20.

Pr(A | Loss of 5000) = (3/8)/(9/20) = 5/6, and, correspondingly, Pr(B | Loss of 5000) = 1/6.

Now we also find E(A) = 0.3*1000 + 0.5*5000 + 0.2*10000 = 4800 and

E(B) = 0.1*1000 + 0.3*5000 + 0.6*10000 = 7600.

Our posterior expected value is thus (5/6)*E(A) + (1/6)*E(B) =

(5/6)*4800 + (1/6)*7600 = **5266.66666667.**

Section 97

Assorted Exam-Style Questions for Exam 4/C – Part 13

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C97-1. Similar to Question 137 of the Exam C Sample Questions from the Society of Actuaries. You know that the random variable X follows a distribution with probability density function $f(x) = (6 + p)*x^p$, where $0 < x < 1$, and $p > -6$.

You observe the following sample of 4 values:

0.46, 0.47, 0.86, 0.98.

Find the maximum likelihood estimate of p.

Solution S4C97-1. We first find the likelihood function L(p).

$L(p) = ((6 + p)*0.46^p)((6 + p)*0.47^p)((6 + p)*0.86^p)((6 + p)*0.98^p) \rightarrow$

$L(p) = (6 + p)^4*0.18221336^p$.

We can now find the loglikelihood function l(p):
$l(p) = \ln((6 + p)^4*0.18221336^p) \rightarrow$

$l(p) = 4*\ln(6+p) + p*\ln(0.18221336) \rightarrow$

$l'(p) = 4/(6+p) + \ln(0.18221336) = 0$ at the maximum $\rightarrow$

$4/(6+p) = -\ln(0.18221336) \rightarrow$

p = 4/-ln(0.18221336) - 6 = **p = -3.650620167**.

Problem S4C97-2. Similar to Question 138 of the Exam C Sample Questions from the Society of Actuaries. You are given the following sample of values:

0, 0, 0, 3, 6, 6.

You fit a binomial distribution with parameters m and q to this sample, with the following conditions:

- The mean of the binomial distribution must equal the sample mean.

- The smoothed empirical estimate of the 3^{rd} percentile must be equal to 3^{rd} percentile of the fitted binomial distribution.

What is the minimum value of m that must be used in order for the above two conditions to hold?

Relevant properties of binomial distributions: E(N) = mq.

Solution S4C97-2. The sample mean is (0 + 0 + 0 + 3 + 6 + 6)/6 = 2.5, which is also equal to mq. The 3^{rd} percentile smoothed empirical estimate is found by examining the 0.03*(6+1)th = 0.21st sample value. Since the first 3 sample values are 0, the 3^{rd} percentile smoothed empirical estimate is also 0. We want to find the smallest m such that Pr(N = 0) is at least 0.03.

In a binomial distribution, Pr(N = 0) = $(1\text{-}q)^m$. As mq = 2.5, q = 2.5/m, and thus
Pr(N = 0) = $(1\text{-}2.5/m)^m$ > 0.3.
Since mq > 2, m must also be 3 or greater.
We test m = 3: $(1\text{-}2.5/3)^3$ = 0.00462962963.
We test m = 4: $(1\text{-}2.5/4)^4$ = 0.019775390625
We test m = 5: $(1\text{-}2.5/5)^5$ = 0.03125 > 0.03. Thus, our minimum value of m is **m = 5.**

Problem S4C97-3. Similar to Question 141 of the Exam C Sample Questions from the Society of Actuaries. The 92% log-transformed confidence interval for the cumulative hazard rate function H(t) at time t is (0.456, 0.645). The cumulative hazard rate function is estimated via the Nelson-Åalen estimator. Find the Nelson-Åalen estimate of the survival function.

Solution S4C97-3. First, we find the Nelson-Åalen estimate of the cumulative hazard rate function. We know that for the log-transformed confidence interval, the lower limit is $\hat{H}(t)*U^-$, and the upper limit is $\hat{H}(t)*U^+$, where U^- = $1/U^+$. Thus, multiplying the two bounds together should give us $\hat{H}(t)$ = 0.456*0.645 = 0.29412.

The estimate of the survival function is $\hat{S}(t)$ = exp(-$\hat{H}(t)$) = exp(-0.29412) = **$\hat{S}(t)$ = 0.7451870636**.

Problem S4C97-4. Similar to Question 143 of the Exam C Sample Questions from the Society of Actuaries. Random variable X follows an inverse Pareto distribution with parameters θ and τ. You have the following sample of 3 values:

2, 4, 8.

Use the method of moment matching and the (-1)st and (-2)nd moments to find an estimate for θ.

Relevant property of inverse Pareto distributions: If k is a negative integer,
$E(X^k) = \theta^k*(-k)!/((\tau-1)...(\tau+k))$.

Solution S4C97-4. We find the empirical average of the values of X^{-1}: = ((1/2) + (1/4) + (1/8))/3 = 7/24.

We find the empirical average of the values of X^{-2}: = $((1/2^2) + (1/4^2) + (1/8^2))/3 = 7/64$.
We find $E(X^{-1}) = \theta^{-1}*1!/(\tau-1) = 1/(\theta(\tau-1))$.
We find $E(X^{-1}) = \theta^{-2}*2!/((\tau-1)(\tau-2)) = 2/(\theta^2(\tau-1)(\tau-2))$.

We have the following equations:
i) $1/(\theta(\tau-1)) = 7/24$;
ii) $2/(\theta^2(\tau-1)(\tau-2)) = 7/64$.

We square both sides of equation i) to get
iii) $1/(\theta^2(\tau-1)^2) = 49/576$.

We divide iii) by ii) to get
$(\tau-2)/2(\tau-1) = 7/9$, and so $(\tau-2)/(\tau-1) = 14/9$.
$(\tau-2) = (14/9)\tau - 14/9 \rightarrow$
$(5/9)\tau = -4/9 \rightarrow \tau = (-4/5) = -0.8$.
Thus, as $7/24 = 1/(\theta(\tau-1)) = 1/(\theta(-0.8-1)) = 1/(-1.8\theta)$, and so $-1.8\theta = 24/7$ and **θ = -1.904761905**.

Problem S4C97-5. Similar to Question 148 of the Exam C Sample Questions from the Society of Actuaries. You are using classical (limited fluctuation) credibility theory and a distribution known to be binomial with parameters m and q. You know that the number of claims originating from this distribution required for full credibility is 6000. You want the actual number of claims to be within 0.02 of the expected number of claims 97% of the time. Find q for the binomial distribution in question.

Relevant properties of binomial distributions: E(X) = mq; Var(X) = mq(1-q).

Solution S4C97-5. We know that $6000 = n = \lambda_0(\text{Var}(X)/E(X))$, where $\lambda_0 = (\Phi^{-1}((1+0.97)/2)/0.02)^2 = (\Phi^{-1}(0.985)/0.02)^2$, which we can find via the Excel input "=(NORMSINV(0.985)/0.02)^2", giving us $\lambda_0 = 11773.23062$. Thus, we know that
6000 = 11773.23062(Var(X)/E(X)) = 11773.23062(mq(1-q)/(mq)) and so
(1-q) = 6000/11773.23062 = 0.5096307202, implying that **q = 0.4903692798**.

Section 98

Assorted Exam-Style Questions for Exam 4/C – Part 14

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C98-1. Similar to Question 146 of the Exam C Sample Questions from the Society of Actuaries. There are 7 insureds with losses in Territory A and 5 insureds with losses in Territory B. In each territory, a different single-parameter Pareto distribution with $\theta = 1$ is used to model losses. Expected losses in territory B are three times the expected losses in territory A. Let α be the maximum likelihood estimate of the parameter in Territory A, calculated using the data from both territories.

Let x_i for i = 1, 2, ..., 7 denote the individual losses in Territory A.
Let y_i for i = 1, 2, ..., 5 denote the individual losses in Territory B.

Find the equation that must be solved for α, expressed in terms of α and summation operations performed on the x_i and y_i and with 0 on one side of the equation.

Relevant properties of single-parameter Pareto distributions: $f(x) = \alpha\theta^\alpha/x^{\alpha+1}$.
$E(X) = \alpha\theta/(\alpha-1)$.

Solution S4C98-1. Let a be the unknown parameter for Territory B.

Since $\theta = 1$, $E(X) = \alpha/(\alpha-1)$ for Territory A and $a/(a-1) = 3\alpha/(\alpha-1)$ for Territory B.

We want to find a in terms of α:

$3\alpha(a-1) = a(\alpha-1) \rightarrow 3\alpha a - 3\alpha = \alpha a - a \rightarrow$
$2\alpha a = 3\alpha - a \rightarrow$
$2\alpha a + a = 3\alpha \rightarrow$

$a(2\alpha + 1) = 3\alpha \rightarrow$
$a = 3\alpha/(2\alpha + 1)$.

Since $\theta = 1$, each individual $f(x) = \alpha/x^{\alpha+1}$ or a/x^{a+1}.

Now we find the likelihood function $L(\alpha)$:

$L(\alpha) = {}_{i=1}{}^{7}\Pi(\alpha/x_i{}^{\alpha+1})*{}_{i=1}{}^{5}\Pi(a/y_i{}^{a+1}) \rightarrow$
$l(\alpha) = \ln({}_{i=1}{}^{7}\Pi(\alpha/x_i{}^{\alpha+1})*{}_{i=1}{}^{5}\Pi(a/y_i{}^{a+1})) \rightarrow$
$l(\alpha) = {}_{i=1}{}^{7}\Sigma(\ln(\alpha/x_i{}^{\alpha+1})) + {}_{i=1}{}^{5}\Sigma(\ln(a/y_i{}^{a+1})) \rightarrow$
$l(\alpha) = {}_{i=1}{}^{7}\Sigma(\ln(\alpha)) - (\alpha+1){}_{i=1}{}^{7}\Sigma(\ln(x_i)) + {}_{i=1}{}^{5}\Sigma(\ln(a)) - (a+1){}_{i=1}{}^{5}\Sigma(\ln(y_i))$.

Now we substitute $3\alpha/(2\alpha + 1)$ for a:
$l(\alpha) = {}_{i=1}{}^{7}\Sigma(\ln(\alpha)) - (\alpha+1){}_{i=1}{}^{7}\Sigma(\ln(x_i)) + {}_{i=1}{}^{5}\Sigma(\ln(3\alpha/(2\alpha + 1))) - (3\alpha/(2\alpha + 1)+1){}_{i=1}{}^{5}\Sigma(\ln(y_i)) \rightarrow$

$l(\alpha) = 7*\ln(\alpha) - (\alpha+1){}_{i=1}{}^{7}\Sigma(\ln(x_i)) + 5*(\ln(3\alpha/(2\alpha + 1))) - (3\alpha/(2\alpha + 1)+1){}_{i=1}{}^{5}\Sigma(\ln(y_i)) \rightarrow$

To find $l'(\alpha)$, it would first be useful to find the derivative of $3\alpha/(2\alpha + 1) =$

$(3\alpha)'/(2\alpha + 1) + (1/(2\alpha + 1))'*3\alpha = 3/(2\alpha + 1) - 2*3\alpha/(2\alpha + 1)^2 =$

$(3(2\alpha + 1) - 2*3\alpha)/(2\alpha + 1)^2 = 3/(2\alpha + 1)^2$.

$l'(\alpha) = 7/\alpha - {}_{i=1}{}^{7}\Sigma(\ln(x_i)) + 5*((2\alpha + 1)/(3\alpha))*(3/(2\alpha + 1)^2) - (3/(2\alpha + 1)^2){}_{i=1}{}^{5}\Sigma(\ln(y_i)) \rightarrow$

$l'(\alpha) = 7/\alpha - {}_{i=1}{}^{7}\Sigma(\ln(x_i)) + 5/(\alpha(2\alpha + 1)) - (3/(2\alpha + 1)^2){}_{i=1}{}^{5}\Sigma(\ln(y_i)) = 0$ at the maximum.

Thus, the equation that must be solved is

$\mathbf{7/\alpha - {}_{i=1}{}^{7}\Sigma(\ln(x_i)) + 5/(\alpha(2\alpha + 1)) - (3/(2\alpha + 1)^2){}_{i=1}{}^{5}\Sigma(\ln(y_i)) = 0}$.

Problem S4C98-2. Similar to Question 150 of the Exam C Sample Questions from the Society of Actuaries. You observe a sample of 4 insurance company loss payments:

35, 37, 50, 50.

The insurance company only pays losses up to a policy limit of 50. You know that losses are uniformly distributed on the interval from 0 to θ. You use the mean of the sample to estimate expected insurance company loss payments. Use this procedure to find θ.

Solution S4C98-2. Let X denote the losses. The sample mean is (35+37+50+50)/4 = 43. Note that this is not E(X), but rather E(X Λ 50), i.e., the expected value of what the insurance company has to pay. As the distribution of X is uniform from 0 to θ, $f(x) = 1/\theta$. We can calculate E(X Λ 50) in terms of θ: $E(X \Lambda 50) = {}_0{}^{50}\int x(1/\theta)dx + S(50)*50$.

For this uniform distribution, $S(50) = (\theta - 50)/\theta$.

Thus, E(X Λ 50) = $x^2/(2\theta) \Big|_0^{50}$ + 50(θ - 50)/θ = 1250/θ + 50(θ - 50)/θ = 43 by our estimation method.

Thus, 1250 + 50(θ - 50) = 43θ →
50θ - 1250 = 43θ →
1250 = 7θ → θ = 1250/7 = **θ = 178.5714286**.

Problem S4C98-3. Similar to Question 151 of the Exam C Sample Questions from the Society of Actuaries. There are two classes of risks, each with the same number of members.

Frequency of losses in class A per year follows a Poisson distribution with λ = 6.

Frequency of losses in class B per year follows a binomial distribution with m = 10 and q = 0.3.

You randomly select a risk and observe it to have 6 losses in year 1, 6 losses in year 2, s losses in year 3, and 5 losses in year 4. The credibility-weighted prediction for the number of losses for this risk in year 5, calculated using Bühlmann credibility, is 4.858565737.

Find the value of s.

Relevant properties of binomial distributions: E(X) = mq; Var(X) = mq(1-q).

Solution S4C98-3. The Bühlmann credibility formula is Estimate = Z*X⁻ + (1-Z)*M, where M is the mean frequency of losses for the entire group of risks.

M = (E(A) + E(B))/2, where E(A) = 6 and E(B) = 10*0.3 = 3, so M = (6+3)/2 = 4.5.
We also want to find Z = N/(N + K), where N = the number of observed years = 4 and K = EPV/VHM.
We find EPV = (Var(A) + Var(B))/2 = (5 + 10*0.3*0.7)/2 = EPV = 3.55.
We find VHM = E(Total^2) - E$(\text{Total})^2$. We already calculated E(Total) = M = 4.5.
E(Total^2) = $(6^2+3^2)/2$ = 22.5. Thus, VHM = 22.5 - 4.5^2 = 2.25.
Thus, K = 3.55/2.25 = 1.577777778 and Z = 4/(4+1.577777778) = Z = 0.7171314741.
Thus, 4.858565737 = 0.7171314741*X⁻ + (1-0.7171314741)*4.5 →
4.858565737 - (1-0.7171314741)*4.5 = 0.7171314741*X⁻ →
X⁻ = (4.858565737 - (1-0.7171314741)*4.5)/0.7171314741 = 5.
We know that X⁻ = 5 = (6 + 6 + s + 5)/4, and so 20 = 6 + 6 + s + 5 = 17 + s, implying that **s = 3.**

Problem S4C98-4. Similar to Question 152 of the Exam C Sample Questions from the Society of Actuaries. You have a sample of observations greater than 70:

90, 200, 230, 500.

No information is available on observations less than 70.

You know that observations follow an exponential distribution with mean θ. Find the maximum likelihood estimate of θ.

Solution S4C98-4. Because of the memoryless property of exponential distributions, losses in excess of 70 follow the same exponential distribution as total losses.

Thus, the maximum likelihood estimate of θ is the same as the mean of the following sample (since the mean of the sample is the maximum likelihood estimate for the mean of an exponential distribution): 20, 130, 160, 430.

Our desired estimate is (20 + 130 + 160 + 430)/2 = **θ = 370**.

Problem S4C98-5. Similar to Question 155 of the Exam C Sample Questions from the Society of Actuaries. You have the following sample of values:
4, 5, 7, 9, 11

You fit the values to an inverse Weibull distribution by matching at the 20^{th} and 60^{th} percentiles. Find the parameter θ of the fitted inverse Weibull distribution.

Relevant properties of inverse Weibull distributions:
$F(x) = \exp(-(\theta/x)^\tau)$.

Solution S4C98-5. As there are 5 values in the sample, the 20^{th} percentile is the first sample value or 4.
The 60^{th} percentile is the 3^{rd} sample value or 7.

Thus, we have the following system of equations:
i) $F(4) = 0.2 = \exp(-(\theta/4)^\tau)$;
ii) $F(7) = 0.6 = \exp(-(\theta/7)^\tau)$;

We take the natural logarithms of both equations:
i)' $\ln(0.2) = -(\theta/4)^\tau$;
ii)' $\ln(0.6) = -(\theta/7)^\tau$.

We divide i)' by ii)' to get $3.150660103 = (7/4)^\tau$, implying that $\tau = \ln(3.150660103)/\ln(7/4) = 2.050714101$.

Therefore, $\ln(0.2) = -(\theta/4)^{2.050714101} \rightarrow$
$(-\ln(0.2))^{1/2.050714101} = \theta/4 = 1.261193111$, implying that $\theta = 4*1.261193111 =$ **θ = 5.044772445**.

Section 99

Assorted Exam-Style Questions for Exam 4/C – Part 15

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C99-1. Similar to Question 156 of the Exam C Sample Questions from the Society of Actuaries. A random variable N follows a Poisson distribution with mean λ. All observations greater than 1 have been removed from the data. You observe a sample where the number of values of 0 is twice the number of values of 1. Find the maximum likelihood estimate for λ. Refer to the Exam 4 / C Tables as necessary.

Solution S4C99-1. Let n be the number of values of 1 in the sample. Then 2n is the number of values of 0 in the sample. In a Poisson distribution, $\Pr(N = 0) = e^{-\lambda}$ and $\Pr(N = 1) = \lambda e^{-\lambda}$. We cannot say anything about data for N > 1, since it has been removed from the sample. Thus, in our likelihood functions, we can only consider $\Pr(N = 0 \mid N = 0 \text{ or } 1)$ and $\Pr(N = 1 \mid N = 0 \text{ or } 1)$.

$\Pr(N = 0 \text{ or } 1) = e^{-\lambda} + \lambda e^{-\lambda} = e^{-\lambda}(1+\lambda)$.
Thus, $\Pr(N = 0 \mid N = 0 \text{ or } 1) = (e^{-\lambda})/(e^{-\lambda}(1+\lambda)) = 1/(1+\lambda)$.
$\Pr(N = 1 \mid N = 0 \text{ or } 1) = \lambda e^{-\lambda}/(e^{-\lambda}(1+\lambda)) = \lambda/(1+\lambda)$.
Our likelihood function is thus $L(\lambda) = (1/(1+\lambda))^{2n}*(\lambda/(1+\lambda))^n = \lambda^n/(1+\lambda)^{3n}$.
The loglikelihood function is $l(\lambda) = n*\ln(\lambda) - 3n*\ln(1+\lambda)$.
$l'(\lambda) = n/\lambda - 3n/(1+\lambda) = 0$ at the maximum →
$n/\lambda = 3n/(1+\lambda) \rightarrow 1+\lambda = 3\lambda \rightarrow 2\lambda = 1 \rightarrow \boldsymbol{\lambda = ½ = 0.5}$.

Problem S4C99-2. Similar to Question 158 of the Exam C Sample Questions from the Society of Actuaries. You have the following sample of 7 values:

5, 5, 6, 8, 8, 8, 20

You estimate the hazard rate function in two ways:

$\hat{H}_1(x)$ is the Nelson-Åalen estimate of the hazard rate function.
$\hat{H}_2(x)$ is the maximum likelihood estimate of the hazard rate function, where the sample is assumed to follow an exponential distribution.

Find $|\hat{H}_2(9) - \hat{H}_1(9)|$. Refer to the Exam 4 / C Tables as necessary.

Solution S4C99-2. We first find $\hat{H}_1(9)$. There are 2 observations of 5, and the risk set size associated with 5 is 7. There is 1 observation of 6, and the risk set size associated with 6 is 5.

There are 3 observations of 8, and the risk set size associated with 8 is 4. There are no other observations less than or equal to 9.

Thus, we apply the following formula:

$\hat{H}(x) = 0$ if $x < y_1$;
$\hat{H}(x) = {}_{i=1}{}^{j-1}\Sigma(s_i/r_i)$ if $y_{j-1} \leq x < y_j$, for $j = 2, ..., k$.
$\hat{H}(x) = {}_{i=j}{}^{k}\Sigma(s_i/r_i)$ if $x \geq y_k$.
$\hat{H}_1(9) = 2/7 + 1/5 + ¾ = 173/140 = 1.235714286$.

Now we find $\hat{H}_2(9)$. To do this, we first find our maximum likelihood estimate of θ. Since the underlying distribution is assumed to be exponential, our estimate is just the sample mean: $(5+5+6+8+8+8+20)/7 = \theta = 60/7 = 8.571428571$.

The estimate of the survival function is $\hat{S}(9) = \exp(-9/8.571428571) = 0.3499377491$.
$\hat{H}_2(9) = -\ln(\hat{S}(9)) = -\ln(0.3499377491) = \hat{H}_2(9) = 1.05$.
Thus, we find $|\hat{H}_2(9) - \hat{H}_1(9)| = |1.05 - 1.235714286| =$ **0.1857142857.**

Problem S4C99-3. Similar to Question 159 of the Exam C Sample Questions from the Society of Actuaries. The number of earthquakes follows a conditional Poisson distribution. For a single year and 2840 regions, you observe the following data:

1000 regions had 0 earthquakes.
1500 regions had 1 earthquake.
300 regions had 2 earthquakes.
40 regions had 3 earthquakes.
No region had more than 3 earthquakes.

Use semiparametric empirical Bayes estimation to find the Bühlmann credibility factor Z for the next year of observations.

Solution S4C99-3. Z = N/(N+K), where N = number of time periods under observation = 1. K = EPV^/VHM^, where EPV^ = E(Total) = (1000*0 + 1500*1 + 300*2 + 40*3)/2840 = 0.7816901408.

To find VHM^, we use the formula

VHM^ = $(1/(R-1))*_i\Sigma(E(R_i) - E(Total))^2$ -EPV^/N.

Here, R = 2840 and N = 1, so VHM^ = $(1/2839)*(1000(0-0.7816901408)^2 + 1500(1-0.7816901408)^2 + 300(2-0.7816901408)^2 + 40(3-0.7816901408)^2)$ = VHM^ = 0.4665896046.

Thus, K = 0.7816901408/0.4665896046 = 1.675326954, and Z = 1/(1+1.675326954) = **Z = 0.3737860894**.

Problem S4C99-4. Similar to Question 162 of the Exam C Sample Questions from the Society of Actuaries. You are examining a Pareto distribution where α = 2 and θ is unknown. You do know, however, that E(X - 1100 │ X > 1100) = 3*E(X - 60 │ X > 60).

Use this information to find E(X - 2000 │ X > 2000).

Relevant properties regarding Pareto distributions: E(X) = θ/(α-1);

$E(X \wedge k) = (\theta/(\alpha-1))(1 - (\theta/(k+\theta))^{\alpha-1})$; $S(x) = (\theta/(x+\theta))^{\alpha}$

Solution S4C99-4. First, it would be useful to have a general expression for

E(X - k │ X > k) in cases where α = 2:

$E(X - k \mid X > k) = (E(X) - E(X \wedge k))/S(k) = (\theta/(\alpha-1) - (\theta/(\alpha-1))(1 - (\theta/(k+\theta))^{\alpha-1}))/(\theta/(k+\theta))^{\alpha} =$

$(\theta/(2-1) - (\theta/(2-1))(1 - (\theta/(k+\theta))^{2-1}))/(\theta/(k+\theta))^2 =$

$(\theta - \theta(1 - (\theta/(k+\theta)))/(\theta/(k+\theta))^2 = (\theta^2/(k+\theta)))/(\theta/(k+\theta))^2 = k + \theta.$

Thus, E(X - 1100 │ X > 1100) = 1100 + θ, and E(X - 60 │ X > 60) = 60 + θ.

We need to solve the following equation: 1100 + θ = 3(60 + θ)→ 1100 + θ = 180 + 3θ → 2θ = 920 and θ = 460. Thus, E(X - 2000 │ X > 2000) = 2000 + 460 = **2460.**

Problem S4C99-5. Similar to Question 163 of the Exam C Sample Questions from the Society of Actuaries. A normal distribution of random variable X has mean λ and variance 64. We also know that λ follows another normal distribution with mean 25 and variance 225. Find the conditional probability that X is less than 30, given that X is greater than 20.

Solution S4C99-5. First, we want to find the mean and varianceof the unconditional distribution of X: $E(X) = E_\Lambda(E(X│\Lambda)) = E_\Lambda(\Lambda) = 25$.
$Var(X) = Var_\Lambda(E(X│\Lambda)) + E_\Lambda(Var(X│\Lambda)) = Var_\Lambda(\Lambda) + E_\Lambda(64) = 225 + 64 = 289$, implying that SD(X) = 17.
Thus, our desired probability is (Φ((30-25)/17) - Φ((20-25)/17))/(1 - Φ((20-25)/17)) = (Φ(5/17) - Φ(-5/17))/(1 - Φ(-5/17)). We can find the answer in MS Excel using the input "=(NORMSDIST(5/17)-NORMSDIST(-5/17))/(1-NORMSDIST(-5/17))". Our answer is **0.375742683**.

Section 100

Assorted Exam-Style Questions for Exam 4/C – Part 16

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C100-1. Similar to Question 164 of the Exam C Sample Questions from the Society of Actuaries. You are working with a collective risk model, where the number of losses, N, follows a Poisson distribution with mean 46. You also know the following attributes of loss severity, X, for each individual loss:

$E(X) = 100$;
$E(X \wedge 50) = 45$;
$\Pr(X > 50) = 0.6$;
$E(X^2 \mid X > 30) = 12000$.

Every individual loss is insured and is subject to a deductible of 50. Find the expected value of aggregate payments for this insurance.

Solution S4C100-1. A deductible of 50 will reduce the frequency of losses that result in payments. Namely, only the 0.6 of insureds whose losses exceed 50 will receive payment. This means that the Poisson parameter for frequency will change from 46 to $46*0.6 = \lambda^* = 27.6$.

The relevant variable for severity is the size of each payment or $(X - 50) \mid X > 50$.

Because the frequency random variable is Poisson, our variance of aggregate payments (P) is $\mathrm{Var}(P) = (\lambda^*)E((X-50)^2 \mid X > 50)$.

We find $E((X-50)^2 \mid X > 50) = E((X^2-100X + 2500) \mid X > 50) =$
$E(X^2 \mid X > 50) - 100E(X \mid X > 50) + 2500$.
We know that $E(X^2 \mid X > 50) = 12000$.

$E(X | X > 50)$ is the mean excess loss function $E(X-50 | X > 50)$ plus 50. That is,
$E(X | X > 50) = 50 + (E(X) - E(X \wedge 50))/Pr(X > 50) = 50 + (100 - 45)/0.6 = 141.666666667$.
Thus, $E((X-50)^2 | X > 50) = 12000 - 100*141.666666667 + 2500 = 333.333333333$,
and $Var(P) = 27.6*333.333333333 =$ **Var(P) = 9200.**

Problem S4C100-2. Similar to Question 165 of the Exam C Sample Questions from the Society of Actuaries. The number of losses follows a Poisson distribution with mean 3.

The size of each loss is either 4 with probability 0.5 or 8 with probability 0.5. The insurance company pays all losses except a deductible of 12 imposed on the aggregate losses. Find the expected value of the insurance company payment.

Solution S4C100-2. Let N denote frequency, let X denote severity, and let S denote the aggregate losses. We want to find $E((S-12)_+)$, which is the left-censored and shifted random variable for S with 12 as the point of censoring.

We note that possible aggregate losses can change by no less than increments of 4.
The possible values of S are thus 0, 4, 8, 12, 16, 20, ...
Possible values of $((S-12)_+)$ are 0, 0, 0, 0, 4, 8, ...
We also consider values of S - 12: -12, -8, -4, 0, 4, 8...
Then $E((S-12)_+) = E(S - 12) + 12*Pr(S = 0) + 8*Pr(S = 4) + 4*Pr(S = 8) =$
$E(S) - 12 + 12*Pr(S = 0) + 8*Pr(S = 4) + 4*Pr(S = 8)$.
We find $E(S) = E(N)*E(X) = 3(0.5*4 + 0.5*8) = E(S) = 18$.
$S = 0$ can only occur if $N = 0$, so $Pr(S = 0) = e^{-3}$.
$S = 4$ can only occur if $N = 1$ and $X = 4$, so $Pr(S = 4) = (3e^{-3})*0.5 = 1.5e^{-3}$.
$S = 8$ can occur if $N = 1$ and $X = 2$ or if $N = 2$ and each $X = 1$, so $Pr(S = 8) =$
$(3e^{-3})*0.5 + (4.5e^{-3})*(0.5^2) = 1.125e^{-3}$.
Our answer is thus $18 - 12 + 12e^{-3} + 8*1.5e^{-3} + 4*1.125e^{-3} = 6 + 28.5e^{-3} =$
$E((S-12)_+) = 7.418931448$.

Problem S4C100-3. Similar to Question 166 of the Exam C Sample Questions from the Society of Actuaries. You have a distribution where $p(k) = (c + 2c/k)*p(k-1)$, where $c \neq 0$. You are given that $p(0) = 0.3$. Find c. Refer to the Exam 4 / C Tables as necessary.

Solution S4C100-3. We note that $p(k)/p(k-1) = (c + 2c/k)$, which means that the distribution belongs to the (a, b, 0) class with $a = c$ and $b = 2c$. We know that the distribution cannot be Poisson, because $a \neq 0$. The distribution cannot be geometric, because $b \neq 0$. The distribution cannot be binomial, because a and b do not have opposite signs. Thus, the distribution must be negative binomial. For a negative binomial distribution, $a = \beta/(1+ \beta)$ and $b = (r-1)\beta/(1+ \beta)$. Since here, $b = 2a$, it follows that $(r-1) = 2 \rightarrow r = 3$.

We also know that $p(0) = (1+ \beta)^{-r} = 0.3 = (1+ \beta)^{-3} \rightarrow 1+ \beta = 0.3^{-1/3} = 1.493801582$, and so $\beta =$ 0.493801582. We can now find $c = a = \beta/(1+ \beta) = 0.493801582/1.493801582 =$ **c = 0.3305670498**.

Problem S4C100-4. Similar to Question 167 of the Exam C Sample Questions from the Society of Actuaries. You have the following information:

There are 80 machines of type A. The probability that a machine of type A will need repairs is 0.2. If a machine of type A needs repairs, then cost of repairs has a mean of 500 and a variance of 64000.

There are 50 machines of type B. The probability that a machine of type A will need repairs is 0.4. If a machine of type B needs repairs, then cost of repairs has a mean of 1000 and a variance of 75000.

Find the sum of the mean of aggregate repair costs and the standard deviation of aggregate repair costs.

Solution S4C100-4. Let N denote frequency of repairs, let X denote severity of repairs, and let S denote the aggregate cost of repairs.

For type A, E(N) = 80*0.2 = 16 and Var(N) = $16*(1 - 0.2)^2 + 64*(0 - 0.2)^2 = 12.8$.
E(X) = 500 and Var(X) = 64000.
E(S) = E(N)*E(X) = 16*500 = 8000, and Var(S) = E(N)*Var(X) + Var(N)*$E(X)^2$ = 16*64000 + $12.8*500^2$ = 4224000.

For type B, E(N) = 50*0.4 = 20 and Var(N) = $20(1 - 0.4)^2 + 30(0 - 0.4)^2 = 12$.
E(X) = 1000 and Var(X) = 75000.
E(S) = E(N)*E(X) = 20*1000 = 20000, and Var(S) = E(N)*Var(X) + Var(N)*$E(X)^2$ = 20*75000 + $12*1000^2$ = 13500000.

For the entire population of machines, E(S) = 8000 + 20000 = 28000, and
Var(S) = 4224000 + 13500000 = 17724000, implying that SD(S) = √(17724000) = 4209.988123.
Thus, our desired answer, E(S) + SD(S) = 28000 + 4209.988123 = **32209.98812.**

Problem S4C100-5. Similar to Question 168 of the Exam C Sample Questions from the Society of Actuaries. Losses are 650 with probability 0.4, 780 with probability 0.3, and 860 with probability 0.3. All claims have a deductible of 700 imposed on them. Let Y^P be the per-payment payment random variable. Calculate Var(Y^P).

Solution S4C100-5. With a deductible of 700, the only payments that are made are (780-700) = 80 with *per-loss* probability 0.3 and (860-700) = 160 with *per-loss* probability 0.3. To get the respective per-payment probabilities, we need to divide the per-loss probabilities by the probability that a payment will be made on a loss, which, in this case, is 1 - 0.4 = 0.6. Thus, the per-payment probability of each of the two possible payments is 0.3/0.6 = 0.5.

Hence, E(Y^P) = 0.5*80 + 0.5*160 = 120, and Var(Y^P) = $0.5*(80-120)^2 + 0.5*(160-120)^2$ =
Var(Y^P) = 1600.

Section 101

Assorted Exam-Style Questions for Exam 4/C – Part 17

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C101-1. Similar to Question 169 of the Exam C Sample Questions from the Society of Actuaries. The distribution of X is a two-point mixture of two Pareto distributions. One of the Pareto distributions occurs with probability 0.4 and has parameters $\alpha = 3$ and $\theta = 90$. The other Pareto distribution occurs with probability 0.6 and has parameters $\alpha = 5$ and $\theta = 130$.

Find $\Pr(X \leq 70)$.

Relevant property of Pareto distributions: $F(x) = 1 - (\theta/(x+\theta))^\alpha$.

Solution S4C101-1. $\Pr(X \leq 70)$ is the probability-weighted average of F(70) for each of the two Pareto distributions of which the mixture consists. Thus,

$\Pr(X \leq 70) = 0.4*(1 - (90/(70+90))^3) + 0.6*(1 - (130/(70+130))^5) =$ **$\Pr(X \leq 70) = 0.8591911563$**.

Problem S4C101-2. Similar to Question 170 of the Exam C Sample Questions from the Society of Actuaries. The number of peanuts eaten by elephants varies by the groups to which the elephants belong.

The probability of an elephant being in group A is 0.6. Elephants in group A eat a number of peanuts that is Poisson-distributed with mean 5.

The probability of an elephant being in group B is 0.2. Elephants in group B eat a number of peanuts that is Poisson-distributed with mean 2.

The probability of an elephant being in group C is 0.2. Elephants in group C eat a number of peanuts that is Poisson-distributed with mean 6.

Given that an elephant is observed eating exactly 4 peanuts, what is the conditional probability that the elephant belongs to group B?

Relevant property of Poisson distributions: $\Pr(N = k) = e^{-\lambda}*\lambda^k/k!$

Solution S4C101-2. We want to find Pr(B│4) = Pr(B and 4)/Pr(4).
Pr(4) = Pr(A and 4) + Pr(B and 4) + Pr(C and 4).
Pr(A and 4) = $0.6*e^{-5}*5^4/24$ = 0.1052804219.
Pr(B and 4) = $0.2*e^{-2}*2^4/24$ = 0.0180447044.
Pr(C and 4) = $0.2*e^{-6}*6^4/24$ = 0.0267705235.

Thus, Pr(B│4) = 0.0180447044/(0.1052804219 + 0.0180447044 + 0.0267705235) = **Pr(B│4) = 0.1202213685**.

Problem S4C101-3. Similar to Question 171 of the Exam C Sample Questions from the Society of Actuaries. Loss frequency (N) follows a negative binomial distribution with mean 7 and variance 9.12. Loss severity (X) follows a uniform distribution from 0 to 60. Loss frequency and severity are independent. Use a normal approximation to find the 90th percentile of aggregate losses (S).

Solution S4C101-3. We find E(X) = (0+60)/2 = 30 and Var(X) = $(60-0)^2/12$ = 300.
We find the mean and variance of aggregate losses.
E(S) = E(N)*E(X) = 7*30 = 210.
Var(S) = $E(N)*Var(X) + Var(N)*E(X)^2 = 7*300 + 9.12*30^2$ = Var(S) = 10308.
Thus, SD(S) = $10308^{1/2}$ = 101.5283212.
We want to find a value s such that 0.90 = $\Phi((s-210)/101.5283212)$.

We find $\Phi^{-1}(0.90)$ via the Excel input "=NORMSINV(0.90)", getting as our result 1.281551566.
Thus, 1.281551566 = (s-210)/101.5283212, and s = 1.281551566*101.5283212 + 210 = **340.113779**.

Problem S4C101-4. Similar to Question 173 of the Exam C Sample Questions from the Society of Actuaries. Loss frequency (N) follows a negative binomial distribution with parameters r and β = 2. Loss severity (X) is 100 with probability 0.5, 200 with probability 0.4, and 300 with probability 0.1. Loss frequency and severity are independent. Use classical credibility theory to determine the expected number of losses needed for aggregate losses to be within 12% of expected aggregate losses with 93% probability.

Relevant properties of negative binomial distributions: E(N) = rβ; Var(N) = rβ(1+β).

Solution S4C101-4. First we find the value $\lambda_0 = (\Phi^{-1}((1+0.93)/2)/0.12)^2$ =

$(\Phi^{-1}(0.965)/0.12)^2$, which we obtain via the Excel input "=(NORMSINV(0.965)/0.12)^2", getting 227.9875199 as our result. The number of *observations* needed for full credibility is thus $\lambda_0(\text{Var}(S)/E(S)^2)$.

We find E(N) = rβ = 2r, and Var(N) = 2r(1+2) = 6r.

We find E(X) = 0.5*100 + 0.4*200 + 0.1*300 = 160 and Var(X) = $0.5*(100-160)^2 + 0.4(200-160)^2 + 0.1(300-160)^2 = 4400$.

We find E(S) = E(N)*E(X) = 2r*160 = 320r.

Var(S) = $E(N)*\text{Var}(X) + \text{Var}(N)*E(X)^2 = 2r*4400 + 6r*160^2$ = Var(S) = 162400r.

Thus, $(\text{Var}(S)/E(S)^2) = 162400r/(320r)^2 = 1.5859375/r$, and

$\lambda_0(\text{Var}(S)/E(S)^2) = 227.9875199(1.5859375/r) = 361.5739573/r$. Since each observation has expected loss number of 2r, the expected number of losses we desire is (361.5739573/r)*2r = 723.1479147 = **at least 724 losses.**

Problem S4C101-5. Similar to Question 176 of the Exam C Sample Questions from the Society of Actuaries. You are analyzing the following sample:

4, 4, 7, 7, 8, 8, 8, 10

You make a p-p plot where you are attempting to fit the data to an exponential distribution with mean 7, whose cumulative distribution function is F(x). You plot the values of the empirical distribution function $F_n(x)$ on the horizontal axis and the values of F(x) on the vertical axis. How many points on the plot can you expect to be above the 45-degree line (slope = 1 line, starting at the origin)? How many can you expect to be above the 45-degree line?

Solution S4C101-5. Here, n = 8, and $F_n(4) = 2/8 = ¼$. $F(4) = 1 - e^{-4/7} = 0.435281878$.

Thus, $F(4) > F_n(4)$, meaning that the corresponding point is above the 45-degree line.

$F_n(7) = 2/8 + 2/8 = ½$. $F(7) = 1 - e^{-7/7} = 0.6321205588$. Thus, $F(7) > F_n(7)$, meaning that the corresponding point is above the 45-degree line.

$F_n(8) = ½ + 3/8 = 7/8$. $F(8) = 1 - e^{-8/7} = 0.6810934427$. Thus, $F(8) < F_n(8)$, meaning that the corresponding point is below the 45-degree line.

As $F_n(10) = 1$, it is clear that $F(10) < F_n(10)$, meaning that the corresponding point is below the 45-degree line. Thus, **2 points are below the 45-degree line, and 2 points are above the 45-degree line.**

Section 102

Assorted Exam-Style Questions for Exam 4/C – Part 18

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C102-1. Similar to Question 154 of the Exam C Sample Questions from the Society of Actuaries. Loss frequency (N) follows a Poisson distribution with mean λ. Loss severity (X) follows a lognormal distribution with parameters μ and σ. The prior distribution of λ, μ, and σ has a joint probability function of f(λ, μ, σ) = 3λσ for each parameter λ, μ, and σ ranging between 0 and 1. Let S denote aggregate losses. Use this information and Bühlmann credibility to find Bühlmann's K with respect to S.

Useful information regarding lognormal distributions:

$E(X) = \exp(\mu + 0.5\sigma^2)$

$E(X^2) = \exp(2\mu + 2\sigma^2)$

Solution S4C102-1. The formula for K is K = EPV/VHM, so we will need to find both EPV and VHM.

$EPV = E_{\lambda, \mu, \sigma}(Var(S \mid \lambda, \mu, \sigma))$.

$VHM = Var_{\lambda, \mu, \sigma}(E(S \mid \lambda, \mu, \sigma))$.

First, we find $E(S \mid \lambda, \mu, \sigma)$. Since N is Poisson,

$E(S \mid \lambda, \mu, \sigma) = \lambda * E(X \mid \mu, \sigma) = \lambda * \exp(\mu + 0.5\sigma^2)$.

Now we find $Var(S \mid \lambda, \mu, \sigma)$. Since N is Poisson,

$Var(S | \lambda, \mu, \sigma) = \lambda*E(X^2 | \mu, \sigma) = \lambda*exp(2\mu + 2\sigma^2)$.

When we find expected value with respect to three random variables, we have no choice but to use a triple integral.

Thus, $E_{\lambda, \mu, \sigma}(Var(S | \lambda, \mu, \sigma)) = {}_0{}^1\int_0{}^1\int_0{}^1\int(\lambda*exp(2\mu + 2\sigma^2))(3\lambda\sigma)d\lambda*d\mu*d\sigma$.

$= {}_0{}^1\int_0{}^1\int_0{}^1\int(\sigma*exp(2\mu + 2\sigma^2))(3\lambda^2)d\lambda*d\mu*d\sigma =$

${}_0{}^1\int_0{}^1\int (\lambda^3 | {}_0{}^1) (\sigma*exp(2\mu + 2\sigma^2))*d\mu*d\sigma =$

${}_0{}^1\int_0{}^1\int(\sigma*exp(2\mu + 2\sigma^2))*d\mu*d\sigma =$

${}_0{}^1\int_0{}^1\int(\sigma*exp(2\mu)*exp(2\sigma^2))*d\mu*d\sigma =$

${}_0{}^1\int(1/2)\sigma*exp(2\sigma^2)(exp(2\mu) | {}_0{}^1)*d\sigma =$

${}_0{}^1\int(1/2)(e^2-1)\sigma*exp(2\sigma^2)*d\sigma =$

$(1/8)(e^2-1)exp(2\sigma^2) | {}_0{}^1 =$

$(1/8)(e^2-1)^2 = EPV = 5.102504729$.

$VHM = Var_{\lambda, \mu, \sigma}(E(S | \lambda, \mu, \sigma)) = Var_{\lambda, \mu, \sigma}(\lambda*exp(\mu + 0.5\sigma^2)) =$

$E_{\lambda, \mu, \sigma}((\lambda*exp(\mu + 0.5\sigma^2))^2) - E_{\lambda, \mu, \sigma}(\lambda*exp(\mu + 0.5\sigma^2))^2$.

It would be best to separate the calculation of this value into two distinct calculations.

First, we find $E_{\lambda, \mu, \sigma}(\lambda*exp(\mu + 0.5\sigma^2)) =$

${}_0{}^1\int_0{}^1\int_0{}^1\int(\lambda*exp(\mu + 0.5\sigma^2))(3\lambda\sigma)d\lambda*d\mu*d\sigma =$

${}_0{}^1\int_0{}^1\int (\lambda^3 | {}_0{}^1)(\sigma*exp(\mu + 0.5\sigma^2))*d\mu*d\sigma =$

${}_0{}^1\int_0{}^1\int(\sigma*exp(\mu)*exp(0.5\sigma^2))*d\mu*d\sigma =$

${}_0{}^1\int(e-1)\sigma*exp(0.5\sigma^2)*d\sigma =$

$(e-1)exp(0.5\sigma^2) | {}_0{}^1 = (e-1)(e^{1/2}-1) = E_{\lambda, \mu, \sigma}(\lambda*exp(\mu + 0.5\sigma^2)) =$

1.114685971. Thus, $E_{\lambda, \mu, \sigma}(\lambda*exp(\mu + 0.5\sigma^2))^2 = 1.242524814$.

Now we find $E_{\lambda, \mu, \sigma}((\lambda*exp(\mu + 0.5\sigma^2))^2) =$

${}_0{}^1\int_0{}^1\int_0{}^1\int(\lambda*exp(\mu + 0.5\sigma^2))^2(3\lambda\sigma)d\lambda*d\mu*d\sigma =$

${}_0{}^1\int{}_0{}^1\int{}_0{}^1\int(\exp(2\mu + \sigma^2)(3\lambda^3\sigma)d\lambda*d\mu*d\sigma =$

${}_0{}^1\int{}_0{}^1\int((3\lambda^4/4)\,|\,{}_0{}^1)\sigma*\exp(2\mu + \sigma^2)*d\mu*d\sigma =$

${}_0{}^1\int{}_0{}^1\int(3/4)\sigma*\exp(2\mu + \sigma^2)*d\mu*d\sigma =$

${}_0{}^1\int((1/2)\exp(2\mu)\,|\,{}_0{}^1)(3/4)\sigma*\exp(\sigma^2)*d\sigma =$

${}_0{}^1\int(3/8)(e^2-1)\sigma*\exp(\sigma^2)*d\sigma =$

$(3/16)(e^2-1)\exp(\sigma^2)\,|\,{}_0{}^1 = (3/16)(e^2-1)(e-1) = E_{\lambda, \mu, \sigma}((\lambda*\exp(\mu + 0.5\sigma^2))^2) = 2.058412312.$

VHM = 2.058412312 - 1.242524814 = 0.8158874974.

Thus, K = EPV/VHM = 5.102504729/0.8158874974 = **K = 6.253931756**.

Problem S4C102-2. Similar to Question 177 of the Exam C Sample Questions from the Society of Actuaries. Losses (N) are Poisson-distributed with mean Λ, where Λ follows a Pareto distribution with parameters α = 5 and θ = 300. There are 8 losses observed. Find the associated Bühlmann credibility factor Z.

Relevant properties of Poisson distributions: E(N) = Var(N) = λ.

Relevant properties of Pareto distributions: $E(X) = \theta/(\alpha-1)$; $E(X^2) = 2\theta^2/((\alpha-1)(\alpha-2))$

Solution S4C102-2. The formula for Z is Z = N/(N + K), where K = EPV/VHM. We are given that 8 losses were observed. This is our value of N.

We find EPV = $E_\Lambda(Var(N\,|\,\Lambda)) = E_\Lambda(\Lambda)$, since N is Poisson-distributed.

$E_\Lambda(\Lambda) = \theta/(\alpha-1) = 300/4 = 75$ = EPV.

VHM = $Var_\Lambda(E(N\,|\,\Lambda)) = Var_\Lambda(\Lambda)$, since N is Poisson-distributed.

$Var_\Lambda(\Lambda) = E(\Lambda^2) - E(\Lambda)^2 = 2*300^2/((5-1)(5-2)) - 75^2 = 15000 - 5625 = 9375$ = VHM.

Thus, K = EPV/VHM = 75/9375 = K = 0.008, and Z = 8/8.008 = **Z = 0.99900999**.

Problem S4C102-3. Similar to Question 181 of the Exam C Sample Questions from the Society of Actuaries. Frequency of losses (N) follows a Poisson distribution with mean Λ. The prior distribution of Λ is exponential with mean 5. Severity of losses follows an exponential distribution with mean Θ. The prior distribution of Θ is Poisson with mean 7. Let S denote aggregate losses. Use this information and Bühlmann credibility to find Bühlmann's K with respect to S.

Relevant properties of Poisson distributions: $E(N) = Var(N) = \lambda$.
Relevant properties of exponential distributions: $E(X) = \theta$; $E(X^2) = 2\theta^2$.

Solution S4C102-3. The formula for K is K = EPV/VHM, so we will need to find both EPV and VHM.

First, we find $E(S \mid \Lambda, \Theta) = E(N)*E(X) = \Lambda*\Theta$.

We also find $Var(S \mid \Lambda, \Theta) = E(N)*E(X^2)$ (since N is Poisson-distributed) $= \Lambda*2\Theta^2$.

$EPV = E_{\Lambda, \Theta}(Var(S \mid \Lambda, \Theta)) = E_{\Lambda, \Theta}(\Lambda*2\Theta^2) = E(\Lambda)*2E(\Theta^2)$.

Since Λ is exponential with mean 5, $E(\Lambda) = 5$. Since Θ is Poisson with mean 7, $E(\Theta^2) = Var(\Theta) + E(\Theta)^2 = 7 + 7^2 = 56$. Thus, EPV = 5*2*56 = EPV = 560.

$VHM = Var_{\Lambda, \Theta}(E(S \mid \Lambda, \Theta)) = Var_{\Lambda, \Theta}(\Lambda*\Theta) = E((\Lambda*\Theta)^2) - E(\Lambda*\Theta)^2 =$

$E(\Lambda^2)*E(\Theta^2) - E(\Lambda)^2*E(\Theta)^2$.

Since Λ is exponential with mean 5, $E(\Lambda^2) = 2*5^2 = 50$, and thus

VHM = 50*56 - 25*49 = VHM = 1575.

Thus, K = 560/1575 = **K = 0.35555555556**.

Problem S4C102-4. Similar to Question 183 of the Exam C Sample Questions from the Society of Actuaries. You are creating a mixture of two independent Poisson distributions, A with mean 6 and B with mean 8. Each distribution occurs in the mixture with probability 0.5. Find the difference between the variance of the mixture and the mean of the mixture.

Solution S4C102-4. The mean of the mixture distribution is the probability-weighted average of the component means or 0.5*6 + 0.5*8 = 7.

The variance of the mixture, however, is the sum of the mean of the variances of the component distributions and the variance of the means of the component distributions. Since the component distributions are Poisson, the mean of the variances is equal to the mean of the means, or 7.

The variance of the means is thus the difference between the variance and the mean of the mixture. This difference is $0.5(6-7)^2 + 0.5(8-7)^2 = 0.5 + 0.5 =$ **1.**

Problem S4C102-5. Similar to Question 185 of the Exam C Sample Questions from the Society of Actuaries. You are given the following data:
10 observations occur at time t = 1.

23 observations occur at time t = 4.

12 observations occur at time t = 7.

19 observations occur at time t = 10.

Using this information, find the 90% linear confidence interval for $\hat{H}(8)$, the Nelson-Åalen estimate of the cumulative hazard rate function at t = 8.

Solution S4C102-5. First, we find the center of the interval, our estimate of $\hat{H}(8)$:

Observations prior to t = 8 occur at times 1, 4, and 7.

At t = 1, the risk set is (10 + 23 + 12 + 19) = 64.

At t = 4, the risk set is (23 + 12 + 19) = 54.

At t = 7, the risk set is (12 + 19) = 31.

The estimate of $\hat{H}(8)$ is the sum of the observations divided by the risk set for each time period above: $\hat{H}(8)$ = 10/64 + 23/54 + 12/31 = 0.9692727001.

The formula for our confidence interval is $\hat{H}(t) \pm z_{1-\alpha/2}*\sqrt{(Var\text{^}(\hat{H}(y_j)))}$, where $Var\text{^}(\hat{H}(y_j)) = {}_{i=1}{}^{j}\Sigma(s_i/r_i^2)$, where s_i is the number of times the ith unique observation occurs and r_i is the size of the risk set associated with the ith unique observation.

We find $Var\text{^}(\hat{H}(8)) = 10/64^2 + 23/54^2 + 12/31^2$ = 0.0228159161.

We find $z_{1-0.1/2} = z_{0.95} = \Phi^{-1}(0.95)$, which we find via the Excel input "=NORMSINV(0.95)", getting as our result 1.644853627. Thus, the bounds of our confidence interval are

0.9692727001 ± 1.644853627*√(0.0228159161), and the interval is

(0.7208185746, 1.217726826).

Section 103

Assorted Exam-Style Questions for Exam 4/C – Part 19

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C103-1. Review of Section 45. Similar to Question 186 of the Exam C Sample Questions from the Society of Actuaries. You know the following information about a random variable X:

$E(X) = \theta$.
$Var(X) = \theta^2/49$.
θ^ is an estimator of X, such that $\theta\text{^} = kX/(k+2)$ for $k > 0$.
$MSE_{\theta\text{^}}(\theta) = 3*(bias_{\theta\text{^}}(\theta))^2$.

Use this information to find k.

Solution S4C103-1. To solve this problem, we need to invoke the following formula:

$MSE_{\theta\text{^}}(\theta) = Var(\theta^{\text{^}} \mid \theta) + (bias_{\theta\text{^}}(\theta))^2$.

Thus, $MSE_{\theta\text{^}}(\theta) = Var(kX/(k+2)) + (k\theta/(k+2) - \theta)^2$.
$MSE_{\theta\text{^}}(\theta) = (k^2/(k+2)^2)*Var(X) + (-2\theta/(k+2))^2$.
$MSE_{\theta\text{^}}(\theta) = k^2\theta^2/(49(k+2)^2) + 4\theta^2/(k+2)^2$.
As $(bias_{\theta\text{^}}(\theta))^2 = 4\theta^2/(k+2)^2$, we know that
$k^2\theta^2/(49(k+2)^2) + 4\theta^2/(k+2)^2 = 3*(bias_{\theta\text{^}}(\theta))^2 = 12\theta^2/(k+2)^2$
$k^2\theta^2/(49(k+2)^2) = 8\theta^2/(k+2)^2$
$k^2\theta^2/49 = 8\theta^2$
$k^2/49 = 8$
$k = \sqrt{(8*49)}$ = **k = 19.79898987**.

Problem S4C103-2. Similar to Question 187 of the Exam C Sample Questions from the Society of Actuaries. The number of claims (C) follows a geometric distribution with parameter β. The prior distribution of β follows a Pareto distribution with θ = 30 and α = 3. During year 1, 20 claims are observed. Find the Bühlmann credibility estimate of the number of claims in year 2.

Relevant properties of geometric distributions: $E(N) = \beta$; $Var(N) = \beta(1+\beta)$.

Relevant properties of Pareto distributions: $E(X) = \theta/(\alpha-1)$; $E(X^2) = 2\theta^2/((\alpha-1)(\alpha-2))$.

Solution S4C103-2. The formula for the Bühlmann credibility factor Z is Z = N/(N + K), where K = EPV/VHM. Here, N = number of years under observation = 1.

We find EPV = $E_\beta(Var(C|\beta))$. $Var(C|\beta) = \beta(1+\beta)$, so

$EPV = E_\beta(\beta(1+\beta)) = E(\beta) + E(\beta^2) = \theta/(\alpha-1) + 2\theta^2/((\alpha-1)(\alpha-2)) = 30/(3-1) + 2*30^2/((3-1)(3-2)) =$ EPV = 915.

$VHM = Var_\beta(E(C|\beta))$. $E(C|\beta) = \beta$, so $VHM = Var_\beta(\beta) = E(\beta^2) - E(\beta)^2 = 2*30^2/((3-1)(3-2)) - (30/(3-1))^2 =$ VHM = 675.

Thus, K = 915/675 = 1.3555555556, and Z = (1/2.3555555556) = 0.4245283019.

Our observed value of claims is 20, whereas our prior mean is

$E(C) = E_\beta(E(C|\beta)) = E_\beta(\beta) = \theta/(\alpha-1) = 15$.

Thus, our credibility estimate is 0.4245283019(20) + (1-0.4245283019)15 = **17.12264151.**

Problem S4C103-3. Similar to Question 190 of the Exam C Sample Questions from the Society of Actuaries. The number of claims (N) is 3 with probability θ and 6 with probability (1- θ). The prior distribution of Θ is such that θ = 0.5 with probability 0.3, and θ = 0.25 with probability 0.7. 5 claims are observed in year 1. Find the Bühlmann credibility estimate of the number of claims in year 2.

Solution S4C103-3. The formula for the Bühlmann credibility factor Z is Z = N/(N + K), where K = EPV/VHM. Here, N = number of years under observation = 1.

First, we find $E(N|\Theta) = 3\theta + 6(1-\theta) = E(N|\Theta) = 6 - 3\theta$.
$E(N^2|\Theta) = 3^2\theta + 6^2(1-\theta) = 36 - 27\theta$.
$Var(N|\Theta) = E(N^2|\Theta) - E(N|\Theta)^2 = 36 - 27\theta - (6 - 3\theta)^2 = 36 - 27\theta - (36 - 36\theta + 9\theta^2) =$
$- 27\theta + 36\theta - 9\theta^2 = -9\theta^2 + 9\theta$.

We also find $E(\Theta) = 0.5*0.3 + 0.25*0.7 = 0.325$ and

$E(\Theta^2) = 0.5^2*0.3 + 0.25^2*0.7 = 0.11875$. Thus, $Var(\Theta) = E(\Theta^2) - E(\Theta)^2 = 0.11875 - 0.325^2 =$ $Var(\Theta) = 0.013125$.

We find EPV = $E_\Theta(Var(N \mid \Theta)) = E_\Theta(-9\theta^2 + 9\theta) = -9E(\Theta^2) + 9E(\Theta) =$

$-9*0.11875 + 9*0.325 =$ EPV = 1.85625.

We find VHM = $Var_\Theta(E(N \mid \Theta)) = Var_\Theta(6 - 3\theta) = Var_\Theta(-3\theta) = 9Var(\Theta) = 9*0.013125 =$ VHM = 0.118125. Thus, K = 1.85625/0.118125 = 15.71428571, and Z = 1/16.71428571 = Z = 0.0598290598. Our observed value of claims is 5, whereas our prior mean is $E_\Theta(E(N \mid \Theta)) =$

$E(6 - 3\theta) = 6 - 3E(\Theta) = 6 - 3*0.325 = 5.025$.

Our answer is thus 5*0.0598290598 + 5.025(1-0.0598290598) = **5.023504274.**

Problem S4C103-4. Review of Section 53. Similar to Question 192 of the Exam C Sample Questions from the Society of Actuaries. You use the following kernel function: $k_y(x) = (2/\pi)\sqrt{(1- (x-y)^2)}$, for $y-1 \leq x \leq y+1$ and 0 otherwise. You are also given a sample of values of x: 4, 6, 6, 6, 8, 10. The kernel function can be used to estimate the probability function f(x).

Draw the shape of this estimated function using axes like the ones below.

f^(x)
|
(8/π)
|
(6/π)
|
(4/π)
|
(2/π)
|
0
|__________3___4___5__6___7___8___9___10___11________ x

Solution S4C103-4. The formula for the kernel density estimator is $\hat{f}(x) = {}_{j=1}\Sigma^{k}(p(y_j)*k_{y\_j}(x))$.

We note that for values of x and y that are farther apart than 1, $k_y(x) = 0$.

For values of x and y that are equal, $k_y(x) = (2/\pi)\sqrt{(1- (x-x)^2)} = (2/\pi)\sqrt{(1-0)} = (2/\pi)$.

All of the given sample values are either equal to one another or more than one unit apart. Thus, the value of f^(x) at any x in the sample is equal to (2/π) multiplied by the number of occurrences of x. This means that f^(4) = 1*(2/π) = (2/π); f^(6) = 3*(2/π) = 6/π; f^(8) = 2*(2/π) = 8/π;

f^(10) = 1*(2/π) = (2/π). What happens in between these points? If x and y are 1 apart. Since for values of x and y that are farther apart than 1, $k_y(x) = 0$, this means that the sample values result in local maxima of f^(x), and f^(x) is 0 one unit to either direction of these local maxima. What is the shape of the curve between the local maxima and the zeroes? The kernel function is based on a square root, which has a smooth, concave-down function. The resulting graph is (very roughly) as follows:

```
f^(x)
|
(8/π)
|
(6/π)..........................................*
|..............................................*...*
(4/π)......................................*.......*...........*
|..........................................*..........*.......*......*
(2/π).........................*..........*............*.....*.........*..........*...
|...........................*..........*..*..............*...*...........*...*.............*
0........................*...............*................*................*.....................*
|___________3___4___5___6___7___8___9___10___11________ x
```

Problem S4C103-5. Similar to Question 193 of the Exam C Sample Questions from the Society of Actuaries. You have the following sample from a Pareto distribution of random variable X:

4, 6, 7, 10, 15, 48.

You use the method of moments to estimate the first and second moments of the distributions. Use the results of this procedure to find the limited expected value of X at 12.

Relevant properties of Pareto distributions: $E(X) = \theta/(\alpha-1)$; $E(X^2) = 2\theta^2/((\alpha-1)(\alpha-2))$;
$E(X \wedge k) = (\theta/(\alpha-1))(1 - (\theta/(x+\theta))^{\alpha-1})$

Solution S4C103-5. We estimate E(X) via the sample mean, or (4+6+7+10+15+48)/6 = 15.

We estimate $E(X^2)$ via the mean of sample squares, or $(4^2+6^2+7^2+10^2+15^2+48^2)/6 = 455$.

Thus, we set up the following system of equations:
i) $15 = \theta/(\alpha-1)$;
ii) $455 = 2\theta^2/((\alpha-1)(\alpha-2)) \rightarrow$ ii') $227.5 = \theta^2/((\alpha-1)(\alpha-2))$.

We divide ii') by i)^2 to get $1.0111111 = (\alpha-1)^2/((\alpha-1)(\alpha-2)) = (\alpha-1)/(\alpha-2)$.

Thus, $1.01111111(\alpha-2) = (\alpha-1) \rightarrow 1.01111111\alpha - 2.02222222 = \alpha-1 \rightarrow$
$0.011111111111\alpha = 1.02222222222 \rightarrow \alpha = 92$.
Thus, $E(X) = 15 = \theta/91$, and so $\theta = 91*15 = 1365$.
Our estimate of $E(X \wedge 12)$ is thus $15*(1 - (1365/1377)^{91})$ = **$E(X \wedge 12) = 8.23646143$**.

Section 104

Assorted Exam-Style Questions for Exam 4/C – Part 20

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C104-1. For a Pareto distribution, find the expression for $f(k)/(1-F(d))$, where $k > d$, in terms of parameters α and θ.

Relevant properties of Pareto distributions: $f(x) = \alpha\theta^{\alpha}/(x+\theta)^{\alpha+1}$; $S(x) = \theta^{\alpha}/(x+\theta)^{\alpha}$.

Solution S4C104-1. $f(k)/(1-F(d)) = f(k)/S(d) = (\alpha\theta^{\alpha}/(k+\theta)^{\alpha+1})/(\theta^{\alpha}/(d+\theta)^{\alpha}) = f(k)/(1-F(d)) =$ $\boldsymbol{\alpha(d+\theta)^{\alpha}/(k+\theta)^{\alpha+1}}$.

Problem S4C104-2. Similar to Question 196 of the Exam C Sample Questions from the Society of Actuaries. You observe the following three losses:
A loss of 600 occurred on a policy with a deductible of 300. The loss amount is 600 before the deductible was applied.

A loss of 500 occurred on a policy with a limit of 1000.

A loss in excess of 2000 occurred on a policy with a limit of 2000.
You know that losses originate from a Pareto distribution with parameters $\theta = 3000$ and α unknown. Find the maximum likelihood estimate of α.

Relevant properties of Pareto distributions: $f(x) = \alpha\theta^{\alpha}/(x+\theta)^{\alpha+1}$; $S(x) = \theta^{\alpha}/(x+\theta)^{\alpha}$.

Solution S4C104-2. We first find the likelihood function $L(\alpha)$. For the loss of 600 with a deductible of 300, the contributing factor to the likelihood function will be $f(600)/(1-F(300))$, which, according to Solution S4C104-1, is equal to $\alpha(300+3000)^{\alpha}/(600+3000)^{\alpha+1} =$

$\alpha(3300^\alpha/3600^{\alpha+1})$. For the loss of 500, the loss is below the policy limit, so the limit needs not be considered. The contributing factor to the likelihood function is thus $f(500) = \alpha*3000^\alpha/(500+3000)^{\alpha+1} = \alpha(3000^\alpha/3500^{\alpha+1})$.

For the loss in excess of the policy limit of 2000, the contributing factor is $S(2000) = 3000^\alpha/(2000+3000)^\alpha = 0.6^\alpha$.

Thus, the likelihood function is $L(\alpha) = (\alpha(3300^\alpha/3600^{\alpha+1}))(\alpha(3000^\alpha/3500^{\alpha+1}))*0.6^\alpha$.

$L(\alpha) = \alpha^2*5940000^\alpha/(12600000^\alpha(12600000)) = \alpha^2*(33/70)^\alpha/12600000$.

We find the loglikelihood function $l(\alpha) = \ln(L(\alpha)) = \ln(\alpha^2*(33/70)^\alpha/12600000) =$

$2*\ln(\alpha) + \alpha*\ln(33/70) - \ln(12600000)$.

Thus, $l'(\alpha) = 2/\alpha + \ln(33/70) = 0$ at the maximum.

Thus, $2/\alpha = -\ln(33/70)$, and $\alpha = 2/(-\ln(33/70)) =$ **α = 2.659618038**.

Problem S4C104-3. Similar to Question 199 of the Exam C Sample Questions from the Society of Actuaries. Losses follow a Weibull distribution with the cumulative distribution $F(x) = 1 - \exp(-(x/\theta)^{0.5})$. You observe the following losses:
2 losses of 40, 1 loss of 60, and 3 losses known to be in excess of the policy limit of 100.

Use this information to find the maximum likelihood estimate of θ.

Solution S4C104-3. First, we find f(x) for this distribution.

$f(x) = F'(x) = -\exp(-(x/\theta)^{0.5})*(-(x/\theta)^{0.5})' = -\exp(-(x/\theta)^{0.5})*(-0.5x^{-0.5}*\theta^{-0.5}) =$

$(0.5x^{-0.5}*\theta^{-0.5})*\exp(-(x/\theta)^{0.5})$. Now we can find $L(\theta)$.

The 2 losses of 40 each contribute a factor of $f(40) = (0.5*40^{-0.5}*\theta^{-0.5})*\exp(-(40/\theta)^{0.5})$.

The 1 loss of 60 contributes a factor of $f(60) = (0.5*60^{-0.5}*\theta^{-0.5})*\exp(-(60/\theta)^{0.5})$.

The 3 losses in excess of the policy limit of 100 each contribute a factor of $S(100) = 1-F(100) = \exp(-(100/\theta)^{0.5})$.

Thus, $L(\theta) = (0.5*40^{-0.5}*\theta^{-0.5})*\exp(-(40/\theta)^{0.5})*(0.5*40^{-0.5}*\theta^{-0.5})*\exp(-(40/\theta)^{0.5})*(0.5*60^{-0.5}*\theta^{-0.5})*\exp(-(60/\theta)^{0.5})*\exp(-(100/\theta)^{0.5})*\exp(-(100/\theta)^{0.5})*\exp(-(100/\theta)^{0.5})$.

To simplify this expression, it is best to work with one aspect at a time.

There are 3 factors of 0.5, which simplifies to a factor of $0.5^3 = 0.125$.

We can also simplify ($40^{-0.5}*40^{-0.5}*60^{-0.5}$) to $96000^{-0.5}$ = 309.83866677.

The three factors of $\theta^{-0.5}$ multiply to $\theta^{-1.5}$.

Finally, we have the exponentiated expression. The negative sign does not change in the course of the multiplication, nor does the denominator of $\theta^{0.5}$. The numerator (absent consideration of the negative sign) becomes $2*40^{0.5} + 60^{0.5} + 3*100^{0.5}$ = 50.39507733.

Thus, $L(\theta) = 0.125*309.83866677*\theta^{-1.5}*\exp(-50.39507733/\theta^{0.5})$;

$L(\theta) = 38.72983346*\theta^{-1.5}*\exp(-50.39507733/\theta^{0.5})$.

Now we can find the loglikelihood function $l(\theta) = \ln(L(\theta))$ =

$\ln(38.72983346*\theta^{-1.5}*\exp(-50.39507733/\theta^{0.5}))$ =

$l(\theta) = \ln(38.72983346) + -1.5*\ln(\theta) - 50.39507733/\theta^{0.5}$.

$l'(\theta) = -1.5/\theta + 0.5*50.39507733/\theta^{1.5} = 0$ at the maximum.
Thus, $0.5*50.39507733/\theta^{1.5} = 1.5/\theta$ and $\theta^{0.5} = 0.5*50.39507733/1.5 = 16.79835911$, implying that $\theta = 16.79835911^2$ = **θ = 282.1848688**.

Problem S4C104-4. Similar to Question 200 of the Exam C Sample Questions from the Society of Actuaries. You are using both Bühlmann credibility and Bayesian estimation to estimate next year's data based on observed data from this year.

You know the following data:

The expected number of observations in a year for the risks being observed ranges from 3 to 7.

The number of observations follows a Poisson distribution for each risk.

Which of the following situations are possible? More than one answer may be correct.

(a) A Bayesian estimate can be higher than 7.

(b) A Bayesian estimate can be lower than 3.

(c) The Bayesian estimates are consistently higher than the Bühlmann credibility estimates.

(d) The Bayesian estimates are consistently lower than the Bühlmann credibility estimates.

(e) The Bayesian estimates are linear, but the Bühlmann credibility estimates are not.

(f) The Bühlmann credibility estimates are linear, but the Bayesian estimates are not.

Solution S4C104-4. A Bayesian estimate can never produce estimates outside the range of the prior distribution, so both (a) and (b) are incorrect. Bühlmann credibility provides the best linear least-squares approximation to the Bayesian estimates, and this cannot be the case if it is either consistently higher or consistently lower than the Bayesian estimates, so (c) and (d) are incorrect. Also, Bühlmann credibility estimates are always linear, because the observed mean X^- varies linearly in the expression $Z*X^- + (1-Z)*M$. Thus, (e) is incorrect, but (f) could be correct, as the Bayesian estimates can vary nonlinearly around the Bühlmann credibility estimates, provided that they are not consistently higher or lower and that Bühlmann credibility provides the best linear least-squares approximation to the Bayesian estimates. Thus, **only (f) is correct.**

Problem S4C104-5. Similar to Question 203 of the Exam C Sample Questions from the Society of Actuaries. Observations have probability 0.6 of coming from geometric distribution A with β = 3 and probability 0.4 of coming from a geometric distribution B with β = 8. An observation of 5 is observed. Use Bayesian analysis to find the posterior expected value of an observation from the same distribution.

Relevant properties of geometric distributions: E(N) = β; $Pr(N = k) = \beta^k/(1+\beta)^{k+1}$.

Solution S4C104-5. We first want to find each of the posterior probabilities of an observation coming from one of the two distributions.

We find Pr(k = 5) = 0.4*(Pr(β = 3 and k = 5)) + 0.6*(Pr(β = 8 and k = 5)) =

$0.4*3^5/4^6 + 0.6*8^5/9^6 = 0.0237304688 + 0.0369952638 = 0.0607257325$.

Now, we find Pr(A│5) = Pr(A and 5)/Pr(5) = 0.0237304688/0.0607257325 = 0.3907811029.

We find Pr(B│5) = Pr(B and 5)/Pr(5) = 0.0369952638/0.0607257325 = 0.6092188971. These are our desired posterior probabilities.
Now, the posterior expected value is E(A)*0.3907811029 + E(B)*0.6092188971 = 3*0.3907811029 + 8*0.6092188971 = **6.046094485.**

Section 105

Assorted Exam-Style Questions for Exam 4/C – Part 21

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C105-1. Similar to Question 204 of the Exam C Sample Questions from the Society of Actuaries. The random variable X is exponentially distributed with mean 1/Y. The random variable Y follows a gamma distribution with parameters $\alpha = \theta = 5$. Find $\Pr(X < 1/3)$.

Relevant properties of gamma distributions: $f(x) = (x/\theta)^{\alpha}*(e^{-x/\theta})/(x*\Gamma(\alpha))$.

Solution S4C105-1. The conditional cumulative distribution of X is $F(x \mid Y = y) = e^{-xy}$. We are in particular interested in $F(1/3 \mid Y = y) = 1-e^{-y/3}$. $F(1/3) = {}_0{}^{\infty}\int F(1/3 \mid Y = y)*f_Y(y)*dy =$ ${}_0{}^{\infty}\int e^{-y/3}*f_Y(y)*dy$. We find $f_Y(y) = (y/5)^5*(e^{-y/5})/(y*\Gamma(5)) = (y^4/3125)*(e^{-y/5})/4 =$ $y^4*(e^{-y/5})/12500$. Thus, our integral is ${}_0{}^{\infty}\int(1-e^{-y/3})*y^4*(e^{-y/5})/12500)dy =$ ${}_0{}^{\infty}\int(y^4*(e^{-y/5})/12500)dy - {}_0{}^{\infty}\int((e^{-8y/15}*y^4)/12500)dy$.

The former integral is the integral of the gamma pdf over all of its possible values, and so has to equal 1.

To solve the latter integral, we use the Tabular Method of integration by parts:

Sign...........u...............dv
+............. y^4.............$(1/12500)e^{-8y/15}$
-...............$4y^3$...........$-(3/20000)e^{-8y/15}$
+.............$12y^2$...........$(9/32000)e^{-8y/15}$
-...............$24y$.............$-(27/51200)e^{-8y/15}$
+...............24..............$(81/81920)e^{-8y/15}$
-.................0..............$-(243/131072)e^{-8y/15}$

Thus, the integral is

$(-y^4(3/20000)e^{-8y/15} - y^3(36/32000)e^{-8y/15} - y^2(324/51200)e^{-8y/15} - y(1944/81920)e^{-8y/15} - (5832/131072)e^{-8y/15}) \Big|_0^{\infty} = 1 - F(1/3) = 5832/131072 = 0.0444946289.$

Thus, F(1/3) = 1 - 0.0444946289 = **F(1/3) = 0.9555053711**.

Problem S4C105-2. Similar to Question 205 of the Exam C Sample Questions from the Society of Actuaries.

There are 340 elephants in group A. The probability that an elephant in group A is hungry and will want to eat some peanuts is 0.4. If an elephant in group A eats some peanuts, the number of peanuts eaten follows a geometric distribution with mean 5.

There are 500 elephants in group B. The probability that an elephant in group B is hungry and will want to eat some peanuts is 0.1. If an elephant in group B eats some peanuts, the number of peanuts eaten follows a geometric distribution with mean 10.

Use a normal approximation to find the probability that the number of peanuts eaten by all elephants exceeds 1200. Refer to the Exam 4 / C Tables, as necessary. You do not need to use a continuity correction.

Solution S4C105-2. Let N be the number of occurrences of elephants eating peanuts, and let X be the amount of peanuts eaten per such occurrence. Let S be aggregate peanuts eaten.

For group A,

E(N) = 340*0.4 = 136.

We note that N is binomially distributed with m = 340 and q = 0.4, as elephants can either eat peanuts or not eat them.

Thus, for group A, Var(N) = 340*0.4*(1-0.4) = 81.6.

For group A, E(X) = β for the geometric distribution = 5; Var(X) = β(1+β) = 5*6 = 30.

Thus, for group A, E(S) = E(N)*E(X) = 136*5 = 680, and

$\text{Var}(S) = E(N)*\text{Var}(X) + \text{Var}(N)*E(X)^2 = 136*30 + 81.6*5^2 = \text{Var}(S) = 6120.$

For group B,

E(N) = 500*0.1 = 50.

We note that N is binomially distributed with m = 500 and q = 0.1, as elephants can either eat peanuts or not eat them.

Thus, for group A, Var(N) = 500*0.1*(1-0.1) = 45.

For group A, E(X) = β for the geometric distribution = 10; Var(X) = β(1+β) = 10*11 = 110.

Thus, for group A, E(S) = E(N)*E(X) = 50*10 = 500, and

Var(S) = E(N)*Var(X) + Var(N)*E(X)^2 = 50*110 + 45*10^2 = Var(S) = 10000.

For the two groups, we add the values of E(S) from each group to get E(S) = 680 + 500 = 1180.

We perform an analogous procedure to get Var(S) = 6120 + 10000 = 16120, implying that SD(S) = 126.964562. We want to find Pr(S > 1200) ≈ 1 - Φ((1200-1180)/126.964562) = 1 - Φ(0.1575242705), which we find via the Excel input "=1-NORMSDIST(0.1575242705)", getting as our answer **0.437415841**.

Problem S4C105-3. Similar to Question 206 of the Exam C Sample Questions from the Society of Actuaries. Frequency of losses (N) follows a geometric distribution with mean 6. Severity of each loss (X) follows a distribution with the following probabilities:
Pr(X = 4) = 0.6;
Pr(X = 8) = 0.3;
Pr(X = 12) = 0.1.

Let S denote aggregate losses.

Find E((S-12)$_+$), the expected value of the left-censored and shifted random variable at 12.

Relevant properties of geometric distributions: E(N) = β; Pr(N = k) = $\beta^k/(1+\beta)^{k+1}$.

Solution S4C105-3. We note that S can only occur in intervals of 4. We compare possible values of three random variables:
S:..........0, 4, 8, 12, 16, 20, 24, ...
(S-12)$_+$: 0, 0, 0, 0, 4, 8, 12,....
S-12: -12, -8, -4, 0, 4, 8, 12, ...

We note that to turn E(S-12) into E((S-12)$_+$), we need to add 12*f_S(0) + 8*f_S(4) + 4*f_S(8).

Thus, E((S-12)$_+$) = E(S) - 12 + 12*f_S(0) + 8*f_S(4) + 4*f_S(8).

First, we find E(X) = 4*0.6 + 8*0.3 + 12*0.1 = 6.

E(S) = E(N)*E(X) = 6*6 = 36.

Now we find f_S(0) = Pr(N = 0) = 1/(1+6) = 1/7.

We find f_S(4) = Pr(N = 1)*Pr(X = 4) = (6/7^2)(0.6) = 18/245.

We find $f_S(8)$ = Pr(N = 1)*Pr(X = 8) + Pr(N = 2)*Pr(Both X are 4) =

$(6/7^2)(0.3) + (6^2/7^3)(0.6^2) = f_S(8) = 639/8575$.

Thus, $E((S-12)_+)$ = 36 - 12 + 12*(1/7) + 8*(18/245) + 4*(639/8575) = **$E((S-12)_+)$ = 26.60011662**.

Problem S4C105-4. Similar to Question 207 of the Exam C Sample Questions from the Society of Actuaries. The size of each loss (X) follows a distribution with probability density function f(x) = 0.05x, for 0 < x < √(40). A deductible of 2 is applied to each loss. Let Y^P be the per-payment loss payment random variable. Find $E(Y^P)$.

Solution S4C105-4. We first find the pdf of X, given that X exceeds the deductible of 2. This is f(x)/(1-F(2)), where $F(2) = \int_0^2 0.05x*dx = 0.025*2^2 = 0.1$. Thus, our pdf of X, given that X exceeds the deductible of 2, is 0.05x/(1-0.1) = x/18. We find the expected value of X, given that X exceeds 2: $\int_2^{\sqrt{(40)}}(x)(x/18)*dx = (x^3/54)\Big|_2^{\sqrt{(40)}} = (40^{3/2}/54) - 8/54 = 4.536707645$. To find $E(Y^P)$, we subtract the deductible of 2 from this value and get **$E(Y^P)$ = 2.536707645**.

Problem S4C105-5. Similar to Question 212 of the Exam C Sample Questions from the Society of Actuaries. Frequency of losses (N) follows a Poisson distribution with mean 30. Severity of each loss (X) follows a uniform distribution on the interval from 0 to 50. A deductible of 10 is applied to each loss. Find the variance of aggregate payments after the deductible has been applied.

Solution S4C105-5. The distribution of payment severity is different from the distribution of loss severity; it is still uniform, but it is uniform from 0 to (50-10) = 40, corresponding to the distribution of losses from 10 to 50. The pdf of this uniform distribution is 1/40. Since (1/5) of losses now do not result in payment, payment frequency is (4/5) of loss frequency, and so expected number of payments is 30*4/5 = 24.

Let R denote number of payments, and let Q denote payment severity. Let S denote aggregate payments. Since R is Poisson, $Var(S) = E(R)*E(Q^2)$. We find $E(Q^2) = \int_0^{40}(q^2/40)dq =$

$(q^3/120)\Big|_0^{40} = 533.3333333$.

Our answer is thus 24*533.33333333 = **Var(S) = 12800**.

Section 106

Assorted Exam-Style Questions for Exam 4/C – Part 22

This section provides additional exam-style practice with a variety of syllabus topics.

A **compound frequency model** is characterized by a random variable S whose probability generating function can be represented as $P_S(z) = P_N(P_M(z))$, where N is called the **primary distribution**, and M is called the **secondary distribution.**

Let the value g_k be the probability that S = k.

For a Poisson distribution, the value g_0 is defined as $g_0 = \exp(-\lambda(1-f_0))$, where λ is the mean of the distribution, and is the probability of a zero value of the random variable which is Poisson-distributed.

The following theorem may be of use in some exam questions:

Theorem 106.1. "For any compound distribution, $g_0 = P_N(f_0)$, where $P_N(z)$ is the probability generating function of the primary distribution and f_0 is the probability that the secondary distribution takes on the value zero."

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Sources: Exam C Sample Questions and Solutions from the Society of Actuaries.

Loss Models: From Data to Decisions, (Third Edition), 2008, by Klugman, S.A., Panjer, H.H. and Willmot, G.E., Chapter 6, pp. 126-129.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C106-1. Similar to Question 208 of the Exam C Sample Questions from the Society of Actuaries. You are given the following compound frequency model:
The primary distribution is negative binomial and has the following probability generating function: $P_N(z) = (1 - 4(z-1))^{-5}$. The secondary distribution is Poisson with mean λ and has the following probability generating function: $P_M(z) = \exp(\lambda(z-1))$. You know that the probability that the compound random variable S = 0 is 0.588.

Use this information to find the value of λ.

Solution S4C106-1. We apply the equation from Theorem 106.1: $g_0 = P_N(f_0)$. We are given that $g_0 = 0.588$. Because M is Poisson-distributed, $f_0 = \Pr(M = 0) = e^{-\lambda}$.

Thus, $0.588 = (1 - 4(e^{-\lambda} -1))^{-5} \to 1.112050564 = 1 - 4(e^{-\lambda} -1) \to$

$4(e^{-\lambda} -1) = -0.112050564 \to -0.0280126411 = (e^{-\lambda} -1) \to e^{-\lambda} = 0.9719873589 \to$

$\lambda = -\ln(0.9719873589) = \boldsymbol{\lambda = 0.0284124798}$.

Problem S4C106-2. Similar to Question 209 of the Exam C Sample Questions from the Society of Actuaries. Losses in the year 4300 follow a Pareto distribution with parameters $\alpha = 2$ and $\theta = 340$. Every year, losses increase by 10%.

An insurance company pays all losses above a deductible of 100. The deductible always remains the same.

The premium charged by an insurance company, denoted P_i for year i, is 1.05 times the expected loss payment by the company.

The company reinsures loss amounts above a certain deductible that always remains. The reinsurance premium is equal to 1.1 times the expected reinsurance payment. Let this reinsurance premium be denoted as R_i for year i.

It is known that $R_{4300}/P_{4300} = 0.46$. Find R_{4301}/P_{4301}.

Relevant properties of Pareto distributions: $E(X) = \theta/(\alpha-1)$;

$E(X \wedge k) = (\theta/(\alpha-1))(1 - (\theta/(k+\theta))^{\alpha-1})$.

Solution S4C106-2. First, we note a useful property of Pareto distributions:
$E(X) - E(X \wedge k) = \theta/(\alpha-1) - (\theta/(\alpha-1))(1 - (\theta/(k+\theta))^{\alpha-1}) = (\theta/(\alpha-1))((\theta/(k+\theta))^{\alpha-1}) =$

$E(X)*((\theta/(k+\theta))^{\alpha-1})$. Using this formula will greatly increase the speed at which this problem can be solved. We also note that for $\alpha = 2$, $E(X) - E(X \wedge k) = E(X)*(\theta/(k+\theta))$.

The Pareto distribution scale parameter is θ, so in year 4301, as losses increase by 10%, θ will become 340*1.1 = 374.

We first find expected losses in the year 4300: $(\theta/(\alpha-1)) = 340/(2-1) = 340$.

We find expected claim payments in the year 4300, which are equal to

$E(X) - E(X \wedge 100) = 340(340/440) = 262.72727272$.

P_{4300} = 1.05*262.72727272 = P_{4300} = 275.8636364.

As R_{4300}/P_{4300} = 0.46, it follows that R_{4300} = 0.46*P_{4300} = 126.8972727.

This means that the expected reinsurance payment in 4300 is

126.8972727/1.1 = 115.361157.

Let d be the reinsurance deductible.

Then 115.361157 = E(X) - E(X Λ d) = E(X)*(θ/(d+θ)) = 340*(340/(d+340)) →

d+340 = 340^2/115.361157 = 1002.070393 → 1002.070393 - 340 = d = 662.0703934.

For year 4301, expected loss payments are E(X) - E(X Λ 100) = 374*(374/474) = 295.0970464.

Thus, P_{4301} = 1.05*295.0970464 = P_{4301} = 309.8518987.

Expected reinsurance payment is 374*(374/(374+662.0703934)) = 135.0062707.

Thus, R_{4301} = 1.1*135.0062707 = 148.5068978.

Therefore, R_{4301}/P_{4301} = 148.5068978/309.8518987 = **0.4792834848.**

Problem S4C106-3. Similar to Question 210 of the Exam C Sample Questions from the Society of Actuaries. A company insures 10 people against cows falling from the sky. The number of cows that fall from the sky follows a Poisson distribution with mean λ = 0.03. 1/3 of the cows that fall from the sky are of type A, and the rest are of type B.

When a cow of type A falls from the sky, the damage done to an individual's house has mean 50 and standard deviation 100.

When a cow of type B falls from the sky, the damage done to an individual's house has mean 30 and standard deviation 150.

The insurance company charges a premium such that the probability (using a normal approximation) that aggregate losses will exceed the premium is 0.15.

A new house shield is developed that prevents only cows of type A from damaging the house. The cost of the house shield per user is 2.

Let Q be the aggregate premium assuming that no one obtains the new house shield.

Let R be the aggregate premium assuming that everyone obtains the new house shield and the cost of the house shield is a covered loss.

Find Q/R.

Solution S4C106-3. First, we find the z value of the standard normal distribution associated with the aggregate premium level. We are given that $0.15 = 1 - \Phi(z) \rightarrow z = \Phi^{-1}(0.85)$, which can find via the Excel input "=NORMSINV(0.85)". The result is z = 1.036433389.

Because loss frequency is Poisson-distributed and 1/3 of the cows are of type A, loss frequency solely due to type A cows is also Poisson-distributed with $\lambda = (1/3)*0.03 = 0.01$.

By implication, loss frequency solely due to type B cows is Poisson-distributed with $\lambda = 0.02$.

Let N_A, X_A, and S_A, respectively, denote loss frequency, loss severity, and aggregate losses from cows of type A.

Let N_B, X_B, and S_B, respectively, denote loss frequency, loss severity, and aggregate losses from cows of type B.

We find $E(S_A) = (\text{Number of insureds})*E(N_A)*E(X_A) = 10*0.01*50 = E(S_A) = 5$.

Since N is Poisson-distributed, we find $Var(S_A) = (\text{Number of insureds})*E(N_A)*E(X_A)^2 = 10*0.01*(50^2 + 100^2) = Var(S_A) = 1250$.

We find $E(S_B) = (\text{Number of insureds})*E(N_B)*E(X_B) = 10*0.02*30 = E(S_B) = 6$.

Since N is Poisson-distributed, we find $Var(S_B) = (\text{Number of insureds})*E(N_B)*E(X_B)^2 =$

$10*0.02*(30^2 + 150^2) = Var(S_B) = 4680$.

If nobody gets the house shield, then we can let S denote the aggregate loss amount.

$E(S) = E(S_A) + E(S_B) = 5 + 6 = 11$.

$Var(S) = Var(S_A) + Var(S_B) = 1250 + 4680 = 5930$.

The aggregate premium is then $Q = E(S) + z*\sqrt{(Var(S))} = 11 + 1.036433389*\sqrt{(5930)} = Q = 90.81210076$.

If everybody gets the house shield, then $E(S) = E(S_B) = 6$, $Var(S) = Var(S_B) = 4680$, but there is added to the premium the total cost of all the house shields, which is 10*2 = 20.

Thus, $R = 20 + E(S) + z*\sqrt{(Var(S))} = 20 + 6 + 1.036433389*\sqrt{(4680)} = R = 96.90295279$.
Thus, Q/R = 90.81210076/96.90295279 = **Q/R = 0.9371448252**.

Problem S4C106-4. Similar to Question 213 of the Exam C Sample Questions from the Society of Actuaries. Frequency of claims (N) follows a discrete distribution with the following properties:

Pr(N = 0) = 0.3; Pr(N = 1) = 0.3; Pr(N = 2) = 0.2; Pr(N = 3) = 0.2.

Severity of claims (X) follows a Poisson distribution with λ = 300. Find Var(S), the variance of aggregate claims (S).

Solution S4C106-4. We first find E(N) = 0*0.3 + 1*0.3 + 2*0.2 + 3*0.2 = 1.3.

We find Var(N) = $0.3*(0-1.3)^2 + 0.3*(1-1.3)^2 + 0.2*(2-1.3)^2 + 0.2*(3-1.3)^2$ = 1.21.

Since X is Poisson-distributed, E(X) = Var(X) = 300.

Thus, Var(S) = $E(N)*Var(X) + Var(N)*E(X)^2 = 1.3*300 + 1.21*300^2$ = **Var(S) = 109290**.

Problem S4C106-5. Similar to Question 214 of the Exam C Sample Questions from the Society of Actuaries. You have the following sample of values:
35, 36, 86, 86, 96, 97, 100, 100, 100, 230, 230, 450, 523, 849.

Determine the empirical estimate of H(100), the cumulative hazard rate function at 100.

Important: The empirical estimate is *not* the same as the Nelson-Åalen estimate!

Solution S4C106-5. We use the formula H(x) = -ln(S(100)). There are 14 total sample values, of which 5 are greater than 100. Thus, the empirical estimate of S(100) is 5/14, and so the estimate of H(100) is -ln(5/14) = **1.029619417.**

Section 107

Assorted Exam-Style Questions for Exam 4/C – Part 23

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C107-1. Similar to Question 215 of the Exam C Sample Questions from the Society of Actuaries. Random variable N - denoting the number of claims - follows a conditional Poisson distribution with mean λ. The prior distribution of λ is gamma, with parameters α and θ. Using Bühlmann credibility, it is estimated that policyholders with an average of 3 claims per year over 3 years will have a 2.5 claims in year 4. It is also estimated that policyholders with an average of 4 claims per year over 5 years will have 3 claims in year 6. Use this information to find θ.

Relevant properties of gamma distributions: $E(X) = \alpha\theta$; $Var(X) = \alpha\theta^2$.

Solution S4C107-1. The Bühlmann credibility factor is N/(N + K), where N is the number of time periods, and K = EPV/VHM.

We first find EPV = $E_\lambda(Var(N | \lambda)) = E_\lambda(\lambda)$, since $(N | \lambda)$ is Poisson-distributed.

Since λ is gamma-distributed, $E_\lambda(\lambda) = EPV = \alpha\theta$.

We find VHM = $Var_\lambda(E(N | \lambda)) = Var_\lambda(\lambda)$, since $(N | \lambda)$ is Poisson-distributed.

Since λ is gamma-distributed, $Var_\lambda(\lambda) = VHM = \alpha\theta^2$.

This means that K = EPV/VHM = $(\alpha\theta)/(\alpha\theta^2) = 1/\theta$.

We also want to find the external mean that is multiplied by the complement of credibility:

$M = E_\lambda(E(N|\lambda)) = E_\lambda(\lambda)$, since $(N|\lambda)$ is Poisson-distributed.

Since λ is gamma-distributed, $E_\lambda(\lambda) = M = \alpha\theta$.

Thus, the Bühlmann credibility estimate for number of years N and observed mean X^- is

$ZX^- + (1\text{-}Z)M =$
$NX^-/(N + 1/\theta) + (1/\theta)(\alpha\theta)/(N + 1/\theta) =$

$NX^-/(N + 1/\theta) + \alpha/(N + 1/\theta)$.

For N = 3 and $X^- = 3$, we are given that the following is the case:
$2.5 = 3*3/(3 + 1/\theta) + \alpha/(3 + 1/\theta) \rightarrow$

i) $2.5 = (9 + \alpha)/(3 + 1/\theta)$.

For N = 5 and $X^- = 4$, we are given that the following is the case:
$3 = 5*4/(5 + 1/\theta) + \alpha/(5 + 1/\theta) \rightarrow$
ii) $3 = (20 + \alpha)/(5 + 1/\theta) \rightarrow$
$3(5 + 1/\theta) = 20 + \alpha \rightarrow$
$15 + 3/\theta - 20 = \alpha \rightarrow$
$3/\theta - 5 = \alpha$.

We can therefore substitute $3/\theta - 5$ for α in equation i):

$2.5 = (9 + 3/\theta - 5)/(3 + 1/\theta) \rightarrow$
$2.5 = (4 + 3/\theta)/(3 + 1/\theta) \rightarrow$
$2.5(3 + 1/\theta) = (4 + 3/\theta) \rightarrow$
$7.5 + 2.5/\theta = 4 + 3/\theta \rightarrow$
$3.5 = 0.5/\theta \rightarrow$
$\theta = 0.5/3.5 =$ **$\theta = 1/7 = 0.1428571429$**.

Problem S4C107-2. Similar to Question 217 of the Exam C Sample Questions from the Society of Actuaries. You are given the following sample of claims, where a "+" superscript indicates that the claim has exceeded the policy limit:
4, 4, 7, 7^+, 7^+, 9, 9, 9^+, 10, 10, 10, 20, 20^+, 25.

You estimate S(15), the survival function at 15, using two methods:
$\hat{S}_1(15)$ is the Kaplan-Meier product-limit estimate.

$\hat{S}_2(15)$ is the maximum likelihood estimate, assuming that claim amounts follow an exponential distribution.

Find $|\hat{S}_2(15) - \hat{S}_1(15)|$.

Solution S4C107-2. First we find $\hat{S}_1(15)$.The Kaplan-Meier estimate uses the formula

$S_n(t) = 1$ if $0 \leq t < y_1$;

$S_n(t) = {}_{i=1}{}^{j-1}\Pi((r_i - s_i)/r_i)$ if $y_{j-1} \leq t < y_j$ for j = 2, ..., k;

$S_n(t) = {}_{i=1}{}^{k}\Pi((r_i - s_i)/r_i)$ or 0 for $y_k \leq t$.

Here, the claims in excess of the limit are considered if the limit is at or above the amount whose survival function is being estimated, but not if they are below that amount.

The observations less than 15 are 4, 7, 9, and 10.

For 4, $s_i = 2$, and $r_i = 14$.

For 7, $s_i = 1$, and $r_i = 12$.

For 9, $s_i = 2$, and $r_i = 9$.

For 10, $s_i = 3$, and $r_i = 6$.

Thus, $\hat{S}_1(15) = (12/14)(11/12)(7/9)(3/6) = \hat{S}_1(15) = 11/36 = 0.30555555556$.

Now we find $\hat{S}_2(15)$.

Every claim of size x and not in excess of the policy limit contributes a factor of

$f(x) = (1/\theta)e^{-x/\theta}$ to the likelihood function. Every claim in excess of a policy limit of x contributes a factor of $S(x) = e^{-x/\theta}$ to the likelihood function.

There are 10 claims not in excess of the policy limit, and 4 claims in excess of the policy limit. Note that there will be a factor of $e^{-x/\theta}$ for each claim, irrespective of whether it exceeds the policy limit. Thus, the total exponentiated factor in the likelihood function will be

$e^{-\Sigma(x)/\theta}$, where $\Sigma(x) = 2*4 + 3*7 + 3*9 + 3*10 + 2*20 + 25 = 151$.

However, for each of the claims that do not exceed the limit, there is an additional factor of $(1/\theta)$.

Thus, $L(\theta) = (1/\theta^{10})e^{-151/\theta}$.

We find the loglikelihood function is $l(\theta) = \ln((1/\theta^{10})e^{-151/\theta}) = -10\ln(\theta) - 151/\theta$.

$l'(\theta) = -10/\theta + 151/\theta^2 = 0$ at the maximum, so

$10/\theta = 151/\theta^2 \rightarrow \theta = 151/10 = 15.1$.

Thus, $\hat{S}_2(15)= e^{-15/15.1} = \hat{S}_2(15) = 0.3703238139$, and $|\hat{S}_2(15)- \hat{S}_1(15)| =$

│0.3703238139- 0.30555555556│ = **0.0647682583.**

Problem S4C107-3. Similar to Question 218 of the Exam C Sample Questions from the Society of Actuaries. Random variable X has the following survival function: $S(x) = \theta^6/(\theta^3 + x^3)^2$. You observe one value of X to be 5, and you know that another value is greater than 5. Calculate the maximum likelihood estimate of θ.

Solution S4C107-3. First, we find $f(x) = F'(x) = (1-S(x))' = -6x^2*\theta^6/(\theta^3 + x^3)^3$.

The likelihood function is $L(\theta) = f(5)*S(5) = (-6*5^2*\theta^6/(\theta^3 + 5^3)^3)(\theta^6/(\theta^3 + 5^3)^2) =$

$-150\theta^{12}/(\theta^3 + 125)^5$.

The loglikelihood function is $l(\theta) = \ln(-150\theta^{12}/(\theta^3 + 125)^5) =$

$\ln(-150\theta^{12}) - 5\ln((\theta^3 + 125))$.

Thus, $l'(\theta) = -150*12\theta^{11}/(-150\theta^{12}) - 5*3\theta^2/(\theta^3 + 125) = 0$ at the maximum.

$12/\theta = 15\theta^2/(\theta^3 + 125) \rightarrow 12(\theta^3 + 125) = 15\theta^3$.

$12\theta^3 + 1500 = 15\theta^3 \rightarrow 3\theta^3 = 1500 \rightarrow \theta^3 = 500 \rightarrow$ **θ = 22.36067977**.

Problem S4C107-4. Similar to Question 219 of the Exam C Sample Questions from the Society of Actuaries. The random variable X has a conditional probability density function $f(x|\theta) = 3x^2/\theta^3$ for $0 < x < \theta$. The prior distribution of θ is $\pi(\theta) = 6\theta^5$ for $0 < \theta < 1$.

During a single time period, you make one observation of X, which is x = 0.5. For the same entity that was observed, determine the Bühlmann credibility estimate of the observation in the next time period.

Solution S4C107-4. The Bühlmann credibility factor is N/(N + K), where N is the number of time periods, and K = EPV/VHM. Here, we are given that N = 1.

$EPV = E_\theta(Var(x|\theta))$, and $VHM = Var_\theta(E(x|\theta))$.

It is useful to find the following:

$E(x|\theta)$, $Var(x|\theta)$, $E_\theta(\theta)$, $E_\theta(\theta^2)$, and $Var_\theta(\theta)$.
$E(x|\theta) = {}_0{}^\theta\int x*(3x^2/\theta^3)dx = 3x^4/4\theta^3 |_0{}^\theta = 3\theta/4$.
$E(x^2|\theta) = {}_0{}^\theta\int x^2*(3x^2/\theta^3)dx = 3x^5/5\theta^3 |_0{}^\theta = 3\theta^2/5$.
$Var(x|\theta) = E(x^2|\theta) - E(x|\theta)^2 = 3\theta^2/5 - 9\theta^2/16 = 3\theta^2/80$.
$E_\theta(\theta) = {}_0{}^1\int\theta*6\theta^5*d\theta = (6\theta^7/7) |_0{}^1 = 6/7$.
$E_\theta(\theta^2) = {}_0{}^1\int\theta^2*6\theta^5*d\theta = (6\theta^8/8) |_0{}^1 = ¾$.
$Var_\theta(\theta) = E_\theta(\theta^2) - E_\theta(\theta)^2 = ¾ - (6/7)^2 = Var_\theta(\theta) = 3/196$.
$EPV = E_\theta(Var(x|\theta)) = E_\theta(3\theta^2/80) = (3/80)E_\theta(\theta^2) = (3/80)(3/4) = EPV = 9/320$.

VHM = $\text{Var}_\theta(E(x | \theta))$ = $\text{Var}_\theta(3\theta/4)$ = $(9/16)\text{Var}_\theta(\theta)$ = (9/16)(3/196) = VHM = 27/3136.
K = EPV/VHM = (9/320)/(27/3136) = K = 49/15 = 3.266666667.
Thus, Z = 1/(1 + 49/15) = Z = 15/64 = 0.234375.

The external mean is M = $E_\theta(E(x | \theta))$ = $E_\theta(3\theta/4)$ = $(3/4)E_\theta(\theta)$ = (3/4)(6/7) = M = 9/14.

Thus, our credibility estimate is (15/64)(0.5) + (49/64)(9/14) = **39/64 = 0.609375.**

Problem S4C107-5. Similar to Question 221 of the Exam C Sample Questions from the Society of Actuaries. You have the following sample of values:

4, 5, 6, 7, 8, 8, 8, 8.

You use two kernel density estimators for the cumulative distribution function:

$F^{\wedge}_1(x)$ is an estimator that uses a uniform kernel of bandwidth 1.
$F^{\wedge}_2(x)$ is an estimator that uses a triangular kernel of bandwidth 1.

In the interval for x from 3 to 9, find all of the sub-intervals for which $F^{\wedge}_1(x) = F^{\wedge}_2(x)$.

Solution S4C107-5. There are 8 values in the sample, so the empirical probability of any one of them (repeats included) occurring is 1/8. The pdf using a kernel density estimator is f^(x) = $_{j=1}{}^{k}\Sigma(p(y_j)*k_{y\_j}(x))$. For a uniform kernel, $k_y(x) = 1/(2b)$ if $y-b \leq x \leq y + b$;
$k_y(x) = 0$ if $x > y + b$.
$k_y(x) = 0$ if $x < y - b$;

For a uniform kernel

A uniform kernel with bandwidth 1 will contribute to the pdf (1/8)/2 = 1/16 within 1 unit of x in both directions for every occurrence of x within the sample.

We consider the various intervals:
In the interval 3 < x < 4, only the one occurrence of 4 contributes 1/16 to $f^{\wedge}_1(x)$.

In the interval 4 < x < 5, the one occurrence of 4 contributes 1/16 to $f^{\wedge}_1(x)$, and the one occurrence of 5 contributes another 1/6, so $f^{\wedge}_1(x)$ = 1/8.

In the interval 5 < x < 6, the one occurrence of 5 contributes 1/16 to $f^{\wedge}_1(x)$, and the one occurrence of 6 contributes another 1/6, so $f^{\wedge}_1(x)$ = 1/8.

In the interval 6 < x < 7, the one occurrence of 6 contributes 1/16 to $f^{\wedge}_1(x)$, and the one occurrence of 7 contributes another 1/6, so $f^{\wedge}_1(x)$ = 1/8.

In the interval 7 < x < 8, the one occurrence of 7 contributes 1/16 to $f^{\wedge}_1(x)$, and the four occurrences of 8 contribute 4/16, so $f^{\wedge}_1(x)$ = 5/16.

In the interval 8 < x < 9, the four occurrences of 8 contribute 4/16, so $f\hat{}_1(x)$ = 4/16 = ¼.

For a triangular kernel, every occurrence of x within the sample contributes 1/8 to the pdf at x, and the contribution to the pdf decreases linearly in both directions until reaching 0 at x-1 and x + 1.
In the interval 3 < x < 4, the $f\hat{}_2(x)$ increases linearly from 0 at 3 to 1/8 at 4. Note that

$F\hat{}_1(4) = F\hat{}_2(4)$, since ${}_a{}^{a+1}\int(1/16)dx = {}_a{}^{a+1}\int(x/8 - a/8)dx$.

In the interval 4 < x < 5, the one occurrence of 4 contributes a linearly decreasing function from 4 to 5, while the occurrence of 5 contributes a linearly increasing function from 4 to 5. These add to a constant pdf of 1/8, so $f\hat{}_1(x) = f\hat{}_2(x)$ = 1/8 over this interval. Since the starting cdfs for this interval were equal, so are the cdfs throughout the interval.

In the interval 5 < x < 6, the one occurrence of 5 contributes a linearly decreasing function from 5 to 6, while the occurrence of 6 contributes a linearly increasing function from 5 to 6. These add to a constant pdf of 1/8, so $f\hat{}_1(x) = f\hat{}_2(x)$ = 1/8 over this interval. Since the starting cdfs for this interval were equal, so are the cdfs throughout the interval.

In the interval 6 < x < 7, the one occurrence of 6 contributes a linearly decreasing function from 6 to 7, while the occurrence of 7 contributes a linearly increasing function from 6 to 7. These add to a constant pdf of 1/8, so $f\hat{}_1(x) = f\hat{}_2(x)$ = 1/8 over this interval. Since the starting cdfs for this interval were equal, so are the cdfs throughout the interval.

In the interval 7 < x < 8, the one occurrence of 7 contributes a linearly decreasing function from 7 to 8, but the four occurrences of 8 contribute a linearly increasing function from 7 to 8 that is four times as steep. The resulting density function is not the same as $f\hat{}_1(x)$ = 5/16, and so the cdfs diverge from this point on.

Thus, $\mathbf{F\hat{}_1(x) = F\hat{}_2(x)}$ **in the interval 4 < x < 7**.

Section 108

Assorted Exam-Style Questions for Exam 4/C – Part 24

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C108-1. Similar to Question 226 of the Exam C Sample Questions from the Society of Actuaries. The number of claims (N) is 3 with probability θ and 6 with probability (1-θ). The prior distribution of Θ is such that θ = 0.5 with probability 0.3, and θ = 0.25 with probability 0.7. 6 claims are observed in year 1. Find the Bayesian estimate of the number of claims in year 2.

Solution S4C108-1. We want to find the posterior probabilities Pr(θ = 0.5 | N = 6) and

Pr(θ = 0.25 | N = 6).

Pr(θ = 0.5 and N = 6) = Pr(θ = 0.5)*Pr(N = 6) = 0.3*(1-0.5) = 0.15.

Pr(θ = 0.25 and N = 6) = Pr(θ = 0.25)*Pr(N = 6) = 0.7*(1-0.25) = 0.525.

Pr(N = 6) = 0.15 + 0.525 = 0.675.

Pr(θ = 0.5 | N = 6) = Pr(θ = 0.5 and N = 6)/Pr(N = 6) = 0.15/0.675 = 2/9.

By implication, Pr(θ = 0.25 | N = 6) = 1 - 2/9 = 7/9.

We also want to find E(N | θ = 0.5) = 3*0.5 + 6*0.5 = 4.5 and

E(N | θ = 0.25) = 3*0.25 + 6*0.75 = 5.25.

Thus, the Bayesian estimate of the number of claims in year 2 is

(2/9)*4.5 + (7/9)*5.25 = **5.083333333.**

Problem S4C108-2. Similar to Question 228 of the Exam C Sample Questions from the Society of Actuaries. Bank failures occurred at times $y_1 < y_2 < y_3 < y_4$.

You know the following Nelson-Åalen estimates of the hazard rate function:
$\hat{H}(y_2) = 0.65$; $\hat{H}(y_3) = 0.87222222$.

You also know the estimated variances of these estimates:
Var^($\hat{H}(y_2)$) = 0.039166666; Var^($\hat{H}(y_3)$) = 0.0638580247.

Determine the number of banks that were observed to fail at time y_3.

Solution S4C108-2. Let s_3 denote the number of bank failures at time y_3. Let r_3 denote the risk set size at time y_3.

Then $\hat{H}(y_2) + s_3/r_3 = \hat{H}(y_3)$

and Var^($\hat{H}(y_2)$) + s_3/r_3^2= Var^($\hat{H}(y_3)$).

We have a system of two equations:
i) $0.65 + s_3/r_3 = 0.87222222$

ii) $0.039166666 + s_3/r_3^2 = 0.0638580247$

We rearrange i) to get i)' $s_3/r_3 = 0.22222222 = 2/9$.

We rearrange ii) to get ii)' $s_3/r_3^2 = 2/81$.

We take (i)')^2/ii)': $(4/81)/(2/81) = (s_3/r_3)^2/(s_3/r_3^2) \rightarrow$

$4/2 = s_3 = 2$. So **2 banks** were observed to fail at time y_3.

Problem S4C108-3. Similar to Question 229 of the Exam C Sample Questions from the Society of Actuaries. You are given the probability density function of random variable X:
$f(x) = 2\theta/(\theta + x)^3$ for $0 < x < \infty$ and $\theta > 0$. For a sample size of n, find the asymptotic variance of the maximum likelihood estimate of θ. Express the answer in terms of n and θ.

Solution S4C108-3. Since only one parameter is being estimated, we can eventually use the formula for Fisher information for one parameter: $I(\theta) = -E((\partial^2(l(\theta)/\partial\theta^2))$.

Then we can apply the formula $1/I(\theta) = Var(\theta\hat{}_n)$ for sample size n.

However, developing a likelihood function for a sample size of n is quite complicated. It would be easier to assume a sample size of 1, get the resulting value for I(θ), and multiply that value by n.

If the sample size is 1 and the sample value is x, then

$L(\theta) = f(x) = 2\theta/(\theta + x)^3$.

$l(\theta) = \ln(L(\theta)) = \ln(2\theta/(\theta + x)^3) = \ln(2\theta) - 3*\ln(\theta + x) = \ln(2) + \ln(\theta) - 3*\ln(\theta + x)$

$l'(\theta) = 1/\theta - 3/(\theta + x)$

$l''(\theta) = -1/\theta^2 + 3/(\theta + x)^2$.

$E((\partial^2(l(\theta)/\partial\theta^2))$ is the expected value of $l''(\theta)$ *with respect to random variable X*.

Thus, $E((\partial^2(l(\theta)/\partial\theta^2)) = {}_0{}^\infty\int(-1/\theta^2 + 3/(\theta + x)^2)*f(x)*dx =$

$-1/\theta^2 + {}_0{}^\infty\int(3/(\theta + x)^2)*2\theta/(\theta + x)^3*dx =$

$-1/\theta^2 + {}_0{}^\infty\int(6\theta/(\theta + x)^5)*dx = -1/\theta^2 + (-1.5\theta/(\theta + x)^4)\Big|_0{}^\infty =$

$-1/\theta^2 + 1.5\theta/\theta^4 = -1/\theta^2 + 3/(2\theta^3)$.

$I(\theta) = -E((\partial^2(l(\theta)/\partial\theta^2)) = -(-1/\theta^2 + 3/(2\theta^3)) = 1/\theta^2 - 3/(2\theta^3)$ for a sample of 1 value.

For a sample of n values, $I(\theta) = n/\theta^2 - 3n/(2\theta^3) = (2n\theta - 3n)/(2\theta^3)$,

and so $1/I(\theta) =$ **$\mathrm{Var}(\theta\hat{}_n) = (2\theta^3)/(2n\theta - 3n)$**.

Problem S4C108-4. Similar to Question 230 of the Exam C Sample Questions from the Society of Actuaries. There are three groups of elephants: A, B, and C.

The number of peanuts eaten by each elephant in a given group of elephants follows a Poisson distribution with mean λ.

There are 60 elephants in group A, and the group A value of λ is 4.

There are 30 elephants in group B, and the group B value of λ is 12.

There are 10 elephants in group C, and the group C value of λ is 20.

You observe x peanuts eaten by a randomly selected elephant in year 1. The Bühlmann credibility estimate for the number of peanuts eaten by the same elephant in year 2 is 9.743589744. Find the value of x.

Solution S4C108-4. The Bühlmann credibility factor is Z = N/(N+K), where N = 1, as only one time period was observed. K = EPV/VHM.

We find the external mean M, which is the mean of peanuts eaten by the whole group.

There are 100 elephants in all, so we can find the probabilities that a randomly selected elephant will belong to each of the groups: Pr(A) = 60/100 = 0.6, Pr(B) = 30/100 = 0.3; Pr(C) = 10/100 = 0.1.

M = 0.6*4 + 0.3*12 + 0.1*20 = M = 8.

The value of EPV is also 8, since each group's number of peanuts eaten follows a Poisson distribution, and the Poisson distribution variance is equal to the mean. The mean of the variances is therefore the same as the mean of the means.

Now we find VHM. This is the variance of the group means around the overall mean of 8:
$0.6(8-4)^2 + 0.3(12-4)^2 + 0.1(20-4)^2$ = VHM = 54.4.

Thus, K = 8/54.4 = 5/34, and Z = 1/(1+5/34) = 34/39.

We know that 9.743589744 = (34/39)x + (1-34/39)*8, implying that

9.743589744 - (5/39)*8 = 8.717948719 = (34/39)x.

Thus, x = (39/34)*8.717948719 = **x = 10 peanuts**.

Problem S4C108-5. Similar to Question 232 of the Exam C Sample Questions from the Society of Actuaries. You are given the following sample of losses:
For Year 1: 5, 6, 8

For Year 2: 4, 5, 12

For Year 3: 9, 10, 10

In each year, losses follow a lognormal distribution with parameters μ and σ which are specific to that year. The annual inflation in expected losses is 20%.

Using losses trended for Year 4 (losses in each year that are inflation-adjusted so as to be at the level at which they would have been had they occurred in Year 4) and the method of moments to estimate the parameter μ in Year 4.

Relevant properties of lognormal distribution:

$E(X) = \exp(\mu + 0.5\sigma^2)$;

$E(X^2) = \exp(2\mu + 2\sigma^2)$.

Solution S4C108-5. First, we develop our sample of trended losses.
Each loss that occurred in Year 1 gets multiplied by 1.2^3. The trended losses from Year 1 are thus 8.64, 10.368, and 13.824.

Each loss that occurred in Year 2 gets multiplied by 1.2^2. The trended losses from Year 2 are thus 5.76, 7.2, and 17.28.

Each loss that occurred in Year 3 gets multiplied by 1.2. The trended losses from Year 3 are thus 10.8, 12, and 12.

The sample mean is (8.64 + 10.368 + 13.824 + 5.76 + 7.2 + 17.28 + 10.8 + 12 + 12)/9 = 10.87466667.

The sample mean of squares is

$(8.64^2 + 10.368^2 + 13.824^2 + 5.76^2 + 7.2^2 + 17.28^2 + 10.8^2 + 12^2 + 12^2)/9 = 129.056$.

Using the method of moments, we set

$10.87466667 = E(X) = \exp(\mu + 0.5\sigma^2)$;

$129.056 = E(X^2) = \exp(2\mu + 2\sigma^2)$;

Thus, we have a system of two equations:

i) $\ln(10.87466667) = 2.386435925 = \mu + 0.5\sigma^2$;

ii) $\ln(129.056) = 4.860246419 = 2\mu + 2\sigma^2$.

We take ii) - 2*i) to get

$\sigma^2 = 0.0873745684$, implying that $\mu = 2.386435925 - 0.5*0.0873745684 =$

μ = 2.342748641.

Section 109

Assorted Exam-Style Questions for Exam 4/C – Part 25

This section provides additional exam-style practice with a variety of syllabus topics.

Questions about **covariance**, as applied to other syllabus topics,may appear on the exam. The following equation is a reminder to the student of how covariance can be calculated in many cases. Let X and Y be two random variables. Then Equation 109.1 holds:

Equation 109.1. Cov(X, Y) = E(XY) - E(X)*E(Y).

The following property is useful with regards to p-p plots and D(x) (difference function) plots.

Property 109.2. Let there be a sample of n values, and let F(x) be the actual cumulative distribution of random variable X. It is assumed that the sample values are hypothesized to follow the distribution of X. Let $F_n(x)$ be the empirical cumulative distribution function of X, based on the sample of n values. Then the following are the case:
i) A D(x) plot plots $F_n(x)$ on the horizontal axis against F(x) on the vertical axis.

ii) A p-p plot plots $(n/(n+1))F_n(x)$ on the horizontal axis against F(x) on the vertical axis.

This is the only difference between the two plots.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C109-1. Similar to Question 233 of the Exam C Sample Questions from the Society of Actuaries. There are three groups of elephants: A, B, and C.

The number of peanuts eaten by each elephant in a given group of elephants follows a Poisson distribution with mean λ.

There are 60 elephants in group A, and the group A value of λ is 4.

There are 30 elephants in group B, and the group B value of λ is 12.

There are 10 elephants in group C, and the group C value of λ is 20.

You observe the elephants for one time period.

Determine the Bühlmann-Straub semiparametric empirical Bayes estimate of the credibility factor *Z* for group A.

Solution S4C109-1. Since we are using semiparameteric estimation, and the frequency distribution is Poisson, we know that E_{total} = EPV^ = sample mean = (Expected Number of Peanuts)/(Number of Elephants) = (4*60 + 12*30 + 20*10)/(60+30+10) = EPV^ = 8.

We find VHM Numerator = $_i\Sigma(N_i*(E(R_i) - E(Total))^2) - (R-1)EPV$^, where R = the number of groups = 3.

Thus, VHM Numerator = $60*(4-8)^2 + 30*(12-8)^2 + 10*(20-8)^2 - (3-1)*8 = 2864$.

VHM Denominator = $N_{total} - (1/N_{total})*_i\Sigma(N_i^2)$, where N_{total} = 60+30+10 = 100,

so VHM Denominator = $100 - (1/100)(60^2+30^2+10^2) = 54$.

VHM^ = 2864/54 = 53.03703704.

Thus, K = EPV^/VHM^ = 8/53.03703704 = 0.1508379888, and

Z for group A is $N_A/(N_A + K)$ = 60/(60 + 0.1508379888) = **Z = 0.9974923377**.

Problem S4C109-2. Similar to Question 236 of the Exam C Sample Questions from the Society of Actuaries. There are n random variables $X_1, ..., X_n$, which are independently and identically distributed in their conditional distributions that depend on random variable Θ. Each random variable X_j has conditional mean $\mu(\theta) = E(X_j \mid \Theta = \theta)$ and conditional variance

$v(\theta) = Var(X_j \mid \Theta = \theta)$.

You estimate the value of X_{12} using the observed values of variables X_1 through X_{11} and Bühlmann credibility. Your estimate for the credibility factor Z is 0.67, and your estimate for EPV, the expected value of the process variance, is 4. Find $Cov(X_i, X_j)$ for $i \neq j$.

Solution S4C109-2. Z = N/(N + K), where N is the number of observations, or 11.

K = EPV/VHM, and we are given EPV = 4. We also know that Z = 0.67.

Thus, 0.67 = 11/(11 + 4/VHM), and 11 + 4/VHM = 11/0.67 = 16.41791045.

Thus, 4/VHM = 5.41791045, and VHM = 4/5.41791045 = 0.738292011.

VHM is the variance of hypothetical means or $Var_\theta(E(X_j \mid \Theta = \theta)) = Var_\theta(\mu(\theta))$.

Now we try to find what $Cov(X_i, X_j)$ is equal to using the formula

$Cov(X_i, X_j) = E(X_i*X_j) - E(X_i)*E(X_j)$.

We can relate the absolute expected values to the given conditional expected values as follows: $E(X_i) = E_\theta(E(X_i \mid \Theta = \theta))$, and likewise for the identically distributed random variable X_j.

$Cov(X_i, X_j) = E_\theta(E(X_i*X_j \mid \Theta = \theta)) - E_\theta(E(X_i \mid \Theta = \theta))*E_\theta(E(X_j \mid \Theta = \theta))$

Since X_i and X_j are independently distributed *in their conditional distributions* (but not in their absolute distributions), $E(X_i*X_j \mid \Theta = \theta) = E(X_i \mid \Theta = \theta)*E(X_j \mid \Theta = \theta)= \mu(\theta)*\mu(\theta)$.

Thus, $Cov(X_i, X_j) = E_\theta(\mu(\theta)^2) - E_\theta(\mu(\theta))*E_\theta(\mu(\theta)) = E_\theta(\mu(\theta)^2) - E_\theta(\mu(\theta))^2 = Var_\theta(\mu(\theta))$ = VHM =

$Cov(X_i, X_j)$ = 0.738292011.

Problem S4C109-3. Similar to Question 238 of the [Exam C Sample Questions] from the Society of Actuaries. A random variable X follows an exponential distribution with mean θ. You use X^3 as estimator of θ^3. Calculate the mean square error of this estimator.

Relevant properties of exponential distributions: $E(X^k) = k!*\theta^k$.

Solution S4C109-3. The mean square error is the expected value of the squared difference between the estimator and the value being estimated.

Thus, the MSE here is $E((X^3- \theta^3)^2) = E(X^6 - 2X^3\theta^3 + \theta^6) = E(X^6) - 2\theta^3E(X^3) + \theta^6 =$

$(6!)\theta^6 - 2\theta^3*(3!)\theta^3 + \theta^6 = 720\theta^6 - 12\theta^6 + \theta^6 =$ **MSE = $709\theta^6$**.

Problem S4C109-4. Similar to Question 241 of the [Exam C Sample Questions] from the Society of Actuaries. You are given the following sample of values:

3, 6, 8, 9, 10, 10, 20, 23, 28.

You hypothesize that the sample values follow a Pareto distribution with parameters α = 2 and θ = 13. You then draw both a p-p plot and a D(x) plot for this sample and the hypothesized distribution.
Let (a, b) be the coordinates of the p-p plot for a value of x = 20.

Find (a - b) - D(20).

Solution S4C109-4. We note that there are 9 values in the sample. By Property 109.2, the only difference between the p-p plot and the D(x) plot is that the empirical cumulative distribution function (cdf) in the p-p plot is multiplied by (n/(n+1)) = 9/10.

Since the horizontal axis contains the empirical distribution function $F_n(x)$ for the D(x) plot, and the vertical coordinate (F(x)) is the same on both plots, it follows that

(a - b) - D(20) = (a - b) - (10a/9 - b) = -a/9, where a = (9/10)$F_n(20)$, and $F_n(20)$ = 7/9. Thus, a = 7/10.

Thus, our answer is -a/9 = **-7/90 = -0.077777778**.

Problem S4C109-5. Similar to Question 243 of the Exam C Sample Questions from the Society of Actuaries. You are given the following distribution:
In the interval from 0 to 50, there are 230 values.

In the interval from 50 to 100, there are 90 values.

In the interval from 100 to 200, there are 100 values.

In the interval from 200 to 300, there are x values.

In the interval from 300 to 500, there are y values.

There is an unknown number of values greater than 500.

You know that there are 1000 values in all.

You also know, from the ogive corresponding to the distribution, that

$F_{500}(250)$ = 0.58, and $F_{500}(400)$ = 0.78. Find the value of y.

Solution S4C109-5. First, we can find the empirical cdf value $F_{500}(200)$ = (230 + 90 + 100)/1000 = 0.42. Assuming a uniform distribution of values within each class, we know that

0.58 - 0.42 = 0.16 of all the values are in the interval from 200 to 250, and, by implication 0.16*2 = 0.32 of the values are in the interval from 200 to 300. x = 0.32*1000 = 320.

We thus derive $F_{500}(300)$ = 0.42 + 0.32 = 0.74.

$F_{500}(400)$ - $F_{500}(300)$ = 0.78 - 0.74 = 0.04, which is the fraction of all values in the interval between 300 and 400. The fraction of all values in the interval between 300 and 500 is twice that, or 0.04*2 = 0.08. Therefore, y = 0.08*1000 = **y = 80.**

Section 110

Assorted Exam-Style Questions for Exam 4/C – Part 26

This section provides additional exam-style practice with a variety of syllabus topics.

The following property is true when the delta method of estimating functions of parameters is applied to multiple parameters.

Property 110.1.

Let X be a function of two parameters, a and o. Let â and ô be the maximum likelihood estimates of these parameters. Let these maximum likelihood estimates be independent of one another. Let F(x) be the cumulative distribution function of X.

Then we estimate the variance of F(x) as
Var^(F(x)) = Var(â)*(dF(x)/da)^2 + Var(ô)*(dF(x)/do)^2.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C110-1. Similar to Question 225 of the Exam C Sample Questions from the Society of Actuaries. Random variable X follows a lognormal distribution with parameters μ and σ. The maximum likelihood estimate of μ is 6, and the maximum likelihood estimate of σ is 2. The covariance matrix of these maximum likelihood estimates of μ and σ is as follows:

(0.003........0)
(0.........0.005).

Let F(x) be the lognormal cumulative distribution function.

Then d(F(x))/dμ = -φ(z)/σ, and d(F(x))/dσ = -z*φ(z)/σ. φ(z) = (1/√(2π))exp(-z^2/2) is the standard normal density function.

Using this information, find the 90% confidence interval for F(8000), the probability that X ≤ 8000.

Solution S4C110-1. First, we want to find the point estimate at the center of the confidence interval. This is Φ((ln(8000)-μ)/σ) = Φ((ln(8000)-6)/2) = Φ(1.49359841), which we find in Excel via the input "=NORMSDIST(1.49359841)", getting as our result 0.932359692.

We also note that the z value associated with x = 8000 is 1.49359841.

We can thus find d(F(8000))/dμ = -φ(1.49359841)/σ = -φ(1.49359841)/2 =

-(1/√(2π))exp(-1.49359841^2/2) = d(F(8000))/dμ = -0.1307645843.

d(F(8000))/dσ = -z*φ(z)/σ = z*(-0.1307645843) = 1.49359841*(-0.1307645843) = d(F(8000))/dσ = -0.1953097752.

Now we apply Property 110.1:

Var^(F(x)) = Var(μ)*$(dF(x)/d\mu)^2$ + Var(σ)*$(dF(x)/d\sigma)^2$.

The covariance matrix has zero entries for Cov(σ, μ), which indicates that the two parameters are independent. It also indicates that Var(μ) = 0.003, and Var(σ) = 0.005.

Thus, Var^(F(8000)) = 0.003$(-0.1307645843)^2$ + 0.005*$(-0.1953097752)^2$ =

Var^(F(8000)) = 0.000242027671.

For a 90% confidence interval, the appropriate z value by which to multiply the square root of the variance is Φ^{-1}((1+0.9)/2) = Φ^{-1}(0.95), which we find via the Excel input "=NORMSINV(0.95)", getting 1.644853627 as our result.

Thus, the endpoints of our confidence interval are

0.932359692 ± 1.644853627√(0.000242027671), and the interval itself is

(0.9067703118, 0.9579490722).

Problem S4C110-2. Similar to Question 246 of the [Exam C Sample Questions] from the Society of Actuaries. You have the following sample of 9 values:

3, 4, 7, 9, 10, 20, 25, 29, 30

You attempt to fit the sample to a Burr distribution with parameters α = 3, θ, and γ.

You estimate θ and γ via percentile matching, using the 20^{th} and 80^{th} empirically smoothed percentile estimates. Find the estimated value of γ.

Relevant property of Burr distributions:

$F(x) = 1 - (1/(1 + (x/\theta)^\gamma))^\alpha$.

Solution S4C110-2. Since the sample has n = 9 values, the pth empirically smoothed percentile estimate is the (n+1)*(p/100)th value in the sample. The 20^{th} empirically smoothed percentile estimate is the (9+1)*0.2 = 2^{nd} value in the sample, or 4.

The 80^{th} empirically smoothed percentile estimate is the (9+1)*0.8 = 8^{th} value in the sample, or 29.

Thus, we have a system of two equations:

i) $F(4) = 0.2 = 1 - (1/(1 + (4/\theta)^\gamma))^3$.

ii) $F(29) = 0.8 = 1 - (1/(1 + (29/\theta)^\gamma))^3$.

We rearrange the equations as follows:
i)' $0.8 = (1/(1 + (4/\theta)^\gamma))^3$.

ii)' $0.2 = (1/(1 + (29/\theta)^\gamma))^3$.

i)" $0.9283177667 = 1/(1 + (4/\theta)^\gamma)$.

ii)" $0.5848035476 = 1/(1 + (29/\theta)^\gamma)$.

i)"' $1.077217345 = 1 + (4/\theta)^\gamma$.

ii)"' $1.7099759467 = 1 + (29/\theta)^\gamma$.

i)"" $0.077217345 = (4/\theta)^\gamma$.

ii)"" $0.7099759467 = (29/\theta)^\gamma$.

We now divide ii)"" by i)"" to get

$0.7099759467/0.077217345 = (29/4)^\gamma$.

$9.19451383 = (29/4)^\gamma$.

$\ln(9.19451383) = \gamma*\ln(29/4)$.

$\gamma = \ln(9.19451383)/\ln(29/4) =$ $\boldsymbol{\gamma = 0.3060147563}$.

Problem S4C110-3. Similar to Question 245 of the Exam C Sample Questions from the Society of Actuaries. You know that claim frequency follows a Poisson distribution, and claim severity follows a gamma distribution with parameters $\alpha = 3$ and $\theta = 450$. You want the full credibility standard for *aggregate* losses to be within 4% of expected aggregate losses 94% of the time. Use limited fluctuation credibility theory to find the number of claims needed to achieve full credibility.

Relevant properties of gamma distributions: $E(X) = \alpha\theta$; $Var(X) = \alpha\theta^2$.

Solution S4C110-3. Here, the full credibility threshold is $\lambda_0(\sigma_f^2/\mu_f + \sigma_s^2/\mu_f^2)$. Note that we need to incorporate the mean and standard deviation of both frequency and severity, because our full credibility standard pertains to aggregate losses.

$\lambda_0 = (\Phi^{-1}((1+0.94)/2)/0.04)^2 = (\Phi^{-1}(0.97)/0.04)^2$, which we can find via the Excel input

"=(NORMSINV(0.97)/0.04)^2", getting as our result $\lambda_0 = 2210.865373$.

Since frequency is Poisson-distributed, $\sigma_f^2 = \mu_f$, and so $\sigma_f^2/\mu_f = 1$.

Severity is gamma-distributed, so $\sigma_s^2/\mu_f^2 = (\alpha\theta^2)/(\alpha\theta)^2 = 1/\alpha = 1/3$.

Thus, our full credibility threshold is 2210.865373(1 + 1/3) = 2947.820497 = **2498 claims**.

Problem S4C110-4. Similar to Question 247 of the Exam C Sample Questions from the Society of Actuaries. There are two groups of rabbits, A and B. 30% of the rabbits are in group A, and 70% are in group B. The number of carrots eaten per day by a rabbit in group A follows a Poisson distribution with mean 6. The number of carrots eaten per day by a rabbit in group B follows a Poisson distribution with mean 10. A randomly chosen rabbit is observed to eat 50 carrots over 6 days. Find the Bayesian estimate of the number of carrots this rabbit will eat during day 7.

Relevant properties of Poisson distributions: $Pr(N = k) = e^{-\lambda}*\lambda^k/k!$.

Solution S4C110-4. We note that the number of carrots eaten over 6 days follows a Poisson process with $\lambda_{6\text{ days}} = 6*\lambda_{1\text{ day}}$. Thus, for group A, $\lambda_{6\text{ days}} = 6*6 = 36$, and for group B, $\lambda_{6\text{ days}} = 6*10 = 60$.

First, we want to find Pr(Group A and 50 carrots in 6 days) = $0.3*e^{-\lambda}*\lambda^k/k! = 0.3*e^{-36}*36^{50}/50! = 0.0014947552$, and Pr(Group B and 6 carrots in 6 days.) = $0.7*e^{-\lambda}*\lambda^k/k! = 0.7*e^{-60}*60^{50}/50! = 0.0162898383$.

Thus, Pr(Group A │50 carrots in 6 days) =

Pr(Group A and 50 carrots in 6 days)/Pr(50 carrots in 6 days) =

0.0014947552/(0.0014947552 + 0.0162898383) = 0.0840477574, and

Pr(Group B | 50 carrots in 6 days) = 1 - 0.0840477574 = 0.9159522426. Given these posterior probabilities, the expected number of carrots eaten during the next *single* day is

0.0840477574*6 + 0.9159522426*10 = **9.66380897 carrots**.

Problem S4C110-5. Similar to Question 250 of the Exam C Sample Questions from the Society of Actuaries. You are given a sample of the following values from an inverse exponential distribution:

30, 40, 50.

You also know that 3 other values are below 20.

Find the maximum likelihood estimate of the mode of this distribution.

Relevant properties of inverse exponential distributions: $F(x) = e^{-\theta/x}$; $f(x) = \theta e^{-\theta/x}/x^2$; Mode = $\theta/2$.

Solution S4C110-5. Every sample value x that is exactly known contributes a factor of f(x) to the likelihood function. Every sample value below 20 contributes a factor of $F(20) = e^{-\theta/20}$.

Thus, the likelihood function $L(\theta) = (e^{-\theta/20})^3*(\theta e^{-\theta/30}/30^2)*(\theta e^{-\theta/40}/40^2)*(\theta e^{-\theta/50}/50^2)$.

$L(\theta) = \theta^3*e^{-137\theta/600}/(30^2*40^2*50^2)$.

The loglikelihood function is

$l(\theta) = \ln(\theta^3*e^{-137\theta/600}/(30^2*40^2*50^2)) = 3\ln(\theta) - 137\theta/600 - \ln(30^2*40^2*50^2)$.

$l'(\theta) = 3/\theta - 137/600 = 0$ at the maximum.

Thus, $3/\theta = 137/600$, and $\theta = 1800/137$, meaning that the mode = $\theta/2$ = **900/137 = 6.569343066**.

Section 111

Assorted Exam-Style Questions for Exam 4/C – Part 27

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C111-1. Similar to Question 251 of the Exam C Sample Questions from the Society of Actuaries. Loss amounts (X) follow a gamma distribution with parameters $\alpha = 3$ and θ. The prior distribution of θ has mean 40.

During years 1, 2, and 3, you observe loss amounts of 40, 20, and 90 for a particular policy. Loss data for year 4 has been lost and is unavailable.

Using only the available loss data and Bühlmann credibility, you estimate losses in year 5 to be 60. Then you locate the missing data for year 4 and now know that there were 85 losses in that year. Find the Bühlmann credibility estimate of losses in year 5 using this more complete information.
Relevant properties of gamma distributions: $E(X) = \alpha\theta$.

Solution S4C111-1. The external mean M here is $E_\theta(E(X|\theta)) = E_\theta(3\theta) = 3*E_\theta(\theta) = 3*40 = 120$.

The average loss amount for the first three years is $(40 + 20 + 90)/3 = 50$.

Let Z_3 be the Bühlmann credibility factor using data from only the first three years.

We are given that $(Z_3)*50 + (1-Z_3)*120 = 60 \rightarrow 120 - (Z_3)*70 = 60 \rightarrow$

$60 = 70(Z_3) \rightarrow Z_3 = 6/7 = 3/(3+K)$, so $3 + K = 3/(6/7) = 3.5$, and thus $K = 0.5$.

The average loss amount for the first four years is $(40 + 20 + 90 + 85)/4 = 58.75$, and the credibility factor Z_4 for this data is $4/(4+K) = 4/4.5 = 8/9$.

Thus, our desired credibility estimate is (8/9)*58.75 + (1/9)*120 = **65.5555555556.**

Problem S4C111-2. Review of Section 65. Similar to Question 254 of the Exam C Sample Questions from the Society of Actuaries. There are 500independent and identically distributed risks. The number of claims for each risks is Poisson-distributed with mean λ. The prior distribution of λ is gamma with parameters α = 20 and θ = 1/90. During year 1, you observe the following:

402 risks had 0 claims.

50 risks had 1 claim.

40 risks had 2 claims.

3 risks had 3 claims.

5 risks had 4 claims.

Using this information, find the Bayesian estimate of the expected number of claims for the entire group of risks in year 2.

Relevant properties of gamma distributions: E(X) = αθ.

Solution S4C111-2. If the model distribution is Poisson, and the prior distribution of λ is gamma, then we have an instance of Case 1 of conjugate prior distributions from Section 65. Thus, the posterior distribution is gamma with parameters $\alpha + \Sigma x_i$ and θ/(nθ+1). Here, n is the number of observations, or 500, and Σx_i is the number of claims observed or 1*50 + 2*40 + 3*3 + 5*4 = 159. Thus, the posterior distribution has parameters α* = 20 + 159 = 179, and θ* = (1/90)/(500/90 + 1) = 1/590. Hence, the Bayesian estimate of the expected number of claims for *one* risk is 179*(1/590) = 179/590, and the Bayesian estimate of the expected number of claims for the entire pool or risks is 500(179/590) = **8950/59 = 151.6949153 claims**.

Problem S4C111-3. Similar to Question 256 of the Exam C Sample Questions from the Society of Actuaries. You are examining 304 insurance policies, of which 204 policies had 0 claims, 50 policies had 1 claim, and 50 policies had 2 claims. You attempt to fit this data to a binomial distribution with m = 2 and q estimated via the method of maximum likelihood. Find the maximum value of the loglikelihood function l(q).

Solution S4C111-3. First, we create our likelihood function l(q).

Pr(0 claims) = $(1-q)^2$, and there are 204 such instances, so we have a factor of $((1-q)^2)^{204}$ =

$(1-q)^{408}$ contributing to the likelihood function.

Pr(1 claim) = 2q*(1-q), and there are 50 such instances, so we have a factor of $(2q*(1-q))^{50}$ =

$2^{50}*q^{50}*(1-q)^{50}$ contributing to the likelihood function.

Pr(2 claims) = q^2, and there are 50 such instances, so we have a factor of $(q^2)^{50} = q^{100}$ contributing to the likelihood function.

Thus, $L(q) = (1-q)^{408}*2^{50}*q^{50}*(1-q)^{50}*q^{100} = 2^{50}*(1-q)^{458}*q^{150}$.

$l(q) = \ln(2^{50}*(1-q)^{458}*q^{150}) = 50*\ln(2) + 458*\ln(1-q) + 150*\ln(q)$.

$l'(q) = -458/(1-q) + 150/q = 0$ at the maximum.

Thus, $458/(1-q) = 150/q$, and $458q = 150(1-q) \rightarrow 458q = 150 - 150q \rightarrow 608q = 150 \rightarrow q =$ 0.2467105263.

The maximum value we seek is

$l(0.2467105263) = 50*\ln(2) + 458*\ln(1-0.2467105263) + 150*\ln(0.2467105263) =$

-305.0275888.

Problem S4C111-4. Review of Section 42. Similar to Question 258 of the Exam C Sample Questions from the Society of Actuaries. You are given the following sample of values:
3, 4, 5, 6, 6, X.

You also know the following:
$s_3 = 1$;

$r_4 = 2$;

$k = 4$;

The Nelson-Åalen estimate $\hat{H}(7) < 1.91$.

Which of the following could be a value of X?

(a) 3
(b) 4
(c) 5
(d) 6
(e) 7

Solution S4C111-4. 5 is the third-lowest value in the sample, and we are given that $s_3 = 1$. Since there is already one occurrence of 5 in the sample, X cannot be 5, so (c) is not true. We are given that $r_4 = 2$, and, as 6 is the fourth-lowest value in the sample, we know that only 2 values can be greater than or equal to 6. We already have 2 values of 6 in the sample, so X cannot be 6 or greater, and so (d) and (e) are not true.

If X were 3, then $\hat{H}(7)$ = 2/6 + ¼ + 1/3 + 2/2 = 1.91666666667.

If X were 4, then $\hat{H}(7)$ = 1/6 + 2/5 + 1/3 + 2/2 = 1.9.

$\hat{H}(7)$ is only less than 1.91 if X = 4, so the correct answer is **(b) 4**.

Problem S4C111-5. Similar to Question 259 of the Exam C Sample Questions from the Society of Actuaries. You have the following sample of values:
5, 5, 6, 6, 7, 7, 7, 8, 8, 8, 8, 9, 9, 9, 9, 9, 10, 10, 10, 10.

You attempt to fit these values to a Poisson distribution using the method of maximum likelihood. Find the coefficient of variation of the maximum likelihood estimator.

Hint: Sample size matters here, and not just in the most obvious way!

Solution S4C111-5. There are 20 values in the sample. The maximum likelihood estimate of the mean λ is just the sample mean: (5*2 + 6*2 + 7*3 + 8*4 + 9*5 + 10*4)/20 = 8.

The variance of the maximum likelihood estimator is *not* λ, but is rather λ/n, because the maximum likelihood estimator is based on n occurrences of values from a Poisson distribution. Here, n = 20. Thus, the standard deviation of the maximum likelihood estimator is √(8/20), and the coefficient of variation is √(8/20)/8 = 1/√(8*20) = 1/√(160) = **0.0790569415**.

Section 112

Assorted Exam-Style Questions for Exam 4/C – Part 28

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C112-1. Similar to Question 260 of the Exam C Sample Questions from the Society of Actuaries. Number of peanuts eaten by an elephant in a day have 0.3 probability of following an exponential distribution with θ = 10 and 0.7 probability of following an exponential distribution with θ = 6. An elephant was observed on day 1 to have eaten 8 peanuts. What is the Bayesian estimate of the expected number of peanuts the elephant will eat on day 2?

Solution S4C112-1. We first find Pr(θ = 10 and 8 peanuts eaten) = 0.3*f(8) = 0.3*(1/10)$e^{-8/10}$ = 0.0134798689.

We also find Pr(θ = 6 and 8 peanuts eaten) = 0.7(1/6)$e^{-8/6}$ = 0.0307529994.

Thus, Pr(θ = 10 │ 8 peanuts eaten) = Pr(θ = 10 and 8 peanuts eaten)/Pr(8 peanuts eaten) =

0.0134798689/(0.0134798689 + 0.0307529994) = 0.304747791. By implication,

Pr(θ = 6 │ 8 peanuts eaten) = 1 - 0.304747791 = 0.695252209.

Since the expected value for an exponential distribution is θ, our desired Bayesian estimate is the posterior-probability-weighted average of the possible values of θ:

0.304747791*10 + 0.695252209*6 = **7.218991164.**

Problem S4C112-2. Similar to Question 262 of the Exam C Sample Questions from the Society of Actuaries. You know the following about the number of peanuts eaten by a group of elephants: All elephants ate more than 10 peanuts. Two of the elephants ate 13 peanuts. One elephant ate 20 peanuts. Four more elephants ate in excess of 10 + p peanuts, but the exact quantity is unknown. The number of peanuts eaten is assumed to be uniformly distributed on the interval $(0, \omega)$. The maximum likelihood estimate of ω is 40. Find the value of p.

Solution S4C112-2. First, we find the likelihood function $L(\omega)$.

We note that because it is known that every elephant ate more than 10 peanuts, the observed values can be assumed to be uniformly distributed not on $(0, \omega)$ but on $(10, \omega)$. This corresponds to a probability density function of $1/(\omega - 10)$. The three elephants who ate known values of peanuts each contribute a factor of $1/(\omega - 10)$ to the likelihood function.

The four elephants who ate more than 10 + p peanuts each contribute a factor of $S(10 + p) =$

$(\omega - 10 - p)/(\omega - 10)$ to the likelihood function.

Thus, $L(\omega) = (1/(\omega - 10))^3((\omega - 10 - p)/(\omega - 10))^4 = (\omega - 10 - p)^4/(\omega - 10)^7$.

The loglikelihood function $l(\omega) = \ln((\omega - 10 - p)^4/(\omega - 10)^7) = 4\ln(\omega - 10 - p) - 7\ln(\omega - 10)$.

$l'(\omega) = 4/(\omega - 10 - p) - 7/(\omega - 10) = 0$ at the maximum.

Thus, $4/(\omega - 10 - p) = 7/(\omega - 10)$. We already know that $\omega = 40$, so we can solve for p:

$4/(30 - p) = 7/30 \rightarrow 4/(7/30) = 30 - p = 120/7$. Thus, $p = 30 - 120/7 =$ **$p = 90/7 = 12.85714286$**.

Problem S4C112-3. Similar to Question 264 of the Exam C Sample Questions from the Society of Actuaries. You have the following sample of values:

3, 3, 5, 5, 6, 7, 8, 8, 8, 12.

You make two estimates of the cumulative hazard rate function H(x).

$\hat{H}_1(x)$ is the Nelson-Åalen estimate assuming that all the data points reflect uncensored values.

$\hat{H}_2(x)$ is the Nelson-Åalen estimate assuming that all data points of 5 and 8 are actually points of right censoring, and so the actual values are greater than 5 or 8, as applicable.

Find $|\hat{H}_1(10) - \hat{H}_2(10)|$.

Solution S4C112-3. First, we find $\hat{H}_1(10)$. Our initial risk set is the sample size 10. For each unique sample value, the risk set is (10 - number of observations smaller than the value in question). Only 12 is greater than 10, so we consider all the other sample values in finding $\hat{H}_1(10)$.

Thus, $\hat{H}_1(10) = 2/10 + 2/8 + 1/6 + 1/5 + ¾ = 1.5666666667$.

In finding $\hat{H}_2(10)$, we consider the right-censored values as part of the risk set when they are greater than the value we are looking at, but we do not consider them as observed occurrences s_i, because we do not know the actual, uncensored values. Thus, we leave out the contributions that considering values of 5 and 8 made to $\hat{H}_1(10)$:

$\hat{H}_2(10)$ = 2/10 + 1/6 + 1/5 = 0.5666666667.

Hence, $|\hat{H}_1(10) - \hat{H}_2(10)|$ = 1.5666666667 - 0.5666666667 = **1.**

Problem S4C112-4. Similar to Question 267 of the [Exam C Sample Questions] from the Society of Actuaries. Claims can follow one of two probability distributions. They follow a Poisson distribution with mean λ = 1 with probability 0.4 and a Poisson distribution with mean λ = 10 with probability 0.6. In year 1, there were k claims reported. Based on this information, the Bayesian estimate of the number of claims in year 2 is 1.163575218. Find the Bühlmann credibility estimate for the number of claims in year 2.

Relevant properties of Poisson distributions: Pr(N = k) = $e^{-\lambda}*\lambda^k/k!$

Solution S4C112-4. First, we find the expression for the Bayesian estimate of the number of claims in year 2.

Pr(λ = 1 and k claims) = $0.4*e^{-1}*1^k/k! = 0.4*e^{-1}/k!$.

Pr(λ = 10 and k claims) = $0.6*e^{-10}*10^k/k!$.

Pr(λ = 1 | k claims) = Pr(λ = 1 and k claims)/Pr(k claims) =

$0.4*e^{-1}/k!/(0.4*e^{-1}/k! + 0.6*e^{-10}*10^k/k!) = 0.4/(0.4 + 0.6*e^{-9}*10^k)$.

Pr(λ = 10 | k claims) = Pr(λ = 10 and k claims)/Pr(k claims) =

$(0.6*e^{-10}*10^k/k!)/(0.4*e^{-1}/k! + 0.6*e^{-10}*10^k/k!) = (0.6*e^{-9}*10^k)/(0.4 + 0.6*e^{-9}*10^k)$.

Thus, the Bayesian estimate is

1*Pr(λ = 1 | k claims) + 10*Pr(λ = 10 | k claims) =

$0.4/(0.4 + 0.6*e^{-9}*10^k) + 10*(0.6*e^{-9}*10^k)/(0.4 + 0.6*e^{-9}*10^k)$ =

$(0.4 + 6*e^{-9}*10^k)/(0.4 + 0.6*e^{-9}*10^k)$, which we know to be 1.163575218.

Thus,

$1.163575218(0.4 + 0.6*e^{-9}*10^k) = 0.4 + 6*e^{-9}*10^k \rightarrow$

$0.163575218*0.4 = 5.301864869e^{-9}*10^k \rightarrow$

$10^k = (0.163575218*0.4)/(5.301864869e^{-9}) = 100$, implying that k = 2.

In the Bühlmann credibility estimate, k = 2 will be X¯, the observed mean.

We find M, the external mean = 0.4*1 + 0.6*10 = 6.4.

We also find EPV = M = 6.4, because the underlying distribution is a mixture of Poisson distributions.

We find VHM = $0.4(1-6.4)^2 + 0.6(10-6.4)^2$ = VHM = 19.44.

Thus, K = EPV/VHM = 6.4/19.44 = 0.329218107. Because only one period was observed, N = 1, and so Z = 1/(1 + 0.329218107) = 0.7523219814, and our credibility estimate is

0.7523219814*2 + (1-0.7523219814)*6.4 = **3.089783282.**

Problem S4C112-5. Review of Section 53. Similar to Question 268 of the Exam C Sample Questions from the Society of Actuaries. You have the following sample of 5 values:

4, 6, 6, 8, 10.

You use a kernel density estimator with bandwidth 4 on this sample. Using this estimator, estimate Ŝ(9), the survival function at 9.

Solution S4C112-5. For a kernel density estimator with bandwidth 4, each sample value contributes a uniform probability density function of 1/(2*4) = 1/8 spanning 4 units to each side of the sample value.

Thus, the sample value of 4 contributes a pdf of 1/8 spanning from 0 to 8, and contributes nothing above 9.

Each sample value of 6 contributes a pdf of 1/8 spanning from 2 to 10, so the amount contributed to values above 9 is (10-9)*1/8 = 1/8. The two sample values of 6 thus contribute 2/8 = ¼ to the survival function estimate.

The sample value of 8 contributes a pdf of 1/8 spanning from 4 to 12, so the amount contributed to values above 9 is (12-9)*1/8 = 3/8.

The sample value of 10 contributes a pdf of 1/8 spanning from 6 to 14, so the amount contributed to values above 9 is (14-9)*1/8 = 5/8.

We cannot just add up these contributions, however. We need to weigh them by the empirical probabilities that each of the relevant sample values occurs. Thus, our answer is

Ŝ(9) = (2/5)*(1/4) + (1/5)*(3/8) + (1/5)*(5/8) = **Ŝ(9) = 0.3**.

Section 113

Assorted Exam-Style Questions for Exam 4/C – Part 29

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C113-1. Similar to Question 269 of the Exam C Sample Questions from the Society of Actuaries. You have a sample of 8 values, drawn from n independent and identically distributed random variables, each with mean $\theta = 30$. Let Q be the sample mean.

Find $E((Q - 5)^2)$.

Relevant properties of exponential distributions: $Var(X) = \theta^2$.

Solution S4C113-1. $E((Q - 5)^2) = E(Q^2 - 10Q + 25) = E(Q^2) - 10E(Q) + 25$.

The sample mean is an unbiased estimator of the population mean, so $E(Q) = \theta = 30$.

We only need to find $E(Q^2)$.

We note that $E(Q^2) = Var(Q) + E(Q)^2$.

The variance of a *sample mean* is not the same as the variance of the distribution; to get the variance of the sample mean, it is necessary to divide the variance of the distribution by the sample size, which in this case is 8. Thus, $Var(Q) = \theta^2/8$, and

$E(Q^2) = \theta^2/8 + \theta^2 = 9\theta^2/8 = 9*30^2/8 = 1012.5$.

Thus, $E(Q^2) - 10E(Q) + 25 = 1012.5 - 10*30 + 25 =$ **737.5.**

Problem S4C113-2. Review of Section 73. Similar to Question 270 of the Exam C Sample Questions from the Society of Actuaries. There are two policyholders: A and B.

In year 1, A had 4 claims; in year 2, A had 6 claims; in year 3, A had 5 claims.

In year 1, B had 8 claims; in year 2, B had 12 claims; in year 3, B had 10 claims.

Use nonparametric empirical Bayes estimation of the Bühlmann credibility factor to find an estimate for the variance of hypothetical means.

Solution S4C113-2. First, we find EPV^ = $({}_i\Sigma(\text{SampleVar}(R_i))/R)$.

For policyholder A, the expected number of claims is (4 + 6 + 5)/3 = 5, and so the sample variance of claims is

$((4-5)^2 + (6-5)^2 + (5-5)^2)/(3-1) = 2/2 = 1$.

For policyholder B, the expected number of claims is (8 + 12 + 10)/3 = 10, and so the sample variance of claims is $((8-10)^2 + (12-10)^2 + (10-10)^2)/(3-1) = 2$.

EPV^ is the average of these sample variances or (1+2)/2 = EPV^ = 1.5.

Now we use the formula VHM^ = $(1/(R-1))*{}_i\Sigma(E(R_i) - E(\text{Total}))^2$ -EPV^/N,

where N is the number of years, 3, and R is the number of risks, 2. E(Total) = (5+10)/2 = 7.5.

Thus, VHM^ = $(1/(2-1))((5-7.5)^2 + (10-7.5)^2) - 1.5/3$ = **VHM^ = 12**.

Problem S4C113-3. Similar to Question 272 of the Exam C Sample Questions from the Society of Actuaries. The number of peanuts eaten by an elephant in an hour has a binomial distribution with parameters m = 5 and q. The prior distribution of q is π(q) = 5q(1-q), for 0 < q < 1. You observe an elephant eating 3 peanuts. Find the mode of the posterior distribution of q. You may use a calculator to solve any integrals.

Solution S4C113-3. First, we find the model distribution, which is the probability in terms of q that the number of peanuts is 3, i.e., $C(5,3)*q^3*(1-q)^2 = 10*q^3*(1-q)^2$.

Joint Distribution = (Prior Distribution)*(Model Distribution) = $5q(1-q)*10*q^3*(1-q)^2$ =

$50q^4*(1-q)^3$.

Marginal Distribution = ∫(Joint Distribution) = ${}_0{}^1\int 50q^4*(1-q)^3*dq = 5/28$.

Thus, Posterior Distribution = (Joint Distribution)/(Marginal Distribution) = $(50q^4*(1-q)^3)/(5/28)$ = $280q^4*(1-q)^3$. To find the mode of this distribution, we simply find its maximum.

The derivative of the posterior distribution is
$1120q^3*(1-q)^3 - 840q^4*(1-q)^2 = 0$ at the maximum. Thus,
$840q^4*(1-q)^2 = 1120q^3*(1-q)^3$.
$840q = 1120(1-q)$.
$1-q = 0.75q$.
$1 = 1.75q$.

q = 4/7 = 0.571428574.

Problem S4C113-4. Similar to Question 273 of the [Exam C Sample Questions] from the Society of Actuaries. You are using limited fluctuation credibility theory, and claim frequency follows a Poisson distribution. You know that the full-credibility standard for the number of claims is 500 claims, if you desire the number of claims to be within 12% of the true value with probability p. Now the standard of credibility changes. You want the standard to apply to *aggregate* claims, where claim severity follows a uniform distribution on the interval from 0 to 500. You want aggregate claims to be within 5% of the true value with probability p.

Solution S4C113-4. We recall the formula for the minimum number of observations required for full credibility of frequency data:
$n = \lambda_0*(\sigma_f^2/\mu_f)$. For a Poisson distribution the mean is equal to the variance, so $\sigma_f^2/\mu_f = 1$.

Here, $\lambda_0 = \Phi^{-1}((1+p)^2/2)/0.12^2$. When the credibility standard changes, the new λ_0 becomes $\Phi^{-1}((1+p)^2/2)/0.05^2$. This change necessitates a multiplication of the original n by $(0.12^2/0.05^2) = 5.76$. Also, the formula for the credibility threshold changes to $\lambda_0*(\sigma_f^2/\mu_f + \sigma_s^2/\mu_s^2)$. The mean of the uniform severity distribution is 500/2 = 250, and the variance is $500^2/12 = 20833.3333333$, and so $\sigma_s^2/\mu_s^2 = 20833.3333333/250^2 = 1/3$, and thus, $(\sigma_f^2/\mu_f + \sigma_s^2/\mu_s^2) = 1 + 1/3 = 4/3$, which is also a factor by which the original n is multiplied.

The new credibility standard is thus 500*5.76*(4/3) = **3840 claims.**

Problem S4C113-5. Similar to Question 274 of the [Exam C Sample Questions] from the Society of Actuaries. Automobiles are driving on an extremely dangerous road, where many of them crash.

There are 50 automobiles at risk at time t = 1, and 6 of them crash.
There are 35 automobiles at risk at time t = 4, and 5 of them crash.
There are 20 automobiles at risk at time t = 9, and 11 of them crash.
There are 7 automobiles at risk at time t = 11, and k of them crash.

You know that the Nelson-Åalen estimate of the *survival* function at time t = 11 is $\hat{S}(11) = 0.3333469536$. Find the value of k.

Solution S4C113-5. The estimate of the hazard rate function, $\hat{H}(11)$ is $-\ln(\hat{S}(11)) = -\ln(0.3333469536) = 1.098571429$.
We know that 6/50 + 5/35 + 11/20 + k/7 = 1.098571429, meaning that
k/7 = 0.2857142857 = 2/7, and thus **k = 2.**

Section 114

Assorted Exam-Style Questions for Exam 4/C – Part 30

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C114-1. Similar to Question 275 of the Exam C Sample Questions from the Society of Actuaries. An insurer applies a deductible of 300 to each claim. Then the insurer pays 50% of the amount of loss amount above 300, subject to a maximum payment of 400.

It is hypothesized the amount of each loss follows an exponential distribution with mean 500. Values of this distribution are simulated using an inversion (inverse transformation) method, where the random variable U has a uniform distribution on the interval (0, 1). Higher values of U correspond to higher values of the simulated loss amount.

Four values of U were generated: 0.12, 0.34, 0.56, 0.89.

Using this pseudorandom sample data, find the average amount that the *insurer* would have to pay.

Solution S4C114-1. First, we can find the raw loss amounts.

Let u be any of the four given simulated values of U. Let X denote the loss amount.

We know that $u = F(x) = 1 - e^{-x/500}$.

Thus, $e^{-x/500} = 1 - u$,

and $x = -500\ln(1 - u)$.

Thus, our simulated loss amounts are as follows:

-500*ln(1-0.12) = 63.91668575.
-500*ln(1-0.34) = 207.757722.
-500*ln(1-0.56) = 410.490276.
-500*ln(1-0.89) = 1103.637457.

Loss amounts of 63.91668575 and 207.757722 are each below the deductible of 300, so they result in payments of 0. The loss amount of 410.490276 results in a payment of

(410.490276 - 300)*0.5 = 55.245138.

The insurer will pay 50% of losses up to a total payment of 400. The total payment of 400 will occur when losses exceed the deductible of 300 by 400/0.5 = 800, so any loss amount above 1100 will result in a payment of 400. Thus, a loss amount of 1103.637457 results in a payment of 400.

The average payment is thus (0 + 0 + 55.245138 + 400)/4 = **113.8112845.**

Problem S4C114-2. Similar to Question 276 of the Exam C Sample Questions from the Society of Actuaries. You know that number of peanuts eaten by elephants follows the distribution function $F(x) = 1 - \theta/x$ for $\theta < x < \infty$.

You also know the following:

8 elephants ate fewer than 60 peanuts.
7 elephants ate between 60 and 90 peanuts.
10 elephants ate more than 90 peanuts.
Find the maximum likelihood estimate of θ.

Solution S4C114-2. Each elephant that ate fewer than 60 peanuts contributes a factor of F(60) = 1 - θ/60 to the likelihood function. Each elephant that ate between 60 and 90 peanuts contributes a factor of F(90) - F(60) = (1 - θ/90) - (1 - θ/60) = θ/60 - θ/90 = θ/180 to the likelihood function.

Each elephant that ate more than 90 peanuts contributes a factor of S(90) = θ/90 to the likelihood function.

Thus, $L(\theta) = (1 - \theta/60)^8*(\theta/180)^7*(\theta/90)^{10}$.

The loglikelihood function is $l(\theta) = \ln((1 - \theta/60)^8*(\theta/180)^7*(\theta/90)^{10}) =$

$8*\ln(1 - \theta/60) + 7*\ln(\theta/180) + 10*\ln(\theta/90)$.

$l'(\theta) = (-8/60)/(1 - \theta/60) + 7/\theta + 10/\theta = 0$ at the maximum.

Thus, $17/\theta = (2/15)/(1 - \theta/60) \rightarrow$

$2\theta = 255(1 - \theta/60) \rightarrow$

$2\theta = 255 - 255\theta/60 \rightarrow$

$6.25\theta = 255 \rightarrow$ **θ = 40.8**.

Problem S4C114-3. Similar to Question 278 of the Exam C Sample Questions from the Society of Actuaries. You know the following data:
There are n total losses.

6 of the losses are below 100.

x of the losses are between 100 and 300.

y of the losses are between 300 and 500.

30 of the losses are between 500 and 800.

There are no losses greater than 800.

Assume that losses are uniformly distributed within each of the intervals given above. From the ogive constructed from this data, you know that

$F_n(200) = 0.18$, and $F_n(350) = 0.40$. Use this information to find x.

Solution S4C114-3. We can find an expression for n: n = 6 + x + y + 30 = 36 + x + y.

Thus, $F_n(200) = (6 + (200-100)x/(300-100))/(36 + x + y) = (6 + 0.5x)/(36 + x + y) = 0.18$.

$F_n(350) = (6 + x + (350-300)y/(500-300))/(36 + x + y) = (6 + x + 0.25y)/(36 + x + y) = 0.40$.

Thus, $(36 + x + y)*0.18 = 6 + 0.5x \rightarrow$

$6.48 + 0.18x + 0.18y = 6 + 0.5x \rightarrow$

$0.18y = -0.48 + 0.32x \rightarrow$

$y = -8/3 + (16/9)x$.

Thus, $(6 + x + 0.25(-8/3 + (16/9)x))/(36 + x + (-8/3 + (16/9)x)) = 0.40 \rightarrow$

$(6 + x - 2/3 + (4/9)x)/(100/3 + (25/9)x) = 0.40 \rightarrow$

$(16/3 + (13/9)x) = 0.40(100/3 + (25/9)x) \rightarrow$

$16/3 + (13/9)x = 40/3 + (10/9)x \rightarrow$

$8 = x/3 \rightarrow$ **$x = 24$**.

Problem S4C114-4. Similar to Question 279 of the Exam C Sample Questions from the Society of Actuaries. Loss amounts (X) follow a distribution function $F(x) = (x/200)^2$ for $0 \leq x \leq 200$. An insurance company pays 50% of losses in excess of an ordinary deductible of 50, up to a maximum payment amount of 50. Given that a payment has been made, find the expected value of that payment.

Solution S4C114-4. If a payment has been made, then we know that the loss is in excess of the deductible of 50, so the *per-payment* expected value of payments is the *per-loss* expected value of payments divided by $S(50) = 1 - (50/200)^2 = 0.9375$.

Also, the company will only make additional payments for losses less than 50 + 50/0.5 = 150.

We can find the per-loss expected value of payments as follows:
The amount paid per loss is $Y = 0.5(X - 50) \wedge 150$.
$f(x) = 2x/200^2 = x/20000$.
$E(Y) = {}_{50}{}^{150}\int 0.5(x-50)*(x/20000)dx + S(150)*50 =$
${}_{50}{}^{150}\int (x^2/40000 - x/800)dx + (1-(150/200)^2)*50 =$
$(x^3/120000 - x^2/1600) \big|_{50}^{150} + 21.875 = 14.5833333 + 21.875 = 36.4583333333$.
The per-payment expected payment is 36.4583333333/0.9375 = **38.8888888889.**

Problem S4C114-5. Similar to Question 280 of the Exam C Sample Questions from the Society of Actuaries. Claim frequency (N) follows a Poisson distribution with mean $\lambda = 8$.

Claim severity (X) is either 7 with probability 0.3 or k with probability 0.7. It is known that k > 7. A policy of aggregate stop-loss insurance has a deductible of 7 on the loss amount. The expected value of a payment under this policy is 65.80234824. Find the value of k.

Solution S4C114-5. Let S be aggregate payments. We want to find an expression for $E(S) - E(S \wedge 7)$.
$E(S) = E(N)*E(X) = 8*(7*0.3 + k*0.7) = 16.8 + 5.6k$.
$E(S \wedge 7)$ is $0*\Pr(S = 0) + 7*(1 - \Pr(S = 0))$. Why is this the case? If there is one claim, then that claim is of size 7 or of size k > 7, so the next-largest value of S after 0 is 7.
$\Pr(S = 0) = \Pr(N = 0)$, i.e., the probability of no claims. $\Pr(N = 0) = e^{-8}$, so
$E(S \wedge 7) = 7(1 - e^{-8})$.
Thus, $16.8 + 5.6k - 7(1 - e^{-8}) = 65.80234824 \rightarrow 56 = 5.6k \rightarrow$ **$k = 10$**.

Section 115

Assorted Exam-Style Questions for Exam 4/C – Part 31

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Source: Exam C Sample Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C115-1. Similar to Question 281 of the Exam C Sample Questions from the Society of Actuaries. Elephants get paid interest (in peanuts) on the number of peanuts they lay aside for future consumption. The rate of interest is based on a market index return R and is 80% of R, subject to a minimum rate of 5%. R is normally distributed with mean 10% and standard deviation 20%. You also know that for a random variable X with a mean of 8% and a standard deviation of 16%, E(X Λ 5%) = -0.34%. Find the expected annual rate of interest an elephant would receive.

Solution S4C115-1. Let X = 0.8R, i.e., the portion of the interest rate that is based on R.

R is normally distributed with mean 10% and standard deviation 20%, so X is normally distributed with mean 0.8*10% = 8% and standard deviation of 0.8*20% = 16%. Thus, from the given information, we know that E(X Λ 5%) = -0.34%.

Let Q be the actual rate of interest an elephant would receive.

Q is either X = 0.8R or 5%, whichever is greater, so Q = Max(0.8R, 5%).

We want to develop an expression for Q that incorporates limited expected value.

We note that if X < 5%, X Λ 5% = X, and so X - X Λ 5% = 0.

Thus, we could say that if X < 5%, Q = (X - X Λ 5%) + 5% = 5%.

If X > 5%, X Λ 5% = 5%, and so (5% - X Λ 5%) = 0.

Thus, we could say that if X > 5%, Q = (5% - X Λ 5%) + X = X.

In each of these cases, Q = (X - X Λ 5%) + 5%.

Thus, our answer, E(Q) = E(X - X Λ 5% + 5%) = E(X) - E(X Λ 5%) + 5%.

We know that E(X) is the mean of X's normal distribution, or 8%. Thus,

E(Q) = 8% - (-0.34%) + 5% = **13.34%.**

Problem S4C115-2. Similar to Question 282 of the Exam C Sample Questions from the Society of Actuaries. Loss frequency (N) follows a Poisson distribution with mean λ = 30. Loss severity (X) follows a Burr distribution with parameters θ = 5, γ = 2, and α = 4. Let S denote aggregate losses. Find Var(S).

Relevant properties of Burr distributions:

$E(X) = \theta*\Gamma(1 + 1/\gamma)*\Gamma(\alpha - 1/\gamma)/\Gamma(\alpha)$.

$E(X^2) = \theta^2*\Gamma(1 + 2/\gamma)*\Gamma(\alpha - 2/\gamma)/\Gamma(\alpha)$.

Solution S4C115-2. Since N is Poisson-distributed, $Var(S) = E(N)*E(X^2)$.

E(N) = λ = 30.

$E(X^2) = \theta^2*\Gamma(1 + 2/\gamma)*\Gamma(\alpha - 2/\gamma)/\Gamma(\alpha) = 5^2*\Gamma(1 + 2/2)*\Gamma(4 - 2/2)/\Gamma(4) =$

25*1!*2!/3! = 25/3.

Thus, Var(S) = 30*(25/3) = **Var(S) = 250.**

Problem S4C115-3. Similar to Question 283 of the Exam C Sample Questions from the Society of Actuaries. There are 5 llamas in a zoo. Each llama eats llama food at a rate of X units per day, where X is geometrically distributed with mean 8. The amount of food eaten by each llama is independent. A llama food subsidization plan will pay $10 for every unit of llama food beyond the third unit eaten by all the llamas. Find the expected payment under this plan. Refer to the Exam 4 / C Tables as necessary.

Solution S4C115-3. Let S be the number of food units eaten by all the llamas. Then E(S) = E(N)*E(X) = 5*8 = 40.

We first want to find the expected number of food units that will be paid for, which is

E(S) - E(S Λ 3).
A geometric distribution is a special case of a negative binomial distribution, where r = 1. If we have five independent random variables, each following a geometric distribution, then the

aggregate distribution will be negative binomial with r = 5 and β = the mean of the geometric distribution = 8 in this case.

For a negative binomial distribution, $\Pr(S = k) = r(r+1)...(r+k-1)*\beta^k/(k!*(1+\beta)^{r+k})$.

Thus, $\Pr(S = 0) = 1/(1+\beta)^r = 1/9^5$.

$\Pr(S = 1) = 5*8/9^6 = 40/9^6$.

$\Pr(S = 2) = 5*6*8^2/(2*9^7) = 960/9^7$.

$\Pr(S = 3) = 5*6*7*8^3/(6*9^8) = 17920/9^8$.

$E(S \wedge 3) = 0*\Pr(S=0) + 1*\Pr(S=1) + 2*\Pr(S=2) + 3*(1-\Pr(S=0)-\Pr(S=1)-\Pr(S=2)) =$

$1/9^5 + 80/9^6 + 1920/9^7 + 3(1- 1/9^5 - 40/9^6- 960/9^7) = 2.999690151$.

Thus, $E(S) - E(S \wedge 3) = 40 - 2.999690151 = 37.00030985$ units of food subsidized, and so the expected subsidy will be 10*37.00030985 = **$370.0030985**.

Problem S4C115-4. Similar to Question 284 of the Exam C Sample Questions from the Society of Actuaries. Loss amounts follow a Poisson distribution with mean λ = 12.

An insurer offers two insurance options for policyholders with this distribution of losses:

A) An ordinary deductible of 2 is applied by an insurer to each loss.

B) There is no deductible, but a coinsurance of 100p% is applied to the loss amount. The coinsurance is proportion of the loss that the insurer will pay.

The expected per-loss loss payment is the same according to each of these options. Find the value of p.

Solution S4C115-4. First, we find the expected loss payment under option A.

This is $E(X) - E(X \wedge 2)$. We are given that $E(X) = 12$.

We find $E(X \wedge 2)$ as follows:
$E(X \wedge 2) = 0*\Pr(X = 0) + 1*\Pr(X = 1) + 2*(1-\Pr(X=0)-\Pr(X=1))$.

$\Pr(X = 0) = e^{-12}$, and $\Pr(X = 12) = 12e^{-12}$.

Thus, $E(X \wedge 2) = 12e^{-12} + 2(1-13e^{-12}) = 1.999913981$, and

$E(X) - E(X \wedge 2) = 12 - 1.999913981 = 10.00008602$, which is also equal to $p*E(X)$, so p = 10.00008602/12 = **p = 0.8333405016**.

Problem S4C115-5. Similar to Question 285 of the Exam C Sample Questions from the Society of Actuaries. There are 50 groups of archaeologists each seeking to unearth a large treasure. The probability that a group of archaeologists will arrive on site and will be able to work is 0.6. Each group of archaeologists has a mean number of 30 members, with a variance of 30. The total cost of all of the expeditions is equal to 100 times the mean number of archaeologists that actually arrive on site, plus 100 times the standard deviation of this number. Find this total cost.

Solution S4C115-5. Let N denote the number of archaeologist groups that participate, and let X denote the number of members per group. Let S denote the total (aggregate) number of participating archaeologists. We want to find 100(E(S) + SD(S)).

E(S) = E(N)*E(X).

Since a group of archaeologists can either arrive with probability 0.6 or not arrive with probability 1-0.6 = 0.4, N follows a binomial distribution with parameters m = 50 and q = 0.6.

Thus, E(N) = mq = 50*0.6 = 30, and Var(N) = mq(1-q) = 30*0.4 = 12.

Thus, E(S) = 50*30 = 1500.

Var(S) = E(N)*Var(X) + Var(N)*E(X)^2 = 30*30 + 12*30^2 = 11700, meaning that SD(S) = √(11700) = 108.1665383. Our answer is 100(1500 + 108.1665383) = **160816.6538.**

Section 116

Assorted Exam-Style Questions for Exam 4/C – Part 32

This section provides additional exam-style practice with a variety of syllabus topics.

Some of the problems in this section were designed to be similar to problems from past versions of Exam 4/C, offered jointly by the Casualty Actuarial Society and the Society of Actuaries. They use original exam questions as their inspiration - and the specific inspiration is cited to give students an opportunity to see the original. All of the original problems are publicly available, and students are encouraged to refer to them. But all of the values, names, conditions, and calculations in the problems here are the original work of Mr. Stolyarov.

Sources: Exam C Sample Questions and Solutions from the Society of Actuaries.

May 2007 Exam C Questions and Solutions from the Society of Actuaries.

Original Problems and Solutions from The Actuary's Free Study Guide

Problem S4C116-1. Similar to Question 286 of the Exam C Sample Questions from the Society of Actuaries. You know the following details about a policy of insurance:

Losses follow a Pareto distribution with parameters $\theta = 300$ and $\alpha = 2$.

The insured pays an ordinary deductible of 40.

The insured pays 50% of the loss amount between 40 and 80.

The insured pays 100% of the loss amount above 80, subject to a maximum out-of-pocket payment of 100.

After the insured has paid the out-of-pocket maximum, the insurer pays the remaining loss amount. Find the expected insurance company payment.

Relevant properties of Pareto distributions:

$E(X) = \theta/(\alpha-1)$; $E(X \wedge k) = (\theta/(\alpha-1))(1 - (\theta/(x+\theta))^{\alpha-1})$.

Solution S4C116-1.

First, we want to find the threshold beyond which the insurer will pay for all losses.

For a loss of 80, the insured will pay the deductible of 40, plus 0.5(80-40) = 20, for a total of 60. For a loss of 120, the insured will pay 40 more, for a total of 100. Thus, the insurer will pay for all losses above 120. The expected value of this payment is E(X) - E(X Λ 120).

The insurer also pays 50% of the loss amount between 40 and 80. The expected value of this payment is E(X Λ 80) - E(X Λ 40). Thus, the expected value of the total payment is

E(X) - E(X Λ 120) + E(X Λ 80) - E(X Λ 40).

We find E(X) = (300)/(2-1) = 300. We note, in general, that α-1 = 1 in this problem.

E(X Λ 120) = 300*(1 - 300/420) = 85.71428571.

E(X Λ 80) = 300*(1 - 300/380) = 63.15789474.

E(X Λ 40) = 300*(1 - 300/340) = 35.29411765.

Thus, our answer is

300 - 85.71428571 + 63.15789474 - 35.29411765 = **242.1494914.**

Problem S4C116-2. Similar to Question 287 of the Exam C Sample Questions from the Society of Actuaries. The number of claims (N) follows a negative binomial distribution with parameters r = 5 and β = 2. Claim severity (X) follows a uniform distribution on the interval from 0 to 6. Find the premium such that the normal approximation of the probability that aggregate losses exceed the premium is 0.01.

Relevant properties of negative binomial distributions: E(N) = rβ; Var(N) = rβ(1+β).

Solution S4C116-2. First we find E(N) = rβ = 5*2 = 10 and Var(N) = rβ(1+β) = 10*3 = 30.

Then we find E(X) = (6+0)/2 = 3 and Var(X) = $(6-0)^2$/12 = 3.

Let S denote aggregate losses. The E(S) = E(N)*E(X) = 10*3 = 30.

Var(S) = E(N)*Var(X) + Var(N)*$E(X)^2$ = 10*3 + 30*3^2 = Var(S) = 300, so SD(S) = √(300) = 17.32050808.

What is the z value associated with 1-Φ(z) = 0.01 → Φ(z) = 0.99. We find the answer via the Excel input "=NORMSINV(0.99)", which gives us a result of 2.326347874.

Let P denote the premium.

Thus,

2.326347874 = (P - 30)/17.32050808, and so P = 2.326347874*17.32050808 + 30 = **P = 70.29352714**.

Problem S4C116-3. Similar to Question 288 of the Exam C Sample Questions from the Society of Actuaries. Random variable N has probability p of following a binomial distribution A with parameters m = 7 and q = 0.3. N has probability (1-p) of following a binomial distribution B with parameters m = 5 and q = 0.6. Find an expression for Pr(N = 4) in terms of p. Refer to the Exam 4 / C Tables as necessary.

Solution S4C116-3. For binomial distribution A,

Pr(N = 4) would be C(7, 4)*0.3^4*0.7^3 = 0.0972405.

For binomial distribution B,

Pr(N = 4) would be C(5, 4)*0.6^4*0.4 = 0.2592.

For the mixture distribution, Pr(N = 4) is the probability-weighted average of the probabilities for the component distributions:

Pr(N = 4) = 0.0972405p + 0.2592(1-p) = **Pr(N = 4) = 0.2592 - 0.1619595p**.

Problem S4C116-4. Similar to Question 289 of the Exam C Sample Questions from the Society of Actuaries. Claim frequency (N) follows a Poisson distribution with mean λ = 3. There are three possible claim amounts (X) for each individual claim:

Pr(X = 20) = 0.4;

Pr(X = 60) = 0.5;

Pr(X = 600) = 0.1.

What is the probability of total claims being exactly 80?

Relevant properties of Poisson: Pr(N = k) = $e^{-\lambda}$*λ^k/k!

Solution S4C116-4. There are only two ways to get total claims of 80.

There can be two claims, one of which is 60, and the other of which is 20.

Pr(2 claims) = e^{-3}*3^2/2! = 0.2240418077.

Pr(One claim is 20, the other claim is 60 | 2 claims) = 2*0.4*0.5 = 0.4.

Thus, Pr(2 claims and one claim is 20, the other claim is 60) = 0.2240418077*0.4 = 0.0896167231.

The second possibility is that there are 4 claims, each of which is 20.

Pr(4 claims) = $e^{-3}*3^4/4!$ = 0.1680313557.

Pr(4 claims are 20 │4 claims) = 0.4^4 = 0.0256.

Pr(4 claims and 4 claims are 20) = 0.1680313557*0.0256 = 0.0043016027.

Thus, our answer is 0.0896167231 + 0.0043016027 = **0.0939183258.**

Problem S4C116-5. Similar to Question 3 of the May 2007 Exam 4/C.

Conditional on Q = q, m independent and identically distributed random variables are follow a Bernoulli distribution with parameter q.

The prior distribution of Qis beta with parameters a = 2, b = 40, and θ = 2.

Let S_{56} be the random variable denoting the sum of 56 independent and identically distributed random variables as described above.

Determine the variance of the marginal distribution of S_{56}.

Relevant properties of beta distributions: E(X) = aθ/(a+b); $E(X^2) = a(a+1)\theta^2/((a+b)(a+b+1))$.

Relevant properties of binomial distributions: E(N) = mq; Var(N) = mq(1-q).

Solution S4C116-5. The sum of m Bernoulli random variables, each with parameter q, is a binomial random variable with parameters m and q. Thus, the conditional distribution of S_{56} is binomial with parameters m = 56 and q.

$Var(S_{56}) = E_Q(Var(S_{56} | q)) + Var_Q(E(S_{56} | q))$.
$E(S_{56} | q) = 56q$ and $Var(S_{56} | q) = 56q(1-q) = 56q - 56q^2$.
Thus, $Var(S_{56}) = E_Q(56q - 56q^2) + Var_Q(56q)$ =
$56E(Q) - 56E(Q^2) + 3136Var(Q)$.
Since Q is beta-distributed, we find E(Q) = aθ/(a+b) = 2*2/42 = 2/21.
We find $E(Q^2) = a(a+1)\theta^2/((a+b)(a+b+1))$ = 2*3*4/(42*43) = 4/301.
Thus, $Var(Q) = E(Q^2) - E(Q)^2 = 4/301 - (2/21)^2$ = 0.0042187418.
Our answer is therefore 56*(2/21) - 56*(4/301) + 3136*(0.0042187418) =
$Var(S_{56})$ = 17.81912145.

About Mr. Stolyarov

Gennady Stolyarov II (G. Stolyarov II) is an actuary, science-fiction novelist, independent philosophical essayist, poet, amateur mathematician, composer, and Editor-in-Chief of The Rational Argumentator, a magazine championing the principles of reason, rights, and progress.

In December 2013, Mr. Stolyarov published *Death is Wrong*, an ambitious children's book on life extension illustrated by his wife Wendy. *Death is Wrong* can be found on Amazon in paperback and Kindle formats.

Mr. Stolyarov has contributed articles to the Institute for Ethics and Emerging Technologies (IEET), The Wave Chronicle, Le Quebecois Libre, Brighter Brains Institute, Immortal Life, Enter Stage Right, Rebirth of Reason, The Liberal Institute, and the Ludwig von Mises Institute. Mr. Stolyarov also published his articles on Associated Content (subsequently the Yahoo! Contributor Network) from 2007 until its closure in 2014, in an effort to assist the spread of rational ideas. He held the highest Clout Level (10) possible on the Yahoo! Contributor Network and was one of its Page View Millionaires, with over 3.1 million views.

Mr. Stolyarov holds the professional insurance designations of Associate of the Society of Actuaries (ASA), Associate of the Casualty Actuarial Society (ACAS), Member of the American Academy of Actuaries (MAAA), Chartered Property Casualty Underwriter (CPCU), Associate in Reinsurance (ARe), Associate in Regulation and Compliance (ARC), Associate in Personal Insurance (API), Associate in Insurance Services (AIS), Accredited Insurance Examiner (AIE), and Associate in Insurance Accounting and Finance (AIAF).

Mr. Stolyarov has written a science fiction novel, *Eden against the Colossus*, a philosophical treatise, *A Rational Cosmology*, a play, *Implied Consent*, and a free self-help treatise, *The Best Self-Help is Free*. You can watch his YouTube Videos. Mr. Stolyarov can be contacted at gennadystolyarovii@gmail.com.