2

Let the moments of the primary distribution be μ_1, μ_2, μ_3 (and similar notation for raw moments). Let the moments of the secondary distribution be ν_1, ν_2, ν_3 (and similar notation for raw moments).

Recall that $P(z) = M(logz)$, so $P'(z) = \frac{M'(log(z))}{z}$, $P''(z) = \frac{M''(log(z))-M'(log(z))}{z^2}$, and $P'''(z) = \frac{M'''(log(z))-3M''(log(z))+2M'(log(z))}{z^3}$ In particular, $P'(1) = \mu$, $P''(1) = \mu_2' - \mu$ and $P'''(1) = \mu_3' - 3\mu_2' + 2\mu$.

m.g.f. of compound model is $P(M(z))$ first 3 derivatives of this at 0 are:

$$M'(0)P'(M(0)) = M'(0)P'(1) = \mu\nu$$
$$M''(0)P'(1) + M'(0)^2 P''(1) = \mu\nu_2' + (\mu_2' - \mu)\nu^2$$
$$M'''(0)P'(1) + 3M''(0)M'(0)P''(1) + M'(0)^3 P'''(1) = \mu\nu_3' + 3(\mu_2' - \mu)\nu\nu_2' + (\mu_3' - 3\mu_2' + 2\mu)\nu^3$$

3

For a given claim, the amount reimbursed has mean
$1000 + 0.8 \times 500 = 1400$, and variance $500^2 + 0.8^2 \times 300^2 + 2 \times 100,000 = 507,600$.
The mean of the aggregate claims is therefore: $4 \times 1400 = 5600$. The raw second moment is

$$4 \times (507600 + 1400^2) + (20 - 4) \times 1400^2 = 2,030,400 + 7,840,000 + 31,360,000 = 41,230,400$$

The variance is this minus 5600^2, which is $9,870,400$. The standard deviation is the square root of this or $3,141.72$.

4

We have

n	$P(N = n)$	Z	$P(A > 130 \mid N = n)$
0	0.4	∞	0
1	0.3	0.8571	0.8051
2	0.2	-1.414	0.0793
3	0.1	-2.804	0.0026

So probability$=0 + 0.3 \times 0.8051 + 0.2 \times 0.0793 + 0.1 \times 0.0026 = 0.24153 + 0.01586 + 0.00026 = 0.2577$.

5

mean$=4 \times 6 \times 16 = 384$. Variance$=\mu\nu_2 + \mu_2\nu^2 = 6 \times 16 \times \frac{8^2}{12} + 4^2 \times 16 \times 6 \times 7 = 512 + 512 \times 21$ The standard deviation is therefore $32\sqrt{11}$. 95th percentile is 1.645 standard deviations above the mean or $384 + 52.64\sqrt{11} = 558.59$.

6

Prob of stop-loss is $e^{-1.25}$. Expected stop-loss claim conditional on claim is θ, so expected stop-loss claim$=e^{-1.25}\theta$. Premium is $2e^{-1.25}\theta$.

in fact value 0.9θ was used instead of theta, so premium is $1.8e^{-1.25}\theta$, and stop loss is really set at $1.25 \times 0.9\theta = 1.125\theta$, so expected payment on stop-loss is $e^{-1.125}\theta$. Percentage loading is therefore $\frac{1.8e^{-1.25}\theta}{e^{-1.125}\theta} - 1 = 1.8e^{-0.125} - 1 = 1.588494 - 1 = 58.85\%$.

9.4 Analytic Results

7

The severity is exponential with mean θ. The frequency is negative binomial with parameters r and β. The aggregate severity of n losses therefore follows a gamma distribution with $\alpha = n$. We therefore have that the probability that the aggregate loss is zero is $p_0 = \left(\frac{1}{1+\beta}\right)^r$, while if it is non-zero, the pdf of the aggregate loss is

$$f(x) = \sum_{n=1}^{\infty} \frac{p_n}{1-p_0} \frac{x^{n-1}e^{-\frac{x}{\theta}}}{\theta^n (n-1)!} = \frac{\left(\frac{1}{1+\beta}\right)^r}{\theta\left(1-\left(\frac{1}{1+\beta}\right)^r\right)} e^{-\frac{x}{\theta}} \sum_{m=0}^{\infty} \binom{r+m}{m+1} \left(\frac{\beta}{1+\beta}\right)^{m+1} \frac{x^m}{\theta^m m!}$$

If r is a positive integer, then

$$\binom{m+r}{m+1} = \frac{(m+2)(m+3)\cdots(m+r)}{r!}$$

Furthermore, a sum of the form

$$\sum_{m=0}^{\infty} (m+2)(m+3)\cdots(m+r)\frac{a^m}{m!}$$

is the rth derivative of

$$\sum_{m=0}^{\infty} \frac{a^{m+r}}{m!} = a^r e^a$$

This derivative is

$$\left(\sum_{k=0}^{r} \binom{r}{k}^2 k! a^{r-k}\right) e^a$$

Therefore, the pdf of the aggregate loss is

$$\begin{aligned} f(x) &= \frac{\beta\left(\frac{1}{1+\beta}\right)^{r+1}}{\theta\left(1-\left(\frac{1}{1+\beta}\right)^r\right)} e^{-\frac{x}{\theta}} \sum_{k=0}^{r} \binom{r}{k}^2 k! \left(\frac{\beta x}{\theta(1+\beta)}\right)^{r-k} e^{\left(\frac{\beta x}{\theta(1+\beta)}\right)} \\ &= \frac{\beta}{\theta\left((1+\beta)^{r+1}-(1+\beta)\right)} e^{-\frac{x}{\theta(1+\beta)}} \sum_{k=0}^{r} \binom{r}{k}^2 k! \left(\frac{\beta x}{\theta(1+\beta)}\right)^{r-k} \end{aligned}$$

8

If there are n claims, the aggregate loss follows a gamma distribution with $\alpha = n$ and $\theta = 3000$. The expected payment on the stop-loss insurance is then $3000 \int_{\frac{500000}{3000}}^{\infty} \left(x - \frac{500000}{3000}\right) \frac{x^{n-1}e^{-x}}{(n-1)!} dx$.

We have

$$\int_a^{\infty} \frac{x^n e^{-x}}{n!} dx = e^{-a}\left(1 + a + \frac{a^2}{2} + \ldots + \frac{a^n}{n!}\right) e^{-a}$$

so the expected payment on the stop-loss insurance if there are n claims is

$$3000 e^{-\frac{500000}{3000}} \left(\left(n - \frac{500000}{3000}\right)\left(1 + a + \frac{a^2}{2} + \ldots + \frac{a^{n-1}}{(n-1)!}\right) + n\frac{a^n}{n!}\right)$$

The overall expected payment on the stop-loss insurance is therefore

$$3000e^{-a}\sum_{n=0}^{\infty}p_n\left(n\sum_{m=0}^{n}\frac{a^m}{m!}-a\sum_{m=0}^{n}\frac{a^m}{m!}\right)=3000e^{-a}\sum_{m=0}^{\infty}\sum_{m=0}^{n}p_n(n-m)\frac{a^m}{m!}$$
$$=3000e^{-a}\sum_{m=0}^{\infty}\frac{a^m}{m!}\sum_{n=m}^{\infty}p_n(n-m)$$

Now

$$p_n=\binom{n+14}{n}\left(\frac{12}{17}\right)^n\left(\frac{5}{17}\right)^{15}$$

So

$$\sum_{n=m}^{\infty}p_n(n-m)=\left(\frac{5}{17}\right)^{15}\sum_{n=m}^{\infty}(n-m)\binom{n+14}{n}\left(\frac{12}{17}\right)^n$$

9

For the first company, the pdf of the aggregate claims at values more than 0 is

$$e^{-0.4}\sum_{n=1}^{\infty}\frac{0.4^n}{n!(3n)!}\frac{x^{3n-1}}{30000^3}e^{-\frac{x}{30000}}=\frac{e^{-0.4}}{x}(e^{0.4\left(\frac{x}{30000}\right)^3}-1)e^{-\frac{x}{30000}}$$

For the second company, the pdf of the aggregate claims at values more than 0 is

$$e^{-3.6}\sum_{n=1}^{\infty}\frac{3.6^n}{n!(1.4n)!}\frac{x^{1.4n-1}}{200000^{1.4}}e^{-\frac{x}{200000}}=\frac{e^{-3.6}}{x}e^{3.6\left(\frac{x}{200000}\right)^{1.4}}e^{-\frac{x}{200000}}$$

For the third company, the pdf of the aggregate claims at values more than 0 is

$$e^{-85}\sum_{n=1}^{\infty}\frac{85^n}{n!(2.2n)!}\frac{x^{2.2n-1}}{45000^{2.2}}e^{-\frac{x}{45000}}=\frac{e^{-85}}{x}e^{85\left(\frac{x}{45000}\right)^{2.2}}e^{-\frac{x}{45000}}$$

9.5 Computing the Aggregate Claims Distribution

10

ETNB $r=-0.6$ $\beta=7$,
$a=\frac{\beta}{1+\beta}=\frac{7}{8}$, $b=\frac{(r-1)\beta}{1+\beta}=-1.4$
$q_1=\frac{-0.6\times 7}{8(8^{-0.6}-1)}=0.7365057$ $q_2=\left(\frac{7}{8}-\frac{1.4}{2}\right)q_1=0.1288885$ $q_3=\left(\frac{7}{8}-\frac{1.4}{3}\right)q_2=0.05262947$
$(n+1)\left(\frac{3}{4}\right)^n$ 0.0625 0.0625 0.09375 0.1054687890625 0.1054687890625

$p_0=0.0625p_1 \quad =0.09375*0.7365057=0.06904741p_2=0.09375*0.1288885+0.1054687890625*0.7365057^2=0.06929383p_3 \quad =0$

So the total probability that the aggregate loss is at most 3 is $0.0625+0.06904741+0.06929383+0.06709359=0.2679348$.

9.6 The Recursive Method

11

The recurrence formula gives $f_S(n) = \frac{\sum_{i=1}^{n}\left(a+\frac{bi}{n}\right)f_X(i)f_S(n-i)}{1-af_X(0)}$ $f_S(n) = \sum_{i=1}^{n}\frac{2.4i}{n}\binom{i+9}{i}\left(\frac{2.3}{3.3}\right)^i\left(\frac{1}{3.3}\right)^{10}f_S(n-i)$
$f_X(0) = \frac{1}{3.3^{10}}$

$$f_S(0) = e^{-2.4}\sum_{n=0}^{\infty}\frac{2.4^n}{3.3^{10n}n!} = e^{\frac{2.4}{3.3^{10}}-2.4} = 0.09071937$$

$$f_S(1) = \frac{2.4}{3.3^{10}}\left(10\times\frac{2.3}{3.3}\times f_S(0)\right) = 0.000009907995$$

$$f_S(2) = \frac{2.4}{3.3^{10}}\left(\frac{10}{2}\times\frac{2.3}{3.3}\times f_S(1)+\frac{10\times 11}{2}\times\left(\frac{2.3}{3.3}\right)^2\times f_S(0)\right) = 0.00003798119$$

$$f_S(3) = \frac{2.4}{3.3^{10}}\left(\frac{10}{3}\times\frac{2.3}{3.3}\times f_S(2)+\frac{2\times 10\times 11}{3\times 2}\times\left(\frac{2.3}{3.3}\right)^2\times f_S(1)+\frac{10\times 11\times 12}{6}\times\left(\frac{2.3}{3.3}\right)^3\times f_S(0)\right) = 0.0001058901$$

So the probability that the aggregate loss is at most 3 is therefore $0.09071937+0.000009907995+0.00003798119+0.0001058901 = 0.09087315$

12

For the zero-truncated ETNB distribution, we have that
$a = \frac{\beta}{1+\beta} = \frac{3}{4}$, $b = \frac{(r-1)\beta}{1+\beta} = -1.2$

$$
\begin{aligned}
q_1 &= \frac{-0.6 \times 3}{4(4^{-0.6} - 1)} = 0.7968484 \\
q_2 &= \left(\frac{3}{4} - \frac{1.2}{2}\right) 0.7968484 = 0.119527260 \\
q_3 &= \left(\frac{3}{4} - \frac{1.2}{3}\right) 0.119527260 = 0.04183454100 \\
q_4 &= \left(\frac{3}{4} - \frac{1.2}{4}\right) 0.04183454100 = 0.018825543450 \\
q_5 &= \left(\frac{3}{4} - \frac{1.2}{5}\right) 0.018825543450 = 0.009601027159 \\
q_6 &= \left(\frac{3}{4} - \frac{1.2}{6}\right) 0.009601027159 = 0.005280564937 \\
q_7 &= \left(\frac{3}{4} - \frac{1.2}{7}\right) 0.005280564937 = 0.003055183999 \\
q_8 &= \left(\frac{3}{4} - \frac{1.2}{8}\right) 0.003055183999 = 0.001833110399 \\
q_9 &= \left(\frac{3}{4} - \frac{1.2}{9}\right) 0.001833110399 = 0.0011304180793 \\
q_{10} &= \left(\frac{3}{4} - \frac{1.2}{10}\right) 0.0011304180793 = 0.0007121633899 \\
q_{11} &= \left(\frac{3}{4} - \frac{1.2}{11}\right) 0.0007121633899 = 0.0004564319907995 \\
q_{12} &= \left(\frac{3}{4} - \frac{1.2}{12}\right) 0.0004564319907995 = 0.0002966807940196 \\
q_{13} &= \left(\frac{3}{4} - \frac{1.2}{13}\right) 0.0002966807940196 = 0.0001951246760669
\end{aligned}
$$

With the deductible set at 10, the probability that a loss does not lead to a claim is $0.7968484 + 0.119527260 + 0.04183454100 + 0.018825543450 + 0.009601027159 + 0.005280564937 + 0.003055183999 + 0.001833110399 + 0.0011304180793 = 0.9981311736993669$ The distribution of the claim value is therefore

$$
\begin{aligned}
q_0 &= 0.9981311736993669 \\
q_1 &= 0.0004564319907995 \\
q_2 &= 0.0002966807940196 \\
q_3 &= 0.0001951246760669
\end{aligned}
$$

For the primary distribution $a = \frac{\beta}{1+\beta} = \frac{3}{8}$ and $b = \frac{(r-1)\beta}{1+\beta} = -0.3$.

Now we can use the recursive formula

$$f_S(n) = \frac{\sum_{i=1}^{n} \left(0.375 - \frac{0.3i}{n}\right) q_i f_S(n-i)}{1 - \frac{3}{8} \times 0.99813}$$

We calculate

$$\begin{aligned} f_S(0) &= \sum_{n=0}^{\infty} p_n (f_X(0))^n = \sum_{n=0}^{\infty} \binom{n-0.8}{n} \left(\frac{3}{8}\right)^n \left(\frac{5}{8}\right)^{0.2} (f_X(0))^n \\ &= \left(\frac{5}{8}\right)^{0.2} \left(1 - \frac{3}{8} f_X(0)\right)^{-0.2} = 0.9997759 \end{aligned}$$

Now using the recurrence, we get

$$\begin{aligned} f_S(1) &= \frac{0.075 \times 0.000456 \times 0.9998}{0.6257008} = 0.00005469823 \\ f_S(2) &= \frac{0.225 \times 0.000456 \times 0.0000547 + 0.075 \times 0.000297 \times 0.9998}{0.6257008} = 0.00003556283 \\ f_S(3) &= \frac{0.275 \times 0.000456 \times 0.0000356 + 0.175 \times 0.000297 \times 0.0000547 + 0.075 \times 0.000195 \times 0.9998}{0.6257008} = 0.00002339517 \end{aligned}$$

The probability of paying out at least \$400 to a single driver is therefore $1 - 0.9997759 - 0.00005469823 - 0.00003556283 - 0.00002339517 = 0.0001104438$

9.6.2 Applications to Compound Frequency Models

13

The zero-truncated logarithmic distribution has $a = 0.8$, $b = -0.8$. This gives $p_1 \left(\sum_{n=0}^{\infty} \frac{0.8^n}{n+1}\right) = 1$ $p_1 = -\frac{0.8}{\log(1-0.8)} = 0.4970679$.

$$\begin{aligned} p_1 &= 0.4970679 \\ p_2 &= 0.1988272 \\ p_3 &= 0.1060412 \end{aligned}$$

Now we compound with a Poisson with $\lambda = 0.1$. The probability of the total being 0 is $e^{-0.1}$ (since the secondary distribution is zero-truncated). The recurrence is

$$f_S(n) = \sum_{i=1}^{n} \frac{0.1i}{n} p_i f_S(n-i)$$

So we calculate:

$$f_S(1) = 0.1 \times 0.497 \times e^{-0.1} = 0.04970679e^{-0.1}$$
$$f_S(2) = 0.05 \times 0.497 \times 0.04970679e^{-0.1} + 0.1 \times 0.199e^{-0.1} = 0.0211181e^{-0.1}$$
$$f_S(3) = (0.0333 \times 0.497 \times 0.0211 + 0.0667 \times 0.199 \times 0.0497 + 0.1 \times 0.106)\, e^{-0.1} = 0.010705e^{-0.1}$$

Now for the overall compound distribution, we have $f_A(0) = e^{-6} \sum_{n=0}^{\infty} \frac{6^n}{n!} e^{-0.1n} = e^{6e^{-0.1}-6} = e^{-0.5709755} = 0.564974$.

The recurrence is

$$f_A(n) = \sum_{i=1}^{n} \frac{6i}{n} f_S(i) f_A(n-i)$$

So we calculate:

$$f_A(1) = 6 \times 0.0497e^{-0.1} \times 0.564974 = 0.1524635$$
$$f_A(2) = 3 \times 0.04970679e^{-0.1} \times 0.1524635 + 6 \times 0.0211181e^{-0.1} \times 0.564974 = 0.08534651$$
$$f_A(3) = 2 \times 0.04970679e^{-0.1} \times 0.08534651 + 4 \times 0.0211181e^{-0.1} \times 0.1524635 + 6 \times 0.010705e^{-0.1} \times 0.564974 = 0.05216554$$

So the probability that the total claimed is more than 3000 is

$$1 - 0.564974 - 0.1524635 - 0.08534651 - 0.05216554 = 0.1450505$$

9.6.2 Underflow Problems

14

The recurrence is

$$f_S(n) = \sum_{i=1}^{n} \frac{\lambda i}{n} \frac{\binom{n+3}{n} 0.6875^i 0.3125^4}{1 - 0.3125^4} f_S(n-i)$$

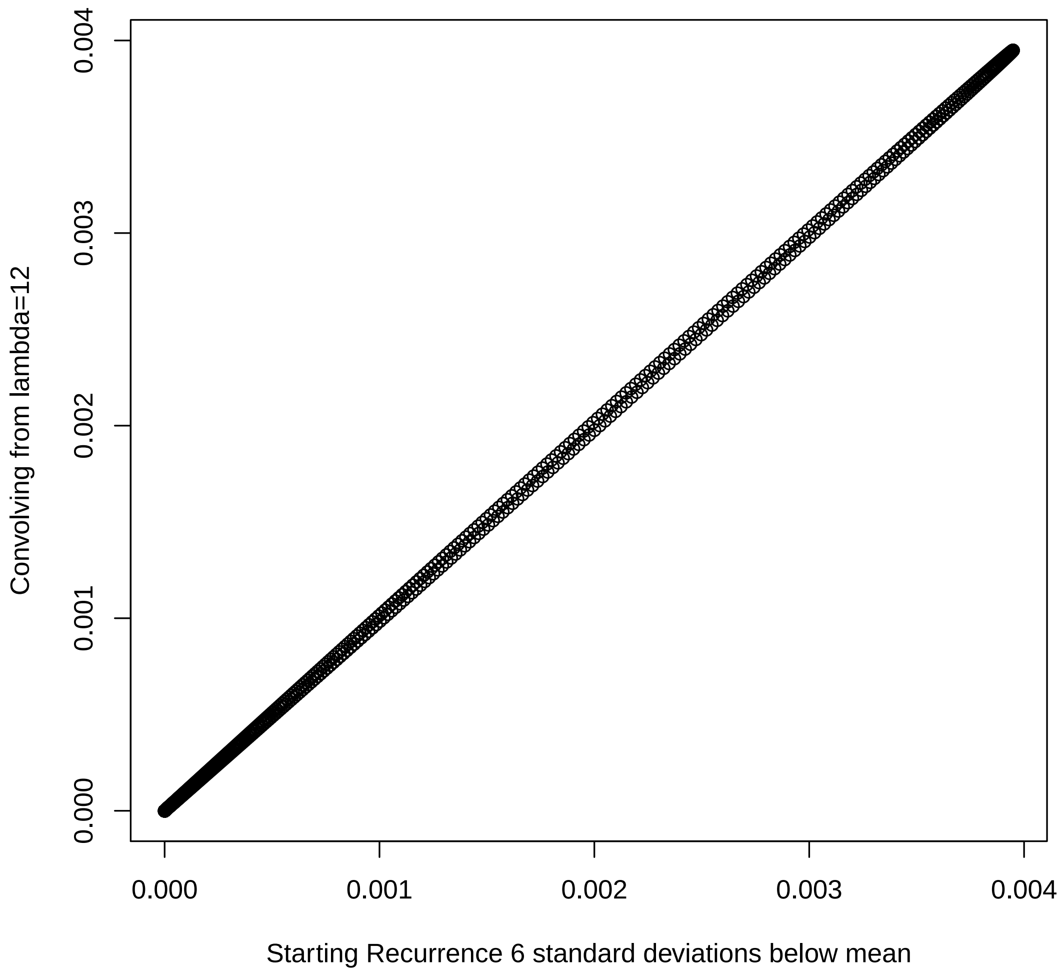
(a)

The mean of the distribution is $96 \times 8.8 = 844.8$, and the variance is $96 \times 28.16 + 96 \times 8.8^2 = 48 \times (28.16 + 77.44) = 96 \times 105.6 = 10137.6$ The standard deviation is therefore $\sqrt{10137.6} = 100.6856$, so six standard deviations below the mean is $422.4 - 6 \times 100.6856 = 240.6864$. We will start the recurrence at 241. If we assume that $f_s(240) = 0$ and $f_2(241) = 1$, then we can calculate the values

$$f_S(n) = \sum_{i=1}^{n} \frac{96i}{n} \frac{\frac{(i+1)(i+2)(i+3)}{6} 0.6875^i 0.3125^4}{1 - 0.3125^4} f_S(n-i)$$

(b) Solution for $\lambda = 12$:

$$f_S(n) = \sum_{i=1}^{n} \frac{96i}{n} \frac{\frac{(i+1)(i+2)(i+3)}{6} 0.6875^i 0.3125^4}{1 - 0.3125^4} f_S(n-i)$$



9.6.3 Numerical Stability

9.6.4 Continuous Severity

9.6.5 Constructing Arithmetic Distributions

15

(a) Using the method of rounding, we have

$p_0 = 1 - e^{-\frac{1}{2\theta}}$ $p_n = e^{-\frac{2n-1}{2\theta}} \left((1 - e^{-\frac{1}{\theta}} \right)$

This is a zero modified geometric distribution.

(b) On the interval $[a,b]$, the moments of the exponential distribution are given by $\mu = a + \frac{\int_0^{b-a} \frac{x}{\theta} e^{-\frac{x}{\theta}} dx}{1-e^{-\frac{b-a}{\theta}}}$. We have that $\int_0^{b-a} \frac{x}{\theta} e^{-\frac{x}{\theta}} dx = [-xe^{-\frac{x}{\theta}}]_0^{b-a} + \int_0^{b-a} e^{-\frac{x}{\theta}} dx = \theta - (b-a+\theta)e^{-\frac{b-a}{\theta}}$. So $\mu = a + \theta - (b-a+\theta)e^{-\frac{b-a}{\theta}}$.

Similarly,

$$
\begin{aligned}
\mu_2' &= \frac{\int_0^{b-a} \frac{(x+a)^2}{\theta} e^{-\frac{x}{\theta}}\, dx}{1-e^{-\frac{b-a}{\theta}}} \\
&= \frac{\left[-(x+a)^2 e^{-\frac{x}{\theta}}\right]_0^{b-a} + \int_0^{b-a} 2(x+a)e^{-\frac{x}{\theta}}\, dx}{1-e^{-\frac{b-a}{\theta}}} \\
&= \frac{a^2 - b^2 e^{-\frac{b-a}{\theta}} + \left[-2\theta(x+a)e^{-\frac{x}{\theta}}\right]_0^{b-a} + \int_0^{b-a} 2\theta e^{-\frac{x}{\theta}}\, dx}{1-e^{-\frac{b-a}{\theta}}} \\
&= \frac{a^2 - b^2 e^{-\frac{b-a}{\theta}} + 2\theta a - 2\theta b e^{-\frac{b-a}{\theta}} + \left[2\theta^2 e^{-\frac{x}{\theta}}\right]_0^{b-a}}{1-e^{-\frac{b-a}{\theta}}} \\
&= \frac{2\theta^2 + 2\theta a + a^2 - (2\theta^2 + 2\theta b + b^2)e^{-\frac{b-a}{\theta}}}{1-e^{-\frac{b-a}{\theta}}} \\
&= \theta^2 + \frac{(\theta+a)^2 - (\theta+b)^2 e^{-\frac{b-a}{\theta}}}{1-e^{-\frac{b-a}{\theta}}} \\
&= \theta^2 + (\theta+a)^2 + \frac{(b-a)(2\theta+a+b)e^{-\frac{b-a}{\theta}}}{1-e^{-\frac{b-a}{\theta}}}
\end{aligned}
$$

Now we want to set the values p_n, p_{n+1}, p_{n+2} to match these moments on the interval $\left[n-\frac{1}{2}, n+2+\frac{1}{2}\right]$. The moments of a distribution with probabilities p_1, p_2, p_3 on the points $\frac{1}{2}$, $\frac{3}{2}$, and $\frac{5}{2}$ are

$$
\begin{aligned}
\mu &= \frac{p_1 + 3p_2 + 5p_3}{2} \\
\mu_2' &= \frac{p_1 + 9p_2 + 25p_3}{4}
\end{aligned}
$$

Substituting the values gives:

$$
\begin{aligned}
\frac{p_1 + 3p_2 + 5p_3}{2} &= \frac{1}{2} + \frac{\theta - (3+\theta)e^{-\frac{3}{\theta}}}{1-e^{-\frac{3}{\theta}}} \\
\frac{p_1 + 9p_2 + 25p_3}{4} &= \theta^2 + \left(\theta + \frac{1}{2}\right)^2 + \frac{3(2\theta+4)e^{-\frac{3}{\theta}}}{1-e^{-\frac{3}{\theta}}} \\
p_1 + p_2 + p_3 &= 1
\end{aligned}
$$

$$
\begin{aligned}
p_2 + 2p_3 &= \frac{\theta - (3+\theta)e^{-\frac{3}{\theta}}}{1 - e^{-\frac{3}{\theta}}} \\
2p_2 + 6p_3 &= 2\theta^2 + \theta + \frac{3(2\theta+4)e^{-\frac{3}{\theta}}}{1 - e^{-\frac{3}{\theta}}} \\
2p_3 &= 2\theta^2 + \theta + \frac{3(2\theta+4)e^{-\frac{3}{\theta}}}{1 - e^{-\frac{3}{\theta}}} - 2\frac{\theta - (3+\theta)e^{-\frac{3}{\theta}}}{1 - e^{-\frac{3}{\theta}}} \\
p_3 &= \theta^2 - \frac{\theta}{2} + 3(\theta+3)\frac{e^{-\frac{3}{\theta}}}{1 - e^{-\frac{3}{\theta}}} \\
p_2 &= \theta - 3\frac{e^{-\frac{3}{\theta}}}{1 - e^{-\frac{3}{\theta}}} - \left(2\theta^2 - \theta + 6(\theta+3)\frac{e^{-\frac{3}{\theta}}}{1 - e^{-\frac{3}{\theta}}}\right) \\
&= 2\theta(1-\theta) - (6\theta+21)\frac{e^{-\frac{3}{\theta}}}{1 - e^{-\frac{3}{\theta}}} \\
p_1 &= 1 - \left(2\theta(1-\theta) - (6\theta+21)\frac{e^{-\frac{3}{\theta}}}{1 - e^{-\frac{3}{\theta}}}\right) - \left(\theta^2 - \frac{\theta}{2} + 3(\theta+3)\frac{e^{-\frac{3}{\theta}}}{1 - e^{-\frac{3}{\theta}}}\right) \\
&= \theta^2 - \frac{3}{2}\theta + 1 + 3(\theta+4)\frac{e^{-\frac{3}{\theta}}}{1 - e^{-\frac{3}{\theta}}}
\end{aligned}
$$

9.7 Individual Policy Modifications

16

Conditional on n losses, aggregate loss follows gamma distribution with $\theta = 2000$ and $r = n$.
(a) The density of the aggregate loss at non-zero values is therefore given by

$$
f(x) = \sum_{n=1}^{\infty} \frac{(n+1)(n+2)(n+3)}{6} \left(\frac{2.1}{3.1}\right)^n \left(\frac{1}{3.1}\right)^4 \frac{x^{n-1}e^{-\frac{x}{2000}}}{2000^n(n-1)!}
$$

Now we note that $(n+1)(n+2)(n+3) = (n-1)(n-2)(n-3) + 12(n-1)(n-2) + 36(n-1) + 18$, so

$$
\begin{aligned}
f(x) &= \frac{2.1e^{-\frac{x}{2000}}}{12000 \times 3.1^5}\left(\sum_{n=4}^{\infty}\frac{\left(\frac{2.1x}{2000\times 3.1}\right)^{n-1}}{(n-4)!} + 12\sum_{n=3}^{\infty}\frac{\left(\frac{2.1x}{2000\times 3.1}\right)^{n-1}}{(n-3)!} + 36\sum_{n=2}^{\infty}\frac{\left(\frac{2.1x}{2000\times 3.1}\right)^{n-1}}{(n-2)!} + 18\sum_{n=1}^{\infty}\frac{\left(\frac{2.1x}{2000\times 3.1}\right)^{n-1}}{(n-1)!}\right) \\
&= \frac{2.1e^{-\frac{x}{2000}}}{12000 \times 3.1^5}\left(\left(\frac{2.1x}{6200}\right)^3 + 12\left(\frac{2.1x}{6200}\right)^2 + 36\left(\frac{2.1x}{6200}\right) + 18\right)e^{\left(\frac{2.1x}{6200}\right)} \\
&= \frac{2.1}{12000 \times 3.1^5}\left(\left(\frac{2.1x}{6200}\right)^3 + 12\left(\frac{2.1x}{6200}\right)^2 + 36\left(\frac{2.1x}{6200}\right) + 18\right)e^{-\frac{x}{6200}}
\end{aligned}
$$

The probability mass at zero is $\frac{1}{3.1^4}$.

(b) With a deductible of \$500, the distribution per loss has probability mass $1 - e^{-0.25}$ at zero. The distribution of the number of claims is negative binomial with $r = 4$ and $\beta = 2.1e^{-0.25}$. The probability mass at zero is therefore $\frac{1}{(1+2.1e^{-0.25})^4}$. If there is a claim, by the memoryless property of the exponential distribution, severity has the same distribution as in part (a), so the aggregate loss has density function

$$f(x) = \left(\frac{2.1e^{-0.25}}{12000 \times (1 + 2.1e^{-0.25})^5}\right) e^{-\frac{x}{2000(1+2.1e^{-0.25})}}$$
$$\left(\left(\frac{2.1e^{-0.25}x}{2000(1 + 2.1e^{-0.25})}\right)^3 + 12\left(\frac{2.1e^{-0.25}x}{2000(1 + 2.1e^{-0.25})}\right)^2 + 36\left(\frac{2.1e^{-0.25}x}{2000(1 + 2.1e^{-0.25})}\right) + 18\right)$$

17

Aggregate loss distribution conditional on n losses is gamma with $\alpha = 2n$ and $\theta = 2000$. The continuous part of the aggregate loss distribution is therefore

$$f(x) = e^{-20} \sum_{n=1}^{\infty} \frac{20^n}{n!} \frac{x^{2n-1}e^{-\frac{x}{2000}}}{2000^{2n}\Gamma(2n)}$$

Now we have $\sum_{n=1}^{\infty} \frac{a^{2n-1}}{(2n-1)!} = \frac{e^a - e^{-a}}{2}$,

$$\sum_{n=1}^{\infty} \frac{a^{2n-1}}{n!(2n-1)!}$$

so the continuous part of the aggregate loss distribution is

$$f(x) = \frac{e^{-20}e^{-\frac{x}{2000}}}{2000} (e\ )$$

For a gamma distribution with $\alpha = 3.4n$, the expected excess loss is

$$\int_{1000000}^{\infty} (x - 1000000) \frac{x^{3.4n-1}e^{-\frac{x}{2000}}}{2000^{3.4n}\Gamma(3.4n)} dx = 2000 \int_{500}^{\infty} (x - 500) \frac{x^{3.4n-1}e^{-x}}{\Gamma(3.4n)} dx$$

Integrating by parts gives

$$\int_{500}^{\infty} x^{3.4n} e^{-x} dx = [-x^{3.4n}e^{-x}]_{500}^{\infty} + \int_{500}^{\infty} 3.4n x^{3.4n-1} e^{-x} dx$$

so the expected excess loss is

$$\frac{2000}{\Gamma(3.4n)} \left(500^{3.4n}e^{-500} + (3.4n - 500) \int_{500}^{\infty} x^{3.4n-1} e^{-x} dx\right)$$

9.8 Individual Risk Model

18

For the first 2 types of worker, the total benefits can be well approximated by a normal distribution with mean $\$5,259,200$ and variance $4622000 \times 99000 + 637200 \times 89820 = 514,811,304,000$, which gives a standard deviation of 717503.5.

The number of managers dying can be well approximated by a Poisson distribution with $\lambda = 8.02$ and the number of Senior managers dying is a binomial distribution.

If we approximate the number of deaths of managers by a normal with mean and variance 8.02, then the aggregate losses of managers have mean $1,604,000, and variance 320,800,000,000, so the aggregate from the first three types of employee has mean $6,863,200 and variance 835,611,304,000, so standard deviation is $914,117.773593753071

We can consider the various cases in a table

Senior Managers	Probability	Z-statistic	Probability aggregate more than 10,000,000	P
0	0.4832131282	3.4315053	0.0003001207	0.0001450223
1	0.3550137268	2.3375544	0.0097051886	0.0034454752
2	0.1267906167	1.2436034	0.1068227763	0.0135441257
3	0.0293257209	0.1496525	0.4405193971	0.0129185489
4	0.0049374938	-0.9442985	0.8274914220	0.0040857338
5	0.0006448972	-2.0382494	0.9792375011	0.0006315075
6	0.0000679994	-3.1322003	0.9991324928	0.0000679404
7	0.0000059475	-4.2261513	0.9999881139	0.0000059474
8	0.0000004400	-5.3201022	0.9999999481	0.0000004400
9	0.0000000279	-6.4140532	0.9999999999	0.0000000279

Total probability 0.03484477

[Using a Poisson for managers we get 0.0355314.]

19

The mean aggregate loss is $6863200 + 720000 = \$7,583,200$, and the variance of the aggregate loss is $835611304000 + 705600000000 = 1,541,211,304,000$, (so the standard deviation is 1241455.316956675650)

(a) Using a normal distribution, the probability that the aggregate loss exceeds 10,000,000 is $1-\Phi\left(\frac{10000000-7583200}{1241455.316956675650}\right) = 1 - \Phi(1.946747472091) = 0.02578251$.

(b) Using a gamma distribution, we have $\theta = \frac{1541211304000}{7583200} = 203240.228927102014$ and $\alpha = \frac{7583200}{203240.228927102014} = 37.311510816689$ We are trying to calculate the probability that the distribution is more than 49.2028573926648θ, which is given by $\frac{\int_{49.2028573926648}^{\infty} x^{\alpha-1}e^{-x}dx}{\Gamma(\alpha)} = 0.03401596$

(c) Using a log-normal distribution, the mean of a log-normal distribution is $e^{\mu+\frac{\sigma^2}{2}}$, while the variance is $e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$. We therefore have $e^{\sigma^2}-1 = \frac{1541211304000}{7583200^2} = 0.026801380542$, so $\sigma^2 = \log(1.026801380542) = 0.02644851$. This gives $e^{\mu} = \frac{7583200}{\sqrt{1.026801380542}} = 7483577.977791868549$, so $\mu = 15.82822$.

The probability that this is greater than 10,000,000 is therefore $1 - \Phi\left(\frac{\log(10000000)-15.82822}{\sqrt{0.02644851}}\right) = 1 - \Phi(1.782415) = 0.03734079$.

20

For individual j, we let I_j be 1 if that individual makes a claim, zero otherwise. We approximate I_j as a Poisson distribution with parameter λ_j.

The aggregate loss is then

$$100000\sum_{j=1}^{4622} I_j + 90000\sum_{j=4623}^{8162} I_j + 200000\sum_{j=8163}^{8964} I_j + 1000000\sum_{j=8965}^{9000} I_j$$

(a) Here we set

$$\lambda_i = \begin{cases} 0.01 & \text{if } 1 \leqslant i \leqslant 4622 \\ 0.002 & \text{if } 4623 \leqslant i \leqslant 8162 \\ 0.01 & \text{if } 8163 \leqslant i \leqslant 8964 \\ 0.02 & \text{if } 8965 \leqslant i \leqslant 9000 \end{cases}$$

This gives us that $\sum_{j=1}^{4622} I_j$, $\sum_{j=4623}^{8162} I_j$, $\sum_{j=8163}^{8964} I_j$ and $\sum_{j=8965}^{9000} I_j$ all follow Poisson distributions with parameters 46.22, 7.08, 8.02 and 0.72 respectively. We approximate the first sum as normal with mean 46.22 and variance 46.22. We summarise the probabilities in the following table:

The total probability is therefore approximated as

$$\sum_{i,j,k} e^{-0.72}e^{-8.02}e^{-7.08}\frac{(0.72)^i(8.02)^j(7.08)^k}{i!j!k!}\Phi\left(\frac{10i+2j+0.9k+46.22-100}{\sqrt{46.22}}\right)$$

$$= e^{-15.82}\sum_{i,j,k}\frac{(0.72)^i(8.02)^j(7.08)^k}{i!j!k!}\Phi\left(1.470906i+0.2941813j+0.1323816k-7.910534\right)$$

$\sum_{j=8965}^{9000} I_j$	$\sum_{j=8163}^{8964} I_j$	$\sum_{j=4623}^{8162} I_j$	Probability	$P(A > 10000000)$	
4	7	8	0.0006294833	0.3521605	0.0002216792
4	7	9	0.0005609396	0.4024106	0.0002257280
4	7	10	0.0004498735	0.4543292	0.0002043907
4	8	7	0.0005557035	0.4138336	0.0002299688
4	8	8	0.0005570927	0.4660064	0.0002596088
4	8	9	0.0004964315	0.5187708	0.0002575342
4	8	10	0.0003981381	0.5712080	0.0002274196
4	9	7	0.0004371534	0.5304841	0.0002319029
4	9	8	0.0004382463	0.5827246	0.0002553769
3	9	9	0.0016271922	0.1293106	0.0002104132
4	9	9	0.0003905261	0.6335423	0.0002474148
3	9	10	0.0013050082	0.1593157	0.0002079083
4	9	10	0.0003132020	0.6821187	0.0002136409
3	10	8	0.0012928265	0.1665581	0.0002153307
4	10	8	0.0003102784	0.6925354	0.0002148787
3	10	9	0.0011520521	0.2017253	0.0002323980
4	10	9	0.0002764925	0.7374265	0.0002038929
3	10	10	0.0009239458	0.2409989	0.0002226699
3	11	8	0.0008321102	0.2502590	0.0002082431
3	11	9	0.0007415026	0.2941527	0.0002181150
3	11	10	0.0005946851	0.3413022	0.0002029674

Adding these and all other terms gives 0.03072908.

(b) Here we set

$$\lambda_i = \begin{cases} 1 - e^{-0.01} = 0.009950166 & \text{if } 1 \leqslant i \leqslant 4622 \\ 1 - e^{-0.002} = 0.001998001 & \text{if } 4623 \leqslant i \leqslant 8162 \\ 1 - e^{-0.01} = 0.009950166 & \text{if } 8163 \leqslant i \leqslant 8964 \\ 1 - e^{-0.02} = 0.019801327 & \text{if } 8965 \leqslant i \leqslant 9000 \end{cases}$$

This gives us that $\sum_{j=1}^{4622} I_j$, $\sum_{j=4623}^{8162} I_j$, $\sum_{j=8163}^{8964} I_j$ and $\sum_{j=8965}^{9000} I_j$ all follow Poisson distributions with parameters 45.9896684, 7.0729247, 7.9800333 and 0.7128478 respectively. We approximate the first sum as normal with mean 45.9896684 and variance 45.9896684. We summarise the probabilities in the following table:

The total probability is therefore approximated as

$$\sum_{i,j,k} e^{-0.7128} e^{-7.9800} e^{-7.0729} \frac{(0.7128)^i (7.9800)^j (7.0729)^k}{i!j!k!} \Phi\left(\frac{10i + 2j + 0.9k + 45.9896684 - 100}{\sqrt{45.9896684}}\right)$$

$$= e^{-15.76581} \sum_{i,j,k} \frac{(0.7128)^i (7.9800)^j (7.0729)^k}{i!j!k!} \Phi\left(1.4745852i + 0.2949170j + 0.1327127k - 7.9642834\right)$$

$\sum_{j=8965}^{9000} I_j$	$\sum_{j=8163}^{8964} I_j$	$\sum_{j=4623}^{8162} I_j$	Probability	$P(A > 10000000)$	
4	7	8	0.0006153420	0.3392880	0.0002087781
4	7	9	0.0005456055	0.3890889	0.0002122890
4	8	7	0.0005453946	0.4004421	0.0002183990
4	8	8	0.0005440334	0.4524435	0.0002461444
4	8	9	0.0004823783	0.5052749	0.0002437336
4	8	10	0.0003849395	0.5580137	0.0002148015
4	9	7	0.0004286150	0.5170353	0.0002216091
4	9	8	0.0004275453	0.5696279	0.0002435417
4	9	9	0.0003790917	0.6210119	0.0002354205
4	9	10	0.0003025165	0.6703400	0.0002027889
3	10	8	0.0012726401	0.1576311	0.0002006077
4	10	8	0.0003023996	0.6809453	0.0002059176
3	10	9	0.0011284122	0.1917328	0.0002163536
3	10	10	0.0009004767	0.2300094	0.0002071181
3	11	9	0.0007255613	0.2821009	0.0002046815

Adding these and all other terms gives 0.02886643.

21

Mean total loss $= 800 \times 0.02 \times 3000 + 2100 \times 0.05 \times 4000 + 500 \times 0.12 \times 5000 = 48000 + 420000 + 300000 =$ \$768,000. The variance of the total loss is

$$800 \times 0.02 \times 0.98 \times 3000^2 + 800 \times 0.02^2 + 2100 \times 0.05 \times 0.95 \times 4000^2 + 2100 \times 0.05 \times 1600^2 + 500 \times 0.12 \times 0.88 \times 5000^2 + 500 \times 0.12 \times 1500^2 = 177,1$$

(a) The gamma approximation therefore has $\theta = \frac{3496920000}{768000} = \frac{1165640}{256} = 4553.28125$ and $\alpha = \frac{768000}{4553.28125} =$ 168.669572080573.

We get $\frac{800000}{\theta} = 175.697470917264$

The expected payment on the stop-loss insurance is therefore

$$\frac{\theta}{\Gamma(\alpha)} \int_{175.697470917264}^{\infty} (x^{\alpha} - 175.697470917264 x^{\alpha-1}) e^{-x} dx = \$11,234.2$$

The expected square of the payment on the stop-loss insurance is therefore

$$\frac{\theta^2}{\Gamma(\alpha)}\int_{175.697470917264}^{\infty}(x^{\alpha+1}-2\times 175.697470917264x^{\alpha}+175.697470917264^2x^{\alpha-1})e^{-x}dx = 740555835$$

so the variance of the stop-loss payment is 614348585, and the standard deviation is $\$24,786.06$

The reinsurance premium is therefore \$36,020.26.

(b) The normal approximation has $\mu = 768000$ and $\sigma^2 = 3496920000$, so the standard deviation is 59134.761350664128 and the cut-off for the stop-loss is 0.541136875656 standard deviations above the mean. The expected payment of the stop-loss is therefore

$$\begin{aligned}
&59134.761350664128\frac{\int_{0.541136875656}^{\infty}(x-0.541136875656)e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\\
&=59134.761350664128\left(\frac{[e^{-\frac{x^2}{2}}]_{0.541136875656}^{\infty}}{\sqrt{2\pi}}-0.541136875656(1-\Phi(0.541136875656))\right)\\
&=59134.761350664128\left(\frac{e^{-\frac{0.541136875656^2}{2}}}{\sqrt{2\pi}}-0.541136875656(1-\Phi(0.541136875656))\right)=10963.59
\end{aligned}$$

The expected square of the payment is

$$\begin{aligned}
&59134.761350664128^2\frac{\int_{0.541136875656}^{\infty}(x-0.541136875656)^2e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\\
&=59134.761350664128^2\frac{\int_{0.541136875656}^{\infty}(x^2-1.082273751312x+0.292829118194)e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\\
&=59134.761350664128\left(\frac{[e^{-\frac{x^2}{2}}]_{0.541136875656}^{\infty}}{\sqrt{2\pi}}-0.541136875656(1-\Phi(0.541136875656))\right)\\
&=59134.761350664128\left(\frac{e^{-\frac{0.541136875656^2}{2}}}{\sqrt{2\pi}}-0.541136875656(1-\Phi(0.541136875656))\right)=10963.59
\end{aligned}$$

Now we have that

$$\int_a^{\infty}x^2e^{-\frac{x^2}{2}}dx=[-xe^{-\frac{x^2}{2}}]_a^{\infty}+\int_a^{\infty}e^{-\frac{x^2}{2}}dx=ae^{-\frac{a^2}{2}}+\sqrt{2\pi}(1-\Phi(a))$$

So the variance is

$$59134.76135^2\left(\frac{0.5411^{-\frac{0.5411^2}{2}}}{\sqrt{2\pi}}+1-\Phi(0.5411)-1.0823\frac{e^{-\frac{0.5411^2}{2}}}{\sqrt{2\pi}}+0.2928(1-\Phi(0.5411))\right)=677982038.478915225279$$

The standard deviation is 26038.088226267980, so the premium is $10963.59 + 26038.088226267980 = \$37,001.68$.

22

if 20% are smokers, the expected number of claims per policy is $0.2\times0.02+0.8\times0.01 = 0.012$, so the premium is set to $1.1 \times 12 = 1.32$. If 30% are smokers, the expected number of claims is per policy is 0.013. Therefore, the claims exceed the premium if the average number of claims per policy is more than 0.002 larger than the expected number. We are told that this has probability at most 0.2, and we have that $\Phi(0.8416212) = 0.8$, so the standard deviation of the number of claims per policy is at most $\frac{0.002}{0.8416212} = 0.002376366$. The variance of the number of claims per policy is therefore 0.000005647115. If the total number of lives is n, then $0.3n$ are smokers and $0.7n$ are non-smokers. The variance of the number of claims is $\frac{0.3\times0.02\times0.98+0.7\times0.01\times0.99}{n} = 0.000005647115$, so we get $n = \frac{0.00588+0.00693}{0.000005647115} = \frac{0.01281}{0.000005647115} = 2268.415$.

So at least 2269 lives.

11 Estimation for Complete Data

11.2 The Empirical Distribution for Complete Individual Data

23

The probability mass function is

n	$P(X = n)$
0	0.2
1	0.1333
2	0.0667
3	0.2667
4	0.1333
6	0.1333
7	0.0667

The cumulative Hazard rate is

x	$H(x)$	$H(x)$
0	0.2231436	0.2
1	0.4054651	0.3667
2	0.5108256	0.4667
3	1.0986123	0.9111
4	1.6094379	1.3111
6	2.7080502	1.9778
7		2.9778

24

The Nelson-Åalen estimate is $H(5) = 1.31111111$, so this gives $S(5) = e^{-1.311111} = 0.2695204$.

11.3 Empirical Distributions for Grouped Data

25

The total number of policies is 1099. 194 are less than \$100,000, and 558 are less than \$500,000, so the empirical estimates are $F(100000) = \frac{194}{1099}$ and $F(500000) = \frac{558}{1099}$. The ogive then gives $F(300000) = \frac{1}{2}\left(\frac{194}{1099} + \frac{558}{1099}\right) = \frac{376}{1099} = 0.3421292$. So the probability that a random policy would be affected by this tax is 0.6578708.

26

See slides.

27

over the 2000 observations, the total of all values of $X \wedge 6000$ is $2000 \times 1810 = 3,620,000$. There are 300 observations for which $X \wedge 6000 = 6000$. The sum of these is therefore $300 \times 6000 = 1,800,000$. The total of the 1700 observations where X is less than $6,000$ is therefore $3,620,000 - 1,800,000 = 1,820,000$. The total of the 30 observations between 6,000 and 7,000 is 200,000, so the total of the 1,730 observations below 7,000 is $2,020,000$. The total of $X \wedge 7000$ for the 270 observations above 7,000 is $7000 \times 270 = 1,890,000$ so the total of all 2000 observations of $X \wedge 7000$ is $2020000 + 1890000 = 3910000$, so the average $\mathbb{E}(X \wedge 7000) = \frac{3910000}{2000} = 1,955$.

28

The total number of observations is $200+x+y$. The number of observations less than 50 is 36. The number of observations less than 150 is $36 + x$. The number of observations less than 250 is $36 + x + y$. Therefore

$$\begin{aligned}
F_n(50) &= \frac{36}{200+x+y} \\
F_n(150) &= \frac{36+x}{200+x+y} \\
F_n(250) &= \frac{36+x+y}{200+x+y} \\
F_n(90) &= \frac{36+0.4x}{200+x+y} = 0.21 \\
F_n(210) &= \frac{36+x+0.6y}{200+x+y} = 0.51
\end{aligned}$$

We therefore need to solve the equations

$$\begin{aligned}
36 + 0.4x &= 0.21(200 + x + y) \\
19x - 21y &= 600 \\
36 + x + 0.6y &= 0.51(200 + x + y) \\
49x + 9y &= 6600 \\
1200x &= 144000 \\
x &= 120 \\
y &= 80
\end{aligned}$$

12 Estimation for Modified Data

12.1 Point Estimation

29

The probability that a randomly chosen individual survives to more than 1.6 is expressed as the product

$$\frac{11}{12} \times \frac{13}{14} \times \frac{15}{16} \times \frac{14}{15} \times \frac{11}{13} \times \frac{9}{11} \times \frac{7}{8} = \frac{11 \times 9 \times 7}{16 \times 12 \times 8} = \frac{231}{512} = 0.451171925$$

30

$$\begin{aligned}\frac{8}{9} &= \frac{8}{9} \geqslant \frac{1}{2}\\ \frac{8}{9} \times \frac{8}{9} &= \frac{64}{81} \geqslant \frac{1}{2}\\ \frac{64}{81} \times \frac{10}{12} &= \frac{160}{243} \geqslant \frac{1}{2}\\ \frac{160}{243} \times \frac{10}{11} &= \frac{1600}{2673} \geqslant \frac{1}{2}\\ \frac{1600}{2673} \times \frac{8}{10} &= \frac{1280}{2673} < \frac{1}{2}\end{aligned}$$

So the median is $y_5 = 0.8$.

31

The cumulative hazard rate funtion is given by $H(1.6) = \frac{1}{12}+\frac{1}{14}+\frac{1}{16}+\frac{1}{15}+\frac{1}{13}+\frac{2}{11}+\frac{1}{8} = \frac{17160+15015+16016+18480+43680+30030}{240240} = \frac{140381}{240240}$. The survival function is therefore $S(1.6) = e^{-\frac{140381}{240240}} = 0.5574756$

12.2 Means, Variances and Interval Estimation

32

The Kaplan-Meier estimator gives $S_n(0.5) = \frac{8}{9} \times \frac{8}{9} \times \frac{10}{12} = \frac{160}{243}$ and $S_n(1) = \frac{8}{9} \times \frac{8}{9} \times \frac{10}{12} \times \frac{10}{11} \times \frac{8}{10} = \frac{1280}{2673}$

So the conditional probability is

$$\frac{\frac{160}{243} - \frac{1280}{2673}}{\frac{160}{243}} = \frac{3}{11}$$

33

Suppose we are estimating the survival function at x which is in the interval $(c_1, c_2]$. The estimate is

$$S(x) = \frac{c_2 - x}{c_2 - c_1} S(c_1) + \frac{x - c_1}{c_2 - c_1} S(c_2)$$

Let X be the number of observations from a sample of n observations that are less than c_1, and let Y be the number that are between c_1 and c_2. We then have

$$\hat{S}(x) = \frac{c_2 - x}{c_2 - c_1}\frac{n - X}{n} + \frac{x - c_1}{c_2 - c_1}\frac{n - X - Y}{n} = 1 - \frac{X(c_2 - x) + (X+Y)(x - c_1)}{c_2 - c_1} = 1 - \frac{X}{n} - \frac{x - c_1}{c_2 - c_1} Y$$

We therefore have that

$$\mathrm{Var}(\hat{S}(x)) = \frac{\mathrm{Var}(X) + \left(\frac{x-c_1}{c_2-c_1}\right)^2 \mathrm{Var}(Y) + 2\left(\frac{x-c_1}{c_2-c_1}\right)\mathrm{Cov}(X,Y)}{n^2}$$

We also have that X and Y are multinomially distributed with probabilities $1 - S(c_1)$ and $S(c_1) - S(c_2)$ respectively. This means

$$\begin{aligned}\mathrm{Var}(X) &= nS(c_1)(1-S(c_1))\\ \mathrm{Var}(Y) &= n(S(c_1)-S(c_2))(1+S(c_2)-S(c_1))\\ \mathrm{Cov}(X,Y) &= -n(1-S(c_1))(S(c_1)-S(c_2))\end{aligned}$$

This gives that

$$\mathrm{Var}(\hat{S}(x)) = \frac{(c_2-c_1)^2S(c_1)(1-S(c_1)) - 2(c_2-c_1)(x-c_1)(1-S(c_1))(S(c_1)-S(c_2)) + (x-c_1)^2(S(c_1)-S(c_2))(1+S(c_2)-S(c_1}{n(c_2-c_1)^2}$$

$$(c_2-c_1)(1-S(c_1))\big((c_2-c_1)S(c_1)-(x-c_1)(S(c_1)-S(c_2)\big)+(x-c_1)(S(c_1)-S(c_2))\big((x-c_1)(1+S(c_2)-S(c_1))-(c_2-c_1)(1-S(c_1))\big)$$

$$(c_2-c_1)(1-S(c_1))\big((c_2-x)S(c_1)+(x-c_1)S(c_2)\big)+(x-c_1)(S(c_1)-S(c_2))\big((x-c_1)(1-S(c_1))-(c_2-x)(1-S(c_1))\big)$$

34

We compute $\hat{S}(10000) = \frac{1446}{4356}$ and $\hat{S}(100000) = \frac{683}{4356}$. This gives $\hat{S}(50000) = \frac{5}{9}\times\frac{1446}{4356} + \frac{4}{9}\times\frac{683}{4356} = \frac{9962}{9\times 4356} =$ 0.2541067

The variances are given by

$$\begin{aligned}\mathrm{Var}(\hat{S}(10000)) &= \frac{1446\times 2910}{4356^3}\\ \mathrm{Var}(\hat{S}(100000)) &= \frac{683\times 3673}{4356^3}\\ \mathrm{Cov}(\hat{S}(10000),\hat{S}(100000)) &= -\frac{2910\times 683}{4356^3}\\ \mathrm{Var}(\hat{S}(50000)) &= \frac{9^2\times 1446\times 2910 + 4^2\times 683\times 3673 - 2\times 9\times 4\times 1446\times 683}{4356^3\times 9^2} = \frac{237873044}{4356^3\times 81}\\ &= 0.00003553011\end{aligned}$$

The standard deviation is $\sqrt{0.00003553011} = 0.005960714$.

A 95% confidence interval is therefore $0.2541067 \pm 1.96\times 0.005960714 = [0.2424237, 0.2657897]$.

35

Let the dying times be $t_1,\ldots,t_n$, and let the corresponding risk sets be $r_1,\ldots,r_n$. Let the number of people surviving at each dying time be X_i. Suppose that the true probability of surviving at time t_i is p_i. The Kaplan-Meier estimate of the survival probability is therefore $\prod_{i=1}^n \frac{X_i}{r_i}$. Since the X_i are independent, we have

$$\begin{aligned}\mathbb{E}\left(\prod_{i=1}^n \frac{X_i}{r_i}\right) &= \prod_{i=1}^n \mathbb{E}\left(\frac{X_i}{r_i}\right)\\ &= \prod_{i=1}^n p_i\end{aligned}$$

so the Kaplan-Meier estimate is unbiassed.
We also have:

$$\begin{aligned}
\mathbb{E}\left(\prod_{i=1}^{n}\frac{X_i}{r_i}\right)^2 &= \prod_{i=1}^{n}\mathbb{E}\left(\frac{X_i}{r_i}\right)^2 \\
&= \prod_{i=1}^{n}\left(\frac{p_i(1-p_i)}{r_i}+p_i^2\right)
\end{aligned}$$

so the variance is

$$\prod_{i=1}^{n}\left(\frac{p_i(1-p_i)}{r_i}+p_i^2\right)-\prod_{i=1}^{n}{p_i}^2=\left(\prod_{i=1}^{n}p_i\right)^2\left(\prod_{i=1}^{n}\left(1+\frac{(1-p_i)}{p_ir_i}\right)-1\right)$$

If we let s_i be the total survivial probability up to time i, so that $s_i=\prod_{j=1}^{i}p_j$, then this becomes

$${s_n}^2\left(\prod_{i=1}^{n}\left(1+\frac{(s_{i-1}-s_i)}{s_ir_i}\right)-1\right)$$

36
Greenwoods formula gives that the variance is

$$\begin{aligned}
\mathrm{Var}(S_n(y_j)) &\approx \hat{S}(y_j)^2\sum_{i=1}^{j}\frac{s_i}{r_i(r_i-s_i)} \\
&= \left(\frac{1280}{2673}\right)^2\left(\frac{1}{9\times 8}+\frac{1}{9\times 8}+\frac{2}{12\times 10}+\frac{1}{11\times 10}+\frac{2}{10\times 8}\right) \\
&= \left(\frac{1280}{2673}\right)^2\left(\frac{55+55+66+36+99}{3960}\right) \\
&= 0.0180089
\end{aligned}$$

The 95% confidence interval is therefore

$$0.4788627\pm 1.96\sqrt{0.0180089}=[0.2158361, 0.7418893]$$

37
The confidence interval is

$$[S_n(1)^{\frac{1}{U}}, S_n(1)^U]$$

where

$$U=e^{1.96\frac{\sqrt{0.0180089}}{0.4788627\log(0.4788627)}}=0.4742837$$

So the confidence interval is

$$[0.4788627^{2.108443}, 0.4788627^{0.4742837}] = [0.2117109, 0.7052276]$$

38
The Nelson-Åalen estimator is $H(5) = \frac{226}{1641} + \frac{387}{1415} + \frac{290}{1028} + \frac{215}{738} + \frac{176}{523} = 1.321168$
The variance of this estimator is then $\frac{226}{1641^2} + \frac{387}{1415^2} + \frac{290}{1028^2} + \frac{215}{738^2} + \frac{176}{523^2} = 0.001589823$
We therefore have

$$\log(\hat{H}(5)) = 0.2785162$$
$$\mathrm{Var}(\log(\hat{H}(5))) = \frac{0.001589823}{1.321168} = 0.001203347$$

So a 95% confidence interval for $\log(H(5))$ is

$$0.2785162 \pm 1.96\sqrt{0.001203347} = [0.2754001, 0.2816323]$$

The corresponding interval for $H(5)$ is

$$[1.317058, 1.325291]$$

and the corresponding interval for $S(5)$ is

$$[0.2657256, 0.2679224]$$

12.3 Kernel Density Models

39
See slide.
40
See slides.

12.4 Approximations for Large Data Sets

41
(a) The exact exposure is $1+0.7+1+0.2+0.8+1+0.4+1+0.4+0.8+0.2+1+0.4+1+0.5+0.1+0.9+0.6+0.2+0.4 = 12.2$ years. There are two deaths in the interval. The estimate for the hazard rate is therefore $\frac{3}{12.2} = 0.2459016$, and the probability of dying in the year is $1 - e^{-0.2459016} = 0.2180008$.
(b) The actuarial exposure is $1+0.7+1+0.2+0.8+1+0.4+1+0.4+0.8+0.2+1+0.4+1+0.5+0.1+0.9+0.6+1+1 = 13.6$, so the estimate for the probability of dying is $\frac{3}{13.6} = 0.2205882$.
42
Using insuring ages, the table looks like this:

entry	death	exit	entry	death	exit
61.2	-	64.2	63.0	-	64.0
61.7	-	63.0	61.8	-	64.0
62.4	-	64.1	61.4	-	63.0
60.1	-	62.3	62.6	-	65.6
62.8	-	65.8	61.0	62.4	-
62.0	-	64.3	62.0	63.2	-
63.6	-	66.6	62.0	64.9	-
61.7	-	64.7	62.1	-	63.5
60.2	-	63.0	62.2	62.7	-
60.4	-	62.9	62.8	65.0	-

(a) Now the exact exposure is given by $1+0+1+0+1+1+0.4+1+0+0+1+1+0+1+0+0.2+1+0.5+0+1=11.1$, so the estimated hazard rate is $\frac{1}{11.1}=0.09009009$ and the estimated probability of dying is $1-e^{-0.09009009}=0.08615115$.

The actuarial exposure is given by $1+0+1+0+1+1+0.4+1+0+0+1+1+0+1+0+1+1+0.5+0+1=11.9$ so the estimated probability of dying is $\frac{1}{11.9}=0.08403361$.

(b) Using an anniversary-to-anniversary study, we ignore all partial units of exposure, so the exposure is 11, which makes $q_{63}=\frac{1}{11}=0.0909$.

43

See next slide.

44

(a) The exact exposure is $\frac{15+11}{2}=13$. The hazard rate is therefore $\frac{3}{13}=0.2307692$ and the probability of dying during the year is therefore $1-e^{-0.2307692}=0.2060773$.

(b) The actuarial exposure is $15+\frac{5-6}{2}=14.5$ and the probability of dying during the year is therefore $\frac{3}{14.5}=0.2068966$.

45

Now death is the censoring event and withdrawl is the event we are trying to estimate. The exposure is $15+\frac{5-3}{2}=16$ and the probability of withdrawing is therefore $\frac{6}{16}=0.375$.

15 Bayesian Estimation

15.1 Bayesian Estimation

46

(a) The marginal density function is given by

$$
\begin{aligned}
f_X(x) &= \int_0^\infty \frac{\theta^4}{1000^5 \times 24} e^{-\frac{\theta}{1000}} \frac{\theta^3}{3x^4} e^{-\frac{\theta}{x}}\, d\theta \\
&= \int_0^\infty \frac{\theta^7}{1000^5 \times 72x^4} e^{-\theta\left(\frac{1}{1000}+\frac{1}{x}\right)}\, d\theta \\
&= \frac{\left(\frac{1}{1000^5 \times 72x^4}\right)}{\left(\frac{1}{1000}+\frac{1}{x}\right)^8} \Gamma(8) \\
&= \frac{1000^3 x^4 \times 70}{(1000+x)^8}
\end{aligned}
$$

(b) For a fixed value of θ, the likelihood of the sample is

$$
\frac{\theta^3}{3 \times 132^4} e^{-\frac{\theta}{132}} \cdots \frac{\theta^3}{3 \times 4422^4} e^{-\frac{\theta}{4422}} = \frac{\theta^{13} e^{-\theta\left(\frac{1}{132}+\cdots+\frac{1}{4422}\right)}}{(132 \times \cdots \times 4422)^4}
$$

The posterior distribution of θ is therefore given by

$$
\begin{aligned}
\pi_{\Theta|X}(\theta|x) &= \frac{\frac{\theta^{13} e^{-\theta\left(\frac{1}{132}+\cdots+\frac{1}{4422}\right)} \pi(\theta)}{(132 \times \cdots \times 4422)^4}}{\int_0^\infty \pi(\theta) \frac{\theta^{13} e^{-\theta\left(\frac{1}{132}+\cdots+\frac{1}{4422}\right)}}{(132 \times \cdots \times 4422)^4} d\theta} \\
&= \frac{\theta^{13} e^{-\theta\left(\frac{1}{132}+\cdots+\frac{1}{4422}\right)} \pi(\theta)}{\int_0^\infty \pi(\theta) \theta^{13} e^{-\theta\left(\frac{1}{132}+\cdots+\frac{1}{4422}\right)} d\theta} \\
&= \frac{\theta^{13} e^{-\theta\left(\frac{1}{132}+\cdots+\frac{1}{4422}\right)} \theta^4 e^{-\frac{\theta}{1000}}}{\int_0^\infty \theta^4 e^{-\frac{\theta}{1000}} \theta^{13} e^{-\theta\left(\frac{1}{132}+\cdots+\frac{1}{4422}\right)} d\theta} \\
&= \frac{\theta^{17} e^{-\theta\left(\frac{1}{132}+\cdots+\frac{1}{4422}+\frac{1}{1000}\right)}}{\int_0^\infty \theta^{17} e^{-\theta\left(\frac{1}{132}+\cdots+\frac{1}{4422}+\frac{1}{1000}\right)}\, d\theta}
\end{aligned}
$$

So the posterior distribution of θ is a gamma distribution with $\alpha = 17$ and $\theta = \frac{1}{\frac{1}{132}+\cdots+\frac{1}{4422}+\frac{1}{1000}} = 28.62476$.
(c) The predictive distribution of X is calculated in the same way as the marginal distribution. That is:

$$
\begin{aligned}
f_{Y|X}(y|x) &= \int_0^\infty \frac{\theta^{16}}{28.62476^{17}\times 16!}e^{-\frac{\theta}{28.62476}}\frac{\theta^3}{3y^4}e^{-\frac{\theta}{y}}\,d\theta \\
&= \int_0^\infty \frac{\theta^{20}}{28.62476^{17}\times 16!\times 3y^4}e^{-\theta\left(\frac{1}{28.62476}+\frac{1}{y}\right)}\,d\theta \\
&= \frac{\left(\frac{1}{28.62476^{17}\times 16!\times 3y^4}\right)}{\left(\frac{1}{1000}+\frac{1}{y}\right)^{21}}\Gamma(21) \\
&= C\frac{y^{17}}{(1000+y)^{21}}
\end{aligned}
$$

15.2 Inference and Prediction

47

(a) The mean of a gamma distribution with $\alpha = 17$ and $\theta = 28.62476$ is $17 \times 28.62476 = 486.6209$.

(b) We want to calculate the expected value of the predictive distribution of X. This is given by

$$
\begin{aligned}
\mathbb{E}(Y|X) &= \frac{\int_0^\infty \frac{x^{18}}{(1000+x)^{21}}\,dx}{\int_0^\infty \frac{x^{17}}{(1000+x)^{21}}\,dx} \\
&= \frac{\int_0^\infty \frac{x^{18}}{(1000+x)^{21}}\,dx}{\int_0^\infty \frac{x^{18}}{(1000+x)^{21}}\,dx}
\end{aligned}
$$

We have

$$
\begin{aligned}
\int_0^\infty \frac{x^{17}}{(1000+x)^{21}}\,dx &= \int_{1000}^\infty \frac{(u-1000)^{17}}{u^{21}}\,du \\
&= \int_{1000}^\infty (u^{-4} - 17000u^{-5} + 136000000u^{-6} - \cdots - 1000^{17})\,du \\
&= \frac{1}{1000^3}\left(\frac{1}{3} - \frac{17}{4} + \frac{136}{5} - \cdots - \frac{1}{21}\right)\,du \\
&= 2.923977e - 12
\end{aligned}
$$

and

$$\int_0^\infty \frac{x^{18}}{(1000+x)^{21}}\,dx = \int_{1000}^\infty \frac{(u-1000)^{18}}{u^{21}}\,du$$
$$= \frac{1}{1000^2}\left(\frac{1}{2} - \frac{18}{3} + \frac{153}{4} - \cdots + \frac{1}{21}\right)du$$
$$= 2.631579e - 09$$

So the expected value of the predictive distribution is

$$\frac{2.631579e-09}{2.923977e-12} = 900.00$$

[We can also get this (exactly 900) by expanding the series as a beta function and integrating by parts.]

(c) For a fixed value of θ, the expected value of x is $\frac{\theta}{3-1}$, so for $\theta = 486.6209$, the expected value of x is $\frac{486.6209}{2} = 243.310$.

48

The posterior distribution is given by

$$\pi_{\Lambda|X}(\lambda) = \frac{\frac{20^4}{6}\lambda^3 e^{-20\lambda} e^{-400000\lambda}\lambda^{62310}}{\int_0^\infty \frac{20^4}{6}\lambda^3 e^{-20\lambda} e^{-400000\lambda}\lambda^{62310}} \propto \lambda^{62313} e^{-400020\lambda}$$

This is a Gamma distribution with $\alpha = 62314$ and $\theta = \frac{1}{400020}$.

(a) To get a HPD interval, we need to calculate which values of λ have the same probability density. That is, for a fixed λ, find the other value of λ' such that $62313\log\left(\frac{\lambda'}{\lambda}\right) = 400020(\lambda' - \lambda)$

Let $r = \frac{\lambda'-\lambda}{\lambda}$. Then this equation becomes

$62313\log(1+r) = 400020\lambda r$

We want to find the solution to this such that the total probability of the interval is 0.95

R-code:

```
PDdiff<-function(x,alpha){
    xprime<-qgamma(0.95+pgamma(x,alpha),alpha)
    return(dgamma(x,alpha)-dgamma(xprime,alpha))
}
```

A search reveals that to solve PDdiff(x)=0.95, we get 61825.02157621. The corresponding upper point is 62803.54. The interval is then

$$\left[\frac{61825.02157621}{400020}, \frac{62803.54}{400020}\right] = [0.1545548, 0.157001]$$

(b) We are looking to calculate the 97.5th percentile of this Gamma distribution. We can compute this numerically.

The interval is

R-code:

```
qgamma(0.025,62314)/400020
qgamma(0.975,62314)/400020
```

This gives the interval

$$[0.1545565, 0.1570027]$$

(c) The posterior distribution has mean $\frac{62314}{400020}$ and variance $\left(\frac{62314}{400020}\right)^2$. Using a normal approximation, the 95% confidence interval is

$$\frac{62314}{400020} \pm 1.96\frac{\sqrt{62314}}{400020} = [0.1545541, 0.1570003]$$

15.3 Conjugate Priors

49

If the prior distribution is $\pi(\theta)$, then the posterior distribution is proportional to

$$\frac{\pi(\theta)e^{r(\theta)\sum X_i}}{q(\theta)^N} = \pi(\theta)e^{r(\theta)\sum X_i - N\log(q(\theta))}$$

The prior distribution must therefore take the form $\pi(\theta) \propto h(\theta)e^{\alpha r(\theta)-\beta\log(q(\theta))}$

We can choose $h(\theta) = \frac{r'(\theta)}{\beta}$

16 Model Selection

16.3 Graphical Comparison

50

The log-likelihood of this Pareto distribution is

$$14\left(\log(\alpha)+\alpha\log(\theta)\right)-(\alpha+1)\left(\log(\theta+325)+\log(\theta+692)+\log(\theta+1340)+\log(\theta+1784)+\log(\theta+1920)+\log(\theta+2503)+\log(\theta+$$

Differentiating with respect to α and θ give

$$\frac{14}{\alpha} = (\log(\theta+325)+\log(\theta+692)+\log(\theta+1340)+\log(\theta+1784)+\log(\theta+1920)+\log(\theta+2503)+\log(\theta+3238)+\log(\theta+4054)$$

$$\frac{\alpha}{\theta} = (\alpha+1)\left(\frac{1}{\theta+325}+\frac{1}{\theta+692}+\frac{1}{\theta+1340}+\frac{1}{\theta+1784}+\frac{1}{\theta+1920}+\frac{1}{\theta+2503}+\frac{1}{\theta+3238}+\frac{1}{\theta+4054}+\frac{1}{\theta+5862}+\frac{1}{\theta+6304}+\right.$$

$\theta = 4156615$ $\alpha = 934.25$

51

See next slide.

16.4 Hypothesis Tests

52

(a) $D = 0.1605338$

- At the 95% level, the critical value is $\frac{1.36}{\sqrt{14}} = 0.3634753$.

- At the 95% level, the critical value is $\frac{1.22}{\sqrt{14}} = 0.3260587$.

so we cannot reject the model.
(b) We have that $F(x) = 1 - \frac{\theta^\alpha}{(x+\theta)^\alpha}$, so the statistic is

$$n\int_t^u \frac{\left(F_n(x) + \frac{\theta^\alpha}{(x+\theta)^\alpha} - 1\right)^2}{\left(\frac{\theta^\alpha((x+\theta)^\alpha - \theta^\alpha)}{(x+\theta)^{2\alpha}}\right)} \left(\frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}\right) dx$$

$$n\int_t^u \alpha(x+\theta)^{\alpha-1} \frac{\left(F_n(x) + \frac{\theta^\alpha}{(x+\theta)^\alpha} - 1\right)^2}{((x+\theta)^\alpha - \theta^\alpha)} dx$$

$$n\int_t^u \alpha \frac{(F_n(x)(x+\theta)^\alpha - ((x+\theta)^\alpha - \theta^\alpha))^2}{(x+\theta)^{\alpha+1}((x+\theta)^\alpha - \theta^\alpha)} dx$$

$$n\int_t^u \frac{\alpha}{(x+\theta)} \left(F_n(x)^2 \frac{(x+\theta)^\alpha}{((x+\theta)^\alpha - \theta^\alpha)} - 2F_n(x) + \frac{((x+\theta)^\alpha - \theta^\alpha)}{(x+\theta)^\alpha}\right) dx$$

For a constant value $F_n(x) = c$, we have

$$\begin{aligned}
&\int_a^b c^2 \frac{\alpha(x+\theta)^{\alpha-1}}{((x+\theta)^\alpha - \theta^\alpha)} - \frac{2\alpha c}{x+\theta} + \frac{\alpha((x+\theta)^\alpha - \theta^\alpha)}{(x+\theta)^{\alpha+1}} dx \\
&= c^2 \left[\log((x+\theta)^\alpha - \theta^\alpha)\right]_a^b - 2\alpha c \left[\log(x+\theta)\right]_a^b + \alpha \left[\log(x+\theta)\right]_a^b - \left[-\frac{\theta^\alpha}{(x+\theta)^\alpha}\right]_a^b \\
&= c^2 \log\left(\frac{(b+\theta)^\alpha - \theta^\alpha}{(a+\theta)^\alpha - \theta^\alpha}\right) + \alpha(1-2c)\log\left(\frac{b+\theta}{a+\theta}\right) + \frac{\theta^\alpha}{(b+\theta)^\alpha} - \frac{\theta^\alpha}{(a+\theta)^\alpha} \\
&= c^2 \log\left(\frac{(a+\theta)^\alpha((b+\theta)^\alpha - \theta^\alpha)}{(b+\theta)^\alpha((a+\theta)^\alpha - \theta^\alpha)}\right) + \alpha(1-c)^2 \log\left(\frac{b+\theta}{a+\theta}\right) + \frac{\theta^\alpha}{(b+\theta)^\alpha} - \frac{\theta^\alpha}{(a+\theta)^\alpha}
\end{aligned}$$

For our example, if we let $t = 0$ and $u = \infty$, we have the following:

$$14\left(\alpha\left(1^2\log\left(\frac{325+\theta}{\theta}\right)+\left(\frac{13}{14}\right)^2\log\left(\frac{692+\theta}{325+\theta}\right)+\left(\frac{12}{14}\right)^2\log\left(\frac{1340+\theta}{692+\theta}\right)+\left(\frac{11}{14}\right)^2\log\left(\frac{1784+\theta}{1340+\theta}\right)\right.\right.$$
$$+\left(\frac{10}{14}\right)^2\log\left(\frac{1920+\theta}{1784+\theta}\right)+\left(\frac{9}{14}\right)^2\log\left(\frac{2503+\theta}{1920+\theta}\right)+\left(\frac{8}{14}\right)^2\log\left(\frac{3238+\theta}{2503+\theta}\right)+\left(\frac{7}{14}\right)^2\log\left(\frac{4054+\theta}{3238+\theta}\right)$$
$$+\left(\frac{6}{14}\right)^2\log\left(\frac{5862+\theta}{4054+\theta}\right)+\left(\frac{6}{14}\right)^2\log\left(\frac{6304+\theta}{5862+\theta}\right)+\left(\frac{5}{14}\right)^2\log\left(\frac{6926+\theta}{6304+\theta}\right)+\left(\frac{4}{14}\right)^2\log\left(\frac{6926+\theta}{6304+\theta}\right)$$
$$\left.+\left(\frac{3}{14}\right)^2\log\left(\frac{8120+\theta}{6926+\theta}\right)+\left(\frac{2}{14}\right)^2\log\left(\frac{9176+\theta}{8120+\theta}\right)+\left(\frac{1}{14}\right)^2\log\left(\frac{9984+\theta}{9176+\theta}\right)\right)$$
$$\left.+\left(\frac{1}{14}\right)^2\log\left(\frac{1-\left(\frac{\theta}{692+\theta}\right)^\alpha}{1-\left(\frac{\theta}{325+\theta}\right)^\alpha}\right)+\cdots+\left(\frac{14}{14}\right)^2\log\left(\frac{1}{1-\left(\frac{\theta}{9984+\theta}\right)^\alpha}\right)-1\right)$$
$$=0.3873562$$

So the model cannot be rejected.

53

If the parameter of the Exponential distribution is θ, then the log-likelihood of the data is

$$742\log(1-e^{-\left(\frac{5000}{\theta}\right)})+1304\log(e^{-\left(\frac{5000}{\theta}\right)}-e^{-\left(\frac{10000}{\theta}\right)})+1022\log(e^{-\left(\frac{10000}{\theta}\right)}-e^{-\left(\frac{15000}{\theta}\right)})+$$
$$830\log(e^{-\left(\frac{15000}{\theta}\right)}-e^{-\left(\frac{20000}{\theta}\right)})+211\log(e^{-\left(\frac{20000}{\theta}\right)}-e^{-\left(\frac{25000}{\theta}\right)})-143\left(\frac{25000}{\theta}\right)$$

Taking the derivative with respect to θ, we get

$$742\frac{5000e^{-\left(\frac{5000}{\theta}\right)}}{\theta^2(1-e^{-\left(\frac{5000}{\theta}\right)})}+1304\frac{(5000e^{-\left(\frac{5000}{\theta}\right)}-10000e^{-\left(\frac{10000}{\theta}\right)})}{\theta^2(e^{-\left(\frac{5000}{\theta}\right)}-e^{-\left(\frac{10000}{\theta}\right)})}+$$
$$1022\frac{(10000e^{-\left(\frac{10000}{\theta}\right)}-15000e^{-\left(\frac{15000}{\theta}\right)})}{\theta^2(e^{-\left(\frac{10000}{\theta}\right)}-e^{-\left(\frac{15000}{\theta}\right)})}+830\frac{(15000e^{-\left(\frac{10000}{\theta}\right)}-20000e^{-\left(\frac{20000}{\theta}\right)})}{\theta^2(e^{-\left(\frac{15000}{\theta}\right)}-e^{-\left(\frac{20000}{\theta}\right)})}+$$
$$211\frac{(20000e^{-\left(\frac{20000}{\theta}\right)}-25000e^{-\left(\frac{25000}{\theta}\right)})}{\theta^2(e^{-\left(\frac{20000}{\theta}\right)}-e^{-\left(\frac{25000}{\theta}\right)})}-143\frac{25000}{\theta^2}=0$$

Multiplying by $\frac{\theta^2(1-e^{-\frac{5000}{\theta}})}{5000}$ gives

$$
\begin{aligned}
742e^{-\left(\frac{5000}{\theta}\right)} + 1304(1-2e^{-\left(\frac{5000}{\theta}\right)}) + 1022(2-3e^{-\left(\frac{5000}{\theta}\right)}) + 830(3-4e^{-\left(\frac{5000}{\theta}\right)})+ \\
211(4-5e^{-\left(\frac{5000}{\theta}\right)}) - 143(5-5e^{-\left(\frac{5000}{\theta}\right)}) &= 0 \\
5967 - 10076e^{-\left(\frac{5000}{\theta}\right)} &= 0 \\
e^{-\left(\frac{5000}{\theta}\right)} &= \frac{5967}{10076} \\
\theta &= \frac{5000}{\log\left(\frac{10076}{5967}\right)} \\
&= 9543.586
\end{aligned}
$$

This gives the following table

Claim Amount	O_i	E_i	$\frac{(O_i-E_i)^2}{E_i}$
0–5,000	742	1733.969	567.49
5,000–10,000	1304	1026.855	74.80
10,000–15,000	1022	608.103	281.71
15,000–20,000	830	360.118	613.10
20,000–25,000	211	213.262	0.02
More than 25,000	143	309.694	89.72
total			1626.85

This should be compared to a Chi-square with 5 degrees of freedom, so the model is rejected at all significance levels.

54

For the exponential distribution, the log-likelihood is

$$
-\left(\frac{382}{\theta}+\frac{596}{\theta}+\frac{920}{\theta}+\frac{1241}{\theta}+\frac{1358}{\theta}+\frac{1822}{\theta}+\frac{2010}{\theta}+\frac{2417}{\theta}+\frac{2773}{\theta}+\frac{3002}{\theta}+\frac{3631}{\theta}+\frac{4120}{\theta}+\frac{4692}{\theta}+\frac{5123}{\theta}+14\log(\theta)\right)
$$

This is maximised by

$$
\theta = \frac{382+596+920+1241+1358+1822+2010+2417+2773+3002+3631+4120+4692+5123}{14} = 2434.786
$$

Which gives a log-likelihood of $-(14+14\log(2434.786)) = -123.1666$.

For the Weibull distribution, the log-likelihood is

$$
14\log(\tau) + (\tau-1)(\log(382)+\cdots+\log(5213)) - \left(\left(\frac{382}{\theta}\right)^{\tau} + \cdots + \left(\frac{5123}{\theta}\right)^{\tau} + 14\tau\log(\theta)\right)
$$

Setting the derivatives with respect to θ and τ equal to zero gives:

$$\tau\left(\frac{382^\tau}{\theta^{\tau+1}}+\cdots+\frac{5123^\tau}{\theta^{\tau+1}}-\frac{14}{\theta}\right)=0$$
$$\frac{382^\tau+\cdots+5123^\tau}{14}=\theta^\tau$$
$$\frac{14}{\tau}+(\log(382)+\cdots+\log(5213))-\left(\left(\frac{382}{\theta}\right)^\tau\log\left(\frac{382}{\theta}\right)+\cdots+\left(\frac{5123}{\theta}\right)^\tau\log\left(\frac{5123}{\theta}\right)\right)-14\log(\theta)=0$$
$$\frac{14}{\tau}+\left(1-\left(\frac{382}{\theta}\right)^\tau\right)\log(382)+\cdots+\left(1-\left(\frac{5123}{\theta}\right)^\tau\right)\log(5123)=0$$

This gives the solution $\tau = 1.695356$ and $\theta = 2729.417$
$l(x;\tau,\theta) = -120.7921$
The log-likelihood ratio statistic is therefore

$$2(-120.7921-(-123.1666)) = 4.749$$

For a Chi-square with 1 degree of freedom, this has a p-value 0.04703955, so the Weibull model is preferred at the 5% significance level.

17 Introduction and Limited Fluctuation Credibility

17.2 Limited Fluctuation Credibility Theory

17.3 Full Credibility

55

(a) The number of claims made is a binomial distribution with $n = 372 \times 7 = 2604$ and some unknown p. The expected number of claims is np and the variance is $np(1-p)$, so the relative error $\frac{\overline{X}-\xi}{\xi}$ is approximately normally distributed with mean zero and variance $\frac{1-p}{np}$. We therefore want to check whether $\Phi\left(\frac{0.05}{\sqrt{\frac{1-p}{np}}}\right) \geqslant 0.975$ (two-sided confidence interval).

In this example, the total number of claims in seven years of experience is 9. This sets $p = \frac{9}{2604}$, and

$$\Phi\left(\frac{0.05}{\sqrt{\frac{1-p}{np}}}\right) = \Phi\left(\frac{0.15}{\sqrt{1-\frac{9}{2604}}}\right) = 0.5597202 < 0.975$$

So the company should not assign full credibility.

(b) Suppose we continue with the assumption that $p = \frac{9}{2604}$. Then we want to find the n such that

$$\Phi\left(\frac{0.05}{\sqrt{\frac{1-p}{np}}}\right)=\Phi\left(\frac{0.15\sqrt{n}}{\sqrt{2595}}\right)=0.975$$
$$\frac{0.15\sqrt{n}}{\sqrt{2595}}=1.96$$
$$n=\frac{1.96^2\times 2595}{0.15^2}=443064.5$$

If the company continues to employ 372 employees, then this equates to 1191.034 years.

Alternatively, since p is small, we can approximate $1-p=0$, so we are looking to solve

$$0.05\sqrt{np}=1.96$$
$$np=400\times 1.96^2=1568.64$$

So the standard for full credibility is 1568.64 claims.

56

(a)

Based on the data, the coefficient of variation is $\frac{3605.52}{962.14}=3.747396$. Assuming the number of claims is large enough to use a normal approximation, we have that the critical value is 1.96 at the 95% confidence level. This means that the coefficient of variation for the average $\overline{X}$ is $\frac{3.747396}{\sqrt{41876}}=0.01831247$. Multiplying by 1.96 gives us the relative 95% confidence interval as 0.03589244. Since this is less than 0.05, the company should assign full credibility to this data.

(b) The insurance company will assign full credibility if

$$\frac{3.747396}{\sqrt{n}}\times 1.96\leqslant 0.05$$
$$n\geqslant\left(\frac{1.96\times 3.747396}{0.05}\right)^2=21579$$

17.4 Partial Credibility

57

The partial credibility assigned is $Z=\sqrt{\frac{7}{1191.034}}=0.0766632$

The credibility premium is therefore

$$0.0766632\times 338.7097+0.9233368\times 1000=\$949.40$$

58

(a) The credibility for claim frequency is $Z=\sqrt{\frac{19}{421}}=0.2124397$, so the credibility estimate for claim frequency is $0.2124397\times 1.9+0.7875603\times 1.2=1.348708$.

The credibility for claim severity is $Z = \sqrt{\frac{19}{1240}} = 0.1237844$, so the credibility estimate for claim severity is $0.1237844 \times \frac{5822}{19} + 0.8762156 \times 230 = 239.4597$. The credibility estimate for aggregate claims is therefore $1.348708 \times 239.4597 = \$322.9613$.

(b) The credibility for claim frequency is $Z = \sqrt{\frac{19}{1146}} = 0.128761$, so the credibility estimate for claim frequency is $0.128761 \times 1.9 + 0.871239 \times 1.2 = 1.290133$.

The credibility for claim severity is $Z = \sqrt{\frac{19}{611}} = 0.1763422$, so the credibility estimate for claim severity is $0.1763422 \times \frac{5822}{19} + 0.8236578 \times 230 = 243.4763$. The credibility estimate for aggregate claims is therefore $1.290133 \times 243.4763 = \$314.1168$.

(c) The credibility for aggregate losses is $Z = \sqrt{\frac{10}{400}} = 0.1581139$. The credibility premium is therefore $0.1581139 \times 582.2 + 0.8418861 \times 276 = \$324.4145$.

(d) The credibility for aggregate losses is $Z = \sqrt{\frac{10}{1000}} = 0.1$. The credibility premium is therefore $0.1 \times 582.2 + 0.9 \times 276 = \$306.62$.

17.5 Problems with this Approach

59

Using a normal approximation, the standard for full credibility is

$$\Phi\left(\frac{r\sqrt{n}}{\tau}\right) \geqslant 1 - \frac{p}{2}$$

where τ is the coefficient of variation of X. For our data, we have

$$\tau = \frac{\sqrt{8240268} \times 3722}{3506608} = 3.046911$$

The standard for full credibility is therefore given by

$$n = \frac{3.046911}{r}\left(\Phi^{-1}\left(1 - \frac{p}{2}\right)\right)^2$$

The credibility is

$$Z = \sqrt{\frac{3722}{n}} = \sqrt{\frac{3722}{3.046911\Phi^{-1}\left(1 - \frac{p}{2}\right)}r} = \frac{\sqrt{1221.565r}}{\Phi^{-1}\left(1 - \frac{p}{2}\right)}$$

18 Greatest Accuracy Credibility

18.2 Conditional Distributions and Expectation

60

(a) Let $\Theta = 1$ for frequent drivers, and $\Theta = 0$ for infrequent drivers. Then

$$\begin{aligned}\mathbb{E}(X|\Theta=1)&=0.4\\ \mathbb{E}(X|\Theta=0)&=0.1\\ \mathrm{Var}(X|\Theta=1)&=0.4\\ \mathrm{Var}(X|\Theta=0)&=0.1\end{aligned}$$

so

$$\mathbb{E}(X)=\mathbb{E}(\mathbb{E}(X|\Theta)=0.75\times 0.4+0.25\times 0.1=0.325$$

and

$$\mathrm{Var}(X)=\mathbb{E}(\mathrm{Var}(X|\theta))+\mathrm{Var}(\mathbb{E}(X|\Theta))=0.325+0.3^2\times 0.25\times 0.75=0.325+0.016875=0.341875$$

(b)

$$P(X=0|\Theta)=\begin{cases}e^{-0.4} & \text{if }\Theta=1\\ e^{-0.1} & \text{if }\Theta=0\end{cases}$$

So

$$P(\Theta=1|X=0)=\frac{0.75e^{-0.4}}{0.75e^{-0.4}+0.25e^{-0.1}}=0.6896776$$

Therefore the new expectation and variance are:

$$\mathbb{E}(X)=\mathbb{E}(\mathbb{E}(X|\Theta)=0.6896776\times 0.4+0.3103224\times 0.1=0.3069033$$

and

$$\mathrm{Var}(X)=\mathbb{E}(\mathrm{Var}(X|\theta))+\mathrm{Var}(\mathbb{E}(X|\Theta))=0.3069033+0.3^2\times 0.3103224\times 0.6896776=0.325+0.016875=0.3261653$$

18.3 Bayesian Methodology

61

(a) We have $\mathbb{E}(X|\Theta=\theta)=\frac{\theta}{2}$, so

$$\mathbb{E}(X)=\mathbb{E}(\mathbb{E}(X|\Theta=\theta))=\mathbb{E}\left(\frac{\Theta}{2}\right)=150$$

(b) The joint density function is

$$f_{X,\Theta}(x,\theta)=\left(\frac{\theta^2}{2\times 100^3}e^{-\frac{\theta}{100}}\right)\left(\frac{\theta^3}{2x^4}e^{-\frac{\theta}{x_1}}\right)$$

For samples, x_1 and x_2, the joint density is therefore

$$\left(\frac{\theta^2}{2000000}e^{-\frac{\theta}{100}}\right)\left(\frac{\theta^3}{2{x_1}^4}e^{-\frac{\theta}{x_1}}\right)\left(\frac{\theta^3}{2{x_2}^4}e^{-\frac{\theta}{x_2}}\right)$$
$$= \frac{\theta^8}{8000000{x_1}^4{x_2}^4}e^{-\theta\left(\frac{1}{100}+\frac{1}{x_1}+\frac{1}{x_2}\right)}$$

The posterior distribution of Θ is therefore a gamma distribution with $\alpha = 8$ and $\theta = \frac{1}{\frac{1}{100}+\frac{1}{x_1}+\frac{1}{x_2}} = 43.29897$.

The expected aggregate losses are given by

$$\begin{aligned}\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|\Theta)) &= \frac{\mathbb{E}(\Theta)}{2}\\ &= 4 \times 43.29897\\ &= 173.1959\end{aligned}$$

62

(a) We have

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|\Lambda)) = \mathbb{E}\left(\Lambda\right) = 0.4$$

(b) The posterior distribution is a Gamma distribution with $\alpha = 4 + m$ and $\theta = \frac{1}{10+n}$. We therefore have

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(\mathbb{E}(X|\Lambda)) = \mathbb{E}\left(\Lambda\right)\\ &= \frac{m+4}{10+n}\\ &= \left(\frac{n}{10+n}\right)\left(\frac{m}{n}\right) + \left(\frac{10}{10+n}\right)0.4\end{aligned}$$

18.4 The Credibility Premium

63

We are trying to choose α_i to minimise

$$\mathbb{E}\left(\mu(\Theta) - \left(\alpha_0 + \sum_{i=1}^{n}\alpha_i X_i\right)\right)^2 = \mathbb{E}\left(\mu(\Theta)^2 - 2\mu(\Theta)\left(\alpha_0 + \sum_{i=1}^{n}\alpha_i X_i\right) + \left(\alpha_0 + \sum_{i=1}^{n}\alpha_i X_i\right)^2\right)$$
$$= \mathbb{E}\left(\mu(\Theta)^2\right) - 2\alpha_0\mathbb{E}\mu(\Theta) - \mathbb{E}\left(\mu(\Theta)\left(\sum_{i=1}^{n}\alpha_i X_i\right)\right) + {\alpha_0}^2 + 2\alpha_0\mathbb{E}\left(\sum_{i=1}^{n}\alpha_i X_i\right) + \mathbb{E}\left(\left(\sum_{i=1}^{n}\alpha_i X_i\right)^2\right)$$
$$= \mu^2 + v^2 - 2\alpha_0\mu + {\alpha_0}^2 + 2\alpha_0\mathbb{E}\left(\sum_{i=1}^{n}\alpha_i X_i\right) - \mathbb{E}\left(\mu(\Theta)\left(\sum_{i=1}^{n}\alpha_i X_i\right)\right) + \mathbb{E}\left(\left(\sum_{i=1}^{n}\alpha_i X_i\right)^2\right)$$

Setting the derivative with respect to α_0 equal to zero yields

$$2\left(\alpha_0 + \mathbb{E}\left(\sum_{i=1}^{n} \alpha_i X_i\right) - \mu\right) = 0$$

That is, α_0 should be chosen to make the estimate unbiassed. Now we differentiate with respect to α_j, and set the derivative equal to zero:

$$\begin{aligned}
2\left(\alpha_0 \mathbb{E}(X_j) - \mathbb{E}(\mu(\Theta)X_j) + \mathbb{E}\left(X_j \sum_{i=1}^{n} \alpha_i X_i\right)\right) &= 0 \\
2\left(\mathbb{E}\left(\mu(\Theta) - \sum_{i=1}^{n} \alpha_i X_i\right)\mathbb{E}(X_j) - \mathbb{E}\left(X_j\left(\mu(\Theta) - \sum_{i=1}^{n} \alpha_i X_i\right)\right)\right) &= 0 \\
\operatorname{Cov}\left(X_j, \mu(\Theta) - \sum_{i=1}^{n} \alpha_i X_i\right) &= 0 \\
\operatorname{Cov}\left(X_j, \mu(\Theta)\right) &= \sum_{i=1}^{n} \alpha_i \operatorname{Cov}\left(X_j, X_i\right)
\end{aligned}$$

Since X_i and X_{n+1} are conditionally independent given $\mu(\Theta)$, we have that $\operatorname{Cov}(X_j, \mu(\Theta)) = \operatorname{Cov}(X_j, X_{n+1})$
64
In this situation, the second normal equations becomes:

$$\begin{aligned}
\rho &= \left(\sum_{i=1}^{n} \alpha_i \rho\right) + \alpha_j \sigma^2 \\
\alpha_j &= \frac{\rho\left(1 - \sum_{i=1}^{n} \alpha_i\right)}{\sigma^2}
\end{aligned}$$

So all the α_j are equal to a common value α, and we get

$$\alpha = \frac{\rho(1 - n\alpha)}{\sigma^2}$$

Now we have $\mathbb{E}(X_{n+1}) = \mu = \mathbb{E}(X_i)$ The first normal equation then becomes

$$\begin{aligned}
\mu &= \alpha_0 + \left(\sum_{i=1}^{n} \alpha_i\right)\mu \\
&= \alpha_0 + n\alpha\mu \\
\alpha_0 &= (1 - n\alpha)\mu
\end{aligned}$$

We can therefore rewrite our credibility estimate as

$$Z\overline{X} + (1 - Z)\mu$$

where $Z = n\alpha$. We can then solve:

$$\begin{aligned}
\frac{Z}{n} &= \frac{\rho\,(1-Z)}{\sigma^2} \\
\sigma^2 Z &= n\rho(1-Z) \\
(\sigma^2 + n\rho)Z &= n\rho \\
Z &= \frac{n}{n + \frac{\sigma^2}{\rho}}
\end{aligned}$$

65

Let the coefficients of the X_i be α_i, and let the coefficients of Y_i be β_i. The normal equations are:

$$\begin{aligned}
\mu + \nu &= \alpha_0 + \sum_{j=1}^{n} \alpha_i \mu + \sum_{k=1}^{m} \beta_i \nu \\
\rho + \xi &= \sum_{j \neq i} \alpha_j \rho + \sum_{k=1}^{m} \beta_k \xi + \alpha_i \sigma^2 \\
\zeta + \xi &= \sum_{j=1}^{n} \alpha_j \xi + \sum_{k \neq i} \beta_k \zeta + \beta_i \tau^2
\end{aligned}$$

From these, we deduce that $\beta_i(\tau - \zeta) = \beta_j(\tau - \zeta)$, and so $\beta_i = \beta_j = \beta$ (assuming the Y_i are not perfectly correlated). Similarly, $\alpha_i = \alpha_j = \alpha$. Substituting these into the normal equations gives:

$$\begin{aligned}
\mu + \nu &= \alpha_0 + n\alpha\mu + m\beta\nu \\
\rho + \xi &= \alpha((n-1)\rho + \sigma^2) + m\beta\xi \\
\zeta + \xi &= n\alpha\xi + \beta((m-1)\zeta + \tau^2)
\end{aligned}$$

This gives

$$\left(\frac{(n-1)\rho+\tau}{n\xi}\right)(\zeta+\xi)-(\rho+\xi)=\left(\frac{(n-1)\rho+\sigma^2}{n\xi}\right)\beta((m-1)\zeta+\tau^2)-m\beta\xi$$

$$\beta=\frac{\left(\frac{(n-1)\rho+\sigma^2}{n\xi}\right)(\zeta+\xi)-(\rho+\xi)}{\left(\frac{(n-1)\rho+\sigma^2}{n\xi}\right)((m-1)\zeta+\tau^2)-m\xi}$$

$$=\frac{\left((n-1)\rho+\sigma^2\right)(\zeta+\xi)-n\xi(\rho+\xi)}{\left((n-1)\rho+\sigma^2\right)\left((m-1)\zeta+\tau^2\right)-mn\xi^2}$$

$$\alpha=\frac{((m-1)\zeta+\tau^2)-m\xi(\zeta+\xi)}{\left((n-1)\rho+\sigma^2\right)\left((m-1)\zeta+\tau^2\right)-mn\xi^2}$$

$$\alpha_0=(1-n\alpha)\mu+(1-m\beta)\nu$$

18.5 The Buhlmann Model

66

We have $Z=\frac{851}{851+\frac{84036}{23804}}=0.9958687$, and $\overline{X}=\frac{121336}{851}=\$142.58$ so the credibility premium is

$$0.9958687\times 142.58+0.0041313\times 326=143.34$$

67

We have $Z=\frac{10}{10+\frac{732403}{28822}}=0.2823961$, and $\overline{X}=\frac{3224}{10}=322.40$ so the credibility premium is

$$0.2823961\times 322.40+0.7176039\times 990=\$801.47$$

18.6 The Buhlmann-Straub Model

68

The weighted mean is $\frac{1000000}{1242}=805.153$. The credibility is $Z=\frac{1242}{1242+\frac{81243100}{120384}}=0.6479325$. The credibility premium is therefore

$$0.6479325\times 805.153+0.4520675\times 1243=\$959.30$$

69

The weighted mean is $\frac{14000}{\left(\frac{49}{12}\right)}=\$3,428.57$. The credibility is $Z=\frac{\left(\frac{49}{12}\right)}{\left(\frac{49}{12}\right)+\left(\frac{34280533}{832076}\right)}=0.09017537$. The credibility premium is therefore

$$0.09017537\times 3428.57+0.90981463\times 600=\$855.07$$

18.7 Exact Credibility

70

The Bayes premium is the conditional expectation of X_{n+1} given $X_1,\ldots,X_n$. We are given that it is a linear function of X_i. That is

$$\mathbb{E}(X_{n+1}|X_1,\ldots,X_n) = \alpha_0 + \sum_{i=1}^{n} \alpha_i X_i$$

Now recall that

$$\begin{aligned}
\mathrm{Cov}(X_i, X_{n+1}) &= \mathbb{E}(X_i X_{n+1}) - \mathbb{E}(X_i)\mathbb{E}(X_{n+1}) \\
&= \mathbb{E}(\mathbb{E}(X_i X_{n+1}|X_1,\ldots,X_n)) - \mathbb{E}(X_i)\mathbb{E}(\mathbb{E}(X_{n+1}|X_1,\ldots,X_n)) \\
&= \mathbb{E}(X_i \sum_{j=1}^{n} \alpha_j X_j) - \mathbb{E}(X_i)\mathbb{E}(\sum_{j=1}^{n} \alpha_j X_j) \\
&= \sum_{j=1}^{n} \alpha_j \left(\mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j)\right) \\
&= \sum_{j=1}^{n} \alpha_j \,\mathrm{Cov}(X_i, X_j)
\end{aligned}$$

This means that the second normal equation is satisfied by the Bayes premium. We also have

$$\mathbb{E}(X_{n+1}) = \mathbb{E}(\mathbb{E}(X_{n+1}|X_1,\ldots,X_n)) = \mathbb{E}\left(\alpha_0 + \sum_{i=1}^{n} \alpha_i X_i\right)$$

So the first normal equation is satisfied. Thus the Bayes premium is the credibility premium. [Technically, need to show this is the only solution].

71

Recall from Question 49 that the conjugate prior is

$$\pi(\theta) \propto h(\theta) e^{\alpha r(\theta) - \beta \log(q(\theta))}$$

We choose $h(\theta) = Cr'(\theta)$. The posterior distribution is

$$\frac{\pi(\theta) e^{r(\theta)\sum X_i}}{q(\theta)^N} = Cr'(\theta) e^{\alpha r(\theta) - \beta \log(q(\theta))} e^{r(\theta)\sum X_i - N\log(q(\theta))} = \frac{Cr'(\theta) e^{r(\theta)\left(\alpha + \sum_{i=1}^{n} X_i\right)}}{q(\theta)^{\beta+N}}$$

Recall that the mean of a distribution from the linear-exponential family is

$$\frac{q'(\theta)}{r'(\theta) q(\theta)}$$

The posterior mean is therefore

$$\mathbb{E}\left(\frac{q'(\Theta)}{r'(\Theta)q(\Theta)}\right) = \int_{\theta_0}^{\theta_1} \frac{Cr'(\theta)e^{r(\theta)\left(\alpha+\sum_{i=1}^n X_i\right)}}{q(\theta)^{\beta+N}} \frac{q'(\theta)}{r'(\theta)q(\theta)}\,d\theta$$
$$= C\int_{\theta_0}^{\theta_1} \frac{q'(\theta)e^{r(\theta)\left(\alpha+\sum_{i=1}^n X_i\right)}}{q(\theta)^{\beta+N+1}}\,d\theta$$
$$= C\left(\left[\frac{e^{r(\theta)\left(\alpha+\sum_{i=1}^n X_i\right)}}{q(\theta)^{\beta+N}}\right]_{\theta_0}^{\theta_1} - \int_{\theta_0}^{\theta_1} \frac{r'(\theta)e^{r(\theta)\left(\alpha+\sum_{i=1}^n X_i\right)}}{q(\theta)^{\beta+N}}\,d\theta\right)$$

$$\iint \frac{h(\theta)e^{r(\theta)\left(\alpha+\sum_{i=1}^n X_i\right)}}{q(\theta)^{\beta+N}} x\frac{p(x)e^{r(\theta)x}}{q(\theta)}\,dx\,d\theta = \iint x\frac{p(x)h(\theta)e^{r(\theta)\left(\alpha+x+\sum_{i=1}^n X_i\right)}}{q(\theta)^{\beta+N+1}}\,dx\,d\theta$$

19 Empirical Bayes Parameter Estimation

19.2 Nonparametric Estimation

72

(a) The overall mean is $\frac{2770.8}{8} = 346.35$

The EPV is $\frac{60595.2+1225822.8+62760.2+192962.3+0.0+30505.0+140653.7+56385.3}{8} = \frac{1769684.5}{8} = 221210.5625$

The total variance is $\frac{(172.80-346.35)^2+(671.60-346.35)^2+(177.80-346.35)^2+(635.40-346.35)^2+(0.00-346.35)^2+(247.00-346.35)^2+(633.60-346.35)^2+(232.60-346.35)^2}{7} = 67592.36$

The VHM is $67592.36 - \frac{221210.5625}{5} = 23350.25$

(b) The credibility of 5 years of experience is

$$Z = \frac{5}{5+\frac{221210.5625}{23350.25}} = 0.3454569$$

The premiums are

$$0.3454569 \times 172.80 + 0.6545431 \times 346.35 = \$286.40$$
$$0.3454569 \times 671.60 + 0.6545431 \times 346.35 = \$458.71$$
$$0.3454569 \times 177.80 + 0.6545431 \times 346.35 = \$288.12$$
$$0.3454569 \times 635.40 + 0.6545431 \times 346.35 = \$446.20$$
$$0.3454569 \times 0.00 + 0.6545431 \times 346.35 = \$226.70$$
$$0.3454569 \times 247.00 + 0.6545431 \times 346.35 = \$312.03$$
$$0.3454569 \times 633.60 + 0.6545431 \times 346.35 = \$445.58$$
$$0.3454569 \times 232.60 + 0.6545431 \times 346.35 = \$307.05$$

73

In total the aggregate claims were 15.7 million, and the total exposure was 14,693 lives. The average claim per life is therefore $\frac{15700000}{14693} = 1068.54$. The averages for the three companies are:

$$\begin{aligned}\frac{5300000}{3623} &= 1,462.88\\ \frac{4000000}{4908} &= \quad 815.00\\ \frac{6400000}{6162} &= 1,038.62\end{aligned}$$

The total variance is therefore

$$\frac{3}{2}\times\frac{3623(1462.88-1068.54)^2+4908(815.00-1068.54)^2+6162(1038.62-1068.54)^2}{14693} = 90288.51$$

The variances for the three companies are:

$$\frac{4}{3}\times\frac{769(1690.51-1462.88)^2+928(1616.38-1462.88)^2+880(909.09-1462.88)^2+1046(1625.24-1462.88)^2}{3623} = 132179.27$$

$$\frac{4}{3}\times\frac{1430(699.30-815)^2+1207(745.65-815)^2+949(632.24-815)^2+1322(1134.64-815)^2}{4908} = 52082.51$$

$$\frac{4}{3}\times\frac{942(1167.73-1038.62)^2+1485(942.76-1038.62)^2+2031(935.50-1038.62)^2+1704(1173.71-1038.62)^2}{6162} = 17752.09$$

The expected process variance is therefore:

$$\frac{3623\times 132179.27+4908\times 52082.51+6162\times 17752.09}{14693} = 57435.16$$

The variance of hypothetical means is $90288.51 - 57435.16\times\frac{3}{14693} = 90276.78$
The credibilities of the three companies' experiences are therefore

$$\begin{aligned}Z_1 &= \frac{3623}{3623+\frac{57435.16}{90276.78}} = 0.9998244\\ Z_2 &= \frac{4908}{4908+\frac{57435.16}{90276.78}} = 0.9998704\\ Z_3 &= \frac{6162}{6162+\frac{57435.16}{90276.78}} = 0.9998968\end{aligned}$$

The credibility premiums per unit of exposure are therefore:

$$
\begin{aligned}
0.9998244 \times 1462.88 + 0.0001756 \times 1068.54 &= \$1462.81\\
0.9998704 \times 815.00 + 0.0001296 \times 1068.54 &= \$815.03\\
0.9998968 \times 1038.62 + 0.0001032 \times 1068.54 &= \$1038.62
\end{aligned}
$$

19.3 Semiparametric Estimation

74

There are a total of 3193 claims from 6210 policyholders, so the estimate for μ is $\frac{3193}{6210} = 0.5141707$. Since for a Poisson distribution the mean and variance are equal, this gives the expected process variance is also $v = 0.5141707$. We calculate the sample variance

$$\frac{6210}{6209}\left(\frac{1406 + 740 \times 4 + 97 \times 9 + 13 \times 16 + 3 \times 25}{6210} - 0.5141707^2\right) = 0.6249401$$

so the variance of hypothetical means is $0.6249401 - 0.5141707 = 0.1107694$ and the credibility of 3 years of experience is

$$Z = \frac{3}{3 + \frac{0.5141707}{0.1107694}} = 0.3925771$$

so the credibility estimate is

$$0.3925771 \times 2 + 0.6074229 \times 0.5141707 = 1.097473$$

75

11073*1 11073 6181*2 12362 2433*3 7299 1598*4 6392 589*5 2945 329*6 1974 65*7 455 9*8 72 2*9 18

42590

42590 claims were made in

6210*1 6210 8041*2 16082 11207*3 33621 8827*4 35308

91221 years.

34285 policyholders

The global mean is therefore $\mu = \frac{42590}{91221} = 0.4668881$ claims per year. (There are slightly better estimators for the global mean.)

For the Poisson distribution, the mean is equal to the variance, so the expected process variance is also 0.4668881.

We now aim to estimate the variance of hypothetical means.

Weighting each individual equally, the variance of the means of these individuals is

$$\frac{1}{34284}\big(11073\times(0-0.467)^2+2828\times(0.25-0.467)^2+4032\times(0.333-0.467)^2+5011\times(0.5-0.467)^2+$$
$$2214\times(0.667-0.467)^2+985\times(0.75-0.467)^2+4066\times(1-0.467)^2+358\times(1.25-0.467)^2+$$
$$734\times(1.333-0.467)^2+655\times(1.5-0.467)^2+215\times(1.667-0.467)^2+43\times(1.75-0.467)^2+$$
$$983\times(2-0.467)^2+22\times(2.333-0.467)^2+13\times(2.5-0.467)^2+103\times(3-0.467)^2+14\times(4-0.467)^2$$
$$+3\times(5-0.467)^2\big)=0.2576804$$

The expected variance of the means is

$$\frac{\left(6210\times 1+\frac{8041}{2}+\frac{11207}{3}+\frac{8827}{4}\right)\times 0.4668881}{34285}=0.2202404$$

The variance of hypothetical means is therefore

$$0.2576804-0.2202404=0.03743999$$

Therefore the credibility is

$$Z=\frac{3}{3+\frac{0.4668881}{0.03743999}}=0.1939199$$

so the credibility estimate is

$$0.1939199\times 0.66666666667+0.8060801\times 0.4668881=0.5056291$$

20 Simulation

20.1 Basics of Simulation

76

(a) We can simulate a normal with $\mu=2$ and $\sigma^2=9$ by transforming a standard normal. We obtain:

$$\Phi^{-1}(0.1850620)=-0.8962411$$
$$\Phi^{-1}(0.8613517)=1.0864124$$
$$\Phi^{-1}(0.3607076)=-0.3565680$$

Our sample from the normal with mean 2 and standard deviation 3 is therefore:

$$2-3\times 0.8962411=-0.6887234$$
$$2+3\times 1.0864124=5.259237$$
$$2-3\times 0.3565680=0.9302959$$

(b) The density function of this Pareto distribution is

$$F(x) = 1 - \frac{2400^4}{(2400 + x)^4}$$

To find the value corresponding to u, we must solve

$$\begin{aligned}
1 - \frac{2400^4}{(2400 + x)^4} &= u \\
\frac{2400}{(2400 + x)} &= (1 - u)^{\frac{1}{4}} \\
2400 + x &= 2400(1 - u)^{-\frac{1}{4}} \\
x &= 2400\left((1 - u)^{-\frac{1}{4}} - 1\right)
\end{aligned}$$

The three simulated numbers are therefore

$$\begin{aligned}
2400\left(0.8149380^{-\frac{1}{4}} - 1\right) &= 125.9811 \\
2400\left(0.1386483^{-\frac{1}{4}} - 1\right) &= 1533.0784 \\
2400\left(0.6392924^{-\frac{1}{4}} - 1\right) &= 284.0238
\end{aligned}$$

(c) The distribution function of the Poisson distribution is:

n	$P(X \leqslant n)$
0	0.09071795
1	0.30844104
2	0.56970875
3	0.77872291
4	0.90413141

We see where our simulated u values fit in this table (taking the upper bound in each case), to get the simulated values:

```
1  4  2
```

20.2 Simulation for Specific Distributions

77

(a) The first random number is 0.29351756, which is between 0.02 and 0.88, so the first driver is average. The loss amount is therefore given by solving

$$
\begin{aligned}
1-\frac{4000^4}{(4000+x)^4} &= 0.11768610\\
\frac{4000^4}{(4000+x)^4} &= 0.88231390\\
\frac{4000}{(4000+x)} &= 0.88231390^{\frac{1}{4}}\\
(4000+x) &= \frac{4000}{0.88231390^{\frac{1}{4}}}\\
x &= 4000\left(0.88231390^{-\frac{1}{4}}-1\right) = 127.1876
\end{aligned}
$$

The third random number is 0.47362823, which is again between 0.02 and 0.88, so the second driver is also average. The loss amount is generated from the fourth random number, and is given by solving

$$
\begin{aligned}
1-\frac{4000^4}{(4000+x)^4} &= 0.13843535\\
\frac{4000^4}{(4000+x)^4} &= 0.86156465\\
\frac{4000}{(4000+x)} &= 0.86156465^{\frac{1}{4}}\\
(4000+x) &= \frac{4000}{0.86156465^{\frac{1}{4}}}\\
x &= 4000\left(0.86156465^{-\frac{1}{4}}-1\right) = 151.8153
\end{aligned}
$$

(b) The first simulated loss is the same as in (a), but for the second, we don't need to simulate the driver's type again, so we use the 0.47362823 to simulate the second loss.

$$
\begin{aligned}
1-\frac{4000^4}{(4000+x)^4} &= 0.47362823\\
\frac{4000^4}{(4000+x)^4} &= 0.52637177\\
\frac{4000}{(4000+x)} &= 0.52637177^{\frac{1}{4}}\\
(4000+x) &= \frac{4000}{0.52637177^{\frac{1}{4}}}\\
x &= 4000\left(0.52637177^{-\frac{1}{4}}-1\right) = 696.0947
\end{aligned}
$$

78

We start by simulating the number of deaths as binomial with $n = 720$, and $p = 0.01$. We can use a normal approximation to obtain our simulated number of deaths. $\Phi^{-1}(0.3876723) = -0.2853910$, so

$$720 \times 0.01 - 0.2853910\sqrt{720 \times 0.01 \times 0.99} = 6.438054$$

Rounding this produces 6 deaths.

Conditional on 6 deaths, the number of disabilities has a binomial distribution with $n = 714$ and $p = \frac{0.04}{0.99} = 0.04040404$.

$\Phi^{-1}(0.2534800) = -0.6635787$, so

$$714 \times 0.0404040404 - 0.6635787\sqrt{714 \times 0.0404040404 \times 0.9595959595} = 25.3571$$

Rounding this produces 25 disabilities.

Conditional on 6 deaths and 25 disabilities, the number of lapses has a binomial distribution with $n = 695$ and $p = \frac{0.12}{0.95} = 0.1263158$.

$\Phi^{-1}(0.2954348) = -0.5375763$, so

$$695 \times 0.1263158 - 0.5375763\sqrt{695 \times 0.1263158 \times 0.8736842} = 83.08145$$

Rounding this produces 83 lapses.

79

The possible outcomes are listed with their probabilities in the following table:

Event	Probability
Dies first year	0.01
Lapses first year	0.02
Dies second year	0.01455
Lapses second year	0.0388
Dies third year	0.018333
Survives	0.898317

We convert the simulated numbers to standard normal:

$$\Phi^{-1}(0.8579075) = 1.0709654$$
$$\Phi^{-1}(0.8193713) = 0.9129717$$
$$\Phi^{-1}(0.4031135) = -0.2452963$$
$$\Phi^{-1}(0.7313493) = 0.6168989$$
$$\Phi^{-1}(0.9613431) = 1.7664893$$
$$\Phi^{-1}(0.7735622) = 0.7506296$$
$$\Phi^{-1}(0.9745215) = 1.9518417$$
$$\Phi^{-1}(0.6261118) = 0.3215727$$

We therefore get

$$2 + 1.0709654\sqrt{200 \times 0.01 \times 0.99} = 3.506982$$

so we simulate 4 deaths in the first year. Then

$$196 \times \frac{0.02}{0.99} + 0.9129717\sqrt{196 \times \frac{0.02}{0.99} \times \frac{0.97}{0.99}} = 5.75785$$

so we simulate 6 lapses in the first year. Then

$$190 \times \frac{0.01455}{0.97} - 0.2452963\sqrt{190 \times \frac{0.01455}{0.97} \times \frac{0.95545}{0.97}} = 2.43901$$

so we simulate 2 deaths in the second year. Then

$$188 \times \frac{0.0388}{0.95545} + 0.6168989\sqrt{188 \times \frac{0.0388}{0.95545} \times \frac{0.91665}{0.95545}} = 9.30408$$

so we simulate 9 lapses in the second year. Finally

$$179 \times \frac{0.018333}{0.91665} + 1.7664893\sqrt{179 \times \frac{0.018333}{0.91665} \times \frac{0.898317}{0.91665}} = 6.888762$$

so we simulate 7 deaths in the third year.

80

For the Poisson distribution, this is almost the definition of a Poisson distribution — the number of events happening in unit time for a Poisson process.

For a binomial with parameters n and p, we can consider n separate Poisson processes with rate $log((1-p))$. The probability that an event happens in such a process before time 1 is p, so the number of processes in which events have occurred by time 1 follows a binomial distribution with parameters n and p as required. Now we will instead simulate the events in the order in which they occur, eliminating each process once it yields an event. After k events, the number of processes remaining is $n-k$, so the rate is $\lambda_k = (n-k)\log(1-p)$, so in the notation of the table $c = n\log(1-p)$ and $d = -\log(1-p)$.

For the negative binomial with parameters r and β, we can prove the result by induction. We first observe that the probability of zero is $e^{-r\log(1+\beta)} = (1+\beta)^{-r}$, as required. Now suppose the result holds for k. We now have the distribution of $T_1+\cdots+T_k$, since $P(T_1+\cdots+T_k < x) = P\left(\frac{T_1}{x}+\cdots+\frac{T_k}{x} < 1\right)$. The probability that $\frac{T_1}{x}+\cdots+\frac{T_k}{x} < 1$ is the probability that a negative binomial distribution with parameters r and β', where $\log(1+\beta') = x\log(1+\beta)$ has value more than k. Now the probability that our negative binomial distribution has value at least $k+1$ is

$$
\begin{aligned}
&\int_0^1 P\left(\frac{T_1}{x}+\cdots+\frac{T_k}{x}<1\right)(r+k)\log(1+\beta)e^{(1-x)(r+k)\log(1+\beta)}\,dx \\
&=\int_0^1\left(\sum_{n=k}^{\infty}\binom{r+k-1}{k}\left(\frac{\beta'}{1+\beta'}\right)^k\left(\frac{1}{1+\beta'}\right)^r\right)(r+k)\log(1+\beta)e^{(1-x)(r+k)\log(1+\beta)}\,dx \\
&=\int_0^1\left(\sum_{n=k}^{\infty}\binom{r+n-1}{n}\frac{(1-(1+\beta)^x)^n}{(1+\beta)^{(r+n)x}}\right)(r+k)\log(1+\beta)(1+\beta)^{(1-x)(r+k)}\,dx \\
&=\int_0^1\left(\sum_{m=0}^{\infty}\binom{r+k-1+m}{k+m}\frac{(1-(1+\beta)^x)^{k+m}}{(1+\beta)^{(r+k+m)x}}\right)(r+k)\log(1+\beta)(1+\beta)^{(1-x)(r+k)}\,dx \\
&=(r+k)(1+\beta)^{(r+k)}\log(1+\beta)\int_0^1\left(\sum_{m=0}^{\infty}\binom{r+k-1+m}{k+m}\frac{(1-(1+\beta)^x)^{k+m}}{(1+\beta)^{(2r+2k+m)x}}\right)\,dx \\
&=(r+k)(1+\beta)^{(r+k)}\sum_{m=0}^{\infty}\binom{r+k-1+m}{k+m}\log(1+\beta)\int_0^1\frac{(1-(1+\beta)^x)^{k+m}}{(1+\beta)^{(2r+2k+m)x}}\,dx
\end{aligned}
$$

substituting $w=(1+\beta)^x$, we get $\frac{dw}{dx}=w\log(1+\beta)$, so

$$
\begin{aligned}
\log(1+\beta)\int_0^1\frac{(1-(1+\beta)^x)^{k+m}}{(1+\beta)^{(2r+2k+m)x}}\,dx &=\int_1^{(1+\beta)}\frac{(1-w)^{k+m}}{w^{(2r+2k+m+1)}}\,dx \\
&=\int_1^{(1+\beta)}\sum_{i=0}^{k+m}\binom{k+m}{i}\frac{(-w)^i}{w^{(2r+2k+m+1)}}\,dx \\
&=\sum_{i=0}^{k+m}\binom{k+m}{i}(-1)^i\int_1^{(1+\beta)}w^{i-(2r+2k+m+1)}\,dx \\
&=\sum_{i=0}^{k+m}\binom{k+m}{i}(-1)^i\frac{\left((1+\beta)^{i-(2r+2k+m)}-1\right)}{i-(2r+2k+m)}
\end{aligned}
$$

This gives that the probability that our negative binomial distribution has value at least $k+1$ is

$$
\begin{aligned}
&(r+k)(1+\beta)^{(r+k)}\sum_{m=0}^{\infty}\binom{r+k-1+m}{k+m}\sum_{i=0}^{k+m}\binom{k+m}{i}(-1)^i\frac{\left((1+\beta)^{i-(2r+2k+m)}-1\right)}{i-(2r+2k+m)}\,dx \\
&=(r+k)(1+\beta)^{(r+k)}\sum_{i=0}^{\infty}\sum_{n=\max(i,k)}^{\infty}\binom{r+n-1}{n}\binom{n}{i}(-1)^i\frac{\left((1+\beta)^{i-(2r+k+n)}-1\right)}{i-(2r+k+n)}\,dx
\end{aligned}
$$

81

To generate an exponential distribution with rate λ from the uniform random variable u, we want to solve

$$1 - e^{-\lambda t} = u$$
$$t = \frac{-\log(1-u)}{\lambda}$$

For our random values we have

u	$-\log(1-u)$
0.9587058	3.1870332
0.4975469	0.6882530
0.7957639	1.5884786
0.1762183	0.1938497
0.8649957	2.0024486
0.4639014	0.6234372
0.4426729	0.5846030
0.4197114	0.5442297
0.4212635	0.5469080
0.3984598	0.5082619
0.4043391	0.5180837
0.3122119	0.3742745

(a) We have $c = -20\log(0.86)$ and $d = \log(0.86)$.

This gives

k	λ_k	T_k	Cumulative Sum
1	3.0164578	1.05654826	1.05654826

So we simulate $X = 0$.

(b) We have $c = 6$ and $d = 0$.

This gives

k	λ_k	T_k	Cumulative Sum
1	6	0.53117220	0.5311722
2	6	0.11470883	0.6458810
3	6	0.26474643	0.9106275
4	6	0.03230829	0.9429358
5	6	0.33374144	1.2766772

So we simulate $X = 4$.

(c) We have $c = 3\log(3)$ and $d = \log(3)$.

This gives

k	λ_k	T_k	Cumulative Sum
1	6	0.96698755	0.9669876
2	6	0.15661871	1.1236063

So we simulate $X = 1$.

82

(a) Using the Box-Muller method, we take

$$
\begin{aligned}
Z_1 &= \sqrt{-2\log(0.9974532)}\cos(2\times 0.4429451\pi) = -0.06687504\\
Z_2 &= \sqrt{-2\log(0.9974532)}\sin(2\times 0.4429451\pi) = 0.02505647
\end{aligned}
$$

(b) Using the polar method, we first calculate

$$
\begin{aligned}
X_1 &= 2\times 0.9974532 - 1 = 0.9949064\\
X_2 &= 2\times 0.4429451 - 1 = -0.1141098\\
W &= 0.9949064^2 + 0.1141098^2 = 1.00286
\end{aligned}
$$

Since $W > 1$, we reject this sample and take the next sample:

$$
\begin{aligned}
X_1 &= 2\times 0.6159707 - 1 = 0.2319414\\
X_2 &= 2\times 0.6626078 - 1 = 0.3252156\\
W &= 0.2319414^2 + 0.3252156^2 = 0.3104239\\
Y &= \sqrt{\frac{-2\log(0.3104239)}{0.3104239}} = 2.745341\\
Z_1 &= 0.2319414\times 2.745341 = 0.6367582\\
Z_2 &= 0.3252156\times 2.745341 = 0.8928277
\end{aligned}
$$

20.3 Determining the Sample Size

83

The estimate is $\hat{p} = \frac{N}{n}$, where n is the number of simulations performed, and N is the number larger than \$200,000. We know N has a binomial distribution with parameters n and p. This means that for large n, $\hat{p}$ approximately follows a normal distribution with mean p and variance $\frac{p(1-p)}{n}$. The probability that the error is at most 0.0001 is therefore

$$
2\Phi\left(\frac{0.0001n}{p(1-p)}\right) - 1
$$

To have a 95% probability of this we need

$$
\begin{aligned}
\Phi\left(\frac{0.0001n}{p(1-p)}\right) &= 0.975\\
\frac{0.0001n}{p(1-p)} &= 1.96\\
n &= 19600p(1-p)
\end{aligned}
$$

The largest this can be is 4900, when $p = 0.5$, so this many will always be sufficient.

We can alternatively use $\hat{p}$ as an estimator for p, and stop when $n \geqslant 19600\hat{p}(1-\hat{p})$.

84

The estimate for the 95th percentile is the 95th percentile of the sample. If the sample size is n, then taking the pth percentile of the sample, we estimate the variance is $\frac{p(1-p)}{n}$. We can use a normal approximation for our estimated distribution function. The 95% confidence interval is therefore $p \pm \sqrt{\frac{p(1-p)}{n}}$. Let π be our estimate of the 95th percentile. We want to have 95% confidence that the true 95th percentile is between $\pi + 100$ and $\pi - 100$. Alternatively, we can demand that the probability is at least 0.975 that the distribution function $F_X(\pi + 100) > 0.95$ and probability at least 0.975 that the distribution function $F_X(\pi + 100) < 0.95$. A one-sided 97.5% confidence interval for $F_X(x)$ is $(-\infty, \frac{N}{n} + 1.96\sqrt{\frac{N(n-N)}{n^3}}]$ or $[\frac{N}{n} - 1.96\sqrt{\frac{N(n-N)}{n^3}}, \infty)$. We see that 0.95 is not within this confidence interval whenever $\hat{p} > 0.95 + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. That is, if we let $\hat{p}$ be the estimated density function at $\pi + 100$, we must have

$$\hat{p} - 0.95 > 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

For large n, we will have $\hat{p} \approx 0.95$, so we will use this value to get

$$\hat{p} - 0.95 > 1.96\sqrt{\frac{0.95 \times 0.05}{n}} = \frac{0.4271721}{\sqrt{n}}$$

Letting N be the number of observations between π and $\pi + 100$, we have

$$\frac{N}{n} > \frac{0.4271721}{\sqrt{n}}$$

giving

$$\frac{N}{\sqrt{n}} > 0.4271721$$

and a similar condition for the number of observations between π and $\pi + 100$.

If we simulate a uniform random variable U, the corresponding Pareto loss is found by solving

$$\begin{aligned}
\frac{4300^{2.5}}{(4300 + x)^{2.5}} &= 1 - U \\
4300 + x &= 4300(1 - U)^{-0.4} \\
x &= 4300((1 - U)^{-0.4} - 1)
\end{aligned}$$

To simulate the number of losses, recall that the negative binomial is the number of exponential random variables that need to be added together to total more than 1, where

$$\lambda_k = (r + k)\log(1 + \beta)$$

Using this we simulate 1000000 aggregate losses.

For the first n simulated aggregate losses, we calculate the estimated 95th percentile π, and the quantity $\frac{N}{\sqrt{n}}$, the number of simulated values between π and $\pi + 100$.

We find that the first time that we have $\frac{N}{\sqrt{n}} \geqslant 0.4271721$ is when $n = 4950000$.

20.4 Examples of Simulation in Actuarial Modelling

85

As usual, we simulate a uniform distribution U, then invert to get

$$X = 6000((1-U)^{\frac{1}{2.2}} - 1)$$

For the ETNB distribution, we have

n	p_n	$F(n)$
0	0.91	0.91
1	0.07148907	0.98148907
2	0.01081054	0.9922996
3	0.003814452	0.9961141
4	0.001730458	0.9978445
5	0.0008897089	0.9987342
6	0.0004933183	0.9992275
7	0.0002877403	0.9995153
8	0.0001740478	0.9996893
9	0.0001082021	0.9997975
10	0.00006872152	0.9998663
11	0.00004.440233	0.9999107
12	0.00002.909616	0.9999398
13	0.00001.929190	0.9999591
14	0.00001.291952	0.9999720
15	0.000008.726456	0.9999807
16	0.000005.938247	0.9999867
17	0.000004.067316	0.9999907
18	0.000002.801929	0.9999935
19	0.000001.940129	0.9999955
20	0.000001.349572	0.9999968

We simulate another uniform distribution and use this lookup table to invert the distribution and get the number of losses. We then simulate claim sizes for all the losses and sum to get aggregate losses. We finally apply the stop-loss insurance and take the mean to estimate the expected aggregate claims.

For the simulation I performed, I get the mean as $1,733,716.

86

We simulate from the Gamma distribution, estimate the 95th percentile empirically, then take the average of all values above this percentile.

For one simulation we get 21540.21.

87

We simulate from a Gamma distribution with $\alpha = 3.7$ and $\theta = 1,352$. We then reestimate α and θ from the data using maximum likelihood. We then calculate the Anderson-Darling test statistic for a large number of samples of size 186. We then count the number of these samples on which the Anderson-Darling test statistic is at least 1.84, divided by the total number of samples. This is the p-value.

In the simulation I performed, 10 out of 10000 datasets had a statistic larger than 1.84. We therefore estimate the p-value as 0.001.

If the p-value is 0.001, then the number of datasets with statistic larger than 1.84 can be approximated by a normal with mean 10 and variance $10000 \times 0.001 \times 0.999 = 9.99$. A 95% confidence interval for this p-value is therefore $0.001 \pm 0.000196\sqrt{9.99} = [0.0003805036, 0.001619496]$.