Non-Life Insurance: Mathematics and Statistics

Solution sheet 1

Solution 1.1 Discrete Distribution

(a) Note that N only takes values in $\mathbb{N} \setminus \{0\}$ and that $p \in (0,1)$. Hence we calculate

$$\mathbb{P}[N \in \mathbb{R}] = \sum_{k=1}^{\infty} \mathbb{P}[N = k] = \sum_{k=1}^{\infty} (1-p)^{k-1} p = p \sum_{k=0}^{\infty} (1-p)^k = p \frac{1}{1-(1-p)} = p \frac{1}{p} = 1,$$

from which we can conclude that the geometric distribution indeed defines a probability distribution on $\mathbb{R}$.

(b) For $n \in \mathbb{N} \setminus \{0\}$, we get

$$\mathbb{P}[N \geq n] = \sum_{k=n}^{\infty} \mathbb{P}[N = k] = \sum_{k=n}^{\infty} (1-p)^{k-1} p = (1-p)^{n-1} p \sum_{k=0}^{\infty} (1-p)^k = (1-p)^{n-1},$$

where we used that $\sum_{k=0}^{\infty}(1-p)^k = \frac{1}{p}$, as was shown in (a).

(c) The expectation of a discrete random variable that takes values in $\mathbb{N} \setminus \{0\}$ can be calculated as

$$\mathbb{E}[N] = \sum_{k=1}^{\infty} k \cdot \mathbb{P}[N = k].$$

Thus we get

$$\mathbb{E}[N] = \sum_{k=1}^{\infty} k(1-p)^{k-1} p = \sum_{k=0}^{\infty} (k+1)(1-p)^k p = \sum_{k=0}^{\infty} k(1-p)^k p + \sum_{k=0}^{\infty} (1-p)^k p = (1-p)\mathbb{E}[N] + 1,$$

where we used that $\sum_{k=0}^{\infty}(1-p)^k p = 1$, as was shown in (a). We conclude that $\mathbb{E}[N] = \frac{1}{p}$.

(d) Let $r \in \mathbb{R}$. Then we calculate

$$\begin{aligned}
\mathbb{E}[\exp\{rN\}] &= \sum_{k=1}^{\infty} \exp\{rk\} \cdot \mathbb{P}[N = k] \\
&= \sum_{k=1}^{\infty} \exp\{rk\} (1-p)^{k-1} p \\
&= p \exp\{r\} \sum_{k=1}^{\infty} [(1-p)\exp\{r\}]^{k-1} \\
&= p \exp\{r\} \sum_{k=0}^{\infty} [(1-p)\exp\{r\}]^{k}.
\end{aligned}$$

Since $(1-p)\exp\{r\}$ is strictly positive, the sum on the right hand side is convergent if and only if $(1-p)\exp\{r\} < 1$, which is equivalent to $r < -\log(1-p)$. Hence $\mathbb{E}[\exp\{rN\}]$ exists if and only if $r < -\log(1-p)$ and in this case we have

$$M_N(r) = \mathbb{E}[\exp\{rN\}] = p \exp\{r\} \frac{1}{1-(1-p)\exp\{r\}} = \frac{p \exp\{r\}}{1-(1-p)\exp\{r\}}.$$

(e) For $r < -\log(1-p)$, we have

$$\begin{aligned}\frac{d}{dr} M_N(r) &= \frac{d}{dr} \frac{p \exp\{r\}}{1-(1-p)\exp\{r\}} \\ &= \frac{p\exp\{r\}[1-(1-p)\exp\{r\}] + p\exp\{r\}(1-p)\exp\{r\}}{[1-(1-p)\exp\{r\}]^2} \\ &= \frac{p\exp\{r\}}{[1-(1-p)\exp\{r\}]^2}.\end{aligned}$$

Hence we get

$$\frac{d}{dr} M_N(r)|_{r=0} = \frac{p\exp\{0\}}{[1-(1-p)\exp\{0\}]^2} = \frac{p}{[1-(1-p)]^2} = \frac{p}{p^2} = \frac{1}{p}.$$

We observe that $\frac{d}{dr} M_N(r)|_{r=0} = \mathbb{E}[N]$, which holds in general for all random variables if the moment generating function exists in an interval around 0.

Solution 1.2 Absolutely Continuous Distribution

(a) We calculate

$$\mathbb{P}[Y \in \mathbb{R}] = \int_{-\infty}^{\infty} f_Y(x)\,dx = \int_0^{\infty} \lambda \exp\{-\lambda x\}\,dx = [-\exp\{-\lambda x\}]_0^{\infty} = [-0-(-1)] = 1,$$

from which we can conclude that the exponential distribution indeed defines a probability distribution on $\mathbb{R}$.

(b) For $0 < y_1 < y_2$, we calculate

$$\begin{aligned}\mathbb{P}[y_1 \le Y \le y_2] &= \int_{y_1}^{y_2} f_Y(x)\,dx \\ &= \int_{y_1}^{y_2} \lambda \exp\{-\lambda x\}\,dx \\ &= [-\exp\{-\lambda x\}]_{y_1}^{y_2} \\ &= \exp\{-\lambda y_1\} - \exp\{-\lambda y_2\}.\end{aligned}$$

(c) The expectation and the second moment of an absolutely continuous random variable can be calculated as

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} x f_Y(x)\,dx \qquad \text{and} \qquad \mathbb{E}[Y^2] = \int_{-\infty}^{\infty} x^2 f_Y(x)\,dx.$$

Thus, using partial integration, we get

$$\begin{aligned}\mathbb{E}[Y] &= \int_0^{\infty} x\lambda \exp\{-\lambda x\}\,dx \\ &= [-x\exp\{-\lambda x\}]_0^{\infty} + \int_0^{\infty} \exp\{-\lambda x\}\,dx \\ &= 0 + \left[-\frac{1}{\lambda}\exp\{-\lambda x\}\right]_0^{\infty} \\ &= \frac{1}{\lambda}.\end{aligned}$$

The variance $\mathrm{Var}(Y)$ can be calculated as

$$\mathrm{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = \mathbb{E}[Y^2] - \frac{1}{\lambda^2}.$$

For the second moment $\mathbb{E}[Y^2]$ we get, again using partial integration,

$$\begin{aligned}
\mathbb{E}[Y^2] &= \int_0^\infty x^2 \lambda \exp\{-\lambda x\}\, dx \\
&= \left[-x^2 \exp\{-\lambda x\}\right]_0^\infty + \int_0^\infty 2x \exp\{-\lambda x\}\, dx \\
&= 0 + \frac{2}{\lambda}\mathbb{E}[Y] \\
&= \frac{2}{\lambda^2},
\end{aligned}$$

from which we can conclude that

$$\mathrm{Var}(Y) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Note that for the exponential distribution both the expectation and the variance exist. The reason is that $\exp\{-\lambda x\}$ goes much faster to 0 than x or x^2 go to infinity, for all $\lambda > 0$.

(d) Let $r \in \mathbb{R}$. Then we calculate

$$\mathbb{E}[\exp\{rY\}] = \int_0^\infty \exp\{rx\}\lambda \exp\{-\lambda x\}\, dx = \int_0^\infty \lambda \exp\{(r-\lambda)x\}\, dx.$$

The integral on the right hand side and therefore also $\mathbb{E}[\exp\{rY\}]$ exist if and only if $r < \lambda$. In this case we have

$$M_Y(r) = \mathbb{E}[\exp\{rY\}] = \frac{\lambda}{r-\lambda}\left[\exp\{(r-\lambda)x\}\right]_0^\infty = \frac{\lambda}{r-\lambda}(0-1) = \frac{\lambda}{\lambda - r}$$

and therefore

$$\log M_Y(r) = \log\left(\frac{\lambda}{\lambda - r}\right).$$

(e) For $r < \lambda$, we have

$$\frac{d^2}{dr^2}\log M_Y(r) = \frac{d^2}{dr^2}\log\left(\frac{\lambda}{\lambda - r}\right) = \frac{d^2}{dr^2}[\log(\lambda) - \log(\lambda - r)] = \frac{d}{dr}\frac{1}{\lambda - r} = \frac{1}{(\lambda - r)^2}.$$

Hence we get

$$\frac{d^2}{dr^2}\log M_Y(r)|_{r=0} = \frac{1}{(\lambda - 0)^2} = \frac{1}{\lambda^2}.$$

We observe that $\frac{d^2}{dr^2}\log M_Y(r)|_{r=0} = \mathrm{Var}(Y)$, which holds in general for all random variables if the moment generating function exists in an interval around 0.

Solution 1.3 Conditional Distribution

(a) For $y > \theta > 0$, we get

$$\begin{aligned}
\mathbb{P}[Y \geq y] &= \mathbb{P}[Y \geq y, I = 0] + \mathbb{P}[Y \geq y, I = 1] \\
&= \mathbb{P}[Y \geq y | I = 0]\mathbb{P}[I = 0] + \mathbb{P}[Y \geq y | I = 1]\mathbb{P}[I = 1] \\
&= 0 \cdot (1-p) + \mathbb{P}[Y \geq y | I = 1] \cdot p \\
&= p \cdot \mathbb{P}[Y \geq y | I = 1],
\end{aligned}$$

since $Y|I=0$ is equal to 0 almost surely and thus $\mathbb{P}[Y \geq y|I=0]=0$. Since $Y \mid I=1 \sim$ Pareto(θ, α), we can calculate

$$\mathbb{P}[Y \geq y|I=1] = \int_y^\infty f_{Y|I=1}(x)\,dx = \int_y^\infty \frac{\alpha}{\theta}\left(\frac{x}{\theta}\right)^{-(\alpha+1)}\,dx = \left[-\left(\frac{x}{\theta}\right)^{-\alpha}\right]_y^\infty = \left(\frac{y}{\theta}\right)^{-\alpha}.$$

We conclude that

$$\mathbb{P}[Y \geq y] = p\left(\frac{y}{\theta}\right)^{-\alpha}.$$

(b) Using that $Y|I=0$ is equal to 0 almost surely and thus $\mathbb{E}[Y|I=0]=0$, we get

$$\mathbb{E}[Y] = \mathbb{E}[Y\cdot 1_{\{I=0\}}]+\mathbb{E}[Y\cdot 1_{\{I=1\}}] = \mathbb{E}[Y|I=0]\mathbb{P}[I=0]+\mathbb{E}[Y|I=1]\mathbb{P}[I=1] = p\cdot\mathbb{E}[Y|I=1].$$

Since $Y \mid I=1 \sim$ Pareto(θ, α), we can calculate

$$\mathbb{E}[Y|I=1] = \int_{-\infty}^\infty x f_{Y|I=1}(x)\,dx = \int_\theta^\infty x\frac{\alpha}{\theta}\left(\frac{x}{\theta}\right)^{-(\alpha+1)}\,dx = \alpha\theta^\alpha \int_\theta^\infty x^{-\alpha}\,dx$$

We see that the integral on the right hand side and therefore also $\mathbb{E}[Y]$ exist if and only if $\alpha > 1$. In this case we get

$$\mathbb{E}[Y|I=1] = \alpha\theta^\alpha \left[-\frac{1}{\alpha-1}x^{-(\alpha-1)}\right]_\theta^\infty = \alpha\theta^\alpha\frac{1}{\alpha-1}\theta^{-(\alpha-1)} = \theta\frac{\alpha}{\alpha-1}.$$

We conclude that, if $\alpha > 1$, we get

$$\mathbb{E}[Y] = p\theta\frac{\alpha}{\alpha-1}.$$

If $0 < \alpha \leq 1$, $\mathbb{E}[Y]$ does not exist.

Non-Life Insurance: Mathematics and Statistics

Solution sheet 2

Solution 2.1 Gaussian Distribution

(a) The moment generating function of $a + bX$ can be calculated as

$$M_{a+bX}(r) = \mathbb{E}\left[\exp\left\{r(a+bX)\right\}\right] = \exp\left\{ra\right\}\mathbb{E}\left[\exp\left\{rbX\right\}\right] = \exp\left\{ra\right\}M_X(rb),$$

for all $r \in \mathbb{R}$. Using the formula for the moment generating function of X given on the exercise sheet, we get

$$M_{a+bX}(r) = \exp\left\{ra\right\}\ \exp\left\{rb\mu + \frac{(rb)^2\sigma^2}{2}\right\} = \exp\left\{r(a+b\mu) + \frac{r^2b^2\sigma^2}{2}\right\},$$

which is equal to the moment generating function of a Gaussian random variable with expectation $a + b\mu$ and variance $b^2\sigma^2$. Since the moment generating function uniquely determines the distribution, we conclude that

$$a + bX \sim \mathcal{N}(a + b\mu, b^2\sigma^2).$$

(b) Using the independence of $X_1, \ldots, X_n$, the moment generating function of $Y = \sum_{i=1}^n X_i$ can be calculated as

$$M_Y(r) = \mathbb{E}\left[\exp\left\{rY\right\}\right] = \mathbb{E}\left[\exp\left\{r\sum_{i=1}^n X_i\right\}\right] = \prod_{i=1}^n \mathbb{E}\left[\exp\left\{rX_i\right\}\right] = \prod_{i=1}^n M_{X_i}(r),$$

for all $r \in \mathbb{R}$. Using the formula for the moment generating function of a Gaussian random variable given on the exercise sheet, we get

$$M_Y(r) = \prod_{i=1}^n \exp\left\{r\mu_i + \frac{r^2\sigma_i^2}{2}\right\} = \exp\left\{r\sum_{i=1}^n \mu_i + \frac{r^2\sum_{i=1}^n \sigma_i^2}{2}\right\},$$

which is equal to the moment generating function of a Gaussian random variable with expectation $\sum_{i=1}^n \mu_i$ and variance $\sum_{i=1}^n \sigma_i^2$. Since the moment generating function uniquely determines the distribution, we conclude that

$$\sum_{i=1}^n X_i \sim \mathcal{N}\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right).$$

Solution 2.2 Maximum Likelihood and Hypothesis Test

(a) Since $\log Y_1, \ldots, \log Y_8$ are independent random variables, the joint density $f_{\mu,\sigma^2}(x_1, \ldots, x_8)$ of $\log Y_1, \ldots, \log Y_8$ is given by product of the marginal densities of $\log Y_1, \ldots, \log Y_8$. We have

$$f_{\mu,\sigma^2}(x_1, \ldots, x_8) = \prod_{i=1}^8 \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right\},$$

since $\log Y_1, \ldots, \log Y_8$ are Gaussian random variables with mean μ and variance σ^2.

(b) By taking the logarithm, we get

$$\begin{aligned}\log f_{\mu,\sigma^2}(x_1,\ldots,x_8) &= \sum_{i=1}^{8} -\log\left(\sqrt{2\pi}\right) - \log(\sigma) - \frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2} \\ &= -8\log\left(\sqrt{2\pi}\right) - 8\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{8}(x_i-\mu)^2.\end{aligned}$$

(c) We have $\log f_{\mu,\sigma^2}(x_1,\ldots,x_8) < -8\log(\sigma)$ for all $\mu \in \mathbb{R}$. Hence, independently of μ, $\log f_{\mu,\sigma^2}(x_1,\ldots,x_8) \to -\infty$ if $\sigma^2 \to \infty$. Moreover, since for example $x_1 \neq x_2$, there exists a $c > 0$ with $\sum_{i=1}^{8}(x_i-\mu)^2 > c$ and thus $\log f_{\mu,\sigma^2}(x_1,\ldots,x_8) < -8\log(\sigma) - \frac{c}{2\sigma^2}$ for all $\mu \in \mathbb{R}$. Since $\frac{c}{2\sigma^2}$ goes much faster to ∞ than $8\log(\sigma)$ goes to $-\infty$ if $\sigma^2 \to 0$, we have $\log f_{\mu,\sigma^2}(x_1,\ldots,x_8) \to -\infty$ if $\sigma^2 \to 0$, independently of μ. Finally, if $\sigma^2 \in [c_1, c_2]$ for some $0 < c_1 < c_2$, we have $\log f_{\mu,\sigma^2}(x_1,\ldots,x_8) < -\frac{1}{2c_2}\sum_{i=1}^{8}(x_i-\mu)^2$. Hence, independently of the value of σ^2 in the interval $[c_1,c_2]$, $\log f_{\mu,\sigma^2}(x_1,\ldots,x_8) \to -\infty$ if $|\mu| \to \infty$. Since $\log f_{\mu,\sigma^2}(x_1,\ldots,x_8)$ is continuous in μ and σ^2, we can conclude that it attains its global maximum somewhere in $\mathbb{R} \times \mathbb{R}_{>0}$. Thus $\hat{\mu}$ and $\hat{\sigma}^2$ as defined on the exercise sheet have to satisfy the first order conditions

$$\frac{\partial}{\partial\mu}\log f_{\mu,\sigma^2}(x_1,\ldots,x_8)|_{(\mu,\sigma^2)=(\hat{\mu},\hat{\sigma}^2)} = 0 \qquad \text{and}$$
$$\frac{\partial}{\partial(\sigma^2)}\log f_{\mu,\sigma^2}(x_1,\ldots,x_8)|_{(\mu,\sigma^2)=(\hat{\mu},\hat{\sigma}^2)} = 0.$$

We calculate

$$\frac{\partial}{\partial\mu}\log f_{\mu,\sigma^2}(x_1,\ldots,x_8) = \frac{1}{\sigma^2}\sum_{i=1}^{8}(x_i-\mu),$$

which is equal to 0 if and only if $\mu = \frac{1}{8}\sum_{i=1}^{8} x_i$. Moreover, we have

$$\frac{\partial}{\partial(\sigma^2)}\log f_{\mu,\sigma^2}(x_1,\ldots,x_8) = -\frac{8}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{8}(x_i-\mu)^2 = \frac{1}{2\sigma^2}\left[-8 + \frac{1}{\sigma^2}\sum_{i=1}^{8}(x_i-\mu)^2\right],$$

which is equal to 0 if and only if $\sigma^2 = \frac{1}{8}\sum_{i=1}^{8}(x_i-\mu)^2$. Since there is only tuple in $\mathbb{R}\times\mathbb{R}_{>0}$ that satisfies the first order conditions, we conclude that

$$\hat{\mu} = \frac{1}{8}\sum_{i=1}^{8} x_i = 7 \qquad \text{and} \qquad \hat{\sigma}^2 = \frac{1}{8}\sum_{i=1}^{8}(x_i-\hat{\mu})^2 = \frac{1}{8}\sum_{i=1}^{8}(x_i-7)^2 = 7.$$

Note that the MLE $\hat{\sigma}^2$ is not unbiased. Indeed, if we replace $x_1,\ldots,x_8$ by independent Gaussian random variables $X_1,\ldots,X_8$ with expectation $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$ and write $\hat{\mu}$ for $\frac{1}{8}\sum_{i=1}^{8} X_i$, we can calculate

$$\mathbb{E}[\hat{\sigma}^2] = \mathbb{E}[\hat{\sigma}^2(X_1,\ldots,X_8)] = \mathbb{E}\left[\frac{1}{8}\sum_{i=1}^{8}(X_i-\hat{\mu})^2\right] = \frac{1}{8}\mathbb{E}\left[\sum_{i=1}^{8}(X_i^2 - 2X_i\hat{\mu} + \hat{\mu}^2)\right].$$

By noting that $\sum_{i=1}^{8} X_i = 8\hat{\mu}$ and that $\mathbb{E}[X_1^2] = \cdots = \mathbb{E}[X_8^2]$, we get

$$\mathbb{E}[\hat{\sigma}^2] = \frac{1}{8}\mathbb{E}\left[\sum_{i=1}^{8} X_i^2 - 2\cdot 8\cdot\hat{\mu}^2 + 8\hat{\mu}^2\right] = \mathbb{E}[X_1^2] - \mathbb{E}[\hat{\mu}^2] = \sigma^2 - \mathbb{E}[X_1]^2 - \mathrm{Var}(\hat{\mu}) + \mathbb{E}[\hat{\mu}]^2.$$

By inserting

$$\mathrm{Var}(\hat{\mu}) = \mathrm{Var}\left(\frac{1}{8}\sum_{i=1}^{8} X_i\right) = \left(\frac{1}{8}\right)^2 \sum_{i=1}^{8} \mathrm{Var}(X_i) = \frac{1}{8}\sigma^2 \qquad \text{and}$$

$$\mathbb{E}[\hat{\mu}]^2 = \mathbb{E}\left[\frac{1}{8}\sum_{i=1}^{8} X_i\right]^2 = \left(\frac{1}{8}\sum_{i=1}^{8} \mathbb{E}[X_i]\right)^2 = \mathbb{E}[X_1]^2,$$

we can conclude that

$$\mathbb{E}[\hat{\sigma}^2] = \sigma^2 - \mathbb{E}[X_1]^2 - \frac{1}{8}\sigma^2 + \mathbb{E}[X_1]^2 = \frac{7}{8}\sigma^2 \neq \sigma^2,$$

hence $\hat{\sigma}^2$ is not unbiased.

(d) Since our data is assumed to follow a Gaussian distribution and the variance is unknown, we perform a t-test. The test statistic is given by

$$T = T(\log Y_1, \ldots, \log Y_8) = \sqrt{8}\,\frac{\frac{1}{8}\sum_{i=1}^{8} \log Y_i - \mu}{\sqrt{S^2}},$$

where

$$S^2 = \frac{1}{7}\sum_{i=1}^{8}\left(\log Y_i - \frac{1}{8}\sum_{i=1}^{8} \log Y_i\right)^2.$$

Under H_0, T follows a Student-t distribution with 7 degrees of freedom. With the data given on the exercise sheet, the random variable S^2 attains the value

$$\frac{1}{7}\sum_{i=1}^{8}\left(x_i - \frac{1}{8}\sum_{i=1}^{8} x_i\right)^2 = \frac{1}{7}\sum_{i=1}^{8}(x_i - 7)^2 = 8,$$

and thus for T we get the observation

$$\sqrt{8}\,\frac{\frac{1}{8}\sum_{i=1}^{8} x_i - \mu}{\sqrt{S^2}} = \sqrt{8}\,\frac{7-6}{\sqrt{8}} = 1,$$

where we use that $\mu = 6$ under H_0. Now the probability under H_0 to observe a T that is at least as extreme as the observation 1 we got above, is

$$\mathbb{P}[|T| \geq 1] = \mathbb{P}[T \geq 1] + \mathbb{P}[T \leq -1] = 1 - \mathbb{P}[T < 1] + 1 - \mathbb{P}[T < 1] = 2 - 2\mathbb{P}[T < 1],$$

where we used the symmetry of the Student-t distribution around 0. The probability $\mathbb{P}[T < 1]$ is approximately 0.83, thus the p-value is given by

$$\mathbb{P}[|T| \geq 1] = 2 - 2\mathbb{P}[T < 1] \approx 2 - 2 \cdot 0.83 = 0.34.$$

This p-value is fairly high, hence we conclude that we can not reject the null hypothesis, for example, at significance level of 5% or 1%.

Solution 2.3 Variance Decomposition

By definition of the random variable X, the second moments exist. Hence, we have

$$\mathbb{E}[\mathrm{Var}(X|\mathcal{G})] = \mathbb{E}\left[\mathbb{E}[X^2|\mathcal{G}] - (\mathbb{E}[X|\mathcal{G}])^2\right] = \mathbb{E}[X^2] - \mathbb{E}\left[(\mathbb{E}[X|\mathcal{G}])^2\right]$$

and

$$\mathrm{Var}(\mathbb{E}[X|\mathcal{G}]) = \mathbb{E}\left[(\mathbb{E}[X|\mathcal{G}])^2\right] - \mathbb{E}\left[\mathbb{E}[X|\mathcal{G}]\right]^2 = \mathbb{E}\left[(\mathbb{E}[X|\mathcal{G}])^2\right] - \mathbb{E}[X]^2.$$

Combining these two results, we get

$$\mathbb{E}[\mathrm{Var}(X|\mathcal{G})] + \mathrm{Var}(\mathbb{E}[X|\mathcal{G}]) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathrm{Var}(X).$$

Non-Life Insurance: Mathematics and Statistics

Solution sheet 3

Solution 3.1 No-Claims Bonus

(a) We define the following events:

$$A = \{\text{“no claims in the last six years”}\},$$
$$B = \{\text{“no claims in the last three years but at least one claim in the last six years”}\},$$
$$C = \{\text{“at least one claim in the last three years”}\}.$$

Note that since the events A, B and C are disjoint and cover all possible outcomes, we have

$$\mathbb{P}[A] + \mathbb{P}[B] + \mathbb{P}[C] = 1,$$

i.e. it is sufficient to calculate two out of the three probabilities. Since the calculation of $\mathbb{P}[B]$ is slightly more involved, we will look at $\mathbb{P}[A]$ and $\mathbb{P}[C]$. Let $N_1, \ldots, N_6$ be the number of claims of the last six years of our considered car driver, where N_6 corresponds to the most recent year. By assumption, $N_1, \ldots, N_6$ are independent Poisson random variables with frequency parameter $\lambda = 0.2$. Therefore, we can calculate

$$\mathbb{P}[A] = \mathbb{P}[N_1 = 0, \ldots, N_6 = 0] = \prod_{i=1}^{6} \mathbb{P}[N_i = 0] = \prod_{i=1}^{6} \exp\{-\lambda\} = \exp\{-6\lambda\} = \exp\{-1.2\}$$

and, similarly,

$$\mathbb{P}[C] = 1 - \mathbb{P}[C^c] = 1 - \mathbb{P}[N_4 = 0, N_5 = 0, N_6 = 0] = 1 - \exp\{-3\lambda\} = 1 - \exp\{-0.6\}.$$

For the event B we get

$$\mathbb{P}[B] = 1 - \mathbb{P}[A] - \mathbb{P}[C] = 1 - \exp\{-1.2\} - (1 - \exp\{-0.6\}) = \exp\{-0.6\} - \exp\{-1.2\}.$$

Thus the expected proportion q of the base premium that is still paid after the grant of the no-claims bonus is given by

$$\begin{aligned} q &= 0.8 \cdot \mathbb{P}[A] + 0.9 \cdot \mathbb{P}[B] + 1 \cdot \mathbb{P}[C] \\ &= 0.8 \cdot \exp\{-1.2\} + 0.9 \cdot (\exp\{-0.6\} - \exp\{-1.2\}) + 1 - \exp\{-0.6\} \\ &\approx 0.915. \end{aligned}$$

If s denotes the surcharge on the base premium, then it has to satisfy the equation

$$q(1+s) \cdot \text{base premium} = \text{base premium},$$

which leads to

$$s = \frac{1}{q} - 1.$$

We conclude that the surcharge on the base premium is given by approximately 9.3%.

(b) We use the same notation as in (a). Since this time the calculation of $\mathbb{P}[B]$ is considerably more involved, we again look at $\mathbb{P}[A]$ and $\mathbb{P}[C]$. By assumption, conditionally given Θ, $N_1, \dots, N_6$ are independent Poisson random variables with frequency parameter $\Theta\lambda$, where $\lambda = 0.2$. Therefore, we can calculate

$$\begin{aligned}\mathbb{P}[A] &= \mathbb{P}\left[N_1 = 0, \dots, N_6 = 0\right] \\ &= \mathbb{E}\left[\mathbb{P}\left[N_1 = 0, \dots, N_6 = 0 | \Theta\right]\right] \\ &= \mathbb{E}\left[\prod_{i=1}^{6} \mathbb{P}\left[N_i = 0 | \Theta\right]\right] \\ &= \mathbb{E}\left[\prod_{i=1}^{6} \exp\{-\Theta\lambda\}\right] \\ &= \mathbb{E}\left[\exp\{-6\Theta\lambda\}\right] \\ &= M_\Theta(-6\lambda),\end{aligned}$$

where M_Θ denotes the moment generating function of Θ. Since $\Theta \sim \Gamma(1,1)$, M_Θ is given by

$$M_\Theta(r) = \frac{1}{1-r},$$

for all $r < 1$, which leads to

$$\mathbb{P}[A] = \frac{1}{1+6\lambda} = \frac{1}{2.2}.$$

Similarly, we get

$$\mathbb{P}[C] = 1 - \mathbb{P}[C^c] = 1 - \mathbb{P}\left[N_4 = 0, N_5 = 0, N_6 = 0\right] = 1 - \frac{1}{1+3\lambda} = 1 - \frac{1}{1.6} = \frac{0.6}{1.6}.$$

For the event B we get

$$\mathbb{P}[B] = 1 - \mathbb{P}[A] - \mathbb{P}[C] = 1 - \frac{1}{2.2} - \frac{0.6}{1.6} = \frac{1}{1.6} - \frac{1}{2.2}.$$

Thus the expected proportion q of the base premium that is still paid after the grant of the no-claims bonus is given by

$$\begin{aligned} q &= 0.8 \cdot \mathbb{P}[A] + 0.9 \cdot \mathbb{P}[B] + 1 \cdot \mathbb{P}[C] \\ &= 0.8 \cdot \frac{1}{2.2} + 0.9 \cdot \left(\frac{1}{1.6} - \frac{1}{2.2}\right) + \frac{0.6}{1.6} \\ &\approx 0.892. \end{aligned}$$

We conclude that the surcharge s on the base premium is given by

$$s = \frac{1}{q} - 1 \approx 12.1\%,$$

which is considerably bigger than in (a).

Solution 3.2 Central Limit Theorem

Let σ^2 be the variance of the claim sizes and $x > 0$. Then we have

$$\begin{aligned}\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} Y_i - \mu\right| < \frac{x}{\sqrt{n}}\right] &= \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i - \mu < \frac{x}{\sqrt{n}}\right] - \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i - \mu \leq -\frac{x}{\sqrt{n}}\right] \\ &= \mathbb{P}\left[\sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n} Y_i - \mu}{\sigma} < \frac{x}{\sigma}\right] - \mathbb{P}\left[\sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n} Y_i - \mu}{\sigma} \leq -\frac{x}{\sigma}\right] \\ &= \mathbb{P}\left[Z_n < \frac{x}{\sigma}\right] - \mathbb{P}\left[Z_n \leq -\frac{x}{\sigma}\right],\end{aligned}$$

where

$$Z_n = \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^n Y_i - \mu}{\sigma}.$$

According to the Central Limit Theorem, Z_n converges in distribution to a standard Gaussian random variable. Hence, if we write Φ for the distribution function of a standard Gaussian random variable, we have the approximation

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^n Y_i - \mu\right| < \frac{x}{\sqrt{n}}\right] \approx \Phi\left(\frac{x}{\sigma}\right) - \Phi\left(-\frac{x}{\sigma}\right).$$

On the one hand, as we are interested in a probabilty of at least 95%, we have to choose $x > 0$ such that

$$\Phi\left(\frac{x}{\sigma}\right) - \Phi\left(-\frac{x}{\sigma}\right) = 0.95,$$

which implies

$$\frac{x}{\sigma} = 1.96.$$

It follows that

$$x = 1.96 \cdot \sigma = 1.96 \cdot \operatorname{Vco}(Y_1) \cdot \mu = 1.96 \cdot 4 \cdot \mu. \tag{1}$$

On the other hand, as we want the deviation of the empirical mean from μ to be less than 1%, we set

$$\frac{x}{\sqrt{n}} = 0.01 \cdot \mu,$$

which implies

$$n = \frac{x^2}{0.01^2 \cdot \mu^2}. \tag{2}$$

Combining (1) and (2), we conclude

$$n = \frac{(1.96 \cdot 4 \cdot \mu)^2}{0.01^2 \cdot \mu^2} = 1.96^2 \cdot 4^2 \cdot 10'000 = 614'656.$$

Solution 3.3 Compound Binomial Distribution

For $\tilde{S} \sim \operatorname{CompBinom}(\tilde{v}, \tilde{p}, \tilde{G})$ with the random variable $\tilde{Y}_1$ having distribution function $\tilde{G}$ and moment generating function $M_{\tilde{Y}_1}$, the moment generating function $M_{\tilde{S}}$ of $\tilde{S}$ is given by

$$M_{\tilde{S}}(r) = \left(\tilde{p} M_{\tilde{Y}_1}(r) + 1 - \tilde{p}\right)^{\tilde{v}},$$

for all $r \in \mathbb{R}$ for which $M_{\tilde{Y}_1}$ is defined. We calculate the moment generating function of S_{lc} and show that it is exactly of the form given above. Since $S_{\mathrm{lc}} \geq 0$ almost surely, its moment generating function is defined at least for all $r < 0$. Thus, for $r < 0$, we have

$$\begin{aligned}
M_{S_{\mathrm{lc}}}(r) &= \mathbb{E}\left[\exp\left\{r \sum_{i=1}^N Y_i \; 1_{\{Y_i > M\}}\right\}\right] \\
&= \mathbb{E}\left[\prod_{i=1}^N \exp\left\{r Y_i \; 1_{\{Y_i > M\}}\right\}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\prod_{i=1}^N \exp\left\{r Y_i \; 1_{\{Y_i > M\}}\right\} \middle| N\right]\right] \\
&= \mathbb{E}\left[\prod_{i=1}^N \mathbb{E}\left[\exp\left\{r Y_i \; 1_{\{Y_i > M\}}\right\}\right]\right],
\end{aligned}$$

where in the third equality we used the tower property of conditional expectation and in the fourth equality the independence between N and Y_i. For the inner expectation we get

$$\begin{aligned}\mathbb{E}\left[\exp\left\{rY_i\ 1_{\{Y_i>M\}}\right\}\right] &= \mathbb{E}\left[\exp\left\{rY_i\right\}\cdot 1_{\{Y_i>M\}} + 1_{\{Y_i\le M\}}\right] \\ &= \mathbb{E}\left[\exp\left\{rY_i\right\}|Y_i>M\right]\mathbb{P}[Y_i>M] + \mathbb{P}[Y_i\le M] \\ &= \mathbb{E}\left[\exp\left\{rY_i\right\}|Y_i>M\right][1-G(M)] + G(M).\end{aligned}$$

First note that the distribution function of the random variable $Y_i \mid Y_i > M$ is G_{lc}. Moreover, since $Y_i \mid Y_i > M$ is greater than 0 almost surely, its moment generating function $M_{Y_1|Y_1>M}$ is defined for all $r<0$ and thus we can write

$$\mathbb{E}\left[\exp\left\{rY_i\ 1_{\{Y_i>M\}}\right\}\right] = M_{Y_1|Y_1>M}(r)[1-G(M)] + G(M).$$

Hence we get

$$\begin{aligned}M_{S_{\mathrm{lc}}}(r) &= \mathbb{E}\left[\prod_{i=1}^{N}\left(M_{Y_1|Y_1>M}(r)[1-G(M)]+G(M)\right)\right] \\ &= \mathbb{E}\left[\left(M_{Y_1|Y_1>M}(r)[1-G(M)]+G(M)\right)^N\right] \\ &= \mathbb{E}\left[\exp\left\{N\log\left(M_{Y_1|Y_1>M}(r)[1-G(M)]+G(M)\right)\right\}\right] \\ &= M_N(\rho),\end{aligned}$$

where M_N is the moment generating function of N and

$$\rho = \log\left(M_{Y_1|Y_1>M}(r)[1-G(M)]+G(M)\right).$$

Since we have $N\sim\mathrm{Binom}(v,p)$, $M_N(r)$ is given by

$$M_N(r) = (p\exp\{r\}+1-p)^v.$$

Therefore, we get

$$\begin{aligned}M_{S_{\mathrm{lc}}}(r) &= \left[p\left(M_{Y_1|Y_1>M}(r)[1-G(M)]+G(M)\right)+1-p\right]^v \\ &= \left(p[1-G(M)]M_{Y_1|Y_1>M}(r)+1-p[1-G(M)]\right)^v.\end{aligned}$$

Applying Lemma 1.3 of the lecture notes, we conclude that $S_{\mathrm{lc}}\sim\mathrm{CompBinom}(\tilde{v},\tilde{p},\tilde{G})$ with $\tilde{v}=v$, $\tilde{p}=p[1-G(M)]$ and $\tilde{G}=G_{\mathrm{lc}}$.

Non-Life Insurance: Mathematics and Statistics

Solution sheet 4

Solution 4.1 Poisson Model and Negative-Binomial Model

(a) Let $v = v_1 = \cdots = v_{10} = 10'000$. In the Poisson model we assume that $N_1, \ldots, N_{10}$ are independent with $N_t \sim \text{Poi}(\lambda v_t)$ for all $t \in \{1, \ldots, 10\}$. We use Estimator 2.32 of the lecture notes to estimate the claims frequency parameter λ by

$$\hat{\lambda}_{10}^{\text{MLE}} = \frac{\sum_{t=1}^{10} N_t}{\sum_{t=1}^{10} v_t} = \frac{\sum_{t=1}^{10} N_t}{10v} = \frac{10'224}{100'000} \approx 10.22\%.$$

Note that a random variable $N \sim \text{Poi}(\lambda v)$ can be understood as

$$N \overset{(\text{d})}{=} \sum_{i=1}^{v} N_i, \tag{1}$$

where $N_1, \ldots, N_v$ are independent random variables that all follow a $\text{Poi}(\lambda)$-distribution. If we define $\hat{\lambda} = N/v$, then we have

$$\mathbb{E}\left[\hat{\lambda}\right] = \mathbb{E}\left[\frac{N}{v}\right] = \frac{\mathbb{E}[N]}{v} = \frac{\lambda v}{v} = \lambda,$$

hence $\hat{\lambda}$ can be seen as an estimator for λ. Moreover, we have

$$\text{Var}\left(\hat{\lambda}\right) = \text{Var}\left(\frac{N}{v}\right) = \frac{\text{Var}(N)}{v^2} = \frac{\lambda v}{v^2} = \frac{\lambda}{v}$$

and, because of (1), we can use the Central Limit Theorem to get

$$\frac{N/v - \mathbb{E}\left[N/v\right]}{\sqrt{\text{Var}\left(N/v\right)}} = \frac{\hat{\lambda} - \lambda}{\sqrt{\lambda/v}} \longrightarrow Z,$$

as $v \to \infty$, where Z is a random variable following a standard normal distribution. Hence, we have the approximation

$$\mathbb{P}\left[\hat{\lambda} - \sqrt{\frac{\lambda}{v}} \le \lambda \le \hat{\lambda} + \sqrt{\frac{\lambda}{v}}\right] = \mathbb{P}\left[-1 \le \frac{\hat{\lambda} - \lambda}{\sqrt{\lambda/v}} \le 1\right] \approx \mathbb{P}(-1 \le Z \le 1) \approx 0.7,$$

i.e. with a probability of roughly 70%, λ lies in the interval $\left[\hat{\lambda} - \sqrt{\lambda/v}, \hat{\lambda} + \sqrt{\lambda/v}\right]$. Since a confidence interval for λ is not allowed to depend on λ itself, we also replace it by the estimator $\hat{\lambda}$ to get an approximate, roughly 70%-confidence interval $\left[\hat{\lambda} - \sqrt{\hat{\lambda}/v}, \hat{\lambda} + \sqrt{\hat{\lambda}/v}\right]$ for λ. If we look at the estimator $\hat{\lambda}_{10}^{\text{MLE}}$ as the random variable $\left(\sum_{t=1}^{10} N_t\right) / (10v)$, we see that

$$\mathbb{E}\left[\hat{\lambda}_{10}^{\text{MLE}}\right] = \frac{\sum_{t=1}^{10} \mathbb{E}[N_t]}{10v} = \frac{\sum_{t=1}^{10} \lambda v_t}{10v} = \lambda = \mathbb{E}\left[\hat{\lambda}\right]$$

and

$$\text{Var}\left(\hat{\lambda}_{10}^{\text{MLE}}\right) = \frac{\sum_{t=1}^{10} \text{Var}(N_t)}{(10v)^2} = \frac{\sum_{t=1}^{10} \lambda v_t}{(10v)^2} = \frac{\lambda}{10v} < \frac{\lambda}{v} = \text{Var}\left(\hat{\lambda}\right).$$

Because of the smaller variance it makes sense to replace $\hat{\lambda}$ by $\hat{\lambda}_{10}^{\text{MLE}}$ to get the approximate, roughly 70%-confidence interval

$$\left[\hat{\lambda}_{10}^{\text{MLE}} - \sqrt{\frac{\hat{\lambda}_{10}^{\text{MLE}}}{v}}, \hat{\lambda}_{10}^{\text{MLE}} + \sqrt{\frac{\hat{\lambda}_{10}^{\text{MLE}}}{v}}\right] \approx [9.90\%, 10.54\%]$$

for λ. If we define $\lambda_t = N_t/v_t$ for all $t \in \{1,\dots,10\}$, we have the following observations $\lambda_1,\dots,\lambda_{10}$ of the frequency parameter λ:

t	1	2	3	4	5	6	7	8	9	10
$\lambda_t = \frac{N_t}{v_t}$	10%	9.97%	9.85%	9.89%	10.56%	10.70%	9.94%	9.86%	10.93%	10.54%

Table 1: Observed claims frequencies $\lambda_t = N_t/v_t$.

We observe that instead of the expected, roughly seven observations, only four observations lie in the estimated confidence interval. We conclude that the assumption of having Poisson distributions might not be reasonable.

(b) By equation (2.8) of the lecture notes, the test statistic $\hat{\chi}^*$ is given by

$$\hat{\chi}^* = \sum_{t=1}^{10} v_t \frac{\left(N_t/v_t - \hat{\lambda}_{10}^{\text{MLE}}\right)^2}{\hat{\lambda}_{10}^{\text{MLE}}}$$

and is approximately χ^2-distributed with $10-1=9$ degrees of freedom. By inserting the numbers and $\hat{\lambda}_{10}^{\text{MLE}}$ calculated in (a), we get

$$\hat{\chi}^* \approx 14.84.$$

The probability that a random variable with a χ^2-distribution with 9 degrees of freedom is greater than 14.84 is approximately equal to 9.55%. Hence we can reject the null hypothesis of having Poisson distributions only at significance levels that are higher than 9.55%. In particular, we can not reject the null hypothesis at the significance level of 5%.

(c) As in part (a), let $v = v_1 = \dots = v_{10} = 10\text{'}000$. In the negative-binomial model we assume that $N_1,\dots,N_{10}$ are independent with $N_t \sim \text{Poi}(\Theta_t \lambda v_t)$ for all $t \in \{1,\dots,10\}$, where $\Theta_1,\dots,\Theta_{10} \overset{\text{i.i.d.}}{\sim} \Gamma(\gamma,\gamma)$ for some $\gamma > 0$. We use Estimator 2.28 of the lecture notes to estimate the claims frequency parameter λ by

$$\hat{\lambda}_{10}^{\text{NB}} = \frac{\sum_{t=1}^{10} N_t}{\sum_{t=1}^{10} v_t} = \frac{\sum_{t=1}^{10} N_t}{10v} = \frac{10\text{'}224}{100\text{'}000} \approx 10.22\%.$$

As in equation (2.7) of the lecture notes, we define

$$\hat{V}_{10}^2 = \frac{1}{9}\sum_{t=1}^{10} v_t \left(\frac{N_t}{v_t} - \hat{\lambda}_{10}^{\text{NB}}\right) \approx 16.9\%.$$

Now we can use Estimator 2.30 of the lecture notes to estimate the dispersion parameter γ by

$$\hat{\gamma}_{10}^{\text{NB}} = \frac{\left(\hat{\lambda}_{10}^{\text{NB}}\right)^2}{\hat{V}_{10}^2 - \hat{\lambda}_{10}^{\text{NB}}} \frac{1}{9}\left(\sum_{t=1}^{10} v_t - \frac{\sum_{t=1}^{10} v_t^2}{\sum_{t=1}^{10} v_t}\right) = \frac{\left(\hat{\lambda}_{10}^{\text{NB}}\right)^2}{\hat{V}_{10}^2 - \hat{\lambda}_{10}^{\text{NB}}} \frac{\left(10v - \frac{10v^2}{10v}\right)}{9} = \frac{\left(\hat{\lambda}_{10}^{\text{NB}}\right)^2 v}{\hat{V}_{10}^2 - \hat{\lambda}_{10}^{\text{NB}}} \approx 1576.15.$$

For a random variable $N \sim \text{Poi}(\Theta\lambda v)$, conditionally given Θ, we have

$$\mathbb{E}\left[\frac{N}{v}\right] = \frac{\mathbb{E}[N]}{v} = \frac{\mathbb{E}[\mathbb{E}[N|\Theta]]}{v} = \frac{\mathbb{E}[\Theta\lambda v]}{v} = \frac{\lambda v}{v} = \lambda,$$

since $\mathbb{E}[\Theta] = 1$, and

$$\text{Var}\left(\frac{N}{v}\right) = \frac{\mathbb{E}[\text{Var}(N|\Theta)] + \text{Var}(\mathbb{E}[N|\Theta])}{v^2} = \frac{\mathbb{E}[\Theta\lambda v] + \text{Var}(\Theta\lambda v)}{v^2} = \frac{\lambda v + \frac{\lambda^2 v^2}{\gamma}}{v^2} = \frac{\lambda + \frac{\lambda^2 v}{\gamma}}{v},$$

since $\text{Var}(\Theta) = 1/\gamma$. Similarly as in the Poisson case in part (a), we get the approximate, roughly 70%-confidence interval

$$\left[\hat{\lambda}_{10}^{\text{NB}} - \sqrt{\frac{\hat{\lambda}_{10}^{\text{NB}} + \left(\hat{\lambda}_{10}^{\text{NB}}\right)^2 v/\hat{\gamma}_{10}^{\text{NB}}}{v}}, \hat{\lambda}_{10}^{\text{NB}} + \sqrt{\frac{\hat{\lambda}_{10}^{\text{NB}} + \left(\hat{\lambda}_{10}^{\text{NB}}\right)^2 v/\hat{\gamma}_{10}^{\text{NB}}}{v}}\right] \approx [9.81\%, 10.63\%].$$

for λ. Looking at the observations $\lambda_1, \ldots, \lambda_{10}$ given in Table 1 above, we see that eight of them lie in the estimated confidence interval, which is clearly better than in the Poisson case in part (a). In conclusion, the negative-binomial model seems more reasonable than the Poisson model.

Solution 4.2 Compound Poisson Distribution

(a) Since $S \sim \text{CompPoi}(\lambda v, G)$, we can write S as

$$S = \sum_{i=1}^{N} Y_i,$$

where $N \sim \text{Poi}(\lambda v)$, $Y_1, Y_2, \ldots$ are i.i.d. with distribution function G and N and $Y_1, Y_2, \ldots$ are independent. Now we can define S_{sc}, S_{mc} and S_{lc} as

$$S_{\text{sc}} = \sum_{i=1}^{N} Y_i 1_{\{Y_i \leq 1'000\}}, \quad S_{\text{mc}} = \sum_{i=1}^{N} Y_i 1_{\{1'000 < Y_i \leq 1'000'000\}} \quad \text{and} \quad S_{\text{lc}} = \sum_{i=1}^{N} Y_i 1_{\{Y_i > 1'000'000\}}.$$

(b) Note that according to Table 2 given on the exercise sheet, we have

$$\begin{aligned}
\mathbb{P}[Y_1 \leq 1'000] &= \mathbb{P}[Y = 100] + \mathbb{P}[Y = 300] + \mathbb{P}[Y = 500] = \frac{3}{20} + \frac{4}{20} + \frac{3}{20} = \frac{1}{2}, \\
\mathbb{P}[1'000 < Y_1 \leq 1'000'000] &= \mathbb{P}[Y = 6'000] + \mathbb{P}[Y = 100'000] + \mathbb{P}[Y = 500'000] \\
&= \frac{2}{15} + \frac{2}{15} + \frac{1}{15} \\
&= \frac{1}{3} \quad \text{and} \\
\mathbb{P}[Y_1 > 1'000'000] &= 1 - \mathbb{P}[Y_1 \leq 1'000'000] = 1 - \frac{1}{2} - \frac{1}{3} = \frac{1}{6}.
\end{aligned}$$

Thus, using Theorem 2.14 of the lecture notes (disjoint decomposition of compound Poisson distributions), we get

$$S_{\text{sc}} \sim \text{CompPoi}\left(\frac{\lambda v}{2}, G_{\text{sc}}\right), \quad S_{\text{mc}} \sim \text{CompPoi}\left(\frac{\lambda v}{3}, G_{\text{mc}}\right) \quad \text{and} \quad S_{\text{lc}} \sim \text{CompPoi}\left(\frac{\lambda v}{6}, G_{\text{lc}}\right),$$

where

$$\begin{aligned} G_{\mathrm{sc}}(y) &= \mathbb{P}[Y_1 \le y | Y_1 \le 1'000], \\ G_{\mathrm{mc}}(y) &= \mathbb{P}[Y_1 \le y | 1'000 < Y_1 \le 1'000'000] \quad \text{and} \\ G_{\mathrm{lc}}(y) &= \mathbb{P}[Y_1 \le y | Y_1 > 1'000'000] \end{aligned}$$

for all $y \in \mathbb{R}$. In particular, for a random variable Y_{sc} having distribution function G_{sc}, we have

$$\begin{aligned} \mathbb{P}[Y_{\mathrm{sc}} = 100] &= \frac{\mathbb{P}[Y = 100]}{\mathbb{P}[Y_1 \le 1'000]} = \frac{3/20}{1/2} = \frac{3}{10}, \\ \mathbb{P}[Y_{\mathrm{sc}} = 300] &= \frac{\mathbb{P}[Y = 300]}{\mathbb{P}[Y_1 \le 1'000]} = \frac{4/20}{1/2} = \frac{4}{10} \quad \text{and} \\ \mathbb{P}[Y_{\mathrm{sc}} = 500] &= \frac{\mathbb{P}[Y = 500]}{\mathbb{P}[Y_1 \le 1'000]} = \frac{3/20}{1/2} = \frac{3}{10} \end{aligned}$$

Analogously, for random variables Y_{mc} and Y_{lc} having distribution functions G_{mc} and G_{lc}, respectively, we get

$$\mathbb{P}[Y_{\mathrm{mc}} = 6'000] = \frac{2}{5}, \quad \mathbb{P}[Y_{\mathrm{mc}} = 100'000] = \frac{2}{5} \quad \text{and} \quad \mathbb{P}[Y_{\mathrm{mc}} = 500'000] = \frac{1}{5},$$

as well as

$$\mathbb{P}[Y_{\mathrm{lc}} = 2'000'000] = \frac{1}{2}, \quad \mathbb{P}[Y_{\mathrm{lc}} = 5'000'000] = \frac{1}{4} \quad \text{and} \quad \mathbb{P}[Y_{\mathrm{lc}} = 10'000'000] = \frac{1}{4}.$$

(c) According to Theorem 2.14 of the lecture notes, S_{sc}, S_{mc} and S_{lc} are independent.

(d) In order to find $\mathbb{E}[S_{\mathrm{sc}}]$, we need $\mathbb{E}[Y_{\mathrm{sc}}]$, which can be calculated as

$$\mathbb{E}[Y_{\mathrm{sc}}] = 100 \cdot \mathbb{P}[Y_{\mathrm{sc}} = 100] + 300 \cdot \mathbb{P}[Y_{\mathrm{mc}} = 300] + 500 \cdot \mathbb{P}[Y_{\mathrm{lc}} = 500] = \frac{300}{10} + \frac{1200}{10} + \frac{1500}{10} = 300.$$

Now we can apply Proposition 2.11 of the lecture notes to get

$$\mathbb{E}[S_{\mathrm{sc}}] = \frac{\lambda v}{2} \mathbb{E}[Y_{\mathrm{sc}}] = 0.3 \cdot 300 = 90.$$

Similarly, we get

$$\mathbb{E}[Y_{\mathrm{mc}}] = 142'400 \quad \text{and} \quad \mathbb{E}[Y_{\mathrm{lc}}] = 4'750'000.$$

Thus we find

$$\mathbb{E}[S_{\mathrm{mc}}] = \frac{\lambda v}{3} \mathbb{E}[Y_{\mathrm{mc}}] = 28'480 \quad \text{and} \quad \mathbb{E}[S_{\mathrm{lc}}] = \frac{\lambda v}{6} \mathbb{E}[Y_{\mathrm{lc}}] = 475'000.$$

Since $S = S_{\mathrm{sc}} + S_{\mathrm{mc}} + S_{\mathrm{lc}}$, we get

$$\mathbb{E}[S] = \mathbb{E}[S_{\mathrm{sc}}] + \mathbb{E}[S_{\mathrm{mc}}] + \mathbb{E}[S_{\mathrm{lc}}] = 503'570.$$

In order to find $\mathrm{Var}(S_{\mathrm{sc}})$, we need $\mathbb{E}[Y_{\mathrm{sc}}^2]$, which can be calculated as

$$\begin{aligned} \mathbb{E}[Y_{\mathrm{sc}}^2] &= 100^2 \cdot \mathbb{P}[Y_{\mathrm{sc}} = 100] + 300^2 \cdot \mathbb{P}[Y_{\mathrm{mc}} = 300] + 500^2 \cdot \mathbb{P}[Y_{\mathrm{lc}} = 500] \\ &= \frac{30'000}{10} + \frac{360'000}{10} + \frac{750'000}{10} = 114'000. \end{aligned}$$

Now we can apply Proposition 2.11 of the lecture notes to get

$$\mathrm{Var}(S_{\mathrm{sc}}) = \frac{\lambda v}{2} \mathbb{E}[Y_{\mathrm{sc}}^2] = 0.3 \cdot 114'000 = 34'200.$$

Similarly, we get

$$\mathbb{E}[Y_{\mathrm{mc}}^2] = 54\text{'}014\text{'}400\text{'}000 \quad \text{and} \quad \mathbb{E}[Y_{\mathrm{lc}}^2] = 33\text{'}250\text{'}000\text{'}000\text{'}000.$$

Thus we find

$$\mathrm{Var}(S_{\mathrm{mc}}) = \frac{\lambda v}{3}\mathbb{E}[Y_{\mathrm{mc}}^2] = 10\text{'}802\text{'}880\text{'}000 \quad \text{and} \quad \mathrm{Var}(S_{\mathrm{lc}}) = \frac{\lambda v}{6}\mathbb{E}[Y_{\mathrm{lc}}^2] = 3\text{'}325\text{'}000\text{'}000\text{'}000.$$

Since $S = S_{\mathrm{sc}} + S_{\mathrm{mc}} + S_{\mathrm{lc}}$ and S_{sc}, S_{mc} and S_{lc} are independent, we get

$$\mathrm{Var}(S) = \mathrm{Var}(S_{\mathrm{sc}}) + \mathrm{Var}(S_{\mathrm{mc}}) + \mathrm{Var}(S_{\mathrm{lc}}) = 3\text{'}335\text{'}802\text{'}914\text{'}200.$$

(e) First, we define the random variable N_{lc} as

$$N_{\mathrm{lc}} \sim \mathrm{Poi}\left(\frac{\lambda v}{6}\right).$$

The probability that the total claim in the large claims layer exceeds 5 millions can be calculated by looking at the complement, i.e. at the probability that the total claim in the large claims layer does not exceed 5 millions. Since with three claims in the large claims layer we already exceed 5 millions, it is enough to consider only up to two claims. Then we get

$$\begin{aligned}\mathbb{P}\left[S_{\mathrm{lc}} \le 5\text{'}000\text{'}000\right] &= \mathbb{P}[N_{\mathrm{lc}} = 0] + \mathbb{P}[N_{\mathrm{lc}} = 1]\mathbb{P}[Y_{\mathrm{lc}} \le 5\text{'}000\text{'}000] + \mathbb{P}[N_{\mathrm{lc}} = 2]\mathbb{P}[Y_{\mathrm{lc}} = 2\text{'}000\text{'}000]^2 \\ &= \exp\left\{-\frac{\lambda v}{6}\right\} + \exp\left\{-\frac{\lambda v}{6}\right\}\frac{\lambda v}{6}\left(\frac{1}{2}+\frac{1}{4}\right) + \exp\left\{-\frac{\lambda v}{6}\right\}\left(\frac{\lambda v}{6}\right)^2 \frac{1}{2}\frac{1}{4} \\ &= \exp\{-0.1\}\left(1 + 0.075 + 0.00125\right) \\ &\approx 97.4\%.\end{aligned}$$

Hence we can conclude

$$\mathbb{P}\left[S_{\mathrm{lc}} > 5\text{'}000\text{'}000\right] = 1 - \mathbb{P}\left[S_{\mathrm{lc}} \le 5\text{'}000\text{'}000\right] \approx 2.6\%.$$

Solution 4.3 Method of Moments

If $Y \sim \Gamma(\gamma, c)$, then we have

$$\mathbb{E}[Y] = \frac{\gamma}{c} \quad \text{and} \quad \mathrm{Var}(Y) = \frac{\gamma}{c^2}.$$

We define the sample mean $\hat{\mu}_8$ and the sample variance $\hat{\sigma}_8^2$ of the eight observations given on the exercise sheet as

$$\hat{\mu}_8 = \frac{1}{8}\sum_{i=1}^{8} x_i = \frac{64}{8} = 8 \quad \text{and} \quad \hat{\sigma}_8^2 = \frac{1}{7}\sum_{i=1}^{8}(x_i - \hat{\mu}_8)^2 = \frac{28}{7} = 4.$$

The method of moments estimates $(\hat{\gamma}, \hat{c})$ of (γ, c) are defined to be those values that solve the equations

$$\hat{\mu}_8 = \frac{\hat{\gamma}}{\hat{c}} \quad \text{and} \quad \hat{\sigma}_8^2 = \frac{\hat{\gamma}}{\hat{c}^2}.$$

We see that $\hat{\gamma} = \hat{\mu}_8 \hat{c}$ and thus

$$\hat{\sigma}_8^2 = \frac{\hat{\mu}_8 \hat{c}}{\hat{c}^2} = \frac{\hat{\mu}_8}{\hat{c}},$$

which is equivalent to

$$\hat{c} = \frac{\hat{\mu}_8}{\hat{\sigma}_8^2} = \frac{8}{4} = 2.$$

Moreover, we get

$$\hat{\gamma} = \frac{\hat{\mu}_8^2}{\hat{\sigma}_8^2} = \frac{64}{4} = 16.$$

Thus we conclude that the method of moments estimate are given by $(\hat{\gamma}, \hat{c}) = (16, 2)$.

Non-Life Insurance: Mathematics and Statistics

Solution sheet 5

Solution 5.1 Kolmogorov-Smirnov Test

The distribution function G_0 of a Weibull distribution with shape parameter $\tau = \frac{1}{2}$ and scale parameter $c = 1$ is given by

$$G_0(y) = 1 - \exp\{-y^{1/2}\}$$

for all $y \geq 0$. Note that since G_0 is continuous, we are allowed to apply a Kolmogorov-Smirnov test. If $x = (-\log u)^2$ for some $u \in (0,1)$, we have

$$G_0(x) = 1 - \exp\left\{-\left[(-\log u)^2\right]^{1/2}\right\} = 1 - \exp\{\log u\} = 1 - u.$$

Hence, if we apply G_0 to $x_1, \ldots, x_5$, we get

$$G_0(x_1) = \frac{2}{40}, \quad G_0(x_2) = \frac{3}{40}, \quad G_0(x_3) = \frac{5}{40}, \quad G_0(x_4) = \frac{6}{40}, \quad G_0(x_5) = \frac{30}{40}.$$

Moreover, the empirical distribution function $\hat{G}_5$ of the sample $x_1, \ldots, x_5$ is given by

$$\hat{G}_5(y) = \begin{cases} 0 & \text{if } y < x_1, \\ 1/5 & \text{if } x_1 \leq y < x_2, \\ 2/5 & \text{if } x_2 \leq y < x_3, \\ 3/5 & \text{if } x_3 \leq y < x_4, \\ 4/5 & \text{if } x_4 \leq y < x_5, \\ 1 & \text{if } y \geq x_5. \end{cases}$$

Now the Kolmogorov-Smirnov test statistic D_5 is defined as

$$D_5 = \sup_{y \in \mathbb{R}} \left| \hat{G}_5(y) - G_0(y) \right|.$$

Since G_0 is continuous and strictly monotonically increasing with range $(0,1)$ and $\hat{G}_5$ is piecewise constant and attains both the values 0 and 1, it is sufficient to consider the discontinuities of $\hat{G}_5$ to find D_5. We define

$$f(s-) = \lim_{r \nearrow s} f(r)$$

for all $s \in \mathbb{R}$, where the function f stands for G_0 and $\hat{G}_5$. Since G_0 is continuous, we have $G_0(s-) = G_0(s)$ for all $s \in \mathbb{R}$. The values of G_0 and $\hat{G}_5$ and their differences can be summarized in the following table:

x_i, x_i-	x_1-	x_1	x_2-	x_2	x_3-	x_3	x_4-	x_4	x_5-	x_5
$\hat{G}_5(\cdot)$	0	8/40	8/40	16/40	16/40	24/40	24/40	32/40	32/40	1
$G_0(\cdot)$	2/40	2/40	3/40	3/40	5/40	5/40	6/40	6/40	30/40	30/40
$\lvert\hat{G}_5(\cdot) - G_0(\cdot)\rvert$	2/40	6/40	5/40	13/40	11/40	19/40	18/40	26/40	2/40	10/40

From this table we see that $D_5 = 26/40 = 0.65$. Let $q = 5\%$. By writing $K^{\leftarrow}(1-q)$ for the $(1-q)$-quantile of the Kolmogorov distribution, we have $K^{\leftarrow}(1-q) = 1.36$. Since

$$\frac{K^{\leftarrow}(1-q)}{\sqrt{5}} \approx 0.61 < 0.65 = D_5,$$

we can reject the null hypothesis of having a Weibull distribution with shape parameter $\tau = \frac{1}{2}$ and scale parameter $c = 1$ as claim size distribution.

Solution 5.2 Large Claims

(a) The density of a Pareto distribution with threshold $\theta = 50$ and tail index $\alpha > 0$ is given by

$$f(x) = f_\alpha(x) = \frac{\alpha}{\theta}\left(\frac{x}{\theta}\right)^{-(\alpha+1)}$$

for all $x \geq \theta$. Using the independence of $Y_1, \ldots, Y_n$, the joint likelihood function $\mathcal{L}_{\mathbf{Y}}(\alpha)$ for the observation $\mathbf{Y} = (Y_1, \ldots, Y_n)$ can be written as

$$\mathcal{L}_{\mathbf{Y}}(\alpha) = \prod_{i=1}^n f_\alpha(Y_i) = \prod_{i=1}^n \frac{\alpha}{\theta}\left(\frac{Y_i}{\theta}\right)^{-(\alpha+1)} = \prod_{i=1}^n \alpha\theta^\alpha Y_i^{-(\alpha+1)},$$

whereas the joint log-likelihood function $\ell_{\mathbf{Y}}(\alpha)$ is given by

$$\ell_{\mathbf{Y}}(\alpha) = \log \mathcal{L}_{\mathbf{Y}}(\alpha) = \sum_{i=1}^n \log\alpha + \alpha\log\theta - (\alpha+1)\log Y_i = n\log\alpha + n\alpha\log\theta - (\alpha+1)\sum_{i=1}^n \log Y_i.$$

Now the MLE $\hat{\alpha}_n^{\mathrm{MLE}}$ is defined as

$$\hat{\alpha}_n^{\mathrm{MLE}} = \arg\max_{\alpha>0} \mathcal{L}_{\mathbf{Y}}(\alpha) = \arg\max_{\alpha>0} \ell_{\mathbf{Y}}(\alpha).$$

Calculating the first and the second derivative of $\ell_{\mathbf{Y}}(\alpha)$ with respect to α, we get

$$\frac{\partial}{\partial\alpha}\ell_{\mathbf{Y}}(\alpha) = \frac{n}{\alpha} + n\log\theta - \sum_{i=1}^n \log Y_i \quad \text{and}$$
$$\frac{\partial^2}{\partial\alpha^2}\ell_{\mathbf{Y}}(\alpha) = \frac{\partial}{\partial\alpha}\left(\frac{n}{\alpha} + n\log\theta - \sum_{i=1}^n \log Y_i\right) = -\frac{n}{\alpha^2} < 0,$$

for all $\alpha > 0$, from which we can conclude that $\ell_{\mathbf{Y}}(\alpha)$ is strictly concave in α. Thus $\hat{\alpha}_n^{\mathrm{MLE}}$ can be found by setting the first derivative of $\ell_{\mathbf{Y}}(\alpha)$ equal to 0. We get

$$\frac{n}{\hat{\alpha}_n^{\mathrm{MLE}}} + n\log\theta - \sum_{i=1}^n \log Y_i = 0 \qquad \Longleftrightarrow \qquad \hat{\alpha}_n^{\mathrm{MLE}} = \left(\frac{1}{n}\sum_{i=1}^n \log Y_i - \log\theta\right)^{-1}.$$

(b) Let $\hat{\alpha}$ denote the unbiased version of the MLE for the storm and flood data given on the exercise sheet. Since we observed 15 storm and flood events, we have $n = 15$. Thus $\hat{\alpha}$ can be calculated as

$$\hat{\alpha} = \frac{n-1}{n}\left(\frac{1}{n}\sum_{i=1}^n \log Y_i - \log\theta\right)^{-1} = \frac{14}{15}\left(\frac{1}{15}\sum_{i=1}^{15} \log Y_i - \log 50\right)^{-1} \approx 0.98,$$

where for $Y_1, \ldots, Y_{15}$ we plugged in the observed claim sizes given on the exercise sheet. Note that with $\hat{\alpha} = 0.98 < 1$, the expectation of the claim sizes does not exist.

(c) We define $N_1, \ldots, N_{20}$ to be the number of yearly storm and flood events during the twenty years $1986 - 2005$. By assumption, we have

$$N_1, \ldots, N_{20} \overset{\text{i.i.d.}}{\sim} \mathrm{Poi}(\lambda).$$

Using Estimator 2.32 of the lecture notes with $v_1 = \cdots = v_{20} = 1$, the MLE $\hat{\lambda}$ of λ is given by

$$\hat{\lambda} = \frac{1}{\sum_{i=1}^{20} 1}\sum_{i=1}^{20} N_i = \frac{1}{20}\sum_{i=1}^{20} N_i.$$

Since we observed 15 storm and flood events in total, we get

$$\hat{\lambda} = \frac{15}{20} = 0.75.$$

(d) Using Proposition 2.11 of the lecture notes, the expected yearly claim amount $\mathbb{E}[S]$ of storm and flood events is given by

$$\mathbb{E}[S] = \lambda \mathbb{E}[\min\{Y_1, M\}].$$

The expected value of $\min\{Y_1, M\}$ can be calculated as

$$\begin{aligned}\mathbb{E}[\min\{Y_1, M\}] &= \mathbb{E}[\min\{Y_1, M\} 1_{\{Y_1 \le M\}}] + \mathbb{E}[\min\{Y_1, M\} 1_{\{Y_1 > M\}}] \\ &= \mathbb{E}[Y_1 1_{\{Y_1 \le M\}}] + \mathbb{E}[M 1_{\{Y_1 > M\}}] \\ &= \mathbb{E}[Y_1 1_{\{Y_1 \le M\}}] + M\mathbb{P}[Y_1 > M],\end{aligned}$$

where for $\mathbb{E}[Y_1 1_{\{Y_1 \le M\}}]$ and $\mathbb{P}[Y_1 > M]$ we have

$$\begin{aligned}\mathbb{E}[Y_1 1_{\{Y_1 \le M\}}] &= \int_\theta^\infty x 1_{\{x \le M\}} f(x)\, dx \\ &= \int_\theta^M x \frac{\alpha}{\theta} \left(\frac{x}{\theta}\right)^{-(\alpha+1)} dx \\ &= \alpha \theta^\alpha \left[\frac{1}{1-\alpha} x^{1-\alpha}\right]_\theta^M \\ &= \frac{\alpha}{1-\alpha} \theta^\alpha M^{1-\alpha} - \frac{\alpha}{1-\alpha}\theta \\ &= \frac{\alpha}{1-\alpha} \theta \left(\frac{M}{\theta}\right)^{1-\alpha} - \frac{\alpha}{1-\alpha}\theta \\ &= \theta \frac{\alpha}{1-\alpha} \left[\left(\frac{M}{\theta}\right)^{1-\alpha} - 1\right] \\ &= \theta \frac{\alpha}{\alpha-1} \left[1 - \left(\frac{M}{\theta}\right)^{1-\alpha}\right]\end{aligned}$$

and

$$M\mathbb{P}[Y_1 > M] = M\left(1 - \mathbb{P}[Y_1 \le M]\right) = M\left(1 - \left[1 - \left(\frac{M}{\theta}\right)^{-\alpha}\right]\right) = \theta \left(\frac{M}{\theta}\right)^{1-\alpha}.$$

Hence we get

$$\mathbb{E}[\min\{Y_1, M\}] = \theta \frac{\alpha}{\alpha-1}\left[1 - \left(\frac{M}{\theta}\right)^{1-\alpha}\right] + \theta\left(\frac{M}{\theta}\right)^{1-\alpha} = \theta\frac{\alpha}{\alpha-1} - \frac{\theta}{\alpha-1}\left(\frac{M}{\theta}\right)^{1-\alpha}.$$

Replacing the unknown parameters by their estimates, we get for the estimated expected total yearly claim amount $\hat{\mathbb{E}}[S]$:

$$\hat{\mathbb{E}}[S] = \hat{\lambda}\left[\frac{\theta}{1-\hat{\alpha}}\left(\frac{M}{\theta}\right)^{1-\hat{\alpha}} - \frac{\hat{\alpha}}{1-\hat{\alpha}}\theta\right] \approx 0.75\left[\frac{50}{1-0.98}\left(\frac{2'000}{50}\right)^{1-0.98} - \frac{0.98 \cdot 50}{1-0.98}\right] \approx 180.4.$$

(e) Since $S \sim \text{CompPoi}(\lambda, G)$, we can write S as

$$S = \sum_{i=1}^N Y_i,$$

where $N \sim \text{Poi}(\lambda)$, $Y_1, Y_2, \ldots$ are i.i.d. with distribution function G and N and $Y_1, Y_2, \ldots$ are independent. Since we are only interested in events that exceed the level of $M = 2$ billions CHF, we define S_M as

$$S_M = \sum_{i=1}^N Y_i 1_{\{Y_i > M\}}.$$

Due to Theorem 2.14 of the lecture notes, we have $S_M \sim \text{CompPoi}(\lambda_M, G_M)$ for some distribution function G_M and

$$\lambda_M = \lambda \mathbb{P}[Y_1 > M] = \lambda\left(1 - \mathbb{P}[Y_1 \leq M]\right) = \lambda\left(1 - \left[1 - \left(\frac{M}{\theta}\right)^{-\alpha}\right]\right) = \lambda\left(\frac{M}{\theta}\right)^{-\alpha}.$$

Defining a random variable $N_M \sim \text{Poi}(\lambda_M)$, the probability that we observe at least one storm and flood event in a particular year is given by

$$\mathbb{P}[N_M \geq 1] = 1 - \mathbb{P}[N_M = 0] = 1 - \exp\{-\lambda_M\} = 1 - \exp\left\{-\lambda\left(\frac{M}{\theta}\right)^{-\alpha}\right\}.$$

If we replace the unknown parameters by their estimates, this probability can be estimated by

$$\hat{\mathbb{P}}[N_M \geq 1] = 1 - \exp\left\{-\hat{\lambda}\left(\frac{M}{\theta}\right)^{-\hat{\alpha}}\right\} \approx 1 - \exp\left\{-0.75\left(\frac{2\text{'}000}{50}\right)^{-0.98}\right\} \approx 0.02.$$

Note that in particular such a flood storm and flood event that exceeds the level of 2 billions CHF is expected roughly every $1/0.02 = 50$ years.

Solution 5.3 Pareto Distribution

The density g and the distribution function G of Y are given by

$$g(x) = \frac{\alpha}{\theta}\left(\frac{x}{\theta}\right)^{-(\alpha+1)} \qquad \text{and} \qquad G(x) = 1 - \left(\frac{x}{\theta}\right)^{-\alpha}$$

for all $x \geq \theta$.

(a) The survival function $\bar{G} = 1 - G$ of Y is

$$\bar{G}(x) = 1 - G(x) = \left(\frac{x}{\theta}\right)^{-\alpha}$$

for all $x \geq \theta$. Hence, for all $t > 0$ we have

$$\lim_{x\to\infty} \frac{\bar{G}(xt)}{\bar{G}(x)} = \lim_{x\to\infty} \frac{(xt/\theta)^{-\alpha}}{(x/\theta)^{-\alpha}} = t^{-\alpha}.$$

Thus, by definition, the survival function of Y is regularly varying at infinity with tail index α.

(b) Let $\theta \leq u_1 < u_2$ and $\alpha \neq 1$. Then the expected value of Y within the layer $(u_1, u_2]$ can be calculated as

$$\mathbb{E}[Y 1_{\{u_1 < Y \leq u_2\}}] = \int_\theta^\infty x 1_{\{u_1 < x \leq u_2\}} g(x)\, dx = \int_{u_1}^{u_2} x \frac{\alpha}{\theta}\left(\frac{x}{\theta}\right)^{-(\alpha+1)} dx = \alpha\theta \int_{u_1}^{u_2} \frac{1}{\theta}\left(\frac{x}{\theta}\right)^{-\alpha} dx.$$

In the case $\alpha \neq 1$, we get

$$\mathbb{E}[Y 1_{\{u_1 < Y \leq u_2\}}] = \alpha\theta\left[-\frac{1}{\alpha-1}\left(\frac{x}{\theta}\right)^{-\alpha+1}\right]_{u_1}^{u_2} = \theta\frac{\alpha}{\alpha-1}\left[\left(\frac{u_1}{\theta}\right)^{-\alpha+1} - \left(\frac{u_2}{\theta}\right)^{-\alpha+1}\right],$$

and if $\alpha = 1$, we get

$$\mathbb{E}[Y 1_{\{u_1 < Y \leq u_2\}}] = \theta \int_{u_1}^{u_2} \frac{1}{x}\, dx = \theta \log\left(\frac{u_2}{u_1}\right).$$

(c) Let $\alpha > 1$ and $y > \theta$. Then the expected value μ_Y of Y is given by

$$\mu_Y = \theta \frac{\alpha}{\alpha - 1}$$

and, similarly as in part (b), we get

$$\mathbb{E}[Y1_{\{Y \leq y\}}] = \mathbb{E}[Y1_{\{\theta < Y \leq y\}}] = \theta \frac{\alpha}{\alpha - 1} \left[\left(\frac{\theta}{\theta}\right)^{-\alpha+1} - \left(\frac{y}{\theta}\right)^{-\alpha+1} \right] = \mu_Y \left[1 - \left(\frac{y}{\theta}\right)^{-\alpha+1} \right].$$

Hence, for the loss size index function for level $y > \theta$ we have

$$\mathcal{I}[G(y)] = \frac{1}{\mu_Y} \mathbb{E}[Y1_{\{Y \leq y\}}] = 1 - \left(\frac{y}{\theta}\right)^{-\alpha+1} \in [0, 1].$$

(d) Let $\alpha > 1$ and $u > \theta$. The mean excess function of Y above u can be calculated as

$$e(u) = \mathbb{E}[Y - u | Y > u] = \mathbb{E}[Y | Y > u] - u = \frac{\mathbb{E}[Y1_{\{Y > u\}}]}{\mathbb{P}[Y > u]} - u = \frac{\mathbb{E}[Y1_{\{Y > u\}}]}{\bar{G}(u)} - u,$$

where for $\mathbb{E}[Y1_{\{Y > u\}}]$ we have, similarly as in part (b),

$$\mathbb{E}[Y1_{\{Y > u\}}] = \alpha\theta \left[-\frac{1}{\alpha - 1} \left(\frac{x}{\theta}\right)^{-\alpha+1} \right]_u^\infty = \frac{\alpha}{\alpha - 1} \theta \left(\frac{u}{\theta}\right)^{-\alpha+1} = \frac{\alpha}{\alpha - 1} u \bar{G}(u).$$

Thus we get

$$e(u) = \frac{\alpha}{\alpha - 1} u - u = \frac{1}{\alpha - 1} u.$$

Note that the mean excess function $u \mapsto e(u)$ has slope $\frac{1}{\alpha - 1} > 0$.

Non-Life Insurance: Mathematics and Statistics

Solution sheet 6

Solution 6.1 Goodness-of-Fit Test

Let Y be a random variable following a Pareto distribution with threshold $\theta = 200$ and tail index $\alpha = 1.25$. Then the distribution function G of Y is given by

$$G(x) = 1 - \left(\frac{x}{\theta}\right)^{-\alpha} = 1 - \left(\frac{x}{200}\right)^{-1.25}$$

for all $x \geq \theta$. For example for the interval I_2 we then have

$$\mathbb{P}[Y \in I_2] = \mathbb{P}[239 \leq Y < 301] = G(301) - G(239) = 1 - \left(\frac{301}{200}\right)^{-1.25} - \left[1 - \left(\frac{239}{200}\right)^{-1.25}\right] \approx 0.2.$$

By analogous calculations for the other four intervals, we get

$$\mathbb{P}[Y \in I_1] \approx 0.2, \quad \mathbb{P}[Y \in I_2] \approx 0.2, \quad \mathbb{P}[Y \in I_3] \approx 0.2, \quad \mathbb{P}[Y \in I_4] \approx 0.2, \quad \mathbb{P}[Y \in I_5] \approx 0.2.$$

Let E_i and O_i denote respectively the expected number of observations in I_i and the observed number of observations in I_i, for all $i \in \{1, \ldots, 5\}$. As we have 20 observations in our data, we can calculate for example E_2 as

$$E_2 = 20 \cdot \mathbb{P}[Y \in I_2] \approx 4.$$

The values of the expected number of observations and the observed number of observations in the five intervals as well as their squared differences are summarized in the following table:

i	1	2	3	4	5
O_i	4	0	8	6	2
E_i	4	4	4	4	4
$(O_i - E_i)^2$	0	16	16	4	4

Now the test statistic of the χ^2-goodness-of-fit test using 5 intervals and 20 observations is given by

$$X^2_{20,5} = \sum_{i=1}^{5} \frac{(O_i - E_i)^2}{E_i} = \frac{0}{4} + \frac{16}{4} + \frac{16}{4} + \frac{4}{4} + \frac{4}{4} = 10.$$

Let $\alpha = 5\%$. Then the $(1-\alpha)$-quantile of the χ^2-distribution with $5 - 1 = 4$ degrees of freedom is given by approximately 9.49. Since this is smaller than $X^2_{20,5}$, we can reject the null hypothesis of having a Pareto distribution with threshold $\theta = 200$ and tail index $\alpha = 1.25$ as claim size distribution at the significance level of 5%.

Solution 6.2 Log-Normal Distribution and Deductible

(a) Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Then the moment generating function M_X of X is given by

$$M_X(r) = \mathbb{E}\left[\exp\{rX\}\right] = \exp\left\{r\mu + \frac{r^2\sigma^2}{2}\right\}$$

for all $r \in \mathbb{R}$. Since Y_1 has a log-normal distribution with mean parameter μ and variance parameter σ^2, we have

$$Y_1 \stackrel{\mathrm{d}}{=} \exp\{X\}.$$

Hence, the expectation, the variance and the coefficient of variation of Y_1 can be calculated as

$$\begin{aligned}
\mathbb{E}[Y_1] &= \mathbb{E}\left[\exp\{X\}\right] = \mathbb{E}\left[\exp\{1 \cdot X\}\right] = M_X(1) = \exp\left\{\mu + \frac{\sigma^2}{2}\right\}, \\
\operatorname{Var}(Y_1) &= \mathbb{E}[Y_1^2] - \mathbb{E}[Y_1]^2 = \mathbb{E}\left[\exp\{2X\}\right] - M_X(1)^2 = M_X(2) - M_X(1)^2 \\
&= \exp\left\{2\mu + \frac{4\sigma^2}{2}\right\} - \exp\left\{2\mu + 2\frac{\sigma^2}{2}\right\} = \exp\left\{2\mu + \sigma^2\right\}\left(\exp\left\{\sigma^2\right\} - 1\right) \quad \text{and} \\
\operatorname{Vco}(Y_1) &= \frac{\sqrt{\operatorname{Var}(Y_1)}}{\mathbb{E}[Y_1]} = \frac{\exp\left\{\mu + \sigma^2/2\right\}\sqrt{\exp\left\{\sigma^2\right\} - 1}}{\exp\left\{\mu + \sigma^2/2\right\}} = \sqrt{\exp\left\{\sigma^2\right\} - 1}.
\end{aligned}$$

(b) From part (a), we know that

$$\begin{aligned}
\sigma &= \sqrt{\log[\operatorname{Vco}(Y_1)^2 + 1]} \quad \text{and} \\
\mu &= \log \mathbb{E}[Y_1] - \frac{\sigma^2}{2}.
\end{aligned}$$

Since $\mathbb{E}[Y_1] = 3\text{'}000$ and $\operatorname{Vco}(Y_1) = 4$, we get

$$\begin{aligned}
\sigma &= \sqrt{\log(4^2 + 1)} \approx 1.68 \quad \text{and} \\
\mu &\approx \log 3\text{'}000 - \frac{(1.68)^2}{2} \approx 6.59.
\end{aligned}$$

(i) The claims frequency λ is given by $\lambda = \mathbb{E}[N]/v$. With the introduction of the deductible $d = 500$, the number of claims changes to

$$N^{\text{new}} = \sum_{i=1}^{N} 1_{\{Y_i > d\}}.$$

Using the independence of N and $Y_1, Y_2, \ldots$, we get

$$\mathbb{E}[N^{\text{new}}] = \mathbb{E}\left[\sum_{i=1}^{N} 1_{\{Y_i > d\}}\right] = \mathbb{E}[N]\mathbb{E}[1_{\{Y_1 > d\}}] = \mathbb{E}[N]\mathbb{P}[Y_1 > d].$$

Let Φ denote the distribution function of a standard Gaussian distribution. Since $\log Y_1$ has a Gaussian distribution with mean μ and variance σ^2, we have

$$\mathbb{P}[Y_1 > d] = 1 - \mathbb{P}\left[\frac{\log Y_1 - \mu}{\sigma} \leq \frac{\log d - \mu}{\sigma}\right] = 1 - \Phi\left(\frac{\log d - \mu}{\sigma}\right).$$

Hence, the new claims frequency λ^{new} is given by

$$\lambda^{\text{new}} = \mathbb{E}[N_{\text{new}}]/v = \mathbb{E}[N]\mathbb{P}[Y_1 > d]/v = \lambda\mathbb{P}[Y_1 > d] = \lambda\left[1 - \Phi\left(\frac{\log d - \mu}{\sigma}\right)\right].$$

Inserting the values of d, μ and σ, we get

$$\lambda^{\text{new}} \approx \lambda\left[1 - \Phi\left(\frac{\log 500 - 6.59}{1.68}\right)\right] \approx 0.59 \cdot \lambda.$$

Note that the introduction of this deductible reduces the administrative burden a lot, because 41% of (small) claims disappear.

(ii) With the introduction of the deductible $d = 500$, the claim sizes change to

$$Y_i^{\text{new}} = Y_i - d \,|\, Y_i > d.$$

Thus, the new expected claim size is given by

$$\mathbb{E}[Y_1^{\text{new}}] = \mathbb{E}[Y_1 - d | Y_1 > d] = e(d),$$

where $e(d)$ is the mean excess function of Y_1 above d. According to the lecture notes, $e(d)$ is given by

$$e(d) = \mathbb{E}[Y_1] \left[\frac{1 - \Phi\left(\frac{\log d - \mu - \sigma^2}{\sigma}\right)}{1 - \Phi\left(\frac{\log d - \mu}{\sigma}\right)} \right] - d.$$

Inserting the values of d, μ, σ and $\mathbb{E}[Y_1]$, we get

$$\mathbb{E}[Y_1^{\text{new}}] \approx 3'000 \left[\frac{1 - \Phi\left(\frac{\log 500 - 6.59 - 1.68^2}{1.68}\right)}{1 - \Phi\left(\frac{\log 500 - 6.59}{1.68}\right)} \right] - 500 \approx 4'456 \approx 1.49 \cdot \mathbb{E}[Y_1].$$

(iii) According to Proposition 2.2 of the lecture notes, the expected total claim amount $\mathbb{E}[S]$ is given by

$$\mathbb{E}[S] = \mathbb{E}[N]\mathbb{E}[Y_1].$$

With the introduction of the deductible $d = 500$, the total claim amount S changes to S^{new}, which can be written as

$$S^{\text{new}} = \sum_{i=1}^{N^{\text{new}}} Y_i^{\text{new}}.$$

Hence, the expected total claim amount changes to

$$\begin{aligned} \mathbb{E}\left[S^{\text{new}}\right] &= \mathbb{E}\left[N^{\text{new}}\right] \mathbb{E}\left[Y_1^{\text{new}}\right] \\ &= \mathbb{E}[N]\mathbb{P}[Y_1 > d] e(d) \\ &= \lambda v \left[1 - \Phi\left(\frac{\log d - \mu}{\sigma} \right) \right] \cdot \left(\mathbb{E}[Y_1] \left[\frac{1 - \Phi\left(\frac{\log d - \mu - \sigma^2}{\sigma}\right)}{1 - \Phi\left(\frac{\log d - \mu}{\sigma}\right)} \right] - d \right). \end{aligned}$$

Inserting the values of d, μ, σ and $\mathbb{E}[Y_1]$, we get

$$\begin{aligned} \mathbb{E}\left[S^{\text{new}}\right] &\approx \lambda v \left[1 - \Phi\left(\frac{\log 500 - 6.59}{1.68} \right) \right] \cdot \left(3'000 \left[\frac{1 - \Phi\left(\frac{\log 500 - 6.59 - 1.68^2}{1.68}\right)}{1 - \Phi\left(\frac{\log 500 - 6.59}{1.68}\right)} \right] - 500 \right) \\ &\approx \lambda v \cdot 0.59 \cdot 4'456 \\ &= 0.88 \cdot \mathbb{E}[S]. \end{aligned}$$

In particular, the insurance company can grant a discount of roughly 12% on the pure risk premium. Note that also the administrative expenses on claims handling will reduce substantially because we only have 59% of the original claims, see the result in (i).

Solution 6.3 Inflation and Deductible

Let Y be a random variable following a Pareto distribution with threshold $\theta > 0$ and tail index $\alpha > 1$. Then the expectation $\mathbb{E}[Y]$ of Y and the mean excess function $e_Y(u)$ of Y above $u > \theta$ are given by

$$\mathbb{E}[Y] = \frac{\alpha}{\alpha-1}\theta \quad \text{and} \quad e_Y(u) = \frac{1}{\alpha-1}u.$$

Since the insurance company only has to pay the part that exceeds the threshold θ, this year's average claim payment z is

$$z = \mathbb{E}[Y] - \theta = \frac{\alpha}{\alpha-1}\theta - \theta = \frac{\theta}{\alpha-1}.$$

For the total claim size $\tilde{Y}$ of a claim next year we have

$$\tilde{Y} \stackrel{d}{=} (1+r)Y \sim \text{Pareto}([1+r]\theta, \alpha).$$

Let $\rho\theta$ for some $\rho > 0$ denote the increase of the threshold that is needed such that the average claims payment remains unchanged. Then next year's average claim payment is given by

$$\tilde{z} = \mathbb{E}[(\tilde{Y} - [1+\rho]\theta)_+].$$

Let's first assume that we can choose a $\rho < r$ such that $z = \tilde{z}$. In this case we get

$$\tilde{Y} \geq (1+r)\theta \quad \text{a.s.} \qquad \Longrightarrow \qquad \tilde{Y} \geq (1+\rho)\theta \quad \text{a.s.}$$

and thus

$$\tilde{z} = \mathbb{E}[\tilde{Y} - (1+\rho)\theta] = \mathbb{E}[\tilde{Y}] - (1+\rho)\theta = \frac{\alpha}{\alpha-1}(1+r)\theta - (1+\rho)\theta.$$

Now we have $z = \tilde{z}$ if and only if

$$\begin{aligned}
\frac{\alpha}{\alpha-1}\theta - \theta &= \frac{\alpha}{\alpha-1}(1+r)\theta - (1+\rho)\theta \\
\Longleftrightarrow \qquad 0 &= \frac{\alpha}{\alpha-1}r\theta - \rho\theta \\
\Longleftrightarrow \qquad \rho &= \frac{\alpha}{\alpha-1}r > r,
\end{aligned}$$

which is a contradiction to the assumption $\rho < r$. Hence, we conclude that $\rho \geq r$, i.e. the percentage increase in the deductible has to be bigger than the inflation. Assuming $\rho \geq r$, we can calculate

$$\begin{aligned}
\tilde{z} &= \mathbb{E}[(\tilde{Y} - [1+\rho]\theta) \cdot 1_{\{\tilde{Y} - (1+\rho)\theta\}}] \\
&= \mathbb{E}[\tilde{Y} - (1+\rho)\theta \,|\, \tilde{Y} > (1+\rho)\theta] \cdot \mathbb{P}[\tilde{Y} > (1+\rho)\theta] \\
&= e_{\tilde{Y}}([1+\rho]\theta) \cdot \mathbb{P}[\tilde{Y} > (1+\rho)\theta] \\
&= \frac{1}{\alpha-1}(1+\rho)\theta \cdot \left[\frac{(1+\rho)\theta}{(1+r)\theta}\right]^{-\alpha} \\
&= \frac{\theta}{\alpha-1}(1+r)^{\alpha}(1+\rho)^{-\alpha+1} \\
&= z \cdot (1+r)^{\alpha}(1+\rho)^{-\alpha+1}.
\end{aligned}$$

Now we have $z = \tilde{z}$ if and only if

$$(1+r)^{\alpha}(1+\rho)^{-\alpha+1} = 1 \quad \Longleftrightarrow \quad \rho = (1+r)^{\frac{\alpha}{\alpha-1}} - 1.$$

We conclude that if we want the claim payment to remain unchanged, we have to increase the deductible θ by the amount

$$\theta\left[(1+r)^{\frac{\alpha}{\alpha-1}} - 1\right].$$

Non-Life Insurance: Mathematics and Statistics

Solution sheet 7

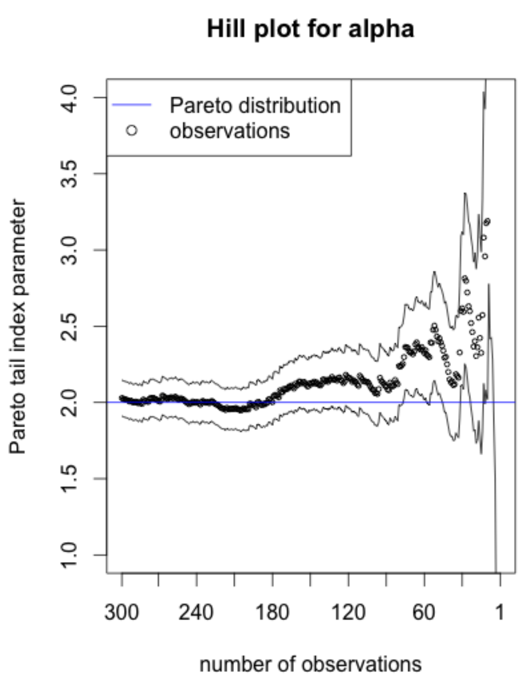
Solution 7.1 Hill Estimator

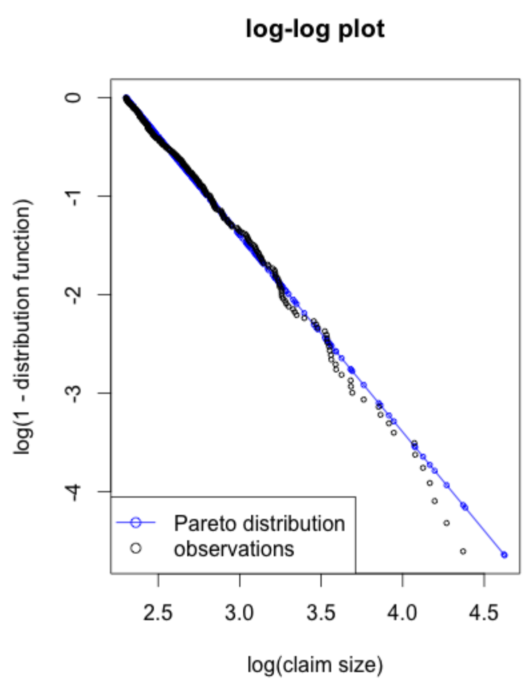
An example of a possible R-code is given below.

```
### Generate 300 independent observations coming from a
### Pareto distribution with threshold theta = 10 millions
### and tail index alpha = 2.
### We use that if Z ~ Gamma(1,alpha),
### then theta*exp{Z} ~ Pareto(theta, alpha).
### Note that for the Gamma distribution we have:
### scale parameter in R = 1/(scale parameter in lecture notes)
n <- 300
theta <- 10   ### in millions
alpha <- 2
set.seed(100)   ### for reproducibility
data.1 <- rgamma(n, shape = 1, scale = 1 / alpha)
data <- theta * exp(data.1)

### Order the data
data.ordered <- data[order(data, decreasing = FALSE)]

### Take the logarithm
log.data.ordered <- log(data.ordered)

### Number of observations
n.obs <- n:1

### Hill estimator
hill.estimator <- ((sum(log.data.ordered)
    - cumsum(log.data.ordered) + log.data.ordered) / n.obs
    - log.data.ordered)^(-1)

### Confidence bounds (see Lemma 3.7 of the lecture notes)
upper.bound <- hill.estimator + sqrt(n.obs^2 / ((n.obs - 1)^2
    * (n.obs - 2)) * hill.estimator^2)
lower.bound <- hill.estimator - sqrt(n.obs^2 / ((n.obs - 1)^2
    * (n.obs - 2)) * hill.estimator^2)

### Hill plot and log-log plot next to each other
par(mfrow=c(1,2))

### Hill plot
plot(hill.estimator, ylim = c(alpha-1,alpha+2), xaxt="n",
    xlab = "number of observations",
    ylab = "Pareto tail index parameter", cex = 0.5)
```

```
title(main = "Hill plot for alpha")
axis(1,at=c(1,seq(from = n / 10, to = n, by = n / 10)),
     c(seq(from = n, to = n / 10, by = -n / 10),1))
lines(upper.bound)
lines(lower.bound)
abline(h = alpha, col = "blue")
legend("topleft", col = c("blue","black"), lty = c(1,NA),
     pch = c(NA,1),
     legend = c("Pareto distribution","observations"))

### True survival function (= 1 - true distribution function)
true.sf <- (data.ordered / theta)^(-alpha)

### Empirical survival function (= 1 - empirical survival function)
empirical.sf <- 1 - (1:n) / n

### Log-log plot
plot(log.data.ordered,log(true.sf), xlab = "log(claim size)",
     ylab = "log(1 - distribution function)",
     cex= 0.5, col = "blue")
title(main = "log-log plot")
lines(log.data.ordered, log(true.sf), col = "blue")
points(log.data.ordered, log(empirical.sf), col = "black",
     cex= 0.5)
legend("bottomleft", col = c("blue","black"), lty = c(1,NA),
     pch = c(1,1), legend = c("Pareto distribution","observations"))
```

The Hill plot (on the left) and the log-log plot (on the right) look as follows:





Note that even though we sampled from a Pareto distribution with tail index $\alpha = 2$, it is not at all clear to see that the data comes from a Pareto distribution. In the Hill plot we see that, first, the estimates of α seem more or less correct, but starting from the 180 largest observations, the plot suggests a higher α or even another distribution. In the log-log plot we see that for small-sized and medium-sized claims the fit seems to be fine. But looking at the largest claims, we would conclude that our data is not as heavy-tailed as a true Pareto distribution with threshold $\theta = 10$ millions and tail index $\alpha = 2$ would suggest. We are confronted with these problems even though we sampled directly from a Pareto distribution. This might indicate the difficulties one faces when trying to fit such a distribution to a real data set, which, to make matters even worse, often contains far less than 300 observations as in this example and moreover the observations may be contaminated by other distributions.

Solution 7.2 Approximations

Note that if $Y \sim \Gamma(\gamma = 100, c = \frac{1}{10})$, then

$$\begin{aligned}
\mathbb{E}[Y] &= \frac{\gamma}{c} = \frac{100}{1/10} = 1\text{'}000, \\
\mathbb{E}[Y^2] &= \frac{\gamma(\gamma+1)}{c^2} = \frac{100 \cdot 101}{1/100} = 1\text{'}010\text{'}000 \quad \text{and} \\
\mathbb{E}[Y^3] &= \frac{\gamma(\gamma+1)(\gamma+2)}{c^3} = \frac{100 \cdot 101 \cdot 102}{1/1000} = 1\text{'}030\text{'}200\text{'}000.
\end{aligned}$$

Let M_Y denote the moment generating function of Y. According to formula (1.3) of the lecture notes, we have

$$M_Y'''(0) = \frac{d^3}{dr^3} M_Y(r)\bigg|_{r=0} = \mathbb{E}[Y^3].$$

For the total claim amount S, we can use Proposition 2.11 of the lecture notes to get

$$\begin{aligned}
\mathbb{E}[S] &= \lambda v \mathbb{E}[Y] = 1\text{'}000 \cdot 1\text{'}000 = 1\text{'}000\text{'}000, \\
\operatorname{Var}(S) &= \lambda v \mathbb{E}[Y^2] = 1\text{'}000 \cdot 1\text{'}010\text{'}000 = 1\text{'}010\text{'}000\text{'}000 \quad \text{and} \\
M_S(r) &= \exp\{\lambda v [M_Y(r) - 1]\}.
\end{aligned}$$

In order to get the skewness ς_S of S, which we will need for the translated gamma and the log-normal approximations, we can use the third equation given in the formulas (1.5) of the lecture notes:

$$\varsigma_S \cdot \operatorname{Var}(S)^{3/2} = \frac{d^3}{dr^3} \log M_S(r)\bigg|_{r=0} = \lambda v \frac{d^3}{dr^3} M_Y(r)\bigg|_{r=0} = \lambda v M_Y'''(0) = \lambda v \mathbb{E}[Y^3],$$

from which we can conclude that

$$\varsigma_S = \frac{\lambda v \mathbb{E}[Y^3]}{(\lambda v \mathbb{E}[Y^2])^{3/2}} = \frac{\mathbb{E}[Y^3]}{\sqrt{\lambda v}\,\mathbb{E}[Y^2]^{3/2}} = \frac{1\text{'}030\text{'}200\text{'}000}{\sqrt{1\text{'}000}(1\text{'}010\text{'}000)^{3/2}} \approx 0.0321.$$

Let F_S denote the distribution function of S. Then, since F_S is continuous and strictly increasing, the quantiles $q_{0.95}$ and $q_{0.99}$ can be calculated as

$$q_{0.95} = F_S^{-1}(0.95) \qquad \text{and} \qquad q_{0.99} = F_S^{-1}(0.99).$$

(a) According to Section 4.1.1 of the lecture notes, the normal approximation is given by

$$F_S(x) \approx \Phi\left(\frac{x - \lambda v \mathbb{E}[Y]}{\sqrt{\lambda v \mathbb{E}[Y^2]}}\right),$$

for all $x \in \mathbb{R}$, where Φ is the standard Gaussian distribution function. For all $\alpha \in (0,1)$, we have

$$\begin{aligned} F_S^{-1}(\alpha) &= \lambda v \mathbb{E}[Y] + \sqrt{\lambda v \mathbb{E}[Y^2]} \cdot \Phi^{-1}(\alpha) \\ &= 1'000 \cdot 1'000 + \sqrt{1'000 \cdot 1'010'000} \cdot \Phi^{-1}(\alpha) \\ &\approx 1'000'000 + 31'780.5 \cdot \Phi^{-1}(\alpha). \end{aligned}$$

In particular, we get

$$q_{0.95} = F_S^{-1}(0.95) \approx 1'000'000 + 31'780.5 \cdot \Phi^{-1}(0.95) \approx 1'000'000 + 31'780.5 \cdot 1.645 = 1'052'279$$

and

$$q_{0.99} = F_S^{-1}(0.99) \approx 1'000'000 + 31'780.5 \cdot \Phi^{-1}(0.99) \approx 1'000'000 + 31'780.5 \cdot 2.325 = 1'073'890.$$

Note that the normal approximation also allows for negative claims S, which under our model assumption is excluded. The probability for negative claims S in the normal approximation can be calculated as

$$F_S(0) \approx \Phi\left(\frac{0 - \lambda v \mathbb{E}[Y]}{\sqrt{\lambda v \mathbb{E}[Y^2]}}\right) \approx \Phi\left(-\frac{1'000'000}{31'780.5}\right) \approx \Phi(-31.5) \approx 4.34 \cdot 10^{-218},$$

which of course is positive, but very close to 0.

(b) According to Section 4.1.2 of the lecture notes, in the translated gamma approximation we model S by the random variable

$$X = k + Z,$$

where $k \in \mathbb{R}$ and $Z \sim \Gamma(\tilde{\gamma}, \tilde{c})$. The three parameters $k, \tilde{\gamma}$ and $\tilde{c}$ can be determined by solving the equations

$$\mathbb{E}[X] = \mathbb{E}[S], \qquad \mathrm{Var}(X) = \mathrm{Var}(S) \qquad \text{and} \qquad \varsigma_X = \varsigma_S, \tag{1}$$

where ς_X is the skewness parameter of X. Since $Z \sim \Gamma(\tilde{\gamma}, \tilde{c})$, we can use the results given in Section 3.2.1 of the lecture notes to calculate

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[k+Z] = k + \mathbb{E}[Z] = k + \frac{\tilde{\gamma}}{\tilde{c}}, \\ \mathrm{Var}(X) &= \mathrm{Var}(k+Z) = \mathrm{Var}(Z) = \frac{\tilde{\gamma}}{\tilde{c}^2} \quad \text{and} \\ \varsigma_X &= \frac{\mathbb{E}\left[(X - \mathbb{E}[X])^3\right]}{\mathrm{Var}(X)^{3/2}} = \frac{\mathbb{E}\left[(k + Z - \mathbb{E}[k+Z])^3\right]}{\mathrm{Var}(k+Z)^{3/2}} = \frac{\mathbb{E}\left[(Z - \mathbb{E}[Z])^3\right]}{\mathrm{Var}(Z)^{3/2}} = \varsigma_Z = \frac{2}{\sqrt{\tilde{\gamma}}}. \end{aligned}$$

Using equations (1), we get

$$\begin{aligned} \frac{2}{\sqrt{\tilde{\gamma}}} = \varsigma_S \quad &\Longleftrightarrow \quad \tilde{\gamma} = \frac{4}{\varsigma_S^2} \approx 3'883, \\ \frac{\tilde{\gamma}}{\tilde{c}^2} = \mathrm{Var}(S) \quad &\Longleftrightarrow \quad \tilde{c} = \sqrt{\frac{\tilde{\gamma}}{\mathrm{Var}(S)}} \approx 0.002 \quad \text{and} \\ k + \frac{\tilde{\gamma}}{\tilde{c}} = \mathbb{E}[S] \quad &\Longleftrightarrow \quad k = \mathbb{E}[S] - \frac{\tilde{\gamma}}{\tilde{c}} = \mathbb{E}[S] - \sqrt{\tilde{\gamma}\mathrm{Var}(S)} \approx -980'392. \end{aligned}$$

If we write F_Z for the distribution function of $Z \sim \Gamma(\tilde{\gamma} \approx 3'883, \tilde{c} \approx 0.002)$, using the translated gamma approximation, we get

$$F_S(x) = \mathbb{P}[S \leq x] \approx \mathbb{P}[X \leq x] = \mathbb{P}[k + Z \leq x] = \mathbb{P}[Z \leq x - k] = F_Z(x - k),$$

for all $x \in \mathbb{R}$. Now, for all $\alpha \in (0,1)$, we have

$$F_S^{-1}(\alpha) \approx k + F_Z^{-1}(\alpha)$$

In particular, we get

$$q_{0.95} = F_S^{-1}(0.95) \approx k + F_Z^{-1}(0.95) \approx -980'392 + 2'032'955 = 1'052'563$$

and

$$q_{0.99} = F_S^{-1}(0.99) \approx k + F_Z^{-1}(0.99) \approx -980'392 + 2'055'074 = 1'074'682.$$

Note that since $k < 0$, the translated gamma approximation in this example also allows for negative claims S, which under our model assumption is excluded. The probability for negative claims S can be calculated as

$$F_S(0) \approx F_Z(0-k) \approx F_Z(980'392) \approx 4.87 \cdot 10^{-320},$$

which is basically 0.

(c) According to Section 4.1.2 of the lecture notes, in the translated log-normal approximation we model S by the random variable

$$X = k + Z,$$

where $k \in \mathbb{R}$ and $Z \sim \mathrm{LN}(\mu, \sigma^2)$. Similarly as in part (b), the three parameters k, μ and σ^2 can be determined by solving the equations

$$\mathbb{E}[X] = \mathbb{E}[S], \qquad \mathrm{Var}(X) = \mathrm{Var}(S) \qquad \text{and} \qquad \varsigma_X = \varsigma_S. \tag{2}$$

Since $Z \sim \mathrm{LN}(\mu, \sigma^2)$, we can use the results given in Section 3.2.3 of the lecture notes to calculate

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[k+Z] = k + \mathbb{E}[Z] = k + \exp\left\{\mu + \sigma^2/2\right\}, \\
\mathrm{Var}(X) &= \mathrm{Var}(k+Z) = \mathrm{Var}(Z) = \exp\left\{2\mu + \sigma^2\right\}\left(\exp\left\{\sigma^2\right\} - 1\right) \quad \text{and} \\
\varsigma_X &= \varsigma_Z = \left(\exp\left\{\sigma^2\right\} + 2\right)\left(\exp\left\{\sigma^2\right\} - 1\right)^{1/2}.
\end{aligned}$$

Using the third equation in (2), we get

$$\left(\exp\left\{\sigma^2\right\} + 2\right)\left(\exp\left\{\sigma^2\right\} - 1\right)^{1/2} = \varsigma_S \approx 0.0321 \qquad \Longleftrightarrow \qquad \sigma^2 \approx 0.00012,$$

which was found using a computer software. Using the second equation in (2), we get

$$\exp\left\{2\mu + \sigma^2\right\}\left(\exp\left\{\sigma^2\right\} - 1\right) = \mathrm{Var}(S) \iff \mu = \frac{1}{2}\left(\log\left[\left(\exp\left\{\sigma^2\right\} - 1\right)^{-1}\mathrm{Var}(S)\right] - \sigma^2\right),$$

which implies

$$\mu \approx 14.875.$$

Finally, using the first equation in (2), we get

$$k + \exp\left\{\mu + \sigma^2/2\right\} = \mathbb{E}[S] \qquad \Longleftrightarrow \qquad k = \mathbb{E}[S] - \exp\left\{\mu + \sigma^2/2\right\} \approx -2'391'769.$$

If we write F_W for the distribution function of $W = \log Z \sim \mathcal{N}(\mu \approx 14.875, \sigma^2 \approx 0.00012)$, using the translated log-normal approximation, we get

$$F_S(x) = \mathbb{P}[S \le x] \approx \mathbb{P}[X \le x] = \mathbb{P}[k + Z \le x] = \mathbb{P}[\log Z \le \log(x-k)] = F_W(\log[x-k]),$$

for all $x \in \mathbb{R}$. Now, for all $\alpha \in (0,1)$, we have

$$F_S^{-1}(\alpha) \approx k + \exp\{F_W^{-1}(\alpha)\}.$$

In particular, we get

$$q_{0.95} = F_S^{-1}(0.95) \approx k + \exp\{F_W^{-1}(0.95)\} \approx -2'391'769 + 3'444'295 = 1'052'527$$

and

$$q_{0.99} = F_S^{-1}(0.99) \approx k + \exp\{F_W^{-1}(0.99)\} \approx -2'391'769 + 3'466'359 = 1'074'590.$$

Note that since $k < 0$, the translated log-normal approximation in this example also allows for negative claims S, which under our model assumption is excluded. The probability for negative claims S can be calculated as

$$F_S(0) \approx F_Z(0-k) = F_W(\log[-k]) \approx F_W(\log 2'391'769) \approx 1.92 \cdot 10^{-304},$$

which is basically 0.

(d) We observe that with all the three approximations applied in parts (a) - (c) we get almost the same results. In particular, the normal approximation does not provide estimates that deviate significantly from the ones we get using the translated gamma and the translated log-normal approximations. This is due to the fact, that $\lambda v = 1'000$ is large enough and the gamma distribution assumed for the claim sizes is not a heavy tailed distribution. Moreover, the skewness $\varsigma_S = 0.0321$ of S is rather small, hence the normal approximation is a valid model in this example. Note that in all the three approximations we allow for negative claims S, which actually should not be possible under our model assumption. However, the probability of observe a negative claim S is vanishingly small.

Solution 7.3 Akaike Information Criterion and Bayesian Information Criterion

(a) By definition, the MLEs $(\hat{\gamma}^{\text{MLE}}, \hat{c}^{\text{MLE}})$ maximize the log-likelihood function $\ell_{\mathbf{Y}}$. In particular, we have

$$\ell_{\mathbf{Y}}\left(\hat{\gamma}^{\text{MLE}}, \hat{c}^{\text{MLE}}\right) \geq \ell_{\mathbf{Y}}(\gamma, c),$$

for all $(\gamma, c) \in \mathbb{R}_+ \times \mathbb{R}_+$.

If we write $d^{(\text{MLE})}$ and $d^{(\text{MM})}$ for the number of estimated parameters in the MLE model and in the method of moments model, respectively, we have $d^{(\text{MLE})} = d^{(\text{MM})} = 2$. The AIC value $\text{AIC}^{(\text{MLE})}$ of the MLE model and the AIC value $\text{AIC}^{(\text{MM})}$ of the method of moments model are then given by

$$\text{AIC}^{(\text{MLE})} = -2\ell_{\mathbf{Y}}\left(\hat{\gamma}^{\text{MLE}}, \hat{c}^{\text{MLE}}\right) + 2d^{(\text{MLE})} = -2 \cdot 1264.013 + 2 \cdot 2 = -2524.026 \quad \text{and}$$
$$\text{AIC}^{(\text{MM})} = -2\ell_{\mathbf{Y}}\left(\hat{\gamma}^{\text{MM}}, \hat{c}^{\text{MM}}\right) + 2d^{(\text{MM})} = -2 \cdot 1264.171 + 2 \cdot 2 = -2524.342.$$

According to the AIC, the model with the smallest AIC value should be preferred. Since $\text{AIC}^{(\text{MLE})} < \text{AIC}^{(\text{MM})}$, we choose the MLE fit.

(b) If we write $d^{(\text{gam})}$ and $d^{(\text{exp})}$ for the number of estimated parameters in the gamma model and in the exponential model, respectively, we have $d^{(\text{gam})} = 2$ and $d^{(\text{exp})} = 1$. The AIC value $\text{AIC}^{(\text{gam})}$ of the gamma model and the AIC value $\text{AIC}^{(\text{exp})}$ of the exponential model are then given by

$$\text{AIC}^{(\text{gam})} = -2\ell_{\mathbf{Y}}^{(\text{gam})}\left(\hat{\gamma}^{\text{MLE}}, \hat{c}^{\text{MLE}}\right) + 2d^{(\text{gam})} = -2 \cdot 1264.013 + 2 \cdot 2 = -2524.026 \quad \text{and}$$
$$\text{AIC}^{(\text{exp})} = -2\ell_{\mathbf{Y}}^{(\text{exp})}\left(\hat{c}^{\text{MLE}}\right) + 2d^{(\text{exp})} = -2 \cdot 1264.169 + 2 \cdot 1 = -2526.338.$$

Since $\mathrm{AIC}^{(\mathrm{gam})} > \mathrm{AIC}^{(\mathrm{exp})}$, we choose the exponential model.

The BIC value $\mathrm{BIC}^{(\mathrm{gam})}$ of the gamma model and the BIC value $\mathrm{BIC}^{(\mathrm{exp})}$ of the exponential model are given by

$$\begin{aligned}\mathrm{BIC}^{(\mathrm{gam})} &= -2\ell_{\mathbf{Y}}^{(\mathrm{gam})}\left(\hat{\gamma}^{\mathrm{MLE}}, \hat{c}^{\mathrm{MLE}}\right) + d^{(\mathrm{gam})} \cdot \log 1000 \\ &= -2 \cdot 1264.013 + 2 \cdot \log 1000 \\ &\approx -2514.21\end{aligned}$$

and

$$\begin{aligned}\mathrm{BIC}^{(\mathrm{exp})} &= -2\ell_{\mathbf{Y}}^{(\mathrm{exp})}\left(\hat{c}^{\mathrm{MLE}}\right) + d^{(\mathrm{exp})} \cdot \log 1000 \\ &= -2 \cdot 1264.169 + \log 1000 \\ &\approx -2521.43.\end{aligned}$$

According to the BIC, the model with the smallest BIC value should be preferred. Since $\mathrm{BIC}^{(\mathrm{gam})} > \mathrm{BIC}^{(\mathrm{exp})}$, we choose the exponential model. Note that the gamma model gives the better in-sample fit than the exponential model. But if we adjust this in-sample fit by the number of parameters used, we conclude that the exponential model probably has the better out-of-sample performance (better predictive power).

Non-Life Insurance: Mathematics and Statistics

Solution sheet 8

Solution 8.1 Panjer Algorithm

For the expected yearly claim amount π_0 we have

$$\pi_0 = \mathbb{E}[S] = \mathbb{E}[N]\,\mathbb{E}[Y_1] = 1 \cdot \mathbb{E}[k+Z] = k + \mathbb{E}[Z] = k + \exp\left\{\mu + \frac{\sigma^2}{2}\right\} \approx 4123.872.$$

Let Y_i^+ denote the discretized claim sizes using a span of $s = 10$, where we put all the probability mass to the upper end of the intervals. If we write $g_m = \mathbb{P}[Y_1^+ = sm]$ for $m \in \mathbb{N}$, then we have

$$g_1 = g_2 = \cdots = g_{10} = 0,$$

since $\mathbb{P}[Y_1^+ \leq k] = \mathbb{P}[Z \leq 0] = 0$ and $k = 10s$. For all $l \geq 11$, we get

$$\begin{aligned} g_l &= \mathbb{P}[Y_1^+ = sl] \\ &= \mathbb{P}[Y_1^+ = k + s(l-10)] \\ &= \mathbb{P}[k + s(l-11) < Y_1 \leq k + s(l-10)] \\ &= \mathbb{P}[Y_1 \leq k + s(l-10)] - \mathbb{P}[Y_1 \leq k + s(l-11)] \\ &= \mathbb{P}[Z \leq s(l-10)] - \mathbb{P}[Z \leq s(l-11)] \\ &= \mathbb{P}\left[\log Z \leq \log(s[l-10])\right] - \mathbb{P}\left[\log Z \leq \log(s[l-11])\right] \\ &= \Phi\left(\frac{\log[s(l-10)] - \mu}{\sigma}\right) - \Phi\left(\frac{\log[s(l-11)] - \mu}{\sigma}\right), \end{aligned}$$

where Φ is the distribution function of the standard Gaussian distribution and where we define $\log 0 = -\infty$. From now on we will replace the claim sizes Y_i with the discretized claim sizes Y_i^+. In particular, we will still write S for the yearly claim amount that changed to

$$S = \sum_{i=1}^{N} Y_i^+.$$

Note that $N \sim \text{Poi}(1)$ has a Panjer distribution with parameters $a = 0$ and $b = 1$, see the proof of Lemma 4.7 of the lecture notes. Applying the Panjer algorithm given in Theorem 4.9 of the lecture notes, we have for $r \in \mathbb{N}_0$

$$f_r \stackrel{\text{def.}}{=} \mathbb{P}[S = sr] = \begin{cases} \mathbb{P}[N=0] & \text{for } r = 0, \\ \sum_{l=1}^{r} \frac{l}{r} g_l f_{r-l} & \text{for } r > 0. \end{cases}$$

Since the yearly amount that the client has to pay by himself is given by

$$S_{\text{ins}} = \min\{S, d\} + \min\{\alpha \cdot (S-d)_+, M\} = \min\{S, d\} + \alpha \cdot \min\left\{(S-d)_+, \frac{M}{\alpha}\right\},$$

$M/\alpha = 7'000$ and the maximal possible franchise is 2'500, we have to apply the Panjer algorithm until we reach $\mathbb{P}[S = 9'500] = f_{950}$. Here we limit ourselves to determine the values of $f_0, \ldots, f_{12}$ to illustrate how the algorithm works. In particular, we have

$$f_0 = \mathbb{P}[N = 0] = e^{-1} \approx 0.36$$

and

$$f_1 = f_2 = \cdots = f_{10} = 0,$$

since $g_1 = g_2 = \cdots = g_{10} = 0$. For $r = 11$ and $r = 12$, we get

$$f_{11} = \sum_{l=1}^{11} \frac{l}{11} g_l f_{11-l} = g_{11} f_0 = \left[\Phi\left(\frac{\log s - \mu}{\sigma} \right) - \Phi\left(\frac{\log 0 - \mu}{\sigma} \right) \right] e^{-1} \approx 7.089 \cdot 10^{-9}$$

and

$$f_{12} = \sum_{l=1}^{12} \frac{l}{12} g_l f_{12-l} = g_{12} f_0 = \left[\Phi\left(\frac{\log 2s - \mu}{\sigma} \right) - \Phi\left(\frac{\log s - \mu}{\sigma} \right) \right] e^{-1} \approx 2.786 \cdot 10^{-7}.$$

Using the discretized claim sizes, the yearly expected amount π_{ins} paid by the client is given by

$$\pi_{\text{ins}} = \mathbb{E}[S_{\text{ins}}] = \mathbb{E}\left[\min\{S, d\}\right] + \alpha\, \mathbb{E}\left[\min\left\{ (S-d)_+, \frac{M}{\alpha} \right\} \right],$$

where we have

$$\mathbb{E}\left[\min\{S, d\}\right] = \sum_{r=0}^{d/s} f_r s r + d \left(1 - \sum_{r=0}^{d/s} f_r \right) = d + \sum_{r=0}^{d/s} f_r (sr - d)$$

and

$$\begin{aligned} \mathbb{E}\left[\min\left\{ (S-d)_+, \frac{M}{\alpha} \right\} \right] &= \sum_{r=d/s+1}^{d/s+M/s\alpha} f_r (sr - d) + \frac{M}{\alpha} \left(1 - \sum_{r=0}^{d/s+M/\alpha} f_r \right) \\ &= \frac{M}{\alpha} + \sum_{r=d/s+1}^{d/s+M/s\alpha} f_r \left(sr - d - \frac{M}{\alpha} \right) - \frac{M}{\alpha} \sum_{r=0}^{d/s} f_r. \end{aligned}$$

Therefore, we get

$$\begin{aligned} \pi_{\text{ins}} &= d + \sum_{r=0}^{d/s} f_r (sr - d) + \alpha \left[\frac{M}{\alpha} + \sum_{r=d/s+1}^{d/s+M/s\alpha} f_r \left(sr - d - \frac{M}{\alpha} \right) - \frac{M}{\alpha} \sum_{r=0}^{d/s} f_r \right] \\ &= d + M + \sum_{r=0}^{d/s} f_r (sr - d - M) + \sum_{r=d/s+1}^{d/s+M/s\alpha} \alpha f_r \left(sr - d - \frac{M}{\alpha} \right). \end{aligned}$$

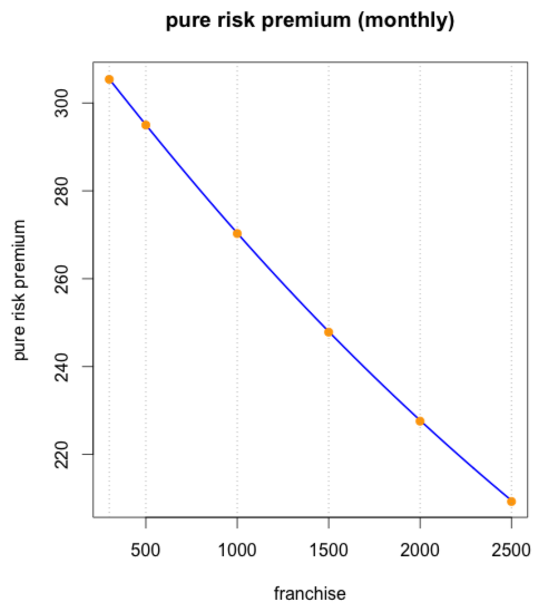
Finally, if the client has chosen franchise d, then the monthly pure risk premium π is given by

$$\begin{aligned} \pi &= \frac{\pi_0 - \pi_{\text{ins}}}{12} \\ &= \frac{1}{12} \left[k + \exp\left\{ \mu + \frac{\sigma^2}{2} \right\} - d - M - \sum_{r=0}^{d/s} f_r (sr - d - M) - \sum_{r=d/s+1}^{d/s+M/s\alpha} \alpha f_r \left(sr - d - \frac{M}{\alpha} \right) \right]. \end{aligned}$$

In the end, we get the following monthly pure risk premiums for the different franchises:

d	300	500	1’000	1’500	2’000	2’500
π	307	297	274	253	233	216

More generally, the monthly pure risk premium as a function of the franchise, which is allowed to vary between 300 CHF and 2’500 CHF, looks as follows:



Note that the above values only represent the pure risk premiums. In order get the premiums that the customer has to pay in the end, we would need to add an appropriate risk-loading, which may vary between different health insurance companies. The above plot can be created by the R-code given below, where we calculated the premiums using two different discretizations of the claim sizes: in one we put the probability mass to the upper end of the intervals and in the other to the lower end of the intervals. However, the resulting premiums for these two versions are basically the same.

```
### Define the function KK_premium with the variables:
### lambda = mean number of claims
### mu = mean parameter of log-normal distribution
### sigma2 = variance parameter of log-normal distribution
### span = span size used in the Panjer algorithm
### shift = shift of the translated log-normal distribution
KK_premium <- function(lambda, mu, sigma2, span, shift){
  ### we will calculate the distribution of S until M (M = 2500 +
      7000)
  M <- 9500

  ### number of steps
  m <- M/span

  ### we won't have any mass until we reach shift, which happens at
       the k0-th step
  k0 <- shift/span

  ### initialize array where mass is put to the lower end of the
      interval
  g_min <- array(0, dim=c(m+1,1))

  ### initialize array where mass is put to the upper end of the
      interval
  g_max <- array(0, dim=c(m+1,1))
```

```
### discretize the log-normal distribution putting the mass to the lower end of the interval
for (k in (k0+1):(m+1)){g_min[k,1] <- pnorm(log((k-k0)*span), mean=mu, sd=sqrt(sigma2))-pnorm(log((k-k0-1)*span), mean=mu, sd=sqrt(sigma2))}

### discretize the log-normal distribution putting the mass to the upper end of the interval
g_max[2:(m+1),1] <- g_min[1:m,1]

### initialize matrix, where we will store the probability distribution of S
f1 <- matrix(0, nrow=m+1, ncol=3)

### store the probability of getting zero claims (in both lower bound and upper bound)
f1[1,1] <- exp(-lambda*(1-g_min[1,1]))
f1[1,2] <- exp(-lambda*(1-g_max[1,1]))

### calculate the values "l * g_{l}" of the discretized claim sizes (lower bound and upper bound), we need these values in the Panjer algorithm
h1 <- matrix(0, nrow=m, ncol=3)
for (i in 1:m){
  h1[i,1] <- g_min[i+1,1]*(i+1)
  h1[i,2] <- g_max[i+1,1]*(i+1)
}

### Panjer algorithm (note that in the Poisson case we have a = 0 and b = lambda*v, which is just lambda here)
for (r in 1:m){
  f1[r+1,1] <- lambda/r*(t(f1[1:r,1])%*%h1[r:1,1])
  f1[r+1,2] <- lambda/r*(t(f1[1:r,2])%*%h1[r:1,2])
  f1[r+1,3] <- r * span
}

### maximal and minimal franchise
m1 <- 2500
m0 <- 300

### number of iterations needed to get to m1 and m0
i1 <- m1/span+1
i0 <- m0/span+1

### calculate the part that the insured pays by himself
franchise <- array(NA, c(i1, 3))
for (i in i0:i1){
  franchise[i,1] <- f1[i,3]   ### this represents the franchise
  franchise[i,2] <- sum(f1[1:i,1]*f1[1:i,3]) + f1[i,3] * (1-sum(f1[1:i,1]))
  franchise[i,2] <- franchise[i,2] + sum(f1[(i+1):(i+7000/span)
```

```
      ,1]*f1[2:(7000/span+1),3])*0.1 + 700 * (1-sum(f1[1:(i+7000/
      span),1]))
    franchise[i,3] <- sum(f1[1:i,2]*f1[1:i,3]) + f1[i,3] * (1-sum(
      f1[1:i,2]))
    franchise[i,3] <- franchise[i,3] + sum(f1[(i+1):(i+7000/span)
      ,2]*f1[2:(7000/span+1),3])*0.1 + 700 * (1-sum(f1[1:(i+7000/
      span),2]))
  }

  ### calculate the price of the monthly premium
  price <- array(NA, c(i1, 3))
  price[,1] <- franchise[,1]    ### this represents the franchise
  price[,2:3] <- (lambda*(exp(mu+sigma2/2)+shift) - franchise
    [,2:3])/12
  price
}

### Load the add-on packages stats and MASS
require(stats)
require(MASS)

### Determine values for the input parameters of the function KK_
   premium
lambda <- 1
mu <- 7.8
sigma2 <- 1
span <- 10
shift <- 100

### The coefficient of variation of the translated log-normal
   distribution is given by
exp(mu+sigma2/2)*sqrt(exp(sigma2)-1)/(shift+exp(mu+sigma2/2))

### Run the function KK_premium
price <- KK_premium(lambda, mu, sigma2, span, shift)

### Plot the monthly pure risk premium as a function of the
   franchise
plot(x=price[,1], y=price[,2], lwd=2, col="blue", type='l', ylab="
   pure risk premium", xlab="franchise", main="pure risk premium (
   monthly)")
lines(x=price[,1], y=price[,2], lwd=1, col="blue")
points(x=c(300,500, 1000, 1500, 2000, 2500), y=price[c(300,500,
   1000, 1500, 2000, 2500)/span+1,3], pch=19, col="orange")
abline(v=c(300, 500, 1000, 1500, 2000, 2500), col="darkgray", lty
   =3)

### Give the monthly pure risk premiums for the six franchises
   listed on the exercise sheet
round(price[c(300,500, 1000, 1500, 2000, 2500)/span+1,2])
round(price[c(300,500, 1000, 1500, 2000, 2500)/span+1,3])
```

Solution 8.2 Variance Loading Principle

(a) Let S_1, S_2, S_3 be the total claim amounts of the passenger cars, delivery vans and trucks, respectively. Then, according Proposition 2.11 of the lecture notes, for the expected total claim amounts we have

$$\mathbb{E}[S_i] = \lambda_i v_i \, \mathbb{E}\left[Y_1^{(i)}\right],$$

for all $i \in \{1, 2, 3\}$. Using the data given in the table on the exercise sheet, we get

$$\begin{aligned}\mathbb{E}[S_1] &= 0.25 \cdot 40 \cdot 2\text{'}000 = 20\text{'}000,\\ \mathbb{E}[S_2] &= 0.23 \cdot 30 \cdot 1\text{'}700 = 11\text{'}730 \quad \text{and}\\ \mathbb{E}[S_3] &= 0.19 \cdot 10 \cdot 4\text{'}000 = 7\text{'}600.\end{aligned}$$

If we write S for the total claim amount of the car fleet, we can conclude that

$$\mathbb{E}[S] = \mathbb{E}[S_1 + S_2 + S_3] = \mathbb{E}[S_1] + \mathbb{E}[S_2] + \mathbb{E}[S_3] = 39\text{'}330.$$

(b) Again using Proposition 2.11 of the lectures notes, we get

$$\mathrm{Var}[S_i] = \lambda_i v_i \, \mathbb{E}\left[\left(Y_1^{(i)}\right)^2\right] = \lambda_i v_i \left(\mathrm{Var}\left(Y_1^{(i)}\right) + \mathbb{E}\left[Y_1^{(i)}\right]^2\right) = \lambda_i v_i \, \mathbb{E}\left[Y_1^{(i)}\right]^2 \left(\mathrm{Vco}(Y_1^{(i)})^2 + 1\right),$$

for all $i \in \{1, 2, 3\}$. Using the data given in the table on the exercise sheet, we find

$$\begin{aligned}\mathrm{Var}(S_1) &= 0.25 \cdot 40 \cdot 2\text{'}000^2(2.5^2 + 1) = 290\text{'}000\text{'}000,\\ \mathrm{Var}(S_2) &= 0.23 \cdot 30 \cdot 1\text{'}700^2(2^2 + 1) = 99\text{'}705\text{'}000 \quad \text{and}\\ \mathrm{Var}(S_3) &= 0.19 \cdot 10 \cdot 4\text{'}000^2(3^2 + 1) = 304\text{'}000\text{'}000.\end{aligned}$$

Since S_1, S_2 and S_3 are independent by assumption, we get for the variance of the total claim amount S of the car fleet

$$\mathrm{Var}(S) = \mathrm{Var}(S_1) + \mathrm{Var}(S_2) + \mathrm{Var}(S_3) = 693\text{'}705\text{'}000.$$

Using the variance loading principle with $\alpha = 3 \cdot 10^{-6}$, we get for the premium π of the car fleet

$$\pi = \mathbb{E}[S] + \alpha \mathrm{Var}(S) = 39\text{'}330 + 3 \cdot 10^{-6} \cdot 693\text{'}705\text{'}000 \approx 39\text{'}330 + 2\text{'}081 = 41\text{'}411.$$

Note that we have

$$\frac{\pi - \mathbb{E}[S]}{\mathbb{E}[S]} = \frac{\alpha \mathrm{Var}(S)}{\mathbb{E}[S]} \approx \frac{2\text{'}081}{39\text{'}330} \approx 5.3\%.$$

Thus, the loading $\pi - \mathbb{E}[S]$ is given by 5.3% of the pure risk premium.

Solution 8.3 Panjer Distribution

If we write

$$p_k = \mathbb{P}[N = k]$$

for all $k \in \mathbb{N}$, then, by definition of the Panjer distribution, we have

$$p_k = p_{k-1}\left(a + \frac{b}{k}\right),$$

for all k in the range of N. We can use this recursion to calculate $\mathbb{E}[N]$ and $\mathrm{Var}(N)$. Note that the range of N is $\mathbb{N}$ if $a \geq 0$ and it is $\{0, 1, \ldots, n\}$ for some $n \in \mathbb{N}_{\geq 1}$ if $a < 0$.

First, we consider the case where $a < 0$, i.e. where the range of N is $\{0, 1, \dots, n\}$. According to the proof of Lemma 4.7 of the lecture notes, we have

$$n = -\frac{a+b}{a}. \tag{1}$$

For the expectation of N, we get

$$\begin{aligned}
\mathbb{E}[N] &= \sum_{k=0}^{n} k\, p_k \\
&= \sum_{k=1}^{n} k\, p_k \\
&= \sum_{k=1}^{n} k\, p_{k-1} \left(a + \frac{b}{k}\right) \\
&= a \sum_{k=1}^{n} k\, p_{k-1} + b \sum_{k=1}^{n} p_{k-1} \\
&= a \sum_{k=0}^{n-1} (k+1)\, p_k + b \sum_{k=0}^{n-1} p_k \\
&= a \sum_{k=0}^{n-1} k\, p_k + (a+b) \sum_{k=0}^{n-1} p_k \\
&= a\,(\mathbb{E}[N] - np_n) + (a+b)(1-p_n) \\
&= a\,\mathbb{E}[N] + a + b + p_n(-an - a - b).
\end{aligned}$$

Using (1), we get

$$-an - a - b = a\,\frac{a+b}{a} - a - b = 0. \tag{2}$$

Hence, the above expression for $\mathbb{E}[N]$ simplifies to

$$\mathbb{E}[N] = a\mathbb{E}[N] + a + b,$$

from which we can conclude that

$$\mathbb{E}[N] = \frac{a+b}{1-a}.$$

In order to get the variance of N, we first calculate the second moment of N:

$$
\begin{aligned}
\mathbb{E}[N^2] &= \sum_{k=0}^{n} k^2 \, p_k \\
&= \sum_{k=1}^{n} k^2 \, p_k \\
&= \sum_{k=1}^{n} k^2 \, p_{k-1} \left(a + \frac{b}{k}\right) \\
&= a \sum_{k=1}^{n} k^2 \, p_{k-1} + b \sum_{k=1}^{n} k \, p_{k-1} \\
&= a \sum_{k=0}^{n-1} (k+1)^2 \, p_k + b \sum_{k=0}^{n-1} (k+1) \, p_k \\
&= a \sum_{k=0}^{n-1} k^2 \, p_k + (2a+b) \sum_{k=0}^{n-1} k \, p_k + (a+b) \sum_{k=0}^{n-1} p_k \\
&= a \, (\mathbb{E}[N^2] - n^2 p_n) + (2a+b)(\mathbb{E}[N] - n p_n) + (a+b)(1-p_n) \\
&= a \, \mathbb{E}[N^2] + (2a+b) \, \mathbb{E}[N] + a + b + p_n[-an^2 - (2a+b)n - a - b].
\end{aligned}
$$

Using (1), we get

$$
\begin{aligned}
-an^2 - (2a+b)n - a - b &= -a \left(\frac{a+b}{a}\right)^2 + (2a+b)\frac{a+b}{a} - a - b \\
&= -\frac{a^2 + 2ab + b^2}{a} + \frac{2a^2 + 3ab + b^2}{a} - \frac{a^2 + ab}{a} \\
&= 0.
\end{aligned}
\tag{3}
$$

Hence, the above expression for $\mathbb{E}[N^2]$ simplifies to

$$
\mathbb{E}[N^2] = a \, \mathbb{E}[N^2] + (2a+b) \, \mathbb{E}[N] + a + b,
$$

from which we get

$$
\begin{aligned}
\mathbb{E}[N^2] &= \frac{(2a+b) \, \mathbb{E}[N] + a + b}{1-a} \\
&= \frac{(2a+b) \, (a+b) + (a+b)(1-a)}{(1-a)^2} \\
&= \frac{2a^2 + 3ab + b^2 + a - a^2 + b - ab}{(1-a)^2} \\
&= \frac{(a+b)^2 + a + b}{(1-a)^2}.
\end{aligned}
$$

Finally, the variance of N then is

$$
\mathrm{Var}(N) = \mathbb{E}[N^2] - \mathbb{E}[N]^2 = \frac{(a+b)^2 + a + b}{(1-a)^2} - \frac{(a+b)^2}{(1-a)^2} = \frac{a+b}{(1-a)^2}.
$$

In the case where $a \geq 0$, i.e. where the range of N is $\mathbb{N}$, we can perform analogous calculations with the only difference that the index of summation in all the sums involved goes up to ∞ instead of stopping at n. As a consequence, the calculations in (2) and in (3) aren't necessary anymore. The formulas for $\mathbb{E}[N]$ and $\mathrm{Var}(N)$, however, remain the same.

The ratio of $\mathrm{Var}(N)$ to $\mathbb{E}[N]$ is given by

$$\frac{\mathrm{Var}(N)}{\mathbb{E}[N]} = \frac{a+b}{(1-a)^2}\,\frac{1-a}{a+b} = \frac{1}{1-a}.$$

Note that if $a < 0$, i.e. if N has a binomial distribution, we have $\mathrm{Var}(N) < \mathbb{E}[N]$. If $a = 0$, i.e. if N has a a Poisson distribution, we have $\mathrm{Var}(N) = \mathbb{E}[N]$. Finally, in the case of $a > 0$, i.e. for a negative-binomial distribution, we have $\mathrm{Var}(N) > \mathbb{E}[N]$.

Non-Life Insurance: Mathematics and Statistics

Solution sheet 9

Solution 9.1 Utility Indifference Price

(a) Suppose that there exist two utility indifference prices $\pi_1 = \pi_1(u, S, c_0)$ and $\pi_2 = \pi_2(u, S, c_0)$ with $\pi_1 \neq \pi_2$. By definition of a utility indifference price, we have

$$\mathbb{E}[u(c_0 + \pi_1 - S)] = u(c_0) = \mathbb{E}[u(c_0 + \pi_2 - S)]. \tag{1}$$

Without loss of generality, we assume that $\pi_1 < \pi_2$. Then we have

$$c_0 + \pi_1 - S < c_0 + \pi_2 - S \qquad \text{a.s.},$$

which implies

$$u(c_0 + \pi_1 - S) < u(c_0 + \pi_2 - S) \qquad \text{a.s.},$$

since u is a utility function and, thus, strictly increasing by definition. Finally, by taking the expectation, we get

$$\mathbb{E}[u(c_0 + \pi_1 - S)] < \mathbb{E}[u(c_0 + \pi_2 - S)],$$

which is a contradiction to (1). We conclude that if the utility indifference price π exists, then it is unique. Moreover, being a utility function, u is strictly concave by definition. Hence, we can apply Jensen's inequality to get

$$u(c_0) = \mathbb{E}[u(c_0 + \pi - S)] < u(\mathbb{E}[c_0 + \pi - S]) = u(c_0 + \pi - \mathbb{E}[S]).$$

Note that we used that S is non-deterministic and, thus, Jensen's inequality is strict. Since u is strictly increasing, this implies $\pi - \mathbb{E}[S] > 0$, i.e. $\pi > \mathbb{E}[S]$.

(b) Note that

$$\mathbb{E}\left[Y_1^{(1)}\right] = \frac{\gamma}{c} = \frac{20}{0.01} = 2\text{'}000$$

and that

$$\mathbb{E}\left[Y_1^{(2)}\right] = \frac{1}{0.005} = 200.$$

Since S_1 and S_2 both have a compound Poisson distribution, Proposition 2.11 of the lecture notes gives

$$\mathbb{E}[S_1] = \lambda_1 v_1 \mathbb{E}\left[Y_1^{(1)}\right] = \frac{1}{2} \cdot 2\text{'}000 \cdot 2\text{'}000 = 2\text{'}000\text{'}000$$

and

$$\mathbb{E}[S_2] = \lambda_2 v_2 \mathbb{E}\left[Y_1^{(2)}\right] = \frac{1}{10} \cdot 10\text{'}000 \cdot 200 = 200\text{'}000.$$

We conclude that

$$\mathbb{E}[S] = \mathbb{E}[S_1 + S_2] = \mathbb{E}[S_1] + \mathbb{E}[S_2] = 2\text{'}200\text{'}000.$$

(c) The utility indifference price $\pi = \pi(u, S, c_0)$ is defined through the equation

$$u(c_0) = \mathbb{E}[u(c_0 + \pi - S)].$$

Using that the utility function u is given by

$$u(x) = 1 - \frac{1}{\alpha} \exp\{-\alpha x\},$$

for all $x \in \mathbb{R}$, with $\alpha = 1.5 \cdot 10^{-6}$, we get

$$\begin{aligned} u(c_0) = \mathbb{E}[u(c_0 + \pi - S)] &\iff 1 - \frac{1}{\alpha}\exp\{-\alpha c_0\} = \mathbb{E}\left[1 - \frac{1}{\alpha}\exp\{-\alpha(c_0 + \pi - S)\}\right] \\ &\iff \exp\{-\alpha c_0\} = \mathbb{E}\left[\exp\{-\alpha(c_0 + \pi - S)\}\right] \\ &\iff \exp\{\alpha\pi\} = \mathbb{E}\left[\exp\{\alpha S\}\right] \\ &\iff \pi = \frac{1}{\alpha}\log \mathbb{E}\left[\exp\{\alpha S\}\right]. \end{aligned}$$

Note that we can write $S = S_1 + S_2$ and use the independence of S_1 and S_2 to get

$$\begin{aligned} \pi &= \frac{1}{\alpha}\log \mathbb{E}\left[\exp\{\alpha(S_1 + S_2)\}\right] \\ &= \frac{1}{\alpha}\log\left(\mathbb{E}\left[\exp\{\alpha S_1\}\right]\mathbb{E}\left[\exp\{\alpha S_2\}\right]\right) \\ &= \frac{1}{\alpha}\left(\log \mathbb{E}\left[\exp\{\alpha S_1\}\right] + \log \mathbb{E}\left[\exp\{\alpha S_2\}\right]\right) \\ &= \frac{1}{\alpha}\left[\log M_{S_1}(\alpha) + \log M_{S_2}(\alpha)\right], \end{aligned}$$

where M_{S_1} and M_{S_2} denote the moment generating functions of S_1 and S_2, respectively. Moreover, since S_1 and S_2 both have a compound Poisson distribution, Proposition 2.11 of the lecture notes gives

$$\pi = \frac{1}{\alpha}\left(\lambda_1 v_1 \left[M_{Y_1^{(1)}}(\alpha) - 1\right] + \lambda_2 v_2 \left[M_{Y_1^{(2)}}(\alpha) - 1\right]\right),$$

where $M_{Y_1^{(1)}}$ and $M_{Y_1^{(2)}}$ denote the moment generating functions of $Y_1^{(1)}$ and $Y_1^{(2)}$, respectively. Using that $Y_1^{(1)} \sim \Gamma(\gamma = 20, c = 0.01)$ and that $Y_1^{(2)} \sim \text{expo}(0.005)$, we get

$$M_{Y_1^{(1)}}(\alpha) = \left(\frac{c}{c - \alpha}\right)^{\gamma} = \left(\frac{0.01}{0.01 - 1.5 \cdot 10^{-6}}\right)^{20}$$

and

$$M_{Y_1^{(2)}}(\alpha) = \frac{0.005}{0.005 - \alpha} = \frac{0.005}{0.005 - 1.5 \cdot 10^{-6}}.$$

In particular, since $\alpha < c$ and $\alpha < 0.005$, both $M_{Y_1^{(1)}}(\alpha)$ and $M_{Y_1^{(2)}}(\alpha)$ and thus also $M_{S_1}(\alpha)$ and $M_{S_2}(\alpha)$ exist. Inserting all the numerical values, we find the utility indifference price

$$\begin{aligned} \pi &= \frac{2}{3} \cdot 10^6 \left(\frac{1}{2} \cdot 2'000 \cdot \left[\left(\frac{0.01}{0.01 - 1.5 \cdot 10^{-6}}\right)^{20} - 1\right] + \frac{1}{10} \cdot 10'000 \cdot \left[\frac{0.005}{0.005 - 1.5 \cdot 10^{-6}} - 1\right]\right) \\ &= 2'203'213. \end{aligned}$$

Note that we have

$$\frac{\pi - \mathbb{E}[S]}{\mathbb{E}[S]} = \frac{2'203'213 - 2'200'000}{2'200'000} = \frac{3'213}{2'200'000} \approx 0.146\%.$$

Thus, the loading $\pi - \mathbb{E}[S]$ is given by approximately 0.146% of the pure risk premium.

(d) The moment generating function M_X of $X \sim \mathcal{N}(\mu, \sigma^2)$ for some $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ is given by

$$M_X(r) = \exp\left\{ r\mu + \frac{r^2\sigma^2}{2} \right\},$$

for all $r \in \mathbb{R}$. Hence, if we assume Gaussian distributions for S_1 and S_2, then we get

$$\begin{aligned}
\pi &= \frac{1}{\alpha}\left[\log M_{S_1}(\alpha) + \log M_{S_2}(\alpha)\right] \\
&= \frac{1}{\alpha}\left(\alpha\mathbb{E}[S_1] + \frac{\alpha^2}{2}\mathrm{Var}(S_1) + \alpha\mathbb{E}[S_2] + \frac{\alpha^2}{2}\mathrm{Var}(S_2)\right) \\
&= \mathbb{E}[S_1] + \mathbb{E}[S_2] + \frac{\alpha}{2}[\mathrm{Var}(S_1) + \mathrm{Var}(S_2)] \\
&= \mathbb{E}[S] + \frac{\alpha}{2}\mathrm{Var}(S),
\end{aligned}$$

where in the last equation we used that S_1 and S_2 are independent. We see that in this case the utility indifference price is given according to a variance loading principle. Since here we assume Gaussian distributions for S_1 and S_2 with the same corresponding first two moments as in the compound Poisson case in part (c), in order to calculate $\mathrm{Var}(S_1)$ and $\mathrm{Var}(S_2)$, we again assume that S_1 and S_2 have compound Poisson distributions. Note that

$$\mathbb{E}\left[(Y_1^{(1)})^2\right] = \frac{\gamma(\gamma+1)}{c^2} = \frac{20 \cdot 21}{0.01^2} = 4\text{'}200\text{'}000,$$

and that

$$\mathbb{E}\left[(Y_1^{(2)})^2\right] = \frac{2}{0.005^2} = 80\text{'}000.$$

Then Proposition 2.11 of the lecture notes gives

$$\mathrm{Var}(S_1) = \lambda_1 v_1 \mathbb{E}\left[(Y_1^{(1)})^2\right] = \frac{1}{2} \cdot 2\text{'}000 \cdot 4\text{'}200\text{'}000 = 4\text{'}200\text{'}000\text{'}000$$

and

$$\mathrm{Var}(S_2) = \lambda_2 v_2 \mathbb{E}\left[(Y_1^{(2)})^2\right] = \frac{1}{10} \cdot 10\text{'}000 \cdot 80\text{'}000 = 80\text{'}000\text{'}000,$$

which leads to

$$\mathrm{Var}(S) = \mathrm{Var}(S_1 + S_2) = \mathrm{Var}(S_1) + \mathrm{Var}(S_2) = 4\text{'}280\text{'}000\text{'}000.$$

We conclude that the utility indifference price is given by

$$\pi = \mathbb{E}[S] + \frac{\alpha}{2}\mathrm{Var}(S) = 2\text{'}200\text{'}000 + \frac{1.5 \cdot 10^{-6}}{2} \cdot 4\text{'}280\text{'}000\text{'}000 = 2\text{'}203\text{'}210.$$

Note that we have

$$\frac{\pi - \mathbb{E}[S]}{\mathbb{E}[S]} = \frac{2\text{'}203\text{'}210 - 2\text{'}200\text{'}000}{2\text{'}200\text{'}000} = \frac{3\text{'}210}{2\text{'}200\text{'}000} \approx 0.146\%.$$

Thus, as in part (c), the loading $\pi - \mathbb{E}[S]$ is given by approximately 0.146% of the pure risk premium. The reason why we get the same results in (c) and (d) is the Central Limit Theorem. In particular, neither the gamma distribution nor the exponential distribution are heavy-tailed distributions and thus $\lambda_1 v_1 = \lambda_2 v_2 = 1\text{'}000$ are large enough for the normal approximations to be valid approximations for the compound Poisson distributions.

Solution 9.2 Value-at-Risk and Expected Shortfall

(a) Since $S \sim \mathrm{LN}(\mu, \sigma^2)$ with $\mu = 20$ and $\sigma^2 = 0.015$, we have

$$\mathbb{E}[S] = \exp\left\{\mu + \frac{\sigma^2}{2}\right\} \approx 488'817'614.$$

Let z denote the VaR of $S - \mathbb{E}[S]$ at security level $1-q = 99.5\%$. Then, since the distribution function of a lognormal distribution is continuous and strictly increasing, z is defined via the equation

$$\mathbb{P}[S - \mathbb{E}[S] \leq z] = 1 - q.$$

By writing Φ for the distribution function of a standard Gaussian distribution, we can calculate z as follows

$$\begin{aligned}
\mathbb{P}[S - \mathbb{E}[S] \leq z] = 1 - q \quad &\iff \quad \mathbb{P}[S \leq z + \mathbb{E}[S]] = 1 - q \\
&\iff \quad \mathbb{P}\left[\frac{\log S - \mu}{\sigma} \leq \frac{\log(z + \mathbb{E}[S]) - \mu}{\sigma}\right] = 1 - q \\
&\iff \quad \Phi\left[\frac{\log(z + \mathbb{E}[S]) - \mu}{\sigma}\right] = 1 - q \\
&\iff \quad \log(z + \mathbb{E}[S]) = \mu + \sigma \cdot \Phi^{-1}(1-q) \\
&\iff \quad z = \exp\left\{\mu + \sigma \cdot \Phi^{-1}(1-q)\right\} - \mathbb{E}[S] \\
&\iff \quad z = \exp\{\mu\}\left(\exp\{\sigma \cdot \Phi^{-1}(1-q)\} - \exp\left\{\frac{\sigma^2}{2}\right\}\right).
\end{aligned}$$

For $1 - q = 99.5\%$, we have $\Phi^{-1}(1-q) \approx 2.576$. Thus, we get

$$z \approx 176'299'286.$$

In particular, π_{CoC} is then given by

$$\pi_{\mathrm{CoC}} = \mathbb{E}[S] + r_{\mathrm{CoC}} \cdot z \approx 488'817'614 + 0.06 \cdot 176'299'286 \approx 499'395'571.$$

Note that we have

$$\frac{\pi_{\mathrm{CoC}} - \mathbb{E}[S]}{\mathbb{E}[S]} \approx \frac{499'395'571 - 488'817'614}{488'817'614} = \frac{10'577'957}{488'817'614} \approx 2.164\%.$$

Thus, the loading $\pi_{\mathrm{CoC}} - \mathbb{E}[S]$ is given by approximately 2.164% of the pure risk premium.

(b) For all $u \in (0,1)$, let VaR_u and ES_u denote the VaR risk measure and the expected shortfall risk measure, respectively, at security level u. Note that actually in part (a) we found that

$$\mathrm{VaR}_u(S - \mathbb{E}[S]) = \exp\left\{\mu + \sigma \cdot \Phi^{-1}(u)\right\} - \mathbb{E}[S]$$

and that by a similar computation we get

$$\mathrm{VaR}_u(S) = \exp\left\{\mu + \sigma \cdot \Phi^{-1}(u)\right\},$$

for all $u \in (0,1)$. In particular, we have

$$\mathrm{VaR}_u(S - \mathbb{E}[S]) + \mathbb{E}[S] = \mathrm{VaR}_u(S)$$

for all $u \in (0,1)$. Since the distribution function of S is continuous and strictly increasing, according to Example 6.26 of the lecture notes we have

$$\begin{aligned}
\mathrm{ES}_{1-q}(S-\mathbb{E}[S]) &= \mathbb{E}\left[S-\mathbb{E}[S] \,|\, S-\mathbb{E}[S] \geq \mathrm{VaR}_{1-q}(S-\mathbb{E}[S])\right] \\
&= \mathbb{E}\left[S-\mathbb{E}[S] \,|\, S \geq \mathrm{VaR}_{1-q}(S)\right] \\
&= \mathbb{E}\left[S \,|\, S \geq \mathrm{VaR}_{1-q}(S)\right] - \mathbb{E}[S] \\
&= \mathrm{ES}_{1-q}(S) - \mathbb{E}[S].
\end{aligned}$$

By definition of the mean excess function $e_S(\cdot)$ of S we have

$$\mathrm{ES}_{1-q}(S) = \mathbb{E}\left[S-\mathrm{VaR}_{1-q}(S) \,|\, S \geq \mathrm{VaR}_{1-q}(S)\right] + \mathrm{VaR}_{1-q}(S) = e_S[\mathrm{VaR}_{1-q}(S)] + \mathrm{VaR}_{1-q}(S).$$

Moreover, according to the formula given in Chapter 3.2.3 of the lecture notes, the mean excess function $e_S[\mathrm{VaR}_{1-q}(S)]$ above level $\mathrm{VaR}_{1-q}(S)$ is given by

$$e_S[\mathrm{VaR}_{1-q}(S)] = \mathbb{E}[S] \left(\frac{1-\Phi\left[\frac{\log \mathrm{VaR}_{1-q}(S)-\mu-\sigma^2}{\sigma}\right]}{1-\Phi\left[\frac{\log \mathrm{VaR}_{1-q}(S)-\mu}{\sigma}\right]} \right) - \mathrm{VaR}_{1-q}(S).$$

Using the formula calculated above for $\mathrm{VaR}_u(S)$ with $u = 1-q$, we get

$$\begin{aligned}
\mathrm{ES}_{1-q}(S) &= \mathbb{E}[S] \left(\frac{1-\Phi\left[\frac{\log \mathrm{VaR}_{1-q}(S)-\mu-\sigma^2}{\sigma}\right]}{1-\Phi\left[\frac{\log \mathrm{VaR}_{1-q}(S)-\mu}{\sigma}\right]} \right) \\
&= \mathbb{E}[S] \left(\frac{1-\Phi\left[\frac{\mu+\sigma\cdot\Phi^{-1}(1-q)-\mu-\sigma^2}{\sigma}\right]}{1-\Phi\left[\frac{\mu+\sigma\cdot\Phi^{-1}(1-q)-\mu}{\sigma}\right]} \right) \\
&= \mathbb{E}[S] \left(\frac{1-\Phi\left[\Phi^{-1}(1-q)-\sigma\right]}{1-\Phi\left[\Phi^{-1}(1-q)\right]} \right) \\
&= \mathbb{E}[S]\, \frac{1}{q} \left(1-\Phi\left[\Phi^{-1}(1-q)-\sigma\right]\right).
\end{aligned}$$

In particular, we have found

$$\begin{aligned}
\mathrm{ES}_{1-q}(S-\mathbb{E}[S]) &= \frac{1}{q}\,\mathbb{E}[S]\left(1-\Phi\left[\Phi^{-1}(1-q)-\sigma\right]\right) - \mathbb{E}[S] \\
&= \frac{1}{q}\,\mathbb{E}[S]\left(1-q-\Phi\left[\Phi^{-1}(1-q)-\sigma\right]\right) \\
&= \frac{1}{q}\,\exp\left\{\mu+\frac{\sigma^2}{2}\right\}\left(1-q-\Phi\left[\Phi^{-1}(1-q)-\sigma\right]\right).
\end{aligned}$$

For $1-q = 99\%$, we get

$$\mathrm{ES}_{99\%}(S-\mathbb{E}[S]) \approx 184'119'256.$$

Finally, π_{CoC} is then given by

$$\pi_{\mathrm{CoC}} = \mathbb{E}[S] + r_{\mathrm{CoC}} \cdot \mathrm{ES}_{99\%}(S-\mathbb{E}[S]) \approx 488'817'614 + 0.06 \cdot 184'119'256 \approx 499'864'769.$$

Note that we have

$$\frac{\pi_{\mathrm{CoC}}-\mathbb{E}[S]}{\mathbb{E}[S]} \approx \frac{499'864'769 - 488'817'614}{488'817'614} = \frac{11'047'155}{488'817'614} \approx 2.26\%.$$

Thus, the loading $\pi_{\mathrm{CoC}} - \mathbb{E}[S]$ is given by approximately 2.26% of the pure risk premium. In particular, the cost-of-capital price in this example is higher using the expected shortfall risk measure at security level 99% than using the VaR risk measure at security level 99.5%.

(c) In parts (a) and (b) we found that

$$\mathrm{VaR}_{99.5\%}(S-\mathbb{E}[S]) < \mathrm{ES}_{99\%}(S-\mathbb{E}[S]).$$

Let $1-q = 99\%$. Now the goal is to find $u \in [0,1]$ such that

$$\mathrm{VaR}_u(S-\mathbb{E}[S]) = \mathrm{ES}_{1-q}(S-\mathbb{E}[S]) \quad\Longleftrightarrow\quad \mathrm{VaR}_u(S) = \mathrm{ES}_{1-q}(S).$$

Note that from part (b) we know

$$\mathrm{VaR}_u(S) = \exp\left\{\mu + \sigma\cdot\Phi^{-1}(u)\right\},$$

for all $u \in (0,1)$, and

$$\mathrm{ES}_{1-q}(S) = \frac{1}{q}\,\mathbb{E}[S]\left(1-\Phi\left[\Phi^{-1}(1-q)-\sigma\right]\right).$$

Hence, we can solve for u to get

$$\begin{aligned} u &= \Phi\left(\frac{\log\left[\frac{1}{q}\,\mathbb{E}[S]\left(1-\Phi\left[\Phi^{-1}(1-q)-\sigma\right]\right)\right]-\mu}{\sigma}\right) \\ &\approx 99.62\%. \end{aligned}$$

We conclude that in this example the cost-of-capital price using the VaR risk measure at security level 99.62% is approximately equal to the cost-of-capital price using the expected shortfall risk measure at security level 99%.

(d) Since $S \sim \mathrm{LN}(\mu,\sigma^2)$ with $\mu = 20$ and $\sigma^2 = 0.015$ and U and V are assumed to be independent, we have

$$U \sim \mathcal{N}(\mu,\sigma^2), \quad V \sim \mathcal{N}(\mu,\sigma^2) \quad \text{and} \quad U+V \sim \mathcal{N}(2\mu, 2\sigma^2).$$

Let $X \sim \mathcal{N}(\tilde{\mu},\tilde{\sigma}^2)$ for some $\tilde{\mu} \in \mathbb{R}$ and $\tilde{\sigma}^2 > 0$. Then $\mathrm{VaR}_{1-q}(X)$ can be calculated as

$$\begin{aligned} \mathbb{P}\left[X \le \mathrm{VaR}_{1-q}(X)\right] = 1-q \quad &\Longleftrightarrow \quad \mathbb{P}\left[\frac{X-\tilde{\mu}}{\tilde{\sigma}} \le \frac{\mathrm{VaR}_{1-q}(X)-\tilde{\mu}}{\tilde{\sigma}}\right] = 1-q \\ &\Longleftrightarrow \quad \Phi\left[\frac{\mathrm{VaR}_{1-q}(X)-\tilde{\mu}}{\tilde{\sigma}}\right] = 1-q \\ &\Longleftrightarrow \quad \mathrm{VaR}_{1-q}(X) = \tilde{\mu} + \tilde{\sigma}\cdot\Phi^{-1}(1-q). \end{aligned}$$

This implies that

$$\mathrm{VaR}_{1-q}(U) + \mathrm{VaR}_{1-q}(V) = \mu + \sigma\cdot\Phi^{-1}(1-q) + \mu + \sigma\cdot\Phi^{-1}(1-q) = 2\mu + 2\sigma\cdot\Phi^{-1}(1-q)$$

and that

$$\mathrm{VaR}_{1-q}(U+V) = 2\mu + \sqrt{2}\sigma\cdot\Phi^{-1}(1-q).$$

Since $\Phi^{-1}(0.45) \approx -0.126$ and $\Phi^{-1}(0.55) \approx 0.126$, we get

$$\mathrm{VaR}_{0.45}(U+V) \approx 39.978 > 39.969 \approx \mathrm{VaR}_{0.45}(U) + \mathrm{VaR}_{0.45}(V)$$

and

$$\mathrm{VaR}_{0.55}(U+V) \approx 40.022 < 40.031 \approx \mathrm{VaR}_{0.55}(U) + \mathrm{VaR}_{0.55}(V).$$

Note that since

$$\begin{aligned} \mathrm{VaR}_{1-q}(U+V) > \mathrm{VaR}_{1-q}(U) + \mathrm{VaR}_{1-q}(V) &\Longleftrightarrow \Phi^{-1}(1-q) > \sqrt{2}\Phi^{-1}(1-q) \\ &\Longleftrightarrow \Phi^{-1}(1-q) < 0, \end{aligned}$$

one can see that in this example

$$\mathrm{VaR}_{1-q}(U+V) > \mathrm{VaR}_{1-q}(U) + \mathrm{VaR}_{1-q}(V)$$

for all $1-q \in \left(0, \frac{1}{2}\right)$ and that

$$\mathrm{VaR}_{1-q}(U+V) < \mathrm{VaR}_{1-q}(U) + \mathrm{VaR}_{1-q}(V)$$

for all $1-q \in \left(\frac{1}{2}, 1\right)$.

Solution 9.3 Esscher Premium

(a) Let $\alpha \in (0, r_0)$ and M_S' and M_S'' denote the first and second derivative of M_S, respectively. According to the proof of Corollary 6.16 of the lecture notes, the Esscher premium π_α can be written as

$$\pi_\alpha = \frac{M_S'(\alpha)}{M_S(\alpha)}.$$

Hence, the derivative of π_α can be calculated as

$$\begin{aligned}
\frac{d}{d\alpha}\pi_\alpha &= \frac{d}{d\alpha}\frac{M_S'(\alpha)}{M_S(\alpha)} \\
&= \frac{M_S''(\alpha)}{M_S(\alpha)} - \left(\frac{M_S'(\alpha)}{M_S(\alpha)}\right)^2 \\
&= \frac{\mathbb{E}\left[S^2 \exp\{\alpha S\}\right]}{M_S(\alpha)} - \left(\frac{\mathbb{E}\left[S \exp\{\alpha S\}\right]}{M_S(\alpha)}\right)^2 \\
&= \frac{1}{M_S(\alpha)} \int_{-\infty}^{\infty} x^2 \exp\{\alpha x\}\, dF(x) - \left[\frac{1}{M_S(\alpha)} \int_{-\infty}^{\infty} x \exp\{\alpha x\}\, dF(x)\right]^2 \\
&= \int_{-\infty}^{\infty} x^2\, dF_\alpha(x) - \left[\int_{-\infty}^{\infty} x\, dF_\alpha(x)\right]^2,
\end{aligned}$$

where we define the distribution function F_α by

$$F_\alpha(s) = \frac{1}{M_S(\alpha)} \int_{-\infty}^{s} \exp\{\alpha x\}\, dF(x),$$

for all $s \in \mathbb{R}$. Let X be a random variable with distribution function F_α. Then we get

$$\frac{d}{d\alpha}\pi_\alpha = \int_{-\infty}^{\infty} x^2\, dF_\alpha(x) - \left[\int_{-\infty}^{\infty} x\, dF_\alpha(x)\right]^2 = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = \mathrm{Var}(X) \geq 0.$$

Hence, the Esscher premium π_α is always non-decreasing. Moreover, if S is non-deterministic, then also X is non-deterministic. Thus, in this case we get

$$\frac{d}{d\alpha}\pi_\alpha = \mathrm{Var}(X) > 0.$$

In particular, the Esscher premium π_α then is strictly increasing in α.

(b) Let $\alpha \in (0, r_0)$. According to Corollary 6.16 of the lecture notes, the Esscher premium π_α is given by

$$\pi_\alpha = \frac{d}{dr} \log M_S(r)\bigg|_{r=\alpha}.$$

For small values of α, we can use a first-order Taylor approximation around 0 to get

$$\begin{aligned}
\pi_\alpha &\approx \frac{d}{dr}\log M_S(r)\Big|_{r=0} + \alpha \cdot \frac{d^2}{dr^2}\log M_S(r)\Big|_{r=0} \\
&= \frac{M_S'(0)}{M_S(0)} + \alpha\left(\frac{M_S''(0)}{M_S(0)} - \left[\frac{M_S'(0)}{M_S(0)}\right]^2\right) \\
&= \mathbb{E}[S] + \alpha\left(\mathbb{E}\left[S^2\right] - \mathbb{E}[S]^2\right) \\
&= \mathbb{E}[S] + \alpha\mathrm{Var}(S).
\end{aligned}$$

We conclude that for small values of α, the Esscher premium π_α of S is approximately equal to a premium resulting from a variance loading principle.

(c) Since $S \sim \mathrm{CompPoi}(\lambda v, G)$, we can use Proposition 2.11 of the lecture notes to get

$$\log M_S(r) = \lambda v\left[M_G(r) - 1\right],$$

where M_G denotes the moment generating function of a random variable with distribution function G. Since G is the distribution function of a gamma distribution with shape parameter $\gamma > 0$ and scale parameter $c > 0$, we have

$$M_G(r) = \left(\frac{c}{c-r}\right)^\gamma,$$

for all $r < c$. In particular, also $M_S(r)$ is defined for all $r < c$, which implies that the Esscher premium π_α exists for all $\alpha \in (0, c)$.

Now let $\alpha \in (0, c)$. Then the Esscher premium π_α can be calculated as

$$\begin{aligned}
\pi_\alpha &= \frac{d}{dr}\log M_S(r)\Big|_{r=\alpha} \\
&= \frac{d}{dr}\lambda v\left[\left(\frac{c}{c-r}\right)^\gamma - 1\right]\Big|_{r=\alpha} \\
&= \frac{d}{dr}\lambda v\left[\left(1 - \frac{r}{c}\right)^{-\gamma} - 1\right]\Big|_{r=\alpha} \\
&= \lambda v\frac{\gamma}{c}\left(1 - \frac{r}{c}\right)^{-\gamma-1}\Big|_{r=\alpha} \\
&= \lambda v\frac{\gamma}{c}\left(\frac{c}{c-\alpha}\right)^{\gamma+1}.
\end{aligned}$$

Note that since $c > c - \alpha$ and $\gamma > 0$, we have

$$\left(\frac{c}{c-\alpha}\right)^{\gamma+1} > 1,$$

and, thus,

$$\pi_\alpha = \lambda v\frac{\gamma}{c}\left(\frac{c}{c-\alpha}\right)^{\gamma+1} > \lambda v\frac{\gamma}{c} = \mathbb{E}[S].$$

Non-Life Insurance: Mathematics and Statistics

Solution sheet 10

Solution 10.1 Tariffication Methods

In this exercise we work with $K = 2$ tariff criteria. The first criterion (vehicle type) has $I = 3$ risk characteristics:

$$\chi_{1,1} \text{ (passenger car)}, \quad \chi_{1,2} \text{ (delivery van)} \quad \text{and} \quad \chi_{1,3} \text{ (truck)}.$$

The second criterion (driver age) has $J = 4$ risk characteristics:

$$\chi_{2,1} \text{ (21 - 30 years)}, \quad \chi_{2,2} \text{ (31 - 40 years)}, \quad \chi_{2,3} \text{ (41 - 50 years)} \quad \text{and} \quad \chi_{2,4} \text{ (51 - 60 years)}.$$

The claim amounts $S_{i,j}$ for the risk classes $(i,j), 1 \leq i \leq 3, 1 \leq j \leq 4$, are given on the exercise sheet. We work with a multiplicative tariff structure. In particular, we use the model

$$\mathbb{E}[S_{i,j}] = v_{i,j}\, \mu\, \chi_{1,i}\, \chi_{2,j},$$

for all $1 \leq i \leq 3, 1 \leq j \leq 4$, where we set the number of policies $v_{i,j} = 1$. Moreover, in order to get a unique solution, we set $\mu = 1$ and $\chi_{1,1} = 1$. Therefore, there remains to find the risk characteristics $\chi_{1,2}, \chi_{1,3}, \chi_{2,1}, \chi_{2,2}, \chi_{2,3}, \chi_{2,4}$.

(a) In the method of Bailey & Simon, these risk characteristics are found by minimizing

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(S_{i,j} - v_{i,j}\, \mu\, \chi_{1,i}\, \chi_{2,j})^2}{v_{i,j}\, \mu\, \chi_{1,i}\, \chi_{2,j}} = \sum_{i=1}^{3} \sum_{j=1}^{4} \frac{(S_{i,j} - \chi_{1,i}\, \chi_{2,j})^2}{\chi_{1,i}\, \chi_{2,j}}.$$

Let $i \in \{2, 3\}$. Then $\hat{\chi}_{1,i}$ is found by the solution of

$$\begin{aligned}
0 &\overset{!}{=} \frac{\partial}{\partial \chi_{1,i}} X^2 \\
&= \sum_{j=1}^{4} \frac{\partial}{\partial \chi_{1,i}} \frac{(S_{i,j} - \chi_{1,i}\, \chi_{2,j})^2}{\chi_{1,i}\, \chi_{2,j}} \\
&= \sum_{j=1}^{4} \frac{-2(S_{i,j} - \chi_{1,i}\, \chi_{2,j})\chi_{1,i}\, \chi_{2,j} - (S_{i,j} - \chi_{1,i}\, \chi_{2,j})^2}{\chi_{1,i}^2\, \chi_{2,j}} \\
&= \sum_{j=1}^{4} \frac{-2S_{i,j}\chi_{1,i}\, \chi_{2,j} + 2\chi_{1,i}^2\, \chi_{2,j}^2 - S_{i,j}^2 + 2S_{i,j}\chi_{1,i}\, \chi_{2,j} - \chi_{1,i}^2\, \chi_{2,j}^2}{\chi_{1,i}^2\, \chi_{2,j}} \\
&= \sum_{j=1}^{4} \frac{\chi_{1,i}^2\, \chi_{2,j}^2 - S_{i,j}^2}{\chi_{1,i}^2\, \chi_{2,j}} \\
&= \sum_{j=1}^{4} \chi_{2,j} - \frac{1}{\chi_{1,i}^2} \sum_{j=1}^{4} \frac{S_{i,j}^2}{\chi_{2,j}}.
\end{aligned}$$

Thus, we get

$$\hat{\chi}_{1,i} = \left(\frac{\sum_{j=1}^{4} S_{i,j}^2 / \chi_{2,j}}{\sum_{j=1}^{4} \chi_{2,j}} \right)^{1/2}.$$

By an analogous calculation, one finds

$$\hat{\chi}_{2,j} = \left(\frac{\sum_{i=1}^{3} S_{i,j}^2 / \chi_{1,i}}{\sum_{i=1}^{3} \chi_{1,i}} \right)^{1/2},$$

for $j \in \{1,2,3,4\}$. For solving these equations, one has to apply a root-finding algorithm like for example the Newton-Raphson method. We get the following multiplicative tariff structure:

	21-30y	31-40y	41-50y	51-60y	$\hat{\chi}_{1,i}$
passenger car	2'176	1'751	1'491	1'493	1
delivery van	2'079	1'674	1'425	1'427	0.96
truck	2'456	1'977	1'684	1'686	1.13
$\hat{\chi}_{2,j}$	2'176	1'751	1'491	1'493	

We see that the risk characteristics for the classes passenger car and delivery van are close to each other, whereas for trucks we have a higher tariff. Moreover, an insured with age in the class 21 - 30 years gets a considerably higher tariff than an insured with age in the class 31 - 40 years. The smallest tariff is assigned to insureds with age in the classes 41 - 50 years and 51 - 60 years. Note that we have

$$\sum_{i=1}^{3} \sum_{j=1}^{4} \hat{\chi}_{1,i} \hat{\chi}_{2,j} = 21\text{'}320 > 21\text{'}300 = \sum_{i=1}^{3} \sum_{j=1}^{4} S_{i,j},$$

which confirms the (systematic) positive bias of the method of Bailey & Simon shown in Lemma 7.2 of the lecture notes.

(b) In the method of Bailey & Jung, which is also called method of marginal totals, the risk characteristics $\chi_{1,2}, \chi_{1,3}, \chi_{2,1}, \chi_{2,2}, \chi_{2,3}, \chi_{2,4}$ are found by solving the equations

$$\sum_{j=1}^{J} v_{i,j}\, \mu\, \chi_{1,i}\, \chi_{2,j} = \sum_{j=1}^{J} S_{i,j},$$
$$\sum_{i=1}^{I} v_{i,j}\, \mu\, \chi_{1,i}\, \chi_{2,j} = \sum_{i=1}^{I} S_{i,j}.$$

Since $I = 3, J = 4$ and we work with $v_{i,j} = 1$ and set $\mu = 1$, we get the equations

$$\sum_{j=1}^{4} \chi_{1,i}\, \chi_{2,j} = \sum_{j=1}^{4} S_{i,j},$$
$$\sum_{i=1}^{3} \chi_{1,i}\, \chi_{2,j} = \sum_{i=1}^{3} S_{i,j}.$$

Thus, for $i \in \{2,3\}$ and $j \in \{1,2,3,4\}$, we get

$$\hat{\chi}_{1,i} = \sum_{j=1}^{4} S_{i,j} \Big/ \sum_{j=1}^{4} \chi_{2,j},$$
$$\hat{\chi}_{2,j} = \sum_{i=1}^{3} S_{i,j} \Big/ \sum_{i=1}^{3} \chi_{1,i}.$$

Analogously to the method of Bailey & Simon, one has to solve this system of equations using a root-finding algorithm. We get the following multiplicative tariff structure:

	21-30y	31-40y	41-50y	51-60y	$\hat{\chi}_{1,i}$
passenger car	2'170	1'749	1'490	1'490	1
delivery van	2'076	1'673	1'425	1'425	0.96
truck	2'454	1'977	1'684	1'684	1.13
$\hat{\chi}_{2,j}$	2'170	1'749	1'490	1'490	

We see that the results are very close to those in part (a) where we applied the method of Bailey & Simon. However, now we have

$$\sum_{i=1}^{3}\sum_{j=1}^{4}\hat{\chi}_{1,i}\hat{\chi}_{2,j} = 21'300 = 21'300 = \sum_{i=1}^{3}\sum_{j=1}^{4} S_{i,j},$$

which comes as no surprise as we fitted the risk characteristics such that the above equality holds true.

(c) In the log-linear regression model we work with the stochastic model

$$X_{i,j} \overset{\text{def}}{=} \log \frac{S_{i,j}}{v_{i,j}} = \log S_{i,j} \sim \mathcal{N}(\beta_0 + \beta_{1,i} + \beta_{2,j}, \sigma^2),$$

where $\beta_0, \beta_{1,i}, \beta_{2,j} \in \mathbb{R}$ and $\sigma^2 > 0$, for all risk classes $(i,j), 1 \leq i \leq 3, 1 \leq j \leq 4$. The risk characteristics of the two tariff criteria vehicle type and driver age are now given by

$$\beta_{1,1} \text{ (passenger car)}, \quad \beta_{1,2} \text{ (delivery van)} \quad \text{and} \quad \beta_{1,3} \text{ (truck)},$$

and

$$\beta_{2,1} \text{ (21 - 30 years)}, \quad \beta_{2,2} \text{ (31 - 40 years)}, \quad \beta_{2,3} \text{ (41 - 50 years)} \quad \text{and} \quad \beta_{2,4} \text{ (51 - 60 years)}.$$

In order to get a unique solution, we set $\beta_{1,1} = \beta_{2,1} = 0$. Because this will simplify notation considerably, we write $\mathbf{X} = (X_1, \ldots, X_M)'$ with $M = 12$ and

$$X_1 = X_{1,1}, \quad X_2 = X_{1,2}, \quad X_3 = X_{1,3}, \quad X_4 = X_{1,4}, \quad X_5 = X_{2,1}, \quad X_6 = X_{2,2},$$
$$X_7 = X_{2,3}, \quad X_8 = X_{2,4}, \quad X_9 = X_{3,1}, \quad X_{10} = X_{3,2}, \quad X_{11} = X_{3,3}, \quad X_{12} = X_{3,4}.$$

Moreover, we define

$$\boldsymbol{\beta} = (\beta_0, \beta_{1,2}, \beta_{1,3}, \beta_{2,2}, \beta_{2,3}, \beta_{2,4})' \in \mathbb{R}^{r+1},$$

where $r = 5$. Then, we assume that $\mathbf{X}$ has a multivariate Gaussian distribution

$$\mathbf{X} \sim \mathcal{N}(Z\boldsymbol{\beta}, \sigma^2 I),$$

where $I \in \mathbb{R}^{M\times M}$ denotes the identity matrix and $Z \in \mathbb{R}^{M\times(r+1)}$ is the so-called design matrix that satisfies

$$\mathbb{E}[\mathbf{X}] = Z\boldsymbol{\beta}.$$

For example for $m = 1$ we have

$$\mathbb{E}[X_m] = \mathbb{E}[X_1] = \mathbb{E}[X_{1,1}] = \beta_0 + \beta_{1,1} + \beta_{2,1} = \beta_0 = (1,0,0,0,0,0)\,\boldsymbol{\beta},$$

and for $m = 8$

$$\mathbb{E}[X_m] = \mathbb{E}[X_8] = \mathbb{E}[X_{2,4}] = \beta_0 + \beta_{1,2} + \beta_{2,4} = (1,1,0,0,0,1)\,\boldsymbol{\beta}.$$

Doing this for all $m \in \{1, \ldots, 12\}$, we find the design matrix Z:

intercept (β_0)	van ($\beta_{1,2}$)	truck ($\beta_{1,3}$)	31-40y ($\beta_{2,2}$)	41-50y ($\beta_{2,3}$)	51-60y ($\beta_{2,4}$)
1	0	0	0	0	0
1	0	0	1	0	0
1	0	0	0	1	0
1	0	0	0	0	1
1	1	0	0	0	0
1	1	0	1	0	0
1	1	0	0	1	0
1	1	0	0	0	1
1	0	1	0	0	0
1	0	1	1	0	0
1	0	1	0	1	0
1	0	1	0	0	1

Here we would like to point out that we can also use R to find the design matrix, see the R-Code of Exercise 10.2. According to formula (7.9) of the lecture notes, the MLE $\hat{\boldsymbol{\beta}}^{\mathrm{MLE}}$ of the parameter vector $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}^{\mathrm{MLE}} = [Z'(\sigma^2 I)^{-1} Z]^{-1} Z'(\sigma^2 I)^{-1}\mathbf{X} = (Z'Z)^{-1}Z'\mathbf{X}.$$

Note that $\hat{\boldsymbol{\beta}}^{\mathrm{MLE}}$ does not depend on σ^2. Moreover, the design matrix Z has full column rank and, thus, $Z'Z$ is indeed invertible. See the R-Code given at the end of the solution to this exercise for the calculation of $\hat{\boldsymbol{\beta}}^{\mathrm{MLE}}$. We get the following tariff structure:

$\hat{\beta}_0 = 7.688$	21-30y	31-40y	41-50y	51-60y	$\hat{\beta}_{1,i}$
passenger car	2'182	1'758	1'500	1'501	0
delivery van	2'063	1'663	1'417	1'419	-0.056
truck	2'444	1'970	1'680	1'682	0.113
$\hat{\beta}_{2,j}$	0	-0.216	-0.375	-0.374	

We see that the results are very close to those in parts (a) and (b) where we applied the method of Bailey & Simon and the method of Bailey & Jung. However, since we are now working in a stochastic framework, we also get standard errors and we can make statements about the statistical significance of the parameters. According to the R-output, we get the following p-values for the individual parameters:

	$\hat{\beta}_0$	$\hat{\beta}_{1,2}$	$\hat{\beta}_{1,3}$	$\hat{\beta}_{2,2}$	$\hat{\beta}_{2,3}$	$\hat{\beta}_{2,4}$
p-value	≈ 0	0.232	0.036	-0.005	0.0003	0.0003

R gets these p-values by applying a t-test individually to each parameter, whether they are equal to zero. While the p-values for $\hat{\beta}_0, \hat{\beta}_{1,3}, \hat{\beta}_{2,2}, \hat{\beta}_{2,3}, \hat{\beta}_{2,4}$ are smaller than 0.05 and, thus, these parameters are significantly different from zero, the p-value of $\hat{\beta}_{1,2}$ (delivery van) is fairly high. Hence, we might question if we really need the class delivery van.

In order to check whether there is statistical evidence that the classification into different types of vehicles could be omitted, we define the null hypothesis of the reduced model:

$$H_0 : \beta_{1,2} = \beta_{1,3} = 0,$$

i.e. we set $p = 2$ parameters equal to 0. Then we can perform the same analysis as above to get the MLE $\hat{\boldsymbol{\beta}}_{H_0}^{\mathrm{MLE}}$. In particular, let Z_{H_0} be the design matrix Z without the second column van ($\beta_{1,2}$) and the third column truck ($\beta_{1,3}$). Then $\hat{\boldsymbol{\beta}}_{H_0}^{\mathrm{MLE}}$ is given by

$$\hat{\boldsymbol{\beta}}_{H_0}^{\mathrm{MLE}} = (Z'_{H_0} Z_{H_0})^{-1} Z'_{H_0}\mathbf{X}.$$

See the R-Code given below for the calculation of $\hat{\boldsymbol{\beta}}_{H_0}^{\text{MLE}}$. Now, for all $m \in \{1, \ldots, 12\}$, we define the fitted value $\hat{X}_m^{\text{full}}$ of the full model and the fitted value $\hat{X}_m^{H_0}$ of the reduced model. In particular, we have

$$\hat{X}_m^{\text{full}} = \left[Z \hat{\boldsymbol{\beta}}^{\text{MLE}} \right]_m$$

and

$$\hat{X}_m^{H_0} = \left[Z_{H_0} \hat{\boldsymbol{\beta}}_{H_0}^{\text{MLE}} \right]_m,$$

where $[\cdot]_m$ denotes the m-th element of the corresponding vector, for all $m \in \{1, \ldots, 12\}$. Moreover, we define

$$SS_{\text{err}}^{\text{full}} = \sum_{m=1}^{M} \left(X_m - \hat{X}_m^{\text{full}} \right)^2$$

and

$$SS_{\text{err}}^{H_0} = \sum_{m=1}^{M} \left(X_m - \hat{X}_m^{H_0} \right)^2.$$

According to formula (7.15) of the lecture notes, the test statistic

$$T = \frac{SS_{\text{err}}^{H_0} - SS_{\text{err}}^{\text{full}}}{SS_{\text{err}}^{\text{full}}} \frac{M - r - 1}{p} = 3 \frac{SS_{\text{err}}^{H_0} - SS_{\text{err}}^{\text{full}}}{SS_{\text{err}}^{\text{full}}}$$

has an F-distribution with degrees of freedeom given by $\text{df}_1 = p = 2$ and $\text{df}_2 = M - r - 1 = 6$. See the R-Code below for the calculation of T. We get

$$T \approx 8.336,$$

which corresponds to a p-value of approximately 1.85%. Thus, we can reject H_0 at significance level of 5%, i.e. there is no statistical evidence that the classification into different types of vehicles could be omitted.

(d) As we already mentioned above, the method of Bailey & Simon, the method of Bailey & Jung and the MLE method in the log-linear regression model all lead to approximately the same results. The only differences are, that with the method of Bailey & Jung we get coinciding marginal totals and with the log-linear regression model we are in a stochastic framework which allows for calculating parameter uncertainties and hypothesis testing.

```
### c)

### We apply the log-linear regression method to the observed claim amounts given on the exercise sheet

### Load the observed claim amounts into a matrix
S <- matrix(c(2000,2200,2500,1800,1600,2000,1500,1400,1700,1600,1400,1600), nrow = 3)

### Define the design matrix Z
Z <- matrix(c(rep(1,12),rep(0,4),rep(1,4),rep(0,12),rep(1,4),rep(c(0,1,0,0),3),rep(c(0,0,1,0),3),rep(c(0,0,0,1),3)),nrow = 12)

### Store the design matrix Z (without the intercept term) and the dependent variable log(S_{i,j}) in one dataset
```

```
data <- cbind(Z[,-1],matrix(log(t(S)),nrow = 12))
data <- as.data.frame(data)
colnames(data) <- c("van", "truck", "X31_40y", "X41_50y", "X51_
    60y", "observation")

### Apply the regression model
linear.model1 <- lm(formula = observation ~ van + truck + X31_
    40y + X41_50y + X51_60y,data=data)

### Print the output of the regression model
summary(linear.model1)

### Fitted values
fitted(linear.model1)

### We can also get the parameters by applying the formula
    (7.9) of the lecture notes
solve(t(Z)%*%Z)%*%t(Z)%*%matrix(log(t(S)),nrow = 12)


### Apply the regression model under H_{0}
linear.model2 <- lm(formula = observation ~ X31_40y + X41_50y +
     X51_60y,data=data)

### Calculation of the test statistic T which has an F-
    distribution
T <- 3 * (sum((fitted(linear.model2) - data[,6])^2) - sum((
    fitted(linear.model1) - data[,6])^2)) / sum((fitted(linear.
    model1) - data[,6])^2)

### Calculation of the corresponding p-value
pf(T, 2, 6, lower.tail = FALSE)
```

Note that we could also define the covariates of factor type in R which then automatically implies that these covariates are of categorical type and R chooses the design matrix Z accordingly, see the R-Code for the solution of Exercise 10.2 given below.

Solution 10.2 Tariffication Methods

(a) In this exercise we work with $K = 3$ tariff criteria. The first criterion (vehicle class) has 2 risk characteristics:

$$\beta_{1,1} \text{ (weight over 60 kg and more than two gears)} \quad \text{and} \quad \beta_{1,2} \text{ (other)}.$$

The second criterion (vehicle age) also has 2 risk characteristics:

$$\beta_{2,1} \text{ (at most one year)} \quad \text{and} \quad \beta_{2,2} \text{ (more than one year)}.$$

The third criterion (geographic zone) has 3 risk characteristics:

$$\beta_{3,1} \text{ (large cities)}, \quad \beta_{3,2} \text{ (middle-sized towns)} \quad \text{and} \quad \beta_{3,3} \text{ (smaller towns and countryside)}.$$

The observed number of claims N_{l_1,l_2,l_3}, the observed volumes v_{l_1,l_2,l_3} and the observed claim frequencies

$$\lambda_{l_1,l_2,l_3} = \frac{N_{l_1,l_2,l_3}}{v_{l_1,l_2,l_3}}$$

for the risk classes $(l_1, l_2, l_3), 1 \leq l_1 \leq 2, 1 \leq l_2 \leq 2, 1 \leq l_3 \leq 3$, are given on the exercise sheet. Now, for modelling purposes, we assume that all N_{l_1,l_2,l_3} are independent with

$$N_{l_1,l_2,l_3} \sim \text{Poi}(\lambda_{l_1,l_2,l_3} v_{l_1,l_2,l_3})$$

and define

$$X_{l_1,l_2,l_3} = \frac{N_{l_1,l_2,l_3}}{v_{l_1,l_2,l_3}}.$$

Then we use the model Ansatz

$$g(\lambda_{l_1,l_2,l_3}) = g\left(\mathbb{E}\left[\frac{N_{l_1,l_2,l_3}}{v_{l_1,l_2,l_3}}\right]\right) = g\left(\mathbb{E}\left[X_{l_1,l_2,l_3}\right]\right) = \beta_0 + \beta_{1,l_1} + \beta_{2,l_2} + \beta_{3,l_3},$$

where $\beta_0 \in \mathbb{R}$ and where we use the log-link function, i.e. $g(\cdot) = \log(\cdot)$. In order to get a unique solution, we set $\beta_{1,1} = \beta_{2,1} = \beta_{3,1} = 0$. Moreover, we define

$$\boldsymbol{\beta} = (\beta_0, \beta_{1,2}, \beta_{2,2}, \beta_{3,2}, \beta_{3,3})' \in \mathbb{R}^{r+1},$$

where $r = 4$. Similarly as in Exercise 10.1, (c), we will relabel the risk classes with the index $m \in \{1, \dots, M\}$, where $M = 2 \cdot 2 \cdot 3 = 12$, define $\mathbf{X} = (X_1, \dots, X_M)'$ and the design matrix $Z \in \mathbb{R}^{M\times(r+1)}$ that satisfies

$$\log \mathbb{E}[\mathbf{X}] = Z\boldsymbol{\beta},$$

where the logarithm is applied componentwise to $\mathbb{E}[\mathbf{X}]$. Let $m \in \{1, \dots, 12\}$. According to Example 7.10 of the lecture notes, $X_m = N_m/v_m$ belongs to the exponential dispersion family with cumulant function $b(\cdot) = \exp\{\cdot\}$, $\theta_m = \log \lambda_m$, $w_m = v_m$ and dispersion parameter $\phi = 1$, i.e. we have

$$[Z\boldsymbol{\beta}]_m = \log \mathbb{E}[X_m] = \log \mathbb{E}\left[\frac{N_m}{v_m}\right] = \log \lambda_m = \theta_m,$$

where $[Z\boldsymbol{\beta}]_m$ denotes as above the m-th element of the vector $Z\boldsymbol{\beta}$. Thus, we assume that $X_1, \dots, X_M$ are independent with

$$X_m \sim \text{EDF}(\theta_m = [Z\boldsymbol{\beta}]_m, \phi = 1, v_m, b(\cdot) = \exp\{\cdot\}),$$

for all $m \in \{1, \dots, M\}$. According to Proposition 7.11 of the lecture notes, the MLE $\hat{\boldsymbol{\beta}}^{\text{MLE}}$ of $\boldsymbol{\beta}$ is the solution of

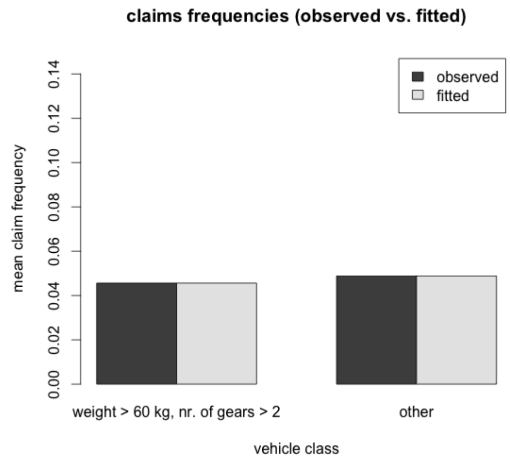
$$Z'V \exp\{Z\boldsymbol{\beta}\} = Z'V\mathbf{X}, \tag{1}$$

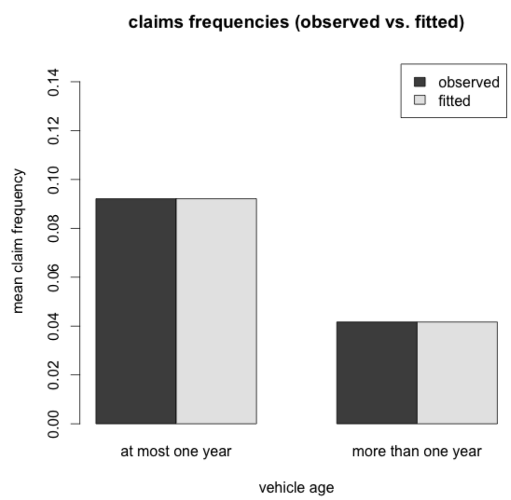
where the weight matrix V is given by $V = \text{diag}(v_1, \dots, v_M)$. This equation has to be solved numerically. See the R-Code at the end of the solution to this exercise for the calculation of $\hat{\boldsymbol{\beta}}^{\text{MLE}}$. We get the following estimates:

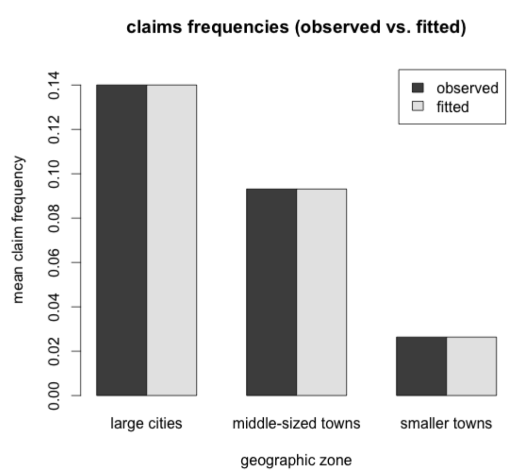
	$\hat{\beta}_0$	$\hat{\beta}_{1,2}$	$\hat{\beta}_{2,2}$	$\hat{\beta}_{3,2}$	$\hat{\beta}_{3,3}$
MLE	-1.435	-0.237	-0.502	-0.404	-1.657

We observe that insureds with a vehicle with weight over 60 kg and more than two gears tend to cause more claims than insureds with other vehicles. Analogoulsy, if the vehicle is at most one year old, we expect more claims than if it was older. Regarding the geographic zone, we see that driving in middle-sized towns leads to fewer claims than driving in large cities. Moreover, driving in smaller towns and countryside leads to even fewer claims than driving in middle-sized towns, where this difference is greater than the difference between large cities and middle-sized towns.

(b) The observed and the fitted claim frequencies against the vehicle class, the vehicle age and the geographical zone look as follows:

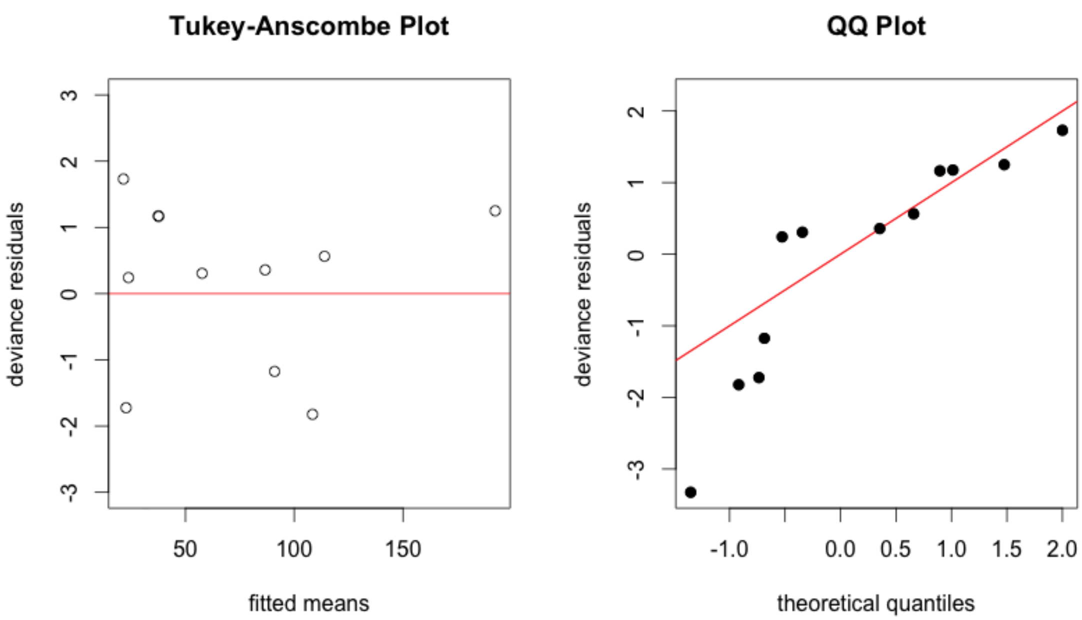






See the R-Code at the end of the solution to this exercise for creating the plots given above. Note that the observed and the fitted marginal claim frequencies are always the same. This is a direct consequence of equation (1) given above which ensures that the observed and the fitted total marginal sums are the same if we use the same volumes again. This is also the reason why in the marginal plot for the vehicle class we don't see that insureds with a vehicle with weight over 60 kg and more than two gears tend to cause more claims than insureds with other vehicles as expected after the discussion at the end of part (a). More precisely, for the vehicles with weight over 60 kg and more than two gears we have a smaller volume for the riskier classes with respect to the other tariff criteria vehicle age and geographic zone than for the other vehicles. This compensates for the fact that vehicles with weight over 60 kg and more than two gears tend to cause more claims than other vehicles, as seen at the end of part (a). For the other variables vehicle age and geographic zone we again see the same results as in part (a).

(c) The Tukey-Anscombe plot and the QQ plot look as follows:



See the R-Code at the end of the solution to this exercise for creating the plots given above. They are both not ideal, but considering that we only have 12 risk classes, we accept them.

(d) We will perform two tests in order to check if there is statistical evidence that the classification into the geographic zones could be omitted. Note that in part (a) we saw that we tend to have considerably fewer claims for drivers in smaller towns and countryside than for drivers in middle-sized towns. The same holds true in a weakened form for middle-sized towns and large cities. Thus, we would expect that the classification into the three different geographic zone is reasonable. Now we will investigate this. To start with, note that the logarithmic probability that a Poisson random variable with frequency parameter α attains the value k, for some $k \in \mathbb{N}$, is equal to

$$\log\left(\exp\{-\alpha\}\frac{\alpha^k}{k!}\right) = -\alpha + k\log\alpha - \log k!.$$

Thus, defining

$$\hat{\boldsymbol{\lambda}}^{\mathrm{MLE}} = \exp\left\{Z\hat{\boldsymbol{\beta}}^{\mathrm{MLE}}\right\},$$

with $\hat{\boldsymbol{\lambda}}^{\mathrm{MLE}} = \left(\hat{\lambda}_1^{\mathrm{MLE}}, \dots, \hat{\lambda}_M^{\mathrm{MLE}}\right)$, the joint log-likelihood function $l_{\mathbf{X}}$ of $\mathbf{X}$ at $\hat{\boldsymbol{\lambda}}^{\mathrm{MLE}}$ is given by

$$l_{\mathbf{X}}\left(\hat{\boldsymbol{\lambda}}^{\mathrm{MLE}}\right) = \sum_{m=1}^{M} -\hat{\lambda}_m^{\mathrm{MLE}} v_m + X_m v_m \log\left(\hat{\lambda}_m^{\mathrm{MLE}} v_m\right) - \log\left[(X_m v_m)!\right].$$

Therefore, we get for the scaled deviance statistics $D^*\left(\mathbf{X}, \hat{\boldsymbol{\lambda}}^{\mathrm{MLE}}\right)$:

$$\begin{aligned}
D^*\left(\mathbf{X}, \hat{\boldsymbol{\lambda}}^{\mathrm{MLE}}\right) &= 2\left[l_{\mathbf{X}}(\mathbf{X}) - l_{\mathbf{X}}\left(\hat{\boldsymbol{\lambda}}^{\mathrm{MLE}}\right)\right] \\
&= 2\sum_{m=1}^{M} -X_m v_m + X_m v_m \log X_m + \hat{\lambda}_m^{\mathrm{MLE}} v_m - X_m v_m \log \hat{\lambda}_m^{\mathrm{MLE}} \\
&= 2\sum_{m=1}^{M} v_m\left(X_m \log X_m - X_m - X_m \log \hat{\lambda}_m^{\mathrm{MLE}} + \hat{\lambda}_m^{\mathrm{MLE}}\right).
\end{aligned}$$

Moreover, since for the Poisson case we have $\phi = 1$, the scaled deviance statistics $D^*\left(\mathbf{X}, \hat{\boldsymbol{\lambda}}^{\mathrm{MLE}}\right)$ and the deviance statistics $D\left(\mathbf{X}, \hat{\boldsymbol{\lambda}}^{\mathrm{MLE}}\right)$ are the same. Now, in order to check whether there is statistical evidence that the classification into the geographic zones could be omitted, we define the null hypothesis

$$H_0 : \beta_{3,2} = \beta_{3,3} = 0.$$

Thus, in the reduced model, we set the above $p = 2$ variables equal to 0. Then we can recalculate $\hat{\boldsymbol{\beta}}_{H_0}^{\mathrm{MLE}}$ for this reduced model and define

$$\hat{\boldsymbol{\lambda}}_{H_0}^{\mathrm{MLE}} = \exp\left\{Z_{H_0}\hat{\boldsymbol{\beta}}_{H_0}^{\mathrm{MLE}}\right\},$$

where Z_{H_0} is the design matrix in the reduced model. According to formula (7.22) of the lecture notes, the test statistic

$$\begin{aligned}
\mathrm{F} &= \frac{D\left(\mathbf{X}, \hat{\boldsymbol{\lambda}}_{H_0}^{\mathrm{MLE}}\right) - D\left(\mathbf{X}, \hat{\boldsymbol{\lambda}}^{\mathrm{MLE}}\right)}{D\left(\mathbf{X}, \hat{\boldsymbol{\lambda}}^{\mathrm{MLE}}\right)} \frac{M - r - 1}{p} \\
&= \frac{7}{2} \frac{D\left(\mathbf{X}, \hat{\boldsymbol{\lambda}}_{H_0}^{\mathrm{MLE}}\right) - D\left(\mathbf{X}, \hat{\boldsymbol{\lambda}}^{\mathrm{MLE}}\right)}{D\left(\mathbf{X}, \hat{\boldsymbol{\lambda}}^{\mathrm{MLE}}\right)}
\end{aligned}$$

has approximately an F-distribution with degrees of freedom given by $\mathrm{df}_1 = p = 2$ and $\mathrm{df}_2 = M - r - 1 = 7$. See the R-Code below for the calculation of F. We get

$$\mathrm{F} \approx 51.239,$$

which corresponds to a p-value of approximately 0.0066%. Thus, we can reject H_0 at significance level of 5%. According to formula (7.23) of the lecture notes, a second test statistic is given by

$$X^2 = D^*\left(\mathbf{X}, \hat{\boldsymbol{\lambda}}_{H_0}^{\mathrm{MLE}}\right) - D^*\left(\mathbf{X}, \hat{\boldsymbol{\lambda}}^{\mathrm{MLE}}\right).$$

The test statistic X^2 has approximately a χ^2-distribution with $\mathrm{df} = p = 2$ degrees of freedom. See the R-Code below for the calculation of X^2. We get

$$X^2 \approx 389.882,$$

which corresponds to a p-value of approximately $2.179 \cdot 10^{-85}$, which is basically 0. Thus, we can reject H_0 at significance level of 5%. Since we can reject H_0 using two different test statistics, we can conclude that there is no statistical evidence that the classification into different types of vehicles could be omitted.

```
### a)

### We perform a GLM analysis for the claim frequencies

### Determine the design matrix Z
class <- factor(c(rep(1,6),rep(2,6)))
age <- factor(c(rep(1,3),rep(2,3),rep(1,3),rep(2,3)))
zone <- factor(c(rep(1:3,4)))
counts <- c(25,15,15,60,90,210,45,45,30,80,120,90)
volumes <- c(1,2,5,4,9,70,2,3,6,8,15,50) * 100
Z <- model.matrix(counts ~ class + age + zone)

### Store the design matrix Z (without the intercept term), the counts and the volumes in one dataset
data <- cbind(Z[,-1],counts,volumes)
data <- as.data.frame(data)

### Apply GLM
d.glm <- glm(counts ~ class2 + age2 + zone2 + zone3, data=data, offset = log(volumes), family = poisson())
d.glm



### b)

### Fitted number of claims
fitted(d.glm)

### Store the features, the observed number of claims and the fitted numer of claims in one data set
data2 <- cbind(class, age, zone, volumes, counts, fitted(d.glm))
data2 <- as.data.frame(data2)
colnames(data2)[5:6] <- c("observed","fitted")

### Marginal claim frequencies for the two class categories
library(plyr)
class.comp <- ddply(data2, .(class), summarise, volumes = sum(volumes), observed = sum(observed), fitted = sum(fitted))
barplot(t(as.matrix(class.comp[,3:4]/class.comp[,2])), beside = TRUE, names.arg = c("weight > 60 kg, nr. of gears > 2", "other"), main = "claims frequencies (observed vs. fitted)", ylim = c(0,0.15), xlab = "vehicle class", ylab = "mean claim frequency",legend.text = TRUE)

### Marginal claim frequencies for the two age categories
```

```
age.comp <- ddply(data2, .(age), summarise, volumes = sum(
    volumes), observed = sum(observed), fitted = sum(fitted))
barplot(t(as.matrix(age.comp[,3:4]/age.comp[,2])), beside =
    TRUE, names.arg = c("at most one year", "more than one year"
    ), main = "claims frequencies (observed vs. fitted)",ylim =
    c(0,0.15), xlab = "vehicle age", ylab = "mean claim
    frequency",legend.text = TRUE)

### Marginal claim frequencies for the three zone categories
zone.comp <- ddply(data2, .(zone), summarise, volumes = sum(
    volumes), observed = sum(observed), fitted = sum(fitted))
barplot(t(as.matrix(zone.comp[,3:4]/zone.comp[,2])), beside =
    TRUE, names.arg = c("large cities", "middle-sized towns", "
    smaller towns"), main = "claims frequencies (observed vs.
    fitted)",ylim = c(0,0.15), xlab = "geographic zone", ylab =
    "mean claim frequency",legend.text = TRUE)



### c)

par(mfrow = c(1, 2))

### Calculate the deviance residuals
dev.red <- sign(data2$observed - data2$fitted) * sqrt(2 * data2
    $observed*(-log(data2$fitted / data2$observed) + data2$
    fitted / data2$observed - 1))

### Tukey-Anscombe plot
plot(data2$fitted, dev.red, main = "Tukey-Anscombe Plot", xlab
    = "fitted means", ylab = "deviance residuals", ylim = c
    (-3,3))
abline(h = 0,col = "red")

### QQ plot
library(mgcv)
qq.gam(d.glm, type = "deviance",rep = 1, pch=19, main = "QQ
    Plot")



### d)

### Calculate the deviance statistics of the full model
X <- data2$observed / data2$volumes
lambda.full <- data2$fitted / data2$volumes
D.full <- 2 * sum(data2$volumes * (X * log(X) - X - X * log(
    lambda.full) + lambda.full))

### Fit the reduced model
d.glm.2 <- glm(counts ~ class2 + age2, data=data, offset = log(
    volumes), family = poisson())
```

```
d.glm.2

### Calculate the deviance statistics of the reduced model
lambda.reduced <- fitted(d.glm.2) / data2$volumes
D.reduced <- 2 * sum(data2$volumes * (X * log(X) - X - X * log(
    lambda.reduced) + lambda.reduced))

### Calculate the test statistic F
F <- 7 / 2 * (D.reduced - D.full) / D.full

### Calculation of the corresponding p-value
pf(F, 2, 7, lower.tail = FALSE)

### Calculate the test statistic X^2
X.2 <- D.reduced - D.full

### Calculation of the corresponding p-value
pchisq(X.2, 2, lower.tail = FALSE)
```

Solution 10.3 Tweedie's Compound Poisson Model

(a) We can write S as

$$S = \sum_{i=1}^{N} Y_i,$$

where $N \sim \text{Poi}(\lambda v)$, $Y_1, Y_2, \ldots \overset{\text{i.i.d.}}{\sim} G$ and N and $(Y_1, Y_2, \ldots)$ are independent. Since G is the distribution function of a gamma distribution, we have $G(0) = 0$ and, thus,

$$\mathbb{P}[S = 0] = \mathbb{P}[N = 0] = \exp\{-\lambda v\}.$$

Let $x \in (0, \infty)$. Then the density f_S of S at x can be calculated as

$$f_S(x) = \frac{d}{dx}\,\mathbb{P}[S \le x],$$

where we have

$$\begin{aligned}
\mathbb{P}[S \le x] &= \sum_{n=0}^{\infty} \mathbb{P}[S \le x, N = n] \\
&= \sum_{n=0}^{\infty} \mathbb{P}[S \le x \,|\, N = n]\,\mathbb{P}[N = n] \\
&= \mathbb{P}[S \le x \,|\, N = 0]\,\mathbb{P}[N = 0] + \sum_{n=1}^{\infty} \mathbb{P}[S \le x \,|\, N = n]\,\mathbb{P}[N = n] \\
&= \mathbb{P}[N = 0] + \sum_{n=1}^{\infty} \mathbb{P}\left[\sum_{i=1}^{n} Y_i \le x\right] \mathbb{P}[N = n].
\end{aligned}$$

Since $Y_1, Y_2, \ldots \overset{\text{i.i.d.}}{\sim} \Gamma(\gamma, c)$, we get

$$\sum_{i=1}^{n} Y_i \sim \Gamma(n\gamma, c).$$

By writing f_n for the density function of $\Gamma(n\gamma, c)$, for all $n \in \mathbb{N}$, we get

$$\begin{aligned}
f_S(x) &= \frac{d}{dx}\left(\mathbb{P}[N=0] + \sum_{n=1}^{\infty} \mathbb{P}\left[\sum_{i=1}^{n} Y_i \le x\right] \mathbb{P}[N=n]\right) \\
&= \sum_{n=1}^{\infty} \frac{d}{dx}\mathbb{P}\left[\sum_{i=1}^{n} Y_i \le x\right] \mathbb{P}[N=n] \\
&= \sum_{n=1}^{\infty} f_n(x)\, \mathbb{P}[N=n] \\
&= \sum_{n=1}^{\infty} \frac{c^{n\gamma}}{\Gamma(n\gamma)} x^{n\gamma-1} \exp\{-cx\} \exp\{-\lambda v\} \frac{(\lambda v)^n}{n!} \\
&= \exp\{-(cx+\lambda v)\} \sum_{n=1}^{\infty} (\lambda v c^{\gamma})^n \frac{1}{\Gamma(n\gamma) n!} x^{n\gamma-1} \\
&= \exp\left\{-(cx+\lambda v) + \log\left[\sum_{n=1}^{\infty} (\lambda v c^{\gamma})^n \frac{1}{\Gamma(n\gamma) n!} x^{n\gamma-1}\right]\right\},
\end{aligned}$$

for all $x \in (0, \infty)$. Note that one can show that interchanging summation and differentiation above is indeed allowed. However, the proof is omitted here.

(b) Let $X \sim f_X$ belong to the exponential dispersion family with $w, \phi, \theta, b(\cdot)$ and $c(\cdot,\cdot,\cdot)$ as given on the exercise sheet. Then we have

$$\frac{x\theta}{\phi/w} = -xv \frac{(\gamma+1)\left(\frac{\lambda v \gamma}{c}\right)^{-\frac{1}{\gamma+1}}}{\frac{\gamma+1}{\lambda\gamma}\left(\frac{\lambda v\gamma}{c}\right)^{\frac{\gamma}{\gamma+1}}} = -x\lambda v\gamma \left(\frac{\lambda v\gamma}{c}\right)^{-1} = -cx,$$

for all $x \ge 0$, and

$$\frac{b(\theta)}{\phi/w} = v\frac{\frac{\gamma+1}{\gamma}\left(\frac{-\theta}{\gamma+1}\right)^{-\gamma}}{\frac{\gamma+1}{\lambda\gamma}\left(\frac{\lambda v\gamma}{c}\right)^{\frac{\gamma}{\gamma+1}}} = \lambda v \frac{\left(\frac{\lambda v\gamma}{c}\right)^{\frac{\gamma}{\gamma+1}}}{\left(\frac{\lambda v\gamma}{c}\right)^{\frac{\gamma}{\gamma+1}}} = \lambda v.$$

Moreover, since

$$\begin{aligned}
\frac{(\gamma+1)^{\gamma+1}}{\gamma}\left(\frac{\phi}{w}\right)^{-\gamma-1} &= \frac{(\gamma+1)^{\gamma+1}}{\gamma}\left[\frac{\gamma+1}{\lambda v\gamma}\left(\frac{\lambda v\gamma}{c}\right)^{\frac{\gamma}{\gamma+1}}\right]^{-\gamma-1} \\
&= \frac{1}{\gamma}(\lambda v\gamma)^{\gamma+1}\left(\frac{\lambda v\gamma}{c}\right)^{-\gamma} \\
&= \frac{1}{\gamma}\lambda v\gamma c^{\gamma} \\
&= \lambda v c^{\gamma},
\end{aligned}$$

we have

$$c(x,\phi,w) = \log\left(\sum_{n=1}^{\infty}\left[\frac{(\gamma+1)^{\gamma+1}}{\gamma}\left(\frac{\phi}{w}\right)^{-\gamma-1}\right]^n \frac{1}{\Gamma(n\gamma)n!}\, x^{n\gamma-1}\right)$$
$$= \log\left[\sum_{n=1}^{\infty}(\lambda v c^{\gamma})^n \frac{1}{\Gamma(n\gamma)n!}\, x^{n\gamma-1}\right],$$

for all $x > 0$. By putting together the above terms, we get

$$f_X(x;\theta,\phi) = \exp\left\{\frac{x\theta - b(\theta)}{\phi/w} + c(x,\phi,w)\right\}$$
$$= \exp\left\{-(cx+\lambda v) + \log\left[\sum_{n=1}^{\infty}(\lambda v c^{\gamma})^n \frac{1}{\Gamma(n\gamma)n!} x^{n\gamma-1}\right]\right\}$$
$$= f_S(x),$$

for all $x > 0$, and

$$f_X(0;\theta,\phi) = \exp\left\{\frac{0\cdot\theta - b(\theta)}{\phi/w} + c(0,\phi,w)\right\} = \exp\{-\lambda v\} = \mathbb{P}[S=0].$$

We conclude that S indeed belongs to the exponential dispersion family. Note that with this result at hand one might be tempted to estimate the shape parameter γ of the claim size distribution and then to do a GLM analysis directly on the compound claim size S. However, there are two reasons to rather perform a separate GLM analysis of the claim frequency and the claim severity instead: First, claim frequency modelling is usually more stable than claim severity modelling and often much of the differences between tariff cells are due to the claim frequency. Second, a separate analysis of the claim frequency and the claim severity allows more insight into the differences between the tariffs.

Non-Life Insurance: Mathematics and Statistics

Solution sheet 11

Solution 11.1 (Inhomogeneous) Credibility Estimators for Claim Counts

We define

$$X_{i,1} = \frac{N_{i,1}}{v_{i,1}},$$

for all $i \in \{1, \ldots, 5\}$. Then we have

$$\mathbb{E}[X_{i,1} \,|\, \Theta_i] = \frac{1}{v_{i,1}}\, \mathbb{E}[N_{i,1} \,|\, \Theta_i] = \frac{1}{v_{i,1}}\, \mu(\Theta_i)\, v_{i,1} = \mu(\Theta_i)$$

and

$$\mathrm{Var}(X_{i,1} \,|\, \Theta_i) = \frac{1}{v_{i,1}^2}\, \mathrm{Var}(N_{i,1} \,|\, \Theta_i) = \frac{1}{v_{i,1}^2}\, \mu(\Theta_i)\, v_{i,1} = \frac{\mu(\Theta_i)}{v_{i,1}} = \frac{\sigma^2(\Theta_i)}{v_{i,1}},$$

with $\sigma^2(\Theta_i) = \mu(\Theta_i) = \Theta_i \lambda_0$, for all $i \in \{1, \ldots, 5\}$. Moreover, since

$$\mathbb{E}[\mu(\Theta_i)^2] = \mathrm{Var}(\mu(\Theta_i)) + \mathbb{E}[\mu(\Theta_i)]^2 = \tau^2 + \lambda_0^2 < \infty$$

and

$$\mathbb{E}[X_{i,1}^2 \,|\, \Theta_i] = \mathrm{Var}(X_{i,1} \,|\, \Theta_i) + \mathbb{E}[X_{i,1} \,|\, \Theta_i]^2 = \frac{\mu(\Theta_i)}{v_{i,1}} + \mu(\Theta_i)^2,$$

we get

$$\mathbb{E}[X_{i,1}^2] = \mathbb{E}\left[\mathbb{E}[X_{i,1}^2 \,|\, \Theta_i]\right] = \mathbb{E}\left[\frac{\mu(\Theta_i)}{v_{i,1}} + \mu(\Theta_i)^2\right] = \frac{\lambda_0}{v_{i,1}} + \tau^2 + \lambda_0^2 < \infty,$$

for all $i \in \{1, \ldots, 5\}$. In particular, the Model Assumptions 8.13 of the lecture notes for the Bühlmann-Straub model are satisfied. The (expected) volatility σ^2 within the regions defined in formula (8.5) of the lecture notes is given by

$$\sigma^2 = \mathbb{E}[\sigma^2(\Theta_i)] = \mathbb{E}[\mu(\Theta_i)] = \lambda_0 = 0.088.$$

(a) Let $i \in \{1, \ldots, 5\}$. Then, according to Theorem 8.17 of the lecture notes, the inhomogeneous credibility estimator $\widehat{\widehat{\mu(\Theta_i)}}$ is given by

$$\widehat{\widehat{\mu(\Theta_i)}} = \alpha_{i,T}\, \widehat{X}_{i,1:T} + (1 - \alpha_{i,T})\, \mu_0,$$

with credibility weight $\alpha_{i,T}$ and observation based estimator $\widehat{X}_{i,1:T}$

$$\alpha_{i,T} = \frac{v_{i,1}}{v_{i,1} + \frac{\sigma^2}{\tau^2}} \qquad \text{and} \qquad \widehat{X}_{i,1:T} = \frac{1}{v_{i,1}}\, v_{i,1}\, X_{i,1} = X_{i,1}.$$

Hence, we get

$$\widehat{\widehat{\mu(\Theta_i)}} = \frac{v_{i,1}}{v_{i,1} + \frac{\sigma^2}{\tau^2}}\, X_{i,1} + \frac{\frac{\sigma^2}{\tau^2}}{v_{i,1} + \frac{\sigma^2}{\tau^2}}\, \mu_0 = \frac{v_{i,1}}{v_{i,1} + \frac{0.088}{0.00024}}\, X_{i,1} + \frac{\frac{0.088}{0.00024}}{v_{i,1} + \frac{0.088}{0.00024}}\, 0.088.$$

The results for the 5 regions are summarized in the following table:

	region 1	region 2	region 3	region 4	region 5
$\alpha_{i,T}$	99.3%	96.5%	99.7%	99.0%	92.0%
$\widehat{X}_{i,1:T}$	7.8%	7.8%	7.4%	9.8%	7.5%
$\widehat{\widehat{\mu(\Theta_i)}}$	7.8%	7.9%	7.4%	9.8%	7.6%

Note that since the credibility coefficient $\kappa = \sigma^2/\tau^2 \approx 367$ is rather small compared to the volumes $v_{1,1}, \ldots, v_{5,1}$, the credibility weights $\alpha_{1,T}, \ldots, \alpha_{5,T}$ are fairly high. Moreover, the observation based estimators are almost the same for the regions $1, 2, 3$ and 5 and only $\widehat{X}_{4,1:T}$ is roughly 2% higher. As a result, only for the smallest two credibility weights $\alpha_{2,T}$ and $\alpha_{5,T}$ we see a slight upwards deviation of the corresponding inhomogeneous credibility estimators $\widehat{\widehat{\mu(\Theta_2)}}$ and $\widehat{\widehat{\mu(\Theta_5)}}$ from the observation based estimators $\widehat{X}_{2,1:T}$ and $\widehat{X}_{5,1:T}$ towards μ_0.

(b) Since the number of policies grows 5% in each region, next year's numbers of policies $v_{1,2}, \ldots, v_{5,2}$ are given by

	region 1	region 2	region 3	region 4	region 5
$v_{i,2}$	52'564	10'642	127'376	36'797	4'402

Similarly to part (a), we define

$$X_{i,2} = \frac{N_{i,2}}{v_{i,2}},$$

for all $i \in \{1, \ldots, 5\}$. Then, according to formula (8.17) of the lecture notes, the mean square error of prediction is given by

$$\mathbb{E}\left[\left(\frac{N_{i,2}}{v_{i,2}} - \widehat{\widehat{\mu(\Theta_i)}}\right)^2\right] = \mathbb{E}\left[\left(X_{i,2} - \widehat{\widehat{\mu(\Theta_i)}}\right)^2\right] = \frac{\sigma^2}{v_{i,2}} + (1 - \alpha_{i,T})\,\tau^2,$$

for all $i \in \{1, \ldots, 5\}$. We get the following square-rooted mean square errors of prediction for the five regions:

	region 1	region 2	region 3	region 4	region 5
$\sqrt{\text{mse of prediction}}$	0.185%	0.408%	0.119%	0.221%	0.627%
in % of the credibility estimators	2.4%	5.2%	1.6%	2.2%	8.3%

Note that we get the highest (square-rooted) mean square errors of prediction for the regions 2 and 5, i.e. exactly for those regions for which we also have the lowest volumes and, consequently, the lowest credibility weights. Of course, this is due to the formula for the mean square error of prediction given above.

Solution 11.2 (Homogeneous) Credibility Estimators for Claim Sizes

We define

$$X_{i,t} = \frac{Y_{i,t}}{v_{i,t}},$$

for all $i \in \{1, 2, 3, 4\}$, $t \in \{1, 2\}$. Then we have

$$\mathbb{E}[X_{i,t} \,|\, \Theta_i] = \frac{1}{v_{i,t}}\,\mathbb{E}[Y_{i,t} \,|\, \Theta_i] = \frac{1}{v_{i,t}}\,\frac{\mu(\Theta_i) c v_{i,t}}{c} = \mu(\Theta_i)$$

and

$$\operatorname{Var}(X_{i,t} \,|\, \Theta_i) = \frac{1}{v_{i,t}^2}\operatorname{Var}(Y_{i,t} \,|\, \Theta_i) = \frac{1}{v_{i,t}^2}\,\frac{\mu(\Theta_i) c v_{i,t}}{c^2} = \frac{\mu(\Theta_i)}{c v_{i,t}} = \frac{\sigma^2(\Theta_i)}{v_{i,t}},$$

with

$$\sigma^2(\Theta_i) = \frac{\mu(\Theta_i)}{c} = \frac{\Theta_i}{c},$$

for all $i \in \{1,2,3,4\}$, $t \in \{1,2\}$. Moreover, using that

$$\mathbb{E}[X_{i,t}^2 \,|\, \Theta_i] = \operatorname{Var}(X_{i,t} \,|\, \Theta_i) + \mathbb{E}[X_{i,t} \,|\, \Theta_i]^2 = \frac{\mu(\Theta_i)}{cv_{i,t}} + \mu(\Theta_i)^2 = \frac{\Theta_i}{cv_{i,t}} + \Theta_i^2$$

we get

$$\mathbb{E}[X_{i,t}^2] = \mathbb{E}\left[\mathbb{E}[X_{i,t}^2 \,|\, \Theta_i]\right] = \mathbb{E}\left[\frac{\Theta_i}{cv_{i,t}} + \Theta_i^2\right] < \infty$$

by assumption, for all $i \in \{1,2,3,4\}$, $t \in \{1,2\}$. In particular, the Model Assumptions 8.13 of the lecture notes for the Bühlmann-Straub model are satisfied.

(a) In order to calculate the homogeneous credibility estimators, we need to estimate the structural parameters $\sigma^2 = \mathbb{E}[\sigma^2(\Theta_1)]$ and $\tau^2 = \operatorname{Var}(\mu(\Theta_1))$. First, following Theorem 8.17 of the lecture notes, we define the observation based estimator $\widehat{X}_{i,1:T}$ as

$$\widehat{X}_{i,1:T} = \frac{1}{\sum_{t=1}^T v_{i,t}} \sum_{t=1}^T v_{i,t} X_{i,t} = \frac{v_{i,1}X_{i,1} + v_{i,2}X_{i,2}}{v_{i,1} + v_{i,2}} = \frac{Y_{i,1} + Y_{i,2}}{v_{i,1} + v_{i,2}},$$

for all $i \in \{1,2,3,4\}$. According to formula (8.15) of the lecture notes, σ^2 can be estimated by

$$\widehat{\sigma}_T^2 = \frac{1}{I} \sum_{i=1}^I \frac{1}{T-1} \sum_{t=1}^T v_{i,t} \left(X_{i,t} - \widehat{X}_{i,1:T}\right)^2 \approx 1.3 \cdot 10^{10}.$$

For the estimator $\widehat{\tau}_T^2$ of τ^2, we define first the weighted sample mean $\bar{X}$ over all observations by

$$\bar{X} = \frac{1}{\sum_{i=1}^I \sum_{t=1}^T v_{i,t}} \sum_{i=1}^I \sum_{t=1}^T v_{i,t} X_{i,t} = \frac{\sum_{i=1}^I Y_{i,1} + Y_{i,2}}{\sum_{i=1}^I v_{i,1} + v_{i,2}} \approx 7004.$$

Then, as on page 219 of the lecture notes, we define $\widehat{v}_T^2$, c_w and $\widehat{t}_T^2$ as

$$\widehat{v}_T^2 = \frac{I}{I-1} \sum_{i=1}^4 \frac{v_{i,1} + v_{i,2}}{\sum_{j=1}^I v_{j,1} + v_{j,2}} \left(\widehat{X}_{i,1:T} - \bar{X}\right)^2 \approx 9.3 \cdot 10^7,$$

$$c_w = \frac{I-1}{I} \left[\sum_{i=1}^I \frac{v_{i,1} + v_{i,2}}{\sum_{j=1}^I v_{j,1} + v_{j,2}} \left(1 - \frac{v_{i,1} + v_{i,2}}{\sum_{j=1}^I v_{j,1} + v_{j,2}}\right)\right]^{-1} \approx 1.425$$

and

$$\widehat{t}_T^2 = c_w \left(\widehat{v}_T^2 - \frac{I\,\widehat{\sigma}_T^2}{\sum_{i=1}^I v_{i,1} + v_{i,2}}\right) \approx 1.25 \cdot 10^8.$$

Then, using formula (8.16) of the lecture notes, τ^2 can be estimated by

$$\widehat{\tau}_T^2 = \max\left\{\widehat{t}_T^2, 0\right\} = \widehat{t}_T^2 \approx 1.25 \cdot 10^8.$$

Now we are ready to calculate the inhomogeneous credibility estimators. Let $i \in \{1,2,3,4\}$. Then, according to Theorem 8.17 of the lecture notes, the inhomogeneous credibility estimator $\widehat{\widehat{\mu(\Theta_i)}}^{\text{hom}}$ is given by

$$\widehat{\widehat{\mu(\Theta_i)}}^{\text{hom}} = \alpha_{i,T}\, \widehat{X}_{i,1:T} + (1 - \alpha_{i,T})\, \widehat{\mu}_T$$

with credibility weight $\alpha_{i,T}$ and estimate $\widehat{\mu}_T$ given by

$$\alpha_{i,T} = \frac{v_{i,1} + v_{i,2}}{v_{i,1} + v_{i,2} + \widehat{\sigma}_T^2 / \widehat{\tau}_T^2} \qquad \text{and} \qquad \widehat{\mu}_T = \frac{1}{\sum_{i=1}^I \alpha_{i,T}} \sum_{i=1}^I \alpha_{i,T} \widehat{X}_{i,1:T}.$$

Hence, we get

$$\widehat{\widehat{\mu(\Theta_i)}}^{\text{hom}} = \alpha_{i,T} \, \widehat{X}_{i,1:T} + \frac{(1-\alpha_{i,T})}{\sum_{i=1}^I \alpha_{i,T}} \sum_{i=1}^I \alpha_{i,T} \widehat{X}_{i,1:T}$$

The results for the 4 risk classes are summarized in the following table:

	risk class 1	risk class 2	risk class 3	risk class 4
$\alpha_{i,T}$	95.4%	98.4%	82.4%	89.6%
$\widehat{X}_{i,1:T}$	10'493	1'907	18'375	29'197
$\widehat{\widehat{\mu(\Theta_i)}}$	10'677	2'107	17'702	27'665

Moreover, we get $\hat{\mu}_T \approx$ 14'538. Looking at the credibility weights $\alpha_{1,1}, \alpha_{2,1}, \alpha_{3,1}$ and $\alpha_{4,1}$, we see that the estimated credibility coefficient $\hat{\kappa} = \hat{\sigma}_T^2/\hat{\tau}_T^2 \approx 104$ has the biggest impact on risk classes 3 and 4 where we have less volumes compared to risk classes 1 and 2. As a result, the smoothing of the observation based estimators $\widehat{X}_{1,1:T}, \widehat{X}_{2,1:T}, \widehat{X}_{3,1:T}$ and $\widehat{X}_{4,1:T}$ towards $\hat{\mu}_T$ is strongest for risk classes 1 and 2.

(b) Since the number of claims grows 5% in each region, next year's numbers of claims $v_{1,3}, \ldots, v_{4,3}$ are given by

	risk class 1	risk class 2	risk class 3	risk class 4
$v_{i,3}$	1'167	3'468	262	479

Similarly to part (a), we define

$$X_{i,3} = \frac{Y_{i,3}}{v_{i,3}},$$

for all $i \in \{1,2,3,4\}$. Then, according to formula (8.17) of the lecture notes, the mean square error of prediction can be estimated by

$$\widehat{\mathbb{E}}\left[\left(\frac{Y_{i,3}}{v_{i,3}} - \widehat{\widehat{\mu(\Theta_i)}}\right)^2\right] = \widehat{\mathbb{E}}\left[\left(X_{i,3} - \widehat{\widehat{\mu(\Theta_i)}}\right)^2\right] = \frac{\widehat{\sigma}_T^2}{v_{i,3}} + (1-\alpha_{i,T})\,\widehat{\tau}_T^2,$$

for all $i \in \{1,2,3,4\}$. We get the following square-rooted mean square errors of prediction for the four risk classes:

	risk class 1	risk class 2	risk class 3	risk class 4
$\sqrt{\text{estimated mse of prediction}}$	4'099	2'390	8'446	6'331
in % of the credibility estimators	38.4%	113.4%	47.7%	22.9%

According to the formula given above for the estimated mean square error of prediction, we observe that, the smaller the volumes of a particular risk class, the bigger the corresponding (square-rooted) estimated mean square error of prediction. Moreover, note that these square-rooted estimated mean square errors of prediction are rather high compared to the credibility estimators, which indicates a high variability within the individual risk classes.

Solution 11.3 Pareto-Gamma Model

(a) Let $f_{\mathbf{Y}|\Lambda}$ and f_Λ denote the density of $\mathbf{Y}\,|\,\Lambda$ and f_Λ, respectively. Then we have

$$f_{\mathbf{Y}|\Lambda}(y_1,\dots,y_T\,|\,\Lambda=\alpha)=\prod_{t=1}^T \frac{\alpha}{\theta}\left(\frac{y_t}{\theta}\right)^{-(\alpha+1)}\cdot 1_{\{y_t\geq\theta\}}=\alpha^T\left(\prod_{t=1}^T\frac{y_t}{\theta}\right)^{-\alpha}\prod_{t=1}^T\frac{y_t}{\theta}\cdot 1_{\{y_t\geq\theta\}}$$

and

$$f_\Lambda(\alpha)=\frac{c^\gamma}{\Gamma(\gamma)}\alpha^{\gamma-1}\exp\{-c\alpha\}\cdot 1_{\{\alpha>0\}}.$$

Let $f_{\Lambda|\mathbf{Y}}$ denote the density of $\Lambda|\mathbf{Y}$. Then, for all $\alpha>0$ and $y_1,\dots,y_T\geq\theta$, we have

$$\begin{aligned}
f_{\Lambda|\mathbf{Y}}(\alpha\,|\,Y_1=y_1,\dots,Y_T=y_T) &= \frac{f_{\mathbf{Y}|\Lambda}(y_1,\dots,y_T\,|\,\Lambda=\alpha)\,f_\Lambda(\alpha)}{\int_0^\infty f_{\mathbf{Y}|\Lambda}(y_1,\dots,y_T\,|\,\Lambda=x)\,f_\Lambda(x)\,dx}\\
&\propto \alpha^T\left(\prod_{t=1}^T\frac{y_t}{\theta}\right)^{-\alpha}\alpha^{\gamma-1}\exp\{-c\alpha\}\\
&= \alpha^{\gamma+T-1}\exp\left\{-\alpha\sum_{t=1}^T\log\frac{y_t}{\theta}\right\}\exp\{-c\alpha\}\\
&= \alpha^{\gamma+T-1}\exp\left\{-\alpha\left(\sum_{t=1}^T\log\frac{y_t}{\theta}+c\right)\right\},
\end{aligned}$$

i.e. we have shown that

$$\Lambda\,|\,\mathbf{Y}\sim\Gamma\left(\gamma+T, c+\sum_{t=1}^T\log\frac{Y_t}{\theta}\right).$$

(b) We calculate

$$\begin{aligned}
\alpha_T\,\widehat{\lambda}_T+(1-\alpha_T)\,\lambda_0 &= \frac{\sum_{t=1}^T\log\frac{Y_t}{\theta}}{c+\sum_{t=1}^T\log\frac{Y_t}{\theta}}\,\frac{T}{\sum_{t=1}^T\log\frac{Y_t}{\theta}}+\frac{c}{c+\sum_{t=1}^T\log\frac{Y_t}{\theta}}\,\frac{\gamma}{c}\\
&= \frac{\gamma+T}{c+\sum_{t=1}^T\log\frac{Y_t}{\theta}}\\
&= \widehat{\lambda}_T^{\text{post}}.
\end{aligned}$$

(c) For the (mean square error) uncertainty of the posterior estimator $\widehat{\lambda}_T^{\text{post}}=\mathbb{E}[\Lambda\,|\,\mathbf{Y}]$ we have

$$\begin{aligned}
\mathbb{E}\left[\left(\Lambda-\widehat{\lambda}_T^{\text{post}}\right)^2\,\middle|\,\mathbf{Y}\right] &= \mathbb{E}\left[\left(\Lambda-\mathbb{E}[\Lambda\,|\,\mathbf{Y}]\right)^2\,\middle|\,\mathbf{Y}\right]\\
&= \operatorname{Var}(\Lambda\,|\,\mathbf{Y})\\
&= \frac{\gamma+T}{\left(c+\sum_{t=1}^T\log\frac{Y_t}{\theta}\right)^2}\\
&= \frac{1}{c+\sum_{t=1}^T\log\frac{Y_t}{\theta}}\,\widehat{\lambda}_T^{\text{post}}\\
&= (1-\alpha_T)\,\frac{1}{c}\,\widehat{\lambda}_T^{\text{post}}.
\end{aligned}$$

(d) Analogously to $\widehat{\lambda}_T^{\text{post}}$, the posterior estimator $\widehat{\lambda}_{T-1}^{\text{post}}$ in the sub-model where we only have observed $(Y_1, \ldots, Y_{T-1})$ is given by

$$\widehat{\lambda}_{T-1}^{\text{post}} = \frac{\gamma + T - 1}{c + \sum_{t=1}^{T-1} \log \frac{Y_t}{\theta}}.$$

Then we can calculate

$$\begin{aligned}
\beta_T \, \frac{1}{\log \frac{Y_T}{\theta}} + (1 - \beta_T) \, \widehat{\lambda}_{T-1}^{\text{post}} &= \frac{\log \frac{Y_T}{\theta}}{c + \sum_{t=1}^{T} \log \frac{Y_t}{\theta}} \, \frac{1}{\log \frac{Y_T}{\theta}} + \frac{c + \sum_{t=1}^{T-1} \log \frac{Y_t}{\theta}}{c + \sum_{t=1}^{T} \log \frac{Y_t}{\theta}} \, \frac{\gamma + T - 1}{c + \sum_{t=1}^{T-1} \log \frac{Y_t}{\theta}} \\
&= \frac{\gamma + T}{c + \sum_{t=1}^{T} \log \frac{Y_t}{\theta}} \\
&= \widehat{\lambda}_T^{\text{post}}.
\end{aligned}$$

Remark: Suppose we would like to use the observations $Y_1, \ldots, Y_{T-1}$ in order to estimate Y_T in a Bayesian sense. Then we have

$$\begin{aligned}
\mathbb{E}[Y_T \,|\, Y_1, \ldots, Y_{T-1}] &= \mathbb{E}\left[\mathbb{E}\left[Y_T \,|\, Y_1, \ldots, Y_{T-1}, \Lambda\right] \,|\, Y_1, \ldots, Y_{T-1}\right] \quad \text{a.s.} \\
&= \mathbb{E}\left[\mathbb{E}\left[Y_T \,|\, \Lambda\right] \,|\, Y_1, \ldots, Y_{T-1}\right] \quad \text{a.s.},
\end{aligned}$$

where in the second equality we used that, conditionally given Λ, $Y_1, \ldots, Y_T$ are independent. Now, by assumption,

$$Y_T \,|\, \Lambda \sim \text{Pareto}(\theta, \Lambda).$$

In particular, $\mathbb{E}[Y_T \,|\, \Lambda] < \infty$ if and only if $\Lambda > 1$. However, according to part (a), we have

$$\Lambda \,|\, (Y_1, \ldots, Y_{T-1}) \sim \Gamma\left(\gamma + T - 1, c + \sum_{t=1}^{T-1} \log \frac{Y_t}{\theta}\right).$$

Since the range of a gamma distribution is the whole positive real line, this implies that

$$0 < \mathbb{P}\left[\Lambda \leq 1 \,|\, Y_1, \ldots, Y_{T-1}\right] = \mathbb{P}\left[\mathbb{E}\left[Y_T \,|\, \Lambda\right] = \infty \,|\, Y_1, \ldots, Y_{T-1}\right] \quad \text{a.s.}$$

We conclude that

$$\mathbb{E}[Y_T \,|\, Y_1, \ldots, Y_{T-1}] = \infty \quad \text{a.s.}$$

Coordinator
A. Gabrielli

Non-Life Insurance: Mathematics and Statistics

Solution sheet 12

Solution 12.1 Chain-Ladder and Bornhuetter-Ferguson

(a) According to formula (9.5) of the lecture notes, the CL factor f_j can be estimated by

$$\widehat{f}_j^{CL} = \frac{\sum_{i=1}^{I-j-1} C_{i,j+1}}{\sum_{i=1}^{I-j-1} C_{i,j}} = \sum_{i=1}^{I-j-1} \frac{C_{i,j}}{\sum_{n=1}^{I-j-1} C_{n,j}} \frac{C_{i,j+1}}{C_{i,j}},$$

for all $j \in \{0, \ldots, 8\}$. Then, for all $i \in \{2, \ldots, 10\}$ and $j \in \{0, \ldots, 9\}$ with $i + j > 10$, $C_{i,j}$ can be predicted by

$$\widehat{C}_{i,j}^{CL} = C_{i,I-i} \prod_{k=I-i}^{j-1} \widehat{f}_k^{CL}.$$

In particular, for the prediction $\widehat{C}_{i,J}^{CL}$ of the ultimate claim $C_{i,J}$ we have

$$\widehat{C}_{i,J}^{CL} = C_{i,I-i} \prod_{j=I-i}^{J-1} \widehat{f}_j^{CL}. \tag{1}$$

The estimates $\widehat{f}_0^{CL}, \ldots, \widehat{f}_8^{CL}$ and the prediction for the lower triangle $\mathcal{D}_{10}^c$ are then given by

accident	development year j									
year i	0	1	2	3	4	5	6	7	8	9
1										
2										10'663'318
3									10'646'884	10'662'008
4								9'734'574	9'744'764	9'758'606
5							9'837'277	9'847'906	9'858'214	9'872'218
6						10'005'044	10'056'528	10'067'393	10'077'931	10'092'247
7					9'419'776	9'485'469	9'534'279	9'544'580	9'554'571	9'568'143
8				8'445'057	8'570'389	8'630'159	8'674'568	8'683'940	8'693'030	8'705'378
9			8'243'496	8'432'051	8'557'190	8'616'868	8'661'208	8'670'566	8'679'642	8'691'971
10		8'470'989	9'129'696	9'338'521	9'477'113	9'543'206	9'592'313	9'602'676	9'612'728	9'626'383
$\widehat{f}_j^{CL}$	1.493	1.078	1.023	1.015	1.007	1.005	1.001	1.001	1.001	

Note that $\widehat{f}_0^{CL} \approx 1.5$ while $\widehat{f}_j^{CL}$ is close to 1, for all $j \in \{1, \ldots, 8\}$, i.e. we observe a rather fast claims settlement in this example. The CL reserves $\widehat{\mathcal{R}}_i^{CL}$ at time $t = I$ are given by

$$\widehat{\mathcal{R}}_i^{CL} = \widehat{C}_{i,J}^{CL} - C_{i,I-i} = C_{i,I-i} \left(\prod_{j=I-i}^{J-1} \widehat{f}_j^{CL} - 1 \right),$$

for all accident years $i \in \{2, \ldots, 10\}$. Moreover, since $C_{1,J} = C_{1,I-1}$ is known, we have $\widehat{\mathcal{R}}_1^{CL} = 0$. Summarizing, we get

accident year i	1	2	3	4	5	6	7	8	9	10
CL reserve $\widehat{\mathcal{R}}_i^{CL}$	0	15'126	26'257	34'538	85'302	156'494	286'121	449'167	1'043'242	3'950'815

By aggregating the CL reserves over all accident years, we get the CL predictor $\widehat{\mathcal{R}}^{CL}$ for the outstanding loss liabilities of past exposure claims:

$$\widehat{\mathcal{R}}^{CL} = \sum_{i=1}^{I} \widehat{\mathcal{R}}_i^{CL} = 6'047'061.$$

(b) For all $j \in \{0, \ldots, J-1\}$, we define $\widehat{\beta}_j^{CL}$ as the proportion paid after the first j development periods according to the estimated CL pattern, i.e.

$$\widehat{\beta}_0^{CL} = \frac{1}{\prod_{l=0}^{J-1} \widehat{f}_l^{CL}} = \prod_{l=0}^{J-1} \frac{1}{\widehat{f}_l^{CL}}$$

and

$$\widehat{\beta}_j^{CL} = \frac{\prod_{l=0}^{j-1} \widehat{f}_l^{CL}}{\prod_{l=0}^{J-1} \widehat{f}_l^{CL}} = \prod_{l=j}^{J-1} \frac{1}{\widehat{f}_l^{CL}},$$

for all $j \in \{1, \ldots, J-1\}$. We get

development period j	0	1	2	3	4	5	6	7	8
proportion $\widehat{\beta}_j^{CL}$ paid so far	0.590	0.880	0.948	0.970	0.984	0.991	0.996	0.998	0.999

According to formula (9.8) of the lecture notes, in the Bornhuetter-Ferguson method the ultimate claim $C_{i,J}$ is predicted by

$$\widehat{C}_{i,J}^{BF} = C_{i,I-i} + \widehat{\mu}_i \left(1 - \widehat{\beta}_{I-i}^{CL}\right),$$

for all accident years $i \in \{2, \ldots, 10\}$. Thus, the Bornhuetter-Ferguson reserves $\widehat{\mathcal{R}}_i^{BF}$ are given by

$$\widehat{\mathcal{R}}_i^{BF} = \widehat{C}_{i,J}^{BF} - C_{i,I-i} = \widehat{\mu}_i \left(1 - \widehat{\beta}_{I-i}^{CL}\right)$$

for all accident years $i \in \{2, \ldots, 10\}$. Moreover, since $C_{1,J} = C_{1,I-1}$ is known, we have $\widehat{\mathcal{R}}_1^{BF} = 0$. Summarizing, we get

accident year i	1	2	3	4	5	6	7	8	9	10
CL reserve $\widehat{\mathcal{R}}_i^{CL}$	0	16'124	26'998	37'575	95'434	178'024	341'305	574'089	1'318'646	4'768'384

By aggregating the BF reserves over all accident years, we get the BF predictor $\widehat{\mathcal{R}}^{BF}$ for the outstanding loss liabilities of past exposure claims:

$$\widehat{\mathcal{R}}^{BF} = \sum_{i=1}^{I} \widehat{\mathcal{R}}_i^{BF} = 7'356'580.$$

(c) Note that for accident year 1 we have

$$\widehat{\mathcal{R}}_1^{CL} = 0 = \widehat{\mathcal{R}}_1^{BF}.$$

Now let $i \in \{2, \ldots, 10\}$. Then, in parts (a) and (b) we can observe that

$$\widehat{\mathcal{R}}_i^{CL} < \widehat{\mathcal{R}}_i^{BF}.$$

This can be explained as follows: Equation (1) can be rewritten as

$$\begin{aligned}
\widehat{C}_{i,J}^{CL} &= C_{i,I-i} \prod_{j=I-i}^{J-1} \widehat{f}_j^{CL} \\
&= C_{i,I-i} + C_{i,I-i} \left(\prod_{j=I-i}^{J-1} \widehat{f}_j^{CL} - 1 \right) \\
&= C_{i,I-i} + C_{i,I-i} \prod_{j=I-i}^{J-1} \widehat{f}_j^{CL} \left(1 - \prod_{j=I-i}^{J-1} \frac{1}{\widehat{f}_j^{CL}} \right) \\
&= C_{i,I-i} + \widehat{C}_{i,J}^{CL} \left(1 - \widehat{\beta}_{I-i}^{CL} \right).
\end{aligned}$$

Comparing this to

$$\widehat{C}_{i,J}^{BF} = C_{i,I-i} + \widehat{\mu}_i \left(1 - \widehat{\beta}_{I-i}^{CL} \right)$$

and noting that for the prior information $\widehat{\mu}_i$ we have $\widehat{\mu}_i > \widehat{C}_{i,J}^{CL}$, we immediately see that

$$\widehat{C}_{i,J}^{CL} < \widehat{C}_{i,J}^{BF},$$

which of course implies that

$$\widehat{\mathcal{R}}_i^{CL} = \widehat{C}_{i,J}^{CL} - C_{i,I-i} < \widehat{C}_{i,J}^{BF} - C_{i,I-i} = \widehat{\mathcal{R}}_i^{BF}.$$

Concluding, we found that choosing a prior information $\widehat{\mu}_i$ bigger than the estimated CL ultimate $\widehat{C}_{i,J}^{CL}$ leads to more conservative, i.e. higher reserves in the Bornhuetter-Ferguson method compared to the chain-ladder method.

Solution 12.2 Mack's Formula and Merz-Wüthrich (MW) Formula (R Exercise)

See the R-Code below for getting the results presented in the following table:

accident year i	CL reserve $\widehat{\mathcal{R}}_i^{CL}$	$\sqrt{\text{total msep}}$ (Mack)	in % reserves	$\sqrt{\text{CDR msep}}$ (MW)	in % $\sqrt{\text{total msep}}$
1	0				
2	15'126	267	1.8 %	267	100 %
3	26'257	914	3.5 %	884	97 %
4	34'538	3'058	8.9 %	2'948	96 %
5	85'302	7'628	8.9 %	7'018	92 %
6	156'494	33'341	21.3 %	32'470	97 %
7	286'121	73'467	25.7 %	66'178	90 %
8	449'167	85'398	19.0 %	50'296	59 %
9	1'043'242	134'337	12.9 %	104'311	78 %
10	3'950'815	410'817	10.4 %	385'773	94 %
total	6'047'061	462'960	7.7 %	420'220	91 %

Mack's square-rooted conditional mean square errors of prediction give us confidence bounds around the estimated CL reserves. We see that for the total claims reserves the one standard deviation confidence bounds are 7.7%. The biggest uncertainties can be found for accident years 6, 7 and 8, where the one standard deviation confidence bounds are roughly 20% or even higher. Moreover, MW's square-rooted conditional mean square errors of prediction measure the contribution of the next accounting year to the total uncertainty given by Mack's square-rooted conditional mean square errors of prediction. We see that 91% of the total uncertainty is due to the next accounting year. This high value can be explained by the fast claims settlement already noticed in Exercise 12.1, (a).

```
### Load the required packages
require(xlsx)
library(ChainLadder)

### Download the data from the link indicated on the exercise sheet
### Store the data under the name "Exercise.12.Data.xls" in the same folder as this R-Code
### Load the data
data <- read.xlsx("Exercise.12.Data.xls", sheetName = "Data_1", rowIndex = c(21:31), colIndex = c(2:11))

### Bring the data in the appropriate triangular form and label the axes
tri <- as.triangle(as.matrix(data))
dimnames(tri)=list(origin=1:nrow(tri),dev=1:ncol(tri))

### Calculate the CL reserves and the corresponding msep's
M <- MackChainLadder(tri, est.sigma = "Mack")

### Cl factors
M$f

### Full triangle
M$FullTriangle

### CL reserves and Mack's square-rooted msep's (including illustrations)
M
plot(M)
plot(M, lattice = TRUE)

### CL reserves, MW's square-rooted msep's and Mack's square-rooted msep's
CDR(M)

### Mack's square-rooted msep's in % of the reserves
round(CDR(M)[,3] / CDR(M)[,1],3) * 100

### MW's square-rooted msep's in % of Mack's square-rooted msep's
round(CDR(M)[,2] / CDR(M)[,3],2) * 100

### Full uncertainty picture
CDR(M, dev="all")
```

Solution 12.3 Conditional MSEP and Claims Development Result

Note that the equalities in this exercise involving a conditional expectation are to be understood in an almost sure sense.

(a) Since X is square-integrable, also $\mathbb{E}[X \mid \mathcal{D}]$ is. Now, by subtracting and adding $\mathbb{E}[X \mid \mathcal{D}]$, we can write

$$\begin{aligned}
\operatorname{msep}_{X|\mathcal{D}}\left(\widehat{X}\right) &= \mathbb{E}\left[\left(X - \widehat{X}\right)^2 \,\middle|\, \mathcal{D}\right] \\
&= \mathbb{E}\left[\left(X - \mathbb{E}[X \mid \mathcal{D}] + \mathbb{E}[X \mid \mathcal{D}] - \widehat{X}\right)^2 \,\middle|\, \mathcal{D}\right] \\
&= \mathbb{E}\left[(X - \mathbb{E}[X \mid \mathcal{D}])^2 \,\middle|\, \mathcal{D}\right] + \mathbb{E}\left[\left(\mathbb{E}[X \mid \mathcal{D}] - \widehat{X}\right)^2 \,\middle|\, \mathcal{D}\right] \\
&\qquad + 2\mathbb{E}\left[(X - \mathbb{E}[X \mid \mathcal{D}])\left(\mathbb{E}[X \mid \mathcal{D}_I] - \widehat{X}\right) \,\middle|\, \mathcal{D}\right] \\
&= \operatorname{Var}(X \mid \mathcal{D}) + \mathbb{E}\left[\left(\mathbb{E}[X \mid \mathcal{D}] - \widehat{X}\right)^2 \,\middle|\, \mathcal{D}\right] \\
&\qquad + 2\,\mathbb{E}\left[(X - \mathbb{E}[X \mid \mathcal{D}])\left(\mathbb{E}[X \mid \mathcal{D}] - \widehat{X}\right) \,\middle|\, \mathcal{D}\right].
\end{aligned}$$

Since $\mathbb{E}[X \mid \mathcal{D}]$ and $\widehat{X}$ are $\mathcal{D}$-measurable, we get

$$\mathbb{E}\left[\left(\mathbb{E}[X \mid \mathcal{D}] - \widehat{X}\right)^2 \,\middle|\, \mathcal{D}\right] = \left(\mathbb{E}[X \mid \mathcal{D}] - \widehat{X}\right)^2$$

and

$$\begin{aligned}
\mathbb{E}\left[(X - \mathbb{E}[X \mid \mathcal{D}])\left(\mathbb{E}[X \mid \mathcal{D}] - \widehat{X}\right) \,\middle|\, \mathcal{D}\right] &= \left(\mathbb{E}[X \mid \mathcal{D}] - \widehat{X}\right)\mathbb{E}\left[(X - \mathbb{E}[X \mid \mathcal{D}]) \,\middle|\, \mathcal{D}\right] \\
&= \left(\mathbb{E}[X \mid \mathcal{D}] - \widehat{X}\right)(\mathbb{E}[X \mid \mathcal{D}] - \mathbb{E}[X \mid \mathcal{D}]) \\
&= 0.
\end{aligned}$$

By collecting the terms, we get the result

$$\operatorname{msep}_{X|\mathcal{D}}\left(\widehat{X}\right) = \mathbb{E}\left[\left(X - \widehat{X}\right)^2 \,\middle|\, \mathcal{D}\right] = \operatorname{Var}(X \mid \mathcal{D}) + \left(\mathbb{E}[X \mid \mathcal{D}] - \widehat{X}\right)^2.$$

(b) For $t \in \mathbb{N}$ with $t \geq I$ and $i > t - J$, the claims development result $\mathrm{CDR}_{i,t+1}$ is defined in formulas (9.27) and (9.29) of the lecture notes by

$$\mathrm{CDR}_{i,t+1} = \widehat{C}_{i,J}^{(t)} - \widehat{C}_{i,J}^{(t+1)} = \mathbb{E}\left[C_{i,J} \mid \mathcal{D}_t\right] - \mathbb{E}\left[C_{i,J} \mid \mathcal{D}_{t+1}\right],$$

which implies, since $\mathcal{D}_t \subset \mathcal{D}_{t+1}$, that $\mathrm{CDR}_{i,t+1}$ is $\mathcal{D}_{t+1}$-measurable. Moreover, using the tower property, we get

$$\begin{aligned}
\mathbb{E}\left[\mathrm{CDR}_{i,t+1} \mid \mathcal{D}_t\right] &= \mathbb{E}\left[\mathbb{E}\left[C_{i,J} \mid \mathcal{D}_t\right] - \mathbb{E}\left[C_{i,J} \mid \mathcal{D}_{t+1}\right] \mid \mathcal{D}_t\right] \\
&= \mathbb{E}\left[C_{i,J} \mid \mathcal{D}_t\right] - \mathbb{E}\left[C_{i,J} \mid \mathcal{D}_t\right] \\
&= 0.
\end{aligned}$$

Note that this result is given in Corollary 9.13 of the lecture notes. In particular, it implies that

$$\mathbb{E}\left[\mathrm{CDR}_{i,t+1}\right] = \mathbb{E}\left[\mathbb{E}\left[\mathrm{CDR}_{i,t+1} \mid \mathcal{D}_t\right]\right] = 0.$$

Now, since $t_1 < t_2$ by assumption, CDR_{i,t_1+1} is $\mathcal{D}_{t_2}$-measurable. Thus, we get

$$\begin{aligned}\mathbb{E}\left[\mathrm{CDR}_{i,t_1+1}\mathrm{CDR}_{i,t_2+1}\right] &= \mathbb{E}\left[\mathbb{E}\left[\mathrm{CDR}_{i,t_1+1}\mathrm{CDR}_{i,t_2+1} \mid \mathcal{D}_{t_2}\right]\right] \\ &= \mathbb{E}\left[\mathrm{CDR}_{i,t_1+1}\mathbb{E}\left[\mathrm{CDR}_{i,t_2+1} \mid \mathcal{D}_{t_2}\right]\right] \\ &= \mathbb{E}\left[\mathrm{CDR}_{i,t_1+1} \cdot 0\right] \\ &= 0.\end{aligned}$$

We can conclude that

$$\mathrm{Cov}\left(\mathrm{CDR}_{i,t_1+1}, \mathrm{CDR}_{i,t_2+1}\right) = \mathbb{E}\left[\mathrm{CDR}_{i,t_1+1}\mathrm{CDR}_{i,t_2+1}\right] - \mathbb{E}\left[\mathrm{CDR}_{i,t_1+1}\right]\mathbb{E}\left[\mathrm{CDR}_{i,t_2+1}\right] = 0.$$