Non-Life Insurance: Mathematics and Statistics

Solution sheet 1

Solution 1.1 Discrete Distribution

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

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

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

(b) For $n \in \mathbb{N}_{>0}$ 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 $p \sum_{k=0}^{\infty} (1-p)^k = 1$, as was shown in (a).

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

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

Thus, we get

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

where we again used that $p \sum_{k=0}^{\infty} (1-p)^k = 1$, 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 converges 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

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

Hence, we get

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

We observe that $\frac{d}{dr}M_N(r)\big|_{r=0} = \mathbb{E}[N]$, which holds in general for all random variables for which 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

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

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

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

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,

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

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 for which the moment generating function exists in an interval around 0.

Solution 1.3 Gaussian Distribution

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

$$\begin{aligned}
M_X(r) \;&=\; \mathbb{E}\left[\exp\{rX\}\right] \;=\; \int_{-\infty}^{\infty} \exp\{rx\}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\}dx \\
&=\; \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\frac{x^2 - 2(\mu + r\sigma^2)x + \mu^2}{\sigma^2}\right\}dx \\
&=\; \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\frac{x^2 - 2(\mu + r\sigma^2)x + \mu^2 + 2r\mu\sigma^2 + r^2\sigma^4 - 2r\mu\sigma^2 - r^2\sigma^4}{\sigma^2}\right\}dx \\
&=\; \exp\left\{r\mu + \frac{r^2\sigma^2}{2}\right\}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\frac{[x-(\mu+r\sigma^2)]^2}{\sigma^2}\right\}dx \\
&=\; \exp\left\{r\mu + \frac{r^2\sigma^2}{2}\right\},
\end{aligned}$$

where the last equality holds true since we integrate the density of a normal distribution with mean $\mu + r\sigma^2$ and variance σ^2.

(b) The moment generating function M_{a+bX} 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 in part (a), 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 (if it exists in an interval around 0) uniquely determines the distribution, see Lemma 1.2 of the lecture notes (version of March 20, 2019), we conclude that

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

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

$$\begin{aligned} M_Y(r) \;&=\; \mathbb{E}\left[\exp\left\{rY\right\}\right] \;=\; \mathbb{E}\left[\exp\left\{r\sum_{i=1}^n X_i\right\}\right] \;=\; \prod_{i=1}^n \mathbb{E}\left[\exp\left\{rX_i\right\}\right] \;=\; \prod_{i=1}^n M_{X_i}(r) \\ &=\; \prod_{i=1}^n \exp\left\{r\mu_i + \frac{r^2\sigma_i^2}{2}\right\} \;=\; \exp\left\{r\sum_{i=1}^n \mu_i + \frac{r^2\sum_{i=1}^n \sigma_i^2}{2}\right\}, \end{aligned}$$

for all $r \in \mathbb{R}$. This is equal to the moment generating function of a Gaussian random variable with expectation $\sum_{i=1}^n \mu_i$ and variance $\sum_{i=1}^n \sigma_i^2$. 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 1.4 χ^2-Distribution

(a) Let $r \in \mathbb{R}$. The moment generating function M_{X_k} of X_k can be calculated as

$$\begin{aligned} M_{X_k}(r) \;&=\; \mathbb{E}\left[\exp\{rX_k\}\right] \;=\; \int_0^\infty \exp\{rx\}\frac{1}{2^{k/2}\Gamma(k/2)}x^{k/2-1}\exp\{-x/2\}\,dx \\ &=\; \int_0^\infty \frac{1}{2^{k/2}\Gamma(k/2)}x^{k/2-1}\exp\{-x(1/2-r)\}\,dx. \end{aligned}$$

This integral (and consequently the moment generating function) exists if and only if $r < 1/2$. Let $r < 1/2$. Then, we use the substitution

$$u \;=\; x(1/2-r), \quad dx \;=\; \frac{1}{1/2-r}du.$$

We get

$$\begin{aligned} M_{X_k}(r) \;&=\; \int_0^\infty \frac{1}{2^{k/2}\Gamma(k/2)}u^{k/2-1}\left(\frac{1}{1/2-r}\right)^{k/2-1}\exp\{-u\}\frac{1}{1/2-r}\,du \\ &=\; \frac{1}{2^{k/2}}\frac{1}{(1/2-r)^{k/2}}\frac{1}{\Gamma(k/2)}\int_0^\infty u^{k/2-1}\exp\{-u\}\,du \\ &=\; \frac{1}{(1-2r)^{k/2}}, \end{aligned}$$

where in the last equality we used the definition of the gamma function

$$\Gamma(z) = \int_0^\infty u^{z-1} \exp\{-u\}\, du, \quad \text{for } z \in \mathbb{R}.$$

(b) For all $r < 1/2$ the moment generating function M_{Z^2} of Z^2 is given by

$$\begin{aligned} M_{Z^2}(r) &= \mathbb{E}\left[\exp\left\{rZ^2\right\}\right] = \int_{-\infty}^\infty \exp\left\{rx^2\right\} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} dx \\ &= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2(1-2r)}{2}\right\} dx \\ &= (1-2r)^{-1/2} \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}(1-2r)^{-1/2}} \exp\left\{-\frac{x^2}{2(1-2r)^{-1}}\right\} dx \\ &= \frac{1}{(1-2r)^{1/2}} \\ &= M_{X_1}(r), \end{aligned}$$

where the second to last equality holds true since we integrate the density of a normal distribution with mean 0 and variance $(1-2r)^{-1} > 0$. We conclude that $Z^2 \overset{(d)}{=} X_1$.

(c) Using that $Z_1, \ldots, Z_k$ are i.i.d., the moment generating function M_Y of $Y = \sum_{i=1}^k Z_i^2$ is given by

$$\begin{aligned} M_Y(r) &= \mathbb{E}\left[\exp\{rY\}\right] = \mathbb{E}\left[\exp\left\{r\sum_{i=1}^k Z_i^2\right\}\right] = \prod_{i=1}^k \mathbb{E}\left[\exp\left\{rZ_i^2\right\}\right] = \Big(M_{Z_1^2}(r)\Big)^k \\ &= \frac{1}{(1-2r)^{k/2}} = M_{X_k}(r), \end{aligned}$$

for all $r < 1/2$. We conclude that $\sum_{i=1}^k Z_i^2 \overset{(d)}{=} X_k$.

Non-Life Insurance: Mathematics and Statistics

Solution sheet 2

Solution 2.1 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

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

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

(b) By taking the logarithm, we get

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

(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 the value 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 $\log f_{\mu,\sigma^2}(x_1, \ldots, x_8) < -8\log c_1 - \frac{1}{2c_2}\sum_{i=1}^{8}(x_i - \mu)^2$. 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, $\widehat{\mu}$ and $\widehat{\sigma}^2$ 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

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

Note that the MLE $\widehat{\sigma}^2$ (considered as an estimator) 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 $\widehat{\mu}$ for $\frac{1}{8}\sum_{i=1}^8 X_i$, we can calculate

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

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

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

By inserting

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

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

we can conclude that

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

i.e. $\widehat{\sigma}^2$ is not unbiased.

(d) Since the logarithms of the claim amounts are assumed to follow a Gaussian distribution and the variance is unknown, we perform a t-test. Under H_0, we have $\mu = 6$. Thus, the test statistic is given by

$$T \;=\; T(\log Y_1, \ldots, \log Y_8) \;=\; \sqrt{8}\,\frac{\frac{1}{8}\sum_{i=1}^8 \log Y_i - 6}{\sqrt{S^2}},$$

where

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

Note that S^2 is an unbiased estimator for the variance σ^2 of the logarithmic claim sizes. 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

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

Thus, for T we get the observation

$$T(x_1, \ldots, x_8) \;=\; \sqrt{8}\,\frac{\frac{1}{8}\sum_{i=1}^8 x_i - 6}{\sqrt{S^2}} \;=\; \sqrt{8}\,\frac{7-6}{\sqrt{8}} \;=\; 1.$$

The probability under H_0 to observe a T that is at least as extreme as the observation 1 we got above is

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

where we used the symmetry of the Student-t distribution around 0, i.e. $\mathbb{P}[T \leq -1] = 1 - \mathbb{P}[T \leq 1]$. The probability $\mathbb{P}[T \leq 1]$ is approximately 0.84, and the p-value is given by

$$\mathbb{P}[|T| \geq 1] \,=\, 2 - 2\mathbb{P}[T \leq 1] \,\approx\, 2 - 2 \cdot 0.84 \,=\, 0.32.$$

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

Solution 2.2 Chebychev's Inequality and Law of Large Numbers

(a) We have $\mu = \mathbb{E}[X_1] = 1\text{'}000 \cdot 0.1 + 0 \cdot 0.9 = 100$.

(b) For $n = 1$ we get

$$\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| \,=\, |X_1 - 100| \,=\, \begin{cases} 900, & \text{with probability } 0.1, \\ 100, & \text{with probability } 0.9. \end{cases}$$

As both values 900 and 100 are bigger than $0.1\mu = 10$, we conclude that $p(1) = 1$. In particular, if we only have $n = 1$ risk in our portfolio, then the corresponding claim amount deviates from the mean claim size by at least 10% with probability equal to 1.

(c) For the n i.i.d. risks $X_1, \ldots, X_n$ we define

$$S(n) \,=\, \sum_{i=1}^{n} \frac{X_i}{1\text{'}000}$$

to be the corresponding (random) number of bikes stolen. We note that $S(n)$ has a binomial distribution with parameters n and $p = 0.1$. In particular, we have

$$\mathbb{P}[S(n) = k] \,=\, \binom{n}{k} p^k \, (1-p)^{n-k},$$

for all $k \in \{0, \ldots, n\}$. For $n \in \mathbb{N}$ we can now write

$$\begin{aligned} p(n) \,&=\, \mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq 0.1\mu\right] \,=\, 1 - \mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < 0.1\mu\right] \\ &=\, 1 - \mathbb{P}\left[-0.1\mu < \frac{1}{n}\sum_{i=1}^{n} X_i - \mu < 0.1\mu\right] \,=\, 1 - \mathbb{P}\left[0.9n\mu < \sum_{i=1}^{n} X_i < 1.1n\mu\right] \\ &=\, 1 - \mathbb{P}\left[\frac{0.9n\mu}{1\text{'}000} < \sum_{i=1}^{n} \frac{X_i}{1\text{'}000} < \frac{1.1n\mu}{1\text{'}000}\right] \,=\, 1 - \mathbb{P}\left[\frac{0.9n\mu}{1\text{'}000} < S(n) < \frac{1.1n\mu}{1\text{'}000}\right]. \end{aligned}$$

For $n = 1\text{'}000$ we get

$$\begin{aligned} p(1\text{'}000) \,&=\, 1 - \mathbb{P}\left[\frac{0.9 \cdot 1\text{'}000 \cdot 100}{1\text{'}000} < S(1\text{'}000) < \frac{1.1 \cdot 1\text{'}000 \cdot 100}{1\text{'}000}\right] \\ &=\, 1 - \mathbb{P}\left[90 < S(1\text{'}000) < 110\right] \\ &=\, 1 - \sum_{k=91}^{109} \binom{1\text{'}000}{k} 0.1^k \, 0.9^{1\text{'}000-k} \\ &\approx\, 0.32. \end{aligned}$$

Thus, if we have $n = 1\text{'}000$ risks in our portfolio, then the sample mean of the claim amounts deviates from the mean claim size by at least 10% with a probability of 0.32. In particular, diversification led to a reduction of this probability.

(d) As

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[X_i\right] = \mathbb{E}[X_1] = \mu$$

and, using the independence of $X_1, \ldots, X_n$,

$$\begin{aligned}\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) &= \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{Var}\left(X_i\right) = \frac{1}{n}\operatorname{Var}(X_1) = \frac{1}{n}\mathbb{E}\left[(X_1-\mu)^2\right] \\ &= \frac{1}{n}\left(900^2\cdot 0.1 + 100^2\cdot 0.9\right) = \frac{90\text{'}000}{n},\end{aligned}$$

Chebychev's inequality leads to

$$p(n) = \mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq 0.1\mu\right] \leq \frac{\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)}{(0.1\mu)^2}\frac{90\text{'}000}{n(0.1\mu)^2} = \frac{900}{n}.$$

We have

$$\frac{900}{n} < 0.01 \qquad \Longleftrightarrow \qquad n > 90\text{'}000.$$

This implies that Chebychev's inequality guarantees that if we have more than 90'000 risks, then the probability that the sample mean of the claim amounts deviates from the mean claim size by at least 10% is smaller than 1%. However, we remark that Chebychev's inequality is very crude. In fact, the true minimum number n of risks such that $p(n) < 0.01$ is given by $n \approx 6\text{'}000$, approximately, while for $n = 90\text{'}000$ we basically have $p(n) \approx 0$.

(e) We have that $X_1, X_2, \ldots$ are i.i.d. and that $\mathbb{E}[|X_1|] = \mathbb{E}[X_1] = \mu < \infty$. Thus, we can apply the strong law of large numbers, and we get

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow \mathbb{E}[X_1] = \mu = 100, \quad \mathbb{P}\text{-a.s.}$$

Solution 2.3 Central Limit Theorem

(a) Let σ^2 be the variance of the claim sizes and $x > 0$. 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

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} Y_i - \mu\right| < \frac{x}{\sqrt{n}}\right] \approx \Phi\left(\frac{x}{\sigma}\right) - \Phi\left(-\frac{x}{\sigma}\right) = 2\Phi\left(\frac{x}{\sigma}\right) - 1,$$

where we used that $\Phi(-\frac{x}{\sigma}) = 1 - \Phi(\frac{x}{\sigma})$. On the one hand, as we are interested in a probabilty of at least 95%, we have to choose $x > 0$ such that $2\Phi(\frac{x}{\sigma}) - 1 = 0.95$. We have

$$2\Phi\left(\frac{x}{\sigma}\right) - 1 = 0.95 \qquad \Longleftrightarrow \qquad \Phi\left(\frac{x}{\sigma}\right) = 0.975.$$

Using $\Phi^{-1}(0.975) = 1.96$, this implies that

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

It follows that

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

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 that

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

(b) In this part we use Chebychev's inequality instead of the Central Limit Theorem in order to derive a minimum number of claims n^{Che} such that with probability of at least 95% the deviation of the sample mean $\frac{1}{n}\sum_{i=1}^n Y_i$ from the mean claim size μ is less than 1%. We have

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^n Y_i - \mu\right| < 0.01\mu\right] \geq 0.95 \qquad \Longleftrightarrow \qquad \mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^n Y_i - \mu\right| \geq 0.01\mu\right] \leq 0.05.$$

Similarly as in Exercise 2.2 we apply Chebychev's inequality to get

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^n Y_i - \mu\right| \geq 0.01\mu\right] \leq \frac{\mathrm{Var}(\frac{1}{n}\sum_{i=1}^n Y_i)}{(0.01\mu)^2} = \frac{\mathrm{Var}(Y_1)}{n \cdot 0.01^2 \cdot \mu^2} = \frac{\mathrm{Vco}(Y_1)^2}{n \cdot 0.01^2} = \frac{160'000}{n}.$$

We have

$$\frac{160'000}{n} \leq 0.05 \qquad \Longleftrightarrow \qquad n \geq 3'200'000.$$

Thus, we get

$$n^{\mathrm{Che}} = 3'200'000 > 614'656 = n^{\mathrm{CLT}}.$$

This comparison confirms that Chebychev's inequality is rather crude, see also Exercise 2.2.

Solution 2.4 Conditional Distribution and Variance Decomposition

(a) First, we write M_Θ for the moment generating function of Θ. As Θ follows an exponential distribution with parameter $\lambda > 0$, we know from Exercise 1.2 that

$$M_\Theta(r) = \mathbb{E}\left[e^{r\Theta}\right] = \frac{\lambda}{\lambda - r},$$

for all $r < \lambda$. As $-v < 0 < \lambda$, we calculate

$$\mathbb{P}[N = 0] = \mathbb{E}[\mathbb{P}[N = 0|\Theta]] = \mathbb{E}\left[e^{-\Theta v}\right] = M_\Theta(-v) = \frac{\lambda}{\lambda + v}.$$

(b) According to the remark on the exercise sheet, we have $\mathbb{E}[N|\Theta] = \Theta v$. The tower property of conditional expectation then leads to

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

as the expectation of an exponential distribution with parameter $\lambda > 0$ is equal to $\frac{1}{\lambda}$, see Exercise 1.2.

(c) Note that

$$\mathbb{E}[N^2] = \mathbb{E}[\mathbb{E}[N^2|\Theta]] = \mathbb{E}[\mathrm{Var}(N|\Theta) + \mathbb{E}[N|\Theta]^2] = \mathbb{E}[\Theta v + (\Theta v)^2] = \frac{v}{\lambda} + \frac{2v^2}{\lambda^2} < \infty,$$

where in the third equation we used that the expectation and the variance of a Poisson distribution are equal to its frequency parameter, and in the fourth equation that the second moment of an exponential distribution with parameter $\lambda > 0$ is equal to $\frac{2}{\lambda^2}$, see Exercise 1.2. In particular, the second moment of N, and thus the variance $\mathrm{Var}(N)$, exist. Now we have

$$\mathbb{E}[\mathrm{Var}(N|\Theta)] = \mathbb{E}[\mathbb{E}[N^2|\Theta] - (\mathbb{E}[N|\Theta])^2] = \mathbb{E}[N^2] - \mathbb{E}[(\mathbb{E}[N|\Theta])^2]$$

and

$$\mathrm{Var}(\mathbb{E}[N|\Theta]) = \mathbb{E}[(\mathbb{E}[N|\Theta])^2] - \mathbb{E}[\mathbb{E}[N|\Theta]]^2 = \mathbb{E}[(\mathbb{E}[N|\Theta])^2] - \mathbb{E}[N]^2.$$

Combining these two results, we get the variance decomposition formula

$$\mathbb{E}[\mathrm{Var}(N|\Theta)] + \mathrm{Var}(\mathbb{E}[N|\Theta]) = \mathbb{E}[N^2] - \mathbb{E}[N]^2 = \mathrm{Var}(N).$$

Using this formula, we can calculate

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

where in the last equation we used that the variance of an exponential distribution with parameter $\lambda > 0$ is equal to $\frac{1}{\lambda^2}$, see Exercise 1.2. In particular, we have

$$\mathrm{Var}(N) = \frac{v}{\lambda} + \frac{v^2}{\lambda^2} > \frac{v}{\lambda} = \mathbb{E}[N],$$

i.e. contrary to the (unconditional) Poisson distribution, the random variable N has a variance which is bigger than the expectation.

Remark: The variance decomposition formula also holds in its general form

$$\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X|\mathcal{G})] + \mathrm{Var}(\mathbb{E}[X|\mathcal{G}]),$$

where X is a square-integrable random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G}$ any sub-σ-algebra of $\mathcal{F}$.

Non-Life Insurance: Mathematics and Statistics

Solution sheet 3

Solution 3.1 No-Claims Bonus

(a) We define the following events:

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

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 $N_1, \dots, N_6$ 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, $N_1, \dots, N_6$ are i.id. Poisson random variables with frequency parameter $\lambda = 0.2$. Therefore, we can calculate

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

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 premium that is still paid after the grant of the no-claims bonus is given by

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

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

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

which leads to

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

We conclude that the surcharge on the 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 i.i.d. 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 Θ has an exponential distribution with parameter $c = 1$, M_Θ is given by

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

for all $r < 1$, see Exercise 1.2, 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 premium that is still paid after the grant of the no-claims bonus is given by

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

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

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

which is considerably bigger than in (a). The reason is that in (b) we introduce dependence between the claim counts of the individual years of the considered car driver. This increases the probability of having no claims in the last six years, and decreases the expected proportion q of the premium that is still paid after the grant of the no-claims bonus.

Solution 3.2 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 $N \sim \text{Poi}(\lambda v)$, $Y_1, Y_2, \dots$ are i.i.d. with distribution function G and N and $Y_1, Y_2, \dots$ are independent. Now we can define S_{sc}, S_{mc} and S_{lc} as

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

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

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

Thus, using Theorem 2.14 (disjoint decomposition of compound Poisson distributions) of the lecture notes (version of March 20, 2019), we get

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

where

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

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

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

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 (version of March 20, 2019), 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

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

Now we can apply Proposition 2.11 of the lecture notes (version of March 20, 2019) 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

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

Now we can apply Proposition 2.11 of the lecture notes (version of March 20, 2019) 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

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

Thus, we find

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

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

$$\sqrt{\mathrm{Var}(S)} \;=\; \sqrt{\mathrm{Var}(S_{\mathrm{sc}}) + \mathrm{Var}(S_{\mathrm{mc}}) + \mathrm{Var}(S_{\mathrm{lc}})} \;=\; \sqrt{3'335'802'914'200} \;\approx\; 1'826'418.$$

(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 million 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 million. Since the smallest claim size for a claim in the large claims layer is given by 2'000'000, with three claims in the large claims layer we already exceed 5 million with probability one. Thus, it is enough to consider only up to two claims. We get

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

We can conclude that

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

Solution 3.3 Compound Distribution

We show that the moment generating function M_S of S is equal to the moment generating function of an exponential distribution with parameter λp. According to Proposition 2.2 of the lecture notes (version of March 20, 2019), M_S is given by (wherever it exists)

$$M_S(r) \,=\, M_N[\log M_{Y_1}(r)],$$

where M_N and M_{Y_1} are the moment generating functions of N and Y_1, respectively. As $S \geq 0$ almost surely, $M_S(r)$ exists at least for all $r < 0$. In Exercises 1.1 and 1.2 we have seen that

$$M_N(r) \,=\, \frac{p\exp\{r\}}{1-(1-p)\exp\{r\}},$$

for all $r < -\log(1-p)$, and that

$$\log M_{Y_1}(r) \,=\, \log\left(\frac{\lambda}{\lambda - r}\right),$$

for all $r < \lambda$. Thus, we get

$$M_S(r) \,=\, M_N\left[\log\left(\frac{\lambda}{\lambda-r}\right)\right] \,=\, \frac{p\exp\left\{\log\left(\frac{\lambda}{\lambda-r}\right)\right\}}{1-(1-p)\exp\left\{\log\left(\frac{\lambda}{\lambda-r}\right)\right\}} \,=\, \frac{\lambda p}{\lambda - r - \lambda(1-p)} \,=\, \frac{\lambda p}{\lambda p - r}.$$

With Lemma 1.3 of the lecture notes (version of March 20, 2019), we conclude that S has indeed an exponential distribution with parameter λp. We remark that for this compound model the corresponding distribution function can be given in closed form. However, usually this is not possible. Therefore, we will consider other methods for the calculation of the distribution function of S in Chapter 4 of the lecture notes (version of March 20, 2019).

Solution 3.4 Compound Binomial Distribution

(a) Let $\tilde{S} \sim \text{CompBinom}(\tilde{v}, \tilde{p}, \tilde{G})$ with the random variable $\tilde{Y}_1$ having distribution function $\tilde{G}$ and moment generating function $M_{\tilde{Y}_1}$. Then, by Proposition 2.6 of the lecture notes (version of March 20, 2019), 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 $M_{S_{\text{lc}}}$ of S_{lc} and show that it is exactly of the form given above. Let $r \in \mathbb{R}$ such that $M_{S_{\text{lc}}}(r)$ exists. Note that since $S_{\text{lc}} \geq 0$ almost surely, its moment generating function is defined at least for all $r < 0$. We have

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

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

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

Note that the distribution function of the random variable $Y_i|Y_i > M$ is G_{lc}. 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

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

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 $N \sim \text{Binom}(v,p)$, $M_N(r)$ is given by

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

Therefore, we get

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

Applying Lemma 1.3 of the lecture notes (version of March 20, 2019), we conclude that $S_{\text{lc}} \sim \text{CompBinom}(\tilde{v}, \tilde{p}, \tilde{G})$ with $\tilde{v} = v$, $\tilde{p} = p[1-G(M)]$ and $\tilde{G} = G_{\text{lc}}$.

(b) In (a) we showed that the number of claims of the compound distribution S_{lc} has a binomial distribution with parameters v and $p[1-G(M)] > 0$. In particular, there is a positive probability that we have v claims with $Y_i > M$. Now suppose that $S_{\text{sc}} > 0$. Then, we know that there is an $i \in \{1,\ldots,N\}$ with $Y_i \le M$. In particular, this claim cannot be part of S_{lc} and there is zero probability that we have v claims with $Y_i > M$. This explains why S_{sc} and S_{lc} cannot be independent. However, note that with the Poisson distribution as claims count distribution such a split in small and large claims leads to independent compound distributions, see Theorem 2.14 of the lecture notes (version of March 20, 2019).

Non-Life Insurance: Mathematics and Statistics

Solution sheet 4

Solution 4.1 Poisson Model and Negative-Binomial Model

(a) In the Poisson model we assume that $N_1, \dots, N_{10}$ are independent with $N_t \sim \mathrm{Poi}(\lambda v_t)$ for all $t \in \{1, \dots, 10\}$. We use Estimator 2.32 of the lecture notes (version of March 20, 2019) to estimate the claim frequency parameter λ by

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

Let $t \in \{1, \dots, 10\}$. We have

$$\mathbb{E}\left[\frac{N_t}{v_t}\right] = \frac{\mathbb{E}[N_t]}{v_t} = \frac{\lambda v_t}{v_t} = \lambda \qquad \text{and} \qquad \mathrm{Var}\left(\frac{N_t}{v_t}\right) = \frac{\mathrm{Var}(N_t)}{v_t^2} = \frac{\lambda v_t}{v_t^2} = \frac{\lambda}{v_t}.$$

Note that for the random variable $N_t \sim \mathrm{Poi}(\lambda v_t)$ we can write

$$N_t \overset{(\mathrm{d})}{=} \sum_{i=1}^{v_t} \widetilde{N}_i,$$

where $\widetilde{N}_1, \dots, \widetilde{N}_{v_t}$ are i.i.d. random variables following a $\mathrm{Poi}(\lambda)$-distribution. Thus, we can use the Central Limit Theorem to get

$$\frac{N_t/v_t - \mathbb{E}\left[N_t/v_t\right]}{\sqrt{\mathrm{Var}\left(N_t/v_t\right)}} = \frac{N_t/v_t - \lambda}{\sqrt{\lambda/v_t}} \Longrightarrow Z,$$

as $v_t \to \infty$, where Z is a random variable following a standard normal distribution. This leads to the approximation

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

i.e. with a probability of roughly 70%, N_t/v_t lies in the interval $[\lambda - \sqrt{\lambda/v_t}, \lambda + \sqrt{\lambda/v_t}]$. Since λ is unknown, we replace it by the estimator $\widehat{\lambda}_{10}^{\mathrm{MLE}}$ to get the approximate prediction interval

$$\left[\widehat{\lambda}_{10}^{\mathrm{MLE}} - \sqrt{\widehat{\lambda}_{10}^{\mathrm{MLE}}/v_t}, \widehat{\lambda}_{10}^{\mathrm{MLE}} + \sqrt{\widehat{\lambda}_{10}^{\mathrm{MLE}}/v_t}\right] \approx [9.90\%, 10.54\%],$$

which should contain roughly 70% of the observed claim frequencies N_t/v_t. We have the following observations of the claim frequencies:

t	1	2	3	4	5	6	7	8	9	10
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 claim frequencies N_t/v_t.

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

(b) By equation (2.9) of the lecture notes (version of March 20, 2019), the test statistic

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

is approximately χ^2-distributed with $10-1=9$ degrees of freedom. By inserting the numbers, we get $\widehat{\chi}^* \approx 14.84$. 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 significance level of 5%.

(c) In the negative-binomial model we assume that $N_1, \dots, N_{10}$ are independent with, conditionally given Θ_t, $N_t \sim \text{Poi}(\Theta_t \lambda v_t)$ 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 (version of March 20, 2019) to estimate the claim frequency parameter λ by

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

As in equation (2.8) of the lecture notes (version of March 20, 2019), we define

$$\widehat{V}_{10}^2 = \frac{1}{9} \sum_{t=1}^{10} v_t \left(\frac{N_t}{v_t} - \widehat{\lambda}_{10}^{\text{NB}}\right)^2 \approx 0.17 > \widehat{\lambda}_{10}^{\text{NB}}.$$

Let $v = v_1 = \dots = v_{10} = 10\text{'}000$. Now we can use Estimator 2.30 of the lecture notes (version of March 20, 2019) to estimate the dispersion parameter γ by

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

For all $t \in \{1, \dots, 10\}$ we have

$$\mathbb{E}\left[\frac{N_t}{v_t}\right] = \frac{\mathbb{E}[N_t]}{v_t} = \frac{\mathbb{E}[\mathbb{E}[N_t|\Theta_t]]}{v_t} = \frac{\mathbb{E}[\Theta_t \lambda v_t]}{v_t} = \frac{\lambda v_t}{v_t} = \lambda,$$

since $\mathbb{E}[\Theta_t] = \gamma/\gamma = 1$, and

$$\text{Var}\left(\frac{N_t}{v_t}\right) = \frac{\mathbb{E}[\text{Var}(N_t|\Theta_t)] + \text{Var}(\mathbb{E}[N_t|\Theta_t])}{v_t^2} = \frac{\mathbb{E}[\Theta_t \lambda v_t] + \text{Var}(\Theta_t \lambda v_t)}{v_t^2} = \frac{\lambda + \lambda^2 v_t/\gamma}{v_t},$$

since $\text{Var}(\Theta_t) = \gamma/\gamma^2 = 1/\gamma$. Similarly as in part (a), we get the prediction interval

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

which should contain roughly 70% of the observed claim frequencies N_t/v_t. Looking at the observations given in Table 1 above, we see that eight of them lie in the estimated interval, which is clearly better than in the Poisson case in part (a). In conclusion, here the negative-binomial model seems more reasonable than the Poisson model.

Solution 4.2 χ^2**-Goodness-of-Fit-Analysis**

(a) The R code used in part (a) is provided in Listing 1.

(i) In Figure 1 (left) we can see that the n MLEs of λ approximately have a Gaussian distribution with mean equal to the true value of $\lambda = 10\%$. On the one hand, this is due to the fact that (under regularity assumptions) the MLE is consistent and asymptotically Gaussian distributed (as $T \to \infty$). For more details we refer to Chapter 6 of the textbook "Theory of Point Estimation" by E.L. Lehmann and G. Casella (2nd edition, 1998). On the other hand, in the Poisson case we directly have an approximate Gaussian distribution of the MLE, independently of the value of T, provided that the volume v is large enough, see also Exercise 4.1.

(ii) From the QQ plot, see Figure 1 (right), we deduce that the test statistic indeed has approximately a χ^2-distribution with $T - 1 = 9$ degrees of freedom. We only observe slightly heavier tails in the observations, compared to a χ^2-distribution with $T - 1 = 9$ degrees of freedom. By increasing the values for n and v, we get even closer to a χ^2-distribution with $T - 1 = 9$ degrees of freedom.

(iii) We observe that we wrongly reject the null hypothesis H_0 of having a Poisson distribution as claim count distribution in 5.16% of the cases. This corresponds almost perfectly to the chosen significance level (indicating the probability of rejecting H_0 even though it is true) of 5%.

Listing 1: R code for Exercise 4.2 (a).

```
### Function generating the data and applying the chi-squared goodness-of-fit test
chi.squared.test.1 <- function(seed1, n, t, lambda, v, alpha){

  ### Generate the claim counts
  set.seed(seed1)
  claim.counts <- array(rpois(n*t,lambda*v), dim=c(t,n))

  ### Distribution of the MLEs of lambda
  lambda_MLE <- colSums(claim.counts)/(t*v)
  plot(density(lambda_MLE), main="Distribution of the MLEs", xlab="Values of the MLEs",
       cex.lab=1.25, cex.main=1.25, cex.axis=1.25)
  abline(v=mean(lambda_MLE), col="red")
  legend("topleft", lty=1, col="red", legend="mean")
  print("1: See plot for the distribution of the MLEs")

  ### Distribution of the test statistic
  lambda_MLE_array <- array(rep(lambda_MLE,each=t), dim=c(t,n))
  test.statistic <- colSums(v*(claim.counts/v-lambda_MLE_array)^2/lambda_MLE_array)
  theoretical.quantiles <- qchisq(p=(1:n)/(n+1), df=t-1)
  empirical.quantiles <- test.statistic[order(test.statistic)]
  lim <- c(min(theoretical.quantiles,empirical.quantiles),
           max(theoretical.quantiles,empirical.quantiles))
  plot(theoretical.quantiles, empirical.quantiles, xlim=lim, ylim=lim,
       xlab="Theoretical Quantiles", ylab="Empirical Quantiles", main="QQ plot", cex.lab=1.25,
       cex.main=1.25, cex.axis=1.25)
  abline(a=0, b=1, col="red")
  print("2: See the QQ plot for a comparison between the empirical quantiles of the test
         statistic and the theoretical quantiles of a chi-squared distribution with t-1
         degrees of freedom")

  ### Result of the hypothesis test
  print(paste("3: How often we wrongly reject the null hypothesis: ",
              sum(test.statistic > qchisq(p=1-alpha, df=t-1))/n,sep=""))
}

### Apply the function with the desired parameters
chi.squared.test.1(seed1=100, n=10000, t=10, lambda=0.1, v=10000, alpha=0.05)
```

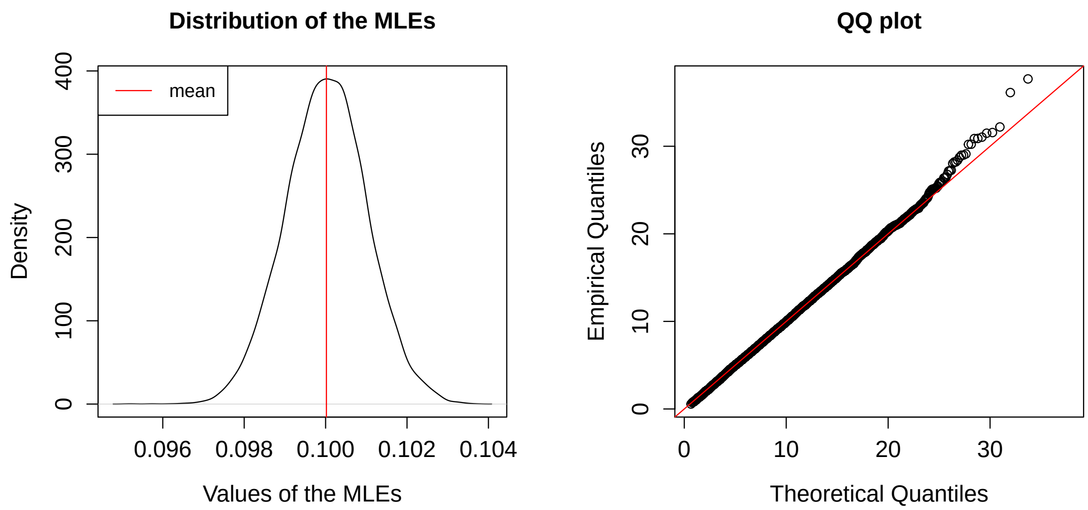

Figure 1: Left: Density plot of the distribution of the MLEs. Right: QQ plot of the theoretical quantiles of a χ^2-distribution with $T-1=9$ degrees of freedom against the empirical quantiles of the values of the test statistic.

(b) The R code used in part (b) is provided in Listing 2.

(i) We observe the following results:

dispersion parameter γ	100	1'000	10'000
Percentage with which we reject H_0	99.78%	48.38%	7.96%

Table 2: Percentage with which we reject H_0 for different values of γ.

(ii) We see that in case of a negative binomial distribution with a comparably small parameter ($\gamma = 100$) for the latent gamma distribution we are almost always able to reject the null hypothesis H_0 of having a Poisson distribution as claim count distribution. The bigger γ, the less we are able to reject H_0. This is because for very large values of γ, the corresponding gamma distribution does not vary a lot, i.e. is almost constantly equal to 1. Thus, for increasing γ, we move back to the Poisson model and, consequently, the χ^2-goodness-of-fit test does not detect the latent variable anymore.

Solution 4.3 Claim Count Distribution

The sample mean $\widehat{\mu}$ and the sample variance $\widehat{\sigma}^2$ of the observed numbers of claims $N_1, \dots, N_{10}$ are given by

$$\widehat{\mu} = \frac{1}{10}\sum_{t=1}^{10} N_t = 21.3 \qquad \text{and} \qquad \widehat{\sigma}^2 = \frac{1}{9}\sum_{t=1}^{10}(N_t - \widehat{\mu})^2 \approx 109.1.$$

We have

$$\widehat{\sigma}^2 \approx 5\widehat{\mu},$$

which suggests $\mathrm{Var}(N_1) > \mathbb{E}[N_1]$. In such a case we would choose a negative binomial distribution, as it allows the variance to exceed the expectation.

Listing 2: R code for Exercise 4.2 (b).

```
### Function generating the data and applying the chi-squared goodness-of-fit test
chi.squared.test.2 <- function(seed1, n, t, lambda, v, alpha, gamma){

  ### Generate the claim counts
  set.seed(seed1)
  claim.counts <- array(rnbinom(n*t, size=gamma, mu=lambda*v), dim=c(t,n))

  ### Calculate the MLEs
  lambda_MLE <- colSums(claim.counts)/(t*v)

  ### Calculate the test statistic
  lambda_MLE_array <- array(rep(lambda_MLE,each=t), dim=c(t,n))
  test.statistic <- colSums(v*(claim.counts/v-lambda_MLE_array)^2/lambda_MLE_array)

  ### Result of the hypothesis test
  print(paste("How often we correctly reject the null hypothesis: ",
              sum(test.statistic > qchisq(p=1-alpha, df=t-1))/n,sep=""))
}

### Apply the function with the desired parameters
chi.squared.test.2(seed1=100, n=10000, t=10, lambda=0.1, v=10000, alpha=0.05, gamma=100)
chi.squared.test.2(seed1=100, n=10000, t=10, lambda=0.1, v=10000, alpha=0.05, gamma=1000)
chi.squared.test.2(seed1=100, n=10000, t=10, lambda=0.1, v=10000, alpha=0.05, gamma=10000)
```

Solution 4.4 Method of Moments

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

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

The sample mean $\widehat{\mu}_8$ and the sample variance $\widehat{\sigma}_8^2$ of the eight observations $y_1, \ldots, y_8$ are given by

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

The method of moments estimates $(\widehat{\gamma}, \widehat{c})$ of (γ, c) solve the equations

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

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

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

which is equivalent to

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

Moreover, we get

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

We conclude that the method of moments estimates are given by $(\widehat{\gamma}, \widehat{c}) = (16, 2)$.

Non-Life Insurance: Mathematics and Statistics

Solution sheet 5

Solution 5.1 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

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

The MLE $\widehat{\alpha}_n^{\mathrm{MLE}}$ is defined as

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

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

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

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

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

(b) Let $\widehat{\alpha}$ denote the unbiased version of the MLE for the storm and flood data given in Table 1 of the exercise sheet. Since we observed 15 storm and flood events, we have $n = 15$. Thus, $\widehat{\alpha}$ can be calculated as

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

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

(c) We define $N_1, \ldots, N_{20}$ to be the numbers of yearly storm and flood events for the twenty years 1986 – 2005. By assumption, $N_1, \ldots, N_{20}$ are i.i.d. Poisson distributed with frequency parameter λ. Using Estimator 2.32 of the lecture notes (version of March 20, 2019) with $v_1 = \cdots = v_{20} = 1$, the MLE $\widehat{\lambda}$ of λ is given by

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

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

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

(d) Using Proposition 2.11 of the lecture notes (version of March 20, 2019), the expected yearly claim amount $\mathbb{E}[S]$ of storm and flood events with maximal claims cover M 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

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

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

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

and

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

Hence, we get

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

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

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

(e) According to our compound Poisson model, next year's total yearly claim amount $S \sim \mathrm{CompPoi}(\lambda, G)$ of storm and flood events with claim amounts exceeding CHF 50 million can be written 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 CHF $M = 2$ billion, 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 (version of March 20, 2019), we have $S_M \sim \text{CompPoi}(\lambda_M, G_M)$ for some distribution function G_M and

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

Defining a random variable $N_M \sim \text{Poi}(\lambda_M)$, the probability that we observe at least one storm and flood event next year which exceeds the level of CHF $M = 2$ billion 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

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

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

Solution 5.2 Claim Size Distributions

The R code used to generate the four i.i.d. samples is given in Listing 1.

Listing 1: R code for Exercise 5.2 (Data generation).

```
### Size of the i.i.d. samples
n <- 10000

### Generate the gamma i.i.d. sample
gamma <- 1/4
c <- 1/40000
set.seed(100)
gamma.sample <- rgamma(n=n, shape=gamma, rate=c)

### Generate the Weibull i.i.d. sample
tau <- 0.54
c <- 0.000175
set.seed(200)
weibull.sample <- rgamma(n=n, shape=1, rate=1)^(1/tau)/c

### Generate the log-normal i.i.d. sample
mu <- log(2000*sqrt(5))
sigma.squared <- log(5)
set.seed(300)
lognormal.sample <- exp(rnorm(n=n, mean=mu, sd=sqrt(sigma.squared)))

### Generate the Pareto i.i.d. sample
theta <- 10000*(sqrt(5)/(2+sqrt(5)))
alpha <- 1+sqrt(5)/2
set.seed(400)
pareto.sample <- theta*exp(rgamma(n=n, shape=1, rate=alpha))
```

In Figure 1 (generated by the R code given in Listing 2) we show the densities (left) of the generated i.i.d. samples as well as the corresponding box plots (right), both on a log scale. We only display logarithmic values starting from 0. We see for example that we have a lot of very small values in case of the gamma distribution and the Weibull distribution. The smallest values observed are considerably bigger for the log-normal and especially the Pareto distribution. Moreover, the value of the biggest value observed increases in going from the gamma over the Weibull and the log-normal to the Pareto distribution. We cannot say much about the tails from looking at these two plots.

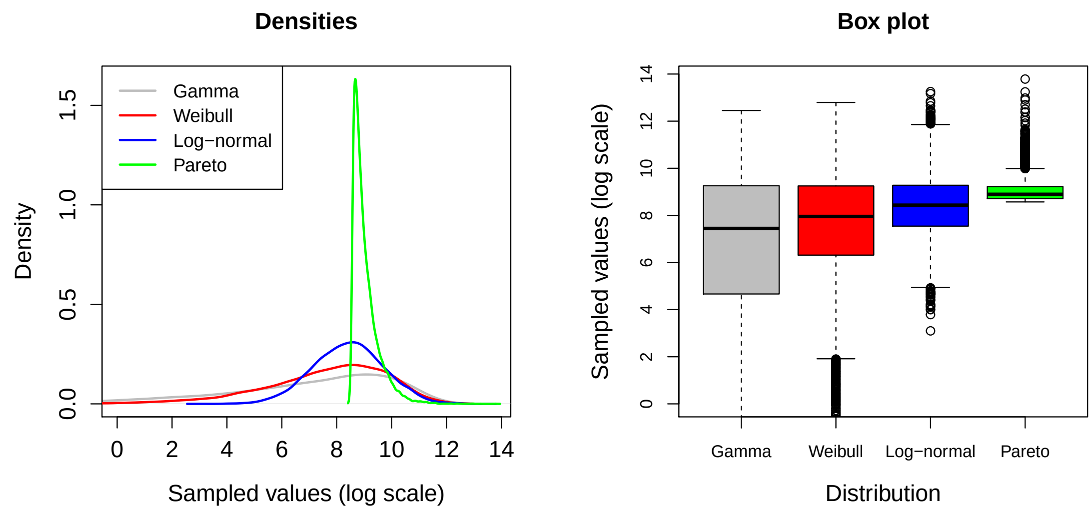

Figure 1: Plot of the densities of the four i.i.d. samples (left). Box plots of the four i.i.d. samples (right).

Listing 2: R code for Exercise 5.2 (Figure 1).

```
### Densities
ymax <- max(density(log(gamma.sample))$y, density(log(weibull.sample))$y,
            density(log(lognormal.sample))$y, density(log(pareto.sample))$y)
ymax2 <- max(log(gamma.sample),log(weibull.sample),log(lognormal.sample),log(pareto.sample))
plot(density(log(gamma.sample)), xlim=c(0,ymax2), col="grey", ylim=c(0,ymax), main="Densities",
             xlab="Sampled values (log scale)", cex.lab=1.25, cex.main=1.25, cex.axis=1.25,
             lwd=2)
lines(density(log(weibull.sample)), col="red", xlim=c(0,ymax2), lwd=2)
lines(density(log(lognormal.sample)), col="blue", xlim=c(0,ymax2), lwd=2)
lines(density(log(pareto.sample)), col="green", xlim=c(0,ymax2), lwd=2)
legend("topleft", lty=1, lwd=2, col=c("grey","red","blue","green"),
       legend=c("Gamma","Weibull","Log-normal","Pareto"))

### Box plots
boxplot(log(gamma.sample), log(weibull.sample), log(lognormal.sample), log(pareto.sample),
        ylim=c(0,ymax2), col=c("grey","red","blue","green"), main="Box plot",
        names=c("Gamma","Weibull","Log-normal","Pareto"), xlab="Distribution",
        ylab="Sampled values (log scale)", cex.lab=1.25, cex.main=1.25, cex.axis=0.95)
```

In Figure 2 (generated by the R code given in Listing 3) we show the plots of the empirical distribution functions (left, on a log scale) and of the empirical loss size index functions (right) of the generated i.i.d. samples. For the plot of the empirical distribution functions we only display logarithmic values starting from 0. We observe that the empirical distribution functions almost perfectly intersect at the point with x-coordinate equal to log(10'000)$\approx$ 9.21. This means that for all of the four considered distributions approximately the same percentage of observations is smaller than the expected value. This percentage is roughly equal to 75%, indicating that three quarters of the observations are smaller than the expected value and one quarter of the observations are above the expected value. Thus, not surprisingly, the large claims are the main driver of the expected value. We get confirmed the observations from Figure 1, namely that the smallest values observed are considerably bigger for the log-normal and especially the Pareto distribution, compared to the gamma and the Weibull distribution. This carries over to the plot of the empirical loss size index function. Also these two plots do not tell us much about the tails of the distributions.

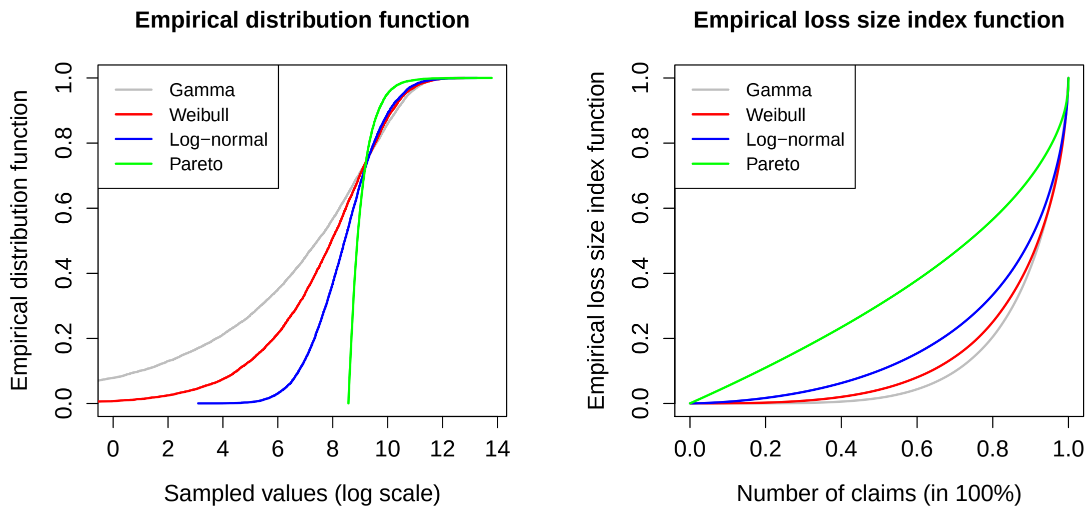

Figure 2: Plot of the empirical distribution functions of the four i.i.d. samples (left). Plot of the empirical loss size index functions of the four i.i.d. samples (right).

Listing 3: R code for Exercise 5.2 (Figure 2).

```
### Empirical distribution functions
plot(log(gamma.sample[order(gamma.sample)]), 1:n/(n+1), xlim=c(0,ymax2), type="l", col="grey",
     main="Empirical distribution function", xlab="Sampled values (log scale)",
     ylab="Empirical distribution function", cex.lab=1.25, cex.main=1.25, cex.axis=1.25, lwd=2)
lines(log(weibull.sample[order(weibull.sample)]), 1:n/(n+1), xlim=c(0,ymax2), col="red", lwd=2)
lines(log(lognormal.sample[order(lognormal.sample)]), 1:n/(n+1), xlim=c(0,ymax2), col="blue",
      lwd=2)
lines(log(pareto.sample[order(pareto.sample)]), 1:n/(n+1), xlim=c(0,ymax2), col="green", lwd=2)
legend("topleft", lty=1, lwd=2, col=c("grey","red","blue","green"),
       legend=c("Gamma","Weibull","Log-normal","Pareto"))

### Empirical loss size index functions
plot(1:n/n, cumsum(gamma.sample[order(gamma.sample)])/sum(gamma.sample), type="l", col="grey",
     main="Empirical loss size index function", xlab="Number of claims (in 100%)",
     ylab="Empirical loss size index function", cex.lab=1.25, cex.main=1.25, cex.axis=1.25,
     lwd=2)
lines(1:n/n, cumsum(weibull.sample[order(weibull.sample)])/sum(weibull.sample), type="l",
      col="red", lwd=2)
lines(1:n/n, cumsum(lognormal.sample[order(lognormal.sample)])/sum(lognormal.sample), type="l",
      col="blue", lwd=2)
lines(1:n/n, cumsum(pareto.sample[order(pareto.sample)])/sum(pareto.sample), type="l",
      col="green", lwd=2)
legend("topleft", lty=1, lwd=2, col=c("grey","red","blue","green"),
       legend=c("Gamma","Weibull","Log-normal","Pareto"))
```

In Figure 3 (generated by the R code given in Listing 4) we show the log-log plots (left) and the plot of the empirical mean excess functions (right) of the generated i.i.d. samples. These two plots can be used for studying the tails of the distributions. We see in both plots that the gamma distribution is the most light-tailed distribution. The Weibull distribution and the log-normal distribution have a similar tail behaviour, with slightly heavier tails of the log-normal distribution. Note that this similar tail behaviour is due to the value of the parameter τ of the Weibull distribution being smaller than 1. With a value $\tau \geq 1$ the distribution gets (even) more light-tailed. The most heavy-tailed distribution among the four distributions we analyzed here is the Pareto distribution.

Listing 4: R code for Exercise 5.2 (Figure 3).

```
### Log-log plots
plot(log(gamma.sample[order(gamma.sample)]), log(1-1:n/(n+1)), xlim=c(0,ymax2), type="l",
     col="grey", main="Log-log plot", xlab="log(sampled values)",
     ylab="log(1-empirical distribution function)", cex.lab=1.25, cex.main=1.25, cex.axis=1.25,
     lwd=2)
lines(log(weibull.sample[order(weibull.sample)]), log(1-1:n/(n+1)), xlim=c(0,ymax2), type="l",
      col="red", lwd=2)
lines(log(lognormal.sample[order(lognormal.sample)]), log(1-1:n/(n+1)), xlim=c(0,ymax2),
      type="l", col="blue", lwd=2)
lines(log(pareto.sample[order(pareto.sample)]), log(1-1:n/(n+1)), xlim=c(0,ymax2), type="l",
      col="green", lwd=2)
legend("bottomleft", lty=1, lwd=2, col=c("grey","red","blue","green"),
       legend=c("Gamma","Weibull","Log-normal","Pareto"))

### Empirical mean excess functions
mean.excess.function <- Vectorize(function(threshold,input.sample){
mean(input.sample[input.sample>threshold])-threshold
},"threshold")
xmax <- pareto.sample[order(pareto.sample)][n-1]
ymax3 <- max(pareto.sample)-xmax
plot(gamma.sample[order(gamma.sample)][-n],
     mean.excess.function(gamma.sample[order(gamma.sample)][-n],gamma.sample), pch=16,
     col="grey", xlim=c(0,xmax), ylim=c(0,ymax3), main="Empirical mean excess function",
     xlab="Threshold", ylab="Mean excess function", cex.lab=1.25, cex.main=1.25, cex.axis=1.25)
points(weibull.sample[order(weibull.sample)][-n],
       mean.excess.function(weibull.sample[order(weibull.sample)][-n],weibull.sample), pch=16,
       col="red", ylim=c(0,ymax3))
points(lognormal.sample[order(lognormal.sample)][-n],
       mean.excess.function(lognormal.sample[order(lognormal.sample)][-n],lognormal.sample),
       pch=16, col="blue", ylim=c(0,ymax3))
points(pareto.sample[order(pareto.sample)][-n],
       mean.excess.function(pareto.sample[order(pareto.sample)][-n],pareto.sample),pch=16,
       col="green", ylim=c(0,ymax3))
legend("topleft", pch=16, col=c("grey","red","blue","green"),
       legend=c("Gamma","Weibull","Log-normal","Pareto"))
```

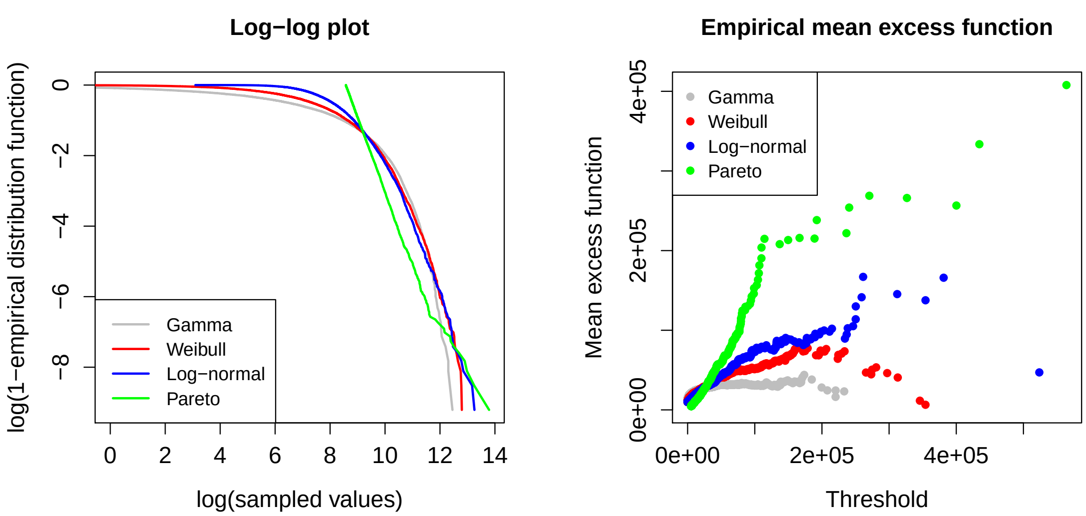

Figure 3: Log-log plots of the four i.i.d. samples (left). Plot of the empirical mean excess functions of the four i.i.d. samples (right).

Summarizing, we can say that although we fixed the mean and the standard deviation to be the same, all of the four considered distributions behave differently. This implies that one has to carefully select the appropriate claim size distribution.

Solution 5.3 Hill Estimator

The Hill plot (on the left, generated by the R code of Listing 5) and the log-log plot (on the right, generated by the R code of Listing 6) are given in Figure 4. 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 $\theta = 10$ and tail index $\alpha = 2$ would suggest. We observe these problems even though we sampled directly from a Pareto distribution. This indicates the difficulties one faces when trying to fit such a distribution to a real data set, where we often have less observations than in this example and the observations may be contaminated by other distributions.

Listing 5: R code for Exercise 5.3 (Hill plot).

```
hill.plot.function <- function(n, theta, alpha, seed1){
  set.seed(seed1)
  data.1 <- rgamma(n, shape=1, scale=1/alpha)
  data <- theta * exp(data.1)
  log.data.ordered <- log(data[order(data, decreasing=FALSE)])
  n.obs <- n:5
  hill.estimator <- ((sum(log.data.ordered)-cumsum(log.data.ordered)
                     +log.data.ordered)[-((n-3):n)]/n.obs-log.data.ordered[-((n-3):n)])^(-1)
  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)
  plot(hill.estimator, ylim=c(min(hill.estimator)-1,max(hill.estimator)+1), xaxt="n",
       xlab="Number of observations", ylab="Pareto tail index parameter",
       main="Hill plot for alpha", cex=0.5, cex.lab=1.25, cex.main=1.25, cex.axis=1.25)
  axis(1, at=c(1,seq(from=n/10+1, to=n*9/10+1, by=n/10), n-5), c(seq(from=n, to=n/10,
       by=-n/10), 5))
  lines(upper.bound)
  lines(lower.bound)
  abline(h=alpha, col="blue", lwd=2)
  legend("topleft", col=c("blue","black"), lty=c(1,NA), pch=c(NA,1), lwd=c(2,NA),
         legend=c("true tail index","estimated tail index"))
}

hill.plot.function(n=300,theta=10,alpha=2,seed1=100)
```

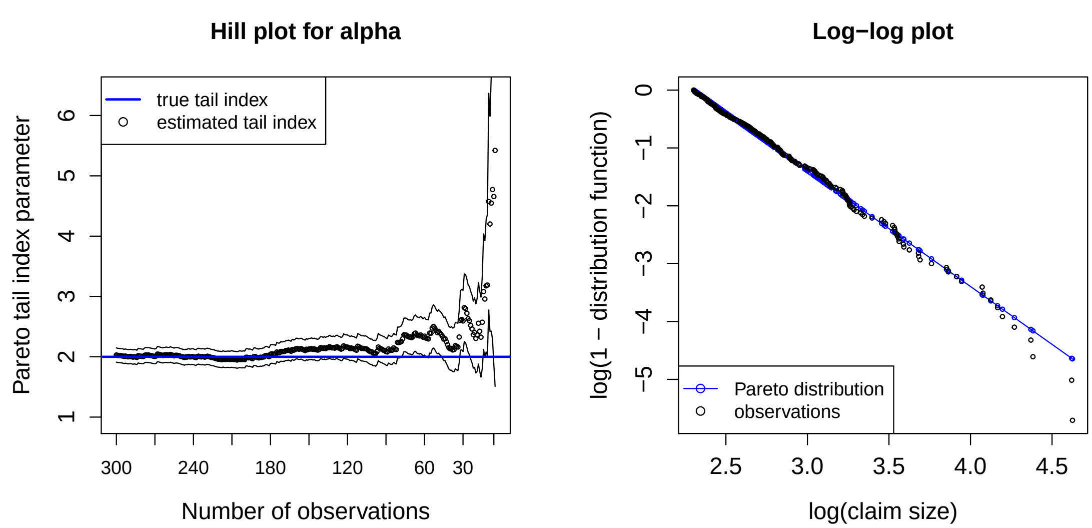

Figure 4: Hill plot for determining the tail index α (left). Log-log plot for the observations and the Pareto distribution (right).

Listing 6: R code for Exercise 5.3 (Log-log plot).

```
log.log.plot.function <- function(n, theta, alpha, seed1){
  set.seed(seed1)
  data.1 <- rgamma(n, shape=1, scale=1/alpha)
  data <- theta*exp(data.1)
  data.ordered <- data[order(data, decreasing=FALSE)]
  log.data.ordered <- log(data.ordered)
  true.sf <- (data.ordered/theta)^(-alpha)
  empirical.sf <- 1-(1:n)/(n+1)
  plot(log.data.ordered, log(true.sf), xlab="log(claim size)",
       ylab="log(1 - distribution function)",
       ylim=c(min(log(true.sf),log(empirical.sf)),max(log(true.sf),log(empirical.sf))),
       main="Log-log plot", cex.lab=1.25, cex.main=1.25, cex.axis=1.25, cex=0.5, col="blue")
  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"))
}

log.log.plot.function(n=300,theta=10,alpha=2,seed1=100)
```

Solution 5.4 Pareto Distribution

The density g and the distribution function G of $Y \sim \text{Pareto}(\theta, \alpha)$ are defined by

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

respectively, for all $x \geq \theta$.

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

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

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 $\theta \leq u_1 < u_2$. Then, the expected value of Y within the layer $(u_1, u_2]$ can be calculated as

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

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

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

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

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

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

$$e(u) = \mathbb{E}[Y - u | Y > u] = \mathbb{E}[Y | Y > u] - u = \frac{\mathbb{E}[Y 1_{\{Y > u\}}]}{\mathbb{P}[Y > u]} - u = \frac{\mathbb{E}[Y 1_{\{Y > u\}}]}{\bar{G}(u)} - u,$$

where for $\mathbb{E}[Y 1_{\{Y > u\}}]$ we have, similarly as in part (b),

$$\begin{aligned} \mathbb{E}[Y 1_{\{Y > u\}}] &= \mathbb{E}[Y 1_{\{u < Y \leq \infty\}}] = \alpha\theta \left[-\frac{1}{\alpha - 1} \left(\frac{x}{\theta}\right)^{-\alpha+1} \right]_u^\infty = \frac{\alpha}{\alpha - 1} \theta \left(\frac{u}{\theta}\right)^{-\alpha+1} \\ &= \frac{\alpha}{\alpha - 1} u \bar{G}(u). \end{aligned}$$

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

(a) 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

$$\mathbb{P}[Y \in I_2] \,=\, \mathbb{P}[239 \leq Y < 301] \,=\, G(301) - G(239) \,=\, 0.2.$$

By analogous calculations for the other four intervals, we get

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

Let O_k denote the actual number of observations and E_k the expected number of observations in interval I_k, for all $k \in \{1, \ldots, 5\}$. The test statistic

$$X^2_{n,5} \,=\, \sum_{k=1}^{5} \frac{(O_k - E_k)^2}{E_k}$$

of the χ^2-goodness-of-fit test using $K = 5$ intervals and n observations converges to a χ^2-distribution with $K - 1 = 5 - 1 = 4$ degrees of freedom, as $n \to \infty$. As we have $n = 20$ observations in our data, we can calculate E_k as

$$E_k \,=\, 20 \cdot \mathbb{P}[Y \in I_k] \,=\, 20 \cdot 0.2 \,\approx\, 4,$$

for all $k = 1, \ldots, 5$. The values of the actual numbers of observations O_k and the expected numbers of observations E_k in the five intervals $k = 1, \ldots, 5$ as well as their squared differences $(O_k - E_k)^2$ are summarized in Table 1.

k	1	2	3	4	5
O_k	4	0	8	6	2
E_k	4	4	4	4	4
$(O_k - E_k)^2$	0	16	16	4	4

Table 1: Actual and expected numbers of observations with squared differences.

With the numbers in Table 1, the test statistic of the χ^2-goodness-of-fit test using 5 intervals in the case of our $n = 20$ observations is given by

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

Let $\alpha = 5\%$. Then, the $(1 - \alpha)$-quantile of the χ^2-distribution with 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 significance level of 5%.

(b) We assume that we have n i.i.d. observations $Y_1, \dots, Y_n$ from the null hypothesis distribution and that we work with $K = 2$ disjoint intervals I_1 and I_2. We define

$$p = \mathbb{P}[Y_1 \in I_1]$$

and

$$X_i = 1_{\{Y_i \in I_1\}},$$

for all $i = 1, \dots, n$. This implies that $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$. Thus, we have

$$\mu \overset{\text{def}}{=} \mathbb{E}[X_1] = p \qquad \text{and} \qquad \sigma \overset{\text{def}}{=} \sqrt{\text{Var}(X_1)} = \sqrt{p(1-p)}.$$

Moreover, we can write

$$O_1 = \sum_{i=1}^n X_i \qquad \text{and} \qquad O_2 = n - O_1 = n - \sum_{i=1}^n X_i$$

as well as

$$E_1 = \mathbb{E}\left[\sum_{i=1}^n X_i\right] = np \qquad \text{and} \qquad E_2 = \mathbb{E}\left[n - \sum_{i=1}^n X_i\right] = n - np = n(1-p).$$

Therefore, we get

$$\begin{aligned} X^2_{n,2} &= \sum_{k=1}^2 \frac{(O_k - E_k)^2}{E_k} = \frac{(O_1 - np)^2}{np} + \frac{[n - O_1 - n(1-p)]^2}{n(1-p)} \\ &= (O_1 - np)^2 \left[\frac{1}{np} + \frac{1}{n(1-p)}\right] = (O_1 - np)^2 \frac{1}{np(1-p)} \\ &= \left(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma}\right)^2. \end{aligned}$$

Let $Z \sim \mathcal{N}(0,1)$ and χ^2_1 follow a χ^2-square distribution with one degree of freedom. According to the central limit theorem, see equation (1.2) of the lecture notes (version of March 20, 2019), we have

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \Longrightarrow Z, \quad \text{as } n \to \infty.$$

As $Z^2 \overset{(d)}{=} \chi^2_1$, see Exercise 1.4, we can conclude that

$$X^2_{n,2} \Longrightarrow Z^2 \overset{(d)}{=} \chi^2_1, \quad \text{as } n \to \infty.$$

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 $r \in \mathbb{R}$, see Exercise 1.3. Since Y_1 has a log-normal distribution with mean parameter μ and variance parameter σ^2, we have

$$Y_1 \overset{(d)}{=} \exp\{X\}.$$

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

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

(b) From part (a) we know that

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

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

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

(i) The claim 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

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

Hence, the new claim 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 we expect that 41% of the 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 page 67 of the lecture notes (version of March 20, 2019), $e(d)$ 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 (version of March 20, 2019), 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

$$\begin{aligned} \mathbb{E}[S^{\text{new}}] \;&=\; \mathbb{E}[N^{\text{new}}]\mathbb{E}[Y_1^{\text{new}}] \;=\; \mathbb{E}[N]\mathbb{P}[Y_1 > d]e(d) \;\approx\; 0.59 \cdot \mathbb{E}[N] \cdot 1.49 \cdot \mathbb{E}[Y_1] \\ &\approx\; 0.87 \cdot \mathbb{E}[S]. \end{aligned}$$

In particular, the insurance company can grant a discount of roughly 13% 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 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

$$G_0(y) \;=\; 1 - \exp\left\{-y^{1/2}\right\},$$

for all $y \geq 0$. Since G_0 is continuous, we are indeed 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 evaluate G_0 at our data points $x_1, \ldots, x_5$, we get

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

We write $\widehat{G}_n$ for the empirical distribution function of a sample with n data points. The Kolmogorov-Smirnov test statistic D_n is then defined as

$$D_n = \sup_{y \in \mathbb{R}} \left| \widehat{G}_n(y) - G_0(y) \right|,$$

and $\sqrt{n} D_n$ converges to the Kolmogorov distribution K, as $n \to \infty$. The empirical distribution function $\widehat{G}_5$ of the sample $x_1, \ldots, x_5$ is given by

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

Since G_0 is continuous and strictly increasing with range $[0, 1)$ and $\widehat{G}_5$ is piecewise constant and attains both the values 0 and 1, it is sufficient to consider the discontinuities of $\widehat{G}_5$ to determine the Kolmogorov-Smirnov test statistic D_5 for our $n = 5$ data points. 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 $\widehat{G}_5$. 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 $\widehat{G}_5$ and their differences (in absolute value) are summarized in Table 2.

x_i, x_i-	x_1-	x_1	x_2-	x_2	x_3-	x_3	x_4-	x_4	x_5-	x_5
$\widehat{G}_5(\cdot)$	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
$\lvert\widehat{G}_5(\cdot) - G_0(\cdot)\rvert$	2/40	6/40	5/40	13/40	11/40	19/40	18/40	26/40	2/40	10/40

Table 2: Values of G_0 and $\widehat{G}_5$ and their differences (in absolute value).

From Table 2 we see for the Kolmogorov-Smirnov test statistic D_5 that

$$D_5 = \sup_{y \in \mathbb{R}} \left| \widehat{G}_5(y) - G_0(y) \right| = 26/40 = 0.65.$$

Let $q = 5\%$. By writing $K^{\leftarrow}(1-q)$ for the $(1-q)$-quantile of the Kolmogorov distribution, we have $K^{\leftarrow}(1-q) = 1.36$, see page 81 of the lecture notes (version of March 20, 2019). Since

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

we can reject the null hypothesis (at significance level of 5%) that the data $x_1, \ldots, x_5$ comes from a Weibull distribution with shape parameter $\tau = \frac{1}{2}$ and scale parameter $c = 1$.

Solution 6.4 Akaike Information Criterion and Bayesian Information Criterion

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

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

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

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

$$\begin{aligned} \text{AIC}^{\text{MM}} &= -2\ell_{\mathbf{Y}}\left(\widehat{\gamma}^{\text{MM}}, \widehat{c}^{\text{MM}}\right) + 2d^{\text{MM}} = -2 \cdot 1'264.013 + 2 \cdot 2 = -2'524.026 \quad \text{and} \\ \text{AIC}^{\text{MLE}} &= -2\ell_{\mathbf{Y}}\left(\widehat{\gamma}^{\text{MLE}}, \widehat{c}^{\text{MLE}}\right) + 2d^{\text{MLE}} = -2 \cdot 1'264.171 + 2 \cdot 2 = -2'524.342. \end{aligned}$$

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

(b) If we write d^{gam} and d^{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 AIC^{gam} of the gamma model and the AIC value AIC^{exp} of the exponential model are then given by

$$\begin{aligned} \text{AIC}^{\text{gam}} &= -2\ell_{\mathbf{Y}}^{\text{gam}}\left(\widehat{\gamma}^{\text{MLE}}, \widehat{c}^{\text{MLE}}\right) + 2d^{\text{gam}} = -2 \cdot 1'264.171 + 2 \cdot 2 = -2'524.342 \quad \text{and} \\ \text{AIC}^{\text{exp}} &= -2\ell_{\mathbf{Y}}^{\text{exp}}\left(\widehat{c}^{\text{MLE}}\right) + 2d^{\text{exp}} = -2 \cdot 1'264.169 + 2 \cdot 1 = -2'526.338. \end{aligned}$$

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

The BIC value BIC^{gam} of the gamma model and the BIC value BIC^{exp} of the exponential model are given by

$$\text{BIC}^{\text{gam}} = -2\ell_{\mathbf{Y}}^{\text{gam}}\left(\widehat{\gamma}^{\text{MLE}}, \widehat{c}^{\text{MLE}}\right) + d^{\text{gam}} \cdot \log n = -2 \cdot 1'264.171 + 2 \cdot \log 1'000 \approx -2'514.53$$

and

$$\text{BIC}^{\text{exp}} = -2\ell_{\mathbf{Y}}^{\text{exp}}\left(\widehat{c}^{\text{MLE}}\right) + d^{\text{exp}} \cdot \log n = -2 \cdot 1'264.169 + \log 1'000 \approx -2'521.43.$$

According to the BIC, the model with the smallest BIC value should be preferred. Since $\text{BIC}^{\text{gam}} > \text{BIC}^{\text{exp}}$, 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 7

Solution 7.1 Re-Insurance Covers and Leverage Effect

(a) Under the assumption that $\mathbb{P}[Y > d] > 0$ and that $\mathbb{E}[Y|Y > d]$ exists, we can generally write

$$\begin{aligned}\mathbb{E}[(Y-d)_+] &= \mathbb{E}[(Y-d)1_{\{Y>d\}}] = \mathbb{E}[Y1_{\{Y>d\}}] - \mathbb{E}[d1_{\{Y>d\}}] \\ &= \frac{\mathbb{E}[Y1_{\{Y>d\}}]}{\mathbb{P}[Y>d]}\mathbb{P}[Y>d] - d\,\mathbb{P}[Y>d] = \mathbb{P}[Y>d](\mathbb{E}[Y|Y>d]-d),\end{aligned}$$

see also formula (3.11) of the lecture notes (version of March 20, 2019). Now we explicitly use that a gamma distribution with shape parameter equal to 1 is an exponential distribution. The characteristic property of an exponential distribution is the so-called memorylessness property

$$\mathbb{P}[Y > t+s|Y>t] = \mathbb{P}[Y>s],$$

for all $t, s > 0$. In particular, this property leads to (see below for the calculation)

$$\mathbb{E}[Y|Y>d] = \mathbb{E}[Y] + d, \tag{1}$$

which, in turn, implies for our loss $Y \sim \Gamma(1, \frac{1}{400})$ that

$$\mathbb{E}[(Y-d)_+] = \mathbb{P}[Y>d]\mathbb{E}[Y].$$

We check equation (1). Indeed, we have

$$\begin{aligned}\mathbb{E}[Y|Y>d] &= \frac{\mathbb{E}[Y1_{\{Y>d\}}]}{\mathbb{P}[Y>d]} = \frac{1}{\mathbb{P}[Y>d]}\int_0^\infty y1_{\{y>d\}}\frac{1}{400}\exp\left\{-\frac{y}{400}\right\}dy \\ &= \frac{1}{\exp\left\{-\frac{d}{400}\right\}}\int_d^\infty y\frac{1}{400}\exp\left\{-\frac{y}{400}\right\}dy \\ &= \exp\left\{\frac{d}{400}\right\}\int_0^\infty (u+d)\frac{1}{400}\exp\left\{-\frac{u}{400}\right\}\exp\left\{-\frac{d}{400}\right\}du \\ &= \int_0^\infty u\frac{1}{400}\exp\left\{-\frac{u}{400}\right\}du + d\int_0^\infty \frac{1}{400}\exp\left\{-\frac{u}{400}\right\}du \\ &= \mathbb{E}[Y]+d,\end{aligned}$$

where in the fourth equality we used the substitution $u = y - d$.

(b) By looking at the graphs in Figure 1 on the exercise sheet, we find the following re-insurance covers:

(i) $(Y-200)_+$,

(ii) $\min\{Y, 400\}$,

(iii) $\min\{Y, 200\} + (Y-400)_+$.

(c) (i) Using part (a), we get

$$\mathbb{E}[(Y-200)_+] = \mathbb{P}[Y>200]\mathbb{E}[Y] = \exp\left\{-\frac{200}{400}\right\}400 = \frac{400}{\sqrt{\exp\{1\}}} \approx 243.$$

(ii) First, we write

$$\begin{aligned}\mathbb{E}[\min\{Y,400\}] &= \mathbb{E}[\min\{Y,400\}1_{\{Y\le 400\}}] + \mathbb{E}[\min\{Y,400\}1_{\{Y>400\}}]\\ &= \mathbb{E}[Y1_{\{Y\le 400\}}] + \mathbb{E}[400\cdot 1_{\{Y>400\}}]\\ &= \mathbb{E}[Y] - \mathbb{E}[Y1_{\{Y>400\}}] + \mathbb{E}[400\cdot 1_{\{Y>400\}}]\\ &= \mathbb{E}[Y] - \mathbb{E}[(Y-400)1_{\{Y>400\}}]\\ &= \mathbb{E}[Y] - \mathbb{E}[(Y-400)_+],\end{aligned}$$

which holds true as $\mathbb{E}[Y]$ exists, see also page 89 the lecture notes (version of March 20, 2019). Using part (a), we then get

$$\begin{aligned}\mathbb{E}[\min\{Y,400\}] &= \mathbb{E}[Y] - \mathbb{P}[Y>400]\mathbb{E}[Y] = \mathbb{E}[Y](1-\mathbb{P}[Y>400])\\ &= 400\left(1-\exp\left\{-\frac{400}{400}\right\}\right) = 400(1-\exp\{-1\}) \approx 253.\end{aligned}$$

(iii) Using the above calculation in (ii) as well as part (a), we have

$$\begin{aligned}\mathbb{E}[\min\{Y,200\}+(Y-400)_+] &= \mathbb{E}[Y] - \mathbb{E}[(Y-200)_+] + \mathbb{E}[(Y-400)_+]\\ &= \mathbb{E}[Y] - \mathbb{P}[Y>200]\mathbb{E}[Y] + \mathbb{P}[Y>400]\mathbb{E}[Y]\\ &= \mathbb{E}[Y](1-\mathbb{P}[Y>200]+\mathbb{P}[Y>400])\\ &= 400\left(1-\exp\left\{-\frac{1}{2}\right\}+\exp\{-1\}\right)\\ &\approx 305.\end{aligned}$$

(d) As $Y_0 \sim \Gamma\left(1, \frac{1}{400}\right)$, formula (3.5) of the lecture notes (version of March 20, 2019) implies

$$Y_1 \overset{(d)}{=} (1+i)Y_0 \sim \Gamma\left(1, \frac{1}{400(1+i)}\right).$$

Using part (a), we get

$$\mathbb{E}[(Y_1-d)_+] = \mathbb{P}[Y_1>d]\mathbb{E}[Y_1] = \exp\left\{-\frac{d}{400(1+i)}\right\}400(1+i)$$

and

$$\mathbb{E}[(Y_0-d)_+] = \mathbb{P}[Y_0>d]\mathbb{E}[Y_0] = \exp\left\{-\frac{d}{400}\right\}400,$$

which leads to

$$\frac{\mathbb{E}[(Y_1-d)_+]}{(1+i)\mathbb{E}[(Y_0-d)_+]} = \frac{\exp\left\{-\frac{d}{400(1+i)}\right\}}{\exp\left\{-\frac{d}{400}\right\}} = \exp\left\{\frac{d}{400}\left(1-\frac{1}{1+i}\right)\right\} > 1,$$

since $i>0$. We conclude that

$$\mathbb{E}[(Y_1-d)_+] > (1+i)\mathbb{E}[(Y_0-d)_+].$$

The reason for this (strict) inequality, which is called leverage effect, is that not only the claim sizes are growing under inflation, but also the number of claims that exceed threshold d increases under inflation, as we do not adapt threshold d to inflation.

Solution 7.2 Inflation and Deductible

Let Y be a random variable following a Pareto distribution with threshold $\theta > 0$ and tail index $\alpha > 1$. Since the insurance company only has to pay the part that exceeds the deductible θ, this year's average claim payment z is

$$z = \mathbb{E}\left[(Y-\theta)_+\right] = \mathbb{E}[Y] - \theta = \frac{\alpha}{\alpha-1}\theta - \theta = \frac{1}{\alpha-1}\theta.$$

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

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

Thus, the mean excess function $e_{\widetilde{Y}}(u)$ of $\widetilde{Y}$ above $u > (1+r)\theta$ is given by

$$e_{\widetilde{Y}}(u) = \frac{1}{\alpha-1}u,$$

see also Exercise 5.4. Let $\rho\theta$ for some $\rho > 0$ denote the increase of the deductible that is needed such that the average claim payment remains unchanged. With the new deductible $(1+\rho)\theta$, next year's average claim payment is given by

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

The goal is to find $\rho > 0$ such that $z = \widetilde{z}$. Assuming $\rho \leq r$, we have

$$\begin{aligned}\widetilde{z} &= \mathbb{E}\left[\left(\widetilde{Y} - [1+\rho]\theta\right)_+\right] \geq \mathbb{E}\left[\left(\widetilde{Y} - [1+r]\theta\right)_+\right] = \mathbb{E}\left[([1+r]Y - [1+r]\theta)_+\right] \\ &= (1+r)\,\mathbb{E}\left[(Y-\theta)_+\right] = (1+r)\,z > z,\end{aligned}$$

i.e. for $\rho \leq r$ it is not possible to get $z = \widetilde{z}$. Hence, we can deduce that $\rho > r$, i.e. the percentage increase in the deductible has to be bigger than the inflation. Assuming $\rho > r$, we can calculate

$$\begin{aligned}\widetilde{z} &= \mathbb{E}\left[\left(\widetilde{Y} - [1+\rho]\theta\right) \cdot 1_{\{\widetilde{Y}-(1+\rho)\theta>0\}}\right] = \mathbb{E}\left[\widetilde{Y} - (1+\rho)\theta \,\middle|\, \widetilde{Y} > (1+\rho)\theta\right] \cdot \mathbb{P}\left[\widetilde{Y} > (1+\rho)\theta\right] \\ &= e_{\widetilde{Y}}([1+\rho]\theta) \cdot \mathbb{P}\left[\widetilde{Y} > (1+\rho)\theta\right] = \frac{1}{\alpha-1}(1+\rho)\theta \cdot \left[\frac{(1+\rho)\theta}{(1+r)\theta}\right]^{-\alpha} \\ &= \frac{1}{\alpha-1}\theta(1+r)^{\alpha}(1+\rho)^{-\alpha+1} = z \cdot (1+r)^{\alpha}(1+\rho)^{-\alpha+1}.\end{aligned}$$

We have

$$z = \widetilde{z} \iff (1+r)^{\alpha}(1+\rho)^{-\alpha+1} = 1 \iff \rho = (1+r)^{\frac{\alpha}{\alpha-1}} - 1 > r.$$

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

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

Solution 7.3 Normal Approximation

As $Y \sim \Gamma(\gamma = 100, c = \frac{1}{10})$, we have

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

For the total claim amount S we can use Proposition 2.11 of the lecture notes (version of March 20, 2019) to get

$$\mathbb{E}[S] = \lambda v \mathbb{E}[Y] = 1'000 \cdot 1'000 = 1'000'000 \quad \text{and}$$
$$\mathrm{Var}(S) = \lambda v \mathbb{E}[Y^2] = 1'000 \cdot 1'010'000 = 1'010'000'000.$$

Let F_S denote the distribution function of S. Then, since F_S is continuous and strictly increasing (above the level $\exp\{-\lambda v\} = \mathbb{P}[S=0]$), the quantiles $q_{0.95}$ and $q_{0.99}$ can be calculated as

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

According to Section 4.1.1 of the lecture notes (version of March 20, 2019), 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 then have

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

In particular, we get

$$q_{0.95} = F_S^{-1}(0.95) \approx 1'000'000 + 31'780.5 \cdot \Phi^{-1}(0.95) \approx 1'000'000 + 31'780.5 \cdot 1.645 \approx 1'052'274$$

and

$$q_{0.99} = F_S^{-1}(0.99) \approx 1'000'000 + 31'780.5 \cdot \Phi^{-1}(0.99) \approx 1'000'000 + 31'780.5 \cdot 2.326 \approx 1'073'932.$$

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

$$F_S(0) \approx \Phi\left(\frac{0 - \lambda v \mathbb{E}[Y]}{\sqrt{\lambda v \mathbb{E}[Y^2]}}\right) \approx \Phi\left(-\frac{1'000'000}{31'780.5}\right) \approx \Phi(-31.5) \approx 1.27 \cdot 10^{-217},$$

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

Solution 7.4 Translated Gamma and Translated Log-Normal Approximation

As $Y \sim \Gamma(\gamma = 100, c = \frac{1}{10})$, we have

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

Let M_Y denote the moment generating function of Y. According to formula (1.3) of the lecture notes (version of March 20, 2019), we have

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

For the total claim amount S we can use Proposition 2.11 of the lecture notes (version of March 20, 2019) to get

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

where M_S denotes the moment generating function of S. In order to get the skewness ς_S of S, we can use the third equation given in formulas (1.5) of the lecture notes (version of March 20, 2019):

$$\varsigma_S \cdot \operatorname{Var}(S)^{3/2} = \frac{d^3}{dr^3} \log M_S(r)\Big|_{r=0} = \lambda v \frac{d^3}{dr^3}[M_Y(r) - 1]\Big|_{r=0} = \lambda v M_Y'''(0) = \lambda v \mathbb{E}[Y^3],$$

from which we can conclude that

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

Let F_S denote the distribution function of S. Then, since F_S is continuous and strictly increasing (above the level $\exp\{-\lambda v\} = \mathbb{P}[S = 0]$), 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 4.1.2 of the lecture notes (version of March 20, 2019), in the translated gamma approximation we model S by the random variable

$$X = k + Z,$$

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

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

where ς_X is the skewness parameter of X. Since $Z \sim \Gamma(\widetilde{\gamma}, \widetilde{c})$, we can use the results given in Section 3.2.1 of the lecture notes (version of March 20, 2019) to calculate

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

Using the equations given in (2), we get

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

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

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

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

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

and

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

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

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

which is basically 0.

(b) According to Section 4.1.2 of the lecture notes (version of March 20, 2019), 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

$$\mathbb{E}[X] = \mathbb{E}[S], \qquad \mathrm{Var}(X) = \mathrm{Var}(S) \qquad \text{and} \qquad \varsigma_X = \varsigma_S. \tag{3}$$

Since $Z \sim \mathrm{LN}(\mu, \sigma^2)$, we can use the results given in Section 3.2.3 of the lecture notes (version of March 20, 2019) 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 (3), we get

$$\left(\exp\left\{\sigma^2\right\} + 2\right)\left(\exp\left\{\sigma^2\right\} - 1\right)^{1/2} = \varsigma_S \approx 0.0321 \quad \Longleftrightarrow \quad \sigma^2 \approx 0.00011444,$$

which was found using a root search algorithm. Using the second equation in (3), 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

$$\mu \approx 14.90425.$$

Finally, using the first equation in (3), we get

$$k + \exp\left\{\mu + \sigma^2/2\right\} = \mathbb{E}[S] \quad \Longleftrightarrow \quad k = \mathbb{E}[S] - \exp\left\{\mu + \sigma^2/2\right\} \approx -1\text{'}970\text{'}704.$$

If we write F_W for the distribution function of

$$W = \log Z \sim \mathcal{N}(\mu \approx 14.90425, \sigma^2 \approx 0.00011444),$$

we get using the translated log-normal approximation

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

for all $x > k$, and $F_S(x) = 0$ for all $x \leq k$. For all $\alpha \in (0, 1)$ we then have

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

In particular, we get

$$q_{0.95} = F_S^{-1}(0.95) \approx k + \exp\left\{F_W^{-1}(0.95)\right\} \approx -1'970'704 + 3'023'266 = 1'052'562$$

and

$$q_{0.99} = F_S^{-1}(0.99) \approx k + \exp\left\{F_W^{-1}(0.99)\right\} \approx -1'970'704 + 3'045'387 = 1'074'684.$$

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

$$F_S(0) \approx F_W(\log[0 - k]) \approx F_W(\log 1'970'704) \approx 3.22 \cdot 10^{-322},$$

which is basically 0.

(c) We observe that with all the three approximations applied in Exercise 7.3 and in parts (a) and (b) above 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 assumptions. However, the probability to observe a negative claim S is vanishingly small in all the three approximations.

Non-Life Insurance: Mathematics and Statistics

Solution sheet 8

Solution 8.1 Panjer Algorithm

For the expected yearly claim amount π_0 we have

$$\pi_0 = \mathbb{E}[S] = \mathbb{E}[N]\,\mathbb{E}[Y_1] = 1 \cdot \mathbb{E}[k+Z] = k + \mathbb{E}[Z] = k + \exp\left\{\mu + \frac{\sigma^2}{2}\right\} \approx 4'124.$$

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. Note that $k = 10s$. If we write $g_l = \mathbb{P}[Y_1^+ = sl]$ for all $l \in \mathbb{N}$, then we have

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

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

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

where Φ is the distribution function of the standard Gaussian distribution and where we define $\log 0 = -\infty$. From now on we replace the original claim sizes Y_i with the discretized claim sizes Y_i^+, but, by a slight abuse of notation, we still write S for the yearly claim amount.
Note that $N \sim \text{Poi}(1)$ has a Panjer distribution with parameters $a = 0$ and $b = 1$, see Corollary 4.8 of the lecture notes (version of March 20, 2019). Applying the Panjer algorithm given in Theorem 4.9 of the lecture notes (version of March 20, 2019), 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. We have

$$f_0 = \mathbb{P}[N = 0] = e^{-1} \approx 0.37$$

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 customer 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

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

and

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

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 customer 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 π for the different franchises d:

franchise d	300	500	1'000	1'500	2'000	2'500
monthly pure risk premium π	307	297	274	253	233	216

Table 1: Monthly pure risk premiums π for the different franchises d.

More generally, the monthly pure risk premium π as a function of the franchise d, which is allowed to vary between 300 CHF and 2'500 CHF, is given in Figure 1. The R code used to calculate the values in Table 1 and to generate Figure 1 is given in Listings 1 and 2. Note that these monthly premiums only represent pure risk premiums. In order to 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.

Listing 1: R code for Exercise 8.1 (Function to calculate risk premium).

```
KK_premium <- function(lambda, mu, sigma2, span, shift){
  M <- 9500
  m <- floor(M/span)
  k0 <- shift/span
  g_min <- array(0, dim=c(m+1,1))   ### mass put 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))
  }
  g_max <- array(0, dim=c(m+1,1))   ### mass is put to the upper end of the interval
  g_max[2:(m+1),1] <- g_min[1:m,1]
  f1 <- matrix(0, nrow=m+1, ncol=3)     ### probability to get zero claims
  f1[1,1] <- exp(-lambda*(1-g_min[1,1]))
  f1[1,2] <- exp(-lambda*(1-g_max[1,1]))
  h1 <- matrix(0, nrow=m, ncol=3)     ### for values "l*g_{l}" of the discretized claim sizes
  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)
  }
  for (r in 1:m){   ### Panjer algorithm (a=0 and b=lambda*v, which is just lambda here)
    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
  }
  m1 <- 2500    ### maximal franchise
  m0 <- 300     ### minimal franchise
  i1 <- floor(m1/span+1)    ### number of iterations to m1
  i0 <- floor(m0/span+1)    ### number of iterations to m0
  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):floor(i+7000/span),1]
                      *f1[2:floor(7000/span+1),3])*0.1
                      +700*(1-sum(f1[1:floor(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):floor(i+7000/span),2]
                      *f1[2:floor(7000/span+1),3])*0.1+700*(1-sum(f1[1:floor(i+7000/span),2]))
  }
  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
}
```

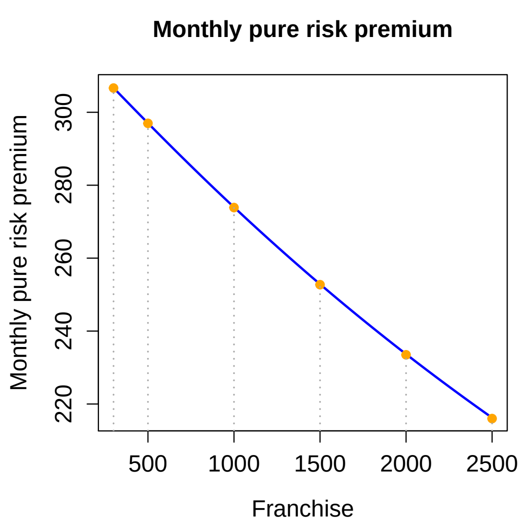

Figure 1: Plot of the monthly pure risk premium π as a function of the franchise d.

Listing 2: R code for Exercise 8.1 (Risk premium).

```
require(stats)
require(MASS)

### Run the function KK_premium
lambda <- 1
mu <- 7.8
sigma2 <- 1
span <- 10
shift <- 100
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="Monthly pure risk premium",
     xlab="Franchise", main="Monthly pure risk premium", cex.lab=1.25, cex.main=1.25,
     cex.axis=1.25)
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")
lines(x=c(300,300), y=c(0,price[300/span+1,3]),lty=3, lwd=1.5, col="darkgray")
lines(x=c(500,500), y=c(0,price[500/span+1,3]),lty=3, lwd=1.5, col="darkgray")
lines(x=c(1000,1000), y=c(0,price[1000/span+1,3]),lty=3, lwd=1.5, col="darkgray")
lines(x=c(1500,1500), y=c(0,price[1500/span+1,3]),lty=3, lwd=1.5, col="darkgray")
lines(x=c(2000,2000), y=c(0,price[2000/span+1,3]),lty=3, lwd=1.5, col="darkgray")
lines(x=c(2500,2500), y=c(0,price[2500/span+1,3]),lty=3, lwd=1.5, col="darkgray")

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

Solution 8.2 Monte Carlo Simulations

(a) We assume that for this comparably simple problem with no heavy tails 100’000 Monte Carlo simulations are enough to provide an empirical distribution function of S which is close to the true distribution function of S. The R codes used for part (a) are given in Listings 3 - 6.

Listing 3: R code for Exercise 8.2 (a) (Monte Carlo simulations).

```
compound.poisson.distribution <- Vectorize(function(n, lambdav, shape, rate){
  number.of.claims <- rpois(n=n, lambda=lambdav)
  sum(rgamma(n=number.of.claims, shape=shape, rate=rate))
},"n")
n <- 100000
lambdav <- 1000
shape <- 100
rate <- 1/10
set.seed(100)
claims <- compound.poisson.distribution(rep(1,n), lambdav, shape, rate)
```

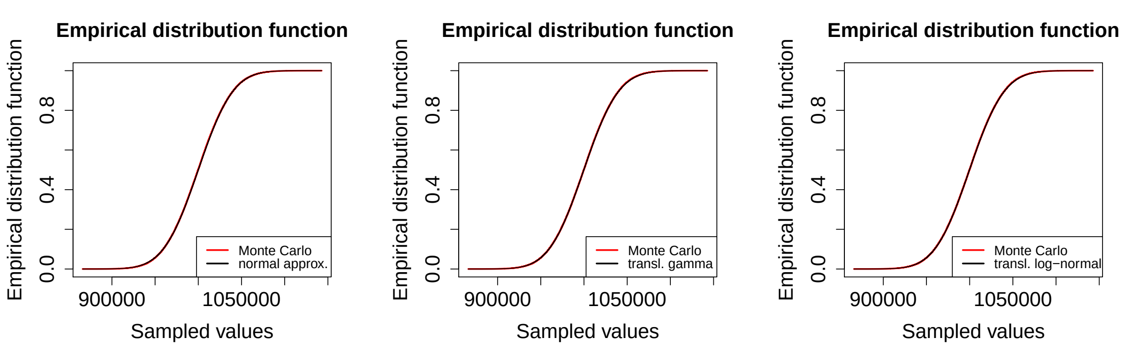

Figure 2: Comparison of the empirical distribution function of S resulting from 100’000 Monte Carlo simulations to the approximate distribution functions when using the normal (left), the translated gamma (middle) and the translated log-normal (right) approximation.

In Figure 2 we compare the empirical distribution function of S resulting from 100'000 Monte Carlo simulations to the approximate distribution functions when using the normal (left), the translated gamma (middle) and the translated log-normal (right) approximation. From these plots we cannot spot any differences between the various distribution functions. In Figure 3 we consider the log-log plot of the 100'000 Monte Carlo simulations of S and compare it to the normal (left), the translated gamma (middle) and the translated log-normal (right) approximation. We observe that all three approximations have a rather good fit to the tail of the distribution of S, but the translated gamma and the translated log-normal approximation seem slightly more accurate than the normal approximation. We conclude that in the absence of heavy tailed distributions the translated gamma and the translated log-normal approximation are very convincing in this example. Moreover, the skewness of S is small enough ($\varsigma_S \approx 0.0321$, see Exercise 7.4) and the expected number of claims large enough ($\lambda v = 1'000$, see Exercise 7.3) for the normal approximation to be a valid approximation, too.

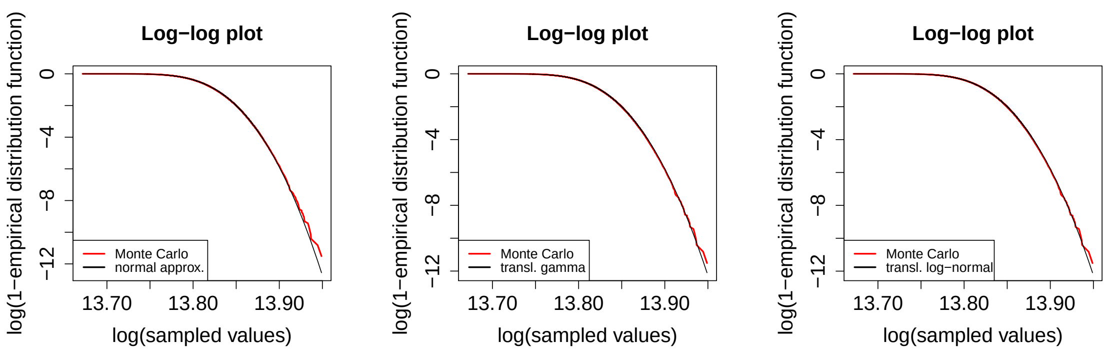

Figure 3: Log-log plot of the 100'000 Monte Carlo simulations of S compared to the normal (left), the translated gamma (middle) and the translated log-normal (right) approximation.

Listing 4: R code for Exercise 8.2 (a) (Normal approximation).

```
mu <- lambdav*shape/rate
sigma <- sqrt(lambdav*shape*(shape+1)/(rate^2))
par(mar=c(5.1, 4.4, 4.1, 2.1))
plot(claims[order(claims)], 1:n/(n+1), xlim=c(min(claims),max(claims)), type="l", col="red",
     main="Empirical distribution function", xlab="Sampled values",
     ylab="Empirical distribution function", cex.lab=1.5, cex.main=1.5, cex.axis=1.5, lwd=2)
lines(claims[order(claims)], pnorm((claims[order(claims)]),mu,sigma))
legend("bottomright", lty=1, lwd=2, col=c("red","black"),
       legend=c("Monte Carlo","normal approx.   "), cex=1)
plot(log(claims[order(claims)]), log(1-1:n/(n+1)), xlim=c(min(log(claims)),max(log(claims))),
     ylim=c(min(log(1-n/(n+1)),log(1-pnorm((claims[order(claims)]),mu,sigma))),0), type="l",
     col="red", main="Log-log plot", xlab="log(sampled values)",
     ylab="log(1-empirical distribution function)", cex.lab=1.5, cex.main=1.5, cex.axis=1.5,
     lwd=2)
lines(log(claims[order(claims)]), log(1-pnorm((claims[order(claims)]),mu,sigma)), col="black")
legend("bottomleft", lty=1, lwd=2, col=c("red","black"),
       legend=c("Monte Carlo","normal approx.   "), cex=1)
```

(b) Replicating 10'000 Monte Carlo simulations 100 times already requires some time. This is also the reason why we chose 10'000 as maximum number of simulations and not 100'000 as in part (a). Note that every single time we use Monte Carlo simulations to derive quantities like for example the quantiles $q_{0.95}$ and $q_{0.99}$, we get different results. This is something one needs to be aware of, and it is in contrast to the normal, the translated gamma and the translated log-normal approximation. In Figure 4 we show the densities of the 100 quantiles $q_{0.95}$ (left) and $q_{0.99}$ (right) resulting from Listing 7 where we replicate the $n \in \{100, 1'000, 10'000\}$ Monte Carlo simulations 100 times. We see that increasing the number of simulations n for every replication, the uncertainty regarding the quantiles $q_{0.95}$ and $q_{0.99}$ is reduced.

Listing 5: R code for Exercise 8.2 (a) (Translated gamma approximation).

```
skews <- (lambdav*shape*(shape+1)*(shape+2)/rate^3)/(lambdav*shape*(shape+1)/rate^2)^(3/2)
shape2 <- 4/skews^2
rate2 <- sqrt(shape2/(lambdav*shape*(shape+1)/rate^2))
k <- lambdav*shape/rate-shape2/rate2
plot(claims[order(claims)], 1:n/(n+1), xlim=c(min(claims),max(claims)), type="l", col="red",
     main="Empirical distribution function", xlab="Sampled values",
     ylab="Empirical distribution function", cex.lab=1.5, cex.main=1.5, cex.axis=1.5, lwd=2)
lines(claims[order(claims)], pgamma((claims[order(claims)])-k,shape=shape2,rate=rate2))
legend("bottomright", lty=1, lwd=2, col=c("red","black"),
       legend=c("Monte Carlo","transl. gamma   "), cex=1)
plot(log(claims[order(claims)]), log(1-1:n/(n+1)), xlim=c(min(log(claims)),max(log(claims))),
     ylim=c(min(log(1-n/(n+1)),log(1-pgamma((claims[order(claims)])-k,shape=shape2,
     rate=rate2))),0), type="l", col="red", main="Log-log plot", xlab="log(sampled values)",
     ylab="log(1-empirical distribution function)", cex.lab=1.5, cex.main=1.5, cex.axis=1.5,
     lwd=2)
lines(log(claims[order(claims)]), log(1-pgamma((claims[order(claims)])-k,shape=shape2,
      rate=rate2)))
legend("bottomleft", lty=1, lwd=2, col=c("red","black"),
       legend=c("Monte Carlo","transl. gamma   "), cex=1)
```

Listing 6: R code for Exercise 8.2 (a) (Translated log-normal approximation).

```
sigma.squared <- 0.00011444
mu2 <- 1/2*(log((exp(sigma.squared)-1)^(-1)*lambdav*shape*(shape+1)/rate^2)-sigma.squared)
k2 <- lambdav*shape/rate-exp(mu2+sigma.squared/2)
plot(claims[order(claims)], 1:n/(n+1), xlim=c(min(claims),max(claims)), type="l", col="red",
     main="Empirical distribution function", xlab="Sampled values",
     ylab="Empirical distribution function", cex.lab=1.5, cex.main=1.5, cex.axis=1.5, lwd=2)
lines(claims[order(claims)], pnorm(log((claims[order(claims)])-k2),mu2,sqrt(sigma.squared)))
legend("bottomright", lty=1, lwd=2, col=c("red","black"),
       legend=c("Monte Carlo","transl. log-normal    "), cex=1)
plot(log(claims[order(claims)]), log(1-1:n/(n+1)), xlim=c(min(log(claims)),max(log(claims))),
     ylim=c(min(log(1-n/(n+1)),log(1-pnorm(log((claims[order(claims)])-k2),mu2,
     sqrt(sigma.squared)))),0), type="l", col="red", main="Log-log plot",
     xlab="log(sampled values)", ylab="log(1-empirical distribution function)", cex.lab=1.5,
     cex.main=1.5, cex.axis=1.5, lwd=2)
lines(log(claims[order(claims)]), log(1-pnorm(log((claims[order(claims)])-k2),mu2,
      sqrt(sigma.squared))))
legend("bottomleft", lty=1, lwd=2, col=c("red","black"),
       legend=c("Monte Carlo","transl. log-normal    "), cex=1)
```

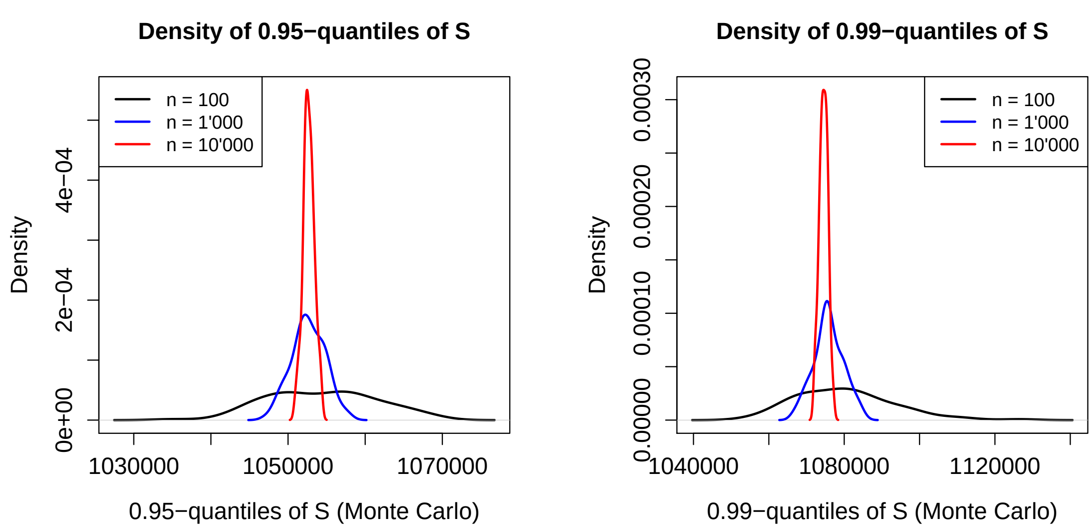

Figure 4: Densities of the 100 quantiles $q_{0.95}$ (left) and $q_{0.99}$ (right) resulting from replicating the $n \in \{100, 1'000, 10'000\}$ Monte Carlo simulations 100 times.

Listing 7: R code for Exercise 8.2 (b) (Quantiles).

```
### Monte Carlo simulations
k <- 100
n <- c(100,1000,10000)
set.seed(100)
claims.1 <- array(compound.poisson.distribution(n=rep(1,k*n[1]), lambdav=1000, shape=100,
                  rate=1/10), dim=c(n[1],k))
set.seed(200)
claims.2 <- array(compound.poisson.distribution(n=rep(1,k*n[2]), lambdav=1000, shape=100,
                  rate=1/10), dim=c(n[2],k))
set.seed(300)
claims.3 <- array(compound.poisson.distribution(n=rep(1,k*n[3]), lambdav=1000, shape=100,
                  rate=1/10), dim=c(n[3],k))

### Function calculating alpha-quantiles of S on the basis of Monte Carlo simulations of S
quantiles.monte.carlo <- function(claims, alpha){
  n <- nrow(claims)
  claims.sorted <- apply(claims, 2, sort)
  quantiles.alpha <- claims.sorted[floor(alpha*n)+1,]
}

### 0.95-quantiles
range(quantiles.1 <- quantiles.monte.carlo(claims=claims.1, alpha=0.95))
range(quantiles.2 <- quantiles.monte.carlo(claims=claims.2, alpha=0.95))
range(quantiles.3 <- quantiles.monte.carlo(claims=claims.3, alpha=0.95))

### Density
ymax <- max(density(quantiles.1)$y, density(quantiles.2)$y, density(quantiles.3)$y)
plot(density(quantiles.1), col="black", ylim=c(0,ymax), main="Density of 0.95-quantiles of S",
     xlab="0.95-quantiles of S (Monte Carlo)", cex.lab=1.25, cex.main=1.25, cex.axis=1.25,
     lwd=2)
lines(density(quantiles.2), col="blue", lwd=2)
lines(density(quantiles.3), col="red", lwd=2)
legend("topleft", col=c("black", "blue", "red"), lwd=2, lty=1,
       legend=c("n = 100","n = 1'000","n = 10'000"))

### 0.99-quantiles
range(quantiles.1 <- quantiles.monte.carlo(claims=claims.1, alpha=0.99))
range(quantiles.2 <- quantiles.monte.carlo(claims=claims.2, alpha=0.99))
range(quantiles.3 <- quantiles.monte.carlo(claims=claims.3, alpha=0.99))

### Density
ymax <- max(density(quantiles.1)$y, density(quantiles.2)$y, density(quantiles.3)$y)
plot(density(quantiles.1), col="black", ylim=c(0,ymax), main="Density of 0.99-quantiles of S",
     xlab="0.99-quantiles of S (Monte Carlo)", cex.lab=1.25, cex.main=1.25, cex.axis=1.25,
     lwd=2)
lines(density(quantiles.2), col="blue", lwd=2)
lines(density(quantiles.3), col="red", lwd=2)
legend("topright", col=c("black", "blue", "red"), lwd=2, lty=1,
       legend=c("n = 100","n = 1'000","n = 10'000"))
```

	$q_{0.95}$		$q_{0.99}$	
Monte Carlo	smallest	largest	smallest	largest
$n = 100$	1'035'018	1'069'209	1'053'719	1'126'533
$n = 1'000$	1'047'186	1'057'829	1'066'770	1'084'902
$n = 10'000$	1'050'955	1'054'282	1'072'045	1'077'195
Approximations				
normal	1'052'274		1'073'932	
translated gamma	1'052'563		1'074'682	
translated log-normal	1'052'562		1'074'684	

Table 2: Smallest and largest observed values of the quantiles $q_{0.95}$ and $q_{0.99}$ among the 100 replications of the $n \in \{100, 1'000, 10'000\}$ Monte Carlo simulations together with the values of the quantiles $q_{0.95}$ and $q_{0.99}$ resulting from the normal, the translated gamma and the translated log-normal approximation.

One can reach the same conclusions from Table 2, where we give the smallest and the largest observed values of the quantiles $q_{0.95}$ and $q_{0.99}$ among the 100 replications of the $n \in \{100, 1'000, 10'000\}$ Monte Carlo simulations. Moreover, we also give the values of the quantiles $q_{0.95}$ and $q_{0.99}$ resulting from the normal, the translated gamma and the translated log-normal approximation, see Exercises 7.3 and 7.4. We see that the quantiles resulting from the approximations are always between the smallest and the largest observed value resulting from the Monte Carlo simulations. Of course, one can argue that we could choose the number of simulations n large enough such that the results do not vary considerably anymore. However, a too high number of simulations n will lead to an excessive computation time. This is especially true if one considers heavy tailed distributions. Therefore, one is often inclined to use other algorithms for compound distributions, such as the Panjer algorithm and fast Fourier transforms.

Solution 8.3 Fast Fourier Transform

Assume that $\widetilde{Y}$ follows the claim size distribution given on the exercise sheet. Let Y denote the discretized version of $\widetilde{Y}$ that takes values in $\mathbb{N}_0$. More precisely, we shift the probability masses of $\widetilde{Y}$ to the right and define

$$\mathbb{P}[Y=0] \,=\, 0 \qquad \text{and} \qquad \mathbb{P}[Y=l] \,=\, \mathbb{P}\left[\widetilde{Y} \leq l\right] - \mathbb{P}\left[\widetilde{Y} \leq l-1\right],$$

for all $l \in \mathbb{N}$. By a slight abuse of notation, we still write S for the compound Poisson distribution with discrete claim size distribution Y. In particular, also S takes values in $\mathbb{N}_0$. We define

$$g_l \,=\, \mathbb{P}[Y=l] \qquad \text{and} \qquad f_l \,=\, \mathbb{P}[S=l],$$

for all $l \in \mathbb{N}_0$. We choose a threshold of $n = 2'000'000$, i.e. we determine the distribution function of S up to $n-1$. Note that n is chosen sufficiently high such that we approximately have

$$P[Y > n-1] \,\approx\, 0. \tag{1}$$

We define $\mathcal{A} = \{0, \ldots, n-1\}$ and calculate the discrete Fourier transform $(\hat{g}_z)_{z\in\mathcal{A}}$ of $(g_l)_{l\in\mathcal{A}}$ by

$$\hat{g}_z \,=\, \sum_{l=0}^{n-1} g_l \exp\left\{2\pi i \frac{zl}{n}\right\}, \tag{2}$$

for all $z \in \mathcal{A}$. Due to (1), we approximately have

$$\hat{g}_z \,\approx\, \mathbb{E}\left[\exp\left\{2\pi i \frac{zY}{n}\right\}\right] \,=\, M_Y\left(2\pi i \frac{z}{n}\right),$$

for all $z \in \mathcal{A}$, where M_Y denotes the moment generating function of Y. Note that we use an extended version of the moment generating function also allowing for complex numbers. If M_S denotes the moment generating function of S, again extended to complex numbers, then, according to Proposition 2.11 of the lecture notes (version of March 20, 2019), we have

$$M_S\left(2\pi i \frac{z}{n}\right) \,=\, \exp\left\{\lambda v \left[M_Y\left(2\pi i \frac{z}{n}\right) - 1\right]\right\} \,\approx\, \exp\left\{\lambda v \left(\hat{g}_z - 1\right)\right\}, \tag{3}$$

for all $z \in \mathcal{A}$. The left hand side of equation (3) can be written as

$$M_S\left(2\pi i \frac{z}{n}\right) \,=\, \sum_{l=0}^{\infty} f_l \exp\left\{2\pi i \frac{zl}{n}\right\} \,=\, \sum_{l=0}^{n-1}\left(f_l + \sum_{k=1}^{\infty} f_{l+kn}\right) \exp\left\{2\pi i \frac{zl}{n}\right\},$$

for all $z \in \mathcal{A}$. Using the approximation

$$f_l \approx f_l + \sum_{k=1}^{\infty} f_{l+kn}, \tag{4}$$

for all $l \in \mathcal{A}$, we compute the discrete Fourier transform $(\hat{f}_z)_{z\in\mathcal{A}}$ of $(f_l)_{l\in\mathcal{A}}$ by

$$\hat{f}_z = \sum_{l=0}^{n-1} f_l \exp\left\{2\pi i \frac{zl}{n}\right\} \approx M_S\left(2\pi i \frac{z}{n}\right) \approx \exp\left\{\lambda v\left(\hat{g}_z - 1\right)\right\},$$

for all $z \in \mathcal{A}$. Applying the inversion formula of the discrete Fourier transform, we finally calculate

$$f_l = \frac{1}{n}\sum_{z=0}^{n-1} \hat{f}_z \exp\left\{-2\pi i \frac{zl}{n}\right\}, \tag{5}$$

for all $l \in \mathcal{A}$. Note that due to the approximation in (4), instead of f_l we actually calculate

$$f_l + \sum_{k=1}^{\infty} f_{l+kn} > f_l,$$

for all $l \in \mathcal{A}$. This error is called wrap around error (or aliasing error), and n should be chosen large enough in order to keep this wrap around error small.

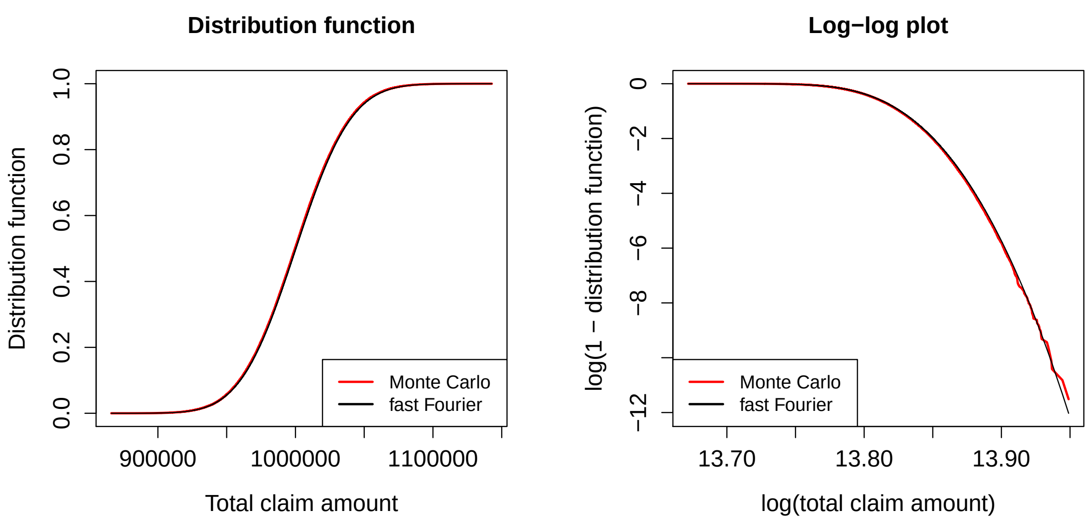

Figure 5: Comparison of the distribution function (left) and the log-log plot (right) of S resulting from the fast Fourier transform algorithm to the Monte Carlo simulations.

In R, the calculations in equations (2) and (5) can be done using the command `fft`. The corresponding R code is given in Listing 8. In Figure 5 we compare the distribution function (left) and the log-log plot (right) of S resulting from the fast Fourier transform algorithm to the Monte Carlo simulations of Exercise 8.2. We see that we get a very good fit. In particular, the threshold $n = 2'000'000$ seems to be high enough. For the 0.95-quantile $q_{0.95}$ and the 0.99-quantile $q_{0.99}$ we get

$$q_{0.95} = 1'053'089 \qquad \text{and} \qquad q_{0.99} = 1'075'215.$$

We see that we get values which are very close to the ones derived in Exercises 7.3 and 7.4, where we used the normal, the translated gamma and the translated log-normal approximation.

Listing 8: R code for Exercise 8.3.

```
### Fast Fourier transfom
n <- 2000000
lambdav <- 1000
claim.size <- c(0, pgamma(1:(n-1), shape=100, rate=1/10)-pgamma(0:(n-2), shape=100, rate=1/10))
claim.size.ft <- fft(claim.size)
total.claim.amount.ft <- exp(lambdav*(claim.size.ft-1))
total.claim.amount <- Re(fft(total.claim.amount.ft,inverse=TRUE)/length(total.claim.amount.ft))

### Monte Carlo simulations from Exercise 8.2
compound.poisson.distribution <- Vectorize(function(n, lambdav, shape, rate){
  number.of.claims <- rpois(n=n, lambda=lambdav)
  sum(rgamma(n = number.of.claims, shape=shape, rate=rate))
},"n")
m <- 100000
set.seed(100)
claim.amounts <- compound.poisson.distribution(n=rep(1,m), lambdav=1000, shape=100, rate=1/10)

### Calculate values of the distribution function of S using the fast Fourier transfrom
probabilities <- cumsum(total.claim.amount)[floor(claim.amounts[order(claim.amounts)])+1]

### Check the fast Fourier transform result
par(mar=c(5.1, 4.4, 4.1, 2.1))
plot(claim.amounts[order(claim.amounts)], 1:m/(m+1),
     xlim=c(min(claim.amounts),max(claim.amounts)), type="l", col="red",
     main="Distribution function", xlab="Total claim amount", ylab="Distribution function",
     cex.lab=1.25, cex.main=1.25, cex.axis=1.25, lwd=2)
lines(claim.amounts[order(claim.amounts)], probabilities, lwd=1)
legend("bottomright", lty=1, lwd=2, col=c("red","black"), legend=c("Monte Carlo","fast Fourier"), cex=1)
plot(log(claim.amounts[order(claim.amounts)]), log(1-1:m/(m+1)),
     xlim=c(min(log(claim.amounts)),max(log(claim.amounts))),
     ylim=c(min(log(1-m/(m+1)),log(1-probabilities)),0), type="l", col="red",
     main="Log-log plot", xlab="log(total claim amount)", ylab="log(1 - distribution function)",
     cex.lab=1.25, cex.main=1.25, cex.axis=1.25, lwd=2)
lines(log(claim.amounts[order(claim.amounts)]), log(1-probabilities), col="black", lwd=1)
legend("bottomleft", lty=1, lwd=2, col=c("red","black"), legend=c("Monte Carlo","fast Fourier"), cex=1)

### Determine the 0.95- and the 0.99-quantiles
which(cumsum(total.claim.amount) > 0.95)[1]-1
which(cumsum(total.claim.amount) > 0.99)[1]-1
```

Solution 8.4 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 $k \in \mathbb{N}$. 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 $\{0, 1, \dots, n\}$ 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 (version of March 20, 2019), we have

$$n = -\frac{a+b}{a}. \tag{6}$$

For the expectation of N we get

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

Using (6), we get

$$-an - a - b \;=\; a\,\frac{a+b}{a} - a - b \;=\; 0. \tag{7}$$

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:

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

Using (6), we get

$$\begin{aligned}
-an^2 - (2a+b)n - a - b \;&=\; -a\left(\frac{a+b}{a}\right)^2 + (2a+b)\frac{a+b}{a} - a - b\\
&=\; -\frac{a^2+2ab+b^2}{a} + \frac{2a^2+3ab+b^2}{a} - \frac{a^2+ab}{a}\\
&=\; 0.
\end{aligned} \tag{8}$$

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 (7) and in (8) 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 risk-averse 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

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

Since S_1 and S_2 both have a compound Poisson distribution, Proposition 2.11 of the lecture notes (version of March 20, 2019) 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

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

In this exercise we use the exponential utility function u given by

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

for all $x \in \mathbb{R}$, with $\alpha = 1.5 \cdot 10^{-6}$. Thus, 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 (version of March 20, 2019) 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, and are given by

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

and

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

In particular, since $\alpha < c$ and $\alpha < \kappa$, 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

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

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 $X \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ is given by

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

for all $r \in \mathbb{R}$, see Exercise 1.3. Thus, if we assume Gaussian distributions for S_1 and S_2, and according to the calculations in part (c), 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

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

and that

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

Then, Proposition 2.11 of the lecture notes (version of March 20, 2019) gives

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

and

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

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 in (d) is the Central Limit Theorem. In particular, neither the gamma distribution nor the exponential distribution are heavy-tailed distributions. Moreover, the skewness ς_{S_1} of S_1 and also the skewness ς_{S_2} of S_2 are rather small ($\varsigma_{S_1} \approx 0.034$ and $\varsigma_{S_2} \approx 0.067$). Thus, the expected numbers of claims $\lambda_1 v_1 = \lambda_2 v_2 = 1\text{'}000$ are large enough for the normal approximations to be valid approximations for the compound Poisson distributions.

(e) On the one hand, in part (c) we have shown that the utility indifference price $\pi = \pi(u, S, c_0)$ is given by

$$\pi \;=\; \frac{1}{\alpha}\lambda v \left[M_{Y_1}(\alpha) - 1\right].$$

On the other hand, according to the calculations in part (c), we also have

$$\widetilde{\pi} \;=\; \widetilde{\pi}\left(u, \widetilde{S}, c_0\right) \;=\; \frac{1}{\alpha}\log \mathbb{E}\left[\exp\left\{\alpha \widetilde{S}\right\}\right].$$

From this we can calculate

$$\widetilde{\pi} \;=\; \frac{1}{\alpha}\log \mathbb{E}\left[\exp\left\{\alpha \sum_{i=1}^{\lambda v} Y_i\right\}\right] \;=\; \frac{1}{\alpha}\sum_{i=1}^{\lambda v}\log \mathbb{E}\left[\exp\left\{\alpha Y_i\right\}\right] \;=\; \frac{1}{\alpha}\lambda v \log M_{Y_1}(\alpha).$$

We then have

$$\pi \;>\; \widetilde{\pi} \quad\Longleftrightarrow\quad M_{Y_1}(\alpha) - 1 \;>\; \log M_{Y_1}(\alpha).$$

We define

$$g(x) \;=\; x - 1 \quad\text{and}\quad h(x) \;=\; \log x.$$

For $x = 1$ we have $g(1) = 0 = h(1)$. Since g is linear and h is strictly concave, we get

$$g(x) \;>\; h(x)$$

for all $x \neq 1$ in the domain of g and h. Since for the claim sizes we assume $Y_1 > 0$, $\mathbb{P}$-a.s., and since $\alpha > 0$, we have $M_{Y_1}(\alpha) > 1$. If follows that

$$M_{Y_1}(\alpha) - 1 \;>\; \log M_{Y_1}(\alpha)$$

and, thus, $\pi > \widetilde{\pi}$. This does not come as a surprise: in the compound Poisson model we have randomness in the number of claims and under risk aversion we do not like this uncertainty. This leads to a higher price in the compound Poisson model and explains why $\pi > \widetilde{\pi}$.

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

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

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

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

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

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

In particular, π_{CoC} is then given by

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

Note that we have

$$\frac{\pi_{\text{CoC}} - \mathbb{E}[S]}{\mathbb{E}[S]} \approx \frac{499'395'571 - 488'817'614}{488'817'614} = \frac{10'577'957}{488'817'614} \approx 2.16\%.$$

Thus, the loading $\pi_{\text{CoC}} - \mathbb{E}[S]$ is given by approximately 2.16% 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 have found that

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

and that by a similar computation we get

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

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

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

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 (version of March 20, 2019), we have

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

By the definition of the mean excess function $e_S(\cdot)$ of S, we can write

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

Moreover, according to the formula given in Chapter 3.2.3 of the lecture notes (version of March 20, 2019), the mean excess function $e_S[\text{VaR}_{1-q}(S)]$ above level $\text{VaR}_{1-q}(S)$ for the log-normal distribution S is given by

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

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

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

In particular, we have found that

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

For $1-q=99\%$ we get $\Phi^{-1}(1-q)\approx 2.326$, and

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

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 have seen that in this example

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

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

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

which is equivalent to

$$\mathrm{VaR}_u(S) = \mathrm{ES}_{1-q}(S).$$

From part (b) we know that

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

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

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

Hence, we can solve for u to get

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

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

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

This implies that

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

and that

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

Since

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

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 Variance Loading Principle

(a) Let S_1, S_2, S_3 denote the total claim amounts of the passenger cars, delivery vans and trucks, respectively. Then, according to Proposition 2.11 of the lecture notes (version of March 20, 2019), 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 Table 2 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 (version of March 20, 2019), we get

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

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

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

Since S_1, S_2 and S_3 are independent by assumption, the variance of the total claim amount S of the car fleet is given by

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

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

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

Note that we have

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

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

Solution 9.4 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 (version of March 20, 2019), 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

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

Hence, the Esscher premium π_α is always non-decreasing in α. 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, if S is non-deterministic, then the Esscher premium π_α is strictly increasing in α.

(b) Let $\alpha \in (0, r_0)$. According to Corollary 6.16 of the lecture notes (version of March 20, 2019), 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

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

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 (version of March 20, 2019) 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

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

Note that since $c > c-\alpha$ and $\gamma + 1 > 1$, 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 Method of Bailey & Simon

In this exercise we work with two 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 in Table 1 on the exercise sheet. The multiplicative tariff structure leads to 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 $\chi_{1,2}, \chi_{1,3}, \chi_{2,1}, \chi_{2,2}, \chi_{2,3}, \chi_{2,4}$. Using 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 $i \in \{2,3\}$ (recall that we set $\widehat{\chi}_{1,1} = 1$). Then, $\widehat{\chi}_{1,i}$ 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, for $i \in \{2,3\}$ we get

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

By an analogous calculation, one finds

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

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	$\widehat{\chi}_{1,i}$
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
$\widehat{\chi}_{2,j}$	2'176	1'751	1'491	1'493	

Table 1: Tariff structure resulting from the method of Bailey & Simon.

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 between 21 and 30 years gets a considerably higher tariff than an insured with a higher age. The smallest tariff is assigned to insureds with age between 41 and 60 years. Note that we have

$$\sum_{i=1}^{3}\sum_{j=1}^{4} v_{i,j}\,\mu\,\widehat{\chi}_{1,i}\,\widehat{\chi}_{2,j} \;=\; 21'320 \;>\; 21'300 \;=\; \sum_{i=1}^{3}\sum_{j=1}^{4} S_{i,j},$$

which confirms the (systematic) positive bias of the method of Bailey & Simon shown in Lemma 7.2 of the lecture notes (version of March 20, 2019).

Solution 10.2 Method of Bailey & Jung

We use the same setup and the same notation as in the solution of Exercise 10.1. In order to get a unique solution, we again set $\mu = 1$ and $\chi_{1,1} = 1$. Using the method of Bailey & Jung, which is also called method of total marginal sums, 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

$$\sum_{j=1}^{J} v_{i,j}\,\mu\,\chi_{1,i}\,\chi_{2,j} \;=\; \sum_{j=1}^{J} S_{i,j}, \qquad i \in \{2, 3\},$$
$$\sum_{i=1}^{I} v_{i,j}\,\mu\,\chi_{1,i}\,\chi_{2,j} \;=\; \sum_{i=1}^{I} S_{i,j}, \qquad j \in \{1, 2, 3, 4\}.$$

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

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

Thus, for $i \in \{2, 3\}$ (recall that we set $\widehat{\chi}_{1,1} = 1$) and $j \in \{1, 2, 3, 4\}$, we get

$$\widehat{\chi}_{1,i} \;=\; \sum_{j=1}^{4} S_{i,j} \Big/ \sum_{j=1}^{4} \widehat{\chi}_{2,j}, \qquad \text{and,}$$
$$\widehat{\chi}_{2,j} \;=\; \sum_{i=1}^{3} S_{i,j} \Big/ \sum_{i=1}^{3} \widehat{\chi}_{1,i}.$$

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	$\widehat{\chi}_{1,i}$
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	1'685	1'685	1.13
$\widehat{\chi}_{2,j}$	2'170	1'749	1'490	1'490	

Table 2: Tariff structure resulting from the method of Bailey & Jung.

We see that the results are very close to those in Exercise 10.1, where we applied the method of Bailey & Simon. However, now we have

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

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

Solution 10.3 Log-Linear Gaussian Regression Model

(a) In the log-linear Gaussian regression model we work with a stochastic model for the claim amounts $S_{i,j}$ for the risk classes $(i,j), 1 \le i \le 3, 1 \le j \le 4$, given in Table 1 on the exercise sheet. We assume that

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

where $\beta_0, \beta_{1,i}, \beta_{2,j} \in \mathbb{R}$ and $\sigma^2 > 0$, for all risk classes $(i,j), 1 \le i \le 3, 1 \le j \le 4$. 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 $\beta_{1,1} = \beta_{2,1} = 0$. Simplifying notation, we write $\boldsymbol{X} = (X_1, \ldots, X_M)'$ 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

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

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

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

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

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

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 $m \in \{1, \dots, 12\}$, we find the design matrix Z given in Table 3. Note that we can also let R find the design matrix by itself, see Listing 1 given below.

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

Table 3: Design matrix Z ($\beta_{1,1} = \beta_{2,1} = 0$).

(b) The R code used for parts (b), (c) and (d) is given in Listing 1 below. According to formula (7.11) of the lecture notes (version of March 20, 2019), the MLE of the parameter vector $\boldsymbol{\beta}$ is given by

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

Note that $\widehat{\boldsymbol{\beta}}^{\text{MLE}}$ does not depend on σ^2. Moreover, the design matrix Z has full column rank and, thus, $Z'Z$ is indeed invertible. We get the following tariff structure:

$\widehat{\beta}_0 = 7.688$	21-30y	31-40y	41-50y	51-60y	$\widehat{\beta}_{1,i}$
passenger car	2'182	1'759	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
$\widehat{\beta}_{2,j}$	0	-0.216	-0.375	-0.374	

Table 4: Tariff structure resulting from the log-linear Gaussian regression model.

If we use the same parametrization as in Exercises 10.1 and 10.2, we get the following table:

$\exp\{\widehat{\beta}_0\} = 1$	21-30y	31-40y	41-50y	51-60y	$\exp\{\widehat{\beta}_{1,i}\}$
passenger car	2'182	1'759	1'500	1'501	1
delivery van	2'063	1'663	1'417	1'419	0.95
truck	2'444	1'970	1'680	1'682	1.12
$\exp\{\widehat{\beta}_{2,j}\}$	2'182	1'759	1'500	1'501	

Table 5: Tariff structure with the same parametrization as in Exercises 10.1 and 10.2.

Note that the tariffs in Tables 4 and 5 do not change with the different parametrization.

(c) We see that the results are very close to those found in Exercises 10.1 and 10.2, where we applied the method of Bailey & Simon and the method of Bailey & Jung. The only differences are that with the method of Bailey & Jung we get coinciding marginal totals and with the log-linear Gaussian regression model we are in a stochastic framework which allows for calculating parameter uncertainties and hypothesis testing, i.e we 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:

	$\widehat{\beta}_0$	$\widehat{\beta}_{1,2}$	$\widehat{\beta}_{1,3}$	$\widehat{\beta}_{2,2}$	$\widehat{\beta}_{2,3}$	$\widehat{\beta}_{2,4}$
p-value	≈ 0	0.2322	0.0366	0.0045	0.0003	0.0003

Table 6: Resulting p-values for the individual parameters.

For every parameter, R calculates the corresponding p-value by applying a t-test to the null hypothesis that the parameter under consideration is equal to 0. While the p-values for $\widehat{\beta}_0, \widehat{\beta}_{1,3}, \widehat{\beta}_{2,2}, \widehat{\beta}_{2,3}, \widehat{\beta}_{2,4}$ are smaller than 0.05 and, thus, these parameters are significantly different from zero, the p-value for $\widehat{\beta}_{1,2}$ (delivery van) is fairly high. Hence, we might question if we really need the class delivery van. This is in line with the observations that the risk characteristics for the classes passenger car and delivery van are close to each other, see Tables 1, 2, 4 and 5.

(d) 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. We can perform the same analysis as above to get the MLE $\widehat{\boldsymbol{\beta}}_{H_0}^{\text{MLE}}$ of the reduced model H_0. 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, we have

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

Now, for all $m \in \{1, \ldots, 12\}$ we define the fitted value $\widehat{X}_m^{\text{full}}$ of the full model and the fitted value $\widehat{X}_m^{H_0}$ of the reduced model. In particular, we have

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

and

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

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

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

and

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

According to formula (7.17) of the lecture notes (version of March 20, 2019), 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 freedom given by $\text{df}_1 = p = 2$ and $\text{df}_2 = M - r - 1 = 6$. 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 seems to be no statistical evidence that the classification into different types of vehicles could be omitted.

Listing 1: R code for Exercise 10.3.

```r
### 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 design matrix Z and log(S_{i,j}) in one dataset
data <- as.data.frame(cbind(Z[,-1],matrix(log(t(S)),nrow=12)))
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)
summary(linear.model1)

### Fitted values
matrix(exp(fitted(linear.model1)), byrow=TRUE, nrow=3)

### We can also get the parameters by applying formula (7.11) of the lecture notes
solve(t(Z)%*%Z) %*% t(Z) %*% matrix(log(t(S)), nrow=12)

### We can also use R directly on the data (it finds the design matrix internally)
car <- c("passenger car", "van", "truck")
age <- c("X21_30y", "X31_40y", "X41_50y", "X51_60y")
dat <- expand.grid(car, age)
colnames(dat) <- c("car","age")
dat$observation <- as.vector(log(S))
linear.model1.direct <- lm(formula = observation ~ car+age, data=dat)
summary(linear.model1.direct)

### Apply the regression model under H_0, calculate the test statistic F and the p-value
linear.model2 <- lm(formula = observation ~ X31_40y+X41_50y+X51_60y, data=data)
test.stat <- 3*(sum((data[,6]-fitted(linear.model2))^2)-sum((data[,6]-fitted(linear.model1))^2))
                /sum((data[,6]-fitted(linear.model1))^2)
pf(test.stat, 2, 6, lower.tail=FALSE)

### We can also directly use anova to test H_0
anova(linear.model1,linear.model2)
```

Solution 10.4 Tweedie's Compound Poisson Model

(a) We can write S as

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

where $N \sim \mathrm{Poi}(\lambda v)$, $Y_1, Y_2, \ldots \overset{\text{i.i.d.}}{\sim} G$ 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 in the second equality 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, for all $x > 0$,

$$\begin{aligned} c(x,\phi,w) &= \log\left(\sum_{n=1}^{\infty}\left[\frac{(\gamma+1)^{\gamma+1}}{\gamma}\left(\frac{\phi}{w}\right)^{-\gamma-1}\right]^n \frac{1}{\Gamma(n\gamma)n!}\, x^{n\gamma-1}\right) \\ &= \log\left[\sum_{n=1}^{\infty} (\lambda v c^{\gamma})^n \frac{1}{\Gamma(n\gamma)n!}\, x^{n\gamma-1}\right]. \end{aligned}$$

By putting the above terms together, we get, for all $x > 0$,

$$\begin{aligned} f_X(x;\theta,\phi) &= \exp\left\{\frac{x\theta - b(\theta)}{\phi/w} + c(x,\phi,w)\right\} \\ &= \exp\left\{-(cx+\lambda v) + \log\left[\sum_{n=1}^{\infty}(\lambda v c^{\gamma})^n \frac{1}{\Gamma(n\gamma)n!} x^{n\gamma-1}\right]\right\} \\ &= f_S(x), \end{aligned}$$

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 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 provides more insights into the differences between the tariffs.

Non-Life Insurance: Mathematics and Statistics

Solution sheet 11

Solution 11.1 Claim Frequency Modeling with GLM

(a) In this exercise we work with three 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:

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

We write N_{l_1,l_2,l_3} for the numbers of claims, v_{l_1,l_2,l_3} for the volumes and λ_{l_1,l_2,l_3} for the claim frequencies of the risk classes $(l_1,l_2,l_3), 1 \leq l_1 \leq 2, 1 \leq l_2 \leq 2, 1 \leq l_3 \leq 3$. 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

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

In particular, we have

$$\lambda_{l_1,l_2,l_3} = \mathbb{E}\left[\frac{N_{l_1,l_2,l_3}}{v_{l_1,l_2,l_3}}\right] = \mathbb{E}\left[X_{l_1,l_2,l_3}\right].$$

We model

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

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 10.3, we relabel the risk classes with the index $m \in \{1, \ldots, M\}$, where $M = 2 \cdot 2 \cdot 3 = 12$, define $\boldsymbol{X} = (X_1, \ldots, X_M)'$ and the design matrix $Z \in \mathbb{R}^{M \times (r+1)}$ that satisfies

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

where the logarithm is applied componentwise to $\mathbb{E}[\boldsymbol{X}]$. Let $m \in \{1, \ldots, 12\}$. According to Example 7.9 of the lecture notes (version of March 20, 2019), $X_m = N_m/v_m$ 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 the m-th element of the vector $Z\boldsymbol{\beta}$. Summarizing, we assume that $X_1, \dots, X_M$ are independent with

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

for all $m \in \{1, \dots, M\}$. As $b(\cdot) = \exp\{\cdot\}$, we also have $b'(\cdot) = \exp\{\cdot\}$, where b' denotes the first derivative of b. In particular, the log-link function $g(\cdot) = \log(\cdot)$ is equal to the canonical link function $h(\cdot) = (b')^{-1}(\cdot) = \log(\cdot)$ in the Poisson model. Therefore, we can use equation (7.26) of the lecture notes (version of March 20, 2019): the MLE $\widehat{\boldsymbol{\beta}}^{\text{MLE}}$ of $\boldsymbol{\beta}$ is the solution of

$$Z'Vb'(Z\boldsymbol{\beta}) \,=\, Z'V\exp\{Z\boldsymbol{\beta}\} \overset{!}{=} Z'V\boldsymbol{X}, \tag{1}$$

where the weight matrix V is given by $V = \text{diag}(v_1, \dots, v_M)$, see also Proposition 7.11 of the lecture notes (version of March 20, 2019). Equation (1) has to be solved numerically. We refer to Listing 1 for the application of this GLM model in R. The resulting MLEs of the parameters $\beta_0, \beta_{1,2}, \beta_{2,2}, \beta_{3,2}, \beta_{3,3}$ are given in the first row of Table 1. 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. Analogously, if the vehicle is at most one year old, we expect more claims than if it is 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. Similarly as the log-linear Gaussian regression model discussed in Exercise 10.3, the GLM framework allows for calculating parameter uncertainties and hypothesis testing. According to the R output, for the individual parameters we get the p-values listed in the second row of Table 1. These p-values are all substantially smaller than 0.05 and, thus, all the parameters are significantly different from zero.

	$\widehat{\beta}_0$	$\widehat{\beta}_{1,2}$	$\widehat{\beta}_{2,2}$	$\widehat{\beta}_{3,2}$	$\widehat{\beta}_{3,3}$
MLE	-1.4351	-0.2371	-0.5019	-0.4036	-1.6571
p-value	≈ 0	0.0009	≈ 0	≈ 0	≈ 0

Table 1: MLEs of the parameters $\beta_0, \beta_{1,2}, \beta_{2,2}, \beta_{3,2}, \beta_{3,3}$ and corresponding p-values.

Listing 1: R code for Exercise 11.1 (a).

```
### 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)))
volumes <- c(1,2,5,4,9,70,2,3,6,8,15,50)*100
counts <- c(25,15,15,60,90,210,45,45,30,80,120,90)
Z <- model.matrix(counts ~ class + age + zone)

### Store design matrix Z (without intercept term), counts and volumes in one dataset
data <- as.data.frame(cbind(Z[,-1],counts,volumes))

### Apply GLM
d.glm <- glm(counts ~ class2 + age2 + zone2 + zone3, data=data, offset=log(volumes),
             family=poisson())
summary(d.glm)
```

(b) The plots of the observed and the fitted claim frequencies against the vehicle class, the vehicle age and the geographic zone are given in Figure 1, the corresponding R code in Listing 2. Note that the observed and the fitted marginal claim frequencies are always the same. This is a direct consequence of equation (1) above, which ensures that the observed and the fitted total marginal sums are the same (if we use the same volumes again), see also the remarks after Proposition 7.11 in the lecture notes (version of March 20, 2019). Moreover, in the marginal plot for the vehicle class we do not 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 we would have expected after the discussion at the end of part (a). The reason for this peculiarity is that the MLE $\widehat{\beta}_{1,2}$ is driven by the risk cells with the biggest volumes ($v_6 = 7'000$ and $v_{12} = 5'000$). However, in these risk cells with the biggest volumes we observe very low claim frequencies. This implies that these risk cells have a small impact on the mean claim frequency. As a consequence, the resulting mean claim frequency is of similar size for both vehicles with weight over 60 kg and more than two gears and for other vehicles. For the other variables vehicle age and geographic zone we again see the same results as in part (a).

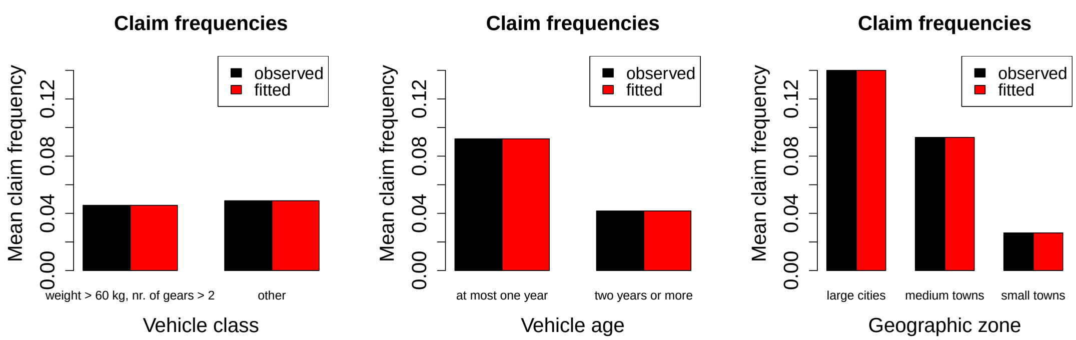

Figure 1: Observed and fitted claim frequencies against the vehicle class, the vehicle age and the geographical zone.

Listing 2: R code for Exercise 11.1 (b).

```
### Store features, observed numbers of claims and fitted numbers of claims in one dataset
data2 <- as.data.frame(cbind(class, age, zone, volumes, counts, fitted(d.glm)))
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))
par(mar=c(5.1, 4.6, 4.1, 2.1))
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="Claim frequencies",
        ylim=c(0,0.15), xlab="Vehicle class", ylab="Mean claim frequency", legend.text=FALSE,
        col=1:2, cex.names=0.95, cex.lab=1.5, cex.main=1.5, cex.axis=1.5)
legend("topright", legend=c("observed  ", "fitted  "), fill=1:2, cex=1.25)

### 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", "two years or more"), main="Claim frequencies",
        ylim=c(0,0.15), xlab="Vehicle age", ylab="Mean claim frequency", legend.text=FALSE,
        col=1:2, cex.names=0.95, cex.lab=1.5, cex.main=1.5, cex.axis=1.5)
legend("topright", legend=c("observed  ", "fitted  "), fill=1:2, cex=1.25)

### 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", "medium towns", "small towns"), main="Claim frequencies",
        ylim=c(0,0.15), xlab="Geographic zone", ylab="Mean claim frequency", legend.text=FALSE,
        col=1:2, cex.names=0.95, cex.lab=1.5, cex.main=1.5, cex.axis=1.5)
legend("topright", legend=c("observed  ", "fitted  "), fill=1:2, cex=1.25)
```

(c) The Tukey-Anscombe plot given in Figure 2 can be generated by the R code of Listing 3. The plot looks rather fine in the sense that we do not observe any structure. However, we remark that we only have 12 observations in this example and, thus, it is difficult to detect possible patterns and to make a clear statement.

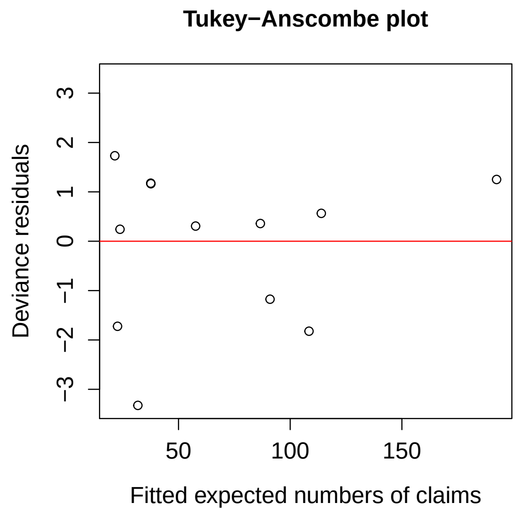

Figure 2: Tukey-Anscombe plot.

Listing 3: R code for Exercise 11.1 (c).

```
### Deviance residuals
dev.red <- residuals.glm(d.glm)

### Tukey-Anscombe plot
par(mar=c(5.1, 4.4, 4.1, 2.1))
plot(data2$fitted, dev.red, main="Tukey-Anscombe plot",
     xlab="Fitted expected numbers of claims", ylab="Deviance residuals",
     ylim=c(-max(abs(dev.red)),max(abs(dev.red))), cex.lab=1.25, cex.main=1.25, cex.axis=1.25)
abline(h=0, col="red")
```

(d) We 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 have seen 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 for middle-sized towns and large cities. Thus, we would expect that the classification into the three different geographic zones is reasonable. Now we investigate this. The estimates of the expected values of X_m are given by

$$\widehat{\mu}_m \,=\, b'(\widehat{\theta}_m) \,=\, \exp\left\{\widehat{\theta}_m\right\} \,=\, \exp\left\{\left[Z\widehat{\boldsymbol{\beta}}^{\text{MLE}}\right]_m\right\},$$

for all $m = 1, \ldots, M$, and we write $\widehat{\boldsymbol{\mu}} = (\widehat{\mu}_1, \ldots, \widehat{\mu}_M)'$. According to page 196 of the lecture notes (version of March 20, 2019), the scaled deviance statistics is given by

$$\begin{aligned} D^*(\boldsymbol{X}, \widehat{\boldsymbol{\mu}}) \,&=\, \frac{2}{\phi}\sum_{m=1}^{M} w_m(X_m h(X_m) - b[h(X_m)] - X_m h(\widehat{\mu}_m) + b[h(\widehat{\mu}_m)]) \\ &=\, 2\sum_{m=1}^{M} v_m\left(X_m \log X_m - X_m - X_m \log \widehat{\mu}_m + \widehat{\mu}_m\right). \end{aligned} \tag{2}$$

Moreover, since for the Poisson case we have $\phi = 1$, the scaled deviance statistics $D^*(\boldsymbol{X}, \widehat{\boldsymbol{\mu}})$ and the deviance statistics $D(\boldsymbol{X}, \widehat{\boldsymbol{\mu}})$ are the same. 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 $\widehat{\boldsymbol{\beta}}_{H_0}^{\text{MLE}}$ for this reduced model and define

$$\widehat{\boldsymbol{\mu}}_{H_0} = \exp\left\{ Z_{H_0} \widehat{\boldsymbol{\beta}}_{H_0}^{\text{MLE}} \right\},$$

where Z_{H_0} is the design matrix in the reduced model. According to formula (7.30) of the lecture notes (version of March 20, 2019), the test statistic

$$\text{F} = \frac{D(\boldsymbol{X}, \widehat{\boldsymbol{\mu}}_{H_0}) - D(\boldsymbol{X}, \widehat{\boldsymbol{\mu}})}{D(\boldsymbol{X}, \widehat{\boldsymbol{\mu}})} \frac{M - r - 1}{p} = \frac{7}{2} \frac{D(\boldsymbol{X}, \widehat{\boldsymbol{\mu}}_{H_0}) - D(\boldsymbol{X}, \widehat{\boldsymbol{\mu}})}{D(\boldsymbol{X}, \widehat{\boldsymbol{\mu}})}$$

has approximately an F-distribution with degrees of freedom given by $\text{df}_1 = p = 2$ and $\text{df}_2 = M - r - 1 = 7$. We get

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

which corresponds to a p-value of approximately 0.007%. Thus, we can reject H_0 at significance level of 5%. According to formula (7.31) of the lecture notes (version of March 20, 2019), a second test statistic is given by

$$X^2 = D^*(\boldsymbol{X}, \widehat{\boldsymbol{\mu}}_{H_0}) - D^*(\boldsymbol{X}, \widehat{\boldsymbol{\mu}}).$$

The test statistic X^2 has approximately a χ^2-distribution with $\text{df} = p = 2$ degrees of freedom. 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 both tests, we can conclude that there seems to be no statistical evidence that the classification into the geographic zones could be omitted. For the R code used in part (d) we refer to Listing 4.

Listing 4: R code for Exercise 11.1 (d).

```
### Deviance statistics of the full model
D.full <- d.glm$deviance

### Fit the reduced model
d.glm.2 <- glm(counts ~ class2 + age2, data=data, offset=log(volumes), family=poisson())
summary(d.glm.2)

### Deviance statistics of the reduced model
D.reduced <- d.glm.2$deviance

### Calculate the test statistic F
test.stat <- 7/2*(D.reduced-D.full)/D.full

### Calculation of the corresponding p-value
pf(test.stat, 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 11.2 Claim Frequency Modeling with Neural Networks

(a) The Poisson deviance statistics calculated on the datasets `trainset` and `testset` with the R code of Listing 5 are given in the first column of Table 2.

	GLM	NN (100 epochs)	NN (1'000 epochs)
deviance statistics `trainset`	1'314.7	709.7	111.6
deviance statistics `testset`	1'454.3	1'070.2	1'523.5

Table 2: Deviance statistics.

Listing 5: R code for Exercise 11.2 (a).

```
### Apply GLM (on the training set)
d.glm <- glm(ClaimNb ~ VehPower+VehAge+DrivAge, data=trainset, offset=log(Exposure),
             family=poisson())
summary(d.glm)

### Deviance statistics on training set
predtrain <- predict(d.glm, trainset, type="response")
obstrain <- trainset$ClaimNb
(Deviancetrain <- 2*sum(log((obstrain/predtrain)^obstrain)-obstrain+predtrain))
d.glm$deviance   ### check deviance statistics on training set

### Deviance statistics on test set
predtestGLM <- predict(d.glm, testset, type="response")
obstest <- testset$ClaimNb
(Deviancetest <- 2*sum(log((obstest/predtestGLM)^obstest)-obstest+predtestGLM))
```

(b) We fit the neural network for 100 gradient descent steps, see the R code given in Listing 6 and use the resulting model to calculate the Poisson deviance statistics on the datasets `trainset` and `testset`, see the second column of Table 2. We observe that the neural network leads to smaller values of the deviance statistics on both the datasets `trainset` and `testset`. This is an indication that the neural network model has better predictive power than the GLM model. We remark that a simple GLM model like the one used in this exercise usually is not able to cope with interactions between the tariff criteria, in contrast to neural network models. This might explain the lower deviance statistics observed for the neural network model on the data `testset`. However, we do not further investigate this here.

(c) We perform the exact same fitting procedure as in part (b), with the only difference that we use 1'000 gradient descent steps instead of only 100. The resulting Poisson deviance statistics on the datasets `trainset` and `testset` are given in the third column of Table 2. On the one hand, we see that the deviance statistics on the dataset `trainset` used during training is smaller than for the GLM model of part (a) and the neural network model with 100 gradient descent steps of part (b). However, this "better" fit is deceiving. In fact, the deviance statistics on the dataset `testset` is bigger than for the GLM model of part (a) and the neural network model with 100 gradient descent steps of part (b). We emphasize that the dataset `testset` has not been seen during training and, thus, is the correct dataset to analyze the predictive power of a fitted model. We conclude that with 1'000 gradient descent steps we are in the situation of overfitting to the training data `trainset`. Therefore, the number of gradient descent steps has to be chosen carefully. Usually, one splits the available dataset into a learning set and a validation set. The learning set is then used to perform the gradient descent steps and to fit the model. The validation set can be used to track over-fitting to the learning set. As long as the deviance statistics on the validation set decreases, we are learning additional model structure. Once the deviance statistics on the validation set starts to increase again, we reach the phase of over-fitting where we are not learning (true) model structure anymore but rather peculiarities of the learning set, which is undesirable.

Listing 6: R code for Exercise 11.2 (b) and (c) (Neural network model).

```
### Features, volumes, responses and initial estimate
Ztrain <- model.matrix(data=trainset, ClaimNb ~ VehPower+VehAge+DrivAge)
trainset[,6:30] <- as.data.frame(Ztrain[,-1])
Ztest <- model.matrix(data=testset, ClaimNb ~ VehPower+VehAge+DrivAge)
testset[,6:30] <- as.data.frame(Ztest[,-1])
featlearn <- data.matrix(trainset[,6:30])
feattest <- data.matrix(testset[,6:30])
vollearn <- as.vector(log(trainset$Exposure))
voltest <- as.vector(log(testset$Exposure))
resplearn <- as.vector(trainset$ClaimNb)
resptest <- as.vector(testset$ClaimNb)
lambda0 <- sum(trainset$ClaimNb)/sum(trainset$Exposure)

### Keras model
seed1 <- 100
use_session_with_seed(seed1)
Design <- layer_input(shape=c(25), dtype="float32", name="Design")
LogVol <- layer_input(shape=c(1), dtype="float32", name="LogVol")
Network <- Design %>%
  layer_dense(units=20, activation="tanh") %>%
  layer_dense(units=10, activation="tanh") %>%
  layer_dense(units=1, activation="linear", name="Network",
              weights=list(array(0,dim=c(10,1)), array(log(lambda0),dim=c(1))))
Response <- list(Network, LogVol) %>%
  layer_add(name="Add") %>%
  layer_dense(units=1, activation=k_exp, name="Response", trainable=FALSE,
              weights=list(array(1,dim=c(1,1)), array(0,dim=c(1))))
model <- keras_model(inputs=c(Design, LogVol), outputs=c(Response))
model %>% compile(optimizer=optimizer_nadam(), loss="poisson")

### Prepare features and responses for keras and fit the neural network model
xlearn = list(Design=featlearn, LogVol=vollearn)
ylearn = list(Response=resplearn)
xtest = list(Design=feattest, LogVol=voltest)
ytest = list(Response=resptest)
epochs <- 100   ### c) 1000
model %>% fit(x=xlearn, y=ylearn, epochs=epochs, verbose=1)

### Deviance statistics on training set
predtrain <- as.vector(model %>% predict(xlearn))
obstrain <- trainset$ClaimNb
(Deviancetrain <- 2*sum(log((obstrain/predtrain)^obstrain)-obstrain+predtrain))

### Deviance statistics on test set
predtestNN <- as.vector(model %>% predict(xtest))
obstest <- testset$ClaimNb
(Deviancetest <- 2*sum(log((obstest/predtestNN)^obstest)-obstest+predtestNN))
```

Solution 11.3 Claim Severity Modeling with GLM

(a) In this exercise we work with three tariff criteria. The first criterion (area code) has 6 risk characteristics:

$$\beta_{1,1}\text{ (A)},\quad \beta_{1,2}\text{ (B)},\quad \beta_{1,3}\text{ (C)},\quad \beta_{1,4}\text{ (D)},\quad \beta_{1,5}\text{ (E)}\quad\text{and}\quad \beta_{1,6}\text{ (F)}.$$

The second criterion (brand of the vehicle) has 11 risk characteristics:

$$\beta_{2,1}\text{ (B1)},\ \beta_{2,2}\text{ (B10)},\ \dots,\ \beta_{2,6}\text{ (B14)},\ \beta_{2,7}\text{ (B2)},\ \dots,\ \beta_{2,11}\text{ (B6)}.$$

The third criterion (diesel/fuel) has 2 risk characteristics:

$$\beta_{3,1}\text{ (diesel)}\quad\text{and}\quad \beta_{3,2}\text{ (regular fuel)}.$$

Therefore, we consider risk classes $(l_1, l_2, l_3), 1 \le l_1 \le 6, 1 \le l_2 \le 11, 1 \le l_3 \le 2$. We write n_{l_1,l_2,l_3} for the numbers of claims in risk class (l_1, l_2, l_3) and we only consider risk classes with $n_{l_1,l_2,l_3} > 0$. The n_{l_1,l_2,l_3} individual claim sizes in risk class (l_1, l_2, l_3) are denoted by $Y^{(i)}_{l_1,l_2,l_3}$,

$i = 1, \dots, n_{l_1,l_2,l_3}$. We assume that all $Y^{(i)}_{l_1,l_2,l_3}$ are independent with

$$Y^{(i)}_{l_1,l_2,l_3} \sim \Gamma(\gamma, c_{l_1,l_2,l_3}),$$

where $\gamma > 0$ is a global shape parameter and $c_{l_1,l_2,l_3} > 0$ a risk-class dependent scale parameter. The total claim amount Y_{l_1,l_2,l_3} in risk class (l_1, l_2, l_3) is then given by

$$Y_{l_1,l_2,l_3} = \sum_{i=1}^{n_{l_1,l_2,l_3}} Y^{(i)}_{l_1,l_2,l_3} \sim \Gamma(\gamma n_{l_1,l_2,l_3}, c_{l_1,l_2,l_3}).$$

For the average claim amount X_{l_1,l_2,l_3} in risk class (l_1, l_2, l_3) we have

$$X_{l_1,l_2,l_3} = \frac{Y_{l_1,l_2,l_3}}{n_{l_1,l_2,l_3}} \sim \Gamma(\gamma n_{l_1,l_2,l_3}, c_{l_1,l_2,l_3} n_{l_1,l_2,l_3}).$$

We model

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

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

$$\boldsymbol{\beta} = (\beta_0, \beta_{1,2}, \dots, \beta_{1,6}, \beta_{2,2}, \dots, \beta_{2,11}, \beta_{3,2})' \in \mathbb{R}^{r+1},$$

where $r = 16$. Similarly as in Exercises 10.3 and 11.1, we relabel the risk classes with the index $m \in \{1, \dots, M\}$, where $M = 6 \cdot 11 \cdot 2 = 132$, define $\boldsymbol{X} = (X_1, \dots, X_M)'$ and the design matrix $Z \in \mathbb{R}^{M \times (r+1)}$ that satisfies

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

where the logarithm is applied componentwise to $\mathbb{E}[\boldsymbol{X}]$. Let $m \in \{1, \dots, M\}$. According to Section 7.4.4 of the lecture notes (version of March 20, 2019), X_m belongs to the exponential dispersion family with cumulant function $b(\theta) = -\log(-\theta)$ for $\theta < 0$, $\theta_m = -c_m/\gamma$, $w_m = n_m$ and dispersion parameter $\phi = 1/\gamma$, i.e. we have

$$[Z\boldsymbol{\beta}]_m = \log \mathbb{E}[X_m] = \log \frac{\gamma n_m}{c_m n_m} = \log \frac{\gamma}{c_m} = \log\left(-\frac{1}{\theta_m}\right),$$

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

$$X_m \sim \text{EDF}\left(\theta_m = -\exp\{-[Z\boldsymbol{\beta}]_m\}, \phi = 1/\gamma, w_m = n_m, b(\theta) = -\log(-\theta)\right),$$

for all $m \in \{1, \dots, M\}$. As $b(\theta) = -\log(-\theta)$, we have $b'(\cdot) = -1/\theta$, where b' denotes the first derivative of b. In particular, the log-link function $g(\cdot) = \log(\cdot)$ is not equal to the canonical link function $h(\mu) = (b')^{-1}(\mu) = -1/\mu$ in the gamma model. Therefore, we cannot use equation (7.26) of the lecture notes (version of March 20, 2019) in order to determine the MLE $\widehat{\boldsymbol{\beta}}^{\text{MLE}}$ of $\boldsymbol{\beta}$. However, according to Proposition 7.13 of the lecture notes (version of March 20, 2019), the MLE $\widehat{\boldsymbol{\beta}}^{\text{MLE}}$ of $\boldsymbol{\beta}$ is the solution of

$$Z' V_{\boldsymbol{\theta}} \exp\{Z\boldsymbol{\beta}\} \stackrel{!}{=} Z' V_{\boldsymbol{\theta}} \boldsymbol{X}, \tag{3}$$

where the weight matrix $V_{\boldsymbol{\theta}}$ is given by $V_{\boldsymbol{\theta}} = \text{diag}(-\theta_1 n_1, \dots, -\theta_M n_M)$. Note that assuming a constant scale parameter γ for all risk cells $m = 1, \dots M$, the dispersion parameter $\phi = 1/\gamma$ cancels from the weight matrix defined on page 195 of the lecture notes (version of March 20, 2019). Equation (3) has to be solved numerically. We refer to Listing 7 for the R code used in this exercise. The resulting MLEs of the parameters $\beta_0, \beta_{2,3}, \beta_{2,7}, \beta_{3,2}$ that are (statistically) significantly different from 0 (on a 10% level) are given in the first row of Table 3. We observe that we expect higher claim sizes in regions B11 and B2, compared to the reference region B1. Moreover, claim sizes tend to be higher if a car with regular fuel is involved compared to a diesel car. We remark that the parameters corresponding to the individual categorical levels of the covariate area code are not (statistically) significantly different from 0 (on a 10% level). However, this does not mean that the covariate area code itself is not statistically significant, see part (b).

	$\widehat{\beta}_0$	$\widehat{\beta}_{2,3}$	$\widehat{\beta}_{2,7}$	$\widehat{\beta}_{3,2}$
MLE	7.6116	0.5288	0.1991	0.1846
p-value	≈ 0	0.0585	0.0898	0.0321

Table 3: MLEs of the statistically significant parameters and corresponding p-values.

(b) The estimates of the expected values of X_m are given by

$$\widehat{\mu}_m \;=\; b'(\widehat{\theta}_m) \;=\; -\widehat{\theta}_m^{-1} \;=\; \exp\left\{\left[Z\widehat{\boldsymbol{\beta}}^{\text{MLE}}\right]_m\right\},$$

for all $m = 1, \ldots, M$, and we write $\widehat{\boldsymbol{\mu}} = (\widehat{\mu}_1, \ldots, \widehat{\mu}_M)'$. According to page 196 of the lecture notes (version of March 20, 2019), the deviance statistics is given by

$$\begin{aligned} D(\boldsymbol{X}, \widehat{\boldsymbol{\mu}}) \;&=\; 2\sum_{m=1}^{M} w_m(X_m h(X_m) - b[h(X_m)] - X_m h(\widehat{\mu}_m) + b[h(\widehat{\mu}_m)]) \\ &=\; 2\sum_{m=1}^{M} n_m(-1 - \log X_m + X_m/\widehat{\mu}_m + \log \widehat{\mu}_m). \end{aligned}$$

Estimating ϕ by

$$\widehat{\phi}_D \;=\; \frac{D(\boldsymbol{X}, \widehat{\boldsymbol{\mu}})}{M - r - 1},$$

see page 197 of the lecture notes (version of March 20, 2019), we have for the scaled deviance statistics

$$D^*(\boldsymbol{X}, \widehat{\boldsymbol{\mu}}) \;=\; 2(\ell_{\boldsymbol{X}}(\boldsymbol{X}) - \ell_{\boldsymbol{X}}(\widehat{\boldsymbol{\mu}})) \;=\; \frac{2}{\widehat{\phi}_D}\sum_{m=1}^{M} n_m(-1 - \log X_m + X_m/\widehat{\mu}_m + \log \widehat{\mu}_m).$$

We perform two tests in order to check if there is statistical evidence that the area code could be omitted as tariff criterion. We define the null hypothesis

$$H_0 : \; \beta_{1,2} \;=\; \cdots \;=\; \beta_{1,6} \;=\; 0.$$

Thus, in the reduced model we set the above $p = 5$ variables equal to 0. Then, we can recalculate $\widehat{\boldsymbol{\beta}}_{H_0}^{\text{MLE}}$ for this reduced model and define

$$\widehat{\boldsymbol{\mu}}_{H_0} = \exp\left\{Z_{H_0}\widehat{\boldsymbol{\beta}}_{H_0}^{\text{MLE}}\right\},$$

where Z_{H_0} is the design matrix in the reduced model. According to formula (7.30) of the lecture notes (version of March 20, 2019), the test statistic

$$\text{F} \;=\; \frac{D(\boldsymbol{X}, \widehat{\boldsymbol{\mu}}_{H_0}) - D(\boldsymbol{X}, \widehat{\boldsymbol{\mu}})}{D(\boldsymbol{X}, \widehat{\boldsymbol{\mu}})}\frac{M - r - 1}{p} \;=\; \frac{115}{5}\frac{D(\boldsymbol{X}, \widehat{\boldsymbol{\mu}}_{H_0}) - D(\boldsymbol{X}, \widehat{\boldsymbol{\mu}})}{D(\boldsymbol{X}, \widehat{\boldsymbol{\mu}})}$$

has approximately an F-distribution with degrees of freedom given by $\mathrm{df}_1 = p = 5$ and $\mathrm{df}_2 = M - r - 1 = 115$. We get

$$\mathrm{F} \approx 2.983,$$

which corresponds to a p-value of approximately 1.44%. Thus, using the F-test, we can reject H_0 at significance level of 5%. According to formula (7.31) of the lecture notes (version of March 20, 2019), a second test statistic is given by

$$X^2 = D^*(\boldsymbol{X}, \widehat{\boldsymbol{\mu}}_{H_0}) - D^*(\boldsymbol{X}, \widehat{\boldsymbol{\mu}}).$$

The test statistic X^2 has approximately a χ^2-distribution with $\mathrm{df} = p = 5$ degrees of freedom. We get

$$X^2 \approx 14.917,$$

which corresponds to a p-value of approximately 1.07%. Thus, also using the χ^2-test we can reject H_0 at significance level of 5%. We can conclude that there seems to be no statistical evidence that the area code could be omitted as tariff criterion, even though the individual categorical levels of the covariate area code are not (statistically) significantly different from 0 (on a 10% level), see part (a).

Listing 7: R code for Exercise 11.3.

```
### Apply GLM
d.glm <- glm(ClaimAmount ~ Area+VehBrand+VehGas, data=data, weights=ClaimNb,
             family=Gamma(link="log"))
summary(d.glm)

### Calculate the deviance statistics of the full model
D.full <- d.glm$deviance

### Fit the reduced model and calculate the deviance statistics
d.glm.2 <- glm(ClaimAmount ~ VehGas+VehBrand, data=data, weights=ClaimNb,
               family=Gamma(link="log"))
D.reduced <- d.glm.2$deviance

### Calculate the test statistic F and the corresponding p-value
round((test.stat <- d.glm$df.residual/5*(D.reduced-D.full)/D.full),3)
pf(test.stat, 5, d.glm$df.residual, lower.tail=FALSE)

### Calculate the test statistic X^2 and the corresponding p-value
phi.est <- d.glm$deviance/d.glm$df.residual
round((X.2 <- D.reduced/phi.est-D.full/phi.est),3)
pchisq(X.2, 5, lower.tail=FALSE)
```

Solution 11.4 Neural Networks and Gradient Descent

(a) We model the regression function $\alpha : \mathcal{Z} \to \mathbb{R}_+$ with a a single hidden layer neural network with $r_1 \in \mathbb{N}$ hidden neurons. Our feature space is $\mathcal{Z} \subset \mathbb{R}^{r_0+1}$ with $r_0 = 1$, i.e. we have input dimension $r_0 = 1$. We assume that the first component of the covariates $\boldsymbol{z} = (1, z) \in \mathcal{Z}$ is equal to 1 for modeling an intercept. We define the parameter vectors

$$\boldsymbol{\beta}_j^{(1)} = \left(\beta_{j,0}^{(1)}, \beta_{j,1}^{(1)}\right) \in \mathbb{R}^{r_0+1},$$

for all $j = 1, \ldots, r_1$, and

$$\boldsymbol{\beta}^{(2)} = \left(\beta_0^{(2)}, \beta_1^{(2)}, \ldots, \beta_{r_1}^{(2)}\right) \in \mathbb{R}^{r_1+1}.$$

The hyperbolic tangent activation function is given by

$$\psi(x) = \tanh(x) = \frac{e^{2x} - 1}{e^{2x} + 1}, \qquad \text{for } x \in \mathbb{R}.$$

For covariates $\boldsymbol{z} = (1, z) \in \mathcal{Z}$, the activations in the hidden layer are then given by

$$\boldsymbol{q}^{(1)}(\boldsymbol{z}) = \left(1, q_1^{(1)}(\boldsymbol{z}), \ldots, q_{r_1}^{(1)}(\boldsymbol{z})\right),$$

where

$$q_j^{(1)}(\boldsymbol{z}) = \psi\left(\langle \boldsymbol{\beta}_j^{(1)}, \boldsymbol{z}\rangle\right) = \psi\left(\beta_{j,0}^{(1)} + \beta_{j,1}^{(1)} z\right),$$

for all $j = 1, \ldots, r_1$. Since the codomain of $\alpha(\cdot)$ has to be $\mathbb{R}_+$, we define a log-linear regression approach as follows

$$\alpha(\boldsymbol{z}) = \alpha_{\boldsymbol{\beta}}(\boldsymbol{z}) = \exp\left\{\langle \boldsymbol{\beta}^{(2)}, \boldsymbol{q}^{(1)}(\boldsymbol{z})\rangle\right\} = \exp\left\{\beta_0^{(2)} + \sum_{j=1}^{r_1} \beta_j^{(2)} q_j^{(1)}(\boldsymbol{z})\right\},$$

with resulting network parameter

$$\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^{(1)}, \ldots, \boldsymbol{\beta}_{r_1}^{(1)}, \boldsymbol{\beta}^{(2)}\right) \in \mathbb{R}^{\varrho}$$

having dimension $\varrho = (r_0 + 1)r_1 + r_1 + 1 = (1 + 1)r_1 + r_1 + 1 = 3r_1 + 1$.

(b) As we assume independent Pareto distributions with threshold $\theta > 0$ and covariate-dependent tail index $\alpha(\boldsymbol{z}_m) > 0$ for the data $\boldsymbol{Y} = (Y_1, \ldots, Y_M)$ with corresponding covariates $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_M$, the joint log-likelihood function $\ell_{\boldsymbol{Y}}(\boldsymbol{\beta})$ is given by

$$\begin{aligned}\ell_{\boldsymbol{Y}}(\boldsymbol{\beta}) &= \log \prod_{m=1}^{M} \frac{\alpha_{\boldsymbol{\beta}}(\boldsymbol{z}_m)}{\theta}\left(\frac{Y_m}{\theta}\right)^{-\alpha_{\boldsymbol{\beta}}(\boldsymbol{z}_m)-1} \\ &= \sum_{m=1}^{M} \log \alpha_{\boldsymbol{\beta}}(\boldsymbol{z}_m) - \log\theta - [\alpha_{\boldsymbol{\beta}}(\boldsymbol{z}_m) + 1]\log\frac{Y_m}{\theta}.\end{aligned}$$

In the saturated model we assume one parameter α_m per observation m. This parameter α_m is determined by maximizing the individual MLE for observation m, i.e. we have to maximize

$$g(\alpha_m) \stackrel{\text{def}}{=} \log(\alpha_m) - \log\theta - (\alpha_m + 1)\log\frac{Y_m}{\theta}$$

with respect to α_m, for all $m = 1, \ldots, M$. If we take the derivative with respect to α_m, we get

$$\frac{\partial g(\alpha_m)}{\partial \alpha_m} = \frac{1}{\alpha_m} - \log\frac{Y_m}{\theta},$$

for all $m = 1, \ldots, M$. This is equal to 0 if and only if

$$\alpha_m = \frac{1}{\log\frac{Y_m}{\theta}}, \tag{4}$$

for all $m = 1, \ldots, M$. For the second derivative of $g(\alpha_m)$ with respect to α_m we get

$$\frac{\partial^2 g(\alpha_m)}{\partial \alpha_m^2} = -\frac{1}{\alpha_m^2} < 0,$$

for all $m = 1, \ldots, M$. That is, in the saturated model we have parameter $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)$ with α_m given as in (4), for all $m = 1, \ldots, M$. For the log-likelihood of the saturated model we then have

$$\begin{aligned}\ell_{\boldsymbol{Y}}(\boldsymbol{Y}) &= \sum_{m=1}^{M} \log\frac{1}{\log\frac{Y_m}{\theta}} - \log\theta - \left(\frac{1}{\log\frac{Y_m}{\theta}} + 1\right)\log\frac{Y_m}{\theta} \\ &= \sum_{m=1}^{M} -\log\log\frac{Y_m}{\theta} - \log\theta - 1 - \log\frac{Y_m}{\theta}.\end{aligned}$$

Finally, the (scaled) deviance statistics is given by

$$\mathcal{L}_{\boldsymbol{Y}}(\boldsymbol{\beta}) \;=\; 2(\ell_{\boldsymbol{Y}}(\boldsymbol{Y}) - \ell_{\boldsymbol{Y}}(\boldsymbol{\beta})) \;=\; 2\sum_{m=1}^{M} -\log\log\frac{Y_m}{\theta} - 1 - \log\alpha_{\boldsymbol{\beta}}(\boldsymbol{z}_m) + \alpha_{\boldsymbol{\beta}}(\boldsymbol{z}_m)\log\frac{Y_m}{\theta}.$$

(c) A neural network model with a large number of hidden neurons is heavily over-parametrized. Therefore, a maximum likelihood estimator would lead to overfitting of the model to the data (in-sample). Thus, we are only interested in finding a sufficiently good approximation which has also a good out-of-sample performance. We believe that such a 'good' parametrization can be reached for example by the gradient descent method.

(d) For the derivative of the hyperbolic tangent activation function ψ we have

$$\begin{aligned}\frac{\partial\psi(x)}{\partial x} \;&=\; \frac{\partial}{\partial x}\frac{e^{2x}-1}{e^{2x}+1} \;=\; \frac{2e^{2x}(e^{2x}+1) - 2e^{2x}(e^{2x}-1)}{(e^{2x}+1)^2} \;=\; \frac{4e^{2x}}{(e^{2x}+1)^2}\\ &=\; \frac{(e^{2x}+1)^2 - (e^{2x}-1)^2}{(e^{2x}+1)^2} \;=\; 1-\psi^2(x).\end{aligned}$$

In the gradient descent optimization algorithm the goal is to decrease a given loss function by iteratively updating the model parameters. In our case we would like to decrease the (scaled) deviance statistics $\mathcal{L}_{\boldsymbol{Y}}(\boldsymbol{\beta})$ derived in part (b) above. To this end, for a given $\boldsymbol{\beta}$, we move in the direction of the maximal local decrease of the deviance statistics, i.e. in the direction of the negative gradient $\nabla_{\boldsymbol{\beta}}\mathcal{L}_{\boldsymbol{Y}}(\boldsymbol{\beta})$ of the deviance statistics. We calculate

$$\nabla_{\boldsymbol{\beta}}\mathcal{L}_{\boldsymbol{Y}}(\boldsymbol{\beta}) \;=\; \frac{\partial\mathcal{L}_{\boldsymbol{Y}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \;=\; 2\sum_{m=1}^{M}\left[-\frac{1}{\alpha_{\boldsymbol{\theta}}(\boldsymbol{z}_m)} + \log\frac{Y_m}{\theta}\right]\frac{\partial\alpha_{\boldsymbol{\beta}}(\boldsymbol{z}_m)}{\partial\boldsymbol{\beta}},$$

where we have

$$\begin{aligned}\frac{\partial\alpha_{\boldsymbol{\beta}}(\boldsymbol{z}_m)}{\partial\beta_{j,0}^{(1)}} \;&=\; \alpha_{\boldsymbol{\theta}}(\boldsymbol{z}_m)\beta_j^{(2)}\left(1-\left[q_j^{(1)}(\boldsymbol{z})\right]^2\right),\\ \frac{\partial\alpha_{\boldsymbol{\beta}}(\boldsymbol{z}_m)}{\partial\beta_{j,1}^{(1)}} \;&=\; \alpha_{\boldsymbol{\theta}}(\boldsymbol{z}_m)\beta_j^{(2)}\left(1-\left[q_j^{(1)}(\boldsymbol{z})\right]^2\right)z_m,\\ \frac{\partial\alpha_{\boldsymbol{\beta}}(\boldsymbol{z}_m)}{\partial\beta_0^{(2)}} \;&=\; \alpha_{\boldsymbol{\beta}}(\boldsymbol{z}),\\ \frac{\partial\alpha_{\boldsymbol{\beta}}(\boldsymbol{z}_m)}{\partial\beta_j^{(2)}} \;&=\; \alpha_{\boldsymbol{\beta}}(\boldsymbol{z})q_j^{(1)}(\boldsymbol{z}),\end{aligned}$$

for all $m = 1, \ldots, M$ and $j = 1, \ldots, r_1$. In one single step of the gradient descent optimization algorithm we have the update

$$\boldsymbol{\beta} \;\longrightarrow\; \boldsymbol{\beta} - \rho\nabla_{\boldsymbol{\beta}}\mathcal{L}_{\boldsymbol{Y}}(\boldsymbol{\beta}),$$

where $\rho > 0$ is the so-called learning rate. Note that one should carefully choose an appropriate stopping time of the algorithm in order to prevent from overfitting; and one should also carefully choose $\rho > 0$ because the gradient descent steps lead to a decrease locally.

Non-Life Insurance: Mathematics and Statistics

Solution sheet 12

Solution 12.1 (Inhomogeneous) Credibility Estimators for Claim Counts

First, we note that

$$\mu_0 \overset{\text{def}}{=} \mathbb{E}[\mu(\Theta_i)] \,=\, \mathbb{E}[\Theta_i \lambda_0] \,=\, \lambda_0 \,=\, 0.088.$$

Then, we define

$$X_{i,1} \,=\, \frac{N_{i,1}}{v_{i,1}},$$

for all $i \in \{1, \ldots, 5\}$. 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

$$\mathbb{E}[X_{i,1}^2] \,=\, \mathbb{E}[\mathbb{E}[X_{i,1}^2|\Theta_i]] \,=\, \mathbb{E}\left[\frac{\mu(\Theta_i)}{v_{i,1}} + \mu(\Theta_i)^2\right] \,=\, \frac{\lambda_0}{v_{i,1}} + \tau^2 + \lambda_0^2 \,<\, \infty,$$

for all $i \in \{1, \ldots, 5\}$. In particular, Model Assumptions 8.12 of the lecture notes (version of March 20, 2019) for the Bühlmann-Straub model are satisfied. The (expected) volatility σ^2 within the regions defined in formula (8.4) of the lecture notes (version of March 20, 2019) 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 8.16 of the lecture notes (version of March 20, 2019), the inhomogeneous credibility estimator 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 five 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%

Table 1: Estimated credibility weights $\alpha_{i,T}$, observation based estimates $\widehat{X}_{i,1:T}$ and inhomogeneous credibility estimates $\widehat{\widehat{\mu(\Theta_i)}}$ in regions $i = 1, \dots, 5$.

Note that since the credibility coefficient $\kappa = \sigma^2/\tau^2 \approx 367$ is rather small compared to the volumes $v_{1,1}, \dots, v_{5,1}$, the credibility weights $\alpha_{1,T}, \dots, \alpha_{5,T}$ are fairly high. Moreover, the observation based estimates are almost the same for the regions $1, 2, 3$ and 5, 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 estimates $\widehat{\widehat{\mu(\Theta_2)}}$ and $\widehat{\widehat{\mu(\Theta_5)}}$ from the observation based estimates $\widehat{X}_{2,1:T}$ and $\widehat{X}_{5,1:T}$ towards $\mu_0 = 8.8\%$. If we decreased the volatility τ^2 between the risk classes, the credibility coefficient $\kappa = \sigma^2/\tau^2$ would increase and, thus, the credibility weights $\alpha_{1,T}, \dots, \alpha_{5,T}$ would decrease. Consequently, the credibility estimates would move stronger towards $\mu_0 = 8.8\%$.

(b) Since the number of policies grows 5% in each region, next year's numbers of policies $v_{1,2}, \dots, v_{5,2}$ 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

Table 2: Next year's numbers of policies in regions $i = 1, \dots, 5$.

Similarly to part (a), we define

$$X_{i,2} = \frac{N_{i,2}}{v_{i,2}},$$

for all $i \in \{1, \dots, 5\}$. According to the exercise sheet, next year's numbers of claims stay within the Bühlmann-Straub model framework assumed for this year's numbers of claims. Thus, according to formula (8.16) of the lecture notes (version of March 20, 2019), the mean square error of prediction is given by, for all $i \in \{1, \dots, 5\}$,

$$\mathbb{E}\left[\left(\frac{N_{i,2}}{v_{i,2}} - \widehat{\widehat{\mu(\Theta_i)}}\right)^2\right] = \mathbb{E}\left[\left(X_{i,2} - \widehat{\widehat{\mu(\Theta_i)}}\right)^2\right] = \frac{\sigma^2}{v_{i,2}} + (1 - \alpha_{i,T})\,\tau^2. \qquad (1)$$

We get the following root mean square errors of prediction for the five regions:

	region 1	region 2	region 3	region 4	region 5
$\sqrt{\text{mean square errors of prediction}}$	0.185%	0.408%	0.119%	0.221%	0.627%
in % of the credibility estimates	2.4%	5.2%	1.6%	2.2%	8.3%

Table 3: Root mean square errors of prediction in regions $i = 1, \dots, 5$.

Note that we get the highest root mean square errors of prediction for 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 formula (1).

Solution 12.2 (Homogeneous) Credibility Estimators for Claim Sizes

We define

$$X_{i,t} = \frac{Y_{i,t}}{v_{i,t}},$$

for all $i \in \{1,2,3,4\}$ and $t \in \{1,2\}$. 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

$$\mathrm{Var}(X_{i,t}|\Theta_i) = \frac{1}{v_{i,t}^2} \mathrm{Var}(Y_{i,t}|\Theta_i) = \frac{1}{v_{i,t}^2} \frac{\mu(\Theta_i) c v_{i,t}}{c^2} = \frac{\mu(\Theta_i)}{c v_{i,t}} = \frac{\sigma^2(\Theta_i)}{v_{i,t}},$$

with

$$\sigma^2(\Theta_i) = \frac{\mu(\Theta_i)}{c} = \frac{\Theta_i}{c},$$

for all $i \in \{1,2,3,4\}$ and $t \in \{1,2\}$. Moreover, using that

$$\mathbb{E}[X_{i,t}^2|\Theta_i] = \mathrm{Var}(X_{i,t}|\Theta_i) + \mathbb{E}[X_{i,t}|\Theta_i]^2 = \frac{\mu(\Theta_i)}{c v_{i,t}} + \mu(\Theta_i)^2 = \frac{\Theta_i}{c v_{i,t}} + \Theta_i^2,$$

we get

$$\mathbb{E}[X_{i,t}^2] = \mathbb{E}[\mathbb{E}[X_{i,t}^2|\Theta_i]] = \mathbb{E}\left[\frac{\Theta_i}{c v_{i,t}} + \Theta_i^2\right] < \infty$$

by assumption, for all $i \in \{1,2,3,4\}$ and $t \in \{1,2\}$. In particular, Model Assumptions 8.12 of the lecture notes (version of March 20, 2019) for the Bühlmann-Straub model are satisfied.

(a) First, following Theorem 8.16 of the lecture notes (version of March 20, 2019), 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 $i \in \{1,2,3,4\}$. Then, we need to estimate the structural parameters $\sigma^2 = \mathbb{E}[\sigma^2(\Theta_1)]$ and $\tau^2 = \mathrm{Var}(\mu(\Theta_1))$. According to formula (8.14) of the lecture notes (version of March 20, 2019), σ^2 can be estimated by

$$\widehat{\sigma}_T^2 = \frac{1}{I} \sum_{i=1}^I \frac{1}{T-1} \sum_{t=1}^T v_{i,t} \left(X_{i,t} - \widehat{X}_{i,1:T}\right)^2 \approx 1.3 \cdot 10^{10}.$$

In order to estimate τ^2, we define first the weighted sample mean $\bar{X}$ over all observations by

$$\bar{X} = \frac{1}{\sum_{i=1}^I \sum_{t=1}^T v_{i,t}} \sum_{i=1}^I \sum_{t=1}^T v_{i,t} X_{i,t} = \frac{\sum_{i=1}^I Y_{i,1} + Y_{i,2}}{\sum_{i=1}^I v_{i,1} + v_{i,2}} \approx 7'004.$$

Then, following the lecture notes, we define $\widehat{v}_T^2$, c_w and $\widehat{t}_T^2$ as

$$\widehat{v}_T^2 = \frac{I}{I-1} \sum_{i=1}^I \frac{v_{i,1} + v_{i,2}}{\sum_{j=1}^I v_{j,1} + v_{j,2}} \left(\widehat{X}_{i,1:T} - \bar{X}\right)^2 \approx 9.3 \cdot 10^7,$$

$$c_w = \frac{I-1}{I} \left[\sum_{i=1}^I \frac{v_{i,1} + v_{i,2}}{\sum_{j=1}^I v_{j,1} + v_{j,2}} \left(1 - \frac{v_{i,1} + v_{i,2}}{\sum_{j=1}^I v_{j,1} + v_{j,2}}\right)\right]^{-1} \approx 1.425$$

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 (8.15) of the lecture notes (version of March 20, 2019), τ^2 is 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 let $i \in \{1,2,3,4\}$. According to Theorem 8.16 of the lecture notes (version of March 20, 2019), the homogeneous credibility estimator 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 $\widehat{\mu}_T$

$$\alpha_{i,T} \;=\; \frac{v_{i,1} + v_{i,2}}{v_{i,1} + v_{i,2} + \widehat{\sigma}_T^2/\widehat{\tau}_T^2} \qquad \text{and} \qquad \widehat{\mu}_T \;=\; \frac{1}{\sum_{i=1}^I \alpha_{i,T}} \sum_{i=1}^I \alpha_{i,T}\widehat{X}_{i,1:T} \;\approx\; 14\text{'}538.$$

The results for the four 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%	82.5%	89.6%
$\widehat{X}_{i,1:T}$	10'493	1'907	18'375	29'197
$\widehat{\widehat{\mu(\Theta_i)}}^{\text{hom}}$	10'677	2'107	17'702	27'665

Table 4: Estimated credibility weights $\alpha_{i,T}$, observation based estimates $\widehat{X}_{i,1:T}$ and homogeneous credibility estimates $\widehat{\widehat{\mu(\Theta_i)}}^{\text{hom}}$ in risk classes $i = 1,2,3,4$.

Looking at the credibility weights $\alpha_{1,T}, \alpha_{2,T}, \alpha_{3,T}$ and $\alpha_{4,T}$, 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 estimates $\widehat{X}_{1,1:T}, \widehat{X}_{2,1:T}, \widehat{X}_{3,1:T}$ and $\widehat{X}_{4,1:T}$ towards $\hat{\mu}_T \approx 14\text{'}538$ is strongest for risk classes 3 and 4.

(b) Since the number of claims grows 5% in each risk class, 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

Table 5: Next year's numbers of claims in risk classes $i = 1,2,3,4$.

Similarly to part (a), we define

$$X_{i,3} \;=\; \frac{Y_{i,3}}{v_{i,3}},$$

for all $i \in \{1,2,3,4\}$. According to the exercise sheet, next year's total claim sizes stay within the Bühlmann-Straub model framework assumed for the previous year's total claim sizes. Thus, according to formula (8.17) of the lecture notes (version of March 20, 2019), the mean square error of prediction can be estimated by, for all $i \in \{1, 2, 3, 4\}$,

$$\widehat{\mathbb{E}}\left[\left(\frac{Y_{i,3}}{v_{i,3}} - \widehat{\widehat{\mu(\Theta_i)}}^{\text{hom}}\right)^2\right] = \widehat{\mathbb{E}}\left[\left(X_{i,3} - \widehat{\widehat{\mu(\Theta_i)}}^{\text{hom}}\right)^2\right] = \frac{\widehat{\sigma}_T^2}{v_{i,3}} + (1-\alpha_{i,T})\widehat{\tau}_T^2\left(1 + \frac{1-\alpha_{i,T}}{\sum_{i=1}^{I}\alpha_{i,T}}\right). \tag{2}$$

We get the following estimated root mean square errors of prediction for the four risk classes:

	risk class 1	risk class 2	risk class 3	risk class 4
$\sqrt{\text{mean square errors of prediction}}$	4'108	2'392	8'508	6'360
in % of the credibility estimates	38.5%	113.5%	48.1%	23.0%

Table 6: Estimated root mean square errors of prediction in risk classes $i = 1, 2, 3, 4$.

According to formula (2), the smaller the volumes in a particular risk class, the bigger the corresponding estimated root mean square error of prediction. Moreover, note that these estimated root mean square errors of prediction are rather high compared to the credibility estimates, which indicates a high variability within the individual risk classes.

Solution 12.3 Degenerate MLE and the Poisson-Gamma Model

(a) We observe that $N_t = 0$ for all $t = 1, \ldots, T$. In this case, the log-likelihood function $\ell_{\boldsymbol{N}}(\lambda)$ of the data $\boldsymbol{N} = (N_1, \ldots, N_T)$ for the unknown parameter $\lambda > 0$ is given by

$$\ell_{\boldsymbol{N}}(\lambda) = \sum_{t=1}^{T} \log\left(\exp\{-\lambda v_t\}\frac{(\lambda v_t)^{N_t}}{N_t!}\right) = \sum_{t=1}^{T} \log\left(\exp\{-\lambda v_t\}\right) = -\lambda\sum_{t=1}^{T} v_t.$$

As the volumes $v_1, \ldots, v_T$ are positive, we see that $\ell_{\boldsymbol{N}}(\lambda)$ increases as λ decreases, i.e. here we are in the situation of a degenerate Poisson model with MLE $\widehat{\lambda}_T = 0$. If we used this degenerate model for premium calculations, we would get a pure risk premium of 0, as we do not expect any claims. Of course, a model with zero pure risk premium does not make any sense, i.e. we need to circumvent this degenerate case. This can be done for example with the Poisson-gamma model considered in part (b).

(b) (i) The prior estimator λ_0 of the unknown parameter Λ is given by

$$\lambda_0 = \mathbb{E}[\Lambda] = \frac{\gamma}{c} = \frac{1}{50}.$$

According to Theorem 8.2 of the lecture notes (version of March 20, 2019), we have

$$\Lambda|\boldsymbol{N} \sim \Gamma\left(\gamma + \sum_{t=1}^{T} N_t, c + \sum_{t=1}^{T} v_t\right),$$

where we write $\boldsymbol{N} = (N_1, \ldots, N_T)$. Therefore, the posterior estimator $\widehat{\lambda}_T^{\text{post}}$ of the unknown parameter Λ is given by

$$\widehat{\lambda}_T^{\text{post}} = \mathbb{E}[\Lambda|\boldsymbol{N}] = \frac{\gamma + \sum_{t=1}^{T} N_t}{c + \sum_{t=1}^{T} v_t} = \frac{1+0}{50+50} = \frac{1}{100}.$$

(ii) According to Corollary 8.4 of the lecture notes (version of March 20, 2019), we can write

$$\widehat{\lambda}_T^{\text{post}} = \alpha_T \, \widehat{\lambda}_T + (1 - \alpha_T) \, \lambda_0$$

by setting

$$\alpha_T = \frac{\sum_{t=1}^T v_t}{c + \sum_{t=1}^T v_t} \in (0, 1).$$

In our case we get

$$\alpha_T = \frac{50}{50 + 50} = \frac{1}{2}.$$

Indeed, we check

$$\alpha_T \, \widehat{\lambda}_T + (1 - \alpha_T) \, \lambda_0 = \frac{1}{2} \cdot 0 + \left(1 - \frac{1}{2}\right) \cdot \frac{1}{50} = \frac{1}{100} = \widehat{\lambda}_T^{\text{post}}.$$

(iii) Similarly as in item (i), for the posterior estimator $\widehat{\lambda}_{T+1}^{\text{post}}$, conditionally given data $(N_1, v_1), \ldots, (N_{T+1}, v_{T+1})$, we get

$$\widehat{\lambda}_{T+1}^{\text{post}} = \frac{\gamma + \sum_{t=1}^{T+1} N_t}{c + \sum_{t=1}^{T+1} v_t} = \frac{1 + 1}{50 + 60} = \frac{2}{110}.$$

According to Corollary 8.6 of the lecture notes (version of March 20, 2019), we can write

$$\widehat{\lambda}_{T+1}^{\text{post}} = \beta_{T+1} \, \frac{N_{T+1}}{v_{T+1}} + (1 - \beta_{T+1}) \, \widehat{\lambda}_T^{\text{post}}$$

by setting

$$\beta_{T+1} = \frac{v_{T+1}}{c + \sum_{t=1}^{T+1} v_t} \in (0, 1).$$

In our case we get

$$\beta_{T+1} = \frac{10}{50 + 60} = \frac{1}{11}.$$

Indeed, we check

$$\beta_{T+1} \, \frac{N_{T+1}}{v_{T+1}} + (1 - \beta_{T+1}) \, \widehat{\lambda}_T^{\text{post}} = \frac{1}{11} \frac{1}{10} + \left(1 - \frac{1}{11}\right) \frac{1}{100} = \frac{1}{110} + \frac{1}{110} = \frac{2}{110} = \widehat{\lambda}_{T+1}^{\text{post}}.$$

(c) Note that, by definition, a Poisson random variable requires a positive frequency parameter. In case of a frequency parameter which is equal to 0, we are in the degenerate Poisson model, see also part (a). However, if $\Lambda \sim \mathcal{N}(\mu, \sigma^2)$, then the probability that Λ is negative is given by

$$\mathbb{P}[\Lambda < 0] = \mathbb{P}\left[\frac{\Lambda - \mu}{\sigma} < -\frac{\mu}{\sigma}\right] = \Phi\left(-\frac{\mu}{\sigma}\right) > 0,$$

where Φ denotes the distribution function of a standard Gaussian distribution. As there is a positive probability that the frequency parameter Λ is negative, we conclude that a Poisson-normal model is not well-defined and, thus, not reasonable.

Solution 12.4 Pareto-Gamma Model

(a) Let $f_{\boldsymbol{Y}|\Lambda}$ denote the density of $\boldsymbol{Y}|\Lambda$ and f_Λ the density of Λ. Then, we have

$$\begin{aligned} f_{\boldsymbol{Y}|\Lambda}(y_1,\ldots,y_T|\Lambda=\alpha) &= \prod_{t=1}^T \frac{\alpha}{\theta}\left(\frac{y_t}{\theta}\right)^{-(\alpha+1)} \cdot 1_{\{y_t\geq\theta\}} \\ &= \alpha^T\theta^{-T}\left(\prod_{t=1}^T \frac{y_t}{\theta}\right)^{-\alpha}\left(\prod_{t=1}^T \frac{y_t}{\theta}\right)^{-1} \cdot 1_{\{y_t\geq\theta\}} \end{aligned}$$

and

$$f_\Lambda(\alpha) = \frac{c^\gamma}{\Gamma(\gamma)}\alpha^{\gamma-1}\exp\{-c\alpha\}\cdot 1_{\{\alpha>0\}}.$$

Let $f_{\Lambda|\boldsymbol{Y}}$ denote the density of $\Lambda|\boldsymbol{Y}$. Then, for all $\alpha>0$ and $y_1,\ldots,y_T\geq\theta$, we have

$$\begin{aligned} f_{\Lambda|\boldsymbol{Y}}(\alpha|Y_1=y_1,\ldots,Y_T=y_T) &= \frac{f_{\boldsymbol{Y}|\Lambda}(y_1,\ldots,y_T|\Lambda=\alpha)\, f_\Lambda(\alpha)}{\int_0^\infty f_{\boldsymbol{Y}|\Lambda}(y_1,\ldots,y_T|\Lambda=x)\, f_\Lambda(x)\,dx} \\ &\propto \alpha^T\left(\prod_{t=1}^T\frac{y_t}{\theta}\right)^{-\alpha}\alpha^{\gamma-1}\exp\{-c\alpha\} \\ &= \alpha^{\gamma+T-1}\exp\left\{-\alpha\sum_{t=1}^T\log\frac{y_t}{\theta}\right\}\exp\{-c\alpha\} \\ &= \alpha^{\gamma+T-1}\exp\left\{-\alpha\left(\sum_{t=1}^T\log\frac{y_t}{\theta}+c\right)\right\}. \end{aligned}$$

We conclude that

$$\Lambda|\boldsymbol{Y} \sim \Gamma\left(\gamma+T, c+\sum_{t=1}^T\log\frac{Y_t}{\theta}\right).$$

(b) First, we observe that

$$\lambda_0 = \mathbb{E}[\Lambda] = \frac{\gamma}{c} \qquad \text{and} \qquad \widehat{\lambda}_T^{\text{post}} = \mathbb{E}[\Lambda|\boldsymbol{Y}] = \frac{\gamma+T}{c+\sum_{t=1}^T\log\frac{Y_t}{\theta}}.$$

Then, we can write

$$\begin{aligned} \widehat{\lambda}_T^{\text{post}} &= \frac{\gamma+T}{c+\sum_{t=1}^T\log\frac{Y_t}{\theta}} = \frac{\sum_{t=1}^T\log\frac{Y_t}{\theta}}{c+\sum_{t=1}^T\log\frac{Y_t}{\theta}}\,\frac{T}{\sum_{t=1}^T\log\frac{Y_t}{\theta}} + \frac{c}{c+\sum_{t=1}^T\log\frac{Y_t}{\theta}}\,\frac{\gamma}{c} \\ &= \alpha_T\,\widehat{\lambda}_T + (1-\alpha_T)\,\lambda_0, \end{aligned}$$

with

$$\alpha_T = \frac{\sum_{t=1}^T\log\frac{Y_t}{\theta}}{c+\sum_{t=1}^T\log\frac{Y_t}{\theta}}.$$

(c) For the (conditional mean square error) uncertainty of the posterior estimator $\widehat{\lambda}_T^{\text{post}} = \mathbb{E}[\Lambda|\boldsymbol{Y}]$ we have

$$\begin{aligned} \mathbb{E}\left[\left(\Lambda-\widehat{\lambda}_T^{\text{post}}\right)^2\middle|\boldsymbol{Y}\right] &= \mathbb{E}\left[(\Lambda-\mathbb{E}[\Lambda|\boldsymbol{Y}])^2\middle|\boldsymbol{Y}\right] = \mathrm{Var}(\Lambda|\boldsymbol{Y}) = \frac{\gamma+T}{\left(c+\sum_{t=1}^T\log\frac{Y_t}{\theta}\right)^2} \\ &= \frac{1}{c+\sum_{t=1}^T\log\frac{Y_t}{\theta}}\,\widehat{\lambda}_T^{\text{post}} = (1-\alpha_T)\,\frac{1}{c}\,\widehat{\lambda}_T^{\text{post}}. \end{aligned}$$

(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 $(Y_1, \dots, Y_{T-1})$ is given by

$$\widehat{\lambda}_{T-1}^{\text{post}} = \frac{\gamma + T - 1}{c + \sum_{t=1}^{T-1} \log \frac{Y_t}{\theta}}.$$

Thus, we can write

$$\begin{aligned}\widehat{\lambda}_T^{\text{post}} &= \frac{\gamma + T}{c + \sum_{t=1}^{T} \log \frac{Y_t}{\theta}} = \frac{\log \frac{Y_T}{\theta}}{c + \sum_{t=1}^{T} \log \frac{Y_t}{\theta}} \frac{1}{\log \frac{Y_T}{\theta}} + \frac{c + \sum_{t=1}^{T-1} \log \frac{Y_t}{\theta}}{c + \sum_{t=1}^{T} \log \frac{Y_t}{\theta}} \frac{\gamma + T - 1}{c + \sum_{t=1}^{T-1} \log \frac{Y_t}{\theta}} \\ &= \beta_T \frac{1}{\log \frac{Y_T}{\theta}} + (1 - \beta_T)\, \widehat{\lambda}_{T-1}^{\text{post}},\end{aligned}$$

with

$$\beta_T = \frac{\log \frac{Y_T}{\theta}}{c + \sum_{t=1}^{T} \log \frac{Y_t}{\theta}}.$$

<u>Remark:</u> Suppose we want to use the observations $Y_1, \dots, Y_{T-1}$ in order to estimate Y_T in a Bayesian sense. Then, we have

$$\begin{aligned}\mathbb{E}[Y_T | Y_1, \dots, Y_{T-1}] &= \mathbb{E}\left[\mathbb{E}\left[Y_T | Y_1, \dots, Y_{T-1}, \Lambda\right] | Y_1, \dots, Y_{T-1}\right] \quad \text{a.s.} \\ &= \mathbb{E}\left[\mathbb{E}\left[Y_T | \Lambda\right] | Y_1, \dots, Y_{T-1}\right] \quad \text{a.s.},\end{aligned}$$

where in the second equality we used that, conditionally given Λ, $Y_1, \dots, Y_T$ 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) (for $T-1$ instead of T observations), we have

$$\Lambda | (Y_1, \dots, Y_{T-1}) \sim \Gamma\left(\gamma + T - 1, c + \sum_{t=1}^{T-1} \log \frac{Y_t}{\theta}\right).$$

Since the range of a gamma distribution is the whole positive real line, this implies that

$$0 < \mathbb{P}\left[\Lambda \leq 1 | Y_1, \dots, Y_{T-1}\right] = \mathbb{P}\left[\mathbb{E}\left[Y_T | \Lambda\right] = \infty | Y_1, \dots, Y_{T-1}\right] \quad \text{a.s.}$$

We conclude that

$$\mathbb{E}[Y_T | Y_1, \dots, Y_{T-1}] = \infty \quad \text{a.s.}$$

Non-Life Insurance: Mathematics and Statistics

Solution sheet 13

Solution 13.1 Chain-Ladder Algorithm

(a) According to formula (9.5) of the lecture notes (version of March 20, 2019), the CL factor f_j can be estimated by

$$\widehat{f}_j^{\mathrm{CL}} = \frac{\sum_{i=1}^{I-j-1} C_{i,j+1}}{\sum_{i=1}^{I-j-1} C_{i,j}} = \sum_{i=1}^{I-j-1} \frac{C_{i,j}}{\sum_{n=1}^{I-j-1} C_{n,j}} \frac{C_{i,j+1}}{C_{i,j}},$$

for all $j = 0, \ldots, J-1$. Then, for all $i = 2, \ldots, I$ and $j = 1, \ldots, J$ with $i + j > I$, $C_{i,j}$ can be predicted by

$$\widehat{C}_{i,j}^{\mathrm{CL}} = C_{i,I-i} \prod_{k=I-i}^{j-1} \widehat{f}_k^{\mathrm{CL}},$$

see formula (9.6) of the lecture notes (version of March 20, 2019). In particular, for the prediction $\widehat{C}_{i,J}^{\mathrm{CL}}$ of the ultimate claim $C_{i,J}$ we have, for all $i = 2, \ldots, I$,

$$\widehat{C}_{i,J}^{\mathrm{CL}} = C_{i,I-i} \prod_{j=I-i}^{J-1} \widehat{f}_j^{\mathrm{CL}}. \tag{1}$$

The estimates $\widehat{f}_0^{\mathrm{CL}}, \ldots, \widehat{f}_{J-1}^{\mathrm{CL}}$ and the prediction of the lower triangle $\mathcal{D}_I^c$ are given in Table 1. We see that $\widehat{f}_0^{\mathrm{CL}} \approx 1.5$, while $\widehat{f}_j^{\mathrm{CL}}$ is close to 1 for all $j = 1, \ldots, J-1$, i.e. we observe a rather fast claims settlement in this example.

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
$\widehat{f}_j^{\mathrm{CL}}$	1.493	1.078	1.023	1.015	1.007	1.005	1.001	1.001	1.001	

Table 1: Estimates $\widehat{f}_0^{\mathrm{CL}}, \ldots, \widehat{f}_{J-1}^{\mathrm{CL}}$ and prediction of the lower triangle $\mathcal{D}_I^c$.

(b) The CL reserves $\widehat{\mathcal{R}}_i^{\mathrm{CL}}$ at time $t = I$ are given by

$$\widehat{\mathcal{R}}_i^{\mathrm{CL}} = \widehat{C}_{i,J}^{\mathrm{CL}} - C_{i,I-i} = C_{i,I-i} \left(\prod_{j=I-i}^{J-1} \widehat{f}_j^{\mathrm{CL}} - 1 \right),$$

for all accident years $i = 2, \ldots, I$. Moreover, since $C_{1,J} = C_{1,I-1}$ is known, we have $\widehat{\mathcal{R}}_1^{\mathrm{CL}} = 0$.

Summarizing, we get the following CL reserves $\widehat{\mathcal{R}}_i^{\mathrm{CL}}$:

accident year i	1	2	3	4	5	6	7	8	9	10
CL reserves $\widehat{\mathcal{R}}_i^{\mathrm{CL}}$	0	15'126	26'257	34'538	85'302	156'494	286'121	449'167	1'043'242	3'950'815

Table 2: CL reserves $\widehat{\mathcal{R}}_i^{\mathrm{CL}}$ for all accident years $i = 1, \ldots, I$.

By aggregating the CL reserves over all accident years, we get the CL predictor $\widehat{\mathcal{R}}^{\mathrm{CL}}$ for the outstanding loss liabilities of past exposure claims:

$$\widehat{\mathcal{R}}^{\mathrm{CL}} = \sum_{i=1}^{I} \widehat{\mathcal{R}}_i^{\mathrm{CL}} = 6'047'061.$$

Solution 13.2 Bornhuetter-Ferguson Algorithm

(a) Let $C_0 > 0$ be some initial value for development period $j = 0$. Then, for all $j = 0, \ldots, J-1$ we define $\widehat{\beta}_j^{\mathrm{CL}}$ to be the proportion paid after the first j development periods according to the estimated CL pattern from Exercise 13.1. In particular, we calculate

$$\widehat{\beta}_0^{\mathrm{CL}} = \frac{C_0}{C_0 \prod_{l=0}^{J-1} \widehat{f}_l^{\mathrm{CL}}} = \frac{1}{\prod_{l=0}^{J-1} \widehat{f}_l^{\mathrm{CL}}} = \prod_{l=0}^{J-1} \frac{1}{\widehat{f}_l^{\mathrm{CL}}}$$

and

$$\widehat{\beta}_j^{\mathrm{CL}} = \frac{C_0 \prod_{l=0}^{j-1} \widehat{f}_l^{\mathrm{CL}}}{C_0 \prod_{l=0}^{J-1} \widehat{f}_l^{\mathrm{CL}}} = \frac{\prod_{l=0}^{j-1} \widehat{f}_l^{\mathrm{CL}}}{\prod_{l=0}^{J-1} \widehat{f}_l^{\mathrm{CL}}} = \prod_{l=j}^{J-1} \frac{1}{\widehat{f}_l^{\mathrm{CL}}},$$

for all $j = 1, \ldots, J-1$. We get the following proportions:

development period j	0	1	2	3	4	5	6	7	8
proportion $\widehat{\beta}_j^{\mathrm{CL}}$ paid so far	0.590	0.880	0.948	0.970	0.984	0.991	0.996	0.998	0.999

Table 3: Proportions $\widehat{\beta}_j^{\mathrm{CL}}$ paid after the first j development periods according to the estimated CL pattern from Exercise 13.1.

According to formula (9.8) of the lecture notes (version of March 20, 2019), in the Bornhuetter-Ferguson method the ultimate claim $C_{i,J}$ is predicted by

$$\widehat{C}_{i,J}^{\mathrm{BF}} = C_{i,I-i} + \widehat{\mu}_i \left(1 - \widehat{\beta}_{I-i}^{\mathrm{CL}}\right),$$

for all accident years $i = 2, \ldots, I$. Thus, the Bornhuetter-Ferguson reserves $\widehat{\mathcal{R}}_i^{\mathrm{BF}}$ are given by

$$\widehat{\mathcal{R}}_i^{\mathrm{BF}} = \widehat{C}_{i,J}^{\mathrm{BF}} - C_{i,I-i} = \widehat{\mu}_i \left(1 - \widehat{\beta}_{I-i}^{\mathrm{CL}}\right),$$

for all accident years $i = 2, \ldots, I$. Moreover, since $C_{1,J} = C_{1,I-1}$ is known, we have $\widehat{\mathcal{R}}_1^{\mathrm{BF}} = 0$.

Summarizing, we get the following BF reserves $\widehat{\mathcal{R}}_i^{\mathrm{BF}}$:

accident year i	1	2	3	4	5	6	7	8	9	10
BF reserves $\widehat{\mathcal{R}}_i^{\mathrm{BF}}$	0	16'124	26'998	37'575	95'434	178'024	341'305	574'089	1'318'646	4'768'384

Table 4: BF reserves $\widehat{\mathcal{R}}_i^{\mathrm{BF}}$ for all accident years $i = 1, \ldots, I$.

By aggregating the BF reserves over all accident years, we get the BF predictor $\widehat{\mathcal{R}}^{\mathrm{BF}}$ for the outstanding loss liabilities of past exposure claims:

$$\widehat{\mathcal{R}}^{\mathrm{BF}} = \sum_{i=1}^{I} \widehat{\mathcal{R}}_i^{\mathrm{BF}} = 7'356'580.$$

(b) Note that for accident year $i = 1$ we have

$$\widehat{\mathcal{R}}_1^{\mathrm{CL}} = 0 = \widehat{\mathcal{R}}_1^{\mathrm{BF}}.$$

Now let $i = 2, \ldots, I$. Then, we observe that

$$\widehat{\mathcal{R}}_i^{\mathrm{CL}} < \widehat{\mathcal{R}}_i^{\mathrm{BF}}.$$

This can be explained as follows: Equation (1) can be rewritten as

$$\begin{aligned} \widehat{C}_{i,J}^{\mathrm{CL}} &= C_{i,I-i} \prod_{j=I-i}^{J-1} \widehat{f}_j^{\mathrm{CL}} = C_{i,I-i} + C_{i,I-i} \left(\prod_{j=I-i}^{J-1} \widehat{f}_j^{\mathrm{CL}} - 1 \right) \\ &= C_{i,I-i} + C_{i,I-i} \prod_{j=I-i}^{J-1} \widehat{f}_j^{\mathrm{CL}} \left(1 - \prod_{j=I-i}^{J-1} \frac{1}{\widehat{f}_j^{\mathrm{CL}}} \right) = C_{i,I-i} + \widehat{C}_{i,J}^{\mathrm{CL}} \left(1 - \widehat{\beta}_{I-i}^{\mathrm{CL}} \right). \end{aligned}$$

Comparing this to

$$\widehat{C}_{i,J}^{\mathrm{BF}} = C_{i,I-i} + \widehat{\mu}_i \left(1 - \widehat{\beta}_{I-i}^{\mathrm{CL}} \right)$$

and noting that for the prior information $\widehat{\mu}_i$ we have $\widehat{\mu}_i > \widehat{C}_{i,J}^{\mathrm{CL}}$, we immediately see that

$$\widehat{C}_{i,J}^{\mathrm{CL}} < \widehat{C}_{i,J}^{\mathrm{BF}},$$

which of course implies that

$$\widehat{\mathcal{R}}_i^{\mathrm{CL}} = \widehat{C}_{i,J}^{\mathrm{CL}} - C_{i,I-i} < \widehat{C}_{i,J}^{\mathrm{BF}} - C_{i,I-i} = \widehat{\mathcal{R}}_i^{\mathrm{BF}}.$$

We conclude that choosing a prior information $\widehat{\mu}_i$ which is bigger than the estimated CL ultimate $\widehat{C}_{i,J}^{\mathrm{CL}}$ leads to more conservative, i.e. higher reserves in the Bornhuetter-Ferguson method compared to the chain-ladder method.

Solution 13.3 Over-Dispersed Poisson Model

(a) According to Theorem 9.11 of the lecture notes (version of March 20, 2019), the MLEs $\widehat{\mu}_1^{\text{MLE}}, \dots, \widehat{\mu}_I^{\text{MLE}}$ and $\widehat{\gamma}_0^{\text{MLE}}, \dots, \widehat{\gamma}_J^{\text{MLE}}$ of $\mu_1, \dots, \mu_I$ and $\gamma_0, \dots, \gamma_J$ are given by

$$\widehat{\mu}_i^{\text{MLE}} = \widehat{C}_{i,J}^{\text{CL}} \qquad \text{and} \qquad \widehat{\gamma}_j^{\text{MLE}} = \left(1 - \frac{1}{\widehat{f}_{j-1}^{\text{CL}}}\right) \prod_{k=j}^{J-1} \frac{1}{\widehat{f}_k^{\text{CL}}},$$

for all $i = 1, \dots, I$ and $j = 1, \dots, J-1$, where $\widehat{C}_{i,J}^{\text{CL}}$ is the prediction of the ultimate claim $C_{i,J}$ and $\widehat{f}_j$ the estimated CL factor f_j from the chain-ladder model of Exercise 13.1. Moreover, we have

$$\widehat{\gamma}_0^{\text{MLE}} = \prod_{k=0}^{J-1} \frac{1}{\widehat{f}_k^{\text{CL}}} \qquad \text{and} \qquad \widehat{\gamma}_J^{\text{MLE}} = \left(1 - \frac{1}{\widehat{f}_{J-1}^{\text{CL}}}\right).$$

The values of the MLEs $\widehat{\mu}_1^{\text{MLE}}, \dots, \widehat{\mu}_I^{\text{MLE}}$ are given in Table 5, the values of the MLEs $\widehat{\gamma}_0^{\text{MLE}}, \dots, \widehat{\gamma}_J^{\text{MLE}}$ in Table 6.

accident year i	1	2	3	4	5
MLE $\widehat{\mu}_i^{\text{MLE}}$	11'148'124	10'663'318	10'662'008	9'758'606	9'872'218
accident year i	6	7	8	9	10
MLE $\widehat{\mu}_i^{\text{MLE}}$	10'092'247	9'568'143	8'705'378	8'691'971	9'626'383

Table 5: Values of the MLEs $\widehat{\mu}_1^{\text{MLE}}, \dots, \widehat{\mu}_I^{\text{MLE}}$.

development year j	0	1	2	3	4	5	6	7	8	9
MLE $\widehat{\gamma}_j^{\text{MLE}}$	0.590	0.290	0.068	0.022	0.014	0.007	0.005	0.001	0.001	0.001

Table 6: Values of the MLEs $\widehat{\gamma}_0^{\text{MLE}}, \dots, \widehat{\gamma}_J^{\text{MLE}}$.

(b) According to Theorem 9.11 of the lecture notes (version of March 20, 2019), the ODP reserves $\widehat{\mathcal{R}}_i^{\text{ODP}}$ are given by

$$\widehat{\mathcal{R}}_i^{\text{ODP}} = \widehat{\mu}_i^{\text{MLE}} \sum_{j=I-i+1}^{J} \widehat{\gamma}_j^{\text{MLE}},$$

for all accident years $i = 2, \dots, I$. Moreover, since $C_{1,J} = C_{1,I-1}$ is known, we have $\widehat{\mathcal{R}}_1^{\text{ODP}} = 0$. Summarizing, we get the following ODP reserves $\widehat{\mathcal{R}}_i^{\text{ODP}}$:

accident year i	1	2	3	4	5	6	7	8	9	10
ODP reserves $\widehat{\mathcal{R}}_i^{\text{ODP}}$	0	15'126	26'257	34'538	85'302	156'494	286'121	449'167	1'043'242	3'950'815

Table 7: ODP reserves $\widehat{\mathcal{R}}_i^{\text{ODP}}$ for all accident years $i = 1, \dots, I$.

We observe that $\widehat{\mathcal{R}}_i^{\text{ODP}} = \widehat{\mathcal{R}}_i^{\text{CL}}$ for all accident years $i = 1, \dots, I$, where $\widehat{\mathcal{R}}_i^{\text{CL}}$ are the CL reserves from Exercise 13.1. As a matter of fact, this observation holds true in general, see Theorem 9.11 of the lecture notes (version of March 20, 2019). By aggregating the ODP reserves over all accident years, we get the ODP predictor $\widehat{\mathcal{R}}^{\text{ODP}}$ for the outstanding loss liabilities of past exposure claims (which is equal to the CL predictor $\widehat{\mathcal{R}}^{\text{CL}}$):

$$\widehat{\mathcal{R}}^{\text{ODP}} = \sum_{i=1}^{I} \widehat{\mathcal{R}}_i^{\text{ODP}} = 6'047'061 = \sum_{i=1}^{I} \widehat{\mathcal{R}}_i^{\text{CL}} = \widehat{\mathcal{R}}^{\text{CL}}.$$

(c) As the ODP model belongs to the family of GLM models, we can calculate the ODP reserves also using the GLM machinery. In particular, we work with the two risk characteristics accident year i, with parameters $\beta_{1,1}, \dots, \beta_{1,I}$, and development year j, with parameters $\beta_{2,0}, \dots, \beta_{2,J}$, where $\beta_{1,i}$ corresponds to accident year i and $\beta_{2,j}$ to development year j. Compared to the parametrization on the exercise sheet, in order to apply GLM techniques, we use the following re-parametrization. We assume that all $X_{i,j}$ are independent with

$$\frac{X_{i,j}}{\phi} \sim \mathrm{Poi}(\lambda_{i,j}/\phi),$$

for all risk classes $(i,j), 1 \le i \le I, 0 \le j \le J$, where $\lambda_{i,j}$ denotes the mean parameter. Note that we work with volumes which are constantly equal to 1. Moreover, in a Poisson GLM model we would set $\phi = 1$. Here we assume a general dispersion parameter $\phi > 0$. We have

$$\mathbb{E}[X_{i,j}] = \phi\, \mathbb{E}\left[\frac{X_{i,j}}{\phi}\right] = \phi \frac{\lambda_{i,j}}{\phi} = \lambda_{i,j},$$

and we model

$$g(\lambda_{i,j}) = g(\mathbb{E}[X_{i,j}]) = \beta_0 + \beta_{1,i} + \beta_{2,j},$$

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 $\beta_{1,1} = \beta_{2,0} = 0$. We refer to Listing 1 for the application of this over-dispersed Poisson GLM model in R.

Listing 1: R code for Exercise 13.3 (c).

```
### Load the required packages
library(readxl)
library(plyr)

### Download the data from the link indicated on the exercise sheet
### Store the data under the name "Exercise13Data.xls" in the same folder as this R code
### Load the data
data <- read_excel("Exercise13Data.xls", sheet="Data_1", range="B22:K31", col_names=FALSE)

### ODP as GLM Model
data2 <- as.data.frame(data)
data2[,2:10] <- data2[,2:10]-data2[,1:9]
data2 <- stack(data2, select=c("X__1","X__2","X__3","X__4","X__5","X__6","X__7","X__8","X__9",
                               "X__10"))
data2[,2] <- rep(1:10)
data2[,3] <- rep(0:9,each=10)
colnames(data2)[2:3] <- c("AY","DY")
data2$AY <- as.factor(data2$AY)
data2$DY <- as.factor(data2$DY)
lower.ind <- is.na(data2[,1])
upper <- data2[is.na(data2[,1])==FALSE,]
lower <- data2[is.na(data2[,1]),]
ODP <- glm(values ~ AY+DY, data=upper, family=quasipoisson())
lower[,1] <- predict(ODP, newdata=lower, "response")
ODP.GLM.reserves <- rep(0,10)
ODP.GLM.reserves[1] <- 0
ODP.GLM.reserves[2:10] <- ddply(lower, .(AY), summarise, reserves=sum(values))[,2]
round(ODP.GLM.reserves)

### MLEs for the accident years
exp(c(0,ODP$coefficients[2:10])+ODP$coefficients[1])*sum(exp(c(0,ODP$coefficients[11:19])))

### MLEs for the development years
round(exp(c(0,ODP$coefficients[11:19]))/sum(exp(c(0,ODP$coefficients[11:19]))),3)
```

Running the R code of Listing 1, we can confirm that the ODP GLM model leads to the same reserves as the CL method. In order to check the MLE parameters of Tables 5 and 6, we have to go back to the parametrization used on the exercise sheet. We write

$\widehat{\beta}_0^{\text{MLE}}, \widehat{\beta}_{1,1}^{\text{MLE}}, \ldots, \widehat{\beta}_{1,I}^{\text{MLE}}, \widehat{\beta}_{2,0}^{\text{MLE}}, \ldots, \widehat{\beta}_{2,J}^{\text{MLE}}$ for the resulting MLEs of the ODP GLM model. By setting

$$\widetilde{\mu}_i = \exp\left\{\widehat{\beta}_0^{\text{MLE}} + \widehat{\beta}_{1,i}^{\text{MLE}}\right\} \sum_{k=0}^{J} \exp\left\{\widehat{\beta}_{2,k}^{\text{MLE}}\right\},$$

for all $i = 1, \ldots, I$, and

$$\widetilde{\gamma}_j = \frac{\exp\left\{\widehat{\beta}_{2,j}^{\text{MLE}}\right\}}{\sum_{k=0}^{J} \exp\left\{\widehat{\beta}_{2,k}^{\text{MLE}}\right\}},$$

for all $j = 0, \ldots, J$, we get

$$\begin{aligned}
\widehat{\lambda}_{i,j}^{\text{MLE}} &= \exp\left\{\widehat{\beta}_0^{\text{MLE}} + \widehat{\beta}_{1,i}^{\text{MLE}} + \widehat{\beta}_{2,j}^{\text{MLE}}\right\} \\
&= \exp\left\{\widehat{\beta}_0^{\text{MLE}} + \widehat{\beta}_{1,i}^{\text{MLE}}\right\} \sum_{k=0}^{J} \exp\left\{\widehat{\beta}_{2,k}^{\text{MLE}}\right\} \frac{\exp\left\{\widehat{\beta}_{2,j}^{\text{MLE}}\right\}}{\sum_{k=0}^{J} \exp\left\{\widehat{\beta}_{2,k}^{\text{MLE}}\right\}} \\
&= \widetilde{\mu}_i \widetilde{\gamma}_i.
\end{aligned}$$

In particular, we get back to the parametrization used on the exercise sheet. The values of $\widetilde{\mu}_i$, $i = 1, \ldots, I$ and $\widetilde{\gamma}_j$, $j = 0, \ldots, J$, are calculated on lines 26 and 29 of Listing 1. We can confirm that we get the same values as in Tables 5 and 6.

Solution 13.4 Mack's Formula and Merz-Wüthrich (MW) Formula

(a) The R code used in this exercise is provided in Listing 2. We get the following results:

accident year i	CL reserves $\widehat{\mathcal{R}}_i^{CL}$	$\sqrt{\text{total msep}}$ (Mack)	in % of the reserves	$\sqrt{\text{CDR msep}}$ (MW)	in % of the $\sqrt{\text{total msep}}$
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 %

Table 8: CL reserves $\widehat{\mathcal{R}}_i^{CL}$, Mack's square-rooted conditional mean square errors of prediction and MW's square-rooted conditional mean square errors of prediction for all accident years $i = 1, \ldots, I$.

(b) Mack's square-rooted conditional mean square errors of prediction give us confidence bounds around the 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.

(c) MW's square-rooted conditional mean square errors of prediction measure the contribution of the next accounting year to the total (run-off) uncertainty given by Mack's square-rooted conditional mean square errors of prediction. For aggregated accident years, 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 discovered in Exercise 13.1, (a).

Listing 2: R code for Exercise 13.4 (a).

```
### Load the required packages
library(readxl)
library(ChainLadder)

### Download the data from the link indicated on the exercise sheet
### Store the data under the name "Exercise13Data.xls" in the same folder as this R code
### Load the data
data <- read_excel("Exercise13Data.xls", sheet="Data_1", range="B22:K31", col_names=FALSE)

### 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 mseps
M <- MackChainLadder(tri, est.sigma="Mack")

### CL reserves and Mack's square-rooted mseps (including illustrations)
M
plot(M)
plot(M, lattice=TRUE)

### CL reserves, MW's square-rooted mseps and Mack's square-rooted mseps
CDR(M)

### Mack's square-rooted mseps in % of the reserves
round(CDR(M)[,3]/CDR(M)[,1],3)*100

### MW's square-rooted mseps in % of Mack's square-rooted mseps
round(CDR(M)[,2]/CDR(M)[,3],2)*100

### Full uncertainty picture
CDR(M, dev="all")
```