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

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

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 $x_1, \dots, x_8$ by independent Gaussian random variables $X_1, \dots, X_8$ 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

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

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

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

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

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

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. $\mathbb{P}[T \le -1] = 1 - \mathbb{P}[T \le 1]$. The probability $\mathbb{P}[T \le 1]$ is approximately 0.84, and the p-value is given by

$$\mathbb{P}[|T| \ge 1] \;=\; 2 - 2\mathbb{P}[T \le 1] \;\approx\; 2 - 2 \cdot 0.84 \;=\; 0.32.$$

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 $\mu = \mathbb{E}[X_1] = 1'000 \cdot 0.1 + 0 \cdot 0.9 = 100$.

(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 $X_1, \dots, X_n$ we define

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

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

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

For $n = 1'000$ we get

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

Thus, if we have $n = 1'000$ 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 $X_1, \ldots, X_n$,

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

Chebychev's inequality leads to

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

We have

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

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 $n \approx 6'000$, approximately, while for $n = 90'000$ 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 $\mathcal{F}$.