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

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

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

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

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

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

Hence, the new 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 $d^{\mathrm{MM}} = d^{\mathrm{MLE}} = 2$. The AIC value $\mathrm{AIC}^{\mathrm{MM}}$ of the method of moments model and the AIC value $\mathrm{AIC}^{\mathrm{MLE}}$ of the MLE model are then given by

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

According to the AIC, the model with the smallest AIC value should be preferred. Since $\mathrm{AIC}^{\mathrm{MM}} > \mathrm{AIC}^{\mathrm{MLE}}$, 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 $d^{\mathrm{gam}} = 2$ and $d^{\mathrm{exp}} = 1$. The AIC value $\mathrm{AIC}^{\mathrm{gam}}$ of the gamma model and the AIC value $\mathrm{AIC}^{\mathrm{exp}}$ of the exponential model are then given by

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

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

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

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

and

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

According to the BIC, the model with the smallest BIC value should be preferred. Since $\mathrm{BIC}^{\mathrm{gam}} > \mathrm{BIC}^{\mathrm{exp}}$, we choose the exponential model.

Note that the gamma model gives the better in-sample fit than the exponential model. But if we adjust this in-sample fit by the number of parameters used, we conclude that the exponential model probably has the better out-of-sample performance (better predictive power).