Non-Life Insurance: Mathematics and Statistics

Solution sheet 5

Solution 5.1 Kolmogorov-Smirnov Test

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

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

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

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

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

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

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

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

Now the Kolmogorov-Smirnov test statistic D_5 is defined as

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

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

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

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

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

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

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

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

Solution 5.2 Large Claims

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

and

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

Hence we get

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Solution 5.3 Pareto Distribution

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

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

for all $x \ge \theta$.

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

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

for all $x \ge \theta$. 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 $\theta \le u_1 < u_2$ and $\alpha \ne 1$. Then the expected value of Y within the layer $(u_1, u_2]$ can be calculated as

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

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

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

and if $\alpha = 1$, we get

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

(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),

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

Thus we get

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

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