Non-Life Insurance: Mathematics and Statistics

Exercise sheet 5

Exercise 5.1 Large Claims

In this exercise we are interested in storm and flood events with total claim amounts exceeding CHF 50 million. Assume that for the yearly claim amount S of such storm and flood events we have $S \sim \text{CompPoi}(\lambda, G)$, where λ is unknown and G is the distribution function of a Pareto distribution with threshold $\theta = 50$ and unknown tail index $\alpha > 0$. Moreover, the number of claims and the claim sizes and, thus, also the total claim amounts are assumed to be independent across different years. During the years $1986 - 2005$, we observed the following 15 storm and flood events:

date	amount in millions	date	amount in millions
20.06.1986	52.8	18.05.1994	78.5
18.08.1986	135.2	18.02.1999	75.3
18.07.1987	55.9	12.05.1999	178.3
23.08.1987	138.6	26.12.1999	182.8
26.02.1990	122.9	04.07.2000	54.4
21.08.1992	55.8	13.10.2000	365.3
24.09.1993	368.2	20.08.2005	1'051.1
08.10.1993	83.8		

(a) Using n independent claim sizes $Y_1, \ldots, Y_n$, show that the MLE $\widehat{\alpha}_n^{\text{MLE}}$ of α is given by

$$\widehat{\alpha}_n^{\text{MLE}} = \left(\frac{1}{n} \sum_{i=1}^{n} \log Y_i - \log \theta \right)^{-1}.$$

(b) According to Lemma 3.8 of the lecture notes,

$$\frac{n-1}{n} \widehat{\alpha}_n^{\text{MLE}}$$

is an unbiased version of the MLE. Estimate α using the unbiased version of the MLE for the storm and flood data given above.

(c) Calculate the MLE of λ for the storm and flood data given above.

(d) Suppose that we introduce a maximal claims cover of $M = 2$ billion CHF per event, i.e. the individual claims are given by $\min\{Y_i, M\}$. Using the estimates of α and λ found above, calculate the estimated expected total yearly claim amount.

(e) Using the estimates of α and λ found above, calculate the estimated probability that we observe at least one storm and flood event next year which exceeds the level of $M = 2$ billion CHF.

Exercise 5.2 Claim Size Distributions (R Exercise)
Write an R Code that generates i.i.d. samples of size $n = 10'000$ from each of the following distributions:

- $\Gamma(\gamma, c)$ with shape parameter $\gamma = \frac{1}{4}$ and scale parameter $c = \frac{1}{40'000}$,
- Weibull(τ, c) with shape parameter $\tau = 0.54$ and scale parameter $c = 0.000175$,
- LN(μ, σ^2) with mean parameter $\mu = \log\left(2000\sqrt{5}\right)$ and variance parameter $\sigma^2 = \log(5)$,
- Pareto(θ, α) with threshold $\theta = 10'000\frac{\sqrt{5}}{2+\sqrt{5}}$ and tail index $\alpha = 1 + \frac{\sqrt{5}}{2}$.

Note that the parameters are chosen such that the theoretical expectations resp. the theoretical standard deviations are approximately equal to 10'000 resp. 20'000, for all the distributions listed above. For each of these i.i.d. samples consider

- the density plot (on the log scale),
- the box plot (on the log scale),
- the plot of the empirical distribution function (on the log scale),
- the plot of the empirical loss size index function,
- the empirical log-log plot,
- the plot of the empirical mean excess function.

Comment your results.

Exercise 5.3 Pareto Distribution
Suppose the random variable Y follows a Pareto distribution with threshold $\theta > 0$ and tail index $\alpha > 0$.

(a) Show that the survival function of Y is regularly varying at infinity with tail index α.

(b) Show that for $\theta \le u_1 < u_2$ the expected value of Y within the layer $(u_1, u_2]$ is given by

$$\mathbb{E}\left[Y 1_{\{u_1 < Y \le u_2\}}\right] = \begin{cases} \theta\frac{\alpha}{\alpha-1}\left[\left(\frac{u_1}{\theta}\right)^{-\alpha+1} - \left(\frac{u_2}{\theta}\right)^{-\alpha+1}\right], & \text{if } \alpha \neq 1, \\ \theta\log\left(\frac{u_2}{u_1}\right), & \text{if } \alpha = 1. \end{cases}$$

(c) Show that for $\alpha > 1$ and $y > \theta$ the loss size index function for level y is given by

$$\mathcal{I}[G(y)] = 1 - \left(\frac{y}{\theta}\right)^{-\alpha+1}.$$

(d) Show that for $\alpha > 1$ and $u > \theta$, the mean excess function of Y above u is given by

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

Exercise 5.4 Kolmogorov-Smirnov Test
Suppose we are given the following claim size data (in increasing order) coming from independent realizations of an unknown claim size distribution:

$$x_1 = \left(-\log\frac{38}{40}\right)^2, x_2 = \left(-\log\frac{37}{40}\right)^2, x_3 = \left(-\log\frac{35}{40}\right)^2, x_4 = \left(-\log\frac{34}{40}\right)^2, x_5 = \left(-\log\frac{10}{40}\right)^2.$$

Perform a Kolmogorov-Smirnov test at significance level of 5% to test the null hypothesis of having a Weibull distribution with shape parameter $\tau = \frac{1}{2}$ and scale parameter $c = 1$ as claim size distribution. Moreover, explain why the Kolmogorov-Smirnov test is applicable in this example.