Non-Life Insurance: Mathematics and Statistics

Exercise sheet 7

Exercise 7.1 Hill Estimator (R Exercise)
Write an R code that samples 300 independent observations from a Pareto distribution with threshold $\theta = 10$ million and tail index $\alpha = 2$. Create a Hill plot and a log-log plot. What do you observe?

Exercise 7.2 Approximations for Compound Distributions
Assume that S has a compound Poisson distribution with expected number of claims $\lambda v = 1\text{'}000$ and claim sizes following a gamma distribution with shape parameter $\gamma = 100$ and scale parameter $c = \frac{1}{10}$. For all $\alpha \in (0, 1)$, let q_α denote the α-quantile of S.

(a) Use the normal approximation to estimate $q_{0.95}$ and $q_{0.99}$.

(b) Use the translated gamma approximation to estimate $q_{0.95}$ and $q_{0.99}$.

(c) Use the translated log-normal approximation to estimate $q_{0.95}$ and $q_{0.99}$.

(d) Comment on the results found above.

Exercise 7.3 Monte Carlo Simulations (R Exercise)
In this exercise we consider the same setup as in Exercise 7.2 and use Monte Carlo simulations to determine the distribution of the total claim amount S.

(a) Write an R code that simulates $n = 100\text{'}000$ times the total claim amount S. Compare the resulting distribution function of S to the approximate distribution functions found in Exercise 7.2, where we used the normal, the translated gamma and the translated log-normal approximation.

(b) Write an R code that simulates $n = 100, 1\text{'}000, 10\text{'}000$ times the total claim amount S and replicates these simulations 100 times. For each $n \in \{100, 1\text{'}000, 10\text{'}000\}$ discuss the distribution of the resulting 100 values of the quantiles $q_{0.95}$ and $q_{0.99}$ of S.

Exercise 7.4 Akaike Information Criterion and Bayesian Information Criterion
Assume that we fit a gamma distribution to a set of $n = 1\text{'}000$ i.i.d. claim sizes and that we obtain the following method of moments estimators and MLEs:

$$\widehat{\gamma}^{\text{MM}} = 0.9794 \qquad \text{and} \qquad \widehat{c}^{\text{MM}} = 9.4249,$$
$$\widehat{\gamma}^{\text{MLE}} = 1.0013 \qquad \text{and} \qquad \widehat{c}^{\text{MLE}} = 9.6360.$$

The corresponding log-likelihoods are given by

$$\ell_{\mathbf{Y}}\left(\widehat{\gamma}^{\text{MM}}, \widehat{c}^{\text{MM}}\right) = 1264.013 \quad \text{and} \quad \ell_{\mathbf{Y}}\left(\widehat{\gamma}^{\text{MLE}}, \widehat{c}^{\text{MLE}}\right) = 1264.171.$$

(a) Why is $\ell_{\mathbf{Y}}\left(\widehat{\gamma}^{\text{MLE}}, \widehat{c}^{\text{MLE}}\right) > \ell_{\mathbf{Y}}\left(\widehat{\gamma}^{\text{MM}}, \widehat{c}^{\text{MM}}\right)$? Which fit should be preferred according to the Akaike Information Criterion (AIC)?

(b) The estimates of γ are very close to 1 and, thus, we could also use an exponential distribution as claim size distribution. For the exponential distribution we obtain the MLE $\widehat{c}^{\text{MLE}} = 9.6231$ and the corresponding log-likelihood $\ell_{\mathbf{Y}}\left(\widehat{c}^{\text{MLE}}\right) = 1264.169$. According to the AIC and the Bayesian Information Criterion (BIC), should we prefer the gamma model or the exponential model?