Non-Life Insurance: Mathematics and Statistics

Exercise sheet 6

Exercise 6.1 Log-Normal Distribution and Deductible
Assume that the total claim amount

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

in a given line of business has a compound distribution with $\mathbb{E}[N] = \lambda v$, where $\lambda > 0$ denotes the claim frequency, $v > 0$ the volume, and with a claim size distribution being a log-normal distribution with mean parameter $\mu \in \mathbb{R}$ and variance parameter $\sigma^2 > 0$.

(a) Show that

$$\begin{aligned}
\mathbb{E}[Y_1] &= \exp\left\{\mu + \frac{\sigma^2}{2}\right\}, \\
\operatorname{Var}(Y_1) &= \exp\left\{2\mu + \sigma^2\right\}\left(\exp\left\{\sigma^2\right\} - 1\right), \quad \text{and} \\
\operatorname{Vco}(Y_1) &= \sqrt{\exp\left\{\sigma^2\right\} - 1}.
\end{aligned}$$

(b) Suppose that $\mathbb{E}[Y_1] = 3\text{'}000$ and $\operatorname{Vco}(Y_1) = 4$. Up to now, the insurance company was not offering contracts with deductibles. Now it wants to offer a deductible of $d = 500$. Answer the following questions:

(i) How does the claim frequency λ change by the introduction of the deductible?

(ii) How does the expected claim size $\mathbb{E}[Y_1]$ change by the introduction of the deductible?

(iii) How does the expected total claim amount $\mathbb{E}[S]$ change by the introduction of the deductible?

Exercise 6.2 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 (MM) estimates and maximum likelihood estimates (MLE):

$$\begin{aligned}
\widehat{\gamma}^{\text{MM}} &= 0.9794 \quad \text{and} \quad \widehat{c}^{\text{MM}} = 9.4249, \\
\widehat{\gamma}^{\text{MLE}} &= 1.0013 \quad \text{and} \quad \widehat{c}^{\text{MLE}} = 9.6360.
\end{aligned}$$

The corresponding log-likelihoods are given by

$$\ell_{\mathbf{Y}}\left(\widehat{\gamma}^{\text{MM}}, \widehat{c}^{\text{MM}}\right) = 1\text{'}264.013 \quad \text{and} \quad \ell_{\mathbf{Y}}\left(\widehat{\gamma}^{\text{MLE}}, \widehat{c}^{\text{MLE}}\right) = 1\text{'}264.171.$$

(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 $\ell_{\mathbf{Y}}\left(\widehat{c}^{\text{MLE}}\right) = 1\text{'}264.169$. According to the AIC and the Bayesian Information Criterion (BIC), should we prefer the gamma model or the exponential model?

Exercise 6.3 Goodness-of-Fit Test
Suppose we are given the following claim size data (in increasing order) coming from independent realizations of an unknown claim size distribution:

210, 215, 228, 232, 303, 327, 344, 360, 365, 379, 402, 413, 437, 481, 521, 593, 611, 677, 910, 1623.

(a) Use the intervals

$$I_1 = [200, 239), \quad I_2 = [239, 301), \quad I_3 = [301, 416), \quad I_4 = [416, 725), \quad I_5 = [725, +\infty)$$

to perform a χ^2-goodness-of-fit test at significance level of 5% to test the null hypothesis of having a Pareto distribution with threshold $\theta = 200$ and tail index $\alpha = 1.25$ as claim size distribution.

(b) In goodness-of-fit tests with K disjoint intervals and a total of n observations, we use the test statistic

$$X^2_{n,K} = \sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k},$$

where O_k denotes the actual number of observations and E_k the expected number of observations in the k-th interval. We assume that the parameters of the null hypothesis distribution function are given and that the K disjoint intervals are chosen such that $E_k > 0$, for all $k = 1, \ldots, K$. Show that in the case of $K = 2$ disjoint intervals, the test statistic $X^2_{n,2}$ converges to a χ^2-distribution with one degree of freedom, as $n \to \infty$.

Exercise 6.4 Kolmogorov-Smirnov Test
Suppose we are given the following data (in increasing order) coming from independent realizations of an unknown 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 that the data given above comes from a Weibull distribution with shape parameter $\tau = \frac{1}{2}$ and scale parameter $c = 1$. Moreover, explain why the Kolmogorov-Smirnov test is applicable in this example.