Exercises

Section 9.2

Exercise 9.2.1 The cost of settling a claim changes from Z to $Z(1+I)$ if I is the rate of inflation between two time points. **a)** Suppose claim size Z is Gamma(α, ξ) in terms of the *old* price system. What are the parameters under the new, inflated price? **b)** The same same question when the old price is Pareto(α, β). **c)** Again the same question when Z is log-normally distributed. **d)** What is the general rule for incorporating inflation into a parametric model of the form (**??**)?

Exercise 9.2.2 This is a follow-up of the preceding exercise. Let $z_1, \dots, z_n$ be historical data collected over a time span influenced by inflation. We must then associate each claim z_i with a price level $Q_i = 1 + I_i$ where I_i is the rate of inflation. Suppose the claims have been ordered so that z_1 is the first (for which $I_1 = 0$) and z_n the most recent. **a)** Modify the data so that a model that can be fitted from them. **b)** Ensure that the model applies to the time of the most recent claim. Imagine that all inflation rates $I_1, \dots, I_n$ can be read off from some relevant index.

Exercise 9.2.3 Consider n_l observations censored to the **left**. This means that each Z_i is some b_i or smaller (by how much isn't known). With $F_0(z/\beta)$ as the distribution function define a contribution to the likelihood similar to **right** censoring in (**??**).

Exercise 9.2.4 Families of distribution with unknown lower limits b can be defined by taking $Y = b + Z$ where Z starts at the orgin. Let $Y_i = b + Z_i$ be an independent sample $(i = 1, \dots .n)$ and define

$$M_y = \min(Y_1, \dots, Y_n) \qquad \text{and} \qquad M_z = \min(Z_1, \dots, Z_n).$$

a) Show that $E(M_y) = b + E(M_z)$. **b)** Also show that

$$\Pr(M_z > z) = \{1 - F(z)\}^n \qquad \text{so that} \qquad E(M_z) = \int_0^\infty \{1 - F(z)\}^n \, dz,$$

where $F(z)$ is the distribution function of Z [Hint: Use Exercise ??? for the expectation.]. **c)** With $F(z) = F_0(z/\beta)$ deduce that

$$E(M_y) = b + \int_0^\infty \{1 - F_0(z/\beta)\}^n \, dz = b + \beta \int_0^\infty \{1 - F_0(z)\}^n \, dz$$

and explain how this justifies the bias correction (**??**) when $\hat{b} = M_y$ is used as estimate for b.

Exercise 9.2.5 We shall in this exericise consider simulated, log-normal historical data, estimate skewness through the ordinary estimate (**??**) and examine how it works when the answer is known (look it up in Exercise 9.3.5 below). **a)** Generate $n = 30$ log-normal claims using $\theta = 0$ and $\tau = 1$ and compute the skewness coefficient (**??**). **b)** Redo four times and remark on the pattern when you compare with the true value. **c)** Redo a),b) when $\tau = 0.1$. What about the patterns now? **d)** Redo a) and b) for $n = 1000$. What has happened?

Exercise 9.2.6 Consider the pure empirical model $\hat{Z}$ defined in (**??**). Show that third order moment and skewness become

$$\nu_3(\hat{Z}) = \frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^3 \qquad \text{so that} \qquad \text{skew}(\hat{Z}) = \frac{n^{-1} \sum_{i=1}^n (z_i - \bar{z})^3}{s^3},$$

where $\bar{z}$ and s are sample mean and standard deviation.

Exercise 9.2.7 Consider as in (**??**) $Z_h = \hat{Z} + hs\varepsilon$ where $\varepsilon \sim N(0,1)$, s the sample standard deviation and $h > 0$ is fixed. **a)** Show that

$$\Pr(Z_h \leq z | \hat{Z} = z_i) = \Phi\left(\frac{z - z_i}{hs}\right) \qquad (\Phi(z) \text{ the normal integral}).$$

b) Use this to deduce that

$$\Pr(Z_h \leq z) = \frac{1}{n}\sum_{i=1}^{n} \Phi\left(\frac{z - z_i}{hs}\right).$$

c) Differentiate to obtain the density function of Z_h and show that it corresponds to the kernel density estimate (**??**) in Section 2.2.

Exercise 9.2.8 Show that a Monte Carlo simulation of Z_h can be generated from two uniform variables U_1^* and U_2^* through

$$i^* \leftarrow [1 + nU_1^*] \qquad \text{followed by} \qquad Z_h^* \leftarrow z_{i^*} + hs\,\Phi^{-1}(U_2^*)$$

where $\Phi^{-1}(u)$ is the percentile function of the standard normal. [Hint: Look up Algorithms 2.3 and 4.1].

Section 9.3

Exercise 9.3.1 The convolution property of the Gamma distribution is often formulated as follows. Start with an independent Gamma sample $Z_1 = \xi Z_{01}, \ldots, Z_n = \xi Z_{0n}$ with $Z_{01}, \ldots, Z_{0n}$ coming from the standard Gamma(α). **a)** Verify that $S = Z_1 + \ldots + Z_n = (n\xi)\bar{Z}_0$ where $\bar{Z}_0 = (Z_{01}, \ldots, Z_{0n})/n$. **b)** Use the result on $\bar{Z}_0$ cited in Section 9.3 to deduce that S is Gamma distributed too. What are its parameters?

Exercise 9.3.2 The data below, taken from Beirlant, Teugels and Vynckier (1996) were originally compiled by The American Insurance Association and show losses (in the US) due to single hurricanes. They apply to the period from 1949 to 1980 and have been corrected for inflation.

6.766	7.123	10.562	14.474	13.351	16.983	18.383	19.030	25.304
29.112	30.146	33.727	40.596	41.409	47.905	49.397	52.600	59.917
63.123	77.809	102.942	103.217	123.680	140.136	192.013	198.446	227.338
329.511	361.200	421.680	513.586	545.778	750.389	863.881	163.8000	

The money unit is million US$, but they are for the year 1980 and would be much larger today. **a)** Fit a log-normal and check the fit through a Q-Q plot. **b)** Repeat a), but now subtract $b = 5000$ from all the observations prior to fitting the log-normal. **c)** Any comments?

Exercise 9.3.3 Alternatively the hurricane loss data of the preceding exercise might be described through Gamma distributions. You may either use likelihood estimates (software needed) or the moment estimates derived in Section 9.3; see (**??**). **a)** Fit gamma distributions both to the orginal data and when you subtract 5000 first. Check the fit by Q-Q plotting. An alternatively is to fit *transformed* data. $y_1, \ldots, y_n$. One possibility could be $y_i = \log(z_i - 5000)$ where $z_1, \ldots, z_n$ are the original losses. **b)** Fit the Gamma model to $y_1, \ldots, y_n$ and verify the fir though Q-Q plotting. **c)** Which of the models you have tested in this and the preceding exercise should be chosen? Other possibiltities?

Exercise 9.3.4 Consider a log-normal claim $Z = \exp(\theta + \tau\varepsilon)$ where $\varepsilon \sim N(0,1)$ and θ and τ are parameters. **a)** Argue that skew(Z) does *not* depend on θ [Hint: Use a general property of skewness.]. To calculate skew(Z) we may therefore take $\theta = 0$, and we also need the formula $E\{\exp(a\varepsilon)\} = \exp(a^2/2)$. **b)** Show that

$$(Z - e^{\tau^2/2})^3 = Z^3 - 3Z^2 e^{\tau^2/2} + 3Ze^{\tau^2} - e^{3\tau^2/2}$$

so that **c)** the third order moment becomes

$$\nu_3(Z) = E(Z - e^{\tau^2/2})^3 = e^{9\tau^2/2} - 3e^{5\tau^2/2} + 2e^{3\tau^2/2}.$$

d) Use this together with $\text{sd}(Z) = e^{\tau^2/2}\sqrt{e^{\tau^2} - 1}$ to deduce that

$$\text{skew}(Z) = \frac{\exp(3\tau^2) - 3\exp(\tau^2) + 2}{(\exp(\tau^2) - 1)^{3/2}}.$$

e) Show that $\text{skew}(Z) \to 0$ as $\tau \to 0$ and calculate $\text{skew}(Z)$ for $\tau = 0.1, 1, 2$. The value for $\tau = 1$ corresponds to the density function plotted in Figure 2.4 right.

Exercise 9.3.5 This exericise is a follow-up of Exercise 9.2.5, but it is now assumed that that the underlying model is known to be log-normal. The natural estimate of τ is then $\hat{\tau} = s$ where s is the sample standard deviation of $y_1 = \log(z_1), \ldots, y_n = \log(z_n)$ where $z_1, \ldots, z_n$ is the orginal log-normal claims. Skewness is then estimated by inserting $\hat{\tau}$ for τ in the skewness formula in Exercise 9.2.5 d). **a)** Repeat a), b) and c) in the previous exercise with this new estimation method. **b)** Try to draw some conclusions about the patterns in the estimation errors. Does it seem to help that we know what the underlying distribution is?

Section 9.4
Exercise 9.4.1 Let Z be exponentially distributed with mean ξ. **a)** Show that the over-threshold variable Z_b has the same distribution as Z. **b)** Comment on how this result is linked to the over-threshold model for ordinary Pareto variables.

Exercise 9.4.2 Suppose you have concluded that the decay parameter α of a claim size distribution is infinite so that the over-threshold model exponential. We can't use the scale estimate (**??**) now. How will you modify it? Answer: the method in Exercise 9.4.6.

Exercise 9.4.3 a) Simulate $m = 10000$ observations from a Pareto distribution with $\alpha = 1.8$ and $\beta = 1$ and pretend you do not known the model they are coming from. **b)** Use the Hill estimate on the 100 largest observations. **c)** Repeat a) and b) four times. Try to see some pattern in the estimates compared to the true α (which you know after all!) **d)** Redo a), b) and c) with $m = 100000$ simulations and compare with the earlier results.

Exercise 9.4.4 The Burr model, introduced in Exercise 2.5.4, had distribution function

$$F(x) = 1 - \{1 + (x/\beta)^{\alpha_1}\}^{-\alpha_2}, \qquad x > 0.$$

where β, α_1 and α_2 are positive parameters. Sampling was by inversion. **a)** Generate $m = 10000$ observations from this model when $\alpha_1 = 1.5$, $\alpha_2 = 1.2$ and $\beta = 1$. **b)** Compute $\hat{\alpha}$ as the Hill estimate from the 100 largest observations. **c)** Comment on the discrepancy from the product $\alpha_1\alpha_2$. Why is this comparision relevant? **d)** Compute $\hat{\beta}$ from the 100 largest simulations using the method in Exercise 9.4.2. **e)** Q-Q plot the 100 largest observations against the Pareto distribuion with patarmeters $\hat{\alpha}$ and $\hat{\beta}$. Any comments?

Exercise 9.4.5 a) Generate $m = 10000$ observations from the lognormal distribution with mean $\xi = 1$ and $\tau = 0.5$. **b)** Compute the Hill estimate based on the 1000 largest observations **c)** Repeat a) and b) four times. Any patterns? **d)** Explain why the value you try to estimate is infinite. There is a strong bias in the estimation that prevents that to be reached. It doesn't help you much to raise the threshold and go to $m = 100000$!

Exercise 9.4.6 a) As in the preceding exercise generate $m = 10000$ observations from the lognormal distribution with mean $\xi = 1$ and $\tau = 0.5$. **b)** As $\hat{\xi}$ use the sample mean of the 1000 largest observations subtracted a suitable threshold (immediately below the smallest of them) and Q-Q plot the 100 largest observations agains the exponential distribution with $\hat{\xi}$ as mean. Comments?

Section 9.5

Exercise 9.5.1 Consider a mixture model of the form

$$Z = (1 - I_b)\hat{Z} + I_b(b + Z_b) \qquad \text{where} \qquad Z_b \sim \text{Pareto}(\alpha, \beta), \quad \Pr(I_b = 1) = 1 - \Pr(I_b = 0) = p$$

and $\hat{Z}$ is the empirical distribution function over $z_{(1)}, \ldots, z_{(n_1)}$. It is assumed that $b \geq z_{(n_1)}$ and that $\hat{Z}$, I_b and Z_b are independent. **a)** Determine the (upper) percentiles of Z. [Hint: The expression depend on whether $\epsilon < p$ or not.] **b)** Derive $E(Z)$ and $\text{var}(Z)$, [Hint: One way is to use the rules of double expectation and double variance, conditioning on I_b.]

Exercise 9.5.2 a) Redo the following exercise when Z_b is exponential with mean ξ instead of a Pareto proper. **b)** Comment on the connection by letting $\alpha \to \infty$ and keeping $\xi = \beta/(\alpha - 1)$ fixed.

Exercise 9.5.3 a) How is Algorithm 9.2 modified when the over-threshold distribution is exponential with mean ξ? **b)** Implement the algorithm.

Exercise 9.5.4 We shall use the algorithm of the preceding exercise to carry out an experiment based on the log-normal $Z = \exp(-\tau^2/2 + \tau\varepsilon)$ where $\varepsilon \sim N(0, 1)$ and $\tau = 1$. **a)** Generate a Monte Carlo sample of $n = 10000$ and use those as historical data after sorting them as $z_{(1)} \leq \ldots \leq z_{(n)}$. In practice you would not that they are log-normal, but assume that they are known to light-tailed enough for the the over-threshold distribution to be exponential. The empirical distribution function is used to the left of the threshold. **b)** Fit a mixture model by taking $p = 0.05$ and $b = z_{(9500)}$ [Hint: You take the mean of the 500 obervations above the threshold as estimate of the parameter ξ of the exponential.]. **c)** Generate a Monte Carlo sample of $m = 10000$ from the fitted mixture distribution and estimate the upper 10% and 1% percentiles from the simulations. **d)** Do they correspond to the true ones? Compare with their *exact* values you obtain from knowing the underlying distribution in this laboratory experiment.

Section 9.6

Exercise 9.6.1 We shall in this exercise test the Hill estimate $\hat{\alpha}$ defined in (**??**) and the corresponding $\hat{\beta}_b$ in (**??**) on the the Danish fire data (it can be downloaded from the file danishfire.txt.). **a)** Determine the estimates when $p = 50\%$, 10% and $p = 5\%$. **b)** Compare with the values obtained by likelihood estimation in Table 9.1.

Exercise 9.6.2 Consider historical claim data starting at b (known). A useful family of transformations is

$$Y = \frac{(Z - a)^{\theta} - 1}{\theta} \qquad \text{for} \qquad \theta \neq 0,$$

where θ is selected by the user. **a)** Show that $Y \to \log(Z - b)$ as $\theta \to 0$ [Hint: L'hôpital's rule]. This shows that the logarithm is a special case $\theta = 0$. The family is known as the **Box-Cox** transformation. We shall use it to try to improve the fit of the models for the Danish fire data in Section 9.6. Download the data from danishfire.txt. **b)** Use $a = -0.00001$ and $\theta = 0.1$ and fit the Gamma model to the Y-data. [Hint: Either likelihood or moment, as in Section 9.3] **c)** Verfy the fit by Q-Q plotting. **d)** Repeat b) and c) when $\theta = -0.1$. **e)** Which of the transformations appears best, $\theta = 0$ (as in Figure 9.6.3) or one of those in this exercise?

Exercise 9.6.3 Suppose a claim Z starts at some known value b. **a)** How will you select a in the Box-Cox transformation of the preceding exercise if you are going to fit a positive family of distributions (gamma, log-normal) to the transformed Y-data? **b)** The same question if you are going to use a model (for example the normal) extending over the entire real axis.