1 Modelling claim size

1.1 Introduction

Models describing variation in claim size lack the theoretical underpinning provided by the Poisson point process in Chapter 8. The traditional approach is to impose a family of probability distributions and estimate their parameters from historical claims $z_1, \dots, z_n$ (corrected for inflation if necessary). Even the family itself is often determined from experience. An alternative with considerable merit is to throw all prior matematical conditions over board and rely solely on the historical data. This is known as a **non-parametric** approach. Much of this chapter is on the use of historical data.

How we proceed is partly dictated by the size of the historical record, and here the variation is enormous. With automobile insurance the number of observations n is often large, providing a good basis for the probability distribution of the claim size Z. By contrast, major incidents in industry (like the collapse of an oil rig!) are rare, making the historical material scarce. Such diversity in what there is to go on is reflected in the presentation below. The extreme right tail of the distribution warrants special attention. Lack of historical data in the region that matters most financially is a challenge. What can be done is discussed in Section 9.5.

1.2 Parametric and non-parametric modelling

Introduction

Claim size modelling can be **parametric** through families of distributions such as the Gamma, log-normal or Pareto with parameters tuned to historical data or **non-parametric** where each claim z_i of the past is assigned a probability $1/n$ of re-appearing in the future. A new claim is then envisaged as a random variable $\hat{Z}$ for which

$$\Pr(\hat{Z} = z_i) = \frac{1}{n}, \quad i = 1, \dots, n. \tag{1.1}$$

This is an entirely proper distribution (the sum over all i is one). It may appear peculiar, but there are several points in its favour (and one in its disfavour); see below. Note the notation $\hat{Z}$ which is the familiar way of emphasizing that estimation has been involved. The model is known as the **empirical distribution function** and will in Section 9.5 be employed as a brick in an edifice which also involves the Pareto distribution. The purpose of this section is to review parametric and non-parametric modelling on a general level.

Scale families of distributions

All sensible parametric models for claim size are of the form

$$Z = \beta Z_0, \tag{1.2}$$

where $\beta > 0$ is a parameter, and Z_0 is a standardized random variable corresponding to $\beta = 1$. This proportionality is inherited by expectations, standard deviations and percentiles; i.e. if ξ_0, σ_0 and $q_{0\epsilon}$ are expectation, standard deviation and ϵ-percentile for Z_0, then the same quantities for Z are

$$\xi = \beta\xi_0, \qquad \sigma = \beta\sigma_0 \qquad \text{and} \qquad q_\epsilon = \beta q_{0\epsilon}. \tag{1.3}$$

To see what β stands for, suppose currency is changed as a part of some international transaction. With c as the exchange rate the claim quoted in foreign currency becomes cZ, and from (1.2) $cZ = (c\beta)Z_0$. The effect of passing from one currency to another is simply that $c\beta$ replaces β, the shape of the density function remaining what it was. Surely anything else makes little sense. It would be contrived to take a view on risk that differed in terms of US$ from that in British £ or euros, and this applies to inflation too (Exercise 9.2.1).

In statistics β is known as a **parameter of scale** and parametric models for claim size should always include them. Consider the log-normal distribution which has been used repeatedly. If it is on the form $Z = \exp(\theta + \tau\varepsilon)$ where ε is $N(0,1)$, it may also be rewritten

$$Z = \xi Z_0 \qquad \text{where} \qquad Z_0 = \exp(-\frac{1}{2}\tau^2 + \tau\varepsilon) \quad \text{and} \quad \xi = \exp(\theta + \frac{\tau^2}{2}).$$

Here $E(Z_0) = 1$, and ξ serves as both expectation and scale parameter. The mean is often the most important of all quantities associated with a distribution, and it is useful to make it visible in the mathematical notation.

Fitting a scale family

Models for scale families satisfy the relationship

$$\Pr(Z \leq z) = \Pr(Z_0 \leq z/\beta) \qquad \text{or} \qquad F(z|\beta) = F_0(z/\beta).$$

where $F_0(z)$ is the distribution function of Z_0. Differentiating with respect to z yields the family of density functions

$$f(z|\beta) = \frac{1}{\beta} f_0(\frac{z}{\beta}), \quad z > 0 \qquad \text{where} \qquad f_0(z) = F_0'(z). \tag{1.4}$$

Additional parameters describing the shape of the distributions are hiding in $f_0(z)$. All scale families have density functions on this form.

The standard way of fitting such models is through likelihood estimation. If $z_1, \ldots, z_n$ are the historical claims, the criterion becomes

$$\mathcal{L}(\beta, f_0) = -n\log(\beta) + \sum_{i=1}^{n} \log\{f_0(z_i/\beta)\}, \tag{1.5}$$

which is to be maximized with respect to β and other parameters. Numerical methods are usually required. A useful extension covers situations with **censoring**. Typical examples are claims only registered as above or below certain limits, known as censoring **to the right** and **left** respectively. Most important is the situation where the actual loss is only given as some *lower* bound b. The chance of a claim Z exceeding b is $1 - F_0(b/\beta)$, and the probability of n_r such events becomes

$$\{1 - F_0(b_1/\beta)\} \cdots \{1 - F_0(b_{n_r}/\beta)\}.$$

Its *logarithm* is added to the log likelihood (1.5) of the fully observed claims $z_1, \ldots, z_n$ making the criterion

$$\mathcal{L}(\beta, f_0) = \underbrace{-n\log(\beta) + \sum_{i=1}^{n} \log\{f_0(z_i/\beta)\}}_{\textit{complete information}} \quad + \quad \underbrace{\sum_{i=1}^{n_r} \log\{1 - F_0(b_i/\beta)\}}_{\textit{censoring to the right}}, \tag{1.6}$$

which is to be maximized. Censoring to the left is similar and discussed in Exercise 9.2.3. Details for the Pareto family will be developed in Section 9.4.

Shifted distributions

The distribution of a claim may start at some some threshold b instead of at the orgin. Obvious examples are deductibles and contracts in re-insurance. Models can be constructed by adding b to variables Z starting at the origin; i.e.

$$Z_{>b} = b + Z = b + \beta Z_0.$$

Now

$$\Pr(Z_{>b} \le z) = \Pr(b + \beta Z_0 \le z) = \Pr\left(Z_0 \le \frac{z-b}{\beta}\right),$$

and differentiation with respect to z yields

$$f_{>b}(z|\beta) = \frac{1}{\beta} f_0\left(\frac{z-b}{\beta}\right), \qquad z > b, \tag{1.7}$$

which is the density function of random variables starting at b.

Sometimes historical claims $z_1, \dots, z_n$ are known to exceed some *unknown* threshold b. Their *minimum* provides an estimate, precisely

$$\hat{b} = \min(z_1, \dots, z_n) - C, \qquad \text{for unbiasedness: } C = \beta \int_0^\infty \{1 - F_0(z)\}^n \, dz; \tag{1.8}$$

see Exercise 9.2.4 for the unbiasedness correction. It is rarely worth the trouble to take that too seriously, and accuracy is typically high even when it isn't done[1]. The estimate is known to be **super-efficient**, which means that its standard deviation for large sample sizes is proportional to $1/n$ rather than the usual $1/\sqrt{n}$; see Lehmann and Casella (1998). Other parameters can be fitted by applying other methods of this section to the sample $z_1 - \hat{b}, \dots, z_n - \hat{b}$.

Skewness as simple description of shape

A major issue is asymmetry and the right tail of the distribution. A useful, *simple* summary is the **coefficient of skewness**

$$\zeta = \text{skew}(Z) = \frac{\nu_3}{\sigma^3} \qquad \text{where} \qquad \nu_3 = E(Z - \xi)^3. \tag{1.9}$$

The numerator is the **third order moment**. Skewness should *not* depend on the currency being used and doesn't since

$$\text{skew}(Z) = \frac{E(Z-\xi)^3}{\sigma^3} = \frac{E(\beta Z_0 - \beta\xi_0)^3}{(\beta\sigma_0)^3} = \frac{E(Z_0 - \xi_0)^3}{\sigma_0^3} = \text{skew}(Z_0)$$

after inserting (1.2) and (1.3). Nor is the coefficient changed when Z is shifted by a fixed amount; i.e. $\text{skew}(Z + b) = \text{skew}(Z)$ through the same type of reasoning. These properties confirm skewness

[1]The adjustment requires C to be *estimated*. It is in any case sensible to subtract some *small* number $C > 0$ from the minimum to make $z_i - \hat{b}$ strictly positive to avoid software crashes.

as a simplified measure of shape.

The standard estimate of the skewness coefficient ζ from observations $z_1, \dots, z_n$ is

$$\hat{\zeta} = \frac{\hat{\nu}_3}{s^3} \qquad \text{where} \qquad \hat{\nu}_3 = \frac{1}{n - 3 + 2/n} \sum_{i=1}^{n} (z_i - \bar{z})^3. \tag{1.10}$$

Here $\hat{\nu}_3$ is the natural estimate of the third order moment[2] and s the sample standard deviation. The estimate is for low n and heavy-tailed distributions severely biased downwards. *Under*-estimation of skewness, and by implication the risk of large losses, is a recurrent theme with claim size modelling in general and is common even when parametric families are used. Several of the exercises are devoted to the issue.

Non-parametric estimation

The random variable $\hat{Z}$ that attaches probabilities $1/n$ to all claims z_i of the past is a possible model for *future* claims. Its definition in (1.1) as a discrete set of probabilities may seem at odds with the underlying distribution being continuous, but experience in statistics (see Efron and Tibshriani, 1993) suggests that this matters little. Expectation, standard deviation, skewness and percentiles are all closely related to the ordinary sample versions. For example, the mean and standard deviation of $\hat{Z}$ are by definition

$$E(\hat{Z}) = \sum_{i=1}^{n} \frac{1}{n} z_i = \bar{z}, \qquad \text{and} \qquad \text{sd}(\hat{Z}) = \left(\sum_{i=1}^{n} \frac{1}{n} (z_i - \bar{z})^2 \right)^{1/2} \doteq s, \tag{1.11}$$

and the upper percentiles are (approximately) the historical claims in descending order; i.e.

$$\hat{q}_\varepsilon = z_{(\varepsilon n)} \qquad \text{where} \qquad z_{(1)} \geq \dots \geq z_{(n)}.$$

The skewness coefficient of $\hat{Z}$ is similar; see Exercise 9.2.6.

The empirical distribution function can only be visualized as a **dot plot** where the observations $z_1, \dots, z_n$ are recorded on a straight line to make their tightness indicate the underlying distribution. If you want a density function, turn to the kernel estimate in Section 2.2, which is related to $\hat{Z}$ in the following way. Let ε be a random variable with mean 0 and standard deviation 1, and define

$$\hat{Z}_h = \hat{Z} + hs\varepsilon, \qquad \text{where} \qquad h \geq 0. \tag{1.12}$$

Then the distribution of $\hat{Z}_h$ coincides with the estimate (**??**); see Exercise 9.2.7. Note that

$$\text{var}(\hat{Z}_h) = s^2 + (hs)^2 \qquad \text{so that} \qquad \text{sd}(\hat{Z}_h) = s\sqrt{1 + h^2},$$

a slight inflation in uncertainty over that found in the historical data. With the usual choices of h that can be ignored. Sampling is still easy (Exercise 9.2.8), but usually there is not much point in using a positive h for other things than visualization.

[2]Division on $n - 3 + 2/n$ makes it unbiased.

The empirical distribution function is in finance often called **historical simulation**. It is ultra-rapidly set up and simulated (use Algorithm 4.1), and there is no worry as to whether a parametric family fits or not. On the other hand, *no simulated claim can be larger than what has been seen in the past*. This is a serious drawback. It may not matter too much when there is extensive experience to build on, and in the big consumer branches of motor and housing we have presumably seen much of the worst. The empirical distribution function *can* also be used with big claims when the responsibility per event is strongly limited, but if it is not, the method can go seriously astray and under-estimate risk substantially. Even then is it possible to combine the method with specific techniques for tail estimation, see Section 9.5.

1.3 The Log-normal and Gamma families

Introduction

The log-normal and Gamma distributions are among the most frequently applied models for claim size. Both are of the form $Z = \xi Z_0$ where ξ is the expectation. The standard log-normal Z_0 can be defined through its stochastic representation

$$Z_0 = \exp(-\frac{1}{2}\tau^2 + \tau\varepsilon) \qquad \text{where} \qquad \varepsilon \sim N(0,1) \tag{1.13}$$

whereas we for the standard Gamma must be use its density function

$$f_0(z) = \frac{\alpha^\alpha}{\Gamma(\alpha)} z^{\alpha-1}\exp(-\alpha z), \qquad z > 0; \tag{1.14}$$

see (**??**). This section is devoted to a brief exposition of the main properties of these models.

The log-normal: A quick summary

Log-normal density functions were plotted in Figure 2.4. Their shape depends heavily on τ and are highly skewed when τ is not too close zero; see also Figure 9.2 below. Mean, standard deviation and skewness are

$$E(Z) = \xi, \qquad \text{sd}(Z) = \xi\{\exp(\tau^2) - 1\}^{1/2}, \qquad \text{skew}(Z) = \frac{\exp(3\tau^2) - 3\exp(\tau^2) + 2}{(\exp(\tau^2) - 1)^{3/2}};$$

see Section 2.3. The expession for the skewness coefficient is derived in Exercise 9.3.4.

Parameter estimation is usually carried out by noting that

$$Y = \log(Z) = \underset{mean}{\log(\xi) - \tfrac{1}{2}\tau^2} \quad + \quad \underset{sd}{\tau \cdot \varepsilon}.$$

The original log-normal sample $z_1, \ldots, z_n$ is then transformed to a Gaussian one $y_1 = \log(z_1), \ldots, y_n = \log(z_n)$ and its sample mean and variance $\bar{y}$ and s_y computed. The estimates of ξ and τ become

$$\log(\hat{\xi}) - \frac{1}{2}\hat{\tau}^2 = \bar{y}, \quad \hat{\tau} = s_y \qquad \text{which yields} \qquad \hat{\xi} = \exp(\frac{1}{2}s_y^2 + \bar{y}), \quad \hat{\tau} = s_y.$$

The log-normal distribution is used everywhere in this book.

Properties of the Gamma model

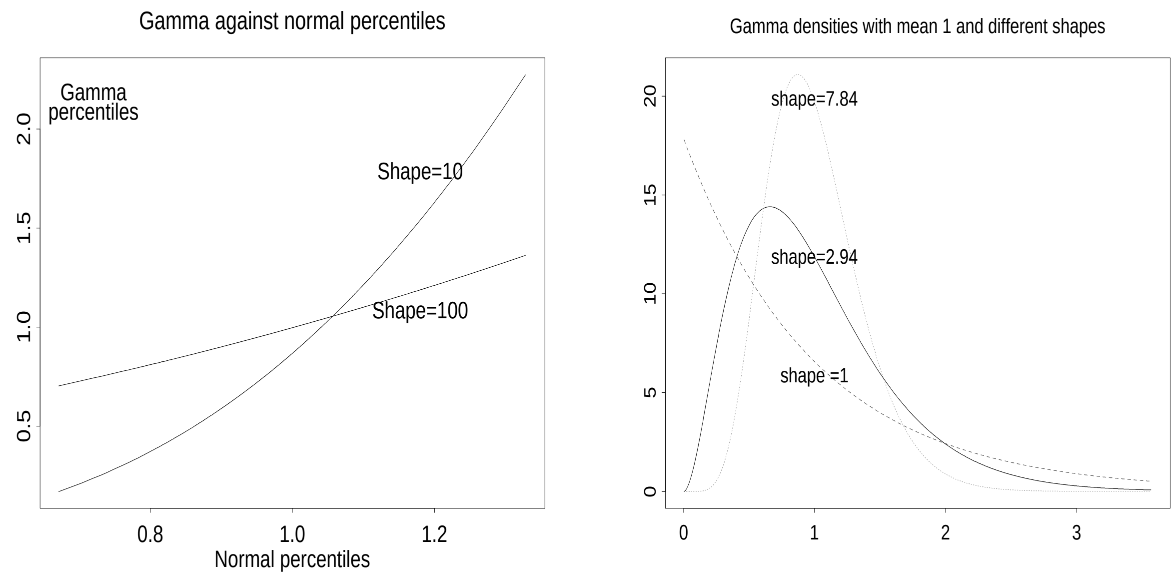


Figure 9.1 Left: *Q-Q plot of standard Gamma percentiles against the normal.* **Right**: *Standard Gxamma density functions.*

Good operational qualities and flexible shape makes the Gamma model useful in many contexts. Mean, standard deviation and skewness are

$$E(Z) = \xi, \qquad \mathrm{sd}(Z) = \xi/\sqrt{\alpha} \qquad \text{and} \qquad \mathrm{skew}(Z) = 2/\sqrt{\alpha}, \tag{1.15}$$

and the model possesses a so-called **convolution** property. Let $Z_{01}, \ldots, Z_{0n}$ be an independent sample from Gamma(α). Then

$$\bar{Z}_0 \sim \text{Gamma } (n\alpha) \qquad \text{where} \qquad \bar{Z}_0 = (Z_{01} + \ldots + Z_{0n})/n.$$

The average of independent, standard Gamma variables is another standard Gamma variable, now with shape $n\alpha$. By the central limit theorem $\bar{Z}_0$ also tends to the normal as $n \to \infty$, and this proves that Gamma variables become normal as $\alpha \to \infty$. This is visible in Figure 9.1 left where Gamma percentiles are Q-Q plotted aganst Gaussian ones. The line is much straightened out when $\alpha = 10$ is replaced by $\alpha = 100$. A similar tendency is seen among the density functions in Figure 9.1 right where two of the distributions were used in Section 8.3 to describe stochastic intensies.

Fitting the Gamma familiy

The method of moments (Section 7.3) is the simplest way to determine Gamma parameters ξ and α from a set of historical data $z_1, \ldots, z_n$. Sample mean and standard deviation $\bar{z}$ and s are then matched the theoretical expressions. This yields

$$\bar{z} = \hat{\xi}, \quad s = \hat{\xi}/\sqrt{\alpha} \qquad \text{with solution} \qquad \hat{\xi} = \bar{z}, \quad \hat{\alpha} = (\bar{z}/s)^2.$$

Likelihood estimation (a little more accurate) is available in commercial software, but is not difficult to implement on your own. The logarithm of the density function of the standard Gamma is

$$\log\{f_0(z)\} = \alpha \log(\alpha) - \log\{\Gamma(\alpha)\} + (\alpha - 1)\log(z) - \alpha z$$

which can be inserted into the log likelihood function (1.5). After some simple manipulations this yields

$$\mathcal{L}(\xi,\alpha) = n\alpha\log(\alpha/\xi) - n\log\Gamma(\alpha) + (\alpha-1)\sum_{j=1}^{n}\log(z_j) - \frac{\alpha}{\xi}\sum_{j=1}^{n} z_j. \tag{1.16}$$

Note that

$$\frac{\partial\mathcal{L}}{\partial\xi} = -\frac{n\alpha}{\xi} + \frac{\alpha}{\xi^2}\sum_{i=1}^{n} z_i = 0 \qquad \text{when} \qquad \xi = (z_1 + \ldots + z_n)/n = \bar{z}.$$

It follows that $\hat{\xi} = \bar{z}$ is the likelihood estimate and $\mathcal{L}(\bar{z},\alpha)$ can be tracked under variation of α for the maximizing value $\hat{\alpha}$. A better way is by the bracketing method in Appendix C.4. Use the approximation in Table 8.2 for the Gamma function.

Regression for claims size

Sometimes you may want to examine whether claim size tend to be systematically higher with certain customers than with others. To the author's experience the issue is often less important than for claim frequency, but we should at least know how it's done. Basis are historical data similar to those in Section 8.4, now of the form

$$\begin{array}{cc}
z_1 & x_{11}\cdots x_{1v} \\
z_2 & x_{21}\cdots x_{2v} \\
\cdot & \cdot\ \cdots\ \cdot \\
\cdot & \cdot\ \cdots\ \cdot \\
z_n & x_{n1}\cdots x_{nv}, \\
\textit{losses} & \textit{covariates}
\end{array}$$

and the question is how we use them to understand how a future, reported loss Z are connected to explanatory variables $x_1, \ldots, x_v$. The standard approach is through

$$Z = \xi Z_0 \qquad \text{where} \qquad \log(\xi) = b_0 + b_1 x_1 + \ldots + b_v x_v,$$

and $E(Z_0) = 1$. As the explanatory variables fluctuate, so does the mean loss ξ.

Frequently applied models for Z_0 are log-normal and Gamma. The former simply boils down to ordinary linear regression and least squares with the logarithm of losses as dependent variable. Gamma regression, member of the family of generalized linear models (Section 8.7) is available in free or commercial software and implemented through an extension of (1.16). An example is given in Section 10.4.

1.4 The Pareto families

Introduction

The Pareto distributions, introduced in Section 2.6, are among the most heavy-tailed of all models in practical use and potentially a conservative choice when evaluating risk in property insurance. Density and distribution functions are

$$f(z) = \frac{\alpha/\beta}{(1+z/\beta)^{1+\alpha}} \quad \text{and} \quad F(z) = 1 - \frac{1}{(1+z/\beta)^{\alpha}}, \qquad z > 0.$$

Simulation ss easy (Algorithm 2.9), and the model was used for illustration in several of the earlier chapters. Pareto distributions also play a special role in the mathematical description of the extreme right tail (see Section 9.5). How they were fitted historical data was explained in Section 7.3 (censoring is added below). A useful generalization to the **extended** Pareto family is covered at the end.

Elementary properties

Pareto models are so-heavy-tailed that even the mean may fail to exist (that's why another parameter β represents scale). Formulae for expectation, standard deviation and skewness are

$$\xi = E(Z) = \frac{\beta}{\alpha - 1}, \quad \mathrm{sd}(Z) = \xi \left(\frac{\alpha}{\alpha-2}\right)^{1/2}, \quad \mathrm{skew}(Z) = 2\left(\frac{\alpha-2}{\alpha}\right)^{1/2}\frac{\alpha+1}{\alpha-3}, \tag{1.17}$$

valid for $\alpha > 1$, $\alpha > 2$ and $\alpha > 3$ respectively. It is to the author's experience rare in practice that the mean is infinite, but infinite variances with values of α between 1 and 2 are not uncommon. We shall later need the median too which equals

$$\mathrm{med}(Z) = \beta(2^{1/\alpha} - 1). \tag{1.18}$$

The exponential distribution appears in the limit when the ratio $\xi = \beta/(\alpha - 1)$ is kept fixed and α raised to infinity; see Section 2.6. There is in this sense overlap between the Pareto and the Gamma families. The exponential distribution is a *heavy*-tailed Gamma and the most *light*-tailed Pareto. It is common to include the exponential distribution in the Pareto family.

Likelihood estimation

The Pareto model was used as illustration in Section 7.3, and likelihood estimation was developed there. **Censored** information are now added. Suppose observations are in two groups, either the ordinary, fully observed claims $z_1, \dots, z_n$ or those (n_r of them) known to have exceeded certain thresholds $b_1, \dots, b_{n_r}$ (by how much isn't known). The log likelihood function for the first group is as in Section 7.3

$$n\log(\alpha/\beta) - (1+\alpha)\sum_{i=1}^{n}\log(1+\frac{z_i}{\beta}),$$

whereas the censored part adds contributions from knowing that $Z_i > b$. The probability that $\Pr(Z_i > b_i)$ is

$$\Pr(Z_i > b_i) = \frac{1}{(1+b_i/\beta)^{\alpha}} \qquad \text{or} \qquad \log\{\Pr(Z_i > b_i)\} = -\alpha\log(1+\frac{b_i}{\beta}),$$

and the full log likelihood becomes

$$\mathcal{L}(\alpha,\beta) = \underbrace{n\log(\alpha/\beta) - (1+\alpha)\sum_{i=1}^{n}\log(1+\frac{z_i}{\beta})}_{\textit{complete information}} \quad - \quad \underbrace{\alpha\sum_{i=1}^{n_r}\log(1+\frac{b_i}{\beta})}_{\textit{Censoring to the right}}.$$

This is to be maximized with respect to α and β, a numerical problem very much the same as the one discussed in Section 7.3

Over-threshold under Pareto
One of the most important properties of the Pareto family is its behaviour at the extreme right tail. The issue is defined by the **over-threshold** model which is the distribution of $Z_b = Z - b$ given $Z > b$. Its density function (derived in Section 6.2) is

$$f_b(z) = \frac{f(b+z)}{1-F(b)}, \qquad z > 0;$$

see (**??**). Over-threshold distributions becomes particularly simple for Pareto models. Inserting the expressions for $f(z)$ and $F(z)$ yields

$$f_b(z) \;=\; \frac{(1+b/\beta)^\alpha \alpha/\beta}{(1+(z+b)/\beta)^{1+\alpha}} \quad = \quad \underbrace{\frac{\alpha/(\beta+b)}{\{1+z/(\beta+b)\}^{1+\alpha}}}_{\textit{Pareto density function}},$$

which is another Pareto density. The shape α is the same as before, but the parameter of scale has now changed to $\beta_b = \beta + b$. Over-threshold distributions preserve the Pareto model and the shape. The mean (if it exists) is known as the **mean excess function**, and becomes

$$E(Z_b|Z>b) = \frac{\beta_b}{\alpha-1} = \frac{\beta+b}{\alpha-1} = \xi + \frac{b}{\alpha-1} \qquad \text{(requires } \alpha > 1\text{)}. \tag{1.19}$$

It is larger than the original ξ and increases linearly with b.

These results hold for infinite α as well. Insert $\beta = \xi(\alpha - 1)$ into the expression for $f_b(z)$, and it follows as in Section 2.6 that

$$f_b(z) \to \frac{1}{\xi}\exp(-z/\xi) \qquad \text{as} \qquad \alpha \to \infty.$$

The over-threshold model of an exponential distribution is the same exponential. Now the mean excess function is a constant which follows from (1.19) when $\alpha \to \infty$.

The extended Pareto family
A valuable addition to the Pareto family is to include a polynomial term in the numerator so that the density function reads

$$f(z) = \frac{\Gamma(\alpha+\theta)}{\Gamma(\alpha)\Gamma(\theta)}\frac{1}{\beta}\frac{(z/\beta)^{\theta-1}}{(1+z/\beta)^{\alpha+\theta}} \qquad \text{where} \qquad \beta,\alpha,\theta > 0. \tag{1.20}$$

Shape is now determined by two parameters (α and θ) which creates useful flexibility; see below. The density function is either decreasing over the entire real line (if $\theta \leq 1$) or has a single maximum (if $\theta > 1$). Mean and standard deviation are

$$\xi = E(Z) = \frac{\theta\beta}{\alpha-1} \qquad \text{and} \qquad \text{sd}(Z) = \xi\left(\frac{\alpha+\theta-1}{\theta(\alpha-2)}\right)^{1/2}, \tag{1.21}$$

which are valid when $\alpha > 1$ and $\alpha > 2$ respectively whereas skewness is

$$\text{skew}(Z) = 2\left(\frac{\alpha-2}{\theta(\alpha+\theta-1)}\right)^{1/2}\frac{\alpha+2\theta-1}{\alpha-3}, \tag{1.22}$$

provided $\alpha > 3$. These results, verified in Section 9.7, reduce to those for the ordinary Pareto distribution when $\theta = 1$; the conditions for their validity are the same too.

The model has an interesting limit when $\alpha \to \infty$. Suppose θ and the mean ξ are kept fixed. By (1.21) left $\beta = \xi(\alpha - 1)/\theta$, and when this is inserted into (1.20), it emerges (proof in Section 9.7) that

$$f(z) \to \frac{\theta^\theta}{\Gamma(\theta)\xi}(z/\xi)^{\theta-1}e^{-\theta z/\xi} \qquad \text{as} \qquad \alpha \to \infty,$$

which is the density function of a Gamma model with mean ξ and shape θ! This generalizes a similar result for the ordinary Pareto distribution and is testemony to the versatility of the extended family which comprises heavy-tailed Pareto ($\theta = 1$ and α small) *and* light-tailed (almost Gaussian) Gamma (both α and θ large). In practice you let historical experience decide by fitting the model to past claims $z_1, \dots, z_n$. The likelihood function is

$$\begin{aligned}\mathcal{L}(\alpha,\theta,\beta) = n[\log\{\Gamma(\alpha+\theta)\} - \log\{\Gamma(\alpha)\} - \log\{\Gamma(\theta)\} - \theta\log(\beta)] \\ +(\theta-1)\sum_{i=1}^n \log(z_i) - (\alpha+\theta)\sum_{i=1}^n \log(1 + z_i/\beta)\end{aligned}$$

which follows from (1.20). To determine the likelihood estimates this criterion must be maximed numerically.

The distribution functions are complicated (unless $\theta = 1$), and it is not covenient to simulate by inversion. A satisfactory alternative is to utilize that an extended Pareto variable with parameters (α, θ, β) can be represented as

$$Z = \frac{\theta\beta}{\alpha}\frac{G_1}{G_2} \qquad \text{where} \qquad G_1 \sim \text{Gamma}(\theta), \quad G_2 \sim \text{Gamma}(\alpha). \tag{1.23}$$

Here G_1 and G_2 are two independent Gamma variables with mean one. The result which is proved in Section 9.7, leads to the following algorithm:

Algorithm 9.1 The extended Pareto sampler
0 Input: α,θ,β and $\eta = \theta\beta/\alpha$
1 Draw $G_1^* \sim$ Gamma(θ) *%Standard Gamma, Algorithm 2.13 or 2.14*
2 Draw $G_2^* \sim$ Gamma(α) *%Standard Gamma, Algorithm 2.13 or 2.14*
3 Return $Z^* \leftarrow \eta\, G_1^*/G_2^*$

1.5 Extreme value methods

Introduction

Large claims play a special role because of their importance financially. It is also hard to assess their distribution. They (luckily!) do not occur very often, and historical experience is therefore limited. Insurance companies may even cover claims *larger* than anything that has been seen before. How should such situations be tackled? The simplest would be to fit a parametric family and try to extrapolate beyond past experience. That may not be a very good idea. A Gamma distribution may fit well in the central regions without being reliable at all at the extreme right tail, and such a procedure may easily underestimate big claims severely; more on this in Section 9.6. The purpose of this section is to enlist help from a theoretical charaterization of the extreme right tail of *all* distributions.

Over-threshold distributions in general

It was established earlier that over-threshold distributions of Pareto models remain Pareto. There is, perhaps surprisingly, a *general* extension. All random variation exceeding a *very large* threshold b is approximately of the Pareto type, no matter (almost) what the distribution is! The prerequisite is that the random variable Z has no upper limit and is continuously distributed. There is even a theory when Z is bounded by some given maximum; for that extension consult Embrects, Klüppelberg and Mikosch (1997).

To express the result in mathematical terms let $P(z|\alpha,\beta)$ be the distribution function of the Pareto model with parameters α and β and define

$$F_b(z) = \Pr(Z_b \leq z|Z > b) = \Pr(Z \leq b + z|Z > b).$$

as the over threshold distribution function of an *arbitrary* random variable Z. Let Z be unlimited with continuous distribution. Then *there exists a positive parameter* α *(possibly infinite) such that there is for all thresholds* b *a parameter* β_b *that makes*

$$\max_{z\geq 0} |F_b(z) - P(z|\alpha,\beta_b)| \to 0, \qquad \text{as} \qquad b \to \infty.$$

This tells us that discrepancies between the two distribution functions vanish as the threshold grows. At the end they are equal, and the over-threshold distribution has become a member of the Pareto family. The result is exact and applies for *finite* b (with $\beta_b = \beta + b$) when the original model is Pareto itself.

Whether we get a Pareto proper (with finite α) or an exponential (infinite α) depends on the right tail of the distribution function $F(z)$. The determining factor is how fast $1 - F(z) \to 0$ as $z \to \infty$. A decay of order $1/z^\alpha$ leads to Pareto models with shape α. A simple example of such **polynomial** decay is the Burr distribution of Exercise 2.5.4 for which the distribution function is

$$F(z) = 1 - \{1 + (z/\beta)^{\alpha_1}\}^{-\alpha_2} \qquad \text{or for } z \text{ large} \qquad 1 - F(z) \doteq \{(z/\beta)^{\alpha_1}\}^{-\alpha_2} = (z/\beta)^{-\alpha_1\alpha_2},$$

and $\alpha = \alpha_1\alpha_2$. Many distributions have lighter tails. The Gamma and the log-normal are two examples where the limiting over-threshold model is the exponential; see Exercises 9.4.3-6 for illustrations.

The Hill estimate

The decay rate α can be determined from historical data (though they have to be plenty). One possibility is to select observations exceeding some threshold, impose the Pareto distribution and use likelihood estimation as explained in Section 9.4. This line is tested in the next section. An alternative is the **Hill** estimate

$$\hat{\alpha}^{-1} = \frac{1}{n - n_1} \sum_{i=n_1+1}^{n} \log\left(\frac{z_{(i)}}{z_{(n_1)}}\right) \qquad (1.24)$$

where $z_{(1)} \leq \ldots \leq z_{(n)}$ are the data sorted in ascending order and n_1 is user selected. Note that the estimate is non-parametric (no model assumed). It is thoroughly discussed in Embrects, Klüppelberg and Mikosch (1997) where it is shown to converge to the true value when $n \to \infty$ and $n_1/n \to 1$. In other words, n_1/n should be close to one and $n - n_1$ large, and this requires n huge. A simple justification of the Hill estimate is given in Section 9.7.

We may want to use $\hat{\alpha}$ as an estimate of α in a Pareto distribution imposed over the threshold $b = z_{(n_1)}$, and would then need an estimate of the scale parameter β_b. The likelihood method is a possibility, but requires numerical computation, and a simpler way is

$$\hat{\beta}_b = \frac{z_{(n_2)} - z_{(n_1)}}{2^{1/\hat{\alpha}} - 1} \qquad \text{where} \qquad n_2 = 1 + \frac{n_1 + n}{2}. \tag{1.25}$$

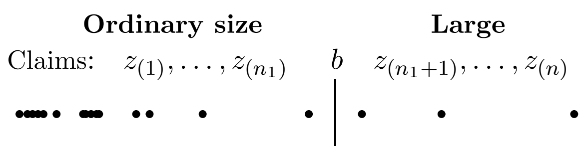
To justify it note that the the over-theshold observations

$$z_{(n_1+1)} - z_{(n_1)}, \ldots, z_{(n)} - z_{(n_1)} \qquad \text{have median} \qquad z_{(n_2)} - z_{(n_1)},$$

and the median (1.18) under Pareto distributions suggests that $\hat{\beta}_b$ can be determined from the equation $\hat{\beta}_b(2^{1/\hat{\alpha}} - 1) = z_{(n_2)} - z_{(n_1)}$ which yields (1.25).

The entire distribution through mixtures

How can the tail characterisation result be utilized to model the entire distribution? Historical claims look schematically like the following:



There are many values in the small and medium range to the left of the vertical bar and just a few (or none!) large ones to the right of it. What is actually meant by 'large' is not clear-cut, but let us say that 'large' claims are those exceeding some threshold b. If the original claims $z_1, \ldots, z_n$ are ranked in ascending order as

$$z_{(1)} \leq z_{(2)} \cdots \leq z_{(n)},$$

then observations from $z_{(n_1)}$ and smaller are below the threshold. How b is chosen in practice is discussed below; see also Section 9.6 for numerical illustrations.

One strategy is to divide modelling into separate parts defined by the threshold. A random variable (or claim) Z may always be written

$$Z = (1 - I_b)Z_{\leq b} + I_b Z_{>b} \tag{1.26}$$

where

$$\underset{\textit{central region}}{Z_{\leq b} = Z|Z \leq b,} \qquad \underset{\textit{extreme right tail}}{Z_{>b} = Z|Z > b} \qquad \text{and} \qquad \begin{aligned} I_b &= 0 \quad \text{if } Z \leq b \\ &= 1 \quad \text{if } Z > b. \end{aligned} \tag{1.27}$$

The random variable $Z_{\leq b}$ is Z confined to the region to the left of b, and $Z_{>b}$ is similar to the right. It is easy to check that the two sides of (1.26) are equal, but at first sight this merely looks complicated. Why on earth can it help? The point is that we reach out to two different sources of information. There is to the left of the threshold historical data with which a model may be identified. On the right the result due to Pickands suggests a Pareto distribution. This defines a modelling strategy which will now be developed.

The empirical distribution mixed with Pareto

The preceding two-component approach can be implemented in more ways than one. For moderate claims to the left of b a parametric family of distributions can be used. There would be data to fit it, and when it was sampled, simulations exceeding the theshold would be discarded. An alternative is non-parametric modelling, and this is the method that will be detailed. *Choose* some small probability p and let $n_1 = n(1-p)$ and $b = z_{(n_1)}$. Then take

$$Z_{\leq b} = \hat{Z} \qquad \text{and} \qquad Z_{>b} = z_{(n_1)} + \text{Pareto}(\alpha, \beta), \tag{1.28}$$

where $\hat{Z}$ is the empirical distribution function over $z_{(1)}, \ldots, z_{(n_1)}$; i.e.

$$\Pr(\hat{Z} = z_{(i)}) = \frac{1}{n_1}, \quad i = 1, \ldots, n_1. \tag{1.29}$$

The remaining part (the delicate one!) are the parameters α and β of the Pareto distribution and the choice of p. Plenty of historical data would deal with everything. Under such circumstances p can be determined low enough (and hence b high enough) for the Pareto approximation to be a good one, and historical data to the right of b would provide accurate estimates $\hat{\alpha}$ and $\hat{\beta}$.

This rosy picture is not the common one, and limited experience often makes it hard to avoid a subjective element. One of the advantages of dividing modelling into two components is that it clarifies the domain where personal judgement enters. You make take the view that a degree of conservatism is in order when there is insufficient information for accuracy. If so, that can be achieved by selecting b relatively low and use Pareto modelling to the right of it. Numerical experiments that supports such a strategy are carried out in the next section. Much material on modelling extremes can be found in Embrects, Klüppelberg and Mikosch (1997).

Sampling mixture models

As usual a sampling algorithm is also a summary of how the model is constructed. With the empirical distribution used for the central region it runs as follows:

Algorithm 9.2 Claims by mixtures

0 Input: Sorted claims $z_{(1)} \leq \ldots \leq z_{(n)}$,
and p, $n_1 = n(1-p)$, α and β.
1 Draw uniforms U_1^*, U_2^*
2 If $U_1^* > p$ then
3 $i^* \leftarrow 1 + [n_1 U_2^*]$ and $Z^* \leftarrow z_{(i^*)}$ *%The empirical distribution, Algorithm 4.1*
else
4 $Z^* \leftarrow b + \beta\{(U_2^*)^{-1/\alpha} - 1\}$ *%Pareto, Algorithm 2.8*
5 Return Z^*

The algorithm operates by testing whether the claim comes from the central part of the distribution or from the extreme, right tail over b. Other distributions could have been used on Line 3. The preceding version is extremely quick to implement.

1.6 Searching for the model

Introduction

How is the final model for claim size selected? There is no single recepy, and it is the kind of issue that can only be learned by example. How we go about is partly dictated by the amount of historical data. Useful tools are transformations and Q-Q plots. The first example below is the so-called Danish fire claims, used by many authors as a test case; see Embrechts, Klüppelberg and Mikosch (1997). These are losses from more than two thousand industrial fires and serve our need for a big example that offers many opportunities for modelling. Several distributions will be fitted and used later (Section 10.3) to calculate reserves.

But what about cases such as the Norwegian fund for natural disasters in chapter 7 where there were just $n = 21$ historical incidents? It is from records this size quite impossible to determine the underlying distribution, and yet we have to come up with a solution. The errors involved and what strategies to employ are discussed in the second half of this section.

Using transformations

A useful tool is to change data by means of **transformations**. These are monotone, continuous functions which will be denoted $H(z)$. The situation is then as follows:

$$\underset{\text{original data}}{z_1, \ldots, z_n} \qquad\qquad \underset{\text{new data}}{y_1 = H(z_1), \ldots, y_n = H(z_n),}$$

and modelling is attacked through $y_1, \ldots, y_n$. The idea is to make standard families of distributions fit the transformed variable $Y = H(Z)$ better than the orginal Z. At the end re-transform through $Z = H^{-1}(Y)$ with $Z^* = H^{-1}(Y^*)$ for the Monte Carlo.

A familiar example is the log-normal. Now $H(z) = \log(z)$ and $H^{-1}(y) = \exp(y)$ with Y normal. General constructions with logarithms are

$$\underset{Y \text{ positive}}{Y = \log(1 + Z)} \qquad\qquad \underset{Y \text{ over the entire real line}}{Y = \log(Z),}$$

opening for entirely different families of distributions for Y. The logarithm is arguably the most commonly applied transformation. Alternatives are powers $Y = Z^\theta$ where $\theta \neq 0$ is some given index; see also Exercise 9.6.2. The final choice of transformations is often made by trial and error.

Example: The Danish fire claims

Many authors have used the Danish fire claims as testing ground for their methods. There are $n = 2167$ industrial fires with damages starting at one million Danish kroner (around eight Danish kroner in one euro). The largest among them is 263, the average $\bar{z} = 3.39$ and the standard deviation $s = 8.51$. A huge skewness coefficient $\hat{\zeta} = 18.7$ signals that the right tail is heavy with

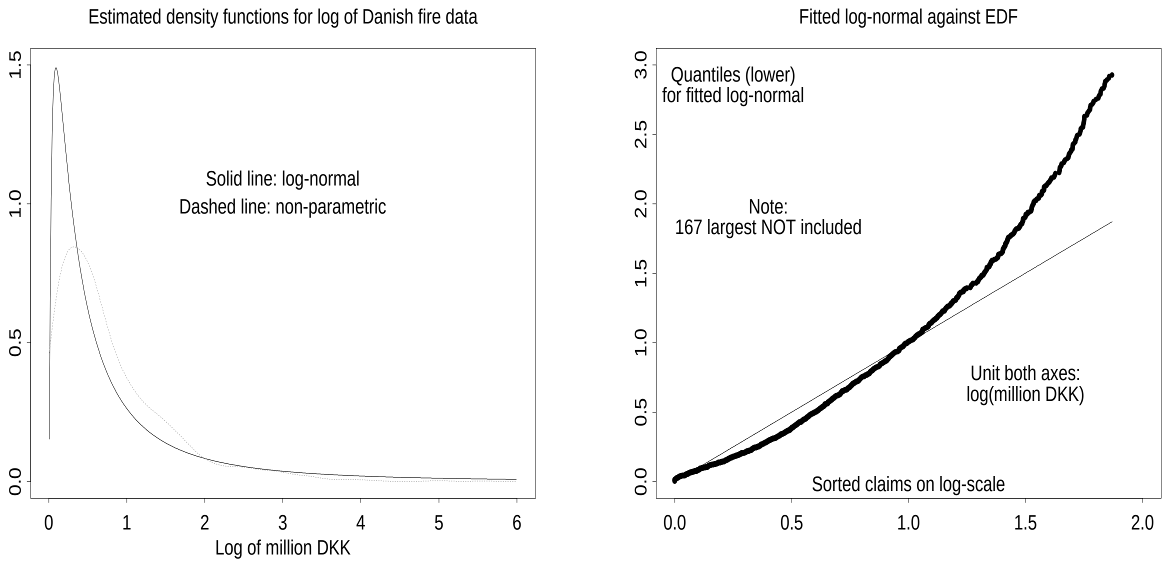


Figure 9.2 *The* **log-normal** *model fitted the Danish fire data on* **log**-*scale. Density function with kernel density estimate (left) and Q-Q plot (right).*

considerable scope for large losses. This is confirmed in Figure 9.2 left where the non-parametric density estimate is shown as the dashed line.

Standard models such as Gamma, log-normal or Pareto do not fit these data (Pareto is attempted below), but matters may be improved by using transformations. The logarithm is often a first choice. The claims start at 1 so that $Y = \log(Z)$ is positive. Could the log-normal be a possibility? With $\varepsilon \sim N(0,1)$ the model reads

$$Z = e^Y, \quad Y = \xi e^{-\tau^2/2+\tau\varepsilon} \qquad \text{with likelihood estimates} \qquad \hat{\xi} = 1.19,\ \hat{\tau} = 1.36,$$

but this doesn't work. Neither the the estimated density function on the left in Figure 9.2 nor the Q-Q plot on the right matches the historical data. The right tail is too heavy and exaggerates the risk of large claims[3].

An second attempt with more success is the extended Pareto family (still applied on log-scale). The best-fitting among those turned out to be a Gamma distribution, i.e.

$$Z = e^Y, \quad Y = \xi\,\text{Gamma}(\alpha) \qquad \text{with likelihood estimates} \qquad \hat{\xi} = 0.79,\ \hat{\alpha} = 1.16,$$

where Gamma(α) has mean one. Note the huge discrepancy in the estimated mean $\hat{\xi}$ from the log-normal mean, indicating that one of the models (or both) is poor. Actually the Gamma fit is none too bad as the Q-Q plot on the right in Figure 9.3 bears evidence. Perhaps the extreme right tail is slightly too light, but fit isn't an end in itself, and consequences when the reserve is evaluated is not necessarily serious; see Section 10.3. A slight modification in Exercise 9.6.2 improves the fit.

[3] The 167 largest observations have been left out of the Q-Q plot to make the resolution in other parts of the plot better

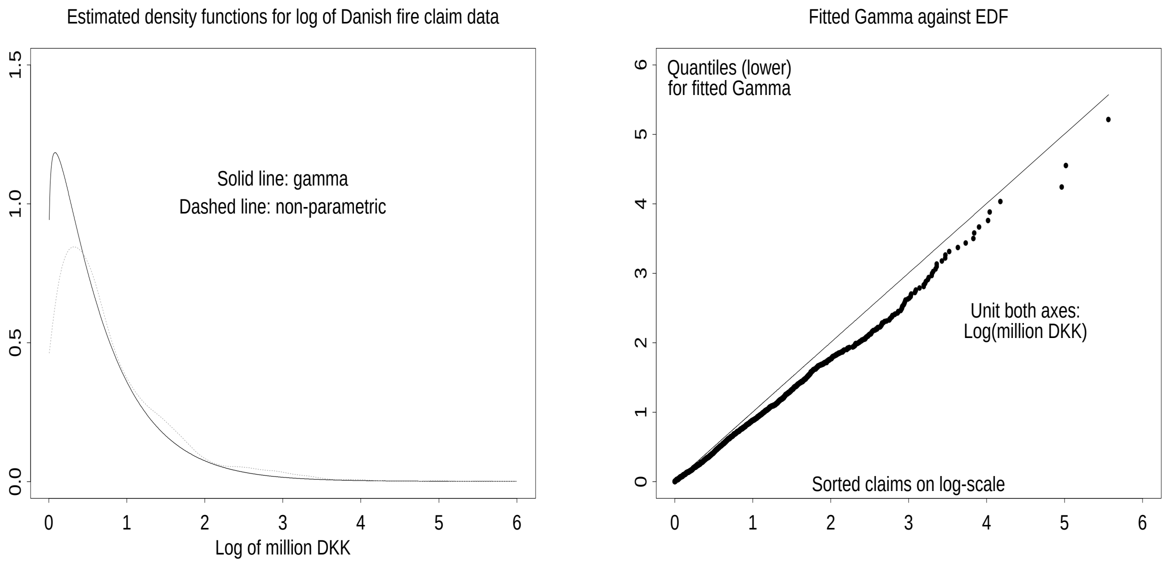


Figure 9.3 *The* **Gamma** *model fitted the Danish fire data on* **log***-scale. Density function with kernel density estimate (left) and Q-Q plot (right).*

Pareto mixing

With so much historical data it is tempting to forget all about parametric families and use the strategy advocated in Section 9.5 instead. The central part is then described by the empirical distribution function and the extreme right tail by Pareto. Table 9.2 shows the results of fitting Pareto distributions over different thresholds (maximum likelihood used). If the parent distribution is Pareto, the shape parameter α is the same for all thresholds b whereas the scale parameter depends on b through $\beta_b = \xi + \beta/(\alpha - 1)$, and there are reminiscences of this in Table 9.1; details in Exercise 9.4.1.

But it would be a gross exaggeration to proclaim the Pareto model for these data. Consider the Q-Q plots in Figure 9.4 where the upper half of the observations have been plotted on the left and the 5% largest on the right. There is a reasonable fit on the right, and this accords with the theoretical result for large thresholds, but it is different on the left where the Pareto distribution overstates the risk of large claims. Table 9.1 tells us why. The shape parameters 1.42 for the 50% largest observations and 2.05 over the 5% threshold correspond to quite different distributions, the

Unit: Million Danish kroner

	Part of data fitted			
	All	50% largest	10% largest	5% largest
Threshold (b)	1.00	1.77	5.56	10.01
Shape (α)	1.64	1.42	1.71	2.05
Scale (β)	1.52	1.82	7.75	14.62

Table 9.1 *Pareto parameters for the over threshold distribution of the fire claims.*

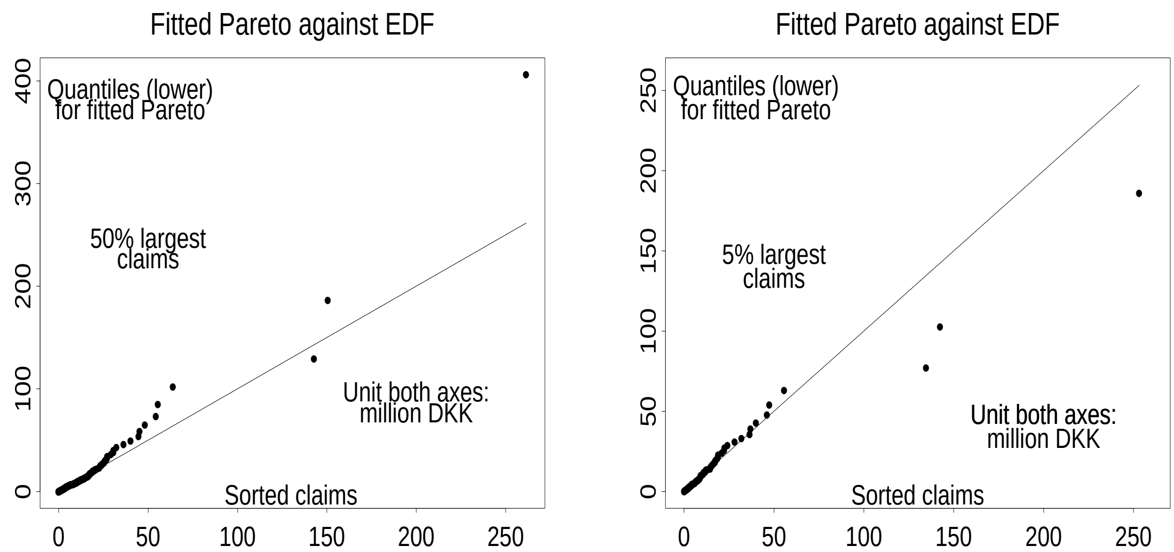


Figure 9.4 *Q-Q plots of fitted* **Pareto** *distributions against the empirical distribution function,* 50% *largest observations (left) and* 5% *largest (right).*

tails of the former being heavier.

When data are scarce

How should we confront a situation like the one in Table 7.1 (the Norwegian natural disasters) where there were no more than $n = 21$ claims and where the phenomenon itself surely is heavy-tailed with potential losses much larger than those on record? The underlying distribution can't be determined with any accuracy, yet somehow a model must be found. Geophysical modelling where natural disasters and their cost are simulated in the computer is a possibility, but such methods are outside our natural range of topics, and we shall concentrate on what can be extracted from historical losses.

The parameters are bound to be highly inaccurate, but how important is the family of distributions? When Gamma and Pareto distribution are fitted the natural disasters (by maximum likelihood), the results look like this:

		percentiles	
Shape (α)	Mean	5%	1%
0.72	179	603	978

Gamma family

		percentiles	
Shape (α)	Mean	5%	1%
1.71	200	658	1928

Pareto family

These are *very* unequal models, yet their discrepancies, though considerable, are not enormous in the central region (differing around 10% up to the upper 5% percentile). *Very* large claims are different, and the Pareto 1% percentile is twice that of Gamma. There is a lesson here. Many families fit reasonably well up to some moderate threshold. *That makes modelling easier when there are strong limits on responsibilities.* If it isn't, the choice between parametric families becomes more delicate.

Shapes in true models: 1.71 in Pareto, 0.72 in Gamma. 1000 *repetitions.*

True model	*Historical record:* $n = 21$			*Historical record:* $n = 80$		
	Models found			*Models found*		
	Pareto	Gamma	log-normal	Pareto	Gamma	log-normal
Pareto	**.49**	.29	.22	**.72**	.12	.16
Gamma	.44	**.51**	.05	.34	**.66**	0

Table 9.2 *Probabilites of selecting given models (***Bold face***: Correct selection).*

Can the right family be detected?
Nothing prevents us from using Q-Q plots to identify parametric families of distributions even with small amounts of data. Is that futile? Here is an experiment throwing light on the issue. The entire process of selecting models according to the Q-Q fit must first be formalized. Percentiles $\hat{q}_i$ under some fitted distribution function $\hat{F}(z)$ are then compared to the observations, sorted in ascending order as $z_{(1)} \leq \ldots \leq z_{(n)}$. What is actually done isn't clearcut (different ways for different people), but suppose we try to minimize

$$Q = \sum_{i=1}^{n} |\hat{q}_i - z_{(i)}| \qquad \text{where} \qquad \hat{q}_i = \hat{F}^{-1}\left(\frac{i-1/2}{n}\right), \quad i = 1, \ldots, n, \tag{1.30}$$

a criterion that has been proposed as basis for formal goodness of fit tests; see Devroye and László (1985). Competing models are then judged accoring to their Q-score and the one with the smallest value selected.

The Monte Carlo experiments in Table 9.2 reports on results by following this strategy. Simulated historical data were generated under Pareto and Gamma distributions, and parametric families of distributions (possibly different from the true one) fitted. The main steps of the scheme are:

$$\begin{array}{ccccccc} \textbf{True model} & & & & \text{Parametric family tried} & & \\ \text{Pareto} & & & & \textit{fitting}\ \hat{q}_1^* \leq \ldots \leq \hat{q}_n^* & & \\ \text{or} & \longrightarrow & z_1^* \leq \ldots \leq z_n^* & \longrightarrow & & \longrightarrow & Q^* = \sum_i |z_{(i)}^* - \hat{q}_i^*|. \\ \text{Gamma} & & \textit{historical data} & & \textit{sorting}\ z_{(1)}^* \leq \ldots \leq z_{(n)}^* & & \end{array}$$

Three parametric families (Pareto, Gamma, log-normal) were applied to the same historical data, and the one with the smallest Q^*-value picked. How often the right model was found could then be investigated. Table 9.2 shows the selection statistics. It is hard to choose between the three models when there are only $n = 21$ claims. Prospects are improved with $n = 80$ and with $n = 400$ (not shown) the success probability was in the range $90 - 95\%$.

Consequences of being wrong
The preceding experiment met (not surprisingly) with mixed success, and when historical data are that scarce, the fitted model is not likely to be very accurate. The impact on how risk is projected has much to do with the maximum responsibility b per claim. The smaller it is, the better the prospects since many distributions fit in the central region. If b is less than the largest observation $z_{(n)}$, a case can be made for for the empirical distribution function. Another factor is the sign of the error. If inaccuracy is inevitable, perhaps our risk strategy should invite over-estimation? Recall that error in estimated parameters typically has the opposite effect. Now there is a tendency of

$m = 1000$ *replications*

	True model: **Pareto**, shape = 1.71						True model: **Gamma**, shape = 0.72					
	Record: n=21			*Record: n=80*			*Record: n=21*			*Record: n=80*		
Percentiles (%)	25	75	90	25	75	90	25	75	90	25	75	90
Fitted Pareto	**0.4**	**1.5**	**2.9**	**0.7**	**1.3**	**1.7**	0.8	1.4	2.2	0.9	1.3	1.6
Fitted Gamma	0.3	0.6	1.0	0.4	0.7	0.9	**0.8**	**1.1**	**1.3**	**0.9**	**1.1**	**1.2**
Best-fitting	0.4	1.2	2.3	0.6	1.2	1.6	0.8	1.2	1.5	0.9	1.1	1.3

Table 9.3 *The distribution (as* 25 70 *and* 90 *percentiles) of* $\hat{q}_{0.01}/q_{0.01}$ *where* $\hat{q}_{0.01}$ *is fitted and* $q_{0.01}$ *true* 1% *percentiles of claims.* **Bold face***: Correct parametric family used.*

under-estimation of risk even if the the right family of distributions has been picked; see Section 7.3.

It is tempting to promote Pareto models in situations with large claims and limited historical experience, though the inherent conservatism in that choice is threatened by estimation error. This is illustrated in Table 9.3 where the true (upper) percentile q_ϵ of the loss distribution is compared to an estimated $\hat{q}_\epsilon$. The criterion used is the ratio

$$\hat{\psi}_\epsilon = \frac{\hat{q}_\varepsilon}{q_\varepsilon}, \qquad \text{where} \qquad \epsilon = 0.01.$$

Historical data were generated under a Pareto distribution (on the left of Table 9.3) and under a Gamma distribution (on the right). Both models were fitted both sets of data, reflecting that in practice you do not know which one to use. The error in $\hat{q}_\epsilon$ is conveyed by the 25%, 75% and 90% percentiles of $\hat{\psi}_\epsilon$. What are the consequences of choosing the wrong family of distributions? Using Gamma when true model is Pareto is almost sure to underestimate risk (the 90% percentile of $\hat{\psi}_\epsilon$ being *less* than one). The opposite (Pareto when the truth is Gamma) is prone to over-estimation, though *not* so much when $n = 21$.

Pareto modelling as a conservative choice seems confirmed by this, but we could also let the data decide. Several families of distributions are then tried, and the best-fitting one picked as the model. This strategy has been followed on the last row of Table 9.3 where the computer chose between Pareto and Gamma distributions according to the criterion (1.30). Errors in the estimated percentiles are still huge, but the method does come out slightly superior.

1.7 Mathematical arguments

Section 9.4

Moments of the extended Pareto Note that

$$E(Z^i) = \int_0^\infty z^i f(z)\,dz = \beta^i \frac{\Gamma(\alpha+\theta)}{\Gamma(\alpha)\Gamma(\theta)} \int_0^\infty \frac{1}{\beta} \frac{(z/\beta)^{\theta+i-1}}{(1+z/\beta)^{\alpha+\theta}}\,dz$$

when inserting for $f(z)$ from (1.20). The integrand is (except for the constant) an extended Pareto density function with shape parameters $\alpha - i$ and $\theta + i$. It follows that that the integral equals $\Gamma(\alpha-i)\Gamma(\theta+i)/\Gamma(\alpha+\theta)$ and

$$E(Z^i) = \beta^i \frac{\Gamma(\alpha-i)\Gamma(\theta+i)}{\Gamma(\alpha)\Gamma(\theta)}$$

which can be simplified by utilizing that $\Gamma(s) = (s-1)\Gamma(s-1)$. This yields

$$E(Z) = \beta\frac{\theta}{\alpha-1} = \xi \qquad \text{and} \qquad E(Z^2) = \beta^2\frac{(\theta+1)\theta}{(\alpha-1)(\alpha-2)}$$

and the left hand shows that the expecation is as claimed in (1.21). To derive the standard deviation we must utilize that $\text{var}(Z) = E(Z^2) - (EZ)^2$ and simplify. In a similar vein

$$E(Z^3) = \beta^3\frac{(\theta+2)(\theta+1)\theta}{(\alpha-1)(\alpha-2)(\alpha-3)}$$

which is combined with

$$E(Z-\xi)^3 = E(Z^3 - 3Z^2\xi + 3Z\xi^2 - \xi^3) = E(Z^3) - 3E(Z^2)\xi + 2\xi^3.$$

When the expressions for $E(Z^3)$, $E(Z^2)$ and ξ are inserted, it follows after some tedious calculations (which are omitted) that

$$E(Z-\xi)^3 = \beta^3\theta\,\frac{2\alpha^2 + (6\theta-4)\alpha + 4\theta^2 - 6\theta + 2}{(\alpha-1)^3(\alpha-2)(\alpha-3)} = 2\beta^3\theta\,\frac{(\alpha+\theta-1)(\alpha+2\theta-1)}{(\alpha-1)^3(\alpha-2)(\alpha-3)}$$

and the formula (1.22) for $\text{skew}(Z)$ follows by diving this expression on $\text{sd}(Z)^3$.

Gamma distribution in the limit To show that extended Pareto density functions become Gamma as $\alpha \to \infty$, insert $\beta = \xi(\alpha-1)/\theta$ into (1.20) which after some reorganisation may be written

$$f(z) = \left(\frac{\theta}{\xi}\right)^\theta \frac{z^{\theta-1}}{\Gamma(\theta)} \times \frac{\Gamma(\alpha+\theta)}{\Gamma(\alpha)(\alpha-1)^\theta} \times \left(1 + \frac{\theta z/\xi}{\alpha-1}\right)^{-(\alpha+\theta)}$$

The first factor on the right is a constant, the second tend to one as $\alpha \to \infty$ (use the expression in the heading of Table 8.2 to see this) and the third becomes $e^{-\theta z/\xi}$ (after the limit $(1+x/a)^{-a} \to e^{-x}$ as $a \to \infty$). It follows that

$$f(z) \to \frac{\theta^\theta}{\Gamma(\theta)\xi}\,(z/\xi)^{\theta-1}e^{-\theta z/\xi} \qquad \text{as} \qquad \alpha \to \infty,$$

as claimed in Section 9.4.

The ratio of Gamma variables Let $Z = X/Y$ where X and Y are two independent and positive random variables with density functions $g_1(x)$ and $g_2(y)$ respectively. Then

$$\Pr(Z \le z) = \Pr(X \le zY) = \int_0^\infty \Pr(X \le zy)g_2(y)\,dy$$

and when this is differentiated with respect to z, the density function of Z becomes

$$f(z) = \int_0^\infty yg_1(zy)g_2(y)\,dy$$

Let $X = \theta G_1$ and $Y = \alpha G_2$ where G_1 and G_2 are Gamma variables with mean one and shape θ and α respectively. Then $g_1(x) = x^{\theta-1}e^{-x}/\Gamma(\theta)$ and $g_2(y) = y^{\alpha-1}e^{-y}/\Gamma(\alpha)$ so that

$$f(z) = \frac{z^{\theta-1}}{\Gamma(\theta)\Gamma(\alpha)}\int_0^\infty y^{\alpha+\theta+-1}e^{-y(1+z)}\,dy = \frac{z^{\theta-1}}{\Gamma(\theta)\Gamma(\alpha)}\frac{1}{(1+z)^{\alpha+\theta}}\int_0^\infty x^{\alpha+\theta-1}e^{-x}\,dx$$

after substituting $x = y(1+z)$ in the last integral. This is the same as

$$f(z) = \frac{\Gamma(\alpha+\theta)}{\Gamma(\theta)\Gamma(\alpha)}\frac{z^{\theta-1}}{(1+z)^{\alpha+\theta}}$$

which is the extended Pareto density when $\beta = 1$.

Section 9.5
Justification of the Hill estimate Suppose first that $z_1, \ldots, z_n$ come from a pure Pareto distribution with *known* scale parameter β. The likelihood estimate of α was derived in Section 7.3 as

$$\hat{\alpha}_\beta^{-1} = \frac{1}{n}\sum_{i=1}^n \log(1 + \frac{z_i}{\beta}).$$

We may apply this result to observations exceeding some *large* threshold b, say to $z_{(n_1+1)} - b, \ldots, z_{(n)} - b$. For large enough b this sample is approximate Pareto with scale parameter $b + \beta$. It follows that the likelihood estimate becomes

$$\hat{\alpha}_\beta^{-1} = \frac{1}{n-n_1}\sum_{i=n_1+1}^n \log\left(1 + \frac{z_{(i)} - b}{b+\beta}\right) = \frac{1}{n-n_1}\sum_{i=n_1+1}^n \log\left(\frac{z_{(i)} + \beta}{b+\beta}\right).$$

But we are assuming that b (and by consequence all $z_{(i)}$) is much larger than β. Hence

$$\log\left(\frac{z_{(i)}+\beta}{b+\beta}\right) \doteq \log\left(\frac{z_{(i)}}{b}\right) = \log\left(\frac{z_{(i)}}{z_{(n_1)}}\right) \qquad \text{if} \qquad b = z_{(n_1)},$$

which leads to $\hat{\alpha}$ in (1.24).

1.8 Bibliographic notes

Specialist monogrpahs on claim size distributions are Klugman, Panjer and Willmot (1998) and Kleiber and Kotz (2003), but all textbooks on general include contain this topic. The literature contains many distributions that have been neglected here, and but there are not many problems that are not solved well with what *has* been presented. Curiously, the empirical distribution function often fails to be mentioned at all. Daykin, Pentikäinen and Pesonen (1994) (calling it the **tabular method**) and Mikosch (2004) are exceptions.

Large claims and extremes are big topics in actuarial science and elsewhere. If you want to study the mathematical and statistical theory in depth, a good place to start is Embrects, Klüppelberg and Mikosch (1997). Other reviews at about the same mathematical level are Beirlant, Goegebeur, Segers, and Teugels (2004) and Resnik (2006). You might also try De Haan and Ferreira (2004) or even the short article Beirlant (2004). Beirlant, Teugels and Vynckier (1996) treats extremes in general insurance only. Extreme value distributions are reviewed in Kotz and Nadarahaj (2000), and Falk, Hüssler and Reiss (2004) are also principally preoccupied with the mathematical side. Finkelstädt and Rootzén (2004) and Castillo, Hadi, Balakrishnan and Sarabia (2005) contain many applications outside insurance and finance.

The idea of using a Pareto model to the right of a certain limit (Section 9.5) goes back at least to Pickands (1975). Modelling over thresholds is discussed in Davison and Smith (1990) from a statistical point of view. Automatic methods for threshold selection are developed in Dupuis (1999), Frigessi, Haug and Rue (2002) and Beirlant and Goegebur (2004). This is what you need if asked to design computerized systems where the computer handles a large number of portfolios on its own, but it is unlikely that accuracy is improved much over trial and error thresholding. The real problem is usually lack of data, and how you then proceed is rarely mentioned. Section 9.6 (second half) was an attempt to attack model selection when there is little to go on.

Beirlant, J. (2004). Extremes. In *Encyclopedia of Actuarial Science*, Teugels, J, and Sundt, B. (eds), John Wiley & Sons, Chichester, 654-661.

Beirlant,J., Teugels, J.L and Vynckier, P. (1996). *Practical Analysis of Extreme Values.* Leuven University Press, Leuven.

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications.* John Wiley & Sons, Chichester.

Beirlant, J. and Goegebur, Y. (2004). Local Polynomial Maximum Likelihood Estimation for Pareto Type Distributions. *Journal of Multivariate Analysis*, 89, 97-118.

Castillo, E., Hadi, A.S., Balakrishnan, N. and Sarabia, J.M. (2005). *Extreme Value and Related Models with Applications in Engineering and Science.* John Wiley & Sons, Hoboken, New Jersey.

Daykin, C.D., Pentikäinen, T. and Pesonen, M. (1994). *Practical Risk theory for Actuaries.* Chapman & Hall/CRC, London.

De Haan, L. and Ferreira, F. (2004). *Extreme Value Theory. an Introduction.* Springer Verlag, New York.

Davison, A.C. and Smith, R.L. (1990). Models for Exceedances over High Thesholds. *Journal of the Royal Statistical Society, Series B*, 5, 393-442.

Devroye, L. and László, G. (1985). *Nonparametric Density Estimation: the L1 View.* John Wiley & Sons, New York

Dupuis, D.J. (1999). Exceedances over High Thresholds: a Guide to Threshold Selection. *Extremes*, 1, 251-261.

Efron. B and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap.* Chapman & Hall, New York.

Embrects, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extreme Events for Insurance and Finance.* Springer Verlag, Heidelberg.

Falk, M., Hüssler, J. and Reiss, R-D (2004), second ed. *Laws of Small Numbers: Extremes and rare Events.* Birkhauser verlag, Basel.

Finkelstädt, B. and Rootzén, H., eds. (2004). *Extreme Values in Finance, Telecommunications and the Environment.* Chapmann & Hall/CRC, Boca raton, Florida.

Frigessi, A., Haug, O. and Rue, H. (2002). A Dynamic Mixture Model for Unsupervised Tail Estimation without Threshold Selection. *Extremes*, 5, 219-235.

Kleiber, C. and Kotz, S. (2003). *Statistical Size Distributions in Economic and Actuarial Sciences.*

John Wiley & Sons, Hoboken, New Jersey.
Klugman, S.A., Panjer, H. and Willmot, G.E. (1998). *Loss Models: from Data to Decisions.* John Wiley & Sons, New York.
Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications.* Imperial College Press, London.
Lehmann, E. and Casella, G. (1998), second ed. *Theory of Point estimation.* Springer Verlag, New York.
Pickands, J. III (1975). Statistical Inference Using Extreme Order Statistics. *Annals of Statistics*, 31, 119-131.
Mikosch, T. (2004). *Non-Life Insurance Mathematics with Stochastic Processes.* Springer Verlag Berlin Heidelberg.
Resnick, S.I. (2006). *Heavy-Tail Phenomena. Probabilistic and Statistical Modeling.* Springer Verlag, New York.

1.9 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 (1.4)?

Exercise 9.2.2 This is a follow-up of the preceding exercise. Let $z_1, \ldots, z_n$ 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 $I_1, \ldots, I_n$ 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 (1.6).

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 $(i = 1, \ldots .n)$ and define

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

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 (1.8) 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 (1.10) 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 (1.10). **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 (1.1). 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 (1.12) $Z_h = \hat{Z} + hs\varepsilon$ 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 in terms of an independent Gamma sample of the form $Z_1 = \xi Z_{01}, \ldots, Z_n = \xi Z_{0n}$ where $Z_{01}, \ldots, Z_{0n}$ are distributed as 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 due to single hurricanes in the US over the period from 1949 to 1980 (in money unit million US$).

6.766	7.123	10.562	14.474	15.351	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	1638.000	

Correction for inflation has been undertaken up to the year 1980 which means that losses would have been 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 (1.15). **a)** Fit gamma distributions both to the orginal data and when you subtract 5000 first. Check the fit by Q-Q plotting. Another way is to fit *transformed* data, say $y_1, \ldots, y_n$. One possibility is to take $y_i = \log(z_i - 5000)$ where $z_1, \ldots, z_n$ are the original losses. **b)** Fit the Gamma model to $y_1, \ldots, y_n$ and verify the fit 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 skew$(Z) \to 0$ as $\tau \to 0$ and calculate 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 exercise 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 $y_1 = \log(z_1), \ldots, y_n = \log(z_n)$. As usual $z_1, \ldots, z_n$ is the orginal log-normal claims. Skewness is then estimated by inserting $\hat{\tau}$ for τ in the skewness formula in Exercise 9.3.4 d). **a)** Repeat a), b) and c) in Exercise 9.2.5 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 similar one when Z is Pareto with finite α.

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 (1.25) 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 $\hat{\beta}_b$ from the 100 largest simulations using (1.25). **e)** Q-Q plot the 100 largest observations against the Pareto distribution with parameters $\hat{\alpha}$ 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 $\tau = 0.5$. The over-threshold distribution is now for large b exponential. **b)** Estimate its mean ξ through the sample mean of the 1000 largest observations subtracted $b = z_{9000}$ and Q-Q plot the 1000 largest observations against this fitted exponential distribution. 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 (1.15) and the corresponding $\hat{\beta}_b$ in (1.16) on the the Danish fire data (downloadable from the file danishfire.txt.). **a)** Determine the estimates when $p = 50\%$, 10% and $p = 5\%$. **b)** Compare with the values in Table 9.1 whivh were obtained by likelihood estimation.

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 the special case $\theta = 0$. The family is known as the **Box-Cox** transformations. We shall use it to try to improve the modelling of 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.