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 $z_1, \ldots, z_n$ (corrected for inflation if necessary). Even the family itself is often determined from experience. An alternative with considerable merit in the computer age 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 go about is partly dictated by the size of the record, and here the variablity from one case to another is enormous. With automobile insurance the number of observations n may be huge, providing a good basis for deducing the probability distribution of the claim size Z. By contrast, major incidents in industry (like the collapse of an oil rig!) are rare, making historical material scarce. This span of variation is reflected in the presentation below. Basic issues are parametric versus non-parametric methods and (above all) the extreme right tail of distributions. Lack of historical data in the region that matters most financially suggests that this problem deserves special attention. The mathematical framework is **mixing** (Section 9.5) in combination with the tail characterization theorem from the theory of extremes (Section 9.4).

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

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

As model for Z this is an entirely proper distribution (since its sum over all i is one). If it appears peculiar, there are actually several points in its favour (one in its disfavour too); 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 employed as a brick in an edifice that 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, for example, be contrived to take a view on risk that differed in terms of US$ from that in British £ or euros, and the same point applies to inflation (Exercise 9.2.1).

In statistics β is known as a **parameter of scale** and parametric models for claim size should always include them. An example worth commenting is the log-normal distribution used in earlier chapters. If it is on the form $Z = \exp(\theta + \tau\varepsilon)$ where ε is $N(0,1)$, we may also write it

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

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 as the scale parameter. Such tactics has in this book been followed whenever practical.

Fitting a scale family

Models for scale families satisfy the relationship

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

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 probably the situation where the actual loss is only given as some *lower* bound b. The probability of this happening is $1 - F_0(b/\beta)$ leading to

$$\{1 - F_0(b_1/\beta)\} \cdots \{1 - F_0(b_n/\beta)\}$$

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

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

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

Shifted distributions

Sometimes the distribution of a claim starts 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

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

and differentiating with respect to z yields

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

as density function for variables starting at b.

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

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

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 the methods below to the sample $z_1 - \hat{b}, \ldots, z_n - \hat{b}$.

Skewness as simple description of shape

A major issue in claim size modelling is the degree of asymmetry towards the right tail of the distribution. A useful, *simple* summary is the **coefficient of skewness** defined as

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

[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. Software may crash otherwise.

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). Neither 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 as a (simplified) representation of the shape of a distribution.

The standard estimate of the skewness coefficient ζ from observations $z_1, \ldots, z_n$ 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 typically 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, 1994) suggests that this matters little. As with other distributions there are an expectation, a standard deviation, a skewness coefficient and also percentiles. All those are closely related to the ordinary sample versions. For example, the mean and standard deviation of $\hat{Z}$ are by definition

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

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

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

The skewness coefficient is also similar; see Exercise 9.2.8.

The empirical distribution function can only be visualized as **dot plot** where the observations $z_1, \ldots, z_n$ 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

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

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

The distribution of $\hat{Z}_h$ coincides with the estimate (**??**); see Exercise 9.2.9. Note that

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

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 9.2.10), but usually there is not much point in using a positive h for other things than visualization.

In finance the empirical distribution function is often called **historical simulation**. It is ultra-rapid to set up and to simulate (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 observed in the past*. How serious that drawback is depends on the situation. It may not matter too much when there is extensive experience to build on. 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 as in Section 9.5.

1.3 The Log-normal and Gamma families of distributions

Introduction

Two of the most frequently applied descriptions of claim size uncertainty are the log-normal and Gamma models. Both are of the form $Z = \xi Z_0$ where ξ is 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 depended heavily on τ and had a highly skewed form when τ was not too close zero; see also Figure 9.2 below. Mean, standard deviation and skewness are

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

see Section 2.3. The expession for the skewness coefficient is derived in Exercise 9.3.5.

Parameter estimation is usually carried out by noting that

$$Y = \log(Z) = \underbrace{\log(\xi) - \tfrac{1}{2}\tau^2}_{mean} + \underbrace{\tau\ \varepsilon}_{sd},$$

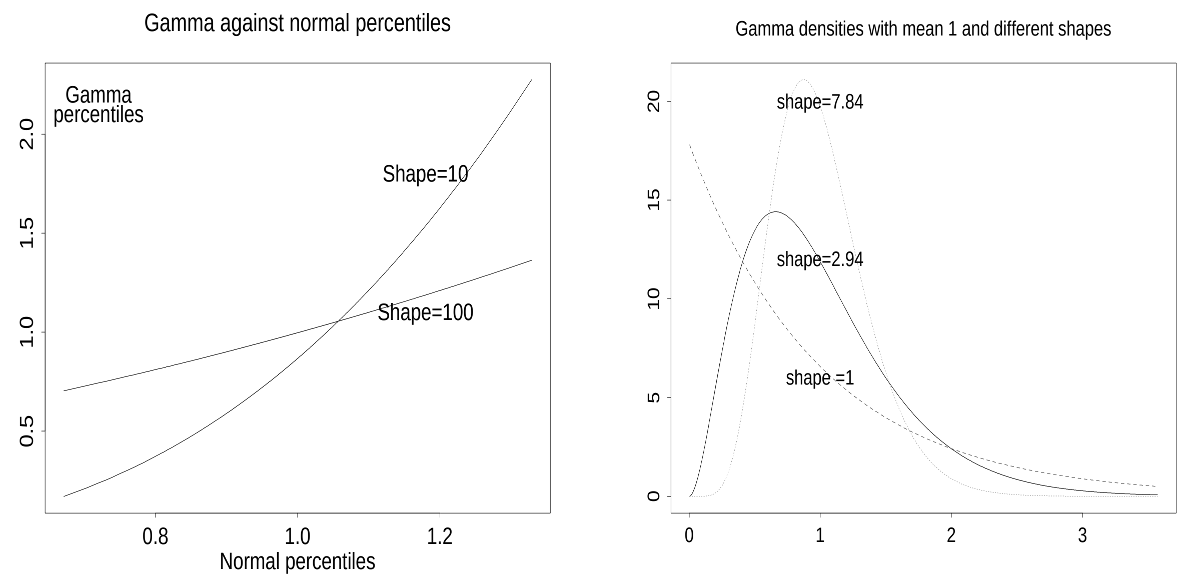


Figure 9.1 **Left**: *Q-Q plot of standard Gamma percentiles against the normal.* **Right**: *Standard Gxamma density functions.*

an immediate consequence of the definition above. A 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 the sample mean and variance $\bar{y}$ and s_y of the latter produce estimates of ξ and τ through

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

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

Properties of the Gamma model

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

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

see Appendix A. In other words, the average is another standard Gamma variable, the shape now being $n\alpha$. 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 normal ones. The line is much straightened out as $\alpha = 10$ is replaced by $\alpha = 100$. A similar tendency is seen among the density functions in Figure 9.1 right where two of the shapes were used with stochastic intensies in Section 8.5. More general versions of the convolution property are given among the exercises.

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 $z_1, \ldots, z_n$. If the theoretical expressions are matched sample mean and standard deviation $\bar{z}$ and s, we obtain

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

Likelihood estimation is slightly more accurate, and is available in commercial software, but it 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 (1.5). After some simple manipulations this yields the log likelihood function

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

Note that

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

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 $\hat{\alpha}$; see also the bisection method in Appendix B.

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 not so important as it was with claim frequency, but we should 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. The logarithm of the claim in then used as as dependent variable and the explanatory variables fitted through ordinary least squares. Gamma regression is available in commercial software and implemented as likelihood fitting through an extension of (1.16). For an example, see Section 10.3.

1.4 The Pareto family and extremes

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 was easy (Algorithm 2.8), and the model was used for illustration in several of the earlier chapters. But Pareto distributions also play a special role in the mathematical description of the extreme right tail. There are, perhaps surprisingly, *general* results in that direction. That is the main topic of this section. How Pareto models were fitted historical data was explained in Section 7.3; censoring is added below.

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

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

valid for $\alpha > 1$, $\alpha > 2$ and $\alpha > 3$ respectively. It is to the author's experience rare in practice that the mean doesn't exist, but infinite variances with values of α between 1 and 2 are not unfrequent. 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. This result is of some importance for the extreme value theory cited below. In this sense the Pareto and the Gamma families intersect. The exponential distribution is a *heavy*-tailed Gamma and the most *light*-tailed 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

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

see (**??**). It becomes particularly simple with Pareto models. Inserting the expressions for $f(z)$ and $F(z)$ yields

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

after some simple manipulations. This is again a Pareto density. The shape α is the same as before, whereas the parameter of scale has become $\beta_b = \beta + b$. In other words, over-threshold models for Pareto variables remain Pareto with shape unaltered. The mean (if it exists) is known as the **mean excess function**, and becomes

$$E(Z_b|Z > b) = \frac{\beta_b}{\alpha-1} = \frac{\beta+b}{\alpha-1} = \xi + \frac{b}{\alpha-1} \qquad \text{(requires } \alpha > 1\text{)}. \tag{1.18}$$

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

Over-threshold modelling in general

The tail property of Pareto models has a *general* extension. When b becomes infinite, only this family can appear no matter (almost) what the distribution of Z was in the beginning! The main condition is that the distribution of Z has no upper limit (it must also be continuous). There is even a theory when Z is bounded by some given maximum, but such models are rarely natural to employ. For that extension see Embrects, Klüppelberg and Mikosch (1997) which also detail certain weak regularity conditions that must be satisfied. The result (which will not be proved) goes back at least to Pickands (1975).

For the precise formulation let $P(z|\alpha, \beta)$ be the distribution function of Pareto(α, 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 satisfying the conditions above. The somewhat complicated statement is that *there exists a positive parameter α (possibly infinite) such we can for all thresholds b find parameters β_b that makes*

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

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. We saw above that 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 $1/z^\alpha$ yields Pareto (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

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

and $\alpha = \alpha_1\alpha_2$. Many distributions used in practice have lighter tails. The Gamma and the log-normal are but two examples of distributions dceaying faster than any polynomial α. Now the limiting over-threshold model is the exponential. Illustrations are provided in Exercises 9.4.3-6.

The Hill estimate

The decay rate α can be determined from historical data (though they have to be plenty). A popular method is the **Hill** estimate

$$\hat{\alpha}^{-1} = \frac{1}{n - n_1} \sum_{i=n_1+1}^{n} \log\left(\frac{z_{(i)}}{z_{(n_1)}}\right) \tag{1.19}$$

where $z_{(1)} \leq \ldots \leq z_{(n)}$ are the data sorted in ascending order and n_1 is user selected. Ideally n_1/n should be close to one and $n - n_1$ large which requires n huge. The Hill estimate is used for general distributions, but (as we saw above) α is also the shape of an approximating Pareto model. There is a link here that can be used to derive the 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

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

see also below. 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

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

coincides (almost) with $\hat{\alpha}$ in (1.19). A number of justifications of the Hill estimate can be found in Chapter 6 of Embrects, Klüppelberg and Mikosch (1997). It is **consistent** and converges to the true value as $n \to \infty$ and $n_1/n \to 1$.

The method is tested among the exercises. It does provide the shape of the over-threshold distribution, but there is a scale parameter β_b too, and we can't use the Pareto model to assess uncertainty without it. A simple estimate is

$$\hat{\beta}_b = \frac{z_{(n_2)} - z_{(n_1)}}{2^{1/\hat{\alpha}} - 1} \qquad \text{where} \qquad n_2 = 1 + \frac{n_1+n}{2}, \tag{1.20}$$

which utilizes that $\beta(2^{1/\alpha}-1)$ is the median under Pareto(α, β) whereas the over-theshold data

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

A possible estimate is therefore $\hat{\beta}_b(2^{1/\hat{\alpha}}-1) = z_{(n_2)} - z_{(n_1)}$ which is (1.20).

Likelihood methods
An alternative way of determining the over-threshold distribution is to apply Pareto likelihood estimation to observations exceeding it. Technically it is a little more work than through the Hill estimate; see Section 7.3 for details.

The Pareto model is also a good example with which to show how **censored** information is utilized. Observations are now in two groups, either the ordinary, fully observed claims $z_1, \ldots, z_n$ or those (n_r of them) known to have exceeded certain thresholds $b_1, \ldots, b_{n_r}$, but not by how much. The log likelihood function for the first group is as in Section 7.3; i.e.

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

whereas for the the censored part we must add contributions from knowing that $\Pr(Z_i > b_i)$. The probability of this happening 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 when all those are taken into account, we obtain the full log likelihood

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

which is to be maximized, a numerical problem very much the same as in Section 7.3

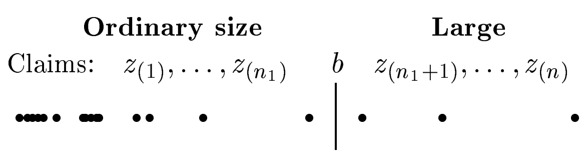
1.5 Large claim situations

Introduction

The big 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. Indeed, insurance companies may give cover to claims *larger* than have been seen earlier. What should our approach be in these situations? The simplest would be to fit a parametric family and extrapolate beyond past experience, but 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. Indeed, such a procedure may easily underestimate big claims risk severely; see Section 9.6. A Pareto model would be more conservative, and then there is the result due to Pickands that points to this distribution as a general description above all large thresholds. There is an idea here, and the purpose of the present section is to develope it.

An approach through mixtures

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 we have selected a threshold b defining 'large' claims as those exceeding it. The original claims $z_1,\dots,z_n$ have been ranked in ascending order as

$$z_{(1)} \le z_{(2)} \dots \le z_{(n)}$$

so that observations from $z_{(n_1)}$ and smaller are below the threshold and those from $z_{(n_1+1)}$ and larger are above. How the threshold b is chosen in practice is discussed below; see also the numerical illustrations in Section 9.6.

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

$$Z = (1 - I_b)Z_{\leq b} + I_b Z_{>b} \tag{1.21}$$

where

$$\underset{\textit{central region}}{Z_{\leq b} = Z|Z \leq b,} \qquad \underset{\textit{extreme right tail}}{Z_{>b} = Z|Z > b} \qquad \text{and} \qquad \begin{aligned} I_b &= 0 \quad \text{if } Z \leq b \\ &= 1 \quad \text{if } Z > b. \end{aligned} \tag{1.22}$$

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 two sides of (1.21) are equal, but at first sight this merely looks complicated. Why on earth can it help us? The point is that we have created a framework reaching out to two different sources of information. To the left of the threshold there is the historical data with which we may identify a model. On the right the result due to Pickands suggests a Pareto distribution. This defines a modelling strategy which will now be developped.

The empirical distribution mixed with Pareto

The preceding argument lead to a two-component approach which can be implemented in many ways. For example, to the left of b we could fit a parametric model. It would extend beyond b, but that may not matter too much; see Exercise ??. Another idea is to use non-parametric modelling, and this is the method that will be developed in detail with the threshold selected as one of the observations. *Choose* some small probability p and let $n_1 = n(1-p)$ and $b = z_{(n_1)}$. Then take

$$Z_{\leq b} = \hat{Z} \qquad \text{and} \qquad Z_{>b} = z_{(n_1)} + \text{Pareto}(\alpha, \beta), \tag{1.23}$$

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

$$\Pr(\hat{Z} = z_{(i)}) = \frac{1}{n_1}, \quad i = 1, \ldots, n_1. \tag{1.24}$$

The remaining part (the delicate one!) are the parameters are α and β 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 provides estimates $\hat{\alpha}$ and $\hat{\beta}$. There are even sophisticated, automated techniques for the selection of p, see ? and ?. In practice you might do just as well with trial and error. An example of this kind is discussed in the next section.

With more limited experience (as is common) is is hard to avoid a subjective element. One of the advantages of dividing modelling into two components is that it clarifies the domain where personal judgment enters. If you take the view that a degree of conservatism is in order when there is insufficient information for accuracy, that can be achieved by placing b low and using 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 $z_{(1)} \leq \ldots \leq z_{(n)}$, 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 present version is extremely quick to implement.

1.6 Searching for the model

Introduction

A final model for claim size is the result of different deliberations. Historical data have typically been utilized through a non-parametric approach or with parametric families. We may also have changed the variable. The idea is then that standard families of distributions may fit a **transformed** variable better than the original one, and with re-transformation afterwards the model again applies to ordinary claims. One of our worries should be model error. Does the distribution selected reflect the uncertainty of real life? If there are small amounts of data to go on, the discrepancy could be huge. Should that lean us towards concervative choices? If accurate mathematical descriptions are beyond reach anyway, it could be an argument in favour of heavy-tailed distributions like Pareto.

The purpose of this section is to indicate how these themes enter by means of two very different examples. We have already met the Norwegian fund for natural disasters in chapter 7 where there were just $n = 21$ historical incidents to rely on. By contrast the so-called Danish fire claims will serve our needs for a 'large' data set. Many authors on actuarial science have used it as a test case; see Embrechts, Klüppelberg and Mikosch(1997) where more on their orgin is given. The historical record comprises $n = 2167$ industrial fires. Damages start at one million Danish kroner (DKK)[3] with 263 as a maximum and with average $\bar{z} = 3.39$, standard deviation $s = 8.51$ and skewness coefficient $\gamma = 18.7$. The latter indicates very heavy tails and strong skewness towards the right. This also emerges clearly from the plots in Figure 9.2 and 9.3 below.

Working with transformations

A useful tool for modelling is to change data by means of a transformation, say $H(z)$. The situation is then as follows:

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

[3] There are about eight Danish kroner in one euro.

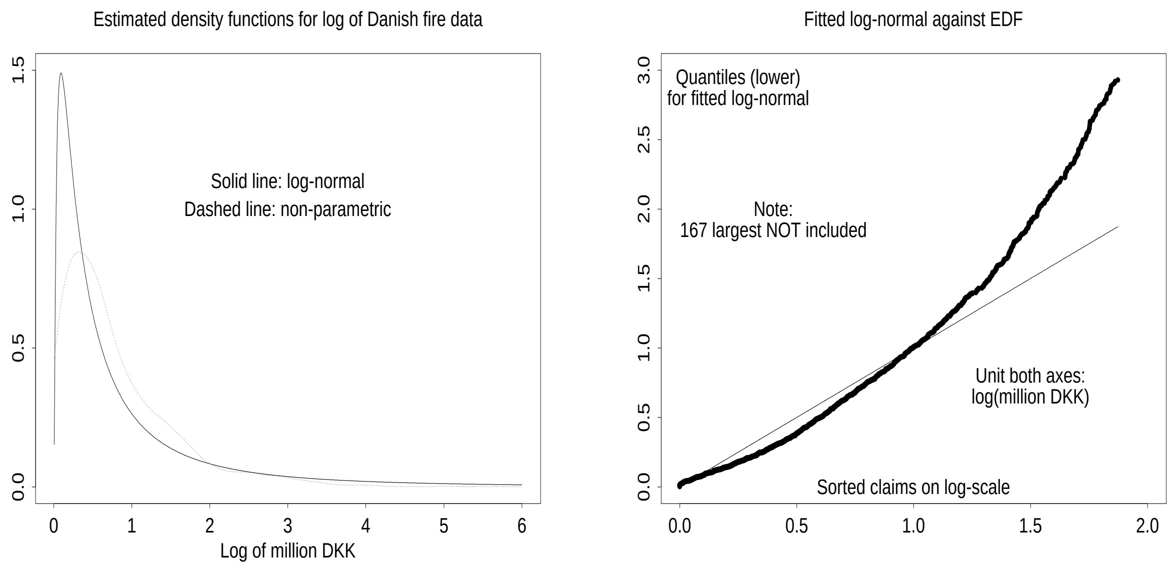


Figure 9.2 The **log-normal** *model fitted the Danish fire data on* **log**-*scale. Density function with kernel density estimate (left) and Q-Q plot (right).*

Modelling is then attacked through $y_1, \ldots, y_n$ and $Y = H(Z)$ instead of the original Z. The idea is to make one of the simple models fit better than could be achieved with Z itself. At the end we re-transform back through $Z = H^{-1}(Y)$ with $Z^* = H^{-1}(Y^*)$ for the Monte Carlo. The log-normal is a familiar example. Then $H(z) = \log(z)$ and $H^{-1}(y) = \exp(y)$ with Y normal. The logarithm is the most commonly used transformation of all. Frequently applied alternatives are powers $Y = Z^\theta$ where $\theta \neq 0$ is a some given index; see also Exercise 9.6.2. The choice of transformations (typically made by trial and error) is a *second* feature that adds flexibility to the usual families of distributions.

Variations on this theme are indeed many. With logaritms we might take

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

and entirely different families of distributions would be used for Y. As an example consider the Danish fire claims where we must take into account that they run from 1 and upwards (in milllion DKK). That makes $Y = \log(Z)$ positive, and one possibility could be the log-normal through

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

Here ε is $N(0,1)$. An alternative is the Gamma familiy. Let $Y_{0\alpha}$ be Gamma distributed with mean one and shape α and consider

$$Z = e^Y, \quad Y = \xi_y Y_{0\alpha} \qquad \text{with estimates} \qquad \hat{\xi_y} = 0.79,\ \hat{\alpha} = 1.16.$$

Both pairs of estimates are likelihood ones.

What is immediately clear from the huge discrepancy in the estimated means ξ_y is that both models can't fit. Indeed, the log-normal doesn't work. Its estimated density function (Figure 9.2

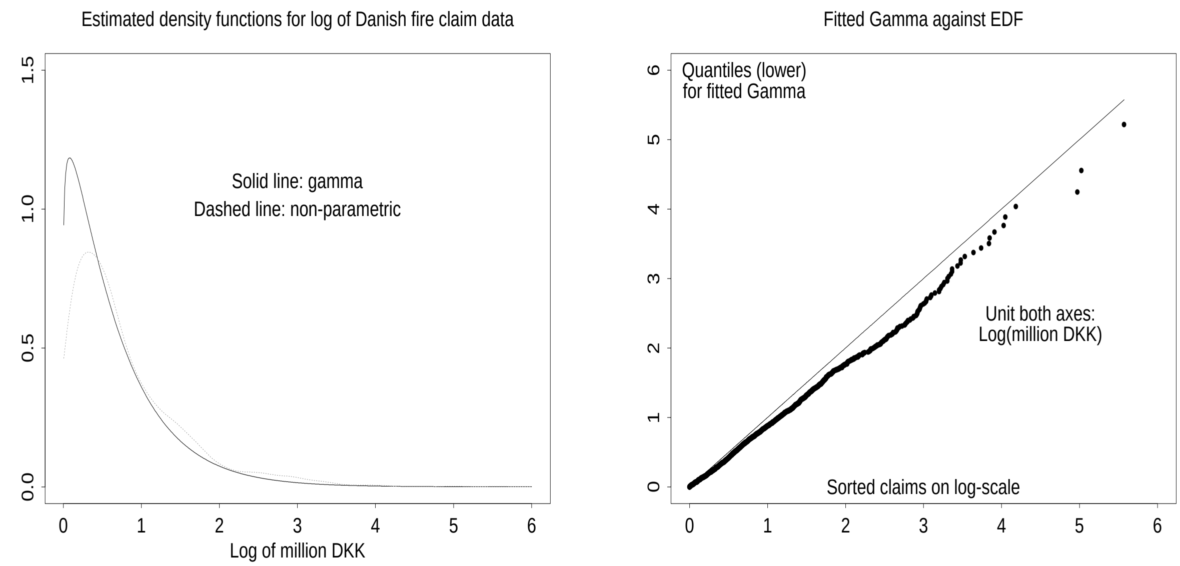


Figure 9.3. The **Gamma** *model fitted the Danish fire data on* **log***-scale. Density function with kernel density estimate (left) and Q-Q plot (right).*

left, horizontal axis on *logarithmic* scale) matches the kernel density estimate poorly, but (as usual) Q-Q plotting (Figure 9.2 right) provides a better view. The right tail of the log-normal is too heavy and exaggerates the risk of extreme claims grossly[4]. By contrast the Gamma fit as displayed in Figure 9.3 is much better. Perhaps the extreme right tail is slightly too light, but the fit isn't an end in itself, and consequences for the evaluation of the reserve is not necessarily serious. That will be examined in Section 10.3; see also Exercise 9.6.2 where a slight modification will improve the fit.

Pareto and Pareto mixing

The Pareto model is so heavy-tailed on its own that it could be tried on the raw Danish fire data directly (without log-transform). It is also a strong candidate for the extreme right tail (Section 9.4). Indeed, with such an extensive data record it is tempting to forget all about parametric families and use the strategy advocated in Section 9.5 using the empirical distribution function for the central part and Pareto on the right. Table 9.2 shows the results of fitting Pareto distributions (through maximum likelihood) over various thresholds b. As b is being raised, the situation should become more and more Pareto-like (Pickand's theorem). Under a strict Pareto regime, the shape parameter α is the same for all b whereas the scale parameter depends on b through $\beta_b = b - 1 + \beta/(\alpha - 1)$; see Exercise ?. Stretching the imagination a bit there are reminiscences of this in Table 9.1 where α is more stable than β; see Exercise 9.4.1 for detailed calculations.

But it would be a gross exaggeration to proclaim the data to be Pareto distributed. Q-Q plots for two of the over threshold distributions is shown in Figure 9.4. There is a reasonable fit on the right (above 5%), but not on the left (above 50%) where the Pareto distribution fitted has heavier tails

[4]Note that the 167 largest observations have been left out to make the resolution in other parts of the plot better

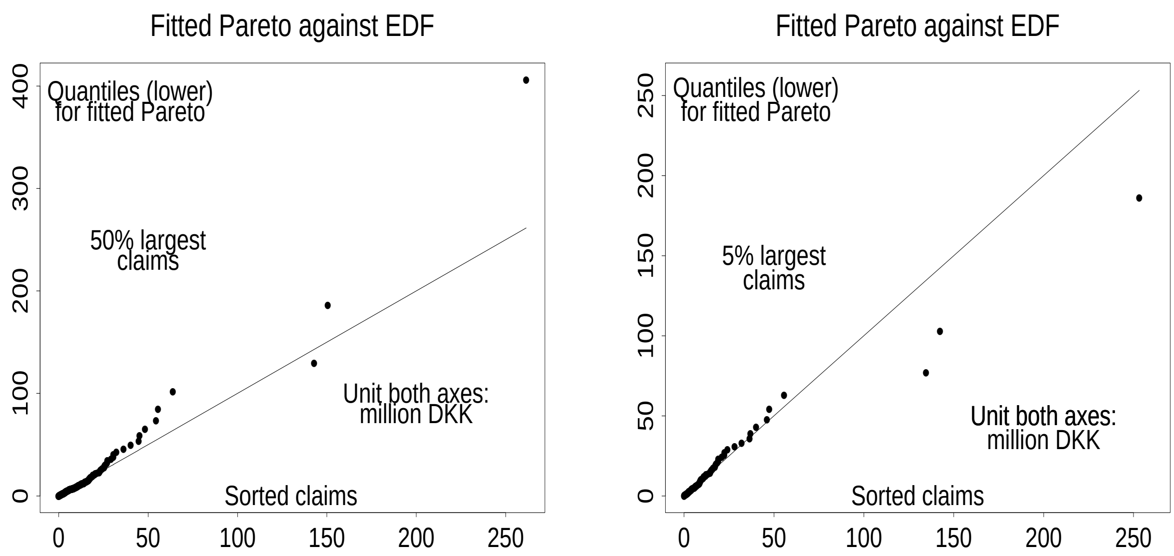


Figure 9.4 Q-Q plots of fitted **Pareto** *distributions against the empirical distribution function,* 50% *largest observations (left) and* 5% *largest (right).*

that the empirical counterpart. Table 9.1 tell us why. The two shape parameters estimated (1.42 and 2.05) deliver quite unequal extreme uncertainty.

Tiny historical records
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 in total, 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 much accuracy, yet somehow a model must be found. One possibility is geophysical modelling. Natural disasters are then simulated in the computer and their cost counted from detailed, physical descriptions of houses and installations. Evaluations of this kind are carried out around the world, but they are outside out natural range of topics, and we shall concentrate on what can be extracted from historical losses.

If you fit the Gamma and Pareto family to the natural disasers by maximum likelihood, the results look like this:

Unit: Million DKK

	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.

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).*

Shape	Mean	5%	1%
0.72	179	603	978

Gamma family

Shape	Mean	5%	1%
1.71	200	658	1928

Pareto family

These are *very* different families of distributions, yet their discrepancies, though considerable, are not enormous in the central region (say up to the upper 5% percentile). For the *very* large claims that changes, and the Pareto 1% percentile is twice that of the 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 a more delicate one.

The right family: Impossible?

Incidentally, how impossible is it to determine the family from small amounts of data? Suppose a Q-Q plot is used. A given family such as Gamma or Pareto is then evaluated by comparing their estimated percentiles $\hat{q}_i$ to empirical ones $z_{(i)}$ where the former correspond to distributions fitted the data. What is actually done when the two sequences are matched, is unclear (different ways for different people), but perhaps some try to minimize

$$Q = \sum_{i=1}^{n} |\hat{q}_i - z_{(i)}|. \tag{1.25}$$

This criterion has been proposed as basis for formal goodness of fit tests in Devroye (1971). It could be that humans do it a little better, but results using other critera didn't deviate that much from those in Table 9.2.

Monte Carlo experiments were run with $m = 1000$ replications according to the following scheme:

True model			Parametric family tried		
Pareto			*fitting* $\hat{q}_1^* \geq \ldots \geq \hat{q}_n^*$		
or	$\longrightarrow$	$z_1^*, \ldots, \geq z_n^*$	$\longrightarrow$	$\longrightarrow$	$Q^* = \sum_i \|z_{(i)}^* - \hat{q}_i^*\|.$
Gamma		*historical data*	*sorting* $z_{(1)}^* \geq \ldots z_{(n)}^*$		

Simulated historical data were drawn from the Pareto or Gamma model on the left and the model (possibly a different one!) fitted. That gave estimated percentiles $\hat{q}_i^*$ which could be compared to purely empirical ones $z_{(i)}^*$ and a value of the criterion Q^* computed for the parametric model tried. When Pareto, Gamma and log-normal were fitted to the same historical data, we obtain three different evaluations Q^*, and the distribution corresponding to the smallest, best-fitting one was picked. The selection statistics is shown in Table 9.2. It is clearly impossible to choose between the three models when there are only $n = 21$ claims. The chance is improved with $n = 80$ and with

$m = 1000$ *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**
Model selected	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* $\hat{\theta} = \hat{q}_{0.01}/q_{0.01}$ *where* $\hat{q}_{0.01}$ *is fitted and* $q_{0.01}$ *true* 1% *percentiles of claims.* **Bold face***: Correct parametric family used.*

$n = 400$ (not shown) the success probability was about $0.90 - 0.95$.

Data in short supply: What then?

The preceding experiment showed the futility of trying to identify models from small amounts of historical data, but when faced with such situations, how should they be attacked? Here are some tentative suggestions. A good deal hinges on the maximum responsibility b per claim. If it is *smaller* than the *largest* observation $z_{(1)}$, it could be a case for the empirical distribution function. That doesn't help us much with the Norwegian natural disasters from Section 7.4 where b is *much larger* than $z_{(1)}$, and risk would be grossly under-estimated by that method. Surely the Pareto distribution is one of the leading contenders now. It is a conservative choice (which seems sensible), possibly estimation errors undermine some of that caution.

These points are illustrated by the experiment in Table 9.3 where the issue is the consequences of being wrong. For example, if the underlying distribution is a member of the Gamma family, how does a Pareto fit perform? Or what about estimated Gamma percentiles when the true model is Pareto? Clauses of maximum payments have much bearing on this (as mentioned), but these problems can also be inspected through

$$\hat{\theta} = \frac{\hat{q}_\varepsilon}{q_\varepsilon} \qquad \text{for} \qquad \varepsilon = 1\%.$$

Patterns in how $\hat{\theta}$ deviate from 1 reveal the impact of model and estimation error jointly. Suppose the Gamma family is fitted to claims that are actually Pareto distributed. It then emerges from Table 9.3 (Line two from bottom) that the 90% percentile of $\hat{\theta}$ is at most one; i.e. $q_{0.01}$ is almost certain to be *under*-estimated! The tendency is reversed when the Pareto model is applied to Gamma-distributed losses. Now the percentile is *over*-estimated. Certainly, we *are* doing something silly, and yet in practice we might not know. The method that comes on top in Table 3 is the last one where the percentiles are computed form the best-fitting of both the Gamma and Pareto distributions, i.e. the alternative minimizing (1.25) has been picked. Now the the distribution of $\hat{\theta}$ varies around one, though with huge errors.

In summary it seems sensible to try determine the family empirically even for small data sets (though we often guess wrong). If we go for conservatism and caution, the Pareto model may be the answer despite the huge uncertainty of the fitted parameters.

1.7 Further reading

1.8 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 $z_1, \dots, z_n$ 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 (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 $(i = 1, \dots .n)$ 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 (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	163.8000	

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

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

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 $\mathrm{sd}(Z) = e^{\tau^2/2}\sqrt{e^{\tau^2} - 1}$ to deduce that

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

e) Show that $\mathrm{skew}(Z) \to 0$ as $\tau \to 0$ and calculate $\mathrm{skew}(Z)$ 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 (1.20) 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 (1.20). **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 $z_{(1)} \leq \cdots \leq z_{(n)}$. 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.