20

Extreme value theory

Syllabus objectives

4.5 Explain the importance of the tails of distributions, tail correlations and low frequency / high severity events.

4.6 Demonstrate how extreme value theory can be used to help model risks that have a low probability.

0 Introduction

In this module we look more closely at the tails of distributions. In Section 2, we introduce the concept of extreme value theory and discuss its importance in helping to manage risk.

The importance of tail distributions and correlations has already been mentioned in Module 15. In this module, the idea is extended to consider the modelling of risks with low frequency but high severity, including the use of extreme value theory.

Note the following advice in the Core Reading:

Beyond the Core Reading, students should be able to recommend a specific approach and choice of model for the tails based on a mixture of quantitative analysis and graphical diagnostics. They should also be able to describe how the main theoretical results can be used in practice.

Module 20 – Task list

	Task	*Completed*
1	**Read** Section 1 of this module and **answer** the self-assessment questions.	
2	**Read:** • **Sweeting, Chapter 12, pages 286 – 293**	
3	**Read** the remaining sections of this module and **answer** the self-assessment questions. This includes relevant **Core Reading** for this module.	
4	**Attempt** the practice questions at the end of this module.	
5	**Review** the Module Summary.	

This page is deliberately left blank
so that you can easily remove and use the task list.

1 Tail events and modelling

Events that occur with low frequency and high severity can have a devastating impact on companies, investment funds *etc*. These events are modelled using the tail of a distribution. Given their high impact, it is essential that they are modelled accurately. For this reason, much research has been focused on this modelling.

Coefficients of tail dependence have already been introduced in Module 18.

Recall that a coefficient of tail dependence measures the strength of dependence between two or more variables in the tail of a multivariate distribution.

1.1 Low frequency / high severity events

Low frequency / high severity events can have a devastating impact on companies and investment funds. However, their low frequency means that little data exists to model their effects accurately.

The financial crisis ('credit crunch') that started in 2007 is an example of a low frequency / high severity event. Stock market falls occur with regularity and credit spreads have also widened historically (*eg* after the LTCM bailout in 1998). However, the financial crisis generated more extreme movements in equity values and credit spreads than had been seen for over 20 years previously. Indeed, there have been few events of such extremity in the last 100 years.

Question

List other low-frequency high-severity events you can recall.

Solution

On 19 October 1987 ('Black Monday') the S&P500 index fell by the unprecedented amount of 20.5%. Estimation of the return period for a 20.5% loss indicates that this is an event which we would expect to occur once in every 2,100 years.

Other examples of such events include terrorism (*eg* the 9/11 attack on the twin towers) and natural disasters (*eg* the 2013 Philippines typhoon Haiyan).

Modelling the full distribution allows the tail to be modelled.

This can overcome the lack of data from severely stressed time periods, *ie* the data from both stressed and non-stressed periods is used to fit the distribution.

However, it is important to ensure that the form of the distribution is correct in the tails. Typically financial data is much more narrowly peaked and has fatter tails than the normal distribution (that is, the data is *leptokurtic*). Thus, when equity values are modelled, extreme events occur more frequently than predicted by the normal distribution: hence, the normal distribution may be of limited use for modelling low frequency / high severity events.

As an example, in Module 16, we found that the t-distribution has fatter tails than the normal distribution.

The fat tails that are observed in financial data are normally the result of two factors:

1. **returns are *heteroscedastic* (that is, the volatility varies over time in a stochastic way)**
2. **the innovations in a heteroscedastic model are best modelled using a fat-tailed distribution.**

However, even if a fat-tailed distribution is fitted to the whole of the dataset, a poor job might be done of fitting the tails of the data, since the parameter estimates are so heavily influenced by the main bulk of the data in the middle of the distribution.

Fortunately, better modelling of the tails of the data can be done through the application of extreme value theory.

2 Extreme value theory

2.1 Defining what is extreme

One approach is to consider an 'extreme value' as being the maximum value in a set of n losses, $X_M = \max(X_1, X_2, \ldots, X_n)$, being referred to as a *block maximum*.

Alternatively, we can consider all losses that exceed a certain threshold amount as being 'extreme'.

2.2 Limiting behaviour

If we are dealing with losses that have typical sizes, *ie* ones whose values come from the central part of the distribution, we can make use of the Central Limit Theorem, which tells us that the standardised value of the average loss, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, can be approximated using a normal distribution.

Here, $\bar{X}$ is the mean of a set of n values taken from a loss distribution that has mean μ and variance σ^2.

However, the most financially significant part of a loss distribution is usually the right-hand tail where the large losses occur. These are the *extreme values* of the distribution. So, is there a similar way to approximate the behaviour of the extreme values in the tail of the distribution?

One approach is to look at $X_M = \max(X_1, X_2, \ldots, X_n)$, the maximum value in a set of n losses – referred to as a *block maximum*.

If we look at a number of such blocks then these maxima can be standardised in a similar way, *ie* we calculate expressions of the form $\frac{X_M - \alpha_n}{\beta_n}$. These standardised values can be approximated by a particular type of distribution – an *extreme value distribution*.

A similar result also holds if we consider the expected size of losses that exceed a certain threshold amount.

The key idea of extreme value theory is that there are certain families of distributions (the Generalised Extreme Value and Generalised Pareto) which describe the behaviour of the tails of many distributions.

Extreme value theory has two main results, which may be roughly stated as follows:

Result 1 – block maxima

The distribution of the standardised block maxima $X_M = \max(X_1, X_2, \ldots, X_n)$ is approximately described by the Generalised Extreme Value (GEV) family of distributions if n is sufficiently large.

Result 2 – threshold exceedances

The tail of the distribution above a threshold, $P(X > x + u \mid X > u)$, can be approximated, for large values of u, by the Generalised Pareto Distribution (GPD).

It is these results and the properties and applications of the resulting distributions that we will consider in the following sections.

3 The Generalised Extreme Value (GEV) distribution

Just as the normal distribution proves to be the important limiting distribution for sample sums or averages, as is made explicit in the Central Limit Theorem, the Generalised Extreme Value (GEV) family of distributions proves to be important in the study of the limiting behaviour of sample extremes.

3.1 Extreme values theorem (EVT)

We are interested in the distribution of $X_M = \max(X_1, X_2, ..., X_n)$ where each X_i is an observed loss. If the losses are independent and identically distributed (*iid*), with cumulative distribution function F:

$$\begin{aligned} P(X_M \leq x) &= P(X_1 \leq x,\ X_2 \leq x,\ \dots,\ X_n \leq x) \\ &= P(X_1 \leq x)P(X_2 \leq x)...P(X_n \leq x) \\ &= \left[P(X_1 \leq x)\right]^n \\ &= \left[F(x)\right]^n \\ &= F^n(x) \end{aligned}$$

We can standardise this for a sequence of real constants $\beta_1, ..., \beta_n > 0$ and $\alpha_1, ..., \alpha_n$, and consider the limit as n increases:

$$H(x) = \lim_{n \to \infty} P\left(\frac{X_M - \alpha_n}{\beta_n} \leq x\right) = \lim_{n \to \infty} F^n(\beta_n x + \alpha_n)$$

This result applies for all commonly used statistical distributions from which the X_i may originate.

Question

Determine $H(x)$, the limiting distribution for the standardised maximum loss from a set of observations, where the individual losses are distributed exponentially (*ie* $F(x) = 1 - e^{-\lambda x}$), by setting $\beta = \frac{1}{\lambda}$ (*ie* $\beta_n = \beta$ for all n) and $\alpha_n = \frac{1}{\lambda}\ln n$. (*Hint:* $e^x = \lim_{n \to \infty}\left(1 + \frac{x}{n}\right)^n$ by definition.)

Solution

If we set $\beta = \frac{1}{\lambda}$ (*ie* $\beta_n = \beta$ for all n) and $\alpha_n = \frac{1}{\lambda}\ln n$ then we have:

$$\beta x + \alpha_n = \frac{x}{\lambda} + \frac{\ln n}{\lambda}$$

Hence, since $F(x)=1-e^{-\lambda x}$, we have:

$$F(\beta x+\alpha_n)=F\left(\frac{x+\ln n}{\lambda}\right)$$

$$=1-e^{-\lambda\left(\frac{x+\ln n}{\lambda}\right)}$$

$$=1-\frac{1}{n}e^{-x}$$

Therefore, $F^n(\beta x+\alpha_n)=\left(1+\frac{-e^{-x}}{n}\right)^n$, where $x\geq -\ln n$

and so, since $e^x=\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n$ by definition, the distribution of the standardised maxima is:

$$\lim_{n\to\infty}F^n(\beta x+\alpha_n)=e^{-e^{-x}}.$$

3.2 GEV distribution

The *generalised extreme value* (GEV) family of distributions has three parameters:

1. α is the location parameter.

2. $\beta>0$ is a scale parameter.

   These two parameters just rescale (shift and stretch) the distribution. They are analogous to, but do not usually correspond to, the mean and standard deviation.

3. γ is the shape parameter.

The cumulative distribution function for the GEV family is:

$$H(x)=\lim_{n\to\infty}P\left(\frac{X_M-\alpha_n}{\beta_n}\leq x\right)=\begin{cases}\exp\left(-\left(1+\dfrac{\gamma(x-\alpha)}{\beta}\right)^{\frac{-1}{\gamma}}\right) & \gamma\neq 0\\[2ex] \exp\left(-\exp\left(-\dfrac{(x-\alpha)}{\beta}\right)\right) & \gamma=0\end{cases}$$

The CDF for the *standard GEV distribution* is given by $\alpha=0$ and $\beta=1$.

The sign of γ (positive, negative or zero) is important and leads to three corresponding three distributions (which are named after their original discoverers).

Question

Describe the three distributions that form the GEV family. (*Hint: name them, sketch examples, refer to any bounds and give examples of their applications*.)

Solution

If $\gamma = 0$, the distribution is a *Gumbel GEV distribution*. This has a tail that falls exponentially.

For $\gamma < 0$, the distribution is known as a *Weibull GEV distribution*. This has a finite upper bound indicating an absolute maximum. We might expect to fit such a distribution to natural phenomenon, for example:

- temperature
- wind-speed
- the ages of a human population (indicating an upper bound to possible age)

or where a loss was certain not to exceed a certain value (for example, if such losses were reinsured).

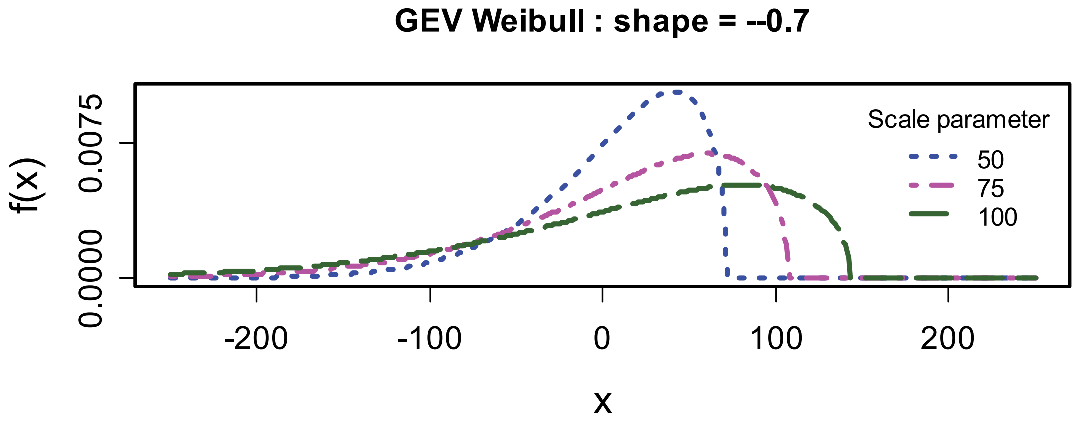
Note that this is not the same Weibull distribution as the one in the Actuarial Tables, which you may be familiar with from earlier subjects.

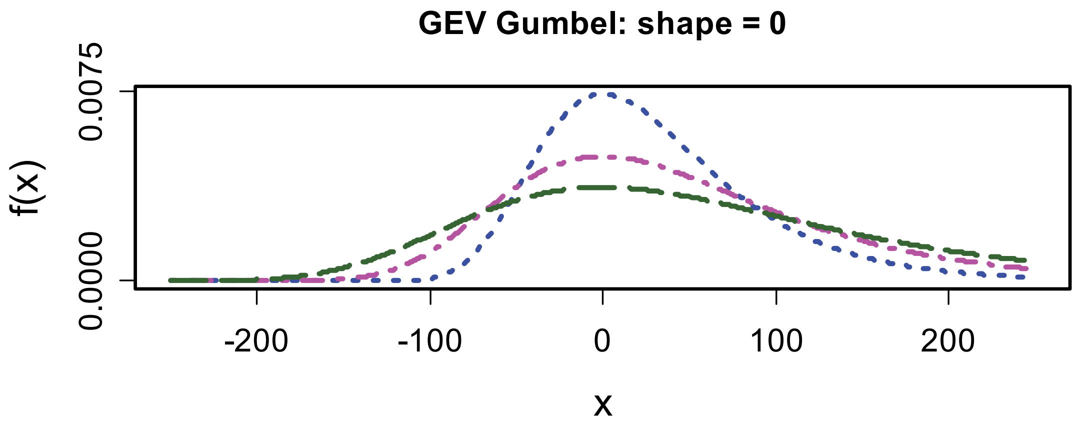
For $\gamma > 0$, the distribution is a *Fréchet-type GEV distribution*. The tails of the distribution become heavier (with infinite variance if` $\gamma > \frac{1}{2}$) and follow a power law. This is typically the subset of distributions most suitable for modelling extreme financial (loss) events. These distributions have a lower bound since:

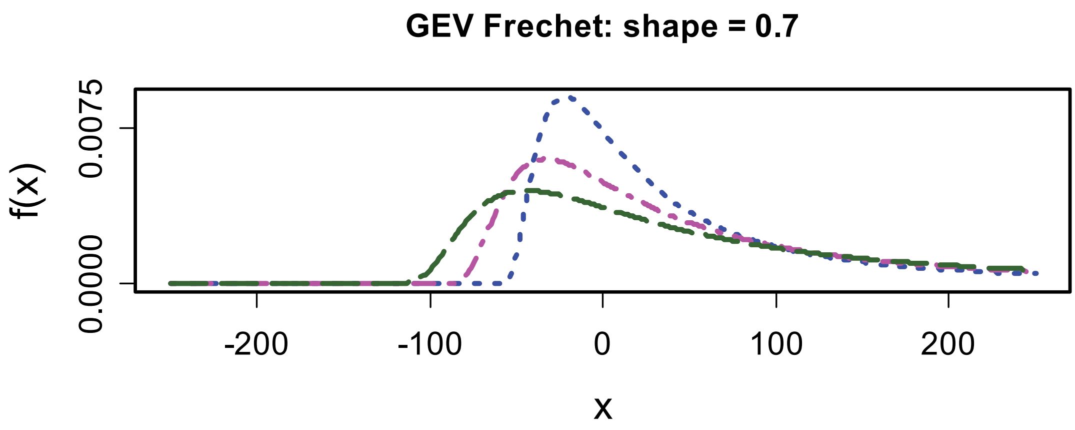
$$x \text{ takes values such that } 1 + \frac{\gamma(x-\alpha)}{\beta} > 0 \;\Rightarrow\; x - \alpha > \frac{-\beta}{\gamma} \;\Rightarrow\; x > \alpha - \frac{\beta}{\gamma}.$$

Example graphs are sketched overleaf.

GEV density function examples







3.3 Underlying distributions

If we know the form of the underlying loss distribution (*eg* exponential), it is possible to work out the limiting distribution of the maximum value.

We can then classify the resulting distribution into one of the three types according to the sign of γ, as shown in the table below.

	GEV distributions (for the maximum value) corresponding to common loss distributions		
Type *Shape parameter*	*WEIBULL* $\gamma < 0$	*GUMBEL* $\gamma = 0$	*FRÉCHET* $\gamma > 0$
Underlying Distribution	Beta Uniform Triangular	Chi-square Exponential Gamma Log-normal Normal Weibull (*)	Burr *F* Log-gamma Pareto *t*
Range of values permitted	$x < \alpha - \dfrac{\beta}{\gamma}$	$-\infty < x < \infty$	$x > \alpha - \dfrac{\beta}{\gamma}$

** Unhelpfully, the extreme value distribution corresponding to the Weibull distribution from the Tables is actually of the Gumbel type!*

Mathematicians have devised various sets of criteria that can be used to predict which family a particular distribution belongs to. As a rough guide:

- underlying distributions that have finite upper limits (*eg* the uniform distribution) are of the Weibull type (which also has a finite upper limit).
- 'light tail' distributions that have finite moments of all orders (*eg* exponential, normal, log-normal) are typically of the Gumbel type
- 'heavy tail' distributions whose higher moments can be infinite are of the Fréchet type.

3.4 Fitting a GEV distribution

Since the GEV distribution describes the distribution of maximum values, we need to subdivide the available data into groups (or blocks) and calculate the maximum for each group.

Return levels and return periods

The GEV distribution can be used to analyse a set of observed losses in two different ways.

Question

Describe the return-level and return-period approaches, and discuss the choice of the number of blocks under the return-level approach.

Solution

The GEV distribution can be used to analyse a set of observed losses in two different ways.

1. select the maximum observation in each block (the *return-level* approach)
2. count the observations in each block that exceed some set level (the *return-period* approach).

The larger the number of blocks, the fewer observations there are in each block. If using the return-level approach this gives less information about extreme values (1-in-a-hundred events rather than 1-in-a-thousand, say). However, it gives more 'extreme' values to fit against and so reduces the variance of parameter estimates.

Alternatively, using fewer blocks (with more observations in each) gives more information about the extreme values under the return-level approach, but the variance of parameter estimates is greater.

Parameterisation

Once the data selection has been done, we can estimate the parameters for the GEV distribution using maximum likelihood estimation or the method of moments.

For example, if we have data divided into m blocks of size n and denote the maximum of the *j*th block by M_{nj}, *ie* we have observed maximum data $M_{n1},\ldots,M_{nm}$.

Using the density function of the GEV distribution, $h(x)$, we can calculate the log-likelihood to be:

$$\ln L(\gamma,\alpha,\beta;\mathbf{M})=\sum_{i=1}^{m}\ln h(M_{ni})$$

This can be maximised subject to the constraints that $\beta>0$ and $\left(1+\gamma\dfrac{M_{ni}-\alpha}{\beta}\right)>0$.

3.5 Advantages and disadvantages of the GEV distributions

We can also use GEV distributions to investigate the limiting distributions for the minimum values of a distribution. If $H(x)$ is the limiting distribution for the standardised maximum value for a particular γ, then $1-H(-x)$ is the limiting distribution for the standardised minimum value from the same original distribution.

Question

Outline two key limitations of the GEV approach.

Solution

1. A key disadvantage of the GEV approach is that a lot of data (and hence information) is lost (as everything apart from the maxima in each block is effectively ignored).

2. The choice of block size can be subjective, and represent a compromise between granularity (*eg* 1-in-100 or 1-in-1000 estimates) and parameter uncertainty.

4 The Generalised Pareto distribution (GPD)

As an alternative to focusing upon a single maximum value, we can consider *all* the claim values that *exceed* some threshold as extreme values. For large samples, the distribution of these extreme values converges to the Generalised Pareto Distribution (GPD). In order to fit the tail of a distribution we need to select a suitably high threshold and then fit the GPD to the values in excess of that threshold.

4.1 Threshold exceedances

If we let X be a random variable with the distribution function F then the *excess distribution over the threshold u* has the distribution function:

$$F_u(x) = P(X - u \le x | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)}$$

for $0 \le x < x_F - u$ where $x_F \le \infty$ is the right endpoint of X.

If the losses are independent and identically distributed (*iid*) then, as the threshold increases, the distribution of the conditional losses, $\lim_{u \to \infty} F_u(x)$, will converge (whatever the underlying distribution of the data) to a *Generalised Pareto distribution*, $G(x)$.

The GPD is a two-parameter distribution with CDF:

$$G(x) = \begin{cases} 1 - \left(1 + \dfrac{x}{\gamma\beta}\right)^{-\gamma} & \gamma \ne 0 \\ 1 - \exp\left(-\dfrac{x}{\beta}\right) & \gamma = 0 \end{cases}$$

$\beta > 0$ is a scale parameter and γ is the shape parameter.

The CDF for the *standardised GPD* is given by setting $\beta = 1$.

Question

Describe the GPD distribution. (*Hint: sketch examples and refer to any bounds.*)

Solution

For the GPD:

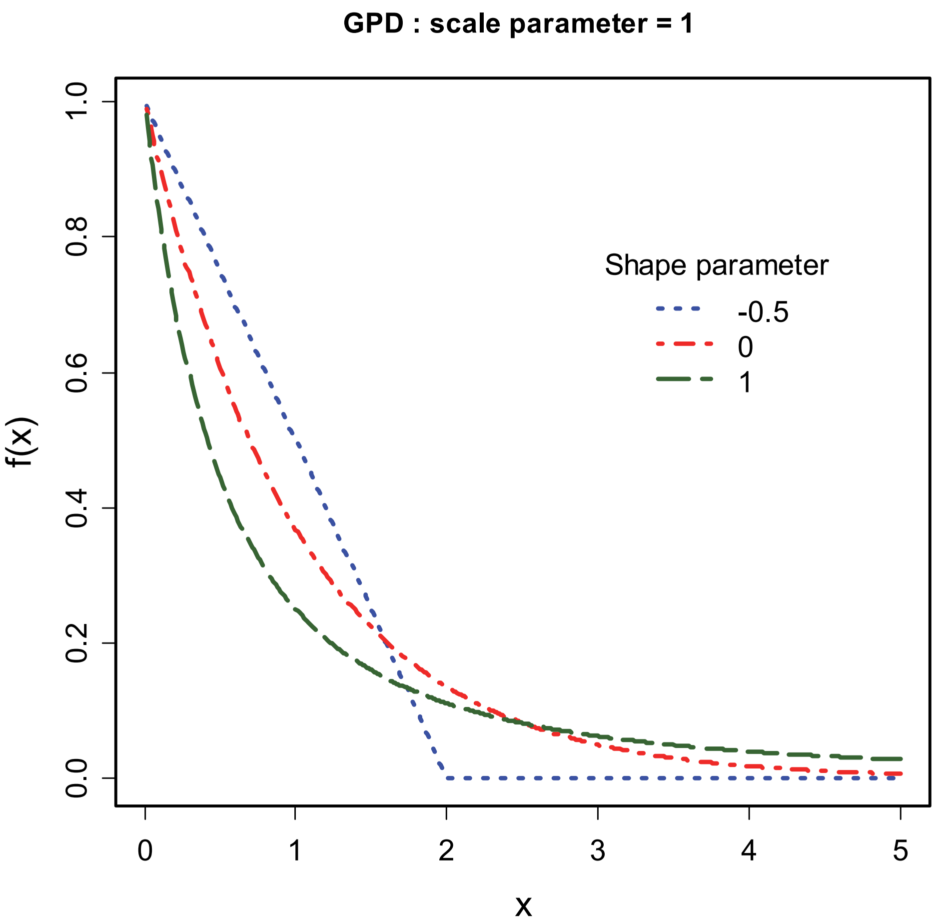
- there is a lower bound ($x \ge 0$) when $\gamma > 0$
- there is also an upper bound ($0 \le x \le -\gamma\beta$) when $\gamma < 0$
- if $\gamma > 0$ then the GPD becomes the Pareto distribution.

The mean of the GPD can be calculated as follows (provided $\gamma > 1$):

$$E(X) = \frac{\gamma\beta}{\gamma - 1}.$$

It can also be shown that if $\gamma > 0$ then $E\left(X^k\right) = \infty$ for $k \geq \gamma$. For example if $\gamma \leq 2$ then we have $E\left(X^2\right) = \infty$, *ie* we have infinite variance.

Examples of GPD density functions are sketched below:



4.2 Asymptotic property

It can be shown that when the standardised *maxima* of a distribution converge to a GEV distribution (as discussed in the preceding section, this is true for all commonly used statistical distributions), the *excess* distribution converges to a GPD distribution with an equivalent shape parameter γ.

4.3 Mean excess function

A useful function, which will later help us fit a GPD, is the *mean excess function*.

The *mean excess function* is defined as:

$$e(u) = E(X - u | X > u).$$

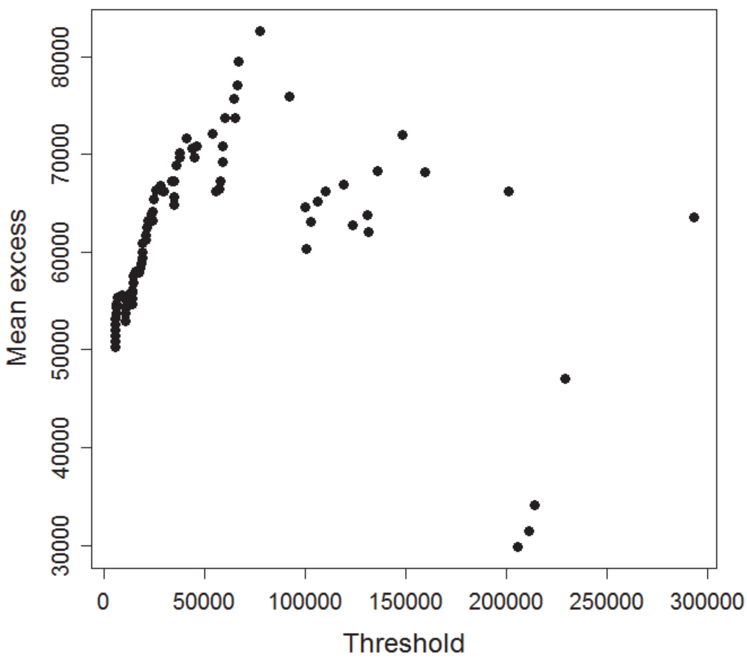
The *empirical mean excess function,* $e_N(u)$, is calculated for each of the N observed data points (using the following formula), by setting $u = X_i$ for $i = 1,...,N$:

$$e_N(u) = \frac{\sum_{i=1}^{N}(X_i - u)\, I(X_i > u)}{\sum_{i=1}^{N} I(X_i > u)}$$

Recall that $I(X_i > u)$ *is an indicator function that takes the value 1 if* $X_i > u$ *and 0 otherwise.*

Example – mean excess plot

An example mean excess plot is shown in the following graph:



On this graph, each of the observed losses has been treated as a potential threshold (on the *x*-axis), and the mean excess loss based on that threshold has been calculated and plotted (on the *y*-axis).

Question

Derive the mean excess function if $X \sim Exp(\lambda)$. (*Hint: the exponential distribution exhibits the 'memoryless' property*.)

Solution

The exponential distribution exhibits the 'memoryless' property, *ie* the expected future waiting time for an event to occur is independent of the waiting time already elapsed.

So, if $X \sim Exp(\lambda)$, then:

$$F_u(x) = P(X \leq x + u | x > u) = P(X \leq x) = F(x)$$

So the excess distribution is also $Exp(\lambda)$ and the mean excess function is simply $e(u) = \frac{1}{\lambda}$, which is independent of the value of u.

4.4 Fitting a GPD

Here, rather than deciding on the periods over which we are taking maxima (which is the key decision on the underlying data for the GEV) we are required to decide on the threshold u, above which we will approximate the underlying distribution with the GPD.

The choice of u should reflect the context. For example, it might represent the excess over which a reinsurer might be liable to cover claims against an insurer.

Typically u is likely to be around the 90-95th percentile of the complete distribution, as a lower threshold would bring into consideration some values that are not in the tail (*ie* not extreme values).

Question

Discuss how to choose the threshold above which the GPD should be fitted to the observed excesses. (*Hint: consider trade-offs and a graphical method of selection*.)

Solution

There is a trade-off between the quality of the approximation to the GPD (good for high u) against the level of bias in the fit we achieve (good for low u).

If we pick a high value for u, the asymptotic properties apply more accurately (because we are 'closer to infinity'). However, we will have relatively few data values in this region to work with, so our estimates may be unreliable.

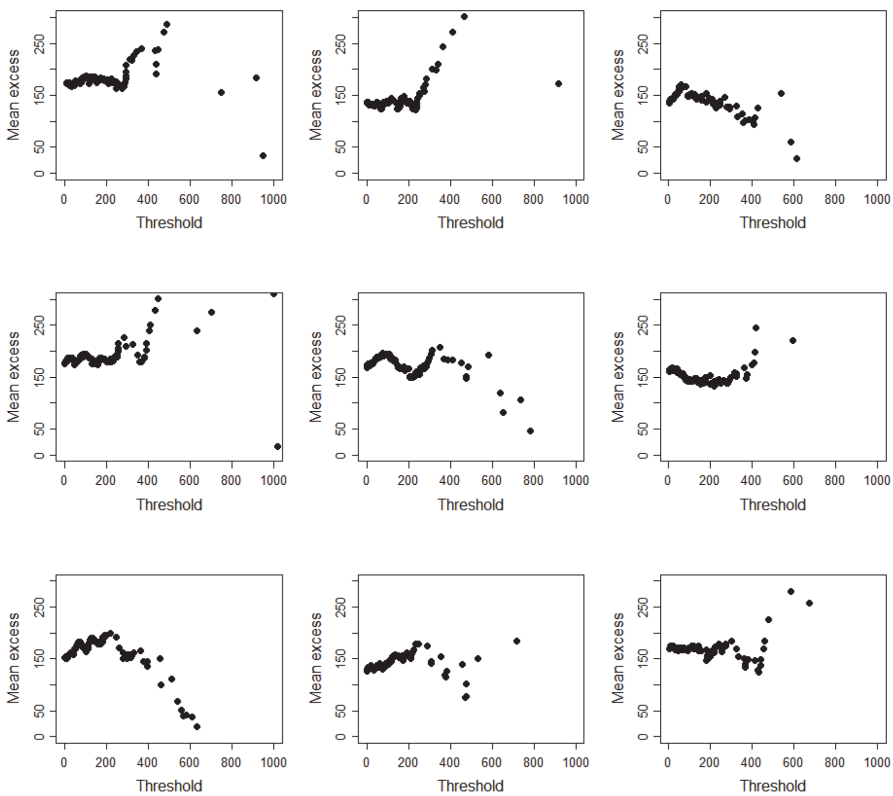
Recall that the asymptotic property is that when the *maxima* of a distribution, the M_{ni}, converge to a GEV distribution, then the *excess* distribution, $F_u(x)$, converges to a GPD distribution with an equivalent shape parameter γ. It can be shown that under these circumstances the mean excess function, $e(u)$, will become a linear function of u. So, if we plot $e(u)$ against u, then we should see linearity appear for higher values of u. It can be shown that the slope of this line is related to the shape parameter γ therefore facilitating the fitting of a GPD distribution to the empirical excess distribution.

So, in order to select a suitable threshold (above which we will fit a GPD to the data), we are looking for the mean excess to be linear in u. A graphical test for tail behaviour can then be based on the empirical mean excess function.

In the previous solution, the mean excess function was constant (*ie* its slope was zero) for an exponential loss distribution (which has shape parameter $\gamma=0$). However, for other loss distributions, the mean excess function will slope upwards if γ is positive and downwards if γ is negative.

Example – mean excess simulations

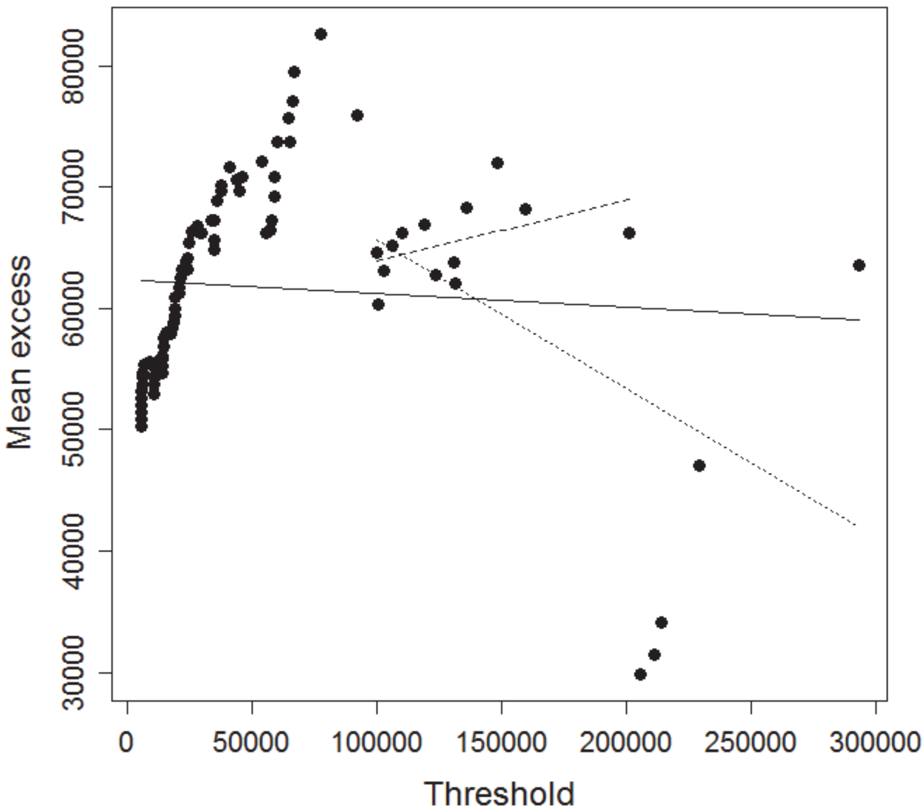
Here is a selection of mean excess loss plots we have simulated. Each of the simulations is based on the same underlying exponential distribution. As you can see, there is a large random element to their appearance and many of them are far from horizontal. This illustrates the difficulty of selecting a suitable threshold and hence value of γ.



Example – fitting a GPD

The graph below illustrates possibilities for choices of thresholds for the mean excess example sketched earlier.

The longest (solid) line is a least squares linear regression on the whole data set. The shorter (dotted) line is a regression on just the tail (above 100,000). However, if the final five data points are ignored, as they distort the picture, then a regression produces the shortest (dashed) line. The slopes, and therefore the fitted GPD parameters, are very different – demonstrating the subjectivity of this part of the fitting process.



4.5 Parameterisation

Above the chosen threshold, we can fit a GPD to the selected data by using standard techniques such as maximum likelihood estimation or the method of moments to calculate γ and β.

Assume that our original data, before the application of the cut-off threshold u, is $X = \{X_1, \ldots, X_n\}$, and is assumed to be independent and identically distributed.

Let $Y_j = X_j - u$ denote the amount of the excess loss when the threshold is exceeded.

We assume the Y_j follow a GPD, not the X_j. We will assume a random number of observations N_u exceed the threshold.

If we attempt to fit a GPD with parameters γ and β, and probability density function $g(x)$, to the data $(Y_1, Y_2 \ldots Y_n)$ then we get the following log-likelihood function:

$$\ln L(\gamma, \beta; Y) = \sum_{j=1}^{N_u} \ln g(Y_j)$$

The domain of valid values for x will depend on the values of the parameters γ and β.

The maximum likelihood methodology is used in practice for both independent and dependent data. Care must be taken to understand the shortcomings and limitations of the method when dependent data is used.

Module 20 Summary – Extreme value theory

Low frequency / high severity events

Lack of data (especially from severely stressed time periods) makes such extreme events hard to model accurately. Modelling the full distribution can help overcome this difficulty. But, the form of distribution may still be incorrect in the tails, *eg* where:

- the 'true' distribution is more skewed or leptokurtic than is indicated by the available data
- the parameter estimates are inappropriately influenced by the main bulk of the data in the middle of the distribution
- features change over time, *eg* heteroscedasticity, structural breaks.

Better modelling of the tails of the data can be done through the application of extreme value theory.

EVT – the GEV distribution

If:

- losses X_i are *iid* with cumulative distribution F
- $X_M = \max(X_1, X_2, \ldots, X_n)$ are the *block maxima*
- $\beta_1, \ldots, \beta_n > 0$ and $\alpha_1, \ldots, \alpha_n$ are a suitable sequence of real constants

then, if n is sufficiently large, the distribution of the standardised block maxima $\dfrac{X_M - \alpha_n}{\beta_n}$ is approximately described by the *Generalised Extreme Value* (GEV) family of distributions:

$$H(x) = \lim_{n\to\infty} P\left(\frac{X_M - \alpha_n}{\beta_n} \le x\right) = \lim_{n\to\infty} F^n(\beta_n x + \alpha_n) \ .$$

The three parameters of the GEV family determine:

- location (α)
- scale (β)
- shape (γ).

The GEV distribution can be used to analyse a set of observed losses in two different ways.

1. select the maximum observation in each block (the return-level approach)
2. count the observations in each block that exceed some set level (the return-period approach).

Parameters for the GEV can be estimated using MLE or the method of moments. The choice of the number of blocks determines the trade-off between the granularity of information provided and the variance of parameter estimates.

A disadvantage of this approach is that it ignores a lot of data (all non-maxima).

EVT – the GPD distribution

Let X be a random variable with the distribution function F. If the losses are independent and identically distributed (*iid*) then, as the threshold increases, the distribution of the conditional losses (the *exceedances*), will converge (whatever the underlying distribution of the data) to a Generalised Pareto distribution, *ie*:

$$G(x) = \lim_{u\to\infty} F_u(x) = \lim_{u\to\infty} P\left(X-u \le x \middle| X>u\right) = \lim_{u\to\infty} \frac{F(x+u)-F(u)}{1-F(u)}$$

The two parameters of the GPD family determine scale (β) and shape (γ).

To select a suitable threshold (above which a GPD is fitted to the data), determine the lowest threshold above which the mean excess function, $e(u)=E\left(X-u \middle| X>u\right)$, is linear in u. The slope of this line is related to the shape parameter γ. Typically the chosen threshold is likely to be around the 90-95th percentile of the complete distribution. Above this threshold, a GPD can be fitted to the selected data by using MLE or the method of moments to determine the parameters.

There is a trade-off between the quality of approximation to the GPD (good for high thresholds) and level of bias in the fit (lower for lower thresholds).

Asymptotic property

When the *maxima* of a distribution converge to a GEV distribution (which is the case for all commonly used statistical distributions), the *excess* distribution converges to a GPD distribution with an equivalent shape parameter γ.

Module 20 Practice Questions

20.1 (i) Explain what is meant by an extreme event and give two examples.

(ii) Explain why it is important, in an enterprise risk management context, to consider extreme events separately from other types.

20.2 (i) Describe the generalised extreme value (GEV) distribution, given that the key formula is:

$$H(x) = \lim_{n\to\infty} P(\frac{X_M - \alpha_n}{\beta_n} \le x) = \exp\left(-\left(1+\frac{\gamma(x-\alpha)}{\beta}\right)^{\frac{-1}{\gamma}}\right) \qquad \gamma \ne 0$$

(ii) Describe how the GEV distribution can be used to model extreme events in an ERM context.

(iii) Outline an alternative approach that can be used in place of the GEV methodology. (You are not required to give formulae.)

20.3 (i) Define the mean excess function, $e(u)$.

(ii) Derive a formula for the mean excess loss function when the underlying loss distribution is exponential with mean $1/\lambda$.

(iii) Describe the overall shape that the mean excess loss function would exhibit for each of the following underlying loss distributions:

(a) exponential

(b) normal

(c) uniform.

20.4 The claim amounts in a general insurance portfolio are independent and follow an exponential distribution with mean £1,250.

Exam style

(i) Calculate the probability that an individual claim will exceed £5,000. [2]

(ii) Calculate the probability that, in a sample of 100 claims, at least one claim will exceed £5,000 using:

(a) an exact method

(b) an approximation based on the Gumbel (extreme value) distribution. [5]

You are given that, for an exponential distribution with parameter λ*, the approximate distribution of the standardised* $\max(X_1,\ldots,X_n)$ *for large n is a Gumbel distribution with distribution function* $H(x) = \exp\left(-\exp\left(-(x-\mu_n)/\sigma\right)\right)$ *where* $\mu_n = \frac{1}{\lambda}\ln(n)$ *and* $\sigma = \frac{1}{\lambda}$.

(iii) Describe the extent to which the two key assumptions made in the calculations in (ii) are likely to be borne out in practice. [3]

[Total 10]

The solutions start on the next page so that you can separate the questions and solutions.

Module 20 Solutions

20.1 (i) ***What is an extreme event?***

Extreme events are outcomes that have a low probability of occurrence but involve very large sums of money.

These are high severity events that occur in the right-hand tail of the loss distribution.

In an insurance context, they may arise as a result of a single cause that has a high financial cost (*eg* a personal injury claim or complete destruction of a building), or as an accumulation of events with a related cause (*eg* flood damage to a large number of houses in one town).

Extreme events can occur in contexts other than insurance claims, *eg* financial losses caused by a stock market crash or default of a company or the 'credit crunch'.

(ii) ***Why are extreme events considered separately?***

The majority of risk events fall within the main body of the fitted loss distribution and can usually be modelled accurately by one of the standard statistical distributions.

However, losses caused by extreme events fall in the right-hand tail of the distribution and often arise through a different cause than the smaller losses (*eg* many small operational losses are caused by human error; extreme operational loss events often have other causes *eg* terrorism).

As a result, the extreme events are usually considered to come from a different statistical distribution, *ie* they are outliers to the main loss distribution.

The extreme events usually involve the highest monetary amounts and therefore it is particularly important to assess them correctly.

Models of the extreme events are important for assessing the impact on a company of unexpected conditions (*eg* the 9/11 terrorist attacks and their impact on the world's financial markets) and can help with disaster recovery; business continuity planning *etc*.

There is usually a lack of past data on extreme events and so a different approach to modelling needs to be taken, *eg* dreaming up 'doomsday' type scenarios.

20.2 (i) ***GEV distribution***

The standardised maximum value (X_M) in a sample of n *iid* random variables ($X_1, X_2, ..., X_n$) tends to a particular distribution as the sample size increases. This is called the generalised extreme value (GEV) distribution.

The key parameter is the shape parameter γ .

When $\gamma > 0$, we have the Fréchet class of distributions, which is the limiting form for 'heavy tailed' underlying distributions with a finite lower bound, such as the Pareto distribution.

When $\gamma < 0$, we have the Weibull class of distributions, which is the limiting form for underlying distributions with a finite upper bound, such as the uniform distribution.

When $\gamma = 0$, we have the class of Gumbel distributions, which is the limiting form for most other underlying distributions that have finite moments, such as the normal and lognormal distributions.

In this case, the distribution function is found by taking the limit as $\gamma \to 0$, which gives:

$$P(X_M \leq x) = \exp\left(-\exp\left[-(x-\alpha)/\beta\right]\right)$$

The parameters α and β are location and scaling factors, respectively. (These will differ depending on the underlying distribution.)

(ii) ***Using the GEV distribution to model extreme events***

Extreme loss events correspond to the maximum values experienced over a period, so we might expect them to conform to a GEV distribution.

We can calculate the maximum loss event from past data by dividing it into blocks (*eg* one block for each year) and calculating the maximum within each block.

We can analyse a set of observed losses in two different ways:

1. divide the data into blocks (*eg* one block for each year) and calculating the maximum within each block (the return-level approach)
2. count the observations in each block that exceed some set level (the return-period approach).

We can then estimate the parameters for a GEV distribution using standard statistical methods, such as the method of moments or maximum likelihood estimation.

The fitted distributions can then be used to calculate percentiles, means and variances for the distribution of the required variable (*eg* maximum future claim amounts) and could also be used in simulations.

(iii) ***Alternative approach***

As an alternative to focusing upon a single maximum value, we can consider all the losses that exceed some threshold.

A similar theory to GEV predicts that, for large samples, these amounts should conform to a generalised Pareto distribution (GPD).

In order to fit the tail of a distribution we need to select a suitably high threshold and then fit the GPD to the values in excess of that threshold.

This method has the advantage that it uses a larger part of the data and models all the large claims above the threshold, not just the single highest value.

Various methods, such the method of moments and maximum likelihood estimation, have been devised for estimating the parameters of the distribution and for testing the results for goodness-of-fit.

20.3 (i) ***Mean excess loss function***

The mean excess function for a random variable X is defined by:

$$e(u) = E\left[X - u \mid X > u\right]$$

where u is the threshold amount.

Typically, X denotes the size of the loss event that is assumed to come from a particular loss distribution.

(ii) ***Formula***

If X has an exponential distribution with mean $1/\lambda$, then:

$$P(X > u) = \int_u^\infty \lambda e^{-\lambda x} dx = e^{-\lambda u}$$

The conditional probabilities above the threshold are then:

$$P(X > x + u \mid X > u) = \frac{P(X > x + u \ \& \ X > u)}{P(X > u)}$$

$$= \frac{P(X > x + u)}{P(X > u)}$$

$$= \frac{e^{-\lambda(x+u)}}{e^{-\lambda u}}$$

$$= e^{-\lambda x}$$

It follows that:

$$\Pr(X - u \le x \mid X > u) = 1 - e^{-\lambda x}$$

So the conditional distribution of the excess loss $X - u \le x \mid X > u$ is $Exp(\lambda)$, which has mean $1/\lambda$.

So in this case, the mean excess loss function has a constant value $e(u) = \dfrac{1}{\lambda}$.

(iii) ***Shape of the mean excess loss function***

(a) *Exponential*

We saw in part (ii), that when the underlying distribution is exponential, the mean excess loss function has a constant value. So its graph would be horizontal.

(b) *Normal*

Unlike the exponential distribution, which is positively skewed, the normal distribution is symmetrical and tails off more quickly than the exponential distribution on the right-hand side.

So if the underlying distribution is normal, the mean excess will also tail off (decrease) as the threshold increases (but always remaining positive).

(c) *Uniform*

The uniform distribution has a finite upper limit. Once the threshold reaches this point, the excess will always be zero. Before that point, the mean excess will decrease linearly.

20.4 (i) ***Probability that an individual claim will exceed £5,000***

The claims X follow an exponential distribution with mean $\frac{1}{\lambda}=1,250$.

Using the formula for the distribution function of an exponential random variable from the Tables, we find that:

$$P(X>5,000)=1-\mathrm{P}(X\le 5,000)=1-F_X(5,000)=e^{-5,000\lambda}=e^{-4}=0.018$$

So the required probability is 1.8%. [Total 2]

(ii) ***Probability that at least one claim will exceed £5,000***

(a) *Exact method*

We require the probability that $\max\left(X_1,\ldots,X_{100}\right)$ is greater than £5,000. We can calculate this exactly by considering the complementary event:

$$P\left[\max\left(X_1,\ldots,X_{100}\right)\le 5,000\right]=P\left[X_1\le 5,000,\ldots,X_{100}\le 5,000\right]$$

$$=P\left[X_1\le 5,000\right]\times\cdots\times P\left[X_{100}\le 5,000\right] \quad [1]$$

$$=\left(1-e^{-4}\right)^{100}=0.157$$

So the required (exact) probability is $1-0.157=84.3\%$. [1]

(b) *Approximate method*

As the approximate distribution of $\max\left(X_1,\ldots,X_{100}\right)$ is a Gumbel EV distribution with distribution function $H(x)=\exp\left(-\exp\left(-\left(x-\mu_n\right)/\sigma\right)\right)$, this means that:

$$\mathrm{P}\left[\max\left(X_1,\ldots,X_{100}\right)\le x\right]\approx\exp\left(-\exp\left(-\left(x-\mu_{100}\right)/\sigma\right)\right)$$

The required parameter values are:

$$\mu_{100}=1,250\ln 100=5,756.46 \quad \text{and} \quad \sigma=1,250 \quad [1]$$

So: $$\Pr\left[\max\left(X_1,\ldots,X_{100}\right)\le 5,000\right]\approx\exp\left(-\mathrm{e}^{-(5,000-5,756.46)/1,250}\right) \quad [1]$$

$$=\exp\left(-e^{0.6052}\right)=0.160$$

So the approximate probability is $1-0.160=84.0\%$. [1]

[Total 5]

(iii) *Assumptions*

The two key assumptions here are that the claims come from an exponential distribution with mean £1,250 and that they are statistically independent. [½]

In practice we cannot know the precise form of the true distribution. Both the type of distribution and the parameters must be estimated based on past claims data. [1]

These calculations concern the maximum claim amounts and it is often found that these follow a different distribution from claims in the main body of the data. [½]

The claims within a portfolio may not be independent. The severity of claims is likely to be affected by a number of underlying causes (*eg* the weather) that will affect the whole portfolio. However, some claims may be independent, *eg* claims from car accidents across a large geographically-diversified portfolio. [1]

[Total 3]

