Chapter 2

Classical models for extremes

Suppose that $X_1, X_2, \ldots, X_n$ is a sequence of independent and identically distributed (IID) random variables with common distribution function F.

One way of characterising extremes is by considering the distribution of the **maximum order statistic**

$$M_n = \max\{X_1, X_2, \ldots, X_n\}. \tag{2.1}$$

Think back to the example in Section 1.4 concerning sea surges at Wassaw Island.

Sea surge measurements were taken every hour; for each year (1955–2004) you were presented with the **annual maximum** sea surge.

Thus, here $n = 365 \times 24 = 8760$ (for non–leap years, anyway), and we might use the notation:

$$M_{8760,i}, \qquad i = 1, \ldots, 50,$$

to denote, generally, the 50 annual maxima given in Table 1.1 (of course, for leap years the notation would be $M_{8784,i}$).

The assumption of IID might be reasonable here.

For example, each annual maximum is likely to occur during the hurricane season (which is usually at its peak in September/October), and so it seems likely that maximum hourly observations from one year to the next will be far enough apart to be independent.

Issues of non–stationarity, however, might arise in long–range datasets owing to the effects of climate change, for example, and we will come back to this in Chapter 4.

Assuming, for now, that our maxima are IID, how can we obtain the distribution of M_n?

This is trivial (in principle), since

$$\begin{aligned}\Pr\{M_n \le x\} &= \Pr\{X_1 \le x, X_2 \le x, \ldots, X_n \le x\}\\ &= \Pr\{X_1 \le x\} \times \Pr\{X_2 \le x\} \times \cdots \times \Pr\{X_n \le x\}\\ &= \{F(x)\}^n.\end{aligned}$$

However, in practice the distribution function F is unknown. This leads to an approach based on asymptotic arguments – specifically, we look for limiting distributions for $\{F(x)\}^n$ as $n \to \infty$ — this is where the field of **Extreme Value Theory** was born.

One of the earliest books on the statistics of extreme values is **E.J. Gumbel** (1958, see Figure 2.1).

Gumbel traces the origins back to 1709, when **N. Bernoulli** considers the problem of estimating the age of the longest survivor in a group of people.

Research into extreme values as a subject in it's own right began much later, between 1920 and 1940, when work by **E.L. Dodd**, **M. Fréchet**, **E.J. Gumbel**, **R. von Mises** and **L.H.C. Tippett** investigated the asymptotic distribution of the *largest order statistic*.

This led to the main theoretical result: the **Extremal Types Theorem** (see Section 2.1.2), which was developed in stages by Fisher, Tippett and von Mises, and eventually proved in general by **B. Gnedenko** in 1943.

Until 1950, development was largely theoretical.

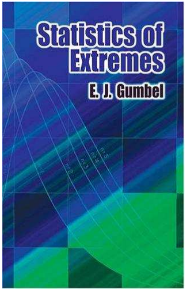
In 1958, Gumbel started applying theory to problems in engineering.

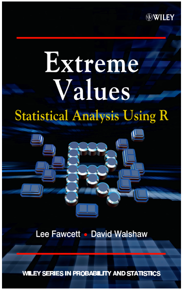
In the 1970s, **L. de Haan** and **J. Pickands** generalised the theoretical results, giving a better basis for statistical models.

Since the 1980s, methods for the application of Extreme Value Theory have become much more widespread.

Current researchers who have played a significant role in developing applications and methodology include **Richard Smith**, **Anthony Davison**, **Jonathan Tawn** and **Stuart Coles**.

There are still gaps between the theory and the models, and also between the models and common practice in applications – this is where our work fits in (**Fawcett** and **Walshaw**).





The obvious questions now are:

- What possible distributions might be considered candidates for the distribution for $M_n = \{F(x)\}^n$ as $n \to \infty$?
- Can we formulate this set of candidate distributions into a single class – say G – which is independent of F?
- Can we estimate the distribution of M_n using G, without any reference to F?

Clearly, the limiting distribution of M_n is **degenerate**:

- The distribution converges to a single point on the real line with probability 1
- In this case, this single point is the **upper endpoint** of F
- In some applications, this will be ∞

This is analogous to the sample mean $\bar{X}$ converging to the population mean μ with certainty in the **Central Limit Theorem**.

Here, the degenerate limit is prevented by allowing a linear rescaling, so that

$$\frac{\bar{X} - b_n}{a_n} \xrightarrow{D} N(0, 1)$$

where $b_n = \mu$ and $a_n = \sigma/\sqrt{n}$, where σ and n are the population standard deviation and sample size, respectively.

Can we apply a similar linear rescaling to M_n to avoid convergence of the distribution to a single point?

The answer, of course, is “yes”, and is provided by the main result in classical extreme value theory – the **Extremal Types Theorem** – a result for the maximum M_n which is analogous to the Central Limit Theorem for the mean μ.

Theorem *(The Extremal Types Theorem)*

If there exist sequences of constants $a_n > 0$ and b_n such that, as $n \to \infty$,

$$\Pr\left\{(M_n - b_n)/a_n \leq x\right\} \to G(x) \tag{2.2}$$

for some non–degenerate distribution G, then G is of the same type as one of the following distributions:

$$I : G(x) = \exp\left\{-\exp(-x)\right\} \qquad -\infty < x < \infty; \tag{2.3}$$

$$II : G(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0, \alpha > 0; \end{cases} \tag{2.4}$$

$$III : G(x) = \begin{cases} \exp\left\{-(-x)^{\alpha}\right\} & x < 0, \alpha > 0 \\ 1 & x \geq 0. \end{cases} \tag{2.5}$$

2.1.2 The Extremal Types Theorem

The three types of distribution – *I*, *II* and *III* – have become known as the **Gumbel**, **Fréchet** and **Weibull** types (respectively), and are known collectively as the **extreme value distributions**.

For both the Gumbel and Fréchet distributions the limiting distribution G is **unbounded**; that is, the upper–endpoint tends to ∞. Of the two, the Fréchet distribution gives **heavier tails**.

For the Weibull distribution, the limiting distribution is **bounded**.

It should be noted that this Theorem does not *ensure* the existence of a non–degenerate limit for M_n.

Nor does it specify *which* of types *I*, *II* or *III* is applicable *if* a limit distribution *does* exist (i.e. in which **domain of attraction** the distribution of G lies).

However, when such a distribution does exist, we find that, by analogy with the Central Limit Theorem, the limiting distribution of sample maxima follows one of the distributions given by the Extremal Types Theorem, *no matter what the parent distribution F.*

So we know that

$$\frac{M_n - b_n}{a_n} \xrightarrow{D} G,$$

where – if it exists – G is given by one of the extreme value distributions.

But how do we know which one of these distributions to use?

2.1.3 The Generalised Extreme Value distribution

In practice, working with — and having to choose between — three distributions is inconvenient.

However, there exists a parameterisation which encompasses all three types of extreme value distribution.

2.1.3 The Generalised Extreme Value distribution

Von Mises (1954) and **Jenkinson (1955)** independently derived the **generalised extreme value distribution** (GEV), often denoted $\mathcal{G}(\mu, \sigma, \xi)$, with CDF:

$$G(x; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_{+}^{-1/\xi}\right\}, \qquad (2.6)$$

where $a_{+} = \max(0, a)$. The situation where $\xi = 0$ is not defined in (2.6), but is taken as the limit as $\xi \to 0$, given by

$$G(x; \mu, \sigma) = \exp\left\{-\exp\left(-\left[\frac{x-\mu}{\sigma}\right]\right)\right\}. \qquad (2.7)$$

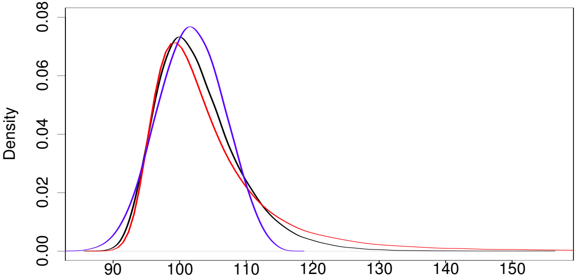
2.1.3 The Generalised Extreme Value distribution

The parameters μ $(-\infty < \mu < \infty)$, σ (> 0) and ξ $(-\infty < \xi < \infty)$ are **location**, **scale** and **shape** parameters, respectively.

Shape parameter ξ:

- $\xi = 0$: Gumbel (type I extreme value) distribution
- $\xi > 0$: Fréchet (type II extreme value) distribution
- $\xi < 0$: Weibull (type III extreme value) distribution
- Through inference for ξ, the data themselves determine the most appropriate type of tail behaviour – no need for any *a priori* judgements
- The standard error for ξ accounts for our uncertainty in choosing between the 3 EV distributions

2.1.3 The Generalised Extreme Value distribution



But what about the **constants** a_n and b_n?

We know that

$$\frac{M_n - b_n}{a_n} \xrightarrow{D} \mathcal{G}(\mu, \sigma, \xi), \qquad \text{as } n \to \infty.$$

After some algebra, it turns out that

$$M_n \xrightarrow{D} \mathcal{G}(\mu^*, \sigma^*, \xi), \qquad \text{approximately, as } n \to \infty,$$

with a_n and b_n being absorbed into μ^* and σ^*.

Since the GEV parameters need to be estimated anyway, in practice we just ignore the normalisation constants and fit the GEV directly to our set of maxima $M_{n,i}$.

However, before we consider applications of the GEV to real data, let us first consider some theoretical examples which demonstrate that, with careful choices of a_n and b_n, one of the three extreme value distributions is always achieved when the parent distribution F is known.

Example 2.1

Suppose $X_1, X_2, \ldots, X_n$ is a sequence of independent $Exp(1)$ variables, that is

$$F(x) = 1 - e^{-x}, \qquad x > 0.$$

By letting $a_n = 1$ and $b_n = \log n$, show that the limit distribution of $(M_n - b_n)/a_n$ is of extreme value type, and identify the distribution.

Example 2.1: Solution (1/2)

We want the distribution of $(M_n - b_n)/a_n$, i.e.

$$\begin{aligned}
\Pr\left\{\frac{M_n - b_n}{a_n} \le z\right\} &= \Pr\left\{\frac{\text{Max}(X_1, \ldots, X_n) - b_n}{a_n} \le z\right\} \\
&= \Pr\left\{\frac{X_1 - b_n}{a_n} \le z, \ldots, \frac{X_n - b_n}{a_n} \le z\right\} \\
&= \Pr\{X_1 \le a_n z + b_n\} \times \cdots \\
&= \left[1 - e^{-(a_n z + b_n)}\right]^n,
\end{aligned}$$

as $X_1, \ldots, X_n$ are I.I.D.

Example 2.1: Solution (2/2)

Now using $a_n = 1$ and $b_n = \log n$, we get

$$\begin{aligned} \left[1 - e^{-(z+\log n)}\right]^n &= \left[1 - e^{-z} e^{\log n^{-1}}\right]^n \\ &= \left[1 - n^{-1} e^{-z}\right]^n. \end{aligned}$$

From Stage 1 methods courses, you should know that

$$\exp(y) = \lim_{n\to\infty} \left(1 + \frac{y}{n}\right)^n.$$

Thus, we have

$$\left[1 + \frac{-e^{-z}}{n}\right]^n \longrightarrow \exp\left(-e^{-z}\right), \qquad \text{as } n \to \infty.$$

This is the **Gumbel** (type I extreme value) distribution.

Example 2.2

Suppose $X_1, X_2, \dots, X_n$ is a sequence of independent *Fréchet*(1) variables, that is

$$F(x) = e^{-1/x}, \qquad x > 0.$$

By letting $a_n = n$ and $b_n = 0$, show that the limit distribution of $(M_n - b_n)/a_n$ is of extreme value type, and identify the distribution.

Example 2.2: Solution (1/1)

Following the solution to Example 2.1, we find:

$$\begin{aligned}\Pr\left\{\frac{M_n - b_n}{a_n} \leq z\right\} &= \Pr\{X_1 \leq a_n z + b_n\} \times \cdots \\ &= \left[\exp\left\{-\frac{1}{a_n z + b_n}\right\}\right]^n.\end{aligned}$$

Letting $a_n = n$ and $b_n = 0$ gives

$$\left[\exp\left\{-\frac{1}{nz}\right\}\right]^n = e^{-1/z},$$

which is the **Fréchet** (type II extreme value) distribution with $\alpha = 1$, i.e. a unit Fréchet distribution.

2.1.4 Typical application – 1: Data pre–processing

- Choose your **block length** n (usually the number of observations in a calendar year)
- Discard all but the largest observation within each block
- n too small: limiting arguments will not hold
- n too large: not enough maxima to work with!

Fit the GEV to your set of block maxima $M_{n,i}$ – numerical maximum likelihood estimation is the most common approach here.

Assuming independence, we form the likelihood in the usual way:

$$L(\mu, \sigma, \xi; \boldsymbol{x}) = \prod_{i=1}^{m} g(x_i; \mu, \sigma, \xi),$$

where g is the GEV probability density function and can be found, after differentiation of the distribution function (2.6), to be

$$\frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_{+}^{-(1/\xi+1)} \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_{+}^{-1/\xi}\right\}. \tag{2.8}$$

(try this yourself!)

Use the GEV probability density function in (2.8) to form the likelihood function $L(\mu, \sigma, \xi; \boldsymbol{x})$. Also, obtain the GEV log–likelihood function $\ell(\mu, \sigma, \xi; \boldsymbol{x})$ and the corresponding (log) likelihood equations.

The likelihood is given by

$$L(\mu, \sigma, \xi; \boldsymbol{x}) = \prod_{i=1}^{m} \frac{1}{\sigma} \left[1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right]_{+}^{-(1/\xi+1)} \times \exp\left\{-\left[1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right]_{+}^{-1/\xi}\right\}$$

The log–likelihood is given by

$$\begin{aligned}
\ell(\mu,\sigma,\xi;\boldsymbol{x}) &= \sum_{i=1}^{m}\log\sigma^{-1} + \sum_{i=1}^{m}\log\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]_+^{-(1/\xi+1)} \\
&\quad -\sum_{i=1}^{m}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]_+^{-1/\xi} \\
&= -m\log\sigma - (1/\xi+1)\sum_{i=1}^{m}\log\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]_+ \\
&\quad -\sum_{i=1}^{m}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]_+^{-1/\xi}.
\end{aligned}$$

2.1.4 Typical application – 2: Parameter estimation

The (log)–likelihood equations are $\frac{\partial \ell}{\partial \mu} = \frac{\partial \ell}{\partial \sigma} = \frac{\partial \ell}{\partial \xi} = 0$. For example:

$$\begin{aligned}\frac{\partial \ell}{\partial \mu} &= \frac{\xi + 1}{\sigma} \sum_{i=1}^{m} \left[1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right]^{-1} \\ &\quad + \frac{1}{\sigma} \sum_{i=1}^{m} \left[1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right]^{-(1/\xi + 1)} = 0\end{aligned}$$

The other two (log)–likelihood equations would be found in exactly the same way.

How would you use the (log) likelihood equations in order to obtain maximum likelihood estimates of μ, σ and ξ?

Why can't we obtain closed form solutions for $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$?

How can we get around this?

2.1.4 Typical application – 2: Parameter estimation

The next step would be to replace μ, σ and ξ with their corresponding estimators $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$ (respectively), and then solve for $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$.

In fact, for the GEV there are no closed–form solutions for $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$ – the (log) likelihood equations cannot be solved analytically.

We can get around this by adopting a numerical method to obtain (approximate) solutions to the (log) likelihood equations – R uses a **Newton–Raphson type algorithm**.

As with all statistical models, there are various **goodness–of–fit properties** that should be considered to check the overall adequacy of the fitted GEV.

These include **probability plots**, **quantile–quantile** (*Q*-*Q* plots) and simply plotting a histogram of the data against the fitted density.

Again, these will be reviewed shortly via a real–life data demonstration in R.

2.1.4 Typical application – 4: Return level estimation

As discussed in Section 1.3, interest usually lies not in estimates of the GEV parameters themselves, but in how we can use the fitted model to estimate other quantities – such as

- The height of a sea wall to protect against the **once in a hundred year sea–surge**;
- The "**fifty year wind speed**" to provide new structures enough protection against wind damage.

Such quantities, in extreme value terminology, are usually referred to as **return levels**.

If we have faith in our fitted model being suitable beyond the range of our observed data, we can estimate the r–year return level z_r for any period by setting the GEV distribution function equal to $1 - 1/r$ and solving for $x = \hat{z}_r$ (provided we have annual maxima).

For example, suppose we fit the GEV to the set of annual maxima given in Table 1.1 and obtain estimates of the location, scale and shape as $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$ (respectively).

Suppose further that the authorities require an estimate of z_{100}, the sea surge we might expect to be exceeded once in a hundred years.

Then we can write down the following probability statement:

$$\text{Pr(annual maximum} > z_{100}) \quad = \quad \frac{1}{100},$$

i.e.

$$1 - \text{Pr(annual maximum} \leq z_{100}) = \frac{1}{100}. \tag{2.9}$$

Now the left–hand–side of (2.9), in terms of our fitted GEV, is

$$1 - G(\hat{z}_{100}; \hat{\mu}, \hat{\sigma}, \hat{\xi}),$$

giving

$$\begin{aligned} 1 - \exp\left\{ - \left[1 + \hat{\xi} \left(\frac{\hat{z}_{100} - \hat{\mu}}{\hat{\sigma}} \right) \right]^{-1/\hat{\xi}} \right\} &= 0.01 \qquad \text{i.e.} \\ \exp\left\{ - \left[1 + \hat{\xi} \left(\frac{\hat{z}_{100} - \hat{\mu}}{\hat{\sigma}} \right) \right]^{-1/\hat{\xi}} \right\} &= 0.99. \end{aligned}$$

Solving for $\hat{z}_{100}$ gives an estimate of the **100–year return level** as

$$\hat{z}_{100} = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\left[(-\log(0.99))^{-\hat{\xi}} - 1\right];$$

more generally, estimates of the r–year return level z_r are given by

$$\hat{z}_r = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\left[\left(-\log\left(1 - r^{-1}\right)\right)^{-\hat{\xi}} - 1\right]. \qquad (2.10)$$

2.1.4 Typical application – 4: Return level estimation

Care needs to be taken when $\xi = 0$; in this case, as with forming the likelihood in the first place, we should work with the limiting form (as $\xi \to 0$) of the distribution function (i.e. Equation (2.7)).

Of course, a point estimate of the return level alone is not enough; any estimate requires a corresponding standard error.

To obtain standard errors for return levels, we can use the **delta method**.

However, as we shall see later, we do not recommend using these standard errors to form confidence intervals for z_r.

2.2 Sea surge at Wassaw Island

Recall the example from Section 1.4.

- The southeastern coast of the USA is an area often hit by hurricanes
- The historic city of Savannah, Georgia, has suffered direct hits from 22 Hurricanes since 1871
- The city can expect to ‘brushed’ or directly hit once every other year
- Predictions show that the city is three years overdue it’s next direct hit

8.5	8.9	9.1	8.9	8.4	9.7	9.1	9.6	8.7	9.3
9.6	9.3	8.7	9.0	8.8	8.9	8.9	12.2	7.8	7.7
8.3	8.1	7.3	6.8	6.7	7.3	7.6	8.2	8.6	9.8
9.5	7.4	7.3	10.2	10.3	10.4	8.8	9.7	10.0	10.8
11.1	12.7	11.5	11.8	12.6	13.0	10.5	10.5	10.0	9.4

Recall that, in Chapter 1, we tried to estimated quantities beyond the range of our observed data.

For example, using simple empirical arguments,

$$\text{Pr(Sea surge exceeds 14 feet)} = 0,$$

simply because, over the time-frame we have data for, we have not observed such an extreme event.

Does this really mean this event is **impossible**?

None necessary — we are given a set of annual maxima (in the first computer practical session, we will consider how to pre–process a dataset to obtain the set of block maxima).

Figure 2.2 shows a time series plot and histogram of the 50 annual maxima:

- No obvious trend in our dataset
- Issue of dependence?

In R:

```
> wassaw=scan('wassaw.txt')
```

Or:

```
> wassaw=c(8.5, 8.9, ..., 9.4)
```

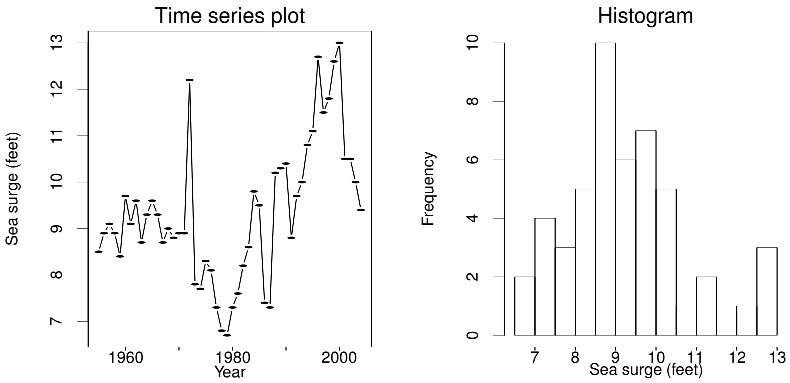
In R:

```
> year=seq(1955,2004,1)
```
Sequence of values from 1995 to 2004

Then, to produce the plots shown in Figure 2.2:

```
> par(mfrow=c(1,2))
```
Partitions the plotting space

```
> plot(wassaw~year,type='b')
> hist(wassaw)
```



Time series plot
Sea surge (feet)
7
8
9
10
11
12
13
1960
1980
2000
Year
Histogram
Frequency
0
2
4
6
8
10
7
8
9
10
11
12
13
Sea surge (feet)

We now use R to maximise

$$\ell(\mu,\sigma,\xi;\boldsymbol{x}) = -50\log\sigma - (1+1/\xi)\sum_{i=1}^{50}\log\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]_+ - \sum_{i=1}^{50}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]_+^{-1/\xi}$$

with respect to μ, σ and ξ.

In R:

```
> theta=c(mean(wassaw), sd(wassaw), 0.1)
```

Sets up parameter vector $\theta = (\mu, \sigma, \xi)$

2.2.2 Fitting the GEV: Parameter estimation

```
> gev.loglik=function(theta){
   mu=theta[1]
   sigma=theta[2]
   xi=theta[3]
   m=min((1+(xi*(dataset-mu)/sigma)))
   if(m<0.00001)return(as.double(1000000))
   if(sigma<0.00001)return(as.double(1000000))
   if(xi==0){
    loglik=-length(dataset)*log(sigma)
+     -sum((dataset-mu)/sigma)
+     -sum(exp(-((dataset-mu)/sigma)))}
  else{
    loglik=-length(dataset)*log(sigma)
+     -(1/xi+1)*sum(log(1+(xi*(dataset-mu)/sigma)))
+     -sum((1+(xi*(dataset-mu)/sigma))**(-1/xi))}
     return(-loglik)}
```

2.2.2 Fitting the GEV: Parameter estimation

```
> gev.loglik=function(theta){
  mu=theta[1]
  sigma=theta[2]
  xi=theta[3]
  m=min((1+(xi*(dataset-mu)/sigma)))
  if(m<0.00001)return(as.double(1000000))
  if(sigma<0.00001)return(as.double(1000000))
  if(xi==0){
   loglik=-length(dataset)*log(sigma)
+    -sum((dataset-mu)/sigma)
+    -sum(exp(-((dataset-mu)/sigma)))}
  else{
   loglik=-length(dataset)*log(sigma)
+    -(1/xi+1)*sum(log(1+(xi*(dataset-mu)/sigma)))
+    -sum((1+(xi*(dataset-mu)/sigma))**(-1/xi))}
    return(-loglik)}
```

2.2.2 Fitting the GEV: Parameter estimation

```
> gev.loglik=function(theta){
  mu=theta[1]
  sigma=theta[2]
  xi=theta[3]
  m=min((1+(xi*(dataset-mu)/sigma)))
  if(m<0.00001)return(as.double(1000000))
  if(sigma<0.00001)return(as.double(1000000))
  if(xi==0){
   loglik=-length(dataset)*log(sigma)
+     -sum((dataset-mu)/sigma)
+     -sum(exp(-((dataset-mu)/sigma)))}
  else{
   loglik=-length(dataset)*log(sigma)
+     -(1/xi+1)*sum(log(1+(xi*(dataset-mu)/sigma)))
+     -sum((1+(xi*(dataset-mu)/sigma))**(-1/xi))}
     return(-loglik)}
```

```
> gev.loglik=function(theta){
  mu=theta[1]
  sigma=theta[2]
  xi=theta[3]
  m=min((1+(xi*(dataset-mu)/sigma)))
  if(m<0.00001)return(as.double(1000000))
  if(sigma<0.00001)return(as.double(1000000))
  if(xi==0){
   loglik=-length(dataset)*log(sigma)
+     -sum((dataset-mu)/sigma)
+     -sum(exp(-((dataset-mu)/sigma)))}
  else{
   loglik=-length(dataset)*log(sigma)
+     -(1/xi+1)*sum(log(1+(xi*(dataset-mu)/sigma)))
+     -sum((1+(xi*(dataset-mu)/sigma))**(-1/xi))}
    return(-loglik)}
```

2.2.2 Fitting the GEV: Parameter estimation

```
> gev.loglik=function(theta){
  mu=theta[1]
  sigma=theta[2]
  xi=theta[3]
  m=min((1+(xi*(dataset-mu)/sigma)))
  if(m<0.00001)return(as.double(1000000))
  if(sigma<0.00001)return(as.double(1000000))
  if(xi==0){
   loglik=-length(dataset)*log(sigma)
+      -sum((dataset-mu)/sigma)
+      -sum(exp(-((dataset-mu)/sigma)))}
 else{
   loglik=-length(dataset)*log(sigma)
+     -(1/xi+1)*sum(log(1+(xi*(dataset-mu)/sigma)))
+     -sum((1+(xi*(dataset-mu)/sigma))**(-1/xi))}
    return(-loglik)}
```

```
> gev.loglik=function(theta){
  mu=theta[1]
  sigma=theta[2]
  xi=theta[3]
  m=min((1+(xi*(dataset-mu)/sigma)))
  if(m<0.00001)return(as.double(1000000))
  if(sigma<0.00001)return(as.double(1000000))
  if(xi==0){
   loglik=-length(dataset)*log(sigma)
+     -sum((dataset-mu)/sigma)
+     -sum(exp(-((dataset-mu)/sigma)))}
  else{
   loglik=-length(dataset)*log(sigma)
+     -(1/xi+1)*sum(log(1+(xi*(dataset-mu)/sigma)))
+     -sum((1+(xi*(dataset-mu)/sigma))**(-1/xi))}
    return(-loglik)}
```

```
> gev.loglik=function(theta){
   mu=theta[1]
   sigma=theta[2]
   xi=theta[3]
   m=min((1+(xi*(dataset-mu)/sigma)))
   if(m<0.00001)return(as.double(1000000))
   if(sigma<0.00001)return(as.double(1000000))
   if(xi==0){
    loglik=-length(dataset)*log(sigma)
+      -sum((dataset-mu)/sigma)
+      -sum(exp(-((dataset-mu)/sigma)))}
  else{
    loglik=-length(dataset)*log(sigma)
+      -(1/xi+1)*sum(log(1+(xi*(dataset-mu)/sigma)))
+      -sum((1+(xi*(dataset-mu)/sigma))**(-1/xi))}
     return(-loglik)}
```

In R:

```
> dataset=wassaw
```

Attaches the sea–surge extremes to “dataset”

```
> nlm(gev.loglik,theta)
```

Minimisation routine

```
$minimum
[1] 89.52412

$estimate
[1] 8.7112735 1.3114836 -0.1084451

$gradient
[1] 3.350727e-06 2.316675e-05 2.145839e-06

$code
[1] 1
```

2.2.2 Fitting the GEV: Parameter estimation

In R:

```
> dataset=wassaw
```

Attaches the sea–surge extremes to "dataset"

```
> nlm(gev.loglik,theta)
```

Minimisation routine

```
$minimum
[1] 89.52412

$estimate
[1] 8.7112735 1.3114836 -0.1084451

$gradient
[1] 3.350727e-06 2.316675e-05 2.145839e-06

$code
[1] 1
```

In R:

```
> A=nlm(gev.loglik,theta,hessian=TRUE)
      # Stores output, including Hessian, in A
```

Now:

```
> A
    ⋮
$hessian
            [,1]      [,2]      [,3]
[1,]   27.375570 -4.837481  18.06140
[2,]   -4.837481 56.151740  29.89004
[3,]   18.061397 29.890039 118.18086
```

2.2.2 Fitting the GEV: Standard errors

The **Hessian** is the matrix of second–order partial derivatives.

In the case of our log–likelihood function, this is

$$\text{Hessian} = \begin{pmatrix} \frac{\partial^2 \ell}{\partial \mu^2} & & \\ \frac{\partial^2 \ell}{\partial \sigma \partial \mu} & \frac{\partial^2 \ell}{\partial \sigma^2} & \\ \frac{\partial^2 \ell}{\partial \xi \partial \mu} & \frac{\partial^2 \ell}{\partial \xi \partial \sigma} & \frac{\partial^2 \ell}{\partial \xi^2} \end{pmatrix} = -I_{\text{obs}},$$

evaluated at the MLEs.

However, since `gev.loglik` returns the *negative* log–likelihood, in this example

$$\text{Hessian} = I_{\text{obs}}!!$$

2.2.2 Fitting the GEV: Standard errors

Recall from MAS2302 that inversion of the Information matrix gives the **Variance–Covariance matrix**.

Thus, we have a 3×3 matrix to invert!

You should be able to do this (e.g. assignments), but let's use the `solve` command in **R**.

2.2.2 Fitting the GEV: Standard errors

In R, `solve` will solve the system of equations:

$$\boldsymbol{a}X = \boldsymbol{b}.$$

In our example, $\boldsymbol{a} = I_{\text{obs}}$. Omitting $\boldsymbol{b}$ in the execution of `solve` assumes the identity matrix for $\boldsymbol{b}$, giving:

$$I_{\text{obs}} \cdot X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Thus, solving for X will return the inverse of I_{obs} – the **variance–covariance matrix**!

In R:

```
> solve(A$hessian)
             [,1]         [,2]         [,3]
[1,]  0.043869792  0.008491490 -0.008852201
[2,]  0.008491490  0.022223135 -0.006918367
[3,] -0.008852201 -0.006918367  0.011564254

> varcovar=solve(A$hessian)

> sqrt(diag(varcovar))

[1] 0.2094512 0.1490743 0.1075372
```

2.2.2 Fitting the GEV: Standard errors

Thus, we now have the following inference for our annual maximum sea surges, in terms of the GEV distribution:

$$\hat{\mu} = 8.711(0.209) \qquad \hat{\sigma} = 1.311(0.149) \qquad \hat{\xi} = -0.108(0.108)$$

From this, we can construct **confidence intervals** in the usual way ("Wald" intervals).

For example: parameter estimate $\pm$ 1.96 $\times$ standard error for a 95% CI, owing to the normality of maximum likelihood estimators — giving

$$(8.301, 9.121) \qquad (1.019, 1.603) \qquad \text{and} \qquad (-0.320, 0.104)$$

for μ, σ and ξ (respectively).

Note that the confidence interval for ξ **includes zero** — a Gumbel–type tail for our data could be appropriate.

Probability plots compare *empirical cumulative probabilities* with corresponding values from the fitted model.

For example, the ordered sea–surges at Wassaw, $x_{(i)}$, $i = 1, \ldots, 50$, are:

6.7	6.8	7.3	7.3	7.3	7.4	7.6	7.7	7.8	8.1
8.2	8.3	8.4	8.5	8.6	8.7	8.7	8.8	8.8	8.9
8.9	8.9	8.9	9.0	9.1	9.1	9.3	9.3	9.4	9.5
9.6	9.6	9.7	9.7	9.8	10.0	10.0	10.2	10.3	10.4
10.5	10.5	10.8	11.1	11.5	11.8	12.2	12.6	12.7	13.0

$x_{(i)}$	Fitted: $G(x_{(i)}; \hat{\mu}, \hat{\sigma}, \hat{\xi})$	Empirical: $i/(n+1)$
6.7	0.016	0.020
6.8	0.021	0.039
7.3	0.063	0.059
⋮	⋮	⋮

2.2.3 Model adequacy: Probability plots

Probability plots compare *empirical cumulative probabilities* with corresponding values from the fitted model.

For example, the ordered sea–surges at Wassaw, $x_{(i)}$, $i = 1, \ldots, 50$, are:

6.7	6.8	7.3	7.3	7.3	7.4	7.6	7.7	7.8	8.1
8.2	8.3	8.4	8.5	8.6	8.7	8.7	8.8	8.8	8.9
8.9	8.9	8.9	9.0	9.1	9.1	9.3	9.3	9.4	9.5
9.6	9.6	9.7	9.7	9.8	10.0	10.0	10.2	10.3	10.4
10.5	10.5	10.8	11.1	11.5	11.8	12.2	12.6	12.7	13.0

$x_{(i)}$	Fitted: $G(x_{(i)}; \hat{\mu}, \hat{\sigma}, \hat{\xi})$	Empirical: $i/(n+1)$
6.7	**0.016**	**0.020**
6.8	**0.021**	**0.039**
7.3	**0.063**	**0.059**
$\vdots$	$\vdots$	$\vdots$

Quantile plots compare empirical quantiles with corresponding values from the fitted model:

Empirical: $x_{(i)}$	Prob: $i/(n+1)$	Fitted: $G^{-1}(i/(n+1); \hat{\mu}, \hat{\sigma}, \hat{\xi})$
6.7	0.020	6.78
6.8	0.039	7.07
7.3	0.059	7.27
$\vdots$	$\vdots$	$\vdots$

Quantile plots compare empirical quantiles with corresponding values from the fitted model:

Empirical: $x_{(i)}$	Prob: $i/(n+1)$	Fitted: $G^{-1}(i/(n+1); \hat{\mu}, \hat{\sigma}, \hat{\xi})$
6.7	0.020	**6.78**
6.8	0.039	**7.07**
7.3	0.059	**7.27**
$\vdots$	$\vdots$	$\vdots$

2.2.3 Probability plot in R

In R:

```
>ordered=sort(dataset)
     # Orders the data

> empirical=vector('numeric',length(ordered))

> for(i in 1:length(empirical)){
  empirical[i]=i/(length(dataset)+1)}
```

The function `GEV.DF` defines the distribution function for the GEV:

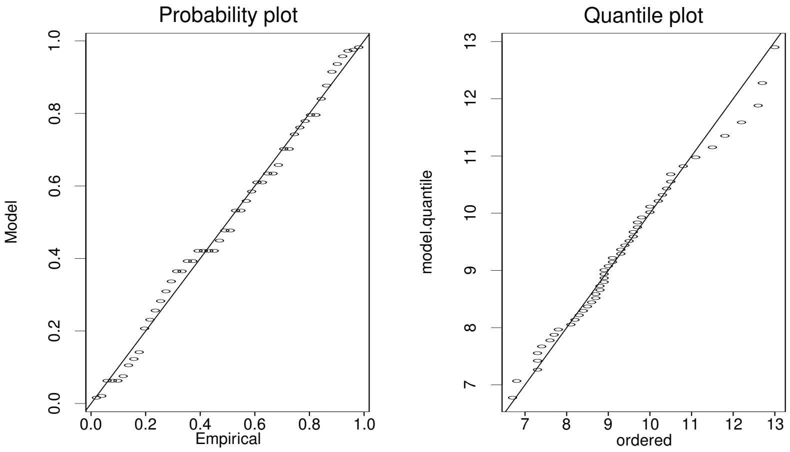
```
GEV.DF=function(data,mu,sigma,xi){
   if(xi==0){
   GEV=exp(-exp(-((data-mu)/sigma)))}
   else{
   GEV=exp(-(1+xi*((data-mu)/sigma))**(-1/xi))}
  return(GEV)}
```

Then:

```
> model=vector('numeric',length(dataset))
> for(i in 1:length(model)){
   model[i]=GEV.DF(ordered[i],A$est[1],A$est[2],A$est[3])}
```

Plotting `model` against `empirical` produces the corresponding probability plot.

2.2.3 Probability/Quantile plots in R



In question 2 of Section 1.4, you were asked to provide an estimate of the height of a new sea wall to protect the city of Savannah against the storm surge we would expect to see

(i) once in ten years;

(ii) once in a hundred years.

2.2.4 Return level estimation

Using the data alone we could not obtain a meaningful estimate of (ii) because we only have 50 years of data.

However, we can now use our fitted GEV to extrapolate beyond the range of our data to estimate such return levels.

Using Equation (2.10):

$$\hat{z}_r = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\left[\left(-\log\left(1 - r^{-1}\right)\right)^{-\hat{\xi}} - 1\right], \qquad (2.10)$$

we find that

$$\hat{z}_{10} = 8.711 - \frac{1.311}{0.108}\left[\left(-\log\left(1 - 10^{-1}\right)\right)^{0.108} - 1\right] = 11.33 \text{ feet;}$$

similarly,

$$\hat{z}_{10} = 8.711 - \frac{1.311}{0.108}\left[\left(-\log\left(1 - 100^{-1}\right)\right)^{0.108} - 1\right] = 13.46 \text{ feet.}$$

Thus, an estimate of the height of the sea–wall might be about 11.5 feet or 13.5 feet to protect against the once in ten year, or once in a hundred year, storm surges (respectively).

In fact, due to the **invariance property** of maximum likelihood estimators, our estimates of z_{10} and z_{100} are also the maximum likelihood estimators of these quantities.

As with our inference for the GEV parameters, it is preferable to quote estimates of return levels with their estimated standard errors.

Since z_r is a function of the GEV parameters μ, σ and ξ, we can use the **delta method** to find the approximate variance of $\hat{z}_r$.

Specifically,

$$\text{Var}(\hat{z}_r) \approx \nabla z_r^T V \nabla z_r,$$

where V is the variance–covariance matrix of $(\hat{\mu}, \hat{\sigma}, \hat{\xi})^T$ and

$$\begin{aligned} \nabla z_r^T &= \left[\frac{\partial z_r}{\partial \mu}, \frac{\partial z_r}{\partial \sigma}, \frac{\partial z_r}{\partial \xi}\right] \\ &= \left[1, -\xi^{-1}(1 - y_r^{-\xi}), \sigma\xi^{-2}(1 - y_r^{-\xi}) - \sigma\xi^{-1} y_r^{-\xi} \log y_r\right], \end{aligned}$$

where $y_r = -\log(1 - r^{-1})$, evaluated at $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$.

2.2.4 Return level standard errors in R

Recall that we previously stored V in the matrix `varcovar` in R.

Also, recall that our estimates of the GEV parameters are stored in `A$est`.

We can define ∇z_{10}, for example, in R as:

```
>y10=-log(1-(1/10))
>del=matrix(ncol=1,nrow=3)
>del[1,1]=1
>del[2,1]=-((A$est[3])**(-1))*(1-(y10**(-A$est[3])))
>del[3,1]=((A$est[2])*((A$est[3])**(-2))*(1-((y10)**(-A$est[3]))))
-((A$est[2])*((A$est[3])**(-1))*((y10)**(-(A$est[3])))*log(y10))
>del.transpose=t(del)
```

Recall that we previously stored V in the matrix `varcovar` in R.

Also, recall that our estimates of the GEV parameters are stored in `A$est`.

We can define ∇z_{10}, for example, in R as

```
>y10=-log(1-(1/10))
>del=matrix(ncol=1,nrow=3)
>del[1,1]=1
>del[2,1]=-((A$est[3])**(-1))*(1-(y10**(-A$est[3])))
>del[3,1]=((A$est[2])*((A$est[3])**(-2))*(1-((y10)**(-A$est[3]))))
-((A$est[2])*((A$est[3])**(-1))*((y10)**(-(A$est[3])))*log(y10))
>del.transpose=t(del)
```

2.2.4 Return level standard errors in R

Recall that we previously stored V in the matrix `varcovar` in R.

Also, recall that our estimates of the GEV parameters are stored in `A$est`.

We can define ∇z_{10}, for example, in R as

```
>y10=-log(1-(1/10))
>del=matrix(ncol=1,nrow=3)
>del[1,1]=1
>del[2,1]=-((A$est[3])**(-1))*(1-(y10**(-A$est[3])))
>del[3,1]=((A$est[2])*((A$est[3])**(-2))*(1-((y10)**(-A$est[3]))))
-((A$est[2])*((A$est[3])**(-1))*((y10)**(-(A$est[3])))*log(y10))

>del.transpose=t(del)
```

2.2.4 Return level standard errors in R

Recall that we previously stored V in the matrix `varcovar` in R.

Also, recall that our estimates of the GEV parameters are stored in `A$est`.

We can define ∇z_{10}, for example, in R as

```
>y10=-log(1-(1/10))
>del=matrix(ncol=1,nrow=3)
>del[1,1]=1
>del[2,1]=-((A$est[3])**(–1))*(1–(y10**(–A$est[3])))
>del[3,1]=((A$est[2])*((A$est[3])**(-2))*(1-((y10)**(-A$est[3]))))
-((A$est[2])*((A$est[3])**(-1))*((y10)**(-(A$est[3])))*log(y10))

>del.transpose=t(del)
```

2.2.4 Return level standard errors in R

Recall that we previously stored V in the matrix `varcovar` in R.

Also, recall that our estimates of the GEV parameters are stored in `A$est`.

We can define ∇z_{10}, for example, in R as

```
>y10=-log(1-(1/10))
>del=matrix(ncol=1,nrow=3)
>del[1,1]=1
>del[2,1]=-((A$est[3])**(-1))*(1-(y10**(-A$est[3])))
>del[3,1]=((A$est[2])*((A$est[3])**(-2))*(1-((y10)**(-A$est[3]))))
-((A$est[2])*((A$est[3])**(-1))*((y10)**(-(A$est[3])))*log(y10))
>del.transpose=t(del)
```

2.2.4 Return level standard errors in R

Recall that we previously stored V in the matrix `varcovar` in R.

Also, recall that our estimates of the GEV parameters are stored in `A$est`.

We can define ∇z_{10}, for example, in R as

```
>y10=-log(1-(1/10))
>del=matrix(ncol=1,nrow=3)
>del[1,1]=1
>del[2,1]=-((A$est[3])**(-1))*(1-(y10**(-A$est[3])))
>del[3,1]=((A$est[2])*((A$est[3])**(-2))*(1-((y10)**(-A$est[3]))))
-((A$est[2])*((A$est[3])**(-1))*((y10)**(-(A$est[3])))*log(y10))
```

>**del.transpose=t(del)**

Then the R command for matrix multiplication – `%*%` – can be used to obtain an estimate of the standard error for $\hat{z}_{10}$ in the following way:

```
> sqrt(del.transpose%*%varcovar%*%del)
          [,1]
 [1,]  0.3614568
```

Estimated standard errors for other return levels can be obtained in a similar way.

For example, for the standard error for the 100–year return level, we would replace

```
> y10=-log(1-(1/10))
```

with

```
> y100=-log(1-(1/100))
```

and then `y10` would be replaced with `y100` throughout.

Table 2.2 shows a range of estimated return levels, with associated standard errors in parentheses.

Of course, we could use these standard errors to construct confidence intervals for our return level estimates; however, as we shall shortly discuss, such confidence intervals **can be misleading**.

Return level: MLE (s.e.)			
z_{10}	z_{100}	z_{200}	z_{1000}
11.33 (0.361)	13.46 (0.938)	13.99 (1.182)	15.09 (1.821)

Class demonstration in R

Example 2.3: Rainfall in the Lake District

The data shown in Table 2.3 are annual maximum rainfall accumulations, obtained from daily records, for a period of 21 years (1991–2011 inclusive) at Eskdale in the Lake District.

333	213	790	343	351	521	307	305	352	277	319
319	339	262	285	297	327	620	350	545	258	

Example 2.3: Rainfall in the Lake District

Shown below are the maximum likelihood estimates of the GEV parameters, obtained using R; also shown is the observed information matrix I_O.

$$\hat{\mu} = 304.242; \qquad \hat{\sigma} = 68.977; \qquad \hat{\xi} = 0.249;$$

$$I_O = \begin{pmatrix} 0.0062 & -0.0046 & 0.1962 \\ -0.0046 & 0.0091 & -0.2114 \\ 0.1962 & -0.2114 & 48.4122 \end{pmatrix}.$$

Example 2.3: Rainfall in the Lake District

(a) Estimate the standard errors for $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$, and use these to estimate the 95% confidence intervals for the GEV parameters. Comment.

(b) What is the estimated correlation between σ and ξ?

(c) Estimate the 100 and 1000 year return levels for daily rainfall totals at Eskdale.

(d) Obtain standard errors for your estimates in (b), and use these to construct 95% confidence intervals in the usual way. Comment.

The standard errors are found from the variance–covariance matrix, which is I_O^{-1}.

1. **Find the determinant of I_O:**

$$\begin{aligned}\det(I_O) &= 0.0062(0.0091 \times 48.4122 - 0.2114^2) \\ &\quad +0.0046(-0.0046 \times 48.4122 - 0.1962 \times (-0.2114)) \\ &\quad +0.1962(0.0046 \times 0.2114 - 0.1962 \times 0.0091) \\ &= 0.00146.\end{aligned}$$

2. **Find the transpose of I_O:**

$$I_O^T = \begin{pmatrix} 0.0062 & -0.0046 & 0.1962 \\ -0.0046 & 0.0091 & -0.2114 \\ 0.1962 & -0.2114 & 48.4122 \end{pmatrix}$$

Example 2.3: Solution to part(a) (3/8)

3. Find the determinants of each of the 2×2 minor matrices of I_O^T:

$$\begin{vmatrix} 0.0091 & -0.2114 \\ -0.2114 & 48.4122 \end{vmatrix} = 0.395861$$

$$\begin{vmatrix} -0.0046 & -0.2114 \\ 0.1962 & 48.4122 \end{vmatrix} = -0.181219$$

$$\begin{vmatrix} -0.0046 & 0.0091 \\ 0.1962 & -0.2114 \end{vmatrix} = -0.000813$$

$$\begin{vmatrix} -0.0046 & 0.1962 \\ -0.2114 & 48.4122 \end{vmatrix} = -0.181219$$

$$\begin{vmatrix} 0.0062 & 0.1962 \\ 0.1962 & 48.4122 \end{vmatrix} = 0.261661$$

$$\begin{vmatrix} 0.0062 & -0.0046 \\ 0.1962 & -0.2144 \end{vmatrix} = -0.000408$$

Example 2.3: Solution to part(a) (5/8)

$$\begin{vmatrix} -0.0046 & 0.1962 \\ 0.0091 & -0.2114 \end{vmatrix} = -0.000813$$

$$\begin{vmatrix} 0.0062 & 0.1962 \\ -0.0046 & -0.2114 \end{vmatrix} = -0.000408$$

$$\begin{vmatrix} 0.0062 & -0.0046 \\ -0.0046 & 0.0091 \end{vmatrix} = 0.000035$$

Example 2.3: Solution to part(a) (6/8)

4. This gives us the matrix of cofactors – we multiply each term in this by the signs indicated below, to get the **adjoint matrix**:

$$\begin{pmatrix} 0.395861 & -0.181219 & -0.000813 \\ -0.181219 & 0.261661 & -0.000408 \\ -0.000813 & -0.000408 & 0.000035 \end{pmatrix} \times \begin{pmatrix} + & - & + \\ - & + & - \\ + & - & + \end{pmatrix}$$

$$= \begin{pmatrix} 0.395861 & 0.181219 & -0.000813 \\ 0.181219 & 0.261661 & 0.000408 \\ -0.000813 & 0.000408 & 0.000035 \end{pmatrix}$$

5. Final step:

$$
\begin{aligned}
I_O^{-1} &= \frac{1}{\det(I_O)}\left(\text{adj}(I_O)\right) \\
&= 0.00146^{-1} \times \begin{pmatrix} 0.395861 & 0.181219 & -0.000813 \\ 0.181219 & 0.261661 & 0.000408 \\ -0.000813 & 0.000408 & 0.000035 \end{pmatrix} \\
&= \begin{pmatrix} 271.1377 & 124.1226 & -0.5568 \\ 124.1226 & 179.2199 & 0.2795 \\ -0.5568 & 0.2795 & 0.0240 \end{pmatrix}.
\end{aligned}
$$

The standard errors are thus $\sqrt{271.1377} = 16.466$, $\sqrt{179.2199} = 13.387$ and $\sqrt{0.0240} = 0.155$ for $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$, respectively.

Example 2.3: Solution to part(a) (8/8)

Using the estimated standard errors, we can form confidence intervals for the GEV parameters in the usual way:

$$\mu \quad : \quad 304.242 \pm 1.96 \times 16.466 \longrightarrow (271.969, 336.515)$$

$$\sigma \quad : \quad 68.977 \pm 1.96 \times 13.387 \longrightarrow (42.738, 95.216)$$

$$\xi \quad : \quad 0.249 \pm 1.96 \times 0.155 \longrightarrow (-0.055, 0.553)$$

Comment: $\xi > 0$ (i.e. we have Fréchet tails), suggesting unbounded, heavy tails – seems plausible for rainfall. However, the 95% CI for ξ passes through zero (only just, though!).

Example 2.3: Solution to part(b) (1/1)

The correlation between σ and ξ can be estimated using

$$\begin{aligned} \text{corr}(\hat{\sigma}, \hat{\xi}) &= \frac{\text{cov}(\hat{\sigma}, \hat{\xi})}{\sqrt{\text{var}(\hat{\sigma}) \times \text{var}(\hat{\xi})}} \\ &= \frac{0.2795}{\sqrt{179.2119 \times 0.0240}} \\ &= 0.1348. \end{aligned}$$

Example 2.3: Solution to part(c) (1/1)

Using Equation (2.10):

$$\hat{z}_r = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}} \left[\left(-\log \left(1 - r^{-1} \right) \right)^{-\hat{\xi}} - 1 \right],$$

we get:

$$\begin{aligned} \hat{z}_{100} &= 304.242 + \frac{68.977}{0.249} \left[\left(-\log \left(1 - 100^{-1} \right) \right)^{-0.249} - 1 \right] \\ &= 898.1133 \text{ mm}; \qquad \text{similarly,} \\ \hat{z}_{1000} &= 1574.085 \text{ mm}. \end{aligned}$$

Example 2.3: Solution to part(d) (1/3)

From earlier, we know that

$$\begin{aligned}\nabla z_r^T &= \left[\frac{\partial z_r}{\partial \mu}, \frac{\partial z_r}{\partial \sigma}, \frac{\partial z_r}{\partial \xi}\right] \\ &= \left[1, -\xi^{-1}(1 - y_r^{-\xi}), \sigma\xi^{-2}(1 - y_r^{-\xi}) - \sigma\xi^{-1} y_r^{-\xi} \log y_r\right],\end{aligned}$$

where $y_r = -\log(1 - r^{-1})$; evaluated at the MLEs for μ, σ and ξ, we get

$$\nabla z_{100}^T = [1, 8.6097, 1621.187]$$

Example 2.3: Solution to part(d) (2/3)

Thus,

$$
\begin{aligned}
\text{var}(\hat{z}_{100}) &= \nabla z_{100}^T V \nabla z_{100} \\
&= [1, 8.6097, 1621.187] \\
&\quad \times \begin{pmatrix} 271.1377 & 124.1266 & -0.5568 \\ 124.1266 & 179.2199 & 0.2795 \\ -0.5568 & 0.2795 & 0.0240 \end{pmatrix} \\
&\quad \times \begin{bmatrix} 1 \\ 8.6097 \\ 1621.187 \end{bmatrix} \\
&= 84768.61
\end{aligned}
$$

Therefore, s.e.$(\hat{z}_{100}) = \sqrt{84768.61} = 291.1505$.

Example 2.3: Solution to part(d) (3/3)

Similar calculations for $\hat{z}_{1000}$ give

$$\text{s.e.}(\hat{z}_{1000}) = 932.325.$$

From these standard errors, we can construct 95% confidence intervals:

$$\hat{z}_{100} \quad : \quad 898.1133 \pm 1.96 \times 291.1505 \longrightarrow (327.5, 1468.8) \text{ mm};$$

$$\hat{z}_{1000} \quad : \quad 1574.085 \pm 1.96 \times 932.325 \longrightarrow (-253.3, 3401.4) \text{ mm}.$$

Comment: CI's are very wide; also, the CI for $\hat{z}_{1000}$ is partly negative – not sensible!

Example 2.4: Solution

See R demo

Example 2.5

The magnitudes of the vertical forces produced by seismic *degassing bursts* are known to be associated with destructive volcanoes.

Table 2.5 shows the maximum force produced by degassing bursts, every quarter in the years 2005–2011 (inclusive), for the Kilauea Volcano in Hawaii.

	Jan–Mar	Apr–Jun	Jul–Sep	Oct–Dec
2005	99983	100067	99905	100367
2006	99980	99970	100086	99988
2007	99912	100084	100432	99921
2008	100123	99913	100240	100448
2009	100162	100296	100015	100122
2010	99930	99997	100710	100136
2011	99946	99918	99962	100052

Example 2.5

These data were stored the in the vector `haw` in R, and the following output obtained:

```
gev.fit(haw)

$ conv
[1] 0

$nllh
[1] 178.2502

$mle
[1] 9.998028e+04 8.710824e+01 5.921945e-01

$se
[1] 21.6185212 21.3598264 0.3027163
```

Example 2.5

(a) Use this information to estimate the degassing bursts we would expect to see (i) once every year, (ii) once every fifty years, and (iii) once every 100 years at this volcano (do not attempt to obtain standard errors for these estimates). How does your estimate in (i) compare to it's empirical counterpart?

(b) Use Figure 2.5 to assess the goodness–of–fit of the GEV to these data.

Example 2.5: Solution to part (a) (1/5)

From Equation (2.10) we know that

$$\hat{z}_r = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\left[\left(-\log\left(1 - r^{-1}\right)\right)^{-\hat{\xi}} - 1\right].$$

The quantity $\hat{z}_r$ is an estimate of the level that is exceeded, on average, once every r **observations** – if we have only one observation per year, this is exactly the the same as the r–year return level.

In this example, we have **four** observations per year!

Example 2.5: Solution to part (a) (2/5)

Thus, an estimate of the one–year return level is given by

$$\hat{z}_4 = 99980.28 + \frac{87.10824}{0.5921945}\left[\left(-\log(1-4^{-1})\right)^{-0.5921945} - 1\right]$$

$$= 100140.8 \text{ kg.}$$

Similarly, an estimate of the 50–year return level is given by

$$\hat{z}_{200} = 99980.28 + \frac{87.10824}{0.5921945}\left[\left(-\log(1-200^{-1})\right)^{-0.5921945} - 1\right]$$

$$= 103218.6 \text{ kg.}$$

Example 2.5: Solution to part (a) (3/5)

Also, an estimate of the 100–year return level is given by

$$\begin{aligned}\hat{z}_{400} &= 99980.28 + \frac{87.10824}{0.5921945}\left[\left(-\log(1-400^{-1})\right)^{-0.5921945} - 1\right] \\ &= 104940.6 \text{ kg.}\end{aligned}$$

Example 2.5: Solution to part (a) (4/5)

We can obtain the empirical estimate of the one–year return level by looking at the **empirical distribution function** (see Section 2.2.3 on quantile–quantile plots).

For example, ordering the data gives:

$x_{(i)}$	$i/(n+1)$
⋮	⋮
99946	0.241
99962	0.276
⋮	⋮

An empirical estimate of the one–year return level is the value which is exceeded once every 4 observations – the value for which $i/(n+1) = 0.25$.

Example 2.5: Solution to part (a) (5/5)

$i/(n+1) = 0.25$ is 25.71% of the way between 0.241 and 0.276; 25.71% of the way between 99946 and 99962 gives

$$\hat{z}_{4,\text{emp}} = 99950.11 \text{ kg.}$$

The model gives a slight over–estimate.

Example 2.5: Solution to part (b) (1/1)

Both the probability plot and *Q*-*Q* plots indicate the suitability of the fitted GEV to the quarterly maxima, with points lying close to the unit diagonal.

The fitted density also follows the histogram very closely.

All points in the return level plot fall within the 95% confidence bands (although these bands have been constructed using Wald intervals – see Section 2.4).

2.4 Profile likelihood

As we saw in our solution to Example 2.3, standard errors for long range return levels, obtained via the delta method, can often be so large that confidence intervals become difficult to work with – or even meaningless.

Actually, it turns out that constructing confidence intervals in the standard way:

$$\text{estimate} \pm 1.96 \times \text{s.e.}$$

is not advisable for return levels.

This is because of the **severe asymmetry** often observed in the likelihood surface for the return level, suggesting that the assumption of normality may not be valid for z_r.

An alternative, and often more accurate, method for making inferences on a particular parameter can be found using the **profile log–likelihood**.

Formally, the log–likelihood for a parameter vector $\boldsymbol{\theta}$ can be written as $\ell(\boldsymbol{\theta}_j, \boldsymbol{\theta}_{-j})$, where $\boldsymbol{\theta}_{-j}$ corresponds to all components of $\boldsymbol{\theta}$ excluding $\boldsymbol{\theta}_j$.

The **profile** log–likelihood for $\boldsymbol{\theta}_j$ is defined as

$$\ell_p(\boldsymbol{\theta}_j) \quad = \quad \max_{\boldsymbol{\theta}_{-j}} \ell(\boldsymbol{\theta}_j, \boldsymbol{\theta}_{-j}).$$

Thus, for each value of $\boldsymbol{\theta}_j$ the profile log–likelihood is the maximised log–likelihood with respect to all other components of $\boldsymbol{\theta}$.

2.4 Profile likelihood

For return levels using the GEV model:

1. Re–parameterise the GEV model so that z_r becomes one of the model parameters.

   For example, re–arrange equation (2.10) to write μ in terms of σ, ξ and z_r:

$$\mu = z_r - \frac{\sigma}{\xi}\left[\left(-\log\left(1-r^{-1}\right)\right)^{-\xi} - 1\right];$$

   then obtain an expression for the log–likelihood $\ell(\sigma, \xi, z_r)$ by substitution of (2.12) into

$$-m\log\sigma - (1+1/\xi)\sum_{i=1}^{m}\log\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]_{+} - \sum_{i=1}^{m}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]_{+}^{-1/\xi},$$

2. For some fixed lower value of $z_r = z_{r,\text{low}}$, maximise the GEV log–likelihood

$$\ell(\sigma, \xi, z_r = z_{r,\text{low}})$$

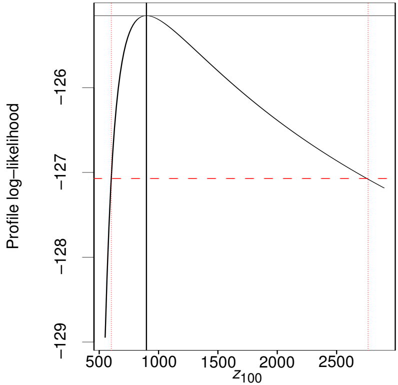
with respect to the two remaining parameters (σ and ξ), to obtain $\ell_p(z_r)$ at $z_{r,\text{low}}$.

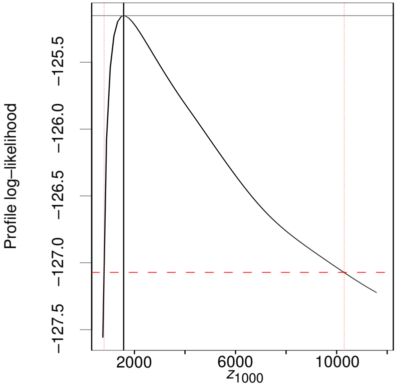
Here, in terms of the more general notation above, $\boldsymbol{\theta}_j = z_r$ and $\boldsymbol{\theta}_{-j} = (\sigma, \xi)$.

3. Repeat step (2) for a range of values z_r for $z_{r,\text{low}} \leq z_r \leq z_{r,\text{up}}$, and plot $\ell_p(z_r)$ against z_r to show the profile log–likelihood curve for z_r.

Figure 2.6 shows plots of the profile log likelihood for the 100 and 1000 year return levels ($\ell_p(z_{100})$ and $\ell_p(z_{1000})$ respectively) for the Lake District rainfall data in Example 2.3.

2.4 Profile likelihood





Both plots reveal strong asymmetry in the (profile) log–likelihood for the return levels, and it should be clear from these plots that constructing return levels in the usual way (as we did in Example 2.3, part (c)) will be misleading.

So how can we use these plots?

2.4 Profile likelihood

In this example, we are partitioning the GEV parameter vector $\boldsymbol{\theta} = (z_r, \sigma, \xi)$ into two components $(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)})$, where

$$\begin{aligned} \boldsymbol{\theta}^{(1)} &= z_r \quad \text{and} \\ \boldsymbol{\theta}^{(2)} &= (\sigma, \xi), \end{aligned}$$

and the profile log–likelihood is now defined as

$$\ell_p(\boldsymbol{\theta}^{(1)}) = \max_{\boldsymbol{\theta}^{(2)}}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}).$$

The following result leads to a procedure for making inferences on the maximum likelihood estimator of $\boldsymbol{\theta}^{(1)}$.

2.4 Profile likelihood: Result

Let $x_1, \ldots, x_n$ be independent realisations from a distribution within a parametric family $\mathcal{F}$, and let $\hat{\boldsymbol{\theta}}_0$ be the maximum likelihood estimator of the d–dimensional model parameter $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)})$, where $\boldsymbol{\theta}^{(1)}$ is a k–dimensional subset of $\boldsymbol{\theta}_0$. Then, under suitable regularity conditions, for large n

$$D_p(\boldsymbol{\theta}^{(1)}) \;\; = \;\; 2\left\{\ell(\hat{\boldsymbol{\theta}}_0) - \ell_p(\boldsymbol{\theta}^{(1)})\right\} \sim \chi^2_k.$$

Thus, for our single component $\boldsymbol{\theta}^{(1)} = z_r$, the set of values C_α for which $\{z_r : D_p(z_r) \leq c_\alpha\}$ provides a $(1-\alpha)$ confidence interval for z_r, where c_α is the $(1-\alpha)$ quantile of the χ^2_1 distribution.

2.4 Profile likelihood: Result

In *practice*:

- Obtain a plausible range of values for $D_p(z_r)$ from the χ_1^2 distribution.

  For example, working at the 5% level of significance, we would have values

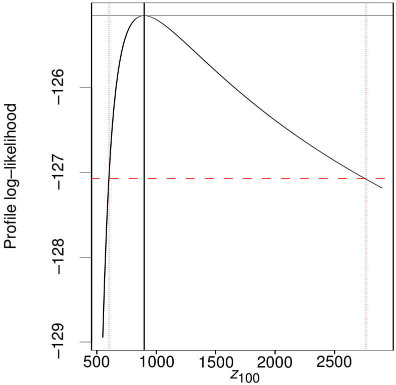
$$D_p(z_r) \leq \chi_1^2(0.05) = 3.842.$$

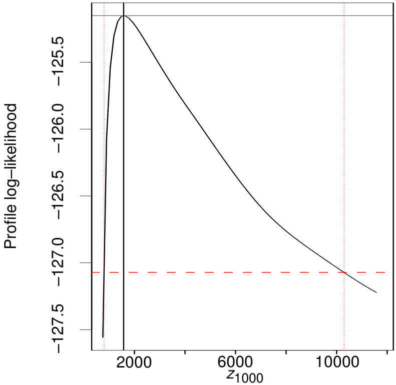
- Convert this into a range of plausible values for the profile log–likelihood:

$$\begin{aligned} \ell(\hat{\theta}_0) - \ell_p(z_r) &\leq 1.921 \\ \ell_p(z_r) &\geq \ell(\hat{\theta}_0) - 1.921 \end{aligned}$$

- Use this to obtain a plausible range of values for z_r.

2.4 Profile likelihood: Result





Profile Likelihood

	$\hat{z}_{100}$	$\hat{z}_{1000}$
Standard 95% CI	(327.5, 1468.8)mm	(–253.3, 3401.4)mm
Profiled 95% CI	(602.5, 2765)mm	(800, 10300)mm

Comments

- The profiled confidence intervals more accuratley capture the asymmetry in the log–likelihood for the return levels
- This manifests in much higher upper bounds
- Notice that the lower bounds are also affected, and are much closer to the MLE for z_r
- More realistic – and no negative rainfall values!
- Problem: confidence intervals often *extremely* wide, especially for long return periods (e.g. 1000 years)

Let $\bar{M}_n = \min\{X_1, \ldots, X_n\}$.

If we can assume the X_i are independent and identically distributed, we can apply similar arguments to $\bar{M}_n$ as we applied to M_n in Section 2.1.2.

If there exist sequences of constants $a_n > 0$ and b_n such that, as $n \to \infty$,

$$\Pr\left\{(\bar{M}_n - b_n)/a_n \leq x\right\} \quad \to \quad \bar{G}(x)$$

for some non–degenerate distribution $\bar{G}$, then $\bar{G}$ is a member of the GEV family of distributions for minima:

$$\bar{G}(x; \bar{\mu}, \sigma, \xi) \quad = \quad 1 - \exp\left\{-\left[1 - \xi\left(\frac{x - \bar{\mu}}{\sigma}\right)\right]_+^{-1/\xi}\right\},$$

where $a_+ = \max(0, a)$, $-\infty < \bar{\mu} < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$.

This result can be useful where we are interested in modelling extremely small, rather than extremely large, observations (e.g. annual minimum air temperatures).

Alternatively, we could negate our set of block minima and then model the corresponding set of maxima, giving identical maximum likelihood estimates of the GEV parameters but for the sign correction $\hat{\tilde{\mu}} = -\hat{\mu}$.

2.6 The GEV: Words of warning

- Block size?
  - Convention is to work with blocks equal in length to the calendar year, but what if you have hourly data collected over just a few years?
  - Block length too small – limiting arguments will not hold (the GEV is a limiting result, which holds *approximately* for large n)
  - Block length is too large – not enough maxima to work with!
  - Possible sensitivity of GEV parameter estimates to block length

2.6 The GEV: Words of warning

- Extremely wasteful of data:
  - Discard all but the block maxima
  - Often results in throwing away tens of thousands of observations – some of which might be 'extreme' – just not as extreme as the block maxima!
  - Results in large standard errors and extremely wide confidence intervals for return levels

2.6 The GEV: Words of warning

- Standard asymptotic likelihood results not always applicable for the GEV:
  - When $\xi < 0$, the end–points of the GEV are functions of the parameter values: e.g. $\mu = \sigma/\xi$ is the upper end–point
  - This violates the usual regularity conditions of MLEs
  - Effect: When $-1 < \xi \leq -0.5$, MLEs for ξ can be obtained, but are **super–efficient** (i.e. have variance smaller than the Cramér–Rao lower bound)
  - Another consequence: Cannot obtain MLE for ξ when $\xi \leq -1$