### **Chapter 2**

#### **Classical models for extremes**

Dr. Lee Fawcett MAS8304: Environmental Extremes

Suppose that  $X_1, X_2, ..., X_n$  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** 

$$M_n = \max \{X_1, X_2, \dots, X_n\}.$$
 (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:

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

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

$$Pr \{M_n \le x\} = Pr \{X_1 \le x, X_2 \le x, \dots, X_n \le x\}$$
  
= Pr \{X\_1 \le x\} × Pr \{X\_2 \le x\} × \dots × Pr \{X\_n \le x\}  
= {F(x)}<sup>n</sup>.

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<sub>n</sub> 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  $\infty$

This is analogous to the sample mean  $\bar{X}$  converging to the population mean  $\mu$  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  $\sigma$  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  $\mu$ .

#### 2.1.2 The Extremal Types Theorem

#### Theorem (The Extremal Types Theorem)

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

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

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

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

$$II: G(x) = \begin{cases} 0 & x \le 0\\ \exp(-x^{-\alpha}) & x > 0, \alpha > 0; \end{cases}$$
(2.4)  
$$III: G(x) = \begin{cases} \exp\{-(-x)^{\alpha}\} & x < 0, \alpha > 0\\ 1 & x \ge 0. \end{cases}$$
(2.5)

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  $\infty$ . 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

$$\frac{M_n-b_n}{a_n} \stackrel{D}{\longrightarrow} \mathbf{G},$$

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?

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.

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

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

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

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

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

Shape parameter  $\xi$ :

- $\xi = 0$ : 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



Dr. Lee Fawcett MAS8304: Environmental Extremes

But what about the **constants** *a<sub>n</sub>* and *b<sub>n</sub>*?

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

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

with  $a_n$  and  $b_n$  being absorbed into  $\mu^*$  and  $\sigma^*$ .

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.

Suppose  $X_1, X_2, ..., X_n$  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.

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

$$\Pr\left\{\frac{M_n - b_n}{a_n} \le z\right\} = \Pr\left\{\frac{\operatorname{Max}(X_1, \dots, X_n) - b_n}{a_n} \le z\right\}$$
$$= \Pr\left\{\frac{X_1 - b_n}{a_n} \le z, \dots, \frac{X_n - b_n}{a_n} \le z\right\}$$
$$= \Pr\left\{X_1 \le a_n z + b_n\right\} \times \cdots$$
$$= \left[1 - e^{-(a_n z + b_n)}\right]^n,$$

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

$$\begin{bmatrix} 1 - e^{-(z + \log n)} \end{bmatrix}^n = \begin{bmatrix} 1 - e^{-z} e^{\log n^{-1}} \end{bmatrix}^n$$
$$= \begin{bmatrix} 1 - n^{-1} e^{-z} \end{bmatrix}^n.$$

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

$$\left[1+rac{-e^{-z}}{n}
ight]^n\longrightarrow \exp\left(-e^{-z}
ight), \qquad ext{as } n o\infty.$$

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

Dr. Lee Fawcett MAS8304: Environmental Extremes

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

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

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:

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

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.

- 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!

#### 2.1.4 Typical application – 2: Parameter estimation

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:

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

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

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

(try this yourself!)

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

The likelihood is given by

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

Dr. Lee Fawcett MAS8304: Environmental Extremes

#### 2.1.4 Typical application – 2: Parameter estimation

The log-likelihood is given by

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

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

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

$$\frac{\partial \ell}{\partial \mu} = \frac{\xi + 1}{\sigma} \sum_{i=1}^{m} \left[ 1 + \xi \left( \frac{\mathbf{x}_i - \mu}{\sigma} \right) \right]^{-1} + \frac{1}{\sigma} \sum_{i=1}^{m} \left[ 1 + \xi \left( \frac{\mathbf{x}_i - \mu}{\sigma} \right) \right]^{-(1/\xi + 1)} = 0$$

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

The next step would be to replace  $\mu$ ,  $\sigma$  and  $\xi$  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.

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:

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

i.e.

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

#### Dr. Lee Fawcett MAS8304: Environmental Extremes

### 2.1.4 Typical application – 4: Return level estimation

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

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

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

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

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

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

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** (see MAS2305).

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

Dr. Lee Fawcett MAS8304: Environmental Extremes

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

For example, using simple empirical arguments,

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:

Then, to produce the plots shown in Figure 2.2:

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

### 2.2.1 Data pre-processing



Dr. Lee Fawcett MAS8304: Environmental Extremes

We now use R to maximise

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

with respect to  $\mu$ ,  $\sigma$  and  $\xi$ .

### In R:

> theta=c(mean(wassaw), sd(wassaw), 0.1) # Sets up parameter vector  $\theta = (\mu, \sigma, \xi)$ 

#### > 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)}
```

```
> 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)}
```

#### > 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
```

### 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: Standard errors

### In R:

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

Now:

> A

1

\$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

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

evaluated at the MLEs.

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

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

Recall from MAS2305 that inversion of the Information matrix gives the **Variance–Covariance matrix**.

Thus, we have a  $3 \times 3$  matrix to invert!

You should be able to do this (e.g. assignments), but let's use the solve command in  $\mathbf{R}$ .

In R, solve will solve the system of equations:

$$aX = b$$
.

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

$$I_{\rm obs} \cdot X = \left( \begin{array}{rrr} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right)$$

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:

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

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) (1.019, 1.603) and (-0.320, 0.104)

for  $\mu$ ,  $\sigma$  and  $\xi$  (respectively).

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

# 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, ..., 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

<b>x</b> <sub>(i)</sub>	Fitted: $G(\mathbf{x}_{(i)}; \hat{\mu}, \hat{\sigma}, \hat{\xi})$	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, ..., 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

<b>x</b> <sub>(i)</sub>	Fitted: $G(\mathbf{x}_{(i)}; \hat{\mu}, \hat{\sigma}, \hat{\xi})$	Empirical: $i/(n+1)$
6.7	0.016	0.020
6.8	0.021	0.039
7.3	0.063	0.059
		•
:		

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

Empirical: x <sub>(i)</sub>	Prob: <i>i</i> /( <i>n</i> + 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
	÷	:

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

Empirical: x <sub>(i)</sub>	Prob: <i>i</i> /( <i>n</i> + 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
		:

### 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  ${\tt GEV}\,.\,{\tt DF}$  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



#### Dr. Lee Fawcett MAS8304: Environmental Extremes

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):

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

we find that

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

similarly,

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

#### 2.2.4 Return level estimation

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  $\mu$ ,  $\sigma$  and  $\xi$ , we can use the **delta method** (see MAS2305) to find the approximate variance of  $\hat{z}_r$ .

Specifically,

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

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

$$\nabla z_r^T = \left[ \frac{\partial z_r}{\partial \mu}, \frac{\partial z_r}{\partial \sigma}, \frac{\partial z_r}{\partial \xi} \right]$$
  
=  $\left[ 1, -\xi^{-1} (1 - y_r^{-\xi}), \sigma \xi^{-2} (1 - y_r^{-\xi}) - \sigma \xi^{-1} y_r^{-\xi} \log y_r \right],$   
where  $y_r = -\log(1 - r^{-1})$ , evaluated at  $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ .

MAS8304: Environmental Extremes

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

We can define  $\nabla z_{10}$ , 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)
```

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

We can define  $\nabla z_{10}$ , 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)
```

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

We can define  $\nabla z_{10}$ , 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)
```

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

We can define  $\nabla z_{10}$ , 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)
```

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

We can define  $\nabla z_{10}$ , 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)

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

We can define  $\nabla z_{10}$ , 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  $- \$ * \$ - can be used to obtain an estimate of the standard error for <math>\hat{z}_{10}$  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.)					
<i>z</i> <sub>10</sub>	<i>Z</i> <sub>100</sub>	Z <sub>200</sub>	<i>Z</i> <sub>1000</sub>		
11.33 (0.361)	13.46 (0.938)	13.99 (1.182)	15.09 (1.821)		

# Class demonstration in R

Dr. Lee Fawcett MAS8304: Environmental Extremes

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	

Shown below are the maximum likelihood estimates of the GEV parameters, obtained using R; also shown is the observed information matrix  $I_{O}$ .

 $\hat{\mu} = 304.242;$   $\hat{\sigma} = 68.977;$   $\hat{\xi} = 0.249;$  $I_0 = \begin{pmatrix} 0.0062 & -0.0046 & 0.1962 \\ -0.0046 & 0.0091 & -0.2114 \\ 0.1962 & -0.2114 & 48.4122 \end{pmatrix}.$ 

### 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  $\sigma$  and  $\xi$ ?
- (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  $I_0^{-1}$ .

#### **1.** Find the determinant of $I_{\Omega}$ :

$$det(I_0) = 0.0062(0.0091 \times 48.4122 - 0.2114^2) +0.0046(-0.0046 \times 48.4122 - 0.1962 \times (-0.2114)) +0.1962(0.0046 \times 0.2114 - 0.2962 \times 0.0091) - 0.00146$$

0.00140

#### **2.** Find the transpose of $I_0$ :

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

Dr. Lee Fawcett MAS8304: Environmental Extremes

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

3. Find the determinants of each of the 2  $\times$  2 minor matrices:

$$\begin{vmatrix} 0.0091 & -0.2114 \\ -0.2114 & 48.4122 \end{vmatrix} = 0.395861$$
$$\begin{vmatrix} -0.0046 & -0.2114 \\ 0.1962 & 48.4122 \end{vmatrix} = -0.181219$$
$$\begin{vmatrix} -0.0046 & 0.0091 \\ 0.1962 & -0.2114 \end{vmatrix} = -0.000813$$

 $\begin{vmatrix} -0.0046 & 0.1962 \\ -0.2114 & 48.4122 \end{vmatrix} = -0.181219 \\ \begin{vmatrix} 0.0062 & 0.1962 \\ 0.1962 & 48.4122 \end{vmatrix} = 0.261661 \\ \begin{vmatrix} 0.0062 & -0.0046 \\ 0.1962 & -0.2144 \end{vmatrix} = -0.000408$ 

Dr. Lee Fawcett MAS8304: Environmental Extremes

-0.0046 0.0091	0.1962 0.2114	=	-0.000813
0.0062 -0.0046	0.1962 -0.2114	=	-0.000408
0.0062 0.0046	-0.0046 0.0091	=	0.000035

Dr. Lee Fawcett MAS8304: Environmental Extremes

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

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

 $\left( \begin{array}{cccc} 0.395861 & -0.181219 & -0.000813 \\ -0.181219 & 0.261661 & -0.000408 \\ -0.000813 & -0.000408 & 0.000035 \end{array} \right) \times \left( \begin{array}{cccc} + & - & + \\ - & + & - \\ + & - & + \end{array} \right)$ 

	/ 0.395861	0.181219	-0.000813
=	0.181219	0.261661	0.000408
	\ −0.000813	0.000408	0.000035

## Example 2.3: Solution to part(a) (7/8)

#### 5. Final step:

$$\begin{array}{rcl} & -1 & = & \displaystyle \frac{1}{\det(I_{obs})} \left( \operatorname{adj}(I_{obs}) \right) \\ & = & \displaystyle 0.00146^{-1} \times \left( \begin{array}{ccc} 0.395861 & 0.181219 & -0.000813 \\ 0.181219 & 0.261661 & 0.000408 \\ -0.000813 & 0.000408 & 0.000035 \end{array} \right) \\ & = & \displaystyle \left( \begin{array}{ccc} 271.1377 & 124.1226 & -0.5568 \\ 124.1226 & 179.2199 & 0.2795 \\ -0.5568 & 0.2795 & 0.0240 \end{array} \right). \end{array}$$

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:

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

$$\sigma$$
 : 68.977 ± 1.96 × 13.387  $\longrightarrow$  (42.738, 95.216)

$$\xi$$
 : 0.249 ± 1.96 × 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  $\xi$  passes through zero (only just, though!).

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

The correlation between  $\sigma$  and  $\xi$  can be estimated using

$$\operatorname{corr}(\hat{\sigma}, \hat{\xi}) = \frac{\operatorname{cov}(\hat{\sigma}, \hat{\xi})}{\sqrt{\operatorname{var}(\hat{\sigma}) \times \operatorname{var}(\hat{\xi})}}$$
$$= \frac{0.2795}{\sqrt{179.2119 \times 0.0240}}$$
$$= 0.1348$$

### 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:

$$\hat{z}_{100} = 304.242 + \frac{68.977}{0.249} \left[ \left( -\log \left( 1 - 100^{-1} \right) \right)^{-0.249} - 1 \right]$$

$$= 898.1133 \text{ mm}; \quad \text{similarly,}$$

 $\hat{z}_{1000} = 1574.085$  mm.

From earlier, we know that

$$\nabla \boldsymbol{z}_{r}^{T} = \begin{bmatrix} \frac{\partial \boldsymbol{z}_{r}}{\partial \mu}, \frac{\partial \boldsymbol{z}_{r}}{\partial \sigma}, \frac{\partial \boldsymbol{z}_{r}}{\partial \xi} \end{bmatrix}$$
$$= \begin{bmatrix} 1, -\xi^{-1}(1 - \boldsymbol{y}_{r}^{-\xi}), \sigma\xi^{-2}(1 - \boldsymbol{y}_{r}^{-\xi}) - \sigma\xi^{-1}\boldsymbol{y}_{r}^{-\xi} \log \boldsymbol{y}_{r} \end{bmatrix},$$

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

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

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

Thus,

$$\operatorname{var}(\hat{z}_{100}) = \nabla z_{100}^T V \nabla z_{100}$$

$$= \begin{bmatrix} 1, 8.6097, 1621.187 \end{bmatrix} \\ \times \begin{pmatrix} 271.1377 & 124.1266 & -0.5568 \\ 124.1266 & 179.2199 & 0.2795 \\ -0.5568 & 0.2795 & 0.0240 \end{pmatrix} \\ \times \begin{bmatrix} 1 \\ 8.6097 \\ 1621.187 \end{bmatrix}$$

= 84768.61

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

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

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

 $\hat{z}_{100}$  : 898.1133 ± 1.96 × 291.1505  $\longrightarrow$  (327.5, 1468.8) mm;

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

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