The Distribution of the Extreme from a Normal Sample
M2R Project, Group 41, Prof. A.T. Walden

Yifan Chen, Nishant Desai, Chang Goh, Maitanki Sutharson, Tudor Trita Trita

Imperial College London, 18th June 2018

Abstract

In this project, we look at the properties of maxima arising from normal random samples. We will find that asymptotically, there are only three types of distribution (Gumbel, Fréchet and Weibull) that describe the data and these can be combined into one; the Generalised Extreme Value (GEV) distribution. We will begin by proving this result, known as the Fisher-Tippett Theorem, and then we present methods (MoM, MLE, PWM) for estimating the parameters of the Gumbel distribution and GEV distribution. We then look at Quantile-Quantile (QQ) plots and Bootstrapping techniques, which find the model with the best fit.

The project then focuses on analysing earthquake data from Greece to be able to derive statistical properties and make future predictions. We do this by fitting a Gumbel distribution and GEV distribution to the Greece earthquake data, and estimating parameters for each model using the methods above. We then use QQ plots and Bootstrap Confidence Intervals to find the model that fits the Greek earthquake data the most. Finally, we look at return periods which will give predictions as to when to expect the next earthquake of a given magnitude.

Contents

1 Introduction

Modelling extreme events has been crucial in seismography, oceanography and insurance. One of the first applications of Extreme Value distributions was modelling flood flows, made by Fuller in 1914. In 1920, Griffith applied extreme value theory to the behaviour of flow and fracture in solids. In the 1940s, Gumbel played a crucial role introducing more exciting applications of extrema, including radioactive emissions and human lifespans. (Kotz & Nadarajah, 2000)

The maxima are taken from a large sample of independent and identically distributed random variables. As the sample size goes to infinity, the behaviour of the maxima is described by the three Extreme Value distributions - Type 1 (Gumbel), Type 2 (Fréchet) and Type 3 (Weibull), which can be written as a unified distribution, the Generalised Extreme Value (GEV) distribution.

Gumbel, Fréchet and Weibull consist of 2 parameters, μ and σ, with $k = 0$, $k > 0$ and $k < 0$ for Gumbel, Fréchet and Weibull respectively. GEV distribution has 3 parameters: μ, σ and k. (Coles, 2001)

Extreme Value distributions have several applications in the environmental science. The GEV distribution has been used for flood frequency analysis in the UK to model daily or annual maxima of natural occurrences, for example rainfall, sea levels and river lengths. Gumbel distribution has been used heavily in hydrology for modelling extreme events, engineering and annual flood flows in 1958. Fréchet distribution has had many applications in finance including modelling market returns, which generally have heavy tails. Weibull distribution initially was developed to observe minima in material analysis. (Alves & Neves, 2008) The GEV distribution is usually used to model the extrema of long finite sequences of random variables.

In 1990, de Haan estimated the parameters, ψ in the distribution based on the high-tide water levels at Hoek van Holland, the Dutch Station during 1887 until 1984.

$$G_\psi(x) = \begin{cases} \exp(-(1+\psi x)^{1/\psi}), & 1+\psi x > 0, \\ \exp(-(\exp(-x))), & \psi = 0. \end{cases}$$

If $\psi = 0$, the water levels follow a Gumbel distribution (under GEV with $\psi = 0$) with $\mu = 0$ and $\sigma = 1$, and if $1 + \psi x > 0$, it is a GEV distribution (Fréchet and Weibull) with $\mu = 0$ and $\sigma = 1$. (Kotz & Nadarajah, 2000)

In 1995, Robinson and Tawn used the GEV distribution to see whether an athlete's record is better than the ultimate performance predicted by the data. The performance of Wang Junxia, a Chinese 3000m flat track athlete in Beijing national championship ran her personal best of 486.11 seconds (1993). Her time was 6.08 seconds faster than the day before, and 10.43 seconds faster than her previous record, hence there was suspicion the record was drug assisted.

Annual minima for the women's 3000m race from 1972-1992 with Junxia's time in 1993 was considered in order to determine whether her performance was drug-enhanced. A confidence interval with the GEV distribution was constructed to find the minimum time possible time- this was insufficient to reach a conclusion. Therefore the five best annual times for 3000m is included, meaning Wang's time is within the 90% confidence interval (430.1, 493.8). Alongside this, incorporating relative performances in 1500m events (478.4, 495.0) and Olympic and World championship years (486.3, 497.0), Wang's time is within the 90% confidence interval. Therefore, Robinson and Tawn concluded that there is no legal case that can say Wang's performance in 1993 was drug-enhanced; her time is within the 90% confidence interval for the ultimate time. (Smith, 2009)

The above applications are based on the data of the maxima will fit an Extreme Value distribution. In order to find this distribution, the parameters (location, scale, shape) need to be estimated for Gumbel distribution and GEV distribution. There are three main methods of parameter estimation, Method of Moments, Maximum Likelihood Estimation and Probability Weighted Moments. It is important to compare these methods of parameter estimation as Figure 1 highlights that different sets of parameters give very different PDFs.

These estimation methods will be discussed and compared in detail for Extreme Value distributions, and put into practice the worked example. After estimating the parameters for Gumbel distribution and GEV distribution, the goodness of fit needs to be tested with Quantile-Quantile plots (QQ plots) and Bootstrap Confidence Intervals.

Using these results, the return value is calculated, which estimates the frequency of extreme quantiles occurring with a certain level. This will be useful in risk management, as information about the likelihood of rare events can be gathered. (National Aeronautics and Space Administration, 2018)

In the second part of the report, the worked example will analyse earthquake data in Greece from 1901-2017. The data will be modelled as Extreme Value distributions, either Gumbel distribution or GEV distribution, with its parameters estimated by Method of Moments, Maximum Likelihood Estimation and Probability Weighted Moments. The goodness of fit for each of the models will tested by QQ plots and Parametric Bootstrap to find the best model. A conclusion will then be made taking the return level into account.

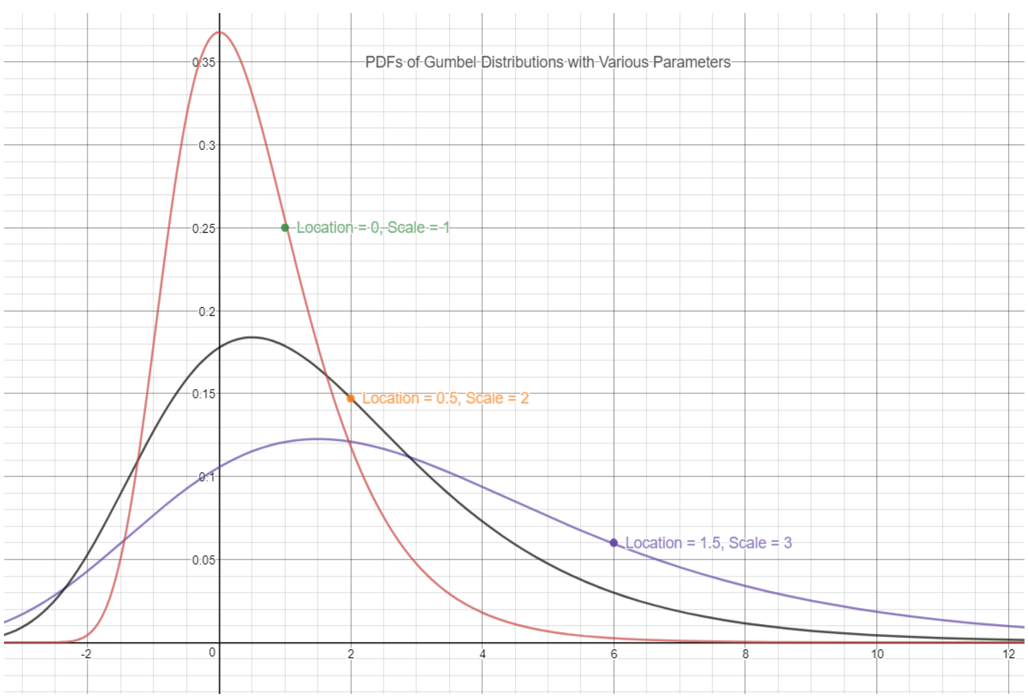

Figure 1: PDFs of Gumbel Distributions with Various Parameters

2 Extreme Value Distributions

2.1 Model Formulation

We consider a model focused on the statistical behaviour of $M_n = \max\{X_1, X_2, \ldots, X_n\}$, where $X_1, X_2, \ldots, X_n$ is a sequence of independent random variables having a common distribution function F.

The X_is are the measurements on a regular time period and M_n is the maximum value over the n observations in the entire time period.

As the X_i are i.i.d., we know that

$$\Pr\{M_n \leqslant x\} = \Pr\{X_1 \leqslant x, \ldots, X_n \leqslant x\} = \Pr\{X_1 \leqslant x\} * \ldots * \Pr\{X_n \leqslant x\} = \{F\{x\}\}^n.$$

However, as the distribution function of F is unknown, $\Pr\{M_n \leqslant x\} = \{F\{x\}\}^n$ is not useful yet at this point.

We choose to accept that F is unknown and look at the families of models of F^n instead. We start by considering the behaviour of F^n as n tends to infinity. For any $X < X_+$, where X_+ is the upper end point of F, (x_+ is the smallest value of x such that $F(x) = 1$, $F^n(x) \to 0$ as $n \to \infty$ and so distribution of M_n degenerates to a point mass on x_+. This difficulty is avoided by allowing a linear renormalisation of the variable M_n:

$$M_n^* = \frac{M_n - b_n}{a_n}$$

with appropriate choices of the sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ which stabilise the location and scale of M_n^* as n increases respectively. (Page 46, Coles, 2001).

If there exist sequences of constants $\{a_n\}$ and $\{b_n\}$ such that

$$\Pr\left(\frac{M_n - b_n}{a_n} \leqslant x\right) \to G(x), \quad \text{as} \quad n \to \infty,$$

where G is a non-degenerate distribution function, then G belongs to one of the following families:

$$\textbf{Type I}: G(x) = \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right), \quad \forall x \in \mathbb{R}. \tag{1}$$

$$\textbf{Type II}: G(x) = \exp\left(-\left(\frac{x-\mu}{\sigma}\right)^{-k}\right), \text{ if } x \geq \mu, \text{ and } G(x) = 0 \text{ otherwise.} \tag{2}$$

$$\textbf{Type III}: G(x) = \exp\left(-\left(-\frac{x-\mu}{\sigma}\right)^{k}\right), \text{ if } x \leq \mu, \text{ and } G(x) = 1 \text{ otherwise.} \tag{3}$$

for parameters $\sigma > 0$, $\mu \in \mathbb{R}$ and in the case of types II and III, $k > 0$. (Page 46, Coles, 2001)

The three families are the extreme value distributions known as **Gumbel, Fréchet and Weibull** respectively with location and scale parameters μ, σ resp. and k is the shape parameter.

2.2 The GEV Distribution

It is in general difficult to know which distribution to adopt and subsequently this may produce errors in the estimates of the parameters of the distributions. Furthermore, once the choice is made, subsequent inference presume the choice to be correct and not allowing for uncertainty in such a choice would make errors substantial. (Page 47, Coles, 2001)

We therefore generalise the three families and combine them into a single family of models CDF of the form

$$G(z) = \begin{cases} \exp\left(-\left(1 + k\ \dfrac{x-\mu}{\sigma}\right)^{-1/k}\right), & \text{if } k \neq 0, \\ \exp\left(-\exp\left(-\dfrac{x-\mu}{\sigma}\right)\right), & \text{if } k = 0, \end{cases} \tag{4}$$

defined on set $\{x : 1 + k((z-\mu)/\sigma) > 0\}$ for $-\infty < \mu < \infty,\ \sigma > 0,\ -\infty < k < \infty$. (Page 47,48, Coles, 2001)

This is called the **Generalised Extreme Value** (GEV) family of distributions. Here μ is the location parameter, σ is the scale parameter and k is the shape parameter. Particularly, if when $k > 0$ resp. $k < 0$, then a Fréchet resp. Weibull family occurs.

We now consider the case when $k \to 0$ and we find that

$$G(x) \to \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right),$$

which is exactly the same as a Gumbel family.

Generally, we say G is a member of the GEV family if there exist sequences of constants $a_n > 0$ and b_n such that

$$\Pr\left(\frac{M_n - b_n}{a_n} \leqslant x\right) \to G(x), \quad \text{as} \quad n \to \infty,$$

and equivalently, for large enough n, $\Pr(M_n \leqslant x) \approx G((x - b_n)/a_n) = G^*(x)$ where $G^*(x)$ is another member of the GEV family. (Page 48, Coles, 2001)

2.3 Return Levels and Return Periods

Additionally, we consider z_p, the return level associated with the return period $1/p$, the level z_p is precisely exceeded by the annual maximum in any particular year with probability p, where

$$z_p = \begin{cases} \mu - (\sigma/k)[1 - (-\log(1-p))^{-k}], & k \neq 0, \\ \mu - \sigma \log\{-\log(1-p)\}, & k = 0. \end{cases}$$

It is said in (Coles, 2001) that "since quantiles enable probability models to be expressed on the scale of data, the relationship of the GEV model to its parameters is most easily interpreted in terms of the quantile expressions. In particular, defining $y_p = -\log(1-p)$, such that

$$z_p = \begin{cases} \mu - (\sigma/k)[1 - {y_p}^{-k}], & k \neq 0, \\ \mu - \sigma \log y_p, & k = 0. \end{cases}$$

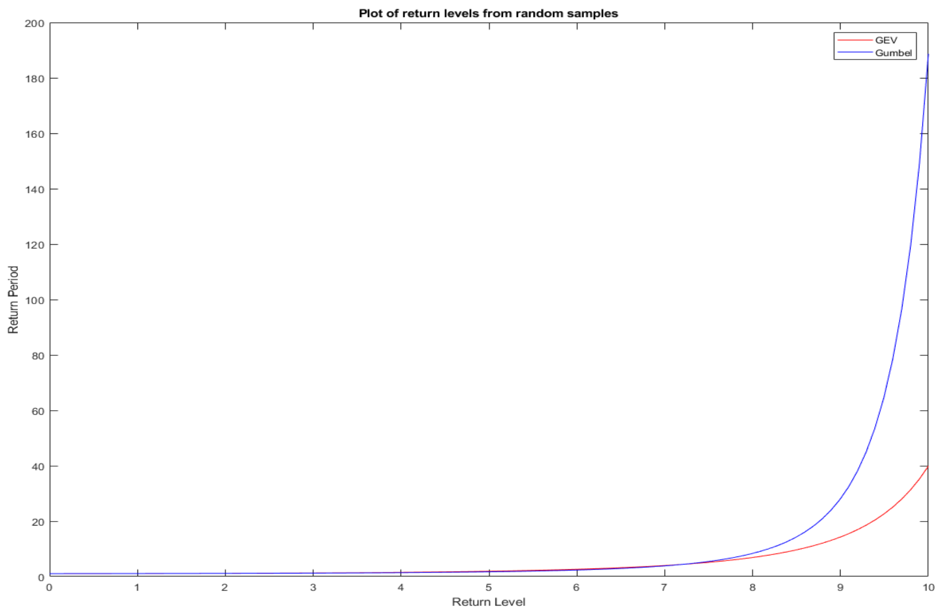

Figure 2: Plot of Return Levels from Random Samples

It follows that if z_p is plotted against y_p on a logarithmic scale, the plot is linear in the case $k = 0$. If $k < 0$, the plot is convex with asymptotic limit as $p \to 0$ at $(\mu - \sigma)/k$. If $k > 0$ the plot is concave and has no finite bound and this graph is a return level plot."

We illustrate this by looking at Figure 2 above.

We generate 100 random samples and fit them into a model of Gumbel and GEV and then we create the plot of return level against return period. For the sake of simplicity, assume that the return level represents magnitude and the return period is measured in years in the graph above. We can therefore make predictions; for the GEV model, an event of magnitude 10 will on average occur every 180 years, and for the Gumbel model, an event of magnitude 10 will on average occur every 40 years.

In particular, one can find out the specific return period T of a particular level by solving the equation

$$T = \frac{1}{1 - G(Z_p)}.$$

For example, if we want the return period for an event of magnitude 5, it would occur in 2.0163 years for GEV and 1.7462 for Gumbel.

3 Asymptotic Properties

The goal of this section is to formalise the claims of Section 2.1.

We begin with some definitions:

Definition 3.1 (Type of Distribution)**.** *Let X and Y be two random variables with distributions μ and ν respectively. We say that μ and ν are of the same* ***type*** *if there exists $a > 0$ and $b \in \mathbb{R}$ such that $aX + b$ has the same distribution as Y.*

Definition 3.2 (Max-Stable Distributions, (Sun, 2018))**.** *A non-degenerate probability distribution μ on $\mathbb{R}$ is called max-stable if for a sequence of i.i.d random variables $(X_i)_{i\in\mathbb{N}}$ with distribution μ and for each $n \in \mathbb{N}$, there exists $a_n > 0$ and $b_n \in \mathbb{R}$ such that $(M_n - b_n)/a_n$ also has distribution μ.*

Equivalently, μ is max-stable if its CDF $F(x) := \mu(-\infty, x]$ satisfies: for each $n \in \mathbb{N}$, there exists $a_n > 0$ and $b_n \in \mathbb{R}$ such that:

$$F^n(a_n x + b_n) = F(x), \forall x \in \mathbb{R}.$$

Remark. *Note that $M_n := \max\limits_{1\leq i\leq n} X_i$.*

The following lemma is useful in proving the main result that follows.

Lemma 3.1. *If G is max-stable, then there exist real-valued functions $a(s) > 0$, and $b(s)$ defined for $s > 0$, such that:*

$$G^n(a(s)x + b(s)) = G(x).$$

The main theorem of this section, the Extremal Types Theorem, also known as the Fisher-Tippett-Gnedenko Theorem:

Theorem 3.2 (Extremal Types Theorem)**.** *Let $X_1, \dots, X_n$ be i.i.d. with CDF F and let $X_{(n)} = \max\limits_{1\leq i\leq n} X_{(i)}$. If there exist constants $a_n > 0$ and b_n and a non-degenerate distribution function G such that*

$$\Pr\left(\frac{X_{(n)} - b_n}{a_n} \leq x\right) \xrightarrow{D} G(x),$$

then G must be of the same type as one of the three extreme value distributions (Gumbel, Fréchet, Weibull).

Conversely, any distribution function of the same type as one of these extreme value classes can appear as such a limit. (Nadarajah, 2018a)

Proof. This proof very closely follows (Nadarajah, 2018a). To prove the theory, it suffices to check that the set of max-stable distribution functions is the same as the set of distribution functions of the same type as the Gumbel, Fréchet and Weibull distributions.

Step 1: Check that the Gumbel, Fréchet and Weibull distributions are max-stable.

For Gumbel, if $a_n = 1, b_n = \log n$, then

$$\begin{aligned} G^n(a_n x + b_n) &= \exp(-n \exp -(a_n x + b_n)), \\ &= G(x). \end{aligned}$$

For Fréchet, if $a_n = n^{1/\alpha}, b_n = 0$, then:

$$\begin{aligned} G^n(a_nx+b_n) &= \exp(-n(a_nx+b_n)^{-\alpha}), \text{ if } x>0, \text{ and zero otherwise}, \\ &= G(x). \end{aligned}$$

For Weibull, if $a_n = n^{-1/\alpha}, b_n = 0$, then:

$$\begin{aligned} G^n(a_nx+b_n) &= \exp(-n(-a_nx-b_n)^{\alpha}), \text{ if } x<0, \text{ and one otherwise}, \\ &= G(x). \end{aligned}$$

Thus we have checked that $G^n(a_nx+b_n) = G(x)$ in all three cases, thus they are max-stable.

Step 2: Now suppose G is max-stable, then by **Lemma 3.1**,

$$G^s(a(s)x+b(s)) = G(x).$$

Taking logs, for $0 < G(x) < 1$,

$$-\log(\log(G(a(s)x+b(s)))) - \log s = \log(-\log G(x)).$$

The max-stability property with $n = 2$ implies that $G^2(ax + b) = G(x)$ for some $a > 0$ and $b \in \mathbb{R}$, so that G does not jump at $x_- = \sup\{x : G(x) = 0\}$ or $x_+ = \inf\{x : G(x) = 1\}$ if these are finite. (Nadarajah, 2018a)

Thus, the non-decreasing function $\phi(x) = -\log(-\log G(x))$ is such that

$$\lim_{x\to x_-} \phi(x) = -\infty, \qquad \lim_{x\to x_+} \phi(x) = +\infty.$$

Therefore, ϕ has an inverse function $U(y) = \inf\{x \in \mathbb{R} : \phi(x) \geq y\}, \forall y \in \mathbb{R}$. Now, $\phi(a(s)x+b(s)) - \log s = \phi(x)$, it follows that:

$$\begin{aligned} U(y) &= \inf\{x : \phi a(a(s)x+b(s)) - \log s \geq y\}, \\ &= \frac{1}{a(s)}[\inf\{x' : \phi(x') \geq y + \log s\} - b(s)], \\ &= \frac{U(y+\log s) - b(s)}{a(s)}. \end{aligned}$$

Subtracting for $y = 0$,

$$\frac{U(y+\log s) - U(\log s)}{a(s)} = U(y) - U(0),$$

Write $z = \log s, \quad \tilde{\eta}(z) = a(e^z), \quad \tilde{U}(y) = U(y) - U(0)$.

$$\tilde{U}(y+z) - \tilde{U}(z) = \tilde{U}(y)\tilde{\eta}(z), \quad \forall y, z \in \mathbb{R}. \tag{5}$$

Interchanging y and z and subtracting,

$$\tilde{U}(y)(1-\tilde{\eta}(z)) = \tilde{U}(z)(1-\tilde{\eta}(y)). \tag{6}$$

Case 1: $\tilde{\eta}(z_0) \neq 1$ for some $z_0 > 0$.

Then $\tilde{\eta}(z) \neq 1, \forall z > 0$, because otherwise $\exists z > 0$, s.t. $\tilde{U}(z) = 0$. But then $\tilde{U}(y+z) = \tilde{U}(y), \forall y$ by (5), so $U(y+z) = U(y)$ $\forall y \in \mathbb{R}$. This is a contradiction.

Fixing $z > 0$, writing $c = \tilde{U}(z)/(1-\tilde{\eta}(z))$ and noting from (6) that this is constant, we have from (5) that

$$c(1-\tilde{\eta}(y+z)) - c(1-\tilde{\eta}(z)) = c(1-\tilde{\eta}(y))\tilde{\eta}(z),$$

such that

$$\tilde{\eta}(y+z) = \tilde{\eta}(y)\tilde{\eta}(z), \quad \forall y \in \mathbb{R}.$$

But $\tilde{\eta}$ is monotone, since $\tilde{U}(y) = c(1-\tilde{\eta}(y))$ from (6), and the only non-zero solutions that are monotone and not identically equal to 1 are $\tilde{\eta}(y) = \exp(\rho y)$ for some $\rho \neq 0$. But then

$$\phi^{-1}(y) = U(y) = \zeta + c(1-e^{\rho y}),$$

where $\zeta = U(0)$. Since ϕ^{-1} is non-decreasing, we must have that $c < 0$ if $\rho > 0$ and $c > 0$ if $\rho < 0$, so in fact ϕ^{-1} is continuous and strictly increasing.

Hence

$$x = \phi^{-1}(\phi(x)) = \zeta + c\left(1 - e^{\rho\phi(x)}\right) = \zeta + c(1-(-\log G(x))^{-\rho}),$$

so

$$G(x) = \exp\left(-\left(1 - \frac{x-\zeta}{c}\right)^{-\frac{1}{\rho}}\right), \quad \text{for } 0 < G(x) < 1.$$

From the continuity of G at any finite endpoints, we see that G is of Type II, with $\alpha = 1/\rho$, if $\rho > 0$, and of Type III, with $\alpha = -1/\rho$, if $\rho < 0$.

Case 2: $\tilde{\eta}(z) = 1,\ \forall z > 0$.

But then, from (5), $\tilde{U}(y+z) = \tilde{U}(y) + \tilde{U}(z)$, for which the only non-constant non-decreasing solutions are $\tilde{U}(y) = \rho y$ for some $\rho > 0$. Thus $\phi^{-1}(y) = U(y) = \zeta + \rho y$, where $\zeta = U(0)$. Since this is continuous and strictly increasing

$$x = \phi^{-1}(\phi(x)) = \rho\phi(x) + \zeta = -\rho\log(-\log G(x)) + \zeta,$$

hence

$$G(x) = \exp\left(-\exp\left(-\frac{x-\zeta}{\rho}\right)\right), \quad \text{for } 0 < G(x) < 1,$$

and since G has no jumps at any finite endpoints, G is of Type I. (Nadarajah, 2018a)

$\square$

4 Parameter Estimation

Estimating parameters is crucial when wanting to fit a distribution to a set of real data. By estimating parameters, we can judge whether the data follows a Gumbel or GEV distribution. In this section we present three methods for performing parameter estimation. The Method of Moments, seen in courses this year is an old and ad-hoc method which performs poorly for small random samples. The Maximum Likelihood Estimators, also seen in courses this year, perform significantly better but as we will see, are hard to derive for the extreme distributions. The final method we will present are the Probability Weighted Moments, not covered in lectures this year. For historical reasons, we will only do parameter estimation on the Gumbel and GEV distributions.

4.1 Method of Moments (MoM)

4.1.1 MoM Estimation for the Gumbel Distribution

The cumulative distribution function (CDF) of a Gumbel distribution is given by

$$F(y) = \Pr(Y \leq y) = \exp(-\exp(-y)).$$

By differentiating $F(y)$ w.r.t. y, we obtain its probability density function (PDF) as

$$f(y) = \begin{cases} \exp(-y - \exp(-y)), & -\infty < x < \infty, \\ 0, & \text{otherwise.} \end{cases}$$

Its moment generating function (MGF) is given by

$$\begin{aligned} \mathrm{M}(t) = \mathrm{E}(e^{ty}) &= \int_{-\infty}^{\infty} \exp(ty) \exp[-y - \exp(y)]\, dy, \\ &= \int_{0}^{\infty} x^{-t} e^{-x} dx, \ \ \text{(by the substitution } x = e^y\text{)}, \\ &= \Gamma(1-t). \ \ \text{(a gamma function)} \end{aligned}$$

By differentiating $\mathrm{M}(t)$ w.r.t. t and setting $t = 0$, we obtain the first and second moment, and the variance as:

$$\begin{aligned} \mathrm{E}(Y) &= -\Gamma'(1) = \gamma, \\ \mathrm{E}(Y^2) &= \Gamma''(1) = \gamma^2 + \frac{\pi^2}{6}, \\ \mathrm{Var}(Y) &= \mathrm{E}(Y^2) - [\mathrm{E}(Y)]^2 = \frac{\pi^2}{6}, \end{aligned}$$

where $\gamma \approx 0.57722$ is the Euler-Mascheroni constant. (Kotz & Nadarajah, 2000)

Let $X = \sigma Y + \mu$, then

$$\begin{aligned} \mathrm{E}(X) &= \sigma\gamma + \mu, \\ \mathrm{Var}(X) &= \frac{1}{6}\sigma^2\pi^2. \end{aligned}$$

Equating the equations above with $\bar{X}$ (sample mean) and S^2 (sample variance), we obtain the moment estimators of μ and σ as

$$\hat{\mu} = \bar{X} - \gamma\hat{\sigma}, \tag{7}$$

$$\hat{\sigma} = \frac{\sqrt{6}}{\pi} S. \tag{8}$$

Remark. *When implementing equation (7) it was necessary to change the equation for the estimator to work. The equation used was* $\hat{\mu} = \bar{X} + \gamma\hat{\sigma}$*.*

4.1.2 MoM Estimation for the GEV Distribution

According to Stedinger et al. (1993) as cited by Bhunya et al. (2007), the moment estimators for the GEV parameters are:

$$\hat{\mu} = \bar{X} - \frac{\hat{\sigma}}{\hat{k}}[1 - \Gamma(1 + \hat{k})], \tag{9}$$

$$\hat{\sigma} = \operatorname{sign}(\hat{k})\ \frac{S\hat{k}}{\{\Gamma(1 + 2\hat{k}) - [\Gamma(1 + \hat{k})]^2\}^{1/2}}, \tag{10}$$

$$\hat{\gamma} = \operatorname{sign}(\hat{k})\ \frac{-\Gamma(1 + 3\hat{k}) + 3\Gamma(1 + \hat{k})\Gamma(1 + 2\hat{k}) - 2[\Gamma(1 + \hat{k})]^3}{\{\Gamma(1 + 2\hat{k}) - [\Gamma(1 + \hat{k})]^2\}^{3/2}}, \tag{11}$$

where $\operatorname{sign}(\hat{k}) = \pm 1$ is the sign of $\hat{k}$, $\Gamma(\cdot)$ is the gamma function, and $\bar{X}, S,$ and $\hat{\gamma}$ are the sample mean, standard deviation, and skewness, respectively.

The equations above contain multiple Gamma functions of $\hat{k}$, whose sign is not known. Therefore, for a given set of data, to evaluate $\hat{k}$, the following steps were used:

1. 10000 values of k in the range $-0.5 \le k \le 0.5$ were generated using MATLAB.
2. The corresponding values of γ for each k were computed using equation (11).
3. The γ computed were compared with $\hat{\gamma}$, which is the skewness of the sample.
4. The value of k which gives the smallest value of $|\hat{\gamma} - \gamma|$ was chosen as the estimate, $\hat{k}$.
5. The corresponding values of μ and σ were computed using the $\hat{k}$ obtained in step 4 and equations (9) and (10).

Remark. *In order for the estimator to work, the final value of* $\hat{k}$ *is taken as* $-\hat{k}$*.*

4.1.3 Properties of Moment Estimators

The validity of moment estimators were tested using the simulation study by the following steps:

1. Four different sample sizes (n) were chosen, namely $n = 5, n = 20, n = 50$, and $n = 100$.
2. 100 random samples of each sample sizes for some selected parameter values were generated using MATLAB.

3. The parameter estimates were computed using equations (9) to (11), and the differences between the actual values and the estimated values, denoted as errors (ϵ) were calculated and plotted in a graph.

Gumbel Distribution:

The parameters (μ, σ) chosen were (0, 1), (0, 5), and (5, 1). Figure 3 shows the plot of the errors (ϵ) of the location parameter (μ) and scale parameter (σ). For small sample size ($n = 5$), the errors of both μ and σ are large for all values of μ and σ. As the sample size increases, the errors decrease for all values of μ and σ. A change in μ generally does not affect the change in errors of both μ and σ. However, a change in σ causes a proportional change in the errors of both μ and σ.

GEV Distribution:

In this case, the parameters μ and σ were fixed as 0 and 1 respectively, while the value of shape parameter (k) were selected from the range -0.4 $\leq k \leq$ 0.4 with step 0.1. Figure 4 shows the plot of errors (ϵ) of the shape parameter (k). For small sample size ($n = 5$), the error are large for all values of k. As the sample size increases, the error decreases. For small values of k (≤ -0.3), the estimator tends to give an estimate which has a wrong sign, especially for small sample size. For $-0.2 \leq k \leq 0.1$, the estimator performs satisfactorily for large sample size, whereas for large values of k (≥ 0.2), there are errors in the estimates regardless of the sample size.

Conclusion:

Summing up the performance of moment estimators as suggested by Figure 3 and 4, we conclude that moment estimators for Gumbel distribution does not perform well for all values of μ and σ in small sample size and for large σ in any sample size. For large sample size, they perform satisfactorily for all values of μ and small σ. On the other hand, the moment estimator for GEV distribution does not perform well for all values of k in all sample sizes, except for k close to 0 in large sample size. Generally, the moment estimators do not perform well in estimating the parameter values. Therefore, we do not suggest the use of method of moments when computing the parameter estimates of a Gumbel or GEV distribution.

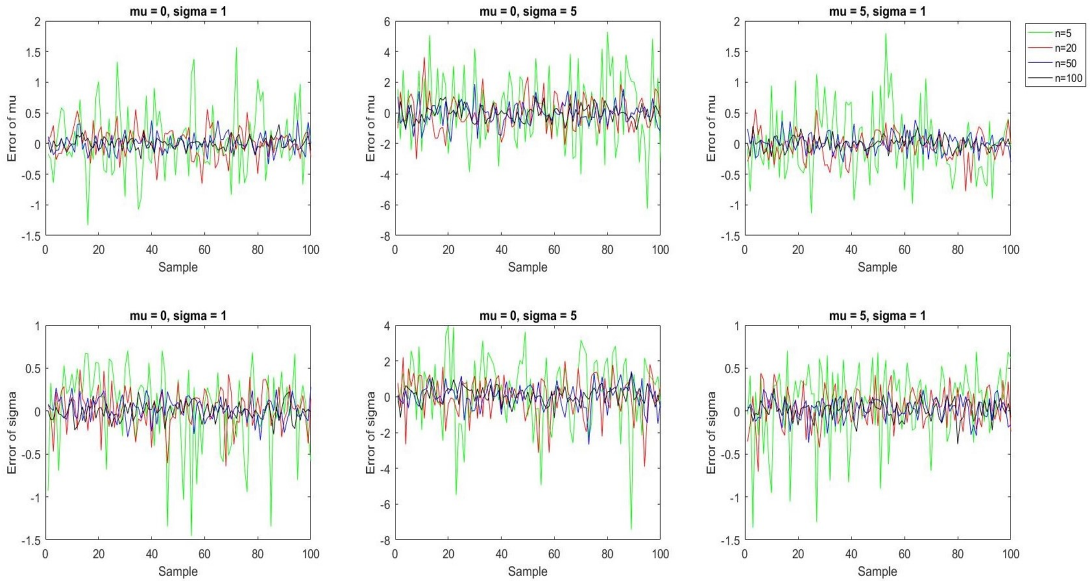

Figure 3: Plot of Errors of Location and Scale Parameters, MoM

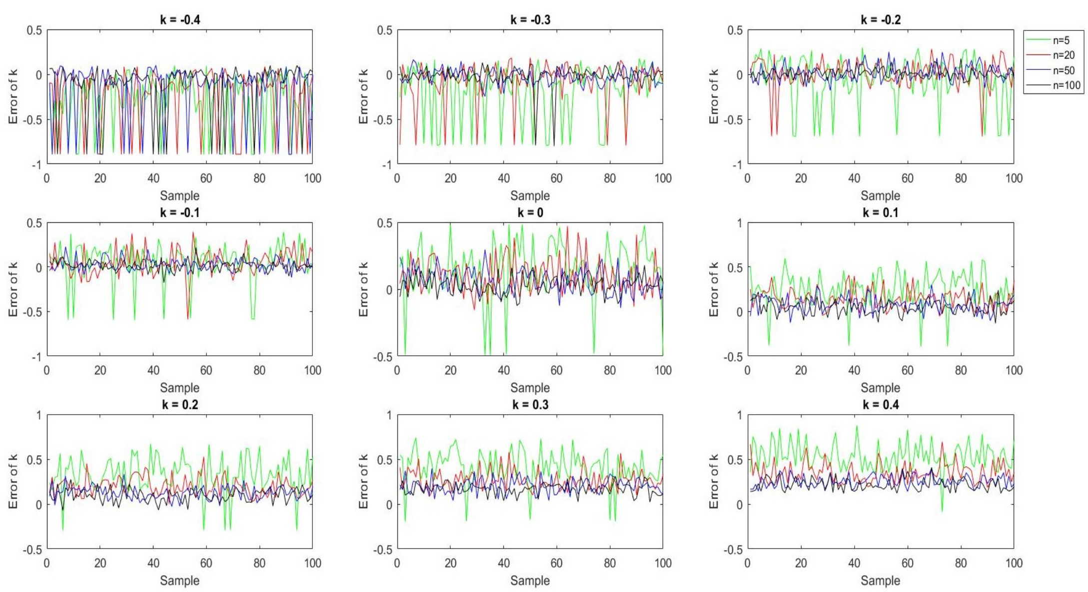

Figure 4: Plot of Errors of Shape Parameter, MoM

4.2 Maximum Likelihood Estimators (MLEs)

4.2.1 MLE Estimation for the Gumbel Distribution

The CDF of the Gumbel distribution is given by $F(X) = \exp(\exp(-(X-\mu)/\sigma)))$, where μ and σ are the location and scale parameters respectively, as usual. Differentiating we can obtain the probability density function:

$$f(X) = \frac{1}{\sigma}\exp\left(-\frac{X-\mu}{\sigma} - \exp\left(-\frac{X-\mu}{\sigma}\right)\right).$$

For a random sample $X_1, \ldots, X_n$ from the previous PDF, the likelihood function is given by

$$\begin{aligned} L(\mu,\sigma) &= \prod_{i=1}^{n} f(x), \\ &= \sigma^{-n}\exp\left[-\sum_{i=1}^{n}\frac{x_i-\mu}{\sigma} - \sum_{i=1}^{n}\exp\left(-\frac{x_i-\mu}{\sigma}\right)\right]. \end{aligned}$$

Setting $\partial(\log L)/\partial\mu = 0$ and $\partial(\log L)/\partial\sigma = 0$ at $\mu = \hat{\mu}$ and $\sigma = \hat{\sigma}$, we obtain after some algebra the equations for the MLEs (Castillo, 1988):

$$\hat{\sigma} + \frac{\sum_{i=1}^{n} x_i \exp\left(-x_i/\hat{\sigma}\right)}{\sum_{i=1}^{n}\exp\left(-x_i/\hat{\sigma}\right)} = \bar{x}, \tag{12}$$

$$\hat{\mu} = \hat{\sigma}\log\left[\frac{1}{n}\sum_{i=1}^{n}\exp\left(-\frac{x_i}{\hat{\sigma}}\right)\right]. \tag{13}$$

The MLEs $\hat{\sigma}$ and $\hat{\mu}$ can be obtained numerically (Mahdi & Cenac, 2004).

Remark. $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

4.2.2 MLE Estimation for the GEV Distribution

The CDF of the General Extreme Value distribution is given by $F(x) = \exp(-(1-k(x-\mu)/\sigma)^{(1/k)})$ if $k \neq 0$ and $F(x) = \exp(-\exp(-(x-\mu)/\sigma))$ if $k = 0$.

If the set $\{X_i\}$ are i.i.d from a GEV distribution, then the log-likelihood function for a sample of n observations $\{x_i\}$ is

$$\begin{aligned} l(\theta|x) &= \log[L(\theta|x)], \\ &= -n\log(\sigma) + \sum_{i=1}^{n}\left[\left(\frac{1}{k}-1\right)\log(y_i) - (y_i)^{1/k}\right], \end{aligned}$$

where $\theta = (\mu,\sigma,k)$ and $y_i = 1-k(x-\mu)/\sigma$.

The MLE of μ, σ, k is obtained by solving the system of equations arising from the equations:

$$\frac{\partial l}{\partial \mu} = 0, \quad \frac{\partial l}{\partial \sigma} = 0, \quad \frac{\partial l}{\partial k} = 0.$$

Solving gives the following equations: (Stedinger & Martins, 2000)

$$\frac{1}{\sigma}\sum_{i=1}^{S}\left[\frac{1-k-(y_i)^{1/k}}{y_i}\right] = 0, \tag{14}$$

$$-\frac{S}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{S}\left[\frac{1-k-(y_i)^{1/k}}{y_i}\left(\frac{x_i-\mu}{\sigma}\right)\right] = 0, \tag{15}$$

$$-\frac{1}{k^2}\sum_{i=1}^{S}\left[\log(y_i)\left(1-k-(y_i)^{1/k}\right) + \frac{1-k-(y_i)^{1/k}}{y_i}k\left(\frac{x_i-\mu}{\sigma}\right)\right] = 0. \tag{16}$$

The equations above can be solved numerically and by an iterative method (Newton-Raphson method). It can be noted that the MLE equations for the GEV are significantly more complex than the MLE equations for the Gumbel.

4.2.3 Working with MLEs

In practice, when evaluating MLEs of Gumbel and GEV distributions, we use the *evfit(X)* and *gevfit(X)* functions in MATLAB, where X is the data.

As seen in the MATLAB documentation: ‘parmhat = evfit(data) returns maximum likelihood estimates of the parameters of the type 1 extreme value distribution given the data in the vector data. parmhat(1) is the location parameter, mu, and parmhat(2) is the scale parameter, sigma.’ and ‘parmhat = gevfit(X) returns maximum likelihood estimates of the parameters for the generalized extreme value (GEV) distribution given the data in X. parmhat(1) is the shape parameter, k, parmhat(2) is the scale parameter, sigma, and parmhat(3) is the location parameter, mu.’ (MathWorks, 2018a and 2018b)

4.3 Probability-Weighted Moments (PWM)

Probability-weighted moments are a generalised version of standard moments of a probability distribution. According to Hosking et al. (1985), they were introduced by Greenwood et al. (1979). PWM have desirable properties which will be discussed further and they provide convenient parameter estimation for a variety of distributions including the logistic, Weibull and Gumbel - and are especially useful for estimation with a small sample.

4.3.1 Introduction to PWM

Suppose we have a random variable X with distribution function F. Probability-weighted moments are given by:

$$M_{p,r,s} = E[X^p\{F(X)\}^r\{1 - F(X)\}^s]. \tag{17}$$

Here we have $p, r, s \in \mathbb{R}$.

If the inverse distribution function $x(F)$ can be written in closed form it may be more convenient to evaluate these moments using the following:

$$M_{p,r,s} = \int_0^1 \{x(F)\}^p F^r (1 - F)^s dF. \tag{18}$$

Notice that $M_{p,0,0}$ are the standard moments of X.

Often it is useful to set $p = 1$ as this allows for a simpler calculation and when r, s are integers it is useful to either consider $M_{1,r,0}$ or $M_{1,0,s}$. For the Generalised Extreme-Value and Gumbel distributions with a sample of size n, it is only necessary to consider the following:

$$\beta_r = M_{1,r,0} = E[X\{F(X)\}^r], \qquad r \in \{0, 1, 2, \dots n\},$$

as suggested by Hosking et al. (1985). Note here that although we can have $p, r, s \in \mathbb{R}$, it is difficult to find estimators for $M_{p,r,s}$ - using β_r is preferable since it has simpler estimators.

Suppose we have a random sample of size n from F, we may choose to use an ordered sample $X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq \dots \leq X_{(n)}$ for estimation of β_r. The statistic:

$$b_r = \frac{1}{n} \sum_{i=1}^{n} \frac{(i-1)(i-2)\dots(i-r)}{(n-1)(n-2)\dots(n-r)} x_{(i)},$$

is unbiased for β_r (Greenwood et al. 1979). Alternatively one can estimate using the following:

$$\hat{\beta}_r = \frac{1}{n} \sum_{i=1}^{n} p_{i,n}^r \; x_{(i)},$$

where $p_{i,n}$ has to be reasonably chosen. Taking $p_{i,n} = (i - a)/n,\ 0 < a < 1$ gives a consistent estimator $\hat{\beta}_r$ of β_r. (Hosking et al. 1985)

4.3.2 PWM Estimation for the Gumbel Distribution

We have already seen the Gumbel distribution, which is given by equation (1). The Gumbel has a closed form inverse distribution function. This allows for an easier calculation of β_r. The inverse distribution function $x(F)$ is given by

$$x(F) = \mu - \sigma \left[\log(-\log F)\right]. \tag{19}$$

This is used for the the computation of the rth order PWM β_r

$$\beta_r = \int_0^1 x(F) F^r dF = \frac{1}{r+1}[\mu + \sigma\{\gamma + \log(r+1)\}]. \tag{20}$$

Here γ is the Euler-Mascheroni constant. It is defined by:

$$\gamma = \lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k} - \log n\right) = 0.5772\ldots\ .$$

The result (20) arises from substitutions $u = -\log f$ and $m = (r+1)u$.

To determine μ, σ another equation is required. This can be done simply by defining $\beta_q, \quad q \neq r$ in the same way. The result is two relatively straightforward equations which can be manipulated to give:

$$\sigma = \frac{(r+1)\beta_r - (q+1)\beta_q}{\log(r+1) - \log(q+1)}, \tag{21}$$

$$\mu = (r+1)\beta_r - \sigma[\log(r+1) + \gamma]. \tag{22}$$

For simplicity, let $(r, q) = (0, 1)$ and replace β_r with an estimator. Here we replace with the consistent estimator $\hat{\beta}_r$ where $p_{i,n} = (i - 0.35)/n$ as suggested by Hosking et al. (1985) for estimation of the Generalised-Extreme Value distribution. In their paper for parameter estimation of the Gumbel distribution - (Mahdi & Cenac, 2003) suggest using the unbiased estimator b_r. Nonetheless we follow the suggestion by Hosking since later we use the consistent estimator for parameter estimation of the GEV.

Thus the following estimators are obtained:

$$\hat{\sigma} = \frac{2\hat{\beta}_1 - \hat{\beta}_0}{\log 2}, \tag{23}$$

$$\hat{\mu} = \hat{\beta}_0 - \hat{\sigma}\gamma. \tag{24}$$

Here (23) and (24) define the classical PWM estimators for the Gumbel distribution (Rasmussen & Gautam, 2003).

Remark. *When implementing equation* (24) *it was necessary to change the equation for the estimator to work. The equation used was* $\hat{\mu} = \hat{\beta}_0 + \hat{\sigma}\gamma$.

4.3.3 PWM Estimation for the GEV Distribution

The CDF of a GEV distribution is given by equation (4). Importantly, for the calculations, the GEV distribution has a closed form inverse distribution function:

$$\begin{aligned} x(F) &= \mu + \sigma\left[\frac{1-(-\log F)^k}{k}\right], & k \neq 0 \\ &= \mu - \sigma\left[\log(-\log F)\right], & k = 0 \end{aligned} \tag{25}$$

Now the calculation of β_r:

$$\beta_r = M_{1,r,0},$$

$$= \int_0^1 \left[\mu + \sigma \left\{\frac{1-(-\log F)^k}{k}\right\}\right] F^r \; dF,$$

Let $u = -\log F$,

$$= \int_0^\infty \left[\mu + \sigma \left\{\frac{1-u^k}{k}\right\}\right] e^{-(r+1)u} \; du,$$

$$= \left(\mu + \frac{\sigma}{k}\right) \int_0^\infty e^{-(r+1)u} \; du - \frac{\sigma}{k} \int_0^\infty u^k e^{-(r+1)u} \; du,$$

$$= \left(\mu + \frac{\sigma}{k}\right) (r+1)^{-1} - \frac{\sigma}{k} (r+1)^{-1-k} \; \Gamma(1+k),$$

$$\Rightarrow \beta_r = (r+1)^{-1} \left[\mu + \frac{\sigma\{1-(r+1)^{-k} \; \Gamma(1+k)\}}{k}\right], \quad \text{if } k > -1. \tag{26}$$

If $k \leq -1$, β_r does not exist.

Now that (26) gives us the equation of β_r, it is necessary to manipulate this into yielding values for μ, σ, k. Using (26) the following equations hold:

$$\beta_0 = \mu + \frac{\sigma\{1-\Gamma(1+k)\}}{k}, \tag{27}$$

$$2\beta_1 - \beta_0 = \frac{\sigma\Gamma(1+k)(1-2^{-k})}{k}, \tag{28}$$

$$\frac{3\beta_2 - \beta_0}{2\beta_1 - \beta_0} = \frac{1-3^{-k}}{1-2^{-k}}. \tag{29}$$

Note here that the probability-weighted moments used are $r = 0, 1, 2$. We do not want the higher deviation from using bigger r values. Also using the lowest values to get three equations for three unknowns allows for a simpler solution.

To yield the estimators $\hat{\mu}, \hat{\sigma}, \hat{k}$, the quantity β_r must be replaced by an estimator, either $\hat{\beta}_r$ or b_r. Hosking et al. (1985) suggests using the consistent estimator $\hat{\beta}_r$ where $p_i, n = (i - 0.35)/n$ for the best performance. Using the unbiased estimator b_r is not always ideal since it has a higher variance.

$$\frac{3\hat{\beta}_2 - \hat{\beta}_0}{2\hat{\beta}_1 - \hat{\beta}_0} = \frac{1-3^{-\hat{k}}}{1-2^{-\hat{k}}} \tag{30}$$

To solve (30) for $\hat{k}$ we require iterative methods. Since $(1-3^{-\hat{k}})/(1-2^{-\hat{k}})$ is almost linear over the range $-1/2 < \hat{k} < 1/2$, which are the values most commonly seen in practice; Hosking et al. (1985) suggest that low-order polynomial approximations are accurate enough. They propose the following estimator:

$$\hat{k} = 7.859c + 2.9554c^2, \qquad c = \frac{2\hat{\beta}_1 - \hat{\beta}_0}{3\hat{\beta}_2 - \hat{\beta}_0} - \frac{\log 2}{\log 3}. \tag{31}$$

Then using $\hat{k}$, estimators for σ and μ can be found:

$$\hat{\sigma} = \frac{(2\hat{\beta}_1 - \hat{\beta}_0)\hat{k}}{\Gamma(1+\hat{k})(1-2^{-\hat{k}})}, \tag{32}$$

$$\hat{\mu} = \hat{\beta}_0 + \frac{\hat{\sigma}\{\Gamma(1+\hat{k}) - 1\}}{\hat{k}}. \tag{33}$$

Here the equations (31), (32) and (33) define PWM estimators of the GEV distribution.

Remark. *When implementing equation* (33) *it was necessary to change the equation for the estimator to work. Instead* $-\hat{k}$ *was used in the code.*

4.3.4 Properties and Performance of the PWM Estimator

PWM estimates of the GEV distribution satisfy criteria, that $\hat{k} > 1$ and that $\hat{\sigma} > 0$. These are desirable properties especially since in practice, $-1/2 < k < 1/2$, and it is important to have a positive scale parameter, σ. For the proof see (Hosking et al. 1985).

The PWM and MLE methods discussed in this report are invariant under linear transformation (Hosking et al. 1985) - thus to assess performance it is conventional and convenient to set $\mu = 0$ and $\sigma = 1$ for the GEV and Gumbel distributions and allow the shape k to vary for the GEV distribution.

The PWM has come under criticism since the method assumes *a priori* that $k < 1$, which can be seen by manipulating (26). Further this implies that the distribution has a finite mean. However the GEV distribution is defined $\forall\ k \in \mathbb{R}$. To broaden the domain of the shape k for PWM, a new class of moments called GPWM (generalised probability-weighted moments) has been introduced. This adds a suitable continuous function ω to the definition. Although this massively increases complexity it allows for a wider domain of k for which these moments can be computed.

Gumbel Estimation:

Here we discuss PWM estimator performance for the Gumbel distribution.

To see the performance of PWM estimators and compare them with MoM estimators and MLEs we use data from (Mahdi & Cenac, 2004). They computed empirical estimates from a random sample of different sizes (which they denote n) of the standard Gumbel distribution with $\mu = 0, \sigma = 1$. They denote the scale by α and the location by ϵ. For PWM they used (r, q) = (0, 1) to use the classical estimators and elected to use the unbiased estimator b_r (Greemwood et al. 1979) for β_r. For MoM - they elected to use the 1$^{\text{st}}$ and 2$^{\text{nd}}$ moments.

n	ML	ML	MM(1,2)	MM(1,2)	PWM(0,1)	PWM(0,1)
	α	ϵ	α	ϵ	α	ϵ
5	0.8267	0.0837	0.7925	0.1107	0.9752	0.0053
10	0.9167	0.0398	0.8954	0.0576	0.9918	0.0019
15	0.9457	0.0199	0.9346	0.0319	1.0001	0.0058
20	0.9653	0.0171	0.9656	0.0269	1.007	0.0024
30	0.9805	0.0135	0.9757	0.0201	1.0103	0.0002
50	0.9835	0.0069	0.9801	0.0116	1.0020	0.0010
100	0.9930	0.0058	0.9905	0.0087	1.002	0.0018

Table of empirical estimates of scale and location using the MoM, PWM and ML methods for the standard Gumbel distribution. (Mahdi & Cenac, 2004: p.5)

As can be seen the PWM estimates are far better than MoM and MLE for small to moderate samples. All three methods perform well for large samples. Although we use this for our example - for the Gumbel distribution - using the classical PWM estimators does not provide the best performance. (Mahdi & Cenac, 2004) suggest using low values for (r, q) to avoid over-weighting large sample observations. A method suggested by (Rasmussen & Gautam, 2003) eliminates q by studying the effects of when $q \to r$ using L'Hôpital's Rule, computing an estimator based solely on r and then using an empirical relationship to best choose r. It provides an improvement of approximately 10%.

GEV Estimation:

Here we discuss PWM estimator performance for the Generalised Extreme-Value distribution. To see the performance for the GEV distribution we use data from Hosking et al. (1985). Here the scale and location parameter have been set to 1 and 0 respectively. For comparison they use different values for shape k for the distribution using $k = -0.4, -0.2, 0, 0.2, 0.4$. The use a variety of sample sizes n and they compare three methods: PWM, MLE and Jenkinson's method of sextiles (JS) - which is not discussed here. For the PWM they elected to use the consistent estimator $\hat{\beta}_r$ where $p_{i,n} = (i - 0.35)/n$ which was chosen since it gave the best overall results. In terms of notation scale is denoted α and location is denoted ξ. The bias and standard deviation of the estimators for the GEV parameters are shown:

Table 5. Bias of Estimators of GEV Parameters

		k														
		Bias ($\hat{\xi}$)					Bias ($\hat{\alpha}$)					Bias ($\hat{k}$)				
n	Method	−.4	−.2	.0	.2	.4	−.4	−.2	.0	.2	.4	−.4	−.2	.0	.2	.4
15	PWM	.10	.05	−.02	.00	−.03	.00	−.06	−.10	−.11	−.12	.11	.03	−.03	−.08	−.12
	ML	.03	.03	.05	.05	.04	−.07	−.07	−.07	−.06	−.07	−.04	−.02	.02	.04	.03
	JS	.11	.08	.06	.04	.02	−.05	−.06	−.07	−.08	−.08	.10	.06	.03	.01	−.01
25	PWM	.06	.03	.01	−.01	−.02	.00	−.04	−.06	−.07	−.07	.08	.02	−.02	−.05	−.07
	ML	.01	.02	.03	.03	.04	−.04	−.04	−.04	−.03	−.03	−.02	−.01	.02	.04	.05
	JS	.06	.04	.03	.02	.01	−.03	−.03	−.04	−.04	−.05	.07	.04	.02	.01	−.00
50	PWM	.04	.02	.01	.00	−.01	.01	−.02	−.03	−.04	−.04	.05	.02	−.01	−.02	−.04
	ML	.01	.01	.02	.02	.02	−.02	−.02	−.02	−.02	−.01	−.01	.00	.01	.02	.03
	JS	.04	.02	.02	.01	.01	−.02	−.02	−.02	−.02	−.02	.04	.02	.01	.00	.00
100	PWM	.02	.01	.00	.00	−.01	.00	−.01	−.02	−.02	−.02	.03	.01	.00	−.01	−.02
	ML	.00	.00	.01	.01	.01	−.01	−.01	−.01	−.01	.00	−.01	.00	.00	.01	.02
	JS	.02	.01	.01	.00	.00	−.01	−.01	−.01	−.01	−.01	.02	.01	.00	.00	.00

NOTE: GEV—Generalized extreme value; PWM—Probability-weighted moment; ML—Maximum likelihood; JS—Jenkinson's (1969) sextiles.

Table showing the bias of estimators for the GEV parameters for different values of k from a GEV distribution with $\mu = 0$ and $\sigma = 0$. (Hosking et al. 1985: p.8)

Table 6. Standard Deviation of Estimators of GEV Parameters

		k														
		Standard Deviation ($\hat{\xi}$)					*Standard Deviation* ($\hat{\alpha}$)					*Standard Deviation* ($\hat{k}$)				
n	*Method*	−.4	−.2	.0	.2	.4	−.4	−.2	.0	.2	.4	−.4	−.2	.0	.2	.4
15	PWM	.32	.30	.29	.28	.28	.33	.25	.21	.19	.19	.20	.19	.18	.18	.19
	ML	.32	.32	.31	.30	.28	.28	.25	.23	.22	.21	.36	.32	.29	.27	.23
	JS	.33	.31	.30	.29	.28	.32	.28	.24	.21	.21	.25	.24	.23	.23	.23
25	PWM	.24	.23	.22	.22	.22	.24	.19	.17	.15	.16	.18	.16	.14	.14	.15
	ML	.24	.24	.23	.23	.22	.21	.19	.17	.16	.17	.24	.21	.20	.18	.17
	JS	.24	.23	.23	.22	.22	.24	.21	.18	.17	.16	.19	.18	.17	.17	.17
50	PWM	.17	.16	.16	.16	.16	.17	.14	.12	.11	.11	.14	.12	.11	.10	.11
	ML	.17	.16	.16	.16	.16	.15	.13	.12	.11	.11	.15	.13	.12	.11	.11
	JS	.17	.16	.16	.16	.16	.17	.14	.13	.12	.11	.14	.13	.12	.11	.12
100	PWM	.12	.12	.11	.11	.11	.12	.10	.09	.08	.08	.11	.09	.07	.07	.08
	ML	.12	.11	.11	.11	.11	.10	.09	.08	.08	.08	.10	.09	.08	.07	.07
	JS	.12	.12	.11	.11	.11	.12	.10	.09	.08	.08	.10	.09	.08	.08	.08

NOTE: GEV—Generalized extreme value; PWM—Probability-weighted moment; ML—Maximum likelihood; JS—Jenkinson's (1969) sextiles.

Table showing the standard deviation of estimators for the GEV parameters for different values of k from a GEV distribution with $\mu = 0$ and $\sigma = 0$. (Hosking et al. 1985: p.8)

In general, it can be seen that the PWM estimator for k has a larger bias than the ML estimator, however this is small near the important value when $k = 0$. Where the advantage of PWM lies is in the low standard deviation particularly for small samples, $n = 15$ and $n = 25$. The bias is relatively insignificant compared to the standard deviation when calculating the mean square error of $\hat{k}$.

With regards to estimators for μ and σ similar results can be seen. Generally, PWM estimators have smaller standard deviation especially for small samples and their bias, although larger than ML (which have the lowest bias in this example), is not enough to significantly impact mean square error. (Hosking et al. 1985)

Conclusion:

PWM offers a fast and straightforward way of computing feasible estimated parameters of both the GEV and Gumbel distributions. They perform better than MLEs for small samples and although tend to have a higher bias, these decrease quickly as the sample size increases. (Hosking et al. 1985). The most useful property of PWM estimators is its low standard deviation which is comparable with MLEs for large samples and far smaller for small samples. This allows for a smaller MSE (mean squared error).

5 Goodness of Fit

5.1 Quantile-Quantile Plots

Construction:

A Quantile-Quantile plot (QQ plot) is a graphical method to determine whether two data sets are from the same distribution, with an emphasis on fitting the tails. As the goodness of fit is tested for an Extreme Value distribution, order statistics and theoretical quantiles will used as the two data sets. No assumptions are made on the distribution of the sample data. Gumbel and Generalised Extreme Value (GEV) distribution will be compared to the data of the maxima by the following method: (Scott, 2018)

Let $Z_1, \ldots, Z_M$ denote M observations of the maximum value:

1. **Ordered Statistics, $z_{(i)}$:**
   The random sample, $z_1, z_2, \ldots, z_M$, is used to create the order statistics so:
   $z_{(1)} < z_{(2)} < \ldots < z_{(M-1)} < z_{(M)}$,
   These M ordered values will form the input sample quantiles, $z_{(i)}$ so:
   $$F(z_{(i)}) \approx i/M, \text{ where}$$
   $z_{(i)}$ - order statistic, i/M - probability associated with each $z_{(i)}$, F - empirical distribution.

2. **Theoretical Quantiles, y_i:**
   The distribution curve, either Gumbel or GEV distribution, will be divided M times into M+1 quantiles, each with equal probability.
   $$\hat{F}(y_i) = p_i$$
   y_i - theoretical quantiles, p_i - probability associated with each y_i,
   $\hat{F}$ - Gumbel or GEV distribution with estimated parameters.

   p_i is set as $(i - 1/2)/M$ to avoid discrepancies: suppose if $p_i = i/M$ and $i = M$, then $\hat{F}(y_M) = 1$.

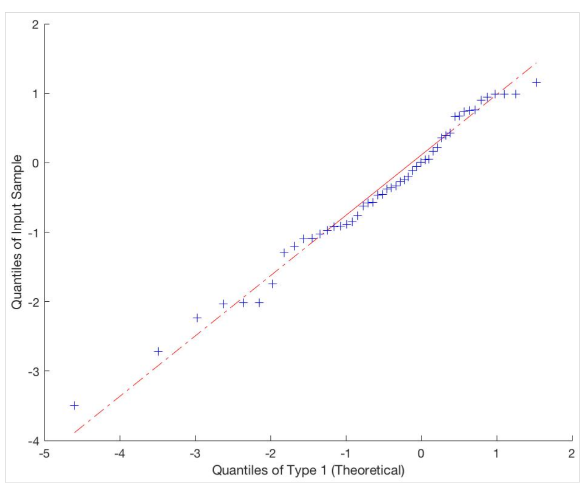

Figure 5: Example of a QQ Plot: Random Sample of 50 from Gumbel distribution

Interpretation of QQ Plots:

A QQ plot has a reference line plotted; the data points needs to be approximately on this line, so that the data sets come from the same distribution. If y_i and $z_{(i)}$ are identically distributed, the QQ plot will be a 45° straight line, with gradient 1, passing through the origin.

This straight line may vary in gradient and y-axis intercept for various data sets, given y_i is a linear function of $z_{(i)}$. This occurs when the data is not standardised, so an alternative straight line is plotted, $z_{(i)} = (y_i - \mu)/\sigma$, where μ is the estimated location parameter and σ is the estimated scale parameter. (Wilk & Gnanadesikan, 1968)

QQ plots may indicate departure from the theoretical distribution. This occurs if the data is skewed or has large kurtosis. (Scott, 2018)

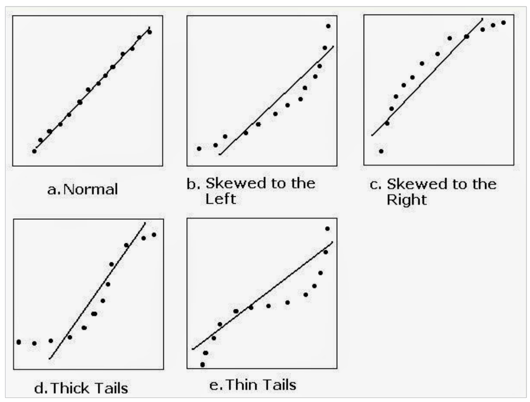

Figure 6: Departures in QQ Plots (Jost, 2018)

Simulating Samples from Gumbel and Generalised Extreme Value Distribution:

The sample quantiles are a random sample from the Extreme Value distribution (Gumbel or GEV). The theoretical quantiles are formed from the same distribution. The deviation of the generated QQ plots will be observed, thereby determining the usefulness of this test.

QQ plots vary from sample to sample - the plots below are of size 100, which is similar to the sample size used in the worked example later.

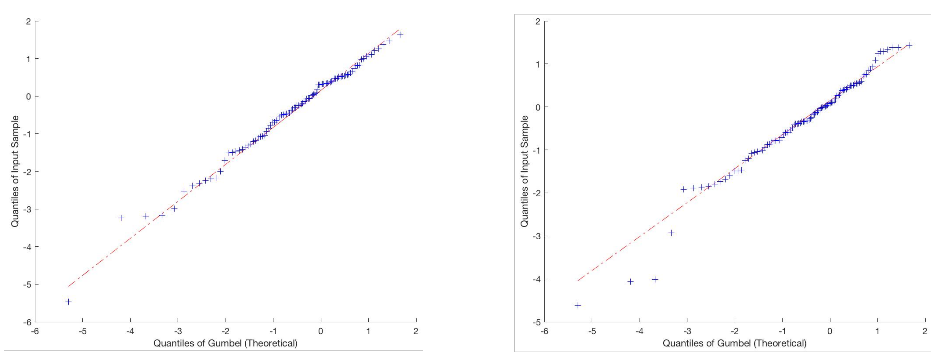

Figure 7: Gumbel QQ Plots

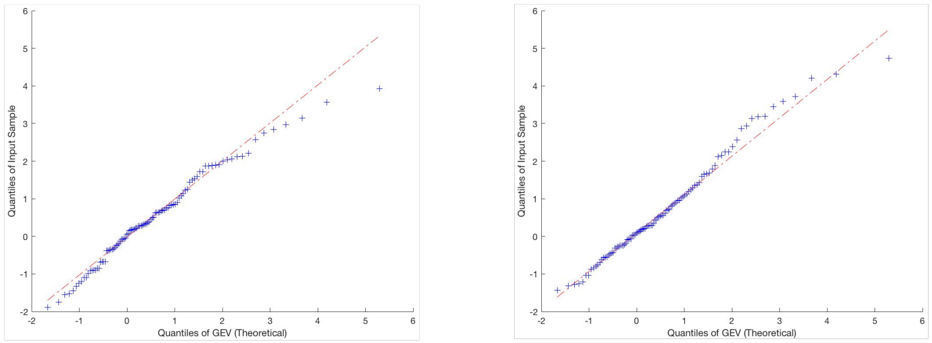

Figure 8: GEV QQ Plots

Gumbel and GEV are sampled using *gevrnd* and *evrnd* respectively on MATLAB, with parameters: $\sigma = 1$, $\mu = 0$ and $k = 0$ (for *gevrnd* only). The points will approximately lie on a 45° straight line if the input sample matches the distribution. In order to not reject a distribution, the points need to be randomly scattered, not having kurtosis or a skew. Outliers are easy to spot in QQ plots.

In both the plots of Figure 7, the majority of the data points fit $y = x$, but the first 3 sample quantiles appear to not fit the Gumbel distribution. These can be taken as outliers as there appears to have no skew or large kurtosis, hence the input sample follows the Gumbel distribution.

The first plot of Figure 8 is skewed slightly to the right, thereby suggesting the sample size of 100 does not fit the GEV distribution. However, the data points aside from the last 4 sample quantiles (outliers) lie closely to the $y = x$ reference line. The other QQ plots approximately lie on $y = x$, so follow the given distribution.

The variations observed in the QQ plots are not significant enough to rule out the distribution that the input sample is compared to. Therefore, a QQ plot is effective in determining whether the input sample follows a certain distribution.

5.2 Parametric Bootstrap

Parametric bootstrap is another goodness of fit test for distributions with estimated parameters, which relies on sample with replacement. In a general sense, it finds properties of an estimator.

Define: (Babu, 2011)

$\{F(\,\cdot\,;\theta) : \theta \in \Theta\}$ - a family of continuous distributions,

Θ - parameter space, an open region in p-dimensional space,

F^* - empirical distribution,

$x_1, \ldots, x_n$ - sample data of size n,

$x_1^*, \ldots, x_n^*$ - resample of size n,

u - statistic computed from the sample,

u^* - statistic computed from the resample.

The Bootstrap principle says (Orloff & Bloom, 2014):

- $F^* \approx F$.
- The variation of u can be predicted by variation of u^*.

Construction:

Parametric bootstrap involves generating a bootstrap sample to estimate a confidence interval for an unknown parameter, θ through the bootstrap difference.

The method is as follows: (Orloff & Bloom, 2014)

1. $x_1, \ldots, x_n$ is a data sample for $F(\theta)$.
2. Statistic, $\hat{\theta}$, is used to estimate θ through parameter estimation methods, which was covered in Section 4.
3. $x_1^*, \ldots, x_n^*$ is a resample of the data with replacement from the empirical distribution, F^*. The size of the sample and resamples are the same so that the variation of the statistic, $\hat{\theta}$ is unaffected.
4. $\hat{\theta}^*$ is computed in the same manner as 2, and the bootstrap difference, δ^* is also calculated, where $\delta^* = \hat{\theta}^* - \hat{\theta}$.
5. The distribution of δ ($\delta = \hat{\theta} - \theta$) is well-approximated by the distribution of δ^*, which is roughly based on the Law of Large Numbers.
6. A $1 - \alpha$ bootstrap confidence interval for θ is constructed by using the confidence limits as $\alpha/2$ and $1 - \alpha/2$ quantiles of the bootstrap difference distribution. This confidence interval for θ can be used to reject the estimated parameter by a hypothesis test. (Geyer, 2012)

6 Worked Example, Earthquakes in Greece

6.1 Introduction

In our worked example, we will look into the dataset of the magnitude (Richter Scale) of all the earthquakes that have occurred in Greece from 1901 to 2017 and we consider the maximum magnitude for each year. (Institute of Geodynamics, National Observatory of Athens, 2018 as cited by Stefopoulos, 2018) By using three different kinds of estimations (MLE, MoM and PWM), we will attempt to fit a GEV or Gumbel distribution to the data. With the parameters estimated, we then create QQ plots and find bootstrap confidence intervals to compare each of the methods and distributions to decide which model gives a better estimate. At the end of our worked example, we will plot the return level against the return period to give a prediction of the maximum magnitude of future earthquakes in Greece over a certain time period.

Major earthquakes have severe social, economic and environmental consequences - there are predicted to be 16 major earthquakes occurring worldwide every year: 15 of magnitude 7.0 - 7.9 and 1 of magnitude greater than 8. (U.S. Geological Survey, 2018) Therefore, it is worth accessing the data of earthquakes and making some predictions on the frequency and number of earthquakes in a certain time period in Greece.

6.2 Parameter Estimates

Methods / Parameters	MoM		MLE		PWM	
	GEV	Gumbel	GEV	Gumbel	GEV	Gumbel
Location, μ	5.945	6.478	5.939	6.519	5.927	6.503
Scale, σ	0.566	0.484	0.547	0.657	0.583	0.526
Shape, k	-0.147	N/A	-0.122	N/A	-0.124	N/A

Table 1: Parameter Estimates for data in Greece

6.3 Analysis of Results

Quantile-Quantile Plots:

The maxima of the Greek Earthquake data from 1901-2017 are used as the order statistics, $z_{(i)}$ so the quantiles of the input sample. The order statistics are transformed by $(z_{(i)} - \mu)/\sigma$, so that the data points would approximately fall on $y = x$ if the data follows the given distribution. The transformed order statistics are plotted against the theoretical quantiles, y_i (Gumbel distribution in the first plot and GEV distribution in the second plot of each Figure).

According to the first QQ plots in Figures 9, 10 and 11 the data is skewed to the right, as the data points have a curved pattern with the slope increasing from left to right. Therefore, we can conclude the data of Greek earthquakes does not fit the Gumbel distribution when the parameters are estimated by MoM, MLE and PWM.

In the second QQ plots in the Figures 9, 10 and 11, the data points roughly fall on $y = x$, aside from the 6 last quantiles which are taken as outliers. The QQ plots here are similar to the QQ plots of the random simulated samples of 100 from the Gumbel distribution and GEV distribution in

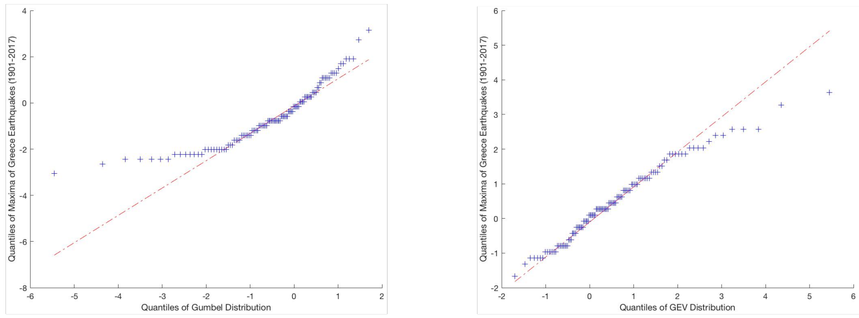

Figure 9: Gumbel and GEV QQ Plots for MoM

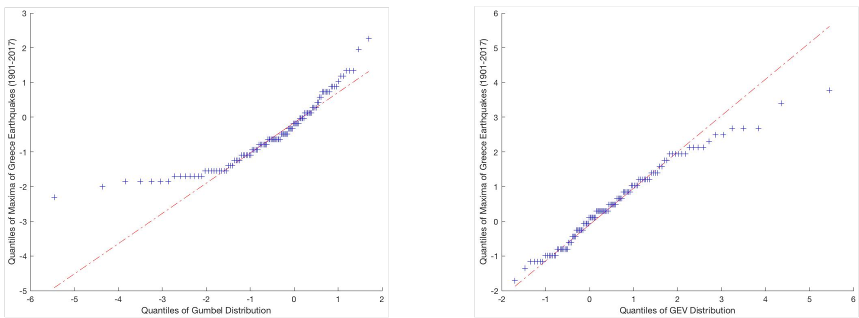

Figure 10: Gumbel and GEV QQ Plots for MLE

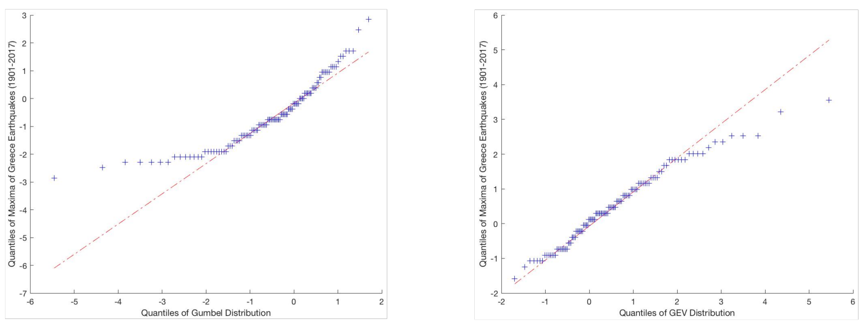

Figure 11: Gumbel and GEV QQ Plots for PWM

Section 5, hence we can conclude the maxima of Greek earthquakes follows the GEV distribution. They also all have a staircase pattern, which occurs as the earthquake data is rounded to 1 decimal place.

To decide which the parameter estimations with MoM, MLE and PWM is the best, the estimation methods will be compared as the QQ plots look very similar. MLE is harder to compute in 3 parameter case, due to the iterative nature of the method, and MoM deviates with higher moments. PWM assumes k between -1 and 1 but in practice k is between -1/2 and 1/2. This is not an issue considering the estimated shape parameter, k for Gumbel and GEV distribution, which is given in Table 1. Therefore, we can conclude PWM is a better parameter estimation method for the data of Greek Earthquakes (1901-2017).

Comparison between Parameter Estimation Methods:

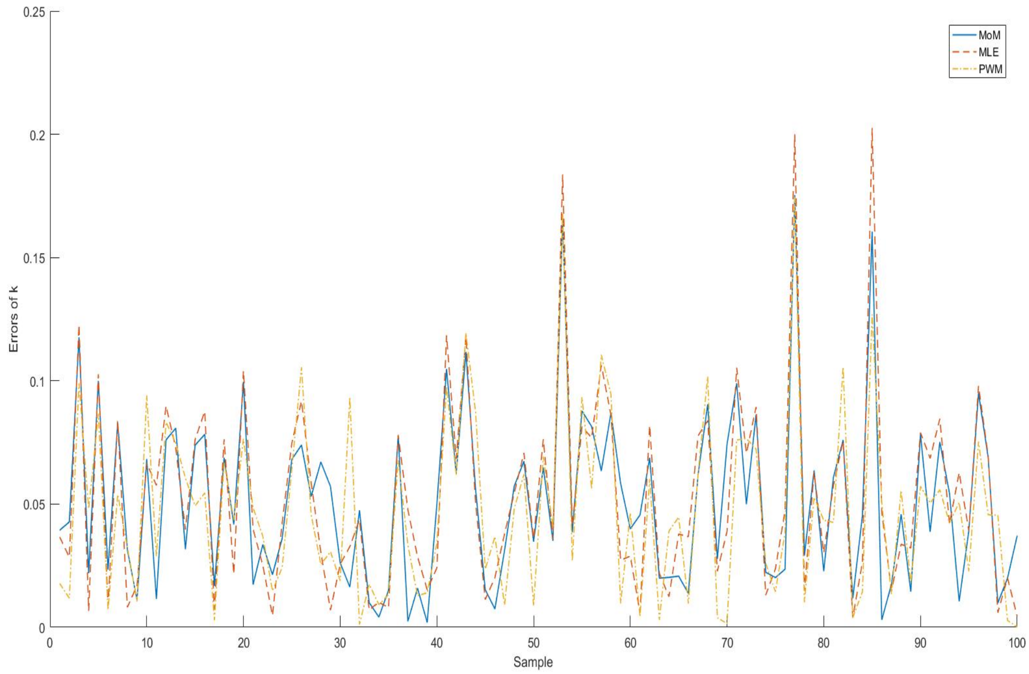

Figure 12: Plot of Absolute Value of Errors of Shape Parameter

To make a further comparison among the three methods of parameter estimation, we generated 100 random sets of data with sample size, $n = 117$, shape parameter, $k = -0.1$, scale parameter, $\sigma = 0.56$, and location parameter, $\mu = 5.94$ using the *gevrnd* function in MATLAB. The estimates for k using three different methods were computed and the absolute value of errors were plotted as shown in Figure 12.

Figure 12 suggests that the errors of the estimates are pretty much the same for all three methods of estimation in most of the sets of data with the same sample size, and similar parameters as the Greek earthquakes dataset that we used. This implies that all three methods agree quite closely on the shape parameter. The mean of the errors of MoM, MLE, and PWM are 0.051, 0.053, and 0.048 respectively. The mean suggests that there is an average error of 0.05 in all three methods of estimation with PWM slightly outperforms both MoM and MLE. Hence, one would expect the true value of k lies between ± 0.05 of the parameter estimate when estimating using these methods.

When studying the parameter estimates from the different methods we see a few interesting facts emerge. When trying to fit a GEV distribution we observe that the PWM and ML estimates are very close together with the shape and location parameters being the same to 2 significant figures. There is only a difference of 0.04 with the scale parameter. This suggests that for this fairly large sample ($n = 117$) both the methods of probability-weighted moments and maximum likelihood perform similarly. Despite the Method of Moments being an old and simplistic method, the estimates are very similar to the PWM and ML estimates. All three estimates agree closely on the location parameter and are similar on the scale and shape parameter with the MoM estimates disagreeing with the other estimates. Nonetheless this is to be expected as, in general, the Method of Moments should only be used as an initial estimate of the parameters. All estimates agree that the shape parameter is negative which suggests that the data follows a Weibull distribution.

When fitting a Gumbel distribution to the data we see some more interesting facts. All three estimates agree quite closely on the location parameter which is higher than that of the GEV estimates. However all three estimates disagree on the scale parameter. With significant disagreement between the ML and MoM estimates. The PWM and MoM also disagree on the scale parameter for the Gumbel distribution but to a much lesser extent. However this is to be expected, as the method of Probability-Weighted Moments builds on the idea of the more simplistic and older Method of Moments.

Most importantly of all, the three estimation methods generally agree on the GEV distribution and in general disagree on the Gumbel distribution. This provides significant evidence that, in fact, the Greece earthquake data follows a GEV distribution rather than a Gumbel distribution.

When looking at the QQ plots and estimates for our Greece earthquake data it seems to suggest that the GEV distribution seems to fit the data better than a Gumbel. Nonetheless here we perform a hypothesis test based on the PWM estimator to help verify this.

We assume that the data follows the GEV distribution with a null hypothesis that the data follows a Gumbel distribution (equivalent to assuming that the shape parameter $k = 0$). We perform a 1-tailed test against the alternative hypothesis that the data follows a Weibull distribution ($k < 0$) to a 5% significance level.

$$H_0 : k = 0, \qquad H_1 : k < 0.$$

It can be shown that the PWM estimator $\hat{k}$ is asymptotically distributed $\mathrm{N}(0, 0.5633/n)$ (Hosking et al. 1985). For our data we have $n = 117$ and the PWM estimate $\hat{k} = -0.124316$. Thus we compare the statistic

$$Z = \frac{\hat{k}}{\sqrt{\frac{0.5633}{n}}} = -1.791632,$$

against the critical value $Z_c = -1.645$. Since $Z < Z_c$ we reject H_0 suggesting there is insufficient evidence to suggest that the data follows a Gumbel distribution to a 5% significance level.

Bootstrap Confidence Intervals:

Bootstrapping consists of resampling the maxima of Greek earthquakes (1901-2017) with replacement. To find the bootstrap confidence interval for each parameter under a certain parameter estimation method, the parameter (say θ) is worked out from the initial data, $\hat{\theta}$. The data is then re-sampled 2000 times and the parameter, $\hat{\theta^*}$ is estimated for each sample under MoM, MLE or PWM. Then the bootstrap difference is calculated, $\delta^* = \hat{\theta^*} - \hat{\theta}$. To find the 95% bootstrap confidence interval ($\alpha = 0.05$) of θ, $\hat{\theta} - (\frac{\alpha}{2}$ quantile of $\delta^*)$ and $\hat{\theta} - (1 - \frac{\alpha}{2}$ quantile of $\delta^*)$ are calculated, which give CI2 and CI1 respectively. This process is used to find the bootstrap confidence limits, CI1 and CI2 for a parameter (in this case μ, σ and k) under each parameter estimation method (MoM, MLE and PWM).

Observing the results on Table 2, 3 and 4, none of the confidence intervals are rejected as the parameter estimates for Gumbel distribution and GEV distribution all lie in its given confidence interval, [CI1, CI2]. After incorporating the results from QQ plots, the GEV distribution (Weibull) with PWM best models the maxima of Greece Earthquakes (1901-2017).

MoM / Parameters	Gumbel			GEV		
	Estimates	CI1	CI2	Estimates	CI1	CI2
Location, μ	6.478	6.355	6.603	5.945	5.827	6.051
Scale, σ	0.484	0.430	0.544	0.566	0.495	0.639
Shape, k	N/A	N/A	N/A	-0.147	-0.226	-0.049

Table 2: MoM Confidence Interval, [CI1, CI2] for Gumbel & GEV

MLE / Parameters	Gumbel			GEV		
	Estimates	CI1	CI2	Estimates	CI1	CI2
Location, μ	6.519	6.395	6.645	5.939	5.808	6.053
Scale, σ	0.657	0.578	0.764	0.547	0.477	0.621
Shape, k	N/A	N/A	N/A	-0.122	-0.256	0.036

Table 3: MLE Confidence Interval, [CI1, CI2] for Gumbel & GEV

PWM / Parameters	Gumbel			GEV		
	Estimates	CI1	CI2	Estimates	CI1	CI2
Location, μ	6.503	6.410	6.675	5.927	5.803	6.048
Scale, σ	0.526	0.514	0.661	0.583	0.514	0.661
Shape, k	N/A	N/A	N/A	-0.124	-0.223	-0.012

Table 4: PWM Confidence Interval, [CI1, CI2] for Gumbel & GEV

6.4 Return Period

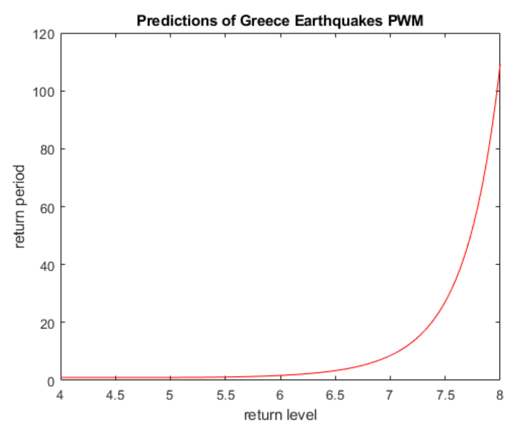

Figure 13: Predictions of Greek Earthquakes (magnitude 4.0 - 8.0)

Figure 13 allows us to predict how often earthquakes of magnitudes 4.0 - 8.0 will occur in Greece on average based on the dataset. We can see from Figure 13 that on average, an earthquake of magnitude 7.5 will occur every 40 years, and an earthquake of magnitude 8 will occur every 120 years. There haven't been any earthquakes of magnitude 9 in Greece since 1901 thus we will not try to predict the average rate of those earthquakes as it would involve extrapolating from the data, rendering the prediction inaccurate. Given these predictions, we recommend that buildings in Greece be built to withstand earthquakes of magnitude 9, so that there is a minimal risk of damage when an earthquake of magnitude 8 occurs.

6.5 Conclusion

In this report, the distribution of Greek earthquakes (1901-2017) was investigated using three parameter estimation methods and goodness of fit. The results suggest that Greek earthquakes follows the GEV distribution with shape parameter, $k \approx -0.12$. In terms of parameters, all three methods perform similarly when estimating μ and σ, with closer agreement in μ than in σ. On the other hand, MLE and PWM outperforms MoM when estimating k, suggesting that both MLE and PWM are better estimators in comparison to MoM. The comparison between estimation methods (in Figure 12) and QQ plots suggest that the estimates of PWM are slightly more accurate than MoM and MLE. Hence, by taking all these results into account, the GEV distribution with PWM gives the best fit of the maxima of Greek earthquakes.

Using this model, the return period analysis suggests that an earthquake of magnitude 8.0 will occur once every 120 years in Greece, which agrees with the data set obtained, where the largest earthquake in Greece over 117 years (1901-2017) was in 1903 with magnitude 8.0. Therefore, we conclude that the methodology used and results obtained in our study are valid and sensible, subject to minor variations dependent on the data set and field.

The GEV distribution has played a vital role in predicting the occurrence of natural catastrophes such as earthquakes and floods over the years. The accuracy of the prediction depends highly on the estimated parameters of the model. Therefore, the process of estimating and justifying the correct parameters is essential to ensure the validity of the model and predictions obtained.

The results in our study suggest that MoM yields unsatisfactory estimates for most of the models in comparison to MLE and PWM. This is reasonable as MoM is the simplest and oldest method for computing the estimates of a distribution. Hence, Hosking (1990) introduced the alternative approach using L-moments, defined as linear combination of order statistics, which provides more accurate estimators of parameters with small samples than MoM and MLE. (Šimková & Picek, 2017) In the future, L-Moments should be used instead of MoM, with MLE and PWM to estimate parameters. Correct justification of the parameters is crucial to obtain a realistic seismic hazard assessment, hence the goodness of fit analysis should always be followed.

7 Appendix: Code Used

7.1 Code for Simulated Return Plots

```
% Code to plot a 'simulated' return level for
% demonstration purposes

x = linspace(0,10);
% Creating a random sample between 0 and 10.
X = 10*rand(100, 1);

% Fitting a GEV distribution to the sample
param = pwmfitgev(X, true, 0.35);

% Fitting a Gumbel distribution to the sample.
param1 = pwmfitgumbel(X, true, 0.35);


% Plotting the Return Levels
plot(x,1./(1-gevcdf(x, param(1), param(2), param(3))),'r');
hold on;
plot(x,1./(1-evcdf(x,double(param1(1)), double(param1(2)))),'b');

% Legend, Title, etc.
legend('GEV', 'Gumbel')
xlabel('Return Level');
ylabel('Return Period');
title('Plot of return levels from random samples');
hold off
```

7.2 MoM Estimation Code for Gumbel and GEV

MoM Estimates for Gumbel:

```
% Function that outputs MoM estimates for Gumbel

function [parameter] = MoMGumbel(X)
% set the second column of the output as the estimator for scale parameter
parameter(2) = sqrt(6*var(X))/pi;
% set the first column of the output as the estimator for location parameter
parameter(1) = mean(X)+0.57722*parameter(2);
end
```

MoM Estimates for GEV:

```
% Function that outputs MoM Estimates for GEV

function [para] = MoMGEV(X)

% set g, v and m as the sample skewness, sample variance and sample mean respectively
g = skewness(X); v = var(X); m = mean(X);

% create initial values of k to be substituted into the equation
h = linspace(-0.5,0.5,10000); y = [];

% use a for loop to compute the corresponding values of y, using the equation in (5)
for i = 1:length(h)
y(i) = sign(h(i))*(-gamma(1+3*h(i))+3*gamma(1+h(i))*gamma(1+2*h(i))...
-2*(gamma(1+h(i)))^3)/((gamma(1+2*h(i))-(gamma(1+h(i)))^2)^1.5) - g;
end

% find the value of h which is closest to zero, i.e. the value of k that satisfies the equation
p = find(abs(y)<=min(abs(y)));
k = h(p);

% compute the corresponding values of sigma and mu using the equations in (3) and (4)
s = (sign(k)*sqrt(v)*k)/((gamma(1+2*k)-(gamma(1+k))^2)^0.5);
mu = m - (s/k)*(1-gamma(1+k));

% set the output of the function as mu, sigma and -k
para(1) = mu; para(2) = s; para(3) = -k;
end
```

7.3 PWM Estimation Code for Gumbel and GEV

PWM Estimates for Gumbel:

```
% Function that outputs PWM Estimates for Gumbel

function paramhat = pwmfitgumbel(X, consistent, a)
% How to use: param = pwmfitgumbel(X, true, 0.35) - write this
% You will get param = [mu, sigma]
% Setting consistent to be true will result in use of a consistent
% estimator.
% Setting consistent to be false will result in use of an unbiased
% estimator.
% The value a must between 0 and 1

n = length(X);
X = sort(X);
b0 = sum(X)/n;

%Unbiased estimator
sum1 = 0;
for j = 1:n
    sum1 = sum1 + (j-1)/(n-1)*X(j);
end

ub1 = sum1/n;

%Consistent estimator - choose value of a between 0 and 1
sum3 = 0;
```

```
for j = 1:n
    sum3 = sum3 + ((j-a)/(n))*X(j);
end

cb1 = sum3/n;

% Which statistic to use?
if consistent
    b1 = cb1;
else
    b1 = ub1;
end

%Calculation of estimates for scale, location - orders used 0,1
sigma = (2*b1 - b0)/log(2);
mu = b0 + sigma*eulergamma;
paramhat = [vpa(mu), vpa(sigma)];
end
```

PWM Estimates for GEV:

```
% Function that outputs PWM Estimates for GEV

function paramhat = pwmfitgev(X, consistent, a)
% How to use: param = pwmfitgev(X, true, 0.35) - write this
% You will get param = [k, mu, sigma]
% Setting consistent to be true will result in use of a consistent
% estimator.
% Setting consistent to be false will result in use of an unbiased
% estimator.
% The value a must between 0 and 1

n = length(X);
X = sort(X);
b0 = sum(X)/n;

%Unbiased estimator
sum1 = 0;
sum2 = 0;
for j = 1:n
    sum1 = sum1 + (j-1)/(n-1)*X(j);
    sum2 = sum2 + (((j-1)*(j-2))/((n-1)*(n-2)))*X(j);
end

ub1 = sum1/n;
ub2 = sum2/n;

%Consistent estimator - choose value of a between 0 and 1
sum3 = 0;
sum4 = 0;
for j = 1:n
    sum3 = sum3 + ((j-a)/(n))*X(j);
    sum4 = sum4 + ((j-a)/(n))^2*X(j);
end

cb1 = sum3/n;
cb2 = sum4/n;

% Which statistic to use?
if consistent
    b1 = cb1;
    b2 = cb2;
```

```
else
    b1 = ub1;
    b2 = ub2;
end

%Calculation of estimates for shape, scale, location - orders used 0,1,2
c = ((2*b1 - b0)/(3*b2 - b0)) - log(2)/log(3) ;
k = 7.8590*c + 2.9554*c^2;
sigma = ((2*b1 - b0)*k)/(gamma(1+k)*(1 - 2^(-k)));
mu = b0 + (sigma/k)*(gamma(1+k) - 1);
paramhat = [-k, sigma, mu];
end
```

7.4 Code for Bootstrap Confidence Intervals of parameters

Bootstrap Confidence Intervals for each parameter in the Gumbel Distribution:

```
%data1 = the maxima of Greek earthquakes data (1901-2017)
%estimates parameters from initial sample [location, scale]
initialparame= [6.519436417; 0.656697667];

n = 117;  %size of each data set
nReps = 2000;  %number of data sets or 'experiments'
id = ceil(rand(n,nReps)*n);
bootstrapData2 = data1(id);
%'empty' matrix to fill in parameter estimates from 2000 resamples
paramresamplee= ones(2,2000);

%work out what parameter estimation method using (MoM, MLE, PWM) for gumbel
%change pwmfitgumbel to momgumbel / 'evfit' function
for i=1:2000
    paramresamplee(:,i)=pwmfitgumbel(bootstrapData2(:,i),true,0.35);
end

%bootstrap difference for each parameter estimate
%deltastar1e for location
deltastar1e= paramresamplee(1,:) - initialparame(1)*ones(1,2000);
%deltastar1e for scale
deltastar2e= paramresamplee(2,:) - initialparame(2)*ones(1,2000);

%work out quantiles of deltastar, 0.025 and 0.975
%95% confidence interval for parameter
%location 95% bootstrap ci [CI11e, CI12e]
CI11e = initialparame(1) - quantile(deltastar1e, 0.975);
CI12e = initialparame(1) - quantile(deltastar1e, 0.025);
%shape 95% bootstrap ci [CI21e, CI22e]
CI21e = initialparame(2) - quantile(deltastar2e, 0.975);
CI22e = initialparame(2) - quantile(deltastar2e, 0.025);
```

(Oxenham, 2011)

Bootstrap Confidence Interval for each parameter in the GEV Distribution:

```
%data1 = the maxima of Greek earthquakes data (1901-2017)
%estimates parameters from initial sample [shape, scale, location]
initialparam= [-0.124315603879923; 0.583335943880307; 5.92695649405941];
```

```
n = 117;  %size of each data set
nReps = 2000;  %number of data sets or 'experiments'
id = ceil(rand(n,nReps)*n);
bootstrapData = data1(id);
%'empty' matrix to fill in parameter estimates from 2000 resamples
paramresample= ones(3,2000);

%work out what parameter estimation method using (MoM, MLE, PWM) for GEV
%change pwmfitgev to momgev / 'gevfit' function
for i=1:2000
    paramresampleg(:,i)=pwmfitgev(bootstrapData(:,i), true, 0.35);
end

%bootstrap difference for each parameter estimate
%deltastar1g for shape
deltastar1g= paramresampleg(1,:) - initialparam(1)*ones(1,2000);
%deltastar2g for scale
deltastar2g= paramresampleg(2,:) - initialparam(2)*ones(1,2000);
%deltastar3g for location
deltastar3g= paramresampleg(3,:) - initialparam(3)*ones(1,2000);

%work out quantiles of deltastar, 0.025 and 0.975
%95% confidence interval for parameter
%shape 95% bootstrap ci [CI11g, CI12g]
CI11g = initialparam(1) - quantile(deltastar1g, 0.975);
CI12g = initialparam(1) - quantile(deltastar1g, 0.025);
%scale 95% bootstrap ci [CI21g, CI22g]
CI21g = initialparam(2) - quantile(deltastar2g, 0.975);
CI22g = initialparam(2) - quantile(deltastar2g, 0.025);
%location 95% bootstrap ci [CI31g, CI32g]
CI31g = initialparam(3) - quantile(deltastar3g, 0.975);
CI32g = initialparam(3) - quantile(deltastar3g, 0.025);
```

(Oxenham, 2011)

8 References

Alves, I. & Neves, C. (2008) *Extreme Value Distribution.* Available from: `https://pdfs.semanticscholar.org/4c0d/a78292aa15dab79446f37b3ee73cb5c5fdec.pdf` [Accessed on 12th June 2018].

Babu, G. (2011) *Model fitting, bootstrap, and model selection.* Available from: `http://astrostatistics.psu.edu/RLectures/model_selection_gof-bootstrap.pdf` [Accessed on 8th June 2018].

Bhunya, P. K., Jain, S. K., Ojha, C. S. & Agarwal, A. (2007) Simple parameter estimation technique for three-parameter generalized extreme value distribution. *Journal of hydrologic engineering.* 12(6), 682-689. Available from: DOI: 10.1061/(ASCE)1084-0699(2007)12:6(682).

Castillo, E. (1988) *Extreme Value Theory in Engineering.* Academic Press, inc.(London) LTD.

Coles, S. (2001) *An introduction to statistical modeling of extreme values.* Springer.

Geyer, C. (2012) *Lecture Slides Deck 8.* [Presentation] Stat 5102. University of Minnesota, 12th March. Available from: `http://www.stat.umn.edu/geyer/s12/5102/slides/s8.pdf` [Accessed on 8th June 2018].

Greenwood, J. A., Landwehr, J. M., Matalas, N. C. & Wallis, J. R. (1979) Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. *Water resources research.* 15(5), 1049-1054. Available from: DOI: 10.1029/WR015i005p01049.

Haan, L. & Ferreira, A. (2006). *Extreme Value Theory, An Introduction.* Springer, New York.

Hosking, J. R. M. (1990) L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B.* 52(1), 105–124. Available from: DOI: 10.2307/2345653.

Hosking, J. R. M., Wallis, J. R. & Wood, E. F. (1985) Estimation of the generalized extreme value distribution by the method of probability-weighted moments. *Technometrics.* 27(3), 251-261. Available from: DOI: 10.2307/1269706.

Institute of Geodynamics, National Observatory of Athens. (2018) *Earthquakes catalogues.* Available from: `http://www.gein.noa.gr/en/seismicity/earthquake-catalogs` [Accessed on 13th June 2018].

Jost, S. (2018) *Normal Plots.* Available from: `http://condor.depaul.edu/sjost/it223/documents/normal-plot.htm` [Accessed on 5th June 2018].

Kotz, S. & Nadarajah, S. (2000) *Extreme value distributions: theory and applications.* Imperial College Press.

Mahdi, S. & Cenac, M. (2004) Estimating parameters of Gumbel distribution using the methods of moments, probability weighted moments and maximum likelihood. *Revista de Matemática: Teoría y Aplicaciones.* 12(1-2), 151-156. Available from: DOI: 10.15517/rmta.v12i1-2.259.

MathWorks. (2018) *QQ Plot.* Available from: `https://uk.mathworks.com/help/stats/qqplot.html` [Accessed on 11th June 2018].

MathWorks. (2018a) *evfit.* Available from: `https://uk.mathworks.com/help/stats/evfit.html`[Accessed on 17th June 2018].

MathWorks. (2018b) *gevfit.* Available form: https://uk.mathworks.com/help/stats/gevfit.html[Accessed on 17th June 2018].

Nadarajah, S. (2018a) *The extremal types theorem.* Available from: http://www.maths.manchester.ac.uk/~saralees/ettproof.pdf [Accessed on 16th June 2018].

Nadarajah, S. (2018b) *Fundamental of extreme value theory.* Available from: http://www.maths.manchester.ac.uk/~saralees/chap1.pdf [Accessed on 16th June 2018].

National Aeronautics and Space Administration. (2018) *Generalized extreme value distribution and calculation of return value.* Available from: https://gmao.gsfc.nasa.gov/research/subseasonal/atlas/GEV-RV-html/GEV-RV-description.html[Accessed on 9th June 2018].

Orloff, J. & Bloom, J. (2014) *Bootstrap confidence intervals.* Available from: https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading24.pdf [Accessed on 8th June 2018].

Oxenham. A. (2011) *Lesson 6: Introduction to the Bootstrap.* Available from: http://www.mbfys.ru.nl/~robvdw/DGCN22/PRACTICUM_2011/LABS_2011/ALTERNATIVE_LABS/Lesson_6.html [Accessed on 13rd June 2018].

Rasmussen, P. F. & Gautam, N. (2003) Alternative PWM-estimators of the Gumbel distribution. *Journal of Hydrology.* 280(1-4), 265-271. Available from: DOI: 10.1016/S0022-1694(03)00241-5.

Scott, D. (2018) *Quantile-Quantile (q-q) plots.* Available from: http://onlinestatbook.com/2/advanced_graphs/q-q_plots.html [Accessed on 1st June 2018].

Šimková, T. & Picek, J. (2017) A comparison of L-, LQ-, TL-moment and maximum likelihood high quantile estimates of the GPD and GEV distribution. *Communications in statistics - Simulation and computation.* 46(8), 5991-6010, Available from: 10.1080/03610918.2016.1188206

Smith, R. (2009) *Extreme value theory.* Available from: https://www.unc.edu/~rls/s890/evtclass.pdf [Accessed on 2nd June 2018].

Stedinger, J. R. & Martins, E. S. (2000). *Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data.* Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.463.9611&rep=rep1&type=pdf[Accessed 16th June 2018].

Stedinger, J. R., Vogel, R. M., & Foufoula-Georgiou, E. (1993) Frequency analysis of extreme events. *Handbook of hydrology.* 18. Available from: https://engineering.tufts.edu/cee/people/vogel/documents/frequencyAnalysis.pdf [Accessed 14th June 2018].

Stefopoulos, A. (2018) *Earthquakes in Greece (1901-2018).* Available from: https://www.kaggle.com/astefopoulos/earthquakes-in-greece-19012018 [Accessed 13th June 2018].

Sun, R. (2018) *Lecture 12.* Available from: http://www.math.nus.edu.sg/~matsr/ProbI/Lecture12.pdf[Accessed 15th June 2018].

U.S. Geological Survey. (2018) *Why are we having so many earthquakes?.* Available from: https://tinyurl.com/ya726dwd [Accessed on 17 June 2018]. (The link has been compressed using http://tinyurl.com)

Wilk, M. B. & Gnanadesikan R. (1968) Probability plotting methods for the analysis of data. *Biometrika.* 55(1), 1-17. Available from: DOI: 10.2307/2334448.