Extreme Value Theory and Analysis

by Nicholas Burke

A project submitted to the Department of Mathematical Sciences in conformity with the requirements for MATH 9801 (Graduate Project)

Abstract

This Graduate project will discuss Extreme Value Theory and Extreme Value Analysis. Extreme Value Theory is a section of statistics pertaining to extreme values. It is used to describes the limiting behavior of extremes $max(X_2, ..., X_n)$ or $min(X_2, ..., X_n)$, the tails of the distributions. This project will explore important theorems and apply using the computer programming R, to further elaborate on this subject matter.

Acknowledgements

It has been a privilege to collaborate with Dr. Deli Li, and Dr. Liping Liu on this project. Their expertise has been greatly appreciated and vital to the completion of this project. I also want to thank everyone who has supported me throughout my graduate school journey.

Contents

1 Introduction

In this project, the origins of the Extreme Value Theory will be explored, related probability theory concepts will be addressed as well. In great detail, the two most important theorems in Extreme Value Theory will be examined as well as their respective related distributions. This project will continue from a practical approach, by outlining two methods in Extreme Value Analysis, concluding with an application using the computer programming software R.

1.1 Historical Review

Extreme Value Theory, EVT for short, was originally discovered in 1928 by two British statisticians Ronald Fisher and Leonard Tippett. The pair conjured three limiting distributions; Frechet, Weibull and Gumbel for the maximum value of an ordered sample. Essentially, these three distributions characterize the function G such that

$$\lim_{n \to \infty} P(\frac{X_{n:n} - b_n}{a_n} \leq x) = \lim_{n \to \infty} F^n(a_n x + b_n) = G(x)$$

where $X_{n:n} = max\{X_1, ..., X_n\}$, a_n, b_n are sequences of real numbers and F_n is a cdf.

At the time, the duo was unable to formulate a rigorous proof.

In 1943, Russian mathematician Boris Gnedenko was able to characterize each of the three types of limiting distributions as heavy tails as Frechet type and bounded support for Weibull. He was able to prove the max domains of attraction for the Frechet, Weibull and Gumbel distributions. Since then, there has been many development on the Fisher-Tippet-Gnedenko theorem to expand the topic of Extreme Value Theory.

2 Extreme Value Theory

The goal of Extreme Value Theory is to estimate the probability of distant outliers, calculating such odds may have a large impact, despite the unlikeliness of it occurring. In order to apply extreme value theory to a data set, must be an order sample of a random variable.

Preliminaries

To help grasp EVT, we will briefly cover some basic probability theory as well as some ordinal statistic concepts.

Definition 2.1. *A **probability density function (pdf)** is a function that describes the likelihood for a random variable to take on a given value.*

$$f(x) = P\{X = x\}$$

where $f(x) \leq 1$ for all x and $\sum f(x) = 1$.

Definition 2.2. *A **cumulative distribution function (cdf)** is the probability that the variable takes a value less than or equal to x.*

$$F(x) = Pr\{X \leq x\}$$

Definition 2.3. *Let $X_1, X_2, ..., X_n$ be i.i.d random variables taken from a discrete or continuous distribution, The **order statistic** of*

$$X_{(1)}, X_{(2)}, ..., X_{(n)}.$$

*The **order sample** is the strict inequality*

$$X_{(1)} < X_{(2)} < ... < X_{(n)}$$

Definition 2.4. *The **first order statistic** is the minimum of the sample:*

$$X_{(1)} = min(X_1, X_2, ..., X_n)$$

Definition 2.5. *The nth **order statistic** is the maximum of the sample:*

$$X_{(n)} = max(X_1, X_2, ..., X_n)$$

Definition 2.6. *An **extreme value** is the minimum or maximum value in a probability distribution.*

Consider the cumulative distribution function of X_1 and let $M_n = max\{X_1, ..., X_n\}$. Then the cumulative distribution function of M_n is

$$\begin{aligned} M_n &= P(X_1 \leq x, ..., X_n \leq x) \\ &= P(X_1 \leq x) \cdots P(X_n \leq x) \\ &= F(x)^n \end{aligned}$$

Definition 2.7. *An **extreme value distribution** is a limiting model for both the maximums and minimums of a particular data set.*

Definition 2.8. *A **limiting distribution** models size the data will reach.*

Definition 2.9. *A non-degenerate distribution with cumulative distribution function $G(x)$ is an **extreme value distribution** if there are sequences of real numbers $a_n \geq 0$ and b_n and cumulative distribution function $F(x)$ such that*

$$\lim_{n \to \infty} F^n(a_n x + b_n) = G(x)$$

at every point x where G is continuous.
*The distribution F is in the **domain of attraction** of G.*

2.1 Fisher-Tippett-Gnedenkno Theorem

The first important theorem in Extreme Value Theory is the **Fisher-Tippett-Gnedenkno Theorem**, also referred to as the **Extreme Value Theorem** is essentially a generalization for the asymptotic distribution of extreme order statistics.

Theorem 2.10. *Let $X_1, X_2, .., X_n$ be a sequence of independent and identically-distributed random variables and $M_n = max(X_1, ..., X_n)$. If as sequence of pairs of real numbers (a_n, b_n) exists such that $a_n > 0$ and*

$$\lim P(\frac{M_n - b_n}{a_n}) \leq x = F(x)$$

*where F is a non-degenerate distribution function, then the limit distribution F belongs to either the **Gumbel**, the **Frechet** or the **Weibull** family, these can be grouped into the **Generalized Extreme Value Distribution**.*

Proof. In order to prove this theorem, we will use various examples to help illustrate the concept. We will start with a basic example on the standard uniform distribution and build upon it.

Example 2.11. *Let X to be a random variable with cumulative distribution function, $F(x) = P(X \leq x)$ and let $x^* = \sup\{x : F(x) < 1\}$.*
Consider X be uniformly distributed, $X \sim U(0, 1)$, then

$$F(x) = 0, x \leq 0$$
$$= x, 0 < x < 1$$
$$= 1, x \geq 1$$

Then in this case we have, $x^* = \sup\{F(x) < 1\} = 1$, thus if $F(x) < 1$ for all x, then $x^* = \sup x : F(x) < 1 = \infty$.

Now lets consider the sequence of *iid* random variables $X_1, X_2, ..X_n, ...$ Let $F(x) = P(X_1 \leq x)$ be the cumulative distribution function $\forall x$ and let $x^* = \sup\{x : F(x) < 1\}$. Then $\max_{1 \leq i \leq n} \to x^*$ in probability.
We will prove this for two cases.
CASE 1: $x^* < \infty$, for any given $\epsilon > 0$

$$P(\max_{1 \leq i \leq n} X_i \leq x^*) = P(X_1 \leq x^* - \epsilon, X_2 \leq x^* - \epsilon, ..., X_n \leq x^* - \epsilon)$$
$$= (F(x^* - \epsilon))^n \to 0$$

Since $F(x^* - \epsilon) < 1$
CASE 2:

$$P(\max_{1 \leq i \leq n} X_i > x^* + \epsilon) = 1 - P(\max_{1 \leq i \leq n} X_i < x^* + \epsilon)$$
$$= 1 - (P(X_1 < x^* + \epsilon))^n$$
$$= 1 - 1^n$$
$$= 0$$

and hence
$P(x^* - \epsilon < \max_{1 \leq i \leq n} X_i < x^* + \epsilon) = 1 - P(\max_{1 \leq i \leq n} X_i \leq x^* - \epsilon) - P(\max_{1 \leq i \leq n} X_i \leq x^* + \epsilon) \to 1$
as $n \to \infty$

Example 2.12. *Let $X_1, X_2, ..., X_n$ be a sequence of iid with common probability distribution function given by*

$$P(X_1 = 1 - \tfrac{1}{m}) = \tfrac{1}{2^m} \text{ for m=1,2,...}$$

Possible values of X_1 are : $0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, ... \to x^* = 1$ and hence $\max_{1 \leq i \leq n} X_1 \to 1$ in probability.

CASE 2: $x^* = \infty$, for any positive number c

$$P(\max_{1 \leq i \leq n} X_i > c) = 1 - P(\max_{1 \leq i \leq n} X_i \leq c)$$
$$= 1 - (F(c))^n$$
$$\to 1 - 0$$
$$= 1$$

Since $F(c) < 1$ and hence $\max_{1 \leq i \leq n} X_i \to \infty$ in probability.
Now we will be able to find two sequences $\{a_n > 0 : n \geq 1\}$ and $\{b_n > 0 : n \geq 1\}$ such that
$$P(\tfrac{\max_{1 \leq i \leq n} X_i - b_n}{a_n} \leq x) \to G(x) \text{ as } n \to \infty$$

where $G(x)$ is a non-degenerate distribution function.
Note that

$$P(\frac{\max_{1 \leq i \leq n} X_i - b_n}{a_n} \leq x) = P(\max_{1 \leq i \leq n} X_i \leq a_n x + b_n)$$
$$= (F(a_n x + b_n))^n, n \geq 1$$

Therefore, we have

$$log(F(a_n x + b_n))^n \to log G(x) \text{ as } n \to \infty$$
$$n log(F(a_n x + b_n)) \to log G(x) \text{ as } n \to \infty$$

$\square$

From *Definition 2.9* if and only if n is large, M_n can then be approximated by $a_n W + b_n$, where W is a random variable with cumulative distribution function $G(x)$, thus

$$\lim_{n \to \infty} P(\tfrac{M_n - b_n}{a_n} \leq x) = G(x)$$

at every continuity point of $G(x)$.
The **Generalized Extreme Value Distribution** and related distributions will be discussed in detail in the subsequent sections.

Basins of attraction

Although not every distribution is in the basin of attraction of an extreme value distribution, there are some basic conditions that can be checked to determine whether a distribution F is attracted to a Frechet or reverse Weibull distribution.

Theorem 2.13. *Let F be a cumulative distribution function. Then F is in the basin of attraction of a Frechet distribution with ε if and only if $F(x) < 1$ for all $x \exists \Re$ and*

$$\lim_{t \to \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{\frac{-1}{\varepsilon}}$$

for every $x > 0$.

Theorem 2.14. *Let F be a cumulative distribution function. Then F is in the basin of attraction of a Weibull distribution with ε if and only if $x^* = \sup\{x : F(x) < 1\} < \infty$ and*

$$\lim_{t \to 0+} \frac{1 - F(x^* - tx)}{1 - F(x^* - t)} = x^{\frac{-1}{\varepsilon}}$$

for every $x > 0$.

2.2 Generalized Extreme Value Distribution

The general idea from the aforementioned theorem, is that the extremes from any set of data with a well-behaved distribution can be modeled by the extreme value distributions. The generalized extreme value distribution is the only limit distribution of re-normalized maxima of a sequence of independent and identically distributed random variables that satisfies the Fisher-Tippett-Gnedenkno Theorem.

Theorem 2.15. *Every extreme value distribution has a cumulative distribution function as follows*
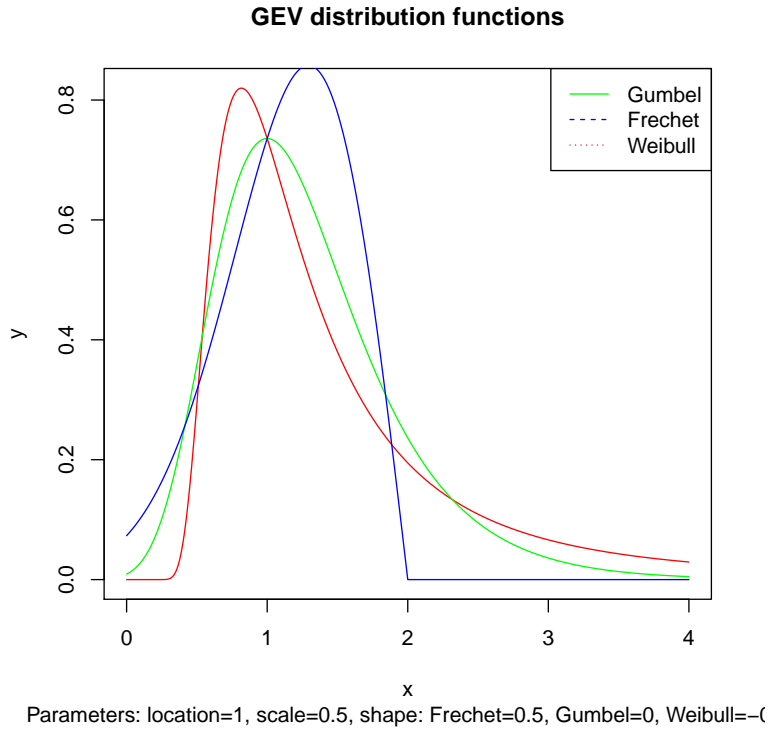
$$F(x : \sigma, \mu, \varepsilon) = e^{\{-[1 + \varepsilon(\frac{x - \mu}{\sigma})]^{-\frac{1}{\varepsilon}}\}}$$

and probability distribution as follows

$$f(x : \sigma, \mu, \varepsilon) = (1 + \varepsilon(\tfrac{x-\mu}{\sigma}))^{(\frac{-1}{\varepsilon}) - 1} e^{\{-[1 + \varepsilon(\frac{x-\mu}{\sigma})]^{-\frac{1}{\varepsilon}}\}}$$

*this known as the **Generalized Extreme Value Distribution** and defined by three parameters, **location** parameter μ, the **scale** parameter σ, and the **shape** parameter ε. σ and $[1 + \varepsilon(\frac{x-\mu}{\sigma})]$ must be greater than 0, ε and μ can take on any real value.*

This development of this family of continuous probability distributions in Extreme Value Theory, essentially combines the **Gumbel, Frechet** and **Weibull** distributions, each will be examined further in the following sections. The maximum of a sample of independent and identically distributed random variables after normalization can only converge to these families of distributions.

GEV distribution functions

Parameters: location=1, scale=0.5, shape: Frechet=0.5, Gumbel=0, Weibull=-(

Gumbel Distribution

The most commonly used extreme value distribution type in Extreme Value Theory, the **Gumbel** distribution is unbounded and defined on the entire set of real numbers. When the **shape** parameter $\varepsilon = 0$, the **Generalized Extreme Value Distribution** is equal to the **Gumbel** distribution. It can be used to model both minimums and maximums.
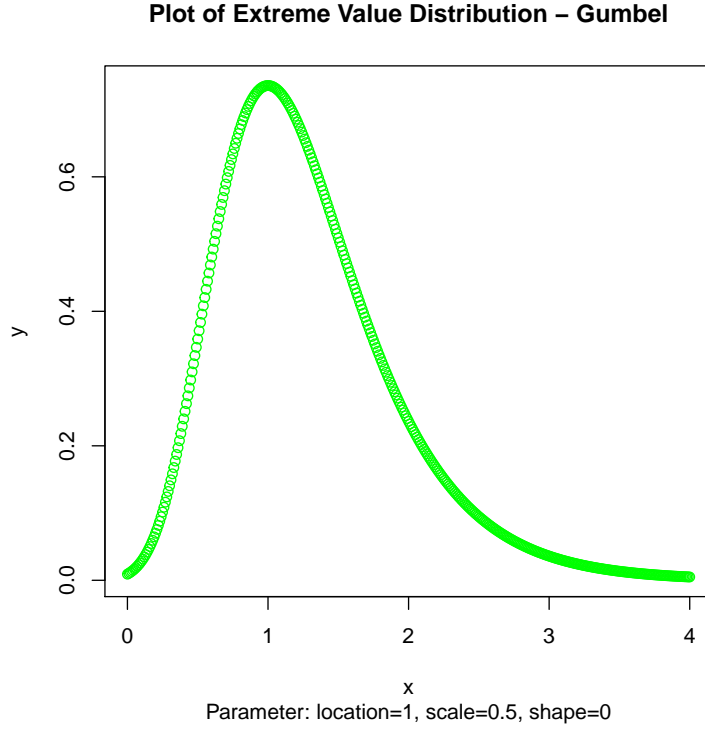
Plot of Extreme Value Distribution – Gumbel

Parameter: location=1, scale=0.5, shape=0

Minimum Gumbel Distribution

The general form of probability distribution function the Minimum Gumbel is as follows

$$F(x:\sigma,\mu) = \frac{1}{\beta}e^{\left(\frac{x-\mu}{\beta}\right)}e^{e^{-\frac{x-\mu}{\beta}}}$$

where μ, β are the **location** and **scale** parameters, respectively.

Standard Gumbel Distribution(Minimum)

When $\mu = 0$ and $\beta = 1$ the Minimum Gumbel distribution becomes standard form with probability distribution function as follows

$$F(x:\beta,\mu) = e^{x}e^{-e^{x}}$$

Maximum Gumbel Distribution

The general form of probability distribution function the Maximum Gumbel is as follows

$$F(x:\sigma,\mu) = \frac{1}{\beta}e^{-\left(\frac{x-\mu}{\beta}\right)}e^{-e^{-\frac{x-\mu}{\beta}}}$$

where μ, β are the **location** and **scale** parameters, respectively.

10

Standard Gumbel Distribution(Maximum)

When $\mu = 0$ and $\beta = 1$ the Maximum Gumbel distribution becomes standard form with probability distribution function as follows

$$F(x : \beta, \mu) = e^{-x}e^{-e^x}$$

In general the probability density function of a Gumbel distribution has only one unchanging shape which shifts according to μ, as μ increases there is a leftward shift, and as μ decreases there is a rightward shift.

The conditions of convergence for the Gumbel distribution is quiet complicated but we will go into further detail for the Frechet and Weibull distributions.

## Simulating the Gumbel Distribution in R

A Gumbel Extreme Value Distribution will be simulated using the data set below. Listed are the first six distribution values as well as the corresponding Gumbel Distribution plot.

```
   Time Observations
1 -10.0           0
2  -9.9           0
3  -9.8           0
4  -9.7           0
5  -9.6           0
6  -9.5           0
```

Generated Plot the Gumbel Distribution
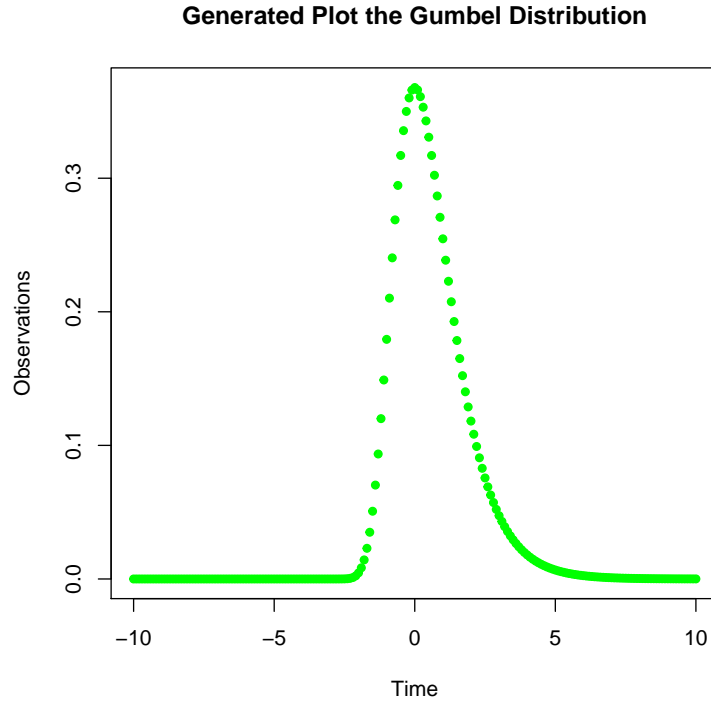
Frechet Distribution

The **Frechet** distribution is typically used to model maximum values since it is only bounded on the lower end and defined on the interval $[\mu : \infty]$, converging slowly to 1. When the shape parameter $\varepsilon > 0$, the **Generalized Extreme Value Distribution** is equal to the **Frechet** Distribution.

Plot of Extreme Value Distribution – Frechet

Parameter: location=1, scale=0.5, shape=0.5
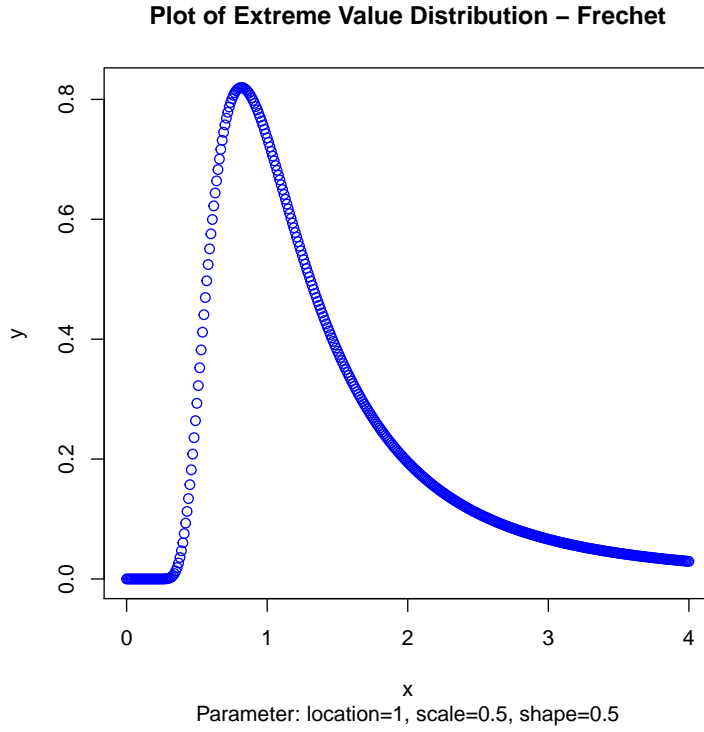
The general form of cumulative distribution function for Frechet is as follows

$$F(x) = e^{-\frac{x}{\alpha}}$$

where α is the **shape** parameter.

Standard Frechet Distribution

The standard form of cumulative distribution function for Frechet is as follows

$$F(x) = e^{-(\frac{x-\mu}{\beta})^\alpha}, \quad \text{if } x > \mu$$

where μ, β, α are the **location, scale** and **shape** parameters, respectively.

Condition of Convergence

If $G(x)$ is the distribution function of X, the M_n can be re-scaled to converge in distribution to a Frechet if and only if

$$G(x) < 1 \text{ for all } x \text{ and } \lim_{t \to \infty} \frac{1 - G(tx)}{1 - G(x)} = x^{-\theta} \text{ for all } x > 0,$$

$$\text{then } \lim_{n \to \infty} P\left(\frac{\max_{1 \leq i \leq n} X_i}{b_n} \leq x\right) = F(x).$$

In this case possible sequences are $a_n = 0$ and $b_n = G^{-1}(1 - \frac{1}{n})$ for $n \geq 1$.

Example 2.16. *Lets consider the function*

$$G(x) = \begin{cases} 1 - \frac{1}{x}, & x \geq 1 \\ 0, & x < 1 \end{cases}$$

As we can see $G(x) < 1$ for all $x > 0$. Now to show that Frechet's condition holds.

$$\begin{aligned}
\lim_{t \to \infty} \frac{1 - G(tx)}{1 - G(x)} &= \lim_{t \to \infty} \frac{1 - (1 - \frac{1}{tx})}{1 - (1 - \frac{1}{t})} \\
&= \lim_{t \to \infty} \frac{\frac{1}{tx}}{\frac{1}{t}} \\
&= \frac{1}{x} \\
&= x^{-\theta}
\end{aligned}$$

where $\theta = 1$.

It is noteworthy that a sequence b_n can be found as

$$\begin{aligned}
b_n &= G^{-1}(1 - \frac{1}{n}) \\
&= \sup\{x : G(x) \leq 1 - \frac{1}{n}\} \\
&= \sup\{x : 1 - \frac{1}{x} \leq 1 - \frac{1}{n}\} \\
&= \sup\{x : \frac{1}{n} \leq \frac{1}{x}\} \\
&= n, \quad n \geq 1
\end{aligned}$$

Thus,

$$\lim_{n \to \infty} P\left(\frac{\max_{1 \leq i \leq n} X_i}{n} \leq x\right) = \begin{cases} e^{-x-1}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

14

Simulating the Frechet Distribution in R

A Frechet Extreme Value Distribution will be simulated using the data set below. Listed are the first six distribution values as well as the corresponding Frechet Distribution plot.

```
    Time Observations
1 -10.0            0
2  -9.9            0
3  -9.8            0
4  -9.7            0
5  -9.6            0
6  -9.5            0
```

Generated Plot of the Frechet Distribution
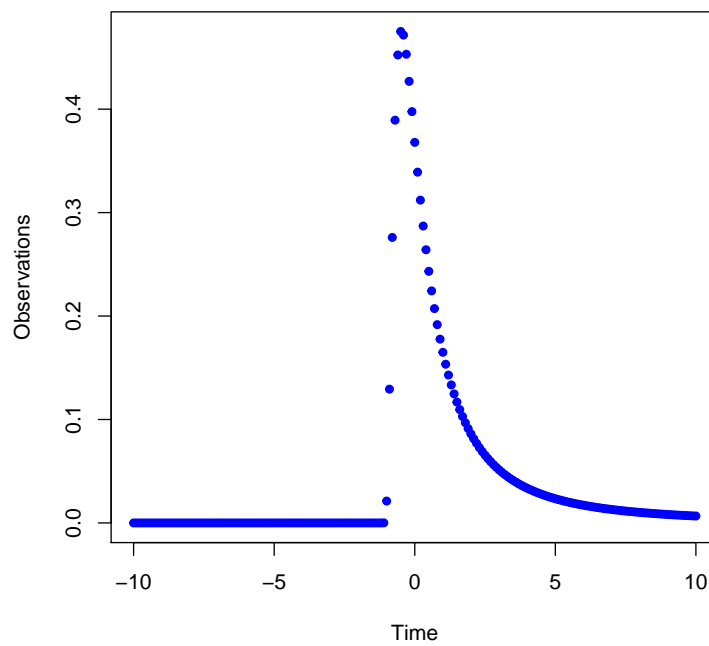
Weibull Distribution

The **Weibull** distribution is typically used to model failure times and real life data analysis. It is entirely dependent on the choice if parameters and is only bounded on the upper end. When the shape parameter $\varepsilon < 0$, the **Generalized Extreme Value Distribution** is equal to the **Weibull Distribution**.

Plot of Extreme Value Distribution – Weibull

Parameters: location=1, scale=0.5, shape=−0.5
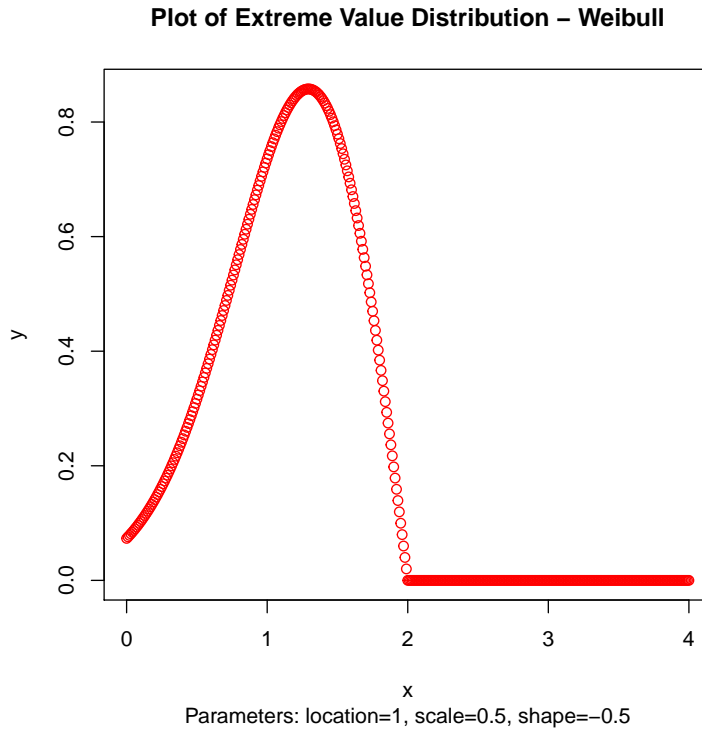
General Weibull Distribution

The general form of the probability distribution function for Weibull is as follows

$$f(x) = \frac{\gamma}{\alpha} \frac{x-\mu}{\alpha}^{(\gamma-1)} e^{(\frac{-x-\mu}{\alpha})^{\gamma}}, \quad x \geq \mu; \gamma, \alpha > 0$$

where γ, α, μ are the **shape, scale** and **location** parameters respectively.

Standard Weibull Distribution

When $\mu = 0$ and $\alpha = 1$, Weibull probability distribution function takes on the standard form as follows

$$f(x) = \gamma x^{(\gamma-1)} e^{(-x-\gamma)}, \quad x \geq 0; \gamma > 0,$$

where γ, α, μ are the **shape, scale** and **location** parameters respectively.

For the two parameters case, the standard Weibull probability distribution function is as follows

$$f(x) = \frac{\gamma}{\alpha}\left(\frac{x}{\alpha}\right)^{\gamma-1} e^{-\left[\frac{x}{\alpha}\right]^{\gamma}}, \quad x \geq 0$$

where γ, α, μ are the **shape, scale** and **location** parameters respectively.

Condition of Convergence

If $G(x)$ is the distribution function of X, the M_n can be re-scaled to converge in distribution to a Weibull if and only if

$$\omega = \sup\{x : G(x) < 1\} < \infty \text{ and}$$

$$\frac{1-G(\omega+tx)}{1-G(\omega-t)} \to_{t \to \infty} x^{-\theta} \text{ for } x < 0.$$

In this case possible sequences are $b_n = \omega$ and $a_n = \omega - G^{-1}(1 - \frac{1}{n})$ for $n \geq 1$

Example 2.17. *Lets consider the function below on the unit interval* $[0, 1]$ *with uniform distribution*

$$G(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x \leq 1 \\ 0, & x > 1 \end{cases}$$

thus we have

$$\begin{aligned} \omega &= \sup\{G(x) < 1\} \\ &= \sup\{x : x < 1\} \\ &= 1 \end{aligned}$$

and

$$\begin{aligned} \frac{1 - G(1 + tx)}{1 - G(1 - t)} &= \frac{1 + (1 + tx)}{1 - (1 - t)} \\ &= \frac{-tx}{t} \\ &= -x^{\theta}, \quad x < 0 \end{aligned}$$

where $\theta = 1$ *and* $0 < t < \min\{1, \frac{-1}{x}\}$

It is noteworthy that a sequence a_n can be found as

$$a_n = \omega - G^{-1}(1 - \frac{1}{n})$$
$$= 1 - (1 - \frac{1}{n})$$
$$= \frac{1}{n}, \quad n \geq 1$$
$$b_n = 1$$

Thus,

$$\lim_{n \to \infty} P(\frac{\max_{1 \leq i \leq n} X_i - 1}{\frac{1}{n}} \leq x) = \begin{cases} e^x, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

Simulating the Weibull Distribution in R

A Weibull Extreme Value Distribution will be simulated using the data set below. Listed are the first six distribution values as well as the corresponding Weibull Distribution plot.

```
   Time Observations
1 -10.0 1.919718e-43
2  -9.9 2.338991e-42
3  -9.8 2.726787e-41
4  -9.7 3.042965e-40
5  -9.6 3.252047e-39
6  -9.5 3.329823e-38
```

Generated Plot of the Weibull Distribution
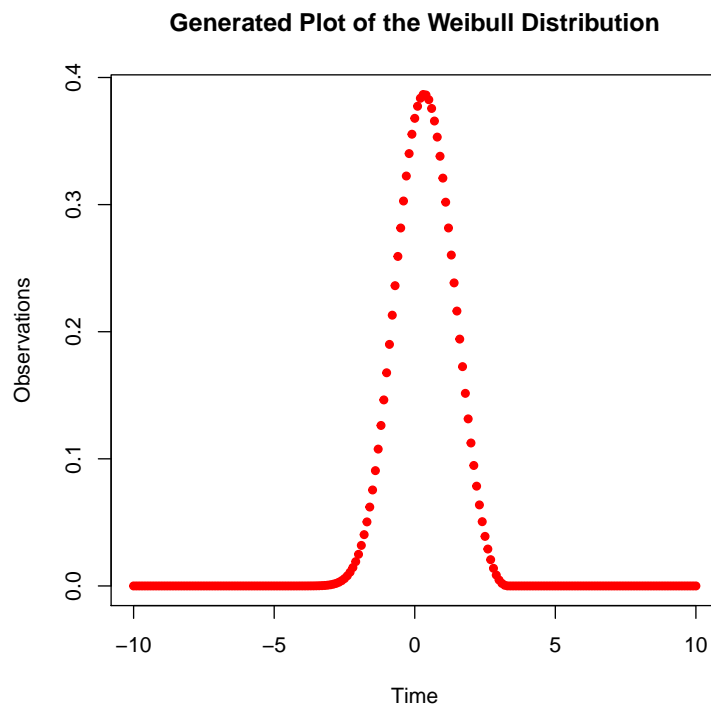
2.3 Pickands-Balkema-deHaan Theorem

The second most important theorem in EVT, the **Pickands-Balkema-deHaan Theorem** is mainly used for threshold exceedances. Suppose the true distribution of a random variable X is unknown, this theorem essentially provides an asymptotic tail distribution for it. The Peaks-Over-Threshold method uses this theorem extensively to provide results.

Conditional Excess Distribution Function

Suppose there is a random variable x from some unknown distribution F. Then the conditional excess distribution function is

$$F_u(y) = P(X - u \leq y | X > u) = \frac{F(u+y) - F(u)}{1 - F(u)}, \quad 0 \leq y \leq x_f - u$$

where x_f is either the finite or infinite right endpoint of the underlying distribution F.

Essentially the function F_u describes the distribution of excess over a given threshold u, given that threshold has been exceeded.

Theorem 2.18. *Let $(X_1, X_2, ...)$ be a sequence of independent and identically distribution random variables and let F_u be their conditional excess distribution function. For a large class of underlying distribution functions F, and large u, F_u is well approximated by the **Generalized Pareto Distribution**. That is*

$$F_u(y) = G_{k,\sigma}(y) \ as \ u \to \infty$$

where

$$G_{k,\sigma}(y) = \begin{cases} 1 - (1 + \frac{ky}{\sigma})^{\frac{-1}{k}}, & \textit{if } k \neq 0 \\ 1 - e^{\frac{-y}{\sigma}}, & \textit{if } k = 0 \end{cases}$$

- $k \geq 0$ *then* $\sigma > 0$ *and* $y \geq 0$

- $k < 0$ *then* $0 \leq y \leq \frac{-\sigma}{k}$

$$\lim_{u \to \infty} \sim GPD(u, \sigma, \varepsilon)$$

Proof. Omitted. □

If the probability distribution F of a random variable is not highly unusual, regardless of F, provided that the threshold u is sufficiently large exceedances of u will be distributed as the **Generalized Pareto Distribution**.

2.4 Generalized Pareto Distribution

The **Generalized Pareto Distribution** is a family of continuous probability distributions defined by three parameters, modelling the tails of another distributions.

The standard cumulative distribution function is as follows

$$F_\varepsilon(x) = \begin{cases} 1 - (1 + \varepsilon x)^{\frac{-1}{\varepsilon}}, & \text{for } \varepsilon \neq 0 \\ 1 - e^{-x}, & \text{for } \varepsilon = 0 \end{cases}$$

where ε is the **shape** parameter

- when $x \geq 0$ for $\varepsilon \geq 0$

- when $0 \leq x \leq \frac{-1}{\varepsilon}$ for $\varepsilon < 0$

The probability distribution function is as follows

$$f_\varepsilon(x) = \begin{cases} \frac{1}{\sigma}(\varepsilon x + 1)^{\frac{-\varepsilon+1}{\varepsilon}}, & \text{for } \varepsilon \neq 0 \\ \frac{1}{\sigma} - e^{-x}, & \text{for } \varepsilon = 0 \end{cases}$$

where ε is the **shape** parameter.

Standard Generalized Pareto Distribution

In order to standardize this distribution we will replace x with $\frac{x-\mu}{\sigma}$ and then adjust the support accordingly.

The cumulative distribution function is as follows

$$F_{\varepsilon,\mu,\sigma}(x) = \begin{cases} 1 - (1 + \varepsilon \frac{x-\mu}{\sigma})^{\frac{-1}{\varepsilon}}, & \text{for } \varepsilon \neq 0 \\ 1 - e^{(-\frac{x-\mu}{\sigma})}, & \text{for } \varepsilon = 0 \end{cases}$$

where μ, σ, ε are the **location, scale** and **shape** parameters, respectively.

- for $x \geq \mu$ when $\varepsilon \geq 0$

- for $\mu \leq x \leq \mu - \frac{\sigma}{\varepsilon}$ when $\varepsilon < 0$

where $\mu \exists \Re, \mu > 0, \varepsilon \exists \Re$.

The probability distribution function is as follows

$$f_{\varepsilon,\mu,\sigma}(x) = \frac{1}{\sigma}(1 + \frac{\varepsilon(x-\mu)}{\sigma})^{\frac{-1}{\varepsilon}-1}$$

where μ, σ, ε are the **location, scale** and **shape** parameters, respectively.

- for $x \geq \mu$ when $\varepsilon \geq 0$

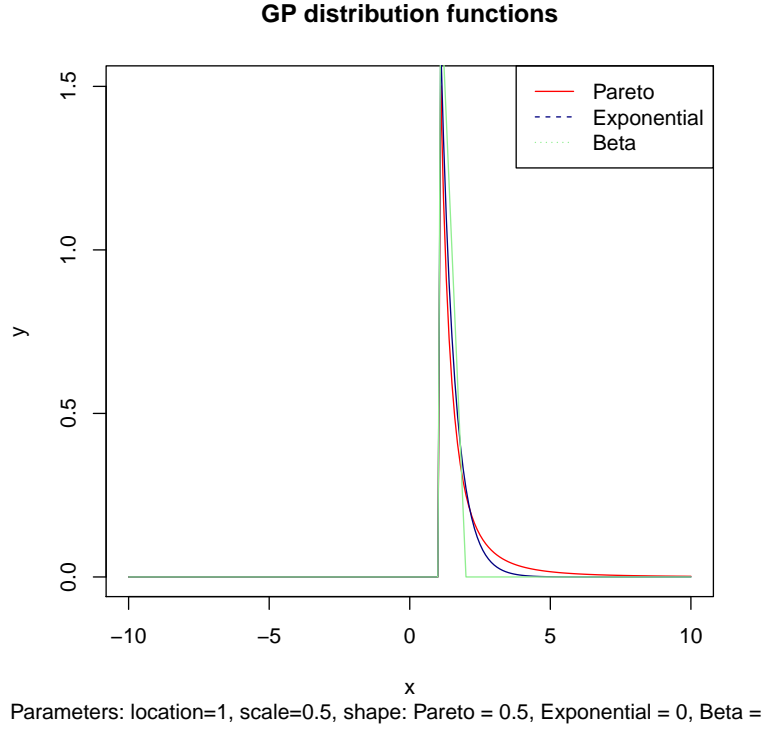- for $\mu \leq x \leq \mu - \frac{\sigma}{\varepsilon}$ when $\varepsilon < 0$

GP distribution functions

Parameters: location=1, scale=0.5, shape: Pareto = 0.5, Exponential = 0, Beta =

Exponential Distribution

The standardized exponential distribution has a distribution function of the form
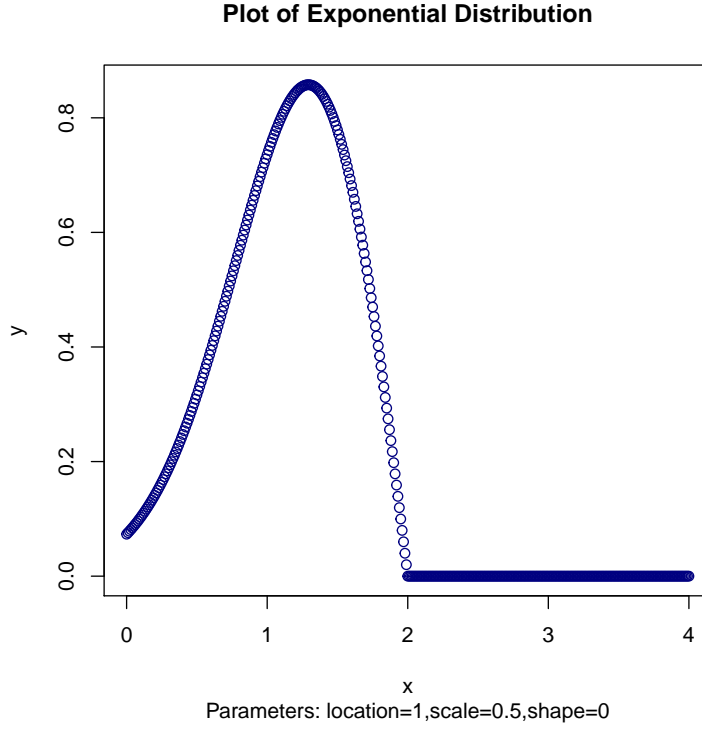
$$F(x) = 1 - e^{-x}, \; x > 0$$

with **location** and **scale** parameters μ and σ included, it has distribution function

$$F(x) = 1 - e^{\left(\frac{-x-\mu}{\sigma}\right)}, \; x > \mu$$

The distribution function of that one-parameter distribution will be denoted by

$$F(x) = 1 - e^{\left(\frac{-x}{\sigma}\right)}$$

When the **shape** parameter $\varepsilon = 0$ and the **location** parameter $\mu = 0$, the **Generalized Pareto Distribution** is equal to the **Exponential** distribution

Plot of Exponential Distribution

Parameters: location=1,scale=0.5,shape=0

Pareto Distribution

The standardized Pareto distribution has a distribution function of the form

$$F(x) = 1 - x^{-\alpha}, \; x \geq 1, \alpha > 0$$

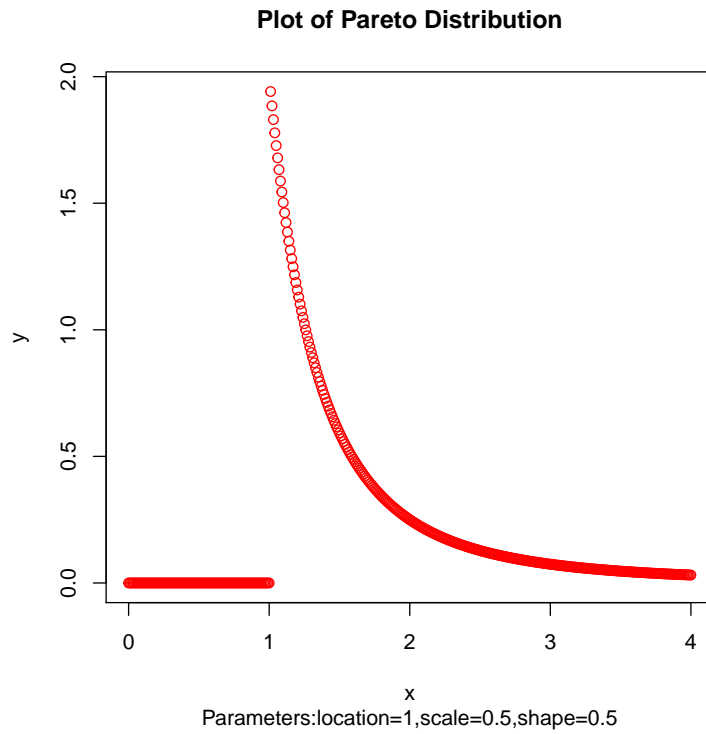with **location** and **scale** parameters μ and σ included, it has distribution function

$$F(x) = 1 - \left(\frac{x-\mu}{\sigma}\right)^{-\alpha}, \; x \geq \mu + \sigma \; \alpha, \sigma > 0$$

The distribution function of that two-parameter Pareto distribution will be denoted by

$$F(x) = 1 - \left(\frac{\sigma}{x+\sigma}\right)^{\alpha} \; x \geq 0, \alpha, \sigma > 0$$

When the **shape** parameter $\varepsilon > 0$ and the **location** parameter $\mu = \frac{\sigma}{\varepsilon}$, the **Generalized Pareto Distribution** is equal to the **Pareto** distribution, with **scale** parameter $x_m = \frac{\sigma}{\varepsilon}$ and **shape** parameter $\alpha = \frac{1}{\varepsilon}$.
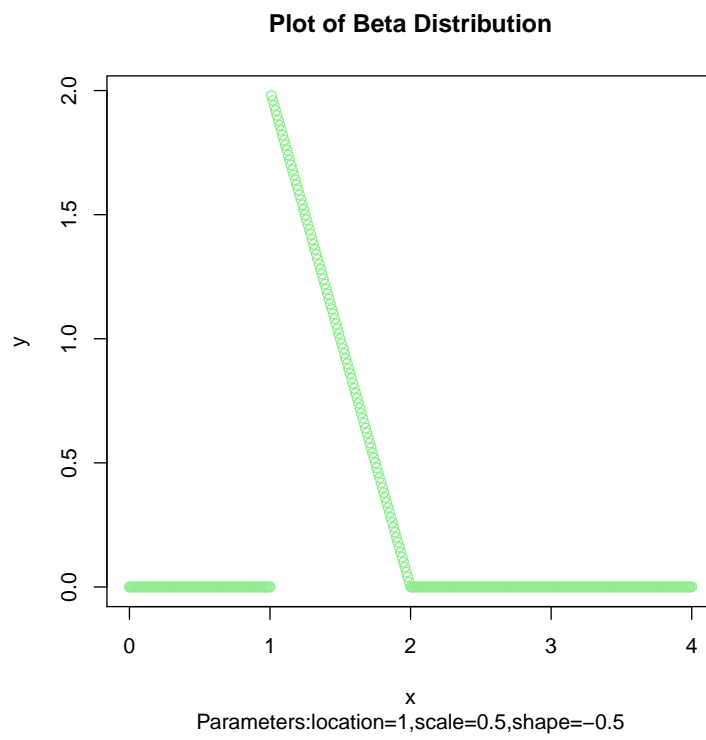
Plot of Pareto Distribution

Parameters:location=1,scale=0.5,shape=0.5

Beta Distribution

The standardized Beta distribution has a distribution function of the form

$$F(x) = 1 - (-x)^{-\alpha}; \; -1 \geq x \geq 0, \alpha < 0$$

with **location** and **scale** parameters μ and σ included, it has distribution function

$$F(x) = 1 - \left(\frac{-x-\mu}{\sigma}\right)^{-\alpha}; \; \mu - \sigma x \geq \sigma \; \alpha < 0, \sigma > 0$$

Plot of Beta Distribution

x
Parameters:location=1,scale=0.5,shape=−0.5

3   Extreme Value Analysis

Statistical inference about rare events is related to observations which are rare in some sense. The statistical inference associated with Extreme Value Analysis is essentially the estimation of the shape, location and scale parameters. This can then be used as a basis for any estimation of other relevant parameters.

Criteria

A set of data must satisfy the following conditions in order to be applicable to EVT

- Continuous probability distribution

- Inverse

- Consist of independent and identically distributed random variables

The general concept behind Extreme Value Analysis is based upon, if you were to generate any number of data sets, take the maximum and minimum value from each; a new distribution can be generated. The limit in this case can be viewed as a failure point, a point that if exceeded a failure or end-of-life event will occur.

3.1   Block Maxima

The **Block Maxima** method separates historical data in k blocks, typically over some time period such months, years or quarter, and examines and collects the maximum value from each, as sample data in order to model the tail risk distributions. This newly created data can be approximated by an extreme value distribution using parametric statistical methods under the extreme value conditions.

Pros

- Simple interpretation and application

- Retains some lower observations

- Can be applied to non *iid* observations

- Block periods often occur naturally in most situations

Cons

- Misses some high observations

- Framework not as useful in a straightforward manner

The **Block Maxima** can be denoted as:

$$M_n := max(X_1, ..., X_n)$$

is the maximum from n observations.

The distribution function of the **Block Maxima**, given the distribution function of the random variables X_i is given by the following,

$$P(M_n \leq x) = P(X_1 \leq x, ..., X_n \leq x) = F^n(x)$$

This distribution depends on the distribution of the underlying random variables, which may not be known in practice. Using the results of the **Fisher-Tippet-Gnedenko theorem**, we can use the asymptotically distribution for a large value of n to the fit the **Generalized Extreme Value Distribution**.

Theorem 3.1. *If appropriately normalized block maxima converge in distribution to a non-trivial limit distribution, then this distribution is the generalized extreme value (GEV) distribution as follows;*

$$H_\varepsilon(x) = \begin{cases} e^{-(1+\varepsilon x)^{\frac{-1}{\varepsilon}}}, & for\ \varepsilon \neq 0 \\ e^{-e^{-x}}, & for\ \varepsilon = 0 \end{cases}$$

*where ε is the **shape** parameter, that determine whether it belongs to **Gumbel, Frechet** or **Weibull** family of distributions .*

3.2   Peaks-Over-Threshold

The **Peaks-Over-Threshold** method sets a specific threshold value and collects data points that exceed such value, to find a conditional distribution. This distribution is approximately a Generalized Pareto Distribution using parametric statistical methods under extreme value conditions. The Generalized Pareto Distribution is primarily used for modelling the tails of distributions.

Pros

- Efficient and relevant use of data

- Picks up all relevant high observations

Cons

- Difficult to implement

- Hard to know if theoretical conditions are satisfied

Let X be the random variable representing losses and let u be a certain threshold. Then the random variable $X - u$ is the value of exceedances over the threshold u, if the threshold has been exceeded. The excess distribution function F_u is given by the following

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

An important result of Extreme Value Theory is that this function can be approximated by the Generalized Pareto Distribution.

Theorem 3.2. *For a class of distributions a function $\beta(u)$ can be found such that*

$$\lim_{u \to \bar{x}} \sup_{0 \leq x < \bar{x} - u} |F_u(x) - G_{\varepsilon, \beta(u)}| = 0$$

where, $\bar{x}$ is the rightmost point of the distribution and u is the threshold point

This theorem holds true for many of the standard distributions, essentially the conditional distribution of sizes of exceedances above threshold u can be asymptotically modeled by the Generalized Pareto Distribution.

4 Application

Extreme Value Theory has applications related to the environment, hydrology, finance and sports as well as other fields of study. As mentioned before this is called Extreme Value Analysis, where the extreme part of a sample is of greater focus. The statistical analysis of extremes can be implemented using various packages depending on a the specific topic. For this implementation of EVT will we be using the "extRemes" package in R. This package consists of Extreme Value Analysis functions used for performing extreme value analysis.

4.1 EVT in R

With the use of the "extRemes" package we will demonstrate how both the Peak-Over-Threshold and the Block Method approach are used. The Block Maxima method in the most basic approach in EVT, essentially the sample is divided into non-overlapping periods of equal size over an observed period of time, then the maximum observations in each period form a set of data. The Peak-Over-Threshold method, involves setting a certain threshold on the sample in which the exceedances form a new set of data of interest.

Block Maxima Method

The Block Maxima Method require a practical selection for the intermediate sequence $k = k_n$ in a sample of size n. The observations created follow domain of attraction conditions that can be approximated by an Extreme Value Distribution, GEV.

Use in R

Using "library(extRemes)", we will be using 12 blocks, with mean set to 5 and standard deviation set to 2. Create a series of maxima, we can simulate blocks of data, which are plotted below.

Block maxima
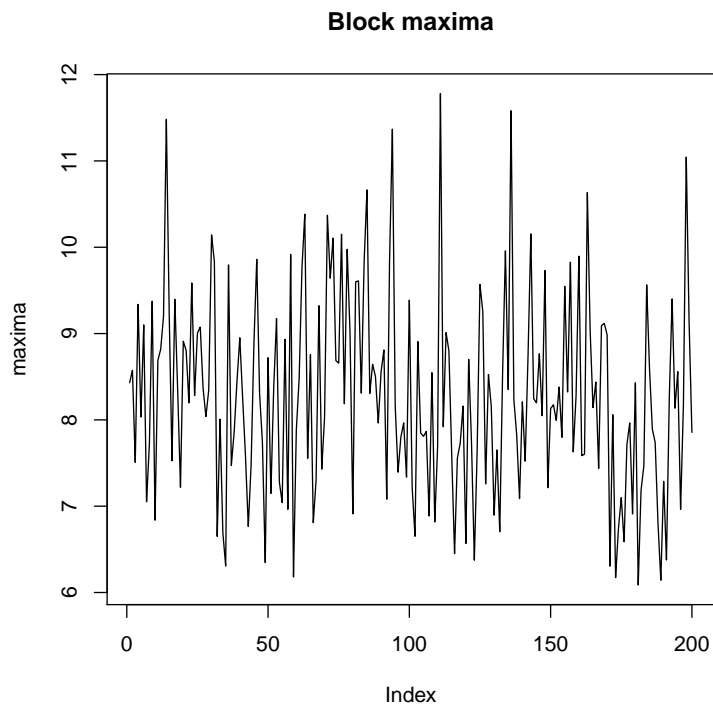
Now using the Maxima data generated, based on the Fisher-Tippet-Gnednko theorem, the density generated will have the same shape of the Gumbel distribution.

**Empirical density vs estimated Gumbel distribution**

N = 200   Bandwidth = 0.3464
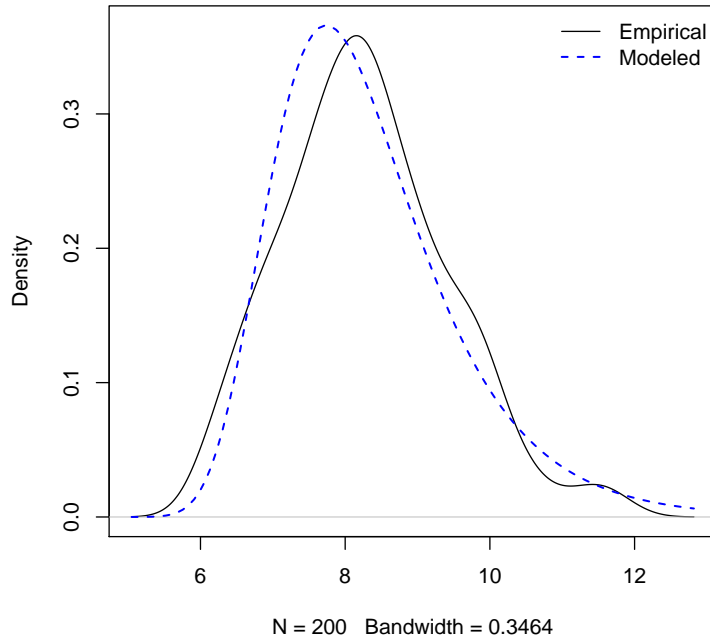
Peak-Over-Threshold Method

The Peaks-Over-Thresholds method uses observations above a given threshold, u. This threshold is determined by using the size of exceedances or the number of events within a period of time. This results in a more optimal sequence k, that minimizes the asymptotic Mean Square Error, as opposed to the Block Maxima method.

Use in R

Based on the data series created in the previous method, will set the threshold to 7.5 to create a plot of the data points that meet this requirement, below.

Peak Over Thresholds

series
threshold = 7.5
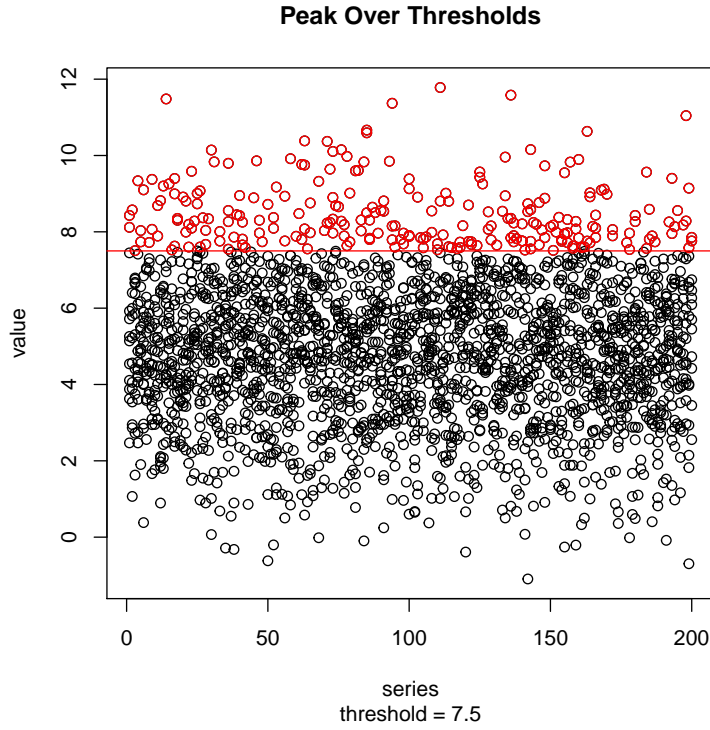
Now we will plot the excess distribution F_u, and based on the conditional distribution of sizes of exceedances it will modeled by the Generalized Pareto Distribution, GPD according to *Theorem 3.2.*

Empirical POT exceedances density vs estimated GP distributio

Density

N = 248   Bandwidth = 0.2451
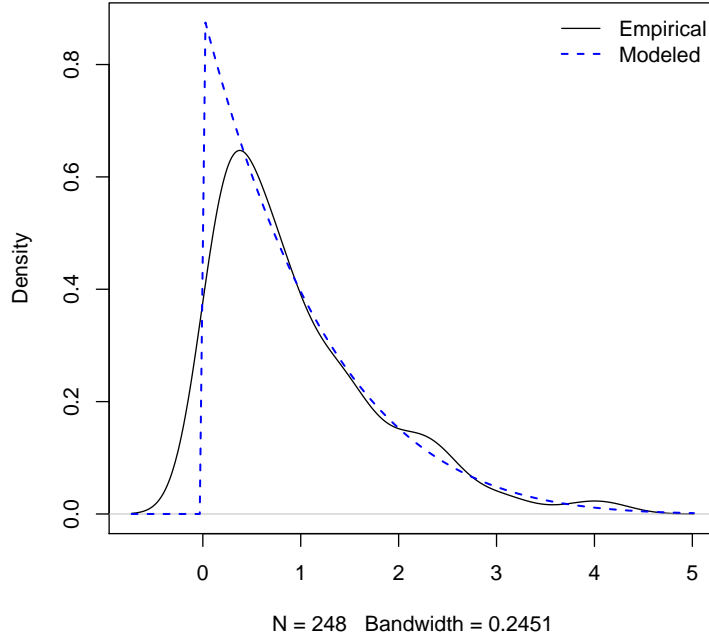
Further Discussion

In this project, we explored EVT from both a theoretical and analytically perspective, and how the two can intertwine to produce real piratical results. The Block Maxima method is primarily used for annual data by generating an Annual Maxima series. For this particular kind of analysis, the Fisher-Tippet-Gnedenko theorem is only partially useful with a greater emphasis and the type distribution for fitting the data, which in most cases differ from the GEV distributions. This is due to the fact that relevant data over the course of a year maybe be limited in a historical sense. The Peaks-Over-Threshold method relies on the Poisson distribution for fitting the number of events and the GP distribution for fitting the size of exceedances.

EVT is very important in risk analysis where the result can have a large traumatic impact. Some real world applications include are not constrained to the parameters of this project, such as environmental and financial catastrophes. In the future, it would be interesting to continue the study of various situations like this apply various EVT tools. In conclusion, this project was only the basis for EVT, with a vast number of applications and outcomes to be further elaborated upon.

4.2 R Code/Output

Below is the supplementary R code used throughout this paper.

```
#GEV Distributions
library(evd)
library(fExtremes)
x <- seq(-10, 10, by=0.1)
Gumbel_density <- exp(-x-exp(-x))
Frechet_density <- dgev(x, xi=0.8, mu=0)
Weibull_density <- dgev(x, xi=-0.3, mu=0)
plot(c(x,x,x), c(Gumbel_density,Frechet_density, Weibull_density),
     type='n', xlab="x", ylab=" ",las=1)
lines(x, Gumbel_density, type='l', lty=1, col='green')
lines(x, Weibull_density, type='l', lty=2, col='blue')
lines(x, Frechet_density, type='l', lty=3, col='red')
legend('topright', legend=c('Gumbel','Frechet','Weibull'), lty=c(1,2,3), col=c('gree

#Gumbel Distribution
GumbelDistribution <- data.frame(x, Gumbel_density)
names(GumbelDistribution) <- c("Time", "Observations")
head(GumbelDistribution)
plot(GumbelDistribution, col="green", pch=19, cex=0.8,
     main="Plot of Extreme Value Distribution - Gumbel")

#Simulating the Gumbel Distribution in R
x <- seq(-10, 10, by=0.1)
Gumbel_density <- exp(-x-exp(-x))
GumbelDistribution <- data.frame(x, Gumbel_density)
names(GumbelDistribution) <- c("Time", "Observations")
head(GumbelDistribution)

#Frechet Distribution
FrechetDistribution <- data.frame(x, Frechet_density)
names(FrechetDistribution) <- c("Time", "Observations")
head(FrechetDistribution)
plot(FrechetDistribution, col="blue", pch=19, cex=0.8,
     main="Plot of Extreme Value Distribution - Frechet")

#Simulating the Frechet Distribution in R}
x <- seq(-10, 10, by=0.1)
Frechet_density <- dgev(x, xi=0.8, mu=0)
FrechetDistribution <- data.frame(x, Frechet_density)
names(FrechetDistribution) <- c("Time", "Observations")
head(FrechetDistribution)
```

```r
plot(FrechetDistribution, col="blue", pch=19, cex=0.8,
     main="Plot of Extreme Value Distribution - Frechet")

#Weibull Distribution
WeibullDistribution <- data.frame(x, Weibull_density)
names(WeibullDistribution) <- c("Time", "Observations")
head(WeibullDistribution)
plot(WeibullDistribution, col="red", pch=19, cex=0.8,
     main="Plot of Extreme Value Distribution - Weibull")

#Simulating the Weibull Distribution in R
x <- seq(-10, 10, by=0.1)
Weibull_density <- dgev(x, xi=-0.3, mu=0)
WeibullDistribution <- data.frame(x, Weibull_density)
names(WeibullDistribution) <- c("Time", "Observations")
head(WeibullDistribution)
plot(WeibullDistribution, col="red", pch=19, cex=0.8,
     main="Generated Plot of the Weibull Distribution")

#Block Maxima Method in R

set.seed(123)
library(extRemes)
## block size
n <- 12
original_mean <- 5
original_sd <- 2

## Create a series of maxima
series_length <- 200
maxima <- c()
## Simulate blocks of data
data_series <- list()
for (i in 1:series_length) {
  data_series[[i]] <- rnorm(n = n, mean = original_mean, sd = original_sd)
}
maxima <- unlist(lapply(data_series, max))
plot(maxima, main = "Block maxima", type = "l")

fit <- fevd(maxima, type = "Gumbel")
plot(fit, type = "density", main = "Empirical density vs estimated Gumbel distributi

#Peak-Over-Threshold Method in R
```

```
## Setting the threshold u
threshold <- 7.5
plot(x = rep(1:series_length, each = n), y = unlist(data_series), main = "Peak Over
      sub = paste("threshold =", threshold), xlab = "series", ylab = "value")
pot_points <- unlist(data_series) > threshold
points(x = rep(1:series_length, each = n)[pot_points], y = unlist(data_series)[pot_p
abline(h = threshold, col = "red")

gp_fit <- fevd(unlist(data_series), threshold = threshold, type = "GP")
plot(gp_fit, type = "density", main = "Empirical POT exceedances density vs estimate

plot(gp_fit, type = "rl", main = "Return level")
```

5 Bibliography

- Limit Distributions of Extreme Order Statistics under Power Normalization and Random Index by Z. Peng, Q. Jiang, S. Nadarajah

- A Simple Point Estimator of the Power of Moments by D. Li, A.Rosalsky, S. Zhang

- Evaluation of Peaks-Over-Threshold Method by Soheil Saeed Far and Ahmad Khairi Abd. Wahab

- Extremes by R.L. Wolpert

- Extreme Value Theory for Risk Factors in R by John Akwei

- Extreme Value Analysis by Shikun Li

- Extreme Value Theory: An Introduction by L. De Haan, A. Ferreira

- EXTREME VALUE THEORY: AN APPLICATION TO SPORTS by Sergio Luis Ganhao Vicente

- Extreme value distributions in R

- Fisher-Tippett Theorem with an Historical Perspective by A. Charpentier

- Operational Risk: Modeling Analytics by Harry H. Panjer

- Outline proof of the extreme value theorem in statistics