MAS3918/8918
Topics in Statistical Modelling: Extremes

Dr. Lee Fawcett, Dr. Vianey Palacios-Ramirez

Semester 2, 2024/25

Contact details

My name: Lee Fawcett

My job: Reader in Applied Statistics, Associate Dean (Global)

Office: Room 2.07 Herschel Building

Phone: 0191 208 7228

Email: `lee.fawcett@ncl.ac.uk`

Timetable and Administrative arrangements

- **Lectures** for the Extremes topic are on **Mondays** at **4pm** (KGVI LT6) and again on **Thursdays** at **2pm** (KGVI LT1).
- Every fortnight, starting this week, there will also be a session on **Fridays** at **12pm** (Herschel LT2); these will usualy be used as **problems classes**.
- **Office hours**: **Wednesdays 2pm–4pm**.

Timetable and Administrative arrangements

- **Lectures** for the Extremes topic are on **Mondays** at **4pm** (KGVI LT6) and again on **Thursdays** at **2pm** (KGVI LT1).
- Every fortnight, starting this week, there will also be a session on **Fridays** at **12pm** (Herschel LT2); these will usualy be used as **problems classes**.
- **Office hours**: **Wednesdays 2pm–4pm**.

Assessment and deadlines

Assessment	Weighting	Date/Deadline
Homework 1	6% (3% from each topic)	4pm, Friday 21st February
Homework 2	7% (3.5% from each topic)	4pm, Friday 7th March
Homework 3	7% (3.5% from each topic)	4pm, Friday 28th March
2.5-hour exam	80% (40% from questions from each topic)	May/June 2025 (TBC)

For this topic, there may be the option to switch out the three assignments for a single project – stay tuned and I'll let you know soon!

Other stuff

All lecture notes will be available in advance on the MAS3918/MAS8918 Canvas page.

Notes will be available in two formats:

- HTML screen-reader-friendly, **accessible lecture notes**
- **Traditional PDF lecture notes** for printing:
  - With gaps, for students who like to write stuff down during lectures
  - With the gaps filled in (in red font)

Other stuff

- A summarised verion of the lecture notes will be used as **slides** in the lectures
- These slides will be made available via Canvas after the lecture, along with the associated **ReCap recordings**
- **Solutions** to problem-solving exercises will be made available via Canvas after the deadline for each assessment has passed
- One last thing... There is an excellent **course textbook** available to buy, or to loan from the University library:

  An Introduction to Statistical Modeling of Extreme Values, by Stuart Coles.

History of this course

- MAS369: **1995–2000** – Dr. Mark Dixon – now the multi–millionaire founder of sports betting company ATASS
- MAS369: resurrected in **2002** for one year only – Dave Walshaw (see OHP slides...)
- Course: **Reading module** for Stage 3 MMath/MMathStat students (and select BSc students!)
- New Stage 4 module in **2011**: material in MAS369 out–of–date with current research – a more advanced "topic" in MAS8306
- Now we want to bring some of this material to stage 3 students – very timely, given the current climate emergency

History of this course

- MAS369: **1995–2000** – Dr. Mark Dixon – now the multi–millionaire founder of sports betting company ATASS
- MAS369: resurrected in **2002** for one year only – Dave Walshaw (see OHP slides...)
- Course: **Reading module** for Stage 3 MMath/MMathStat students (and select BSc students!)
- New Stage 4 module in **2011**: material in MAS369 out–of–date with current research – a more advanced "topic" in MAS8306
- Now we want to bring some of this material to stage 3 students – very timely, given the current climate emergency

History of this course

- MAS369: **1995–2000** – Dr. Mark Dixon – now the multi–millionaire founder of sports betting company ATASS
- MAS369: resurrected in **2002** for one year only – Dave Walshaw (see OHP slides...)
- Course: **Reading module** for Stage 3 MMath/MMathStat students (and select BSc students!)
- New Stage 4 module in **2011**: material in MAS369 out–of–date with current research – a more advanced "topic" in MAS8306
- Now we want to bring some of this material to stage 3 students – very timely, given the current climate emergency

History of this course

- MAS369: **1995–2000** – Dr. Mark Dixon – now the multi–millionaire founder of sports betting company ATASS
- MAS369: resurrected in **2002** for one year only – Dave Walshaw (see OHP slides...)
- Course: **Reading module** for Stage 3 MMath/MMathStat students (and select BSc students!)
- New Stage 4 module in **2011**: material in MAS369 out–of–date with current research – a more advanced "topic" in MAS8306
- Now we want to bring some of this material to stage 3 students – very timely, given the current climate emergency

Aims of this course

- To expose you to an area of **active research** in the School
- To expose you to a very **hands–on**, applied area of Statistics
- To give you a **firm grounding** in the **basics** of techniques for analysing environmental data...
- ... but also to bring you bang up–to–date with **current research ideas**
- Quite a níche area of research, but along the way we will have to get to grips with some more generic statistical tools

Aims of this course

- To expose you to an area of **active research** in the School
- To expose you to a very **hands–on**, applied area of Statistics
- To give you a **firm grounding** in the **basics** of techniques for analysing environmental data...
- ... but also to bring you bang up–to–date with **current research ideas**
- Quite a níche area of research, but along the way we will have to get to grips with some more generic statistical tools

Aims of this course

- To expose you to an area of **active research** in the School
- To expose you to a very **hands–on**, applied area of Statistics
- To give you a **firm grounding** in the **basics** of techniques for analysing environmental data...
- ... but also to bring you bang up–to–date with **current research ideas**
- Quite a níche area of research, but along the way we will have to get to grips with some more generic statistical tools

A word on pre–requisite knowledge

- Statistical "Theory": I will not assume any knowledge past 2nd year Statistics courses (**MAS2901** and **MAS2906** in particular)
- Statistical "Practice": Some basic **Statistical Modelling** techniques will be used, but will be fully demonstrated if they go beyond MAS2901/MAS2902
- We will make use of some techniques from these courses (e.g. **maximum likelihood**) – these will be quickly reviewed, as–and–when they make an appearance, but you may want to consult your old notes for further details

A word on pre–requisite knowledge

- Statistical "Theory": I will not assume any knowledge past 2nd year Statistics courses (**MAS2901** and **MAS2906** in particular)
- Statistical "Practice": Some basic **Statistical Modelling** techniques will be used, but will be fully demonstrated if they go beyond MAS2901/MAS2902
- We will make use of some techniques from these courses (e.g. **maximum likelihood**) – these will be quickly reviewed, as–and–when they make an appearance, but you may want to consult your old notes for further details

Let's start with some (non–mathematical) background and motivation...

Chapter 1

Introduction

Practical Motivation

Finally, there is almost a **global consensus** amongst scientists that our planet's climate is changing.

Evidence for climatic change has been collected from a variety of sources, some of which can be used to reconstruct the earth's changing climates over **tens of thousands of years**.

Reasonable complete global records of the earth's surface temperature since the mid–1800's indicate a **positive trend** in the average annual temperature, and maximum annual temperature, most noticeable at the earth's poles.

Practical Motivation

Finally, there is almost a **global consensus** amongst scientists that our planet's climate is changing.

Evidence for climatic change has been collected from a variety of sources, some of which can be used to reconstruct the earth's changing climates over **tens of thousands of years**.

Reasonable complete global records of the earth's surface temperature since the mid–1800's indicate a **positive trend** in the average annual temperature, and maximum annual temperature, most noticeable at the earth's poles.

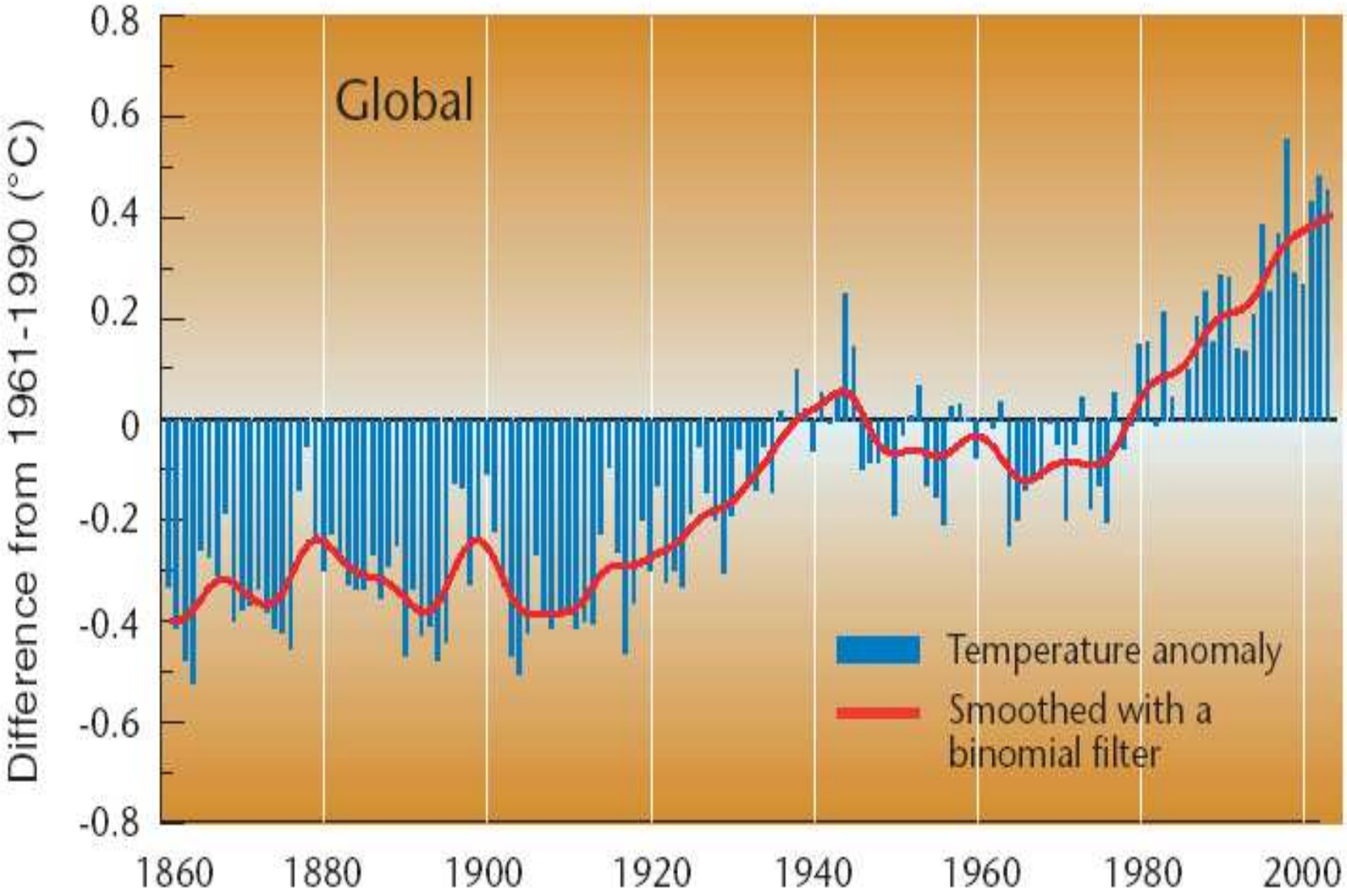
Practical Motivation

Finally, there is almost a **global consensus** amongst scientists that our planet's climate is changing.

Evidence for climatic change has been collected from a variety of sources, some of which can be used to reconstruct the earth's changing climates over **tens of thousands of years**.

Reasonable complete global records of the earth's surface temperature since the mid–1800's indicate a **positive trend** in the average annual temperature, and maximum annual temperature, most noticeable at the earth's poles.

Practical Motivation



Practical Motivation

Glaciers are considered the **most sensitive indicators** of climate change.

As the earth warms, glaciers **retreat** and **ice sheets melt**.

Over the last 30 years or so this has resulted in a gradual **increase in sea and ocean levels**.

Practical Motivation

Glaciers are considered the **most sensitive indicators** of climate change.

As the earth warms, glaciers **retreat** and **ice sheets melt**.

Over the last 30 years or so this has resulted in a gradual **increase in sea and ocean levels**.

Practical Motivation

Glaciers are considered the **most sensitive indicators** of climate change.

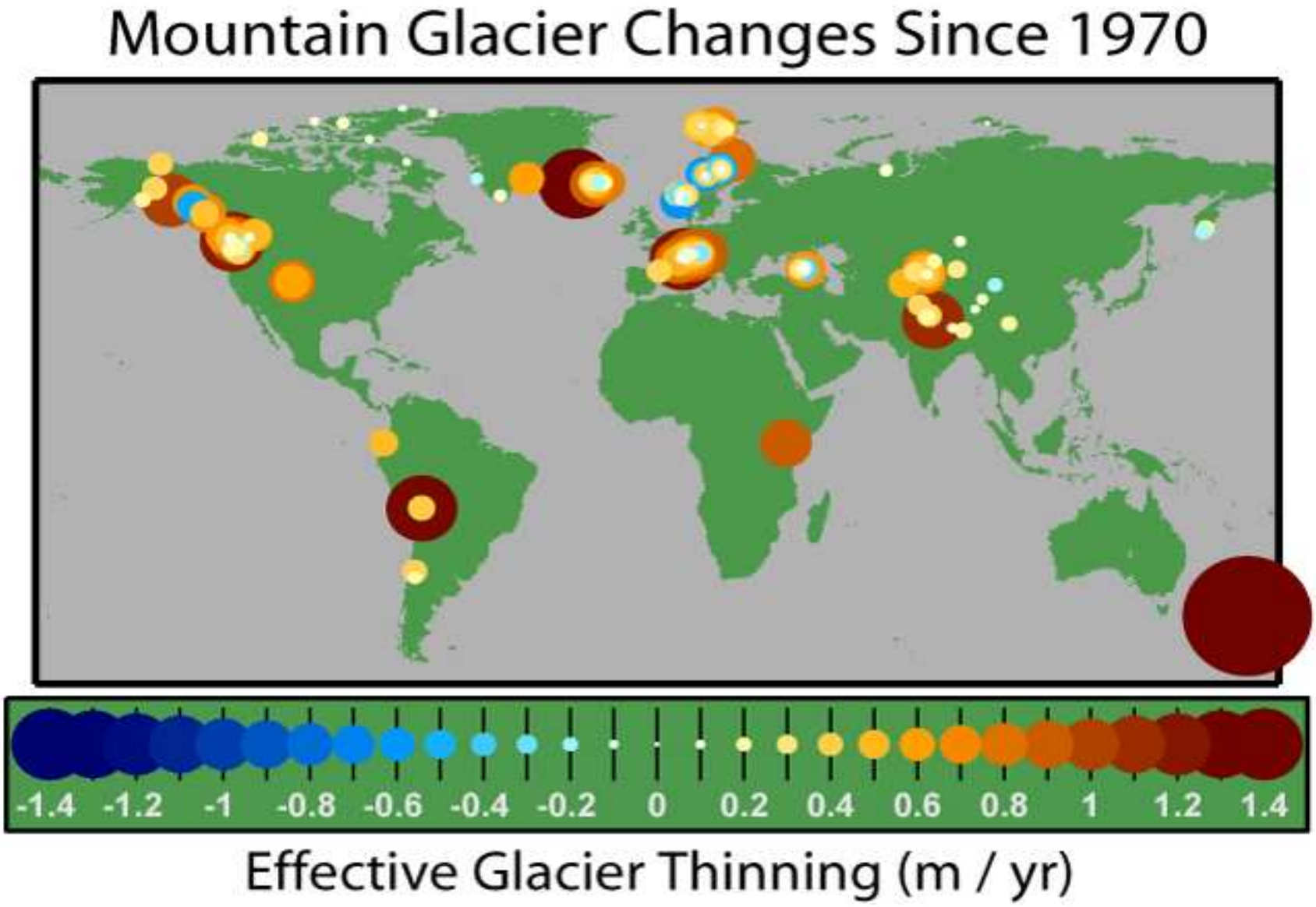
As the earth warms, glaciers **retreat** and **ice sheets melt**.

Over the last 30 years or so this has resulted in a gradual **increase in sea and ocean levels**.

1977

2006

Practical Motivation



Practical Motivation

Consequences

- Irreversible changes to **sea** and **ocean ecosystems**
- Threat of **flooding** to low–lying inhabited areas of land
- Effect on the earth's **weather systems** – for example:
  - Larger amount of warmer water in North Atlantic Ocean $\longrightarrow$ Stronger, more frequent tropical storms/hurricanes (e.g. Katrina, Helene, Milton)
  - More frequent tropical cyclones in the Indian Ocean/Pacific Ocean (e.g. Bangladesh, India, Malaysia)
  - Flooding in the U.K. – 18-month period from October 2022 wettest on record
  - Named storms in the U.K. – e.g. Storm Darragh and **Storm Éowyn**

Practical Motivation

Consequences

- Irreversible changes to **sea** and **ocean ecosystems**
- Threat of **flooding** to low–lying inhabited areas of land
- Effect on the earth's **weather systems** – for example:
  - Larger amount of warmer water in North Atlantic Ocean $\longrightarrow$ Stronger, more frequent tropical storms/hurricanes (e.g. Katrina, Helene, Milton)
  - More frequent tropical cyclones in the Indian Ocean/Pacific Ocean (e.g. Bangladesh, India, Malaysia)
  - Flooding in the U.K. – 18-month period from October 2022 wettest on record
  - Named storms in the U.K. – e.g. Storm Darragh and **Storm Éowyn**

Practical Motivation

Consequences

- Irreversible changes to **sea** and **ocean ecosystems**
- Threat of **flooding** to low–lying inhabited areas of land
- Effect on the earth's **weather systems** – for example:
  - Larger amount of warmer water in North Atlantic Ocean $\longrightarrow$ Stronger, more frequent tropical storms/hurricanes (e.g. Katrina, Helene, Milton)
  - More frequent tropical cyclones in the Indian Ocean/Pacific Ocean (e.g. Bangladesh, India, Malaysia)
  - Flooding in the U.K. – 18-month period from October 2022 wettest on record
  - Named storms in the U.K. – e.g. Storm Darragh and **Storm Éowyn**

Practical Motivation

Consequences

- Irreversible changes to **sea** and **ocean ecosystems**
- Threat of **flooding** to low–lying inhabited areas of land
- Effect on the earth's **weather systems** – for example:
  - Larger amount of warmer water in North Atlantic Ocean $\longrightarrow$ Stronger, more frequent tropical storms/hurricanes (e.g. Katrina, Helene, Milton)
  - More frequent tropical cyclones in the Indian Ocean/Pacific Ocean (e.g. Bangladesh, India, Malaysia)
  - Flooding in the U.K. – 18-month period from October 2022 wettest on record
  - Named storms in the U.K. – e.g. Storm Darragh and **Storm Éowyn**

Practical Motivation

Consequences

- Irreversible changes to **sea** and **ocean ecosystems**
- Threat of **flooding** to low–lying inhabited areas of land
- Effect on the earth's **weather systems** – for example:
  - Larger amount of warmer water in North Atlantic Ocean $\longrightarrow$ Stronger, more frequent tropical storms/hurricanes (e.g. Katrina, Helene, Milton)
  - More frequent tropical cyclones in the Indian Ocean/Pacific Ocean (e.g. Bangladesh, India, Malaysia)
  - Flooding in the U.K. – 18-month period from October 2022 wettest on record
  - Named storms in the U.K. – e.g. Storm Darragh and **Storm Éowyn**

Practical Motivation

Consequences

- Irreversible changes to **sea** and **ocean ecosystems**
- Threat of **flooding** to low–lying inhabited areas of land
- Effect on the earth's **weather systems** – for example:
  - Larger amount of warmer water in North Atlantic Ocean $\longrightarrow$ Stronger, more frequent tropical storms/hurricanes (e.g. Katrina, Helene, Milton)
  - More frequent tropical cyclones in the Indian Ocean/Pacific Ocean (e.g. Bangladesh, India, Malaysia)
  - Flooding in the U.K. – 18-month period from October 2022 wettest on record
  - Named storms in the U.K. – e.g. Storm Darragh and **Storm Éowyn**

Practical Motivation

Consequences

- Irreversible changes to **sea** and **ocean ecosystems**
- Threat of **flooding** to low–lying inhabited areas of land
- Effect on the earth's **weather systems** – for example:
  - Larger amount of warmer water in North Atlantic Ocean $\longrightarrow$ Stronger, more frequent tropical storms/hurricanes (e.g. Katrina, Helene, Milton)
  - More frequent tropical cyclones in the Indian Ocean/Pacific Ocean (e.g. Bangladesh, India, Malaysia)
  - Flooding in the U.K. – 18-month period from October 2022 wettest on record
  - Named storms in the U.K. – e.g. Storm Darragh and **Storm Éowyn**

CLASSES
ASS VIII-XII

“Toon Monsoon”, June 2012

“Toon Monsoon”, June 2012

“Toon Monsoon”, June 2012

“Toon Monsoon”, June 2012

The Great North Sea Flood, 1953

The Great North Sea Flood, 2025?

The Great Storm of 1987

- 15/16 October 1987
- Killed 22 people in England/France
- Met Office criticism:
  - Recent cutbacks – no weather ship in the Southwest Approaches
  - Weather presenter Michael Fish told the British public not to worry, and that there is no "hurricane on the way"

The Great Storm of 1987

- 15/16 October 1987
- Killed 22 people in England/France
- Met Office criticism:
  - Recent cutbacks – no weather ship in the Southwest Approaches
  - Weather presenter Michael Fish told the British public not to worry, and that there is no "hurricane on the way"

The Great Storm of 1987

- 15/16 October 1987
- Killed 22 people in England/France
- Met Office criticism:
  - Recent cutbacks – no weather ship in the Southwest Approaches
  - Weather presenter Michael Fish told the British public not to worry, and that there is no "hurricane on the way"

The Great Storm of 1987

- 15/16 October 1987
- Killed 22 people in England/France
- Met Office criticism:
  - Recent cutbacks – no weather ship in the Southwest Approaches
  - Weather presenter Michael Fish told the British public not to worry, and that there is no "hurricane on the way"

The Great Storm of 1987

- 15/16 October 1987
- Killed 22 people in England/France
- Met Office criticism:
  - Recent cutbacks – no weather ship in the Southwest Approaches
  - Weather presenter Michael Fish told the British public not to worry, and that there is no "hurricane on the way"

Tornadoes in Birmingham, 2005

Tornadoes in Birmingham, 2005

Practical Motivation

The world's changing climate is also likely to give rise to an increase in the frequency of periods of intense **heat** and **drought**.

- Some parts of the world are familiar with the effects of extreme heat...
- ... but we are beginning to see an increase in severe hot spells in more temperate latitudes. For example, the **2003 and 2022 European heatwaves** which resulted in thousands of deaths
- Increase in devastating forest fires (**New South Wales** and **Victoria** in Australia, and of course more recently in **LA**)
- Increase in famine

Practical Motivation

The world's changing climate is also likely to give rise to an increase in the frequency of periods of intense **heat** and **drought**.

- Some parts of the world are familiar with the effects of extreme heat...
- ... but we are beginning to see an increase in severe hot spells in more temperate latitudes. For example, the **2003 and 2022 European heatwaves** which resulted in thousands of deaths
- Increase in devastating forest fires (**New South Wales** and **Victoria** in Australia, and of course more recently in **LA**)
- Increase in famine

Practical Motivation

The world's changing climate is also likely to give rise to an increase in the frequency of periods of intense **heat** and **drought**.

- Some parts of the world are familiar with the effects of extreme heat...
- ... but we are beginning to see an increase in severe hot spells in more temperate latitudes. For example, the **2003 and 2022 European heatwaves** which resulted in thousands of deaths
- Increase in devastating forest fires (**New South Wales** and **Victoria** in Australia, and of course more recently in **LA**)
- Increase in famine

Practical Motivation

The world's changing climate is also likely to give rise to an increase in the frequency of periods of intense **heat** and **drought**.

- Some parts of the world are familiar with the effects of extreme heat...
- ... but we are beginning to see an increase in severe hot spells in more temperate latitudes. For example, the **2003 and 2022 European heatwaves** which resulted in thousands of deaths
- Increase in devastating forest fires (**New South Wales** and **Victoria** in Australia, and of course more recently in **LA**)
- Increase in famine

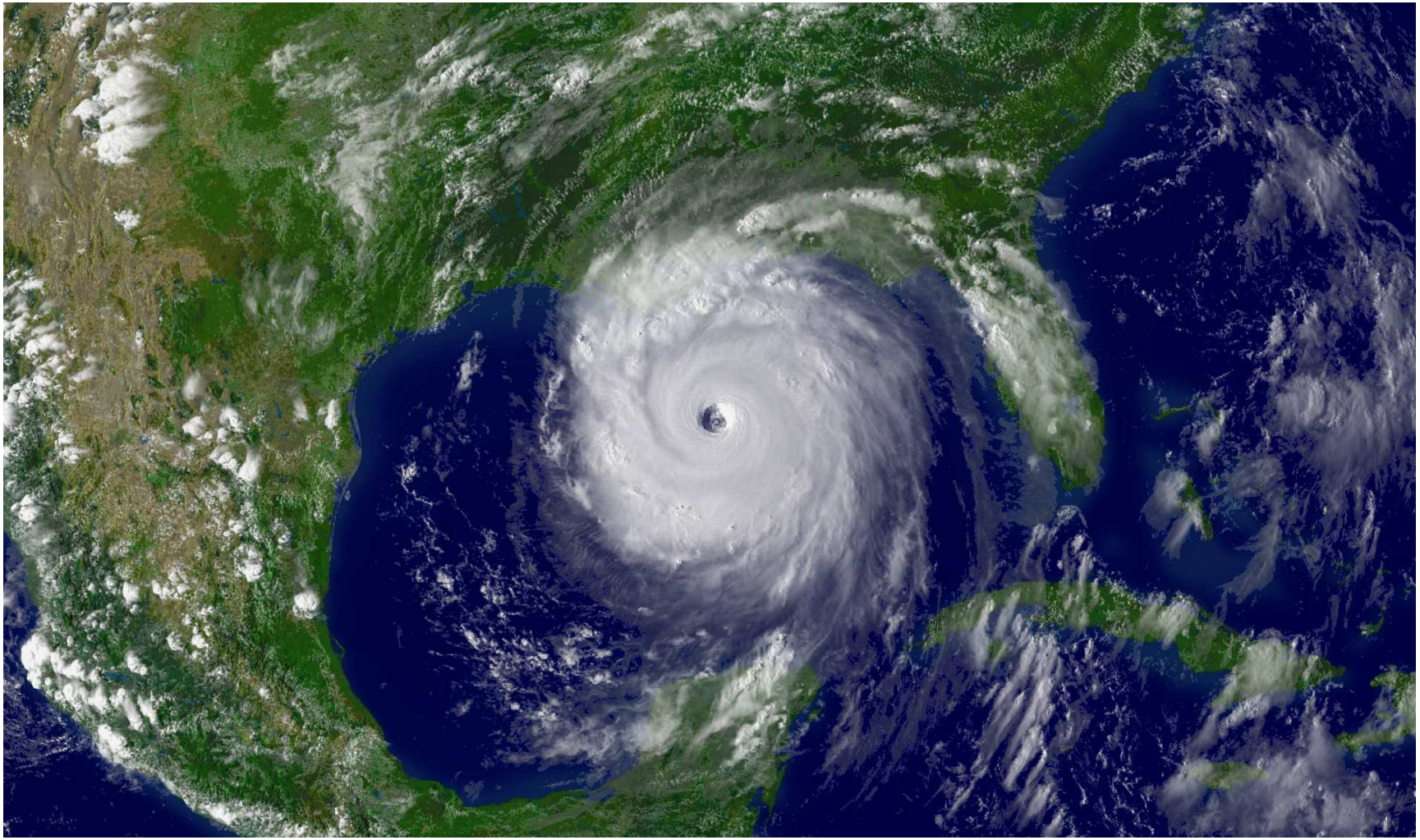
Practical Motivation

The world's changing climate is also likely to give rise to an increase in the frequency of periods of intense **heat** and **drought**.

- Some parts of the world are familiar with the effects of extreme heat...
- ... but we are beginning to see an increase in severe hot spells in more temperate latitudes. For example, the **2003 and 2022 European heatwaves** which resulted in thousands of deaths
- Increase in devastating forest fires (**New South Wales** and **Victoria** in Australia, and of course more recently in **LA**)
- Increase in famine

Schweppes
PRUÉBALA SOLA
REGÍCLAME
JCDecaux
47º

Hurricane Katrina



Hurricane Katrina

Some facts

- **Category 5** Hurricane (Category 3 when it made landfall)
- Killed **1833** people
- Caused **$108 bn** worth of damage
- Most damage/loss of life caused by **storm surge**
- Storm surge reached nearly **14 feet** above sea level
- Lowest air pressure 902 mb
- Political controversy
- Billed as the "**storm of the century**"

Hurricane Katrina

Some facts

- **Category 5** Hurricane (Category 3 when it made landfall)
- Killed **1833** people
- Caused **$108 bn** worth of damage
- Most damage/loss of life caused by **storm surge**
- Storm surge reached nearly **14 feet** above sea level
- Lowest air pressure 902 mb
- Political controversy
- Billed as the "**storm of the century**"

Hurricane Katrina

Some facts

- **Category 5** Hurricane (Category 3 when it made landfall)
- Killed **1833** people
- Caused **$108 bn** worth of damage
- Most damage/loss of life caused by **storm surge**
- Storm surge reached nearly **14 feet** above sea level
- Lowest air pressure 902 mb
- Political controversy
- Billed as the "**storm of the century**"

Hurricane Katrina

Some facts

- **Category 5** Hurricane (Category 3 when it made landfall)
- Killed **1833** people
- Caused **$108 bn** worth of damage
- Most damage/loss of life caused by **storm surge**
- Storm surge reached nearly **14 feet** above sea level
- Lowest air pressure 902 mb
- Political controversy
- Billed as the "**storm of the century**"

Hurricane Katrina

Some facts

- **Category 5** Hurricane (Category 3 when it made landfall)
- Killed **1833** people
- Caused **$108 bn** worth of damage
- Most damage/loss of life caused by **storm surge**
- Storm surge reached nearly **14 feet** above sea level
- Lowest air pressure 902 mb
- Political controversy
- Billed as the "**storm of the century**"

Hurricane Katrina

Some facts

- **Category 5** Hurricane (Category 3 when it made landfall)
- Killed **1833** people
- Caused **$108 bn** worth of damage
- Most damage/loss of life caused by **storm surge**
- Storm surge reached nearly **14 feet** above sea level
- Lowest air pressure 902 mb
- Political controversy
- Billed as the "**storm of the century**"

Hurricane Katrina

Some facts

- **Category 5** Hurricane (Category 3 when it made landfall)
- Killed **1833** people
- Caused **$108 bn** worth of damage
- Most damage/loss of life caused by **storm surge**
- Storm surge reached nearly **14 feet** above sea level
- Lowest air pressure 902 mb
- Political controversy
- Billed as the "**storm of the century**"

Hurricane Katrina

Some facts

- **Category 5** Hurricane (Category 3 when it made landfall)
- Killed **1833** people
- Caused **$108 bn** worth of damage
- Most damage/loss of life caused by **storm surge**
- Storm surge reached nearly **14 feet** above sea level
- Lowest air pressure 902 mb
- Political controversy
- Billed as the "**storm of the century**"

Hurricane Katrina

Some facts

- **Category 5** Hurricane (Category 3 when it made landfall)
- Killed **1833** people
- Caused **$108 bn** worth of damage
- Most damage/loss of life caused by **storm surge**
- Storm surge reached nearly **14 feet** above sea level
- Lowest air pressure 902 mb
- Political controversy
- Billed as the "**storm of the century**"

Hurricane Katrina

Hurricane Katrina

Hurricane Katrina

So why statistical modelling?

Statistical modelling of extreme weather has a very practical motivation: **reliability** —

— anything we build needs to have a good chance of surviving the weather/environment for the whole of its working life.

This has obvious implications for **civil engineers** and **planners**.

So why statistical modelling?

Statistical modelling of extreme weather has a very practical motivation: **reliability** —

— anything we build needs to have a good chance of surviving the weather/environment for the whole of its working life.

This has obvious implications for **civil engineers** and **planners**.

So why statistical modelling?

Statistical modelling of extreme weather has a very practical motivation: **reliability** —

— anything we build needs to have a good chance of surviving the weather/environment for the whole of its working life.

This has obvious implications for **civil engineers** and **planners**.

So why statistical modelling?

For example, in the context of some of the examples we have looked at so far, they need to know:

- How tall to build sea walls/flood defences – for example –
  - the levees surrounding New Orleans or Cedar Key, Florida;
  - flood defences in New York that were breached during Superstorm Sandy;
  - river and sea defences in the U.K.;
- How strong to make buildings – for example, in the U.K. the *British Standards Institute* specifies that new structures need to be strong enough to withstand the "**Fifty year wind speed**";
- How tall to build reservoir dams;
- How much fuel to stockpile.

So why statistical modelling?

For example, in the context of some of the examples we have looked at so far, they need to know:

- How tall to build sea walls/flood defences – for example –
  - the levees surrounding New Orleans or Cedar Key, Florida;
  - flood defences in New York that were breached during Superstorm Sandy;
  - river and sea defences in the U.K.;
- How strong to make buildings – for example, in the U.K. the *British Standards Institute* specifies that new structures need to be strong enough to withstand the "**Fifty year wind speed**";
- How tall to build reservoir dams;
- How much fuel to stockpile.

So why statistical modelling?

For example, in the context of some of the examples we have looked at so far, they need to know:

- How tall to build sea walls/flood defences – for example –
  - the levees surrounding New Orleans or Cedar Key, Florida;
  - flood defences in New York that were breached during Superstorm Sandy;
  - river and sea defences in the U.K.;
- How strong to make buildings – for example, in the U.K. the *British Standards Institute* specifies that new structures need to be strong enough to withstand the "**Fifty year wind speed**";
- How tall to build reservoir dams;
- How much fuel to stockpile.

So why statistical modelling?

For example, in the context of some of the examples we have looked at so far, they need to know:

- How tall to build sea walls/flood defences – for example –
  - the levees surrounding New Orleans or Cedar Key, Florida;
  - flood defences in New York that were breached during Superstorm Sandy;
  - river and sea defences in the U.K.;
- How strong to make buildings – for example, in the U.K. the *British Standards Institute* specifies that new structures need to be strong enough to withstand the "**Fifty year wind speed**";
- How tall to build reservoir dams;
- How much fuel to stockpile.

So why statistical modelling?

For example, in the context of some of the examples we have looked at so far, they need to know:

- How tall to build sea walls/flood defences – for example –
  - the levees surrounding New Orleans or Cedar Key, Florida;
  - flood defences in New York that were breached during Superstorm Sandy;
  - river and sea defences in the U.K.;
- How strong to make buildings – for example, in the U.K. the *British Standards Institute* specifies that new structures need to be strong enough to withstand the "**Fifty year wind speed**";
- How tall to build reservoir dams;
- How much fuel to stockpile.

So why statistical modelling?

For example, in the context of some of the examples we have looked at so far, they need to know:

- How tall to build sea walls/flood defences – for example –
  - the levees surrounding New Orleans or Cedar Key, Florida;
  - flood defences in New York that were breached during Superstorm Sandy;
  - river and sea defences in the U.K.;
- How strong to make buildings – for example, in the U.K. the *British Standards Institute* specifies that new structures need to be strong enough to withstand the "**Fifty year wind speed**";
- How tall to build reservoir dams;
- How much fuel to stockpile.

So why statistical modelling?

For example, in the context of some of the examples we have looked at so far, they need to know:

- How tall to build sea walls/flood defences – for example –
  - the levees surrounding New Orleans or Cedar Key, Florida;
  - flood defences in New York that were breached during Superstorm Sandy;
  - river and sea defences in the U.K.;
- How strong to make buildings – for example, in the U.K. the *British Standards Institute* specifies that new structures need to be strong enough to withstand the "**Fifty year wind speed**";
- How tall to build reservoir dams;
- How much fuel to stockpile.

So why statistical modelling?

For example, in the context of some of the examples we have looked at so far, they need to know:

- How tall to build sea walls/flood defences – for example –
  - the levees surrounding New Orleans or Cedar Key, Florida;
  - flood defences in New York that were breached during Superstorm Sandy;
  - river and sea defences in the U.K.;
- How strong to make buildings – for example, in the U.K. the *British Standards Institute* specifies that new structures need to be strong enough to withstand the "**Fifty year wind speed**";
- How tall to build reservoir dams;
- How much fuel to stockpile.

So why statistical modelling?

This motivates the need to estimate what the:

- Highest tide;
- Heaviest rainfall;
- Strongest wind;
- most severe cold spell;

will be over some fixed period of future time.

So why statistical modelling?

This motivates the need to estimate what the:

- Highest tide;
- Heaviest rainfall;
- Strongest wind;
- most severe cold spell;

will be over some fixed period of future time.

So why statistical modelling?

This motivates the need to estimate what the:

- Highest tide;
- Heaviest rainfall;
- Strongest wind;
- most severe cold spell;

will be over some fixed period of future time.

So why statistical modelling?

The only sensible way to do this is to use **data** on the variable of interest (wind, rain etc.) and fit an appropriate **statistical model**.

The models themselves are motivated by **asymptotic theory**, and this will be our starting point in Chapter 2.

Before this, let's start at a very basic level, and think about why we need statistical models at all.

So why statistical modelling?

The only sensible way to do this is to use **data** on the variable of interest (wind, rain etc.) and fit an appropriate **statistical model**.

The models themselves are motivated by **asymptotic theory**, and this will be our starting point in Chapter 2.

Before this, let's start at a very basic level, and think about why we need statistical models at all.

So why statistical modelling?

The only sensible way to do this is to use **data** on the variable of interest (wind, rain etc.) and fit an appropriate **statistical model**.

The models themselves are motivated by **asymptotic theory**, and this will be our starting point in Chapter 2.

Before this, let's start at a very basic level, and think about why we need statistical models at all.

Simple example: sea surges at Wassaw Island

Wassaw Island is a barrier island off the southeast coast of Georgia, in the USA.

Any hurricanes about to make landfall on the mainland often hit such barrier islands first, substantially subduing the impending storm.

Sea surge data is collected at Wassaw Island every hour; the data shown below are the extracted **annual maxima** for the years 1975–2024.

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

Simple example: sea surges at Wassaw Island

Wassaw Island is a barrier island off the southeast coast of Georgia, in the USA.

Any hurricanes about to make landfall on the mainland often hit such barrier islands first, substantially subduing the impending storm.

Sea surge data is collected at Wassaw Island every hour; the data shown below are the extracted **annual maxima** for the years 1975–2024.

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

Simple example: sea surges at Wassaw Island

Wassaw Island is a barrier island off the southeast coast of Georgia, in the USA.

Any hurricanes about to make landfall on the mainland often hit such barrier islands first, substantially subduing the impending storm.

Sea surge data is collected at Wassaw Island every hour; the data shown below are the extracted **annual maxima** for the years 1975–2024.

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

Simple example: sea surges at Wassaw Island

Wassaw Island is a barrier island off the southeast coast of Georgia, in the USA.

Any hurricanes about to make landfall on the mainland often hit such barrier islands first, substantially subduing the impending storm.

Sea surge data is collected at Wassaw Island every hour; the data shown below are the extracted **annual maxima** for the years 1975–2024.

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

Simple example: sea surges at Wassaw Island

1. Civil engineers are interested in various exceedance probabilities.

   For example, suppose they need to know the probability that sea–surge will exceed 8.75, 11.25 and 14 feet. Use the data to estimate these probabilities.

Solution

Estimating these probabilities using relative frequencies gives:

$$\begin{aligned}
\text{Pr(sea–surge exceeds 8.75 feet)} &= \frac{33}{50} = 0.66 \\
\text{Pr(sea–surge exceeds 11.25 feet)} &= \frac{6}{50} = 0.12 \\
\text{Pr(sea–surge exceeds 14 feet)} &= \frac{0}{50} = 0.
\end{aligned}$$

Do you *really* think it's *impossible* for sea–surge to exceed 14 feet?

Solution

Estimating these probabilities using relative frequencies gives:

$$\begin{aligned}
\text{Pr(sea–surge exceeds 8.75 feet)} &= \frac{33}{50} = 0.66 \\
\text{Pr(sea–surge exceeds 11.25 feet)} &= \frac{6}{50} = 0.12 \\
\text{Pr(sea–surge exceeds 14 feet)} &= \frac{0}{50} = 0.
\end{aligned}$$

Do you *really* think it's *impossible* for sea–surge to exceed 14 feet?

Solution

Estimating these probabilities using relative frequencies gives:

$$\begin{aligned}
\text{Pr(sea–surge exceeds 8.75 feet)} &= \frac{33}{50} = 0.66 \\
\text{Pr(sea–surge exceeds 11.25 feet)} &= \frac{6}{50} = 0.12 \\
\text{Pr(sea–surge exceeds 14 feet)} &= \frac{0}{50} = 0.
\end{aligned}$$

Do you *really* think it's *impossible* for sea–surge to exceed 14 feet?

Solution

Estimating these probabilities using relative frequencies gives:

$$\begin{aligned}
\text{Pr(sea–surge exceeds 8.75 feet)} &= \frac{33}{50} = 0.66 \\
\text{Pr(sea–surge exceeds 11.25 feet)} &= \frac{6}{50} = 0.12 \\
\text{Pr(sea–surge exceeds 14 feet)} &= \frac{0}{50} = 0.
\end{aligned}$$

Do you *really* think it's *impossible* for sea–surge to exceed 14 feet?

Simple example: sea surges at Wassaw Island

2. How high should state authorities build flood defences at the nearby city of Savannah if they want to protect the city against the sea–surge they would expect to see:
   (i) Once every 10 years;
   (ii) Once every hundred years?

Solution (1/2)

In ascending numerical order, the data are:

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

Since we have annual data, an intuitive argument would be to find the value which has 10% of the data above it.

We have 50 observations, and 10% of these observations exceed $11.7999 \approx 11.8$ feet.

Solution (1/2)

In ascending numerical order, the data are:

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

Since we have annual data, an intuitive argument would be to find the value which has 10% of the data above it.

We have 50 observations, and 10% of these observations exceed $11.7999 \approx 11.8$ feet.

Solution (1/2)

In ascending numerical order, the data are:

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

Since we have annual data, an intuitive argument would be to find the value which has 10% of the data above it.

We have 50 observations, and 10% of these observations exceed $11.7999 \approx 11.8$ feet.

Solution (2/2)

For the level which is exceeded once every **hundred** years, we would need the value which has 1% of the data above it.

We don't have enough data – we only have 50 years of data, and we want the value that is exceeded once every *hundred* years! 1% of 50 = 1/2 an observation!

Solution (2/2)

For the level which is exceeded once every **hundred** years, we would need the value which has 1% of the data above it.

We don't have enough data – we only have 50 years of data, and we want the value that is exceeded once every *hundred* years! 1% of 50 = 1/2 an observation!

Further comments

We have no chance of producing any sort of estimate that goes beyond the timeframe of our collected data – however, it is **vitally important** for engineers that we do estimate such quantities!

We are trying to estimate events more extreme than have ever occurred before – an impossible task?

Not *impossible*... To extrapolate beyond the range of our observed data, we need a **statistical model** that we can assume is valid for our observed extremes *and beyond*.

This is where **extreme value theory** comes in!

Further comments

We have no chance of producing any sort of estimate that goes beyond the timeframe of our collected data – however, it is **vitally important** for engineers that we do estimate such quantities!

We are trying to estimate events more extreme than have ever occurred before – an impossible task?

Not *impossible*... To extrapolate beyond the range of our observed data, we need a **statistical model** that we can assume is valid for our observed extremes *and beyond*.

This is where **extreme value theory** comes in!

Further comments

We have no chance of producing any sort of estimate that goes beyond the timeframe of our collected data – however, it is **vitally important** for engineers that we do estimate such quantities!

We are trying to estimate events more extreme than have ever occurred before – an impossible task?

Not *impossible*... To extrapolate beyond the range of our observed data, we need a **statistical model** that we can assume is valid for our observed extremes *and beyond*.

This is where **extreme value theory** comes in!

Further comments

We have no chance of producing any sort of estimate that goes beyond the timeframe of our collected data – however, it is **vitally important** for engineers that we do estimate such quantities!

We are trying to estimate events more extreme than have ever occurred before – an impossible task?

Not *impossible*... To extrapolate beyond the range of our observed data, we need a **statistical model** that we can assume is valid for our observed extremes *and beyond*.

This is where **extreme value theory** comes in!

Chapter 2

Classical Models for Extremes

Background and theoretical motivation

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

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

$$M_n = \max\{X_1, X_2, \ldots, X_n\}. \tag{1}$$

Background and theoretical motivation

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

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

$$M_n = \max\{X_1, X_2, \ldots, X_n\}. \tag{1}$$

Background and theoretical motivation

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

Sea surge measurements were taken every hour; for each year (1975–2024) you were presented with the **annual maximum** sea surge.

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

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

to denote, generally, the 50 annual maxima given in Table 1 (of course, for leap years the notation would be $M_{8784,i}$).

Background and theoretical motivation

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

Sea surge measurements were taken every hour; for each year (1975–2024) you were presented with the **annual maximum** sea surge.

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

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

to denote, generally, the 50 annual maxima given in Table 1 (of course, for leap years the notation would be $M_{8784,i}$).

Background and theoretical motivation

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

Sea surge measurements were taken every hour; for each year (1975–2024) you were presented with the **annual maximum** sea surge.

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

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

to denote, generally, the 50 annual maxima given in Table 1 (of course, for leap years the notation would be $M_{8784,i}$).

Background and theoretical motivation

The assumption of IID might be reasonable here.

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

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

Background and theoretical motivation

The assumption of IID might be reasonable here.

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

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

Background and theoretical motivation

The assumption of IID might be reasonable here.

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

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

Background and theoretical motivation

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

Background and theoretical motivation

This is trivial (in principle), since

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

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

Background and theoretical motivation

This is trivial (in principle), since

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

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

Background and theoretical motivation

This is trivial (in principle), since

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

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

Background and theoretical motivation

This is trivial (in principle), since

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

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

Background and theoretical motivation

This is trivial (in principle), since

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

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

A brief history of Extreme Value Theory

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

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

A brief history of Extreme Value Theory

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

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

A brief history of Extreme Value Theory

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

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

A brief history of Extreme Value Theory

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

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

A brief history of Extreme Value Theory

Until 1950, development was largely theoretical.

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

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

A brief history of Extreme Value Theory

Until 1950, development was largely theoretical.

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

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

A brief history of Extreme Value Theory

Until 1950, development was largely theoretical.

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

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

A brief history of Extreme Value Theory

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

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

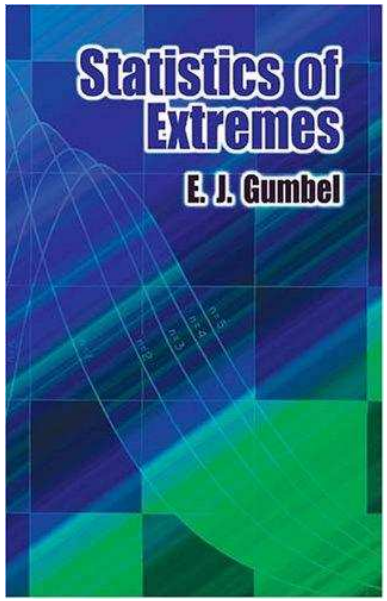
A brief history of Extreme Value Theory

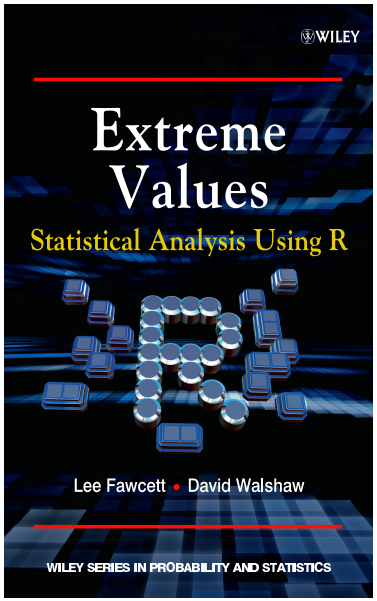
Since the 1980s, methods for the application of Extreme Value Theory have become much more widespread.

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

A brief history of Extreme Value Theory

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





The Extremal Types Theorem

The obvious questions now are:

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

The Extremal Types Theorem

The obvious questions now are:

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

The Extremal Types Theorem

The obvious questions now are:

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

The Extremal Types Theorem

The obvious questions now are:

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

The Extremal Types Theorem

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

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

The Extremal Types Theorem

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

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

The Extremal Types Theorem

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

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

The Extremal Types Theorem

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

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

The Extremal Types Theorem

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

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

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

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

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

The Extremal Types Theorem

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

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

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

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

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

The Extremal Types Theorem

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

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

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

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

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

The Extremal Types Theorem

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

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

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

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

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

The Extremal Types Theorem

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

The Extremal Types Theorem

Theorem *(The Extremal Types Theorem)*

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

$$\Pr\{(M_n - b_n)/a_n \leq x\} \to G(x) \tag{2}$$

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

$$I : G(x) = \exp\{-\exp(-x)\} \qquad -\infty < x < \infty; \tag{3}$$

$$II : G(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0, \alpha > 0; \end{cases} \tag{4}$$

$$III : G(x) = \begin{cases} \exp\{-(-x)^{\alpha}\} & x < 0, \alpha > 0 \\ 1 & x \geq 0. \end{cases} \tag{5}$$

The Extremal Types Theorem

Theorem *(The Extremal Types Theorem)*

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

$$\Pr\{(M_n - b_n)/a_n \leq x\} \to G(x) \tag{2}$$

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

$$I : G(x) = \exp\{-\exp(-x)\} \qquad -\infty < x < \infty; \tag{3}$$

$$II : G(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0, \alpha > 0; \end{cases} \tag{4}$$

$$III : G(x) = \begin{cases} \exp\{-(-x)^{\alpha}\} & x < 0, \alpha > 0 \\ 1 & x \geq 0. \end{cases} \tag{5}$$

The Extremal Types Theorem

Theorem *(The Extremal Types Theorem)*

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

$$\Pr\{(M_n - b_n)/a_n \leq x\} \to G(x) \tag{2}$$

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

$$I : G(x) = \exp\{-\exp(-x)\} \qquad -\infty < x < \infty; \tag{3}$$

$$II : G(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0, \alpha > 0; \end{cases} \tag{4}$$

$$III : G(x) = \begin{cases} \exp\{-(-x)^{\alpha}\} & x < 0, \alpha > 0 \\ 1 & x \geq 0. \end{cases} \tag{5}$$

The Extremal Types Theorem

Theorem *(The Extremal Types Theorem)*

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

$$\Pr\{(M_n - b_n)/a_n \leq x\} \to G(x) \tag{2}$$

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

$$I : G(x) = \exp\{-\exp(-x)\} \qquad -\infty < x < \infty; \tag{3}$$

$$II : G(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0, \alpha > 0; \end{cases} \tag{4}$$

$$III : G(x) = \begin{cases} \exp\{-(-x)^{\alpha}\} & x < 0, \alpha > 0 \\ 1 & x \geq 0. \end{cases} \tag{5}$$

The Extremal Types Theorem

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

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

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

The Extremal Types Theorem

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

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

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

The Extremal Types Theorem

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

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

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

The Extremal Types Theorem

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

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

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

The Extremal Types Theorem

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

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

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

The Extremal Types Theorem

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

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

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

The Extremal Types Theorem

So we know that

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

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

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

The Extremal Types Theorem

So we know that

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

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

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

The Generalised Extreme Value distribution

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

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

The Generalised Extreme Value distribution

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

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

The Generalised Extreme Value distribution

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

$$G(x; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}, \tag{6}$$

where $a_+ = \max(0, a)$. The situation where $\xi = 0$ is not defined in (6), but is taken as the limit as $\xi \to 0$, given by

$$G(x; \mu, \sigma) = \exp\left\{-\exp\left(-\left[\frac{x-\mu}{\sigma}\right]\right)\right\}. \tag{7}$$

The Generalised Extreme Value distribution

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

$$G(x; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}, \tag{6}$$

where $a_+ = \max(0, a)$. The situation where $\xi = 0$ is not defined in (6), but is taken as the limit as $\xi \to 0$, given by

$$G(x; \mu, \sigma) = \exp\left\{-\exp\left(-\left[\frac{x-\mu}{\sigma}\right]\right)\right\}. \tag{7}$$

The Generalised Extreme Value distribution

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

$$G(x; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}, \tag{6}$$

where $a_+ = \max(0, a)$. The situation where $\xi = 0$ is not defined in (6), but is taken as the limit as $\xi \to 0$, given by

$$G(x; \mu, \sigma) = \exp\left\{-\exp\left(-\left[\frac{x-\mu}{\sigma}\right]\right)\right\}. \tag{7}$$

The Generalised Extreme Value distribution

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

$$G(x; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}, \tag{6}$$

where $a_+ = \max(0, a)$. The situation where $\xi = 0$ is not defined in (6), but is taken as the limit as $\xi \to 0$, given by

$$G(x; \mu, \sigma) = \exp\left\{-\exp\left(-\left[\frac{x-\mu}{\sigma}\right]\right)\right\}. \tag{7}$$

The Generalised Extreme Value distribution

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

Shape parameter ξ:

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

The Generalised Extreme Value distribution

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

Shape parameter ξ:

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

The Generalised Extreme Value distribution

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

Shape parameter ξ:

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

The Generalised Extreme Value distribution

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

Shape parameter ξ:

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

The Generalised Extreme Value distribution

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

Shape parameter ξ:

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

The Generalised Extreme Value distribution

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

Shape parameter ξ:

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

The Generalised Extreme Value distribution

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

Shape parameter ξ:

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

The Generalised Extreme Value distribution

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

Shape parameter ξ:

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

The Generalised Extreme Value distribution

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

Shape parameter ξ:

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

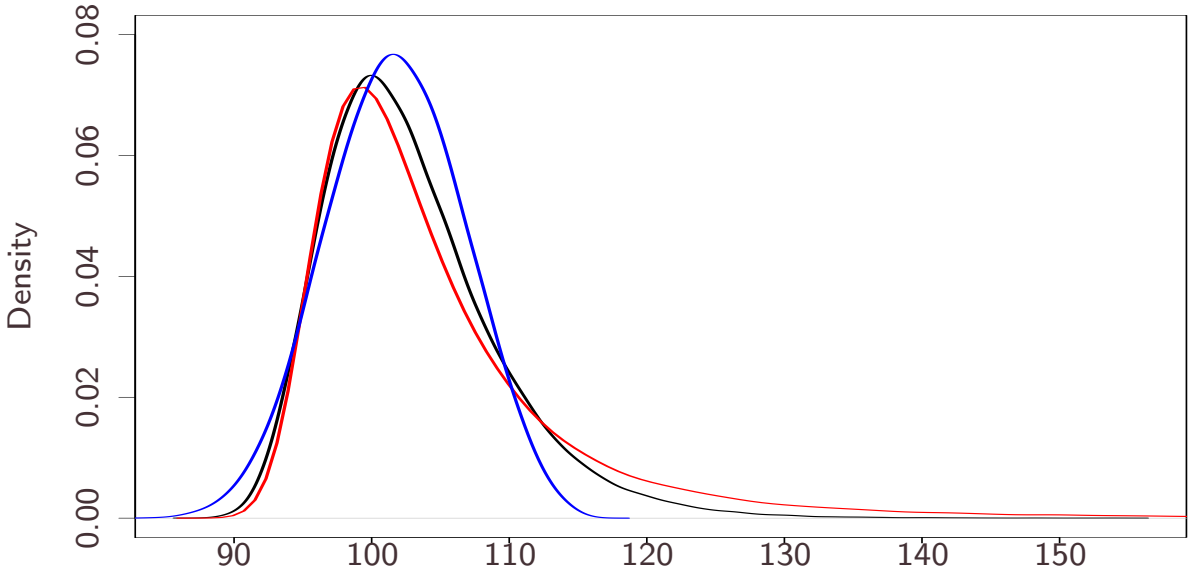
The Generalised Extreme Value distribution

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

Shape parameter ξ:

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

The Generalised Extreme Value distribution



The Generalised Extreme Value distribution

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

We know that

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

After some algebra, it turns out that

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

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

The Generalised Extreme Value distribution

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

We know that

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

After some algebra, it turns out that

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

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

The Generalised Extreme Value distribution

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

We know that

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

After some algebra, it turns out that

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

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

The Generalised Extreme Value distribution

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

We know that

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

After some algebra, it turns out that

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

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

The Generalised Extreme Value distribution

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

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

The Generalised Extreme Value distribution

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

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

Example 2.1

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

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

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

Example 2.1: Solution (1/2)

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

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

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

Example 2.1: Solution (1/2)

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

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

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

Example 2.1: Solution (1/2)

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

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

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

Example 2.1: Solution (1/2)

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

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

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

Example 2.1: Solution (1/2)

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

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

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

Example 2.1: Solution (1/2)

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

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

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

Example 2.1: Solution (2/2)

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

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

From Stage 1 methods courses, you should know that

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

Thus, we have

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

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

Example 2.1: Solution (2/2)

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

$$\left[1 - e^{-(z+\log n)}\right]^n = \left[1 - e^{-z} e^{\log n^{-1}}\right]^n$$

$$= \left[1 - n^{-1} e^{-z}\right]^n.$$

From Stage 1 methods courses, you should know that

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

Thus, we have

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

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

Example 2.1: Solution (2/2)

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

$$\left[1 - e^{-(z+\log n)}\right]^n = \left[1 - e^{-z} e^{\log n^{-1}}\right]^n$$

$$= \left[1 - n^{-1} e^{-z}\right]^n.$$

From Stage 1 methods courses, you should know that

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

Thus, we have

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

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

Example 2.1: Solution (2/2)

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

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

From Stage 1 methods courses, you should know that

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

Thus, we have

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

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

Example 2.1: Solution (2/2)

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

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

From Stage 1 methods courses, you should know that

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

Thus, we have

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

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

Example 2.1: Solution (2/2)

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

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

From Stage 1 methods courses, you should know that

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

Thus, we have

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

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

Example 2.1: Solution (2/2)

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

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

From Stage 1 methods courses, you should know that

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

Thus, we have

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

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

Example 2.1: Solution (2/2)

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

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

From Stage 1 methods courses, you should know that

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

Thus, we have

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

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

Example 2.1: Solution (2/2)

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

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

From Stage 1 methods courses, you should know that

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

Thus, we have

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

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

Example 2.2

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

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

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

Example 2.2: Solution (1/1)

Following the solution to Example 2.1, we find:

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

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

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

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

Example 2.2: Solution (1/1)

Following the solution to Example 2.1, we find:

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

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

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

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

Example 2.2: Solution (1/1)

Following the solution to Example 2.1, we find:

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

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

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

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

Example 2.2: Solution (1/1)

Following the solution to Example 2.1, we find:

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

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

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

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

Example 2.2: Solution (1/1)

Following the solution to Example 2.1, we find:

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

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

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

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

Example 2.2: Solution (1/1)

Following the solution to Example 2.1, we find:

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

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

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

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

Typical application – 1: Data pre–processing

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

Typical application – 1: Data pre–processing

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

Typical application – 1: Data pre–processing

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

Typical application – 1: Data pre–processing

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

Typical application – 1: Data pre–processing

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

Typical application – 2: Parameter estimation

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

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

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

where g is the GEV probability density function and can be found, after differentiation of the distribution function (6), to be

$$\frac{1}{\sigma}\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-(1/\xi+1)} \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}. \tag{8}$$

(try this yourself!)

Typical application – 2: Parameter estimation

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

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

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

where g is the GEV probability density function and can be found, after differentiation of the distribution function (6), to be

$$\frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-(1/\xi+1)} \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}. \tag{8}$$

(try this yourself!)

Typical application – 2: Parameter estimation

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

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

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

where g is the GEV probability density function and can be found, after differentiation of the distribution function (6), to be

$$\frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-(1/\xi+1)} \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}. \tag{8}$$

(try this yourself!)

Typical application – 2: Parameter estimation

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

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

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

where g is the GEV probability density function and can be found, after differentiation of the distribution function (6), to be

$$\frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-(1/\xi+1)} \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}. \tag{8}$$

(try this yourself!)

Typical application – 2: Parameter estimation

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

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

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

where g is the GEV probability density function and can be found, after differentiation of the distribution function (6), to be

$$\frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-(1/\xi+1)} \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}. \tag{8}$$

(try this yourself!)

Typical application – 2: Parameter estimation

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

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

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

where g is the GEV probability density function and can be found, after differentiation of the distribution function (6), to be

$$\frac{1}{\sigma}\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-(1/\xi+1)} \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}. \tag{8}$$

(try this yourself!)

Typical application – 2: Parameter estimation

Use the GEV probability density function in (8) to form the likelihood function $L(\mu, \sigma, \xi; \boldsymbol{x})$. Also, obtain the GEV log–likelihood function $\ell(\mu, \sigma, \xi; \boldsymbol{x})$ and the corresponding (log) likelihood equations.

Typical application – 2: Parameter estimation

The likelihood is given by

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

Typical application – 2: Parameter estimation

The likelihood is given by

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

Typical application – 2: Parameter estimation

The log–likelihood is given by

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

Typical application – 2: Parameter estimation

The log–likelihood is given by

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

Typical application – 2: Parameter estimation

The log–likelihood is given by

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

Typical application – 2: Parameter estimation

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

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

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

Typical application – 2: Parameter estimation

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

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

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

Typical application – 2: Parameter estimation

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

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

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

Typical application – 2: Parameter estimation

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

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

How can we get around this?

Typical application – 2: Parameter estimation

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

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

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

Typical application – 2: Parameter estimation

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

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

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

Typical application – 2: Parameter estimation

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

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

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

Typical application – 3: Model adequacy

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

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

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

Typical application – 3: Model adequacy

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

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

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

Typical application – 3: Model adequacy

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

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

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

Typical application – 3: Model adequacy

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

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

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

Typical application – 4: Return level estimation

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

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

Typical application – 4: Return level estimation

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

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

Typical application – 4: Return level estimation

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

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

Typical application – 4: Return level estimation

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

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

Typical application – 4: Return level estimation

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

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

Then we can write down the following probability statement:

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

i.e.

$$1 - \Pr(\text{annual maximum} \leq z_{100}) = \frac{1}{100}. \tag{9}$$

Typical application – 4: Return level estimation

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

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

Then we can write down the following probability statement:

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

i.e.

$$1 - \Pr(\text{annual maximum} \leq z_{100}) = \frac{1}{100}. \tag{9}$$

Typical application – 4: Return level estimation

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

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

Then we can write down the following probability statement:

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

i.e.

$$1 - \Pr(\text{annual maximum} \leq z_{100}) = \frac{1}{100}. \tag{9}$$

Typical application – 4: Return level estimation

Now the left–hand–side of (9), in terms of our fitted GEV, is

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

giving

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

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

Typical application – 4: Return level estimation

Now the left–hand–side of (9), in terms of our fitted GEV, is

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

giving

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

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

Typical application – 4: Return level estimation

Now the left–hand–side of (9), in terms of our fitted GEV, is

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

giving

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

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

Typical application – 4: Return level estimation

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

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

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

$$\hat{z}_r = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}} \left[\left(-\log\left(1 - r^{-1}\right) \right)^{-\hat{\xi}} - 1 \right]. \tag{10}$$

Typical application – 4: Return level estimation

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

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

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

$$\hat{z}_r = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}} \left[\left(-\log\left(1 - r^{-1}\right)\right)^{-\hat{\xi}} - 1 \right]. \tag{10}$$

Typical application – 4: Return level estimation

Care needs to be taken when $\xi = 0$; in this case, as with forming the likelihood in the first place, we should work with the limiting form (as $\xi \to 0$) of the distribution function (i.e. Equation (7)).

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

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

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

Typical application – 4: Return level estimation

Care needs to be taken when $\xi = 0$; in this case, as with forming the likelihood in the first place, we should work with the limiting form (as $\xi \to 0$) of the distribution function (i.e. Equation (7)).

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

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

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

Typical application – 4: Return level estimation

Care needs to be taken when $\xi = 0$; in this case, as with forming the likelihood in the first place, we should work with the limiting form (as $\xi \to 0$) of the distribution function (i.e. Equation (7)).

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

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

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

Typical application – 4: Return level estimation

Care needs to be taken when $\xi = 0$; in this case, as with forming the likelihood in the first place, we should work with the limiting form (as $\xi \to 0$) of the distribution function (i.e. Equation (7)).

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

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

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

Sea surge at Wassaw Island

Recall the example from Section 1.3:

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

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

Sea surge at Wassaw Island

Recall the example from Section 1.3:

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

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

Sea surge at Wassaw Island

Recall the example from Section 1.3:

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

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

Sea surge at Wassaw Island

Recall the example from Section 1.3:

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

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

Sea surge at Wassaw Island

Recall the example from Section 1.3:

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

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

Sea surge at Wassaw Island

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

For example, using simple empirical arguments,

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

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

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

Sea surge at Wassaw Island

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

For example, using simple empirical arguments,

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

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

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

1. Data pre–processing

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

Figure 6 shows a time series plot and histogram of the 50 annual maxima:

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

1. Data pre–processing

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

Figure 6 shows a time series plot and histogram of the 50 annual maxima:

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

1. Data pre–processing

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

Figure 6 shows a time series plot and histogram of the 50 annual maxima:

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

1. Data pre–processing

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

Figure 6 shows a time series plot and histogram of the 50 annual maxima:

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

1. Data pre–processing

In R:

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

Or:

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

1. Data pre–processing

In R:

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

Or:

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

1. Data pre–processing

In R:

```
year = seq(1975, 2024, 1)
      # Sequence of values from 1975 to 2024
```

Then, to produce the plots shown in Figure 6:

```
> par(mfrow=c(1,2))
      # Partitions the plotting space
> plot(wassaw~year,type='b')
> hist(wassaw)
```

1. Data pre–processing

In R:

```
year = seq(1975, 2024, 1)
      # Sequence of values from 1975 to 2024
```

Then, to produce the plots shown in Figure 6:

```
> par(mfrow=c(1,2))
      # Partitions the plotting space
> plot(wassaw~year,type='b')
> hist(wassaw)
```

1. Data pre–processing

In R:

```
year = seq(1975, 2024, 1)
      # Sequence of values from 1975 to 2024
```

Then, to produce the plots shown in Figure 6:

```
> par(mfrow=c(1,2))
      # Partitions the plotting space
> plot(wassaw~year,type='b')
> hist(wassaw)
```

1. Data pre–processing

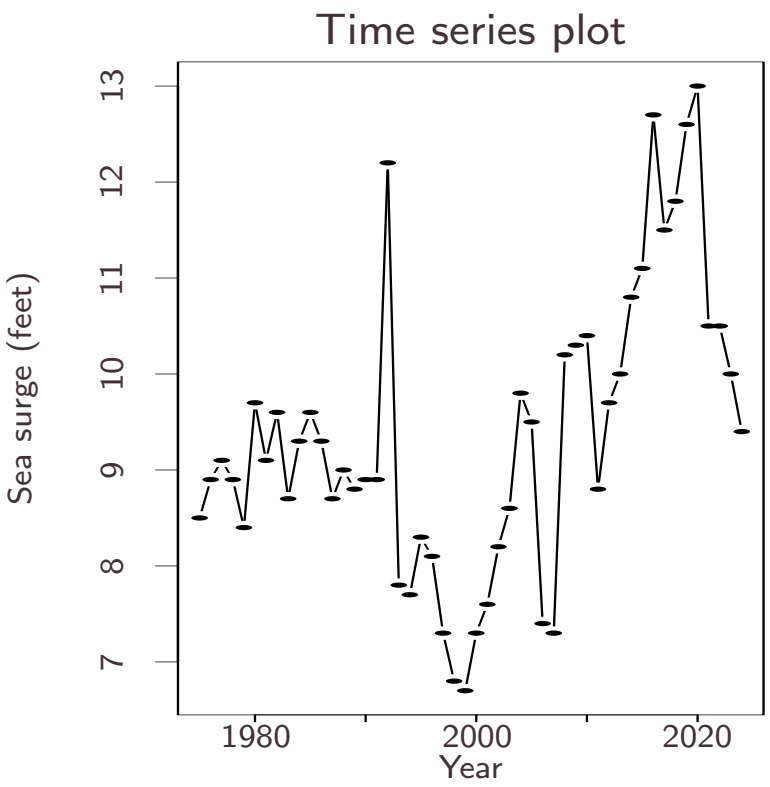
In R:

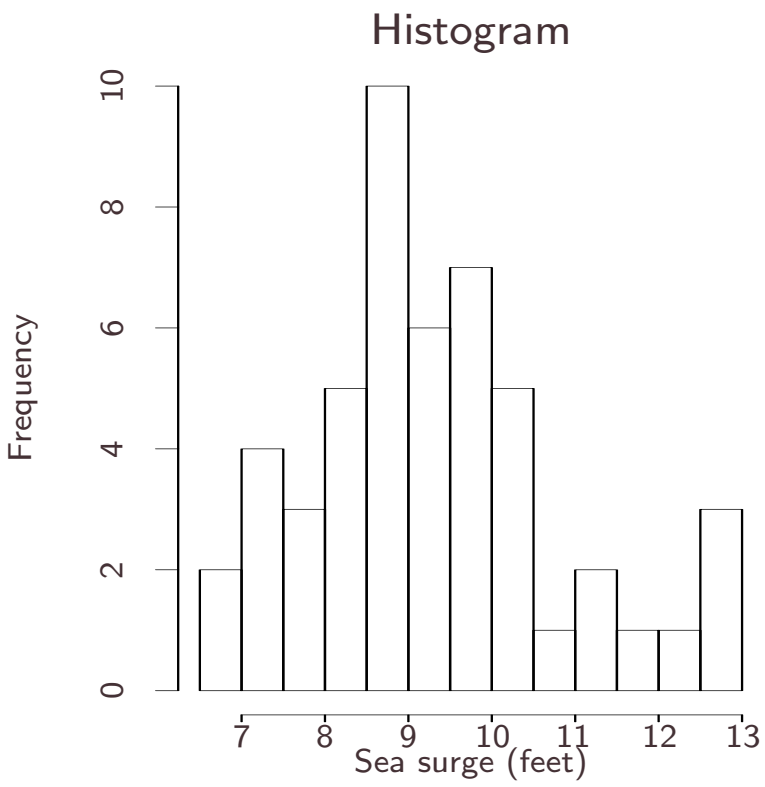
```
year = seq(1975, 2024, 1)
      # Sequence of values from 1975 to 2024
```

Then, to produce the plots shown in Figure 6:

```
> par(mfrow=c(1,2))
      # Partitions the plotting space
> plot(wassaw~year,type='b')
> hist(wassaw)
```

1. Data pre–processing





2. Fitting the GEV: Parameter estimation

We now use R to maximise

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

with respect to μ, σ and ξ.

In R:

```
> theta=c(mean(wassaw), sd(wassaw), 0.1)
      # Sets up parameter vector θ = (μ, σ, ξ)
```

2. Fitting the GEV: Parameter estimation

We now use R to maximise

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

with respect to μ, σ and ξ.

In R:

```
> theta=c(mean(wassaw), sd(wassaw), 0.1)
      # Sets up parameter vector θ = (μ, σ, ξ)
```

2. Fitting the GEV: Parameter estimation

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

2. Fitting the GEV: Parameter estimation

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

2. Fitting the GEV: Parameter estimation

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

2. Fitting the GEV: Parameter estimation

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

2. Fitting the GEV: Parameter estimation

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

2. Fitting the GEV: Parameter estimation

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

2. Fitting the GEV: Parameter estimation

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

2. Fitting the GEV: Parameter estimation

In R:

```
> dataset=wassaw
      # Attaches the sea–surge extremes to "dataset"

> nlm(gev.loglik,theta)
      # Minimisation routine

$minimum
[1] 89.52412

$estimate
[1] 8.7112735 1.3114836 -0.1084451

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

$code
[1] 1
```

2. Fitting the GEV: Parameter estimation

In R:

```
> dataset=wassaw
      # Attaches the sea–surge extremes to "dataset"

> nlm(gev.loglik,theta)
      # Minimisation routine

$minimum
[1] 89.52412

$estimate
[1] 8.7112735 1.3114836 -0.1084451

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

$code
[1] 1
```

2. Fitting the GEV: Parameter estimation

In R:

```
> dataset=wassaw
      # Attaches the sea–surge extremes to "dataset"

> nlm(gev.loglik,theta)
      # Minimisation routine

$minimum
[1] 89.52412

$estimate
[1] 8.7112735 1.3114836 -0.1084451

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

$code
[1] 1
```

2. Fitting the GEV: Parameter estimation

In R:

```
> dataset=wassaw
      # Attaches the sea–surge extremes to "dataset"

> nlm(gev.loglik,theta)
      # Minimisation routine

$minimum
[1] 89.52412

$estimate
[1] 8.7112735 1.3114836 -0.1084451

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

$code
[1] 1
```

2. Fitting the GEV: Parameter estimation

In R:

```
> dataset=wassaw
      # Attaches the sea–surge extremes to "dataset"

> nlm(gev.loglik,theta)
      # Minimisation routine

$minimum
[1] 89.52412

$estimate
[1] 8.7112735 1.3114836 -0.1084451

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

$code
[1] 1
```

2. Fitting the GEV: Parameter estimation

In R:

```
> dataset=wassaw
      # Attaches the sea–surge extremes to "dataset"

> nlm(gev.loglik,theta)
      # Minimisation routine

$minimum
[1] 89.52412

$estimate
[1] 8.7112735 1.3114836 -0.1084451

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

$code
[1] 1
```

2. Fitting the GEV: Standard errors

In R:

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

Now:

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

2. Fitting the GEV: Standard errors

In R:

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

Now:

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

2. Fitting the GEV: Standard errors

In R:

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

Now:

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

2. Fitting the GEV: Standard errors

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

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

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

evaluated at the MLEs.

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

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

2. Fitting the GEV: Standard errors

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

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

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

evaluated at the MLEs.

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

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

2. Fitting the GEV: Standard errors

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

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

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

evaluated at the MLEs.

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

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

2. Fitting the GEV: Standard errors

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

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

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

evaluated at the MLEs.

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

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

2. Fitting the GEV: Standard errors

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

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

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

evaluated at the MLEs.

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

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

2. Fitting the GEV: Standard errors

Recall from MAS2901 that inversion of the Information matrix gives the **Variance–Covariance matrix**.

Thus, we have a 3×3 matrix to invert!

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

2. Fitting the GEV: Standard errors

Recall from MAS2901 that inversion of the Information matrix gives the **Variance–Covariance matrix**.

Thus, we have a 3×3 matrix to invert!

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

2. Fitting the GEV: Standard errors

Recall from MAS2901 that inversion of the Information matrix gives the **Variance–Covariance matrix**.

Thus, we have a 3×3 matrix to invert!

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

2. Fitting the GEV: Standard errors

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

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

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

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

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

2. Fitting the GEV: Standard errors

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

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

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

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

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

2. Fitting the GEV: Standard errors

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

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

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

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

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

2. Fitting the GEV: Standard errors

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

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

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

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

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

2. Fitting the GEV: Standard errors

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

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

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

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

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

2. Fitting the GEV: Standard errors

In R:

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

> varcovar=solve(A$hessian)

> sqrt(diag(varcovar))

[1] 0.2094512 0.1490743 0.1075372
```

2. Fitting the GEV: Standard errors

In R:

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

> varcovar=solve(A$hessian)

> sqrt(diag(varcovar))

[1] 0.2094512 0.1490743 0.1075372
```

2. Fitting the GEV: Standard errors

In R:

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

> varcovar=solve(A$hessian)

> sqrt(diag(varcovar))

[1] 0.2094512 0.1490743 0.1075372
```

2. Fitting the GEV: Standard errors

In R:

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

> varcovar=solve(A$hessian)

> sqrt(diag(varcovar))

[1] 0.2094512 0.1490743 0.1075372
```

2. Fitting the GEV: Standard errors

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

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

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

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

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

for μ, σ and ξ (respectively).

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

2. Fitting the GEV: Standard errors

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

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

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

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

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

for μ, σ and ξ (respectively).

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

2. Fitting the GEV: Standard errors

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

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

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

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

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

for μ, σ and ξ (respectively).

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

2. Fitting the GEV: Standard errors

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

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

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

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

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

for μ, σ and ξ (respectively).

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

2. Fitting the GEV: Standard errors

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

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

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

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

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

for μ, σ and ξ (respectively).

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

2. Fitting the GEV: Standard errors

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

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

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

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

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

for μ, σ and ξ (respectively).

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

3. Model adequacy: Probability plots

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

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

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

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

3. Model adequacy: Probability plots

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

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

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

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

3. Model adequacy: Probability plots

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

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

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

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

3. Model adequacy: Probability plots

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

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

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

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

3. Model adequacy: Probability plots

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

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

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

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

3. Model adequacy: Probability plots

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

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

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

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

3. Model adequacy: Quantile plots

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

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

3. Model adequacy: Quantile plots

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

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

3. Model adequacy: Quantile plots

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

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

3. Model adequacy: Quantile plots

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

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

3. Model adequacy: Quantile plots

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

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

Probability plot in R

In R:

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

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

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

Probability plot in R

In R:

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

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

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

Probability plot in R

In R:

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

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

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

Probability plot in R

In R:

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

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

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

Probability plot in R

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

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

Probability plot in R

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

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

Probability plot in R

Then:

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

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

Probability plot in R

Then:

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

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

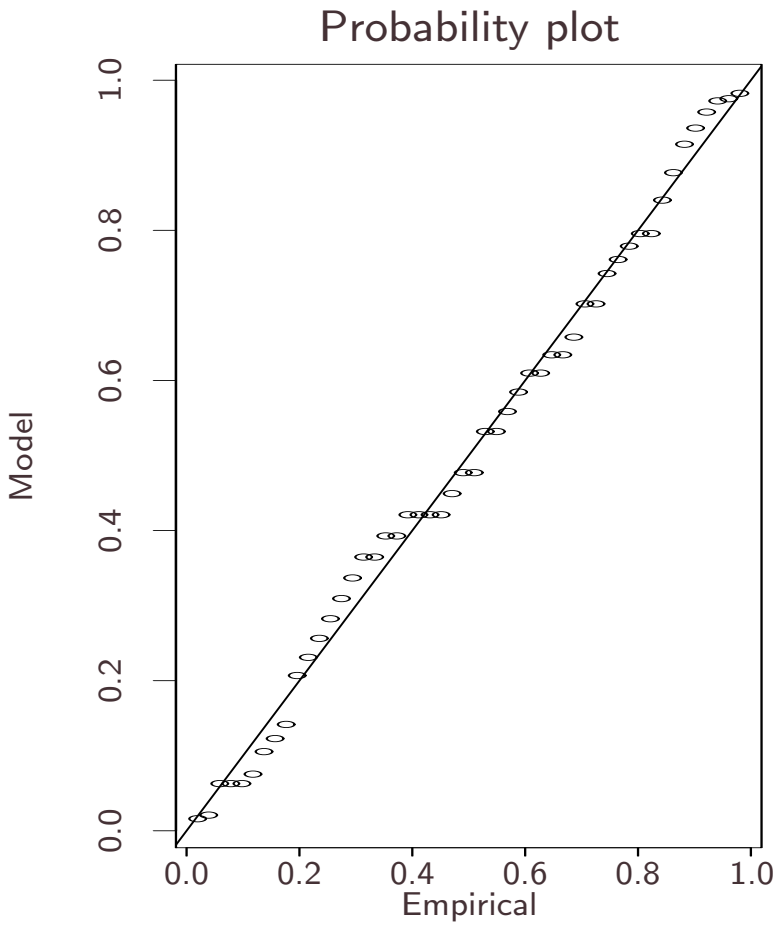
Probability plot in R

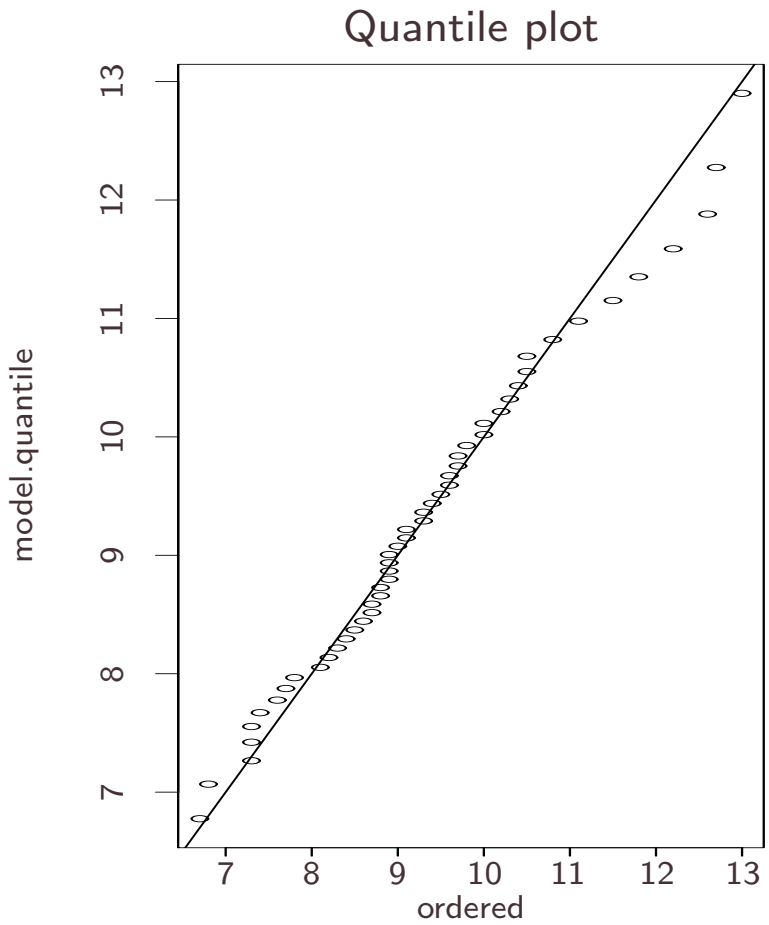
Then:

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

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

Probability/Quantile plots in R





4. Return level estimation

In question 2 of Section 1.3, you were asked to provide an estimate of the height of a new sea wall to protect the city of Savannah against the storm surge we would expect to see

(i) once in ten years;

(ii) once in a hundred years.

4. Return level estimation

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

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

Using Equation (10):

$$\hat{z}_r = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\left[\left(-\log\left(1 - r^{-1}\right)\right)^{-\hat{\xi}} - 1\right], \tag{10}$$

we find that

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

similarly,

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

4. Return level estimation

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

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

Using Equation (10):

$$\hat{z}_r = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\left[\left(-\log\left(1 - r^{-1}\right)\right)^{-\hat{\xi}} - 1\right], \tag{10}$$

we find that

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

similarly,

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

4. Return level estimation

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

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

Using Equation (10):

$$\hat{z}_r = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\left[\left(-\log\left(1 - r^{-1}\right)\right)^{-\hat{\xi}} - 1\right], \tag{10}$$

we find that

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

similarly,

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

4. Return level estimation

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

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

Using Equation (10):

$$\hat{z}_r = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\left[\left(-\log\left(1 - r^{-1}\right)\right)^{-\hat{\xi}} - 1\right], \tag{10}$$

we find that

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

similarly,

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

4. Return level estimation

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

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

4. Return level estimation

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

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

4. Return level estimation

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

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

Specifically,

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

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

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

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

4. Return level estimation

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

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

Specifically,

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

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

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

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

4. Return level estimation

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

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

Specifically,

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

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

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

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

4. Return level estimation

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

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

Specifically,

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

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

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

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

Return level standard errors in R

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

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

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

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

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

Return level standard errors in R

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

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

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

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

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

Return level standard errors in R

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

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

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

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

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

Return level standard errors in R

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

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

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

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

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

Return level standard errors in R

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

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

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

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

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

Return level standard errors in R

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

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

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

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

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

Return level standard errors in R

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

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

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

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

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

Return level standard errors in R

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

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

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

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

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

Return level standard errors in R

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

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

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

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

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

Return level standard errors in R

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

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

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

Return level standard errors in R

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

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

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

Return level standard errors in R

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

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

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

Return level standard errors in R

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

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

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

Return level standard errors in R

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

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

with

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

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

Return level standard errors in R

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

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

with

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

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

Return level standard errors in R

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

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

with

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

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

4. Return level estimation

Table 3 shows a range of estimated return levels, with associated standard errors in parentheses.

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

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

4. Return level estimation

Table 3 shows a range of estimated return levels, with associated standard errors in parentheses.

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

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

4. Return level estimation

Table 3 shows a range of estimated return levels, with associated standard errors in parentheses.

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

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

Using the ismev package in R

Class demonstration in R

Example 2.3: Rainfall in the Lake District

The data shown in Table 4 are annual maximum rainfall accumulations, obtained from daily records, for a period of 21 years (2001–2021 inclusive) at Eskdale in the Lake District.

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

Example 2.3: Rainfall in the Lake District

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

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

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

Example 2.3: Rainfall in the Lake District

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

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

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

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

Example 2.3: Solution to part(a) (1/8)

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

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

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

Example 2.3: Solution to part(a) (1/8)

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

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

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

Example 2.3: Solution to part(a) (1/8)

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

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

$$\begin{aligned}
\det(I_O) &= 0.0062(0.0091 \times 48.4122 - 0.2114^2) \\
&\quad +0.0046(-0.0046 \times 48.4122 - 0.1962 \times (-0.2114)) \\
&\quad +0.1962(0.0046 \times 0.2114 - 0.\mathbf{1}962 \times 0.0091) \\
&= 0.00146.
\end{aligned}$$

Example 2.3: Solution to part(a) (2/8)

2. Find the transpose of I_O:

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

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

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

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

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

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

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

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

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

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

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

Example 2.3: Solution to part(a) (4/8)

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

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

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

Example 2.3: Solution to part(a) (4/8)

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

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

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

Example 2.3: Solution to part(a) (4/8)

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

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

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

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

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

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

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

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

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

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

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

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

5. Final step:

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

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

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

5. Final step:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The correlation between σ and ξ can be estimated using

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

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

The correlation between σ and ξ can be estimated using

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

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

The correlation between σ and ξ can be estimated using

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

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

Using Equation (10):

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

we get:

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

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

Using Equation (10):

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

we get:

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

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

Using Equation (10):

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

we get:

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

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

Using Equation (10):

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

we get:

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

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

Using Equation (10):

$$\hat{z}_r = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\left[\left(-\log\ \left(1 - r^{-1}\right)\right)^{-\hat{\xi}} - 1\right],$$

we get:

$$\begin{aligned}
\hat{z}_{100} &= 304.242 + \frac{68.977}{0.249}\left[\left(-\log\ \left(1 - 100^{-1}\right)\right)^{-0.249} - 1\right] \\
&= 898.1133 \text{ mm}; \qquad \text{similarly,} \\
\hat{z}_{1000} &= 1574.085 \text{ mm.}
\end{aligned}$$

Example 2.3: Solution to part(d) (1/3)

From earlier, we know that

$$\begin{aligned}\nabla z_r^T &= \left[\frac{\partial z_r}{\partial \mu}, \frac{\partial z_r}{\partial \sigma}, \frac{\partial z_r}{\partial \xi}\right] \\ &= \left[1, -\xi^{-1}(1 - y_r^{-\xi}), \sigma\xi^{-2}(1 - y_r^{-\xi}) - \sigma\xi^{-1}y_r^{-\xi}\log y_r\right],\end{aligned}$$

where $y_r = -\log(1 - r^{-1})$; evaluated at the MLEs for μ, σ and ξ, we get

$$\nabla z_{100}^T = [1, 8.6097, 1621.187]$$

Example 2.3: Solution to part(d) (1/3)

From earlier, we know that

$$\begin{aligned}\nabla z_r^T &= \left[\frac{\partial z_r}{\partial \mu}, \frac{\partial z_r}{\partial \sigma}, \frac{\partial z_r}{\partial \xi}\right] \\ &= \left[1, -\xi^{-1}(1 - y_r^{-\xi}), \sigma\xi^{-2}(1 - y_r^{-\xi}) - \sigma\xi^{-1}y_r^{-\xi}\log y_r\right],\end{aligned}$$

where $y_r = -\log(1 - r^{-1})$; evaluated at the MLEs for μ, σ and ξ, we get

$$\nabla z_{100}^T = [1, 8.6097, 1621.187]$$

Example 2.3: Solution to part(d) (1/3)

From earlier, we know that

$$\begin{aligned}\nabla z_r^T &= \left[\frac{\partial z_r}{\partial \mu}, \frac{\partial z_r}{\partial \sigma}, \frac{\partial z_r}{\partial \xi}\right] \\ &= \left[1, -\xi^{-1}(1 - y_r^{-\xi}), \sigma\xi^{-2}(1 - y_r^{-\xi}) - \sigma\xi^{-1}y_r^{-\xi}\log y_r\right],\end{aligned}$$

where $y_r = -\log(1 - r^{-1})$; evaluated at the MLEs for μ, σ and ξ, we get

$$\nabla z_{100}^T = [1, 8.6097, 1621.187]$$

Example 2.3: Solution to part(d) (2/3)

Thus,

$$
\begin{aligned}
\text{var}(\hat{z}_{100}) &= \nabla z_{100}^T V \nabla z_{100} \\
&= [1, 8.6097, 1621.187] \\
&\quad \times \begin{pmatrix} 271.1377 & 124.1266 & -0.5568 \\ 124.1266 & 179.2199 & 0.2795 \\ -0.5568 & 0.2795 & 0.0240 \end{pmatrix} \\
&\quad \times \begin{bmatrix} 1 \\ 8.6097 \\ 1621.187 \end{bmatrix} \\
&= 84768.61
\end{aligned}
$$

Therefore, s.e.$(\hat{z}_{100}) = \sqrt{84768.61} = 291.1505.$

Example 2.3: Solution to part(d) (2/3)

Thus,

$$
\begin{aligned}
\text{var}(\hat{z}_{100}) &= \nabla z_{100}^T V \nabla z_{100} \\
&= [1, 8.6097, 1621.187] \\
&\quad \times \begin{pmatrix} 271.1377 & 124.1266 & -0.5568 \\ 124.1266 & 179.2199 & 0.2795 \\ -0.5568 & 0.2795 & 0.0240 \end{pmatrix} \\
&\quad \times \begin{bmatrix} 1 \\ 8.6097 \\ 1621.187 \end{bmatrix} \\
&= 84768.61
\end{aligned}
$$

Therefore, s.e.$(\hat{z}_{100}) = \sqrt{84768.61} = 291.1505$.

Example 2.3: Solution to part(d) (3/3)

Similar calculations for $\hat{z}_{1000}$ give

$$\text{s.e.}(\hat{z}_{1000}) = 932.325.$$

From these standard errors, we can construct 95% confidence intervals:

$$\hat{z}_{100} \quad : \quad 898.1133 \pm 1.96 \times 291.1505 \longrightarrow (327.5, 1468.8) \text{ mm};$$

$$\hat{z}_{1000} \quad : \quad 1574.085 \pm 1.96 \times 932.325 \longrightarrow (-253.3, 3401.4) \text{ mm}.$$

Comment: CI's are very wide; also, the CI for $\hat{z}_{1000}$ is partly negative – not sensible!

Example 2.3: Solution to part(d) (3/3)

Similar calculations for $\hat{z}_{1000}$ give

$$\text{s.e.}(\hat{z}_{1000}) = 932.325.$$

From these standard errors, we can construct 95% confidence intervals:

$$\hat{z}_{100} \quad : \quad 898.1133 \pm 1.96 \times 291.1505 \longrightarrow (327.5, 1468.8) \text{ mm};$$

$$\hat{z}_{1000} \quad : \quad 1574.085 \pm 1.96 \times 932.325 \longrightarrow (-253.3, 3401.4) \text{ mm}.$$

Comment: CI's are very wide; also, the CI for $\hat{z}_{1000}$ is partly negative – not sensible!

Example 2.3: Solution to part(d) (3/3)

Similar calculations for $\hat{z}_{1000}$ give

$$\text{s.e.}(\hat{z}_{1000}) = 932.325.$$

From these standard errors, we can construct 95% confidence intervals:

$$\hat{z}_{100} \quad : \quad 898.1133 \pm 1.96 \times 291.1505 \longrightarrow (327.5, 1468.8) \text{ mm};$$

$$\hat{z}_{1000} \quad : \quad 1574.085 \pm 1.96 \times 932.325 \longrightarrow (-253.3, 3401.4) \text{ mm}.$$

Comment: CI's are very wide; also, the CI for $\hat{z}_{1000}$ is partly negative – not sensible!

Example 2.3: Solution to part(d) (3/3)

Similar calculations for $\hat{z}_{1000}$ give

$$\text{s.e.}(\hat{z}_{1000}) = 932.325.$$

From these standard errors, we can construct 95% confidence intervals:

$$\hat{z}_{100} \quad : \quad 898.1133 \pm 1.96 \times 291.1505 \longrightarrow (327.5, 1468.8) \text{ mm};$$

$$\hat{z}_{1000} \quad : \quad 1574.085 \pm 1.96 \times 932.325 \longrightarrow (-253.3, 3401.4) \text{ mm}.$$

Comment: CI's are very wide; also, the CI for $\hat{z}_{1000}$ is partly negative – not sensible!

Example 2.3: Solution to part(d) (3/3)

Similar calculations for $\hat{z}_{1000}$ give

$$\text{s.e.}(\hat{z}_{1000}) = 932.325.$$

From these standard errors, we can construct 95% confidence intervals:

$$\hat{z}_{100} \quad : \quad 898.1133 \pm 1.96 \times 291.1505 \longrightarrow (327.5, 1468.8) \text{ mm};$$

$$\hat{z}_{1000} \quad : \quad 1574.085 \pm 1.96 \times 932.325 \longrightarrow (-253.3, 3401.4) \text{ mm}.$$

Comment: CI's are very wide; also, the CI for $\hat{z}_{1000}$ is partly negative – not sensible!

Example 2.3: Solution to part(d) (3/3)

Similar calculations for $\hat{z}_{1000}$ give

$$\text{s.e.}(\hat{z}_{1000}) = 932.325.$$

From these standard errors, we can construct 95% confidence intervals:

$$\hat{z}_{100} \quad : \quad 898.1133 \pm 1.96 \times 291.1505 \longrightarrow (327.5, 1468.8) \text{ mm};$$

$$\hat{z}_{1000} \quad : \quad 1574.085 \pm 1.96 \times 932.325 \longrightarrow (-253.3, 3401.4) \text{ mm}.$$

Comment: CI's are very wide; also, the CI for $\hat{z}_{1000}$ is partly negative – not sensible!

Example 2.3: Solution to part(d) (3/3)

Similar calculations for $\hat{z}_{1000}$ give

$$\text{s.e.}(\hat{z}_{1000}) = 932.325.$$

From these standard errors, we can construct 95% confidence intervals:

$$\hat{z}_{100} \quad : \quad 898.1133 \pm 1.96 \times 291.1505 \longrightarrow (327.5, 1468.8) \text{ mm};$$

$$\hat{z}_{1000} \quad : \quad 1574.085 \pm 1.96 \times 932.325 \longrightarrow (-253.3, 3401.4) \text{ mm}.$$

Comment: CI's are <u>very</u> wide; also, the CI for $\hat{z}_{1000}$ is partly negative – not sensible!

Example 2.4

The magnitudes of the vertical forces produced by seismic *degassing bursts* are known to be associated with destructive volcanoes.

Table 5 shows the maximum force produced by degassing bursts, every quarter in the years 2015–2021 (inclusive), for the Kilauea Volcano in Hawaii.

	Jan–Mar	Apr–Jun	Jul–Sep	Oct–Dec
2015	99983	100067	99905	100367
2016	99980	99970	100086	99988
2017	99912	100084	100432	99921
2018	100123	99913	100240	100448
2019	100162	100296	100015	100122
2020	99930	99997	100710	100136
2021	99946	99918	99962	100052

Example 2.4

These data were stored the in the vector `haw` in R, and the following output obtained:

```
gev.fit(haw)

$ conv
[1] 0

$nllh
[1] 178.2502

$mle
[1] 9.998028e+04 8.710824e+01 5.921945e-01

$se
[1] 21.6185212 21.3598264 0.3027163
```

Example 2.4

Use this information to estimate the degassing bursts we would expect to see (i) once every year, (ii) once every fifty years, and (iii) once every 100 years at this volcano (do not attempt to obtain standard errors for these estimates).
How does your estimate in (i) compare to it's empirical counterpart?

Example 2.4: Solution (1/5)

From Equation (10) we know that

$$\hat{z}_r = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}} \left[\left(-\log\left(1 - r^{-1}\right) \right)^{-\hat{\xi}} - 1 \right].$$

The quantity $\hat{z}_r$ is an estimate of the level that is exceeded, on average, once every r **observations** – if we have only one observation per year, this is exactly the the same as the r–<u>year</u> return level.

In this example, we have **four** observations per year!

Example 2.4: Solution (1/5)

From Equation (10) we know that

$$\hat{z}_r = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}} \left[\left(-\log\left(1 - r^{-1}\right) \right)^{-\hat{\xi}} - 1 \right].$$

The quantity $\hat{z}_r$ is an estimate of the level that is exceeded, on average, once every r **observations** – if we have only one observation per year, this is exactly the the same as the r–year return level.

In this example, we have **four** observations per year!

Example 2.4: Solution (2/5)

Thus, an estimate of the one–year return level is given by

$$\begin{aligned}\hat{z}_4 &= 99980.28 + \frac{87.10824}{0.5921945}\left[\left(-\log(1-4^{-1})\right)^{-0.5921945} - 1\right] \\ &= 100140.8 \text{ kg.}\end{aligned}$$

Similarly, an estimate of the 50–year return level is given by

$$\begin{aligned}\hat{z}_{200} &= 99980.28 + \frac{87.10824}{0.5921945}\left[\left(-\log(1-200^{-1})\right)^{-0.5921945} - 1\right] \\ &= 103218.6 \text{ kg.}\end{aligned}$$

Example 2.4: Solution (2/5)

Thus, an estimate of the one–year return level is given by

$$\begin{aligned}\hat{z}_4 &= 99980.28 + \frac{87.10824}{0.5921945}\left[\left(-\log(1-4^{-1})\right)^{-0.5921945} - 1\right] \\ &= 100140.8 \text{ kg.}\end{aligned}$$

Similarly, an estimate of the 50–year return level is given by

$$\begin{aligned}\hat{z}_{200} &= 99980.28 + \frac{87.10824}{0.5921945}\left[\left(-\log(1-200^{-1})\right)^{-0.5921945} - 1\right] \\ &= 103218.6 \text{ kg.}\end{aligned}$$

Example 2.4: Solution (2/5)

Thus, an estimate of the one–year return level is given by

$$\begin{aligned}\hat{z}_4 &= 99980.28 + \frac{87.10824}{0.5921945}\left[\left(-\log(1-4^{-1})\right)^{-0.5921945} - 1\right] \\ &= 100140.8 \text{ kg.}\end{aligned}$$

Similarly, an estimate of the 50–year return level is given by

$$\begin{aligned}\hat{z}_{200} &= 99980.28 + \frac{87.10824}{0.5921945}\left[\left(-\log(1-200^{-1})\right)^{-0.5921945} - 1\right] \\ &= 103218.6 \text{ kg.}\end{aligned}$$

Example 2.4: Solution (2/5)

Thus, an estimate of the one–year return level is given by

$$\begin{aligned}\hat{z}_4 &= 99980.28 + \frac{87.10824}{0.5921945}\left[\left(-\log(1-4^{-1})\right)^{-0.5921945} - 1\right] \\ &= 100140.8 \text{ kg.}\end{aligned}$$

Similarly, an estimate of the 50–year return level is given by

$$\begin{aligned}\hat{z}_{200} &= 99980.28 + \frac{87.10824}{0.5921945}\left[\left(-\log(1-200^{-1})\right)^{-0.5921945} - 1\right] \\ &= 103218.6 \text{ kg.}\end{aligned}$$

Example 2.4: Solution (3/5)

Also, an estimate of the 100–year return level is given by

$$\begin{aligned}\hat{z}_{400} &= 99980.28 + \frac{87.10824}{0.5921945}\left[\left(-\log(1-400^{-1})\right)^{-0.5921945} - 1\right] \\ &= 104940.6 \text{ kg.}\end{aligned}$$

Example 2.4: Solution (3/5)

Also, an estimate of the 100–year return level is given by

$$\begin{aligned}\hat{z}_{400} &= 99980.28 + \frac{87.10824}{0.5921945}\left[\left(-\log(1-400^{-1})\right)^{-0.5921945} - 1\right] \\ &= 104940.6 \text{ kg.}\end{aligned}$$

Example 2.4: Solution (4/5)

We can obtain the empirical estimate of the one–year return level by looking at the **empirical distribution function** (see Section 2.2.3 on quantile–quantile plots).

For example, ordering the data gives:

$x_{(i)}$	$i/(n+1)$
$\vdots$	$\vdots$
99946	0.241
99962	0.276
$\vdots$	$\vdots$

An empirical estimate of the one–year return level is the value which is exceeded once every 4 observations – the value for which $i/(n+1) = 0.25$.

Example 2.4: Solution (4/5)

We can obtain the empirical estimate of the one–year return level by looking at the **empirical distribution function** (see Section 2.2.3 on quantile–quantile plots).

For example, ordering the data gives:

$x_{(i)}$	$i/(n+1)$
$\vdots$	$\vdots$
99946	0.241
99962	0.276
$\vdots$	$\vdots$

An empirical estimate of the one–year return level is the value which is exceeded once every 4 observations – the value for which $i/(n+1) = 0.25$.

Example 2.4: Solution (4/5)

We can obtain the empirical estimate of the one–year return level by looking at the **empirical distribution function** (see Section 2.2.3 on quantile–quantile plots).

For example, ordering the data gives:

$x_{(i)}$	$i/(n+1)$
$\vdots$	$\vdots$
99946	0.241
99962	0.276
$\vdots$	$\vdots$

An empirical estimate of the one–year return level is the value which is exceeded once every 4 observations – the value for which $i/(n+1) = 0.25$.

Example 2.4: Solution (5/5)

$i/(n+1) = 0.25$ is 25.71% of the way between 0.241 and 0.276;
25.71% of the way between 99946 and 99962 gives

$$\hat{z}_{4,\text{emp}} = 99950.11 \text{ kg}.$$

The model gives a slight over–estimate.

Example 2.4: Solution (5/5)

$i/(n+1) = 0.25$ is 25.71% of the way between 0.241 and 0.276; 25.71% of the way between 99946 and 99962 gives

$$\hat{z}_{4,\text{emp}} = 99950.11 \text{ kg}.$$

The model gives a slight over–estimate.

Example 2.4: Solution (5/5)

$i/(n+1) = 0.25$ is 25.71% of the way between 0.241 and 0.276; 25.71% of the way between 99946 and 99962 gives

$$\hat{z}_{4,\text{emp}} = 99950.11 \text{ kg}.$$

The model gives a slight over–estimate.

Example 2.4: Solution (5/5)

$i/(n+1) = 0.25$ is 25.71% of the way between 0.241 and 0.276; 25.71% of the way between 99946 and 99962 gives

$$\hat{z}_{4,\text{emp}} = 99950.11 \text{ kg}.$$

The model gives a slight over–estimate.

Example 2.5

The table below shows the annual maximum number of sunshine hours (per day) for Santiago de Compostela, Spain, for the years 1993–2017.

12.8	10.7	11.9	13.1	13.3	10.7	11.1	11.9	12.1
11.3	12.5	12.1	12.0	12.4	11.8	10.5	10.9	10.1
11.7	12.7	10.6	11.6	9.9	9.6	12.0		

(a) Use the ismev package in R to fit the GEV distribution to these data. Report estimates of the location, scale and shape parameters, with associated standard errors.

(b) Explain why the output in part (a) shows that the Weibull (Type *III* extreme value distribution) is most appropriate here, and give a practical justification for this.

Example 2.5: Solution

See R demo

The GEV: Words of warning

- Block size?
  - Convention is to work with blocks equal in length to the calendar year, but what if you have hourly data collected over just a few years?
  - Block length too small – limiting arguments will not hold (the GEV is a limiting result, which holds *approximately* for large n)
  - Block length is too large – not enough maxima to work with!
  - Possible sensitivity of GEV parameter estimates to block length

The GEV: Words of warning

- Block size?
  - Convention is to work with blocks equal in length to the calendar year, but what if you have hourly data collected over just a few years?
  - Block length too small – limiting arguments will not hold (the GEV is a limiting result, which holds *approximately* for large n)
  - Block length is too large – not enough maxima to work with!
  - Possible sensitivity of GEV parameter estimates to block length

The GEV: Words of warning

- Block size?
  - Convention is to work with blocks equal in length to the calendar year, but what if you have hourly data collected over just a few years?
  - Block length too small – limiting arguments will not hold (the GEV is a limiting result, which holds *approximately* for large n)
  - Block length is too large – not enough maxima to work with!
  - Possible sensitivity of GEV parameter estimates to block length

The GEV: Words of warning

- Block size?
  - Convention is to work with blocks equal in length to the calendar year, but what if you have hourly data collected over just a few years?
  - Block length too small – limiting arguments will not hold (the GEV is a limiting result, which holds *approximately* for large n)
  - Block length is too large – not enough maxima to work with!
  - Possible sensitivity of GEV parameter estimates to block length

The GEV: Words of warning

- Block size?
  - Convention is to work with blocks equal in length to the calendar year, but what if you have hourly data collected over just a few years?
  - Block length too small – limiting arguments will not hold (the GEV is a limiting result, which holds *approximately* for large n)
  - Block length is too large – not enough maxima to work with!
  - Possible sensitivity of GEV parameter estimates to block length

The GEV: Words of warning

- Extremely wasteful of data:
  - Discard all but the block maxima
  - Often results in throwing away tens of thousands of observations – some of which might be 'extreme' – just not as extreme as the block maxima!
  - Results in large standard errors and extremely wide confidence intervals for return levels

The GEV: Words of warning

- Extremely wasteful of data:
  - Discard all but the block maxima
  - Often results in throwing away tens of thousands of observations – some of which might be 'extreme' – just not as extreme as the block maxima!
  - Results in large standard errors and extremely wide confidence intervals for return levels

The GEV: Words of warning

- Extremely wasteful of data:
  - Discard all but the block maxima
  - Often results in throwing away tens of thousands of observations – some of which might be 'extreme' – just not as extreme as the block maxima!
  - Results in large standard errors and extremely wide confidence intervals for return levels

The GEV: Words of warning

- Extremely wasteful of data:
  - Discard all but the block maxima
  - Often results in throwing away tens of thousands of observations – some of which might be 'extreme' – just not as extreme as the block maxima!
  - Results in large standard errors and extremely wide confidence intervals for return levels

The GEV: Words of warning

- Standard asymptotic likelihood results not always applicable for the GEV:
  - When $\xi < 0$, the end–points of the GEV are functions of the parameter values: e.g. $\mu = \sigma/\xi$ is the upper end–point
  - This violates the usual regularity conditions of MLEs
  - Effect: When $-1 < \xi \leq -0.5$, MLEs for ξ can be obtained, but are **super–efficient** (i.e. have variance smaller than the Cramér–Rao lower bound)
  - Another consequence: Cannot obtain MLE for ξ when $\xi \leq -1$

The GEV: Words of warning

- Standard asymptotic likelihood results not always applicable for the GEV:
  - When $\xi < 0$, the end–points of the GEV are functions of the parameter values: e.g. $\mu = \sigma/\xi$ is the upper end–point
  - This violates the usual regularity conditions of MLEs
  - Effect: When $-1 < \xi \leq -0.5$, MLEs for ξ can be obtained, but are **super–efficient** (i.e. have variance smaller than the Cramér–Rao lower bound)
  - Another consequence: Cannot obtain MLE for ξ when $\xi \leq -1$

The GEV: Words of warning

- Standard asymptotic likelihood results not always applicable for the GEV:
  - When $\xi < 0$, the end–points of the GEV are functions of the parameter values: e.g. $\mu = \sigma/\xi$ is the upper end–point
  - This violates the usual regularity conditions of MLEs
  - Effect: When $-1 < \xi \leq -0.5$, MLEs for ξ can be obtained, but are **super–efficient** (i.e. have variance smaller than the Cramér–Rao lower bound)
  - Another consequence: Cannot obtain MLE for ξ when $\xi \leq -1$

The GEV: Words of warning

- Standard asymptotic likelihood results not always applicable for the GEV:
  - When $\xi < 0$, the end–points of the GEV are functions of the parameter values: e.g. $\mu = \sigma/\xi$ is the upper end–point
  - This violates the usual regularity conditions of MLEs
  - Effect: When $-1 < \xi \leq -0.5$, MLEs for ξ can be obtained, but are **super–efficient** (i.e. have variance smaller than the Cramér–Rao lower bound)
  - Another consequence: Cannot obtain MLE for ξ when $\xi \leq -1$

The GEV: Words of warning

- Standard asymptotic likelihood results not always applicable for the GEV:
  - When $\xi < 0$, the end–points of the GEV are functions of the parameter values: e.g. $\mu = \sigma/\xi$ is the upper end–point
  - This violates the usual regularity conditions of MLEs
  - Effect: When $-1 < \xi \leq -0.5$, MLEs for ξ can be obtained, but are **super–efficient** (i.e. have variance smaller than the Cramér–Rao lower bound)
  - Another consequence: Cannot obtain MLE for ξ when $\xi \leq -1$