MAS8306: Environmental Extremes

Dr. Lee Fawcett

Semester 2, 2015/16

Contact details

My name: Lee Fawcett

Office: Room 2.07 Herschel Building

Phone: 0191 208 7228

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

www: `www.mas.ncl.ac.uk/~nlf8`

Timetable and Administrative arrangements

- **Lectures** are in KGVI LT5 (Herschel TR4 after Easter) on **Tuesdays at 2pm** and Daysh G.05 on **Wednesdays at 12 noon**.
- **Tutorial time** will be given in some lectures. I will notify you in advance if a lecture/part of a lecture is to be replaced with tutorial/reading/self–study time – these slots will be optional.
- **Computer practicals**: **Mondays at 9am** (Chart PC cluster, Barbara Strang). These will happen in teaching weeks 3 and 6 and will help you with the assignments.
- **Office hours**: **Mondays 1pm**, **Tuesdays 1pm**.

Timetable and Administrative arrangements

- **Lectures** are in KGVI LT5 (Herschel TR4 after Easter) on **Tuesdays at 2pm** and Daysh G.05 on **Wednesdays at 12 noon**.
- **Tutorial time** will be given in some lectures. I will notify you in advance if a lecture/part of a lecture is to be replaced with tutorial/reading/self–study time – these slots will be optional.
- **Computer practicals**: **Mondays at 9am** (Chart PC cluster, Barbara Strang). These will happen in teaching weeks 3 and 6 and will help you with the assignments.
- **Office hours**: **Mondays 1pm**, **Tuesdays 1pm**.

Timetable and Administrative arrangements

- **Lectures** are in KGVI LT5 (Herschel TR4 after Easter) on **Tuesdays at 2pm** and Daysh G.05 on **Wednesdays at 12 noon**.
- **Tutorial time** will be given in some lectures. I will notify you in advance if a lecture/part of a lecture is to be replaced with tutorial/reading/self–study time – these slots will be optional.
- **Computer practicals**: **Mondays at 9am** (Chart PC cluster, Barbara Strang). These will happen in teaching weeks 3 and 6 and will help you with the assignments.
- **Office hours**: **Mondays 1pm**, **Tuesdays 1pm**.

Assessment and deadlines

	Assessment	Weighting	Date/Deadline
Extremes (LF)	Homework 1		4pm, Thu 25 Feb
	Homework 2	10% total	4pm, Thu 10 Mar
	Homework 3		4pm, Thu 12 May
	Mid–semester test	5%	12–1, Wed 13 Apr
	Exam (half of MAS8306)	35%	May/June
Longitudinal (PJA)	Homeworks	10%	See PJA
	Mid–semester test	5%	See PJA
	Exam (half of MAS8306)	35%	May/June
		100%	

Other stuff

- Notes (with gaps) will be handed out in lectures – you should fill in the gaps during lectures.
- A summarised version of the notes will be used in lectures as slides.
  - **Listen and learn**
  - **Write down**
  - **Announcements**
- These notes and slides will be posted on the **course website** after each section is finished, along with any other course material – such as problems sheets, model solutions to assignment questions, supplementary handouts etc.

Other stuff

- Notes (with gaps) will be handed out in lectures – you should fill in the gaps during lectures.
- A summarised version of the notes will be used in lectures as slides.
  - **Listen and learn**
  - **Write down**
  - **Announcements**
- These notes and slides will be posted on the **course website** after each section is finished, along with any other course material – such as problems sheets, model solutions to assignment questions, supplementary handouts etc.

Other stuff

- Notes (with gaps) will be handed out in lectures – you should fill in the gaps during lectures.
- A summarised version of the notes will be used in lectures as slides.
  - **Listen and learn**
  - **Write down**
  - **Announcements**
- These notes and slides will be posted on the **course website** after each section is finished, along with any other course material – such as problems sheets, model solutions to assignment questions, supplementary handouts etc.

Other stuff

- The course website can be found via the “Additional teaching information” link on the School’s webpage, or directly via:

  http://www.mas.ncl.ac.uk/~nlf8

- Please check your University email account regularly, as course announcements will often be made via email.

Other stuff

- The course website can be found via the “Additional teaching information” link on the School’s webpage, or directly via:

  http://www.mas.ncl.ac.uk/~nlf8

- Please check your University email account regularly, as course announcements will often be made via email.

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!)
- MAS369 out–of–date with current research – now a new, more advanced "topic" in MAS8304/6

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!)
- MAS369 out–of–date with current research – now a new, more advanced "topic" in MAS8304/6

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!)
- MAS369 out–of–date with current research – now a new, more advanced "topic" in MAS8304/6

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!)
- MAS369 out–of–date with current research – now a new, more advanced "topic" in MAS8304/6

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
- **Independent project**: will be based on a recently–published article in the field

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
- **Independent project**: will be based on a recently–published article in the field

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
- **Independent project**: will be based on a recently–published article in the field

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
- **Independent project**: will be based on a recently–published article in the field

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
- **Independent project**: will be based on a recently–published article in the field

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
- **Independent project**: will be based on a recently–published article in the field

A word on pre–requisite knowledge

- I will not assume any knowledge past 2nd year Statistics courses (**MAS2302**)
- We will make use of some techniques from these courses (e.g. **maximum likelihood**, **the delta method**, **likelihood ratio tests**) – these will be quickly reviewed, as–and–when they make an appearance, but you may want to consult your old notes for further details
- Some basic **Statistical Modelling** techniques will be used, but will be fully demonstrated if they go beyond MAS2302/3320.

A word on pre–requisite knowledge

- I will not assume any knowledge past 2nd year Statistics courses (**MAS2302**)
- We will make use of some techniques from these courses (e.g. **maximum likelihood**, **the delta method**, **likelihood ratio tests**) – these will be quickly reviewed, as–and–when they make an appearance, but you may want to consult your old notes for further details
- Some basic **Statistical Modelling** techniques will be used, but will be fully demonstrated if they go beyond MAS2302/3320.

A word on pre–requisite knowledge

- I will not assume any knowledge past 2nd year Statistics courses (**MAS2302**)
- We will make use of some techniques from these courses (e.g. **maximum likelihood**, **the delta method**, **likelihood ratio tests**) – these will be quickly reviewed, as–and–when they make an appearance, but you may want to consult your old notes for further details
- Some basic **Statistical Modelling** techniques will be used, but will be fully demonstrated if they go beyond MAS2302/3320.

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

Chapter 1

Background and motivation

1.1 Introduction

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.

1.1 Introduction

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.

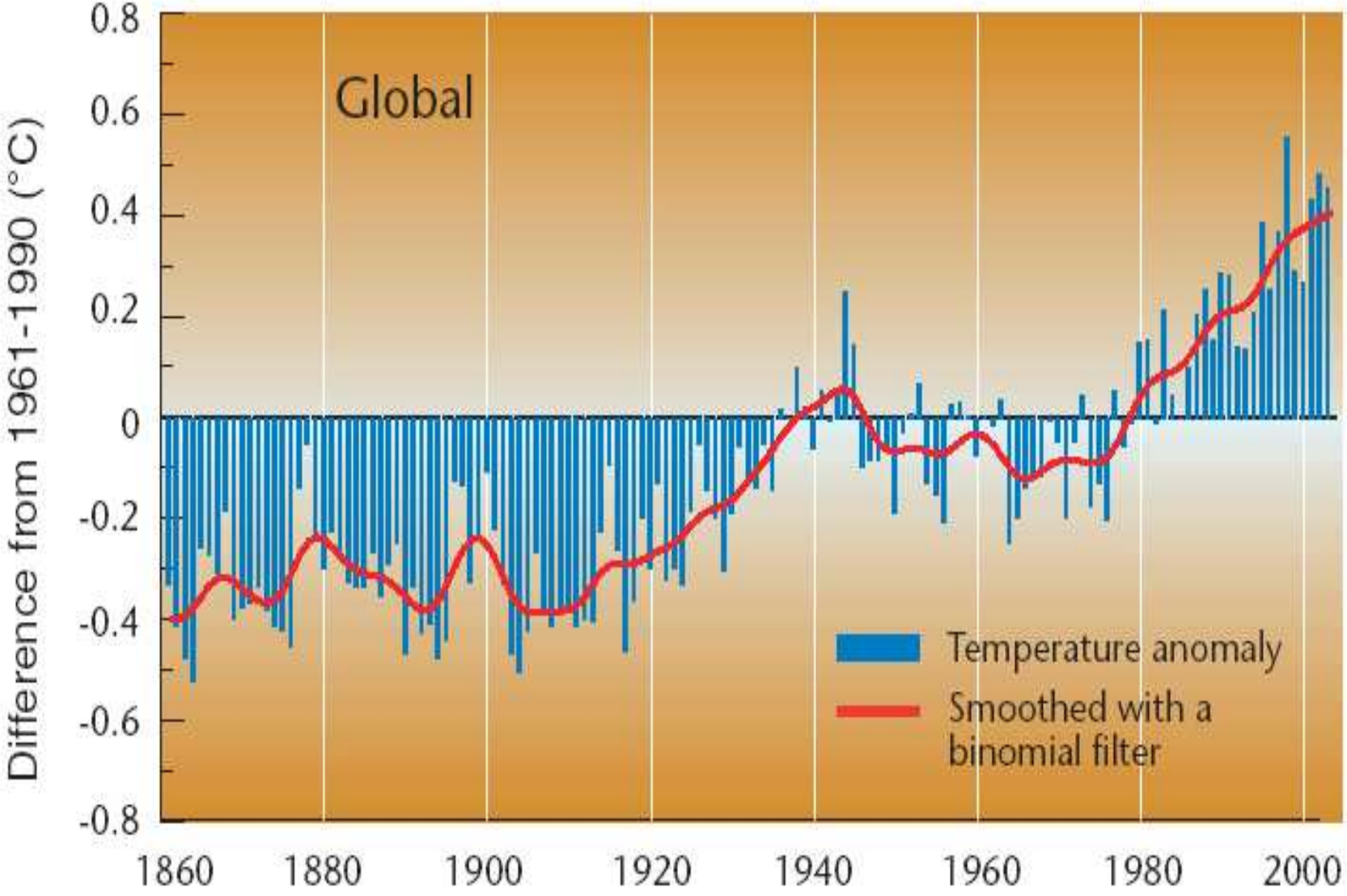
1.1 Introduction

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

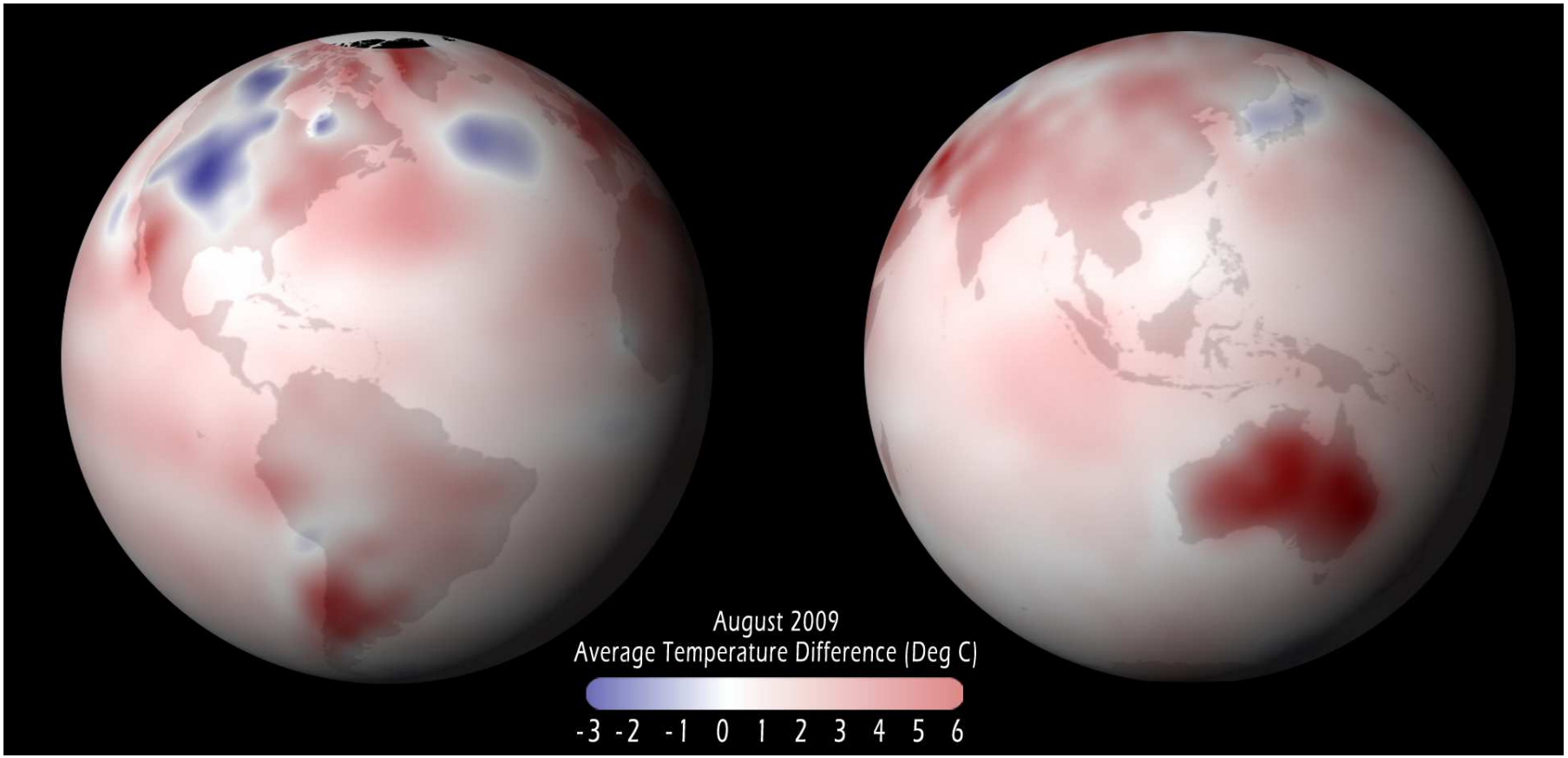
Evidence for climatic change has been collected from a variety of sources, some of which can be used to reconstruct the earth's changing climates over **tens of thousands of years**.

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.

1.1 Introduction



1.1 Introduction



1.1 Introduction

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**.

1.1 Introduction

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**.

1.1 Introduction

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

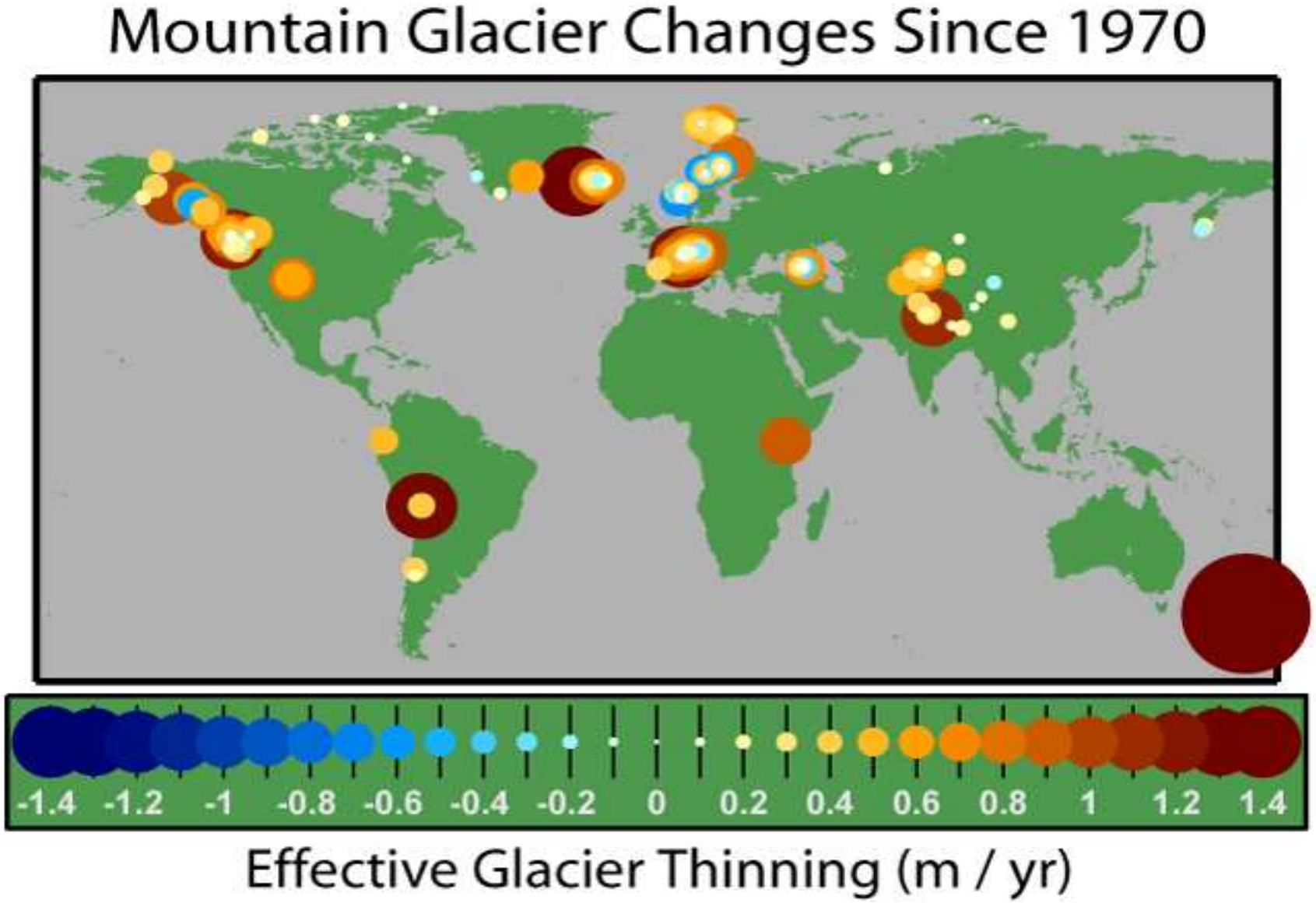
As the earth warms, glaciers **retreat** and **ice sheets melt**.

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

1.1 Introduction

Retreat of Lyall Glacier, California

1.1 Introduction



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, Sandy)
  - More frequent tropical cyclones in the Indian Ocean/Pacific Ocean (e.g. Bangladesh, India, Malaysia)
  - Flooding in the U.K. – 2012 being the second wettest year on record

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, Sandy)
  - More frequent tropical cyclones in the Indian Ocean/Pacific Ocean (e.g. Bangladesh, India, Malaysia)
  - Flooding in the U.K. – 2012 being the second wettest year on record

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, Sandy)
  - More frequent tropical cyclones in the Indian Ocean/Pacific Ocean (e.g. Bangladesh, India, Malaysia)
  - Flooding in the U.K. – 2012 being the second wettest year on record

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, Sandy)
  - More frequent tropical cyclones in the Indian Ocean/Pacific Ocean (e.g. Bangladesh, India, Malaysia)
  - Flooding in the U.K. – 2012 being the second wettest year on record

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, Sandy)
  - More frequent tropical cyclones in the Indian Ocean/Pacific Ocean (e.g. Bangladesh, India, Malaysia)
  - Flooding in the U.K. – 2012 being the second wettest year on record

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, Sandy)
  - More frequent tropical cyclones in the Indian Ocean/Pacific Ocean (e.g. Bangladesh, India, Malaysia)
  - Flooding in the U.K. – 2012 being the second wettest year on record

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, Sandy)
  - More frequent tropical cyclones in the Indian Ocean/Pacific Ocean (e.g. Bangladesh, India, Malaysia)
  - Flooding in the U.K. – 2012 being the second wettest year on record

1.1 Introduction

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

Though some parts of the world are familiar with the effects of extreme heat, we are beginning to see an increase in severe hot spells in more temperate latitudes – for example, the **2003 European heatwave** which resulted in thousands of deaths.

In drier parts of the world, this has resulted in an increase in devastating forest fires (**New South Wales** and **Victoria**), and in developing counties an increase in famine.

1.1 Introduction

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

Though some parts of the world are familiar with the effects of extreme heat, we are beginning to see an increase in severe hot spells in more temperate latitudes – for example, the **2003 European heatwave** which resulted in thousands of deaths.

In drier parts of the world, this has resulted in an increase in devastating forest fires (**New South Wales** and **Victoria**), and in developing counties an increase in famine.

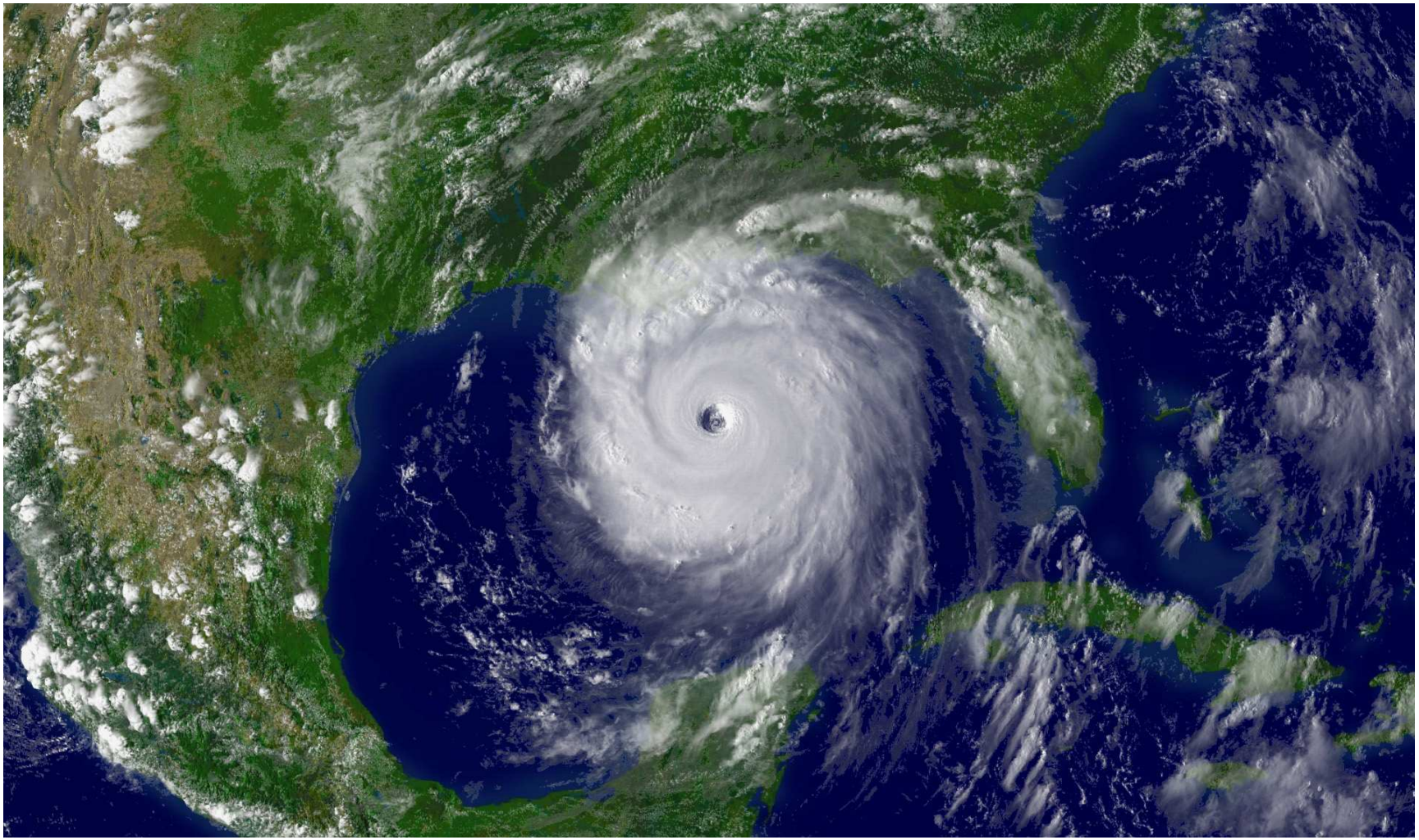
1.1 Introduction

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

Though some parts of the world are familiar with the effects of extreme heat, we are beginning to see an increase in severe hot spells in more temperate latitudes – for example, the **2003 European heatwave** which resulted in thousands of deaths.

In drier parts of the world, this has resulted in an increase in devastating forest fires (**New South Wales** and **Victoria**), and in developing counties an increase in famine.

1.2 Examples: Hurricane Katrina



1.2 Examples: 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**"

1.2 Examples: 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**"

1.2 Examples: 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**"

1.2 Examples: 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**"

1.2 Examples: 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**"

1.2 Examples: 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**"

1.2 Examples: 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**"

1.2 Examples: 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**"

1.2 Examples: 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**"

1.2 Examples: Hurricane Katrina

1.2 Examples: Hurricane Katrina

1.2 Examples: Hurricane Katrina

1.2 Examples: Hurricane Katrina

1.2 Examples: Hurricane Sandy

1.2 Examples: Hurricane Sandy

1.2 Examples: Hurricane Sandy

1.2 Examples: Hurricane Sandy

1.2 Examples: Hurricane Sandy

- Devastated parts of the **Caribbean**, **Mid–Atlantic** and **New Jersey/New York**
- Tenth hurricane of **busy 2012 hurricane season**
- Caused an estimated **$ 65.6 bn** worth of damage
- Second costliest hurricane in US history (after Katrina)
- Killed **253** people
- Sea surge reached **10 feet**

1.2 Examples: Hurricane Sandy

- Devastated parts of the **Caribbean**, **Mid–Atlantic** and **New Jersey/New York**
- Tenth hurricane of **busy 2012 hurricane season**
- Caused an estimated **$ 65.6 bn** worth of damage
- Second costliest hurricane in US history (after Katrina)
- Killed **253** people
- Sea surge reached **10 feet**

1.2 Examples: Hurricane Sandy

- Devastated parts of the **Caribbean**, **Mid–Atlantic** and **New Jersey/New York**
- Tenth hurricane of **busy 2012 hurricane season**
- Caused an estimated **$ 65.6 bn** worth of damage
- Second costliest hurricane in US history (after Katrina)
- Killed **253** people
- Sea surge reached **10 feet**

1.2 Examples: Hurricane Sandy

- Devastated parts of the **Caribbean**, **Mid–Atlantic** and **New Jersey/New York**
- Tenth hurricane of **busy 2012 hurricane season**
- Caused an estimated **$ 65.6 bn** worth of damage
- Second costliest hurricane in US history (after Katrina)
- Killed **253** people
- Sea surge reached **10 feet**

1.2 Examples: Hurricane Sandy

- Devastated parts of the **Caribbean**, **Mid–Atlantic** and **New Jersey/New York**
- Tenth hurricane of **busy 2012 hurricane season**
- Caused an estimated **$ 65.6 bn** worth of damage
- Second costliest hurricane in US history (after Katrina)
- Killed **253** people
- Sea surge reached **10 feet**

1.2 Examples: Hurricane Sandy

- Devastated parts of the **Caribbean**, **Mid–Atlantic** and **New Jersey/New York**
- Tenth hurricane of **busy 2012 hurricane season**
- Caused an estimated **$ 65.6 bn** worth of damage
- Second costliest hurricane in US history (after Katrina)
- Killed **253** people
- Sea surge reached **10 feet**

1.2 Examples: Hurricane ????, USA, 2016?

1.2 Examples: Extreme weather in the UK

The UK is no stranger to changing climate.

- 2012 was the **second wettest** year in the UK since records began
- 1998, 1999, 2000, 2002, 2008 and 2012 in the **top ten** wettest years
- Examples:
  - Flash–flooding in **Glasgow** (2002)
  - Flooding in **Boscastle** (2004)
  - **Entire UK** (2007, 2008, 2009, 2012) – including severe flash–flooding in Newcastle in June 2012

1.2 Examples: Extreme weather in the UK

The UK is no stranger to changing climate.

- 2012 was the **second wettest** year in the UK since records began
- 1998, 1999, 2000, 2002, 2008 and 2012 in the **top ten** wettest years
- Examples:
  - Flash–flooding in **Glasgow** (2002)
  - Flooding in **Boscastle** (2004)
  - **Entire UK** (2007, 2008, 2009, 2012) – including severe flash–flooding in Newcastle in June 2012

1.2 Examples: Extreme weather in the UK

The UK is no stranger to changing climate.

- 2012 was the **second wettest** year in the UK since records began
- 1998, 1999, 2000, 2002, 2008 and 2012 in the **top ten** wettest years
- Examples:
  - Flash–flooding in **Glasgow** (2002)
  - Flooding in **Boscastle** (2004)
  - **Entire UK** (2007, 2008, 2009, 2012) – including severe flash–flooding in Newcastle in June 2012

1.2 Examples: Extreme weather in the UK

The UK is no stranger to changing climate.

- 2012 was the **second wettest** year in the UK since records began
- 1998, 1999, 2000, 2002, 2008 and 2012 in the **top ten** wettest years
- Examples:
  - Flash–flooding in **Glasgow** (2002)
  - Flooding in **Boscastle** (2004)
  - **Entire UK** (2007, 2008, 2009, 2012) – including severe flash–flooding in Newcastle in June 2012

1.2 Examples: Extreme weather in the UK

The UK is no stranger to changing climate.

- 2012 was the **second wettest** year in the UK since records began
- 1998, 1999, 2000, 2002, 2008 and 2012 in the **top ten** wettest years
- Examples:
  - Flash–flooding in **Glasgow** (2002)
  - Flooding in **Boscastle** (2004)
  - **Entire UK** (2007, 2008, 2009, 2012) – including severe flash–flooding in Newcastle in June 2012

1.2 Examples: Extreme weather in the UK

The UK is no stranger to changing climate.

- 2012 was the **second wettest** year in the UK since records began
- 1998, 1999, 2000, 2002, 2008 and 2012 in the **top ten** wettest years
- Examples:
  - Flash–flooding in **Glasgow** (2002)
  - Flooding in **Boscastle** (2004)
  - **Entire UK** (2007, 2008, 2009, 2012) – including severe flash–flooding in Newcastle in June 2012

1.2 Examples: Extreme weather in the UK

The UK is no stranger to changing climate.

- 2012 was the **second wettest** year in the UK since records began
- 1998, 1999, 2000, 2002, 2008 and 2012 in the **top ten** wettest years
- Examples:
  - Flash–flooding in **Glasgow** (2002)
  - Flooding in **Boscastle** (2004)
  - **Entire UK** (2007, 2008, 2009, 2012) – including severe flash–flooding in Newcastle in June 2012

1.2 Examples: Flooding in Newcastle, June 2012

1.2 Examples: Flooding in Newcastle, June 2012

1.2 Examples: Flooding in Newcastle, June 2012

1.2 Examples: Flooding in Newcastle, June 2012

1.2 Examples: The Great North Sea Flood, 1953

1.2 Examples: The Great North Sea Flood, 2025?

1.2 Examples: 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”

1.2 Examples: 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”

1.2 Examples: 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”

1.2 Examples: 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”

1.2 Examples: 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”

1.2 Examples: Tornadoes in Birmingham, 2005

1.2 Examples: Tornadoes in Birmingham, 2005

1.2 Examples: Heatwaves and drought

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

- Not a new thing for some parts of the world!
- Hot spells becoming more frequent in other areas:
- **Heatwaves** in Europe 2003, 2006
- Increase in frequency of **forest fires** in parts of Europe, Canada and Australia

1.2 Examples: Heatwaves and drought

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

- Not a new thing for some parts of the world!
- Hot spells becoming more frequent in other areas:
- **Heatwaves** in Europe 2003, 2006
- Increase in frequency of **forest fires** in parts of Europe, Canada and Australia

1.2 Examples: Heatwaves and drought

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

- Not a new thing for some parts of the world!
- Hot spells becoming more frequent in other areas:
- **Heatwaves** in Europe 2003, 2006
- Increase in frequency of **forest fires** in parts of Europe, Canada and Australia

1.2 Examples: Heatwaves and drought

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

- Not a new thing for some parts of the world!
- Hot spells becoming more frequent in other areas:
- **Heatwaves** in Europe 2003, 2006
- Increase in frequency of **forest fires** in parts of Europe, Canada and Australia

1.2 Examples: Heatwaves and drought

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

- Not a new thing for some parts of the world!
- Hot spells becoming more frequent in other areas:
- **Heatwaves** in Europe 2003, 2006
- Increase in frequency of **forest fires** in parts of Europe, Canada and Australia

1.3 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**.

1.3 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**.

1.3 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**.

1.3 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;
  - 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.

1.3 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;
  - 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.

1.3 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;
  - 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.

1.3 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;
  - 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.

1.3 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;
  - 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.

1.3 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;
  - 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.

1.3 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;
  - 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.

1.3 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;
  - 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.

1.3 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.

1.3 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.

1.3 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.

1.3 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.

1.3 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.

1.3 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.

1.4 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 1955–2004.

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

1.4 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 1955–2004.

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

1.4 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 1955–2004.

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

1.4 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 1955–2004.

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

1.4 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:

$$\Pr(\text{sea–surge exceeds 8.75 feet}) \quad =$$

Solution

Estimating these probabilities using relative frequencies gives:

$$\Pr(\text{sea–surge exceeds 8.75 feet}) \quad =$$

Solution

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**

Solution

Estimating these probabilities using relative frequencies gives:

$$\Pr(\text{sea–surge exceeds 8.75 feet}) = \frac{33}{50} = 0.66$$

$$\Pr(\text{sea–surge exceeds 11.25 feet}) = \frac{6}{50} = 0.12$$

$$\Pr(\text{sea–surge exceeds 14 feet}) = \frac{0}{50} = 0.$$

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

Solution

Estimating these probabilities using relative frequencies gives:

$$\Pr(\text{sea–surge exceeds 8.75 feet}) = \frac{33}{50} = 0.66$$

$$\Pr(\text{sea–surge exceeds 11.25 feet}) = \frac{6}{50} = 0.12$$

$$\Pr(\text{sea–surge exceeds 14 feet}) = \frac{0}{50} = 0.$$

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

Solution

Estimating these probabilities using relative frequencies gives:

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

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

Solution

Estimating these probabilities using relative frequencies gives:

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

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

1.4 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

2.1 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**

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

2.1 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**

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

2.1 Background and theoretical motivation

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

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

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

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

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

2.1 Background and theoretical motivation

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

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

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

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

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

2.1 Background and theoretical motivation

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

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

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

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

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

The assumption of IID might be reasonable here.

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

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

2.1 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.

2.1 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.

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

This is trivial (in principle), since

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

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.

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.

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.

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.

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.

2.1.1 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.

2.1.1 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.

2.1.1 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.

2.1.1 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.

2.1.1 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.

2.1.1 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.

2.1.1 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.

2.1.1 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**.

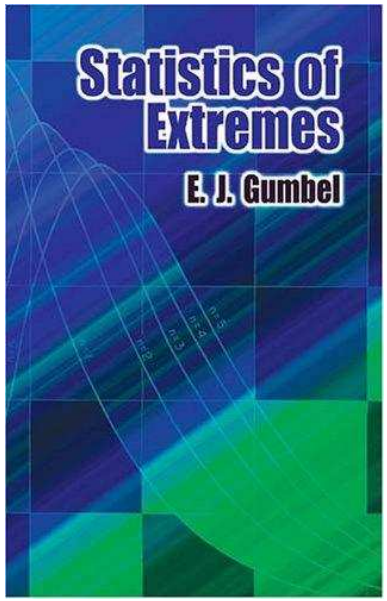
2.1.1 A brief history of Extreme Value Theory

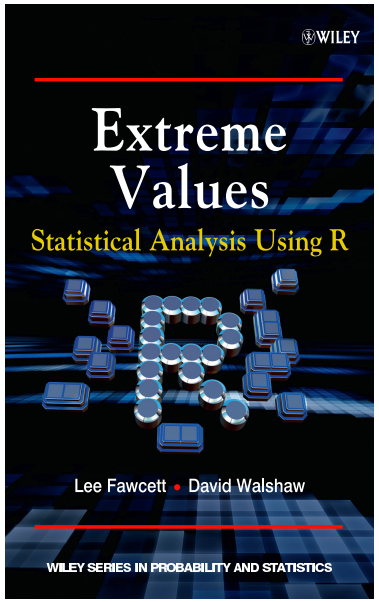
Since the 1980s, methods for the application of Extreme Value Theory have become much more widespread.

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

2.1.1 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**).





2.1.2 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?

2.1.2 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?

2.1.2 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?

2.1.2 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?

2.1.2 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 ∞

2.1.2 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 ∞

2.1.2 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 ∞

2.1.2 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 ∞

2.1.2 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?

2.1.2 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?

2.1.2 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?

2.1.2 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?

2.1.2 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.

Theorem *(The Extremal Types Theorem)*

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

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

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

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

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

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

Theorem *(The Extremal Types Theorem)*

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

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

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

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

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

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

Theorem *(The Extremal Types Theorem)*

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

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

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

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

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

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

Theorem *(The Extremal Types Theorem)*

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

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

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

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

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

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

2.1.2 The Extremal Types Theorem

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

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

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

2.1.2 The Extremal Types Theorem

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

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

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

2.1.2 The Extremal Types Theorem

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

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

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

2.1.2 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.*

2.1.2 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.*

2.1.2 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.*

2.1.2 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?

2.1.2 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?

2.1.3 The Generalised Extreme Value distribution

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

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

2.1.3 The Generalised Extreme Value distribution

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

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

2.1.3 The Generalised Extreme Value distribution

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

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

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

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

2.1.3 The Generalised Extreme Value distribution

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

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

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

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

2.1.3 The Generalised Extreme Value distribution

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

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

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

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

2.1.3 The Generalised Extreme Value distribution

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

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

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

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

2.1.3 The Generalised Extreme Value distribution

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

Shape parameter ξ:

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

2.1.3 The Generalised Extreme Value distribution

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

Shape parameter ξ:

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

2.1.3 The Generalised Extreme Value distribution

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

Shape parameter ξ:

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

2.1.3 The Generalised Extreme Value distribution

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

Shape parameter ξ:

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

2.1.3 The Generalised Extreme Value distribution

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

Shape parameter ξ:

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

2.1.3 The Generalised Extreme Value distribution

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

Shape parameter ξ:

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

2.1.3 The Generalised Extreme Value distribution

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

Shape parameter ξ:

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

2.1.3 The Generalised Extreme Value distribution

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

Shape parameter ξ:

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

2.1.3 The Generalised Extreme Value distribution

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

Shape parameter ξ:

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

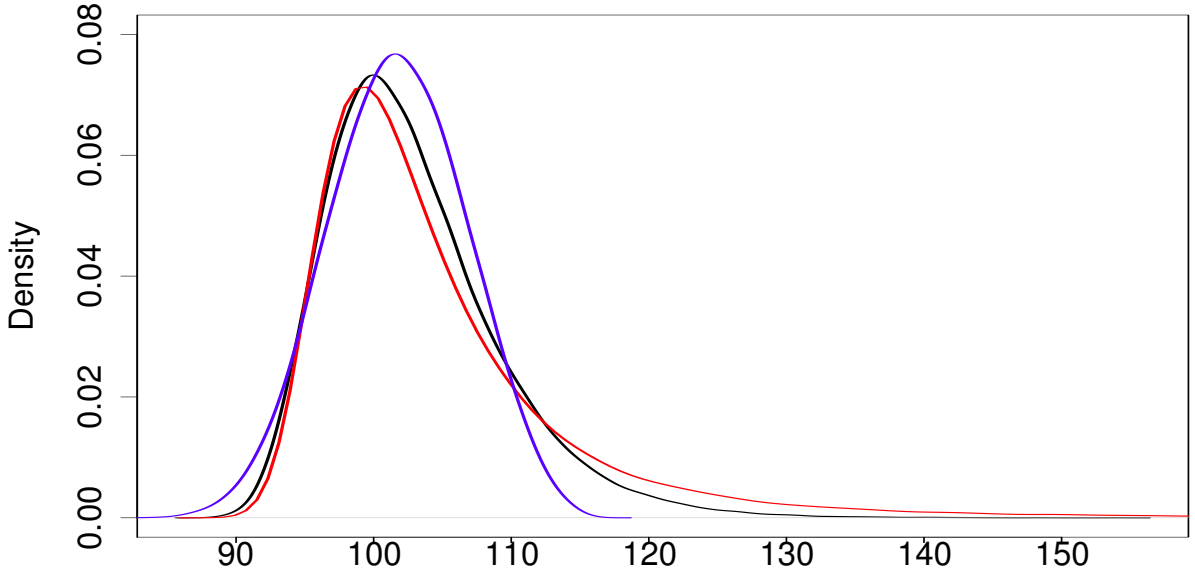
2.1.3 The Generalised Extreme Value distribution

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

Shape parameter ξ:

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

2.1.3 The Generalised Extreme Value distribution



2.1.3 The Generalised Extreme Value distribution

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

We know that

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

After some algebra, it turns out that

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

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

2.1.3 The Generalised Extreme Value distribution

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

We know that

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

After some algebra, it turns out that

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

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

2.1.3 The Generalised Extreme Value distribution

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

We know that

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

After some algebra, it turns out that

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

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

2.1.3 The Generalised Extreme Value distribution

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

We know that

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

After some algebra, it turns out that

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

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

2.1.3 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.

2.1.3 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.

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

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.

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

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.

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

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.

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

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.

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

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.

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

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

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

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

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

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

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

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:

$$\Pr\left\{\frac{M_n - b_n}{a_n} \leq z\right\} = \Pr\left\{X_1 \leq a_n z + b_n\right\} \times \cdots$$

$$= \left[\exp\left\{-\frac{1}{a_n z + b_n}\right\}\right]^n.$$

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:

$$\Pr\left\{\frac{M_n - b_n}{a_n} \leq z\right\} = \Pr\left\{X_1 \leq a_n z + b_n\right\} \times \cdots$$

$$= \left[\exp\left\{-\frac{1}{a_n z + b_n}\right\}\right]^n.$$

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:

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

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.

2.1.4 Typical application – 1: Data pre–processing

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

2.1.4 Typical application – 1: Data pre–processing

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

2.1.4 Typical application – 1: Data pre–processing

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

2.1.4 Typical application – 1: Data pre–processing

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

2.1.4 Typical application – 1: Data pre–processing

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

2.1.4 Typical application – 2: Parameter estimation

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

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

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

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

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

(try this yourself!)

2.1.4 Typical application – 2: Parameter estimation

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

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

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

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

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

(try this yourself!)

2.1.4 Typical application – 2: Parameter estimation

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

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

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

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

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

(try this yourself!)

2.1.4 Typical application – 2: Parameter estimation

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

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

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

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

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

(try this yourself!)

2.1.4 Typical application – 2: Parameter estimation

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

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

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

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

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

(try this yourself!)

2.1.4 Typical application – 2: Parameter estimation

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

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

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

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

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

(try this yourself!)

2.1.4 Typical application – 2: Parameter estimation

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

2.1.4 Typical application – 2: Parameter estimation

The likelihood is given by

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

2.1.4 Typical application – 2: Parameter estimation

The likelihood is given by

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

2.1.4 Typical application – 2: Parameter estimation

The log–likelihood is given by

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

2.1.4 Typical application – 2: Parameter estimation

The log–likelihood is given by

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

2.1.4 Typical application – 2: Parameter estimation

The log–likelihood is given by

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

2.1.4 Typical application – 2: Parameter estimation

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

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

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

2.1.4 Typical application – 2: Parameter estimation

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

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

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

2.1.4 Typical application – 2: Parameter estimation

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

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

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

2.1.4 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?

2.1.4 Typical application – 2: Parameter estimation

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

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

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

2.1.4 Typical application – 2: Parameter estimation

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

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

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

2.1.4 Typical application – 2: Parameter estimation

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

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

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

2.1.4 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.

2.1.4 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.

2.1.4 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.

2.1.4 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.

2.1.4 Typical application – 4: Return level estimation

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

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

2.1.4 Typical application – 4: Return level estimation

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

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

2.1.4 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).

2.1.4 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).

2.1.4 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.

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

2.1.4 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.

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

2.1.4 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.

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

2.1.4 Typical application – 4: Return level estimation

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

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

giving

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

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

2.1.4 Typical application – 4: Return level estimation

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

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

giving

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

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

2.1.4 Typical application – 4: Return level estimation

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

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

giving

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

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

2.1.4 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

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

2.1.4 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

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

2.1.4 Typical application – 4: Return level estimation

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

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

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

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

2.1.4 Typical application – 4: Return level estimation

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

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

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

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

2.1.4 Typical application – 4: Return level estimation

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

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

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

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

2.1.4 Typical application – 4: Return level estimation

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

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

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

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

2.2 Sea surge at Wassaw Island

Recall the example from Section 1.4.

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

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

2.2 Sea surge at Wassaw Island

Recall the example from Section 1.4.

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

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

2.2 Sea surge at Wassaw Island

Recall the example from Section 1.4.

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

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

2.2 Sea surge at Wassaw Island

Recall the example from Section 1.4.

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

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

2.2 Sea surge at Wassaw Island

Recall the example from Section 1.4.

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

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

2.2 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**?

2.2 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**?

2.2.1 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 2.2 shows a time series plot and histogram of the 50 annual maxima:

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

2.2.1 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 2.2 shows a time series plot and histogram of the 50 annual maxima:

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

2.2.1 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 2.2 shows a time series plot and histogram of the 50 annual maxima:

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

2.2.1 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 2.2 shows a time series plot and histogram of the 50 annual maxima:

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

2.2.1 Data pre–processing

In R:

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

Or:

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

2.2.1 Data pre–processing

In R:

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

Or:

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

In R:

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

Then, to produce the plots shown in Figure 2.2:

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

In R:

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

Then, to produce the plots shown in Figure 2.2:

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

In R:

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

Then, to produce the plots shown in Figure 2.2:

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

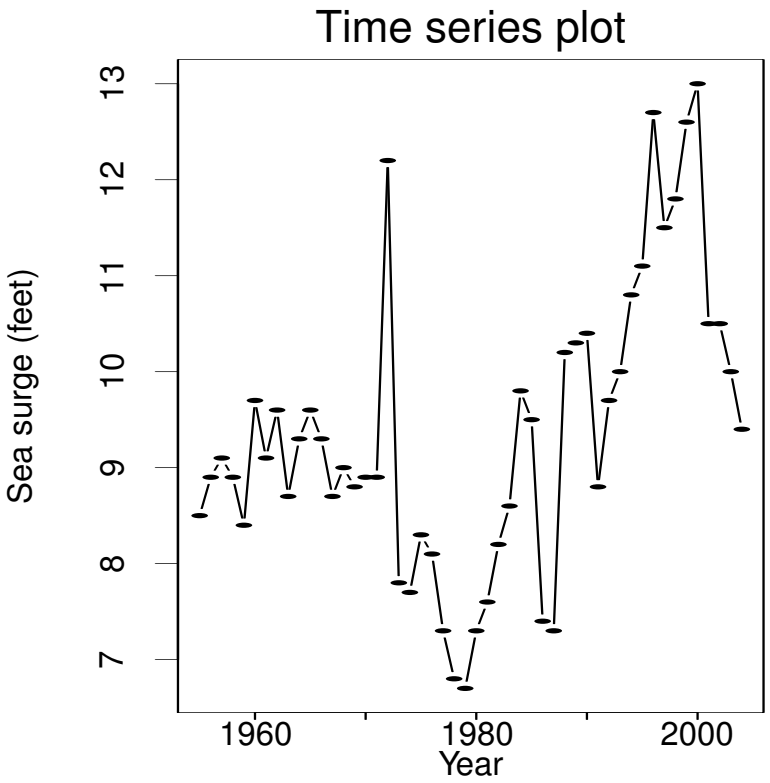
In R:

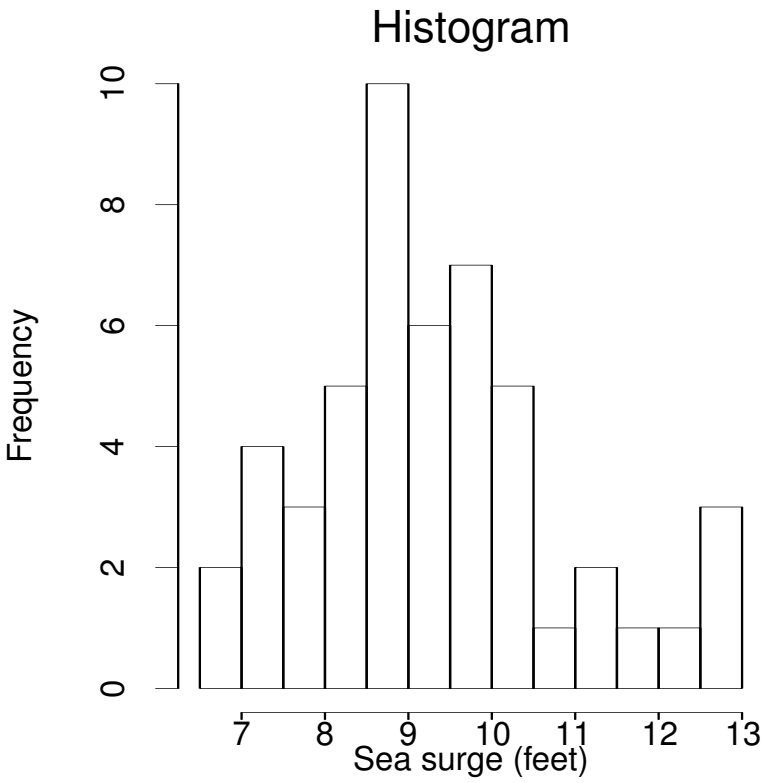
```
> year=seq(1955,2004,1)
    # Sequence of values from 1995 to 2004
```

Then, to produce the plots shown in Figure 2.2:

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

2.2.1 Data pre–processing





2.2.2 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)
```

2.2.2 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 $\theta = (\mu,\sigma,\xi)$

2.2.2 Fitting the GEV: Parameter estimation

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

2.2.2 Fitting the GEV: Parameter estimation

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

2.2.2 Fitting the GEV: Parameter estimation

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

2.2.2 Fitting the GEV: Parameter estimation

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

2.2.2 Fitting the GEV: Parameter estimation

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

2.2.2 Fitting the GEV: Parameter estimation

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

2.2.2 Fitting the GEV: Parameter estimation

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

In R:

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

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

$minimum
[1] 89.52412

$estimate
[1] 8.7112735 1.3114836 -0.1084451

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

$code
[1] 1
```

In R:

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

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

$minimum
[1] 89.52412

$estimate
[1] 8.7112735 1.3114836 -0.1084451

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

$code
[1] 1
```

2.2.2 Fitting the GEV: Parameter estimation

In R:

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

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

$minimum
[1] 89.52412

$estimate
[1] 8.7112735 1.3114836 -0.1084451

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

$code
[1] 1
```

In R:

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

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

$minimum
[1] 89.52412

$estimate
[1] 8.7112735 1.3114836 -0.1084451

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

$code
[1] 1
```

In R:

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

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

$minimum
[1] 89.52412

$estimate
[1] 8.7112735 1.3114836 -0.1084451

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

$code
[1] 1
```

In R:

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

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

$minimum
[1] 89.52412

$estimate
[1] 8.7112735 1.3114836 -0.1084451

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

$code
[1] 1
```

2.2.2 Fitting the GEV: Standard errors

In R:

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

Now:

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

2.2.2 Fitting the GEV: Standard errors

In R:

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

Now:

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

2.2.2 Fitting the GEV: Standard errors

In R:

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

Now:

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

2.2.2 Fitting the GEV: Standard errors

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

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

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

evaluated at the MLEs.

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

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

2.2.2 Fitting the GEV: Standard errors

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

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

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

evaluated at the MLEs.

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

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

2.2.2 Fitting the GEV: Standard errors

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

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

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

evaluated at the MLEs.

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

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

2.2.2 Fitting the GEV: Standard errors

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

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

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

evaluated at the MLEs.

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

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

2.2.2 Fitting the GEV: Standard errors

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

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

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

evaluated at the MLEs.

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

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

2.2.2 Fitting the GEV: Standard errors

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

Thus, we have a 3×3 matrix to invert!

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

2.2.2 Fitting the GEV: Standard errors

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

Thus, we have a 3×3 matrix to invert!

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

2.2.2 Fitting the GEV: Standard errors

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

Thus, we have a 3×3 matrix to invert!

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

2.2.2 Fitting the GEV: Standard errors

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

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

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

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

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

2.2.2 Fitting the GEV: Standard errors

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

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

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

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

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

2.2.2 Fitting the GEV: Standard errors

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

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

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

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

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

2.2.2 Fitting the GEV: Standard errors

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

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

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

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

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

2.2.2 Fitting the GEV: Standard errors

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

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

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

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

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

2.2.2 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
```

2.2.2 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
```

2.2.2 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
```

2.2.2 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
```

2.2.2 Fitting the GEV: Standard errors

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

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

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

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

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

for μ, σ and ξ (respectively).

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

2.2.2 Fitting the GEV: Standard errors

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

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

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

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

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

for μ, σ and ξ (respectively).

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

2.2.2 Fitting the GEV: Standard errors

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

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

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

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

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

for μ, σ and ξ (respectively).

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

2.2.2 Fitting the GEV: Standard errors

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

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

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

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

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

for μ, σ and ξ (respectively).

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

2.2.2 Fitting the GEV: Standard errors

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

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

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

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

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

for μ, σ and ξ (respectively).

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

2.2.2 Fitting the GEV: Standard errors

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

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

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

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

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

for μ, σ and ξ (respectively).

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

2.2.3 Model adequacy: Probability plots

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

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

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

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

2.2.3 Model adequacy: Probability plots

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

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

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

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

2.2.3 Model adequacy: Probability plots

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

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

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

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

2.2.3 Model adequacy: Probability plots

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

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

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

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

2.2.3 Model adequacy: Probability plots

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

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

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

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

2.2.3 Model adequacy: Probability plots

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

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

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

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

2.2.3 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
$\vdots$	$\vdots$	$\vdots$

2.2.3 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
$\vdots$	$\vdots$	$\vdots$

2.2.3 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
$\vdots$	$\vdots$	$\vdots$

2.2.3 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
$\vdots$	$\vdots$	$\vdots$

2.2.3 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**
$\vdots$	$\vdots$	$\vdots$

2.2.3 Probability plot in R

In R:

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

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

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

2.2.3 Probability plot in R

In R:

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

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

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

2.2.3 Probability plot in R

In R:

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

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

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

2.2.3 Probability plot in R

In R:

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

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

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

2.2.3 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)}
```

2.2.3 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)}
```

2.2.3 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.

2.2.3 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.

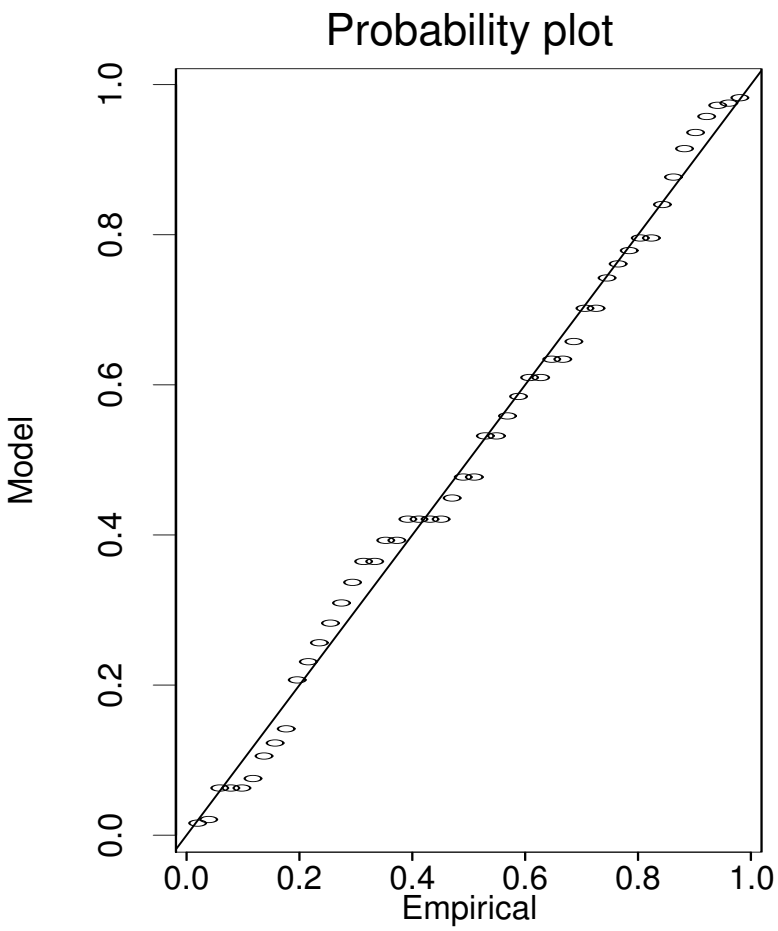
2.2.3 Probability plot in R

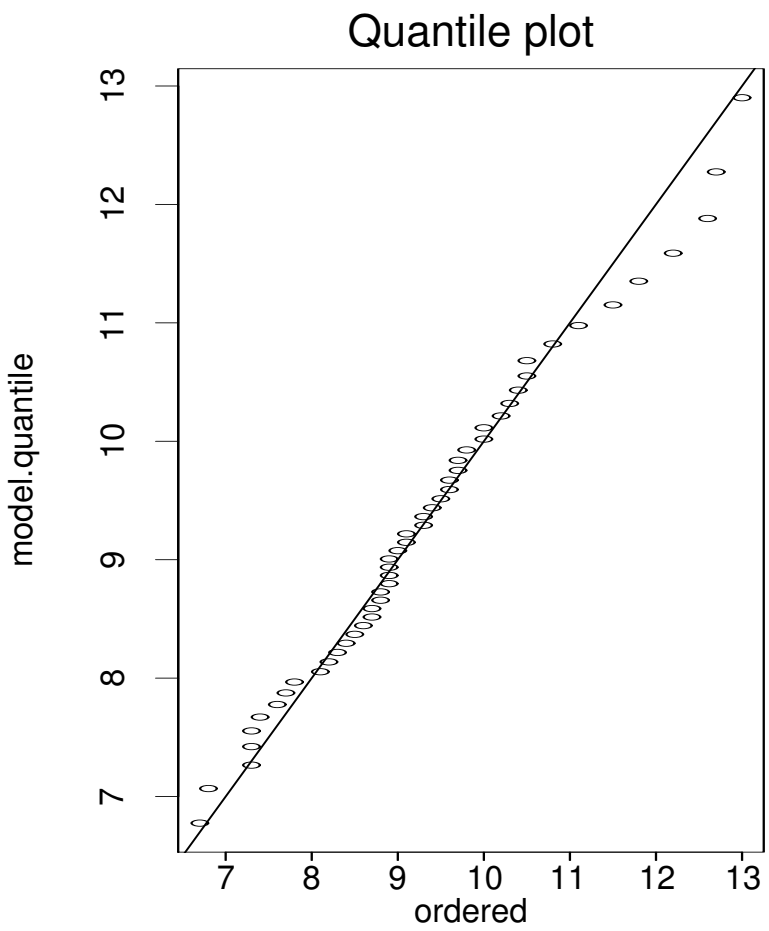
Then:

```
> 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.

2.2.3 Probability/Quantile plots in R





2.2.4 Return level estimation

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

(i) once in ten years;

(ii) once in a hundred years.

2.2.4 Return level estimation

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

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

Using Equation (2.10):

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

we find that

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

similarly,

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

2.2.4 Return level estimation

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

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

Using Equation (2.10):

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

we find that

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

similarly,

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

2.2.4 Return level estimation

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

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

Using Equation (2.10):

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

we find that

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

similarly,

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

2.2.4 Return level estimation

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

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

Using Equation (2.10):

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

we find that

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

similarly,

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

2.2.4 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.

2.2.4 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.

2.2.4 Return level estimation

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

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

Specifically,

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

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

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

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

2.2.4 Return level 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,

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

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

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

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

2.2.4 Return level 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,

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

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

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

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

2.2.4 Return level 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,

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

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

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

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

2.2.4 Return level standard errors in R

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

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

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

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

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

2.2.4 Return level standard errors in R

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

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

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

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

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

2.2.4 Return level standard errors in R

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

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

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

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

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

2.2.4 Return level standard errors in R

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

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

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

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

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

2.2.4 Return level standard errors in R

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

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

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

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

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

2.2.4 Return level standard errors in R

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

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

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

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

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

2.2.4 Return level standard errors in R

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

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

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

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

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

2.2.4 Return level standard errors in R

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

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

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

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

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

2.2.4 Return level standard errors in R

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

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

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

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

2.2.4 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.

2.2.4 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.

2.2.4 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.

2.2.4 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.

2.2.4 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.

2.2.4 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.

2.2.4 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.

2.2.4 Return level estimation

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

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

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

2.2.4 Return level estimation

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

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

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

2.2.4 Return level estimation

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

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

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

Class demonstration in R

Example 2.3: Rainfall in the Lake District

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

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

Example 2.3: Rainfall in the Lake District

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

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

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

Example 2.3: Rainfall in the Lake District

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

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

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

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

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:**

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

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 $\boldsymbol{I}_O^T$:

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

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

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

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

3. **Find the determinants of each of the 2×2 minor matrices of $\boldsymbol{I}_O^T$:**

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

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

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

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:

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

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:

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

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 (2.10):

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

we get:

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

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

Using Equation (2.10):

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

we get:

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

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

Using Equation (2.10):

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

we get:

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

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

Using Equation (2.10):

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

we get:

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

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

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

Using Equation (2.10):

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

we get:

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

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

From earlier, we know that

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

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

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

Example 2.3: Solution to part(d) (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 very wide; also, the CI for $\hat{z}_{1000}$ is partly negative – not sensible!

Example 2.4: Solution

See R demo

Example 2.5

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

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

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

Example 2.5

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

```
gev.fit(haw)

$ conv
[1] 0

$nllh
[1] 178.2502

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

$se
[1] 21.6185212 21.3598264 0.3027163
```

Example 2.5

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

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

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

From Equation (2.10) we know that

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

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

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

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

From Equation (2.10) we know that

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

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

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

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

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

$$\begin{aligned}\hat{z}_4 &= 99980.28 + \frac{87.10824}{0.5921945}\left[\left(-\log(1-4^{-1})\right)^{-0.5921945} - 1\right] \\ &= 100140.8 \text{ kg.}\end{aligned}$$

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

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

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

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

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

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

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

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

For example, ordering the data gives:

$x_{(i)}$	$i/(n+1)$
$\vdots$	$\vdots$
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 2.5: Solution to part (a) (4/5)

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

For example, ordering the data gives:

$x_{(i)}$	$i/(n+1)$
$\vdots$	$\vdots$
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 2.5: Solution to part (a) (4/5)

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

For example, ordering the data gives:

$x_{(i)}$	$i/(n+1)$
$\vdots$	$\vdots$
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 2.5: Solution to part (a) (5/5)

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

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

The model gives a slight over–estimate.

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

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

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

The model gives a slight over–estimate.

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

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

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

The model gives a slight over–estimate.

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

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

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

The model gives a slight over–estimate.

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

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

The fitted density also follows the histogram very closely.

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

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

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

The fitted density also follows the histogram very closely.

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

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

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

The fitted density also follows the histogram very closely.

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

2.4 Profile likelihood

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

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

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

is not advisable for return levels.

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

2.4 Profile likelihood

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

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

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

is not advisable for return levels.

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

2.4 Profile likelihood

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

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

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

is not advisable for return levels.

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

2.4 Profile likelihood

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

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

2.4 Profile likelihood

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

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

2.4 Profile likelihood

The **profile** log–likelihood for θ_j is defined as

$$\ell_p(\theta_j) \quad = \quad \max_{\theta_{-j}} \ell(\theta_j, \theta_{-j}).$$

Thus, for each value of θ_j the profile log–likelihood is the maximised log–likelihood with respect to all other components of θ.

2.4 Profile likelihood

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

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

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

2.4 Profile likelihood

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

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

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

2.4 Profile likelihood

For return levels using the GEV model:

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

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

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

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

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

2.4 Profile likelihood

For return levels using the GEV model:

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

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

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

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

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

2.4 Profile likelihood

For return levels using the GEV model:

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

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

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

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

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

2.4 Profile likelihood

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

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

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

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

2.4 Profile likelihood

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

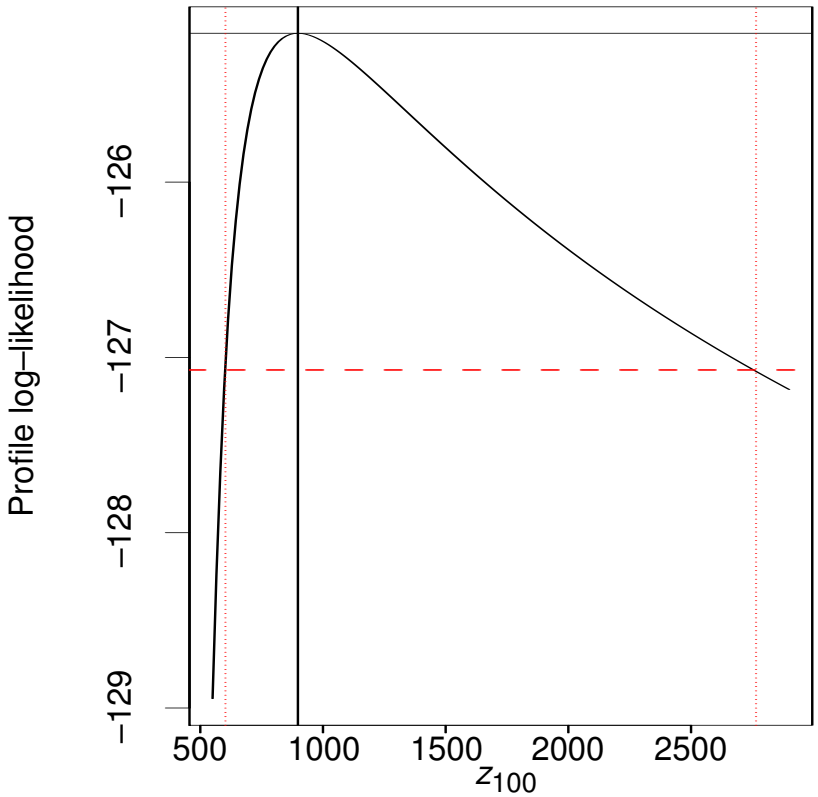
Figure 2.6 shows plots of the profile log likelihood for the 100 and 1000 year return levels ($\ell_p(z_{100})$ and $\ell_p(z_{1000})$ respectively) for the Lake District rainfall data in Example 2.3.

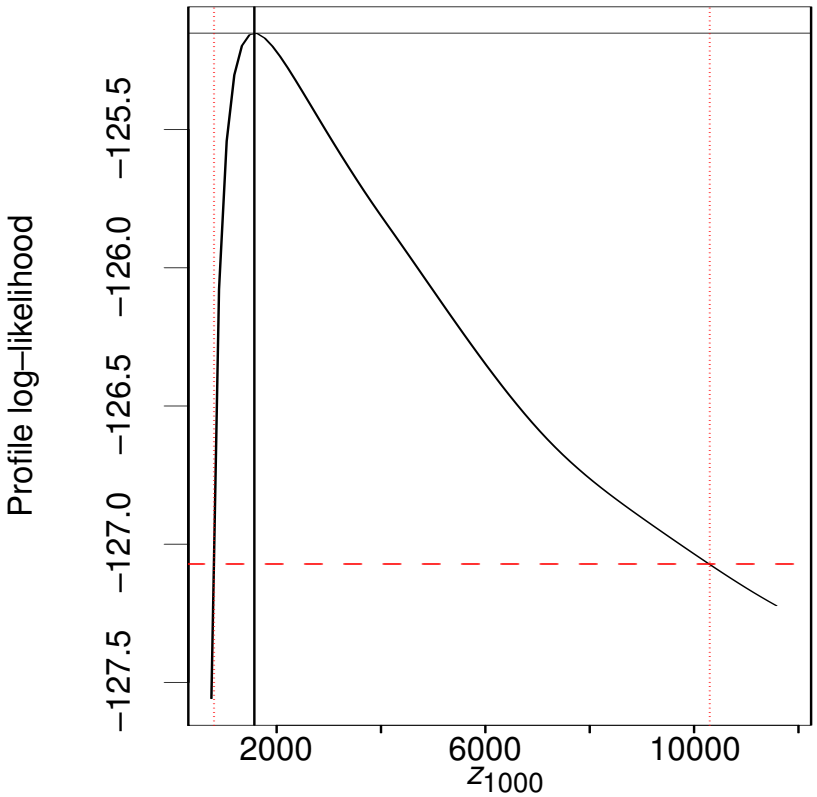
2.4 Profile likelihood

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

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

2.4 Profile likelihood





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

So how can we use these plots?

2.4 Profile likelihood

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

So how can we use these plots?

2.4 Profile likelihood

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

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

and the profile log–likelihood is now defined as

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

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

2.4 Profile likelihood

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

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

and the profile log–likelihood is now defined as

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

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

2.4 Profile likelihood

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

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

and the profile log–likelihood is now defined as

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

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

2.4 Profile likelihood

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

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

and the profile log–likelihood is now defined as

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

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

2.4 Profile likelihood

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

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

and the profile log–likelihood is now defined as

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

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

2.4 Profile likelihood: Result

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

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

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

2.4 Profile likelihood: Result

In *practice*:

- Obtain a plausible range of values for $D_p(z_r)$ from the χ^2_1 distribution.

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

$$D_p(z_r) \leq \chi^2_1(0.05) = 3.842.$$

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

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

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

2.4 Profile likelihood: Result

In *practice*:

- Obtain a plausible range of values for $D_p(z_r)$ from the χ^2_1 distribution.

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

$$D_p(z_r) \leq \chi^2_1(0.05) = 3.842.$$

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

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

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

2.4 Profile likelihood: Result

In *practice*:

- Obtain a plausible range of values for $D_p(z_r)$ from the χ^2_1 distribution.

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

$$D_p(z_r) \leq \chi^2_1(0.05) = 3.842.$$

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

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

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

2.4 Profile likelihood: Result

In *practice*:

- Obtain a plausible range of values for $D_p(z_r)$ from the χ^2_1 distribution.

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

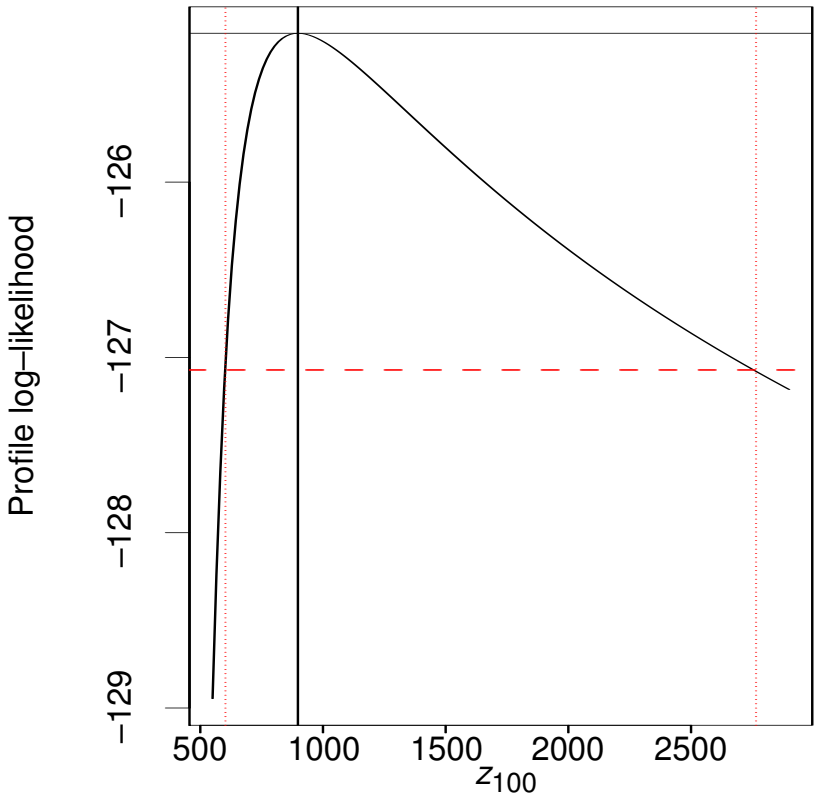
$$D_p(z_r) \leq \chi^2_1(0.05) = 3.842.$$

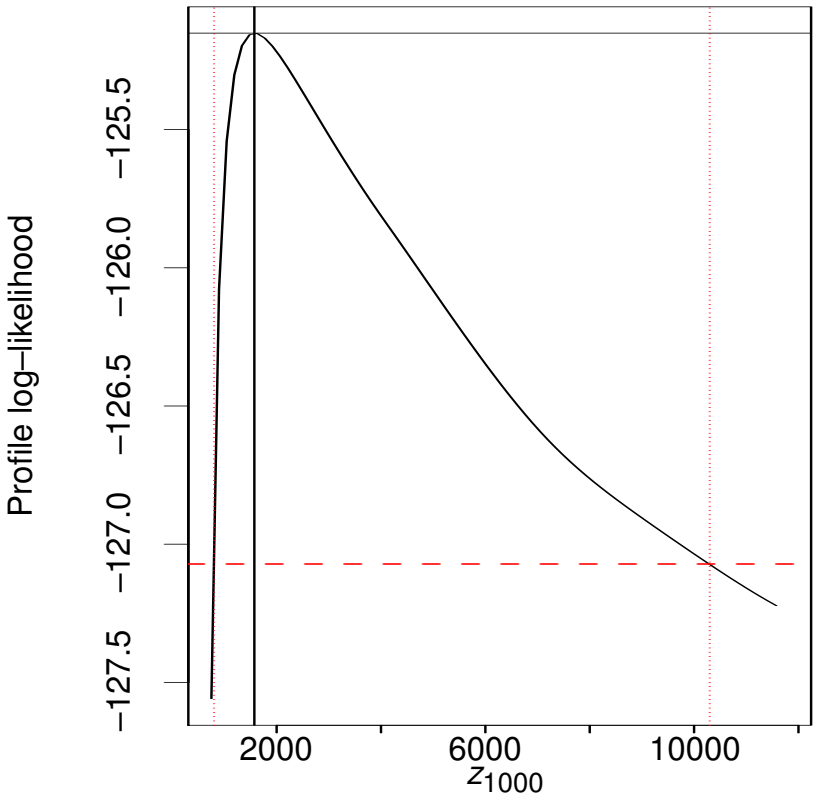
- Convert this into a range of plausible values for the profile log–likelihood:

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

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

2.4 Profile likelihood: Result





Profile Likelihood

	$\hat{z}_{100}$	$\hat{z}_{1000}$
Standard 95% CI	(327.5, 1468.8)mm	(–253.3, 3401.4)mm
Profiled 95% CI	(602.5, 2765)mm	(800, 103000)mm

Comments

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

Profile Likelihood

	$\hat{z}_{100}$	$\hat{z}_{1000}$
Standard 95% CI	(327.5, 1468.8)mm	(–253.3, 3401.4)mm
Profiled 95% CI	(602.5, 2765)mm	(800, 103000)mm

Comments

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

Profile Likelihood

	$\hat{z}_{100}$	$\hat{z}_{1000}$
Standard 95% CI	(327.5, 1468.8)mm	(–253.3, 3401.4)mm
Profiled 95% CI	(602.5, 2765)mm	(800, 103000)mm

Comments

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

Profile Likelihood

	$\hat{z}_{100}$	$\hat{z}_{1000}$
Standard 95% CI	(327.5, 1468.8)mm	(–253.3, 3401.4)mm
Profiled 95% CI	(602.5, 2765)mm	(800, 103000)mm

Comments

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

Profile Likelihood

	$\hat{z}_{100}$	$\hat{z}_{1000}$
Standard 95% CI	(327.5, 1468.8)mm	(–253.3, 3401.4)mm
Profiled 95% CI	(602.5, 2765)mm	(800, 103000)mm

Comments

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

Profile Likelihood

	$\hat{z}_{100}$	$\hat{z}_{1000}$
Standard 95% CI	(327.5, 1468.8)mm	(–253.3, 3401.4)mm
Profiled 95% CI	(602.5, 2765)mm	(800, 103000)mm

Comments

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

2.5 Models for minima

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

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

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

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

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

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

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

2.5 Models for minima

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

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

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

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

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

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

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

2.5 Models for minima

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

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

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

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

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

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

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

2.5 Models for minima

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

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

2.5 Models for minima

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

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

2.6 The GEV: Words of warning

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

2.6 The GEV: Words of warning

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

2.6 The GEV: Words of warning

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

2.6 The GEV: Words of warning

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

2.6 The GEV: Words of warning

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

2.6 The GEV: Words of warning

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

2.6 The GEV: Words of warning

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

2.6 The GEV: Words of warning

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

2.6 The GEV: Words of warning

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

2.6 The GEV: Words of warning

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

2.6 The GEV: Words of warning

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

2.6 The GEV: Words of warning

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

2.6 The GEV: Words of warning

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

2.6 The GEV: Words of warning

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

Announcements

- I will mark your work for problems sheet 1 this weekend and return the scripts to you in Monday's lecture
- Problems sheet 2 – quite alot of work: deadline?
  - 4pm Friday 15th March (end of term)?
  - 4pm Monday 15th April (first day back)?
  - Up to you, but you have the mid–semester test to revise for over the Easter, and you'll have to wait longer for feedback!

Announcements

- I will mark your work for problems sheet 1 this weekend and return the scripts to you in Monday's lecture
- Problems sheet 2 – quite alot of work: deadline?
  - 4pm Friday 15th March (end of term)?
  - 4pm Monday 15th April (first day back)?
  - Up to you, but you have the mid–semester test to revise for over the Easter, and you'll have to wait longer for feedback!

Announcements

- I will mark your work for problems sheet 1 this weekend and return the scripts to you in Monday's lecture
- Problems sheet 2 – quite alot of work: deadline?
  - 4pm Friday 15th March (end of term)?
  - 4pm Monday 15th April (first day back)?
  - Up to you, but you have the mid–semester test to revise for over the Easter, and you'll have to wait longer for feedback!

Announcements

- I will mark your work for problems sheet 1 this weekend and return the scripts to you in Monday's lecture
- Problems sheet 2 – quite alot of work: deadline?
  - 4pm Friday 15th March (end of term)?
  - 4pm Monday 15th April (first day back)?
  - Up to you, but you have the mid–semester test to revise for over the Easter, and you'll have to wait longer for feedback!

Announcements

- Mid–semester test: 4pm, Thursday 18th April
  - By the end of the second week of the Easter holidays, I will post a specimen test to the course webpage
  - I will also upload lots of specimen exam questions, with solutions
- Project: to be done *after* the Easter holidays (no more homeworks after problems sheet 2) – however, I will try to get details of this on the course webpage over the Easter holidays too

  Revised deadline: **4pm, Friday 10th May**

Announcements

- Mid–semester test: 4pm, Thursday 18th April
  - By the end of the second week of the Easter holidays, I will post a specimen test to the course webpage
  - I will also upload lots of specimen exam questions, with solutions
- Project: to be done *after* the Easter holidays (no more homeworks after problems sheet 2) – however, I will try to get details of this on the course webpage over the Easter holidays too

  Revised deadline: **4pm, Friday 10th May**

Announcements

- Mid–semester test: 4pm, Thursday 18th April
  - By the end of the second week of the Easter holidays, I will post a specimen test to the course webpage
  - I will also upload lots of specimen exam questions, with solutions

- Project: to be done *after* the Easter holidays (no more homeworks after problems sheet 2) – however, I will try to get details of this on the course webpage over the Easter holidays too

  Revised deadline: **4pm, Friday 10th May**

Announcements

- Mid–semester test: 4pm, Thursday 18th April
  - By the end of the second week of the Easter holidays, I will post a specimen test to the course webpage
  - I will also upload lots of specimen exam questions, with solutions
- Project: to be done *after* the Easter holidays (no more homeworks after problems sheet 2) – however, I will try to get details of this on the course webpage over the Easter holidays too

  Revised deadline: **4pm, Friday 10th May**

Announcements

- Mid–semester test: 4pm, Thursday 18th April
  - By the end of the second week of the Easter holidays, I will post a specimen test to the course webpage
  - I will also upload lots of specimen exam questions, with solutions
- Project: to be done *after* the Easter holidays (no more homeworks after problems sheet 2) – however, I will try to get details of this on the course webpage over the Easter holidays too

  Revised deadline: **4pm, Friday 10th May**

Chapter 3

Threshold methods

3.1 Background and theoretical motivation

Threshold methods use a more natural way of determining whether an observation is extreme – *all* observations greater than some high value (**threshold**) are considered.

This allows more efficient use of data and avoids the problems that can arise as a result of blocking...

...but brings its own problems (see later).

We must first go back and consider the asymptotic theory appropriate for this new situation.

3.1 Background and theoretical motivation

Threshold methods use a more natural way of determining whether an observation is extreme – *all* observations greater than some high value (**threshold**) are considered.

This allows more efficient use of data and avoids the problems that can arise as a result of blocking...

...but brings its own problems (see later).

We must first go back and consider the asymptotic theory appropriate for this new situation.

3.1 Background and theoretical motivation

Threshold methods use a more natural way of determining whether an observation is extreme – *all* observations greater than some high value (**threshold**) are considered.

This allows more efficient use of data and avoids the problems that can arise as a result of blocking...

...but brings its own problems (see later).

We must first go back and consider the asymptotic theory appropriate for this new situation.

3.1 Background and theoretical motivation

Threshold methods use a more natural way of determining whether an observation is extreme – *all* observations greater than some high value (**threshold**) are considered.

This allows more efficient use of data and avoids the problems that can arise as a result of blocking...

...but brings its own problems (see later).

We must first go back and consider the asymptotic theory appropriate for this new situation.

3.1 Background and theoretical motivation

Suppose once more that $X_1, X_2, \ldots, X_n$ is a sequence of IID random variables having marginal distribution F, and – once again – let

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

We know from Chapter 2 that

$$\Pr\{M_n \leq x\} \approx G(x),$$

where

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

is the **Generalised Extreme Value** distribution with location, scale and shape μ, σ and ξ respectively.

3.1 Background and theoretical motivation

Suppose once more that $X_1, X_2, \ldots, X_n$ is a sequence of IID random variables having marginal distribution F, and – once again – let

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

We know from Chapter 2 that

$$\Pr\{M_n \leq x\} \approx G(x),$$

where

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

is the **Generalised Extreme Value** distribution with location, scale and shape μ, σ and ξ respectively.

3.1 Background and theoretical motivation

Suppose once more that $X_1, X_2, \ldots, X_n$ is a sequence of IID random variables having marginal distribution F, and – once again – let

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

We know from Chapter 2 that

$$\Pr\{M_n \leq x\} \approx G(x),$$

where

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

is the **Generalised Extreme Value** distribution with location, scale and shape μ, σ and ξ respectively.

Theorem *(Distribution of threshold excess)*

For a large enough threshold u, the distribution function of $(X - u)$, conditional on $X > u$, is approximately

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)_+^{-1/\xi}, \tag{3.1}$$

defined on $y > 0$, where

$$\tilde{\sigma} = \sigma + \xi(u - \mu). \tag{3.2}$$

Theorem *(Distribution of threshold excess)*

For a large enough threshold u, the distribution function of $(X - u)$, conditional on $X > u$, is approximately

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)_+^{-1/\xi}, \tag{3.1}$$

defined on $y > 0$, where

$$\tilde{\sigma} = \sigma + \xi(u - \mu). \tag{3.2}$$

Outline proof (1/5)

Denote an arbitrary term in the X_i sequence by X. Then it follows that a description of the behaviour of extreme events is given by

$$\begin{aligned}
\Pr\{X > u + y | X > u\} &= \frac{\Pr(X > u + y, X > u)}{\Pr(X > u)} \\
&= \frac{\Pr(X > u + y)}{\Pr(X > u)} \\
&= \frac{1 - F(u + y)}{1 - F(u)}. \qquad (\dagger)
\end{aligned}$$

Outline proof (1/5)

Denote an arbitrary term in the X_i sequence by X. Then it follows that a description of the behaviour of extreme events is given by

$$\begin{aligned}
\Pr\{X > u + y | X > u\} &= \frac{\Pr(X > u + y, X > u)}{\Pr(X > u)} \\
&= \frac{\Pr(X > u + y)}{\Pr(X > u)} \\
&= \frac{1 - F(u + y)}{1 - F(u)}. \qquad (\dagger)
\end{aligned}$$

Outline proof (1/5)

Denote an arbitrary term in the X_i sequence by X. Then it follows that a description of the behaviour of extreme events is given by

$$\begin{aligned}
\Pr\{X > u + y | X > u\} &= \frac{\Pr(X > u + y, X > u)}{\Pr(X > u)} \\
&= \frac{\Pr(X > u + y)}{\Pr(X > u)} \\
&= \frac{1 - F(u + y)}{1 - F(u)}. \qquad (\dagger)
\end{aligned}$$

Outline proof (1/5)

Denote an arbitrary term in the X_i sequence by X. Then it follows that a description of the behaviour of extreme events is given by

$$\begin{aligned}
\Pr\{X > u + y | X > u\} &= \frac{\Pr(X > u + y, X > u)}{\Pr(X > u)} \\
&= \frac{\Pr(X > u + y)}{\Pr(X > u)} \\
&= \frac{1 - F(u + y)}{1 - F(u)}. \qquad (\dagger)
\end{aligned}$$

Outline proof (1/5)

Denote an arbitrary term in the X_i sequence by X. Then it follows that a description of the behaviour of extreme events is given by

$$\begin{aligned}
\Pr\{X > u + y | X > u\} &= \frac{\Pr(X > u + y, X > u)}{\Pr(X > u)} \\
&= \frac{\Pr(X > u + y)}{\Pr(X > u)} \\
&= \frac{1 - F(u + y)}{1 - F(u)}. \qquad (\dagger)
\end{aligned}$$

Outline proof (2/5)

From Section 2.1, we know that

$$\Pr\{M_n \leq x\} = F^n(x) \approx G(x) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\},$$

for some parameters μ, σ and ξ. Hence

$$n\log F(x) \approx -\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi} \qquad (\ddagger)$$

Outline proof (2/5)

From Section 2.1, we know that

$$\Pr\{M_n \leq x\} = F^n(x) \approx G(x) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\},$$

for some parameters μ, σ and ξ. Hence

$$n\log F(x) \approx -\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi} \qquad (\ddagger)$$

Outline proof (2/5)

From Section 2.1, we know that

$$\Pr\{M_n \leq x\} = F^n(x) \approx G(x) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\},$$

for some parameters μ, σ and ξ. Hence

$$n\log F(x) \approx -\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi} \qquad (\ddagger)$$

Outline proof (3/5)

For large values of x, a Taylor series expansion implies that

$$\log F(x) \approx -\{1 - F(x)\}.$$

Substitution into (‡) gives

$$\begin{aligned} n(-\{1 - F(u)\}) &\approx -\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi} \\ 1 - F(u) &\approx \frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi} \end{aligned}$$

for large u. Similarly, for $y > 0$,

$$1 - F(u+y) \approx \frac{1}{n}\left[1 + \xi\left(\frac{u+y-\mu}{\sigma}\right)\right]_+^{-1/\xi}.$$

Outline proof (3/5)

For large values of x, a Taylor series expansion implies that

$$\log F(x) \approx -\{1 - F(x)\}.$$

Substitution into (‡) gives

$$\begin{aligned} n(-\{1 - F(u)\}) &\approx -\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi} \\ 1 - F(u) &\approx \frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi} \end{aligned}$$

for large u. Similarly, for $y > 0$,

$$1 - F(u+y) \approx \frac{1}{n}\left[1 + \xi\left(\frac{u+y-\mu}{\sigma}\right)\right]_+^{-1/\xi}.$$

Outline proof (3/5)

For large values of x, a Taylor series expansion implies that

$$\log F(x) \approx -\{1 - F(x)\}.$$

Substitution into (‡) gives

$$\begin{aligned} n(-\{1 - F(u)\}) &\approx -\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi} \\ 1 - F(u) &\approx \frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi} \end{aligned}$$

for large u. Similarly, for $y > 0$,

$$1 - F(u+y) \approx \frac{1}{n}\left[1 + \xi\left(\frac{u+y-\mu}{\sigma}\right)\right]_+^{-1/\xi}.$$

Outline proof (3/5)

For large values of x, a Taylor series expansion implies that

$$\log F(x) \approx -\{1 - F(x)\}.$$

Substitution into (‡) gives

$$n(-\{1 - F(u)\}) \approx -\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi}$$

$$1 - F(u) \approx \frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi}$$

for large u. Similarly, for $y > 0$,

$$1 - F(u+y) \approx \frac{1}{n}\left[1 + \xi\left(\frac{u+y-\mu}{\sigma}\right)\right]_+^{-1/\xi}.$$

Outline proof (3/5)

For large values of x, a Taylor series expansion implies that

$$\log F(x) \approx -\{1 - F(x)\}.$$

Substitution into (‡) gives

$$\begin{aligned} n(-\{1 - F(u)\}) &\approx -\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi} \\ 1 - F(u) &\approx \frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi} \end{aligned}$$

for large u. Similarly, for $y > 0$,

$$1 - F(u+y) \approx \frac{1}{n}\left[1 + \xi\left(\frac{u+y-\mu}{\sigma}\right)\right]_+^{-1/\xi}.$$

Outline proof (3/5)

For large values of x, a Taylor series expansion implies that

$$\log F(x) \approx -\{1 - F(x)\}.$$

Substitution into (‡) gives

$$\begin{aligned} n(-\{1 - F(u)\}) &\approx -\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi} \\ 1 - F(u) &\approx \frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi} \end{aligned}$$

for large u. Similarly, for $y > 0$,

$$1 - F(u+y) \approx \frac{1}{n}\left[1 + \xi\left(\frac{u+y-\mu}{\sigma}\right)\right]_+^{-1/\xi}.$$

Outline proof (3/5)

For large values of x, a Taylor series expansion implies that

$$\log F(x) \approx -\{1 - F(x)\}.$$

Substitution into (‡) gives

$$\begin{aligned} n(-\{1 - F(u)\}) &\approx -\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi} \\ 1 - F(u) &\approx \frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi} \end{aligned}$$

for large u. Similarly, for $y > 0$,

$$1 - F(u+y) \approx \frac{1}{n}\left[1 + \xi\left(\frac{u+y-\mu}{\sigma}\right)\right]_+^{-1/\xi}.$$

Outline proof (4/5)

Substitution into (†) then gives

$$
\begin{aligned}
\Pr\{X > u + y | X > u\} &= \frac{1 - F(u+y)}{1 - F(u)} \\
&\approx \frac{\frac{1}{n}\left[1 + \xi\left(\frac{u+y-\mu}{\sigma}\right)\right]_+^{-1/\xi}}{\frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi}} \\
&= \left[\frac{1 + \xi\left(\frac{u-\mu}{\sigma}\right) + \xi\frac{y}{\sigma}}{1 + \xi\left(\frac{u-\mu}{\sigma}\right)}\right]_+^{-1/\xi} \\
&= \left[\frac{1 + \xi\left(\frac{u-\mu}{\sigma}\right)}{1 + \xi\left(\frac{u-\mu}{\sigma}\right)} + \frac{\xi\frac{y}{\sigma}}{1 + \xi\left(\frac{u-\mu}{\sigma}\right)}\right]_+^{-1/\xi} \\
&= \left[1 + \frac{\xi y}{\sigma + \xi(u-\mu)}\right]_+^{-1/\xi}
\end{aligned}
$$

Outline proof (4/5)

Substitution into (†) then gives

$$\Pr\{X > u + y | X > u\} = \frac{1 - F(u+y)}{1 - F(u)}$$

$$\approx \frac{\frac{1}{n}\left[1 + \xi\left(\frac{u+y-\mu}{\sigma}\right)\right]_+^{-1/\xi}}{\frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi}}$$

$$= \left[\frac{1 + \xi\left(\frac{u-\mu}{\sigma}\right) + \xi\frac{y}{\sigma}}{1 + \xi\left(\frac{u-\mu}{\sigma}\right)}\right]_+^{-1/\xi}$$

$$= \left[\frac{1 + \xi\left(\frac{u-\mu}{\sigma}\right)}{1 + \xi\left(\frac{u-\mu}{\sigma}\right)} + \frac{\xi\frac{y}{\sigma}}{1 + \xi\left(\frac{u-\mu}{\sigma}\right)}\right]_+^{-1/\xi}$$

$$= \left[1 + \frac{\xi y}{\sigma + \xi(u-\mu)}\right]_+^{-1/\xi}$$

Outline proof (4/5)

Substitution into (†) then gives

$$\begin{aligned}
\Pr\{X > u + y | X > u\} &= \frac{1 - F(u+y)}{1 - F(u)} \\
&\approx \frac{\frac{1}{n}\left[1 + \xi\left(\frac{u+y-\mu}{\sigma}\right)\right]_+^{-1/\xi}}{\frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi}} \\
&= \left[\frac{1 + \xi\left(\frac{u-\mu}{\sigma}\right) + \xi\frac{y}{\sigma}}{1 + \xi\left(\frac{u-\mu}{\sigma}\right)}\right]_+^{-1/\xi} \\
&= \left[\frac{1 + \xi\left(\frac{u-\mu}{\sigma}\right)}{1 + \xi\left(\frac{u-\mu}{\sigma}\right)} + \frac{\xi\frac{y}{\sigma}}{1 + \xi\left(\frac{u-\mu}{\sigma}\right)}\right]_+^{-1/\xi} \\
&= \left[1 + \frac{\xi y}{\sigma + \xi(u-\mu)}\right]_+^{-1/\xi}
\end{aligned}$$

Outline proof (4/5)

Substitution into (†) then gives

$$
\begin{aligned}
\Pr\{X > u + y | X > u\} &= \frac{1 - F(u+y)}{1 - F(u)} \\
&\approx \frac{\frac{1}{n}\left[1 + \xi\left(\frac{u+y-\mu}{\sigma}\right)\right]_+^{-1/\xi}}{\frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi}} \\
&= \left[\frac{1 + \xi\left(\frac{u-\mu}{\sigma}\right) + \xi\frac{y}{\sigma}}{1 + \xi\left(\frac{u-\mu}{\sigma}\right)}\right]_+^{-1/\xi} \\
&= \left[\frac{1 + \xi\left(\frac{u-\mu}{\sigma}\right)}{1 + \xi\left(\frac{u-\mu}{\sigma}\right)} + \frac{\xi\frac{y}{\sigma}}{1 + \xi\left(\frac{u-\mu}{\sigma}\right)}\right]_+^{-1/\xi} \\
&= \left[1 + \frac{\xi y}{\sigma + \xi(u-\mu)}\right]_+^{-1/\xi}
\end{aligned}
$$

Outline proof (4/5)

Substitution into (†) then gives

$$
\begin{aligned}
\Pr\{X > u + y | X > u\} &= \frac{1 - F(u+y)}{1 - F(u)} \\
&\approx \frac{\frac{1}{n}\left[1 + \xi\left(\frac{u+y-\mu}{\sigma}\right)\right]_+^{-1/\xi}}{\frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi}} \\
&= \left[\frac{1 + \xi\left(\frac{u-\mu}{\sigma}\right) + \xi\frac{y}{\sigma}}{1 + \xi\left(\frac{u-\mu}{\sigma}\right)}\right]_+^{-1/\xi} \\
&= \left[\frac{1 + \xi\left(\frac{u-\mu}{\sigma}\right)}{1 + \xi\left(\frac{u-\mu}{\sigma}\right)} + \frac{\xi\frac{y}{\sigma}}{1 + \xi\left(\frac{u-\mu}{\sigma}\right)}\right]_+^{-1/\xi} \\
&= \left[1 + \frac{\xi y}{\sigma + \xi(u-\mu)}\right]_+^{-1/\xi}
\end{aligned}
$$

Outline proof (4/5)

Substitution into (†) then gives

$$
\begin{aligned}
\Pr\{X > u + y | X > u\} &= \frac{1 - F(u+y)}{1 - F(u)} \\
&\approx \frac{\frac{1}{n}\left[1 + \xi\left(\frac{u+y-\mu}{\sigma}\right)\right]_+^{-1/\xi}}{\frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi}} \\
&= \left[\frac{1 + \xi\left(\frac{u-\mu}{\sigma}\right) + \xi\frac{y}{\sigma}}{1 + \xi\left(\frac{u-\mu}{\sigma}\right)}\right]_+^{-1/\xi} \\
&= \left[\frac{1 + \xi\left(\frac{u-\mu}{\sigma}\right)}{1 + \xi\left(\frac{u-\mu}{\sigma}\right)} + \frac{\xi\frac{y}{\sigma}}{1 + \xi\left(\frac{u-\mu}{\sigma}\right)}\right]_+^{-1/\xi} \\
&= \left[1 + \frac{\xi y}{\sigma + \xi(u-\mu)}\right]_+^{-1/\xi}
\end{aligned}
$$

Outline proof (5/5)

This gives

$$\Pr\{X > u + y | X > u\} = \left[1 + \frac{\xi y}{\tilde{\sigma}}\right]_{+}^{-1/\xi},$$

where $\tilde{\sigma} = \sigma + \xi(u - \mu)$, as required (compare with Equation 3.1 on page 45).

The family of distributions defined by Equation (3.1) is known as the **Generalised Pareto family**; the distribution itself is often referred to as the **Generalised Pareto Distribution**, or GPD for short.

Outline proof (5/5)

This gives

$$\Pr\{X > u + y | X > u\} = \left[1 + \frac{\xi y}{\tilde{\sigma}}\right]_{+}^{-1/\xi},$$

where $\tilde{\sigma} = \sigma + \xi(u - \mu)$, as required (compare with Equation 3.1 on page 45).

The family of distributions defined by Equation (3.1) is known as the **Generalised Pareto family**; the distribution itself is often referred to as the **Generalised Pareto Distribution**, or GPD for short.

Outline proof (5/5)

This gives

$$\Pr\{X > u + y | X > u\} = \left[1 + \frac{\xi y}{\tilde{\sigma}}\right]_{+}^{-1/\xi},$$

where $\tilde{\sigma} = \sigma + \xi(u - \mu)$, as required (compare with Equation 3.1 on page 45).

The family of distributions defined by Equation (3.1) is known as the **Generalised Pareto family**; the distribution itself is often referred to as the **Generalised Pareto Distribution**, or GPD for short.

3.1 Background and theoretical motivation

- If block maxima have approximate distribution G, then threshold excesses have a corresponding distribution given by the Generalised Pareto family.
- The parameters of the GPD are uniquely determined by those of the GEV:
  - The parameter ξ in Equation (3.1) is equal to that of the corresponding GEV
  - The GPD scale parameter is a function of the GEV location and shape parameters

3.1 Background and theoretical motivation

- If block maxima have approximate distribution G, then threshold excesses have a corresponding distribution given by the Generalised Pareto family.
- The parameters of the GPD are uniquely determined by those of the GEV:
  - The parameter ξ in Equation (3.1) is equal to that of the corresponding GEV
  - The GPD scale parameter is a function of the GEV location and shape parameters

3.1 Background and theoretical motivation

- If block maxima have approximate distribution G, then threshold excesses have a corresponding distribution given by the Generalised Pareto family.
- The parameters of the GPD are uniquely determined by those of the GEV:
  - The parameter ξ in Equation (3.1) is equal to that of the corresponding GEV
  - The GPD scale parameter is a function of the GEV location and shape parameters

3.1 Background and theoretical motivation

- If block maxima have approximate distribution G, then threshold excesses have a corresponding distribution given by the Generalised Pareto family.
- The parameters of the GPD are uniquely determined by those of the GEV:
  - The parameter ξ in Equation (3.1) is equal to that of the corresponding GEV
  - The GPD scale parameter is a function of the GEV location and shape parameters

3.1 Background and theoretical motivation

- Estimates of the GEV parameters are sensitive to the size of block chosen to identify extremes; estimates of the GPD parameters are ‘stable’.

- The duality between the GEV and GPD means that the shape parameter ξ is dominant in determining the qualitative behaviour of the GPD:
  - If $\xi < 0$ the distribution of excesses has an upper bound
  - if $\xi > 0$ the distribution has no upper limit
  - the case $\xi = 0$ is also unbounded, and is taken as the limit $\xi \to 0$, giving

$$H(y) = 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right), \qquad y > 0;$$

  i.e. an exponential distribution with rate $1/\tilde{\sigma}$.

3.1 Background and theoretical motivation

- Estimates of the GEV parameters are sensitive to the size of block chosen to identify extremes; estimates of the GPD parameters are 'stable'.
- The duality between the GEV and GPD means that the shape parameter ξ is dominant in determining the qualitative behaviour of the GPD:
  - If $\xi < 0$ the distribution of excesses has an upper bound
  - if $\xi > 0$ the distribution has no upper limit
  - the case $\xi = 0$ is also unbounded, and is taken as the limit $\xi \to 0$, giving

$$H(y) = 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right), \qquad y > 0;$$

  i.e. an exponential distribution with rate $1/\tilde{\sigma}$.

3.1 Background and theoretical motivation

- Estimates of the GEV parameters are sensitive to the size of block chosen to identify extremes; estimates of the GPD parameters are ‘stable’.
- The duality between the GEV and GPD means that the shape parameter ξ is dominant in determining the qualitative behaviour of the GPD:
  - If $\xi < 0$ the distribution of excesses has an upper bound
  - if $\xi > 0$ the distribution has no upper limit
  - the case $\xi = 0$ is also unbounded, and is taken as the limit $\xi \to 0$, giving

$$H(y) = 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right), \qquad y > 0;$$

i.e. an exponential distribution with rate $1/\tilde{\sigma}$.

3.1 Background and theoretical motivation

- Estimates of the GEV parameters are sensitive to the size of block chosen to identify extremes; estimates of the GPD parameters are 'stable'.
- The duality between the GEV and GPD means that the shape parameter ξ is dominant in determining the qualitative behaviour of the GPD:
  - If $\xi < 0$ the distribution of excesses has an upper bound
  - if $\xi > 0$ the distribution has no upper limit
  - the case $\xi = 0$ is also unbounded, and is taken as the limit $\xi \to 0$, giving

$$H(y) = 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right), \qquad y > 0;$$

i.e. an exponential distribution with rate $1/\tilde{\sigma}$.

3.1 Background and theoretical motivation

- Estimates of the GEV parameters are sensitive to the size of block chosen to identify extremes; estimates of the GPD parameters are 'stable'.
- The duality between the GEV and GPD means that the shape parameter ξ is dominant in determining the qualitative behaviour of the GPD:
  - If $\xi < 0$ the distribution of excesses has an upper bound
  - if $\xi > 0$ the distribution has no upper limit
  - the case $\xi = 0$ is also unbounded, and is taken as the limit $\xi \to 0$, giving

$$H(y) = 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right), \qquad y > 0;$$

i.e. an exponential distribution with rate $1/\tilde{\sigma}$.

3.1 Background and theoretical motivation

- Estimates of the GEV parameters are sensitive to the size of block chosen to identify extremes; estimates of the GPD parameters are 'stable'.
- The duality between the GEV and GPD means that the shape parameter ξ is dominant in determining the qualitative behaviour of the GPD:
  - If $\xi < 0$ the distribution of excesses has an upper bound
  - if $\xi > 0$ the distribution has no upper limit
  - the case $\xi = 0$ is also unbounded, and is taken as the limit $\xi \to 0$, giving

$$H(y) = 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right), \qquad y > 0;$$

  i.e. an exponential distribution with rate $1/\tilde{\sigma}$.

3.1 Background and theoretical motivation

- Until this point, we have used the notation $\tilde{\sigma}$ to denote the scale parameter of the GPD, so as to distinguish it from the corresponding parameter of the GEV...
- ... For notational convenience we now drop this distinction, using σ to denote the scale parameter within either family.

3.1 Background and theoretical motivation

- Until this point, we have used the notation $\tilde{\sigma}$ to denote the scale parameter of the GPD, so as to distinguish it from the corresponding parameter of the GEV...
- ... For notational convenience we now drop this distinction, using σ to denote the scale parameter within either family.

Example 3.1

Suppose $X_1, X_2, \ldots, X_n$ is a sequence of independent $exp(1)$ random variables.

Show that the limiting distribution of threshold excesses belongs to the generalised Pareto family.

Example 3.1: Solution

If $X_i \sim exp(1)$, then $F(x) = 1 - e^{-x}$ for $x > 0$.

By direct calculation,

$$\begin{aligned}\frac{1 - F(u+y)}{1 - F(u)} &= \frac{1 - \left(1 - e^{-(u+y)}\right)}{1 - (1 - e^{-u})} \\ &= \frac{e^{-(u+y)}}{e^{-u}} = e^{-y}, \qquad y > 0.\end{aligned}$$

Thus, the limit distribution is an exponential distribution; i.e. a GPD with $\xi = 0$. Further, we know that when $\xi = 0$, we have

$$1 - H(y) = \exp\left(-\frac{y}{\sigma}\right);$$

hence, we also have $\sigma = 1$. That is, the limit distribution of threshold exceedances is $GPD(1, 0)$.

Example 3.1: Solution

If $X_i \sim exp(1)$, then $F(x) = 1 - e^{-x}$ for $x > 0$.

By direct calculation,

$$\begin{aligned}\frac{1 - F(u+y)}{1 - F(u)} &= \frac{1 - \left(1 - e^{-(u+y)}\right)}{1 - (1 - e^{-u})} \\ &= \frac{e^{-(u+y)}}{e^{-u}} = e^{-y}, \qquad y > 0.\end{aligned}$$

Thus, the limit distribution is an exponential distribution; i.e. a GPD with $\xi = 0$. Further, we know that when $\xi = 0$, we have

$$1 - H(y) = \exp\left(-\frac{y}{\sigma}\right);$$

hence, we also have $\sigma = 1$. That is, the limit distribution of threshold exceedances is $GPD(1, 0)$.

Example 3.1: Solution

If $X_i \sim exp(1)$, then $F(x) = 1 - e^{-x}$ for $x > 0$.

By direct calculation,

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{1 - \left(1 - e^{-(u+y)}\right)}{1 - (1 - e^{-u})}$$

$$= \frac{e^{-(u+y)}}{e^{-u}} = e^{-y}, \qquad y > 0.$$

Thus, the limit distribution is an exponential distribution; i.e. a GPD with $\xi = 0$. Further, we know that when $\xi = 0$, we have

$$1 - H(y) = \exp\left(-\frac{y}{\sigma}\right);$$

hence, we also have $\sigma = 1$. That is, the limit distribution of threshold exceedances is $GPD(1, 0)$.

Example 3.1: Solution

If $X_i \sim exp(1)$, then $F(x) = 1 - e^{-x}$ for $x > 0$.

By direct calculation,

$$\begin{aligned} \frac{1 - F(u+y)}{1 - F(u)} &= \frac{1 - \left(1 - e^{-(u+y)}\right)}{1 - (1 - e^{-u})} \\ &= \frac{e^{-(u+y)}}{e^{-u}} = e^{-y}, \qquad y > 0. \end{aligned}$$

Thus, the limit distribution is an exponential distribution; i.e. a GPD with $\xi = 0$. Further, we know that when $\xi = 0$, we have

$$1 - H(y) = \exp\left(-\frac{y}{\sigma}\right);$$

hence, we also have $\sigma = 1$. That is, the limit distribution of threshold exceedances is $GPD(1, 0)$.

Example 3.1: Solution

If $X_i \sim exp(1)$, then $F(x) = 1 - e^{-x}$ for $x > 0$.

By direct calculation,

$$\begin{aligned} \frac{1 - F(u+y)}{1 - F(u)} &= \frac{1 - \left(1 - e^{-(u+y)}\right)}{1 - (1 - e^{-u})} \\ &= \frac{e^{-(u+y)}}{e^{-u}} = e^{-y}, \qquad y > 0. \end{aligned}$$

Thus, the limit distribution is an exponential distribution; i.e. a GPD with $\xi = 0$. Further, we know that when $\xi = 0$, we have

$$1 - H(y) = \exp\left(-\frac{y}{\sigma}\right);$$

hence, we also have $\sigma = 1$. That is, the limit distribution of threshold exceedances is $GPD(1, 0)$.

Example 3.1: Solution

If $X_i \sim exp(1)$, then $F(x) = 1 - e^{-x}$ for $x > 0$.

By direct calculation,

$$\begin{aligned} \frac{1 - F(u+y)}{1 - F(u)} &= \frac{1 - \left(1 - e^{-(u+y)}\right)}{1 - (1 - e^{-u})} \\ &= \frac{e^{-(u+y)}}{e^{-u}} = e^{-y}, \qquad y > 0. \end{aligned}$$

Thus, the limit distribution is an exponential distribution; i.e. a GPD with $\xi = 0$. Further, we know that when $\xi = 0$, we have

$$1 - H(y) = \exp\left(-\frac{y}{\sigma}\right);$$

hence, we also have $\sigma = 1$. That is, the limit distribution of threshold exceedances is $GPD(1, 0)$.

Example 3.2

Suppose $X_1, X_2, \ldots, X_n$ is a sequence of independent $U(0, 1)$ random variables.

Show that the limiting distribution of threshold excesses belongs to the generalised Pareto family.

Example 3.2: Solution

If $X_i \sim U(0,1)$, then $F(x) = x$ for $0 \leq x \leq 1$.

Hence

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{1 - u + y}{1 - u} = 1 - \frac{y}{1 - u}.$$

For the GPD with $\xi \neq 0$, we have

$$1 - H(y) = \left[1 + \frac{\xi y}{\sigma}\right]^{-1/\xi}.$$

Thus, we have $\xi = -1$ and $\sigma = 1 - u$, i.e. the limiting distribution for threshold exceedances is $GPD(1-u,-1)$.

Example 3.2: Solution

If $X_i \sim U(0,1)$, then $F(x) = x$ for $0 \leq x \leq 1$.

Hence

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{1 - u + y}{1 - u} = 1 - \frac{y}{1 - u}.$$

For the GPD with $\xi \neq 0$, we have

$$1 - H(y) = \left[1 + \frac{\xi y}{\sigma}\right]^{-1/\xi}.$$

Thus, we have $\xi = -1$ and $\sigma = 1 - u$, i.e. the limiting distribution for threshold exceedances is $GPD(1-u,-1)$.

Example 3.2: Solution

If $X_i \sim U(0,1)$, then $F(x) = x$ for $0 \leq x \leq 1$.

Hence

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{1 - u + y}{1 - u} = 1 - \frac{y}{1 - u}.$$

For the GPD with $\xi \neq 0$, we have

$$1 - H(y) = \left[1 + \frac{\xi y}{\sigma}\right]^{-1/\xi}.$$

Thus, we have $\xi = -1$ and $\sigma = 1 - u$, i.e. the limiting distribution for threshold exceedances is $GPD(1-u,-1)$.

Example 3.2: Solution

If $X_i \sim U(0,1)$, then $F(x) = x$ for $0 \leq x \leq 1$.

Hence

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{1 - u + y}{1 - u} = 1 - \frac{y}{1-u}.$$

For the GPD with $\xi \neq 0$, we have

$$1 - H(y) = \left[1 + \frac{\xi y}{\sigma}\right]^{-1/\xi}.$$

Thus, we have $\xi = -1$ and $\sigma = 1 - u$, i.e. the limiting distribution for threshold exceedances is $GPD(1-u,-1)$.

Example 3.2: Solution

If $X_i \sim U(0,1)$, then $F(x) = x$ for $0 \leq x \leq 1$.

Hence

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{1 - u + y}{1 - u} = 1 - \frac{y}{1 - u}.$$

For the GPD with $\xi \neq 0$, we have

$$1 - H(y) = \left[1 + \frac{\xi y}{\sigma}\right]^{-1/\xi}.$$

Thus, we have $\xi = -1$ and $\sigma = 1 - u$, i.e. the limiting distribution for threshold exceedances is $GPD(1-u,-1)$.

Example 3.2: Solution

If $X_i \sim U(0,1)$, then $F(x) = x$ for $0 \leq x \leq 1$.

Hence

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{1 - u + y}{1 - u} = 1 - \frac{y}{1 - u}.$$

For the GPD with $\xi \neq 0$, we have

$$1 - H(y) = \left[1 + \frac{\xi y}{\sigma}\right]^{-1/\xi}.$$

Thus, we have $\xi = -1$ and $\sigma = 1 - u$, i.e. the limiting distribution for threshold exceedances is $GPD(1-u,-1)$.

Example 3.2: Solution

If $X_i \sim U(0,1)$, then $F(x) = x$ for $0 \leq x \leq 1$.

Hence

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{1 - u + y}{1 - u} = 1 - \frac{y}{1 - u}.$$

For the GPD with $\xi \neq 0$, we have

$$1 - H(y) = \left[1 + \frac{\xi y}{\sigma}\right]^{-1/\xi}.$$

Thus, we have $\xi = -1$ and $\sigma = 1 - u$, i.e. the limiting distribution for threshold exceedances is $GPD(1–u,–1)$.

3.2 Application: rainfall extremes in NYC

The file `newyork.txt`, available to download from the course webpage, gives daily rainfall accumulations (in mm) for New York City for the years 1914–1961 (inclusive).

Much of the northeastern United States is relatively low–lying and so prone to flooding; this is made much worse on the odd occasion that a hurricane travels this far north (for example, "**Superstorm Sandy**" in 2012).

Thus, analysing extreme rainfall data has a real practical motivation here, in terms of **river** and **sea flood defence systems**.

In this Section, we will illustrate a complete threshold–based analysis of the rainfall extremes observed at New York.

3.2 Application: rainfall extremes in NYC

The data have been scanned into R from the course webpage and stored in the vector `rain`:

```
rain=scan('newyork.txt')
```

3.2 Application: rainfall extremes in NYC

The data have been scanned into R from the course webpage and stored in the vector `rain`:

```
rain=scan('newyork.txt')
```

3.2.1 Threshold choice

The **threshold stability property** of the GPD means that if the GPD is a valid model for excesses over some threshold u_0, then it is valid for excesses over all thresholds $u > u_0$.

Denoting by σ_{u_0} the GPD scale parameter for excesses over threshold u_0, the expected value of our threshold excesses, conditional on being greater than the threshold, is

$$E[X - u | X > u] = \frac{\sigma_{u_0} + \xi u}{1 - \xi}. \tag{3.3}$$

Thus, for all $u > u_0$, $E[X - u | X > u]$, is a linear function of u.

3.2.1 Threshold choice

The **threshold stability property** of the GPD means that if the GPD is a valid model for excesses over some threshold u_0, then it is valid for excesses over all thresholds $u > u_0$.

Denoting by σ_{u_0} the GPD scale parameter for excesses over threshold u_0, the expected value of our threshold excesses, conditional on being greater than the threshold, is

$$E[X - u|X > u] = \frac{\sigma_{u_0} + \xi u}{1 - \xi}. \tag{3.3}$$

Thus, for all $u > u_0$, $E[X - u|X > u]$, is a linear function of u.

3.2.1 Threshold choice

The **threshold stability property** of the GPD means that if the GPD is a valid model for excesses over some threshold u_0, then it is valid for excesses over all thresholds $u > u_0$.

Denoting by σ_{u_0} the GPD scale parameter for excesses over threshold u_0, the expected value of our threshold excesses, conditional on being greater than the threshold, is

$$E[X - u|X > u] = \frac{\sigma_{u_0} + \xi u}{1 - \xi}. \tag{3.3}$$

Thus, for all $u > u_0$, $E[X - u|X > u]$, is a linear function of u.

3.2.1 Threshold choice

Furthermore, $E[X - u|X > u]$ is simply the mean of the excesses of the threshold u, for which the sample mean of the threshold excesses of u provides an estimate.

This leads to the **mean residual life plot**, a graphical procedure for identifying a suitably high threshold for modelling extremes via the GPD.

In this plot, for a range of candidate values for u we identify the corresponding mean threshold excess; we then plot this mean threshold excess against u, and look for the value u_0 above which we can see linearity in the plot.

3.2.1 Threshold choice

Furthermore, $E[X - u|X > u]$ is simply the mean of the excesses of the threshold u, for which the sample mean of the threshold excesses of u provides an estimate.

This leads to the **mean residual life plot**, a graphical procedure for identifying a suitably high threshold for modelling extremes via the GPD.

In this plot, for a range of candidate values for u we identify the corresponding mean threshold excess; we then plot this mean threshold excess against u, and look for the value u_0 above which we can see linearity in the plot.

3.2.1 Threshold choice

Furthermore, $E[X - u|X > u]$ is simply the mean of the excesses of the threshold u, for which the sample mean of the threshold excesses of u provides an estimate.

This leads to the **mean residual life plot**, a graphical procedure for identifying a suitably high threshold for modelling extremes via the GPD.

In this plot, for a range of candidate values for u we identify the corresponding mean threshold excess; we then plot this mean threshold excess against u, and look for the value u_0 above which we can see linearity in the plot.

3.2.1 Threshold choice

We can easily do this from first principles in R.

First of all, we set up a vector of possible thresholds, starting at zero and going up to the maximum value in our dataset:

```
> u=seq(0,max(rain),0.1)
```

The vector `x` is then set up to take the corresponding values for the mean excess over each value in `u`:

```
> x=vector('numeric', length(u))
```

3.2.1 Threshold choice

We can easily do this from first principles in R.

First of all, we set up a vector of possible thresholds, starting at zero and going up to the maximum value in our dataset:

```
> u=seq(0,max(rain),0.1)
```

The vector `x` is then set up to take the corresponding values for the mean excess over each value in `u`:

```
> x=vector('numeric', length(u))
```

3.2.1 Threshold choice

We can easily do this from first principles in R.

First of all, we set up a vector of possible thresholds, starting at zero and going up to the maximum value in our dataset:

```
> u=seq(0,max(rain),0.1)
```

The vector `x` is then set up to take the corresponding values for the mean excess over each value in `u`:

```
> x=vector('numeric', length(u))
```

3.2.1 Threshold choice

Then the following code computes the mean excess for each value in `u` and stores it in `x`:

```
> for(i in 1:length(x))
    {
      threshold.exceedances=rain[rain>u[i]]
      x[i]=mean(threshold.exceedances-u[i])
    }
```

The MRL plot is then produced using the following code, giving the plot in Figure 3.1:

```
plot(u,x,type='l', main='MRL plot',ylab='mean
excess')
```

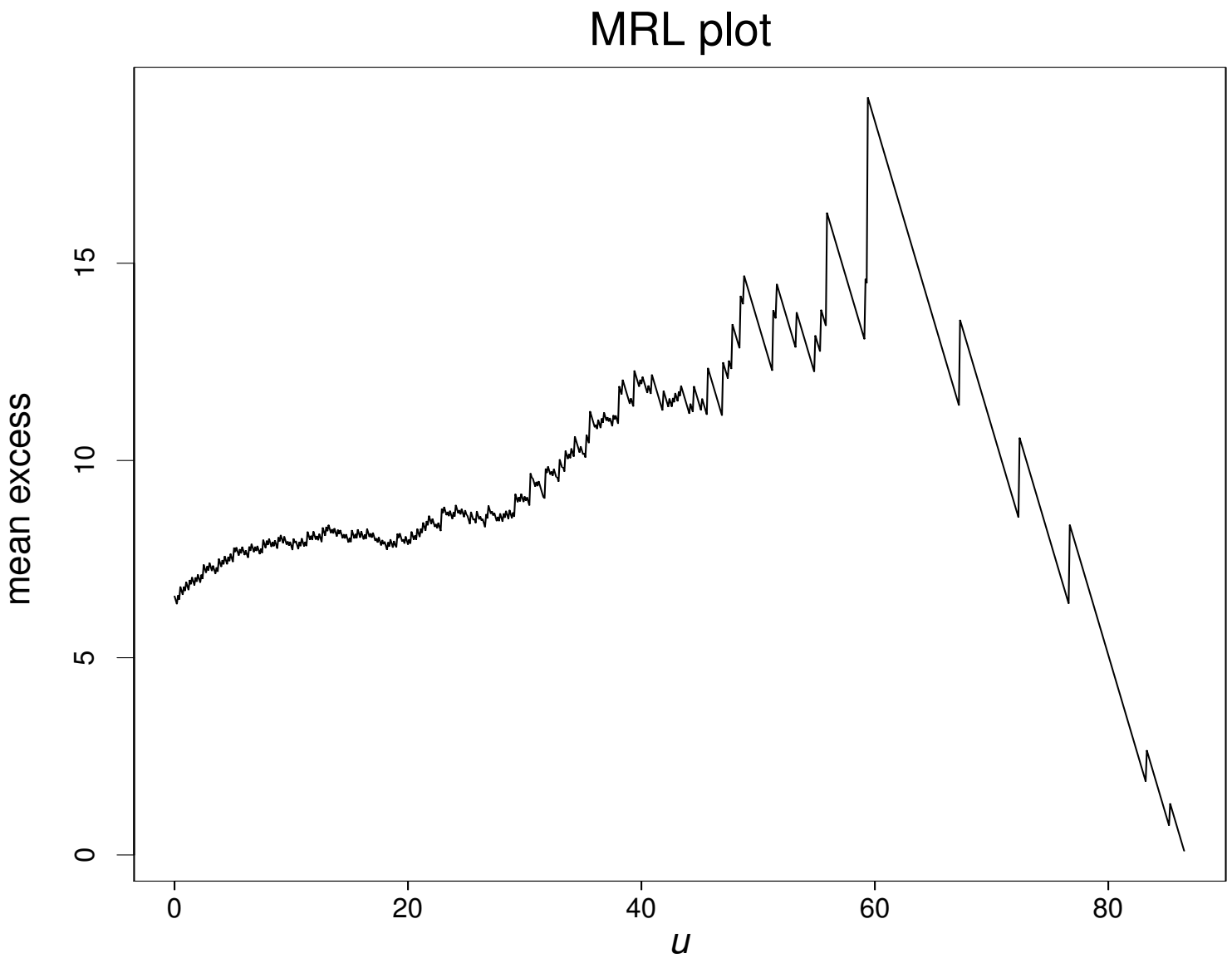
Then the following code computes the mean excess for each value in `u` and stores it in `x`:

```
> for(i in 1:length(x))
    {
      threshold.exceedances=rain[rain>u[i]]
      x[i]=mean(threshold.exceedances-u[i])
    }
```

The MRL plot is then produced using the following code, giving the plot in Figure 3.1:

```
plot(u,x,type='l', main='MRL plot',ylab='mean
excess')
```

3.2.1 Threshold choice



3.2.2 Data pre–processing

In problems sheet 1, we considered how to use R to extract the set of block maxima to be modelled by the generalised extreme value distribution.

Writing R code to do this can be a time–consuming process, and the code needs to written specifically for the data being analysed.

In a threshold–based analysis the data pre–processing is far more straightforward.

Using $u_0 = 30$ as our threshold for identifying extremes (see Figure 3.1), we can easily obtain our set of threshold exceedances for modelling with the generalised Pareto distribution:

```
above.threshold=rain[rain>30]
threshold.exceedances=above.threshold-30
```

3.2.2 Data pre–processing

In problems sheet 1, we considered how to use R to extract the set of block maxima to be modelled by the generalised extreme value distribution.

Writing R code to do this can be a time–consuming process, and the code needs to written specifically for the data being analysed.

In a threshold–based analysis the data pre–processing is far more straightforward.

Using $u_0 = 30$ as our threshold for identifying extremes (see Figure 3.1), we can easily obtain our set of threshold exceedances for modelling with the generalised Pareto distribution:

```
above.threshold=rain[rain>30]
threshold.exceedances=above.threshold-30
```

3.2.2 Data pre–processing

In problems sheet 1, we considered how to use R to extract the set of block maxima to be modelled by the generalised extreme value distribution.

Writing R code to do this can be a time–consuming process, and the code needs to written specifically for the data being analysed.

In a threshold–based analysis the data pre–processing is far more straightforward.

Using $u_0 = 30$ as our threshold for identifying extremes (see Figure 3.1), we can easily obtain our set of threshold exceedances for modelling with the generalised Pareto distribution:

```
above.threshold=rain[rain>30]
threshold.exceedances=above.threshold-30
```

3.2.2 Data pre–processing

In problems sheet 1, we considered how to use R to extract the set of block maxima to be modelled by the generalised extreme value distribution.

Writing R code to do this can be a time–consuming process, and the code needs to written specifically for the data being analysed.

In a threshold–based analysis the data pre–processing is far more straightforward.

Using $u_0 = 30$ as our threshold for identifying extremes (see Figure 3.1), we can easily obtain our set of threshold exceedances for modelling with the generalised Pareto distribution:

```
above.threshold=rain[rain>30]
threshold.exceedances=above.threshold-30
```

3.2.3 Fitting the GPD

The most commonly–used approach to fit the GPD to the set of threshold excesses is that of maximum likelihood.

The GPD log–likelihood function can be derived in the usual way; this is left as an exercise, but can be shown to be:

$$\ell(\sigma, \xi; \boldsymbol{y}) = -152 \log \sigma - (1 + 1/\xi) \sum_{i=1}^{152} \log \left(1 + \frac{\xi y_i}{\sigma}\right)_+, \quad (3.4)$$

where $\boldsymbol{y} = (y_1, \ldots, y_{152})^T$ are the set of exceedances above threshold $u_0 = 30$.

For the case $\xi = 0$, interpreted as $\xi \to 0$, we have the log–likelihood f or an exponential distribution with rate $1/\sigma$ (again, see problems sheet 2).

3.2.3 Fitting the GPD

The most commonly–used approach to fit the GPD to the set of threshold excesses is that of maximum likelihood.

The GPD log–likelihood function can be derived in the usual way; this is left as an exercise, but can be shown to be:

$$\ell(\sigma, \xi; \boldsymbol{y}) = -152 \log \sigma - (1 + 1/\xi) \sum_{i=1}^{152} \log \left(1 + \frac{\xi y_i}{\sigma}\right)_+, \quad (3.4)$$

where $\boldsymbol{y} = (y_1, \ldots, y_{152})^T$ are the set of exceedances above threshold $u_0 = 30$.

For the case $\xi = 0$, interpreted as $\xi \to 0$, we have the log–likelihood f or an exponential distribution with rate $1/\sigma$ (again, see problems sheet 2).

3.2.3 Fitting the GPD

R demo, including assessment of model adequacy

3.2.5 Return level estimation

Figure 3.2 seems to indicate that the GPD is suitable for our set of threshold exceedances $y_1, \ldots, y_{152}$; that is,

$$\Pr(X > u + y | X > u) \approx \left[1 + \frac{\hat{\xi} y}{\hat{\sigma}}\right]_+^{-1/\hat{\xi}}, \tag{3.5}$$

for $\xi \neq 0$. Working with the LHS of (3.5), we see that

$$\Pr(X > u + y | X > u) = \frac{\Pr(X > u + y)}{\Pr(X > u)},$$

giving

$$\Pr(X > u + y) = \Pr(X > u)\Pr(X > u + y | X > u).$$

3.2.5 Return level estimation

Figure 3.2 seems to indicate that the GPD is suitable for our set of threshold exceedances $y_1, \ldots, y_{152}$; that is,

$$\Pr(X > u + y | X > u) \approx \left[1 + \frac{\hat{\xi} y}{\hat{\sigma}}\right]_+^{-1/\hat{\xi}}, \tag{3.5}$$

for $\xi \neq 0$. Working with the LHS of (3.5), we see that

$$\Pr(X > u + y | X > u) = \frac{\Pr(X > u + y)}{\Pr(X > u)},$$

giving

$$\Pr(X > u + y) = \Pr(X > u)\Pr(X > u + y | X > u).$$

3.2.5 Return level estimation

Figure 3.2 seems to indicate that the GPD is suitable for our set of threshold exceedances $y_1, \ldots, y_{152}$; that is,

$$\Pr(X > u + y | X > u) \approx \left[1 + \frac{\hat{\xi} y}{\hat{\sigma}}\right]_+^{-1/\hat{\xi}}, \tag{3.5}$$

for $\xi \neq 0$. Working with the LHS of (3.5), we see that

$$\Pr(X > u + y | X > u) = \frac{\Pr(X > u + y)}{\Pr(X > u)},$$

giving

$$\Pr(X > u + y) = \Pr(X > u)\Pr(X > u + y | X > u).$$

3.2.5 Return level estimation

Figure 3.2 seems to indicate that the GPD is suitable for our set of threshold exceedances $y_1, \ldots, y_{152}$; that is,

$$\Pr(X > u + y | X > u) \approx \left[1 + \frac{\hat{\xi} y}{\hat{\sigma}}\right]_+^{-1/\hat{\xi}}, \tag{3.5}$$

for $\xi \neq 0$. Working with the LHS of (3.5), we see that

$$\Pr(X > u + y | X > u) = \frac{\Pr(X > u + y)}{\Pr(X > u)},$$

giving

$$\Pr(X > u + y) = \Pr(X > u)\Pr(X > u + y | X > u).$$

3.2.5 Return level estimation

Figure 3.2 seems to indicate that the GPD is suitable for our set of threshold exceedances $y_1, \ldots, y_{152}$; that is,

$$\Pr(X > u + y | X > u) \approx \left[1 + \frac{\hat{\xi} y}{\hat{\sigma}}\right]_+^{-1/\hat{\xi}}, \tag{3.5}$$

for $\xi \neq 0$. Working with the LHS of (3.5), we see that

$$\Pr(X > u + y | X > u) = \frac{\Pr(X > u + y)}{\Pr(X > u)},$$

giving

$$\Pr(X > u + y) = \Pr(X > u)\Pr(X > u + y | X > u).$$

3.2.5 Return level estimation

After substitution of (3.5), we get

$$\Pr(X > u + y) \approx \hat{\lambda}_u \left[1 + \frac{\hat{\xi} y}{\hat{\sigma}} \right]_+^{-1/\hat{\xi}}, \tag{3.6}$$

where $\lambda_u = \Pr(X > u)$ and is estimated as the empirical threshold exceedance rate $\hat{\lambda}_u$.

3.2.5 Return level estimation

After substitution of (3.5), we get

$$\Pr(X > u + y) \approx \hat{\lambda}_u \left[1 + \frac{\hat{\xi} y}{\hat{\sigma}} \right]_+^{-1/\hat{\xi}}, \tag{3.6}$$

where $\lambda_u = \Pr(X > u)$ and is estimated as the empirical threshold exceedance rate $\hat{\lambda}_u$.

3.2.5 Return level estimation

After substitution of (3.5), we get

$$\Pr(X > u + y) \approx \hat{\lambda}_u \left[1 + \frac{\hat{\xi} y}{\hat{\sigma}} \right]_+^{-1/\hat{\xi}}, \tag{3.6}$$

where $\lambda_u = \Pr(X > u)$ and is estimated as the empirical threshold exceedance rate $\hat{\lambda}_u$.

3.2.5 Return level estimation

Now each y_i, $i = 1, \ldots, 152$, are the raw rainfall observations (exceeding the threshold) minus the threshold ($x_i - u$), as the GPD models the magnitude of excess over u.

Substitution of $y_i = x_i - u$ into (3.6) gives

$$\Pr(X > x) \approx \hat{\lambda}_u \left[1 + \hat{\xi} \left(\frac{x - u}{\hat{\sigma}} \right) \right]_+^{-1/\hat{\xi}}. \tag{3.7}$$

3.2.5 Return level estimation

Now each y_i, $i = 1, \ldots, 152$, are the raw rainfall observations (exceeding the threshold) minus the threshold ($x_i - u$), as the GPD models the magnitude of excess over u.

Substitution of $y_i = x_i - u$ into (3.6) gives

$$\Pr(X > x) \approx \hat{\lambda}_u \left[1 + \hat{\xi} \left(\frac{x - u}{\hat{\sigma}} \right) \right]_+^{-1/\hat{\xi}}. \tag{3.7}$$

3.2.5 Return level estimation

Now each y_i, $i = 1, \ldots, 152$, are the raw rainfall observations (exceeding the threshold) minus the threshold ($x_i - u$), as the GPD models the magnitude of excess over u.

Substitution of $y_i = x_i - u$ into (3.6) gives

$$\Pr(X > x) \approx \hat{\lambda}_u \left[1 + \hat{\xi} \left(\frac{x - u}{\hat{\sigma}} \right) \right]_+^{-1/\hat{\xi}}. \tag{3.7}$$

3.2.5 Return level estimation

Thus, an estimate of the level z_t that is exceeded on average once every t **observations** is obtained as the solution of

$$\hat{\lambda}_u \left[1 + \hat{\xi} \left(\frac{\hat{z}_t - u}{\hat{\sigma}} \right) \right]_+^{-1/\hat{\xi}} = \frac{1}{t},$$

giving

$$\hat{z}_t = u + \frac{\hat{\sigma}}{\hat{\xi}} \left[(t\hat{\lambda}_u)^{\hat{\xi}} - 1 \right] \tag{3.8}$$

for $\xi \neq 0$, and

$$\hat{z}_t = u + \hat{\sigma} \log(t\hat{\lambda}_u)$$

when $\xi = 0$ (see problems sheet 2).

3.2.5 Return level estimation

Thus, an estimate of the level z_t that is exceeded on average once every t **observations** is obtained as the solution of

$$\hat{\lambda}_u \left[1 + \hat{\xi}\left(\frac{\hat{z}_t - u}{\hat{\sigma}}\right)\right]_+^{-1/\hat{\xi}} = \frac{1}{t},$$

giving

$$\hat{z}_t = u + \frac{\hat{\sigma}}{\hat{\xi}}\left[(t\hat{\lambda}_u)^{\hat{\xi}} - 1\right] \tag{3.8}$$

for $\xi \neq 0$, and

$$\hat{z}_t = u + \hat{\sigma}\log(t\hat{\lambda}_u)$$

when $\xi = 0$ (see problems sheet 2).

3.2.5 Return level estimation

Thus, an estimate of the level z_t that is exceeded on average once every t **observations** is obtained as the solution of

$$\hat{\lambda}_u \left[1 + \hat{\xi}\left(\frac{\hat{z}_t - u}{\hat{\sigma}}\right)\right]_+^{-1/\hat{\xi}} = \frac{1}{t},$$

giving

$$\hat{z}_t = u + \frac{\hat{\sigma}}{\hat{\xi}}\left[(t\hat{\lambda}_u)^{\hat{\xi}} - 1\right] \tag{3.8}$$

for $\xi \neq 0$, and

$$\hat{z}_t = u + \hat{\sigma}\log(t\hat{\lambda}_u)$$

when $\xi = 0$ (see problems sheet 2).

3.2.5 Return level estimation

By construction, z_t is the t–**observation** return level.

However, it is often more convenient to give return levels on an annual scale, so that the r–**year** return level is the level expected to be exceeded once every r **years**.

3.2.5 Return level estimation

By construction, z_t is the t–**observation** return level.

However, it is often more convenient to give return levels on an annual scale, so that the r–**year** return level is the level expected to be exceeded once every r **years**.

3.2.5 Return level estimation

By construction, z_t is the t–**observation** return level.

However, it is often more convenient to give return levels on an annual scale, so that the r–**year** return level is the level expected to be exceeded once every r **years**.

3.2.5 Return level estimation

If there are n_y observations per year, this corresponds to the t–observation return level with $t = r \times n_y$.

Hence, an estimate of the r–year return level z_r is defined by

$$\hat{z}_r = u + \frac{\hat{\sigma}}{\hat{\xi}}\left[(rn_y\hat{\lambda}_u)^{\hat{\xi}} - 1\right],$$

unless $\xi = 0$, in which case

$$\hat{z}_r \quad = \quad u + \hat{\sigma}\log(rn_y\hat{\lambda}_u).$$

3.2.5 Return level estimation

If there are n_y observations per year, this corresponds to the t–observation return level with $t = r \times n_y$.

Hence, an estimate of the r–year return level z_r is defined by

$$\hat{z}_r = u + \frac{\hat{\sigma}}{\hat{\xi}} \left[(r n_y \hat{\lambda}_u)^{\hat{\xi}} - 1 \right],$$

unless $\xi = 0$, in which case

$$\hat{z}_r = u + \hat{\sigma} \log(r n_y \hat{\lambda}_u).$$

3.2.5 Return level estimation

If there are n_y observations per year, this corresponds to the t–observation return level with $t = r \times n_y$.

Hence, an estimate of the r–year return level z_r is defined by

$$\hat{z}_r = u + \frac{\hat{\sigma}}{\hat{\xi}} \left[(r n_y \hat{\lambda}_u)^{\hat{\xi}} - 1 \right],$$

unless $\xi = 0$, in which case

$$\hat{z}_r \quad = \quad u + \hat{\sigma} \log(r n_y \hat{\lambda}_u).$$

3.2.5 Return level estimation

Thus, for the New York City rainfall extremes, we have

$$\hat{z}_{50} = 30 + \frac{7.44}{0.184}\left[(50 \times 365.25 \times 0.00867)^{0.184} - 1\right] = 92.24\text{mm}$$

as an estimate of the 50–year return level, where $n_y = 365.25$ to account for leap years (we have daily observations).

3.2.5 Return level estimation

The delta method can, once again, be used to obtain estimated standard errors for such return levels.

We should, however, also include uncertainty in our estimate of λ_u in the calculation (since z_r is a function of λ_u).

Now the number of threshold exceedances follows a binomial distribution $\text{Bin}(N, \lambda_u)$, where N is the total number of observations in the series. We know (from MAS2302) that

$$Var(\hat{\lambda}_u) = \hat{\lambda}_u(1 - \hat{\lambda}_u)/N = 0.0007^2$$

in our rainfall example.

3.2.5 Return level estimation

The delta method can, once again, be used to obtain estimated standard errors for such return levels.

We should, however, also include uncertainty in our estimate of λ_u in the calculation (since z_r is a function of λ_u).

Now the number of threshold exceedances follows a binomial distribution $\text{Bin}(N, \lambda_u)$, where N is the total number of observations in the series. We know (from MAS2302) that

$$Var(\hat{\lambda}_u) = \hat{\lambda}_u(1 - \hat{\lambda}_u)/N = 0.0007^2$$

in our rainfall example.

3.2.5 Return level estimation

The delta method can, once again, be used to obtain estimated standard errors for such return levels.

We should, however, also include uncertainty in our estimate of λ_u in the calculation (since z_r is a function of λ_u).

Now the number of threshold exceedances follows a binomial distribution $\text{Bin}(N, \lambda_u)$, where N is the total number of observations in the series. We know (from MAS2302) that

$$Var(\hat{\lambda}_u) = \hat{\lambda}_u(1 - \hat{\lambda}_u)/N = 0.0007^2$$

in our rainfall example.

3.2.5 Return level estimation

The delta method can, once again, be used to obtain estimated standard errors for such return levels.

We should, however, also include uncertainty in our estimate of λ_u in the calculation (since z_r is a function of λ_u).

Now the number of threshold exceedances follows a binomial distribution $\text{Bin}(N, \lambda_u)$, where N is the total number of observations in the series. We know (from MAS2302) that

$$Var(\hat{\lambda}_u) = \hat{\lambda}_u(1 - \hat{\lambda}_u)/N = 0.0007^2$$

in our rainfall example.

3.2.5 Return level estimation

Recall that, by the delta method,

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

Here, V is now the variance–covariance matrix of the triple $(\hat{\lambda}_u, \hat{\sigma}, \hat{\xi})^T$; in our rainfall example, this is

$$V = \begin{pmatrix} 0.0007^2 & & \\ 0 & 0.958^2 & \\ 0 & -0.0655 & 0.101^2 \end{pmatrix},$$

assuming that $Var(\hat{\lambda}_u, \hat{\sigma}) = Var(\hat{\lambda}_u, \hat{\xi}) = 0$.

3.2.5 Return level estimation

Recall that, by the delta method,

$$\mathrm{Var}(\hat{z}_r) \quad \approx \quad \nabla z_r^T V \nabla z_r.$$

Here, V is now the variance–covariance matrix of the triple $(\hat{\lambda}_u, \hat{\sigma}, \hat{\xi})^T$; in our rainfall example, this is

$$V \quad = \quad \begin{pmatrix} 0.0007^2 & & \\ 0 & 0.958^2 & \\ 0 & -0.0655 & 0.101^2 \end{pmatrix},$$

assuming that $Var(\hat{\lambda}_u, \hat{\sigma}) = Var(\hat{\lambda}_u, \hat{\xi}) = 0$.

3.2.5 Return level estimation

Recall that, by the delta method,

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

Here, V is now the variance–covariance matrix of the triple $(\hat{\lambda}_u, \hat{\sigma}, \hat{\xi})^T$; in our rainfall example, this is

$$V = \begin{pmatrix} 0.0007^2 & & \\ 0 & 0.958^2 & \\ 0 & -0.0655 & 0.101^2 \end{pmatrix},$$

assuming that $Var(\hat{\lambda}_u, \hat{\sigma}) = Var(\hat{\lambda}_u, \hat{\xi}) = 0$.

3.2.5 Return level estimation (1/3)

As in Chapter 2,

$$\nabla z_r^T = \left[\frac{\partial z_r}{\partial \lambda_u}, \frac{\partial z_r}{\partial \sigma}, \frac{\partial z_r}{\partial \xi}\right];$$

this can be shown to give

$$\begin{aligned}\nabla z_r^T = & \left[\sigma (r n_y)^{\xi} \lambda_u^{\xi-1}, \xi^{-1}\left\{(r n_y \lambda_u)^{\xi} - 1\right\},\right. \\ & \left. -\sigma \xi^{-2}\left\{(r n_y \lambda_u) - 1\right\} + \sigma \xi^{-1} (r n_y \lambda_u)^{\xi} \log(r n_y \lambda_u)\right],\end{aligned}$$

which we evaluate at $(\hat{\lambda}_u, \hat{\sigma}, \hat{\xi})$.

3.2.5 Return level estimation (1/3)

As in Chapter 2,

$$\nabla z_r^T = \left[\frac{\partial z_r}{\partial \lambda_u}, \frac{\partial z_r}{\partial \sigma}, \frac{\partial z_r}{\partial \xi}\right];$$

this can be shown to give

$$\begin{aligned}\nabla z_r^T &= \Big[\sigma (rn_y)^{\xi} \lambda_u^{\xi-1}, \xi^{-1}\left\{(rn_y\lambda_u)^{\xi} - 1\right\}, \\ &\quad -\sigma\xi^{-2}\left\{(rn_y\lambda_u) - 1\right\} + \sigma\xi^{-1}(rn_y\lambda_u)^{\xi}\log(rn_y\lambda_u)\Big],\end{aligned}$$

which we evaluate at $(\hat{\lambda}_u, \hat{\sigma}, \hat{\xi})$.

3.2.5 Return level estimation (1/3)

As in Chapter 2,

$$\nabla z_r^T = \left[\frac{\partial z_r}{\partial \lambda_u}, \frac{\partial z_r}{\partial \sigma}, \frac{\partial z_r}{\partial \xi}\right];$$

this can be shown to give

$$\begin{aligned}\nabla z_r^T &= \Big[\sigma(rn_y)^{\xi}\lambda_u^{\xi-1}, \xi^{-1}\left\{(rn_y\lambda_u)^{\xi} - 1\right\}, \\ &\quad -\sigma\xi^{-2}\left\{(rn_y\lambda_u) - 1\right\} + \sigma\xi^{-1}(rn_y\lambda_u)^{\xi}\log(rn_y\lambda_u)\Big],\end{aligned}$$

which we evaluate at $(\hat{\lambda}_u, \hat{\sigma}, \hat{\xi})$.

3.2.5 Return level estimation (2/3)

This gives:

$$\nabla z_{50}^T = (2179.096, 8.366, 181.757),$$

and so

$$\begin{aligned} Var(\hat{z}_{50}) &= (2179.096, 8.366, 181.757) \\ &\quad \times \begin{pmatrix} 0.0007^2 & & \\ 0 & 0.958^2 & \\ 0 & -0.0655 & 0.101^2 \end{pmatrix} \\ &\quad \times \begin{pmatrix} 2179.096 \\ 8.366 \\ 181.757 \end{pmatrix} \\ &= 397.1853 \end{aligned}$$

3.2.5 Return level estimation (2/3)

This gives:

$$\nabla z_{50}^T = (2179.096, 8.366, 181.757),$$

and so

$$\begin{aligned} Var(\hat{z}_{50}) &= (2179.096, 8.366, 181.757) \\ &\quad \times \begin{pmatrix} 0.0007^2 & & \\ 0 & 0.958^2 & \\ 0 & -0.0655 & 0.101^2 \end{pmatrix} \\ &\quad \times \begin{pmatrix} 2179.096 \\ 8.366 \\ 181.757 \end{pmatrix} \\ &= 397.1853 \end{aligned}$$

3.2.5 Return level estimation (2/3)

This gives:

$$\nabla z_{50}^T = (2179.096, 8.366, 181.757),$$

and so

$$\begin{aligned} Var(\hat{z}_{50}) &= (2179.096, 8.366, 181.757) \\ &\quad \times \begin{pmatrix} 0.0007^2 & & \\ 0 & 0.958^2 & \\ 0 & -0.0655 & 0.101^2 \end{pmatrix} \\ &\quad \times \begin{pmatrix} 2179.096 \\ 8.366 \\ 181.757 \end{pmatrix} \\ &= 397.1853 \end{aligned}$$

3.2.5 Return level estimation (2/3)

This gives:

$$\nabla z_{50}^T = (2179.096, 8.366, 181.757),$$

and so

$$\begin{aligned} Var(\hat{z}_{50}) &= (2179.096, 8.366, 181.757) \\ &\quad \times \begin{pmatrix} 0.0007^2 & & \\ 0 & 0.958^2 & \\ 0 & -0.0655 & 0.101^2 \end{pmatrix} \\ &\quad \times \begin{pmatrix} 2179.096 \\ 8.366 \\ 181.757 \end{pmatrix} \\ &= 397.1853 \end{aligned}$$

3.2.5 Return level estimation (2/3)

This gives:

$$\nabla z_{50}^T = (2179.096, 8.366, 181.757),$$

and so

$$\begin{aligned} Var(\hat{z}_{50}) &= (2179.096, 8.366, 181.757) \\ &\quad \times \begin{pmatrix} 0.0007^2 & & \\ 0 & 0.958^2 & \\ 0 & -0.0655 & 0.101^2 \end{pmatrix} \\ &\quad \times \begin{pmatrix} 2179.096 \\ 8.366 \\ 181.757 \end{pmatrix} \\ &= 397.1853 \end{aligned}$$

3.2.5 Return level estimation (3/3)

And so

$$s.e.(\hat{z}_{50}) = \sqrt{397.1853} = 19.930.$$

3.2.6 Profile likelihood

As discussed in Chapter 2, confidence intervals for return levels are better constructed via the method of profile likelihood, owing to the asymmetry in the likelihood surface often observed.

A plot of the profile log–likelihood for the 50–year return level for daily rainfall accumulations at New York is shown in Figure 3.3 of the lecture notes.

This gives

$$(74.1 \text{ mm}, 143 \text{ mm}).$$

Interpretation: "Once every fifty years, we might expect daily rainfall accumulations in New York City to reach up to about 143mm".

3.2.6 Profile likelihood

Compare the 95% confidence interval for the 50–year return level obtained from profiling the log–likelihood to that you would obtain using the standard error. Comment.

3.2.6 Profile likelihood

From the profiled log–likelihood:

$$(74.1 \text{ mm}, 143 \text{ mm}).$$

Using the standard error, we get:

$$92.24 \pm 1.96 \times 19.93 \longrightarrow (53.2 \text{ mm}, 131.3 \text{ mm}).$$

Comment: the confidence interval based on the profiled log–likelihood is much more conservative, giving a substantially higher upper bound for the return level.

Chapter 4

Dependent extremes

4.1 Introduction

The threshold–based approach to modelling extremes has obvious practical advantages over the more traditional ‘block maxima’ approach.

Since extremes are (by their very nature) scarce, a modelling procedure which allows the inclusion of more data in the analysis has got to be a good thing.

Indeed, the whole point of a threshold–based analysis is that we include *all* extremes in the analysis — extreme in the sense that the observations used have all exceeded some pre–determined threshold u.

Advantages

- Data pre–processing dead easy
- Estimates with **smaller standard errors** (including return levels!)
- **Narrower confidence intervals** (standard or profile–likelihood based)

Disadvantages

- **Temporal dependence** (e.g. temperature extremes, wind gusts,...)
- **Non–stationarity** (e.g. trends and seasonal effects)

Advantages

- Data pre–processing dead easy
- Estimates with **smaller standard errors** (including return levels!)
- **Narrower confidence intervals** (standard or profile–likelihood based)

Disadvantages

- **Temporal dependence** (e.g. temperature extremes, wind gusts,...)
- **Non–stationarity** (e.g. trends and seasonal effects)

Advantages

- Data pre–processing dead easy
- Estimates with **smaller standard errors** (including return levels!)
- **Narrower confidence intervals** (standard or profile–likelihood based)

Disadvantages

- **Temporal dependence** (e.g. temperature extremes, wind gusts,...)
- **Non–stationarity** (e.g. trends and seasonal effects)

Advantages

- Data pre–processing dead easy
- Estimates with **smaller standard errors** (including return levels!)
- **Narrower confidence intervals** (standard or profile–likelihood based)

Disadvantages

- **Temporal dependence** (e.g. temperature extremes, wind gusts,...)
- **Non–stationarity** (e.g. trends and seasonal effects)

4.1 Introduction

Advantages

- Data pre–processing dead easy
- Estimates with **smaller standard errors** (including return levels!)
- **Narrower confidence intervals** (standard or profile–likelihood based)

Disadvantages

- **Temporal dependence** (e.g. temperature extremes, wind gusts,...)
- **Non–stationarity** (e.g. trends and seasonal effects)

4.1 Introduction

Advantages

- Data pre–processing dead easy
- Estimates with **smaller standard errors** (including return levels!)
- **Narrower confidence intervals** (standard or profile–likelihood based)

Disadvantages

- **Temporal dependence** (e.g. temperature extremes, wind gusts,...)
- **Non–stationarity** (e.g. trends and seasonal effects)

4.1 Introduction

Advantages

- Data pre–processing dead easy
- Estimates with **smaller standard errors** (including return levels!)
- **Narrower confidence intervals** (standard or profile–likelihood based)

Disadvantages

- **Temporal dependence** (e.g. temperature extremes, wind gusts,...)
- **Non–stationarity** (e.g. trends and seasonal effects)

4.2 Example: Hourly wind speeds at High Bradfield

Hourly maximum wind gusts (in **knots**) were collected at High Bradfield, in the Peak District, over a period of 10 years from January 1st 2003 to December 31st 2012.

This gives total of 87,672 observations (including any missing values: 81,835 after the missing values have been removed).

Suppose these hourly observations are in the vector `brad` in R; the series without the missing values has been stored in the vector `brad2`.

4.2 Example: Hourly wind speeds at High Bradfield

Hourly maximum wind gusts (in **knots**) were collected at High Bradfield, in the Peak District, over a period of 10 years from January 1st 2003 to December 31st 2012.

This gives total of 87,672 observations (including any missing values: 81,835 after the missing values have been removed).

Suppose these hourly observations are in the vector `brad` in R; the series without the missing values has been stored in the vector `brad2`.

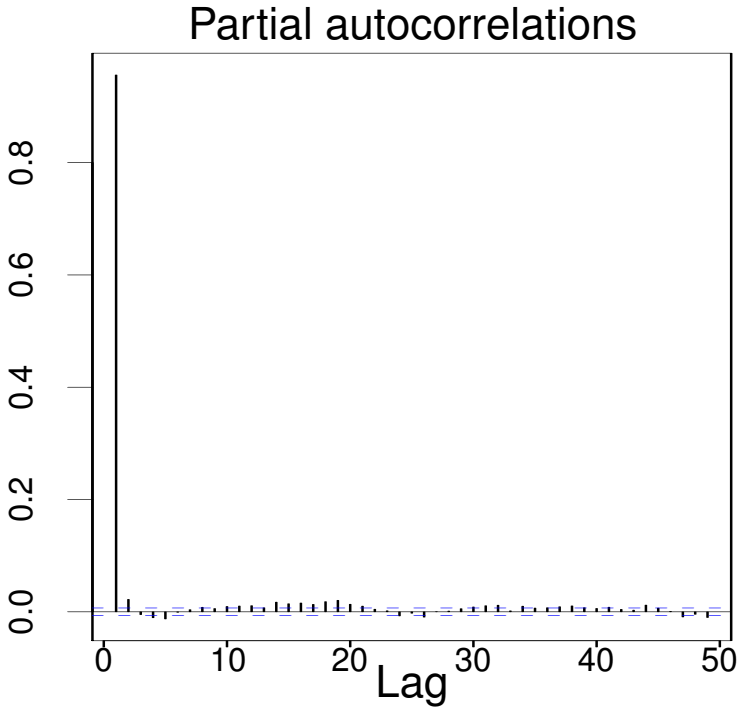
4.2 Example: Hourly wind speeds at High Bradfield

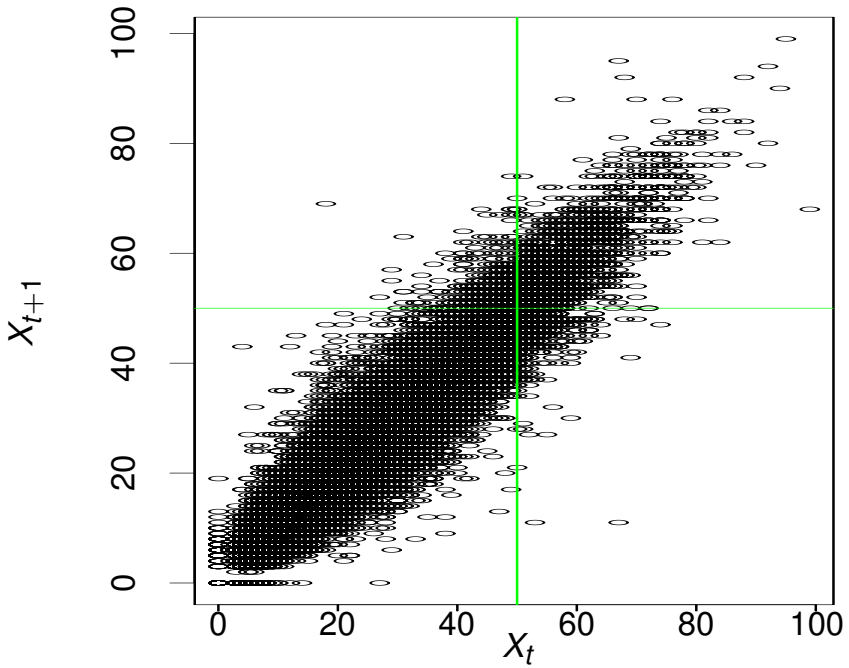
Hourly maximum wind gusts (in **knots**) were collected at High Bradfield, in the Peak District, over a period of 10 years from January 1st 2003 to December 31st 2012.

This gives total of 87,672 observations (including any missing values: 81,835 after the missing values have been removed).

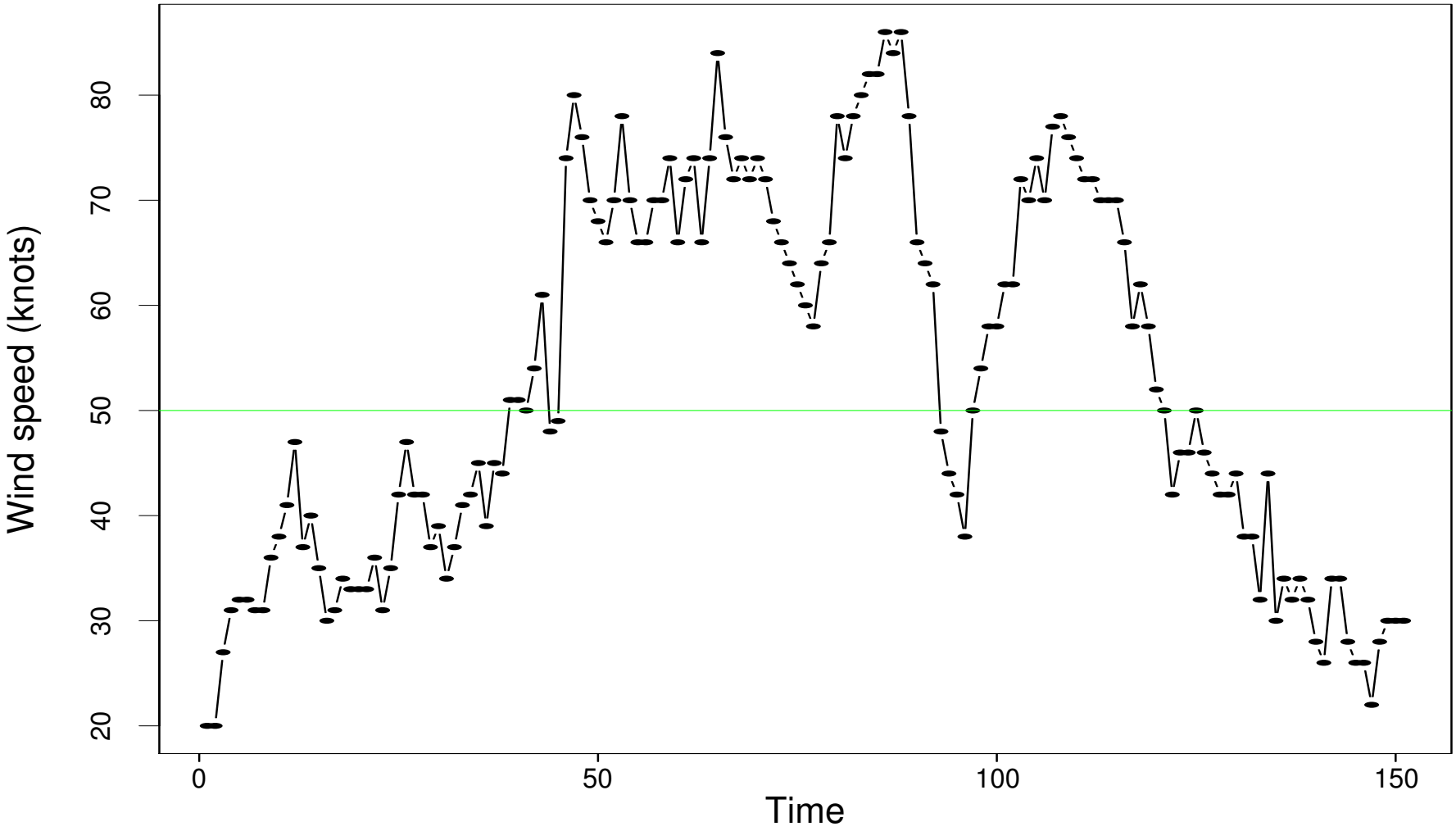
Suppose these hourly observations are in the vector `brad` in R; the series without the missing values has been stored in the vector `brad2`.

4.2 Example: Hourly wind speeds at High Bradfield





4.2 Example: Hourly wind speeds at High Bradfield



4.2 Example: Hourly wind speeds at High Bradfield

It is clear that we have (fairly strong) dependence between consecutive wind speeds recordings – even at extreme levels.

How can we proceed?

- **Ignore** this dependence – proceed as normal?

  Possible consequence: Standard errors under–estimated as there are fewer independence observations than the likelihood assumes

- **Model** this dependence – see Chapter 6

- **Filter out** this dependence

4.2 Example: Hourly wind speeds at High Bradfield

It is clear that we have (fairly strong) dependence between consecutive wind speeds recordings – even at extreme levels.

How can we proceed?

- **Ignore** this dependence – proceed as normal?

  Possible consequence: Standard errors under–estimated as there are fewer independence observations than the likelihood assumes

- **Model** this dependence – see Chapter 6

- **Filter out** this dependence

4.2 Example: Hourly wind speeds at High Bradfield

It is clear that we have (fairly strong) dependence between consecutive wind speeds recordings – even at extreme levels.

How can we proceed?

- **Ignore** this dependence – proceed as normal?

  Possible consequence: Standard errors under–estimated as there are fewer independence observations than the likelihood assumes

- **Model** this dependence – see Chapter 6

- **Filter out** this dependence

4.2 Example: Hourly wind speeds at High Bradfield

It is clear that we have (fairly strong) dependence between consecutive wind speeds recordings – even at extreme levels.

How can we proceed?

- **Ignore** this dependence – proceed as normal?

  Possible consequence: Standard errors under–estimated as there are fewer independence observations than the likelihood assumes

- **Model** this dependence – see Chapter 6

- **Filter out** this dependence

4.2 Example: Hourly wind speeds at High Bradfield

It is clear that we have (fairly strong) dependence between consecutive wind speeds recordings – even at extreme levels.

How can we proceed?

- **Ignore** this dependence – proceed as normal?

  Possible consequence: Standard errors under–estimated as there are fewer independence observations than the likelihood assumes

- **Model** this dependence – see Chapter 6

- **Filter out** this dependence

4.2 Example: Hourly wind speeds at High Bradfield

It is clear that we have (fairly strong) dependence between consecutive wind speeds recordings – even at extreme levels.

How can we proceed?

- **Ignore** this dependence – proceed as normal?

  Possible consequence: Standard errors under–estimated as there are fewer independence observations than the likelihood assumes

- **Model** this dependence – see Chapter 6

- **Filter out** this dependence

4.3 Maxima of dependent (but stationary) series

The book by **Leadbetter *et al.* (1983)** considers, in great detail, properties of extremes of dependent processes.

A key result often used is '**Leadbetter's $D(u_n)$ condition**', which ensures that long–range dependence is sufficiently weak so as not to affect the asymptotics of an extreme value analysis.

This condition is stated more formally in the definition below.

4.3 Maxima of dependent (but stationary) series

The book by **Leadbetter *et al.* (1983)** considers, in great detail, properties of extremes of dependent processes.

A key result often used is '**Leadbetter's $D(u_n)$ condition**', which ensures that long–range dependence is sufficiently weak so as not to affect the asymptotics of an extreme value analysis.

This condition is stated more formally in the definition below.

4.3 Maxima of dependent (but stationary) series

The book by **Leadbetter *et al.* (1983)** considers, in great detail, properties of extremes of dependent processes.

A key result often used is '**Leadbetter's $D(u_n)$ condition**', which ensures that long–range dependence is sufficiently weak so as not to affect the asymptotics of an extreme value analysis.

This condition is stated more formally in the definition below.

Definition *(Leadbetter's $D(u_n)$ condition)*

A stationary series $\tilde{X}_1, \tilde{X}_2, \ldots$ is said to satisfy the $D(u_n)$ condition if, for all $i_1 < \ldots < i_p < j_1 < \ldots < j_q$ with $j_1 - i_p > l$,

$$\left| \Pr\left\{ \tilde{X}_{i_1} \leq u_n, \ldots, \tilde{X}_{i_p} \leq u_n, \tilde{X}_{j_1} \leq u_n, \ldots, \tilde{X}_{j_q} \leq u_n \right\} - \Pr\left\{ \tilde{X}_{i_1} \leq u_n, \ldots, \tilde{X}_{i_p} \leq u_n \right\} \Pr\left\{ \tilde{X}_{j_1} \leq u_n, \ldots, \tilde{X}_{j_q} \leq u_n \right\} \right| \leq \alpha(n, l),$$

where $\alpha(n, l) \to 0$ for some sequence l_n such that $l_n/n \to 0$ as $n \to \infty$.

Definition *(Leadbetter's $D(u_n)$ condition)*

A stationary series $\tilde{X}_1, \tilde{X}_2, \ldots$ is said to satisfy the $D(u_n)$ condition if, for all $i_1 < \ldots < i_p < j_1 < \ldots < j_q$ with $j_1 - i_p > l$,

$$\left| \Pr\left\{\tilde{X}_{i_1} \leq u_n, \ldots, \tilde{X}_{i_p} \leq u_n, \tilde{X}_{j_1} \leq u_n, \ldots, \tilde{X}_{j_q} \leq u_n\right\} \right.$$
$$\left. -\Pr\left\{\tilde{X}_{i_1} \leq u_n, \ldots, \tilde{X}_{i_p} \leq u_n\right\} \Pr\left\{\tilde{X}_{j_1} \leq u_n, \ldots, \tilde{X}_{j_q} \leq u_n\right\} \right| \quad \leq \quad \alpha(n, l),$$

where $\alpha(n, l) \to 0$ for some sequence l_n such that $l_n/n \to 0$ as $n \to \infty$.

4.3 Maxima of dependent (but stationary) series

For sequences of independent variables, the difference in probabilities in the above expression is exactly zero for *any* sequence u_n.

More generally, we will require that the $D(u_n)$ condition holds only for a specific sequence of thresholds u_n that increases with n.

For such a sequence, the $D(u_n)$ condition ensures that, for sets of variables that are far enough apart, the difference in probabilities expressed in (4.1), while not zero, is sufficiently close to zero to have no effect on the limit laws for extremes.

4.3 Maxima of dependent (but stationary) series

Theorem *(Extremes of dependent sequences)*

Let $\tilde{X}_1, \tilde{X}_2, \ldots$ be a stationary series satisfying Leadbetter's $D(u_n)$ condition, and let $\tilde{M}_n = \max\{\tilde{X}_1, \ldots, \tilde{X}_n\}$.

Now let $X_1, X_2, \ldots$ be an **independent** series with X having the same distribution as $\tilde{X}$, and let $M_n = \max\{X_1, \ldots, X_n\}$. Then if M_n has a non–degenerate limit law given by $\Pr\{(M_n - b_n)/a_n \leq x\} \to G(x)$, it follows that

$$\Pr\left\{(\tilde{M}_n - b_n)/a_n \leq x\right\} \to G^{\theta}(x) \tag{4.2}$$

for some $0 \leq \theta \leq 1$.

4.3 Maxima of dependent (but stationary) series

The parameter θ is known as the **extremal index**, and quantifies the extent of extremal dependence

- $\theta = 1$: completely independent process
- $\theta \to 0$: increasing levels of (extremal) dependence

4.3 Maxima of dependent (but stationary) series

The parameter θ is known as the **extremal index**, and quantifies the extent of extremal dependence

- $\theta = 1$: completely independent process
- $\theta \to 0$: increasing levels of (extremal) dependence

4.3 Maxima of dependent (but stationary) series

The parameter θ is known as the **extremal index**, and quantifies the extent of extremal dependence

- $\theta = 1$: completely independent process
- $\theta \to 0$: increasing levels of (extremal) dependence

4.3 Maxima of dependent (but stationary) series

Since G in the above theorem is necessarily an extreme value distribution, and due to the **max–stability** property (see Leadbetter *et al.*, 1983), then the distribution of maxima in processes displaying short–range temporal dependence (characterised by the extremal index θ) is also a GEV distribution.

The powering of the limit distribution by θ only affects the location and scale parameters of this distribution.

4.3 Maxima of dependent (but stationary) series

Since G in the above theorem is necessarily an extreme value distribution, and due to the **max–stability** property (see Leadbetter *et al.*, 1983), then the distribution of maxima in processes displaying short–range temporal dependence (characterised by the extremal index θ) is also a GEV distribution.

The powering of the limit distribution by θ only affects the location and scale parameters of this distribution.

4.3 Maxima of dependent (but stationary) series

What does this all mean in **practical terms**?

- If maxima of a stationary series converge in distribution (and we know they do – the GEV), and
- if Leadbetter's $D(u_n)$ condition holds (i.e. long–range dependence is negligible), then
- the limit distribution is related to that of the independent series...
- ...in fact, it's $G^\theta(x)$, where G is the GEV for the independent series (with parameters μ, σ and ξ)

4.3 Maxima of dependent (but stationary) series

What does this all mean in **practical terms**?

- If maxima of a stationary series converge in distribution (and we know they do – the GEV), and
- if Leadbetter's $D(u_n)$ condition holds (i.e. long–range dependence is negligible), then
- the limit distribution is related to that of the independent series...
- ...in fact, it's $G^{\theta}(x)$, where G is the GEV for the independent series (with parameters μ, σ and ξ)

4.3 Maxima of dependent (but stationary) series

What does this all mean in **practical terms**?

- If maxima of a stationary series converge in distribution (and we know they do – the GEV), and
- if Leadbetter's $D(u_n)$ condition holds (i.e. long–range dependence is negligible), then
- the limit distribution is related to that of the independent series...
- ...in fact, it's $G^{\theta}(x)$, where G is the GEV for the independent series (with parameters μ, σ and ξ)

4.3 Maxima of dependent (but stationary) series

What does this all mean in **practical terms**?

- If maxima of a stationary series converge in distribution (and we know they do – the GEV), and
- if Leadbetter's $D(u_n)$ condition holds (i.e. long–range dependence is negligible), then
- the limit distribution is related to that of the independent series...
- ...in fact, it's $G^{\theta}(x)$, where G is the GEV for the independent series (with parameters μ, σ and ξ)

4.3 Maxima of dependent (but stationary) series

What does this all mean in **practical terms**?

- If maxima of a stationary series converge in distribution (and we know they do – the GEV), and
- if Leadbetter's $D(u_n)$ condition holds (i.e. long–range dependence is negligible), then
- the limit distribution is related to that of the independent series...
- ...in fact, it's $G^\theta(x)$, where G is the GEV for the independent series (with parameters μ, σ and ξ)

4.3 Maxima of dependent (but stationary) series

Further, it can be shown that

$$\begin{aligned} G^{\theta}(x) &= \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}^{\theta} \\ &= \exp\left\{-\left[1+\xi\left(\frac{x-\mu^*}{\sigma^*}\right)\right]^{-1/\xi}\right\}, \end{aligned}$$

where $\mu^* = \mu - \frac{\sigma}{\xi}\left(1-\theta^{-\xi}\right)$ and $\sigma^* = \sigma\theta^{\xi}$.

4.3.1 Modelling block maxima

We can proceed as we did in Chapter 2 – even in the presence of dependence...

... provided long range dependence is weak (i.e. **Leadbetter's $D(u_n)$ condition** can be assumed)

The only difference is the parameters of the GEV: $(\mu, \sigma, \xi) \longrightarrow (\mu^*, \sigma^*, \xi)$.

Since we estimates these parameters anyway, who cares?!

4.3.1 Modelling block maxima

We can proceed as we did in Chapter 2 – even in the presence of dependence...

... provided long range dependence is weak (i.e. **Leadbetter's $D(u_n)$ condition** can be assumed)

The only difference is the parameters of the GEV: $(\mu, \sigma, \xi) \longrightarrow (\mu^*, \sigma^*, \xi)$.

Since we estimates these parameters anyway, who cares?!

4.3.1 Modelling block maxima

We can proceed as we did in Chapter 2 – even in the presence of dependence...

... provided long range dependence is weak (i.e. **Leadbetter's $D(u_n)$ condition** can be assumed)

The only difference is the parameters of the GEV: $(\mu, \sigma, \xi) \longrightarrow (\mu^*, \sigma^*, \xi)$.

Since we estimates these parameters anyway, who cares?!

4.3.1 Modelling block maxima

We can proceed as we did in Chapter 2 – even in the presence of dependence...

... provided long range dependence is weak (i.e. **Leadbetter's $D(u_n)$ condition** can be assumed)

The only difference is the parameters of the GEV: $(\mu, \sigma, \xi) \longrightarrow (\mu^*, \sigma^*, \xi)$.

Since we estimates these parameters anyway, who cares?!

4.3.2 Modelling threshold exceedances

Though the modelling procedure for fitting the GEV to a set of annual maxima is unchanged for series which display short–term temporal dependence, some revision is needed of the threshold exceedance approach.

If all threshold exceedances are used in our analysis, and the GPD fitted to the set of threshold excesses, the likelihoods we use will be incorrect since they assume independence of sample observations.

4.3.2 Modelling threshold exceedances

Though the modelling procedure for fitting the GEV to a set of annual maxima is unchanged for series which display short–term temporal dependence, some revision is needed of the threshold exceedance approach.

If all threshold exceedances are used in our analysis, and the GPD fitted to the set of threshold excesses, the likelihoods we use will be incorrect since they assume independence of sample observations.

4.3.2 Modelling threshold exceedances

In practice, several techniques have been developed to circumvent this problem, including:

1. filtering out an (approximately) independent set of threshold exceedances
2. fitting the GPD to *all* exceedances, ignoring dependence, but then appropriately adjusting the inference (usually an inflation of standard errors) to take into account the reduction in information
3. Explicitly modelling the temporal dependence in the process

4.3.2 Declustering

Since the mid–1990s, various methods for **declustering** a series of extremes, to extract a set of independent extremes, have been discussed in the literature.

The most natural, commonly–used method of declustering is that of **runs declustering**. This is how it works:

1. Choose an auxiliary 'declustering parameter' (which we call κ)
2. A cluster of threshold excesses is then deemed to have terminated as soon as at least κ consecutive observations fall below the threshold
3. Go through the entire series identifying clusters in this way
4. The maximum (or 'peak') observation from each cluster is then extracted, and the GPD fitted to the set of cluster peak excesses.

4.3.2 Declustering

Since the mid–1990s, various methods for **declustering** a series of extremes, to extract a set of independent extremes, have been discussed in the literature.

The most natural, commonly–used method of declustering is that of **runs declustering**. This is how it works:

1. Choose an auxiliary 'declustering parameter' (which we call κ)
2. A cluster of threshold excesses is then deemed to have terminated as soon as at least κ consecutive observations fall below the threshold
3. Go through the entire series identifying clusters in this way
4. The maximum (or 'peak') observation from each cluster is then extracted, and the GPD fitted to the set of cluster peak excesses.

4.3.2 Declustering

Since the mid–1990s, various methods for **declustering** a series of extremes, to extract a set of independent extremes, have been discussed in the literature.

The most natural, commonly–used method of declustering is that of **runs declustering**. This is how it works:

1. Choose an auxiliary 'declustering parameter' (which we call κ)
2. A cluster of threshold excesses is then deemed to have terminated as soon as at least κ consecutive observations fall below the threshold
3. Go through the entire series identifying clusters in this way
4. The maximum (or 'peak') observation from each cluster is then extracted, and the GPD fitted to the set of cluster peak excesses.

4.3.2 Declustering

Since the mid–1990s, various methods for **declustering** a series of extremes, to extract a set of independent extremes, have been discussed in the literature.

The most natural, commonly–used method of declustering is that of **runs declustering**. This is how it works:

1. Choose an auxiliary 'declustering parameter' (which we call κ)
2. A cluster of threshold excesses is then deemed to have terminated as soon as at least κ consecutive observations fall below the threshold
3. Go through the entire series identifying clusters in this way
4. The maximum (or 'peak') observation from each cluster is then extracted, and the GPD fitted to the set of cluster peak excesses.

4.3.2 Declustering

This approach is often referred to as the **peaks over threshold** approach (POT, **Davison and Smith, 1990**) and is widely accepted as the main pragmatic approach for dealing with clustered extremes.

Easy to implement, but if:

- κ is too small, the cluster peaks will not be far enough apart to safely assume independence
- κ is too large, there will be too few cluster exceedances on which to form our inference

4.3.2 Declustering

This approach is often referred to as the **peaks over threshold** approach (POT, **Davison and Smith, 1990**) and is widely accepted as the main pragmatic approach for dealing with clustered extremes.

Easy to implement, but if:

- κ is too small, the cluster peaks will not be far enough apart to safely assume independence
- κ is too large, there will be too few cluster exceedances on which to form our inference

4.3.2 Declustering

This approach is often referred to as the **peaks over threshold** approach (POT, **Davison and Smith, 1990**) and is widely accepted as the main pragmatic approach for dealing with clustered extremes.

Easy to implement, but if:

- κ is too small, the cluster peaks will not be far enough apart to safely assume independence
- κ is too large, there will be too few cluster exceedances on which to form our inference

4.3.2 Declustering

This approach is often referred to as the **peaks over threshold** approach (POT, **Davison and Smith, 1990**) and is widely accepted as the main pragmatic approach for dealing with clustered extremes.

Easy to implement, but if:

- κ is too small, the cluster peaks will not be far enough apart to safely assume independence
- κ is too large, there will be too few cluster exceedances on which to form our inference

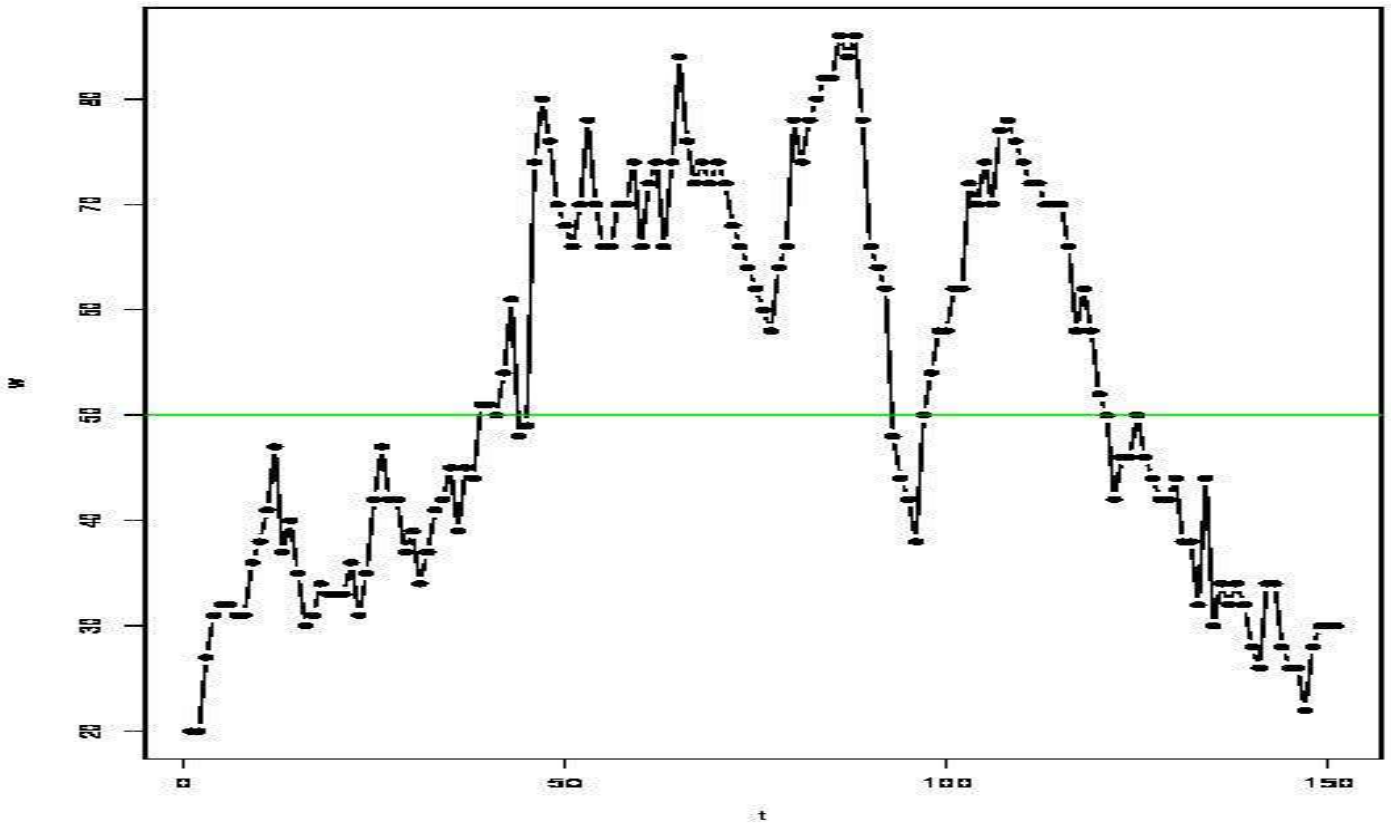
4.3.2 Declustering

It has also been shown (**Fawcett and Walshaw, 2012**)[1] that parameter estimates can be sensitive to the choice of κ, and κ is all too often chosen arbitrarily.

[1] Estimating return levels from serially dependent extremes, *Environmetrics* **23**(3), pp 272–283

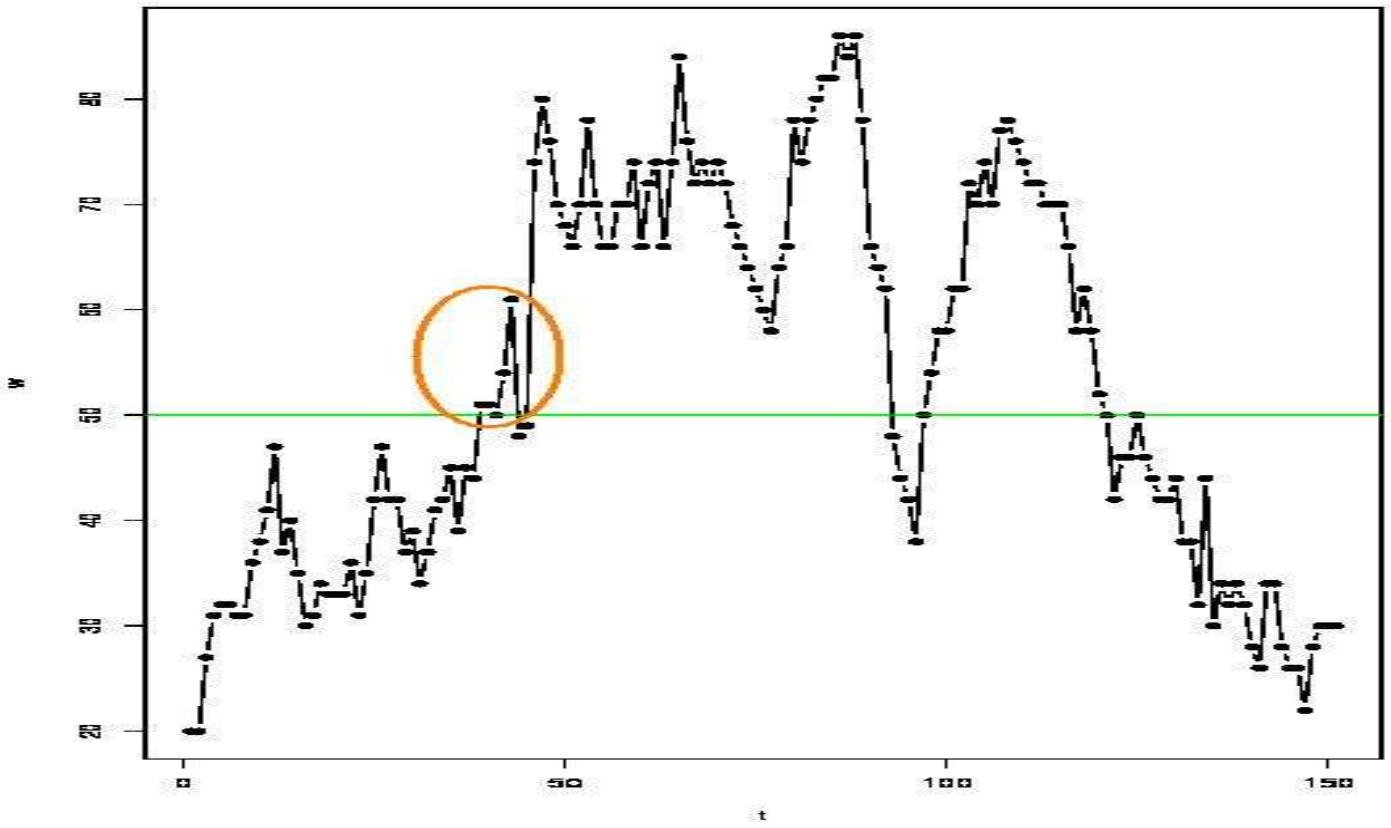
4.3.2 Declustering

Using $\kappa = 2$:



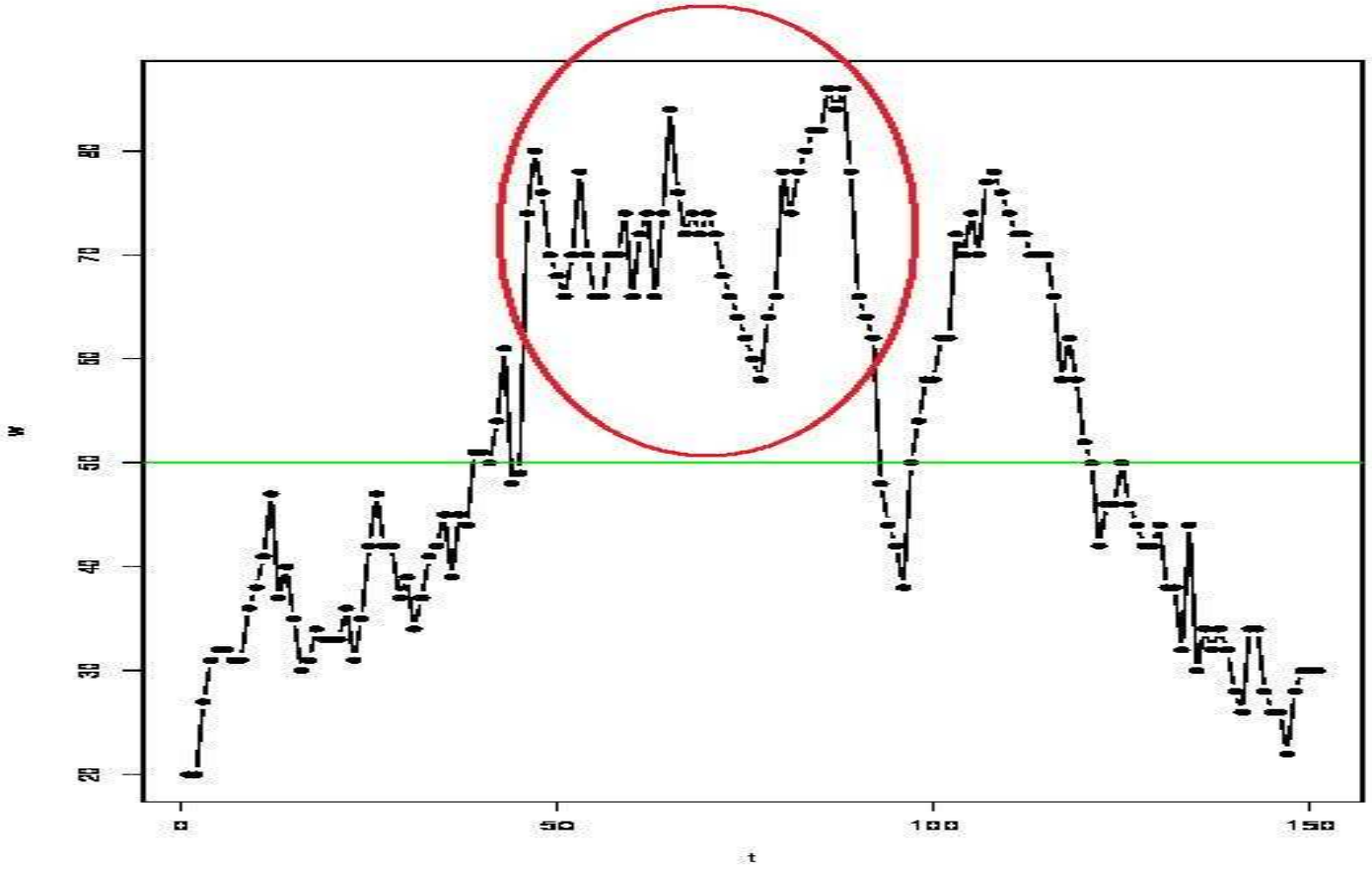
4.3.2 Declustering

Using $\kappa = 2$:



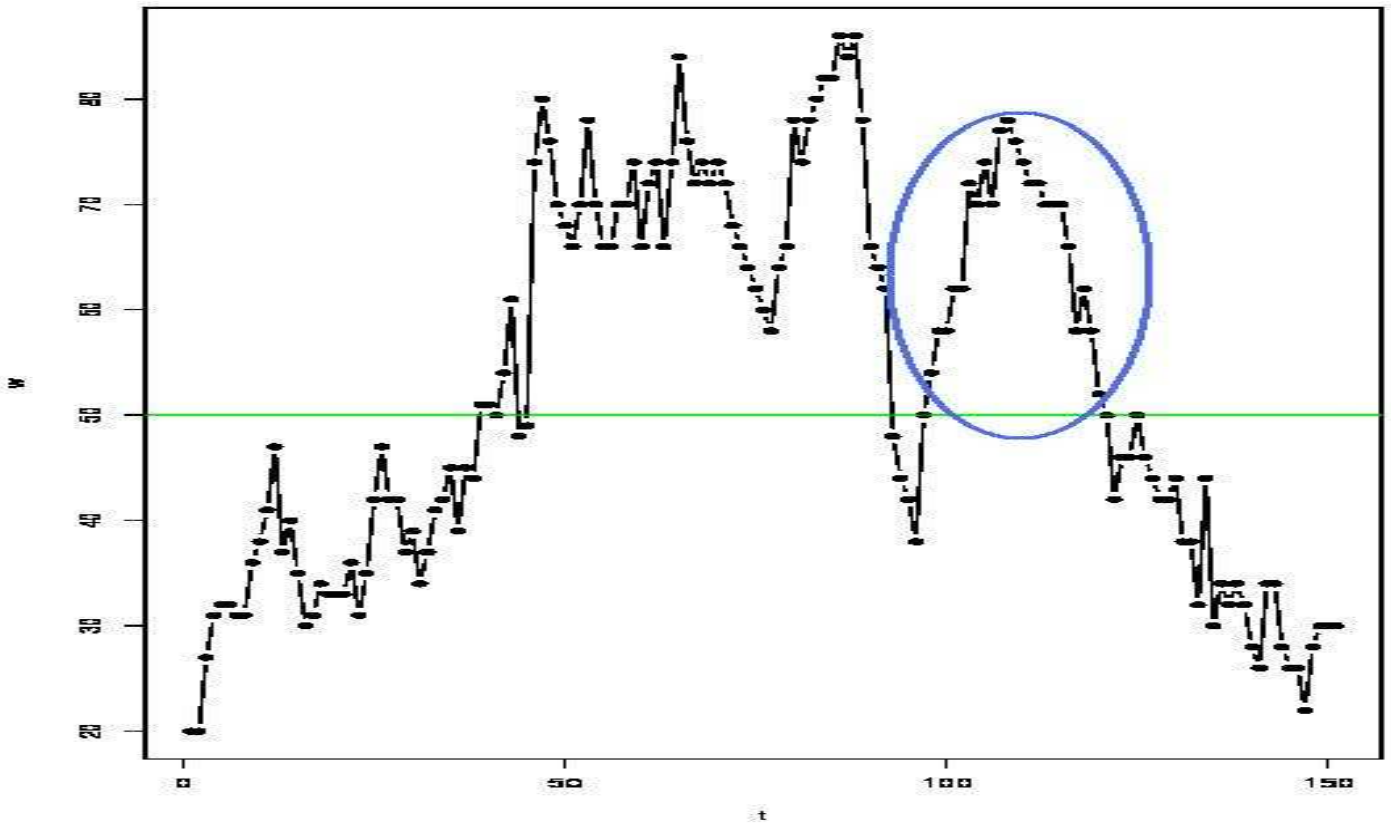
4.3.2 Declustering

Using $\kappa = 2$:



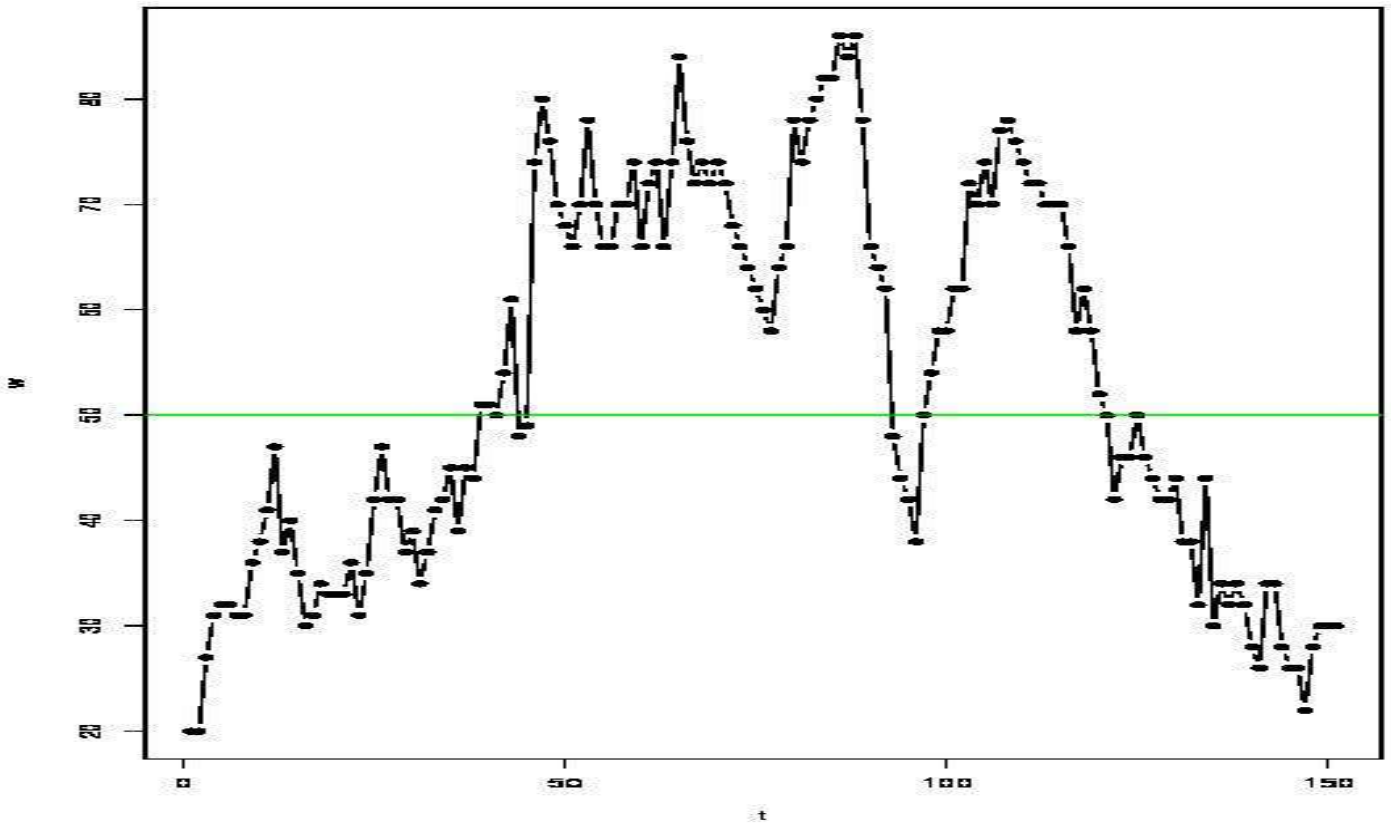
4.3.2 Declustering

Using $\kappa = 2$:



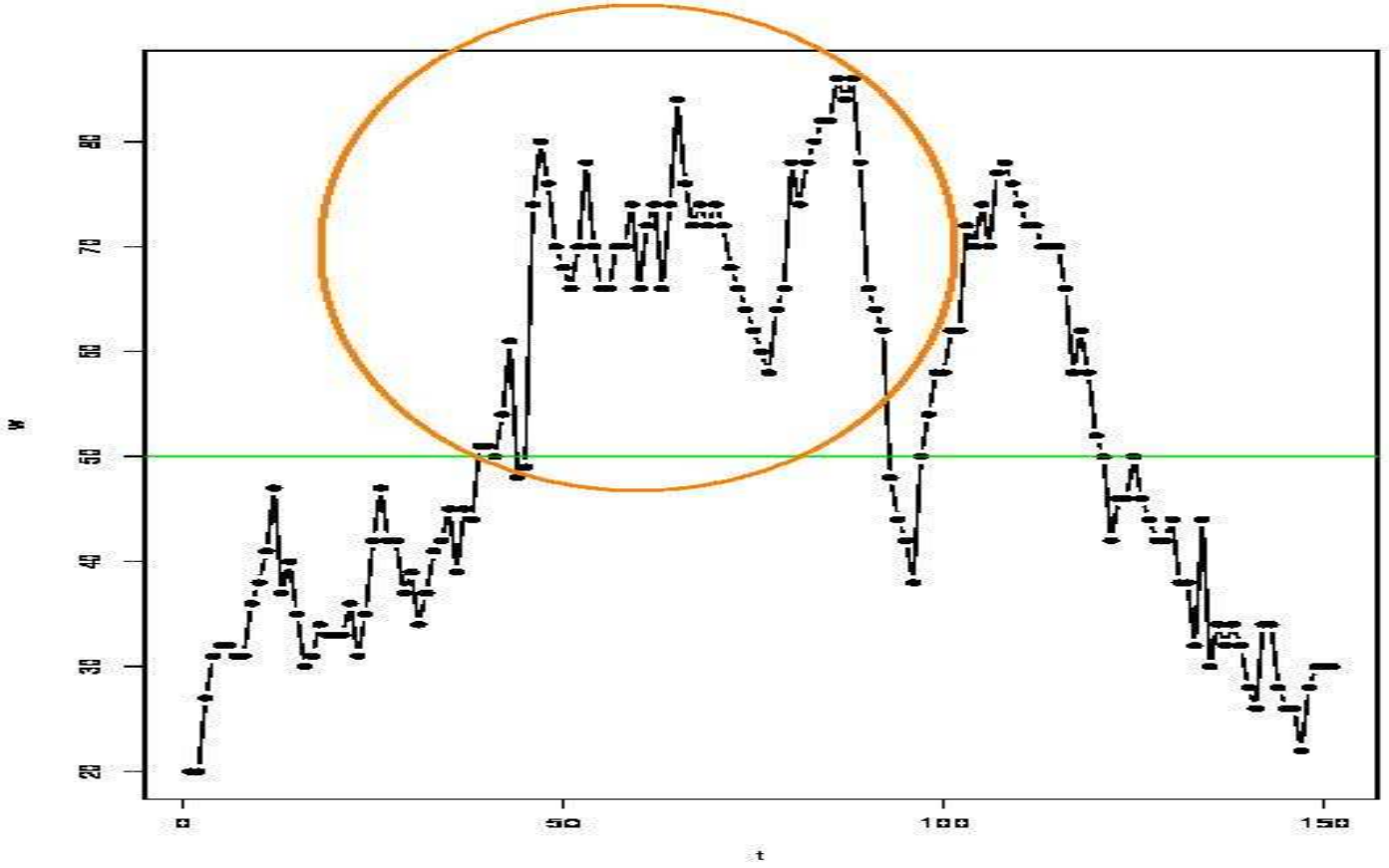
4.3.2 Declustering

Using $\kappa = 4$:



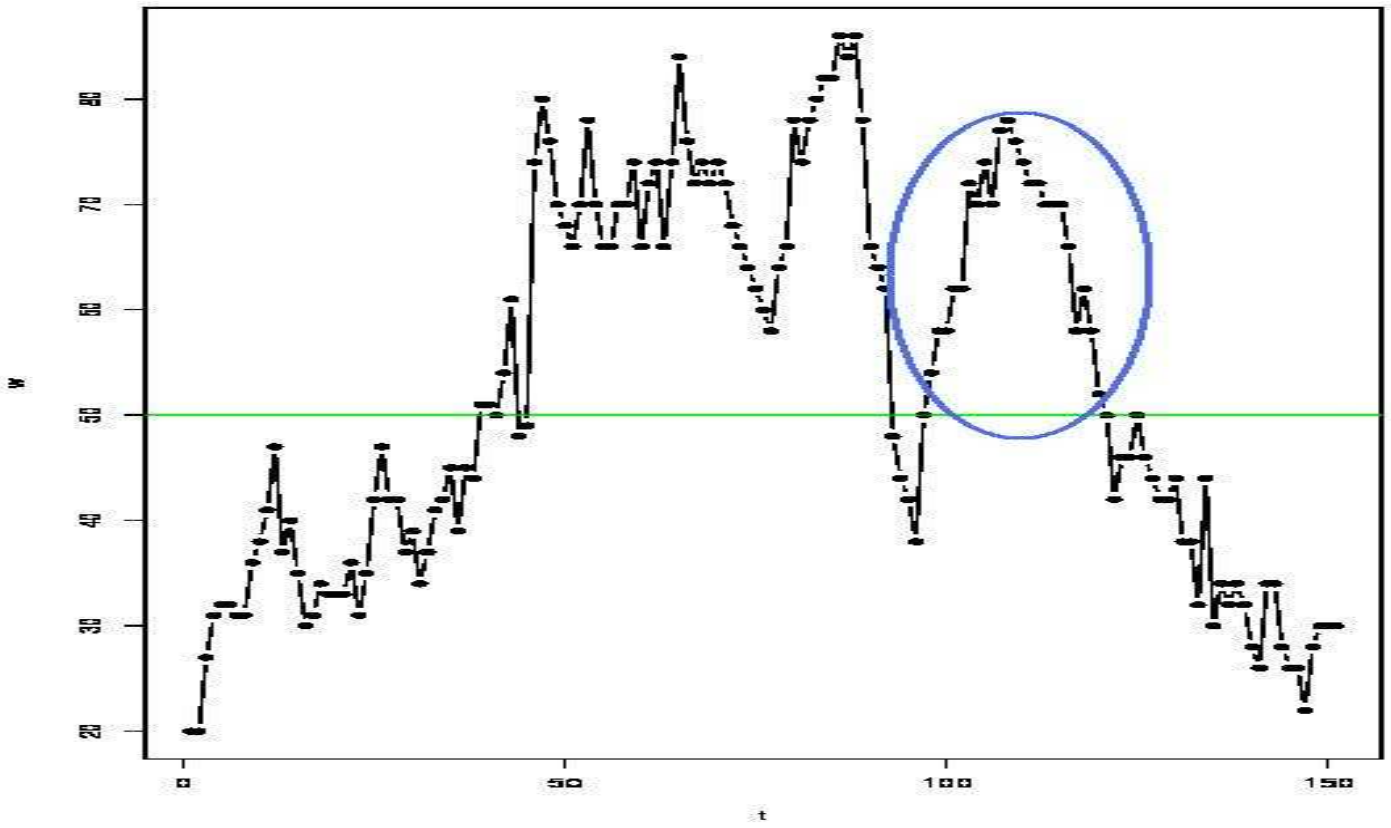
4.3.2 Declustering

Using $\kappa = 4$:



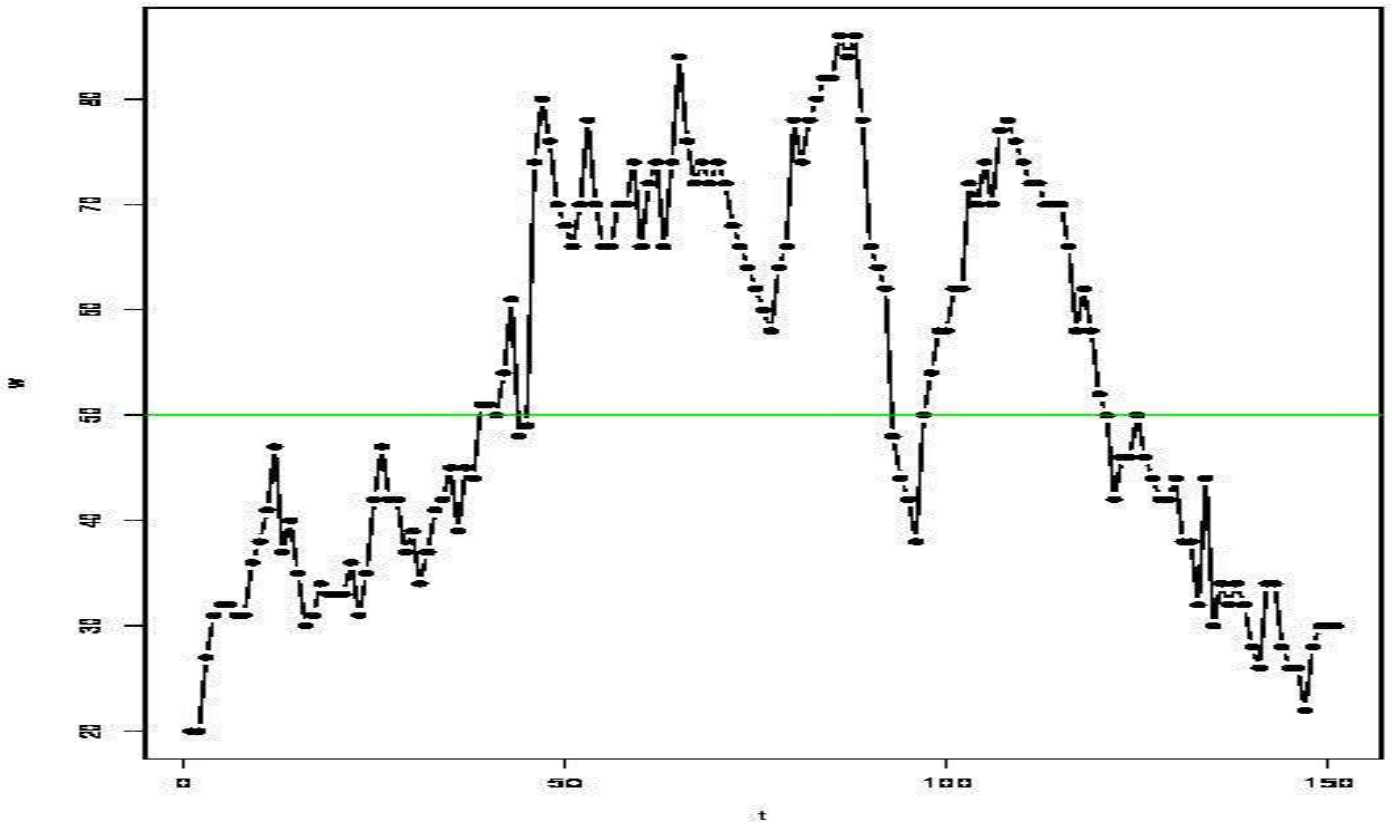
4.3.2 Declustering

Using $\kappa = 4$:



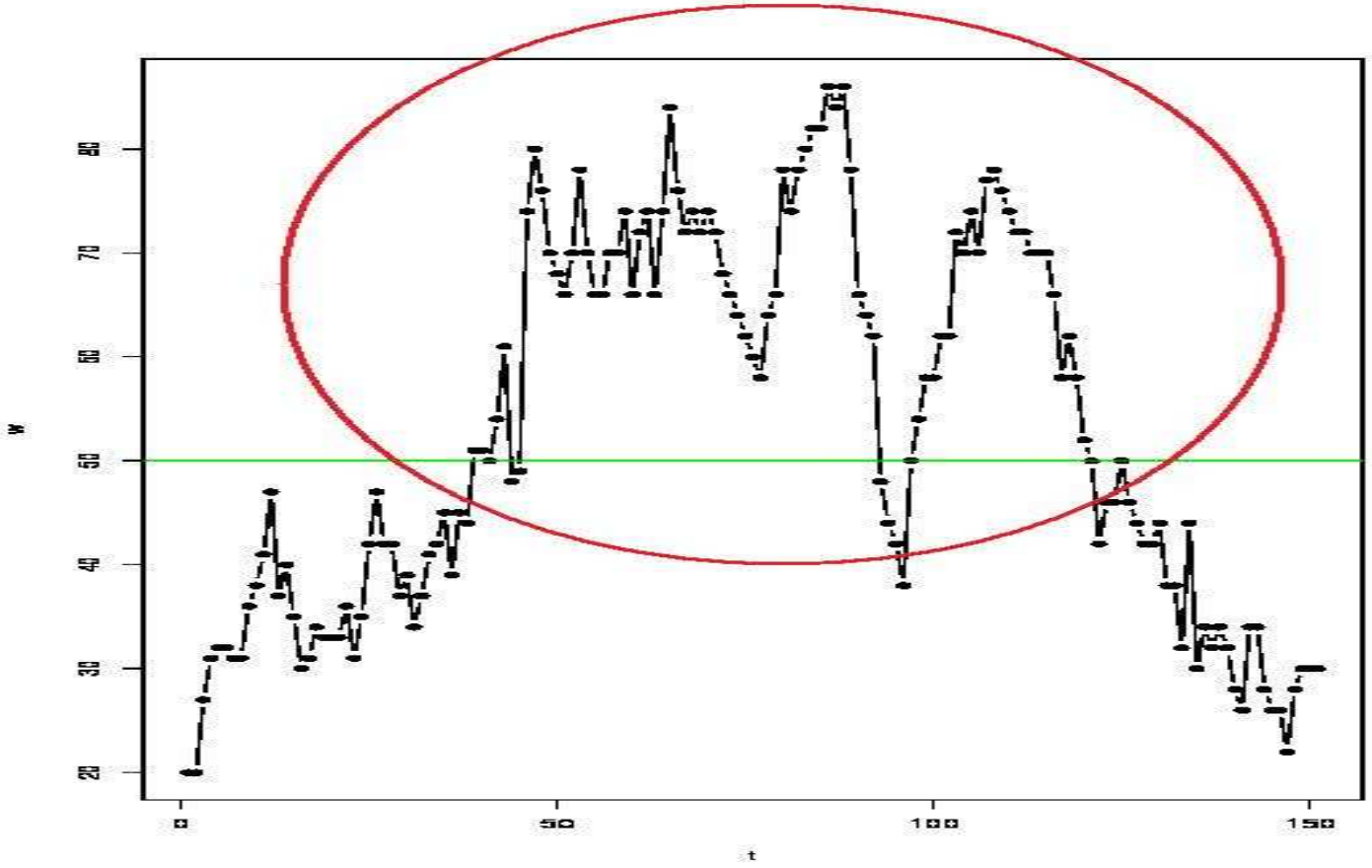
4.3.2 Declustering

Using $\kappa = 10$:



4.3.2 Declustering

Using $\kappa = 10$:



4.3.2 Declustering

Suppose we use $\kappa = 10$.

Obviously, we wouldn't want to identify clusters by hand for the full Bradfield wind speed series (recall that we have 10 years of hourly observations!).

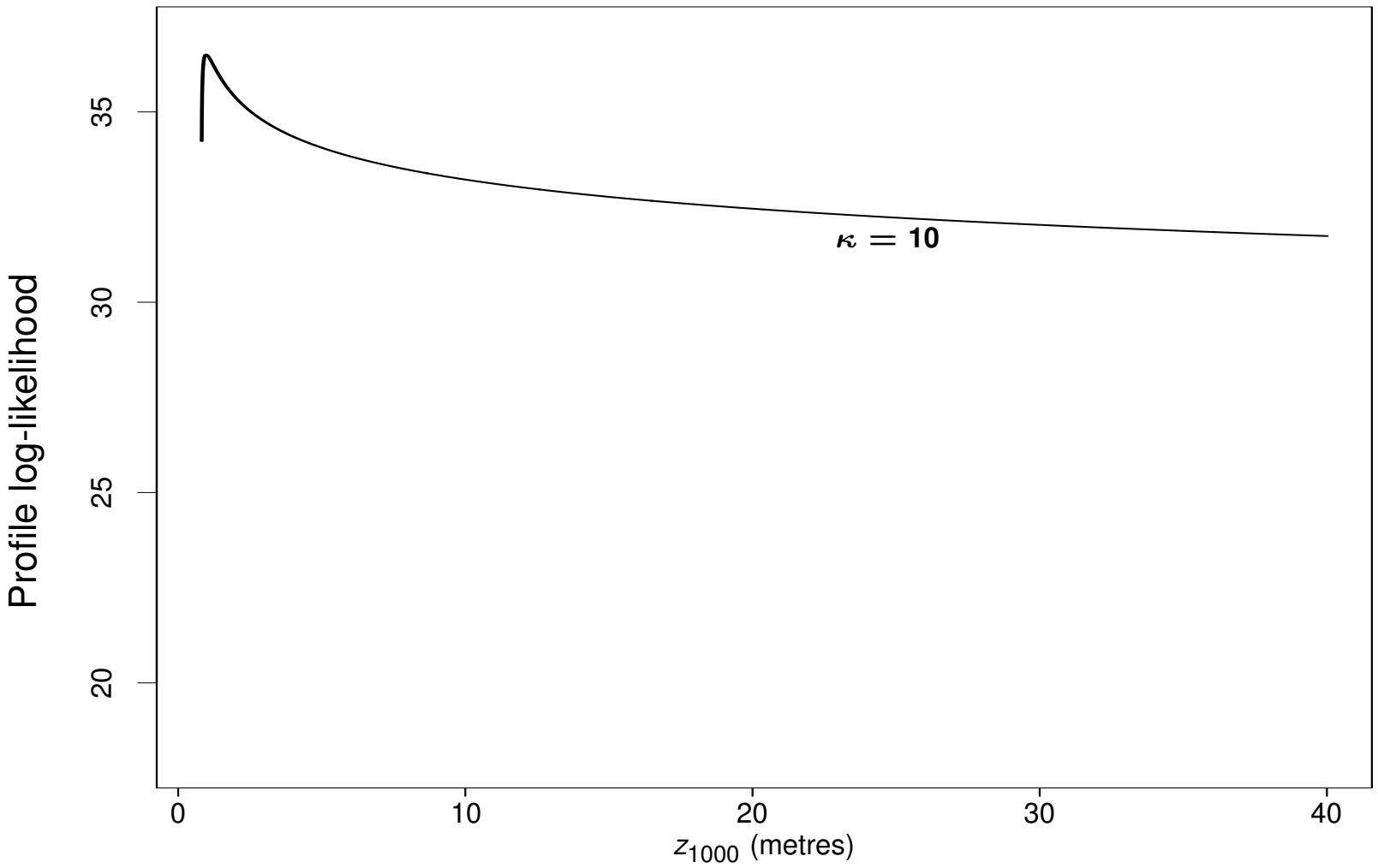
Unfortunately, there is no function in ismev to perform runs declustering.

I have written the following R code to perform this declustering, for $\kappa = 10$.

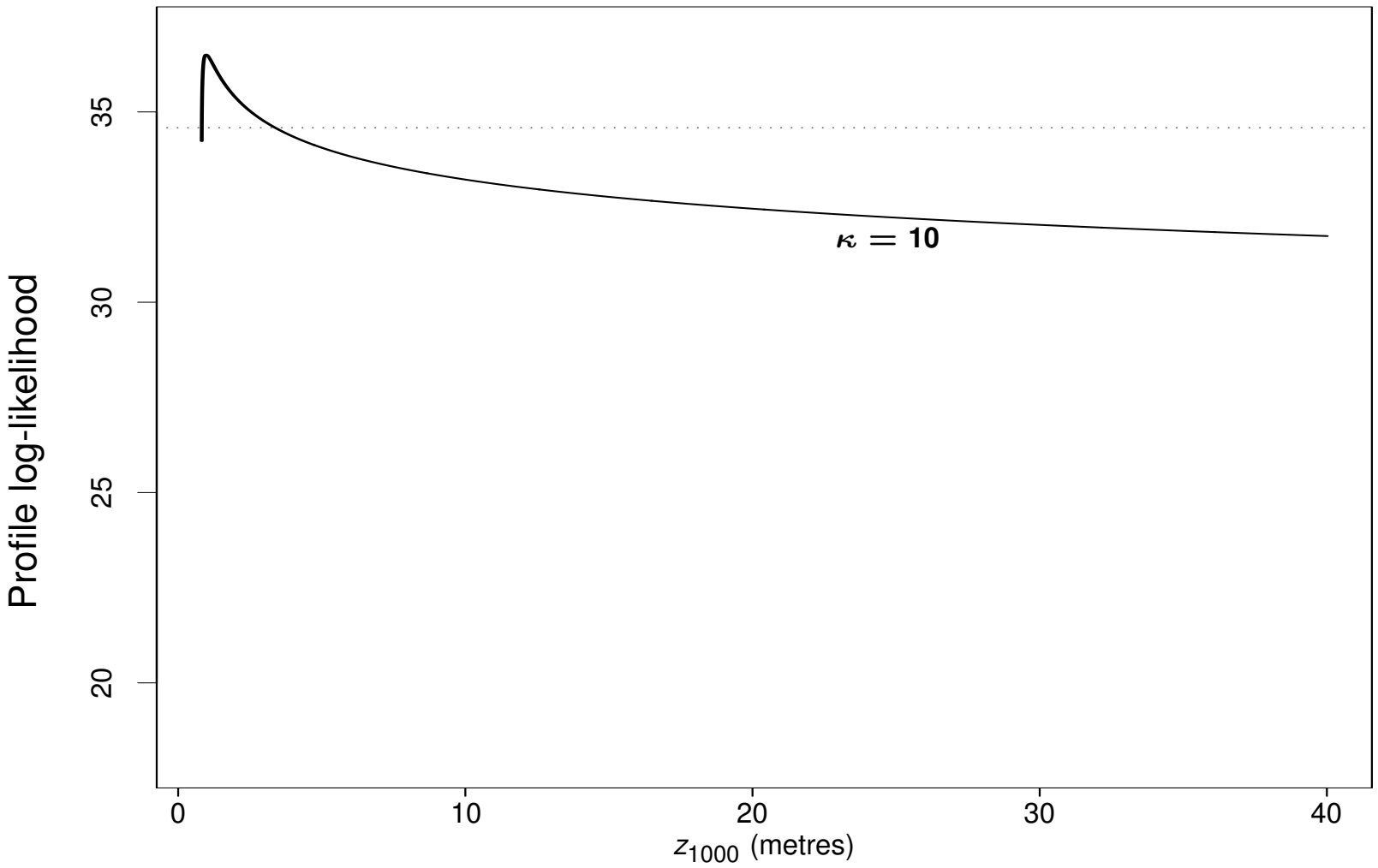
4.3.2 Declustering

Class demonstration in R

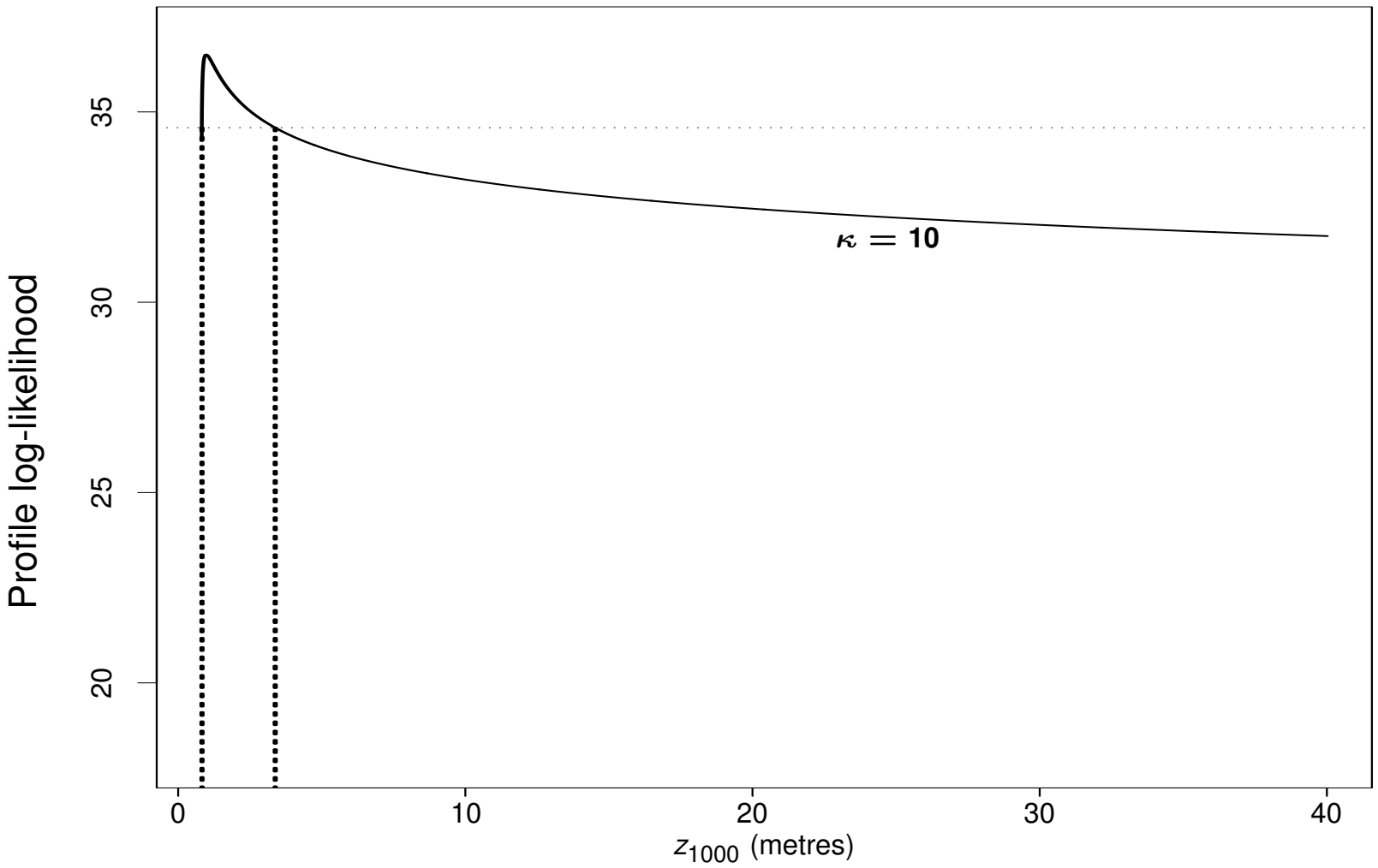
4.4 Words of warning (*Fawcett & Walshaw, 2012*)



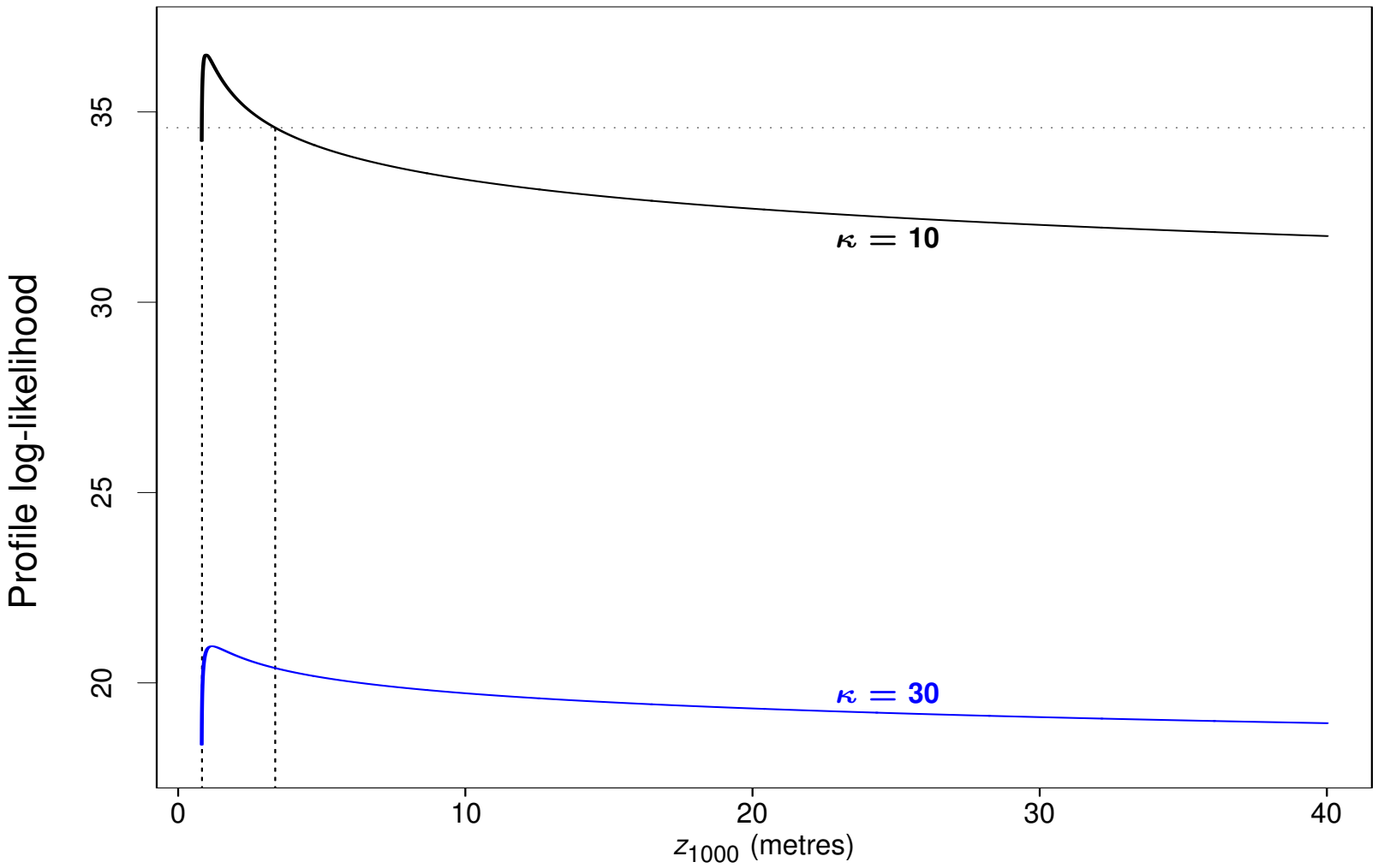
4.4 Words of warning (*Fawcett & Walshaw, 2012*)



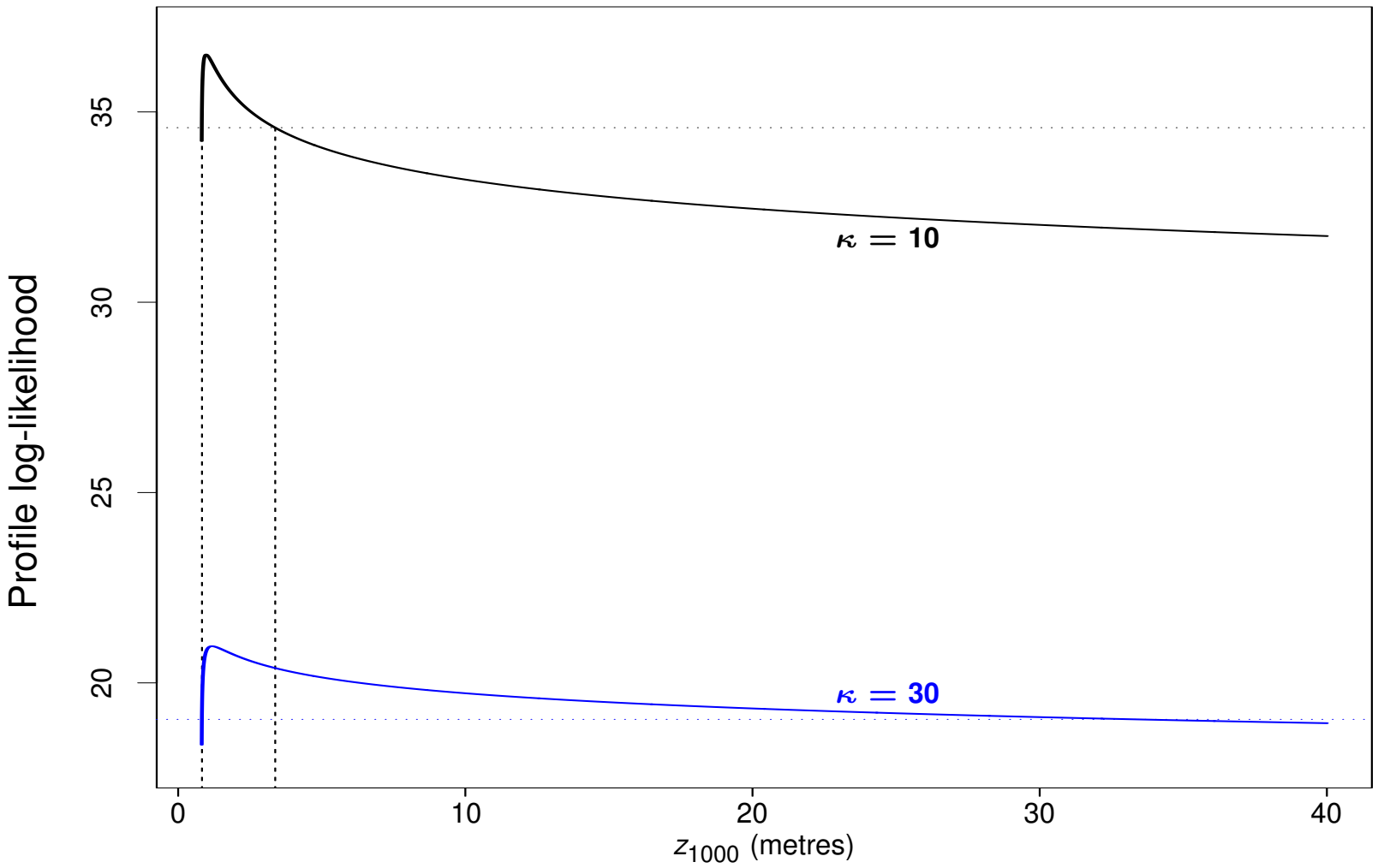
4.4 Words of warning (*Fawcett & Walshaw, 2012*)



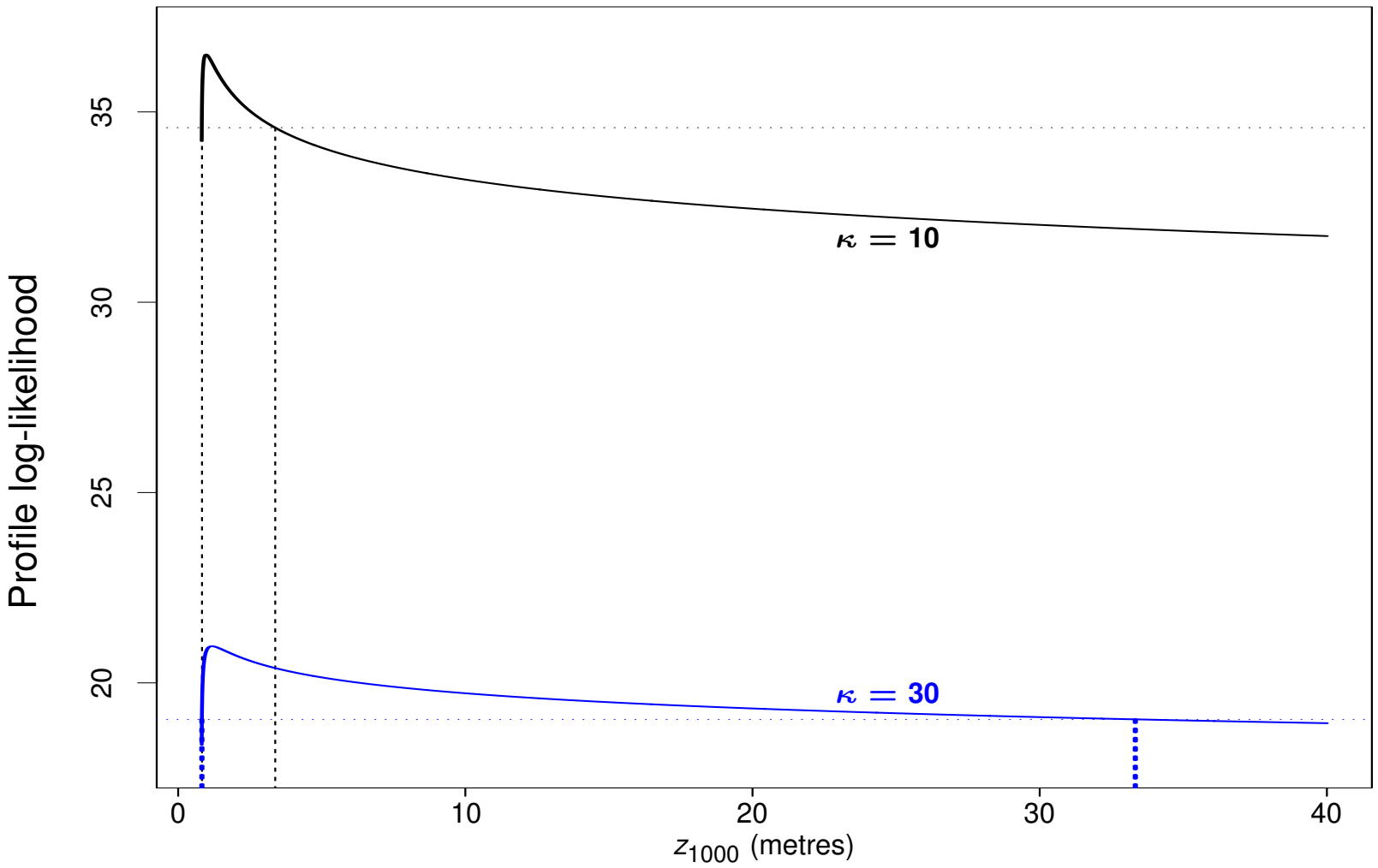
4.4 Words of warning (*Fawcett & Walshaw, 2012*)



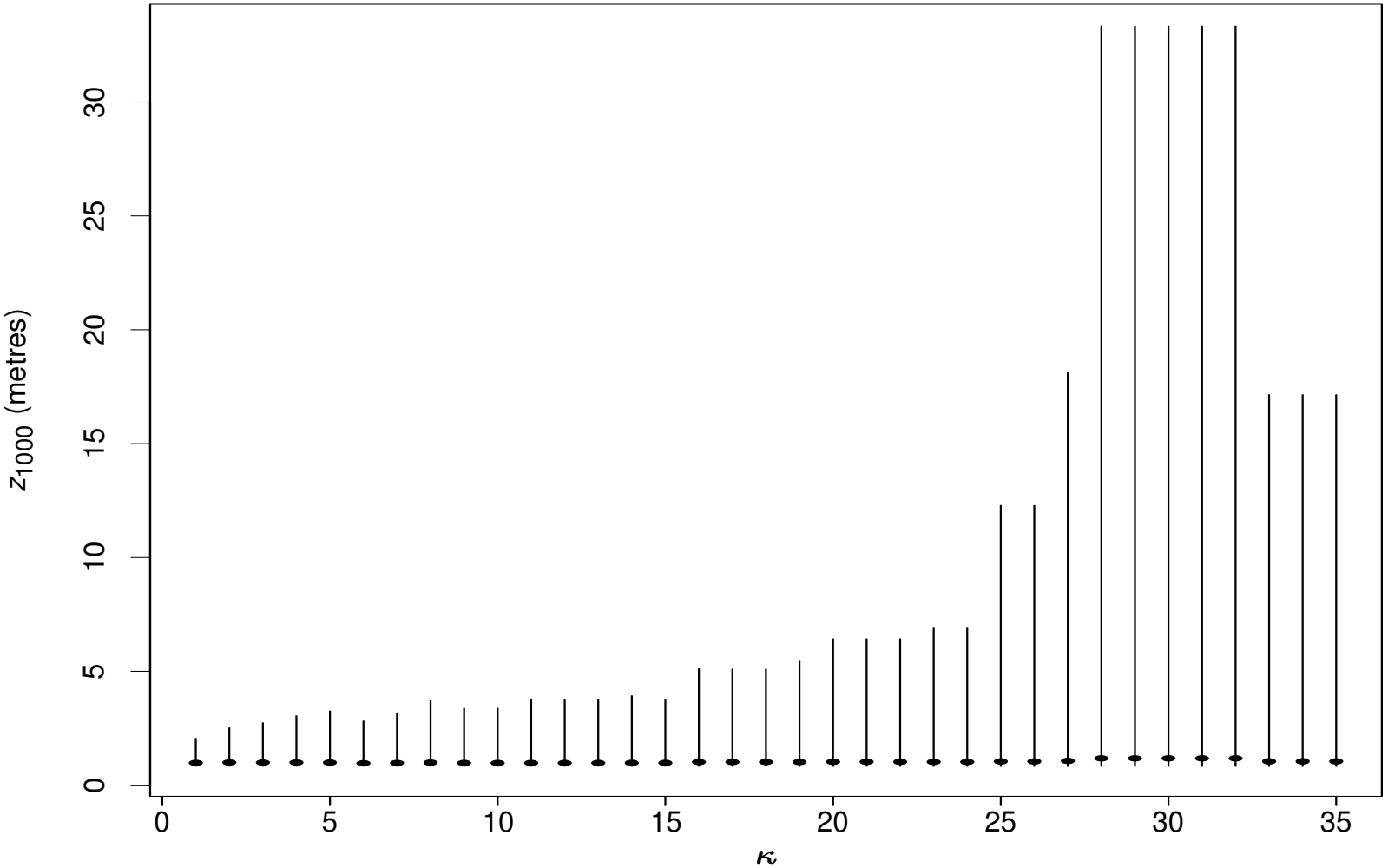
4.4 Words of warning (*Fawcett & Walshaw, 2012*)



4.4 Words of warning (*Fawcett & Walshaw, 2012*)



4.4 Words of warning (*Fawcett & Walshaw, 2012*)



4.4 Words of warning

- Similar to the block maxima approach, this is **wasteful of data**! We are throwing away observations we have identified as extreme!
- How do we choose κ? Estimates of GPD parameters – and return levels/confidence intervals for return levels – can be sensitive to the choice of κ (often hugely so)
- What if we are interested in the dependence?

4.4 Words of warning

- Similar to the block maxima approach, this is **wasteful of data**! We are throwing away observations we have identified as extreme!
- How do we choose κ? Estimates of GPD parameters – and return levels/confidence intervals for return levels – can be sensitive to the choice of κ (often hugely so)
- What if we are interested in the dependence?

4.4 Words of warning

- Similar to the block maxima approach, this is **wasteful of data**! We are throwing away observations we have identified as extreme!
- How do we choose κ? Estimates of GPD parameters – and return levels/confidence intervals for return levels – can be sensitive to the choice of κ (often hugely so)
- What if we are interested in the dependence?

Welcome back!

- Assignments
- Mid–semester test
  - This Thursday, **prompt 4.05pm** start
  - Specimen, with solutions, available online
  - Can cover anything from **Chapters 1–4**
  - Key formulae given on paper (and exam paper in May/June): CDF for GEV and GPD
  - Tips: should know the **delta method**; should know the link between the **Hessian matrix**, the **information matrix** and the **variance–covariance matrix**; should be familiar with R output for things like `gev.fit` and `gpd.fit`; worthwhile memorising formulae for **return levels**

Welcome back!

- Assignments
- Mid–semester test
  - This Thursday, **prompt 4.05pm** start
  - Specimen, with solutions, available online
  - Can cover anything from **Chapters 1–4**
  - Key formulae given on paper (and exam paper in May/June): CDF for GEV and GPD
  - Tips: should know the **delta method**; should know the link between the **Hessian matrix**, the **information matrix** and the **variance–covariance matrix**; should be familiar with R output for things like `gev.fit` and `gpd.fit`; worthwhile memorising formulae for **return levels**

Welcome back!

- Assignments
- Mid–semester test
  - This Thursday, **prompt 4.05pm** start
  - Specimen, with solutions, available online
  - Can cover anything from **Chapters 1–4**
  - Key formulae given on paper (and exam paper in May/June): CDF for GEV and GPD
  - Tips: should know the **delta method**; should know the link between the **Hessian matrix**, the **information matrix** and the **variance–covariance matrix**; should be familiar with R output for things like `gev.fit` and `gpd.fit`; worthwhile memorising formulae for **return levels**

Welcome back!

- Assignments
- Mid–semester test
  - This Thursday, **prompt 4.05pm** start
  - Specimen, with solutions, available online
  - Can cover anything from **Chapters 1–4**
  - Key formulae given on paper (and exam paper in May/June): CDF for GEV and GPD
  - Tips: should know the **delta method**; should know the link between the **Hessian matrix**, the **information matrix** and the **variance–covariance matrix**; should be familiar with R output for things like `gev.fit` and `gpd.fit`; worthwhile memorising formulae for **return levels**

Welcome back!

- Assignments
- Mid–semester test
  - This Thursday, **prompt 4.05pm** start
  - Specimen, with solutions, available online
  - Can cover anything from **Chapters 1–4**
  - Key formulae given on paper (and exam paper in May/June): CDF for GEV and GPD
  - Tips: should know the **delta method**; should know the link between the **Hessian matrix**, the **information matrix** and the **variance–covariance matrix**; should be familiar with R output for things like `gev.fit` and `gpd.fit`; worthwhile memorising formulae for **return levels**

Welcome back!

- Assignments
- Mid–semester test
  - This Thursday, **prompt 4.05pm** start
  - Specimen, with solutions, available online
  - Can cover anything from **Chapters 1–4**
  - Key formulae given on paper (and exam paper in May/June): CDF for GEV and GPD
  - Tips: should know the **delta method**; should know the link between the **Hessian matrix**, the **information matrix** and the **variance–covariance matrix**; should be familiar with R output for things like `gev.fit` and `gpd.fit`; worthwhile memorising formulae for **return levels**

Welcome back!

- Assignments
- Mid–semester test
  - This Thursday, **prompt 4.05pm** start
  - Specimen, with solutions, available online
  - Can cover anything from **Chapters 1–4**
  - Key formulae given on paper (and exam paper in May/June): CDF for GEV and GPD
  - Tips: should know the **delta method**; should know the link between the **Hessian matrix**, the **information matrix** and the **variance–covariance matrix**; should be familiar with R output for things like `gev.fit` and `gpd.fit`; worthwhile memorising formulae for **return levels**

Welcome back!

- Project
  - Due in 4pm, **Friday 10th May**
  - Based on your own unique dataset, and the paper I gave out before Easter
- Next few weeks
  - Chapter 5: non–stationarity (e.g. trend and/or seasonality)
  - Chapter 6: Multivariate considerations

Welcome back!

- Project
  - Due in 4pm, **Friday 10th May**
  - Based on your own unique dataset, and the paper I gave out before Easter
- Next few weeks
  - Chapter 5: non–stationarity (e.g. trend and/or seasonality)
  - Chapter 6: Multivariate considerations

Welcome back!

- Project
  - Due in 4pm, **Friday 10th May**
  - Based on your own unique dataset, and the paper I gave out before Easter
- Next few weeks
  - Chapter 5: non–stationarity (e.g. trend and/or seasonality)
  - Chapter 6: Multivariate considerations

Welcome back!

- Project
  - Due in 4pm, **Friday 10th May**
  - Based on your own unique dataset, and the paper I gave out before Easter
- Next few weeks
  - Chapter 5: non–stationarity (e.g. trend and/or seasonality)
  - Chapter 6: Multivariate considerations

Welcome back!

- Project
  - Due in 4pm, **Friday 10th May**
  - Based on your own unique dataset, and the paper I gave out before Easter
- Next few weeks
  - Chapter 5: non–stationarity (e.g. trend and/or seasonality)
  - Chapter 6: Multivariate considerations

Welcome back!

- Project
  - Due in 4pm, **Friday 10th May**
  - Based on your own unique dataset, and the paper I gave out before Easter
- Next few weeks
  - Chapter 5: non–stationarity (e.g. trend and/or seasonality)
  - Chapter 6: Multivariate considerations

Chapter 5

Non–stationary extremes

5.1 Introduction

In the context of environmental processes, it is common to observe **non–stationarity**:

- different **seasons** having different climate patterns
- **long term trends** owing to climate change

The models from chapters 2 and 3 assume that the observations used are IID!

5.1 Introduction

In the context of environmental processes, it is common to observe **non–stationarity**:

- different **seasons** having different climate patterns
- **long term trends** owing to climate change

The models from chapters 2 and 3 assume that the observations used are IID!

5.1 Introduction

In the context of environmental processes, it is common to observe **non–stationarity**:

- different **seasons** having different climate patterns
- **long term trends** owing to climate change

The models from chapters 2 and 3 assume that the observations used are IID!

5.1 Introduction

In the context of environmental processes, it is common to observe **non–stationarity**:

- different **seasons** having different climate patterns
- **long term trends** owing to climate change

The models from chapters 2 and 3 assume that the observations used are IID!

5.1 Introduction

We examined the effects of **dependence** in Chapter 4:

- Likely that consecutive extremes of wind or rain will be **correlated**
- Block maxima: Doesn't matter – provided we can assume long–range independence (Leadbetter's $D(u_n)$ condition), then the GEV still appropriate
- Threshold exceedances: Not quite so straightforward, although we can **decluster**, and fit the GPD to **cluster peak excesses**? Problems?

5.1 Introduction

We examined the effects of **dependence** in Chapter 4:

- Likely that consecutive extremes of wind or rain will be **correlated**
- Block maxima: Doesn't matter – provided we can assume long–range independence (Leadbetter's $D(u_n)$ condition), then the GEV still appropriate
- Threshold exceedances: Not quite so straightforward, although we can **decluster**, and fit the GPD to **cluster peak excesses**? Problems?

5.1 Introduction

We examined the effects of **dependence** in Chapter 4:

- Likely that consecutive extremes of wind or rain will be **correlated**
- Block maxima: Doesn't matter – provided we can assume long–range independence (Leadbetter's $D(u_n)$ condition), then the GEV still appropriate
- Threshold exceedances: Not quite so straightforward, although we can **decluster**, and fit the GPD to **cluster peak excesses**? Problems?

5.1 Introduction

We examined the effects of **dependence** in Chapter 4:

- Likely that consecutive extremes of wind or rain will be **correlated**
- Block maxima: Doesn't matter – provided we can assume long–range independence (Leadbetter's $D(u_n)$ condition), then the GEV still appropriate
- Threshold exceedances: Not quite so straightforward, although we can **decluster**, and fit the GPD to **cluster peak excesses**? Problems?

5.1 Introduction

To date, no general theory for **non–stationary extremes** has been established.

In practice, it is common to adopt pragmatic ‘workarounds’ based on the type of non–stationarity observed.

For this reason, in this Chapter we will give some specific examples of how practitioners have dealt with non–stationarity in recent work and publications.

5.1 Introduction

To date, no general theory for **non–stationary extremes** has been established.

In practice, it is common to adopt pragmatic ‘workarounds’ based on the type of non–stationarity observed.

For this reason, in this Chapter we will give some specific examples of how practitioners have dealt with non–stationarity in recent work and publications.

5.1 Introduction

To date, no general theory for **non–stationary extremes** has been established.

In practice, it is common to adopt pragmatic ‘workarounds’ based on the type of non–stationarity observed.

For this reason, in this Chapter we will give some specific examples of how practitioners have dealt with non–stationarity in recent work and publications.

5.2 Annual maximum sea levels in Venice

Recall **question 4** in **problems sheet 2**, which investigated the use of the r–largest order statistics model for extreme sea levels at Venice (1961–2011).

These data are available to download from the course webpage.

5.2 Annual maximum sea levels in Venice

Recall **question 4** in **problems sheet 2**, which investigated the use of the r–largest order statistics model for extreme sea levels at Venice (1961–2011).

These data are available to download from the course webpage.

5.2 Annual maximum sea levels in Venice

Demonstration in R

5.2 Annual maximum sea levels in Venice

One way of capturing the trend observed in the Venice annual maxima is to allow the GEV location parameter μ to vary across time.

From Figure 5.1, a **simple linear trend** in time seems plausible for our annual maximum sea levels X, and so we could use the model

$$X_t \sim \text{GEV}(\mu(t), \sigma, \xi),$$

where

$$\mu(t) = \beta_0 + \beta_1 t \tag{5.1}$$

and t is an indicator of year.

5.2 Annual maximum sea levels in Venice

One way of capturing the trend observed in the Venice annual maxima is to allow the GEV location parameter μ to vary across time.

From Figure 5.1, a **simple linear trend** in time seems plausible for our annual maximum sea levels X, and so we could use the model

$$X_t \sim \text{GEV}(\mu(t), \sigma, \xi),$$

where

$$\mu(t) = \beta_0 + \beta_1 t \tag{5.1}$$

and t is an indicator of year.

5.2 Annual maximum sea levels in Venice

One way of capturing the trend observed in the Venice annual maxima is to allow the GEV location parameter μ to vary across time.

From Figure 5.1, a **simple linear trend** in time seems plausible for our annual maximum sea levels X, and so we could use the model

$$X_t \sim \text{GEV}(\mu(t), \sigma, \xi),$$

where

$$\mu(t) = \beta_0 + \beta_1 t \tag{5.1}$$

and t is an indicator of year.

5.2 Annual maximum sea levels in Venice

$$\mu(t) = \beta_0 + \beta_1 t :$$

- Variations over time modelled as a linear trend in the **location parameter**
- β_1 represents the **slope** – in this case, the annual rate of change in sea–level maxima at Venice
- The time–homogeneous model is a special case of this time dependent model, with $\beta_1 = 0$...
- ...since this is **nested** within the model which allows for a time dependence, the deviance statistic can be used to formally compare models (see later)

5.2 Annual maximum sea levels in Venice

$$\mu(t) = \beta_0 + \beta_1 t :$$

- Variations over time modelled as a linear trend in the **location parameter**
- β_1 represents the **slope** – in this case, the annual rate of change in sea–level maxima at Venice
- The time–homogeneous model is a special case of this time dependent model, with $\beta_1 = 0$...
- ...since this is **nested** within the model which allows for a time dependence, the deviance statistic can be used to formally compare models (see later)

5.2 Annual maximum sea levels in Venice

$$\mu(t) = \beta_0 + \beta_1 t :$$

- Variations over time modelled as a linear trend in the **location parameter**
- β_1 represents the **slope** – in this case, the annual rate of change in sea–level maxima at Venice
- The time–homogeneous model is a special case of this time dependent model, with $\beta_1 = 0$...
- ...since this is **nested** within the model which allows for a time dependence, the deviance statistic can be used to formally compare models (see later)

5.2 Annual maximum sea levels in Venice

$$\mu(t) = \beta_0 + \beta_1 t :$$

- Variations over time modelled as a linear trend in the **location parameter**
- β_1 represents the **slope** – in this case, the annual rate of change in sea–level maxima at Venice
- The time–homogeneous model is a special case of this time dependent model, with $\beta_1 = 0$...
- ...since this is **nested** within the model which allows for a time dependence, the deviance statistic can be used to formally compare models (see later)

5.2.1 Parameter estimation

Recall the log–likelihood function for the GEV:

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

where m is the number of block maxima $x_1, x_2, \ldots, x_m$.

5.2.1 Parameter estimation

We simply replace μ in the above expression with equation (5.1), giving

$$\begin{aligned}\ell(\beta_0, \beta_1, \sigma, \xi; \boldsymbol{x}, t) &= -m\log\sigma \\ &\quad -(1/\xi + 1)\sum_{i=1}^{m} \log\left[1 + \xi\left(\frac{x_i - (\beta_0 + \beta_1 t)}{\sigma}\right)\right]_+ \\ &\quad -\sum_{i=1}^{m}\left[1 + \xi\left(\frac{x_i - (\beta_0 + \beta_1 t)}{\sigma}\right)\right]_+^{-1/\xi},\end{aligned}$$

with the usual replacement when $\xi = 0$.

5.2.1 Parameter estimation

- *Could* maximise $\ell(\beta_0, \beta_1, \sigma, \xi; \boldsymbol{x}, t)$ 'from first principles' using `nlm`
- However, easier to do in the ismev package!

Demonstration in R

5.2.1 Parameter estimation

- *Could* maximise $\ell(\beta_0, \beta_1, \sigma, \xi; \boldsymbol{x}, t)$ 'from first principles' using `nlm`
- However, easier to do in the ismev package!

Demonstration in R

5.2.1 Parameter estimation

Thus, we have

$$\hat{\beta}_0 = 96.986(4.249) \quad \hat{\beta}_1 = 0.564(0.139)$$

$$\hat{\sigma} = 14.584(1.578) \quad \hat{\xi} = -0.027(0.083)$$

(with standard errors in parentheses).

This gives:

$$\hat{\mu}(t) = 96.986 + 0.564t,$$

giving an estimated increase in maximum sea levels at Venice of about 0.564cm per year.

5.2.1 Parameter estimation

Thus, we have

$$\hat{\beta}_0 = 96.986(4.249) \quad \hat{\beta}_1 = 0.564(0.139)$$

$$\hat{\sigma} = 14.584(1.578) \quad \hat{\xi} = -0.027(0.083)$$

(with standard errors in parentheses).

This gives:

$$\hat{\mu}(t) = 96.986 + 0.564t,$$

giving an estimated increase in maximum sea levels at Venice of about 0.564cm per year.

5.2.1 Parameter estimation

For example, the estimated value for μ in the year 2013 would be

$$\hat{\mu}(53) = 96.986 + 0.564 \times 13 = 116.878;$$

we could, of course, use the delta method to obtain the corresponding standard error.

Using this simple linear model, we can estimate μ for $t = 1, 2, \ldots, 51$ to cover the years for which we have data (i.e. 1961, 1962, ..., 2011).

5.2.1 Parameter estimation

For example, the estimated value for μ in the year 2013 would be

$$\hat{\mu}(53) = 96.986 + 0.564 \times 13 = 116.878;$$

we could, of course, use the delta method to obtain the corresponding standard error.

Using this simple linear model, we can estimate μ for $t = 1, 2, \ldots, 51$ to cover the years for which we have data (i.e. 1961, 1962, ..., 2011).

5.2.1 Parameter estimation

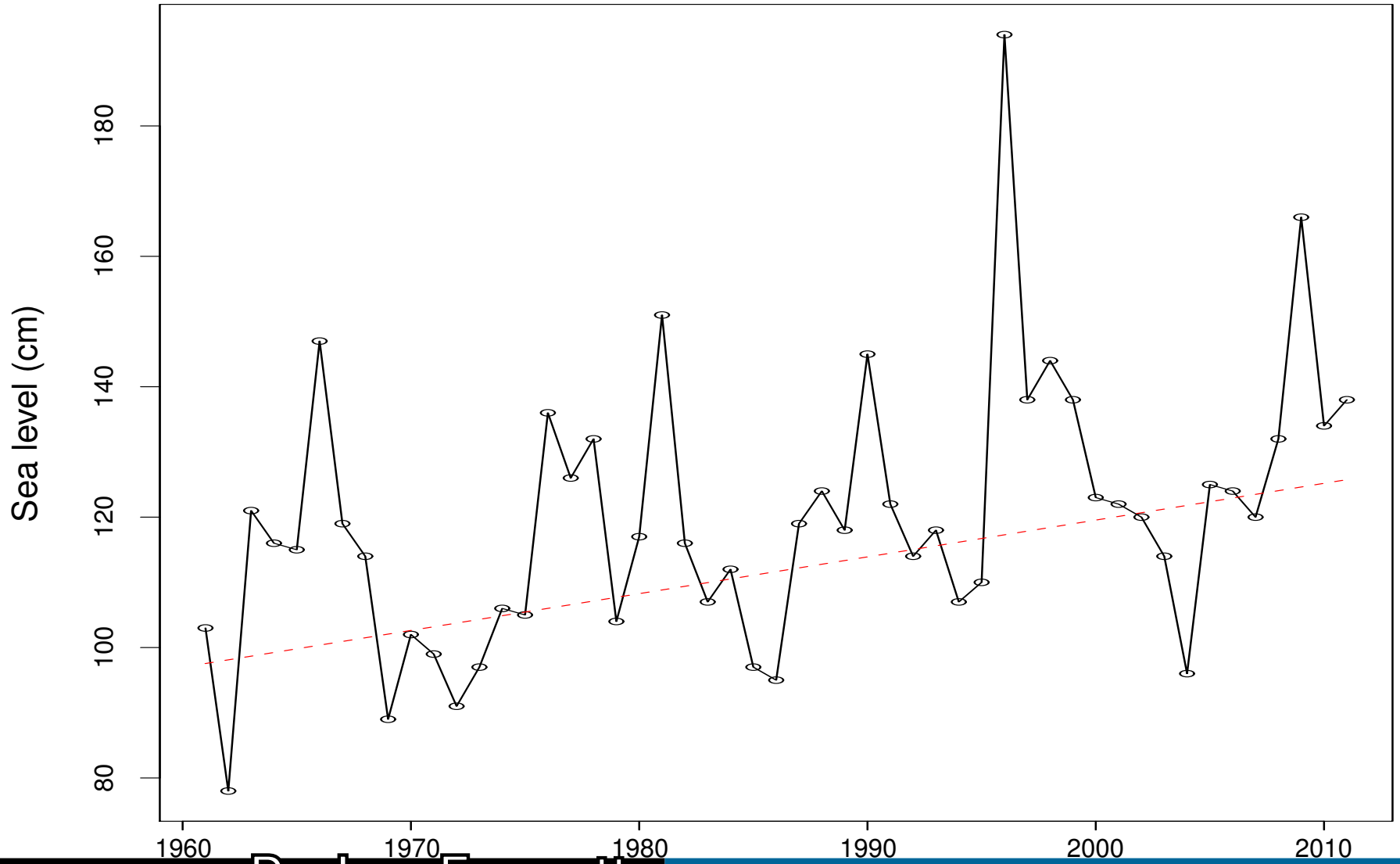
For example, the estimated value for μ in the year 2013 would be

$$\hat{\mu}(53) = 96.986 + 0.564 \times 13 = 116.878;$$

we could, of course, use the delta method to obtain the corresponding standard error.

Using this simple linear model, we can estimate μ for $t = 1, 2, \ldots, 51$ to cover the years for which we have data (i.e. 1961, 1962, . . . , 2011).

5.2.1 Parameter estimation



5.2.2 Model choice

Fitting a completely stationary model to the set of annual maxima, as you did in question 4(a) of problems sheet 2, gives:

```
gev.fit(venice.anmax)

$conv
[1] 0

$nllh
[1] 222.7145

$mle
[1] 111.09925486 17.17548761 -0.07673265

$se
[1] 2.6280071 1.8033672 0.0735214
```

5.2.2 Model choice

Questions:

- Is the non–stationary model **worthwhile?**
- Is the trend we observe in Figure 5.1 **significant**?
- In other words, does the non–stationary model provide an **improvement** in fit over the simpler model shown here?

5.2.2 Model choice

Questions:

- Is the non–stationary model **worthwhile?**
- Is the trend we observe in Figure 5.1 **significant**?
- In other words, does the non–stationary model provide an **improvement** in fit over the simpler model shown here?

5.2.2 Model choice

Questions:

- Is the non–stationary model **worthwhile?**
- Is the trend we observe in Figure 5.1 **significant**?
- In other words, does the non–stationary model provide an **improvement** in fit over the simpler model shown here?

5.2.2 Model choice

Questions:

- Is the non–stationary model **worthwhile?**
- Is the trend we observe in Figure 5.1 **significant**?
- In other words, does the non–stationary model provide an **improvement** in fit over the simpler model shown here?

5.2.2 Model choice

We can use a version of the result in Section 2.4 (page 41) to address this question.

Generally, maximum likelihood estimation of nested models leads to a simple test procedure of one model against the other.

With models $\mathcal{M}_0 \subset \mathcal{M}_1$, we define the deviance statistic as

$$D = 2\left\{\ell_1(\mathcal{M}_1) - \ell_0(\mathcal{M}_0)\right\},$$

where $\ell_1(\mathcal{M}_1)$ and $\ell_0(\mathcal{M}_0)$ are the maximised log–likelihood under models $\mathcal{M}_1$ and $\mathcal{M}_0$ respectively.

Asymptotically, $D \sim \chi^2_k$, where $k = \dim(\mathcal{M}_1) - \dim(\mathcal{M}_0)$.

5.2.2 Model choice

We can use a version of the result in Section 2.4 (page 41) to address this question.

Generally, maximum likelihood estimation of nested models leads to a simple test procedure of one model against the other.

With models $\mathcal{M}_0 \subset \mathcal{M}_1$, we define the deviance statistic as

$$D = 2\left\{\ell_1(\mathcal{M}_1) - \ell_0(\mathcal{M}_0)\right\},$$

where $\ell_1(\mathcal{M}_1)$ and $\ell_0(\mathcal{M}_0)$ are the maximised log–likelihood under models $\mathcal{M}_1$ and $\mathcal{M}_0$ respectively.

Asymptotically, $D \sim \chi^2_k$, where $k = \dim(\mathcal{M}_1) - \dim(\mathcal{M}_0)$.

5.2.2 Model choice

We can use a version of the result in Section 2.4 (page 41) to address this question.

Generally, maximum likelihood estimation of nested models leads to a simple test procedure of one model against the other.

With models $\mathcal{M}_0 \subset \mathcal{M}_1$, we define the deviance statistic as

$$D = 2\left\{\ell_1(\mathcal{M}_1) - \ell_0(\mathcal{M}_0)\right\},$$

where $\ell_1(\mathcal{M}_1)$ and $\ell_0(\mathcal{M}_0)$ are the maximised log–likelihood under models $\mathcal{M}_1$ and $\mathcal{M}_0$ respectively.

Asymptotically, $D \sim \chi^2_k$, where $k = \dim(\mathcal{M}_1) - \dim(\mathcal{M}_0)$.

5.2.2 Model choice

We can use a version of the result in Section 2.4 (page 41) to address this question.

Generally, maximum likelihood estimation of nested models leads to a simple test procedure of one model against the other.

With models $\mathcal{M}_0 \subset \mathcal{M}_1$, we define the deviance statistic as

$$D = 2\left\{\ell_1(\mathcal{M}_1) - \ell_0(\mathcal{M}_0)\right\},$$

where $\ell_1(\mathcal{M}_1)$ and $\ell_0(\mathcal{M}_0)$ are the maximised log–likelihood under models $\mathcal{M}_1$ and $\mathcal{M}_0$ respectively.

Asymptotically, $D \sim \chi^2_k$, where $k = \dim(\mathcal{M}_1) - \dim(\mathcal{M}_0)$.

5.2.2 Model choice

Formally,

$$H_0 : \mathcal{M}_1 = \mathcal{M}_0.$$

Thus, large values of D suggest that the increase in model size has been worthwhile, and that model $\mathcal{M}_1$ explains substantially more of the variation in the data than $\mathcal{M}_0$.

5.2.2 Model choice

Formally,

$$H_0 : \mathcal{M}_1 = \mathcal{M}_0.$$

Thus, large values of D suggest that the increase in model size has been worthwhile, and that model $\mathcal{M}_1$ explains substantially more of the variation in the data than $\mathcal{M}_0$.

5.2.2 Model choice

Suppose

- $\mathcal{M}_1$: model allowing for **linear trend** in μ
- $\mathcal{M}_0$: **stationary** model

5.2.2 Model choice

Suppose

- $\mathcal{M}_1$: model allowing for **linear trend** in μ
- $\mathcal{M}_0$: **stationary** model

5.2.2 Model choice

Suppose

- $\mathcal{M}_1$: model allowing for **linear trend** in μ
- $\mathcal{M}_0$: **stationary** model

5.2.2 Model choice

Thus,

$$D = 2\left\{-216.0626 - (-222.7145)\right\} = 13.3038.$$

The difference in dimensionality is 1, so the critical value for our test is

$$\chi^2_1(0.05) = 3.841.$$

Relative to this, our deviance statistic is large: the model which allows for a linear trend in μ explains substantially more of the variation in the data than does the simpler stationary model.

5.2.2 Model choice

Thus,

$$D = 2\left\{-216.0626 - (-222.7145)\right\} = 13.3038.$$

The difference in dimensionality is 1, so the critical value for our test is

$$\chi^2_1(0.05) = 3.841.$$

Relative to this, our deviance statistic is large: the model which allows for a linear trend in μ explains substantially more of the variation in the data than does the simpler stationary model.

5.2.2 Model choice

Thus,

$$D = 2\left\{-216.0626 - (-222.7145)\right\} = 13.3038.$$

The difference in dimensionality is 1, so the critical value for our test is

$$\chi^2_1(0.05) = 3.841.$$

Relative to this, our deviance statistic is large: the model which allows for a linear trend in μ explains substantially more of the variation in the data than does the simpler stationary model.

5.2.2 Model choice

Thus,

$$D = 2\left\{-216.0626 - (-222.7145)\right\} = 13.3038.$$

The difference in dimensionality is 1, so the critical value for our test is

$$\chi^2_1(0.05) = 3.841.$$

Relative to this, our deviance statistic is large: the model which allows for a linear trend in μ explains substantially more of the variation in the data than does the simpler stationary model.

5.2.2 Model choice

We could, of course, use this method to check for a more complex association through time.

For example, to check for a **quadratic trend** we might use a model $\mathcal{M}_2$ with the following form for μ:

$$\mu(t) = \beta_0 + \beta_1 t + \beta_2 t^2.$$

Demonstration in R

5.2.2 Model choice

We could, of course, use this method to check for a more complex association through time.

For example, to check for a **quadratic trend** we might use a model $\mathcal{M}_2$ with the following form for μ:

$$\mu(t) = \beta_0 + \beta_1 t + \beta_2 t^2.$$

Demonstration in R

5.2.2 Model choice

We could, of course, use this method to check for a more complex association through time.

For example, to check for a **quadratic trend** we might use a model $\mathcal{M}_2$ with the following form for μ:

$$\mu(t) = \beta_0 + \beta_1 t + \beta_2 t^2.$$

Demonstration in R

5.2.2 Model choice

We could, of course, use this method to check for a more complex association through time.

For example, to check for a **quadratic trend** we might use a model $\mathcal{M}_2$ with the following form for μ:

$$\mu(t) = \beta_0 + \beta_1 t + \beta_2 t^2.$$

Demonstration in R

5.2.2 Model choice

Comparing model $\mathcal{M}_2$ with model $\mathcal{M}_1$ gives

$$D = 2\left\{-216.0555 - (-216.0626)\right\} = 0.0142,$$

which is *small* compared to $\chi^2_1(0.05) = 3.841$.

Thus, allowing for a quadratic dependence in time does not improve on our model which allows for a linear trend through time, and so we would reject model $\mathcal{M}_2$.

5.2.2 Model choice

Comparing model $\mathcal{M}_2$ with model $\mathcal{M}_1$ gives

$$D = 2\left\{-216.0555 - (-216.0626)\right\} = 0.0142,$$

which is *small* compared to $\chi^2_1(0.05) = 3.841$.

Thus, allowing for a quadratic dependence in time does not improve on our model which allows for a linear trend through time, and so we would reject model $\mathcal{M}_2$.

5.2.3 Model diagnostics

Before estimating return levels, we should check the goodness–of–fit of our model which allows for a linear trend in μ.

The lack of homogeneity in the distributional assumptions for each observation, however, mean some modification of the standard procedures (e.g. **probability plots** and **quantile plots**) is required.

5.2.3 Model diagnostics

Before estimating return levels, we should check the goodness–of–fit of our model which allows for a linear trend in μ.

The lack of homogeneity in the distributional assumptions for each observation, however, mean some modification of the standard procedures (e.g. **probability plots** and **quantile plots**) is required.

5.2.3 Model diagnostics

For example, for the Venice annual maximum sea levels, we have

$$X_t \sim \text{GEV}(\mu(t), \sigma, \xi), \quad t = 1, 2, \ldots, 51,$$

giving a different GEV in each year indicated by t.

What we need to do is standardise so that we can assume the X_t *are* IID across all years t.

Usually, the set of non–stationary annual maxima are transformed to a common Gumbel distribution with distribution function $F(y) = \exp\{-e^{-y}\}$.

5.2.3 Model diagnostics

For example, for the Venice annual maximum sea levels, we have

$$X_t \sim \text{GEV}(\mu(t), \sigma, \xi), \quad t = 1, 2, \ldots, 51,$$

giving a different GEV in each year indicated by t.

What we need to do is standardise so that we can assume the X_t *are* IID across all years t.

Usually, the set of non–stationary annual maxima are transformed to a common Gumbel distribution with distribution function $F(y) = \exp\{-e^{-y}\}$.

5.2.3 Model diagnostics

For example, for the Venice annual maximum sea levels, we have

$$X_t \sim \text{GEV}(\mu(t), \sigma, \xi), \quad t = 1, 2, \ldots, 51,$$

giving a different GEV in each year indicated by t.

What we need to do is standardise so that we can assume the X_t *are* IID across all years t.

Usually, the set of non–stationary annual maxima are transformed to a common Gumbel distribution with distribution function $F(y) = \exp\{-e^{-y}\}$.

5.2.3 Model diagnostics

We can obtain the required transformation by equating $F(y_t)$ to $\text{GEV}(x_t; \hat{\mu}(t), \hat{\sigma}, \hat{\xi})$ and solving for y_t:

$$\begin{aligned}
\exp\left\{-e^{-y_t}\right\} &= \exp\left\{-\left[1+\hat{\xi}\left(\frac{x_t-(\hat{\beta}_0+\hat{\beta}_1 t)}{\hat{\sigma}}\right)\right]^{-1/\hat{\xi}}\right\} \\
e^{-y_t} &= \left[1+\hat{\xi}\left(\frac{x_t-(\hat{\beta}_0+\hat{\beta}_1 t)}{\hat{\sigma}}\right)\right]^{-1/\hat{\xi}} \\
y_t &= \frac{1}{\hat{\xi}}\log\left[1+\hat{\xi}\left(\frac{x_t-(\hat{\beta}_0+\hat{\beta}_1 t)}{\hat{\sigma}}\right)\right],
\end{aligned}$$

5.2.3 Model diagnostics

We can obtain the required transformation by equating $F(y_t)$ to $\text{GEV}(x_t; \hat{\mu}(t), \hat{\sigma}, \hat{\xi})$ and solving for y_t:

$$\begin{aligned}
\exp\left\{-e^{-y_t}\right\} &= \exp\left\{-\left[1+\hat{\xi}\left(\frac{x_t-(\hat{\beta}_0+\hat{\beta}_1 t)}{\hat{\sigma}}\right)\right]^{-1/\hat{\xi}}\right\} \\
e^{-y_t} &= \left[1+\hat{\xi}\left(\frac{x_t-(\hat{\beta}_0+\hat{\beta}_1 t)}{\hat{\sigma}}\right)\right]^{-1/\hat{\xi}} \\
y_t &= \frac{1}{\hat{\xi}}\log\left[1+\hat{\xi}\left(\frac{x_t-(\hat{\beta}_0+\hat{\beta}_1 t)}{\hat{\sigma}}\right)\right],
\end{aligned}$$

5.2.3 Model diagnostics

We can obtain the required transformation by equating $F(y_t)$ to $\text{GEV}(x_t; \hat{\mu}(t), \hat{\sigma}, \hat{\xi})$ and solving for y_t:

$$\begin{aligned}
\exp\left\{-e^{-y_t}\right\} &= \exp\left\{-\left[1+\hat{\xi}\left(\frac{x_t-(\hat{\beta}_0+\hat{\beta}_1 t)}{\hat{\sigma}}\right)\right]^{-1/\hat{\xi}}\right\} \\
e^{-y_t} &= \left[1+\hat{\xi}\left(\frac{x_t-(\hat{\beta}_0+\hat{\beta}_1 t)}{\hat{\sigma}}\right)\right]^{-1/\hat{\xi}} \\
y_t &= \frac{1}{\hat{\xi}}\log\left[1+\hat{\xi}\left(\frac{x_t-(\hat{\beta}_0+\hat{\beta}_1 t)}{\hat{\sigma}}\right)\right],
\end{aligned}$$

5.2.3 Model diagnostics

This gives gives the following transformation to common Gumbel margins for our Venice annual maxima:

$$y_t = -\frac{1}{0.027}\log\left[1 - 0.027\left(\frac{x_t - (96.986 + 0.564t)}{14.584}\right)\right].$$

Demonstration in R

5.2.3 Model diagnostics

Now that we have transformed the original data to a single common distribution, we can apply the standard graphical diagnostics.

For example, we can compare the **empirical probabilities** and **quantiles** of y_t to their **theoretical** counterparts from the Gumbel distribution.

Fortunately, the usual command in ismev can be used to produce these plots for us.

Demonstration in R

5.2.3 Model diagnostics

Now that we have transformed the original data to a single common distribution, we can apply the standard graphical diagnostics.

For example, we can compare the **empirical probabilities** and **quantiles** of y_t to their **theoretical** counterparts from the Gumbel distribution.

Fortunately, the usual command in ismev can be used to produce these plots for us.

Demonstration in R

5.2.3 Model diagnostics

Now that we have transformed the original data to a single common distribution, we can apply the standard graphical diagnostics.

For example, we can compare the **empirical probabilities** and **quantiles** of y_t to their **theoretical** counterparts from the Gumbel distribution.

Fortunately, the usual command in ismev can be used to produce these plots for us.

Demonstration in R

5.2.3 Model diagnostics

Now that we have transformed the original data to a single common distribution, we can apply the standard graphical diagnostics.

For example, we can compare the **empirical probabilities** and **quantiles** of y_t to their **theoretical** counterparts from the Gumbel distribution.

Fortunately, the usual command in ismev can be used to produce these plots for us.

Demonstration in R

5.2.4 Return level estimation

Recall Equation (2.10) (**memorise!**) from Chapter 2 for estimating return levels from the GEV:

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

Since we have a time–varying location parameter $\hat{\mu}(t) = \hat{\beta}_0 + \hat{\beta}_1 t$, we will clearly have time–varying estimates of return levels $\hat{z}_r(t)$.

For example, an estimate of the sea level we might expect to see in Venice once every 100 years is given by

$$\hat{z}_r(t) = (96.986 + 0.564t) - \frac{14.584}{0.027}\left[\left(-\log\left(1 - 100^{-1}\right)\right)^{0.027} - 1\right].$$

5.2.4 Return level estimation

Recall Equation (2.10) (**memorise!**) from Chapter 2 for estimating return levels from the GEV:

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

Since we have a time–varying location parameter $\hat{\mu}(t) = \hat{\beta}_0 + \hat{\beta}_1 t$, we will clearly have time–varying estimates of return levels $\hat{z}_r(t)$.

For example, an estimate of the sea level we might expect to see in Venice once every 100 years is given by

$$\hat{z}_r(t) = (96.986 + 0.564t) - \frac{14.584}{0.027}\left[\left(-\log\left(1 - 100^{-1}\right)\right)^{0.027} - 1\right].$$

5.2.4 Return level estimation

Recall Equation (2.10) (**memorise!**) from Chapter 2 for estimating return levels from the GEV:

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

Since we have a time–varying location parameter $\hat{\mu}(t) = \hat{\beta}_0 + \hat{\beta}_1 t$, we will clearly have time–varying estimates of return levels $\hat{z}_r(t)$.

For example, an estimate of the sea level we might expect to see in Venice once every 100 years is given by

$$\hat{z}_r(t) = (96.986 + 0.564t) - \frac{14.584}{0.027}\left[\left(-\log\left(1 - 100^{-1}\right)\right)^{0.027} - 1\right].$$

5.2.4 Return level estimation

Figure 5.4 shows how we might expect this estimate to vary for $t = 52, 53, \ldots$, i.e. for the years 2012, 2013, ...

We could treat these as forecasts of the 100–year return levels as we move through time.

obviously, such forecasts will assume the linear trend for μ continues beyond the range of data we have observed and will, of course be subject to error (which we can estimate by constructing point–wise 95% confidence intervals using the profile log–likelihood, for example).

5.2.4 Return level estimation

Figure 5.4 shows how we might expect this estimate to vary for $t = 52, 53, \ldots$, i.e. for the years 2012, 2013, ...

We could treat these as forecasts of the 100–year return levels as we move through time.

obviously, such forecasts will assume the linear trend for μ continues beyond the range of data we have observed and will, of course be subject to error (which we can estimate by constructing point–wise 95% confidence intervals using the profile log–likelihood, for example).

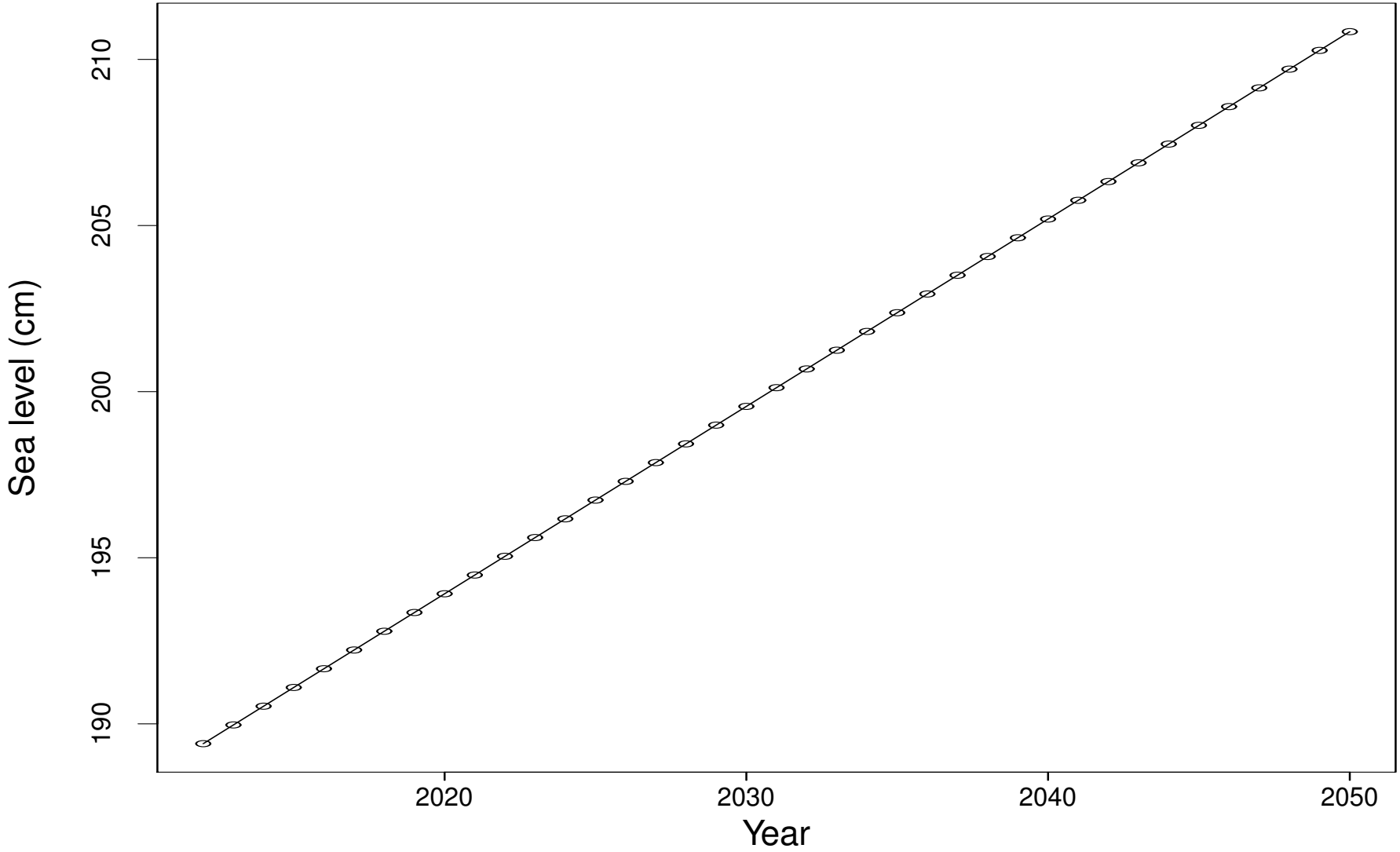
5.2.4 Return level estimation

Figure 5.4 shows how we might expect this estimate to vary for $t = 52, 53, \ldots$, i.e. for the years 2012, 2013, ...

We could treat these as forecasts of the 100–year return levels as we move through time.

obviously, such forecasts will assume the linear trend for μ continues beyond the range of data we have observed and will, of course be subject to error (which we can estimate by constructing point–wise 95% confidence intervals using the profile log–likelihood, for example).

5.2.4 Return level estimation



5.3 Sea levels and the SOI

A different situation which could use the same approach as that in the previous section is where the extremal behaviour of a series is related to another variable, rather than time.

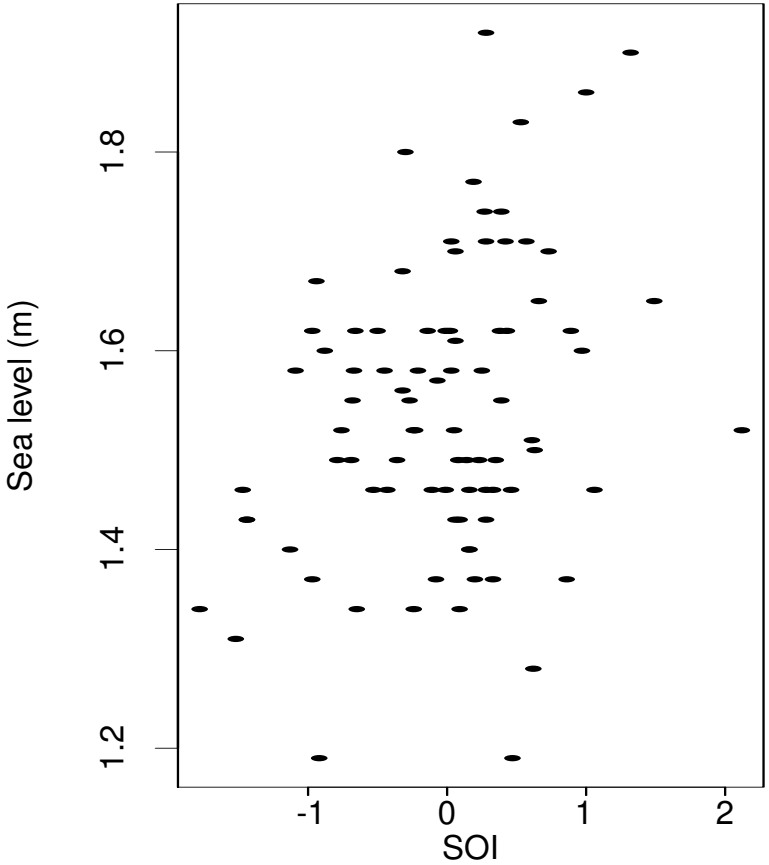
For example, studies have revealed a link between annual maximum sea levels at Fremantle, Australia, and the mean value of the **Southern Oscillation Index** (SOI, an indicator of meteorological volatility due to effects such as El Niño).

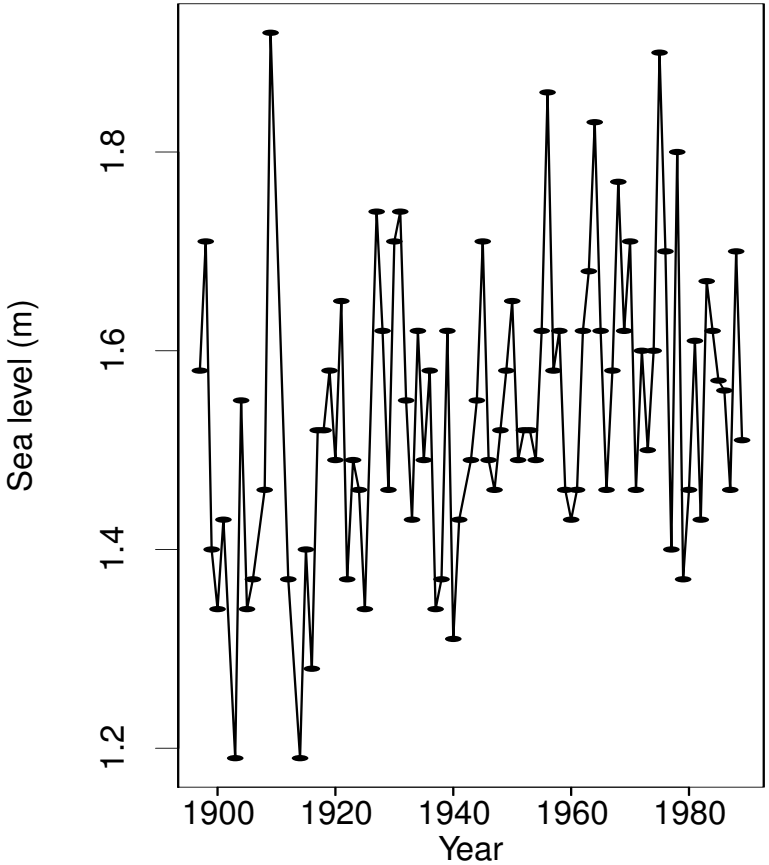
5.3 Sea levels and the SOI

A different situation which could use the same approach as that in the previous section is where the extremal behaviour of a series is related to another variable, rather than time.

For example, studies have revealed a link between annual maximum sea levels at Fremantle, Australia, and the mean value of the **Southern Oscillation Index** (SOI, an indicator of meteorological volatility due to effects such as El Niño).

5.3 Sea levels and the SOI





5.3 Sea levels and the SOI

Thus, the following model for X_t, the annual maximum sea level at Fremantle in year t, might be suitable:

$$X_t \sim \text{GEV}(\mu(t), \sigma, \xi),$$

where

$$\mu(t) = \beta_0 + \beta_1 \text{SOI}(t), \tag{5.3}$$

where $\text{SOI}(t)$ denotes the mean value of SOI in year t.

5.3 Sea levels and the SOI

Thus, the following model for X_t, the annual maximum sea level at Fremantle in year t, might be suitable:

$$X_t \sim \text{GEV}(\mu(t), \sigma, \xi),$$

where

$$\mu(t) = \beta_0 + \beta_1 \text{SOI}(t), \tag{5.3}$$

where $\text{SOI}(t)$ denotes the mean value of SOI in year t.

5.3 Sea levels and the SOI

However, the plot in the right–hand–side of Figure 5.5 also reveals a possible trend in sea levels through time, suggesting

$$\mu(t) = \beta_0 + \beta_1 t, \tag{5.4}$$

where $t = 1, 2, \ldots,$ as in Example 5.2.

5.3 Sea levels and the SOI

We can combine Equations (5.3) and (5.4) to allow for a dependence on time *and* SOI by letting

$$\mu(t) = \beta_0 + \beta_1 \text{SOI}(t) + \beta_2 t; \tag{5.5}$$

however, a technique of **forward selection** should be used to check whether or not any of Equations (5.3), (5.4) or (5.5) give significant improvement over the stationary model.

5.3 Sea levels and the SOI

Class demonstration in R

5.4 Rainfall extremes in New York City

Recall Section 3.2 in which we modelled rainfall extremes in New York using a threshold–based approach.

The Generalised Pareto distribution was applied to rainfall exceedances over a threshold of $u = 30$mm, giving estimates of the scale and shape as

$$\hat{\sigma} = 7.44(0.958) \quad \hat{\xi} = 0.184(0.101)$$

(respectively).

5.4 Rainfall extremes in New York City

Recall Section 3.2 in which we modelled rainfall extremes in New York using a threshold–based approach.

The Generalised Pareto distribution was applied to rainfall exceedances over a threshold of $u = 30$mm, giving estimates of the scale and shape as

$$\hat{\sigma} = 7.44(0.958) \quad \hat{\xi} = 0.184(0.101)$$

(respectively).

5.4 Rainfall extremes in New York City

Recall that the GPD(σ, ξ) arises from the GEV(μ, σ, ξ), where the GPD scale parameter is a function of the GEV location and shape parameters.

Thus, attempting to model any trend in our threshold exceedances is usually done through linear modelling of the scale parameter σ (the GPD doesn't have a location parameter *per se*).

5.4 Rainfall extremes in New York City

Recall that the GPD(σ, ξ) arises from the GEV(μ, σ, ξ), where the GPD scale parameter is a function of the GEV location and shape parameters.

Thus, attempting to model any trend in our threshold exceedances is usually done through linear modelling of the scale parameter σ (the GPD doesn't have a location parameter *per se*).

5.4 Rainfall extremes in New York City

Since the scale parameter σ must be positive, we might choose to model a trend through time as

$$\sigma = \exp\{\beta_0 + \beta_1 t\}, \tag{5.6}$$

where t is, once again, an indicator of time.

Class demonstration in R

5.5 Generalisation

With reference to the examples discussed so far, we could model non–stationarity through *any* of the parameters in our extremal model.

For example, take a non–stationary GEV model to describe the distribution of X_t, for $t = 1, 2, \ldots, m$:

$$X_t \sim \text{GEV}(\mu(t), \sigma(t), \xi(t)),$$

where each of the model parameters have an expression in terms of a parameter vector β and some covariates.

The likelihood is then

$$L(x_t; \beta) = \prod_{i=1}^{m} g(x_t; \mu(t), \sigma(t), \xi(t)),$$

where g is the GEV density function.

5.5 Generalisation

With reference to the examples discussed so far, we could model non–stationarity through *any* of the parameters in our extremal model.

For example, take a non–stationary GEV model to describe the distribution of X_t, for $t = 1, 2, \ldots, m$:

$$X_t \sim \text{GEV}(\mu(t), \sigma(t), \xi(t)),$$

where each of the model parameters have an expression in terms of a parameter vector $\boldsymbol{\beta}$ and some covariates.

The likelihood is then

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

where g is the GEV density function.

5.5 Generalisation

With reference to the examples discussed so far, we could model non–stationarity through *any* of the parameters in our extremal model.

For example, take a non–stationary GEV model to describe the distribution of X_t, for $t = 1, 2, \ldots, m$:

$$X_t \sim \text{GEV}(\mu(t), \sigma(t), \xi(t)),$$

where each of the model parameters have an expression in terms of a parameter vector $\boldsymbol{\beta}$ and some covariates.

The likelihood is then

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

where g is the GEV density function.

5.5 Generalisation

From this, we can form the log–likelihood, and then maximise in the usual way (for example, using `nlm` in R).

In terms of threshold exceedances Y_t, $t = 1, 2, \ldots, k$, we could replace the GEV with the GPD:

$$Y_t \sim \text{GPD}(\sigma(t), \xi(t)),$$

with σ being defined as in Equation (5.6) to retain the positivity of the GPD scale.

5.5 Generalisation

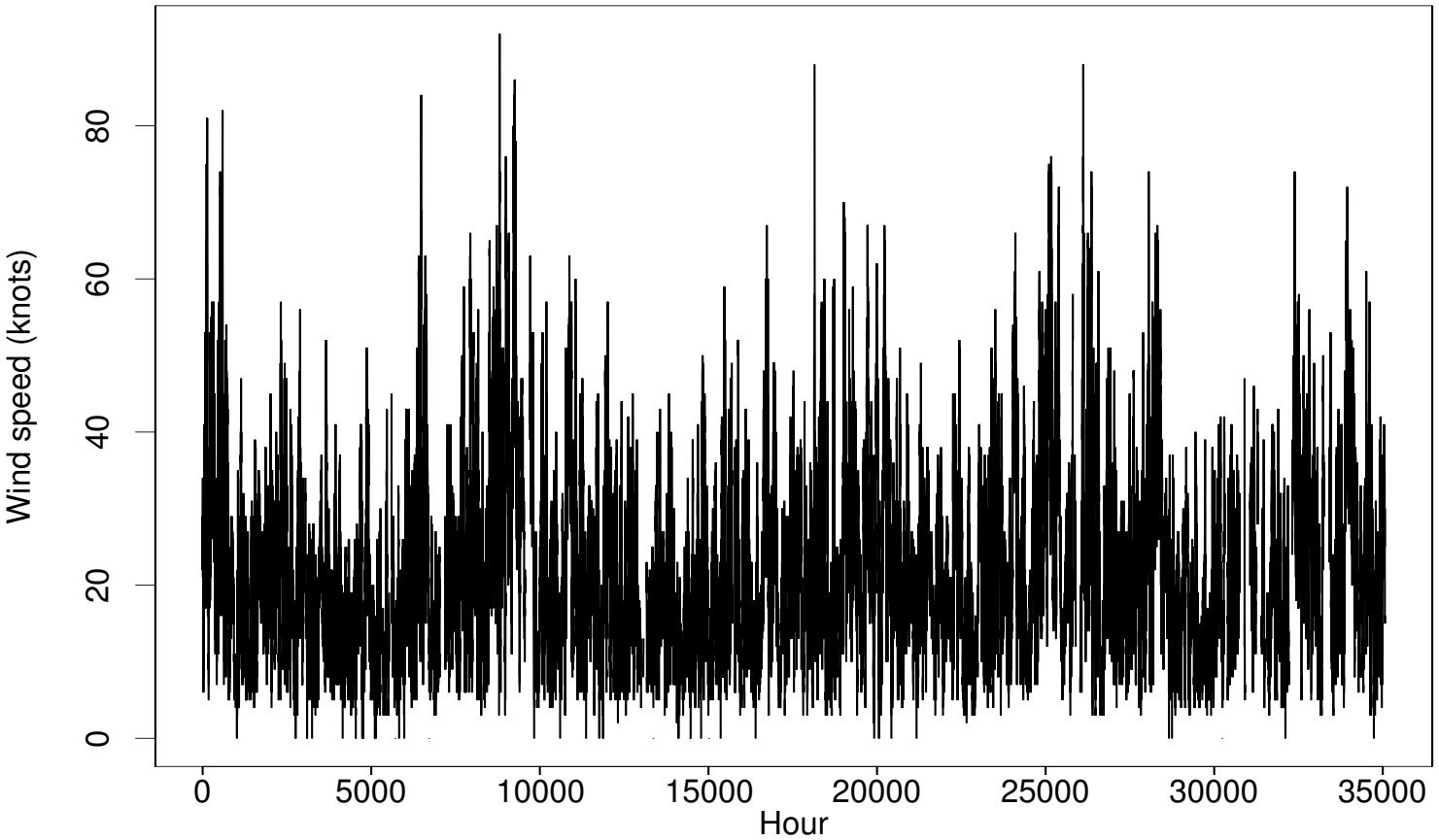
From this, we can form the log–likelihood, and then maximise in the usual way (for example, using `nlm` in R).

In terms of threshold exceedances Y_t, $t = 1, 2, \ldots, k$, we could replace the GEV with the GPD:

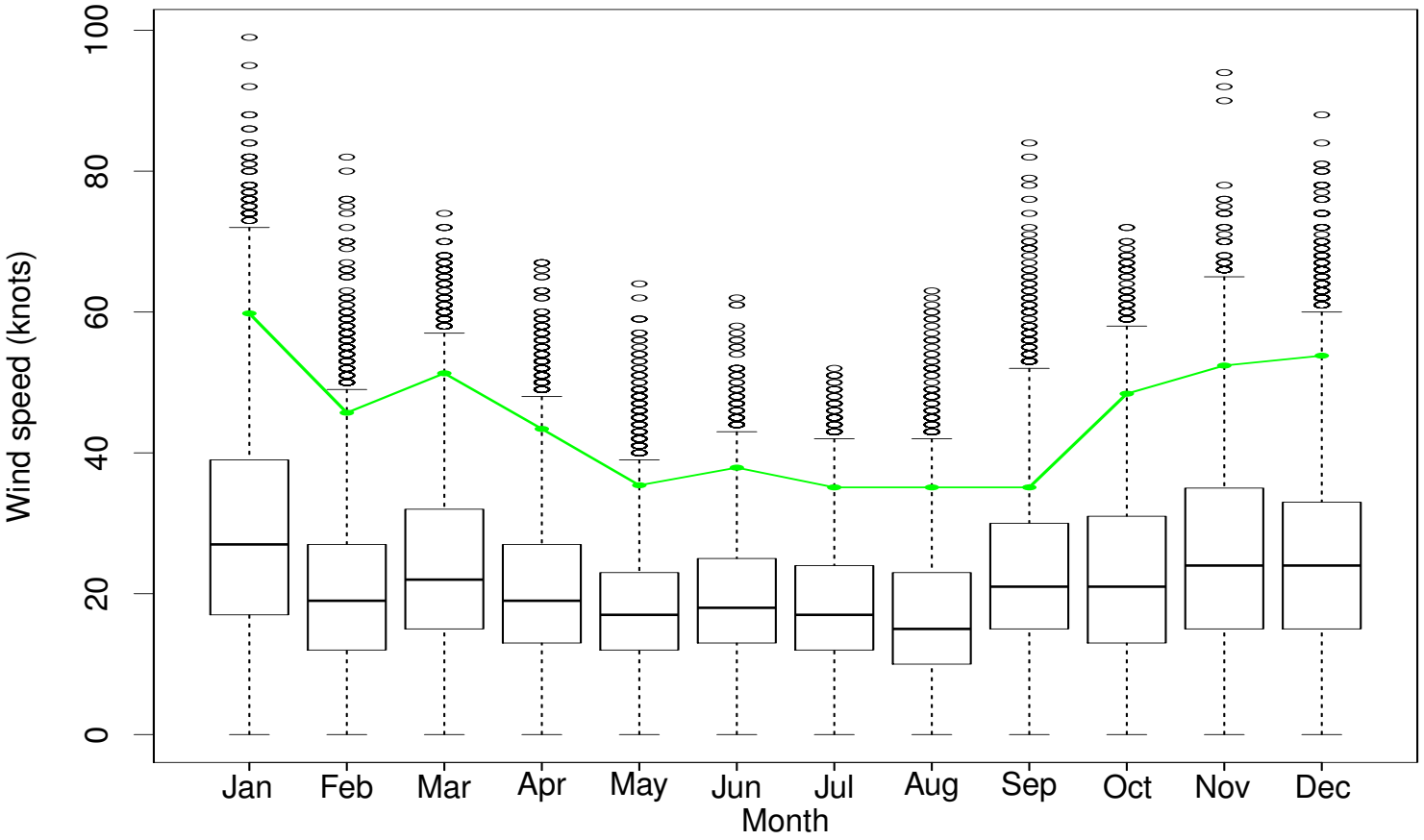
$$Y_t \sim \text{GPD}(\sigma(t), \xi(t)),$$

with σ being defined as in Equation (**5.6**) to retain the positivity of the GPD scale.

5.6 Wind speed extremes at High Bradfield



5.6 Wind speed extremes at High Bradfield



5.6.1. Single season approach

Under the single season approach, an extremal model is fitted to the extremes of an environmental process from the season which gives rise to the ‘most extreme’ extremes.

- Bradfield wind speeds – use January only
- Some appeal: **easy to implement**, focuses on the **most extreme** extremes
- But: **wasteful** of data – especially if we decluster as well!

5.6.1. Single season approach

Under the single season approach, an extremal model is fitted to the extremes of an environmental process from the season which gives rise to the 'most extreme' extremes.

- Bradfield wind speeds – use January only
- Some appeal: **easy to implement**, focuses on the **most extreme** extremes
- But: **wasteful** of data – especially if we decluster as well!

5.6.1. Single season approach

Under the single season approach, an extremal model is fitted to the extremes of an environmental process from the season which gives rise to the 'most extreme' extremes.

- Bradfield wind speeds – use January only
- Some appeal: **easy to implement**, focuses on the **most extreme** extremes
- But: **wasteful** of data – especially if we decluster as well!

5.6.1. Single season approach

Under the single season approach, an extremal model is fitted to the extremes of an environmental process from the season which gives rise to the 'most extreme' extremes.

- Bradfield wind speeds – use January only
- Some appeal: **easy to implement**, focuses on the **most extreme** extremes
- But: **wasteful** of data – especially if we decluster as well!

5.6.2 Seasonal piecewise approach

It is usual in strongly seasonal climates for the occurrence of extreme winds to be confined to a certain part of the yearly cycle.

Example: U.K. $\longrightarrow$ unusual for wind damage to occur outside the period **October–March**.

5.6.2 Seasonal piecewise approach

It is usual in strongly seasonal climates for the occurrence of extreme winds to be confined to a certain part of the yearly cycle.

Example: U.K. $\longrightarrow$ unusual for wind damage to occur outside the period **October–March**.

5.6.2 Seasonal piecewise approach

The seasonal variation observed at Bradfield might be expected: Prolonged, **anticyclonic** periods more prevalent during June, July and August.

There is only a point to modelling the extremes which occur in summer months if we believe that they can help us understand what happens in winter months, where genuinely large events *can* occur.

5.6.2 Seasonal piecewise approach

The seasonal variation observed at Bradfield might be expected: Prolonged, **anticyclonic** periods more prevalent during June, July and August.

There is only a point to modelling the extremes which occur in summer months if we believe that they can help us understand what happens in winter months, where genuinely large events *can* occur.

5.6.2 Seasonal piecewise approach

The seasonal variation observed at Bradfield might be expected: Prolonged, **anticyclonic** periods more prevalent during June, July and August.

There is only a point to modelling the extremes which occur in summer months if we believe that they can help us understand what happens in winter months, where genuinely large events *can* occur.

5.6.2 Seasonal piecewise approach

For this to be realistic, we must assume that the **same mechanism** is responsible for the generation of large gusts throughout the year $\longrightarrow$ it is just the **scale** of this mechanism which changes.

Indeed, in temperate climates (such as that of the U.K.), the same alternating sequence of anticyclones and depressions leads to most of the storms which occur throughout the year.

The seasonal variability comes from the **severity** of these systems.

5.6.2 Seasonal piecewise approach

For this to be realistic, we must assume that the **same mechanism** is responsible for the generation of large gusts throughout the year $\longrightarrow$ it is just the **scale** of this mechanism which changes.

Indeed, in temperate climates (such as that of the U.K.), the same alternating sequence of anticyclones and depressions leads to most of the storms which occur throughout the year.

The seasonal variability comes from the **severity** of these systems.

5.6.2 Seasonal piecewise approach

For this to be realistic, we must assume that the **same mechanism** is responsible for the generation of large gusts throughout the year $\longrightarrow$ it is just the **scale** of this mechanism which changes.

Indeed, in temperate climates (such as that of the U.K.), the same alternating sequence of anticyclones and depressions leads to most of the storms which occur throughout the year.

The seasonal variability comes from the **severity** of these systems.

5.6.2 Seasonal piecewise approach

For this to be realistic, we must assume that the **same mechanism** is responsible for the generation of large gusts throughout the year $\longrightarrow$ it is just the **scale** of this mechanism which changes.

Indeed, in temperate climates (such as that of the U.K.), the same alternating sequence of anticyclones and depressions leads to most of the storms which occur throughout the year.

The seasonal variability comes from the **severity** of these systems.

5.6.2 Seasonal piecewise approach

Taking the **calendar month** as our seasonal unit:

- For the U.K., experience suggests this might be OK
- Strikes a good balance between:
  - reflecting the continuous nature of seasonal changes in climate
  - retaining a substantial amount of data for analysis within each season

5.6.2 Seasonal piecewise approach

Taking the **calendar month** as our seasonal unit:

- For the U.K., experience suggests this might be OK
- Strikes a good balance between:
  - reflecting the continuous nature of seasonal changes in climate
  - retaining a substantial amount of data for analysis within each season

5.6.2 Seasonal piecewise approach

Taking the **calendar month** as our seasonal unit:

- For the U.K., experience suggests this might be OK
- Strikes a good balance between:
  - reflecting the continuous nature of seasonal changes in climate
  - retaining a substantial amount of data for analysis within each season

5.6.2 Seasonal piecewise approach

Taking the **calendar month** as our seasonal unit:

- For the U.K., experience suggests this might be OK
- Strikes a good balance between:
  - reflecting the continuous nature of seasonal changes in climate
  - retaining a substantial amount of data for analysis within each season

5.6.2 Seasonal piecewise approach

Taking the **calendar month** as our seasonal unit:

- For the U.K., experience suggests this might be OK
- Strikes a good balance between:
  - reflecting the continuous nature of seasonal changes in climate
  - retaining a substantial amount of data for analysis within each season

5.6.2 Seasonal piecewise approach

Month (m)	u_m	n_m	$\hat{\sigma}_m$	$\hat{\xi}_m$
1	55.341	28	21.373 (5.358)	−0.420 (0.183)
2	41.531	24	15.130 (4.635)	−0.226 (0.233)
3	48.100	29	23.277 (6.316)	−0.894 (0.259)
4	39.910	29	14.853 (4.448)	−0.440 (0.249)
5	31.943	46	9.456 (1.990)	−0.158 (0.147)
6	35.670	35	12.329 (2.592)	−0.409 (0.143)
7	32.290	36	12.517 (2.609)	−0.605 (0.161)
8	32.639	34	10.199 (2.361)	−0.203 (0.159)
9	33.232	49	18.772 (3.668)	−0.255 (0.138)
10	44.914	34	11.669 (3.533)	−0.274 (0.254)
11	48.394	33	14.991 (3.381)	−0.225 (0.149)
12	49.341	35	18.681 (4.229)	−0.416 (0.166)

5.6.2 Seasonal piecewise approach

In terms of **return level inference**, it would not make practical sense to have monthly varying estimates of the r–year return level z_r.

To include information from *all* months in our return level estimation procedure, we solve

$$\prod_{m=1}^{12} H(\hat{z}_r; \hat{\lambda}_{u_m}, \hat{\sigma}_m, \hat{\xi}_m) \quad = \quad 1 - \frac{1}{rn_y}$$

for $\hat{z}_r$, where H is the GPD distribution function and n_y is the (average) number of observations per year.

5.6.2 Seasonal piecewise approach

In terms of **return level inference**, it would not make practical sense to have monthly varying estimates of the r–year return level z_r.

To include information from *all* months in our return level estimation procedure, we solve

$$\prod_{m=1}^{12} H(\hat{z}_r; \hat{\lambda}_{u_m}, \hat{\sigma}_m, \hat{\xi}_m) \quad = \quad 1 - \frac{1}{rn_y}$$

for $\hat{z}_r$, where H is the GPD distribution function and n_y is the (average) number of observations per year.

5.6.2 Seasonal piecewise approach

Thus, we need to solve

$$\prod_{m=1}^{12} \left\{ 1 - \hat{\lambda}_u \left[1 + \hat{\xi} \left(\frac{\hat{z}_r - u_m}{\hat{\sigma}_m} \right) \right]_+^{-1/\hat{\xi}_m} \right\} = 1 - \frac{1}{rn_y}$$

for $\hat{z}_r$, that is

$$\prod_{m=1}^{12} \left\{ 1 - \hat{\lambda}_u \left[1 + \hat{\xi} \left(\frac{\hat{z}_r - u_m}{\hat{\sigma}_m} \right) \right]_+^{-1/\hat{\xi}_m} \right\} - \left(1 - \frac{1}{rn_y} \right) = 0.$$

5.6.2 Seasonal piecewise approach

Thus, we need to solve

$$\prod_{m=1}^{12} \left\{ 1 - \hat{\lambda}_u \left[1 + \hat{\xi} \left(\frac{\hat{z}_r - u_m}{\hat{\sigma}_m} \right) \right]_+^{-1/\hat{\xi}_m} \right\} = 1 - \frac{1}{rn_y}$$

for $\hat{z}_r$, that is

$$\prod_{m=1}^{12} \left\{ 1 - \hat{\lambda}_u \left[1 + \hat{\xi} \left(\frac{\hat{z}_r - u_m}{\hat{\sigma}_m} \right) \right]_+^{-1/\hat{\xi}_m} \right\} - \left(1 - \frac{1}{rn_y} \right) = 0.$$

5.6.2 Seasonal piecewise approach

This equation cannot be solved analytically.

Rather, a numerical procedure must be used.

This is easy to do in R using the `uniroot` procedure.

[Class demonstration in R*]*

5.6.2 Seasonal piecewise approach

This equation cannot be solved analytically.

Rather, a numerical procedure must be used.

This is easy to do in R using the `uniroot` procedure.

[Class demonstration in R*]*

5.6.2 Seasonal piecewise approach

This equation cannot be solved analytically.

Rather, a numerical procedure must be used.

This is easy to do in R using the `uniroot` procedure.

[Class demonstration in R*]*

5.6.2 Seasonal piecewise approach

	Return period (r years)			
	10	50	200	1000
$\hat{z}_r$ (st. err.)	102.33 (3.970)	104.59 (15.951)	106.21 (23.793)	108.89 (44.865)

5.6.2 Seasonal piecewise approach

For the standard errors, we use the delta method. However, we now have:

$$V = \begin{pmatrix} \frac{\hat{\lambda}_{u_1}(1-\hat{\lambda}_{u_1})}{N_1} & 0 & \dots & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \frac{\hat{\lambda}_{u_{12}}(1-\hat{\lambda}_{u_{12}})}{N_{12}} & 0 & \dots & 0 \\ 0 & \dots & 0 & v_{1,1} & \dots & v_{1,24} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & v_{24,1} & \dots & v_{24,24} \end{pmatrix},$$

where $v_{i,j}$ denotes the (i,j)–th term of the variance–covariance matrix of $\hat{\sigma}_m$ and $\hat{\xi}_m$, $m = 1, \dots, 12$.

5.6.2 Seasonal piecewise approach

For the standard errors, we use the delta method. However, we now have:

$$V = \begin{pmatrix} \frac{\hat{\lambda}_{u_1}(1-\hat{\lambda}_{u_1})}{N_1} & 0 & \dots & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \frac{\hat{\lambda}_{u_{12}}(1-\hat{\lambda}_{u_{12}})}{N_{12}} & 0 & \dots & 0 \\ 0 & \dots & 0 & v_{1,1} & \dots & v_{1,24} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & v_{24,1} & \dots & v_{24,24} \end{pmatrix},$$

where $v_{i,j}$ denotes the (i,j)–th term of the variance–covariance matrix of $\hat{\sigma}_m$ and $\hat{\xi}_m$, $m = 1, \dots, 12$.

5.6.2 Seasonal piecewise approach

Hence, by the delta method,

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

where

$$\nabla z_r^T = \left[\frac{\partial z_r}{\partial \lambda_{u_1}}, \ldots, \frac{\partial z_r}{\partial \lambda_{u_{12}}}, \frac{\partial z_r}{\partial \sigma_1}, \ldots, \frac{\partial z_r}{\partial \sigma_{12}}, \frac{\partial z_r}{\partial \xi_1}, \ldots, \frac{\partial z_r}{\partial \xi_{12}}\right],$$

evaluated at $(\hat{\lambda}_{u_1}, \hat{\sigma}_1, \hat{\xi}_1, \ldots, \hat{\lambda}_{u_{12}}, \hat{\sigma}_{12}, \hat{\xi}_{12})$.

5.6.2 Seasonal piecewise approach

As an aside, notice how the estimate of the 50–year return level wind speed here differs to that when we assumed stationarity in Section 4.3.2:

$$\text{Assuming stationarity} \quad : \quad \hat{z}_{50} = 101.53 \text{ knots}$$

$$\text{Seasonal piecewise} \quad : \quad \hat{z}_{50} = 104.59 \text{ knots}$$

Such discrepancies become more pronounced for longer–period return levels.

5.6.2 Seasonal piecewise approach

As an aside, notice how the estimate of the 50–year return level wind speed here differs to that when we assumed stationarity in Section 4.3.2:

$$\text{Assuming stationarity} \quad : \quad \hat{z}_{50} = 101.53 \text{ knots}$$

$$\text{Seasonal piecewise} \quad : \quad \hat{z}_{50} = 104.59 \text{ knots}$$

Such discrepancies become more pronounced for longer–period return levels.

5.6.2 Seasonal piecewise approach

As an aside, notice how the estimate of the 50–year return level wind speed here differs to that when we assumed stationarity in Section 4.3.2:

$$\text{Assuming stationarity} \quad : \quad \hat{z}_{50} = 101.53 \text{ knots}$$

$$\text{Seasonal piecewise} \quad : \quad \hat{z}_{50} = 104.59 \text{ knots}$$

Such discrepancies become more pronounced for longer–period return levels.

5.6.2 Seasonal piecewise approach

As an aside, notice how the estimate of the 50–year return level wind speed here differs to that when we assumed stationarity in Section 4.3.2:

$$\text{Assuming stationarity} \quad : \quad \hat{z}_{50} = 101.53 \text{ knots}$$

$$\text{Seasonal piecewise} \quad : \quad \hat{z}_{50} = 104.59 \text{ knots}$$

Such discrepancies become more pronounced for longer–period return levels.

5.6.3 Smoothly varying seasonal parameters

Various authors (e.g. Fawcett and Walshaw (2006)) have investigated the use of continuously varying parameters for the GPD when seasonal variation is present.

For example, **Fourier forms** can be used to allow the GPD scale and shape to vary smoothly through time.

However, such analyses for the Bradfield wind speed data yielded little, if any, improvement over the seasonal piecewise approach.

5.6.3 Smoothly varying seasonal parameters

Various authors (e.g. Fawcett and Walshaw (2006)) have investigated the use of continuously varying parameters for the GPD when seasonal variation is present.

For example, **Fourier forms** can be used to allow the GPD scale and shape to vary smoothly through time.

However, such analyses for the Bradfield wind speed data yielded little, if any, improvement over the seasonal piecewise approach.

5.6.3 Smoothly varying seasonal parameters

Various authors (e.g. Fawcett and Walshaw (2006)) have investigated the use of continuously varying parameters for the GPD when seasonal variation is present.

For example, **Fourier forms** can be used to allow the GPD scale and shape to vary smoothly through time.

However, such analyses for the Bradfield wind speed data yielded little, if any, improvement over the seasonal piecewise approach.

Chapter 6

Multivariate extremes

6.1 Introduction

In this section we consider the problems we face if we wish to model the extremal behaviour of two or more (dependent) processes **simultaneously**.

There are several reasons why we may wish to do this:

- to model the extreme behaviour of a **particular variable** over **several nearby locations** (e.g. rainfall over a network of sites – simultaneous flooding at several locations could cause devastation);

6.1 Introduction

In this section we consider the problems we face if we wish to model the extremal behaviour of two or more (dependent) processes **simultaneously**.

There are several reasons why we may wish to do this:

- to model the extreme behaviour of a **particular variable** over **several nearby locations** (e.g. rainfall over a network of sites – simultaneous flooding at several locations could cause devastation);

6.1 Introduction

In this section we consider the problems we face if we wish to model the extremal behaviour of two or more (dependent) processes **simultaneously**.

There are several reasons why we may wish to do this:

- to model the extreme behaviour of a **particular variable** over **several nearby locations** (e.g. rainfall over a network of sites – simultaneous flooding at several locations could cause devastation);

6.1 Introduction

- to model the joint extremes of two or more **different variables** at a **particular location** (e.g. wind and rain at a site – the combined effects of wind and rain during a hurricane can result in extreme storm surge);
- to model the joint behaviour of extremes which occur as **consecutive observations** in a time–series (e.g. consecutive hourly maximum wind gusts during a storm).

6.1 Introduction

- to model the joint extremes of two or more **different variables** at a **particular location** (e.g. wind and rain at a site – the combined effects of wind and rain during a hurricane can result in extreme storm surge);
- to model the joint behaviour of extremes which occur as **consecutive observations** in a time–series (e.g. consecutive hourly maximum wind gusts during a storm).

6.1. Introduction

All of these problems suggest fitting an appropriate **limiting multivariate distribution** to the relevant data.

However, as we shall see, the derivation of such a multivariate distribution is not as easy as we might hope.

The analogy with the Normal distribution as a model for means breaks down as we move into n dimensions!

It is not even clear what the '**relevant data**' should be!

Most of the increased complexity is apparent in the move from 1 to 2 dimensions, so we will focus largely on **bivariate problems**.

6.1. Introduction

All of these problems suggest fitting an appropriate **limiting multivariate distribution** to the relevant data.

However, as we shall see, the derivation of such a multivariate distribution is not as easy as we might hope.

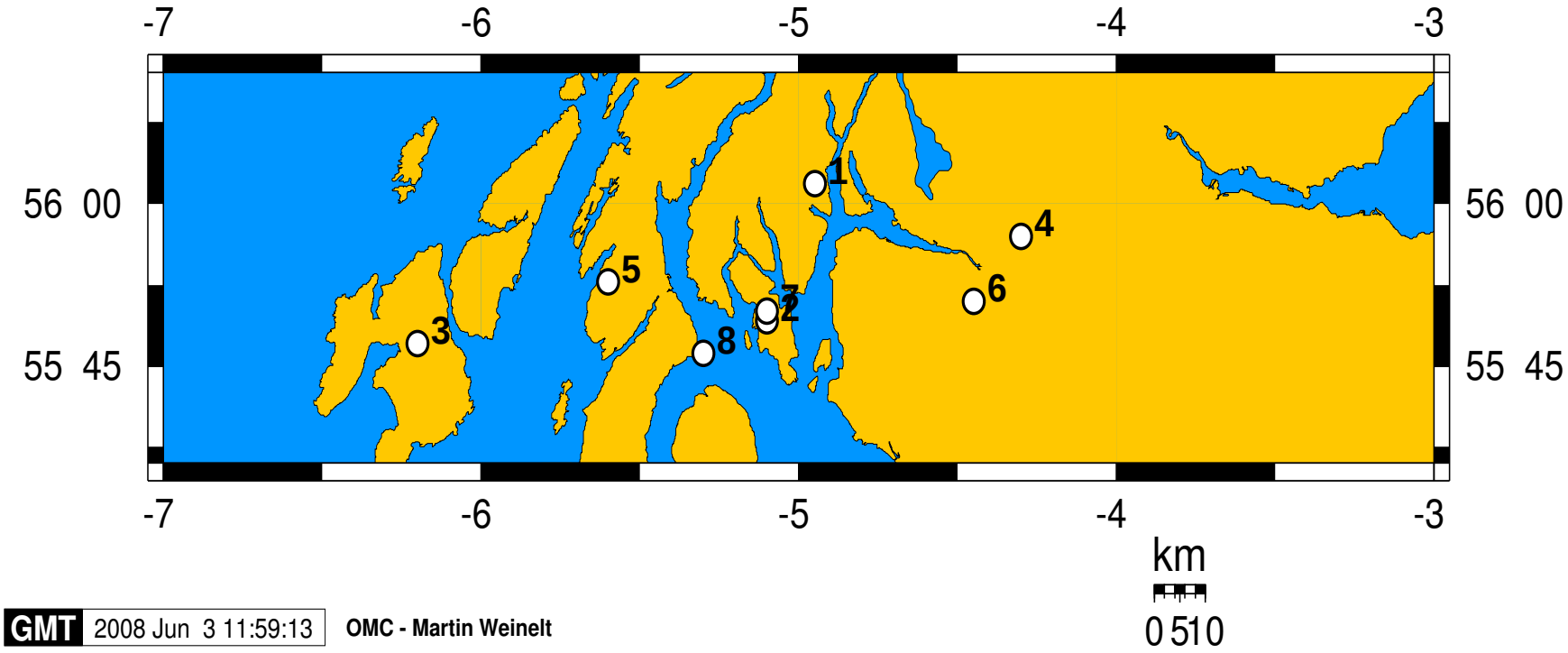
The analogy with the Normal distribution as a model for means breaks down as we move into n dimensions!

It is not even clear what the '**relevant data**' should be!

Most of the increased complexity is apparent in the move from 1 to 2 dimensions, so we will focus largely on **bivariate problems**.

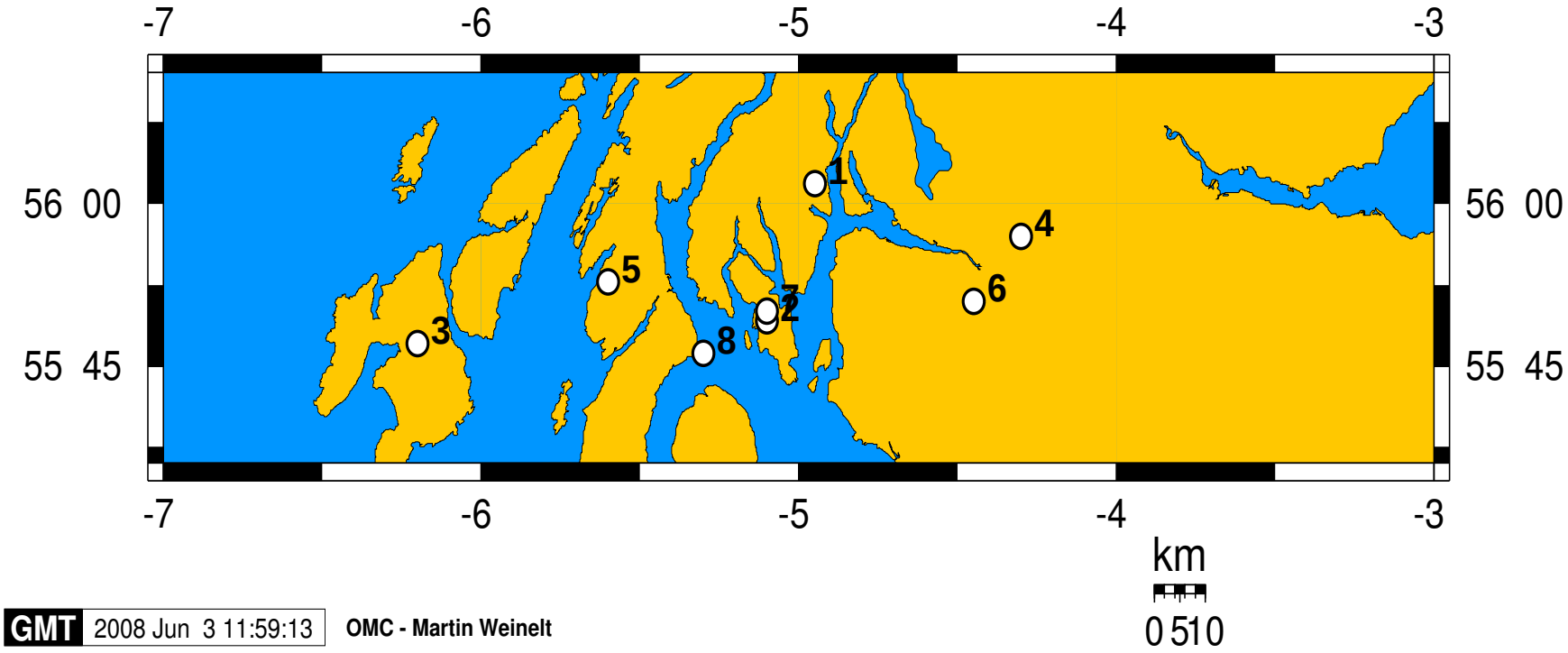
6.2.1 Example: Rainfall at 8 locations in Scotland

Suppose we want to study the joint extremes of daily rainfall accumulations at the network of 8 sites is southwest Scotland shown in Figure 6.1.



6.2.1 Example: Rainfall at 8 locations in Scotland

Suppose we want to study the joint extremes of daily rainfall accumulations at the network of 8 sites is southwest Scotland shown in Figure 6.1.



6.2.1 Example: Rainfall at 8 locations in Scotland

Such issues are of great interest, especially currently – given the **severe flooding** experienced in the U.K.

Suppose we have sequences of daily total rainfall at each location.

There is likely to be **strong inter–site dependence** extremes, in the sense that days with heavy rain are liable to occur simultaneously across locations.

The raw multivariate observations are 8–dimensional vectors of the daily rainfall over the eight sites.

6.2.1 Example: Rainfall at 8 locations in Scotland

Now suppose we wish to take a block–maxima approach, with 'blocks' being years.

For any given year, the 8–dimensional vector of annual maxima is unlikely to be one of the raw multivariate observations.

Let's simplify to the bivariate case: Let $(X_1, Y_1), (X_2, Y_2), \ldots$ be i.i.d. vectors with distribution function $F(x, y)$.

6.2.1 Example: Rainfall at 8 locations in Scotland

Now consider the **componentwise block maxima**

$$M_{x,n} = \max_{i=1,\ldots,n} \{X_i\} \qquad \text{and} \qquad M_{y,n} = \max_{i=1,\ldots,n} \{Y_i\}.$$

We define the **vector of componentwise maxima** to be

$$\mathbf{M_n} = (M_{x,n}, M_{y,n}).$$

$\mathbf{M_n}$ is not necessarily one of the original observations (X_i, Y_i). Nevertheless, we are interested in the limiting behaviour of $\mathbf{M_n}$ as $n \to \infty$.

6.2.1 Example: Rainfall at 8 locations in Scotland

Now consider the **componentwise block maxima**

$$M_{x,n} = \max_{i=1,\ldots,n} \{X_i\} \qquad \text{and} \qquad M_{y,n} = \max_{i=1,\ldots,n} \{Y_i\}.$$

We define the **vector of componentwise maxima** to be

$$\mathbf{M_n} = (M_{x,n}, M_{y,n}).$$

$\mathbf{M_n}$ is not necessarily one of the original observations (X_i, Y_i). Nevertheless, we are interested in the limiting behaviour of $\mathbf{M_n}$ as $n \to \infty$.

6.2.1 Example: Rainfall at 8 locations in Scotland

The first point to note is that **standard univariate extreme value results** apply in each margin.

When considering the dependence, this allows us to make a simplifying assumption.

6.2.1 Example: Rainfall at 8 locations in Scotland

The first point to note is that **standard univariate extreme value results** apply in each margin.

When considering the dependence, this allows us to make a simplifying assumption.

6.2.1 Example: Rainfall at 8 locations in Scotland

We assume that the X_i and Y_i variables have a known marginal distribution. It is convenient to assume a unit Fréchet distribution (see Chapter 2), which has CDF

$$F(z) = \exp(-1/z), \qquad z > 0.$$

This gives rise to a very simple normalisation of maxima:

$$\Pr(X_i < x) = \Pr(M_{x,n}/n < x) = \exp(-1/x), \qquad x > 0,$$

(and similarly for Y_i). So if we consider the re–scaled vector

$$\mathbf{M}_\mathbf{n}^* = \left(\max_{i=1,\ldots,n} \{X_i\}/n, \max_{i=1,\ldots,n} \{Y_i\}/n \right),$$

the margins are unit Fréchet for all n, and hence we can characterise the limiting joint behaviour of $\mathbf{M}_\mathbf{n}^*$ without having to worry about the margins.

6.2.1 Example: Rainfall at 8 locations in Scotland

We assume that the X_i and Y_i variables have a known marginal distribution. It is convenient to assume a unit Fréchet distribution (see Chapter 2), which has CDF

$$F(z) = \exp(-1/z), \qquad z > 0.$$

This gives rise to a very simple normalisation of maxima:

$$\Pr(X_i < x) = \Pr(M_{x,n}/n < x) = \exp(-1/x), \qquad x > 0,$$

(and similarly for Y_i). So if we consider the re–scaled vector

$$\mathbf{M}_\mathbf{n}^* = \left(\max_{i=1,\ldots,n} \{X_i\}/n, \max_{i=1,\ldots,n} \{Y_i\}/n \right),$$

the margins are unit Fréchet for all n, and hence we can characterise the limiting joint behaviour of $\mathbf{M}_\mathbf{n}^*$ without having to worry about the margins.

6.2.1 Example: Rainfall at 8 locations in Scotland

Unfortunately no limiting parametric family exists (for bivariate extremes, or multivariate extremes in general)!

6.2.2 Limiting distributions for bivariate extremes

Let $\mathbf{M}_\mathbf{n}^* = (M_{x,n}^*, M_{y,n}^*)$ be the normalised maxima as above, where the (X_i, Y_i) are i.i.d. with standard Fréchet marginal distributions.

Then if

$$\Pr(M_{x,n}^*, M_{y,n}^*) \to G(x, y),$$

where G is non–degenerate, then G has the form

$$G(x, y) = \exp\{-V(x, y)\}; \quad x > 0, \quad y > 0 \tag{6.1}$$

where:

$$V(x, y) = 2\int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) dH(\omega) \tag{6.2}$$

and H is a distribution function on $[0, 1]$ satisfying the mean constraint:

$$\int_0^1 \omega \, dH(\omega) = 0.5. \tag{6.3}$$

6.2.2 Limiting distributions for bivariate extremes

Let $\mathbf{M_n^*} = (M_{x,n}^*, M_{y,n}^*)$ be the normalised maxima as above, where the (X_i, Y_i) are i.i.d. with standard Fréchet marginal distributions.

Then if

$$\Pr(M_{x,n}^*, M_{y,n}^*) \to G(x, y),$$

where G is non–degenerate, then G has the form

$$G(x, y) = \exp\{-V(x, y)\}; \quad x > 0, \quad y > 0 \tag{6.1}$$

where:

$$V(x, y) = 2\int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) dH(\omega) \tag{6.2}$$

and H is a distribution function on $[0, 1]$ satisfying the mean constraint:

$$\int_0^1 \omega \, dH(\omega) = 0.5. \tag{6.3}$$

6.2.2 Limiting distributions for bivariate extremes

Let $\mathbf{M_n^*} = (M_{x,n}^*, M_{y,n}^*)$ be the normalised maxima as above, where the (X_i, Y_i) are i.i.d. with standard Fréchet marginal distributions.

Then if

$$\Pr(M_{x,n}^*, M_{y,n}^*) \to G(x, y),$$

where G is non–degenerate, then G has the form

$$G(x, y) = \exp\{-V(x, y)\}; \quad x > 0, \quad y > 0 \tag{6.1}$$

where:

$$V(x, y) = 2\int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) dH(\omega) \tag{6.2}$$

and H is a distribution function on $[0, 1]$ satisfying the mean constraint:

$$\int_0^1 \omega \, dH(\omega) = 0.5. \tag{6.3}$$

6.2.2 Limiting distributions for bivariate extremes

Hence the class of bivariate extreme value distributions is in one-to-one correspondence with distribution functions H satisfying the constraint (6.3).

If H is differentiable with density h, then (6.1) becomes

$$V(x, y) = 2 \int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) h(\omega) d\omega.$$

However some simple models arise when H is not differentiable. For example, if H places mass 0.5 on each of $\omega = 0$ and $\omega = 1$, then we get

$$G(x, y) = \exp\{-(x^{-1} + y^{-1})\}, \qquad x > 0, y > 0,$$

corresponding to independent x and y.

6.2.2 Limiting distributions for bivariate extremes

Hence the class of bivariate extreme value distributions is in one-to-one correspondence with distribution functions H satisfying the constraint (6.3).

If H is differentiable with density h, then (6.1) becomes

$$V(x, y) = 2 \int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) h(\omega) d\omega.$$

However some simple models arise when H is not differentiable. For example, if H places mass 0.5 on each of $\omega = 0$ and $\omega = 1$, then we get

$$G(x, y) = \exp\{-(x^{-1} + y^{-1})\}, \qquad x > 0, y > 0,$$

corresponding to independent x and y.

6.2.2 Limiting distributions for bivariate extremes

Hence the class of bivariate extreme value distributions is in one-to-one correspondence with distribution functions H satisfying the constraint (6.3).

If H is differentiable with density h, then (6.1) becomes

$$V(x, y) = 2 \int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) h(\omega) d\omega.$$

However some simple models arise when H is not differentiable. For example, if H places mass 0.5 on each of $\omega = 0$ and $\omega = 1$, then we get

$$G(x, y) = \exp\{-(x^{-1} + y^{-1})\}, \qquad x > 0, y > 0,$$

corresponding to independent x and y.

6.2.2 Limiting distributions for bivariate extremes

Since the GEV provides the complete class of **marginal** limit distributions, then the complete class of **bivariate** extreme value distributions is obtained as follows.

If we suppose X and Y are GEV with parameters (μ_x, σ_x, ξ_x) and (μ_y, σ_y, ξ_y) respectively, then the transformations

$$\tilde{x} = \left[1 + \xi_x \left(\frac{x - \mu_x}{\sigma_x}\right)\right]^{1/\xi_x} \quad \text{and} \quad \tilde{y} = \left[1 + \xi_y \left(\frac{y - \mu_y}{\sigma_y}\right)\right]^{1/\xi_y}$$

obtain unit Fréchet margins.

6.2.2 Limiting distributions for bivariate extremes

Since the GEV provides the complete class of **marginal** limit distributions, then the complete class of **bivariate** extreme value distributions is obtained as follows.

If we suppose X and Y are GEV with parameters (μ_x, σ_x, ξ_x) and (μ_y, σ_y, ξ_y) respectively, then the transformations

$$\tilde{x} = \left[1 + \xi_x \left(\frac{x - \mu_x}{\sigma_x}\right)\right]^{1/\xi_x} \quad \text{and} \quad \tilde{y} = \left[1 + \xi_y \left(\frac{y - \mu_y}{\sigma_y}\right)\right]^{1/\xi_y}$$

obtain unit Fréchet margins.

6.2.2 Limiting distributions for bivariate extremes

Hence

$$G(x, y) = \exp\{-V(\tilde{x}, \tilde{y})\}$$

is a bivariate extreme value distribution with the appropriate margins for valid $V(.)$.

6.3.1 Example: wave–surge data at Newlyn, Cornwall

Here, we choose a different type of example of dependence to the rainfall problem considered in Section 6.2.1.

Specifically, we consider **two** variables recorded concurrently at the **same site**.

A series of 3–hourly measurements on sea–surge were obtained from Newlyn, southwest England, giving, at each time point,

(i) measurements of the **wave height** (in metres)

(ii) measurements of the **surge height** (in metres)

6.3.1 Example: wave–surge data at Newlyn, Cornwall

Here, we choose a different type of example of dependence to the rainfall problem considered in Section 6.2.1.

Specifically, we consider **two** variables recorded concurrently at the **same site**.

A series of 3–hourly measurements on sea–surge were obtained from Newlyn, southwest England, giving, at each time point,

(i) measurements of the **wave height** (in metres)

(ii) measurements of the **surge height** (in metres)

6.3.1 Example: wave–surge data at Newlyn, Cornwall

Here, we choose a different type of example of dependence to the rainfall problem considered in Section 6.2.1.

Specifically, we consider **two** variables recorded concurrently at the **same site**.

A series of 3–hourly measurements on sea–surge were obtained from Newlyn, southwest England, giving, at each time point,

(i) measurements of the **wave height** (in metres)

(ii) measurements of the **surge height** (in metres)

6.3.1 Example: wave–surge data at Newlyn, Cornwall

Here, we choose a different type of example of dependence to the rainfall problem considered in Section 6.2.1.

Specifically, we consider **two** variables recorded concurrently at the **same site**.

A series of 3–hourly measurements on sea–surge were obtained from Newlyn, southwest England, giving, at each time point,

(i) measurements of the **wave height** (in metres)

(ii) measurements of the **surge height** (in metres)

6.3.1 Example: wave–surge data at Newlyn, Cornwall

Here, we choose a different type of example of dependence to the rainfall problem considered in Section 6.2.1.

Specifically, we consider **two** variables recorded concurrently at the **same site**.

A series of 3–hourly measurements on sea–surge were obtained from Newlyn, southwest England, giving, at each time point,

(i) measurements of the **wave height** (in metres)

(ii) measurements of the **surge height** (in metres)

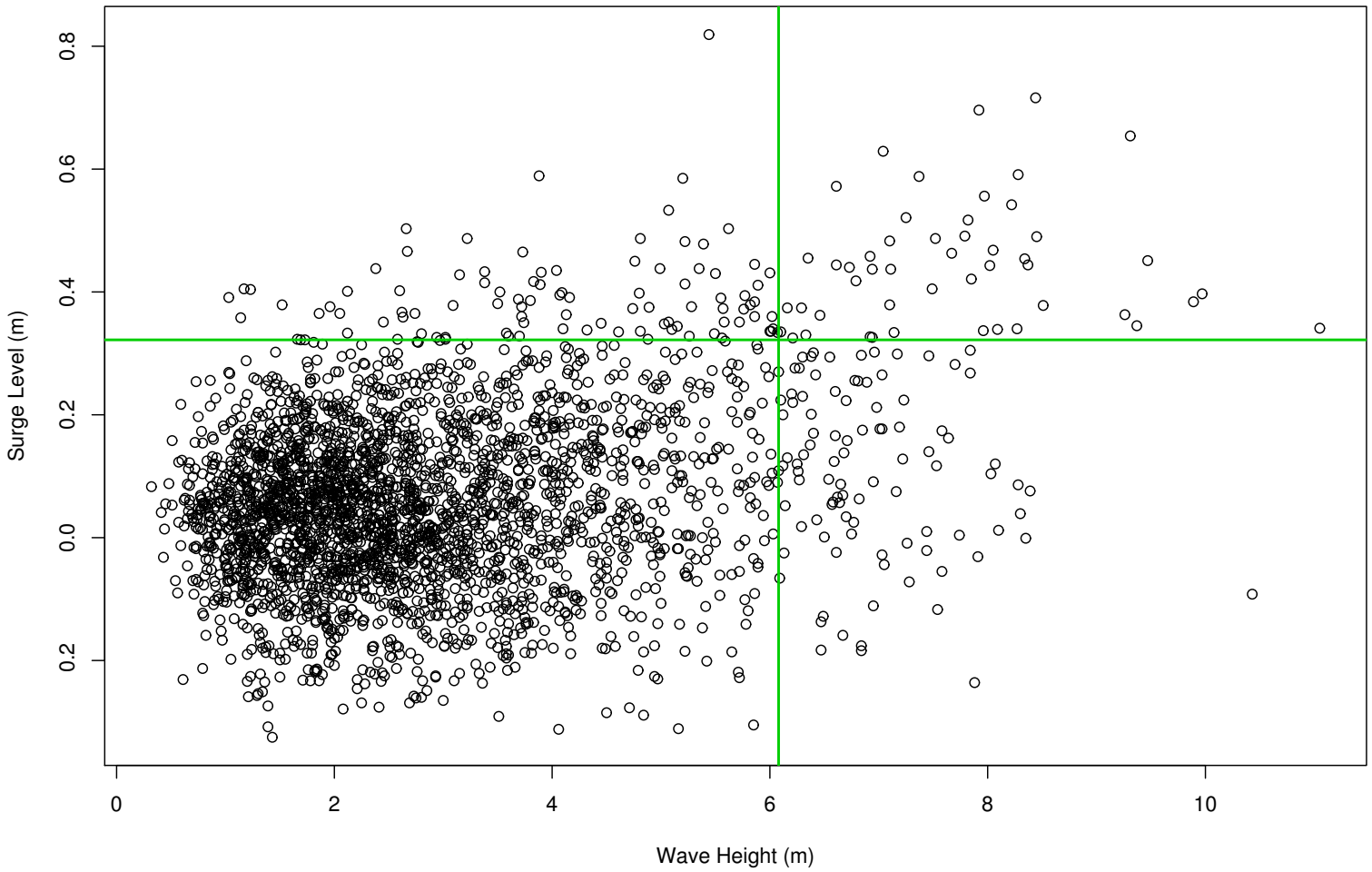
6.3.1 Example: wave–surge data at Newlyn, Cornwall

Figure 6.2 shows these two variables plotted against each other.

This plot suggests a tendency for extremes of one variable to coincide with extremes of the other.

This dependence could be important – the impact of an event that is **simultaneously extreme** in both variables is likely to be much greater than if extremes of either component occurred in isolation.

6.3.1 Example: wave–surge data at Newlyn, Cornwall



6.3.2 Componentwise maxima to threshold excesses

Now we want to define our bivariate extremes as those observations which exceed a threshold in one or other margin.

For our bivariate observation (X, Y), let's focus on X.

We have already seen that the distribution function for the exceedances of a threshold u by a variable X, conditional on $X > u$ for large enough u, is given by:

$$H(x) = 1 - \lambda_{u_x} \left\{ 1 + \frac{\xi_x (x - u_x)}{\sigma_x} \right\}^{-1/\xi_x}$$

defined on $\{x - u_x : x - u_x > 0 \text{ and } (1 + \xi_x (x - u_x) / \sigma_x) > 0\}$, where $\xi_x \neq 0$, $\sigma_x > 0$, and $\lambda_{u_x} = \Pr(X > u_x)$.

6.3.2 Componentwise maxima to threshold excesses

Now we want to define our bivariate extremes as those observations which exceed a threshold in one or other margin.

For our bivariate observation (X, Y), let's focus on X.

We have already seen that the distribution function for the exceedances of a threshold u by a variable X, conditional on $X > u$ for large enough u, is given by:

$$H(x) = 1 - \lambda_{u_x} \left\{ 1 + \frac{\xi_x (x - u_x)}{\sigma_x} \right\}^{-1/\xi_x}$$

defined on $\{x - u_x : x - u_x > 0 \text{ and } (1 + \xi_x (x - u_x) / \sigma_x) > 0\}$, where $\xi_x \neq 0$, $\sigma_x > 0$, and $\lambda_{u_x} = \Pr(X > u_x)$.

6.3.2 Componentwise maxima to threshold excesses

Now we want to define our bivariate extremes as those observations which exceed a threshold in one or other margin.

For our bivariate observation (X, Y), let's focus on X.

We have already seen that the distribution function for the exceedances of a threshold u by a variable X, conditional on $X > u$ for large enough u, is given by:

$$H(x) = 1 - \lambda_{u_x} \left\{ 1 + \frac{\xi_x (x - u_x)}{\sigma_x} \right\}^{-1/\xi_x}$$

defined on $\{x - u_x : x - u_x > 0 \text{ and } (1 + \xi_x (x - u_x) / \sigma_x) > 0\}$, where $\xi_x \neq 0$, $\sigma_x > 0$, and $\lambda_{u_x} = \Pr(X > u_x)$.

6.3.2 Componentwise maxima to threshold excesses

Now we can obtain a unit Fréchet margin with the **transformation**:

$$\tilde{x} = -\left(\log\left\{1 - \lambda_{u_x}\left[1 + \frac{\xi_x\left(x - u_x\right)}{\sigma_x}\right]^{-1/\xi_x}\right\}\right)^{-1}. \qquad (6.4)$$

If we apply the analogous transformation in the Y margin, we obtain

$$G\left(x, y\right) = \exp\left\{-V\left(\tilde{x}, \tilde{y}\right)\right\}; \quad x > u_x, \quad y > u_y,$$

where $V(x, y)$ is as defined in Equation (6.1), again satisfying the mean constraint in Equation (6.3).

6.3.2 Componentwise maxima to threshold excesses

Now we can obtain a unit Fréchet margin with the **transformation**:

$$\tilde{x} = -\left(\log\left\{1 - \lambda_{u_x}\left[1 + \frac{\xi_x\left(x - u_x\right)}{\sigma_x}\right]^{-1/\xi_x}\right\}\right)^{-1}. \qquad (6.4)$$

If we apply the analogous transformation in the Y margin, we obtain

$$G\left(x, y\right) = \exp\left\{-V\left(\tilde{x}, \tilde{y}\right)\right\}; \quad x > u_x, \;\; y > u_y,$$

where $V(x, y)$ is as defined in Equation (6.1), again satisfying the mean constraint in Equation (6.3).

6.4 Modelling bivariate extremes in practice

In practice, modelling usually involves identifying a parametric sub–family with appropriate flexibility to handle the structure inherent in the data.

Models can be fitted, e.g. by maximum–likelihood estimation, either:

- **two steps** (marginal components followed by dependence function), or
- or in a **single sweep**

All of these procedures, including the choice of models, are handled in a very similar way when dealing with either bivariate componentwise maxima or bivariate threshold exceedances.

6.4 Modelling bivariate extremes in practice

In practice, modelling usually involves identifying a parametric sub–family with appropriate flexibility to handle the structure inherent in the data.

Models can be fitted, e.g. by maximum–likelihood estimation, either:

- **two steps** (marginal components followed by dependence function), or
- or in a **single sweep**

All of these procedures, including the choice of models, are handled in a very similar way when dealing with either bivariate componentwise maxima or bivariate threshold exceedances.

6.4 Modelling bivariate extremes in practice

In practice, modelling usually involves identifying a parametric sub–family with appropriate flexibility to handle the structure inherent in the data.

Models can be fitted, e.g. by maximum–likelihood estimation, either:

- **two steps** (marginal components followed by dependence function), or
- or in a **single sweep**

All of these procedures, including the choice of models, are handled in a very similar way when dealing with either bivariate componentwise maxima or bivariate threshold exceedances.

6.4.1 Modelling the dependence structure

The class of bivariate extreme value models contains many families of distributions which can be used to model the dependence structure in the data.

The dependence structure must satisfy the conditions on $H(\omega)$.

Possible choices are:

- **Logistic Model** — symmetric
- Negative Logistic Model
- **Bilogistic Model** — asymmetric
- Dirichlet Model

Here we will focus on the logistic model and the bilogistic model as two commonly used but contrasting choices.

6.4.1 Modelling the dependence structure

The class of bivariate extreme value models contains many families of distributions which can be used to model the dependence structure in the data.

The dependence structure must satisfy the conditions on $H(\omega)$.

Possible choices are:

- **Logistic Model** — symmetric
- Negative Logistic Model
- **Bilogistic Model** — asymmetric
- Dirichlet Model

Here we will focus on the logistic model and the bilogistic model as two commonly used but contrasting choices.

6.4.1 Modelling the dependence structure

The class of bivariate extreme value models contains many families of distributions which can be used to model the dependence structure in the data.

The dependence structure must satisfy the conditions on $H(\omega)$.

Possible choices are:

- **Logistic Model** — symmetric
- Negative Logistic Model
- **Bilogistic Model** — asymmetric
- Dirichlet Model

Here we will focus on the logistic model and the bilogistic model as two commonly used but contrasting choices.

6.4.1 Modelling the dependence structure

Symmetric dependence

- X depends on Y to exactly the same degree that Y depends on X
- They both have the same **influence** over each other

Asymmetric dependence

- X has greater influene over Y than Y has over X (or *vice–versa*)
- Example: Wind speeds at two nearby locations in the U.K.

6.4.1 Modelling the dependence structure

Symmetric dependence

- X depends on Y to exactly the same degree that Y depends on X
- They both have the same **influence** over each other

Asymmetric dependence

- X has greater influene over Y than Y has over X (or *vice–versa*)
- Example: Wind speeds at two nearby locations in the U.K.

6.4.1 Modelling the dependence structure

Symmetric dependence

- X depends on Y to exactly the same degree that Y depends on X
- They both have the same **influence** over each other

Asymmetric dependence

- X has greater influene over Y than Y has over X (or *vice–versa*)
- Example: Wind speeds at two nearby locations in the U.K.

6.4.1 Modelling the dependence structure

Symmetric dependence

- X depends on Y to exactly the same degree that Y depends on X
- They both have the same **influence** over each other

Asymmetric dependence

- X has greater influene over Y than Y has over X (or *vice–versa*)
- Example: Wind speeds at two nearby locations in the U.K.

6.4.1 Modelling the dependence structure

Symmetric dependence

- X depends on Y to exactly the same degree that Y depends on X
- They both have the same **influence** over each other

Asymmetric dependence

- X has greater influene over Y than Y has over X (or *vice–versa*)
- Example: Wind speeds at two nearby locations in the U.K.

1. The Logistic model

Here, for $V(x, y)$ in Equation (6.1), we have

$$\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha},$$

where $x > 0$, $y > 0$ and $\alpha \in (0, 1)$, giving

$$G(x, y) = \exp\left\{-\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}\right\}.$$

- $\alpha \to 1$ corresponds to independent variables.
- $\alpha \to 0$ corresponds to perfectly dependent variables.
- This model is symmetric — the variables are *exchangeable*.

1. The Logistic model

Here, for $V(x, y)$ in Equation (6.1), we have

$$\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha},$$

where $x > 0$, $y > 0$ and $\alpha \in (0, 1)$, giving

$$G(x, y) = \exp\left\{-\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}\right\}.$$

- $\alpha \to 1$ corresponds to independent variables.
- $\alpha \to 0$ corresponds to perfectly dependent variables.
- This model is symmetric — the variables are *exchangeable*.

1. The Logistic model

Here, for $V(x, y)$ in Equation (6.1), we have

$$\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha},$$

where $x > 0$, $y > 0$ and $\alpha \in (0, 1)$, giving

$$G(x, y) = \exp\left\{-\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}\right\}.$$

- $\alpha \to 1$ corresponds to independent variables.
- $\alpha \to 0$ corresponds to perfectly dependent variables.
- This model is symmetric — the variables are *exchangeable*.

1. The Logistic model

Here, for $V(x, y)$ in Equation (6.1), we have

$$\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha},$$

where $x > 0$, $y > 0$ and $\alpha \in (0, 1)$, giving

$$G(x, y) = \exp\left\{-\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}\right\}.$$

- $\alpha \to 1$ corresponds to independent variables.
- $\alpha \to 0$ corresponds to perfectly dependent variables.
- This model is symmetric — the variables are *exchangeable*.

2. The Bilogistic model

Now we have the following form for $V(x, y)$:

$$-x\gamma^{1-\alpha} - y\,(1-\gamma)^{1-\beta},$$

where $0 < \alpha < 1$, $0 < \beta < 1$ and $\gamma = \gamma\,(x, y; \alpha, \beta)$ is the solution of:

$$(1-\alpha)\,x\,(1-\gamma)^{\beta} = (1-\beta)\,y\gamma^{\alpha}$$

- Independence is obtained when $\alpha = \beta \to 1$ and when one of α or β is fixed and the other approaches 1.
- When $\alpha = \beta$ the model reduces to the logistic model.
- The value of $\alpha - \beta$ determines the extent of **asymmetry** in the dependence structure.

2. The Bilogistic model

Now we have the following form for $V(x, y)$:

$$-x\gamma^{1-\alpha} - y\,(1-\gamma)^{1-\beta},$$

where $0 < \alpha < 1$, $0 < \beta < 1$ and $\gamma = \gamma\,(x, y; \alpha, \beta)$ is the solution of:

$$(1-\alpha)\,x\,(1-\gamma)^{\beta} = (1-\beta)\,y\gamma^{\alpha}$$

- Independence is obtained when $\alpha = \beta \to 1$ and when one of α or β is fixed and the other approaches 1.
- When $\alpha = \beta$ the model reduces to the logistic model.
- The value of $\alpha - \beta$ determines the extent of **asymmetry** in the dependence structure.

2. The Bilogistic model

Now we have the following form for $V(x, y)$:

$$-x\gamma^{1-\alpha} - y\,(1-\gamma)^{1-\beta},$$

where $0 < \alpha < 1$, $0 < \beta < 1$ and $\gamma = \gamma\,(x, y; \alpha, \beta)$ is the solution of:

$$(1-\alpha)\,x\,(1-\gamma)^{\beta} = (1-\beta)\,y\gamma^{\alpha}$$

- Independence is obtained when $\alpha = \beta \to 1$ and when one of α or β is fixed and the other approaches 1.
- When $\alpha = \beta$ the model reduces to the logistic model.
- The value of $\alpha - \beta$ determines the extent of **asymmetry** in the dependence structure.

2. The Bilogistic model

Now we have the following form for $V(x, y)$:

$$-x\gamma^{1-\alpha} - y\,(1-\gamma)^{1-\beta},$$

where $0 < \alpha < 1$, $0 < \beta < 1$ and $\gamma = \gamma\,(x, y; \alpha, \beta)$ is the solution of:

$$(1-\alpha)\,x\,(1-\gamma)^{\beta} = (1-\beta)\,y\gamma^{\alpha}$$

- Independence is obtained when $\alpha = \beta \to 1$ and when one of α or β is fixed and the other approaches 1.
- When $\alpha = \beta$ the model reduces to the logistic model.
- The value of $\alpha - \beta$ determines the extent of **asymmetry** in the dependence structure.

6.4.2 Likelihood calculations

After transformation to unit Fréchet margins, we can obtain the probability density function of the chosen dependence model by **differentiation** of Equation (6.1) to give $g(x, y)$.

From this, the **likelihood** can be formed (and maximised) in the usual way.

6.4.2 Likelihood calculations

After transformation to unit Fréchet margins, we can obtain the probability density function of the chosen dependence model by **differentiation** of Equation (6.1) to give $g(x, y)$.

From this, the **likelihood** can be formed (and maximised) in the usual way.

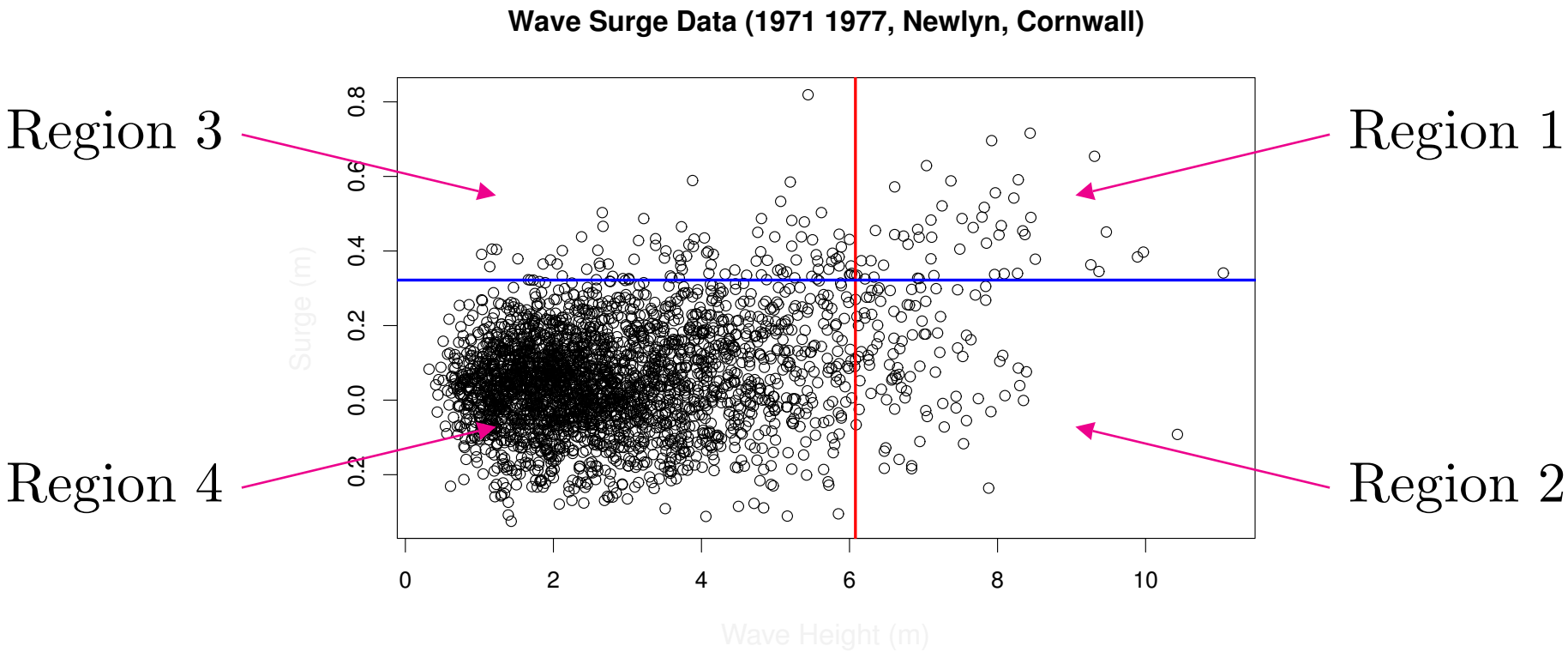
6.4.2 Likelihood calculations

After transformation to unit Fréchet margins, we can obtain the probability density function of the chosen dependence model by **differentiation** of Equation (6.1) to give $g(x, y)$.

From this, the **likelihood** can be formed (and maximised) in the usual way.

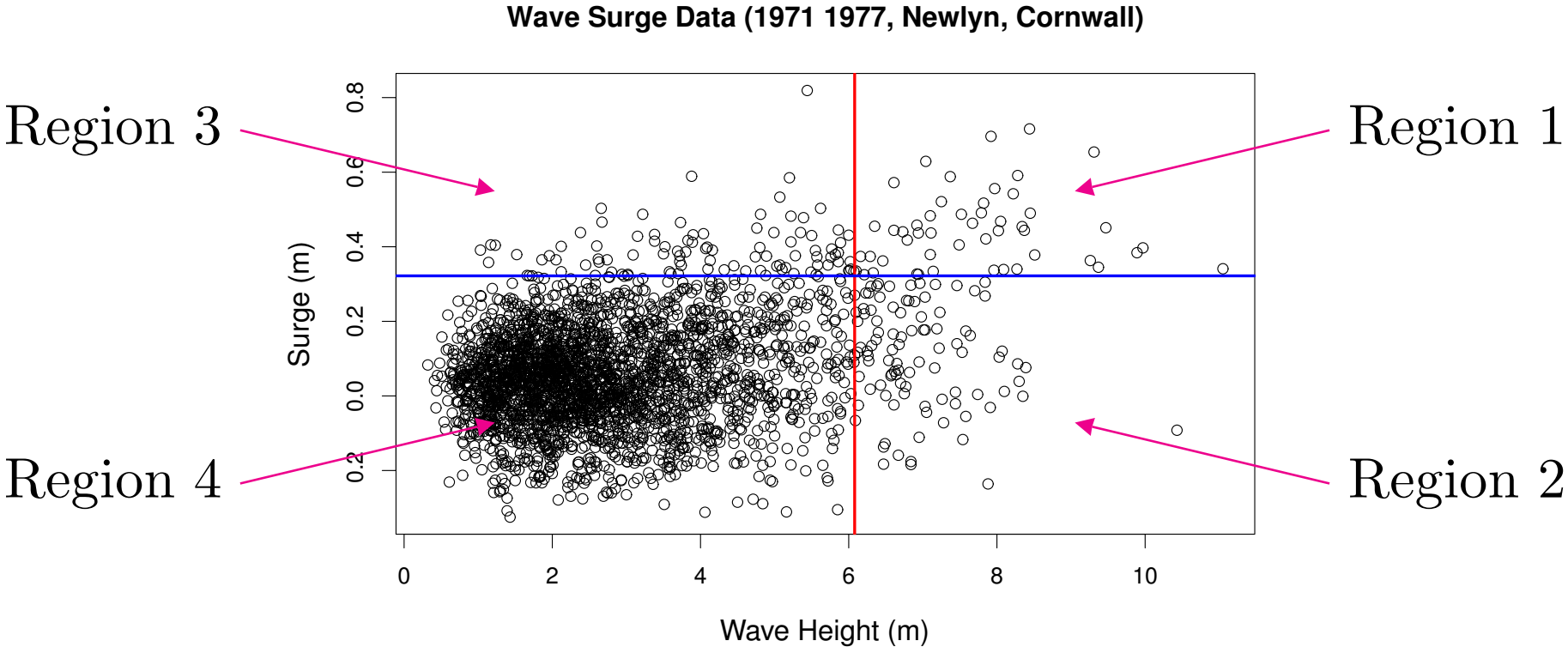
6.4.2 Likelihood calculations

However, inference for the bivariate threshold excess setup is complicated by the fact that a bivariate pair may exceed a specified threshold in just one of its components:



6.4.2 Likelihood calculations

However, inference for the bivariate threshold excess setup is complicated by the fact that a bivariate pair may exceed a specified threshold in just one of its components:



6.4.2 Likelihood calculations

We obtain contributions to the likelihood function for a pair of observations in the following way, where θ represents the parameter(s) in our dependence model:

$$g(x, y; \theta) = \begin{cases} \left.\dfrac{\partial^2 G}{\partial x \partial y}\right|_{(\tilde{x}, \tilde{y})} & \text{if } (x, y) \in \text{Region 1} \\ \left.\dfrac{\partial G}{\partial x}\right|_{(\tilde{x}, \tilde{u}_y)} & \text{if } (x, y) \in \text{Region 2} \\ \left.\dfrac{\partial G}{\partial y}\right|_{(\tilde{u}_x, \tilde{y})} & \text{if } (x, y) \in \text{Region 3} \\ G(\tilde{u}_x, \tilde{u}_y) & \text{if } (x, y) \in \text{Region 4} \end{cases}$$

6.4.2 Likelihood calculations

We obtain contributions to the likelihood function for a pair of observations in the following way, where θ represents the parameter(s) in our dependence model:

$$g(x, y; \theta) = \begin{cases} \left.\dfrac{\partial^2 G}{\partial x \partial y}\right|_{(\tilde{x}, \tilde{y})} & \text{if } (x, y) \in \text{Region 1} \\ \left.\dfrac{\partial G}{\partial x}\right|_{(\tilde{x}, \tilde{u}_y)} & \text{if } (x, y) \in \text{Region 2} \\ \left.\dfrac{\partial G}{\partial y}\right|_{(\tilde{u}_x, \tilde{y})} & \text{if } (x, y) \in \text{Region 3} \\ G(\tilde{u}_x, \tilde{u}_y) & \text{if } (x, y) \in \text{Region 4} \end{cases}$$

6.4.2 Likelihood calculations

Then, we have

$$L(\theta; x, y) = \prod_{i=1}^{n} g(\tilde{x}_i, \tilde{y}_i).$$

Example 6.1: Wave–surge analysis at Newlyn

Consider the wave–surge data from Newlyn, Cornwall, shown in Figures 6.2 and 6.3.

Flood defences in Newlyn have been designed to withstand a sea swell resulting from, at most, a **wave height** of $x = 9$metres combined with a **surge height** of $y = 0.7$metres.

A threshold–based approach to modelling is to be used for the wave–surge data shown in Figures 6.2 and 6.3.

Mean residual life plots suggest marginal thresholds of $u_x = 6.1$metres and $u_y = 0.32$metres (as shown in Figure 6.2) for identifying wave height and surge height as extreme.

Example 6.1: Wave–surge analysis at Newlyn

Consider the wave–surge data from Newlyn, Cornwall, shown in Figures 6.2 and 6.3.

Flood defences in Newlyn have been designed to withstand a sea swell resulting from, at most, a **wave height** of $x = 9$metres combined with a **surge height** of $y = 0.7$metres.

A threshold–based approach to modelling is to be used for the wave–surge data shown in Figures 6.2 and 6.3.

Mean residual life plots suggest marginal thresholds of $u_x = 6.1$metres and $u_y = 0.32$metres (as shown in Figure 6.2) for identifying wave height and surge height as extreme.

Example 6.1: Wave–surge analysis at Newlyn

(a) Assuming extreme wave heights occur **independently** of extreme surge heights, find the probability that the flood defence system in Newlyn will be overwhelmed.

Example 6.1: Solution to part (a)

Assuming independence, we have

Pr(flood defence fails)

$$
\begin{aligned}
&= \Pr(X > 9) \times \Pr(Y > 0.7) \\
&= \hat{\lambda}_{u_x}\left[1 + \hat{\xi}_x\left(\frac{9 - u_x}{\hat{\sigma}_x}\right)\right]_+^{-1/\hat{\xi}_x} \times \hat{\lambda}_{u_y}\left[1 + \hat{\xi}_y\left(\frac{0.7 - u_y}{\hat{\sigma}_y}\right)\right]_+^{-1/\hat{\xi}_y} \\
&= 0.049\left[1 - 0.188\left(\frac{9 - 6.1}{1.334}\right)\right]_+^{1/0.188} \\
&\quad \times 0.051\left[1 - 0.041\left(\frac{0.7 - 0.32}{0.093}\right)\right]_+^{1/0.041} \\
&= 0.002995 \times 0.000583 = 0.00000175.
\end{aligned}
$$

Example 6.1: Solution to part (a)

Assuming independence, we have

Pr(flood defence fails)

$$
\begin{aligned}
&= \Pr(X > 9) \times \Pr(Y > 0.7) \\
&= \hat{\lambda}_{u_x} \left[1 + \hat{\xi}_x \left(\frac{9 - u_x}{\hat{\sigma}_x}\right)\right]_+^{-1/\hat{\xi}_x} \times \hat{\lambda}_{u_y} \left[1 + \hat{\xi}_y \left(\frac{0.7 - u_y}{\hat{\sigma}_y}\right)\right]_+^{-1/\hat{\xi}_y} \\
&= 0.049 \left[1 - 0.188 \left(\frac{9 - 6.1}{1.334}\right)\right]_+^{1/0.188} \\
&\quad \times 0.051 \left[1 - 0.041 \left(\frac{0.7 - 0.32}{0.093}\right)\right]_+^{1/0.041} \\
&= 0.002995 \times 0.000583 = 0.00000175.
\end{aligned}
$$

Example 6.1: Solution to part (a)

Assuming independence, we have

Pr(flood defence fails)

$$
\begin{aligned}
&= \Pr(X > 9) \times \Pr(Y > 0.7) \\
&= \hat{\lambda}_{u_x} \left[1 + \hat{\xi}_x \left(\frac{9 - u_x}{\hat{\sigma}_x}\right)\right]_+^{-1/\hat{\xi}_x} \times \hat{\lambda}_{u_y} \left[1 + \hat{\xi}_y \left(\frac{0.7 - u_y}{\hat{\sigma}_y}\right)\right]_+^{-1/\hat{\xi}_y} \\
&= 0.049 \left[1 - 0.188 \left(\frac{9 - 6.1}{1.334}\right)\right]_+^{1/0.188} \\
&\quad \times 0.051 \left[1 - 0.041 \left(\frac{0.7 - 0.32}{0.093}\right)\right]_+^{1/0.041} \\
&= 0.002995 \times 0.000583 = 0.00000175.
\end{aligned}
$$

Example 6.1: Solution to part (a)

Assuming independence, we have

Pr(flood defence fails)

$$
\begin{aligned}
&= \Pr(X > 9) \times \Pr(Y > 0.7) \\
&= \hat{\lambda}_{u_x} \left[1 + \hat{\xi}_x \left(\frac{9 - u_x}{\hat{\sigma}_x}\right)\right]_+^{-1/\hat{\xi}_x} \times \hat{\lambda}_{u_y} \left[1 + \hat{\xi}_y \left(\frac{0.7 - u_y}{\hat{\sigma}_y}\right)\right]_+^{-1/\hat{\xi}_y} \\
&= 0.049 \left[1 - 0.188 \left(\frac{9 - 6.1}{1.334}\right)\right]_+^{1/0.188} \\
&\quad \times 0.051 \left[1 - 0.041 \left(\frac{0.7 - 0.32}{0.093}\right)\right]_+^{1/0.041} \\
&= 0.002995 \times 0.000583 = 0.00000175.
\end{aligned}
$$

Example 6.1: Solution to part (a)

Assuming independence, we have

Pr(flood defence fails)

$$
\begin{aligned}
&= \Pr(X > 9) \times \Pr(Y > 0.7) \\
&= \hat{\lambda}_{u_x} \left[1 + \hat{\xi}_x \left(\frac{9 - u_x}{\hat{\sigma}_x}\right)\right]_+^{-1/\hat{\xi}_x} \times \hat{\lambda}_{u_y} \left[1 + \hat{\xi}_y \left(\frac{0.7 - u_y}{\hat{\sigma}_y}\right)\right]_+^{-1/\hat{\xi}_y} \\
&= 0.049 \left[1 - 0.188 \left(\frac{9 - 6.1}{1.334}\right)\right]_+^{1/0.188} \\
&\quad \times 0.051 \left[1 - 0.041 \left(\frac{0.7 - 0.32}{0.093}\right)\right]_+^{1/0.041} \\
&= 0.002995 \times 0.000583 = 0.00000175.
\end{aligned}
$$

Example 6.1: Wave–surge analysis at Newlyn

(b) Now assume there is **extremal dependence** between wave height and surge height.

(i) Assuming the logistic model for this dependence, obtain the likelihood contributions to $L(\alpha; x_i, y_i)$ if (1) both $x > u_x$ and $y > u_y$; (2) only $x > u_x$; (3) only $y > u_y$; (4) neither x nor y exceed their marginal thresholds.

Example 6.1: Solution to (b)(i)

For the logistic model, we have

$$G(x, y) = \exp\left\{-\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}\right\}.$$

Region 1:

If $x > u_x$ and $y > u_y$, then we are in **Region 1** of Figure 6.3, and so

$$g(x, y; \alpha) = \frac{\partial^2 G}{\partial x \partial y}.$$

Example 6.1: Solution to (b)(i)

For the logistic model, we have

$$G(x, y) = \exp\left\{-\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}\right\}.$$

Region 1:

If $x > u_x$ and $y > u_y$, then we are in **Region 1** of Figure 6.3, and so

$$g(x, y; \alpha) = \frac{\partial^2 G}{\partial x \partial y}.$$

Example 6.1: Solution to (b)(i)

Differentiating w.r.t. x gives

$$\begin{aligned}\frac{\partial G}{\partial x} &= \exp\left\{-\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha}\right\} \\ &\quad \times -\alpha\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha-1} \times (-1/\alpha)x^{-(1/\alpha+1)} \\ &= \exp\left\{-\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha}\right\}\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha-1} x^{-(1/\alpha-1)}.\end{aligned}$$

Example 6.1: Solution to (b)(i)

Differentiating w.r.t. x gives

$$\begin{aligned}\frac{\partial G}{\partial x} &= \exp\left\{-\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha}\right\} \\ &\quad \times -\alpha\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha-1} \times (-1/\alpha)x^{-(1/\alpha+1)} \\ &= \exp\left\{-\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha}\right\}\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha-1} x^{-(1/\alpha-1)}.\end{aligned}$$

Example 6.1: Solution to (b)(i)

Differentiating w.r.t. x gives

$$\begin{aligned}\frac{\partial G}{\partial x} &= \exp\left\{-\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha}\right\} \\ &\quad \times -\alpha\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha-1} \times (-1/\alpha)x^{-(1/\alpha+1)} \\ &= \exp\left\{-\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha}\right\}\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha-1} x^{-(1/\alpha-1)}.\end{aligned}$$

Example 6.1: Solution to (b)(i)

Differentiating w.r.t. x gives

$$\begin{aligned}\frac{\partial G}{\partial x} &= \exp\left\{-\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha}\right\} \\ &\quad \times -\alpha\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha-1} \times (-1/\alpha)x^{-(1/\alpha+1)} \\ &= \exp\left\{-\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha}\right\}\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha-1} x^{-(1/\alpha-1)}.\end{aligned}$$

Example 6.1: Solution to (b)(i)

Differentiating the result w.r.t. y then gives (after some algebra)

$$\begin{aligned}\frac{\partial^2 G}{\partial x \partial y} &= (xy)^{-(1/\alpha+1)}\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha-2} \\ &\quad \times\left[\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha}-(1-1/\alpha)\right] \\ &\quad \times\exp\left\{-\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha}\right\} \\ &= (\tilde{x}\tilde{y})^{-(1/\alpha+1)}\left(\tilde{x}^{-1/\alpha}+\tilde{y}^{-1/\alpha}\right)^{\alpha-2} \\ &\quad \times\left[\left(\tilde{x}^{-1/\alpha}+\tilde{y}^{-1/\alpha}\right)^{\alpha}-(1-1/\alpha)\right] \\ &\quad \times\exp\left\{-\left(\tilde{x}^{-1/\alpha}+\tilde{y}^{-1/\alpha}\right)^{\alpha}\right\},\end{aligned}$$

evaluated at $x = \tilde{x}$ and $y = \tilde{y}$.

Example 6.1: Solution to (b)(i)

Differentiating the result w.r.t. y then gives (after some algebra)

$$\begin{aligned}\frac{\partial^2 G}{\partial x \partial y} &= (xy)^{-(1/\alpha+1)}\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha-2} \\ &\quad \times\left[\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha}-(1-1/\alpha)\right] \\ &\quad \times\exp\left\{-\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha}\right\}\end{aligned}$$

$$\begin{aligned}&= (\tilde{x}\tilde{y})^{-(1/\alpha+1)}\left(\tilde{x}^{-1/\alpha}+\tilde{y}^{-1/\alpha}\right)^{\alpha-2} \\ &\quad \times\left[\left(\tilde{x}^{-1/\alpha}+\tilde{y}^{-1/\alpha}\right)^{\alpha}-(1-1/\alpha)\right] \\ &\quad \times\exp\left\{-\left(\tilde{x}^{-1/\alpha}+\tilde{y}^{-1/\alpha}\right)^{\alpha}\right\},\end{aligned}$$

evaluated at $x = \tilde{x}$ and $y = \tilde{y}$.

Example 6.1: Solution to (b)(i)

Differentiating the result w.r.t. y then gives (after some algebra)

$$\begin{aligned}\frac{\partial^2 G}{\partial x \partial y} &= (xy)^{-(1/\alpha+1)}\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha-2} \\ &\quad \times\left[\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha}-(1-1/\alpha)\right] \\ &\quad \times\exp\left\{-\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha}\right\} \\ &= (\tilde{x}\tilde{y})^{-(1/\alpha+1)}\left(\tilde{x}^{-1/\alpha}+\tilde{y}^{-1/\alpha}\right)^{\alpha-2} \\ &\quad \times\left[\left(\tilde{x}^{-1/\alpha}+\tilde{y}^{-1/\alpha}\right)^{\alpha}-(1-1/\alpha)\right] \\ &\quad \times\exp\left\{-\left(\tilde{x}^{-1/\alpha}+\tilde{y}^{-1/\alpha}\right)^{\alpha}\right\},\end{aligned}$$

evaluated at $x = \tilde{x}$ and $y = \tilde{y}$.

Example 6.1: Solution to (b)(i)

Differentiating the result w.r.t. y then gives (after some algebra)

$$\begin{aligned}
\frac{\partial^2 G}{\partial x \partial y} &= (xy)^{-(1/\alpha+1)} \left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha-2} \\
&\quad \times \left[\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha} - (1 - 1/\alpha)\right] \\
&\quad \times \exp\left\{-\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}\right\} \\
&= (\tilde{x}\tilde{y})^{-(1/\alpha+1)} \left(\tilde{x}^{-1/\alpha} + \tilde{y}^{-1/\alpha}\right)^{\alpha-2} \\
&\quad \times \left[\left(\tilde{x}^{-1/\alpha} + \tilde{y}^{-1/\alpha}\right)^{\alpha} - (1 - 1/\alpha)\right] \\
&\quad \times \exp\left\{-\left(\tilde{x}^{-1/\alpha} + \tilde{y}^{-1/\alpha}\right)^{\alpha}\right\},
\end{aligned}$$

evaluated at $x = \tilde{x}$ and $y = \tilde{y}$.

Example 6.1: Solution to (b)(i)

Region 2:

If only $x > u_x$, then we are in **Region 2** of Figure 6.3, and so

$$g(x, y; \alpha) = \frac{\partial G}{\partial x}.$$

Example 6.1: Solution to (b)(i)

Region 2:

If only $x > u_x$, then we are in **Region 2** of Figure 6.3, and so

$$g(x, y; \alpha) = \frac{\partial G}{\partial x}.$$

Example 6.1: Solution to (b)(i)

Region 2:

If only $x > u_x$, then we are in **Region 2** of Figure 6.3, and so

$$g(x, y; \alpha) = \frac{\partial G}{\partial x}.$$

Example 6.1: Solution to (b)(i)

Thus, we have

$$\begin{aligned}\frac{\partial G}{\partial x} &= \exp\left\{-\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha}\right\}\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha-1}x^{-(1/\alpha+1)} \\ &= \exp\left\{-\left(\tilde{x}^{-1/\alpha}+\tilde{u}_y^{-1/\alpha}\right)^{\alpha}\right\}\left(\tilde{x}^{-1/\alpha}+\tilde{u}_y^{-1/\alpha}\right)^{\alpha-1}\tilde{x}^{-(1/\alpha+1)}\end{aligned}$$

evaluated at $x = \tilde{x}$ and $y = \tilde{u}_y$.

Example 6.1: Solution to (b)(i)

Thus, we have

$$\begin{aligned}\frac{\partial G}{\partial x} &= \exp\left\{-\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha}\right\}\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha-1} x^{-(1/\alpha+1)} \\ &= \exp\left\{-\left(\tilde{x}^{-1/\alpha}+\tilde{u}_y^{-1/\alpha}\right)^{\alpha}\right\}\left(\tilde{x}^{-1/\alpha}+\tilde{u}_y^{-1/\alpha}\right)^{\alpha-1} \tilde{x}^{-(1/\alpha+1)}\end{aligned}$$

evaluated at $x = \tilde{x}$ and $y = \tilde{u}_y$.

Example 6.1: Solution to (b)(i)

Thus, we have

$$\begin{aligned}\frac{\partial G}{\partial x} &= \exp\left\{-\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha}\right\}\left(x^{-1/\alpha}+y^{-1/\alpha}\right)^{\alpha-1}x^{-(1/\alpha+1)} \\ &= \exp\left\{-\left(\tilde{x}^{-1/\alpha}+\tilde{u}_y^{-1/\alpha}\right)^{\alpha}\right\}\left(\tilde{x}^{-1/\alpha}+\tilde{u}_y^{-1/\alpha}\right)^{\alpha-1}\tilde{x}^{-(1/\alpha+1)}\end{aligned}$$

evaluated at $x = \tilde{x}$ and $y = \tilde{u}_y$.

Example 6.1: Solution to (b)(i)

Region 3:

If only $y > u_y$, then we are in **Region 3** of Figure 6.3, and so

$$g(x, y; \alpha) = \frac{\partial G}{\partial y}.$$

Example 6.1: Solution to (b)(i)

Thus, in a similar fashion, we have:

$$\frac{\partial G}{\partial y} = \exp\left\{-\left(\tilde{u}_x^{-1/\alpha} + \tilde{y}^{-1/\alpha}\right)^{\alpha}\right\}\left(\tilde{u}_x^{-1/\alpha} + \tilde{y}^{-1/\alpha}\right)^{\alpha-1}\tilde{y}^{-(1/\alpha+1)}$$

evaluated at $x = \tilde{u}_x$ and $y = \tilde{y}$.

Example 6.1: Solution to (b)(i)

Thus, in a similar fashion, we have:

$$\frac{\partial G}{\partial y} = \exp\left\{-\left(\tilde{u}_x^{-1/\alpha} + \tilde{y}^{-1/\alpha}\right)^{\alpha}\right\}\left(\tilde{u}_x^{-1/\alpha} + \tilde{y}^{-1/\alpha}\right)^{\alpha-1}\tilde{y}^{-(1/\alpha+1)}$$

evaluated at $x = \tilde{u}_x$ and $y = \tilde{y}$.

Example 6.1: Solution to (b)(i)

Region 4:

If both x and y are sub–threshold, then we are in **Region 4** of Figure 6.3. Thus,

$$g(x, y; \alpha) = G(\tilde{u}_x, \tilde{u}_y) = \exp\left\{-\left(\tilde{u}_x^{-1/\alpha} + \tilde{u}_y^{-1/\alpha}\right)^{\alpha}\right\}.$$

Example 6.1: Solution to (b)(i)

Region 4:

If both x and y are sub–threshold, then we are in **Region 4** of Figure 6.3. Thus,

$$g(x, y; \alpha) = G(\tilde{u}_x, \tilde{u}_y) = \exp\left\{-\left(\tilde{u}_x^{-1/\alpha} + \tilde{u}_y^{-1/\alpha}\right)^{\alpha}\right\}.$$

Example 6.1: Wave–surge analysis at Newlyn

(ii) Use the evd function `fbvpot` to fit the logistic model with likelihood contributions you identified in part (i). Show your estimated dependence parameter α, with it's standard error.

[Demo in R]

Example 6.1: Wave–surge analysis at Newlyn

(ii) Use the evd function `fbvpot` to fit the logistic model with likelihood contributions you identified in part (i). Show your estimated dependence parameter α, with it's standard error.

[Demo in R*]*

Example 6.1: Wave–surge analysis at Newlyn

(iii) Using your fitted model in (ii), find the probability that the flood defence system in Newlyn will be overwhelmed. Compare your answer to that in part (a).

Example 6.1: Solution to part (b)(iii)

Assuming extremal dependence, we have

Pr(Flood defence fails) =

$$1 - G\left(\widetilde{9}, \widetilde{0.7}\right)$$

$$= 1 - \exp\left\{-\left(\widetilde{9}^{-1/0.764} + \widetilde{0.7}^{-1/0.764}\right)^{0.764}\right\}.$$

Using our estimated marginal GPD parameters from the output above, and the transformation given by Equation (6.4), we have

$$\widetilde{9} = 321.4721 \qquad \widetilde{0.7} = 1161.417.$$

Example 6.1: Solution to part (b)(iii)

Assuming extremal dependence, we have

Pr(Flood defence fails) =

$$1 - G\left(\widetilde{9}, \widetilde{0.7}\right)$$

$$= \quad 1 - \exp\left\{-\left(\widetilde{9}^{-1/0.764} + \widetilde{0.7}^{-1/0.764}\right)^{0.764}\right\}.$$

Using our estimated marginal GPD parameters from the output above, and the transformation given by Equation (6.4), we have

$$\widetilde{9} = 321.4721 \qquad \widetilde{0.7} = 1161.417.$$

Example 6.1: Solution to part (b)(iii)

Assuming extremal dependence, we have

Pr(Flood defence fails) =

$$1 - G\left(\widetilde{9}, \widetilde{0.7}\right)$$

$$= \quad 1 - \exp\left\{-\left(\widetilde{9}^{-1/0.764} + \widetilde{0.7}^{-1/0.764}\right)^{0.764}\right\}.$$

Using our estimated marginal GPD parameters from the output above, and the transformation given by Equation (6.4), we have

$$\widetilde{9} = 321.4721 \qquad \widetilde{0.7} = 1161.417.$$

Example 6.1: Solution to part (b)(iii)

This gives, on substitution into the above expression,

Pr(Flood defence fails)

$$= 1 - \exp\left\{-\left(321.4721^{-1/0.764} + 1161.417^{-1/0.764}\right)^{0.764}\right\} = 0.00354.$$

Although this probability is small:

- It's more than 2000 times larger than the probability obtained assuming independence!
- Assuming independence gives us a **false sense of security** $\longrightarrow$ **grossly underestimates** the chance of the flood defence being overwhelmed
- This is typical of what happens when we ignore extremal dependence!

Example 6.1: Solution to part (b)(iii)

This gives, on substitution into the above expression,

Pr(Flood defence fails)

$$= 1 - \exp\left\{-\left(321.4721^{-1/0.764} + 1161.417^{-1/0.764}\right)^{0.764}\right\} = 0.00354.$$

Although this probability is small:

- It's more than 2000 times larger than the probability obtained assuming independence!
- Assuming independence gives us a **false sense of security** $\longrightarrow$ **grossly underestimates** the chance of the flood defence being overwhelmed
- This is typical of what happens when we ignore extremal dependence!

Example 6.1: Solution to part (b)(iii)

This gives, on substitution into the above expression,

Pr(Flood defence fails)

$$= \quad 1 - \exp\left\{-\left(321.4721^{-1/0.764} + 1161.417^{-1/0.764}\right)^{0.764}\right\} = 0.00354.$$

Although this probability is small:

- It's more than 2000 times larger than the probability obtained assuming independence!
- Assuming independence gives us a **false sense of security** $\longrightarrow$ **grossly underestimates** the chance of the flood defence being overwhelmed
- This is typical of what happens when we ignore extremal dependence!

Example 6.1: Solution to part (b)(iii)

This gives, on substitution into the above expression,

Pr(Flood defence fails)

$$= \quad 1 - \exp\left\{-\left(321.4721^{-1/0.764} + 1161.417^{-1/0.764}\right)^{0.764}\right\} = 0.00354.$$

Although this probability is small:

- It's more than 2000 times larger than the probability obtained assuming independence!
- Assuming independence gives us a **false sense of security** $\longrightarrow$ **grossly underestimates** the chance of the flood defence being overwhelmed
- This is typical of what happens when we ignore extremal dependence!

Example 6.1: Solution to part (b)(iii)

This gives, on substitution into the above expression,

Pr(Flood defence fails)

$$= \quad 1 - \exp\left\{-\left(321.4721^{-1/0.764} + 1161.417^{-1/0.764}\right)^{0.764}\right\} = 0.00354.$$

Although this probability is small:

- It's more than 2000 times larger than the probability obtained assuming independence!
- Assuming independence gives us a **false sense of security** $\longrightarrow$ **grossly underestimates** the chance of the flood defence being overwhelmed
- This is typical of what happens when we ignore extremal dependence!

Example 6.1: Solution to part (b)(iii)

This gives, on substitution into the above expression,

Pr(Flood defence fails)

$$= \quad 1 - \exp\left\{-\left(321.4721^{-1/0.764} + 1161.417^{-1/0.764}\right)^{0.764}\right\} = 0.00354.$$

Although this probability is small:

- It's more than 2000 times larger than the probability obtained assuming independence!
- Assuming independence gives us a **false sense of security** $\longrightarrow$ **grossly underestimates** the chance of the flood defence being overwhelmed
- This is typical of what happens when we ignore extremal dependence!

Example 6.1: Solution to part (b)(iii)

This gives, on substitution into the above expression,

Pr(Flood defence fails)

$$= \quad 1 - \exp\left\{-\left(321.4721^{-1/0.764} + 1161.417^{-1/0.764}\right)^{0.764}\right\} = 0.00354.$$

Although this probability is small:

- It's more than 2000 times larger than the probability obtained assuming independence!
- Assuming independence gives us a **false sense of security** $\longrightarrow$ **grossly underestimates** the chance of the flood defence being overwhelmed
- This is typical of what happens when we ignore extremal dependence!

Example 6.1: Solution to part (b)(iii)

This gives, on substitution into the above expression,

Pr(Flood defence fails)

$$= \quad 1 - \exp\left\{-\left(321.4721^{-1/0.764} + 1161.417^{-1/0.764}\right)^{0.764}\right\} = 0.00354.$$

Although this probability is small:

- It's more than 2000 times larger than the probability obtained assuming independence!
- Assuming independence gives us a **false sense of security** $\longrightarrow$ **grossly underestimates** the chance of the flood defence being overwhelmed
- This is typical of what happens when we ignore extremal dependence!

Example 6.1: Wave–surge analysis at Newlyn

(c) Check for the presence of asymmetry in the dependence structure.

[Demo in R*]*