Extreme Value Statistics

MINES Paris PSL
Athens Network

March 14–18, 2022

Course details

Lecturers:

- Anthony DAVISON (EPFL),
- Raphaël DE FONDEVILLE (Federal Statistical Office),
- Thomas OPITZ (INRAE),
- Hans WACKERNAGEL (MINES ParisTech)

Time schedule: from 9h to 12h30, 14h-17h30

(with lunch break from 12h30-14h)

Date	Program subjects
Mon 14 March	Introduction to extreme value statistics
Tue 15March	Bivariate extremes: extremal dependence
Wed 16 March	Multivariate extremes
Thu 17 March	Spatial extremes
Fri 18 March	Data science, machine learning and extreme value theory

Course details

Practicals: participants bring their own laptop and
will use dedicated packages in the R language
(free download at **www.r-project.org**).
Also needed: the RStudio Desktop
which is freely available at
www.rstudio.com

Web platform: Renku https://renkulab.io

Course location: Mines Paris PSL

Contacts: Hans WACKERNAGEL, Raphaël DE FONDEVILLE

hans.wackernagel@minesparis.psl.eu

raphael.defondeville@bfs.admin.ch

Evaluation by project

Several space-time data sets:

Different subsets will be proposed to pairs of participants.

Study groups:

A list will be established and participants should form groups of two that will hand out a common report.

Deadline for report:

Thursday 24 March 2022

Climate extremes

IPCC 2001 definition:

- an extreme weather event is an event that is rare within its statistical reference distribution at a particular place.
- Definitions of "rare" vary, but an extreme weather event would normally be as rare or rarer than the 10th or 90th percentile.

Objective: estimate the probability of events that lie in the *tail of the distribution*.

Financial extremes

Example: Dow Jones index

Source: `www.spiegel.de`

On *black monday* 29 september 2008 the Dow Jones had its largest historical daily drop of -777,65 points: -6.98%.

Financial extremes: DJI for last 30 years

Example: Dow Jones index

Motivation: extreme value analysis

- **Extreme value analysis**
  We need a rationale to compute the probability and level of events
  within and beyond the range of past measurements.
- **Frequency and intensity of extremes**
  Return period:
  - how likely is the advent of an unusual weather event of a given type within the next month/year/decade/century?
  Return level:
  - what level could the event reach as compared to past events?

Need to assess:

- **Ecological consequences**
- **Socio-economic consequences**

Rationale of extreme value statistics

Simplest case:

- i.i.d. random variables $X_1, ..., X_n$ following a distribution F.
- Require accurate inferences on tail of F .

Key issues:

- there are very few observations in the tail of the distribution;
- estimates are often required beyond the largest observed data value;
- standard density estimation techniques fit well where the data have greatest density, but tend to be biased in estimating tail probabilities.

Lack of physical or empirical basis for extrapolation leads to the *extreme value paradigm*:

Use tail models that are based on asymptotically-motivated distributions.

Software: R

Main site: `http://www.r-project.org`

- A powerful high-level language for statistical exploration, analysis and modeling of large data sets.
- Simple incorporation of C, Python, Fortran code.
- Available for Linux, Windows and Mac systems.
- Public domain software with intense development activity
- A large community and a great many contributed packages:
  e.g. have a look at the **Task View** *Extreme Value Analysis*
  on the **CRAN** site `https://cran.r-project.org/`

Aims of the Course

- Introduction to statistical modelling of extreme values
- Discussion of applications in climate, environmental science and insurance/finance
- Introduction to R and a few extreme value analysis packages (`ismev, evir,`...)
- Some computer practice on standard data sets

References

1. The main reference for this course is the introductory text by Stuart Coles [3], which has many application examples using data sets from climate, environmental science and finance. They can easily be recomputed with the R package `ismev`.
2. The book by Beirlant et al. is good as a second reading [2] and contains also application examples from environmental science, insurance and finance (there is a Web page providing the data sets).
3. The book by Embrechts et al. is focussed on insurance and finance applications.[6]

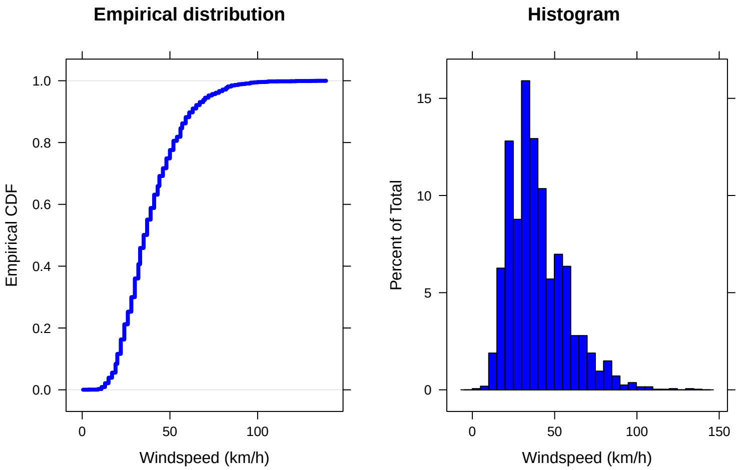
Exploratory Data Analysis

We review a few tools available in R (www.r-project.org) for generating statistical graphics using the package **lattice**. We load **lattice** and the package **latticeExtra** available on CRAN:

```
install.packages(latticeExtra)
require(lattice)
require(latticeExtra)
Windspeed=scan("zaventem.txt")  # load Zaventem windspeeds
```

Daily maximal windspeed

Zaventem airport (1985-1992)



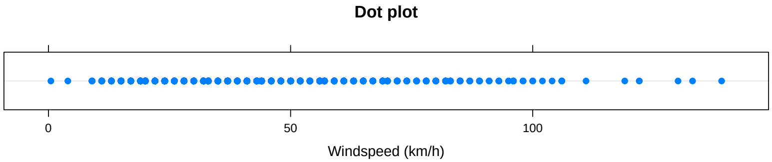
Plots of:

- empirical probatility distribution function
- histogram (empirical density function)

Dot plot and box plot

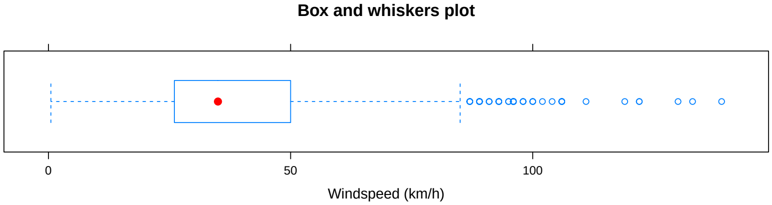
Daily maximal windspeed: Zaventem airport (1985-1992)

Ordered values along a line:



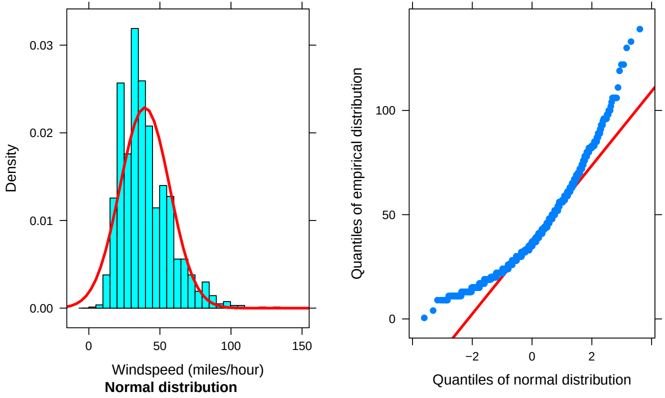
Five numbers that characterize the distribution:

Min.	1st Quart.	Median	3rd Quart.	Max.
0.5	26.0	35.0	50.0	139.0



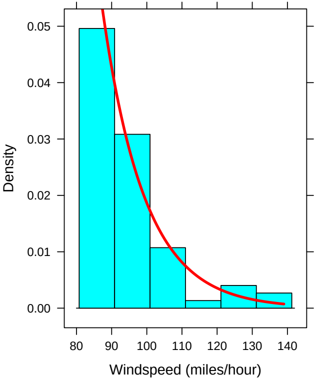
Normal distribution and QQ-plot

Daily maximal windspeed: Zaventem airport (1985-1992)

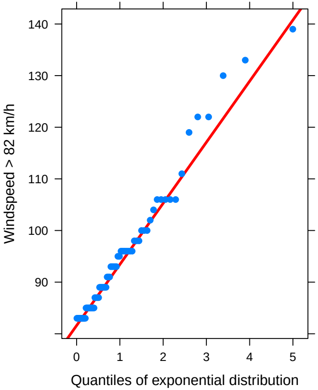


Exponential distribution and QQ-plot for tail

Daily maximal windspeed: Zaventem airport (1985-1992)



Conditional exponential distribution



Exploratory Extreme Value Analysis

(following the book by Beirlant et al.)

Statistical Model

Model: samples are drawn from
independent identically distributed random variables
(iid model):

$$
\begin{array}{cccl}
X_1 & \dots & X_n & \text{iid random variables} \\
\downarrow & & \downarrow & \\
x_1 & \dots & x_n & \text{samples}
\end{array}
$$

We consider ordered samples (drawn from ordered RV):

$$X_1 \leq X_2 \leq X_3 \leq \cdots \leq X_n$$

Distribution function

The distribution function:

$$F(x) = P(X \leq x)$$

Its inverse is the quantile function:

$$Q(p) = \inf\{x : F(x) \geq p\}$$

i.e. the smallest x for which $F(x) \geq p$.

Example: maximal wind speed and wind shed

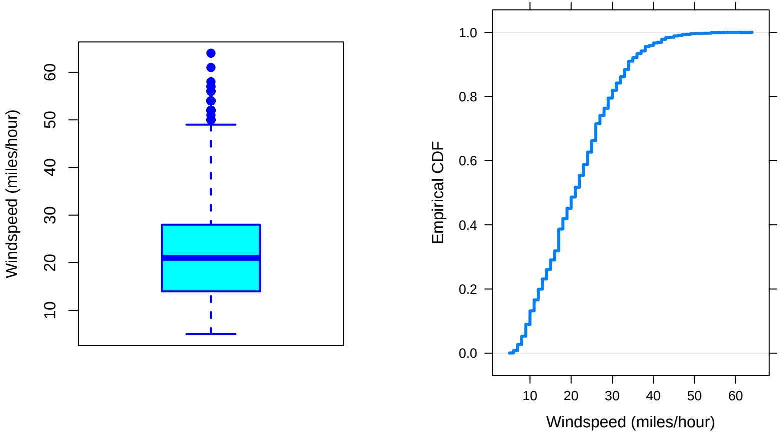
Albuquerque (USA)

Direct problem: a shed breaks down if the wind speed is larger than 30 miles per hour. We are interested in the probability of that event, $p = P(X > 30)$.

To solve it, we compute the empirical distribution function:

$$\widehat{F}_n(x) = \frac{i}{n} \quad \text{if } x \in [x_i, x_{i+1})$$

Albuquerque daily maximal wind speeds



Wind shed: breakdown probability

Albuquerque daily max windspeeds

Daily fastest-mile wind speed data:

$$\widehat{p} = 1 - \widehat{F}_n(30) = 0.18$$

i.e. there is one chance out of five that it breaks down.

- The shed will break down one day out of five in a city like Albuquerque.

Wind shed breakdown

Albuquerque daily max windspeeds

- The shed should resist strong winds, perhaps the strongest wind that might occur.

Problem: winds may occur that are stronger than what was ever measured.

Modeling problem: how can we compute the probability for an event x for $x > x_{max}$, i.e. when x is larger than the largest ever measured value?

The empirical distribution function is of no help as $\widehat{p} = 0$ for $x > x_{max}$. This corresponds actually to the narrow-minded statement: "*it was never measured so far, so it is impossible*".

We need a mathematical model for the distribution of wind-speed maxima, i.e. we need extreme value theory.

Wind shed breakdown

Albuquerque daily max windspeeds

- The shed should resist strong winds, perhaps the strongest wind that might occur.

Problem: winds may occur that are stronger than what was ever measured.

Modeling problem: how can we compute the probability for an event x for $x > x_{max}$, i.e. when x is larger than the largest ever measured value?

The empirical distribution function is of no help as $\widehat{p} = 0$ for $x > x_{max}$. This corresponds actually to the narrow-minded statement: "*it was never measured so far, so it is impossible*".

We need a mathematical model for the distribution of wind-speed maxima, i.e. we need extreme value theory.

Quantile-Quantile plots

Distribution F and its inverse Q

Direct problem: what is the probability $p = 1 - F(x)$ of a wind with speed x?

Inverse problem: which wind speed $x = Q(p)$ corresponds to a given p?

Model: exponential distribution

Exponential distribution: plays an important role in extreme value theory.

- Standard exponential distribution:

$$F(x) = 1 - \exp(-\lambda x) \qquad x > 0, \lambda > 0$$

- Survival function:

$$1 - F(x)$$

- Exponential survival function:

$$\exp(-\lambda x)$$

Quantile function

Exponential distribution

The quantile function for the exponential distribution is:

$$Q_\lambda(p) = -\frac{1}{\lambda}\log(1-p) \qquad 0 < p < 1$$

Linear relation with standard exponential quantiles Q_1:

$$Q_\lambda(p) = \frac{1}{\lambda}\, Q_1(p) \qquad 0 < p < 1$$

Quantile-Quantile plot

Exponential Quantiles against Empirical Quantiles

Given a set of ordered samples $x_1, \ldots, x_n$ the empirical quantiles are:

$$\widehat{Q}_n(p) = \inf\{x : \widehat{F}_n(x) \geq p\}$$

In the QQ-plot they are plotted against exponential quantiles.

Abscissa: the standard exponential quantiles $-\log(1-p)$

Ordinate: the empirical quantiles $\widehat{Q}_n(p)$

QQ-plot: estimation of extreme quantiles

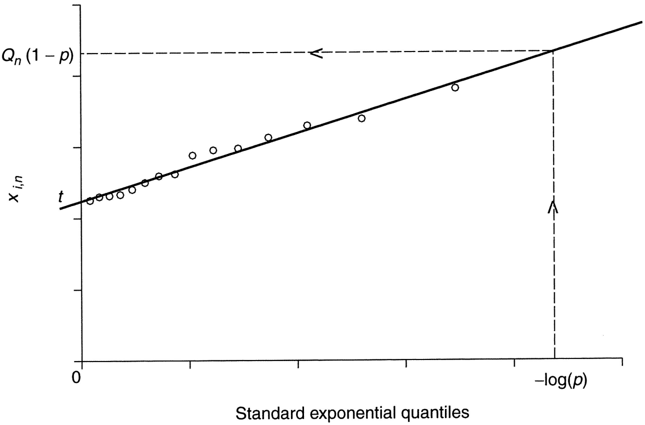


Figure 1.2 Exponential QQ-plot: estimation of extreme quantiles.

QQ-plot: exponential model

- We expect a straight line if the exponential model provides a plausible fit.
- For a straight line pattern, the slope of the fitted straight line is an estimate of λ^{-1}as:
$$Q_\lambda(p) = -\frac{1}{\lambda}\log(1-p)$$
with $p \in \{\frac{1}{n+1}, \frac{2}{n+1}, \dots, \frac{n}{n+1}\}$.
- The slope λ^{-1} can be estimated by least-squares.

Exponential QQ-plot: threshold t

If data is only available above a threshold t, we handle the conditional distribution of X (given $X > t$):

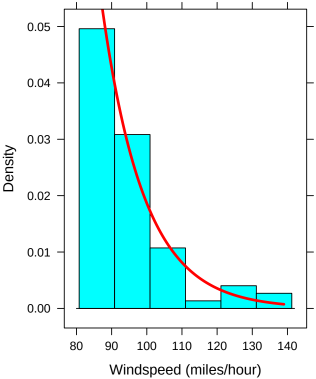
$$P(X > x \mid X > t) = \frac{P(X > x)}{P(X > t)} = \exp(-\lambda(x - t)) \qquad x > t$$

and the corresponding quantile function is:

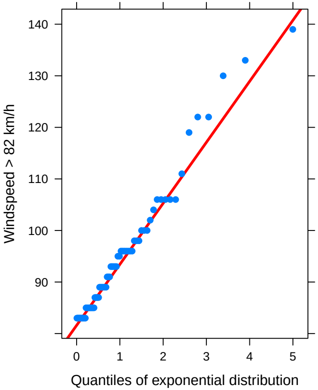
$$Q_{\lambda,t}(p) = t - \frac{1}{\lambda}\log(1 - p) \qquad 0 < p < 1$$

where t is the intercept of the straight line

Zaventem: maximal wind speeds



Conditional exponential distribution

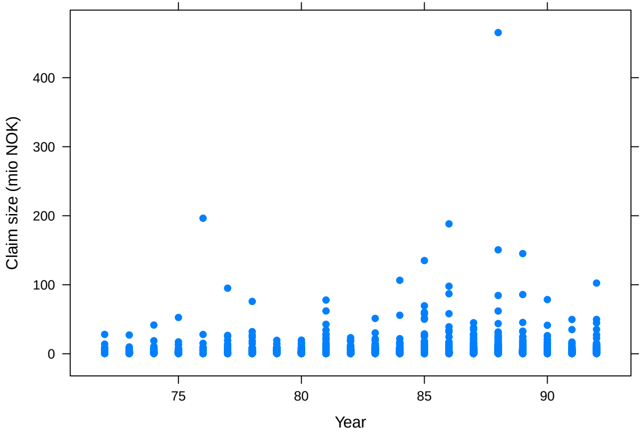


Pareto QQ-plot

Distribution	$F(x)$	Abscissa	Ordinate
Exponential	$1-\exp(-\lambda x)$ $_{x>0,\lambda>0}$	$-\log(1-p_{i,n})$	$x_{i,n}$
Pareto	$1-x^{-\alpha}$ $_{x>1,\alpha>0}$	$-\log(1-p_{i,n})$	$\log x_{i,n}$

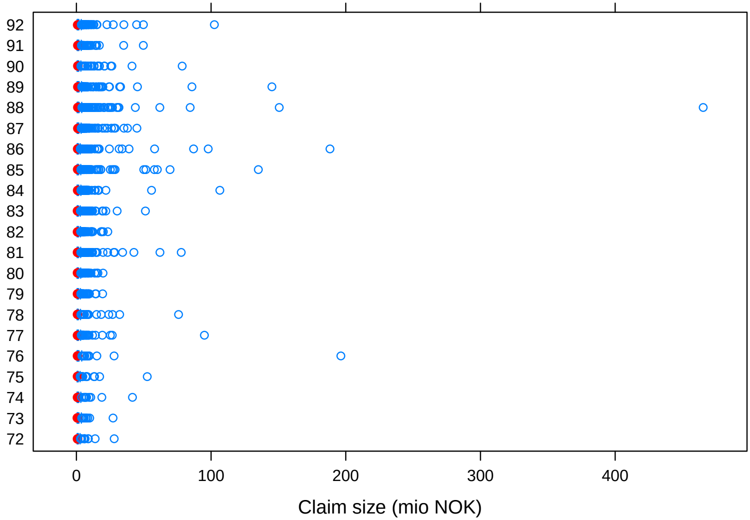
Pareto QQ-plot: obtained by log-transforming
the ordinate of an exponential QQ-plot.

Norwegian fire insurance: max. claim sizes



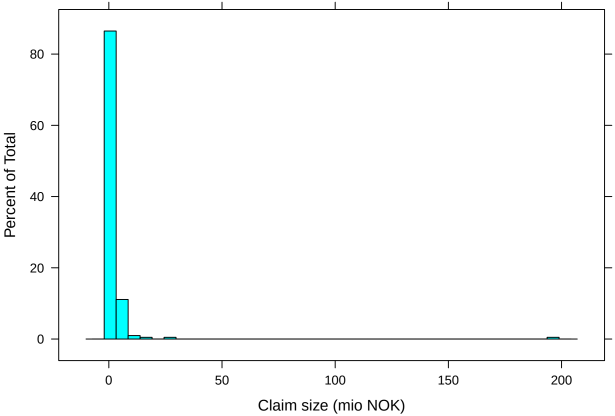
Maximal claim sizes: box plots

Norwegian fire insurance



Claim sizes for 1976

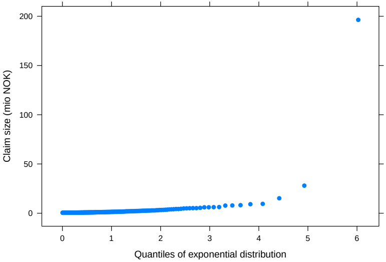
Norwegian fire insurance

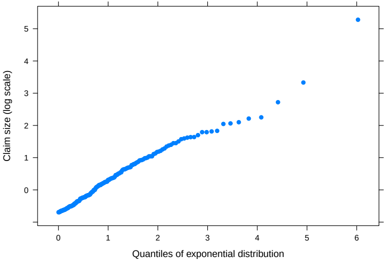


Min.	1st Qu.	Median	3rd Qu.	Mean	Max.
0.5	0.6	1.0	2.0	2.8	196.4

Claims (1976): exponential and Pareto QQ-plots

Norwegian fire insurance





Excess plots

Mean-excess function: motivation

Conditioning of a random variable X on the event $X > t$ is important in many applications.

Example

Excess-of-loss treaty with a retention t in reinsurance.
The reinsurer has to pay a random amount $X - t$*,*
but only if $X > t$*.*
To decide on the priority level t through simulation,
the **expected amount to be paid out per client**
for a given level t has to be calculated.
The net premium principle depends on
the mean claim size $E[X]$.
For the overshoot we consider: $E[X - t \mid X > t]$.

Mean-excess function

The mean-excess function (or mean residual life function) is:

$$e(t) = E[X - t \mid X > t]$$

Reference: the exponential model

The mean excess function for the exponential distribution is constant:

$$e(t) = \frac{1}{\lambda} \qquad \text{for all} \quad t > 0$$

Interpretation of excess plots
When the tail is

HTE: heavier than exponential, the excess function increases,

LTE: lighter than exponential, the excess function decreases.

The exponential model: a reference

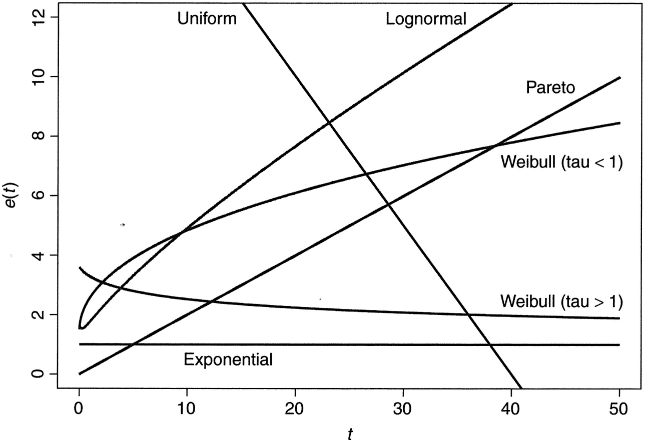


Figure 1.5 Shapes of some mean excess functions.

Empirical mean-excess function

The empirical mean-excess function is:

$$\widehat{e}_n(t) = \frac{\sum_{i=1}^{n} x_i \, \mathbf{1}_{x_i > t}}{\sum_{i=1}^{n} \mathbf{1}_{x_i > t}} - t$$

where the test function $\mathbf{1}_{x_i > t}$ equals 1 for $x_i > t$, and zero otherwise.

Excess plots

Often $\widehat{e}_n(t)$ is plotted at the values $t = x_{n-k}$ for $k = 1, \ldots, n-1$ which yields:

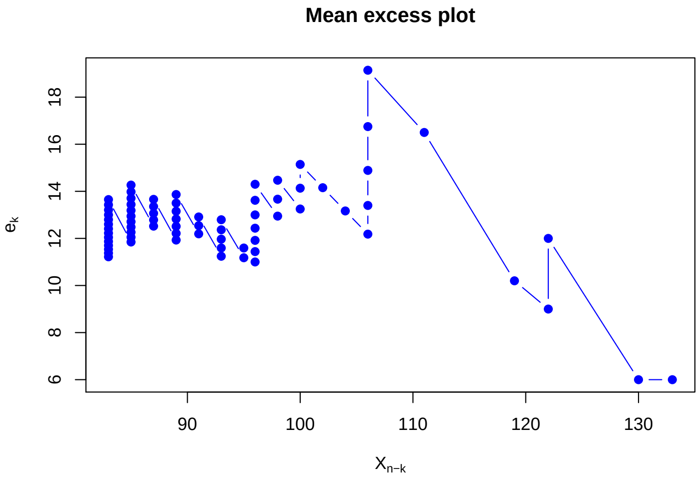
$$e_k = \frac{1}{k} \sum_{j=1}^{k} x_{n-j+1} - x_{n-k}$$

- Excess plot: e_k is usually plotted against x_{n-k}.
- (Some authors plot it against k)

Interpretation of excess plots

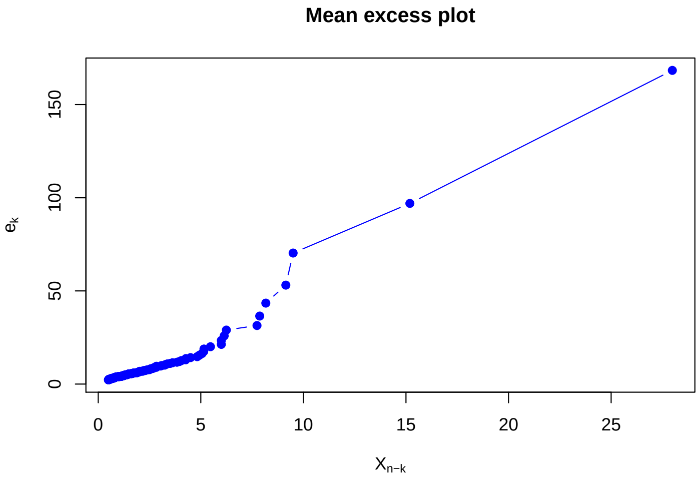
- Excess plot e_k vs x_{n-k}:
  examine the behavior for increasing values x_{n-k}.
- (When plotting e_k vs k:
  look at the behavior for decreasing k)

Zaventem: maximal wind speeds > 82km/h



Shape: constant → exponential

Norwegian fire insurance



Shape: linear increase $\rightarrow$ Pareto

Exponential QQ-plots and mean-excess plots

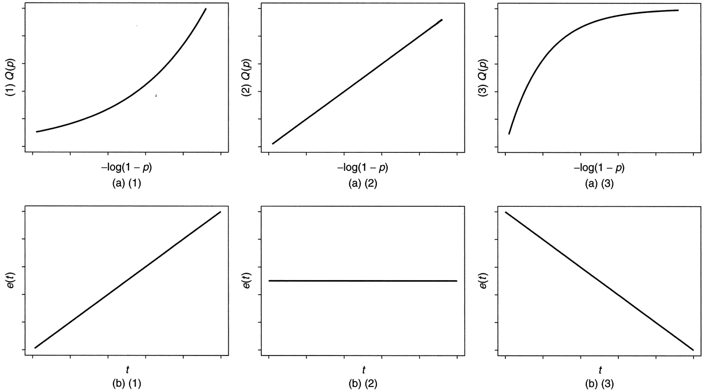
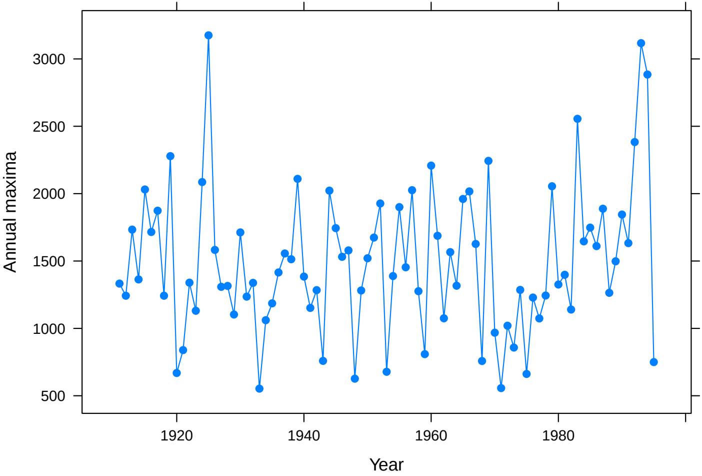


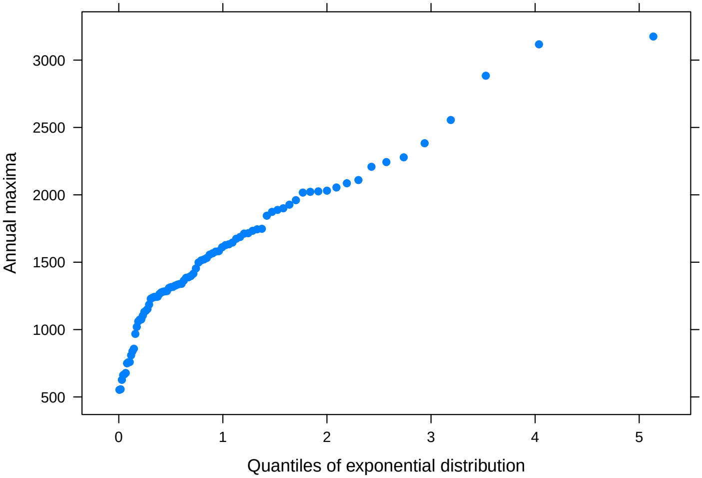
Figure 1.8 (a) Exponential QQ-plot and (b) mean excess plot for simple examples from (1) *HTE*-type, (2) exponential-type ($\bar{F}(x) \approx \exp(-x)$ for large) and (3) *LTE*-type distributions.

Meuse river: maximal annual discharges



Exponential QQ-plot

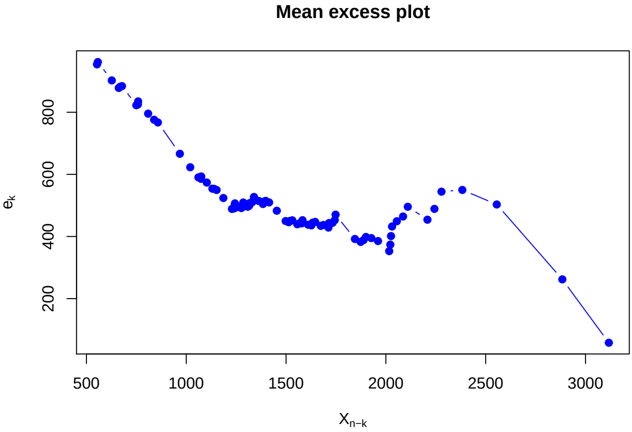
Meuse river: maximal annual discharges



Shape: linear increase → exponential

Excess plot

Meuse river: maximal annual discharges



Shape: constant $\rightarrow$ exponential

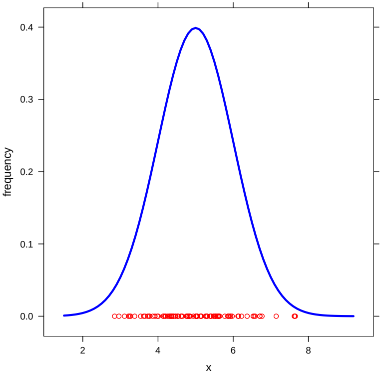
Statistical modeling of extreme values

(following the book by Stuart COLES and slides by Stuart COLES & Anthony DAVISON)

"Extreme value theory is unique as a statistical technique in that it develops techniques and models for describing the unusual rather than the usual."

Basics of extreme value theory

Suppose we have 100 independent observations following a distribution F:



We wish to estimate the right tail of F. We need to consider the following paradigm…

The extreme value paradigm

- There are only few observations in the right tail.
- Estimates are often required beyond the largest observed value.
- Standard density estimation techniques fit well where the data have the greatest density, but may be severely biased in estimating the tail probabilities.

The extreme value paradigm consists in modeling the tails using asymptotically motivated distributions.

It will *not* be necessary to know F, the distribution of the complete data set.

Applications

Historically there have been two main application areas of extreme value theory:

- Environmental:
  - sea-level,
  - wind speeds,
  - river flow…
- Reliability modeling:
  - weakest-link type models.

Nowadays there are a great variety:

- Financial modeling, insurance modeling, telecommunications,…

History

- 1920's: Foundations of asymptotic argument by Fisher and Tippet.
- 1940's: Unification and extension of asymptotic theory by Gnedenko and later by von Mises.
- 1950's: use of asymptotic distributions for statistical modeling by Gumbel and Jenkinson.
- 1970's: classic limit laws generalized by Pickands.
- 1980's: Leadbetter and others extend theory to stationary processes.
- 1990's: Multivariate and other techniques explored as a means to improve inference relying on covariates.
- **Currently:** geostatistical models for spatial extreme values.
- Main conference:
  Extreme Value Analysis 2021 (28 June - 02 July: online)

Probabilistic framework

Let $X_1, X_2, \ldots, X_n$ be a sequence of *independent* identically-distributed (iid) random variables with distribution function F. Let:

$$M_n = \max(X_1, X_2, \ldots, X_n)$$

Then the distribution function of the **maximum** M_n is:

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

But: F is unknown (and we do not wish to rely on it).

Instead, the distribution of M_n is rather approximated by limit distributions as $n \to \infty$.

Question: what limit distributions can arise ?

- With $F(x) < 1$, then $(F(x))^n \longrightarrow 0$ as $n \longrightarrow \infty$. Quite trivial, so far!

Probabilistic framework

Let $X_1, X_2, \dots, X_n$ be a sequence of *independent* identically-distributed (iid) random variables with distribution function F. Let:

$$M_n = \max(X_1, X_2, \dots, X_n)$$

Then the distribution function of the **maximum** M_n is:

$$\begin{aligned} P(M_n \leq x) &= P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) \\ &= P(X_1 \leq x)\, P(X_2 \leq x) \dots P(X_n \leq x) \\ &= (F(x))^n \end{aligned}$$

But: F is unknown (and we do not wish to rely on it).

Instead, the distribution of M_n is rather approximated by limit distributions as $n \rightarrow \infty$.

Question: what limit distributions can arise ?

- With $F(x) < 1$, then $(F(x))^n \longrightarrow 0$ as $n \longrightarrow \infty$. Quite trivial, so far!

The classical limit laws

- Now, recall the Central Limit Theorem, for iid random X_n:

$$P\left(\frac{\overline{X_n} - \mu_n}{\sigma_n} \leq x\right) \to \mathcal{N}(0,1) \qquad \text{as } n \to \infty$$

where $\mu_n = \mu$ and $\sigma_n = \sigma/\sqrt{n}$ act as normative coefficients.

- In the same way we look for limits of

$$\frac{(M_n - b_n)}{a_n}$$

for suitable sequences of a_n and b_n, so that the limit is a **non-degenerate** distribution.

The classical limit laws

- Now, recall the Central Limit Theorem, for iid random X_n:

$$P\left(\frac{\overline{X_n}-\mu_n}{\sigma_n}\leq x\right)\to\mathcal{N}(0,1)\qquad\text{as } n\to\infty$$

where $\mu_n=\mu$ and $\sigma_n=\sigma/\sqrt{n}$ act as normative coefficients.

- In the same way we look for limits of

$$\frac{(M_n-b_n)}{a_n}$$

for suitable sequences of a_n and b_n, so that
the limit is a **non-degenerate** distribution.

Equivalence class of distributions

Definition

The distributions F and F^* are of the *same type,* if there are constants a and b such that $F^*(ax+b) = F(x)$ for all x.

Example

$\mathcal{N}(\mu_1, \sigma_1)$ and $\mathcal{N}(\mu_2, \sigma_2)$ are of the same type.

Extremal Types Theorem

Theorem

If there exist sequences of constants $a_n > 0$ and b_n such that:

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \to G(x) \qquad \text{as } n \to \infty$$

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

1. **Gumbel**: $G(x) = \exp(-\exp(-x))$ for $-\infty < x < \infty$
2. **Fréchet**:
$$G(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0, \alpha > 0 \end{cases}$$
3. **Weibull**:
$$G(x) = \begin{cases} \exp(-(-x)^{\alpha}) & x < 0,\ \alpha > 0 \\ 1 & x \geq 0 \end{cases}$$

Example: exponential distribution

Suitable normative coefficients

Take the standard exponential distribution $F(x) = 1 - \exp(-x)$.
Then $F(x)^n = (1 - e^{-x})^n$.
With $a_n = 1$ and $b_n = \log n$,

$$P(M_n - \log n \le x) = P(M_n \le x + \log n) = F(x + \log n)^n$$

Then

$$\begin{aligned} F(x+\log n)^n &= (1 - e^{-x-\log n})^n \\ &= (1 - \frac{1}{n} e^{-x})^n \\ &\to \exp(-\exp(-x)) \qquad \text{as } n \to \infty \end{aligned}$$

so that the limiting distribution is the **Gumbel** distribution.

- We say that the exponential distribution is in the domain of attraction of the **Gumbel** distribution.

Example: exponential distribution

Suitable normative coefficients

Take the standard exponential distribution $F(x) = 1 - \exp(-x)$.
Then $F(x)^n = (1 - e^{-x})^n$.
With $a_n = 1$ and $b_n = \log n$,

$$P(M_n - \log n \le x) = P(M_n \le x + \log n) = F(x + \log n)^n$$

Then

$$\begin{aligned} F(x + \log n)^n &= (1 - e^{-x - \log n})^n \\ &= (1 - \frac{1}{n} e^{-x})^n \\ &\to \exp(-\exp(-x)) \qquad \text{as } n \to \infty \end{aligned}$$

so that the limiting distribution is the **Gumbel** distribution.

- We say that the exponential distribution is in the domain of attraction of the **Gumbel** distribution.

Example: exponential distribution

Suitable normative coefficients

Take the standard exponential distribution $F(x) = 1 - \exp(-x)$.
Then $F(x)^n = (1 - e^{-x})^n$.
With $a_n = 1$ and $b_n = \log n$,

$$P(M_n - \log n \leq x) = P(M_n \leq x + \log n) = F(x + \log n)^n$$

Then

$$\begin{aligned} F(x + \log n)^n &= (1 - e^{-x - \log n})^n \\ &= (1 - \frac{1}{n} e^{-x})^n \\ &\to \exp(-\exp(-x)) \qquad \text{as } n \to \infty \end{aligned}$$

so that the limiting distribution is the **Gumbel** distribution.

- We say that the exponential distribution is in the domain of attraction of the **Gumbel** distribution.

Example: normal distribution

Suitable normative coefficients

In the case of the standard normal distribution and with coefficicients

$$
\begin{aligned}
a_n &= (2\log n)^{-1/2} \\
b_n &= (2\log n)^{1/2} - \frac{(2\log n)^{-1/2}(\log\log n + \log 4\pi)}{2}
\end{aligned}
$$

the normalized maximum $(M_n - b_n)/a_n$ of n independent normal variables again converges to a **Gumbel** distribution.

Domains of attraction

A few examples

Gumbel domain: exponential, normal, lognormal, logistic, gamma distributions,...

Fréchet domain: Pareto, Cauchy, t, F distributions,...

Weibull domain: uniform, beta, Burr distributions,...

Using the limit law in practice

For a given n, regarded as large enough, the limit law may be used as an approximation:

$$P\left(\frac{M_n - b}{a} \leq x\right) \approx G(x) \qquad \text{for some } a > 0,\ b.$$

Equivalently,

$$\begin{aligned} P(M_n \leq x) &\approx G\left(\frac{x - b}{a}\right) \\ &= G^*(x) \end{aligned}$$

where G^* is of the same type as G.

- Thus the family of extreme value distributions may be fitted directly to a series of observations of M_n.

Generalized Extreme Value distribution

- The extremal types theorem distinguishes three families of distributions.
- The extreme value distribution families can be united in one expression:

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

defined on $\{x : 1+\xi(x-\mu)/\sigma > 0\}$.

We denote: $x_{+} = max(x, 0)$

μ is the location parameter,

σ is the scale parameter,

ξ is the shape parameter,
determining the rate of decay in the tail.

The Generalized Extreme Value (GEV) distribution is denoted $\mathscr{G}(\mu, \sigma, \xi)$.

Shape parameter

- The GEV distribution $\mathscr{G}(\mu,\sigma,\xi)$ splits up into the three families:

  $\xi > 0$ the **Fréchet** family
  characterized by a heavy upper tail,

  $\xi = 0$ the **Gumbel** family
  with an exponential upper tail,

  $\xi < 0$ the **Weibull** family
  whose tail has a finite upper limit.

- Actually the relation to the extremal types theorem is through $\xi = 1/\alpha$.
- The GEV distribution is actually not defined for $\xi = 0$.
  The subset of the GEV family with $\xi = 0$ is interpreted as the limit obtained for $\xi \to 0$, leading to the Gumbel family.

Return levels

The quantiles of the GEV distribution are computed as:

$$x_p = \begin{cases} \mu - \sigma \log[-\log(1-p)] & \text{for } \xi = 0, \\ \mu - \frac{\sigma}{\xi}\left(1 - [-\log(1-p)]^{-\xi}\right) & \text{for } \xi \neq 0. \end{cases}$$

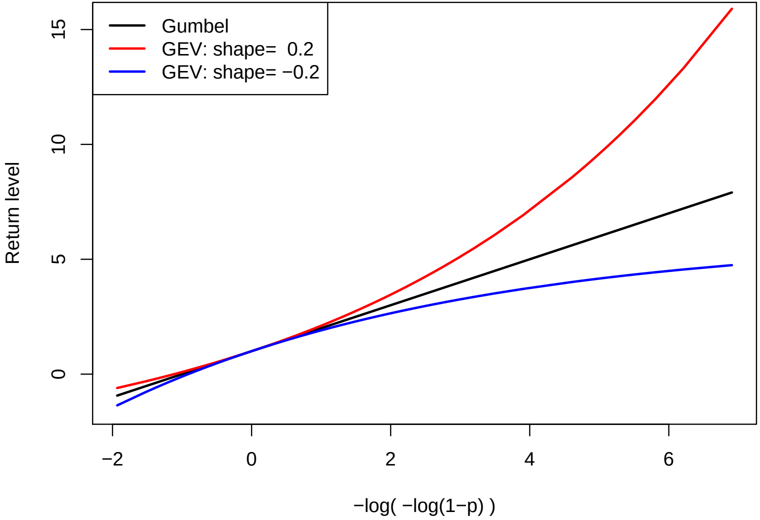
where $G(x_p) = 1 - p$.
In extreme value terminology,

- return period :
  mean waiting time between two extreme events,
- return level :
  the level x_p associated with a return period $T = 1/p$.

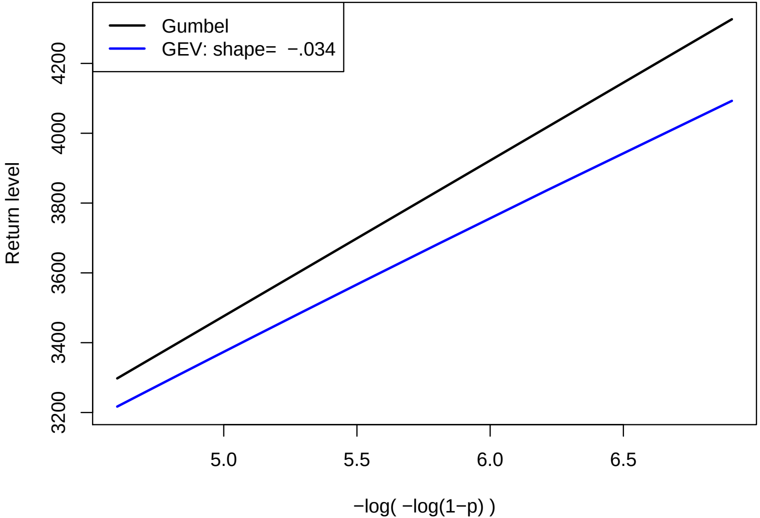
Example

A typical application is to fit the GEV to annual maxima.
We wish to compute the return level x_p that is expected to be exceeded on average once every T years.

GEV return level plot: different shapes ξ



Meuse discharge: from 100 to 1000-year return period



Max-stable distributions

Insight is gained using the concept of max-stability.

Definition

A distribution G is said to be max-stable, if for any k

$$G^k(a_k x + b_k) = G(x)$$

for suitable constants a_k and b_k.
Taking powers of G results only in a change of location and scale.

The connection with extremes is that :

- a distribution is max-stable, if and only if it is a GEV distribution.

Domains of Attraction

- Given a distribution function F, how can we determine suitable sequences a_n and b_n, and how can we know what limit G will occur?

For sufficiently smooth distributions, define

$$h(x) = \frac{1 - F(x)}{f(x)}$$

and let

$$b_n = F^{-1}\left(1 - \frac{1}{n}\right) \qquad a_n = h(b_n) \qquad \xi = \lim_{x \to \infty} h'(x)$$

then the limit distribution of $(M_n - b_n)/a_n$ is

$$\begin{array}{ll} \exp(-e^{-x}) & \text{if} \quad \xi = 0, \\ \exp\left(-[1 + \xi x]_+^{-1/\xi}\right) & \text{if} \quad \xi \neq 0. \end{array}$$

Domain of attraction: examples

Example

For the exponential distribution $F(x) = 1 - \exp(-x)$,
we have $h(x) = 1$.

Thus $\xi = \lim_{x\to\infty} h'(x) = 0$: this confirms the Gumbel domain.

Example

Suppose $1 - F(x) \sim c\,x^{-\alpha}$ as $x \to \infty$ for constants c, $\alpha > 0$.

This family includes the Pareto, Cauchy, t and F distributions.

Then

$$f(x) \sim \alpha c\,x^{-\alpha-1} \qquad h(x) \sim \alpha^{-1}x, \quad h'(x) \sim \alpha^{-1}$$

So $\xi = \alpha^{-1}$: so we are in the Fréchet domain of attraction.

Domain of attraction: examples

Example

For the exponential distribution $F(x) = 1 - \exp(-x)$,
we have $h(x) = 1$.

Thus $\xi = \lim_{x\to\infty} h'(x) = 0$: this confirms the Gumbel domain.

Example

Suppose $1 - F(x) \sim c\,x^{-\alpha}$ as $x \to \infty$ for constants c, $\alpha > 0$.

This family includes the Pareto, Cauchy, t and F distributions.

Then

$$f(x) \sim \alpha c\,x^{-\alpha-1} \qquad h(x) \sim \alpha^{-1}x, \quad h'(x) \sim \alpha^{-1}$$

So $\xi = \alpha^{-1}$: so we are in the Fréchet domain of attraction.

Block maxima approach

In classical EVA applications,
typically a sequence of annual maxima is analyzed:

- the sequence is partitioned into blocks (=years) and
- only the maximal value for each block (=year) is retained.

All other data are ignored.

The aim is to make inferences on the GEV parameters μ, σ, ξ.
Inference methods include:

- graphical techniques,
- moment-based estimators,
- **likelihood-based** techniques,
- Bayesian techniques.

Maximum likelihood

For theoretical and practical purposes, likelihood-based techniques are generally preferable. They are implemented in the R package `ismev`.

However, there is a potential difficulty that the endpoint of the distribution is a function of the parameters, so usual regularity conditions do not hold. Smith (1985) established that:

- When $\xi > -0.5$, maximum likelihood estimators are completely regular;
- When $-1 < \xi < -0.5$, maximum likelihood estimators exist, but are non-regular;
- When $\xi < -1$ maximum likelihood estimators do not exist.

In most environmental problems $\xi > -1$, so maximum likelihood works fine.

Modeling procedure

- Specification of log-likelihood function:

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

- Numerical maximization of log-likelihood.
- Calculation of standard errors from inverse of observed information matrix (also obtained numerically).
- Profile likelihood functions of parameters.
- Diagnostic checks: probability plots, quantile plots, return level plots.
- Calculation of confidence intervals for return levels.

Profile likelihood function

The log-likelihood for a parameter θ_i can be written $l(\theta_i, \theta_{-i})$, where θ_{-i} is the vector of all parameters excluding θ_i.
The profile likelihood is defined as

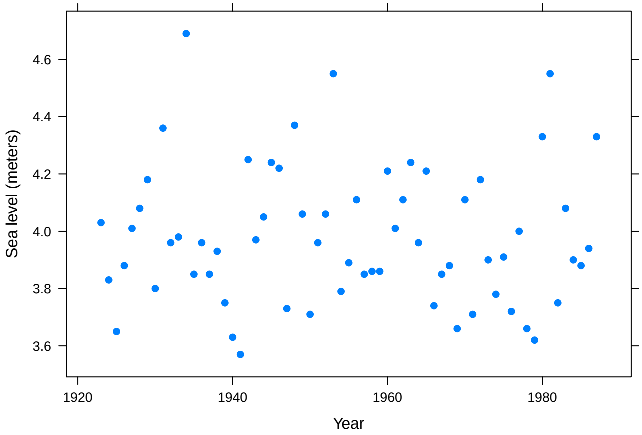
$$l_p(\theta_i) = \max_{\theta_{-i}} l(\theta_i, \theta_{-i})$$

For each value of θ_i the profile log-likelihood is the log-likelihood maximized for all other components of the vector θ.

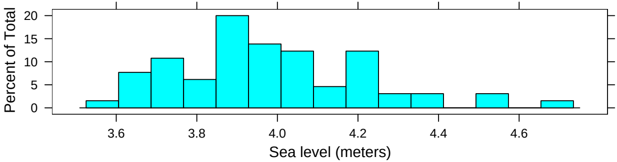
The profile likelihood plot shows the profile of the log-likelihood surface along the axis θ_i.

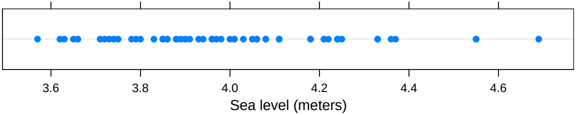
Port Pirie sea levels (1923 to 1987)

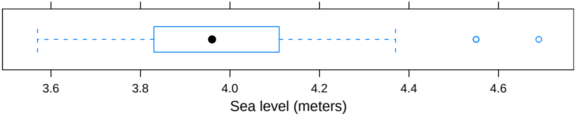
South Australia



Port Pirie sea levels








```
> summary(SeaLevel)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  3.570   3.830   3.960   3.981   4.110   4.690
```

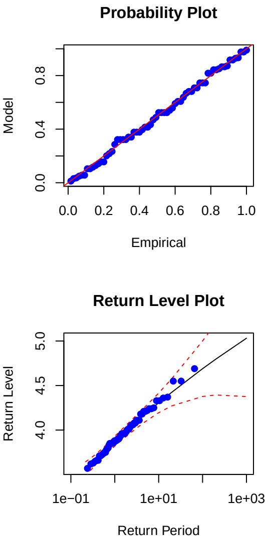
GEV fit

```
> data(portpirie)
> attach(portpirie)
> pirie.fit = gev.fit(SeaLevel)
# The negative logarithm of the likelihood
# evaluated at the maximum likelihood estimates.
$nllh [1] -4.339058
#          mu          sigma        xi
$mle [1]  3.87474692  0.19804120 -0.05008773
# standard errors
$se [1] 0.02793211 0.02024610 0.09825633
# diagnostic plots
> gev.diag(pirie.fit)
# profile likelihood plot for xi
> gev.profxi(pirie.fit,-0.242,0.242)
# profile likelihood plot for return level
> gev.prof(pirie.fit,100,4.4,6.0)
```

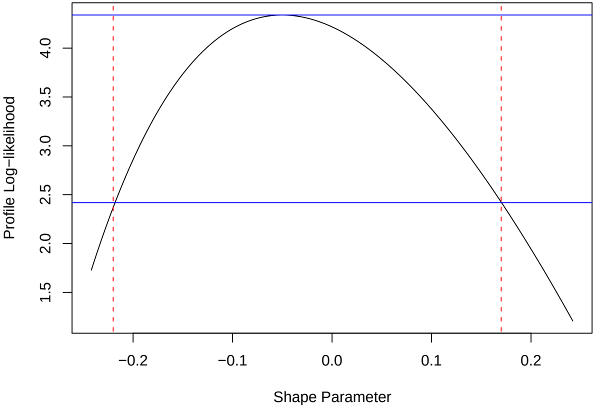
GEV fit

```
> data(portpirie)
> attach(portpirie)
> pirie.fit = gev.fit(SeaLevel)
# The negative logarithm of the likelihood
# evaluated at the maximum likelihood estimates.
$nllh [1] -4.339058
#          mu         sigma       xi
$mle [1]  3.87474692  0.19804120 -0.05008773
# standard errors
$se [1] 0.02793211 0.02024610 0.09825633
# diagnostic plots
> gev.diag(pirie.fit)
# profile likelihood plot for xi
> gev.profxi(pirie.fit,-0.242,0.242)
# profile likelihood plot for return level
> gev.prof(pirie.fit,100,4.4,6.0)
```

```
gev.diag(pirie.fit)
```

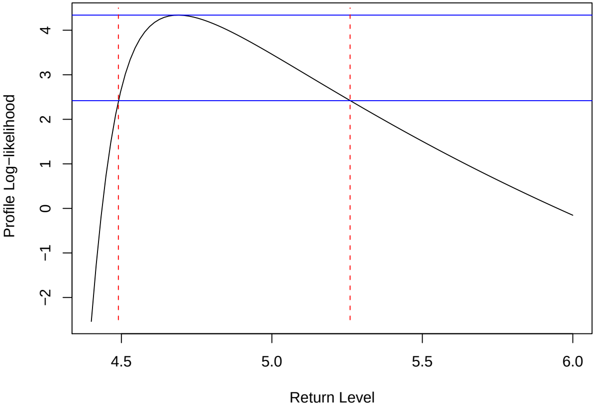


Profile likelihood for ξ



The 95% confidence interval for $\xi = -.05$ is: [-.22, .17].
Could be Gumbel!

Profile likelihood: 100-year return level



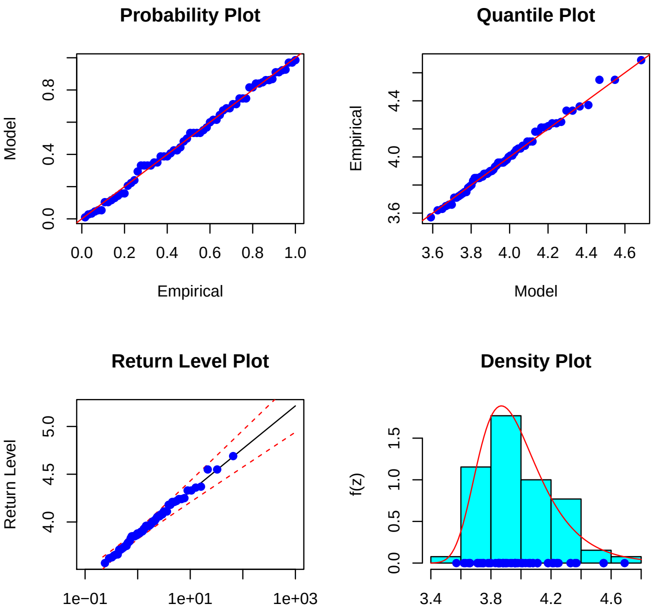
The 95% confidence interval for 100-year level of 4.69 is: [4.5, 5.27]

Gumbel fit: less parameters!

```
> pirie.Gumfit = gum.fit(SeaLevel)
$nllh [1] -4.217682
#        mu        sigma
$mle [1] 3.8694426 0.1948867
#       standard errors
$se [1] 0.02549356 0.01885190
```

A likelihood ratio test statistic for the reduction to Gumbel model suggests the Gumbel is more adequate.

gum.diag(pirie.Gumfit)



- Modeling only block maxima is a wasteful approach,
  if other data besides the maxima are available within the blocks.
- If an entire time series of, say, hourly or daily observations is available, then better use is made of the data by avoiding altogether the procedure of blocking.

Threshold method: model

Let $X_1, X_2, \ldots$ be a sequence of iid RV with marginal distribution F.
We now regard as extreme events those of the X_i that exceed some high threshold u.

A description of the stochastic behavior of extreme events is provided by the conditional probability:

$$P(X > u + y \mid X > u) = \frac{1 - F(u+y)}{1 - F(u)} \qquad \text{with} \quad y > 0$$

In applications we do not wish to specify F.
We can rely on the fact that:

- the distribution of threshold exceedances belongs to the **generalized Pareto family**.

Generalized Pareto distribution

Let $X_1, X_2, \ldots$ be a sequence of iid RV with marginal distribution F and let

$$M_n = \max(X_1, X_2, \ldots, X_n)$$

Supposing for large n that the M_n are approximately GEV(μ, σ, ξ) distributed, then for large enough u
the distribution of $X - u$ conditional on $X > u$
is approximately:

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi}$$

with $\{y : y > 0 \, \text{and} \, (1 + \xi y/\tilde{\sigma}) > 0\}$, where $\tilde{\sigma} = \sigma + \xi(u - \mu)$.

The distributions $H(y)$ form the generalized Pareto family.

Generalized Pareto distributions

Example. For the exponential distribution $F(x) = 1 - e^{-x}$ for $x > 0$,

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{e^{-(u+y)}}{e^{-u}} = e^{-y} \qquad \text{for} \quad y > 0$$

is GPD with $\xi = 0$ and $\tilde{\sigma} = 1$.

Example. For a Fréchet distribution $F(x) = \exp(-1/x)$ for $x > 0$,

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{1 - \exp[-(u+y)^{-1}]}{1 - \exp(-u^{-1})} \sim \left(1 + \frac{y}{u}\right)^{-1} \quad \text{as } u \to \infty, \text{ for } y > 0$$

is GPD with $\xi = 1$ and $\tilde{\sigma} = u$.

Example. For the uniform distribution $F(x) = x$ for $0 \leq x \leq 1$,

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{1 - (u+y)}{1 - u} = 1 - \frac{y}{1-u} \qquad \text{for} \quad 0 \leq y \leq 1 - u$$

is GPD with $\xi = -1$ and $\tilde{\sigma} = 1 - u$.

Generalized Pareto distributions

Example. For the exponential distribution $F(x) = 1 - e^{-x}$ for $x > 0$,

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{e^{-(u+y)}}{e^{-u}} = e^{-y} \qquad \text{for} \quad y > 0$$

is GPD with $\xi = 0$ and $\tilde{\sigma} = 1$.

Example. For a Fréchet distribution $F(x) = \exp(-1/x)$ for $x > 0$,

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{1 - \exp[-(u+y)^{-1}]}{1 - \exp(-u^{-1})} \sim \left(1 + \frac{y}{u}\right)^{-1} \quad \text{as } u \to \infty, \text{ for } y > 0$$

is GPD with $\xi = 1$ and $\tilde{\sigma} = u$.

Example. For the uniform distribution $F(x) = x$ for $0 \leq x \leq 1$,

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{1 - (u+y)}{1 - u} = 1 - \frac{y}{1-u} \qquad \text{for} \quad 0 \leq y \leq 1 - u$$

is GPD with $\xi = -1$ and $\tilde{\sigma} = 1 - u$.

Generalized Pareto distributions

Example. For the exponential distribution $F(x) = 1 - e^{-x}$ for $x > 0$,

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{e^{-(u+y)}}{e^{-u}} = e^{-y} \qquad \text{for} \quad y > 0$$

is GPD with $\xi = 0$ and $\tilde{\sigma} = 1$.

Example. For a Fréchet distribution $F(x) = \exp(-1/x)$ for $x > 0$,

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{1 - \exp[-(u+y)^{-1}]}{1 - \exp(-u^{-1})} \sim \left(1 + \frac{y}{u}\right)^{-1} \quad \text{as } u \to \infty, \text{ for } y > 0$$

is GPD with $\xi = 1$ and $\tilde{\sigma} = u$.

Example. For the uniform distribution $F(x) = x$ for $0 \le x \le 1$,

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{1 - (u+y)}{1 - u} = 1 - \frac{y}{1-u} \qquad \text{for} \quad 0 \le y \le 1 - u$$

is GPD with $\xi = -1$ and $\tilde{\sigma} = 1 - u$.

Threshold choice

The issue of threshold choice is analogous to that of block size in the block maxima approach: a balance between bias and variance

- Too low a threshold may violate the asymptotic basis of the model, leading to bias.
- Too high, it will generate few excesses for estimating the model, leading to high variance.

Discussion. *In some applications the threshold is prescribed by a convention.*

Example

In the last IPCC report (2001) the following threshold for extremes was defined: 90th (or 10th) percentile for hot (cold) extremes.

Threshold choice

The issue of threshold choice is analogous to that of block size in the block maxima approach: a balance between bias and variance

- Too low a threshold may violate the asymptotic basis of the model, leading to bias.
- Too high, it will generate few excesses for estimating the model, leading to high variance.

Discussion. *In some applications the threshold is prescribed by a convention.*

Example

In the last IPCC report (2001) the following threshold for extremes was defined: 90th (or 10th) percentile for hot (cold) extremes.

Two methods for threshold choice

1. Exploratory technique: the mean excess plot.
2. Estimation of the model at a range of thresholds.

Once a threshold has been selected the parameters of the GPD are estimated by maximum likelihood.
The algorithm should avoid numerical instabilities for $\xi \approx 0$ and the evaluation has to take place within allowable parameter space.

Threshold choice: 1st method

Mean excess plot

Return levels for Pareto models

Suppose that the exceedances of X over a threshold u are GPD(σ, ξ).

$$P(X > x \mid X > u) = \left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi}$$

As $P(A) = P(B)\,P(A \mid B)$, defining $\zeta_u = P(X > u)$,

$$P(X > x) = \zeta_u\left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi}$$

A level x_m exceeded on average once every m observations is solution of:

$$\zeta_u\left[1 + \xi\left(\frac{x_m-u}{\sigma}\right)\right]^{-1/\xi} = \frac{1}{m}$$

so that:

$$\begin{cases} x_m = u + \frac{\sigma}{\xi}\left[(m\,\zeta_u)^{\xi} - 1\right] & \text{for} \quad \xi \neq 0 \\ x_m = u + \sigma\log(m\,\zeta_u) & \text{for} \quad \xi = 0 \end{cases}$$

provided m is sufficiently large to ensure that $x_m > u$.

Return levels for Pareto models

Suppose that the exceedances of X over a threshold u are GPD(σ, ξ).

$$P(X > x \mid X > u) = \left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi}$$

As $P(A) = P(B)\,P(A \mid B)$, defining $\zeta_u = P(X > u)$,

$$P(X > x) = \zeta_u \left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi}$$

A level x_m exceeded on average once every m observations is solution of:

$$\zeta_u \left[1 + \xi\left(\frac{x_m-u}{\sigma}\right)\right]^{-1/\xi} = \frac{1}{m}$$

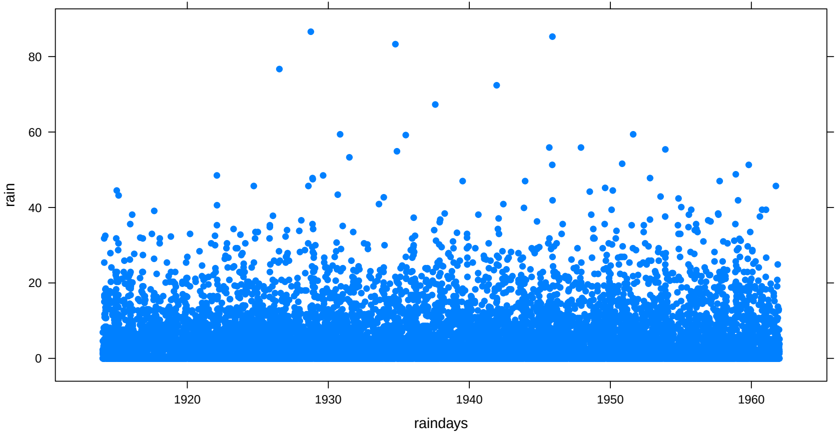
so that:

$$\begin{cases} x_m = u + \frac{\sigma}{\xi}\left[(m\,\zeta_u)^{\xi} - 1\right] & \text{for} \quad \xi \neq 0 \\ x_m = u + \sigma \log(m\,\zeta_u) & \text{for} \quad \xi = 0 \end{cases}$$

provided m is sufficiently large to ensure that $x_m > u$.

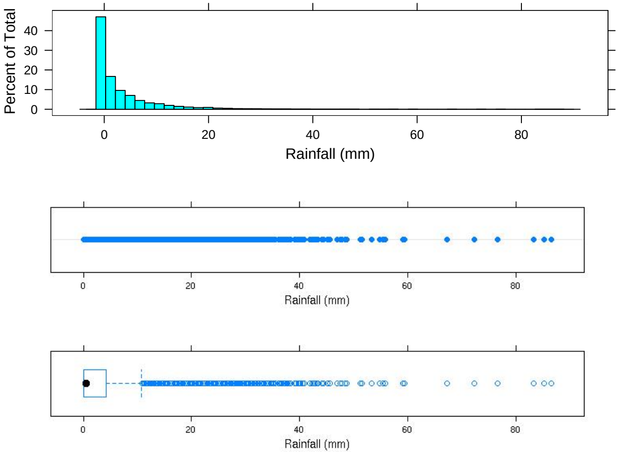
Daily rainfall in mm (1914-1961)

SW England




```
raindays=seq(as.Date("1914/1/1"),as.Date("1961/12/30"), "days")
xyplot(rain ~ raindays)
```

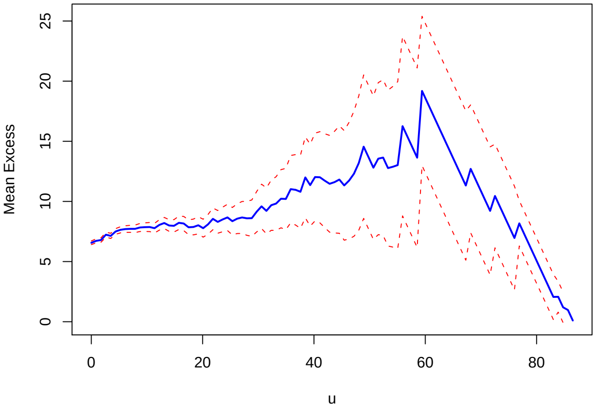
Rainfall in SW England (17531 days)




```
> summary(rain)
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.000   0.000   0.500   3.476   4.300  86.600
```

Rainfall data: mean excess plot

`mrl.plot(rain)`



Interpretation is not easy, but discarding the few points at the end, $u = 30$ seems a reasonable choice.

Threshold choice: 2nd method

Testing a range of thresholds

Range of thresholds

Idea: fit the GPD at a range of thresholds and examine parameter stability.

Modified scale parameter

Remember the scale parameter $\tilde{\sigma} = \sigma + \xi\,(u - \mu)$ of the GPD for a GEV(μ, σ, ξ) .
Denoting by σ_u the value for a threshold of $u > u_0$, it follows:

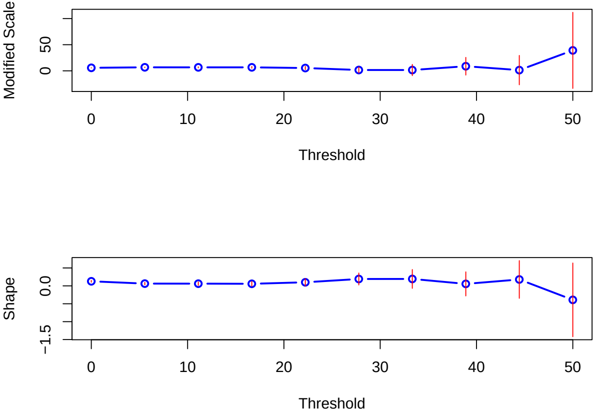
$$\sigma_u = \sigma_{u_0} + \xi\,(u - u_0)$$

where the scale parameter changes with u (unless $\xi = 0$).
Therefore the GPD scale parameter is modified to:

$$\sigma^* = \sigma_u - \xi\, u$$

Rainfall data: range of thresholds

`gpd.fitrange(rain,0,50)`



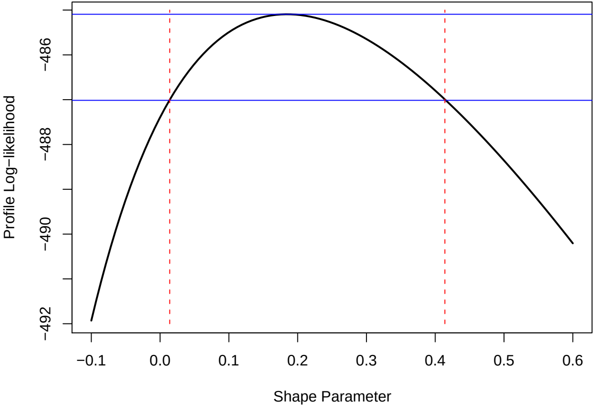
Conclusion: $u = 30$ seems a reasonable choice.

GPD fit

```
> rain.fit = gpd.fit(rain,30)
# number of data above the threshold
$nexc [1] 152
# zero means successful convergence of algorithm
$conv [1] 0
# negative logarithm of likelihood at maximum
$nllh [1] 485.0937
#        sigma     xi
$mle [1] 7.4406505 0.1843329
#         proportion of data above threshold
$rate [1] 0.008670355
#  standard errors
$se [1] 0.958432 0.101151
```

Profile likelihood for ξ

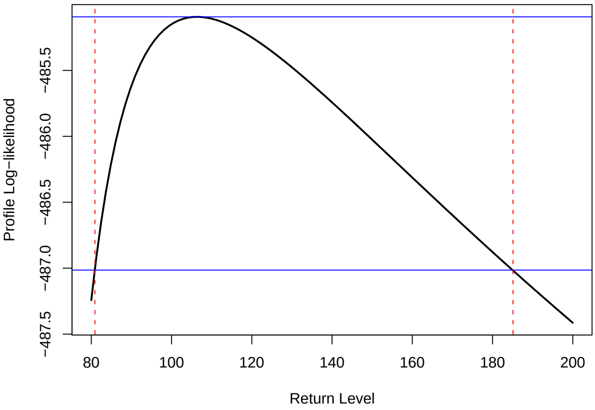
`gpd.profxi(rain.fit, -0.1, 0.6)`



The 95% confidence interval for $\xi = 0.184$ is: $[0.014, 0.414]$.

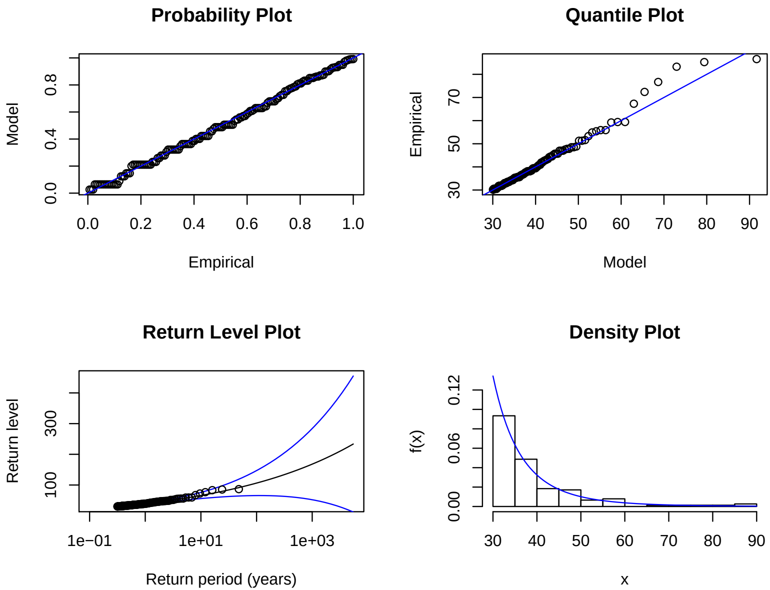
Profile likelihood: 100-year return level

`gpd.prof(rain.fit,m=100,80,200)`



The 95% confidence interval for 100-year level of 106.3 is: [80.9, 185.1]

```
gpd.diag(rain.fit)
```



Dependence – Non-stationarity

(following the book and the slides of Stuart COLES)

Time series of Geophysical Data

Extreme value data usually demonstrate:

- Short term dependence (storms for example);
- Dependence on covariate effects;
- Seasonality (due to annual cycles in meteorology);
- Long-term trends (due to gradual climatic change);
- Other forms of non-stationarity (for example, the deterministic effect of tides on sea-levels).

For temporal dependence there is a sufficiently wide-ranging theory which can be invoked.

Other aspects have to be handled at a later stage.

Temporal Dependence

It is useful to distinguish between:

1. Long-range dependence,
2. Short-term dependence.

The second is more problematic in environmental applications as extreme events may tend to cluster (storms, for example).

- The usual approach is to specify a condition which restricts the impact of long-range dependence on extremes – hoping physical processes behave this way – and to study the effect of short-term dependence.
- This is equivalent to assume that two events $X_i > u$ and $X_j > u$ are approximately independent provided the threshold u is high enough and time points i and j have a large separation.

The $D(u_n)$ condition

The first step is to formulate a condition that makes precise the notion of extreme events being near-independent if they are sufficiently separated in time.

Definition

A stationary series $X_1, X_2, \ldots$ is said to satisfy the $D(u_n)$ condition if, for all $i_1 < \cdots < i_p < j_1 < \cdots < j_q$ with $j_1 - i_p > l$,

$$
\begin{aligned}
&|P(X_{i_1} \le u_n, \ldots, X_{i_p} \le u_n, X_{j_1} \le u_n, \ldots, X_{j_q} \le u_n) \\
&\quad - P(X_{i_1} \le u_n, \ldots, X_{i_p} \le u_n)\ P(X_{j_1} \le u_n, \ldots, X_{j_q} \le u_n)| \quad \le \quad \alpha(n, l)
\end{aligned}
$$

where $\alpha(n, l_n) \to 0$ for some sequences l_n such that $l_n / n \to 0$ as $n \to \infty$.

Stationary series with short dependence

Theorem

Let $X_1.X_2, \ldots$ be a stationary process and define $M_n = \max(X_1, X_2, \ldots, X_n)$. If $a_n > 0$ and b_n are sequences of constants such that:

$$P\left(\frac{M_n - b_n}{a_n} \leq z\right) \to G(z)$$

where G is a non-degenerate distribution function, and if the $D(u_n)$ condition is satisfied with $u_n = a_n z + b_n$ for every real z, G is a member of the generalized extreme value distributions.

Short-term dependence

Simulated example of a series with dependent values

Let $Y_1, Y_2, \ldots$ be iid RV with a standard exponential distribution.
A dependent sequence is obtained as:

$$X_i = \max(Y_i, Y_{i+1})$$

```
# simulate 101 values
# from an exponential distribution
y = rexp(101)
# compute the "parallel" maximum (R jargon)
x = pmax(y[1:100], y[2:101])
xyplot(x ~ 1:100)
```

Short-term dependence

Simulated example of a series with dependent values

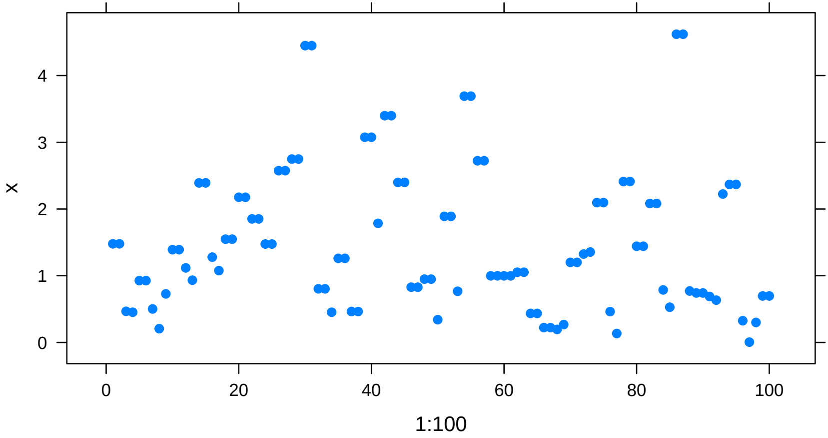
Let $Y_1, Y_2, \ldots$ be iid RV with a standard exponential distribution.
A dependent sequence is obtained as:

$$X_i = \max(Y_i, Y_{i+1})$$

```
# simulate 101 values
# from an exponential distribution
y = rexp(101)
# compute the "parallel" maximum (R jargon)
x = pmax(y[1:100], y[2:101])
xyplot(x ~ 1:100)
```

Dependent series

Simulated example



The maxima tend to cluster in pairs, i.e. the average cluster size is 2.

Dependent series: 2-point cluster

Comparison with independent case

Let $Y_1, Y_2, \dots$ be iid RV with standard exponential distribution.
$X_i = \max(Y_i, Y_{i+1})$.

The marginal distribution of the dependent random variable X is:

$$P(X_i < x) = P(Y_i < x, Y_{i+1} < x) = (1 - e^{-x})^2$$

Dependent series: 2-point cluster

Comparison with independent case

Let $X_1^*, X_2^*, \ldots$ be a series of independent random variables with the same marginal distribution and $M_n^* = \max(X_1^*, \ldots, X_n^*)$. With $a_n = 1, b_n = \log(2n)$:

$$
\begin{aligned}
P(M_n^* - \log(2n) < x) &= (1 - \exp[-x - \log(2n)])^{2n} \\
&= (1 - \frac{1}{2n} e^{-x})^{2n} \quad \to \quad \exp\left(-e^{-x}\right) = G_1(x)
\end{aligned}
$$

In the dependent case, for $M_n = \max(X_1, \ldots, X_n)$:

$$
\begin{aligned}
P(M_n - \log(2n) < x) &= P(Y_1 < x + \log(2n), \ldots, Y_{n+1} < x + \log(2n)) \\
&= (1 - \frac{1}{2n} e^{-x})^{n+1} \quad \to \quad \exp\left(-\frac{1}{2} e^{-x}\right) = G_2(x)
\end{aligned}
$$

So:

$$G_2(x) = (G_1(x))^{1/2}$$

- **Note:** the power is the inverse of the cluster size (which is 2).

Dependent series: 2-point cluster

Comparison with independent case

Let $X_1^*, X_2^*, \ldots$ be a series of independent random variables with the same marginal distribution and $M_n^* = \max(X_1^*, \ldots, X_n^*)$. With $a_n = 1, b_n = \log(2n)$:

$$
\begin{aligned}
P(M_n^* - \log(2n) < x) &= (1 - \exp[-x - \log(2n)])^{2n} \\
&= (1 - \frac{1}{2n} e^{-x})^{2n} \quad \to \quad \exp\left(-e^{-x}\right) = G_1(x)
\end{aligned}
$$

In the dependent case, for $M_n = \max(X_1, \ldots, X_n)$:

$$
\begin{aligned}
P(M_n - \log(2n) < x) &= P(Y_1 < x + \log(2n), \ldots, Y_{n+1} < x + \log(2n)) \\
&= (1 - \frac{1}{2n} e^{-x})^{n+1} \quad \to \quad \exp\left(-\frac{1}{2} e^{-x}\right) = G_2(x)
\end{aligned}
$$

So:

$$
G_2(x) = (G_1(x))^{1/2}
$$

- **Note:** the power is the inverse of the cluster size (which is 2).

Extremal Index

Theorem

Let $X_1, X_2, \dots$ be a stationary process and $X_1^*, X_2^*, \dots$ be a sequence of independent variables with the same marginal distribution.
Define $M_n = \max(X_1, X_2, \dots)$ and $M_n^* = \max(X_1^*, \dots, X_n^*)$. Under suitable regularity conditions

$$P\left(\frac{M_n^* - b_n}{a_n} \leq z\right) \to G_1(z)$$

as $n \to \infty$ for normalizing sequences $\{a_n > 0\}$ and $\{b_n\}$, where G_1 is a non-degenerate distribution function, if and only if

$$P\left(\frac{M_n - b_n}{a_n} \leq z\right) \to G_2(z)$$

where

$$G_2(z) = (G_1(z))^\theta$$

for a constant θ, called the extremal index, such that $0 < \theta \leq 1$.

Extremal index: parameters of G_2

$$
\begin{aligned}
G_1^{\theta}(z) &= \left[\exp\left(-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right)\right]^{\theta} \\
&= \exp\left(-\theta\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right) \\
&= \exp\left(-\left[1+\xi\left(\frac{x-\mu^*}{\sigma^*}\right)\right]^{-1/\xi^*}\right)
\end{aligned}
$$

where the parameters of the independent case are obtained from those of the stationary process by:

$$
\mu^* = \mu - \frac{\sigma}{\xi}(1-\theta^{-\xi}) \qquad \sigma^* = \sigma\,\theta^{\xi} \qquad \xi^* = \xi
$$

Extremal index: Gumbel case

The parameters of the independent case are related to those of the stationary process by:

$$\mu^* = \mu + \sigma \log \theta \qquad \sigma^* = \sigma$$

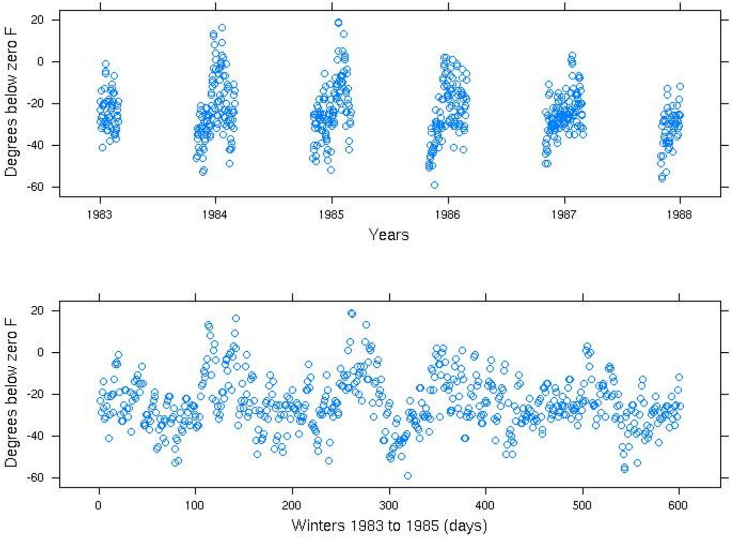
when G_1 is a Gumbel distribution.
Then G_2 is also Gumbel and differs only in the location parameter.

Modeling Techniques

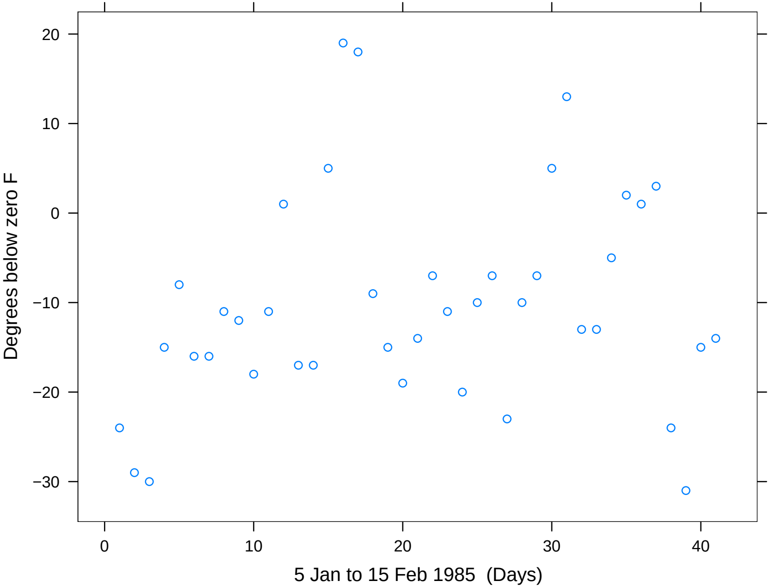
1. Identify clusters and model cluster maxima only.
2. As in 1, but also estimate the extremal index empirically.
3. Ignore dependence since marginal model is valid, but inflate standard errors to account for reduction in independent information.
4. Specify explicit model for dependence, such as a first-order Markov chain.

Wooster minimum winter temperatures

Negated daily minimum temperature in degrees Fahrenheit

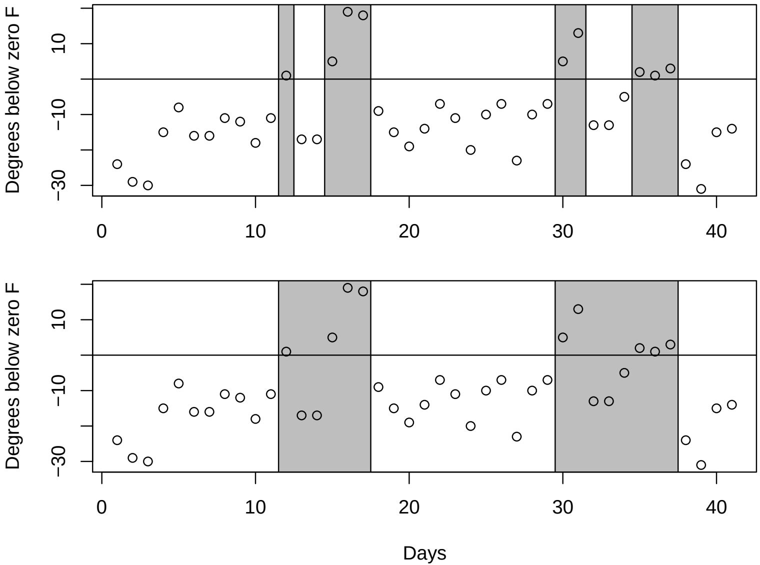


Wooster temperatures: zoom on 40 days



Wooster temperatures: clusters of size 2 and 4

`clusters(wooWin85,u=0,2,plot=T);  clusters(wooWin85,u=0,4,plot=T)`



Non-stationarity

In applications we often need to include model trends, seasonality and covariate effects by parametric models.

Examples

1. Trend in the location parameter:

$$\mu(t) = \alpha + \beta\, t$$

2. Trend in the scale parameter:

$$\sigma(t) = \alpha + \beta\, t$$

3. Non-stationary shape parameter:

$$\xi = \begin{cases} \xi_1 & \text{for} \quad t \leq t_0 \\ \xi_2 & \text{for} \quad t > t_0 \end{cases}$$

4. Location parameter depends on covariate:

$$\mu(t) = \alpha + \beta\, Y(t)$$

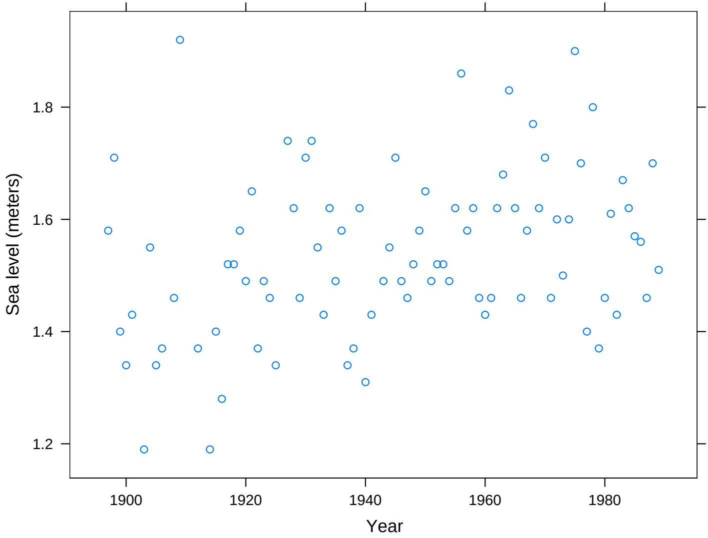
Modeling in R

Functions such as `gev.fit` enable time trends and covariates to be fitted by specifying a design matrix:

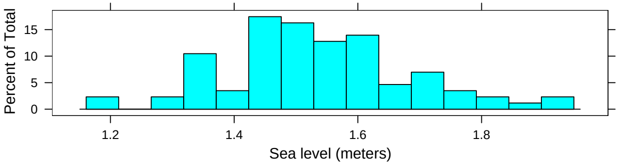
- Construct in `ydat` a design matrix corresponding to the required forms of time or covariate dependence,
- Specify how each of the parameters μ, σ, ξ is to be modelled as a linear function of the columns of the design matrix. e.g. `mul=c(1,3)`.
- Specify the link function for each of the parameters μ, σ, ξ, e.g. `mulink=identity`.

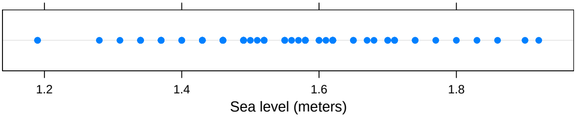
Fremantle sea levels (1897 to 1989)

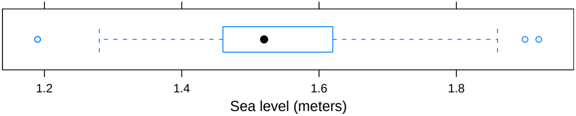
Western Australia



Fremantle sea levels







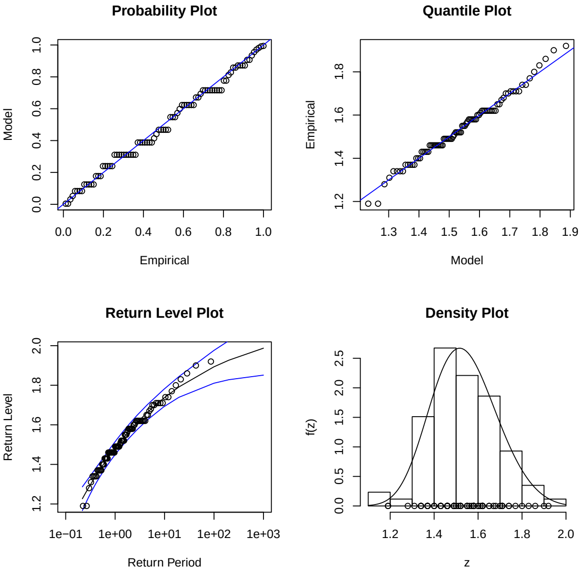

```
> summary(SeaLevel)
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.190   1.460   1.520   1.538   1.620   1.920
```

GEV fit and diagnostic plots

```
data(fremantle)
dim(fremantle)
names(fremantle)
# access directly to variable names
attach(fremantle)
# gev fit and diagnostics
fm.gev = gev.fit(SeaLevel)
gev.diag(fm.gev)
gev.profxi(fm.gev,-0.4,0.0)
gev.prof(fm.gev,100,1.83,2.05)
```

GEV fit: diagnostic plots

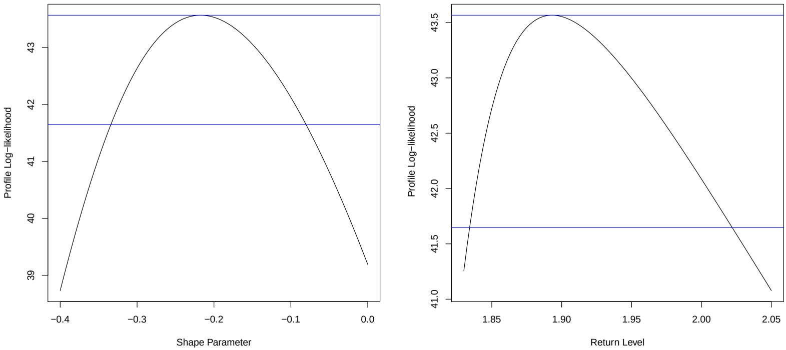
gev.diag(fm.gev)



GEV fit: profiles

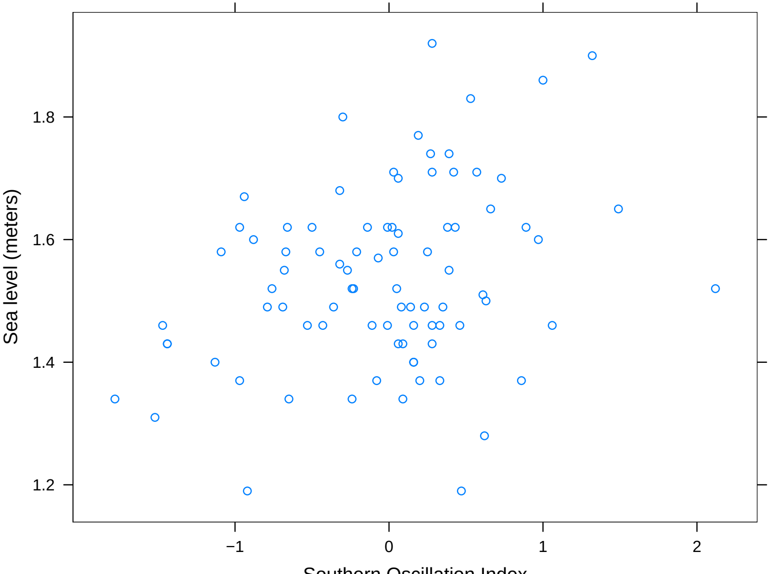
gev.profxi(fm.gev,-0.4,0.0)

gev.prof(fm.gev,100,1.83,2.05)



Introducing a covariate: SOI

`xyplot(SeaLevel ~ SOI)`



Design matrix

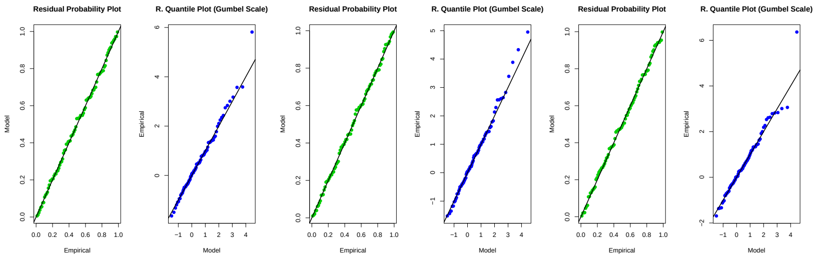
```
## matrix of covariates
## Year is linearly rescaled for numerical reasons
fm.covar = cbind( (Year-1943)/46, SOI )
colnames(fm.covar) = c( "YEAR","SOI")
head(fm.covar)
           YEAR   SOI
[1,] -1.0000000 -0.67
[2,] -0.9782609  0.57
[3,] -0.9565217  0.16
[4,] -0.9347826 -0.65
[5,] -0.9130435  0.06
[6,] -0.8695652  0.47
```

GEV fit with 3 different models

```
# Model M1
# mu depends linearly on YEAR...
fm.gevM1 = gev.fit(SeaLevel, ydat=fm.covar, mul=1 )
# Model M2
# ... and with sigma depending exponentially of year
fm.gevM2 = gev.fit(SeaLevel, ydat=fm.covar, mul=1,
                                   sigl=1,siglink=exp )
# Model M3
# linear dependence of mu on YEAR and SOI
fm.gevM3 = gev.fit(SeaLevel, ydat=fm.covar,
                                            mul=c(1,2) )
gev.diag(fm.gevM1)
gev.diag(fm.gevM2)
gev.diag(fm.gevM3)
```

GEV fit: 3 models

gev.diag(fm.gevM1) gev.diag(fm.gevM2) gev.diag(fm.gevM3)



Clearly M2 fits best.

Threshold models

The same techniques can also be applied to GPD models (`gpd.fit`).

However, there is an extra complication: if, for example, there is a time trend or seasonality, is it appropriate to use a constant threshold?
Hence `gpd.fit` also allows user-specified time-dependent thresholds to be incorporated.

The Point Process approach

to extreme value analysis

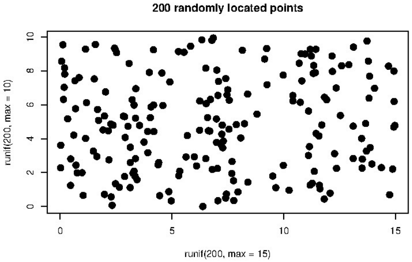
Point process

How can a set of points randomly scattered in space be characterized?

Example: uniform points in a rectangle

`plot(runif(200,max=10),runif(200,max=15))`

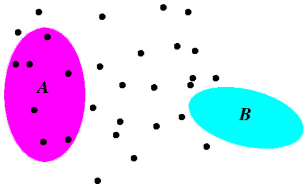
To obtain n points $\mathbf{u} = (u_1, u_2)$ with random locations in a rectangle $a \times b$ we simulate independently n values of $u_1 \sim \mathscr{U}[0, a]$ and $u_2 \sim \mathscr{U}[0, b]$.



- How can such a set of points be characterized ?

Spatial characterization of a point process

Analysing finite subsets of points *per se* is not a fruitful way to go.
The spatial point process should rather be examined using objects (subsets of $\mathbb{R}^d$) called *Borelians* :



Counting approach: count the points covered by an object

$$n(A) = 7, \quad n(B) = 0$$

Avoiding functional: test for whether object fills a void space

$$q(A) = 0, \quad q(B) = 1.$$

Spatial distribution of the counting measure

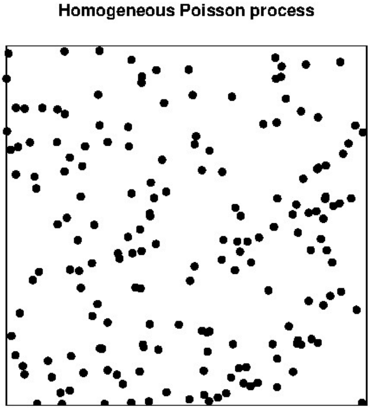
To characterize the point process we can use the spatial distribution of the counting measure $N(A)$:

$$
\begin{aligned}
&P(\quad N(A_1) = n_1 \quad)\\
&P(\quad N(A_1) = n_1, \quad N(A_2) = n_2 \ )\\
&\qquad\qquad\vdots\\
&P(\quad N(A_1) = n_1, \quad \ldots\ , N(A_p) = n_p \ )
\end{aligned}
$$

for any $p \in \mathbb{N}$, $n_1 \ldots n_p \in \mathbb{N}$, $A_1 \ldots A_p \in \mathscr{B}(\mathbb{R}^d)$.

$\mathscr{B}(\mathbb{R}^d)$ is a family of Borel sets, i.e.
the smallest σ-algebra spanned by the open subsets of $\mathbb{R}^d$.

Homogeneous Poisson process



Modeling the counts: Poisson distribution

A random variable N is Poisson distributed of mean λ if its probability mass function is

$$P(N = n) = \mathrm{e}^{-\lambda} \frac{\lambda^n}{n!} \qquad \text{with} \quad \lambda > 0, n \in \mathbb{N}.$$

The variance is also equal to λ.

Homogeneous Poisson process in $\mathbb{R}^d$

Definition

The spatial distribution of a homogeneous Poisson process has the two properties:

- for $A \in \mathscr{B}(\mathbb{R}^d)$ the counts $N(A)$ are Poisson distributed with mean $\lambda|A|$:

$$P(N(A) = n) = \mathrm{e}^{-\lambda|A|}\,\frac{(\lambda|A|)^n}{n!},$$

- for pairwise disjoint $A_1, \ldots, A_p \in \mathscr{B}(\mathbb{R}^d)$ the corresponding $N(A_1), \ldots, N(A_p)$ are mutually independent.

The parameter λ is the intensity of the process.

Homogeneous Poisson process in $\mathbb{R}^d$

Definition

The spatial distribution of a homogeneous Poisson process has the two properties:

- for $A \in \mathscr{B}(\mathbb{R}^d)$ the counts $N(A)$ are Poisson distributed with mean $\lambda|A|$:

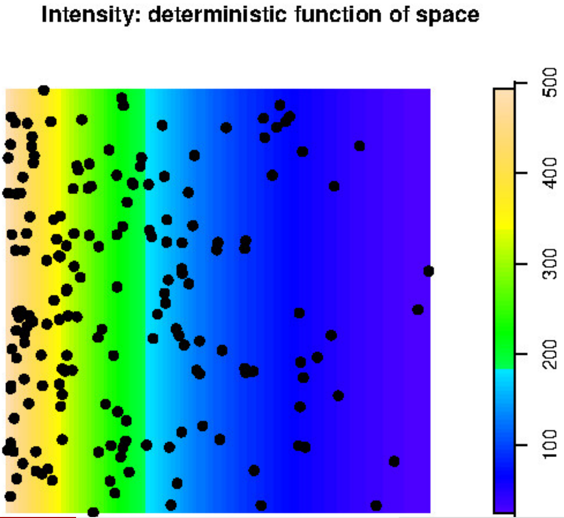
$$P(N(A) = n) = \mathrm{e}^{-\lambda|A|} \frac{(\lambda|A|)^n}{n!},$$

- for pairwise disjoint $A_1, \dots, A_p \in \mathscr{B}(\mathbb{R}^d)$ the corresponding $N(A_1), \dots, N(A_p)$ are mutually independent.

The parameter λ is the intensity of the process.

Heterogeneous Poisson process

The intensity is not constant: it is a function $\lambda(\mathbf{x})$ of space.



Heterogeneous Poisson point process

Definition

The process intensity $\lambda = (\lambda(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d)$ varies through space.
The spatial distribution of a heterogeneous Poisson process has the two properties:

- for $A \in \mathscr{B}(\mathbb{R}^d)$ the counts $N(A)$ are Poisson distributed with mean $\lambda(A) = \int_A \lambda(\mathbf{x})\, d\mathbf{x}$:

$$P(N(A) = n) = \mathrm{e}^{-\lambda(A)} \frac{(\lambda(A))^n}{n!}$$

- for pairwise disjoint $A_1, \ldots, A_p \in \mathscr{B}(\mathbb{R}^d)$ the corresponding $N(A_1), \ldots, N(A_p)$ are mutually independent.

Heterogeneous Poisson point process

Definition

The process intensity $\lambda = (\lambda(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d)$ varies through space.
The spatial distribution of a heterogeneous Poisson process has the two properties:

- for $A \in \mathscr{B}(\mathbb{R}^d)$ the counts $N(A)$ are Poisson distributed with mean $\lambda(A) = \int_A \lambda(\mathbf{x})\, d\mathbf{x}$:

$$P(N(A) = n) = \mathrm{e}^{-\lambda(A)} \frac{(\lambda(A))^n}{n!}$$

- for pairwise disjoint $A_1, \ldots, A_p \in \mathscr{B}(\mathbb{R}^d)$ the corresponding $N(A_1), \ldots, N(A_p)$ are mutually independent.

Heterogeneous Poisson process

Fundamental property

For a Poisson process with $N(A) = n > 0$
the n points are independently distributed in A with the same probability density function:

$$f(\mathbf{x}) \quad = \quad \frac{\lambda(\mathbf{x})}{\lambda(A)} \qquad \text{for} \quad \mathbf{x} \in A$$

Point process characterization of extremes

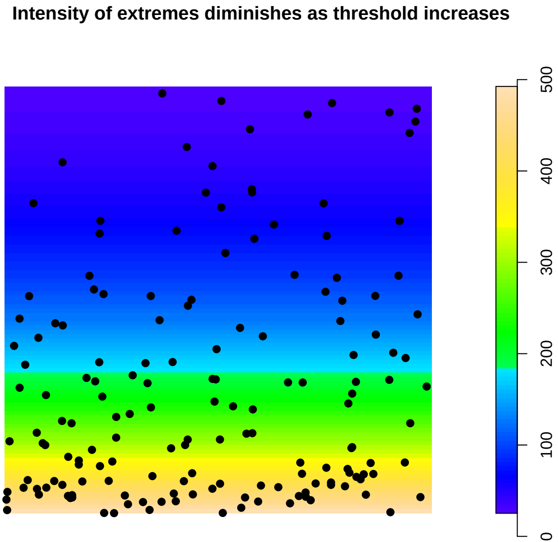
- The theory of point processes opens the door to an *elegant characterization* of extreme value behavior.
- The point process characterization leads to nothing new: all inferences using PP methodology could be obtained with methods already presented.

There are however two good reasons for the PP approach:

1. a unifying framework for the models previously studied,
2. likelihood with more natural formulation of non-stationarity in threshold excesses.

Basic idea

The extremes are viewed as a point process in a region $(0,t) \times [u,\infty)$.



Point process limit for extremes

We assume the $X_1, X_2, \dots$ to be iid random variables for which there are sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$P\left(\frac{M_n - b_n}{a_n} \leq z\right) \to G(z) \qquad \text{as } n \to \infty$$

where

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

Let z_- and z_+ be the lower and upper endpoints of G. Then the sequence of point processes

$$N_n = \left\{\left(\frac{i}{n+1}, \frac{X_i - b_n}{a_n}\right) : i = 1, \dots, n\right\}$$

converges on regions of the form $(0,1) \times [u, \infty)$, for any $u > z_-$, to a Poisson process with intensity function on $A = [t_1, t_2] \times [z, z_+)$ given by:

$$\Lambda(A) = (t_2 - t_1)\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}$$

The intensity function is expressed in terms of parameters of the GEV.

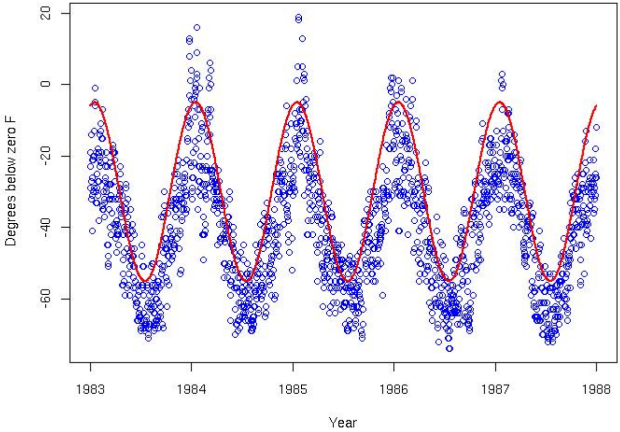
Wooster minimum temperatures

Defining a time-varying threshold

```
data(wooster)
x = seq(along = wooster)
usin = function(x,a,b,d){
       a+ b * sin((( x-d )*2*pi)/365.25) }
wu = usin(x,-30,25,-75)
plot(-wooster ~ x)
lines(wu ~ x)
```

Wooster minimum temperatures

Negated daily minimum temperature in degrees Fahrenheit



- Proposed time-varying threshold

Wooster minimum temperatures

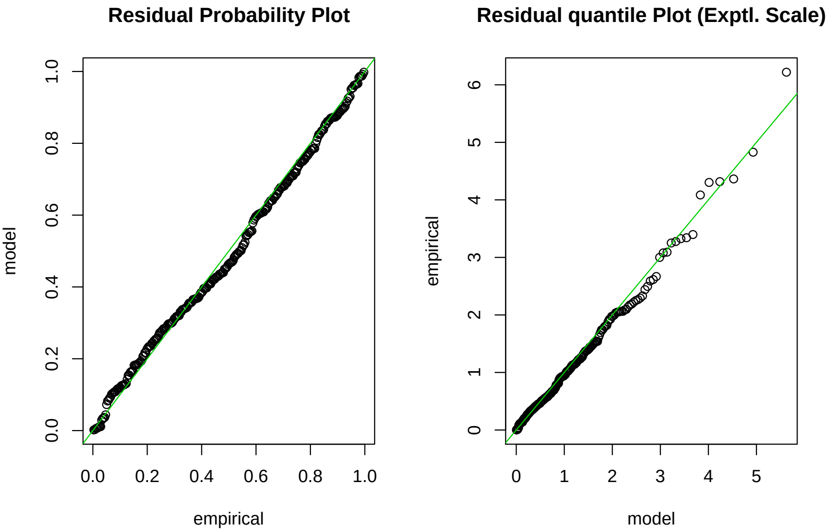
Point process model fit

- Time-varying threshold
- $\mu(t) = \beta_0 + \beta_1 \sin(2\pi t/365) + \beta_2 \cos(2\pi t/365)$

```
# design matrix
ydat=cbind( sin(x*2*pi/365.25), cos(x*2*pi/365.25) )
wooster.pp =
    pp.fit(-wooster,threshold=wu,ydat=ydat,mul=1:2,
    sigl = 1:2, siglink = exp, method = "BFGS")
pp.diag(wooster.pp)
```

Wooster minimum temperatures

Point process model fit



Multivariate extreme value modeling

Lack of data means the precision of extreme value estimates is often poor.
To overcome this, additional information can be incorporated, which requires the formulation of multivariate models. Questions include:

- What issues are important when contemplating multivariate extremes?
- What are appropriate ways to summarize **dependence** in extremes?
- What models are suggested by asymptotic theory?
- How should inference be carried out?

The multivariate extreme value theory will be develpped in some detail.
A central question is the **dependence** of bivariate extremes.

Multivariate extremes: bivariate case

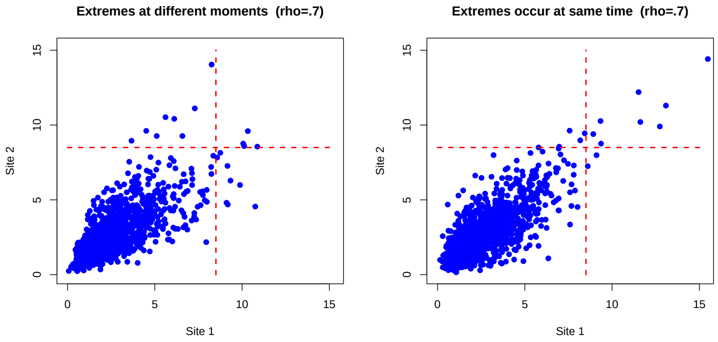
In a multi-variate or a multi-location setting we may wonder:

- how likely is it that an extreme event occurs simultaneously for **two** (or more) **variables**?

- how likely is it that an extreme event occurs simultaneously at **two** (or more) **geographical locations**?

The bivariate distributions contain the answer.

Example: scatterplots between two sites



- The overall correlation coefficient is the same for the two realizations:

$$\rho = 0.7$$

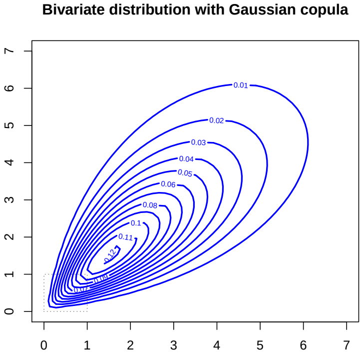
wheras the behaviour of bivariate extremes differs between left and right

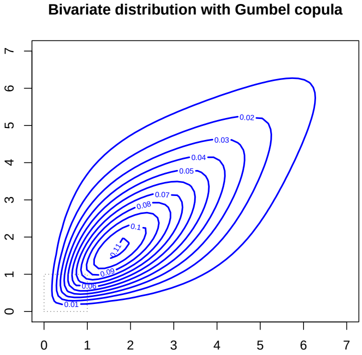
Actually different dependence functions were used to construct these examples.

(for details see EMBRECHTS et al. 2002, p22)

Plot of the two bivariate distributions

Marginals are Gamma(3,1)





Using **Gauss** copula

Using **Gumbel** copula

Discussion

Copulas are a convenient tool for representing bivariate distributions and separating the dependence structure from the marginal distributions.

- The **Gaussian copula** belongs to the family of elliptic copulas, and implies asymptotically independent extremes.
  This may be unrealistic!!!

- The **Gumbel copula** does not belong to that family and is suitable for extreme value analysis/simulation.

Thus, paradoxically, from the point of view of extreme value theory a Gaussian dependence structure is generally not desirable!

An anecdote about financial extremes

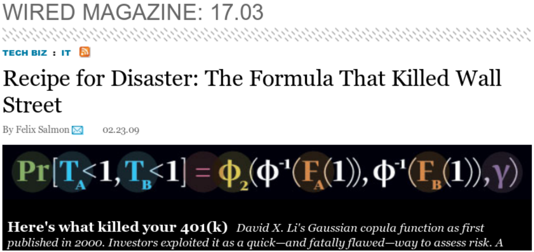
Gaussian copulas were the main ingredient of a formula proposed by Li for financial analysis in 2000.

It has been widely used by financial industry due to its simplicity.

Its inherent **underevaluation of joint risks** was deemed to be partly responsible for the unforeseen advent of the financial crisis of 2007-2009.

From a web paper

(2009):



WIRED MAGAZINE: 17.03

TECH BIZ : IT

Recipe for Disaster: The Formula That Killed Wall Street

By Felix Salmon 02.23.09

$$\Pr[T_A<1, T_B<1] = \phi_2(\phi^{-1}(F_A(1)), \phi^{-1}(F_B(1)), \gamma)$$

Here's what killed your 401(k) *David X. Li's Gaussian copula function as first published in 2000. Investors exploited it as a quick—and fatally flawed—way to assess risk. A*

Bibliography

BACRO, J.-N., AND TOULEMONDE, G.
Measuring and modelling multivariate and spatial dependence of extremes.
Journal de la Société Française de Statistique 154, 2 (2013), 139–155.

BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J., AND TEUGELS, J.
Statistics of Extremes: Theory and Applications.
Wiley, Chichester, 2004.

COLES, S.
An Introduction to Statistical Modeling of Extreme Values.
Springer, London, 2001.

COOLEY, D., CISEWSKI, J., ERHARDT, R. J., JEON, S., MANNSHARDT, E., OMOLO, B. O., AND SUN, Y.
A survey of spatial extremes: measuring spatial dependence and modeling spatial effects.
Revstat 10 (2012), 135–165.

DAVISON, A. C., PADOAN, S. A., AND RIBATET, M.
Statistical modeling of spatial extremes.
Statistical Science 27 (2012), 161–186.

EMBRECHTS, P., KLÜPPELBERG, C., AND MIKOSCH, T.
Modelling Extremal Events for insurance and finance.
Springer, Berlin, 1997.

EMBRECHTS, P., MCNEIL, A., AND STRAUMANN, D.
Correlation and dependence in risk management: properties and pitfalls.
In *Risk Management: Value at Risk and Beyond* (2002), M. A. H. Dempster, Ed., Cambridge University Press, pp. 176–223.