Week 9: Extreme Value Theory

MATH-516 Applied Statistics

Linda Mhalla

2025-04-14

Section 1

Introduction

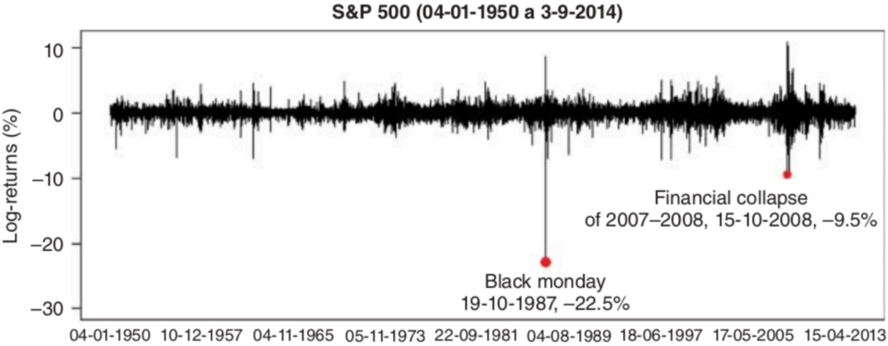
Motivation for modelling extreme events

Modelling extremes in environmental sciences

- **Temperatures** → heat waves (Europe, 2003 → 40'000 deaths and €13.1 billion of crop damages)
- **Water heights** → floods (hurricane Harvey, 2017 → 107 deaths and $125 billion in damages)
- **Concentrations of air pollutants** → health problems

Motivation for modelling extreme events

Modelling extremes in finance



Growing areas of application include: insurance, athletic records, networks

Basic problem

- Let X be a random variable of interest with cdf F
- We are interested in cases where X is “extremely” large or “extremely” low, i.e.,

$$\Pr(X > x) \text{ when } x \text{ is large, or } \Pr(X < x) \text{ when } x \text{ is low}$$

Therefore, we require accurate inference on the tails of F. But...

- There are very few observations in the tails of the distribution $\rightarrow$ standard techniques can result in severely biased estimates
- We often require estimates that are beyond the observed values

$\rightarrow$ **Rely on the extreme value paradigm**: base tail models on asymptotically-motivated distributions!

How bad does it get?

We want to study the worst case scenario

Two classical approaches

- Block-maxima: $\max(X_1, \dots, X_n)$ (maximum over, e.g., a year)
- Peaks over threshold: $X | X > u$ for a large threshold u

References

- **Books**
  - *Resnick (1987)*: *Extreme Values, Regular Variation, and Point Processes*, Springer
  - *de Haan and Ferreira (2006)*: *Extreme Value Theory: An Introduction*, Springer
  - *Embrechts, Klüppelberg and Mikosch (1997)*: *Modelling Extreme Events for Insurance and Finance*, Springer
  - *Coles (2001)*: *An Introduction to Statistical Modeling of Extreme Values*, Springer
  - *Beirlant, Goegebeur, Segers, and Teugels (2004)*: *Statistics of Extremes: Theory and Applications*, Wiley
  - *Finkenstädt and Rootzén (2004)*: *Extreme Values in Finance, Telecommunications and the Environment*, CRC
  - *Embrechts, Hofert, and Chavez-Demoulin (2024)*: *Risk revealed*, Cambridge University Press

- **R Packages**
  `evd`, `evdbayes`, `evir`, `extRemes`, `fExtremes`, `POT`, `SpatialExtremes`

- **Journal**: *Extremes* (published by Springer)

Section 2

Block-maxima Approach

Notations

- Let $X_1, X_2, \ldots$ be iid random variables with distribution function F
- We seek approximations to the distribution of the maximum of the X_i
- Let $M_n = \max(X_1, \ldots, X_n)$ be the worst-case value in a sample of n values. Clearly

$$\mathbb{P}(M_n \leq x) = \mathbb{P}(X_1 \leq x, \ldots, X_n \leq x) = F^n(x)$$

- F is unknown, so approximate F^n by some limit distribution, but as $n \to \infty$,

$$F(x)^n \to \begin{cases} 0, & F(x) < 1, \\ 1, & F(x) = 1, \end{cases}$$

  so $M_n \xrightarrow{d} x^*$, where $x^* = \sup\{x : F(x) < 1\}$ is the upper end point of F
- This is not useful, because the distribution is concentrated at x_F
- But what about normalized maxima?

Limiting Behaviour of Sums or Averages

- We are familiar with the central limit theorem
- Let $X_1, X_2, \ldots$ be iid with finite mean μ and finite variance σ^2. Let $S_n = X_1 + \ldots + X_n$. Then

$$\mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq x\right) \xrightarrow{n \to \infty} \Phi(x)$$

where Φ is the cdf of the standard normal distribution

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du$$

- More generally, the limiting distributions for appropriately normalized sample sums are the class of α-stable distributions; Gaussian distribution is a special case

Limiting Behaviour of Sample Maxima

- Let $X_1, X_2, \ldots$ be iid from F and let $M_n = \max(X_1, \ldots, X_n)$

Extremal types theorem

Suppose we can find sequences of real numbers $a_n > 0$ and b_n such that $(M_n - b_n)/a_n$, the sequence of normalized maxima, converges in distribution, i.e.,

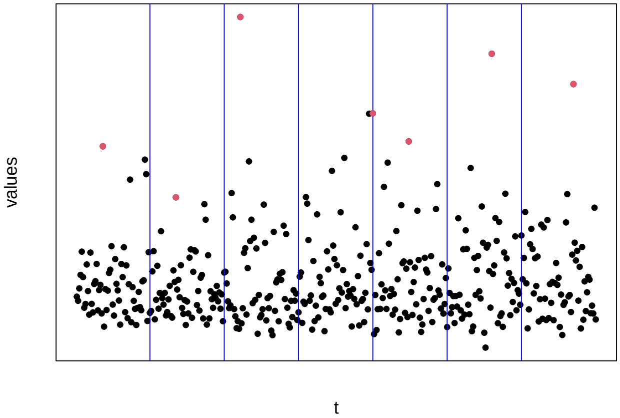
$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n) \xrightarrow{n\to\infty} G(x)$$

for some non-degenerate df $G(x)$. Then, this must be the **generalized extreme-value distribution (GEV)**

$$G_{\xi,\mu,\sigma}(x) = \begin{cases} \exp\left[-\left\{1 + \xi(x-\mu)/\sigma\right\}_+^{-1/\xi}\right], & \xi \neq 0, \\ \exp\left[-\exp\left\{-(x-\mu)/\sigma\right\}\right], & \xi = 0, \end{cases} \quad x \in \mathbb{R},$$

where $a_+ = \max(a, 0)$ for any real a, and with $\xi, \mu \in \mathbb{R}$ and $\sigma > 0$. Put another way, $(M_n - b_n)/a_n \xrightarrow{d} Z$ as $n \to \infty$, where Z has distribution function G

Block maxima

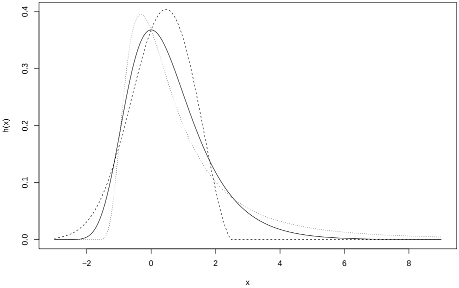


$\rightsquigarrow$ *generalized extreme value limit distribution* for rescaled maxima

Generalized Extreme Value Distribution

The parametrization is continuous in the shape parameter ξ which determines the rate of tail decay. For

- $\xi > 0$: the heavy-tailed Fréchet (Type II) (dotted line)
- $\xi = 0$: the light-tailed Gumbel, Type I, with support on $\mathbb{R}$ (solid line)
- $\xi < 0$: the short-tailed (reverse) Weibull, Type III (dashed line)



Examples: Rainfall or financial data (usually $\xi > 0$), temperature data (usually $\xi < 0$), and Gaussian data ($\xi = 0$)

Generalized Extreme Value Distribution

- If ETT applies, we say that F is in the **maximum domain of attraction** of G, abbreviated $F \in MDA(G)$
- μ and σ are location and scale parameters: not crucial as they can be absorbed by the normalizing sequences, i.e., $G_{\xi,\mu,\sigma}(x) := G_\xi\left(\frac{x-\mu}{\sigma}\right)$. Thus, we can always choose normalizing sequences a_n and b_n so that the limit law G_ξ appears in standard form (without relocation or rescaling)
- The rth moment of the GEV exists only if $\xi < 1/r$, so the mean exists only if $\xi < 1$, the variance only if $\xi < 1/2$, etc. In applications (particularly in finance) some moments may not exist
- Essentially, all commonly encountered continuous distributions are in the maximum domain of attraction of an extreme value distribution

ETT - Fisher–Tippett Theorem (1928): Examples

Recall: $F \in \text{MDA}(G_\xi)$, iff there are sequences a_n and b_n with

$$\mathbb{P}\left\{(M_n - b_n)/a_n \leq x\right\} = F^n(a_n x + b_n) \stackrel{n\to\infty}{\longrightarrow} G(x)$$

- The **exponential distribution**

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

  is in MDA(G_0) (Gumbel). Take $a_n = 1/\lambda$, $b_n = (\log n)/\lambda$

- The **Pareto distribution**

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

  is in MDA($G_{1/\alpha}$) (Fréchet). Take $a_n = \kappa n^{1/\alpha}/\alpha$, $b_n = \kappa n^{1/\alpha} - \kappa$

When does $F \in \mathsf{MDA}(G_\xi)$ hold?

Fréchet case: $(\xi > 0)$

- Gnedenko (1943) showed that for $\xi > 0$

$$F \in \mathsf{MDA}(G_\xi) \iff 1 - F(x) = x^{-1/\xi} L(x)$$

for some slowly varying function $L(x)$

- A function L on $(0, \infty)$ is slowly varying if

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1, \quad t > 0$$

Summary: If the tail of the distribution function F decays like a power function, then the distribution is in $\mathsf{MDA}(G_\xi)$ for $\xi > 0$
Examples: Heavy-tailed distributions such as Pareto, Burr, log-gamma, Cauchy, and t-distributions as well as various mixture models. Not all moments are finite

When does $F \in \text{MDA}(G_\xi)$ hold?

Gumbel case: $F \in \text{MDA}(G_0)$

- The characterization of this class is more complicated. Essentially, it contains distributions whose tails decay roughly exponentially and we call these distributions light-tailed. All moments exist for distributions in the Gumbel class
- Examples are the normal, log-normal, exponential, and gamma

Using Fisher–Tippett on data: Block Maxima Method

If you are given n values, use the limiting distribution to model M_n:

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) \approx G_{\xi,0,1}(x)$$

or

$$\mathbb{P}(M_n \leq y) = G_{\xi,b_n,a_n}(y)$$

- All that's left is to estimate three parameters: ξ, b_n, and a_n
- Need repeated values of $M_n \Rightarrow$ required data is a multiple of n

The values b_n and a_n are equivalent to the parameters μ and σ in the formula, respectively

ML Inference for Maxima

We have block maxima data $\mathbf{y} = \left(M_n^{(1)}, \dots, M_n^{(m)}\right)^\top$ from m blocks of size $n \rightarrow$ want to estimate $\theta = (\xi, \mu, \sigma)^\top$

We construct a **log-likelihood** by assuming we have independent observations from a GEV with density g_θ,

$$l(\theta; \mathbf{y}) = \log \left\{ \prod_{i=1}^{m} g_\theta \left(M_n^{(i)} \right) \mathbb{1}_{\left\{ 1+\xi\left(M_n^{(i)}-\mu\right)/\sigma>0 \right\}} \right\}$$

and (numerically) maximize this w.r.t. θ to obtain the MLE $\hat{\theta} = (\hat{\xi}, \hat{\mu}, \hat{\sigma})^\top$

ML Inference for Maxima

- When $\xi > -0.5$, maximum likelihood estimator obeys the standard theory. In particular
  - standard errors can be computed from inverse of the observed information matrix
  - likelihood ratio test applies to nested models
  - profile log-likelihood preferred to construct CIs and perform tests for quantiles
- If $\xi \leq -0.5$, Bayesian methods may be preferable (this is very rare in practice!)

Clearly, when defining blocks, **bias** and **variance** must be traded off

- we reduce bias by increasing the block size n
- we reduce variance by increasing the number of blocks m

Risk Measures

- We have a time series of daily values $X_1, X_2, \ldots$, assumed to be independent and identically distributed from F
- We aim to estimate some measure of risk of high (or low) values of X
- Common risk measures:
  - **Probability**: $\Pr(X > v) = 1 - F(v)$ for some high threshold v
  - High **quantile**: x_{1-p} corresponding to some small p, i.e., $x_{1-p} = F^{-1}(1-p)$
  - **Value-at-Risk (VaR$_{1-p}$)**: high quantile x_{1-p} used for financial losses
    - where X denotes 1-day or 10-day losses (negative returns) and typically $p = 0.01$ or 0.05
  - **Expected Shortfall (ES$_{1-p}$)**: $\mathbb{E}(X \mid X > x_{1-p})$

Return Levels

- Aim: What is the 40-year return level $R_{365,40}$?
- We define a rare **stress** $R_{n,k}$, the k n-block return level, as

$$\mathbb{P}(M_n > R_{n,k}) = \frac{1}{k}$$

i.e., it is the level that is exceeded in one out of every k n-blocks, on average

In extreme value terminology, $R_{n,k}$ is the **return level** associated with **return period** $1/k$ (small as k is typically large)

If M_n are yearly maxima, then $R_{n,k}$ represents the level that is expected to be exceeded once every k years

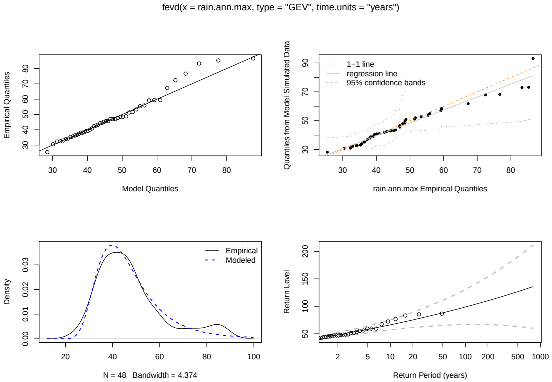
- We use the approximation

$$R_{n,k} \approx G^{-1}_{\xi,\mu,\sigma}\left(1 - \frac{1}{k}\right) = \mu + \frac{\sigma}{\xi}\left[\left\{-\log(1 - \frac{1}{k})\right\}^{-\xi} - 1\right]$$

The interest is then in estimating this functional of the unknown parameters of our GEV model for maxima of n-blocks

GEV in practice: Daily rainfall in south-west England

```
library(ismev)
library(extRemes)

data(rain) #from ismev
years <- rep(1:48, rep(c(365,365,366,365), times = 12))[-17532] #period 1914 to 1962
rain.ann.max <- unlist(lapply(X = split(rain,years), FUN = max)) # annual maxima

mod <- fevd(rain.ann.max, type="GEV", time.units="years")
plot(mod)
```



GEV in practice: Daily rainfall in south-west England

```r
mod $results$par #gives parameters of the GEV
```

```
  location      scale      shape 
40.7830335  9.7284060  0.1072355 
```

```r
# suggests heavy-tailed model here, but only point estimates
# What about building confidence intervals?
ci.fevd(mod, alpha=0.05, type="parameter")
```

```
fevd(x = rain.ann.max, type = "GEV", time.units = "years")

[1] "Normal Approx."

         95% lower CI   Estimate 95% upper CI
location   37.6941916 40.7830335   43.8718754
scale       7.3991505  9.7284060   12.0576614
shape      -0.1055497  0.1072355    0.3200207
```

```r
# the CI includes 0, so not sure we're that heavy-tailed
gev.rl <- return.level(x = mod, return.period = c(10,100,1000),
                       do.ci = TRUE, alpha = 0.05)
gev.rl
```

```
fevd(x = rain.ann.max, type = "GEV", time.units = "years")

[1] "Normal Approx."

                        95% lower CI  Estimate 95% upper CI
10-year return level        56.67333  65.54301     74.41268
100-year return level       66.85346  98.63615    130.41884
1000-year return level      58.37286 140.34002    222.30718
```

GEV in practice: Daily rainfall in south-west England

What is the 10–period **return level** $R_{365,10}$? i.e., the level that is exceeded once every 10 years, on average

$$\hat{R}_{365,10} \approx \hat{G}^{-1}_{\hat{\xi},\hat{\mu},\hat{\sigma}}(1-1/k) = 40.78 + 9.73 \frac{\left[\{-\log(1-1/10)\}^{-0.11} - 1\right]}{0.11}$$

$\approx$ 65.62 mm is the estimated value of daily rainfall that can be exceeded once every 10 years

Confidence intervals:

- Rely on the normal approximation of the distribution of MLE + Delta method (or profile likelihood)
- Rely on parametric or non-parametric bootstrap

Section 3

Threshold Exceedances

Exceedance Theorem

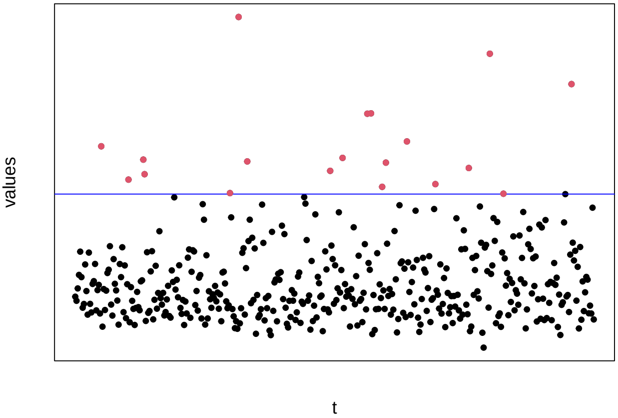
Theorem (Exceedance)

Let X be a random variable having distribution function F, and suppose that a function $c(u)$ can be chosen so that the limiting distribution of $(X-u)/c(u)$, conditional on $X > u$, is non-degenerate as u approaches the upper support value $x^* = \sup\{x : F(x) < 1\}$ of X.
If such a limiting distribution exists, it must be of *generalized Pareto* form, i.e.,

$$H(x) = \begin{cases} 1 - (1 + \xi x/\sigma)_+^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - \exp(-x/\sigma) & \text{if } \xi = 0, \end{cases} \quad x > 0,$$

where $\xi \in \mathbb{R}$ and $\sigma > 0$. This is known as the *generalized Pareto distribution (GPD)*

Threshold exceedances



$\rightsquigarrow$ *generalized Pareto distribution* for rescaled exceedances

Remarks on the Exceedance Theorem

- There is a close connection with the **Extreme Type Theorem (ETT)**, which applies for maxima under the same conditions as the **Exceedance Theorem (ET)** applies for exceedances, and with the same ξ and $\sigma = \sigma_{GEV} + \xi(u - \mu)$
- The GPD is a natural model for exceedances over high thresholds (and under low ones, using $1 - H(-x)$)
- The GPD is the only **threshold-stable** distribution, satisfying

$$\frac{1 - H(x+u)}{1 - H(u)} = 1 - H(x/\sigma_u), \quad 0 < u < u + x < x_H,$$

  for some function $\sigma_u > 0$, where x_H is the upper support point of the density of H

Threshold choice

The GPD approach requires a threshold u to be chosen

- Choosing u too low leads to **bias** (model inappropriate), while too high a u increases **variance** (too few exceedances)

If $X \sim \text{GPD}(\sigma, \xi)$, then the conditional distribution satisfies
$X - u \mid X > u \sim \text{GPD}(\sigma + \xi u, \xi)$,
which implies:

$$\mathbb{E}(X - u \mid X > u) = \frac{\sigma + \xi u}{1 - \xi}, \quad \xi < 1,$$

so a **mean excess plot** (or **mean residual life plot**) of

$$\frac{\sum_j (x_j - u)\mathbb{I}(x_j > u)}{\sum_j \mathbb{I}(x_j > u)} \quad \text{against} \quad u$$

should be approximately straight with slope $\xi/(1-\xi)$ above $u_{\min}$

- You can also test for equal shape parameters above u using the **Northrop–Coleman test**

Daily rainfall: Threshold analysis

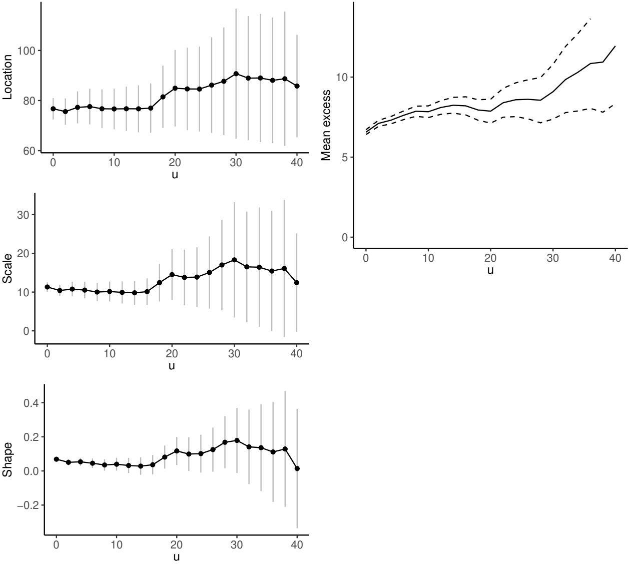


Figure 1: Threshold selection plots

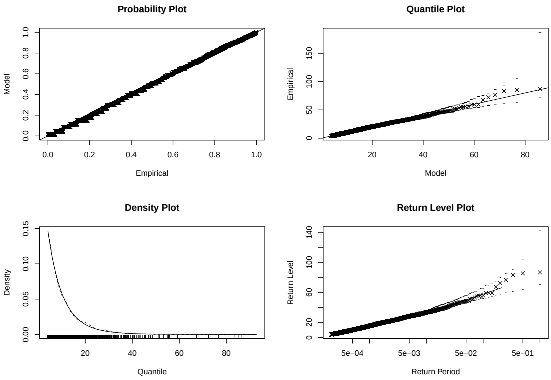
Daily rainfall: GPD fit

```
fit.gpd <- fpot(rain, threshold= 4) #likelihood-based estimation
fit.gpd
```

```
Call: fpot(x = rain, threshold = 4)
Deviance: 27950.18

Threshold: 4
Number Above: 4681
Proportion Above: 0.267

Estimates
  scale    shape
6.70792  0.08208

Standard Errors
  scale    shape
0.14735  0.01644

Optimization Information
  Convergence: successful
  Function Evaluations: 21
  Gradient Evaluations: 6
```

Daily rainfall: GPD fit

```r
par(mfrow = c(2,2))
plot(fit.gpd)
```



Section 4

Non-stationary Extremes

Modelling Issues

Extreme value data usually show:

- Short term dependence (storms for example); clustering effect and extremal index $\rightarrow$ not covered in this short course about EVT
- Seasonality (due to annual cycles in meteorology)
- Long-term trends (due to gradual climatic change)
- Dependence on covariate effects
- Other forms of non-stationarity

For (short-term) temporal dependence, there is a sufficiently wide-ranging theory which can be invoked (requires some sort of mixing conditions at extreme levels of a stationary series). Other aspects have to be handled at the modelling stage

Non-stationarity Example: Daily mean temperature in Lausanne

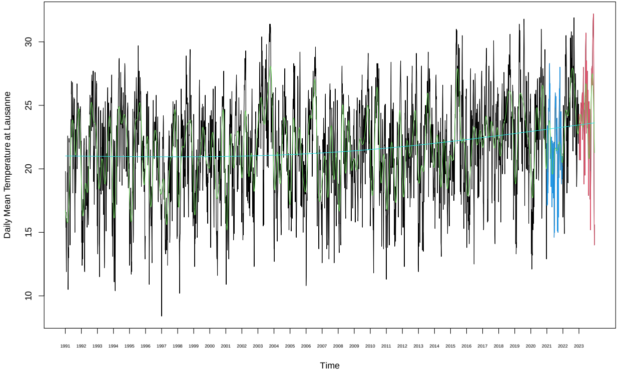


Figure 2: Daily mean temperature in Lausanne

Non-stationarity Example: Dailymean temperature during summer

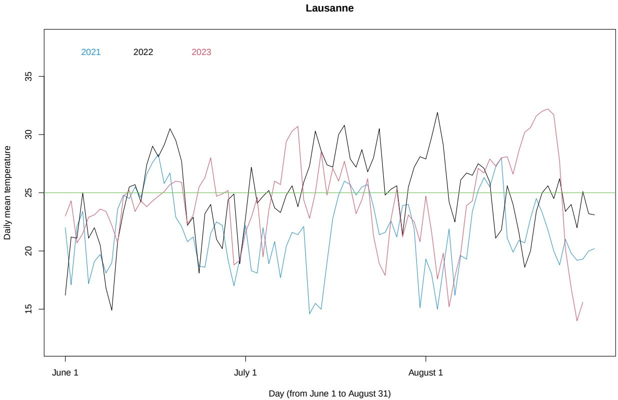


Figure 3: Daily mean temperature during Summer

Non-stationarity

Model trends, seasonality and covariate effects by parametric or nonparametric models for the usual extreme value model parameters

Some possibilities for parametric modelling:

- $\mu(t) = \alpha + \beta t$
- $\sigma(t) = \exp(\alpha + \beta t)$
- $\xi(t) = \begin{cases} \xi_1, & t \leq t_0 \\ \xi_2, & t > t_0 \end{cases}$
- $\mu(t) = \alpha + \beta y(t)$

Parameter Estimation

- **Model specification (example):**

$$Z_t \sim \mathsf{GEV}\{\mu(t), \sigma(t), \xi(t)\}$$

- **Likelihood (for complete parameter set β):**

$$L(\beta) = \prod_{t=1}^{m} g\{z_t; \mu(t), \sigma(t), \xi(t)\},$$

  where h is GEV model density

- Maximization of L yields maximum likelihood estimates
- Standard likelihood techniques also yield standard errors, confidence intervals, etc

Model Reduction

- For nested models $\mathcal{M}_0 \subset \mathcal{M}_1$, the deviance statistic is:

$$D = 2\{\ell_1(\mathcal{M}_1) - \ell_0(\mathcal{M}_0)\}$$

- Based on asymptotic likelihood theory, $\mathcal{M}_0$ is rejected by a test at the α-level of significance if $D > c_\alpha$, where c_α is the $(1-\alpha)$ quantile of the χ^2_k distribution, and k is the difference in the dimensionality of $\mathcal{M}_1$ and $\mathcal{M}_0$

Example: Race times[1]

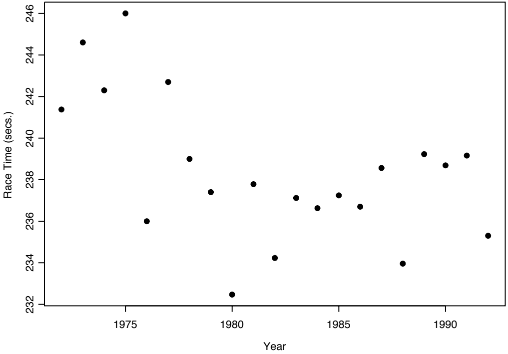


Figure 4: Annual fastest race times for women's 1500m event, with an obvious time trend

[1] From the excellent introductory book: Coles, 2001

Example: Race times

Model	Log-likelihood	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\xi}$
Constant	-54.5	239.3 (0.9)	3.63 (0.64)	-0.469 (0.141)
Linear	-51.8	$(242.9, -0.311)$ (1.4, 0.101)	2.72 (0.49)	-0.201 (0.172)
Quadratic	-48.4	$(247.0, -1.395, 0.049)$ (2.3, 0.420, 0.018)	2.28 (0.45)	-0.182 (0.232)

Quadratic model appears preferable

Example: Race times

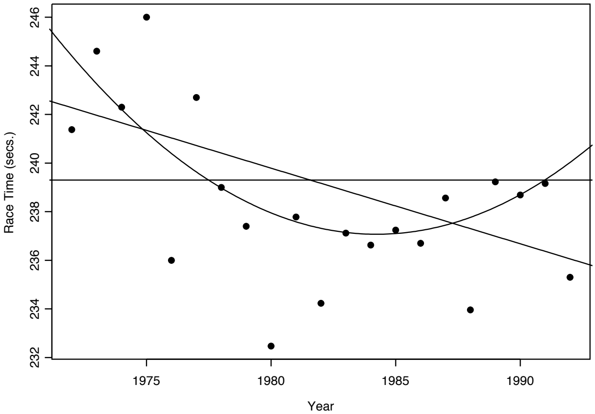


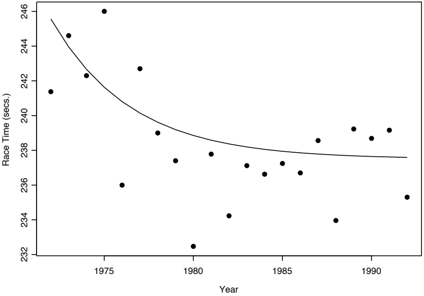
Figure 5: Fitted models for location parameter in women's 1500 metre race times. Note quadratic model would lead to slower races in recent and future events

Example: Race times

Alternative exponential model

$$\tilde{\mu}(t) = \beta_0 + \beta_1 e^{-\beta_2 t}$$

has log-likelihood -49.5. Not as good as the quadratic model, though comparison via likelihood ratio test is invalid as models are not nested. Better behaviour for large t suggests a preferable model though



Example: Spatial modelling of rainfall extremes

```r
library(evgam)
library(knitr)

data("COprcp", package = "evgam")
COprcp      <- cbind(COprcp, COprcp_meta[COprcp$meta_row, ])
COprcp$year <- format(COprcp$date, "%Y")
COprcp_gev  <- aggregate(prcp ~ year + meta_row, COprcp, max)
COprcp_gev  <- cbind(COprcp_gev, COprcp_meta[COprcp_gev$meta_row, ])

head(COprcp_gev)
```

```
    year meta_row prcp          id        name       lon     lat   elev
1   1990        1 43.2 USC00050263 ANTERO RSVR -105.8919 38.9933 2718.8
1.1 1991        1 14.7 USC00050263 ANTERO RSVR -105.8919 38.9933 2718.8
1.2 1992        1 44.7 USC00050263 ANTERO RSVR -105.8919 38.9933 2718.8
1.3 1993        1 11.2 USC00050263 ANTERO RSVR -105.8919 38.9933 2718.8
1.4 1994        1 30.5 USC00050263 ANTERO RSVR -105.8919 38.9933 2718.8
1.5 1995        1 26.7 USC00050263 ANTERO RSVR -105.8919 38.9933 2718.8
```

```r
tail(COprcp_gev)
```

```
      year meta_row prcp          id          name       lon   lat   elev
64.24 2014       64 21.6 USW00093058 PUEBLO MEM AP -104.4983 38.29 1438.7
64.25 2015       64 41.9 USW00093058 PUEBLO MEM AP -104.4983 38.29 1438.7
64.26 2016       64 27.7 USW00093058 PUEBLO MEM AP -104.4983 38.29 1438.7
64.27 2017       64 38.4 USW00093058 PUEBLO MEM AP -104.4983 38.29 1438.7
64.28 2018       64 16.8 USW00093058 PUEBLO MEM AP -104.4983 38.29 1438.7
64.29 2019       64 42.4 USW00093058 PUEBLO MEM AP -104.4983 38.29 1438.7
```

Example: Spatial modelling of rainfall extremes

```r
library(evgam)

fmla_gev <- list(prcp ~ s(lon, lat, k = 30) + s(elev, bs = "cr"),
                 ~ s(lon, lat, k = 20), ~ 1) #formula for each GEV parameter
m_gev    <- evgam(fmla_gev, COprcp_gev, family = "gev") #fit the model
summary(m_gev)
```

```
** Parametric terms **

location
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    28.56       0.26  111.89   <2e-16

logscale
            Estimate Std. Error t value Pr(>|t|)
(Intercept)     2.24       0.02  118.07   <2e-16

shape
            Estimate Std. Error t value Pr(>|t|)
(Intercept)     0.08       0.02    5.08 1.92e-07

** Smooth terms **

location
              edf max.df Chi.sq Pr(>|t|)
s(lon,lat) 19.27     29 178.23   <2e-16
s(elev)     5.19      9  19.39  0.00139

logscale
              edf max.df Chi.sq Pr(>|t|)
s(lon,lat) 13.94     19 211.15   <2e-16
```

Example: Spatial modelling of rainfall extremes