4 Complications slide 107

4.1 Introduction slide 108

Basic ideas

- ☐ Chapter 3 described 'vanilla' statistical analyses for rare events using the GEV, GPD and point process methods.
- ☐ The basic derivations of these models assume that

$$X_1, \ldots, X_m \overset{\text{iid}}{\sim} F, \quad m \to \infty.$$

- ☐ In applications these assumptions are generally false:
  - m is finite;
  - the background data may show trend, seasonality or other forms of **non-stationarity**, so $X_j \sim F_j$;
  - time series are typically **dependent**, as cold weather, heatwaves, ... occur over several days;
  - some (maybe subtle) **selection** mechanism may apply, e.g., when an analysis is performed immediately after a rare event.
- ☐ This chapter will describe methods for detecting and dealing with these problems.

http://stat.epfl.ch slide 109

4.2 Nonstationarity slide 110

Vanilla analysis of maxima

- ☐ Our previous analyses supposed that
  - block maxima satisfy $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \text{GEV}(\eta, \tau, \xi)$,
  - exceedances of a threshold u satisfy $X_1 - u, \ldots, X_n - u \overset{\text{iid}}{\sim} \text{GPD}(\sigma, \xi)$,

  but often we observe additional variation, either due to
  - **systematic** changes in the background data (e.g., due to trend or seasonality), or to
  - **haphazard** variation (e.g., due to weather conditions) that we have not accounted for.
- ☐ We'll pass most time looking at systematic changes.
- ☐ For an example of haphazard variation, consider annual maximum daily rainfall $M = \max(X_1, \ldots, X_{365})$, where X_j is total rainfall on day j. On many days $X_j = 0$, so

$$M = \max(X_1, \ldots, X_N),$$

where $N \ll 365$ is the (random) number of rainy days. If N varies a lot from year to year, then M might be much smaller in some years than in others, so the GEV is a poor model (remember we derive it assuming that $X_j \sim F$, where F is continuous ...).

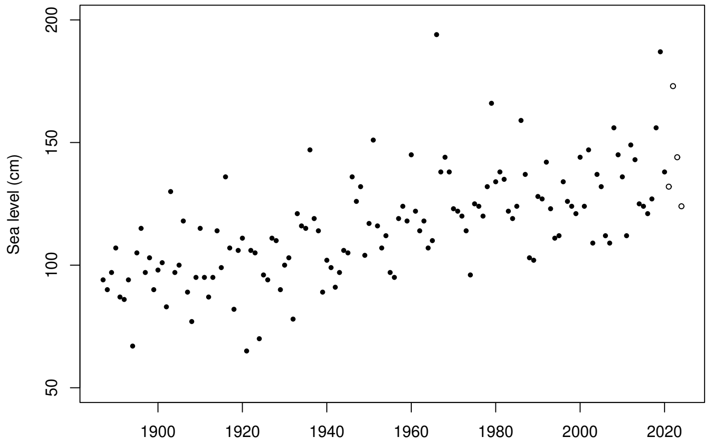
http://stat.epfl.ch slide 111

A damp day in Venice

Punta della Dogana and Santa Maria della Salute

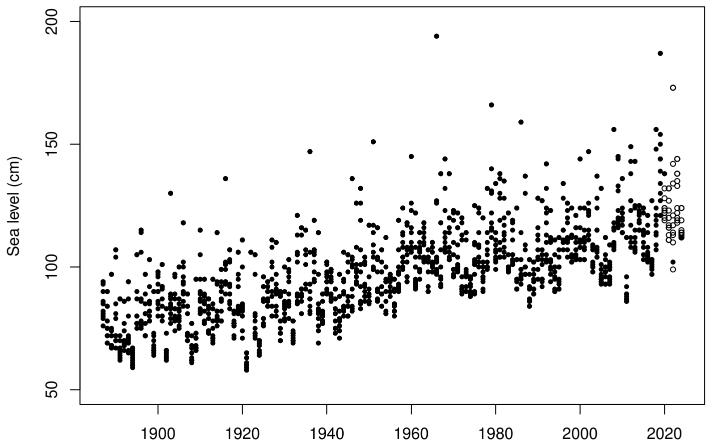
Annual maximum sea levels, 1887–2024

In October 2020, the MOdulo Sperimentale Elettromeccanico (MOSE) system was inaugurated: rows of mobile gates are raised when particularly high tides are predicted, in order to limit how much water from the Adriatic Sea can enter the Venetian lagoon. Data with MOSE operational are shown by circles. The record: 196 cm in 1996.



Ten largest annual sea levels, 1887–2024

In 1935, only the six largest values are available, and in 1922 only the largest value is available. The data sources for 1887–1981 and 1982 onwards are different.



Non-stationarity

- Obvious approach is to suppose that the GEV parameters can depend on external factors, i.e., $Y_t \sim \text{GEV}(\eta_t, \tau_t, \xi_t)$, where the dependence might be specified as

$$\begin{aligned}
\eta_t(\beta) &= \beta_0 + \beta_1 t, \\
\eta_t(\beta) &= \beta_0 + \sum_{k=1}^{K} \{\beta_{2k-1} \cos(2\pi k t/365) + \beta_{2k} \sin(2\pi k t/365)\}, \\
\eta_t(\beta) &= \beta_0 + \beta_1 x(t), \\
\tau_t(\beta) &= \exp(\beta_0 + \beta_1 t), \\
\xi_t(\beta) &= \begin{cases} \beta_1, & t \le t_0, \\ \beta_2, & t > t_0, \end{cases}
\end{aligned}$$

  where $x(t)$ is some physical quantity that varies over time (e.g., ENSO, NAO, or global average temperature).
- In applications we typically find that
  - the location parameter η varies,
  - the scale parameter τ might or might not vary,
  - the shape parameter ξ is constant (it is difficult to estimate, and anyway often is regarded as an intrinsic aspect of the background process).

Parametric inference

- Example model specification: $y_t \overset{\text{ind}}{\sim} \text{GEV}(\eta_t, \tau_t, \xi_t)$, where η_t, τ_t, ξ_t depend on parameters β.
- If $y_1, \ldots, y_n$ are assumed to be independent, then the log likelihood for β is

$$\ell(\beta) = \sum_{t=1}^{n} \log g\{y_t; \eta_t(\beta), \tau_t(\beta), \xi_t(\beta)\},$$

  where g is the GEV density.
- Maximization of $\ell(\beta)$ yields maximum likelihood estimates and the observed information matrix, from which we compute standard errors, confidence intervals, etc.
- We say that model $\mathcal{M}_0$ is nested within a model $\mathcal{M}_1$ if $\mathcal{M}_1$ reduces to $\mathcal{M}_0$ by fixing (say) d parameters. Then the corresponding maximised log likelihoods satisfy $\widehat{\ell}_1 \ge \widehat{\ell}_0$, and the likelihood ratio statistic (or equivalently difference in deviances) is

$$W = 2(\widehat{\ell}_1 - \widehat{\ell}_0).$$

- If $\mathcal{M}_0$ is adequate, then asymptotic likelihood theory implies that $W \dot{\sim} \chi^2_d$, so values of W larger than the $1-\alpha$ quantile of the χ^2_d distribution would lead to a rejection of $\mathcal{M}_0$ in favour of $\mathcal{M}_1$, at significance level α.

Model diagnostics

- If $Y_t \sim \text{GEV}(\eta_t, \tau_t, \xi_t)$ for $t = 1, \ldots, n$, then

$$Z_t = \frac{1}{\xi_t} \log\left(1 + \xi_t \frac{y_t - \eta_t}{\tau_t}\right) \overset{\text{iid}}{\sim} \text{standard Gumbel},$$

  i.e.,

$$\mathrm{P}(Z_t \le z) = \exp\{-\exp(-z)\}, \quad z \in \mathbb{R}, \quad t = 1, \ldots, n.$$

- If we replace the parameters by their estimates $\widehat{\eta}_t = \eta_t(\widehat{\beta})$, etc., these results should still hold (approximately) for the **Gumbel residuals**

$$\widehat{z}_t = \frac{1}{\widehat{\xi}_t} \log\left(1 + \widehat{\xi}_t \frac{y_t - \widehat{\eta}_t}{\widehat{\tau}_t}\right), \quad t = 1, \ldots, n.$$

- We use the $\widehat{z}_t$ in diagnostic plots, e.g.,
  - the **probability plot**, showing $\left\{j/(n+1), \exp\{-\exp(-\widehat{z}_{(j)})\};\ j = 1, \ldots, n\right\}$, or
  - the **quantile plot**, showing $\left\{\left(-\log\left[-\log\{j/(n+1)\}\right], \widehat{z}_{(j)}\right);\ j = 1, \ldots, n\right\}$, or

  plots of the $\widehat{z}_j$ against appropriate variables, to see if any patterns remain after fitting the model.

Example: Venice sea levels, 1887–2019

- We ignore the data from 2020 onwards, when MOSE is operational.
- Analysis of maxima uses straight-line regression model,

$$\eta_t = \beta_0 + \beta_1 x_t, \quad t = 1, \ldots, n = 133,$$

  with $(x_1, \ldots, x_{133}) = (1887 - 1900, \ldots, 2019 - 1900)/100$ chosen so that
  - β_0 equals the location parameter in the year 1900,
  - β_1 denotes the change in maximum sea level over 100 years,
- We fit two nested models, both with constant scale and shape parameters, i.e.,

$$\mathcal{M}_0: \quad \eta_t = \beta_0, \quad \tau_t \equiv \tau, \quad \xi_t \equiv \xi,$$
$$\mathcal{M}_1: \quad \eta_t = \beta_0 + \beta_1 x_t, \quad \tau_t \equiv \tau, \quad \xi_t \equiv \xi.$$

- The code prints a 'deviance' $D = -2\widehat{\ell}$ (or `nllh`$= -\widehat{\ell}$) for the fitted model, which allows model comparison using the likelihood ratio statistic:

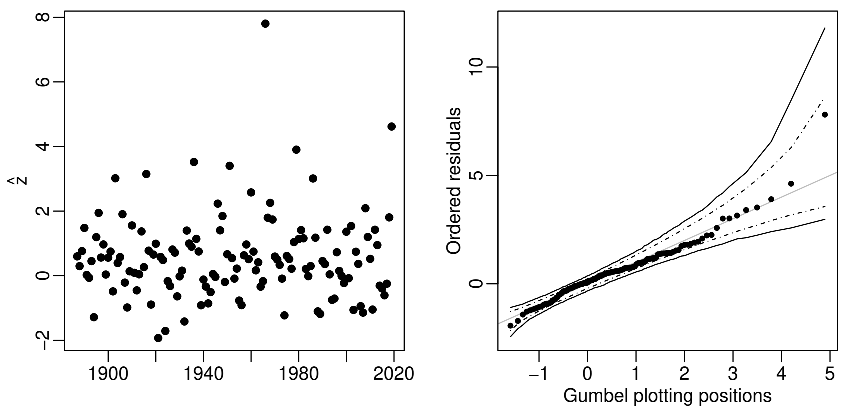
$$w = 2(\widehat{\ell}_1 - \widehat{\ell}_0) = D_0 - D_1.$$

Example: Fitting models

```
y <- venice$y[venice$year<2020,1]
x <- (venice$year[venice$year<2020]-1900)/100
(fit0 <- evd::fgev(y))

Call: evd::fgev(x = y)
Deviance: 1193.487

Estimates
    loc    scale    shape
106.517   20.050   -0.139

Standard Errors
    loc    scale    shape
1.89487  1.29297  0.04412

Optimization Information
  Convergence: successful ...

(fit1 <- evd::fgev(y,nsloc=x))  # nsloc specifies the x variable for the non-stationary locat

Call: evd::fgev(x = y, nsloc = x)
Deviance: 1122.072

Estimates
     loc  loctrend     scale     shape
 89.8087   35.0291   15.0816   -0.1023

Standard Errors
     loc  loctrend     scale     shape
 2.34431   3.51218   0.96584   0.04071

Optimization Information
  Convergence: successful ...
```

Example: Venice sea levels, 1887–2019

Model-checking for fit to Venice maximum sea-level data. Left panel: Gumbel-scale residuals, $\widehat{z}_t$. Right: ordered $\widehat{z}_t$ plotted against Gumbel plotting positions, with pointwise (dot-dash) and overall (solid) 95% confidence bands obtained by simulating 10,000 Gumbel samples.



r-largest analysis

- The limiting joint density of the largest r order statistics $Y_1 > \cdots > Y_r$ in a large sample is

$$\exp\{-\Lambda(y_r)\} \times \prod_{j=1}^{r} \left\{-\dot{\Lambda}(y_j)\right\}, \quad y_1 > \cdots > y_r,$$

  where $\Lambda(y;\eta,\tau,\xi) = \{1+\xi(y-\eta)/\tau\}_+^{-1/\xi}$ and $\dot{\Lambda}(y) = \mathrm{d}\Lambda(y)/\mathrm{d}y$.
- In the Venice data there are r_t (usually 10) largest values in each year, say

$$y_{t,1} > \cdots > y_{t,r_t}, \quad t = 1,\ldots,n,$$

  so if the data for different years are independent, and if we again use parameters $\eta_t(\beta)$, $\tau_t(\beta)$, $\xi_t(\beta)$ in year t, the likelihood is

$$L(\beta) = \prod_{t=1}^{n} \exp\left[-\Lambda\{y_{t,r_t};\eta_t(\beta),\tau_t(\beta),\xi_t(\beta)\}\right] \times \prod_{j=1}^{r_t} \left[-\dot{\Lambda}\{y_{t,j};\eta_t(\beta),\tau_t(\beta),\xi_t(\beta)\}\right].$$

- To fit the model we just maximise the corresponding log likelihood, compute the observed information matrix, and proceed as before ...
- Of course this assumes that this is a reasonable model, which might be questioned: are the background data really independent?

r-largest: model-checking

- ☐ The transformed values $\Lambda(Y_j)$ form a Poisson process of unit rate on the positive half-line:

$$0 < \Lambda(Y_1) < \Lambda(Y_2) < \Lambda(Y_3) < \cdots ,$$

so

$$\Lambda(Y_1), \quad \Lambda(Y_2) - \Lambda(Y_1) \quad \Lambda(Y_3) - \Lambda(Y_2), \ldots \stackrel{\text{iid}}{\sim} \exp(1).$$

- ☐ Recall that if $E \sim \exp(1)$, then $-\log E$ has a standard Gumbel distribution.
- ☐ Hence if the model is adequate and we replace the parameters by their estimates, the

$$-\log\left\{\widehat{\Lambda}(Y_j) - \widehat{\Lambda}(Y_{j-1})\right\}, \quad j = 2, \ldots,$$

should be approximately independent Gumbel variables.
- ☐ The theory above is OK in principle, but if the observations are heavily rounded, the values of $\widehat{\Lambda}(Y_j)$ might be very similar, so that the approach above fails. Alternatively we might note that

$$\Lambda(Y_j) = \Lambda(Y_1) + \sum_{i=1}^{j-1}\{\Lambda(Y_{i+1}) - \Lambda(Y_i)\} \sim \text{Gamma}(j, 1), \quad j = 1, \ldots, r,$$

and then compare the ordered values of the $\Lambda(Y_j)$ for different years with quantiles of the gamma distribution.

Venice data

- ☐ We shall use the r largest observations to fit several models:

$$\begin{aligned}
\mathcal{M}_0: &\quad \eta_t = \beta_0, \quad \tau_t \equiv \tau, \quad \xi_t \equiv \xi,\\
\mathcal{M}_1: &\quad \eta_t = \beta_0 + \beta_1 x_t, \quad \tau_t \equiv \tau, \quad \xi_t \equiv \xi,\\
\mathcal{M}_2: &\quad \eta_t = \beta_0 + \beta_1 x_t + \beta_2 I(t > 1981), \quad \tau_t \equiv \tau, \quad \xi_t \equiv \xi,\\
\mathcal{M}_3: &\quad \eta_t = \beta_0 + \beta_1 x_t + \beta_3 \cos(2\pi t/18.6) + \beta_4 \sin(2\pi t/18.6), \quad \tau_t \equiv \tau, \quad \xi_t \equiv \xi.
\end{aligned}$$

- ☐ Reasoning:
  - we expect the baseline IID model $\mathcal{M}_0$ to be terrible (there is an obvious trend);
  - we expect $\mathcal{M}_1$ to be much better than $\mathcal{M}_1$, as it allows for the trend;
  - if $\mathcal{M}_2$ improves significantly on $\mathcal{M}_1$ then the data sources pre- and post-1982 disagree;
  - $\mathcal{M}_3$ is suggested by the discussion in Pirazzoli (1982, *Acqua Aria*) who suggested that an 18.6-year astronomical cycle may influence the maxima.
- ☐ We could add trend in the scale and shape parameters, but will avoid this here.
- ☐ We take $r = 2$ for now.

Example: Venice sea levels

Code to fit the models using the function `rlarg.fit` of the `ismev` package:

```
y <- venice$y[venice$year<2020,]
year <- venice$year[venice$year<2020]
X <- cbind(x,(year>=1982),cos(2*pi*year/18.6),sin(2*pi*year/18.6) )

> head(y)  # 10 largest values for each year, starting in 1887
   y1  y2 y3 y4 y5 y6 y7 y8 y9 y10
1  94  93 90 86 85 82 81 80 79  76
2  90  84 84 78 75 75 72 72 69  69
3  97  75 74 72 72 68 68 68 67  67
4 107 104 85 81 79 72 72 70 70  67
5  87  72 70 67 66 66 65 64 63  62
6  86  77 74 70 70 69 69 68 68  66

> head(X)  # matrix of covariates for different fits
         x
[1,] -0.13 0 -0.9541393  0.29936312
[2,] -0.12 0 -0.9994295 -0.03377414
[3,] -0.11 0 -0.9317526 -0.36309386
[4,] -0.10 0 -0.7587581 -0.65137248
[5,] -0.09 0 -0.5000000 -0.86602540
[6,] -0.08 0 -0.1847261 -0.98279005

fit0 <- ismev::rlarg.fit(xdat=y, r=2)

fit1 <- ismev::rlarg.fit(xdat=y, r=2, ydat=X, mul=c(1)) # mul says which columns of X to use

fit2 <- ismev::rlarg.fit(xdat=y, r=2, ydat=X, mul=c(1,2))

fit3 <- ismev::rlarg.fit(xdat=y, r=2, ydat=X, mul=c(1,3,4))
```

Fits of $\mathcal{M}_0$ and $\mathcal{M}_1$

```
> fit0 <- ismev::rlarg.fit(xdat=y, r=2)
$conv
[1] 0
$nllh
[1] 1035.521
$mle
[1] 112.1400606  18.3991733  -0.1485854
$se
[1] 1.48482972 0.80384786 0.03128211

> fit1 <- ismev::rlarg.fit(xdat=y, r=2, ydat=X, mul=c(1))
$model
$model[[1]]
[1] 1
$model[[2]]
NULL
$model[[3]]
NULL

$link
[1] "c(identity, identity, identity)"
$conv
[1] 0
$nllh
[1] 973.297
$mle
[1] 93.946101 31.725728 14.160470 -0.103406
$se
[1] 1.68660910 2.42460867 0.65001011 0.03125164
```

Fits of $\mathcal{M}_2$ and $\mathcal{M}_3$

```
fit2 <- ismev::rlarg.fit(xdat=y, r=2, ydat=X, mul=c(1,2))
$model
$model[[1]]
[1] 1 2
...
$conv
[1] 0
$nllh
[1] 969.336
$mle
[1] 91.7010233 41.0735757 -9.7734237 13.8874345 -0.1025608
$se
[1] 1.84707056 4.05490122 3.40265002 0.64056193 0.03303069

> fit3 <- ismev::rlarg.fit(xdat=y, r=2, ydat=X, mul=c(1,3,4))
$model
$model[[1]]
[1] 1 3 4
...
$conv
[1] 0
$nllh
[1] 973.0675
$mle
[1] 93.8443140 31.9199292 -0.4317245 -0.7943716 14.1372635 -0.1050128
$se
[1] 1.69570726 2.44300955 1.36448615 1.34041204 0.64842570 0.03166116
```

Example: Venice sea levels, 1887–2019

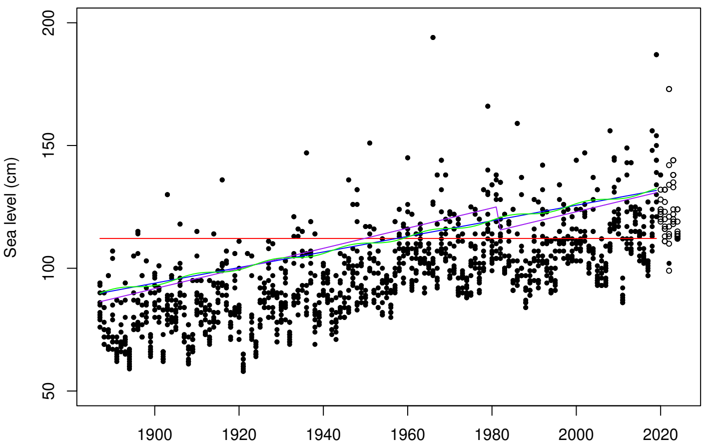
Summaries of fitted models for Venice sea level data analysis, with estimates$_{\rm SEs}$:

Model	r	$-2\widehat{\ell}$	ξ	τ (cm)	β_0 (cm)	β_1 (cm/century)	β_2 (cm)	β_3 (cm)	β_4 (cm)
$\mathcal{M}_1$	1	1122.07	$-0.102_{0.041}$	$15.1_{0.97}$	$89.8_{2.3}$	$35.0_{3.5}$			
$\mathcal{M}_1$	2	1946.59	$-0.103_{0.031}$	$14.2_{0.65}$	$94.0_{1.7}$	$31.7_{2.4}$			
$\mathcal{M}_1$	3	2605.49	$-0.106_{0.025}$	$13.3_{0.52}$	$95.9_{1.4}$	$31.5_{1.9}$			
$\mathcal{M}_1$	4	3185.07	$-0.104_{0.022}$	$12.8_{0.46}$	$96.8_{1.3}$	$31.3_{1.7}$			
$\mathcal{M}_1$	9	5263.20	$-0.090_{0.016}$	$11.5_{0.36}$	$98.2_{1.0}$	$30.3_{1.1}$			
$\mathcal{M}_0$	2	2071.04	$-0.149_{0.031}$	$18.4_{0.80}$	$112.1_{1.5}$				
$\mathcal{M}_1$	2	1946.59	$-0.103_{0.031}$	$14.2_{0.65}$	$94.0_{1.7}$	$31.7_{2.4}$			
$\mathcal{M}_2$	2	1938.67	$-0.103_{0.033}$	$13.9_{0.64}$	$91.7_{1.9}$	$41.1_{4.0}$	$-9.8_{3.4}$		
$\mathcal{M}_3$	2	1946.14	$-0.105_{0.032}$	$14.1_{0.65}$	$93.8_{1.7}$	$31.9_{2.4}$		$-0.43_{1.46}$	$-0.79_{1.34}$
$\mathcal{M}_4$	2	1924.37	$0.057_{0.063}$	$10.6_{0.87}$	$93.2_{2.0}$	$40.0_{4.3}$	$-8.4_{3.6}$		

Note that:

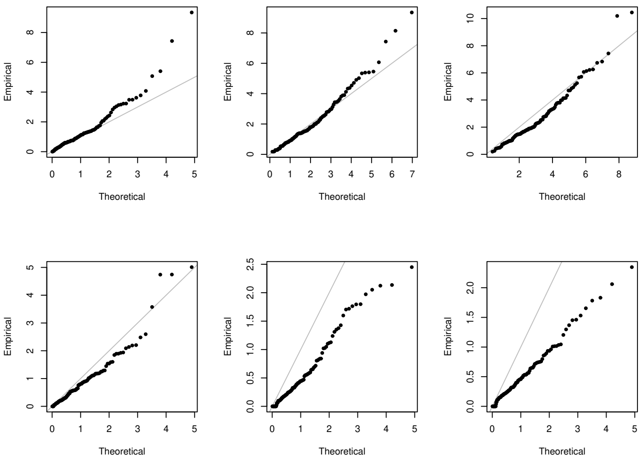
- ☐ the log likelihoods are only comparable for the same values of r, because different values of r use different subsets of the data;
- ☐ if the data were independent, we'd expect the SEs for $r = 1$ to reduce by factors of roughly 2 and 3 for $r = 4$ and $r = 9$;
- ☐ there is strong evidence for trend (surprise!), a change due to the data sources, and $\xi < 0$;
- ☐ there is no evidence of the astronomical cycle.

Example: Venice sea levels, 1887–2019



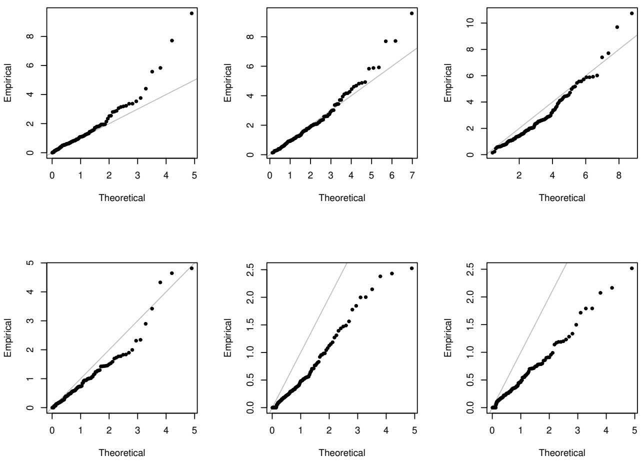
Largest ten annual sea levels at Venice, with fits from models $\mathcal{M}_0$ (red), $\mathcal{M}_1$ (blue), $\mathcal{M}_2$ (purple), and $\mathcal{M}_3$ (green), when $r = 2$.

Example: Venice sea levels, 1887–2019



Residual plots for fit of $\mathcal{M}_1$ with $r = 2$: Top row: comparison of $\widehat{\Lambda}(y_j)$ with corresponding gamma distributions for $j = 1, 2, 3$. Bottom row: comparison of $\widehat{\Lambda}(y_{j+1}) - \widehat{\Lambda}(y_j)$ for $j = 1, 2, 3$. The model does not seem to fit very well!

Example: Venice sea levels, 1887–2019



Residual plots for fit of $\mathcal{M}_2$ with $r = 2$: Top row: comparison of $\widehat{\Lambda}(y_j)$ with corresponding gamma distributions for $j = 1, 2, 3$. Bottom row: comparison of $\widehat{\Lambda}(y_{j+1}) - \widehat{\Lambda}(y_j)$ for $j = 1, 2, 3$. The model does not seem to fit very well!

Haphazard variation

- Above we modelled **systematic variation** by allowing the parameters to depend on known quantities.
- There may be **haphazard variation** that can be modelled by adding extra randomness.
- Suppose that conditional on ε, the data have rate $\varepsilon\Lambda(y)$, where $\varepsilon \sim$ Gamma$(\nu, 1/\nu)$, i.e.,

$$f(\varepsilon) = \frac{\nu^\nu}{\Gamma(\nu)}\varepsilon^{\nu-1}e^{-\nu\varepsilon}, \quad \varepsilon > 0, \quad \nu > 0,$$

  which is the usual gamma density with parameters $\alpha = \nu$, $\lambda = \nu$, so $\mathrm{E}(\varepsilon) = \alpha/\lambda = 1$, $\mathrm{var}(\varepsilon) = \alpha/\lambda^2 = 1/\nu \to 0$ as $\nu \to \infty$. Hence the baseline model corresponds to $\nu = \infty$.
- The marginal density for $Y_1 > \cdots > Y_r$ is then

$$f(y_1, \ldots, y_r) = \int_0^\infty f(y_1, \ldots, y_r \mid \varepsilon) f(\varepsilon)\, \mathrm{d}\varepsilon = \cdots = \prod_{j=1}^r \{-\dot{\Lambda}(y_j)\} \frac{\Gamma(\nu+r)}{\Gamma(\nu)\nu^r} \frac{1}{\{1+\Lambda(y_r)/\nu\}^{\nu+r}},$$

  so in particular the maximum has density

$$f(y_1) = \{-\dot{\Lambda}(y_1)\} \frac{1}{\{1+\Lambda(y_1)/\nu\}^{\nu+1}},$$

  i.e., a model with parameters (η, τ, ξ, ν), where $\nu \to \infty$ gives the basic model.

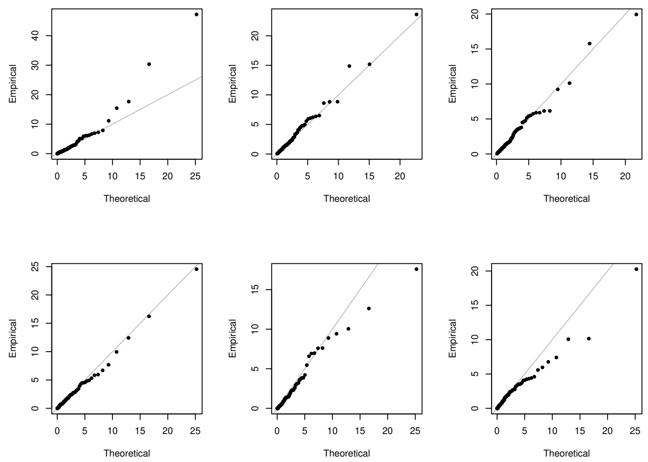
A better model?

- Take $r = 2$ and fit

$$\mathcal{M}_4: \quad \eta_t = \beta_0 + \beta_1 x_t + \beta_2 I(t > 1981), \quad \tau_t \equiv \tau, \quad \xi_t \equiv \xi, \quad \nu,$$

  gives the results shown on slide 128:
  - there is a big reduction in $-2\widehat{\ell}$, so the model is a clear improvement on the others: it is worthwhile to include ν;
  - $\widehat{\nu} = 1.72_{0.63}$, giving strong evidence of overdispersion relative to the baseline model, with ε having standard deviation $\widehat{\nu}^{-1/2} = 0.76$;
  - the estimates of the βs are similar, but $\widehat{\tau}$ is smaller and now $\widehat{\xi} \approx 0$, because including ν accounts for some of the variation not accounted for in the other models;
  - most of the standard errors are larger, because of the additional variation that ν accommodates.
- Residual plots on the next slide are (somewhat) better, though the largest values are still too big relative to the model.

Example: Venice sea levels, 1887–2019



Residual plots for Venice data fit of $\mathcal{M}_4$ with $r = 2$: Top row: comparison of $\widehat{\Lambda}(y_j)$ with corresponding F distributions for $j = 1, 2, 3$. Bottom row: comparison of $\widehat{\Lambda}(y_{j+1}) - \widehat{\Lambda}(y_j)$ for $j = 1, 2, 3$. The model fits better than before, but still the largest values are not well modelled.

Measures of risk

- ☐ Under non-stationarity the quantiles and thus the return levels vary with time, so the interpretation as 'the level exceeded once on average every T years' needs thought . . .
- ☐ Suppose there are m IID background observations in each block (year, say), but that their distributions F_t differ for the different blocks, and let M_T be the maximum for T blocks.
- ☐ If the maximum in year t has GEV distribution G_t, then $F_t \approx G_t^{1/m}$, and we solve

$$1 - p = \left\{ \prod_{t=1}^{T} F_t(x_p) \right\}^{1/T} = \left\{ \prod_{t=1}^{T} G_t^{1/m}(x_p) \right\}^{1/T} = \mathrm{P}(M_T \le x_p)^{1/(mT)}.$$

- ☐ Likewise, in the POT setup, we suppose that independent observations X_j have thresholds u_j, exceedance probabilities p_{u_j} and GP distributions $H_j(x) = 1 - \overline{H}_j(x)$, and then solve

$$1 - p = \left[\prod_{j=1}^{mT} \left\{ 1 - p_{u_j} \overline{H}_j(x_p - u_j) \right\} \right]^{1/mT} = \mathrm{P}(M_T \le x_p)^{1/(mT)}.$$

- ☐ Note that
  - $\mathrm{P}(X > x_p)$ will vary over time, so x_p may not be a very useful summary of risk,
  - both formulae reduce to the previous ones when the data are stationary,
  - there are no explicit formulae for x_p, which must be found numerically.

Comments

- ☐ Similar techniques are applicable for the threshold exceedance and point process models, but the threshold may need to be time-varying and thus needs care.
- ☐ Using the r-largest model may be preferable, as the threshold is replaced with a choice of r.
- ☐ Under the GPD, changing the threshold $u \mapsto u'$ changes the scale parameter:

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

so, for example, the formulation

$$(\sigma_u, \xi) = (g(x_1), h(x_2))$$

at threshold u will become

$$(\sigma_{u'}, \xi) = \left(g(x_1) + h(x_2)(u - u'), h(x_2)\right),$$

at threshold u', so interpretation depends on threshold—undesirable.

- ☐ GEV, r-largest and Poisson process fits use the parameters (η, τ, ξ), invariant to type of model, which is preferable.
- ☐ In some investigations it is preferable to use the GP model: do increasing rainfall maxima come from increases in the number of days with heavy rainfall, but no changes in the amounts, or increased amounts when it rains?

4.3 Dependence slide 137

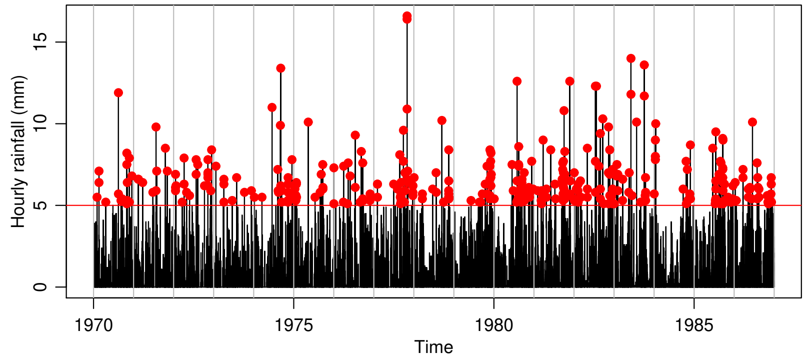
Modelling issues

- ☐ Environmental time series data typically show:
  - long–term trends (e.g., gradual climatic change);
  - seasonality (e.g., annual cycles in meteorology);
  - other forms of non–stationarity (e.g., the effect of ENSO or NAO); and
  - short term dependence (due to volatility, storms, ...).
- ☐ We have discussed non-stationarity. Now we discuss dependence. In brief:
  - the previous limiting theory for maxima also applies, with small changes, provided long-range dependence of extremes is sufficiently weak; but
  - clustering of extremes due to short range dependence arises and must be dealt with.
- ☐ If the background data were independent, then the indicators $I(X_t > u)$ would be IID Bernoulli variables with probability p_u, say, and thus
  - for any $h = 1, 2, \ldots$ we would see

$$\mathrm{P}(X_{t+h} > u \mid X_t > u) = \mathrm{P}(X_{t+h} > u) = p_u;$$

  - intervals between exceedances would be IID geometric variables with mean $1/p_u$ (approximately exponential for small p_u).

Example: Eskdalemuir rainfall

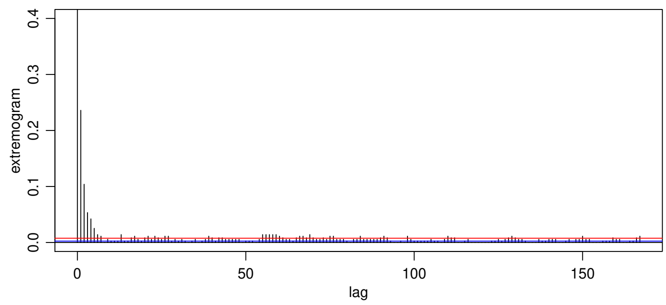


These data show:

- ☐ apparent stationarity (with small seasonal changes?);
- ☐ long-range independence (rain in 1975 is independent of rain in 1980 ...);
- ☐ short-range dependence, owing to clustering of hours with heavy rain?

It seems safe to assume weak dependence of extremes at long ranges, but we need to allow for local dependence.

Extremogram



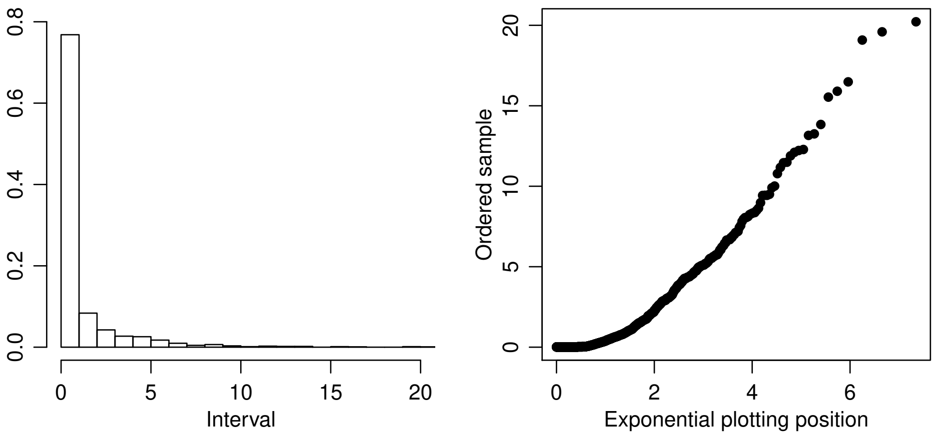
The **extremogram** for a stationary time series $\{X_t\}$ estimates

$$\pi_h(u) = \mathrm{P}(X_{t+h} > u \mid X_t > u), \quad h = 1, 2, \ldots.$$

Independent data would have $\pi_h(u) \equiv \mathrm{P}(X_t > u)$ for all h (blue line in picture, upper 95% point is red line).

- ☐ This is the analogue of the ACF in conventional time series analysis,
- ☐ estimated using frequencies in place of probabilities —
- ☐ beware poor sampling properties of $\widehat{\pi}_h(u)$ (so don't worry about values for high lags).

Intervals between exceedances



The intervals between exceedances should be approximately exponentially distributed, but we see too many small intervals, due to clustering.

Definitions

Definition 19 *A time series* $\{X_j\}$ *is said to be* **(strictly) stationary** *if, for any finite subset* $\mathcal{A}$ *of* $\mathbb{Z}$, *the sets of variables* $X_{\mathcal{A}}$ *and* $X_{h+\mathcal{A}}$ *have the same distribution for all* $h \in \mathbb{Z}$. *In particular this means that the marginal distribution of* X_j *is invariant to location shifts, i.e.,*

$$\mathrm{P}(X_j \le x) = F(x), \quad j \in \mathbb{Z}, x \in \mathbb{R}.$$

Definition 20 *The* **matching series** *for a stationary time series* $\{X_j\}$ *with* $X_j \sim F$ *is the independent series* $\{X_j^*\}$ *for which* $X_j^* \stackrel{\text{iid}}{\sim} F$.

Definition 21 *If* F *is a continuous CDF then* $\{u_m\}$ *is a* **threshold sequence (for** F**)** *if there exists* $\Lambda \in (0,\infty)$ *such that* $\lim_{m\to\infty} m\{1-F(u_m)\} = \Lambda$.

- ☐ If $M = \max(X_1, \ldots, X_m)$ where $X_j \stackrel{\text{iid}}{\sim} F$, and if the extremal types theorem (ETT) applies for sequences $\{a_m\} > 0$ and $\{b_m\}$, then taking $u_m = b_m + a_m x$ gives

$$\Lambda_m(x) = m\{1-F(u_m)\} = m\{1-F(b_m + a_m x)\} \to \left(1+\xi\frac{x-\eta}{\tau}\right)_+^{-1/\xi} = \Lambda(x),$$

  say, so $\{u_m\}$ is then a threshold sequence if $\Lambda(x) > 0$.
- ☐ If there is no Λ for which a threshold sequence exists, then the ETT does not apply.

$D(u_n)$

The usual condition used to impose near-independence of distant extremes is $D(u_n)$:

Definition 22 *Let $\mathcal{A}, \mathcal{B}$ be subsets of $\{1, \ldots, n\}$ such that $\max \mathcal{A} < \min \mathcal{B} - l$ for some positive integer l, and let $M_{\mathcal{A}} \le u$ denote the event $\max_{i \in \mathcal{A}} X_i \le u$, etc. Then $D(u_n)$ is satisfied if*

$$|\mathrm{P}(M_{\mathcal{A}} \le u_n, M_{\mathcal{B}} \le u_n) - \mathrm{P}(M_{\mathcal{A}} \le u_n)\mathrm{P}(M_{\mathcal{B}} \le u_n)| \le \alpha(n, l),$$

where $\alpha(n, l_n) \to 0$ for some sequence $l_n = o(n)$ as $n \to \infty$.

Under $D(u_n)$, maxima of subsets that are sufficiently separated are almost independent, where 'sufficiently separated' means that as $n \to \infty$, the gap l_n between $\mathcal{A}$ and $\mathcal{B}$ satisfies $l_n/n \to 0$.

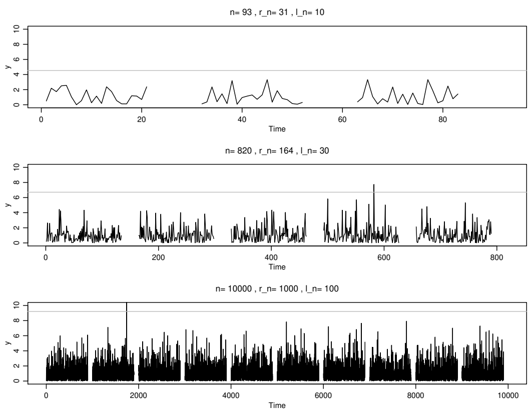
Theorem 23 *Let $X_1, \ldots, X_n$ be a sequence from a stationary series with marginal distribution F that satisfies $D(u_n)$ for a threshold sequence $u_n = b_n + a_n x$. Then if*

$$\mathrm{P}\{\max(X_1, \ldots, X_n) \le u_n\} \to G(x), \quad n \to \infty,$$

where G is non-degenerate, G is a GEV distribution.

Idea of Theorem 23

- ☐ We split $X_1, \ldots, X_n$ into k_n blocks of lengths r_n, where $k_n, r_n \to \infty$ as $n \to \infty$;
- ☐ we ensure that the block maxima are at least l_n observations apart, where $l_n \to \infty$, so if $D(u_n)$ applies these maxima become independent for large n;
- ☐ then we apply the ETT to the k_n (nearly independent) block maxima, and show that if these have a limiting distribution, it must be GEV.



Implications of Theorem 23

- ☐ The assumptions of Theorem 23 are weak, so it should hold in many applications.
- ☐ Hence we aim to understand the effect of local dependence by studying the properties of the maximum of a 'short' block $X_1, \ldots, X_{r_n}$ of neighbouring observations, which we compare with the maximum of an independent series $\{X_j^*\}$ with the same marginal distribution as $\{X_j\}$.
- ☐ Let $X_j^* \stackrel{\text{iid}}{\sim} F$, where $X_j \sim F$, i.e., F is the marginal distribution of $\{X_j\}$, and let

$$M_n^* = \max(X_1^*, \ldots, X_n^*), \quad M_n = \max(X_1, \ldots, X_n).$$

- ☐ We first consider an example.

Example 24 (Moving maximum process) *Let* $Z_j \stackrel{\text{iid}}{\sim} F(z) = \exp(-1/z)$ *for* $z > 0$*, and for* $a \geq 0$ *define*

$$X_0 = Z_0, \quad X_j = (a+1)^{-1} \max(aZ_{j-1}, Z_j), \quad j = 1, \ldots, n.$$

Show that

$$\mathrm{P}(M_n/n \leq x) \to \mathrm{P}(M_n^*/n \leq x)^\theta, \quad n \to \infty,$$

where $\theta = \max(1, a)/(a+1)$ *lies in the interval* $[1/2, 1]$.

Note to Example 24

☐ The marginal distribution of X_j is unit Fréchet:

$$\begin{aligned} P(X_j \le x) &= P\{aZ_{j-1} \le (a+1)x, Z_j \le (a+1)x\} \\ &= \exp\left\{-\frac{a}{(a+1)x}\right\}\exp\left\{-\frac{1}{(a+1)x}\right\} = \exp(-1/x), \quad x > 0, \end{aligned}$$

☐ If $X_1^*, X_2^*, \ldots$ are independent unit Fréchet variables and $M_n^* = \max(X_1^*, \ldots, X_n^*)$, then

$$P(M_n^*/n \le x) = [\exp\{-1/(nx)\}]^n = \exp(-1/x),$$

whereas $M_n = \max(X_1, \ldots, X_n)$ satisfies

$$\begin{aligned} P(M_n/n \le x) &= P(X_1 \le nx, \ldots, X_n \le nx) \\ &= P\{aZ_0 \le (a+1)nx, Z_1 \le (a+1)nx, aZ_1 \le (a+1)nx, \ldots, Z_n \le (a+1)nx\} \\ &= P\{aZ_0 \le (a+1)nx\}\left[\prod_{j=1}^{n-1} P\{Z_j \le (a+1)nx/\max(1,a)\}\right] P\{Z_n \le (a+1)nx\} \end{aligned}$$

because the Z_j are independent. This implies that

$$\begin{aligned} P(M_n/n \le x) &= \exp\left\{-\frac{a}{(a+1)nx}\right\}\left[\exp\left\{-\frac{\max(1,a)}{(a+1)nx}\right\}\right]^{n-1}\exp\left\{-\frac{1}{(a+1)nx}\right\} \\ &= \exp\left(-\theta_n/x\right), \end{aligned}$$

where

$$\theta_n = \frac{a+1+(n-1)\max(1,a)}{n(a+1)} \to \theta = \frac{\max(1,a)}{a+1} \in [1/2, 1], \quad n \to \infty.$$

☐ Hence

$$P(M_n/n \le x) \to \exp(-\theta/x) = P(M_n^*/n \le x)^\theta, \quad n \to \infty,$$

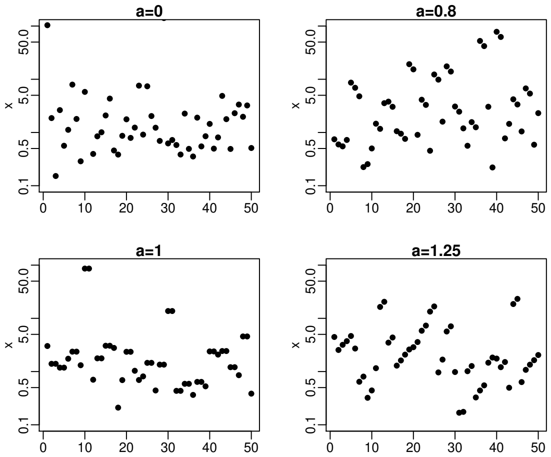
so although $X_j \stackrel{D}{=} X_j^*$,

$$P(M_n^* \le z) \doteq P(M_n \le z)^{1/\theta} \le P(M_n \le z),$$

i.e., M_n is stochastically smaller than M_n^*.

Moving maxima

Realisations of the moving maximum process of Example 24 with $a = 0, 0.8, 1, 1.25$. In each case the marginal distribution is unit Fréchet. The maxima show increasing clustering as $a \to 1$.



Effect of local dependence

Theorem 25 *Let* $\{X_j\}$ *be a stationary process such that* $X_j \sim F$ *and let* $X_j^* \stackrel{\text{iid}}{\sim} F$. *Set* $M_n = \max(X_1, \ldots, X_n)$, $M_n^* = \max(X_1^*, \ldots, X_n^*)$ *and let* $\{a_n\} > 0$ *and* $\{b_n\}$ *be sequences of real numbers. Then there exists a non-degenerate distribution function* G *such that*

$$\mathrm{P}\left\{(M_n - b_n)/a_n \le y\right\} \to G(y) = \exp\{-\Lambda(y)\}, \quad n \to \infty,$$

if and only if

$$\mathrm{P}\left\{(M_n^* - b_n)/a_n \le y\right\} \to G^*(y) = \exp\{-\Lambda^*(y)\}, \quad n \to \infty.$$

If so, $G(y) = \{G^*(y)\}^\theta$ *or equivalently* $\Lambda(y) = \theta\Lambda^*(y)$. *We call* $\theta \in (0, 1]$ *the* ***extremal index***.

☐ As G^* must be GEV(η^*, τ^*, ξ^*), say, G is also GEV, with parameters

$$\xi = \xi^*, \quad \tau = \tau^*\theta^\xi, \quad \eta = \eta^* + \tau^*(\theta^\xi - 1)/\xi \le \eta^* :$$

- the shape parameter is unchanged by the dependence but $\eta < \eta^*$, and
- M_n is stochastically smaller than M_n^*, i.e., dependence tends to reduce the sizes of the extremes for a series of given length, because

$$\lim_{n\to\infty} \mathrm{P}(M_n \le b_n + a_n y) = G(y) = \{G^*(y)\}^\theta \ge G^*(y) = \lim_{n\to\infty} \mathrm{P}(M_n^* \le b_n + a_n y).$$

Implications

- ☐ As $m \to \infty$ for independent data, the rescaled intervals T_m/m between exceedances are independent and Poisson process properties imply that

$$P(T_m/m \le s) \to P(S \le s) = 1 - e^{-\lambda s}, \quad s \ge 0.$$

- ☐ In the corresponding dependent case it can be shown that

$$P(T_m/m \le s) \to P(S \le s) = 1 - \theta e^{-\lambda\theta s}, \quad s \ge 0,$$

  i.e., in the limit,
  - exceedances arise in clusters of mean size $1/\theta \in [1, \infty)$,
  - θ is the probability that a randomly-chosen observation is the last of a cluster;
  - the expected interval $E(S)$ between exceedances is unchanged,
  - but $E(S \mid S > 0) \to 1/(\theta\lambda)$, so the mean interval between clusters increases by $1/\theta$,
  - the maximum of m dependent data has the same limiting distribution as the maximum of $m\theta \le m$ independent data.
- ☐ In fact a cluster maximum has the same limiting distribution as a randomly-chosen exceedance, so there is no bias in fitting the GPD only to cluster maxima, **if we can identify clusters . . .**

Note to Implications

- In the independent case, consider exceedances of a threshold sequence $u_m = b_m + a_m u$. As the X_j are independent and the process is stationary, the interval T_m between two successive exceedances satisfies

$$\begin{aligned} P(T_m > k) &= P(X_1 \le u_m, \ldots, X_k \le u_m \mid X_0 > u_m) \\ &= P(X_1 \le u_m, \ldots, X_k \le u_m) \\ &= F(u_m)^k \\ &= \{1 - \Lambda_m(u)/m\}^k, \quad k \in \mathbb{N}, \end{aligned}$$

where $\Lambda_m(u) = m\{1 - F(u_m)\} \to \Lambda(u) \equiv \lambda$, say, as $m \to \infty$. For any $s > 0$, $\lfloor ms \rfloor / m \to s$ as $m \to \infty$, so

$$\begin{aligned} P(T_m/m > s) &= P(T_m > ms) \\ &= P(T_m > \lfloor ms \rfloor) \\ &= \{1 - \Lambda_m(u)/m\}^{\lfloor ms \rfloor} \\ &\to \exp(-\lambda s), \quad s > 0. \end{aligned}$$

Hence $T_m/m \xrightarrow{D} S \sim \exp(\lambda)$.

- In the dependent case, we argue heuristically as follows. Let $\mathcal{C}$ denote the event that the exceedance at $j = 0$ is the last exceedance in a cluster. Then

$$\begin{aligned} P(T_m > k) &= P(X_1 \le u_m, \ldots, X_k \le u_m \mid X_0 > u_m) \\ &= P(X_1 \le u_m, \ldots, X_k \le u_m \mid \mathcal{C}, X_0 > u_m) P(\mathcal{C} \mid X_0 > u_m) \\ &\quad + P(X_1 \le u_m, \ldots, X_k \le u_m \mid \mathcal{C}^c, X_0 > u_m) P(\mathcal{C}^c \mid X_0 > u_m). \end{aligned}$$

As the data are dependent, $P(X_1 \le u_m, \ldots, X_k \le u_m \mid \mathcal{C}, X_0 > u_m)$ is approximately the probability that $\max(X_1, \ldots, X_k) \le u_m$, conditional on $\mathcal{C} \cap \{X_0 > u_m\}$, and for large k we therefore have

$$P(X_1 \le u_m, \ldots, X_k \le u_m \mid \mathcal{C}, X_0 > u_m) \doteq F(u_m)^{k\theta} = \{1 - \Lambda_m(u)/m\}^{k\theta},$$

whereas for large k,

$$P(X_1 \le u_m, \ldots, X_k \le u_m \mid \mathcal{C}^c, X_0 > u_m) \approx 0,$$

because an observation that is not the last of cluster is highly likely to be followed by another exceedance. Thus if we let $a = \lim_{m\to\infty} P(\mathcal{C} \mid X_0 > u_m)$, we have

$$P(T_m/m > s) = P(T_m > \lfloor ms \rfloor) \doteq \{1 - \Lambda_m(u)/m\}^{\lfloor ms \rfloor \theta} P(\mathcal{C} \mid X_0 > u_m) \to a \exp(-\theta \lambda s), \quad s > 0.$$

This distribution puts a mass of $1 - a$ at $s = 0$ and therefore has mean

$$(1 - a)0 + a/(\theta\lambda) = a/(\theta\lambda).$$

- But as the expected number of exceedances is $1/\lambda$, it must be the case that $a = \theta$, which gives the stated distribution.
- A ‘corrected’ argument is much messier, but is essentially the same as that above.

Statistical consequences of clustering

- Clustering affects the return levels and their interpretation:
  - if $\theta = 1$, then annual maxima are independent but the 'T-year-event' has probability
$$(1 - 1/T)^T \doteq e^{-1} \doteq 0.368$$
    of not appearing in any period of T years;
  - if $\theta < 1$, then the T-year event has probability
$$(1 - 1/T)^{T\theta} \doteq e^{-\theta}$$
    of not appearing in a period of T years, giving (for example) $e^{-0.1} \doteq 0.905$. The same number of events will occur, on average, but they will occur together when $\theta < 1$.
- Various estimators of θ exist. A simple procedure is
  - identify clusters, e.g., by declaring that clusters are separated by runs of more than r non-exceedances of u,
  - let $\widehat{\theta}_u = n_c/n_u$, i.e., the number of clusters divided by the number of exceedances of u.

POT fit to the Eskdalemuir data

```
> fpot(esk.rain, threshold=5)

Call: fpot(x = esk.rain, threshold = 5)
Deviance: 1058.954

Threshold: 5
Number Above: 356
Proportion Above: 0.0024

Estimates
  scale    shape
1.52239  0.06702

Standard Errors
  scale    shape
0.11488  0.05383

Optimization Information
  Convergence: successful
  Function Evaluations: 18
  Gradient Evaluations: 6
```

The fit above does not allow for any clustering.

POT fit to the Eskdalemuir data, allowing for clustering

```
> fpot(esk.rain, threshold=5, r=1, cmax=TRUE)  # fit only to cluster maxima

Call: fpot(x = esk.rain, threshold = 5, cmax = TRUE, r = 1)
Deviance: 835.234

Threshold: 5
Number Above: 356
Proportion Above: 0.0024

Clustering Interval: 1
Number of Clusters: 272
Extremal Index: 0.764

Estimates
  scale    shape
1.63808  0.04183

Standard Errors
  scale    shape
0.14343  0.06322

Optimization Information
  Convergence: successful
  Function Evaluations: 18
  Gradient Evaluations: 5
```

The fit above uses a simple (simplistic) approach to identifying clusters, which end when there are r values below the threshold. Note that $\widehat{\theta} = n_c/n_u = 272/356$, and that as the estimation of the GP parameters is based on the n_c cluster maxima, the standard errors are appreciably larger than in the other fit.

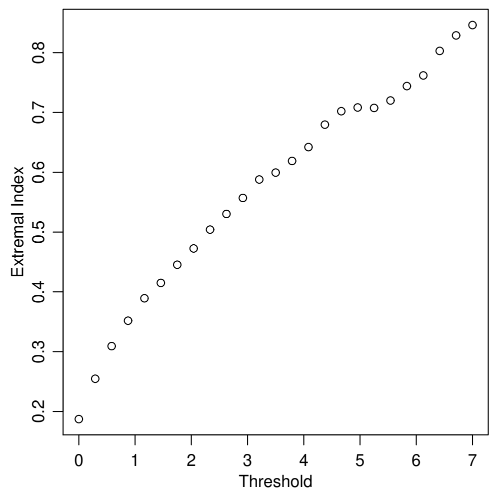
POT fit to the Eskdalemuir data, allowing for clustering

```
> fpot(esk.rain,threshold=5, r=4, cmax=TRUE)

Call: fpot(x = esk.rain, threshold = 5, cmax = TRUE, r = 4)
Deviance: 777.1452

Threshold: 5
Number Above: 356
Proportion Above: 0.0024

Clustering Interval: 4
Number of Clusters: 243
Extremal Index: 0.6826

Estimates
  scale    shape
1.79613  0.01343

Standard Errors
  scale    shape
0.16476  0.06557

Optimization Information
  Convergence: successful
  Function Evaluations: 18
  Gradient Evaluations: 4
```

The fit above uses $r = 4$. Note that $\widehat{\theta} = 0.68$ is smaller than when $r = 1$, and that the standard errors for the GP parameters are still larger than before.

Dependence of $\widehat{\theta}$ on u

Unfortunately $\widehat{\theta}$ (here estimated with $r = 1$) depends on u. The lack of a limit might throw doubt on the theory ...



Clustering and return levels

- The consequences for estimation of return levels are that m dependent background observations correspond to $m\theta$ matching observations, so in the previous formulae on slide 102 we replace m by $m\theta$ (for maxima) and probability $p = 1/N_p$ for dependent observations by $p/\theta = 1/(N_p\theta)$ matching observations (for exceedances), solving $1 - F_X(x_p) = p/\theta$ in the threshold case and

$$1 - F_X(x_p) = 1 - G^{1/(m\theta)}(x_p)$$

  when fitting maxima.
- To estimate the return level

$$x_p \doteq u + \frac{\sigma}{\xi}\left[(p_u\theta/p)^{\xi} - 1\right],$$

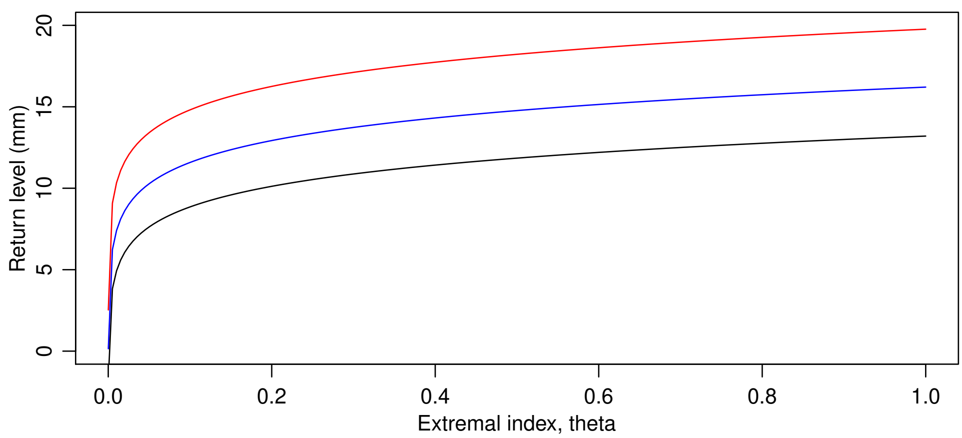
  we estimate σ and ξ by fitting the GPD to threshold exceedances, and use

$$\widehat{p}_u = \frac{n_u}{n}, \quad \widehat{\theta} = \frac{n_c}{n_u},$$

  where n_c is number of clusters and n_u is number of exceedances; thus $\widehat{p}_u\widehat{\theta} = n_c/n$.

Return level estimation

- The figure below shows how θ affects estimates of the 5- (black), 20- (blue) and 100- (red) year return levels for the Eskdalemuir data with threshold $u = 5$ mm.



Ignoring clustering is dangerous . . .

- We fit the GEV to block maxima of a dependent series, and obtain fitted model $\widehat{G}$.
- If we ignore (or are ignorant of) any clustering, then we find the return level by solving

$$1-p = \mathrm{P}(X \le x_p) = F_X(x_p) \doteq \widehat{G}^{1/m}(x_p) \implies x_p = \widehat{G}^{-1}\{(1-p)^m\}.$$

- But we <u>should</u> solve

$$1-p = \mathrm{P}(X \le x_p) = F_X(x_p) \doteq \widehat{G}^{1/(m\theta)}(x_p) \implies x_p = \widehat{G}^{-1}\{(1-p)^{m\theta}\}.$$

  and clearly

$$\widehat{G}^{-1}\{(1-p)^m\} \le \widehat{G}^{-1}\{(1-p)^{m\theta}\},$$

  because $(1-p)^m \le (1-p)^{m\theta}$.
- Hence ignoring the clustering would lead to under-estimation of the return level.

Summary

- Under a weak (and often plausible) condition $D(u_n)$ on the dependence of distant extremes, the GEV is the limiting distribution for the maximum of a stationary dependent process.
- We compare a stationary dependent series $\{X_j\}$ such that $X_j \sim F$ with a **matching series** $\{X_j^*\} \stackrel{\text{iid}}{\sim} F$.
- The effect of local dependence is that extremes arise in clusters whose properties depend on the **extremal index** θ, and
  - the mean cluster size is $1/\theta \ge 1$,
  - the probability that a randomly chosen large event is the last in a cluster is θ,
  - the mean interval between clusters is $1/\theta$ times larger than for the matching series,
  - the GPD marginal distribution of a threshold exceedance is the same as that of a cluster maximum,
  - the maximum of m observations in the dependent series is approximately that of $m\theta$ observations in the matching series,
  - estimates of return levels must be modified to allow for θ.
- Various empirical estimates of θ can be computed.