Risk and Environmental Sustainability

Linda Mhalla

2 Statistics of Extremes slide 47

2.1 Basic Methods for Maxima slide 48

Probability framework for maxima

- Let $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} F$ and define the maximum $M_n = \max(X_1, \ldots, X_n)$, giving

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

- 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 $M_n \stackrel{D}{\longrightarrow} x^*$, where $x^* = \sup\{x : F(x) < 1\}$ is the upper support point of F. Not a useful limit.

- All the argument below applies equally to minima, because

$$\min(X_1, \ldots, X_n) = -\max(-X_1, \ldots, -X_n).$$

Our general discussion is for maxima, and we make this transformation without comment when we model minima.

slide 49

Central limit theorem

- Recall that to get convergence of a sum $\sum_{j=1}^n X_j$ in the central limit theorem, we consider the centred and scaled quantities

$$Z_n = \frac{\sum_{j=1}^n X_j - b_n}{a_n},$$

and with the choices $b_n = n\mu$ and $a_n = n^{1/2}\sigma$, where μ and σ are the mean and standard deviation of X_j, assumed finite, we then obtain

$$Z_n \stackrel{D}{\longrightarrow} Z \sim \mathcal{N}(0,1), \quad n \to \infty.$$

- This suggests studying the convergence of the centered and scaled quantities

$$\frac{M_n - b_n}{a_n},$$

for suitable series $\{b_n\} \subset \mathbb{R}$ and $\{a_n\} \subset \mathbb{R}_+$.

slide 50

Extremal Types Theorem

Theorem 2 (Extremal types) *Let* $M_n = \max(X_1, \ldots, X_n)$ *be the maximum of a random sample* $X_1, \ldots, X_n$*. If sequences of real numbers* $\{a_n\} > 0$ *and* $\{b_n\}$ *can be chosen so that the centred and scaled sample maximum,* $(M_n - b_n)/a_n$*, has a non-degenerate limiting distribution* G*, then this must be the* generalized extreme-value distribution (GEV),

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

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

- ☐ The 'types', which arise for $\xi = 0$, $\xi > 0$ and $\xi < 0$, are now usually subsumed into (1), and are discussed below.
- ☐ This theorem provides a single distribution for maxima, and is in some ways stronger than the CLT, since we only assume that linear rescaling can result in a non-degenerate distribution, without other assumptions on F.

slide 51

Minima

- ☐ In general discussion we consider maxima and large values — what about minima and small values?
- ☐ As

$$Y = \min(X_1, \ldots, X_m) = -\max(-X_1, \ldots, -X_m) = -Y^-,$$

say, then

$$\tilde{G}(y) \approx \mathrm{P}(Y \leq y) = \mathrm{P}(Y^- \geq -y) \approx 1 - G(-y),$$

where G is the GEV approximation for $\max(-X_1, \ldots, -X_m)$. Hence

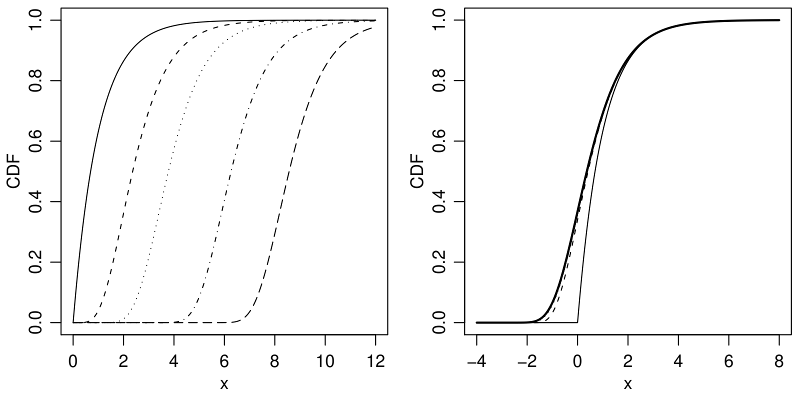
$$\tilde{G}(y; \tilde{\eta}, \tilde{\tau}, \tilde{\xi}) = 1 - G(-y; -\eta, \tau, \xi),$$

where G is estimated from the negative minima.

slide 52

Examples

Example 3 *Find sequences $\{a_m\}$ and $\{b_m\}$ such that maxima of independent variables from the (a) uniform, (b) exponential, and (c) Pareto distributions have non-degenerate limiting distributions.*



Distributions of maxima (left) and renormalized maxima (right) of $m = 1$, 7, 30, 365, 3650 standard exponential variables (from left to right), with limiting Gumbel distribution (heavy).

slide 53

Note I to Example 3

- Note that

$$\mathrm{P}\{(M-b_m)/a_m \le y\} = \mathrm{P}\{M \le b_m + a_m y\} = F^m(b_m + a_m y),$$

and we need to choose a_m and b_m such that this has a limit as $m \to \infty$. We saw from Theorem 2 that a limit $G(y) = \exp\{-\Lambda(y)\}$, so it is equivalent to identify Λ.

- (a) In the uniform case, $F(x) = x$ for $x \in [0,1]$. Provided $0 \le b_m + a_m y \le 1$, we therefore have

$$F(b_m + a_m y)^m = (b_m + a_m y)^m,$$

so if we set $b_m = 1$, $a_m = 1/m$ and $-m \le y \le 0$, we have $(b_m + a_m y)^m \to e^y$. Hence

$$\Lambda(y) = \begin{cases} -y, & y \le 0, \\ 0, & y > 0, \end{cases}$$

i.e., $\Lambda(y) = (-y)_+$, where $a_+ = \max(a, 0)$ for real a. Clearly Λ is decreasing on $(-\infty, 0)$. Hence

$$G(y) = \exp\{-\Lambda(y)\} = \begin{cases} e^y, & y \le 0, \\ 1, & y > 0, \end{cases}$$

which is the distribution function of $-W$, where $W \sim \exp(1)$. It is straightforward to check that this G is (1) with $\eta = 1$, $\tau = 1$ and $\xi = -1$.

- (b) In the exponential case, $F(x) = 1 - \exp(-x)$ for $x > 0$. Provided $b_m + a_m y > 0$,

$$F(b_m + a_m y)^m = \left[1 - \exp\{-(b_m + a_m y)\}\right]^m,$$

so if we set $b_m = \log m$ and $a_m = 1$, and if $y > -\log m$,

$$G(y) = \lim_{m\to\infty} F(b_m + a_m y)^m = \lim_{m\to\infty}\left(1 - \frac{e^{-y}}{m}\right)^m = \exp\left(-e^{-y}\right), \quad y \in \mathbb{R},$$

which is (1) with $\eta = 0$, $\tau = 1$ and $\xi = 0$. Here $\Lambda(y) = e^{-y}$ with support in $\mathbb{R}$.

- (c) In the Pareto case, $F(x) = 1 - x^{-\alpha}$ for $x > 1$ and $\alpha > 0$. Provided $b_m + a_m y > 1$, we have

$$F(b_m + a_m y)^m = \left\{1 - (b_m + a_m y)^{-\alpha}\right\}^m$$

so if we set $b_m = 0$ and $a_m = m^{1/\alpha}$, and if $y > m^{-1/\alpha}$, we have

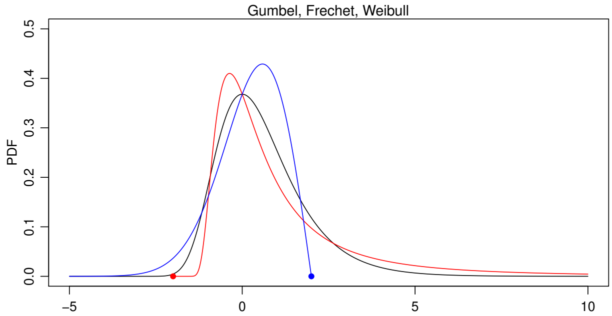
$$G(y) = \lim_{m\to\infty} F(b_m + a_m y)^m = \lim_{m\to\infty}\left(1 - \frac{y^{-\alpha}}{m}\right)^m = \exp\left(-y^{-\alpha}\right), \quad y \ge 0,$$

which is (1) with $\eta = 1$, $\tau = 1/\alpha$ and $\xi = 1/\alpha$. In this case

$$\Lambda(y) = \begin{cases} \infty, & y \le 0, \\ y^{-\alpha}, & y > 0. \end{cases}$$

- Note that we have not shown that the three limits above are the only ones possible, just that we can choose a_m and b_m to obtain these limits.

GEV and 'three types'



- ξ is a shape parameter determining the rate of tail decay, with:
  - $\xi > 0$ giving the heavy-tailed **Fréchet (Type II)** density with support $(\eta - \tau/\xi, \infty)$;
  - $\xi = 0$ giving the light-tailed **Gumbel (Type I)** density, with support $\mathbb{R}$;
  - $\xi < 0$ giving the short-tailed **(reverse) Weibull (Type III)** density, with support $(-\infty, \eta - \tau/\xi)$.
- The usual Weibull distribution gives a model for minima.
- η and τ are location and scale parameters (not so crucial as the shape parameter ξ).

slide 54

Properties of the GEV

- **Support**: If $\xi > 0$ then $Y > \eta - \tau/\xi$, and if $\xi < 0$ then $Y < \eta - \tau/\xi$.
- **Moments**: $\mathrm{E}(Y^r)$ 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.
- **Quantiles**: solve $G(y) = p$ for $0 < p < 1$, but usually we use the **return levels** given by solving $G(y_p) = 1 - p$ (next slide) — so y_p is the $(1-p)$ quantile (careful!)
- **Maximum likelihood estimation**: is regular only if $\xi > -1/2$. Not usually a problem in applications.
- **Max-stability**: if $Y_1, \ldots, Y_T \stackrel{\text{iid}}{\sim}$ GEV(η, τ, ξ) then $\max(Y_1, \ldots, Y_T) \sim$ GEV(η_T, τ_T, ξ_T), i.e.,

$$G(y; \eta, \tau, \xi)^T = G(y; \eta_T, \tau_T, \xi_T)$$

  where

$$\eta_T = \begin{cases} \eta + \tau(T^\xi - 1)/\xi, & \xi \neq 0, \\ \eta + \tau \log T, & \xi = 0, \end{cases} \qquad \tau_T = \tau T^\xi, \quad \xi_T = \xi,$$

  so the distribution type and shape parameter are unchanged by taking maxima.
- In fact the GEV is the only max-stable class of distributions.

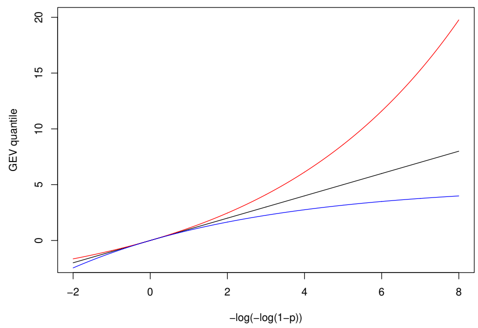
slide 55

Quantiles and return levels

- Define the **return level** associated to the **return period** $T = 1/p$ (blocks) as

$$y_p = \eta + \tau \frac{\{-\log(1-p)\}^{-\xi} - 1}{\xi}, \quad 0 < p < 1,$$

  i.e., the solution to $G(y_p) = 1 - p = 1 - 1/T$.
- Informally, y_p is the level expected to be exceeded once every T blocks.
- The plot below compares the quantiles for $\xi = -0.2$ (blue) and $\xi = 0.2$ (red) with the Gumbel quantiles (black).



slide 56

Statistical approach

- Assume **background data** $x_1, x_2, \ldots$ are IID realisations from some continuous distribution F to which the GEV approximation applies.
- Take maxima $y = \max(x_1, \ldots, x_m)$ of blocks of size m from the background data.
  - for environmental time series, typically $m \approx 365$ for annual maxima, $m \approx 30$ for monthly maxima, ...
  - in finance, typically $m = 250$ for annual maxima, $m = 20$ for monthly maxima, ...
- Suppose the resulting series of maxima $y_1, \ldots, y_n$ are IID GEV(η, τ, ξ).

slide 57

Exploratory plot for maxima

- □ For GEV with $\xi \neq 0$,

$$F^{-1}\{j/(n+1)\} = G^{-1}\{j/(n+1)\} = \eta + \frac{\tau}{\xi}\left[\{-\log(j/(n+1))\}^{-\xi} - 1\right], \quad j = 1, \ldots, n,$$

  which involves all three parameters (which must therefore be estimated!).
- □ By taking $\eta = 0$, $\tau = 1$, and the limit as $\xi \to 0$, we get the **Gumbel plotting positions**

$$-\log\left[-\log\{j/(n+1)\}\right], \quad j = 1, \ldots, n.$$

- □ Plot ordered block maxima $y_{(1)} \leq \cdots \leq y_{(n)}$ against Gumbel plotting positions.
- □ After allowing for noise,
  - convex shape suggests $\xi > 0$,
  - straight line suggests $\xi \approx 0$,
  - concave shape suggests $\xi < 0$.
- □ Outliers, heavy rounding or other issues with data should be visible.
- □ Comparison of these plots for different block sizes may also suggest a minimum block size for the GEV to apply.

slide 58

Estimation

- □ Mostly we use maximum likelihood estimation according to the recipe on slide 24.
- □ This has theoretical and practical advantages:
  - it is efficient (has the smallest possible variance) in large samples — in regular situations;
  - likelihood ratio tests are generally fairly powerful;
  - there's a simple recipe to follow — write down the likelihood and maximise it — which works in many situations;
  - lots of code already exists and can be readily applied. Hooray!
- □ Other methods of estimation are also used:
  - method of moments estimation to get initial values for maximising a likelihood;
  - probability-weighted (or L-) moments estimation is widely used in hydrology and some other domains, because it can beat ML estimation in small samples;
  - in more complex problems the likelihood can be awkward, and then other methods must be used.

slide 59

Moment estimation

- Define moments for random variable X as $\mu'_r = \mathrm{E}(X^r)$ for $r = 1, \ldots$ (if μ'_r finite).
- If X depends on $p \times 1$ parameter vector θ, then $\mu'_r = \mu'_r(\theta)$, and we estimate θ by solving the equations

$$\mu'_r(\theta) = n^{-1} \sum_j X_j^r, \quad r = 1, \ldots, p.$$

- Moment estimators usually simple but inefficient (variance larger than for competing approaches)
- For GEV, μ'_r exists only if $\xi r < 1$, so must have $\xi < 1/3$ to estimate all three parameters, and $\xi < 1/6$ for them to have finite variances. Much too restrictive for use in practice.
- Useful for finding starting-values for ML estimation.

slide 60

L-moment estimation

- Define **probability-weighted moments** as $\mu'_{r,s,t} = \mathrm{E}[X^r F(X)^s \{1 - F(X)\}^t]$ for $r, s, t = 0, 1, 2, \ldots$, or equivalently

$$\mu'_{r,s,t} = \int_0^1 x_p^r p^s (1-p)^t dp, \quad \text{where} \quad F(x_p) = p;$$

  ordinary moments have $s = t = 0$.
- Use $\mu'_{1,s,0}$ for $s = 0, 1, \ldots$ to fit GEV and $\mu'_{1,0,t}$ with $t = 0, 1$ to fit GPD.
- In practice estimate the **L-moments** $\lambda_1, \lambda_2, \ldots$, linear combinations of the μ's, by expressions like

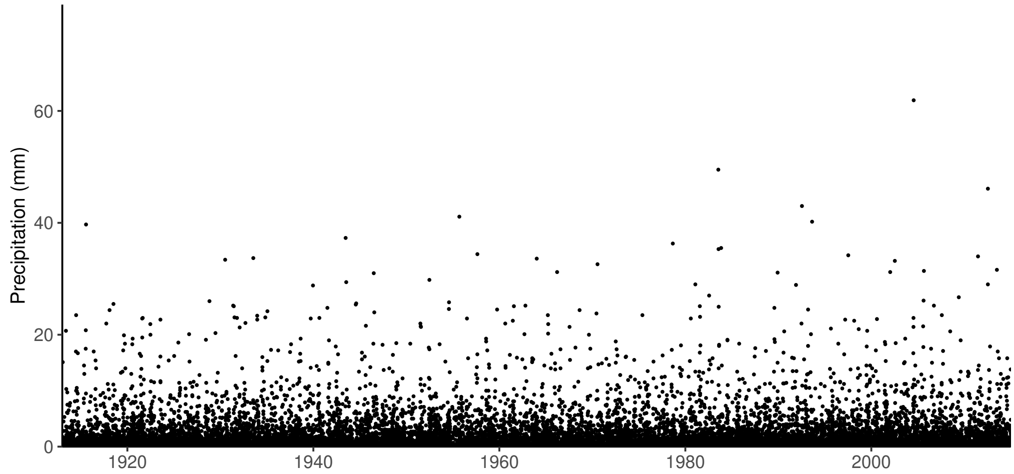
$$\widehat{\lambda}_1 = \frac{1}{\binom{n}{1}} \sum_{j=1}^n X_{(j)}, \quad \widehat{\lambda}_2 = \frac{1}{2\binom{n}{2}} \sum_{j=1}^n \left\{ \binom{j-1}{1} - \binom{n-j}{1} \right\} X_{(j)}, \quad \ldots,$$

- L-moment estimators of η, τ and ξ based on $\widehat{\lambda}_1$, $\widehat{\lambda}_2$ and $\widehat{\lambda}_3$ are linear in the observations, so are more robust than the ordinary moment estimators.
- Have good small-sample properties, but don't generalise to complex settings.

slide 61

Abisko daily rainfall data

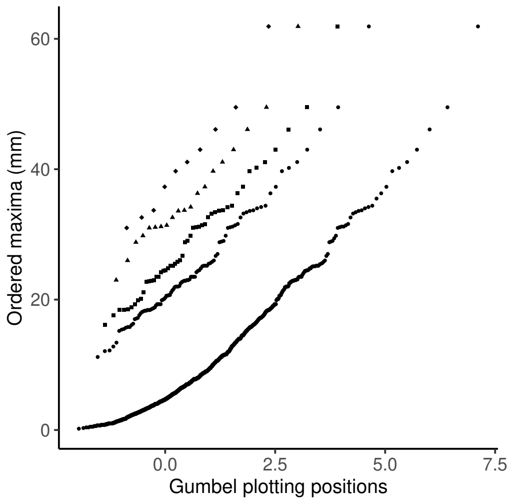
☐ Daily precipitation in Abisko, in northern Sweden, 1913–2014. The largest value is 61.9 mm, but many values are zero and most of the positive values are quite small.



slide 62

Abisko block maxima

☐ Gumbel QQplot of maxima for blocks of lengths (from bottom) one month and one, two, five and ten years.



slide 63

Abisko annual maxima

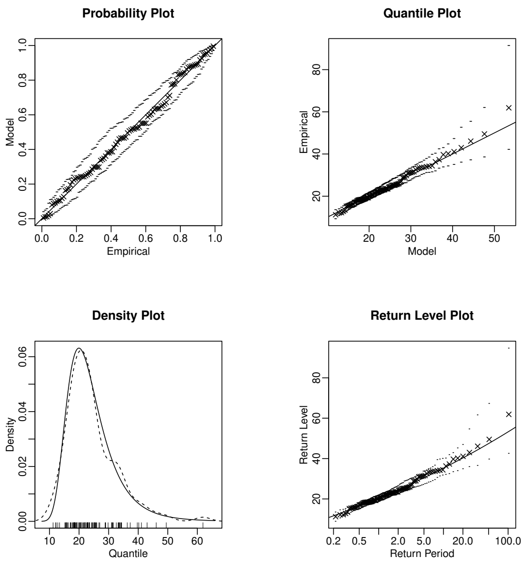
- QQplot suggests stability from one year onwards, with slight convexity, so let's fit the GEV to annual maxima:

```
library(evd)
(fit <- fgev(year.max))

Call: fgev(x = year.max)
Deviance: 691.9509

Estimates
     loc     scale     shape
20.40530   5.84596   0.08353

Standard Errors
    loc    scale    shape
0.64854  0.48317  0.07193

Optimization Information
  Convergence: successful
  Function Evaluations: 27
  Gradient Evaluations: 7
```

slide 64

Abisko annual maxima

- Let's check the fit using `plot(fit)`:



slide 65

Commentary

- ☐ These (horrible!) plots use the fitted GEV CDF $\widehat{G} \equiv G(\cdot; \widehat{\eta}, \widehat{\tau}, \widehat{\xi})$ and are the
  - **probability plot** showing $\left\{(j/(n+1), \widehat{G}(y_{(j)})) : j = 1, \ldots, n\right\}$, which should be a straight line of unit gradient if $\widehat{G}$ is a good fit;
  - **quantile plot** showing $\left\{(\widehat{G}^{-1}\{j/(n+1)\}, y_{(j)}) : j = 1, \ldots, n\right\}$, which should be a straight line of unit gradient if $\widehat{G}$ is a good fit;
  - **return level plot** showing (solid line) $(-1/\log(1-p), \widehat{G}^{-1}(1-p))$, for $0 < p < 1$, and the points $\{(-1/\log\{j/(n+1)\}, y_{(j)}) : j = 1, \ldots, n\}$, which should lie on the line if $\widehat{G}$ is a good fit;
  - **density plot** showing a kernel density estimate based on $y_1, \ldots, y_n$ (shown by the rug) and the fitted GEV density.
- ☐ Some of the plots have pointwise 95% limits for individual points.
- ☐ They show essentially the same information but on different scales to highlight different aspects of the fit.
- ☐ In this case the fit seems reasonable.

slide 66

2.2 Basic Methods for Exceedances

slide 67

Exceedance Theorem

Theorem 4 (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.,*

$$H(x) = \begin{cases} 1 - (1 + \xi x/\sigma)_+^{-1/\xi} & \xi \neq 0, \\ 1 - \exp(-x/\sigma), & \xi = 0, \end{cases} \quad x > 0, \tag{2}$$

where $\xi \in \mathbb{R}$ and $\sigma > 0$. Expression (2) *is commonly known as the* generalized Pareto distribution (GPD).

- ☐ There is a close connection with the ETT, which applies for maxima under the same conditions as the ET applies for exceedances, and with the same ξ.
- ☐ The GPD is a natural model for exceedances over high thresholds (and under low ones, using $1 - H(-x)$).

Example 5 *Find a limiting distribution for threshold exceedances for $Z \sim N(0,1)$. Recall that $1 - \Phi(z) \sim \phi(z)/z$ as $z \to \infty$.*

slide 68

Note to Example 5

- Here $x^* = \infty$ and for large z we have $1 - \Phi(z) \sim \phi(z)/z$.
- By analogy with renormalising maxima we aim to find a function $c_u > 0$ such that

$$\lim_{u\to\infty} \mathrm{P}\{(Z-u)/c_u > x \mid Z > u\}$$

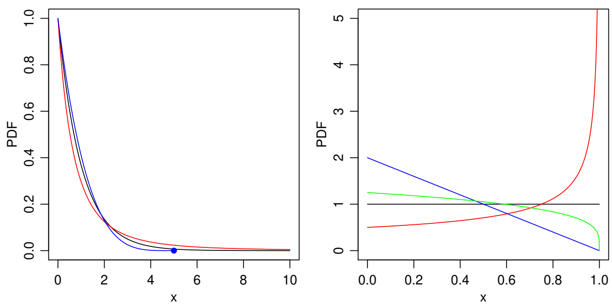
is non-degenerate. The hint gives that for fixed $x > 0$ and large u,

$$\begin{aligned}
\mathrm{P}\{(Z-u)/c_u > x \mid Z > u\} &= \frac{\mathrm{P}(Z > u + c_u x)}{\mathrm{P}(Z > u)} \\
&= \frac{1 - \Phi(u + c_u x)}{1 - \Phi(u)} \\
&\sim \frac{\phi(u + c_u x)/(u + c_u x)}{\phi(u)/u} \\
&= \frac{u}{u + c_u x}\exp\{u^2/2 - (u + c_u x)^2/2\} \\
&= \frac{1}{1 + c_u x/u}\exp(-c_u u x - c_u^2 x^2/2),
\end{aligned}$$

so if we choose $c_u = 1/u$ then the ratio tends to unity and the exponent tends to $-x$, i.e., the limiting distribution for an appropriately rescaled exceedance is standard exponential.
- If we had chosen $c_u = 1/(\sigma u)$ for any fixed $\sigma > 0$ we would have an exponential limit, with mean σ, as in (2), so we can think of the parameter σ as arising because we don't know the ideal scaling function.

note 1 of slide 68

Generalized Pareto distribution



- A flexible distribution whose density can take a variety of shapes.
- Left: exponential density ($\xi = 0$, black), heavy-tailed density ($\xi = 0.5$, red) and light-tailed density ($\xi = -0.2$, blue, with upper terminal shown); all have $\sigma = 1$.
- Right: densities with negative shape parameter and upper terminal at $x = 1$, with $\xi = -1$ (black), $\xi = -2$ (red), $\xi = -0.5$ (blue) and $\xi = -0.8$ (green).

slide 69

Stability and threshold choice

- This approach relying on exceedances is termed the **peaks over threshold (POT)** approach. It is easy to explain and understand but requires the choice of a threshold u.
  - u too low will lead to bias (model inappropriate) and u too high will increase variance (too few exceedances).
- If $X \sim \mathrm{GPD}(\sigma, \xi)$, then $X - u \mid X > u \sim \mathrm{GPD}(\sigma + \xi u, \xi)$, and this implies that

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

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

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

should be approximately straight with slope $\xi/(1-\xi)$ above $u_{\min}$.
- Can also test for equal shape parameters above u (Northrop–Coleman test).

slide 70

Statistical implications of stability

- If $\mathcal{A}_u = [u, \infty)$ and u is sufficiently large that the GP approximation holds, we may write

$$\begin{aligned} \mathrm{P}(X \in \mathcal{A}_v \mid X \in \mathcal{A}_u) &= \mathrm{P}\{X > u + (v - u) \mid X > u\} \\ &\doteq \{1 + \xi(v - u)/\sigma_u\}_+^{-1/\xi}, \quad v > u, \end{aligned}$$

giving the estimate

$$\mathrm{P}(X \in \mathcal{A}_v) \doteq \mathrm{P}(X \in \mathcal{A}_u) \times \{1 + \xi(v - u)/\sigma_u\}_+^{-1/\xi} \doteq \frac{n_u}{n} \times \left\{1 + \widehat{\xi}(v - u)/\widehat{\sigma}_u\right\}_+^{-1/\xi},$$

where n_u is the number of exceedances of u.
- These calculations presuppose that the stability relations apply to exceedances of u, i.e., that the corresponding data are already in the asymptotic regime, which is generally hard to verify due to a lack of data.

slide 71

Relation between GPD and GEV

- We have

$$\mathrm{P}\left(\frac{M_n - b_n}{a_n} \le x\right) = F^n(b_n + a_n x) = \left[1 - \frac{n\{1 - F(b_n + a_n x)\}}{n}\right]^n,$$
$$= \left\{1 - \frac{\Lambda_n(x)}{n}\right\}^n,$$

where

$$\Lambda_n(x) = n\{1 - F(b_n + a_n x)\}, \quad x \in \mathbb{R}.$$

- Since $(1 + d_n/n)^n \to e^d$ for $\{d_n\}$ iff $d_n \to d$, we see that $(M_n - b_n)/a_n$ has a non-degenerate limiting distribution $\exp\{-\Lambda(x)\}$ iff

$$\lim_{n\to\infty} \Lambda_n(x) = \Lambda(x) \text{ exists for all } x \in \mathbb{R},$$

and $\Lambda(x)$ must be decreasing (and must take at least three values).

- If F is continuous, and we choose b_n such that $1 - F(b_n) = 1/n$, then for $x \ge 0$,

$$\Lambda_n(x) = n\{1 - F(b_n + a_n x)\} = \frac{1 - F(b_n + a_n x)}{1 - F(b_n)} = \mathrm{P}(X > b_n + a_n x \mid X > b_n),$$

so $\mathrm{P}(X > b_n + a_n x \mid X > b_n) \to \Lambda(x)$ is equivalent to convergence of $(M_n - b_n)/a_n$ to a non-degenerate limiting random variable.

slide 72

Limit Λ

Lemma 6 *The only possible non-degenerate limit for $\Lambda_n(x)$ as $n \to \infty$ is*

$$\Lambda(x) = \begin{cases} \left(1 + \xi \frac{x-\eta}{\tau}\right)_+^{-1/\xi}, & \xi \neq 0, \\ \exp\left(-\frac{x-\eta}{\tau}\right), & \xi = 0. \end{cases}$$

- The location and scale parameters η and τ are not needed for the limit result, but are needed for statistical applications in which F is unknown.

Later we'll need

$$-\dot{\Lambda}(x) = -\frac{\mathrm{d}\Lambda(x)}{\mathrm{d}x} = \begin{cases} \tau^{-1}\left\{1 + \xi(x-\eta)/\tau\right\}_+^{-1/\xi-1}, & \xi \neq 0, \\ \tau^{-1}\exp\left\{-(x-\eta)/\tau\right\}, & \xi = 0, \end{cases}$$

which is non-negative because $\Lambda(x)$ is decreasing,

slide 73

Proof of Lemma 6

- ☐ If the support of f is a finite union of separated intervals, then as $t \to \infty$ we need only consider the right-most interval. If the support is an infinite union, then $r'(u)$ cannot have a limit.
- ☐ For $t \ge 1$, set $b_t = F^{-1}(1 - 1/t) \in \mathcal{I}$ and $a_t = r(b_t) > 0$. If $x > 0$ satisfies $b_t + a_t x \in \mathcal{I}$, then

$$\begin{aligned} -\log t\{1 - F(b_t + a_t x)\} &= -\log\{1 - F(b_t + a_t x)\} - [-\log\{1 - F(b_t)\}] \qquad (3)\\ &= \mathcal{H}(b_t + a_t x) - \mathcal{H}(b_t) \qquad (4)\\ &= a_t \int_0^x \mathcal{H}'(b_t + a_t u)\,\mathrm{d}u\\ &= a_t \int_0^x \frac{\mathrm{d}u}{r(b_t + a_t u)}. \end{aligned}$$

Taylor's theorem implies that for each u there exists $s \equiv s(u) \in (0, u)$ such that

$$r(b_t + a_t u) = r(b_t) + a_t u r'\{b_t + a_t s(u)\}, \qquad (5)$$

so, as $r(b_t) = a_t$,

$$\frac{r(b_t + a_t u)}{a_t} = \frac{r(b_t) + a_t u r'(b_t + a_t s)}{a_t} = 1 + u r'(b_t + a_t s), \quad 0 < u < x. \qquad (6)$$

If $r'\{b_t + a_t s(u)\}$ did not depend on u, then **(??)** would exactly equal

$$a_t \int_0^x \frac{\mathrm{d}u}{r(b_t + a_t u)} = \int_0^x \frac{\mathrm{d}u}{1 + u r'(b_t + a_t s)} = \xi^{-1}\log(1 + \xi x),$$

with $\xi \equiv \xi(x, t) = r'(b_t + a_t s)$ depending on t and x, and the integral equalling x if $\xi = 0$.

- ☐ To show that such a ξ exists, we define

$$I(s) = \int_0^x \frac{\mathrm{d}u}{1 + u r'(b_t + a_t s)},$$

and note that since r and r' are continuous, the function $s(u)$ defined implicitly by **(??)** is continuous for $u \in (0, x)$. Hence there exist $s_1, s_2 \in [0, x]$ such that

$$r'(b_t + a_t s_1) \le r'\{b_t + a_t s(u)\} \le r'(b_t + a_t s_2), \quad 0 \le u \le x,$$

and thus that

$$\frac{1}{1 + u r'(b_t + a_t s_2)} \le \frac{1}{1 + u r'\{b_t + a_t s(u)\}} \le \frac{1}{1 + u r'(b_t + a_t s_1)}, \quad 0 \le u \le x, \qquad (7)$$

yielding

$$I(s_2) \le \int_0^x \frac{\mathrm{d}u}{1 + u r'\{b_t + a_t s(u)\}} \le I(s_1).$$

Now $I(s)$ is continuous, so the intermediate value theorem implies that there exists some $s^* \equiv s^*(x, t)$ such that $I(s^*)$ equals the integral. Hence

$$\int_0^x \frac{a_t}{r(b_t + a_t u)}\,\mathrm{d}u = \int_0^x \frac{\mathrm{d}u}{1 + u r'(b_t + a_t s^*)} = \frac{1}{\xi(t, x)}\log\{1 + \xi(t, x) x\}, \qquad (8)$$

with $\xi(t, x) = r'\{b_t + a_t s^*(x, t)\}$, as required. The integral equals x if $\xi(t, x) = 0$.

- ☐ The above argument also applies if $x < 0$ is such that $b_t + a_t x \in \mathcal{I}$, with $x < u < 0$, etc.

note 1 of slide 73

Proof of Lemma 6, II

☐ To show that $\xi = \lim_{t\to\infty} \xi(t,x) = \lim_{t\to\infty} r'\{b_t + a_t s^*(x,t)\}$ does not depend on x, we show that as $t \to \infty$ or equivalently $b_t \to x^*$, $b_t + a_t x \to x^*$, which implies that $b_t + a_t s^*(x,t) \to x^*$.

☐ First suppose that $x^* = \infty$, and write

$$b_t + a_t x = b_t(1 + x a_t/b_t),$$

so b_t and $b_t + a_t x$ have the same limit if a_t/b_t is bounded as $t \to \infty$. But for any fixed $u < b_t$ and some u_1 between u and b_t,

$$\frac{a_t}{b_t} = \frac{r(b_t)}{b_t} = \frac{r(u) + (b_t - u)r'(u_1)}{b_t}$$

is bounded because $r'(u_1) \to \xi$ as $t \to \infty$. Hence as $b_t \to \infty$, $b_t + a_t x \to \infty$ also.

☐ When $x^* < \infty$, we must have $F(x) \to 1$ as $x \to x^*$, since otherwise there is positive mass on x^* and the limiting distribution will be degenerate, which is not allowed by hypothesis. Hence

$$-\log\{1 - F(x)\} = \int_x^{x_*} \frac{\mathrm{d}u}{r(u)} \to \infty, \quad x \to x^*,$$

and this implies that $r(x^*) = 0$, since if $r(x^*) > 0$, the assumptions that $r(x)$ is positive and continuous in $\mathcal{I}$ would imply that the integral is finite. Hence l'Hôpital's rule gives

$$\xi = \lim_{b_t\to x^*} r'(b_t) = \lim_{b_t\to x^*} \frac{r(x^*) - r(b_t)}{x^* - b_t} = \lim_{b_t\to x^*} \frac{0 - a_t}{x^* - b_t} = -\lim_{b_t\to x^*} \frac{a_t}{x^* - b_t}. \quad (9)$$

Notice that if $x^* < \infty$, then $\xi \le 0$, since the rightmost expression in (**??**) cannot be positive. Now

$$x^* - b_t - a_t x = (x^* - b_t)\left(1 - \frac{a_t x}{x^* - b_t}\right),$$

where the second term on the right-hand side converges to $1 + \xi x$ as $b_t \to x^*$, and hence is bounded, implying that the entire expression tends to zero.

☐ Hence if x is such that $\lim_{t\to\infty} 1 + a_t x/b_t = 1 + \xi x > 0$, then

$$\lim_{t\to\infty} \xi(t,x) = \lim_{t\to\infty} r'\{b_t + a_t s^*(x,t)\} = \lim_{t\to\infty} r'(b_t) = \xi$$

does not depend on x, and

$$t\{1 - F(b_t + a_t x)\} \to \begin{cases} (1+\xi x)_+^{-1/\xi}, & \xi \ne 0, \\ \exp(-x), & \xi = 0, \end{cases}$$

for both positive and negative x; here the $(\cdot)_+$ gives a formula valid for all $x \in \mathbb{R}$.

☐ To see that the limit is unique up to location and scale, note that the value of ξ depends on the limit for $r'(x)$ as $x \to x^*$. This does not depend on location and scale changes, since replacing x and x^* by $y = \eta + \tau x$ and $y^* = \eta + \tau x^*$ simply puts $r'(x) = r'\{(y-\eta)/\tau\} \to r'\{(y^*-\eta)/\tau\} = r'(x^*)$ (exercise!), and only leads to location and scale changes in Λ, yielding

$$\Lambda(x) = \Lambda\{(y-\eta)/\tau\},$$

which is the general form given.

note 2 of slide 73

Exceedance theorem

- Theorem 4 follows directly from Lemma 6. When $x > 0$, the probability that a rescaled version of X exceeds $x + u$, conditional on it exceeding u, i.e.,

$$\mathrm{P}\left\{(X - b_n)/a_n > x + u \mid (X - b_n)/a_n > u\right\}$$

may be written as

$$\begin{aligned}
\frac{\mathrm{P}\left\{X > b_n + a_n(x+u)\right\}}{\mathrm{P}\left(X > b_n + a_n u\right)} &= \frac{n\left[1 - F\left\{b_n + a_n(x+u)\right\}\right]}{n\left\{1 - F\left(b_n + a_n u\right)\right\}} \\
&= \frac{\Lambda_n(x+u)}{\Lambda_n(u)} \\
&\to \frac{\Lambda(x+u)}{\Lambda(u)}, \quad n \to \infty, \\
&= \frac{\{1 + \xi(x + u - \eta)/\tau\}_+^{-1/\xi}}{\{1 + \xi(u - \eta)/\tau\}_+^{-1/\xi}} \\
&= \begin{cases} (1 + \xi x/\sigma_u)_+^{-1/\xi}, & \xi \neq 0, \\ \exp(-x/\sigma_u), & \xi = 0, \end{cases}
\end{aligned}$$

provided that $\sigma_u = \tau + \xi(u - \eta) > 0$, so that $\Lambda(u) > 0$. Thus the limiting probability that $(X - b_n)/a_n < x + u$, conditional on $(X - b_n)/a_n > u$, is given by the GPD.

Largest order statistics

- Consider the r largest order statistics $Y_r \leq \cdots \leq Y_1$ of the rescaled variables $\{(X_j - b_n)/a_n : j = 1, \ldots, n\}$, and suppose that $n \to \infty$.
- As Y_1 has the limiting distribution of the rescaled maximum $(M_n - b_n)/a_n$,

$$\Pr\left(Y_1 \leq y_1\right) = \exp\left\{-\Lambda\left(y_1\right)\right\}, \quad y_1 \in \mathbb{R}$$

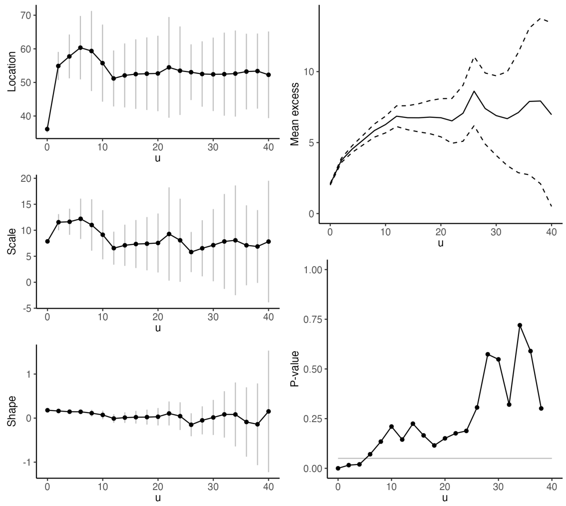
and $f_{Y_1}(y_1) = \left\{-\dot{\Lambda}\left(y_1\right)\right\} \exp\left\{-\Lambda\left(y_1\right)\right\}$.
- The second-largest variable Y_2 is also the largest of an infinite number of these rescaled variables, so its distribution is also G, but conditioned on $Y_2 < Y_1$. Hence

$$\mathrm{P}(Y_2 \leq y_2 \mid Y_1 = y_1) = \exp\left\{\Lambda(y_1) - \Lambda(y_2)\right\}, \quad y_2 < y_1,$$

and it follows that the limiting joint density of the **r-largest order statistics** $Y_r < \cdots < Y_1$ is

$$f(y_1, \ldots, y_r) = \exp\{-\Lambda(y_r)\} \prod_{j=1}^{r} \{-\dot{\Lambda}(y_j)\}, \quad y_r < \cdots < y_1. \qquad (10)$$

Abisko threshold analysis

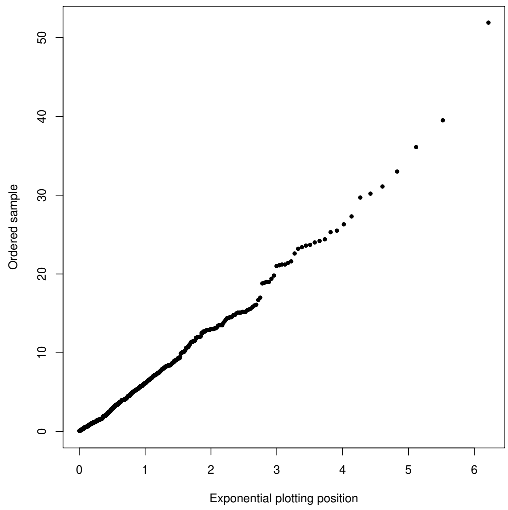


☐ All panels suggest that $u_{\min} = 10$ is reasonable.

slide 76

Exploratory plot

☐ The natural plot here is of ordered exceedances against exponential plotting positions:



slide 77

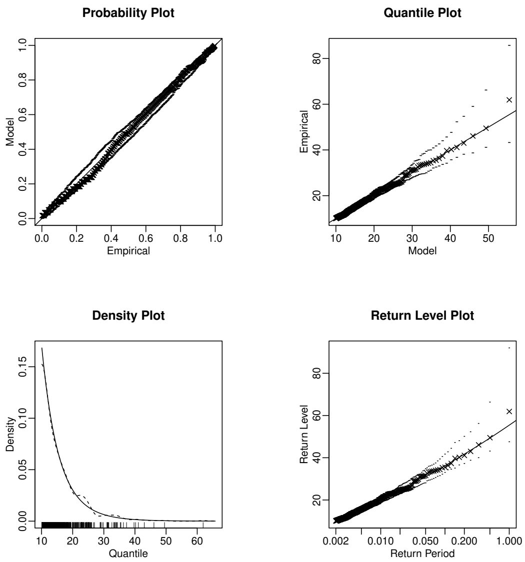
GPD fit

```
(fit.gpd <- fpot(abisko$precip,threshold=10))

Deviance: 2828.05

Threshold: 10
Number Above: 499
Proportion Above: 0.033

Estimates
  scale    shape
5.83261  0.07025

Standard Errors
  scale    shape
0.39483  0.05088

Optimization Information
  Convergence: successful
  Function Evaluations: 16
  Gradient Evaluations: 6
```

slide 78

Abisko POT fit

- ☐ Let's check the fit using `plot(fit.gpd)`:



slide 79

Summary

- ☐ The two fits agree fairly well:
  - Maxima: $\widehat{\eta} = 20.4_{0.649}$, $\widehat{\tau} = 5.84_{0.483}$, $\widehat{\xi} = 0.08_{0.072}$;
  - POT: $\widehat{p}_u = 0.033$, $\widehat{\sigma}_u = 5.83_{0.394}$, $\widehat{\xi} = 0.07_{0.051}$.
- ☐ The location and scale parameters are estimated quite well, but the shape much less well.
- ☐ The shape parameter estimate is slightly positive, but not significantly so (some hydrologists claim that rainfall has $\xi \approx 0.1 \ldots$).
- ☐ The fit appears to be good.
- ☐ In applications one would need to check that the threshold fits are robust to the choice of u (above $u_{\min}$).
- ☐ It is tempting to fit the model with $\xi = 0$, which will give much smaller standard errors for the other parameters. But as we do not <u>know</u> that $\xi = 0$, this reduction in uncertainty may be unrealistic, and it may introduce bias in extrapolation.

slide 80

2.3 Targets of Inference slide 81

Return levels and return periods

- ☐ In basic analyses, typically aim to estimate risk measures such as

$$\mathrm{P}(X > x) = 1 - F_X(x), \quad x_p = F_X^{-1}(1-p),$$

  where $X \sim F_X$ is a background observation and x and x_p are larger than any value yet observed.
- ☐ We often express risk in terms of blocks of m background observations, often daily measurements, with the blocks being years; then $m = 365.25$.
- ☐ We then call x_p a **T-year return level** with a **return period** of $1/p$ observations or T years (i.e., $N_p = Tm$ background observations),
  - e.g., the law states that nuclear installations should withstand the highest windspeed in $T = 10^7$ years(!), so if X is a daily maximum windspeed, then $N_p = 365.25 \times T$ and $p = 1/(365.25T)$.
- ☐ Hence a return level solves the equation

$$\mathrm{P}(X > x_p) = 1 - F_X(x_p) = p = 1/N_p \tag{11}$$

  for some small p.

slide 82

Return levels and return periods II

- ☐ If $x_p > u$, so p is less than the probability $\mathrm{P}(X > u) = p_u$ that a background observation exceeds threshold u, then solving $1 - F_X(x_p) = p$ in the POT model gives

$$x_p = \begin{cases} u + \frac{\sigma_u}{\xi}\left\{(p_u/p)^{\xi} - 1\right\}, & \xi \neq 0, \\ u + \sigma_u \log(p_u/p), & \xi = 0. \end{cases} \tag{12}$$

- ☐ The GEV applies to maxima of blocks of m background observations, so we approximate the upper tail of F_X by $G^{1/m}$, giving

$$1 - p = G^{1/m}(x_p), \tag{13}$$

which yields

$$x_p = \begin{cases} \eta + \frac{\tau}{\xi}\left[\{-m\log(1-p)\}^{-\xi} - 1\right], & \xi \neq 0, \\ \eta - \tau \log\{-m\log(1-p)\}, & \xi = 0. \end{cases} \tag{14}$$

- ☐ In both cases
  - $-\log(1-p) \doteq p = 1/N_p$ for large N_p, giving simpler expressions,
  - point estimates are obtained by replacing the unknown parameters by their estimates,
  - uncertainty is best assessed using the profile log likelihood for x_p.

slide 83

Note on computation of return levels

- ☐ For the POT model in which the GPD is fitted to exceedances of u, and provided $x_p > u$, we have

$$\begin{aligned} \mathrm{P}(X > x_p) &= \mathrm{P}(X > x_p \mid X > u)\mathrm{P}(X > u) \\ &= \mathrm{P}(X - u > x_p - u \mid X > u)\mathrm{P}(X > u) \\ &= \{1 + \xi(x_p - u)/\sigma_u\}_+^{-1/\xi} \times p_u, \end{aligned}$$

and we seek x_p such that

$$1 - p = \mathrm{P}(X \le x_p) = 1 - p_u\{1 + \xi(x_p - u)/\sigma_u\}_+^{-1/\xi},$$

which leads to the stated expression for x_p.

- ☐ If the GEV model is fitted to the maxima of blocks of m background observations then we have

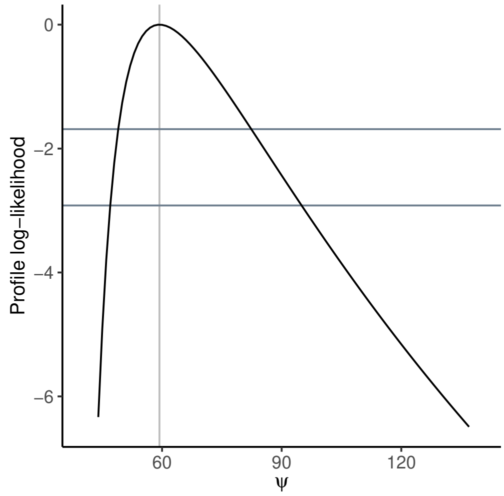
$$1 - p = G^{1/m}(x_p) = \exp\left[-\{1 + \xi(x_p - \eta)/\tau\}_+^{-1/\xi}/m\right],$$

which gives the stated expression for x_p.

note 1 of slide 83

Profile log-likelihood

- ☐ Here $\psi = x_p$ is the 100-year return level for daily precipitation at Abisko based on the GEV fit.
- ☐ The strong asymmetry means that symmetric confidence intervals could be very misleading.



slide 84

Other measures of risk

- ☐ In environmental applications it may be important to estimate amounts of rain falling into an entire catchment area, or the length and impact of a heatwave, or ...
- ☐ The Basel Accords regulate measures of risk to be used by financial institutions:
  - the **Value at Risk** VaR_p is another name for a quantile/return level x_p;
  - the **Expected Shortfall** is defined as the expected loss conditional on VaR_p being exceeded,

$$\mathrm{E}(X - \text{VaR}_p \mid X > \text{VaR}_p),$$

    where in both cases X represents a potential loss.
  - More sophisticated measures such as **expectiles** are also used.

slide 85

Comments

- ☐ The T-year return level is often called 'the level exceeded once on average every T years', and is easily misinterpreted:
  - 'on average' does not mean that disasters arise at regular T-year intervals!
  - selection is often discounted — if M independent time series are monitored, then we expect M/T T-year events each year;
  - the assumption of stationarity is rarely true, so large events may cluster together in periods of elevated risk.
- ☐ Preferable to refer to quantiles — but probably impossible to change a cultural icon!
- ☐ Return levels and return periods are parameters of distributions, but future events are as-yet unobserved random variables, and it may be useful to consider their distributions. The distribution of the largest value X_T to be observed over T blocks of future background observations is $G^T(y)$, and it may be better to use this for risk analysis, in a Bayesian approach (later, probably).