5 Multivariate Extremes slide 158

5.1 Introduction slide 159

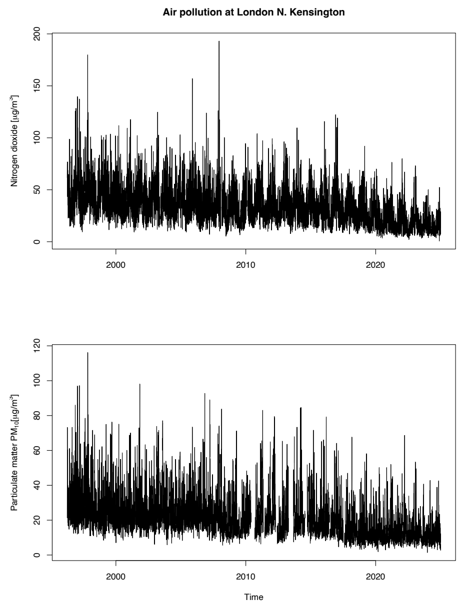
Multivariate extremes

☐ Many extremal problems are essentially multivariate in nature:
  - overwhelming of sea defences by a combination of high tides and high winds;
  - lots of traffic at different servers on a communication network;
  - flooding at many locations of a river system;
  - several successive very hot days (heat waves);
  - (near-)simultaneous downturns in several stock markets.

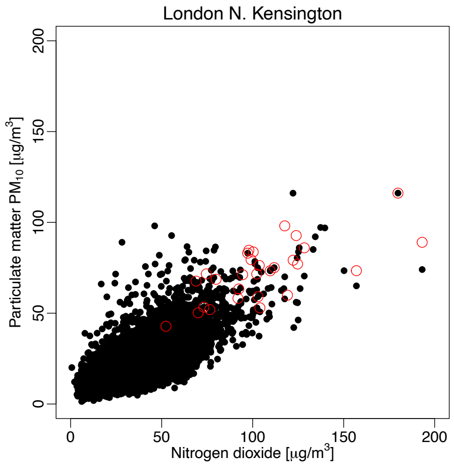
☐ Also, the (often) large variability of extreme value estimates may be reduced by incorporating information via multivariate models.

☐ In one dimension it's obvious what is 'extreme', but in addition to previous questions about suitable asymptotic models, inference, and complications, we must now consider:
  - what is 'extreme' in two or more dimensions?
  - how can we summarize extremal dependence of different variables?

Air pollution in London



Air pollution in London

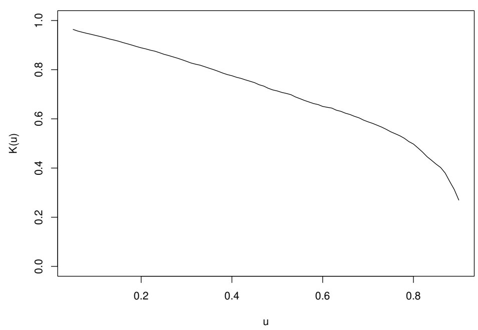


Air pollution in London: Empirical tail dependence

- Each pollutant has its own marginal distribution: F_1 for NO_2 and F_2 for PM_{10}, say.
- To see empirically how their large values behave we first make the margins comparable.
- If X is a continuous random variable with CDF F_X, the random variable $U = F_X(X)$, the **probability integral transform** of X, is uniformly distributed on the unit interval.
- We assess the **joint behaviour** of NO_2 and PM_{10} by estimating

$$K(u) = \mathrm{P}(F_1(NO_2) > u \mid F_2(PM_{10}) > u), \quad 0 < u < 1.$$

- Large values for one variable lead to large values of the other. What happens when $u \to 1$?



Structure variables

- ☐ Multivariate analysis is difficult, so perhaps we could simplify?
- ☐ Could consider a scalar **structure variable** $S = s(X_1, \ldots, X_D) \in \mathbb{R}$, e.g.,
  - insurance loss

$$S = \sum_d a_d(X_d),$$

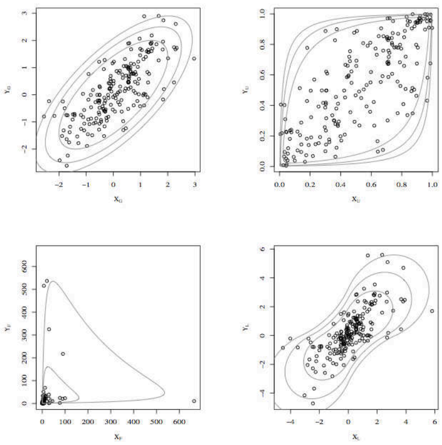
  where increasing (possibly non-linear) functions $a_d(\cdot)$ express damages to properties d due to risks X_d.

- ☐ Then, we have scalar losses $S_1, \ldots, S_n$ to which previous ideas apply, using block maxima or threshold exceedances.
- ☐ **Advantages**: simple analysis, ignores dependence between $X_1, \ldots, X_D$.
- ☐ **Disadvantages**:
  - analysis changes with S, so if new structure variable is introduced, new analysis is needed—which may disagree with original;
  - missing values of X_d not allowed;
  - don't learn which combinations of $X_1, \ldots, X_D$ yield extreme events.
- ☐ So we should study the joint distributions. First we have to look at dependence in general ...

http://stat.epfl.ch slide 164

Choice of margins

- ☐ Different aspects of dependence between multivariate data are visually highlighted by taking different marginal distributions.
- ☐ Bivariate normal data on (clockwise from top left) Gaussian, uniform, Fréchet and Laplace scales, with density contours (grey).



http://stat.epfl.ch slide 165

Standardizing margins

- When studying dependence it helps to remove the effect of marginal transformations.
- Here we consider only continuous random variables, and apply the **probability integral transformation** to obtain variables with uniform margins.
- Suppose that $X \sim F$ is continuous, and takes values everywhere in an interval of $\mathbb{R}$, so the inverse

$$F^{-1}(p) = \inf\{x : F(x) \geq p\}, \quad 0 < p < 1,$$

  satisfies $F\{F^{-1}(p)\} = p$ and $F^{-1}\{F(x)\} = x$.
- Then

$$\mathrm{P}\{F(X) \leq u\} = \mathrm{P}\{X \leq F^{-1}(u)\} = F\{F^{-1}(u)\} = u, \quad 0 < u < 1:$$

  i.e., $F(X) \sim U(0,1)$.
- Equivalently, if $U \sim U(0,1)$, then $X = F^{-1}(U) \sim F$.
- If $X = (X_1, \ldots, X_D) \sim F$ has strictly monotone increasing marginal distributions $F_1, \ldots, F_D$, we therefore have $U_d = F_d(X_d) \sim U(0,1)$ for each $d \in \{1, \ldots, D\}$, corresponding to the top right panel on slide 165.

Copulas

- If $X \sim F$ is continuous with margins F_d then $U_d = F_d(X_d) \sim U(0,1)$ for every $d \in \{1, \ldots, D\}$, and there exists $C : [0,1]^D \to [0,1]$ such that

$$\mathrm{P}(U_1 \leq u_1, \ldots, U_D \leq u_D) = F\{F_1^{-1}(u_1), \ldots, F_D^{-1}(u_D)\} = C(u_1, \ldots, u_D),$$

  for $0 \leq u_1, \ldots, u_D \leq 1$, where

$$C(0, u_2, \ldots, u_D) = 0, \quad C(u, 1, \ldots, 1) = u$$

  for any permutation of the indices.
- Similarly

$$F(x) = C\{F_1(x_1), \ldots, F_D(x_D)\}.$$

- The **copula** C
  - determines the dependence structure of $U = (U_1, \ldots, U_D) = (F_1(X_1), \ldots, F_d(X_d))$,
  - is a cumulative distribution function with uniform margins,
  - is unique (Sklar's theorem) if F is continuous (as we assume), and
  - its must derivatives yield joint density functions and thus constrain C.

Examples

Example 26 (Independence copula) *If* $(U_1,\ldots,U_D)$ *are independent, then*

$$C(u_1,\ldots,u_D)=\prod_{d=1}^{D} u_d, \quad , \quad 0<u_1,\ldots,u_D<1.$$

Example 27 (Co-monotone copula) *If* $(U_1,\ldots,U_D)$ *are totally dependent, then* $U_1=\cdots=U_D$ *with probability one, and*

$$C(u_1,\ldots,u_D)=\min(u_1,\ldots,u_D), \quad , \quad 0<u_1,\ldots,u_D<1.$$

Example 28 (Gaussian copula) *In terms of the* D*-dimensional Gaussian CDF* Φ_D *and the univariate Gaussian CDF* Φ*, the Gaussian copula is*

$$C_\Omega(u_1,\ldots,u_D)=\Phi_D\left\{\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_D);\Omega\right\}, \quad 0<u_1,\ldots,u_D<1,$$

which corresponds to the top-right panel on slide 165.

Non-extremal dependence

- Measures such as the usual (Pearson) correlation coefficient depend on the margins, and we seek to avoid this.
- Let (U_1,U_2) and (V_1,V_2) be independent pairs of variables with copula C.
- A standard measure of dependence is **Kendall's tau**,

$$\tau=\mathrm{corr}\{I(U_1>V_1),I(U_2>V_2)\}=4\mathrm{E}\{C(U_1,U_2)\}-1,$$

  which measures the extent to which the event $(U_1-V_1)(U_2-V_2)>0$ is more probable than the event $(U_1-V_1)(U_2-V_2)<0$.
- τ measures the dependence of the entire distribution, but we wish to focus on the tails.

Example 29 *Compute Kendall's tau for the bivariate independence and co-monotone copulas.*

Note to Example 29

- The bivariate independence copula is $C(u_1,u_2)=u_1u_2$, and

$$\mathrm{E}\{C(U_1,U_2)\}=\mathrm{E}(U_1U_2)=\int_0^1\int_0^1 u_1u_2\,\mathrm{d}u_1\,\mathrm{d}u_2=(1/2)^2=1/4,$$

  so $\tau=0$.
- The bivariate co-monotone copula is $C(u_1,u_2)=\min(u_1,u_2)$ and under this model $U_1=U_2$ with probability one, so we integrate over just one of the variables:

$$\mathrm{E}\{C(U_1,U_2)\}=\mathrm{E}\{\min(U_1,U_2)\}=\int_0^1 u_1\,\mathrm{d}u_1=1/2,$$

  so $\tau=1$.

χ

- The probability that equally rare values of two variables occur simultaneously is a key extremal property. When they have copula C we define the **extremal correlation** to be

$$\chi = \lim_{u \to 1} \mathrm{P}(U_2 > u \mid U_1 > u) = \lim_{u \to 1} \frac{1 - 2u + C(u,u)}{1-u}, \qquad (13)$$

  if it exists, or equivalently on general margins,

$$\chi = \lim_{u \to 1} \mathrm{P}\left\{X_2 > F_2^{-1}(u) \mid X_1 > F_1^{-1}(u)\right\}.$$

- X_1 and X_2 are **asymptotically dependent (AD)** if $\chi > 0$ and **asymptotically independent (AI)** if $\chi = 0$.
- For statistical purposes, as $u \to 1$ we replace $1-u$ and $1-C(u,u)$ in (13) by the approximations $-\log u$ and $-\log C(u,u)$ and obtain

$$\chi(u) = 2 - \frac{\log C(u,u)}{\log u}, \quad 0 < u < 1,$$

  with the focus on $\chi(u)$ for $u \approx 1$.

Example 30 (Logistic copula) *Compute $\chi(u)$ for the logistic copula.*

Note to Example 30

- The bivariate logistic copula

$$C(u_1, u_2) = \exp\left[-\left\{(-\log u_1)^{1/\alpha} + (-\log u_2)^{1/\alpha}\right\}^{\alpha}\right], \quad \alpha \in (0, 1],$$

  evaluated for $u_1 = u_2 = u$ gives

$$C(u,u) = \exp\left[-\left\{2\,(-\log u)^{1/\alpha}\right\}^{\alpha}\right] = \exp\left\{-2^{\alpha}\,(-\log u)\right\} = u^{2^{\alpha}},$$

  which results in the extremal correlation

$$\chi(u) = 2 - \frac{\log(u^{2^{\alpha}})}{\log u} = 2 - \frac{2^{\alpha}\log u}{\log u} = 2 - 2^{\alpha}.$$

  This is constant in u due to the max-stability of the logistic copula (more later).
- We have the following limiting cases:

$$\lim_{\alpha \to 1} \chi(u) = 2 - 2^1 = 0,$$

  that corresponds to independence and

$$\lim_{\alpha \to 0} \chi(u) = 2 - 2^0 = 2 - 1 = 1,$$

  that corresponds to perfect dependence (the co-monotone copula).

$\overline{\chi}$

- ☐ χ can distinguish the strength of dependence for AD distributions, but not the different rates at which $\chi(u) \to 0$ for AI distributions as $u \to 1$.
- ☐ We define

$$\overline{\chi}(u) = 2\frac{\log \mathrm{P}(U_1 > u)}{\log \mathrm{P}(U_2 > u, U_1 > u)} - 1 = 2\frac{\log(1-u)}{\log\{1 - 2u + C(u,u)\}} - 1, \quad 0 < u < 1, \qquad (14)$$

  and use $\overline{\chi} = \lim_{u\to 1} \overline{\chi}(u)$ to measure the degree of AI.
- ☐ The scaling is chosen so that if
  - X and Y are independent, $\overline{C}(u,u) = \mathrm{P}\{F_X(X) > u, F_Y(Y) > u\} = (1-u)^2$, and then $\overline{\chi}(u) \equiv 0$;
  - if X and Y are perfectly dependent, $\overline{C}(u,u) = 1 - u$, and then $\overline{\chi}(u) \equiv 1$;
  - if X and Y are asymptotically dependent, $\overline{\chi}(u) \to 1$ as $u \to 1$;
  - $-1 < \overline{\chi}(u) \leq 1$, and $\overline{\chi}(u)$ increases with increasing dependence.

Example 31 (Bivariate normal) *$\overline{\chi} = \rho$ for the bivariate normal distribution.*

Note to Example 31

☐ As both probabilities tend to zero as $u \to 1$, we first use l'Hôpital's rule to obtain

$$\begin{aligned}\lim_{u\to 1} \mathrm{P}(U_2 > u \mid U_1 > u) &= \lim_{u\to 1} \frac{\mathrm{P}(U_2 > u, U_1 > u)}{\mathrm{P}(U_1 > 0)} \\ &= \lim_{u\to 1} \frac{\mathrm{P}(U_2 > u, U_1 = u) + \mathrm{P}(U_2 = u, U_1 > u)}{\mathrm{P}(U_1 = u)} \\ &= \lim_{u\to 1} 2\mathrm{P}(U_2 > u \mid U_1 = u)\end{aligned}$$

if the distribution is symmetric (as here).

☐ Note that Mill's ratio

$$\mathrm{P}(X_1 > x) = \overline{\Phi}(x) \sim \phi(x)/x, \quad x \to \infty,$$

implies that $\log \mathrm{P}(X_1 > x) \sim -x^2/2 - \log x$ as $x \to \infty$.

☐ In

$$\frac{\log \mathrm{P}(U_1 > u)}{\log \mathrm{P}(U_1 > u) + \log \mathrm{P}(U_2 > u \mid U_1 > u)}, \tag{15}$$

U_1 and U_2 can be replaced by standard normal variables X_1 and X_2, and as the distribution is symmetric, for large x we have

$$\begin{aligned}\mathrm{P}(X_2 > x \mid X_1 > x) \sim 2\mathrm{P}(X_2 > x \mid X_1 = x) &= 2\overline{\Phi}\left\{\frac{x(1-\rho)}{(1-\rho^2)^{1/2}}\right\} \\ &\sim \phi\left\{\frac{x(1-\rho)}{(1-\rho^2)^{1/2}}\right\} \div \frac{x(1-\rho)}{(1-\rho^2)^{1/2}}\end{aligned}$$

Hence

$$\begin{aligned}\frac{\log \mathrm{P}(X_1 > x)}{\log \mathrm{P}(X_1 > x) + \log \mathrm{P}(X_2 > x \mid X_1 > x)} &\sim \frac{-x^2/2 - \log x}{-x^2/2 - \log x - \frac{x^2(1-\rho)^2}{2(1-\rho^2)} - \log\left\{\frac{x(1-\rho)}{(1-\rho^2)^{1/2}}\right\}} \\ &\to \frac{1}{1 + (1-\rho)/(1+\rho)}, \quad x \to \infty, \\ &= (1+\rho)/2.\end{aligned}$$

Hence $\overline{\chi} = \rho$, thus stressing the interpretation of positive, negative and zero values of $\overline{\chi}$.

Comments

- ☐ Copulas allow the dependence between several variables to be studied without reference to their marginal distributions. They are very widely used in finance, insurance and other areas.
- ☐ A key distinction for rare events is between asymptotic dependence (AD) and asymptotic independence (AI), which correspond to $\chi > 0$ and $\chi = 0$ for

$$\chi = \lim_{u \to 1} \chi(u) = \lim_{u \to 1} \mathrm{P}\{F_2(X_2) > u \mid F_1(X_1) > u\}, \quad 0 < u < 1,$$

in practice replaced by the asymptotically (as $u \to 1$) equivalent

$$\chi(u) = 2 - \frac{\log C(u,u)}{\log u}, \quad 0 < u < 1,$$

where C is the copula for (X_1, X_2).

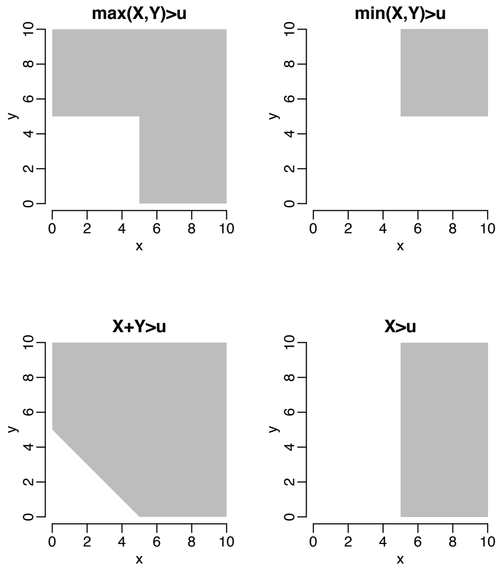
- ☐ A similar quantity $\overline{\chi}(u)$ is used to distinguish different strengths of AI.
- ☐ Plots of $\chi(u)$ and $\overline{\chi}(u)$ are essential graphical tools for looking at joint extremes, and are produced by the R command `evd::chiplot`.

5.3 Multivariate Models for Extremes slide 174

Extremes for $D = 2$

- ☐ Given variables (X, Y) with the same marginal distributions, and a high threshold u, we might consider any of the following scenarios:
  - at least one of X and Y exceeds u, i.e., $\max(X, Y) > u$;
  - both X and Y exceed u, i.e., $\min(X, Y) > u$;
  - a function $s(X, Y)$ exceeds u, e.g., $X + Y > u$, though $s(\cdot)$ could also measure distance from some multivariate centre for the data; or
  - given that $X > u$, we consider the distribution of Y, where Y is called a **concomitant** of X; the extremal set is $X > u$.
- ☐ There are other possibilities, but these already make life complicated enough.
- ☐ The grey regions on the next slide are considered to be extreme under these four scenarios.

Extremes for $D = 2$



http://stat.epfl.ch slide 176

Componentwise maxima

- If $(X_1, Y_1), (X_2, Y_2), \ldots \overset{\text{iid}}{\sim} F(x, y)$, define the **componentwise maxima**,

$$M_{X,n} = \max_{j=1,\ldots,n} \{X_j\}, \qquad M_{Y,n} = \max_{j=1,\ldots,n} \{Y_j\};$$

  note that $M_n = (M_{X,n}, M_{Y,n})$ may not correspond to an actual observation (e.g., NO_2, PM_{10} in pollution example).

- Limiting distributions must exist for maxima of X and Y individually, because otherwise any limiting joint distribution will be degenerate, so we ask

  If non-degenerate limiting distributions exist for maxima of rescaled pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ as $n \to \infty$, what forms can they have?

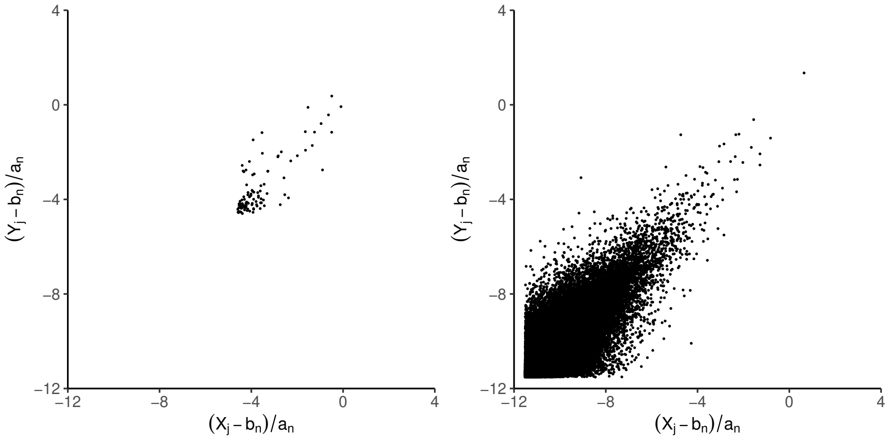
- Considered separately, $\{X_j\}$ and $\{Y_j\}$ are sequences of independent, univariate random variables, to which our previous theory applies: if a limiting distribution exists for each margin, then we can consider the sequence of point patterns

$$\mathcal{P}_n = \{((X_j - b_{X,n})/a_{X,n}, (Y_j - b_{Y,n})/a_{Y,n}) : j = 1, \ldots, n\}, \quad n = 1, 2, \ldots.$$

- As $n \to \infty$ we will have convergence to a Poisson process, $\mathcal{P}_n \xrightarrow{D} \mathcal{P}$, with state space $\mathcal{E} = \mathbb{R}^2$ and measure μ, say, which we can use for inference on extreme values.

http://stat.epfl.ch slide 177

Point patterns for $D = 2$



Rescaled bivariate exponential datasets of sizes $n = 100$ (left) and $n = 10^4$ (right).
In this case $\Lambda(x) = e^{-x}$ on each margin, so the transformation $1/\Lambda(x)$ to unit Fréchet margins for maxima would exponentiate both axes and give many observations close to the origin.

Marginal transformation

☐ On the X margin as $n \to \infty$ we can choose the sequences $\{b_n\}$ and $\{a_n\}$ so that

$$\max\{(X_1 - b_n)/a_n, \ldots, (X_n - b_n)/a_n\} \xrightarrow{D} S \sim \exp\{-\Lambda_X(s)\},$$

where $\Lambda_X(s) = (1 + \xi_X s)_+^{-1/\xi_X}$ is monotone decreasing, so $Z = 1/\Lambda_X(S)$ has the unit Fréchet distribution

$$\begin{aligned} \mathrm{P}(Z \le z) &= \mathrm{P}\{1/\Lambda_X(S) \le z\} \\ &= \mathrm{P}\{\Lambda_X(S) \ge 1/z\} \\ &= \mathrm{P}\{S \le \Lambda_X^{-1}(1/z)\} \\ &= \exp[-\Lambda_X\{\Lambda_X^{-1}(1/z)\}] \\ &= \exp(-1/z), \quad z > 0. \end{aligned}$$

☐ The same argument applies on the Y margin, so we can apply the transformation

$$g(x, y) = (1/\Lambda_X(x), 1/\Lambda_Y(y)), \quad (x, y) \in \mathcal{E} = \mathbb{R}^2$$

to $\mathcal{P}_n$, giving sequences of point patterns

$$\mathcal{P}_n^* = g(\mathcal{P}_n) \subset \mathcal{E}^* \subset \mathbb{R}_+^2, \quad n = 1, \ldots,$$

such that the limiting maxima on each margin have unit Fréchet distributions.

Limiting Poisson process

- ☐ The function g is invertible, so Poisson convergence for $\mathcal{P}_n$ and for $\mathcal{P}_n^* = g(\mathcal{P}_n)$ is equivalent.
- ☐ As $n \to \infty$, the point pattern $\mathcal{P}_n^*$ converges to a Poisson process $\mathcal{P}^*$ on $\mathcal{E}^*$ whose mean measure μ^* has margins

$$\mu^*\{\mathbb{R}_+ \times (z,\infty)\} = \mu^*\{(z,\infty) \times \mathbb{R}_+\} = 1/z, \quad z > 0,$$

so we cannot compute the measure of any set containing the origin. To avoid this we define $\mathcal{P}^*$ on the 'punctured set' $\mathcal{E}^* = [0,\infty)^2 - \{(0,0)\}$.

- ☐ For any $z \equiv (z_1, z_2) \in \mathcal{E}^*$ it is convenient to define

$$\mathcal{A}_z^* = \{(x,y) \in \mathcal{E}^* : x > z_1 \text{ or } y > z_2\},$$

so that the joint maxima satisfy

$$G^*(z) = \mathrm{P}(Z_1 \le z_1, Z_2 \le z_2) = \mathrm{P}\{N^*(\mathcal{A}_z^*) = 0\} = \exp\{-\mu^*(\mathcal{A}_z^*)\},$$

where $N^*(\mathcal{A}^*)$ is the number of points of $\mathcal{P}^*$ in $\mathcal{A}^* \subset \mathcal{E}^*$.

- ☐ In terms of the original process $\mathcal{P}$ and $\mathcal{A}^* = g(\mathcal{A})$ we have

$$\mu(\mathcal{A}) = \mu^*(\mathcal{A}^*) = \mu^*\{g(\mathcal{A})\},$$

which will be useful later.

Exponent function

- ☐ For ease of notation define the **exponent function**

$$V(z) = V(z_1, z_2) = \mu^*(\mathcal{A}_z^*), \quad z \in \mathcal{E}^*.$$

- ☐ In the scalar case we saw that the GEV is max-stable, i.e., for any $t > 0$ there exist $a_t > 0$ and b_t such that

$$G^t(b_t + a_t x) = G(x), \quad x \in \mathbb{R},$$

and when G is unit Fréchet we have $b_t = 0$ and $a_t = t$.

- ☐ In the multivariate case the same argument applies, giving

$$\{G^*(tz)\}^t = G^*(z) \quad \Longrightarrow \quad tV(tz) = V(z), \quad z \in \mathcal{E}^*, t > 0,$$

i.e., the function V is homogeneous of order -1.

- ☐ The marginal unit Fréchet distributions yield

$$V(z', \infty) = V(\infty, z') = 1/z', \quad z' > 0.$$

- ☐ The homogeneity of V suggests a change of variables to **angular coordinates**

$$R = Z_1 + Z_2, \quad W = (Z_1, Z_2)/R \quad \Longleftrightarrow \quad Z = (Z_1, Z_2) = RW,$$

which allow us to state the joint distribution of the maxima.

Limit distribution of componentwise maxima

Theorem 32 *If $X_1, X_2, \ldots,$ are independent copies of a D-dimensional random variable whose componentwise maxima can be linearly renormalised to converge as $n \to \infty$ to a random variable $Z = (Z_1, \ldots, Z_D)$ that has a non-degenerate distribution with unit Fréchet margins, then*

$$\mathrm{P}(Z_1 \leq z_1, \ldots, Z_D \leq z_D) = \exp\left[-D\mathrm{E}\left\{\max_{d=1}^{D}(W_d/z_d)\right\}\right], \quad z_1, \ldots, z_D > 0, \qquad (16)$$

where the **angular variable** *$W = (W_1, \ldots, W_D)$ lies in the $(D-1)$-dimensional simplex, i.e.,*

$$W \in \mathbb{S}_{D-1} = \left\{(w_1, \ldots, w_D) : w_d \geq 0, \sum_{d=1}^{D} w_d = 1\right\}$$

and satisfies the marginal mean constraints

$$\mathrm{E}(W_d) = 1/D, \quad d = 1, \ldots, D.$$

The **angular probability distribution** *ν of W is otherwise arbitrary — it may or may not have an* **angular density function** *$\dot{\nu}$.*

Note: Proof of Theorem 32

- ☐ The proof is not more complicated in D dimensions.
- ☐ The preceding argument implies that after the marginal transformations the D-dimensional point processes $\mathcal{P}_n^* = g(\mathcal{P}_n)$ converge to a Poisson process on $\mathcal{E}^* = [0,\infty)^D - \{(0,\dots,0)\}$ with measure μ^* which is homogeneous of order -1, so we just have to establish the form of the distribution for maxima.
- ☐ Let $(q_1,\dots,q_D)$ be a point in $\mathcal{E}^*$. To see the effect of the change of variables, consider the transformation $T : \mathcal{E}^* \to (0,\infty) \times \mathbb{S}_{D-1}$ to angular variables defined by

$$T(q_1,\dots,q_D) = (r,w), \quad r = q_1 + \cdots + q_D, \quad w_d = q_d/r, \quad d = 1,\dots,D,$$

and with inverse $T^{-1}(r,w) = rw$. As T is invertible, $T(\mathcal{P}^*)$ is also a Poisson process.
- ☐ To compute the measure $\mu^* \circ T^{-1}$ that T induces for the angular variables (R,W), note that as $rw \in \mathcal{A}_z^*$ if and only if at least one of the rw_d exceeds z_d, or equivalently $\max_d rw_d/z_d > 1$,

$$\begin{aligned} T(\mathcal{A}_z^*) &= \{(r,w) : \max(rw_1/z_1,\dots,rw_D/z_D) > 1\} \\ &= \{(r,w) : ra_z(w) > 1\}, \end{aligned} \qquad (17)$$

where $a_z(w) = \max(w/z_1,\dots,w_D/z_D)$. Moreover,

$$\mu^*(\mathcal{A}_z^*) = V(z) = V(rw) = \frac{D}{r} \times D^{-1}V(w) = \frac{D}{r} \times \nu(w), \quad r > 0, w \in \mathbb{S}_{D-1},$$

say, implying that

$$\mu^* \circ T^{-1}\{(\mathrm{d}r,\mathrm{d}w)\} = \frac{D}{r^2}\mathrm{d}r \times \nu(\mathrm{d}w). \qquad (18)$$

- ☐ The appearance of D in (**??**) ensures that ν has unit measure, as we shall see below.
- ☐ Expression (**??**) is a product, so R and W are independent, and (**??**) yields

$$\begin{aligned} V(z) = \mu^*(\mathcal{A}_z^*) = \mu^* \circ T^{-1} \circ T(\mathcal{A}_z^*) &= \mu^* \circ T^{-1}\left[\{(r,w) : r > 1/a_z(w)\}\right] \\ &= D \iint_{\{(r,w):r>1/a_z(w)\}} r^{-2}\mathrm{d}r\,\nu(\mathrm{d}w) \\ &= D \int \Big[-r^{-1}\Big]_{1/a_z(w)}^{\infty} \nu(\mathrm{d}w) \\ &= D \int_{\mathbb{S}_{D-1}} \max_d \left(\frac{w_d}{z_d}\right) \nu(\mathrm{d}w) \\ &= D\mathrm{E}\left\{\max_d (W_d/z_d)\right\}. \end{aligned}$$

- ☐ As the margins of G^* are unit Fréchet, when all but one of the z_d are set to infinity we have

$$D\mathrm{E}\left\{\max_d (W_d/z_d)\right\} = D\mathrm{E}(W_d)/z_d = 1/z_d, \quad d = 1,\dots,D.$$

Hence $\mathrm{E}(W_d) = 1/D$, concluding the proof.
- ☐ To check that ν is a probability measure, note that

$$\int_{\mathbb{S}_{D-1}} \nu(\mathrm{d}w) = \int_{\mathbb{S}_{D-1}} (w_1 + \cdots + w_D)\,\nu(\mathrm{d}w) = \sum_{d=1}^{D} \mathrm{E}(W_d) = D \times 1/D = 1.$$

Bivariate maxima

- If $D = 2$, then $W_1 = 1 - W_2 = W$, say,

$$V(z_1, z_2) = 2\mathrm{E}\left\{\max\left(\frac{W}{z_1}, \frac{1-W}{z_2}\right)\right\} = 2\int_0^1 \max\left(\frac{w}{z_1}, \frac{1-w}{z_2}\right) \nu(\mathrm{d}w),$$

and $W \sim \nu$, an **angular (or spectral) distribution function** on $[0, 1]$, such that

$$\mathrm{E}(W) = \int_0^1 w\,\nu(\mathrm{d}w) = 1/2.$$

- If ν has an angular density function $\dot{\nu}$, then

$$V(z_1, z_2) = 2\int_0^1 \max\left(\frac{w}{z_1}, \frac{1-w}{z_2}\right) \dot{\nu}(w)\,\mathrm{d}w.$$

Example 33 *Find the limiting distributions for maxima when (a) $W \in \{0, 1\}$ with equal probabilities, (b) $W = 1/2$ with probability one, (c) $W \sim U(0, 1)$.*

These cases are not useful statistically, but they illustrate why a general treatment must allow ν to have a mixture of point masses and density.

Note to Example 33

- Let $D = 2$ and let $\mathrm{P}(W = 0) = \mathrm{P}(W = 1) = 1/2$. Then

$$V(z_1, z_2) = 2\mathrm{E}\left[\max\{W/z_1, (1-W)/z_2\}\right] = 1/z_1 + 1/z_2,$$

yielding $G^*(z_1, z_2) = \exp(-1/z_1)\exp(-1/z_2)$, corresponding to independence of Z_1 and Z_2. In this case (Z_1, Z_2) have a joint density function.
In the corresponding D-dimensional case, W falls at the D corners of $\mathbb{S}_{D-1}$ with equal probabilities $1/D$.

- Let $D = 2$ and suppose that $\mathrm{P}(W = 1/2) = 1$. Then

$$V(z_1, z_2) = 2\mathrm{E}\left[\max\{W/z_1, (1-W)/z_2\}\right] = \max(1/z_1, 1/z_2),$$

and hence $G^*(z_1, z_2) = \exp\{-1/\min(z_1, z_2)\}$, corresponding to total dependence of Z_1 and Z_2. Here $Z_1 = Z_2$ with probability one, so they have no joint density function; all the mass of their joint distribution lies on the line $z_1 = z_2$.
In the corresponding D-dimensional case, W equals the barycentre $D^{-1}1_D$ of $\mathbb{S}_{D-1}$ with probability one.

- Let $D = 2$ and let W have the uniform distribution on $[0, 1]$. Then $w/z_1 \geq (1-w)/z_2$ when $w \geq z_1/(z_1 + z_2)$, and it is easy to check that

$$V(z_1, z_2) = 2\mathrm{E}\left[\max\{W/z_1, (1-W)/z_2\}\right] = 1/z_1 + 1/z_2 - 1/(z_1 + z_2).$$

In this case both W and (Z_1, Z_2) have density functions, with no atoms.

Comments

☐ The strategy above was:

– use g to transform the original data to standard (unit Fréchet) margins;

– show that on these standard margins the limiting distribution for the transformed data has a specific nonparametric form, subject only to restrictions on the marginal means.

☐ We saw that $\mu(\mathcal{A})$ (for the original data) equals $\mu^*\{g(\mathcal{A})\}$ (for the transformed data), so

– since g is monotone on each axis, $\mathcal{A}_z^* = g(\mathcal{A}_z)$ has the same shape as $\mathcal{A}_z$, and

– if we set $z = (z_1, z_2) = ((x - b_{X,n})/a_{X,n}, (y - b_{Y,n})/a_{Y,n})$, then

$$\begin{aligned} \mathrm{P}(M_{X,n} \le x, M_{Y,n} \le y) &= \mathrm{P}\{(M_{X,n} - b_{X,n})/a_{X,n} \le z_1, (M_{Y,n} - b_{Y,n})/a_{Y,n} \le z_2\} \\ &\approx \exp\{-\mu(\mathcal{A}_z)\} \\ &= \exp[-\mu^*\{g(\mathcal{A}_z)\}] \\ &= \exp[-V\{g(\mathcal{A}_z)\}], \end{aligned}$$

gives an approximate distribution function for $(M_{X,n}, M_{Y,n})$, which we can use to find their joint density.

☐ For statistical purposes we need models for V, or equivalently ν or $\dot{\nu}$, ...

5.4 Statistical Modelling slide 185

Introduction

☐ The previous section gave a general framework for multivariate extremes, in which a general nonparametric model appears.

☐ Two approaches to modelling are:

– estimating the full nonparametric model — difficult because of limited data;

– fitting a parametric sub-family of models — may be restrictive, but often is good enough.

☐ To do this we need

– some (reasonably flexible) parametric families of models,

– methods to fit them (likelihood),

– methods to assess their fit and to compare them.

Parametric models

It is tricky to formulate parametric models that satisfy the mean constraints in $D \geq 3$, but numerous models exist for $D = 2$.

Example 34 *The* **logistic model** *for general* D *has*

$$V(z_1, \ldots, z_D) = \left(\sum_{d=1}^{D} z_d^{-1/\alpha} \right)^{\alpha}, \quad z_1, \ldots, z_D > 0, 0 < \alpha \leq 1,$$

and for $D = 2$,

$$\dot{\nu}(w) = \tfrac{1}{2}(\alpha^{-1} - 1)\{w(1-w)\}^{-1-1/\alpha}\{w^{-1/\alpha} + (1-w)^{-1/\alpha}\}^{\alpha-2}, \quad 0 < w < 1.$$

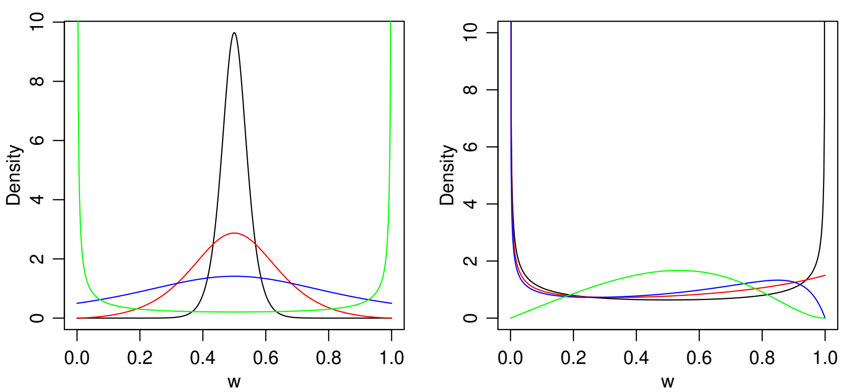
Independence and perfect dependence arise as limits as $\alpha \uparrow 1$ *and* $\alpha \downarrow 0$ *respectively.*
This model is limited by having only one parameter, which makes it symmetric and too inflexible for most purposes. The same applies to the bivariate **negative logistic model**, *which has*

$$V(z_1, z_2) = 1/z_1 + 1/z_2 - (z_1^{\alpha} + z_2^{\alpha})^{-1/\alpha}, \quad \alpha > 0,$$

for which independence and perfect dependence arise when $\alpha \to 0$ *and* $\alpha \to \infty$.

Logistic and Dirichlet densities

Left: logistic densities $\dot{\nu}$ with $\alpha = 0.1$ (black), 0.3 (red), 0.5 (blue), 0.9 (green).
Right: Dirichlet densities $\dot{\nu}$ with parameters $(\alpha, \beta) = (0.5, 0.5)$ (black), (0.5, 1) (red), (0.5, 2) (blue) and (2,3) (green).



Hüsler–Reiss distribution

Example 35 (Hüsler–Reiss distribution) *This is a natural analogue of the normal distribution in extremal contexts. The bivariate version has a scalar parameter $\lambda > 0$ and exponent function*

$$V(z_1, z_2) = \frac{1}{z_1}\Phi\left\{\frac{\lambda}{2} + \frac{1}{\lambda}\log\left(\frac{z_2}{z_1}\right)\right\} + \frac{1}{z_2}\Phi\left\{\frac{\lambda}{2} + \frac{1}{\lambda}\log\left(\frac{z_1}{z_2}\right)\right\}, \quad z_1, z_2 > 0, \tag{19}$$

where Φ denotes the standard normal cumulative distribution function.
For this model the angular variable W has density

$$\dot\nu(w) = \frac{e^{-\lambda^2/8}}{2\lambda\{w(1-w)\}^{3/2}}\phi\left\{\frac{1}{\lambda}\log\left(\frac{w}{1-w}\right)\right\}, \quad 0 < w < 1. \tag{20}$$

- □ Recall the angular coordinates $r = z_1 + \cdots + z_D$ and $w = (w_1, \ldots, w_D) = z/r \in \mathbb{S}_{D-1}$ defined in terms of $z = (z_1, \ldots, z_D) \in \mathcal{E}^*$.
- □ If it exists, we obtain the angular density $\dot\nu$ from V using the formula

$$\dot\nu(w) = -\frac{r^{D+1}}{D}\left.\frac{\partial^D V(z_1, \ldots, z_D)}{\partial z_1 \cdots \partial z_D}\right|_{z_1 = rw_1, \ldots, z_D = rw_D}, \quad w \in \mathbb{S}_{D-1}, \quad r > 0.$$

Asymmetric models

Asymmetric models include:

- □ the **bilogistic** model

$$\dot\nu(w) = \tfrac{1}{2}(1-\alpha)(1-w)^{-1}w^{-2}(1-u)u^{1-\alpha}\{\alpha(1-u) + \beta u\}^{-1}, \quad 0 < w < 1,$$

  where $0 < \alpha, \beta < 1$, and $u = u(w, \alpha, \beta)$ satisfies

$$(1-\alpha)(1-w)(1-u)^\beta - (1-\beta)wu^\alpha = 0;$$

- □ and the **Dirichlet** model

$$\dot\nu(w) = \frac{\alpha\beta\Gamma(\alpha+\beta+1)(\alpha w)^{\alpha-1}\{\beta(1-w)\}^{\beta-1}}{2\Gamma(\alpha)\Gamma(\beta)\{\alpha w + \beta(1-w)\}^{\alpha+\beta+1}}, \quad 0 < w < 1,$$

  for parameters $\alpha, \beta > 0$.
- □ The `R` function `evd::fbvevd` (see also `evd::dbvevd`) fits several bivariate models, including all those above.

Pickands' dependence function

☐ **Pickands' dependence function** A, determined by

$$V(z_1, z_2) = \left(\frac{1}{z_1} + \frac{1}{z_2}\right) A\left(\frac{z_1}{z_1 + z_2}\right),$$

gives a useful summary of dependence in bivariate problems. We have

(a) $\max(t, 1-t) \le A(t) \le 1$ for $t \in [0, 1]$;

(b) $A(t) = 1$ for independent data, and $A(t) = \max(t, 1-t)$ for perfectly dependent data;

(c) $A(t)$ is convex in t; and

(d) we can write

$$A(t) = 1 - t + 2\int_0^t \nu([0, w])\,\mathrm{d}w, \quad 0 \le t \le 1.$$

☐ This last formula enables the computation of ν from A, since

$$\nu([0, w]) = \begin{cases} \{1 + A'(w)\}/2, & 0 \le w < 1, \\ 1, & w = 1, \end{cases}$$

where A' is the right-hand derivative of A. Further differentiation gives $\dot{\nu}$, if it exists.

Note: Pickands' dependence function

(a) First, note that with $t = z_1/(z_1+z_2)$, we have

$$A(t) = z_1 z_2 V(z_1,z_2)/(z_1+z_2) = V\{(z_1+z_2)/z_2, (z_1+z_2)/z_1\} = V\{1/(1-t), 1/t\}, \quad 0 \le t \le 1.$$

Now since $V > 0$, comparison of the sets $\mathcal{A}_{(z_1,z_2)}$, $\mathcal{A}_{(z_1,\infty)}$ and $\mathcal{A}_{(\infty,z_2)}$ shows that

$$V(z_1,z_2) \le V(z_1,\infty) + V(\infty,z_2) = 1/z_1 + 1/z_2,$$

and hence $A(t) \le 1$, and likewise

$$V(z_1,z_2) \ge \max\{V(z_1,\infty), V(\infty,z_2)\} = \max(1/z_1, 1/z_2),$$

giving $A(t) \ge \max(t, 1-t)$.

(b) To check the values of A for dependent and independent data, note that

$$A(t) = 2\int_0^1 \max\{w(1-t), (1-w)t\}\,\nu(\mathrm{d}w),$$

and insert the appropriate ν. For example, if $\nu(\{0\}) = \nu(\{1\}) = 1/2$, then $A(t) = 2\{t/2 + (1-t)/2\} = 1$, and if $\nu(\{1/2\}) = 1$, then $A(t) = 2\max\{(1-t)/2, t/2\} = \max(t, 1-t)$, corresponding to the independent and fully dependent models respectively.

(c) For the convexity, note that the function $\max(ax, by)$ is convex for $a, b \ge 0$ and $x, y > 0$, and that linear combinations (with positive coefficients) of convex functions are convex. Thus the (possibly infinite) linear combination of such functions, $A(t)$, is convex in t.

(d) The final part is a bit more delicate. We can write

$$A(t) = 2\left\{(1-t)\int_{(t,1]} w\,\nu(\mathrm{d}w) + t\int_{[0,t]} (1-w)\,\nu(\mathrm{d}w)\right\},$$

and the first integral may be expressed as

$$\begin{aligned}\int_{(t,1]} w\,\nu(\mathrm{d}w) &= \int_{(t,1]} \{1 + (w-1)\}\,\nu(\mathrm{d}w)\\ &= 1 - \nu([0,t]) - \left\{\tfrac{1}{2} - \int_{[0,t]} (1-w)\,\nu(\mathrm{d}w)\right\}\\ &= \tfrac{1}{2} - \nu([0,t]) + \int_{[0,t]} (1-w)\,\nu(\mathrm{d}w).\end{aligned}$$

We can write the remaining integral as

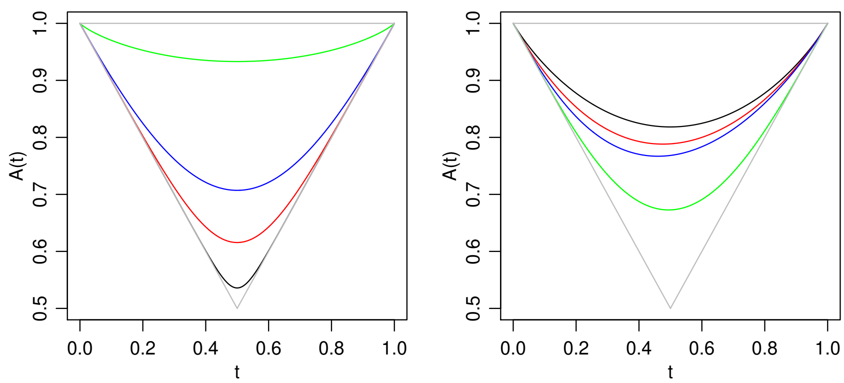
$$\begin{aligned}\int_{[0,t]} (1-w)\,\nu(\mathrm{d}w) &= \int_{[0,t]}\int_w^1 \mathrm{d}u\,\nu(\mathrm{d}w) = \int_0^1\int_{[0,\min(u,t)]} \nu(\mathrm{d}w)\mathrm{d}u\\ &= \int_0^t \nu([0,u])\,\mathrm{d}u + (1-t)\nu([0,t]).\end{aligned}$$

Putting the bits together we get

$$A(t) = 1 - t + 2\int_0^t \nu([0,u])\,\mathrm{d}u, \quad 0 \le t \le 1.$$

Pickands' dependence functions

Left: for logistic density with $\alpha = 0.1$ (black), 0.3 (red), 0.5 (blue), 0.9 (green).
Right: for Dirichlet density with parameters $(\alpha, \beta) = (0.5, 0.5)$ (black), (0.5, 1) (red), (0.5, 2) (blue) and (2,3) (green).



Extremal coefficient

- A common scalar summary of dependence between $(Z_1, \ldots, Z_D)$ with CDF

$$P(Z_1 \leq z_1, \ldots, Z_D \leq z_D) = \exp\{-V(z_1, \ldots, z_D)\}, \quad z_1, \ldots, z_D > 0,$$

  is the so-called **extremal coefficient** $\theta = V(1, \ldots, 1)$, which
  - satisfies $1 \leq \theta \leq D$, and
  - can be interpreted as the 'number of independent maxima' underlying $Z_1, \ldots, Z_D$, because the homogeneity of V gives

$$\begin{aligned} P\{\max(Z_1, \ldots, Z_D) \leq z\} &= P(Z \leq z, \ldots, Z_D \leq z) \\ &= \exp\{-V(z, \ldots, z)\} \\ &= \exp\{-V(1, \ldots, 1)/z\} \\ &= (e^{-1/z})^{\theta}, \end{aligned}$$

    so smaller θ corresponds to stronger dependence.
- For asymptotically dependent models and $D = 2$,

$$\chi = \lim_{z \to \infty} P(Z_2 > z \mid Z_1 > z) = 2 - \theta = 2\{1 - A(1/2)\}.$$

Marginal transformation

- The models for multivariate maxima have unit Fréchet margins, but data do not, so for each margin we have

$$Z_d = \left(1 + \xi_d \frac{Y_d - \eta_d}{\tau_d}\right)_+^{1/\xi_d}, \quad d = 1, \ldots, D,$$

in terms of the original component-wise maxima $Y = (Y_1, \ldots, Y_D)^{\mathrm{T}}$, or in vector form,

$$Z_{D\times 1} = \left(1 + \xi \frac{Y - \eta}{\tau}\right)_+^{1/\xi},$$

where η, τ, ξ are vectors and addition, etc. are component-wise.

- The distribution of the maxima Y is therefore assumed to be

$$\mathrm{P}(Y \le y) = G^* \left\{ \left(1 + \xi \frac{y - \eta}{\tau}\right)_+^{1/\xi} \right\}, \quad x \in \mathbb{R}^D,$$

where $G^*(z) = \exp\{-V(z)\}$ is a simple extreme-value distribution.

- There are at least $3D + 1$ parameters (3 for each margin, and at least 1 for V).

Inference for multivariate maxima

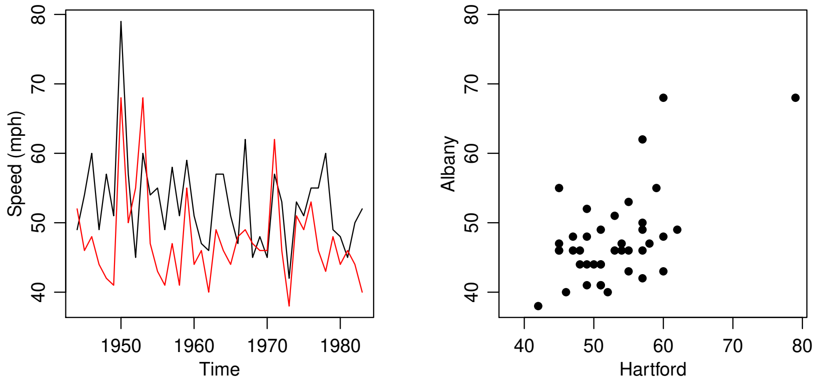
- Inference involves:
  - fitting of marginal GEV distributions and transformation to standard Fréchet;
  - choice of dependence model V (or equivalently ν);
  - estimation of ν (by maximum likelihood in parametric cases);
  - model checking;
  - computation of probabilities for events of interest.
- Ideally all the estimation is performed at once, by fitting marginal and dependence models together (not always feasible in complex cases).
- In the bivariate case, the joint density for the maxima (Y_1, Y_2) can be written as

$$f(y_1, y_2) = \frac{\partial z_1}{\partial y_1}\frac{\partial z_2}{\partial y_2} \times \left\{ \frac{\partial V(z_1, z_2)}{\partial z_1}\frac{\partial V(z_1, z_2)}{\partial z_2} - \frac{\partial^2 V(z_1, z_2)}{\partial z_1 \partial z_2} \right\} \times \exp\left\{-V(z_1, z_2)\right\},$$

where the first two (Jacobian) terms depend on the marginal parameters and the remainder depend both on those parameters and those of V (or equivalently ν).

- For larger D the number of terms with derivatives of V increases very rapidly, but the structure of the density is the same.

Example: Wind data



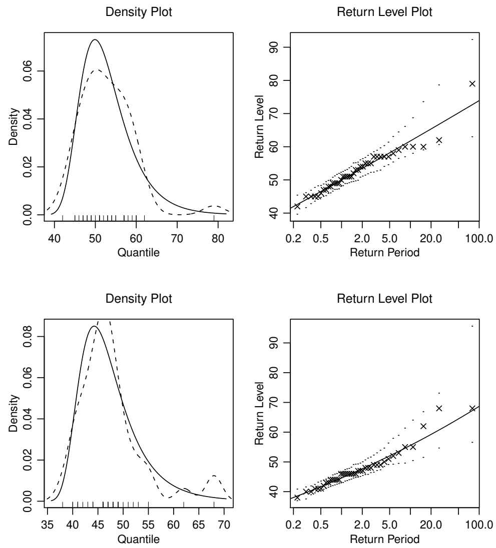
Annual maximum wind speeds at Albany, New York and Hartford, Connecticut respectively, from 1944–1983.

Example: Wind data

```
(fit1<-fbvevd(wind,model="log"))  # Fit logistic dependence function

Call: fbvevd(x = wind, model = "log")
Deviance: 492.1304
AIC: 506.1304
Dependence: 0.3658468

Estimates
    loc1     scale1     shape1       loc2     scale2     shape2        dep
49.96955    5.03097    0.01413   44.58484    4.33938    0.07879    0.70854

Standard Errors
   loc1    scale1    shape1      loc2    scale2    shape2       dep
0.87434   0.63662   0.08826   0.76813   0.56747   0.11101   0.09742

plot(fit1,mar=1,which=c(3,4))
plot(fit1,mar=2,which=c(3,4))
plot(fit1,which=c(3:6))
```

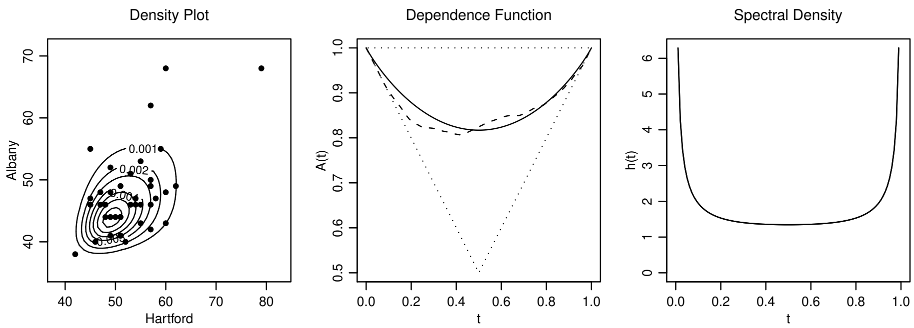
Example: Wind data

Diagnostic plots for the two marginal fits:



Example: Wind data

Diagnostic plots for the fitted logistic model:



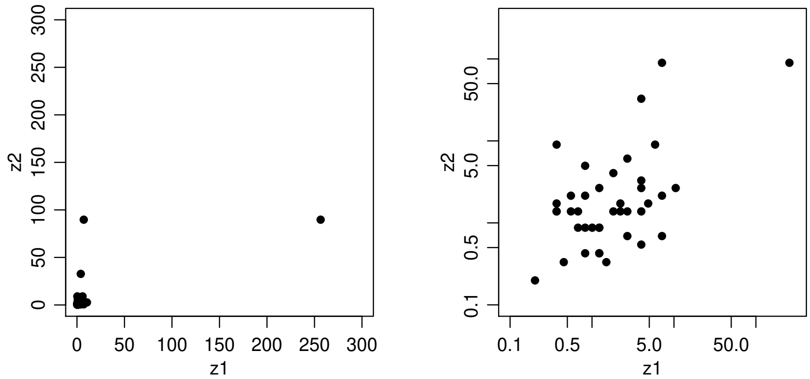
What is wrong with (a) the empirical Pickands function? (b) the spectral density?

Example: Wind data

Residuals

$$\widehat{z}_1 = \left\{1 + \widehat{\xi}_1(y_1 - \widehat{\eta}_1)/\widehat{\tau}_1\right\}_+^{1/\widehat{\xi}_1}, \quad \widehat{z}_2 = \left\{1 + \widehat{\xi}_2(y_2 - \widehat{\eta}_2)/\widehat{\tau}_2\right\}_+^{1/\widehat{\xi}_2}$$

from the fitted model, on the unit Fréchet (left) and Gumbel (right) scales. Note the size of the event in 1950.



Example: Wind data

```
# fit without 1950

(fit2<-fbvevd(wind[-7,],model="log"))

Call: fbvevd(x = wind[-7, ], model = "log")
Deviance: 466.0202
AIC: 480.0202
Dependence: 0.2858570

Estimates
    loc1    scale1    shape1      loc2    scale2    shape2       dep
50.45888   4.98736  -0.31263  44.41348   4.16650   0.08284   0.77749

Standard Errors
  loc1  scale1  shape1    loc2  scale2  shape2     dep
0.9011  0.6727  0.1355  0.7471  0.5383  0.1072  0.1004
```

The dependence parameter has increased, but still is significantly less than $\alpha = 1$. The shape parameter estimate $\widehat{\xi}_1$ is now significantly negative.

Comments

We see that:

- ☐ the marginal fits appear to be adequate (density plots not very useful);
- ☐ the joint fit appears to be reasonable (though there is not much data);
- ☐ the empirical Pickands function is not convex (!), but matches the fitted one fairly well;
- ☐ the angular density suggests that the windspeeds are not very dependent (the fitted angular density shows spikes at $w = 0, 1$), though the standard error of around 0.1 for $\widehat{\alpha} = 0.71$ shows that the data are clearly not completely independent;
- ☐ the 1950 event is quite influential—is it special in any way?

Example: Wind data

Comparison with other models: recall that the deviance equals $-2\widehat{\ell}$ and that $\mathrm{AIC} = -2\widehat{\ell} + 2p$, for a fitted model with p parameters and maximised log likelihood value $\widehat{\ell}$. Small AIC is better.

Dependence function	Paras	Deviance	AIC	$2\{1 - A(1/2)\}$
Logistic	7	492.13	506.13	0.37
Hüsler–Reiss	7	491.20	505.20	0.37
Negative logistic	7	491.57	505.57	0.36
Asymmetric negative logistic	9	491.75	509.75	0.38
Bilogistic	9	489.79	505.79	0.29
Coles–Tawn	8	489.97	505.97	0.38
Asymmetric logistic	—	—	—	—
Asymmetric mixed	—	—	—	—

More comments

- ☐ The Hüsler–Reiss model seems to be best of those that could be fitted, but there is very little difference among them.
- ☐ There is no evidence of a need for an asymmetric dependence function.
- ☐ The dependence is not very strong; this measure is $2 - \theta$, so $\widehat{\theta} \approx 1.63$, corresponding to

$$\mathrm{P}(Z_2 > z \mid Z_1 > z) \approx 2 - \theta \approx 0.37;$$

  this is appreciable but not strong dependence.
- ☐ There is probably a big loss of information due to using only the annual maxima, but at least these can (probably) be treated as independent (need to check the dates to be sure).

Poisson process approach

- If $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F$ lie in $\mathbb{R}^D$, then under the conditions for convergence of the maxima, we can define component-wise transformations

$$g(X) = \{1 + \xi(X - b_n)/a_n\}_+^{1/\xi}, \quad X, \xi, b_n \in \mathbb{R}^D, \quad a_n \in \mathbb{R}_+^D,$$

  such that the sequence of point processes

$$\mathcal{P}_n^* = \{g(X_1), \ldots, g(X_n)\}, \quad n = 1, 2, \ldots,$$

  converges to a Poisson process $\mathcal{P}^*$ on $\mathcal{E}^* = \mathbb{R}_+^D - \{0\}$ with exponent function V and measure μ^*, where $V(z) = \mu^*(\mathcal{A}_z^*)$, and $\mathcal{A}_z^* = \mathcal{E}^* - [0, z_1] \times \cdots \times [0, z_D]$
- The corresponding density based on the points $z_1, \ldots, z_{n_{\mathcal{A}^*}}$ in an 'extreme' set $\mathcal{A}^* \subset \mathcal{E}^*$ is

$$\exp\{-\mu^*(\mathcal{A}^*)\} \times \prod_{j=1}^{n_{\mathcal{A}^*}} \mu^*(\mathrm{d}z_j),$$

  with $\mu^*(\mathrm{d}z_j)$ replaced by $\dot{\mu}^*(z_j)$ when μ^* has a density function.
- If the limit process is assumed to be exact for events in $\mathcal{A}^*$ (see below), we can base inference for μ^* on this expression.

Marginal transformation

- If we are interested in an extreme region $\mathcal{B}$ in $\mathcal{E}$, we choose a region $\mathcal{A}^* \subset \mathcal{E}^*$ in the transformed data space such that $g(\mathcal{B}) \subset \mathcal{A}^*$ and use the Poisson process model.
- For the **marginal transformation** g, we take thresholds $u_1, \ldots, u_D$, usually corresponding to the same quantile (e.g., 0.95) of each margin; n_{u_d} observations exceed these thresholds, and $\widehat{p}_d = n_{u_d}/n$ is the estimated exceedance probability for dimension d.
- We fit GPDs above these thresholds, giving fitted marginal distributions

$$\widehat{F}_d(x) = \begin{cases} \#\{j : x_{j,d} \le x\}/n, & x \le u_d, \\ 1 - \widehat{p}_d \left\{1 + \widehat{\xi}_d(x - u_d)/\widehat{\sigma}_d\right\}_+^{-1/\widehat{\xi}_d}, & x > u_d, \end{cases} \quad d = 1, \ldots, D,$$

  based on the D marginal GP parameter estimates $(\widehat{\sigma}_d, \widehat{\xi}_d)$.
- We then apply this estimated probability integral transformation component-wise to $x_j = (x_{j,1}, \ldots, x_{j,D}) \in \mathcal{E}$ to get

$$z_j = -1/\log \widehat{F}(x_j), \quad j = 1, \ldots, n,$$

  which lie in $\mathcal{E}^*$ and have approximate unit Fréchet margins.
- There are corresponding angular variables

$$r_j = \|z_j\|_1 > 0, \quad w_j = z_j/r_j \in \mathbb{S}_{D-1}, \quad j = 1, \ldots, n.$$

Extremal region $\mathcal{A}^*$

- We have to choose an 'extreme' region $\mathcal{A}^*$ on which to base the likelihood

$$\exp\{-\mu^*(\mathcal{A}^*)\} \times \prod_{j=1}^{n_{\mathcal{A}^*}} \mu^*(\mathrm{d}z_j), \quad \mathcal{A}^* \subset \mathcal{E}^*.$$

- In most cases μ^* has a tractable density $\dot\mu^*$, so the bottleneck is computation of $\mu^*(\mathcal{A}^*)$.
- In terms of the angular coordinates, $\mu^*(\mathrm{d}z_j) = r_j^{-2}\mathrm{d}r \times D\,\nu(\mathrm{d}w_j) \propto \dot\nu_\theta^*(w_j)$, if there is a parametric angular density.
- If $\mathcal{A}^* = \{(x, y) : x + y > r_0\}$ for some large r_0, then

$$\mu^*(\mathcal{A}^*) = 2\int_{\mathcal{A}^*} \frac{\mathrm{d}r}{r^2}\,\nu(\mathrm{d}w) = 2\int_{r=r_0}^{\infty} \frac{\mathrm{d}r}{r^2} \int_{w=0}^{1} \nu(\mathrm{d}w) = 2/r_0,$$

  does not depend on parameters of ν^*. In this case the likelihood is

$$L(\theta) = \exp\{-\mu^*(\mathcal{A}^*)\} \prod_{j=1}^{n_{\mathcal{A}^*}} \dot\mu^*(z_j) \propto \prod_{j=1}^{n_{\mathcal{A}^*}} \dot\nu_\theta^*(w_j),$$

  since none of the other terms depend on $\dot\nu_\theta^*$ (or θ).
- This likelihood is simple but not much used, because non-extreme data can corrupt the fit.

Censored likelihood

- For simplicity suppose $D = 2$ and $z = (z_1, z_2) \in \mathcal{E}^*$.
- Even if r_0 is large, sets such as

$$\mathcal{A}^* = \{(z_1, z_2) : z_1 + z_2 > r_0 > 0\}$$

  contain values for which one of z_1 and z_2 is small, so asymptotic models may not apply.
- To fix this we split $\mathcal{E}^*$ into subsets

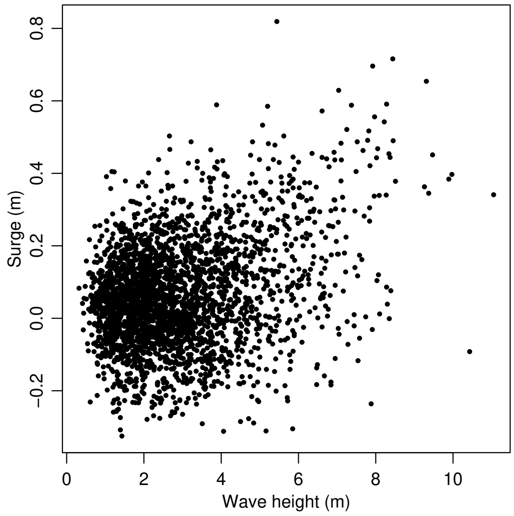
$$\begin{aligned} \mathcal{E}_{00}^* &= \{(z_1, z_2) : z_1 \le u_1, z_2 \le u_2\}, & \mathcal{E}_{10}^* &= \{(z_1, z_2) : z_1 > u_1, z_2 \le u_2\}, \\ \mathcal{E}_{01}^* &= \{(z_1, z_2) : z_1 \le u_1, z_2 > u_2\}, & \mathcal{E}_{11}^* &= \{(z_1, z_2) : z_1 > u_1, z_2 > u_2\}, \end{aligned}$$

  with respective likelihood contributions based on

$$\begin{aligned} &\mathrm{P}(Z_1 \le u_1, Z_2 \le u_2), && \frac{\partial \mathrm{P}(Z_1 \le z_1, Z_2 \le u_2)}{\partial z_1}, \\ &\frac{\partial \mathrm{P}(Z_1 \le u_1, Z_2 \le z_2)}{\partial z_2}, && \frac{\partial^2 \mathrm{P}(Z_1 \le z_1, Z_2 \le z_2)}{\partial z_1 \partial z_2}. \end{aligned}$$

  This uses the full information about the values of (z_1, z_2) only in $\mathcal{E}_{11}^*$, and otherwise just uses the information that $z_{1,j}$ or $z_{2,j}$ falls below the appropriate threshold.
- This **censored likelihood** is the default for fitting Poisson process models to multivariate data.

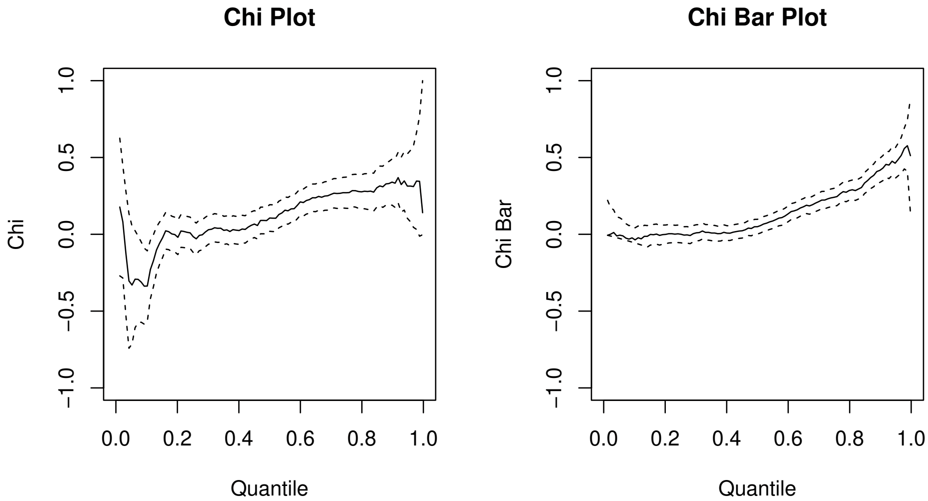
Example: Oceanographic data



Simultaneous values of wave and surge height.

http://stat.epfl.ch slide 209

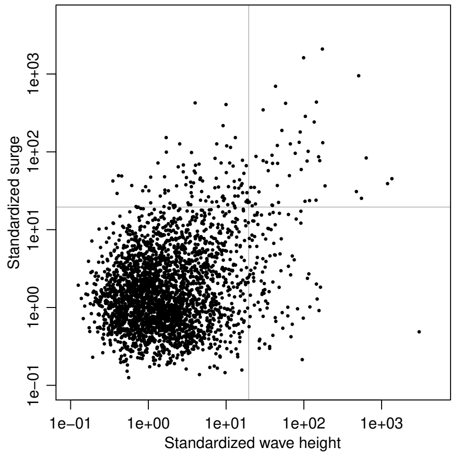
Example: Oceanographic data



Estimates of $\chi(u)$ and $\overline{\chi}(u)$. The wide confidence intervals as $u \to 1$ are typical (and indeed inevitable) and complicate the interpretation of such plots.

http://stat.epfl.ch slide 210

Example: Oceanographic data



Simultaneous values of wave and surge height, transformed to unit Fréchet scale, with regions $\mathcal{E}^*_{00}, \ldots, \mathcal{E}^*_{11}$ determined by the grey lines marking the marginal 0.95 quantiles.

Example: Oceanographic data

```
> (fit <- fbvpot(wavesurge,apply(wavesurge,2,quantile,0.95),model="log"))

Call: fbvpot(x = wavesurge, threshold = apply(wavesurge, 2, quantile,0.95), model = "log")
Deviance: 2036.076
AIC: 2046.076
Dependence: 0.3072850

Threshold: 6.08 0.322
Marginal Number Above: 144 144
Marginal Proportion Above: 0.0498 0.0498
Number Above: 49
Proportion Above: 0.0169

Estimates
   scale1     shape1     scale2     shape2        dep
 1.261341  -0.134651   0.091877   0.008904   0.759339

Standard Errors
 scale1   shape1   scale2   shape2      dep
0.13162  0.06908  0.01067  0.08568  0.02945
```

Example: Oceanographic data

Results for fits: recall that the deviance equals $-2\widehat{\ell}$ and that $\mathrm{AIC} = -2\widehat{\ell} + 2p$, for a fitted model with p parameters and maximised log likelihood value $\widehat{\ell}$. Small AIC is better.

Dependence function	Paras	Deviance	AIC	$2\{1 - A(1/2)\}$
Logistic	5	2036.08	2046.08	0.31
Hüsler–Reiss	5	2035.38	2045.38	0.30
Negative logistic	5	2034.91	2044.91	0.30
Bilogistic	6	2035.80	2047.80	0.31
Coles–Tawn	6	2035.35	2047.35	0.31
Negative bilogistic	6	2034.85	2046.85	0.31
Asymmetric mixed	6	2044.15	2056.15	0.33
Asymmetric logistic	7	2036.60	2050.60	0.31
Asymmetric negative logistic	7	2035.78	2049.78	0.31

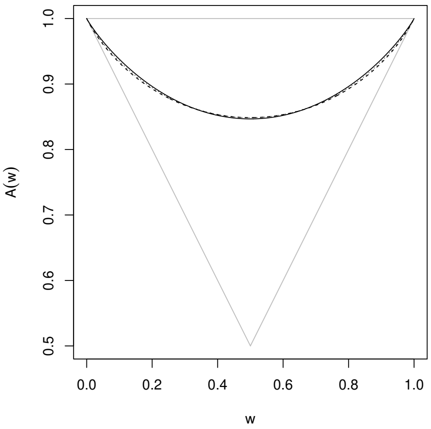
The first three fits seem best, based on the AIC values, but the differences are not large overall; it seems that the data are not complex enough to warrant a very complex (or asymmetric) model.

Example: Oceanographic data

Model	Deviance	χ	σ_1	ξ_1	σ_2	ξ_2	Dep
Logistic	2036.08	0.31	$1.26_{0.13}$	$-0.13_{0.07}$	$0.09_{0.01}$	$0.01_{0.09}$	$0.76_{0.03}$
Hüsler–Reiss	2035.38	0.30	$1.25_{0.13}$	$-0.11_{0.07}$	$0.09_{0.01}$	$0.03_{0.08}$	$0.97_{0.07}$
Negative logistic	2034.91	0.30	$1.25_{0.13}$	$-0.12_{0.07}$	$0.09_{0.01}$	$0.01_{0.08}$	$0.58_{0.06}$

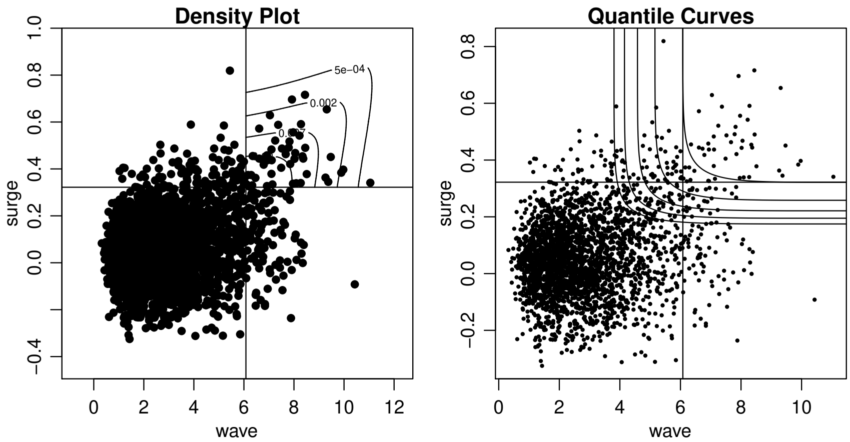
- ☐ The fits are very similar, showing data whose extremes are clearly dependent, even if the probability of high extremes on one variable given them on the other is not very high.
- ☐ The dependence parameters are not comparable (because the models are not the same), but they all give similar values of χ.
- ☐ The same censored fitting approach can be used when $D \geq 3$, but the coding becomes quite complex, with 2^D cases.

Example: Oceanographic data



Fitted Pickands functions for the logistic (solid), Hüsler–Reiss (dots) and negative logistic (dashed) models. They are essentially identical.

Example: Oceanographic data



Diagnostic plots (from `plot(fit)`) for threshold fit to oceanographic data. (A bug prevents the points on the left from being smaller.) According to the help, the quantiles are at 0.75, 0.8, . . . , 0.95 by default.