A conditional approach for multivariate extreme values

by Janet Heffernan and Jonathan Tawn

J.R. Statist. Soc. B **66**, 497–546 (with discussion)

Discussion by Richard Smith

Presented to NCAR Extremes Working Group

February 23 2006

A set $C \in \mathcal{R}^d$ is an "extreme set" if

- we can partition $C = \cup_{i=1}^{d} C_i$ where

$$C_i = C \cap \left\{ x \in \mathcal{R}^d : \; F_{X_i}(x_i) > F_{X_j}(x_j) \right\} \text{ for all } i \neq j$$

  (C_i is the set on which X_i is "most extreme" among $X_1, ..., X_d$, extremeness being measured by marginal tail probabilities)

- The set C_i satisfies the following property: there exists a v_{X_i} such that

$$X \in C_i \text{ if and only if } X \in C_i \text{ and } X_i > v_{X_i}.$$

  (or: if $\mathbf{X} = (X_1, ..., X_d) \in C_i$ then so is any other $\mathbf{X}$ for which X_i is more extreme)

The objective of the paper is to propose a general methodology for estimating probabilities of extreme sets.

Marginal distributions

- Standard "GPD" model fit: define a threshold u_{X_i} and assume

$$\Pr\{X_i > u_{X_i} + x \mid X_i > u_{X_i}\} = \left(1 + \xi_i \frac{x}{\beta_i}\right)_+^{-1/\xi_i}$$

  for $x > 0$.

- Also estimate $F_{X_i}(x)$ by empirical CDF for $x < u_{X_i}$.

- Combine together for an estimate $\widehat{F}_{X_i}(x)$ across the entire range of x.

- Apply componentwise probability integral transform — $Y_i = -\log[-\log\{\widehat{F}_{X_i}(X_i)\}]$ to have approximately Gumbel margins (i.e. $\Pr\{Y_i \le y_i\} \approx \exp(-e^{-y})$).

- Henceforth assume marginal distributions are exactly Gumbel and concentrate on dependence among $Y_1, ..., Y_d$.

Existing techniques

Most existing extreme value methods with Gumbel margins reduce to

$$\Pr\{\mathbf{Y} \in t + A\} \approx e^{-t/\eta_{\mathbf{Y}}} \Pr\{\mathbf{Y} \in A\}$$

for some $\eta_{\mathbf{Y}} \in (0, 1]$. Ledford-Tawn classification:

- $\eta_{\mathbf{Y}} = 1$ is *asymptotic dependence* (includes all conventional multivariate extreme value distributions)
- $\frac{1}{d} < \eta_{\mathbf{Y}} < 1$ — positive extremal dependence
- $0 < \eta_{\mathbf{Y}} < \frac{1}{d}$ — negative extremal dependence
- $\eta_{\mathbf{Y}} = \frac{1}{d}$ — near extremal independence

Disadvantage: doesn't work for extreme sets that are not simultaneously extreme in all components

The key assumption of this paper

Define $\mathbf{Y}_{-i}$ to be the vector $\mathbf{Y}$ with i'th component omitted.

We assume that for each y_i, there exist vectors of normalizing constants $\mathbf{a}_{|i}(y_i)$ and $\mathbf{b}_{|i}(y_i)$ and a limiting $(d-1)$-dimensional CDF $G_{|i}$ such that

$$\lim_{y_i \to \infty} \Pr\{\mathbf{Y}_{-i} \leq \mathbf{a}_{|i}(y_i) + \mathbf{b}_{|i}(y_i)\mathbf{z}_{|i}\} \; = \; G_{|i}(\mathbf{z}_{|i}). \qquad (1)$$

Put another way: as $u_i \to \infty$ the variables $Y_i - u_i$ and

$$\mathbf{Z}_{-i} \; = \; \frac{\mathbf{Y}_{-i} - \mathbf{a}_{|i}(Y_i)}{\mathbf{b}_{|i}(Y_i)}$$

are asymptotically independent, with distributions unit exponential and $G_{|i}(\mathbf{z}_{|i})$.

Examples in case $d = 2$

Distribution	η	$a(y)$	$b(y)$
Perfect pos. dependence	1	y	1
Bivariate EVD	1	y	1
Bivariate normal, $\rho > 0$	$\frac{1+\rho}{2}$	$\rho^2 y$	$y^{1/2}$
Inverted logistic, $\alpha \in (0, 1]$	$2^{-\alpha}$	0	$y^{1-\alpha}$
Independent	$\frac{1}{2}$	0	1
Morganstern	$\frac{1}{2}$	0	1
Bivariate normal, $\rho < 0$	$\frac{1+\rho}{2}$	$-\log(\rho^2 y)$	$y^{-1/2}$
Perfect neg. dependence	0	$-\log(y)$	1

Key observation: in all cases $b(y) = y^b$ for some b and $a(y)$ is either 0 or a linear function of y or $-\log y$ (for more precise conditions see equation (3.8) of the paper)

The results so far suggest the *conditional dependence model* (Section 4.1) where we assume the asymptotic relationships conditional on $Y_i = y_i$ are exact above a given threshold u_{Y_i}, or in other words

$$\Pr\{\mathbf{Y}_{-i} < \mathbf{a}_{|i}(y_i) + \mathbf{b}_{|i}(y_i)\mathbf{z}_{|i} \mid Y_i = y_i\} \;=\; G_{|i}(\mathbf{z}_{|i}), \qquad y_i > u_{Y_i}.$$

Here $\mathbf{a}_{|i}(y_i)$ and $\mathbf{b}_{|i}(y_i)$ are assumed to be parametrically dependent on y_i through one of the alternative forms given by (3.8), and $G_{|i}(\mathbf{z}_{|i})$ is estimated nonparametrically from the empirical distribution of the standardized variables

$$\hat{\mathbf{Z}}_{-i} \;=\; \frac{\mathbf{Y}_{-i} - \hat{\mathbf{a}}_{|i}(y_i)}{\hat{\mathbf{b}}_{|i}(y_i)} \text{ for } Y_i = y_i > u_{Y_i}.$$

N.B. The threshold u_{Y_i} does not have to correspond to the threshold u_{X_i} used for estimating the marginal GPDs.

Some issues raised by this representation:

- Self-consistency of separate conditional models? (Section 4.2)

- Extrapolation — critical to have parametric forms for $\mathbf{a}_{|i}(y_i)$ and $\mathbf{b}_{|i}(y_i)$ (Section 4.3)

- Diagnostics — combine standard diagnostics for marginal extremes with tests of independence of $\mathbf{Z}_{-i}$ and Y_i. Also test whether the separate components of $\mathbf{Z}_{-i}$ are independent, since estimation via empirical distribution is much simpler in this case.

Inference

1. *Estimation of marginal parameters*

If ψ denotes the collection of all (β_i, ξ_i) parameters for the individual GPDs, maximize

$$\log L(\psi) \;=\; \sum_{i=1}^{d} \sum_{k=1}^{n_{u_{X_i}}} \log \widehat{f}_{X_i}(x_{i|i,k})$$

Here $n_{u_{X_i}}$ is the number of threshold exceedances in the i'th component and $\widehat{f}_{X_i}(x_{i|i,k})$ is the GP density evaluated at the k'th exceedance.

[In essence, if there is no functional dependence among the (β_i, ξ_i) for different i then this is just the usual marginal estimation of the GPD in each component. But if there is functional dependence, we estimate the parameters jointly by combining the individual likelihood estimation equations, ignoring dependence among the components.]

2. *Single conditional*

i.e. How do we estimate $\mathbf{a}_{|i}(y_i)$ and $\mathbf{b}_{|i}(y_i)$ for a single i, assuming parametric representation?

The problem: don't know the distribution of $\mathbf{Z}_{|i}$.

The solution: do it as if the $\mathbf{Z}_{|i}$ were Gaussian with known means $\boldsymbol{\mu}_{|i}$ and standard deviations $\boldsymbol{\sigma}_{|i}$

This leads to the formulae

$$\begin{aligned}\boldsymbol{\mu}_{|i}(y) &= \mathbf{a}_{|i}(y) + \boldsymbol{\mu}_{|i}\mathbf{b}_{|i}(y),\\ \boldsymbol{\sigma}_{|i}(y) &= \boldsymbol{\sigma}_{|i}\mathbf{b}_{|i}(y),\end{aligned}$$

and estimating equation

$$Q_i = -\sum_{j\neq i}\sum_{k=1}^{n_{u_{Y_i}}}\left[\log \sigma_{j|i}(y_{i|i,k}) + \frac{1}{2}\left\{\frac{y_{j|i,k} - \mu_{j|i}(y_{i|i,k})}{\sigma_{j|i}(y_{i|i,k})}\right\}^2\right].$$

3. *All conditionals*

To estimate all $\mathbf{a}_{|i}(y_i)$ and $\mathbf{b}_{|i}(y_i)$ jointly maximize

$$Q = \sum_{i=1}^{d} Q_i.$$

Analogous with *pseudolikelihood estimation* (Besag 1975)

4. *Uncertainty*

Bootstrap.....

Air quality application

The data: daily values of five air pollutants (O_3, NO_2, NO, SO_2, PM_{10}) in Leeds, U.K., during 1994–1998.

Two seasons: winter (NDJF), early summer (AMJJ)

Omit values around November 5 and some clear outliers

Marginal model: fit GPD above a (somewhat) high threshold, estimate 99% quantile with standard error (Table 4)

Dependence model

Transform margins to Gumbel, select threshold for dependence modeling. Selected to be 70% quantile (for all five variables)

Estimate $(a_{j|i}, b_{j|i})$ and $(a_{i|j}, b_{i|j})$ for each combination of $i \neq j$, with sampling variability represented by convex hull of 100 bootstrap simulations (Fig. 5).

Several pairs do not exhibit weak pairwise exchangeability (e.g. PM_{10}, O_3 in summer)

Components of $\mathbf{Z}_{|i}$ are typically dependent

Some pairs exhibit negative dependence (e.g. O_3, NO in winter — consistent with chemical reactions)

Fig. 6 shows pseudosamples of other variables given NO over threshold ($C^5(23)$ are points for which sum of all 5 variables on Gumbel scale exceeds 23)

Estimation of critical functionals

Contrast "joint probability" with "structure variable" approach (Coles and Tawn 1994)

(a) Estimate conditional mean of each other variable given that NO is about 95% or 99% threshold (Table 5)

(b) Sums of variables on Gumbel scale, e.g. for a subset $\mathcal{M} \subseteq \{1, ..., d\}$ with $|\mathcal{M}| = m$, define $C^m(v) = \{\mathbf{y} : \sum_{i \in \mathcal{M}} y_i > v\}$; define p-quantile v_p by property $\Pr\{\mathbf{Y} \in C(v_p)\} = p$.

Compute return-level estimates for $C^m(v)$ with $\mathcal{M}$ corresponding to (O_3,NO_2) and (NO_2,SO_2,PM_{10}) (Fig. 7)

Comments on the whole approach

(based on my "vote of thanks")

1. Is it general enough? Consider independent Gumbel (Z_1, Z_2) and define

$$\begin{aligned} Y_1 &= Z_1, \\ Y_2 &= \max(Z_1, Z_2) - \log 2. \end{aligned}$$

The conditional distribution of $Y_1 \mid Y_2 = y_2$ is a mixture: either $Y_1 = Y_2$ with probability $\frac{1}{2}$, or Y_1 is Gumbel independently of y_2. This sort of behavior is not captured by renormalizing to an asymptotic distribution

However Heffernan and Tawn in their response pointed out that this sort of degeneracy arises with other MEV models as well

2. Form of $\mathbf{a}_{|i}(y_i)$ and $\mathbf{b}_{|i}(y_i)$

A very interesting preprint has appeared by Janet Heffernan and Sid Resnick (see either of their webpages)

For bivariate (X, Y), they consider conditions under which

$$\lim_{t\to\infty} t \, \mathrm{Pr}\left\{\frac{X-\beta(t)}{\alpha(t)} \leq x, \frac{Y-b(t)}{a(t)} > y\right\} \;=\; \mu([-\infty, x] \times [y, \infty]) \tag{2}$$

for functions $a(t), b(t), \alpha(t), \beta(t)$ and a nondegenerate measure μ.

For example if $b(t) = 0, a(t) = t$ (which they call *standard form*), (2) leads to

$$\lim_{t\to\infty} \mathrm{Pr}\left\{\frac{X-\beta(t)}{\alpha(t)} \leq x \mid Y > t\right\} \;=\; \mu([-\infty, x] \times [0, \infty]) \tag{3}$$

Under this condition they show

$$\lim_{t\to\infty} \frac{\alpha(tc)}{\alpha(t)} \;=\; \psi_1(c), \quad \lim_{t\to\infty} \frac{\beta(tc)-\beta(t)}{\alpha(t)} \;=\; \psi_2(c),$$

from which it follows that $\psi_1(x) = x^\rho$ for some ρ, and

$$\psi_2(x) = \begin{cases} k\frac{x^\rho-1}{\rho} & \text{if } \rho \neq 0 \\ k\log x & \text{if } \rho = 0 \end{cases}$$

(but k may be 0, which causes problems)

They consider cases for which limits do or do not exist. For example, if (X, Y) is bivariate normal, transforming X to log Pareto form is OK, but transforming X to Pareto is not. They also show much stronger connections with traditional multivariate regular variation theory.

3. Estimation methods — why so many different techniques?

4. Application — needs a punch line.

My suggestion was to use chemical models to determine combinations of pollutants that would be likely to result in violations of standards — could have an impact on control strategies.

5. Nobody really took them to task on the possible lack of consistency of conditional distributions — still a potential problem with this whole approach.