REVIEW OF MULTIVARIATE EXTREMES

Richard L. Smith
Department of Statistics and Operations Research
University of North Carolina
Chapel Hill, NC 27599-3260
rls@email.unc.edu

SAMSI Risk Program,
Working Group on Multivariate
Extremes Methodology
October 11, 2007

- I. BASIC THEORY
- II. EXAMPLES OF MULTIVARIATE EXTREME VALUE DISTRIBUTIONS
- III. ESTIMATION
- IV. DEPENDENCE MEASURES

I. BASIC THEORY

Three viewpoints of extreme values theory:

1. Limit theorems for sample maxima
   - Three types theorem
   - Generalized Extreme Value distribution

2. Exceedances Over Thresholds
   - Generalized Pareto distribution

3. Point process approach
   - Joint distribution of exceedance time and excess values approximated by a nonhomogeneous Poisson process

Limit theorems for multivariate sample maxima —

Let $\mathbf{Y}_i = (Y_{i1}...Y_{id})^T$ be i.i.d. d-dimensional vectors, $i = 1, 2, ...$

$M_{nj} = \max\{Y_{1j}, ..., Y_{nj}\} (1 \le j \le d)$ — j'th-component maximum

Look for constants a_{nj}, b_{nj} such that

$$\Pr\left\{\frac{M_{nj} - b_{nj}}{a_{nj}} \le x_j,\ j = 1, ..., d\right\} \ \rightarrow\ G(x_1, ..., x_d).$$

Vector notation:

$$\Pr\left\{\frac{\mathbf{M}_n - \mathbf{b}_n}{\mathbf{a}_n} \le \mathbf{x}\right\} \ \rightarrow\ G(\mathbf{x}).$$

Before going on, two rather easy points:

1. If we fix some $j' \in \{1, ..., d\}$ and define $x_j = +\infty$ for $j \neq j'$, we deduce

$$\Pr\left\{\frac{M_{nj'} - b_{nj'}}{a_{nj'}}\right\} \;\to\; G(\infty, \infty, ..., x_{j'}, ..., \infty).$$

   Therefore, all the marginal distributions of G are GEV.

2. If we know the joint distribution of maxima from $\{Y_{ij},\ i = 1, 2, 3, ..., j = 1, ..., d\}$, then we immediately know also the joint distribution of $\{g_j(Y_{ij})\}$ for any monotone increasing functions $\{g_j,\ j = 1, ..., d\}$. This is true because

$$\max\{g_j(Y_{1j}), ..., g_j(Y_{nj})\} \;=\; g_j(\max\{Y_{1j}, ..., Y_{nj}\}).$$

Therefore, without loss of generality, we may restrict the marginal distributions of G to be any member of the GEV family. Common choices are the Gumbel law, $e^{-e^{-x}}$, and the Fréchet law, $e^{-x^{-\alpha}}$ for some $\alpha > 0$. Here we use Fréchet, often with $\alpha = 1$.

Basic of Multivariate Regular Variation

(following Resnick (2006), Chapter 6)

After transformation of margins,

$$\lim_{t\to\infty} t \,\text{Pr}\left\{\frac{\mathbf{X}_i}{b(t)} \in A\right\} \;=\; \nu(A)$$

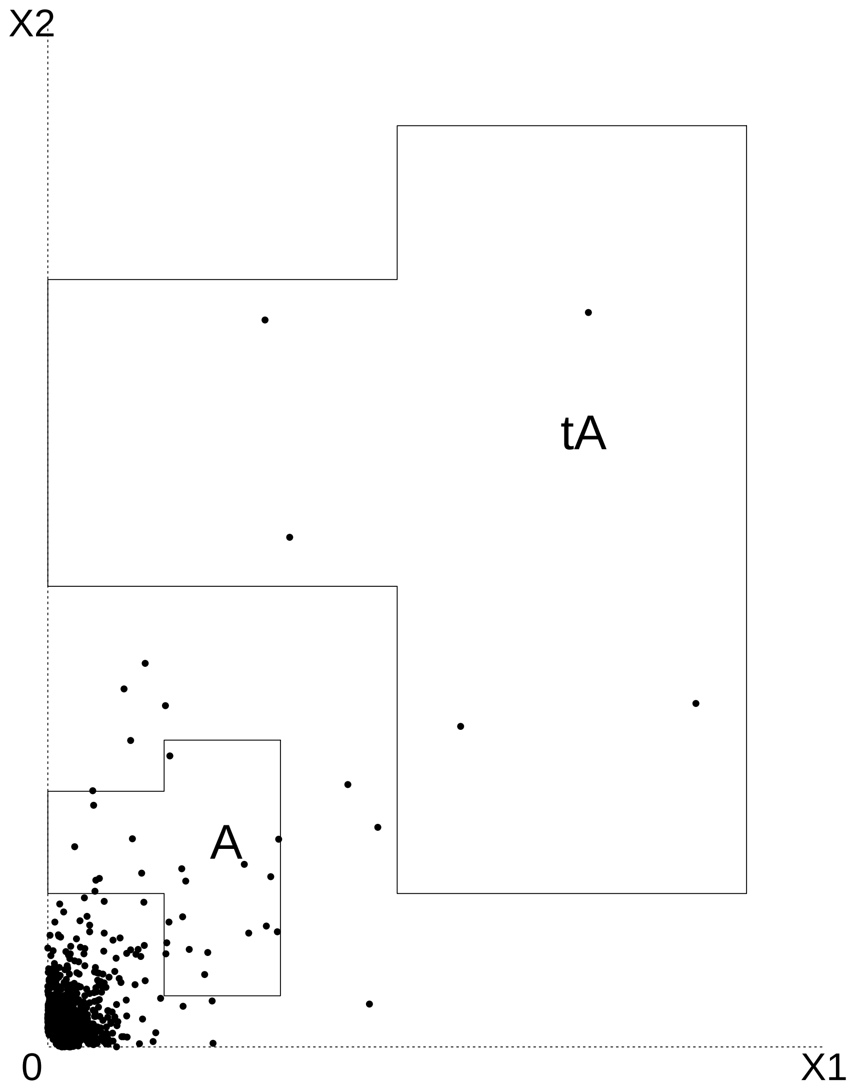
b regularly varying function of index $\alpha > 0$ (w.l.o.g. $\alpha = 1$), ν a measure on the cone

$$\mathcal{E} \;=\; [0,\infty]^d - \{0\}$$

satisfying

$$\nu(tA) \;=\; t^{-\alpha}\nu(A)$$

for any scalar $t > 0$.


X2
tA
A
0
X1

The last statement implies that ν can be decomposed into a product of *radial* and *angular* components. Define

$$\mathcal{S}_d = \{(x_1, ..., x_d) : \; x_1 \geq 0, ..., x_d \geq 0, x_1 + ... + x_d = 1\}.$$

Consider sets A of form

$$A = \left\{ \mathbf{x} \in \mathcal{E} : \; ||\mathbf{x}|| > r, \; \frac{\mathbf{x}}{||\mathbf{x}||} \in S \right\}$$

for some $S \in \mathcal{S}_d$.

Then

$$\nu(A) = r^{-\alpha} H(S)$$

for some measure H on $\mathcal{S}_d$.

$||\cdot||$ can be any norm but the choice of norm affects the definition of H. Henceforth assume $||\mathbf{x}|| = \sum_{j=1}^{d} x_j$. Also, assume $\alpha = 1$.

First Interpretation:

Consider i.i.d. vectors $\mathbf{X}_i, \ \ i = 1, 2, ...\}$ whose distribution is MRV.

Let P_n be a measure on $[0, \infty]^d$ consisting of the points $\left\{\frac{\mathbf{X}_1}{b(n)}, ..., \frac{\mathbf{X}_n}{b(n)}\right\}$.

Let A be a measurable set on $\mathcal{E}$, then the expected number of points of P_n in A is

$$n \, \mathrm{Pr}\left\{\frac{\mathbf{X}_i}{b(n)} \in A\right\} \ \rightarrow \ \nu(A) \ \text{ as } n \rightarrow \infty.$$

With some measure-theoretic formalities, this shows that P_n converges vaguely to a nonhomogeneous Poisson process on $\mathcal{E}$ with intensity measure ν.

Second Interpretation:

Fix $x_1 \geq 0, ..., x_d \geq 0$, $\sum_{j=1}^{d} x_j > 0$. Let A be the complement of

$$[0, x_1] \times [0, x_2] \times ... \times [0, x_d].$$

Then

$$\Pr^n \left\{ \frac{\mathbf{X}_1}{b(n)} \leq x_1, ..., \frac{\mathbf{X}_n}{b(n)} \leq x_d \right\} \tag{1}$$

is the probability that P_n places no points in the set A. By Poisson limit theorem, this probability tends to $e^{-\nu(A)}$ as $n \to \infty$. Therefore, the limit of (1) is

$$G(\mathbf{x}) \ = \ \exp\{-V(\mathbf{x})\} \tag{2}$$

where $V(\mathbf{x}) = \nu(A)$.

Moreover, using the radial-spectral decomposition of ν,

$$V(\mathbf{x}) \ = \ \int_{\mathcal{S}_d} \max_{j=1,...,d} \left(\frac{w_j}{x_j} \right) dH(w). \tag{3}$$

The function $V(\mathbf{x})$ is called the *exponent measure* and formula (3) is the *Pickands representation*. If we fix $j' \in \{1, ..., d\}$ with $0 < x_{j'} < \infty$, and define $x_j = +\infty$ for $j \neq j'$, then

$$\begin{aligned} V(\mathbf{x}) &= \int_{\mathcal{S}_d} \max_{j=1,...,d} \left(\frac{w_j}{x_j} \right) dH(w) \\ &= \frac{1}{x_{j'}} \int_{\mathcal{S}_d} w_{j'} dH(w) \end{aligned}$$

so we must have

$$\int_{\mathcal{S}_d} w_j dH(w) = 1, \quad j = 1, ..., d, \tag{4}$$

to ensure that the marginal distributions are correct.

Note that

$$kV(\mathbf{x}) \;=\; V\left(\frac{\mathbf{x}}{k}\right)$$

(which is in fact another characterization of V) so

$$\begin{aligned} G^k(\mathbf{x}) &= \exp(-kV(\mathbf{x})) \\ &= \exp\left(-V\left(\frac{\mathbf{x}}{k}\right)\right) \\ &= G\left(\frac{\mathbf{x}}{k}\right). \end{aligned}$$

Hence G is *max-stable*. In particular, if $\mathbf{X}_1, ..., \mathbf{X}_k$ are i.i.d. from G, then max$\{\mathbf{X}_1, ..., \mathbf{X}_k\}$ (vector of componentwise maxima) has the same distribution as $k\mathbf{X}_1$.

II. EXAMPLES OF MULTIVARIATE EXTREME VALUE DISTRIBUTIONS

Logistic (Gumbel and Goldstein, 1964)

$$V(\mathbf{x}) = \left(\sum_{j=1}^{d} x_j^{-r}\right)^{1/r}, \quad r \geq 1.$$

Check:

1. $V(\mathbf{x}/k) = kV(\mathbf{x})$
2. $V((+\infty, +\infty, ..., x_j, ..., +\infty, +\infty) = x_j^{-1}$
3. $e^{-V(\mathbf{x})}$ is a valid c.d.f.

Limiting cases:

- $r = 1$: independent components
- $r \to \infty$: the limiting case when $X_{i1} = X_{i2} = ... = X_{id}$ with probability 1.

Asymmetric logistic (Tawn 1990)

$$V(\mathbf{x}) \;=\; \sum_{c\in C}\left\{\sum_{i\in c}\left(\frac{\theta_{i,c}}{x_i}\right)^{r_c}\right\}^{1/r_c},$$

where C is the class of non-empty subsets of $\{1,...,d\}$, $r_c \geq 1$, $\theta_{i,c} = 0$ if $i \notin c$, $\theta_{i,c} \geq 0$, $\sum_{c\in C}\theta_{i,c} = 1$ for each i.

Negative logistic (Joe 1989)

$$V(\mathbf{x}) \;=\; \sum\frac{1}{x_j} + \sum_{c\in C:\ |c|\geq 2}(-1)^{|c|}\left\{\sum_{i\in c}\left(\frac{\theta_{i,c}}{x_i}\right)^{r_c}\right\}^{1/r_c},$$

$r_c \leq 0$, $\theta_{i,c} = 0$ if $i \notin c$, $\theta_{i,c} \geq 0$, $\sum_{c\in C}(-1)^{|c|}\theta_{i,c} \leq 1$ for each i.

Bilogistic (Smith 1990 — only for $d = 2$)

Tilted Dirichlet (Coles and Tawn 1991)

A general construction: Suppose h^* is an arbitrary positive function on $\mathcal{S}_d$ with $m_j = \int_{\mathcal{S}_d} u_j h^*(\mathbf{u}) d\mathbf{u} < \infty$, then define

$$h(\mathbf{w}) \;=\; (\sum m_k w_k)^{-(d+1)} \prod_{j=1}^{d} m_j h^* \left(\frac{m_1 w_1}{\sum m_k w_k}, ..., \frac{m_d w_d}{\sum m_k w_k} \right).$$

h is density of positive measure H satisfying $\int_{\mathcal{S}_d} u_j dH(\mathbf{u}) = 1$.

As a special case of this, they considered Dirichlet density

$$h^*(\mathbf{u}) \;=\; \frac{\Gamma(\sum \alpha_j)}{\prod_j \Gamma(\alpha_j)} \prod_{j=1}^{d} u_j^{\alpha_j - 1}.$$

Leads to

$$h(\mathbf{w}) \;=\; \prod_{j=1}^{d} \frac{\alpha_j}{\Gamma(\alpha_j)} \cdot \frac{\Gamma(\sum \alpha_j + 1)}{(\sum \alpha_j w_j)^{d+1}} \prod_{j=1}^{d} \left(\frac{\alpha_j w_j}{\sum \alpha_k w_k} \right)^{\alpha_j - 1}.$$

Disadvantage: need for numerical integration

III. ESTIMATION

Coles and Tawn (1991):

1. Fix thresholds $u_1, ..., u_d$.

2. Transform margins to unit Fréchet. Typically this involves fitting a GPD to $Y_j > u_j$, an empirical CDF to $Y_j \leq u_j$, and applying the probability integral transformation.

3. Likelihood based on the Poisson process approximation on $([0, u_1] \times ... \times [0, u_d])^c$.

Joe, Smith and Weissman (1992) proposed a somewhat similar method.

Smith (1994) and Smith, Tawn and Coles (1997) proposed a more direct threshold approach. Suppose the raw data points are represented by vectors $(Y_{i1},...,Y_{id})$, $i = 1,...,n$. However in the spirit of threshold methods we replace Y_{ij} by (δ_{ij}, X_{ij}) where $\delta_{ij} = I(Y_{ij} > u_j)$, $X_{ij} = \delta_{ij}(Y_{ij} - u_j)$. We use the limiting multivariate EVT to propose an approximation to $F(y_1,...,y_d)$ when $y_1 > u_1,...,y_d > u_d$. This allows us to calculate the contribution to the likelihood from all $(Y_{i1},...,Y_{id})$ for which $Y_{i1} > u_1,...,Y_{id} > u_d$. All other cases (observations $\mathbf{Y}_i$ where some Y_{ij} are above the threshold and others are below) are approximated by adding and subtracting terms based on $y_1 \geq u_1,...,y_d \geq u_d$.

Ledford and Tawn (1996) proposed an alternative approximation which is more cumbersome but performs better.

More recent authors have thought about the problem directly in terms of *multivariate Generalized Pareto distributions*, see e.g. Rootzén and Tajvidi, *Bernoulli* **12** 917–930 (2006).

IV. DEPENDENCE MEASURES

The first paper to suggest that multivariate extreme value theory (as defined so far) might not be general enough was Ledford and Tawn (1996).

Suppose (Z_1, Z_2) are a bivariate random vector with unit Fréchet margins. Traditional cases lead to

$$\Pr\{Z_1 > r,\ Z_2 > r\} \sim \begin{cases} r^{-1} & \text{dependent cases} \\ r^{-2} & \text{exact independent} \end{cases}$$

They showed by example that for a number of cases of practical interest,

$$\Pr\{Z_1 > r,\ Z_2 > r\} \sim \mathcal{L}(r) r^{-1/\eta},$$

where $\mathcal{L}$ is a slowly varying function and $\eta \in \left(\frac{1}{2}, 1\right)$.

Estimation: used fact that $1/\eta$ is Pareto index for $\min(Z_1, Z_2$.

More general case (Ledford and Tawn 1997):

$$\Pr\{Z_1 > z_1, Z_2 > z_2,\} \;=\; \mathcal{L}(z_1, z_2) z_1^{-c_1} z_2^{-c_2},$$

$0 < \eta \le 1$; $c_1 + c_2 = \frac{1}{\eta}$; $\mathcal{L}$ slowly varying in sense that

$$g(z_1, z_2) \;=\; \lim_{t \to \infty} \frac{\mathcal{L}(tz_1, tz_2)}{\mathcal{L}(t, t)}.$$

They showed $g(z_1, z_2) = g_* \left(\frac{z_1}{z_1 + z_2} \right)$ but were unable to estimate g_* directly — needed to make parametric assumptions about this.

More recently, Resnick and co-authors were able to make a more rigorous mathematical theory using concept of *hidden regular variation* (see e.g. Resnick 2002, Maulik and Resnick 2005, Heffernan and Resnick 2005; see also Section 9.4 of Resnick (2006)).

The latest?

Heffernan, Tawn and Zhang (*Extremes*, 2007) have proposed an approximation based on moving maxima processes that incorporates these dependence measures.

$$Y_{i,d} = \max_{\ell} \max_{k} a_{\ell,k,d}^{-1} W_{\ell,i-k}$$

with $\{W_{\ell,i-k}\}$ independent GEV.

The representation generalizes earlier work by Smith and Weissman (1996), Zhang (and Smith) (2001.......).

The representation is quite general (not restricted to $d = 2$).

But estimation is much harder for these types of processes (MLE doesn't work).

Open question to find better estimation methods for these models, or better models!