Multivariate Extremes

Classical formulation:

$$\underset{\sim}{Y}_i = (Y_{i1}, \ldots, Y_{ip}) \quad i = 1, 2, \ldots \text{ IID } p\text{-vectors}$$

$$M_{n,j} = \max_{1 \le i \le n} Y_{ij} \quad j = 1, \ldots, p$$

$$\underset{\sim}{M}_n = (M_{n,1}, \ldots, M_{n,p})$$

Find multivariate G, constants $\underset{\sim}{a}_n$, $\underset{\sim}{b}_n$ s.t.

$$\frac{\underset{\sim}{M}_n - \underset{\sim}{b}_n}{\underset{\sim}{a}_n} \xrightarrow{d} G$$

meaning

$$P\left\{ \frac{M_{n,j} - b_{n,j}}{a_{n,j}} \le x_j, \, j = 1, \ldots, p \right\} \to G(x_1, \ldots, x_p) \quad (*)$$

$$\forall \, \underset{\sim}{x} = (x_1, \ldots, x_p) \in \mathbb{R}^p$$

Side comment: if $(*)$ holds, each marginal distribution of G is GEV.

Special case of Sklar's Theorem

Sklar (1959): if G is any multivariable CDF then we can write

$$G(x_1,\dots,x_p) = D_G(G_1(x_1),\dots, G_p(x_p))$$

where G_j is marginal CDF of x_j and D_G is copula.

For MEVD: necessary & sufficient that

$$D_G(u_1,\dots,u_p) = D_G^k(u_1^{1/k},\dots,u_p^{1/k}), \text{ any } k \geq 1$$

equivalent to:

$$G^k(\underset{\sim}{x}) = G(\underset{\sim}{A}_k \underset{\sim}{x} + \underset{\sim}{B}_k)$$

characterizes all possible MEVDs.

Early literature: Geffroy (1958), Tiago de Oliveria (1958), Sibuya (1960), Gumbel (1960+ ...)

Later: Pickands (1976), de Haan & Resnick (1977), Deheuvels (1978) .. general representations

STOR 834 p-89

Pichands representation

Pichands: recast as min-stable, $P(X_j > x) = e^{-x}$

Define $S(x_1, \dots, x_p) = P(X_1 > x_1, \dots, X_p > x_p)$

such that

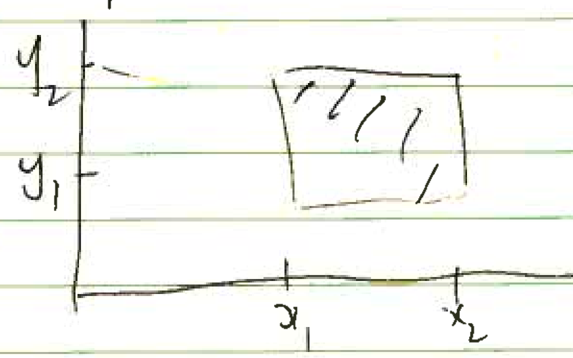
(i) $S(0, \dots, 0, x, 0, \dots, 0) = e^{-x}$ for $j = 1, \dots, p$

x in j'th position

(ii) $S^a(x_1, \dots, x_p) = S(ax_1, \dots, ax_p)$
any $a > 0$

(iii) Non-neg condition: must given non-neg mass to any subset of $\mathbb{R}^p$

eg $p=2$:



$$S(x_1, y_1) - S(x_2, y_1) - S(x_1, y_2) + S(x_2, y_2) \geq 0$$

As an example, try

$$X_j = \min_{i=1,\dots,q} \frac{Z_i}{c_{ij}}, \quad j=1,\dots,p$$

where $Z_1,\dots,Z_q$ ind. exp $\left(P(Z_i > z) = e^{-z} \ \forall i \ \forall z \geq 0,\right.$

$$0 \leq c_{ij} \leq 1, \quad \sum_i c_{ij} = 1 \ \forall j$$

leads to

$$P\{X_j > x_j \ \forall j\} = \exp\left[-\sum_{i=1}^{q} \max_{1\leq j\leq p} c_{ij} x_j\right]$$

of form $S(x_1,\dots,x_p) = \exp\left[-\int_{S_p} \max_j (w_j x_j) dH(\underset{\sim}{w})\right]$ (A)

where $S_p = \{\underset{\sim}{w} = (w_1,\dots,w_p),\ w_j \geq 0,\ \sum_{j=1}^{p} w_j = 1\}$

and H is a (discrete) non-neg measure on S_p

satisfying $\int_{S_p} w_j \, dH(\underset{\sim}{w}) = 1$ for $j=1,\dots,p$ (B)

Pickands rep: any MV function S satisfying (A) and (B) is the survivor function of a MEVD.

Alternative representation:

$$S(x_1,\dots,x_p) = \exp\left[-\sum x_j\, A\left(\frac{x_1}{\sum x_j},\dots,\frac{x_p}{\sum x_j}\right)\right]$$

A called "dependence function"

$p=2$: ~~$S(x,y)$~~ $S(x,y) = \exp\left[-(x+y)\,A\left(\frac{y}{x+y}\right)\right]$

A convex on $[0,1]$, $A(0)=A(1)=1$, $A(\frac{1}{2})\geq\frac{1}{2}$.

Estimation: skip to parametric cases.
Assume A or S is differentiable

[p. 113 of text]

Mixed Model $\quad A(w) = \theta w^2 - \theta w + 1 \qquad 0\leq\theta\leq 1$

$$S(x,y) = \exp\left[-x-y+\frac{\theta xy}{x+y}\right]$$

Logistic $\quad A(w) = \{(1-w)^r + w^r\}^{1/r}, \quad r\geq 1$

Hüsler-Reiss: $A(w) = (1-w)\Phi\left(\frac{a}{2}+\frac{1}{a}\log\frac{1-w}{w}\right) + w\,\Phi\left(\frac{a}{2}+\frac{1}{a}\log\frac{w}{1-w}\right)$

$0\leq a\leq\infty$, Φ is standard normal cdf.

Tilted Dirichlet see p. 115-116 of text

Asymptotic Results: MLE is usually regular on interior of parameter space but may have non-standard properties on the boundary. (Tawn '88, '90)

Example: test $r=1$ in logistic model

$$G(x,y;r) = 1-\exp\{-(x^r+y^r)^{1/r}\} \quad r\geq 1$$

Density $g(x,y;r) = \dfrac{\partial^2 G}{\partial x \partial y}$

Obs. (x_i, y_i), $i=1,\dots,n$

Let $\ell_n(r) = \sum_{i=1}^{n} \log g(x_i, y_i; r)$

$U_n(r) = \dfrac{d}{dr}\ell_n(r)$ (score statistic)

When $r=1$, $U_n(1) = \sum_1^n u(x_i, y_i)$

where $u(x,y) = \log(xy) + (x+y-2)\log(x+y) - x\log x - y\log y + (x+y)^{-1}$.

Check: $E\,u(X,Y)=0$, var $=+\infty$ when (X,Y) indep. unit exponential

We can show $\sqrt{\frac{2}{n\log n}}\ U_n(1) \xrightarrow{d} N[0,1]$

Create a test from this.

Many similar results in the literature.

Threshold method : Coles & Tawn (1991)

Convenient to transform to unit Fréchet

Assume IID sequence $\underset{\sim}{X}_i = (X_{i1},\dots, X_{ip})$

$$P(X_{ij} \le x) = e^{-1/x}, \quad 0<x<\infty$$

Point process $P_n = \{\frac{1}{n}\underset{\sim}{X}_i,\ i=1,\dots,n\}$

Convergence: $P_n \xrightarrow{d} P$ point process on $\mathbb{R}_+^p \setminus \{0\}$

Alternatively: $r_i = \frac{1}{n}\sum_{j=1}^{p} X_{ij},\ w_{ij} = \frac{X_{ij}}{nr_i}$

$$\underset{\sim}{w}_i = (w_{i1}\cdots w_{ip}) \in S_p$$

Intensity of P is $\mu(dr \times d\underset{\sim}{w}) = \frac{dr}{r}\cdot dH(\underset{\sim}{w})$

H a measure on S_p with $\int_{S_p} w_j\, dH(\underset{\sim}{w}) = 1,\ j=1,\dots,p$

If $M_{nj} = \max(X_{1j}, \ldots, X_{nj})$, $j = 1, \ldots, p$ then

$$P\{M_{nj} \le n x_j,\ j=1,\ldots,p\} \to G(\underset{\sim}{x}) = e^{-V(\underset{\sim}{x})}$$

$$V(\underset{\sim}{x}) = \int_{S_p} \max_{j=1,\ldots,p} \left(\frac{w_j}{x_j}\right) dH(\underset{\sim}{w})$$

V is "exponent measure"

Estimation: Two approaches to parametric estimation

(a) original Coles-Tawn (1991) approach

Uses point process representation directly.

(b) censored data approach

Based on Prop. 5.15 of Resnick
Leads to explicit formulas for $F(x_1, \ldots, x_p)$
when each $x_j >$ some high u_j

Formulas (4.40), (4.41) p.124

Construct "censored data likelihood" from this