Detection of the Conditional Extreme Value Model

Sidney Resnick
School of Operations Research and Information Engineering
Rhodes Hall, Cornell University
Ithaca NY 14853 USA

http://legacy.orie.cornell.edu/∼sid
sir1@cornell.edu

October 12, 2008

Work with: J. Heffernan, B. Das (Cornell)

1. Introduction

- The conditional (multivariate) extreme value model (CEV).
  - What is it?

    Short, slightly crude answer (more later): (X, Y) satisfy a conditional extreme value model if

    * Y is in a domain of attraction of an extreme value distribution and
    * $\exists\, \alpha(t) > 0,\ \beta(t) \in \mathbb{R}$ such that

      $$P\Big[\frac{X-\beta(t)}{\alpha(t)} \in \cdot \,\Big|\, Y > t\Big] \Rightarrow H(\cdot),$$

      for a non-degenerate distribution H. Given Y is large, the distribution of X is approximately the type of H.
  - How is it positioned vis a vis usual theory? What is its relationship to usual multivariate EVT and theory of multivariate regular variation.
  - Is it really applicable? (Early days. Hopefully. Maybe...yea.)

- Can we detect when the model is appropriate and plausible for a data set? (We think so and this is promising. See Hillish, Pickandsish, Kendall's tau plots later.)

- Problems with traditional multivariate EVT.
  - Usual formulation of multivariate EVT has the observation vectors $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ iid random vectors in $\mathbb{R}^d$ and each component of the d-dimensional vector $\boldsymbol{X}_i$ should be in a one dimensional domain of attraction.

    May not be true. See QQ plot later.
  - Even if traditional theory's assumptions satisfied, may have asymptotic independence which hinders sensible estimates.

- So CEV model applicable if either
  - Not all components of a vector are in a domain of attraction.
  - Multivariate EVT applies but asymptotic independence prevents sensible estimates of the probability of risk events and a supplementary assumption of CEV useful.

2. Background: Regular variation on cones.

Regular variation is a unifying idea providing a common framework for several theories.

Suppose $\mathbb{CONE}$ is a cone centered at $\mathbf{0}$:

$$\mathbf{x} \in \mathbb{CONE} \quad \Rightarrow t\mathbf{x} \in \mathbb{CONE}, \qquad t > 0.$$

Suppose $\boldsymbol{Z}^*$ is a random vector. $\boldsymbol{Z}^*$ has a regularly varying distribution in standard form on $\mathbb{CONE}$ if

$$tP\left[\frac{\boldsymbol{Z}^*}{t} \in \cdot\right] \xrightarrow{v} \nu^*(\cdot), \qquad \text{in } M_+(\mathbb{CONE}).$$

Here $M_+(\mathbb{CONE})$ all Radon non-negative measures on $\mathbb{CONE}$. A Radon measure is finite on compact sets.

2.1. Different cones $\Rightarrow$ different theories.

CONE	Application
$\mathbb{E} = [\mathbf{0}, \boldsymbol{\infty}] \setminus \{\mathbf{0}\}$	multivariate extreme value theory
$\mathbb{E}_0 = (\mathbf{0}, \boldsymbol{\infty}]$	hidden regular variation, coefficient of tail dependence;
$\mathbb{E}_\sqcap = [0, \infty] \times (0, \infty]$	Conditioned limit theorems when one component is extreme.
$[-\boldsymbol{\infty}, \boldsymbol{\infty}] \setminus \{\mathbf{0}\}$	weak conv to stable laws

Table 1: Theories stemming from standard multivariate regular variation on different cones.

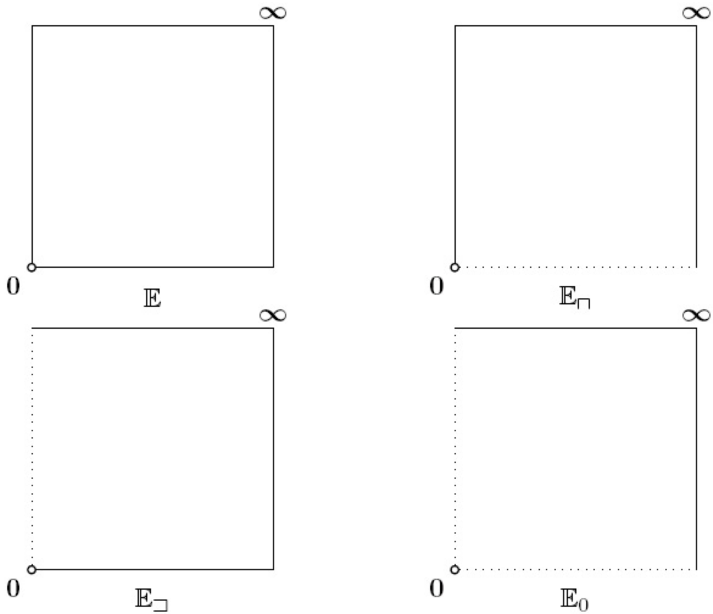

FIGURE 1. The different cones in 2-dimensions

The different cones have different compacta and hence *Radon* means something different on each space.

2.2. Consequences of the regular variation definition.

Recall the definition of standard regular variation:

$$tP\left[\frac{\boldsymbol{Z}^*}{t} \in \cdot\,\right] \xrightarrow{v} \nu^*(\cdot), \qquad \text{in } M_+(\mathbb{CONE}).$$

- For our cones, a scaling argument $\Rightarrow$

$$\nu^*(t\cdot) = t^{-1}\nu^*(\cdot), \quad t > 0.$$

- Translated scaling property via polar coordinates transform:

$$\mathbf{x} \overset{T}{\mapsto} \left(\|\mathbf{x}\|, \frac{\mathbf{x}}{\|\mathbf{x}\|}\right)$$

and we get

$$\nu^* \circ T^{-1} = \nu_1 \times S(\cdot)$$

where

$\nu_1(r, \infty] = r^{-1}, \quad r > 0,$
S is a measure on the unit sphere intersection $\mathbb{CONE}$.

Depending on the cone, S is finite (pm) or not.

3. Warmup: EVT, the cone $\mathbb{E} = [0,\infty] \setminus \{0\}$ and regular variation.

3.1. Formulate the domain of attraction problem in multivariate EVT

Central issue in multivariate EVT: Given $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ iid random vectors in $\mathbb{R}^d$ with common distribution F.

3.1.1. Problems:

- When do there exist

$$\boldsymbol{a}(n) = (a^{(1)}(n), \dots, a^{(d)}(n)) \in \mathbb{R}^d_+, \quad \boldsymbol{b}(n) = (b^{(1)}(n), \dots, b^{(d)}(n)) \in \mathbb{R}^d,$$

  and a probability distribution G such that [DOA]

$$P[(\bigvee_{j=1}^{n} \boldsymbol{X}_j - \mathbf{b}(n))/\boldsymbol{a}(n) \le \mathbf{x}] = F^n(\boldsymbol{a}(n)\mathbf{x} + \mathbf{b}(n))$$

$$= P[(\bigvee_{j=1}^{n} X_j^{(i)} - b^{(i)}(n))/a^{(i)}(n) \le x^{(i)}; i = 1, \dots, d] \to G(\mathbf{x})? \tag{1}$$

- What is the family of possible limits G?

- For a given G how do you characterize $\boldsymbol{a}(n)$ and $\boldsymbol{b}(n)$?
- For a given G in the family of possible limits, what properties does F have to satisfy in order for (1) to hold.

  If (1) holds, we say F is in the domain of attraction of G and write $F \in MDA(G)$.
- The important message: $\boldsymbol{X}$ is in the domain of attraction of the multivariate EV distribution $G(\mathbf{x})$ iff $\exists$ monotone transformations $b^{(i)}(t)$, $i = 1, \dots, d$ (satisfying limiting properties) such that

$$\boldsymbol{Z}^* = \left((b^{(i)})^{\leftarrow}(X^{(i)}),\ i = 1, \dots, d\right)$$

is standard regularly varying on $\mathbb{E} = [\mathbf{0}, \boldsymbol{\infty}] \setminus \{\mathbf{0}\}$. Thus

$$\boldsymbol{X} = \left(b^{(i)}(Z^{*(i)}),\ i = 1, \dots, d\right).$$

3.1.2. Harvest some quick facts.

- From (1), we take logarithms to get

$$n\Big(1 - F\big(\boldsymbol{a}(n)\mathbf{x} + \boldsymbol{b}(n)\big)\Big) \to -\log G(\mathbf{x}), \quad (G(\mathbf{x}) > 0).$$

Re-write this as

$$nP\Big\{\Big[\frac{\boldsymbol{X}_1 - \boldsymbol{b}(n)}{\boldsymbol{a}(n)} \le \mathbf{x}\Big]^c\Big\} \to -\log G(\mathbf{x}),$$

or [DOA≡Measure]

$$nP\Big[\frac{\boldsymbol{X}_1 - \boldsymbol{b}(n)}{\boldsymbol{a}(n)} \in \cdot\Big] \xrightarrow{v} \nu(\cdot), \tag{2}$$

where

$$\nu([-\boldsymbol{\infty}, \mathbf{x}]^c) = -\log G(\mathbf{x}).$$

The measure ν is called the *exponent measure* of G or the *limit measure*, since

$$G(\mathbf{x}) = \exp\{-\nu([-\boldsymbol{\infty}, \mathbf{x}]^c)\}.$$

- Equivalently ([DOA≡POT])

$$P\left[\frac{\boldsymbol{X}_1 - \boldsymbol{b}(n)}{\boldsymbol{a}(n)} \in \cdot \,\middle|\, \cup_{i=1}^d \,[X_1^{(i)} > b^{(i)}(n)]\right] \xrightarrow{v} \nu(\cdot), \tag{3}$$

- Joint convergence implies marginal convergence (marginal [DOA]):

$$F_i^n\big(a^{(i)}(n)x^{(i)}+b^{(i)}(n)\big)\to G_i(x^{(i)}),\quad (n\to\infty).$$

  – So if we know how to find normalizing constants in one dimension, we can find them in d-dimensions.

- Finding the limiting properties of $b^{(i)}(n), a^{(i)}(n)$:

  – Take -log in (marginal [DOA]) and instead of $-\log F_i$ put $1-F_i$; take reciprocals to get

$$\frac{1}{n}\frac{1}{1-F_i}\big(a^{(i)}(n)x^{(i)}+b^{(i)}(n)\big)\to\frac{1}{-\log G_i(x)}.$$

  Invert to get

$$\frac{\left(\frac{1}{1-F_i}\right)^{\leftarrow}(ny)-b^{(i)}(n)}{a^{(i)}(n)}\to\left(\frac{1}{-\log G_i}\right)^{\leftarrow}(y).$$

  Identify

$$b^{(i)}(t)=\left(\frac{1}{1-F_i}\right)^{\leftarrow}(t),$$

  and then

$$\frac{b^{(i)}(ty)-b^{(i)}(t)}{a^{(i)}(t)}\to\left(\frac{1}{-\log G_i}\right)^{\leftarrow}(y),\qquad t\to\infty.$$

4. Standardization

Standardization is the process of marginally transforming

$$\boldsymbol{X} \mapsto \boldsymbol{Z}^*$$

so that the distribution of $\boldsymbol{Z}^*$ is standard regularly varying on a cone $\mathbb{CONE}$: For some Radon measure $\nu^*(\cdot)$

$$tP\left[\frac{\boldsymbol{Z}^*}{t} \in \cdot\;\right] \xrightarrow{v} \nu^*(\cdot), \qquad \text{in } M_+(\mathbb{CONE}).$$

For EVT,

$$\mathbb{CONE} = \mathbb{E} = [\mathbf{0}, \boldsymbol{\infty}] \setminus \{\mathbf{0}\}.$$

In general, depending on the cone, this says one or more components of $\boldsymbol{Z}^*$ are asymptotically Pareto. For the EVT case, each is asymptotically Pareto.

4.1. Theoretical advantages of standardization:

- Standardization is analogous to the copula transformation but is better suited to studying limit relations (Klüppelberg and Resnick, 2008).
- In Cartesian coordinates, the limit measure has scaling property:

$$\nu^*(t\,\cdot\,) = t^{-1}\nu^*(\cdot), \quad t > 0.$$

- The scaling in Cartesian coordinates allows transformation to polar coordinates to yield a product measure: An angular measure exists allowing characterization of limits:

$$\nu^*\{\mathbf{x} : \|\mathbf{x}\| > r, \frac{\mathbf{x}}{\|\mathbf{x}\|} \in \Lambda\} = r^{-1}S(\Lambda),$$

  for Borel subsets Λ of the unit sphere in $\mathbb{CONE}$.

- For EVT, S is a finite measure (wlog taken to be a pm) but when $\mathbb{CONE} = \mathbb{E}_{\sqcap}$, S is NOT necessarily finite.
- See de Haan and Resnick (1977), Resnick (1987), Mikosch (2005, 2006), de Haan and Ferreira (2006), Resnick (2007).

4.2. How to Standardize?

Theoretical formulations in EVT often assume the *standard case*.

- Standard case almost never happens in practice.
- A vector which is standard regularly varying has each component having the same (asymptotically equivalent) tail.

How to transform to the standard case in practice?

- In heavy tail analysis, the simplest method: Hope $1 - F_{(i)}(x) \sim x^{-\alpha_i}$ for all i and then power up.
  BUT: Must estimate α's.

- More generally, for EVT: estimate marginals (somehow; POT?) and transform using the marginals.
  BUT: Difficult to quantify the error made when estimating marginal distributions.

- Use ranks method (de Haan and de Ronde (1998), de Haan and Ferreira (2006), Huang (1992), Resnick (2007)).

  BUT: Lose independence among observations; probably lose efficiency in favor of robustness.

5. Asymptotic Independence in EVT

If (X, Y) is in a bivariate domain of attraction of a multivariate extreme value distribution $G(x, y)$,

$$tP\Big\{\Big[\frac{X-\beta(t)}{\alpha(t)} \le x, \frac{Y-b(t)}{a(t)} \le y\Big]^c\Big\} \to -\log G(x,y), \quad t \to \infty,$$

then *asymptotic independence* of (X, Y) means the above plus

$$tP\Big[\frac{X-\beta(t)}{\alpha(t)} > x, \frac{Y-b(t)}{a(t)} > y\Big] \to 0, \quad t \to \infty.$$

This says, (de Haan and Ferreira, 2006, Resnick, 1987)

- The probability of both X and Y being biggish is smallish.
- Componentwise maxima (normalized) of iid samples of (X, Y) are asymptotically distributed as a product measure.
- The limit measure concentrates on lines; for example, if $\boldsymbol{Z}^* \in \mathbb{R}_+^d$ has a standard regularly varying distribution, then asymptotic independence means

$$tP[\frac{\boldsymbol{Z}^*}{t} \in \cdot\,] \xrightarrow{v} \nu^*(\cdot)$$

where

$$\nu^*\{\mathbf{x} \in \mathbb{E} : x^{(i)} \wedge x^{(j)} > 0, \text{ some } i, j\} = 0.$$

5.1. Examples

5.1.1. Asymptotic independence with Pareto marginals:

Let $U \sim U(0,1)$ and

$$(X,Y) = \Big(\frac{1}{U}, \frac{1}{1-U}\Big).$$

Then (X,Y) has a standard regularly varying distribution which possesses asymptotic independence.

5.1.2. Multivariate normal with correlations less than 1.

If (X,Y) are normal with $\rho(X,Y) < 1$, then asy indep holds.

5.2. Conclusion.

Asymptotic independence has little to do with

- independence
- rational nomenclature.

5.3. Why asymptotic independence creates problems.

- Estimators of various parameters may behave badly under asymptotic independence; eg, estimator of the spectral measure S. Estimators may be asymptotically normal with an asymptotic variance of 0 (oops!).
- Estimators of risk probabilities given by asymptotic theory may be uninformative.

Scenario: Estimate the probability of simultaneous non-compliance.

Supppose $\boldsymbol{Z} = (Z^{(1)}, Z^{(2)})$ = concentrations of different pollutants. Environmental agencies set critical levels $\boldsymbol{t}_0 = (t_0^{(1)}, t_0^{(2)})$ which not be exceeded. Imagine simultaneous *non-compliance* creates a health hazard. Worry about

$$[\text{ health hazard }] = [\boldsymbol{Z} > \boldsymbol{t}_0] = [Z^{(j)} > t_0^{(j)};\ j = 1, 2].$$

Asymptotic independence might lead one to report an estimate one does not believe:

$$P[\text{ health hazard }] = P[\boldsymbol{Z} > \boldsymbol{t}_0] = 0.$$

6. Conditional EV model.

CEV model applicable if either

- Not all components of a vector are in a domain of attraction.
- Multivariate EVT applies but asymptotic independence prevents sensible estimates of the probability of risk events and a supplementary assumption of CEV useful.

Auckland

http://pma.nlanr.net/traces/long/auck2.html

- Packets transmitted to and from Auckland server; measure: packet size, arrival time, source & destination IP address, port numbers, Internet protocol.
- Cluster packets into e2e sessions. Definition: cluster packets with same source and destination IP address such that delay between 2 successive packets in a cluster is less than a threshold (≤ 2 seconds).
- Compute
  - $F = \#$ bytes in a session.
  - $L =$ duration of a session.
  - $R =$ average rate associated to a session; defined to be F/L.

The variable R does not appear to be in a domain of attraction and resists characterization of its distribution. Hence difficult to decide on the joint distribution of (F, L, R).

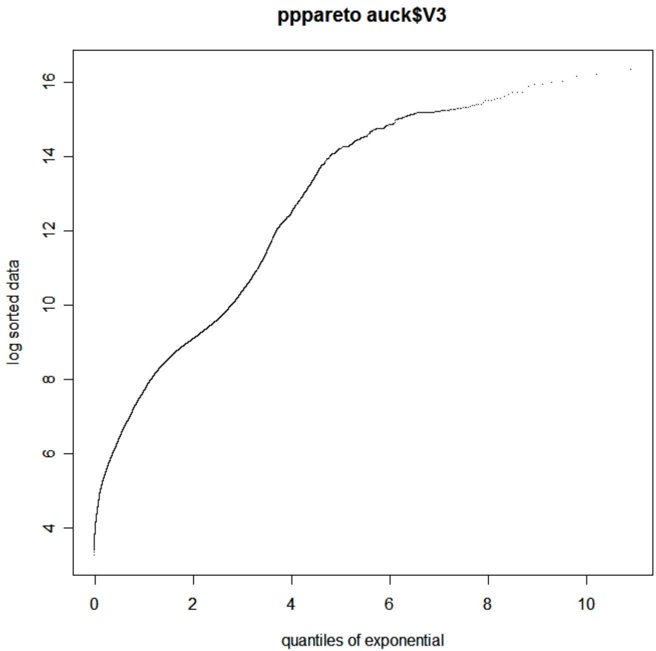

Figure 1: QQ plot for R; quantiles of exponential vs quantiles of $\log R$.

6.1. Conditional EV models; antecedents.

- Heffernan & Tawn models (Heffernan and Tawn, 2004):

$$P\Big[\frac{X-\beta(t)}{\alpha(t)} \le x \big| Y = t\Big] \to G(x), \quad t \to \infty. \tag{4}$$

- Alternate approaches to asymptotic independence consider

$$P[X \le x | Y > t] \to G(x), \quad t \to \infty,$$

which comes from

$$tP[\big(X, \frac{Y}{t}\big) \in \cdot] \to H \times \nu_1$$

where H is a pm, $\nu_1(x,\infty] = x^{-1}$, $x > 1$. See Maulik et al. (2002).

- With Jan Heffernan, meld 2 approaches (Heffernan and Resnick, 2007) and reformulate as

$$tP\Big[\Big(\frac{X-\beta(t)}{\alpha(t)}, \frac{Y-b(t)}{a(t)}\Big) \in \cdot\Big] \xrightarrow{v} \mu$$

where μ satisfies non-degeneracy assumptions. This relates to regular variation on the cone $\mathbb{E}_\sqcap$.

6.2. Basic Convergence ($d = 2$)

Given a random vector (X, Y) with

$$F_Y(x) := P[Y \leq x] \in MDA(G_\gamma),$$

and $\exists\, b(\cdot) \in \mathbb{R}, a(\cdot) > 0$ such that for some $\gamma \in \mathbb{R}$, as $t \to \infty$,

$$\left(P\Big[\frac{Y - b(t)}{a(t)} \leq x\Big]\right)^t \to G_\gamma(x) = \exp\{-(1+\gamma x)^{-1/\gamma}\}, \quad t \to \infty.$$

Further assume $\exists\, \beta(\cdot) \in \mathbb{R}, \alpha(\cdot) > 0$ and a Radon measure μ such that

$$tP\Big[\Big(\frac{X - \beta(t)}{\alpha(t)}, \frac{Y - b(t)}{a(t)}\Big) \in \cdot\Big] \xrightarrow{v} \mu(\cdot), \tag{5}$$

in $M_+\big([-\infty, \infty] \times (-\infty, \infty]\big)$, and where μ is non-null and satisfies non-degeneracy conditions: for each fixed $y \in \{x : (1+\gamma x)^{-1/\gamma} > 0\}$,

1. $\mu\big((-\infty, x] \times (y, \infty]\big)$ is not a degenerate distribution function in x;
2. $\mu\big((-\infty, x] \times (y, \infty]\big) < \infty$.

6.3. Observations:

- The Basic Convergence (5) implies the conditioned limit

$$P\Big[\frac{X-\beta(t)}{\alpha(t)}\le x\Big|Y>b(t)\Big]\to\mu\big([-\infty,x]\times(0,\infty]\big),$$

  where the limit is assumed to be a proper probability distribution in x.

- WLOG can assume Y is heavy tailed and reduce the basic convergence to a more standard form:

$$tP\Big[\Big(\frac{X-\beta(t)}{\alpha(t)},\frac{Y^*}{t}\Big)\in\cdot\Big]\xrightarrow{v}\mu^*(\cdot) \qquad (6)$$

  in $M_+([-\infty,\infty]\times(0,\infty])$ (μ^* is modified from μ). For instance,

$$Y^*=b^{\leftarrow}(Y)\quad\text{and}\quad b(t)=\Big(\frac{1}{1-F_Y}\Big)^{\leftarrow}(t).$$

- Suppose $(X, Y) \in MDA(G)$.
  - With no asymptotic independence in the EVT sense, Basic Convergence automatically holds and in this case
$$DOA \Rightarrow \text{ Basic Convergence.}$$
  - With asymptotic independence in EVT sense, Basic Convergence with the same normalizing constants fails because non-degeneracy conditions fail. BUT, Basic Convergence with different normalizing constants could still hold.

Example

Let X and Z be iid Pareto(1) random variables and define

$$Y = X^2 \wedge Z^2.$$

Then in $\mathbb{E}$ and $\mathbb{E}_\sqcap$ check convergence on representative relatively compact sets:

- In $M_+(\mathbb{E})$, asymptotic independence $\quad (\mathbb{E} = [\mathbf{0}, \boldsymbol{\infty}] \setminus \{\mathbf{0}\})$:

$$t\mathbb{P}\left[\left(\frac{X}{t}, \frac{Y}{t}\right) \in ([0,x] \times [0,y])^c\right] \to \frac{1}{x} + \frac{1}{y}, \quad x \vee y > 0.$$

- In $M_+(\mathbb{E}_\sqcap)$ $\quad (\mathbb{E}_\sqcap = [0,\infty] \times (0,\infty])$:

$$t\mathbb{P}\left[\left(\frac{X}{t^{1/2}}, \frac{Y}{t}\right) \in [0,x] \times (y,\infty]\right] \to \frac{1}{y} - \frac{1}{\sqrt{y}} \times \frac{1}{x \vee \sqrt{y}}, \quad x \geq 0, y > 0,$$

or in standard form,

$$t\mathbb{P}\left[\left(\frac{X^2}{t}, \frac{Y}{t}\right) \in [0,x] \times (y,\infty]\right] \to \frac{1}{y} - \frac{1}{\sqrt{y}} \times \frac{1}{\sqrt{x} \vee \sqrt{y}}, \quad x \geq 0, y > 0.$$

6.4. Other Consequences.

- A convergence to types argument implies variation properties of $\alpha(\cdot)$ and $\beta(\cdot)$: Suppose (X, Y^*) satisfy the condition (6). $\exists$ two functions $\psi_1(\cdot), \psi_2(\cdot)$, such that for all $c > 0$,

$$\lim_{t\to\infty} \frac{\alpha(tc)}{\alpha(t)} = \psi_1(c), \quad \lim_{t\to\infty} \frac{\beta(tc) - \beta(t)}{\alpha(t)} \to \psi_2(c). \tag{7}$$

  locally uniformly.

- This means

$$\alpha(\cdot) \in RV_\rho,\ \rho \in \mathbb{R} \quad \text{and} \quad \psi_1(c) = c^\rho,\ c > 0.$$

- $\exists$ important cases where $\psi_2 \equiv 0$ (bivariate normal). However, if $\psi_2 \not\equiv 0$, then ((Geluk and de Haan, 1987, page 16), Bingham et al. (1987))

$$\psi_2(x) = \begin{cases} k\frac{(x^\rho - 1)}{\rho}, & \text{if } \rho \neq 0,\ x > 0, \\ k \log x, & \text{if } \rho = 0,\ x > 0, \end{cases} \tag{8}$$

  for $k \neq 0$.

6.5. When can both components in the basic convergence be standardized to get regular variation on $\mathbb{E}_\sqcap$?

- Can sometimes also standardize the X variable so that

$$tP\Big[\frac{\beta^{\leftarrow}(X)}{t} \le x, \frac{Y^*}{t} > y\Big] = tP\Big[\frac{X^*}{t} \le x, \frac{Y^*}{t} > y\Big]$$
$$\to \mu^*\big([-\infty, \psi_2(x)] \times (y, \infty]\big)$$
$$= \mu^{**}\big([-\infty, x] \times (y, \infty]\big) \qquad (t \to \infty), \qquad (9)$$

giving standard regular variation on $\mathbb{E}_\sqcap$.

When?? Short version: When and only when μ^* (or μ) is not a product measure (Das and Resnick, 2008).

  - When is μ^* a product measure?
    Answer: $\mu^* = H \times \nu_1$ iff $\psi_1 \equiv 1$ ($\alpha(\cdot)$ is sv) and $\psi_2 \equiv 0$.
  - If you can standardize, how do you do it?
    * One answer: If $\beta(t) \ge 0$ and $\beta^{\leftarrow}$ is non-decreasing on the range of X, then (9) is possible (provided μ is NOT a product).
    * If the condition $\beta(t) \ge 0$ and $\beta^{\leftarrow}$ is non-decreasing fails, then a transformation of X allows one to reduce the problem to the previous case.

6.6. Form of the limit.

6.7. Case 1. Assume μ is not a product.

Then can standardize X and for the case that $\beta(t) \geq 0$ and $\beta(t) \uparrow$

$$tP\Big[\frac{\beta^{\leftarrow}(X)}{t} \leq x, \frac{Y^*}{t} > y\Big] = tP\Big[\frac{X^*}{t} \leq x, \frac{Y^*}{t} > y\Big]$$
$$\to \mu^*\big([0,\psi_2(x)] \times (y,\infty]\big) = \mu^{**}([0,x] \times (y,\infty]), \quad (t \to \infty)$$

on $[0,\infty] \times (0,\infty]$. This is standard regular variation on the cone $[0,\infty] \times (0,\infty]$ so

$$\mu^{**}(t\Lambda) = t^{-1}\mu^{**}(\Lambda), \quad t > 0.$$

$\exists$ angular measure: Let

$$\|(x,y)\| = x + y, \quad \aleph = \{(w, 1-w) : 0 \leq w < 1\}$$

and

$$\mu^{**}\{\mathbf{x} : \|\mathbf{x}\| > r, \frac{\mathbf{x}}{\|\mathbf{x}\|} \in A\} = r^{-1}S(A),$$

where S is a measure on $\aleph$ or $[0,1)$.

Warning:

$$S \text{ does not have to be finite}$$

but to guarantee

$$P[\frac{X^*}{t} \le x | Y^* > t] \to H^{**}(x) = \mu^{**}([0,x] \times (1,\infty])$$

is proper pm, need

$$\int_{[0,1)} (1-w) S(dw) = 1. \tag{10}$$

Conclusions for Case 1:

- Can write $\mu^{**}([0,x] \times (y,\infty])$ as function of S and get characterization of the class of limit measures.
- Hence, limit measures μ^{**} are in 1-1 correspondence with class of measures on $[0,1)$ satisfying (10).
- Alternatively, a scaling argument gives class of μ^{**} with following description:

$$\mu^{**}([0,x] \times (y,\infty]) = y^{-1} H^{**}(\frac{x}{y}), \quad (x,y) \in [0,\infty] \times (0,\infty]$$

for any proper prob distribution H^{**}.

6.8. Case 2: Assume μ is a product.

Any measure of the form

$$\mu^{**}([0,x] \times (y,\infty]) = y^{-1} H^{**}(x),$$

for a prob distribution H^{**} is a possible limit.

End of story for Case 2.

7. Consistency

- In practice, should one condition on X or Y?
- What if one could do either?

Then the distribution is in the domain of attraction of an EV distribution.

Conclusion: So the conditioned limit theory is only different than classical EVT if we assume we can condition only on one variable but not on the other.

8. Detecting when the model is appropriate

Estimators to help us decide if this model consistent with the data:

- Hillish (Hill-like).
- Pickandsish (suggested by the Pickands estimator of the EV index).
- Kendall's tau.

Rank based methods bypass need to estimate centering and scaling functions:
Notation:

$(X_1, Y_1), \ldots, (X_n, Y_n)$;	iid bivariate sample.
$Y_{(1)} \geq \ldots Y_{(n)}$;	order statistics of Y's in decreasing order.
X_i^*, $1 \leq i \leq n$;	X_i^* is the X-variable corresponding to $Y_{(i)}$; concomitant of $Y_{(i)}$.
R_I^k, $1 \leq i \leq k \leq n$;	Rank of X_i^* among $X_1^*, \ldots, X_k^*$; often write $R_i = R_i^k$.
$X_{1:k}^* \leq X_{2:k}^* \leq \ldots X_{2:k}^*$;	The order statistics in increasing order of $X_1^*, \ldots, X_k^*$.

Assume Basic Convergence:

$$tP\left[\left(\frac{X_1-\beta(t)}{\alpha(t)},\frac{Y_1-b(t)}{a(t)}\right)\in\cdot\right]\to\mu(\cdot),\quad t\to\infty.$$

Leads to

$$\frac{1}{k}\sum_{i=1}^{n}\epsilon_{\left(\frac{X_i-\beta(n/k)}{\alpha(n/k)},\frac{Y_i-b(n/k)}{a(n/k)}\right)}(\cdot)\Rightarrow\mu(\cdot),$$

as $n\to\infty$ and $k=k(n)\to\infty$ and $k/n\to 0$.
Assume the distribution of Y_1 is in $\mathrm{MDA}(G_\gamma)$. Scaling and weak convergence arguments yield

$$\begin{aligned}\mu_n^*\big([0,x]\times(y,\infty]\big):=&\frac{1}{k}\sum_{i=1}^{k}\epsilon_{\left(\frac{R_i}{k},\frac{k+1}{i}\right)}\big([0,x]\times(y,\infty]\big)\\ \Rightarrow&\mu^*\big((-\infty,H^{\leftarrow}(x)]\times(y,\infty]\big),\end{aligned}$$

for $0<x<1$ and $y>1$; γ is the EV index for Y and

$$H(x)=\mu\big([-\infty,x]\times(0,\infty]\big),$$

assumed to be a pm, and

$$\mu^*\big([-\infty,x]\times(y,\infty]\big)=\mu\big([-\infty,x]\times(\frac{y^\gamma-1}{\gamma},\infty]\big).$$

8.1. Hillish estimator.

Define

$$H_{k,n} = \frac{1}{k}\sum_{j=1}^{k} \log \frac{k}{R_j^k} \log \frac{k}{j}.$$

Then, as $n \to \infty$, $k \to \infty$, $k/n \to 0$,

$$H_{k,n} \xrightarrow{P} I^*$$

where

$$I^* = \int_1^\infty \int_1^\infty \mu^*\big([-\infty, H^{\leftarrow}(\frac{1}{x})] \times (y,\infty]\big) \frac{dx}{x}\frac{dy}{y}.$$

Method:

Use

$$\mu_n^*\big([0,x] \times (y,\infty]\big) \Rightarrow \mu\big([0,x] \times (y,\infty]\big)$$

and integrate to limit.

Detect product measure.

If μ is product measure then

$$H_{k,n} \xrightarrow{P} 1 = I^*,$$

and otherwise

$$H_{k,n} \xrightarrow{P} I^* \neq 1.$$

8.2. Pickandsish estimator.

Based on ratios of differences of order statistics of the concomitants. Let $0 < p < 1$.

$$R_p = \frac{X^*_{pk:k} - X^*_{pk/2:k/2}}{X^*_{pk:k} - X^*_{pk/2:k}}.$$

Then

$$R_p \xrightarrow{P} \frac{H^{\leftarrow}(p)(1-2^{\rho}) - \psi_2(2)}{H^{\leftarrow}(p) - H^{\leftarrow}(p/2)}.$$

Data example1: e2e sessions; (R,L) top and (R,F) bottom

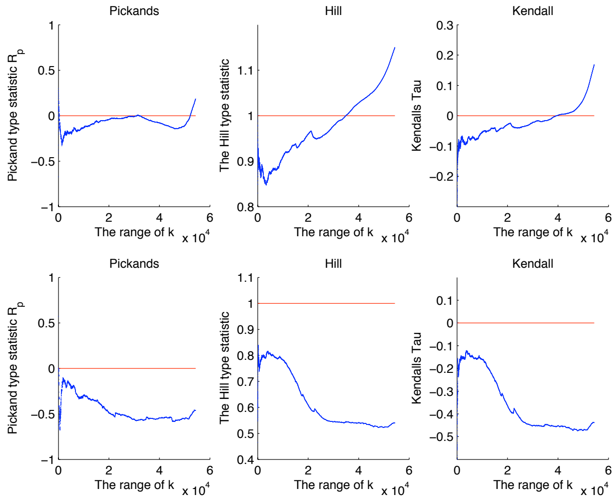

Figure 2: Pickandsish, Hillish, Kendall for (top) Auckland (R,L)–yech–and (bottom) (R,F)–not bad.

Data example 2

- Model 0: X is $N(0,1)$, Y is Pareto(1); $X \perp\!\!\!\perp Y$. Theoretically Pickandsish, $R_p = 0$, Hillish, $H = 1$, Kendalls tau, $K = 0$.
- Model 1: X and Z are Pareto(1), $X \perp\!\!\!\perp Z$, $Y = X^2 \wedge Z^2$. Theoretically Pickandsish, $R_p = -3(\sqrt{(2)} - 1)$ Hillish, $H = 0.5$.

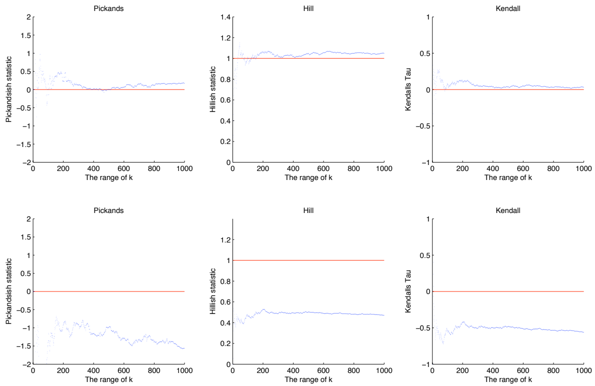

Figure 3: Pickandsish, Hillish, Kendall for (top) model0 and (bottom) model 1.

9. Final thoughts

- How practical is all this?
- When should you try to use this theory rather than EVT or to supplement EVT. Need a couple of earth shaking examples.
- It would be nice to prove the Hillish and Pickandsish estimators are asymptotically normal or else think about bootstrap CI's.
- Crucial pact with the devil: We avoided having to estimate $\alpha(\cdot)$, $\beta(\cdot)$, $a(\cdot)$, $b(\cdot)$ by switching to the rank based methods. BUT

$$H(x) = \mu\big([-\infty, x] \times (0, \infty]\big)$$

  appears in the limits and $H(x)$ is, of course, unknown. Oy!

We are thinking about all this.

Contents

References

N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular Variation*. Cambridge University Press, 1987.

B. Das and S. Resnick. Conditioning on an extreme component: Model consistency and regular variation on cones. Technical report, Cornell University, School of ORIE, 2008. http://arxiv.org/abs/0805.4373.

L. de Haan. A characterization of multidimensional extreme-value distributions. *Sankhyā Ser. A*, 40(1):85–88, 1978. ISSN 0581-572X.

L. de Haan and J. de Ronde. Sea and wind: multivariate extremes at work. *Extremes*, 1(1):7–46, 1998.

L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer-Verlag, New York, 2006.

L. de Haan and S. I. Resnick. Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 40(4):317–337, 1977.

J. L. Geluk and L. de Haan. *Regular Variation, Extensions and Tauberian Theorems*, volume 40 of *CWI Tract*. Stichting Mathematisch Centrum, Centrum voor Wiskunde en Informatica, Amsterdam, 1987. ISBN 90-6196-324-9.

J.E. Heffernan and S. Resnick. Limit laws for random vectors with an extreme component. *Ann. Appl. Probab.*, 17(2):537–571, 2007. ISSN 1050-5164. doi: 10.1214/105051606000000835.

J.E. Heffernan and J.A. Tawn. A conditional approach for multivariate extreme values (with discussion). *JRSS B*, 66(3):497–546, 2004.

Xin Huang. *Statistics of Bivariate Extreme Values*. Ph.D. thesis, Tinbergen Institute Research Series 22, Erasmus University Rotterdam, Postbus 1735, 3000DR, Rotterdam, The Netherlands, 1992.

C. Klüppelberg and S. Resnick. The Pareto copula, aggregation of risks and the emperor's socks. *Journal of Applied Probability*, 45(1): 67–84, 2008.

K. Maulik, S.I. Resnick, and H. Rootzén. Asymptotic independence and a network traffic model. *J. Appl. Probab.*, 39(4):671–699, 2002. ISSN 0021-9002.

T. Mikosch. How to model multivariate extremes if one must? *Statist. Neerlandica*, 59(3):324–338, 2005. ISSN 0039-0402.

T. Mikosch. Copulas: Tales and facts. *Extremes*, 9(1):3–20, 2006. ISSN 1386-1999 (Print) 1572-915X (Online).

S. Resnick. *Heavy Tail Phenomena: Probabilistic and Statistical Modeling*. Springer Series in Operations Research and Financial Engineering. Springer-Verlag, New York, 2007. ISBN: 0-387-24272-4.

S.I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New York, 1987.