Modeling Data Network Sessions and the Conditional Extreme Value Model

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

http://people.orie.cornell.edu/~sid
sir1@cornell.edu

October 12, 2009

MSU

Work with: Luis Lopez-Oliveros (Cornell), B. Das (Cornell, ETH), J. Heffernan

1. Introduction: Models for data networks.

When a file is sent through the Internet:

1. Divided into *packets*.
2. Headers: Packets numbered and labeled with
   - Source and destination ip addresses
   - Souce and destination port numbers
   - Packet size
   - Internet protocol (TCP?, UDP?, ...)
   - Time stamp of arrival at sniffer.
3. Stored and forwarded through routers.
4. Reassembled into the original file upon delivery.

Plausible goals:

- How to model Internet transmissions based on collected data of individual packet headers?
- Since the Internet behaves partly as a result of human stimulation, hope somewhere in this mess of data there lurk Poisson points.
- How to amalgamate packets into higher order entities which simplifiy modeling and allow fluid or continuous type models?
- What entity arrival times can plausibly be regarded as derived from Poisson?
- Teach a computer to mimic Internet sessions and hence end user Internet behavior.

2. Data sessions

Definition 1 (Sarvotham et al. (2005)) A *session* is a cluster of packets with same source and destination network addresses, such that the delay between any two succesive packets in the cluster is less than a threshold $t(=2s)$.

Other definitions possible.

2.1. Session descriptors:

For each session, compute the following descriptors:

- S : Number of bytes transmitted (size).
- D : Duration of the session.
- $R = S/D$: Average transfer rate.
- Γ: Starting time.
- For studying burstiness, some measure of peak rate.

2.2. Sample data set.

http://wand.cs.waikato.ac.nz/wits/

- TCP traffic (www, email, FTP)
- Traffic sent to a University of Auckland server on December 8, 1999, between 3 and 4 pm.
- Raw data: 1,177,497 packet headers.
- Harvest working data set of the form $\{(S_i, D_i, R_i, \Gamma_i) : 1 \leq i \leq 44,136\}$.

Originally used by Sarvotham et al. (2005) to study of sources of burstiness: Burstiness is important in order to understand congestion because of the sudden peak loads it introduces to the network; qos concerns.

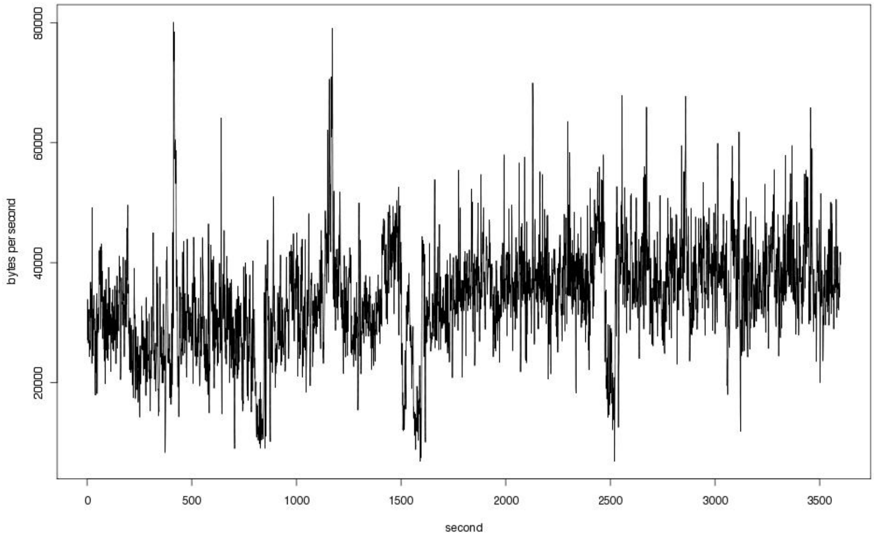


Figure 1: Bytes per time (seconds) process.

2.3. The alpha-beta split

Sarvotham et al. (2005)

Definition 2 (δ-maximum input) Divide each session in l consecutive intervals of length δ. Let

$$B_i = \#\text{ bytes transmitted over the } i\text{th subinterval}, \; i = 1, \ldots, l.$$

The δ-maximum input of a session is defined as $M_\delta = \bigvee_{i=1}^{l} B_i$.

Definition 3 (Alpha-beta split) Choose a high threshold u. A session with a δ-maximum input M_δ is called

- *alpha*, if $M_\delta \geq u$,
- *beta*, if $M_\delta < u$.

Empirical features

Sarvotham et al. (2005) found:

- Alpha-sessions are the major source of burstiness.
- In alpha-sessions: $R \perp\!\!\!\perp S$ (sort of).
- In beta-sessions: $R \perp\!\!\!\perp D$ (sort of).
- Split usually produces huge beta-group ($\approx$ tens of thousands) vs. tiny alpha-group (< 100).

Does further segmentation of the beta-group produce meaningful information?

Goals:

- Better description of dependence structure of (S, D, R) within segment
- Simulation model.

2.4. Finer segmentation

Split the data into m groups of approximately equal size according to the empirical quantiles of the burstiness predictor or covariate; we had to define a new definition of peak rate.

Will use $m = 10$ and speak of the decile groups. Split into decile groups.

Features:

- Rather than a beta-group, we have 9 groups each with the peak rate covariate in a given decile range.
- Claim the alpha-beta split masks further structure and it is informative to take into account the explicit level of the covariate.

2.5. Peak rate

Definition 4 For a session with n packets:

$$B'_i = \#\text{ bytes in } i\text{th packet},$$
$$T'_i = \text{interarrival time between } i\text{th and } (i+1)\text{th packet},$$

$i = 1, \dots, n-1$. For $k = 2, \dots, n$, the *peak rate of order k* is

$$P^{(k)} = \bigvee_{i=1}^{n-k+1} \frac{\sum_{j=i}^{i+k-1} B'_j}{\sum_{j=i}^{i+k-2} T'_j}.$$

The $P^{(k)}$ is the maximum transfer rate using only k consecutive packets. The peak rate is defined as

$$P^{\vee} = \bigvee_{k=2}^{n} P^{(k)}.$$

[Makes sense empirically but would be difficult to work with analytically.]

CORNELL
Intro
Sessions
Distn Structure
Dependence
CEV: Intro
CEV: Def
CEV↔EVT
Detection
Finale
Title Page
Page 10 of 45
Go Back
Full Screen
Close
Quit

Outline

- Divide the 44,136 sessions into 10 groups according to the deciles of $P^\vee$.
- Study the marginals of (S, D, R) in the 10 decile groups. (Heavy tails?)
- Study dependence structure of (S, D) using EVT across the decile groups.
- For our definition of peak rate, $P^\vee$, within a decile group, data sessions are initiated according to a homogeneous Poisson process.
  - Not true for other peak rate definitions of Sarvotham et al. (2005).]

3. Structure of (S, D, R).

3.1. Heavy tails

Definition 5 (Heavy tails) Call Y has *heavy tailed* if its cdf F satisfies

$$\bar{F}(y) = y^{-1/\gamma}\ell(y),$$

where ℓ is slowly varying and $\gamma > 0$.

Quickie summary:

- (S, D) appear to be jointly heavy tailed in each decile group.
- R is only heavy tailed for the highest decile group;
- R does *not* appear to be even in a domain of attraction for any of the 9 lower decile groups.

3.2. Estimation of γ's

Definition 6 (Hill estimator) Let $\{X_1, \ldots, X_n\}$ be iid (or stationary + mixing) with order statistics

$$X_{1:n} \leq \cdots \leq X_{n:n}.$$

The *Hill estimator* of $\gamma > 0$ is

$$\hat{\gamma}_{k,n} = \frac{1}{k} \sum_{i=n-k+1}^{n} \log \frac{X_{i:n}}{X_{n-k:n}}. \tag{1}$$

Theorem 1 (Consistency of Hill) *If the distribution is heavy tailed +additional second order condition, as $k \to \infty, n \to \infty, k/n \to 0$:*

$$\sqrt{k}(\hat{\gamma}_{k,n} - \gamma) \xrightarrow{d} N(0, \gamma^2). \tag{2}$$

Equivalent to peaks over threshold method and MLE.

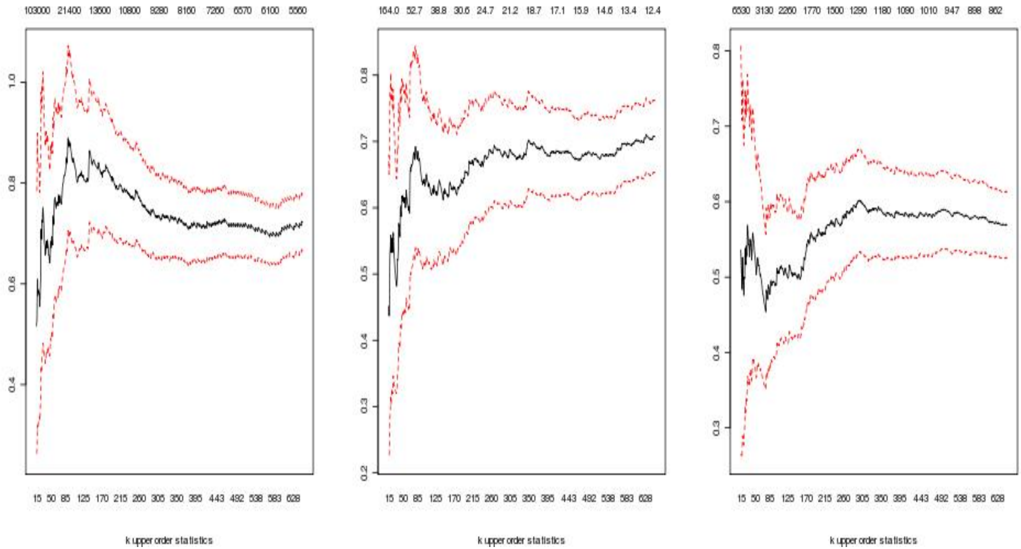


Figure 2: Size, duration and rate in the 10th decile group

decile	γ_S	s.e.	γ_D	s.e.	γ_R	s.e.
1	0.56	0.056	0.60	0.028		
2	0.55	0.061	0.47	0.023		
3	0.62	0.044	0.63	0.034		
4	0.62	0.036	0.62	0.029		
5	0.61	0.035	0.55	0.029		
6	0.69	0.040	0.55	0.028		
7	0.88	0.042	0.73	0.037		
8	0.77	0.045	0.71	0.033		
9	0.70	0.037	0.69	0.032		
10	0.73	0.034	0.68	0.032	0.58	0.027

Table 1: Summary of Hill estimates with asymptotic standard errors for the shape parameter of S, D and R.

Conclude: It appears that marginal distributions vary by decile.

4. Dependence structure of (S, D)

- Dependence structure varies by decile. Seen already in simple scatter plots.
- Assess dependence by computing *angular measures* which give favored directions for big values of (S, D).
  - Standardize the pairs to have the same tails by *ranks method*.
  - Threshold the resulting pairs and keep only those data pairs outside a large circle.
  - Convert to polar coordinates.
  - Make density plot of θ-coordinate.

Simple scatter plot.

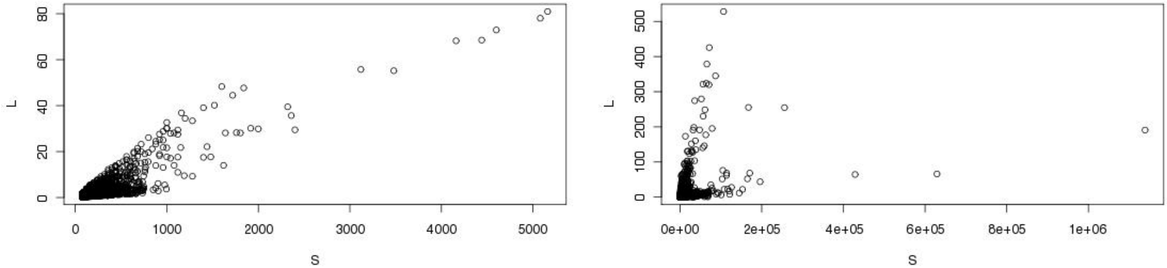


Figure 3: S vs. D for (left) 1st $P^{\vee}$ decile and (right) 6th $P^{\vee}$ decile.

Title Page
Page 17 of 45
Go Back
Full Screen
Close
Quit

Alpha-beta-like split; angular measures (S, D)

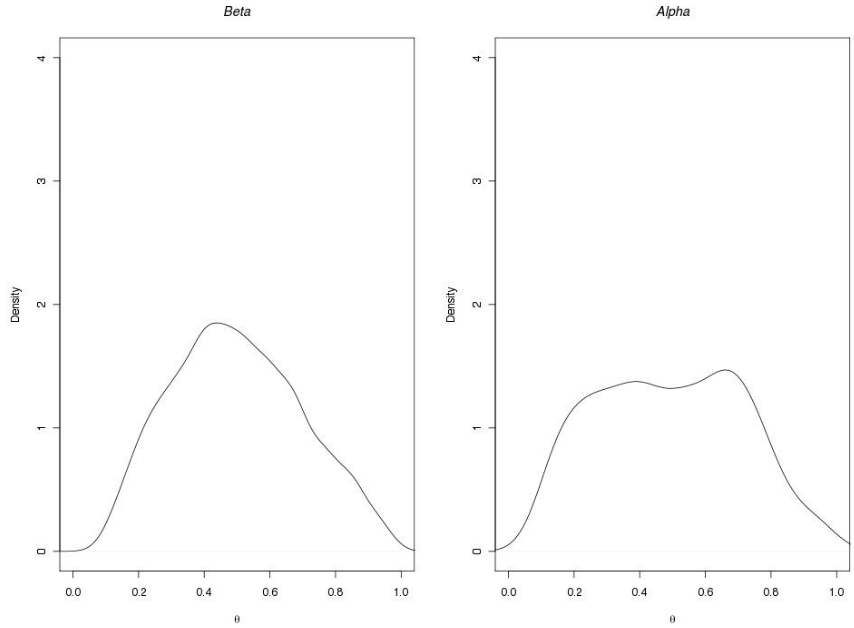


Figure 4: Non-parametric estimates of the spectral density of $\mathbb{S}$ for (left) a beta aggregate of 9 deciles and (right) an alpha group.

Finer split: plots reasonably symmetric, unimodal

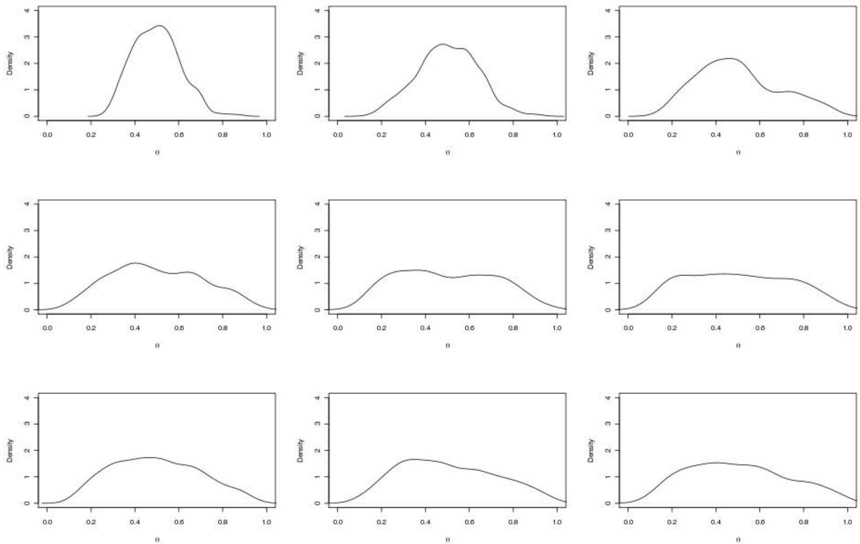


Figure 5: Non-parametric estimates of the spectral density of $\mathbb{S}$ from left to right and top to bottom: 1st to 9th decile groups.

Comments

- Seek to relate the explicit level of $P^{\vee}$ with the dependence structure of (S, D).

  Seek global model: Hope the spectral measure $\mathbb{S}$ can be approximated by some $\mathbb{S}_{\psi}$ where a (generalized linear) model links $g(\psi)$ $\sim$decile group.

- Using QQ plots and sample acf's can check within decile groups, session initiation times look Poisson. This is not true across the whole data set– only when the data is segmented by decile group; also not true with other definitions of peak rate.

4.1. Global model: Toward a parametric model for the spectral density $\mathbb{S}$

Logistic model:

$$h_\psi(t) = \frac{1}{2}\left(\frac{1}{\psi} - 1\right) t^{-1-1/\psi}(1-t)^{-1-1/\psi}[t^{-1/\psi} + (1-t)^{-1/\psi}]^{\psi-2}, \qquad (3)$$
$$= h(t), \quad 0 \le t \le 1,$$

with $\psi \in (0,1)$.

Features:

- Symmetric.
- For $\psi < 0.5$: h is unimodal and as $\psi \to 0$ we obtain perfect dependence.
- For $\psi > 0.5$: h is bimodal and as $\psi \to 1$ we obtain asymptotic independence.

This allows us to quantify the effect of $P^\vee$ on the dependence between S and D.

Parametric vs non-parametric density estimates.

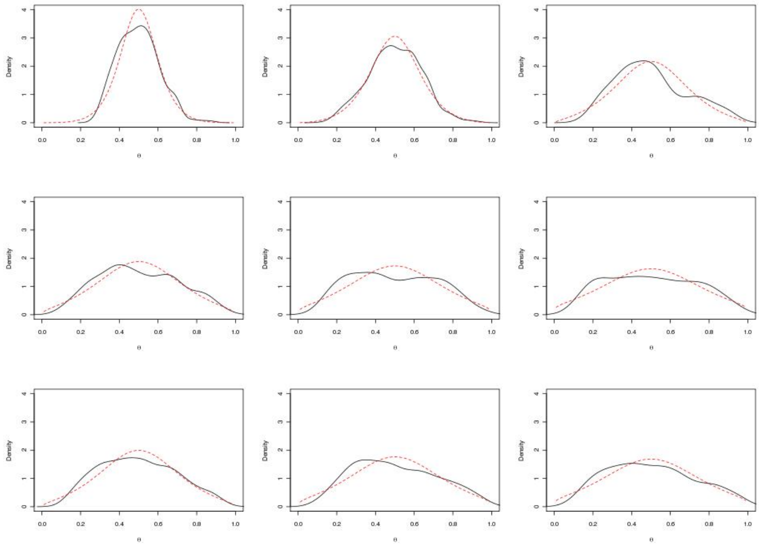


Figure 6: Parametric estimates of the spectral density $\mathbb{S}$ superimposed to non-parametric counterparts, from left to right and top to bottom: 1st to 9th decile groups.

Intro
Sessions
Distn Structure
Dependence
CEV: Intro
CEV: Def
CEV↔EVT
Detection
Finale

Dependence of (S, D) as a function of $P^{\vee}$

Fit a global trend logistic model where

$$g^{-1}(\psi) = \beta_0 + \beta_1 \log(P^{\vee}).$$

After some experimenting choose link function g

$$g(x) = \frac{1/2}{1 + e^{-x}}.$$

Sketch of simulation:

1. In the existing data set, each session has an associated $P^{\vee}$. Form the EDF. Get a bootstrap sample of $P^{\vee}$ from this EDF and divide into $m = 10$ samples according to the quantiles.
2. For each group, simulate the starting times of the sessions via homogeneous Poisson process.
3. For each $P_j^{\vee}$, compute the corresponding value of ψ_j from the GLM and use it to simulate an angle Θ_j from the logistic distribution.
4. Simulate the radius component N_j; use Pareto for the heavy tail.
5. Transform to Cartesian coordinates and then invert using fitted marginal distributions to get back to the original scale where (S_j, D_j) do not have same tails.Compute $R_j = S_j/D_j$.

What about R?

- Except for highest decile, R is not in a domain of attraction and not heavy tailed.
- Evidence $R|S$ can be modeled.
- Evidence that $R|D$ cannot be modeled.

Credit:

Lopez-Oliveros and Resnick (2009) + ideas of Jan Heffernan.

5. CEV: 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.
  - Applicable? (Cautious but firm "yes".)
  - Can we detect when the model is appropriate and plausible for a data set? (Cautious "yes". Rank transformations + graphical plotting methods: Hillish, Pickandsish, Kendall's tau plots.)

- 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 (and usually do) have asymptotic independence which hinders making sensible estimates of risk regions where several coordinates are simultaneously large.
- So CEV model may be applicable if either
  - Not all components of a vector are in a domain of attraction.
  - Multivariate EVT applies but asymptotic independence prevents helpful estimates of the probability of risk events; CEV–if appropriate–provides a supplementary assumption.

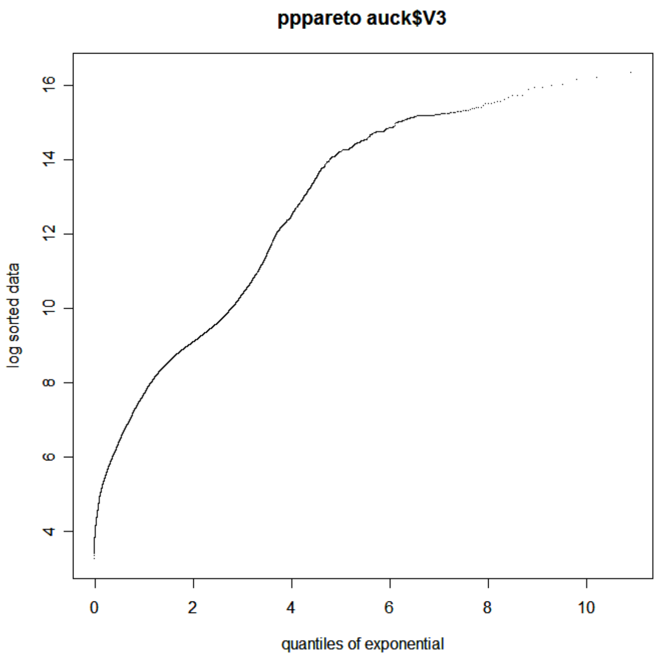


Figure 7: QQ plot for R; quantiles of exponential vs quantiles of $\log R$.

6. Conditional EV model.

CEV model may be applicable if either

- Not all components of a vector are in a domain of attraction.
- Multivariate EVT holds but asymptotic independence prevents estimates of the probability of risk events requiring two or more components be large simultaneously. A way forward is to
  - Assume more: A supplementary assumption such as
    - Hidden regular variation
    - Conditional extreme value model.
    - Check the additional assumption is statistically warranted.

6.1. CEV Model definition: Basic Convergence ($d = 2$)

Given a random vector (X, Y) with

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

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

$$tP\Big[\frac{Y - b(t)}{a(t)} \ge x\Big] \to -\log G_\gamma(x) = (1 + \gamma x)^{-1/\gamma}, \quad t \to \infty.$$

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

$$tP\Big[\Big(\frac{X - \beta(t)}{\alpha(t)}, \frac{Y - b(t)}{a(t)}\Big) \in \cdot\Big] \xrightarrow{v} \mu(\cdot), \tag{4}$$

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. $\mu\big((-\infty, x] \times (y, \infty]\big) < \infty$.

6.2. Observations:

- The Basic Convergence (4) implies the conditioned limit

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

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

- Suppose $(X, Y) \in MDA(G)$.
  - With no asymptotic independence in the EVT sense, Basic Convergence automatically holds and in this case no value added:

$$DOA + \text{No Asy Indep} \;\Rightarrow\; \text{Basic Convergence=CEV}.$$

  - With asymptotic independence in EVT sense,
    * Basic Convergence with the same EVT normalizing constants fails because non-degeneracy conditions fail.
    * BUT, Basic Convergence with different normalizing constants could still hold. [Have theoretical examples; need data examples.]

- Antecedents and related: Balkema and Embrechts (2007), Das and Resnick (2008a,b), Fougéres and Soulier (2008), Heffernan and Resnick (2007), Heffernan and Tawn (2004), Maulik et al. (2002).

Intro
Sessions
Distn Structure
Dependence
CEV: Intro
CEV: Def
CEV↔EVT
Detection
Finale

7. Connection to classical EVT and Regular variation.

Connection to classical theory of multivariate extremes comes from regular variation–unifying idea providing a common framework for several theories.

Suppose $\mathbb{CONE}$ is a cone (or just star shaped) 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

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

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

7.1. 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,

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

- Standardization analogous to copula transformation but better suited to studying limit relations (Klüppelberg and Resnick, 2008).
- Characterizations of limit relations rely on characterizations of the standardized form.
- Definition of compact set dependent on the cone.
- Data never come in standard form; to transform data to standard case use (anti)-ranks method: replace vectors by vectors of componentwise anti-ranks (de Haan and de Ronde, 1998, de Haan and Ferreira, 2006, Huang, 1992, Resnick, 2007)).

7.2. Different cones $\Rightarrow$ different theories I–table.

$\mathbb{CONE}$	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 2: Theories stemming from standard multivariate regular variation on different cones.

7.3. Different cones ⇒ different theories II–artwork.

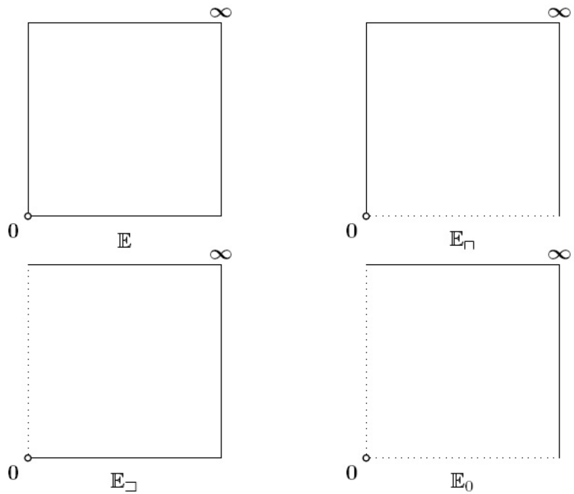


FIGURE 1. The different cones in 2-dimensions

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

7.4. Contrast EVT vs CEV

- **EVT:** Standardization to regular variation on $\mathbb{E}$ is always possible:
  - The distribution of the random vector $\boldsymbol{X}$ is in the domain of attraction of the multivariate EV distribution $G(\mathbf{x})$ iff $\exists$ monotone marginal transformations $b^{(i)}(t)$, $i = 1, \ldots, d$ (satisfying limiting properties) such that

$$\boldsymbol{Z}^* = \big((b^{(i)})^{\leftarrow}(X^{(i)}),\, i = 1, \ldots, d\big)$$

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

$$\boldsymbol{X} = \big(b^{(i)}(Z^{*(i)}),\, i = 1, \ldots, d\big).$$

- **CEV:** Suppose the basic convergence of the CEV holds.
  - Standardization is not always possible
  - It is possible iff the limit measure μ is NOT a product measure (B. Das)
  - If one can standardize, then standardization is to regular variation on $\mathbb{E}_{\sqcap}$.

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.

Note

- Rank based methods bypass need to estimate centering $(b(t),\ \beta(t))$ and scaling $(a(t),\ \alpha(t))$ functions.
- Asymptotics suggest thresholding the data according to Y's which are large and using these (X, Y)'s to make inference.
- Strong suspicion that distribution of X not in a univariate MDA but distribution of Y is in MDA helps lead to suspicion we might apply CEV.

Notation:

$(X_1, Y_1), \dots, (X_n, Y_n)$;	iid bivariate sample.
$Y_{(1)} \geq \dots Y_{(n)}$;	order statistics of Y's in decreasing order.
$Y_{(1)} \geq \dots Y_{(k)}$;	use the part of the sample corresponding to k largest Y's–corresponds to thresholding data according to large Y's.
$X_i^*,\ 1 \leq i \leq n$;	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 $X_1^*, \dots, X_k^*$; often write $R_i = R_i^k$.
$X_{1:k}^* \leq X_{2:k}^* \leq \dots X_{k:k}^*$;	The order statistics in increasing order of $X_1^*, \dots, X_k^*$.

Assume *basic convergence*:

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

This implies convergence of empirical measures:

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

as $n\to\infty$, $k=k(n)\to\infty$, $\frac{k}{n}\to 0$.
Scaling and weak convergence arguments yield ($0<x<1$, $y>1$)

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

where

- $H(x)=\mu\big([-\infty,x]\times(0,\infty]\big)$, assumed to be a pm,
- $\mu^*\big([-\infty,x]\times(y,\infty]\big)=\mu\big([-\infty,x]\times(\frac{y^\gamma-1}{\gamma},\infty]\big)$.
- γ is the EV index for Y;

Intro
Sessions
Distn Structure
Dependence
CEV: Intro
CEV: Def
CEV↔EVT
Detection
Finale

8.1. Hillish estimator.

Define

$$\mathrm{Hillish}_{k,n} = \mathrm{Hillish}_{k,n}(X,Y) := \frac{1}{k}\sum_{j=1}^{k} \log \frac{k}{R_j^k} \log \frac{k}{j}.$$

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

$$\mathrm{Hillish}_{k,n} \xrightarrow{P} I^*$$

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
$$\mu_n^*\big([0,x] \times (y,\infty]\big) \Rightarrow \mu^*\big((-\infty, H^{\leftarrow}(x)] \times (y,\infty]\big).$$
- Integrate to limit in this convergence.
- This is the same general method as used to prove Hill estimator converges.

Detect product measure.

Criterion: μ is product measure iff

$$\text{Hillish}_{k,n}(X,Y) \xrightarrow{P} 1 = I^*,$$

and

$$\text{Hillish}_{k,n}(-X,Y) \xrightarrow{P} 1 = I^*.$$

8.2. Pickandsish estimator.

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

$$\text{Pickandsish}_{p,k} = \frac{X^*_{pk:k} - X^*_{pk/2:k/2}}{X^*_{pk:k} - X^*_{pk/2:k}}.$$

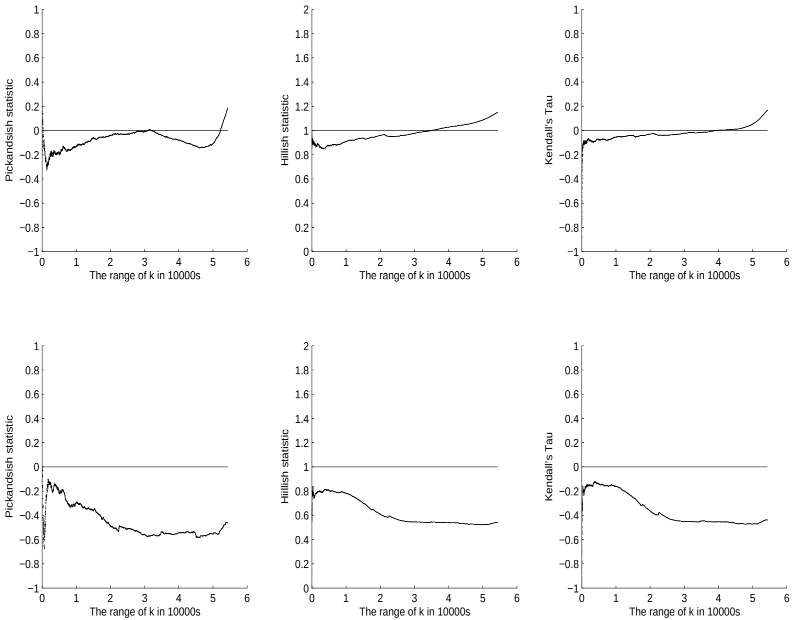
Then

$$\text{Pickandsish}_{p,k} \xrightarrow{P} \frac{H^{\leftarrow}(p)(1-2^{\rho}) - \psi_2(2)}{H^{\leftarrow}(p) - H^{\leftarrow}(p/2)}.$$

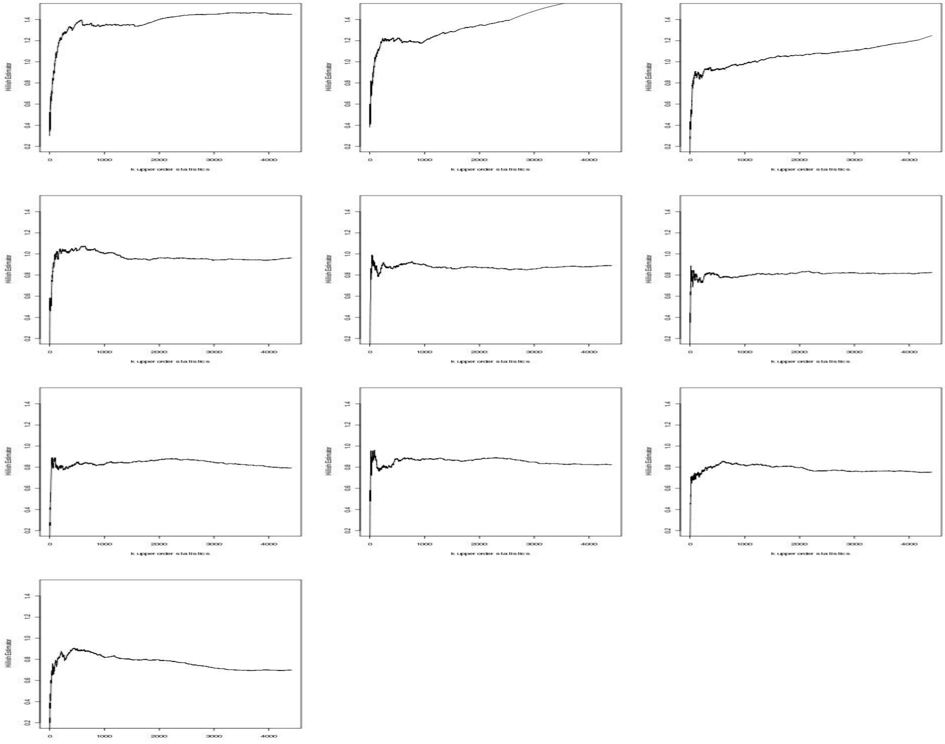
8.3. Kendall's tau

Classical Kendall's tau also converges when model holds.

Data example 1: e2e sessions; (R, D) top [yech] and (R, S) bottom [not bad]–Pickandsish, Hillish, Kendall's Tau



Hillish for (R, S) data segmented by peak rate



CORNELL
Intro
Sessions
Distn Structure
Dependence
CEV: Intro
CEV: Def
CEV↔EVT
Detection
Finale
Title Page
Page 42 of 45
Go Back
Full Screen
Close
Quit

9. Final thoughts

- Need more killer apps; especially when $(X, Y) \in DOA$ with asymptotic independence.
- It would be nice to prove Hillish and Pickandsish estimators are asymptotically normal or else think about bootstrap CI's. Must deal with dependence of the ranks.
- Crucial pact with the devil: We avoided having to estimate $\alpha(\cdot)$, $\beta(\cdot)$, $a(\cdot)$, $b(\cdot)$, γ 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.
- Made some progress in dimensions higher than two.
- Consistency issues for (S, D, R) where $S/D = R$. Currently squaring theory with empirical observation giving headaches.

We are thinking about all this.

References

AA Balkema and P. Embrechts. *High risk scenarios and extremes: a geometric approach*. European Mathematical Society, 2007.

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

B. Das and S.I. Resnick. Detecting a conditional extreme value model. Technical report, Cornell University, School of ORIE, 2008b. http://arxiv.org/abs/0902.2996; submitted to *Extremes*.

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.

A. Fougéres and P. Soulier. Modeling and estimation of conditional excesses. Technical report, Université Paris Ouest-Nanterre, 2008. http://arxiv.org/abs/0806.2426v1.

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

L. Lopez-Oliveros and S.I. Resnick. Extremal dependence analysis of network sessions. Technical report, School of ORIE, Cornell University, 2009. URL http://arxiv.org/pdf/0905.1983v1.

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.

S.I. 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. Sarvotham, R. Riedi, and R. Baraniuk. Network and user driven on-off source model for network traffic. *Computer Networks*, 48:335–350, 2005. Special Issue on *Long-range Dependent Traffic.*