Generation and Detction of Models with Multivariate Heavy Tails

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

http://people.orie.cornell.edu/∼sid
607 255 1210 sir1@cornell.edu

ACAS DC

October 20, 2014

Work with: B. Das

1. MURI team

- Cornell (Resnick, Samorodnitsky–ORIE)
- Columbia (Davis–Stat)
- University of Massachusetts (Gong–ECE, Towsley–CS)
- American University (Nolan–Math)
- Ohio State University (Shroff–ECE & CS)
- University of Illinois (Srikant–ECE)
- University of Minnesota (Zhang–CS)

2. Scientific Objectives

Goal: Develop and apply tools to models of multivariate heavy tail phenomena:

- applied probability modeling,
- statistical modeling, simulation, numerical analysis.
- control and optimization; algorithms.

Synthesize core discipline strengths:

> applied probability, statistics, simulation, numerical analysis, computer science, electrical engineering, operations research and optimization.

Apply to significant application areas:

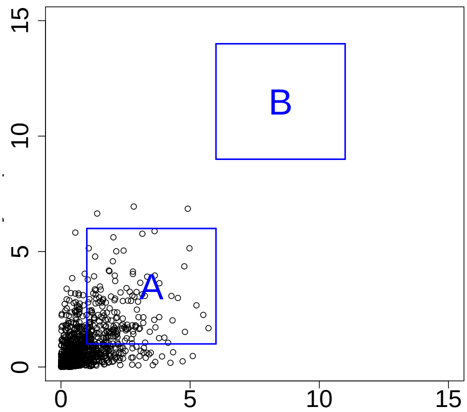
- risk estimation,
- social networks,
- cloud computing,
- scheduling and control, eg cloud computing,
- anomaly detection.

3. Heavy Tailed Phenomena

3.1. Description?

- Rough: The probability of observing large multivariate values is relatively large.

  Large usually means beyond the range of the data.



- Associated with *power laws*: In one dimension, if $X > 0$, roughly

$$P[X > x] \approx x^{-\alpha}, \quad x > x_0.$$

- Need to specify a dependence structure; correlations may not exist and are vague information.

- Generalize to higher dimensions d or even sequence or stochastic processes. If $\boldsymbol{X} = (X_1, \dots, X_d) \in \mathbb{R}_+^d$, $\boldsymbol{X}$ has a multivariate heavy tail if
  - $\exists\, b(t) \to \infty$ as $t \to \infty$, and
  - $\exists$ measure $\nu(\cdot)$, such that for nice sets A (thought of as *tail regions*,

$$tP\left[\frac{\boldsymbol{X}}{b(t)} \in A\right] \to \nu(A). \tag{HT}$$

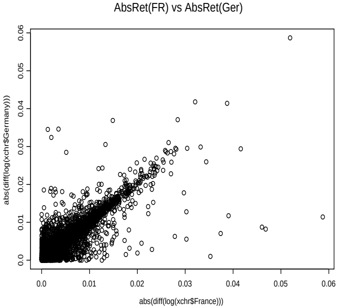
- To infer beyond the range of the data, we make the reasonably robust assumption that (HT) holds so that for tail region $\mathcal{R}$,

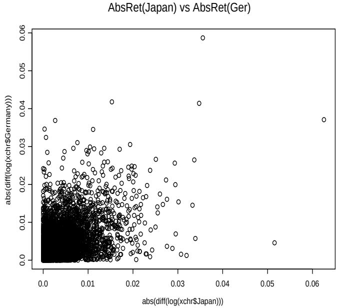
$$P[\boldsymbol{X} \in \mathcal{R}] \approx \frac{1}{t}\nu(\mathcal{R}/b(t)) \approx \frac{1}{t}\hat{\nu}(\mathcal{R}/\hat{b}(t)).$$

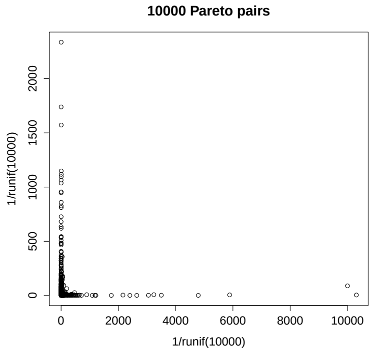
  Replacement of a converging family by the limit is *peaks over threshold* (POT) philosophy.
  - Estimates based on asymptotic methods depend on a convergence rate as a threshold gets large.
  - There could be more than one relevant asymptotic regime. Ouch!

3.2. How to model different dependence structures in heavy tailed data

- (Left) Large values occur together (strong extremal dependence
- (Middle) Large value of one variable occurs with range of values in other.
- (Right) No risk contagion or extremal dependence.







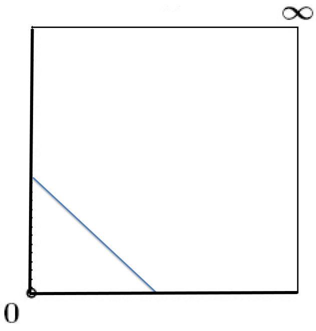
4. Model Generation

4.1. A general construction of a standardized multivariate heavy tailed distribution

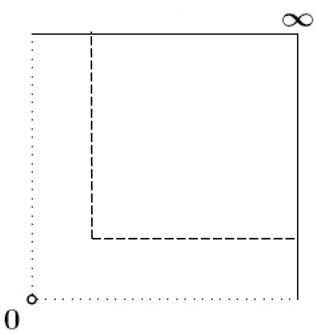
- On $\mathbb{R}_+^d$, delete a closed cone F; for example:
  - $F = \{\mathbf{0}\}$ or
  - $F =$[axes].
- Regions away from F are considered *tail regions*.
- Write
$$\aleph_F = \{\mathbf{x} : d(\mathbf{x}, F) = 1\}.$$
Take
$$\boldsymbol{\Theta} \text{ random element in } \aleph_F, \quad R \sim \text{ Pareto}, \quad \boldsymbol{\Theta} \perp\!\!\!\perp R.$$
- Set
$$\boldsymbol{X} = R\boldsymbol{\Theta}$$
and $\boldsymbol{X} \in \text{MRV}$ on $R_+^d \setminus F$.

Can apply this construction to successive choices of deleted F:

- first delete F_1 (eg, origin)
  - $\aleph_0 = [\|\mathbf{x}\| = 1]$
  - tail regions bounded away from $\mathbf{0}$.
  - $tP[\boldsymbol{X}/b(t) \in A] \to \nu(A)$ for A bounded away from $\mathbf{0}$.



- then deleting $F_1 \cup F_2$; ie, delete 2nd cone F_2 (eg, axes).
  - $\aleph_{[\text{axes}]}$ = dashed lines.
  - tail regions bounded away from axes; both components big
  - $tP[\boldsymbol{X}/b_1(t) \in A] \to \nu_1(A)$ for A bounded away axes.



4.2. Hidden regular variation

When $\boldsymbol{X}$ has regular variation on both $\mathbb{R}_+^2 \setminus \{\mathbf{0}\}$ and $\mathbb{R}_+^2 \setminus [\text{axes}]$, and

$$b(t)/b_1(t) \to \infty,$$

we say $\boldsymbol{X}$ has **hidden regular variation (HRV)**.

? How do the 2 regular variation properties interact? Statistically identifiable? Das and Resnick (2014).

4.2.1. Methods to generate models having both MRV & HRV:

- Product method described above:
  - Construct $R\boldsymbol{\Theta}$, MRV on $\mathbb{R}_+^2 \setminus [\text{axes}]$.
  - Moment conditions ensure $R\boldsymbol{\Theta}_i$ are one-dimensional regularly varying.
  - Once marginals are regularly varying, $R\boldsymbol{\Theta}$ has a multivariate distribution that is also regularly varying on $\mathbb{R}_+^2 \setminus \{\mathbf{0}\}$.
- Mixture method (Maulik and Resnick, 2005).

$$\boldsymbol{X} = B\boldsymbol{Y} + (1-B)\boldsymbol{V}, \quad B \perp\!\!\!\perp \boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{V},$$

where

  - B is a Bernoulli switching variable: $P[B=0] = P[B=1] = 1/2$.
  - $\boldsymbol{Y}$ is regularly varying on $\mathbb{R}_+^2 \setminus \{\mathbf{0}\}$.
  - $\boldsymbol{V}$ is regularly varying on $\mathbb{R}_+^2 \setminus \{[\text{axes}]\}$.

- Additive models (Weller and Cooley, 2014):

$$\boldsymbol{X} = \boldsymbol{Y} + \boldsymbol{V}, \quad \boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{V},$$

where

  - $\boldsymbol{Y}$ is MRV on $\mathbb{R}_+^2 \setminus \{\mathbf{0}\}$
  - $\boldsymbol{V}$ is MRV on $\mathbb{R}_+^2 \setminus \{[\text{axes}]\}$.

This model has severe identification issues:

  - Does HRV of $\boldsymbol{X}$ come from $\boldsymbol{Y}$ (sometimes) or $\boldsymbol{V}$ (sometimes)?
  - Is the hidden index of regular variation of $\boldsymbol{X}$ (the scaling property) what one would predict from $\boldsymbol{V}$ (not necessarily).

5. Model Detection Diagnostics

When should MRV or HRV be applied to data?

1. Reduction to one dimension:

   - $\boldsymbol{X} \in$ MRV on $\mathbb{R}_+^2 \setminus \{\mathbf{0}\}$ iff $aX_1 \vee bX_2 \in RV(\alpha)$ for all $a \geq 0$, $b \geq 0$.
   - $\boldsymbol{X} \in$ HRV on $\mathbb{R}_+^2 \setminus$[axes] iff $aX_1 \wedge bX_2 \in RV(\alpha_0)$ for $a \wedge b > 0$.

   [Hint: Cannot check $\forall a, b$.]

2. Use GPOLAR to convert to the CEV model and then use CEV diagnostics (Das and Resnick, 2011) using the *Hillish* and *Pickandsish* plots.

   - A CEV model for (ξ, η) has the form

   $$tP\left[\left(\frac{\xi}{b(t)}, \eta\right) \in \cdot\ \right] \to \mu(\cdot),$$

   on $(0, \infty) \times [0, \infty)$.

- MRV on $\mathbb{R}_+^2 \setminus \{\mathbf{0}\}$, after transformation via GPOLAR is of the form

$$tP\Big[\big(\underbrace{\|\boldsymbol{X}\|}_{\xi}/b(t), \underbrace{\boldsymbol{X}/\|\boldsymbol{X}\|}_{\eta}\big) \in \cdot\Big] \to \underbrace{\nu_\alpha \times S(\cdot)}_{\text{product measure}}, \quad \text{on } (0,\infty)\times\aleph_{\mathbf{0}}.$$

- HRV on $\mathbb{R}_+^2 \setminus [\text{axes}]$ after transformation by

$$\text{GPOLAR} : \mathbf{x} \mapsto \left(d(\mathbf{x}, \aleph_{[\text{axes}]}), \frac{\mathbf{x}}{d(\mathbf{x}, \aleph_{[\text{axes}]})}\right),$$

is of the form

$$tP\left[\left(\frac{X_1 \wedge X_2}{b_0(t)}, \frac{\boldsymbol{X}}{X_1 \wedge X_2}\right) \in \cdot\right] \to \nu_{\alpha_0}\times S_0(\cdot) \qquad \text{on } \big((0,\infty)\times\aleph_{[\text{axes}]}\big).$$

5.0.2. Hillish statistic.

Suppose $(\xi_i, \eta_i); 1 \leq i \leq n$ are iid samples in $\mathbb{R}_+^2$ and $(\xi_1, \eta_1) \in \text{CEV}(b, \mu)$. Notation:

$\xi_{(1)} \geq \ldots \geq \xi_{(n)}$	The decreasing order statistics of $\xi_1, \ldots, \xi_n$.
$\eta_i^*,\ 1 \leq i \leq n$	The η-variable corresponding to $\xi_{(i)}$, also called the concomitant of $\xi_{(i)}$.
$N_i^k = \sum_{l=i}^{k} \mathbf{1}_{\{\eta_l^* \leq \eta_i^*\}}$	Rank of η_i^* among $\eta_1^*, \ldots, \eta_k^*$. We write $N_i = N_i^k$.

Hillish statistic. For $1 \leq k \leq n$, the *Hillish statistic* is

$$\text{Hillish}_{k,n} = \text{Hillish}_{k,n}(\xi, \eta) := \frac{1}{k} \sum_{j=1}^{k} \log \frac{k}{j} \log \frac{k}{N_j^k} \qquad (1)$$

Properties (Das and Resnick, 2011): If,

- (ξ_i, η_i); $1 \leq i \leq n$ are iid observations from the $\text{CEV}(b, \mu)$;
- Mild regularity.
- $k = k(n) \to \infty$, $n \to \infty$ and $k/n \to 0$.

then

$$\text{Hillish}_{k,n} \xrightarrow{P} I_\mu = \text{ ugly integral.}$$

Moreover μ is a product measure if and only if both

$$\text{Hillish}_{k,n}(\xi, \eta) \xrightarrow{P} 1 \quad \text{and} \quad \text{Hillish}_{k,n}(\xi, -\eta) \xrightarrow{P} 1.$$

Usefulness: Detect either MRV or HRV after applying GPOLAR.

5.0.3. Example: BU data; HTTP downloads: MRV with asymptotic independence + HRV

- HTTP downloads in sessions from 1995.
- 8 hours 20 minutes worth of downloads after applying an aggregation rule to downloads to associate machine triggered actions with human requests. See Guerin, Nyberg, Perrin, Resnick, Rootzén, and Stărică (2003).
- 4161 downloads.

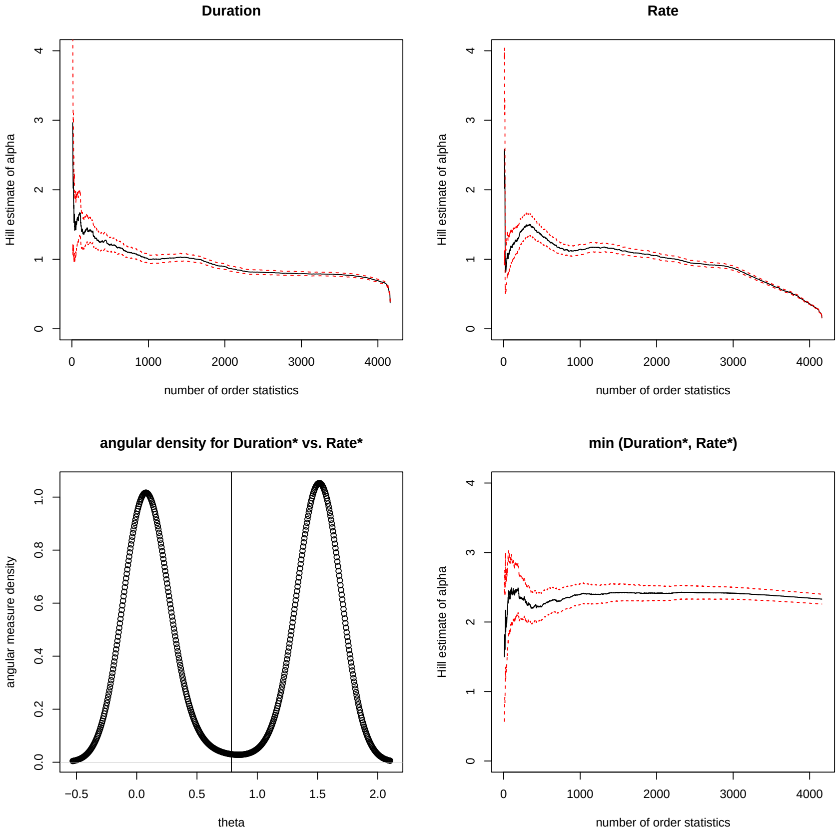
Consider the variables:

- S = the size of the download in kilobytes,
- D = the duration of the download in seconds,
- R = throughput of the download; that is, $= S/D$.

Concentrate on (D, R) and *standardize* with rank transformed variables:

$$D_i^* = \sum_{j=1}^{4161} \mathbf{1}_{\{D_i \geq D_j\}}, R_i^* = \sum_{j=1}^{4161} \mathbf{1}_{\{R_i \geq R_j\}}.$$

One dimensional analysis.



Conclusions so far:

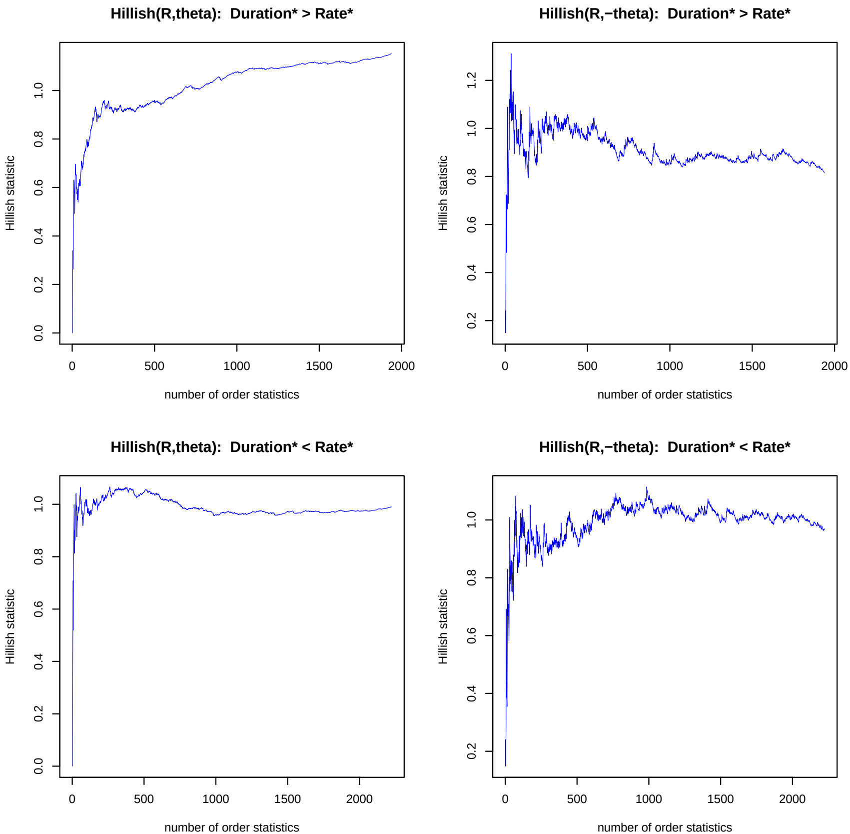
- Hill plots for marginals D^* and R^* consistent with marginal heavy tails.
- Evidence that the MRV on $\mathbb{R}_+^2 \setminus \{\mathbf{0}\}$ exists with asymptotic independence and limit measure concentrates on [axes]:
  - Spectral density plot seems to concentrate on $\{0\}$ and $\{\pi/2\}$.
  - Hill plot for $\min(D^*, R^*)$ is heavy tailed but with index

$$\alpha_0 \approx 2.4 > 1 = \text{ marginal indices}$$

    which is evidence for regular variation on $\mathbb{R}_+^2 \setminus \{[\text{axes}]\}$.

- Will Hillish confirm existence of HRV on $\mathbb{R}_+^2 \setminus \{[\text{axes}]\}$?

Hillish analysis for HRV.



6. Challenges.

- Practical?
  - Limitations of asymptotic methods: rates of convergence?
- Need for more formal inference for estimation including confidence statements.
- General HRV technique in higher dimensions requires knowing the support of the limit measure. Estimate support?
- High dimension problems? How to sift through different possible subcones? There could be a sequence of cones with regular variation on each. How to teach a computer to find the cones?
- How to go from standard to more realistic non-standard case; still some inference problems.

Contents

References

B. Das and S. Resnick. Generation and detection of multivariate regular variation and hidden regular variation. *ArXiv e-prints*, March 2014. URL http://adsabs.harvard.edu/abs/2014arXiv1403.5774D. Accepted pending revision in Stochastic Systems.

B. Das and S.I. Resnick. Detecting a conditional extreme value model. *Extremes*, 14(1):29–61, 2011.

C.A. Guerin, H. Nyberg, O. Perrin, S.I. Resnick, H. Rootzén, and C. Stărică. Empirical testing of the infinite source poisson data traffic model. *Stochastic Models*, 19(2):151–200, 2003.

K. Maulik and S.I. Resnick. Characterizations and examples of hidden regular variation. *Extremes*, 7(1):31–67, 2005.

G.B. Weller and D. Cooley. A sum characterization of hidden regular variation with likelihood inference via expectation-maximization. *Biometrika*, 101(1):17–36, 2014. ISSN 0006-3444. doi: 10.1093/biomet/ast046. URL http://dx.doi.org/10.1093/biomet/ast046.

CORNELL