Multivariate Heavy Tails, Preferential Attachment

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 sr2382@columbia.edu

MURI SpringFest April 2016, CFEM

April 13, 2016

1. Outline

- Strong dependence and hidden regular variation. (With B. Das.) John: Graphics in higher dimensions?
- The rth largest in an infinite sequence of iid random variables as a family of $\mathbb{R}^\infty$ valued stochastic processes indexed by r. What happens as $r \to \infty$? (With Ross Maller and Boris Buchmann.)
- Asymptotic normality of the number of nodes with degree counts in preferential attachment.
  - Undirected case. (with Gena)
  - Directed case. (with Tiandong)
  - Need to use AN in formal math stat techniques for model calibration.
- Relation of regular variation of measure and regular variation of density or mass function. (with Tiandong)
  - If a measure is regularly varying, is the density or mass function?
  - In dimensions more than 1, if the density or mass function is regularly varying, is the measure?
  - Application to preferential attachment.

2. Strong Dependence and HRV

2.1. Regular variation on the first quadrant.

$\boldsymbol{Z} \geq \mathbf{0}$ has a distribution which is regularly varying if

- $\exists\, b \in RV_{1/\alpha}$;
- $\exists$ Radon limit measure $\nu(\cdot)$ on $\mathbb{R}_+^2 \setminus \{\mathbf{0}\}$;
- such that as $t \to \infty$,

$$tP[\boldsymbol{Z}/b(t) \in \cdot\,] \to \nu(\cdot).$$

The limit measure always concentrates on a cone $\mathbb{C}$.

- What if $\mathbb{C} \subsetneq \mathbb{R}_+^2$?
- If $A \cap \mathbb{C} = \emptyset$, risk estimation of being in A is 0:

$$P\widehat{[\boldsymbol{Z} \in A]} \approx \frac{1}{t}\hat{\nu}(A/\hat{b}(t) = 0.$$

2.2. Strong Dependence

Consider two cases:

- Asymptotic full dependence: limit measure concentrates on diagonal.
  - Hard to find data examples.
- Asymptotic strong dependence: limit measure concentrates on a narrow wedge. Can look for 2nd regular variation property on $\mathbb{R}_+^2 \setminus$ [small wedge].

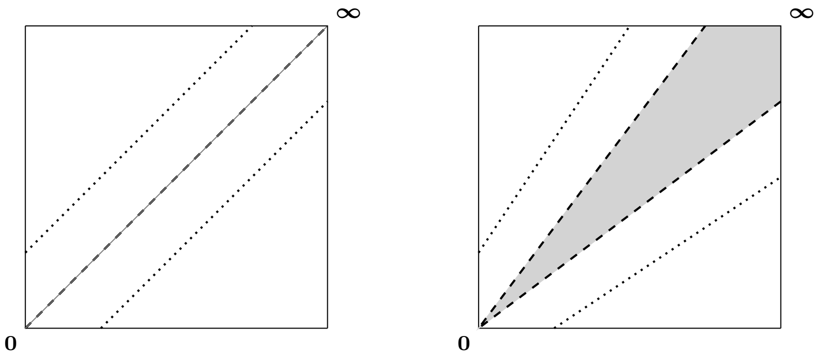


Figure 1: Left: $\mathbb{R}_+^2 \setminus \{\mathbf{0}\}$ and then $[diag]$ removed Right: $\mathbb{R}_+^2 \setminus \{\mathbf{0}\}$ and then [small wedge] is removed. The dotted lines represent the locus of points at distance one from the forbidden zone.

2.3. HRV

- When the limit measure concentrates on [small wedge], delete it from the state space.
- Look for 2nd regular variation property on $\mathbb{R}_+^2 \setminus$ [small wedge] using GPOLAR:

$$\mathrm{GPOLAR}(\mathbf{x}) = \left(d(\mathbf{x}, [\text{small wedge}]), \frac{\mathbf{x}}{d(\mathbf{x}, [\text{small wedge}])} \right).$$

- Diagnostics to find 2nd regular variation property such Hillish estimator apply.
- If [small wedge] has boundaries $y = a_l x$ and $y = a_u x$ consider the region $\{(v, w) : w - 2a_u v > x\}$; ie compute

$$P[Z_2 - 2a_u Z_1 > x],$$

ie, buy

  - 1 unit of security I_2 with risk Z_2 per unit; and
  - sell $2a_u$ units of security I_1 with risk Z_1.

2.4. (exxonr,chevronr)

- 1316 daily prices of Exxon and Chevron.
- October 10, 2001 to December 29, 2006 daily returns.
- Called (exxonr, chevronr).
- One expects strong dependence from two big companies engaged in similar activities.

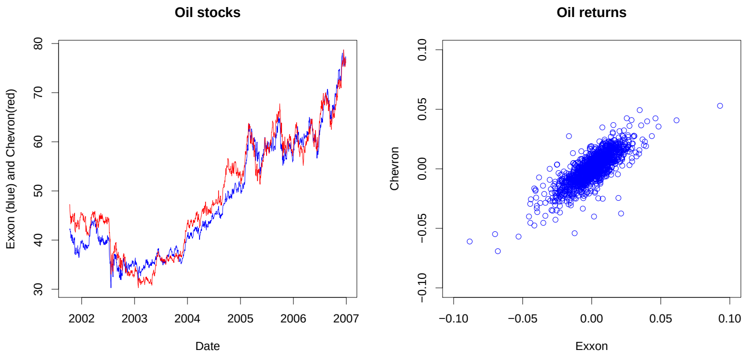


Figure 2: Stock prices and scatterplot of Chevron and Exxon returns.

2.4.1. Diamond plots

- Map (exxonr,chevronr) onto L_1 unit sphere;
- Use

$$(x, y) \mapsto \left(\frac{x}{|x| + |y|}, \frac{y}{|x| + |y|}\right) = \boldsymbol{\theta} = (\theta_1, \theta_2).$$

  from

$$\mathbb{R}^2 \mapsto \aleph_{\mathbf{0}} = [\text{diamond}] \subset \mathbb{R}^2.$$

- where the L_1 unit sphere is

$$[\text{diamond}] = \{(\theta_1, \theta_2) : |\theta_1| + |\theta_2| = 1\}.$$

- Experiment with mapping at various thresholds determined by k, the number of order statistics of the norms $|x| + |y|$.
- Use thresholds $k = 400$ and $k = 70$.
- Model for the angular measure S of limit measure ν is that S concentrates in the first and third quadrants.
- Use range of θ_1 in these quadrants as estimators. Get

  1. for the first quadrant

$$(\hat{\theta}_1, \hat{\theta}_2) = (0.312, 0.701)$$

  and

2. in the third quadrant

$$(\hat{\theta}_1, \hat{\theta}_2) = (-0.814, -0.284).$$

- These $\hat{\theta}$'s correspond to slopes of rays in Cartesian coordinates of $(\hat{a}_1, \hat{a}_2) = (0.429, 2.226)$ for the first quadrant.

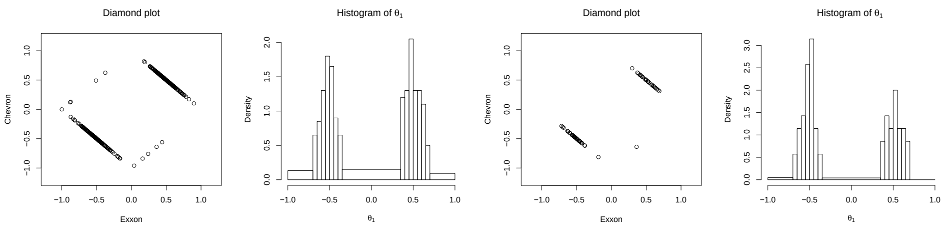


Figure 3: Empirical angles (diamond plot) for 400 largest values under L_1 norm for (exxonr,chevronr) with histogram (left two plots) and the same for 70 largest values (right two plots).

3. The rth largest of an iid sequence

- Let $\{X_n, n \geq 1\}$ be iid random variables with common distribution function $F(x)$
- Set $R(x) = -\log(1 - F(x))$, the integrated hazard function.
- Suppose F and R are continuous.
- Let $M_n^{(r)}$ be the rth largest among $X_1, \ldots, X_n$ and set

$$\boldsymbol{M}^{(r)} = \{M_n^{(r)}, n \geq r\}. \tag{1}$$

3.1. Facts

- By Ignatov's theorem (Engelen et al., 1988; Goldie and Rogers, 1984; Ignatov, 1976/77; Resnick, 2008; Stam, 1985), $\mathcal{R}_r$, the range of $\boldsymbol{M}^{(r)}$ is a sum of r independent PRM(R) processes and therefore the range of $\boldsymbol{M}^{(r)}$ is PRM(rR).
- $\mathcal{R}_r$, the range of $\boldsymbol{M}^{(r)}$, converges as a random closed set in the Fell topology to $\mathcal{R}$, the support of the measure R:

$$\mathcal{R}_r \Rightarrow \mathcal{R}, \tag{2}$$

as $r \to \infty$.

- How to get a random limit? Domain of attraction for minimum condition: Assume

$$rR(a_r x - b_r) \to g(x), \qquad (r \to \infty)$$

or equivalently

$$(\bar{F}(a_r x - b_r))^r = \exp\{-rR(a_r x - b_r)\} \to e^{-g(x)}$$

where

$$e^{-g(x)} = G_\gamma(-x)$$

and

$$G_\gamma(x) = \exp\{-(1+\gamma x)^{-1/\gamma}\},\ 1+\gamma x > 0$$

is the shape parameter family of extreme value distributions for maxima (de Haan and Ferreira, 2006; Resnick, 2008).

- Then

$$(\mathcal{R}_r + b_r)/a_r \Rightarrow PRM(m_\gamma).$$

where $m_\gamma(\cdot)$ is the measure with density

$$\frac{d}{dx}\Big(-\log G_\gamma(-x)\Big).$$

- Under the same domain of attraction condition for minima: in $\mathbb{R}^\infty$, as $r \to \infty$,

$$\frac{\boldsymbol{M}^{(r)} + b_r}{a_r} = \Big(\frac{M_{r+j}^{(r)} + b_r}{a_r}, j \geq 0\Big) \Rightarrow \Big(g_\gamma^{\leftarrow}(\Gamma_l), l \geq 1\Big),$$

  where $\{\Gamma_l, l \geq 1\}$ are the points of a homogeneous Poisson process on $\mathbb{R}_+$.

- Defining $\{\boldsymbol{M}^{(r)}, r \geq 1\}$ slightly differently yields that this family indexed by r is Markov on the space $\mathbb{R}^\infty$.

- Use?

Contents

References

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

R. Engelen, P. Tommassen, and W. Vervaat. Ignatov's theorem: a new and short proof. *J. Appl. Probab.*, Special Vol. 25A:229–236, 1988. ISSN 0021-9002. A celebration of applied probability.

C. M. Goldie and L. C. G. Rogers. The k-record processes are i.i.d. *Z. Wahrsch. Verw. Gebiete*, 67(2):197–211, 1984. ISSN 0044-3719. doi: 10.1007/BF00535268. URL http://dx.doi.org/10.1007/BF00535268.

Z. Ignatov. Ein von der Variationsreihe erzeugter Poissonscher Punktprozeß. *Annuaire Univ. Sofia Fac. Math. Méc.*, 71(2):79–94 (1986), 1976/77. ISSN 0205-0811.

S.I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer, New York, 2008. ISBN 978-0-387-75952-4. Reprint of the 1987 original.

A. J. Stam. Independent Poisson processes generated by record values and inter-record times. *Stochastic Process. Appl.*, 19(2):315–325, 1985. ISSN 0304-4149. doi: 10.1016/0304-4149(85)90033-X. URL http://dx.doi.org/10.1016/0304-4149(85)90033-X.

MURI Update

Tiandong Wang

School of ORIE, Cornell University

April 15th, 2016

Analysis of the joint mass function

Suppose $U(\cdot)$ is a measure on $\mathbb{R}^2$ with mass function $p(i,j)$:

- If $p(i,j)$ is a regularly varying array-indexed function, can it always be embedded in a regularly varying function $g(x,y)$ of continuous arguments so that
$$p(i,j) = g(i,j).$$
- If the measure U is regularly varying, is the mass function p also regularly varying?
- If the mass function p is regularly varying, is U a regularly varying measure?

Regularly varying array-indexed functions

Definition 1.1

A doubly indexed function $f : \mathbb{Z}^2 \setminus \{\mathbf{0}\} \mapsto \mathbb{R}_+$ is regularly varying with scaling functions b_1 and b_2 and limit function $\lambda(x, y)$ if for some $h \in RV_\alpha$ for some $\alpha \in \mathbb{R}$, $b_i \in RV_{\beta_i}$, $\beta_i > 0$, we have

$$\lim_{n\to\infty} \frac{f([b_1(n)x], [b_2(n)y])}{h(n)} = \lambda(x, y) > 0, \quad \forall x, y > 0. \tag{1.1}$$

- A function $g : \mathbb{R}_+^2 \to \mathbb{R}_+^2$ is regularly varying if the same limit holds without the greatest integer function square brackets $[], []$.
- When f satisfies (1.1), we say $f(i, j)$ is *embeddable* if there exists a bivariate regularly varying function $g(x, y)$ such that $g(x, y) := f([x], [y])$.
- In one dimension, a regularly varying sequence c_n can always be embedded in a regularly varying function $g(x)$ of a continuous argument.

Results

Suppose $u(i,j) > 0$ is a regularly varying mass function and satisfies some *extra condition*, then

- The function
$$g(x,y) := u([x],[y])$$
is regularly varying as function of continuous variables and therefore $u(i,j)$ is embeddable.
- If $u(i,j) = p(i,j)$ is a pmf corresponding to (X,Y), then
$$P[(X,Y) \in \cdot\,]$$
is a regularly varying measure.

One choice of *extra condition*:
$u(i,j)$ is eventually decreasing in both i and j. – Easy assumption but hard to check, can only show this hold for standard preferential attachment models.

Alternatively, assume

- $h(\cdot) \in RV_\rho$, $\rho < 0$, and $u : \mathbb{Z}_+^2 \mapsto \mathbb{R}_+$,
- Scaling functions: $b_i(t) = t^{1/\alpha_i}$, $i = 1, 2$.
- There exists a limit function $\lambda_0 > 0$ defined on

$$\mathcal{E}_0 := \{(x, y) : \|(x^{\alpha_1}, y^{\alpha_2})\| = 1\}, \tag{2.1}$$

such that u satisfies

$$\lim_{t\to\infty} \frac{u([t^{1/\alpha_1}x], [t^{1/\alpha_2}y])}{h(t)} = \lambda_0(x, y), \quad \forall (x, y) \in \mathcal{E}_0. \tag{2.2}$$

Then

- The doubly indexed function $u(i,j)$ is regularly varying: For all $x,y>0$, define $\mathbf{w}=\mathbf{w}(x,y):=(x^{\alpha_1},y^{\alpha_2})$ and

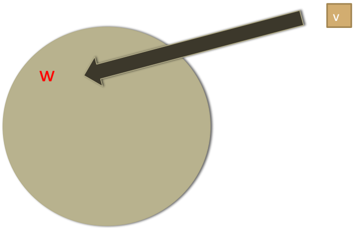
$$\lim_{n\to\infty}\frac{u([n^{1/\alpha_1}x],[n^{1/\alpha_2}y])}{h(n)}=\lambda(x,y):=\lambda_0\left(\frac{x}{\|\mathbf{w}\|^{1/\alpha_1}},\frac{y}{\|\mathbf{w}\|^{1/\alpha_2}}\right)\|\mathbf{w}\|^{\rho};$$

- The doubly indexed function $u(i,j)$ is embeddable in a non-standard regularly varying function $f:\mathbb{R}_+^2\mapsto\mathbb{R}$ with limit function $\lambda(\cdot)$ such that $f(x,y)=u([x],[y])$;
- If convergence in (2.2) is *uniform* on $\mathcal{E}_0$, then also the measure corresponding to $u(i,j)$ is a (discretely supported) regularly varying measure.

Model description

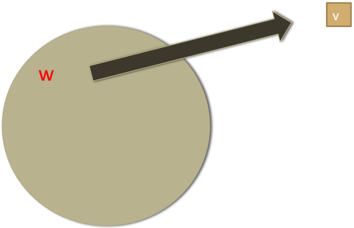
See Bollobás, Borgs, Chayes and Riordan (2003) and Krapivsky and Redner (2001).

- Model parameters: α, β, γ, $\delta_{in}, \delta_{\text{out}}$ with $\alpha + \beta + \gamma = 1$.
- $G(n)$ is a directed random graph with n edges, $N(n)$ nodes.
- Set of nodes of $G(n)$ is V_n; so $|V_n| = N(n)$.
- Set of edges of $G(n)$ is $E_n = \{(u, v) \in V_n \times V_n : (u, v) \in E_n\}$.
- In-degree of v is $D_{\text{in}}(v)$; out-degree of v is $D_{\text{out}}(v)$. Dependence on n is suppressed.
- Obtain graph $G(n)$ from $G(n-1)$ in a Markovian way as follows:



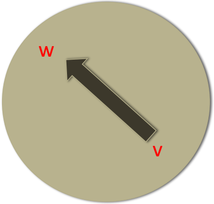
1. With probability α, append to $G(n-1)$ a new node $v \notin V_{n-1}$ and create directed edge $v \mapsto w \in V_{n-1}$ with probability

$$\frac{D_{\text{in}}(w) + \delta_{\text{in}}}{n - 1 + \delta_{\text{in}} N(n-1)}.$$



2. With probability γ, append to $G(n-1)$ a new node $v \notin V_{n-1}$ and create directed edge $w \in V_{n-1} \mapsto v \notin V_{n-1}$ with probability

$$\frac{D_{\text{out}}(w) + \delta_{\text{out}}}{n - 1 + \delta_{\text{out}} N(n-1)}.$$



3. With probability β, create new directed edge between existing nodes

$$v \in V_{n-1} \mapsto w \in V_{n-1}$$

with probability

$$\Big(\frac{D_{\text{out}}(v) + \delta_{\text{out}}}{n - 1 + \delta_{\text{out}} N(n-1)}\Big)\Big(\frac{D_{\text{in}}(w) + \delta_{\text{in}}}{n - 1 + \delta_{\text{in}} N(n-1)}\Big).$$

Applications to preferential attachment models

For $i, j = 0, 1, 2, \ldots$ and $n \geq n_0$, let $N_{ij}(n)$ be the random number of nodes in $G(n)$ with in-degree i and out-degree j. There exist non-random constants $p(i, j)$ such that

$$\lim_{n \to \infty} \frac{N_{ij}(n)}{N(n)} = p(i, j) \quad \text{a.s. for } i, j = 0, 1, 2, \ldots. \tag{3.1}$$

Define two random variables (I, O) such that

$$P[I = i, O = j] = p(i, j), \quad i, j = 0, 1, 2, \ldots$$

and the distribution generated by (I, O) is a non-standard regularly varying measure. The pair (I, O) has representation

$$(I, O) \stackrel{d}{=} B(1 + X_1, Y_1) + (1 - B)(X_2, 1 + Y_2), \tag{3.2}$$

where B is a Bernoulli switching variable independent of X_j, Y_j, $j = 1, 2$ with

$$\mathbb{P}(B = 1) = 1 - \mathbb{P}(B = 0) = \frac{\gamma}{\alpha + \gamma}.$$

Let $T_\delta(p)$ be a negative binomial integer valued random variable with parameters $\delta > 0$ and $p \in (0, 1)$. Now suppose $\{T_{\delta_1}(p), p \in (0,1)\}$ and $\{\tilde{T}_{\delta_2}(p), p \in (0,1)\}$ are two independent families of negative binomial random variables and define

$$c_1 = \frac{\alpha+\beta}{1+\delta_{\text{in}}(\alpha+\gamma)}, \quad c_2 = \frac{\beta+\gamma}{1+\delta_{\text{out}}(\alpha+\gamma)} \quad \text{and } a = c_2/c_1.$$

X_j, Y_j, $j = 1, 2$ in (3.2) can be written as

$$\begin{aligned}(X_1, Y_1) =& (T_{\delta_{\text{in}}+1}(Z^{-1}), \tilde{T}_{\delta_{\text{out}}}(Z^{-a})),\\ (X_2, Y_2) =& (T_{\delta_{\text{in}}}(Z^{-1}), \tilde{T}_{\delta_{\text{out}}+1}(Z^{-a})),\end{aligned}$$

where Z is a Pareto random variable on $[1, \infty)$ with index c_1^{-1}, independent of the negative binomial random variables.

From the representations:

$$\lim_{n\to\infty}\frac{p([n^{c_1}x],[n^{c_2}y])}{n^{-(1+c_1+c_2)}} = \frac{\gamma}{\alpha+\gamma}f_1(x,y)+\frac{\alpha}{\alpha+\gamma}f_2(x,y)$$
$$= \frac{\gamma}{\alpha+\gamma}\frac{x^{\delta_{in}}y^{\delta_{out}-1}}{c_1\Gamma(\delta_{in}+1)\Gamma(\delta_{out})}\int_0^\infty z^{-(2+1/c_1+\delta_{in}+a\delta_{out})}e^{-\left(\frac{x}{z}+\frac{y}{z^a}\right)}\mathrm{d}z$$
$$+\frac{\alpha}{\alpha+\gamma}\frac{x^{\delta_{in}-1}y^{\delta_{out}}}{c_1\Gamma(\delta_{in})\Gamma(\delta_{out}+1)}\int_0^\infty z^{-(1+a+1/c_1+\delta_{in}+a\delta_{out})}e^{-\left(\frac{x}{z}+\frac{y}{z^a}\right)}\mathrm{d}z.$$

- This convergence can be shown to be uniform on $\mathcal{E}_0$.
- Therefore, this uniform convergence implies

$$P[(I,O)\in\cdot\,]$$

is a regularly varying measure.

Threshold Selection

For power-law distributed data, we want to estimate

1. the scaling parameter α
2. the lower-limit on the scaling region x_{min} from empirical data.

Clauset (2004):

1. For $k = 1 \ldots n$, compute the Kolmogorov-Smirnov distance

$$D_k = \sup_{y \geq 1} \left| \frac{1}{k} \sum_{i=1}^{k} \epsilon_{\frac{X_{(i)}}{X_{(k+1)}}}(y, \infty] - y^{-\hat{\alpha}(k)} \right|,$$

where

$$\hat{\alpha}(k)^{-1} = \frac{1}{k} \sum_{i=1}^{k} \log \frac{X_{(i)}}{X_{(k+1)}}.$$

2. Choose

$$k^* = argmin D_k,$$

then $\hat{x}_{min} = X_{(k^*+1)}$ and $\hat{\alpha} = \hat{\alpha}(k^*)$.

Question: Is $\hat{\alpha}(k^*)$ consistent?
We can show that $k^* \xrightarrow{P} \infty$.
Asymptotically, under the assumption of second order regular variation $\overline{F} \in 2RV_{-\alpha,\rho}$, D_k is bounded by

$$\frac{1}{\sqrt{k}} \sup_{t\in(0,1]} |W(t) - tW(1) + t\log tW(1)| \\ + Const.g(b(n/k)) + o(k^{-1/2} + g(b(n/k))),$$

for some $g \in RV_\rho$, $\rho < 0$.
Then k^* satisfies

$$\sqrt{k^*} g(b(n/k^*)) \to 1,$$

and it follows that $k^* = h(n)$, with $h \in RV_{\frac{2|\rho|}{2|\rho|+\alpha}}$. This shows that k_n^* is an intermediate sequence so the corresponding hill estimator $\hat{\alpha}(k_n^*)^{-1}$ is consistent.

Further Questions:

- In practice, given a certain data set, how can we tell whether the underlying distribution has second order regular variation?
  Naive approach: look at hill plots, but can we do better??
- If the data is in fact Pareto or for example, log-gamma (with $\rho = 0$), what shall we do?
  Experimentally, Clauset's algorithm will lead us to choose the whole sample and do MLE. What about theoretically proving this??