# $r$-largest jump or observation; trimming; PA models

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

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

May 10, 2017

**Work with:** Tiandong Wang, R. Maller, B. Buchmann, Y. Fan Ipsen

# 1. Outline

- Preferential attachment models
  - Fitting the directed linear PA model (Wan, Wang, Davis, and Resnick (2017)).
    * MLE vs asymptotic EVT methods.
    * MLE superior on simulated data.
    * Model simplistic for real data but can alert you to interventions and deviations from the model.
  - Tiandong: Embedding undirected (and hopefully directed) PA model in birth (Birth-Immigration, Markov branching) processes. Goals:
    * Seek methodology more robust to changes in PA assumptions.
    * Seek justifications for using EVT methods on data that is far from iid.
- $r$th largest of an iid sequence or $r$th highest point in a Poisson random measure or $r$th largest jump of a Lévy process; trim a Lévy process; joint distribution of

  (trimmed Lévy, r-th largest jump).

## 2. The $r$th largest of an iid sequence

Buchmann, Maller, and Resnick (2016)

- 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 $r$th largest among $X_1, \ldots, X_n$ and set

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

### 2.1. 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 $R$ or $F$:

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

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 $r \to \infty$,

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

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 $\mathbb{R}^\infty$. Set,

$$\boldsymbol{X}^{(r)} = \big(\underbrace{-\infty, \dots, -\infty,}_{r-1 \text{ entries}} M_n^{(r)}, n \geq r\big).$$

Then in $\mathbb{R}^\infty$,

$$\left(\boldsymbol{X}^{(r+1)} | \boldsymbol{X}^{(r)} \dots \boldsymbol{X}^{(1)}\right) \stackrel{d}{=} \left(\boldsymbol{X}^{(r+1)} | \boldsymbol{X}^{(r)}\right).$$

- Use?

- Similar but not identical results for $r$th order extremal processes: Let

$$N = \sum_k \epsilon_{(t_k, j_k)},$$

be Poisson random measure on $[0, \infty) \times (x_l, x_r)$, with mean measure $ds \times \Pi$. Set

$$Q(x) = \Pi(x, x_r) < \infty, \quad x_l < x < x_r$$

and assume $Q(x_l) = \Pi(x_l, x_r) = \infty$. Define

$$Y^{(r)}(t) := \inf\{x > x_l : N\big([0, t] \times (x, x_r)\big) < r\}, \quad t > 0.$$

CORNELL

Set $t = 1$.

- Q?: When does a limit law exist for $Y^{(r)}(1)$ as $r \to \infty$.
- A: N&S condition: $\exists\ a(r) > 0,\ b(r)$

$$\lim_{r\to\infty} \frac{r - Q\big(a(r)x + b(r)\big)}{\sqrt{r}} = h(x),$$

for a non-decreasing limit function $h(x) \in \mathbb{R}$ with at least two points of increase. Requires

$$G(x) = e^{-Q^{1/2}(x)}$$

to be in a domain of attraction for minima.

# 3. Trimming a Lévy subordinator

Setup: Let $X = X(1) \geq 0$ be a Lévy subordinator, Lévy measure $\nu(\cdot)$ with $Q(x) = \nu(x, \infty)$. Define $N = \sum_k \epsilon_{j_k}(\cdot) = \mathrm{PRM}(\nu)$ and

$$X = \int_0^\infty u N(du) = \sum_{l=1}^{\infty} Q^{\leftarrow}(\Gamma_l)$$

$$= \text{sum of Poisson jumps written in decreasing order,}$$

$$^{(r)}X = \sum_{l=r+1}^{\infty} Q^{\leftarrow}(\Gamma_l),$$

$$= r\text{-largest jumps peeled off the Lévy process at } t = 1;$$

$$Y^{(r)} = Q^{\leftarrow}(\Gamma_r) = r\text{th largest jump of Lévy process.}$$

When does

$$({}^{(r)}X, Y^{(r)})$$

have a limit distribution (with appropriate centering and scaling)?

## 3.1. Joint limits.

Note: We always have

$$\frac{{}^{(r)}X - \mu(Y^{(r)})}{\sigma(Y^{(r)})} = \frac{{}^{(r)}X - \int_0^{Y^{(r)}} u\nu(du)}{\int_0^{Y^{(r)}} u^2\nu(du)} \Rightarrow N_X = N(0,1)$$

since $r \to \infty$ means we mash down the size of the jumps.

Assuming $Y^{(r)}$ has a limit law we get jointly

$$\Big(\frac{{}^{(r)}X - \mu(Y^{(r)})}{\sigma(Y^{(r)})}, \frac{Y^{(r)} - b(r))}{a(r)}\Big) \Rightarrow \big(N_X, h^{\leftarrow}(N_\Gamma)\big),$$

where $(N_X, N_\Gamma)$ are independent standard normal random variables. Proceed conditionally on $Y^{(r)}$ and then uncondition.

Can we get deterministic centering and scaling for $X$?

- Yes if $X$ is stable or if $Q$ is regularly varying at 0 and then the limit is of the form

$$\Big(\frac{{}^{(r)}X-\mu(b(r))}{\sigma(b(r))},\frac{Y^{(r)}-b(r)}{a(r)}\Big)\Rightarrow\Big(N_X+\frac{N_\Gamma}{\sqrt{2c}},N_\Gamma\Big)$$

=dependent normal rv's,

where $(N_X, N_\Gamma)$ are independent standard normal random variables.

- In general the answer depends on $\gamma$, the EV parameter for

$$G=e^{-Q^{1/2}(x)}.$$

- Limit may not be normal but it will be a function of independent normals.

# Contents

## References


B. Buchmann, R. Maller, and S. Resnick. Processes of rth Largest. *ArXiv e-prints*, July 2016. http://adsabs.harvard.edu/abs/2016arXiv160708674B.

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.

P. Wan, T. Wang, R. A. Davis, and S. I. Resnick. Fitting the Linear Preferential Attachment Model. *ArXiv e-prints, 1703.03095*, March 2017. URL https://arxiv.org/abs/1703.03095. Submitted: Electronic J. Statistics.

CORNELL