Extreme Value Limit Theory without Extreme Value Distributions

O'Brien's theory for maxima of stationary sequences (Draft, version 1.0)

LABEX MME - DII CHAIRE INTERNATIONALE
Paris, January 8th, 2014

Adam Jakubowski
Nicolaus Copernicus University
Toruń, Poland

Content

I. An overview.

II. O'Brien's theory:

- Phantom distribution function and criteria for its existence.
- Extremal index of a stationary sequence.

III. A multi-sequence method for multivariate extremes.

IV. Calculating limits for maxima of random fields. Phantom distribution function and extremal index for stationary random fields.

V. An asymptotic $(r-1)$-dependent representation for r-th order statistics from a stationary sequence. Understanding joint distributions of order statistics.

VI. Extremes on $\mathbb{D}([0,1])$.

Limit theorems for extrema due to Gnedenko (with complements due to de Haan)

- $\{X_j\}_{j\in\mathbb{N}}$ - an i.i.d. sequence of random variables, $M_n = \max_{1\leqslant j\leqslant n} X_j$, $n \in \mathbb{N}$.
- Problem: find $a_n > 0, b_n \in \mathbb{R}$ and a non-degenerate distribution function H such that

$$(*) \qquad \lim_{n\to\infty} P((M_n - b_n)/a_n \leqslant x) = H(x), \quad x \in \mathbb{R}.$$

- Question I: identify possible types of limiting distributions (3 types or one-parameter family indexed with $\gamma \in \mathbb{R}$).
- Question II: describe when a particular distribution F of X_j leads to the given limiting H (“domains of attraction”).
- Question III: for given F and H determine asymptotically a_n and b_n.
- Question IV: find a rate of convergence in $(*)$.

LABEX MME-DII

Comments on the extreme value limit theory

- $(+)$ Parallels the limit theory for sums.
- $(\pm)$ The exists a variety of limiting laws with quite different properties.
- $(-)$ Limiting laws in general do not scale: if X_j's are standard normal and

$$\lim_{n\to\infty} P((M_n - b_n)/a_n \leqslant x) = \exp(-e^{-x}),$$

  and if $X_j' \sim \mathcal{N}(0, \sigma^2)$, where $\sigma^2 \neq 1$, then $\lim_{n\to\infty} P((M_n' - b_n)/a_n \leqslant x) =$ either 0 or 1.
- $(-)$ The last property makes difficult deducing limit properties on the base of conditional distributions.
- $(-)$ In many practical tasks only a single sequence $\{v_n\}$ is of interest. For example:
  - For given αclose to 1 find

$$\lim_{n\to\infty} P(M_n \leqslant v_n) = \alpha \in (0,1).$$

  - Find exact asymptotics (large deviations) of

$$P(M_n \geqslant v_n) \to 0.$$

O'Brien's theory for maximas of stationary sequences (1974, 1987)

- Basic observation. If $\{X_j\}$ are i.i.d. and $X_j \sim F$, then for arbitrary sequence $\{v_n\}$

$$P(M_n \leqslant v_n) = F(v_n)^n = \exp\big(-n(1-F(v_n))\big) + o(1).$$

- O'Brien's (1974) observation: Given $\alpha \in (0,1)$, it is possible to find constants $\{v_n = v_n(\alpha)\}$ such that

$$P(M_n \leqslant v_n) = F(v_n)^n \to \alpha, \text{ as } n \to \infty,$$

if, and only if,

$$F(F_*-) = 1 \quad \text{and} \quad \lim_{x \to F_*-} \frac{1-F(x)}{1-F(x-)} = 1,$$

where

$$F_* = \sup\{x : F(x) < 1\}.$$

- If for some $\alpha \in (0,1)$ then for every $\alpha \in (0,1)$!

LABEX MME-DII

O'Brien's theory for maxima of stationary sequences

- By definition, a distribution function G is regular (in the sense of O'Brien) if

$$G(G_*-) = 1 \quad \text{and} \quad \lim_{x \to G_*-} \frac{1 - G(x)}{1 - G(x-)} = 1.$$

- Now suppose that $\{X_j\}$ is a stationary sequence of random variables, with marginal distribution F.
- Following O'Brien (1987) we call any distribution function G satisfying

$$(*) \qquad P(M_n \leqslant v_n) - G^n(v_n) \to 0, \text{ as } n \to \infty,$$

  for all sequences $\{v_n\}$, a phantom distribution function for $\{X_j\}$.
- Clearly $(*)$ is equivalent to

$$sup_v |P(M_n \leqslant v) - G^n(v)| \to 0, \text{ as } n \to \infty.$$

- G is not uniquely determined!

O’Brien’s theory for maxima of stationary sequences

- O’Brien (1987) gave sufficient condition for the existence of a regular phantom distribution function.
- In J. (1991) and J. (1993) necessary and sufficient conditions were given.

LABEX
MME-DII

O'Brien's theory for maxima of stationary sequences

Theorem

Let $\{X_j\}$ be stationary. The following are equivalent:

1. The sequence $\{X_j\}$ admits a regular phantom distribution function.
2. There exists a sequence $\{v_n\}$ and $\alpha \in (0,1)$ such that
$$P(M_n \leqslant v_n) \to \alpha.$$
and the following Condition $B_\infty(v_n)$ holds: as $n \to \infty$
$$\sup_{p,q\in\mathbb{N}} \left| P(M_{p+q} \leqslant v_n) - P(M_p \leqslant v_n) P(M_q \leqslant v_n) \right| \to 0,$$
3. There exists $\alpha \in (0,1)$ such that for some dense subset $\mathbb{Q} \subset \mathbb{R}^+$
$$P(M_{[nt]} \leqslant v_n) \to \alpha^t,\ \ t \in \mathbb{Q}.$$

O'Brien's theory for maxima of stationary sequences

- In fact, given $\alpha \in (0,1)$ and $\{v_n\}$ we can construct a regular phantom distribution function G for $\{X_j\}$.
- First, we can replace $\{v_n\}$ with

$$v_n^* = \begin{cases} \max\{v_k \,:\, 1 \leqslant k \leqslant n, v_k < F_*\} & \text{if} \neq \emptyset, \\ \inf\{v_n \,:\, n \in N\} & \text{otherwise,} \end{cases}$$

  which is nondecreasing.
- Then one can define

$$G(x) = \begin{cases} 0, & \text{if } x < v_1^*, \\ \alpha^{1/n}, & \text{if } v_n^* \leqslant x < v_{n+1}^*, \\ 1, & \text{if } x \geqslant \sup\{v_n^* \,:\, n \in N\}. \end{cases}$$

LABEX MME-DII

O'Brien's theory for maxima of stationary sequences

- How to check $\lim_{n\to\infty} P(M_n \leqslant v_n) = \alpha$?
- By mixing, for some $k_n \to \infty$,

$$\begin{aligned} P(M_n \leqslant v_n) &= \Big(P(M_{[n/k_n]} \leqslant v_n)\Big)^{k_n} + o(1), \\ &= \exp\Big(-k_n P(M_{[n/k_n]} > v_n)\Big) + o(1). \end{aligned}$$

- O'Brien (1987) gives conditions for k_n which allow to write

$$= \exp\Big(-nP(X_0 > v_n, M_{[n/k_n]-1} \leqslant v_n)\Big) + o(1).$$

- This extends earlier results of Newell (1964) for m-dependent sequences and many results for Markov chains (see e.g. Chernick et al. (1991)).

Managing clusters of big values

- Let us observe that in O'Brien's formula (writing r_n for $[n/k_n]$)

$$P(X_0 > v_n, M_{r_n-1} \leqslant v_n) = P(M_{r_n-1} \leqslant v_n) - P(M_{r_n} \leqslant v_n).$$

- This means that on the one hand

$$P(M_n \leqslant v_n) = \exp\Big(- k_n P(M_{r_n} > v_n)\Big) + o(1),$$

while on the other hand

$$P(M_n \leqslant v_n) = \exp\Big(- nP(X_0 > v_n, M_{r_n-1} \leqslant v_n)\Big) + o(1)$$
$$= \exp\Big(- k_n\Big(r_n\Big(P(M_{r_n-1} \leqslant v_n) - P(M_{r_n} \leqslant v_n)\Big)\Big)\Big) + o(1).$$

- Hence it must be

$$k_n\Big(P(M_{r_n} < v_n) - r_n\Big(P(M_{r_n-1} \leqslant v_n) - P(M_{r_n} \leqslant v_n)\Big)\Big) \to 0.$$

Managing clusters of big values

Lemma (J. 1997, also BJMW 2011, Balan & Louichi 2010

Let $Z_1, Z_2, \ldots$ be strictly stationary random vectors. Set $T_0 = 0,\ T_k = \sum_{j=1}^k Z_j,\ k \in \mathbb{N}$. If $0 \notin U$, then for every $n \in \mathbb{N}$ and every $m \in \mathbb{N}$, $m \leqslant n$, the following inequality holds:

$$\Big| P(T_n \in U) - n\big(P(T_{m+1} \in U) - P(T_m \in U)\big)\Big| \leqslant 3mP(Z_1 \neq 0) + 2 \sum_{\substack{1 \leqslant i < j \leqslant n \\ j-i > m}} P(Z_i \neq 0, Z_j \neq 0).$$

Set $n = r_n$, $Z_k = I\{X_k > v_n\}$, $U = (0, +\infty)$, to get

$$\Big| P(M_{r_n} > v_n) - r_n\big(P(M_{m+1} > v_n) - P(M_m > v_n)\big)\Big|$$
$$= \Big| P(M_{r_n} > v_n) - r_n\big(P(X_0 > v_n, M_m \leqslant v_n)\big)\Big|$$
$$\leqslant 3mP(X_1 > v_n) + 2 \sum_{\substack{1 \leqslant i < j \leqslant n \\ j-i > m}} P(X_i > v_n, X_j > v_n).$$

LABEX MME-DII

Managing clusters of big values

Corollary

If $k_n \to \infty$ and m_n are such that

$$\begin{aligned} \lim_{n\to\infty} k_n m_n P(X_1 > v_n) &= 0, \\ \lim_{n\to\infty} k_n \sum_{\substack{1\leqslant i<j\leqslant [n/k_n] \\ j-i>m_n}} P(X_i > v_n, X_j > v_n) &= 0, \end{aligned}$$

(plus mixing into k_n blocks) then

$$P(M_n \leqslant v_n) = \exp\Big(- nP(X_0 > v_n, M_{m_n} \leqslant v_n)\Big) + o(1).$$

LABEX MME-DII

Managing clusters of big values

Corollary

If $k_n \to \infty$ is such that

$$P(M_n \leqslant v_n) = \Big(P(M_{[n/k_n]} \leqslant v_n)\Big)^{k_n} + o(1),$$

$$\begin{aligned} \lim_{n\to\infty} k_n P(X_1 > v_n) &= 0, \\ \lim_{m\to\infty} \limsup_{n\to\infty} k_n \sum_{\substack{1\leqslant i<j\leqslant[n/k_n]\\ j-i>m}} P(X_i > v_n, X_j > v_n) &= 0, \end{aligned}$$

if for each $m \in \mathbb{N}$ there exists

$$\lim_{n\to\infty} nP(X_0 > v_n, M_m \leqslant v_n) = \beta(m),$$

then

$$\lim_{n\to\infty} P(M_n \leqslant v_n) = \exp\big(-\lim_{m\to\infty} \beta(m)\big).$$

Extremal index due to Leadbetter(1983), also Loynes(1965), O'Brien (1974)

- The extremal index of a stationary sequence $\{X_j\}$ is a number $\theta \in (0,1)$, such that for all $\tau > 0$,

$$(*) \qquad P(M_n \leqslant u_n(\tau)) \to e^{-\theta\tau}$$

  whenever

$$(**) \qquad nP(X_1 > u_n(\tau)) \to \tau.$$

- Let $\{\widehat{X}_j\}$ be the sequence "associated" to $\{X_n\}$, i.e. $\widehat{X}_j$'s are i.i.d. with the same marginal distributions as $X_j : \mathcal{L}(\widehat{X}_j) = \mathcal{L}(X_j)$. Then (**) means $P(\widehat{M}_n \leqslant u_n(\tau)) \to e^{-\tau}$ and (*) and (**) imply

$$P(M_n \leqslant u_n) - P(\widehat{M}_n \leqslant u_n)^{\theta} \to 0,$$

  at least for sequences $u_n = u_n(\tau)$ defined by (**).

Extremal index due to Leadbetter

In fact, Leadbetter (1983) proved that if $\theta > 0$ then the relation

$$P(M_n \leqslant u_n) - P(\widehat{M}_n \leqslant u_n)^{\theta} \to 0,$$

holds for all sequences $\{u_n\}$. Hence $G(x) = F^{\theta}(x)$ is a phantom distribution function for $\{X_j\}$.

Theorem

Let $\{X_j\}$ be a stationary sequence. Then $\{X_j\}$ has the extremal index $\theta > 0$ if and only if there exists a sequence $\{v_n\}$ such that Condition $B_{\infty}(v_n)$ holds and for some $\alpha, \widehat{\alpha} \in (0, 1)$

$$\begin{aligned} P(M_n \leqslant v_n) &\to \alpha, \\ nP(X_1 > v_n) &\to -\log \widehat{\alpha}. \end{aligned}$$

In such a case

$$\theta = \frac{\log \alpha}{\log \widehat{\alpha}}.$$

A standard example

- Let $Y_1, Y_2, \ldots$ be an i.i.d. sequence with regular distribution function F. Define

$$X_j = Y_j \vee Y_{j+1},\ j \in \mathbb{N}.$$

- Then $\{X_j\}$ is a stationary and 1-dependent sequence with extremal index $\theta = 1/2$.
- This is because

$$P(\max_{1\leqslant j\leqslant n} X_j \leqslant v_n) = P(\max_{1\leqslant j\leqslant n+1} Y_j \leqslant v_n)$$
$$= F(v_n)^{n+1} = \exp\Big(-(n+1)(1-F(v_n))\Big) + o(1),$$

while

$$nP(X_1 > v_n) \sim 2nP(Y_1 > v_n) = 2n(1-F(v_n)).$$

- A simple task: given a number $\theta \in (0,1)$ find a stationary sequence with the extremal index θ.

LABEX MME-DII

A model for order statistics

- Let $\beta_1, \beta_2, \ldots, \beta_r \geqslant 0$ be such that $\sum_{q=1}^{r} \beta_q = 1$, and let G be a (regular) distribution function.
- For each $1 \leqslant q \leqslant r$, let $\{\tilde{Y}_{qj}\}_{j\in\mathbb{N}}$ be independent, identically distributed with

$$\tilde{Y}_{qj} \sim G^{\beta_q}.$$

- Let sequences $\{\tilde{Y}_{1j}\}_{j\in\mathbb{N}}, \{\tilde{Y}_{2j}\}_{j\in\mathbb{N}}, \ldots, \{\tilde{Y}_{rj}\}_{j\in\mathbb{N}}$ be mutually independent.
- Define a new, $(r-1)$ - dependent sequence:

$$\begin{aligned}
\tilde{X}_j \quad = \quad & \tilde{Y}_{1j} \\
& \vee \big(\tilde{Y}_{2j} \vee \tilde{Y}_{2,j+1}\big) \\
& \vee \big(\tilde{Y}_{3j} \vee \tilde{Y}_{3,j+1} \vee \tilde{Y}_{3,j+2}\big) \\
& \vdots \\
& \vee \big(\tilde{Y}_{rj} \vee \tilde{Y}_{r,j+1} \vee \ldots \vee \tilde{Y}_{r,j+r-1}\big).
\end{aligned}$$

A model for order statistics

- Let $\tilde{M}_n^{(1)}, \tilde{M}_n^{(2)}, \ldots, \tilde{M}_n^{(r)}$ be the highest order statistics for $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n$ defined above.
- If v_n is such that

$$P(\tilde{M}_n^{(1)} \leqslant v_n) \longrightarrow \alpha_1, \text{ as } n \to \infty,$$

where $\alpha_1 \in (0, 1)$,
- then also for $q = 2, 3, \ldots, r$

$$P(\tilde{M}_n^{(q)} \leqslant v_n) \longrightarrow \alpha_q, \text{ as } n \to \infty,$$

where α_q are functions of α_1 and $\beta_1, \beta_2, \ldots \beta_r$.

Non-stationarity - some motivating examples

R. Ballerini and S. Resnick, Records from improving populations, *J. Appl. Probab.*, 22 (1985) 487–502.

$\{X_j\}$ - i.i.d. sequence, F_{X_j} - continuous, $Y_j = X_j + c \cdot j$.
What are the records of a nonstationary sequence $\{Y_j\}$?
If $F(x) = exp(-exp(-(x-a)/b))$, then

$$P(Y_j \leqslant x) = P(X_j \leqslant x - c \cdot j) = F(x)^{e^{c \cdot j/b}}.$$

A. Kukush, Y. Chernikov and D. Pfeifer, Maximum Likelihood Estimators in a Statistical Model of Natural Catastrophe Claims with Trend, *Extremes*, 7 (2004) 309–336.

$\{X_j\}$ - independent, $X_j \sim F^{\gamma_j}$, where $\gamma_j = \gamma^{j-1}$ for some $\gamma > 1$.

LABEX
MME-DII

Non-stationarity - some motivating examples

- Common denominator: non-stationary model built on independent variables, with distributions designed according to some rule.
- Problem: What are the reasonable rules?

Asymptotic Independent Representation for Maxima

- Let $\{X_j\}_{k\in\mathbb{N}}$ be an arbitrary sequence of random variables. Define as before $M_n = \max_{1\leqslant j\leqslant n} X_j$.
- Suppose one can find a sequence $\{\tilde{X}_j\}$ of independent random variables such that

$$\sup_{x\in\mathbb{R}^1} \Big| P(M_n \leqslant x) - P(\tilde{M}_n \leqslant x) \Big| \longrightarrow 0, \text{ as } n \to \infty,$$

  where $\tilde{M}_n$ is the n-th partial maximum of X_j's.
- Then $\{\tilde{X}_j\}$ is said to provide an asymptotic independent representation (a.i.r.) for maxima of $\{X_j\}_{j\in\mathbb{N}}$.

LABEX MME-DII

LABEX
MME-DII

Asymptotic Independent Representation for Maxima

- We suggest studying the whole path

$$\mathbb{R}^+ \ni t \mapsto P(M_{[nt]} \leqslant v_n)$$

and its limit behavior.

- The idea: Assume that

$$P(M_{[nt]} \leqslant v_n) \longrightarrow \alpha_t, \text{ as } n \to \infty, \ t \in \mathbb{Q},$$

for some dense subset $\mathbb{Q} \subset \mathbb{R}^+$ and we recover an a.i.r. from the limiting function α_t, provided the latter is of a special form.

- Note that α_t is non-increasing and can be regularized to the right-continuous function

$$\tilde{\alpha}_t = \sup_{\mathbb{Q} \ni u > t} \alpha_u,$$

for which $P(M_{[nt]} \leqslant v_n) \longrightarrow \tilde{\alpha}_t$ at every point of continuity.

Asymptotic Independent Representation for Maxima

Theorem, J. (1993)

Assume there is a sequence $\{v_n\}$ such that for some dense subset $\mathbb{Q} \subset \mathbb{R}^+ = (0, +\infty)$

$$P(M_{[nt]} \leqslant v_n) \longrightarrow \alpha_t, \text{ as } n \to \infty, \ t \in \mathbb{Q},$$

where the limiting function α_t possesses the following properties:

$$\begin{aligned} \alpha_t &> 0, \quad t \in \mathbb{Q} \\ \sup_{t \in \mathbb{Q}} \alpha_t &= 1, \\ \inf_{t \in \mathbb{Q}} \alpha_t &= 0. \end{aligned}$$

Then $\{X_k\}$ admits an asymptotic independent representation for maxima if, and only if, the function $g_\alpha = \log \circ \widetilde{\alpha} \circ \exp$ is concave.

Bibliography

- Balan, R. & Louhichi, S. (2010), Explicit conditions for the convergence of point processes associated to stationary arrays, *Elect. Comm. in Probab.*, **15**, 428—44.
- Chernick, M.R., Hsing, T., McCormick, W.P. (1991), Calculating the extremal index for a class of stationary sequences, *Adv. Appl. Probab.* **23**, 835–850.
- Bartkiewicz, K., Jakubowski, A., Mikosch, T. & Wintenberger, O. (2011), Stable limits for sums of dependent infinite variance random variables, *Probab. Theory Rel. Fields*, **150**, 337–372.
- de Haan, L. & Ferreira, A. (2006), **Extreme Value Theory. An Introduction**, Springer, New York.
- Galambos, J. (1978), **The Asymptotic Theory of Extreme Order Statistics**, Wiley, New York.

Bibliography

- Jakubowski, A. (1991), Relative extremal index of two stationary processes, *Stochastic Process. Appl.*, **37**, 281–297.
- Jakubowski, A. (1993), An asymptotic independent representation in limit theorems for maxima of nonstationary random sequences, *Ann. Probab.*, **21**, 819–830.
- Jakubowski, A. (1997), Minimal conditions in p-stable limit theorems II, *Stochastic Process. Appl.*, **68**, 1–20.
- Leadbetter, M. R. (1983) Extremes and local dependence in stationary sequences, *Z. Wahr. verw. Gebiete*, **65**, 291–306.
- Leadbetter, M. R., Lindgren, G. & Rootzén, H. (1983) **Extremes and related properties of random sequences and processes**, Springer, Berlin.

LABEX
MME-DII

Bibliography

- Loynes, R.M. (1965), Extreme values in uniformly mixing stationary processes, *Ann. Math. Statist.*,**36**, 993–999.
- Newell, G. F. (1964) Asymptotic extremes for m-dependent random variables, *Ann. Math. Statist.*, **35**, 1322–1325.
- O'Brien, G. L. (1974) The maximum term of uniformly mixing stationary processes, *Z. Wahr. verw. Gebiete*, **30**, 57–63.
- O'Brien, G. L. (1987) Extreme values for stationary and Markov sequences, *Ann. Probab.*,**15**, 281–292.