Extreme value theory for dependent random series with application to random graphs

Igor Rodionov
Trapeznikov Institute of Control Sciences, Moscow, Russia

joint work with Mikhail Isaev, Rui Zhang and Maxim Zhukovskii

8PCC, 2020

Introduction

Let $\{\xi_n\}_{n\geq 1}$ be iid random variables with a cumulative distribution function (cdf) F. Denote $M_n = \max\{\xi_1, \ldots, \xi_n\}$.

Fisher-Tippet-Gnedenko theorem states that if for some sequences $a_n > 0$, b_n, $n \geq 1$ it holds

$$\Pr(M_n \leq a_n x + b_n) \to G(x), \quad n \to \infty$$

for some non-degenerate G, then G belongs to either the Gumbel, the Fréchet or the Weibull family of distributions.

Introduction

Let $\mathbf{X}(n) = (X_1(n), \ldots, X_{d(n)}(n))^T$ be an arbitrary sequence of real-valued random vectors. Clearly, results on maxima of independent random variables can be applied to $\mathbf{X}(n)$, if, for every $x \in \mathbb{R}$,

$$\left| \Pr\left(\max_{i \in [d(n)]} X_i(n) \leq x \right) - \prod_{i \in [d(n)]} \Pr\left(X_i(n) \leq x\right) \right| \to 0, \quad (1)$$

as $n \to \infty$, where $[m]$ denotes $\{1, \ldots, m\}$. Notice that, in these general framework, for every n, the i-th element of the random vector (n) depends on n. Hereafter we write X_i instead of $X_i(n)$ and d instead of $d(n)$ for shortening.

Previous results

Our aim is to find new sufficient conditions of (1). Previous results in this field were obtained by:

- Leadbetter (1974, for stationary sequences)
- Hüsler (1983, 1986, for non-stationary sequences)
- Pereira and Ferreira (2006, for non-stationary random fields on $\mathbb{Z}^2$)
- Galambos (1972, in our setting, but his conditions are much more restrictive than ours)
- ...

Our approach generalizes all known similar results obtained for sequences of random variables (both for stationary and non-stationary sequences) and random fields on $\mathbb{Z}_+^k$.

Leadbetter-type φ-mixing

For two families of events $\mathcal{A}, \mathcal{B}$, we, as usual, denote

$$\phi(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |\Pr(B|A) - \Pr(B)|, \qquad (2)$$

where we follow the usual convention that $\Pr(A) = 0$ implies $\Pr(B|A) = \Pr(B)$. For simplicity, we let $\phi(A, \mathcal{B}) := \phi(\{A\}, \mathcal{B})$. For a family of events , we denote

$$\mathrm{U}(\mathcal{F}) = \left\{ \bigcup_{B \in \mathcal{F}'} B,\ \mathcal{F}' \subseteq \mathcal{F} \right\}.$$

Main result

Theorem

Let $\mathbf{X}(n) = (X_1, \ldots, X_d)^T$, $d = d(n)$, be a sequence of real-valued random vectors. Let $x \in \mathbb{R}$, $A_i = A_i(n) = \{X_i > x\}$. Assume there are sets $W_i = W_i(n) \subseteq [d] \setminus \{i\}$, $i \in [d]$, and $\varphi = \varphi(n) = o(1)$ such that,

(A1) $\sum_{i\in[d]} \Pr(\cup_{j\in[i-1]\setminus W_i} A_j) \Pr(A_i) = o(1)$;

(A2) *for every* $i \in [d]$, $\phi(A_i, \mathsf{U}(\{A_j\}_{j\in W_i})) \leq \varphi$;

(A3) $\sum_{i\in[d]} \sum_{j\in[i-1]\setminus W_i} \Pr(A_i \cap A_j) = o(1)$.

Then, (1) holds.

Notice that the same result holds for sequences of arbitrary events $\{A_i(n)\}$, $i \in [d]$, without any reference to random vectors.

Applications in random graphs: preliminaries

First we briefly discuss an extremal behavior of binomial random variables, which will be utilized to study certain extremal characteristics of random structures.

Assume $N = N(n) \in \mathbb{N}$, $N \to \infty$ as $n \to \infty$; $p = p(n) \in (0, 1)$ be such that $(\log n)^3 = o(Np(1-p))$ and $x \in \mathbb{R}$ do not depend on n. Denote

$$a_n = pN + \sqrt{2 \log n Np(1-p)} \left(1 - \frac{\log \log n}{4 \log n} - \frac{\log(2\sqrt{\pi})}{2 \log n}\right) \quad (3)$$

and $b_n = \sqrt{\frac{Np(1-p)}{2 \log n}}$.

Then de Moivre–Laplace theorem implies the following

Lemma (Nadaraja and Mitov (2002))

Let $X_1, \ldots, X_n \sim \mathrm{Bin}(N, p)$ be independent. Then we get

$$\Pr\left(\max_{i\in[n]} X_i \leq a_n + b_n x\right) \to e^{-e^{-x}}, \quad n \to \infty. \tag{4}$$

However, our asymptotic independence results allow to derive the results of (4) type for dependent random variables.

Maximum degree distribution

Let $\mathcal{H}_{n,k,p}$ be the k-uniform binomial random hypergraph with vertex set $[n]$. Denote the degree of a vertex in the complete k-uniform hypergraph on $[n]$ by $N := \binom{n-1}{k-1}$. Our asymptotic independence results imply the following result on asymptotic distribution of the maximum degree of $\mathcal{H}_{n,k,p}$.

Theorem

Assume $p = p(n) \in (0,1)$ and $k = k(n) \in \{2, \ldots, n\}$ is such that

$$(\log n)^3 = o\left(\max\left\{\binom{n}{k-2}, n\right\} p(1-p)\right), \quad k = o\left(\frac{n}{\log^2 n}\right).$$

Let X_i, $i \in [n]$, denote the degree of the vertex i in $\mathcal{H}_{n,k,p}$. Then (4) holds for all $x \in \mathbb{R}$ with a_n, b_n defined in (3).

Distribution of the maximum number of k-clique extensions

Let $k \geq 2$. For $i \in [n]$, let X_i be the number of k-cliques containing i. Below, we show that our main result implies the asymptotic distribution of the maximum value of X_i over $i \in [n]$. Denote

$$a_n^k = \frac{(pn)^{k-2} p^{\binom{k-1}{2}}}{(k-1)!} \Big[pn + (k-1)\sqrt{2np(1-p)\log n} \times \left(1 - \frac{\log\log n}{4\log n} - \frac{\log(2\sqrt{\pi})}{2\log n}\right) \Big],$$

$$b_n^k = \frac{1}{(k-2)!}(pn)^{k-2} p^{\binom{k-1}{2}} \sqrt{\frac{np(1-p)}{2\log n}}.$$

Theorem

Let $p = p(n) \in (0, 1)$, $k = k(n) \in \{2, \ldots, n\}$ *be such that*

$$\log^3 n = o(np(1-p)), \quad \log^2 n = o\left(\frac{np^{\binom{k-1}{2}+1}(1-p)}{k^2}\right).$$

Then for every $x \in \mathbb{R}$

$$\Pr\left(\max_{i \in [n]} X_i \le a_n^k + b_n^k x\right) \to e^{-e^{-x}}, \quad n \to \infty.$$

Distribution of maximum of common neighbors

The below result for constant h was proved in Rodionov Zhukovskii (2020). Let us show that it is a more or less direct corollary of our main result.
For $h \in \mathbb{N}$ and $\mathbf{x} \in \binom{[n]}{h}$, denote the number of common neighbors of vertices in $\mathbf{x}$ in $\mathcal{G}_{n,p}$ by $X_{\mathbf{x}}$. Define $a_{h,n} = a_n$, $b_{h,n} = b_n$ by replacing n with $\binom{n}{h}$, N with n and p with p^h in (3).

Theorem

Let $h = h(n) = o(\log n / \log\log n)$ and $p = p(n) \in (0,1)$ be such that

$$p^h/h \gg \log^3 n/n, \quad 1 - p \gg \sqrt{\log\log n/\log n}.$$

Then, for every $x \in \mathbb{R}$,

$$\Pr\Big(\max_{\mathbf{x}\in\binom{[n]}{h}} X_{\mathbf{x}} \le a_{h,n} + b_{h,n}x\Big) \to e^{-e^{-x}}, \quad n \to \infty.$$

References

[1] J. Galambos. On the distribution of the maximum of random variables. *The Annals of Mathematical Statistics*, 43(2): 516–521, 1972.

[2] J. Hüsler. Asymptotic approximation of crossing probabilities of random sequences. *Z. Wahrsch. Verw. Gebiete*, 63(2): 257–270, 1983.

[3] J. Hüsler. Extreme values of non-stationary random sequences. *J. Appl. Probab.*, 23(4): 937–950, 1986.

[4] M.R. Leadbetter. On extreme values in stationary sequences. *Z. Wahrsch. verw. Gebiete*, 28: 289–303, 1974.

[5] S. Nadarajah, K. Mitov. Asymptotics of maxima of discrete random variables. *Extremes*, 5(3): 287–294, 2002.

[6] L. Pereira, H. Ferreira. Limiting crossing probabilities of random fields. *Journal of Applied Probability*, 3: 884–891, 2006.

[7] I. Rodionov, M. Zhukovskii. The distribution of the maximum number of common neighbors in the random graph. *arXiv:1804.04430*, 2018.

Thank you

for your attention!