Краткое введение в теорию экстремальных величин и ее применения

Natalia Markovich

Institute of Control Sciences
Russian Academy of Sciences, Moscow, Russia

2023, Moscow

Analysis of extremal events

- Activities:

1. International journal Extremes, many journal include extreme value analysis in the topics
2. International conferences EVA-1999,2001,...,2023
3. Free software EVIS, POT packages,...
4. One World Extremes Seminar (OWE)

A broad spectrum of topics

ranging from

- Heavy-tailed phenomena;
- Time series; Large deviations;
- Multivariate modelling; Rare events;
- Spatio-temporal processes;

to

- Machine Learning;
- Graphical Modelling;
- Big data;
- Random networks

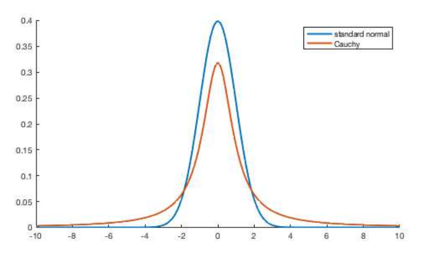
Applications

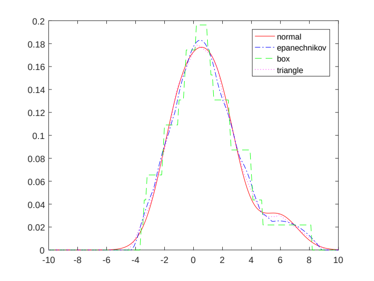
1. Climate change detection: (risk of global warming, large and low temperatures, waves, winds, overflooding, fire, ...)
2. Finance (insurance claims, portfolio, Value-at-Risk,...)
3. Telecommunication (file sizes, duration of subsessions, overloading and loss control,...)
4. Ecology (ozon and polution concentrations,...)
5. Geology (concentrations of rare minerals,...)
6. Astrophysics (sizes of comet dust, space debris, ...)
7. Graphical Modelling, Extremes of Energy Systems, Cause Inferences, Public Health, Epidemiology, Life Sciences and Life Lengths

Теория экстремумов занимается изучением максимумов и минимумов (а также других порядковых статистик) систем случайных величин.

- Фундаментальная работа: Б.В.Гнеденко (1943)
- Предшественники: М.Фреше (1927), Р.Фишер и Л.Типпетт (1928), Р. фон Мизес (1936)
- Классические монографии: Я.И.Галамбош (1984), N.Bingham, C. Goldie, J. Teugels (1987), S.Resnick (1987,2007), R.Leadbetter, G.Lingren, H.Rootzen (1989), P.Embrechts, C.Klűppelberg, T.Mikosh (1997), J.Beirlant et al. (2004), L.de Haan, A.Ferreira (2007)
- Современные исследователи в России: В.И.Питербарг, В.Б.Невзоров, А.В.Степанов, С.Ю.Новак, Н.М.Маркович, А.В.Лебедев, А.А.Голдаева, И.В.Родионов и др.

Heavy-tailed distributions





Cauchy distribution is super heavy-tailed: no one distribution moment exists.

Density kernel estimator $\widehat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x-X_i}{h})$,

$h > 0$ is bandwidth: consistency conditions $h \to 0$, $nh \to \infty$ as sample size $n \to \infty$

Devroye L., Györfi L. (1985) Nonparametric density estimation. The L_1 view. Wiley, New York.

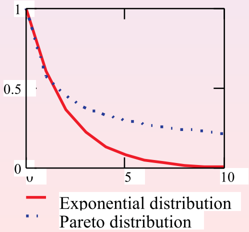
Definitions and main assumptions

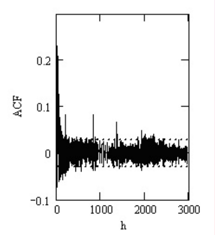
Let $X_1, \ldots, X_n$ be a sample of independent, identically distributed (i.i.d.) r.v.s X_i governed by the distribution function (DF) $F(x) = P\{X \leq x\}$ with probability density function (PDF) $f(x) = dF(x)/dx$.

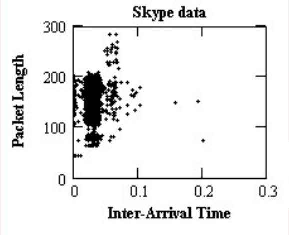
Definition

A DF $F(x)$ (or the r.v. X) is called heavy-tailed if its tail $\overline{F}(x) = 1 - F(x) > 0, \quad x \geq 0$ satisfies for all $y \geq 0$

$$\lim_{x\to\infty} P\{X > x + y | X > x\} = \lim_{x\to\infty} \overline{F}(x+y)/\overline{F}(x) = 1.$$







Examples of heavy- and light-tailed distributions

Heavy-tailed distributions:	Subexponential: Pareto, Lognormal, Weibull with shape parameter less than 1. With regularly varying tails: Pareto, Cauchy, Burr, Frechét, Zipf-Mandelbrot law. Super heavy-tailed
Light-tailed distributions	exponential, gamma, Weibull with shape parameter more than 1, normal, finite distributions.

Example of non-heavy tailed distribution: Exponential distribution $F(x) = 1 - e^{-\lambda x}$, $x \geq 0$ satisfies

$$\bar{F}(x+y)/\bar{F}(x) = e^{-\lambda(x+y)}/e^{-\lambda x} = e^{-\lambda y} \nrightarrow 1$$

as $x \to \infty$, $x \geq 0$, $y > 0$

Problems arising from heaviness of tails

1. Non-existence of some or all moments
2. Violation of Cramer's condition (the existence of moment generation function)
3. Sparse observations at the tails (outliers)

As consequence we cannot use

1. Likely Central limit theorem
   Likely Berry-Essen, Chebyshev's inequalities
2. Likely empirical averages and standard deviations
3. Histogramms, project and other estimators of the PDF which require a finite support

Heavy-tailed distributions have been accepted as realistic models for various phenomena:

- WWW-session characteristics
  - sizes and durations of sub-sessions; sizes of responses
  - inter-response time intervals
- on/off-periods of packet traffic
- file sizes
- service-time in queueing model
- flood levels of rivers
- major insurance claims
- extreme levels of ozon concentrations
- high wind-speed values
- wave heights during a storm
- low and high temperatures

Распределение максимума

Пусть $M_n = \max\{X_1, ..., X_n\}$ - максимум из n случайных величин $X_1, ..., X_n$.
Каково его распределение?
Если $X_1, ..., X_n$ независимы и одинаково распределены с функцией распределения $F(x)$, то

$$P\{M_n \le x\} = P\{X_1 \le x, ..., X_n \le x\} = P^n\{X_1 \le x\}.$$

Вопросы:

- Каково будет предельное распределение максимума при объеме выборки $n \to \infty$?
- Каково будет предельное распределение максимума, если $X_1, ..., X_n$ зависимы?

Basic results: Fisher-Tippett-Gnedenko theorem

Let $X^n = \{X_1, \ldots, X_n\}$ be a sample of i.i.d. r.v. distributed with the DF $F(x)$ and $M_n = \max(X_1, X_2, \ldots, X_n)$.
Fisher-Tippett-Gnedenko theorem states that for a suitable choice of normalizing constants $a_n > 0$, $b_n \in R$ it holds

$$P\{(M_n - b_n)/a_n \le x\} = F^n(b_n + a_n x) \to_{n\to\infty} H_\gamma(x), x \in R,$$

and an Extreme Value DF $H_\gamma(x)$ is of the following type:

$$H_\gamma(x) = \begin{cases} \exp(-x^{-1/\gamma}), & x > 0, \gamma > 0 \quad \text{'Fréchet'}, \\ \exp(-(-x)^{-1/\gamma}), & x < 0, \gamma < 0 \quad \text{'Weibull'}, \\ \exp(-e^{-x}), & \gamma = 0, x \in R \quad \text{'Gumbel'}. \end{cases}$$

Definition

The parameter γ is called the extreme value index (EVI) and defines the shape of the tail of the r.v. X.
The parameter $\alpha = 1/\gamma$ is called tail index.

1. $H_\gamma(x)$, $\gamma < 0 \hookrightarrow$ short tails with finite right endpoint

2. $H_\gamma(x)$, $\gamma = 0 \hookrightarrow$ exponential decaying tails light-tailed

$$\overline{F}(x) \sim \exp(-x), \qquad x \to +\infty$$

3. $H_\gamma(x)$, $\gamma > 0 \hookrightarrow$ polynomial decaying tails heavy-tailed

$$\overline{F}(x) \sim x^{-\alpha}, \qquad x \to +\infty$$

Distribution of maxima for dependent sequence

The extremal index $\theta \in [0,1]$ allows to extend

classical results those are valid for independent r.v.s to dependent r.v.s.

Theorem (Leadbetter, 1983)

Under the condition $D(u_n)$ it holds

$$\lim_{n\to\infty} P\{(M_n - b_n)/a_n \le x\} = H_\gamma^\theta(x)$$

for some normalizing constants $a_n > 0$ and $b_n \in R$, if

$$\lim_{n\to\infty} P\{(\widetilde{M}_n - b_n)/a_n \le x\} = H_\gamma(x),$$

where $H_\gamma(x)$ is called a generalized extreme value (GEV) distribution (see Gnedenko's theorem).

Limit distribution of maximum of dependent observations

Gnedenko's theorem shows what limit distribution may have

a sequence of independent or weak dependent random variables.

What could be a limit distribution of a sequence of dependent random variables?

An extremal index θ may help to get it.

Tail and extremal indices are the most important characteristics of extreme value theory

The tail index α shows the heaviness of tail.
The extremal index θ shows the dependence in extremes and relates to clusters of exceedances over a high threshold.

Basic results: Pickands theorem

The limit distribution of the excess distribution of the X_i's is necessarily of the Generalized Pareto form

$$\lim_{u \uparrow x_F, u+x<x_F} P(X_1 - u > x | X_1 > u) \to (1+\gamma x)_+^{-1/\gamma}, \qquad x \in R,$$

where

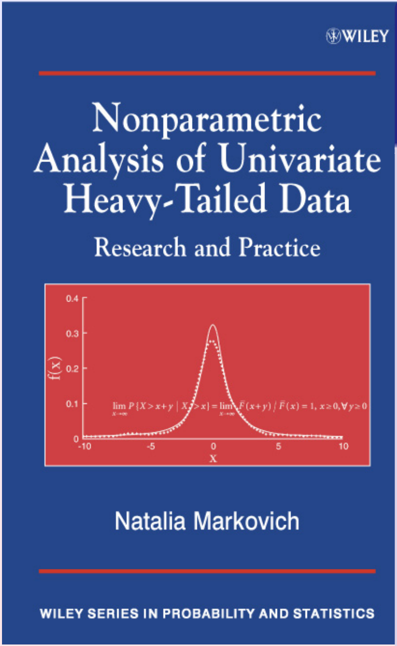
$$x_F = \sup\{x \in R : F(x) < 1\}$$

is the right endpoint of the distribution F and the shape parameter $\gamma \in R$.

Therefore, the Generalized Pareto distribution (GPD) with DF

$$\Psi_{\sigma,\gamma}(x) = \begin{cases} 1-(1+\gamma x/\sigma)_+^{-1/\gamma}, & \gamma \neq 0, \\ 1-\exp(-x/\sigma), & \gamma = 0, \end{cases}$$

where $\sigma \geq 0$, $x \geq 0$ for $\gamma \geq 0$; $0 \leq x \leq -\sigma/\gamma$ for $\gamma < 0$, is often used as a model of the tail of the distribution.



The book provides a survey of classical results and explores the recent developments in the theory of nonparametric estimation of

- the heavy-tailed probability density function and its application to the classification,
- the tail index,
- high quantiles,
- the hazard rate, and
- the renewal function

It demonstrates how to detect heavy tails and dependence, features exercises and examples of applications in teletraffic theory and population analysis.

Пусть $X_1, X_2, ..., X_n$ независимые одинаково распределенные случайные величины.

1. Оценивание хвостового индекса α или индекса экстремальной величины $\gamma = 1/\alpha$.
2. График функции среднего превышения уровня u (the mean excess function) $(u, e(u))$.
3. График отношения максимума к сумме $(n, R_n(p))$,

$$R_n(p) = M_n(p)/S_n(p) = \frac{\max(|X_1|^p, \ldots, |X_n|^p)}{|X_1|^p + \ldots |X_n|^p}, \quad p > 0$$

4. График квантиль-квантиль (QQ-plot).

Estimators of extreme value index (EVI) γ

$$X_{(1)} \le X_{(2)} \le ... \le X_{(n)}$$

are order statistics of the sample $X^n = \{X_1, X_2, ..., X_n\}$

For $\gamma > 0$:

- Hill's estimator

$$\hat{\gamma}^H(n,k) = \frac{1}{k}\sum_{i=1}^{k} \ln X_{(n-i+1)} - \ln X_{(n-k)}$$

- Ratio estimator Goldie, Smith, (1987)

$$a_n = a_n(x_n) = \sum_{i=1}^{n} \ln(X_i/x_n) I\{X_i > x_n\} / \sum_{i=1}^{n} I\{X_i > x_n\},$$

x_n is am arbitrary threshold level, e.g., $x_n = X_{(n-k)}$

The Hill's estimator is biased, i.e. $E\widehat{\gamma}^H(n,k) - \gamma \neq 0$. A bias reduced modification is the generalized Jackknife estimator

$$\widehat{\gamma}_k^{GJ} = 2\widehat{\gamma}_k^{V} - \widehat{\gamma}^H(n,k),$$

where $\widehat{\gamma}^H(n,k)$ is the Hill's estimator of the extreme value index $\gamma = 1/\alpha$,

$$\widehat{\gamma}_k^{V} = \frac{M_{n,k}}{2\widehat{\gamma}^H(n,k)}, \ M_{n,k} = \frac{1}{k}\sum_{i=1}^{k} Y_{(i,k)}^2, \ Y_{(i,k)} = \log\left(\frac{X_{(n-i+1)}}{X_{(n-k)}}\right).$$

is proposed in Gomes et al. (2000)[a]

[a] Gomes, I., Martins, J., Neves, M. Alternatives to a Semi-Parametric Estimator of Parameters of Rare Events - The Jackknife Methodology. Extremes (2000) 3, 207-229

For $\gamma \in R$:

- "Moment estimator Dekkers, Einmahl, de Haan, (1989):

$$\hat{\gamma}^{M}(n,k) = \hat{\gamma}^{H}(n,k) + 1 - 0.5\left(1 - (\hat{\gamma}^{H}(n,k))^2/S_{n,k}\right)^{-1},$$

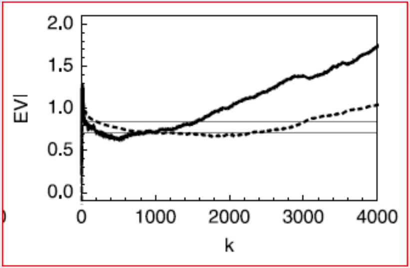
$$S_{n,k} = (1/k)\sum_{i=1}^{k}\left(\ln X_{(n-i+1)} - \ln X_{(n-k)}\right)^2$$

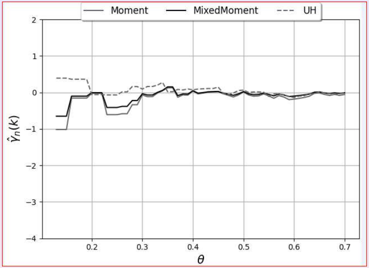
- "UH estimator Berlinet, (1998)

$$\hat{\gamma}^{UH}(n,k) = (1/k)\sum_{i=1}^{k}\ln UH_i - \ln UH_{k+1},\ UH_i = X_{(n-i)}\hat{\gamma}^{H}(n,i)$$

- Pickands's estimator

$$\hat{\gamma}^{P}(n,k) = \frac{1}{\ln 2}\ln\frac{X_{(n-k+1)} - X_{(n-2k+1)}}{X_{(n-2k+1)} - X_{(n-4k+1)}},\quad k \le n/4$$





A heavy-tail distribution

since $\gamma > 0$.

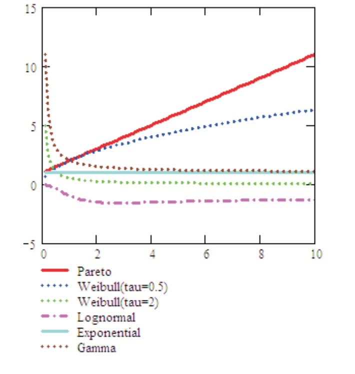
A light-tail distribution

since $\gamma \approx 0$.

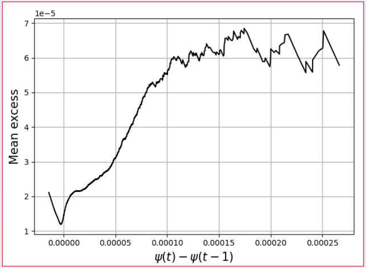
Plot of the mean excess function

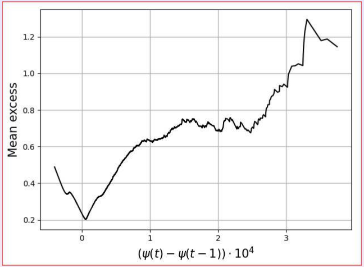
$$e_n(u) = \frac{\sum_{i=1}^{n}(X_i - u)1\{X_i > u\}}{\sum_{i=1}^{n} 1\{X_i > u\}}$$

1. Heavy-tailed distributions: $e(u) \to \infty$, $u \to \infty$.
2. A Pareto distribution: a linear $e(u)$.
3. An exponential distribution: the constant $e(u) = 1/\lambda$.
4. Light-tailed distributions: $e(u) \to 0$, $u \to \infty$.



Examples of mean excess function estimation





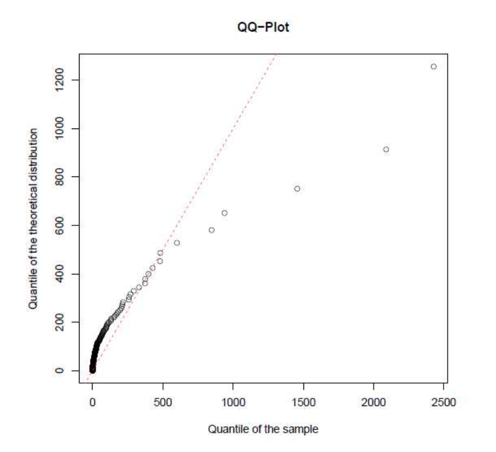
A heavy-tail distribution

since e(u) linearly increases.

A mixed light-heavy-tailed distribution

since e(u) is constant and linear increasing.

QQ plots



QQ plot for a Fréchet distributed sample with $\gamma = 0.5$ against the quantiles of an lognormal distribution with $\mu = 2.5$ and $\sigma = 1.5$.

The QQ plot is below the line $y = x$.

$F^{\leftarrow}_{hyp}(q) < F^{\leftarrow}_{n}(q)$, where $F_n(q)$ and $F_{hyp}(q)$ denote the empirical and model distribution functions, and $F^{\leftarrow}$ denotes the inverse function, or $F_{hyp}(x_q) > F_n(x_q)$, or $\overline{F}_{hyp}(x_q) < \overline{F}_n(x_q)$. The sample has a heavier tail than the model.

Направления, задачи

1. Изучения кластеров превышения уровня для случайных последовательностей (временных рядов), теория и практика.
2. Эволюционирующие случайные сети и моделирующие их случайные графы, экстремумы, кластеры превышения уровня на них. Влияние стратегии эволюционирования, удаления узлов и связей в ходе эволюции на распределения мер влиятельности узлов.
3. Машинное обучение для разбиения на сообщества, распознавания аномальных наблюдений (Machine learning for anomaly detection).
4. Анализ рисков в климатологии, гидрологии, финансах, страховании, медицине и др.

Кластеры превышения уровня

Кластеры превышений стохастической последовательностью высоких уровней

Средняя величина кластера определяется $1/\theta$,
θ - экстремальный индекс

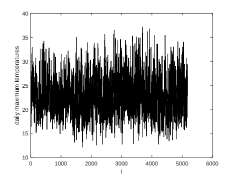
Причины возникновения кластеров

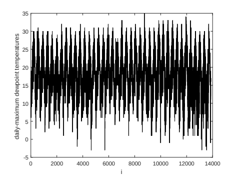
долговременные зависимости в наблюдаемых процессах и тяжелые хвосты распределений процессов

Приложения

гидрология, сейсмология, климатология, финансовые рынки, телекоммуникационные, социальные и энергетические сети ...

Кластеры превышения уровня: примеры

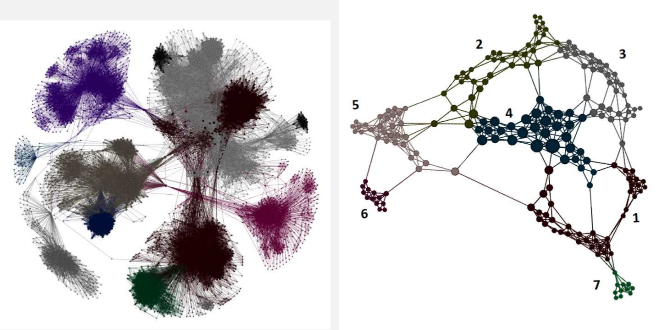




Толпа зевак на пляже Кони-Айленд, собравшихся понаблюдать, как делают искусственное дыхание, 1951 год.

Максимумы температуры по дням в июле в Uccle, Бельгия с размером выборки $n = 5177$ (слева); температура точки росы на станции Dhahran, Саудовская Аравия с $n = 13866$ (середина); кластер случайного графа (справа).

Communities (clusters) in random graphs



Facebook graph (left); Simulated geometric graph (right). Communities can be determined as clusters (classes).

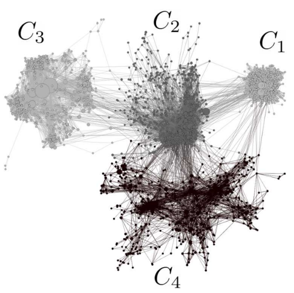
Communities in graphs

The community is a set of nodes

highly connected between each other and weakly connected with nodes of other communities.

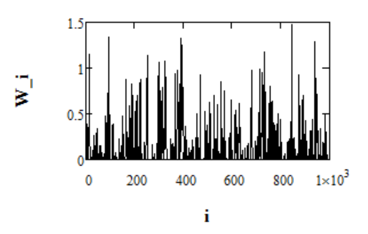
Community detection

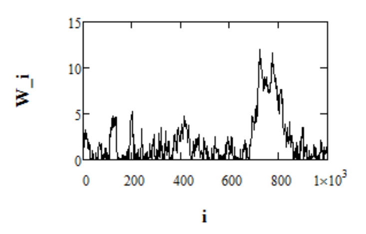
is provided here by Directed Louvain's Algorithm proposed, a modification of classical Louvain's Algorithm for directed graphs.



The Berkeley-Stanford seed network

divided into communities $\{C_i\}$, where point sizes are proportional to the nodes' PRs.



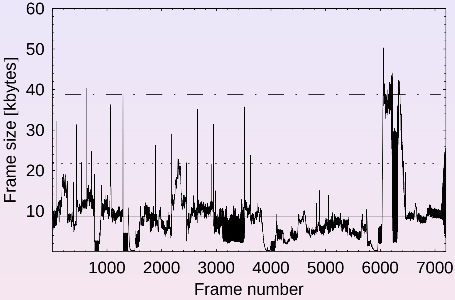


Lindley time series for exponentially distributed service time B_n of customer n and the inter-arrival time A_n with exponential intensities $(\mu, \lambda) = (5, 2)$ (left) and $(\mu, \lambda) = (2.5, 2)$ (right).

Lindley process
$W_{n+1} = (W_n + D_n)^+ = \max(0, W_n + D_n)$, $n = 0, 1, \ldots,$ W_n is the waiting time of the nth customer until he is served;
$D_n = (B_n - A_n)^+$, B_n is the service time of customer n, A_n is the inter-arrival time between customer n and $n + 1$.

Clusters of exceedances: Video traffic analysis

Slice-based encoded MPEG4 video data are analyzed



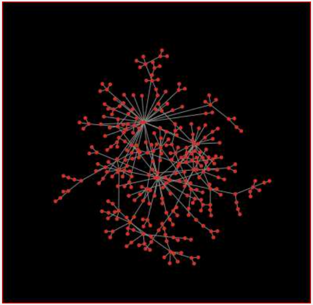
The frame-sizes of the slice-based encoded video stream,

together with the mean frame size u =8.781 Kbytes (solid horizontal line) and the 95% (dotted) and 99% (dot-dashed) empirical quantiles.

Предпочтительное присоединение (preferential attachment (PA))

Новый узел с большей вероятностью присоединяется к существующим наиболее влиятельным узлам, которые быстро превращаются в узлы-гиганты с большим числом связей (node degrees).

Распределение числа связей узлов имеет тяжелый хвост (power-law):
$P(X = k) \sim Ck^{-(1+\alpha)}, k \to \infty,$
for some positive constants C and α.



A new edge to an existing graph G_{n-1} is generated by one of three scenarios with probabilities α, β, γ.

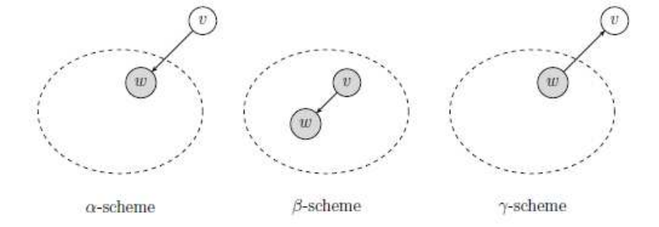
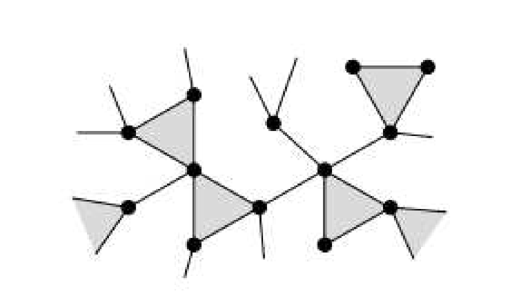


Рис.: The α, β, γ schemes of linear PA attachment of new edges.

Кластерное присоединение (clustering attachment)

Новый узел присоединяется к существующему узлу i с числом треугольников связей Δ_i среди его соседних узлов, с вероятностью, пропорциональной кластерному коэффициенту $c_i = 2\Delta_i/(k_i(k_i - 1))$ узла i, k_i - число связей узла i. Распределение k_i имеет легкий хвост.



Распознавание аномальных наблюдений

Предполагается, что в обучающей выборке присутствуют элементы не всех возможных классов. Аномальные наблюдения редкие и могут появиться в будущем.

Vignotto & Engelke (2020) Тестирование гипотезы:

$H_0 : x_0$ is normal
$H_1 : x_0$ is abnormal

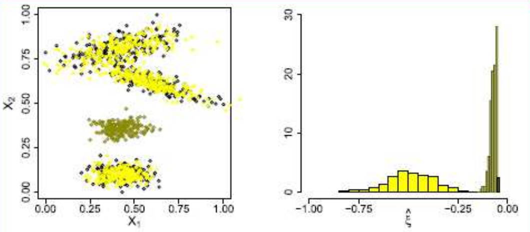


Fig. 1 Left panel: the simulated dataset for the toy example: training data (in black), the normal (bright yellow) and abnormal examples (dark yellow) from the test set. Right panel: the $\hat{\xi}_n$ estimates for normal (bright yellow) and abnormal test data (dark yellow) for the simulated dataset

Обзор: Wang, J., Lu, S., Wang, SH., Zhang, Y.-D. A review on extreme learning machine. Multimedia Tools and Applications 2022, 81, 41611–41660

Анализ рисков в финансах

1. Оценивание Value-at-Risk at probability level $p \in (0,1)$[a]

$$\mathrm{VaR}_p = F_x^{-1}(p) = \inf\{x_p \in R : P\{X > x_p\} \le 1 - p\}$$

2. Оценивание Expected shortfall at confidence level $p \in (0,1)$ [b]

$$\mathrm{ES}_p = E\{X | X \ge \mathrm{VaR}_p\}$$

3. Моделирование и прогноз волатильности:
P_t, $t = 0, 1, 2, \ldots$ (time units) are financial data,

$$X_t = \log P_t - \log P_{t-1} = \log\left(1 + \frac{P_t - P_{t-1}}{P_{t-1}}\right)$$

(нелинейные модели GARCH).[c]

---

[a] I. Gomes et al. (2015, 2017, 2020), A.McNeil

[b] John Einmahl

[c] T. Mikosch (2002) Modeling Dependence and Tails of Financial Time Series

Modeling clusters of extreme values

- We study clusters of rare events, i.e. conglomerates of exceedances of the process over a threshold.
- Main theorem: geometric-like asymptotical distributions of the cluster and inter-cluster sizes.
- Distribution tails of the duration of clusters and return intervals between clusters.

[1]

[1] Markovich, Extremes 2014, 2016, 2017

Problem

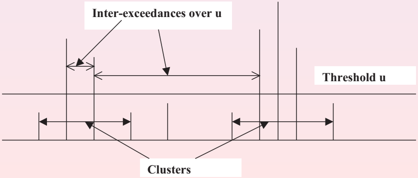
Let $\{R_n : n \geq 1\}$ be a stationary sequence of r.v.s with marginal cdf $F(x)$ and the extremal index θ, $M_n = \max\{R_1, ..., R_n\}$.

The number of inter-arrival times (IATs) between events arising between two consequent exceedances of the process $\{R_n\}_{n\geq 1}$ over u

$$T_1(u) = \min\{j \geq 1 : M_{1,j} \leq u, R_{j+1} > u | R_1 > u\}$$

The number of IATs between two consecutive non-exceedances

$$T_2(u) = \min\{j \geq 1 : L_{1,j} > u, R_{j+1} \leq u | R_1 \leq u\}$$



$$M_{1,j} = \max\{R_2, ..., R_j\},$$
$$M_{1,1} = -\infty$$

$$L_{1,j} = \min\{R_2, ..., R_j\},$$
$$L_{1,1} = +\infty$$

Problem (Cont.)

In practice we exclude the cases $T_1(u) = 1$ and $T_2(u) = 1$ since they correspond to single IATs between consecutive events $\{R_i\}$:

$$T_1^*(u) = \min\{j > 1 : M_{1,j} \le u, R_{j+1} > u | R_1 > u\},$$

$$T_2^*(u) = \min\{j > 1 : L_{1,j} > u, R_{j+1} \le u | R_1 \le u\}.$$

Important cluster characteristics

- Return interval: the time between two consecutive exceedances
- Duration of a cluster: the time between two consecutive nonadjacent non-exceedances
- First hitting time: the minimal time to reach a first exceedance

Let $\{R_n\}_{n\geq 1}$ be a stationary process with the extremal index $\theta \in (0,1]$.
Let $\{x_{\rho_n}\}$ be a sequence of quantiles of R_1 of the levels $\{1-\rho_n\}$, that satisfies the conditions $\lim_{n\to\infty} n(1-F(x_{\rho_n})) = \tau$ and $\lim_{n\to\infty} P\{M_n \leq x_{\rho_n}\} = e^{-\tau\theta}$ for each $0 < \tau < \infty$ and, $q_n = 1-\rho_n$.

Under specific mixing conditions

$$|P\{T_1(x_{\rho_n}) = j\}/(\theta^2 \rho_n (1-\rho_n)^{(j-1)\theta}) - 1| < \varepsilon,$$
$$|P\{T_2(x_{\rho_n}) = j\}/(\theta^2 q_n (1-q_n)^{(j-1)\theta}) - 1| < \varepsilon$$

for all n and j sufficiently large.

$P\{T_1(x_{\rho_n}) = j\}$ of ARMAX and MM processes and geometric approximation, Markovich 2014

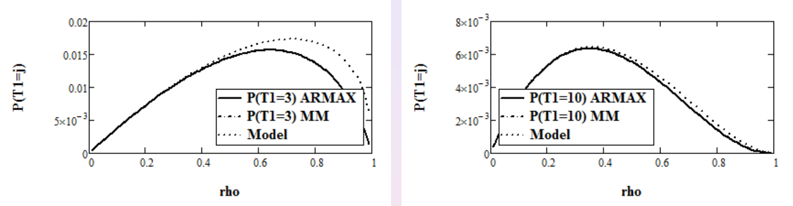


Рис.: The distribution of $T_1(u)$ of the ARMAX process (solid line), the MM process (dashed line) and the geometric-like model $\theta^2\rho(1-\rho)^{(j-1)\theta}$ (points) against ρ for $j = 3$ (left) and $j = 10$ (right) and $\theta = 0.2$.

The ARMAX and MM processes have the same $P\{T_1(x_{\rho_n}) = j\}$.

Maxima and Sums of Non-Stationary Random Length Sequences (M. & Rodionov 2020)

Let $\{Y_{n,i} : n, i \geq 1\}$ be a double-indexed array of nonnegative r.v.s:

- the "row index"n corresponds to time, and the "column index"i enumerates the series.
- The length N_n of "row"sequences $\{Y_{n,i} : i \geq 1\}$ for each n is generally random.
- For each i the "column"sequence $\{Y_{n,i} : n \geq 1\}$ is assumed to be strict-sense stationary with extremal index θ_i having a regularly varying distribution tail

$$P\{Y_{n,i} > x\} = \ell_i(x)x^{-k_i}$$

with tail index $k_i > 0$ and a slowly varying function $\ell_i(x)$.
- There are no assumptions on the dependence structure in i.

The matrix (1) of the array $\{Y_{n,i} : n, i \geq 1\}$ and the matrix (2) of tail indices $\{k_i\}$ and extremal indices $\{\theta_i\}$ of its "column" series:

$$\begin{pmatrix} cY_{1,1} & cY_{1,2} & cY_{1,N_1} & \dots & 0 & 0 \\ cY_{2,1} & 0 & cY_{2,3} & \dots & cY_{n,N_2} & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & cY_{n,2} & cY_{n,3} & \dots & 0 & cY_{n,N_n} \end{pmatrix}, \qquad (1)$$

$$\begin{pmatrix} k_1 & k_2 & k_3 & \dots & \dots & k_{N_n} \\ \theta_1 & \theta_2 & \theta_3 & \dots & \dots & \theta_{N_n} \end{pmatrix}. \qquad (2)$$

Sums and maxima over "row" random-length sequences are considered.

Weighted sums and maxima

$$Y_n^*(z, N_n) = \max(z_1 Y_{n,1}, ..., z_{N_n} Y_{n,N_n}),$$
$$Y_n(z, N_n) = z_1 Y_{n,1} + ... + z_{N_n} Y_{n,N_n}, \quad z_1, z_2, ... > 0$$

Obtained results:

- If the "column" series with minimum tail index k_1 is unique, then $Y_n^*(z, N_n)$ and $Y_n(z, N_n)$ have the tail index k_1 and extremal index θ_1, Markovich, Rodionov, 2020.
- If there are a random number d of the "column" series with minimum tail index k_1 and they are independent or weakly dependent, then $Y_n^*(z, N_n)$ and $Y_n(z, N_n)$ have the tail index k_1; their extremal index is the same as for the "column" with the largest maximum among d "columns", Markovich, 2022b.

Other results

- Extremal properties of evolving networks: local dependence and heavy tails. Markovich (2023)
- Threshold selection for extremal index estimation by discrepancy method. Markovich, Rodionov (2023)

References

1. Markovich N.M. Modeling clusters of extreme values. Extremes, 17(1), 97–125, 2014.
2. Markovich N.M. Erratum to: modeling clusters of extreme values. Extremes, 19(1), 139–142, 2016.
3. Markovich N.M. Clusters of extremes: modeling and examples. Extremes, 20, 519–538, 2017.
4. Markovich N.M., Rodionov I.V. Threshold selection for extremal index estimation. In Appear in Journal of Nonparametric Statistics arXiv:2009.02318v1

References

1. Markovich, N. M., Rodionov, I. V. Maxima and Sums of Non-Stationary Random Length Sequences. Extremes, 2020, 23(3), 451-464
2. Markovich, N. M., Ryzhov M.S., Vaičiulis M. Tail Index Estimation of PageRanks in Evolving Random Graphs. Mathematics (Q1) 2022, 10(16), 3026.
3. Markovich, N.M. Extremal properties of evolving networks: local dependence and heavy tails. Annals of Operation Research (Q1) DOI: 10.1007/s10479-023-05175-y
4. Markovich N.M., Vaičiulis M. Extreme Value Statistics for Evolving Random Networks. Mathematics 2023, 11(9), 2171
5. Markovich, N.M. Weighted maxima and sums of non-stationary random length sequences in heavy-tailed models. arXiv: 2209.08485v [math.ST] 18 Sep 2022

Контакты:

Маркович Наталья Михайловна, док. физ.-мат. наук

Институт проблем управления им. В.А.Трапезникова РАН
Лаборатория 38, комн. 429

Тел. (495) 334-88-20, 1-429

E-mail: nat.markovich@gmail.com