Об оценке параметров сдвига и масштаба хвостов распределений

П. И. АХТЯМОВ
Московский физико-технический институт
(*государственный университет*)
e-mail: akhtyamovpavel@gmail.com

И. В. РОДИОНОВ
Институт проблем управления им. В. А. Трапезникова РАН,
Московский физико-технический институт
(*государственный университет*)
e-mail: vecsell@gmail.com

УДК 519.218

Ключевые слова: параметр сдвига, параметр масштаба, порядковые статистики, хвост распределения, область притяжения Гумбеля.

Аннотация

Предложены асимптотически нормальные оценки параметров сдвига и масштаба хвоста функции распределения, принадлежащей области максимального притяжения Гумбеля или Фреше, по k максимальным членам вариационного ряда выборки. Для распределений из области максимального притяжения Гумбеля настоящая задача рассмотрена впервые.

Abstract

P. I. Akhtyamov, I. V. Rodionov, On estimation of the scale and location parameters of distribution tails, Fundamentalnaya i prikladnaya matematika, vol. 23 (2020), no. 1, pp. 25—49.

Estimators of the location and scale parameters are proposed for tails of distributions belonging to the Gumbel or Frechèt maximum domain of attraction using only higher order statistics of the sample. The related problem for the Gumbel domain of attraction is considered for the first time.

1. Введение

Пусть $X = (X_1, \ldots, X_n)$ — независимые одинаково распределённые случайные величины с непрерывной функцией распределения F. Обозначим $x_F^* = \sup\{x\colon F(x) < 1\}$, далее полагаем $x_F^* = +\infty$. Будем говорить, что семейство одномерных функций распределения $\{F_a,\ a \in A\}$ обладает *параметром сдвига*, если для каждых $a \in A$, $x \in \mathbb{R}$ справедливо $F_a(x) = F_0(x - a)$. Семейство функций распределения $\{F_b,\ b \in B\}$ обладает *параметром масштаба*, если для каждых $b \in B$, $x \in \mathbb{R}$ справедливо $F_b(x) = F_1(x/b)$.

Фундаментальная и прикладная математика, 2020, том 23, № 1, с. 25—49.

В настоящей работе рассматривается задача оценивания параметров сдвига и масштаба хвоста функции распределения F по максимальным членам вариационного ряда выборки X. Указанная задача для распределений из области максимального притяжения Гумбеля рассматривается впервые. Состоятельность и асимптотическая нормальность оценки параметра сдвига доказаны для распределений с хвостами легче, чем у экспоненциального распределения. Для параметра масштаба указанные свойства установлены для всех распределений из областей максимального притяжения Гумбеля и Фреше (определения см. ниже). Асимптотические свойства оценок будут проиллюстрированы результатами численного моделирования.

В статистике часто возникает задача оценки функции распределения и её параметров по неполным или цензурированным данным, в частности в областях, связанных со страхованием, с задачами надёжности, телекоммуникациями, медициной и computer science. Задача, когда известны лишь наблюдения, которые выше какого-то заранее заданного порога, хорошо изучена, в частности, классической является оценка Каплана—Мейера [21] хвоста функции распределения (см. [3, 10, 24]). С другой стороны, статистика экстремумов утверждает (см. [16]), что для оценки хвоста функции распределения могут быть использованы только максимальные члены вариационного ряда выборки, тогда как центральные значения выборки могут быть промоделированы с помощью стандартных статистических процедур.

Центральным результатом стохастической теории экстремумов является теорема Фишера—Типпета—Гнеденко (см. [15]). Она утверждает, что если существуют такие последовательности констант $a_n > 0$ и b_n, что функция распределения нормированного максимума $M_n = \max(X_1, \ldots, X_n)$ стремится к некоторой невырожденной функции распределения G, т. е.

$$\lim_{n\to+\infty} P(M_n \leqslant a_n x + b_n) = G(x), \tag{1}$$

то существуют такие константы $a > 0$ и b, что $G(ax + b) = G_\gamma(x)$, где

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

$\gamma \in \mathbb{R}$, и при $\gamma = 0$ правую часть следует понимать как $\exp(-e^{-x})$. Параметр γ называется индексом экстремального значения [16]. Говорят, что функция распределения F выборки $(X_1, \ldots, X_n)$ принадлежит области максимального притяжения Фреше, если соотношение (1) выполнено для $\gamma > 0$, области максимального притяжения Вейбулла, если соотношение (1) выполнено для $\gamma < 0$, и области максимального притяжения Гумбеля (пишем $F \in D(G_\gamma)$ для всех трёх областей), если соотношение (1) выполнено для $\gamma = 0$.

В большинстве работ, связанных с оцениванием поведения хвоста функции распределения, так или иначе используются оценки индекса экстремального значения γ (см. подробнее [16]). В этой связи упомянем [4, 8, 11, 12, 19, 22, 23] и многие другие работы. Иной подход состоит в оценивании хвостов распределений напрямую с помощью максимальных членов вариационного ряда (см. [9, 18, 28]). Популярен также подход получения оценок индекса γ в случае $k/n \to C$, где $C > 0$ (POT-метод, см. [3]).

С помощью оценок индекса экстремального значения γ можно оценить хвост функции распределения, принадлежащей областям максимального притяжения Фреше и Вейбулла. В частности, для хвоста функции распределения из области притяжения Фреше является состоятельной оценка

$$\big(\bar{F}(x)\big)_n = \frac{k}{n}\left(\frac{x}{X_{(n-k)}}\right)^{-1/\hat{\gamma}},$$

где $\hat{\gamma}$ — некая состоятельная оценка индекса экстремального значения, а $X_{(1)} \leqslant \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}$ — вариационный ряд выборки $X_1, \ldots, X_n$ (подробнее см. [16]). Но этот метод неприменим для огромного класса распределений, к примеру принадлежащих области максимального притяжения Гумбеля (поскольку $\gamma = 0$ и оценки хвоста функции распределения, полученные с помощью оценок индекса экстремального значения, будут крайне неточными) или так называемых распределений с супертяжёлыми хвостами [3] (параметр γ не определён).

Важными классами распределений, принадлежащих области максимального притяжения Гумбеля, являются вейбулловский и лог-вейбулловский классы распределений. Будем говорить, что функция распределения F принадлежит вейбулловскому классу распределений, если существует такое $\theta > 0$, что для всех $\lambda > 0$ выполнено

$$\lim_{x\to+\infty} \frac{\ln\big(1-F(\lambda x)\big)}{\ln\big(1-F(x)\big)} = \lambda^{\theta}.$$

Параметр θ также называют вейбулловским хвостовым индексом. Вейбулловскому классу распределений принадлежат, в частности, нормальное, экспоненциальное, гамма- и другие важные для статистики распределения. Если для некоторой функции распределения F функция распределения $F(e^x)$ принадлежит вейбулловскому классу с $\theta > 1$, то говорят, что F принадлежит лог-вейбулловскому классу распределений.

Несмотря на то что во многих прикладных задачах (наиболее часто в задачах надёжности и теории страхования) требуется оценивать хвосты распределений (т. е. вероятности разорения) из области максимального притяжения Гумбеля, работ, посвящённых оценке параметров хвостов распределений, принадлежащих этой области, не так много. В [14] предложены оценки вейбулловского и лог-вейбулловского хвостовых индексов, построенные с помощью обобщения известной оценки Хилла [20]. Оценке вейбулловского индекса также посвящены работы [2, 5, 7, 26, 27].

Работа устроена следующим образом. Основные результаты, а также примеры выполнения условий теорем работы сформулированы в разделе 2. Раздел 3 посвящён численному моделированию поведения полученных оценок и демонстрации их асимптотических свойств. Леммы и их доказательства вынесены в раздел 4, доказательства основных результатов приведены в разделе 5.

2. Основные результаты. Примеры

Обозначим через $X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}$ вариационный ряд случайных величин $X_1, \ldots, X_n$. Предположим, что функция распределения F случайной величины X_1 имеет третью производную и дважды непрерывно дифференцируема. Определим $x_F^* = \inf\{x\colon F(x) = 1\}$. Далее будет достаточно рассматривать только распределения с $x_F^* = +\infty$. Действительно, если $x_G < +\infty$ для некоторой функции распределения $G(x)$, то $H(y) = G(x_G^* - 1/y)$ является функцией распределения с $x_H^* = +\infty$.

Пусть $F \in D(G_\gamma)$. В этом случае [16] F, с учётом существования у неё второй производной, удовлетворяет условию фон Мизеса

$$\lim_{t \to x_F^*} \frac{\big(1 - F(t)\big)F''(t)}{\big(F'(t)\big)^2} = -\gamma - 1. \tag{2}$$

Более того, известно, что если функция распределения F удовлетворяет условию фон Мизеса с параметром γ, то $F \in D(G_\gamma)$ (см. подробнее [16]).

Далее будем считать, что функция распределения F принадлежит семейству распределений $\{F_{a,b}(x),\ a \in \mathbb{R},\ b > 0\}$, где a и b — параметры сдвига и масштаба соответственно, т. е. существуют такие a_0 и b_0, что

$$F\left(\frac{x-a}{b}\right) = F_{0,1}(x).$$

В дальнейшем вместо $F_{0,1}(x)$ пишем $F_0(x)$. Легко убедиться, что F_0 удовлетворяет условию фон Мизеса с тем же значением параметра γ, что и F.

Введём $Y_1, \ldots, Y_n$ — выборку из распределения с функцией распределения F_0. Очевидно, что для каждого $l \in \{1, \ldots, n\}$ выполнено $X_l \stackrel{d}{=} bY_l + a$. Рассмотрим случай $b = 1$. Легко видеть, что при условии $E|X_1| < +\infty$ оценка

$$\bar{a} = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} Y_i$$

является несмещённой и состоятельной оценкой параметра сдвига a. По аналогии с предыдущей оценкой рассмотрим

$$\hat{a} = \frac{1}{k}\sum_{i=0}^{k-1} X_{(n-i)} - \frac{1}{k}\sum_{i=0}^{k-1} Y_{(n-i)}.$$

Как показано в следующей теореме, при некоторых условиях на последовательность $k = k(n)$, если семейства распределений $\{F_{a,b}(x),\ a \in \mathbb{R},\ b > 0\}$ принадлежат области притяжения Гумбеля, оценка является состоятельной оценкой параметра a.

Теорема 1. *Предположим, что функция распределения F_0 удовлетворяет условию фон Мизеса* (2) *при $\gamma = 0$. Пусть последовательность $k = k(n)$ такова, что*

$$k \to +\infty, \quad \frac{k}{n} \to 0 \ \ \text{при} \ \ n \to +\infty \tag{3}$$

и при этом существует такое $M > 0$*, что*

$$\limsup_{n\to+\infty} \frac{1 - F_0(M\sqrt{k})}{k/n} \leqslant 1.$$

Тогда оценка $\hat{a}$ *является состоятельной оценкой параметра сдвига* a.

Теперь рассмотрим случай $a = 0$. Тогда оценка

$$\hat{b} = \exp\left\{\sum_{i=0}^{k-1} \frac{\ln X_{(n-i)}}{k} - \sum_{i=0}^{k-1} \frac{\ln Y(n-i)}{k}\right\}$$

при некоторых условиях является состоятельной оценкой параметра масштаба b, как показано в следующей теореме.

Теорема 2. *Пусть функция распределения* F_0 *удовлетворяет условию фон Мизеса для* $\gamma \geqslant 0$. *Пусть последовательность* $k = k(n)$ *удовлетворяет условиям* (3)*, причём существует такое* $M > 1$*, что*

$$\limsup_{n\to+\infty} \frac{1 - F_0(\exp\{M\sqrt{k}\})}{k/n} \leqslant 1.$$

Тогда оценка $\hat{b}$ *является состоятельной оценкой параметра масштаба* b.

Обозначим

$$U(x) = \inf_t \left\{t\colon\ \frac{1}{1 - F_0(t)} = x\right\}.$$

Функция $U(x)$ играет огромную роль в стохастической теории экстремумов, в частности в задачах, связанных со скоростью сходимости в предельной теореме Гнеденко [17]. Для того чтобы введённые нами оценки параметров сдвига и масштаба были асимптотически нормальными, необходимо потребовать выполнения более жёстких условий на функцию распределения F_0. Ниже мы приведём примеры, показывающие, что хвосты многих используемых в статистике распределений удовлетворяют этим условиям. Обозначим $T(x) = -\ln\bigl(1 - F_0(x)\bigr)$. Наложим на функцию $T(x)$ следующие условия регулярности.

(R1) $T(x) \in C^{(3)}$ для каждого $x > x_0$, причём $T(x)$, $T'(x)$ и $T''(x)$ монотонны на бесконечности и имеют там предел (возможно, бесконечный). Дополнительно для $m \in \{1, 2, 3\}$ существуют пределы

$$a_m := \lim_{x\to+\infty} \frac{T'(x)}{x^m}.$$

(R2) Выполнено

$$\lim_{x\to+\infty} \frac{\ln\bigl|\bigl(T(\sqrt{x})\bigr)''\bigr|}{T(\sqrt{x})} = 0.$$

Для таких функций распределений F_0 верна следующая теорема.

Теорема 3. *Пусть для функции распределения F_0 выполнено условие фон Мизеса при $\gamma = 0$, а последовательность $k = k(n)$ удовлетворяет условиям* (3). *Если для функции T выполнены условия* (R1), (R2) *и $\sqrt{k}T'\big(U(n/k)\big) \to +\infty$ при выполнении* (3), *то*

$$\sqrt{k}T'\left(U\left(\frac{n}{k}\right)\right)(\hat{a}-a) \xrightarrow{d} \mathcal{N}(0,4), \quad n \to +\infty. \tag{4}$$

Введём функцию $Z(x) = -\ln\big(1 - F_0(e^x)\big)$. В следующей теореме указаны условия, в которых оценка $\hat{b}$ является асимптотически нормальной оценкой параметра масштаба b.

Теорема 4. *Пусть для функции распределения F_0 выполнено условие фон Мизеса при $\gamma \geqslant 0$, а последовательность $k = k(n)$ удовлетворяет условиям* (3). *Если функция $Z(x)$ удовлетворяет условиям регулярности* (R1), (R2) *и*

$$\sqrt{k}Z'\left(\ln\left(U\left(\frac{n}{k}\right)\right)\right) \to +\infty, \quad n \to +\infty,$$

то

$$\sqrt{k}Z'\left(\ln\left(U\left(\frac{n}{k}\right)\right)\right)(\hat{b}-b) \xrightarrow{d} \mathcal{N}(0,4b^2), \quad n \to +\infty.$$

Укажем теперь семейства распределений, для которых условия теорем 3 и 4 будут выполняться.

Пример 1. Пусть функция $T(x)$ такова, что $T(x)/x \to +\infty$ при $x \to +\infty$ (т. е. хвост функции распределения F_0 легче, чем у экспоненциального закона) и $T(x)$ трижды дифференцируема. В этом случае условия теоремы 3 выполнены.

Действительно, поскольку $T'(x) \to +\infty$ при $x \to +\infty$, получаем, что $\sqrt{k}T'\big(U(n/k)\big) \to +\infty$ при $n \to +\infty$.

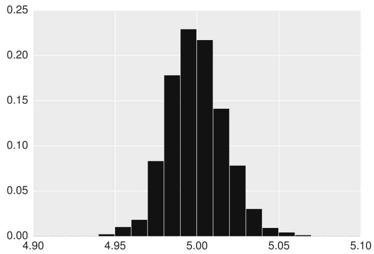
Пример 2. Пусть функция $T(x)$ такова, что $T(x)/x^{\alpha} \to +\infty$ при $x \to +\infty$ и $T(x)$ трижды дифференцируема. В этом случае условия теоремы 3 выполнены для последовательности $k = k(n)$, которая растёт не быстрее, чем n^{α}, $\alpha \in (0,1)$. Доказательство этого утверждения дано в разделе 5.5.

Легко видеть, что оценка $\hat{b}$ параметра масштаба b получена как оценка параметра сдвига функции распределения $F(e^x)$, поэтому приведённые примеры легко переносятся на случай оценивания параметра масштаба. В частности, мы получаем, что если F трижды дифференцируема, то оценка $\hat{b}$ является асимптотически нормальной оценкой параметра масштаба b для всех распределений из области максимального притяжения Гумбеля с $x_F^* = +\infty$.

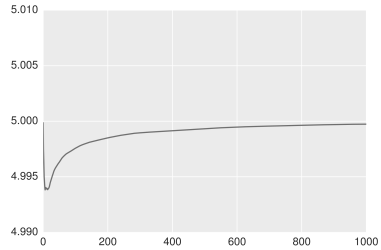
3. Моделирование

Настоящий раздел призван продемонстрировать асимптотические свойства изучаемых оценок. Рассмотрим поведение статистики $\hat{a}$ для выборок из следующих распределений:

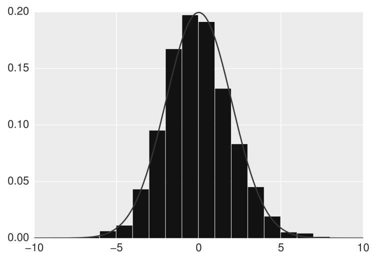
1) $X_1 \sim \mathcal{N}(a,1)$, параметр сдвига $a = 5$ (рис. 1);



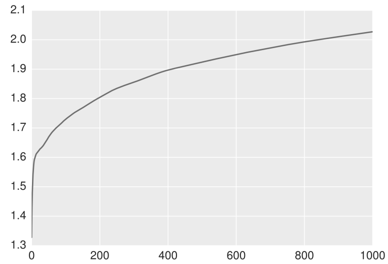
Гистограмма значений $\hat{a}$



Поведение среднего оценки $\hat{a}$



Гистограмма предельного закона $\hat{a}$



Стандартное отклонение

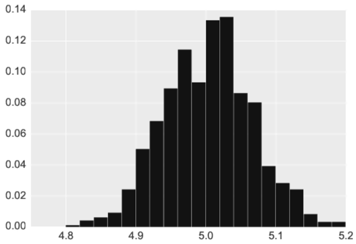
Рис. 1. Данные по выборке из нормального распределения, параметр сдвига $a = 5$

2) X_1 имеет экспоненциальную функцию распределения $F(x) = (1-e^{-(x-5)})\cdot I(x > 5)$ (рис. 2)
3) X_1 имеет функцию распределения типа Вейбулла $F(x) = (1 - e^{-\sqrt{x-5}})\cdot I(x > 5)$ (рис. 3).

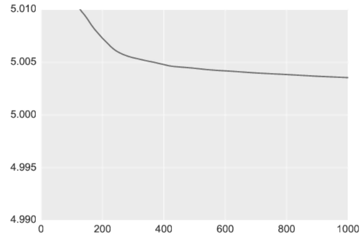
По умолчанию во всех экспериментах моделируем $m = 1000$ выборок размера $n = 1\,000\,000$ из каждого распределения, меняем значение k от 1 до 1000. Как видно из первых двух графиков для каждого распределения, оценка $\hat{a}$ близка к несмещённой. Кроме того, при $k = 1000$ она довольно точно оценивает параметр сдвига для экспоненциального и нормального распределений, для распределения вейбулловского типа наблюдается существенный разброс значений статистики. Построим в каждом случае гистограмму значений статистики (4). Для случаев нормального, экспоненциального распределений и распределений типа Вейбулла

1) среднее стандартное отклонение статистики (4) — 2,028, среднее значение статистики $\hat{a}$ — 5;
2) среднее стандартное отклонение — 2,034, среднее значение $\hat{a}$ — 4,99;
3) среднее стандартное отклонение — 2,523, среднее значение $\hat{a}$ — 5.

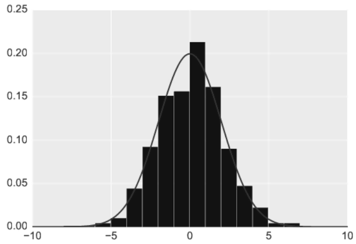
Из графиков значений оценки параметра сдвига видно, что в рассмотренных распределениях гистограмма похожа на график плотности нормального распределения. Дополнительно полученные значения дисперсии предельного закона близки к теоретическим.



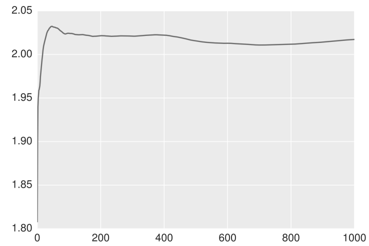
Гистограмма значений $\hat{a}$



Поведение среднего оценки $\hat{a}$

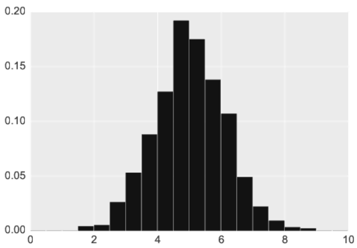


Гистограмма предельного закона $\hat{a}$

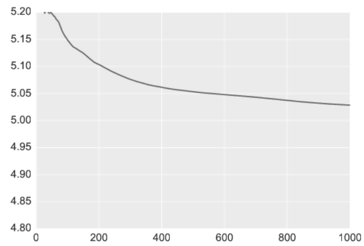


Стандартное отклонение

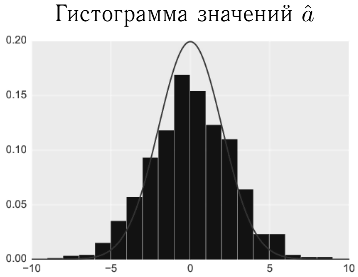
Рис. 2. Данные по выборке из экспоненциального распределения, параметр сдвига $a = 5$



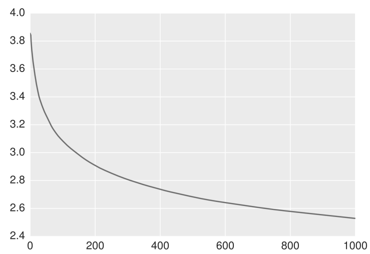
Гистограмма значений $\hat{a}$



Поведение среднего оценки $\hat{a}$



Гистограмма предельного закона $\hat{a}$

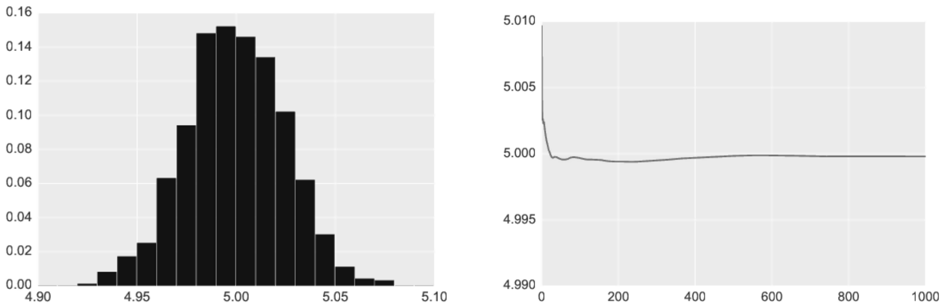


Стандартное отклонение

Рис. 3. Данные по выборке из распределения $F(x) = 1 - e^{-\sqrt{x-a}}$, параметр сдвига $a = 5$

Аналогично предыдущему случаю построим графики выборочного среднего и гистограммы значений оценки параметра масштаба для следующих распределений:

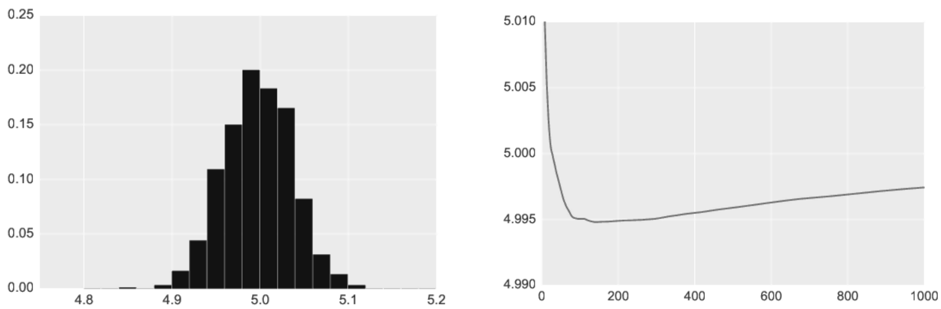
1) $X_l \sim \mathcal{N}(0, 25)$, параметр масштаба $b = 5$ (рис. 4);



Гистограмма значений $\hat{b}$ Поведение среднего оценки $\hat{b}$

Рис. 4. Данные по выборке из распределения $N(0, b)$, параметр $b = 5$

2) $X_l \sim \mathrm{Exp}(5)$ с параметром масштаба $b = 5$ (рис. 5);

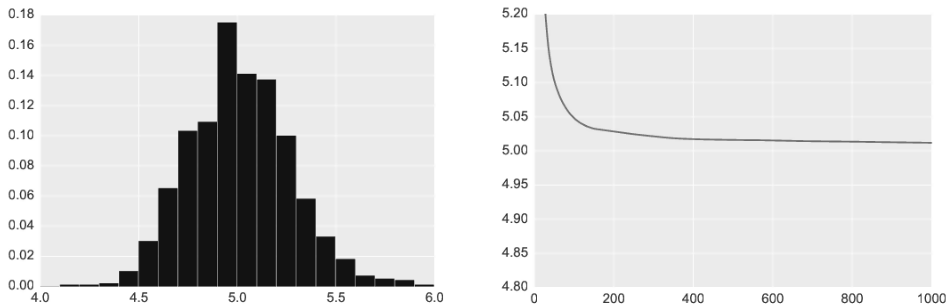


Гистограмма значений $\hat{b}$ Поведение среднего $\hat{b}$

Рис. 5. Данные по выборке из распределения $\mathrm{Exp}(1/b)$, параметр $b = 5$

3) $X_l \sim \mathrm{Pareto}(1.2)$, параметр масштаба $b = 5$ (рис. 6).

Как и для оценок параметра сдвига, гистограммы значений $\hat{b}$ при $k = 1000$ похожи на гистограммы плотности нормального распределения.



Гистограмма значений $\hat{b}$ Поведение $\hat{b}$

Рис. 6. Данные по выборке с функцией распределения $F(x) = (1 - b/x)I(x > b)$, параметр $b = 5$

4. Леммы

Прежде всего заметим, что из условия фон Мизеса (2) вытекает

$$\lim_{t \to +\infty} -\frac{T''(t)}{\big(T'(t)\big)^2} = -\gamma.$$

Поскольку введённые нами оценки зависят только от максимальных членов вариационного ряда выборки, мы не можем сразу использовать независимость рассматриваемых случайных величин $(X_1, \ldots, X_n)$. Поэтому будем рассматривать условное распределение оценок при условии $X_{(n-k)} = q$, используя следующую лемму.

Лемма 1 [16, лемма 3.4.1]. *Пусть $X, X_1, \ldots, X_n$ — независимые одинаково распределённые случайные величины с функцией распределения F и $X_{(1)} \leqslant \ldots \leqslant X_{(n)}$ — вариационный ряд данной выборки. Для всех $k = 1, \ldots, n-1$ совместное условное распределение статистик $\{X_{(i)}\}_{i=n-k+1}^{n}$ при условии $X_{(n-k)} = q$ совпадает с (безусловным) совместным распределением множества $\{X^*_{(i)}\}_{i=1}^{k}$ порядковых статистик независимых одинаково распределённых случайных величин $\{X^*_i\}_{i=1}^{k}$, имеющих функцию распределения*

$$F_q(x) = P(X \leqslant x \mid X > q) = \frac{F(x) - F(q)}{1 - F(q)}, \quad x > q.$$

Для доказательства результатов работы будут применяться классические предельные теоремы для сумм случайных величин при условии, что $(n-k)$-й член вариационного ряда выборки равен q. Таким образом, предельные распределения в указанных теоремах будут зависеть от значения $X_{(n-k)}$, и необходимо знать асимптотику распределения данной порядковой статистики. Этот вопрос решается с помощью следующей леммы.

Лемма 2 [16, теорема 2.4.1]. *Пусть функция распределения F удовлетворяет условию фон Мизеса. Тогда в условиях* (3) *выполнено*

$$\sqrt{k}\frac{X_{(n-k)} - U(n/k)}{(n/k)U'(n/k)} \xrightarrow{d} \mathcal{N}(0,1). \tag{5}$$

Следующая техническая лемма используется для поиска асимптотики условного математического ожидания и условной дисперсии различных случайных величин, возникающих в настоящей работе. Данный результат является развитием известного метода Лапласа оценки интегралов (см. [1]).

Лемма 3 [25]. *Обозначим $F(q) = \int\limits_q^{+\infty} \exp\bigl(-S(x)\bigr)\,dx$. Пусть функция $S(x)$ удовлетворяет следующим условиям регулярности.*

I1. $S(x)/\ln x \to +\infty$ *при* $x \to +\infty$.

I2. $S(x)$ *монотонна и* $S(\cdot) \in C^4$ *на бесконечности. Кроме того, существуют пределы*
$$a_k := \lim_{x\to\infty} S^{(k)}(x) \in [0,\infty], \quad k = 1,2,3,4.$$

I3. *Существуют пределы*
$$\lim_{x\to+\infty} \frac{\ln\bigl|S^{(k)}(x)\bigr|}{S(x)} = 0, \quad \lim_{x\to+\infty} \frac{S^{(k+1)}(x)}{\bigl(S^{(k)}(x)S'(x)\bigr)} \leqslant \infty, \quad k = 1,2,3.$$

Тогда
$$F(q) = \exp\bigl(-S(q)\bigr)\left(\sum_{k=0}^{2} c_k + o(c_2)\right) \quad \textit{при} \quad q \to +\infty,$$
где $c_k = M^k\bigl(\bigl(S'(x)\bigr)^{-1}\bigr)\big|_{x=q}$ *и* $M = \bigl(S'(x)\bigr)^{-1}\frac{d}{dx}$.

Замечание 1. Для получения соотношения
$$F(q) = \frac{\exp\bigl(-S(q)\bigr)}{S'(q)} \cdot \bigl(1 + o(1)\bigr), \quad q \to +\infty$$
можно не рассматривать условие наличия предела второй производной и условие I3 для $k = 2$.

Следующее важное утверждение позволяет перейти от условных распределений к безусловным.

Лемма 4. *Пусть для любого x из множества точек непрерывности некоторой функции распределения F ($C(F)$) и для любого q выполнено следующее соотношение*:
$$F_n(x,q) := \mathrm{P}(\eta_n \leqslant x \mid \xi_n = q) \to F(x), \quad n \to +\infty,$$
где $\{\eta_n\}_{n=1}^{+\infty}$, $\{\xi_n\}_{n=1}^{+\infty}$ — последовательности случайных величин, причём $\xi_n \uparrow x^*$ почти наверное при $n \to \infty$. Тогда для каждого $x \in C(F)$ $\mathrm{P}(\eta_n \leqslant x) \to F(x)$ при $n \to +\infty$.

Данный результат означает, что сходимость по распределению последовательности случайных величин $\{\eta_n\}$ будет сохраняться, если предельная функция распределения не будет зависеть от параметра q.

Доказательство. По формуле полной вероятности

$$\mathrm{P}(\eta_n \leqslant x) = \int\limits_{\mathbb{R}} \mathrm{P}(\eta_n \leqslant x \mid \xi_n = q) dF_{\xi_n}(q),$$

где $F_{\xi_n}(q)$ — функция распределения случайной величины ξ_n. По условию существует такой компакт $[x_0, x^*]$, что $P(\xi_n \in [x_0, x^*]) \to 1$ при $n \to +\infty$, а на компакте сходимость $\mathrm{P}(\eta_n \leqslant x \mid \xi_n = q)$ к $F(x)$ будет равномерной по q, что и приводит к требуемому утверждению. Заметим, что утверждение доказывается аналогично, если $\{\xi_n\}$ — это последовательность случайных векторов. □

5. Доказательства

5.1. Доказательство теоремы 1

Условное распределение статистики $\hat{a}$ при условии $X_{(n-k)} = q$, $Y_{(n-k)} = r$ будет иметь вид

$$\frac{X^*_{(n)} + \ldots + X^*_{(n-k+1)}}{k} - \frac{Y^*_{(n)} + \ldots + Y^*_{(n-k+1)}}{k} \overset{d}{=} \frac{X^*_1 + \ldots + X^*_k}{k} - \frac{Y^*_1 + \ldots + Y^*_k}{k},$$

где $\{X^*_i\}$ и $\{Y^*_j\}$ — случайные величины из леммы 1. По усиленному закону больших чисел выражение справа почти наверное стремится к $\delta := \mathrm{E}X^*_1 - \mathrm{E}Y^*_1$. Обозначим через f_0 плотность случайной величины Y_1, через f_a обозначим плотность случайной величины X_1. Имеем

$$\mathrm{E}X^*_1 - \mathrm{E}Y^*_1 = \frac{\int\limits_q^{+\infty} x f_0(x-a)\, dx}{1 - F_0(q-a)} - \frac{\int\limits_r^{+\infty} x f_0(x)\, dx}{1 - F_0(r)} =$$

$$= \frac{\int\limits_{q-a}^{+\infty} (x+a) f_0(x)\, dx}{1 - F_0(q-a)} - \frac{\int\limits_r^{+\infty} x f_0(x)\, dx}{1 - F_0(r)} = a + \frac{\int\limits_{q-a}^{+\infty} x f_0(x)\, dx}{1 - F_0(q-a)} - \frac{\int\limits_r^{+\infty} x f_0(x)\, dx}{1 - F_0(r)}.$$

Рассмотрим числитель каждой из дробей, сделаем преобразования при помощи формулы интегрирования по частям:

$$\int\limits_r^{+\infty} x f_0(x)\, dx = -\int\limits_r^{+\infty} x d\big(1 - F_0(x)\big) =$$

$$= x\big(1 - F_0(x)\big)\big|_r^{+\infty} + \int\limits_r^{+\infty} \big(1 - F_0(x)\big)\, dx = r\big(1 - F_0(r)\big) + \int\limits_r^{+\infty} \big(1 - F_0(x)\big)\, dx.$$

Тогда искомая разность математических ожиданий равняется

$$a + (q-a) - r + \left(\frac{\int\limits_{q-a}^{+\infty} \big(1 - F_0(x)\big)\,dx}{1 - F_0(q-a)} - \frac{\int\limits_{r}^{+\infty} \big(1 - F_0(x)\big)\,dx}{1 - F_0(r)} \right).$$

Применяя теорему 2 и соотношение $U_a(t) = U(t) + a$ к случайным величинам $X_{(n-k)}$ и $Y_{(n-k)}$, имеем

$$\sqrt{k} \cdot \frac{X_{(n-k)} - a - U(n/k)}{(n/k) \cdot U'(n/k)} \xrightarrow{d} \mathcal{N}(0,1), \quad \sqrt{k} \cdot \frac{Y_{(n-k)} - U(n/k)}{(n/k) \cdot U'(n/k)} \xrightarrow{d} \mathcal{N}(0,1)$$

при $k \to +\infty$, $k/n \to 0$, $n \to +\infty$. Применим дельта-метод для функции

$$h(r) = r + \frac{\int\limits_{r}^{+\infty} \big(1 - F(x)\big)\,dx}{1 - F(r)}$$

и выражений выше. Положим $\alpha = U(n/k)$, $b_n = U'(n/k)/\sqrt{k}$. Докажем, что в условиях теоремы 1 $h'(r) \to 1$ при $r \to +\infty$.

Проверим, что для функции $S(x) = -\ln\big(1 - F(x)\big) = T(x)$ выполнены предположения леммы 3.

Шаг 1. По критерию принадлежности области максимального притяжения Гумбеля [16] $T(x)/\ln x \to +\infty$ при $x \to +\infty$.

Шаг 2. Покажем, что

$$\lim_{x \to +\infty} \frac{\ln T'(x)}{T(x)} = 0.$$

Заметим, что $T(x) \to +\infty$ при $x \to +\infty$. Поэтому по правилу Лопиталя имеем

$$\lim_{x \to +\infty} \frac{\ln T'(x)}{T(x)} = \lim_{x \to +\infty} \frac{-T''(x)}{\big(T'(x)\big)^2} = 0$$

из условия фон Мизеса. Наконец, покажем, что $h'(r) \to 1$ при $r \to +\infty$. Действительно,

$$h'(r) = 1 - \frac{1 - F(r)}{1 - F(r)} + \frac{f(r) \int\limits_{r}^{+\infty} \big(1 - F(x)\big)\,dx}{\big(1 - F(r)\big)^2}. \tag{6}$$

В силу леммы 3 имеем

$$\int\limits_{r}^{+\infty} \big(1 - F(x)\big)\,dx = \frac{\big(1 - F(r)\big)^2}{f(r)} \cdot \big(1 + o(1)\big). \tag{7}$$

Подставляя выражение (7) в выражение (6), имеем $h'(r) = 1 + o(1)$ при $r \to +\infty$, что и завершает доказательство утверждения.

Применяя дельта-метод и лемму 2, имеем

$$\sqrt{k}\frac{h(X_{(n-k)}-a)+h\big(U(n/k)\big)}{n/k\cdot U'(n/k)} \xrightarrow{d} \mathcal{N}(0,1), \tag{8}$$

$$\sqrt{k}\frac{h(Y_{(n-k)})-h\big(U(n/k)\big)}{n/k\cdot U'(n/k)} \xrightarrow{d} \mathcal{N}(0,1) \tag{9}$$

при $k\to+\infty$, $k/n\to 0$, $n\to+\infty$. Вычитая из выражения (8) выражение (9), имеем в силу независимости выборок X и Y

$$\sqrt{k}\frac{h(X_{(n-k)}-a)-h(Y_{(n-k)})}{n/k\cdot U'(n/k)} \xrightarrow{d} \mathcal{N}(0,2) \tag{10}$$

при $k\to+\infty$, $k/n\to 0$, $n\to+\infty$. С учётом того что разность математических ожиданий равняется

$$\mathrm{E}(X_1^*-Y_1^*)=a+h(q-a)-h(r), \tag{11}$$

получаем предельное условное распределение (т. е. при условии $X_{(n-k)}$)

$$\sqrt{k}\frac{\mathrm{E}X_1^*-\mathrm{E}Y_1^*-a}{n/k\cdot U'(n/k)} \xrightarrow{d} \mathcal{N}(0,2) \tag{12}$$

при $k\to+\infty$, $k/n\to 0$, $n\to+\infty$. В итоге, применяя лемму 4 к последовательностям пар случайных величин $X_{(n-k)}$, $Y_{(n-k)}$ (лемму можно применять, так как k является функцией от n) и случайным величинам

$$\eta_{n,k}=\sqrt{k}\frac{\mathrm{E}X_1^*-\mathrm{E}Y_1^*-a}{n/k\cdot U'(n/k)},$$

где разность математических ожиданий является пределом последовательности условных математических ожиданий $\mathrm{E}(\hat{a}\mid X_{(n-k)},\ Y_{(n-k)})$, получаем аналогичное утверждение для $\hat{a}$:

$$\sqrt{k}\frac{\hat{a}-a}{n/k\cdot U'(n/k)} \xrightarrow{d} \mathcal{N}(0,2), \quad n\to+\infty, \quad k\to+\infty, \quad \frac{k}{n}\to 0. \tag{13}$$

Для завершения доказательства достаточно проверить, что в условиях теоремы

$$\frac{\sqrt{k}}{n/k\cdot U'(n/k)}\to+\infty$$

при $n\to+\infty$, $k\to+\infty$, $k/n\to 0$. Заметим, что по определению функции U выполнено равенство $1/\Big(1-F\big(U(t)\big)\Big)=t$ для всех t, лежащих в некоторой окрестности x_F^*. Преобразуем выражение $n/k\cdot U'(n/k)$,

$$F\big(U(t)\big)=1-\frac{1}{t}, \quad f\big(U(t)\big)U'(t)=\frac{1}{t^2}, \quad tU'(t)=\frac{1}{t\cdot f\big(U(t)\big)}.$$

Выясним, при каких условиях на последовательность $k=k(n)$ выполнено

$$\frac{1}{\sqrt{k}\cdot\frac{n}{k}\cdot f\left(U\left(\frac{n}{k}\right)\right)}\to 0.$$

Для этого докажем следующее утверждение.

Утверждение 1. *Пусть $F \in D(G_0)$ и выполнено условие фон Мизеса* (2). *Тогда*

$$\left(\frac{1-F(t)}{f(t)}\right)' \to 0, \quad t \to +\infty.$$

Доказательство. Из условия фон Мизеса вытекает, что

$$\left(\frac{1-F(t)}{f(t)}\right)' = \\ = \frac{-f(t)f(t) - f'(t)\big(1-F(t)\big)}{f(t)^2} = -1 - \frac{f'(t)\big(1-F(t)\big)}{f(t)^2} \to 0, \quad t \to +\infty. \quad \square$$

Лемма 5. *Пусть $F \in D(G_0)$. Тогда*

$$\frac{1-F(t)}{tf(t)} \to 0, \quad t \to +\infty,$$

где f — плотность случайной величины с функцией распределения F.

Доказательство. Заметим, что из утверждения 1 следует, что

$$\left(\frac{1-F(t)}{f(t)}\right)' \to 0, \quad t \to +\infty.$$

Функция

$$\frac{f(t)}{1-F(t)} = S'(t)$$

имеет конечный или бесконечный предел в силу леммы 3 при $t \to +\infty$. Рассмотрим два случая:

1) $\big(1-F(t)\big)/f(t) \to +\infty$ при $t \to +\infty$: можно применить правило Лопиталя и получить требуемое;
2) $\big(1-F(t)\big)/f(t)$ имеет конечный предел — очевидно.

Тогда имеем

$$\frac{n}{k} \cdot f\left(U\left(\frac{n}{k}\right)\right) \cdot U\left(\frac{n}{k}\right) \to +\infty, \quad n \to +\infty.$$

Тем самым достаточным условием сходимости является ограниченность выражения $U(n/k)/\sqrt{k}$. Пусть существует такое $M > 0$, что при достаточно больших n, k выполнено $U(n/k)/\sqrt{k} \leqslant M$. Тогда $1 - F_0(M\sqrt{k}) \leqslant k/n$, что и завершает доказательство. $\square$

5.2. Доказательство теоремы 2

Теорема 2 представляет собой несложное следствие теоремы 1. Сначала докажем следующую лемму.

Лемма 6. *Пусть случайная величина X имеет функцию распределения, удовлетворяющую условию фон Мизеса для $\gamma \geqslant 0$. Тогда случайная величина $\ln X$ имеет функцию распределения, удовлетворяющую условию фон Мизеса для области притяжения Гумбеля.*

Доказательство. Из условий леммы следует, что

$$\lim_{t\to+\infty}\frac{\big(1-F(t)\big)F''(t)}{\big(F'(t)\big)^2}=-\gamma-1.$$

Обозначим через $G(t)$ функцию распределения случайной величины $\ln X$. Тогда $G(t)=F(e^t)$ для любого $t\in\mathbb{R}$. Заметим, что $G'(t)=e^tF'(e^t)$, $G''(t)=$ $=e^{2t}F''(e^t)+e^tF'(e^t)$. Тогда

$$\frac{\big(1-G(t)\big)G''(t)}{\big(G'(t)\big)^2}=\frac{\big(1-F(s)\big)\big(s^2F''(s)+sF'(s)\big)}{s^2\big(F'(s)\big)^2}\bigg|_{s=e^t}=$$
$$=\frac{\big(1-F(t)\big)F''(t)}{\big(F'(t)\big)^2}+\frac{\big(1-F(s)\big)}{sF'(s)}.$$

Первое слагаемое выражения стремится к $-\gamma-1$ по условию леммы, второе слагаемое стремится к нулю по лемме 5, что и требовалось. □

Таким образом, переписывая условия теоремы 1 в терминах функции распределения для логарифмов порядковых статистик, получаем утверждение теоремы 2, поскольку $X_{(n-k)}\to+\infty$ п. н. при $n\to+\infty$, $k\to+\infty$, $k/n\to0$.

5.3. Доказательство теоремы 3

Для условных распределений выполнена центральная предельная теорема:

$$\sqrt{k}\frac{\frac{1}{k}\sum\limits_{i=1}^{k}X_i^*-\mathrm{E}X_1^*}{\sqrt{\mathrm{D}X_1^*}}\xrightarrow{d}\mathcal{N}(0,1),\quad k\to+\infty. \tag{14}$$

Оценим асимптотику математического ожидания и дисперсии X_1^* при условии $X_{(n-k)}=q$. Из доказательства теоремы 1 имеем

$$\mathrm{E}X_1^*=a+(q-a)+\frac{\int\limits_{q-a}^{+\infty}\big(1-F_0(x)\big)\,dx}{1-F_0(q-a)},\quad \mathrm{E}Y_1^*=r+\frac{\int\limits_{r}^{+\infty}\big(1-F_0(x)\big)\,dx}{1-F_0(r)}.$$

Функция $T(x)$ удовлетворяет условиям леммы 3. Действительно, выполнение условий I1 и I2 установлено при доказательстве теоремы 1. Проверим второе условие I3: $\ln|T''(x)|/T(x)\to0$ при $x\to+\infty$. Из условия (R1) имеем, что $T''(x)$ имеет конечный или бесконечный предел при $x\to+\infty$, поэтому возможны два случая:

1) $T''(x)\to0$ или $T''(x)\to\infty$: применяя правило Лопиталя, достаточно показать, что $T'''(x)/\big(T''(x)T'(x)\big)\to0$ при $x\to+\infty$, поскольку $\ln|T''(x)|\to+\infty$;
2) $T''(x)$ имеет конечный и ненулевой предел. Тогда числитель дроби стремится к конечному значению, а знаменатель стремится к 0, что и даёт выполнение условий.

Докажем, что
$$\frac{T'''(x)}{T''(x)T'(x)} \to 0 \quad \text{при} \quad x \to +\infty.$$
Из условия фон Мизеса следует, что $T''(x)/\big(T'(x)\big)^2 \to 0$ при $x \to +\infty$. Рассмотрим три случая:

1) $T'(x) \to 0$, $T''(x) \to 0$. По правилу Лопиталя имеем
$$\frac{T'''(x)}{T''(x)T'(x)} \to 0;$$
2) $T'(x) \to +\infty$, $T''(x) \to +\infty$ (аналогичен первому случаю);
3) $T''(x) \to c$, $T'(x) \to \infty$ для некоторого числа $c \in \mathbb{R}$.

Рассмотрим третий случай. Покажем, что в этом случае верно соотношение
$$\frac{T'''(x)}{T''(x)} \to 0 \quad \text{при} \quad x \to +\infty.$$
Действительно,
$$\frac{T'''(x)}{T''(x)} = \frac{(\ln|T''(x)|)'}{x'}.$$
Если $T''(x)$ имеет ненулевой предел, то $T'''(x)/T''(x) \to 0$, поскольку в этом случае $(\ln|T''(x)|)' \to 0$. Если $T''(x) \to 0$ при $x \to +\infty$, то тогда покажем, что для каждого $\varepsilon > 0$ найдётся x_ε, такой что для каждого $x > x_\varepsilon$ справедливо $\ln|T''(x)|/x < \varepsilon$. Предположим, что существует $\varepsilon > 0$, такой что для каждого x найдётся $x^* > x$, для которого $\ln|T''(x^*)|/x^* \leqslant -\varepsilon$; противоположный случай невозможен из-за наличия нулевого предела у функции $T''(x)$. Тогда $|T''(x)| \leqslant \exp\{-\varepsilon x/2\}$ в окрестности точки x^*, что говорит о конечном пределе функции $T'(x)$.

Применяя лемму 3, имеем
$$\frac{\int\limits_r^{+\infty} e^{-T_0(x)}\,dx}{e^{-T_0(r)}} = \frac{1}{T_0'(r)} \cdot \left(1 - \frac{T_0''(r)}{\big(T_0'(r)\big)^2} + o\left(\frac{T_0''(r)}{\big(T_0'(r)\big)^2}\right)\right), \quad r \to +\infty.$$
В итоге получаем
$$\mathrm{E}X_1^* = a + (q-a) + \frac{1}{T_0'(q-a)} - \frac{T_0''(q-a)}{T_0'(q-a)^3} + o\left(\frac{T_0''(q-a)}{T_0'(q-a)^3}\right), \tag{15}$$
$$\mathrm{E}Y_1^* = r + \frac{1}{T_0'(r)} - \frac{T_0''(r)}{T_0'(r)^3} + o\left(\frac{T_0''(r)}{T_0'(r)^3}\right), \quad r \to +\infty, \quad q \to +\infty. \tag{16}$$
Для того чтобы оценить DX_1^*, найдём асимптотику $\mathrm{E}X_1^{*2}$:
$$\mathrm{E}X_1^{*2} = \frac{\int\limits_q^{+\infty} x^2 f_a(x)\,dx}{1 - F_a(q)} = a^2 + 2a \cdot \frac{\int\limits_{q-a}^{+\infty} x f_0(x)\,dx}{1 - F_0(q-a)} + \frac{\int\limits_{q-a}^{+\infty} x^2 f_0(x)\,dx}{1 - F_0(q-a)}.$$

Тогда искомая дисперсия будет равняться

$$\mathrm{D}X_1^* = a^2 + 2a \cdot \frac{\int\limits_{q-a}^{+\infty} x f_0(x)\,dx}{1 - F_0(q-a)} + \frac{\int\limits_{q-a}^{+\infty} x^2 f_0(x)\,dx}{1 - F_0(q-a)} - \left(a + \frac{\int\limits_{q-a}^{+\infty} x f_0(x)\,dx}{1 - F_0(q-a)} \right)^2$$

$$= \frac{\int\limits_{q-a}^{+\infty} x^2 f_0(x)\,dx}{1 - F_0(q-a)} - \left(\frac{\int\limits_{q-a}^{+\infty} x f_0(x)\,dx}{1 - F_0(q-a)} \right)^2,$$

т. е. не зависит от параметра a. Используя формулу интегрирования по частям, получаем

$$\frac{\int\limits_{q-a}^{+\infty} x^2 f_0(x)\,dx}{1 - F_0(q-a)} = \frac{\int\limits_{q-a}^{+\infty} x^2 d\big(1 - F_0(x)\big)}{1 - F_0(q-a)} = (q-a)^2 + \frac{\int\limits_{q-a}^{+\infty} 2x\big(1 - F_0(x)\big)\,dx}{1 - F_0(q-a)}. \tag{17}$$

Обозначим $W(x) = T(\sqrt{x})$. Тогда выражение (17) будет иметь следующий вид:

$$\frac{\int\limits_{q-a}^{+\infty} x^2 f_0(x)\,dx}{1 - F_0(q-a)} = (q-a)^2 + \frac{\int\limits_{(q-a)^2}^{+\infty} \exp\{-W(x)\}\,dx}{\exp\{-W\big((q-a)^2\big)\}}.$$

Покажем, что для функции W выполнены условия леммы 3. Заметим, что производные функции W равны

$$W'(x) = \frac{T'(\sqrt{x})}{2\sqrt{x}}, \quad W''(x) = \frac{T''(\sqrt{x})}{4x} - \frac{T'(\sqrt{x})}{4x^{3/2}}.$$

Проверим условие I1: $W(x)/\ln x \to +\infty$ при $x \to +\infty$. Действительно,

$$\lim_{x\to+\infty} \frac{W(x)}{\ln x} = \lim_{x\to+\infty} \frac{T(x)}{2\ln x} = 0.$$

Проверим условие I2: функции $W'(x)$ и $W''(x)$ имеют конечный или бесконечный предел при $x \to +\infty$. Для функции $W'(x^2) = T'(x)/(2x)$ это показывается применением правила Лопиталя при условии $\lim\limits_{x\to+\infty} T'(x) = \infty$. Покажем, что $W''(x)$ имеет конечный или бесконечный предел при $x \to +\infty$. Действительно, заметим, что поскольку $T'(x) \to +\infty$, то

$$\lim_{x\to+\infty} \frac{T'(x)}{x^3} = \lim_{x\to+\infty} \frac{1}{3} \cdot \frac{T''(x)}{x^2}.$$

Тем самым если $W'(x)$ имеет конечный предел, то $W''(x) \to 0$ при $x \to +\infty$. Иначе заметим, что

$$W''(x^2) = \frac{\big(T'(x)/(2x)\big)'}{(x^2)'} = \frac{\big(W'(x^2)\big)'}{(x^2)'}$$

и по правилу Лопиталя существование предела для $W''(x)$ при $x \to +\infty$ эквивалентно наличию предела $T'(x)/x^3$ при $x \to +\infty$, поскольку $W'(x^2) \to +\infty$ в рассматриваемом случае, а наличие предела функции $T'(x)/x^3$ при $x \to +\infty$ следует из условия регулярности.

Проверим первое условие I3. Заметим, что

$$\lim_{x\to+\infty} \frac{\ln|W'(x)|}{W(x)} = \lim_{x\to+\infty} \frac{\ln|T'(x)| - \ln 2 - \ln x}{T(x)} = 0$$

(это следует из условий I1 и I3 леммы 3 для функции $T'(x)$).

Второе условие I3 следует из условий регулярности (R1), (R2).

Таким образом, условия утверждения 3 выполнены, поэтому имеем

$$I(s) := \frac{\int\limits_s^{+\infty} \exp\{-W(x)\}\,dx}{\exp\{-W(s)\}} = \frac{1}{W'(s)} \cdot \left(1 - \frac{W''(s)}{\big(W'(s)\big)^2} + o\left(\frac{W''(s)}{\big(W'(s)\big)^2}\right)\right) \tag{18}$$

при $s \to +\infty$, где $s = (q-a)^2$. Преобразуем выражение $W''(s)/W'(s)^2$:

$$H(s) := \frac{W''(s)}{\big(W'(s)\big)^2} = \frac{T''(\sqrt{s}) - T'(\sqrt{s})/\sqrt{s}}{\big(T'(\sqrt{s})\big)^2} = \frac{T''(\sqrt{s})}{\big(T'(\sqrt{s})\big)^2} - \frac{1}{\sqrt{s}T'(\sqrt{s})}. \tag{19}$$

В итоге из (18) имеем

$$I(s) = \frac{2\sqrt{s}}{T'(\sqrt{s})} \cdot \left(1 - \frac{T''(\sqrt{s})}{\big(T'(\sqrt{s})\big)^2} + \frac{1}{\sqrt{s}T'(\sqrt{s})} + o\big(H(s)\big)\right) =$$

$$= \frac{2(q-a)}{T'(q-a)} - \frac{2(q-a)T''(q-a)}{\big(T'(q-a)\big)^3} + \frac{2}{\big(T'(q-a)\big)^2} + o\left(H\big((q-a)^2\big) \cdot \frac{2(q-a)}{T'(q-a)}\right).$$

Применяя соотношение для оценивания асимптотики интегралов, получаем (оценка о-малых опускается для простоты записи выражения и будет произведена далее)

$$\mathrm{D}X_1^* \sim (q-a)^2 + \frac{2(q-a)}{T'(q-a)} - \frac{2(q-a)T''(q-a)}{\big(T'(q-a)\big)^3} + \frac{2}{\big(T'(q-a)\big)^2} -$$

$$- \left(q - a + \frac{1}{T'(q-a)} - \frac{T''(q-a)}{\big(T'(q-a)\big)^3}\right)^2 =$$

$$= \frac{1}{\big(T'(q-a)\big)^2} - 2\frac{T''(q-a)}{\big(T'(q-a)\big)^4} - \left(\frac{T''(q-a)}{\big(T'(q-a)\big)^2}\right)^2 =$$

$$= \left(\frac{1}{T'(q-a)}\right)^2 \cdot \left[\left(1 - \frac{T''(q-a)}{\big(T'(q-a)\big)^2}\right)^2 - \left(\frac{T''(q-a)}{\big(T'(q-a)\big)^2}\right)^2\right].$$

Произведём анализ о-малых. Из условия фон Мизеса для области максимального притяжения Гумбеля следует, что значимыми слагаемыми, стоящими под знаком о-малого в математическом ожидании (15), являются

$$\frac{(q-a)T''(q-a)}{\big(T'(q-a)\big)^3}, \quad \frac{T''(q-a)}{\big(T_0'(q-a)\big)^4}.$$

Под знаком о-малого для I таковыми слагаемыми являются

$$\frac{(q-a)T''(q-a)}{\big(T'(q-a)\big)^3}, \quad \frac{1}{\big(T'(q-a)\big)^2}.$$

Однако по правилу Лопиталя

$$\lim_{x\to+\infty}\frac{1/\,T'(x-a)}{(x-a)'} = \lim_{x\to+\infty}\frac{T''(x-a)}{\big(T'(x-a)\big)^2}.$$

В итоге значения функций под знаком о-малого в выражении для I можно оценить как

$$\frac{1}{\big(T'(q-a)\big)^2}\big(1+o(1)\big), \quad x\to+\infty.$$

Поскольку в силу условия фон Мизеса

$$\frac{T''(x-a)}{\big(T'(x-a)\big)^2}\to 0, \quad x\to+\infty,$$

асимптотика дисперсии X_1^* имеет вид

$$\mathrm{D}X_1^* = \left(\frac{1}{T'(q-a)}\right)^2\cdot\big(1+o(1)\big), \quad q\to+\infty. \tag{20}$$

Проводя аналогичные рассуждения для выборки $\{Y_i^*\}$, получаем

$$\mathrm{D}Y_1^* = \left(\frac{1}{T'(r)}\right)^2\cdot\big(1+o(1)\big), \quad r\to+\infty. \tag{21}$$

Найдём асимптотическую дисперсию оценки. Применим для этого центральную предельную теорему для случайных величин $\{X_i^*\}_{i=1}^k$ и $\{Y_i^*\}_{i=1}^k$. Обозначим

$$M_{X,k} = \frac{1}{k}\sum_{i=1}^k X_i^*, \quad M_{Y,k} = \frac{1}{k}\sum_{i=1}^k Y_i^*.$$

Получаем

$$\sqrt{k}(M_{X,k}-\mathrm{E}X_1^*) \xrightarrow{d} \mathcal{N}(0,\mathrm{D}X_1^*), \quad k\to+\infty, \tag{22}$$

$$\sqrt{k}(M_{Y,k}-\mathrm{E}Y_1^*) \xrightarrow{d} \mathcal{N}(0,\mathrm{D}Y_1^*), \quad k\to+\infty. \tag{23}$$

Вычитая из соотношения (22) соотношение (23), имеем

$$\sqrt{k}(M_{X,k}-M_{Y,k}-a)+\sqrt{k}(a-\mathrm{E}X_1^*+\mathrm{E}Y_1^*) \xrightarrow{d} \mathcal{N}(0,\mathrm{D}X_1^*+\mathrm{D}Y_1^*), \quad k\to+\infty. \tag{24}$$

Обозначим $V(X,Y) = \mathrm{E}X_1^* - \mathrm{E}Y_1^* - a$. Тогда предельный закон (12) представляется в виде

$$D(X,Y,n,k) := \sqrt{k}\frac{V(X,Y)}{n/k \cdot U_0'(n/k)} \xrightarrow{d} \mathcal{N}(0,2), \quad n \to +\infty.$$

Докажем лемму, которая позволит оценить поведение асимптотической дисперсии.

Лемма 7. *Для любых $n > k$ выполнено*

$$\frac{1}{T'\big(U(n/k)\big)} = \frac{n}{k}U'\left(\frac{n}{k}\right).$$

Доказательство. Ранее было показано, что

$$\frac{n}{k}U'\left(\frac{n}{k}\right) = \frac{1}{f\big(U(n/k)\big)\cdot(n/k)}.$$

Заметим, что

$$f\left(U\left(\frac{n}{k}\right)\right) = T'\left(U\left(\frac{n}{k}\right)\right)e^{-T(U(n/k))}.$$

С другой стороны,

$$F\left(U\left(\frac{n}{k}\right)\right) = 1 - \frac{k}{n} = 1 - e^{-T(U(n/k))},$$

поэтому $e^{-T(U(n/k))} = k/n$. Следовательно,

$$f\left(U\left(\frac{n}{k}\right)\right)\cdot\frac{n}{k} = T'\left(U\left(\frac{n}{k}\right)\right),$$

что и приводит к доказываемому утверждению. □

Таким образом, имеем при $n \to +\infty$

$$\sqrt{k}(M_{X,k} - M_{Y,k} - a) - D(X,Y,n,k)\frac{1}{T'\big(U(n/k)\big)} \xrightarrow{d} \mathcal{N}(0, \mathrm{D}X_1^* + \mathrm{D}Y_1^*). \quad (25)$$

Покажем, что

$$\frac{T'(X(n-k)-a)}{T'\big(U(n/k)\big)} \xrightarrow{P} 1, \quad \frac{T'(Y_{(n-k)})}{T'\big(U(n/k)\big)} \xrightarrow{P} 1, \quad n \to +\infty.$$

Применим дельта-метод для функции $g(x) = 1/T'(x)$ и соотношения (5) к выборкам X и Y (метод применим, поскольку производная имеет нулевой предел в силу условия фон Мизеса, $g'(x) = -T''(x)/\big(T'(x)\big)^2$). В итоге получаем при $n \to +\infty$

$$\sqrt{k}T'\left(U\left(\frac{n}{k}\right)\right)\cdot\frac{\big(T'\big(U(n/k)\big)\big)^2}{T''\big(U(n/k)\big)}\left(\frac{1}{T'(X_{(n-k)}-a)} - \frac{1}{T'\big(U(n/k)\big)}\right) \xrightarrow{d} \mathcal{N}(0,1),$$

что может быть преобразовано к следующему виду:

$$\sqrt{k}\frac{\big(T'\big(U(n/k)\big)\big)^2}{T''\big(U(n/k)\big)}\left(\frac{T'\big(U(n/k)\big)}{T'(X_{(n-k)}-a)} - 1\right) \xrightarrow{d} \mathcal{N}(0,1), \quad n \to +\infty. \quad (26)$$

Поскольку верно
$$\sqrt{k}\frac{\big(T'\big(U(n/k)\big)\big)^2}{T''\big(U(n/k)\big)} \to \infty, \quad n \to +\infty,$$
то получаем требуемое. Для дроби $T'\big(Y(n-k)\big)/T'\big(U(n/k)\big)$ рассуждения аналогичны. Таким образом, выполнены предельные соотношения
$$\frac{1/T'\big(U(n/k)\big)}{\sqrt{\mathrm{D}X_1^*}} \to 1, \quad \frac{1/T'\big(U(n/k)\big)}{\sqrt{\mathrm{D}Y_1^*}} \to 1,$$
где n, k удовлетворяют условию теоремы, откуда следует, что
$$\frac{1/T'\big(U(n/k)\big)}{\sqrt{\mathrm{D}X_1^* + \mathrm{D}Y_1^*}} \to \frac{1}{\sqrt{2}}, \quad n \to +\infty.$$
Из леммы Слуцкого и формулы (25) имеем
$$\sqrt{k}\frac{M_{X,k} - M_{Y,k} - a}{\sqrt{\mathrm{D}X_1^* + \mathrm{D}Y_1^*}} - \frac{1}{\sqrt{2}} \cdot D(X, Y, n, k) \xrightarrow{d} \mathcal{N}(0,1), \quad n \to +\infty,$$
откуда следует, что
$$\sqrt{k}T'\big(U(n/k)\big)(M_{X,k} - M_{Y,k} - a) \xrightarrow{d} \mathcal{N}(0,4), \quad n \to +\infty. \tag{27}$$
Применяя лемму 4 аналогично доказательству теоремы 1, получаем
$$\sqrt{k}T'\big(U(n/k)\big)(\hat{a} - a) \xrightarrow{d} \mathcal{N}(0,4), \quad n \to +\infty, \tag{28}$$
что и доказывает теорему 3.

5.4. Доказательство теоремы 4

В данном разделе будем считать, что функция T задана для случайной величины $\ln Y_i$. Тогда с учётом лемм выполнен предельный закон
$$\sqrt{k}T'\left(U\left(\frac{n}{k}\right)\right)(\ln\hat{b} - \ln b) \xrightarrow{d} \mathcal{N}(0,4), \quad n \to +\infty. \tag{29}$$
Переход к параметру сдвига осуществляется посредством применения теоремы о наследовании асимптотической нормальности к формуле (29) для функции $m(x) = e^x$. Тогда асимптотическая дисперсия станет равной $4\big(m'(\ln b)\big)^2 = 4b^2$, откуда следует, что
$$\sqrt{k}T'\left(U\left(\frac{n}{k}\right)\right)(\hat{b} - b) \xrightarrow{d} \mathcal{N}(0, 4b^2), \quad n \to +\infty.$$

5.5. Доказательство примера 2

Проверим условия теоремы 3. Во-первых, наличие конечных или бесконечных пределов функции $T'(x) = \alpha x^{\alpha-1}$, $T''(x) = \alpha \cdot (\alpha - 1)x^{\alpha-2}$, $T'''(x)$ и монотонность этих функций в окрестности точки $+\infty$ очевидны. Дополнительно

$T'(x)/x^m = \alpha \cdot x^{\alpha-1-m}$ имеет конечный предел при $x \to +\infty$ для $\alpha \leqslant m+1$ и бесконечный предел при $x \to +\infty$ иначе. Далее, проверим условие фон Мизеса:

$$\frac{T''(x)}{\big(T'(x)\big)^2} = \frac{\alpha \cdot (\alpha-1)x^{\alpha-2}}{\alpha^2 \cdot x^{2\alpha-2}} = \frac{\alpha-1}{\alpha}x^{-\alpha} \to 0, \quad x \to +\infty.$$

Условие регулярности (R1) будет выполнено при $\alpha \neq 2$, поскольку

$$\frac{\ln|T'(x)|}{\ln x} = \frac{\ln\alpha + (\alpha-1)\ln x}{\ln x} \to \alpha - 1, \quad x \to +\infty.$$

Проверим выполнение ограничений, которые наложены на параметры n и k. Для этого вычислим $U(n/k)$. Поскольку $1 - F\big(U(n/k)\big) = k/n$, то $\exp\{U(n/k)^\alpha\} = n/k$, поэтому $U(n/k) = \big(\ln(n/k)\big)^{1/\alpha}$. Проверим условия теоремы 1.

1. $\sqrt{k}T'\big(U(n/k)\big) \to \infty$ при $n \to +\infty$. Тогда должен выполняется следующий предельный переход:

$$\sqrt{k}\alpha \cdot \left(\ln\left(\frac{n}{k}\right)\right)^{(\alpha-1)/\alpha} \to +\infty, \quad n \to +\infty, \quad k \to +\infty, \quad \frac{k}{n} \to 0.$$

Подставляя $k = n^\alpha$, получаем

$$\sqrt{k}\alpha \cdot \left(\ln\left(\frac{n}{k}\right)\right)^{(\alpha-1)/\alpha} = n^{\alpha/2} \cdot \big((1-\beta)\cdot \ln n\big)^{(\alpha-1)/\alpha} \to +\infty, \quad n \to +\infty;$$

что и требовалось.

2. Для некоторого $M > 0$

$$\frac{1 - F(M\sqrt{k})}{k/n} \leqslant 1.$$

Заметим, что $1 - F(M\sqrt{k}) = e^{-(M^\alpha n^{\beta/2})}$, $k/n = n^{\beta-1}$. Поскольку для любых многочленов $r > 0$, $s \in \mathbb{R}$ справедливо $e^{-n^r}n^s \to 0$ при $n \to +\infty$, то условие будет выполнено при любом $M > 0$.

Таким образом, условия теоремы 1 выполнены.

Литература

[1] Федорюк М. В. Метод перевала. — М.: Наука, 1977.

[2] Beirlant J., Bouquiaux C., Werker B. Semiparametric lower bounds for tail-index estimation // J. Statist. Plann. Inference. — 2006. — Vol. 136, No. 3. — P. 705—729.

[3] Beirlant T., Goegebeur Y., Teugels J. Statistics of Extremes. Theory and Applications. — London: Wiley, 2004. — (Wiley Ser. Probab. Stat.).

[4] Beirlant T., Teugels J. L. Asymptotics of Hill's estimator // Theory Probab. Appl. — 1986. — Vol. 31. — P. 463—469.

[5] Berred M. Record values and the estimation of the Weibull tail-coefficient // Compt. Rend. Acad. Sci. — 1991. — Vol. 312, No. 1. — P. 943—946.

[6] Bingham N. H., Goldie C. M., Teugels J. L. Regular Variation. — (Encyclopedia of Mathematics and Its Application; Vol. 27). — Cambridge: Cambridge Univ. Press, 1987.

[7] Diebolt J., Gardes L., Girard S., Guillou A. Bias-reduced estimators of the Weibull tail-coefficient // Test. — 2008. — Vol. 17. — P. 311—331.

[8] Drees H., Ferreira A., de Haan L. On maximum likelihood estimation of the extreme value index // Ann. Appl. Probab. — 2003. — Vol. 14. — P. 1179—1201.

[9] Drees H., de Haan L., Li D. Approximations to the tail empirical distribution function with application to testing extreme value conditions // J. Statist. Plann. Inference. — 2006. — Vol. 136, No. 10. — P. 3498—3538.

[10] Elandt-Johnson R., Johnson N. Survival Models and Data Analysis. — New York: Wiley, 1999.

[11] Fraga Alves M. I., Gomes M. I., de Haan L. A new class of semi-parametric estimators of the second order parameter // Portugal. Math. — 2003. — Vol. 60. — P. 193—213.

[12] Fraga Alves M. I., de Haan L., Lin T. Estimation of the parameter controlling the speed of convergence in extreme value theory // Math. Methods Statist. — 2003. — Vol. 12. — P. 155—176.

[13] Fraga Alves M. I., de Haan L., Neves C. A test procedure for detecting super-heavy tails // J. Statist. Plann. Inference. — 2009. — Vol. 138, No. 2. — P. 213—227.

[14] Gardes L., Girard S., Guillou A. Weibull tail-distributions revisited: a new look at some tail estimators // J. Statist. Plann. Inference. — 2009. — Vol. 141, No. 4. — P. 429—444.

[15] Gnedenko B. V. Sur la distribution limite du terme maximum d'une serie aleatoire // Ann. Math. — 1943. — Vol. 44. — P. 423—453.

[16] De Haan L., Ferreira A. Extreme Value Theory: An Introduction. — Springer, 2006.

[17] De Haan L., Resnick S. Second-order regular variation and rates of convergence in extreme value theory // Ann. Probab. — 1996. — Vol. 24, No. 1. — P. 97—124.

[18] De Haan L., Sinha A. K. Estimating the probability of a rare event // Ann. Statist. — 1999. — Vol. 27. — P. 732—759.

[19] Hall P. On some simple estimates of an exponent of regular variation // J. Roy. Statist. Soc., Ser. B. — 1982. — Vol. 44. — P. 37—42.

[20] Hill B. M. A simple general approach to inference about the tail of a distribution // Ann. Statist. — 1975. — Vol. 3. — P. 1163—1174.

[21] Kaplan E. L., Meier P. Nonparametric estimation from incomplete observations // J. Amer. Statist. Assn. — 1958. — Vol. 53, No. 282. — P. 457—481.

[22] Martins M. J. Heavy tails estimation — variants to the Hill estimator: PhD thesis. — Univ. of Lisbon, Portugal, 2000.

[23] Pickands III J. Statistical inference using extreme order statistics // Ann. Statist. — 1975. — Vol. 3. — P. 119—131.

[24] Rausand M., Hoyland A. System Reliability Theory: Models, Statistical Methods, and Applications. — Hoboken: Wiley, 2004.

[25] Rodionov I. V. Discrimination of close hypotheses about the distribution tails using higher order statistic // Theory Probab. Its Appl. — 2019. — Vol. 63, No. 3. — P. 364—380.

[26] Rodionov I. V. Inferences on parametric estimation of distribution tails // Dokl. Math. — 2019. — Vol. 100, No. 2. — P. 456—458.

[27] Rodionov I. V. On estimation of Weibull-tail and log-Weibull-tail distributions for modeling end-to-end delay // Distributed Computer and Communication Networks, 22nd

Int. Conf., DCCN 2019, Moscow, Russia, September 23—27, 2019, Revised Selected Papers. — Berlin: Springer, 2019. — (Commun. Comput. Inform. Sci.; Vol. 1141). — P. 302—314.

[28] Smith R. L. Estimating tails of probability distributions // Ann. Statist. — 1987. — Vol. 15. — P. 1174—1207.