МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М. В. ЛОМОНОСОВА

МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ

На правах рукописи
УДК 519.21

Родионов Игорь Владимирович

СТАТИСТИЧЕСКИЙ АНАЛИЗ И ПРОВЕРКА ГИПОТЕЗ О РАСПРЕДЕЛЕНИИ ЭКСТРЕМУМОВ ВРЕМЕННОГО РЯДА

01.01.05 — теория вероятностей и математическая статистика

Автореферат

диссертации на соискание учёной степени
кандидата физико-математических наук

Москва 2014

Работа выполнена на кафедре теории вероятностей механико-математического факультета Московского государственного университета имени М. В. Ломоносова.

Научный руководитель:

доктор физико-математических наук,
главный научный сотрудник Питербарг Владимир Ильич.

Официальные оппоненты:

доктор физико-математических наук
Маркович Наталья Михайловна,
Институт проблем управления РАН,
главный научный сотрудник лаборатории №38.

кандидат физико-математических наук,
Кудров Александр Владимирович,
Центральный экономико-математический институт РАН,
научный сотрудник лаборатории
вероятностно-статистических методов и моделей в экономике.

Ведущая организация:

Санкт-Петербургский государственный университет.

Защита диссертации состоится 14 марта 2014 года в 16 часов 45 минут на заседании диссертационного совета Д 501.001.85 при Московском государственном университете имени М. В. Ломоносова по адресу: Российская Федерация, 119991, Москва, ГСП-1, Ленинские горы, д. 1, МГУ, механико-математический факультет, аудитория 16-24.
С диссертацией можно ознакомиться в Фундаментальной библиотеке МГУ имени М. В. Ломоносова.

Автореферат разослан

Учёный секретарь
диссертационного совета
Д 501.001.85 при МГУ
доктор физико-математических наук,
профессор В. Н. Сорокин.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Классическая теория экстремумов — асимптотическая теория распределения максимума

$$M_n = \max(X_1, \ldots, X_n),$$

где $X_1, \ldots, X_n$ - независимые одинаково распределённые случайные величины, — начала активно развиваться около полувека назад, хотя её корни уходят глубоко в историю математики. В основе теории лежит фундаментальный результат, полученный Фишером и Типпетом[1] и позднее обобщённый Б.В. Гнеденко[2] — теорема об экстремальных типах, также известная как теорема Фишера-Типпета-Гнеденко, которая описывает все возможные формы распределения M_n при линейной нормировке. Строго говоря, Б.В. Гнеденко доказал, что если для некоторых числовых последовательностей $a_n > 0$, b_n случайная последовательность $a_n(M_n - b_n)$ имеет невырожденное предельное распределение, т.е. существует такая невырожденная функция распределения $G(x)$, что

$$P\left(a_n(M_n - b_n) \leq x\right) \xrightarrow[n \to \infty]{} G(x), \tag{1}$$

то G принадлежит одному из трёх экстремальных типов (т.е. существуют такие константы $a > 0$ и b, что функция распределения $H(x) := G(ax + b)$ в точности равна одной из указанных ниже функций распределения):

$$\text{Тип I:}\;\; G_0(x) = \exp(-e^{-x}),\; x \in \mathbf{R}.$$

$$\text{Тип II:}\;\; G_\gamma(x) = \begin{cases} \exp(-x^{-1/\gamma}),\; x > 0, \gamma > 0, \\ 0,\; x \leq 0. \end{cases}$$

$$\text{Тип III:}\;\; G_{-\gamma}(x) = \begin{cases} \exp(-(-x)^{1/\gamma}),\; x < 0, \gamma > 0, \\ 1,\; x \geq 0, \end{cases}$$

где $G_0(x)$ - так называемое распределения Гумбеля, $G_\gamma(x)$ - класс распределений Фреше, а $G_{-\gamma}(x)$ - класс распределений Вейбулла. Параметр γ, возникающий в теореме Фишера-Типпета-Гнеденко, называется экстремальным индексом. Пусть $F(x)$ - маргинальная функция распределения случайной последовательности $\{X_i\}_{i=1}^{+\infty}$. Тогда в случае независимых наблюдений верно

$$P(M_n \leq x) = F^n(x),$$

[1] Fisher R.A., Tippet, L.H.C.. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Phil. Soc.*, **24** (1928), 180-190.

[2] Gnedenko B.V.. Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.*, **44** (1943), 423-453.

т.е. (1) можно переписать в следующем виде:

$$F^n(a_n^{-1}x + b_n) \xrightarrow[n\to\infty]{} G(x). \quad (2)$$

Если для некоторых числовых последовательностей $a_n > 0$ и b_n выполнено (2), то говорят, что F принадлежит области притяжения распределения G, и пишут $F \in D(G)$. Известны критерии, дающие необходимые и достаточные условия принадлежности F области притяжения того или иного экстремального типа (см. например [3] или [4]). Однако есть распределения, которые не принадлежат ни одной из трёх областей максимального притяжения, например, пуассоновское распределение.

К сегодняшнему дню классическая теория экстремумов сформировалась полностью и имеет большое количество различных приложений (см. например книгу Гумбеля[5]). Позднее, начиная с работ Крамера, Лойнса, Бермана, Лидбеттера и Уотсона, возник интерес к расширению классической теории на последовательности зависимых случайных величин и на стационарные процессы с непрерывным временем. Развитие шло в следующих направлениях: создание теории для гауссовских процессов с непрерывным временем (Крамер) и стационарных последовательностей (Берман, Лидбеттер), а также доказательство результатов классической теории для некоторых типов зависимостей случайных величин (Уотсон, Лойнс).

Впоследствии Лидбеттер, Лингрен и Ротцен[3], объединив два этих направления, создали достаточно общую теорию, которая включала и полученные к тому моменту результаты для гауссовских стационарных последовательностей и процессов. Так, Лидбеттер[6] [7] показал, что результаты классической теории, при некоторых ограничениях на зависимость далеко отстоящих членов последовательности, остаются в силе для стационарных последовательностей и некоторых важных нестационарных случаев. При этом предложенное Лидбеттером условие перемешивания оказалось слабее, чем такая существовавшая на тот момент форма ограничения зависимости, как предложенное Розенблаттом[8] в 1956г. условие сильного перемешивания, использованное им при доказательстве центральной предельной теоремы для так называемых "слабо зависящих" случайных величин. Итак, приведем то

[3] Leadbetter M.R, Lindgren G., Rootźen H.. Extremes and related properties of random sequences and processes. *Springer Statistics Series.* Berlin-Heidelberg-New York: Springer, (1983).

[4] Fereira A., Haan L. de.. Extreme value theory. An introduction. *Springer Series in Operations Research and Financial Engineering. N. Y.: Springer*, (2006).

[5] Gumbel E.J.. Statistics of Extremes. *New York, Columbia Univ. Press*, год!

[6] Leadbetter M.R.. On extreme values in stationary sequences. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **28** (1974), 289-303.

[7] Leadbetter M.R.. Extremes and local dependence in stationary sequences. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **65** (1983), 291-306.

[8] Rozenblatt M.. A central limit theorem and strong mixing condition. *Proceedings of the National Academy of Sciences USA*, **42** (1956), 43-47.

условия перемешивания, о котором идет речь. Пусть

$$F_{i_1,\ldots,i_n}(x_1,\ldots,x_n) = P(X_{i_1} \le x_1,\ldots,X_{i_n} \le x_n),$$

где $\{X_i\}_{i\ge 1}$ - стационарная в узком смысле случайная последовательность. Обозначим $F_{i_1,\ldots,i_n}(u) = F_{i_1,\ldots,i_n}(u,\ldots,u)$ для любого набора индексов $i_1,\ldots,i_n$. Пусть u_n - некая числовая последовательность. По определению, случайная последовательность удовлетворяет условию $D(u_n)$, если найдётся такая последовательность чисел $\{\alpha_{n,l}\}$, $n,l = 1,2,...$, и последовательность натуральных чисел l_n, где $l_n = o(n)$ и $\alpha_{n,l_n} \to 0$ при $n \to \infty$, что для любого набора индексов $i_1,\ldots,i_p,j_1,\ldots,j_q$ такого, что

$$1 \le i_1 < \ldots < i_p < j_1 < \ldots < j_q \le n \ \text{ и } \ j_1 - i_p \ge l_n,$$

выполнено неравенство

$$\left|F_{i_1,\ldots,i_p,j_1,\ldots,j_q}(u_n) - F_{i_1,\ldots,i_p}(u_n)F_{j_1,\ldots,j_q}(u_n)\right| \le \alpha_{n,l_n}.$$

Для исследования областей притяжения в теореме об экстремальных типах в случае зависимых последовательностей (см., например, работу Лойнса[9]), а также в задаче нахождения вероятности превышения заданного уровня максимумом по зависимой последовательности (см. работу Уотсона[10]) используется условие $D'(u_n)$, гарантирующее отсутствие предельной кластеризации высоких экстремумов:

$$\limsup_{n\to\infty} n \sum_{i=2}^{[n/k]} P(X_1 > u_n, X_i > u_n) \to 0 \ \text{ при } k \to \infty,$$

где [.] обозначает целую часть числа. Точечные процессы превышения заданного уровня в применении к стохастической теории экстремумов впервые были рассмотрены Лидбеттером[11]. Он показал, что моменты превышения очень высоких уровней носят стохастический характер и что для таких уровней этот точечный процесс при выполнении условий $D(u_n)$ и $D'(u_n)$ сходится по распределению к пуассоновскому процессу. В дальнейшем метод Лидбеттера в применении к стационарным процессам был использован в таких работах, как [12] [13].

[9] Loynes R.M.. Extreme values in uniformly mixing stationary stochastic processes. *Ann. Math. Statist.*, **36** (1965), 993—999.

[10] Watson G.S.. Extreme values in samples from m-dependent stationary stochastic processes. *Ann. Math. Statist.*, **25** (1954), 798-800.

[11] Leadbetter M.R.. Point processes generated by level crossings. *In: Stochastic point processes, Ed. P. A. W. Lewis*. New York: Wiley, (1973).

[12] Кузнецов Д.С.. Предельные теоремы для максимума случайных величин. *Вестник МГУ, Сер. Матем. Механ.,* **3** (2005), 6-9.

[13] Kudrov A.V., Piterbarg V. I.. On maxima of partial samples in Gaussian sequences with pseudo-stationary trends. *Liet. matem. rink.*, **47(1)** (2007), 1-10.

Задача оценки экстремального индекса с использованием выборки как независимых, так и зависимых случайных величин, была предметом исследования большого количества авторов (см. [14] [15] [16] [17] [18]). Одной из наиболее известных оценок экстремального индекса является оценка Хилла[14]. Пусть последовательность независимых одинаково распределённых случайных величин имеет функцию распределения F, которая принадлежит области максимального притяжения Фреше. Определим

$$\widehat{\gamma}_H := \frac{1}{k}\sum_{i=0}^{k-1} \ln X_{(n-i)} - \ln X_{(n-k)},$$

где $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ - вариационный ряд выборки $\{X_i\}_{i=1}^n$. Тогда при $n \to \infty$, $k = k(n) \to \infty$ и $k/n \to 0$ оценка Хилла $\widehat{\gamma}_H$ сходится по вероятности к экстремальному индексу.

Другой, не менее известной оценкой экстремального индекса является оценка Пикандса[19]:

$$\widehat{\gamma}_P := (\ln 2)^{-1} \frac{X_{(n-k)} - X_{(n-2k)}}{X_{(n-2k)} - X_{(n-4k)}}.$$

Если $F \in D(G_\gamma)$, где $\gamma \in \mathbf{R}$, то при тех же условиях на n и k, что и для оценки Хилла, оценка Пикандса $\widehat{\gamma}_P$ сходится по вероятности к экстремальному индексу. Известно также[20], что оценка максимального правдоподобия, построенная по первым (максимальным) k членам вариационного ряда, при тех же условиях на n и k сходится по вероятности к экстремальному индексу, при условии, что $\gamma > -1/2$.

В работе[16] показано, что при выполнении условия второго порядка на функцию распределения F

$$\lim_{t\to\infty} \frac{\frac{1-F(tx)}{1-F(t)} - x^{-1/\gamma}}{A\left(\frac{1}{1-F(t)}\right)} = x^{-1/\gamma}\frac{x^{\rho/\gamma}-1}{\gamma\rho},$$

где параметры $\gamma > 0$, $\rho \leq 0$, а функция $A(t)$ — знакопостоянная, причём

[14] Hill B.M.. A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, **3** (1975), 1163-1174.

[15] Haal P.. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society*, **B44**(1982), 37-42.

[16] Hausler E., Teugels J.L.. On the asymptotic normality of Hill's estimator for the exponent or regular variation. *Ann. Statist.*, **13** (1985), 743-756.

[17] Dekkers A.L.M., de Haan L.. On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist.*, **17** (1989), 1795-1832.

[18] Hsing T.. On tail index estimation using dependent data. *Ann. Statist.*, **19** (1991), 1547-1569.

[19] Pickands J. III. Statistical inference using order statistics. *Ann. Statist.*, **3** (1975), 119-131.

[20] Drees H., Ferreyra A., de Haan L.. On maximum likelihood estimation of the extreme value index. *Ann. Appl. Probab.*, **14** (2003), 1179-1201.

$\lim\limits_{t\to\infty} A(t) = 0$, при $n \to \infty$, $k = k(n) \to \infty$ и $k/n \to 0$ оценка Хилла $\widehat{\gamma}_H$ является асимптотически нормальной

$$\sqrt{k}(\widehat{\gamma}_H - \gamma) \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \gamma^2\right),$$

где $N(0,1)$ - стандартная нормальная случайная величина, λ - конечный параметр и $\lim\limits_{n\to\infty} \sqrt{k}A\left(\frac{n}{k}\right) = \lambda$. Известно, что при похожих условиях оценки Пикандса и максимального правдоподобия также являются асимптотически нормальными оценками экстремального индекса (см. [4]).

Задача различения распределений - одна из важнейших задач математической статистики, которая являлась предметом изучения множества авторов (см. [21] [22] [23] [24]). Одним из ключевых инструментов при решении данной задачи является лемма Неймана-Пирсона, которая позволяет найти равномерно наиболее мощный критерий для различения двух произвольных распределений при использовании отношения правдоподобий. Довольно часто используется также метод отношения максимальных правдоподобий (RML-test), впервые использованный в работе[21] (см. [25] [26]). Решение некоторых задач различения распределений, принадлежащих области максимального притяжения Гумбеля, можно найти в работах [27] [28].

Метод Лапласа асимптотических оценок интегралов, содержащих большой параметр, широко известен и применяется для решения многочисленных задач математики, физики, механики и техники (см. монографию [29]). В отличие от классического метода Лапласа, когда изучается асимптотическое разложение интеграла $L(\lambda) = \int\limits_a^b f(x)e^{\lambda S(x)}dx$ при $\lambda \to +\infty$, в диссертации исследовано асимптотическое разложение интеграла $\widetilde{L}(\lambda) = \int\limits_{\lambda}^{+\infty} e^{-S(x)}dx$ при $\lambda \to +\infty$ и $\lim\limits_{x\to+\infty} \frac{S(x)}{\ln x} = +\infty$, что позволяет оценивать стохастические инте-

[21] Cox D.R.. Tests of separate families of hypotheses. *Proceedings of the Fourth Berkeley Symposium in Mathematical Statistics and Probability*, Berkeley, University of California Press, (1961), 105–123.

[22] Atkinson A.. A method for discriminating between models (with discussions), *J. R. Statist. Soc.* Ser. B **32** (1970), 323–353.

[23] Dyer A.R.. Discrimination procedure for separate families of hypotheses, *J. Am. Statist. Assoc.*, **68** (1973), 970–974.

[24] Bain L.J., Engelhardt M.. Probability of correct selection of Weibull versus gamma based on likelihood ratio. *Commun. Stat.* Ser. A **9** (1980), 375–381.

[25] Dumonceaux R., Antle C.E.. Discrimination between the log-normal and the Weibull distributions. *Technometrics*, **15 (4)** (1973), 923–926.

[26] Dumonceaux R., Antle C.E., Haas G.. Likelihood ratio test for discrimination between two models with unknown location and scale parameters. *Technometrics*, **15 (1)** (1975), 19–27.

[27] Gupta R.D., Kundu D.. Discriminating between Weibull and generalized exponential distributions. *Computational Statistics and Data Analysis*, **43 (2)** (2003), 179 – 196.

[28] Kundu D., Raqab M.Z.. Discriminating Between the Generalized Rayleigh and Log-Normal Distribution. *Statistics*, **41 (6)** (2007), 505-515.

[29] Федорюк М.В.. Метод перевала. М., (1977).

гралы, возникающие в задаче различения близких гипотез.

Задача различения близких гипотез о распределениях является ключевой для теории контигуальных мер (см. монографию[30], а также [31]). Пусть $\{f(x,\theta),\ \theta \in \Theta\}$ - некое семейство плотностей распределений. При выполнении некоторых условий регулярности, наложенных на это семейство плотностей (подробнее см. [32]), распределение логарифма отношения правдоподобий при условии, что выборка взята из распределения с плотностью $f(x,\theta)$, слабо сходится к нормальному закону при $n \to \infty$

$$\mathcal{L}\left(\ln \frac{\prod\limits_{i=1}^{n} f(X_i, \theta + t(n)h)}{\prod\limits_{i=1}^{n} f(X_i, \theta)}\right) \Rightarrow N\left(-\frac{1}{2}\Gamma(\theta)h^2, \Gamma(\theta)h^2\right),$$

где $t(n) = n^{-1/2}$, h — произвольная константа, а $\Gamma(\theta)$ — функция, зависящая от распределения.

Цель и задачи исследования.

Целью диссертации является изучение предельного поведения статистических оценок экстремального индекса в условиях зависимости и загрязнения наблюдений и проверка близких статистических гипотез об экстремальном индексе распределения. Среди задач исследования выделяются следующие.

- Асимптотический анализ поведения точечных процессов высоких экстремумов при условии загрязнения данных.
- Изучение асимптотических свойств оценок экстремального индекса в условиях растущего загрязнения.
- Построение общих критериев различения гипотез о распределениях из области максимального притяжения Гумбеля.

Научная новизна.

Представленные в диссертации результаты являются новыми, полученными автором самостоятельно. Основные результаты диссертации следующие.

- Доказана пуассоновская предельная теорема для высоких экстремумов временного ряда с сезонной составляющей и монотонным трендом.

[30]Русас Дж.. Контигуальность вероятностных мер. М.: Мир, (1975).

[31]Чибисов Д.М.. Лекции по асимптотической теории ранговых критериев. *МИАН, Лекц. курсы НОЦ*, **14** (2009), 3–174.

[32]Le Cam L.. On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *Ann. Math. Stat.*, **41** (1970), 802-828.

- Доказана состоятельность и асимптотическая нормальность оценки Хилла экстремального индекса для модели с асимптотически растущим аддитивным загрязнением.

- Найдено асимптотическое распределение отношения правдоподобий, построенного по первым членам вариационного ряда выборки, для близких гипотез о распределениях вейбулловского и логвейбулловского типов.

Методика исследования.

В диссертации используются как классические методы теории вероятностей и теории случайных процессов, такие как предельная теорема Фишера-Типпета-Гнеденко[1,2], метод отношения правдоподобий[30], метод представления Реньи[33], теорема Калленберга[34], так и методы математического анализа, такие, как метод Лапласа и метод перевала[29], а также специальные методы стохастической теории экстремумов, такие как лемма Лидбеттера[11].

Теоретическая и практическая значимость.

Диссертация носит теоретический характер. Ее результаты могут найти применение в стохастической теории экстремумов, теории контигуальных мер, а также использоваться в финансовой сфере.

Апробация работы.

По теме диссертации были сделаны доклады на следующих семинарах механико-математического факультета МГУ им. М. В. Ломоносова:

- Большом семинаре кафедры теории вероятностей под руководством действительного члена РАН, профессора А. Н. Ширяева (2013 г.),

- Семинаре "Гауссовские случайные процессы"под руководством проф. В.И. Питербарга (2009-2013 гг., неоднократно),

а также на семинаре “Статистика экстремальных значений” под руководством д.ф.-м.н., гл.н.с. Н.М. Маркович, Институт проблем управления РАН (2013).

Результаты диссертации докладывались на:

- Международной конференции ”Ломоносов-2011” (МГУ, Москва, 2011)

- Международной конференции ”Теория вероятностей и её приложения”, посвященной столетию со дна рождения Б.В.Гнеденко (МГУ, Москва, июнь 2012)

- Международной конференции ”Ломоносов-2013” (МГУ, Москва, 2013)

[33]Reiss R. Approximate Distributions of Order Statistics. *Springer, Berlin*, (1989).

[34]Kallenberg O. Random measures. *N.Y.: Academic Press*, (1983).

- International Workshop "Stochastic, Analysis and Geometry" (Ульм, сентябрь 2013);

Публикации.

Результаты диссертации опубликованы в пяти работах, из которых две — в журналах из перечня ВАК, одна депонирована в ВИНИТИ. Список работ приведен в конце автореферата [1-5].

Структура и объем работы.

Диссертация изложена на 70 страницах и состоит из введения, двух глав и списка литературы, включающего 46 наименований.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** приводится краткий обзор исследований, посвященных вопросам статистической теории экстремумов. Историческая справка подкрепляется соответствующими ссылками на научные работы. Кроме того, во введении объясняется актуальность темы и научная новизна предпринятого автором исследования.

В **первой главе** диссертации рассматриваются статистические оценки характеристик экстремумов в модели выборок с загрязнениями. В первом параграфе рассматривается модель

$$Y_{i,n} = X_i + a_n m_i + a_n h\left(\frac{i}{n}\right),\ i, n \in \mathbf{N}, \mathbf{n} \to \infty \tag{3}$$

где $\{X_i,\ i = 1, 2, ...\}$ — строго стационарная случайная последовательность, $\{m_i,\ i = 1, 2, ...\}$ — тренд, ведущий себя стационарным образом в определенном ниже смысле (например, сезонная составляющая), h — монотонная функция, a_n — последовательность нормировочных констант из теоремы Фишера-Типпета-Гнеденко[35]. Предполагается, что функция распределения $F(x)$ случайной величины X_1 является максимально устойчивой. Далее, определяется случайный загрязненный процесс высоких экстремумов ограниченных борелевских подмножествах B действительной прямой:

$$\xi_z^n(B) = \#\{k : Y_k > a_n z + b_n, k \in nB\},\quad n = 1, 2, ..., \tag{4}$$

где символом $\#$ обозначается число точек конечного множества; a_n, b_n — те же последовательности нормировочных констант из теоремы Фишера—Типпета—Гнеденко, соответствующих функции распределения F; z — действительное число. На точечный процесс высоких экстремумов налагаются следующие условия.

[35] Fereira A., Haan L. de.. Extreme value theory. An introduction. *Springer Series in Operations Research and Financial Engineering. N. Y.: Springer*, (2006).

Условие 1. Последовательность $\{m_i,\ i=1,2,...\}$ ограничена сверху:

$$m := \sup_{i=1,2,...} m_i < \infty.$$

Обозначим

$$w_i = a_n x + b_n - a_n m_i - a_n h(\frac{i}{n}),$$
$$w^n = a_n x + b_n - a_n m - a_n h,$$

где $h = \sup_{x \in B} h(x)$. Обозначим также $\xi_z(I) = \#\{k : Y_k > a_n z + b_n, k \in I\}$. Введем условие типа Лидбеттера[36] слабой зависимости далеко отстоящих больших значений в модели (3).

Условие 2. Найдутся семейство чисел $\{\alpha_{n,l}\}$, $n,l = 1,2,...$, и последовательность натуральных чисел $\{l_n\}$, такие, что $l_n = o(n)$, $\alpha_{n,l_n} \to 0$, для любых z, для любых неотрицательных целых r и s и произвольных множеств натуральных чисел $I = \{i_1,...,i_p\}$, $J = \{j_1,...,j_q\}$, таких, что

$$1 \leqslant i_1 < i_2 < ... < i_p < j_1 < ... < j_q \leqslant n, \quad j_1 - i_p \geqslant l_n,$$

выполняется неравенство

$$|P(\xi_z(I) = r, \xi_z(J) = s) - P(\xi_z(I) = r)P(\xi_z(J) = s)| \leq \alpha_{n,l_n}.$$

Следующее условие, как и предыдущее, являющееся стандартным для вероятностной теории экстремумов, гарантирует отсутствие предельной кластеризации высоких экстремумов[36].

Условие 3. Выполняется равенство

$$\lim_{k\to\infty} \limsup_{n\to\infty} n \sum_{2 \leqslant j \leqslant n/k} P\{X_1 > w^n; X_j > w^n\} = 0.$$

Далее, вводится "выборочная функция распределения" значений сезонной составляющей для наблюдаемых Y_i, а также аналогичную функцию для монотонного тренда

$$\Phi_n(x) = \frac{\#\{i : m_i \leqslant x,\ 1 \leqslant i \leqslant n\}}{n}, \quad H_n^B(x) = \frac{\#\{i : h(\frac{i}{n}) \leqslant x,\ i \in nB\}}{n}.$$

Обозначим

$$\Phi(x) = \lim_{n\to\infty} \Phi_n(x);\ H^B(x) = \lim_{n\to\infty} H_n^B(x).$$

[36]Leadbetter M.R, Lindgren G., Rootźen H.. Extremes and related properties of random sequences and processes. *Springer Statistics Series.* Berlin-Heidelberg-New York: Springer, (1983).

Обозначим также

$$f_\gamma(z) = \begin{cases} e^{-z},\ z \in R \text{ если } \gamma = 0; \\ z^{-1/\gamma} I\{z \geq 0\},\ \text{если } \gamma > 0; \\ (-z)^{-1/\gamma} I\{z \leq 0\},\ \text{если } \gamma < 0. \end{cases}$$

Основной результат первого параграфа касается предельного распределения случайного точечного процесса $\xi_z^n(B)$.

Теорема 1. *Пусть в модели (4) $F \in D(G_\gamma)$, где $\gamma \in R$. Предположим, что выполнены условия 1—3. Тогда если $\gamma \geq 0$, то для всех z, а если $\gamma < 0$, то для всех $z > m$ случайный точечный процесс экстремумов $\xi_z^n(B)$ слабо сходится при $n \to \infty$ к пуассоновскому точечному процессу с функцией интенсивности*

$$\lambda_z(B) = \int\limits_{-\infty}^{+\infty} f_\gamma(z-t) d(\mu(B) \cdot \Phi(t) + H^B(t)),$$

где μ — мера Лебега на действительной прямой.

Во втором параграфе первой главы исследуется состоятельность и асимптотическая нормальность оценки Хилла экстремального индекса[37] в модели

$$Y_{i,n} = X_i + m_{i,n},\ n = 1, 2, ..., \tag{5}$$

где X_i — последовательность независимых, одинаково распределенных случайных величин, функция распределения которых F принадлежит области максимального притяжения Фреше с индексом γ; $m_{i,n}$, $1 \leq i \leq n$, $n = 1, 2, \ldots,$ — аддитивное загрязнение, максимальная величина которого может расти вместе с n. Обозначим $c_n := \max\{|m_{i,n}|, 1 \leq i \leq n\}$, будем считать, что асимптотика последовательности c_n нам известна. Оценкой Хилла экстремального индекса γ для модели (5) назовём статистику

$$\hat{\gamma}_H^{(y)} = \frac{1}{k} \sum_{i=0}^{k-1} \ln(Y_{(n-i)}) - \ln(Y_{(n-k)}), \tag{6}$$

где $(Y_{(1)} \leq Y_{(2)} \leq ... \leq Y_{(n)})$ — вариационный ряд выборки $\{Y_{i,n}\}_{i=1}^n$. Предположим, что функция распределения F имеет плотность, обозначим ее F'. Автором показано, что в случае, если загрязнение $m_{i,n}$ существенно меньше a_n, где a_n - последовательность нормировочных констант из теоремы Фишера-Типпета-Гнеденко, оценка Хилла (6) сохраняет свойства состоятельности и асимптотической нормальности.

Основным результатом параграфа является следующая

[37]Hill B.M.. A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, **3** (1975), 1163-1174.

Теорема 2. *Пусть для некоторой последовательности* k_n, *такой, что* $k_n/n \to 0$ *при* $n \to \infty$ *и* $k_n \to \infty$, *выполняется соотношение* $\frac{c_n}{Y_{(n-k_n)}} \xrightarrow{P} 0$ *(или, что эквивалентно, существует* $\varepsilon > 0$, *такое, что* $c_n/a_n = O(k(n)^{-\gamma-\varepsilon})$*). Тогда оценка Хилла (6) сходится по вероятности к* γ *при* $n \to \infty$. *Кроме того, если функция распределения* F *удовлетворяет условию*

$$\lim_{t\to\infty} \frac{\frac{1-F(tx)}{1-F(t)} - x^{-1/\gamma}}{A(\frac{1}{1-F(t)})} = x^{-1/\gamma}\frac{x^{\rho/\gamma}-1}{\gamma\rho}$$

для всех $x > 0$, *где* $\rho \le 0$ *и* $A(t)$ *— положительная или отрицательная функция с* $\lim_{t\to\infty} A(t) = 0$, *то при всех* k_n, *таких, что* $\frac{c_n k_n^{1/2}}{Y_{(n-k_n)}} \xrightarrow{P} 0$ *(или, что эквивалентно,* $\frac{c_n}{a_n} = O\left(k_n^{-\gamma-\frac{1}{2}-\varepsilon}\right)$ *для некоторого* $\varepsilon > 0$ *при* $n \to \infty$*), верно соотношение*

$$\sqrt{k_n}(\hat{\gamma}_H^{(y)} - \gamma) \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \gamma^2\right),$$

где N *— стандартное нормальное распределение,* $k_n \to \infty, k_n/n \to 0, n \to \infty$ *и*

$$\lim_{n\to\infty} \sqrt{k_n} A\left(\frac{n}{k_n}\right) = \lambda$$

с конечным λ.

Вторая глава диссертации посвящена задаче нахождения критериев различения близких гипотез о хвостах распределений из области максимального притяжения Гумбеля. Пусть $\{X_i, i = 1, \ldots, n\}$ - независимые одинаково распределенные случайные величины, имеющие плотность распределения $f(x,\gamma)$. Исследуется асимптотическое поведение отношения правдоподобий

$$R_n(u) = \frac{L(X_{n,n}, \ldots, X_{n-k_n+1,n}; \gamma + t(k_n, u))}{L(X_{n,n}, \ldots, X_{n-k_n+1,n}; \gamma)}, \tag{7}$$

построенного по верхним k_n членам вариационного ряда выборки при $n \to \infty, k_n \to \infty, \frac{n}{k_n} \to \infty$ для семейства плотностей

$$f(x,\gamma) = C(x,\gamma)\exp(-V(x,\gamma)). \tag{8}$$

Обозначим

$$S(x,\gamma) = V(x,\gamma) - \ln C(x,\gamma).$$

Будем считать, что $C(x,\gamma) = C_1(\gamma) + C_2(\gamma)x^{-\beta} + R(x)$, $\beta > 0$, при $x \to \infty$, где остаточный член $R(x) = o(x^{-\beta})$ не зависит от γ и четырежды непрерывно дифференцируем, а функция $S(x,\gamma)$ строго монотонна по x и четырежды непрерывно дифференцируема по паре переменных при $x > x_0(\gamma) > 0$. Следующие условия регулярности, налагаемые нами на семейство распределений $f(x,\gamma)$, назовём условиями регулярности первого типа.

1. Существует $\varepsilon = \varepsilon(\gamma) > 0$ такой, что $\frac{S(x,\gamma)}{x^{1+\varepsilon}} \to +\infty$ при $x \to +\infty$.

2. Любая из частных производных функции $S(x,\gamma)$ вплоть до четвертого порядка $\neq 0$ или $\equiv 0$ при $x > x_0(\gamma)$, а также имеет конечный или бесконечный предел при $x \to +\infty$.

3. Существует конечный предел при $x \to +\infty$ выражений вида $\frac{F_x(x,\gamma)S(x,\gamma)}{F(x,\gamma)S_x(x,\gamma)}$, где $F(x,\gamma)$ — любая частная производная функции $S(x,\gamma)$ вплоть до третьего порядка и $F(x,\gamma) \neq 0$ при $x > x_0(\gamma)$.

4. Для $k = 1,2,3$ выполнено $\lim\limits_{x\to+\infty} \frac{\ln\left|\frac{\partial^k S(x,\gamma)}{\partial\gamma^k}\right|}{\ln S(x,\gamma)} = 1$, если эти частные производные не равны 0 тождественно.

Стоит заметить, что класс распределений, удовлетворяющий условиям регулярности первого типа, довольно широк и содержит такие семейства распределений, как, например, семейство вейбулловского типа или нормальное семейство распределений. Обозначим

$$a_{n/k_n} = F^{\leftarrow}\left(\frac{n-k_n}{n}, \gamma\right),$$

где $F(x,\gamma)$ - функция распределения, имеющая плотность $f(x,\gamma)$, а $F^{\leftarrow}(x,\gamma) = \inf\{t:\ F(x,\gamma) \leq t\}$. Также будем писать $S_{x\gamma}(a_{n/k_n}, \gamma + t(k_n,u))$ вместо $\frac{\partial^2 S(x,\gamma)}{\partial x \partial\gamma}\Big|_{(x,\gamma)=(a_{n/k_n},\gamma+t(k_n,u))}$ и аналогично для других частных производных функции $S(x,\gamma)$. В следующей теореме найдено предельное распределение отношения правдоподобий $R_n(u)$ для лёгких хвостов при медленно растущей последовательности k_n:

Теорема 3. *Пусть для семейства плотностей $f(x,\gamma)$ выполнены условия регулярности первого типа, $n \to \infty$, $k_n \to \infty$, причём выполнено следующее условие: $\exists\varepsilon$, $0 < \varepsilon < 2$, такой, что*

$$\lim_{n\to\infty} \frac{k_n}{(\ln\frac{n}{k_n})^{\varepsilon}} = 1.$$

Тогда при $t(k_n,u) = \frac{u}{\sqrt{k_n}} \frac{S_x(a_{n/k_n},\gamma)}{S_{x\gamma}(a_{n/k_n},\gamma)}$ и $n \to \infty$ выполнено

$$\ln R_n(u) \xrightarrow{d} N\left(-\frac{u^2}{2}, u^2\right).$$

Если же исследовать асимптотическое поведение отношения правдоподобий $R_n(u)$ (7) при $n \to \infty, k_n \to \infty$ и быстро растущей последовательности k_n ($k_n \sim n^{\alpha}$, где $\alpha \in (0,1)$) для семейства плотностей $f(x,\gamma)$ (8), то нам потребуются несколько другие условия регулярности, налагаемые нами на это семейство:

1. Существует $\varepsilon = \varepsilon(\gamma) > 0$ такой, что $\frac{S(x,\gamma)}{(\ln x)^{1+\varepsilon}} \to +\infty$ при $x \to +\infty$.

2. Существует $\delta = \delta(\gamma)$, $\delta \in [0, 1)$, такое, что $\lim\limits_{x\to+\infty} \frac{\ln S_x(x,\gamma)}{(S(x,\gamma))^{1-\delta}} = 0$.

3. Любая из частных производных функции $S(x, \gamma)$ вплоть до четвертого порядка $\neq 0$ или $\equiv 0$ при $x > x_0(\gamma)$, а также имеет конечный или бесконечный предел при $x \to +\infty$.

4. Существует конечный предел при $x \to +\infty$ выражений вида $\frac{F_x(x,\gamma)S_x(x,\gamma)}{F(x,\gamma)S_xx(x,\gamma)}$, где $F(x, \gamma)$ — любая частная производная функции $S(x, \gamma)$ вплоть до третьего порядка и $F(x, \gamma) \neq 0$ при $x > x_0(\gamma)$.

5. Для $k = 1, 2, 3$ выполнено $\lim\limits_{x\to+\infty} \frac{\ln\left|\frac{\partial^k S(x,\gamma)}{\partial\gamma^k}\right|}{\ln S(x,\gamma)} = 1$, если эти частные производные не равны 0 тождественно.

Класс распределений, удовлетворяющий условиям регулярности второго типа, содержит такие семейства распределений, как, например, семейства лог-вейбулловского и вейбулловского типов или лог-нормальное семейство распределений. Обозначим

$$H(x) = \sqrt{k_n}t(k_n)\left(\frac{\int\limits_x^\infty S_\gamma(y,\gamma)\exp(-S(y,\gamma))dx}{\int\limits_x^\infty \exp(-S(y,\gamma))dx} - S_\gamma(x,\gamma) - \right.$$

$$\left. -\frac{S_{x\gamma}(x,\gamma)}{S_x(x,\gamma)} - \frac{S_{xx\gamma}(x,\gamma)}{(S_x(x,\gamma))^2} + 2\frac{S_{xx}(x,\gamma)S_{x\gamma}(x,\gamma)}{(S_x(x,\gamma))^3}\right).$$

При этом функция $H(x) = -\frac{S_{xx}(x,\gamma)}{S_x^2(x,\gamma)}(1 + o(1))$ при $x \to +\infty$.

Теорема 4. *Пусть для семейства плотностей $f(x, \gamma)$ выполнены условия регулярности второго типа, $n \to \infty$, $k_n \to \infty$, причём выполнено следующее условие: $\exists\varepsilon$, $0 < \varepsilon < 2$, такой, что*

$$\lim_{n\to\infty} \frac{k_n}{n^\varepsilon} = 1.$$

Тогда при $t(k_n) = \frac{u}{\sqrt{k_n}} \frac{S_x(a_{n/k_n},\gamma)}{S_{x\gamma}(a_{n/k_n},\gamma)}$ выполнено

$$\ln R_n(u) - \sqrt{k_n}H(a_{n/k_n}) \xrightarrow{d} \exp\left(-N\left(\frac{u^2}{2}, u^2\right)\right)$$

при $n \to +\infty$.

Стоит заметить, что как в теореме 3, так и в теореме 4 предельное распределение логарифма отношения правдоподобий $R_n(u)$ не зависит от γ, что весьма удобно для построения критерия.

Далее, во второй главе также исследовалась задача нахождения вероятностей редких событий для рассматриваемых типов распределений. Изучается поведение интеграла $F(q) = \int\limits_q^{+\infty} \exp(-S(x))dx$ при $q \to +\infty$. Пусть для функции $S(x)$ выполнены следующие условия регулярности:

1. $\frac{S(x)}{\ln x} \to +\infty$ при $x \to +\infty$.
2. Начиная с некоторого $x_0 > 0$, $S(x)$ – строго монотонная функция.
3. Начиная с некоторого $x_1 > 0$, $S(x)$ четырежды непрерывно дифференцируема. Кроме того, первая, вторая и третья производные функции $S(x,\gamma)$ имеют конечный или бесконечный предел при $x \to +\infty$.
4. Для $k = 1, 2, 3$ $\lim\limits_{x \to +\infty} \frac{\frac{d^{k+1}S(x)}{dx^{k+1}}}{S'(x)\frac{d^kS(x)}{dx^k}} = 0$.

Легко видеть, что как условия регулярности первого типа, так и условия регулярности второго типа оказываются слабее данных условий, поэтому следующий результат может применяться для оценки стохастических интегралов, возникающих в задаче различения близких гипотез во второй главе диссертации.

Лемма 1. *Пусть функция $S(x)$ удовлетворяет условиям регулярности 1)-4).*
Тогда при $q \to +\infty$

$$F(q) = \exp(-S(q)) \left(\sum_{k=0}^{2} c_k + o(c_2) \right),$$

где $c_k = M^k \left(\frac{1}{S'(x)} \right) |_{x=q}$, $M = \frac{1}{S'(x)} \frac{d}{dx}$.

Автор благодарит научного руководителя доктора физико-математических наук, профессора Питербарга Владимира Ильича за постановку задач и постоянное внимание к работе. Автор высоко ценит содействие, оказанное работе сотрудниками кафедры теории вероятностей механико-математического факультета МГУ имени М. В. Ломоносова.

РАБОТЫ АВТОРА ПО ТЕМЕ ДИССЕРТАЦИИ

[1] Родионов, И. В., “Пуассоновская предельная теорема для высоких экстремумов временного ряда с сезонной составляющей и строго монотонным трендом”. *Вестн. Моск. ун-та. Сер. 1. Матем. Механ.*, **1**, 12–17 (2012).

[2] Родионов, И. В., “О свойствах оценки Хилла экстремального индекса для выборок с загрязнениями”. *Вестн. Моск. ун-та. Сер. 1. Матем. Механ.*, **1**, 3–6 (2014).

[3] Родионов, И. В., “Асимптотическое разложение Лапласа вероятностей редких событий для одного класса распределений из области максимального притяжения Гумбеля”. *Современные проблемы математики и механики, Математика*, **8(3)**, 108–115 (2013).

[4] Родионов, И. В., “О различении близких гипотез о распределениях вейбулловского и логнормального типов по первым членам вариационного ряда”. *Деп. в ВИНИТИ* 03.12.2013, №349-В2013, 37 (2013).

[5] Родионов, И. В., “Discrimination between close hypotheses about distributions of the Weibull and the log-Weibull types by the first order statistics”. Международная конференция “Теория Вероятностей и ее Приложения”, посвященная 100-летию со дня рождения Б.В. Гнеденко, Тезисы докладов, 172–173, Ленанд, Москва, 2012.