На правах рукописи

Новак Сергей Юрьевич

ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ И ОЦЕНКИ СКОРОСТИ СХОДИМОСТИ В ТЕОРИИ ЭКСТРЕМАЛЬНЫХ ЗНАЧЕНИЙ

01.01.05 – теория вероятностей и математическая статистика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

доктора физико-математических наук

Санкт-Петербург

2013

Работа выполнена на кафедре экономики и статистики факультета науки и технологии Мидлсекского университета.

Официальные оппоненты:

Гущин Александр Александрович
доктор физико-математических наук, профессор, ведущий научный сотрудник
ФГБУН Математический институт им. В.А. Стеклова
Российской академии наук

Невзоров Валерий Борисович
доктор физико-математических наук, профессор, профессор
Математико-механического факультета ФГБОУ ВПО
«Санкт-Петербургский государственный университет»

Питербарг Владимир Ильич
доктор физико-математических наук, профессор, главный научный сотрудник
ФГБОУ ВПО «Московский государственный университет им. М.В. Ломоносова».

Ведущая организация:

ФГБУН Институт математики им. С.Л. Соболева
Сибирского отделения Российской академии наук.

Защита состоится « ____ » ____________ 2014 г. в « ____ » часов на заседании диссертационного совета Д 002.202.01 в ФГБУН Санкт-Петербургского отделения Математического института им. В.А. Стеклова Российской академии наук по адресу:
191023, Санкт-Петербург, наб. р. Фонтанки, д. 27, к. 311.

С диссертацией можно ознакомиться в библиотеке ФГБУН Санкт-Петербургское отделение Математического института им. В.А. Стеклова Российской академии наук.

Автореферат разослан « ____ » ____________ 2013 г.

Ученый секретарь
диссертационного совета Д 002.202.01,
доктор физико-математических наук А.Ю. Зайцев

Общая характеристика работы

Актуальность темы. Теория экстремальных значений является одним из наиболее динамично развивающихся разделов теории вероятностей и математической статистики. Её истоком можно считать классическую теорему Пуассона об асимптотике распределения числа редких событий; ряд задач имеет более глубокую историю (см., к примеру, Муавр (1738), Задача LXXIV).

Актуальность исследования асимптотических свойств распределений экстремальных значений связана с приложениями в страховом деле, финансах, метеорологии, гидрологии (см. Эмбрехтс, Клюпельберг, Микош (1997), Бейрлант, Гогебер, Тойгельс, Сегерс (2004)). К примеру, популярной мерой риска, используемой крупнейшими банками, является VaR (экстремальная квантиль). Задача оценивания вероятности выхода за высокий уровень имеет приложения в страховом деле.

Основы современной теории экстремальных значений заложили в начале 20-го века Мизес (1923, 1936), Фреше (1927), Фишер и Типет (1928), Гнеденко (1943). Работа де Хаана (1970) завершает классический период развития теории, посвящённый изучению распределений экстремальных значений в последовательностях независимых одинаково распределённых случайных величин (с.в.).

В то время как классическая теория экстремальных значений имеет дело с последовательностями независимых одинаково распределённых с.в., финансовые приложения часто демонстрируют зависимость наблюдений. Это делает актуальным изучение асимптотических свойств распределений экстремальных значений в стационарных последовательностях случайных величин.

Значительный вклад в развитие теории экстремальных значений для последовательностей стационарно связанных случайных величин внесли Ньюэл (1964) и Лойнес (1965), которые фактически ввели понятие экстремального индекса. Дальнейшее развитие теории связано с работами Бермана (1962), Лидбеттера (1974), О'Брайена (1974, 1987), Мори (1977), Хсина (1987) и др..

Хсин, Хюслер и Лидбеттер (1988) установили, что предельным распределением одномерного эмпирического точечного процесса выходов за высокий уровень, учитывающего местоположение экстремумов, является сложно-пуассоновское распределение. Это связано с тем, что в последовательностях зависимых случайных величин экстремальные значения обычно появляются кластерами.

Мори (1977) показал, что класс распределений общих процессов выходов за высокий уровень в последовательностях стационарно связанных с.в. богаче класса сложно-пуассоновских процессов. Хсин (1987) охарактеризовал предельное распределение общего двумерного процесса выходов за высокий уровень в последовательностях стационарно зависимых случайных величин в терминах двумерных точечных процессов.

Диссертация посвящена исследованию асимптотики распределения случайных величин и процессов, возникающих в теории экстремальных значений для последовательностей стационарно связанных с.в.. Рассматриваются такие задачи, как характеризация класса $\mathcal{P}$ предельных распределений общих точечных процессов, возникающих в теории экстремальных значений, оценивание скорости сходимости в соответствующих предельных теоремах, статистическое оценивание характеристик распределений, рассматриваемых в теории экстремальных значений, установление нижних границ точности оценивания характеристик распределений с тяжёлыми хвостами.

В диссертации получена характеризация распределений двумерных точечных процессов из класса $\mathcal{P}$ в терминах одномерных точечных процессов, описаны свойства распределений из класса $\mathcal{P}$, установлены свойства маргинальных распределений.

Важную роль при изучении асимптотики распределения экстремальных значений играет задача установления оценок скорости сходимости в соответствующих предельных теоремах. Вопрос является нетривиальным даже в случае теоремы Пуассона. Многие известные авторы работали над указанной задачей, в том числе Прохоров (1952), Лекам (1965), Серфлин (1975), Чен (1975), Шоргин (1977), Барбур и Иглсон (1983), Барбур и Холл (1984), Деовельс и Пфайфер (1986, 1988).

Асимптотику расстояния по вариации в теореме Пуассона в случае независимых одинаково распределённых случайных величин установил Прохоров (1952). Роос (2001) получил оценку точности пуассоновской аппроксимации в терминах расстояния по вариации с неулучшаемой константой. Однако вопрос о точности сложно-пуассоновской аппроксимации долгое время оставался открытым, равно как и вопрос о точности пуассоновской аппроксимации в ряде задач теории экстремальных значений для выборок случайного объёма. Решению этих задач посвящена одна из глав диссертации.

В статистике экстремальных значений основное внимание уделяется задачам оценивания характеристик распределений с тяжёлыми хвостами. Актуальность указанной тематики связана с приложениями к финансам и страховому делу, где наблюдения зачастую оказываются зависимыми, а их распределения имеют тяжёлый хвост.

Основной характеристикой распределения с тяжёлым хвостом является показатель скорости убывания хвоста распределения. Оценка показателя скорости убывания хвоста распределения входит в конструкцию оценок экстремальной квантили и вероятности выхода за высокий уровень в последовательности стационарно связанных случайных величин.

В последние десятилетия эта тематика развивается весьма интенсивно (см. Хилл (1975), Холл (1982), Хойслер и Тойгельс (1985), Голди и Смит (1987), Декерс, Айнмаль, де Хаан (1989), Эмбрехтс, Клюпельберг, Микош (1997), Бейрлант, Гогебер, Тойгельс, Сегерс (2004)). В диссертации предложены новые оценки показателя скорости убывания хвоста распределения, экстремальной квантили, вероятности выхода за высокий уровень, доказана их состоятельность и асимптотическая нормальность при минимальных ограничениях на коэффициенты перемешивания, построены под-асимптотические доверительные интервалы, предложен алгоритм выбора управляющего параметра непараметрических оценок.

Важным направлением в статистике экстремальных значений является тема нижних границ точности оценивания характеристик неизвестного распределения. Этой тематике посвящены, в частности, работы Холл и Вэлш (1984), Донохо и Лю (1991), Пфанцаль (2000), Дреес (2001), Бейрлант, Буко, Веркер (2006). Однако имеющаяся литература даёт лишь частичное решение указанной задачи: найден порядок скорости убывания нижней границы, асимптотическая нижняя граница выводится при ограничениях на класс рассматриваемых оценок.

В диссертации впервые получены неасимптотические нижние границы точности оценивания характеристик распределений с тяжёлыми хвостами, выявлены соответствующие информационные функционалы.

Многие оценки в статистике экстремальных значений входят в группу статистик, являющихся самонормированными суммами (СНС) случайных величин. Таковы ряд оценок показателя скорости убывания хвоста распределения, экстремального индекса, элементы конструкции оценок экстремальной квантили и вероятности выхода за высокий уровень. Группа СНС статистик включает также статистику Стьюдента, ядерную оценку функции регрессии, оценку функции интенсивности отказов.

Раздел статистики, связанный с самонормированными суммами случайных величин, интенсивно развивается в последние десятилетия (см. Чун (1946), Эфрон (1969), Малер (1981), Славова (1985), Холл (1987), Бенткус и Гётце (1996), Жине, Гётце, Мейсон (1997), Шао (1997), Чистяков (2001)).

В диссертации получены оценки скорости сходимости в ЦПТ для распределений самонормированных сумм независимых и стационарно связанных случайных величин; решена долго остававшаяся открытой задача получения оценок скорости сходимости с явными константами; показано, что в неравенстве типа Берри–Эссеена для статистики Стьюдента константа не может быть лучше, чем $1/\sqrt{2e}$; установлено, что аналог неравномерного неравенства Берри–Эссеена, вообще говоря, не имеет места для самонормированных сумм случайных величин.

Цель работы. Основная цель работы — исследование асимптотических свойств распределений случайных величин и процессов, применяемых в задачах теории экстремальных значений, характеризация класса предельных распределений соответствующих случайных величин и процессов, получение оценок скорости сходимости в указанных предельных теоремах, разработка статистических методов оценивания характеристик распределений с тяжёлыми хвостами по выборкам стационарно связанных случайных величин, установление нижних границ точности оценивания характеристик распределений с тяжёлыми хвостами, выявление соответствующих информационных функционалов.

Научная новизна. Все основные результаты диссертации являются новыми и состоят в следующем:

1) Дана характеризация класса $\mathcal{P}$ предельных распределений эмпирического точечного процесса выходов за высокий уровень (ПВВУ). Элемент $P \in \mathcal{P}$, являющийся двумерным точечным процессом, охарактеризован как процесс, являющийся композицией двух одномерных точечных процессов: пуассоновского $\pi(\cdot)$ и процесса $\gamma(\cdot)$ со стохастически непрерывными траекториями и маргинальными распределениями, удовлетворяющими условию (15). Найдены необходимые и достаточные условия слабой сходимости распределения эмпирического точечного процесса выхода за высокий уровень к распределению произвольного заданного процесса $P \in \mathcal{P}$.

2) Получены оценки скорости сходимости в предельных теоремах для процесса выходов за высокий уровень. Точность аппроксимации распределения процесса выходов за высокий уровень соответствующим кластер-пуассоновским распределением оценена в терминах расстояния по вариации.
3) Предложен метод оценивания и получены оценки точности сложно-пуассоновской аппроксимации для распределения вектора количеств выходов за высокий уровень, получены утверждения типа ЗПЛ и оценки точности пуассоновской аппроксимации распределения числа длинных "повторов" в случайных последовательностях (задача имеет приложения к анализу последовательностей ДНК), а также оценки точности пуассоновской аппроксимации в ряде других задач теории экстремальных значений, получено обобщение на многомерный случай теоремы Бредли (1983) о задании независимой копии случайного вектора на одном вероятностном пространстве.
4) Предложен новый подход к изучению асимптотики распределения самонормированных сумм (СНС) случайных величин, решена долго остававшаяся открытой задача получения оценок типа Берри–Эссеена с явными константами для СНС случайных величин, в том числе для статистики Стьюдента. Оценки с явными константами получены впервые. На основе указанных оценок построены под-асимптотические доверительные интервалы для оценок показателя скорости убывания хвоста распределения. Установлено, что неравномерное неравенство типа Берри–Эссеена в общем случае не имеет места для самонормированных сумм с.в.. Впервые получены оценки скорости сходимости в ЦПТ для распределений самонормированных сумм стационарно связанных с.в..
5) В задачах статистического оценивания характеристик распределений с тяжёлыми хвостами предложены непараметрические оценки показателя скорости убывания хвоста распределения, вероятности выхода за высокий уровень и экстремальной квантили, доказана их состоятельность и асимптотическая нормальность в условиях слабой зависимости при минимальных ограничениях на коэффициенты перемешивания, предложена процедура выбора управляющего параметра непараметрических оценок.
6) Получены нижние границы точности оценивания характеристик распределений с тяжёлыми хвостами, выявлены соответствующие информационные функционалы.

Основные результаты работы.

1. Дана характеризация класса предельных распределений общих эмпирических точечных процессов, возникающих в теории экстремальных значений.

2. Установлены оценки скорости сходимости в предельных теоремах для распределений процесса выходов за высокий уровень, вектора количеств выходов за высокий уровень в последовательности стационарно связанных случайных величин, ряда других задач теории экстремальных значений.

3. Получены оценки точности сложно-пуассоновской аппроксимации для распределения вектора количеств выходов за высокий уровень, получены утверждения типа ЗПЛ и оценки точности пуассоновской аппроксимации распределения числа длинных "повторов" в случайных последовательностях, оценки точности пуассоновской аппроксимации в ряде других задач теории экстремальных значений, получено обобщение на многомерный случай теоремы Бредли о задании независимой копии случайного вектора на одном вероятностном пространстве.

4. Предложен новый подход к оцениванию точности нормальной аппроксимации для распределений стьюдентизованных сумм независимых и стационарно связанных случайных величин, получены оценки с явными константами точности нормальной аппроксимации для распределений СНС, доказана невозможность неравномерной оценки типа Берри–Эссеена для статистики Стьюдента в общем случае.

5. Статистическое оценивание характеристик распределений с тяжёлыми хвостами по выборке стационарно связанных случайных величин: предложены новые оценки показателя скорости убывания хвоста распределения, экстремальной квантили и вероятности выхода за высокий уровень для распределений с тяжёлыми хвостами, доказана их состоятельность и асимптотическая нормальность в условиях слабой зависимости при минимальных ограничениях на коэффициенты перемешивания, предложен алгоритм выбора управляющего параметра.

6. Получены нижние границы точности оценивания характеристик распределений с тяжёлыми хвостами, выявлены соответствующие информационные функции.

Методы исследования. В работе применяются методы теории вероятностей и математической статистики. Кроме того, используется ряд конструкций, предложенных автором.

Теоретическая и практическая значимость работы. Работа носит преимущественно теоретический характер. Её результаты являются вкладом в развитие теории экстремальных значений и ряда других разделов теории вероятностей и математической статистики. Разработанные методики могут быть использованы за пределами круга задач теории экстремальных значений. Практическая ценность работы определяется созданием методики оценивания характеристик распределений с тяжёлыми хвостами по выборкам стационарно связанных случайных величин. Предложены новые оценки показателя скорости убывания хвоста распределения, экстремальной квантили и вероятности выхода за высокий уровень, а также алгоритм выбора управляющего параметра. Указанные задачи имеют приложения к проблеме оценивания финансовых рисков. Результаты диссертации могут найти применение в исследованиях, проводимых в МГУ им. М.В. Ломоносова, МИ РАН им. В.А. Стеклова, ПОМИ РАН им. В.А. Стеклова, СПбГУ, Новосибирском государственном университете, ИМ СО РАН, МГТУ им. Н.Э. Баумана, ДВНЦ РАН.

Апробация работы. Основные результаты диссертации докладывались на многих международных конференциях и семинарах: международная конференция "Measuring Risk in Complex Stochastic Systems", Берлин, 1999; международный научный семинар, ETH, Цюрих, 2001; международный научный семинар, Технический Университет, Эйндховен, 2001; международная конференция "Recent advances in Probability and Statistics", Лондон, Brunel University, 2002; международный научный семинар по теории вероятностей и математической статистике Оксфордского университета, Оксфорд, 2003; международный научный семинар по теории вероятностей и математической статистике, Royal Holloway University, Лондон, 2003; 11-й симпозиум международного финансового общества, Стамбул, 2004; международная конференция "Financial Stochastics", Лондон, Brunel University, 2005; международная конференция “Modern stochastics: theory and applications”, Киевский национальный университет, Киев, 2006; международная конференция “Recent advances in Probability, Statistics and Financial Stochastics”, Middlesex University, Лондон, 2007; семинар кафедры теории вероятностей и математической статистики, университет Мельбурна, Мельбурн, 2007; международный научный семинар по теории вероятностей и математической статистике, Leeds University, 2007; международная конференция “Combinatorial and probabilistic inequalities”, Институт Ньютона Кембриджского Университета, Кембридж, 2008; международный статистический симпозиум, Университет Джорджии, США, 2009; международный научный семинар кафедры теории вероятностей и математической статистики, Университет Мельбурна, 2010; 6$^{\text{th}}$ Finance Conference, Португалия, 2010; научный семинар кафедры теории вероятностей и математической статистики, Киевский национальный университет, Киев, 2011; научный семинар отдела случайных процессов, Институт Математики, Киев, 2011; 5-я международная конференция “Предельные теоремы теории вероятностей и их приложения”, Новосибирск, 2011; научный семинар ИПУ, Москва, 2011; научный семинар по теории вероятностей и математической статистике Санкт-Петербургского отделения Математического института РАН под руководством академика РАН И.А. Ибрагимова, Санкт-Петербург, 2011; международная конференция CFE-11, Лондон, 2011; международная конференция “Теория вероятностей и её приложения”, Москва, 2012; международная конференция “Statistics and Probability IMS-SWUFE”, Ченду, Китай, 2013.

Публикации. Список работ по теме диссертации приведен в конце реферата. Основные работы, в которых отражены результаты диссертации: [1]–[23].

Личный вклад автора. Все основные результаты, выносимые на защиту, принадлежат соискателю. Из совместных публикаций в диссертацию включен лишь материал, который был получен непосредственно соискателем. Все работы, за исключением [22], выполнены без соавторов. Вкладом автора в работу [22] является теорема 3.8.

Структура и объём диссертации. Диссертация состоит из введения, шести глав и списка литературы из 417 наименований. Общий объем диссертации – 232 страницы.

Содержание работы

Во **введении** обсуждаются методы и задачи исследования, даётся обзор истории вопроса, описывается структура и содержание работы, вводятся основные обозначения.

В **главе 1** описывается предложенный автором подход к задаче получения оценок скорости сходимости в предельных теоремах теории экстремальных значений (метод рекуррентных неравенств), а также рассматривается ряд задач теории экстремальных значений, связанных с выборками случайного объёма.

Пусть $X, X_1, X_2, \ldots$ – стационарная последовательность случайных величин,

$$M_n = \max_{1 \le i \le n} X_i \,, \; p \equiv p(u) := \mathbb{P}(X > u)\,.$$

Метод рекуррентных неравенств позволяет получать для вероятности $\mathbb{P}(M_n \le u)$ оценки сверху и снизу (теорема 1.1), с помощью которых выводятся оценки точности аппроксимации

$$\mathbb{P}(M_n \le u) \approx e^{-nb}$$

в предельных теоремах для M_n, где $b \equiv b(r,u) = \mathbb{P}(X_r > u, M_{r-1} \le u)$. В частности, теорема 1.2 устанавливает оценку скорости сходимости вида $O(n^{-1}+p)$ в случае m-зависимых с.в., что точнее оценки $O(n^{-1/2}+n^{1/2}p)$, полученной по методу блоков.

Важную роль в описании асимптотики распределения выборочного максимума последовательности стационарно связанных случайных величин играет понятие экстремального индекса. Необходимые и достаточные условия существования экстремального индекса устанавлены в теореме 1.5.

Рассмотрим строго стационарную последовательность $X, X_1, X_2, \ldots$ случайных величин. Пусть $(K^*; K^*) = \operatorname{supp}\mathcal{L}(X)$ и $u \equiv u_n < K^*$. Обозначим $M_{m,n} = \max_{m<k\le n} X_k$,

$$\theta^{R} \;\equiv\; \theta^{R}(r,u) = \mathbb{P}(M_{1,r} \le u | X_1 > u)\,,\; \theta^{B} \equiv \theta^{B}(r,u) = \mathbb{P}(M_r > u)/r\mathbb{P}(X > u)\,,$$

где $r = r_n \in \{1,\ldots,n\}$.

Теорема 1.5 *Пусть последовательность* $\{u_n\}$ *удовлетворяет соотношению*

$$0 < \liminf_{n\to\infty} n\mathbb{P}(X > u_n) \le \limsup_{n\to\infty} n\mathbb{P}(X > u_n) < \infty$$

и условию перемешивания $(D\{u_n\})$. *Последовательность* $\{X_i, i \ge 1\}$ *имеет экстремальный индекс* θ, *если и только если*

$$\lim_{r\to\infty}\limsup_{n\to\infty} |\theta - \theta^{R}(r,u_n)| = 0\,.$$

Последнее соотношение равносильно следующему:

$$\lim_{r\to\infty}\limsup_{n\to\infty} |\theta - \theta^{B}(r,u_n)| = 0\,.$$

Далее метод рекуррентных неравенств применяется при изучении ряда специальных задач, привлёкших внимание исследователей. Теоремы 2.1–2.11 посвящены задаче о распределении максимума частичных сумм (МЧС) случайных величин. Эту задачу изучали многие известные авторы, в т.ч. Эрдёш и Реньи[1], Деовельс, Эрдёш, Грил, Ревес[2], Деовельс, Девройс, Линч[3], Чаки, Фёлдеш, Комлош[4], Фролов, Мартикайнен, Штай-

[1]Erdös P., Rényi A. (1970) On a new law of large numbers. — J. Anal. Math., v. 22, 103–111.

[2]Deheuvels P., Erdös P., Grill K. and Révész P. (1987) Many heads in a short block. — Math. Statist. Probab. Theory (M.L.Puri et al., eds.), v. A, 53–67.

[3]Deheuvels P., Devroye L. and Lynch J. (1986) Exact convergence rates in the limit theorems of Erdös-Rényi and Shepp. — Ann. Probab., v. 14, № 1, 209–223.

[4]Csaki E., Földes A., Komlos J. (1987) Limit theorems for Erdös-Rényi type problems. — Studia Sci. Math. Hung., v. 22, 321–332.

небах[5], Питербарг[6], Штайнебах[7]. В работе найдено предельное распределение МЧС в спектре ситуаций $\ln n \ll k = k(n) \ll n$, получены оценки скорости сходимости в предельной теореме для максимума частичных сумм, а также результат типа ЗПЛ.

Пусть $X, X_1, X_2, \ldots$ – последовательность независимых одинаково распределённых случайных величин,

$$\zeta_m \equiv \zeta_m(k) = X_{m+1} + \ldots + X_{m+k}\,,$$

$(m \ge 0, k \ge 1)$. Обозначим

$$R_n^* \equiv R_n^*(k) = \max_{0 \le i \le n} \zeta_i\,.$$

Случайная величина $R_n^*(k)$ есть максимум частичных сумм Эрдёша–Реньи.

В основе результатов о МЧС лежит теорема 1.6. Она устанавливает оценки $\mathbb{P}(R_n^* < x)$ сверху и снизу.

Теорема 1.6 получена методом рекуррентных неравенств. Этот метод оказался эффективным средством получения оценок скорости сходимости в ряде предельных теорем теории экстремальных значений. Так, с помощью этого метода получена неулучшаемая по порядку оценка скорости сходимости в предельной теореме для распределения длины наибольшей серии “успехов”, а также впервые оценена скорость сходимости в предельной теореме для МЧС (теорема 1.14).

Теорема 1.13 устанавливает предельное распределение МЧС в спектре ситуаций $\ln n \ll k = k(n) \ll n$. Обозначим

$$\begin{aligned} u_n &= \sqrt{2\ln\Big(nk^{-1}\sqrt{2\ln(nk^{-1})}\,\Big)}\ , \ \alpha_n = u_n/\sqrt{k}\ , \\ \psi(y,\alpha) &= -y - y^2/2 + (1+y)^3\alpha G((1+y)\alpha)\,, \end{aligned}$$

где $G(x)$ есть ряд Крамера. В теореме 1.13 рассматриваются следующие возможные ситуации: при $n \to \infty$,

$$\begin{aligned} u_n^2\alpha_n^{m-1} &\to \infty\,,\ u_n^2\alpha_n^m \to 0 \qquad && (A_m), \\ u_n^2\alpha_n^m &\to \text{const} > 0 && (A_m^*), \end{aligned}$$

где $m \ge 1$. Отметим, что

$$(\ln(n/k))^{1+2/m} \ll k \ll (\ln(n/k))^{1+2/(m-1)}\ ,$$

если (A_m) выполнено;

$$(\ln(n/k))^{2+m}\,k^{-m} \to \text{const} \qquad (n \to \infty),$$

[5]Frolov A.N., Martikainen A. and Steinebach J. (2001) On the maximal excursion over increasing runs. — In: Asymptotic methods in probability and statistics with applications (St.-Petersburg, 1998), 225–242. Stat. Ind. Technol., Birkhaüser: Boston, MA.

[6]Питербарг В.И. (1991) О больших скачках случайного блуждания. — Теория вероятн. и её примен., т. 36, № 1, 54–64.

[7]Steinebach J. (1998) On a conjecture of Révész and its analogue for renewal processes. — In: Asymp. Methods Probab. Statist. (B.Szyszkowicz, ed.), 311–322.

если имеет место (A_m^*).

Приведём следующие следствие теоремы 2.11: если условие (A_2^*) выполнено, то

$$\mathbb{P}\left(R_n^*(k)/\sqrt{k} - u_n(1+c_1\alpha_n+c_2\alpha_n^2) < z/u_n\right) \to \exp\left(-e^{-z}/\sqrt{2\pi}\right) \qquad (n\to\infty),$$

где $c_1 = G(0)$, $c_2 = G'(0) + 2.5G^2(0)$.

Далее в главе 1 рассмотрен ряд задач, в которых объём выборки является случайным. Обозначим

$$M_t = \max\{t - S_{\mu(t)}; \max_{1\le i\le\mu(t)} X_i\},$$

где $S_0 = 0$, $S_n = X_1 + \ldots + X_n$ $(n \ge 1)$, $\mu(t) = \max\{n \ge 0 : S_n \le t\}$.

Задача имеет финансовые приложения. К примеру, банк открыл кредитную линию на сумму в t единиц, выбираемую траншами $X_1, X_2, \ldots$. Сведения об асимптотике распределения M_t позволяют ответить на вопрос о величине наибольшей из выплат.

В теоремах 1.17–1.19 получены оценки скорости сходимости и асимптотические разложения в предельной теореме для M_t. Лемма 1.20 устанавливает экспоненциальное неравенство для $|\mu(t) - \mathbb{E}\mu(t)|$.

Теоремы 1.21–1.23 посвящены распределению числа

$$N_t(x) = \sum_{j=1}^{\mu(t)} \mathbb{I}\{X_j \ge x\} + \mathbb{I}\{t - S_{\mu(t)} \ge x\}$$

выходов за высокий уровень до момента $\mu(t)$. Отметим, что $\{X_{k,t} < x\} = \{N_t(x) < k\}$, где $X_{k,t}$ является k-м по порядку неубывания элементом $\{X_1, \ldots, X_{\mu(t)}, t - S_{\mu(t)}\}$. В теоремах 1.21–1.23 получены оценки скорости сходимости и асимптотические разложения в предельной теореме для $N_t(x)$.

Пусть $X, X_1, \ldots, X_m, Y_1, \ldots, Y_n$ – независимые одинаково распределенные случайные величины с дискретным множеством состояний.

Будем говорить, что в последовательности $X_1, \ldots, X_n$ имеет место “повтор” длины $\ge k$, если $X_{i+1} = X_{j+1}, \ldots, X_{i+k} = X_{j+k}$ $(\exists i, j \le n-k)$. Будем говорить, что в последовательностях $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ имеет место “общий фрагмент” длины $\ge k$, если $X_{i+1} = Y_{j+1}, \ldots, X_{i+k} = Y_{j+k}$ $(\exists i \le m-k, j \le n-k)$.

Обозначим через N_n^* число повторов длины $\ge k$ в последовательности $X_1, \ldots, X_n$ и через $W_{m,n}$ число общих фрагментов длины $\ge k$ среди $X_1, \ldots, X_m, Y_1, \ldots, Y_n$. Пусть M_n^* – длина наибольшего повтора в последовательности $X_1, \ldots, X_n$, и пусть $M_{m,n}$ – длина наибольшего общего фрагмента в последовательностях $X_1, \ldots, X_m$, $Y_1, \ldots, Y_n$. Поскольку количество повторов длины $\ge k$ случайно, это задача о числе выходов за высокий уровень в выборке случайного объёма.

Задача изучения распределения длины наибольшего повтора и распределения числа длинных общих фрагментов в дискретных последовательностях имеет приложения в биологии при анализе “значимых” фрагментов последовательностей ДНК. Многие известные авторы работали над указанной задачей, в т.ч. Зубков и Михайлов[8], Михайлов[9], Аррати я[10], Гордон[11], Ватерман[12], Карлин[13], Ост[14] .

Зубков и Михайлов[8] доказали теорему Пуассона для N_n^*. Аррати я и др.[15] нашли предельное распределение $M_{m,n}$:

$$\Delta_{m,n} = \max_k |\mathbb{P}(M_{m,n} < k) - \exp\left(-mn(1-p)p^k\right)| \to 0$$

если $m \to \infty, n \to \infty, m \sim n$, где $p = \mathbb{P}(X = Y)$. Аррати я и Ватерман[12] получили результат типа ЗБЧ для M_n. Аррати я и др.[11] показали, что если $1 - c_+ < (\ln m)/\ln mn < c_+$, где

$$c_+ = \log_p q - 1\,,\ p = \mathbb{P}(X = Y)\,,\ q = \mathbb{P}(X_1 = X_2 = X_3),$$

то существует $\varepsilon > 0$ такое, что $\Delta_{m,n} = O(n^{-\varepsilon})$ при $m \to \infty, n \to \infty$.

В теоремах 1.28–1.35 установлены оценки скорости сходимости в предельных теоремах для числа длинных “повторов”, получены утверждения типа ЗПЛ. Указанные результаты уточняют соответствующие утверждения Зубкова, Михайлова, Аррати и, Гольдштейна, Гордона, Ватермана.

Обозначим $m' = m - k + 1$, $n' = n - k + 1$, $\lambda = (m'-1)(n'-1)(1-p)p^k + (m'+n'-1)p^k$,

$$p_* = \max_{a \in A} p_a\,,\ c_* = \log(1/p_*)\,,\ \Delta_{m,n}(k) = |\mathbb{P}(M_{m,n} < k) - \exp(-\lambda)|\,,$$

$\Delta^*(n,k)$ есть аналог $\Delta_{m,n}(k)$ с заменой $M_{m,n}$ на M_n^*, символ d_{TV} означает расстояние по вариации, $\mathcal{L}(\pi_{m,n}) = \mathbf{\Pi}(\lambda)$. Отметим, что $p_*^2 < p$, $p^2 \le q \le p_*$, $1 \ge c_+ \ge c_* > 1/2$; если распределение $\mathcal{L}(X)$ является равномерным, то $c_+ = c_* = 1$.

Теорема 1.25 *Если* $n \ge k$ *и* $m \ge k \ge 1$, *то*

$$d_{TV}(W_{m,n}; \pi_{m,n}) \le \lambda^{-1}\left(1 - e^{-\lambda}\right) m'n'(2k+1)\left(2kq_{2k} + (m'+n'-1)(p^{2k} + q^k)\right)\,.$$

Следствие 1.26 *Если* $m \to \infty$, $n \to \infty$, $(\ln mn)/(\min\{m,n\}) \to 0$, *то*

$$\max_{1 \le k \le m \wedge n} \Delta_{m,n}(k) = O\Big((m+n)(mn)^{-c_+}(\ln mn)^{1+c_+} + (mn)^{1-2c_*}(\ln mn)^{1+2c_*}\Big)\,.$$

[8] Зубков А.М., Михайлов В.Г. (1979) О повторениях s-цепочек в последовательности независимых величин. — Теория вероятн. и её примен., т. 24, № 2, 267–279.

[9] Михайлов В.Г. (2001) Оценка точности сложной пуассоновской аппроксимации для распределения числа совпадающих цепочек. — Теория вероятн. и её примен., т. 46, № 4, 713–723.

[10] Arratia R., Goldstein L. and Gordon L. (1989) Two moments suffice for Poisson approximation. — Ann. Probab., v. 17, № 1, 9–25.

[11] Arratia R., Gordon L. and Waterman M.S. (1990) The Erdös–Rényi law in distribution, for coin tossing and sequence matching. — Ann. Statist., v. 18, № 2, 539–570.

[12] Arratia R. and Waterman M.S. (1989) The Erdös–Rényi strong law for pattern matching with a given proportion of mismatches. — Ann. Probab., v. 17, № 4, 1152–1169.

[13] Karlin S. and Ost F. (1987) Counts of long aligned word matches among random letter sequences. — Adv. Appl. Probab., v. 19, № 2, 293–351.

[14] Karlin S. and Ost F. (1988) Maximal length of common words among random letter sequences. — Ann. Probab., v. 16, № 3, 535–563.

[15] Arratia R., Gordon L. and Waterman M.S. (1986) An extreme value theory for sequence matching. — Ann. Statist., v. 14, № 3, 971–993.

Теорема 1.27 *Если* $n > 3k \geq 3$, *то*

$$d_{TV}(N_n^*; \pi_{n,k}^*) \leq \lambda^{*-1}\left(1-e^{-\lambda^*}\right)\left((n^*)^3(2k+1)\left(p^{2k}+q^k\right) + 2(kn^*)^2 q_{2k}\right) + 2kn^*p^k\,,$$

где $\lambda^* \equiv \lambda_{n,k}^* = (n-3k+1)\,p^k(1+(n-3k)(1-p)/2)$, $\mathcal{L}(\pi_{n,k}^*) = \mathbf{\Pi}(\lambda^*)$, $n^* = n-k$.

Следствие 1.28 *При* $n \to \infty$,

$$\max_{1\leq k<n/3} \Delta^*(n,k) = O\Big(n^{1-2c_+}(\ln n)^{1+c_+} + n^{2-4c_*}(\ln n)^{1+2c_*}\Big)\,.$$

Пусть $\log$ означает логарифм по основанию $1/p$,

$$f_n = \log n^2 - \log\ln\ln n + \log((1-p)/2)\,,\ g_n = \log n^2 + \log\log n\,.$$

Теорема 1.30 *С вероятностью 1*

$$\liminf_{n\to\infty}(M_n^* - f_n) = -1\,,$$
$$\limsup_{n\to\infty}(M_n^* - g_n)\,/\log\ln\ln n = 1\,.$$

Рассмотрим ситуацию, когда внутри "повтора" допускается не более r несовпадений, и пусть $N_n^*(r)$ – число таких повторов длины $\geq k$ в последовательности $X_1,\ldots,X_n$, $M_n^*(r)$ – длина наибольшего повтора в последовательности $X_1,\ldots,X_n$, $W_{m,n}(r)$ – число повторов длины $\geq k$ в последовательностях $X_1,\ldots,X_m$, $Y_1,\ldots,Y_n$, $M_{m,n}(r)$ – длина наибольшего "повтора" в последовательностях $X_1,\ldots,X_m$, $Y_1,\ldots,Y_n$. Аналоги теорем 1.25 – 1.30 получены также для $N_n^*(r)$, $M_n^*(r)$, $M_{m,n}(r)$ и $W_{m,n}(r)$.

Исходным объектом теории экстремальных значений является число выходов за высокий уровень

$$N_n \equiv N_n(x) = \sum_{i=1}^{n} \mathbb{1}\{X_i > x\},$$

где $X_1, X_2, \ldots, X_n$ – рассматриваемая выборка. Результаты о распределении порядковых статистик следуют из результатов о распределении $N_n(x)$: если выборка записана в порядке неубывания как $X_{1,n} \geq \ldots \geq X_{n,n}$, то

$$\{X_{m,n} \leq x\} = \{N_n(x) < m\}.$$

В **главе 2** изучается точность пуассоновской и сложно-пуассоновской аппроксимации распределения числа выходов за высокий уровень. История вопроса восходит к теореме Пуассона.

Если $X_1,\ldots,X_n$ – независимые одинаково распределенные случайные величины, то число выходов за высокий уровень (число экстремальных значений) N_n имеет Биномиальное распределение $\mathbf{B}(n,p)$, где

$$p = \mathbb{P}(X > x)\,.$$

В задачах теории экстремальных значений p обычно мало, и $\mathcal{L}(N_n)$ допускает аппроксимацию распределением Пуассона $\mathbf{\Pi}(\lambda)$, где $\lambda = \mathbb{E}N_n$.

Задачу оценивания точности пуассоновской аппроксимации изучали многие известные авторы, в т.ч. Барбур, Деовельс, Лекам, Прохоров, Пфайфер, Серфлин, Холл.

Прохоров[16] установил асимптотику расстояния по вариации $d_{TV}(\mathbf{B}(n,p);\mathbf{\Pi}(np))$:

$$d_{TV}(\mathbf{B}(n,p);\mathbf{\Pi}(np)) = p/\sqrt{2\pi e}\Big(1 + O\big(p+1/\sqrt{np}\,\big)\Big). \tag{1}$$

Роос[17], опираясь на результаты Деовельса и Пфайфера[18], обобщил (1) на случай независимых неодинаково распределённых 0-1 случайных величин.

Пусть

$$N_n = \sum_{i=1}^{n} I_i\,,$$

где $\{I_i, i \geq 1\}$ – независимые случайные величины со значениями в $\{0,1\}$. Положим

$$p_i = \mathbb{E}I_i\,,\ \lambda = \sum_{i=1}^{n} p_i\,,\ \theta = \sum_{i=1}^{n} p_i^2/\lambda\,.$$

Обозначим через π_λ случайную величину с распределением Пуассона $\mathbf{\Pi}(\lambda)$. Барбуру и Иглсону[19] принадлежит оценка

$$d_{TV}(N_n;\pi_\lambda) \leq \lambda^{-1}(1-e^{-\lambda})\sum_{i=1}^{n} p_i^2\,. \tag{2}$$

Роос[17] (см. также Чяканавичюс и Роос[20]) получил оценку с неулучшаемой константой:

$$d_{TV}(N_n;\pi_\lambda) \leq 3\theta/4e\big(1-\sqrt{\theta}\,\big)^{3/2}\,. \tag{3}$$

Барбур[21] и Борисов и Ружанкин[22] получили асимптотические разложения для $\mathbb{E}h(N_n)$ для класса функций h.

Теорема 2.1 уточняет асимптотику второго порядка в оценке (3).

---

[16] Прохоров Ю.В. (1952) Асимптотическое поведение биномиального распределения. — УМН, т. 8, № 3(55), 135—142.

[17] Roos B. (1999) Asymptotic and sharp bounds in the Poisson approximation to the Poisson-binomial distribution. — Bernoulli, v. 5, № 6, 1021–1034.

[18] Deheuvels P. and Pfeifer D. (1986) A semigroup approach to Poisson approximation. — Ann. Probab., v. 14, № 2, 663–676.

[19] Barbour A.D. and Eagleson G.K. (1983) Poisson approximation for some statistics based on exchangeable trials. — Adv. Appl. Probab., v. 15, № 3, 585–600.

[20] Čekanavičius V. and Roos B. (2006) An expansion in the exponent for compound binomial approximations. Liet. Matem. Rink., v. 46, 67–110.

[21] Barbour A.D. (1987) Asymptotic expansions in the Poisson limit theorem. — Ann. Probab., v. 15, No 2, 748–766.

[22] Borisov I.S. and Ruzankin P.S. (2002) Poisson approximation for expectations of unbounded function of independent random variables. — Ann. Probab., v. 30, No 4, 1657–1680.

Пусть $[\cdot]$ означает целую часть, и положим

$$\begin{aligned} p_n^* &= \max_{i\le n} p_i\,, \quad \varepsilon=\min\Big\{1;(2\pi[\lambda-p_n^*])^{-1/2}+2\delta/(1-p_n^*/\lambda)\Big\}\,,\\ \delta &= \lambda^{-1}(1-e^{-\lambda})\sum_{i=1}^n p_i^2\,, \quad \delta^*=\lambda^{-1}(1-e^{-\lambda})\sum_{i=1}^n p_i^3\,. \end{aligned}$$

Теорема 2.1 *Справедливо неравенство*

$$d_{TV}(N_n;\pi_\lambda)\le 3\theta/4e+2\delta^*\varepsilon+2\delta^2\,. \tag{4}$$

Если $X,X_1,\ldots,X_n$ – одинаково распределенные случайные величины, то правая часть (4) имеет вид $3p/4e(1+O(p))$, тогда как правая часть (3) есть $3p/4e(1+O(\sqrt{p}))$.

В теореме 2.3 предложено новое доказательство результата Прохорова–Рооса об асимптотике расстояния по вариации $d_{TV}(N_n;\pi_\lambda)$, показано, что главный член асимптотики $d_{TV}(N_n;\pi_\lambda)$ может быть выражен в терминах специального преобразования распределения Пуассона.

Далее рассматривается более общая ситуация, когда $N_n=\sum_{i=1}^n\zeta_i$, где $\{\zeta_i\}$ – независимые случайные величины со значениями в $\mathbf{Z}_+=\{0,1,2,\ldots\}$. Барбур и Йенсен[23] оценили $|\mathbb{E}h(N_n)-\mathbb{E}h(\pi_\lambda)|$ для функций h, таких что $\|h\|_1=1$. В теореме 2.5 получены оценки точности пуассоновской аппроксимации $\mathcal{L}(\sum_{i=1}^n\zeta_i)$ в терминах расстояния по вариации d_{TV} и расстояния Джини–Канторовича–Васерштейна d_G. В теоремах 2.6 и 2.7 результат теоремы 2.5 обобщен на случай слабо зависимых случайных величин.

Теоремы 2.12 и 2.13 посвящены проблеме сложно-пуассоновской аппроксимации распределения числа выходов за высокий уровень в случае стационарно связанных случайных величин. Отметим, что данная задача является частным случаем проблемы аппроксимации распределений сумм случайных величин безгранично делимыми законами, рассматривавшейся многими известными авторами (см. Зайцев[24]).

Пусть $X,X_1,X_2,\ldots$ – строго стационарная последовательность случайных величин, $\{\pi,\zeta_{r,1},\ldots\}$ – независимые случайные величины, причём $\mathcal{L}(\pi)=\mathbf{\Pi}(kq)$,

$$\mathcal{L}(\zeta_{r,i})=\mathcal{L}(N_r|N_r>0)\quad(i\ge1)\,,\quad \zeta_{r,0}=0,$$

$$q=\mathbb{P}(N_r(u)>0)\,,\ k=[n/r]\,,\ r'=n-rk\,,$$

где $u=u_n$, и пусть $\tilde N_n=\sum_{i=0}^{\pi}\zeta_{r,i}$. Отметим, что случайная величина $\tilde N_n$ имеет сложно-пуассоновское распределение. Коэффициенты перемешивания $\alpha(l),\beta(l),\varphi(l)$ определены в Приложении.

Теорема 2.12 устанавливает, что при естественных предположениях слабая сходимость $\mathcal{L}(N_r|N_r>0)$, где $n\gg r=r_n\gg1$, является необходимым и достаточным

[23]Barbour A.D., Jensen J.L. (1989) Local and tail approximations near the Poisson limit. — Scand. J. Statist., v. 16, 75–87.

[24]Зайцев А.Ю. (2003) Об аппроксимации выборки пуассоновским точечным процессом. — Записки научных семинаров ПОМИ, т. 298, 111–125.

условием слабой сходимости $\mathcal{L}(N_n)$ к сложно-пуассоновскому закону. Оценка точности сложно-пуассоновской аппроксимации $\mathcal{L}(N_n)$ дана в теореме 2.13.

Теорема 2.13 *Если* $n > r > l \geq 0$, *то*

$$\begin{aligned} d_{TV}\big(N_n\,;\,\tilde{N}_n\big) &\leq c_{n,r}rp + (2kl + r')p + nr^{-1}\gamma_{n,l}\,, \\ d_{G}\big(N_n\,;\,\tilde{N}_n\big) &\leq rp\min\Big\{np\,;\tfrac{4}{3}\sqrt{2np/e}\Big\} + (2kl + r')p + n\gamma_{n,l}\,, \end{aligned} \tag{5}$$

где $c_{n,r} = \min\{1-e^{-np}\,;3/4e + (1-e^{-np})rp\}$ *и* $\gamma_{n,l} = \min\{4\alpha(l)\sqrt{r}\,;\beta(l)\}$.

Оценка (5) точна в следующем смысле: в случае независимых наблюдений ($r = 1, l = 0$) она совпадает с наилучшей оценкой точности пуассоновской аппроксимации.

Теоремы 2.16, 2.18 и 2.20 посвящены характеризация класса предельных распределений и оцениванию точности сложно-пуассоновской аппроксимации распределения вектора количеств выходов за высокие уровни в последовательности стационарно связанных случайных величин. Обозначим

$$\bar{N}_n = \bar{\xi}_1 + \ldots + \bar{\xi}_n\,,$$

где $\bar{\xi}_i = (\mathbb{I}\{X_i > x_1\}, \mathbb{I}\{x_1 \geq X_i > x_2\}, \ldots, \mathbb{I}\{x_{m-1} \geq X_i > x_m\})$.

В теореме 2.16 показано, что условие $C_{\bar{t}}$ является необходимым и достаточным для слабой сходимости $\bar{N}_n$ к сложно-пуассоновскому случайному вектору с независимыми компонентами (ввиду ограниченности места мы не приводим здесь условия $C_{\bar{t}}$, C, Δ, $\Delta\{u_n\}$). Утверждение 2.17 и теорема 2.18 устанавливают необходимые и достаточные условия слабой сходимости $\mathcal{L}(\bar{N}_n)$ в общем случае, когда условие $C_{\bar{t}}$ не предполагается выполненным. Теорема 2.20 оценивает точность аппроксимации распределения вектора $\mathcal{L}(\bar{N}_n)$ многомерным сложно-пуассоновским распределением.

Распределение вектора $\bar{N}_n$ аппроксимируется многомерным сложно-пуассоновским распределением $\mathbf{\Pi}(kq, \mathcal{L}(\bar{\zeta}))$, где $k = [n/r]$, $r \in \{1, \ldots, n\}$, $q = \mathbb{P}(N_r(u_m) > 0)$,

$$\mathcal{L}(\bar{\zeta}) = \mathcal{L}(\bar{N}_r | N_r(x_m) > 0)\,.$$

Положим $r' = n - rk$, и пусть $\mathcal{L}(\bar{Y}) = \mathbf{\Pi}(kq, \mathcal{L}(\bar{\zeta}))$.

Теорема 2.20. *Если* $n > r > l \geq 0$, *то*

$$d_{TV}\big(\bar{N}_n; \bar{Y}\big) \leq (1 - e^{-np})rp + (2nr^{-1}l + r')p + nr^{-1}\min\{\beta(l); \kappa(l)\}\,,$$

где $\kappa(l) = 2(1+2/m)\,(2^{m-1}m^2\alpha^2(l))^{1/(2+m)}$ *если* $m2^{(m-1)/2}\alpha(l) \leq 1$, *иначе* $\kappa(l) = 1$.

Слагаемое $(1 - e^{-np})rp$ здесь связано с применением оценки (2); вместо (2) можно использовать любую другую оценку точности пуассоновской аппроксимации биномиального распределения, например, (4).

При выводе оценок точности сложно-пуассоновской аппроксимации для распределения вектора количеств выходов за высокий уровень используется полученное в лемме 2.22 обобщение на многомерный случай теоремы Бредли о задании независимой копии случайного вектора на одном вероятностном пространстве.

Пусть (X, Y) – случайные векторы, принимающие значениями в $\mathbb{R}^l \times \mathbb{R}^m$, и пусть α – коэффициент равномерного перемешивания, соответствующий σ-алгебрам $\sigma(X)$ и $\sigma(Y)$. Обозначим $|v| = \max_{i \le m} |v_i|$ $(v \in \mathbb{R}^m)$.

Лемма 2.22 *Случайные векторы* X, Y *и* $\hat{Y}$ *могут быть заданы на одном вероятностном пространстве так, что* $\hat{Y}$ *не зависит от* X, $\hat{Y} \stackrel{d}{=} Y$ *и*

$$\mathbb{P}(|\hat{Y} - Y| > y) \le 2^{(m+3)/2} K^{m/2} \alpha + 2\mathbb{P}(|Y| > Ky) \qquad (y > 0, K \in \mathbb{N}).$$

Глава 3 посвящена точечным процессам выходов за высокий уровень (ПВВУ). Дана характеризации класса предельных распределений ПВВУ, оценена скорость сходимости в соответствующих предельных теоремах.

Если $\{X_i\}$ – строго стационарная последовательность случайных величин, то экстремальные значения, вообще говоря, появляются не по одиночке, а гроздьями (кластерами). Обозначим через $\mathcal{L}(\zeta)$ предельное распределение кластера. Тогда

$$\mathcal{L}(N_r \mid N_r > 0) \Rightarrow \mathcal{L}(\zeta) \tag{6}$$

для некоторой последовательности $\{r = r_n\}$ натуральных чисел, такой что

$$n \gg r_n \gg 1.$$

Если условие перемешивания $\Delta\{u_n\}$ выполнено, то существуют последовательности $\{r_n\}$ и $\{l_n\}$, такие что

$$n \gg r_n \gg l_n \gg 1\,,\ n r_n^{-1} \alpha_n^{2/3} \to 0\,, \tag{7}$$

где $\alpha_n = \alpha_n(l_n, u_n)$ – коэффициент α-перемешивания. Распределение $\mathcal{L}(\zeta)$ не зависит от выбора последовательности $\{r_n\}$, удовлетворяющей (7).

Определим точечный процесс $N_n(\cdot, u_n)$ равенством

$$N_n(B, u_n) = \sum_{i=1}^{n} \mathbb{I}\{\, i/n \in B, X_i > u_n \,\} \qquad (B \subset \mathcal{B}(0;1]). \tag{8}$$

Случайная мера $N_n(\cdot, u_n)$ учитывает местоположение экстремумов (выходов за уровень u_n), но не их размах.

Хсин, Хюслер и Лидбеттер[25] показали, что если (6) выполнено для последовательности $\{r_n\}$, удовлетворяющей соотношению (7), то

$$N_n(\cdot, u_n) \Rightarrow N(\cdot)\,,$$

[25] Hsing T., Hüsler J., Leadbetter M.R. (1988) On the exceedance point process for stationary sequence. — Probab. Theory Rel. Fields, v. 78, 97–112.

где $N(\cdot)$ – сложно-пуассоновский точечный процесс с распределением кластера $\mathcal{L}(\zeta)$ и интенсивностью λ, удовлетворяющей соотношению

$$\lim_{n\to\infty} \mathbb{P}(N_n = 0) = e^{-\lambda} \qquad (\exists\lambda \geq 0). \tag{9}$$

И наоборот, если $N_n(\cdot, u_n)$ слабо сходится к точечному процессу $N(\cdot)$, то $N(\cdot)$ есть сложно-пуассоновский процесс на $(0;1]$ с интенсивностью λ, удовлетворяющей соотношению (9). Если $\lambda > 0$, то (6) выполнено для некоторой случайной величины ζ и последовательности $\{r_n\}$, таких что имеет место (7).

В то время как процесс (8) подсчитывает местоположения выходов за один и тот же уровень u_n, процесс

$$\{N_n(u_n(t)), t \geq 0\} \tag{10}$$

выходов за высокий уровень имеет дело с размахом выбросов.

В теореме 3.2 найдены необходимые и достаточные условия слабой сходимости ПВВУ (10) к сложно-пуассоновскому процессу.

Пусть $\{u_n(\cdot), n \geq 1\}$ – последовательность функций на $[0;\infty)$, такая что функция $u_n(\cdot)$ возрастает для всех достаточно больших n, $u_n(0) = \infty$,

$$\limsup_{n\to\infty} n\mathbb{P}(X_n > u_n(t)) < \infty \qquad (0 < t < \infty),$$
$$\lim_{n\to\infty} \mathbb{P}(M_n \leq u_n(t)) = e^{-t} \qquad (t \geq 0).$$

Пусть $\{\pi(s), s \geq 0\}$ – пуассоновский процесс с интенсивностью 1, и пусть $\zeta, \zeta_1, \zeta_2, \ldots$ – последовательность н.о.р.с.в. со значениями в $\mathbb{N}$. Обозначим

$$Q(t) = \sum_{j=1}^{\pi(t)} \zeta_j\,. \tag{11}$$

Отметим, что $\{Q(t), t \geq 0\}$ – сложно-пуассоновский скачкообразный процесс (иными словами, $\tilde{Q}(B) := \int_B Q(dt)$ является сложно-пуассоновским точечным процессом с лебеговской мерой интенсивности и распределением кластера $\mathcal{L}(\zeta)$).

Теорема 3.2 устанавливает необходимые и достаточные условия слабой сходимости ПВВУ (10) к сложно-пуассоновскому процессу (11).

Теорема 3.2 *Предположим, что условие перемешивания Δ выполнено. Тогда*

$$N_n(u_n(\cdot)) \Rightarrow Q(\cdot) \qquad (n \to \infty),$$

если и только если условие C выполнено.

Процессы (8) и (10) – одномерные. Наиболее общим процессом, рассматриваемым в теории экстремальных значений, является процесс N_n^*, учитывающий как расположение экстремумов, так и их размах. Теорема 3.4 описывают необходимые и достаточные условия слабой сходимости эмпирического процесса выходов за высокий уровень N_n^* к сложно-пуассоновскому процессу.

Для всякого борелевского множества $A \subset (0;1] \times [0;\infty)$ обозначим

$$N_n^*(A) = \sum_{i=1}^{n} \mathbb{1}\{\, (i/n, u_n^{-1}(X_i)) \in A \,\} .$$

Отметим, что $N_{[sn]}(u_n(t)) = N_n^*((0;s] \times [0;t))$.

Пусть N^* – сложно-пуассоновский точечный процесс на $(0;1] \times [0;\infty)$ с лебеговой мерой интенсивности и распределением кластера $\mathcal{L}(\zeta)$.

Теорема 3.4 *Пусть выполнено условие перемешивания* Δ. *Тогда*

$$N_n^* \Rightarrow N^* ,$$

если и только если условие C *выполнено.*

Результаты о сходимости к сложно-пуассоновскому процессу играют роль ЦПТ в теории экстремальных значений. Однако класс предельных распределений указанных эмпирических процессов богаче класса сложно-пуассоновских процессов.

Далее рассматривается задача аппроксимации распределений эмпирических процессов выходов за высокий уровень в общем случае, когда предельное распределение, вообще говоря, не является сложно-пуассоновским.

Теоремы 3.6–3.8 посвящены одномерному процессу выходов за высокий уровень, учитывающему размах выбросов. Общий процесс N_n^*, учитывающий как расположение экстремумов, так и их размах, рассматривается в теореме 3.9 и следствии 3.10.

Положим $N_n(r,\bar t) = \big(N_r(u_n(t_1)), \ldots, N_r(u_n(t_m))\big)$, и пусть $\bar t = (t_1, \ldots, t_m)$, где $0 < t_1 < \ldots < t_m \le 1$. Обозначим через $\zeta(\bar t, n)$ случайный вектор с распределением

$$\mathcal{L}(\zeta(\bar t, n)) = \mathcal{L}\left(N_n(r,\bar t\,) \,|\, N_r(u_n(t_m)) > 0\right) .$$

Теорема 3.6 *Предположим, что существует скачкообразный процесс* $\{\gamma(t), t \in [0;1]\}$ *со стохастически непрерывными траекториями такой, что*

$$\zeta(\bar t, n) \Rightarrow (\gamma(t_1), \ldots, \gamma(t_m)) \qquad (12)$$

для любых $m \ge 1$ *и* $\bar t$, *где последовательность натуральных чисел* $\{r = r_n\}$ *удовлетворяет* (7). *Тогда*

$$\{N_{[sn]}(u_n(tT)), t \in [0;1]\} \Rightarrow \{N_{_T}(s,t), t \in [0;1]\} \qquad (13)$$

для любых $s > 0, T > 0$, *где* $N_{_T}(s,t) = \sum_{j=1}^{\pi_T(s)} \gamma_j(t)$, $\{\gamma_j(\cdot), 0 \le t \le 1\}$ *– независимые копии* $\gamma(\cdot)$, *пуассоновский процесс* $\pi_T(s)$ *не зависит от* $\{\gamma_j(\cdot)\}$.

Процесс $\{N_{_T}(s,t)\}$ обладает свойством

$$N_T(as, \cdot) \stackrel{d}{=} N_T(s, a\cdot) \qquad (\forall a \in [0;1]) .$$

Отметим, что соотношение (13) может быть представлено в виде

$$\{N_{[sn]}(u_n(t))\,,\,t\le T\}\Rightarrow\left\{\sum_{j=1}^{\pi_T(s)}\gamma_j(t/T)\,,\,t\le T\right\}. \qquad (13^*)$$

Положим

$$Z(t)\stackrel{d}{=}\zeta\xi(t) \qquad (0\le t\le 1)\,, \qquad (14)$$

где случайная величина $\xi(t)$ не зависит от ζ и имеет распределение Бернулли $\mathbf{B}(t)$.

Теорема 3.7 *Если* $\{N_n(u_n(t)),t\in[0;T]\}$ *слабо сходится к скачкообразному процессу, то тогда существует скачкообразный процесс* $\{\gamma(t),t\in[0;1]\}$ *со стохастически непрерывными траекториями, такой что* (12) *и* (13*) *выполнены. Маргинальные распределения процесса* γ *удовлетворяют соотношению*

$$\gamma(t)\stackrel{d}{=}Z(t) \qquad (0\le t\le 1)\,, \qquad (15)$$

где случайные величины $\{Z(t)\}$ *определены уравнением* (14).

Точность аппроксимации распределения процесса $\{N_n(u_n(\cdot))\}$ оценена в следующей теореме.

Теорема 3.8 *Для всякого* $n\ge r>l\ge 0$,

$$d_{TV}\Big(N_n(u_n(\cdot));\sum_{i=1}^{\pi}\gamma_i(\cdot/T)\Big)\le(1-e^{-np})rp+r'p+2npl/r+\beta(l)n/r\,. \qquad (16)$$

В доказательстве теоремы 3.8 используется оценка точности пуассоновской аппроксимации биномиального распределения, и слагаемое $(1-e^{-np})rp$ в (16) связано с применением оценки (2); вместо (2) можно использовать любую другую оценку точности пуассоновской аппроксимации биномиального распределения, например, (4).

Мори[26] и Хсин[27] охарактеризовали класс предельных распределений ПВВУ $N_n^*=N_{n,1}^\star$ в терминах двумерных точечных процессов. Теорема 3.9 и следствие 3.10 устанавливают, что класс предельных процессов для ПВВУ $N_{n,T}^\star$ есть класс процессов $N_{T,\gamma}^\star$, образованых одномерными точечными процессами.

Пусть $\{\gamma(t),t\in[0;1]\}$ – скачкообразный процесс. Для всякого борелевского множества $A\subset(0;1]\times[0;1)$ обозначим

$$N_{T,\gamma}^\star(A)=\int_A\sum_{\pi_T(x)<j\le\pi_T(x+dx)}(\gamma_j(y+dy)-\gamma_j(y))\,.$$

[26] Mori T. (1977) Limit distributions of two-dimensional point processes generated by strong–mixing sequences. — Yokohama Math. J., v. 25, 155–168.

[27] Hsing T. (1987) On the characterization of certain point processes. — Stochastic Processes Appl., v. 26, 297–316.

Точечный процесс $N^\star_{_T} \equiv N^\star_{_{T,\gamma}}$ имеет следующие свойства:
(P1) приращения $N^\star_{_T}$ вдоль горизонтальной оси независимы,
(P2) $N^\star_{_T}((a;b] \times B) \stackrel{d}{=} N^\star_{_T}((0;b-a] \times B)$ для любого борелевского множества $B \subset [0;1)$,
(P3) $\{N^\star_{_T}((0;a] \times [0;t)),\ t \in [0;1)\} \stackrel{d}{=} \{N_{_T}(a,t),\ t \in [0;1)\}$.

Теорема 3.9 *Предположим, что выполнено условие перемешивания* Δ. *Если существует скачкообразный процесс* $\{\gamma(t),\ t \in [0;1]\}$ *со стохастически непрерывными траекториями, такой что* (12) *выполнено, то* $N^\star_{n,T} \Rightarrow N^\star_{_{T,\gamma}}$.

Из теоремы 3.7 и теоремы 3.9 вытекает

Следствие 3.10 *Предположим, что выполнено условие* Δ. *Если* $N^\star_{n,T}$ *слабо сходится к точечному процессу* $N^\star$, *то существует скачкообразный процесс* $\{\gamma(t),\ t \in [0;1]\}$ *со стохастически непрерывными траекториями, такой что* (12) *выполнено и* $N^\star = N^\star_{_{T,\gamma}}$.

Глава 4 посвящена задачам статистического оценивания характеристик распределений с тяжёлыми хвостами. Данная тематика имеет приложения к проблеме оценивания финансовых рисков.

Распределение $\mathcal{L}(X)$ имеет тяжёлый правый хвост, если

$$\mathbb{P}(X > x) = L(x)x^{-1/a} \qquad (\alpha > 0),$$

где функция L медленно меняется: $\lim\limits_{x\to\infty} L(xt)/L(x) = 1$ $(\forall t > 0)$. Обозначим через $\mathcal{H}$ класс таких распределений.

Число $\alpha = 1/a$ является показателем скорости убывания хвоста распределения (ПСУХР). Если $L(x) = C + o(1)$, то C называется константой хвоста распределения.

Распределение с тяжёлым хвостом принадлежит к одному их трёх типов предельных распределений для выборочного максимума. Интерес к задаче оценивания характеристик распределения с тяжёлым хвостом связан также с тем, что в финансовых и страховых приложениях распределение наблюдений часто имеет тяжёлый хвост.

Класс распределений с тяжёлым правым хвостом является непараметрическим, что делает задачу оценивания характеристик распределения $\mathcal{L}(X)$ нетривиальной. Многие известные авторы работали над указанной задачей, в т.ч. Хилл[28], Холл[29], Голди и Смит[30], Смит[31], Декерс, Айнмаль, де Хаан[32], Хойслер и Тойгельс[33], Бейрлант, Гогебер,

[28]Hill B.M. (1975) A simple general approach to inference about the tail of a distribution. — Ann. Statist., v. 3, 1163–1174.

[29]Hall P. (1982) On estimating the endpoint of a distribution. — Ann. Statist., v. 10, № 2, 556–568.

[30]Goldie C.M. and Smith R.L. (1987) Slow variation with remainder: theory and applications. — Quart. J. Math. Oxford, v. 38, 45–71.

[31]Smith R.L. (1987) Estimating tails of probability distributions. — Ann. Statist., v. 15, № 3, 1174–1207.

[32]Dekkers A.L.M., Einmahl J.H.J. and de Haan L. (1989) A moment estimator for the index of an extreme–value distribution. — Ann. Statist., v. 17, № 4, 1833–1855.

[33]Haeusler E. and Teugels J.L. (1985) On asymptotic normality of Hill's estimator for the exponent of regular regulation. — Ann. Statist., v. 13, № 2, 743–756.

Тойгельс, Сегерс[34].

Голди и Смит[30] предложили RE-оценку

$$a_n^{RE} \equiv a_n^{RE}(x) = \sum_{i=1}^{n} \ln(X_i/x) \mathbb{I}\{X_i > x\} \Big/ \sum_{i=1}^{n} \mathbb{I}\{X_i > x\} \tag{17}$$

индекса a (соответственно, $1/a_n^{RE}$ является RE-оценкой показателя скорости убывания хвоста распределения α). Она относится к типу оценок, построенных по элементам выборки, выходящим за высокий неслучайный уровень. Другой известной оценкой является оценка Хилла[28]. Она относится к типу оценок, построенных по верхним порядковым статистикам. Приведены аргументы, свидетельствующие в пользу оценок первого типа.

Для оценивания показателем скорости убывания хвоста распределения предложена статистика $a_{n,m}$, являющаяся обобщением RE-оценки. Для RE-оценки и статистики $a_{n,m}$ найдены асимптотика средне-квадратичной ошибки и теоретически оптимальное значение управляющего параметра (в случае выборки из последовательности независимых одинаково распределённых случайных величин). Установлена состоятельность статистики $a_{n,m}$ и оценки константы хвоста распределения в случае строго стационарной последовательности случайных величин при условии перемешивания

$$\sum_{i \geq 1} i^{-1} \rho(i) < \infty. \tag{18}$$

Асимптотическая нормальность RE-оценки доказана при дополнительном условии

$$\lim_{l \to \infty} \varphi(l) = 0 \tag{19}$$

(теоремы 4.2 и 4.4). Построены асимптотические и (в случае независимых случайных величин) под-асимптотические доверительные интервалы. Предложена процедура выбора управляющего параметра указанных непараметрических оценок. Результаты моделирования выборок из распределений с тяжёлыми хвостами свидетельствуют в пользу предложенной процедуры.

Теоретически оптимальное значение управляющего параметра зависит от индексов второго порядка в асимптотическом разложении $\mathbb{P}(X > x)$. Ряд авторов рассматривали задачу оценивания параметров второго порядка по выборке из распределения вида

$$\mathbb{P}(X > x) = c x^{-1/a} \left(1 + d x^{-\beta} + o(x^{-\beta})\right) , \tag{20}$$

где $a > 0, b > 0, c > 0, d \neq 0$. Класс распределений (20) является непараметрическим. Предложены оценки индексов β и $w_x = d x^{-\beta}$, доказана их состоятельность (утверждения 4.5 и 4.6).

Задача оценивания экстремальных квантилей имеет важные приложения к финансам: с точностью до знака, экстремальная квантиль является мерой риска, известной

[34] Beirlant J., Goegebeur Y., Teugels J., Segers J. (2004) *Statistics of extremes.* — Chichester: Wiley.

как “Value-at-Risk” (VaR). Результаты моделирования выборок из распределений с тяжёлыми хвостами показывают, что эмпирическая оценка экстремальной квантили не точна. Другие имеющиеся в литературе оценки (см., к примеру, Макнил[35]) также оказались не точны.

Предложена оценка экстремальной квантили и оценка функции $M_n(x) = \mathbb{E}\{X - x|X > x\}$, используемой при построении другой популярной меры риска, известной как “Expected Shortfall”. Предложена процедура выбора управляющего параметра оценок, в теореме 4.7 доказана состоятельность оценок при условии (18). Асимптотическая нормальность оценок экстремальной квантили и функции $M_n(x)$ доказана при дополнительном условии (19) (теоремы 4.8 и 4.10). Результаты моделирования выборок из распределений с тяжёлыми хвостами свидетельствуют о предпочтительности предложенной оценки экстремальной квантили по сравнению с эмпирической оценкой.

Далее рассматривается задача оценивания вероятности $F_c(y) := \mathbb{P}(X > y)$ выхода за высокий уровень в последовательности стационарно связанных случайных величин. Предложена оценка вероятности $F_c(y)$, доказана её состоятельность при условии (18) (теорема 4.11) и асимптотическая нормальность при дополнительном условии (19) (теорема 4.12), предложена процедура выбора управляющего параметра.

Важным направлением в статистике экстремальных значений является задача определения нижних границ точности оценивания характеристик неизвестного распределения с тяжёлым хвостом. Этой тематике посвящены работы Бейрлант, Буко, Веркер[36], Холл и Вэлш[37], Донохо и Лю[38], Дреес[39], Пфанцаль[40] и др.. Однако указанная литература даёт лишь частичное решение задачи: найден порядок скорости убывания нижних границ[37,38,40], а также асимптотические нижние границы при тех или иных ограничениях на класс рассматриваемых оценок и распределений. Например, Бейрлант, Буко и Веркер[36] рассматривают класс оценок показателя скорости убывания хвоста распределения, являющихся $O_p(1)$ равномерно по классу рассматриваемых распределений, а возможные значения показателя скорости убывания хвоста распределения должны принадлежать интервалу фиксированной длины. Дреес[39] получил асимптотическую нижнюю границу для класса аффинных оценок показателя скорости убывания хвоста распределения, возможные значения показателя скорости убывания хвоста распределения должны принадлежать интервалу фиксированной длины.

В работе найдены нижние границы точности оценивания характеристик распределений с тяжёлыми хвостами, при этом не требуется, чтобы возможные значения показателя скорости убывания хвоста распределения принадлежали интервалу фиксиро-

[35]McNeil A.J. (1998) On extremes and crashes. — Risk, v. 11, p. 99–104.

[36]Beirlant J., Bouquiaux C., Werker B.J.M. (2006) Semiparametric lower bounds for tail index estimation. — J. Statist. Plann. Inference, v. 136, № 3, 705–729.

[37]Hall P. and Welsh A.H. (1984) Best attainable rates of convergence for estimates of parameters of regular variation. — Ann. Statist., v. 12, № 3, 1079–1084.

[38]Donoho D.L. and Liu R.C. (1991) Geometrizing rates of convergence II, III. — Ann. Statist., v. 19, 633–667, 668–701.

[39]Drees H. (2001) Minimax risk bounds in extreme values theory. — Ann. Statist., v. 29, № 1, 266–294.

[40]Pfanzagl J. (2000) On local uniformity for estimators and confidence limits. — J. Statist. Plann. Inference, v. 84, 27–53.

ванной длины; выявлены соответствующие информационные функционалы; получен аналог теоремы Пфанцаля[40] о невозможности оценивания с оптимальной точностью равномерно по шарам в метрике Хеллингера (теорема 4.15); найдены нижние границы точности оценивания функции распределения выборочного максимума (теоремы 4.19–4.21). Нижние границы точности оценивания экстремальных квантилей получены впервые.

Семейство $\mathcal{H}$ распределений с тяжёлыми хвостами весьма богато, и в литературе по статистике распределений с тяжёлыми хвостами принято[37] рассматривать непараметрический класс

$$\mathcal{F}(b) = \{F \in \mathcal{H} : \sup_{x>1} |c_F^{-1} x^{\alpha_F}(1 - F(x)) - 1| x^{b\alpha_F} < \infty\},$$

где b, c_F – положительные числа. Если функция распределения (ф.р.) $F \in \mathcal{F}(b)$, то

$$1 - F(x) = c_F x^{-\alpha_F}\left(1 + O(x^{-b\alpha_F})\right) \qquad (x \to \infty).$$

Теорема 4.16 устанавливает нижнюю границу точности оценивания ПСУХР. Используемый подход основан на построении двух "близких" ф.р. F_0 и F_1. Обозначим $\alpha_i = \alpha_{F_i}$, $a_i = 1/\alpha_i$, $\mathbb{E}_i$ есть матожидание, соответствующее F_i. Пусть

$$r = b/(1+2b), \; J_n = ((4/n)^b; \sqrt{n}/2], \; z_n = (8r/ne)^r e^{-r^2/n}/2.$$

Теорема 4.16 *Для любых $b > 0$ и $\alpha \in J_n$ существуют ф.р. $F_0, F_1 \in \mathcal{F}(b)$, такие что для любой оценки $\hat{\alpha}_n$ показателя скорости убывания хвоста распределения и любой оценки $\hat{a}_n$ индекса a*

$$\max_{i\in\{0;1\}} \alpha_i^{r/b} c_{F_i}^r \left(1-(4/n)^r \alpha_i^{-r/b}\right)^{-1} \mathbb{E}_i^{1/2}(\hat{\alpha}_n/\alpha_{F_i} - 1)^2 \geq z_n,$$
$$\max_{i\in\{0;1\}} a_i^{-r/b} c_{F_i}^r \left(1-(4/n)^r a_i^{r/b}\right)^{-1} \mathbb{E}_i^{1/2}(\hat{a}_n/a_{F_i} - 1)^2 \geq z_n. \qquad (21)$$

Если $\hat{a}_n$ в (21) заменить на RE-оценку, то левая часть (21) будет отличаться от правой части лишь на множитель $e^r/\sqrt{2r}$.

Аналогичные теореме 4.16 нижние границы установлены также для оценок константы хвоста распределения и оценок экстремальной квантили.

Значительный класс статистик, возникающих в теории экстремальных значений, имеет вид самонормированных сумм случайных величин. Таковыми являются, к примеру, ряд оценок показателя скорости убывания хвоста распределения и экстремального индекса. Конструкция предложенной оценки экстремальной квантили также включает СНС. К классу статистик, являющихся СНС, относятся также статистика Стьюдента и ряд непараметрических оценок плотности[41] и функции регрессии.

[41] Terrel G.R. and Scott D.W. (1980) On improving convergence rates for nonnegative kernel density estimators. — Ann. Statist., v. 8, № 5, p. 1160–1163.

В **главе 5** решена долго остававшаяся открытой задача получения оценок типа Берри–Эссеена с явными константами для самонормированных сумм (СНС) случайных величин.

Задача получения оценок скорости сходимости в ЦПТ восходит к Ляпунову[42]. Поскольку сумма случайных величин нормируется стандартным уклонением, которое в практических ситуациях обычно не известно и заменяется его оценкой, приходим к задаче о точности нормальной аппроксимации для статистики Стьюдента. Многие известные авторы работали над указанной задачей, в т.ч. Чун[43], Бенткус[44], Эфрон[45], Жине, Гётце, Мейсон[46], Гётце и Бенткус[47], Холл[48], Чистяков[49].

В случае независимых наблюдений автором впервые получены оценки *с явными константами* точности нормальной аппроксимации для распределений самонормированных сумм случайных величин, включая случай неодинаково распределённых случайных величин (теоремы 5.2 и 5.3). Оценки скорости сходимости в ЦПТ для самонормированных сумм получены также в случае стационарной последовательности зависимых случайных величин (теоремы 5.4 и 5.7). На основе указанных оценок предложены под-асимптотические доверительные интервалы для оценок показателя скорости убывания хвоста распределения. Получены оценки точности нормальной аппроксимации распределений квадратичных функционалов сумм независимых случайных векторов. Неравенство типа Берри–Эссеена с явными константами для распределения статистики Стьюдента получено впервые (теоремы 5.13 и 5.18). Показано, что константа в неравенстве типа Берри–Эссеена для статистики Стьюдента не может быть лучше, чем $1/\sqrt{2e}$ (утверждение 5.16). Установлено, что неравномерная оценка типа Берри–Эссеена для статистики Стьюдента не может имеет место без дополнительных предположений. Предложено новое доказательство характеризации нормального распределения в терминах самонормированных случайных величин. Свойство распределения быть симметричным охарактеризовано в терминах матожиданий самонормированных случайных величин.

Квадратичные функционалы от сумм независимых случайных векторов. Пусть (X, Y), $(X_1, Y_1), \ldots$ – независимые одинаково распределенные пары случайных величин, такие

[42] Liapunov A.M. (1901) Nouvelle forme du théorème sur la limite des probabilités. — Mem. Acad. Imp. Sci. St.-Petersburg, v. 12, 1–24.

[43] Chung K.-L. (1946) The approximate distribution of Student's statistic. — Ann. Math. Statistics, v. 17, 447–465.

[44] Bentkus V. (1994) On the asymptotical behavior of the constant in the Berry–Esseen inequality. — J. Theor. Probab., v. 7, № 2, 211–224.

[45] Efron B. (1969) Student's t–test under symmetry conditions. — J. Amer. Statist. Assoc., v. 64, 1278–1302.

[46] Giné E., Götze F. and Mason D.M. (1997) When is the Student t-statistic asymptotically standard normal? — Ann. Probab., v. 25, № 3, 1514–1531.

[47] Bentkus V. and Götze F. (1996) Berry–Esseen bound for Student's statistic. — Ann. Probab., v. 24, № 1, 491–503.

[48] Hall P. (1987) Edgeworth expansion for Student's t–statistic under minimal moment conditions. — Ann. Probab., v. 15, № 3, 920–931.

[49] Chistyakov G.P. (2001) A new asymptotic expansion and asymptotically best constants in Lyapunov's theorem. — Theory Probab. Appl., v. 46, 226–242, 516–522.

что $\mathbb{E}X = \mathbb{E}Y = 0$, $\mathbb{D}S_{n,X} = 1$, где

$$S_{n,X} = \sum_{i=1}^{n} X_i\,,\ S_{n,Y} = \sum_{i=1}^{n} Y_i\,,\ Z_n{=}S_{n,X} + cS_{n,Y}^2 \tag{22}$$

($c \in \mathbb{R}$). Обозначим

$$\begin{aligned} r_1 &= \sqrt{\pi/8}\,\left(n\mathbb{E}|X|^3/2 + |c|\mathbb{D}S_{n,Y}\right),\ r_2 = |c|n\left(\mathbb{E}|X|Y^2{+}2\mathbb{E}|X||Y|\mathbb{E}|S_{n,Y}|\right), \\ r_3 &= \left(\mathbb{E}|X| + n\mathbb{E}|X|^3/2 + |c|\mathbb{E}Y^2 + 2|c|\mathbb{E}|Y|\mathbb{E}^{1/3}|S_{n,X}|^3\mathbb{E}^{2/3}|S_{n,Y}|^{3/2}\right). \end{aligned}$$

Теорема 5.1 *Для любого* $x \in \mathbb{R}$

$$\Delta_n \equiv |\mathbb{P}\left(Z_n < x\right) - \Phi(x)| \le 9n\mathbb{E}|X|^3/\sqrt{2\pi}\, + 2(r_1 + r_2 + r_3)\,. \tag{23}$$

Само-нормированные суммы случайных величин. Пусть $(X,Y),(X_1,Y_1),\ldots$ – пары случайных величин, $Y_i \ge 0$. Один из разделов посвящен оцениванию точности нормальной аппроксимации для распределений статистик

$$S_n/T_n\,,\ S_n/\sqrt{T_n}\,,$$

где $S_n = \sum_{i=1}^n X_i$, $T_n = \sum_{i=1}^n Y_i$.

Теорема 5.2 устанавливает оценку типа Берри–Эссеена с явными константами для распределения статистики $Z_{n,1} = S_n/T_n$. В качестве приложения теоремы 5.2 получены оценки типа Берри–Эссеена в ЦПТ для ядерной оценки функции регрессии (следствия 5.8 и 5.9) и функции интенсивности отказов (следствия 5.10 и 5.11). В теореме 5.3 найдена асимптотика $\mathbb{E}S_n/T_n$ вплоть до слагаемого порядка $O(n^{-2})$ и асимптотика $\mathbb{E}(S_n/T_n)^2$ вплоть до слагаемого порядка $O(n^{-3})$.

Теоремы 5.4 и 5.7 устанавливают оценки скорости сходимости в ЦПТ для само-нормированных сумм стационарно связанных случайных величин. Для m-зависимых случайных величин при $\mathbb{E}|X|^3 + \mathbb{E}Y^{3/2} < \infty$ из результата теоремы 5.4 следует оценка порядка $O(n^{-1/2})$.

Поскольку суммы случайных величин в приложениях обычно нормируются оценкой стандартного уклонения, естественно встаёт вопрос о точности нормальной аппроксимации для распределения статистики $t_n^* = S_n/T_n^{1/2}$, которая тесно связана со статистикой Стьюдента t_n: если $Y = X^2$, то

$$\{t_n \ge x\} = \left\{\, t_n^* \ge x/\sqrt{1 + x^2/n}\,\right\}.$$

Теорема 5.13 устанавливает оценку с явными константами точности нормальной аппроксимации для распределения статистики t_n^*. Приведём следующее следствие теоремы 5.13.

Следствие 5.14 *Пусть* $(X,Y),(X_1,Y_1),\ldots$ *– независимые одинаково распределенные пары случайных величин, такие что* $Y > 0$,

$$\mathbb{E}X = 0\,,\ \mathbb{E}X^2 = 1\,,\ \mathbb{E}|X|^3 + \mathbb{E}Y^{3/2} < \infty\,.$$

Тогда равномерно по $x \geq 0$

$$-An^{-1/2}(1+o(1)) \leq \mathbb{P}\big(S_n/(T_n/\mathbb{E}Y)^{1/2} < x\big) - \Phi(x) \leq Bn^{-1/2}(1+o(1)) \qquad (24)$$

при $n \to \infty$, *где*

$$\begin{aligned} A &= \mathbb{E}|X|^3\left(1+9/\sqrt{2\pi}+\sqrt{\pi/8}\right)+2\mathbb{E}|X|+8|\mathbb{E}XY|\Big/e\sqrt{2\pi}\,\mathbb{E}Y\,, \\ B &= C_*\mathbb{E}|X|^3+4|\mathbb{E}XY|\Big/e\sqrt{2\pi}\,\mathbb{E}Y\,. \end{aligned}$$

В случае $Y = X^2$ из (24) следует соответствующий результат для t_n.

Утверждение 5.16 устанавливает, что константа в неравенстве Берри–Эссеена для t_n и t_n^* не может быть лучше, чем $1/\sqrt{2e}$.

Показано, что неравномерная оценка вида

$$\Big|\mathbb{P}(t_n^* < x) - \Phi(x)\Big| \leq Cg(x)\mathbb{E}|X|^3 n^{-1/2} \qquad (x \in \mathbb{R}), \qquad (25)$$

где C – абсолютная постоянная и $0 < g(\cdot) \downarrow 0$, вообще говоря, не имеет место. Теорема 5.18 предлагает неравномерную оценку скорости сходимости в ЦПТ для t_n^* при более жёстких моментных ограничениях. Константы в неравенствах теорем 5.13 и 5.18 точны в следующем смысле: показано, что для определённых распределений оценки теорем 5.13 и 5.18 для t_n отличаются от оценок в равномерном и неравномерном неравенствах Берри–Эссеена для $S_n/\sqrt{n}$ лишь на множитель $(1+o(1))$.

Заключительная глава содержит необходимые сведения по теории вероятностей и математической статистике, используемые в диссертации.

Публикации автора по теме диссертации

[1] Новак С.Ю. (1991) О распределении максимума случайного числа случайных величин. — Теория вероятн. и её примен., т. 36, № 4, 675–681.

[2] Novak S.Y. (1993) On the asymptotic distribution of the number of random variables exceeding a given level. — Siberian Adv. Math., v. 3, № 4, 108–122.

[3] Novak S.Y. (1994) Asymptotic expansions for the maximum of random number of random variables. — Stochastic Process. Appl., v. 51, № 2, 297–305.

[4] Новак С.Ю. (1994) Пуассонова аппроксимация числа длинных "повторов" в случайных последовательностях. — Теория вероятн. и её примен., т. 39, № 4, 731–742.

[5] Novak S.Y. (1995) Long match patterns in random sequences. — Siberian Adv. Math., v. 5, № 3, 128–140.

[6] Новак С.Ю. (1996) О распределении отношения сумм случайных величин. — Теория вероятн. и её примен., т. 41, № 3, 533–560.

[7] Novak S.Y. (1996) On extreme values in stationary sequences. — Siberian Adv. Math., v. 6, № 3, 68–80.

[8] Новак С.Ю. (1997) О распределении максимума частичных сумм Эрдёша–Реньи. — Теория вероятн. и её примен., т. 42, в. 2, 274–293.

[9] Новак С.Ю. (1997) Статистическое оценивание максимального собственного числа матрицы. — Известия ВУЗов (Матем.), т. 41, № 5, 49–52.
[10] Novak S.Y. (1998) On the limiting distribution of extremes. — Siberian Adv. Math., v. 8, № 2, 70–95.
[11] Novak S.Y. (2000) On self-normalized sums. — Math. Methods Statist., v. 9, № 4, 415–436.
[12] Novak S.Y. (2002) On self-normalized sums. Supplement. — Math. Methods Statist., v. 11, № 2, 256–258.
[13] Novak S.Y. (2002) Multilevel clustering of extremes. — Stochastic Process. Appl., v. 97, № 1, 59–75.
[14] Novak S.Y. (2002) Inference on heavy tails from dependent data. — Siberian Adv. Math., v. 12, № 2, 73–96.
[15] Novak S.Y. (2003) On the accuracy of multivariate compound Poisson approximation. — Statist. Probab. Lett., v. 62, № 1, 35–43.
[16] Новак С.Ю. (2004) О самонормированных суммах случайных величин и статистике Стьюдента. — Теория вероятн. и её примен., т. 49, № 2, 365–373.
[17] Novak S.Y. (2006) A new characterization of the normal law. — Statist. Probab. Letters, v. 77, № 1, 95–98.
[18] Novak S.Y. (2007) Measures of financial risks and market crashes. — Theory Stochast. Processes, v. 13, № 1, 182–193.
[19] Novak S.Y. (2010) Lower bounds to the accuracy of sample maximum estimation. — Theory Stochast. Processes, v. 15(31), № 2, 156–161.
[20] Novak S.Y. (2010) Impossibility of consistent estimation of the distribution function of a sample maximum. — Statistics, v. 44, № 1, 25–30.
[21] Novak S.Y. (2011) Extreme value methods with applications to finance. — London: Chapman & Hall/CRC Press. ISBN 9781439835746
[22] Novak S.Y. and Xia A. (2012) On exceedances of high levels. — Stochastic Process. Appl., v. 122, 582–599.
[23] Новак С.Ю. (2013) О точности оценивания характеристик распределений с тяжёлыми хвостами. — Теория вероятн. и её примен., т. 58, № 3, 1–13.