Московский государственный университет им. М. В. Ломоносова
Факультет вычислительной математики и кибернетики

На правах рукописи

ЛЯМИН Олег Олегович

НЕКОТОРЫЕ ЗАДАЧИ ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ, СВЯЗАННЫЕ С РАСПРЕДЕЛЕНИЕМ ЛАПЛАСА

01.01.05 — теория вероятностей и математическая статистика

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата физико–математических наук

Москва — 2010

Работа выполнена на кафедре математической статистики факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова.

Научный руководитель:	доктор физико-математических наук, профессор Бенинг Владимир Евгеньевич
Официальные оппоненты:	доктор физико-математических наук, профессор Хохлов Юрий Степанович
	доктор физико-математических наук, профессор Зейфман Александр Израилевич
Ведущая организация:	Институт проблем информатики РАН

Защита состоится «12» ноября 2010 г. в 11 часов на заседании диссертационного совета Д 501.001.44 при Московском государственном университете имени М. В. Ломоносова по адресу: 119991, ГСП-1, Москва, Ленинские горы, МГУ, 2-й учебный корпус, факультет ВМиК, аудитория 685.

С диссертацией можно ознакомиться в библиотеке факультета ВМиК МГУ. С текстом автореферата можно ознакомиться на официальном сайте факультета ВМиК МГУ `http://www.cmc.msu.ru` в разделе «Наука» — «Работа диссертационных советов» — «Д 501.001.44».

Автореферат разослан «12» октября 2010 г.

Ученый секретарь
диссертационного совета
профессор Н. П. Трифонов

Общая характеристика работы

Актуальность темы. Классическое распределение Лапласа с нулевым средним и дисперсией σ^2 было введено П. С. Лапласом в 1774 году. С тех пор оно стало одной из наиболее активно используемых симметричных вероятностных моделей. Это распределение задается плотностью

$$l(x) = \frac{1}{\sigma\sqrt{2}} \exp\left\{-\frac{\sqrt{2}|x|}{\sigma}\right\}, \quad \sigma > 0, \quad x \in \mathbb{R}.$$

Распределение Лапласа находит широкое применение при математическом моделировании многих процессов в телекоммуникационных системах, в экономике, финансовом деле, технике и других областях, например, в задачах выделения полезного сигнала на фоне помех. Популярность распределения Лапласа как математической (вероятностной) модели обусловлена тем, что его хвосты тяжелее, чем у нормального распределения (см., например, работу Р. Истерлинга[1], где обосновывается целесообразность использования распределения Лапласа как модели распределения погрешностей измерений в энергетике; статью Д. Хсу[2], посвященную применению распределения Лапласа для моделирования ошибок в навигации; работу Т. Окубо[3], в которой распределение Лапласа применяется в метеорологии). Во многих работах описано успешное применение распределения Лапласа для моделирования распределения приращений логарифмов финансовых индексов[4], для моделирования распределения логарифма размера частиц при дроблении[5], при моделировании статистических закономерностей поведения некоторых характеристик атмосферной[6] и плазменной[7] турбулентности. В работах Н. Джонсона[8] и С.

[1] Easterling R. J. Exponential responses with double exponential measurement error. A model for steam generator inspection // In: Proceedings of DOE Statistics Symposium. — U.S., Department of Energy. — 1978. — P. 90–100.

[2] Hsu D. A. Long-tailed distribution for position errors in navigation // Applied Statistics. — 1979. — Vol. 28. — P. 62–72.

[3] Okubo T., Narita N. On the distribution of extreme winds expected in Japan // In: National Bureau of Standards Special Publication 560-1. — 1980. — P. 12.

[4] Kozubowski T. J., Podgorski K. Asymmetric Laplace laws and modeling financial data // Math. Comput. Modelling. — 2001. — Vol. 34. — P. 1003–1021.

[5] Bagnold R. A. The physics of blown sand desert dunes. — London, Methuen. — 1954.

[6] Barndorff-Nielsen O. E. Models for non-Gaussian variation, with applications to turbulence // Proc. Royal Soc. A. — 1979. — Vol. 353. — P. 401–419.

[7] Королёв В. Ю. Вероятностно-статистический анализ хаотических процессов с помощью смешанных Гауссовских моделей. Декомпозиция волатильности финансовых индексов и турбулентной плазмы. — М.: Изд-во ИПИРАН. — 2007.

[8] Johnson N. L., Kotz S., Balakrishnan N. Continuous uni-variate distributions. Vol. II. 2nd ed.. — N. Y.: Wiley. — 1995.

Котца[9] можно найти дальнейшие ссылки на работы, в которых описывается применение распределения Лапласа к решению прикладных задач в самых разнообразных областях. Привлекательность распределения Лапласа в качестве вероятностной модели при решении конкретных прикладных задач во многом обусловливается также его экстремальными энтропийными свойствами. Этим свойством часто мотивируется выбор распределения Лапласа в качестве распределения погрешностей измерений, в которых точность (параметр масштаба) изменяется от измерения к измерению случайным образом (см., например, работу Г. Л. Шевлякова[10]). Последний результат стоит отметить особо.

Можно признать, что в подавляющем большинстве ситуаций, связанных с анализом экспериментальных данных, число случайных факторов, влияющих на наблюдаемые величины, само является случайным и изменяется от наблюдения к наблюдению. Примеры прикладных статистических задач, в которых объем выборки существенно случаен, можно найти, например, в книгах В. Ю. Королёва[11] и В. Е. Бенинга[12]. Поэтому вместо различных версий центральной предельной теоремы, обосновывающих нормальность распределения наблюдаемых случайных величин в классической статистике, в таких ситуациях следует опираться на их аналоги для выборок случайного объема. Здесь следует отметить недавние результаты[13] В. Е. Бенинга и В. Ю. Королёва, в которых была получена довольно простая асимптотическая схема, приводящая к распределению Лапласа как к предельному и, как следствие, дающая обоснование возможности более широкого использования распределения Лапласа в задачах описательной статистики.

Первая часть диссертации посвящена дальнейшему развитию идей этой работы, а именно получению оценок скорости сходимости распределений асимптотически нормальных статистик, построенных по выборкам случайного объема, к распределению Лапласа.

---

[9] Kotz S., Kozubowski T. J., Podgorski K. The Laplace distribution and generalizations: A revisit with applications to communications, economics, engineering and finance. — Boston: Birkhauser. — 2001.

[10] Shevlyakov G. L., Vilchevski N. O. Robustness in data analysis: Criteria and methods. — Utrecht: VSP. — 2002.

[11] Королёв В. Ю., Бенинг В. Е., Шоргин С. Я. Математические основы теории риска. — М.: Физматлит. — 2007.

[12] Королёв В. Ю., Бенинг В. Е., Соколов И. А., Шоргин С. Я. Рандомизированные модели и методы теории надежности информационных и технических систем. — М.: Торус Пресс. — 2007.

[13] Бенинг В. Е., Королёв В. Ю. Некоторые статистические задачи, связанные с распределением Лапласа // Информатика и ее Применения. — 2008. — Т. 2, №2. — С. 19–34.

Вторая часть диссертации связана с исследованиями Д. М. Чибисова[14,15] и В. Е. Бенинга[16] в области задачи проверки простой гипотезы против последовательности сложных близких альтернатив. В указанной работе В. Е. Бенинга была получена общая теорема, дающая достаточные условия для существования предела отклонения функции мощности асимптотически наиболее мощного (АНМ) критерия от функции мощности наилучшего критерия. В том типичном случае, когда соблюдены условия регулярности, можно ожидать (см. работы Д. М. Чибисова[14,15]), что мощность АНМ критерия отличается от мощности наилучшего критерия на величину порядка n^{-1}. Отсутствие регулярности может приводить к нарушению естественного порядка n^{-1} и приводить к другим порядкам. Факт нарушения обычных порядков в случае распределения Лапласа с параметром сдвига был отмечен в работе Р. А. Королёва[17]. Там же на эвристическом уровне была получена формула для предела отклонения мощностей. При этом в работе В. Е. Бенинга[18] прямым методом были получены асимптотические разложения для мощностей критериев, из которых непосредственно следует, что отсутствие регулярности распределения Лапласа приводит к порядку $n^{-1/2}$. Однако, как выяснилось позже, достаточные условия, сформулированные в общей теореме В. Е. Бенинга не выполнены, поэтому распределение Лапласа не может являться примером использования общей теоремы для случая нерегулярного распределения. Отсутствие такого примера может говорить в пользу того, что условия общей теоремы слишком сильны и выполняются только в регулярном случае, что существенным образом ограничивает множество ситуаций, в которых эта теорема может быть применима, и уменьшает ее прикладное значение. Таким образом, невыполнимость условий достаточности общей теоремы для случая распределения Лапласа оставляет актуальным вопрос поиска подходящего примера. Во второй части диссертационной работы далее исследуется возможность использования общей теоремы для нерегулярного распределения на примере случая обобщенного распределения Лапласа.

---

[14] Chibisov D. M. Asymptotic expansions and deficiencies of tests // In: Proc. Intern. Congr. Math., 2. — Warszawa. — 1983. — P. 1063–1079.

[15] Чибисов Д. М. Вычисление дефекта асимптотически эффективных критериев // Теор. вероятн. и ее прим. — 1985. — Т. 30, №2. — С. 269–288.

[16] Bening V. E. Asymptotic Theory of Testing Statistical Hypotheses. — Utrecht: VSP. — 2000.

[17] Королёв Р. А., Тестова А. В., Бенинг В. Е. О мощности асимптотически оптимального критерия в случае распределения Лапласа // Вестник Тверского Государственного Университета. — 2008. — Т. 28, №. 1. — С. 7–27.

[18] Бенинг В. Е., Королёв Р. А. Асимптотические разложения для мощностей критериев в случае распределения Лапласа // Вестник Тверского государственного университета, серия Прикладная математика. — 2008. — Т. 3(10), №26(86). — P. 97–107.

Цель работы. Цель первой части данной работы состоит в обосновании возможности использования распределения Лапласа в задачах теории вероятностей и математической статистики, возникающего в качестве предельного в случае выборок случайного объема. Задачами первой части диссертации являются:

1. Описание асимптотической схемы, приводящей к распределению Лапласа как к предельному.

2. Получение оценок скорости сходимости распределений асимптотически нормальных статистик, построенных по выборкам случайного объема специального вида, к распределению Лапласа.

Целью второй части диссертации является изучение возможности применения исследований В. Е. Бенинга, связанных с задачей проверки простой гипотезы против последовательности сложных близких альтернатив, в нерегулярных случаях. Задачами второй части диссертации являются:

1. Проверка условий общей теоремы из работы В. Е. Бенинга в случае нерегулярного распределения (случай обобщенного распределения Лапласа, предложенный в работе).

2. Демонстрация того, что отсутствие регулярности может приводить к нарушению естественного порядка разности функций мощности наилучшего и асимптотически наиболее мощного критериев.

Научная новизна. Все основные результаты работы новые и заключаются в следующем:

1. Получены новые оценки скорости сходимости распределения нормированного максимума от n случайных величин с дискретным распределением Парето к обратному показательному распределению с ростом n. Получены новые оценки скорости сходимости распределения асимптотически нормальных статистик к распределению Лапласа в случае, когда объем выборки случаен и равен указанному максимуму.

2. Предложено обобщенное распределение Лапласа, для которого рассмотрена задача проверки простой гипотезы против последовательности

сложных близких альтернатив. С применением общей теоремы показано, как отсутствие регулярности у этого распределения приводит к нарушению естественного порядка разности функций мощности наилучшего и асимптотически наиболее мощного критериев. Получена формула для предела отклонения мощности наилучшего критерия от мощности асимптотически наиболее мощного критерия. Обоснована возможность использования общей теоремы для случая нерегулярных распределений.

Методы исследования. В работе использованы аналитические методы математического анализа, неравенства и предельные теоремы теории вероятностей, аппарат математической статистики, а также метод характеристической функции.

Теоретическая и практическая значимость. Работа имеет теоретический характер. Результаты, относящиеся к оценке скорости сходимости распределений статистик к распределению Лапласа, могут найти применение в теории оценивания, а также в прикладных исследованиях, связанных с теорией риска. Результаты, касающиеся обобщенного распределения Лапласа, могут применяться в задачах о различении близких гипотез, выделении полезного сигнала на фоне помех.

Апробация работы. Основные результаты диссертации докладывались на научном семинаре «Теория риска и смежные вопросы» под руководством профессора В. Е. Бенинга, профессора В. Ю. Королёва и стар. преп. А. А. Кудрявцева, на X Всероссийском симпозиуме по прикладной и промышленной математике (1 – 8 октября 2009 г., Сочи – Дагомыс).

Публикации. Основные результаты диссертации опубликованы в 4 печатных работах, из них 3 статьи [1, 3, 4] в журналах, входящих в список ВАК «*Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертации на соискание ученой степени доктора и кандидата наук*», и 1 работа в сборниках трудов конференций [2].

Структура и объем диссертации. Диссертация состоит из введения, двух глав, заключения и списка литературы, содержащего 64 наименования. Общий объем работы составляет 99 страниц.

Благодарности. Автор выражает глубокую благодарность профессору Бенингу Владимиру Евгеньевичу, под руководством которого проходила работа над диссертацией, за постановку задачи и постоянное внимание к работе.

Содержание работы

Введение содержит общую характеристику работы, описание объектов исследования и основных результатов.

В первой главе получена оценка скорости сходимости распределения случайной величины, равной максимуму от n независимых случайных величин с одним и тем же дискретным распределением Парето, к обратному показательному распределению при $n \to \infty$. Здесь также получена оценка скорости сходимости распределений асимптотически нормальных статистик к распределению Лапласа в случае выборок случайного объема с распределением указанной случайной величины.

Рассмотрим случайные величины $N_1, N_2, \ldots, X_1, X_2, \ldots$, определенные на одном и том же измеримом пространстве $(\Omega,\ \mathcal{A})$. Пусть на $\mathcal{A}$ задана вероятностная мера P. Предположим, что при каждом $n \geq 1$ случайные величины N_n принимают только натуральные значения и не зависят от последовательности $X_1, X_2, \ldots$. Пусть $T_n = T_n(X_1, \ldots, X_n)$ — некоторая статистика, то есть измеримая функция от случайных величин $X_1, \ldots, X_n$. Для каждого $n \geq 1$ определим случайную величину T_{N_n}, положив

$$T_{N_n}(\omega) \ = \ T_{N_n(\omega)}\left(X_1(\omega),\ \ldots\ , X_{N_n(\omega)}(\omega)\right)$$

для каждого элементарного исхода $\omega \in \Omega$. Будем говорить, что статистика T_n асимптотически нормальна, если существуют числа $\sigma > 0$ и $\mu \in \mathbb{R}$ такие, что

$$\mathsf{P}\left(\sigma\sqrt{n}(T_n \ - \ \mu) \ < \ x\right) \ \Longrightarrow \ \Phi(x), \quad n \ \to \ \infty, \tag{1}$$

где $\Phi(x)$ — функция распределения стандартного нормального закона.

Примеры асимптотически нормальных статистик хорошо известны. Свойством асимптотической нормальности обладают, например, выборочное среднее (при условии существования дисперсий), центральные порядковые статистики, оценки максимального правдоподобия (при достаточно общих условиях регулярности) и многие другие статистики.

Ранее была доказана лемма (см., например, [[13]]) о необходимых и достаточных условиях сходимости распределений асимптотически нормальных статистик, построенных по выборкам случайного объема, к заданному распределению $F(x)$.

Лемма 1. *(Бенинг, Королёв 2008) Пусть $\{d_n\}_{n\geq 1}$ — неограниченно возрастающая последовательность положительных чисел. Предположим, что $N_n \to \infty$ по вероятности при $n \to \infty$. Пусть статистика T_n асимптотически нормальна в смысле (1). Для того чтобы существовала такая функция распределения $F(x)$, что*

$$\mathsf{P}\left(\sigma\sqrt{d_n}(T_{N_n} - \mu) < x\right) \Longrightarrow F(x), \quad n \to \infty,$$

необходимо и достаточно, чтобы существовала функция распределения $H(x)$, удовлетворяющая условиям

$$H(x) = 0, \quad x < 0;$$

$$F(x) = \int\limits_0^\infty \Phi\left(x\sqrt{y}\right) dH(y), \quad x \in \mathbb{R};$$

$$\mathsf{P}(N_n < d_n x) \Longrightarrow H(x), \quad n \to \infty.$$

Распределение Лапласа может быть представлено в виде масштабной смеси нормальных законов с нулевым средним при обратном показательном смешивающем распределении, а именно: для любого $x \in \mathbb{R}$

$$\Lambda(x) = \int\limits_0^\infty \Phi(x\sqrt{y})dQ(y),$$

где $Q(x)$ — функция распределения обратного показательного распределения

$$Q(x) = e^{-\delta/x}, \quad \delta > 0, \quad x > 0 \tag{2}$$

и $\Lambda(x)$ — функция распределения распределения Лапласа, соответствующая плотности

$$\lambda(x) = \sqrt{\frac{\delta}{2}}\, e^{-\sqrt{2\delta}|x|}, \quad x \in \mathbb{R}. \tag{3}$$

Обратное показательное распределение — это распределение случайной величины

$$V = \frac{1}{U},$$

где случайная величина U имеет показательное распределение. Обратное показательное распределение является частным случаем распределения Фреше, хорошо известного в асимптотической теории экстремальных порядковых статистик как предельное распределение типа II.

Из леммы 1 непосредственно следует следующая теорема (см. [13]), дающая необходимые и достаточные условия сходимости распределений асимптотически нормальных статистик, построенных по выборкам случайного объема, к распределению Лапласа.

Теорема 2. *(Бенинг, Королёв 2008) Пусть $\delta > 0$ произвольно и $\{d_n\}_{n\geq 1}$ — некоторая неограниченно возрастающая последовательность положительных чисел. Предположим, что $N_n \to \infty$ по вероятности при $n \to \infty$. Пусть статистика T_n асимптотически нормальна в смысле (1). Для справедливости соотношения*

$$\mathsf{P}(\sigma\sqrt{d_n}(T_{N_n} - \mu) < x) \Longrightarrow \Lambda(x), \quad n \to \infty$$

необходимо и достаточно, чтобы

$$\mathsf{P}(N_n < d_n x) \Longrightarrow Q(x), \quad n \to \infty.$$

Приведем пример ситуации, в которой случайный объем выборки имеет предельное обратное показательное распределение $Q(x)$. Пусть $Y_1, Y_2, \ldots$ — независимые одинаково распределенные случайные величины с одной и той же непрерывной функцией распределения. Пусть m — произвольное натуральное число. Обозначим

$$N(m) = \min\{n \geq 1 : \max_{1\leq j\leq m} Y_j < \max_{m+1\leq k\leq m+n} Y_k\}.$$

Случайная величина $N(m)$ имеет смысл количества дополнительных наблюдений, которые надо произвести, чтобы текущий (по m наблюдениям) максимум был перекрыт. Распределение случайной величины $N(m)$ было найдено С. Уилксом, который показал, что распределение величины $N(m)$ является дискретным распределением Парето:

$$\mathsf{P}\left(N(m) \geq k\right) = \frac{m}{m + k - 1}, \quad k \geq 1. \tag{4}$$

Пусть теперь $N^{(1)}(m), N^{(2)}(m), \ldots$ — независимые случайные величины с одним и тем же распределением (4). В работе [13] было доказано, что для любого $x > 0$

$$\mathsf{P}\left(\frac{1}{n} \cdot \max_{1\leq j\leq n} N^{(j)}(m) < x\right) \Longrightarrow e^{-m/x}, \quad n \to \infty.$$

Заметим, что в правой части предельного выражения стоит функция распределения обратного показательного распределения $Q(x)$ с параметром формы $\delta = m$. Поэтому, если положить

$$N_n \;=\; \max_{1 \le j \le n} N^{(j)}(m), \tag{5}$$

теорема 2 с $d_n = n$ дает иллюстрацию того, как вместо ожидаемого в соответствии с утверждениями классической асимптотической статистики нормального распределения при замене объема выборки случайной величиной в качестве предельного распределения регулярных статистик возникает распределение Лапласа.

Теорема 3. *(Бенинг, Королёв 2008) Пусть m — произвольное натуральное число. Предположим, что $N^{(1)}(m), N^{(2)}(m), \ldots$ — независимые случайные величины с одним и тем же распределением (4), и случайная величина N_n определяется формулой (5). Пусть статистика T_n асимптотически нормальна в смысле (1). Тогда*

$$\mathsf{P}\left(\sigma\sqrt{n}(T_{N_n} \;-\; \mu) \;<\; x\right) \Longrightarrow \Lambda(x), \quad n \to \infty,$$

где $\Lambda(x)$ — функция распределения распределения Лапласа с плотностью вида (3) с $\delta = m$.

Таким образом, основной вывод из приведенных выше результатов можно сформулировать следующим образом. Если число случайных факторов, определяющих наблюдаемое значение случайной величины, само является случайной величиной, распределение которой может быть приближено обратным показательным распределением (например, является случайной величиной вида (5)), то те функции от значений случайных факторов, которые в классической ситуации считаются асимптотически нормальными, в действительности являются асимптотически лапласовскими.

В приводимой ниже теореме изучается вопрос о предельном поведении распределения случайной величины N_n/n при $n \to \infty$, где N_n — случайная величина вида (5) с некоторым натуральным параметром m.

Теорема 4. *Для каждого натурального m существует константа $C_m > 0$ такая, что*

$$\sup_{x>0}\left|\mathsf{P}\left(\frac{N_n}{n} \;<\; x\right) \;-\; Q(x)\right| \;\le\; \frac{C_m}{n}, \quad n \to \infty, \tag{6}$$

где $Q(x)$ — функция распределения обратного показательного распределения, определяемая формулой (2) с $\delta = m$. При этом

$$C_m = \begin{cases} 8e^{-2}/3, & m = 1, \\ 2e^{-2}, & m \geq 2. \end{cases}$$

Далее приведена общая теорема, позволяющая автоматически получать оценки скорости сходимости распределений статистик, построенных по выборкам случайного объема, к распределению Лапласа из оценок скорости сходимости к нормальному закону этих же статистик, но уже построенных по обычным, неслучайным выборкам. При этом предполагается, что объем выборок является случайной величиной вида (5).

Теорема 5. *Пусть $T_n = T_n(X_1, \ldots, X_n)$ — асимптотически нормальная статистика, и для ее распределения справедлива оценка скорости сходимости вида: для некоторого $0 < s \leq s_0 < 2$ существуют числа $\sigma > 0$ и $\mu \in \mathbb{R}$ такие, что*

$$\sup_x \left| \mathsf{P}\left(\sigma\sqrt{n}(T_n - \mu) < x\right) - \Phi(x) \right| = \mathcal{O}(n^{-s}), \quad n \to \infty\,. \tag{7}$$

Пусть N_n — случайная величина вида (5) с некоторым натуральным параметром m, которая не зависит от исходных случайных величин $X_1, X_2, \ldots$ и $n \to \infty$. Пусть $\Lambda(x)$ — функция распределения распределения Лапласа с плотностью вида (3) с $\delta = m$. Тогда распределение статистики $T_{N_n} = T_{N_n}(X_1, \ldots, X_{N_n})$ удовлетворяет равенству

$$\sup_x \left| \mathsf{P}\left(\sigma\sqrt{n}(T_{N_n} - \mu) < x\right) - \Lambda(x) \right| = \mathcal{O}(n^{-\min\{1,\, s\}}), \quad n \to \infty.$$

Далее приведены два примера использования теоремы 5. Первый пример касается U–статистик, а второй — линейных комбинаций порядковых статистик, широко применяемых в статистике (см., например, [[16]]).

Пусть $X_1, X_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин, имеющих общую функцию распределения $F(x)$. Определим U–статистику $T_n^{(1)}$ по формуле

$$T_n^{(1)} = \frac{1}{n\,(n-1)} \sum_{1 \leq i < j \leq n} h(X_i,\, X_j), \quad n \geq 2,$$

где ядро $h(x,\, y)$ — симметричная измеримая функция двух переменных.

Определим теперь линейные комбинации порядковых статистик. Пусть $(X_{1:n} \leq \ldots \leq X_{n:n})$ — вариационный ряд, построенный по исходной выборке $(X_1, \ldots, X_n)$. Тогда линейная комбинация порядковых статистик определяется по формуле

$$T_n^{(2)} = \frac{1}{n} \sum_{i=1}^{n} c_{in} X_{i:n},$$

где c_{in} — некоторые числа.

Пусть для статистик $T_n^{(i)}$, $i = 1, 2$ выполнены условия регулярности, сформулированные в работах Р. Хелмерса [19, 20, 21]. В этом случае справедливы следующие соотношения

$$\sup_x \left| \mathsf{P}(\sigma_i \sqrt{n} (T_n^{(i)} - \mu_i) < x) - \Phi(x) \right| = \mathcal{O}(n^{-1/2}), \quad n \to \infty,$$

где величины σ_i, $i = 1, 2$; $\mu_1 = 0$, $\mu_2 = \mu$ определены в указанных работах Р. Хелмерса.

Пусть теперь N_n – случайная величина вида (5) с некоторым натуральным m, которая не зависит от исходных случайных величин $X_1, X_2 \ldots$ и $n \to \infty$. Рассмотрим статистики $T_{N_n}^{(i)}$, $i = 1, 2$, построенные по выборке случайного объема N_n, то есть

$$T_{N_n}^{(i)} = T_{N_n}^{(i)}(X_1, \ldots, X_{N_n}).$$

Следующая теорема является непосредственным следствием теоремы 5.

Теорема 6. *Пусть для статистик $T_n^{(i)}$, $i = 1, 2$ выполнены условия регулярности, сформулированные в работах Р. Хелмерса. Тогда распределения статистик $T_{N_n}^{(i)}$, $i = 1, 2$ удовлетворяют равенству*

$$\sup_x \left| \mathsf{P}(\sigma_i \sqrt{n} (T_{N_n}^{(i)} - \mu_i) < x) - \Lambda(x) \right| = \mathcal{O}(n^{-1/2}), \quad n \to \infty,$$

где величины σ_i, $i = 1, 2$; $\mu_1 = 0$, $\mu_2 = \mu$ определены в работах Р. Хелмерса, а $\Lambda(x)$ – функция распределения распределения Лапласа с плотностью вида (3) с $\delta = m$.

[19] Helmers R., van Zwet W. R. The Berry–Esseen bound for U-statistics // In: Statistical decision theory and related topics, III / Ed. by S. S. Gupta, J. O. Berger. — New York: 1982. — P. 497–512.

[20] Helmers R. Edgeworth Expansions for Linear Combinations of Order Statistics. — Amsterdam: Mathematisch Centrum. — 1984.

[21] Helmers R., Berry A. Esseen bound for linear combinations of order statistics // Ann. Probab. — 1981. — P. 342–347.

Утверждение теоремы 5 можно усилить, если несколько изменить условие, наложенное на скорость сходимости распределения асимптотически нормальной статистики T_n к нормальному распределению. Вместо условия (7) потребуем, чтобы для некоторого $0 < s \leq s_0 < 2$ была справедлива оценка

$$\sup_x \left| P(\sigma\sqrt{n}(T_n - \mu) < x) \, - \, \Phi(x) \right| \, \leq \, Cn^{-s}, \quad n = 1, 2, \ldots, \tag{8}$$

и перепишем результат теоремы 5 при $n \to \infty$ в виде

$$\sup_x \left| \mathsf{P}\left(\sigma\sqrt{n}(T_{N_n} \, - \, \mu) \, < \, x\right) \, - \, \Lambda(x) \right| \, \leq \, C_s n^{-\min(1,s)} + o(n^{-\min(1,s)}). \tag{9}$$

В диссертации доказана следующая теорема.

Теорема 7. *Для константы C_s из (9) справедливо соотношение*

$$C_s \, = \begin{cases} \frac{C\Gamma(1+s)}{m^s}, & s < 1, \\ \frac{C}{m} + C_{1,m}, & s = 1, \\ C_{1,m}, & s > 1, \end{cases}$$

где

$$C_{1,m} \, = \begin{cases} 4e^{-2}/3, & m = 1, \\ e^{-2}, & m \geq 2 \end{cases}$$

и C — константа из неравенства (8).

Воспользуемся теоремой для получения оценки скорости сходимости распределения выборочного среднего, построенного по выборке объема, равного случайной величине (5), к распределению Лапласа. Пусть $X_1, \ldots, X_n$ — независимые одинаково распределенные случайные величины с математическим ожиданием 0 и дисперсией 1. Положим $\mu_3 = \mathsf{E}|X_1|^3 < \infty$. Тогда для статистики $T_n = \sum_{i=1}^n X_i/n$ по неравенству Берри-Эссеена справедлива оценка

$$\sup_x \left| \mathsf{P}(\sqrt{n}\, T_n < x) \, - \, \Phi(x) \right| \, \leq \, \frac{0.4784\mu_3}{\sqrt{n}}, \quad n = 1, 2, \ldots$$

Пусть N_n — случайная величина вида (5) с некоторым натуральным m. Применяя теорему 7 для случая $s = 1/2$, получим

$$\sup_x \left| \mathsf{P}\left(\sqrt{n}\, T_{N_n} < x\right) \, - \, \Lambda(x) \right| \, \leq \, \frac{0.4784\sqrt{\pi}\mu_3}{2\sqrt{m}} \cdot \frac{1}{\sqrt{n}} \, + \, o\left(\frac{1}{\sqrt{n}}\right).$$

Справедлива также более грубая, но не зависящая от m, оценка:

$$\sup_x \left| \mathsf{P}\left(\sqrt{n}\, T_{N_n} < x\right) \, - \, \Lambda(x) \right| \, \leq \, 0.2392\sqrt{\pi}\mu_3 \cdot \frac{1}{\sqrt{n}} \, + \, o\left(\frac{1}{\sqrt{n}}\right).$$

Во второй главе предложено обобщенное распределение Лапласа. Для него доказано, что отсутствие регулярности распределения приводит к нарушению естественного порядка n^{-1} разности функций мощности наилучшего и асимптотически наиболее мощного критериев и приводит к порядку $n^{-1/2}$. Получена формула предела отклонения функций мощности.

Пусть имеются независимые наблюдения $\mathbf{X}_n = (X_1, \ldots, X_n)$, каждое из которых принимает значения в измеримом пространстве $(\mathcal{X}, \mathcal{A})$ и имеет неизвестную с точностью до параметра θ плотность $p(x, \theta)$ относительно некоторой σ-конечной меры $\nu(\cdot)$ на $\mathcal{A}$. Предположим, что неизвестный параметр θ принадлежит открытому множеству $\Theta \subset \mathbb{R}$, содержащему 0. Обозначим через $\mathsf{P}_{n,\theta}, \mathsf{P}_\theta$ соответственно распределения $\mathbf{X}_n$ и X_1, а через $\mathsf{E}_{n,\theta}, \mathsf{E}_\theta$ — соответствующие математические ожидания.

Рассмотрим задачу проверки простой гипотезы

$$\mathsf{H}_0 : \theta \;=\; 0$$

против последовательности сложных близких альтернатив вида

$$\mathsf{H}_{n,1} : \theta \;=\; \frac{t}{\sqrt{n}}, \quad 0 < t \le C, \quad C > 0,$$

где параметр t неизвестен. Согласно фундаментальной лемме Неймана-Пирсона для любого фиксированного $t \in (0, C]$ наилучший (наиболее мощный) критерий для проверки гипотезы H_0 против простой альтернативы

$$\mathsf{H}_{n,t} : \theta \;=\; \frac{t}{\sqrt{n}}$$

основан на логарифме отношения правдоподобия

$$\Lambda_n(t) \;\equiv\; \sum_{i=1}^{n} \left(l\left(X_i, \frac{t}{\sqrt{n}}\right) - l(X_i, 0) \right),$$

где $l(x, \theta) = \log p(x, \theta)$, и отвергает гипотезу H_0, если

$$\Lambda_n(t) \;>\; c_{n,t},$$

причем критическое значение выбирается из условия

$$\mathsf{P}_{n,0}(\Lambda_n(t) \;>\; c_{n,t}) \;=\; \alpha,$$

где $\alpha \in (0,1)$ — фиксированный уровень значимости, и мы для простоты предполагаем непрерывность распределения $\Lambda_n(t)$ при гипотезе H_0, то есть считаем, что

$$\mathsf{P}_{n,\,0}(\Lambda_n(t) = c_{n,t}) \;=\; 0.$$

Обозначим через $\beta_n^*(t)$ мощность такого критерия, то есть положим

$$\beta_n^*(t) \;=\; \mathsf{P}_{n,t/\sqrt{n}}\,(\Lambda_n(t) > c_{n,t}).$$

Поскольку t неизвестно, мы не можем использовать статистику $\Lambda_n(t)$ для построения критерия проверки гипотезы H_0 против альтернативы $\mathsf{H}_{n,1}$. Однако $\beta_n^*(t)$ дает верхнюю границу для мощности любого критерия при проверке гипотезы H_0 против фиксированной альтернативы $\mathsf{H}_{n,\,t}, t > 0$ и может служить стандартом при сравнении различных критериев.

Ранее (см., например, [16]) при естественных условиях регулярности, состоящих в существовании необходимого числа моментов случайных величин $l(X_1,\theta)$ и всех необходимых производных по θ функций $l(x,\theta)$, при $n \to \infty$ были получены следующие соотношения

$$\begin{aligned}
\Lambda_n(t) \;&=\; tL_n^{(1)} - \frac{1}{2}\,t^2 I + \frac{1}{2}\,\frac{t^2}{\sqrt{n}}\,L_n^{(2)} + \ldots,\\
\mathcal{L}(\Lambda_n(t)|\mathsf{H}_0) \;&\longrightarrow\; \mathcal{N}\big(-t^2I/2, t^2I\big),\\
\mathcal{L}(\Lambda_n(t)|\mathsf{H}_{n,t}) \;&\longrightarrow\; \mathcal{N}\big(t^2I/2, t^2I\big),\\
c_{n,t} \;\longrightarrow\; c_t \;&=\; t\sqrt{I}u_\alpha - \frac{1}{2}\,t^2 I,\\
\beta_n^*(t) \;\longrightarrow\; \beta^*(t) \;&=\; \Phi(t\sqrt{I} - u_\alpha),
\end{aligned} \tag{10}$$

где

$$L_n^{(j)} \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(l^{(j)}(X_i) - \mathsf{E}_0 l^{(j)}(X_i)), \quad j = 1,2,\ldots,$$

$$l^{(j)}(x) \;=\; \frac{\partial^j}{\partial\theta^j}\,l(x,\theta)\bigg|_{\theta=0}, \quad j = 1,2,\ldots,$$

$I \;=\; \mathsf{E}_0(l^{(1)}(X_1))^2$ — фишеровская информация, $\mathcal{L}(\Lambda_n(t)|\mathsf{H}_0)$, $\mathcal{L}(\Lambda_n(t)|\mathsf{H}_{n,t})$ — распределения $\Lambda_n(t)$ при соответствующих гипотезах, $\mathcal{N}$ — нормальное распределение с соответствующими параметрами и $u_\alpha = \Phi^{-1}(1-\alpha)$.

Для проверки гипотезы H_0 против альтернативы $\mathsf{H}_{n,t}$ существуют критерии, основанные на статистиках, отличных от $\Lambda_n(t)$ и не зависящих от t, и имеющие ту же предельную мощность $\beta^*(t)$. Такие критерии называются асимптотически наиболее мощными (АНМ) (точнее локально АНМ, поскольку альтернатива $\mathsf{H}_{n,1}$ имеет локальный характер). Таковы, например, критерии, основанные на статистиках $L_n^{(1)}$, $\Lambda_n(t_0)$, где $t_0 > 0$ фиксировано, оценках максимального правдоподобия. Все эти статистики не зависят от неизвестного параметра t, и поэтому могут быть использованы при проверке гипотезы H_0 против альтернативы $\mathsf{H}_{n,1}$.

Соотношение (10) дает естественную основу для асимптотического сравнения различных АНМ критериев, однако для различения критериев такого рода, то есть удовлетворяющих соотношению

$$\beta_n(t) \longrightarrow \beta^*(t), \quad n \to \infty,$$

где $\beta_n(t)$ — мощность конкретного рассматриваемого критерия, нужны следующие члены асимптотического разложения $\beta_n(t)$, то есть представление типа

$$\beta_n(t) = \beta^*(t) + \frac{1}{\sqrt{n}} h_1(t) + \frac{1}{n} h_2(t) + \dots .$$

Было обнаружено, что для широкого класса АНМ критериев мощность $\beta_n(t)$ АНМ критерия отличается от мощности $\beta_n^*(t)$ наилучшего критерия на величину порядка $1/n$. При этом величина

$$r(t) = \lim_{n\to\infty} n(\beta_n^*(t) - \beta_n(t)),$$

допускает статистическую интерпретацию в терминах необходимого числа наблюдений и позволяет находить асимптотический дефект (см. работы Д. М. Чибисова[14,15]).

Первоначально выражения для величины $r(t)$ строились с помощью асимптотических разложений для $\beta_n^*(t)$ и $\beta_n(t)$ (см., например, [14,22]). Этот подход технически очень трудоемкий и громоздкий. Однако в работе В. Е. Бенинга[16] был рассмотрен общий случай в терминах общего статистического эксперимента и приведена теорема, дающая достаточные условия для существования предела

$$r(t) \equiv \lim_{n\to\infty} \tau_n^{-2} (\beta_n^*(t) - \beta_n(t)),$$

где $\tau_n \to 0$ — малый параметр.

Теорема 8. *(Bening, 2000) Пусть выполнены условия регулярности (см. теорему 3.2.1 из работы [16]), и $0 < \alpha < 1$. Тогда*

$$r(t) \equiv \lim_{n\to\infty} \tau_n^{-2} (\beta_n^*(t) - \beta_n(t)) = \frac{1}{2} e^d p(d)\, \mathsf{D}(\Delta \,|\Lambda = d),$$

[22] Hodges J. L., Lehmann E. L. Deficiency // Ann. Math. Statist. — 1970. — Vol. 41, N. 5. — P. 783–801.

где $d = \Phi_1^{-1}(1-\alpha)$, $\Phi_1(x)$ — *функция распределения, предельная для логарифма отношения правдоподобия* $\Lambda_n(t)$ *при гипотезе* H_0, $p(x) = \Phi_1'(x)$, (Δ, Λ) — *случайный вектор, предельный для* $(\tau_n^{-1}\Delta_n(t), \Lambda_n(t))$ *при гипотезе* H_0, $\Delta_n(t) = S_n(t) - \Lambda_n(t)$, $S_n(t)$ — *монотонное (не меняющее мощности критерия) преобразование статистики критерия* T_n.

Как уже было отмечено, в типичном случае, когда соблюдены естественные условия регулярности, $\tau_n = n^{-1/2}$ и $\beta_n(t)$ отличается от $\beta_n^*(t)$ на величину порядка n^{-1}. Заметим, что здесь следует различать естественные условия регулярности и условия регулярности, указанные в теореме 8. Естественные условия регулярности заключаются в существовании всех необходимых моментов случайных величин $l(X_1, \theta)$ и существовании необходимых производных по θ для функций $l(x, \theta)$. Отсутствие регулярности может приводить к нарушению естественного порядка n^{-1} разности $\beta_n^*(t) - \beta_n(t)$ и приводить к другим порядкам.

Назовем обобщенным распределением Лапласа распределение с плотностью вида

$$p(x) \;=\; C_{a,\,b}e^{-ax^2 - b|x|}, \quad a \geq 0, \quad b > 0, \quad x \in \mathbb{R}, \tag{11}$$

где $C_{a,\,b}$ — константа нормировки такая, что

$$C_{a,\,b} \;=\; \begin{cases} \frac{b}{2}, & a = 0; \\ \frac{\sqrt{a}}{\sqrt{\pi}\exp\left\{\frac{b^2}{4a}\right\}\mathsf{erfc}\left(\frac{b}{2\sqrt{a}}\right)}, & a > 0 \end{cases}$$

и

$$\mathsf{erfc}\,(x) \;=\; \frac{2}{\sqrt{\pi}}\int\limits_x^\infty e^{-y^2}dy.$$

Обобщенное распределение Лапласа может оказаться полезным в тех случаях, когда необходим более тонкий контроль за поведением функции плотности, чем может быть предоставлен однопараметрическими по параметру формы нормальным и лапласовским распределениями. Так, правильно подобрав параметры, можно получить распределение с хвостами менее тяжелыми, чем у соответствующего распределения Лапласа, при этом сохранив существенную особенность негладкости функции плотности, которая отсутствует у нормального распределения.

Обычное распределение Лапласа получается при $a = 0$. Для этого случая известно, что условия регулярности теоремы 8 не выполнены. Далее всюду будем предполагать $a > 0$.

Для обобщенного распределения Лапласа с параметром сдвига $p(x,\theta) = p(x-\theta)$ рассмотрим задачу проверки простой гипотезы против последовательности близких сложных альтернатив в описанной постановке. Следующая лемма описывает асимптотическое поведение распределения логарифма отношения правдоподобия при основной гипотезе и альтернативе $\mathsf{H}_{n,t}$.

Лемма 9. *В случае распределения (11) справедливы следующие соотношения*

$$I_{a,\,b} \;=\; 2(bC_{a,\,b} + a) \;-\; \textit{фишеровская информация},$$
$$\mathcal{L}(\Lambda_n(t)|\mathsf{H}_0) \;\longrightarrow\; \mathcal{N}\big(-I_{a,\,b}t^2/2, I_{a,\,b}t^2\big),$$
$$\mathcal{L}(\Lambda_n(t)|\mathsf{H}_{n,t}) \;\longrightarrow\; \mathcal{N}\big(I_{a,\,b}t^2/2, I_{a,\,b}t^2\big), \quad n\to\infty.$$

Из этой леммы вытекает следующее утверждение.

Следствие 10. *В случае распределения (11) справедливо соотношение*

$$\beta_n^*(t) \;\longrightarrow\; \beta^*(t) \;=\; \Phi\big(t\sqrt{I_{a,\,b}} - u_\alpha\big), \quad n \;\to\; \infty.$$

Ранее было отмечено, что это соотношение справедливо в регулярном случае. Заметим, что семейство (11), также как и семейство плотностей обычного распределения Лапласа, не является регулярным, поскольку у $p(x,\theta)$ не существует производной по θ в точке $\theta = x$. Тем самым показано, что отсутствие дифференцируемости по θ функции плотности $p(x,\theta)$ в точке $\theta = x$ качественно не влияет на порядок альтернатив θ_n (равный $\frac{1}{\sqrt{n}}$) и вид предельной мощности $\beta^*(t)$.

Рассмотрим критерий, основанный на асимптотически эффективной статистике

$$T_n \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[2aX_i + b\,\mathsf{sign}\,(X_i)].$$

Заметим, что статистика T_n не зависит от t, и потому может быть использована для проверки гипотезы H_0 против альтернативы $\mathsf{H}_{n,1}$. Пусть $\beta_n(t)$ — функция мощности этого критерия заданного уровня $\alpha \in (0,1)$. Следующая лемма показывает, что критерий, основанный на статистике T_n, является АНМ критерием.

Лемма 11. *В случае распределения (11) для любого* $0 \le \delta < 1/2$ *справедливо соотношение*

$$n^\delta(\beta_n^*(t) - \beta_n(t)) \;\longrightarrow\; 0, \quad n\to\infty.$$

Основным результатом второй главы является получение предела выражения $\sqrt{n}(\beta_n^*(t)-\beta_n(t))$ с использованием теоремы 8. Рассмотрим монотонное преобразование статистики T_n вида $S_n(t) = tT_n - t^2 I_{a,\,b}/2$ и положим $\Delta_n(t) = S_n(t) - \Lambda_n(t)$. В следующей лемме устанавливаются предельные распределения случайной величины $\sqrt[4]{n}\Delta_n(t)$ и случайного вектора $(\sqrt[4]{n}\Delta_n(t), \Lambda_n(t))$ при основной гипотезе.

Лемма 12. *В случае распределения (11) при $n \to \infty$ справедливы следующие соотношения*

$$\mathcal{L}(\sqrt[4]{n}\Delta_n(t)|\mathsf{H}_0) \;\longrightarrow\; \mathcal{N}\big(0, 4C_{a,\,b}b^2t^3/3\big),$$
$$\mathcal{L}((\sqrt[4]{n}\Delta_n(t), \Lambda_n(t))|\mathsf{H}_0) \;\longrightarrow\; \mathcal{N}_2\big(0, 4C_{a,\,b}b^2t^3/3, 0, -I_{a,\,b}t^2/2, I_{a,\,b}t^2\big),$$

где $\mathcal{N}_2$ — двумерный нормальный закон с соответствующими параметрами.

Эта лемма показывает, что случайные величины $\sqrt[4]{n}\Delta_n(t)$ и $\Lambda_n(t)$ асимптотически независимы. Полагая Δ и Λ независимыми случайными величинами с распределениями, равными предельным распределениям случайных величин $\sqrt[4]{n}\Delta_n(t)$ и $\Lambda_n(t)$, то есть

$$\Delta \sim \mathcal{N}(0, 4C_{a,\,b}b^2t^3/3), \quad \Lambda \sim \mathcal{N}(-I_{a,\,b}t^2/2, I_{a,\,b}t^2),$$

и $\tau_n = n^{-1/4}$, можно на эвристическом уровне из теоремы 8 получить формулу для предела выражения $\sqrt{n}(\beta_n^*(t) - \beta_n(t))$. Для формального доказательства этой формулы в диссертации проверены условия регулярности теоремы 8. Таким образом, установлена следующая теорема.

Теорема 13. *Для распределения (11) справедлива следующая формула*

$$r(t) \;\equiv\; \lim_{n\to\infty} \sqrt{n}(\beta_n^*(t) - \beta_n(t)) \;=\; \frac{2C_{a,\,b}b^2t^2}{3\sqrt{I_{a,\,b}}}\,\varphi\left(t\sqrt{I_{a,\,b}} - u_\alpha\right),$$

где $\varphi(x)$ — плотность стандартного нормального закона.

Тем самым показано, что обобщенное распределение Лапласа может служить примером применения теоремы 8 для случая нерегулярного распределения и что отсутствие регулярности этого распределения приводит к порядку $n^{-1/2}$ отклонения мощности асимптотически наиболее мощного критерия от мощности наилучшего критерия, в отличие от регулярного случая, для которого этот порядок равен n^{-1}.

В заключении приводятся выводы диссертационной работы и возможные перспективы дальнейших исследований.

Основные результаты работы

В диссертационной работе получено обоснование возможности использования распределения Лапласа в задачах теории вероятности и математической статистики, возникающего в качестве предельного в случае выборок случайного объема, а также построение примера применения общей теоремы, дающей достаточные условия для существования предела отклонения мощности асимптотически наиболее мощного критерия от мощности наилучшего критерия, в случае нерегулярного распределения.

1. Получена оценка скорости сходимости распределения случайной величины, равной максимуму от n независимых случайных величин с одним и тем же дискретным распределением Парето, к обратному показательному распределению при $n \to \infty$.

2. Построены оценки скорости сходимости распределений асимптотически нормальных статистик к распределению Лапласа в случае выборок случайного объема с распределением указанного максимума.

3. Предложены варианты использования полученных результатов в исследованиях с применением U–статистик, а также линейных комбинаций порядковых статистик.

4. Предложено обобщенное распределение Лапласа, которое при правильном подборе параметров обладает хвостами менее тяжелыми, чем у соответствующего распределения Лапласа, сохраняя при этом существенную особенность негладкости функции плотности, которая отсутствует у нормального распределения.

5. Для этого распределения описана задача проверки простой гипотезы против последовательности сложных близких альтернатив. С применением общей теоремы из работы [16] показано, что отсутствие регулярности приводит к нарушению естественного порядка n^{-1} разности функций мощности наилучшего и асимптотически наиболее мощного критериев и приводит к порядку $n^{-1/2}$.

Список публикаций по теме диссертации

1. Бенинг В. Е., Лямин О. О. О мощности критериев в случае обобщенного распределения Лапласа // Информатика и ее применения. — 2009. — Т. 3, № 3. — С. 79–85.

2. Лямин О. О. О скорости сходимости распределений некоторых статистик к распределению Лапласа // Обозрение прикладной и промышленной математики. — 2009. — Т. 16, № 6. — С. 1090–1091.

3. Лямин О. О. О предельном поведении мощностей критериев в случае обобщенного распределения Лапласа // Информатика и ее применения. — 2010. — Т. 4, № 3. — С. 49–59.

4. Лямин О. О. О скорости сходимости распределений некоторых статистик к распределению Лапласа // Вестн. Моск. Ун-та. Сер. 15. Вычисл. матем. и киберн. — 2010. — № 3. — С. 30–38.