

МИНИСТЕРСТВО ОБЩЕГО И ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

На правах рукописи

Лемешко Борис Юрьевич

**СТАТИСТИЧЕСКИЙ АНАЛИЗ
ГРУППИРОВАННЫХ, ЧАСТИЧНО ГРУППИРОВАННЫХ И
НЕГРУППИРОВАННЫХ НАБЛЮДЕНИЙ ОДНОМЕРНЫХ
НЕПРЕРЫВНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН**

Специальность 05.13.16 - применение вычислительной техники,
математического моделирования и математических методов в
научных исследованиях (в области технических наук)

Диссертация на соискание ученой степени
доктора технических наук

Новосибирск - 1997 г.

Содержание

ВВЕДЕНИЕ	5
1. СТРУКТУРА ПРЕДСТАВЛЕНИЯ ДАННЫХ, МЕТОДЫ ОЦЕНИВАНИЯ И УСЛОВИЯ СУЩЕСТВОВАНИЯ ОЦЕНОК	15
1.1. Структура представления данных	15
1.2. Виды оценок и методы оценивания	18
1.3. Условия существования ОМП по частично группированным данным.....	24
1.4. Оценивание параметров распределений по интервальным наблюдениям	43
1.5. Выводы	46
2. АСИМПТОТИЧЕСКИ ОПТИМАЛЬНОЕ ГРУППИРОВАНИЕ И ОЦЕНИВАНИЕ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЙ	47
2.1. Асимптотическая эффективность оценок и асимптотически оптимальное группирование данных	47
2.2. Решение задачи асимптотически оптимального группирования	50
2.3. О возможности практического использования асимптотически оптимального группирования в задачах оценивания параметров	70
2.4. Оценки параметров распределений по выборочным асимптотически оптимальным квантилям	75
2.5. Оптимальные L-оценки параметров сдвига и масштаба по выборочным квантилям	84
2.6. Выбор числа интервалов	96
2.7. Выводы	98
3. РОБАСТНОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЙ.....	99
3.1. Способы вычисления робастных оценок	99
3.2. Группирование наблюдений как способ получения робастных оценок .	100
3.3. Функции влияния и робастность оценок	113
3.4. Робастное оценивание и проблема отбраковки аномальных наблюдений	121
3.5. Выводы	125
4. ВОПРОСЫ ПРИМЕНЕНИЯ КРИТЕРИЕВ СОГЛАСИЯ	127
4.1. Проверка гипотез о согласии	127
4.2. Асимптотически оптимальное группирование наблюдений в критериях согласия	131
4.3. О зависимости предельных распределений статистик χ^2 Пирсона и отношения правдоподобия от способа группирования.....	147

4.4. Применение непараметрических критериев согласия в случае потери ими свойства “свободы от распределения”	166
4.4.1. Влияние оценивания параметров закона на предельные распределения статистик непараметрических критериев	166
4.4.2. Моделирование и идентификация законов распределения статистики типа Колмогорова	173
4.4.3. Моделирование и идентификация законов распределения статистики типа Смирнова	181
4.4.4. Моделирование и идентификация законов распределения статистики Крамера-Мизеса-Смирнова	185
4.4.5. Моделирование и идентификация законов распределения статистики Андерсона-Дарлинга	189
4.4.6. Предельные распределения статистик для выборок из гамма-распределения	191
4.4.7. Предельные распределения статистик при оценивании по выборке параметров смеси распределений	193
4.5. Использование критериев согласия в случае частично группированных и интервальных наблюдений	196
4.6. Идентификация закона распределения случайной величины как многокритериальная задача	206
4.7. Выводы	209
5. ПРОГРАММНАЯ СИСТЕМА “СТАТИСТИЧЕСКИЙ АНАЛИЗ ОДНОМЕРНЫХ НАБЛЮДЕНИЙ СЛУЧАЙНЫХ ВЕЛИЧИН”	214
5.1. Назначение программной системы	214
5.2. Алгоритмы вычисления оценок максимального правдоподобия	216
5.3. Алгоритмическое обеспечение программной системы	218
5.3.1. Экспоненциальное распределение	218
5.3.2. Полунормальное распределение	220
5.3.3. Распределение Рэлея	222
5.3.4. Распределение Максвелла	222
5.3.5. Распределение модуля многомерной нормальной величины	224
5.3.6. Распределение Парето	226
5.3.7. Распределение Эрланга	227
5.3.8. Распределение Лапласа	229
5.3.9. Нормальное распределение	231
5.3.10. Логарифмически (ln) нормальное распределение	232
5.3.11. Логарифмически (lg) нормальное распределение	233

5.3.12. Распределение Коши.....	234
5.3.13. Логистическое распределение	237
5.3.14. Распределение Вейбулла	240
5.3.15. Распределение минимального значения	242
5.3.16. Распределение максимального значения	245
5.3.17. Двойное показательное распределение.....	247
5.3.18. Распределение Накагами	249
5.3.19. Гамма-распределение.....	254
5.3.20. Бета-распределение I-го рода.....	257
5.3.21. Стандартное бета-распределение II-го рода.....	262
5.3.22. Бета-распределение II-го рода	264
5.3.23. Распределение Sb -Джонсона.....	267
5.3.24. Распределение Sl -Джонсона.....	270
5.3.25. Распределение Su -Джонсона.....	273
5.3.26. Класс экспоненциальных распределений	278
5.4. Имитация выборочных значений	282
5.5. Статистический анализ смесей распределений по частично группированным данным	285
5.6. Статистический анализ независимых и зависимых случайных величин в геодезии	292
5.7. Выводы	300
ЗАКЛЮЧЕНИЕ	303
Список литературы	305
ПРИЛОЖЕНИЕ 1. Таблицы асимптотически оптимального группирования данных	322
ПРИЛОЖЕНИЕ 2. Таблицы коэффициентов оптимальных L -оценок для больших выборок	395
ПРИЛОЖЕНИЕ 3. Акты внедрения результатов	434

ВВЕДЕНИЕ

Современное состояние и актуальность темы исследований. Бурное развитие статистических методов в последние десятилетия, расширение круга специалистов, разрабатывающих статистические алгоритмы, и специалистов, широко использующих статистические методы в различных прикладных областях, количественный и качественный рост программного обеспечения задач статистического анализа определяют как неослабевающий интерес к тем возможностям, которые вытекают из получаемых результатов, так и вечную неудовлетворенность тем, что с использованием известных методов не удастся решить имеющуюся задачу, так как её условия не укладываются в ограничения, налагаемые этими методами.

Несмотря на определенную насыщенность рынка программными системами статистического анализа, реализуемые в них методы и алгоритмы отстают от последних теоретических достижений в области прикладной статистики. Имеющийся отрыв с одной стороны объясняется тем, что далеко не все результаты вошли в литературу справочного, нормативного и учебного характера и остались недоступными для разработчиков программного обеспечения. С другой стороны при создании программных систем пытаются охватить максимально большой круг решаемых задач статистического анализа. Это приводит к тому, что далеко не всегда к разработке программного обеспечения привлекается специалист высокого класса из соответствующей области статистики. Вследствие этого во многих разработках либо отсутствуют возможности, которые уже полностью обоснованы в специальной литературе, либо статистические решения принимаются в условиях, при которых, что тоже уже доказано, их выводы некорректны. В частности, трудно назвать примеры программных систем статистического анализа, где обеспечивается обработка группированных, частично группированных и цензурированных наблюдений. И совсем нетрудно указать примеры неправильного применения непараметрических критериев согласия.

Перспективы программного обеспечения по статистическому анализу данных, в том числе для персональных ЭВМ, обсуждались в работах С.А. Айвазяна [3, 4], современные проблемы внедрения прикладной статистики и статистических методов поднимались А.И. Орловым [125]. Отметим, что использование ЭВМ и их совершенствование отражается на развитии статистических методов [9, 133], взгля-

дах на точность вычисления оценок [108, 121], расширяет использование статистических методов в приложениях [160].

Форма представления исходных данных, часто определяемая условиями и точностью регистрации, неспособность существующих методов предложить приемлемое решение, тупиковые ситуации приводят к появлению новых методов анализа, подходов, моделей, иногда резкому отрицанию многого из ранее наработанного аппарата. Очевидно, что методы статистического анализа, алгоритмы, точность вычислений должны соответствовать форме представления и точности исходных данных. Методы статистического анализа группированных, частично группированных и интервальных наблюдений также требуют учета их специфики.

Вопросы оценивания параметров распределений, условия существования оценок максимального правдоподобия по группированным и частично группированным данным рассматривались Г. Куллдорфом [65], Н.А. Бодины [17, 18], В.П. Артамоновским [7, 8]. Вопросы оценивания параметров по случайно цензурированным наблюдениям, определения по таким выборкам параметров долговечности изделий рассматривались Ю.Н. Благовещенским [1, 11, 13-16, 105], по группированным и цензурированным выборкам М.С. Тиховым [161-169]. Статистическому анализу цензурированных и группированных данных посвящены работы [21, 30, 45, 53, 186, 187, 189, 192, 195, 196, 203, 204, 206, 211, 212, 214, 218, 220, 226, 228, 232, 234, 235, 236, 243, 245, 246, 249, 254, 257, 266-270, 276, 279]. С обработкой цензурированных и группированных данных в задачах надежности связаны работы [2, 12, 24, 58, 131, 132, 176, 229, 231, 237, 238, 239, 247, 248, 252, 271].

Трудности вычислительного характера, связанные с оцениванием параметров распределений непосредственно по группированным выборкам, заставляют заменять группированные наблюдения (при равных интервалах) негруппированными, применяя в дальнейшем к полученной оценке различные поправки на группировку [62, 242, 42, 179]. Поправки такого вида для многомерного случая исследуются в [268, 118].

Интерес к задачам, содержащим группированные и цензурированные наблюдения не снижается, так как появление таких выборок оказывается естественным и обычно порождается спецификой проведения экспериментов и условиями регистрации наблюдений.

Если в задачах, связанных с оценкой параметров распределений, группирование и цензурирование чаще всего является следствием того, как осуществляется регистрация данных, то в критериях согласия типа χ^2 группирование исходной (негруппированной) выборки - следствие используемой статистики.

Исследованием критериев типа χ^2 для непрерывных распределений при сложных гипотезах и оценивании параметров по негруппированным наблюдениям занимались W.G. Cochran [201, 202], Н. Chernoff и E.L. Lehman [199], G.S. Watson [273-275], Д.М. Чибисов [180], D.S. Moore [250], М.С. Никулин [107, 114, 115], М. Мирвалиев [107].

Очевидно, что качество статистических выводов зависит от того как группируются данные. Это относится и к оцениванию параметров распределений, и к критериям проверки гипотез. Разбиение на интервалы равной длины кочует по различным источникам, чаще всего используемых в качестве практических руководств. Разбиение на интервалы равных вероятностей предложили Н.В. Mann и А. Wald [244], и такой вариант группирования в качестве предпочтительного обычно рассматривается в работах теоретического характера. Но является ли он оптимальным с точки зрения качества статистических выводов?

Ясно, что группирование данных сопровождается потерей информации о наблюдаемом законе. И естественно требовать, чтобы эти потери были минимальны. Информационный подход рассматривается, например, в работах [153-155] при выборе оптимального размера интервалов группирования в случае равномерного квантования применительно к различным аспектам функционирования систем связи. В [265, 266, 235] анализируется величина потерь при группировании в информации Фишера. Задача асимптотически оптимального группирования при оценивании параметров распределений, в результате решения которой минимизируется асимптотическая дисперсия асимптотически эффективной оценки, впервые ставится и решается для частных случаев в [205, 65]. Некоторые вопросы, связанные с оптимальным группированием, рассматриваются в [195, 226, 257]. В [212, 46] подчеркивается, что построение оптимального разбиения является задачей планирования экспериментов, и решение ее не простое, так как количество потерянной информации для заданного разбиения является функцией неизвестного параметра θ .

Широкое применение на практике классических непараметрических критериев Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса [19] связано с именами А.Н. Колмогорова [59] и Н.В. Смирнова [143-145] и при проверке простых гипотез давно не вызывает затруднений. А вот возможность их применения для проверки сложных гипотез высветилась как серьезная проблема с появлением работы [233], в которой М. Кас, J. Kiefer и J. Wolfowitz показали, что предельные распределения этих статистик зависят от оцениваемых параметров и вида наблюдаемого закона. С тех пор к решению этой проблемы подходили с разных позиций. Например, Н.В. Lilliefors [240, 241] и М.А. Stephens [261-263] оценивали процентные точки для распределений статистик типа Колмогорова методами Монте-Карло. Ю.Н. Тюриным [171-173] предложены для этих же целей полученные приближенные зависимости. Г.В. Мартыновым [103] получен и протабулирован предельный закон распределения статистики ω^2 Мизеса для нормального закона. Отметим также связанные с проверкой гипотез работы [119, 128, 129]. Но острота проблемы не снята до настоящего времени. Более того, требуют разрешения вопросы применения критериев согласия при частично группированных и интервальных наблюдениях.

Этапы развития и разделы прикладной статистики отмечены степенью близости используемых моделей к реальной действительности. Например, в [156] неослабевающий интерес к непараметрической статистике объясняется тем, что она исходит из более широкой и реалистической модели. То же самое по отношению уже к интервальной статистике утверждается в работах [122, 123, 126].

Ограниченность круга параметрических моделей, используемых для описания наблюдаемых случайных величин в различных приложениях (порядка нескольких десятков законов распределения), невозможность с их помощью адекватно описать наблюдаемые данные является одним из узких мест прикладной параметрической статистики. К тому же в приложениях часто и почти всегда необоснованно стремятся еще более сузить это множество [124]. В известных программных системах количество включенных законов распределения обычно не превышает полутора десятков.

Параметрические модели становятся ближе к реальности, если учитывать, что наблюдаемые данные могут содержать частично группированные, группированные и интервальные наблюдения. Существенное расширение множества параметрических моделей можно по-

лучить, если соответствующее программное обеспечение позволит анализировать не только определенный набор семейств законов распределения, но и смеси усеченных и неусеченных законов.

Цель и задачи исследований. Основной целью диссертационной работы является дальнейшее развитие прикладных методов статистического анализа наблюдений одномерных непрерывных случайных величин, обеспечивающих качественные выводы при частично группированных, группированных и негруппированных наблюдениях, создание программного обеспечения, ориентированного на представление исходных данных в виде частично группированных (и интервальных) выборок и предназначенного для широкого использования.

Для достижения поставленной цели решаются следующие задачи:

- исследуются условия существования и единственности оценок максимального правдоподобия по частично группированным и группированным выборкам;
- исследуется влияние группирования на робастность оценок;
- исследуется влияние способов группирования на мощность критериев типа χ^2 ;
- исследуется влияние способов группирования на предельные распределения статистик типа χ^2 в случае простых и сложных гипотез;
- исследуются предельные распределения статистик типа Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса при проверке сложных гипотез;
- разрабатывается программное обеспечение задач статистического анализа частично группированных и интервальных наблюдений.

Методы исследования. Для решения поставленных задач использовался аппарат теории вероятностей, математической статистики, вычислительной математики, математического программирования, статистического моделирования.

Научная новизна диссертационной работы заключается в:

- в использовании асимптотически оптимального группирования данных в критериях типа χ^2 ;
- в использовании асимптотически оптимального группирования для построения оптимальных L-оценок параметров для больших выборок;
- в использовании группирования данных для вычисления робастных оценок;

- в построении законов распределения, аппроксимирующих предельные законы распределения статистик критериев типа Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса.

Основные положения, выносимые на защиту.

1. Условия существования и единственности оценок максимального правдоподобия по частично группированным выборкам.
2. Решение задачи асимптотически оптимального группирования наблюдений для совокупности непрерывных распределений, включенных в программную систему.
3. Оценки, использующие выборочные квантили, соответствующие оптимальному группированию. В том числе, оптимальные L-оценки параметров сдвига и масштаба для больших выборок.
4. Применение асимптотически оптимального группирования данных в критериях согласия отношения правдоподобия, χ^2 Пирсона и типа χ^2 , обеспечивающее максимальную мощность этих критериев при близких конкурирующих гипотезах.
5. Применение группирования наблюдений для построения робастных оценок максимального правдоподобия и реализация эффективной параметрической процедуры отбраковки аномальных измерений.
6. Результаты исследований зависимости предельных распределений статистик критериев отношения правдоподобия и χ^2 Пирсона от способа группирования.
7. Результаты исследований предельных распределений статистик критериев типа Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса при проверке сложных гипотез и вычислении оценок параметров по негруппированным данным.
8. Программная система статистического анализа одномерных наблюдений случайных величин, в которой реализованы все полученные результаты исследований по обработке частично группированных, группированных и негруппированных данных.

Обоснованность и достоверность научных положений, выводов и рекомендаций обеспечивается:

- применением аналитических методов исследования свойств оценок и критериев;
- подтверждением аналитических выводов и рекомендаций результатами статистического моделирования.

Практическая ценность и реализация результатов. Сформированные таблицы асимптотически оптимального группирования исполь-

зуются в задачах оценивания параметров распределений и проверки статистических гипотез. Использование полученных таблиц оптимального группирования в критериях согласия типа χ^2 обеспечивает их максимальную мощность при близких альтернативах. Полученные таблицы коэффициентов для оптимальных L-оценок для больших выборок в совокупности с таблицами оптимального группирования позволяют просто вычислять оценки по выборочным квантилям. Полученные предельные распределения статистик непараметрических критериев согласия типа Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса (при проверке сложных гипотез) и соответствующие таблицы процентных точек обеспечивают корректность применения этих критериев.

Все основные результаты реализованы в программной системе “Статистический анализ одномерных наблюдений случайных величин”, которая используется в учебном процессе студентов факультета прикладной математики и информатики НГТУ в дисциплинах “Исследование операций”, “Методы моделирования в экономике” и др., в научных исследованиях, при обработке данных различных приложений и вместе с системой “Корреляционный анализ многомерных наблюдений случайных величин” передана в ряд организаций.

Апробация работы. Результаты исследований докладывались на Межреспубликанском семинаре “Исследования и расчеты надежности энергосистем на этапах проектирования и эксплуатации” (Фрунзе, 1978); V Всесоюзном симпозиуме “Эффективность, качество и надежность “систем человек-техника”” (Ленинград, 1978); Международной конференции “Идентификация, измерение характеристик и имитация случайных сигналов” (Новосибирск, 1994); Российской конференции “Повышение эффективности производства и использования энергии в условиях Сибири” (Иркутск, 1994); Международной конференции “Актуальные проблемы электронного приборостроения АПЭП-94” (Новосибирск, 1994); Российской НТК “Информатика и проблемы телекоммуникаций” (Новосибирск, 1994, 1996); Международной НТК “Информатика и проблемы телекоммуникаций” (Новосибирск, 1995, 1997); VIII Международном симпозиуме по непараметрическим и робастным методам в кибернетике (Красноярск, 1995); Международной конференции “Актуальные проблемы электронного приборостроения АПЭП-96” (Новосибирск, 1996); III международной научно-технической конференции “Микропроцессорные системы автоматики” (Новосибирск, 1996); Втором Сибир-

ском Конгрессе по Прикладной и Индустриальной Математике ИН-ПРИМ-96 (Новосибирск, 1996); Fifth Conference of International Federation of Classification Societies IFCS-96 "Data Scitnce, Classification and Related Methods" (Japan, Kobe, 1996); Межреспубликанском совещании по интервальному анализу (Новосибирск, 1996); Международной конференции "Информационные технологии в моделировании и управлении" (С.-Петербург, 1996); Международной конференции "Новые информационные технологии в университетском образовании" (Новосибирск, 1997).

Пакет программ по обработке группированных наблюдений демонстрировался на ВДНХ СССР (Москва, 1986). Программная система "Статистический анализ одномерных наблюдений случайных величин" демонстрировалась на выставке "Научно-технические достижения и интеллектуальная собственность высшей школы" (Москва, 1994), на выставке "Наука и образование Сибири - 95" (Новосибирск, 1995).

Публикации. Основные результаты исследований по теме диссертации опубликованы в 52 печатных работах и 5 зарегистрированных отчетах по НИР.

Структура работы. Диссертация состоит из введения, 5 глав основного содержания, заключения, списка литературы, приложений.

Краткое содержание работы.

В первой главе диссертации рассматривается структура представления наблюдений случайных величин с наиболее общих позиций в форме частично группированных и интервальных выборок. Рассматриваются условия существования и единственности оценок максимального правдоподобия параметров распределений непрерывных случайных величин по частично группированным выборкам. Рассматриваются подходы к вычислению оценок параметров по интервальным наблюдениям.

Во второй главе решается задача асимптотически оптимального группирования, при которой минимизируется асимптотическая дисперсия оценок по группированным данным (минимизируются потери в информации Фишера, связанные с группированием наблюдений), строятся таблицы асимптотически оптимального группирования. Рассматриваются возможности использования асимптотически оптимального группирования в задачах оценивания параметров. Предлагаются оценки, использующие выборочные квантили, соответствующие асимптотически оптимальному группированию. Предлагаются

оптимальные L-оценки параметров сдвига и масштаба, строятся таблицы коэффициентов для этих оценок.

В третьей главе диссертации рассматриваются подходы к вычислению робастных оценок параметров. На основании анализа функций влияния и экспериментально исследуется робастность оценок максимального правдоподобия по негруппированным и группированным данным. Предлагается в случае необходимости проводить предварительное группирование данных для получения робастных оценок. Показывается, что совокупность применения критериев, способных обнаружить такие отклонения в выборке, как наличие аномальных наблюдений, и вычисления робастных оценок по группированным наблюдениям позволяют реализовать эффективный параметрический алгоритм отбраковки.

В четвертой главе показывается, что при минимизации потерь в информации Фишера, связанных с группированием данных, обеспечивается максимальная мощность критериев типа χ^2 при близких конкурирующих гипотезах. Методами статистического моделирования исследуются предельные распределения статистик типа χ^2 в зависимости от способа группирования данных при простых и сложных гипотезах (при оценивании параметров по негруппированным наблюдениям). Методами статистического моделирования исследуются предельные распределения статистик типа Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса при потере ими свойства “свободы от распределения”. Приводятся полученные законы распределения, аппроксимирующие предельные, и таблицы верхних процентных точек. Рассматривается применение непараметрических критериев согласия в случае частично группированных и интервальных наблюдений. Задача идентификации закона распределения, наиболее хорошо описывающего наблюдаемые данные, рассматривается как многокритериальная задача принятия решения.

В пятой главе диссертации рассматривается алгоритмическое обеспечение разработанной программной системы статистического анализа данных, все задачи которой рассматриваются с точки зрения наиболее общего представления экспериментальных наблюдений в виде частично группированных выборок. Приводятся соотношения, используемые при вычислении оценок максимального правдоподобия по частично группированным наблюдениям. Рассматривается расширение множества моделей для описания реальных наблюдений за счет использования смесей усеченных и неусеченных законов распределе-

ний, возможность обработки интервальных наблюдений, являющиеся отличительной особенностью разрабатываемой версии программной системы. Приводятся результаты анализа выборок случайных величин, возникающих при обработке различных геодезических измерений.

В приложении П1 приведены полученные таблицы асимптотически оптимального группирования, в приложении П2 - таблицы коэффициентов для оптимальных L-оценок параметров сдвига и масштаба, в приложении П3 - акты внедрения результатов.

1. СТРУКТУРА ПРЕДСТАВЛЕНИЯ ДАННЫХ, МЕТОДЫ ОЦЕНИВАНИЯ И УСЛОВИЯ СУЩЕСТВОВАНИЯ ОЦЕНОК

1.1. Структура представления данных

В настоящей работе анализируемые наблюдения рассматриваются (в основном) с позиций наиболее общего представления в форме *частично группированных выборок*. Понятие частично группированной выборки объединяет практически все возможные типы выборок случайных величин. Выборка может быть *негруппированной*, т.е. состоять только из известных индивидуальных измерений, *группированной*, когда индивидуальные значения измерений неизвестны, а зафиксированы лишь границы интервалов и количества наблюдений, попавших в каждый интервал, или *частично группированной*. В последнем случае часть индивидуальных наблюдений известна, а для других отмечен лишь факт попадания наблюдения в соответствующий интервал значений.

Простейшими случаями частично группированных выборок являются *цензурированные* выборки, когда имеется один или два крайних интервала группирования.

В частности, большинство измерений, проводимых с помощью различных приборов, представляют собой, вообще говоря, группированные или поразрядно группированные наблюдения, что определяется конечностью цены деления прибора.

Следуя [65], введём определения.

Определение 1. Выборка называется негруппированной, если выборочные значения представляют собой индивидуальные значения наблюдений из области определения случайной величины:

$$X_1, X_2, \dots, X_N,$$

где N - объем выборки.

Определение 2. Выборка называется группированной, если область определения случайной величины разбита на k непересекающихся интервалов граничными точками

$$x_{(0)} < x_{(1)} < \dots < x_{(k-1)} < x_{(k)},$$

где $x_{(0)}$ - нижняя грань области определения случайной величины X , $x_{(k)}$ - верхняя грань области определения случайной величины X , и зафиксированы количества наблюдений n_i , попавших в i -й интервал значений. Объем выборки

$$N = \sum_{i=1}^k n_i.$$

Определение 3. Выборка называется частично группированной, если имеющаяся в нашем распоряжении информация связана с множеством непересекающихся интервалов, которые делят область определения случайной величины так, что каждый интервал принадлежит к одному из двух типов:

- а) i -ый интервал принадлежит к первому типу, если число n_i известно, но индивидуальные значения $x_{ij}, j = \overline{1, n_i}$, неизвестны;
- б) i -й интервал принадлежит ко второму типу, если известно не только число n_i , но и все индивидуальные значения $x_{ij}, j = \overline{1, n_i}$.

В последующем, когда речь будет идти о частично группированной выборке, суммирование по интервалам первого и второго типов будем обозначать соответственно через $(\sum_{(1)})$ и $(\sum_{(2)})$.

Дальнейшим обобщением частично группированной выборки является *интервальная* выборка, в которой каждое наблюдение представлено интервалом $[a_i, b_i]$, которому принадлежит неизвестное точно значение X_i . Классификация одномерных выборок показана на рис. 1.1.1. Группированная выборка задана непересекающимися интервалами, негруппированная - вырожденными интервалами, у которых $a_i = b_i$. Интервальное представление наблюдения можно интерпретировать как неточное измерение случайной величины, связанное либо с заведомо известной погрешностью измерительного прибора, либо с особенностями измеряемой величины. Интервальные наблюдения можно получить также в результате обработки измерений, сопровождающихся усечением части данных (например, при группировании или цензурировании). Приведенная на рис. 1.1.1. классификация не может претендовать на полноту, но охватывает достаточно широкий спектр вариантов регистрации наблюдаемых величин.

Естественно, что форма представления наблюдений (регистрации) случайных величин отражается на используемых методах и алгоритмах статистического анализа: методах оценивания параметров законов распределения и задачах проверки статистических гипотез.

Основное содержание данной работы касается вопросов обработки частично группированных наблюдений. Именно с этим связаны

полученные теоретические и практические результаты, реализованные в разработанной программной системе [76]. Вопросы анализа интервальных наблюдений только исследуются. Разрабатываемая новая объектно-ориентированная версия программной системы должна обеспечивать, в том числе, и обработку интервальных наблюдений [117,90]. Подробная библиография работ, связанных с вопросами интервальной статистики, приведена в [126].

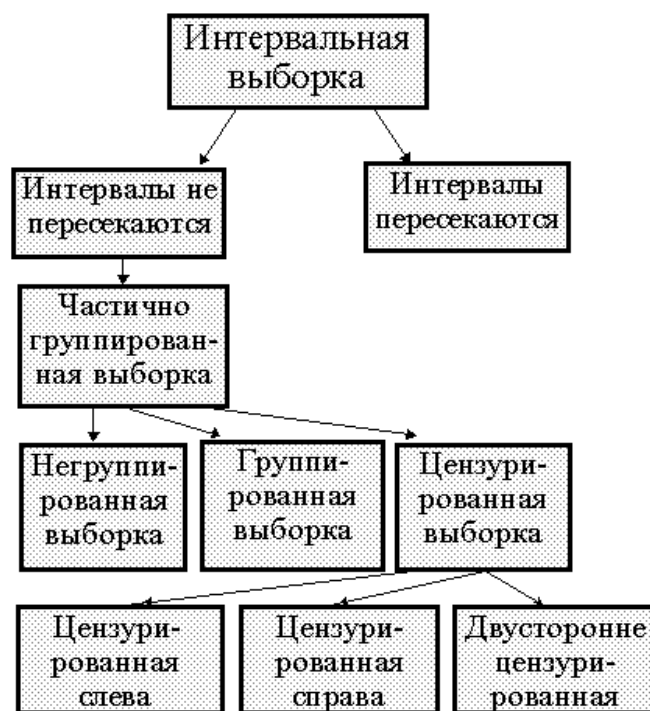


Рис. 1.1.1. Классификация выборочных наблюдений

Понятно, что не только форма представления наблюдений, обычно связанная с технологией их регистрации, влияет на методы обработки, но и методы анализа и свойства используемых статистических процедур иногда определяют преобразование исходных данных. Например, наблюдения группируют при применении критериев согласия χ^2 Пирсона и отношения правдоподобия. К группированию можно прибегать с целью получения робастных оценок, как это будет показано в дальнейшем.

В работе рассматриваются два вида задач математической статистики применительно к анализу одномерных непрерывных случайных величин: задачи оценивания параметров распределений и задачи проверки гипотез о согласии наблюдаемых данных с теоретическим законом распределения.

1.2. Виды оценок и методы оценивания

Среди всего множества различных оценок можно выделить три основных класса оценок [184]. Это М-оценки, L-оценки и R-оценки.

Пусть $X_1, X_2, \dots, X_N$ последовательность независимых одинаково распределенных случайных величин с функцией распределения $F(x, \theta)$ и функцией плотности $f(x, \theta)$.

В классе М-оценок искомая оценка $\hat{\theta}_N$ определяется как решение задачи вида

$$\min_{\theta_N} \sum_{i=1}^N \rho(X_i, \theta_N),$$

или как решение уравнения

$$\sum_{i=1}^N \frac{\partial \rho(X_i, \theta_N)}{\partial \theta_N} = 0.$$

Начало исследований свойств этого класса оценок связывают с работой [230]. Асимптотические свойства М-оценок и их свойства при конечных объемах выборок достаточно полно изложены в [178]. К М-оценкам относятся, например, оценки максимального правдоподобия и наименьших квадратов. Основное содержание данной работы связано с оценками максимального правдоподобия. В случае негруппированных данных для метода максимального правдоподобия $\rho(x, \theta) = -\ln f(x, \theta)$.

L-оценки формируются как линейные комбинации порядковых статистик

$$\tilde{\theta}_N = \sum_{i=1}^N a_{Ni} X_{(i)},$$

где $X_{(i)}$ - i -я порядковая статистика. Оптимальные L-оценки для случая больших выборок, обладающие хорошими свойствами робастности, рассматриваются во втором разделе работы.

R-оценки, основанные на использовании ранговых критериев, в данной работе не рассматриваются.

Еще один класс оценок образуют методы минимизирующие расстояния (MD-оценки) [277]. Если в пространстве функций распределения для двух функций распределения F и G ввести расстояние $d(F, G)$, и если случайная величина принадлежит параметрическому семейству с функцией распределения $F(x, \theta)$, то MD-оценка определяется выражением

$$\hat{\theta}_N = \arg \min_{\theta} d(F_N(x), F(x, \theta)),$$

где $F_N(x)$ - эмпирическая функция распределения. Вообще говоря, любая статистика, из используемых в критериях согласия, где различным образом измеряются расстояния между законами распределения, может быть положена в основу метода оценивания. Интерес к MD-оценкам вызывается тем, что они отличаются хорошими робастными свойствами [184]. Свойства MD-оценок исследовались в [194, 255]. В программной системе [76] вычисляются в основном оценки максимального правдоподобия, но в разрабатываемой версии [117, 90] строятся и MD-оценки.

Определяющими факторами при выборе метода оценивания являются структура представления наблюдаемых данных и качество оценок.

Качество оценок определяется такими свойствами, как состоятельность и асимптотическая эффективность.

Определение 4. Оценка неизвестного параметра $\hat{\theta}_N$ называется состоятельной, если $\forall \delta > 0, \lim_{N \rightarrow \infty} P[|\hat{\theta}_N - \theta| < \delta] = 1, |\hat{\theta} - \theta| = \max_{1 \leq j \leq m} |\hat{\theta}_j - \theta_j|$.

Определение 5. Оценка $\hat{\theta}_N$ называется асимптотически эффективной, если а) $\lim_{N \rightarrow \infty} E[\sqrt{N}(\hat{\theta}_N - \theta)] = 0$;

б) $\lim_{N \rightarrow \infty} E\{[\sqrt{N}(\hat{\theta}_N - \theta)]^T [\sqrt{N}(\hat{\theta}_N - \theta)]\} = M^{-1}$,

где $M = E\left[\frac{\partial \ln L(\theta)}{\partial \theta_s} \frac{\partial \ln L(\theta)}{\partial \theta_l}\right]$ - информационная матрица Фишера;

с) векторная случайная величина $\sqrt{N}(\hat{\theta}_N - \theta)$ при $N \rightarrow \infty$ распределена по нормальному закону с нулевым вектором математического ожидания и ковариационной матрицей, равной M^{-1} .

Рассмотрим некоторые наиболее часто используемые на практике методы вычисления оценок.

Метод моментов. Предполагается, что исходная выборка негруппирована. Если она группирована, то всем наблюдениям, попавшим в интервал, присваивают значения, равные, например, середине интервала, а затем вычисляют выборочные значения моментов распределения. Пусть первые r моментов распределения существуют и явно выражаются функциями $E[X^i] = m_i(\theta_1, \theta_2, \dots, \theta_r), i = \overline{1, r}, r$

неизвестных параметров, где $E[\cdot]$ - оператор математического ожидания. Выборочные значения моментов вычисляются по формулам

$$m_i^* = \frac{1}{N} \sum_{j=1}^N X_j^i, i = \overline{1, r}.$$

Тогда оценки параметров $\theta_1, \theta_2, \dots, \theta_r$ по методу моментов определяются как решение системы уравнений

$$m_i(\theta_1, \theta_2, \dots, \theta_r) = m_i^*, i = \overline{1, r}.$$

Метод моментов приводит при определённых условиях к состоятельным оценкам. Однако он неприменим, когда теоретические моменты нужного порядка не существуют. Кроме того, эти оценки очень чувствительны к наличию аномальных наблюдений и не всегда эффективны [138]. Если исходная выборка является группированной, то процедура присваивания всем наблюдениям в группе одинаковых значений представляет собой приближение, которое приводит к значительным систематическим ошибкам и требует поправок. Например, если интервалы равны по длине, зачастую используют поправки Шеппарда для моментов, определяемые соотношениями [62]

$$\begin{aligned} m_1 &= m_1^*, \\ m_2 &= m_2^* - \frac{1}{12} h^2, \\ m_3 &= m_3^* - \frac{1}{4} m_1^* h^2, \\ m_4 &= m_4^* - \frac{1}{2} m_2^* h^2 + \frac{7}{240} h^4, \\ &\dots, \end{aligned}$$

где h - длина интервала. Введение поправок не всегда приводит к удовлетворительным результатам. Иногда оценка, полученная с применением поправки, оказывается дальше от истинного значения, чем оценка без поправки. Особенно неудовлетворительные результаты получаются, когда мало число групп, т.е. происходит грубое группирование, или область определения случайной величины разбита на интервалы неравной длины.

Вообще говоря, оценки параметров, получаемые по методу моментов, рационально использовать в качестве начального приближения при поиске оценок более эффективными методами.

Отметим, что обобщение поправок Шеппарда для моментов и случая многомерного распределения получено в [118].

Целый ряд методов предполагает использование группированных выборок.

Метод минимума χ^2 [138]. Здесь предполагается, что выборка, по которой оцениваются параметры, полностью группирована. В данном методе оценка определяется как значение параметра, минимизирующее статистику

$$\chi^2 = N \sum_{i=1}^k \frac{(n_i / N - P_i(\theta))^2}{P_i(\theta)},$$

где $P_i(\theta)$ - вероятность попадания наблюдения в i -ый интервал.

Модифицированный метод минимума χ^2 [138]. При вычислении оценки минимизируется статистика

$$\text{mod} \chi^2 = N \sum_{i=1}^k \frac{(n_i - NP_i(\theta))^2}{n_i},$$

где n_i заменяется на 1, если $n_i = 0$.

Расстояние Хеллингера [138]. В данном случае при поиске оценки минимизируется статистика

$$H_D = \arccos \sum_{i=1}^k \sqrt{(n_i / N) P_i(\theta)}.$$

Дивергенция Кульбака-Лейблера [138]. Минимизируемая статистика имеет вид

$$S_{KL} = \sum_{i=1}^k P_i(\theta) \ln [P_i(\theta) / (n_i / N)].$$

Мера расхождения Холдейна [138]. Оценки находятся в результате минимизации статистики

$$D_j = \frac{(N+j)!}{N!} \sum_{i=1}^k \frac{n_i! P_i^{j+1}(\theta)}{(n_i+j)!}, \quad j \neq -1,$$

$$D_{-1} = \frac{1}{N} \sum_{i=1}^k n_i \ln P_i(\theta).$$

Все эти методы при соответствующих условиях регулярности, как показал Рао, дают состоятельные и асимптотически эффективные оценки [138]. Однако имеются и различия между этими методами, возникающие при учете введенной Рао эффективности второго порядка [258]. Им показано, что асимптотическая дисперсия оценки определяется соотношением

$$D(\theta) = \frac{1}{NJ(\theta)} + \frac{\psi(\theta)}{N^2} + o\left(\frac{1}{N^2}\right),$$

где $J(\theta)$ - информационное количество Фишера о параметре, а величина $\psi(\theta)$ определяется методом оценивания и неотрицательна. Если обозначить $\psi(\theta)$ для метода максимального правдоподобия, описанного в следующем пункте, как ψ_M , то для метода минимума χ^2 $\psi(\theta) = \psi_M + \delta$, где δ неотрицательная величина, равная нулю только в специальных случаях, для модифицированного χ^2 $\psi(\theta) = \psi_M + 4\delta$, для расстояния Хеллингера $\psi(\theta) = \psi_M + \delta/4$, для дивергенции Кульбака-Лейблера $\psi(\theta) = \psi_M + \delta$, для меры расхождения Холдейна $\psi(\theta) = \psi_M + \delta(J+1)^2$. Следовательно, метод максимального правдоподобия при учёте эффективности второго порядка является наилучшим.

Кроме того, метод является наиболее универсальным по отношению к форме представления выборочных данных (структуре выборки), по которым оцениваются параметры. Метод моментов требует преобразования группированных данных к негруппированным, только после чего оцениваются параметры с использованием при необходимости (или возможности) поправок на группирование. Напротив, метод минимума χ^2 и родственные с ним используют только группированные данные: если в распоряжении исследователя имеются индивидуальные наблюдения, выборку следует преобразовывать в полностью группированную. Метод максимального правдоподобия в отличие от других позволяет определять оценки максимального правдоподобия (ОМП) параметров по негруппированным, частично группированным и группированным данным, т.е. дает возможность исследователю самому определять, в каком виде регистрировать и в каком виде хранить экспериментальную информацию в зависимости от характеристик приборов, регистрирующих наблюдения, и объема экспериментальной информации.

Метод максимального правдоподобия. Оценкой максимального правдоподобия неизвестного параметра по группированным наблюдениям называется такое значение параметра, при котором функция правдоподобия

$$L(\theta) = \gamma \prod_{i=1}^k P_i^{n_i}(\theta), \quad (1.2.1)$$

где γ - некоторая константа и $P_i(\theta) = \int_{x_{(i-1)}}^{x_{(i)}} f(x, \theta) dx$ - вероятность попадания наблюдения в i -й интервал значений, достигает максимума на множестве возможных значений параметра. Здесь предпола-

гается, что для всех $\theta \in \Omega$ $P_i(\theta) > 0, i = \overline{1, k}$. Для вычисления ОМП дифференцируют функцию правдоподобия по θ и, приравнявая производные нулю, получают систему уравнений правдоподобия

$$\sum_{(1)} n_i \frac{\partial \ln P_i(\theta)}{\partial \theta_l} = 0, l = \overline{1, m},$$

где m - размерность вектора параметров $\theta = (\theta_1, \theta_2, \dots, \theta_m)^T$.

Функция правдоподобия для частично группированной выборки имеет вид

$$L(\theta) = \prod_{(1)} P_i^{n_i}(\theta) \prod_{(2)} \prod_{j=1}^{n_i} f(x_{ij}, \theta),$$

система уравнений правдоподобия

$$\sum_{(1)} n_i \frac{\partial \ln P_i(\theta)}{\partial \theta_l} + \sum_{(2)} \sum_{j=1}^{n_i} \frac{\partial \ln f(x_{ij}, \theta)}{\partial \theta_l} = 0, l = \overline{1, m},$$

где $f(x, \theta)$ - функция плотности случайной величины, (1) и (2) означают, что суммирование и умножение осуществляются по интервалам с группированными и негруппированными данными соответственно.

В случае достаточно больших N при определенных условиях регулярности для функции плотности ОМП существует практически всегда, состоятельна и асимптотически эффективна [138]. Условия существования, асимптотической эффективности и состоятельности по группированным и цензурированным выборкам рассматривались в работах Г. Куллдорфа [65] и Н.А. Бодина [17, 18]. Г.Г. Зачепой [48] были получены условия существования ОМП основных параметров распределения Вейбулла и гамма-распределения. В наших работах [30, 69, 70, 208] получены условия существования и единственности ОМП для параметров ряда непрерывных законов распределения случайных величин, а в [42] приведены сводные результаты. Кратко эти результаты изложены в следующем параграфе. Следует отметить, что условия существования и единственности ОМП параметров сдвига и масштаба по группированным выборкам рассматривались в работе [8].

Вообще говоря, метод максимального правдоподобия требует значительного объема вычислений. А в случае группированных или частично группированных данных возникает необходимость в решении задач численного интегрирования, в том числе, и вычисления несобственных интегралов. Именно трудности вычислительного характера, особенно в ситуации группированных и частично группирован-

ных данных, ограничивали использование метода максимального правдоподобия.

Существует большое число работ, например [224, 242, 269], в которых рассматривается вычисление приближенных оценок максимального правдоподобия. В этом случае исходная группированная выборка заменяется негруппированной, в которой индивидуальным значениям присваиваются значения центров интервалов группирования при их равной длине. Далее вычисляются оценки, а затем выводятся выражения для поправок к полученным оценкам. Для ряда законов распределения выражения для оценок и соответствующих поправок получены в [42, 179].

Описанные в данном разделе методы вычисления оценок параметров распределений далеко не представляют собой полный перечень всех возможных методов, да эта цель и не преследовалась.

1.3. Условия существования ОМП по частично группированным данным

При решении уравнений правдоподобия по частично группированным выборкам, особенно по группированным, приходится останавливаться на вопросах существования решения уравнения правдоподобия и его единственности. Проверка условий существования и единственности позволяет отказаться от порой трудоемкого и бесполезного процесса вычисления ОМП, если она не существует. При экспериментальных исследованиях, например, надежности, знание условий существования и единственности даёт возможность принять решение о прекращении эксперимента или его продолжении, если по полученным данным нельзя найти оценку параметра θ исследуемого распределения.

Ниже рассматриваются условия существования и единственности ОМП по частично группированным выборкам для скалярных параметров, наиболее часто встречающихся в приложениях одномерных непрерывных распределений. Все рассматриваемые законы включены в программную систему [76].

Функция плотности *распределения Рэлея* имеет вид

$$f(x, \theta) = \frac{x}{\theta^2} \exp \left\{ -\frac{x^2}{2\theta^2} \right\} \quad (1.3.1)$$

на множестве $X = (0, \infty)$, параметр $\theta \in \Omega = (0, \infty)$, $x_{(0)} = 0$, $x_{(k)} = \infty$.

Вероятность попадания наблюдения i -й интервал

$$P_i(\theta) = \exp\{-t_{i-1}^2/2\} - \exp\{-t_i^2/2\},$$

где $t_i = x_{(i)}/\theta$.

Теорема 1.1. *ОМП параметра θ распределения Рэлея по частично группированной выборке существует при условии*

$$\sum_{(2)} n_i > 0 \text{ или } n_1 < N \text{ и } n_k < N \quad (1.3.2)$$

и получается в качестве единственного решения уравнения правдоподобия

$$\frac{1}{\theta} \left\{ \sum_{(1)} n_i \frac{t_{i-1}^2 e^{-t_{i-1}^2/2} - t_i^2 e^{-t_i^2/2}}{e^{-t_{i-1}^2/2} - e^{-t_i^2/2}} + \sum_{(2)} \sum_{j=1}^{n_i} (t_{ij}^2 - 2) \right\} = 0,$$

где $t_{ij} = x_{ij}/\theta$.

Для доказательства существования достаточно показать, что при условии непрерывности первой и второй производной от логарифма функции правдоподобия первая производная на области определения параметра меняется от некоторого положительного значения до некоторого отрицательного. При этом ОМП будет единственна, если для любого $\hat{\theta}$, являющегося решением уравнения правдоподобия, вторая производная будет отрицательна.

Доказательство. 1. Покажем, что функции $\ln P_i(\theta)$ и $\ln f(\theta)$ дважды дифференцируемы по θ :

$$\begin{aligned} \frac{\partial}{\partial \theta} \ln P_i(\theta) &= \frac{1}{\theta} \frac{t_{i-1}^2 e^{-t_{i-1}^2/2} - t_i^2 e^{-t_i^2/2}}{e^{-t_{i-1}^2/2} - e^{-t_i^2/2}}, \\ \frac{\partial^2}{\partial \theta^2} \ln P_i(\theta) &= \frac{1}{\theta^2} \frac{1}{e^{-t_{i-1}^2/2} - e^{-t_i^2/2}} \left[t_{i-1}^2 e^{-t_{i-1}^2/2} (t_{i-1}^2 - 3) - t_i^2 e^{-t_i^2/2} (t_i^2 - 3) - \right. \\ &\quad \left. - \frac{(t_{i-1}^2 e^{-t_{i-1}^2/2} - t_i^2 e^{-t_i^2/2})^2}{e^{-t_{i-1}^2/2} - e^{-t_i^2/2}} \right], \end{aligned} \quad (1.3.3)$$

$$\frac{\partial}{\partial \theta} \ln f(\theta) = \frac{1}{\theta} (t_{ij}^2 - 2), \quad \frac{\partial^2}{\partial \theta^2} \ln f(\theta) = \frac{1}{\theta^2} (2 - 3t_{ij}^2).$$

2. Так как $\theta > 0$, то решение уравнения (1.3.2) эквивалентно решению уравнения

$$\left\{ \sum_{(1)} n_i \frac{t_{i-1}^2 e^{-t_{i-1}^2/2} - t_i^2 e^{-t_i^2/2}}{e^{-t_{i-1}^2/2} - e^{-t_i^2/2}} + \sum_{(2)} \sum_{j=1}^{n_i} (t_{ij}^2 - 2) \right\} = 0. \quad (1.3.4)$$

Рассмотрим поведение $\frac{\partial}{\partial \theta} \ln P_i(\theta)$ на области определения параметра. Применяя теорему Коши, имеем

$$\frac{t_{i-1}^2 e^{-t_{i-1}^2/2} - t_i^2 e^{-t_i^2/2}}{e^{-t_{i-1}^2/2} - e^{-t_i^2/2}} = \tau_i^2 - 2,$$

где $t_{i-1} < \tau_i < t_i$, $i = \overline{2, (k-1)}$. Отметим, что при $\theta \rightarrow 0$ $t_i \rightarrow +\infty$. Отсюда для i -го интервала, $i = \overline{2, (k-1)}$, принадлежащего первому типу, получаем

$$\lim_{\theta \rightarrow 0} \theta \frac{\partial}{\partial \theta} \ln P_i(\theta) = \lim_{t_i \rightarrow +\infty} (\tau_i^2 - 2) = +\infty. \quad (1.3.5)$$

Для $i = 1$, применяя правило Лопиталя,

$$\lim_{\theta \rightarrow 0} \theta \frac{\partial}{\partial \theta} \ln P_1(\theta) = \lim_{t_1 \rightarrow +\infty} \frac{-t_1^2}{e^{-t_1^2/2} - 1} = 0. \quad (1.3.6)$$

Для $i = k$

$$\lim_{\theta \rightarrow 0} \theta \frac{\partial}{\partial \theta} \ln P_k(\theta) = \lim_{t_{k-1} \rightarrow +\infty} t_{k-1}^2 = +\infty. \quad (1.3.7)$$

Для всех интервалов второго типа имеем

$$\lim_{t_{ij} \rightarrow +\infty} \sum_{(2)} \sum_{j=1}^{n_i} (t_{ij}^2 - 2) = +\infty. \quad (1.3.8)$$

Из (1.3.5-1.3.8) очевидно, что при условии $n_1 < N$, если 1-й интервал относится к первому типу, найдется некоторое $\underline{\theta}$, для которого выполняется неравенство

$$\theta \left\{ \sum_{(1)} n_i \frac{\partial \ln P_i(\theta)}{\partial \theta} + \sum_{(2)} \sum_{j=1}^{n_i} \frac{\partial \ln f(\theta)}{\partial \theta} \right\} \bigg|_{\theta = \underline{\theta}} > 0. \quad (1.3.9)$$

Далее при $\theta \rightarrow +\infty$ для i -го интервала, $i = \overline{2, (k-1)}$, принадлежащего первому типу, имеем

$$\lim_{\theta \rightarrow +\infty} \theta \frac{\partial}{\partial \theta} \ln P_i(\theta) = \lim_{t_i \rightarrow 0} (\tau_i^2 - 2) = -2. \quad (1.3.10)$$

Для $i = 1$, применяя правило Лопиталя,

$$\lim_{\theta \rightarrow +\infty} \theta \frac{\partial}{\partial \theta} \ln P_1(\theta) = \lim_{t_1 \rightarrow 0} \frac{-t_1^2}{e^{-t_1^2/2} - 1} = -2. \quad (1.3.11)$$

Для $i = k$

$$\lim_{\theta \rightarrow +\infty} \theta \frac{\partial}{\partial \theta} \ln P_k(\theta) = \lim_{t_{k-1} \rightarrow 0} t_{k-1}^2 = 0. \quad (1.3.12)$$

Для всех интервалов второго типа имеем

$$\lim_{t_{ij} \rightarrow 0} \sum_{(2)} \sum_{j=1}^{n_i} (t_{ij}^2 - 2) = -2 \sum_{(2)} n_i. \quad (1.3.13)$$

Из (1.3.10-1.3.13) следует, что при условии $n_k < N$, если k -й интервал относится к первому типу, найдется некоторое $\bar{\theta}$, для которого выполняется неравенство

$$\theta \left\{ \sum_{(1)} n_i \frac{\partial \ln P_i(\theta)}{\partial \theta} + \sum_{(2)} \sum_{j=1}^{n_i} \frac{\partial \ln f(\theta)}{\partial \theta} \right\} \left(\theta = \bar{\theta} \right) < 0. \quad (1.3.14)$$

Из (1.3.9) и (1.3.14) следует существование ОМП параметра θ распределения Рэлея.

3. Для доказательства единственности ОМП достаточно показать, что

$$\left\{ \sum_{(1)} n_i \frac{\partial^2 \ln P_i(\theta)}{\partial \theta^2} + \sum_{(2)} \sum_{j=1}^{n_i} \frac{\partial^2 \ln f(\theta)}{\partial \theta^2} \right\} \left(\theta = \check{\theta} \right) < 0, \quad (1.3.15)$$

где $\check{\theta}$ является решением уравнения правдоподобия (1.3.2).

Рассмотрим $\frac{\partial^2 \ln L}{\partial \theta^2}$. Из (1.3.3) имеем

$$\frac{\partial^2}{\partial \theta^2} \ln P_i(\theta) = \frac{1}{\theta^2} \frac{A_i}{e^{-t_{i-1}^2/2} - e^{-t_i^2/2}},$$

где

$$A_i = t_{i-1}^2 e^{-t_{i-1}^2/2} (t_{i-1}^2 - 3) - t_i^2 e^{-t_i^2/2} (t_i^2 - 3) - \frac{(t_{i-1}^2 e^{-t_{i-1}^2/2} - t_i^2 e^{-t_i^2/2})^2}{e^{-t_{i-1}^2/2} - e^{-t_i^2/2}}.$$

Применяя теорему Коши, получаем для всех i

$$\begin{aligned} A_i &= t_{i-1}^2 e^{-t_{i-1}^2/2} (t_{i-1}^2 - 3) - t_i^2 e^{-t_i^2/2} (t_i^2 - 3) - (d_i^2 - 2)(t_{i-1}^2 e^{-t_{i-1}^2/2} - t_i^2 e^{-t_i^2/2}) = \\ &= B_i - (t_{i-1}^2 e^{-t_{i-1}^2/2} - t_i^2 e^{-t_i^2/2}), \end{aligned}$$

где $t_{i-1} < d_i < t_i$ и

$$B_i = t_{i-1}^2 e^{-t_{i-1}^2/2} (t_{i-1}^2 - d_i^2) - t_i^2 e^{-t_i^2/2} (t_i^2 - d_i^2) < 0.$$

Далее,

$$\begin{aligned}
& \frac{\partial^2 \ln L}{\partial \theta^2} \bigg|_{\theta = \tilde{\theta}} = \left\{ \sum_{(1)} \frac{n_i B_i}{\theta^2} \bigg/ (e^{-t_{i-1}^2/2} - e^{-t_i^2/2}) - \right. \\
& \quad \left. - \sum_{(1)} \frac{n_i}{\theta^2} \frac{t_{i-1}^2 e^{-t_{i-1}^2/2} - t_i^2 e^{-t_i^2/2}}{e^{-t_{i-1}^2/2} - e^{-t_i^2/2}} + \frac{2}{\theta^2} \sum_{(2)} n_i - \frac{3}{\theta^2} \sum_{(2)} \sum_{j=1}^{n_i} t_{ij}^2 \right\} \bigg|_{\theta = \tilde{\theta}} = \\
& = \left\{ \frac{1}{\theta^2} \left[\sum_{(1)} \frac{n_i B_i}{e^{-t_{i-1}^2/2} - e^{-t_i^2/2}} - 2 \sum_{(2)} \sum_{j=1}^{n_i} t_{ij}^2 \right] - \right. \\
& \quad \left. - \frac{1}{\theta^2} \left[\sum_{(1)} n_i \frac{t_{i-1}^2 e^{-t_{i-1}^2/2} - t_i^2 e^{-t_i^2/2}}{e^{-t_{i-1}^2/2} - e^{-t_i^2/2}} + \sum_{(2)} \sum_{j=1}^{n_i} (t_{ij}^2 - 2) \right] \right\} \bigg|_{\theta = \tilde{\theta}} = \\
& = \left\{ \frac{1}{\theta^2} \left[\sum_{(1)} \frac{n_i B_i}{e^{-t_{i-1}^2/2} - e^{-t_i^2/2}} - 2 \sum_{(2)} \sum_{j=1}^{n_i} t_{ij}^2 \right] \right\} \bigg|_{\theta = \tilde{\theta}} < 0,
\end{aligned}$$

так как $B_i < 0$.

Следовательно, ОМП распределения Рэлея по частично группированной выборке существует и единственна.

Как следствие этого результата в частном случае справедлива теорема для полностью группированной выборки [69]. Для группированной выборки $k=2$ оценка параметра θ определяется выражением

$$\hat{\theta} = x_{(1)} / \sqrt{2 \ln(n_1/n_2 + 1)}.$$

ОМП параметра распределения Рэлея по группированным данным состоятельна и асимптотически эффективна, если она существует и единственна.

Техника доказательства остальных теорем, определяющих условия существования и единственности и приводимых в данном разделе, аналогична. Поэтому в дальнейшем изложении доказательства теорем опущены.

Функция плотности **распределения Максвелла** определяется выражением

$$f(x, \theta) = 2x^2 \varphi(x/\theta) / \theta^3, \quad (1.3.16)$$

где $\varphi(z) = e^{-z^2/2} / \sqrt{2\pi}$ на множестве $X = (0, \infty)$; параметр $\theta \in \Omega = (0, \infty)$; $x_{(0)} = 0$; $x_{(k)} = \infty$. Вероятность попадания наблюдения в i -й интервал

$$P_i(\theta) = 2[\Phi_0(t_i) - t_i\varphi(t_i) - \Phi_0(t_{i-1}) + t_{i-1}\varphi(t_{i-1})],$$

где $t_i = x_{(i)}/\theta$; $\Phi_0(t) = \int_0^t \varphi(z) dz$.

Теорема 1.2. *ОМП параметра θ распределения Максвелла по частично группированной выборке существует при условии (1.3.2) и получается в качестве единственного решения уравнения правдоподобия*

$$\frac{1}{\theta} \left\{ \sum_{(1)} n_i \frac{t_{i-1}^3 \varphi(t_{i-1}) - t_i^3 \varphi(t_i)}{\Phi_0(t_i) - t_i \varphi(t_i) - \Phi_0(t_{i-1}) + t_{i-1} \varphi(t_{i-1})} + \sum_{(2)} \sum_{j=1}^{n_i} (t_{ij}^2 - 3) \right\} = 0.$$

В частном случае получаем теорему для полностью группированной выборки [69].

В условиях теоремы 1.2 ОМП параметра распределения Максвелла по группированным данным состоятельна и асимптотически эффективна.

Функция плотности **распределения Вейбулла** имеет вид

$$f(x, \theta, \theta_1) = \frac{\theta}{\theta_1} \left(\frac{x}{\theta_1} \right)^{\theta-1} \exp \left\{ - (x/\theta_1)^\theta \right\} \quad (1.3.17)$$

на множестве $X = (0, \infty)$, где θ - основной параметр; θ_1 - масштабный параметр; $\theta, \theta_1 \in (0, \infty)$. Вероятность попадания наблюдения в интервал

$$P_i(\theta, \theta_1) = \exp(-t_{i-1}^\theta) - \exp(-t_i^\theta),$$

где $t_i = (x_{(i)}/\theta_1)^\theta$.

В [48] приведены условия существования ОМП основного параметра распределения по группированным данным. В данном случае [42] представим более общие условия ОМП по частично группированным выборке.

Теорема 1.3. *ОМП основного параметра распределения Вейбулла по частично группированной выборке существует тогда и*

только тогда, когда $\sum_{(2)} n_i > 0$ или для интервалов первого типа выполняется одно из следующих условий:

а) при $k = 2$, $n_1 > N(1 - e^{-1})$ для $t_1 > 1$

или $n_1 < N(1 - e^{-1})$ для $t_1 < 1$;

б) при $k > 2$, $n_1 < N$, $n_k < N$, $n_1 + n_k = N$ и

$n_1 > n_k(e - 1) \cdot \ln t_{k-1} / \ln t_1$ для $t_1 > 1$ или

$n_1 < n_k(e - 1) \cdot \ln t_{k-1} / \ln t_1$ для $t_1 < 1$;

с) при $k > 2$, $n_1 + n_k < N$ и для некоторого i такого, что $t_i < 1$ или $t_{i-1} > 1$, $n_i > 0$.

При этом ОМП определяется как решение уравнения правдоподобия

$$\frac{1}{\theta} \left\{ \sum_{(1)} n_i \frac{t_i e^{-t_i} \ln t_i - t_{i-1} e^{-t_{i-1}} \ln t_{i-1}}{e^{-t_{i-1}} - e^{-t_i}} + \sum_{(2)} \sum_{j=1}^{n_i} [(1 - t_{ij}) \ln t_{ij} + 1] \right\} = 0,$$

где $t_{ij} = (x_{ij} / \theta_1)^\theta$.

Существует положительная вероятность того, что ОМП основного параметра не единственна, если i -й интервал такой, что $t_{i-1} \leq 1 \leq t_i$, принадлежит к первому типу и в него попадает значительное число наблюдений. В случае, если этот интервал относится ко второму типу, единственность оценки обеспечивается.

Из теоремы 1.3 в качестве частного случая могут быть получены условия существования ОМП при полностью группированной выборке. Для $k = 2$ ОМП параметра θ по группированной выборке определяется выражением

$$\hat{\theta} = \ln(\ln(n_1/n_2 + 1)) / \ln(x_{(1)} / \theta).$$

Теорема 1.4. ОМП масштабного параметра θ_1 распределения Вейбулла по частично группированным данным существует при условиях (1.3.2) и определяется как единственный корень уравнения правдоподобия

$$\frac{\theta}{\theta_1} \left\{ \sum_{(1)} n_i \frac{t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i}}{e^{-t_{i-1}} - e^{-t_i}} - \sum_{(2)} \sum_{j=1}^{n_i} (1 - t_{ij}) \right\} = 0,$$

где $t_i = (x_{(i)} / \theta_1)^\theta$; $t_{ij} = (x_{ij} / \theta_1)^\theta$.

В частном случае как следствие получаем теорему для группированных данных [69]. Для числа интервалов $k = 2$ при группированной выборке получаем явное выражение для ОМП параметра θ_1 :

$$\hat{\theta}_1 = x_{(1)} / [\ln(n_1/n_2 + 1)]^{1/\theta}.$$

В условиях теоремы 1.4 ОМП масштабного параметра распределения Вейбулла по группированным данным состоятельна и асимптотически эффективна.

Плотность распределения *наибольшего экстремального значения* имеет вид

$$f(x, \mu, \sigma) = \frac{1}{\sigma} \exp \left\{ \frac{\mu - x}{\sigma} - \exp \left[\frac{\mu - x}{\sigma} \right] \right\} \quad (1.3.18)$$

на множестве $X \in (-\infty, \infty)$, параметр сдвига $\mu \in (-\infty, \infty)$, масштабный параметр $\sigma \in (0, \infty)$, $x_{(0)} = -\infty$, $x_{(k)} = \infty$. Вероятность попадания наблюдения в i -й интервал

$$P_i(\mu, \sigma) = \exp\{-t_i\} - \exp\{-t_{i-1}\},$$

где $t_i = \exp\left\{\left(\mu - x_{(i)}\right)/\sigma\right\}$.

Теорема 1.5. *ОМП параметра сдвига μ распределения наибольшего экстремального значения по частично группированной выборке существует при условиях (1.3.2) и получается в качестве единственного решения уравнения правдоподобия*

$$\frac{1}{\sigma} \left\{ \sum_{(1)} n_i \frac{t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i}}{e^{-t_i} - e^{-t_{i-1}}} + \sum_{(2)} \sum_{j=1}^{n_i} (1 - t_{ij}) \right\} = 0$$

где $t_{ij} = \exp\left\{\left(\mu - x_{ij}\right)/\sigma\right\}$.

В частном случае получаем теорему для полностью группированной выборки [69]. Для группированной выборки при $k = 2$ выражение для ОМП параметра μ принимает вид

$$\hat{\mu} = x_{(1)} + \sigma \ln[\ln(n_2/n_1 + 1)].$$

В условиях теоремы 1.5 ОМП параметра μ распределения наибольшего экстремального значения состоятельна и асимптотически эффективна.

Теорема 1.6. *ОМП масштабного параметра σ распределения наибольшего экстремального значения по частично группированной выборке существует тогда и только тогда, когда $\sum_{(2)} n_i > 0$ или для*

интервалов первого типа выполняется одно из следующих условий:

а) *при $k = 2$, $n_1 < Ne^{-1}$ для $x_{(1)} < \mu$ или*

$$n_1 > Ne^{-1} \text{ для } x_{(1)} > \mu;$$

б) *при $k > 2$, $n_1 < N$, $n_k < N$, $n_1 + n_k = N$,*

$$n_1 < \frac{n_k}{e-1} \frac{\mu - x_{(k-1)}}{\mu - x_{(1)}} \text{ для } x_{(k-1)} < \mu \text{ или}$$

$$n_1 > \frac{n_k}{e-1} \frac{\mu - x_{(k-1)}}{\mu - x_{(1)}} \text{ для } x_{(1)} > \mu;$$

в) *при $k > 2$, $n_1 + n_k < N$ и для некоторого i такого, что $x_{(i)} < \mu$ или $\mu < x_{(i-1)}$, $n_i > 0$.*

В этом случае ОМП определяется как решение уравнения правдоподобия

$$\frac{1}{\sigma} \left\{ \sum_{(1)} n_i \frac{s_i e^{s_i} e^{-e^{s_i}} - s_{i-1} e^{s_{i-1}} e^{-e^{s_{i-1}}}}{e^{-e^{s_i}} - e^{-e^{s_{i-1}}}} + \sum_{(2)} \sum_{j=1}^{n_i} \left[s_{ij} (e^{s_{ij}} - 1)^{-1} \right] \right\} = 0$$

где $s_i = (\mu - x_{(i)})/\sigma$; $s_{ij} = (\mu - x_{ij})/\sigma$.

Для частного случая полностью группированной выборки получаем теорему, доказанную в [69]. В случае $k = 2$ ОМП параметра σ по группированной выборке может быть получена в явном виде

$$\hat{\sigma} = (\mu - x_{(1)}) / \ln[\ln(n_2/n_1 + 1)].$$

ОМП параметра σ распределения наибольшего экстремального значения по группированным данным состоятельна и асимптотически эффективна, если она существует и единственна.

Распределение **наименьшего экстремального значения** имеет функцию плотности

$$f(x, \mu, \sigma) = \frac{1}{\sigma} \exp \left\{ \frac{x - \mu}{\sigma} - \exp \left\{ \frac{x - \mu}{\sigma} \right\} \right\} \quad (1.3.19)$$

на множестве $X = (-\infty, \infty)$, параметр сдвига $\mu \in (-\infty, \infty)$, масштабный параметр $\sigma \in (0, \infty)$, $x_{(0)} = -\infty$, $x_{(k)} = \infty$. Вероятность попадания наблюдения в i -й интервал

$$P_i(\mu, \sigma) = \exp(-t_{i-1}) - \exp(-t_i),$$

где $t_i = \exp \left\{ (x_{(i)} - \mu) / \sigma \right\}$.

Теорема 1.7. ОМП параметра сдвига μ распределения наименьшего экстремального значения по частично группированной выборке существует при условиях (1.3.2) и получается в качестве единственного решения уравнения правдоподобия

$$\frac{1}{\sigma} \left\{ \sum_{(1)} n_i \frac{t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i}}{e^{-t_{i-1}} - e^{-t_i}} + \sum_{(2)} \sum_{j=1}^{n_i} (t_{ij} - 1) \right\} = 0$$

где $t_{ij} = \exp \left\{ (x_{ij} - \mu) / \sigma \right\}$.

Теорема 1.8. ОМП параметра σ распределения наименьшего экстремального значения по частично группированной выборке существует тогда и только тогда, когда $\sum_{(2)} n_i > 0$ или для

интервалов первого типа выполняется одно из следующих условий:

а) при $k = 2$, $n_2 < Ne^{-1}$ для $x_{(1)} > \mu$ или

$$n_2 > Ne^{-1} \text{ для } x_{(1)} < \mu;$$

б) при $k > 2$, $n_1 < N$, $n_k < N$, $n_1 + n_k = N$,

$$n_k < \frac{n_1}{e-1} \frac{x_{(1)} - \mu}{x_{(k-1)} - \mu} \text{ для } x_{(1)} > \mu \text{ для}$$

$$n_k > \frac{n_1}{e-1} \frac{x_{(1)} - \mu}{x_{(k-1)} - \mu} \text{ для } x_{(k-1)} < \mu;$$

с) при $k > 2$, $n_1 + n_k < N$ и для некоторого i такого, что $x_{(i)} < \mu$

$$\text{или } \mu < x_{(i-1)}, n_i > 0.$$

В этом случае ОМП определяется как решение уравнения правдоподобия

$$\frac{1}{\sigma} \left\{ \sum_{(1)} n_i \frac{s_{i-1} e^{s_{i-1}} e^{-e^{s_{i-1}}} - s_i e^{s_i} e^{-e^{s_i}}}{e^{-e^{s_{i-1}}} - e^{-e^{s_i}}} + \sum_{(2)} \sum_{j=1}^{n_i} \left[s_{ij} (e^{s_{ij}} - 1)^{-1} \right] \right\} = 0,$$

где $s_i = (x_{(i)} - \mu) / \sigma$; $s_{ij} = (x_{ij} - \mu) / \sigma$.

Функция плотности **гамма-распределения** имеет вид

$$f(x, \theta, \theta_1) = \theta_1^\theta x^{\theta-1} \exp\{-\theta_1 x\} / \Gamma(\theta) \quad (1.3.20)$$

на множестве $X = (0, \infty)$, θ - основной параметр, θ_1 - масштабный, $\theta, \theta_1 \in (0, \infty)$, $x_{(0)} = 0$, $x_{(k)} = \infty$, $\Gamma(\theta)$ - гамма-функция Эйлера. Вероятность попадания в i -й интервал группирования

$$P_i(\theta, \theta_1) = \frac{1}{\Gamma(\theta)} \int_{t_{i-1}}^{t_i} t^{\theta-1} e^{-t} dt,$$

where $t_i = \theta_1 x_{(i)}$.

В [30,48] приведены условия существования для ОМП основного параметра гамма-распределения по группированным наблюдениям, доказана состоятельность и асимптотическая эффективность оценки.

Теорема 1.9. *ОМП основного параметра гамма-распределения по группированным наблюдениям существует тогда и только тогда, когда $n_1 < N$, $n_k < N$, и получается в качестве решения уравнения правдоподобия*

$$\sum_{i=1}^k n_i \int_{t_{i-1}}^{t_i} t^{\theta-1} e^{-t} \ln t dt \Bigg/ \int_{t_{i-1}}^{t_i} t^{\theta-1} e^{-t} dt - N \psi(\theta) = 0,$$

где $\psi(\theta)$ - логарифмическая производная гамма-функции [149].

ОМП основного параметра гамма-распределения по группированным данным состоятельна и асимптотически эффективна, если она существует и единственна.

Теорема 1.10. *ОМП масштабного параметра θ_1 гамма-распределения по частично группированной выборке существует при условиях (1.3.2) и определяется как единственное решение уравнения правдоподобия*

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i (t_i^\theta e^{-t_i} - t_{i-1}^\theta e^{-t_{i-1}}) \Bigg/ \int_{t_{i-1}}^{t_i} t^{\theta-1} e^{-t} dt + \sum_{(2)} \sum_{j=1}^{n_i} (\theta - t_{ij}) \right\} = 0,$$

где $t_i = \theta_1 x_{(i)}$, $t_{ij} = \theta_1 x_{ij}$.

Теорема для полностью группированной выборки получается как частный случай. ОМП масштабного параметра гамма-распределения в условиях теоремы 1.10 состоятельна и асимптотически эффективна.

В частном случае гамма-распределения, когда основной параметр принимает целые значения, имеем **распределение Эрланга** с функцией плотности

$$f(x, \theta_1) = \theta_1^n x^{n-1} \exp\{-\theta_1 x\} / \Gamma(n). \quad (1.3.21)$$

Условия существования и единственности ОМП параметра распределения Эрланга определяются теоремой 1.10 [6].

Функция плотности **распределения Коши** описывается выражением [150]

$$f(x, \theta, \theta_1) = \frac{1}{\pi} \frac{\theta}{\theta^2 + (x - \theta_1)^2} \quad (1.3.22)$$

на множестве $X = (-\infty, \infty)$, где θ - масштабный параметр, $\theta \in (0, \infty)$, θ_1 - параметр сдвига, $\theta_1 \in (-\infty, \infty)$, $x_{(0)} = -\infty$, $x_{(k)} = \infty$. Вероятность попадания наблюдения в i -й интервал

$$P_i(\theta, \theta_1) = (\arctg t_i - \arctg t_{i-1}) / \pi,$$

где $t_i = (x_{(i)} - \theta_1) / \theta$.

Теорема 1.11. ОМП масштабного параметра θ распределения Коши по частично группированной выборке существует тогда и только тогда, когда $\sum_{(2)} n_i > 0$ или для интервалов первого типа:

- а) при $k \geq 2$, $n_1 < N$, $n_k < N$, $n_1 + n_k = N$ для $t_1 < t_{k-1} < 0$
- $$n_1 < n_k (1 + t_1^2) t_{k-1} (\arctg t_1 + \pi/2) / [t_1 (1 + t_{k-1}^2) (\pi/2 - \arctg t_{k-1})],$$
- для $0 < t_1 < t_{k-1}$
- $$n_1 > n_k (1 + t_1^2) t_{k-1} (\arctg t_1 + \pi/2) / [t_1 (1 + t_{k-1}^2) (\pi/2 - \arctg t_{k-1})];$$
- б) при $k > 2$, $n_1 + n_k < N$, для некоторого i такого, что $t_i < 0$ или $t_{i-1} > 0$ ($x_{(i)} < \theta_1$ или $x_{(i-1)} > \theta_1$), $n_i > 0$.

В этом случае ОМП получается в качестве корня уравнения правдоподобия

$$\frac{1}{\theta} \left\{ \sum_{(1)} n_i \frac{\frac{t_{i-1}}{1+t_{i-1}^2} - \frac{t_i}{1+t_i^2}}{\operatorname{arctg} t_i - \operatorname{arctg} t_{i-1}} + \sum_{(2)} \sum_{j=1}^{n_i} \left(1 - \frac{2}{1+t_{ij}^2} \right) \right\} = 0,$$

где $t_{ij} = (x_{ij} - \theta_1)/\theta$.

В частном случае для группированной выборки при $k = 2$, ОМП параметра θ имеет вид

$$\hat{\theta} = (x_{(1)} - \theta_1) \operatorname{ctg} \left(\frac{\pi}{2} \frac{2n_1 - N}{N} \right).$$

Теорема 1.12. *ОМП масштабного параметра θ распределения Коши по группированным данным состоятельна и асимптотически эффективна, если она существует и единственна.*

Доказательство. Первые частные производные функций $p_i(\theta, \theta_1)$ по θ имеют вид

$$\frac{\partial}{\partial \theta} P_i(\theta, \theta_1) = \frac{1}{\pi \theta} \left(\frac{t_{i-1}}{1+t_{i-1}^2} - \frac{t_i}{1+t_i^2} \right),$$

и, следовательно, так как $k \geq 2$ не все $\partial p_i(\theta, \theta_1)/\partial \theta$ равны нулю при истинном значении параметра θ . Вторые частные производные, определяемые выражением

$$\frac{\partial^2 P_i(\theta, \theta_1)}{\partial \theta^2} = \frac{1}{\pi \theta^2} \left[\frac{2t_i}{(1+t_i^2)^2} - \frac{2t_{i-1}}{(1+t_{i-1}^2)^2} \right],$$

также непрерывны. Справедливость теоремы следует из теорем 3 и 5 [18, §1].

Теорема 1.13. *ОМП параметра сдвига θ_1 распределения Коши по частично группированной выборке существует при условии (1.3.2) и получается в качестве решения уравнения правдоподобия*

$$\frac{1}{\theta} \left\{ \sum_{(1)} n_i \frac{\frac{1}{1+t_{i-1}^2} - \frac{1}{1+t_i^2}}{\operatorname{arctg} t_i - \operatorname{arctg} t_{i-1}} + 2 \sum_{(2)} \sum_{j=1}^{n_i} \frac{t_{ij}}{1+t_{ij}^2} \right\} = 0,$$

Численными методами показано, что при оценивании параметра сдвига по небольшому числу независимых одинаково распределенных величин (даже в негруппированном случае) с достаточно боль-

шой вероятностью возникают функции правдоподобия с несколькими максимумами.

Для группированной выборки при $k=2$ ОМП параметра θ_1 определяется выражением

$$\hat{\theta}_1 = x_{(1)} - \theta \operatorname{tg} \left(\frac{\pi}{2} \frac{2n_1 - N}{N} \right).$$

ОМП параметра сдвига θ_1 распределения Коши по группированным данным состоятельна и асимптотически эффективна, если она существует и единственна.

Функция плотности *логистического распределения* описывается выражением

$$f(x, \mu, \sigma) = \frac{\pi}{\sigma\sqrt{3}} \exp \left\{ -\frac{\pi}{\sqrt{3}} \frac{x - \mu}{\sigma} \right\} / \left[1 + \exp \left\{ -\frac{\pi}{\sqrt{3}} \frac{x - \mu}{\sigma} \right\} \right]^2 \quad (1.3.23)$$

на множестве $X = (-\infty, \infty)$, где $\mu \in (-\infty, \infty)$ - параметр сдвига; $\sigma \in (0, \infty)$ - масштабный параметр; $x_{(0)} = -\infty$; $x_{(k)} = \infty$. Вероятность попадания наблюдения в i -й интервал

$$P_i(\mu, \sigma) = (1 + e^{-t_i})^{-1} - (1 + e^{-t_{i-1}})^{-1},$$

где $t_i = \pi(x_{(i)} - \mu) / \sigma\sqrt{3}$.

Теорема 1.14. *ОМП параметра сдвига μ логистического распределения существует при условиях (1.3.2) и получается в качестве единственного решения уравнения правдоподобия*

$$\frac{\pi}{\sigma\sqrt{3}} \left\{ \sum_{(1)} n_i \frac{\frac{e^{-t_{i-1}}}{(1 + e^{-t_{i-1}})^2} - \frac{e^{-t_i}}{(1 + e^{-t_i})^2}}{(1 + e^{-t_i})^{-1} - (1 + e^{-t_{i-1}})^{-1}} + \sum_{(2)} \sum_{j=1}^{n_i} \left(1 - 2 \frac{e^{-t_{ij}}}{1 + e^{-t_{ij}}} \right) \right\} = 0,$$

where $t_{ij} = \pi(x_{ij} - \mu) / \sigma\sqrt{3}$.

В условиях теоремы 1.14 ОМП параметра сдвига логистического распределения по группированным наблюдениям состоятельна и асимптотически эффективна.

При $k = 2$ для группированной выборки ОМП параметра μ находится из соотношения

$$\hat{\mu} = x_{(1)} - \frac{\sigma\sqrt{3}}{\pi} \ln \frac{n_1}{n_2}.$$

Теорема 1.15. *ОМП масштабного параметра σ логистического распределения по частично группированной выборке существует тогда и только тогда, когда $\sum_{(2)} n_i > 0$ или для интервалов первого типа выполняется одно из следующих условий:*

а) *при $k = 2$, $n_1 + n_2 = N$,*

$$n_1 > n_2, \text{ если } x_{(1)} > \mu, \quad n_1 < n_2, \text{ если } x_{(1)} < \mu;$$

б) *при $k > 2$, $n_1 < N$, $n_k < N$, $n_1 + n_k = N$ и*

$$n_1 > n_k (x_{(k-1)} - \mu) / (x_{(1)} - \mu), \text{ если } x_{(1)} > \mu, \text{ или}$$

$$n_1 < n_k (x_{(k-1)} - \mu) / (x_{(1)} - \mu), \text{ если } x_{(k-1)} < \mu;$$

с) *при $k > 2$, $n_1 + n_k < N$ и для некоторого i такого, что $y_{(i)} < \mu$,*

$$\text{или } x_{(i-1)} > \mu, \quad n_i > 0.$$

При этом ОМП получается в качестве решения уравнения правдоподобия

$$\frac{1}{\sigma} \left\{ \sum_{(1)} n_i \frac{\frac{t_{i-1} e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2} - \frac{t_i e^{-t_i}}{(1+e^{-t_i})^2}}{(1+e^{-t_i})^{-1} - (1+e^{-t_{i-1}})^{-1}} + \sum_{(2)} \sum_{j=1}^{n_i} \left(t_{ij} - 1 - 2 \frac{t_{ij} e^{-t_{ij}}}{1+e^{-t_{ij}}} \right) \right\} = 0,$$

$$\text{где } t_i = \pi(x_{(i)} - \mu) / \sigma\sqrt{3}; \quad t_{ij} = \pi(x_{ij} - \mu) / \sigma\sqrt{3}.$$

По группированной выборке для $k = 2$, ОМП параметра σ определяется из выражения

$$\hat{\sigma} = \frac{\pi}{\sqrt{3}} (x_{(1)} - \mu) \cdot \ln \frac{n_1}{n_2}.$$

ОМП параметра σ по группированным наблюдениям состоятельна и асимптотически эффективна, если она существует и единственна.

Функция плотности **распределения Парето** имеет вид

$$f(x, \alpha, \sigma) = \alpha^\theta \theta x^{-(\theta+1)} \quad (1.3.24)$$

на множестве $X \in (\alpha, \infty)$, α - параметр сдвига, $\alpha \in (0, \infty)$, θ - основной параметр, $\theta \in (0, \infty)$, $x_{(0)} = \alpha$, $x_{(k)} = \infty$. Вероятность попадания наблюдения в i -й интервал

$$P_i(\theta) = s_{i-1} - s_i,$$

где $s_i = (x_{(i)}/\alpha)^{-\theta}$.

Теорема 1.16. *ОМП параметра θ распределения Парето по частично группированной выборке существует при условиях (1.3.2) и получается в качестве единственного решения уравнения правдоподобия*

$$\frac{1}{\theta} \left\{ \sum_{(1)} n_i \frac{s_{i-1} \ln s_{i-1} - s_i \ln s_i}{s_{i-1} - s_i} + \sum_{(2)} \sum_{j=1} (1 + \ln s_{ij}) \right\} = 0,$$

где $s_{ij} = (x_{ij}/\alpha)^{-\theta}$.

В частном случае для группированной выборки и $k=2$, ОМП параметра θ определяется соотношением

$$\hat{\theta} = \ln(n_1/n_2 + 1) / \ln(x_{(1)}/\alpha).$$

Условия существования и единственности ОМП параметров **нормального** и **экспоненциального** распределений по группированным и частично группированным выборкам подробно рассмотрены Куллдорфом [65].

Наиболее часто **логарифмически нормальное распределение** используется в виде

$$f(x, \mu, \sigma) = \frac{1}{\sigma x \sqrt{2\pi}} \exp \left\{ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right\} \quad (1.3.25)$$

на множестве $X \in (0, \infty)$ или в виде

$$f(x, \mu, \sigma) = \frac{1}{\ln 10 \sigma x \sqrt{2\pi}} \exp \left\{ -\frac{(\lg x - \mu)^2}{2\sigma^2} \right\}, \quad (1.3.26)$$

где $\mu \in (-\infty, \infty)$, $\sigma \in (0, \infty)$ параметры распределения. Вероятность попадания наблюдения в i -й интервал равна

$$P_i(\mu, \sigma) = \Phi(t_i) - \Phi(t_{i-1}),$$

где $t_i = (\ln x_{(i)} - \mu) / \sigma$ для закона (1.3.25) и $t_i = (\lg x_{(i)} - \mu) / \sigma$ для распределения (1.3.26).

Теорема 1.17. *ОМП параметра μ логарифмически нормального распределения по частично группированной выборке существует при условиях (1.3.2) и получается в качестве единственного решения уравнения правдоподобия*

$$\frac{1}{\sigma} \left\{ \sum_{(1)} n_i \frac{\varphi(t_{i-1}) - \varphi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} \sum_{j=1}^{n_i} t_{ij} \right\} = 0, \quad (1.3.27)$$

где $\varphi(z) = e^{-z^2/2} / \sqrt{2\pi}$; $\Phi(t) = \int_{-\infty}^t \varphi(z) dz$;

$t_i = (\ln x_{(i)} - \mu) / \sigma$; $t_{ij} = (\ln x_{ij} - \mu) / \sigma$ для распределения (1.3.25);

$t_i = (\lg y_{(i)} - \mu) / \sigma$; $t_{ij} = (\lg y_{ij} - \mu) / \sigma$ для распределения (1.3.26).

Справедливость теоремы следует из теоремы 12.1 [65], так как уравнение (1.3.27) совпадает с уравнением правдоподобия для параметра μ нормального распределения.

Теорема 1.18. *ОМП параметра σ логарифмически нормального распределения по частично группированной выборке существует тогда и только тогда, когда $\sum_{(2)} n_i > 0$ или для интервалов первого типа*

выполняется одно из следующих условий: для некоторого i удовлетворяющего неравенствам $t_{i-1} > 0$ или $t_i > 0$, $n_1 + n_k < N$ и $n_i > 0$; $n_1 + n_k = N$, $0 < n_1 < N$ и $n_1 t_1 > n_k t_{k-1}$. И в этом случае ОМП получается в качестве единственного решения уравнения правдоподобия

$$\frac{1}{\sigma} \left\{ \sum_{(1)} n_i \frac{t_{i-1} \varphi(t_{i-1}) - t_i \varphi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} \sum_{j=1}^{n_i} (t_{ij}^2 - 1) \right\} = 0,$$

где $t_i = (\ln x_{(i)} - \mu) / \sigma$; $t_{ij} = (\ln x_{ij} - \mu) / \sigma$ для распределения (1.3.25);

$t_i = (\lg x_{(i)} - \mu) / \sigma$; $t_{ij} = (\lg x_{ij} - \mu) / \sigma$ для распределения (1.3.26).

Справедливость теоремы следует из теоремы 13.1 [65].

Функция плотности **полунормального распределения** описывается выражением

$$f(x, \theta) = \frac{2}{\theta\sqrt{2\pi}} e^{-x^2/2\theta^2}, \quad (1.3.28)$$

где $X \in (0, \infty)$, $\theta \in (0, \infty)$. Вероятность попадания наблюдения в интервал

$$P_i(\theta) = \Phi_0(t_i) - \Phi_0(t_{i-1}),$$

где $t_i = x_{(i)}/\theta$; $\Phi_0(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-z^2/2} dz$.

Теорема 1.19. *ОМП масштабного параметра θ полунормального распределения существует при условиях (1.3.2) и получается в качестве единственного решения уравнения правдоподобия*

$$\frac{1}{\theta} \left\{ \sqrt{\frac{2}{\pi}} \sum_{(1)} n_i \frac{t_{i-1} e^{-t_{i-1}^2/2} - t_i e^{-t_i^2/2}}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} \sum_{j=1}^{n_i} (t_{ij}^2 - 1) \right\} = 0,$$

где $t_{ij} = x_{ij}/\theta$. В этих условиях ОМП состоятельна и асимптотически эффективна.

Функция плотности **распределения Лапласа** имеет вид

$$f(x, \theta) = \theta e^{-\theta|x|/2}, \quad (1.3.29)$$

где $X \in (-\infty, \infty)$; $\theta \in (0, \infty)$; $x_{(0)} = -\infty$; $x_{(k)} = \infty$.

Теорема 1.20. [71]. *ОМП масштабного параметра θ распределения Лапласа по частично группированной выборке существует при условиях (1.3.2) и получается в качестве решения уравнения правдоподобия*

$$\sum_{(1)} n_i B_i + \sum_{(2)} \sum_{j=1}^{n_i} (1 - t_{ij}) = 0.$$

где $t_i = \theta x_{(i)}$; $t_{ij} = \theta x_{ij}$;

$$B_i = \begin{cases} \frac{t_i e^{t_i} - t_{i-1} e^{t_{i-1}}}{e^{t_i} - e^{t_{i-1}}}, & t_i < 0, \\ \frac{t_i e^{-t_i} - t_{i-1} e^{-t_{i-1}}}{e^{-t_{i-1}} - e^{-t_i}}, & t_{i-1} > 0, \\ \frac{t_i e^{-t_i} - t_{i-1} e^{t_{i-1}}}{2 - (e^{-t_i} - e^{t_{i-1}})}, & t_{i-1} < 0, \quad t_i > 0. \end{cases}$$

При оценивании двух и более параметров распределений методом максимального правдоподобия необходимо отыскивать решение системы уравнений правдоподобия, которое доставляет максимум (или локальный максимум) функции правдоподобия. Н.А. Бодины доказана следующая теорема для группированных выборок [18].

Теорема 1.21. Пусть для всякого i ($i = \overline{1, k}$), для которого $n_i > 0$, функции $P_i = P_i(\theta_1, \dots, \theta_m)$ имеют непрерывные первые частные производные на открытой области Ω . Предположим, что существуют точки $\underline{\theta} = (\underline{\theta}_1, \underline{\theta}_2, \dots, \underline{\theta}_m)^T$ и $\bar{\theta} = (\bar{\theta}_1, \bar{\theta}_2, \dots, \bar{\theta}_m)$ области Ω , удовлетворяющие следующим условиям для каждого индекса $l = \overline{1, m}$:

$$\sum_{i=1}^k n_i \frac{\partial \ln P_i}{\partial \theta_l} \bigg|_{\theta_l = \underline{\theta}_l} > 0; \quad (1.3.30)$$

$$\sum_{i=1}^k n_i \frac{\partial \ln P_i}{\partial \theta_l} \bigg|_{\theta_l = \bar{\theta}_l} < 0 \quad (1.3.31)$$

для всех $\theta_1, \dots, \theta_l, \theta_{l+1}, \dots, \theta_m$, принадлежащих Ω . Тогда существует ОМП, являющаяся решением системы уравнений правдоподобия, доставляющая локальный максимум функции правдоподобия.

Справедливость приведенной теоремы сохраняется и для частично группированной выборки. В этом случае добавляется требование существования первых частных производных функции плотности распределения на области Ω . Условия (1.3.30-1.3.31) преобразуются к виду

$$\left\{ \sum_{(1)} n_i \frac{\partial \ln p_i}{\partial \theta_l} + \sum_{(2)} \sum_{j=1}^{n_i} \frac{\partial \ln f}{\partial \theta_l} \right\} \bigg|_{\theta_l = \underline{\theta}_l} > 0,$$

$$\left\{ \sum_{(1)} n_i \frac{\partial \ln p_i}{\partial \theta_l} + \sum_{(2)} \sum_{j=1}^{n_i} \frac{\partial \ln f}{\partial \theta_l} \right\} \bigg|_{\theta_l = \bar{\theta}_l} < 0$$

для $l = \overline{1, m}$. Это расширение не нуждается в доказательстве, так как совершенно не меняет ход доказательства теоремы 1.21.

Опираясь на приведенную модификацию теоремы, из анализа полученных выше условий существования ОМП параметров для распределений Вейбулла, нормального, логарифмически нормальных, гамма-распределения, распределений экстремальных значений, Коши, логистического, можно утверждать следующее: ОМП векторного параметра перечисленных распределений существует, если $\sum_{(2)} n_i > 0$ или для $k \geq 3$ по крайней мере для трех интервалов первого

типа $n_i > 0$. При этом ОМП получается в качестве решения системы уравнений правдоподобия. При попадании наблюдений в два крайних интервала первого типа или в какой-то один не крайний существует положительная вероятность отсутствия решения системы уравнений правдоподобия. Естественно, что при попадании всех наблюдений в один из интервалов ОМП не существует.

Таким образом, получены условия существования и единственности ОМП по частично группированным данным для ряда одномерных непрерывных распределений, наиболее часто используемых на практике при описании реально наблюдаемых случайных величин. Условия существования легко проверяемы и в совокупности позволяют предположить существование ОМП по группированным данным при распределении числа попаданий наблюдений в число интервалов, большее количества оцениваемых по выборке параметров.

1.4. Оценивание параметров распределений по интервальным наблюдениям

В случае частично группированной выборки (группированной или негруппированной) одним из наиболее эффективных методов является метод максимального правдоподобия.

При интервальной выборке наблюдения задаются интервалами, которые в отличие от группированного случая могут пересекаться. Пусть задана интервальная выборка [83]:

$$\mathbf{X}_k = \left\{ (a_i, b_i, n_i) \left(\left(a_i \leq x_{ij} \leq b_i, j=1, \dots, n_i; a_i \in R, b_i \in R, i=1, \dots, k \right) \right. \right\},$$

где k - число интервалов, $N = \sum_{i=1}^k n_i$ - объём выборки, n_i - число наблюдений в i -м интервале, a_i и b_i - границы i -го интервала, x_{ij} - точные значения наблюдений. И пусть

$$N_{a_{(i)}} = \sum_{j=1}^i n_{a_{(j)}}, N_{b_{(i)}} = \sum_{j=1}^i n_{b_{(j)}}, i = 1, \dots, k.$$

При таком представлении наблюдаемой выборки мы можем пытаться вычислить точечную или интервальную оценку неизвестного параметра распределения. К определению точечной оценки по интервальной выборке пожалуй наиболее естественно адаптируется метод максимального правдоподобия. Искомая оценка параметра вычисляется как значение θ , максимизирующее выражение

$$L(\theta) = \prod_{i=1}^k \left(\frac{F(b_i, \theta) - F(a_i, \theta)}{b_i - a_i} \right)^{n_i},$$

где $F(x, \theta)$ - функция распределения, параметр которой оценивается по интервальной выборке. Последнее выражение совпадает с выражением (1.2.1) для функции правдоподобия по группированным данным.

Используя различные статистики, измеряющие расстояния между теоретической функцией распределения и эмпирической, можно строить точечные или интервальные оценки параметров, которые минимизируют эти расстояния (MD-оценки). Эмпирическая функция распределения $F_N(x)$, построенная по индивидуальным наблюдениям x_{ij} (неизвестным нам), будет ограничена снизу и сверху функциями $\underline{F}_N(x)$ и $\overline{F}_N(x)$, которые имеют следующий вид:

$$\overline{F}_N(x) = \begin{cases} 0, & x < a_{(1)} \\ \frac{N_{a_{(i)}}}{N}, & a_{(i)} \leq x < a_{(i+1)}, \quad i=1, \dots, k-1, \\ 1, & x \geq a_{(k)} \end{cases}$$

$$\underline{F}_N(x) = \begin{cases} 0, & x < b_{(1)} \\ \frac{N_{b_{(i)}}}{N}, & b_{(i)} \leq x < b_{(i+1)}, \quad i=1, \dots, k-1. \\ 1, & x \geq b_{(k)} \end{cases}$$

То есть,

$$\underline{F}_N(x) \leq F_N(x) \leq \overline{F}_N(x), \quad \forall x \in R.$$

Если исследователем априори задана некоторая параметрическая модель $F(x, \theta)$, то верхняя и нижняя граница искомой функции распределения также должны принадлежать этой модели:

$$\underline{F}(x) = F(x, \theta_1), \quad \overline{F}(x) = F(x, \theta_2).$$

В случае скалярного параметра θ мы можем, используя $\underline{F}_N(x, \theta)$ и $\overline{F}_N(x, \theta)$, естественным образом получить интервальную оценку параметра, а в случае векторного параметра - оценить область допустимых значений

$$T = \{\theta \in \Omega \mid (\forall x) F(x, \theta_1) \leq F(x, \theta) \leq F(x, \theta_2)\}.$$

Так как эмпирическая функция распределения принадлежит некоторому интервалу неопределенности, то и значение S^* любой статистики S , измеряющей расстояние между эмпирической и истинной функциями распределения также принадлежит соответствующему интервалу $[\underline{S}^*, \overline{S}^*]$, границы которого определяются неравенством:

$$\underline{S}^* = \inf_{\underline{F}_N \leq F_N \leq \overline{F}_N} S^*(F_N, F) \leq S^*(F_N, F) \leq \sup_{\underline{F}_N \leq F_N \leq \overline{F}_N} S^*(F_N, F) = \overline{S}^*$$

Если использовать выражения для оценок сверху и снизу [81-83,98] для статистик, например, применяемых в критериях согласия типа Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса, можно построить различные процедуры оценивания параметров.

Из-за неполноты информации об исходных данных, статистические оценки будут зависеть от степени “оптимизма” [63] в отношении согласия модели с наблюдениями случайной величины. Получение точечных оценок параметров распределений является процедурой, в значительной степени зависящей от степени оптимизма исследователя относительно соответствия выбранной модели исходным интервальным данным. Оценку сверху для статистики можно рассматривать как случай наихудшего расположения индивидуальных значений наблюдений в интервалах (“крайний пессимизм”), а оценку снизу - как случай наилучшего расположения точных значений наблюдений (“крайний оптимизм”). Оценки параметров, соответствующие этим двум крайностям,

$$\begin{aligned}\hat{\theta}_1 &= \arg \min_{\theta} \overline{S}^*(\underline{F}_N, \overline{F}_N, F(\theta)), \\ \hat{\theta}_2 &= \arg \min_{\theta} \underline{S}^*(\underline{F}_N, \overline{F}_N, F(\theta))\end{aligned}$$

можно рассматривать как интервальную оценку неизвестного параметра.

Заметим, что использование параметрической модели для описания интервальной выборки может оказаться не всегда приемлемым, так как верхняя и нижняя границы эмпирической функции распреде-

ления могут сходиться в общем случае к законам из разных параметрических семейств.

Рассмотренные подходы к оцениванию параметров по интервальной выборке реализуются в разрабатываемом программном обеспечении [84, 90, 91, 98].

1.5. Выводы

Экспериментальные наблюдения случайных величин, являющиеся объектом статистического анализа рассмотрены с позиций наиболее общего представления экспериментальных наблюдений в виде частично группированных и интервальных выборок, частными случаями которых являются группированные, негруппированные и цензурированные выборки.

Рассмотрены методы оценивания параметров распределений для частично группированных выборок.

Получены условия существования и единственности ОМП по частично группированным данным для ряда одномерных непрерывных распределений, наиболее часто используемых на практике при описании реально наблюдаемых случайных величин.

Рассмотрены подходы к оцениванию параметров распределений по интервальным выборкам.

2. АСИМПТОТИЧЕСКИ ОПТИМАЛЬНОЕ ГРУППИРОВАНИЕ И ОЦЕНИВАНИЕ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЙ

2.1. Асимптотическая эффективность оценок и асимптотически оптимальное группирование данных

Группирование наблюдений используется как при оценке параметров распределений, так и в задачах проверки статистических гипотез. До настоящего момента при группировании область определения случайной величины разбивается на интервалы равной длины или равной вероятности. Такой подход рассматривается в наиболее часто используемых для руководства источниках [10, 23, 45, 46, 55, 66, 67, 109, 136, 139, 146, 150, 151, 182, 185]. Всякая *группировка данных* по сравнению с негруппированной выборкой ведет к *потере информации*, понимаемой в общем широком смысле. И это, естественно, отражается на качестве статистических выводов. Очевидно, что чем меньше потери информации, тем надежнее получаемые результаты. Постановка задачи асимптотически оптимального группирования вытекает непосредственно из асимптотических свойств оценок и критериев. Её решение, минимизирующее потери в количестве информации Фишера, повышает качество статистических выводов как при оценивании, так и при проверке гипотез. В данной главе задача асимптотически оптимального группирования наблюдений рассматривается относительно оценивания параметров. Вопросы асимптотически оптимального группирования данных в связи с критериями согласия подробно рассматривается в главе 4.

ОМП параметров распределений по группированным данным являются асимптотически эффективными (если они существуют и единственны). И их асимптотическая дисперсия определяется соотношением

$$D(\theta) = N^{-1} M_{\Gamma}^{-1}(\theta),$$

где $M_{\Gamma}(\theta) = \left[\sum_{i=1}^k \left(\frac{\partial \ln P_i(\theta)}{\partial \theta_l} \frac{\partial \ln P_i(\theta)}{\partial \theta_r} \right) P_i(\theta) \right] = \sum_{i=1}^k \frac{\nabla P_i(\theta) \nabla P_i^T(\theta)}{P_i(\theta)}$ - информационная матрица Фишера по группированным наблюдениям.

Это же справедливо [138] для оценок, получаемых в результате минимизации статистик χ^2

$$\chi^2 = N \sum_{i=1}^k \frac{(n_i / N - P_i(\theta))^2}{P_i(\theta)},$$

модифицированный статистики χ^2

$$\text{mod } \chi^2 = N \sum_{i=1}^k \frac{(n_i - NP_i(\theta))^2}{n_i},$$

где n_i заменяется на 1, если $n_i = 0$. расстояния Хеллингера

$$H_D = \arccos \sum_{i=1}^k \sqrt{(n_i / N) P_i(\theta)},$$

дивергенции Кульбака-Лейблера

$$S_{KL} = \sum_{i=1}^k P_i(\theta) \ln [P_i(\theta) / (n_i / N)],$$

меры расхождения Холдейна

$$D_j = \frac{(N+j)!}{N!} \sum_{i=1}^k \frac{n_i! P_i^{j+1}(\theta)}{(n_i+j)!}, \quad j \neq -1,$$

$$D_{-1} = \frac{1}{N} \sum_{i=1}^k n_i \ln P_i(\theta).$$

Все эти методы, как уже упоминалось, при соответствующих условиях регулярности дают состоятельные и асимптотически эффективные оценки [138].

Аналогично, асимптотически эффективными являются оценки параметров сдвига и масштаба по выборочным квантилям [141], рассматриваемые ниже в данной главе.

Элементы информационной матрицы зависят от граничных точек интервалов, так как $P_i(\theta) = \int_{x_{(i-1)}}^{x_{(i)}} f(x, \theta) dx$. В случае, когда функция

плотности распределения определяется скалярным параметром или осуществляется оценивание только одного параметра при известных остальных, целью задачи асимптотически оптимального группирования является минимизация асимптотической дисперсии ОМП по группированным данным. И эта задача сводится к максимизации информационного количества Фишера о параметре по группированной выборке, т.е. к решению задачи нелинейного программирования

$$\max_{x_{(1)} < x_{(2)} < \dots < x_{(k-1)}} \sum_{i=1}^k \left(\frac{\partial \ln P_i(\theta)}{\partial \theta} \right)^2 P_i(\theta) = \max_{x_{(1)} < x_{(2)} < \dots < x_{(k-1)}} J_{\Gamma}(\theta), \quad (2.1.1)$$

где $J_{\Gamma}(\theta)$ - количество информации Фишера о параметре по группированным наблюдениям.

Решение такой задачи для параметра экспоненциального распределения при $k = 2, \dots, 6$ впервые было получено в [205], а для параметров нормального распределения при известном одном из них решения приведены в [65].

При оценивании вектора параметров мы имеем дело с информационной матрицей. В этом случае в качестве критериев оптимальности могут быть выбраны различные функционалы от асимптотической дисперсионной матрицы, но наиболее естественно минимизировать обобщенную асимптотическую дисперсию (максимизировать $\det M_{\Gamma}(\theta) = |M_{\Gamma}(\theta)|$), т.е. решать задачу нелинейного программирования вида

$$\max_{x_{(0)} < x_{(1)} < \dots < x_{(k-1)} < x_{(k)}} |M_{\Gamma}(\theta)|. \quad (2.1.2)$$

Как будет показано в главе 4, стремление использовать критерии согласия, предусматривающие группировку данных, таким образом, чтобы они обладали максимальной мощностью (по крайней мере, при близких альтернативах), приводит к этой же самой задаче асимптотически оптимального группирования. И, таким образом, полученные ее решения могут эффективно использоваться как в задачах оценивания, так и в задачах проверки статистических гипотез.

В то же время до настоящего момента при оценивании параметров распределений с использованием группировки данных или проверке согласия по критерию Пирсона или родственному ему, а также при построении гистограмм область определения случайной величины разбивают на интервалы равной длины или равной вероятности. Причем в основном осуществляется разбиение на интервалы равной длины с последующим объединением тех смежных интервалов, в которые попало малое число наблюдений. Такое группирование рассматривается в наиболее часто используемых для руководства источниках.

Задачи (2.1.1-2.1.2) представляют собой задачи нелинейного программирования с допустимым множеством решений, задаваемым линейной системой ограничений, и нелинейной целевой функцией.

При вогнутой целевой функции эти задачи имеют единственное решение. Но в таких задачах вогнутость целевой функции обеспечивается не всегда, и в некоторых случаях задача асимптотически оптимального группирования оказывается многоэкстремальной. Задачи асимптотически оптимального группирования для конкретных законов распределения решались нами в работах [29, 31, 32, 37, 38, 39, 69, 70, 71, 72]. Аналогичная задача, связанная с оцениванием параметров регрессионных моделей по группированным наблюдениям и асимптотически оптимальным группированием при их оценивании, решалась в работах [33, 34, 35].

Вид системы ограничений позволяет достаточно просто, используя метод штрафных функций [175, 181, 61] свести условную задачу оптимизации (2.1.1) или (2.1.2) к безусловной. А для решения последней использовать различные алгоритмы поиска минимума функции n переменных [25, 26, 44, 47, 55, 111, 130, 135, 142,], в том числе методы сопряженных градиентов [137, 175, 68], вращающихся координат Розенброка [177], алгоритмы переменной метрики [177]. В процессе решения задач асимптотически оптимального группирования был сформирован комплекс программ оптимизации, который применялся для решения задач оптимизации из различных областей [49], используются в учебном процессе [80], а некоторые алгоритмы реализованы в программной системе [76].

В следующем разделе рассматривается задача асимптотически оптимального группирования для конкретных законов распределения, в основном, для тех, для которых решение получается в виде инвариантном относительно параметров распределения.

2.2. Решение задачи асимптотически оптимального группирования

В данном разделе представлены результаты решения задачи асимптотически оптимального группирования для законов распределения, наиболее часто используемых в приложениях, для которых решение удалось получить в виде, инвариантном относительно параметров распределения.

В прил. П1.1 приведены асимптотически оптимальные граничные точки для **экспоненциального распределения** в виде $t_i = \theta x_{(i)}$, максимизирующие информационное количество Фишера по группированным данным

$$J_{\Gamma}(\theta) = \frac{1}{\theta^2} \sum_{i=1}^k \frac{(t_i e^{-t_i} - t_{i-1} e^{-t_{i-1}})^2}{e^{-t_{i-1}} - e^{-t_i}},$$

где $t_i = \theta x_{(i)}$, а функция плотности экспоненциального распределения описывается выражением $f(y, \theta) = \theta e^{-\theta y}$. Для числа интервалов $k = 2, \dots, 6$ асимптотически оптимальные граничные точки приведены [65], в [69] эти данные были проверены и таблица продолжена до $k = 11$. В этой же таблице приложения даны значения относительной асимптотической информации, равной

$$A = J_{\Gamma}(\theta) / J_{\text{H}}(\theta),$$

где $J_{\text{H}}(\theta) = E \left(\frac{\partial}{\partial \theta} \ln [\theta e^{-\theta x}] \right)^2 = \frac{1}{\theta^2}$ - информационное количество Фишера о θ по негруппированным наблюдениям. В прил. П1.2 представлены соответствующие значения оптимальных вероятностей. Заметим, что величина A позволяет делать заключение о качестве проводимого группирования. Для сравнения дадим вычисленные значения относительной асимптотической информации при разбиении на интервалы равной вероятности. Оказалось, что $A = 0,8928$ для $k = 10$ и $A = 0,9462$ для $k = 20$, в то время как оптимальному группированию уже при $k = 10$ соответствует значение $A = 0,9798$.

На рис. 2.2.1 представлен график относительных потерь в информационном количестве Фишера при равновероятном группировании для экспоненциального распределения в зависимости от числа интервалов. Видно, что потери особенно велики при малом числе интервалов. Этот же график отражает и относительную величину возрастания асимптотической дисперсии оценки.

Информационное количество Фишера о параметре **распределения Рэлея** на одно наблюдение по группированным данным описывается выражением

$$J_{\Gamma}(\theta) = \frac{1}{\theta^2} \sum_{i=1}^k \frac{(t_{i-1}^2 e^{-t_{i-1}^2/2} - t_i^2 e^{-t_i^2/2})}{e^{-t_{i-1}^2/2} - e^{-t_i^2/2}},$$

где $t_i = x_{(i)} / \theta$. Информационное количество Фишера на одно наблюдение по негруппированным данным

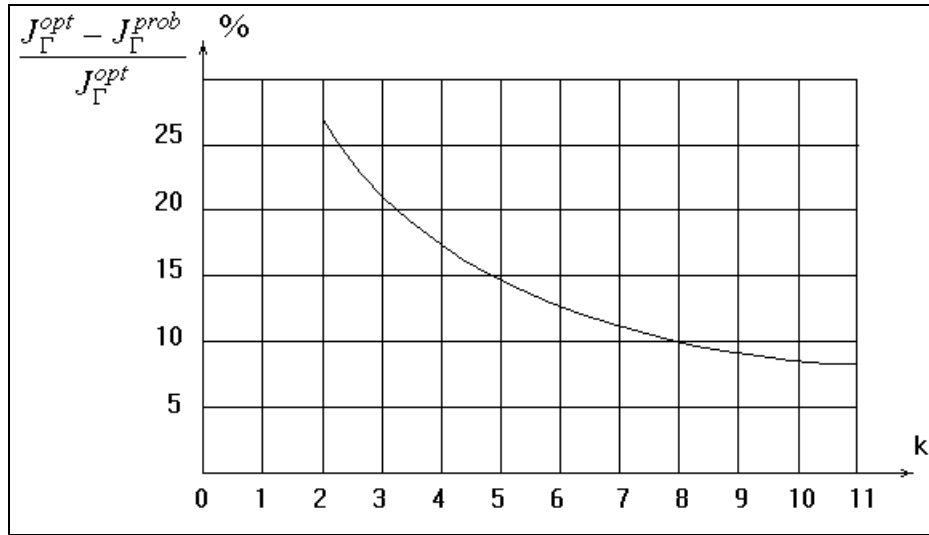


Рис. 2.2.1. Относительные потери в информационном количестве Фишера при равновероятном группировании по сравнению с асимптотически оптимальным группированием для экспоненциального распределения

$$J_H(\theta) = E \left(\frac{\partial}{\partial \theta} \ln \left[\frac{x}{\theta^2} e^{-x^2/2\theta^2} \right] \right)^2 = \frac{4}{\theta^2}.$$

Асимптотически оптимальные граничные точки интервалов группирования [69,29] в форме $t_i = x_{(i)}/\theta$, максимизирующие $J_\Gamma(\theta)$, приводятся в прил. П1.3. В нем же даны соответствующие значения относительной асимптотической информации $A = J_\Gamma(\theta)/J_H(\theta)$, позволяющей судить о качестве группирования. Значения вероятностей, которые соответствуют граничным точкам, содержатся в прил. П1.2.

Выражение для информационного количества Фишера о параметре **распределения Максвелла** по группированным наблюдениям имеет вид

$$J_\Gamma(\theta) = \frac{2}{\theta^2} \sum_{i=1}^k \frac{[t_{i-1}^3 \varphi(t_{i-1}) - t_i^3 \varphi(t_i)]}{\Phi_0(t_i) - t_i \varphi(t_i) - \Phi_0(t_{i-1}) + t_{i-1} \varphi(t_{i-1})},$$

где $t_i = x_{(i)}/\theta$; $\varphi(z) = e^{-z^2/2}/\sqrt{2\pi}$; $\Phi_0(t) = \int_0^t \varphi(z) dz$.

По негруппированным данным имеем

$$J_H(\theta) = E \left(\frac{\partial}{\partial \theta} \ln \left[\frac{2x^2}{\theta^3} \varphi\left(\frac{x}{\theta}\right) \right] \right)^2 = \frac{6}{\theta^2}.$$

Полученные асимптотически оптимальные граничные точки интервалов группирования [62,69] в форме $t_i = x_{(i)}/\theta$ и соответствующие значения относительной асимптотической информации $A = J_\Gamma(\theta)/J_H(\theta)$ приведены в прил. П1.4, а значения соответствующих вероятностей - в прил. П1.5.

Рассмотрим задачу асимптотически оптимального группирования для **распределения модуля многомерного нормального вектора**. В общем случае функция плотности этого распределения имеет вид

$$f(x, \theta) = \frac{2x^{m-1}e^{-x/2\theta^2}}{(2\theta^2)^{m/2}\Gamma(m/2)},$$

где $x \in (0, \infty)$; m - размерность соответствующего многомерного нормального распределения. При известном m распределение полностью определяется параметром $\theta \in (0, \infty)$. Для $m=1$ имеем полунормальное распределение, для $m=2$ - распределение Рэлея, для $m=3$ - распределение Максвелла.

В общем случае функция распределения этого закона при четном и нечетном m представляется соответственно выражениями:

$$F(x) = 1 - e^{-t^2/2} \left[1 + \frac{t^2}{2} + \frac{t^4}{8} + \dots + \frac{t^{m-2}}{2^{\frac{m}{2}-1}\Gamma(m/2)} \right]$$

и

$$F(x) = \Phi_0(t) - e^{-t^2/2} \left[\frac{t}{2^{1/2}\Gamma(3/2)} + \frac{t^3}{2^{3/2}\Gamma(5/2)} + \dots + \frac{t^{m-2}}{2^{\frac{m}{2}-1}\Gamma(m/2)} \right]$$

где $t = x/\theta$; $\Phi_0(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-z^2/2} dz$.

Информационное количество Фишера по группированным наблюдениям определяется соотношением

$$J_\Gamma(\theta) = \frac{1}{\theta^2 2^{m-2} [\Gamma(m)]^2} \sum_{i=1}^k \frac{\left(t_{i-1}^m e^{-t_{i-1}^2/2} - t_i^m e^{-t_i^2/2} \right)^2}{F(t_i) - F(t_{i-1})},$$

а для негруппированных - $J_H(\theta) = 2m/\theta^2$.

При $m=1$ для **полунормального распределения** информационное количество Фишера по группированным данным описывается соотношением

$$J_{\Gamma}(\theta) = \frac{2}{\pi\theta^2} \sum_{i=1}^k \frac{\left(t_{i-1}e^{-t_{i-1}^2/2} - t_i e^{-t_i^2/2}\right)^2}{\Phi_0(t_i) - \Phi_0(t_{i-1})},$$

по негруппированным - $J_H(\theta) = 2/\theta^2$.

Асимптотически оптимальные граничные точки интервалов группирования в виде $t_i = x_{(i)}/\theta$ представлены в прил. П1.6 вместе с соответствующими значениями $A = J_{\Gamma}(\theta)/J_H(\theta)$. Оптимальные вероятности приведены в прил. П1.7.

Для $m=4$ функция плотности примет вид

$$f(x, \theta) = x^3 e^{-x^2/2\theta^2} / 2\theta^4,$$

функция распределения

$$F(x) = 1 - e^{-t^2/2} \left(1 + t^2/2\right),$$

где $t = x/\theta$. Информационное количество Фишера по группированным данным о параметре θ записывается в форме

$$J_{\Gamma}(\theta) = \frac{1}{4\theta^2} \sum_{i=1}^k \frac{\left(t_{i-1}^4 e^{-t_{i-1}^2/2} - t_i^4 e^{-t_i^2/2}\right)^2}{e^{-t_{i-1}^2/2} \left(1 + \frac{t_{i-1}^2}{2}\right) - e^{-t_i^2/2} \left(1 + \frac{t_i^2}{2}\right)},$$

где $t_i = x_{(i)}/\theta$; по негруппированным наблюдениям - $J_H(\theta) = 8/\theta^2$.

Асимптотически оптимальные граничные точки, максимизирующие $J_{\Gamma}(\theta)$, представлены в прил. П1.8, соответствующие вероятности - в прил. П1.9.

Решения задачи асимптотически оптимального группирования при $m=5$ приведены в прил. П1.10-П1.11, для $m=6$ - в прил. П1.12-П1.13, для $m=7$ - в прил. П1.14-П1.15, для $m=8$ - в прил. П1.16-П1.17, для $m=9$ - в прил. П1.18-П1.19.

Функция плотности **распределения Вейбулла** (Вейбулла-Гнеденко) определяется двумя параметрами: основным и масштабным. В прил. П1.20 приведены асимптотически оптимальные граничные точки интервалов для оценивания основного параметра при известном масштабном. Они максимизируют информационное количество Фишера о θ по группированным данным

$$J_{\Gamma}(\theta) = \frac{1}{\theta^2} \sum_{i=1}^k \frac{\left(t_i \ln t_i e^{-t_i} - t_{i-1} \ln t_{i-1} e^{-t_{i-1}} \right)^2}{e^{-t_{i-1}} - e^{-t_i}},$$

где $t_i = \left(x_{(i)} / \theta_1 \right)^\theta$. Соответствующие вероятности даны в прил. П1.21.

Информационное количество Фишера по негруппированным данным о параметре θ определяется соотношением

$$\begin{aligned} J_{\text{H}}(\theta) &= E \left(\frac{\partial}{\partial \theta} \ln \left[\frac{\theta}{\theta_1} \left(\frac{x}{\theta_1} \right)^{\theta-1} e^{-(x/\theta_1)^\theta} \right] \right)^2 = \\ &= \frac{1}{\theta^2} \left(1 + \frac{\pi}{6} + c^2 - 2c \right) \cong \frac{1,8237}{\theta^2}, \end{aligned}$$

где $c = 0,5772$ - постоянная Эйлера. Значения относительной асимптотической информации A представлены в этих же приложениях.

Следует отметить, что задача асимптотически оптимального группирования в данном случае оказалась многоэкстремальной. В прил. П1.20 приведены значения граничных точек, соответствующие *глобальному* экстремуму.

Информационное количество Фишера по группированным наблюдениям о масштабном параметре θ_1 при известном параметре θ описываются выражением

$$J_{\Gamma}(\theta) = \frac{\theta^2}{\theta_1^2} \sum_{i=1}^k \frac{\left(t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i} \right)^2}{e^{-t_{i-1}} - e^{-t_i}},$$

где $t_i = \left(x_{(i)} / \theta_1 \right)^\theta$; по негруппированным данным -

$$J_{\text{H}}(\theta) = E \left(\frac{\partial}{\partial \theta_1} \ln \left[\frac{\theta}{\theta_1} \left(\frac{x}{\theta_1} \right)^{\theta-1} e^{-(x)^\theta} \right] \right)^2 = \frac{\theta^2}{\theta_1^2}.$$

Асимптотически оптимальные граничные точки в виде $t_i = \left(x_{(i)} / \theta_1 \right)^\theta$ для оценивания масштабного параметра θ_1 и соответствующие значения относительной асимптотической информации $A = J_{\Gamma}(\theta_1) / J_{\text{H}}(\theta_1)$ могут быть взяты из прил. П1.1, значения вероятностей - из прил. П1.2. [62,29].

При одновременном оценивании двух параметров распределения Вейбулла дисперсионная матрица оценки параметров определя-

ется информационной матрицей Фишера по группированным наблюдениям. Эта матрица имеет вид

$$M_{\Gamma}(\theta) = \begin{bmatrix} J_{\Gamma}(\theta) & J_{\Gamma}(\theta, \theta_1) \\ J_{\Gamma}(\theta, \theta_1) & J_{\Gamma}(\theta_1) \end{bmatrix},$$

$$\text{где } J_{\Gamma}(\theta, \theta_1) = \frac{1}{\theta_1} \sum_{i=1}^k \frac{(t_{i-1}e^{-t_{i-1}} - t_i e^{-t_i})(t_i e^{-t_i} \ln t_i - t_{i-1} e^{-t_{i-1}} \ln t_{i-1})}{e^{-t_{i-1}} - e^{-t_i}}.$$

Информационная матрица Фишера по негруппированным наблюдениям

$$M_{\text{H}}(\theta) = \begin{bmatrix} J_{\text{H}}(\theta) & J_{\text{H}}(\theta, \theta_1) \\ J_{\text{H}}(\theta, \theta_1) & J_{\text{H}}(\theta_1) \end{bmatrix},$$

$$\text{где } J_{\text{H}}(\theta, \theta_1) = E \left(\frac{\partial}{\partial \theta} \ln f(x, \theta) \cdot \frac{\partial}{\partial \theta_1} \ln f(x, \theta) \right) = \frac{1}{\theta_1} (c - 1), \quad c - \text{постоянная}$$

Эйлера. Асимптотически оптимальные граничные точки, максимизирующие определитель матрицы $M_{\Gamma}(\theta)$, и соответствующие значения относительной асимптотической информации

$$A = |M_{\Gamma}(\theta)| / |M_{\text{H}}(\theta)| = |M_{\Gamma}(\theta)| / 1,644934 \theta_1^2$$

представлены в прил. П1.22, а соответствующие оптимальному группированию вероятности - в прил. П1.23.

На рис. 2.2.2 показан выигрыш в относительной асимптотической информации при использовании оптимального группирования по сравнению с разбиением на интервалы равной вероятности для распределения Вейбулла.

Информационное количество Фишера по группированным данным о параметре μ нормального распределения описывается выражением

$$J_{\Gamma}(\mu) = \frac{1}{\sigma^2} \sum_{i=1}^k \frac{[\varphi(t_i) - \varphi(t_{i-1})]^2}{\Phi(t_i) - \Phi(t_{i-1})},$$

где $t_i = (x_{(i)} - \mu) / \sigma$; $\varphi(z) = e^{-z^2/2} / \sqrt{2\pi}$; $\Phi(t) = \int_{-\infty}^t \varphi(z) dz$. Асимптотически оптимальные граничные точки, максимизирующие $J_{\Gamma}(\mu)$, приве-

дены в прил. П1.24. Здесь же приведены значения относительной асимптотической информации $A = J_{\Gamma}(\mu)/J_{\text{H}}(\mu)$. Решение данной задачи впервые было дано в [65]. Вычисленные значения соответствующих оптимальных вероятностей представлены в прил. П1.25.

Для стандартного отклонения нормального распределения информационное количество Фишера по группированным наблюдениям равно

$$J_{\Gamma}(\sigma) = \frac{1}{\sigma^2} \sum_{i=1}^k \frac{[t_i \varphi(t_i) - t_{i-1} \varphi(t_{i-1})]^2}{\Phi(t_i) - \Phi(t_{i-1})}$$

где $t_i = (x_{(i)} - \mu)/\sigma$. Асимптотически оптимальные граничные точки, максимизирующие $J_{\Gamma}(\sigma)$ были получены в [65]. Эти значения проверены, исправлены при $k = 9$ и представлены в прил. П1.26. Здесь же даны значения относительной асимптотической информации $A = J_{\Gamma}(\sigma)/J_{\text{H}}(\sigma)$, где $J_{\text{H}}(\sigma) = 2/\sigma^2$ - информационное количество Фишера о σ по негруппированным данным. Соответствующие значения оптимальных вероятностей представлены в прил. П1.27.

Асимптотически оптимальные граничные точки интервалов, максимизирующие определитель информационной матрицы по группированным данным,

$$M_{\Gamma}(\mu, \sigma) = \begin{bmatrix} J_{\Gamma}(\mu) & J_{\Gamma}(\mu, \sigma) \\ J_{\Gamma}(\mu, \sigma) & J_{\Gamma}(\sigma) \end{bmatrix},$$

где

$$J_{\Gamma}(\mu, \sigma) = \frac{1}{\sigma^2} \sum_{i=1}^k \frac{[\varphi(t_i) - \varphi(t_{i-1})][t_i \varphi(t_i) - t_{i-1} \varphi(t_{i-1})]}{\Phi(t_i) - \Phi(t_{i-1})},$$

в форме $t_i = (x_{(i)} - \mu)/\sigma$ даны в прил. П1.28. Здесь же приведены соответствующие значения относительной асимптотической информации

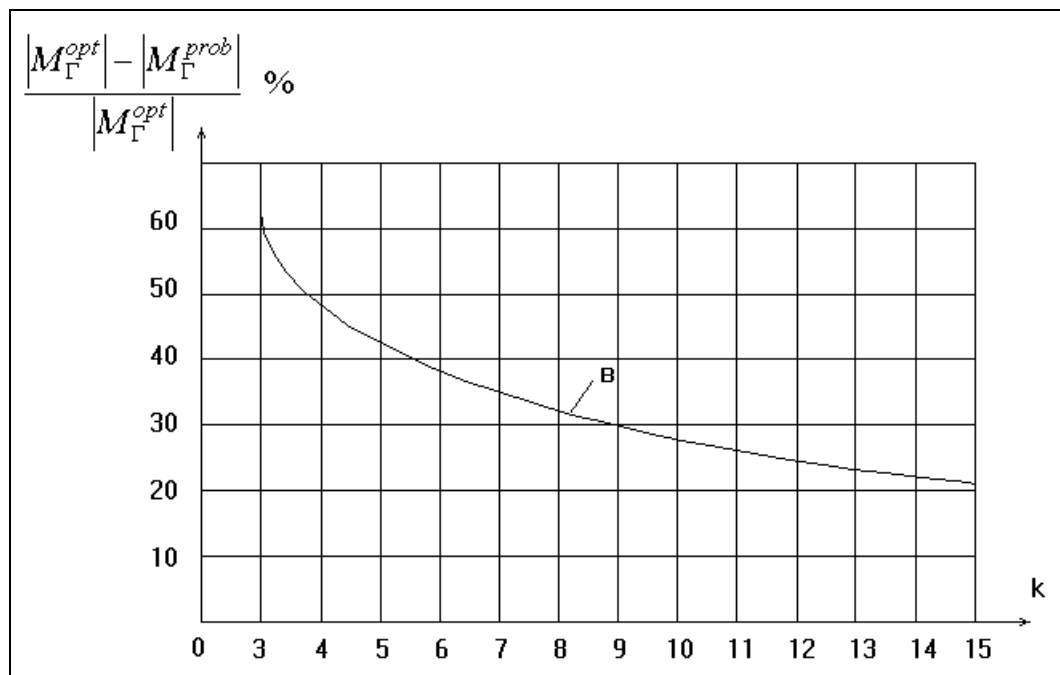
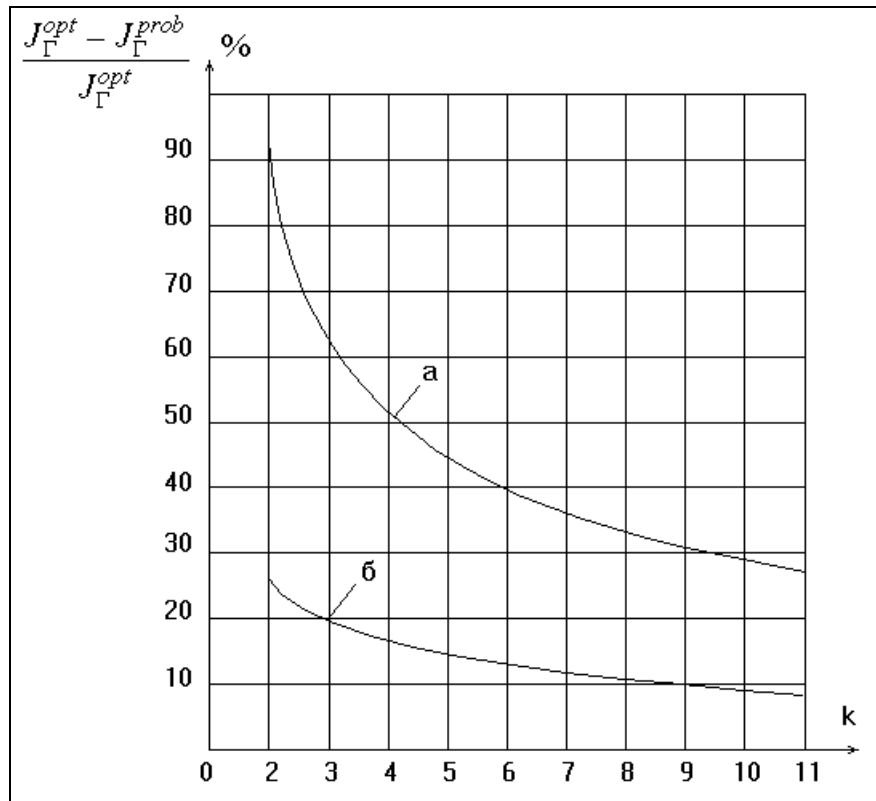


Рис. 2.2.2. Относительные потери в информационном количестве Фишера при равновероятном группировании по сравнению с асимптотически оптимальным группированием для распределения Вейбулла:
 а - при оценивании основного параметра;
 б - при оценивании масштабного параметра;
 в - при оценивании двух параметров.

$$A = |M_{\Gamma}(\mu, \sigma)| / |M_{\text{H}}(\mu, \sigma)|,$$

где $M_{\text{H}}(\mu, \sigma)$ - информационная матрица параметров нормального распределения по негруппированной выборке

$$M_{\text{H}}(\mu, \sigma) = \begin{bmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{bmatrix}.$$

Значения оптимальных вероятностей приведены в прил. П1.29.

Для **логарифмически нормальных распределений** асимптотически оптимальные граничные точки для оценивания параметров μ , σ или одновременно μ и σ (или для проверки гипотез) соответственно те же самые, что и для нормального распределения, и даны в прил. П1.24-П1.29. Но для распределения с плотностью

$$f(x, \mu, \sigma) = \frac{1}{\sigma x \sqrt{2\pi}} \exp \left\{ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right\}$$

они представлены там в виде $t_i = (\ln x_{(i)} - \mu)/\sigma$, а для распределения с плотностью

$$f(x, \mu, \sigma) = \frac{1}{\ln 10 \sigma x \sqrt{2\pi}} \exp \left\{ -\frac{(\lg x - \mu)^2}{2\sigma^2} \right\}$$

в форме $t_i = (\lg x_{(i)} - \mu)/\sigma$.

Приведенные на рис. 2.2.3 результаты сравнения отражают степень предпочтительности оптимального группирования по отношению к равновероятному для нормального закона распределения.

Информационное количество Фишера о параметре μ **распределения наибольшего экстремального значения** описывается выражением

$$J_{\Gamma}(\mu) = \frac{1}{\sigma^2} \sum_{i=1}^k \frac{(s_{i-1} e^{-s_{i-1}} - s_i e^{-s_i})^2}{e^{-s_i} - e^{-s_{i-1}}},$$

где $s_i = \exp[(\mu - x_{(i)})/\sigma]$ [69]. Асимптотически оптимальные граничные точки, максимизирующие $J_{\Gamma}(\mu)$, в виде $t_i = (x_{(i)} - \mu)/\sigma$ представлены в прил. П1.30. В нем же приведены соответствующие значения относительной асимптотической информации $A = J_{\Gamma}(\mu)/J_{\text{H}}(\mu)$, где

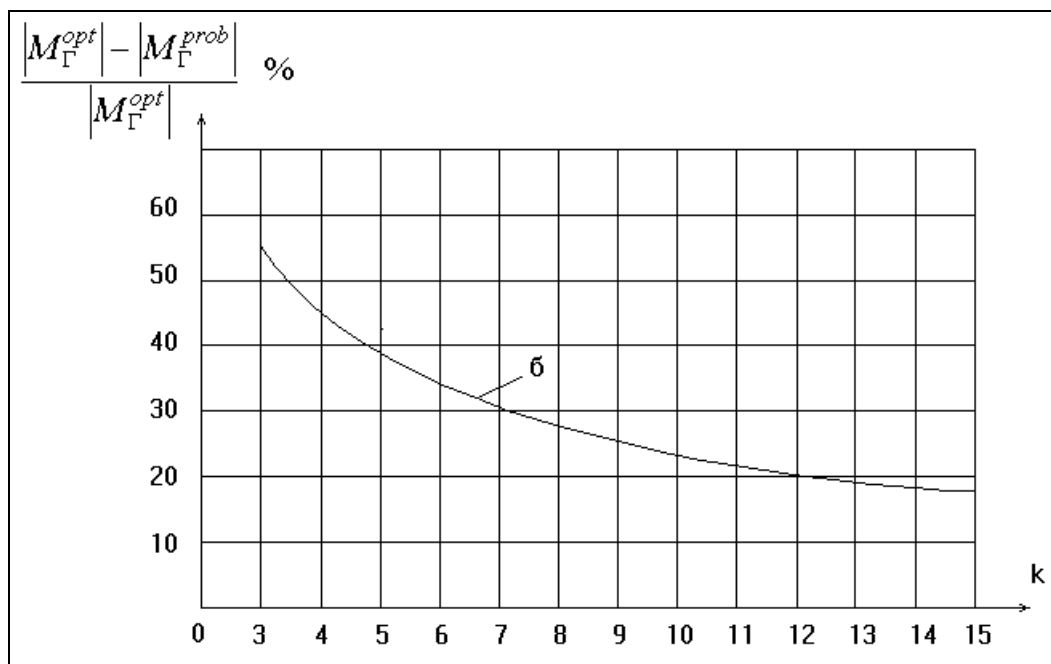
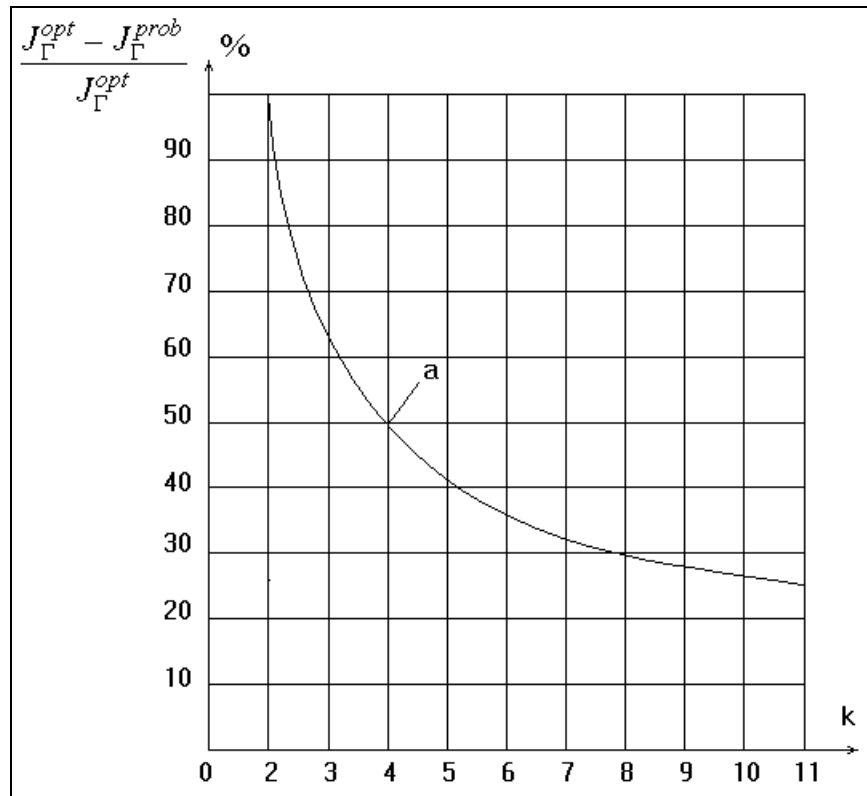


Рис. 2.2.3. Относительные потери в информационном количестве Фишера при равновероятном группировании по сравнению с асимптотически оптимальным группированием для нормального распределения

а - при оценивании стандартного отклонения; **б** - при оценивании математического ожидания и стандартного отклонения.

$$J_H(\mu) = E \left(\frac{\partial}{\partial \mu} \ln \left[\frac{1}{\sigma} \exp \left\{ \frac{\mu - x}{\sigma} - \exp \left(\frac{\mu - x}{\sigma} \right) \right\} \right] \right)^2 = \frac{1}{\sigma^2}$$

- информационное количество по негруппированной выборке. Вероятности, соответствующие оптимальному группированию, содержатся в прил. П1.31.

Информационное количество Фишера о параметре σ распределения наибольшего экстремального значения определяется соотношением

$$J_\Gamma(\sigma) = \frac{1}{\sigma^2} \sum_{i=1}^k \frac{\left(t_{i-1} e^{-t_{i-1}} e^{-e^{-t_{i-1}}} - t_i e^{-t_i} e^{-e^{-t_i}} \right)^2}{e^{-e^{-t_i}} - e^{-e^{-t_{i-1}}}},$$

где $t_i = (x_{(i)} - \mu) / \sigma$. Асимптотически оптимальные граничные точки, максимизирующие $J_\Gamma(\sigma)$, представлены в прил. П1.32 вместе с соответствующими значениями относительной асимптотической информации

$$A = J_\Gamma(\sigma) / J_H(\sigma),$$

где

$$J_H(\sigma) = E \left(\frac{\partial}{\partial \sigma} \ln \left[\frac{1}{\sigma} \exp \left\{ \frac{\mu - x}{\sigma} - \exp \left(\frac{\mu - x}{\sigma} \right) \right\} \right] \right)^2 = \left(1 - 2c + c^2 + \frac{\pi^2}{6} \right) / \sigma^2;$$

c - постоянная Эйлера. Значения оптимальных вероятностей приведены в прил. П1.33. В приложениях приведены значения, соответствующие *глобальному* экстремуму данной задачи асимптотически оптимального группирования, так как она оказалась многоэкстремальной.

Вид информационной матрицы Фишера по группированным наблюдениям для параметров распределения наибольшего экстремального значения определяется выражением

$$M_\Gamma(\mu, \sigma) = \begin{bmatrix} J_\Gamma(\mu) & J_\Gamma(\mu, \sigma) \\ J_\Gamma(\mu, \sigma) & J_\Gamma(\sigma) \end{bmatrix},$$

где

$$J_\Gamma(\mu, \sigma) = \frac{1}{\sigma^2} \sum_{i=1}^k \frac{\left(e^{-t_{i-1}} e^{-e^{-t_{i-1}}} - e^{-t_i} e^{-e^{-t_i}} \right) \left(t_{i-1} e^{-t_{i-1}} e^{-e^{-t_{i-1}}} - t_i e^{-t_i} e^{-e^{-t_i}} \right)}{e^{-e^{-t_i}} - e^{-e^{-t_{i-1}}}};$$

$t_i = (x_{(i)} - \mu) / \sigma$. Информационная матрица по негруппированным данным

$$M_H(\mu, \sigma) = \begin{bmatrix} J_H(\mu) & J_H(\mu, \sigma) \\ J_H(\mu, \sigma) & J_H(\sigma) \end{bmatrix},$$

где $J_H(\mu, \sigma) = E \left(\frac{\partial}{\partial \mu} \ln f(x, \mu, \sigma) \cdot \frac{\partial}{\partial \sigma} \ln f(x, \mu, \sigma) \right)^2 = (c-1)/\sigma^2$; c -

постоянная Эйлера. Асимптотически оптимальные граничные точки в виде $t_i = (x_{(i)} - \mu)/\sigma$, максимизирующие определитель матрицы $M_\Gamma(\mu, \sigma)$, представлены в прил. П1.34. Здесь же даны значения относительной асимптотической информации в виде соотношения

$$A = |M_\Gamma(\mu, \sigma)| / |M_H(\mu, \sigma)| = |M_\Gamma(\mu, \sigma)| / 1,644934\sigma^{-4}.$$

В прил. П1.35 показаны полученные оптимальные вероятности.

Выражение для информационного количества Фишера о параметре μ распределения **наименьшего экстремального значения** по группированной выборке имеет вид

$$J_\Gamma(\mu) = \frac{1}{\sigma^2} \sum_{i=1}^k \frac{(t_{i-1}e^{-t_{i-1}} - t_i e^{-t_i})^2}{e^{-t_{i-1}} - e^{-t_i}},$$

где $t_i = \exp\{(x_{(i)} - \mu)/\sigma\}$. В таком виде асимптотически оптимальные граничные точки, максимизирующие $J_\Gamma(\mu)$, представлены в прил. П1.1, а соответствующие вероятности в прил. П1.2.

Информационное количество о параметре σ имеет вид

$$J_\Gamma(\sigma) = \frac{1}{\sigma^2} \sum_{i=1}^k \frac{(s_{i-1}e^{s_{i-1}}e^{-e^{s_{i-1}}} - s_i e^{s_i}e^{-e^{s_i}})^2}{e^{-e^{s_{i-1}}} - e^{-e^{s_i}}},$$

где $s_i = (x_{(i)} - \mu)/\sigma$. Оптимальные граничные точки для этого случая представлены в прил. П1.20 в виде $t_i = \exp\{(x_{(i)} - \mu)/\sigma\}$, а соответствующие вероятности - в прил. П1.21.

Информационная матрица Фишера параметров распределения наименьшего значения

$$M_\Gamma(\mu, \sigma) = \begin{bmatrix} J_\Gamma(\mu) & J_\Gamma(\mu, \sigma) \\ J_\Gamma(\mu, \sigma) & J_\Gamma(\sigma) \end{bmatrix},$$

где

$$J_\Gamma(\mu, \sigma) = \frac{1}{\sigma^2} \sum_{i=1}^k \frac{(e^{s_{i-1}}e^{-e^{s_{i-1}}} - e^{s_i}e^{-e^{s_i}})(s_{i-1}e^{s_{i-1}}e^{-e^{s_{i-1}}} - s_i e^{s_i}e^{-e^{s_i}})}{e^{-e^{s_{i-1}}} - e^{-e^{s_i}}}$$

для $s_i = (x_{(i)} - \mu)/\sigma$. Информационная матрица по негруппированным данным совпадает с информационной матрицей Фишера для распределения наибольшего значения. Асимптотически оптимальные граничные точки, максимизирующие определитель $M_\Gamma(\mu, \sigma)$, в виде $t_i = \exp\{(x_{(i)} - \mu)/\sigma\}$ представлены в прил. П1.22, а оптимальные вероятности - в прил. П1.23.

Решение задачи асимптотически оптимального группирования для **распределения Коши** рассматривалось в [39]. Информационное количество Фишера по группированным наблюдениям о масштабном параметре θ распределения Коши описывается соотношением

$$J_\Gamma(\theta) = \frac{1}{\pi\theta^2} \sum_{i=1}^k \frac{\left(\frac{t_{i-1}}{1+t_{i-1}^2} - \frac{t_i}{1+t_i^2} \right)^2}{\operatorname{arctg} t_i - \operatorname{arctg} t_{i-1}},$$

где $t_i = (x_{(i)} - \theta_1)/\theta$. Полученные асимптотически оптимальные граничные точки, максимизирующие $J_\Gamma(\theta)$, представлены в прил. П1.36. В нем же приведены значения относительной асимптотической информации $A = J_\Gamma(\theta)/J_H(\theta)$, позволяющие судить о качестве группирования и величине потерь информации, связанных с группированием выборки. Информационное количество Фишера по негруппированной выборке определяется выражением

$$J_H(\theta) = E \left(\frac{\partial}{\partial \theta} \ln \left[\frac{1}{\pi} \cdot \frac{\theta}{\theta^2 + (x - \theta_1)^2} \right] \right)^2 = \frac{1}{2\theta^2}.$$

Значения оптимальных вероятностей, соответствующих рассматриваемому случаю, представлены в прил. П1.37.

Информационное количество о параметре θ_1 распределения Коши по группированной выборке

$$J_\Gamma(\theta_1) = \frac{1}{\pi\theta^2} \sum_{i=1}^k \frac{\left[1/(1+t_{i-1}^2) - 1/(1+t_i^2) \right]^2}{\operatorname{arctg} t_i - \operatorname{arctg} t_{i-1}},$$

где $t_i = (x_{(i)} - \theta_1)/\theta$. По негруппированным данным информационное количество Фишера определяется выражением

$$J_H(\theta_1) = E \left(\frac{\partial}{\partial \theta_1} \ln \left[\frac{1}{\pi} \cdot \frac{\theta}{\theta^2 + (x - \theta_1)^2} \right] \right)^2 = \frac{1}{2\theta^2}.$$

Асимптотически оптимальные граничные точки, максимизирующие $J_\Gamma(\theta_1)$, также получены и представлены в прил. П1.38. Здесь же даны значения относительной асимптотической информации $A = J_\Gamma(\theta_1)/J_H(\theta_1)$. Соответствующие значения оптимальных частот приведены в прил. П1.39.

Информационная матрица Фишера параметров распределения Коши по группированным наблюдениям

$$M_\Gamma(\theta) = \begin{bmatrix} J_\Gamma(\theta) & J_\Gamma(\theta, \theta_1) \\ J_\Gamma(\theta, \theta_1) & J_\Gamma(\theta_1) \end{bmatrix}$$

$$\text{где } J_\Gamma(\theta, \theta_1) = \frac{1}{\pi\theta^2} \sum_{i=1}^k \frac{\left(\frac{t_{i-1}}{1+t_{i-1}^2} - \frac{t_i}{1+t_i^2} \right) \left(\frac{1}{1+t_{i-1}^2} - \frac{1}{1+t_i^2} \right)}{\arctg t_i - \arctg t_{i-1}} \text{ для } t_i = \frac{(x_{(i)} - \theta_1)}{\theta}.$$

Асимптотически оптимальные групповые пределы, максимизирующие определитель матрицы $M_\Gamma(\theta)$, представлены в прил. П1.40. Для сравнения со случаем негруппированной выборки в этом же приложении приведены значения относительной асимптотической информации $A = |M_\Gamma(\theta)|/|M_H(\theta)|$, где определитель информационной матрицы по негруппированным наблюдениям $|M_H(\theta)| = 0,25\theta^4$.

В отличие от всех ранее рассмотренных случаев оптимальное группирование относительно двух параметров распределения Коши разбивает область определения случайной величины на интервалы, которым соответствуют *равные вероятности*.

Информационное количество Фишера о параметре μ **логистического распределения** по группированной выборке определяется выражением

$$J_\Gamma(\mu) = \frac{\pi^2}{3\sigma^2} \sum_{i=1}^k \frac{\left[\frac{e^{-t_i}}{(1+e^{-t_i})^2} - \frac{e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2} \right]}{\left(1+e^{-t_i}\right)^{-1} - \left(1+e^{-t_{i-1}}\right)^{-1}},$$

где $t_i = \pi(x_{(i)} - \mu)/\sigma\sqrt{3}$. Максимизирующие его оптимальные граничные точки приведены в прил. П1.41. Информационное количество по негруппированным наблюдениям

$$J_H(\mu) = E \left[\frac{\partial}{\partial \mu} \ln \left[\frac{\pi}{\sigma\sqrt{3}} \cdot \frac{\exp\{-\pi(x - \mu)/\sigma\sqrt{3}\}}{(1 + \exp\{-\pi(x - \mu)/\sigma\sqrt{3}\})^2} \right] \right]^2 = \frac{\pi^2}{9\sigma^2}.$$

Кроме граничных точек в прил. П1.41 представлены значения относительной асимптотической информации $A = J_\Gamma(\mu)/J_H(\mu)$, позволяющие судить о потерях информации, связанных с группированием наблюдений. Соответствующие значения вероятностей оказались равны: оптимальное группирование совпало с разбиением на интервалы *равной вероятности*.

Информационное количество Фишера о параметре σ логистического распределения по группированной выборке имеет вид

$$J_\Gamma(\sigma) = \frac{1}{\sigma^2} \sum_{i=1}^k \frac{\left[\frac{t_i e^{-t_i}}{(1 + e^{-t_i})^2} - \frac{t_{i-1} e^{-t_{i-1}}}{(1 + e^{-t_{i-1}})^2} \right]}{(1 + e^{-t_i})^{-1} - (1 + e^{-t_{i-1}})^{-1}},$$

а по негруппированным наблюдениям

$$J_H(\sigma) = E \left[\frac{\partial}{\partial \sigma} \ln \left[\frac{\pi}{\sigma\sqrt{3}} \cdot \frac{\exp\{-\pi(x - \mu)/\sigma\sqrt{3}\}}{(1 + \exp\{-\pi(x - \mu)/\sigma\sqrt{3}\})^2} \right] \right]^2 = \frac{\pi^2 + 3}{9\sigma^2}.$$

Значения асимптотически оптимальных граничных точек, максимизирующих $J_\Gamma(\sigma)$, и значения относительной асимптотической информации $A = J_\Gamma(\sigma)/J_H(\sigma)$ представлены в прил. П1.42, а значения соответствующих вероятностей - в прил. П1.43.

Информационная матрица Фишера по группированным наблюдениям о параметрах логистического распределения

$$M_\Gamma(\mu, \sigma) = \begin{bmatrix} J_\Gamma(\mu) & J_\Gamma(\mu, \sigma) \\ J_\Gamma(\mu, \sigma) & J_\Gamma(\sigma) \end{bmatrix},$$

$$\text{где } J_{\Gamma}(\mu, \sigma) = \frac{\pi^2}{\sigma^2 \sqrt{3}} \sum_{i=1}^k \frac{\left[\frac{e^{-t_i}}{(1+e^{-t_i})^2} - \frac{e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2} \right] \left[\frac{t_i e^{-t_i}}{(1+e^{-t_i})^2} - \frac{t_{i-1} e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2} \right]}{(1+e^{-t_i})^{-1} - (1+e^{-t_{i-1}})^{-1}},$$

а информационная матрица по негруппированным данным

$$M_H(\mu, \sigma) = \begin{bmatrix} J_H(\mu) & J_H(\mu, \sigma) \\ J_H(\mu, \sigma) & J_H(\sigma) \end{bmatrix},$$

$$\text{где } J_H(\mu, \sigma) = E \left(\frac{\partial}{\partial \mu} \ln f(x, \mu, \sigma) \frac{\partial}{\partial \sigma} \ln f(x, \mu, \sigma) \right) = 0.$$

Асимптотически оптимальные граничные точки и соответствующие значения $A = |M_G(\mu, \sigma)| / |M_U(\mu, \sigma)|$ относительной асимптотической информации представлены в прил. П1.44, а оптимальные значения вероятностей попадания в интервалы - в прил. П1.45.

Информационное количество Фишера по группированным данным о параметре θ **гамма-распределения** определяется соотношением

$$J_{\Gamma}(\theta) = \sum_{i=1}^k \frac{\left(\int_{t_{i-1}}^{t_i} t^{\theta-1} e^{-t} \ln t \, dt - \Psi(\theta) \int_{t_{i-1}}^{t_i} t^{\theta-1} e^{-t} \, dt \right)^2}{\Gamma(\theta) \int_{t_{i-1}}^{t_i} t^{\theta-1} e^{-t} \, dt},$$

где $t_i = \theta_1 x_{(i)}$. В таком виде оптимальные граничные точки в зависимости от значения основного параметра θ (формы) представлены в прил. П1.53. Для сравнения с негруппированным случаем здесь же приведены значения относительной асимптотической информации $A = J_{\Gamma}(\theta) / J_H(\theta)$, где информационное количество по негруппированным данным $J_H(\theta) = 1,82368 \theta^2$. Эта задача впервые решалась в [31].

В случае гамма-распределения решение задачи оптимального группирования не может быть получено в виде инвариантного относительно параметров распределения. Оптимальные граничные точки являются функциями параметра θ . В принципе, граничные точки при значениях θ , не совпадающих с приводимыми в прил. П1.53, могут быть получены с помощью интерполяции. Это затрудняет использование готового решения в различных постановках задачи оценивания

неизвестных параметров гамма-распределения, делает весьма проблематичной ее необходимость. Но по отношению к критериям согласия актуальность использования асимптотически оптимального группирования сохраняется.

Для масштабного параметра θ_1 информационное количество по группированным данным находится из выражения

$$J_{\Gamma}(\theta_1) = \frac{1}{\theta^2 \Gamma(\theta)} \sum_{i=1}^k \frac{(t_i^\theta e^{-t_i} - t_{i-1}^\theta e^{-t_{i-1}})^2}{\int_{t_{i-1}}^{t_i} t^{\theta-1} e^{-t} dt},$$

где $t_i = \theta_1 x_{(i)}$; $\Gamma(\theta)$ - гамма-функция Эйлера. Информационное количество по негруппированным выборкам -

$$J_{\text{H}}(\theta_1) = E \left(\frac{\partial}{\partial \theta_1} \ln [\theta_1^\theta x^{\theta-1} e^{-\theta_1 x} / \Gamma(\theta)] \right)^2 = \frac{\theta}{\theta_1^2}.$$

Эта задача решалась в работе [37]. Асимптотически оптимальные граничные точки в виде $t_i = \theta_1 x_{(i)}$ в зависимости от основного параметра θ представлены в прил. П1.54. В нем же для сравнения с негруппированной выборкой приведены значения относительной асимптотической информации $A = J_{\Gamma}(\theta_1) / J_{\text{H}}(\theta_1)$.

Информационное количество Фишера по группированным данным о параметре θ **распределения Парето** дается соотношением

$$J_{\Gamma}(\theta) = \frac{1}{\theta^2} \sum_{i=1}^k \frac{(s_{i-1} \ln s_{i-1} - s_i \ln s_i)^2}{s_{i-1} - s_i},$$

где $s_i = (x_{(i)} / \alpha)^{-\theta}$; α - параметр сдвига распределения. По негруппированной выборке информационное количество о параметре θ имеет вид

$$J_{\text{H}}(\theta) = E \left(\frac{\partial}{\partial \theta} \ln [\alpha^\theta \theta x^{-(\theta+1)}] \right)^2 = \frac{1}{\theta^2}.$$

Асимптотически оптимальные граничные точки в виде $t_i = \theta \ln(x_{(i)} / \alpha)$ и соответствующие значения $A = J_{\Gamma}(\theta) / J_{\text{H}}(\theta)$ содержатся в прил. П1.1, а оптимальные вероятности - в прил. П1.2.

Количество информации Фишера о параметре **распределения Лапласа** по группированным наблюдениям определяется выражением

$$J_{\Gamma}(\theta) = \frac{1}{\theta^2} \sum_{i=1}^k A_i,$$

$$\text{где } A_i = \begin{cases} \frac{1}{2} \frac{(t_i e^{t_i} - t_{i-1} e^{-t_{i-1}})^2}{e^{t_i} - e^{-t_{i-1}}}, & t_i < 0, \\ \frac{1}{2} \frac{(t_i e^{-t_i} - t_{i-1} e^{-t_{i-1}})^2}{e^{-t_{i-1}} - e^{-t_i}}, & t_{i-1} > 0, \\ \frac{1}{4} \frac{(t_i e^{-t_i} - t_{i-1} e^{-t_{i-1}})^2}{1 - (e^{-t_i} - e^{t_{i-1}})/2}, & t_{i-1} < 0, t_i > 0; \end{cases}$$

и $t_i = \theta x_{(i)}$. Количество информации Фишера по негруппированным наблюдениям определяется величиной

$$J_H(\theta) = E \left(\frac{\partial}{\partial \theta} \ln \left[\frac{\theta}{2} e^{-\theta(x)} \right] \right)^2 = \frac{1}{\theta^2}.$$

Решение данной задачи оптимального группирования представлено в прил. П1.46, П1.47, П1.48, П1.49.

Функция плотности двойного показательного распределения

$$f(x, \theta) = \theta_1 \theta_2 e^{-\theta_1 x - \theta_2 e^{-\theta_1 x}},$$

где $x \in (-\infty, \infty)$; $\theta_1 \theta_2 \in (0, \infty)$, определяется двумя параметрами. Вероятность попадания наблюдения в интервал значений

$$p_i(\theta) = e^{-t_i} - e^{-t_{i-1}},$$

где $t_i = \theta_2 e^{-\theta_1 x_{(i)}}$. Информационное количество Фишера о параметре θ_2 по группированным наблюдениям имеет вид

$$J_{\Gamma}(\theta_2) = \frac{1}{\theta_2^2} \sum_{i=1}^k \frac{(t_i e^{-t_i} - t_{i-1} e^{-t_{i-1}})^2}{e^{-t_i} - e^{-t_{i-1}}},$$

а по негруппированным - $J_H(\theta_2) = 1/\theta_2^2$.

При решении задачи асимптотически оптимального группирования для параметра θ_2 максимизировалась относительная асимптотическая информация $A = J_{\Gamma}(\theta_2)/J_H(\theta_2)$. Соответствующие асимптотически оптимальные граничные точки представлены в прил. П1.1 в виде $t_i = \theta_2 \exp\{-\theta_1 x_{(i)}\}$ вместе со значениями A , а вероятности попадания в интервал в прил. П1.31.

Информационное количество Фишера о параметре θ_1 по группированным данным при $t_i = \theta_2 \exp\{-\theta_1 x_{(i)}\}$ -

$$J_{\Gamma}(\theta_1) = \frac{1}{\theta_1^2} \sum_{i=1}^k \frac{[t_i \ln(t_i/\theta_2) e^{-t_i} - t_{i-1} \ln(t_{i-1}/\theta_2) e^{-t_{i-1}}]^2}{e^{-t_i} - e^{-t_{i-1}}},$$

а по негруппированным -

$$J_{\text{H}}(\theta_1) = \left[\ln^2 \theta_2 + 2(c-1) \ln \theta_2 + (c-1)^2 + \frac{\pi^2}{6} \right] / \theta_1^2,$$

где c - постоянная Эйлера. Относительная асимптотическая информация $A = J_{\Gamma}(\theta_1)/J_{\text{H}}(\theta_1)$ в данном случае *зависит* от θ_2 . Решения задачи асимптотически оптимального группирования получено при различных значениях θ_2 и представлены в прил. П1.50, а соответствующие вероятности в прил. П1.51.

При одновременном оценивании двух параметров распределения рассматривается информационная матрица Фишера по группированным данным

$$M_{\Gamma}(\theta_1, \theta_2) = \begin{bmatrix} J_{\Gamma}(\theta_1) & J_{\Gamma}(\theta_1, \theta_2) \\ J_{\Gamma}(\theta_1, \theta_2) & J_{\Gamma}(\theta_2) \end{bmatrix},$$

$$\text{где } J_{\Gamma}(\theta_1, \theta_2) = \frac{1}{\theta_1 \theta_2} \sum_{i=1}^k \frac{\left(t_i \ln \frac{t_i}{\theta_2} e^{-t_i} - t_{i-1} \ln \frac{t_{i-1}}{\theta_2} e^{-t_{i-1}} \right) (t_i e^{-t_i} - t_{i-1} e^{-t_{i-1}})}{e^{-t_i} - e^{-t_{i-1}}}.$$

Информационная матрица Фишера по негруппированным наблюдениям -

$$M_{\text{H}}(\theta_1, \theta_2) = \begin{bmatrix} J_{\text{H}}(\theta_1) & J_{\text{H}}(\theta_1, \theta_2) \\ J_{\text{H}}(\theta_1, \theta_2) & J_{\text{H}}(\theta_2) \end{bmatrix},$$

где $J_{\text{H}}(\theta_1, \theta_2) = 1 - c - \ln \theta_2$.

При решении задачи асимптотически оптимального группирования максимизировалась величина $A = |M_{\Gamma}(\theta_1, \theta_2)| / |M_{\text{H}}(\theta_1, \theta_2)|$, причем определитель информационной матрицы по негруппированным данным $|M_{\text{H}}(\theta_1, \theta_2)| = \frac{1,6449}{\theta_1^2 \theta_2^2}$. Решение задачи приведено в прил. П1.52, а

соответствующие вероятности - в прил. П1.35. Подчеркнем, что это решение получается в виде инвариантно относительно параметров распределения: граничные точки в таблице не зависят от θ_2 .

Выводы

Таким образом, для широкого ряда распределений при решении задач асимптотически оптимального группирования удалось получить граничные точки интервалов в виде, инвариантном относительно параметров распределений и на их основе сформировать таблицы асимптотически оптимального группирования. В совокупности были получены таблицы для распределений экспоненциального, полунормального, Рэлея, Максвелла, модуля многомерного нормального вектора, Парето, Эрланга, Лапласа, нормального, логарифмически-нормальных ($\ln$ и $\lg$), Коши, Вейбулла, распределений минимального и максимального значения, двойного показательного, гамма-распределения. В общей сложности, 54 таблицы оптимальных граничных точек и соответствующих вероятностей, которые приведены в приложении П1.

2.3. О возможности практического использования асимптотически оптимального группирования в задачах оценивания параметров

Возможность эффективного использования оптимального группирования при оценивании параметров распределений в первый момент всегда вызывает сомнения.

Первое возражение связано с доводом, а зачем вообще группировать наблюдения и терять информацию? Современный уровень вычислительной техники и запоминающих устройств для хранения информации позволяют не думать о группировании, как способе сжатия данных или ускорения вычислений при оценивании параметров. Всё справедливо, если эти наблюдения принадлежат действительно тому закону, параметры которого мы оцениваем. Но если наблюдаемая выборка содержит отклонения от наших предположений, то это может существенно отразиться на получаемых оценках. В главе 3 предлагается для вычисления робастных оценок максимального правдоподобия использовать предварительное группирование наблюдений и вычислять ОМП уже по группированным данным. Это позволяет получать оценки, устойчивые к наличию аномальных наблюдений или другим отклонениям от предположений. И использование асимптотически оптимального группирования в этом случае приносит положительные результаты. С другой стороны, в случае

больших выборок вычисление ОМП по группированным данным оказывается экономичней по сравнению с поиском оценок непосредственно по несгруппированным данным. Далее, давно известно, что оценки параметров, вычисляемые по выборочным квантилям также являются робастными. Предлагаемые нами оценки, использующие выборочные квантили, соответствующие асимптотически оптимальному группированию, рассматриваются в последующих разделах данной главы.

Второе возражение определяется тем, что истинные значения оцениваемых параметров нам неизвестны, а асимптотически оптимальные граничные точки зависят от параметров. Как же группировать? Если в нашем распоряжении выборка, состоящая из индивидуальных наблюдений, и к группированию мы прибегаем для того, чтобы получить робастные ОМП по группированным данным, то разбивая выборку на интервалы мы должны воспользоваться таблицами оптимальных вероятностей, разделяя выборку на части, пропорциональные им. Если же группирование вызвано технологическими причинами и возможна фиксация только принадлежности наблюдения некоторому интервалу значений, то выбор граничных точек возможен по прогнозируемым значениям параметров. В большинстве практических ситуаций имеются некоторые априорные сведения о параметрах закона. В таком случае существует возможность выбора оптимальных граничных точек (шкалирование) относительно прогнозируемого значения параметра. Эти априорные сведения могут поступать от предыдущих экспериментов или следовать из теоретических или практических предпосылок.

Конечно, используемые априорные сведения о параметре могут существенно отличаться от его истинного значения. Тогда оптимальное группирование по прогнозируемому значению параметра приводит к потере в относительной асимптотической информации. В этом случае интересно знать, при каких ошибках прогноза имеет смысл проводить оптимальное группирование по априорным сведениям о параметре.

Величина потерь в относительной асимптотической информации может быть рассчитана [32,36]. Например, выражение для потерь в относительной асимптотической информации для оценки математического ожидания нормального распределения имеет вид

$$\Delta A(\lambda) = 1 - \frac{\sum_{i=1}^k \frac{[\varphi(t_{i-1} - \lambda) - \varphi(t_i - \lambda)]^2}{\Phi(t_i - \lambda) - \Phi(t_{i-1} - \lambda)}}{\sum_{i=1}^k \frac{[\varphi(t_{i-1}) - \varphi(t_i)]^2}{\Phi(t_i) - \Phi(t_{i-1})}},$$

где $x_{(i)} = (t_i - \lambda)\sigma + \mu$ - оптимальные граничные точки по ошибочному прогнозу;

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}; \quad \Phi(t) = \int_{-\infty}^t \varphi(z) dz.$$

Вычисленные значения потерь в зависимости от ошибки прогнозирования $+\lambda\sigma$ и оцениваемого параметра μ представлены в табл. 2.3.1.

Таблица 2.3.1

Потери в относительной асимптотической информации при оценивании математического ожидания нормального распределения по группированным данным вследствие ошибочного прогноза, %

λ	k								
	2	3	4	5	6	7	8	9	10
2,0	79,40	61,53	48,76	39,64	32,95	27,89	23,27	20,86	18,35
1,9	75,75	57,11	44,43	35,63	29,30	24,59	20,98	18,14	15,87
1,8	71,75	52,57	40,14	31,76	25,85	21,51	18,22	15,66	13,63
1,7	67,37	47,97	35,96	28,07	22,61	18,66	15,70	13,41	11,61
1,6	62,69	43,35	31,91	24,59	19,61	16,05	13,41	11,39	9,81
1,5	57,73	38,78	28,03	21,32	16,84	13,67	11,35	9,59	8,22
1,4	52,56	34,29	24,35	18,30	14,31	11,53	9,51	7,80	6,82
1,3	47,24	29,95	20,90	15,52	12,03	9,62	7,89	6,60	5,61
1,2	41,83	25,79	17,70	12,99	9,98	7,93	6,47	5,38	4,55
1,1	36,42	21,87	14,77	10,72	8,17	6,45	5,23	4,33	3,65
1,0	31,10	18,19	12,11	8,70	6,58	5,16	4,17	3,44	2,89
0,9	15,95	14,81	9,72	6,92	5,19	4,05	3,26	2,68	2,24
0,8	21,05	11,75	7,61	5,37	4,01	3,11	2,49	2,04	1,70
0,7	16,50	9,02	5,78	4,04	3,00	2,32	1,85	1,51	1,25
0,6	12,37	6,64	4,21	2,92	2,16	1,66	1,32	1,07	0,89
0,5	8,74	4,62	2,90	2,00	1,47	1,13	0,89	0,89	0,60
0,4	5,67	2,96	1,84	1,27	0,93	0,71	0,56	0,45	0,38
0,3	3,22	1,66	1,03	0,71	0,51	0,39	0,31	0,25	0,21
0,2	1,44	0,74	0,46	0,31	0,22	0,17	0,14	0,11	0,09
0,1	0,36	0,18	0,11	0,08	0,06	0,04	0,03	0,03	0,02

Потери в относительной асимптотической информации при оценивании среднеквадратичного отклонения вследствие ошибочного прогноза при известном математическом ожидании определяется соотношением

$$\Delta A(\lambda) = 1 - \frac{\sum_{i=1}^k \frac{[\lambda t_i \Phi(\lambda t_i) - \lambda t_{i-1} \Phi(\lambda t_{i-1})]^2}{\Phi(\lambda t_i) - \Phi(\lambda t_{i-1})}}{\sum_{i=1}^k \frac{[t_i \Phi(t_i) - t_{i-1} \Phi(t_{i-1})]^2}{\Phi(t_i) - \Phi(t_{i-1})}}.$$

Оптимальные граничные точки по ошибочному прогнозу в данном случае имеют вид $x_{(i)} = \lambda \sigma t_i + \mu$, где μ - известное математическое ожидание; $\lambda \sigma$ - предполагаемое значение стандартного отклонения. Величины потерь в зависимости от λ представлены в табл. 2.3.2.

Таблица 2.3.2

Потери в относительной асимптотической информации при оценивании стандартного отклонения нормального распределения по группированным данным вследствие ошибочного прогноза, %

λ	k								
	2	3	4	5	6	7	8	9	10
0,1	97,43	89,91	89,31	88,90	88,38	87,92	87,48	87,07	86,70
0,2	89,99	78,03	76,50	75,46	74,22	73,12	72,10	71,14	70,28
0,3	73,47	64,89	62,42	60,71	58,81	57,12	55,61	54,17	62,93
0,4	64,10	51,17	48,05	45,26	43,59	41,56	39,80	38,14	36,74
0,5	48,39	37,67	34,40	32,07	29,82	27,80	26,13	24,54	23,26
0,6	32,95	25,24	22,35	27,27	18,40	16,71	15,39	14,14	13,16
0,7	19,31	14,67	12,58	11,06	9,79	8,65	7,80	6,99	6,40
0,8	8,77	6,44	5,52	4,69	4,06	3,48	3,08	2,69	2,42
0,9	2,20	1,67	1,35	1,10	0,98	0,78	0,68	0,58	0,81
1,0	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
1,1	2,09	1,62	1,24	0,96	0,78	0,62	0,52	0,43	0,38
1,2	7,33	6,26	4,42	3,51	2,84	2,22	1,87	1,53	1,32
1,3	16,69	13,41	10,02	7,27	5,86	4,53	3,80	3,09	2,65
1,4	27,34	22,42	16,68	11,93	9,59	7,38	6,18	5,01	4,30
1,5	38,84	32,54	24,28	17,26	13,88	10,69	8,95	7,25	6,24
1,6	50,28	43,05	32,41	23,10	18,64	14,40	12,08	9,81	8,45
1,7	60,92	53,31	40,71	29,35	23,81	18,49	15,55	12,68	10,93
1,8	70,28	62,82	48,88	35,89	29,31	22,93	19,35	15,83	13,68
1,9	78,11	71,22	56,65	42,60	35,07	27,67	23,44	19,27	16,69
2,0	84,39	78,35	63,84	49,36	40,99	32,67	27,80	22,97	19,95
2,1	89,21	84,16	70,33	55,98	46,97	37,85	32,37	26,90	23,43

Приводимые в таблицах результаты вычислений позволяют сделать вывод, что при умеренных ошибках прогноза о значении оцениваемого параметра потери в относительной асимптотической информации, а, следовательно, и потери в информации, вызываемые груп-

пированием наблюдений, имеют относительно малые величины при малых ошибках прогноза, особенно при росте числа интервалов группирования.

Иногда при группировании пользуются равноотстоящими граничными точками. При таком группировании представляет интерес оптимальная длина интервала. В табл. 2.3.3. приведены оптимальные длины интервалов при равноотстоящем группировании в виде $\Delta t = \Delta x / \theta$ для распределений Рэлея и Максвелла, и в форме $\Delta t = (\Delta x / \theta_1)^0$ для распределения Вейбулла, а также соответствующие значения относительной асимптотической информации A . Если сравнить достигнутые значения A со значениями, соответствующими асимптотически оптимальному группированию, то видно, что и такой способ группирования имеет определенный смысл.

Таблица 2.3.3
Оптимальные длины интервалов группирования
для оценивания параметров распределений
Рэлея, Максвелла, Вейбулла

k	Распределения					
	Рэлея		Максвелла		Вейбулла	
	Δt	A	Δt	A	Δt	A
2	1,7853	0,6476	2,0451	0,6451	1,5936	0,6476
3	1,1288	0,7808	1,2199	0,7639	1,2070	0,8076
4	0,8510	0,8696	0,9179	0,8529	0,9898	0,8749
5	0,6894	0,9113	0,7421	0,9000	0,8474	0,9105
6	0,5827	0,9357	0,6255	0,9276	0,7456	0,9320
7	0,5063	0,9517	0,5422	0,9451	0,6687	0,9461
8	0,4484	0,9616	0,4805	0,9568	0,6080	0,9560
9	0,4080	0,9689	0,4305	0,9651	0,5588	0,9632
10	0,3664	0,9743	0,3910	0,9713	0,5180	0,9687

Таким образом, можно выделить следующие моменты, когда применение асимптотически оптимального группирования может дать положительный эффект в задачах точечного оценивания параметров.

Во-первых, при группировании информации в целях сокращения объемов хранимых или передаваемых по каналам связи данных о наблюдениях случайных величин с минимальными потерями информации об этом законе.

Во-вторых, для предварительного группирования исходных негруппированных выборок с дальнейшим вычислением робастных

ОМП по группированным данным. Подробно этот вопрос рассмотрен в главе 3.

В-третьих, в случае больших выборок для приближенного вычисления квантилей, соответствующих асимптотически оптимальному группированию, и последующем вычислении по ним оценок параметров (также робастных). Соответствующие вопросы рассматриваются в следующих параграфах этой главы.

Вопросы использования асимптотически оптимального группирования в задачах статистического анализа обсуждались в работах [29, 38, 40, 41, 52, 101, 105].

2.4. Оценки параметров распределений по выборочным асимптотически оптимальным квантилям

Для ряда распределений решение задачи асимптотически оптимального группирования удалось получить в виде, инвариантном относительно параметров распределений. Это даёт возможность при больших выборках использовать значения выборочных квантилей для вычисления оценок [41, 42]. В сформированных таблицах асимптотически оптимального группирования граничные точки содержатся в виде $t_i = \varphi(x_i, \theta)$. Если упорядочить выборку и разбить её на интервалы так, чтобы число наблюдений, попавших в i -й интервал было пропорционально оптимальной частоте $n_i = NP_i$, то в результате будут найдены выборочные квантили $\hat{x}_{(i)}$, соответствующие асимптотически оптимальному группированию. Например, в качестве $\hat{x}_{(i)}$ можно брать средние значения между смежными выборочными значениями, попавшими в соседние интервалы. Используя зависимость $t_i = \varphi(x_i, \theta)$, можно найти приближённую оценку параметра (вектора параметров) $\tilde{\theta} = \varphi^{-1}(t_i, \hat{x}_{(i)})$, усреднив её по всем i . Полученные выражения для оценок параметров конкретных распределений приведены в табл. 2.4.1.

Таблица 2.4.1

Оценки параметров по выборочным квантилям асимптотически оптимального группирования
и используемые при оценивании таблицы асимптотически оптимального группирования

№	Распределение	Оценки	Номер формулы	Номера таблиц оптимального группирования в приложении 1
1	2	3	4	5
1	Экспоненциальное $f(x) = \theta e^{-\theta x}$	$\tilde{\theta} = \frac{1}{k-1} \sum_{i=1}^k \frac{t_i}{\hat{x}_{(i)}}$	1	П1.1, П1.2
2	Рэлея $f(x) = \frac{x}{\theta^2} e^{-x^2/2\theta^2}$	$\tilde{\theta} = \frac{1}{k-1} \sum_{i=1}^k \frac{\hat{x}_{(i)}}{t_i}$	2	П1.3, П1.2
3	Максвелла $f(x) = \frac{2x^2}{\theta^3 \sqrt{2\pi}} e^{-\frac{x^2}{2\theta^2}}$	Формула 2	—	П1.4, П1.5
4	Полунормальное $f(x) = \frac{2}{\theta \sqrt{2\pi}} e^{-x^2/2\theta^2}$	Формула 2	—	П1.6, П1.7

Продолжение табл. 2.4.1

1	2	3	4	5
5	Модуля многомерного нормального вектора $f(x) = \frac{2x^{m-1}}{(2\theta^2)^{m/2} \Gamma(m/2)} e^{-x^2/2\theta^2}$ $m = 4$ $m = 5$ $m = 6$ $m = 7$ $m = 8$ $m = 9$	Формула 2 —" —" —" —" —"		П1.8, П1.9 П1.10, П1.11 П1.12, П1.13 П1.14, П1.15 П1.16, П1.17 П1.18, П1.19
6	Вейбулла $f(x) = \frac{\theta x^{\theta-1}}{\theta_1^\theta} \exp\left\{-\left(\frac{x}{\theta_1}\right)^\theta\right\}$	а) $\tilde{\theta} = \frac{1}{k-1} \sum_{i=1}^{k-1} \frac{\ln t_i}{\ln(\hat{x}_{(i)}/\theta_1)}$ б) $\tilde{\theta}_1 = \frac{1}{k-1} \sum_{i=1}^{k-1} \hat{x}_{(i)} / t_i^{1/\theta}$ в) $\tilde{\theta} = \frac{1}{k-2} \sum_{i=2}^{k-1} \frac{\ln t_{i-1} - \ln t_i}{\ln \hat{x}_{(i-1)} - \ln \hat{x}_{(i)}}$ $\tilde{\theta}_1 = \frac{1}{k-2} \sum_{i=2}^{k-1} \left\{ \frac{\ln t_{i-1} \ln \hat{x}_{(i)} - \ln t_i \ln \hat{x}_{(i-1)}}{\ln t_{i-1} - \ln t_i} \right\}$	3 4 5 6	П1.20, П1.21 П1.16, П1.2 П1.22, П1.23

Продолжение табл. 2.4.1

1	2	3	4	5
7	Нормальное $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	a) $\tilde{\mu} = \frac{1}{k-1} \sum_{i=1}^{k-1} (\hat{x}_{(i)} - \sigma t_i)$ b) $\tilde{\sigma} = \frac{1}{k-1} \sum_{i=1}^{k-1} (\hat{x}_{(i)} - \mu) / t_i$ c) $\tilde{\mu} = \frac{1}{k-2} \sum_{i=2}^{k-1} \frac{t_i \hat{x}_{(i-1)} - t_{i-1} \hat{x}_{(i)}}{t_i - t_{i-1}}$ $\tilde{\sigma} = \frac{1}{k-2} \sum_{i=2}^{k-1} \frac{\hat{x}_{(i)} - \hat{x}_{(i-1)}}{t_i - t_{i-1}}$	7 8 13 14	П1.24, П1.25 П1.26, П1.27 П1.28, П1.29
8	Логарифмически нормальное $f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln x - \mu)^2 / 2\sigma^2}$	a) $\tilde{\mu} = \frac{1}{k-1} \sum_{i=1}^{k-1} (\ln \hat{x}_{(i)} - \sigma t_i)$ b) $\tilde{\sigma} = \frac{1}{k-1} \sum_{i=1}^{k-1} (\ln \hat{x}_{(i)} - \mu) / t_i$ c) $\tilde{\mu} = \frac{1}{k-1} \sum_{i=1}^{k-1} \frac{t_i \ln \hat{x}_{(i-1)} - t_{i-1} \ln \hat{x}_{(i)}}{t_i - t_{i-1}}$ $\tilde{\sigma} = \frac{1}{k-2} \sum_{i=2}^{k-1} \frac{\ln \hat{x}_{(i)} - \ln \hat{x}_{(i-1)}}{t_i - t_{i-1}}$	11 12 13 14	П1.24, П1.25 П1.26, П1.27 П1.28, П1.29

Продолжение табл. 2.4.1

1	2	3	4	5
9	Логарифмически нормальное $f(x) = \frac{1}{x\sigma \ln 10 \sqrt{2\pi}} e^{-(\lg x - \mu)^2 / 2\sigma^2}$	Формулы 11-14 с заменой $\ln$ на $\lg$		24-29
10	Наибольшего экстремального значения $f(x) = \frac{1}{\theta_1} \exp \left\{ -\frac{x - \theta_0}{\theta_1} \right\}$	а) $\tilde{\mu}$, Формула 7 б) $\tilde{\sigma}$, Формула 8 в) $\tilde{\mu}$, Формула 9 $\tilde{\sigma}$, Формула 10		П1.30, П1.31 П1.32, П1.33 П1.34, П1.35
11	Наименьшего экстремального значения $f(x) = \frac{1}{\theta_1} \exp \left\{ \frac{x - \theta_0}{\theta_1} \right\}$	а) $\tilde{\mu} = \frac{1}{k-1} \sum_{i=1}^{k-1} (\hat{x}_{(i)} - \sigma \ln t_i)$ б) $\tilde{\sigma} = \frac{1}{k-1} \sum_{i=1}^{k-1} (\hat{x}_{(i)} - \mu) / \ln t_i$ в) $\tilde{\mu} = \frac{1}{k-2} \sum_{i=2}^{k-1} \frac{\hat{x}_{(i-1)} \ln t_i - \hat{x}_{(i)} \ln t_{i-1}}{\ln t_i - \ln t_{i-1}}$ $\tilde{\sigma} = \frac{1}{k-2} \sum_{i=2}^{k-1} \frac{\hat{x}_{(i)} - \hat{x}_{(i-1)}}{\ln t_i - \ln t_{i-1}}$	15 16 17 18	П1.1, П1.2 П1.20, П1.21 П1.22, П1.23

Продолжение табл. 2.4.1

1	2	3	4	5
12	Коши $f(x) = \frac{\theta_0}{\pi[\theta_0^2 + (x - \theta_1)^2]}$	а) $\tilde{\theta}_1 = \frac{1}{k-1} \sum_{i=1}^{k-1} (\hat{x}_{(i)} - \theta t_i)$ б) $\tilde{\theta} = \frac{1}{k-1} \sum_{i=1}^{k-1} (\hat{x}_{(i)} - \theta_1) / t_i$ в) $\tilde{\theta}_1 = \frac{1}{k-2} \sum_{i=2}^{k-1} \frac{t_i \hat{x}_{(i-1)} - t_{i-1} \hat{x}_{(i)}}{t_i - t_{i-1}}$ $\tilde{\theta} = \frac{1}{k-2} \sum_{i=2}^{k-1} \frac{\hat{x}_{(i)} - \hat{x}_{(i-1)}}{t_i - t_{i-1}}$	19 20 21 22	П1.38, П1.39 П1.36, П1.38 П1.40, равные вероятности
13	Логистическое $f(x) = \frac{\pi}{\theta_1 \sqrt{3}} \frac{\exp\left\{ \frac{x - \theta_1}{\theta_1} \right\}}{\left[1 + \exp\left\{ \frac{x - \theta_1}{\theta_1} \right\} \right]^2}$	а) $\tilde{\mu} = \frac{1}{k-1} \sum_{i=1}^{k-1} \left(\hat{x}_{(i)} - \frac{\sqrt{3}}{\pi} \sigma t_i \right)$ б) $\tilde{\sigma} = \frac{1}{k-1} \sum_{i=1}^{k-1} (\hat{x}_{(i)} - \mu) / t_i$ в) $\tilde{\mu} = \frac{1}{k-2} \sum_{i=2}^{k-1} \frac{t_i \hat{x}_{(i-1)} - t_{i-1} \hat{x}_{(i)}}{t_i - t_{i-1}}$ $\tilde{\sigma} = \frac{1}{k-2} \sum_{i=2}^{k-1} \frac{\hat{x}_{(i)} - \hat{x}_{(i-1)}}{t_i - t_{i-1}}$	23 24 25 26	П1.41, равные вероятности П1.42, П1.43 П1.44, П1.45

Продолжение табл. 2.4.1

1	2	3	4	5
14	Парето $f(x) = \theta \alpha^\theta x^{-\theta-1}$	$\tilde{\theta} = \frac{1}{k-1} \sum_{i=1}^{k-1} \frac{t_i}{\ln(\hat{x}_{(i)}/\alpha)}$	27	П1.1, П1.2
15	Лапласа $f(x) = \frac{\theta}{2} e^{-\theta(x)}$	Формула 1		П1.46, П1.48 и П1.47, П1.49
16	Двойное показательное $f(x) = \theta_0 \theta_1 \exp\{-(x - \theta_0) - \theta_1 \exp[-(x - \theta_2)\theta_0]\}$	a) $\tilde{\theta}_2 = \frac{1}{k-1} \sum_{i=1}^{k-1} t_i / \exp(-\theta_1 \hat{x}_{(i)})$	28	П1.1, П1.31
		b) $\tilde{\theta}_1 = \frac{1}{k-1} \sum_{i=1}^{k-1} (-\ln t_i / \hat{x}_{(i)})$	29	П1.50, П1.51
		c) $\tilde{\theta}_2 = \frac{1}{k-2} \sum_{i=2}^{k-1} \exp \left\{ \frac{\hat{x}_{(i)} \ln t_{i-1} - \hat{x}_{(i-1)} \ln t_i}{\hat{x}_{(i)} - \hat{x}_{(i-1)}} \right\}$	30	П1.52, П1.35
		$\tilde{\theta}_1 = \frac{1}{k-2} \sum_{i=2}^{k-1} \frac{\ln t_{i-1} - \ln t_i}{\hat{x}_{(i)} - \hat{x}_{(i-1)}}$	31	

Ниже в табл. 2.4.2 проведены результаты небольшого эксперимента по сравнению предложенных оценок с оценками максимального правдоподобия по группированным данным. Соответствующие выборки объемом в 2000 наблюдений моделировались с масштабными параметрами: экспоненциальное - $\theta=1$, Рэлея и Максвелла - $\theta=2$. В таблице использованы следующие обозначения: $\hat{\theta}$ - ОМП по негруппированным наблюдениям, $\hat{\theta}_Г$ - ОМП по сгруппированным данным, $\tilde{\theta}$ - оценки по выборочным асимптотически оптимальным квантилям. ОМП $\hat{\theta}_Г$ по сгруппированным данным вычислялись при том же самом разбиении на интервалы исходной выборки.

На рис. 2.4.1 представлены результаты проверки согласия смоделированной по экспоненциальному закону выборки с распределением с параметром $\tilde{\theta}=0.972800$, полученной при $k=10$. При проверке согласия в критериях отношения правдоподобия и χ^2 Пирсона использовалось асимптотически оптимальное группирование (см. гл. 4). При вычислении вероятности превышения статистикой полученного по выборке значения для непараметрических критериев Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса учтено оценивание по данной выборке параметра закона (см. гл. 4).

Таблица 2.4.2

k	Экспоненциальное $\hat{\theta}=0.969938$		Рэлея $\hat{\theta}=1.978140$		Максвелла $\hat{\theta}=1.972455$	
	$\hat{\theta}_Г$	$\tilde{\theta}$	$\hat{\theta}_Г$	$\tilde{\theta}$	$\hat{\theta}_Г$	$\tilde{\theta}$
5	0.970761	0.970881	1.976280	1.970031	1.970618	1.973493
6	0.965740	0.970275	1.982693	1.973644	1.965676	1.970536
7	0.971096	0.974537	1.970832	1.960941	1.970248	1.975539
8	0.973889	0.979115	1.980232	1.968719	1.969892	1.979076
9	0.969049	0.972566	1.974299	1.969750	1.971905	1.982197
10	0.967202	0.972800	1.980326	1.976820	1.972612	1.982102

На рис. 2.4.2 представлены результаты проверки согласия смоделированной выборки с распределением Рэлея с параметром $\tilde{\theta}=0.970031$, полученным при $k=5$, а на рис. 2.4.3 - результаты

статистического анализа для закона Рэлея с $\hat{\theta}_T=1.976280$ при том же k .

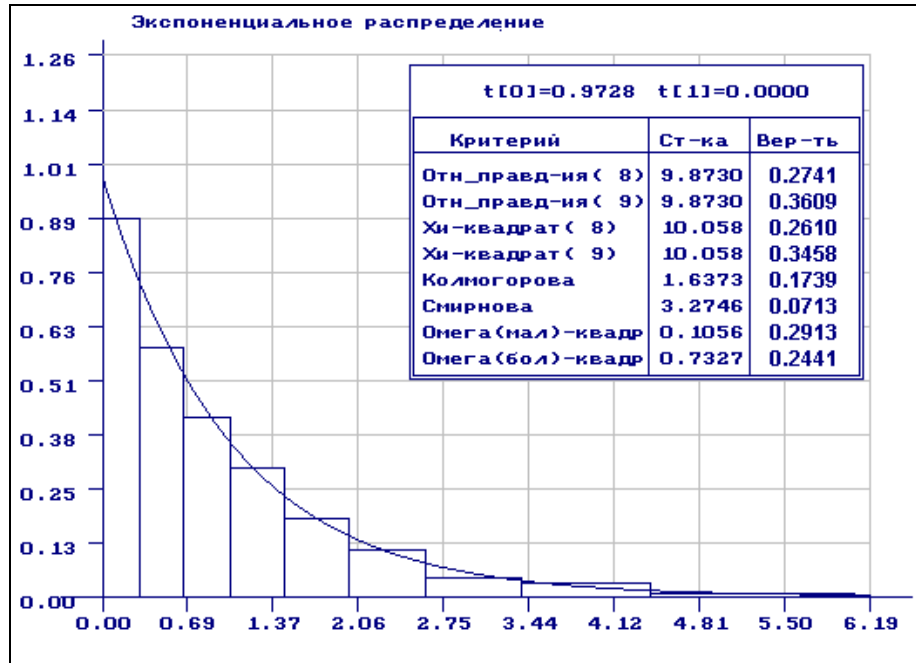


Рис. 2.4.1. Результаты проверки согласия с экспоненциальным законом с параметром $\tilde{\theta}=0.972800$

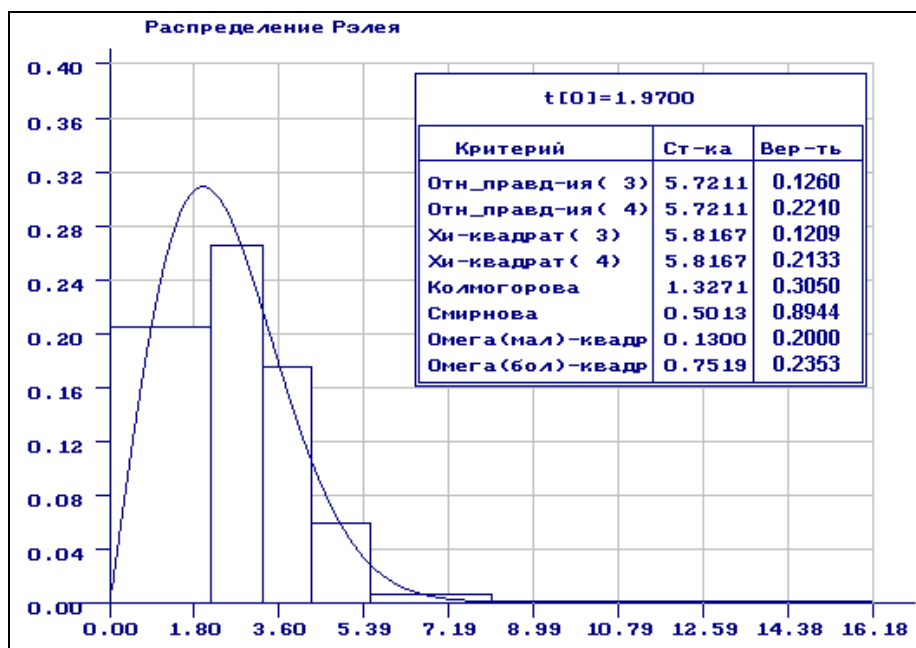


Рис. 2.4.2 Результаты проверки согласия с законом Рэлея с параметром $\tilde{\theta}=0.970031$

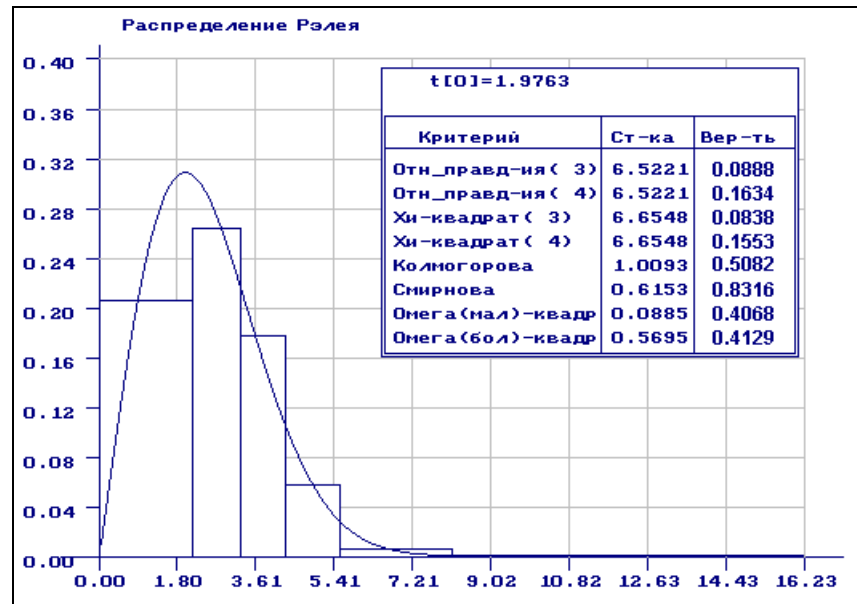


Рис. 2.4.3 Результаты проверки согласия с законом Рэлея с параметром $\hat{\theta}_r = 1.976280$

2.5. Оптимальные L-оценки параметров сдвига и масштаба по выборочным квантилям

Опираясь на асимптотическое распределение $k-1$ выборочных квантилей (при k интервалах) [251], Огавой в работах [253,141] выведено асимптотическое распределение выборочных квантилей для функции плотности, зависящей только от параметра расположения μ и от параметра рассеяния σ . Им же в [141] получены линейные несмещенные оценки параметров сдвига и масштаба по методу наименьших квадратов, в основе которых лежат значения выборочных квантилей (L-оценки).

Пусть μ и σ неизвестные параметры сдвига и масштаба закона с функцией распределения $F\left(\frac{x-\mu}{\sigma}\right)$ и функцией плотности $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$.

При известном параметре σ оценка параметра μ имеет вид

$$\tilde{\mu} = \frac{1}{K_1} Z - \sigma \frac{K_3}{K_1}, \quad (2.5.1)$$

где

$$Z = \sum_{i=1}^k \frac{(f_i - f_{i-1})(f_i x_{(i)} - f_i x_{(i-1)})}{P_i}, \quad (2.5.2)$$

$$K_1 = \sigma^2 J_\Gamma(\mu) = \sum_{i=1}^k \frac{(f_i - f_{i-1})^2}{P_i}, \quad (2.5.3)$$

$$K_3 = \sigma^2 J_\Gamma(\mu, \sigma) = \sum_{i=1}^k \frac{(f_{i-1} - f_i)(f_{i-1}t_{i-1} - f_i t_i)}{P_i}, \quad (2.5.4)$$

$f_i = f(t_i)$, $t = (x - \mu) / \sigma$, $P_i = F(t_i) - F(t_{i-1})$, $f_0 = f_k = 0$. Здесь $x_{(i)}$ - выборочная квантиль, такая, что $F\left(\frac{x_{(i)} - \mu}{\sigma}\right) = F(t_i)$ и t_i - квантиль рас-

пределения с нулевым параметром сдвига и единичным масштабным. Через $J_\Gamma(\cdot)$ - обозначено количество информации Фишера о соответствующем параметре. Информационная матрица Фишера вектора параметров будет иметь вид

$$M_\Gamma(\mu, \sigma) = \begin{bmatrix} J_\Gamma(\mu) & J_\Gamma(\mu, \sigma) \\ J_\Gamma(\mu, \sigma) & J_\Gamma(\sigma) \end{bmatrix}.$$

При известном параметре μ оценка параметра σ имеет вид

$$\hat{\sigma} = \frac{1}{K_2} Y - \mu \frac{K_3}{K_2}, \quad (2.5.5)$$

где

$$Y = \sum_{i=1}^k \frac{(t_i f_i - t_{i-1} f_{i-1})(f_i x_{(i)} - f_i x_{(i-1)})}{P_i}, \quad (2.5.6)$$

$$K_2 = \sigma^2 J_\Gamma(\sigma) = \sum_{i=1}^k \frac{(t_i f_i - t_{i-1} f_{i-1})^2}{P_i}, \quad (2.5.7)$$

При симметричных функциях плотности и симметричных квантилях второе слагаемое в формулах (2.5.1) и (2.5.5) равно нулю.

Если неизвестны оба параметра, то оценки параметров сдвига и масштаба определяются соотношениями

$$\hat{\mu} = \frac{1}{\Delta} (K_2 Z - K_3 Y), \quad (2.5.8)$$

$$\hat{\sigma} = \frac{1}{\Delta} (-K_3 Z + K_1 Y), \quad (2.5.9)$$

где $\Delta = K_1 K_2 - K_3^2$.

Все эти оценки асимптотически эффективны и их асимптотические дисперсии определяются количеством информации Фишера по группированному данным, а в случае векторного параметра информационной матрицей по группированным данным.

Если использовать таблицы асимптотически оптимального группирования, то можно получить оптимальные оценки параметров сдвига и масштаба в виде линейной комбинации выборочных квантилей [88, 117].

Формулу (2.5.1) для оценивания μ при известном σ можно привести к виду

$$\tilde{\mu} = \alpha_0 \sigma + \sum_{i=1}^{k-1} \alpha_i x_{(i)}, \quad (2.5.10)$$

где

$$\begin{aligned} \alpha_0 &= -\frac{K_3}{K_1}, \\ \alpha_1 &= \alpha'_1 / K_1 = \left(\frac{f_1^2}{P_0} + \frac{f_1^2 - f_1 f_2}{P_1} \right) / K_1, \\ \alpha_i &= \alpha'_i / K_1 = \left(\frac{f_i^2 - f_i f_{i-1}}{P_{i-1}} + \frac{f_i^2 - f_i f_{i+1}}{P_i} \right) / K_1, \quad i = \overline{2, (k-2)}; \\ \alpha_{k-1} &= \alpha'_{k-1} / K_1 = \left(\frac{f_{k-1}^2 - f_{k-1} f_{k-2}}{P_{k-2}} + \frac{f_{k-1}^2}{P_{k-1}} \right) / K_1. \end{aligned}$$

А формулу (2.5.5) для оценивания σ при известном μ можно представить в виде

$$\tilde{\sigma} = \beta_0 \mu + \sum_{i=1}^{k-1} \beta_i x_{(i)}, \quad (2.5.11)$$

где

$$\begin{aligned} \beta_0 &= -\frac{K_3}{K_2}, \\ \beta_1 &= \beta'_1 / K_2 = \left(\frac{t_1 f_1^2}{P_0} + \frac{t_1 f_1^2 - t_2 f_1 f_2}{P_1} \right) / K_2, \\ \beta_i &= \beta'_i / K_2 = \left(\frac{t_i f_i^2 - t_{i-1} f_i f_{i-1}}{P_{i-1}} + \frac{t_i f_i^2 - t_{i+1} f_i f_{i+1}}{P_i} \right) / K_2, \quad i = \overline{2, (k-2)}; \\ \beta_{k-1} &= \beta'_{k-1} / K_2 = \left(\frac{t_{k-1} f_{k-1}^2 - t_{k-2} f_{k-1} f_{k-2}}{P_{k-2}} + \frac{t_{k-1} f_{k-1}^2}{P_{k-1}} \right) / K_2. \end{aligned}$$

Аналогично формулы (2.5.8) и (2.5.9) можно преобразовать следующим образом

$$\tilde{\mu} = \sum_{i=1}^{k-1} \gamma_i x_{(i)}, \quad (2.5.12)$$

$$\tilde{\sigma} = \sum_{i=1}^{k-1} v_i x_{(i)}, \quad (2.5.13)$$

где

$$\gamma_i = (\alpha_i' K_1 - \beta_i' K_3) / \Delta,$$

$$v_i = (-\alpha_i' K_3 + \beta_i' K_1) / \Delta.$$

Коэффициенты α_i , β_i , γ_i , v_i зависят от граничных точек t_i (квантилей стандартизованного распределения). Очевидно, что так как рассматриваемые оценки асимптотически эффективны, то использование квантилей (граничных точек интервалов), соответствующих асимптотически оптимальному группированию, обеспечит оптимальные свойства этих оценок: минимум асимптотической дисперсии, а в случае оценивания сразу двух параметров - минимум обобщенной асимптотической дисперсии. Несложно вычислить значения $\alpha_i, \beta_i, \gamma_i, v_i$ при асимптотически оптимальном группировании и сформировать таблицы соответствующих коэффициентов. И если в случае *больших* выборок мы будем выбирать $x_{(i)}$ таким образом, чтобы $n_i \approx nP_i$, где P_i - соответствует вероятности попадания в интервал при асимптотически оптимальном группировании, и использовать соответственно формулы (2.5.10), (2.5.11), (2.5.12) и (2.5.13) с полученными коэффициентами, то будем получать оптимальные оценки.

В частных случаях решение такой задачи рассматривалось в ряде работ. В [141,213] рассматривались оценки параметров для нормального распределения, в [141] - для однопараметрического экспоненциального распределения, в [259] - для двухпараметрического, в [219] - для параметров логистического распределения, в [191] - для параметров распределения Коши, в [225] - для параметров распределения экстремальных значений. Приближенный подход к решению такой задачи рассматривался в [260]. Причем в случае одновременного оценивания параметров μ и σ оптимальные наборы граничных точек определялись исходя из минимума $D[\tilde{\mu}] + cD[\tilde{\sigma}]$, ($c=1,2,3,\dots$), а не минимума $\det M_{\Gamma}^{-1}(\tilde{\mu}, \tilde{\sigma})$.

Опираясь на построенную нами совокупность таблиц асимптотически оптимального группирования, сформированы таблицы коэффициентов для формул вида (2.5.10), (2.5.11), (2.5.12) и (2.5.13) для нормального распределения, для логистического распределения с функцией плотности

$$f(x) = \frac{\pi}{\sigma\sqrt{3}} \exp\left\{-\frac{\pi(x-\mu)}{\sigma\sqrt{3}}\right\} / \left[1 + \exp\left\{-\frac{\pi(x-\mu)}{\sigma\sqrt{3}}\right\}\right]^2,$$

для распределения Коши с плотностью

$$f(x) = \frac{\sigma}{\pi[\sigma^2 + (x - \mu)^2]},$$

для распределения наименьшего экстремального значения с плотностью

$$f(x) = \frac{1}{\sigma} \exp \left\{ \frac{x - \mu}{\sigma} - \exp \left(\frac{x - \mu}{\sigma} \right) \right\},$$

для распределения наибольшего экстремального значения с плотностью

$$f(x) = \frac{1}{\sigma} \exp \left\{ -\frac{x - \mu}{\sigma} - \exp \left(-\frac{x - \mu}{\sigma} \right) \right\},$$

для распределения Лапласа с плотностью

$$f(x) = \frac{1}{2\sigma} \exp \left(-\frac{|x - \mu|}{\sigma} \right).$$

При этом, в зависимости от того, известен ли один из параметров или неизвестны оба параметра, наборам коэффициентов α_i , β_i и паре γ_i , ν_i соответствуют свои таблицы асимптотически оптимального группирования.

Для оценивания параметров μ и σ логарифмически нормальных распределений с плотностями

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right\}$$

и

$$f(x) = \frac{1}{x\sigma \ln 10 \sqrt{2\pi}} \exp \left\{ -\frac{(\lg x - \mu)^2}{2\sigma^2} \right\}$$

с использованием порядковых статистик достаточно осуществить преобразование исходной выборки $X_1, X_2, \dots, X_n$, принадлежащей одному из этих распределений, в соответствии с соотношениями $U_i = \ln X_i$ или $U_i = \lg X_i$. В этом случае величина U будет принадлежать нормальному распределению с параметрами μ и σ , для оценивания которых будут использоваться формулы с коэффициентами и асимптотически оптимальными вероятностями для нормального распределения.

Для распределений экспоненциального с плотностью

$$f(x) = \frac{1}{\sigma} \exp \{ -(x - \mu) / \sigma \},$$

модуля нормального вектора ($m = 1 \div 9$) с плотностью

$$f(x) = \frac{2(x - \mu)^{m-1}}{(2\sigma^2)^{m/2} \Gamma(m/2)} \exp\left\{- (x - \mu)^2 / 2\sigma^2\right\},$$

частными случаями которого являются полунормальное распределение - $m=1$, Рэлея - $m=2$ и Максвелла - $m=3$ таблицы коэффициентов α_i , β_i , γ_i , ν_i опираются на таблицы асимптотически оптимального группирования *только относительно масштабного параметра σ* . Это связано с тем, что область определения этих случайных величин зависит от параметра сдвига μ и, следовательно, в этом случае теряет смысл максимизация соотношения (2.5.3) для построения асимптотически оптимальных граничных точек относительно этого параметра.

Значения $x_{(i)}$, фигурирующие в формулах (2.5.10), (2.5.11), (2.5.12) и (2.5.13), следует выбирать из условия

$$X_{(\lfloor nP^i \rfloor)} \leq x_{(i)} \leq X_{(\lfloor nP^i \rfloor + 1)},$$

где $X_{(l)}$ - члены вариационного ряда $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$, построенного по исходной выборке, $P^i = \sum_{j=1}^i P_j$, $\cdot$ - означает целую часть числа,

а P_j - выбираются из соответствующей строки таблицы оптимальных вероятностей. Например, это могут быть средние значения между соответствующими соседними членами вариационного ряда.

Для удобства для различных законов распределения указания на используемые таблицы с коэффициентами для формул (2.5.10-2.5.13) и соответствующие таблицы вероятностей, содержащиеся в приложениях, сведены в табл. 2.5.1.

Таблица 2.5.1

Распределение	Оцениваемый параметр	№ формулы	Таблица вероятностей	Таблица коэффициентов
Нормальное	μ (известно σ)	2.5.10	П1.25	П2.1
	σ (известно μ)	2.5.11	П1.27	П2.2
	μ	2.5.12	П1.29	П2.3
	σ	2.5.13	П1.29	П2.4
Логарифмически нормальное (ln)	μ (известно σ)	2.5.10	П1.25	П2.1
	σ (известно μ)	2.5.11	П1.27	П2.2
	μ	2.5.12	П1.29	П2.3
	σ	2.5.13	П1.29	П2.4
Логарифмически нормальное (lg)	μ (известно σ)	2.5.10	П1.25	П2.1
	σ (известно μ)	2.5.11	П1.27	П2.2
	μ	2.5.12	П1.29	П2.3
	σ	2.5.13	П1.29	П2.4

Логистическое	μ (известно σ)	2.5.10	Равновероятные П1.43 П1.45 П1.45	П2.5
	σ (известно μ)	2.5.11		П2.6
	μ	2.5.12		П2.7
	σ	2.5.13		П2.8
Коши	μ (известно σ)	2.5.10	П1.39	П2.9
	σ (известно μ)	2.5.11	П1.37	П2.10
	μ	2.5.12	Равновероятные	П2.11
	σ	2.5.13	Равновероятные	П2.12
Минимального значения	μ (известно σ)	2.5.10	П1.2	П2.13
	σ (известно μ)	2.5.11	П1.21	П2.14
	μ	2.5.12	П1.23	П2.15
	σ	2.5.13	П1.23	П2.16
Максимального значения	μ (известно σ)	2.5.10	П1.31	П2.17
	σ (известно μ)	2.5.11	П1.33	П2.18
	μ	2.5.12	П1.35	П2.19
	σ	2.5.13	П1.35	П2.20
Экспоненциальное (показательное)	μ (известно σ)	2.5.10	П1.2	П2.21
	σ (известно μ)	2.5.11	П1.2	П2.22
	μ	2.5.12	П1.2	П2.23
	σ	2.5.13	П1.2	П2.24
Модуля многомерного нормального вектора ($m=1$, полунормальное)	μ (известно σ)	2.5.10	П1.7	П2.25
	σ (известно μ)	2.5.11	П1.7	П2.26
	μ	2.5.12	П1.7	П2.27
	σ	2.5.13	П1.7	П2.28
Модуля многомерного нормального вектора ($m=2$, Рэлея)	μ (известно σ)	2.5.10	П1.2	П2.29
	σ (известно μ)	2.5.11	П1.2	П2.30
	μ	2.5.12	П1.2	П2.31
	σ	2.5.13	П1.2	П2.32
Модуля многомерного нормального вектора ($m=3$, Максвелла)	μ (известно σ)	2.5.10	П1.5	П2.33
	σ (известно μ)	2.5.11	П1.5	П2.34
	μ	2.5.12	П1.5	П2.35
	σ	2.5.13	П1.5	П2.36
Модуля многомерного нормального вектора ($m=4$)	μ (известно σ)	2.5.10	П1.9	П2.37
	σ (известно μ)	2.5.11	П1.9	П2.38
	μ	2.5.12	П1.9	П2.39
	σ	2.5.13	П1.9	П2.40
Модуля многомерного нормального вектора ($m=5$)	μ (известно σ)	2.5.10	П1.11	П2.41
	σ (известно μ)	2.5.11	П1.11	П2.42
	μ	2.5.12	П1.11	П2.43
	σ	2.5.13	П1.11	П2.44

Модуля многомер-	μ (известно σ)	2.5.10	П1.13	П2.45
------------------	----------------------------	--------	-------	-------

ного нормального вектора ($m=6$)	σ (известно μ)	2.5.11	П1.13	П2.46
	μ	2.5.12	П1.13	П2.47
	σ	2.5.13	П1.13	П2.48
Модуля многомерного нормального вектора ($m=7$)	μ (известно σ)	2.5.10	П1.15	П2.49
	σ (известно μ)	2.5.11	П1.15	П2.50
	μ	2.5.12	П1.15	П2.51
	σ	2.5.13	П1.15	П2.52
Модуля многомерного нормального вектора ($m=8$)	μ (известно σ)	2.5.10	П1.17	П2.53
	σ (известно μ)	2.5.11	П1.17	П2.54
	μ	2.5.12	П1.17	П2.55
	σ	2.5.13	П1.17	П2.56
Модуля многомерного нормального вектора ($m=9$)	μ (известно σ)	2.5.10	П1.19	П2.57
	σ (известно μ)	2.5.11	П1.19	П2.58
	μ	2.5.12	П1.19	П2.59
	σ	2.5.13	П1.19	П2.60
Лапласа	μ (известно σ)	2.5.10	П1.48	П2.61
	σ (известно μ)	2.5.11	П1.48	П2.62
	μ	2.5.12	П1.48	П2.63
	σ	2.5.13	П1.48	П2.64

Пример. Для нормального распределения при $k=9$ соотношение (2.5.10) (см. табл. П2.1) принимает вид

$$\tilde{\mu} = 0.056339(x_{(1)} + x_{(8)}) + 0.111523(x_{(2)} + x_{(7)}) + 0.154649(x_{(3)} + x_{(6)}) + \\ + 0.177489(x_{(4)} + x_{(5)}),$$

соотношение (2.5.11) (см. табл. П2.2) -

$$\tilde{\sigma} = 0.031157(-x_{(1)} + x_{(8)}) + 0.072629(-x_{(2)} + x_{(7)}) + 0.116643(-x_{(3)} + x_{(6)}) + \\ + 0.147029(-x_{(4)} + x_{(5)}),$$

соотношения (2.5.12) и (2.5.13) (см. табл. П2.3-П2.4) -

$$\tilde{\mu} = 0.021547(x_{(1)} + x_{(8)}) + 0.068108(x_{(2)} + x_{(7)}) + 0.148605(x_{(3)} + x_{(6)}) + \\ + 0.261739(x_{(4)} + x_{(5)}),$$

$$\tilde{\sigma} = 0.052747(-x_{(1)} + x_{(8)}) + 0.114684(-x_{(2)} + x_{(7)}) + 0.153492(-x_{(3)} + x_{(6)}) + \\ + 0.090860(-x_{(4)} + x_{(5)}).$$

Для смоделированной нормальной выборки объемом 1000 наблюдений оценки максимального правдоподобия (ОМП) по негруппированным данным оказались равны $\hat{\mu} = -0.01672934$, $\hat{\sigma} = 0.9849848$, оценки по формулам (2.5.12) и (2.5.13) при $k=9$ - $\tilde{\mu} = -0.018158$, $\tilde{\sigma} = 0.975442$. При определении $x_{(i)}$ вероятности P_i выбирались из табл. П1.29 и в качестве $x_{(i)}$ были взяты средние значения

между следующими парами членов вариационного ряда $X_{(10)} - X_{(11)}$, $X_{(52)} - X_{(53)}$, $X_{(153)} - X_{(154)}$, $X_{(350)} - X_{(351)}$, $X_{(649)} - X_{(650)}$, $X_{(846)} - X_{(847)}$, $X_{(947)} - X_{(948)}$, $X_{(989)} - X_{(990)}$. Для сравнения вычисленные ОМП по группированным данным с использованием тех же самых граничных точек равны $\hat{\mu}_T = -0.01680601$ и $\hat{\sigma}_T = 0.9766322$. Естественно, что при проверке критериев согласия полученных теоретических законов с исходной выборкой наилучшие результаты получены для ОМП по негруппированным данным. Причем применялись критерии отношения правдоподобия, χ^2 Пирсона, Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса. В то же время заслуживает внимания факт, что по всем этим критериям оценки $\tilde{\mu}$ и $\tilde{\sigma}$ оказались *предпочтительнее* оценок $\hat{\mu}_T$ и $\tilde{\sigma}_T$. Результаты статистического анализа для оценок $\tilde{\mu}$ и $\tilde{\sigma}$ представлены на рис. 2.5.1. На этом и следующем рисунках приведены вычисленные значения статистик соответствующих критериев S^* и вероятности вида $P\{S > S^*\}$, позволяющие судить о степени согласия. Гипотезы о согласии не отвергаются, если задаваемый уровень значимости $\alpha < P\{S > S^*\}$.

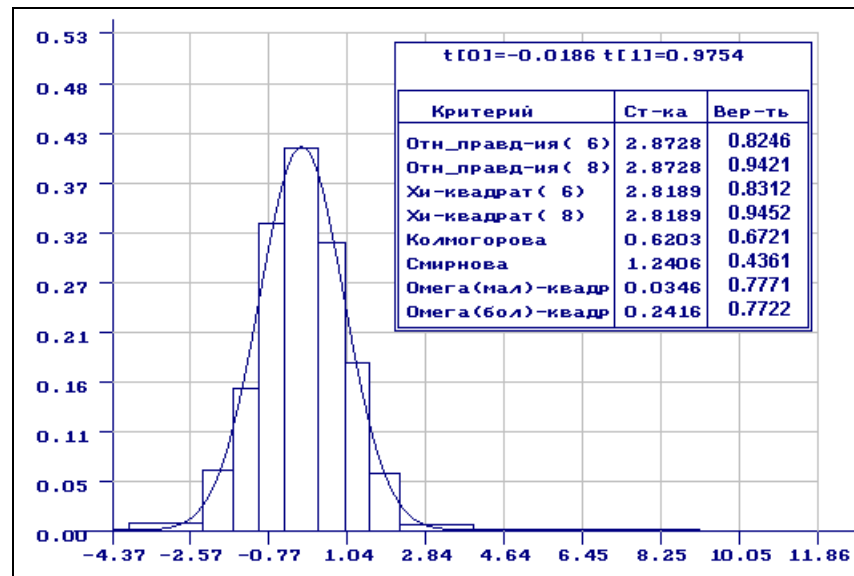


Рис. 2.5.1. Результаты статистического анализа для нормального закона с параметрами $\tilde{\mu}$ и $\tilde{\sigma}$.

Аналогичные результаты характерны для оценок параметров сдвига и масштаба для законов распределений Коши, логистического, наибольшего и наименьшего экстремальных значений. На рис. 2.5.2 представлены данные статистического анализа для распределения

наименьшего экстремального значения, когда по смоделированной выборке определялись параметры $\tilde{\mu}$ и $\tilde{\sigma}$.

Симметричность коэффициентов в формулах (2.5.10-2.5.13) для симметричных распределений определяется симметричностью оптимальных граничных точек интервалов. А для параметров масштаба при известном параметре сдвига и четном k задача асимптотически оптимального группирования обычно имеет два решения с несимметричными значениями квантилей. В таких случаях пара этих решений зеркальна относительно центра симметрии распределения. Поэтому не единственным будет оптимальный набор коэффициентов в формулах (2.5.11). Таким образом, не оправдывается предположение о симметричности оптимальных порядковых статистик для параметра σ нормального распределения, высказанные в [43].

Следует отметить, что рассмотренные оценки, как и все оценки по группированным данным являются робастными. Они устойчивы к наличию аномальных ошибок измерений, к малым отклонениям от исходных предположений о виде наблюдаемого закона распределения.



Рис. 2.5.2. Результаты статистического анализа для распределения наименьшего экстремального значения с параметрами $\tilde{\mu}$ и $\tilde{\sigma}$.

Сказанное можно подтвердить результатами следующего эксперимента. В выборку объемом 1000 наблюдений, смоделированную по нормальному закону с параметрами $\mu = 0$, $\sigma = 1$, было внесено 5 аномальных “ошибок” измерений. Результаты анализа с ОМП по

негруппированным данным приведены на рис. 2.5.3. Оценки параметров $\hat{\mu} = -0.020729$ и $\hat{\sigma} = 1.136284$. Как и ожидалось, налицо влияние аномальных наблюдений.

На рис. 2.5.4 приведены результаты анализа при использовании найденных оптимальных оценок $\tilde{\mu} = -0.02204$, $\tilde{\sigma} = 0.995429$. Как видим, на оценки не повлияло наличие грубых ошибок. Для сравнения на рис. 2.5.5 приведены результаты анализа при использовании ОМП по группированным данным с оценками $\hat{\mu}_T = -0.019573$ и $\hat{\sigma}_T = 0.996555$. При всей близости результатов оценивания критерии согласия показывают предпочтительность оценок $\tilde{\mu}$ и $\tilde{\sigma}$.

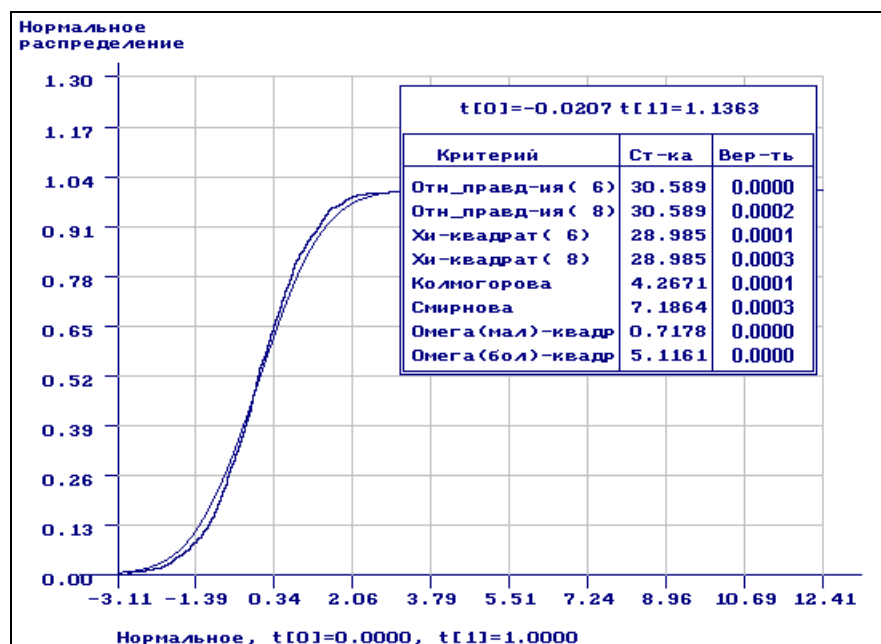


Рис. 2.5.3. Результаты статистического анализа для нормального закона при использовании ОМП по негруппированным наблюдениям

Аналогичные результаты, подтверждающие устойчивость оптимальных оценок параметров сдвига и масштаба по выборочным квантилям, дали эксперименты с моделированием засоренных выборок в соответствии с другими законами распределения с последующим оцениванием параметров. При этом практически всегда оценки $\tilde{\mu}$ и $\tilde{\sigma}$ оказывались предпочтительней (по критериям согласия) оценок $\hat{\mu}_T$ и $\hat{\sigma}_T$. Последний результат имел место как при наличии аномальных измерений, так и без них.

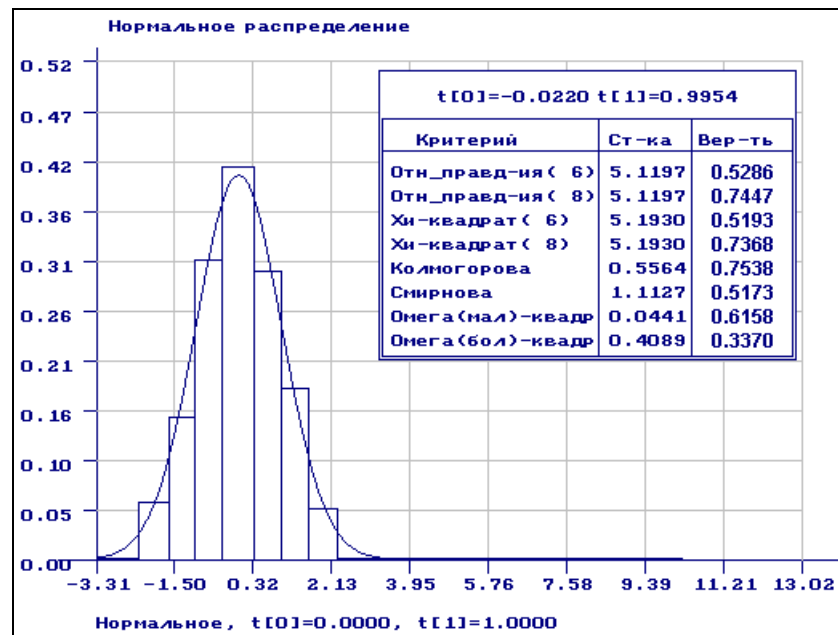


Рис. 2.5.4. Результаты статистического анализа для нормального закона с параметрами $\tilde{\mu}$ и $\tilde{\sigma}$.



Рис. 2.5.5. Результаты статистического анализа для нормального закона с параметрами $\hat{\mu}_T$ и $\hat{\sigma}_T$.

Выводы

1. Применение таблиц вероятностей попадания в интервал, соответствующих асимптотически оптимальному группированию, и формул (2.5.10-2.5.13), опирающихся на вычисленные таблицы коэффициентов, позволяет получать легко вычисляемые оптимальные оценки параметров сдвига и масштаба для больших

выборок. При этом самым трудоемким является процедура упорядочивания исходных наблюдений.

2. Предложенные оценки являются робастными. Они устойчивы к аномальным ошибкам измерений, к отклонениям наблюдаемых данных от предполагаемого закона распределения.

2.6. Выбор числа интервалов

Рекомендуемое количество интервалов группирования, приводимое в различных источниках, и используемое при вычислении оценок параметров, построении гистограмм, вычислении статистик типа отношения правдоподобия или χ^2 Пирсона колеблется в очень широких пределах. Рекомендации связывают с объемом выборки и стремлением, чтобы вид гистограммы как можно более близко был к плавной кривой распределения генеральной совокупности. В [143] приводится целый обзор рекомендаций из различных источников по выбору числа интервалов k . При выборе интервалов равной длины определяющим является требование, чтобы количество наблюдений, попавших в интервалы, было не слишком малым и сравнимым. При этом наиболее часто рекомендуется, чтобы количество наблюдений, попавших в интервал, было не менее 10. В [201] отмечается, что на практике допустимо, чтобы количество наблюдений в крайних интервалах было менее 5. В работе [244], где изучалась мощность критерия χ^2 Пирсона, говорится, что в случае унимодального распределения допускается уменьшение ожидаемых частот попадания наблюдений для одного или двух интервалов до 1 и даже ниже.

Естественно, что определение количества интервалов связывается с объемом выборки. Большинство рекомендуемых формул для оценки числа интервалов k носят эмпирический характер. Во многих источниках можно найти упоминание эвристической формулы Старджесса для определения “оптимального” числа интервалов [264]

$$k = \log_2 N + 1 = 3.3 \lg N + 1.$$

В [183] для определения “оптимального” числа интервалов рекомендуется формула Брукса и Каррузера

$$k = 5 \lg N.$$

В [227] рекомендуют соотношение

$$k = \sqrt{N}.$$

В [245] для равновероятных интервалов их количество устанавливается порядка

$$k \approx 4\sqrt[5]{2}(N/t)^{0.4},$$

где t - квантиль стандартного нормального распределения для заданного уровня значимости. В ряде работ приводятся модификации данной формулы. В [157] предлагается значение

$$k = 4 \lg N,$$

а в [170] дальнейшее развитие этого соотношения

$$k = 5 \lg N - 5.$$

В исследовании [5] получено соотношение

$$k = \frac{4}{\aleph} \lg \frac{N}{10},$$

где $\aleph$ - значение контрэксцесса ($\aleph = 1 / \sqrt{\mu_4 / \sigma^4}$).

Разброс значений k , задаваемых различными формулами, достаточно велик при больших N . Поэтому на практике при выборе числа интервалов больше руководствуются разумными соображениями, выбирая число интервалов так, чтобы в интервалы попадало число наблюдений не менее 10.

Очевидно, что “оптимальная” величина k зависит не только от объёма выборки, но и от вида закона распределения. Она конечно зависит и от способа группирования. Но не следует придавать её выбору слишком большого значения. При асимптотически оптимальном группировании относительно скалярного параметра при 10-11 интервалах в группированной выборке сохраняется около 98% информации, при оптимальном группировании относительно вектора параметров (два параметра) для 15 интервалов - около 95%. Дальнейшее увеличение количества интервалов существенного значения не имеет. Выбор конкретного числа интервалов должен осуществляться из следующих соображений. При оптимальном группировании вероятности попадания в интервалы в общем случае не равны. Обычно минимальны вероятности попадания в крайние интервалы. Поэтому k желательно выбирать из условия $NP_i(\theta) \geq 5 \div 10$ для любого интервала при оптимальном группировании. По крайней мере минимальная ожидаемая частота должна быть больше 1. В случае использования равновероятного группирования порядок k должен быть примерно таким же, как и при асимптотически оптимальном группировании.

2.7. Выводы

1. Для ряда законов распределения вероятностей при решении задач асимптотически оптимального группирования получены граничные точки интервалов в виде, инвариантном относительно параметров распределений, и на их основе сформированы таблицы асимптотически оптимального группирования. В общей сложности получены 54 таблицы оптимальных граничных точек и соответствующих вероятностей для распределений экспоненциального, полунормального, Рэлея, Максвелла, модуля многомерного нормального вектора $m = 4 \div 9$, Парето, Эрланга, Лапласа, нормального, логарифмически нормальных ($\ln$ и $\lg$), Коши, Вейбулла, распределений минимального и максимального значения, двойного показательного, гамма-распределения. Построенные таблицы представлены в приложении П1.
2. Опираясь на таблицы асимптотически оптимального группирования, предложены оценки, использующие выборочные квантили, соответствующие оптимальному группированию. Получены формулы для вычисления данных оценок.
3. Для законов распределения, зависящих только от параметров сдвига и масштаба (нормального, логарифмически нормальных, логистического, Коши, наименьшего экстремального значения, наибольшего экстремального значения, Лапласа, экспоненциального, модуля нормального вектора $m = 1 \div 9$), предложены оптимальные L-оценки параметров сдвига и масштаба для больших выборок, опирающиеся на таблицы асимптотически оптимального группирования. Вычисленные таблицы коэффициентов для этих оценок представлены в приложении П2.
4. Выбираемое число интервалов при группировании зависит от объёма выборки и способа группирования. При асимптотически оптимальном группировании относительно скалярного параметра и 10-11 интервалах в группированной выборке сохраняется около 98% информации, при группировании относительно вектора параметров (два параметра) и 15 интервалов - около 95%. Дальнейшее увеличение количества интервалов практического значения не имеет. Выбор конкретного числа интервалов рекомендуется осуществляться из условия $NP_i(\theta) \geq 5 \div 10$ при оптимальном группировании. По крайней мере ожидаемая частота для интервала с минимальной вероятностью должна быть больше 1.

3. РОБАСТНОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЙ

3.1. Способы вычисления робастных оценок

В статистике под робастностью понимают *нечувствительность к малым отклонениям от предположений* [178]. Естественно, что при условии сохранения хороших качеств оценок лучше всего использовать робастные методы оценивания.

В выборке могут присутствовать отклонения от предположений двух видов. Допустим, что наблюдаемая выборка действительно принадлежит тому закону распределения, оценки которого мы пытаемся найти. В этом случае отклонения могут быть связаны с наличием аномальных наблюдений, появление которых в выборке определяется самыми различными причинами, в том числе засорением значениями, принадлежащими другому закону. Если не учитывать наличие аномальных наблюдений, попытки оценивания параметров распределения могут привести к самым плачевным результатам. Что же делать? Естественно, надо отбраковать аномальные измерения, а затем искать оценки параметров. К сожалению, реализовать отбраковку наблюдений в общем случае оказывается совсем не просто. Наблюдения, аномальные с позиций одного закона распределения, являются естественным проявлением закономерностей второго. Если нет надежной процедуры отбраковки или практических соображений, связанных с физикой наблюдаемой величины, пытаются выйти из положения одним из следующих способов. В первом случае *усекают* выборку, отбрасывая определенную часть минимальных и/или максимальных наблюдений, и по оставшейся части оценивают параметры распределения, то есть используют так называемые α -урезанные оценки (отбрасывается αN наименьших и αN наибольших значений выборки [184]). Во втором - перед процедурой оценивания *винзорируют* выборку [178]: всем наблюдениям левее и/или правее определенных значений присваивают одинаковые значения. Эти два подхода используются при обработке наблюдений ещё с XIX века и связаны с именами Пуанкаре [51] и Винзора [272]. Обе эти процедуры далеко не всегда приводят к положительным результатам. Кроме того в обоих случаях мы имеем дело с новой генеральной совокупностью, которой принадлежит видоизмененная выборка. Более правильным следует считать третий подход, когда выборку цензурируют. Для наблюдений, попавших левее и/или правее опреде-

ленных значений, фиксируют лишь факт попадания в соответствующий интервал, опуская конкретные значения этих наблюдений. По такой цензурированной выборке оценивают параметры закона.

Другая ситуация. В выборке нет аномальных наблюдений, но наблюдаемый закон распределения отличается от предполагаемого. Такая ситуация присутствует практически всегда, так как множество законов распределения вероятностей бесконечно, а количество моделей, используемых на практике для описания наблюдаемых случайных величин, очень ограничено. Чем существенней вид предполагаемой модели отличается от реально наблюдаемого закона, тем сильнее это отражается на оценках параметров.

Очевидно, что в наблюдаемых на практике выборках и закон, пусть мало, но отличается от используемой модели, и обычно налицо аномальные наблюдения. Поэтому применение цензурирования, одностороннего или двустороннего, далеко не всегда приносит желаемый эффект.

Вопросам построения и исследования робастных оценок посвящено очень много работ (см., например [178,147]). В данном случае не преследовалась цель анализа всей совокупности подходов и методов построения робастных методов оценивания. Анализируются только свойства робастности оценок максимального правдоподобия по группированным и негруппированным наблюдениям.

3.2. Группирование наблюдений как способ получения робастных оценок

В работах [76, 81, 85] подчеркивается высокая устойчивость оценок максимального правдоподобия по группированным наблюдениям к наличию в выборке аномальных измерений, к отклонению реально наблюдаемого закона от предполагаемого, к засорению выборки данными, принадлежащими другому закону. Это подтверждается опытом эксплуатации программной системы [76] и многочисленными результатами модельных экспериментов.

Метод максимального правдоподобия является одним из наиболее популярных и эффективных методов оценивания параметров распределений. Достаточно часто приходится сталкиваться с необоснованными утверждениями, что оценки максимального правдоподобия являются робастными и, именно поэтому, предпочтительно использовать именно их. Автор является сторонником широкого использования метода максимального правдоподобия, но вместе с тем, основны-

ваясь на достаточно большом практическом опыте и результатах этого и следующего параграфа, обязан подчеркнуть, возможно очевидное для многих, что в общем случае ОМП параметров распределений не являются робастными.

Проиллюстрируем сказанное следующими примерами. Это можно было бы сделать с одинаковым эффектом на различных законах распределения, но, учитывая роль нормального распределения в теории и приложениях математической статистики, приведем примеры именно с нормальным законом. В первом примере иллюстрируется влияние аномальных ошибок на ОМП параметров нормального распределения, а во втором параметры нормального закона оцениваются по выборке, принадлежащей другому закону распределения. Для чистоты эксперимента выборки моделируются в соответствии с заданными законами.

Пример 3.2.1. Выборка по нормальному закону моделировалась с математическим ожиданием $\mu=0$ и среднеквадратическим отклонением $\sigma=1$. На рис. 3.2.1 приведены результаты статистического анализа смоделированной выборки. Вычисленные значения ОМП $\hat{\mu} = -0.0113$ и $\hat{\sigma} = 1.0198$. На этом и последующем аналогичных рисунках приведены значения статистик отношения правдоподобия, χ^2 Пирсона, Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса, вычисляемые при проверке гипотез о согласии, и соответствующие вероятности вида

$P\{S > S^*\} = \int_{S^*}^{\infty} g(s)ds$, где S^* - вычисленное значение соответствующей

статистики, $g(s)$ - предельное распределение вероятностей для статистики. Гипотеза о согласии не отвергается, если $P\{S > S^*\} > \alpha$, где α - заданный уровень значимости. Для статистик отношения правдоподобия и χ^2 Пирсона значения вероятностей приводятся при двух различных степенях свободы. Разность степеней свободы определяется количеством параметров, оцененных по выборке. При вычислении вероятностей вида $P\{S > S^*\}$ для непараметрических критериев типа Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса учитывается факт потери ими свойства “свободы от распределения” [97]. Как видим, согласие с нормальным законом очень хорошее.

Теперь допустим, что в выборку “вкралось” всего 3 аномальных наблюдения (в результате замены трех первых наблюдений: -0.7783 на -10.7783 , -0.6389 на -10.6389 , 0.1698 на 10.1698). Результаты ана-

лиза с теоретической и эмпирической функциями распределения приведены на рис. 3.2.2. Полученные ОМП параметров нормального распределения $\hat{\mu} = -0.0213$ и $\hat{\sigma} = 1.1712$. Особенно существенно наличие аномальных наблюдений повлияло на оценку среднеквадратичного отклонения. По всем критериям согласие с нормальным законом распределения будет отклонено при уровне значимости $\alpha > 0.0008$.

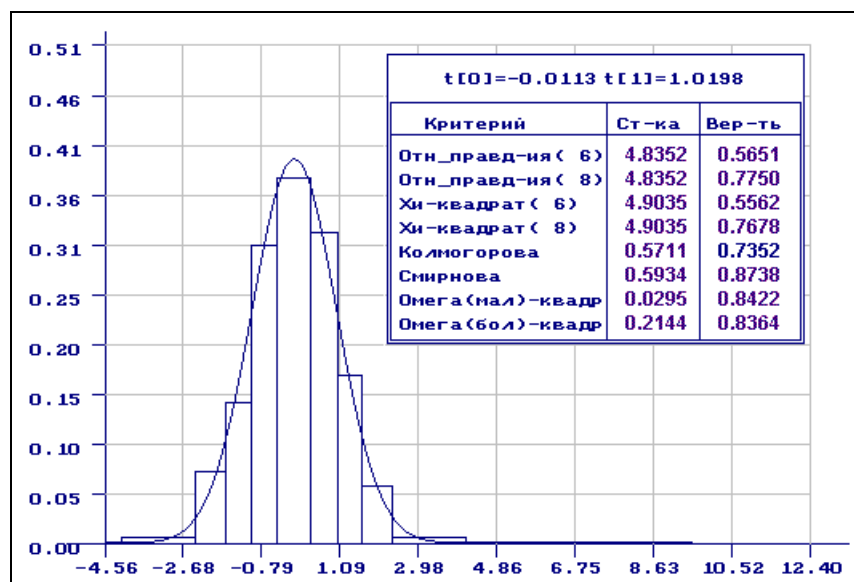


Рис.3.2.1. Результаты статистического анализа исходной выборки по негруппированным данным

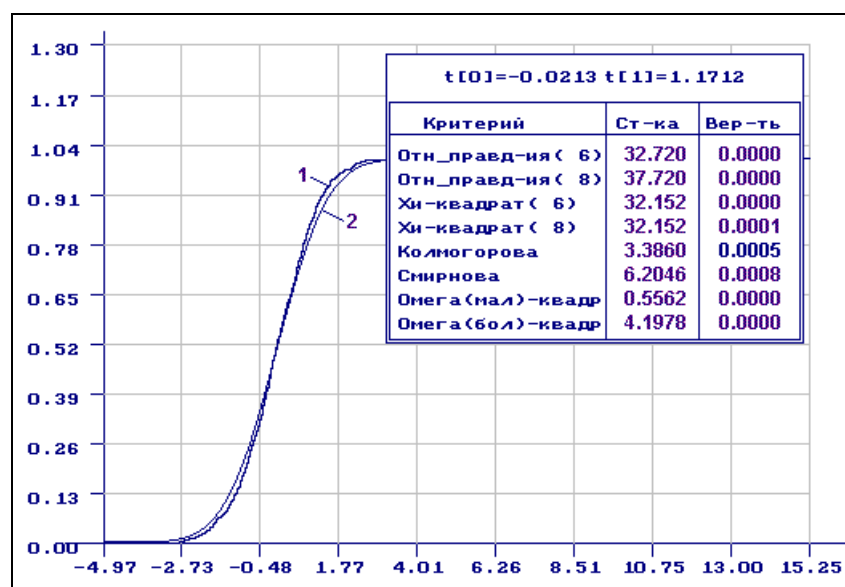


Рис.3.2.2. Эмпирическая функция распределения (1) и теоретическая функция нормального распределения (2), полученная по выборке с аномальными наблюдениями

Пример 3.2.2. Этот пример связан с использованием нормального закона распределения в ситуации, когда на самом деле выборка принадлежит распределению Лапласа. Распределение Лапласа с более “тяжелыми” хвостами, чем у нормального. На рис. 3.3.3 приведены эмпирическая и теоретические функции нормального распределения, когда по выборке, смоделированной в соответствии с распределением Лапласа, оценивались параметры нормального закона ($\hat{\mu} = 0.0033$, $\hat{\sigma} = 1.4391$). Как видно из значений статистик и соответствующих вероятностей, ни о какой близости эмпирической и теоретической функций распределения говорить не приходится.

Естественно, что использование получившихся в этих 2-х примерах нормальных законов в качестве моделей наблюдаемых выборок ни к чему хорошему не приведет.

Что же можно сделать, чтобы снизить влияние аномальных ошибок и отклонений наблюдаемых выборок от предполагаемого закона на оценки вычисляемых параметров? Мы настоятельно рекомендуем использовать перед вычислением оценок параметров процедуру предварительного группирования наблюдений. *Группирование выборки позволяет резко снизить влияние аномальных наблюдений, а иногда практически исключить последствия присутствия их в выборке. Резко снижается влияние на оценки параметров и отклонений вида наблюдаемого закона от предполагаемого.* Продемонстрируем это на выборках приведенных примеров.

На рис. 3.2.4 представлены результаты оценивания параметров нормального распределения и последующего анализа по сгруппированной выборке из примера 3.2.1, содержащей 3 аномальных наблюдения (сравните значения статистик и соответствующих вероятностей с представленными на рис. 3.2.2). Полученные ОМП параметров нормального распределения по группированным данным $\hat{\mu}_д = -0.0056$ и $\hat{\sigma}_д = 1.0428$. Визуального различия между эмпирической и теоретической функцией нормального закона в данном случае нет, поэтому соответствующие графики не приводятся.

На рис. 3.2.5 представлены результаты оценивания по сгруппированным данным параметров нормального закона по выборке из примера 3.2.2, принадлежащей распределению Лапласа (сравните результаты анализа с результатами, представленными на рис. 3.2.3). ОМП параметров нормального распределения по группированным данным $\hat{\mu}_д = 0.0174$ и $\hat{\sigma}_д = 1.2249$. В данном случае в центре области

определения случайной величины наблюдается некоторая близость эмпирической функции распределения и функции распределения нормального закона.

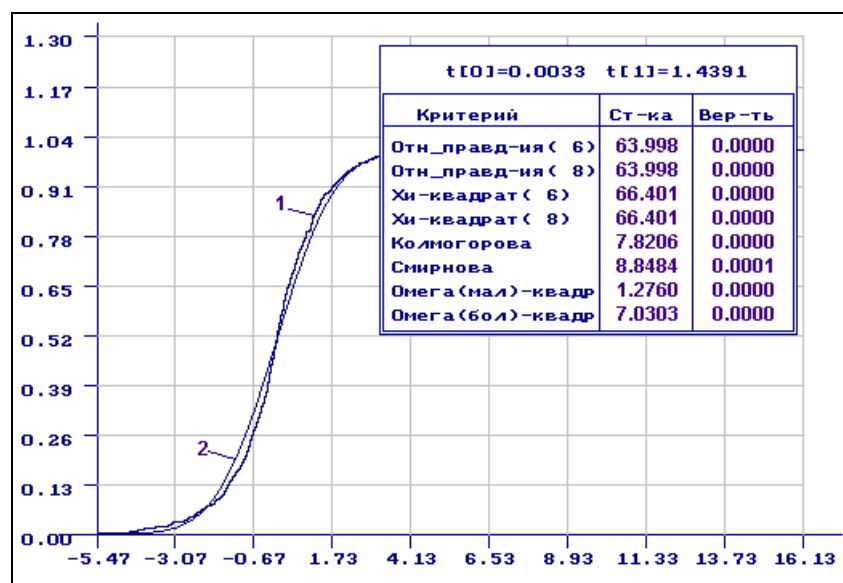


Рис.3.2.3. Эмпирическая функция распределения (1) и функция распределения нормального закона (2), найденного по выборке, принадлежащей распределению Лапласа

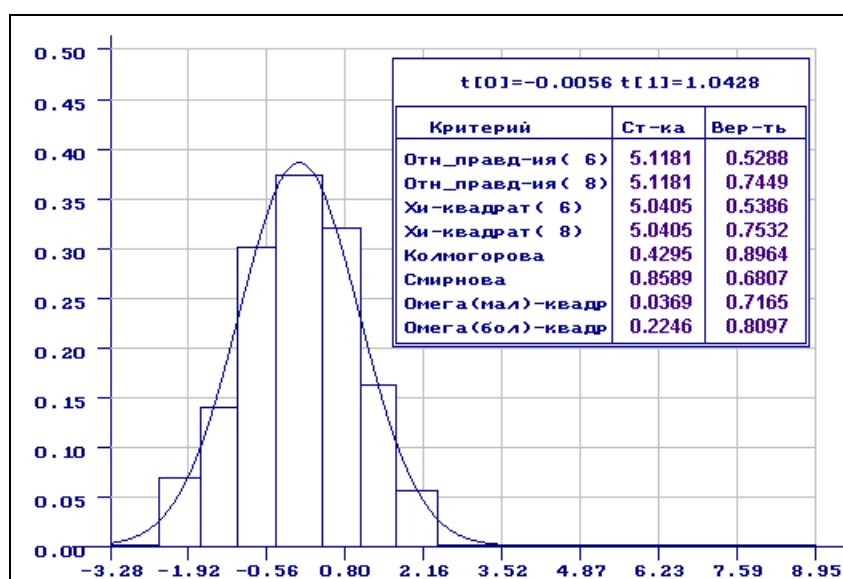


Рис. 3.2.4. Результаты оценивания по сгруппированной выборке и последующего статистического анализа при наличии в выборке аномальных измерений

Пример 3.2.3. Выборка объёмом 1000 наблюдений в была смоделирована соответствии с распределением Вейбулла с плотностью

$$f(x) = \frac{\theta_0(x - \theta_2)^{\theta_0 - 1}}{\theta_1^{\theta_0}} \exp \left\{ - \left(\frac{x - \theta_2}{\theta_1} \right)^{\theta_0} \right\}.$$

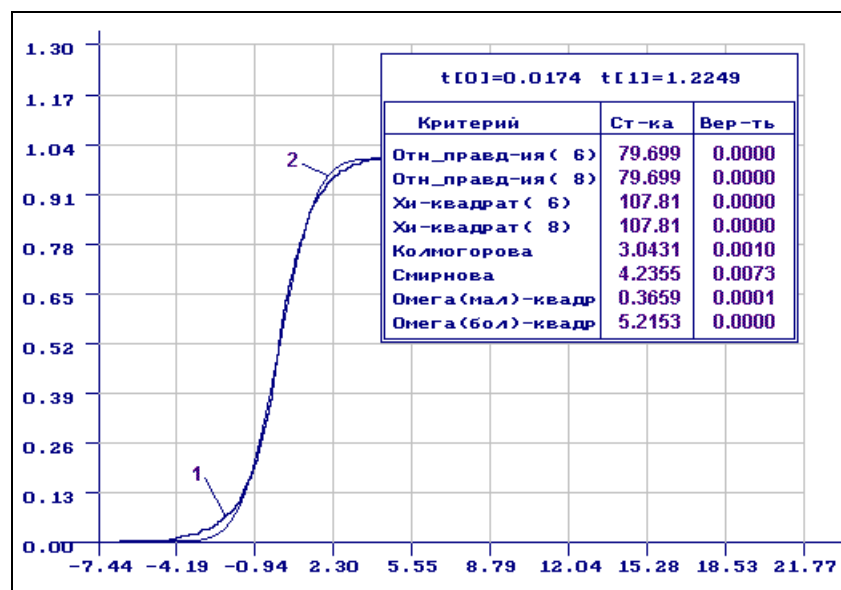


Рис. 3.2.5. Эмпирическая функция, построенная по выборке, принадлежащей распределению Лапласа (1), и теоретическая функция нормального закона (2), найденная по сгруппированной выборке

При моделировании были заданы параметры: $\theta_0 = 2$, $\theta_1 = 1$, $\theta_2 = 0$. В процессе регистрации 8 наблюдений “подверглись” сильным искажениям.

На рис. 3.2.6-3.2.7 приведены результаты статистического анализа полученной выборки. В данном случае получили закон распределения Вейбулла с параметрами $\theta_0 = 1.4433$, $\theta_1 = 1.0613$, $\theta_2 = 0$. Как видим из рис. 3.2.6, согласие по всем критериям отвергается: наличие аномальных наблюдений сыграло свою роль. На рис. 3.2.7 хорошо заметна разница между эмпирической и теоретической функциями распределения.

На рис. 3.2.8 приведены результаты статистического анализа, когда перед оцениванием выборка была разбита на интервалы равной частоты, затем по получившейся группированной выборке были найдены оценки параметров распределения $\theta_0 = 1.9668$, $\theta_1 = 1.0258$, $\theta_2 = 0$, после чего проверены гипотезы о согласии исходной выборки с полученным законом распределения. При проверке гипотез о согласии исходная выборка разбивалась на интервалы равной

вероятности. Как видим, результаты проверки гипотез о согласии по всем критериям очень хорошие.

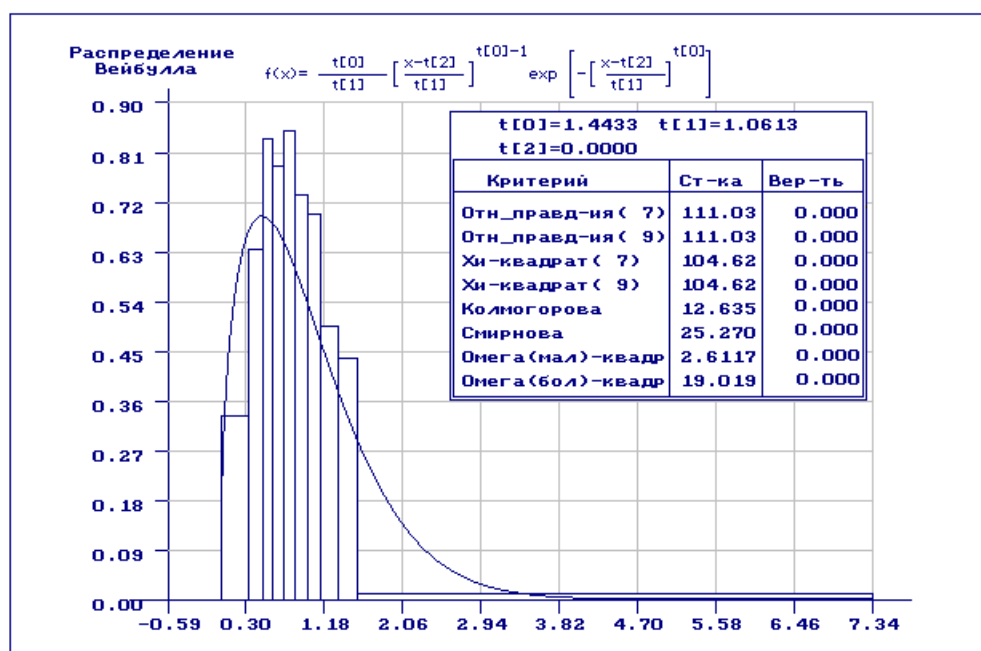


Рис. 3.2.6. Результаты статистического анализа исходной выборки по негруппированным данным

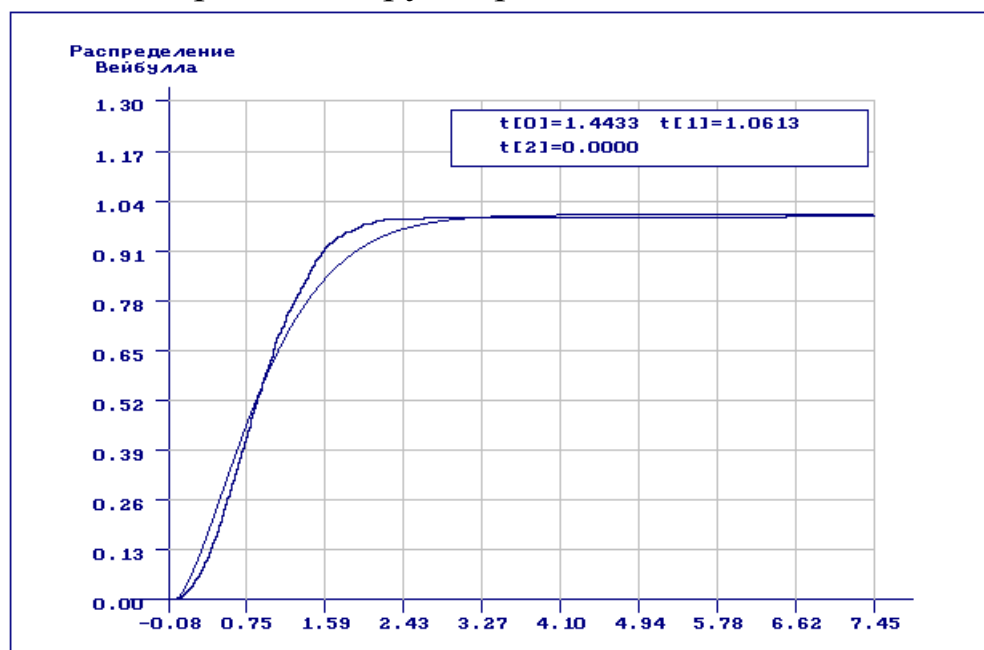


Рис. 3.2.7. Теоретическая и эмпирическая функции распределения

Отличие результатов на рис. 3.2.9 определяется тем, что при проверке гипотез о согласии исходная выборка разбивалась на интервалы в соответствии с асимптотически оптимальным группированием. В данном случае критерии отношения правдоподобия и χ^2 Пирсона оказываются более чувствительными, чем остальные: *улавливают*

наличие аномальных измерений. Гипотезы о согласии при $\alpha \geq 0.0027$ по этим критериям должны быть отвергнуты.

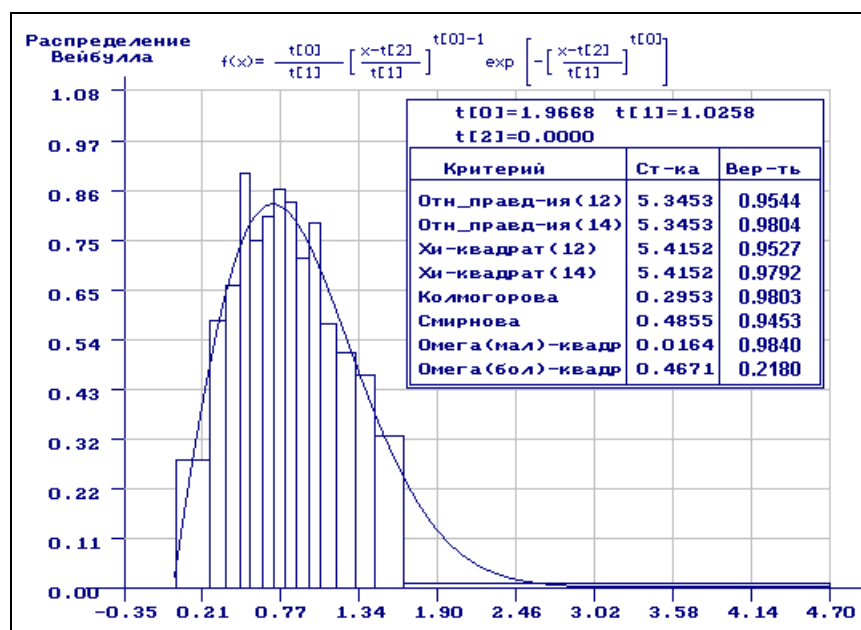


Рис. 3.2.8 Оценивание с предварительным равночастотным группированием и проверкой гипотез о согласии с разбиением на равночастотные интервалы.

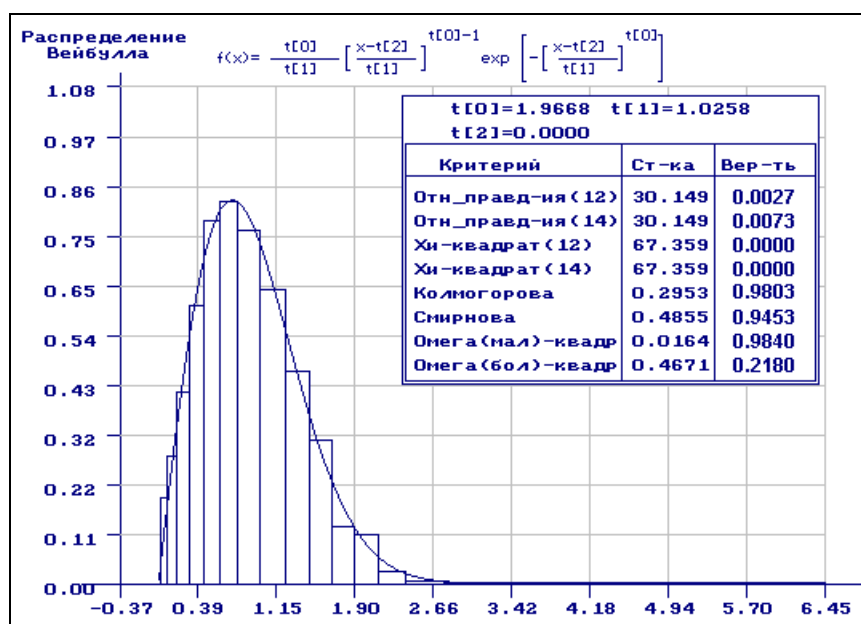


Рис. 3.2.9. Оценивание с предварительным равночастотным группированием и проверкой гипотез о согласии с разбиением на асимптотически оптимальные интервалы

Приведенные на рис. 3.2.10-3.2.11 результаты анализа, аналогичны тем, что представлены на рис. 3.2.8-3.2.9, но перед оцениванием выборка была разбита на асимптотически оптимальные интер-

валы. Получены оценки параметров $\theta_0 = 1.8798$, $\theta_1 = 1.0324$, $\theta_2 = 0$. Если при проверке гипотез исходная выборка разбивалась на интервалы равной вероятности (рис. 3.2.10), то гипотеза о согласии по всем критериям принимается. При использовании асимптотически оптимального группирования гипотеза о согласии по критериям отношения правдоподобия и χ^2 Пирсона должна быть отвергнута (рис. 3.2.11). Если мы сравним эти результаты, с результатами, представленными на рис. 3.2.8-3.2.9, то увидим, что уровень согласия в данном случае ниже. То есть, полученные оценки оказались хуже, а способ их определения *более чувствителен* к аномальным наблюдениям.

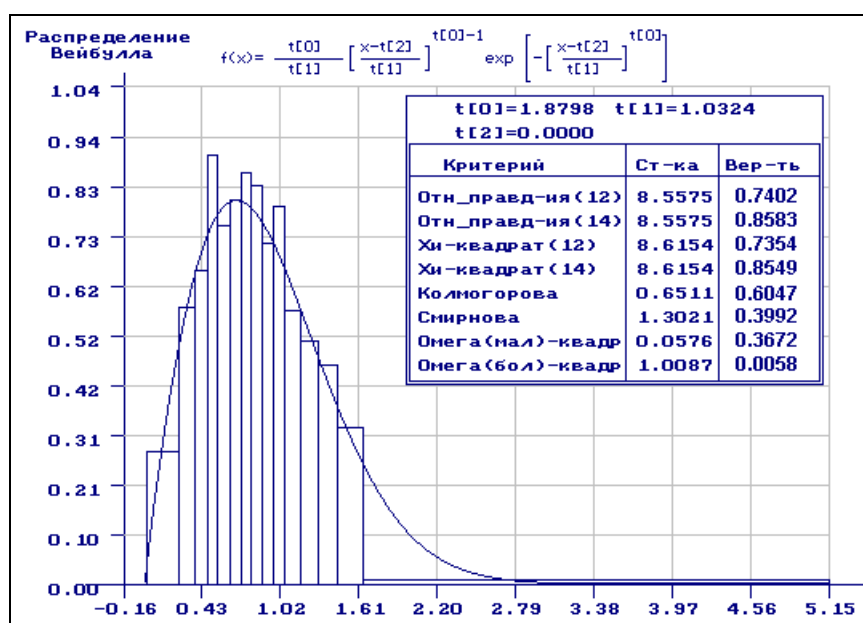


Рис. 3.2.10. Оценивание с предварительным асимптотически оптимальным группированием. При проверке согласия использованы равночастотные интервалы.

Приведем ещё один пример, подчеркивающий устойчивость оценок максимального правдоподобия по группированным данным. Он связан с использованием нормального закона распределения в ситуации, когда на самом деле выборка принадлежит распределению Коши.

Пример 3.2.4. Распределение Коши это распределение с “тяжелыми” хвостами, а такое отклонение от нормальности особенно сильно отражается на оценках параметров нормального закона. На рис. 3.2.12-3.2.13 приведены эмпирическая и теоретические функции нормального распределения при использовании обычных оценок максималь-

ного правдоподобия (рис. 3.2.12, оценки параметров нормального распределения: $\mu = -0.7781$, $\sigma = 7.1575$) и оценок максимального правдоподобия по группированным данным (рис. 3.2.13, оценки: $\mu = -0.2880$, $\sigma = 2.1893$).

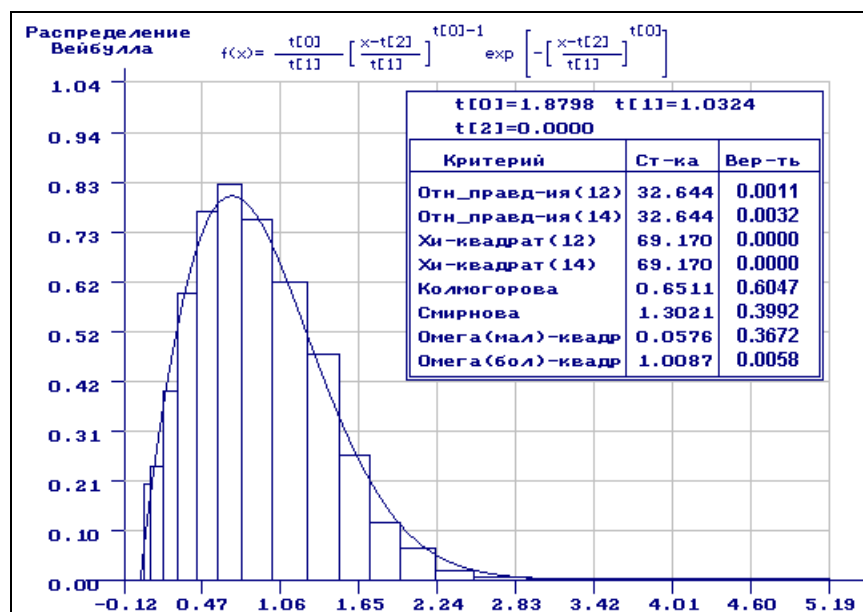


Рис. 3.2.11. Оценивание с предварительным группированием с разбиением на асимптотически оптимальные интервалы. При проверке согласия также использовано асимптотически оптимальное группирование.

Качественная картина, хорошо прослеживаемая на графиках, говорит сама за себя: во втором случае можно даже говорить об определенной близости эмпирической и теоретической функций распределения. Выборка объёмом 100 наблюдений моделировалась по закону Коши с функцией плотности $f(x) = \frac{\theta_0}{\pi[\theta_0^2 + (x - \theta_1)^2]}$ и параметрами $\theta_0 = 1$, $\theta_1 = 0$.

Подведем итоги вышесказанному. Группирование наблюдений перед оцениванием и последующее оценивание параметров по группированной выборке позволяет получать устойчивые оценки. Когда мы говорим об оценках по группированным данным, то имеем ОМП, которые определяются в результате максимизации функции правдоподобия вида

$$L(\theta) = \prod_{i=1}^k P_i^{n_i}(\theta),$$

где $P_i(\theta) = \int_{x_{i-1}}^{x_i} f(x, \theta) dx$ - вероятность попадания наблюдения в i -й интервал значений, k - число интервалов, но только не оценки по методу моментов с последующим использованием поправок типа Шепарда.

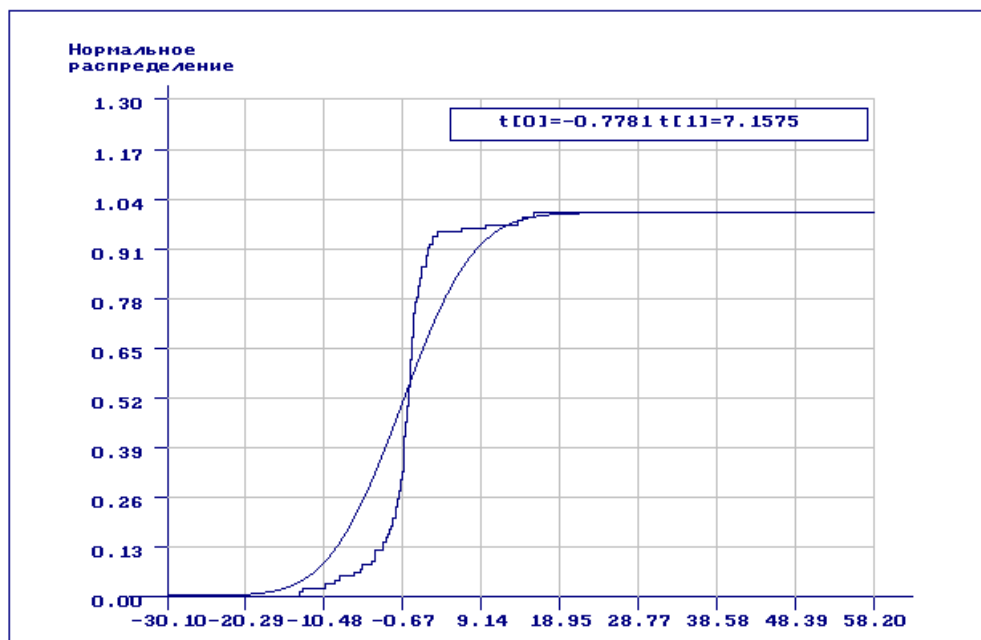


Рис. 3.2.12. Эмпирическая функция распределения и теоретическая функция нормального закона распределения, найденная по выборке, принадлежащей распределению Коши.

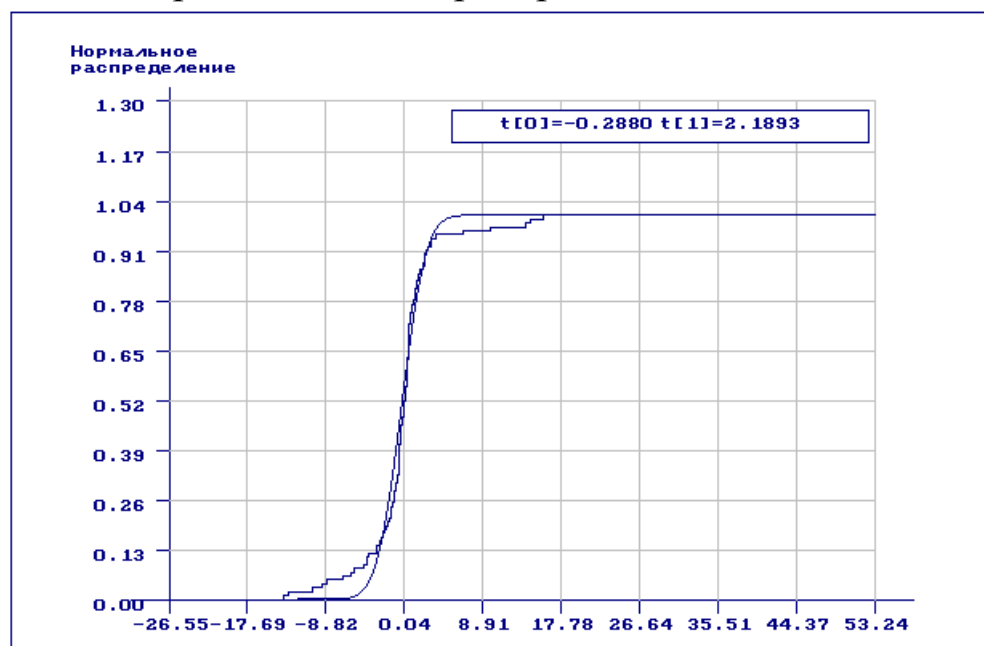


Рис. 3.2.13. Эмпирическая функция распределения и теоретическая функция нормального закона распределения, найденная по сгруппированной выборке, принадлежащей распределению Коши.

Естественно, мы предлагаем использовать оценки по группированным данным не вместо, а вместе с оценками по негруппированным наблюдениям. Качество тех и других зависит от степени засоренности выборки аномальными наблюдениями или близости к предполагаемому закону распределения.

Остаётся вопрос, как группировать? Можно различным образом. Не стоит, конечно, рассматривать крайние случаи: сгруппировать можно и так, что в группированной выборке не останется никакой информации о законе и его параметрах. Реально на интервалы разбивают область, определяемую размахом выборки. Это могут быть или интервалы равной длины, или интервалы равной вероятности (равной частоты), или асимптотически оптимальные интервалы, или интервалы, сформированные по какому-то другому принципу. Обычно наиболее устойчивыми к отклонениям оказываются оценки при разбиении выборки на интервалы равной вероятности. В то же время в случае асимптотически оптимального группирования потери информации о параметрах закона распределения, связанные с группированием, существенно меньше, чем при равновероятном. Если мы знаем, что отклонения от наших предположений в выборке минимальны, то использование полученных таблиц асимптотически оптимального группирования позволяет резко сократить объемы хранимых данных без существенной потери информации о законе распределения. Но всё-таки в общем случае здесь следует ожидать большей чувствительности оценок к отклонениям от предположений.

С другой стороны, достаточно часто мешающая информация, связанная с засорением выборки, оказывает меньшее влияние на оценки, чем потери информации от группирования при асимптотически оптимальном группировании. В некоторых случаях оценки с использованием асимптотически оптимального группирования оказываются так же устойчивыми, как и при равновероятном, и при этом показывают лучшие результаты. Поэтому рекомендуется вычислять две оценки по группированным данным с использованием как оптимального, так и равновероятного группирования, и остановиться на той оценке, которая дает лучшее согласие с исходной выборкой.

Группирование наблюдений приводит к потере в количестве информации Фишера о параметре (параметрах) распределения. При асимптотически оптимальном группировании, как показывают таблицы асимптотически оптимального группирования, эти потери составляют в среднем порядка 2% при оценивании одного параметра и 10-11 интервалах группирования и порядка 5% при оценивании двух

параметров и 15 интервалах группирования [42]. Это вызывает соответствующий рост асимптотической дисперсии эффективных оценок. В то же время, так как ОМП по негруппированным наблюдениям (и многие другие оценки) в общем случае чрезвычайно чувствительны к наличию аномальных наблюдений или отклонению наблюдаемой выборки от предполагаемого закона распределения, то дисперсия таких оценок может быть существенно больше асимптотической. Напротив ОМП по группированным данным являются устойчивыми к таким отклонениям в наблюдениях. Реально, при наличии в выборке неоднократно упоминаемых отклонений, дисперсия оценок по группированным данным оказывается меньше, чем по негруппированным. Т.е. вклад в дисперсию от потерь в асимптотике оказывается несоизмеримо мал по сравнению со вкладом, связанным с наличием отклонений.

Существенное различие в оценках, вычисляемых по негруппированным и сгруппированным данным, может служить сигналом о том, что между имеющимися данными и нашими предположениями (знаниями о виде закона распределения) имеются некоторые разногласия: либо налицо засорение выборки, либо в измерения вкрались ошибки, либо наши предположения о виде закона распределения (модели) неверны.

Выводы

1. Предварительное группирование исходной выборки и последующее вычисление ОМП по группированным данным приводит к робастным оценкам, устойчивым как к наличию в исходной выборке аномальных измерений, так и к отклонениям закона распределения выборки от предполагаемого.

2. Процедура предварительного группирования реализована в программном обеспечении. Возможно использование равномерного, равновероятного и асимптотически оптимального группирования. На основании исходной негруппированной выборки может создаваться соответствующая группированная выборка. Реализован режим предварительного группирования при оценивании, в том числе при идентификации закона распределения.

3. Высокая устойчивость к присутствию в выборке грубых искажений или принадлежности выборки к другому закону распределения оценок максимального правдоподобия по группированной выборке позволяет использовать их в процедурах отбраковки аномальных наблюдений.

3.3. Функции влияния и робастность оценок

В работах [76,81,85,100] и предыдущих разделах подчеркивается высокая устойчивость оценок максимального правдоподобия (ОМП) по группированным наблюдениям к наличию в выборке аномальных измерений, к отклонению реально наблюдаемого закона от предполагаемого, к засорению выборки данными, принадлежащими другому закону. Всё это подтверждается опытом эксплуатации программной системы [76] и многочисленными результатами модельных экспериментов. В данном разделе, основные результаты которого изложены в [93], свойство робастности ОМП исследуется с позиций *функции влияния*, предложенной Хэмпелом [221,222]. Именно анализ функций влияния ОМП параметров различных распределений, в том числе того множества распределений, которое включено в программную систему [76], позволяет утверждать, что ОМП по негруппированным данным, вопреки порой бытующему заблуждению, в большинстве своём являются неробастными. В то же время ОМП по группированным данным всегда оказываются робастными.

Влияние ещё одного наблюдения на очень большую выборку может характеризоваться функцией (кривой) влияния, которая определяется следующим образом [178]

$$IF(x; F, T) = \lim_{s \rightarrow 0} \frac{T((1-s)F + s\delta_x) - T(F)}{s},$$

где δ_x - единичная масса в точке x , F - функция распределения, к которому принадлежит выборка, $T(F)$ - вычисляемая статистика.

Функция влияния позволяет оценить относительное влияние отдельного наблюдения на значение статистики критерия или оценку параметров. Если функция влияния неограничена, то резко выделяющиеся наблюдения могут приводить к существенным изменениям оценок или статистик. *Чувствительность к большой ошибке* может характеризоваться величиной

$$\gamma^* = \sup_x |IF(x; F, T)|.$$

Для асимптотически эффективных оценок, к которым относятся оценки максимального правдоподобия по негруппированным данным, функция влияния удовлетворяет равенству [178]

$$IF(x; F_0, T) = J^{-1}(F_0) \frac{\partial \ln f_0}{\partial \theta}, \quad (3.3.1)$$

где $J(F_0)$ - количество информации Фишера.

Для оценок типа максимального правдоподобия (M -оценок), где всякая оценка T_n определяется как решение экстремальной задачи на минимум вида

$$\sum_{i=1}^n \rho(x_i; T_n) \rightarrow \min$$

или как решение неявного уравнения

$$\sum_{i=1}^n \psi(x_i; T_n) = 0,$$

где ρ - произвольная функция, $\psi(x; \theta) = \partial \rho(x; \theta) / \partial \theta$, функция влияния имеет вид [184]

$$IF(x; F, T) = -\frac{\psi(x; T)}{\lambda_F^{\odot}(T(F))}, \quad -\infty < x < \infty,$$

где

$$\lambda_F'(T(F)) = \frac{d}{dt} \int \psi(x; t) dF(x), \quad -\infty < t < \infty.$$

В случае ОМП по группированным данным

$$\psi(x; \theta) = \frac{\partial \ln P_i(\theta)}{\partial \theta}, \quad x_{i-1} < x < x_i,$$

$$\lambda_F'(\theta) = \frac{d}{d\theta} \int \psi(x; \theta) dF(x) = \sum_{j=1}^k \left(\frac{\partial^2 \ln P_j(\theta)}{\partial \theta^2} \right) P_j(\theta)$$

и функция влияния будет иметь вид

$$IF(x; F, \theta) = -\frac{\frac{\partial \ln P_i(\theta)}{\partial \theta}}{\sum_{j=1}^k \left(\frac{\partial^2 \ln P_j(\theta)}{\partial \theta^2} \right) P_j(\theta)}, \quad x_{i-1} < x < x_i. \quad (3.3.2)$$

Для оценок, использующих квантили, соответствующие асимптотически оптимальному группированию [88], и являющихся одним из частных случаев L -оценок, функция влияния имеет вид [184]

$$IF(x; F, T) = \sum_{j=1}^{k-1} a_j \left(p_j - c(F^{-1}(p_j) - x) \right) / f(F^{-1}(p_j)), \quad (3.3.3)$$

где a_j - коэффициенты при выборочных квантилях в формуле для вычисления L -оценок, $p_j = \sum_{l=1}^j P_l(\theta)$, $c(u) = \begin{cases} 1, & u \geq 0, \\ 0, & u < 0. \end{cases}$

Были рассмотрены функции влияния для оценок параметров множества распределений, включенных в программную систему [76].

Приводимые ниже функции влияния построены при конкретных значениях параметров и характеризуют качественную картину их поведения на области определения случайных величин. На рис. 3.3.1-3.3.2 представлены функции влияния для оценок параметров сдвига и масштаба нормального распределения, определяемых методом максимального правдоподобия по негруппированным и сгруппированным данным. Функция влияния для ОМП параметра сдвига по негруппированным данным имеет вид

$$IF(x; F_{\theta}, \theta_0) = x - \theta_0,$$

для ОМП параметра масштаба -

$$IF(x; F_{\theta}, \theta_1) = \frac{1}{2} \left[\frac{(x - \theta_0)^2}{\theta_1} - \theta_1 \right].$$

Функции влияния неограничены, и этим определяется чувствительность данных оценок к ошибкам измерения и засорению выборки. Напротив, функции влияния оценок параметров нормального распределения по группированным данным ограничены. Это ещё раз подчеркивает высокую устойчивость получаемых по группированным наблюдениям оценок, подтверждаемую практикой. На этих и последующих рисунках функции влияния для ОМП по группированным данным соответствуют случаю использования асимптотически оптимального группирования.

Аналогично, на рис. 3.3.3-3.3.4 приведены функции влияния для параметров распределения Вейбулла. Функция влияния для ОМП основного параметра по негруппированным данным имеет вид

$$IF(x; F_{\theta}, \theta_0) = \frac{\theta_0 [(1-t) \ln t + 1]}{1 + \pi / 6 + C^2 - 2C},$$

где C - постоянная Эйлера и $t = \left(\frac{x - \theta_2}{\theta_1} \right)^{\theta_0}$, для ОМП параметра масштаба -

$$IF(x; F_{\theta}, \theta_1) = \frac{\theta_1}{\theta_0} (t - 1).$$

Для основного параметра функция влияния по негруппированным данным неограничена снизу на левой и правой границе области определения случайной величины, для масштабного параметра - неограничена сверху на правой границе. В то же время для группиро-

ванных наблюдений функции влияния являются ступенчатыми ограниченными функциями.

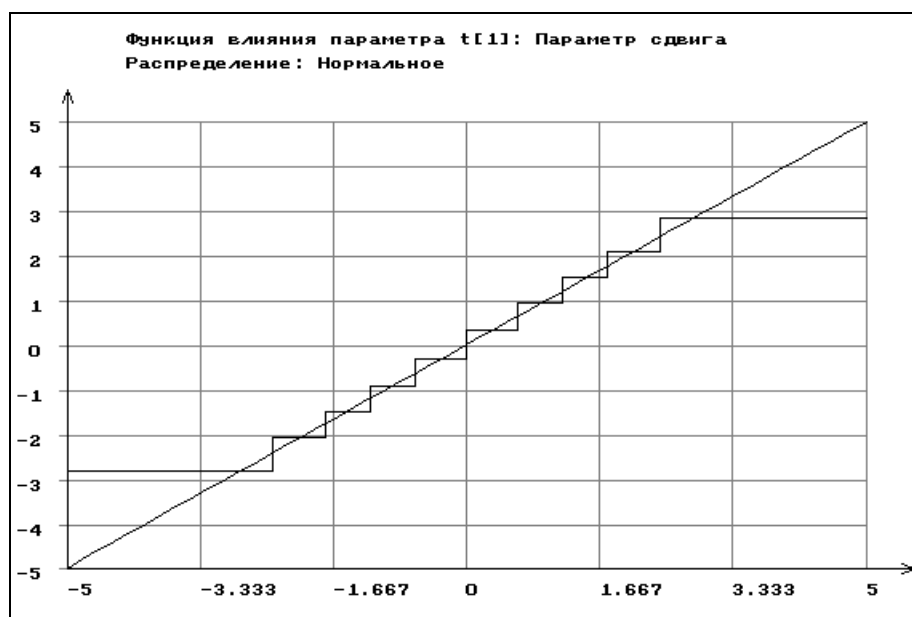


Рис. 3.3.1. Функции влияния для параметра сдвига нормального распределения по негруппированным (прямая) и сгруппированным данным (ступенчатая линия)



Рис. 3.3.2. Функции влияния для параметра масштаба нормального распределения по негруппированным и сгруппированным данным (ступенчатая линия)

Совершенно другую картину мы наблюдаем для ОМП по негруппированным наблюдениям для параметров распределения

Коши (см. рис. 3.3.5-3.3.6). Функция влияния для ОМП параметра сдвига по негруппированным данным имеет вид

$$IF(x; F_0, \theta_1) = \frac{4\theta_0 t}{1+t^2},$$

где $t = \frac{x - \theta_1}{\theta_0}$.

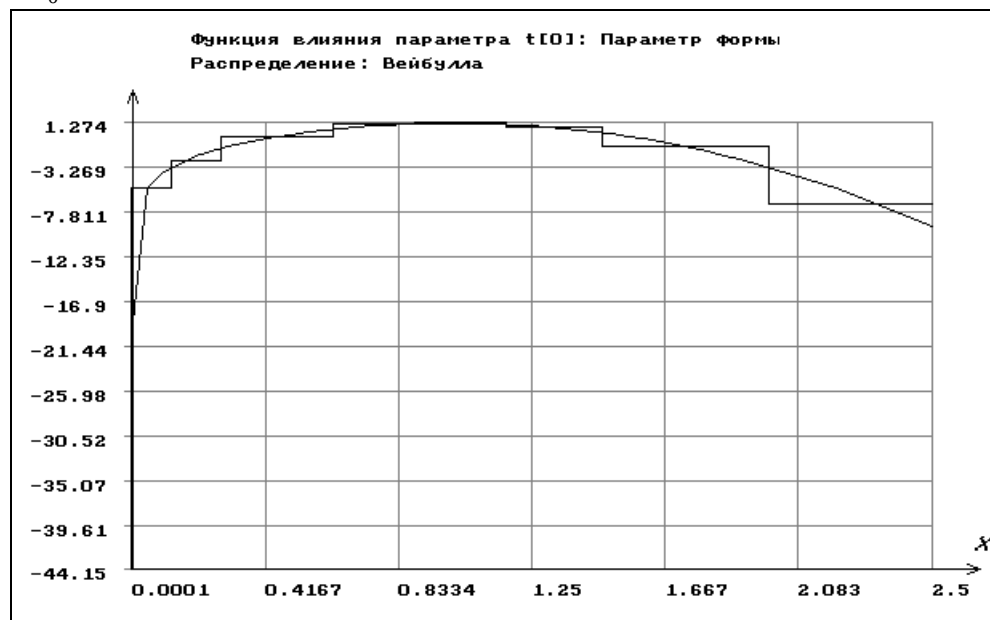


Рис. 3.3.3. Функции влияния для основного параметра распределения Вейбулла по негруппированным и сгруппированным данным (ступенчатая линия)



Рис. 3.3.4. Функции влияния для параметра масштаба распределения Вейбулла по негруппированным и сгруппированным данным

Для ОМП параметра масштаба -

$$IF(x; F_\theta, \theta_0) = 2\theta_0 \left(1 - \frac{2}{1+t^2} \right).$$

Их функции влияния ограничены на области определения случайной величины, что говорит о робастности этих оценок, их устойчивости к грубым ошибкам измерений.

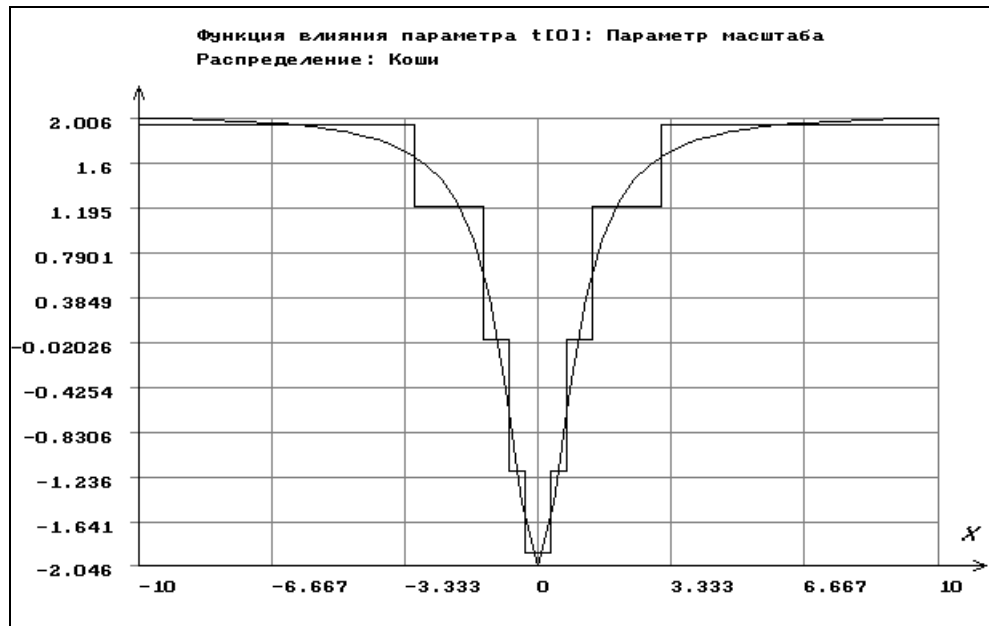


Рис. 3.3.5. Функции влияния для параметра масштаба распределения Коши по негруппированным (непрерывная) и сгруппированным данным (ступенчатая линия)

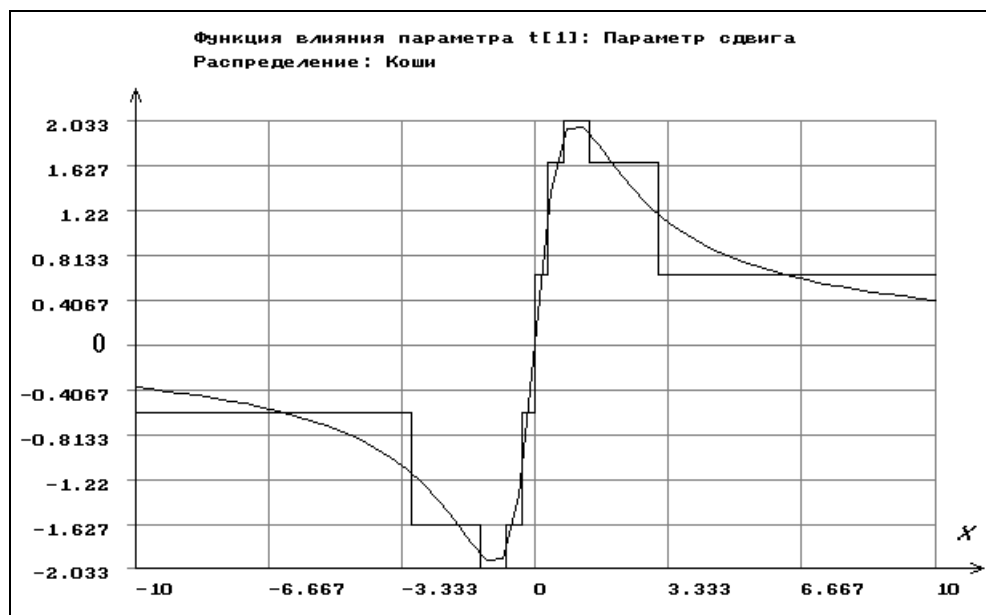


Рис. 3.3.6. Функции влияния для параметра сдвига распределения Коши по негруппированным и сгруппированным данным

Для логистического распределения функция влияния ОМП параметра масштаба по негруппированным данным имеет вид

$$IF(x; F_{\theta}, \theta_1) = \frac{9\theta_1}{\pi^2 + 3} \left(t - 1 - 2 \frac{te^{-t}}{1 - e^{-t}} \right),$$

где $t = \frac{\pi(x - \theta_0)}{\theta_1 \sqrt{3}}$, (см. рис. 3.3.7). Из её неограниченности следует, что соответствующая оценка неробастна. В то же время функция влияния ОМП параметра сдвига

$$IF(x; F_{\theta}, \theta_0) = 3\theta_1 \left(1 - 2 \frac{e^{-t}}{1 + e^{-t}} \right)$$

ограничена сверху и снизу, А это свидетельствует о робастности ОМП этого параметра.



Рис. 3.3.7. Функции влияния для параметра масштаба логистического распределения по негруппированным и сгруппированным данным (ступенчатая линия)

Ситуация, которую мы наблюдаем для функций влияния ОМП по негруппированным наблюдениям параметров распределений Коши и логистического (параметр сдвига), оказывается явно нетипичной. Для ОМП параметров остальных законов распределения, включенных в программную систему [76], а в совокупности это 26 законов и семейств непрерывных распределений, функции влияния неограничены, откуда следует *неробастность* этих оценок. С другой

стороны, функции влияния для ОМП по группированным данным всегда представляют собой ограниченные ступенчатые зависимости, что свидетельствует о робастности этих оценок.

Функции влияния L -оценок с использованием оптимальных порядковых статистик, как следует из вида соотношения (3.3.3), также представляют собой ступенчатые ограниченные зависимости, что говорит о робастности этих оценок. Это же подтверждают и проведенные эксперименты по моделированию выборок, их засорению, оцениванию параметров и анализу.

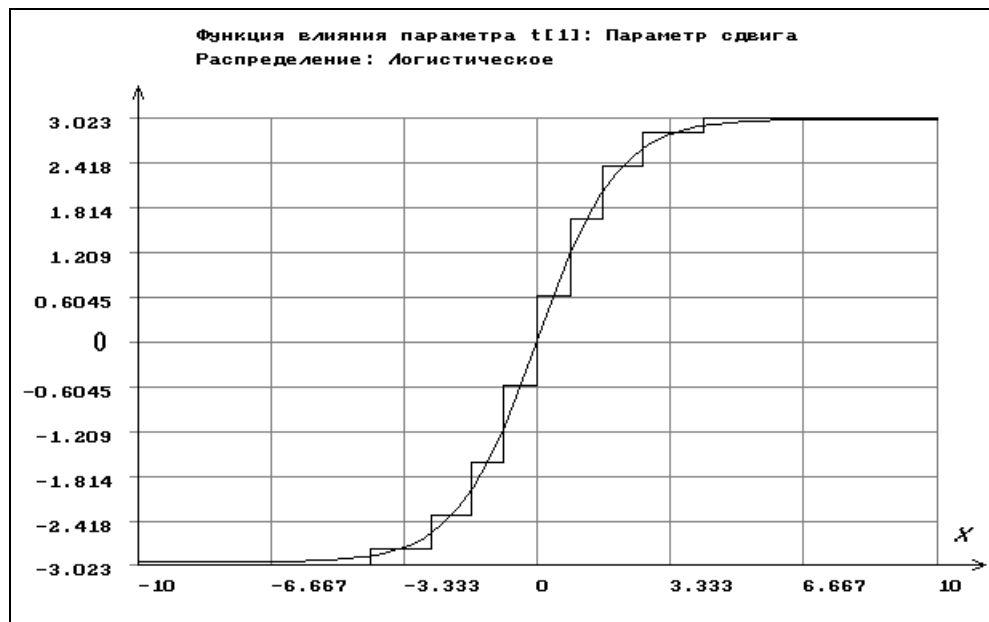


Рис. 3.3.8 Функции влияния для параметра масштаба логистического распределения по негруппированным и сгруппированным данным (ступенчатая линия)

Выводы

Таким образом, анализ функций влияния оценок по негруппированным и группированным выборкам ещё раз позволяет сделать следующие выводы.

1. За редким исключением ОМП по негруппированным наблюдениям являются неробастными.
2. Напротив, ОМП по группированным данным и оптимальные оценки параметров сдвига и масштаба по выборочным квантилям для больших выборок устойчивы как к аномальным ошибкам измерений, так и к отклонениям наблюдаемого закона от предполагаемого.

3.4. Робастное оценивание и проблема отбраковки аномальных наблюдений

При вычислении оценок параметров распределений присутствие единственного аномального наблюдения может приводить к оценкам, которые совершенно не согласуются с выборочными данными.

В борьбе с грубыми погрешностями измерений, если они не были обнаружены в процессе измерений, используют два подхода:

- исключение резко выделяющихся аномальных измерений из дальнейшей обработки;
- использование робастных методов обработки.

Во втором случае обычно подразумевается и дальнейшее выделение аномальных наблюдений, так как именно они могут нести необходимую исследователю информацию.

Понятно желание каждого исследователя, чтобы найденные оценки были как можно менее чувствительны к аномальным наблюдениям. В противном случае прежде чем переходить к оцениванию, приходится использовать процедуры исключения грубых ошибок измерений, что выливается в непростую задачу, которая, как справедливо отмечается в [127], без устойчивых методов надёжно не решается. Соглашаясь в принципе со справедливыми доводами, приводимыми в этой работе в поддержку вывода о неустойчивости параметрических методов отбраковки резко выделяющихся наблюдений, мы не склонны хоронить параметрические методы, так как при определённых условиях получаются очень *хорошие результаты* [85,89,99].

Параметрическая процедура отбраковки грубых ошибок измерений в одномерной ситуации выглядит следующим образом [127]. Рассматривается ситуация, когда $x_1, x_2, \dots, x_N$ числа. Резко выделяется одно наблюдение, для определенности $x_{\max}$. При нулевой гипотезе H_0 наблюдения $x_1, x_2, \dots, x_N$ рассматриваются как реализация независимых одинаково распределённых случайных величин $X_1, X_2, \dots, X_N$ с функцией распределения $F(x)$. При альтернативной гипотезе H_1 случайные величины $X_1, X_2, \dots, X_N$ также независимы, $X_1, X_2, \dots, X_{N-1}$ имеют распределение $F(x)$, а X_N - распределение $G(x)$, которое “существенно сдвинуто вправо” относительно $F(x)$, например $G(x) = F(x - A)$, где A достаточно велико. Если $x_{\max} \leq d$, то принимается гипотеза H_0 , в противном случае - гипотеза H_1 . При справедливости нулевой гипотезы $P\{\max_{1 \leq i \leq n} X_i \leq d\} = [F(d)]^N = 1 - \alpha$, и крити-

ческое значение $d = d(\alpha, N)$ определяется из уравнения $F(d) = \sqrt[N]{1 - \alpha}$. При больших N и малых α

$$\sqrt[N]{1 - \alpha} = 1 - \frac{\alpha}{N} + O\left(\left(\frac{\alpha}{N}\right)^2\right),$$

поэтому в качестве хорошего приближения к $d(\alpha, N)$ можно рассматривать $\left(1 - \frac{\alpha}{N}\right)$ -квантиль распределения $F(x)$. Если рассматриваем принадлежность к выборке $x_{\min}$, то гипотеза H_0 принимается при $x_{\min} \geq d_1$. Если справедлива гипотеза H_0 , то $P\{\min_{1 \leq i \leq n} X_i \geq d_1\} = [1 - F(d_1)]^N = 1 - \alpha$. Тогда значение $d_1 = d_1(\alpha, N)$ определяется из уравнения $1 - F(d_1) = \sqrt[N]{1 - \alpha}$.

Неустойчивость такой процедуры отбраковки в основном связана с возможно неточной идентификацией закона $F(x)$ из-за влияния грубых ошибок на оценки параметров и последующего неверного выбора закона из заданного класса с помощью критериев согласия.

В задаче отбраковки аномальных наблюдений на разных этапах её решения к статистическим процедурам оценивания и проверки гипотез предъявляются прямо противоположные требования. На этапе идентификации закона распределения и оценивании его параметров методы должны быть как можно менее чувствительны к наличию аномальных ошибок измерений. Наоборот, на последующем этапе исключения аномальных измерений критерий должен улавливать их наличие и позволять отсекают. Отметим, что все критерии согласия не достаточно чувствительны к аномальным наблюдениям и не могут использоваться для отбраковки.

Замечание. Наличие в выборке аномальных измерений или отклонение закона распределения, к которому принадлежит выборка, от предполагаемого особенно сильно сказываются на оценках масштабного параметра используемого закона и существенно в меньшей степени на оценках параметра сдвига. Напротив, как показали проведенные нами исследования, на предельные распределения статистик различных критериев согласия (при сложных гипотезах) в большей степени влияет оценивание по выборке параметра сдвига и в меньшей степени - оценивание масштабного параметра.

Таким образом, при идентификации (при оценивании параметров распределений) мы должны использовать робастные

алгоритмы (устойчивые к наличию аномальных наблюдений). В этой связи на первом этапе рекомендуется использовать оценки по группированным данным. Как уже говорилось, их важным достоинством является малая чувствительность к случайным выбросам. Причем для большей устойчивости оценок следует осуществлять разбиение выборки на интервалы равной вероятности (равночастотные интервалы). В то же время необходимо учитывать, что в общем случае в выборке, сгруппированной с использованием асимптотически оптимального группирования содержится существенно больше информации о параметрах распределения, чем при равновероятном группировании. Это значит, что потери информации от группирования при асимптотически оптимальном группировании могут быть меньше, чем мешающая информация, связанная с засорением выборки. Поэтому желательно на первом этапе находить не одну, а две оценки по группированным данным с использованием как оптимального, так и равновероятного группирования, и остановиться на той оценке, которая дает лучшие результаты (лучшее согласие). Это не отрицает и применения разбиения выборки на интервалы равной длины. Правильный выбор закона $F(x)$ и определение робастных оценок его параметров гарантируют успех на последующем этапе отбраковки в соответствии с изложенной выше параметрической процедурой.

Продemonстрируем сказанное на конкретном примере исключения аномальных измерений из исходной негруппированной выборки.

Пример 3.4.1. Выборка объёмом 1000 наблюдений была смоделирована в соответствии с логистическим распределением с плотностью

$$f(x) = \frac{\pi}{\theta_1 \sqrt{3}} \exp \left\{ -\frac{\pi(x - \theta_0)}{\theta_1 \sqrt{3}} \right\} / \left[1 + \exp \left\{ -\frac{\pi(x - \theta_0)}{\theta_1 \sqrt{3}} \right\} \right]^2.$$

При моделировании были заданы параметры: $\theta_0 = 1, \theta_1 = 1$. В процессе регистрации 3 наблюдения “подверглись” сильным искажениям.

На рис. 3.4.1 приведены результаты статистического анализа полученной выборки. В данном случае получили логистический закон распределения с параметрами $\theta_0 = 1.0133, \theta_1 = 1.1508$. Как видим из рис. 3.4.1, согласие по всем критериям плохое: наличие аномальных наблюдений сыграло свою роль.

А на рис. 3.4.2 приведены результаты статистического анализа, когда перед оцениванием выборка была разбита на интервалы, а

затем по получившейся группированной выборке были найдены оценки параметров распределения $\theta_0 = 1.0066$, $\theta_1 = 1.0058$, после чего проверены гипотезы о согласии исходной выборки с полученным законом распределения. Как видим, результаты проверки гипотез о согласии по всем критериям очень хорошие. Это является косвенным подтверждением того, что *все попытки выделения с помощью критериев согласия грубых ошибок измерений заведомо обречены на неудачу.*

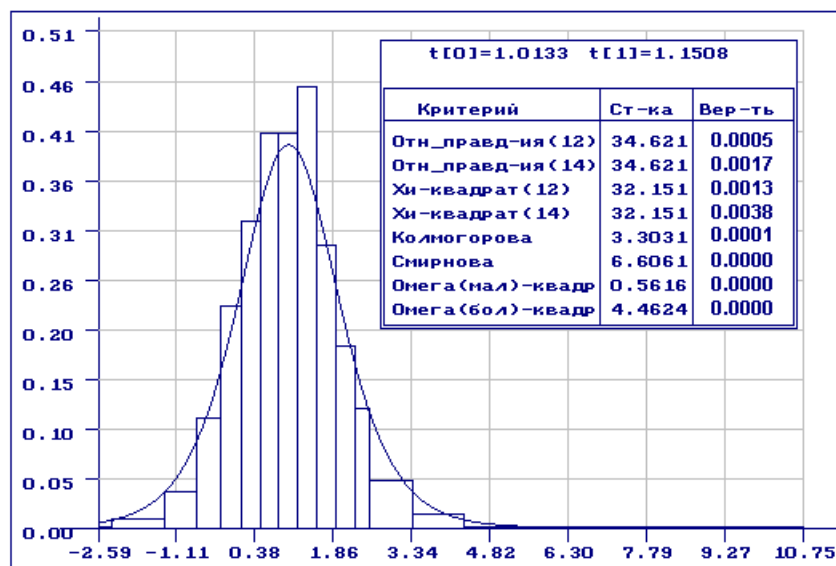


Рис. 3.4.1. Результаты оценивания параметров логистического распределения и проверки согласия с “исходной” выборкой



Рис. 3.4.2. Результаты оценивания параметров логистического распределения по сгруппированной выборке и проверки согласия с “исходной” выборкой

Для отбраковки аномальных наблюдений зададимся уровнем значимости $\alpha = 0.1$, при объёме выборки $N = 1000$ и найденном

векторе параметров $\theta^T = [1.0066, 1.0058]$ логистического распределения определяем критическое значение $d_1 = -4.101$ из условия $F(d_1) \approx 0.0001 = \alpha / N$ и $d = 6.114$ из условия $F(d) \approx 0.9999 = 1 - \alpha / N$ (в программную систему встроена возможность вычисления различных вероятностей для законов распределения). При анализе минимального наблюдения мы должны исключить его, если оно оказывается левее d_1 , а при анализе максимального элемента выборки - правее d . Последовательно используя эту процедуру, были исключены все 3 грубые “ошибки” измерений.

Пример 3.4.2. Отсеем аномальные наблюдения в выборке *примера 3.2.3*. Для отбраковки аномальных наблюдений зададимся уровнем значимости $\alpha = 0.1$ и из условия $F(d) \approx 0.9999 = 1 - \alpha / N$ при объёме выборки $N = 1000$ и векторе параметров $\theta^T = [1.9668, 1.0258, 0]$ распределения Вейбулла найдем $d = F^{-1}(0.9999) = 3.1721$. Далее, мы должны исключить те наблюдения, которые превышают величину 3.1721. Таких наблюдений оказалось 8.

На рис. 3.4.3 отражены результаты статистического анализа выборки после исключения из неё аномальных наблюдений. При проверке согласия использовано разбиение области определения случайной величины на асимптотически оптимальные интервалы.

3.5. Выводы

Методами статистического моделирования показано, что предварительное группирование исходной выборки и последующее вычисление ОМП по группированным данным приводит к робастным оценкам, устойчивым как к наличию в исходной выборке аномальных измерений, так и к отклонениям закона распределения выборки от предполагаемого.

На основании анализа функций влияния оценок по негруппированным и группированным выборкам показано, что за редким исключением ОМП по негруппированным наблюдениям являются *неробастными*. Из множества 26-и законов распределения, включенных в программную систему, исключение составляют ОМП параметров распределения Коши и ОМП параметра сдвига логистического закона, функции влияния которых ограничены. Напротив, ОМП по группированным данным и оптимальные оценки параметров сдвига и масштаба по выборочным квантилям для

больших выборок *устойчивы* как к аномальным ошибкам измерений, так и к отклонениям наблюдаемого закона от предполагаемого.

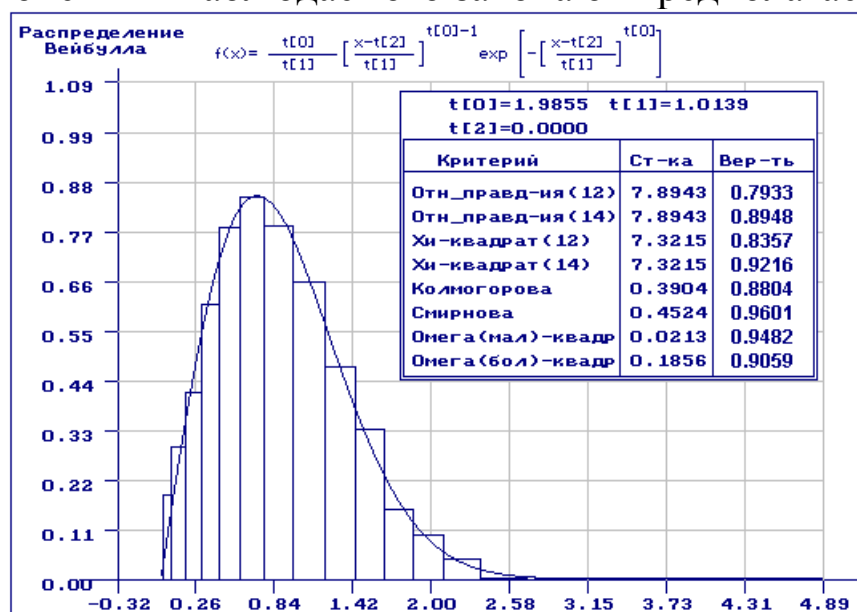


Рис. 3.4.3. Результаты анализа после удаления аномальных наблюдений (при проверке согласия использовано асимптотически оптимальное группирование).

Показано, что предварительное группирование исходной выборки и последующее вычисление ОМП по группированным данным приводит к робастным оценкам, устойчивым как к наличию в исходной выборке аномальных измерений, так и к отклонениям закона распределения выборки от предполагаемого. Процедура предварительного группирования реализована в программном обеспечении. Возможно использование равномерного, равновероятного и асимптотически оптимального группирования. На основании исходной негруппированной выборки может создаваться соответствующая группированная выборка. Реализован режим предварительного группирования при оценивании, в том числе при идентификации закона распределения.

Показано, что совокупность предварительного группирования данных для вычисления робастных оценок и применение асимптотически оптимального группирования в критериях согласия, используемые на этапе идентификации закона распределения по выборке, содержащей аномальные наблюдения, позволяют эффективно отбраковывать грубые ошибки измерений. Алгоритм параметрической отбраковки грубых ошибок измерений реализован в программной системе и хорошо зарекомендовал себя в сочетании с процедурой группирования наблюдений перед оцениванием.

4. ВОПРОСЫ ПРИМЕНЕНИЯ КРИТЕРИЕВ СОГЛАСИЯ

4.1. Проверка гипотез о согласии

Целью первичной обработки экспериментальных наблюдений обычно является выбор закона распределения, наиболее хорошо описывающего случайную величину, выборку которой мы наблюдали. Поэтому, после того как вычислены оценки параметров гипотетического распределения, необходимо проверить, насколько хорошо выборка согласуется с найденным законом. Такие проверки осуществляются с использованием различных критериев согласия.

Множество непрерывных законов распределений, используемых в задачах статистического анализа немногим превышает 100, а для описания наблюдаемых случайных величин в прикладных исследованиях в основном используется порядка 30 параметрических законов и семейств распределений. Естественно, что это не покрывает того многообразия случайных величин, которое встречается на практике. Правильное применение критериев согласия часто приводит и должно приводить к отклонению гипотез о принадлежности выборки удобному и привычному закону распределения, например нормальному. Законы реальных случайных величин, которые являются следствием большого числа причин, сложнее тех моделей, которые используются для их описания. Следовательно и модели должны быть более сложными. В частности, положительных результатов в приложениях можно добиться, используя в качестве моделей смеси различных законов распределений, в том числе усеченных. В то же время ошибки в использовании критериев согласия обычно приводят к необоснованному принятию (*чаще всего*) или отклонению проверяемой гипотезы.

Проверка гипотезы вида $H_0: f(x, \theta) = f(x, \hat{\theta})$, где $\hat{\theta}$ - оценка параметра распределения, осуществляется по следующей схеме. Для выбранного критерия вычисляется значение S^* статистики S как некоторой функции от выборки и закона распределения $f(x, \hat{\theta})$. Для используемых на практике критериев обычно известны асимптотические (предельные) распределения $g(s)$ соответствующих статистик при условии истинности гипотезы H_0 . Далее в принятой практике статистического анализа обычно сравнивают полученное значение статистики S^* с критическим S_α для данного уровня значимости α и

нулевую гипотезу отвергают, если $S^* > S_\alpha$ (см. рис. 4.1.1). Критическое значение S_α , определяемое из уравнения

$$\alpha = \int_{S_\alpha}^{\infty} g(s)ds,$$

обычно берётся из соответствующей статистической таблицы.

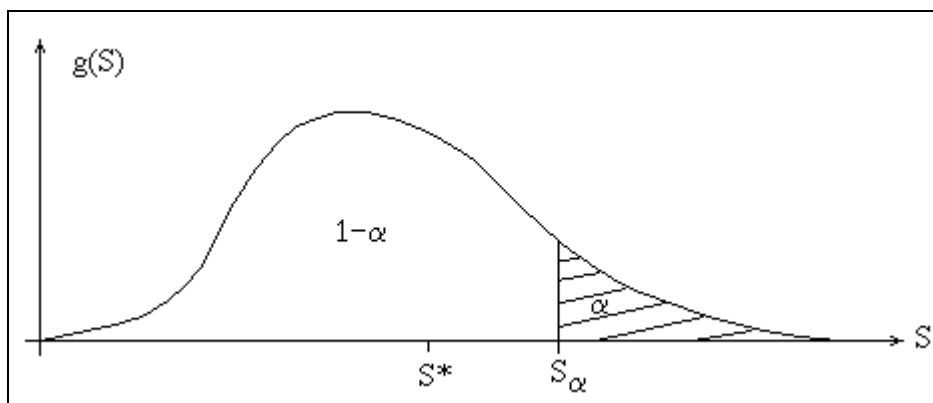


Рис. 4.1.1. Распределение статистики при истинной гипотезе H_0

Естественно, что больше информации о степени согласия можно почерпнуть из величины вероятности превышения полученного значения статистики при истинности нулевой гипотезы

$P\{S > S^*\} = \int_{S^*}^{\infty} g(s)ds$. Именно она позволяет судить о степени согласия,

так как по существу представляет собой вероятность истинности нулевой гипотезы (рис. 4.1.2). Гипотеза о согласии не отвергается, если $P\{S > S^*\} > \alpha$.

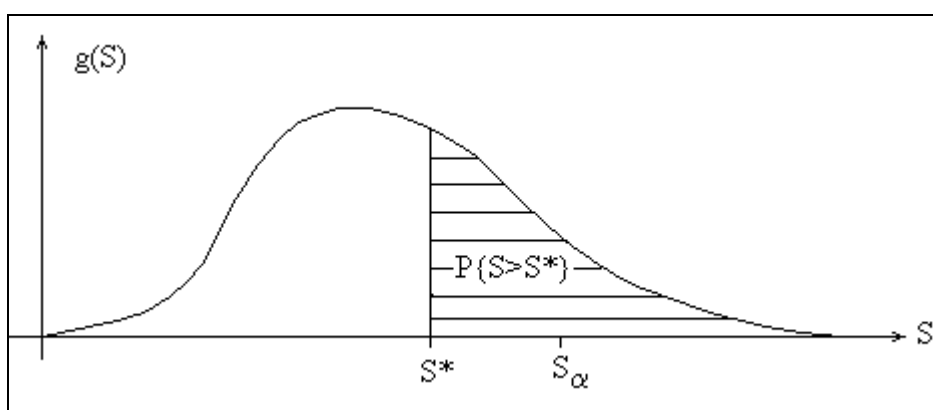


Рис. 4.1.2. Распределение статистики при истинной гипотезе H_0

Задачи оценивания параметров и проверки гипотез опираются на выборки независимых случайных величин. Случайность самой выборки предопределяет, что возможны и ошибки в результатах статистических выводов. С результатами проверки гипотез связывают

ошибки 2 видов: ошибка 1-го рода состоит в том, что отклоняется гипотеза H_0 , когда она верна; ошибка 2-го рода состоит в том, что принимается гипотеза H_0 , в то время как справедлива альтернативная гипотеза H_1 . Величина α задаёт вероятность ошибки 1-го рода. Если гипотеза H_1 определена, то задание α определяет и вероятность ошибки 2-го рода β для используемого критерия проверки гипотез. На рис. 4.1.3 $g(s|H_0)$ отображает плотность распределения статистики S при истинности гипотезы H_0 , а $g(s|H_1)$ - плотность распределения при справедливости гипотезы H_1 .

Мощность критерия представляет собой величину $1-\beta$. Понятно, что чем выше мощность используемого критерия при заданном значении α , тем лучше он различает гипотезы H_0 и H_1 . Особенно важно, чтобы используемый критерий хорошо различал близкие альтернативы. Графически требование максимальной мощности критерия означает, что на рис. 4.1.3 плотности $g(s|H_0)$ и $g(s|H_1)$ должны быть максимально "раздвинуты".

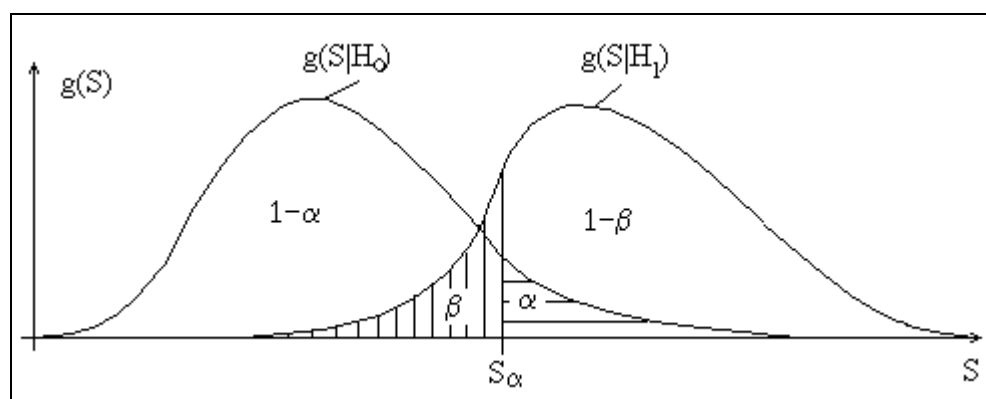


Рис. 4.1.3. Распределения статистик при справедливости гипотез H_0 и H_1

Рассматриваемые в данной главе вопросы применения критериев согласия так или иначе связаны со строгостью их использования: обеспечением максимальной мощности применяемого критерия или обоснованностью выбора плотности $g(s|H_0)$ при вычислении вероятности вида $P\{S > S^*\}$.

Для группы критериев типа χ^2 Пирсона, предусматривающих группирование наблюдений, исследуется зависимость свойств критерия от выбранного способа группировки. Ни для кого не секрет, что результаты проверки гипотез о согласии в соответствии с наиболее

часто используемым критерием χ^2 Пирсона существенно зависят от того, как сгруппированы данные: сгруппировали одним способом, гипотезу о согласии можно принять, сгруппировали другим - гипотеза о согласии должна быть отвергнута. Откуда же взять уверенность в таком выводе?

В [56] отмечается, что искать оптимальный метод определения границ интервалов для критерия χ^2 следует в терминах мощности критерия, т.е. выбирать такую совокупность границ, которая максимизировала бы мощность критерия данного размера. Правило использования интервалов равной вероятности предложено Манном и Вальдом в 1942 г. [242]. При таком разбиении максимизируется энтропия, равная $-\sum_{i=1}^k P_i(\theta) \ln P_i(\theta)$ [60]. В этом случае максимизируется асимптотическая мощность критерия, если не требуется рассматривать какую-либо конкретную альтернативу.

В нашем случае рассматривается использование в критериях согласия асимптотически оптимального группирования, минимизирующего потери в информации Фишера, вызываемые группировкой данных. Показывается, что мощность критерия при близких альтернативах тем выше, чем меньше потери в информации Фишера.

Для группы непараметрических критериев типа Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса методами статистического моделирования исследуются предельные законы распределений этих статистик при проверке сложных гипотез, когда по той же выборке оцениваются и параметры закона, с которым проверяется согласие. В такой ситуации эти критерии теряют свойство “свободы от распределения”.

В конце главы рассматриваются вопросы использования критериев согласия в случае группированных, частично группированных и интервальных наблюдений, а также вопросы принятия решения при идентификации закона распределения по совокупности используемых критериев согласия.

4.2. Асимптотически оптимальное группирование наблюдений в критериях согласия

Целый ряд критериев согласия или критериев проверки гипотез о принадлежности функции распределения независимых одинаково распределенных случайных величин $\xi_1, \dots, \xi_n$ семейству непрерывных функций $F(x, \theta)$, зависящих от параметра θ , в общем случае векторного, предусматривает группирование наблюдений. Область определения случайной величины разбивается на k непересекающихся интервалов граничными точками

$$x_{(0)} < x_{(1)} < \dots < x_{(k-1)} < x_{(k)},$$

где $x_{(0)}$ - нижняя грань области определения случайной величины, $x_{(k)}$ - верхняя грань. В результате получают количества наблюдений n_i , попавших в i -й интервал значений, и соответствующие вероятности попадания в интервал $P_i(\theta) = \int_{x_{(i-1)}}^{x_{(i)}} f(x, \theta) dx$, где $f(x, \theta)$ - функция плотности распределения, $\sum_{i=1}^k n_i = N$, где N - объем выборки, $\sum_{i=1}^k P_i(\theta) = 1$. Измерение отклонений n_i / N от $P_i(\theta)$ лежит в основе статистик, используемых в критериях согласия.

К критериям такого рода, в частности, относятся критерий χ^2 Пирсона, критерий отношения правдоподобия [56] и критерии типа χ^2 [114, 115, 107].

Очевидно, что группирование наблюдений приводит к потере информации, и эти потери зависят от выбора варианта группировки данных. На практике обычно строят интервалы равной длины или, в лучшем случае, интервалы равной вероятности. Потери информации о законе распределения в этих ситуациях различны и различна способность критериев распознавать близкие гипотезы.

Мерой внутренней близости распределений случайных величин служит фишеровская информация, и этот внутренний характер связан с мощностью различения между близкими значениями параметра. Так как в любой статистике не больше информации, чем в исходной выборке, то мощность различения с помощью статистики не больше, чем с помощью всей выборки. Следовательно, если нужно выбирать между несколькими статистиками, следует предпочесть ту, для которой потери фишеровской информации минимальны [138, с.299-300].

Статистика критерия согласия χ^2 Пирсона вычисляется в соответствии с соотношением

$$X_N^2 = N \sum_{i=1}^k \frac{(n_i / N - P_i(\theta))^2}{P_i(\theta)} \quad (4.2.1)$$

и в пределе подчиняется χ^2 -распределению с $(k-1)$ -й степенью свободы, если верна нулевая гипотеза. Эта же статистика подчиняется нецентральному χ^2 -распределению с тем же числом степеней свободы и параметром нецентральности

$$\lambda = N \sum_{i=1}^k \frac{(P_i(\theta_1) - P_i(\theta))^2}{P_i(\theta)} \quad \lambda = N \sum_{i=1}^k \frac{(P_i(\theta_1) - P_i(\theta))^2}{P_i(\theta)},$$

если верна конкурирующая гипотеза, и выборка соответствует распределению того же типа, но с параметром θ_1 .

Несложно показать, разлагая $P_i(\theta_1)$ в ряд Тейлора при малых $\delta\theta = \theta_1 - \theta$ и пренебрегая членами высшего порядка, что

$$\begin{aligned} \lambda &\approx N \sum_{i=1}^k \frac{[P_i(\theta) + \nabla^T P_i(\theta) \delta\theta - P_i(\theta)]^2}{P_i(\theta)} = N \sum_{i=1}^k \frac{\delta\theta^T \nabla P_i(\theta) \nabla^T P_i(\theta) \delta\theta}{P_i(\theta)} = \\ &= N \delta\theta^T \left(\sum_{i=1}^k \frac{\nabla P_i(\theta) \nabla^T P_i(\theta)}{P_i(\theta)} \right) \delta\theta = N \delta\theta^T M_\Gamma(\theta) \delta\theta, \end{aligned} \quad (4.2.2)$$

где $M_\Gamma(\theta) = \sum_{i=1}^k \frac{\nabla P_i(\theta) \nabla^T P_i(\theta)}{P_i(\theta)}$ - информационная матрица Фишера по

группированным данным. Мощность критерия χ^2 Пирсона представляет собой неубывающую функцию от λ . Матрица потерь информации, вызванных группированием, $\Delta M = M(\theta) - M_\Gamma(\theta)$, где $M(\theta)$ - информационная матрица Фишера по негруппированным наблюдениям, является неотрицательно определённой, и, следовательно, $\delta\theta^T \Delta M \delta\theta \geq 0$. А так как $\delta\theta^T M_\Gamma(\theta) \delta\theta = \delta\theta^T M(\theta) \delta\theta - \delta\theta^T \Delta M \delta\theta$, то очевидно, что с ростом потерь информации падает и мощность критерия при близких альтернативных гипотезах.

В критерии отношения правдоподобия используется статистика [56]

$$-2 \ln l = -2 \sum_{i=1}^k n_i \ln \left(\frac{P_i(\theta)}{n_i / N} \right),$$

которая при верной нулевой гипотезе асимптотически распределена как χ^2 с $(k-1)$ -й степенью свободы. Если верна конкурирующая гипотеза и выборка принадлежит распределению того же типа, но с

параметром θ_1 , мерой близости рассматриваемых распределений является величина

$$-2\ln l = 2N \sum_{i=1}^k P_i(\theta_1) \ln \left(\frac{P_i(\theta_1)}{P_i(\theta)} \right).$$

С ростом её увеличивается мощность критерия. Действуя как и в предыдущем случае и пренебрегая членами высшего порядка, будем иметь

$$-2\ln l \approx 2N \sum_{i=1}^k \left(P_i(\theta) + \nabla^T P_i(\theta) \delta\theta \right) \ln \left(1 - \frac{\nabla^T P_i(\theta) \delta\theta}{P_i(\theta)} \right).$$

Далее, раскладывая $\ln(1+x)$ по формуле Тейлора и вновь пренебрегая членами выше второго порядка, получаем

$$\begin{aligned} 2\ln l &\approx 2N \sum_{i=1}^k \left(P_i(\theta) + \delta\theta^T \nabla P_i(\theta) \right) \left[\frac{\nabla^T P_i(\theta) \delta\theta}{P_i(\theta)} - \frac{\delta\theta^T \nabla P_i(\theta) \nabla^T P_i(\theta) \delta\theta}{2P_i^2(\theta)} \right] \approx \\ &\approx 2N \sum_{i=1}^k \left(\delta\theta^T \nabla P_i(\theta) + \frac{\delta\theta^T \nabla P_i(\theta) \nabla^T P_i(\theta) \delta\theta}{2P_i(\theta)} \right) = \\ &= N\delta\theta^T \left[\sum_{i=1}^k \frac{\nabla P_i(\theta) \nabla^T P_i(\theta)}{P_i(\theta)} \right] \delta\theta = N\delta\theta^T M_\Gamma(\theta) \delta\theta \\ &= N\delta\theta^T \left[\sum_{i=1}^k \frac{\nabla P_i(\theta) \nabla^T P_i(\theta)}{P_i(\theta)} \right] \delta\theta = N\delta\theta^T M_\Gamma(\theta) \delta\theta. \end{aligned} \quad (4.2.3)$$

Это соотношение аналогично выражению (4.2.2).

Если оценки параметров находились в результате минимизации статистики X_N^2 , то X_N^2 асимптотически распределена как χ^2 с числом степеней свободы $k-m-1$, где m - количество оцененных по данной выборке параметров. Если же параметры оцениваются по негруппированным наблюдениям, например, методом максимального правдоподобия, то предельное распределение этой статистики лежит между χ_{k-1}^2 и χ_{k-m-1}^2 [19]. В этом случае, принимая нулевую гипотезу, мы должны удостовериться, что статистика X_N^2 не превышает критических значений $\chi_{k-m-1,\alpha}^2$ и $\chi_{k-1,\alpha}^2$, где α - задаваемый уровень значимости. И если $\chi_{k-m-1,\alpha}^2 < X_n^2 < \chi_{k-1,\alpha}^2$, то, принимая или отклоняя гипотезу о согласии, мы можем с одинаковым риском совершить ошибку.

В работах [114,115,107] рекомендуется видоизменение стандартной статистики X_N^2 , при котором предельное распределение есть обычное распределение χ^2 с $(k-1)$ -й степенью свободы и в том слу-

чае, когда в качестве оценок используются оценки максимального правдоподобия параметров по негруппированным данным. При этом вектор $p = (p_1, \dots, p_k)^T$ предполагается заданным и граничные точки интервалов определяются соотношениями $x_{(i)}(\theta) = F^{-1}(p_1 + \dots p_i)$, $i = \overline{1, (k-1)}$. Предложенная статистика имеет вид [115]

$$Y_N^2(\theta) = X_N^2 + N^{-1} a^T(\theta) \Lambda(\theta) a(\theta),$$

где X_N^2 определяется по (4.2.1), матрица $\Lambda(\theta) = \left\| I(\theta_l, \theta_j) - \sum_{i=1}^k \frac{w_{\theta li} w_{\theta ji}}{p_i} \right\|^{-1}$,

элементы и размерность которой определяются оцениваемыми компонентами вектора параметров θ ,

$I(\theta_l, \theta_j) = \int \left(\frac{\partial f(x, \theta)}{\partial \theta_l} \frac{\partial f(x, \theta)}{\partial \theta_j} \right) f(x, \theta) dx$ - элементы информационной

матрицы Фишера по негруппированным данным $M(\theta)$,

$a(\theta_l) = w_{\theta l1} n_1 / p_1 + \dots + w_{\theta lk} n_k / p_k$ - элементы вектора $a(\theta)$,

$w_{\theta li} = -f[x_{(i)}(\theta), \theta] \frac{\partial x_{(i)}(\theta)}{\partial \theta_l} + f[x_{(i-1)}(\theta), \theta] \frac{\partial x_{(i-1)}(\theta)}{\partial \theta_l}$. В данном случае

справедливо соотношение

$$\left\| \sum_{i=1}^k \frac{w_{\theta li} w_{\theta ji}}{p_i} \right\| = \sum_{i=1}^k \frac{\nabla^T P_i(\theta) \nabla P_i(\theta)}{P_i(\theta)} = M_\Gamma(\theta)$$

и, следовательно,

$$\Lambda(\theta) = [M(\theta) - M_\Gamma(\theta)]^{-1}. \quad (4.2.4)$$

Следовательно, как и в соотношениях (4.2.2-4.2.3), в данном случае параметр нецентральности предельного нецентрального χ_{k-1}^2 распределения (см. [115]) зависит от величины потерь информации при группировании.

Аналогично, в случае использования в критерии согласия модифицированной статистики χ^2 мерой близости рассматриваемых распределений является величина

$$\text{mod } \chi^2 = N \sum_{i=1}^k \frac{(P_i(\theta_1) - P_i(\theta))^2}{P_i(\theta_1)}.$$

Раскладывая, как и ранее, в ряд Тейлора и пренебрегая членами при $\delta\theta$ выше второго порядка, получаем

$$\text{mod } \chi^2 \approx N \sum_{i=1}^k \frac{[P_i(\theta) + \delta\theta^T \nabla P_i(\theta) - P_i(\theta)]^2}{P_i(\theta) + \delta\theta^T \nabla P_i(\theta)} =$$

$$= N \sum_{i=1}^k \frac{\delta \theta^T \nabla P_i(\theta) \nabla^T P_i(\theta) \delta \theta}{P_i(\theta) [1 + \delta \theta^T \nabla P_i(\theta) / P_i(\theta)]}.$$

Используя разложение в ряд $(1+x)^{-1}$, имеем

$$\text{mod } \chi^2 \approx N \delta \theta^T \left[\sum_{i=1}^k \frac{\nabla P_i(\theta) \nabla^T P_i(\theta)}{P_i(\theta)} \right] \delta \theta = N \delta \theta^T M_\Gamma(\theta) \delta \theta. \quad (4.2.5)$$

Для меры Хеллингера, действуя аналогично,

$$\begin{aligned} H_D &= \arccos \sum_{i=1}^k \sqrt{P_i(\theta_1) P_i(\theta)} \approx \\ &\approx \arccos \sum_{i=1}^k \sqrt{(P_i(\theta) + \delta \theta^T \nabla P_i(\theta)) P_i(\theta)} = \\ &= \arccos \sum_{i=1}^k P_i(\theta) \sqrt{1 + \delta \theta^T \nabla P_i(\theta) / P_i(\theta)}. \end{aligned}$$

Используя разложение в ряд для $\sqrt{1+x}$, получаем

$$\begin{aligned} H_D &\approx \arccos \sum_{i=1}^k P_i(\theta) \left[1 + \frac{1}{2} \frac{\delta \theta^T \nabla P_i(\theta)}{P_i(\theta)} - \frac{1}{8} \frac{\delta \theta^T \nabla P_i(\theta) \nabla^T P_i(\theta) \delta \theta}{P_i(\theta)} \right] = \\ &= \arccos \left[1 - \frac{1}{8} \delta \theta^T M_\Gamma(\theta) \delta \theta \right]. \end{aligned} \quad (4.2.6)$$

Для дивергенции Кульбака-Лейблера

$$H_{KL} = \sum_{i=1}^k P_i(\theta) \ln \left(\frac{P_i(\theta)}{P_i(\theta_1)} \right).$$

Используя разложение в ряд для $\ln(1+x)$, получаем

$$H_{KL} \approx \sum_{i=1}^k P_i(\theta) \left[\frac{\nabla^T P_i(\theta) \delta \theta}{P_i(\theta)} - \frac{1}{2} \frac{\delta \theta^T \nabla P_i(\theta) \nabla^T P_i(\theta) \delta \theta}{P_i^2(\theta)} \right] = \frac{1}{2} \delta \theta^T M_\Gamma(\theta) \delta \theta. \quad (4.2.7)$$

Таким образом, на основании соотношений (4.2.2-4.2.7) можно утверждать, что чем меньше потери информации, связанные с группированием наблюдений, тем выше мощность соответствующих критериев согласия при близких конкурирующих гипотезах. Потери от группирования можно уменьшить, подбирая граничные точки так, чтобы $M_\Gamma(\theta)$ стремилась к информационной матрице по негруппированным данным $M(\theta)$, т.е., как и при оценивании параметров, решая задачу асимптотически оптимального группирования. В случае скалярного параметра эта задача сводится к максимизации информационного количества Фишера о параметре по группированной выборке

$$\max_{x_{(0)} < x_{(1)} < \dots < x_{(k-1)} < x_{(k)}} \sum_{i=1}^k \left(\frac{\partial \ln P_i(\theta)}{\partial \theta} \right)^2 P_i(\theta).$$

А в случае вектора параметров в качестве критериев оптимальности могут быть выбраны различные функционалы от информационной матрицы Фишера. Наиболее естественно максимизировать определитель информационной матрицы, т.е. решать задачу

$$\max_{x_{(0)} < x_{(1)} < \dots < x_{(k-1)} < x_{(k)}} \det M_{\Gamma}(\theta).$$

К сожалению на практике наиболее часто, применяя критерий χ^2 Пирсона, используют интервалы равной длины или, в лучшем случае, интервалы равной вероятности. Естественно, что в такой ситуации мощность критерия обычно далека от максимально возможной.

Для иллюстрации на рис. 4.2.1 приведены построенные функции мощности критерия χ^2 при проверке согласия с экспоненциальным распределением для числа интервалов $k=3, 10$, объеме выборки $N=1000$, при уровне значимости $\alpha=0.05$ в случае асимптотически оптимального группирования и разбиения на интервалы равной вероятности. При построении кривых были использованы таблицы мощности критерия χ^2 из [19]. Графики наглядно иллюстрируют предпочтительность асимптотически оптимального группирования.

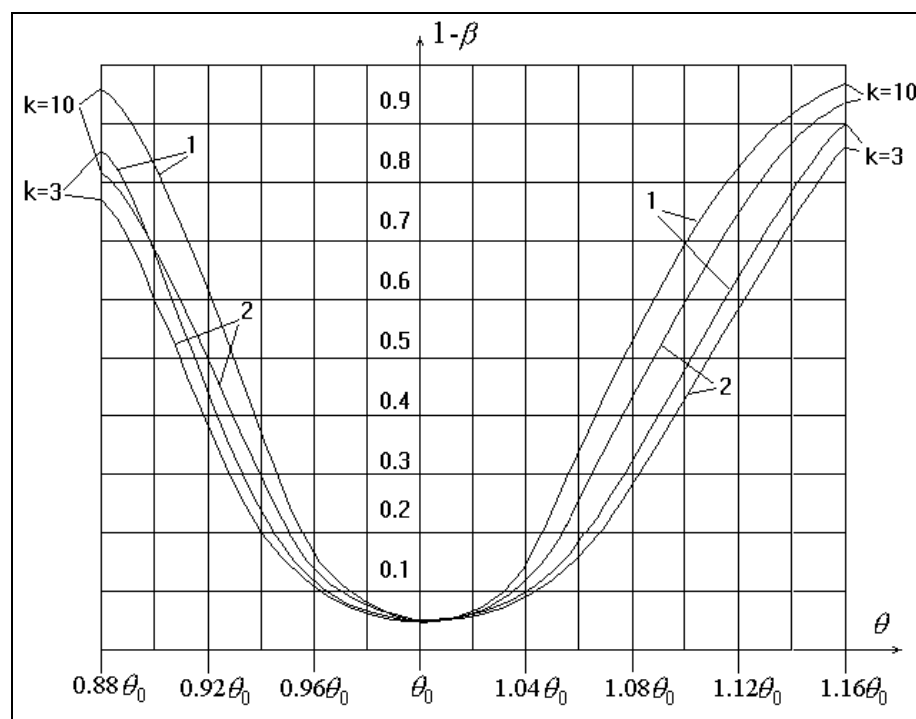


Рис. 4.2.1. Функция мощности критерия χ^2 для уровня значимости $\alpha=0.05$, объема выборки $n=1000$, числа интервалов $k=3, 10$:
1 - для оптимального группирования; 2 - для равновероятного группирования.

На рис. 4.2.2 в аналогичной ситуации приведены функции мощности при проверке гипотезы об основном параметре (формы) распределения Вейбулла. Как видим, в обоих случаях мощность критерия при асимптотически оптимальном группировании существенно выше. Выбор в качестве объекта сравнения случая равновероятного группирования обоснован определённой этой процедуры разбиения и ее оптимальностью при отсутствии конкретных альтернатив [60].

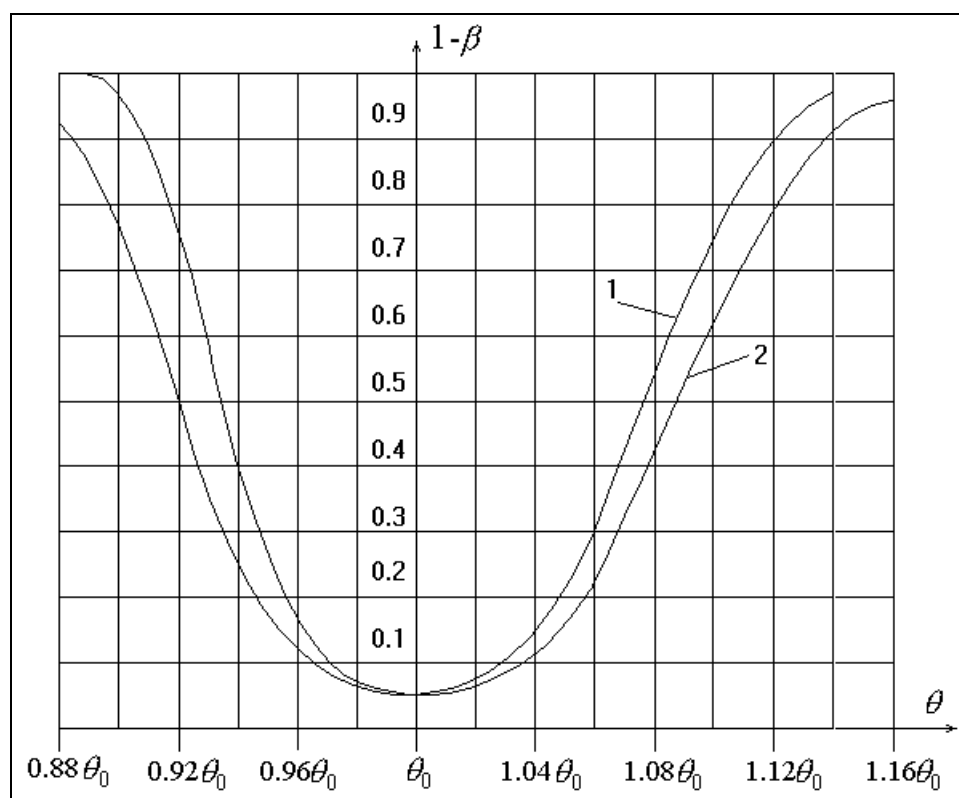


Рис. 4.2.2. Функция мощности критерия χ^2 Пирсона при проверке гипотез об основном (формы) параметре распределения Вейбулла: уровень значимости $\alpha = 0.05$, объем выборки $N = 1000$, количество интервалов $k = 10$;
1 - для оптимального группирования;
2 - для равновероятного группирования.

В общем случае информационная матрица Фишера зависит не только от граничных точек $x_{(i)}$, но и от параметров исследуемого распределения. Однако, как уже говорилось, для достаточно широкого ряда распределений при решении задач асимптотически оптимального группирования удалось получить граничные точки интервалов в виде, инвариантном относительно параметров распределений, и на их основе сформировать таблицы асимптотически оптимального группирования.

В связи с проблемой оценивания параметров по группированным выборкам задача асимптотически оптимального группирования данных рассматривалась в [205,65], неоднократно она возникала при использовании для оценивания параметров выборочных квантилей. Наиболее полная совокупность таблиц асимптотически оптимального группирования для распределений экспоненциального, полунормального, Рэлея, Максвелла, модуля многомерного нормального вектора, Парето, Эрланга, Лапласа, нормального, логарифмически-нормальных ($\ln$ и $\lg$), Коши, Вейбулла, распределений минимального и максимального значения, двойного показательного, гамма-распределения представлена в [42]. В общей сложности получено 54 таблицы оптимальных граничных точек и соответствующих вероятностей. *Эти таблицы могут использоваться как при проверке гипотез, так и при оценивании.* Полученные таблицы используются в программной системе [76] при проверке согласия по критериям χ^2 Пирсона и отношения правдоподобия и при вычислении робастных оценок. Вопросы использования асимптотически оптимального группирования рассматривались в работах [40,104,207,85,52].

В таблице приложения П1.28 представлены асимптотически оптимальные граничные точки интервалов в виде $t_i = (x_i - \theta_0) / \theta_1$ при проверке гипотез о согласии с нормальным распределением с плотностью

$$f(x) = \frac{1}{\theta_1 \sqrt{2\pi}} \exp \left\{ -\frac{(x - \theta_0)^2}{2\theta_1^2} \right\}.$$

Они были получены максимизацией определителя информационной матрицы Фишера по группированным наблюдениям. При проверке гипотезы о согласии с нормальным распределением с параметрами θ_0 и θ_1 конкретные значения x_i определяются соотношением $x_i = t_i \theta_1 + \theta_0$. Вероятности попадания наблюдений в интервалы, соответствующие оптимальному группированию, представлены в табл. П1.29. Содержащиеся в табл. П1.28 граничные точки интервалов рекомендуется использовать в случаях, когда по выборке или не оценивались параметры нормального распределения, или оценивались сразу оба. Если же по наблюдаемой выборке оценивался только один из параметров, то должно быть использовано группирование, максимизирующее количество информации по группированным данным для оцениваемого

параметра. Соответствующие таблицы приведены в [42] и приложении П1.

Эти же таблицы могут использоваться для логарифмически нормального распределения с плотностью

$$f(x) = \frac{1}{x\theta_1\sqrt{2\pi}} e^{-(\ln x - \theta_0)^2 / 2\theta_1^2}.$$

В этом случае граничные точки интервалов будут определяться соотношением $x_i = \exp\{t_i\theta_1 + \theta_0\}$. Для логарифмически нормального закона с плотностью

$$f(x) = \frac{1}{x\theta_1 \ln 10 \sqrt{2\pi}} e^{-(\lg x - \theta_0)^2 / 2\theta_1^2}$$

граничные точки вычисляются в соответствии с выражением $x_i = 10^{t_i\theta_1 + \theta_0}$. В последних колонках приведенных таблиц содержатся значения относительной асимптотической информации $A = \det M_\Gamma(\theta) / \det M(\theta)$, позволяющей судить о качестве группирования.

В табл. П1.22 представлены асимптотически оптимальные граничные точки интервалов в виде $t_i = ((x_i - \theta_2)/\theta_1)^{\theta_0}$ для проверки согласия с распределением Вейбулла с плотностью

$$f(x) = \frac{\theta_0(x - \theta_2)^{\theta_0-1}}{\theta_1^{\theta_0}} \exp\left\{-\left(\frac{x - \theta_2}{\theta_1}\right)^{\theta_0}\right\}.$$

Соответствующие значения вероятностей представлены в табл. П1.23. Эти таблицы рекомендуется использовать в случаях, когда по выборке не оценивались параметры распределения, или одновременно оценивались параметры θ_0 и θ_1 . Конкретные значения граничных точек определяются соотношением $x_i = t_i^{1/\theta_0}\theta_1 + \theta_2$. Если по имеющейся выборке оценивался только один из указанных параметров, то рекомендуется использовать таблицы, приведенные в [42], в которых при группировании максимизировалось количество информации по группированным данным для соответствующего параметра.

Аналогичным образом табл. П1.22-П1.23 могут использоваться в критериях проверки согласия с распределением наименьшего экстремального значения с плотностью

$$f(x) = \frac{1}{\theta_1} \exp \left\{ \frac{x - \theta_0}{\theta_1} - \exp \left(\frac{x - \theta_0}{\theta_1} \right) \right\}.$$

В этом случае конкретные значения граничных точек будут определяться выражением $x_i = \theta_1 \ln t_i + \theta_0$.

Для многих законов распределений граничные точки интервалов не могут быть выражены в виде, инвариантном относительно параметров распределений, т.е. они остаются функциями этих параметров. Это касается, например, таких законов, как гамма- и бета-распределения [31,72], экспоненциального семейства распределений. В этом случае формирование таблиц асимптотически оптимального группирования теряет смысл. Однако возможно решение задачи асимптотически оптимального группирования при конкретных значениях параметров в процессе проверки гипотез о согласии, как это реализуется в таких ситуациях в программной системе [76].

Продемонстрируем, как практически отражается применение асимптотически оптимального группирования на результатах статистического анализа при различных отклонениях выборки от предположений.

Пример 1. На рис. 4.2.3 приведены результаты моделирования выборки объемом $N = 1000$ в соответствии с логистическим распределением с функцией плотности

$$f(x) = \frac{\pi}{\theta_1 \sqrt{3}} \exp \left\{ -\frac{\pi(x - \theta_0)}{\theta_1 \sqrt{3}} \right\} / \left[1 + \exp \left\{ -\frac{\pi(x - \theta_0)}{\theta_1 \sqrt{3}} \right\} \right]^2,$$

и параметрами $\theta_0 = 0, \theta_1 = 1$. При проверке согласия использовано асимптотически оптимальное группирование при числе интервалов $k = 9$. На этом и последующих рисунках для параметров распределений использованы обозначения $t[0] = \theta_0$, $t[1] = \theta_1$, $t[2] = \theta_2$. На рисунках отражаются результаты проверки гипотез о согласии: вычисленные значения S^* соответствующих статистик S и вероятности превышения полученного значения статистики при истинности нулевой гипотезы $P\{S > S^*\}$. Гипотеза о согласии не отвергается, если $P\{S > S^*\} > \alpha$. Проверка согласия осуществляется по критериям отношения правдоподобия, χ^2 Пирсона, применение которых предусматривает группирование наблюдений, а также по непараметрическим критериям Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса [19].

Статистику критерия ω^2 называют статистикой Смирнова-Мизеса, статистику Ω^2 - Андерсона-Дарлинга.

Посмотрим, что получится, если мы будем проверять согласие полученной выборки с нормальным распределением с параметрами $\theta_0 = 0$, $\theta_1 = 1$. На рис. 4.2.4 представлены результаты проверки согласия при использовании в критериях отношения правдоподобия и χ^2 Пирсона равновероятного группирования. Как видим, по всем критериям, в том числе непараметрическим, при уровне значимости $\alpha \leq 0.1327$ нет оснований для отклонения гипотезы о согласии.

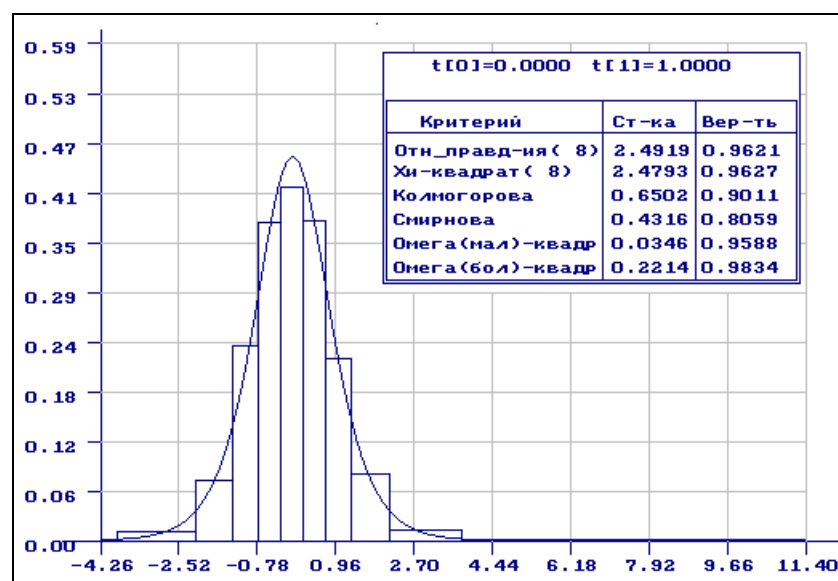


Рис. 4.2.3. Результаты моделирования логистического распределения

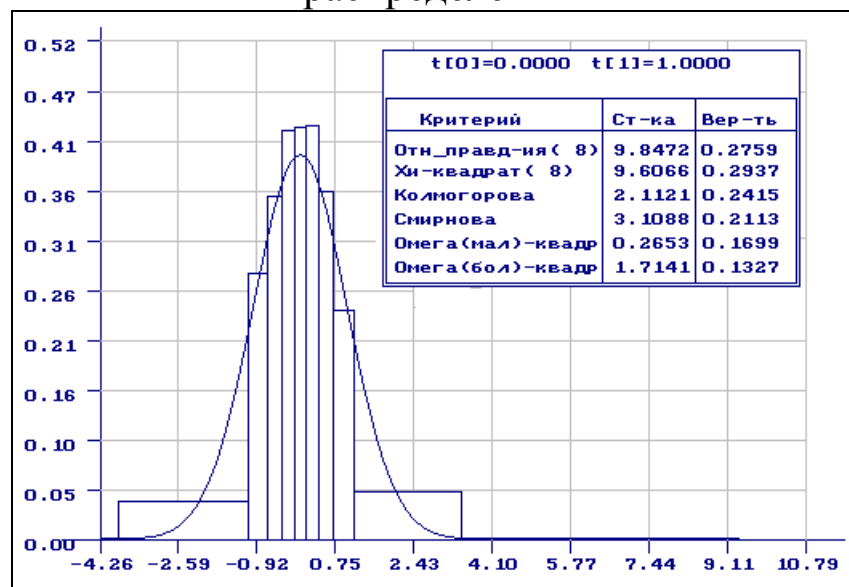


Рис.4.2.4. Результаты проверки согласия выборки, распределенной по логистическому закону, с нормальным распределением (равновероятное группирование)

Результаты анализа, приведенные на рис. 4.2.5 отличаются тем, что в критериях отношения правдоподобия и χ^2 Пирсона использовалось асимптотически оптимальное группирование (табл. П1.28-П1.29, $k = 9$). В этом случае гипотеза о согласии при $\alpha > 0.0861$ должна быть отклонена. В данном примере логистическое и нормальное распределения имеют отличие в середине области определения.

Пример 2. На рис. 4.2.6 приведены результаты моделирования выборки объемом $N = 200$ в соответствии с распределением Лапласа с плотностью

$$f(x) = \frac{1}{2\theta_1} \exp\left(-\frac{(x - \theta_0)}{\theta_1}\right)$$

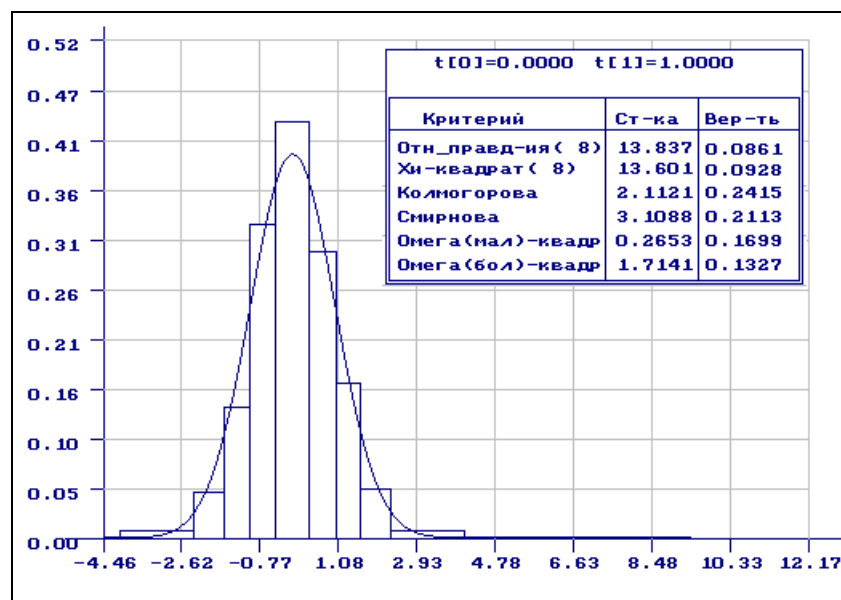


Рис. 4.2.5. Результаты проверки согласия выборки, распределенной по логистическому закону, с нормальным распределением (асимптотически оптимальное группирование)

и параметрами $\theta_0 = 1, \theta_1 = 1$. Было высказано предположение, что наблюдаемая выборка принадлежит нормальному закону с параметрами $\theta_0 = 1, \theta_1 = 1.2$. Результаты проверки согласия с нормальным распределением представлены на рис. 4.2.7-4.2.8. Результатам анализа на рис. 4.2.7 соответствовало равновероятное группирование в критериях согласия при числе интервалов $k = 7$, а на рис. 4.2.8 - асимптотически оптимальное группирование (табл. П1.28-П1.29, $k = 7$). Как видим, в первом случае гипотеза о согласии будет принята, если уровень значимости $\alpha \leq 0.1845$, во втором - по критериям

отношения правдоподобия и χ^2 Пирсона гипотеза о согласии должна быть отклонена. В отличие от предыдущего примера рассматриваемые здесь распределения отличаются на “хвостах”.

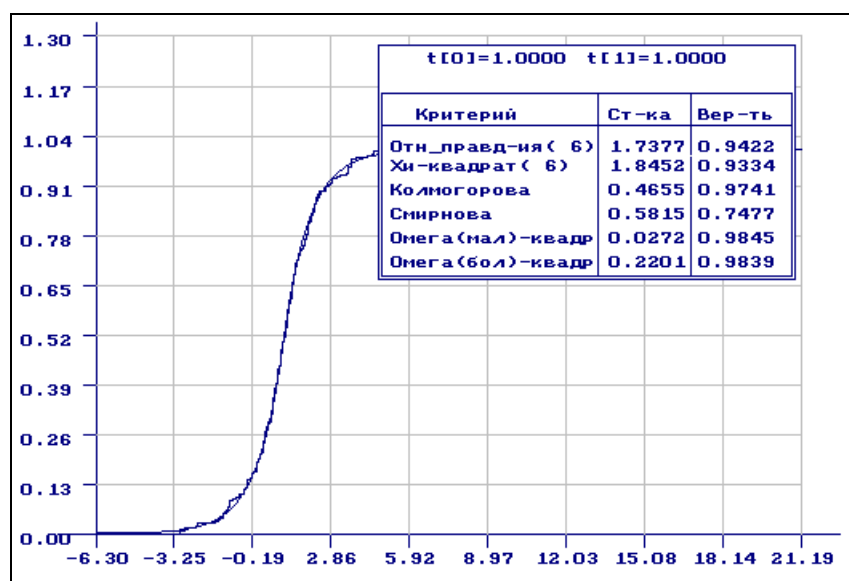


Рис. 4.2.6. Результаты моделирования распределения Лапласа: эмпирическая функция и функция распределения Лапласа

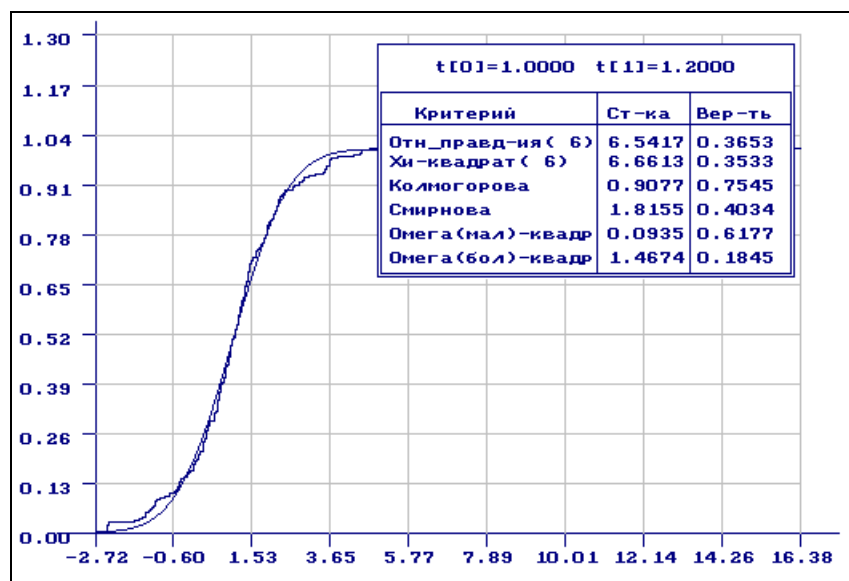


Рис. 4.2.7. Результаты проверки согласия выборки, распределенной по закону Лапласа, с нормальным распределением: эмпирическая функция и функция распределения нормального закона (равновероятное группирование)

Пример 3. Выборка объемом 200 наблюдений, смоделированная в соответствии с распределением Вейбулла с параметрами $\theta_0 = 2$, $\theta_1 = 2$, $\theta_2 = 0$, была “засорена” 10 наблюдениями нормального закона с параметрами $\theta_0 = 4$, $\theta_1 = 1$. При анализе в данном примере осущест-

влялось группирование для числа интервалов $k = 7$. На рис. 4.2.9 отражены результаты проверки согласия “смеси” с исходным распределением Вейбулла при использовании равновероятного группирования, а на рис. 4.2.10 - асимптотически оптимального (табл. П1.22-П1.23, $k = 7$). На рис. 4.2.11 приведены функция распределения Вейбулла и эмпирическая функция “смеси”. Как видим, и в данном случае при асимптотически оптимальном группировании критерии отношения правдоподобия и χ^2 Пирсона улавливают наличие отклонений от предположений.

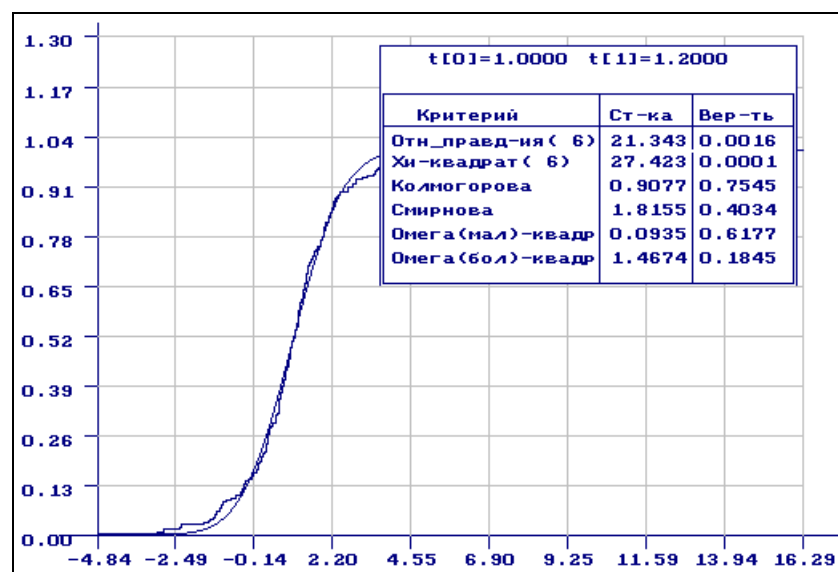


Рис. 4.2.8. Результаты проверки согласия выборки, распределенной по закону Лапласа, с нормальным распределением: эмпирическая функция и функция распределения нормального закона (асимптотически оптимальное группирование)

Выводы

1. Показана связь мощности критериев согласия с информационной матрицей Фишера по группированным данным. Показано, что применение асимптотически оптимального группирования данных в критериях согласия отношения правдоподобия, χ^2 Пирсона и типа χ^2 обеспечивает максимальную мощность этих критериев при близких конкурирующих гипотезах. Следовательно, снижается риск принятия нулевой гипотезы, когда на самом деле выборка принадлежит некоторому другому закону.
2. Одни и те же таблицы асимптотически оптимального группирования могут использоваться как в задачах проверки гипотез о согла-

сии, так и при получении различных оценок параметров, использующих выборочные квантили.

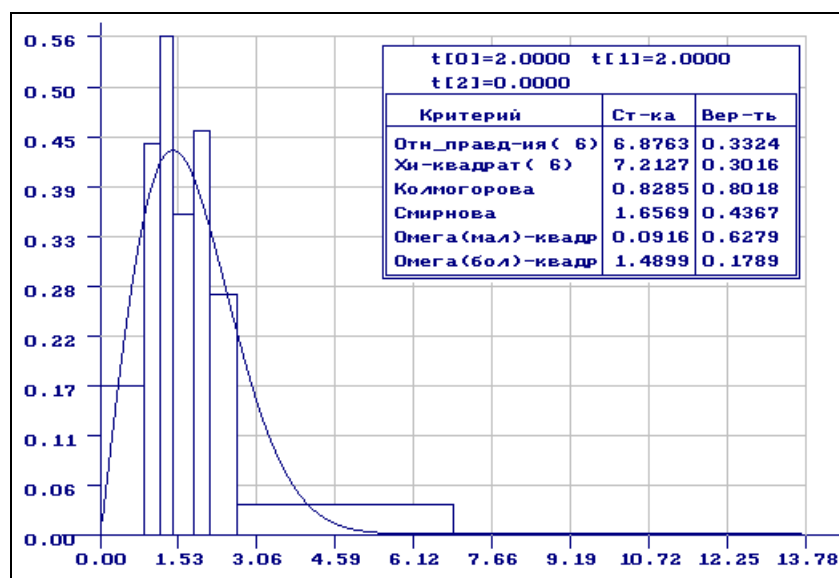


Рис. 4.2.9. Результаты проверки смеси с распределением Вейбулла с параметрами $\theta_0 = 2, \theta_1 = 2, \theta_2 = 0$ (равновероятное группирование)

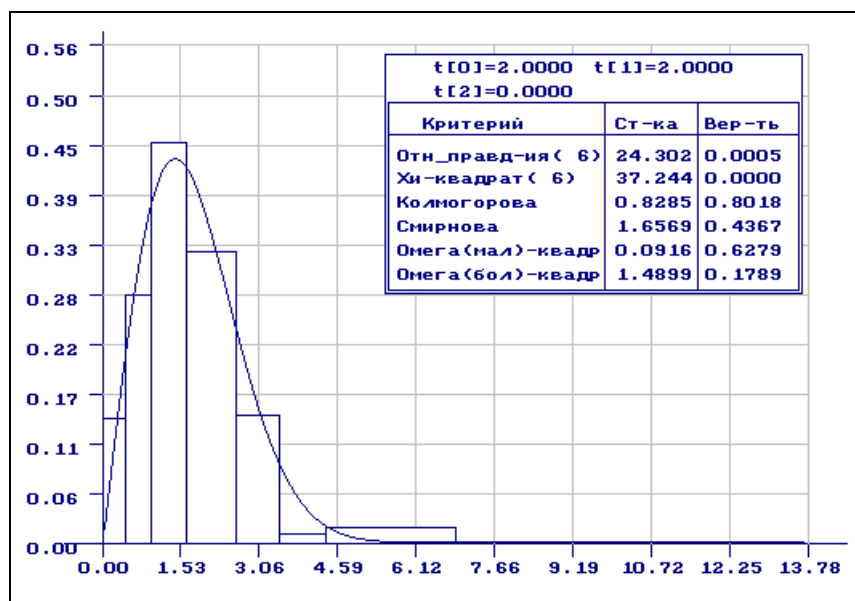


Рис. 4.2.10. Результаты проверки смеси с распределением Вейбулла с параметрами $\theta_0 = 2, \theta_1 = 2, \theta_2 = 0$ (асимптотически оптимальное группирование)

- В реализованной программной системе в критериях отношения правдоподобия и χ^2 Пирсона при проверке гипотез о согласии используются или построенные таблицы асимптотически оптимального группирования, или задача асимптотически оптимального группирования решается непосредственно перед проверкой согласия.

4. Критерии отношения правдоподобия и χ^2 Пирсона при использовании асимптотически оптимального группирования лучше, чем непараметрические критерии Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса улавливают малые отклонения от предположений в наблюдаемых данных.

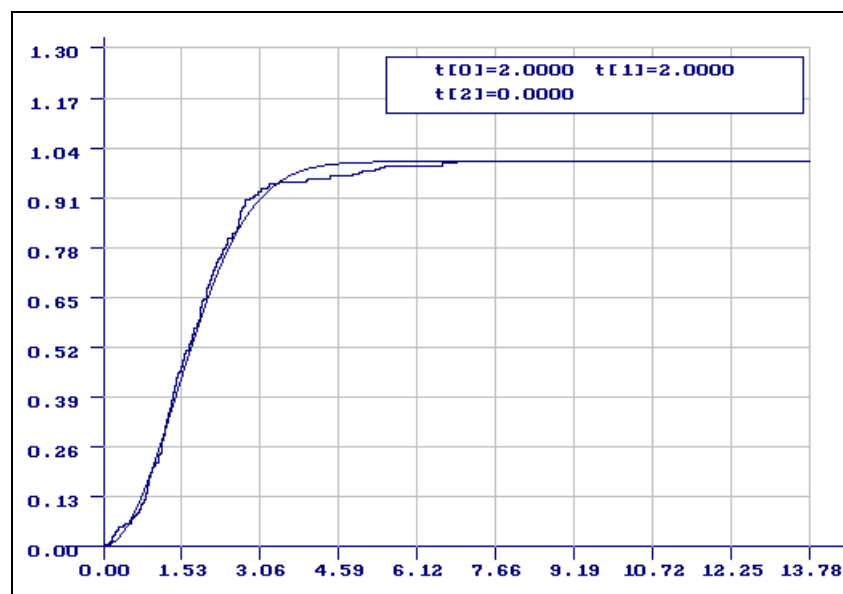


Рис. 4.2.11. Эмпирическая функция распределения “смеси” и теоретическая функция распределения Вейбулла с параметрами $\theta_0 = 2, \theta_1 = 2, \theta_2 = 0$

5. Выбор числа интервалов. По этому поводу в свое время проводилось немало исследований, существует достаточно много противоречивых рекомендаций, часть из которых приведена в [116]. При асимптотически оптимальном группировании относительно скалярного параметра при 10-11 интервалах в группированной выборке сохраняется около 98% информации, при оптимальном группировании относительно вектора параметров (два параметра) для 15 интервалов - около 95%. Дальнейшее увеличение количества интервалов существенного значения не имеет. Выбор конкретного числа интервалов должен осуществляться из следующих соображений. При оптимальном группировании вероятности попадания в интервалы в общем случае не равны. Обычно минимальны вероятности попадания в крайние интервалы. Поэтому k желательно выбирать из условия $NP_i(\theta) \geq 5 \div 10$ для любого интервала при оптимальном группировании. По крайней мере минимальная ожидаемая частота должна быть больше 1.

4.3. О зависимости предельных распределений статистик χ^2 Пирсона и отношения правдоподобия от способа группирования

В критериях согласия χ^2 Пирсона и отношения правдоподобия при вычислении соответствующих им статистик осуществляют группирование наблюдений. Неоспоримо, что вычисленные по конкретной выборке значения рассматриваемых статистик очень сильно зависят от того, как сгруппированы данные: выбрали интервалы группирования одним способом - нулевая гипотеза H_0 о согласии должна быть отвергнута, другим - нет оснований её отвергать. Очевидно, что *предельные распределения* статистик критериев χ^2 Пирсона и отношения правдоподобия *зависят* не только от числа оцененных параметров, но и *от способа группирования*, вида исследуемого закона распределения $f(x, \theta)$, метода оценивания и типа оцениваемого параметра.

Статистика χ^2 Пирсона вычисляется в соответствии с соотношением

$$S_{\chi^2} = N \sum_{i=1}^k \frac{(n_i / N - P_i(\theta))^2}{P_i(\theta)},$$

где n_i - количество наблюдений, попавших в интервал, $P_i(\theta)$ - вероятность попадания наблюдения в i -й интервал. При справедливой гипотезе H_0 ее предельное распределение $g(S(H_0))$ есть χ_r^2 -распределение с числом степеней свободы $r = k - 1$. Если по выборке оценивалось p параметров закона, статистика подчиняется χ_r^2 -распределению с $r = k - p - 1$ степеней свободы. При справедливой альтернативной гипотезе H_1 предельное распределение $g(S(H_1))$ представляет собой нецентральное χ_r^2 -распределение с тем же числом степеней свободы и параметром нецентральности

$$\lambda = N \sum_{i=1}^k \frac{(P_i(\theta_1) - P_i(\theta))^2}{P_i(\theta)}.$$

В случае проверки сложных гипотез и оценивании по выборке параметров распределений использование в качестве предельных χ_{k-p-1}^2 -распределений справедливо лишь при определении оценок параметров по сгруппированным данным [56]. При вычислении оценок максимального правдоподобия по негруппированным данным предельное распределение статистики лежит между χ_{k-1}^2 и χ_{k-p-1}^2 . При за-

данном уровне значимости α гипотеза о согласии не должна отвергаться, если

$$P\{S_{\chi^2} > S_{\chi^2}^*\} = \frac{1}{2^{r/2} \Gamma(r/2)} \int_{S_{\chi^2}^*}^{\infty} s^{r/2-1} e^{-s/2} ds > \alpha.$$

В критерии отношения правдоподобия используется статистика [56]

$$S_{\text{оп}} = -2 \ln l = -2 \sum_{i=1}^k n_i \ln \left(\frac{P_i(\theta)}{n_i / N} \right).$$

При верной гипотезе H_0 эта статистика асимптотически распределена так же, как и статистика χ^2 Пирсона.

χ_r^2 -распределения используются в качестве предельных для статистик χ^2 Пирсона и отношения правдоподобия независимо от того, каким образом сгруппированы данные в интервалы.

При практическом использовании критериев согласия выбирают либо интервалы равной длины, либо интервалы равной вероятности (равной частоты), либо асимптотически оптимальные интервалы [42]. Использование интервалов равной вероятности было предложено ещё в работе [244]. Разбиение области определения случайной величины (размаха выборки) на интервалы равной длины неоднозначно. Более определенными способами являются равновероятное и асимптотически оптимальное группирование. При асимптотически оптимальном группировании мощность этих критериев для близких альтернатив максимальна [42]. Опыт применения данных критериев согласия показывает, что при использовании равновероятного группирования и малых отклонениях выборки от предположений значения статистик существенно меньше, а вероятности “согласия” вида $P\{S > S^*\}$, где S^* - значение статистики, вычисленное по выборке, значительно выше, чем при асимптотически оптимальном. При асимптотически оптимальном группировании лучше улавливаются небольшие отклонения выборки от предположений.

Очевидно, что выбор равновероятного или асимптотически оптимального группирования отражается на предельных распределениях статистик $g(S(H_1))$ при верной альтернативной гипотезе. При использовании асимптотически оптимального группирования распределения $g(S(H_0))$ и $g(S(H_1))$ максимально “сдвинуты” относительно друг друга [42]. Надо полагать, что распределение статистики $g(S(H_0))$ при справедливой нулевой гипотезе также зависит от способа группирования.

Если значения одной и той же статистики так сильно отличаются при различных способах группирования, то можем ли мы при практическом использовании критерия согласия использовать один и тот же предельный закон? Насколько сильно законы распределения статистик отличаются от соответствующего χ_r^2 -распределения, особенно при определении оценок по негруппированным данным? Если действительная функция распределения статистики проходит выше используемого предельного теоретического χ_r^2 , при проверке согласия увеличивается вероятность ошибок второго рода (принятие гипотезы H_0 при справедливой H_1), ниже - вероятность ошибок первого рода (отклонения верной гипотезы H_0).

Целью исследования в данном случае явилось желание выяснить, насколько сильно отличаются предельные распределения рассматриваемых статистик при равновероятном и асимптотически оптимальном группировании (при справедливости гипотезы H_0) в зависимости от того, с каким законом проверяется согласие, сколько параметров оценивалось по выборке, каким образом они оценивались, а также предложить соответствующие рекомендации для практики статистического анализа.

Для этого моделировались и исследовались эмпирические законы распределений статистик критериев χ^2 Пирсона и отношения правдоподобия, вычисляемых при равновероятном и асимптотически оптимальном группировании. По каждому закону распределения $f(x, \theta)$ моделировалась серия из $n = 2000$ выборок объемом $N = 200$. По каждой выборке оценивались параметры этого закона распределения (один параметр, или другой, или два одновременно). Затем вычислялись значения статистик критериев χ^2 Пирсона и отношения правдоподобия при равновероятном и асимптотически оптимальном группировании. По накопленной выборке значений статистики объема n анализировалось полученное эмпирическое распределение статистики и осуществлялась идентификация закона. Оценки параметров находились по методу максимального правдоподобия по негруппированным данным.

В соответствии с изложенным алгоритмом моделировались выборки значений статистик S_{χ^2} , $S_{\text{оп}}$. Ниже приводятся результаты моделирования и анализа, которые вместе с результатами, изложенными в [96], позволяют сделать определенные выводы. На приводимых рисунках с результатами моделирования эмпирических распределений статистик “1” отмечено распределение статистики при асим-

птотически оптимальном группировании, а “2” - при равновероятном группировании. На рис. 4.3.1-4.3.3 представлены распределения статистики S_{χ^2} при 5 интервалах группирования и оценивании параметров нормального распределения. На рис. 4.3.4-4.3.6 - распределения статистики S_{χ^2} при оценивании параметров логистического распределения. На рис. 4.3.7-4.3.9 - распределения статистики при оценивании параметров распределения Коши. На рис. 4.3.10-4.3.12 - распределения статистики при оценивании параметров распределения максимального значения.

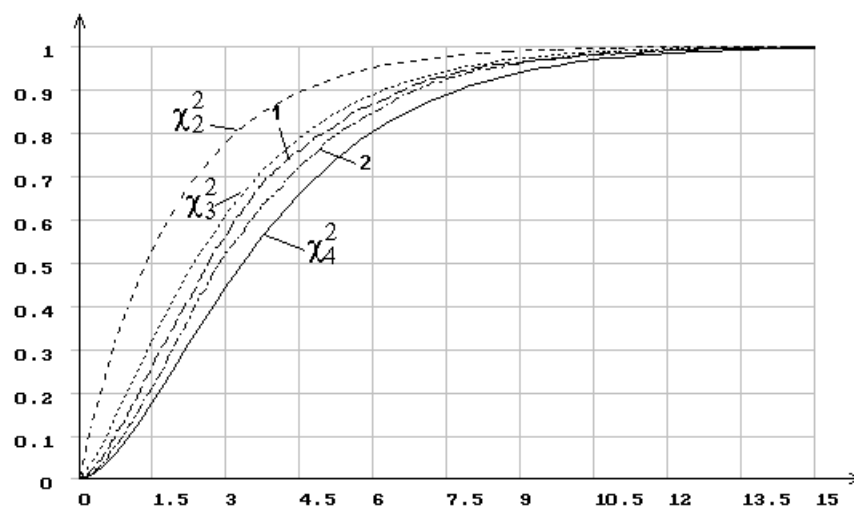


Рис. 4.3.1. Распределение статистики S_{χ^2} при 5 интервалах группирования и оценивании масштабного параметра нормального распределения.

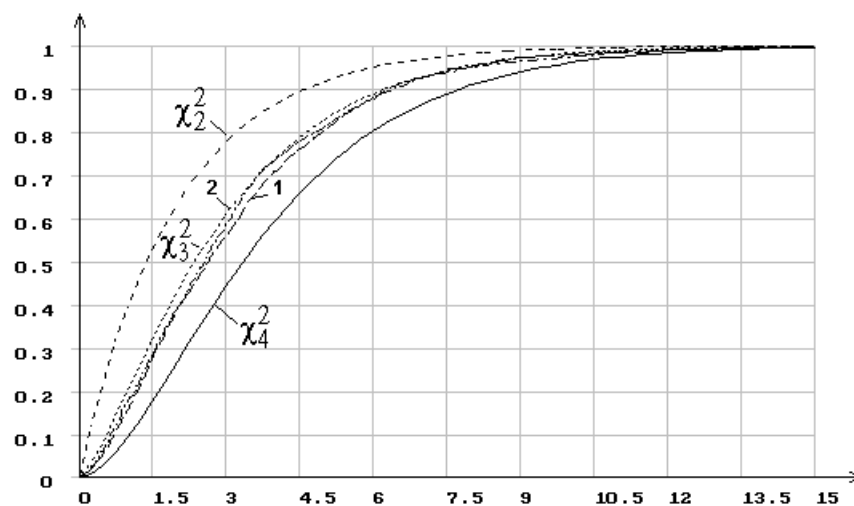


Рис. 4.3.2. Распределение статистики S_{χ^2} при 5 интервалах группирования и оценивании параметра сдвига нормального распределения.

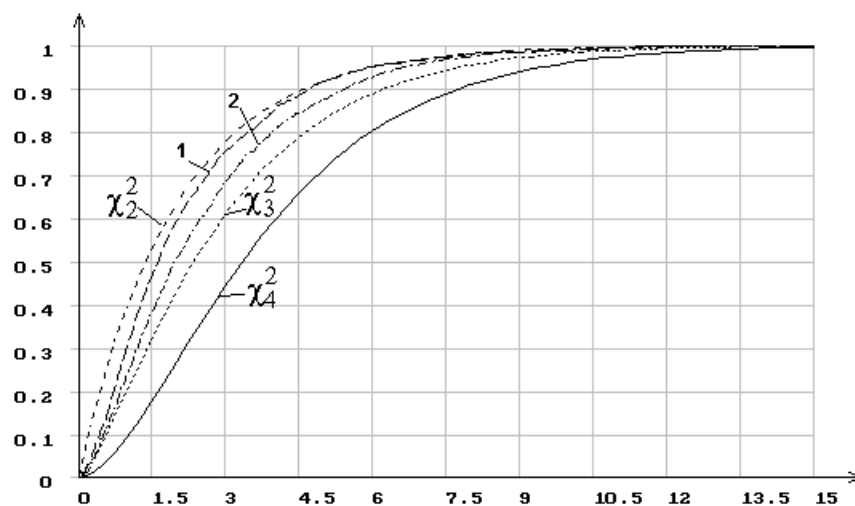


Рис. 4.3.3. Распределение статистики S_{χ^2} при 5 интервалах группирования и оценивании двух параметров нормального распределения.

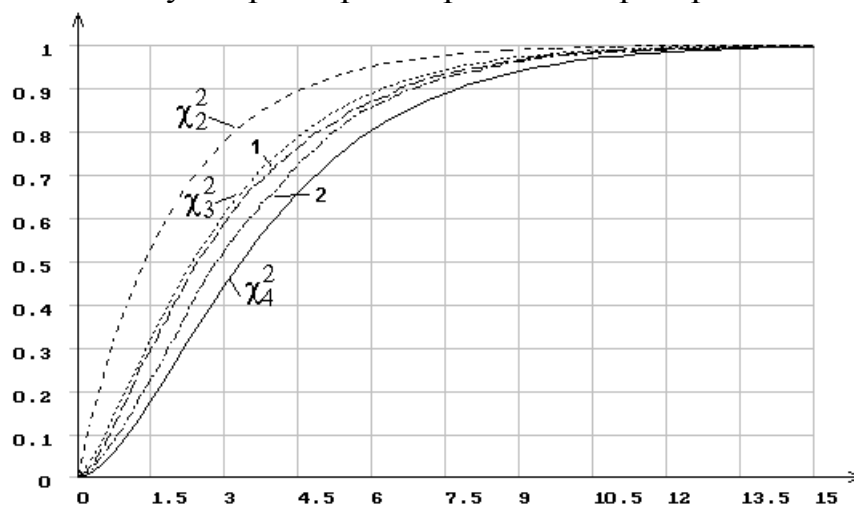


Рис. 4.3.4. Распределение статистики S_{χ^2} при 5 интервалах группирования и оценивании масштабного параметра логистического распределения.

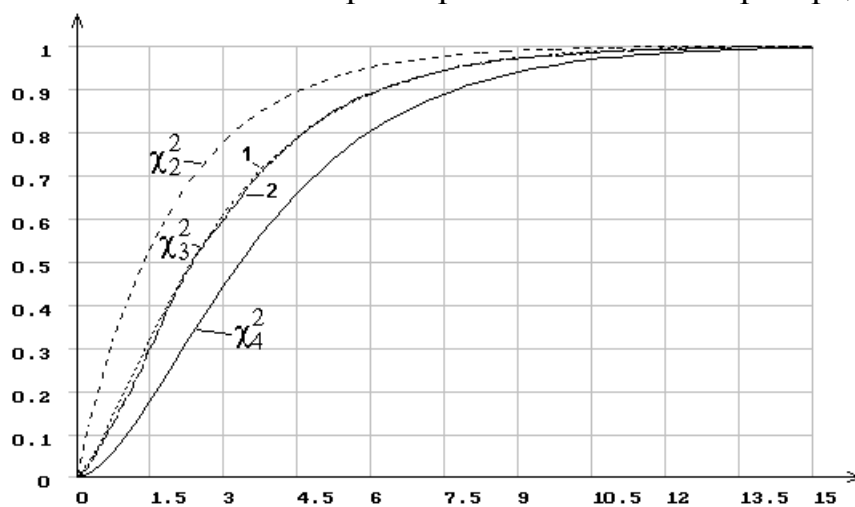


Рис. 4.3.5. Распределение статистики S_{χ^2} при 5 интервалах группирования и оценивании параметра сдвига логистического распределения.

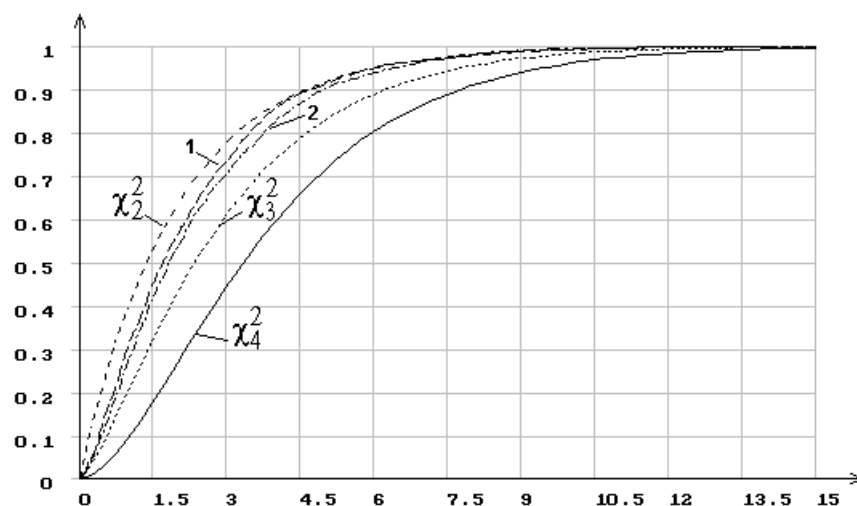


Рис. 4.3.6. Распределение статистики S_{χ^2} при 5 интервалах группирования и оценивании двух параметров логистического распределения.

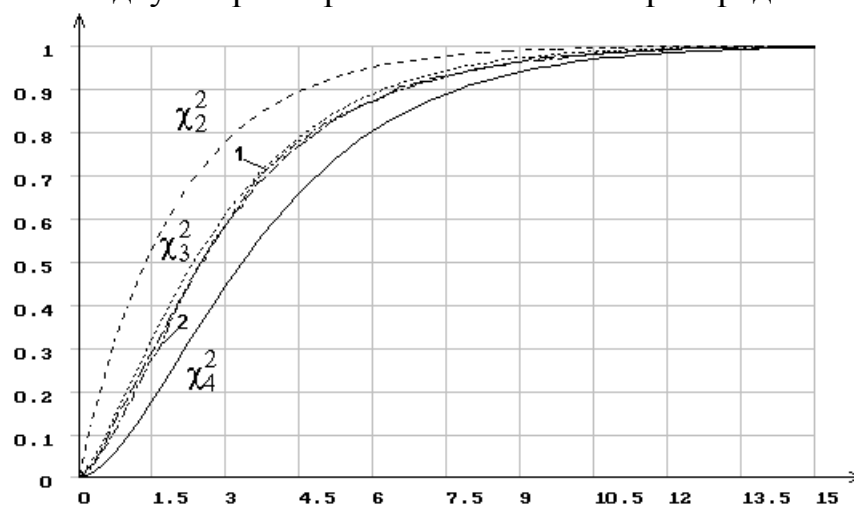


Рис. 4.3.7. Распределение статистики S_{χ^2} при 5 интервалах группирования и оценивании масштабного параметра распределения Коши.

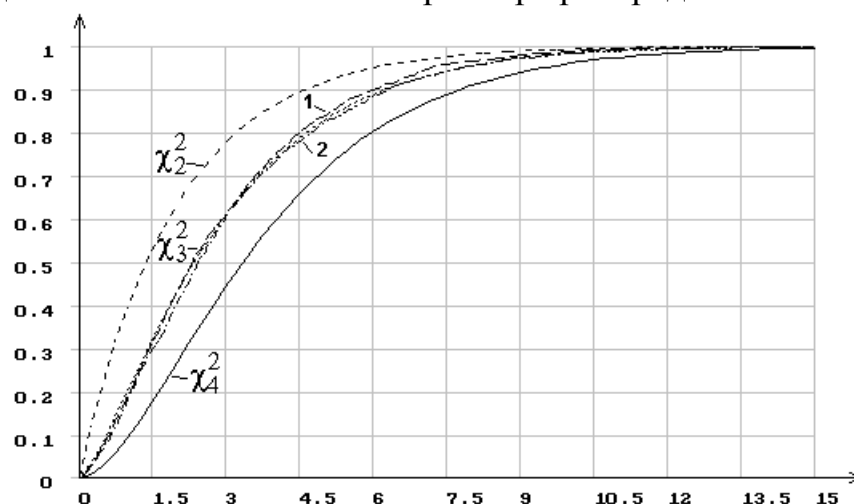


Рис. 4.3.8. Распределение статистики S_{χ^2} при 5 интервалах группирования и оценивании параметра сдвига распределения Коши.

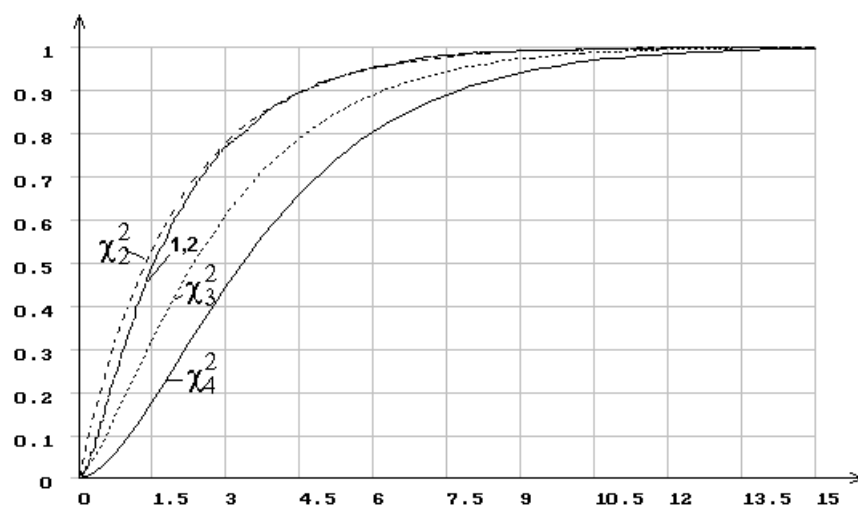


Рис. 4.3.9. Распределение статистики S_{χ^2} при 5 интервалах группирования и оценивании двух параметров распределения Коши.

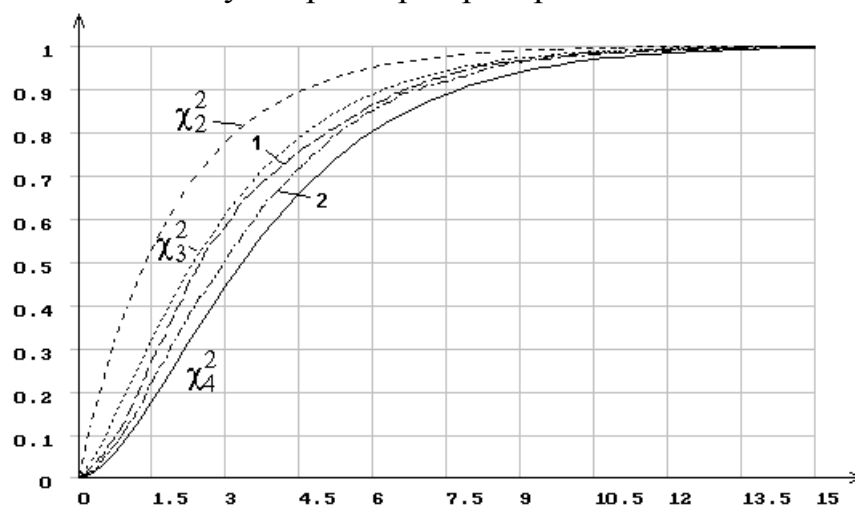


Рис. 4.3.10. Распределение статистики S_{χ^2} при 5 интервалах и оценивании масштабного параметра распределения максимального значения.

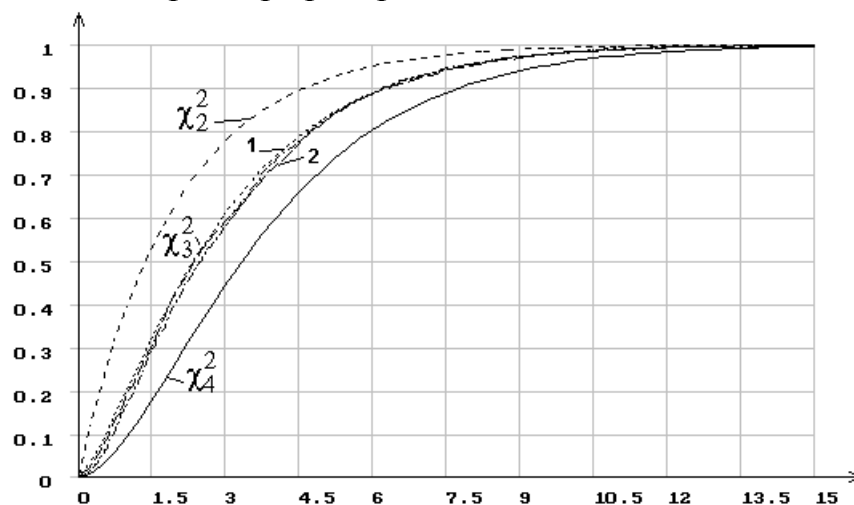


Рис. 4.3.11. Распределение статистики S_{χ^2} при 5 интервалах и оценивании параметра сдвига распределения максимального значения.

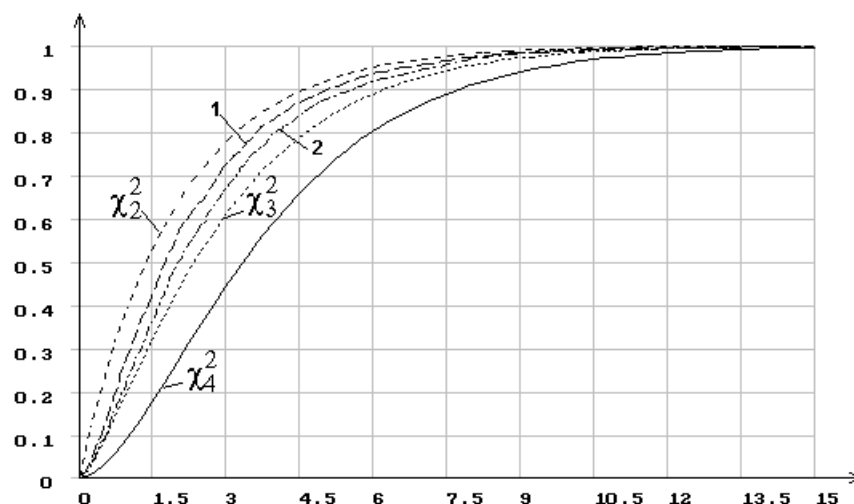


Рис. 4.3.12. Распределение статистики S_{χ^2} при 5 интервалах и оценивании двух параметров распределения максимального значения.

Примечание: Асимптотически оптимальное группирование для параметра сдвига логистического распределения и одновременно 2-х параметров распределения Коши совпадает с равновероятным. Это и отразилось на рис. 4.3.5 и 4.3.9.

Для сравнения на рис. 4.3.13 представлены распределения статистики $S_{\text{н}}$ при 5 интервалах группирования и оценивании обоих параметров нормального распределения при асимптотически оптимальном и равновероятном группировании.

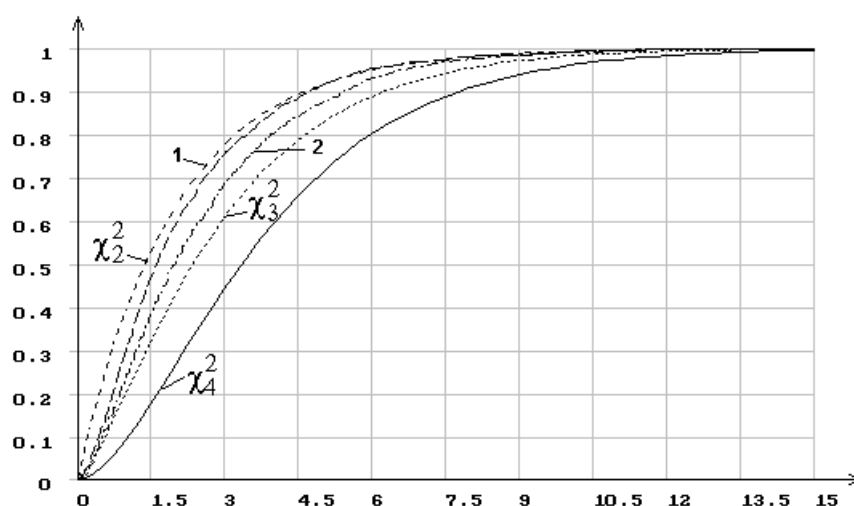


Рис. 4.3.13. Распределение статистики $S_{\text{оп}}$ при 5 интервалах группирования и оценивании двух параметров нормального распределения.

Какие изменения происходят с исследуемыми распределениями статистик с ростом числа интервалов? На рис. 4.3.14-4.3.16 приве-

дены распределения статистики S_{χ^2} при 7 интервалах группирования и оценивании параметров нормального распределения (моделировалась серия из $n = 2000$ выборок объемом $N = 500$.). Распределения статистики S_{χ^2} при 7 интервалах группирования в случае оценивания параметров нормального распределения представлены на рис. 4.3.17-4.3.19, на которых наблюдается аналогичная картина.

Результаты моделирования и анализа указывают на то, что предельные распределения статистик обоих рассматриваемых критериев существенно отличаются при различных способах группирования (при вычислении оценок по негруппированным данным). При этом эмпирический закон распределения статистики при асимптотически оптимальном группировании *ближе* к теоретическому χ^2_{k-p-1} -распределению, чем при равновероятном группировании. Мало того, *распределения статистик зависят* не только от количества оцененных по выборке параметров, но и *от того, какой параметр оценивался*. Например, оценивание параметра сдвига приводит к более значительному изменению распределения статистики, чем оценивание масштабного параметра.

Следует отметить, что эмпирические распределения статистики отношения правдоподобия оказываются ближе к предельному теоретическому χ^2_{k-p-1} -распределению, чем соответствующие эмпирические распределения статистики χ^2 Пирсона.

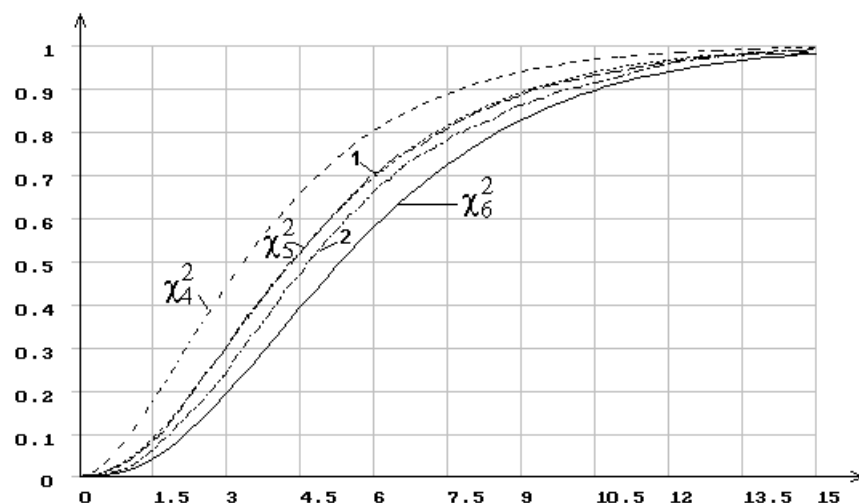


Рис. 4.3.14. Распределение статистики S_{χ^2} при 7 интервалах и оценивании масштабного параметра нормального распределения.

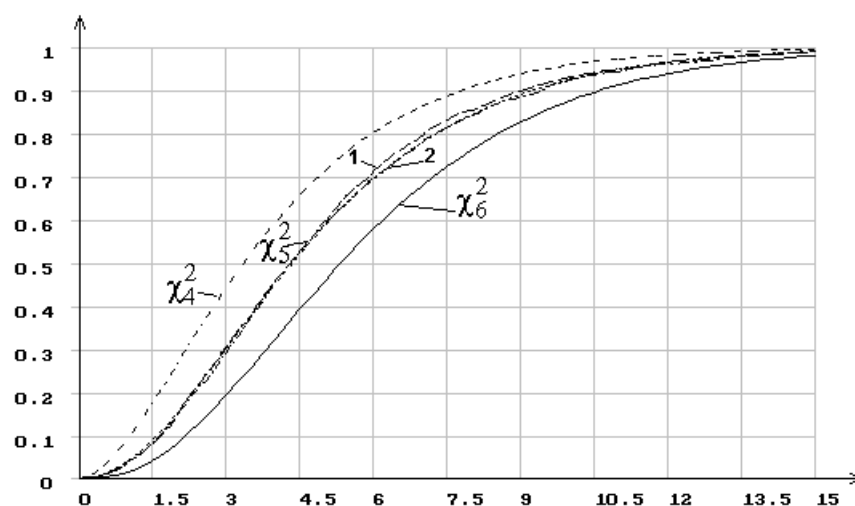


Рис. 4.3.15. Распределение статистики S_{χ^2} при 7 интервалах и оценивании параметра сдвига нормального распределения.

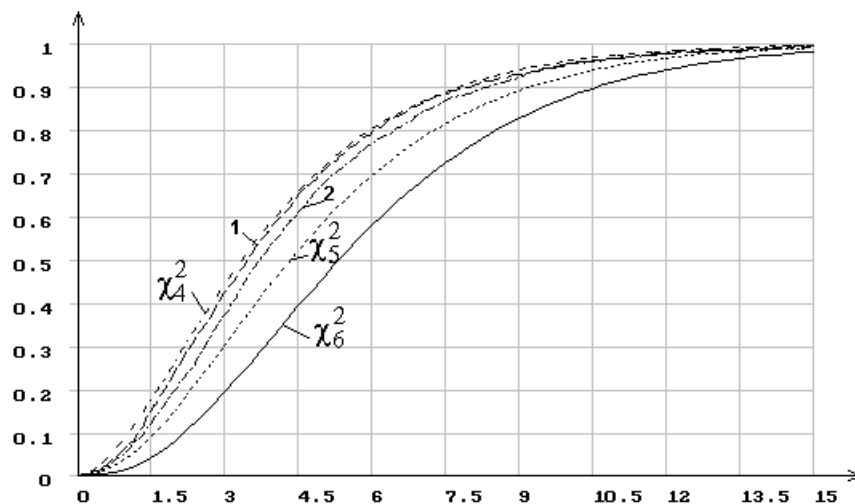


Рис. 4.3.16. Распределение статистики S_{χ^2} при 7 интервалах и оценивании двух параметров нормального распределения.

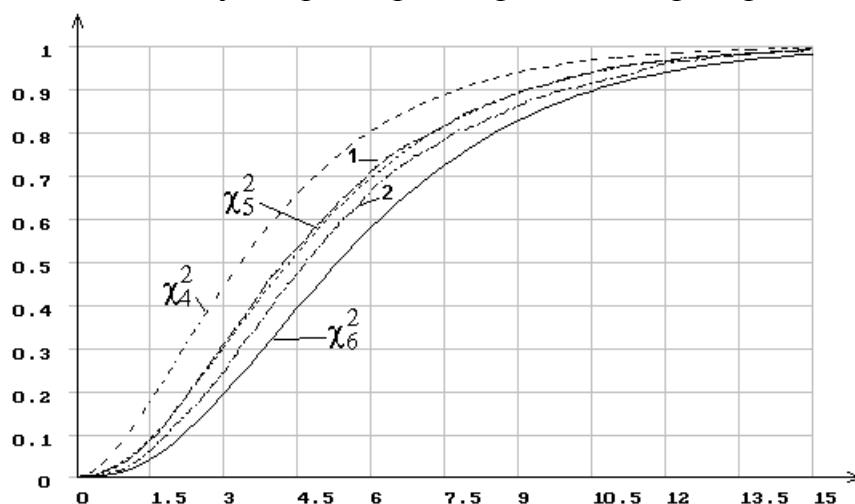


Рис. 4.3.17. Распределение статистики $S_{оп}$ при 7 интервалах и оценивании масштабного параметра нормального распределения.

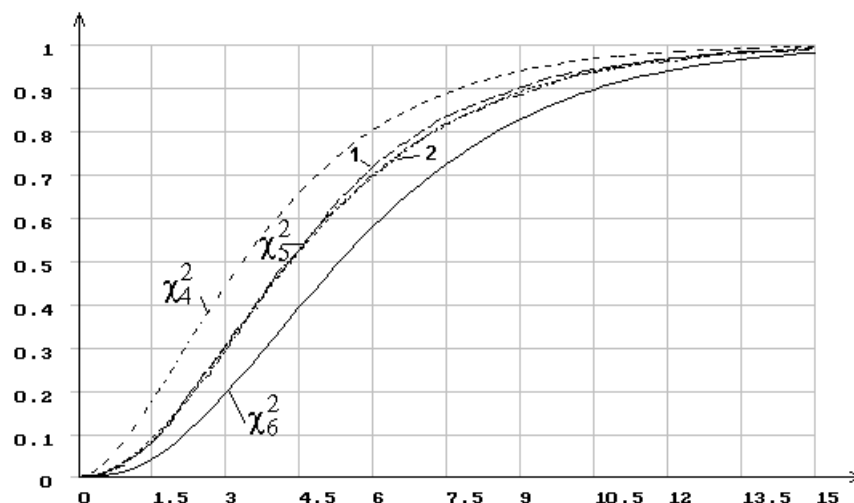


Рис. 4.3.18. Распределение статистики $S_{\text{оп}}$ при 7 интервалах и оценивании параметра сдвига нормального распределения.

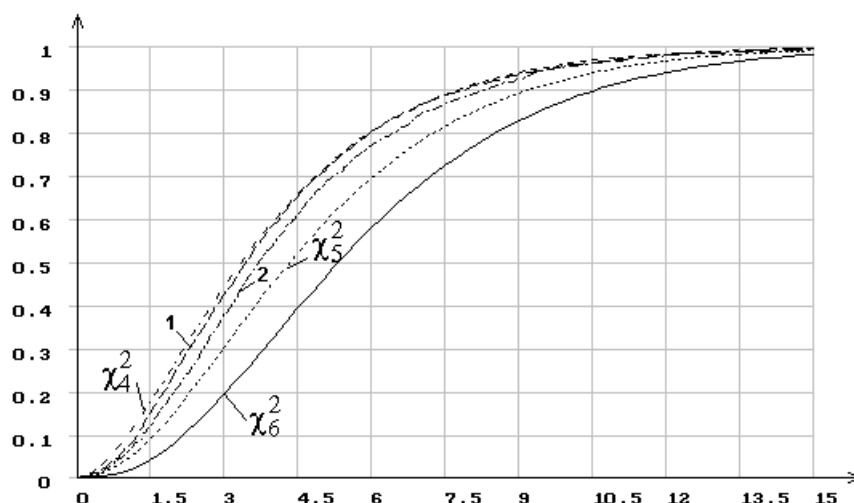


Рис. 4.3.19. Распределение статистики $S_{\text{оп}}$ при 7 интервалах и оценивании двух параметров нормального распределения.

χ_r^2 -распределение с числом степеней свободы r является частным случаем гамма-распределения с основным параметром, равным $r/2$, и масштабным - 0.5. Оценивание одного из параметров учитывается уменьшением числа степеней свободы на 1. На самом деле, если измерять изменение предельного закона “в степенях свободы”, то оказывается, что оценивание даже параметра сдвига обычно приводит к изменению “числа степеней” на величину меньшую 1, еще к меньшему изменению в степенях свободы приводит оценивание масштабного параметра. Это особенно заметно при оценивании 2-х параметров и указывает на то, что использование в критериях согласия χ_{k-p-1}^2 -распределения чревато занижением вероятности вида $P\{S > S^*\}$ и определенным риском отвергнуть верную гипотезу.

тезу H_0 . Занижение $P\{S > S^*\}$ более существенно при меньшем числе интервалов группирования. В целом этот риск существенно меньше в случае применения асимптотически оптимального группирования.

При идентификации типов предельных законов распределения статистик использовалось множество законов и семейств распределений, включенных в программную систему [76]. Эмпирические законы распределения статистик рассматриваемых критериев практически всегда с достаточно высокой точностью описываются *гамма-распределением*, то есть эмпирическая функция статистики очень хорошо согласуется с теоретическим законом по всем используемым в системе критериям согласия.

В табл. 4.3.1-4.3.2 сведены результаты идентификации законов, соответственно, для статистик критериев χ^2 Пирсона и отношения правдоподобия при 5-ти интервалах группирования в случае асимптотически оптимального (АОГ) и равновероятного группирования (РВГ). В соответствующих клетках таблиц приведены значения параметров $\theta_0, \theta_1, \theta_2$ гамма-распределения с функцией плотности

$$f(x) = \frac{\theta_1^{\theta_0}}{\Gamma(\theta_0)} (x - \theta_2)^{\theta_0 - 1} e^{-\theta_1(x - \theta_2)}.$$

На первом месте указаны параметры $\theta_0, 0.5, \theta_2$ гамма-распределения, в котором по эмпирическому распределению оценивался параметр θ_0 , позволяющий судить о действительном “числе степеней свободы”, на втором - параметры гамма-распределения, согласие с которым наилучшее. По приводимым значениям параметров можно судить об отличии предельного распределения от соответствующего χ^2_{k-p-1} .

В работе исследовались распределения статистик при принадлежности наблюдаемой выборки различным законам. В таблицах представлены распределения статистик, когда наблюдаемые случайные величины распределены в соответствии с законами: нормальным, с функцией

плотности нормальным - $f(x) = \frac{1}{\theta_0 \sqrt{2\pi}} e^{-(x - \theta_1)^2 / 2\theta_0^2}$; Коши -

$f(x) = \frac{\theta_0}{\pi[\theta_0^2 + (x - \theta_1)^2]}$; наибольшего значения -

$f(x) = \frac{1}{\theta_0} \exp\left\{-\frac{x - \theta_1}{\theta_0} - \exp\left(-\frac{x - \theta_1}{\theta_0}\right)\right\}$ и логистическим -

$$f(x) = \frac{\pi}{\theta_0 \sqrt{3}} \exp\left\{-\frac{\pi(x - \theta_1)}{\theta_0 \sqrt{3}}\right\} / \left[1 + \exp\left\{-\frac{\pi(x - \theta_1)}{\theta_0 \sqrt{3}}\right\}\right]^2.$$

Таблица 4.3.1.

<i>Предельные распределения статистики χ^2 Пирсона</i>					
№ п/п	Распредел-е случайной величины	Параметры по выборке не оценивались	Оценивался только масштабный параметр	Оценивался только параметр сдвига	Оценивалось два параметра
1.	Нормальное (АОГ)	1.9443, 0.5, 0.0 1.9482, 0.5011, 0.0	1.6810, 0.5, 0.0 1.7471, 0.5267, 0.0	1.6493, 0.5, 0.0 1.7263, 0.5318, 0.0	1.2165, 0.5, 0.0 1.4498, 0.6465, 0.0
2.	Нормальное (РВГ)	2.0421, 0.5, 0.0 1.9940, 0.4851, 0.0	1.8134, 0.5, 0.0 1.9394, 0.5463, 0.0	1.6009, 0.5, 0.0 1.6336, 0.5142, 0.0	1.3635, 0.5, 0.0 1.5358, 0.5922, 0.0
3.	Логистич-кое (АОГ)	1.9898, 0.5, 0.0 1.9090, 0.4737, 0.0	1.6028, 0.5, 0.0 1.6070, 0.5019, 0.0	1.5466, 0.5, 0.0 1.5847, 0.5173, 0.0	1.2215, 0.5, 0.0 1.4158, 0.6212, 0.0
4.	Логистич-кое (РВГ)	2.0352, 0.5, 0.0 1.9662, 0.4785, 0.0	1.7786, 0.5, 0.0 1.8281, 0.5189, 0.0	1.5523, 0.5, 0.0 1.6144, 0.5278, 0.0	1.2904, 0.5, 0.0 1.4974, 0.6196, 0.0
5.	Коши (АОГ)	2.0534, 0.5, 0.0 1.9470, 0.4667, 0.0	1.5906, 0.5, 0.0 1.5730, 0.4925, 0.0	1.5177, 0.5, 0.0 1.6500, 0.5609, 0.0	1.1623, 0.5, 0.0 1.3585, 0.6312, 0.0
6.	Коши (РВГ)	2.0534, 0.5, 0.0 1.9465, 0.4667, 0.0	1.6204, 0.5, 0.0 1.6658, 0.5195, 0.0	1.5681, 0.5, 0.0 1.6482, 0.5353, 0.0	1.1470, 0.5, 0.0 1.2997, 0.6039, 0.0
7.	Наибольшего знач. (АОГ)	1.9942, 0.5, 0.0 2.0244, 0.5100, 0.0	1.6547, 0.5, 0.0 1.7134, 0.5245, 0.0	1.5573, 0.5, 0.0 1.5758, 0.5085, 0.0	1.2721, 0.5, 0.0 1.4362, 0.5965, 0.0
8.	Наибольшего знач. (РВГ)	2.0364, 0.5, 0.0 2.0144, 0.4932, 0.0	1.8152, 0.5, 0.0 1.8991, 0.5308, 0.0	1.5916, 0.5, 0.0 1.6751, 0.5361, 0.0	1.3674, 0.5, 0.0 1.4200, 0.5280, 0.0

Таблица 4.3.2.

<i>Предельные распределения статистики отношения правдоподобия</i>					
№ п/п	Распредел-е случайной величины	Параметры по выборке не оценивались	Оценивался только масштабный параметр	Оценивался только параметр сдвига	Оценивалось два параметра
1.	Нормальное (АОГ)	1.9307, 0.5, 0.0 1.9463, 0.5054, 0.0	1.6568, 0.5, 0.0 1.7959, 0.5572, 0.0	1.6419, 0.5, 0.0 1.7388, 0.5403, 0.0	1.2150, 0.5, 0.0 1.4646, 0.6571, 0.0
2.	Нормальное (РВГ)	2.0357, 0.5, 0.0 1.9958, 0.4877, 0.0	1.8083, 0.5, 0.0 1.9491, 0.5518, 0.0	1.5986, 0.5, 0.0 1.6383, 0.5171, 0.0	1.3458, 0.5, 0.0 1.4769, 0.5713, 0.0
3.	Логистич-кое (АОГ)	1.9808, 0.5, 0.0 1.9159, 0.4789, 0.0	1.5913, 0.5, 0.0 1.6344, 0.5188, 0.0	1.5426, 0.5, 0.0 1.5931, 0.5229, 0.0	1.2176, 0.5, 0.0 1.4276, 0.6316, 0.0
4.	Логистич-кое (РВГ)	2.0300, 0.5, 0.0 1.9715, 0.4818, 0.0	1.7750, 0.5, 0.0 1.8385, 0.5241, 0.0	1.5233, 0.5, 0.0 1.5246, 0.5004, 0.0	1.2878, 0.5, 0.0 1.5059, 0.6266, 0.0
5.	Коши (АОГ)	2.0458, 0.5, 0.0 1.9625, 0.4740, 0.0	1.5889, 0.5, 0.0 1.5688, 0.4916, 0.0	1.5100, 0.5, 0.0 1.6625, 0.5707, 0.0	1.1598, 0.5, 0.0 1.3664, 0.6387, 0.0
6.	Коши (РВГ)	2.0458, 0.5, 0.0 1.9623, 0.4740, 0.0	1.6169, 0.5, 0.0 1.6779, 0.5259, 0.0	1.5649, 0.5, 0.0 1.6506, 0.5378, 0.0	1.1471, 0.5, 0.0 1.3166, 0.6154, 0.0
7.	Наибольшего знач. (АОГ)	1.9734, 0.5, 0.0 2.0228, 0.5165, 0.0	1.6201, 0.5, 0.0 1.7708, 0.5638, 0.0	1.5415, 0.5, 0.0 1.6039, 0.5283, 0.0	1.2642, 0.5, 0.0 1.4643, 0.6190, 0.0
8.	Наибольшего знач. (РВГ)	2.0303, 0.5, 0.0 2.0201, 0.4970, 0.0	1.8109, 0.5, 0.0 1.9077, 0.5357, 0.0	1.5871, 0.5, 0.0 1.6721, 0.5368, 0.0	1.3599, 0.5, 0.0 1.4148, 0.5293, 0.0

Как говорилось выше, использование в критериях согласия χ^2_{k-p-1} -распределения для вычисления вероятности вида $P\{S > S^*\}$ связано с риском отвергнуть верную гипотезу H_0 . В случае применения асимп-

тотически оптимального группирования этот риск существенно меньше. В качестве примера в табл. 4.3.3 приводятся значения вероятностей $P\{S > S^*\}$ для различных значений статистики S^* , вычисленные по χ^2_4 -распределению и в соответствии с “действительными” распределениями статистик при 7 интервалах группирования и оценивании 2-х параметров нормального распределения. Эти данные, вместе с другими результатами, позволяют сделать вывод, что в случае применения асимптотически оптимального группирования и использования χ^2_{k-p-1} -распределения ошибки при вычислении вероятности $P\{S > S^*\}$ практически незначимы.

Таблица 4.3.3.

Значения вероятностей вида $P\{S > S^*\}$					
Знач. S^*	Распределение χ^2_4	Распределение статистики $S_{\text{оп}}$		Распределение статистики S_{χ^2}	
		При АОГ	При РВГ	При АОГ	При РВГ
3	0.5578	0.5876	0.6293	0.5914	0.6304
4	0.4060	0.4303	0.4700	0.4356	0.4716
5	0.2873	0.3049	0.3384	0.3108	0.3403
6	0.1991	0.2108	0.2370	0.2165	0.2389
7	0.1359	0.1430	0.1626	0.1481	0.1643
8	0.0916	0.0956	0.1096	0.0999	0.1111
9	0.0611	0.0632	0.0729	0.0666	0.0741
10	0.0404	0.0413	0.0480	0.0440	0.0489
11	0.0266	0.0268	0.0313	0.0288	0.0320
12	0.0173	0.0173	0.0202	0.0187	0.0207
13	0.0113	0.0110	0.0130	0.0120	0.0133

Посмотрим, что будет происходить при дальнейшем увеличении числа интервалов. На рис. 4.3.20-4.3.22 приведены эмпирические распределения статистики S_{χ^2} при 10 интервалах группирования и оценивании параметров нормального распределения (при справедливой гипотезе H_0). О степени близости эмпирических распределений статистики S_{χ^2} соответствующим χ^2_{k-p-1} -распределениям можно судить по оценкам параметров гамма-распределения, полученным по смоделированным выборкам этой статистики и представленным в табл. 4.3.4 (в соответствии с нормальным распределением моделировалось $n=2000$ выборок объемом $N=1000$, число интервалов $k=10$). На первом месте указаны параметры $\theta_0, \theta_1, \theta_2$ гамма-распределения, в котором по эмпирическому распределению

оценивался параметр θ_0 , на втором - параметры, где оценивались θ_0 и θ_1 . Как видим, при 10 интервалах и оценивании одного параметра распределения статистики и при АОГ, и при РВГ достаточно близки к теоретическому χ^2_{k-p-1} (разницы практически нет: см. рис. 4.3.21-4.3.22 и табл. 4.3.4). При оценивании одновременно 2-х параметров разница в распределениях еще заметна (причем для АОГ эмпирическое распределение даже выше теоретического χ^2_7). Таким образом, при объеме выборки, позволяющем выбрать количество интервалов больше 10, при вычислении ОМП по негруппированным данным можно с уверенностью использовать в рассматриваемых критериях χ^2_{k-p-1} -распределение.

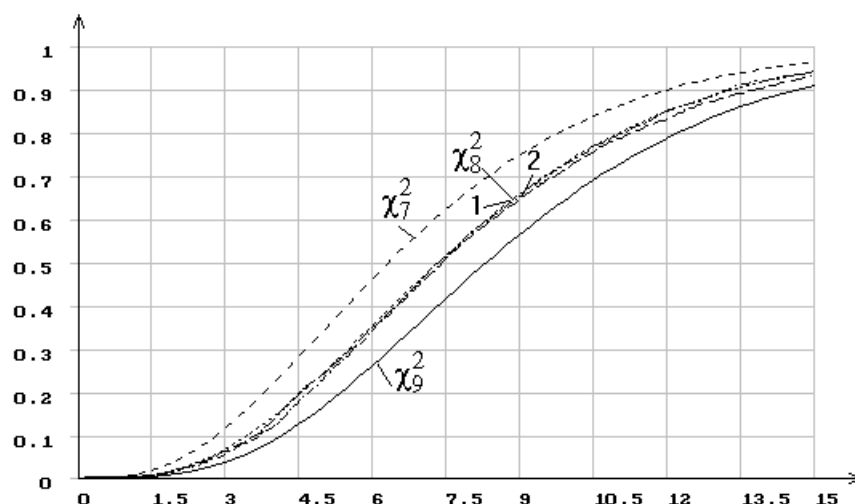


Рис. 4.3.20. Распределение статистики S_{χ^2} при 10 интервалах группирования и оценивании масштабного параметра нормального распределения

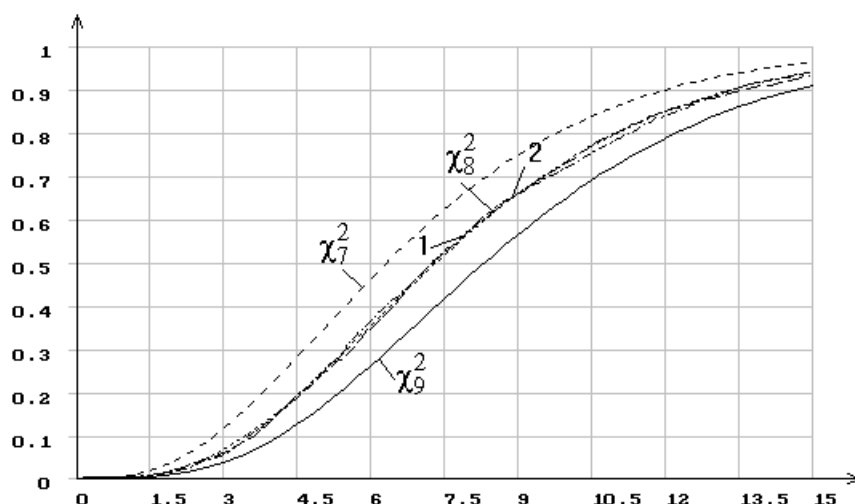


Рис. 4.3.21. Распределение статистики S_{χ^2} при 10 интервалах группирования и оценивании параметра сдвига нормального распределения

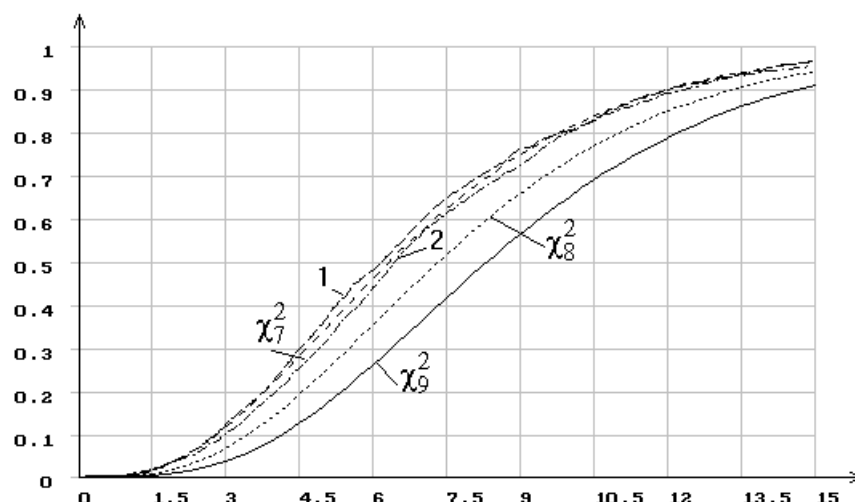


Рис. 4.3.22. Распределение статистики S_{χ^2} при 10 интервалах группирования и оценивании двух параметров нормального распределения

Таблица 4.3.4.

<i>Предельные распределения статистики χ^2 Пирсона при $k=10$</i>		
Оцениваемый параметр нормального закона	Параметры “предельного” гамма-распределения при АОГ	Параметры “предельного” гамма-распределения при РВГ
Параметры не оценивались	4.5483, 0.5, 0.0 4.3343, 0.4732, 0.0	4.3599, 0.5, 0.0 4.3034, 0.4930, 0.0
Масштабный параметр	4.0724, 0.5, 0.0 3.9629, 0.4851, 0.0	4.0751, 0.5, 0.0 4.2553, 0.5253, 0.0
Параметр сдвига	4.0466, 0.5, 0.0 4.0395, 0.4995, 0.0	4.0332, 0.5, 0.0 4.0349, 0.5005, 0.0
Оба параметра	3.4717, 0.5, 0.0 3.4030, 0.4890, 0.0	3.6249, 0.5, 0.0 3.5647, 0.4909, 0.0

Посмотрим, как отражается способ группирования на предельных распределениях статистики S_{χ^2} при справедливости альтернативной гипотезы H_1 . На рис. 4.3.23-4.3.24 приведены полученные в результате моделирования функции распределения $G(S(H_1))$ при использовании асимптотически оптимального и равновероятного группирования, когда гипотеза H_0 соответствует нормальному закону распределения, а гипотеза H_1 - логистическому. Эти два закона достаточны близки и трудно различаются с помощью критериев согласия. На рис. 4.3.23 приведены эмпирические распределения статистики при простой гипотезе H_0 , а на рис. 4.3.24 -

при сложной. Количество интервалов группирования - 7. Из рисунков видно, на сколько выше мощность критерия при асимптотически оптимальном группировании по сравнению равновероятным.

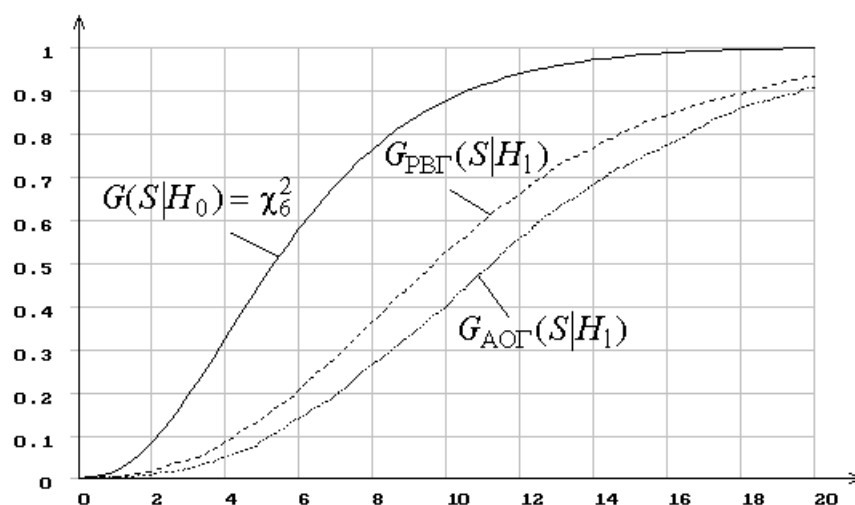


Рис. 4.3.23. Распределения статистики S_{χ^2} при проверке простой гипотезы H_0

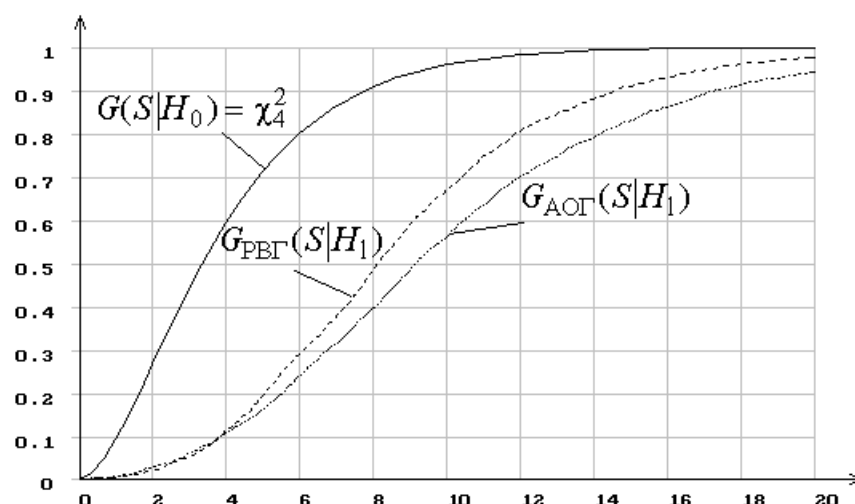


Рис. 4.3.24. Распределения статистики S_{χ^2} при проверке сложной гипотезы H_0

В работах [114, 115, 107] рекомендуется видоизменение стандартной статистики S_{χ^2} , при котором предельное распределение есть обычное распределение χ_{k-1}^2 с $(k-1)$ -й степенью свободы и в том случае, когда в качестве оценок используются оценки максимального правдоподобия параметров по негруппированным данным. При этом вектор $P = (P_1, \dots, P_k)^T$ предполагается заданным и граничные точки интервалов определяются соотношениями $x_i(\theta) = F^{-1}(P_1 + \dots + P_i)$, $i = \overline{1, (k-1)}$. Предложенная статистика имеет вид [115]

$$Y_N^2(\theta) = S_{\chi^2} + N^{-1} a^T(\theta) \Lambda(\theta) a(\theta),$$

где матрица $\Lambda(\theta) = \left\| I(\theta_l, \theta_j) - \sum_{i=1}^k \frac{w_{\theta_i l} w_{\theta_i j}}{P_i} \right\|^{-1}$, и ее элементы и размерность определяются оцениваемыми компонентами вектора параметров θ ; $I(\theta_l, \theta_j) = \int \left(\frac{\partial \ln f(x, \theta)}{\partial \theta_l} \frac{\partial \ln f(x, \theta)}{\partial \theta_j} \right) f(x, \theta) dx$ - элементы информационной матрицы Фишера;

$$w_{\theta_i l} = -f[x_i(\theta), \theta] \frac{\partial x_i(\theta)}{\partial \theta_l} + f[x_{i-1}(\theta), \theta] \frac{\partial x_{i-1}(\theta)}{\partial \theta_l};$$

$$a(\theta_l) = w_{\theta_{l1}} n_1 / P_1 + \dots + w_{\theta_{lk}} n_k / P_k$$
 - элементы вектора $a(\theta)$.

Моделирование распределений статистики $Y_N^2(\theta)$ при вычислении оценок максимального правдоподобия параметров по негруппированным данным показало, что её эмпирические распределения при верной нулевой гипотезе $g(S(H_0))$ также зависят от способа группирования. Но это различие одного и того же порядка для распределений статистик при оценивании и без оценивания параметров. Причем законы распределения, описывающие выборки, достаточно близки χ_{k-1}^2 -распределениям. О степени близости можно судить по оценкам параметров гамма-распределения, полученным по смоделированным и представленным в табл. 4.3.5 выборкам статистики $Y_N^2(\theta)$ (H_0 : нормальное распределение, моделировалось $n = 2000$ выборок объемом $N = 500$, число интервалов $k = 5$).

Таблица 4.3.5.

Оцениваемый параметр нормального закона	Параметры “предельного” гамма-распределения при АОГ	Параметры “предельного” гамма-распределения при РВГ
Параметры не оценивались	2.0508, 0.5302, 0.0	1.9855, 0.4770, 0.0
Масштабный параметр	2.0612, 0.4988, 0.0	1.9195, 0.4827, 0.0
Параметр сдвига	1.9702, 0.5063, 0.0	1.9833, 0.5035, 0.0
Оба параметра	1.9366, 0.5051, 0.0	2.0345, 0.5089, 0.0

Распределения статистики $Y_N^2(\theta)$ при справедливой конкурирующей гипотезе $g(S(H_1))$, также как и распределения статистики S_{χ^2} , существенно зависят от способа группирования.

Выводы

Результаты моделирования и анализа показали, что если по наблюдаемой выборке не оценивались параметры закона, то распределения статистик критериев отношения правдоподобия и χ^2 Пирсона при справедливой гипотезе H_0 достаточно хорошо согласуются с χ^2_{k-1} как при равновероятном, так и при асимптотически оптимальном группировании. Различия между распределениями статистик при равновероятном и асимптотически оптимальном группировании заметны, но не значимы. При этом распределения статистик при равновероятном группировании в целом оказываются ближе к χ^2_{k-1} распределению.

В условиях, когда по негруппированным наблюдениям определяются оценки параметров, распределения этих статистик хорошо согласуются с гамма-распределениями. При этом:

1. Предельные распределения статистик критериев отношения правдоподобия и χ^2 Пирсона существенно зависят от способа группирования, особенно при малом числе интервалов.
2. Распределения статистик зависят не только от количества оцененных по выборке параметров, но и от того, какой параметр оценивался. Оценивание параметра сдвига приводит к более значительному изменению распределений статистик, чем оценивание масштабного параметра.
3. Эмпирические распределения статистики отношения правдоподобия оказываются ближе к предельному теоретическому χ^2_{k-p-1} -распределению, чем соответствующие эмпирические распределения статистики χ^2 Пирсона.
4. В целом, при малом числе интервалов и оценивании p параметров число степеней свободы предельного распределения уменьшается на “число степеней свободы” меньше p . При этом эмпирический закон распределения статистики при асимптотически оптимальном группировании ближе к теоретическому χ^2_{k-p-1} -распределению, чем при равновероятном группировании.
5. С ростом числа интервалов k разность между функцией распределения χ^2_{k-p-1} и действительными функциями распределения статистик отношения правдоподобия и χ^2 Пирсона в случае

асимптотически оптимального группирования убывает существенно быстрее.

6. Расчеты показывают, что при использовании для вычисления $P\{S > S^*\}$ функции распределения χ^2_{k-p-1} при асимптотически оптимальном группировании и малом числе интервалов ($k < 10$) при малых $P\{S > S^*\}$ (именно при малых вероятностях принимается решение отклонить гипотезу H_0 или нет) погрешность имеет величины, которые не существенны для практических задач. То есть в этом случае, используя χ^2_{k-p-1} , мы не совершаем большой ошибки.
7. С ростом количества интервалов (при соответствующем объеме выборки) отличие предельных распределений от распределений χ^2_{k-p-1} и при асимптотически оптимальном и при равновероятном группировании становится несущественным.

Распределения статистик S_{χ^2} и S_{ω} при справедливой гипотезе H_1 всегда (и с оцениванием параметров и без оценивания) сильно зависят от варианта группирования. Разность $G(S(H_0)) - G(S(H_1))$ при близких альтернативах и асимптотически оптимальном группировании максимальна, следовательно, максимальна мощность критерия.

4.4. Применение непараметрических критериев согласия в случае потери ими свойства “свободы от распределения”

4.4.1. Влияние оценивания параметров закона на предельные распределения статистик непараметрических критериев

Наиболее часто в практике статистического анализа с необходимостью использования критериев согласия приходится сталкиваться после оценивания по этой же выборке параметров предполагаемого закона распределения. К сожалению, в этом случае предельные распределения статистик таких непараметрических критериев, как Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса, при справедливости нулевой гипотезы вида $H_0: f(x, \theta_0) = f(x, \hat{\theta})$, где $f(\cdot)$ - плотность распределения наблюдаемого закона, θ_0 - истинное значение параметра, $\hat{\theta}$ - оценка параметра, вычисленная по выборке, отличаются от предельных распределений классических статистик

(когда по выборке не оцениваются параметры). При оценивании по выборке параметров рассматриваемые критерии теряют свойство “свободы от распределения”, и предельные распределения статистик на самом деле зависят как от числа оцененных параметров, так и от вида исследуемого закона распределения $f(x, \theta)$. Широкую известность этот факт приобрел, начиная с работы [233]. Распространенная ошибка, связанная с пренебрежением существующей проблемы, чаще всего приводит к необоснованному принятию нулевой гипотезы, что подчеркивается в работах [120,20], из-за сильно завышенных значений вероятностей “согласия” вида $P\{S > S^*\}$, где S^* - значение статистики, вычисленное по выборке. И как отмечено в [64], это не всегда учитывается в популярных пакетах статистического анализа.

Если объем выборки достаточно велик, можно, опираясь на результаты, полученные в [210], оценивать параметры распределения по одной половине выборки, а проверять согласие по другой половине. В такой ситуации применение предельных распределений рассматриваемых критериев вполне обосновано. Но в этом случае как при оценивании, так и при проверке согласия мы используем только половину имеющейся информации, что, естественно, сказывается на качестве статистических выводов. К тому же, объемы выборок, зачастую имеющиеся в распоряжении исследователя, не настолько велики, чтобы можно было смириться с потерей части информации при оценивании параметров.

Таким образом, с одной стороны некорректность использования классических непараметрических критериев в задачах проверки сложных гипотез из-за слишком больших вероятностей ошибок, с другой стороны свойства асимптотической оптимальности критериев типа Колмогорова-Смирнова и ω^2 (по Ходжесу-Леману) [113] предопределяют необходимость поиска предельных законов распределения непараметрических статистик в зависимости от того, сколько параметров оценивалось по выборке и с каким законом проверяется согласие.

Для случая нормального распределения предельные распределения статистики критерия ω^2 Мизеса при оценивании одного из двух или обоих параметров подробно исследованы в [103], где приведены их таблицы.

Но очевидно, что теоретически найти решение задачи определения предельных распределений непараметрических статистик для множества законов, используемых для описания реальных величин, очень сложно. Именно поэтому большинство существующих таблиц и результатов, связанных с предельными распределениями, получены методом статистического моделирования [190,188,240,2241,256,261-263,209,278]. Таким образом получены, например, таблицы процентных точек для модифицированных статистик критериев Колмогорова (типа Колмогорова) при проверке нормальности и оценивании одного или двух параметров нормального распределения и процентные точки при проверке экспоненциальности и оценивании его масштабного параметра, представленные в [256].

В работах [112,171-173] разработан аналитический метод расчета процентных точек для статистик Колмогорова-Смирнова для регулярных семейств распределений и больших выборок, получены формулы для приближенного вычисления вероятностей вида $P\{S > S^*\}$. С помощью этих приближенных формул, учитывающих отличие предельных распределений непараметрических статистик при оценивании параметров законов от классических, вычисляют вероятности вида $P\{S > S^*\}$ в пакете STADIA [174].

Нам наиболее реальный выход видится в моделировании предельных законов распределения статистик непараметрических критериев и в последующей идентификации полученных эмпирических законов. Под идентификацией понимается выбор такого теоретического закона распределения, который наиболее хорошо согласуется с эмпирической функцией распределения. Результаты такого моделирования позволяют, с одной стороны, количественно оценить возможные ошибки в случае пренебрежения фактом оценивания по выборке параметров наблюдаемого закона, с другой стороны, дают возможность с достаточной для практики точностью использовать эти результаты при проверке гипотез.

Реализованный алгоритм численного моделирования выглядит следующим образом.

1. Моделируется (имитируется) n выборок заданного объема N в соответствии с заданным законом распределения $f(x, \theta)$.
2. По каждой выборке оцениваются параметры этого закона распределения (один параметр, или другой, или два одновременно).

3. Вычисляются значения соответствующих непараметрических статистик.
4. По накопленной выборке значений статистики объема n анализируется полученное эмпирическое распределение статистики и осуществляется его идентификация.

Конечной целью настоящего исследования явилось стремление по возможности найти такие законы распределения вероятностей, которые с практической точки зрения хорошо аппроксимируют предельные распределения статистик непараметрических критериев в тех случаях, когда по выборке оцениваются параметры.

При постановке задачи исследования планировалось осуществить моделирование для различных объемов N и n и, в случае устойчивости полученных результатов, использовать их в программном обеспечении статистического анализа [76], а также протабулировать полученные эмпирические законы и критические значения статистик, если не удастся подобрать подходящего теоретического закона распределения. Подобную процедуру планировалось осуществить для каждого закона распределения, включенного в программную систему статистического анализа [76].

В данном разделе приводятся некоторые результаты моделирования и анализа, которые, с нашей точки зрения, могут с успехом применяться при решении практических задач проверки гипотез о согласии с использованием непараметрических критериев после вычисления оценок параметров распределения по той же выборке [95,97].

Отметим, что результаты моделирования классических непараметрических статистик (без оценивания параметров) показали хорошее согласие получившихся эмпирических распределений с предельными теоретическими законами. Это позволяет судить о точности описания предельных законов полученными эмпирическими распределениями статистик для ситуации с оцениванием параметров, а также служит косвенным подтверждением возможности использования полученных распределений соответствующих статистик при решении практических задач статистического анализа.

Статистика Колмогорова (Колмогорова-Смирнова) определяется выражением [19]

$$S_k = \frac{(6ND_N + 1)^2}{18N},$$

статистика Смирнова [19]

$$S_m = \frac{(6ND_N^+ + 1)^2}{9N},$$

где

$$D_N = \max(D_N^+, D_N^-), \quad D_N^+ = \max_{1 \leq i \leq N} \left\{ \frac{i}{N} - F(x_i) \right\}, \quad D_N^- = \max_{1 \leq i \leq N} \left\{ F(x_i) - \frac{i-1}{N} \right\},$$

N - объем выборки, $x_1, x_2, \dots, x_N$ - упорядоченные по возрастанию выборочные значения, $F(x)$ - функция распределения, согласие с которой проверяется.

Распределение величины $\sqrt{S_k/2}$, если по выборке не оценивались параметры, в пределе подчиняется закону Колмогорова с функцией распределения $K(x)$ [19,61].

Гипотеза о согласии не отвергается, если

$$P\{S_k > S_k^*\} = 1 - K\left(\sqrt{\frac{S_k^*}{2}}\right) > \alpha,$$

где α - задаваемый уровень значимости.

В аналогичной ситуации статистика Смирнова S_m подчиняется в пределе распределению χ^2 с числом степеней свободы, равным 2. Гипотеза о согласии не отвергается, если

$$P\{S_m > S_m^*\} = \int_{S_m^*}^{\infty} \frac{1}{2} e^{-x/2} dx = e^{-S_m^*/2} > \alpha.$$

Статистики Мизеса имеют вид [19]

$$S_\omega = N\omega_N^2 = \frac{1}{12N} + \sum_{i=1}^N \left\{ F(x_i) - \frac{2i-1}{2N} \right\}^2,$$

которую чаще называют статистикой Крамера-Мизеса-Смирнова, и

$$S_\Omega = N\Omega_N^2 = -N - 2 \sum_{i=1}^N \left\{ \frac{2i-1}{2N} \ln F(x_i) + \left(1 - \frac{2i-1}{2N} \right) \ln(1 - F(x_i)) \right\},$$

называемую еще статистикой Андерсона-Дарлинга. Для этих статистик также известны предельные распределения вероятностей [19]

$$\lim_{N \rightarrow \infty} P\{N\omega_N^2 < x\} = a1(x),$$

$$\lim_{N \rightarrow \infty} P\{N\Omega_N^2 < x\} = a2(x),$$

$$a1(x) = \frac{1}{\sqrt{2x}} \sum_{j=0}^{\infty} \frac{\Gamma(j+1/2)\sqrt{4j+1}}{\Gamma(1/2)\Gamma(j+1)} \exp\left\{-\frac{(4j+1)^2}{16x}\right\} \times \\ \times \left\{ I_{-\frac{1}{4}}\left[\frac{(4j+1)^2}{16x}\right] - I_{\frac{1}{4}}\left[\frac{(4j+1)^2}{16x}\right] \right\},$$

где $I_{-\frac{1}{4}}(\cdot)$, $I_{\frac{1}{4}}(\cdot)$ - модифицированные функции Бесселя,

$$I_{\nu}(z) = \sum_{k=0}^{\infty} \frac{\left(\frac{z}{2}\right)^{\nu+2k}}{\Gamma(k+1)\Gamma(k+\nu+1)}, \quad |z| < \infty, \quad |\arg z| < \pi;$$

$$a2(x) = \frac{\sqrt{2\pi}}{x} \sum_{j=0}^{\infty} (-1)^j \frac{\Gamma(j+1/2)(4j+1)}{\Gamma(1/2)\Gamma(j+1)} \exp\left\{-\frac{(4j+1)^2 \pi^2}{8x}\right\} \times \\ \times \int_0^{\infty} \exp\left\{\frac{x}{8(y^2+1)} - \frac{(4j+1)^2 \pi^2 y^2}{8x}\right\} dy.$$

Гипотезы о согласии не отвергаются, если выполняются неравенства

$$P\{S_{\omega} > S_{\omega}^*\} = 1 - a1(S_{\omega}^*) > \alpha \quad \text{и} \quad P\{S_{\Omega} > S_{\Omega}^*\} = 1 - a2(S_{\Omega}^*) > \alpha.$$

В соответствии с изложенным алгоритмом моделировались выборки значений статистик $\sqrt{S_k/2}$, S_m , S_{ω} , S_{Ω} .

По каждому закону распределения $f(x, \theta)$ моделировалась серия из $n=2000$ выборок объемом $N=200$. Псевдослучайная величина, принадлежащая закону с функцией распределения $F(x, \theta)$, имитировалась по методу обратных функций, при котором случайная величина X , подчиняющаяся закону с функцией распределения $F(\theta, x)$, получается в соответствии с соотношением $X = F^{-1}(\theta, Y)$, где $F^{-1}(\cdot)$ - функция, обратная к $F(\cdot)$, а Y - случайная величина, равномерно распределённая на интервале $[0,1]$. В качестве датчика равномерно распределённых псевдослучайных чисел использовался стандартный датчик, реализованный в C++. Оценки параметров находились по методу максимального правдоподобия по негруппированным данным. В данном случае исследовались распределения непараметрических статистик, когда наблюдаемые величины распределены в соответствии с законами, представленными в табл. 4.4.1.

Таблица 4.4.1

№ п/п	Распределение случайной величины	Функция плотности
1.	Экспоненциальное	$\theta_0 e^{-\theta_0 x}$
2.	Полунормальное	$\frac{2}{\theta_0 \sqrt{2\pi}} e^{-x^2/2\theta_0^2}$
3.	Рэлея	$\frac{x}{\theta_0^2} e^{-x^2/2\theta_0^2}$
4.	Максвелла	$\frac{2x^2}{\theta_0^3 \sqrt{2\pi}} e^{-x^2/2\theta_0^2}$
5.	Лапласа	$\frac{\theta_0}{2} e^{-\theta_0 x - \theta_1 }$
6.	Нормальное	$\frac{1}{\theta_0 \sqrt{2\pi}} e^{-(x - \theta_1)^2/2\theta_0^2}$
7.	Логнормальное	$\frac{1}{x \theta_0 \sqrt{2\pi}} e^{-(\ln x - \theta_1)^2/2\theta_0^2}$
8.	Коши	$\frac{\theta_0}{\pi [\theta_0^2 + (x - \theta_1)^2]}$
9.	Логистическое	$\frac{\pi}{\theta_0 \sqrt{3}} \exp\left\{-\frac{\pi(x - \theta_1)}{\theta_0 \sqrt{3}}\right\} / \left[1 + \exp\left\{-\frac{\pi(x - \theta_1)}{\theta_0 \sqrt{3}}\right\}\right]^2$
10.	Наибольшего значения	$\frac{1}{\theta_0} \exp\left\{-\frac{x - \theta_1}{\theta_0} - \exp\left(-\frac{x - \theta_1}{\theta_0}\right)\right\}$
11.	Наименьшего значения	$\frac{1}{\theta_0} \exp\left\{\frac{x - \theta_1}{\theta_0} - \exp\left(\frac{x - \theta_1}{\theta_0}\right)\right\}$
12.	Вейбулла- Гнеденко	$\frac{\theta_0 x^{\theta_0 - 1}}{\theta_1^{\theta_0}} \exp\left\{-\left(\frac{x}{\theta_1}\right)^{\theta_0}\right\}$

4.4.2. Моделирование и идентификация законов распределения статистики типа Колмогорова

На рис. 4.4.1 приведены результаты моделирования величины $\sqrt{S_k/2}$, используемой в критерии Колмогорова, при проверке гипотез о согласии с нормальным распределением при справедливости гипотезы H_0 . На этом и последующих рисунках представлены эмпирические функции распределения статистики, когда по выборке не оценивались параметры (“1”), по выборке оценивался только масштабный параметр (“2”) (в данном случае θ_0), оценивался только параметр сдвига (“3”) (в данном случае θ_1), оценивались одновременно оба параметра (“4”). Здесь же приведена функция распределения Колмогорова (“0”), которому подчиняется статистика $\sqrt{S_k/2}$, если по выборке не оцениваются параметры. Результаты проверки согласия эмпирического распределения “1” с распределением Колмогорова “0” достаточно хорошее. В то же время весьма наглядно отличие эмпирических функций распределения “2”, “3”, “4” от распределения Колмогорова “0”. Это отличие позволяет судить о величине тех ошибок, которые мы допускаем, не учитывая факта оценивания параметров конкретного распределения при использовании критерия Колмогорова. Крестиками, лежащими практически на эмпирическом распределении “4”, обозначены значения, соответствующие процентным точкам при проверке нормальности, взятые из таблицы М.Стефенса [256], а кружочками - значения, полученные по приближенным формулам Ю.Н.Тюрина [171,173] для расчетов вероятностей, соответствующих распределениям “4” и “2”.

Замечание: Процентные точки М.Стефенса соответствуют модифицированной статистике вида $D_N \left(\sqrt{N} - 0.001 + \frac{0.85}{\sqrt{N}} \right)$, а по формулам Ю.Н.Тюрина рассчитывают вероятности для статистики $\sqrt{N}D_N$. В то же время

$$\sqrt{S_k/2} = \sqrt{N}D_N + \frac{1}{6\sqrt{N}}.$$

Аналогичная картина распределения статистики Колмогорова при справедливой гипотезе H_0 наблюдается для логистического распределения на рис. 4.4.2, для распределения Лапласа на рис. 4.4.3, для распределения Коши на рис. 4.4.4 и для экспоненциального рас-

пределения на рис. 4.4.5, где крестиками, обозначены значения, соответствующие процентным точкам при проверке экспоненциальности и неизвестном масштабном параметре, взятые из таблицы процентных точек [256].

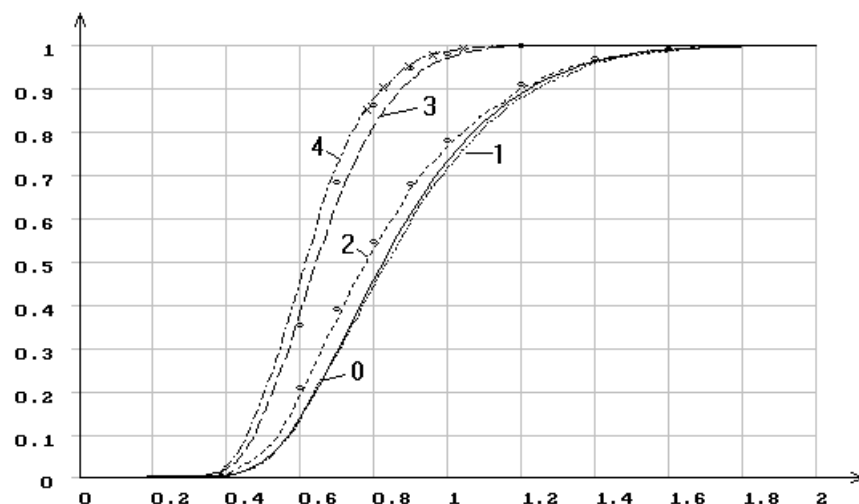


Рис.4.4.1. Эмпирические функции распределения статистики $\sqrt{S_k}/2$

Колмогорова при различном количестве оцениваемых параметров нормального закона: 0 - функция распределения Колмогорова; 1 - по выборке не оценивались параметры; 2 - по выборке оценивался только масштабный параметр θ_0 ; 3 - оценивался только параметр сдвига θ_1 ; 4 - оценивались одновременно оба параметра.

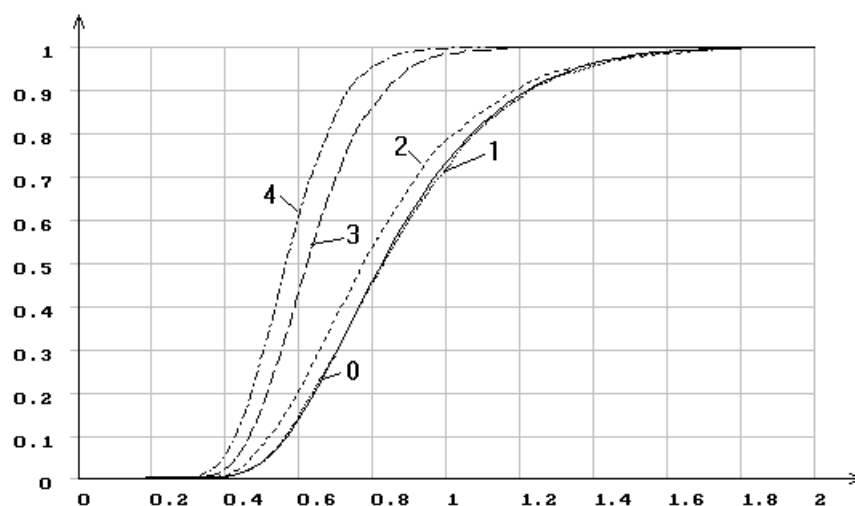


Рис. 4.4.2. Эмпирические функции распределения статистики $\sqrt{S_k}/2$

Колмогорова при различном количестве оцениваемых параметров логистического распределения

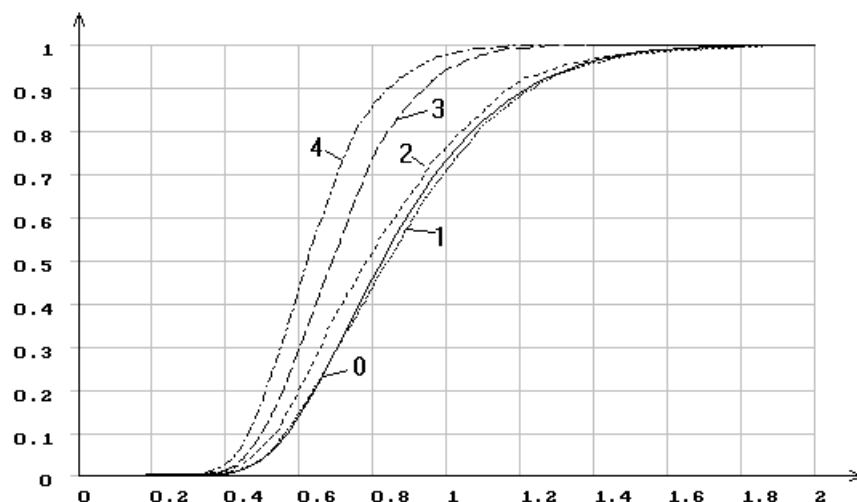


Рис. 4.4.3. Эмпирические функции распределения статистики $\sqrt{S_k}/2$ Колмогорова при различном количестве оцениваемых параметров распределения Лапласа

Результаты моделирования однозначно указывают на то, что предельные распределения статистики $\sqrt{S_k}/2$ при условии оценивания параметров конкретного закона настолько сильно отличаются от распределения Колмогорова, что использование последнего никак не может быть оправдано из-за высокого риска неверных выводов.

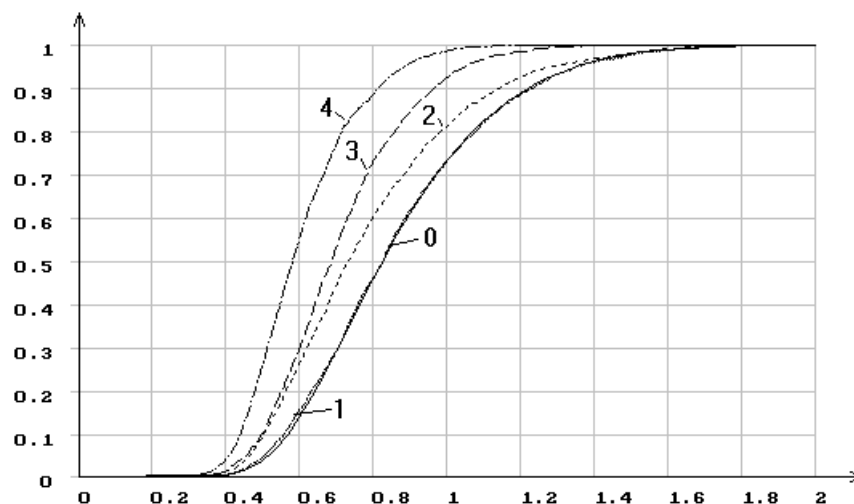


Рис. 4.4.4. Эмпирические функции распределения статистики $\sqrt{S_k}/2$ Колмогорова при различном количестве оцениваемых параметров распределения Коши

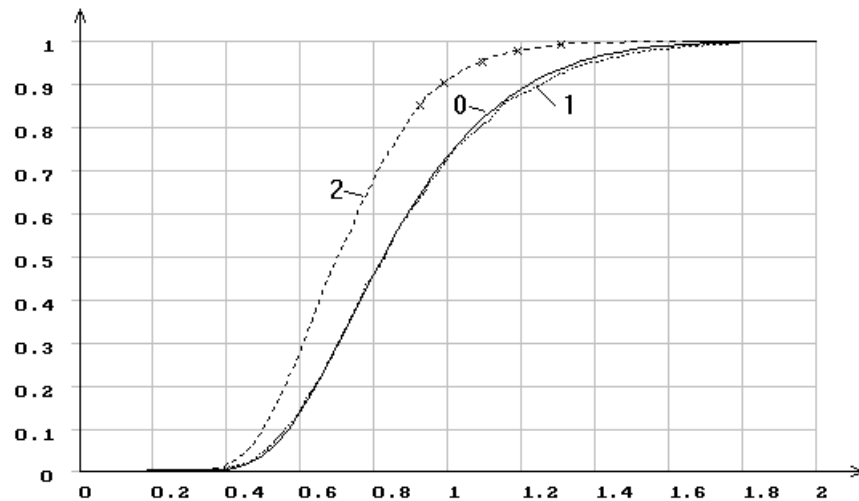


Рис. 4.4.5. Эмпирические функции распределения статистики $\sqrt{S_k}/2$

Колмогорова при различном количестве оцениваемых параметров экспоненциального закона: 0 - функция распределения Колмогорова; 1 - по выборке не оценивались параметры; 2 - по выборке оценивался масштабный параметр.

При идентификации типов предельных законов распределения непараметрических статистик в зависимости от вида закона наблюдаемой случайной величины и количества оцениваемых по выборке параметров использовалось множество из 26 законов и семейств распределений, включенных в программную систему [76]. Оказалось, что почти всегда эмпирические законы распределения статистик непараметрических критериев наиболее хорошо описываются одним из двух законов распределения: *логарифмически нормальным* или *гамма-распределением*. То есть, эмпирическая функция статистики хорошо согласуется с одним из этих законов по всем используемым в системе критериям согласия: отношения правдоподобия, χ^2 Пирсона, Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса.

На рис. 4.4.6 представлены результаты выравнивания распределения статистики Колмогорова при оценивании одновременно двух параметров нормального распределения. На рисунке приведены эмпирическая функция распределения статистики и функция распределения логарифмически нормального распределения с параметрами $\theta_1 = -0.4849$, $\theta_0 = 0.2254$ (они практически совпали). Здесь же отражены значения статистик всех используемых при проверке согласия критериев и соответствующие вероятности вида $P\{S > S^*\}$. Факт оценивания параметров логарифмически нормального распределения в критериях отношения правдоподобия и χ^2 Пирсона

учитывается уменьшением числа степеней свободы χ^2 -распределения. Причем в критериях использовано асимптотически оптимальное группирование наблюдений при 15 интервалах, максимизирующее мощность при близких альтернативах. В непараметрических критериях факт оценивания параметров не учитывался. Если использовать представленные ниже в табл. 4.4.2, 4.4.5, 4.4.7, 4.4.9 результаты по аппроксимации предельных распределений статистик, то вероятность вида $P\{S > S^*\}$ для критерия Колмогорова составит 0.6067, а не 0.8905, для критерия Смирнова - 0.4416, а не 0.5111, для критерия ω^2 Мизеса - 0.6277, а не 0.8897, для критерия Ω^2 Мизеса - 0.4737, а не 0.8397.

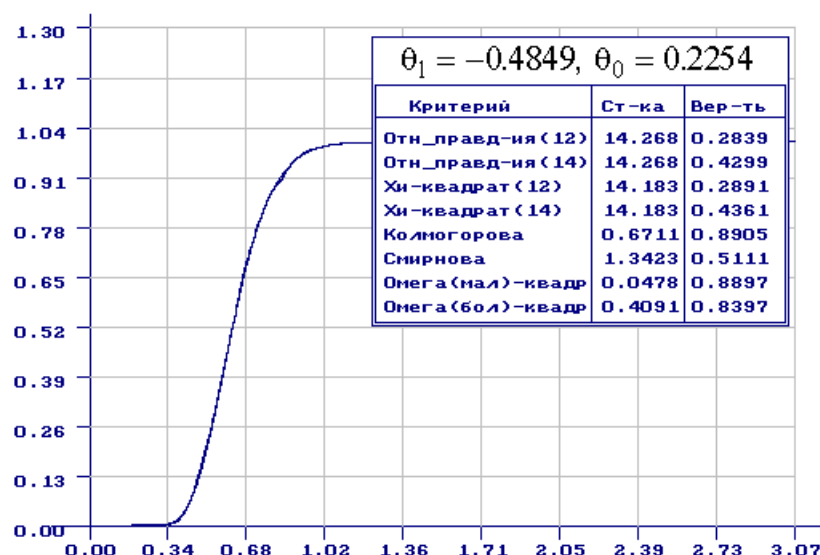


Рис. 4.4.6. Эмпирическая функция распределения статистики Колмогорова при справедливой гипотезе H_0 и оценивании 2-х параметров нормального распределения и выравнивающая её функция распределения логарифмически нормального распределения

В табл. 4.4.2 сведены результаты идентификации законов для статистики критерия Колмогорова. Указание в клетке на конкретное распределение означает, что выборка соответствующей статистики хорошо описывается данным законом (согласуется с законом). В случае если согласие с каким-то законом не очень хорошее (гипотеза о согласии принимается с уровнем значимости $\alpha < 0.1$), то соответствующий закон указан на сером фоне. В таблицах через $\ln N(\theta_1, \theta_0)$ обозначено логарифмически нормальное распределение с функцией плотности

$$f(x) = \frac{1}{x\theta_0\sqrt{2\pi}} e^{-(\ln x - \theta_1)^2 / 2\theta_0^2},$$

через $\gamma(\theta_0, \theta_1, \theta_2)$ - гамма-распределение с функцией плотности

$$f(x) = \frac{\theta_1^{\theta_0}}{\Gamma(\theta_0)} (x - \theta_2)^{\theta_0 - 1} e^{-\theta_1(x - \theta_2)}.$$

Таблица 4.4.2.

<i>Предельные распределения статистики Колмогорова</i>				
№ п/п	Распределение случайной величины	Оценивался только масштабный параметр	Оценивался только параметр сдвига	Оценивалось два параметра
1.	Экспоненциальное	$\ln N(-0.3477, 0.2638)$		
2.	Полунормальное	$\gamma(3.4090, 8.2385, 0.3443)$		
3.	Рэлея	$\ln N(-0.3366, 0.2579)$		
4.	Максвелла	$\gamma(3.4809, 9.0801, 0.3417)$		
5.	Лапласа	$\gamma(3.2121, 6.5137, 0.3400)$	$\ln N(-0.3721, 0.2426)$	$\ln N(-0.4679, 0.2329)$
6.	Нормальное	$\gamma(3.6448, 7.0208, 0.3164)$	$\ln N(-0.4349, 0.2337)$	$\ln N(-0.4849, 0.2254)$
7.	Логнормальное	$\gamma(3.7326, 7.4146, 0.3265)$	$\gamma(3.0857, 8.4464, 0.3532)$	$\gamma(3.7311, 10.045, 0.3062)$
8.	Коши	$\gamma(3.1388, 6.7500, 0.3261)$	$\ln N(-0.3691, 0.2542)$	$\gamma(4.2049, 13.595, 0.2983)$
9.	Логистическое	$\gamma(3.3283, 6.6563, 0.3280)$	$\ln N(-0.4681, 0.2248)$	$\ln N(-0.5684, 0.2111)$
10.	Наибольшего значения	$\gamma(3.5340, 6.9273, 0.3302)$	$\ln N(-0.3588, 0.2555)$	$\ln N(-0.4950, 0.2231)$
11.	Наименьшего значения	$\gamma(3.4746, 6.9386, 0.3362)$	$\gamma(3.7737, 9.3627, 0.3293)$	$\ln N(-0.5009, 0.2221)$
12.	Вейбулла	$\gamma(3.6550, 9.0285, 0.3282)$	$\ln N(-0.2325, 0.2974)$ **	$\ln N(-0.5026, 0.2202)$

** - оценивался параметр формы распределения Вейбулла.

Отметим, что предельные распределения статистики Колмогорова для выборок из распределения Вейбулла-Гнеденко, как это и доказано в [172], не зависят от параметров. В случае распределений экстремальных значений и распределения Вейбулла-Гнеденко могут использоваться одни и те же предельные распределения.

Посмотрим, что будет получаться, если, например, мы будем использовать распределение Колмогорова для вычисления вероятности вида $P\{S > S^*\}$ в случае, когда по выборке предварительно вычисляются оценки параметров нормального распределения. Распределение статистики $\sqrt{S_k}/2$ в этом случае хорошо описывается логарифмически нормальным $\ln N(-0.4849, 0.2254)$. Для сравнения в табл. 4.4.3 для некоторых значений статистик S^* представлены соответствующие значения функции распределения Колмогорова, данного логарифмически нормального распределения и эмпирической функции распределения $F(\sqrt{S}/2)$, полученной при моделировании распределения

статистики. Так, для распределения Колмогорова величина $P\{\sqrt{S^2/2} > 0.9\} = 0.392731$, а для логарифмически нормального $P\{\sqrt{S^2/2} > 0.9\} = 0.04612$. Это означает, что при значении статистики $\sqrt{S/2} = 0.9$, используя распределение Колмогорова, мы безоговорочно примем гипотезу H_0 , когда на самом деле даже при уровне значимости $\alpha = 0.05$ она должна быть отклонена.

Таблица 4.4.3

$\sqrt{S/2}$	0.5	0.9	1.0	1.22	1.36
$K(\sqrt{S/2})$	0.036055	0.607269	0.730000	0.898102	0.990154
$\ln N(\sqrt{S/2})$	0.177708	0.953880	0.984268	0.998791	0.999780
$F(\sqrt{S/2})$	0.19	0.9533	0.98425	0.999	1

Распределения, приводимые в табл. 4.4.2, удобны для реализации в программном обеспечении задач статистического анализа, как это сделано в [21], так как для вычисления вероятностей вида $P\{S > S^*\}$, соответствующих логарифмически нормальному или гамма-распределениям, не обойтись без процедур численного интегрирования или использования приближенных формул. В то же время для большинства практиков, применяющих статистические методы в различных приложениях, более привычно опираться на верхние процентные точки распределений статистик. В табл. 4.4.4 приводятся приближенные значения верхних процентных точек для статистики $\sqrt{S_k/2}$, которые для более высокой точности рассчитаны не по распределениям табл. 4.4.2, а получены сглаживанием “хвостов” эмпирических распределений. Точность этих процентных точек вполне достаточна для корректного использования критерия в задачах статистического анализа.

Таблица 4.4.4

<i>Процентные точки распределения статистики Колмогорова</i>							
№ п/п	Распределение случайной величины	Оцениваемые параметры	Верхние процентные точки				
			0.15	0.1	0.05	0.025	0.01
1.	Экспоненциальное	масштабный	0.9291	0.9872	1.0861	1.1846	1.3145
2.	Полунормальное	масштабный	0.9879	1.0519	1.1539	1.2511	1.3752
3.	Рэля	масштабный	0.9402	0.9999	1.0952	1.1859	1.3017
4.	Максвелла	масштабный	0.9284	0.9890	1.0853	1.1770	1.2938
5.	Лапласа	масштабный	1.1081	1.1897	1.3222	1.4501	1.6147
		сдвиг	0.8914	0.9435	1.0240	1.0992	1.1935
		два параметра	0.7966	0.8467	0.9261	1.0016	1.0978
6.	Нормальное	масштабный	1.1208	1.2081	1.3446	1.4731	1.6356
		сдвиг	0.8330	0.8790	0.9497	1.0156	1.0982
		два параметра	0.7808	0.8255	0.8954	0.9611	1.0442
7.	Логнормальное	масштабный	1.0880	1.1736	1.3147	1.4523	1.6308
		сдвиг	0.9147	0.9875	1.1083	1.2266	1.3805
		два параметра	0.8539	0.9268	1.0500	1.1723	1.3330
8.	Коши	масштабный	1.0590	1.1497	1.2950	1.4339	1.6116
		сдвиг	0.9080	0.9659	1.0589	1.1481	1.2623
		два параметра	0.7620	0.8117	0.8868	0.9557	1.0414
9.	Логистическое	масштабный	1.1034	1.1957	1.3441	1.4864	1.6689
		сдвиг	0.7912	0.8373	0.9109	0.9813	1.0714
		два параметра	0.7060	0.7400	0.7964	0.8516	0.9234
10.	Наибольшего значения	масштабный	1.1157	1.2033	1.3402	1.4689	1.6315
		сдвиг	0.9137	0.9775	1.0794	1.1765	1.3004
		два параметра	0.7705	0.8119	0.8808	0.9485	1.0367
11.	Наименьшего значения	масштабный	1.1027	1.1897	1.3286	1.4612	1.6305
		сдвиг	0.9386	1.0048	1.1103	1.2108	1.3390
		два параметра	0.7655	0.8080	0.8758	0.9405	1.0233
12.	Вейбулла	формы	1.0824	1.1659	1.3043	1.4396	1.6157
		масштаба	0.9425	1.0064	1.1087	1.2065	1.3315
		два параметра	0.7634	0.8022	0.8658	0.9276	1.0074

4.4.3. Моделирование и идентификация законов распределения статистики типа Смирнова

На рис. 4.4.7 представлены результаты моделирования статистики Смирнова S_m при справедливой гипотезе H_0 , соответствующей нормальному распределению, на рис. 4.4.8, соответствующей логистическому распределению, на рис. 4.4.9, соответствующей распределению Лапласа, на рис. 4.4.10. - распределению Коши, а на рис. 4.4.11 - экспоненциальному распределению. На этих рисунках символом “0” отмечена функция распределения χ^2_2 , которому подчиняется статистика Смирнова, если по выборке не оцениваются параметры.

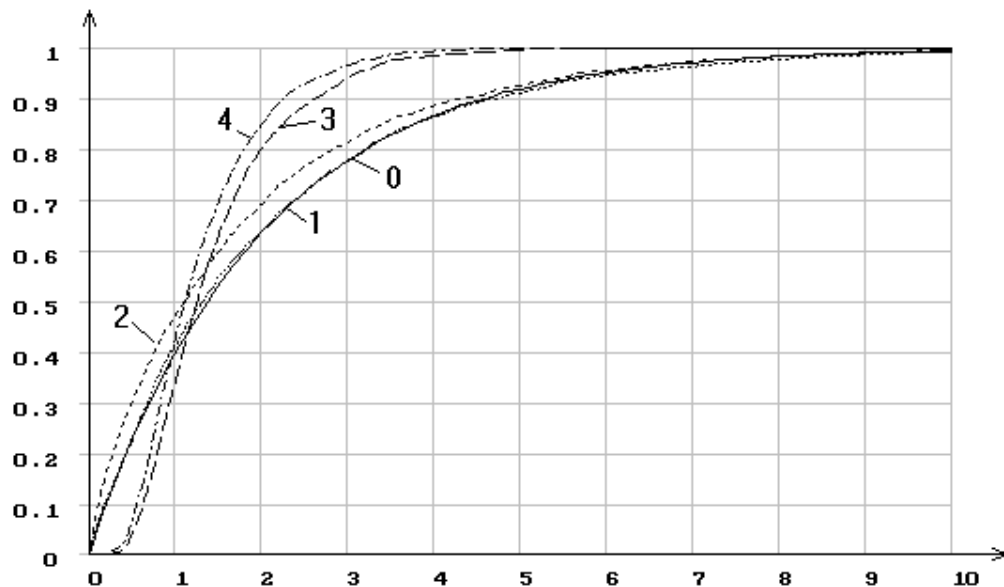


Рис. 4.4.7. Эмпирические функции распределения статистики Смирнова при различном количестве оцениваемых параметров нормального закона: 0 - функция распределения χ^2_2 (с двумя степенями свободы); 1 - по выборке не оценивались параметры; 2 - по выборке оценивался только масштабный параметр θ_0 ; 3 - оценивался только параметр сдвига θ_1 ; 4 - оценивались одновременно оба параметра

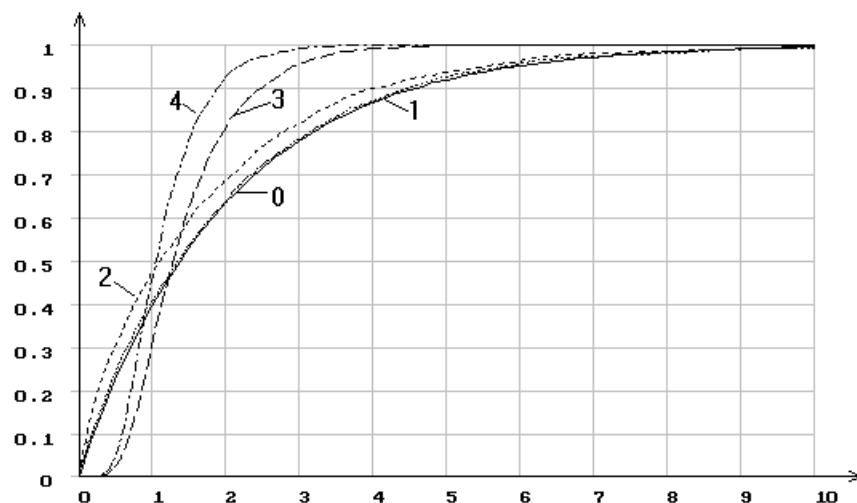


Рис. 4.4.8. Эмпирические функции распределения статистики S_m Смирнова при оценивании параметров логистического распределения

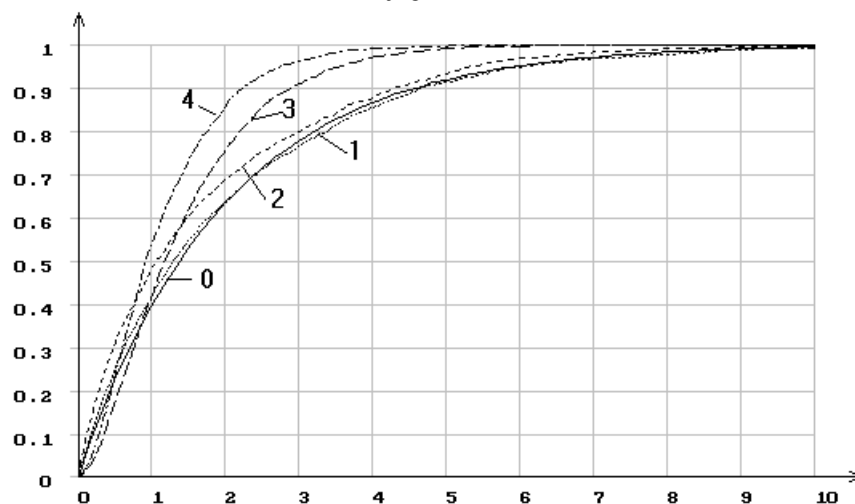


Рис. 4.4.9. Эмпирические функции распределения статистики S_m Смирнова при оценивании параметров распределения Лапласа

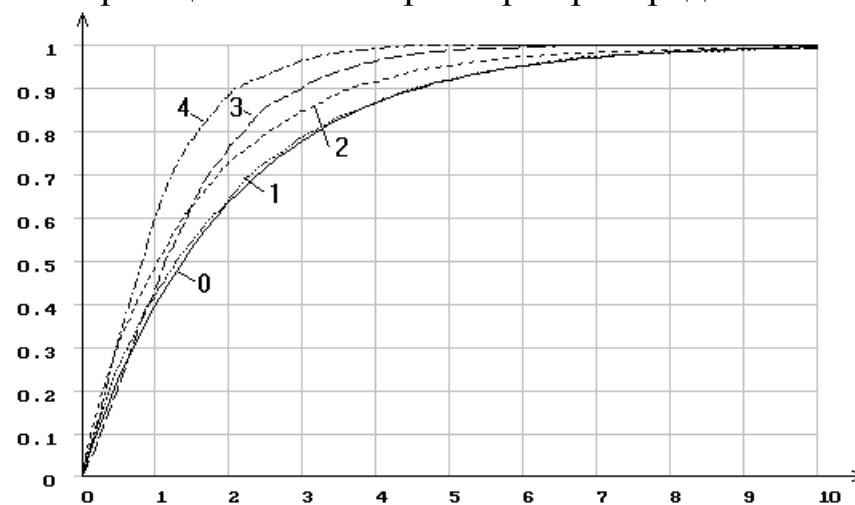


Рис. 4.4.10. Эмпирические функции распределения статистики S_m Смирнова при оценивании параметров распределения Коши

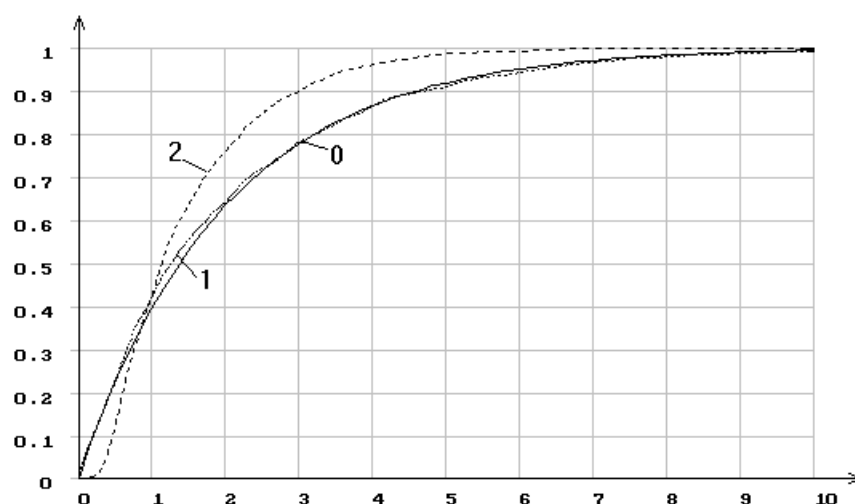


Рис. 4.4.11. Эмпирические функции распределения статистики S_m Смирнова без оценивания параметров (1), при оценивании масштабного параметра (2) экспоненциального распределения

Как видим, и в данном случае предельные распределения статистики Смирнова при условии оценивания параметров конкретного закона кардинально отличаются от распределения χ^2_2 .

В табл. 4.4.5 сведены результаты идентификации предельных законов для статистики критерия Смирнова.

Таблица 4.4.5.

<i>Предельные распределения статистики Смирнова</i>				
№ п/п	Распределение случайной величины	Оценивался только масштабный параметр	Оценивался только параметр сдвига	Оценивалось два параметра
1.	Экспоненциальное	$\ln N(0.1585, 0.7009)$		
2.	Полунормальное	$\ln N(0.1289, 0.7900)$		
3.	Рэлея	$\ln N(0.1936, 0.7073)$		
4.	Максвелла	$\ln N(0.2221, 0.6794)$		
5.	Лапласа	$\gamma(0.8146, 0.4654, 0.0006)$	$\gamma(1.7664, 1.2256, 0.0207)$	$\gamma(1.8235, 1.5842, 0.0058)$
6.	Нормальное	$\gamma(0.8088, 0.4549, 0.0006)$	$\ln N(0.2471, 0.5321)$	$\ln N(0.1299, 0.5331)$
7.	Логнормальное	$\gamma(0.8391, 0.4641, 0.0006)$	$\ln N(0.4252, 0.6481)$	$\ln N(0.1947, 0.6783)$
8.	Коши	$\gamma(0.8570, 0.5348, 0.0006)$	$\gamma(1.4215, 0.9846, 0.0006)$	$\gamma(1.2931, 1.2542, 0.0006)$
9.	Логистическое	$\gamma(0.8164, 0.4709, 0.0)$	$\ln N(0.2684, 0.4856)$	$\ln N(0.0569, 0.4491)$
10.	Наибольшего значения	$\gamma(0.8507, 0.4682, 0.0006)$	$\ln N(0.2667, 0.6821)$	$\ln N(0.1049, 0.5058)$
11.	Наименьшего значения	$\gamma(0.8234, 0.4592, 0.0006)$	$\ln N(0.2022, 0.7015)$	$\ln N(0.1238, 0.5024)$
12.	Вейбулла	$\ln N(0.1615, 0.7186)$	$\gamma(0.8641, 0.4851, 0.0006)$ **	$\gamma(2.3495, 2.3008, 0.2564)$

** - оценивался параметр формы распределения Вейбулла.

В табл. 4.4.6 приводятся приближенные значения верхних процентных точек для статистики S_m .

Таблица 4.4.6.

<i>Процентные точки распределения статистики Смирнова</i>							
№ п/п	Распределение случайной величины	Оцениваемые параметры	Верхние процентные точки				
			0.15	0.1	0.05	0.025	0.01
1.	Экспоненциальное	масштабный	2.5341	2.9847	3.7675	4.5593	5.6145
2.	Полунормальное	масштабный	2.6856	3.2651	4.2439	5.2147	6.4902
3.	Рэля	масштабный	2.6590	3.1390	3.9317	4.7058	5.7115
4.	Максвелла	масштабный	2.5751	3.0644	3.9162	4.7788	5.9296
5.	Лапласа	масштабный	3.5933	4.3253	5.5464	6.7473	8.3154
		сдвиг	2.5139	2.9110	3.5728	4.2230	5.0716
		два параметра	1.9821	2.2836	2.8034	3.3261	4.0201
6.	Нормальное	масштабный	3.4951	4.3824	5.8586	7.3074	9.1966
		сдвиг	2.2762	2.6147	3.1684	3.7055	4.3998
		два параметра	2.0255	2.3001	2.7552	3.2005	3.7801
7.	Логнормальное	масштабный	3.4308	4.2549	5.6948	7.1566	9.1105
		сдвиг	3.0417	3.6974	4.8495	6.0236	7.5972
		два параметра	2.3621	2.9020	3.9162	4.9983	6.4985
8.	Коши	масштабный	3.0648	3.7974	5.0833	6.3929	8.1473
		сдвиг	2.5211	2.9562	3.7225	4.5046	5.5542
		два параметра	1.8510	2.1830	2.7286	3.2596	3.9478
9.	Логистическое	масштабный	3.3625	4.1502	5.5238	6.9163	8.7754
		сдвиг	2.1914	2.4814	2.9781	3.4754	4.1334
		два параметра	1.6973	1.9035	2.2459	2.5814	3.0185
10.	Наибольшего значения	масштабный	3.5767	4.3936	5.7430	7.0609	8.7735
		сдвиг	2.6771	3.1331	3.9274	4.7318	5.8051
		два параметра	1.8947	2.1702	2.6283	3.0778	3.6638
11.	Наименьшего значения	масштабный	3.5733	4.3186	5.5688	6.8028	8.4186
		сдвиг	2.6298	3.1309	3.9946	4.8630	6.0157
		два параметра	1.9043	2.1615	2.6199	3.0917	3.7285
12.	Вейбулла	формы	3.4390	4.1895	5.5179	6.8786	8.7090
		масштаба	2.5623	3.0621	3.9347	4.8203	6.0035
		два параметра	1.9035	2.1477	2.5708	2.9977	3.5659

4.4.4. Моделирование и идентификация законов распределения статистики Крамера-Мизеса-Смирнова

Результаты моделирования статистики S_{ω} Мизеса (Крамера-Мизеса-Смирнова) при справедливой гипотезе H_0 для случая нормального распределения представлены на рис. 4.4.12. На этом рисунке крестиками нанесены соответствующие значения предельных распределений, полученные в работе Г.В.Мартынова [103]. Для случая логистического распределения результаты моделирования приведены на рис. 4.4.13, для случая распределения Лапласа - на рис. 4.4.14, для распределения Коши - на рис. 4.4.15, а на рис. 4.4.16 - для экспоненциального распределения, где для сравнения крестиками обозначены процентные точки при проверке экспоненциальности из таблицы М.Стефенса [256]. На этих рисунках символом “0” отмечена функция распределения $a1(s)$ [19], предельная для этой статистики, если по выборке не оценивались параметры распределения.

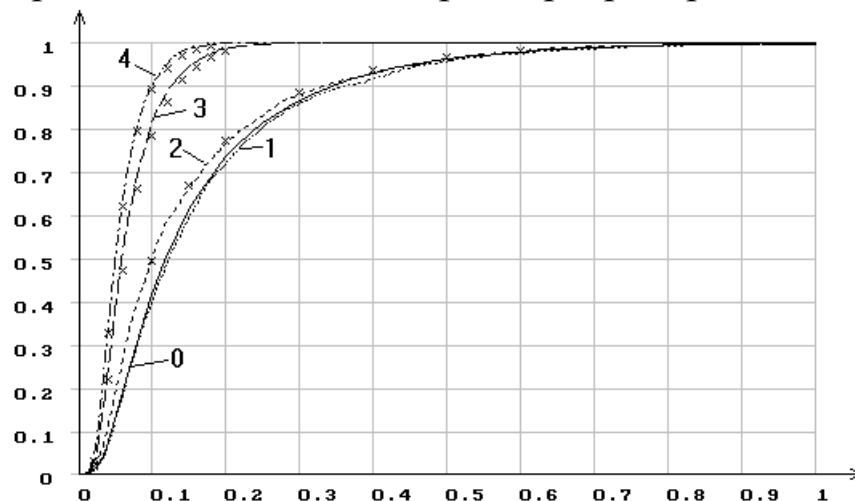


Рис. 4.4.12. Эмпирические функции распределения статистики S_{ω} Мизеса при различном количестве оцениваемых параметров нормального закона: 0 - функция распределения $a1(s)$; 1 - по выборке не оценивались параметры; 2 - по выборке оценивался только масштабный параметр θ_0 ; 3 - оценивался только параметр сдвига θ_1 ; 4 - оценивались оба параметра.

В табл. 4.4.7 сведены результаты идентификации законов для статистики критерия ω^2 Мизеса. Следует отметить, что в данном случае предельные распределения хуже аппроксимируются логарифмически нормальным, которое в каждом случае оказалось наиболее подходящим из всего множества законов, включенных в систему [76].

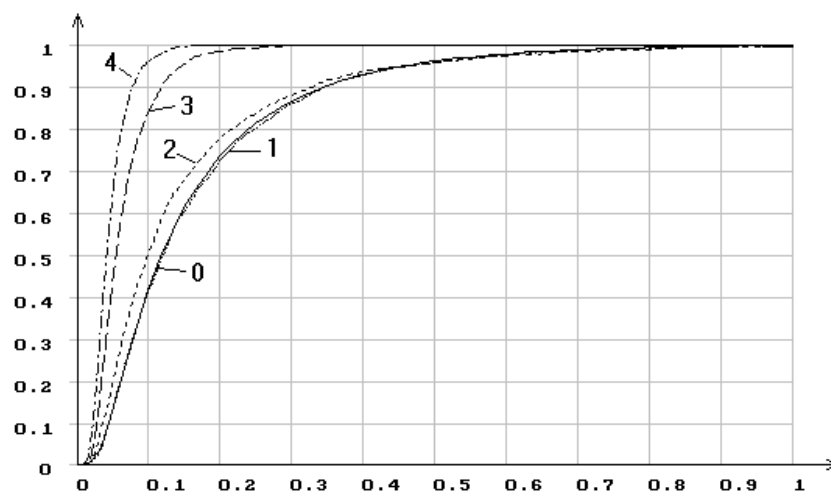


Рис. 4.4.13. Эмпирические функции распределения статистики S_{ω} Мизеса при различном количестве оцениваемых параметров логистического распределения

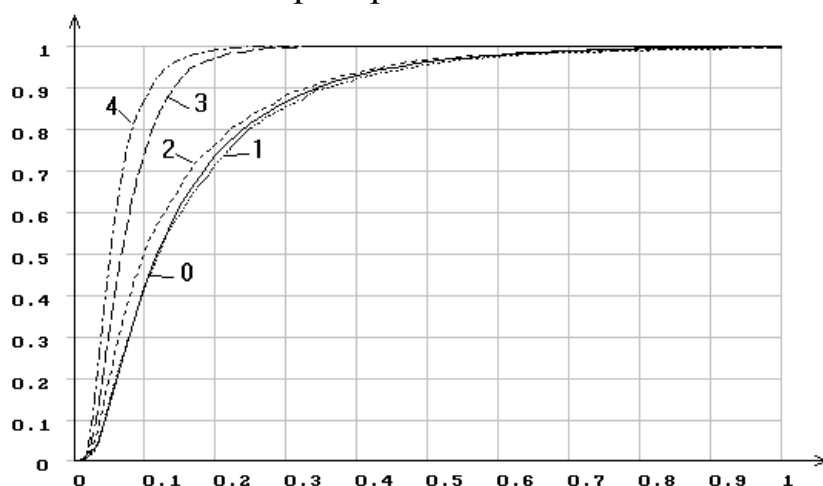


Рис. 4.4.14. Эмпирические функции распределения статистики S_{ω} Мизеса при различном количестве оцениваемых параметров распределения Лапласа

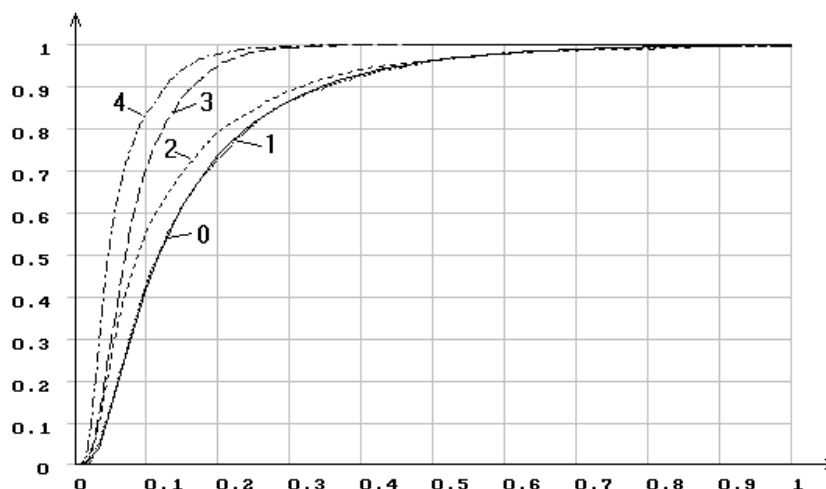


Рис. 4.4.15. Эмпирические функции распределения статистики S_{ω} Мизеса при различном количестве оцениваемых параметров распределения Коши

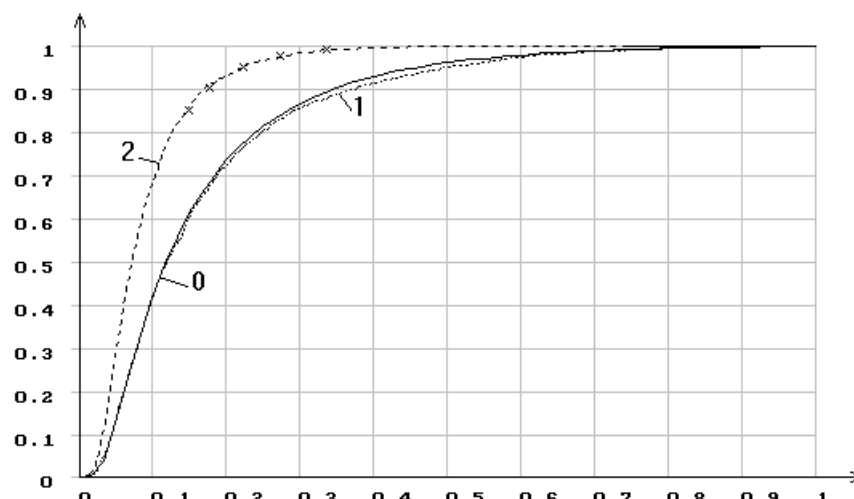


Рис. 4.4.16. Эмпирические функции распределения статистики S_{ω} Мизеса при различном количестве оцениваемых параметров экспоненциального распределения

Таблица 4.4.7.

<i>Предельные распределения статистики ω^2 Мизеса</i>				
№ п/п	Распределение случайной величины	Оценивался только масштабный параметр	Оценивался только параметр сдвига	Оценивалось два параметра
1.	Экспоненциальное	$\ln N(-2.6028, 0.6453)$		
2.	Полунормальное	$\ln N(-2.5046, 0.6814)$		
3.	Рэлея	$\ln N(-2.5743, 0.6345)$		
4.	Максвелла	$\ln N(-2.6147, 0.6361)$		
5.	Лапласа	$\ln N(-2.2328, 0.8302)$	$\ln N(-2.6890, 0.5802)$	$\ln N(-2.9386, 0.5500)$
6.	Нормальное	$\ln N(-2.2290, 0.8284)$	$\ln N(-2.8102, 0.5625)$	$\ln N(-2.9685, 0.5187)$
7.	Логнормальное	$\ln N(-2.2334, 0.7951)$	$\ln N(-2.6226, 0.6972)$	$\ln N(-2.8233, 0.6673)$
8.	Коши	$\ln N(-2.3228, 0.8554)$	$\ln N(-2.6269, 0.6202)$	$\ln N(-2.9746, 0.6493)$
9.	Логистическое	$\ln N(-2.2460, 0.8329)$	$\ln N(-2.8755, 0.5612)$	$\ln N(-3.1713, 0.4841)$
10.	Наибольшего значения	$\ln N(-2.2043, 0.8174)$	$\ln N(-2.6070, 0.6178)$	$\ln N(-3.0093, 0.5078)$
11.	Наименьшего значения	$\ln N(-2.2246, 0.8149)$	$\ln N(-2.5863, 0.6469)$	$\ln N(-3.0200, 0.5173)$
12.	Вейбулла	$\ln N(-2.5875, 0.6669)$	$\ln N(-2.2404, 0.7910)$ **	$\ln N(-3.0258, 0.5058)$

** - оценивался параметр формы распределения Вейбулла.

В табл. 4.4.8 представлены полученные приближенные значения верхних процентных точек для статистики S_{ω} .

Таблица 4.4.8.

<i>Процентные точки распределения статистики ω^2 Мизеса</i>							
№ п/п	Распределение случайной величины	Оцениваемые параметры	Верхние процентные точки				
			0.15	0.1	0.05	0.025	0.01
1.	Экспоненциальное	масштабный	0.1463	0.1752	0.2263	0.2785	0.3486
2.	Полунормальное	масштабный	0.1730	0.2073	0.2673	0.3283	0.4098
3.	Рэля	масштабный	0.1518	0.1790	0.2253	0.2713	0.3319
4.	Максвелла	масштабный	0.1463	0.1738	0.2208	0.2681	0.3306
5.	Лапласа	масштабный	0.2702	0.3394	0.4594	0.5809	0.7427
		сдвиг	0.1249	0.1440	0.1770	0.2103	0.2546
		два параметра	0.0942	0.1103	0.1377	0.1651	0.2014
6.	Нормальное	масштабный	0.2711	0.3414	0.4636	0.5871	0.7515
		сдвиг	0.1096	0.1270	0.1561	0.1847	0.2222
		два параметра	0.0874	0.0998	0.1208	0.1417	0.1691
7.	Логнормальное	масштабный	0.2599	0.3235	0.4347	0.5476	0.6987
		сдвиг	0.1481	0.1902	0.2709	0.3580	0.4799
		два параметра	0.1134	0.1474	0.2154	0.2912	0.3996
8.	Коши	масштабный	0.2565	0.3278	0.4545	0.5848	0.7604
		сдвиг	0.1416	0.1656	0.2060	0.2461	0.2987
		два параметра	0.1062	0.1285	0.1660	0.2031	0.2518
9.	Логистическое	масштабный	0.2648	0.3369	0.4667	0.6012	0.7838
		сдвиг	0.1027	0.1200	0.1507	0.1822	0.2245
		два параметра	0.0692	0.0784	0.0943	0.1161	0.1354
10.	Наибольшего значения	масштабный	0.2752	0.3413	0.4601	0.5829	0.7494
		сдвиг	0.1435	0.1697	0.2133	0.2560	0.3117
		два параметра	0.0840	0.0963	0.1172	0.1380	0.1654
11.	Наименьшего значения	масштабный	0.2649	0.3324	0.4541	0.5804	0.7518
		сдвиг	0.1500	0.1774	0.2269	0.2785	0.3488
		два параметра	0.0848	0.0974	0.1182	0.1383	0.1644
12.	Вейбулла	формы	0.2521	0.3165	0.4292	0.5438	0.6971
		масштаба	0.1535	0.1855	0.2405	0.2956	0.3686
		два параметра	0.0833	0.0940	0.1125	0.1311	0.1557

4.4.5. Моделирование и идентификация законов распределения статистики Андерсона-Дарлинга

На рис. 4.4.17-4.4.21 отражены результаты моделирования предельных законов распределения статистики S_Ω Мизеса (Андерсона-Дарлинга). На рисунках символом “0” отмечена функция распределения $a2(s)$ [19]. Этому распределению подчиняется статистика S_Ω , если по выборке не оцениваются параметры наблюдаемого закона.

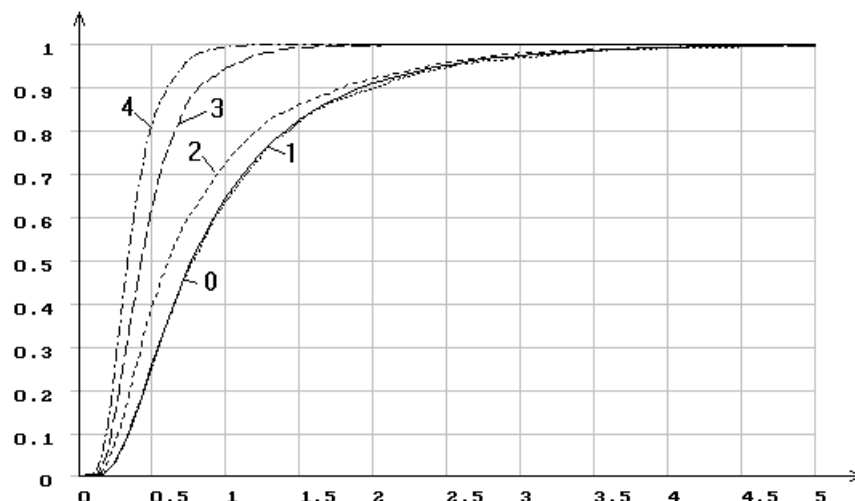


Рис. 4.4.17. Эмпирические функции распределения статистики S_Ω Мизеса при различном количестве оцениваемых параметров нормального закона:

0 - функция распределения $a2(s)$; 1 - не оценивались параметры; 2 - оценивался масштабный параметр θ_0 ; 3 - оценивался параметр сдвига θ_1 ; 4 - оценивались оба параметра.

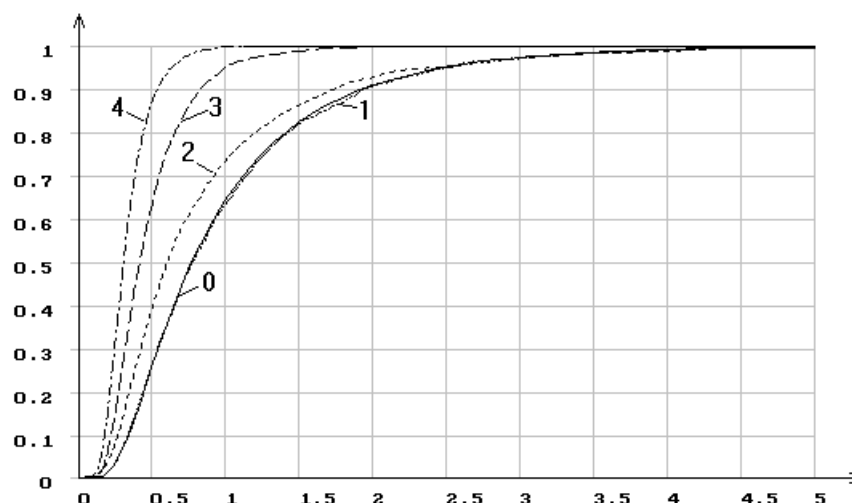


Рис. 4.4.18. Эмпирические функции распределения статистики S_Ω Мизеса при оценивании параметров логистического распределения

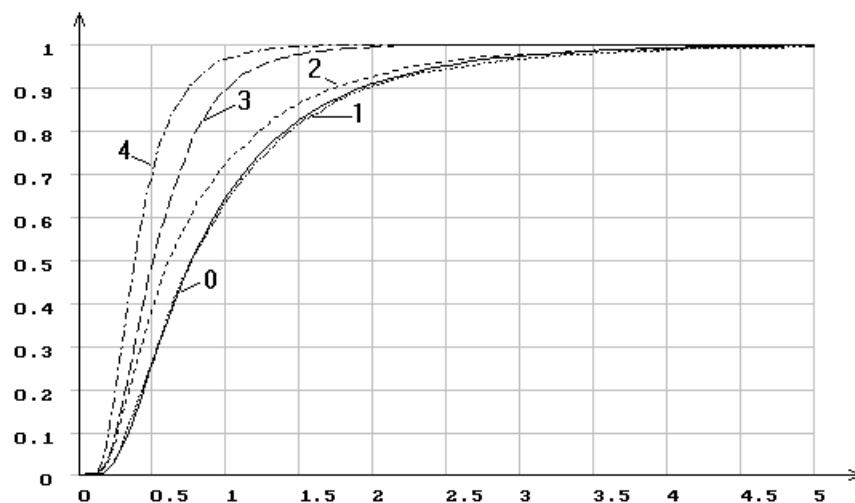


Рис. 4.4.19. Эмпирические функции распределения статистики S_{Ω} Мизеса при оценивании параметров распределения Лапласа

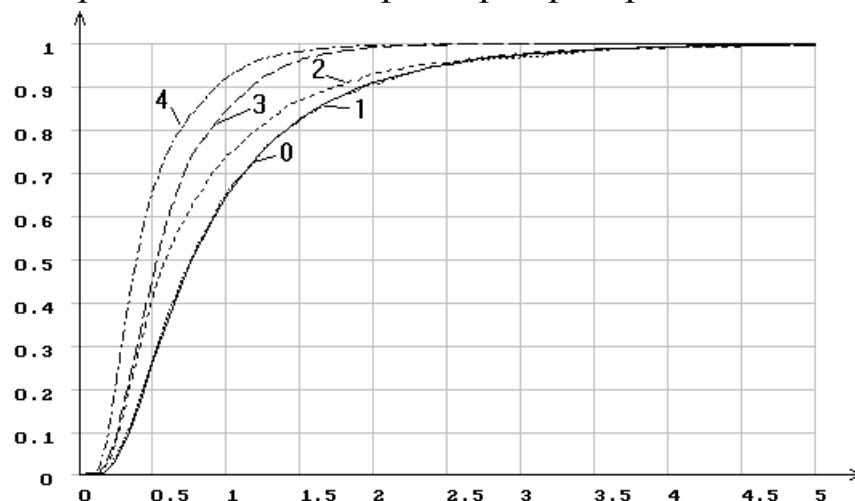


Рис. 4.4.20. Эмпирические функции распределения статистики S_{Ω} Мизеса при оценивании параметров распределения Коши

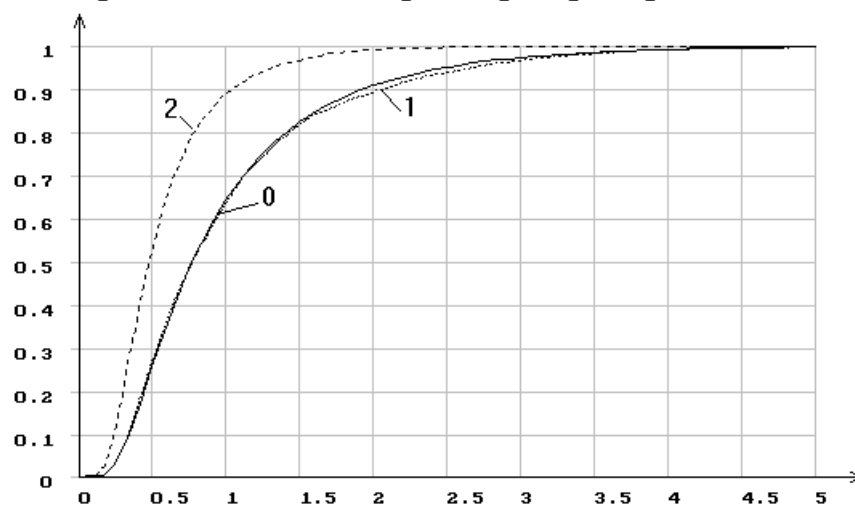


Рис. 4.4.21. Эмпирические функции распределения статистики S_{Ω} Мизеса при оценивании параметра экспоненциального распределения

В табл. 4.4.9 сведены результаты идентификации законов для статистики критерия Ω^2 Мизеса. Как и в предыдущем случае здесь предельные распределения также хуже аппроксимируются логарифмически нормальным распределением.

Таблица 4.4.9.

<i>Предельные распределения статистики Ω^2 Мизеса</i>				
№ п/п	Распределение случайной величины	Оценивался только масштабный параметр	Оценивался только параметр сдвига	Оценивалось два параметра
1.	Экспоненциальное	$\ln N(-0.7055, 0.5690)$		
2.	Полунормальное	$\ln N(-0.6931, 0.5900)$		
3.	Рэлея	$\ln N(-0.6850, 0.5541)$		
4.	Максвелла	$\ln N(-0.7051, 0.5658)$		
5.	Лапласа	$\ln N(-0.4107, 0.7170)$	$\ln N(-0.6654, 0.5345)$	$\ln N(-0.9378, 0.4999)$
6.	Нормальное	$\ln N(-0.4121, 0.7206)$	$\ln N(-0.8363, 0.5096)$	$\ln N(-1.0840, 0.4509)$
7.	Логнормальное	$\ln N(-0.4092, 0.6938)$	$\ln N(-0.6827, 0.6146)$	$\ln N(-0.9322, 0.5819)$
8.	Коши	$\ln N(-0.4326, 0.7164)$	$\ln N(-0.6102, 0.5737)$	$\ln N(-0.8815, 0.5905)$
9.	Логистическое	$\ln N(-0.4243, 0.7224)$	$\ln N(-0.8465, 0.5109)$	$\ln N(-1.1685, 0.4239)$
10.	Наибольшего значения	$\ln N(-0.3863, 0.7097)$	$\ln N(-0.6903, 0.5534)$	$\ln N(-1.0809, 0.4447)$
11.	Наименьшего значения	$\ln N(-0.4007, 0.7108)$	$\ln N(-0.6866, 0.5667)$	$\ln N(-1.0986, 0.4474)$
12.	Вейбулла	$\ln N(-0.6908, 0.5866)$	$\ln N(-0.4177, 0.6846)$ **	$\ln N(-1.1044, 0.4405)$

** - оценивался параметр формы распределения Вейбулла.

В табл. 4.4.10 приведены приближенные значения верхних процентных точек для статистики S_Ω .

4.4.6. Предельные распределения статистик для выборок из гамма-распределения

Как и следовало ожидать, предельные распределения рассматриваемых статистик для выборок из гамма-распределения с плотностью

$$f(x) = \frac{\theta_0^{\theta_0}}{\Gamma(\theta_0)} x^{\theta_0-1} e^{-\theta_0 x}$$

зависят от параметра формы θ_0 [140]. Выборки статистик моделировались при различных значениях параметра θ_0 . Для иллюстрации приведем лишь результаты моделирования статистики Колмогорова $\sqrt{S_k} / 2$.

Таблица 4.4.10.

<i>Процентные точки распределения статистики Ω^2 Мизеса</i>							
№ п/п	Распределение случайной величины	Оцениваемые параметры	Верхние процентные точки				
			0.15	0.1	0.05	0.025	0.01
1.	Экспоненциальное	масштабный	0.8975	1.0599	1.3407	1.6235	1.9995
2.	Полунормальное	масштабный	0.9912	1.1672	1.4791	1.7987	2.2289
3.	Рэля	масштабный	0.9168	1.0629	1.3128	1.5629	1.8938
4.	Максвелла	масштабный	0.8975	1.0569	1.3342	1.6150	1.9896
5.	Лапласа	масштабный	1.4505	1.7830	2.3789	2.9943	3.8275
		сдвиг	0.8979	1.0326	1.2598	1.4851	1.7810
		два параметра	0.6604	0.7593	0.9312	1.1051	1.3369
6.	Нормальное	масштабный	1.4538	1.8026	2.4220	3.0579	3.9146
		сдвиг	0.7300	0.8371	1.0340	1.2410	1.5249
		два параметра	0.5409	0.6095	0.7239	0.8366	0.9837
7.	Логнормальное	масштабный	1.4278	1.7433	2.2896	2.8407	3.5739
		сдвиг	0.9489	1.1734	1.5913	2.0342	2.6453
		два параметра	0.6893	0.8589	1.2006	1.5829	2.1318
8.	Коши	масштабный	1.4282	1.7664	2.3817	3.0238	3.8996
		сдвиг	1.0213	1.1849	1.4549	1.7185	2.0607
		два параметра	0.8019	0.9480	1.1928	1.4345	1.7508
9.	Логистическое	масштабный	1.4373	1.7879	2.4150	3.0617	3.9361
		сдвиг	0.7344	0.8489	1.0486	1.2510	1.5213
		два параметра	0.4849	0.5423	0.6364	0.7277	0.8460
10.	Наибольшего значения	масштабный	1.4936	1.8259	2.4120	3.0108	3.8148
		сдвиг	0.9061	1.0523	1.3054	1.5607	1.9004
		два параметра	0.5443	0.6107	0.7248	0.8392	0.9907
11.	Наименьшего значения	масштабный	1.4538	1.7934	2.3948	3.0111	3.8404
		сдвиг	0.9088	1.0660	1.3426	1.6250	2.0039
		два параметра	0.5398	0.6092	0.7198	0.8251	0.9594
12.	Вейбулла	формы	1.3909	1.7030	2.2511	2.8096	3.5579
		масштаба	0.9355	1.1049	1.4018	1.7039	2.1081
		два параметра	0.5253	0.5908	0.6991	0.8052	0.9432

На рис. 4.4.22 представлены эмпирические распределения статистики при оценивании по выборке параметра формы, на рис. 4.4.23 - масштабного параметра, на рис. 4.4.24 - двух параметров распределения. На этих рисунках “1” помечена эмпирическая функция распределения статистики при $\theta_0=0.5$; “2” - при $\theta_0=1.0$; “3” - при $\theta_0=2.0$; “4” - при $\theta_0=5.0$; “5” - при $\theta_0=10.0$. “0” помечена функция распределения Колмогорова. Отметим, что с ростом θ_0 предельные распределения статистик стремятся к предельным распределениям статистик для выборок из нормального закона. При значениях $\theta_0 > 5$ эмпирические распределения статистик при оценивании двух параметров практически совпадают и хорошо согласуются с распределением соответствующей статистики для нормального закона.

Общая картина принципиально сохраняется и для распределений других непараметрических статистик.

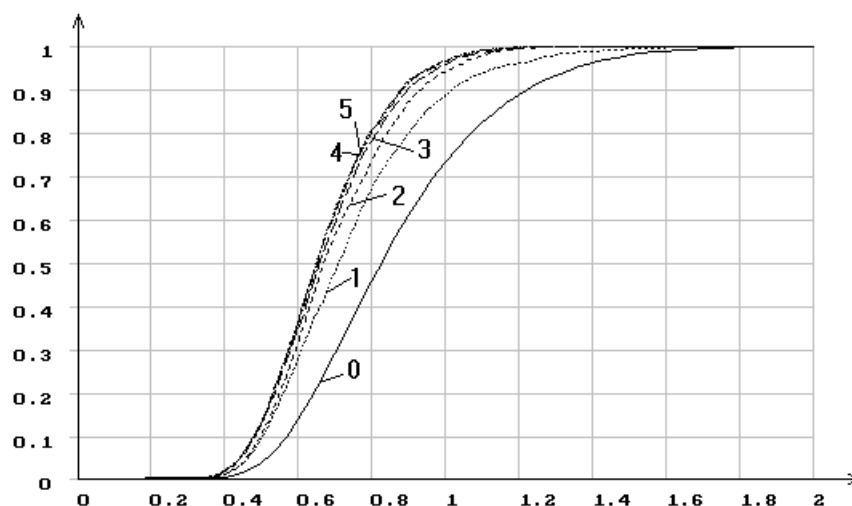


Рис. 4.4.22. Эмпирические функции распределения статистики $\sqrt{S_k} / 2$ Колмогорова при оценивании параметра формы гамма-распределения: 0 - функция распределения Колмогорова; 1 - $\theta_0=0.5$; 2 - $\theta_0=1.0$; 3 - $\theta_0=2.0$; 4 - $\theta_0=5.0$; 5 - $\theta_0=10.0$.

4.4.7. Предельные распределения статистик при оценивании по выборке параметров смеси распределений

В качестве моделей законов распределения реальных случайных величин могут использоваться и усеченные законы, и смеси законов

вида $f(x, \theta) = wf_1(x, \theta_1) + (1-w)f_2(x, \theta_2)$, где w - параметр смеси, θ_1, θ_2 - параметры законов, входящих в смесь, в общем случае векторные. Если по выборке будут оцениваться все или часть параметров, определяющих смесь законов, то это отразится на предельном распределении статистики используемого критерия.

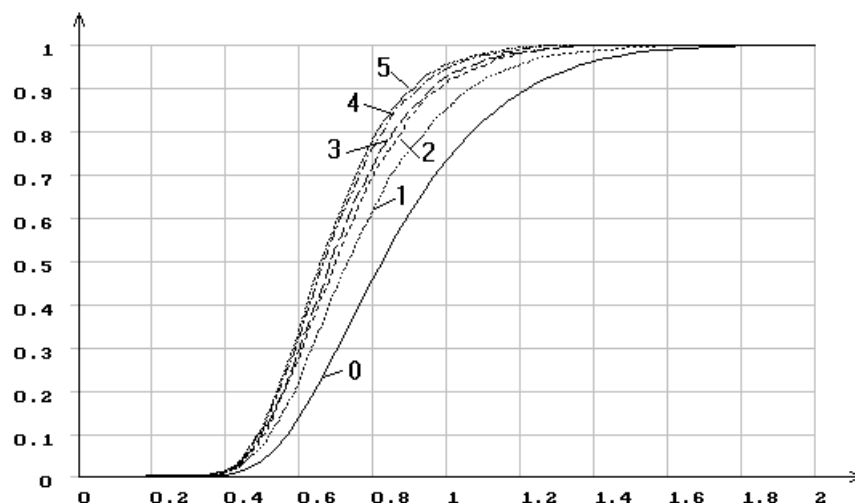


Рис. 4.4.23. Эмпирические функции распределения статистики $\sqrt{S_k} / 2$ Колмогорова при оценивании масштабного параметра гамма-распределения: 0 - функция распределения Колмогорова; 1 - $\theta_0 = 0.5$; 2 - $\theta_0 = 1.0$; 3 - $\theta_0 = 2.0$; 4 - $\theta_0 = 5.0$; 5 - $\theta_0 = 10.0$.

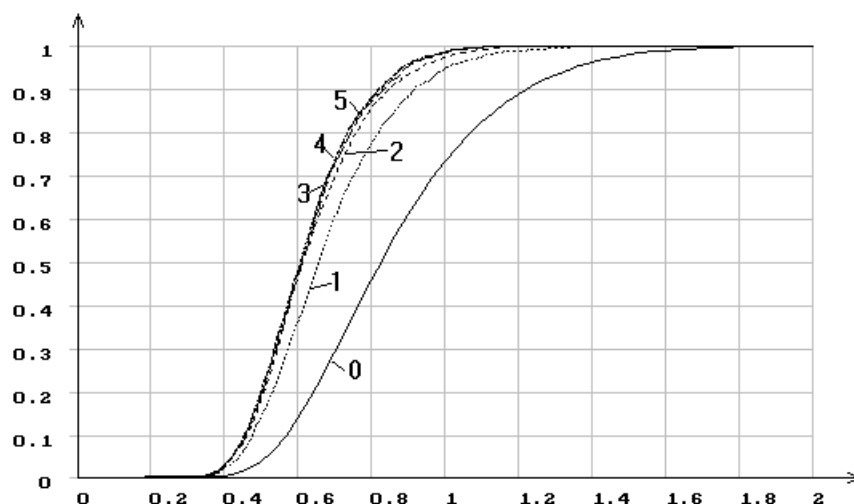


Рис. 4.4.24. Эмпирические функции распределения статистики $\sqrt{S_k} / 2$ Колмогорова при оценивании одновременно двух параметров гамма-распределения: 0 - функция распределения Колмогорова; 1 - $\theta_0 = 0.5$; 2 - $\theta_0 = 1.0$; 3 - $\theta_0 = 2.0$; 4 - $\theta_0 = 5.0$; 5 - $\theta_0 = 10.0$.

На рис. 4.4.25 для иллюстрации приведены смоделированные распределения статистики Колмогорова при оценивании 2-х параметров нормального распределения (“1”) и при оценивании всех 5-ти параметров смеси двух нормальных распределений (“2”). Таким образом, очевидно, что в случае оценивания параметров смеси двух нормальных законов предельные распределения статистики типа Колмогорова также очень существенно отличаются не только от распределения Колмогорова $K(x)$, но и от распределения этой статистики при оценивании по данной выборке параметров описывающего ее нормального закона. То же самое можно утверждать относительно предельных распределений других рассматриваемых в данном разделе статистик непараметрических критериев.

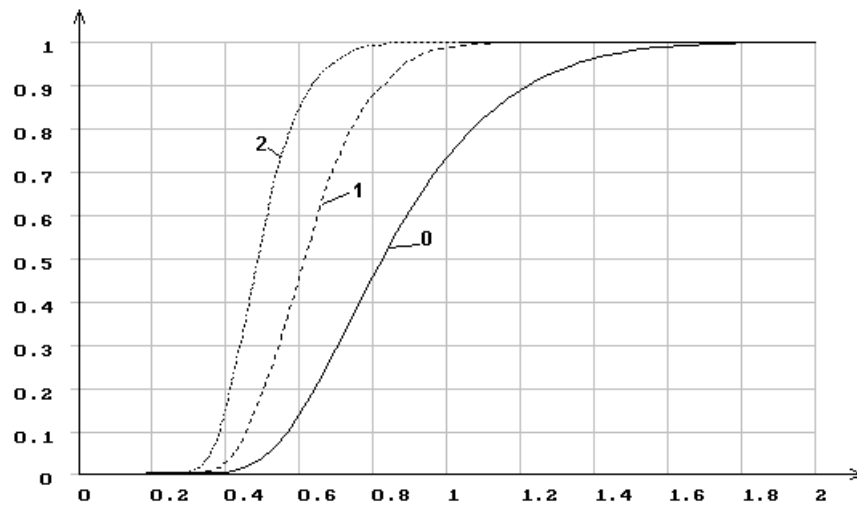


Рис. 4.4.25. Распределения статистики Колмогорова $\sqrt{S_k} / 2$ при оценивании по выборке всех параметров распределения: 1 - нормального; 2 - смеси двух нормальных распределений (5 параметров); 0 - функция распределения Колмогорова $K(x)$.

Выводы

Можно констатировать, что на основании результатов моделирования могут быть получены хорошие приближения предельных распределений статистик непараметрических критериев. Причем достаточно хорошая аппроксимация для реальных распределений статистик непараметрических критериев обычно получается с использованием логарифмически нормального и/или гамма-распределения. Для ряда законов случайных величин идентифицированы распределения статистик непараметрических критериев при различном количестве оцененных параметров, построены таблицы верхних процентных точек.

4.5. Использование критериев согласия в случае частично группированных и интервальных наблюдений

Критерии отношения правдоподобия и χ^2 Пирсона работают с группированными наблюдениями или данными, которые группируются в процессе проверки гипотез. В критериях типа Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса измеряется расстояние между теоретическим и эмпирическим законом распределения, причем подразумевается, что наблюдались индивидуальные значения случайной величины. Если же данные, по которым проверяется гипотеза, представляют собой частично группированную или интервальную выборку, то применение всех рассматриваемых критериев оказывается затруднительным.

В случае частично группированной выборки нет принципиальных трудностей использования критериев отношения правдоподобия и χ^2 , если имеется возможность разбиения на интервалы негруппированной части выборки. Но в случае интервальных наблюдений, которые не могли быть зафиксированы точно (например из-за колебаний) и представляются в виде интервалов $[a_i, b_i]$, содержащих неизвестное точно значение наблюдения, или при нечетком описании наблюдений случайных величин [215] непосредственное применение всех рассматриваемых в работе критериев невозможно.

Подробная библиография работ, связанных с вопросами интервальной статистики, приведена в [126]. Так как классические методы статистического анализа нельзя применить к интервальным выборкам в явном виде, то для адаптации известных методов обычным приемом может служить построение интервала неопределенности интересующей исследователя статистики [54]. В данном разделе рассмотрены модификации параметрических критериев согласия χ^2 Пирсона и отношения правдоподобия и непараметрических критериев Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса, при которых находятся нижние и верхние оценки соответствующих статистик при частично группированных или интервальных наблюдениях, позволяющие делать статистические выводы [77,81-83,98].

Аналогичные подходы к перечисленным выше задачам были рассмотрены в [216,123]. Gastaldi [216] нашел верхнюю и нижнюю границу статистики Колмогорова в случае, когда выборка задана с пропусками данных, но при этом известно количество пропущенных

наблюдений на интервалах между членами вариационного ряда. Орлов [123] сформулировал общие подходы к проверке гипотез в случае интервального представления выборки и в качестве одного из примеров рассмотрел критерий Смирнова однородности двух выборок.

Пусть задана интервальная выборка [83]:

$$\mathbf{X}_k = \left\{ \langle a_i, b_i, n_i \rangle \mid a_i \leq x_{ij} \leq b_i, j=1, \dots, n_i; a_i \in R, b_i \in R, i=1, \dots, k \right\},$$

где k - число интервалов, $N = \sum_{i=1}^k n_i$ - объём выборки, n_i - число наблюдений в i -м интервале, a_i и b_i - границы i -го интервала, x_{ij} - точные значения наблюдений. Упорядочим граничные точки интервалов:

$$a_{(1)} < a_{(2)} < \dots < a_{(k)},$$

$$b_{(1)} < b_{(2)} < \dots < b_{(k)}.$$

Обозначим через $n_{a(i)}$ - число наблюдений, принадлежащих интервалу с левой границей $a_{(i)}$, и $n_{b(i)}$ - число наблюдений, принадлежащих интервалу с правой границей $b_{(i)}$. Тогда

$$N_{a(i)} = \sum_{j=1}^i n_{a(j)}, N_{b(i)} = \sum_{j=1}^i n_{b(j)}, i=1, \dots, k.$$

Эмпирическая функция распределения $F_N(x)$, построенная по точным (индивидуальным) наблюдениям x_{ij} , будет ограничена снизу и сверху функциями $\underline{F}_N(x)$ и $\overline{F}_N(x)$, имеющими следующий вид:

$$\overline{F}_N(x) = \begin{cases} 0, & x < a_{(1)}, \\ \frac{N_{a(i)}}{N}, & a_{(i)} \leq x < a_{(i+1)}, i=1, \dots, k-1, \\ 1, & x \geq a_{(k)}, \end{cases}$$

$$\underline{F}_N(x) = \begin{cases} 0, & x < b_{(1)}, \\ \frac{N_{b(i)}}{N}, & b_{(i)} \leq x < b_{(i+1)}, i=1, \dots, k-1, \\ 1, & x \geq b_{(k)}. \end{cases}$$

Таким образом,

$$\underline{F}_N(x) \leq F_N(x) \leq \overline{F}_N(x), \quad \forall x \in R. \quad (4.5.1)$$

На рис. 4.5.1 приведен вид эмпирической функции распределения ($\underline{F}_N(x)$ и $\overline{F}_N(x)$) для различных типов исходных данных.

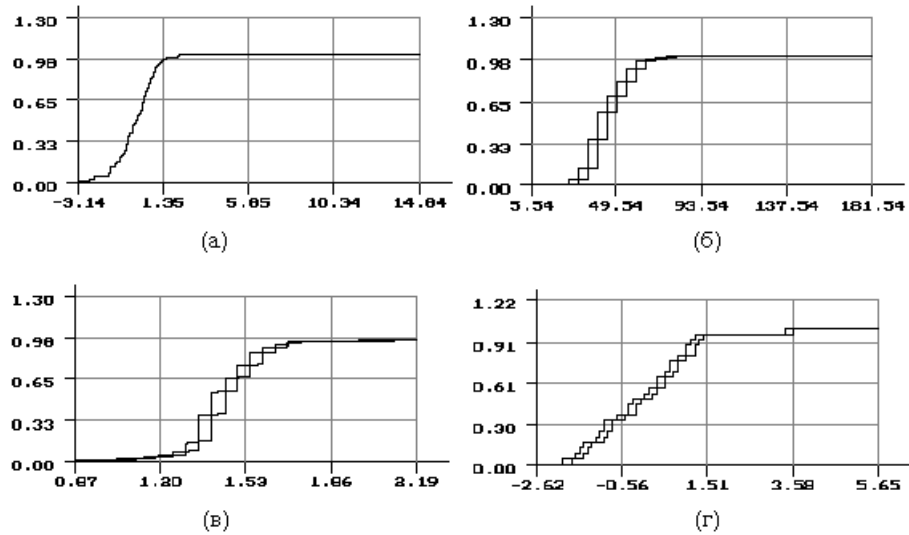


Рис. 4.5.1. Эмпирическая функция распределения для различных типов выборочных данных: (а)-негруппированные, (б)-группированные, (в)-частично группированные, (г)-интервальные с пересекающимися интервалами

При проверке гипотез о согласии для найденного значения соответствующей статистики S^* вычисляется вероятность

$$p = P\{S > S^*\} = \int_{S^*}^{\infty} g(s)ds,$$

где $g(s)$ - плотность распределения статистики при условии истинности нулевой гипотезы. При заданном уровне значимости α гипотеза о согласии не отвергается, если $p > \alpha$. Когда выборка задана неточно, то статистика принадлежит интервалу $[S^*, \bar{S}^*]$, где на основании (4.5.1) границы определяются следующим неравенством:

$$\underline{S}^* = \inf_{F_N \leq F_N \leq \bar{F}_N} S^*(F_N, F) \leq S^*(F_N, F) \leq \sup_{F_N \leq F_N \leq \bar{F}_N} S^*(F_N, F) = \bar{S}^*$$

Вероятность $P\{S > S^*\}$ будет принадлежать интервалу $[p_{\min}, p_{\max}]$, где

$$p_{\min} = \int_{S^*}^{\infty} g(s)ds, \quad p_{\max} = \int_{\bar{S}^*}^{\infty} g(s)ds.$$

Тогда, при заданном уровне значимости α , гипотезу о согласии следует отклонить, если $p_{\max} \leq \alpha$; гипотезу о согласии не следует отвергать, если $p_{\min} > \alpha$.

Рассмотрим критерий χ^2 Пирсона. Перед использованием критерия необходимо сгруппировать исходную выборку. Область определения случайной величины разбивается на K непересекающихся интервалов граничными точками

$$x_{(0)} < x_{(1)} < \dots < x_{(K)},$$

после чего подсчитывается число наблюдений m_i , попавших в интервалы $(x_{(i)}, x_{(i+1)}]$, $i = 0, 1, \dots, K-1$. Процедура подсчета не представляет трудности, если выборка задана точно. В случае же, если граничная точка попадает внутрь интервального наблюдения, $a_i < x_{(j)} < b_i$, процедура становится неоднозначной, так как точное значение можно отнести как к интервалу $(x_{(j-1)}, x_{(j)}]$, так и к интервалу $(x_{(j)}, x_{(j+1)}]$. Возможные значения m_i в соответствии с (1) удовлетворяют ограничениям:

$$N \cdot \underline{F}_N(x_{(i)}) \leq \sum_{j=0}^i m_j \leq N \cdot \overline{F}_N(x_{(i)}), m_i \geq 0, i = 0, 1, \dots, K-1, \quad (4.5.2)$$

$$\sum_{i=0}^{K-1} m_i = n. \quad (4.5.3)$$

Статистика критерия имеет вид

$$\chi^2 = \sum_{i=0}^{K-1} \frac{(m_i - np_i)^2}{p_i}.$$

Если m_i определяются неоднозначно, то можно найти максимум и минимум статистики χ^2 на области, заданной формулами (4.5.2) и (4.5.3):

$$\underline{\chi^2} = \min_{m_i} \chi^2, \quad \overline{\chi^2} = \max_{m_i} \chi^2.$$

Аналогично ограничивается статистика критерия отношения правдоподобия.

Для критерия Колмогорова статистика имеет вид

$$D_N = \sup_x (F_N(x) - F(x)),$$

где $F_N(x)$ - эмпирическая функция распределения, $F(x)$ - теоретическая, согласие с которой проверяется, N - объем выборки. Преобразуем неравенство (4.5.1) к виду:

$$\begin{aligned} \underline{F}_N(x) - F(x) \leq F_N(x) - F(x) \leq \overline{F}_N(x) - F(x), \\ F(x) - \overline{F}_N(x) \geq F(x) - F_N(x) \geq F(x) - \underline{F}_N(x). \end{aligned}$$

Эти неравенства выполняются для всех x , поэтому они сохраняются при взятии супремума:

$$\sup_x (\underline{F}_N(x) - F(x)) \leq \sup_x (F_N(x) - F(x)) \leq \sup_x (\overline{F}_N(x) - F(x)),$$

$$\sup_x \left(F(x) - \overline{F}_n(x) \right) \geq \sup_x \left(F(x) - F_n(x) \right) \geq \sup_x \left(F(x) - \underline{F}_n(x) \right).$$

Объединим эти неравенства в одно:

$$\begin{aligned} \underline{D}_N &= \max \left\{ \sup_x \left(\underline{F}_N(x) - F(x) \right), \sup_x \left(F(x) - \overline{F}_N(x) \right) \right\} \leq \\ &\leq D_N = \max \left\{ \sup_x \left(F_N(x) - F(x) \right), \sup_x \left(F(x) - F_N(x) \right) \right\} \leq \\ &\leq \overline{D}_N = \max \left\{ \sup_x \left(\overline{F}_N(x) - F(x) \right), \sup_x \left(F(x) - \underline{F}_N(x) \right) \right\}. \end{aligned}$$

В случае критерия Смирнова статистика имеет вид

$$D_N^+ = \sup_x \left(F_N(x) - F(x) \right).$$

Из неравенства (4.5.2) следует:

$$\underline{D}_N^+ = \sup_x \left(\underline{F}_N(x) - F(x) \right), \quad \overline{D}_N^+ = \sup_x \left(\overline{F}_N(x) - F(x) \right).$$

В [83] нами были построены также соответствующие оценки для статистик критериев ω^2 и Ω^2 Мизеса. Следующий пример [98] иллюстрирует применение рассмотренного подхода. Выборка из 100 наблюдений генерировалась с абсолютной погрешностью $\varepsilon = 0.05$. Далее по ней было проверено согласие с нормальным распределением с параметрами $\mu = -0.0786$ и $\sigma = 0.9916$ (см. рис. 4.5.2). На диаграмме в правом верхнем углу цифрами обозначена вероятность согласия по критериям: 1,2 - отношения правдоподобия, 3,4 - χ^2 Пирсона, 5 - Колмогорова, 6 - Смирнова, 7,8 - ω^2 и Ω^2 Мизеса. Заштрихованные области показывают интервалы неопределенности вероятности согласия. По результатам проверки согласия можно сделать следующие выводы:

- При уровне значимости $\alpha = 0.3$ гипотеза о согласии *не отвергается* по критериям отношения правдоподобия, χ^2 Пирсона, Колмогорова, Смирнова, ω^2 .
- При уровне значимости $\alpha = 0.5$ гипотеза о согласии *не отвергается* по критериям отношения правдоподобия, χ^2 Пирсона, Смирнова.
- При уровне значимости $\alpha = 0.9$ гипотеза о согласии *отвергается* по критериям отношения правдоподобия, χ^2 Пирсона, Смирнова.
- По остальным критериям однозначного вывода сделать невозможно.

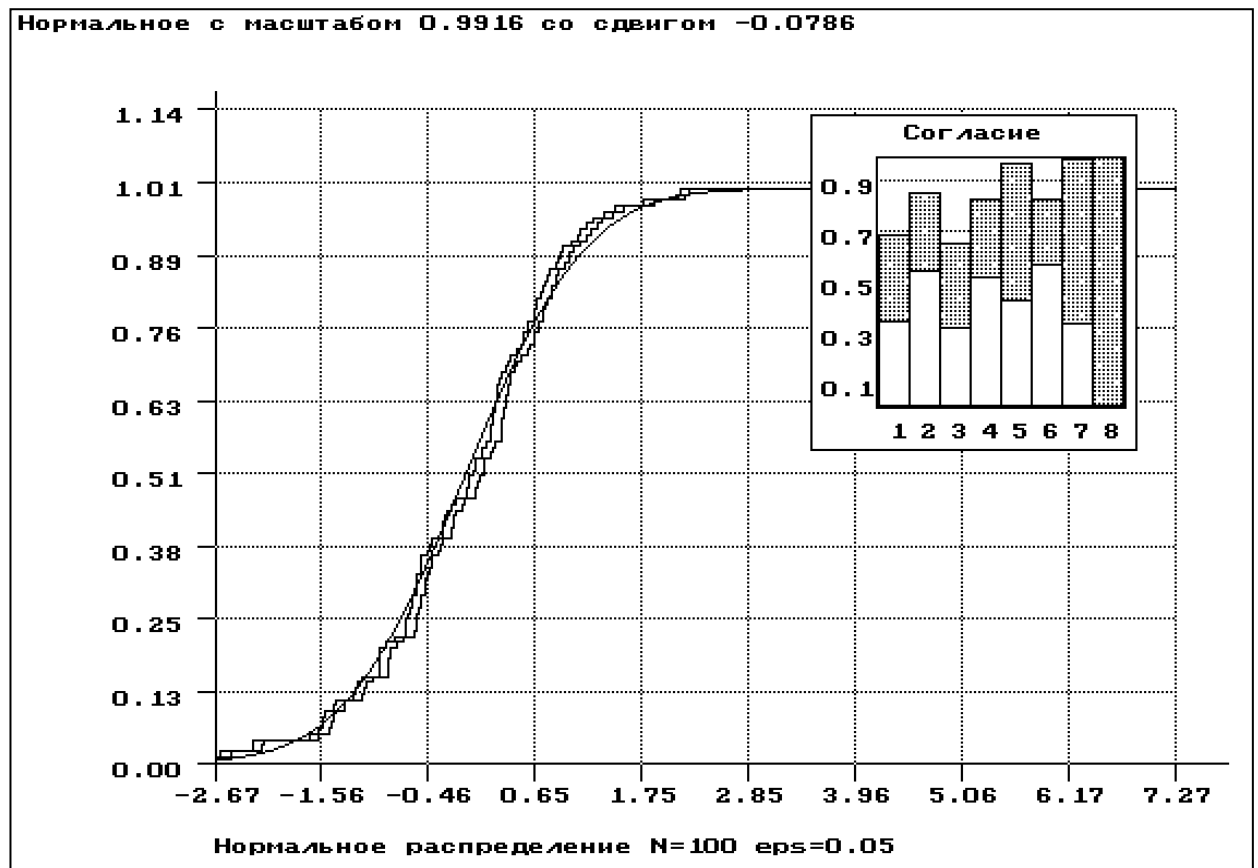


Рис. 4.5.2. Проверка согласия интервальной выборки с нормальным распределением

В [83] был проведен численный эксперимент по исследованию длины интервала $p_{\max} - p_{\min}$ в зависимости от близости к истинному значению масштабного параметра нормального распределения, соответствующего гипотезе H_0 . Для этого была смоделирована негруппированная выборка объемом 100 наблюдений по нормальному закону с параметрами $(0,1)$. Оценки максимального правдоподобия параметров сдвига и масштаба по негруппированной выборке равны, соответственно, $\hat{\mu} = 0.048574$ и $\hat{\sigma} = 1.07144$. Далее, предположили, что исходные наблюдения получены с некоторой ошибкой, не превосходящей по модулю $\Delta = 0.05$. В этом случае каждое i -е наблюдение является интервалом вида $[x_i - \Delta, x_i + \Delta]$.

Примечание: Для данной интервальной выборки при известном значении параметра сдвига, равным ОМП по исходной негруппированной выборке, были вычислены оценки, минимизирующие соответствующие статистики по интервальным данным (MD-оценки), и оценка по методу максимального правдоподобия. Последняя оценка, равная $\hat{\sigma} = 1.0657$, оказалась наиболее близкой к ОМП по исходным негруппированным данным.

Зафиксировав значение параметра сдвига равным оценке максимального правдоподобия по негруппированной выборке, проверялось согласие по всем рассмотренным критериям в зависимости от изменяющегося значения параметра масштаба (см. рис. 4.5.3). На графике изображены функции $p_{\max}(\sigma)$ и $p_{\min}(\sigma)$.

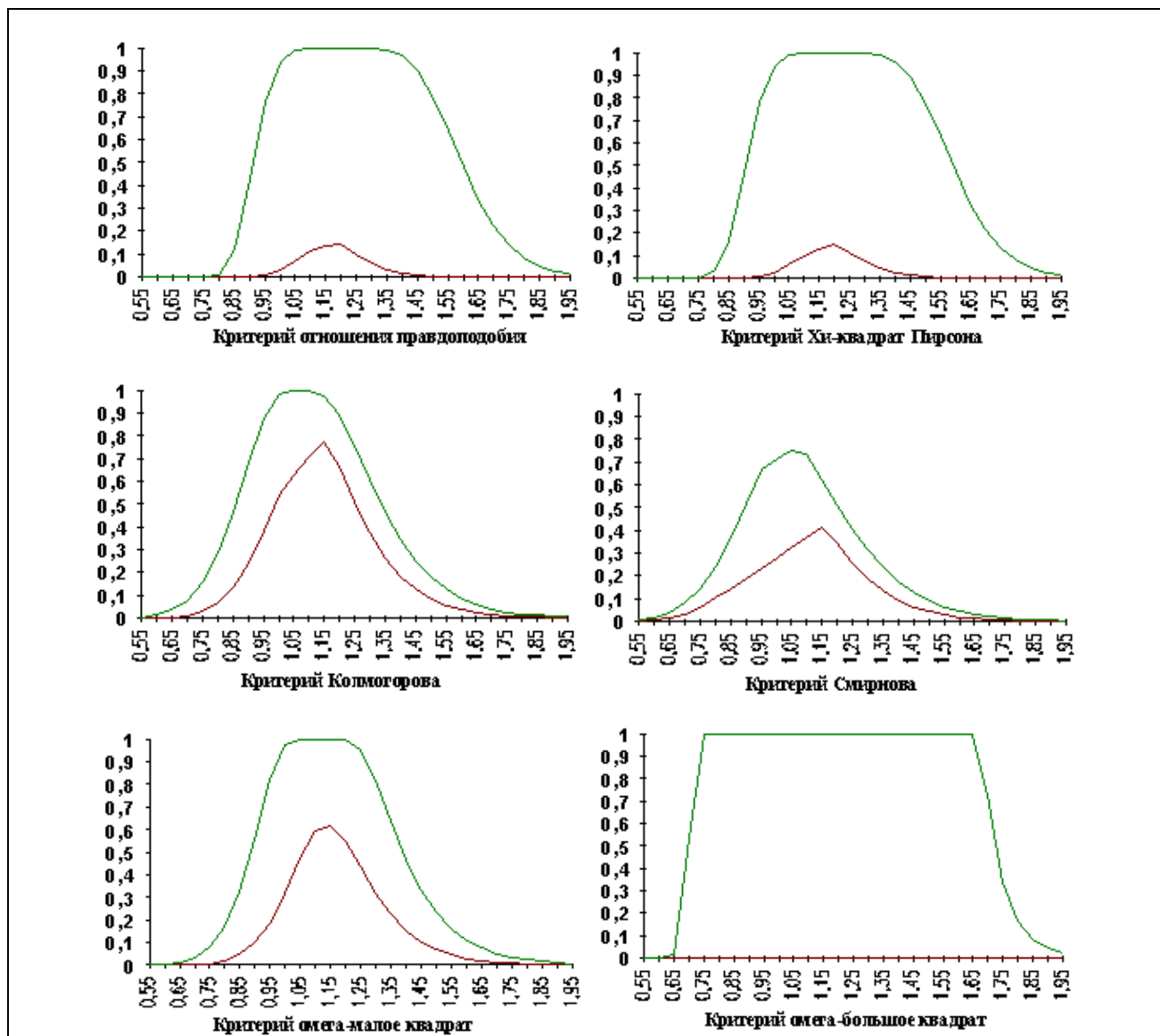


Рис. 4.5.3 Согласие интервальной выборки с нормальным распределением с параметрами (μ, σ) , где $\mu=0.048574$, а σ изменяется от 0.55 до 1.95.

Верхняя линия соответствует $p_{\max}$, нижняя - $p_{\min}$.

На основании той же негруппированной выборки, была исследована зависимость длины интервала $p_{\max} - p_{\min}$ от Δ . Проверялось согласие с нормальным распределением с параметрами, оцененными по методу максимального правдоподобия. Естественно, что с ростом неопределенности исходных данных (Δ), происходило увеличение

неопределенности “вероятности согласия” ($\Delta p = p_{\max} - p_{\min}$). При этом критерии можно упорядочить в порядке возрастания Δp следующим образом: Колмогорова, Смирнова, ω^2 Мизеса, отношения правдоподобия, χ^2 Пирсона, Ω^2 Мизеса. Поэтому, при интервальном задании исходной выборки возможно более предпочтительным при проверке согласия оказывается использование критериев Колмогорова, Смирнова, ω^2 Мизеса.

Очевидно, что чем меньше интервал неопределенности $[p_{\min}, p_{\max}]$ тем более определенные выводы можно сделать. На длину интервала неопределенности $\Delta p = p_{\max} - p_{\min}$ влияют неопределенность в задании исходных данных, выбранная модель, критерий согласия и количество наблюдений. О том как увеличение объема выборки влияет на Δp говорит следующая теорема об асимптотических свойствах оценок границ статистики критерия Колмогорова по интервальной выборке [98] (доказательство получено Постоваловым С.Н.).

Теорема 4.5.1. Пусть задана последовательность интервальных выборок X_N , для которых нижняя и верхняя границы эмпирической функции распределения $\underline{F}_N(x)$ и $\overline{F}_N(x)$ сходятся соответственно к $\underline{F}(x)$ и $\overline{F}(x)$ со скоростью $O(1/N)$, и $\sup_x (\overline{F}(x) - \underline{F}(x)) \geq c > 0$.

Тогда, при $N \rightarrow \infty$

1. $\forall F \in \mathfrak{Z}$, таких что $\forall x \underline{F}(x) \leq F(x) \leq \overline{F}(x)$, следует
 - (a) $p_{\max}(F, X_N) \rightarrow 1$, (b) $p_{\min}(F, X_N) \rightarrow 0$;
2. $\forall F \in \mathfrak{Z}$, таких что $\exists x F(x) < \underline{F}(x) \vee F(x) > \overline{F}(x)$, следует
 - (a) $p_{\max}(F, X_N) \rightarrow 0$, (b) $p_{\min}(F, X_N) \rightarrow 0$;

где $\mathfrak{Z}$ - множество всех функций распределения, непрерывных справа, $p_{\max}(F, X_N)$ и $p_{\min}(F, X_N)$ - соответственно верхняя и нижняя границы вероятности согласия по критерию Колмогорова.

Поведение $p_{\max}$ и $p_{\min}$ в зависимости от объема выборки иллюстрирует следующий пример. Были сгенерированы три интервальные выборки с абсолютной погрешностью $\varepsilon = 0.05$, подчиненные одному и тому же закону распределения, объемом 100, 500 и 1000 наблюдений соответственно. Затем было исследовано поведение $p_{\max}$ и $p_{\min}$ при проверке согласия по критерию Колмогорова с нормальным распределением, у которого параметр μ зафиксирован, а параметр σ

изменялся от 0.5 до 1.5 (см. рис. 4.5.4). Хорошо видно, что с ростом количества наблюдений верхняя кривая согласия ($p_{\max}$) становится более крутой, а нижняя ($p_{\min}$) становится ближе к нулю. Это означает, что множество распределений, не отвергаемых по критерию согласия при заданном уровне значимости уменьшается, но неопределенность при принятии решений о согласии для этих распределений увеличивается.

С учетом всего вышеизложенного можно сделать заключение, что при анализе интервальной выборки возможно применение классических методов проверки гипотез о согласии (в случае простой гипотезы), используя, например, полученные оценки для статистик критериев. Однако выводы становятся менее определенными при уменьшении точности наблюдений, причем эта неопределенность может увеличиваться с ростом количества наблюдений. При увеличении объема интервальной выборки для целого множества априори допустимых для описания данной случайной величины распределений (для множества альтернатив) длина интервала неопределенности вероятности согласия растет и стремится к единице. Это значит, что функцию распределения случайной величины, наблюдения которой фиксируются с неустранимой погрешностью, невозможно определить точно, даже при очень большом числе экспериментов. Отсюда следует практический вывод: если условия регистрации наблюдений позволяют проводить их измерение лишь с какой-то ограниченной точностью, то существует разумный объем выборки, превышение которого ничего не дает ни для точности оценивания параметров по такой выборке, ни для проверки гипотез о согласии.

В случае сложных гипотез, когда по интервальной выборке необходимо оценивать параметры закона, использование рассмотренных модификаций критериев проблематично, так как предельные распределения статистик (или их оценок снизу и сверху) будут, по-видимому, зависеть от метода оценивания и, возможно, конкретного вида выборки. Поэтому, даже найдя оценки снизу и сверху для используемой статистики, нельзя будет вычислить вероятности $p_{\max}$ и $p_{\min}$.

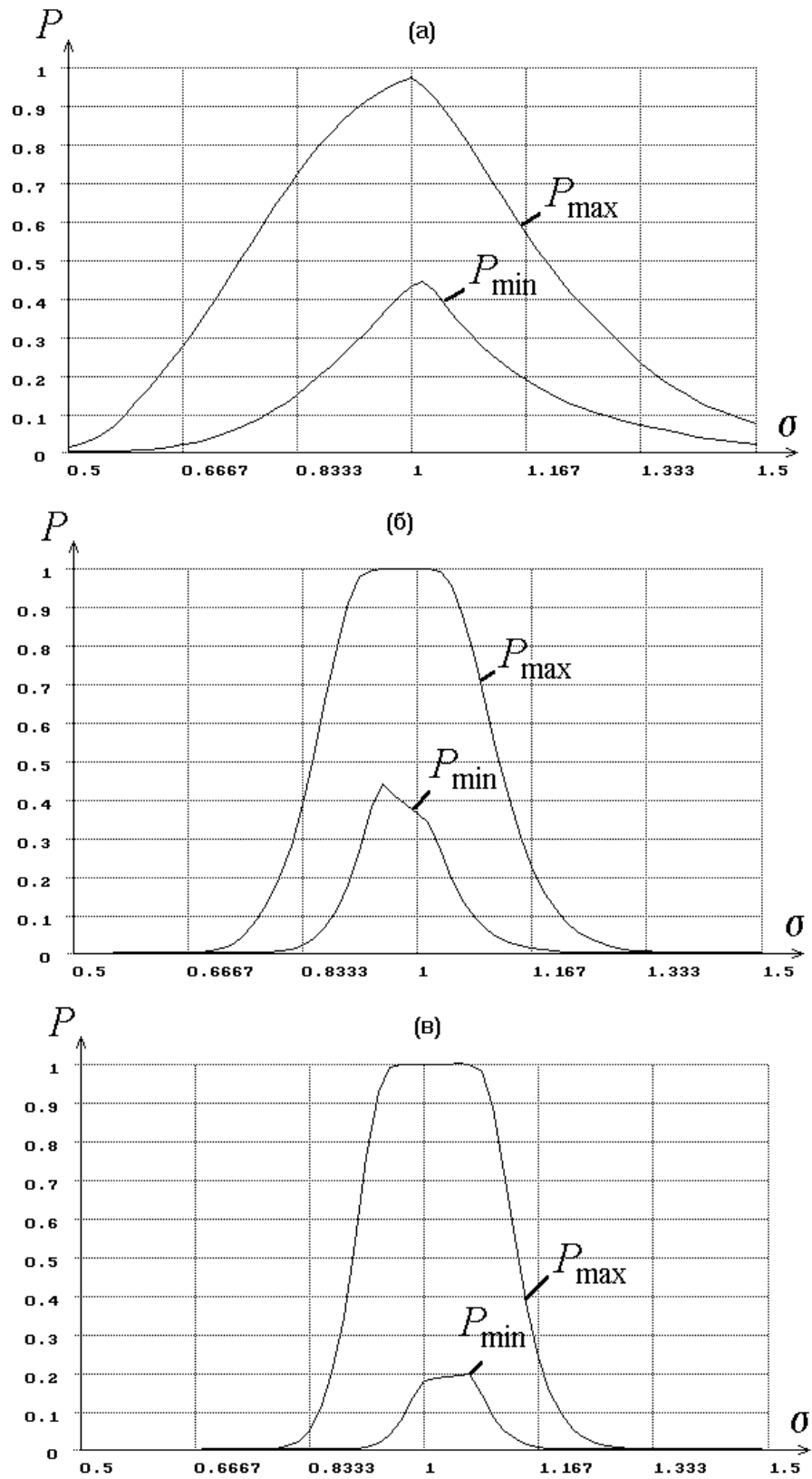


Рис. 4.5.4. Согласие интервальных выборок разного объема с нормальным распределением: (а) - $N=100$, (б) - $N=500$, (в) - $N=1000$

4.6. Идентификация закона распределения случайной величины как многокритериальная задача

Когда перед нами возникает задача выбора распределения, с которым наиболее хорошо согласуются данные экспериментов, то последовательность наших действий описывается следующим алгоритмом.

1. Ограничиваем класс распределений, из которого мы будем выбирать подходящий закон распределения вероятностей. Это может определяться физическими особенностями наблюдаемых величин, областью их определения, теоретическими предпосылками о законе распределения наблюдаемой величины.
2. Далее для выбранных распределений оцениваем параметры и проверяем гипотезы о согласии.
3. Выбираем то распределение, которое наиболее хорошо согласуется с выборкой.

При использовании общепринятой методики проверки гипотез по критериям согласия, когда гипотеза о согласии с данным распределением не отвергается, если вычисленное значение статистики не превышает критического, соответствующего заданному уровню значимости α , обычно оказывается, что нет причин отказаться от целого ряда распределений, то есть сохраняется несколько возможных альтернатив. Мы же должны остановиться на том распределении, согласие с которым наиболее хорошее.

К сожалению на практике довольно часто ограничиваются проверкой по какому-то одному критерию согласия, причем пытаются остановиться на привычном и удобном распределении. И если только проверка показывает, что нет оснований отвергнуть гипотезу о согласии, на этом и останавливаются. Хотя более глубокий анализ показывает, что выбранный закон далеко не лучшим образом описывает наблюдаемые данные, что еще по отношению к целому ряду распределений может быть принято то же решение и с большим основанием в качестве модели должен быть выбран другой закон.

С развитием вычислительной техники и внедрением в практику статистического анализа программного обеспечения при проверке гипотез о согласии уже не сравнивается полученное значение статистики с критическим, а вычисляется вероятность вида $P\{S > S^*\}$, где S^* - значение статистики, вычисленное по данной выборке. По величине этой вероятности можно надежней судить о степени согласия.

Если проверку гипотез осуществляют по ряду критериев, то для каждой используемой статистики $S_i, i = \overline{1, m}$, вычисляются вероятности вида $P\{S_i > S_i^*\} = \int_{S_i^*}^{\infty} g_i(s(H_0))ds$, где S_i^* - найденное по выборке значение соответствующей статистики, $g_i(s(H_0))$ - функция плотности распределения статистики S_i при условии, что гипотеза H_0 является истинной.

Значение S_i^* является функционалом, зависящим от выборки, закона распределения и оценок его параметров, т.е. $S_i^* = S_i^*(\bar{X}, f(x, \hat{\theta}))$, где через $\bar{X}$ обозначена выборка случайной величины. Допустим, что на основании первичных предположений мы выделили множество законов распределений, для которых имеет смысл рассматривать принадлежность к ним анализируемой выборки, пронумеровали эти законы, обозначив через R множество индексов функций плотности $f_j(x, \hat{\theta}), j \in R$, оценили по данной выборке параметры законов распределений, вычислили значения статистик $S_{ij}^* = S_i^*(\bar{X}, f_j(x, \hat{\theta}))$ и вероятности $P\{S_i > S_{ij}^*\} = \alpha_{ij}$. Тогда при проверке гипотезы о согласии с j -м распределением по i -му критерию, если $\alpha_{ij} > \alpha$, где α - задаваемый исследователем уровень значимости, то нет повода отвергать гипотезу о согласии с j -м распределением в соответствии с i -м критерием. Пусть в соответствии с используемыми критериями нет оснований отвергать гипотезу о согласии с множеством законов, помеченных индексами из $R_i \subset R$. Тогда мы должны выбрать тот закон распределения случайной величины $f_i(x, \hat{\theta})$, для которого $\forall i \alpha_{il} = \max_{j \in R_i} \alpha_{ij}$.

Обычно такой вывод можно сделать однозначно, если рассматриваются достаточно существенно отличающиеся законы распределений. Однако вполне возможно (и это бывает довольно часто для различных, но близких законов распределения), что выводы по различным критериям указывают на предпочтительность различных законов, т.е. при общем "согласии" по всем критериям один критерий указывает на предпочтительность одного закона, второй - на предпочтительность некоторого другого и т.д. Это означает, что решения задачи выбора распределения по различным критериям не совпадают, то есть, используя различные критерии мы приходим к различным

решениям [134]. Такая "несогласованность" объясняется различием мер, используемых в критериях. Но решение о выборе закона для описания случайной величины надо принимать.

Возникает естественная многокритериальная задача принятия решения. Так как все критерии измеряются в единой шкале, то решить её можно, сформировав простой компромиссный критерий вида

$$\max_{j \in R_1} \sum_{i=1}^m \omega_i \alpha_{ij},$$

где ω_i - весовой коэффициент i -го критерия, $\sum_{i=1}^m \omega_i = 1$.

Некоторой неопределённостью здесь обладает лишь процедура выбора весовых коэффициентов. При использовании результатов, полученных в данной главе, и корректным применением всех критериев согласия эти веса могут быть равными. Хотя различные критерии по разному "улавливают" в выборке различные отклонения от предположений. И в принципе, если необходимо выявление соответствующих особенностей в наблюдаемом законе, веса можно подбирать соответствующим образом. В частности, в версии программной системы [76], в которой не учитывалось изменение законов распределения статистик непараметрических критериев при оценивании по выборке параметров распределения, а в критериях отношения правдоподобия и χ^2 Пирсона использовалось асимптотически оптимальное группирование, весовые коэффициенты для критериев Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса брались равными, но в 3 раза меньшими, чем для критериев χ^2 и отношения правдоподобия. Эти значения были выбраны на основании многочисленных опытных данных. Последующие исследования предельных законов статистик критериев Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса показали, что этот выбор оказался достаточно удачным, так как примерно в 3 раза завышались вероятности $P\{S > S^*\}$, если не учитывалась потеря этими критериями свойства "свободы от распределения" и вероятности вычислялись в соответствии с законами $K(s)$, χ^2 , $a1(s)$ и $a2(s)$.

Вообще говоря, оценки параметров, определяемые различными методами, не совпадают друг с другом. Естественно, что значения статистик, вычисленных при проверке гипотез о согласии, в этом случае также будут отличаться. Например, если для данной выборки оценивать параметры распределения по группированным данным, то оценке по методу минимума χ^2 соответствует минимум статистики

χ^2 , оценке максимального правдоподобия соответствует минимум статистики отношения правдоподобия. Если мы используем критерий согласия χ^2 , но оценки параметров распределения находили методом максимального правдоподобия, то значение статистики χ^2 будет больше, чем в том случае, если бы мы использовали оценки, найденные по методу минимума χ^2 .

Последним замечанием хотелось ещё раз подчеркнуть необходимость принятия решения при выборе закона, наиболее хорошо описывающего выборочные данные, по совокупности критериев согласия, так как они по разному видят отклонения от проверяемой гипотезы. Это позволит сократить количество необоснованных решений, принимаемых на практике, особенно выбор без достаточных оснований популярных законов распределения (типа нормального), заставит расширять круг параметрических и непараметрических моделей для описания реальных случайных величин.

4.7. Выводы

1. Показана связь мощности критериев согласия с информационной матрицей Фишера по группированным данным. Показано, что применение асимптотически оптимального группирования данных в критериях согласия отношения правдоподобия, χ^2 Пирсона и типа χ^2 обеспечивает максимальную мощность этих критериев при близких конкурирующих гипотезах.

Построенные таблицы асимптотически оптимального группирования используются в реализованной программной системе в критериях отношения правдоподобия и χ^2 Пирсона при проверке гипотез о согласии. В тех случаях, когда решение задачи асимптотически оптимального группирования неинвариантно относительно параметров распределения, задача асимптотически оптимального группирования решается непосредственно перед проверкой согласия.

Экспериментально подтверждено, что при использовании асимптотически оптимального группирования критерии отношения правдоподобия и χ^2 Пирсона лучше, чем непараметрические критерии Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса улавливают малые отклонения от предположений в наблюдаемых данных.

2. Экспериментально исследованы предельные распределения статистик критериев отношения правдоподобия и χ^2 Пирсона в зависимости от способа группирования.

В случае проверки простой гипотезы распределения статистик критериев отношения правдоподобия и χ^2 Пирсона при справедливой гипотезе H_0 достаточно хорошо согласуются с χ^2_{k-1} как при равновероятном, так и при асимптотически оптимальном группировании.

В случае сложной гипотезы и вычислении ОМП по негруппированным данным предельные распределения статистик критериев отношения правдоподобия и χ^2 Пирсона существенно зависят от способа группирования и хорошо согласуются с гамма-распределениями. При этом:

- Зависимость от способа группирования особенно сильно проявляется при малом числе интервалов.
- Распределения статистик зависят не только от количества оцененных по выборке параметров, но и от того, какой параметр оценивался. Оценивание параметра сдвига приводит к более значительному изменению распределений статистик, чем оценивание масштабного параметра.
- В целом, при малом числе интервалов и оценивании p параметров число степеней свободы предельного распределения уменьшается на “число степеней свободы” меньшее p . При этом эмпирический закон распределения статистики при асимптотически оптимальном группировании ближе к теоретическому χ^2_{k-p-1} -распределению, чем при равновероятном группировании.
- Эмпирические распределения статистики отношения правдоподобия оказываются ближе к предельному теоретическому χ^2_{k-p-1} -распределению, чем соответствующие эмпирические распределения статистики χ^2 Пирсона.
- С ростом числа интервалов k разность между функцией распределения χ^2_{k-p-1} и действительными функциями распределения статистик отношения правдоподобия и χ^2 Пирсона в случае асимптотически оптимального группирования убывает существенно быстрее.

- Расчеты показали, что при использовании для вычисления $P\{S > S^*\}$ функции распределения χ^2_{k-p-1} при асимптотически оптимальном группировании и малом числе интервалов ($k < 10$) при малых $P\{S > S^*\}$ погрешность имеет величины, которые не существенны для практических задач.
- С ростом количества интервалов (при соответствующем объеме выборки) отличие предельных распределений от распределений χ^2_{k-p-1} и при асимптотически оптимальном и при равновероятном группировании становится несущественным.

Распределения статистик S_{χ^2} и S_{ω} при справедливой гипотезе H_1 (как в случае проверки простой, так и в случае проверки сложной гипотезы) сильно зависят от варианта группирования. При этом разность $G(S(H_0)) - G(S(H_1))$ при близких альтернативах и асимптотически оптимальном группировании максимальна, следовательно, максимальна мощность критерия.

3. Методами статистического моделирования исследованы предельные распределения статистик непараметрических критериев Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса при проверке сложных гипотез и вычислении оценок параметров по негруппированным данным. Показано, что на основании результатов моделирования могут быть получены хорошие приближения предельных распределений непараметрических статистик.

Показано, что для законов, определяемых только параметрами сдвига и масштаба наиболее значительное влияние на предельные распределения всех непараметрических статистик оказывает оценивание параметра сдвига, в существенно меньшей степени - оценивание масштабного параметра.

Показано, что достаточно хорошая аппроксимация для реальных распределений статистик непараметрических критериев обычно может быть получена с использованием логарифмически нормального распределения и/или гамма-распределения. Для ряда законов распределения случайных величин идентифицированы законы распределения статистик непараметрических критериев при различном количестве оцененных параметров, построены таблицы верхних процентных точек. Полученные распределения при практическом использовании критериев согласия позволяют делать более надежные статистические выводы.

Разработанная методика и программное обеспечение, вообще говоря, позволяют проводить аналогичное моделирование распределений статистик непараметрических критериев для любого закона распределения случайной величины.

Результаты проведенных исследований показывают, что предельные распределения статистик непараметрических критериев согласия Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса при оценивании по выборке параметров в случае справедливости гипотезы H_0 настолько сильно отличаются от распределений классических статистик, соответственно законов $K(s)$, χ^2 , $a1(s)$ и $a2(s)$, что это полностью исключает возможность использования последних в такой ситуации.

4. Рассмотрено использование критериев согласия в случае частично группированных и интервальных наблюдений (при простой гипотезе). Рассмотрены модификации параметрических критериев согласия χ^2 Пирсона и отношения правдоподобия и непараметрических критериев Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса, при которых находятся нижние и верхние оценки соответствующих статистик при частично группированных или интервальных наблюдениях. На основании этих оценок строится интервал неопределенности вероятности согласия $[p_{\min}, p_{\max}]$, позволяющий при заданном уровне значимости делать статистические выводы. Экспериментально исследована зависимость интервала неопределенности от точности наблюдений и объема выборки. Показано, что выводы становятся менее определенными при уменьшении точности наблюдений, причем эта неопределенность может увеличиваться с ростом количества наблюдений. При увеличении объема интервальной выборки длина интервала неопределенности растет и стремится к единице для множества априори допустимых для описания данной случайной величины распределений. Это не позволяет сделать однозначный вывод в пользу какой-то параметрической модели. Следовательно, если условия регистрации наблюдений позволяют проводить их измерение лишь с какой-то ограниченной точностью, то существует разумный объем выборки, превышение которого ничего не дает ни для точности оценивания параметров по такой выборке, ни для проверки гипотез о согласии.

В случае сложных гипотез, когда по интервальной выборке необходимо оценивать параметры закона, использование рассмотрен-

ных модификаций критериев проблематично, так как предельные распределения статистик (или их оценок снизу и сверху) будут, по-видимому, зависеть от метода оценивания и, возможно, конкретного вида выборки. Поэтому, даже найдя оценки снизу и сверху для используемой статистики, нельзя будет вычислить вероятности $p_{\max}$ и $p_{\min}$.

5. Из-за различия используемых мер выводы по различным критериям при близких альтернативах довольно часто указывают на предпочтительность различных законов: решения задачи выбора распределения по различным критериям не совпадают. Поэтому задача идентификации закона распределения, наиболее хорошо описывающего наблюдаемую выборку, сформулирована как естественная многокритериальная задача принятия решения по простому компромиссный критерию (все критерии измеряются в единой шкале). Этот подход использован в программной системе при реализации режима “Идентификация” и позволяет выбирать параметрическую модель по совокупности критериев согласия.

5. ПРОГРАММНАЯ СИСТЕМА “СТАТИСТИЧЕСКИЙ АНАЛИЗ ОДНОМЕРНЫХ НАБЛЮДЕНИЙ СЛУЧАЙНЫХ ВЕЛИЧИН”

5.1. Назначение программной системы

Программная система предназначена для всестороннего статистического анализа наблюдений одномерных непрерывных случайных величин.

Все задачи статистического анализа данных рассматриваются с точки зрения *наиболее общего представления* экспериментальных наблюдений в виде частично группированных выборок, частными случаями которых являются негруппированные, группированные и цензурированные выборки. Это обеспечивает проведение анализа реальных данных независимо от того, в каком виде они регистрировались и в каком виде представлены для обработки.

Система позволяет оценивать параметры законов распределения, проверять гипотезы о согласии по критериям χ^2 Пирсона, отношения правдоподобия, Колмогорова, Смирнова, критериям ω^2 и Ω^2 Мизеса, отбраковывать аномальные наблюдения, идентифицировать закон распределения, выполнять группирование и сортировку выборки, моделировать выборки, подчиненные заданным законам распределения. Режим идентификации обеспечивает выбор из рассматриваемого множества законов распределения, наиболее хорошо согласующегося с исходной выборкой.

Основным методом оценивания параметров распределений, заложенным в системе, является метод максимального правдоподобия. Система позволяет получать оценки параметров 26 наиболее часто используемых в приложениях непрерывных законов и семейств распределений: экспоненциального, полунормального, Рэлея, Максвелла, модуля многомерного нормального вектора, Парето, Эрланга, Лапласа, нормального, логарифмически нормальных (ln и lg), Коши, Вейбулла, Накагами, распределения минимального значения, распределения максимального значения, двойного показательного, гамма-распределения, логистического, бета-распределения 1-го рода, стандартного бетараспределения 2-го рода, бета-распределения 2-го рода, распределений Sb -Джонсона, Sl -Джонсона и Su -Джонсона, экспоненциального семейства распределений.

Графическое обеспечение позволяет наглядно отображать итоги статистического анализа с построением функции плотности и гистограммы или теоретической и эмпирической функций распределения найденного закона с отображением результатов проверки гипотез по различным критериям.

Развитое меню и встроенная справочная система максимально облегчают освоение комплекса неподготовленным пользователем.

По ряду возможностей система не имеет аналогов среди программного обеспечения задач статистического анализа, опирается на ряд оригинальных результатов. *Во-первых*, исходная выборка может быть негруппированной, группированной или частично группированной. *Во-вторых*, при проверке гипотез по критериям χ^2 Пирсона и отношения правдоподобия используется асимптотически оптимальное группирование данных, обеспечивающее максимальную мощность критериев при близких альтернативных гипотезах. *В-третьих*, группирование исходной выборки с последующим оцениванием параметров распределения позволяет получать робастные оценки, менее чувствительные к аномальным результатам наблюдений и к отклонениям наблюдаемого закона распределения от предполагаемого. *В-четвертых*, использование робастных оценок позволило реализовать эффективную параметрическую процедуру отбраковки аномальных ошибок измерений. *В-пятых*, использование совокупности критериев согласия позволяет делать более надежные выводы при идентификации распределения по формируемому компромиссному критерию. *В-шестых*, при оценивании параметров сдвига используются робастные оптимальные оценки по выборочным квантилям, построенные на базе таблиц асимптотически оптимального группирования. *В-седьмых*, применение системы при обработке результатов дает возможность использовать оптимальное группирование при проведении экспериментов и регистрации их результатов, что позволяет резко сократить объем хранимых и передаваемых по каналам связи данных без существенной потери информации о законе распределения наблюдаемой случайной величины. Возможности системы, особенности решаемых задач, использование в приложениях изложены в работах [27, 73-76, 91, 92, 94].

Программное обеспечение может использоваться в любой прикладной области, связанной с регистрацией наблюдений: при испытаниях на надежность, при контроле качества, при обработке измерительной информации и аттестации измерительных приборов, при ана-

лизе результатов социологических обследований, при обработке наблюдений физических, медицинских, биологических и других экспериментов. Программное обеспечение успешно используется в учебном процессе ВУЗов.

5.2. Алгоритмы вычисления оценок максимального правдоподобия

Уравнения правдоподобия в большинстве случаев оказываются слишком сложными, чтобы можно было рассчитывать на их явное решение, а для группированных и частично группированных выборок их решение может быть найдено только в численном виде. Исключения составляют лишь частные случаи для некоторых распределений, когда в группированной выборке вся информация разбита на 2 интервала. Полученные выражения для таких оценок приведены в [42]. Однако такие ситуации не имеют практического значения.

Для решения уравнений правдоподобия при вычислении ОМП скалярных параметров используется модификация метода Ньютона, в котором очередное приближение оценки вычисляется по формуле

$$\theta^{k+1} = \theta^k - \lambda \frac{\partial \ln L(\theta^k)}{\partial \theta} / \frac{\partial^2 \ln L(\theta^k)}{\partial \theta^2}.$$

Следует подчеркнуть, что левая часть уравнения правдоподобия $\partial \ln L(\theta) / \partial \theta$ обычно представляет собой такую зависимость, что метод Ньютона в чистом виде сходится только при очень хорошем начальном приближении.

При оценивании векторного параметра распределения осуществляется поиск максимума логарифма функции правдоподобия по области определения параметров

$$\ln L(\theta) \xrightarrow{\theta \in \Omega} \max.$$

В качестве основного метода поиска выбран метод сопряженных градиентов. Используются модификации Флетчера-Ривса [175] и Пшеничного [137]. Модификация Флетчера-Ривса при поиске максимума функции n переменных $f(\bar{x})$ описывается следующей последовательностью действий. Начиная с некоторого начального приближения $\bar{x}^0$, очередные приближения находятся по формуле

$$\bar{x}^{k+1} = \bar{x}^k + \lambda^k \bar{S}^k,$$

где $\bar{S}^0 = \nabla f(\bar{x}^0)$, при $k \neq 0$

$$\bar{S}^k = \nabla f(\bar{x}^k) + \omega^k \bar{S}^{k-1},$$

$$\omega^k = \frac{(\nabla f(\bar{x}^k))^2}{(\nabla f(\bar{x}^{k-1}))^2},$$

Модификация Пшеничного отличается способом вычислением значения ω^k

$$\omega^k = -\frac{\nabla f(\bar{x}^k) [\nabla f(\bar{x}^k) - \nabla f(\bar{x}^{k-1})]}{\nabla f(\bar{x}^{k-1}) \bar{S}^{k-1}}.$$

Величина λ^k вычисляется в результате решения задачи $\max_{\lambda} f(\bar{x}^k + \lambda \bar{S}^k)$. После $(n+1)$ -й итерации процедура циклически повторяется с заменой $\bar{x}^0$ на $\bar{x}^{n+1}$. Алгоритм заканчивается, как только выполнится соотношение $(\nabla f(\bar{x}^k))^2 < \varepsilon$, где ε - заданная точность. Экспериментальное сравнение этих алгоритмов показали, что модификация Пшеничного обладает более высокой скоростью сходимости при приближении максимизируемой функции к квадратичной. Это означает в общем случае более высокую скорость сходимости в районе экстремума, где обычно оказывается удовлетворительной квадратичная аппроксимация.

Таким образом, для вычисления ОМП скалярных параметров необходимы выражения для $\partial \ln L(\theta) / \partial \theta$ и $\partial^2 \ln L(\theta) / \partial \theta^2$, а для вычисления ОМП векторных параметров - $\ln L(\theta)$ и $\nabla \ln L(\theta)$, компонентами которого являются левые части соответствующих уравнений правдоподобия, т.е. $\partial \ln L(\theta) / \partial \theta_l$, где θ_l - l -я компонента вектора параметров.

Очевидно, что найденная в результате максимизации логарифма функции правдоподобия ОМП векторного параметра будет являться решением соответствующей системы уравнений правдоподобия.

Последующий материал данного раздела охватывает распределения, включенные в программную систему. Изложение связано с реализацией программного обеспечения вычисления оценок максимального правдоподобия и использованием критериев согласия. В каждом случае приводятся плотность закона распределения, выражения для $\partial \ln L(\theta) / \partial \theta$ и $\partial^2 \ln L(\theta) / \partial \theta^2$, необходимые при вычислении ОМП скалярных параметров, выражения для $\ln L(\theta)$ и $\nabla \ln L(\theta)$, компонентами которого являются левые части соответствующих уравнений правдоподобия, требуемые при вычислении ОМП векторных параметров. Далее, так как качество находимых оценок определяется асимптотической дисперсией, описываемой в общем случае частично группированной выборки соотношением

$$D(\hat{\theta}) = M_{PG}^{-1},$$

где M_{PG} - информационная матрица Фишера по частично группированной выборке, элементы которой J_{sl} определяются соотношением

$$J_{sl} = \sum_{(1)} n_i \times \sum_{(1)} \left(\frac{\partial \ln P_i(\hat{\theta})}{\partial \theta_s} \frac{\partial \ln P_i(\hat{\theta})}{\partial \theta_l} \right) P_i(\hat{\theta}) + \\ + \sum_{(2)} n_i \times \int_{-\infty}^{\infty} \left(\frac{\partial \ln f(x, \hat{\theta})}{\partial \theta_s} \frac{\partial \ln f(x, \hat{\theta})}{\partial \theta_l} \right) f(x, \hat{\theta}) dx,$$

то приводятся выражения и для элементов этой матрицы.

При изложении материала приводятся лишь сами выражения, используемые при вычислении оценок и их дисперсий, при решении задач асимптотически оптимального группирования. При необходимости отражаются моменты, связанные с численным интегрированием и вычислением специальных функций.

В процессе разработки в программное обеспечение включено большинство непрерывных законов распределения, используемых на практике для описания выборочных данных. Эти законы заимствованы из различных литературных источников, но наиболее полно их сводка представлена в работе [28].

5.3. Алгоритмическое обеспечение программной системы

5.3.1. Экспоненциальное распределение

Функция плотности экспоненциального (показательного) распределения имеет вид

$$f(x) = \theta_0 e^{-\theta_0(x-\theta_1)},$$

где $\theta_0 \in (0, \infty)$ - основной параметр распределения, $\theta_1 \in (-\infty, \infty)$ - параметр сдвига, область определения случайной величины - $x \in [\theta_1, \infty)$.

Функция распределения - $F(x) = 1 - e^{-\theta_0(x-\theta_1)}$.

Параметр θ_0 оценивается по методу максимального правдоподобия, а параметр θ_1 при необходимости оценивается с использованием асимптотически оптимальных порядковых статистик. Уравнение правдоподобия по θ_0 имеет вид

$$\frac{1}{\theta_0} \left\{ \sum_{(1)} n_i \frac{t_i e^{-t_i} - t_{i-1} e^{-t_{i-1}}}{e^{-t_{i-1}} - e^{-t_i}} + \sum_{(2)} \sum_{j=1}^{n_i} (1 - t_{ij}) \right\} = 0,$$

где $t_i = \theta_0(x_i - \theta_1)$, $t_{ij} = \theta_0(x_{ij} - \theta_1)$. Вторая производная по θ_0 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \left[\frac{t_{i-1}^2 e^{-t_{i-1}} - t_i^2 e^{-t_i}}{e^{-t_{i-1}} - e^{-t_i}} - \left(\frac{t_i e^{-t_i} - t_{i-1} e^{-t_{i-1}}}{e^{-t_{i-1}} - e^{-t_i}} \right)^2 \right] - \sum_{(2)} n_i \right\}.$$

Количество информации Фишера о параметре θ_0 в частично группированной выборке

$$J_0 = J(\theta_0) = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{(t_i e^{-t_i} - t_{i-1} e^{-t_{i-1}})^2}{e^{-t_{i-1}} - e^{-t_i}} + \sum_{(2)} n_i \right\}.$$

Асимптотическая дисперсия оценки θ_0 равна $D(\theta_0) = J^{-1}(\theta_0)$.

Результаты анализа на смоделированной негруппированной выборке объёмом 1000 измерений с параметрами $\theta_0 = 1$, $\theta_1 = 0$ при оценивании только параметра θ_0 в графическом виде представлены на рис. 5.3.1.

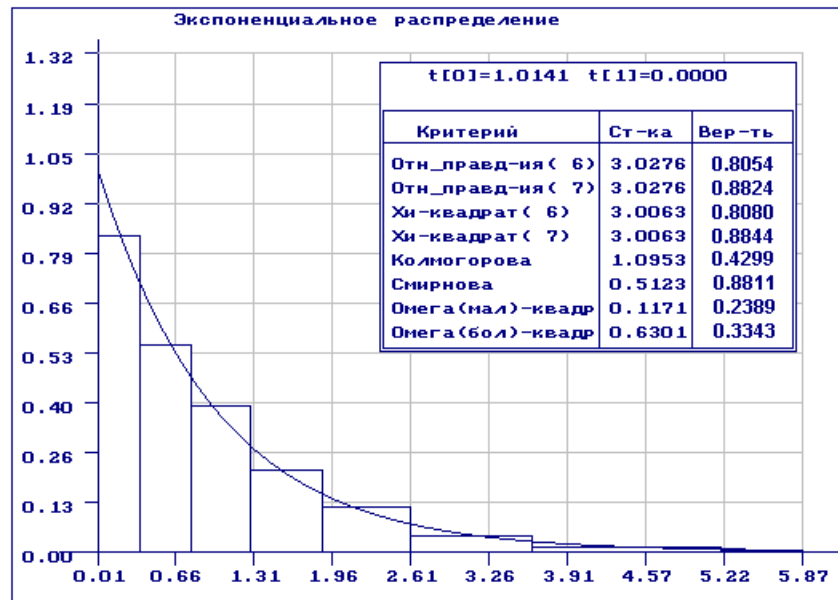


Рис. 5.3.1. Результаты оценивания параметров показательного распределения и проверки гипотез

5.3.2. Полунормальное распределение

Распределение модуля нормальной величины (полунормальное) имеет функцию плотности

$$f(x) = \frac{2}{\theta_0 \sqrt{2\pi}} e^{-(x-\theta_1)^2/2\theta_0^2},$$

где параметр $\theta_0 \in (0, \infty)$, параметр сдвига $\theta_1 \in (-\infty, \infty)$ и аргумент $x \in (\theta_1, \infty)$.

Парметр θ_0 находится как решение уравнения правдоподобия

$$\frac{1}{\theta_0} \left\{ \sum_{(1)} n_i \frac{t_{i-1} e^{-t_{i-1}^2/2} - t_i e^{-t_i^2/2}}{\sqrt{\frac{\pi}{2}} \{\Phi_0(t_i) - \Phi_0(t_{i-1})\}} + \sum_{(2)} \sum_{j=1}^{n_i} (t_{ij}^2 - 1) \right\} = 0,$$

где $t_i = (x_i - \theta_1)/\theta_0$, $t_{ij} = (x_{ij} - \theta_1)/\theta_0$, $\Phi_0(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-t^2/2} dt$, и функция распределения $F(x) = \Phi_0(t)$. Вторая производная по θ_0 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \left[\frac{(t_{i-1}^2 - 2)t_{i-1} e^{-t_{i-1}^2/2} - (t_i^2 - 2)t_i e^{-t_i^2/2}}{\sqrt{\frac{\pi}{2}} \{\Phi_0(t_i) - \Phi_0(t_{i-1})\}} - \right. \right. \\ \left. \left. - \left(\frac{t_{i-1} e^{-t_{i-1}^2/2} - t_i e^{-t_i^2/2}}{\sqrt{\frac{\pi}{2}} \{\Phi_0(t_i) - \Phi_0(t_{i-1})\}} \right)^2 \right] + \sum_{(2)} \sum_{j=1}^{n_i} (1 - 3t_{ij}^2) \right\}.$$

Количество информации Фишера о параметре θ_0 в частично группированной выборке

$$J_0 = J(\theta_0) = \frac{2}{\theta_0^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{(t_{i-1} e^{-t_{i-1}^2/2} - t_i e^{-t_i^2/2})^2}{\pi \{\Phi_0(t_i) - \Phi_0(t_{i-1})\}} + \sum_{(2)} n_i \right\}.$$

Асимптотическая дисперсия оценки θ_0 равна $D(\theta_0) = J^{-1}(\theta_0)$. Параметр сдвига θ_1 при необходимости оценивается с использованием асимптотически оптимальных порядковых статистик.

На рис. 5.3.2а приведен пример с результатами моделирования и последующего оценивания параметра θ_0 по негруппированным данным. Моделировалась выборка объемом 1000 наблюдений со значением $\theta_0 = 2$.

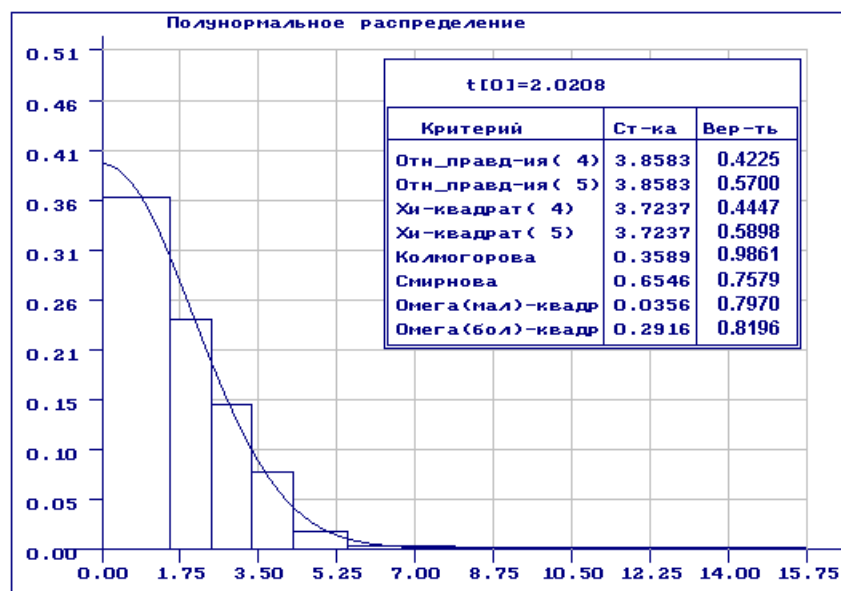


Рис. 5.3.2а. Результаты анализа по негруппированным данным

А на рис. 5.3.2б результаты анализа по той же выборке, но предварительно выборка группировалась. Как видим, существенной разницы в оценках не наблюдается.

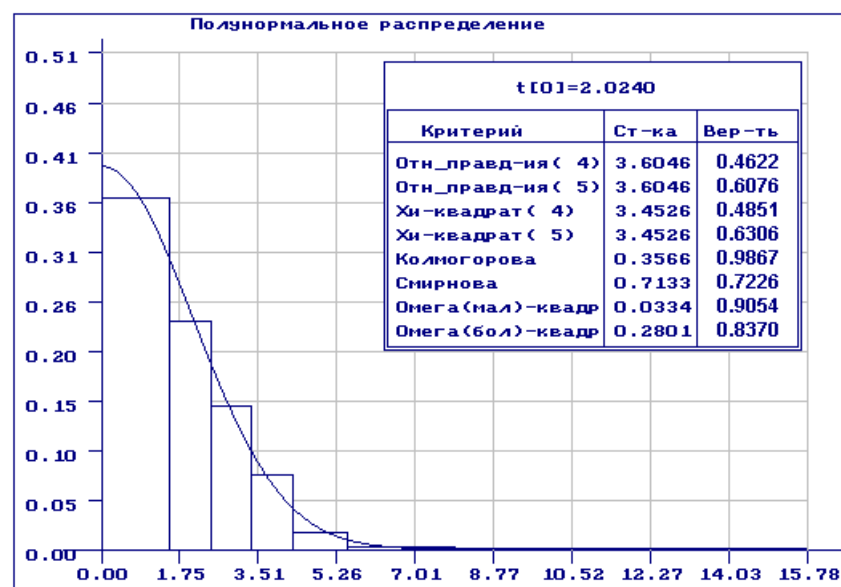


Рис. 5.3.2б. Результаты анализа по группированным данным

5.3.3. Распределение Рэлея

Распределение Рэлея имеет функцию плотности

$$f(x) = \frac{(x - \theta_1)}{\theta_0^2} e^{-(x - \theta_1)^2 / 2\theta_0^2},$$

где параметр распределения $\theta_0 \in (0, \infty)$, параметр сдвига $\theta_1 \in (-\infty, \infty)$, аргумент $x \in [\theta_1, \infty)$, а функция распределения $F(x) = 1 - e^{-(x - \theta_1)^2 / 2\theta_0^2}$.

Параметр сдвига θ_1 при необходимости оценивается с использованием асимптотически оптимальных порядковых статистик.

Оценка максимального правдоподобия параметра определяется как решение уравнения правдоподобия

$$\frac{1}{\theta_0} \left\{ \sum_{(1)} n_i \frac{t_{i-1}^2 e^{-t_{i-1}^2/2} - t_i^2 e^{-t_i^2/2}}{e^{-t_{i-1}^2/2} - e^{-t_i^2/2}} + \sum_{(2)} \sum_{j=1}^{n_i} (t_{ij}^2 - 2) \right\} = 0,$$

где $t_i = (x_i - \theta_1) / \theta_0$, $t_{ij} = (x_{ij} - \theta_1) / \theta_0$. Вторая производная по θ_0 определяется выражением

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \left[\frac{(t_{i-1}^2 - 3)t_{i-1}^2 e^{-t_{i-1}^2/2} - (t_i^2 - 2)t_i^2 e^{-t_i^2/2}}{e^{-t_{i-1}^2/2} - e^{-t_i^2/2}} - \right. \right. \\ \left. \left. - \left(\frac{t_{i-1}^2 e^{-t_{i-1}^2/2} - t_i^2 e^{-t_i^2/2}}{e^{-t_{i-1}^2/2} - e^{-t_i^2/2}} \right)^2 \right] + \sum_{(2)} \sum_{j=1}^{n_i} (2 - 3t_{ij}^2) \right\}. \end{aligned}$$

Количество информации Фишера о параметре θ_0 в частично группированной выборке

$$J_0 = J(\theta_0) = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{(t_{i-1}^2 e^{-t_{i-1}^2/2} - t_i^2 e^{-t_i^2/2})^2}{e^{-t_{i-1}^2/2} - e^{-t_i^2/2}} + 4 \sum_{(2)} n_i \right\}.$$

Асимптотическая дисперсия оценки θ_0 равна $D(\theta_0) = J^{-1}(\theta_0)$.

На рис. 5.3.3 представлены результаты анализа выборки объёмом 1000 наблюдений, смоделированной по закону Рэлея с параметром $\theta_0 = 3$.

5.3.4. Распределение Максвелла

Распределение Максвелла имеет функцию плотности

$$f(x) = \frac{2(x - \theta_1)^2}{\theta_0^3 \sqrt{2\pi}} e^{-\frac{(x - \theta_1)^2}{2\theta_0^2}},$$

где параметр $\theta_0 > 0$ и аргумент $x \in [\theta_1, \infty)$. Функция распределения определяется выражением $F(x) = \Phi_0(t) - 2t\varphi(t)$, где $t = (x - \theta_1) / \theta_0$,

$$\Phi_0(t) = 2 \int_0^t \varphi(y) dy, \quad \varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$

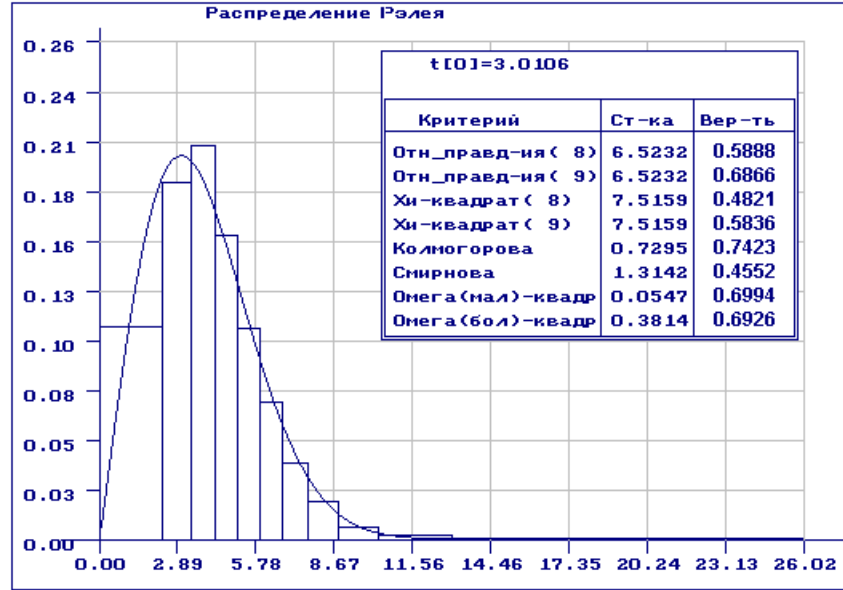


Рис. 5.3.3. Плотность и гистограмма по результатам анализа

Методом максимального правдоподобия оценивается параметр θ_0 . Оценка получается в качестве единственного корня уравнения правдоподобия

$$\frac{1}{\theta_0} \left\{ \sum_{(1)} n_i \frac{2t_{i-1}^3 \varphi(t_{i-1}) - 2t_i^3 \varphi(t_i)}{\Phi_0(t_i) - 2t_i \varphi(t_i) - \Phi_0(t_{i-1}) + 2t_{i-1} \varphi(t_{i-1})} + \sum_{(2)} \sum_{j=1}^{n_i} (t_{ij}^2 - 3) \right\} = 0,$$

где $t_i = (x_i - \theta_1) / \theta_0$, $t_{ij} = (x_{ij} - \theta_1) / \theta_0$. Вторая производная по θ_0 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \left[\frac{(t_{i-1}^2 - 4)t_{i-1}^3 \varphi(t_{i-1}) - (t_i^2 - 4)t_i^3 \varphi(t_i)}{\Phi_0(t_i) - 2t_i \varphi(t_i) - \Phi_0(t_{i-1}) + 2t_{i-1} \varphi(t_{i-1})} - \right. \right. \\ \left. \left. - 4 \left(\frac{t_{i-1}^3 \varphi(t_{i-1}) - t_i^3 \varphi(t_i)}{\Phi_0(t_i) - 2t_i \varphi(t_i) - \Phi_0(t_{i-1}) + 2t_{i-1} \varphi(t_{i-1})} \right)^2 \right] + 3 \sum_{(2)} \sum_{j=1}^{n_i} (1 - t_{ij}^2) \right\}.$$

Количество информации Фишера о параметре θ_0 в частично группированной выборке

$$J(\theta_0) = \frac{1}{\theta_0^2} \left\{ 4 \sum_{(1)} n_i \times \sum_{(1)} \frac{[t_{i-1}^3 \varphi(t_{i-1}) - t_i^3 \varphi(t_i)]^2}{\Phi_0(t_i) - 2t_i \varphi(t_i) - \Phi_0(t_{i-1}) + 2t_{i-1} \varphi(t_{i-1})} + 6 \sum_{(2)} n_i \right\}.$$

Асимптотическая дисперсия оценки θ_0 равна $D(\theta_0) = J^{-1}(\theta_0)$.

Параметр сдвига θ_1 при необходимости оценивается с использованием асимптотически оптимальных порядковых статистик.

На рис. 5.3.4 приведены результаты оценивания параметра θ_0 по смоделированной для данного распределения выборке со значением параметров $\theta_0 = 2, \theta_1 = 0$.

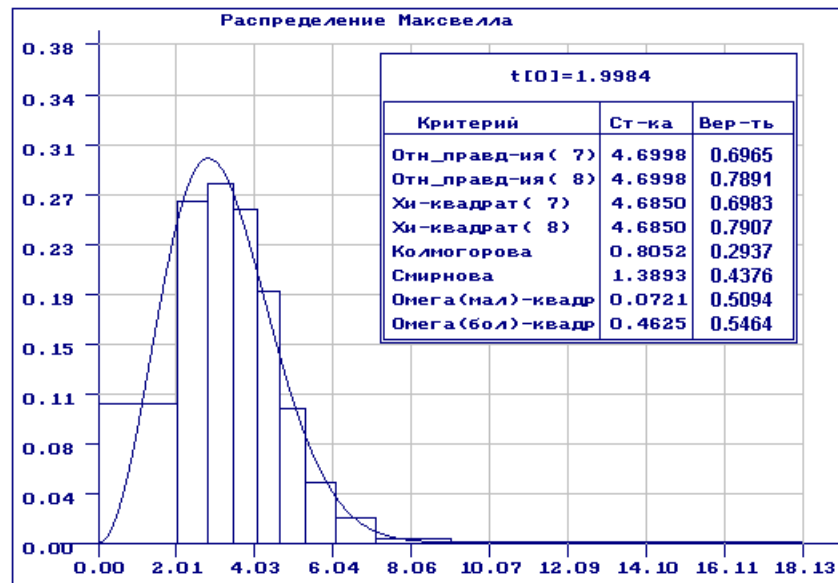


Рис. 5.3.4. Результаты оценивания параметра распределения Максвелла и последующего анализа

5.3.5. Распределение модуля многомерной нормальной величины

Распределение модуля многомерной нормальной случайной величины порядка m имеет функцию плотности

$$f(x) = \frac{2(x - \theta_1)^{m-1}}{(2\theta_0^2)^{m/2} \Gamma(m/2)} e^{-(x - \theta_1)^2 / 2\theta_0^2},$$

где $\theta_0 \in (0, \infty)$, $\theta_1 \in (-\infty, \infty)$, $x \in [\theta_1, \infty)$, $\Gamma(\cdot)$ - гамма-функция Эйлера. Частными случаями этого распределения являются полунормальное распределение ($m=1$), распределение Рэлея ($m=2$), распределение Максвелла ($m=3$). Оценка параметра θ_0 получается в качестве единственного решения уравнения правдоподобия

$$\frac{1}{\theta_0} \left\{ \sum_{(1)} n_i \frac{t_{i-1}^m e^{-t_{i-1}^2/2} - t_i^m e^{-t_i^2/2}}{2^{m/2-1} \Gamma(m/2) [F(t_i) - F(t_{i-1})]} + \sum_{(2)} \sum_{j=1}^{n_i} (t_{ij}^2 - m) \right\} = 0,$$

где $t_i = (x_i - \theta_1) / \theta_0$, $t_{ij} = (x_{ij} - \theta_1) / \theta_0$, а функция распределения при четном m определяется выражением

$$F(t) = 1 - e^{-t^2/2} \left(1 + \frac{t^2}{2} + \frac{t^4}{8} + \dots + \frac{t^{m-2}}{2^{m/2-1} \Gamma(m/2)} \right),$$

а при нечетном -

$$F(t) = \Phi_0(t) - e^{-t^2/2} \left(\frac{t}{2^{1/2} \Gamma(3/2)} + \frac{t^3}{2^{3/2} \Gamma(5/2)} + \dots + \frac{t^{m-2}}{2^{m/2-1} \Gamma(m/2)} \right),$$

и $\Phi_0(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-t^2/2} dt$. Вторая производная по θ_0 определяется выражением

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \left[\frac{(t_{i-1}^2 - m - 1)t_{i-1}^m e^{-t_{i-1}^2/2} - (t_i^2 - m - 1)t_i^m e^{-t_i^2/2}}{\xi [F(t_i) - F(t_{i-1})]} - \right. \right. \\ \left. \left. - \left(\frac{t_{i-1}^m e^{-t_{i-1}^2/2} - t_i^m e^{-t_i^2/2}}{\xi [F(t_i) - F(t_{i-1})]} \right)^2 \right] + \sum_{(2)} \sum_{j=1}^{n_i} (m - 3t_{ij}^2) \right\}, \end{aligned}$$

где $\xi = 2^{m/2-1} \Gamma(m/2)$. Количество информации Фишера о параметре θ_0 в частично группированной выборке

$$J_0 = J(\theta_0) = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{(t_{i-1}^m e^{-t_{i-1}^2/2} - t_i^m e^{-t_i^2/2})^2}{\xi^2 [F(t_i) - F(t_{i-1})]} + 2m \sum_{(2)} n_i \right\}.$$

Асимптотическая дисперсия оценки θ_0 равна $D(\theta_0) = J^{-1}(\theta_0)$.

Параметр сдвига θ_1 при необходимости оценивается с использованием асимптотически оптимальных порядковых статистик.

На рис. 5.3.5 приведены результаты анализа выборки, смоделированной со значением параметра $\theta_0 = 2$ при $m = 7$ и нулевом параметре сдвига. ОМП параметра θ_0 по негруппированным наблюдениям $\hat{\theta}_0 = 1.9964$, по предварительно сгруппированным - $\hat{\theta}_{0\Gamma} = 1.9987$.



Рис. 5.3.5. Результаты проверки согласия выборки, моделированной с распределением модуля 7-мерного нормального вектора

5.3.6. Распределение Парето

Распределение Парето имеет функцию плотности

$$f(x) = \theta_0 \theta_1^{\theta_0} x^{-\theta_0-1},$$

где параметр $\theta_0 \in (0, \infty)$, $\theta_1 \in (0, \infty)$, $x \in [\theta_1, \infty)$. Оценивается по методу максимального правдоподобия только параметр θ_0 , параметр θ_1 задается или оценивается с использованием порядковых статистик. Оценка θ_0 получается в качестве единственного корня уравнения правдоподобия

$$\frac{1}{\theta_0} \left\{ \sum_{(1)} n_i \frac{t_{i-1} \ln t_{i-1} - t_i \ln t_i}{t_{i-1} - t_i} + \sum_{(2)} \sum_{j=1}^{n_j} (1 + \ln t_{ij}) \right\} = 0,$$

где $t_i = (x_i / \theta_1)^{-\theta_0}$, а $t_{ij} = (x_{ij} / \theta_1)^{-\theta_0}$. Вторая производная по θ_0 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \left[\frac{t_{i-1} \ln^2 t_{i-1} - t_i \ln^2 t_i}{t_{i-1} - t_i} - \left(\frac{t_{i-1} \ln t_{i-1} - t_i \ln t_i}{t_{i-1} - t_i} \right)^2 \right] - \sum_{(2)} n_i \right\}.$$

Количество информации Фишера о параметре θ_0 в частично группированной выборке

$$J(\theta_0) = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{(t_{i-1} \ln t_{i-1} - t_i \ln t_i)^2}{t_{i-1} - t_i} + \sum_{(2)} n_i \right\}.$$

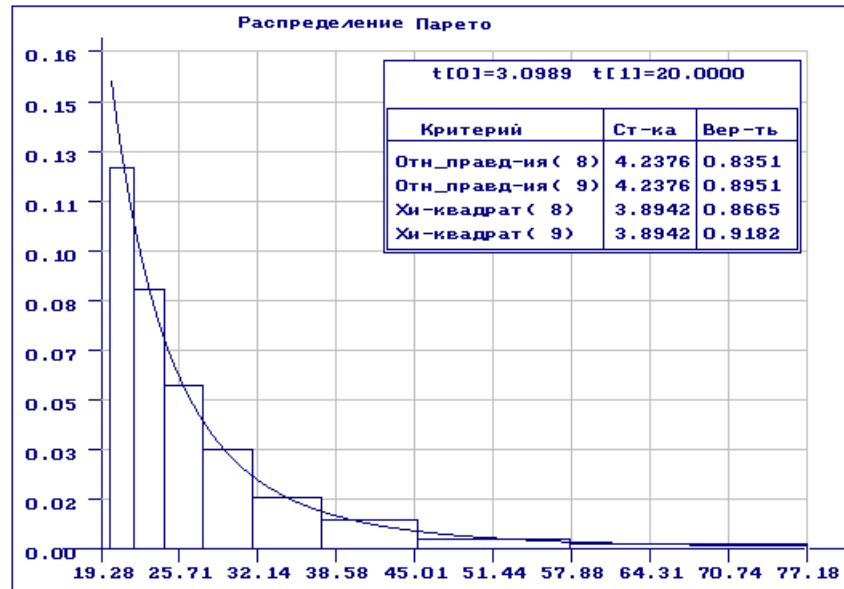


Рис. 5.3.6. Результаты анализа выборки, смоделированной по закону Парето

На рис. 5.3.6 приведены результаты анализа выборки объемом 1000 наблюдений, смоделированной по распределению Парето с параметрами $\theta_0 = 3$, $\theta_1 = 20$. При проверке согласия использовано асимптотически оптимальное группирование.

5.3.7. Распределение Эрланга

Распределение Эрланга порядка m имеет функцию плотности

$$f(x) = \frac{\theta_1^m}{\Gamma(m)} (x - \theta_2)^{m-1} e^{-\theta_1(x-\theta_2)},$$

где параметр $\theta_1 > 0$ и аргумент $x \in (\theta_2, \infty)$, θ_2 - параметр сдвига. Порядок распределения m интерпретируется как параметр θ_0 . Параметр θ_2 при необходимости оценивается с использованием порядковых статистик. Распределение Эрланга является частным случаем гамма-распределения.

Оценка параметра θ_1 определяется как решение уравнения правдоподобия

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{t_i e^{-t_i} - t_{i-1} e^{-t_{i-1}}}{\int_{t_{i-1}}^{t_i} t^{m-1} e^{-t} dt} + \sum_{(2)} \sum_{j=1}^{n_i} (m - t_{ij}) \right\} = 0,$$

где $t_i = \theta_1(x_i - \theta_2)$, $t_{ij} = \theta_1(x_{ij} - \theta_2)$. Вторая производная по θ_1 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{t_i^m (m-1-t_i) e^{-t_i} - t_{i-1}^m (m-1-t_i) e^{-t_{i-1}}}{\int_{t_{i-1}}^{t_i} t^{m-1} e^{-t} dt} - \left(\frac{t_i^m e^{-t_i} - t_{i-1}^m e^{-t_{i-1}}}{\int_{t_{i-1}}^{t_i} t^{m-1} e^{-t} dt} \right)^2 \right] - m \sum_{(2)} n_i \right\}.$$

Количество информации Фишера о параметре θ_1 в частично группированной выборке

$$J(\theta_1) = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{(t_i^m e^{-t_i} - t_{i-1}^m e^{-t_{i-1}})^2}{\Gamma(m) \int_{t_{i-1}}^{t_i} t^{m-1} e^{-t} dt} + m \sum_{(2)} n_i \right\}.$$

На рис. 5.3.7 приведены результаты анализа выборки, распределенной в соответствии с законом Эрланга. Исследуемая выборка моделировалась с параметром $\theta_1 = 2$ и $m = 4$. В процессе проверки согласия по критериям отношения правдоподобия и χ^2 Пирсона находилось и использовалось решение задачи асимптотически оптимального группирования.

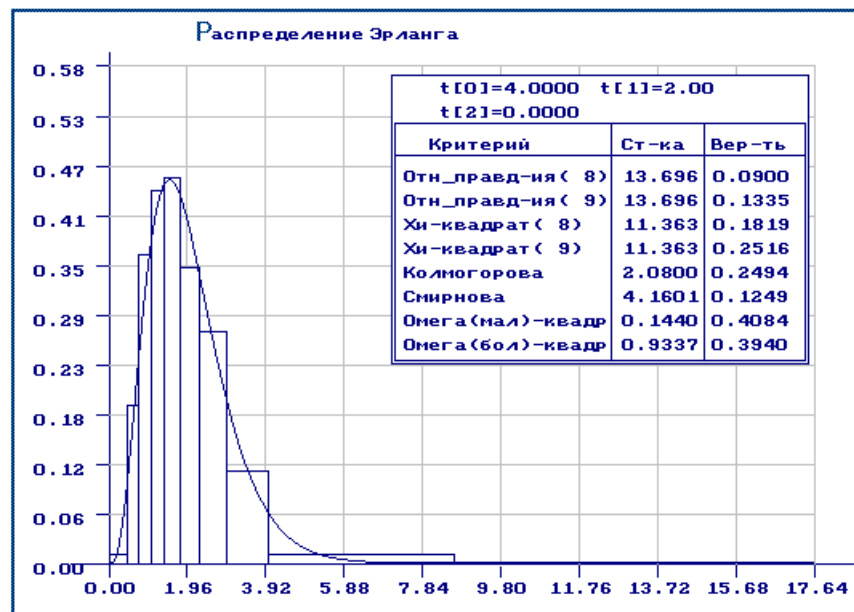


Рис. 5.3.7. Результаты проверки согласия смоделированной выборки с распределением Эрланга

5.3.8. Распределение Лапласа

Распределение Лапласа, или двустороннее экспоненциальное имеет функцию плотности

$$f(x) = \frac{\theta_0}{2} e^{-\theta_0(x-\theta_1)},$$

где параметр $\theta_0 \in (0, \infty)$, параметр сдвига $\theta_1 \in (-\infty, \infty)$, $x \in (-\infty, \infty)$.

Параметр сдвига θ_1 при необходимости оценивается с использованием асимптотически оптимальных квантилей или по методу моментов. Оценка параметра θ_0 определяется как единственное решение уравнения правдоподобия

$$\sum_{(1)} n_i \frac{\partial P_i(\theta) / \partial \theta_0}{P_i(\theta)} + \frac{1}{\theta_0} \sum_{(1)} \sum_{j=1}^{n_i} (1 - t_{ij}) = 0,$$

где $t_i = \theta_0(x_i - \theta_1)$, $t_{ij} = \theta_0(x_{ij} - \theta_1)$,

$$P_i(\theta) = \begin{cases} \frac{1}{2}(e^{t_i} - e^{t_{i-1}}), & t_i < 0; \\ \frac{1}{2}(e^{-t_{i-1}} - e^{-t_i}), & t_{i-1} > 0; \\ 1 - \frac{1}{2}(e^{-t_i} + e^{t_{i-1}}), & t_{i-1} < 0 < t_i; \end{cases}$$

$$\frac{1}{\theta_0} \frac{\partial P_i(\theta)}{\partial \theta_0} = \begin{cases} \frac{1}{2}(t_i e^{t_i} - t_{i-1} e^{t_{i-1}}), & t_i < 0; \\ \frac{1}{2}(-t_{i-1} e^{-t_{i-1}} + t_i e^{-t_i}), & t_{i-1} > 0; \\ \frac{1}{2}(t_i e^{-t_i} - t_{i-1} e^{t_{i-1}}), & t_{i-1} < 0 < t_i. \end{cases}$$

Вторая производная по θ_0 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \sum_{(1)} n_i \left[\frac{\partial^2 P_i(\theta)}{\partial \theta_0^2} - \left(\frac{\partial P_i(\theta)}{\partial \theta_0} \right)^2 / P_i(\theta) \right] / P_i(\theta) - \frac{1}{\theta_0^2} \sum_{(2)} n_i,$$

где
$$\frac{\theta_0^2 \partial^2 P_i(\theta)}{\partial \theta_0^2} = \begin{cases} \frac{1}{2}(t_i^2 e^{t_i} - t_{i-1}^2 e^{t_{i-1}}), t_i < 0; \\ \frac{1}{2}(t_{i-1}^2 e^{-t_{i-1}} - t_i^2 e^{-t_i}), t_{i-1} > 0; \\ \frac{1}{2}(-t_i^2 e^{-t_i} - t_{i-1}^2 e^{t_{i-1}}), t_{i-1} < 0 < t_i. \end{cases}$$

Количество информации Фишера в частично группированной выборке о параметре θ_0 имеет вид

$$J(\theta_0) = \sum_{(1)} n_i \times \sum_{(1)} \left(\frac{\partial P_i(\theta)}{\partial \theta_0} \right)^2 / P_i(\theta) + \frac{1}{\theta_0^2} \sum_{(2)} n_i,$$

а асимптотическая дисперсия оценки θ_0 равна $D(\theta_0) = J^{-1}(\theta_0)$.

Результаты моделирования выборки по закону Лапласа и последующего её анализа представлены на рис. 5.3.8. Выборка объёмом 1000 наблюдений моделировалась с параметрами $\theta_0 = 2, \theta_1 = 0$. Оценивался только масштабный параметр. Этот факт учитывался при вычислении вероятностей согласия $P\{S > S^*\}$ в критериях типа Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса.



Рис. 5.3.8. Анализ выборки, распределённой по закону Лапласа

5.3.9. Нормальное распределение

Нормальное распределение имеет функцию плотности

$$f(x) = \frac{1}{\theta_1 \sqrt{2\pi}} e^{-\frac{(x-\theta_0)^2}{2\theta_1^2}},$$

где математическое ожидание - параметр $\theta_0 \in (-\infty, \infty)$, среднее квадратичное отклонение - параметр $\theta_1 \in (0, \infty)$, $x \in (-\infty, \infty)$.

Оцениваются либо сразу оба параметра, либо один из них. Оценка параметра θ_0 определяется как единственное решение уравнения правдоподобия

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{\varphi(t_{i-1}) - \varphi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} \sum_{j=1}^{n_i} t_{ij} \right\} = 0,$$

где $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$, $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-t^2/2} dt$, $t_i = \frac{x_i - \theta_0}{\theta_1}$, $t_{ij} = \frac{x_{ij} - \theta_0}{\theta_1}$.

Вторая производная логарифма функции правдоподобия по параметру θ_0 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{t_{i-1}\varphi(t_{i-1}) - t_i\varphi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} - \left(\frac{\varphi(t_{i-1}) - \varphi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} \right)^2 \right] - \sum_{(2)} n_i \right\}.$$

Оценка параметра θ_1 находится из уравнения правдоподобия

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{t_{i-1}\varphi(t_{i-1}) - t_i\varphi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} \sum_{j=1}^{n_i} (t_{ij}^2 - 1) \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_1^2} = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{t_i(1-t_i^2)\varphi(t_i) - t_{i-1}(1-t_{i-1}^2)\varphi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} - \right. \right. \\ \left. \left. - \left(\frac{t_{i-1}\varphi(t_{i-1}) - t_i\varphi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} \right)^2 \right] + \sum_{(2)} \sum_{j=1}^{n_i} (1 - 3t_{ij}^2) \right\}. \end{aligned}$$

При одновременном оценивании обоих параметров максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln [\Phi(t_i) - \Phi(t_{i-1})] - \sum_{(2)} \sum_{j=1}^{n_i} \left(\frac{t_{ij}^2}{2} + \ln \theta_1 + \ln \sqrt{2\pi} \right)$$

и оценки являются решением системы уравнений правдоподобия.

Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[\varphi(t_{i-1}) - \varphi(t_i)]^2}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} n_i \right\},$$

$$J_{11} = J(\theta_1) = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[t_{i-1}\varphi(t_{i-1}) - t_i\varphi(t_i)]^2}{\Phi(t_i) - \Phi(t_{i-1})} + 2 \sum_{(2)} n_i \right\},$$

$$J_{01} = J(\theta_0, \theta_1) = \frac{1}{\theta_1^2} \sum_{(1)} n_i \times \sum_{(1)} \frac{[\varphi(t_i) - \varphi(t_{i-1})][t_{i-1}\varphi(t_{i-1}) - t_i\varphi(t_i)]}{\Phi(t_i) - \Phi(t_{i-1})},$$

и $J_{01} = J_{10}$.

На рис. 5.3.9 приведены результаты оценивания параметров нормального распределения и последующего анализа с использованием оптимального группирования при проверке согласия по критерию χ^2 Пирсона. Выборка объёмом 1000 наблюдений моделировалась с параметрами $\theta_0 = 1, \theta_1 = 1$. Количество интервалов в критериях согласия отношения правдоподобия и χ^2 Пирсона равно 11.



Рис. 5.3.9. Гистограмма и плотность нормального распределения

5.3.10. Логарифмически (ln) нормальное распределение

Логарифмически (ln) нормальное распределение имеет функцию плотности

$$f(x) = \frac{1}{x\theta_1\sqrt{2\pi}} e^{-(\ln x - \theta_0)^2 / 2\theta_1^2},$$

где параметр сдвига $\theta_0 \in (-\infty, \infty)$, масштабный параметр $\theta_1 \in (0, \infty)$, аргумент $x \in (0, \infty)$.

Оцениваются либо сразу оба параметра, либо один из них. В данном случае справедливы все соотношения для нормального закона с заменой выражений для t_i, t_{ij} на $t_i = \frac{\ln x_i - \theta_0}{\theta_1}$, $t_{ij} = \frac{\ln x_{ij} - \theta_0}{\theta_1}$. На рис.

5.3.10 представлен результат анализа выборки объемом 1000 измерений, смоделированной по данному закону с параметрами $\theta_0 = 1$, $\theta_1 = 0.4$. Значения вычисленных по выборке ОМП параметров $\hat{\theta}_0 = 0.9912$, $\hat{\theta}_1 = 0.3908$.



Рис. 5.3.10. Результаты статистического анализа выборки, смоделированной в соответствии с логарифмически (ln) нормальным распределением

5.3.11. Логарифмически (lg) нормальное распределение

Логарифмически (lg) нормальное распределение имеет функцию плотности

$$f(x) = \frac{1}{x\theta_1 \ln 10 \sqrt{2\pi}} e^{-(\lg x - \theta_0)^2 / 2\theta_1^2},$$

где параметр сдвига $\theta_0 \in (-\infty, \infty)$, масштабный параметр $\theta_1 \in (0, \infty)$, аргумент $x \in (0, \infty)$.

Оцениваются либо сразу оба параметра, либо один из них. В данном случае также справедливы все соотношения для нормального закона, но соответственно $t_i = \frac{\lg x_i - \theta_0}{\theta_1}$, $t_{ij} = \frac{\lg x_{ij} - \theta_0}{\theta_1}$. На рис. 5.3.11 представлен результат анализа группированной выборки объемом 1000 измерений. Выборка моделировалась по закону с параметрами $\theta_0 = 1$, $\theta_1 = 0.25$, затем была сгруппирована в соответствии с асимптотически оптимальным группированием (разбита пропорционально вероятностям асимптотически оптимального группирования). Количество интервалов равно 9. Для сравнения, ОМП параметров по исходной негруппированной выборке $\hat{\theta}_0 = 0.9925$ и $\hat{\theta}_1 = 0.2436$, а по группированной - $\hat{\theta}_{ог} = 0.9925$ и $\hat{\theta}_{ог} = 0.2436$.

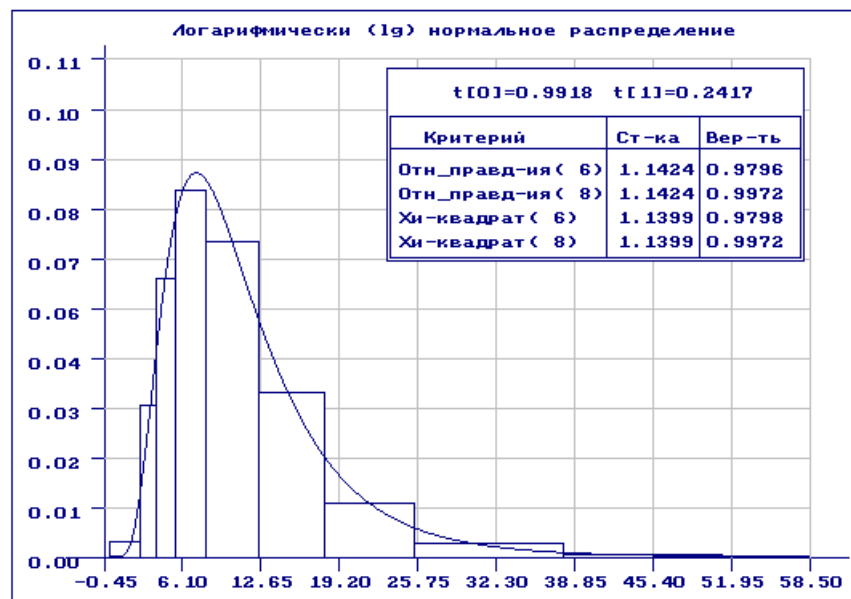


Рис. 5.3.11. Результаты статистического анализа группированной выборки, смоделированной в соответствии с логарифмически (lg) нормальным распределением

5.3.12. Распределение Коши

Распределение Коши, или распределение Брейта-Вигнера, имеет функцию плотности

$$f(x) = \frac{\theta_0}{\pi[\theta_0^2 + (x - \theta_1)^2]},$$

где $\theta_0 \in (0, \infty)$, $\theta_1 \in (-\infty, \infty)$, $x \in (-\infty, \infty)$. Распределение Коши является примером "паталогического" распределения: не существует математическое ожидание, расходится дисперсия. Поэтому, например, метод моментов совсем неприменим.

Методом максимального правдоподобия оцениваются либо сразу оба параметра, либо один из них. Оценка параметра θ_0 определяется как решение уравнения правдоподобия

$$\frac{1}{\theta_0} \left\{ \sum_{(1)} n_i \frac{\frac{t_{i-1}}{1+t_{i-1}^2} - \frac{t_i}{1+t_i^2}}{\arctgt_i - \arctgt_{i-1}} + \sum_{(2)} \sum_{j=1}^{n_i} \left(1 - \frac{2}{1+t_{ij}^2} \right) \right\} = 0,$$

где $t_i = (x_i - \theta_1) / \theta_0$, $t_{ij} = (x_{ij} - \theta_1) / \theta_0$. Вторая производная логарифма функции правдоподобия по параметру θ_0 определяется выражением

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \left[\frac{\frac{2t_i}{(1+t_i^2)^2} - \frac{2t_{i-1}}{(1+t_{i-1}^2)^2}}{\arctgt_i - \arctgt_{i-1}} - \left(\frac{\frac{t_{i-1}}{1+t_{i-1}^2} - \frac{t_i}{1+t_i^2}}{\arctgt_i - \arctgt_{i-1}} \right)^2 \right] - \right. \\ \left. - \sum_{(2)} \sum_{j=1}^{n_i} \left(1 - \frac{2(1-t_{ij}^2)}{(1+t_{ij}^2)^2} \right) \right\}. \end{aligned}$$

Оценка параметра θ_1 находится как решение уравнения правдоподобия

$$\frac{1}{\theta_0} \left\{ \sum_{(1)} n_i \frac{\frac{1}{1+t_{i-1}^2} - \frac{1}{1+t_i^2}}{\arctgt_i - \arctgt_{i-1}} + 2 \sum_{(2)} \sum_{j=1}^{n_i} \frac{t_{ij}}{1+t_{ij}^2} \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \left[\frac{\frac{2t_{i-1}}{(1+t_{i-1}^2)^2} - \frac{2t_i}{(1+t_i^2)^2}}{\arctgt_i - \arctgt_{i-1}} - \left(\frac{\frac{1}{1+t_{i-1}^2} - \frac{1}{1+t_i^2}}{\arctgt_i - \arctgt_{i-1}} \right)^2 \right] - \right. \\ \left. - 2 \sum_{(2)} \sum_{j=1}^{n_i} \frac{1-t_{ij}^2}{(1+t_{ij}^2)^2} \right\}.$$

При одновременном оценивании обоих параметров максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln(\arctgt_i - \arctgt_{i-1}) - \sum_{(2)} \sum_{j=1}^{n_i} [\ln \theta_0 + \ln(1+t_{ij}^2)] - \\ - \ln \pi \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right),$$

и оценка векторного параметра определяется решением системы уравнений правдоподобия.

Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{1}{\pi} \frac{\left[\frac{t_{i-1}}{1+t_{i-1}^2} - \frac{t_i}{1+t_i^2} \right]^2}{\arctgt_i - \arctgt_{i-1}} + \frac{1}{2} \sum_{(2)} n_i \right\},$$

$$J_{11} = J(\theta_1) = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{1}{\pi} \frac{\left[\frac{1}{1+t_{i-1}^2} - \frac{1}{1+t_i^2} \right]^2}{\arctgt_i - \arctgt_{i-1}} + \frac{1}{2} \sum_{(2)} n_i \right\},$$

$$J_{01} = J(\theta_0, \theta_1) = \frac{1}{\pi \theta_0^2} \sum_{(1)} n_i \times \sum_{(1)} \frac{\left(\frac{t_{i-1}}{1+t_{i-1}^2} - \frac{t_i}{1+t_i^2} \right) \left(\frac{1}{1+t_{i-1}^2} - \frac{1}{1+t_i^2} \right)}{\arctgt_i - \arctgt_{i-1}} \text{ и } J_{01} = J_{10}.$$

Некоторая "паталогия" данного распределения заключается и в том, что асимптотически оптимальному группированию при оценивании 2

параметров распределения или при проверке согласия по критериям Пирсона или отношения правдоподобия соответствует разбиение на интервалы равной вероятности.

На рис. 5.3.12 приведены результаты оценивания параметров распределения и последующего анализа с использованием оптимального группирования при проверке согласия по критерию χ^2 Пирсона. Выборка объёмом 1000 наблюдений моделировалась с параметрами $\theta_0 = 3, \theta_1 = 2$. Оценивание осуществлялось по сгруппированным данным.

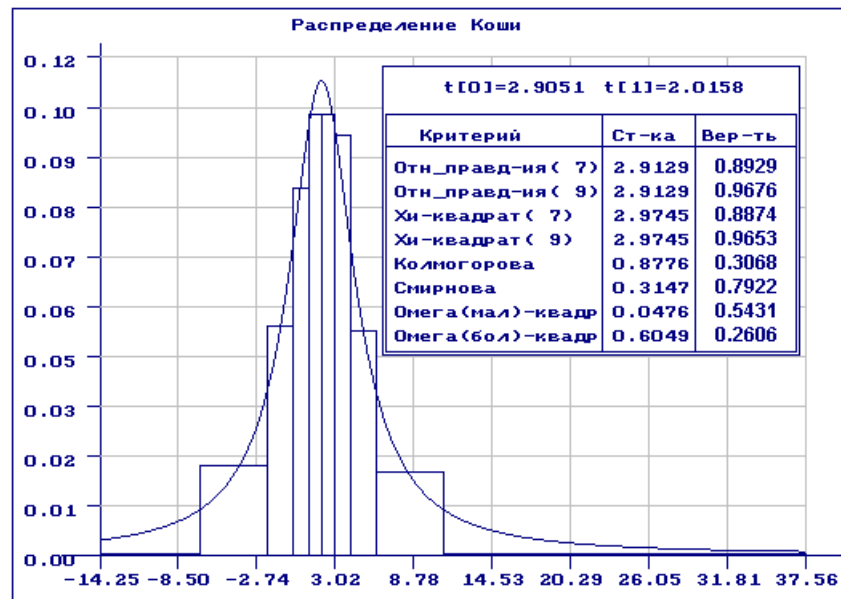


Рис. 5.3.12. Результаты статистического анализа выборки, смоделированной по распределению Коши

5.3.13. Логистическое распределение

Логистическое распределение имеет функцию плотности

$$f(x) = \frac{\pi}{\theta_1 \sqrt{3}} \frac{\exp \left\{ -\frac{\pi(x - \theta_0)}{\theta_1 \sqrt{3}} \right\}}{\left[1 + \exp \left\{ -\frac{\pi(x - \theta_0)}{\theta_1 \sqrt{3}} \right\} \right]^2},$$

где параметр сдвига $\theta_0 \in (-\infty, \infty)$, аргумент $x \in (-\infty, \infty)$, а масштабный параметр $\theta_1 \in (0, \infty)$. Оцениваются либо сразу оба параметра, либо один из них. Оценка максимального правдоподобия параметра сдвига

θ_0 получается в качестве единственного решения уравнения правдоподобия

$$\frac{\pi}{\theta_1 \sqrt{3}} \left\{ \sum_{(1)} n_i \frac{\frac{e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2} - \frac{e^{-t_i}}{(1+e^{-t_i})^2}}{(1+e^{-t_i})^{-1} - (1+e^{-t_{i-1}})^{-1}} + \sum_{(2)} \sum_{j=1}^{n_i} \left(1 - 2 \frac{e^{-t_{ij}}}{1+e^{-t_{ij}}} \right) \right\} = 0,$$

где $t_i = \pi(x_i - \theta_0) / \theta_1 \sqrt{3}$, $t_{ij} = \pi(x_{ij} - \theta_0) / \theta_1 \sqrt{3}$. Вторая производная логарифма функции правдоподобия по параметру θ_0 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{\pi^2}{3\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{\frac{e^{-t_i}(e^{-t_i}-1)}{(1+e^{-t_i})^3} - \frac{e^{-t_{i-1}}(e^{-t_{i-1}}-1)}{(1+e^{-t_{i-1}})^3}}{(1+e^{-t_i})^{-1} - (1+e^{-t_{i-1}})^{-1}} - \left(\frac{\frac{e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2} - \frac{e^{-t_i}}{(1+e^{-t_i})^2}}{(1+e^{-t_i})^{-1} - (1+e^{-t_{i-1}})^{-1}} \right)^2 \right] - \sum_{(2)} \sum_{j=1}^{n_i} \frac{2e^{-t_{ij}}}{(1+e^{-t_{ij}})^2} \right\}.$$

Оценка масштабного параметра θ_1 находится как решение уравнения правдоподобия

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{\frac{t_{i-1}e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2} - \frac{t_i e^{-t_i}}{(1+e^{-t_i})^2}}{(1+e^{-t_i})^{-1} - (1+e^{-t_{i-1}})^{-1}} + \sum_{(2)} \sum_{j=1}^{n_i} \left(t_{ij} - 1 - 2 \frac{t_{ij}e^{-t_{ij}}}{1+e^{-t_{ij}}} \right) \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \frac{\pi^2}{3\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{\frac{t_i e^{-t_i} (2 - t_i + 2e^{-t_i} + t_i e^{-t_i})}{(1+e^{-t_i})^3} - \frac{t_{i-1} e^{-t_{i-1}} (2 - t_{i-1} + 2e^{-t_{i-1}} + t_{i-1} e^{-t_{i-1}})}{(1+e^{-t_{i-1}})^3}}{(1+e^{-t_i})^{-1} - (1+e^{-t_{i-1}})^{-1}} - \right. \right.$$

$$\left. - \left(\frac{t_{i-1} e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2} - \frac{t_i e^{-t_i}}{(1+e^{-t_i})^2} \right)^2 \right] + \sum_{(2)} \sum_{j=1}^{n_i} \left(1 - \frac{2t_{ij} (1+t_{ij} e^{-t_{ij}} - e^{-2t_{ij}})}{(1+e^{-t_{ij}})^2} \right) \Bigg\}.$$

При одновременном оценивании обоих параметров максимизируется логарифм функции правдоподобия по частично группированной выборке, описываемый следующим выражением

$$\ln L = \sum_{(1)} n_i \ln \left(\frac{1}{1+e^{-t_i}} - \frac{1}{1+e^{-t_{i-1}}} \right) + \sum_{(2)} \sum_{j=1}^{n_i} \left[\ln \frac{\pi}{\theta_1 \sqrt{3}} - t_{ij} - 2 \ln(1+e^{-t_{ij}}) \right],$$

и оценка является решением системы уравнений правдоподобия.

Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \frac{\pi}{\theta_1^2 \sqrt{3}} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \left[\frac{e^{-t_i}}{(1+e^{-t_i})^2} - \frac{e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2} \right]^2 \frac{1}{(1+e^{-t_i})^{-1} - (1+e^{-t_{i-1}})^{-1}} + \frac{1}{3} \sum_{(2)} n_i \right\},$$

$$J_{11} = J(\theta_1) = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \left[\frac{t_i e^{-t_i}}{(1+e^{-t_i})^2} - \frac{t_{i-1} e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2} \right]^2 \frac{1}{(1+e^{-t_i})^{-1} - (1+e^{-t_{i-1}})^{-1}} + \frac{\pi^2 + 3}{9} \sum_{(2)} n_i \right\},$$

$$J_{01} = J(\theta_0, \theta_1) = \frac{\pi^2}{\theta_1^2 \sqrt{3}} \sum_{(1)} n_i \times \sum_{(1)} \left[\frac{t_i e^{-t_i}}{(1+e^{-t_i})^2} - \frac{t_{i-1} e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2} \right] \times \frac{1}{(1+e^{-t_i})^{-1} - (1+e^{-t_{i-1}})^{-1}} \times \left[\frac{e^{-t_i}}{(1+e^{-t_i})^2} - \frac{e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2} \right],$$

и $J_{01} = J_{10}$.

На рис. 5.3.13 приведены результаты оценивания параметров логистического распределения и последующего анализа с

использованием оптимального группирования при проверке согласия по критерию χ^2 Пирсона. Выборка объёмом 1000 наблюдений моделировалась с параметрами $\theta_0 = 1, \theta_1 = 2$.



Рис. 5.3.13. Результаты анализа выборки, смоделированной в соответствии с логистическим распределением

5.3.14. Распределение Вейбулла

Распределение Вейбулла имеет функцию плотности

$$f(x) = \frac{\theta_0 (x - \theta_2)^{\theta_0 - 1}}{\theta_1^{\theta_0}} \exp \left\{ - \left(\frac{x - \theta_2}{\theta_1} \right)^{\theta_0} \right\},$$

где основной параметр $\theta_0 \in (0, \infty)$, масштабный параметр $\theta_1 \in (0, \infty)$, и аргумент $x \in (\theta_2, \infty)$. Если параметр сдвига θ_2 не известен, то он оценивается с помощью порядковых статистик. Параметры θ_0 и θ_1 могут оцениваться либо одновременно, либо может оцениваться один из них. Оценка максимального правдоподобия параметра θ_0 определяется как решение уравнения правдоподобия

$$\frac{1}{\theta_0} \left\{ \sum_{(1)} n_i \frac{t_i e^{-t_i} \ln t_i - t_{i-1} e^{-t_{i-1}} \ln t_{i-1}}{e^{-t_{i-1}} - e^{-t_i}} + \sum_{(2)} \sum_{j=1}^{n_i} [(1 - t_{ij}) \ln t_{ij} + 1] \right\} = 0,$$

где $t_i = [(x_i - \theta_2) / \theta_1]^{\theta_0}$, $t_{ij} = [(x_{ij} - \theta_2) / \theta_1]^{\theta_0}$. Вторая производная логарифма функции правдоподобия по параметру θ_0 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \frac{\theta_0^2}{\theta_1^1} \left\{ \sum_{(1)} n_i \left[\frac{e^{-t_i} t_i (1 - t_i + 1/\theta_0) - e^{-t_{i-1}} t_{i-1} (1 - t_{i-1} + 1/\theta_0)}{e^{-t_{i-1}} - e^{-t_i}} - \left(\frac{e^{-t_i} t_i \ln t_i - e^{-t_{i-1}} t_{i-1} \ln t_{i-1}}{e^{-t_{i-1}} - e^{-t_i}} \right)^2 \right] - \sum_{(2)} \sum_{j=1}^{n_i} (t_{ij} \ln^2 t_{ij} + 1) \right\}.$$

Оценка параметра θ_1 находится как единственный корень уравнения правдоподобия

$$\frac{\theta_0}{\theta_1} \left\{ \sum_{(1)} n_i \frac{t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i}}{e^{-t_{i-1}} - e^{-t_i}} - \sum_{(2)} \sum_{j=1}^{n_i} (1 - t_{ij}) \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \frac{\theta_0^2}{\theta_1^1} \left\{ \sum_{(1)} n_i \left[\frac{e^{-t_i} t_i (1 - t_i + 1/\theta_0) - e^{-t_{i-1}} t_{i-1} (1 - t_{i-1} + 1/\theta_0)}{e^{-t_{i-1}} - e^{-t_i}} - \left(\frac{e^{-t_{i-1}} t_{i-1} - e^{-t_i} t_i}{e^{-t_{i-1}} - e^{-t_i}} \right)^2 \right] + \frac{1}{\theta_0} \sum_{(2)} \sum_{j=1}^{n_i} [1 - (\theta_0 + 1) t_{ij}] \right\}.$$

При одновременном оценивании двух параметров θ_0 и θ_1 максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln [e^{-t_{i-1}} - e^{-t_i}] - \sum_{(2)} \sum_{j=1}^{n_i} (t_{ij} - \ln t_{ij} + \ln x_{ij} - \ln \theta_0),$$

и оценки являются решением системы приведенных выше уравнений правдоподобия. Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[t_i \ln t_i e^{-t_i} - t_{i-1} \ln t_{i-1} e^{-t_{i-1}}]^2}{e^{-t_{i-1}} - e^{-t_i}} + \sum_{(2)} n_i \times \left(1 + \frac{\pi}{6} + C^2 - 2C \right) \right\},$$

$$J_{11} = J(\theta_1) = \frac{\theta_0^2}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i}]^2}{e^{-t_{i-1}} - e^{-t_i}} + \sum_{(2)} n_i \right\},$$

$$J_{01} = J(\theta_0, \theta_1) = \frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{(t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i})(t_i \ln t_i e^{-t_i} - t_{i-1} \ln t_{i-1} e^{-t_{i-1}})}{e^{-t_{i-1}} - e^{-t_i}} + \right.$$

$$+ \sum_{(2)} n_i \times (C-1) \Big\},$$

где $C = 0.577215664$ - постоянная Эйлера, и $J_{01} = J_{10}$.

На рис. 5.3.14 приведены результаты оценивания параметров распределения Вейбулла и последующего анализа с использованием оптимального группирования при проверке согласия по критериям отношения правдоподобия и χ^2 Пирсона. Выборка объёмом 1000 наблюдений моделировалась с параметрами $\theta_0 = 3, \theta_1 = 2, \theta_2 = 0$.

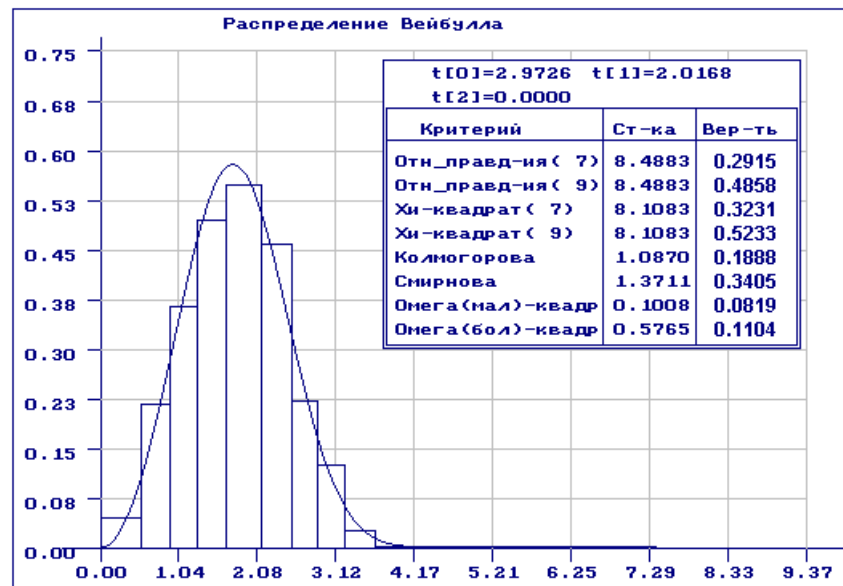


Рис. 5.3.14. Результаты анализа выборки, распределенной по закону Вейбулла

5.3.15. Распределение минимального значения

Распределение минимального значения имеет функцию плотности

$$f(x) = \frac{1}{\theta_1} \exp \left\{ \frac{x - \theta_0}{\theta_1} - \exp \left(\frac{x - \theta_0}{\theta_1} \right) \right\},$$

где параметр сдвига $\theta_0 \in (-\infty, \infty)$, масштабный параметр $\theta_1 \in (0, \infty)$, и аргумент $x \in (-\infty, \infty)$. Параметры θ_0 и θ_1 могут оцениваться либо одновременно, либо может оцениваться один из них. Оценка максимального правдоподобия параметра сдвига θ_0 определяется как решение уравнения правдоподобия

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i}}{e^{-t_{i-1}} - e^{-t_i}} + \sum_{(2)} \sum_{j=1}^{n_i} (t_{ij} - 1) \right\} = 0,$$

где $t_i = \exp\left(\frac{x_i - \theta_0}{\theta_1}\right)$, $t_{ij} = \exp\left(\frac{x_{ij} - \theta_0}{\theta_1}\right)$. Вторая производная логарифма функции правдоподобия по параметру θ_0 имеет вид

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{e^{-t_i} t_i (1 - t_i) - e^{-t_{i-1}} t_{i-1} (1 - t_{i-1})}{e^{-t_{i-1}} - e^{-t_i}} - \right. \right. \\ \left. \left. - \left(\frac{e^{-t_{i-1}} t_{i-1} - e^{-t_i} t_i}{e^{-t_{i-1}} - e^{-t_i}} \right)^2 \right] - \sum_{(2)} \sum_{j=1}^{n_i} t_{ij} \right\}. \end{aligned}$$

Оценка параметра θ_1 находится как единственный корень уравнения правдоподобия

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{s_{i-1} t_{i-1} e^{-t_{i-1}} - s_i t_i e^{-t_i}}{e^{-t_{i-1}} - e^{-t_i}} - \sum_{(2)} \sum_{j=1}^{n_i} [s_{ij} (t_{ij} - 1) - 1] \right\} = 0,$$

где $s_i = \frac{x_i - \theta_0}{\theta_1}$, $s_{ij} = \frac{x_{ij} - \theta_0}{\theta_1}$. Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_1^2} = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{s_{i-1} t_{i-1} e^{-t_{i-1}} (s_{i-1} t_{i-1} - s_{i-1} - 2) - s_i t_i e^{-t_i} (s_i t_i - s_i - 2)}{e^{-t_{i-1}} - e^{-t_i}} - \right. \right. \\ \left. \left. - \left(\frac{s_{i-1} t_{i-1} e^{-t_{i-1}} - s_i t_i e^{-t_i}}{e^{-t_{i-1}} - e^{-t_i}} \right)^2 \right] - \sum_{(2)} \sum_{j=1}^{n_i} [2s_{ij} (t_{ij} - 1) + s_{ij}^2 t_{ij} - 1] \right\}. \end{aligned}$$

При одновременном оценивании двух параметров θ_0 и θ_1 максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln [e^{-t_{i-1}} - e^{-t_i}] - \sum_{(2)} \sum_{j=1}^{n_i} (t_{ij} - s_{ij} + \ln \theta_1),$$

и оценки являются решением системы вышеприведенных уравнений правдоподобия. Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[t_{i-1}e^{-t_{i-1}} - t_i e^{-t_i}]^2}{e^{-t_{i-1}} - e^{-t_i}} + \sum_{(2)} n_i \right\},$$

$$J_{11} = J(\theta_1) = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[t_{i-1}s_{i-1}e^{-t_{i-1}} - t_i s_i e^{-t_i}]^2}{e^{-t_{i-1}} - e^{-t_i}} + \sum_{(2)} n_i \times \left(1 + \frac{\pi}{6} + C^2 - 2C \right) \right\},$$

$$J_{01} = J(\theta_0, \theta_1) = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{(t_{i-1}e^{-t_{i-1}} - t_i e^{-t_i})(t_{i-1}s_{i-1}e^{-t_{i-1}} - t_i s_i e^{-t_i})}{e^{-t_{i-1}} - e^{-t_i}} + \sum_{(2)} n_i \times (C - 1) \right\},$$

где $C = 0.577215664$ - постоянная Эйлера, и $J_{01} = J_{10}$.

На рис. 5.3.15 приведены результаты оценивания параметров распределения минимального значения и последующего анализа с использованием оптимального группирования при проверке согласия по критериям отношения правдоподобия и χ^2 Пирсона. Выборка объемом 1000 наблюдений моделировалась с параметрами $\theta_0 = 1, \theta_1 = 2$.



Рис. 5.3.15. Результаты анализа выборки, распределенной по закону минимального экстремального значения

5.3.16. Распределение максимального значения

Распределение максимального значения имеет функцию плотности

$$f(x) = \frac{1}{\theta_1} \exp \left\{ -\frac{x - \theta_0}{\theta_1} - \exp \left(-\frac{x - \theta_0}{\theta_1} \right) \right\},$$

где параметр сдвига $\theta_0 \in (-\infty, \infty)$, масштабный параметр $\theta_1 \in (0, \infty)$, и аргумент $x \in (-\infty, \infty)$. Оцениваться могут либо оба параметра, либо по одному. Оценка максимального правдоподобия параметра сдвига θ_0 определяется как единственное решение уравнения правдоподобия

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i}}{e^{-t_i} - e^{-t_{i-1}}} + \sum_{(2)} \sum_{j=1}^{n_i} (1 - t_{ij}) \right\} = 0,$$

где $t_i = \exp \left(-\frac{x_i - \theta_0}{\theta_1} \right)$, $t_{ij} = \exp \left(-\frac{x_{ij} - \theta_0}{\theta_1} \right)$. Вторая производная логарифма функции правдоподобия по параметру θ_0 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{e^{-t_{i-1}} t_{i-1} (1 - t_{i-1}) - e^{-t_i} t_i (1 - t_i)}{e^{-t_i} - e^{-t_{i-1}}} - \left(\frac{e^{-t_{i-1}} t_{i-1} - e^{-t_i} t_i}{e^{-t_i} - e^{-t_{i-1}}} \right)^2 \right] - \sum_{(2)} \sum_{j=1}^{n_i} t_{ij} \right\}.$$

Оценка параметра θ_1 находится как корень уравнения правдоподобия

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{s_i t_i e^{-t_i} - s_{i-1} t_{i-1} e^{-t_{i-1}}}{e^{-t_i} - e^{-t_{i-1}}} + \sum_{(2)} \sum_{j=1}^{n_i} [s_{ij} (t_{ij} - 1) - 1] \right\} = 0,$$

где $s_i = -\frac{x_i - \theta_0}{\theta_1}$, $s_{ij} = -\frac{x_{ij} - \theta_0}{\theta_1}$. Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{s_i t_i e^{-t_i} (s_i t_i - s_i - 2) - s_{i-1} t_{i-1} e^{-t_{i-1}} (s_{i-1} t_{i-1} - s_{i-1} - 2)}{e^{-t_i} - e^{-t_{i-1}}} - \left(\frac{s_i t_i e^{-t_i} - s_{i-1} t_{i-1} e^{-t_{i-1}}}{e^{-t_i} - e^{-t_{i-1}}} \right)^2 \right] + \sum_{(2)} \sum_{j=1}^{n_i} [1 + 2s_{ij} (1 - t_{ij}) - s_{ij}^2 t_{ij}] \right\}.$$

При одновременном оценивании двух параметров θ_0 и θ_1 максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln [e^{-t_i} - e^{-t_{i-1}}] + \sum_{(2)} \sum_{j=1}^{n_i} (s_{ij} - t_{ij} - \ln \theta_1),$$

и оценки являются решением системы приведенных выше уравнений правдоподобия. Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[t_{i-1}e^{-t_{i-1}} - t_i e^{-t_i}]^2}{e^{-t_i} - e^{-t_{i-1}}} + \sum_{(2)} n_i \right\},$$

$$J_{11} = J(\theta_1) = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[t_i s_i e^{-t_i} - t_{i-1} s_{i-1} e^{-t_{i-1}}]^2}{e^{-t_i} - e^{-t_{i-1}}} + \sum_{(2)} n_i \times \left(1 + \frac{\pi^2}{6} + C^2 - 2C \right) \right\},$$

$$J_{01} = J(\theta_0, \theta_1) = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{(t_{i-1}e^{-t_{i-1}} - t_i e^{-t_i})(t_i s_i e^{-t_i} - t_{i-1} s_{i-1} e^{-t_{i-1}})}{e^{-t_i} - e^{-t_{i-1}}} + \sum_{(2)} n_i \times (C - 1) \right\},$$

где $C = 0.577215664$ - постоянная Эйлера, и $J_{01} = J_{10}$.

На рис. 5.3.16 приведены результаты оценивания параметров распределения максимального значения по сгруппированным данным и последующего анализа с использованием оптимального группирования. Выборка объемом 1000 наблюдений моделировалась с параметрами $\theta_0 = 1, \theta_1 = 3$.

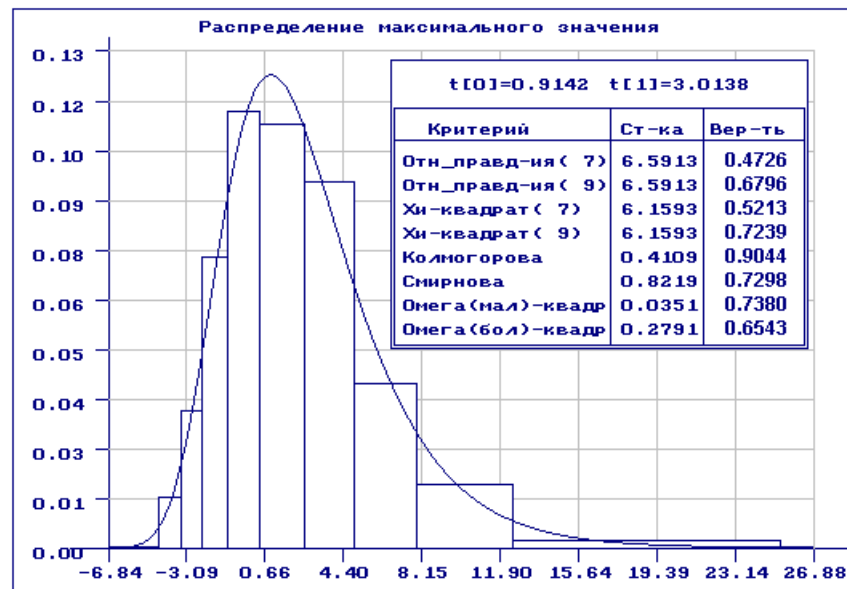


Рис. 5.3.16. Результаты анализа выборки, распределенной по закону максимального экстремального значения

5.3.17. Двойное показательное распределение

Двойное показательное распределение имеет функцию плотности

$$f(x) = \theta_0 \theta_1 \exp\left\{-(x - \theta_2)\theta_0 - \theta_1 \exp\left[-(x - \theta_2)\theta_0\right]\right\},$$

где масштабный параметр $\theta_0 \in (0, \infty)$, основной параметр $\theta_1 \in (0, \infty)$, параметр сдвига $\theta_2 \in (-\infty, \infty)$ и аргумент $x \in (-\infty, \infty)$. Оцениваться могут либо сразу два параметра θ_0 и θ_1 , либо один из них. Оценка максимального правдоподобия масштабного параметра θ_0 определяется как решение уравнения правдоподобия

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{s_i t_i e^{-t_i} - s_{i-1} t_{i-1} e^{-t_{i-1}}}{e^{-t_i} - e^{-t_{i-1}}} + \sum_{(2)} \sum_{j=1}^{n_j} \left[s_{ij} (t_{ij} - 1) + \frac{\theta_1}{\theta_0} \right] \right\} = 0,$$

где $t_i = \theta_1 \exp\{-\theta_0(x_i - \theta_2)\}$, $t_{ij} = \theta_1 \exp\{-\theta_0(x_{ij} - \theta_2)\}$, $s_i = \theta_0(x_i - \theta_2)$, $s_{ij} = \theta_0(x_{ij} - \theta_2)$. Вторая производная логарифма функции правдоподобия по параметру θ_0 имеет вид

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \left[\frac{s_i^2 t_i e^{-t_i} (t_i - 1) - s_{i-1}^2 t_{i-1} e^{-t_{i-1}} (t_{i-1} - 1)}{e^{-t_i} - e^{-t_{i-1}}} - \right. \right. \\ \left. \left. - \left(\frac{s_i t_i e^{-t_i} - s_{i-1} t_{i-1} e^{-t_{i-1}}}{e^{-t_i} - e^{-t_{i-1}}} \right)^2 \right] + \sum_{(2)} \sum_{j=1}^{n_j} [1 - s_{ij}^2 t_{ij}] \right\}. \end{aligned}$$

Оценка параметра θ_1 находится как корень уравнения правдоподобия

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i}}{e^{-t_i} - e^{-t_{i-1}}} + \sum_{(2)} \sum_{j=1}^{n_j} (1 - t_{ij}) \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{t_i^2 e^{-t_i} - t_{i-1}^2 e^{-t_{i-1}}}{e^{-t_i} - e^{-t_{i-1}}} - \left(\frac{t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i}}{e^{-t_i} - e^{-t_{i-1}}} \right)^2 \right] - \sum_{(2)} n_j \right\}.$$

При одновременном оценивании двух параметров θ_0 и θ_1 максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln[e^{-t_i} - e^{-t_{i-1}}] + \sum_{(2)} \sum_{j=1}^{n_j} (\ln \theta_0 \theta_1 - s_{ij} - t_{ij}),$$

и оценки являются решением системы приведенных выше уравнений правдоподобия.

Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[t_i s_i e^{-t_i} - t_{i-1} s_{i-1} e^{-t_{i-1}}]^2}{e^{-t_i} - e^{-t_{i-1}}} + \right. \\ \left. + \sum_{(2)} n_i \times \left(\ln^2 \theta_1 + 2(C-1) \ln \theta_1 + 1 + \frac{\pi}{6} + C^2 - 2C \right) \right\},$$

$$J_{11} = J(\theta_1) = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i}]^2}{e^{-t_i} - e^{-t_{i-1}}} + \sum_{(2)} n_i \right\},$$

$$J_{01} = J(\theta_0, \theta_1) = \frac{1}{\theta_0 \theta_1} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{(t_i s_i e^{-t_i} - t_{i-1} s_{i-1} e^{-t_{i-1}})(t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i})}{e^{-t_i} - e^{-t_{i-1}}} + \right. \\ \left. + \sum_{(2)} n_i \times (1 - C - \ln \theta_2) \right\},$$

где $C = 0.577215664$ - постоянная Эйлера, и $J_{01} = J_{10}$.

На рис. 5.3.17 приведены результаты проверки согласия смоделированной выборки с двойным показательным распределением с параметрами $\theta_0 = 3, \theta_1 = 1, \theta_2 = 0$. Объем смоделированной выборки 1000 наблюдений.

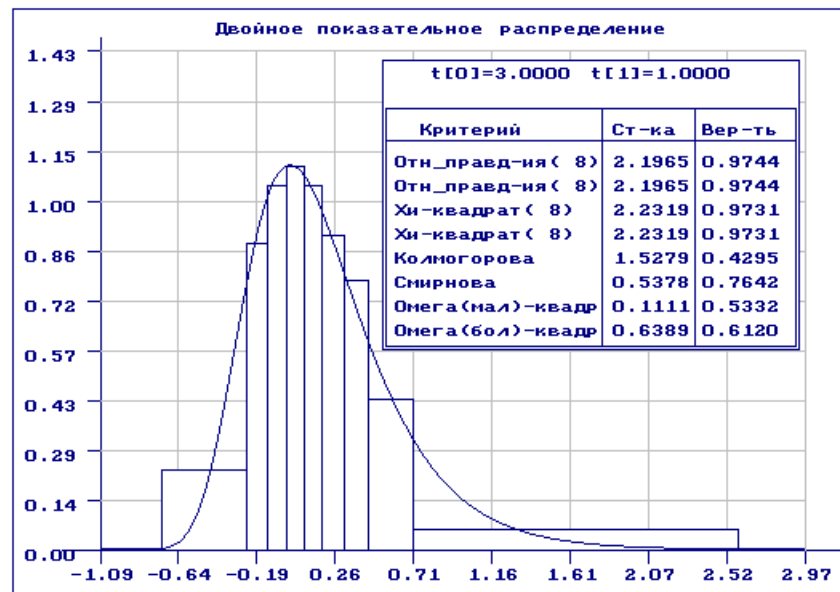


Рис. 5.3.17. Результаты проверки согласия с двойным показательным распределением

5.3.18. Распределение Накагами

Распределение Накагами имеет функцию плотности

$$f(x) = \frac{2}{\Gamma(\theta_1)} \left(\frac{\theta_1}{\theta_0^2} \right)^{\theta_1} (x - \theta_2)^{2\theta_1-1} \exp \left\{ -\frac{\theta_1 (x - \theta_2)^2}{\theta_0^2} \right\},$$

где масштабный параметр $\theta_0 \in (0, \infty)$, основной параметр $\theta_1 \in (0, \infty)$, параметр сдвига $\theta_2 \in (-\infty, \infty)$, а аргумент $x \in (\theta_2, \infty)$. Параметры θ_0 и θ_1 оцениваются методом максимального правдоподобия, а параметр θ_2 при необходимости оценивается с использованием порядковых статистик. Оценка максимального правдоподобия масштабного параметра θ_0 определяется как решение уравнения правдоподобия

$$\frac{2}{\theta_1} \left\{ \sum_{(1)} n_i \frac{t_{i-1}^{\theta_1} e^{-t_{i-1}} - t_i^{\theta_1} e^{-t_i}}{\int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} dt} + \sum_{(2)} \sum_{j=1}^{n_j} (t_{ij} - \theta_1) \right\} = 0,$$

где $t_i = \frac{\theta_1 (x_i - \theta_2)^2}{\theta_0^2}$, $t_{ij} = \frac{\theta_1 (x_{ij} - \theta_2)^2}{\theta_0^2}$. Вторая производная логарифма функции правдоподобия по параметру θ_0 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{2}{\theta_0^2} \left\{ \sum_{(1)} n_i \left[\frac{t_{i-1}^{\theta_1} e^{-t_{i-1}} (2t_{i-1} - 2\theta_1 - 1) - t_i^{\theta_1} e^{-t_i} (2t_i - 2\theta_1 - 1)}{\int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} dt} - \frac{2(t_{i-1}^{\theta_1} e^{-t_{i-1}} - t_i^{\theta_1} e^{-t_i})^2}{\int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} dt} \right] + \sum_{(2)} \sum_{j=1}^{n_j} (\theta_1 - 3t_{ij}) \right\}.$$

Оценка параметра θ_1 находится как корень уравнения правдоподобия

$$\sum_{(1)} n_i \left\{ \frac{\int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} \ln t dt + \frac{1}{\theta_1} (t_i^{\theta_1} e^{-t_i} - t_{i-1}^{\theta_1} e^{-t_{i-1}})}{\int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} dt} - \psi(\theta_1) \right\} +$$

$$+\sum_{(2)} \sum_{j=1}^{n_i} \left(\ln t_{ij} - \frac{t_{ij}}{\theta_1} \right) + (1 - \psi(\theta_1)) \sum_{(2)} n_i = 0,$$

где $\psi(\theta_1) = \Gamma'(\theta_1)/\Gamma(\theta_1)$ - логарифмическая производная гамма-функции Эйлера. Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \sum_{(1)} n_i \left\{ \frac{\int_{t_{i-1}}^{t_i} t^{\theta_1} e^{-t} \ln^2 t dt + \frac{1}{\theta_1} \left[t_i^{\theta_1} e^{-t_i} \left[2 \ln t_i + 1 - \frac{1+t_i}{\theta_1} \right] - t_{i-1}^{\theta_1} e^{-t_{i-1}} \left[2 \ln t_{i-1} + 1 - \frac{1+t_{i-1}}{\theta_1} \right] \right]}{\int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} dt} - \left(\frac{\int_{t_{i-1}}^{t_i} t^{\theta_1} e^{-t} \ln t dt + \frac{1}{\theta_1} (t_i^{\theta_1} e^{-t_i} - t_{i-1}^{\theta_1} e^{-t_{i-1}})}{\int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} dt} \right)^2 \right\} + \frac{1}{\theta_1} \sum_{(2)} n_i - \psi'(\theta_1) \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right).$$

При одновременном оценивании двух параметров θ_0 и θ_1 максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln \left(\int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} dt \right) + \sum_{(2)} \sum_{j=1}^{n_i} (\ln 2 + \theta_0 \ln t_{ij} - t_{ij} - \ln x_{ij}) - \ln \Gamma(\theta_1) \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right),$$

и оценки являются решением системы приведенных уравнений правдоподобия. Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \frac{4}{\theta_0^2} \left\{ \sum_{(1)} n_i \times \frac{1}{\Gamma(\theta_1)} \sum_{(1)} \frac{[t_{i-1}^{\theta_1} e^{-t_{i-1}} - t_i^{\theta_1} e^{-t_i}]^2}{\int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} dt} + \theta_1 \sum_{(2)} n_i \right\},$$

$$J_{11} = J(\theta_1) = \sum_{(1)} n_i \times \frac{1}{\Gamma(\theta_1)} \sum_{(1)} \left\{ \int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} \ln t dt + \frac{1}{\theta_1} (t_i^{\theta_1} e^{-t_i} - t_{i-1}^{\theta_1} e^{-t_{i-1}}) - \frac{[\psi(\theta_1) \int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} dt]^2}{\int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} dt} + \left(\psi'(\theta_1) - \frac{1}{\theta_1} \right) \sum_{(2)} n_i \right\},$$

$$J_{01} = J(\theta_0, \theta_1) = \sum_{(1)} n_i \times \frac{2}{\theta_0 \Gamma(\theta_1)} \sum_{(1)} \left\{ \int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} \ln t dt + \frac{1}{\theta_1} (t_i^{\theta_1} e^{-t_i} - t_{i-1}^{\theta_1} e^{-t_{i-1}}) - \right. \\ \left. - \psi(\theta_1) \int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} dt \right\} \left(t_{i-1}^{\theta_1} e^{-t_{i-1}} - t_i^{\theta_1} e^{-t_i} \right) / \int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} dt,$$

и $J_{01} = J_{10}$.

На рис. 5.3.18 приведены результаты статистического анализа выборки, моделированной в соответствии с распределением Накагами с параметрами $\theta_0 = 3, \theta_1 = 1, \theta_2 = 0$. Объём выборки 1000 наблюдений. Асимптотически оптимальное группирование при проверке согласия осуществлялось непосредственно в процессе проверки гипотез.

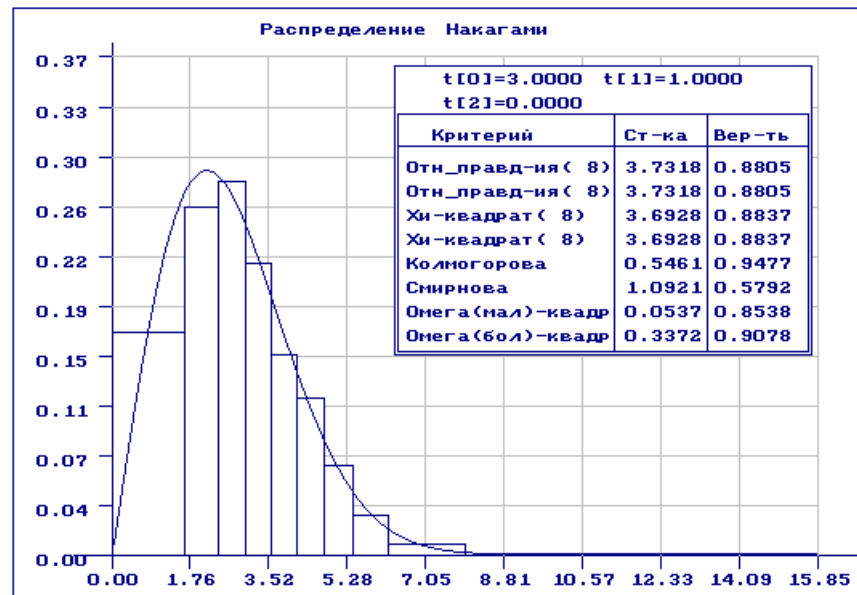


Рис. 5.3.18. Результаты проверки согласия смоделированной выборки с распределением Накагами

Вычисление интегралов.

Интегралы, содержащиеся в вышеприведенных формулах, могут находиться только численно. При $\theta_1 < 1$ и $t \rightarrow 0$ подынтегральные функции $\rightarrow -\infty$. Если не учитывать эту особенность, то погрешности численного интегрирования могут непредсказуемо возрастать, что приведет в свою очередь к некорректности задач оценивания параметров и асимптотически оптимального группирования.

1. При значениях $\theta_1 > 1$ интеграл вида $\int_0^{t_1} t^{\theta_1} e^{-t} dt$ (неполная гамма-функция) не имеет особенностей и для его вычисления может использоваться любая квадратурная формула. При $\theta_1 < 1$ для вычисления несобственного интеграла можно использовать аддитивный метод выделения особенностей Канторовича. В этом случае функция $v(t) = t^{\theta_1} e^{-t}$ представляется в виде $v(t) = \phi(t) + \zeta(t)$, причем интеграл от первого слагаемого интегрируется непосредственно, а интеграл от второго уже не имеет особенностей и при использовании квадратурных формул не возникает проблем с большими погрешностями численного интегрирования. В данной ситуации

$$\begin{aligned} v(t) &= \phi(t) + \zeta(t) = \\ &= \left\{ t^{\theta_1-1} - \frac{t^{\theta_1}}{1!} + \frac{t^{\theta_1+1}}{2!} - \frac{t^{\theta_1+2}}{3!} + \dots + (-1)^n \frac{t^{\theta_1+n-1}}{n!} \dots \right\} + \\ &+ t^{\theta_1-1} \left\{ e^{-t} - \left[1 - \frac{t}{1!} + \frac{t^2}{2!} - \frac{t^3}{3!} + \dots + (-1)^n \frac{t^n}{n!} \dots \right] \right\}. \end{aligned}$$

Тогда

$$\int_0^{t_1} \phi(t) dt = \frac{t_1^{\theta_1}}{\theta_1} - \frac{t_1^{\theta_1+1}}{\theta_1+1} + \frac{t_1^{\theta_1+2}}{2!(\theta_1+2)} + \dots + (-1)^n \frac{t_1^{\theta_1+n}}{n!(\theta_1+n)} + \dots,$$

а

$$\int_0^{t_1} \zeta(t) dt = \int_0^{t_1} t^{\theta_1-1} \left\{ e^{-t} - \left[1 - \frac{t}{1!} + \frac{t^2}{2!} - \frac{t^3}{3!} + \dots + (-1)^n \frac{t^n}{n!} \dots \right] \right\} dt$$

без особенностей и вычисляется с использованием квадратурных формул. При этом в случае $t_1 > 1$ для повышения точности вычислений интеграл может разбиваться на 2: в пределах от 0 до 1 и от 1 до t_1 .

2. Аналогично поступаем при вычислении интеграла вида $\int_0^{t_1} t^{\theta_1} e^{-t} \ln t dt$.

При $\theta_1 < 1$ для вычисления несобственного интеграла с использованием метода выделения особенностей Канторовича представление подынтегральной функции имеет вид

$$\begin{aligned}
v(t) &= \phi(t) + \zeta(t) = \\
&= \ln t \left\{ t^{\theta_1-1} - \frac{t^{\theta_1}}{1!} + \frac{t^{\theta_1+1}}{2!} - \frac{t^{\theta_1+2}}{3!} + \dots + (-1)^n \frac{t^{\theta_1+n-1}}{n!} \dots \right\} + \\
&+ t^{\theta_1-1} \ln t \left\{ e^{-t} - \left[1 - \frac{t}{1!} + \frac{t^2}{2!} - \frac{t^3}{3!} + \dots + (-1)^n \frac{t^n}{n!} \dots \right] \right\}.
\end{aligned}$$

Тогда

$$\begin{aligned}
\int_0^{t_1} \phi(t) dt &= \frac{t_1^{\theta_1}}{\theta_1} \left(\ln t_1 - \frac{1}{\theta_1} \right) - \frac{t_1^{\theta_1+1}}{\theta_1+1} \left(\ln t_1 - \frac{1}{\theta_1+1} \right) + \frac{t_1^{\theta_1+2}}{2!(\theta_1+2)} \left(\ln t_1 - \frac{1}{\theta_1+2} \right) + \\
&+ \dots + (-1)^n \frac{t_1^{\theta_1+n}}{n!(\theta_1+n)} \left(\ln t_1 - \frac{1}{\theta_1+n} \right) + \dots,
\end{aligned}$$

а

$$\int_0^{t_1} \zeta(t) dt = \int_0^{t_1} t^{\theta_1-1} \ln t \left\{ e^{-t} - \left[1 - \frac{t}{1!} + \frac{t^2}{2!} - \frac{t^3}{3!} + \dots + (-1)^n \frac{t^n}{n!} \dots \right] \right\} dt$$

без особенностей и вычисляется с использованием квадратурных формул. При этом в случае $t_1 > 1$ для повышения точности вычислений интеграл может разбиваться на 2: в пределах от 0 до 1 и от 1 до t_1 .

3. При вычислении интеграла вида $\int_0^{t_1} t^{\theta_1} e^{-t} \ln^2 t dt$ для $\theta_1 < 1$ представление подынтегральной функции имеет вид

$$\begin{aligned}
v(t) &= \phi(t) + \zeta(t) = \\
&= \ln^2 t \left\{ t^{\theta_1-1} - \frac{t^{\theta_1}}{1!} + \frac{t^{\theta_1+1}}{2!} - \frac{t^{\theta_1+2}}{3!} + \dots + (-1)^n \frac{t^{\theta_1+n-1}}{n!} \dots \right\} + \\
&+ t^{\theta_1-1} \ln^2 t \left\{ e^{-t} - \left[1 - \frac{t}{1!} + \frac{t^2}{2!} - \frac{t^3}{3!} + \dots + (-1)^n \frac{t^n}{n!} \dots \right] \right\}.
\end{aligned}$$

Отсюда

$$\begin{aligned} \int_0^{t_1} \phi(t) dt &= \frac{t_1^{\theta_1}}{\theta_1} \left(\ln^2 t_1 - \frac{2 \ln t_1}{\theta_1} + \frac{2}{\theta_1^2} \right) - \frac{t_1^{\theta_1+1}}{\theta_1+1} \left(\ln^2 t_1 - \frac{2 \ln t_1}{\theta_1+1} + \frac{2}{(\theta_1+1)^2} \right) + \\ &+ \frac{t_1^{\theta_1+2}}{2!(\theta_1+2)} \left(\ln^2 t_1 - \frac{2 \ln t_1}{\theta_1+2} + \frac{2}{(\theta_1+2)^2} \right) + \\ &+ \dots + (-1)^n \frac{t_1^{\theta_1+n}}{n!(\theta_1+n)} \left(\ln^2 t_1 - \frac{2 \ln t_1}{\theta_1+n} + \frac{2}{(\theta_1+n)^2} \right) + \dots, \end{aligned}$$

а следующий интеграл без особенностей и вычисляется с использованием квадратурных формул

$$\int_0^{t_1} \zeta(t) dt = \int_0^{t_1} t^{\theta_1-1} \ln^2 t \left\{ e^{-t} - \left[1 - \frac{t}{1!} + \frac{t^2}{2!} - \frac{t^3}{3!} + \dots + (-1)^n \frac{t^n}{n!} \dots \right] \right\} dt.$$

При этом как и в предыдущем случае при $t_1 > 1$ для повышения точности вычислений интеграл может разбиваться на 2: в пределах от 0 до 1 и от 1 до t_1 .

5.3.19. Гамма-распределение

Гамма-распределение имеет функцию плотности

$$f(x) = \frac{\theta_1^{\theta_0}}{\Gamma(\theta_0)} (x - \theta_2)^{\theta_0-1} e^{-\theta_1(x-\theta_2)},$$

где основной параметр $\theta_0 \in (0, \infty)$, масштабный параметр $\theta_1 \in (0, \infty)$, параметр сдвига $\theta_2 \in (-\infty, \infty)$, а аргумент $x \in (\theta_2, \infty)$. Параметры θ_0 и θ_1 оцениваются методом максимального правдоподобия, а параметр θ_2 при необходимости оценивается с использованием порядковых статистик. Оценка максимального правдоподобия масштабного параметра θ_0 определяется как решение уравнения правдоподобия

$$\sum_{(1)} n_i \left\{ \frac{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} \ln t dt}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} dt} - \psi(\theta_0) \right\} + \sum_{(2)} \sum_{j=1}^{n_i} (\ln t_{ij} - \psi(\theta_0)) = 0,$$

где $t_i = \theta_1(x_i - \theta_2)$, $t_{ij} = \theta_1(x_{ij} - \theta_2)$, $\psi(\theta_0) = \Gamma'(\theta_0)/\Gamma(\theta_0)$ - логарифмическая производная гамма-функции Эйлера. Вторая

производная логарифма функции правдоподобия по параметру θ_0 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \sum_{(1)} n_i \left\{ \frac{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} \ln^2 t dt}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} dt} - \left(\frac{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} \ln t dt}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} dt} \right)^2 \right\} - \psi'(\theta_0) \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right).$$

Оценка параметра θ_1 находится как корень уравнения правдоподобия

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{t_i^{\theta_0} e^{-t_i} - t_{i-1}^{\theta_0} e^{-t_{i-1}}}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} dt} + \sum_{(2)} \sum_{j=1}^{n_i} (\theta_0 - t_{ij}) \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{t_i^{\theta_0} e^{-t_i} (\theta_0 - t_i - 1) - t_{i-1}^{\theta_0} e^{-t_{i-1}} (\theta_0 - t_{i-1} - 1)}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} dt} - \left(\frac{t_i^{\theta_0} e^{-t_i} - t_{i-1}^{\theta_0} e^{-t_{i-1}}}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} dt} \right)^2 \right] - \theta_0 \sum_{(2)} n_i \right\}.$$

При одновременном оценивании двух параметров θ_0 и θ_1 максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln \left(\int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} dt \right) + \sum_{(2)} \sum_{j=1}^{n_i} \left(\ln \theta_1 + (\theta_0 - 1) \ln t_{ij} - t_{ij} \right) - \ln \Gamma(\theta_0) \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right)$$

и оценки являются решением системы приведенных выше уравнений правдоподобия.

Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \sum_{(1)} n_i \times \frac{1}{\Gamma(\theta_0)} \sum_{(1)} \left\{ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} \ln t dt - \right. \\ \left. - \psi(\theta_0) \int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} dt \right\}^2 / \int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} dt + \psi'(\theta_0) \sum_{(2)} n_i,$$

$$J_{11} = J(\theta_1) = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \frac{1}{\Gamma(\theta_0)} \sum_{(1)} \frac{[t_i^{\theta_0} e^{-t_i} - t_{i-1}^{\theta_0} e^{-t_{i-1}}]^2}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} dt} + \theta_0 \sum_{(2)} n_i \right\},$$

$$J_{01} = J(\theta_0, \theta_1) = \sum_{(1)} n_i \times \frac{1}{\theta_1 \Gamma(\theta_0)} \sum_{(1)} \left\{ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} \ln t dt - \right. \\ \left. - \psi(\theta_0) \int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} dt \right\} (t_i^{\theta_0} e^{-t_i} - t_{i-1}^{\theta_0} e^{-t_{i-1}}) / \int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} dt - \frac{1}{\theta_1} \sum_{(2)} n_i,$$

и $J_{01} = J_{10}$.

Выборка объёмом 1000 наблюдений моделировалась с параметрами $\theta_0 = 2, \theta_1 = 1, \theta_2 = 0$. Результаты проверки согласия по различным критериям с гамма-распределением приведены на рис. 5.3.19. При проверке согласия по критериям отношения правдоподобия и χ^2 Пирсона решалась задача асимптотически оптимального группирования.

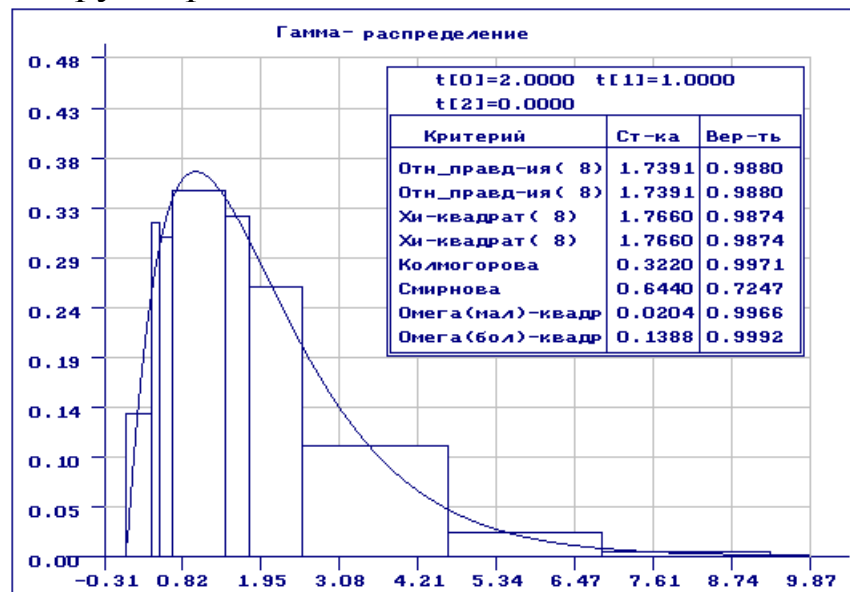


Рис. 5.3.19. Результаты проверки согласия смоделированной выборки с гамма-распределением

5.3.20. Бета-распределение I-го рода

Бета-распределение I-го рода имеет функцию плотности

$$f(x) = \frac{1}{\theta_3 B(\theta_0, \theta_1)} \left(\frac{x - \theta_2}{\theta_3} \right)^{\theta_0 - 1} \left(1 - \frac{x - \theta_2}{\theta_3} \right)^{\theta_1 - 1},$$

где $B(\theta_0, \theta_1) = \Gamma(\theta_0)\Gamma(\theta_1)/\Gamma(\theta_0 + \theta_1)$ - бета-функция, основные параметры $\theta_0, \theta_1 \in (0, \infty)$, параметр сдвига $\theta_2 \in (-\infty, \infty)$, масштабный параметр $\theta_3 \in (0, \infty)$, а аргумент $x \in [\theta_2, \theta_2 + \theta_3]$. Параметры θ_0 и θ_1 оцениваются методом максимального правдоподобия, а параметры сдвига θ_2 и масштаба θ_3 при необходимости оцениваются с использованием порядковых статистик. Оценка максимального правдоподобия параметра θ_0 определяется как решение уравнения правдоподобия

$$\sum_{(1)} n_i \frac{\int_{t_{i-1}}^{t_i} t^{\theta_0 - 1} (1-t)^{\theta_1 - 1} \ln t dt}{\int_{t_{i-1}}^{t_i} t^{\theta_0 - 1} (1-t)^{\theta_1 - 1} dt} + \sum_{(2)} \sum_{j=1}^{n_j} \ln t_{ij} - [\psi(\theta_0) - \psi(\theta_0 + \theta_1)] \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right) = 0,$$

где $t_i = (x_i - \theta_2)/\theta_3$, $t_{ij} = (x_{ij} - \theta_2)/\theta_3$, $\psi(\theta_0) = \Gamma'(\theta_0)/\Gamma(\theta_0)$ - логарифмическая производная гамма-функции Эйлера. Вторая производная логарифма функции правдоподобия по параметру θ_0 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \sum_{(1)} n_i \left\{ \frac{\int_{t_{i-1}}^{t_i} t^{\theta_0 - 1} (1-t)^{\theta_1 - 1} \ln^2 t dt}{\int_{t_{i-1}}^{t_i} t^{\theta_0 - 1} (1-t)^{\theta_1 - 1} dt} - \left(\frac{\int_{t_{i-1}}^{t_i} t^{\theta_0 - 1} (1-t)^{\theta_1 - 1} \ln t dt}{\int_{t_{i-1}}^{t_i} t^{\theta_0 - 1} (1-t)^{\theta_1 - 1} dt} \right)^2 \right\} -$$

$$- [\psi'(\theta_0) - \psi'(\theta_0 + \theta_1)] \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right).$$

Оценка параметра θ_1 находится как корень уравнения правдоподобия

$$\sum_{(1)} n_i \frac{\int_{t_{i-1}}^{t_i} t^{\theta_0 - 1} (1-t)^{\theta_1 - 1} \ln(1-t) dt}{\int_{t_{i-1}}^{t_i} t^{\theta_0 - 1} (1-t)^{\theta_1 - 1} dt} + \sum_{(2)} \sum_{j=1}^{n_j} \ln(1-t_{ij}) -$$

$$- [\psi(\theta_1) - \psi(\theta_0 + \theta_1)] \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right) = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \sum_{(1)} n_i \left\{ \frac{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln^2(1-t) dt}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt} - \left(\frac{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln(1-t) dt}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt} \right)^2 \right\} -$$

$$- [\psi^\circ(\theta_1) - \psi^\circ(\theta_0 + \theta_1)] \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right).$$

При одновременном оценивании двух параметров θ_0 и θ_1 максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln \left(\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt \right) + \sum_{(2)} \sum_{j=1}^{n_i} [(\theta_0 - 1) \ln t_{ij} + (\theta_1 - 1) \ln(1 - t_{ij})] -$$

$$- \ln B(\theta_0, \theta_1) \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right)$$

и оценки являются решением представленной системы уравнений правдоподобия. Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \sum_{(1)} n_i \times \frac{1}{B(\theta_0, \theta_1)} \sum_{(1)} \left\{ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt - \right.$$

$$\left. - [\psi(\theta_0) - \psi(\theta_0 + \theta_1)] \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt \right\}^2 / \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt +$$

$$+ [\psi^\circ(\theta_0) - \psi^\circ(\theta_0 + \theta_1)] \sum_{(2)} n_i,$$

$$J_{11} = J(\theta_1) = \sum_{(1)} n_i \times \frac{1}{B(\theta_0, \theta_1)} \sum_{(1)} \left\{ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln(1-t) dt - \right.$$

$$\left. - [\psi(\theta_1) - \psi(\theta_0 + \theta_1)] \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt \right\}^2 / \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt +$$

$$+ [\psi^\circ(\theta_1) - \psi^\circ(\theta_0 + \theta_1)] \sum_{(2)} n_i,$$

$$J_{01} = J(\theta_0, \theta_1) = \sum_{(1)} n_i \times \frac{1}{B(\theta_0, \theta_1)} \sum_{(1)} \left\{ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt - \right.$$

$$\left. - [\psi(\theta_0) - \psi(\theta_0 + \theta_1)] \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt \right\} \times \left\{ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln(1-t) dt - \right.$$

$$\left. - [\psi(\theta_1) - \psi(\theta_0 + \theta_1)] \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt \right\} / \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt -$$

$$-\psi^{\odot}(\theta_0 + \theta_1) \sum_{(2)} n_i,$$

и $J_{01} = J_{10}$.

Выборка объёмом 1000 наблюдений моделировалась с параметрами $\theta_0 = 1.5, \theta_1 = 2, \theta_2 = 0, \theta_3 = 2$. Результаты проверки согласия этой выборки с моделируемым бета-распределением приведены на рис. 5.20. При проверке согласия по критериям отношения правдоподобия и χ^2 Пирсона выборка разбивалась на интервалы равной вероятности (частоты). ОМП параметров, вычисленные по данной выборке равны $\hat{\theta}_0 = 1.4365, \hat{\theta}_1 = 1.8766$ при неопределяемых значениях параметров $\theta_2 = 0$ и $\theta_3 = 2$.

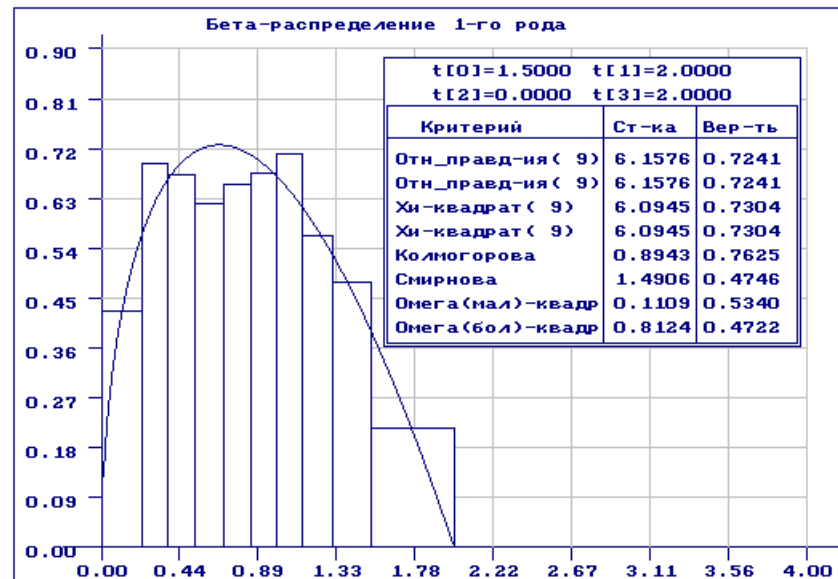


Рис. 5.20. Результаты проверки согласия смоделированной выборки с бета-распределением I-го рода

Вычисление интегралов.

При вычислении интегралов, содержащихся в вышеприведенных соотношениях, необходимо учитывать, что при $\theta_0 < 1$ и $\theta_1 < 1$ соответствующие интегралы имеют особенности, соответственно, на левой и правой границах области определения случайной величины. Если игнорировать этот факт, то погрешности численного интегрирования будут чрезвычайно велики.

1. При значениях $\theta_0 < 1$ интеграл вида $\int_0^{t_1} t^{\theta_0-1} (1-t)^{\theta_1-1} dt$ имеет особенность в нуле. Используя аддитивный метод выделения особенностей, представляем подынтегральную функцию $v(t) = t^{\theta_0-1} (1-t)^{\theta_1-1}$ в виде

$$\begin{aligned} v(t) &= \phi(t) + \zeta(t) = \\ &= \left\{ t^{\theta_0-1} - \frac{(\theta_1-1)t^{\theta_0}}{1!} + \frac{(\theta_1-1)(\theta_1-2)t^{\theta_0+1}}{2!} - \dots \right. \\ &\quad \left. \dots + (-1)^n \frac{(\theta_1-1)\dots(\theta_1-n)t^{\theta_0+n-1}}{n!} \dots \right\} + \\ &\quad + t^{\theta_0-1} \left\{ (1-t)^{\theta_1-1} - \left[1 - \frac{(\theta_1-1)t}{1!} + \frac{(\theta_1-1)(\theta_1-2)t^2}{2!} - \dots \right. \right. \\ &\quad \left. \left. \dots + (-1)^n \frac{(\theta_1-1)\dots(\theta_1-n)t^n}{n!} \dots \right] \right\}. \end{aligned}$$

Тогда

$$\begin{aligned} \int_0^{t_1} \phi(t) dt &= \frac{t_1^{\theta_0}}{\theta_0} - \frac{(\theta_1-1)t_1^{\theta_0+1}}{\theta_0+1} + \frac{(\theta_1-1)(\theta_1-2)t_1^{\theta_0+2}}{2!(\theta_0+2)} - \dots \\ &\quad \dots + (-1)^n \frac{(\theta_1-1)\dots(\theta_1-n)t_1^{\theta_0+n}}{n!(\theta_0+n)} + \dots, \end{aligned}$$

и следующий интеграл

$$\begin{aligned} \int_0^{t_1} \zeta(t) dt &= \int_0^{t_1} t^{\theta_0-1} \left\{ (1-t)^{\theta_1-1} - \left[1 - \frac{(\theta_1-1)t}{1!} + \frac{(\theta_1-1)(\theta_1-2)t^2}{2!} - \dots \right. \right. \\ &\quad \left. \left. \dots + (-1)^n \frac{(\theta_1-1)\dots(\theta_1-n)t^n}{n!} \dots \right] \right\} dt \end{aligned}$$

без особенностей и вычисляется с использованием квадратурных формул.

Вычисление интеграла вида $\int_{t_{k-1}}^1 t^{\theta_0-1} (1-t)^{\theta_1-1} dt$ при $\theta_1 < 1$ сводится к предыдущему случаю очевидной заменой переменной $s = 1 - t$.

2. Аналогично, в случае интеграла вида $\int_0^{t_1} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt$ приходим к преобразованию

$$\begin{aligned} v(t) &= \phi(t) + \zeta(t) = \\ &= \ln t \left\{ t^{\theta_0-1} - \frac{(\theta_1-1)t^{\theta_0}}{1!} + \frac{(\theta_1-1)(\theta_1-2)t^{\theta_0+1}}{2!} - \dots \right. \\ &\quad \left. \dots + (-1)^n \frac{(\theta_1-1)\dots(\theta_1-n)t^{\theta_0+n-1}}{n!} \dots \right\} + \\ &\quad + t^{\theta_0-1} \ln t \left\{ (1-t)^{\theta_1-1} - \left[1 - \frac{(\theta_1-1)t}{1!} + \frac{(\theta_1-1)(\theta_1-2)t^2}{2!} - \dots \right. \right. \\ &\quad \left. \left. \dots + (-1)^n \frac{(\theta_1-1)\dots(\theta_1-n)t^n}{n!} \dots \right] \right\}. \end{aligned}$$

Тогда

$$\begin{aligned} \int_0^{t_1} \phi(t) dt &= \frac{t_1^{\theta_0}}{\theta_0} \left(\ln t_1 - \frac{1}{\theta_0} \right) - \frac{(\theta_1-1)t_1^{\theta_0+1}}{\theta_0+1} \left(\ln t_1 - \frac{1}{\theta_0+1} \right) + \dots \\ &\quad \dots + (-1)^n \frac{(\theta_1-1)\dots(\theta_1-n)t_1^{\theta_0+n}}{n!(\theta_0+n)} \left(\ln t_1 - \frac{1}{\theta_0+n} \right) + \dots, \end{aligned}$$

а интеграл

$$\begin{aligned} \int_0^{t_1} \zeta(t) dt &= \int_0^{t_1} t^{\theta_0-1} \ln t \left\{ (1-t)^{\theta_1-1} - \left[1 - \frac{(\theta_1-1)t}{1!} + \frac{(\theta_1-1)(\theta_1-2)t^2}{2!} - \dots \right. \right. \\ &\quad \left. \left. \dots + (-1)^n \frac{(\theta_1-1)\dots(\theta_1-n)t^n}{n!} \dots \right] \right\} dt \end{aligned}$$

- без особенностей. Вычисление интеграла вида $\int_{t_{k-1}}^1 t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt$ при $\theta_1 < 1$ сводится к предыдущему случаю.

3. В случае интеграла вида $\int_0^{t_1} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt$ имеем

$$\begin{aligned}
v(t) &= \phi(t) + \zeta(t) = \\
&= \ln^2 t \left\{ t^{\theta_0-1} - \frac{(\theta_1-1)t^{\theta_0}}{1!} + \frac{(\theta_1-1)(\theta_1-2)t^{\theta_0+1}}{2!} - \dots \right. \\
&\quad \left. \dots + (-1)^n \frac{(\theta_1-1)\dots(\theta_1-n)t^{\theta_0+n-1}}{n!} \dots \right\} + \\
&\quad + t^{\theta_0-1} \ln^2 t \left\{ (1-t)^{\theta_1-1} - \left[1 - \frac{(\theta_1-1)t}{1!} + \frac{(\theta_1-1)(\theta_1-2)t^2}{2!} - \dots \right. \right. \\
&\quad \left. \left. \dots + (-1)^n \frac{(\theta_1-1)\dots(\theta_1-n)t^n}{n!} \dots \right] \right\}.
\end{aligned}$$

В результате

$$\begin{aligned}
\int_0^{t_1} \phi(t) dt &= \frac{t_1^{\theta_0}}{\theta_0} \left[\ln^2 t_1 - \frac{2}{\theta_0} \left(\ln t_1 - \frac{1}{\theta_0} \right) \right] - \\
&\quad - \frac{(\theta_1-1)t_1^{\theta_0+1}}{\theta_0+1} \left[\ln^2 t_1 - \frac{2}{\theta_0+1} \left(\ln t_1 - \frac{1}{\theta_0+1} \right) \right] + \dots \\
&\quad \dots + (-1)^n \frac{(\theta_1-1)\dots(\theta_1-n)t_1^{\theta_0+n}}{n!(\theta_0+n)} \left[\ln^2 t_1 - \frac{2}{\theta_0+n} \left(\ln t_1 - \frac{1}{\theta_0+n} \right) \right] + \dots,
\end{aligned}$$

а интеграл

$$\begin{aligned}
\int_0^{t_1} \zeta(t) dt &= \int_0^{t_1} t^{\theta_0-1} \ln^2 t \left\{ (1-t)^{\theta_1-1} - \left[1 - \frac{(\theta_1-1)t}{1!} + \frac{(\theta_1-1)(\theta_1-2)t^2}{2!} - \dots \right. \right. \\
&\quad \left. \left. \dots + (-1)^n \frac{(\theta_1-1)\dots(\theta_1-n)t^n}{n!} \dots \right] \right\} dt
\end{aligned}$$

- без особенностей. Вычисление интеграла вида $\int_{t_{k-1}}^1 t^{\theta_0-1} (1-t)^{\theta_1-1} \ln^2 t dt$ при $\theta_1 < 1$ сводится к предыдущему случаю.

5.3.21. Стандартное бета-распределение II-го рода

Стандартное бета-распределение II-го рода имеет функцию плотности вида

$$f(x) = \frac{1}{B(\theta_0, \theta_1)} \frac{(x - \theta_2)^{\theta_0 - 1}}{(1 + x - \theta_2)^{\theta_0 + \theta_1}},$$

где $\theta_0, \theta_1 \in (0, \infty)$, параметр сдвига $\theta_2 \in (-\infty, \infty)$, аргумент $x \in [\theta_2, \infty)$.

Параметры θ_0 и θ_1 оцениваются по методу максимального правдоподобия, а параметр сдвига θ_2 с использованием порядковых статистик. В данном случае для параметров θ_0 и θ_1 уравнения правдоподобия, выражения для вторых производных и информационных количеств имеют тот же вид, что и для бета-распределения I-го рода, но при подстановке в них величин t_i и t_{ij} в виде $t_i = \frac{x_i - \theta_2}{1 + x_i - \theta_2}$ и

$$t_{ij} = \frac{x_{ij} - \theta_2}{1 + x_{ij} - \theta_2}.$$

Результаты оценивания параметров бета-распределения II-го рода и последующего анализа приведены на рис. 5.3.21. Выборка объёмом 1000 наблюдений моделировалась с параметрами $\theta_0 = 2, \theta_1 = 1.5, \theta_2 = 0$. При проверке согласия по критериям отношения правдоподобия и χ^2 Пирсона она разбивалась на интервалы равной вероятности (частоты).

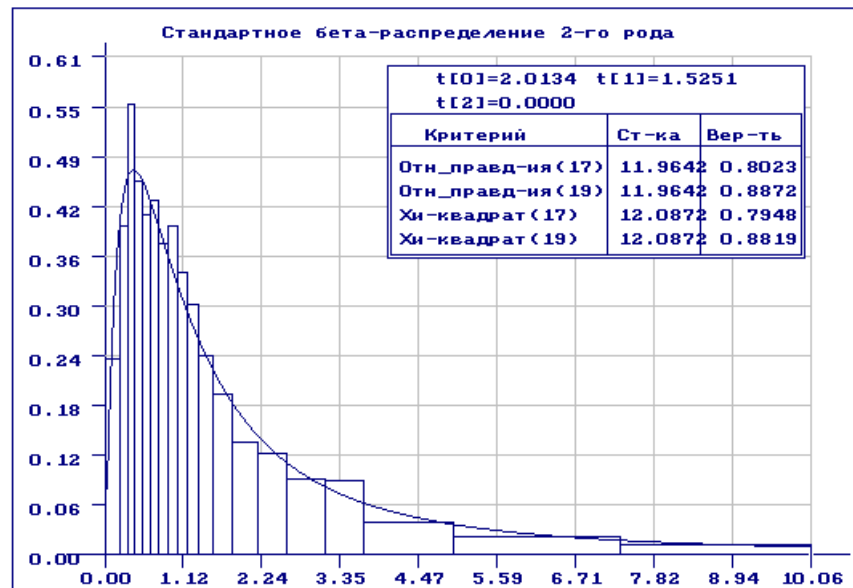


Рис. 5.3.21. Результаты статистического анализа выборки, смоделированной в соответствии со стандартным бета-распределением II-го рода

5.3.22. Бета-распределение II-го рода

Функция плотности бета-распределения II-го рода описывается выражением

$$f(x) = \frac{\theta_2}{B(\theta_0, \theta_1)} \frac{[\theta_2(x - \theta_3)]^{\theta_0-1}}{[1 + \theta_2(x - \theta_3)]^{\theta_0+\theta_1}},$$

где $B(\theta_0, \theta_1) = \Gamma(\theta_0)\Gamma(\theta_1)/\Gamma(\theta_0 + \theta_1)$ - бета-функция, основные параметры $\theta_0, \theta_1 \in (0, \infty)$, масштабный параметр $\theta_2 \in (0, \infty)$, параметр сдвига $\theta_3 \in (-\infty, \infty)$, а аргумент $x \in [\theta_3, \infty)$. Первые три параметра оцениваются методом максимального правдоподобия, а параметр сдвига θ_3 при необходимости оценивается с использованием порядковых статистик. Оценка максимального правдоподобия параметра θ_0 определяется как решение уравнения правдоподобия

$$\sum_{(1)} n_i \frac{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt} + \sum_{(2)} \sum_{j=1}^{n_i} \ln t_{ij} - [\psi(\theta_0) - \psi(\theta_0 + \theta_1)] \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right) = 0,$$

где $t_i = \frac{\theta_2(x_i - \theta_3)}{1 + \theta_2(x_i - \theta_3)}$, $t_{ij} = \frac{\theta_2(x_{ij} - \theta_3)}{1 + \theta_2(x_{ij} - \theta_3)}$. Вторая производная логарифма функции правдоподобия по параметру θ_0 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \sum_{(1)} n_i \left\{ \frac{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln^2 t dt}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt} - \left(\frac{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt} \right)^2 \right\} -$$

$$- [\psi^\circ(\theta_0) - \psi^\circ(\theta_0 + \theta_1)] \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right).$$

Оценка параметра θ_1 находится как корень уравнения правдоподобия

$$\sum_{(1)} n_i \frac{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln(1-t) dt}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt} + \sum_{(2)} \sum_{j=1}^{n_i} \ln(1-t_{ij}) -$$

$$- [\psi(\theta_1) - \psi(\theta_0 + \theta_1)] \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right) = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \sum_{(1)} n_i \left\{ \frac{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln^2(1-t) dt}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt} - \left(\frac{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln(1-t) dt}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt} \right)^2 \right\} -$$

$$- [\Psi^\circ(\theta_1) - \Psi^\circ(\theta_0 + \theta_1)] \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right).$$

Оценка параметра θ_2 находится как корень уравнения правдоподобия

$$\frac{\theta_0}{\theta_2} \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right) - \frac{\theta_0 + \theta_1}{\theta_2} \left\{ \sum_{(1)} n_i \frac{\int_{t_{i-1}}^{t_i} t^{\theta_0} (1-t)^{\theta_1-1} dt}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt} + \sum_{(2)} \sum_{j=1}^{n_i} t_{ij} \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_2 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_2^2} = \frac{\theta_0 + \theta_1}{\theta_2^2} \left\{ \sum_{(1)} n_i \left[\frac{(\theta_0 + \theta_1 + 1) \int_{t_{i-1}}^{t_i} t^{\theta_0+1} (1-t)^{\theta_1-1} dt}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt} - \right. \right.$$

$$\left. - (\theta_0 + \theta_1) \left(\frac{\int_{t_{i-1}}^{t_i} t^{\theta_0} (1-t)^{\theta_1-1} dt}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt} \right)^2 \right] - \sum_{(2)} \sum_{j=1}^{n_i} t_{ij}^2 \right\} - \frac{\theta_0}{\theta_2^2} \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right).$$

При одновременном оценивании трёх параметров θ_0 , θ_1 и θ_2 максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln \left(\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt \right) + \sum_{(2)} \sum_{j=1}^{n_i} [(\theta_0 - 1) \ln t_{ij} + (\theta_1 - 1) \ln(1 - t_{ij})] -$$

$$- \ln B(\theta_0, \theta_1) \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right) + \ln \theta_2 \sum_{(2)} n_i$$

и оценки являются решением системы уравнений правдоподобия.

Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \sum_{(1)} n_i \times \frac{1}{B(\theta_0, \theta_1)} \sum_{(1)} \left\{ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt - \right.$$

$$\begin{aligned}
& -\left[\psi(\theta_0) - \psi(\theta_0 + \theta_1)\right] \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt \Bigg\}^2 \Bigg/ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt + \\
& + \left[\psi^{\odot}(\theta_0) - \psi^{\odot}(\theta_0 + \theta_1)\right] \sum_{(2)} n_i, \\
J_{11} = J(\theta_1) = \sum_{(1)} n_i \times \frac{1}{B(\theta_0, \theta_1)} \sum_{(1)} \Bigg\{ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln(1-t) dt - \\
& - \left[\psi(\theta_1) - \psi(\theta_0 + \theta_1)\right] \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt \Bigg\}^2 \Bigg/ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt + \\
& + \left[\psi^{\odot}(\theta_1) - \psi^{\odot}(\theta_0 + \theta_1)\right] \sum_{(2)} n_i, \\
J_{01} = J(\theta_0, \theta_1) = \sum_{(1)} n_i \times \frac{1}{B(\theta_0, \theta_1)} \sum_{(1)} \Bigg\{ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt - \\
& - \left[\psi(\theta_0) - \psi(\theta_0 + \theta_1)\right] \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt \Bigg\} \times \Bigg\{ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln(1-t) dt - \\
& - \left[\psi(\theta_1) - \psi(\theta_0 + \theta_1)\right] \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt \Bigg\} \Bigg/ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt - \\
& - \psi^{\odot}(\theta_0 + \theta_1) \sum_{(2)} n_i, \\
J_{22} = J(\theta_2) = \sum_{(1)} n_i \times \frac{1}{\theta_2^2 B(\theta_0, \theta_1)} \sum_{(1)} \Bigg\{ \theta_0 \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt - \\
& - (\theta_0 + \theta_1) \int_{t_{i-1}}^{t_i} t^{\theta_0} (1-t)^{\theta_1-1} dt \Bigg\}^2 \Bigg/ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt + \frac{\theta_0 \theta_1}{\theta_2^2 (\theta_0 + \theta_1 + 1)} \sum_{(2)} n_i, \\
J_{02} = J(\theta_0, \theta_2) = \sum_{(1)} n_i \times \frac{1}{\theta_2 B(\theta_0, \theta_1)} \sum_{(1)} \Bigg\{ \frac{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt} - \\
& - (\psi(\theta_0) - \psi(\theta_0 + \theta_1)) \Bigg\} \Bigg\{ \theta_0 \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt - (\theta_0 + \theta_1) \int_{t_{i-1}}^{t_i} t^{\theta_0} (1-t)^{\theta_1-1} dt \Bigg\} - \\
& - \frac{\theta_1}{\theta_2 (\theta_0 + \theta_1)} \sum_{(2)} n_i, \\
J_{12} = J(\theta_1, \theta_2) = \sum_{(1)} n_i \times \frac{1}{\theta_2 B(\theta_0, \theta_1)} \sum_{(1)} \Bigg\{ \frac{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln(1-t) dt}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt} - \\
& - (\psi(\theta_1) - \psi(\theta_0 + \theta_1)) \Bigg\} \Bigg\{ \theta_0 \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt - (\theta_0 + \theta_1) \int_{t_{i-1}}^{t_i} t^{\theta_0} (1-t)^{\theta_1-1} dt \Bigg\} -
\end{aligned}$$

$$+ \frac{\theta_0}{\theta_2(\theta_0 + \theta_1)^{(2)}} \sum n_i,$$

и $J_{01} = J_{10}, J_{02} = J_{20}, J_{12} = J_{21}$.

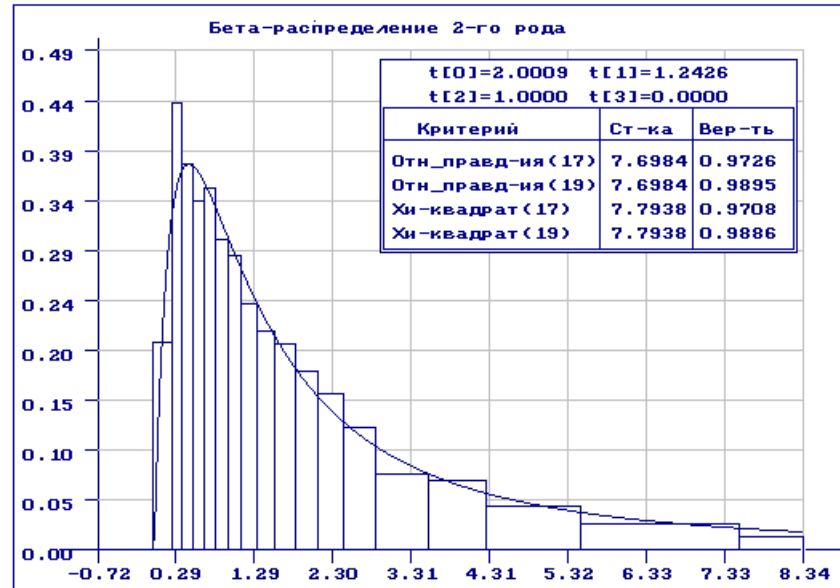


Рис. 5.3.22. Результаты статистического анализа выборки, смоделированной в соответствии с бета-распределением II-го рода

Результаты оценивания параметров бета-распределения II-го рода и последующего анализа приведены на рис. 5.3.22. Выборка объёмом 1000 наблюдений моделировалась с параметрами $\theta_0 = 2, \theta_1 = 1.2, \theta_2 = 1, \theta_3 = 0$. Параметры распределения оценивались по сгруппированной выборке. При группировании она разбивалась на интервалы равной вероятности (частоты).

5.3.23. Распределение Sb-Джонсона

Распределение Sb-Джонсона имеет функцию плотности вида

$$f(x) = \frac{\theta_1 \theta_2}{(x - \theta_3)(\theta_2 + \theta_3 - x)} \exp \left\{ -\frac{1}{2} \left[\theta_0 - \theta_1 \ln \frac{x - \theta_3}{\theta_2 + \theta_3 - x} \right]^2 \right\},$$

где параметр $\theta_0 \in (-\infty, \infty)$, параметры $\theta_1, \theta_2 \in (0, \infty)$, θ_2 задаёт размах области определения случайной величины, θ_3 является параметром сдвига, $\theta_3 \in (-\infty, \infty)$, $x \in [\theta_3, \theta_2 + \theta_3]$. Параметры θ_0 и θ_1 оцениваются по методу максимального правдоподобия, а θ_2 и θ_3 при необходимости с использованием порядковых статистик.

Оценка максимального правдоподобия параметра θ_0 определяется как решение уравнения правдоподобия

$$\sum_{(1)} n_i \frac{\varphi(t_i) - \varphi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} - \sum_{(2)} \sum_{j=1}^{n_i} t_{ij} = 0,$$

где

$$t_i = \theta_0 + \theta_1 \ln \frac{x_i - \theta_3}{\theta_2 + \theta_3 - x_i}, \quad t_{ij} = \theta_0 + \theta_1 \ln \frac{x_{ij} - \theta_3}{\theta_2 + \theta_3 - x_{ij}}, \quad \varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2},$$

$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-t^2/2} dt$. Вторая производная логарифма функции правдоподобия по параметру θ_0 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \sum_{(1)} n_i \left[\frac{t_{i-1} \varphi(t_{i-1}) - t_i \varphi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} - \left(\frac{\varphi(t_i) - \varphi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} \right)^2 \right] - \sum_{(2)} n_i.$$

Оценка параметра θ_1 находится в качестве решения уравнения правдоподобия

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{(t_i - \theta_0) \varphi(t_i) - (t_{i-1} - \theta_0) \varphi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} n_i - \sum_{(2)} \sum_{j=1}^{n_i} t_{ij} (t_{ij} - \theta_0) \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_1^2} = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{t_{i-1} (t_{i-1} - \theta_0)^2 \varphi(t_{i-1}) - t_i (t_i - \theta_0)^2 \varphi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} - \right. \right. \\ \left. \left. - \left(\frac{(t_i - \theta_0) \varphi(t_i) - (t_{i-1} - \theta_0) \varphi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} \right)^2 \right] - \sum_{(2)} \sum_{j=1}^{n_i} \left[1 + (t_{ij} - \theta_0)^2 \right] \right\}. \end{aligned}$$

При одновременном оценивании параметров θ_0 и θ_1 максимизируется логарифм функции правдоподобия

$$\begin{aligned} \ln L = \sum_{(1)} n_i \ln [\Phi(t_i) - \Phi(t_{i-1})] + \ln \frac{\theta_1 \theta_2}{\sqrt{2\pi}} \sum_{(2)} n_i - \\ - \sum_{(2)} \sum_{j=1}^{n_i} \left\{ \frac{t_{ij}^2}{2} + \ln [(x_{ij} - \theta_3)(\theta_2 + \theta_3 - x_{ij})] \right\} \end{aligned}$$

и оценки являются решением системы уравнений правдоподобия. Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \sum_{(1)} n_i \times \sum_{(1)} \frac{[\varphi(t_i) - \varphi(t_{i-1})]^2}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} n_i,$$

$$J_{11} = J(\theta_1) = \sum_{(1)} n_i \times \frac{1}{\theta_1^2} \sum_{(1)} \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]^2}{\Phi(t_i) - \Phi(t_{i-1})} + \frac{2 + \theta_0^2}{\theta_1^2} \sum_{(2)} n_i,$$

$$J_{01} = J(\theta_0, \theta_1) = \sum_{(1)} n_i \times \frac{1}{\theta_1} \sum_{(1)} \left\{ \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]}{\Phi(t_i) - \Phi(t_{i-1})} \times \right. \\ \left. \times [\varphi(t_i) - \varphi(t_{i-1})] \right\} - \frac{\theta_0}{\theta_1} \sum_{(2)} n_i,$$

и $J_{01} = J_{10}$.

На рис. 5.3.23 приведены результаты проверки согласия выборки, смоделированной в соответствии с распределением *Sb*-Джонсона с параметрами $\theta_0 = 0, \theta_1 = 1, \theta_2 = 2, \theta_3 = 1$. Объем выборки 1000 наблюдений. На рис. 5.3.24 приведены результаты анализа после оценивания параметров θ_0 и θ_1 .

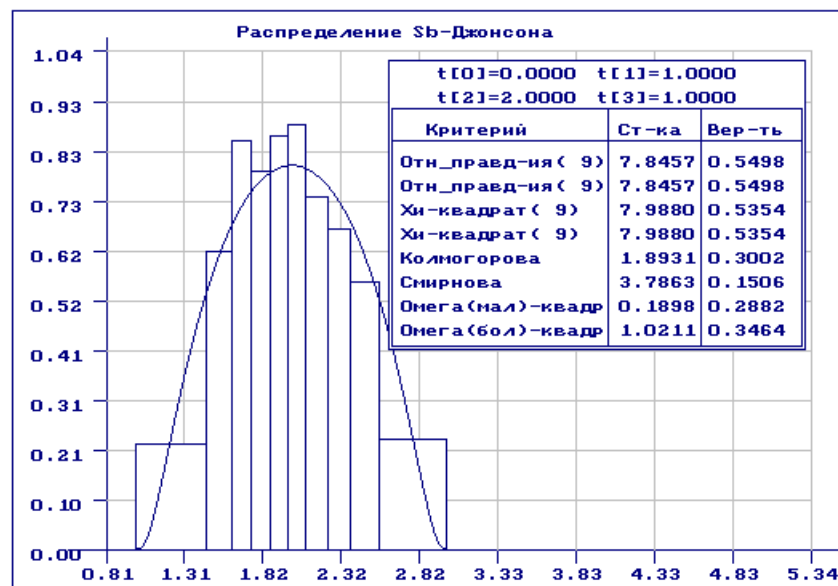


Рис. 5.3.23. Результаты проверки согласия с распределением *Sb*-Джонсона

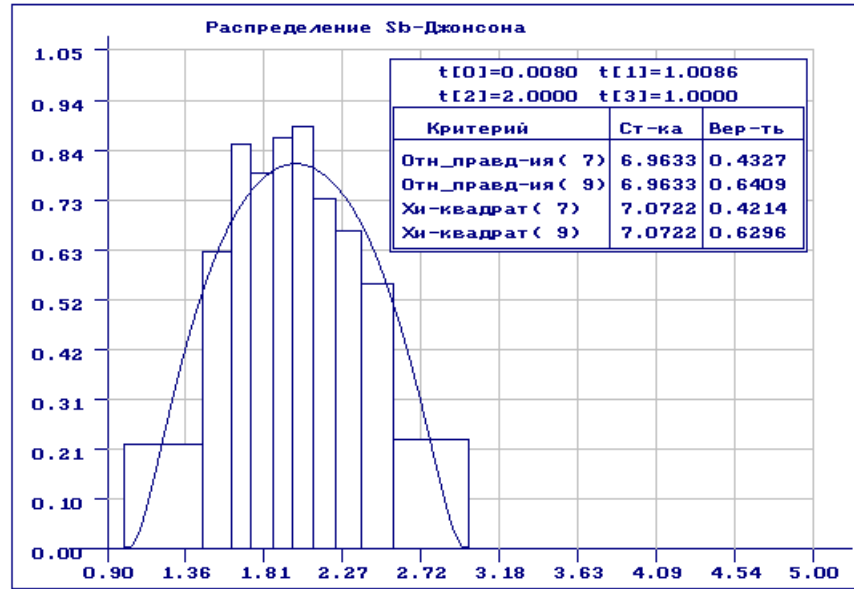


Рис. 5.3.24. Результаты проверки согласия с распределением Sb-Джонсона после оценивания параметров распределения

5.3.24. Распределение Sl-Джонсона

Распределение Sl-Джонсона имеет функцию плотности вида

$$f(x) = \frac{\theta_1}{(x - \theta_3)} \exp \left\{ -\frac{1}{2} \left[\theta_0 + \theta_1 \ln \frac{x - \theta_3}{\theta_2} \right]^2 \right\},$$

где параметр $\theta_0 \in (-\infty, \infty)$, параметры $\theta_1, \theta_2 \in (0, \infty)$, параметр сдвига $\theta_3 \in (-\infty, \infty)$, $x \in [\theta_3, \infty)$. Параметры θ_0 , θ_1 и θ_2 оцениваются по методу максимального правдоподобия, а θ_3 при необходимости с использованием порядковых статистик.

Оценка максимального правдоподобия параметра θ_0 определяется как решение уравнения правдоподобия

$$\sum_{(1)} n_i \frac{\varphi(t_i) - \varphi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} - \sum_{(2)} \sum_{j=1}^{n_j} t_{ij} = 0,$$

$$\text{где } t_i = \theta_0 + \theta_1 \ln \frac{x_i - \theta_3}{\theta_2}, \quad t_{ij} = \theta_0 + \theta_1 \ln \frac{x_{ij} - \theta_3}{\theta_2}, \quad \varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2},$$

$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-t^2/2} dt$. Вторая производная логарифма функции правдоподобия по параметру θ_0 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \sum_{(1)} n_i \left[\frac{t_{i-1} \varphi(t_{i-1}) - t_i \varphi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} - \left(\frac{\varphi(t_i) - \varphi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} \right)^2 \right] - \sum_{(2)} n_i.$$

Оценка параметра θ_1 находится в качестве решения уравнения правдоподобия

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{(t_i - \theta_0) \varphi(t_i) - (t_{i-1} - \theta_0) \varphi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} n_i - \sum_{(2)} \sum_{j=1}^{n_i} t_{ij} (t_{ij} - \theta_0) \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{t_{i-1} (t_{i-1} - \theta_0)^2 \varphi(t_{i-1}) - t_i (t_i - \theta_0)^2 \varphi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} - \right. \right. \\ \left. \left. - \left(\frac{(t_i - \theta_0) \varphi(t_i) - (t_{i-1} - \theta_0) \varphi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} \right)^2 \right] - \sum_{(2)} \sum_{j=1}^{n_i} [1 + (t_{ij} - \theta_0)^2] \right\}.$$

Оценка параметра θ_2 находится в качестве решения уравнения правдоподобия

$$\frac{\theta_1}{\theta_2} \left\{ \sum_{(1)} n_i \frac{\varphi(t_{i-1}) - \varphi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} \sum_{j=1}^{n_i} t_{ij} \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_2 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_2^2} = \frac{\theta_1^2}{\theta_2^2} \left\{ \sum_{(1)} n_i \left[\frac{(t_{i-1} - 1/\theta_1) \varphi(t_{i-1}) - (t_i - 1/\theta_1) \varphi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} - \right. \right. \\ \left. \left. - \left(\frac{\varphi(t_{i-1}) - \varphi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} \right)^2 \right] - \sum_{(2)} n_i - \frac{1}{\theta_1} \sum_{(2)} \sum_{j=1}^{n_i} t_{ij} \right\}.$$

При одновременном оценивании параметров θ_0 , θ_1 и θ_2 максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln [\Phi(t_i) - \Phi(t_{i-1})] + \ln \frac{\theta_1}{\sqrt{2\pi}} \sum_{(2)} n_i - \sum_{(2)} \sum_{j=1}^{n_i} \left\{ \frac{t_{ij}^2}{2} + \ln(x_{ij} - \theta_3) \right\}$$

и оценки являются решением системы уравнений правдоподобия. Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \sum_{(1)} n_i \times \sum_{(1)} \frac{[\varphi(t_i) - \varphi(t_{i-1})]^2}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} n_i,$$

$$J_{11} = J(\theta_1) = \sum_{(1)} n_i \times \frac{1}{\theta_1^2} \sum_{(1)} \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]^2}{\Phi(t_i) - \Phi(t_{i-1})} + \frac{2 + \theta_0^2}{\theta_1^2} \sum_{(2)} n_i,$$

$$J_{01} = J(\theta_0, \theta_1) = \sum_{(1)} n_i \times \frac{1}{\theta_1} \sum_{(1)} \left\{ \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]}{\Phi(t_i) - \Phi(t_{i-1})} \times \right. \\ \left. \times [\varphi(t_i) - \varphi(t_{i-1})] \right\} - \frac{\theta_0}{\theta_1} \sum_{(2)} n_i,$$

$$J_{22} = J(\theta_2) = \frac{\theta_1^2}{\theta_2^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[\varphi(t_{i-1}) - \varphi(t_i)]^2}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} n_i \right\},$$

$$J_{02} = J(\theta_0, \theta_2) = \frac{\theta_1}{\theta_2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[\varphi(t_i) - \varphi(t_{i-1})][\varphi(t_{i-1}) - \varphi(t_i)]}{\Phi(t_i) - \Phi(t_{i-1})} - \sum_{(2)} n_i \right\},$$

$$J_{12} = J(\theta_1, \theta_2) = \sum_{(1)} n_i \times \frac{1}{\theta_2} \sum_{(1)} \left\{ \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]}{\Phi(t_i) - \Phi(t_{i-1})} \times \right. \\ \left. \times [\varphi(t_{i-1}) - \varphi(t_i)] \right\} + \frac{\theta_0}{\theta_2} \sum_{(2)} n_i,$$

и $J_{01} = J_{10}, J_{02} = J_{20}, J_{12} = J_{21}$.

Результаты оценивания параметров распределения *Sl*-Джонсона и последующей проверки гипотез по критериям согласия приведены на рис. 5.3.25. Выборка объемом 1000 наблюдений моделировалась с параметрами $\theta_0 = -1$, $\theta_1 = 1$, $\theta_2 = 2$, $\theta_3 = 1$.

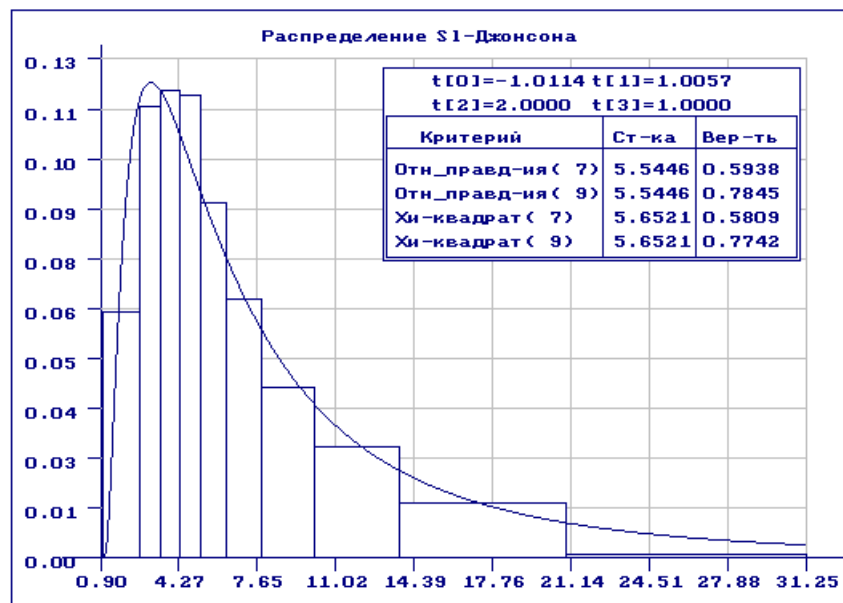


Рис. 5.3.25. Гистограмма и плотность распределения *Sl*-Джонсона

5.3.25. Распределение *Su*-Джонсона

Распределение *Su*-Джонсона имеет функцию плотности вида

$$f(x) = \frac{\theta_1}{\sqrt{2\pi}\sqrt{(x-\theta_3)^2 + \theta_2^2}} \exp \left\{ -\frac{1}{2} \left[\theta_0 + \theta_1 \ln \left\{ \frac{x-\theta_3}{\theta_2} + \sqrt{\left(\frac{x-\theta_3}{\theta_2} \right)^2 + 1} \right\} \right]^2 \right\},$$

где параметр $\theta_0 \in (-\infty, \infty)$, параметры $\theta_1, \theta_2 \in (0, \infty)$, параметр сдвига $\theta_3 \in (-\infty, \infty)$, аргумент $x \in (-\infty, \infty)$. Параметры распределения оцениваются по методу максимального правдоподобия.

Оценка максимального правдоподобия параметра θ_0 определяется как решение уравнения правдоподобия

$$\sum_{(1)} n_i \frac{\varphi(t_i) - \varphi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} - \sum_{(2)} \sum_{j=1}^{n_i} t_{ij} = 0,$$

где

$$t_i = \theta_0 + \theta_1 \ln \left\{ \frac{x_i - \theta_3}{\theta_2} + \sqrt{\left(\frac{x_i - \theta_3}{\theta_2} \right)^2 + 1} \right\}, \quad \varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2},$$

$$t_{ij} = \theta_0 + \theta_1 \ln \left\{ \frac{x_{ij} - \theta_3}{\theta_2} + \sqrt{\left(\frac{x_{ij} - \theta_3}{\theta_2} \right)^2 + 1} \right\}, \quad \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-t^2/2} dt.$$

Вторая производная логарифма функции правдоподобия по параметру θ_0 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \sum_{(1)} n_i \left[\frac{t_{i-1} \varphi(t_{i-1}) - t_i \varphi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} - \left(\frac{\varphi(t_i) - \varphi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} \right)^2 \right] - \sum_{(2)} n_i.$$

Оценка параметра θ_1 находится в качестве решения уравнения правдоподобия

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{(t_i - \theta_0) \varphi(t_i) - (t_{i-1} - \theta_0) \varphi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} n_i - \sum_{(2)} \sum_{j=1}^{n_i} t_{ij} (t_{ij} - \theta_0) \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{t_{i-1}(t_{i-1} - \theta_0)^2 \varphi(t_{i-1}) - t_i(t_i - \theta_0)^2 \varphi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} - \right. \right. \\ \left. \left. - \left(\frac{(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} \right)^2 \right] - \sum_{(2)} \sum_{j=1}^{n_i} \left[1 + (t_{ij} - \theta_0)^2 \right] \right\}.$$

Оценка параметра θ_2 находится в качестве решения уравнения правдоподобия

$$\frac{\theta_1}{\theta_2} \left\{ \sum_{(1)} n_i \frac{\frac{s_{i-1}\varphi(t_{i-1})}{\sqrt{s_{i-1}^2 + 1}} - \frac{s_i\varphi(t_i)}{\sqrt{s_i^2 + 1}}}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} \sum_{j=1}^{n_i} \left[\frac{t_{ij}s_{ij}}{\sqrt{s_{ij}^2 + 1}} - \frac{1}{\theta_1(s_{ij}^2 + 1)} \right] \right\} = 0,$$

где $s_i = (x_i - \theta_3) / \theta_2$, $s_{ij} = (x_{ij} - \theta_3) / \theta_2$. Вторая производная логарифма функции правдоподобия по параметру θ_2 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_2^2} = \frac{\theta_1^2}{\theta_2^2} \left\{ \sum_{(1)} n_i \left[\frac{\varphi(t_{i-1}) \left(\frac{t_{i-1}s_{i-1}^2}{s_{i-1}^2 + 1} - \frac{s_{i-1}(s_{i-1}^2 + 2)}{(s_{i-1}^2 + 1)^{3/2}} \right) - \varphi(t_i) \left(\frac{t_i s_i^2}{s_i^2 + 1} - \frac{s_i(s_i^2 + 2)}{(s_i^2 + 1)^{3/2}} \right)}{\Phi(t_i) - \Phi(t_{i-1})} - \right. \right. \\ \left. \left. - \left(\frac{\frac{s_{i-1}\varphi(t_{i-1})}{\sqrt{s_{i-1}^2 + 1}} - \frac{s_i\varphi(t_i)}{\sqrt{s_i^2 + 1}}}{\Phi(t_i) - \Phi(t_{i-1})} \right)^2 \right] - \sum_{(2)} \sum_{j=1}^{n_i} \left[\frac{t_{ij}s_{ij}(s_{ij}^2 + 2)}{\theta_1(s_{ij}^2 + 1)^{3/2}} + \frac{s_{ij}^2 - 1}{\theta_1^2(s_{ij}^2 + 1)^2} + \frac{s_{ij}^2}{s_{ij}^2 + 1} \right] \right\}.$$

Уравнение правдоподобия по параметру θ_3 имеет вид

$$\frac{\theta_1}{\theta_2} \left\{ \sum_{(1)} n_i \frac{\frac{\varphi(t_{i-1})}{\sqrt{s_{i-1}^2 + 1}} - \frac{\varphi(t_i)}{\sqrt{s_i^2 + 1}}}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} \sum_{j=1}^{n_i} \left[\frac{t_{ij}}{\sqrt{s_{ij}^2 + 1}} + \frac{s_{ij}}{\theta_1(s_{ij}^2 + 1)} \right] \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_3 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_3^2} = \frac{\theta_1^2}{\theta_2^2} \left\{ \sum_{(1)} n_i \left[\frac{\varphi(t_{i-1}) \left(\frac{t_{i-1}}{s_{i-1}^2 + 1} + \frac{s_{i-1}}{\theta_1 (s_{i-1}^2 + 1)^{3/2}} \right) - \varphi(t_i) \left(\frac{t_i}{s_i^2 + 1} + \frac{s_i}{\theta_1 (s_i^2 + 1)^{3/2}} \right)}{\Phi(t_i) - \Phi(t_{i-1})} - \left(\frac{\frac{\varphi(t_{i-1})}{\sqrt{s_{i-1}^2 + 1}} - \frac{\varphi(t_i)}{\sqrt{s_i^2 + 1}}}{\Phi(t_i) - \Phi(t_{i-1})} \right)^2 \right] - \sum_{(2)} \sum_{j=1}^{n_i} \left[\frac{t_{ij} s_{ij}}{\theta_1 (s_{ij}^2 + 1)^{3/2}} + \frac{s_{ij}^2 - 1}{\theta_1^2 (s_{ij}^2 + 1)^2} - \frac{1}{s_{ij}^2 + 1} \right] \right\}.$$

При одновременном оценивании нескольких параметров максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln [\Phi(t_i) - \Phi(t_{i-1})] + \ln \frac{\theta_1}{\theta_2 \sqrt{2\pi}} \sum_{(2)} n_i - \sum_{(2)} \sum_{j=1}^{n_i} \left[\frac{t_{ij}^2}{2} + \frac{\ln(s_{ij}^2 + 1)}{2} \right]$$

и оценки являются решением соответствующей системы уравнений правдоподобия.

Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \sum_{(1)} n_i \times \sum_{(1)} \frac{[\varphi(t_i) - \varphi(t_{i-1})]^2}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} n_i,$$

$$J_{11} = J(\theta_1) = \sum_{(1)} n_i \times \frac{1}{\theta_1^2} \sum_{(1)} \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]^2}{\Phi(t_i) - \Phi(t_{i-1})} + \frac{2 + \theta_0^2}{\theta_1^2} \sum_{(2)} n_i,$$

$$J_{01} = J(\theta_0, \theta_1) = \sum_{(1)} n_i \times \frac{1}{\theta_1} \sum_{(1)} \left\{ \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]}{\Phi(t_i) - \Phi(t_{i-1})} \times \right. \\ \left. \times [\varphi(t_i) - \varphi(t_{i-1})] \right\} - \frac{\theta_0}{\theta_1} \sum_{(2)} n_i,$$

$$J_{22} = J(\theta_2) = \sum_{(1)} n_i \times \frac{\theta_1^2}{\theta_2^2} \sum_{(1)} \frac{\left(\frac{s_{i-1} \varphi(t_{i-1})}{\sqrt{s_{i-1}^2 + 1}} - \frac{s_i \varphi(t_i)}{\sqrt{s_i^2 + 1}} \right)^2}{\Phi(t_i) - \Phi(t_{i-1})} +$$

$$+ \frac{1}{\theta_2^2} \left(1 + \theta_1^2 - \theta_1^3 \int_{-\infty}^{\infty} \frac{\varphi(t)}{(s^2 + 1)^{3/2}} ds \right) \sum_{(2)} n_i,$$

где $t = \theta_0 + \theta_1 \ln(s + \sqrt{s^2 + 1})$,

$$J_{02} = J(\theta_0, \theta_2) = \frac{\theta_1}{\theta_2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[\varphi(t_i) - \varphi(t_{i-1})] \left[\frac{\varphi(t_{i-1}) s_{i-1}}{\sqrt{s_{i-1}^2 + 1}} - \frac{\varphi(t_i) s_i}{\sqrt{s_i^2 + 1}} \right]}{\Phi(t_i) - \Phi(t_{i-1})} - \right.$$

$$\left. - \sum_{(2)} n_i \times \theta_1 \int_{-\infty}^{\infty} \frac{\varphi(t)}{(s^2 + 1)^{3/2}} ds \right\},$$

$$J_{12} = J(\theta_1, \theta_2) = \frac{1}{\theta_2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[(t_i - \theta_0) \varphi(t_i) - (t_{i-1} - \theta_0) \varphi(t_{i-1})]}{\Phi(t_i) - \Phi(t_{i-1})} \times \right.$$

$$\left. \times \left[\frac{\varphi(t_{i-1}) s_{i-1}}{\sqrt{s_{i-1}^2 + 1}} - \frac{\varphi(t_i) s_i}{\sqrt{s_i^2 + 1}} \right] \right\} + \int_{-\infty}^{\infty} \frac{s \varphi(t)}{s^2 + 1} \left(\theta_0 \theta_1 - \frac{2}{\sqrt{s^2 + 1}} \right) ds \sum_{(2)} n_i,$$

$$J_{33} = J(\theta_3) = \sum_{(1)} n_i \times \frac{\theta_1^2}{\theta_2^2} \sum_{(1)} \frac{\left(\frac{\varphi(t_{i-1})}{\sqrt{s_{i-1}^2 + 1}} - \frac{\varphi(t_i)}{\sqrt{s_i^2 + 1}} \right)^2}{\Phi(t_i) - \Phi(t_{i-1})} +$$

$$+ \frac{\theta_1^3}{\theta_2^2} \int_{-\infty}^{\infty} \frac{\varphi(t)}{(s^2 + 1)^{3/2}} ds \sum_{(2)} n_i,$$

$$J_{03} = J(\theta_0, \theta_3) = \frac{\theta_1}{\theta_2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[\varphi(t_i) - \varphi(t_{i-1})] \left[\frac{\varphi(t_{i-1})}{\sqrt{s_{i-1}^2 + 1}} - \frac{\varphi(t_i)}{\sqrt{s_i^2 + 1}} \right]}{\Phi(t_i) - \Phi(t_{i-1})} \right\} -$$

$$\begin{aligned}
& -\frac{\theta_1^2}{\theta_2} \int_{-\infty}^{\infty} \frac{\varphi(t)}{s^2 + 1} ds \sum_{(2)} n_i, \\
J_{13} = J(\theta_1, \theta_3) &= \frac{1}{\theta_2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]}{\Phi(t_i) - \Phi(t_{i-1})} \times \right. \\
& \times \left[\frac{\varphi(t_{i-1})}{\sqrt{s_{i-1}^2 + 1}} - \frac{\varphi(t_i)}{\sqrt{s_i^2 + 1}} \right] + \theta_0 \theta_1 \int_{-\infty}^{\infty} \frac{\varphi(t)}{s^2 + 1} ds \sum_{(2)} n_i \left. \right\}, \\
J_{23} = J(\theta_2, \theta_3) &= \frac{\theta_1^2}{\theta_2^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{\left[\frac{\varphi(t_{i-1})s_{i-1}}{\sqrt{s_{i-1}^2 + 1}} - \frac{\varphi(t_i)s_i}{\sqrt{s_i^2 + 1}} \right]}{\Phi(t_i) - \Phi(t_{i-1})} \times \right. \\
& \times \left[\frac{\varphi(t_{i-1})}{\sqrt{s_{i-1}^2 + 1}} - \frac{\varphi(t_i)}{\sqrt{s_i^2 + 1}} \right] + \int_{-\infty}^{\infty} \frac{s\varphi(t)}{(s^2 + 1)^{3/2}} \left(\theta_1 + \frac{5}{\theta_1(s^2 + 1)} \right) ds \sum_{(2)} n_i \left. \right\},
\end{aligned}$$

и $J_{01} = J_{10}, J_{02} = J_{20}, J_{12} = J_{21}, J_{03} = J_{30}, J_{13} = J_{31}, J_{23} = J_{32}$.

Результаты оценивания параметров распределения *Su*-Джонсона и последующей проверки критериев согласия приведены на рис. 5.3.26. Выборка объёмом 1000 наблюдений моделировалась с параметрами $\theta_0 = 0, \theta_1 = 1, \theta_2 = 2, \theta_3 = 3$.

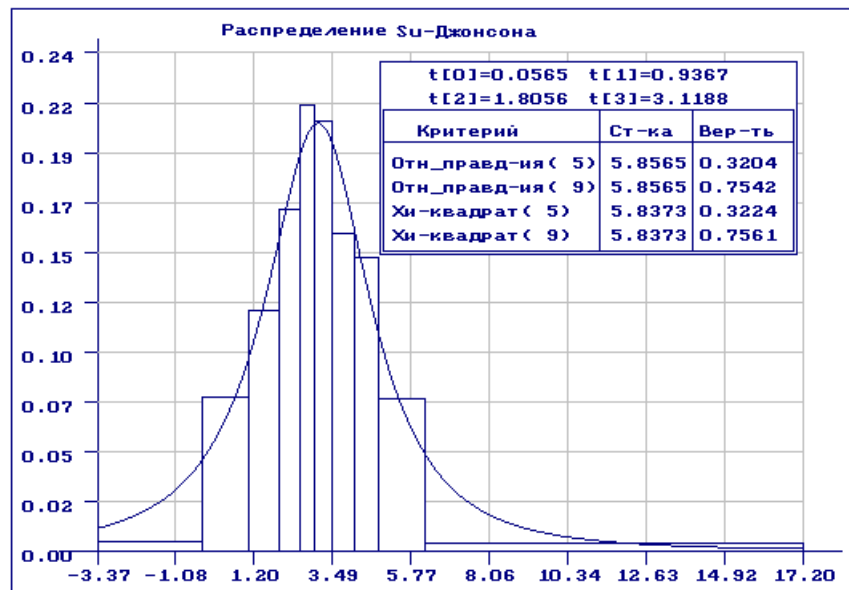


Рис. 5.3.26. Результаты статистического анализа выборки, смоделированной в соответствии с распределением *Su*-Джонсона

5.3.26. Класс экспоненциальных распределений

Функция плотности в этом случае имеет вида

$$f(x) = \frac{\theta_2}{2\theta_1\Gamma(1/\theta_2)} \exp\left\{-\left(\frac{|x-\theta_0|}{\theta_1}\right)^{\theta_2}\right\},$$

где параметр $\theta_0 \in (-\infty, \infty)$, параметры $\theta_1, \theta_2 \in (0, \infty)$ и аргумент $x \in (-\infty, \infty)$. Эта зависимость действительно охватывает целый класс распределений. Частными случаями являются распределения Лапласа и нормальное. Параметры распределения оцениваются по методу максимального правдоподобия.

Оценка максимального правдоподобия параметра θ_0 определяется как решение уравнения правдоподобия

$$\frac{\theta_2}{\theta_1} \left\{ \sum_{(1)} n_i \frac{e^{-|t_{i-1}|\theta_2} - e^{-|t_i|\theta_2}}{\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}} + \sum_{(2)} \sum_{j=1}^{n_i} \frac{|t_{ij}|^{\theta_2}}{t_{ij}} \right\} = 0,$$

где $t_i = \frac{x_i - \theta_0}{\theta_1}$, $t_{ij} = \frac{x_{ij} - \theta_0}{\theta_1}$, $\gamma_i = \gamma(1/\theta_2, |t_i|^{\theta_2}) = \int_0^{|t_i|^{\theta_2}} y^{1/\theta_2-1} e^{-y} dy$ - значения

неполной гамма-функции. Вторая производная логарифма функции правдоподобия по параметру θ_0 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{\theta_2^2}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{\frac{|t_{i-1}|^{\theta_2}}{t_{i-1}} e^{-|t_{i-1}|\theta_2} - \frac{|t_i|^{\theta_2}}{t_i} e^{-|t_i|\theta_2}}{\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}} - \right. \right. \\ \left. \left. - \left(\frac{e^{-|t_{i-1}|\theta_2} - e^{-|t_i|\theta_2}}{\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}} \right)^2 \right] + \left(\frac{1}{\theta_2} - 1 \right) \sum_{(2)} \sum_{j=1}^{n_i} |t_{ij}|^{\theta_2-2} \right\}.$$

Оценка параметра θ_1 находится в качестве решения уравнения правдоподобия

$$\frac{\theta_2}{\theta_1} \left\{ \sum_{(1)} n_i \frac{t_{i-1} e^{-|t_{i-1}|\theta_2} - t_i e^{-|t_i|\theta_2}}{\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}} + \sum_{(2)} \sum_{j=1}^{n_i} \left(|t_{ij}|^{\theta_2} - \frac{1}{\theta_2} \right) \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \frac{\theta_2^2}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{t_{i-1} e^{-|t_{i-1}| \theta_2} \left(|t_{i-1}|^{\theta_2} - \frac{2}{\theta_2} \right) - t_i e^{-|t_i| \theta_2} \left(|t_i|^{\theta_2} - \frac{2}{\theta_2} \right)}{\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}} - \right. \right. \\ \left. \left. - \left(\frac{t_{i-1} e^{-|t_{i-1}| \theta_2} - t_i e^{-|t_i| \theta_2}}{\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}} \right)^2 \right] + \sum_{(2)} \sum_{j=1}^{n_i} \left[\frac{1}{\theta_2^2} - \left(\frac{1}{\theta_2} + 1 \right) |t_{ij}|^{\theta_2} \right] \right\}.$$

Оценка параметра θ_2 находится в качестве решения уравнения правдоподобия

$$\sum_{(1)} n_i \frac{\frac{1}{\theta_2^2} (\gamma'_{i-1} \cdot \text{sign } t_{i-1} - \gamma'_i \cdot \text{sign } t_i) + t_i e^{-|t_i| \theta_2} \ln |t_i| - t_{i-1} e^{-|t_{i-1}| \theta_2} \ln |t_{i-1}|}{\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}} + \\ + \frac{1}{\theta_2^2} \psi \left(\frac{1}{\theta_2} \right) \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right) + \sum_{(2)} \sum_{j=1}^{n_i} \left(\frac{1}{\theta_2} - |t_{ij}|^{\theta_2} \ln |t_{ij}| \right) = 0,$$

где $\gamma'_i = \gamma' \left(1/\theta_2, |t_i|^{\theta_2} \right) = \int_0^{|t_i|^{\theta_2}} y^{1/\theta_2-1} e^{-y} \ln y dy$ - производная неполной гамма-функции по параметру, $\psi(\theta) = \Gamma'(\theta) / \Gamma(\theta)$ - логарифмическая производная гамма-функции (дигамма-функция). Вторая производная логарифма функции правдоподобия по параметру θ_2 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_2^2} = \sum_{(1)} n_i \left\{ \left[-\frac{2}{\theta_2^3} (\gamma'_{i-1} \cdot \text{sign } t_{i-1} - \gamma'_i \cdot \text{sign } t_i) + \right. \right. \\ \left. \left. + \frac{1}{\theta_2^2} \left[\text{sign } t_{i-1} \left(-\frac{\gamma''_{i-1}}{\theta_2^2} + \theta_2 t_{i-1} e^{-|t_{i-1}| \theta_2} \ln^2 |t_{i-1}| \right) - \text{sign } t_i \left(-\frac{\gamma''_i}{\theta_2^2} + \theta_2 t_i e^{-|t_i| \theta_2} \ln^2 |t_i| \right) \right] + \right. \right. \\ \left. \left. + t_{i-1} |t_{i-1}|^{\theta_2} e^{-|t_{i-1}| \theta_2} \ln^2 |t_{i-1}| - t_i |t_i|^{\theta_2} e^{-|t_i| \theta_2} \ln^2 |t_i| \right] / (\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}) + \right. \\ \left. \left. + \left(\frac{\frac{1}{\theta_2^2} (\gamma''_{i-1} \cdot \text{sign } t_{i-1} - \gamma''_i \cdot \text{sign } t_i) + t_i e^{-|t_i| \theta_2} \ln |t_i| - t_{i-1} e^{-|t_{i-1}| \theta_2} \ln |t_{i-1}|}{\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}} \right)^2 \right] \right\} - \\ - \frac{1}{\theta_2^3} \left[2\psi \left(\frac{1}{\theta_2} \right) + \frac{1}{\theta_2} \psi^{\circ} \left(\frac{1}{\theta_2} \right) \right] \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right) - \sum_{(2)} \sum_{j=1}^{n_i} \left(\frac{1}{\theta_2^2} + |t_{ij}|^{\theta_2} \ln^2 |t_{ij}| \right),$$

где $\gamma_i'' = \gamma''(1/\theta_2, |t_i|^{\theta_2}) = \int_0^{|t_i|^{\theta_2}} y^{1/\theta_2-1} e^{-y} \ln^2 y dy$.

При одновременном оценивании нескольких параметров максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln(\gamma_i \cdot \text{sign} t_i - \gamma_{i-1} \cdot \text{sign} t_{i-1}) + \\ + [\ln \theta_2 - \ln(2\theta_1 \Gamma(1/\theta_2))] \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right) - \sum_{(2)} \sum_{j=1}^{n_i} |t_{ij}|^{\theta_2},$$

и оценки являются решением соответствующей системы уравнений правдоподобия. Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \frac{\theta_2^2}{\theta_1^2 \Gamma(1/\theta_2)} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{2(e^{-|t_{i-1}|^{\theta_2}} - e^{-|t_i|^{\theta_2}})^2}{\gamma_i \cdot \text{sign} t_i - \gamma_{i-1} \cdot \text{sign} t_{i-1}} + \Gamma \left(2 - \frac{1}{\theta_2} \right) \sum_{(2)} n_i \right\},$$

$$J_{11} = J(\theta_1) = \frac{\theta_2}{\theta_1^2} \left\{ \frac{\theta_2}{2\Gamma(1/\theta_2)} \sum_{(1)} n_i \times \sum_{(1)} \frac{(t_{i-1} e^{-|t_{i-1}|^{\theta_2}} - t_i e^{-|t_i|^{\theta_2}})^2}{\gamma_i \cdot \text{sign} t_i - \gamma_{i-1} \cdot \text{sign} t_{i-1}} + \sum_{(2)} n_i \right\},$$

$$J_{01} = J(\theta_0, \theta_1) = \frac{\theta_2^2}{2\theta_1^2 \Gamma(1/\theta_2)} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{(e^{-|t_{i-1}|^{\theta_2}} - e^{-|t_i|^{\theta_2}})(t_{i-1} e^{-|t_{i-1}|^{\theta_2}} - t_i e^{-|t_i|^{\theta_2}})}{\gamma_i \cdot \text{sign} t_i - \gamma_{i-1} \cdot \text{sign} t_{i-1}} \right\},$$

$$J_{22} = J(\theta_2) = \frac{1}{2\Gamma(1/\theta_2)} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \left[\frac{1}{\theta_2^2} [\gamma_i' \cdot \text{sign} t_i - \gamma_{i-1}' \cdot \text{sign} t_{i-1} + \right. \right. \\ \left. \left. + \psi \left(\frac{1}{\theta_2} \right) (\gamma_i \cdot \text{sign} t_i - \gamma_{i-1} \cdot \text{sign} t_{i-1}) \right] + \right. \\ \left. + t_i e^{-|t_i|^{\theta_2}} \ln |t_i| - t_{i-1} e^{-|t_{i-1}|^{\theta_2}} \ln |t_{i-1}| \right]^2 / (\gamma_i \cdot \text{sign} t_i - \gamma_{i-1} \cdot \text{sign} t_{i-1}) \right\},$$

$$J_{02} = J(\theta_0, \theta_2) = \frac{\theta_2}{2\theta_1 \Gamma(1/\theta_2)} \left\{ \sum_{(1)} n_i \times \sum_{(1)} (e^{-|t_{i-1}|^{\theta_2}} - e^{-|t_i|^{\theta_2}}) \times \right. \\ \times \left[\frac{1}{\theta_2^2} [\gamma_i' \cdot \text{sign} t_i - \gamma_{i-1}' \cdot \text{sign} t_{i-1} + \psi(1/\theta_2)(\gamma_i \cdot \text{sign} t_i - \gamma_{i-1} \cdot \text{sign} t_{i-1})] + \right. \\ \left. + t_i e^{-|t_i|^{\theta_2}} \ln |t_i| - t_{i-1} e^{-|t_{i-1}|^{\theta_2}} \ln |t_{i-1}| \right] / (\gamma_i \cdot \text{sign} t_i - \gamma_{i-1} \cdot \text{sign} t_{i-1}) \right\},$$

$$J_{12} = J(\theta_1, \theta_2) = \frac{\theta_2}{2\theta_1\Gamma(1/\theta_2)} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \left(t_{i-1} e^{-|t_{i-1}|^{\theta_2}} - t_i e^{-|t_i|^{\theta_2}} \right) \times \right. \\
\times \left[\frac{1}{\theta_2^2} \left[\gamma'_i \cdot \text{sign } t_i - \gamma'_{i-1} \cdot \text{sign } t_{i-1} - \psi(1/\theta_2) (\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}) \right] + \right. \\
\left. \left. + t_i e^{-|t_i|^{\theta_2}} \ln |t_i| - t_{i-1} e^{-|t_{i-1}|^{\theta_2}} \ln |t_{i-1}| \right] / (\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}) \right\} - \\
- \frac{1 + \theta_2 + \psi(1/\theta_2)}{\theta_1 \theta_2} \sum_{(2)} n_i,$$

и $J_{01} = J_{10}$, $J_{02} = J_{20}$, $J_{12} = J_{21}$.

На рис. 5.3.27 приведен пример с результатами статистического анализа выборки объемом в 1000 наблюдений, смоделированной по закону с параметрами $\theta_0 = 0$, $\theta_1 = 2$, $\theta_2 = 3$.

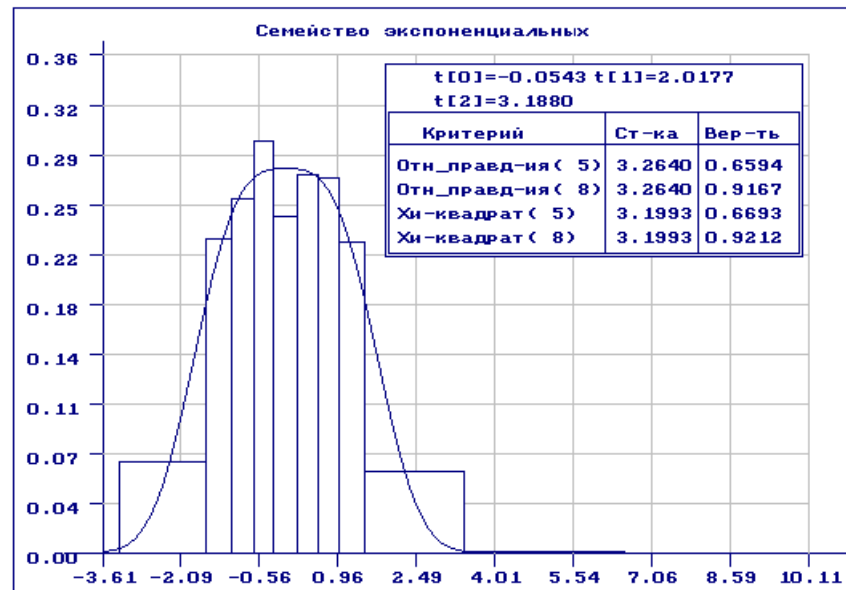


Рис. 5.3.27. Гистограмма и плотность распределения экспоненциального семейства

5.4. Имитация выборочных значений

Имитация выборочных значений (моделирование случайных величин) в программной системе носит вспомогательный характер. При разработке программного обеспечения в основном были использованы методы и алгоритмы, изложенные в [50,110]. Выборки псевдослучайных величин могут моделироваться в соответствии с каждым из включенных в систему распределений. Наиболее часто используемым методом формирования псевдослучайных величин является метод обратных функций, при котором случайная величина X , подчиняющаяся закону с функцией распределения $F(x)$, получается в соответствии с соотношением $X = F^{-1}(Y)$, где $F^{-1}(\cdot)$ - функция, обратная к $F(\cdot)$, а Y - случайная величина, равномерно распределённая на интервале $[0,1]$.

При дальнейшем изложении моделирующих соотношений будем обозначать через Y - случайные величины, равномерно распределённые на интервале $[0,1]$, через Z - случайные величины, распределённые по стандартному нормальному закону с параметрами $(0,1)$.

Псевдослучайная величина, распределённая в соответствии с *экспоненциальным* (показательным) распределением моделируется по формуле

$$X = \theta_1 - \frac{1}{\theta_0} \ln Y .$$

Для *полунормального* распределения -

$$X = \theta_1 + \theta_0 |Z| .$$

Псевдослучайная величина, распределённая по закону *Рэля* -

$$X = \theta_1 + \theta_0 \sqrt{-2 \ln Y} .$$

Для распределения *Максвелла* -

$$X = \theta_1 + \theta_0 \sqrt{\sum_{i=1}^3 Z_i^2} .$$

Псевдослучайные величины, распределённые в соответствии с законом *модуля многомерной нормальной величины порядка t* генерируются в соответствии с соотношением

$$X = \theta_1 + \theta_0 \sqrt{\sum_{i=1}^m Z_i^2} .$$

Для распределения *Парето* -

$$X = \frac{\theta_1}{(1-Y)^{1/\theta_0}}.$$

Распределение *Эрланга* порядка m моделируется в соответствии с соотношением

$$X = \theta_2 - \frac{1}{\theta_1} \ln \left(\prod_{i=1}^m Y_i \right).$$

Псевдослучайные числа по закону *Лапласа* моделируются по формуле

$$X = \begin{cases} \theta_1 + \frac{1}{\theta_0} \ln(2Y), & Y < 0.5; \\ \theta_1 - \frac{1}{\theta_0} \ln(2(1-Y)), & Y > 0.5; \\ \theta_1, & Y = 0. \end{cases}$$

Пара псевдослучайных чисел, распределённых по стандартному нормальному закону с параметрами $(0,1)$, генерируется по формулам

$$Z_1 = \sqrt{-2 \ln Y_1} \sin(2\pi Y_2),$$

$$Z_2 = \sqrt{-2 \ln Y_1} \cos(2\pi Y_2),$$

а нормальное распределение с математическим ожиданием θ_0 и стандартным отклонением θ_1 получается преобразованием стандартной величины

$$X = \theta_0 + \theta_1 Z.$$

Псевдослучайные величины в соответствии с *логарифмически (ln) нормальным* распределением генерируются по формуле

$$X = \exp(\theta_0 + \theta_1 Z),$$

а в соответствии с *логарифмически (lg) нормальным* -

$$X = 10^{\theta_0 + \theta_1 Z}.$$

Для распределения *Коши* псевдослучайные величины получаются по соотношениям

$$X = \theta_1 + \theta_0 \tan[\pi(Y - 0.5)].$$

Для *логистического* распределения

$$X = \theta_0 + \theta_1 \ln \left[\frac{1-Y}{Y} \right].$$

Для распределения *Вейбулла*

$$X = \theta_2 + \theta_1 (-\ln Y)^{1/\theta_0}.$$

Псевдослучайные величины в соответствии с распределением *минимального значения* генерируются по формуле

$$X = \theta_0 + \theta_1 \ln[-\ln(1 - Y)].$$

По закону *максимального значения*

$$X = \theta_0 - \theta_1 \ln[-\ln Y].$$

Числа по *двойному показательному* распределению получаются в соответствии с соотношением

$$X = \theta_2 + -\frac{1}{\theta_0} \ln \left[-\frac{1}{\theta_1} \ln Y \right].$$

Выборка для распределения *Накагами* моделируются по формулам

$$X = \theta_2 + \theta_0 \sqrt{\frac{1}{\theta_1} \left[(\beta - 1) \ln(Y^{1/(1-C)}) - \ln \left(\prod_{i=1}^{[\theta_1]} Y_i \right) \right]},$$

где $C = \theta_1 - [\theta_1]$, $[\theta_1]$ - целая часть θ_1 , и $\beta = Y_1^{1/\theta_1} / (Y_1^{1/\theta_1} + Y_2^{1/\theta_1})$ - случайная величина, отдельно генерируемая в соответствии с бета-распределением.

Для *гамма-распределения* -

$$X = \theta_2 + \frac{1}{\theta_1} \left[(\beta - 1) \ln(Y^{1/(1-C)}) - \ln \left(\prod_{i=1}^{[\theta_0]} Y_i \right) \right],$$

где, аналогично предыдущему распределению Накагами, $C = \theta_0 - [\theta_0]$, $[\theta_0]$ - целая часть θ_0 , и $\beta = Y_1^{1/\theta_0} / (Y_1^{1/\theta_0} + Y_2^{1/\theta_0})$.

Для *бета-распределения первого рода* псевдослучайные величины генерируются по алгоритму Йонка

$$X = \theta_2 + \theta_3 S,$$

где $S = Y_1^{1/\theta_0} / (Y_1^{1/\theta_0} + Y_2^{1/\theta_1})$.

Аналогично, *стандартное бета-распределение второго рода* -

$$X = \theta_2 + \frac{S}{1 - S},$$

Также *бета-распределение второго рода*

$$X = \theta_3 + \frac{S}{\theta_2(1 - S)}.$$

Выборки для распределения *Sb-Джонсона* генерируются на основании величин распределённых по нормальному закону

$$X = \theta_3 + \theta_2 \frac{S}{1+S},$$

где $S = \frac{Z - \theta_0}{\theta_1}$ и Z - стандартная нормальная величина.

Соответственно, для распределения *Sl-Джонсона*

$$X = \theta_3 + \theta_2 e^S,$$

для распределения *Su-Джонсона*

$$X = \theta_3 + \theta_2 \sinh S.$$

Класс *экспоненциальных распределений* (двустороннее экспоненциальное) моделируется с использованием метода обратных функций. И, так как, обратная функция $F^{-1}(Y)$ в этом случае не выражается явно, для вычисления её значений $X = F^{-1}(Y)$ численно решается уравнение $Y = F(X)$.

5.5. Статистический анализ смесей распределений по частично группированным данным

Программная система [76] включает 26 типов и семейств распределений. Но и этого множества далеко недостаточно для адекватного описания реальных случайных величин, встречающихся на практике. Достаточно часто приходится сталкиваться с ситуациями, когда с наблюдаемой выборкой нельзя связать ни одно распределение, из числа включенных в программную систему. Существенно увеличить множество параметрических моделей, используемых для описания реальных случайных величин, можно за счет применения операций над распределениями. Это может быть введение параметров сдвига и масштаба, когда они явно не включены в распределение, усечение закона распределения или формирование смеси распределений.

Вопросы статистического анализа смесей усеченных и неусеченных распределений для частично группированных наблюдений рассматривались в нами в [78, 81, 84, 86, 90]. Использование усечённых законов распределения и смесей законов существенно расширяет множество вероятностных моделей, применяемых для описания реальных данных. Вполне естественным представляется применение смесей распределений в качестве моделей для описания двухмодальных и многомодальных эмпирических распределений.

Усеченные законы распределения возникают и применяются обычно тогда, когда существуют или известны физические ограничения на область определения случайной величины. Смесь законов распределений может образоваться, например, при объединении выборок, распределенных по разным законам, или когда наблюдаемая величина может являться следствием различных причин. Применение смесей распределений в технических приложениях, в задачах надежности рассматривается в работах [193,198, 200, 217, 223, 231].

Естественно, что при использовании для описания реальных величин усеченных законов и смесей различных законов распределений увеличиваются и сложности, в том числе вычислительного характера. Во-первых, увеличивается размерность вектора параметров, который необходимо оценить для идентификации закона, так как добавляются параметры усечения и смеси (см. табл. 5.5.1.). Для усечённых законов распределения решение задачи асимптотически оптимального группирования, имеющее принципиально важное значение для качества статистических выводов, оказывается зависящим от параметров усечения. Это означает, что таблицы асимптотически оптимального группирования для использования в критериях согласия могут формироваться либо для конкретных значений параметров усечения, либо, что более предпочтительно, соответствующая задача должна решаться непосредственно перед проверкой гипотезы о согласии. Остается открытым вопрос решения задачи асимптотически оптимального группирования для смесей неусеченных и усеченных законов распределения. При анализе смесей распределений возникают принципиальные моменты, связанные с оцениванием параметров смеси.

Рассмотрим, что представляет собой область определения параметра смеси. Параметр w выбирают таким образом, чтобы функция плотности была неотрицательной:

$$f(x, w) = wf_1(x) + (1-w)f_2(x) \geq 0, \quad \forall w \in \Omega, \forall x \in X. \quad (5.5.1)$$

Условие (5.5.1) можно преобразовать к виду:

$$w[f_2(x) - f_1(x)] \leq f_2(x).$$

Пусть $X = A \cup B \cup C$, где

Таблица 5.5.1.

Операция	Число дополнит. параметро в	Функция распределения	Функция плотности распределения
Сдвиг	1	$F(x - \mu)$	$f(x - \mu)$
Масштаб	1	$F(x/\sigma)$	$\frac{f(x/\sigma)}{\sigma}$
Усечение слева	1	$\begin{cases} 0, & x < a \\ \frac{F(x) - F(a)}{1 - F(a)}, & x \geq a \end{cases}$	$\begin{cases} 0, & x < a \\ \frac{f(x)}{1 - F(a)}, & x \geq a \end{cases}$
Усечение справа	1	$\begin{cases} \frac{F(x)}{F(b)}, & x \leq b \\ 0, & x > b \end{cases}$	$\begin{cases} \frac{f(x)}{F(b)}, & x \leq b \\ 0, & x > b \end{cases}$
Двустороннее усечение	2	$\begin{cases} 0, & x < a \\ \frac{f(x)}{F(b) - F(a)}, & x \in [a, b] \\ 1, & x > b \end{cases}$	$\begin{cases} 0, & x < a \\ \frac{F(x) - F(a)}{F(b) - F(a)}, & x \in [a, b] \\ 0, & x > b \end{cases}$
Смесь	1	$wF_1(x, \theta_1) + (1 - w)F_2(x, \theta_2)$	$wf_1(x, \theta_1) + (1 - w)f_2(x, \theta_2)$

$$A = \{x \in X : f_2(x) - f_1(x) < 0\};$$

$$B = \{x \in X : f_2(x) - f_1(x) > 0\};$$

$$C = \{x \in X : f_2(x) - f_1(x) = 0\}.$$

Тогда $\Omega = \Omega_A \cap \Omega_B \cap \Omega_C$, где $\Omega_C = R$, а

$$\Omega_A = \left\{ w \in R : w \geq \frac{f_2(x)}{f_2(x) - f_1(x)}, \forall x \in A \right\};$$

$$\Omega_B = \left\{ w \in R : w \leq \frac{f_2(x)}{f_2(x) - f_1(x)}, \forall x \in B \right\}.$$

Обозначим

$$a = \max_{x \in A} \frac{f_2(x)}{f_2(x) - f_1(x)} \leq 0;$$

$$b = \min_{x \in B} \frac{f_2(x)}{f_2(x) - f_1(x)} \geq 1.$$

В результате получим, что область определения параметра смеси имеет вид: $\Omega = \Omega_A \cap \Omega_B \cap \Omega_C = [a, b] \supseteq [0, 1]$.

Когда параметр смеси принадлежит интервалу $0,1$, мы имеем классическую смесь, получаемую, например, объединением выборок. Если же $w \notin [0, 1]$, то одно из распределений входит в смесь со знаком минус и, таким образом, вычитается из другого распределения.

Обобщением смеси двух распределений является смесь из s распределений. Функция распределения в этом случае имеет вид

$$F(x) = \sum_{i=1}^s w_i F_i(x, \theta_i),$$

где s - число распределений в смеси, w_i - параметры смеси, F_i - i -я функция распределения, θ_i - вектор её параметров. Параметры смеси удовлетворяют условию нормировки:

$$\sum_{i=1}^s w_i = 1.$$

Смесь является *идентифицируемой*, если для $\forall w_1, w_2 \in \Omega: w_1 \neq w_2$, следует, что $\exists x \in X: f(x, w_1) \neq f(x, w_2)$. Очевидно, что смесь (5.5.1) неидентифицируема, если $f_1 \equiv f_2$. Численная реализация метода максимального правдоподобия показала, что функция правдоподобия для смеси распределений зачастую является многоэкстремальной, и поэтому получаемые оценки параметров зависят от начального приближения. Качество оценки параметра смеси заметно ухудшается, когда входящие в смесь распределения достаточно близки по форме.

Программное обеспечение, реализующее решение задачи выбора закона распределения или смеси законов, наиболее хорошо описывающих выборочные данные, является дальнейшим развитием программной системы “Статистический анализ одномерных наблюдений случайных величин” [76].

В качестве примера на рис. 5.5.1 приведены результаты статистического анализа негруппированной выборки (оценки параметров и результаты проверки гипотез о согласии по различным критериям), полученной объединением двух выборок: первая распределена по экспоненциальному закону, вторая - по нормальному; каждая объе-

мом по 100 наблюдений (t_0 - параметр сдвига, t_1 - параметр масштаба).

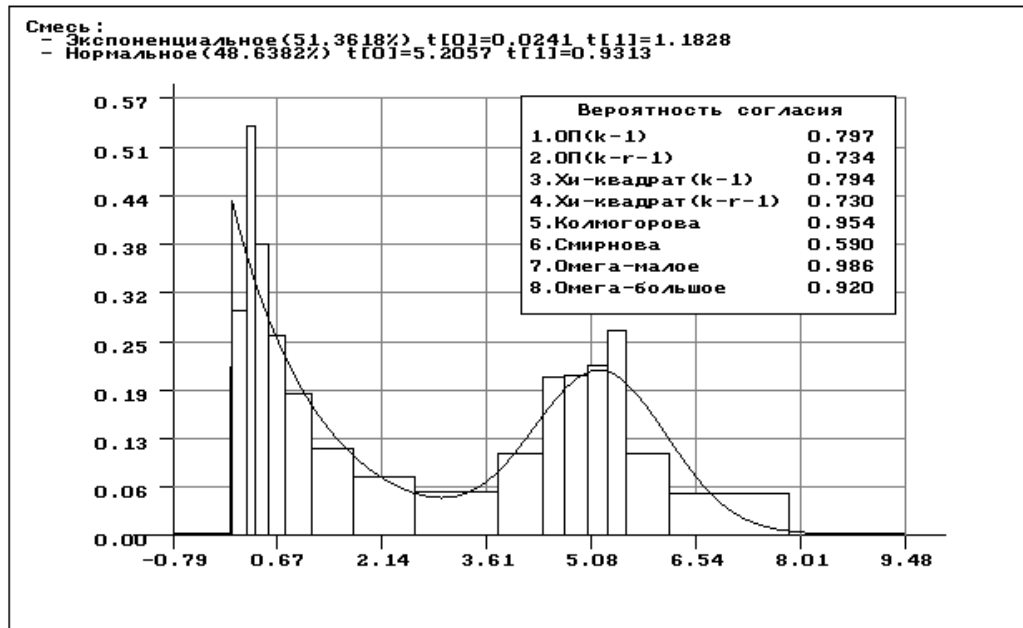


Рис. 5.5.1. Оценивание параметров смеси экспоненциального и нормального распределений

На рис. 5.5.2 приведены результаты статистического анализа выборки, полученной объединением двух выборок, распределённых по нормальному закону: первая объемом 100 наблюдений генерировалась с параметрами (0, 0.5), вторая объемом 900 наблюдений - с параметрами (0, 5).

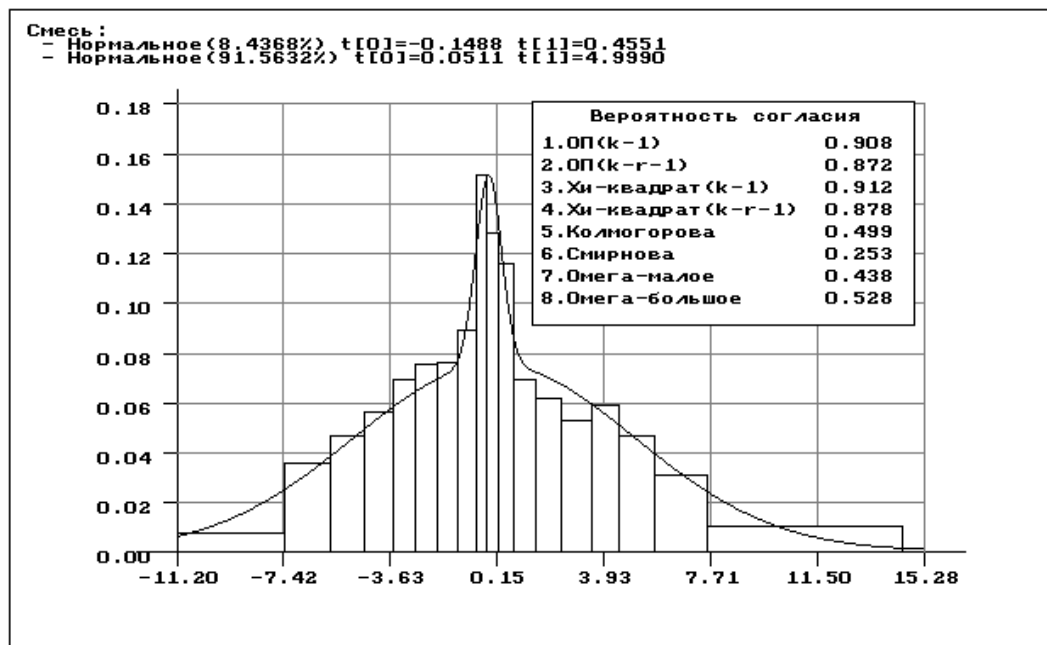


Рис. 5.5.2. Оценивание параметров смеси двух нормальных распределений

На рис. 5.5.3. другой пример смеси двух нормальных законов. Причем в данном случае оценивание параметров смеси осуществлялось по группированной выборке.

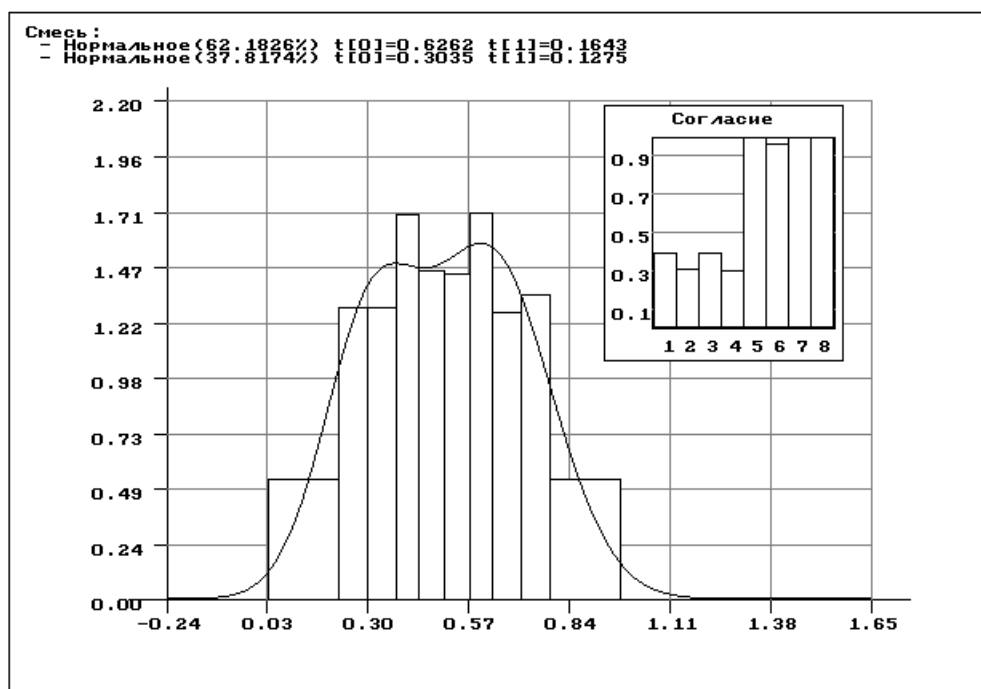


Рис. 5.5.3. Результаты статистического анализа смеси двух нормальных распределений по группированной выборке

В реальных условиях, когда исследуемая выборка действительно является смесью наблюдений двух различных (но с близкими законами) величин, либо, например, наблюдаемая величина по истечении времени несколько меняет свои параметры, а выборка сформирована из наблюдений до и после изменения, при подборе закона распределения иногда удаётся остановиться на каком-то одном законе. Однако, уровень значимости α (вероятность ошибки первого рода), при котором гипотеза о согласии не отвергается, оказывается чрезвычайно малым, за границей разумного. В то же время использование в качестве искомого закона смеси распределений даёт хорошие результаты. Более того, иногда это позволяет подсказать, что наблюдается не одна, а несколько случайных величин.

На рис. 5.5.4 приведены результаты статистического анализа геодезических измерений, где при идентификации их закона распределения в качестве наилучшей модели была выбрана смесь трех нормальных распределений [152]. Оказалось, что выборка была сформирована из 3-х групп аналогичных данных.

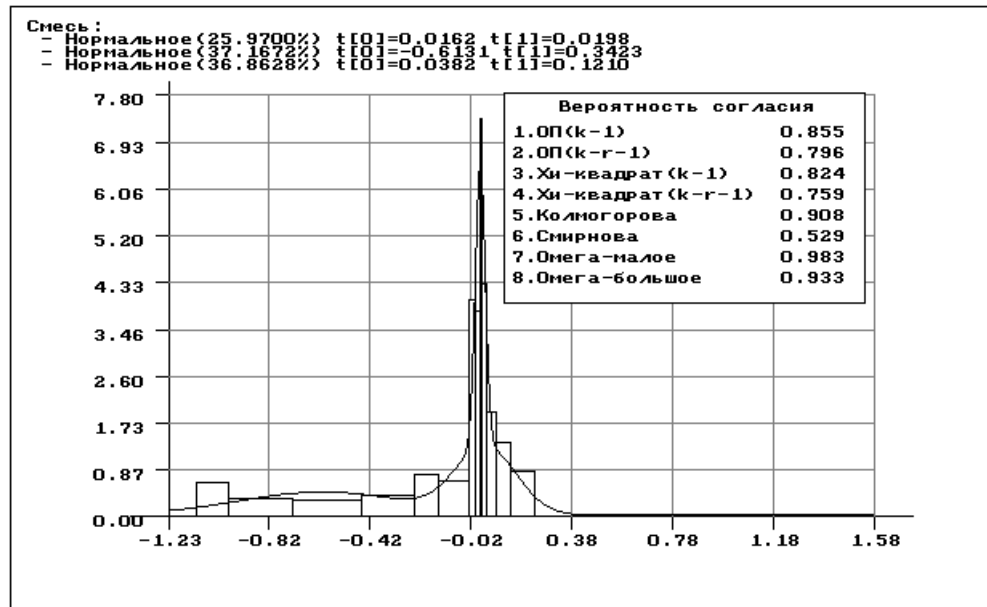


Рис. 5.5.4. Оценивание параметров смеси трех нормальных распределений

На рис. 5.5.5 в качестве примера приведены результаты статистического анализа смеси двух усеченных нормальных распределений по группированным наблюдениям [90].

В [116] по материалам многочисленных публикаций приводятся результаты анализа реальных законов распределений ошибок измерительных приборов. Отмечается, что достаточно часто эти законы оказываются многомодальными. Надо полагать, что использование в качестве модели для описания таких ошибок смесей распределений может оказаться весьма перспективным.

Одним из узких мест параметрической статистики является ограниченность множества параметрических моделей, с помощью которых в различных приложениях пытаются описать наблюдаемые законы случайных величин. Вследствие этой ограниченности часто сталкиваются с ситуацией, когда все попытки использовать один из теоретических законов оказываются неудачными.

Возможность выбора в качестве моделей смесей усеченных и неусеченных распределений наряду с интерпретацией наблюдаемых выборочных данных как интервальных наблюдений, частным случаем которых являются частично группированные наблюдения (в том числе, негруппированные, группированные и цензурированные) представляют собой то основание, на котором базируется разработка новой объектно-ориентированной программной системы статистического анализа одномерных наблюдений.

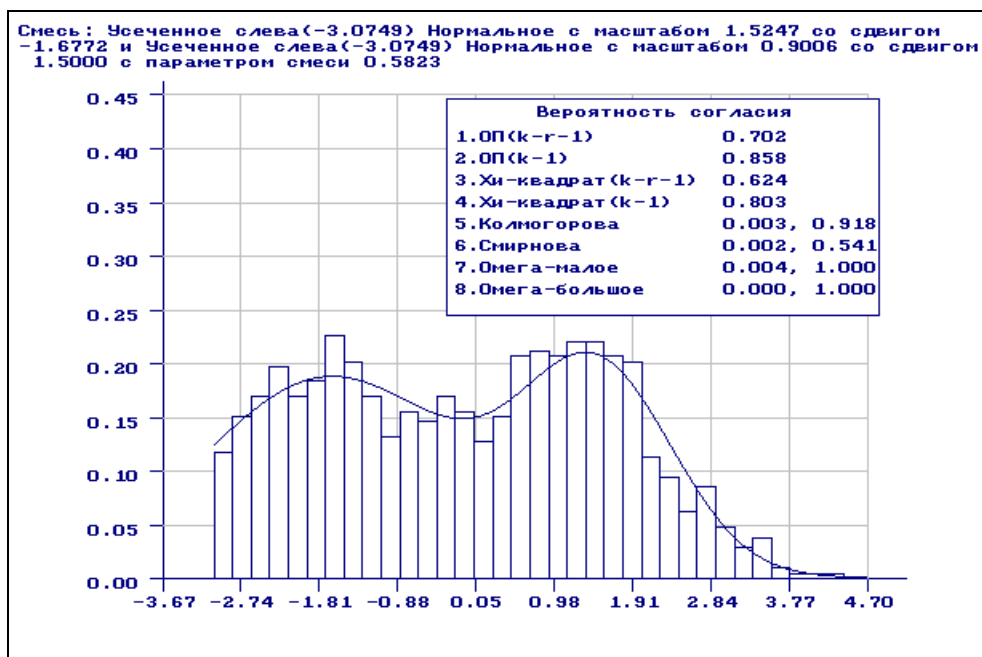


Рис. 5.5.5. Оценивание параметров смеси двух усеченных нормальных распределений по группированной выборке

5.6. Статистический анализ независимых и зависимых случайных величин в геодезии

Исследования, результаты которых обсуждаются в данном параграфе, выполнялись в рамках проекта конкурса грантов по фундаментальным исследованиям в области геодезии и картографии Министерства общего и профессионального образования Российской Федерации. При анализе экспериментальных наблюдений использовались разработанные программные системы [76, 79] и разрабатываемое программное обеспечение объектно-ориентированной системы, обеспечивающее обработку по частично группированным и интервальным наблюдениям смесей усеченных и неусеченных законов распределения [78, 84, 87, 90].

Изучение качества геодезических измерений, законов возникновения и действия неизбежных ошибок измерения является одной из задач математической обработки геодезических измерений. С целью последующего совершенствования методики анализа результатов геодезических измерений проводилось экспериментальное исследование законов распределения ошибок геодезических измерений. Анализ подвергались ошибки измерений, поправки к результатам измерений (превышения), полученные из уравнивания сетей, независимые невязки геометрических условий [158].

На моделях геодезических сетей исследовано влияние аномальных ошибок измерений на законы распределения поправок. Отмечено, что наличие аномальных ошибок измерений приводит к изменению законов распределения поправок. Это сопровождается изменением параметров законов распределения. Следовательно по изменению закона распределения поправок (его вида или параметров) можно судить о наличии грубых ошибок измерений.

Проведено сравнение статистических свойств неравноточных и равноточных поправок, получаемых из уравнивания нивелирных сетей параметрическим способом [159].

С использованием программной системы корреляционного анализа [79], в которой реализованы задачи анализа многомерных наблюдений [6,57], исследована взаимозависимость поправок, получаемых из уравнивания различными способами. Отмечена высокая взаимозависимость рассмотренных способов уравнивания [159].

Проведен статистический анализ разностей средних превышений из двойных ходов [152]. При статистическом анализе использовались данные нивелирования II-го класса на линиях Алтайская - Кош-Агач - Ташанта и Советское - Минусинск - Кызыл. Задачей статистического анализа ставилась идентификация закона распределения разностей средних превышений. На первом этапе идентифицировался закон распределения по всей выборке. Затем, рассматривая разности средних превышений при определенных длинах визирного луча как *равноточные* измерения, были проанализированы законы распределения в этих случаях. Для этих целей разности средних превышений были сгруппированы по длинам визирного луча 10 м, 15 м, 20 м, 25 м, 30 м, 35 м, 40 м, 45 м, 50 м. Разности средних превышений при соответствующих длинах визирного луча приведены в табл. 1-9 [152]. В совокупности данные, приведенные в этих таблицах, образуют исходную полную выборку разностей средних превышений.

Статистический анализ разностей средних превышений для двойных ходов указывает на то, что их реальный закон распределения существенно отличается от нормального. В данном случае наблюдаемое отличие связано с более “тяжелыми хвостами” реального закона распределения по сравнению с нормальным.

На рис. 5.6.1-5.6.2 приведены результаты статистического анализа при попытке выравнивания разностей средних превышений по всей выборке по нормальному закону.

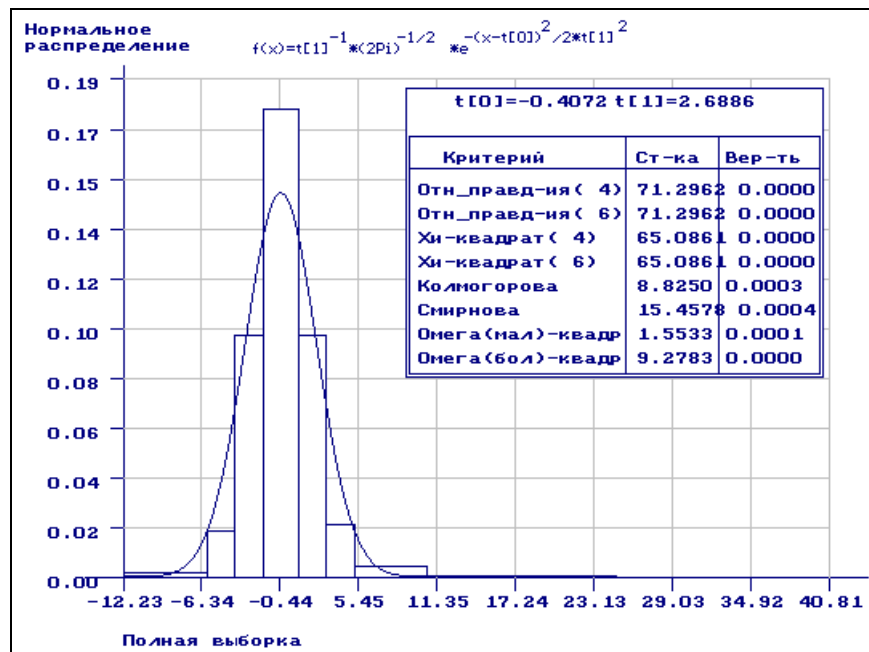


Рис. 5.6.1. Результаты выравнивания разностей средних превышений по всей выборке по нормальному закону

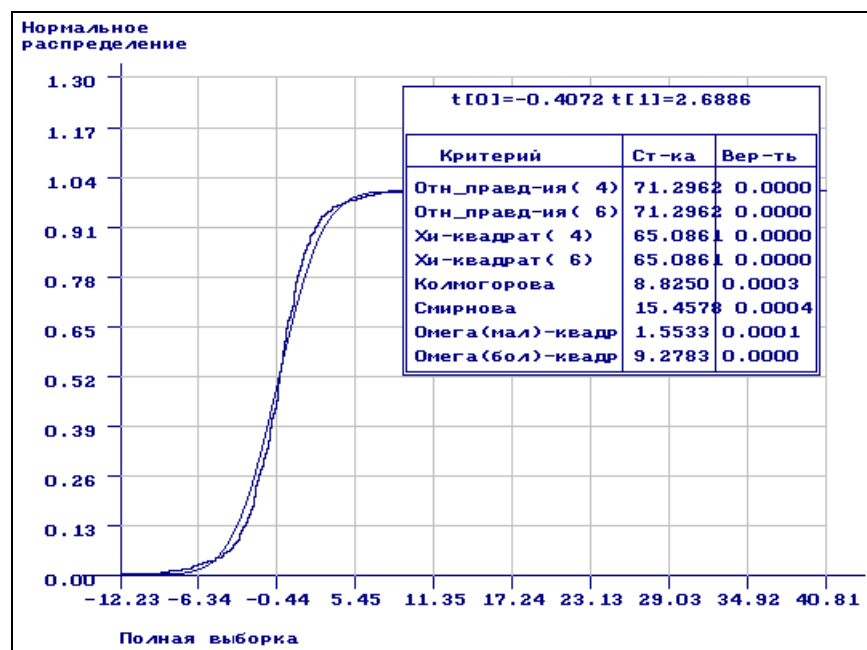


Рис. 5.6.2. Теоретическая и эмпирическая функции распределения при выравнивании разностей средних превышений по всей выборке по нормальному закону

На рис. 5.6.3-5.6.4 приведены результаты статистического анализа при выравнивании разностей средних превышений по нормальному закону с предварительной группировкой наблюдений. В этом случае процедура оценивания устойчива к возможному наличию аномальных наблюдений и к возможным отклонениям от нормальности наблюдаемой величины. При оценивании разбиение выборки

осуществлялось на равночастотные интервалы, так как при этом процедура оценивания обычно наиболее робастна. На рис. 5.6.3 отражены результаты анализа при использовании в критериях согласия равночастотных интервалов, а на рис. 5.6.4 - асимптотически оптимальных интервалов. В последнем случае мощность критериев согласия χ^2 Пирсона и отношения правдоподобия максимальна.

Результаты анализа недвусмысленно показывают, что использование нормального закона для описания разностей средних превышений неправомерно. Реальный закон распределения имеет более тяжелые хвосты. Идентификация закона распределения разностей средних превышений по всей выборке (при любой длине визирного луча), под которой понимается выбор из некоторого множества закона, наиболее хорошо согласующегося с наблюдаемыми значениями случайной величины, приводит к распределению Лапласа. Вид соответствующего распределения и результаты анализа приведены на рис. 5.6.5-5.6.6.

Конечно, результаты анализа не означают, что наблюдаемая случайная величина подчиняется закону Лапласа. Они говорят лишь о том, что в данном случае мы совершим меньшую ошибку, используя в качестве модели этот закон, чем при использовании нормального распределения.

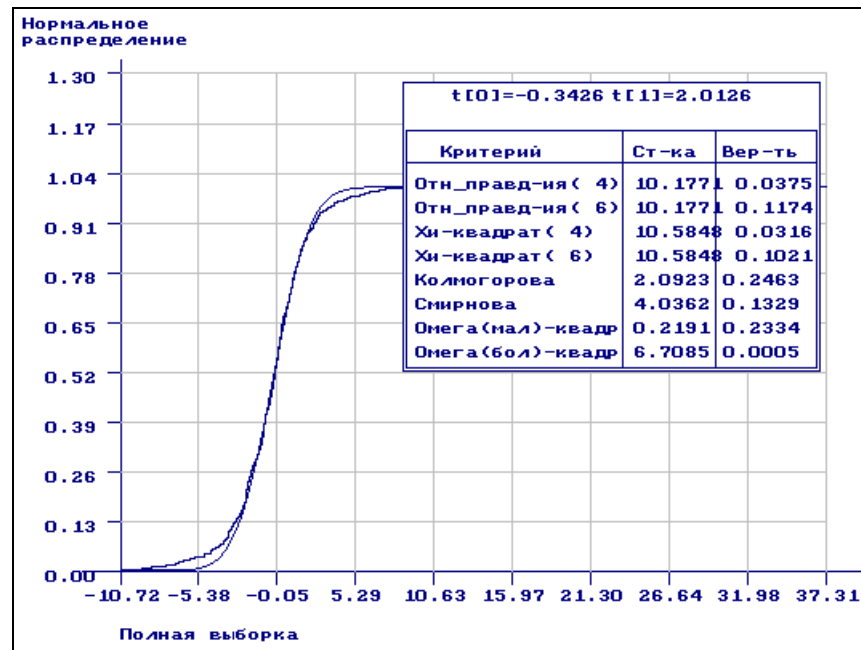


Рис. 5.6.3. Результаты выравнивания разностей средних превышений по нормальному закону с предварительным группированием данных и равночастотным группированием при проверке гипотез о согласии

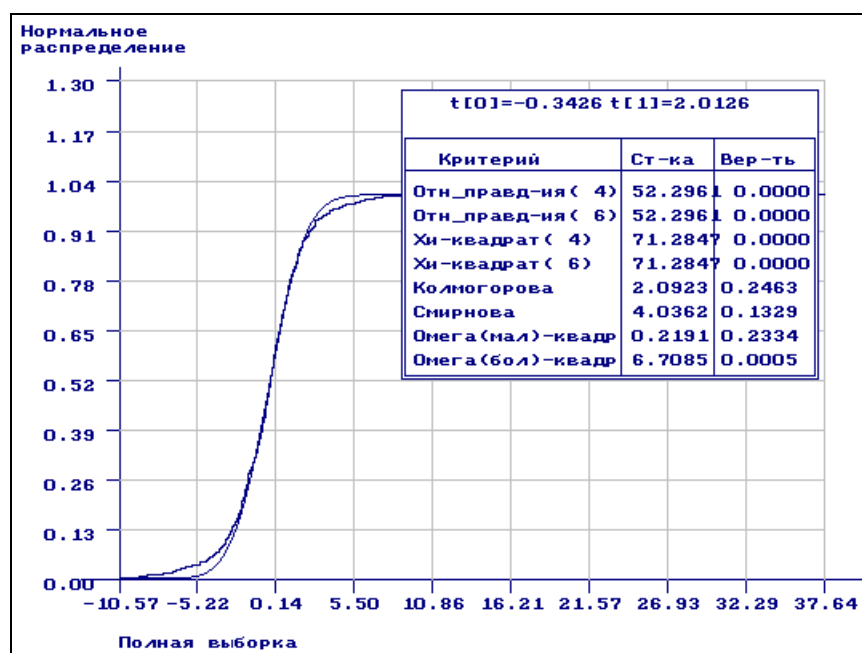


Рис. 5.6.4. Результаты выравнивания разностей средних превышений по нормальному закону с предварительным группированием данных и асимптотически оптимальным группированием при проверке гипотез о согласии

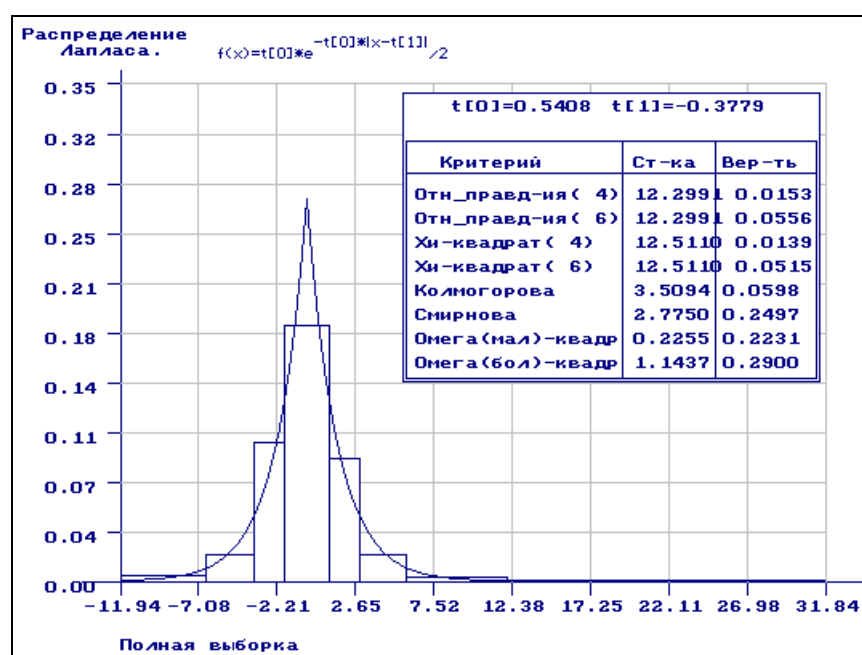


Рис. 5.6.5. Результаты выравнивания разностей средних превышений по закону Лапласа

Полную выборку разностей средних превышений можно рассматривать как смесь совокупности относительно равноточных наблюдений при определенных длинах визирного луча.

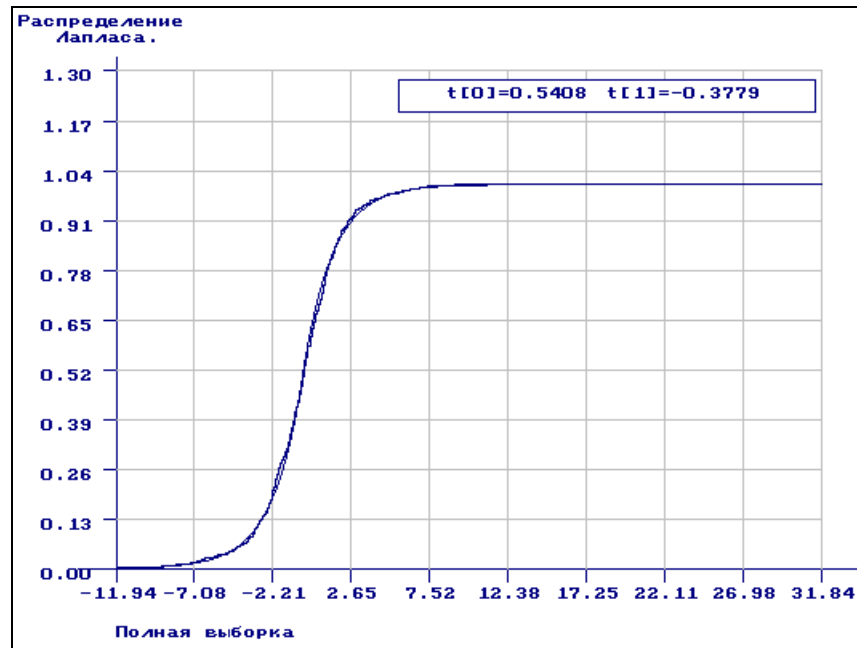


Рис. 5.6.6. Эмпирическая функция распределения и теоретическая функция распределения Лапласа для разностей средних превышений по всей выборке

Рассматривая разности средних превышений при определенных длинах визирного луча как *равноточные* измерения, были проанализированы законы распределения разностей средних превышений при длинах визирного луча 10 м, 15 м, 20 м, 25 м, 30 м, 35 м, 40 м, 45 м, 50 м.

При анализе *равноточных* измерений отличие от нормального закона оказалось более значительным. И это отличие также связано с более “тяжелыми хвостами” реального закона распределения по сравнению с нормальным. К таким законам относятся распределения Лапласа, логистическое и Коши, которые оказались наиболее подходящими моделями для описания закона распределения разностей средних превышений в случае равноточных измерений.

Очевидно, что закон распределения разностей средних превышений должен быть симметричным. Также бесспорным фактом можно считать, что этот закон представляет собой усеченное распределение.

Результаты статистического анализа разностей средних превышений при использовании для описания наблюдаемых величин моделей усеченных законов распределения с указанием идентифицированного закона представлены ниже в табл. 5.6.1.

Наилучшие распределения
(по величине максимума функции правдоподобия)

№	Длина виз. луча	Объе м выб.	Распреде- ление	Пар-тр сдвига	Пар-тр масш-ба	Усеч-е слева	Усеч-е справа	К-во ан. из-й
1.	10 м	62	Усеченно е Лапласа	-0.10	2.09	-3.31	3.41	0
2.	15 м	68	Усеченно е Коши	-0.03	0.49	-2.12	8.72	1
3.	20 м	98	Усеченно е Коши	-0.14	1.45	-10.83	6.33	0
4.	25 м	88	Усеченно е Коши	-0.25	1.31	-7.63	7.43	0
5.	30 м	100	Усеченно е Коши	-0.19	1.82	-24.06	8.76	1
6.	35 м	80	Усеченно е Коши	-0.14	1.38	-7.83	7.53	0
7.	40 м	104	Усеченно е Коши	-0.79	2.41	-8.63	9.13	0
8.	45 м	79	Усеченно е Коши	-0.51	1.86	-10.53	7.43	0
9.	50 м	127	Усеченно е логистическ ое	-0.48	1.50	-9.03	6.53	1

Наилучший закон распределения в данном случае определялся по величине максимума логарифма функции правдоподобия. Т.е. при идентификации закона решалась задача

$$\max_{f \in \Psi} \max_{\theta \in \Theta} \ln L(\theta | x_1, \dots, x_n) = \max_{f \in \Psi} \max_{\theta \in \Theta} \sum_{i=1}^n \ln f(x_i, \theta),$$

где x_i - выборочные наблюдения, f - плотность закона распределения, из определенного множества допустимых (подходящих) законов распределения Ψ , θ - вектор параметров соответствующего закона распределения, принадлежащий множеству его возможных значений Θ .

При использовании для описания наблюдаемых величин усеченных законов распределения ещё более явно прослеживается отличие реального закона от нормального.

В табл. 5.6.1 указано наличие в некоторых выборках аномальных измерений, которые выделялись (отбраковывались) с использованием разработанной параметрической процедуры отбраковки гру-

бых ошибок, базирующейся на робастности оценок, получаемых по предварительно сгруппированным данным.

Реальный закон распределения (выборка) разностей средних превышений при длине визирного луча 15 м достаточно плохо описывается даже усеченным распределением Коши, которое оказалось наилучшим из множества 26 типов непрерывных законов распределения (усеченных и неусеченных). Вид эмпирической функции распределения позволил предположить, что хорошей моделью для реального закона может оказаться смесь законов [81,86]. На рис. 5.6.7 представлены результаты статистического анализа такой смеси. Наиболее хорошо распределение разностей средних превышений при длине визирного луча 15 м описывается смесью законов распределений Коши и Лапласа. Действительно, в этом случае согласие оказывается существенно лучше, чем при описании выборки одним неусеченным или усеченным распределением Коши.

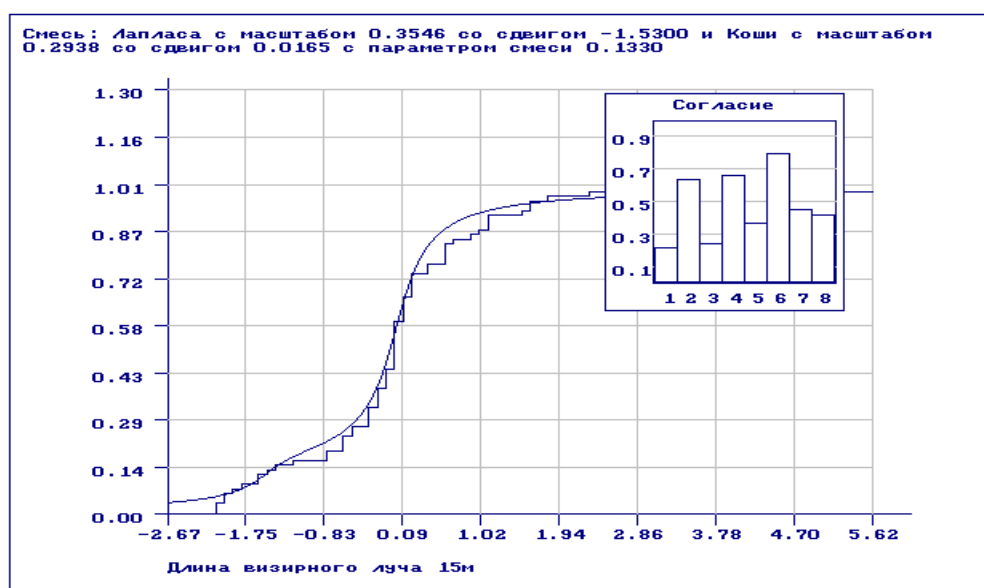


Рис. 5.6.7. Эмпирическая и теоретическая функции распределения смеси законов, описывающие распределение разностей средних превышений при длине визирного луча 15м

В [152] с использованием программной системы [79] был проведен корреляционный анализ поправок, полученных из уравнивания свободных нивелирных сетей различными способами [102]: способом регуляризации - δx_1 и v_1 , для псевдорешения - δx_2 и v_2 , параметрическим способом - δx_3 , v_3 . Здесь δx - поправки к приближенным значениям параметров, v - поправки к результатам измерений.

Анализ матриц парных и частных коэффициентов корреляции дал основание для предположения о высокой взаимозависимости поправок, получаемых различными способами уравнивания. По поводу поправок для псевдорешения $\delta x_2, v_2$ и полученных параметрическим способом $\delta x_3, v_3$ на основании статистического анализа было высказано предположение о равноценности этих способов и их взаимозаменяемости.

Анализ выборок случайных величин, возникающих при обработке различных геодезических измерений, показал:

- Достаточно часто реальные наблюдения представляют собой группированные данные (по-разрядно группированные). Это объясняется ограниченной точностью проводимых измерений или математическими методами обработки. В то же время при статистическом анализе факт естественной группировки наблюдений обычно не учитывается, что и отражается на статистических выводах.
- Предположения о нормальном законе распределения различных поправок и ошибок измерений далеко не всегда соответствуют действительности, так как реально наблюдаемые величины лучше описываются другими законами распределения, обычно с более тяжелыми хвостами, часто усеченными. Иногда удовлетворительное описание наблюдаемой случайной величины параметрической моделью удастся получить только с использованием смесей законов распределения.

5.7. Выводы

Разработана и сопровождается программная система статистического анализа данных, все задачи которой рассматриваются с точки зрения наиболее общего представления экспериментальных наблюдений в виде частично группированных выборок, частными случаями которых являются негруппированные, группированные и цензурированные выборки.

Система позволяет оценивать параметры законов распределения, проверять гипотезы о согласии по критериям χ^2 Пирсона, отношения правдоподобия, Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса, идентифицировать закон распределения, выполнять группирование и сортировку выборки, имитировать выборки, подчиненные заданным законам распределения, отбраковывать аномальные наблюдения. Ре-

жим идентификации обеспечивает выбор из рассматриваемого множества закона, наиболее хорошо согласующегося с исходной выборкой.

Система позволяет вычислять ОМП 26 наиболее часто используемых в приложениях непрерывных законов распределения: экспоненциального, полунормального, Рэлея, Максвелла, модуля многомерного нормального вектора, Парето, Эрланга, Лапласа, нормального, логарифмически нормальных (ln и lg), Коши, Вейбулла, Накагами, распределения минимального значения, распределения максимального значения, двойного показательного, гамма-распределения, логистического, бета-распределения 1-го рода, стандартного бета-распределения 2-го рода, бета-распределения 2-го рода, распределений Sb -Джонсона, Sl -Джонсона и Su -Джонсона, экспоненциального семейства распределений.

При решении всех задач статистического анализа используются полученные таблицы оптимальных граничных точек и соответствующих вероятностей. При проверке согласия с распределениями, для которых отсутствуют таблицы асимптотически оптимального группирования (это касается, например, таких законов, как гамма- и бета-распределения) решение задачи асимптотически оптимального группирования осуществляется непосредственно в процессе проверки гипотез, что обеспечивает максимальную мощность критериев отношения правдоподобия и χ^2 Пирсона.

При проверке сложных гипотез по критериям типа Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса при вычислении вероятностей $P\{S > S^*\}$ используются полученные приближенные предельные законы распределения соответствующих статистик.

Использование в программной системе совокупности критериев согласия даёт возможность принимать более обоснованные решения, а при противоречивости выводов по отдельным критериям формировать компромиссный критерий и делать окончательный вывод с учетом его.

В целях вычисления робастных оценок, устойчивых к наличию аномальных наблюдений и к отклонениям от предположений о виде закона распределения, предусмотрена возможность предварительного группирования данных с последующим определением оценок по группированным данным. При этом возможны различные варианты группирования при оценивании и проверке гипотез о согласии: равномерное, равновероятное и асимптотически оптимальное. Использо-

вание процедуры группирования перед оцениванием нейтрализует влияние грубых ошибок измерений и обеспечивает их выделение.

Программная система используется в учебном процессе, научных исследованиях, обработке данных различных приложений. Различные версии системы были поставлены в ряд вузов страны.

Отличительной особенностью разрабатываемой версии программной системы является расширение множества моделей для описания реальных наблюдений за счет использования смесей усеченных и неусеченных законов распределений, а также возможность обработки интервальных наблюдений. Новая версия реализуется с использованием возможностей объектно-ориентированного программирования. Библиотека классов распределений может наращиваться и использоваться в составе других систем.

В результате анализа выборок случайных величин, возникающих при обработке различных геодезических измерений, показано, что вследствие ограниченной точности измерительных систем и методов обработки реальные наблюдения чаще, чем это считается, представляют собой группированные данные (по-разрядно группированные). В то же время при статистическом анализе факт такой естественной группировки наблюдений обычно не учитывается. Далеко не всегда соответствуют действительности предположения о нормальном законе распределения различных поправок и ошибок измерений, так как обычно выборки реально наблюдаемых величин лучше описываются законами распределения с более тяжелыми хвостами.

ЗАКЛЮЧЕНИЕ

В результате проведенных исследований разработан раздел прикладной статистики по обработке частично группированных и группированных данных, который поднимает на новый уровень решение задач статистического анализа, позволяет получать более надежные статистические выводы. Совокупность созданного алгоритмического и программного обеспечения, построенных статистических таблиц и разработанных методик можно квалифицировать как новое крупное достижение в области прикладной параметрической статистики.

Основные результаты исследований состоят в следующем:

1. Получены условия существования и единственности ОМП по частично группированным данным для ряда одномерных непрерывных распределений.
2. Для широкого ряда распределений при решении задач асимптотически оптимального (по Фишеру) группирования получены граничные точки интервалов в виде, инвариантном относительно параметров распределений, и на их основе сформированы таблицы асимптотически оптимального группирования.
3. Опираясь на таблицы асимптотически оптимального группирования, предложены оценки, использующие выборочные квантили, соответствующие оптимальному группированию. Получены формулы для вычисления данных оценок.
4. Для законов распределения, зависящих только от параметров сдвига и масштаба, предложены оптимальные L-оценки параметров сдвига и масштаба для больших выборок, опирающиеся на таблицы асимптотически оптимального группирования. Вычислены и сформированы таблицы коэффициентов для оптимальных L-оценок.
5. Показана связь мощности критериев согласия с информационной матрицей Фишера по группированным данным. Показано, что применение асимптотически оптимального группирования данных в критериях согласия отношения правдоподобия, χ^2 Пирсона и типа χ^2 обеспечивает максимальную мощность этих критериев при близких конкурирующих гипотезах.
6. Методами статистического моделирования и на основании анализа функций влияния оценок по негруппированным и группированным выборкам показано, что за редким исключением ОМП по негруп-

пированным наблюдениям являются неробастными. Напротив, ОМП по группированным данным и оптимальные оценки параметров сдвига и масштаба по выборочным квантилям устойчивы как к аномальным ошибкам измерений, так и к отклонениям наблюдаемого закона от предполагаемого.

7. Показано, что совокупность вычисления робастных оценок по группированным данным и применение асимптотически оптимального группирования в критериях согласия, используемые на этапе идентификации закона распределения по выборке, содержащей аномальные наблюдения, позволяют эффективно отбраковывать грубые ошибки измерений.
8. Экспериментально исследованы предельные распределения статистик критериев отношения правдоподобия и χ^2 Пирсона в зависимости от способа группирования. Показано, что в случае проверки сложной гипотезы и вычислении ОМП по негруппированным данным предельные распределения статистик критериев отношения правдоподобия и χ^2 Пирсона существенно зависят от способа группирования. При этом в случае применения асимптотически оптимального группирования и справедливой нулевой гипотезе предельные распределения статистик наиболее близки к χ^2_{k-p-1} -распределению.
9. Методами статистического моделирования исследованы предельные распределения статистик непараметрических критериев Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса при проверке сложных гипотез и вычислении оценок параметров по негруппированным данным. Для ряда законов распределения случайных величин идентифицированы законы распределения статистик непараметрических критериев при различном количестве оцененных параметров, построены таблицы верхних процентных точек.
10. Разработана и сопровождается программная система статистического анализа одномерных наблюдений случайных величин, в которой все задачи статистического анализа данных рассматриваются с точки зрения наиболее общего представления экспериментальных наблюдений в виде частично группированных выборок. Реализация в системе всех полученных результатов обеспечивает корректность статистических выводов на всех этапах анализа. Программная система используется в учебном процессе, научных исследованиях, обработке данных различных приложений.

Список литературы

1. Агеев В.В., Благовещенский Ю.Н. О методе максимального правдоподобия для случайно цензурированных интервалов выборок // Статистические методы. - Пермь, 1980. - С. 19-30.
2. Агзамов С.К., Огульник Ю.М. Определение интервальных оценок и точности показателей долговечности по многократно усеченным выборкам // Надежность и контроль качества, 1976. № 9. - С. 49-54.
3. Айвазян С.А. Программное обеспечение персональных ЭВМ по статистическому анализу данных (проблемы, тенденции, перспективы отечественных разработок) // Заводская лаборатория, 1991. Т. 57. № 1. - С. 54-58.
4. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: основы моделирования и первичная обработка данных. - М.: Финансы и статистика, 1983. - 472 с.
5. Алексеева И.У. Теоретическое и экспериментальное исследование законов распределения погрешностей, их классификация и методы оценки их параметров: Автореф. дис. на соиск. учен. степени кан. техн. наук. - Л., 1975. - 20 с.
6. Андерсон Т. Введение в многомерный статистический анализ. - М.: Физматгиз, 1963. - 500 с.
7. Артамоновский В.П., Кордонский Х.Б. Оценка максимального правдоподобия при простейшей группировке данных // Теория вероятностей и ее применение. 1970. Т. XV. № 1. - С. 132-136.
8. Артамоновский В.П. Об оценке максимального правдоподобия параметров сдвига и масштаба по группированным выборкам // Теория вероятностей и ее применение. 1988. Т. XXXIII. № 4. - С. 759-762.
9. Афифи А., Эйзен С. Статистический анализ: Подход с использованием ЭВМ. - М.: Мир, 1982. - 488 с.
10. Барлоу Р., Прошан Ф. Статистическая теория надежности и испытания на безотказность. Пер. с англ. - М.: Наука, 1984. - 328 с.
11. Баталова З.Г., Благовещенский Ю.Н. О точности оценок ресурсов элементов конструкций методом максимума правдоподобия при случайном усечении длительности наблюдений // Надежность и контроль качества, 1979. № 9. - С. 12-20.
12. Баскаков В.Н. Обобщенный метод минимального расстояния для цензурированных данных // Надежность и контроль качества, 1995. № 7. - С. 3-7.
13. Благовещенский Ю.Н. Об асимптотической нормальности одного класса статистик для случайно цензурированных справа выборок // Теория вероятностей и ее применение. 1979. Т. XXIV. № 3. - С. 660-662.
14. Благовещенский Ю.Н. Анализ оценки максимального правдоподобия по случайно цензурированной выборке с малой долей отказов на примере

- экспоненциальных распределений // Заводская лаборатория, 1982. Т. 48. № 3. - С. 50-52.
15. Благовещенский Ю.Н. Оценивание по неполным выборкам. Часть 1. Общая модель // Статистические модели и методы. - Москва, 1984. Вып. 1. - С.4-17.
 16. Благовещенский Ю.Н. Оценивание по неполным выборкам. Часть 2. Модель случайного цензурирования справа // Статистические модели и методы. - Москва, 1984. Вып. 1. - С.17-32.
 17. Бодин Н.А. К теории группированных выборок // Докл. АН СССР, - 1968. - Т. 178. № 1. - С. 17-20.
 18. Бодин Н.А. Оценка параметров распределений по группированным выборкам. // Тр. ин-та им. В.А. Стеклова АН СССР. - 1970. - Т. 111. - С. 110-154.
 19. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. - М.: Наука, 1983. - 416 с.
 20. Бондарев Б.В. О проверке сложных статистических гипотез // Заводская лаборатория. 1986. Т. 52. № 10. - С. 62-63.
 21. Боровков А.А. Математическая статистика. -М.: Наука, 1984. -472 с.
 22. Боровков А.А. Математическая статистика: Дополнительные главы. -М.: Наука, 1984. - 142 с.
 23. Брандт З. Статистические методы анализа наблюдений. - М.: Мир, 1975. - 312 с.
 24. Бурдасов Е.И., Зарифьянц И.Д., Поспелова З.И. Построение функции интенсивности отказов в случае незавершенных испытаний // Надежность и контроль качества, 1976. № 9. - С. 67-71.
 25. Васильев В.П. Численные методы решения экстремальных задач. - М.: Наука, 1980. - 520 с.
 26. Введение в нелинейное программирование / Эльстер К.-Х., Рейнгардт Р., Шойбле М., Донаг Г./ Пер. с нем. - М.: Наука, 1985. - 264 с.
 27. Григорьев Ю.Д., Лемешко Б.Ю., Щеглов Н.В. Оптимизация толщины термореактивной изоляции по кратковременной прочности статистическими методами // Тезисы докладов РНТК "Повышение эффективности производства и использования энергии в условиях Сибири", Иркутск, 1994. Ч. II. - С. 76-78.
 28. Губарев В.В. Вероятностные модели: Справочник. В 2-х ч. / Новосиб. электротехн. ин-т. - Новосибирск, 1992. - 422 с.
 29. Губинский А.И., Денисов В.И., Гречко Ю.П., Лемешко Б.Ю., Цой Е.Б. Методические рекомендации по планированию экспериментов и обработке экспериментальных данных при исследовании надежности и качества функционирования систем "человек-техника". - Препринт / ЛЭТИ им. В.И.Ульянова (Ленина). Л., 1978. - 46 с.
 30. Денисов В.И. Математическое обеспечение системы "ЭВМ-экспериментатор". М.: Наука, 1977. - 251 с.
 31. Денисов В.И., Зачепа Г.Г., Лемешко Б.Ю. Об асимптотически оптимальном группировании при оценивании основного параметра гамма-распределения

- по группированным данным // Применение ЭВМ в оптимальном планировании и проектировании. - Новосибирск, 1974. - С. 50-53.
32. Денисов В.И., Зачепа Г.Г., Лемешко Б.Ю. Об асимптотически оптимальном группировании при оценивании параметров по группированным данным / Новосиб. электротехн. ин-т. - Новосибирск, 1975. - 14 с. - Деп. в ВИНТИ, № 3338-75.
 33. Денисов В.И., Зачепа Г.Г., Лемешко Б.Ю. Об определении максимально допустимой ошибки округления при оценивании регрессионных коэффициентов по округленным наблюдениям / Новосиб. электротехн. ин-т. - Новосибирск, 1975. - 9 с. - Деп. в ВИНТИ, № 3339-75.
 34. Денисов В.И., Зачепа Г.Г., Лемешко Б.Ю. Асимптотически оптимальное группирование при оценивании коэффициентов линейной регрессионной модели по группированным наблюдениям / Новосиб. электротехн. ин-т. - Новосибирск, 1976. - 9 с. - Деп. в ВИНТИ, N 1756-76.
 35. Денисов В.И., Зачепа Г.Г., Лемешко Б.Ю. Оценивание параметров регрессионных моделей при группировании данных // Мат. XIX областной НТК. НТО радиотехники, электроники и связи им. А.С. Попова. Новосибирск, 1976. - с. 5-6.
 36. Денисов В.И., Лемешко Б.Ю. О потерях информации при оптимальном группировании по ошибочному прогнозу / Новосиб. электротехн. ин-т. - Новосибирск, 1978. - 16 с. - Деп. в ВИНТИ, № 1779-78.
 37. Денисов В.И., Лемешко Б.Ю. Асимптотически оптимальное группирование при оценивании масштабного параметра гамма-распределения по группированным данным / Новосиб. электротехн. ин-т. - Новосибирск, 1978. - 19 с. - Деп. в ВИНТИ, № 1778-78.
 38. Денисов В.И., Лемешко Б.Ю. Использование оптимального группирования для оценивания параметров распределения при определении характеристик параметров устройств // Исследования и расчеты надежности энергосистем на этапах проектирования и эксплуатации. - Фрунзе, 1978. - С. 72-75.
 39. Денисов В.И., Лемешко Б.Ю. Оценивание параметров распределения Коши по частично группированным выборкам. - Деп. ВИНТИ, № 249-79. - 1979. - 22 с.
 40. Денисов В.И., Лемешко Б.Ю. Оптимальное группирование при обработке экспериментальных данных // Измерительные информационные системы. - Новосибирск, 1979. - С. 5-14.
 41. Денисов В.И., Лемешко Б.Ю. Вычисление оценок параметров распределений с использованием таблиц асимптотически оптимального группирования // Применение ЭВМ в оптимальном планировании и проектировании. - Новосибирск, 1981. - С. 3-17.
 42. Денисов В.И., Лемешко Б.Ю., Цой Е.Б. Оптимальное группирование, оценка параметров и планирование регрессионных экспериментов. В 2-х ч. / Новосиб. гос. техн. ун-т. - Новосибирск, 1993. - 347 с.
 43. Дэйвид Г. Порядковые статистики. - М.: Наука, 1979. - 336 с.
 44. Дэннис Д., Шнабель Р. Численные методы безусловной оптимизации и решения нелинейных уравнений. - М.: Мир, 1988. - 440 с.

45. Закс Л. Статистическое оценивание. Пер. с нем. - М.: Статистика, 1976. - 598 с.
46. Закс Ш. Теория статистических выводов. - М.: Мир, 1975. - 776 с.
47. Зангвилл У.И. Нелинейное программирование. - М.: Сов. радио, 1973. - 312 с.
48. Зачепа Г.Г. Математическое обеспечение обработки негруппированных и группированных данных на ЭВМ / Диссертация на соискание ученой степени кандидата технических наук. Новосибирск, Новосибирский электротехнический институт, 1975. 165 с.
49. Золотарев И.В., Зотов Л.Г., Лемешко Б.Ю. Анализ, основные исходные уравнения, алгоритм и программа минимизации массы и объема бестрансформаторных низковольтных источников питания каскадного типа на ЭЦВМ. - Деп. НИИЭР, № 3-5863. - 1979. - 96 с.
50. Ермаков С.М., Михайлов Г.А. Статистическое моделирование. М.: Наука, 1982. - 296 с.
51. Ершов А.А. Стабильные методы оценки параметров. Обзор // Автоматика и телемеханика, 1978. № 8. - С. 66-101.
52. Информационно-управляющие человеко-машинные системы: Исследование, проектирование, испытания: Справочник / В.Г. Евграфов, А.И. Губинский, Б.Ю. Лемешко, Е.Б. Цой и др. - М. Машиностроение, 1993. - 528 с.
53. Каган А.М. Об использовании дополнительной информации при оценивании параметров по группированным данным // Материалы семинара "Статистические методы обработки малого числа наблюдений при контроле качества и надежности машин и приборов", 1976. - С. 14-17.
54. Канторович Л.В. О некоторых новых подходах к вычислительным методам и обработке наблюдений // Сиб. мат. журн., 1962. № 3(5). - С. 701-709.
55. Карманов В.П. Математическое программирование. - М.: Наука, 1975. - 272 с.
56. Кендалл М., Стьюарт А.. Статистические выводы и связи. - М.: Наука, 1973. - 900 с.
57. Кендалл М., Стьюарт А.. Многомерный статистический анализ и временные ряды. - М.: Наука, 1976. - 736 с.
58. Ковальков Ю.А., Бойцов Б.В., Череватенко Ю.Л. К вопросу точности оценки закона распределения параметра, используемого для управления надежностью // Надежность и контроль качества, 1976. № 8. - С. 62-68.
59. Колмогоров А.Н. Об эмпирическом определении закона распределения // Избранные труды по ТВ и МС. С. 134-141.
60. Кокс Д., Хинкли Д. Теоретическая статистика. - М.: Мир, 1978. - 560 с.
61. Коробкин А.Д., Лемешко Б.Ю., Цой Е.Б. Математические методы оптимизации. Учебное пособие. - Новосибирск, 1977. Ч.1. - 99 с.
62. Крамер Г. Математические методы статистики. - М.: Мир, 1975. - 648 с.
63. Кузнецов В.П. Интервальные статистические модели. -М.: Радио и связь, 1991. - 348 с.

64. Кулинская Е.В., Саввушкина Н.Е. О некоторых ошибках в реализации и применении непараметрических методов в пакете для IBM PC // Заводская лаборатория, 1990. Т. 56. № 5. - С. 96-99.
65. Куллдорф Г. Введение в теорию оценивания по группированным и частично группированным выборкам. - М.: Наука, 1966. - 176 с.
66. Леман Э. Теория точечного оценивания. - М.: Наука, 1991. - 448 с.
67. Леман Э. Проверка статистических гипотез. - М.: Наука, 1979. - 408 с.
68. Лемешко Б.Ю., Цой Е.Б. О скорости сходимости метода сопряженного градиента // Применение ЭВМ в оптимальном планировании и проектировании. - Новосибирск, 1974. - С. 87-90.
69. Лемешко Б.Ю. Об оценивании параметров распределений по группированным наблюдениям // Вопросы кибернетики. - М., 1977. - Вып. 30. - С. 80-96.
70. Лемешко Б.Ю. Об оценивании по частично группированным выборкам параметров распределения Эрланга // Применение ЭВМ в оптимальном планировании и проектировании. - Новосибирск, 1982. - С. 90-93.
71. Лемешко Б.Ю., Панфилов А.Г., Панфилова Л.В.. Асимптотически оптимальное группирование данных при анализе наблюдений, распределенных по законам Лапласа и двойного показательного // Математическое обеспечение стохастических и детерминированных моделей. - Новосибирск, 1986. - С. 15-23.
72. Лемешко Б.Ю. К вопросу решения задачи асимптотически оптимального группирования данных при обработке наблюдений, подчиняющихся бета-распределению // Машинные методы оптимизации, моделирования и планирования эксперимента. - Новосибирск, 1988. - С. 134-138.
73. Лемешко Б.Ю. Программная система "Оценивание параметров распределений" // Тезисы докладов Российской НТК "Информатика и проблемы телекоммуникаций", Новосибирск, 1994. - С. 128-129.
74. Лемешко Б.Ю. Программная система "Оценивание параметров распределений" // Тезисы докладов международной НТК "Идентификация, измерение характеристик и случайных сигналов", Новосибирск, 1994. С. 36-37.
75. Лемешко Б.Ю. Программная система статистического анализа "Оценивание параметров распределений" // Труды второй международной конференции "Актуальные проблемы электронного приборостроения АПЭП-94", Новосибирск, 1994. Т.2. С. 46-51.
76. Лемешко Б.Ю. Статистический анализ одномерных наблюдений случайных величин: Программная система. - Новосибирск:Изд-во НГТУ. - 1995. - 125 с.
77. Лемешко Б.Ю., Постовалов С.Н. Проверка непараметрических гипотез по группированным данным // Материалы международной НТК "Информатика и проблемы телекоммуникаций", Новосибирск, 1995. Т.1. - С. 63-65.
78. Лемешко Б.Ю., Постовалов С.Н. Статистический анализ смесей распределений по группированным данным // Материалы международной НТК

- “Информатика и проблемы телекоммуникаций”, Новосибирск, 1995. Т.1. - С. 83-85.
79. Лемешко Б.Ю. Корреляционный анализ многомерных наблюдений случайных величин: Программная система. - Новосибирск: Изд-во НГТУ, 1995. - 39 с.
80. Лемешко Б.Ю. Исследование операций. Методические указания. - Новосибирск: Изд-во НГТУ, 1995. - 48 с.
81. Лемешко Б.Ю., Постовалов С.Н. Статистический анализ одномерных наблюдений по частично группированным данным // Изв. вузов. Физика. - Томск, 1995. - № 9. - С. 39-45.
82. Лемешко Б.Ю., Постовалов С.Н. К использованию непараметрических критериев по частично группированным данным // Сб. научных трудов НГТУ. - Новосибирск: изд-во НГТУ. 1995. - №2. -С. 21-30.
83. Лемешко Б.Ю., Постовалов С.Н. Статистический анализ наблюдений, имеющих интервальное представление // Сб. научных трудов НГТУ. - Новосибирск: изд-во НГТУ. - 1996. - №1. - С. 3-12.
84. Лемешко Б.Ю., Постовалов С.Н. Система статистического анализа смесей и усеченных распределений случайных величин // Тезисы докладов Российской НТК “Информатика и проблемы телекоммуникаций”, Новосибирск, 1996. Т.2. - С. 38-39.
85. Лемешко Б.Ю., Постовалов С.Н. Вопросы обработки выборок одномерных случайных величин // Научный вестник НГТУ. - Новосибирск, 1996. № 2. - С. 3-24.
86. Лемешко Б.Ю., Постовалов С.Н. Статистический анализ смесей распределений по частично группированным данным // Сб. научных трудов НГТУ. - Новосибирск: изд-во НГТУ. 1995. - №1. -С. 25-31.
87. Лемешко Б.Ю., Постовалов С.Н. Система статистической обработки случайных наблюдений // Тезисы докладов международной НТК “Информационные технологии в моделировании и управлении”, С.-Петербург, 1996. - С.155-159.
88. Лемешко Б.Ю. Оптимальные оценки параметров сдвига и масштаба по выборочным квантилям для больших выборок // Тр. третьей МНТК “Актуальные проблемы электронного приборостроения АПЭП-96”, Новосибирск, 1996. - Т. 6, ч.1. - С.37-44.
89. Лемешко Б.Ю., Постовалов С.Н. Робастные алгоритмы оценивания и параметрические методы отбраковки аномальных наблюдений // Тр. третьей МНТК “Актуальные проблемы электронного приборостроения АПЭП-96”, Новосибирск, 1996. - Т. 6, ч.1. - С.45-49.
90. Лемешко Б.Ю., Постовалов С.Н. Программное обеспечение статистического анализа смесей случайных величин, представленных частично группированными и интервальными выборками // Тр. третьей МНТК “Актуальные проблемы электронного приборостроения АПЭП-96”, Новосибирск, 1996. - Т. 6, ч.1. - С.50-53.
91. Лемешко Б.Ю., Постовалов С.Н. Система статистического анализа одномерных непрерывных распределений случайных величин (версия 3.0) //

- Мат. III международной НТК “Микропроцессорные системы автоматики”, Новосибирск, 1996. - С. С-16 - С-17.
92. Лемешко Б.Ю., Постовалов С.Н. Статистический анализ одномерных непрерывных распределений случайных величин // Тезисы докладов Второго Сибирского Конгресса по Прикладной и Индустриальной Математике (ИНПРИМ-96), Новосибирск, 1996. - С. 178-179.
 93. Лемешко Б.Ю., Постовалов С.Н. К вопросу о робастности оценок по группированным данным // Сб. научных трудов НГТУ. -Новосибирск: изд-во НГТУ. - 1996. - № 2(4). - С. 9-18.
 94. Лемешко Б.Ю., Постовалов С.Н. Программное обеспечение задач статистического анализа одномерных непрерывных случайных величин // Мат. международной научно-методической конференции “Новые информационные технологии в университетском образовании”, Новосибирск, НИИМИОО, 1997. - С. 44-45.
 95. Лемешко Б.Ю., Постовалов С.Н. О распределениях статистик непараметрических критериев при потере свойства “свободы от распределения” // Мат. международной НТК “Информатика и проблемы телекоммуникаций”. - Новосибирск, 1997. - С. 117-120.
 96. Лемешко Б.Ю., Постовалов С.Н. О влиянии способа группирования данных на распределения статистик χ^2 Пирсона и отношения правдоподобия // Мат. международной НТК “Информатика и проблемы телекоммуникаций”. - Новосибирск, 1997. - С. 120-123.
 97. Лемешко Б.Ю., Постовалов С.Н. К вопросу о распределениях статистик непараметрических критериев согласия // Сб. научных трудов НГТУ. - 1997. - № 1(6). - С. 23-32.
 98. Лемешко Б.Ю., Постовалов С.Н. О решении задач статистического анализа интервальных наблюдений // Вычислительные технологии, 1997. Т.2. № 1. - С. 28-36.
 99. Лемешко Б.Ю. Робастные методы оценивания и отбраковка аномальных измерений // Заводская лаборатория, 1997. Т.63. № 5. - С. 43-49.
 100. Лемешко Б.Ю. Группирование наблюдений как способ получения робастных оценок // Надежность и контроль качества, 1997. № 5. - С. 26-35.
 101. Лемешко Б.Ю. Асимптотически оптимальное группирование наблюдений - это обеспечение максимальной мощности критериев // Надежность и контроль качества, 1997. № 8. - С. 3-14.
 102. Лесных Н.Б., Малиновский А.Л., Мизина Г.И. Анализ результатов уравнивания нивелирных сетей АЭС //Совершенствование методов инженерно-геодезических работ. НИИГАиК. - Новосибирск, 1988.
 103. Мартынов Г.В. Критерии омега-квадрат. - М.: Наука, 1978. - 80 с.
 104. Методические рекомендации по оптимальному группированию данных в задачах обработки статистической информации / Исследование методов статистического моделирования. Новосибирск, 1980. Заключительный отчет по НИР. Научн. рук. Денисов В.И., отв. исп. Лемешко Б.Ю. № гос. рег. 76072132, Инв. № Б941799. - 64 с.

105. Методические указания для определения показателей долговечности изделий по результатам незавершенных испытаний или наблюдений / Научн. рук.: Благовещенский Ю.Н., Кугель Р.В. - М.: ГОНТИ-НАТИ, 1980. - 80 с.
106. Мизина Г.И. О критерии выбора оптимального способа уравнивания // Совершенствование инженерно-геодезических работ. НИИГАиК. - Новосибирск, 1990.
107. Мирвалиев М., Никулин М.С. Критерии согласия типа хи-квадрат / Заводская лаборатория. 1992. Т. 58. № 3. - С.52-58.
108. Миронова Н.Г., Орлов А.И. Одношаговые оценки для параметров гамма-распределения // Надежность и контроль качества, 1988. № 9. - С. 18-22.
109. Митропольский А.К. Техника статистических вычислений. - М.: Наука, 1971. - 479 с.
110. Михайлов Г.А. Некоторые вопросы теории методов Монте-Карло. - Новосибирск: Наука, 1974. - 142 с.
111. Моисеев Н.Н., Иванюков Ю.П., Столярова Е.М. Методы оптимизации. - М.: Наука, 1978. - 352 с.
112. Несененко Г.А., Тюрин Ю.Н. Асимптотика статистики Колмогорова для параметрических семейств // Докл. АН СССР, 1978. Т.239. №6.
113. Никитин Я.Ю. Асимптотическая эффективность непараметрических критериев. - М.: Физматлит. 1995. - 240 с.
114. Никулин М.С. Критерий хи-квадрат для непрерывных распределений с параметрами сдвига и масштаба // Теория вероятностей и ее применение. 1973. Т. XVIII. № 3. - С. 583-591.
115. Никулин М.С. О критерии хи-квадрат для непрерывных распределений // Теория вероятностей и ее применение. 1973. Т. XVIII. № 3. С.675-676.
116. Новицкий П.В., Зограф И.А. Оценка погрешностей результатов измерений. - Л.: Энергоатомиздат, 1991. - 303 с.
117. Объектно-ориентированная программная система статистического анализа: Таблицы коэффициентов для оптимальных L-оценок параметров сдвига и масштаба по выборочным квантилям больших выборок и таблицы асимптотически оптимального группирования наблюдений // Отчет по НИР, НГТУ, 1996. № гос.рег. 01.9.70 000550, инв. № 02.9.70 000190. Научн. рук. Лемешко Б.Ю. - 129 с.
118. Орлов А.И., Орловский И.В. О поправках на группировку / Прикладной многомерный статистический анализ. - М.: Наука, 1978. - С. 339-342.
119. Орлов А.И. О проверке симметрии распределения // Теория вероятностей и ее применения, 1972. Т. 17. № 2. - С. 372-377.
120. Орлов А.И. Распространенная ошибка при использовании критериев Колмогорова и омега-квадрат // Заводская лаборатория, 1985. Т. 51. № 1. - С. 60-62.
121. Орлов А.И. О нецелесообразности использования итеративных процедур нахождения оценок максимального правдоподобия // Заводская лаборатория, 1986. Т. 52. № 5. - С. 67-69.

122. Орлов А.И. О развитии реалистической статистики / Статистические методы оценивания и проверки гипотез, Пермский у-тет, Пермь, 1990. - С. 89-99.
123. Орлов А.И. Некоторые алгоритмы реалистической статистики / Статистические методы оценивания и проверки гипотез, Пермский у-тет, Пермь, 1991. - С. 77-86.
124. Орлов А.И. Часто ли распределение результатов наблюдений является нормальным? // Заводская лаборатория, 1991. Т. 57. № 7. - С. 64-66.
125. Орлов А.И. О современных проблемах внедрения прикладной статистики и других статистических методов // Заводская лаборатория. 1992. Т. 58. № 1. - С. 67-74.
126. Орлов А.И. Интервальная статистика // Интервальные вычисления, 1992. № 1. - С. 44-52.
127. Орлов А.И. Неустойчивость параметрических методов отбраковки резко выделяющихся наблюдений // Заводская лаборатория. 1992. Т. 58. № 7. С. 40-42.
128. Орлов А.И. О критериях Колмогорова и Смирнова // Заводская лаборатория, 1995. Т. 61. № 7. - С. 59-61.
129. Орлов А.И. Проблемы множественных проверок статистических гипотез // Заводская лаборатория, 1996. Т. 62. № 5. - С. 51-54.
130. Ортега Дж., Рейнболдт Г. Итерационные методы решения нелинейных систем уравнений со многими неизвестными. - М.: Мир, 1975. - 558 с.
131. Павлов И.В. Доверительные границы для показателей надежности для цензурированных выборок // Надежность и контроль качества, 1995. № 7. - С. 8-13.
132. Петелин В.А. К вопросу о верхней доверительной границе интенсивности отказов в многократно цензурированной выборке // Надежность и контроль качества, 1990. № 8. - С. 5-9.
133. Петрович М.П., Давидович М.И. Статистическое оценивание и проверка гипотез на ЭВМ. - М.: Финансы и статистика. - 191 с.
134. Подиновский В.В., Ногин В.Д. Парето-оптимальные решения многокритериальных задач. - М.: Наука, 1982. - 254 с.
135. Полак Э. Численные методы оптимизации. Единый подход. - М., Мир, 1974. - 376 с.
136. Пугачев В.С. Теория вероятностей и математическая статистика. - М.: Наука, 1979. - 496 с.
137. Пшеничный Б.Н., Данилин Ю.М. Численные методы экстремальных задач. - М.: Наука, 1975. - 320 с.
138. Рао. С.Р. Линейные статистические методы и их применения. - М.: Наука, 1968. - 548 с.
139. Румшицкий Л.З. Математическая обработка результатов эксперимента: Справочное руководство. - М.: Наука, 1971. - 192 с.
140. Саввушкина Н.Е. Критерий Колмогорова-Смирнова для логистического и гамма-распределения // Сб. тр. / ВНИИ систем. исслед. - 1990. № 8. - С. 50-56.

- 141.Сархан А.Е., Гринберг Б.Г. Введение в теорию порядковых статистик. - М.: Статистика, 1970. - 414 с.
- 142.Сea Ж. Оптимизация. Теория и алгоритмы. - М.: Мир, 1973. -244 с.
- 143.Смирнов Н.В. О распределении ω^2 критерия Мизеса // Математический сборник, 1937, Т. 44. С. 973-994.
- 144.Смирнов Н.В. Приближение законов распределений случайных величин по эмпирическим данным // Успехи математических наук, 1944. - С. 179-206.
- 145.Смирнов Н.В. О критерии Крамера-Мизеса // Успехи математических наук, 1949, 4, вып. 4. С. 196-197.
- 146.Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностей и математической статистики для технических приложений. -М.: Наука, 1969.
- 147.Смоляк С.А., Титаренко Б.П. Устойчивые методы оценивания: (Статистическая обработка неоднородных совокупностей). - М.: Статистика, 1980. - 208 с.
- 148.Справочник по прикладной статистике. В 2-х т. Т.1: Пер с англ. / Под ред. Э.Ллойда, У.Ледермана, Ю.Н.Тюрина. - М.: Финансы и статистика, 1989. - 510 с.
- 149.Справочник по специальным функциям. Под ред. М. Абрамовица и И. Стиган. - М.: Наука, 1979. - 832 с.
- 150.Статистические методы в экспериментальной физике. - М.: Атомиздат, 1976. - Идьe В., Драйард Д., Джеймс Ф., Рус М., Садуле Б. - 335 с.
- 151.Статистические методы обработки эмпирических данных: Рекомендации. - М.: Издательство стандартов. 1978. - 232 с.
- 152.Статистический анализ независимых и зависимых случайных величин в геодезии // Отчет по НИР, НГТУ, 1996. № гос.рег. 01.9.60 003120, инв. № 02.9.60 008130. Научн. рук. Лемешко Б.Ю. - 69 с.
- 153.Тарасенко Ф.П. Некоторые вопросы обработки результатов измерений с точки зрения теории информации // Труды СФТИ. - Томск, 1959. Вып. 37. - С. 212-218.
- 154.Тарасенко Ф.П. Об оптимальном пороге квантования принимаемого сигнала для бинарных систем обнаружения / Изв. вузов. Радиотехника. - Киев, 1959. - № 3. - С. 292-298.
- 155.Тарасенко Ф.П., Захаров В.В. Оптимальное квантование принимаемого сигнала на несколько уровней / Изв. вузов. Радиотехника. - Киев, 1962. - № 9. - С. 11-14.
- 156.Тарасенко Ф.П. Непараметрическая статистика. - Томск: Изд-во Том. ун-та, 1976. - 292 с.
- 157.Таушанов З., Тонева Е., Пенова Р. Вычисление энтропийного коэффициента при малых выборках // Изобретательство, стандартизация и качество, 1973. № 5. - София.
- 158.Теория, методы и программное обеспечение задач статистического анализа независимых и зависимых случайных величин в геодезии / Отчет по НИР, НГТУ, 1994 г. № гос.рег. 01.9.50 001519, инв. № 02.9.50 001199. Денисов В.И., Лемешко Б.Ю., Лесных Н.Б., Постовалов С.Н. - 40 с.

159. Теория, методы и программное обеспечение задач статистического анализа независимых и зависимых случайных величин в геодезии: Вероятностные модели, непараметрические критерии, анализ / Отчет по НИР, НГТУ, 1995. № гос.рег. 01.9.60 003119, инв. № 02.9.60 005332. Денисов В.И., Лемешко Б.Ю., Лесных Н.Б., Постовалов С.Н. - 42 с.
160. Тескин О.И. Диалоговая система НАДИС - инструмент статистического анализа надежности технических изделий // Надежность и контроль качества, 1994. № 3. - С. 32-64.
161. Тихов М.С. Статистический анализ по группированным и цензурированным наблюдениям // Статистические методы. - Пермь, 1978. - С. 122-137.
162. Тихов М.С. О предельных распределениях оценок по двусторонне цензурированным выборкам // Теория вероятностей и ее применение. 1984. Т. XXIX. № 2. С.354-360.
163. Тихов М.С. Последовательное оценивание параметра сдвига равномерного распределения по цензурированным типа II выборкам // Записки научн. семинаров ЛОМИ, 1984. Т.136. - С. 183-192.
164. Тихов М.С. О предельных законах распределения оценок по цензурированным выборкам // Статистические методы. - Пермь, 1984. - С. 145-154.
165. Тихов М.С. О предельных распределениях отношения правдоподобия и оценок по цензурированным выборкам // Статистические методы. - Пермь, 1986. - С. 95-104.
166. Тихов М.С. О сокращении длительности испытаний при цензурировании выборки // Теория вероятностей и ее применение. 1991. Т. XXXVI. № 3. С. 626-629.
167. Тихов М.С. Аддитивные статистики и устойчивость оценок по цензурированным выборкам // Пробл. компьютер. анал. данных и моделир.: Матер. докл. Респ. науч. конф. "Мат. и прогр. обеспеч. анал. данных". - Минск, 1991. - С. 174-179.
168. Тихов М.С. Оценка показателей безотказности по случайно цензурированной выборке // Надежность и контроль качества, 1993. № 2. С. 11-17.
169. Тихов М.С. Оценивание показателей качества по неполным выборкам // Надежность и контроль качества, 1996. № 11. С. 16-24.
170. Тонева Е. Аппроксимация распределений погрешности средств измерений // Измерительная техника, 1981. № 6. - С. 15-16.
171. Тюрин Ю.Н. О предельном распределении статистик Колмогорова-Смирнова для сложной гипотезы / Известия АН СССР. Сер. Матем. 1984. Т. 48. № 6. - С. 1314-1343.
172. Тюрин Ю.Н., Саввушкина Н.Е. Критерий согласия для распределения Вейбулла-Гнеденко / Известия АН СССР. Сер. Техн. кибернетика. 1984. № 3. - С. 109-112.
173. Тюрин Ю.Н. Исследования по непараметрической статистике (непараметрические методы и линейная модель). Автореф. дисс. докт. физ.-мат. наук. - М.: МГУ, 1985. - 33 с.

- 174.Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере. - М.: ИНФРА-М, Финансы и статистика, 1995. - 384 с.
- 175.Фиакко А., Мак-Кормик Г. Нелинейное программирование: Методы последовательной безусловной минимизации. -М.: Мир, 1972. - 240 с.
- 176.Фролов И.А., Макаров Ю.С. Определение показателей надежности в случае приостановления испытаний части изделий // Надежность и контроль качества, 1973. № 9. С. 3-9.
- 177.Химмельблау Д. Прикладное нелинейное программирование. -М.: Мир, 1975. - 534 с.
- 178.Хьюбер П. Робастность в статистике. - М.: Мир, 1984. - 303 с.
- 179.Цой Е.Б. Вероятностное моделирование по группированным данным при исследовании и проектировании человеко-машинных систем / Диссертация в виде научн. доклада на соискание ученой степени доктора технических наук. Новосибирск, НГТУ, 1997. - 49 с.
- 180.Чибисов Д.М. Некоторые критерии типа хи-квадрат для непрерывных распределений // Теория вероятностей и ее применение. 1971. Т. XVI. № 1. - С. 3-20.
- 181.Численные методы условной оптимизации / Под ред. Ф. Гилла, У. Мюррэя/ - М.: Мир, 1977. - 290 с.
- 182.Шор Я.Б. Статистические методы анализа и контроля качества в надежности. - М.: Сов. радио, 1962. - 552 с.
- 183.Шторм Р. Теория вероятностей. Математическая статистика. Статистический контроль качества. - М.: Мир, 1970. - 368 с.
- 184.Шуленин В.П. Введение в робастную статистику. - Томск: Изд-во Том. ун-та, 1993. - 227 с.
- 185.Яноши Л. Теория и практика обработки результатов измерений. - М.: Мир, 1968. - 462 с.
- 186.Aigner D.J. The estimation of moments for a Pareto distribution subjekt to boath sampling and grouping errors // Review of the International Statistical Institute, 1970. V. 38. № 2. - P. 210-219.
- 187.Aigner D.J., Goldberger A.S. Estimation of Pareto's Law from grouped Observations // J. Am. Statist. Assoc., 1970. V.65. - P.712-723.
- 188.Anderson T.W., Darling D.A. Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes // Ann. Math. Stat., 1952. V.23. - P.193-212.
- 189.Billman B.R., Antle C.E., Bein L.J. Statistical Inference from censored Weibull samples // Technometrics. 1972. V.14. - P. 831-840.
- 190.Birnbaum Z.W. Numerical tabulation of the distribution of Kolmogoroff's statistic for finite sample size // J. Am. Statist. Assoc., 1952. V.47. - P.425-441.
- 191.Bloch D. A note on the estimation of the location parameter of the Cauchy distribution. J. Amer. Statist. Ass. 61 (1966), - p. 852-855.
- 192.Boardman T.J. Estimation in compound Exponential Feilure Models when the Data are Grouped // Technometrics. 1973. V.15. - P. 271-277.
- 193.Boland P.J., Proschan F., Tong Y.L. Crossing properties of mixture distribution // Probab. Eng. and Int. Sci., 1989. V. 3. № 3. - P. 355-366.

194. Boss D.D. Minimum distance estimators for location and goodness of fit // J. Am. Statist. Assoc., 1981. V. 76. - P. 663-670.
195. Bühler W., Deutler T. Optimal stratification and grouping by dynamic programming // Metrika, 1975. B. 22. № 3. - S. 56-69.
196. Carter W.H., Bowen J., Myers R.H. Maximum likelihood estimation from grouped Poisson data // J. Am. Statist. Assoc., 1971. V. 66. - P. 351-353.
197. Chandra M., Singpurwalla N.D., Stephens M.A. Statistics for Test of Fit for the Extreme-Value and Weibull Distribution // J. Am. Statist. Assoc., 1981. V. 76. - P. 375.
198. Chen K.-W., Papadopoulos A.S., Tamer P. On bayes estimation for mixtures of two Weibull distributions under type I censoring // Microelectron. and Rel., 1989. V. 29. № 4. - P. 607-617.
199. Chernoff H., Lehmann E.L. The use of maximum likelihood estimates in χ^2 test for goodness of fit // Ann. Math. Stat., 1954. V. 25. - P. 579-586.
200. Chilingarian A.A., Zazian G.Z. A bootstrap method of distribution mixture proportion determination // Pattern Recogn. Lett., 1990. V. 11. № 12. - P. 781-785.
201. Cochran W.G. The χ^2 Test of Goodness of Fit // Ann. Math. Stat., 1952. V. 23. - P. 315-316.
202. Cochran W.G. Some Methods of Strengthening the Common χ^2 Tests // Biometrics, 1954. V. 10. - P. 417.
203. Cohen A.C. Multi-Censored sampling in the Three Parameter Weibull Distribution // Technometrics. 1975. V. 17. - P. 347-351.
204. Cohen A.C. Progressively Censored Sampling in Three Parameter Log-Normal Distribution // Technometrics. 1976. V. 18. - P. 99-103.
205. Cox D.R. Note on grouping // J. of the American Statistical Association, 1957. - Vol. 52, № 280. - P. 543-547.
206. De Grutolla V., Logakos S.W. Analysis of doubly-censored survival data, with application to AIDS // Biometrics, -1989. V. 45. № 1. - P. 1-11.
207. Denisov V.I., Lemeshko B.Yu. Optimal grouping in Estimation and Tests of Goodnes-of-fit Hypotheses // Wissenschaftliche Schriftenreihe der Technischen universitat Karl-Marx-Stadt. - 1989. № 10. - "Statistics for grouped Observations". - S. 63-81.
208. Denisov V.I., Lemeshko B.Yu., Tsoi E.B. Estimation of unknown parameters of onedimensional distributions with partially grouped data // Wissenschaftliche Schriftenreihe der Technischen universitat Karl-Marx-Stadt. - 1989. № 10. - "Statistics for grouped Observations". - S. 6-21.
209. Durbin J. Kolmogorov-Smirnov tests when parameters are estimated with applications to tests of exponentially and tests of spacings // Biometrika, 1975. V. 62, № 1. - P. 5-22.
210. Durbin J. Kolmogorov-Smirnov tests when parameters are estimated // Lect. Notes Math., 1976. V. 566. - P. 33-44.
211. Dwyer P.S. Grouping metohds // Ann. Math. Stat., 1942. V. 13. № 2. - P. 138-155.

- 212.Ehrenfeld S. Some experimental design problems in attribute life testing // J. Amer. Statist. Ass. 1962. V. 57. - P. 668-679.
- 213.Eisenberger J., Posner E.C. Systematic statistics used for data compression in space telemetry // J. Amer. Statist. Ass. 1965. V. 60. - P. 97-133.
- 214.Fisk P.R. Estimation of location and scale parameters in a truncated grouped sexh square distribution // J. Amer. Statist. Ass. 1961. V. 56. - P. 692-702.
- 215.Fruhworth-Schnatter S. On statistical inference for fuzzy data with applications to discriptive statistics // Fuzzy Sets and Systems, 1992. № 50. - C. 143 - 165.
- 216.Gastaldi Tommaso. A Kolmogorov-Smirnov test procedure involving apossibility censored or truncated sample // Communications in statistics. Theory and methods. 1993, V. 22, №1. P. 31-39.
- 217.Gera A.E. Unimodality and bimodality of mixed Weibull distributions // Qual. and Reliab. Eng. Int., 1994. V. 10. № 4. - P. 355-358.
- 218.Grundy P.M. The fitting grouped truncated and grouped censored normal distributions // Biometrika, 1952. V. 39, № 1/2. - P. 252-259.
- 219.Gupta S.S., Gnanadesikan M. Estimation of the parameters of the logistic distribution // Biometrika, 1966. V.53. - P. 565-570.
- 220.Hammersley J.M., Morton K.W. The Estimation of location and scale parameters from grouped data // Biometrika, 1954. V. 41, № 3/4. - P. 296-301.
- 221.Hampel F.R. Contributions to the theory of robust estimation // Ph. D. Thesis. Berkeley: Univ. California.
- 222.Hampel F.R. The influence curve and its role in robust estimation // J. Amer. Statist. Ass. - V. 69, № 346. - P. 383-393.
- 223.Hansen C.K., Thyregod P. Component lifetime models based on Weibull mixtures and competing risks // Qual. and Reliab. Eng. Int., - 1992. V. 8. № 4. - P. 325-333.
- 224.Hartly H.O. Asimplified form of Sheppards correction formulae // Biometrika, 1950. V.37. - P. 145-148.
- 225.Hassanein K.M. Analysis of extreme-value data by sample quantiles for very large samples // J. Amer. Statist. Ass. 1968. V. 63. - P. 877-888.
- 226.Hassanein K.M., Sebaugh J.L. Estimation of the parameters of the logistic distribution from grouped samples // Skandinavisk Akturitidskrift, 1973. B. 56. № 1/2. - S. 1-10.
- 227.Heinhold I., Gaede K.W. Ingeniur statistic. - München; Wien, Springer Verlag, 1964. - 352 s.
- 228.Hoffman L. Zur Berechnung der Streuung einer Hafigkeitsverteilung mit Klassenleitung // Allgemeine Fortzeitung, 1953. Bd. 64. № 11/12. - S. 141-142.
- 229.Hsieh H.K. Average type-II censoring times for the 2-parameter Weibull distribution // IEEE Trans. Reliab., 1994. V. 43. № 1. - P. 91-96.
- 230.Huber P.J. Robust estimation of a location parameter // Ann. Math. Statist., 1964. V. 35. № 1. - P. 73-101.
- 231.Jiang S., Kececioglu D. Maximum likelihood estimates, from censored data, for mixed-Weibull distributions // IEEE Trans. Reliab., 1992. V. 41. № 2. - P. 248-255.

232. Jones E.L. The use of grouped Measurements // J. Am. Statist. Assoc., 1942. V.36. - P. 525-529.
233. Kac M., Kiefer J., Wolfowitz J. On tests of normality and other tests of goodness of fit based on distance methods // Ann. Math. Stat., 1955. V. 26. - P. 189-211.
234. Kale B.K. A note on the loss information due to grouping of observations // Biometrika, 1964. V. 51, № 3/4. - P. 495-497.
235. Kale B.K. Approximations to the maximum-likelihood estimator using grouped data // Biometrika, 1966. V. 53, № 1/2. - P. 282-285.
236. Kharti G.R. A method for estimating approximately the parameters of a certain class of distributions from grouped observations // Annals of the Institute of Statistical Mathematics, 1962. V. 14. № 2. - P. 57-62.
237. Kendall P.J., Anderson R.L. An Estimation Problem in Life-Testing // Technometrics. 1971. V. 13. - P. 289-301.
238. Kielpinski T.J., Nelson W. Optimum Censored Accelerated Life Tests for Normal and Lognormal Life Distributions // IEEE Trans. Reliab., 1975. V. 24. № 5. - P. 310-320.
239. Kielpinski T.J., Nelson W. Theory for Optimum Censored Accelerated Life Tests for Normal and Lognormal Life Distributions // Technometrics. 1976. V. 18. - P. 105-114.
240. Lilliefors H.W. On the Kolmogorov-Smirnov test for normality with mean and variance unknown // J. Am. Statist. Assoc., 1967. V. 62. - P. 399-402.
241. Lilliefors H.W. On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown // J. Am. Statist. Assoc., 1969. V. 64. - P. 387-389.
242. Lindley D.V. Grouping corrections and maximum likelihood equations // Proceedings of the Cambridge Philosophical Society, 1950. V. 46. № 1. - P. 106-110.
243. Lu J.-C., Unal C. Process characterization and optimization based on censored data from highly fractionated experiments // IEEE Trans. Reliab., 1994. V. 43. № 1. - P. 145-155.
244. Mann H.B., Wald A. On the choice of the number of class intervals in the application of the chi square test // Ann. Math. Stat., 1942. V. 13. - P. 306-317.
245. Mann H.B., Wald A. On the choice of the number of intervals in the application of the chi-square test // Ann. Math. Stat., 1942. V. 13. - P. 478-479.
246. McNeil D.R. Consistent statistics for estimating and testing hypotheses from grouped samples // Biometrika, 1966. V.53, № 3/4. - P. 545-557.
247. Meeker Q.W., Nelson W. Optimum Accelerated Life-Tests for the Weibull and Extreme Value Distributions // IEEE Trans. Reliab., 1975. V. 24. № 5. - P. 321-332.
248. Michako D. Chi-square test-of-fit for location-scale families using type-II censored data // IEEE Trans. Reliab., 1993. V. 42. № 1. - P. 76-80.
249. Mohamed E.A. On the problem of estimation of three parameter log-normal distribution in case of grouped data // AMSE rev., 1989. V. 11. № 2. - P. 53-64.
250. Moore D.S. A chi-square statistic with random cell boundaries // Ann. Math. Statist. 1971. V. 42. - P. 147-156.

251. Mosteller F. On some useful inefficient statistics // *Ann. Math. Statist.* 1946. V. 17. - P. 377-407.
252. Nachas J.A., Kumar A. Reliability estimation using doubly-censored field data // *IEEE Trans. Reliab.*, 1993. V. 42. № 2. - P. 268-279.
253. Ogawa J. Contributions to the theory of systematic statistics // *I. Osaka Math. J.* 1951. V. 3. - P. 175-213.
254. O'Carroll F.M. Fitting a negative binomial distribution to coarsely grouped data by maximum likelihood // *Applied Statistics*, 1962. V. 11. № 1/3. - P. 196-201.
255. Parr W.C., De Wet. On minimum weighted Cramer - von Mises statistical estimation // *Comm. Statist.*, 1981. A10(12). - P. 1149-1166.
256. Pearson E.S., Hartley H.O. *Biometrika tables for Statisticians*. V. 2. - Cambridge: University Press. 1972. - 634 p.
257. Rade L. Grouping and Combining: An operations research problem // *Skandinavisk Aktuarietidskrift*, 1963. B. 46. № 1/2. - S. 56-69.
258. Rao C.R. Criteria of estimation in large samples // *Sankhya*, 1962. - V. 25. - P. 189-206.
259. Saleh A.K.M.J., Ali M.M. Asymptotic optimum quantiles for the estimation of the parameters of the negative exponential distribution // *Ann. Math. Statist.* 1966. V. 37. - P. 143-151.
260. Särndal C.E. Estimation of the parameters of the gamma distribution by sample quantiles // *Technometrics*. 1964. V. 6. - P. 405-414.
261. Stephens M.A. Use of Kolmogorov - Smirnov, Cramer - von Mises and related statistics - without extensive tables // *J. R. Stat. Soc.*, 1970. B. 32. - P. 115-122.
262. Stephens M.A. EDF statistics for goodness of fit and some comparisons // *J. Am. Statist. Assoc.*, 1974. V. 69. - P. 730-737.
263. Stephens M.A. Tests of fit for the logistic distribution, based on the empirical distribution function // *Biometrika*, 1979. V. 68. № 3.
264. Sturges H.A. The choice of classic intervals // *J. Am. Statist. Assoc.* - march 1926. - 47 p.
265. Swamy P.S. On the amount of information supplied by censored samples of grouped observations in the estimation of statistical parameters // *Biometrika*, 1962. V. 49, № 1/2. - P. 245-249.
266. Swamy P.S. On the amount of information supplied by truncated samples of grouped observations in the estimation of the parameters of normal population // *Biometrika*, 1963. V. 50, № 1/2. - P. 207-213.
267. Swan A.W. Computing Maximum-likelihood Estimates for parameters of the Normal Distribution from Grouped and Censored Data // *J. R. Stat. Soc., Ser. C.* 1969. V. 18. № 1. - P. 65-69.
268. Tallis G.M. Approximate maximum likelihood estimates from grouped data // *Technometrics*. 1967. V. 9. - P. 599-606.
269. Tallis G.M., Young S.S.Y. Maximum likelihood estimation of parameters of the normal, log-normal, truncated normal and bivariate normal distributions from grouped data // *The Australian Journal of Statistics*, 1962. V. 4. № 2. - P. 49-54.
270. Tocher K.D. A note on the Analyses of grouped probit data // *Biometrika*, 1949. V. 36, № 1/2. - P. 9-17.

271. Tseng S.-T., Wu H.-J. Selecting, under type-II censoring, Weibull populations that are more reliable // IEEE Trans. Reliab., 1990. V. 39. № 2. - P. 193-198.
272. Tukey J.W. The future of data analysis // Ann. Math. Stat., 1962. V. 33. - P. 1-67.
273. Watson G.S. The χ^2 goodness-of-fit test for normal distributions // Biometrika, 1957. V. 44, № 3/4. - P. 336-348.
274. Watson G.S. On chi-square goodness-of-fit tests for continuous distributions // J. R. Stat. Soc., Ser. B. 1958. B. 20. - P. 44-61.
275. Watson G.S. Some recent results in χ^2 goodness-of-fit tests // Biometrics, 1959. V. 15. - P. 440-468.
276. Weibull C. Maximum Likelihood Estimation from Truncated, Censored and Grouped Samples // Skandinavisk Aktuarietidskrift, 1963. B. 46. № 1/2. - S. 70-77.
277. Wolfowitz J. The minimum distance method // Ann. Math. Stat., 1957. V. 28. - P. 75-88.
278. Woodruff B.W., Viviano P.J., Moore A.H., Dunne E.J. Modified goodness-of-fit tests for gamma-distribution with unknown location and scale parameters // IEEE Trans. Reliab., 1984. R-33.
279. Yoneda K., Ushiyama M. Some estimation in the case of relatively large class intervals // Yokohama mathematical Journal, 1956. V. 4. № 2. - P. 99-118.

**ПРИЛОЖЕНИЕ 1. Таблицы асимптотически оптимального
группирования данных**

Таблица П1.1.

Оптимальные граничные точки интервалов группирования при оценивании параметра θ экспоненциального распределения в виде $t_i = \theta \cdot x_{(i)}$, при оценивании параметра θ распределения Парето в виде $t_i = -\theta \ln(x_{(i)}/\alpha)$, при оценивании масштабного параметра θ_1 распределения Вейбулла в виде $t_i = (x_{(i)}/\theta_1)^\theta$, параметра сдвига μ распределения наименьшего экстремального значения в виде $t_i = \exp\{(x_{(i)} - \mu)/\sigma\}$ и при проверке гипотез о них по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	A
2	1.5936										0.6476
3	1.0176	2.6112									0.8203
4	0.7541	1.7716	3.3652								0.8910
5	0.6004	1.3545	2.3720	3.9657							0.9269
6	0.4993	1.0997	1.8538	2.8714	4.4650						0.9606
7	0.4276	0.9269	1.5273	2.2813	3.2989	4.8925					0.9606
8	0.3739	0.8015	1.3008	1.9012	2.6553	3.6729	5.2665				0.9693
9	0.3323	0.7063	1.1338	1.6331	2.2336	2.9876	4.0052	5.5988			0.9754
10	0.2990	0.6314	1.0053	1.4329	1.9322	2.5326	3.2866	4.3042	5.8979		0.9798
11	0.2716	0.5695	0.9014	1.2746	1.7015	2.1989	2.7955	3.5429	4.5480	6.1176	0.9832

Таблица П1.2.

Оптимальные частоты при оценивании параметров распределений экспоненциального, Парето, Рэлея, масштабного параметра θ_1 распределения Вейбулла, параметра сдвига μ распределения наименьшего экстремального значения и при проверке гипотез о них по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	A
2	0.7968	0.2032										0.6476
3	0.6385	0.2880	0.0735									0.8203
4	0.5296	0.3004	0.1355	0.0345								0.8911
5	0.4514	0.2905	0.1648	0.0744	0.0189							0.9269
6	0.3930	0.2740	0.1763	0.1000	0.0451	0.0116						0.9476
7	0.3479	0.2563	0.1787	0.1150	0.0652	0.0294	0.0075					0.9606
8	0.3120	0.2394	0.1763	0.1229	0.0791	0.0449	0.0202	0.0052				0.9693
9	0.2827	0.2238	0.1717	0.1265	0.0882	0.0567	0.0322	0.0145	0.0037			0.9754
10	0.2584	0.2097	0.1659	0.1273	0.0938	0.0654	0.0421	0.0239	0.0107	0.0028		0.9798
11	0.2378	0.1964	0.1598	0.1264	0.0971	0.0715	0.0498	0.0322	0.0183	0.0083	0.0024	0.9832

Таблица П1.3.

Оптимальные граничные точки интервалов группирования в виде $t_i = x_{(i)} / \theta$ при оценивании параметра θ распределения Рэлея и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	A
2	1.7853									0.6476
3	1.4266	2.2853								0.8203
4	1.2280	1.8823	2.5943							0.8910
5	1.0954	1.6459	2.1781	2.8163						0.9269
6	0.9993	1.4831	1.9255	2.3964	2.9883					0.9476
7	0.9247	1.3615	1.7477	2.1360	2.5686	3.1281				0.9606
8	0.8648	1.2661	1.6130	1.9500	2.3045	2.7103	3.2455			0.9693
9	0.8152	1.1884	1.5058	1.8073	2.1136	2.4444	2.8303	3.3463		0.9754
10	0.7734	1.1237	1.4180	1.6928	1.9658	2.2506	2.5638	2.9340	3.4344	0.9798

Таблица П1.4.

Оптимальные граничные точки интервалов группирования в виде $t_i = x_{(i)} / \theta$ при оценивании параметра θ распределения Максвелла и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	A
2	2.0451									0.6451
3	1.6762	2.5366								0.8179
4	1.4689	2.1292	2.8402							0.8892
5	1.3292	1.8879	2.4221	3.0583						0.9254
6	1.2261	1.7205	2.1667	2.6379	3.2274					0.9464
7	1.1458	1.5947	1.9859	2.3759	2.8081	3.3649				0.9596
8	1.0807	1.4952	1.8481	2.1879	2.5431	2.9480	3.4803			0.9685
9	1.0267	1.4138	1.7377	2.0423	2.3499	2.6803	3.0652	3.5789		0.9747
10	0.9798	1.3447	1.6460	1.9252	2.2003	2.4857	2.7984	3.1668	3.6625	0.9792

Таблица П1.5.

Оптимальные частоты при оценивании параметра распределения Максвелла и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	A
2	0.7576	0.2424									0.6451
3	0.5781	0.3296	0.0923								0.8179
4	0.4597	0.3310	0.1647	0.0446							0.8892
5	0.3778	0.3097	0.1942	0.0933	0.0250						0.9254
6	0.3185	0.2837	0.2022	0.1223	0.0579	0.0154					0.9464
7	0.2739	0.2585	0.2000	0.1374	0.0818	0.0383	0.0101				0.9596
8	0.2393	0.2357	0.1931	0.1438	0.0971	0.0573	0.0267	0.0070			0.9685
9	0.2119	0.2155	0.1840	0.1450	0.1063	0.0711	0.0418	0.0194	0.0050		0.9747
10	0.1891	0.1978	0.1745	0.1437	0.1112	0.0806	0.0536	0.0313	0.0145	0.0038	0.9792

Таблица П1.6.

Оптимальные граничные точки интервалов группирования в виде $t_i = x_{(i)} / \theta$ при оценивании параметра θ полунормального распределения (модуля нормальной случайной величины ($\mathbf{m}=1$)) и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	A
2	1.4821									0.6523
3	1.1395	1.9952								0.8244
4	0.9553	1.5993	2.3124							0.8943
5	0.8355	1.3712	1.9002	2.5407						0.9295
6	0.7495	1.2163	1.6526	2.1225	2.7173					0.9496
7	0.6842	1.1025	1.4797	1.8658	2.3009	2.8632				0.9622
8	0.6277	1.0069	1.3432	1.6751	2.0278	2.4343	2.9742			0.9706
9	0.5892	0.9409	1.2488	1.5445	1.8503	2.1861	2.5844	3.1015		0.9765
10	0.5545	0.8838	1.1682	1.4380	1.7113	2.0012	2.3249	2.6993	3.2125	0.9808

Таблица П1.7.

Оптимальные частоты при оценивании параметра θ полунормального распределения (модуля нормальной случайной величины ($m=1$)) и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	A
2	0.8617	0.1383									0.6523
3	0.7455	0.2085	0.0460								0.8244
4	0.6606	0.2297	0.0890	0.0207							0.8943
5	0.5966	0.2331	0.1129	0.0463	0.0111						0.9295
6	0.5464	0.2297	0.1255	0.0646	0.0272	0.0066					0.9496
7	0.5061	0.2236	0.1313	0.0769	0.0407	0.0172	0.0042				0.9622
8	0.4698	0.2162	0.1348	0.0853	0.0513	0.0277	0.0120	0.0029			0.9706
9	0.4443	0.2090	0.1350	0.0893	0.0582	0.0355	0.0190	0.0078	0.0019		0.9765
10	0.4208	0.2024	0.1341	0.0923	0.0634	0.0417	0.0253	0.0131	0.0056	0.0013	0.9808

Таблица П1.8.

Оптимальные граничные точки интервалов группирования в виде $t_i = x_{(i)} / \theta$ при оценивании параметра θ модуля нормальной случайной величины ($m=4$) и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	A
2	2.2759									0.6435
3	1.9011	2.7621								0.8164
4	1.6877	2.3523	3.0619							0.8880
5	1.5409	2.1027	2.6375	3.2731						0.9245
6	1.4320	1.9320	2.3793	2.8420	3.4371					0.9456
7	1.3441	1.7985	2.1899	2.5797	3.0076	3.5655				0.9589
8	1.2773	1.6988	2.0537	2.3944	2.7499	3.1553	3.6866			0.9679
9	1.2376	1.6331	1.9559	2.2567	2.5594	2.8931	3.2773	3.7885		0.9742
10	1.1957	1.5742	1.8805	2.1615	2.4364	2.7194	3.0318	3.3999	3.8922	0.9788

Таблица П1.9.

Оптимальные частоты при оценивании параметра θ модуля нормальной случайной величины ($m=4$) и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	A
2	0.7306	0.2694									0.6435
3	0.5393	0.3546	0.1061								0.8164
4	0.4165	0.3467	0.1844	0.0524							0.8880
5	0.3327	0.3153	0.2138	0.1082	0.0300						0.9245
6	0.2736	0.2830	0.2175	0.1383	0.0688	0.0188					0.9456
7	0.2287	0.2519	0.2105	0.1536	0.0953	0.0427	0.0128				0.9589
8	0.1969	0.2260	0.1997	0.1574	0.1110	0.0678	0.0325	0.0087			0.9679
9	0.1790	0.2060	0.1849	0.1522	0.1163	0.0827	0.0493	0.0234	0.0062		0.9742
10	0.1610	0.1905	0.1762	0.1498	0.1187	0.0875	0.0599	0.0355	0.0165	0.0044	0.9788

Таблица П1.10.

Оптимальные граничные точки интервалов группирования в виде $t_i = x_{(i)} / \theta$ при оценивании параметра θ модуля нормальной случайной величины ($m=5$) и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	A
2	2.4851									0.6424
3	2.1047	2.9667								0.8154
4	1.8861	2.5542	3.2635							0.8871
5	1.7383	2.3089	2.8453	3.4778						0.9238
6	1.6274	2.1357	2.5880	3.0606	3.6463					0.9451
7	1.5434	2.0084	2.4056	2.7983	3.2306	3.7856				0.9585
8	1.4642	1.8929	2.2499	2.5904	2.9449	3.3472	3.8734			0.9676
9	1.4122	1.8192	2.1563	2.4668	2.7770	3.1092	3.4938	4.0021		0.9739
10	1.3523	1.7371	2.0514	2.3376	2.6193	2.9121	3.2278	3.6023	4.1009	0.9786

Таблица П1.11.

Оптимальные частоты при оценивании параметра θ модуля нормальной случайной величины ($m=5$) и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	A
2	0.7105	0.2895									0.6424
3	0.5106	0.3721	0.1173								0.8154
4	0.3853	0.3562	0.1997	0.0588							0.8871
5	0.3033	0.3199	0.2258	0.1175	0.0335						0.9238
6	0.2460	0.2822	0.2276	0.1489	0.0745	0.0208					0.9451
7	0.2059	0.2495	0.2171	0.1617	0.1021	0.0501	0.0136				0.9585
8	0.1711	0.2181	0.2025	0.1653	0.1202	0.0754	0.0371	0.0103			0.9676
9	0.1501	0.1975	0.1923	0.1620	0.1252	0.0876	0.0532	0.0253	0.0068		0.9739
10	0.1277	0.1750	0.1774	0.1580	0.1307	0.0995	0.0675	0.0406	0.0187	0.0049	0.9786

Таблица П1.12.

Оптимальные граничные точки интервалов группирования в виде $t_i = x_{(i)} / \theta$ при оценивании параметра θ модуля нормальной случайной величины ($m=6$) и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	A
2	2.6788									0.6416
3	2.2937	3.1562								0.8146
4	2.0690	2.7373	3.4474							0.8865
5	1.9209	2.4937	3.0329	3.6644						0.9233
6	1.8077	2.3206	2.7794	3.2560	3.8376					0.9446
7	1.7052	2.1822	2.5885	2.9928	3.4297	3.9772				0.9581
8	1.6448	2.0820	2.4474	2.7948	3.1545	3.5620	4.0856			0.9673
9	1.5655	1.9772	2.3189	2.6446	2.9689	3.3159	3.7061	4.2064		0.9736
10	1.5219	1.9135	2.2322	2.5258	2.8116	3.1058	3.4252	3.7974	4.2833	0.9784

Таблица П1.13.

Оптимальные частоты при оценивании параметра θ модуля нормальной случайной величины ($\mathbf{m}=6$) и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	A
2	0.6951	0.3049									0.6416
3	0.4892	0.3845	0.1263								0.8146
4	0.3612	0.3611	0.2131	0.0646							0.8865
5	0.2815	0.3194	0.2364	0.1260	0.0367						0.9233
6	0.2254	0.2791	0.2365	0.1575	0.0790	0.0225					0.9446
7	0.1797	0.2456	0.2253	0.1734	0.1085	0.0528	0.0147				0.9581
8	0.1552	0.2133	0.2071	0.1721	0.1255	0.0785	0.0378	0.0105			0.9673
9	0.1261	0.1850	0.1925	0.1750	0.1371	0.0958	0.0558	0.0256	0.0071		0.9736
10	0.1115	0.1661	0.1764	0.1640	0.1369	0.1048	0.0721	0.0429	0.0199	0.0054	0.9784

Таблица П1.14.

Оптимальные граничные точки интервалов группирования в виде $t_i = x_{(i)} / \theta$ при оценивании параметра θ модуля нормальной случайной величины ($m=7$) и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	A
2	2.8592									0.6410
3	2.4702	3.3335								0.8141
4	2.2424	2.9130	3.6227							0.8861
5	2.0892	2.6637	3.2017	3.8338						0.9229
6	1.9753	2.4906	2.9522	3.4307	4.0111					0.9444
7	1.8697	2.3537	2.7638	3.1754	3.6180	4.1618				0.9579
8	1.8095	2.2496	2.6140	2.9589	3.3148	3.7168	4.2420			0.9671
9	1.7284	2.1442	2.4877	2.8152	3.1416	3.4897	3.8810	4.3795		0.9734
10	1.6738	2.0692	2.3912	2.6935	2.9861	3.2910	3.6147	3.9779	4.4670	0.9782

Таблица П1.15.

Оптимальные частоты при оценивании параметра θ модуля нормальной случайной величины ($m=7$) и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	A
2	0.6826	0.3174									0.6410
3	0.4720	0.3942	0.1338								0.8141
4	0.3435	0.3648	0.2226	0.0691							0.8861
5	0.2631	0.3179	0.2442	0.1347	0.0401						0.9229
6	0.2090	0.3866	0.1307	0.1653	0.0841	0.0243					0.9444
7	0.1643	0.2413	0.2288	0.1816	0.1140	0.0545	0.0155				0.9579
8	0.1414	0.2059	0.2061	0.1758	0.1316	0.0846	0.0426	0.0120			0.9671
9	0.1138	0.1772	0.1910	0.1787	0.1432	0.1012	0.0597	0.0276	0.0076		0.9734
10	0.0972	0.1559	0.1738	0.1702	0.1441	0.1237	0.0645	0.0438	0.0211	0.0057	0.9782

Таблица П1.16.

Оптимальные граничные точки интервалов группирования в виде $t_i = x_{(i)} / \theta$ при оценивании параметра θ модуля нормальной случайной величины ($m=8$) и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	A
2	3.0289									0.6405
3	2.6345	3.4992								0.8136
4	2.4103	3.0825	3.7907							0.8857
5	2.2556	2.8342	3.3745	4.0004						0.9226
6	2.1318	2.6542	3.1120	3.5661	4.1522					0.9441
7	2.0383	2.5119	2.9189	3.3178	3.7560	4.3031				0.9577
8	1.9612	2.4048	2.7742	3.1251	3.4917	3.8964	4.4313			0.9669
9	1.8989	2.3089	2.6432	2.9523	3.2672	3.6044	3.9900	4.4952		0.9734
10	1.8477	2.2411	2.5569	2.8434	3.1248	3.4169	3.7366	4.1059	4.5926	0.9781

Таблица П1.17.

Оптимальные частоты при оценивании параметра θ модуля нормальной случайной величины ($m=8$) и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	A
2	0.6722	0.3278									0.6405
3	0.4569	0.4024	0.1407								0.8136
4	0.3314	0.3668	0.2292	0.0726							0.8857
5	0.2518	0.3179	0.2496	0.1384	0.0423						0.9226
6	0.1950	0.2732	0.2439	0.1659	0.0944	0.0276					0.9441
7	0.1571	0.2303	0.2282	0.1831	0.1223	0.0614	0.0176				0.9577
8	0.1293	0.1992	0.2078	0.1819	0.1390	0.0871	0.0440	0.0117			0.9669
9	0.1092	0.1691	0.1836	0.1713	0.1460	0.1087	0.0686	0.0339	0.0096		0.9734
10	0.0942	0.1506	0.1680	0.1838	0.1215	0.1157	0.0835	0.0511	0.0248	0.0068	0.9781

Таблица П1.18.

Оптимальные граничные точки интервалов группирования в виде $t_i = x_{(i)} / \theta$ при оценивании параметра θ модуля нормальной случайной величины ($m=9$) и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	A
2	3.1896									0.6401
3	2.7889	3.6556								0.8132
4	2.5681	3.2422	3.9494							0.8854
5	2.4156	2.9987	3.5397	4.1663						0.9224
6	2.2791	2.7928	3.2466	3.7175	4.2994					0.9439
7	2.1972	2.6754	3.0764	3.4674	3.8929	4.4386				0.1576
8	2.1200	2.5657	2.9302	3.2748	3.6274	4.0259	4.5434			0.9668
9	2.0507	2.4729	2.8164	3.1248	3.4271	3.7451	4.1148	4.6058		0.9732
10	2.0025	2.3997	2.7147	2.9993	3.2763	3.5582	3.8621	4.2153	4.6964	0.9780

Таблица П1.19.

Оптимальные частоты при оценивании параметра θ модуля нормальной случайной величины ($m=9$) и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	A
2	0.6634	0.3366									0.6401
3	0.4433	0.4098	0.1469								0.8132
4	0.3208	0.3685	0.2349	0.0758							0.8854
5	0.2437	0.3183	0.2529	0.1417	0.0434						0.9224
6	0.1829	0.2626	0.2459	0.1797	0.0989	0.0300					0.9439
7	0.1509	0.2284	0.2251	0.1835	0.1253	0.0669	0.0199				0.9516
8	0.1240	0.1955	0.2041	0.1813	0.1395	0.0929	0.0484	0.0143			0.9668
9	0.1026	0.1691	0.1873	0.1711	0.1418	0.1067	0.0716	0.0381	0.0117		0.9732
10	0.0892	0.1469	0.1651	0.1610	0.1434	0.1158	0.0853	0.0553	0.0293	0.0087	0.9780

Таблица П1.20.

Оптимальные граничные точки интервалов группирования в виде $t_i = (x_{(i)}/\theta_1)^\theta$ при оценивании основного параметра θ распределения Вейбулла, в виде $t_i = \exp\{(x_{(i)} - \mu)/\sigma\}$ при оценивании параметра σ распределения наименьшего экстремального значения и при проверке гипотез о них по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	A
2	3.4903									0.3282
3	0.1418	3.2891								0.6518
4	0.1505	2.6936	4.5643							0.7481
5	0.0516	0.2486	2.6173	4.4970						0.8235
6	0.0535	0.2580	2.3339	3.6005	5.3984					0.8639
7	0.0244	0.1154	0.3260	2.2878	3.5602	5.3523				0.8936
8	0.0251	0.1181	0.3342	2.1205	3.1036	4.2984	6.0540			0.9141
9	0.0136	0.0639	0.1731	0.3843	2.0935	3.0803	4.2767	6.0333		0.9288
10	0.0137	0.0649	0.1760	0.3917	1.9766	2.7906	3.7069	4.8673	6.6006	0.9408

Таблица П1.21.

Оптимальные частоты при оценивании основного параметра θ распределения Вейбулла, при оценивании параметра σ распределения наименьшего экстремального значения и при проверке гипотез о них по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	A
2	0.9695	0.0305									0.3282
3	0.1322	0.8305	0.0373								0.6518
4	0.1397	0.7927	0.0572	0.0104							0.7481
5	0.0503	0.1698	0.7069	0.0619	0.0111						0.8235
6	0.0521	0.1753	0.6757	0.0696	0.0228	0.0045					0.8639
7	0.0241	0.0949	0.1692	0.6203	0.0731	0.0237	0.0047				0.8936
8	0.0248	0.0866	0.1727	0.5959	0.0751	0.0313	0.0113	0.0023			0.9141
9	0.0134	0.0465	0.0970	0.1602	0.5576	0.0774	0.0320	0.1115	0.0024		0.9288
10	0.0136	0.0492	0.0986	0.1627	0.5374	0.0771	0.0368	0.0169	0.0063	0.0014	0.9408

Оптимальные граничные точки интервалов группирования в виде $t_i = (x_{(i)}/\theta_1)^\theta$ при оценивании двух параметров распределения Вейбулла, в виде $t_i = \exp\{(x_{(i)} - \mu)/\sigma\}$ при оценивании двух параметров распределения наименьшего экстремального значения и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
3	0.2731	2.6067						
4	0.2109	1.3979	3.4137					
5	0.1044	0.5123	1.9590	3.8606				
6	0.0772	0.3649	1.2269	2.5726	4.4096			
7	0.0501	0.2318	0.6758	1.7192	2.9922	4.7959		
8	0.0377	0.1740	0.4837	1.1904	2.2041	3.4285	5.2049	
9	0.0275	0.1269	0.3431	0.7829	1.6027	2.5713	3.7667	5.5273
10	0.0213	0.0988	0.2638	0.5770	1.1805	1.9932	2.9269	4.1024
11	0.0165	0.0771	0.2046	0.4359	0.8560	1.5344	2.3192	3.2319
12	0.0123	0.0618	0.1638	0.3434	0.6517	1.1789	1.8570	2.6163
13	0.0106	0.0500	0.1326	0.2754	0.5106	0.9030	1.4807	2.1401
14	0.0087	0.0412	0.1094	0.2261	0.3126	0.7116	1.1798	1.7608
15	0.0072	0.0344	0.0913	0.1881	0.3394	0.5734	0.9387	1.4426

t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}	A
						0.4079
						0.5572
						0.6836
						0.7571
						0.8109
						0.8480
						0.8756
5.8478						0.8963
4.3930	6.1270					0.9123
3.5103	4.6589	6.3853				0.9248
2.8810	3.7623	4.9016	6.6208			0.9349
2.4019	3.1286	3.9997	5.1314	6.8444		0.9431
2.0116	2.6381	3.3538	4.2169	5.3425	7.0506	0.9498

Оптимальные частоты при оценивании двух параметров распределения Вейбулла, при оценивании двух параметров распределения наименьшего экстремального значения и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
3	0.2390	0.6872	0.0738					
4	0.1901	0.5628	0.2142	0.0329				
5	0.0991	0.3018	0.4581	0.1199	0.0211			
6	0.0743	0.2314	0.4011	0.2169	0.0641	0.0122		
7	0.0489	0.1581	0.2843	0.3295	0.1290	0.0419	0.0083	
8	0.0370	0.1227	0.2238	0.3124	0.1938	0.0779	0.0269	0.0055
9	0.0271	0.0921	0.1712	0.2525	0.2557	0.1250	0.0533	0.0191
10	0.0211	0.0729	0.1379	0.2065	0.2545	0.1708	0.0827	0.0371
11	0.0164	0.0578	0.1108	0.1683	0.2218	0.2101	0.1164	0.0589
12	0.0131	0.0468	0.0912	0.1395	0.1882	0.2136	0.1515	0.0830
13	0.0105	0.0383	0.0754	0.1165	0.1592	0.1947	0.1779	0.1099
14	0.0087	0.0317	0.0632	0.0988	0.1357	0.1710	0.1836	0.1354
15	0.0072	0.0266	0.0535	0.0842	0.1163	0.1486	0.1725	0.1548

P_9	P_{10}	P_{11}	P_{12}	P_{13}	P_{14}	P_{15}	A
							0.4079
							0.5572
							0.6836
							0.7572
							0.8109
							0.8480
0.0040							0.8756
0.0136	0.0029						0.8963
0.0271	0.0102	0.0022					0.9123
0.0432	0.0204	0.0078	0.0017				0.9248
0.0615	0.0329	0.0158	0.0061	0.0013			0.9349
0.0814	0.0467	0.0255	0.0124	0.0048	0.0011		0.9431
0.1025	0.0623	0.0365	0.0203	0.0099	0.0039	0.0009	0.9498

Таблица П1.24.

Оптимальные граничные точки интервалов группирования в виде $t_i = (x_{(i)} - \mu) / \sigma$ при оценивании математического ожидания μ нормального распределения, при оценивании параметров μ логарифмически нормальных распределений в виде $t_i = (\ln x_{(i)} - \mu) / \sigma$ и $t_i = (\lg x_{(i)} - \mu) / \sigma$ и при проверке гипотез о них по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	A
2	0.0									0.6336
3	-0.6120	0.6120								0.8098
4	-0.9816	0.0	0.9816							0.8825
5	-1.2444	-0.3823	0.3823	1.2444						0.9201
6	-1.4468	-0.6589	0.0	0.6589	1.4468					0.9420
7	-1.6108	-0.8744	-0.2803	0.2803	0.8744	1.6108				0.9560
8	-1.7479	-1.0499	-0.5005	0.0	0.5005	1.0499	1.7479			0.9655
9	-1.8655	-1.1976	-0.6812	-0.2218	0.2218	0.6812	1.1976	1.8655		0.9721
10	-1.9682	-1.3246	-0.4047	-0.8338	0.0	0.4047	0.8338	1.3246	1.9682	0.9771

Таблица П1.25.

Оптимальные частоты при оценивании математического ожидания μ нормального распределения и параметров μ логарифмически нормальных распределений и при проверке гипотез о них по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	A
2	0.5000	0.5000									0.6336
3	0.2703	0.4594	0.2703								0.8098
4	0.1631	0.3369	0.3369	0.1631							0.8825
5	0.1067	0.2444	0.2978	0.2444	0.1067						0.9201
6	0.0740	0.1810	0.2450	0.2450	0.1810	0.0740					0.9320
7	0.0536	0.1373	0.1987	0.2208	0.1987	0.1373	0.0536				0.9560
8	0.0403	0.1066	0.1615	0.1916	0.1916	0.1615	0.1066	0.0403			0.9655
9	0.0310	0.0845	0.1323	0.1644	0.1756	0.1644	0.1323	0.0845	0.0310		0.9721
10	0.0245	0.0662	0.1095	0.1406	0.1572	0.1572	0.1406	0.1095	0.0662	0.0245	0.9771

Таблица П1.26.

Оптимальные граничные точки интервалов группирования в виде $t_i = (x_{(i)} - \mu)/\sigma$ при оценивании стандартного отклонения σ нормального распределения, при оценивании параметров σ логарифмически нормальных распределений в виде $t_i = (\ln x_{(i)} - \mu)/\sigma$ и $t_i = (\lg x_{(i)} - \mu)/\sigma$ и при проверке гипотез о них по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	A
2	-1.5750									
	1.5750									0.3042
3	-1.4821	1.4821								0.6522
4	-2.0249	-1.1865	1.4520							
	-1.4520	1.1855	2.0249							0.7358
5	-1.9956	-1.1401	1.1401	1.9958						0.8244
6	-2.3269	-1.6190	-0.9837	1.1190	1.9821					0.8588
	-1.9821	-1.1190	0.9837	1.6190	2.3269					
7	-2.3130	-1.6002	-0.9558	0.9558	1.6002	2.3130				0.8943
8	-2.5488	-1.9105	-1.3848	-0.8548	0.9400	1.5897	2.3053			
	-2.3053	-1.5897	0.9400	0.8548	1.9105	1.9105	2.5488			0.9117
9	-2.5408	-1.9003	-1.3715	-0.8355	0.8355	1.3715	1.9003	2.5408		0.9294
10	-2.7223	-2.1286	-1.6602	-1.2260	-0.7634	0.8233	1.3632	1.8939	2.5358	
	-2.5358	-1.8939	-1.3632	-0.8233	0.7634	1.2260	1.6602	2.1286	2.2723	0.9394

Таблица П1.27.

Оптимальные частоты при оценивании стандартного отклонения σ нормального распределения, при оценивании параметров σ логарифмически нормальных распределений и при проверке гипотез о них по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	A
2	0.0576	0.9424									
	0.9424	0.0576									0.3042
3	0.0692	0.8616	0.0692								0.6522
4	0.0214	0.0965	0.8089	0.0732							
	0.0732	0.8089	0.0965	0.0214							0.7358
5	0.0230	0.1041	0.7458	0.1041	0.0230						0.8244
6	0.0100	0.0427	0.1099	0.7058	0.1079	0.0237					
	0.0237	0.1079	0.7058	0.1099	0.0427	0.0100					0.8588
7	0.0104	0.0444	0.1148	0.6608	0.1148	0.0444	0.0104				0.8943
8	0.0054	0.0226	0.0550	0.1133	0.6300	0.1177	0.0454	0.0106			
	0.0106	0.0454	0.1177	0.6300	0.1133	0.0550	0.0226	0.0054			0.9117
9	0.0055	0.0232	0.0564	0.1166	0.5966	0.1166	0.0564	0.0232	0.0055		0.9294
10	0.0032	0.0134	0.0318	0.0617	0.1125	0.5722	0.1188	0.0573	0.0235	0.0056	
	0.0056	0.0235	0.0573	0.1188	0.5722	0.1126	0.0617	0.0318	0.0314	0.0032	0.9394

Оптимальные граничные точки интервалов в виде $t_i = (x_{(i)} - \mu)/\sigma$ при одновременном оценивании двух параметров нормального распределения, в виде $t_i = (\ln x_{(i)} - \mu)/\sigma$ и $t_i = (\lg x_{(i)} - \mu)/\sigma$ при оценивании параметров логарифмически нормальных распределений и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации

A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7
3	-1.1106	1.1106					
4	-1.3834	0.0	1.3834				
5	-1.6961	-0.6894	0.6894	1.6961			
6	-1.8817	-0.9970	0.0	0.9970	1.8817		
7	-2.0600	-1.2647	-0.4918	0.4918	1.2647	2.0600	
8	-2.1954	-1.4552	-0.7863	0.0	0.7863	1.4552	2.1954
9	-2.3188	-1.6218	-1.0223	-0.3828	0.3828	1.0223	1.6218
10	-2.4225	-1.7578	-1.2046	-0.6497	0.0	0.6497	1.2046
11	-2.5167	-1.8784	-1.3602	-0.8621	-0.3143	0.3143	0.8621
12	-2.5993	-1.9028	-1.4914	-1.0331	-0.5334	0.0	0.5334
13	-2.6746	-2.0762	-1.6068	-1.1784	-0.7465	-0.2669	0.2669
14	-2.7436	-2.1609	-1.7092	-1.3042	-0.9065	-0.4818	0.0
15	-2.8069	-2.2378	-1.8011	-1.4150	-1.0435	-0.6590	-0.2325

t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}	A
							0.4065
							0.5527
							0.6826
							0.7557
							0.8103
							0.8474
2.3188							0.8753
1.7578	2.4225						0.8960
1.3602	1.8784	2.5167					0.9121
1.0331	1.4914	1.9028	2.5993				0.9247
0.7465	1.1784	1.6068	2.0762	2.6746			0.9348
0.4818	0.9065	1.3042	1.7092	2.1609	2.7436		0.9430
0.2325	0.6590	1.0435	1.4150	1.8011	2.2378	2.8069	0.9498

Оптимальные частоты при одновременном оценивании двух параметров нормального распределения, параметров логарифмически нормальных распределений и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
3	0.1334	0.7332	0.1334					
4	0.0833	0.4167	0.4167	0.0833				
5	0.0449	0.2004	0.5094	0.2004	0.0449			
6	0.0299	0.1295	0.3406	0.3406	0.1295	0.0299		
7	0.0197	0.0833	0.2084	0.3772	0.2084	0.0833	0.0197	
8	0.0141	0.0587	0.1431	0.2841	0.2841	0.1431	0.0587	0.0141
9	0.0102	0.0422	0.1009	0.1976	0.2982	0.1976	0.1009	0.0422
10	0.0077	0.0317	0.0748	0.1438	0.2420	0.2420	0.1438	0.0748
11	0.0059	0.0243	0.0567	0.1074	0.1823	0.2468	0.1823	0.1074
12	0.0047	0.0190	0.0442	0.0829	0.1392	0.2100	0.2100	0.1392
13	0.0037	0.0152	0.0352	0.0652	0.1085	0.1670	0.2104	0.1670
14	0.0030	0.0124	0.0283	0.0524	0.0862	0.1327	0.1850	0.1850
15	0.0025	0.0101	0.0232	0.0427	0.0698	0.1066	0.1532	0.1838

P_9	P_{10}	P_{11}	P_{12}	P_{13}	P_{14}	P_{15}	A
							0.4065
							0.5527
							0.6826
							0.7557
							0.8103
							0.8474
0.0102							0.8753
0.0317	0.0077						0.8960
0.0567	0.0243	0.0059					0.9121
0.0829	0.0442	0.0190	0.0047				0.9247
0.1085	0.0652	0.0352	0.0152	0.0037			0.9348
0.1327	0.0862	0.0524	0.0283	0.0124	0.0030		0.9430
0.1532	0.1066	0.0698	0.0427	0.0232	0.0101	0.0025	0.9498

Таблица П1.30.

Оптимальные граничные точки интервалов в виде $t_i = (x_{(i)} - \mu) / \sigma$ при оценивании параметра μ распределения наибольшего экстремального значения и при проверке гипотез о нем по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	A
2	-0.4660										0.6476
3	-0.9598	-0.0174									0.8203
4	-1.2135	-0.5719	0.2822								0.8910
5	-1.3777	-0.8637	-0.3034	0.5102							0.9269
6	-1.4963	-1.0548	-0.6172	-0.0950	0.6945						0.9476
7	-1.5877	-1.1936	-0.8247	-0.4235	0.0759	0.8496					0.9606
8	-1.6614	-1.3010	-0.9766	-0.6425	-0.2630	0.2213	0.9838				0.9693
9	-1.7226	-1.3876	-1.0945	-0.8036	-0.1256	0.3477	1.0999				0.9754
10	-1.7746	-1.4596	-1.1899	-0.9292	-0.6587	-0.3597	-0.0053	0.4598	1.2073		0.9798
11	-1.8112	-1.5147	-1.2649	-1.0280	-0.7880	-0.5315	-0.2426	0.1038	0.5630	1.3034	0.9832

Таблица П1.31.

Оптимальные частоты при оценивании параметра μ распределения наибольшего экстремального значения и при проверке гипотез о нем по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	A
2	0.2032	0.7968										0.6476
3	0.0735	0.2880	0.6385									0.8203
4	0.0345	0.1355	0.3004	0.5296								0.8911
5	0.0189	0.0744	0.1648	0.2905	0.4514							0.9269
6	0.0116	0.0451	0.1000	0.1763	0.2740	0.3930						0.9476
7	0.0075	0.0294	0.0652	0.1150	0.1787	0.2563	0.3479					0.9606
8	0.0052	0.0202	0.0449	0.0791	0.1229	0.1763	0.2394	0.3120				0.9693
9	0.0037	0.0145	0.0322	0.0567	0.0882	0.1265	0.1717	0.2238	0.2827			0.9754
10	0.0028	0.0107	0.0239	0.0421	0.0654	0.0938	0.1273	0.1659	0.2097	0.2584		0.9798
11	0.0024	0.0083	0.0183	0.0322	0.0498	0.0715	0.0971	0.1264	0.1598	0.1964	0.2378	0.9832

Таблица П1.32.

Оптимальные граничные точки интервалов в виде $t_i = (x_{(i)} - \mu) / \sigma$ при оценивании масштабного параметра σ распределения наибольшего экстремального значения и при проверке гипотез о нем по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	A
2	-1.2500										0.3282
3	-1.1906	1.9535									0.6518
4	-1.5183	-0.9909	1.8940								0.7481
5	-1.5034	-0.9622	1.3919	2.9638							0.8235
6	-1.6852	-1.2811	-0.8475	1.3546	2.9291						0.8639
7	-1.6788	-1.2709	-0.8295	1.1218	2.1591	3.7091					0.8936
8	-1.8007	-1.4582	-1.1326	-0.7516	1.0959	2.1361	3.6862				0.9141
9	-1.7973	-1.4532	-1.1250	-0.7389	0.9564	1.7540	2.7510	4.3053			0.9288
10	-1.8872	-1.5826	-1.3102	-1.0262	-0.6814	0.9372	1.7374	2.7350	4.2891		0.9408
11	-1.8851	-1.5796	-1.3061	-1.0204	-0.6717	0.8425	1.5010	2.2522	3.2414	4.8031	0.9490

Таблица П1.33.

Оптимальные частоты при оценивании параметра σ распределения наибольшего экстремального значения и при проверке гипотез о нем по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	A
2	0.0305	0.9695										0.3282
3	0.0373	0.8305	0.1322									0.6518
4	0.0104	0.0572	0.7927	0.1397								0.7481
5	0.0111	0.0619	0.7069	0.1698	0.0503							0.8235
6	0.0045	0.0228	0.0696	0.6756	0.1754	0.0521						0.8639
7	0.0047	0.0236	0.0727	0.6210	0.1690	0.0848	0.0242					0.8936
8	0.0023	0.0112	0.0313	0.0751	0.5959	0.1727	0.0867	0.0248				0.9141
9	0.0024	0.0115	0.0321	0.0773	0.5577	0.1601	0.0971	0.0484	0.0134			0.9288
10	0.0014	0.0063	0.0169	0.0369	0.0771	0.5374	0.1628	0.0985	0.0492	0.0136		0.9408
11	0.0014	0.0064	0.0171	0.0374	0.0788	0.5089	0.1501	0.1000	0.0615	0.0302	0.0082	0.9490

Оптимальные граничные точки интервалов в виде $t_i = (x_{(i)} - \mu) / \sigma$ при одновременном оценивании параметров μ и σ распределения наибольшего экстремального значения и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации

A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7
3	-0.9581	1.2978					
4	-1.2278	-0.335	1.5562				
5	-1.3508	-0.6725	0.6688	2.2591			
6	-1.4838	-0.9449	-0.2045	1.0080	2.5608		
7	-1.5676	-1.096	-0.5419	0.3919	1.4614	2.9934	
8	-1.6496	-1.2321	-0.7903	-0.1743	0.7262	1.7490	3.2777
9	-1.7097	-1.3265	-0.9444	-0.4717	0.2448	1.0698	2.0644
10	-1.7660	-1.4115	-1.0740	-0.6897	-0.1659	0.5499	1.3326
11	-1.8119	-1.4787	-1.1715	-0.8397	-0.4262	0.1576	0.5877
12	-1.8540	-1.5388	-1.2557	-0.9617	-0.619	-0.1648	0.4282
13	-1.8902	-1.5896	-1.3250	-1.0581	-0.7608	-0.3925	0.1020
14	-1.9234	-1.6354	-1.3862	-1.1406	-0.8763	-0.5658	-0.1654
15	-1.9531	-1.6757	-1.4391	-1.2101	-0.9701	-0.6990	-0.3665

t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}	A
							0.4079
							0.5572
							0.6836
							0.7571
							0.8109
							0.8480
3.5952							0.8756
2.3149	3.8496						0.8963
1.5877	2.5639	4.1040					0.9123
1.0689	1.8092	2.7833	4.3280				0.9248
0.6723	1.2894	2.0205	2.9948	4.5454			0.9349
0.3401	0.8852	1.4868	2.2126	3.1883	4.7419		0.9431
0.0632	0.5564	1.0811	1.6723	2.3949	3.3713	4.9279	0.9498

Оптимальные частоты при одновременном оценивании параметров μ и σ распределения наибольшего экстремального значения, двух параметров двойного показательного распределения и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации

A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
3	0.0738	0.6872	0.2390					
4	0.0329	0.2142	0.5628	0.1901				
5	0.0211	0.1199	0.4581	0.3018	0.0991			
6	0.0122	0.0641	0.2169	0.4011	0.2314	0.0743		
7	0.0083	0.0419	0.1290	0.3295	0.2843	0.1581	0.0489	
8	0.0055	0.0269	0.0779	0.1938	0.3124	0.2238	0.1227	0.0370
9	0.0040	0.0191	0.0533	0.1250	0.2557	0.2525	0.1712	0.0921
10	0.0029	0.0136	0.0371	0.0827	0.1708	0.2545	0.2065	0.1379
11	0.0022	0.0102	0.0271	0.0589	0.1164	0.2101	0.2218	0.1683
12	0.0017	0.0078	0.0204	0.0432	0.0830	0.1515	0.2136	0.1882
13	0.0013	0.0061	0.0158	0.0329	0.0615	0.1099	0.1779	0.1947
14	0.0011	0.0048	0.0124	0.0255	0.0467	0.0814	0.1354	0.1836
15	0.0009	0.0039	0.0099	0.0203	0.0365	0.0623	0.1025	0.1548

P_9	P_{10}	P_{11}	P_{12}	P_{13}	P_{14}	P_{15}	A
							0.4079
							0.5572
							0.6836
							0.7571
							0.8109
							0.8480
0.0271							0.8756
0.0729	0.0211						0.8963
0.1108	0.0578	0.0164					0.9123
0.1395	0.0912	0.0463	0.0131				0.9248
0.1592	0.1165	0.0754	0.0383	0.0105			0.9349
0.1710	0.1357	0.0988	0.0632	0.0317	0.0087		0.9431
0.1725	0.1486	0.1163	0.0842	0.0535	0.0266	0.0072	0.9498

Таблица П1.36.

Оптимальные граничные точки интервалов в виде $t_i = (x_{(i)} - \theta_1)/\theta$ при оценивании масштабного параметра θ распределения Коши и при проверке гипотез о нем по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	A
2	-1.3274									0.2867
	1.3274									0.2867
3	-1.0000	1.0000								0.8106
4	-0.9597	0.7149	1.5797							0.8630
	-1.5797	-0.7149	0.9597							0.8630
5	-1.5171	-0.6591	0.6591	1.5171						0.9226
6	-1.4959	-0.6398	0.5438	1.0028	1.9562					0.9339
	-1.9562	-1.0228	-0.5438	0.6398	1.4959					0.9339
7	-1.9269	-1.0000	-0.5190	0.5190	1.0000	1.9269				0.9582
8	-1.9130	-0.9891	-0.5070	0.4529	0.8015	1.2806	2.2993			0.9715
	-2.2993	-1.2806	-0.8015	-0.4529	0.5070	0.9801	1.9130			0.9715
9	-2.2814	-1.2674	-0.7891	-0.4384	0.4384	0.7891	1.2674	2.2814		0.9739
10	-2.2712	-1.2600	-0.7819	-0.4300	0.3944	0.6759	1.0078	1.5060	2.6124	0.9780
	-2.6124	-1.5060	-1.0078	-0.6579	-0.3944	0.4300	0.7819	1.2600	2.2712	0.9780
11	-2.6002	-1.4972	-1.0000	-0.6679	-0.3846	0.6679	1.0000	1.4972	2.6002	0.9821

Таблица П1.37.

Оптимальные частоты при оценивании масштабного параметра θ распределения Коши и при проверке гипотез о нем по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	A
2	0.2055	0.7945										0.2867
3	0.7945	0.2055										0.2867
	0.2500	0.5000	0.2500									0.8106
4	0.2565	0.4411	0.1228	0.1796								0.8630
	0.1796	0.1228	0.4411	0.2565								0.8630
5	0.1855	0.1290	0.3710	0.1290	0.1855							0.9226
6	0.1876	0.1312	0.3397	0.0951	0.0960	0.1504						0.9339
	0.1504	0.0960	0.0951	0.3397	0.1312	0.1876						0.9339
7	0.1524	0.0976	0.0976	0.3048	0.0976	0.0976	0.1524					0.9583
8	0.1533	0.0984	0.0989	0.2848	0.0797	0.0739	0.0804	0.1306				0.9715
	0.1306	0.0804	0.0739	0.7978	0.2848	0.0989	0.0984	0.1533				0.9715
9	0.1315	0.0811	0.0748	0.0811	0.2630	0.0811	0.0748	0.0811	0.1315			0.9739
10	0.1320	0.0815	0.0753	0.0819	0.2489	0.0696	0.0620	0.0622	0.0702	0.1164		0.9780
	0.1164	0.0702	0.0622	0.0620	0.0696	0.2489	0.0819	0.0753	0.0815	0.1320		0.9780
11	0.1169	0.0705	0.0626	0.0626	0.0705	0.2338	0.0705	0.0626	0.0626	0.0705	0.1169	0.9821

Таблица П1.38.

Оптимальные граничные точки интервалов в виде $t_i = (x_{(i)} - \theta_1)/\theta$ при оценивании параметра сдвига θ_1 распределения Коши и при проверке гипотез о нем по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	A
2	0.0											0.8106
3	-0.1794	0.1794										0.8577
4	-3.2845	-0.2390	0.1570									0.8991
	-0.1570	0.2390	3.2845									0.8991
5	-3.3338	-0.2137	0.2137	3.3338								0.9393
6	-3.1577	-0.3168	0.0	0.3168	0.1577							0.9582
7	-3.0827	-0.3786	-0.1150	0.1150	0.3786	3.0827						0.9658
8	-5.0956	-2.2911	-0.4018	-0.1275	0.1072	0.3744	3.0863					0.9718
	-3.0863	0.3744	-0.1072	0.1275	0.4018	2.2911	5.0965					0.9718
9	-5.1041	-2.2961	-0.3972	-0.1196	0.1196	0.3972	2.2961	5.1041				0.9779
10	-5.0213	-2.2496	-0.4445	-0.1991	0.0	0.1991	0.4445	0.4445	2.2496	5.0213		0.9821
11	-4.9723	-2.2216	-0.4792	-0.2572	-0.0820	0.0820	0.2572	0.4792	2.2216	4.9723		0.9845
12	-4.9357	-2.2027	-0.5059	-0.3020	-0.1433	0.0	0.1433	0.3020	0.5059	2.2027	4.9357	0.9860

Таблица П1.38.

Оптимальные граничные точки интервалов в виде $t_i = (x_{(i)} - \theta_1)/\theta$ при оценивании параметра сдвига θ_1 распределения Коши и при проверке гипотез о нем по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	A
2	0.0											0.8106
3	-0.1794	0.1794										0.8577
4	-3.2845	-0.2390	0.1570									0.8991
	-0.1570	0.2390	3.2845									0.8991
5	-3.3338	-0.2137	0.2137	3.3338								0.9393
6	-3.1577	-0.3168	0.0	0.3168	0.1577							0.9582
7	-3.0827	-0.3786	-0.1150	0.1150	0.3786	3.0827						0.9658
8	-5.0956	-2.2911	-0.4018	-0.1275	0.1072	0.3744	3.0863					0.9718
	-3.0863	0.3744	-0.1072	0.1275	0.4018	2.2911	5.0965					0.9718
9	-5.1041	-2.2961	-0.3972	-0.1196	0.1196	0.3972	2.2961	5.1041				0.9779
10	-5.0213	-2.2496	-0.4445	-0.1991	0.0	0.1991	0.4445	0.4445	2.2496	5.0213		0.9821
11	-4.9723	-2.2216	-0.4792	-0.2572	-0.0820	0.0820	0.2572	0.4792	2.2216	4.9723		0.9845
12	-4.9357	-2.2027	-0.5059	-0.3020	-0.1433	0.0	0.1433	0.3020	0.5059	2.2027	4.9357	0.9860

Таблица П1.39.

Оптимальные частоты при оценивании параметра сдвига θ_1 распределения Коши и при проверке гипотез о нем по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	A
2	0.5	0.5											0.8106
3	0.4435	0.1130	0.4435										0.8577
4	0.0941	0.3312	0.1243	0.4504									0.8991
	0.4504	0.1243	0.3312	0.0941									0.8991
5	0.0928	0.3402	0.1340	0.3402	0.0928								0.9393
6	0.0976	0.3047	0.0977	0.0977	0.3047	0.0976							0.9582
7	0.0998	0.2850	0.0788	0.0728	0.0788	0.2850	0.0998						0.9658
8	0.0617	0.0693	0.2474	0.0812	0.0744	0.0800	0.2863	0.0997					0.9718
	0.0997	0.2863	0.0800	0.0744	0.0812	0.2474	0.0693	0.0617					0.9718
9	0.0616	0.0691	0.2490	0.0824	0.0758	0.0824	0.2490	0.0691	0.0616				0.9779
10	0.0626	0.0705	0.2338	0.0705	0.0626	0.0626	0.0705	0.2338	0.0705	0.0626			0.9821
11	0.0632	0.0714	0.2232	0.0621	0.0541	0.0520	0.0541	0.0621	0.2232	0.0714	0.0632		0.9845
12	0.0636	0.0721	0.2152	0.0557	0.0481	0.0453	0.0453	0.0481	0.0557	0.2152	0.0721	0.0636	0.9860

Таблица П1.40.

Оптимальные граничные точки интервалов в виде $t_i = (x_{(i)} - \theta_1)/\theta$ при одновременном оценивании параметров θ и θ_1 распределения Коши и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
3	-0.5773	0.5773						
4	-1.0000	0.0	1.0000					
5	-1.3764	-0.3249	0.3249	1.3764				
6	-1.7320	-0.5773	0.0	0.5773	1.7320			
7	-2.0765	-0.7975	-0.2282	0.2282	0.7975	2.0765		
8	-2.4142	-1.0000	-0.4142	0.0	0.4142	1.0000	2.4142	
9	-2.7475	-1.1917	-0.5773	-0.1763	0.1763	0.5773	1.1917	2.7475
10	-3.0778	-1.3764	-0.7266	-0.3249	0.0	0.3249	0.7266	1.3764
11	-3.4062	-1.5561	-0.8665	-0.4567	-0.1438	0.1438	0.4567	0.8665
12	-3.7319	-1.7320	-1.0000	-0.5773	-0.2679	0.0	0.2679	0.5773
13	-4.0573	-1.9053	-1.1288	-0.6903	-0.3793	-0.1214	0.1214	0.3793
14	-4.3809	-2.0763	-1.2538	-0.7974	-0.4815	-0.2282	0.0	0.2282
15	-4.7041	-2.2457	-1.3762	-0.9030	-0.5773	-0.3249	-0.1051	0.1051

t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}	A
						0.4677
						0.6570
						0.7659
						0.8315
						0.8735
						0.9018
						0.9217
3.0778						0.9361
1.5561	3.4062					0.9469
1.0000	1.7320	3.7319				0.9552
0.6903	1.1288	1.9053	4.0573			0.9617
0.4815	0.7974	1.2538	2.0763	4.3809		0.9669
0.3249	0.5773	0.9003	1.3762	2.2457	4.7041	0.9711

Таблица П1.41.

Оптимальные граничные точки интервалов в виде $t_i = \pi(x_{(i)} - \mu) / \sigma\sqrt{3}$ при оценивании параметра μ логистического распределения и при проверке гипотез о нем по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	A
2	0.0										0.7500
3	-0.6931	0.6931									0.8889
4	-1.0986	0.0	1.0986								0.9375
5	-1.3863	-0.4055	0.4055	1.3863							0.9600
6	-1.6094	-0.6931	0.0	0.6931	1.6094						0.9722
7	-1.7918	-0.9163	-0.2877	0.2877	0.9163	1.7918					0.9796
8	-1.9459	-1.0986	-0.5108	0.0	0.5108	1.0986	1.9459				0.9844
9	-2.0794	-1.2528	-0.6931	-0.2231	0.2231	0.6931	1.2528	2.0794			0.9876
10	-2.1972	-1.3863	-0.8473	-0.4055	0.0	0.4055	0.8473	1.3863	2.1972		0.9900
11	-2.3026	-1.5041	-0.9808	-0.5596	-0.1823	0.1823	0.5596	0.9808	1.5041	2.3026	0.9921

Таблица П1.42.

Оптимальные граничные точки интервалов в виде $t_i = \pi(x_{(i)} - \mu) / \sigma\sqrt{3}$ при оценивании параметра σ логистического распределения и при проверке гипотез о нем по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	A
2	-2.3993										0.3074
	2.3993										0.3074
3	-2.1651	2.1651									0.6804
4	-3.2424	-1.6762	2.1049								0.7618
	-2.1049	1.6267	3.2424								0.7618
5	-3.1685	-1.5929	1.5929	3.1385							0.8449
6	-3.9427	-2.4141	-1.3564	1.5566	3.1368						0.8757
	3.1638	-1.5566	1.3564	2.4141	3.9427						0.8757
7	-3.9053	-2.3760	-1.3107	1.3107	2.3760	3.9053					0.9078
8	-4.5191	-2.9874	-1.9953	-1.1639	1.2856	2.3552	3.8845				0.9233
	-3.8845	-2.3552	-1.2866	1.1639	1.9953	2.9874	4.5191				0.9233
9	-4.4961	-2.9648	-1.9713	-1.1341	1.1341	1.9713	2.9648	4.4961			0.9391
10	-4.4819	-2.9508	-1.9565	-1.1156	1.0318	1.7314	2.4887	3.4650	5.0067		0.9478
	-5.0067	-3.4650	-2.4887	-1.7314	-1.0318	1.1156	1.9565	2.9508	4.4819		0.9478
11	-4.9907	-3.4496	-2.4733	-1.7145	-1.0104	1.0104	1.7145	2.4733	3.4496	4.9907	0.9568

Таблица П1.43.

Оптимальные частоты при оценивании параметра σ логистического распределения и при проверке гипотез о нем по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	A
2	0.0832	0.9168										0.3074
	0.9168	0.0832										0.3074
3	0.1029	0.7942	0.1029									0.6844
4	0.0376	0.1200	0.7338	0.1086								0.7618
	0.1086	0.7338	0.1200	0.0376								0.7618
5	0.0404	0.1286	0.6620	0.1286	0.0404							0.8449
6	0.0190	0.0631	0.1227	0.6211	0.1325	0.0416						0.8757
	0.0416	0.1325	0.6211	0.1227	0.0631	0.0190						0.8757
7	0.0197	0.0653	0.1274	0.5752	0.1274	0.0653	0.0197					0.9078
8	0.0108	0.0372	0.0717	0.1183	0.5454	0.1299	0.0666	0.0201				0.9233
	0.0201	0.0666	0.1299	0.5454	0.1183	0.0717	0.0372	0.0108				0.9233
9	0.0110	0.0380	0.0732	0.1212	0.5132	0.1212	0.0732	0.0380	0.0110			0.9391
10	0.0066	0.0237	0.0464	0.0737	0.1123	0.4905	0.1230	0.0741	0.0385	0.0112		0.9478
	0.0112	0.0385	0.0741	0.1239	0.4905	0.1123	0.0737	0.0464	0.0237	0.0066		0.9478
11	0.0068	0.0240	0.0470	0.0748	0.1143	0.4662	0.1143	0.0748	0.0470	0.0240	0.0068	0.9568

Таблица П1.44.

Оптимальные граничные точки интервалов в виде $t_i = \pi(x_{(i)} - \mu) / \sigma\sqrt{3}$

при одновременном оценивании двух параметров логистического распределения и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7
3	-1.1674	1.1674					
4	-1.9797	0.0	1.9797				
5	-2.5090	-0.8470	0.8470	2.5090			
6	-2.9083	-1.3335	0.0	1.3335	2.9083		
7	-3.2797	-1.7426	-0.5955	0.5955	1.7426	3.2797	
8	-3.6023	-2.0786	-1.0165	0.0	1.0165	2.0786	3.6023
9	-3.8981	-2.3777	-1.3604	-0.4631	0.4631	1.3604	2.3777
10	-4.1665	-2.6442	-1.6512	-0.8235	0.0	0.8235	1.6512
11	-4.4148	-2.8883	-1.9085	-1.1227	-0.3805	0.3805	1.1227
12	-4.6437	-3.1123	-2.1394	-1.3798	-0.6929	0.0	0.6929
13	-4.8577	-3.3212	-2.3511	-1.6081	-0.9580	-0.3225	0.3235
14	-5.0564	-3.5150	-2.5460	-1.8138	-1.1888	-0.5984	0.0
15	-5.2452	-3.6987	-2.7289	-2.0032	-1.3953	-0.8364	-0.2818

t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}	A
							0.4426
							0.6026
							0.6861
							0.7890
							0.8369
							0.8701
3.8981							0.8942
2.6442	4.1665						0.9123
1.9085	2.8883	4.4148					0.9261
1.3798	2.1394	3.1123	4.6437				0.9371
0.9580	1.6081	2.3511	3.3212	4.8577			0.9457
0.5984	1.1888	1.8138	2.5460	3.5150	5.0564		0.9526
0.2818	0.8364	1.3953	2.0032	2.4289	3.6987	5.2452	0.9585

Таблица П1.45.

Оптимальные частоты при одновременном оценивании двух параметров логистического распределения и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
3	0.2373	0.5254	0.2373					
4	0.1214	0.3786	0.3786	0.1214				
5	0.0752	0.2249	0.3998	0.2249	0.0752			
6	0.0517	0.1568	0.2915	0.2915	0.1568	0.0517		
7	0.0363	0.1127	0.2064	0.2892	0.2064	0.1127	0.0363	
8	0.0265	0.0847	0.1545	0.2343	0.2343	0.1545	0.0847	0.0265
9	0.0199	0.0650	0.1193	0.1821	0.2274	0.1821	0.1193	0.0650
10	0.0153	0.0510	0.0946	0.1441	0.1950	0.1950	0.1441	0.0946
11	0.0119	0.0408	0.0764	0.1164	0.1605	0.1880	0.1605	0.1164
12	0.0095	0.0331	0.0627	0.0957	0.1324	0.1666	0.1666	0.1324
13	0.0077	0.0271	0.0521	0.0799	0.1104	0.1425	0.1606	0.1425
14	0.0063	0.0226	0.0438	0.0675	0.0933	0.1212	0.1453	0.1453
15	0.0053	0.0189	0.0371	0.0576	0.0797	0.1037	0.1277	0.1400

P_9	P_{10}	P_{11}	P_{12}	P_{13}	P_{14}	P_{15}	A
							0.4426
							0.6026
							0.6861
							0.7890
							0.8369
							0.8701
0.0199							0.8942
0.0510	0.0153						0.9123
0.0764	0.0408	0.0119					0.9261
0.0957	0.0627	0.0331	0.0095				0.9371
0.1104	0.0799	0.0521	0.0271	0.0077			0.9457
0.1212	0.0933	0.0675	0.0438	0.0226	0.0063		0.9526
0.1277	0.1037	0.0797	0.0576	0.0371	0.0189	0.0053	0.9585

Таблица П1.46.

Оптимальные граничные точки интервалов в виде $t_i = \theta \cdot x_{(i)}$ при оценивании параметра распределения Лапласа и при проверке гипотез о нем по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	A
2	1.8416									0.2920
	-1.8416									
3	-1.5936	1.5936								0.6476
4	-2.7044	-1.1121	1.5183							0.7303
	-1.5183	1.1121	2.7044							
5	-2.6112	-1.0176	1.0176	2.6112						0.8203
6	-2.5670	-0.9745	0.8045	1.8233	3.4173					0.8548
	-3.4173	-1.8233	-0.8045	0.9745	2.5670					
7	-3.3652	-1.7716	-0.7544	0.7544	1.7716	3.3652				0.8910
8	-3.3329	-1.7388	-0.7266	0.6311	1.3816	2.3972	3.9879			0.9087
	-3.9879	-2.3972	-1.3816	-0.6311	0.7266	1.7388	3.3329			
9	-3.9657	-2.3720	-1.3545	-0.6004	0.6004	1.3545	2.3720	3.9657		0.9269
10	-3.9468	-2.3487	-1.3335	-0.5817	0.5217	1.1246	1.8811	2.9054	4.5228	0.9371
	-4.5228	-2.9054	-1.8811	-1.1246	-0.5217	0.5817	1.3335	2.3487	3.9468	

Таблица П1.47.

Оптимальные граничные точки интервалов в виде $t_i = \theta \cdot x_{(i)}$ при оценивании параметра распределения Лапласа и при проверке гипотез о нем по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	$\pm t_1$	$\pm t_2$	$\pm t_3$	$\pm t_4$	$\pm t_5$	$\pm t_6$	$\pm t_7$	$\pm t_8$	$\pm t_9$	A
11	± 0.4993	± 1.0997	± 1.8538	± 2.8714	± 4.4650					0.9476
13	± 0.4276	± 0.9269	± 1.5273	± 2.2813	± 3.2989	± 4.8925				0.9606
15	± 0.3739	± 0.8015	± 1.3008	± 1.9012	± 2.6553	± 3.6729	± 5.2665			0.9693
17	± 0.3323	± 0.7063	± 1.1338	± 1.6331	± 2.2336	± 2.9876	± 4.0052	± 5.5988		0.9754
19	± 0.2990	± 0.6314	± 1.0053	± 1.4329	± 1.9322	± 2.5326	± 3.2866	± 4.3042	± 5.8979	0.9798

Примечание. Граничные точки интервалов симметричны относительно нуля.

Таблица П1.48.

Оптимальные частоты при оценивании параметра распределения Лапласа и при проверке гипотез о нем по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	A
2	0.9207	0.0793									0.2920
	0.0793	0.9207									0.2920
3	0.1016	0.7968	0.1016								0.6476
4	0.0335	0.1310	0.7260	0.1095							0.7303
	0.1095	0.7260	0.1310	0.0335							0.7303
5	0.0367	0.1440	0.6386	0.1440	0.0367						0.8203
6	0.0384	0.1503	0.5877	0.1429	0.0643	0.0164					0.8548
	0.0164	0.0643	0.1429	0.5877	0.1503	0.0384					0.8548
7	0.0173	0.0677	0.1502	0.5296	0.1502	0.0677	0.0173				0.8910
8	0.0179	0.0700	0.1539	0.4922	0.1404	0.0801	0.0362	0.0093			0.9087
	0.0093	0.0362	0.0801	0.1404	0.4922	0.1593	0.0700	0.0178			0.9087
9	0.0094	0.0372	0.0824	0.1453	0.4514	0.1453	0.0824	0.0372	0.0094		0.9269
10	0.0097	0.0381	0.0840	0.1477	0.4238	0.1344	0.0862	0.0488	0.0219	0.0054	0.9371
	0.0054	0.0219	0.0488	0.0862	0.1344	0.4328	0.1477	0.0840	0.0381	0.0096	0.9371

Таблица П1.49.

Оптимальные частоты при оценивании параметра распределения Лапласа и при проверке гипотез о нем по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	P_0	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	A
11	0.3930	0.1370	0.0882	0.0500	0.0225	0.0058					0.9476
13	0.3478	0.1282	0.0894	0.0575	0.0326	0.0147	0.0037				0.9606
15	0.3120	0.1197	0.0882	0.0614	0.0396	0.0224	0.0101	0.0026			0.9693
17	0.2826	0.1119	0.0859	0.0633	0.0441	0.0283	0.0161	0.0073	0.0018		0.9754
19	0.2584	0.1049	0.0829	0.0637	0.0469	0.0327	0.0210	0.0119	0.0054	0.0014	0.9798

Примечание. В таблице приведены значения вероятностей попадания наблюдений в центральный интервал P_0 и в интервалы правее центра: $P_1, P_2, \dots$. Значения вероятностей попадания в интервалы левее центра получаются зеркальным отражением.

Таблица П1.50.

Оптимальные граничные точки интервалов в виде $t_i = \exp\{-\theta_1 x_{(i)}\}$ при оценивании параметра θ_1 двойного показательного распределения и при проверке гипотез о нем по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_i	θ_2										
		0.5	1	2	3	4	5	6	7	8	9	10
2	t_1	5.9243	3.4910	0.1058	0.0952	0.0864	0.0792	0.0730	0.0679	0.0634	0.0594	0.0560
	A	0.5239	0.3283	0.5580	0.6494	0.6742	0.6798	0.6795	0.6774	0.6750	0.6725	0.6702
3	t_1	5.7236	3.2901	0.1714	0.1578	0.1458	0.1353	0.1260	0.1178	0.1106	0.1042	0.0984
	t_2	0.1704	0.1417	0.0371	0.0351	0.0330	0.0312	0.0295	0.0279	0.0265	0.0253	0.0242
	A	0.6465	0.6518	0.6790	0.7947	0.8286	0.8384	0.8404	0.8397	0.8382	0.8365	0.8348
4	t_1	8.2073	4.5218	1.9539	0.1979	0.1851	0.1737	0.1630	0.1535	0.1447	0.1369	0.1296
	t_2	4.5709	2.6958	0.1871	0.0737	0.0694	0.0656	0.0620	0.0588	0.0558	0.0530	0.0506
	t_3	0.1986	0.1504	0.0404	0.0166	0.0161	0.0155	0.0149	0.0143	0.0138	0.0133	0.0128
	A	0.7910	0.7481	0.8049	0.8497	0.8879	0.8999	0.9033	0.9036	0.9029	0.9018	0.9006
5	t_1	9.9575	4.5027	1.9227	1.4712	0.2135	0.2012	0.1906	0.1799	0.1709	0.1617	0.1540
	t_2	6.4293	2.6201	0.2361	0.2097	0.1002	0.0950	0.0898	0.0851	0.0810	0.0770	0.0734
	t_3	3.9668	0.2484	0.0860	0.0772	0.0393	0.0378	0.0362	0.0346	0.0331	0.0318	0.0305
	t_4	0.2096	0.0518	0.0189	0.0174	0.0089	0.0088	0.0085	0.0083	0.0081	0.0078	0.0076
	A	0.8484	0.8235	0.8546	0.8919	0.9167	0.9300	0.9343	0.9353	0.9351	0.9344	0.9336

Продолжение табл. П1.50.

k	t_i	θ_2										
		0.5	1	2	3	4	5	6	7	8	9	10
6	t_1	9.8965	5.3402	2.5662	1.4602	0.2346	0.2231	0.2118	0.2007	0.1916	0.1825	0.1738
	t_2	6.3520	3.5979	1.6106	0.2410	0.1257	0.1192	0.1131	0.1071	0.1025	0.0977	0.0931
	t_3	3.8701	2.3295	0.2457	0.1114	0.0622	0.0596	0.0567	0.0540	0.0522	0.0498	0.0476
	t_4	0.3459	0.2582	0.0889	0.0429	0.0245	0.0238	0.0230	0.0221	0.0216	0.0208	0.0199
	t_5	0.0699	0.0533	0.0195	0.0095	0.0055	0.0054	0.0053	0.0051	0.0051	0.0050	0.0048
	A	0.8799	0.8639	0.8891	0.9200	0.9328	0.9469	0.9518	0.9533	0.9534	0.9530	0.9525
7	t_1	11.2128	5.3662	2.5559	1.4460	1.2202	0.2395	0.2292	0.2182	0.2082	0.1988	0.1904
	t_2	7.7564	3.5651	1.5897	0.2642	0.2443	0.1394	0.1332	0.1264	0.1206	0.1150	0.1103
	t_3	5.4226	2.2911	0.2825	0.1393	0.1297	0.0792	0.0762	0.0723	0.0693	0.0662	0.0637
	t_4	3.5152	0.3265	0.1295	0.0676	0.0639	0.0402	0.0391	0.0373	0.0359	0.0345	0.0334
	t_5	0.3610	0.1156	0.0489	0.0260	0.0252	0.0160	0.0158	0.0152	0.0147	0.0142	0.0139
	t_6	0.0727	0.0245	0.0106	0.0057	0.0056	0.0035	0.0035	0.0035	0.0034	0.0033	0.0032
	A	0.9094	0.8935	0.9141	0.9357	0.9493	0.9574	0.9626	0.9644	0.9648	0.9647	0.9643

Продолжение табл. П1.50.

k	t_i	θ_2										
		0.5	1	2	3	4	5	6	7	8	9	10
8	t_1	12.4184	6.0434	2.9963	1.8917	1.2107	0.2534	0.2425	0.2324	0.2216	0.2134	0.2039
	t_2	8.9812	4.2908	2.0838	1.2386	0.2631	0.1575	0.1495	0.1431	0.1359	0.1307	0.1248
	t_3	6.6955	3.0976	1.4411	0.2722	0.1517	0.0974	0.0926	0.0892	0.0843	0.0815	0.0778
	t_4	4.8980	2.1156	0.2887	0.1435	0.0851	0.0565	0.0540	0.0524	0.0495	0.0482	0.0462
	t_5	3.3170	0.3342	0.1321	0.0701	0.0426	0.0287	0.0278	0.0272	0.0257	0.0252	0.0244
	t_6	0.3690	0.1175	0.0499	0.0272	0.0167	0.0113	0.0111	0.0109	0.0104	0.0103	0.0101
	t_7	0.0743	0.0249	0.0109	0.0060	0.0036	0.0025	0.0024	0.0024	0.0023	0.0023	0.0023
	A	0.9261	0.9141	0.9283	0.9477	0.9596	0.9643	0.9698	0.9718	0.9724	0.9724	0.9721
9	t_1	12.2869	6.0573	2.9859	1.8703	1.2022	1.0648	0.2549	0.2449	0.2349	0.2255	0.2148
	t_2	8.8417	4.2945	2.0795	1.2237	0.2765	0.2619	0.1654	0.1574	0.1508	0.1444	0.1370
	t_3	6.5618	3.0962	1.4280	0.2939	0.1697	0.1608	0.1087	0.1037	0.0991	0.0949	0.0897
	t_4	4.7832	2.1001	0.3166	0.1693	0.1038	0.0987	0.0692	0.0661	0.0634	0.0609	0.0574
	t_5	3.2276	0.3840	0.1652	0.0946	0.0593	0.0569	0.0408	0.0391	0.0377	0.0364	0.0343
	t_6	0.4694	0.1733	0.0795	0.0467	0.0296	0.0287	0.0208	0.0202	0.0197	0.0190	0.0180
	t_7	0.1633	0.0642	0.0302	0.0181	0.0114	0.0113	0.0082	0.0080	0.0079	0.0077	0.0073
	t_8	0.0340	0.0136	0.0064	0.0039	0.0025	0.0024	0.0018	0.0018	0.0017	0.0017	0.0016
	A	0.9387	0.9288	0.9425	0.9577	0.9664	0.9713	0.9748	0.9770	0.9777	0.978	0.9777

Окончание табл. П1.50.

k	t_i	θ_2										
		0.5	1	2	3	4	5	6	7	8	9	10
1 0	t_1	13.2458	6.6629	2.9794	1.8704	1.1963	1.0539	0.2636	0.2536	0.2457	0.2350	0.2258
	t_2	9.8558	4.9268	2.0697	1.2180	0.2912	0.2787	0.1769	0.1688	0.1637	0.1554	0.1491
	t_3	7.6321	3.7589	1.4116	0.3107	0.1888	0.1807	0.1220	0.1160	0.1127	0.1066	0.1021
	t_4	5.8941	2.8248	0.3426	0.1908	0.1234	0.1185	0.0826	0.0785	0.0766	0.0722	0.0693
	t_5	4.4362	1.9905	0.1972	0.1164	0.0776	0.0748	0.0530	0.0507	0.0494	0.0468	0.0450
	t_6	3.0918	0.3890	0.1090	0.0660	0.0452	0.0438	0.0311	0.0298	0.0294	0.0279	0.0269
	t_7	0.4764	0.1748	0.0533	0.0326	0.0228	0.0222	0.0158	0.0152	0.0151	0.0145	0.0140
	t_8	0.1658	0.0642	0.0202	0.0124	0.0088	0.0087	0.0061	0.0060	0.0060	0.0059	0.0056
	t_9	0.0346	0.0135	0.0043	0.0026	0.0019	0.0019	0.0013	0.0013	0.0013	0.0013	0.0013
	A	0.9493	0.9407	0.9513	0.9642	0.9712	0.9762	0.9784	0.9807	0.9815	0.9817	0.9817

Таблица П1.51.

Оптимальные частоты при оценивании параметра θ_1 двойного показательного распределения и при проверке гипотез о нем по критериям χ^2 Пирсона и отношения правдоподобия

k	P_i	θ_2										
		0.5	1	2	3	4	5	6	7	8	9	10
2	P_1	0.0517	0.0305	0.8094	0.7514	0.7078	0.6729	0.6452	0.6216	0.6024	0.5857	0.5710
	P_2	0.9483	0.9695	0.1906	0.2486	0.2922	0.3271	0.3548	0.3784	0.3976	0.4143	0.4290
3	P_1	0.0572	0.0373	0.7098	0.6229	0.5582	0.5084	0.4695	0.4383	0.4129	0.3916	0.3738
	P_2	0.8584	0.8306	0.2187	0.2772	0.3181	0.3472	0.3682	0.3841	0.3956	0.4044	0.4114
	P_3	0.0844	0.1321	0.0715	0.0998	0.1237	0.1444	0.1623	0.1776	0.1913	0.2040	0.2148
4	P_1	0.0165	0.0103	0.0201	0.5522	0.4768	0.4195	0.3760	0.3414	0.3143	0.2917	0.2735
	P_2	0.0878	0.0571	0.6677	0.2494	0.2807	0.3009	0.3134	0.3211	0.3255	0.3288	0.3296
	P_3	0.8012	0.7929	0.2346	0.1497	0.1802	0.2050	0.2250	0.2420	0.2557	0.2669	0.2767
	P_4	0.0945	0.1397	0.0776	0.0487	0.0623	0.0746	0.0856	0.0955	0.1045	0.1126	0.1202
5	P_1	0.0069	0.0111	0.0214	0.0121	0.4258	0.3656	0.3187	0.2839	0.2545	0.2334	0.2144
	P_2	0.0333	0.0617	0.6023	0.5209	0.2440	0.2563	0.2649	0.2672	0.2686	0.2664	0.2658
	P_3	0.0974	0.7073	0.2183	0.2602	0.1849	0.2956	0.2214	0.2340	0.2439	0.2513	0.2571
	P_4	0.7629	0.1694	0.1209	0.1559	0.1103	0.1295	0.1453	0.1587	0.1704	0.1809	0.1896
	P_5	0.0995	0.0505	0.0371	0.0509	0.0350	0.0430	0.0497	0.0562	0.0624	0.0680	0.0731

Продолжение табл. П1.51.

k	P_i	θ_2										
		0.5	1	2	3	4	5	6	7	8	9	10
6	P_1	0.0071	0.0046	0.0059	0.0125	0.3912	0.3277	0.2806	0.2453	0.2159	0.1934	0.1758
	P_2	0.0347	0.0228	0.0340	0.4728	0.2136	0.2232	0.2268	0.2272	0.2245	0.2217	0.2184
	P_3	0.1027	0.0700	0.5718	0.2305	0.1750	0.1913	0.2041	0.2127	0.2182	0.2235	0.2269
	P_4	0.6968	0.6751	0.2254	0.1634	0.1268	0.1446	0.1596	0.1714	0.1830	0.1908	0.1981
	P_5	0.1244	0.1756	0.1246	0.0927	0.0717	0.0856	0.0977	0.1080	0.1185	0.1269	0.1339
	P_6	0.0343	0.0519	0.0383	0.0281	0.0217	0.0266	0.0312	0.0354	0.0399	0.0437	0.0469
7	P_1	0.0037	0.0047	0.0060	0.0131	0.0076	0.3019	0.2528	0.2170	0.1891	0.1670	0.1490
	P_2	0.0170	0.0236	0.0356	0.4396	0.3688	0.1963	0.1970	0.1958	0.1919	0.1882	0.1828
	P_3	0.0458	0.0728	0.5268	0.2058	0.2188	0.1748	0.1833	0.1900	0.1936	0.1959	0.1971
	P_4	0.1051	0.6203	0.2034	0.1579	0.1793	0.1451	0.1579	0.1674	0.1758	0.1820	0.1869
	P_5	0.6632	0.1694	0.1350	0.1085	0.1297	0.1051	0.1187	0.1288	0.1385	0.1467	0.1547
	P_6	0.1295	0.0850	0.0722	0.0582	0.0736	0.0592	0.0693	0.0770	0.0843	0.0909	0.0975
	P_7	0.0357	0.0242	0.0210	0.0169	0.0222	0.0175	0.0210	0.0239	0.0268	0.0293	0.0320

Продолжение табл. П1.51.

k	P_i	θ_2										
		0.5	1	2	3	4	5	6	7	8	9	10
8	P_1	0.0020	0.0024	0.0025	0.0034	0.0079	0.2816	0.2334	0.1966	0.1698	0.1465	0.1302
	P_2	0.0092	0.0113	0.0130	0.0209	0.3412	0.1735	0.1744	0.1707	0.1674	0.1619	0.1569
	P_3	0.0239	0.0315	0.0405	0.4176	0.1960	0.1594	0.1659	0.1683	0.1722	0.1718	0.1723
	P_4	0.0512	0.0754	0.5053	0.2083	0.1663	0.1394	0.1494	0.1573	0.1636	0.1681	0.1706
	P_5	0.1041	0.5953	0.2066	0.1602	0.1318	0.1125	0.1232	0.1339	0.1412	0.1488	0.1538
	P_6	0.6411	0.1733	0.1371	0.1111	0.1921	0.0787	0.0891	0.0998	0.1059	0.1143	0.1204
	P_7	0.1320	0.0862	0.0735	0.0607	0.0502	0.0427	0.0500	0.0565	0.0614	0.0677	0.0731
	P_8	0.0365	0.0246	0.0215	0.0178	0.0145	0.0122	0.0146	0.0109	0.0186	0.0209	0.0227
9	P_1	0.0021	0.0023	0.0025	0.0036	0.0082	0.0049	0.2166	0.1801	0.1527	0.1314	0.1167
	P_2	0.0099	0.0113	0.0131	0.0218	0.3227	0.2651	0.1541	0.1521	0.1466	0.1412	0.1374
	P_3	0.0256	0.0316	0.0419	0.3886	0.1763	0.1775	0.1501	0.1517	0.1532	0.1530	0.1537
	P_4	0.0539	0.0772	0.4734	0.1878	0.1530	0.1629	0.1393	0.1458	0.1497	0.1523	0.1555
	P_5	0.1076	0.5587	0.1877	0.1512	0.1285	0.1421	0.1228	0.1308	0.1373	0.1426	0.1465
	P_6	0.5917	0.1597	0.1344	0.1162	0.0995	0.1137	0.0999	0.1075	0.1150	0.1223	0.1257
	P_7	0.1308	0.0970	0.0883	0.0781	0.0672	0.0789	0.0691	0.0774	0.0842	0.0903	0.0945
	P_8	0.0616	0.0487	0.0459	0.0411	0.0348	0.0428	0.0374	0.0423	0.0474	0.0515	0.0538
	P_9	0.0168	0.0135	0.0128	0.0116	0.0098	0.0121	0.0107	0.0123	0.0139	0.0154	0.0162

Окончание табл. П1.51.

k	P_i	θ_2										
		0.5	1	2	3	4	5	6	7	8	9	10
10	P_1	0.0013	0.0013	0.0026	0.0037	0.0083	0.0052	0.2057	0.1694	0.1401	0.1206	0.1045
	P_2	0.0059	0.0060	0.0133	0.0222	0.3037	0.2430	0.1403	0.1372	0.1299	0.1264	0.1207
	P_3	0.0148	0.0161	0.0435	0.3678	0.1578	0.1570	0.1350	0.1372	0.1359	0.1361	0.1351
	P_4	0.0305	0.0360	0.4446	0.1704	0.1406	0.1478	0.1282	0.1334	0.1360	0.1389	0.1397
	P_5	0.0563	0.0773	0.1702	0.1411	0.1228	0.1348	0.1185	0.1238	0.1319	0.1342	0.1377
	P_6	0.1043	0.5411	0.1300	0.1152	0.1015	0.1155	0.1019	0.1104	0.1166	0.1215	0.1263
	P_7	0.5749	0.1619	0.0947	0.0865	0.0780	0.0914	0.0798	0.0875	0.0955	0.1001	0.1052
	P_8	0.1324	0.0982	0.0615	0.0567	0.0526	0.0628	0.0546	0.0602	0.0673	0.0715	0.0761
	P_9	0.0624	0.0488	0.0310	0.0285	0.0273	0.0332	0.0282	0.0320	0.0364	0.0393	0.0423
	P_{10}	0.0172	0.0133	0.0086	0.0079	0.0074	0.0093	0.0078	0.0089	0.0104	0.0114	0.0124

Таблица П1.52.

Оптимальные граничные точки интервалов в виде $t_i = \theta_2 \exp\{-\theta_1 x_{(i)}\}$ при одновременном оценивании двух параметров двойного показательного распределения и при проверке гипотез о согласии по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
3	2.6067	0.2731						
4	3.4137	1.3979	0.2109					
5	3.8606	1.9590	0.5123	0.1044				
6	4.4096	2.5726	1.2269	0.3649	0.0772			
7	4.7959	2.9922	1.7192	0.6758	0.2318	0.0501		
8	5.2049	3.4285	2.2041	1.1904	0.4837	0.1740	0.0377	
9	5.5273	3.7667	2.5713	1.6027	0.7829	0.3431	0.1269	0.0275
10	5.8478	4.1024	2.9269	1.9932	1.1805	0.5770	0.2638	0.0988
11	6.1270	4.3930	3.2319	2.3192	1.5344	0.8560	0.4359	0.2046
12	6.3853	4.6589	3.5103	2.6163	1.8570	1.1789	0.6517	0.3434
13	6.6208	4.9016	3.7623	2.8810	2.1401	1.4807	0.9030	0.5106
14	6.8444	5.1314	3.9997	3.1286	2.4019	1.7608	1.1798	0.7116
15	7.0506	5.3425	4.2169	3.3538	2.6381	2.0116	1.4426	0.9387

t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}	A
						0.4079
						0.5572
						0.6836
						0.7571
						0.8109
						0.8480
						0.8756
0.0213						0.8963
0.0771	0.0165					0.9123
0.1638	0.0618	0.0123				0.9248
0.2754	0.1326	0.0500	0.0106			0.9349
0.3126	0.2261	0.1094	0.0412	0.0087		0.9431
0.5734	0.3394	0.1881	0.0913	0.0344	0.0072	0.9498

Таблица П1.53.

Оптимальные граничные точки интервалов в виде $t_i = \theta_1 x_{(i)}$ при оценивании основного параметра θ гамма-распределения и при проверке гипотез о нем по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_i	θ									
		0.25	1.25	2.25	3.25	4.25	5.25	6.25	7.25	8.25	9.25
2	t_1	0.02560	0.5812	1.5172	2.4968	3.4855	4.4777	5.4731	6.4690	7.4637	8.4717
	A	0.4848	0.6226	0.6271	0.6302	0.6313	0.6315	0.6317	0.6317	0.6317	0.6317
4	t_1	0.00008	0.1360	0.6269	1.2461	1.9435	2.6924	3.4574	4.2433	5.0593	5.8820
	t_2	0.01220	0.5331	1.4647	2.4277	3.4021	4.4033	5.3792	6.3768	7.3641	8.3689
	t_3	0.49990	1.4805	2.9208	4.2491	5.5048	6.7670	7.9626	9.1637	10.3424	11.5244
	A	0.8376	0.8753	0.8775	0.8786	0.8790	0.8795	0.8802	0.8803	0.8803	0.8803
	A	0.8376	0.8753	0.8775	0.8786	0.8790	0.8795	0.8802	0.8803	0.8803	0.8803
6	t_1	0.00008	0.0583	0.3608	0.8350	1.4157	2.0201	2.6980	3.4151	4.1321	4.8957
	t_2	0.00850	0.2154	0.7959	1.5095	2.2969	3.1410	3.9603	4.8533	5.7161	6.6270
	t_3	0.05000	0.5235	1.3956	2.3411	3.3267	4.3395	5.3086	6.3398	7.3265	8.3597
	t_4	0.25000	1.0848	2.2747	3.4367	4.6268	5.8140	6.9482	8.1186	9.2329	10.3876
	t_5	0.50000	2.1895	3.7786	5.2074	6.5847	8.0378	9.3451	10.6917	11.9613	13.2407
	A	0.8669	0.9381	0.9390	0.9394	0.9397	0.9400	0.9400	0.9403	0.9403	0.9403

Окончание табл. П1.53.

k	t_i	θ									
		0.25	1.25	2.25	3.25	4.25	5.25	6.25	7.25	8.25	9.25
8	t_1	0.00016	0.0304	0.2530	0.6386	1.1291	1.6975	2.3342	2.8195	3.7606	4.1631
	t_2	0.00140	0.1106	0.5449	1.1298	1.7958	2.5380	3.3444	3.9386	4.9636	5.5622
	t_3	0.01080	0.2585	0.9211	1.6845	2.5006	3.3657	4.3244	5.0201	5.9979	6.8770
	t_4	0.04980	0.5003	1.4010	2.3498	3.2771	4.3152	5.2773	6.2247	7.2423	8.1584
	t_5	0.10000	0.8795	2.0323	3.1817	4.2394	5.4257	6.3946	7.5463	8.5933	9.7135
	t_6	0.25000	1.4971	2.9308	4.2911	5.4964	6.9133	7.8272	9.2002	10.2585	11.6583
	t_7	0.5000	2.6347	4.4028	6.0154	7.3689	9.1658	9.9220	11.7413	12.7759	14.6371
	A	0.8746	0.9630	0.9634	0.9636	0.9637	0.9638	0.9639	0.9639	0.9639	0.9639

Таблица П1.54.

Оптимальные граничные точки интервалов в виде $t_i = \theta_1 x_{(i)}$ при оценивании масштабного параметра θ_1 гамма-распределения и при проверке гипотез о нем по критериям χ^2 Пирсона и отношения правдоподобия и соответствующие значения относительной асимптотической информации A

k	t_i	θ									
		0.5	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5
2	t_1	1.0987	2.0915	3.0890	4.0878	5.0871	6.0866	7.0863	8.0861	9.0860	10.0859
	A	0.6524	0.6451	0.6424	0.6410	0.6402	0.6396	0.6392	0.6389	0.6386	0.6384
3	t_1	0.6506	1.4054	2.2152	3.0538	3.9113	4.7823	5.6634	6.5527	7.4486	8.3502
	t_2	1.9942	3.2190	4.4022	5.5596	6.7000	7.8276	8.9456	10.0554	11.1591	12.2573
	A	0.8248	0.8181	0.8155	0.8142	0.8134	0.8128	0.8124	0.8121	0.8119	0.8117
4	t_1	0.4577	1.0797	1.7807	2.5257	3.2998	4.0942	4.9047	5.7278	6.5621	7.4039
	t_2	1.2837	2.2689	3.2637	4.2610	5.2596	6.2581	7.2571	8.2561	9.2565	10.2553
	t_3	2.6849	4.0387	5.3295	6.5806	7.8045	9.0077	10.1957	11.3709	12.5373	13.6934
	A	0.8949	0.8894	0.8874	0.8862	0.8855	0.8851	0.8845	0.8845	0.8843	0.8841
5	t_1	0.3504	0.8846	1.5114	2.1913	2.9069	3.6487	4.4101	5.1886	5.9773	6.7785
	t_2	0.9446	1.7846	2.6639	3.5613	4.4721	5.3913	6.3176	7.2504	8.1836	9.1222
	t_3	1.8149	2.9381	4.0469	5.1410	6.2268	7.3053	8.3786	9.4495	10.5121	11.5724
	t_4	3.2512	4.6884	6.0541	7.3678	8.6503	9.9087	11.1458	12.3714	13.5796	14.7771
	A	0.9303	0.9258	0.9241	0.9232	0.9226	0.9222	0.9219	0.9217	0.9216	0.9216

Продолжение табл. П1.54.

6	t_1	0.2821	0.7529	1.3254	1.9574	2.6297	3.3310	4.0538	4.7951	5.5510	6.3207
	t_2	0.7439	1.4830	2.2807	3.1102	3.9600	4.8212	5.6923	6.5722	7.4604	8.3552
	t_3	1.3746	2.3522	3.3462	4.3449	5.3449	6.3431	7.3392	8.3365	9.3358	10.3366
	t_4	2.2714	3.4885	4.6794	5.8475	6.9990	8.1344	9.2559	10.3708	11.4811	12.5871
	t_5	3.7388	5.2300	6.6516	8.0192	9.3495	10.6457	11.9206	13.1738	14.4153	15.6472
	A	0.9506	0.9469	0.9454	0.9447	0.9442	0.9439	0.9436	0.9435	0.9433	0.9432
7	t_1	0.2354	0.6570	1.1864	1.7811	2.4272	3.0858	3.7806	4.4933	5.2218	5.9648
	t_2	0.6118	1.2736	2.0094	2.7851	3.5836	4.4019	5.2345	6.0769	6.9278	7.7862
	t_3	1.1054	1.9779	2.8864	3.8135	4.7464	5.6870	6.6329	7.5809	8.5321	9.4845
	t_4	1.7588	2.8339	3.9079	4.9786	6.0408	7.0973	8.1493	9.1977	10.2445	11.2856
	t_5	2.6765	3.9625	5.2126	6.4331	7.6294	8.8116	9.9762	11.1317	12.2782	13.4140
	t_6	4.1759	5.7057	7.1625	8.5672	9.9313	11.2609	12.5672	13.8542	15.1226	16.3779
	A	0.9635	0.9602	0.9590	0.9583	0.9579	0.9576	0.9574	0.9573	0.9572	0.9571
8	t_1	0.2013	0.5866	1.0791	1.6413	2.2474	2.8905	3.5604	4.2517	4.9591	5.6868
	t_2	0.5196	1.1236	1.8050	2.5367	3.2949	4.0788	4.8797	5.6928	6.5165	7.3538
	t_3	0.9252	1.7181	2.5552	3.4253	4.3063	5.2043	6.1090	7.0198	7.9362	8.8613
	t_4	1.4396	2.4098	3.3916	4.3888	5.3838	6.3867	7.3870	8.3886	9.3878	10.3939
	t_5	2.1110	3.2594	4.3852	5.5099	6.6189	7.7258	8.8239	9.9148	10.9964	12.0860
	t_6	3.0525	4.3859	5.6698	6.9333	8.1646	9.3865	10.5910	11.7773	12.9525	14.1361
	t_7	4.5961	6.1313	7.6087	9.0476	10.4281	11.7946	13.1305	14.4401	15.7335	17.0324
	A	0.9721	0.9691	0.9681	0.9675	0.9672	0.9669	0.9667	0.9666	0.9665	0.9664

Окончание табл. П1.54.

9	t_1	0.1773	0.5313	0.9956	1.5303	2.1150	2.7324	3.3789	4.0449	4.7360	5.4398
	t_2	0.4528	1.0089	1.6506	2.3453	3.0741	3.8233	4.5926	5.3753	6.1746	6.9832
	t_3	0.7983	1.5258	2.3122	3.1347	3.9794	4.8334	5.6993	6.5732	7.4583	8.3495
	t_4	1.2256	2.1095	3.0274	3.9676	4.9157	5.8633	6.8161	7.7715	8.7340	9.6984
	t_5	1.7612	2.7945	3.8423	4.8952	5.9426	6.9826	8.0190	9.0538	10.0914	11.1265
	t_6	2.4580	3.6385	4.8221	5.9940	7.1419	8.2769	9.4001	10.5180	11.6304	12.7398
	t_7	3.4343	4.7622	6.0988	7.4071	8.6596	9.9012	11.1213	12.3304	13.5290	14.7226
	t_8	5.0431	6.5229	8.0419	9.5272	10.9048	12.2812	13.6256	14.9504	16.2603	17.5596
	A	0.9782	0.9755	0.9746	0.9740	0.9737	0.9735	0.9733	0.9732	0.9732	0.9731
10	t_1	0.1564	0.4805	0.9207	1.4330	1.9989	2.6007	3.2306	3.8836	4.5538	5.2395
	t_2	0.3970	0.9054	1.5157	2.1798	2.8852	3.6143	4.3613	5.1244	5.9012	6.6888
	t_3	0.6938	1.3583	2.1046	2.8913	3.7072	4.5343	5.3751	6.2251	7.0817	7.9478
	t_4	1.0518	1.8594	2.7294	3.6246	4.5386	5.4555	6.3773	7.3036	8.2330	9.1644
	t_5	1.4871	2.4307	3.4182	4.4183	5.4245	6.4268	7.4274	8.4243	9.4222	10.4160
	t_6	2.0301	3.1082	4.2105	5.3174	6.4171	7.5033	8.5815	9.6512	10.7163	11.7759
	t_7	2.7376	3.9503	5.1693	6.3900	7.5885	8.7651	9.9240	11.0706	12.2041	13.3339
	t_8	3.7375	5.0909	6.4259	7.7789	9.0836	10.3627	11.6150	12.8468	14.0622	15.2675
	t_9	5.4272	6.8830	8.3437	8.3437	9.8770	11.3142	12.7271	14.0942	15.4375	18.0697
	A	0.9828	0.9801	0.9792	0.9788	0.9785	0.9783	0.9782	0.9782	0.9780	0.9779

ПРИЛОЖЕНИЕ 2. Таблицы коэффициентов оптимальных L-оценок для больших выборок

Таблица П2.1.

Нормальное распределение. Известен масштабный параметр.

Коэффициенты параметра сдвига.

k	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
2	0	1								
3	0	0.5	0.5							
4	0	0.295397	0.409207	0.295397						
5	0	0.191747	0.308253	0.308253	0.191747					
6	0	0.132788	0.232629	0.269166	0.232629	0.132788				
7	0	0.096375	0.178771	0.224854	0.224854	0.178771	0.096375			
8	0	0.072178	0.140209	0.186282	0.202664	0.186282	0.140209	0.072178		
9	0	0.056339	0.111523	0.154649	0.177489	0.177489	0.154649	0.111523	0.056339	
10	0	0.041318	0.097367	0.127902	0.152800	0.161226	0.152800	0.127902	0.097367	0.041318

Таблица П2.2.

Нормальное распределение. Известен параметр сдвига.

Коэффициенты масштабного параметра.

k	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
2	0.634921	-0.634921								
3	0	-0.337359	0.337359							
4	0.064150	-0.122413	-0.252996	0.311258						
5	0	-0.115285	-0.236766	0.236766	0.115285					
6	0.021513	-0.055242	-0.126006	-0.185699	0.231609	0.113824				
7	0	-0.054640	-0.124571	-0.182339	0.182339	0.124571	0.054640			
8	0.010586	-0.030816	-0.073132	-0.117180	-0.148499	0.180589	0.123829	0.054624		
9	0	-0.031157	-0.072629	-0.116643	-0.147029	0.147029	0.116643	0.072629	0.031157	
10	0.005915	-0.019601	-0.046317	-0.077199	-0.106036	-0.123379	0.146340	0.116468	0.072905	0.030905

Таблица П2.3.

Коэффициенты L -оценки параметра сдвига нормального распределения. Неизвестны оба параметра.

k	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9	γ_{10}	γ_{11}	γ_{12}	γ_{13}	γ_{14}
3	0.500000	0.500000												
4	0.224374	0.551252	0.224374											
5	0.108579	0.391421	0.391421	0.108579										
6	0.067815	0.234061	0.396249	0.234061	0.067815									
7	0.043180	0.141936	0.314884	0.314884	0.141936	0.043180								
8	0.029871	0.096902	0.216939	0.312575	0.216939	0.096902	0.029871							
9	0.021547	0.068108	0.148605	0.261739	0.261739	0.148605	0.068108	0.021547						
10	0.016187	0.050213	0.107748	0.196679	0.258345	0.196679	0.107748	0.050213	0.016187					
11	0.012568	0.037666	0.080500	0.145524	0.223741	0.223741	0.145524	0.080500	0.037666	0.012568				
12	0.002356	0.078666	0.032450	0.099837	0.188002	0.197378	0.188002	0.099837	0.032450	0.078666	0.002356			
13	0.008056	0.023716	0.048181	0.086640	0.138225	0.195182	0.195182	0.138225	0.086640	0.048181	0.023716	0.008056		
14	0.006737	0.018173	0.039623	0.068299	0.109776	0.161527	0.191732	0.161527	0.109776	0.068299	0.039623	0.018173	0.006737	
15	0.005076	0.015581	0.032157	0.055371	0.088028	0.130918	0.172869	0.172869	0.130918	0.088028	0.055371	0.032157	0.015581	0.005076

Таблица П2.4.

Коэффициенты L -оценки параметра масштаба нормального распределения. Неизвестны оба параметра.

k	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}	v_{11}	v_{12}	v_{13}	v_{14}
3	-0.450207	0.450207												
4	-0.361428	0	0.361428											
5	-0.201360	-0.229872	0.229872	0.201360										
6	-0.140732	-0.235892	0	0.235892	0.140732									
7	-0.095717	-0.186279	-0.136715	0.136715	0.186279	0.095717								
8	-0.070411	-0.147147	-0.166972	0	0.166972	0.147147	0.070411							
9	-0.052747	-0.114684	-0.153492	-0.090860	0.090860	0.153492	0.114684	0.052747						
10	-0.040995	-0.091463	-0.132388	-0.123812	0	0.123812	0.132388	0.091463	0.040995					
11	-0.032533	-0.073373	-0.111849	-0.124976	-0.064980	0.064980	0.124976	0.111849	0.073373	0.032533				
12	-0.019239	-0.098971	-0.069069	-0.112065	-0.080404	0	0.080404	0.112065	0.069069	0.098971	0.019239			
13	-0.021688	-0.050048	-0.079521	-0.102566	-0.102357	-0.048835	0.048835	0.102357	0.102566	0.079521	0.050048	0.021688		
14	-0.018214	-0.041087	-0.068460	-0.090418	-0.099244	-0.075434	0	0.075434	0.099244	0.090418	0.068460	0.041087	0.018214	
15	-0.014676	-0.035496	-0.058652	-0.079499	-0.092263	-0.085051	-0.038353	0.038353	0.085051	0.092263	0.079499	0.058652	0.035496	0.014676

Таблица П2.5.

Логистическое распределение. Известен масштабный параметр.
Коэффициенты параметра сдвига.

k	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
2	0	1.000000								
3	0	0.500000	0.500000							
4	0	0.300005	0.399990	0.300005						
5	0	0.200003	0.299997	0.299997	0.200003					
6	0	0.142866	0.228583	0.257103	0.228583	0.142866				
7	0	0.107125	0.178585	0.214290	0.214290	0.178585	0.107125			
8	0	0.083334	0.142856	0.178581	0.190456	0.178581	0.142856	0.083334		
9	0	0.066695	0.116598	0.150057	0.166650	0.166650	0.150057	0.116598	0.066695	
10	0	0.054558	0.096957	0.127285	0.145431	0.151541	0.145431	0.127285	0.096957	0.054558

Таблица П2.6.

Логистическое распределение. Известен параметр сдвига.
Коэффициенты масштабного параметра.

k	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
2	0.755970	-0.755970								
3	0	-0.418872	0.418872							
4	0.061267	-0.135300	-0.322464	0.396497						
5	0	-0.130116	-0.305735	0.302100	0.138061					
6	0.019847	-0.058311	-0.152800	-0.243988	0.304532	0.130720				
7	0	-0.058446	-0.152535	-0.241265	0.241265	0.152535	0.058446			
8	0.008962	-0.030625	-0.084527	-0.147981	-0.197890	0.240583	0.152676	0.058803		
9	0	-0.031087	-0.084851	-0.148233	-0.196942	0.196942	0.148233	0.084851	0.031087	
10	0.005069	-0.018618	-0.051311	-0.093647	-0.137254	-0.165922	0.196706	0.148756	0.085222	0.030999

Таблица П2.7.

Логистическое распределение. Неизвестны оба параметра.

Коэффициенты параметра сдвига.

k	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7
3	0.500000	0.500000					
4	0.180032	0.639936	0.180032				
5	0.068759	0.431241	0.431241	0.068759			
6	0.032667	0.235580	0.463506	0.235580	0.032667		
7	0.016276	0.127108	0.356615	0.356615	0.127108	0.016276	
8	0.009070	0.073276	0.235704	0.363901	0.235704	0.073276	0.009070
9	0.005027	0.044244	0.151113	0.299616	0.299616	0.151113	0.044244
10	0.002907	0.027965	0.098968	0.220560	0.299200	0.220560	0.098968
11	0.002076	0.017817	0.066423	0.156705	0.256979	0.256979	0.156705
12	0.001340	0.011776	0.045582	0.111840	0.202450	0.254025	0.202450
13	0.000801	0.008215	0.032047	0.080289	0.154017	0.224845	0.224247
14	0.000669	0.005497	0.022853	0.058798	0.116512	0.185366	0.220609
15	0.000199	0.003813	0.015948	0.041514	0.084367	0.141300	0.189859

Продолжение Табл. П2.7.

k	γ_8	γ_9	γ_{10}	γ_{11}	γ_{12}	γ_{13}	γ_{14}
3							
4							
5							
6							
7							
8							
9	0.005027						
10	0.027965	0.002907					
11	0.066423	0.017817	0.002076				
12	0.111840	0.045582	0.011776	0.001340			
13	0.154184	0.080289	0.032047	0.008215	0.000801		
14	0.185366	0.116512	0.058798	0.022853	0.005497	0.000669	
15	0.189859	0.141300	0.084367	-0.047407	0.181940	-0.027260	0.000199

Таблица П2.8.

Логистическое распределение. Неизвестны оба параметра.

Коэффициенты масштабного параметра.

k	v_1	v_2	v_3	v_4	v_5	v_6	v_7
3	-0.776854	0.776854					
4	-0.458100	0	0.458100				
5	-0.258222	-0.305810	0.305810	0.258222			
6	-0.168553	-0.312483	0	0.312483	0.168553		
7	-0.112256	-0.248569	-0.177290	0.177290	0.248569	0.112256	
8	-0.079479	-0.191015	-0.219920	0	0.219920	0.191015	0.079479
9	-0.057642	-0.146792	-0.205412	-0.116039	0.116039	0.205412	0.146792
10	-0.043187	-0.114395	-0.176647	-0.161259	0	0.161259	0.176647
11	-0.033660	-0.089706	-0.147760	-0.165782	-0.081671	0.081671	0.165782
12	-0.026443	-0.071718	-0.123177	-0.154062	-0.122382	0	0.122382
13	-0.020990	-0.058367	-0.102783	-0.137369	-0.135186	-0.059276	0.059789
14	-0.017295	-0.047587	-0.086118	-0.120354	-0.132429	-0.095578	0
15	-0.013019	-0.037789	-0.069312	-0.099835	-0.117940	-0.105469	-0.044735

Продолжение Табл. П2.8.

k	v_8	v_9	v_{10}	v_{11}	v_{12}	v_{13}	v_{14}
3							
4							
5							
6							
7							
8							
9	0.057642						
10	0.114395	0.043187					
11	0.147760	0.089706	0.033660				
12	0.154062	0.123177	0.071718	0.026443			
13	0.134756	0.137369	0.102783	0.058367	0.020990		
15	0.095578	0.132429	0.120354	0.086118	0.047587	0.017295	
15	0.044735	0.105469	0.117940	0.045811	0.187676	0.018911	0.013019

Таблица П2.9.

Распределение Коши. Известен масштабный параметр.
Коэффициенты параметра сдвига.

k	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
2	0	1.000000								
3	0	0.500000	0.500000							
4	-0.069618	-0.032472	0.502862	0.529609						
5	0	-0.029899	0.529899	0.529899	-0.029899					
6	0	-0.033723	0.336301	0.394844	0.336301	-0.033723				
7	0	-0.035583	0.240362	0.295221	0.295221	0.240362	-0.035583			
8	-0.009075	-0.009071	-0.035351	0.236730	0.298054	0.300385	0.244491	-0.035238		
9	0	-0.008981	-0.034899	0.240933	0.302947	0.302947	0.240933	-0.034899	-0.008981	
10	0	-0.009367	-0.036311	0.183752	0.236240	0.251371	0.236240	0.183752	-0.036311	-0.009367

Таблица П2.10.

Распределение Коши. Известен параметр сдвига.
Коэффициенты масштабного параметра.

k	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
2	0.753352	-0.753352								
3	0	-0.500000	0.500000							
4	-0.019303	-0.507605	0.369455	0.157453						
5	0	-0.164801	-0.379276	0.379276	0.164801					
6	-0.008426	-0.168069	-0.384011	0.287400	0.196815	0.076291				
7	0	-0.078485	-0.197507	-0.291445	0.291445	0.197507	0.078485			
8	-0.001648	-0.079673	-0.200011	-0.293518	0.231220	0.184869	0.114843	0.043918		
9	0	-0.044722	-0.116728	-0.187479	-0.232872	0.232872	0.187479	0.116728	0.044722	
10	-0.000755	-0.045237	-0.117888	-0.189199	-0.233560	0.191379	0.167963	0.124473	0.074757	0.028066

Таблица П2.11.

Распределение Коши. Неизвестны оба параметра.

Коэффициенты параметра сдвига.

k	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7
3	0.500000	0.500000					
4	0	1.000000	0				
5	-0.085415	0.585415	0.585415	-0.085415			
6	-0.083338	0.250058	0.666560	0.250058	-0.083338		
7	-0.067065	0.077672	0.489393	0.489393	0.077672	-0.067065	
8	-0.051776	-0.000008	0.301801	0.499967	0.301801	-0.000008	-0.051776
9	-0.039830	-0.031878	0.166716	0.404992	0.404992	0.166716	-0.031878
10	-0.030901	-0.042682	0.080792	0.292814	0.399953	0.292814	0.080792
11	-0.024278	-0.044162	0.029601	0.197000	0.341839	0.341839	0.197000
12	-0.019338	-0.041657	-0.000054	0.125080	0.269365	0.333208	0.269365
13	-0.015605	-0.037732	-0.016315	0.073863	0.201276	0.294513	0.294513
14	-0.012750	-0.033543	-0.024668	0.038828	0.144688	0.244642	0.285606
15	-0.010537	-0.029522	-0.026876	0.010359	0.103436	0.195122	0.257985

Продолжение Табл. П2.11.

k	γ_8	γ_9	γ_{10}	γ_{11}	γ_{12}	γ_{13}	γ_{14}
3							
4							
5							
6							
7							
8							
9	-0.039830						
10	-0.042682	-0.030901					
11	0.029601	-0.044162	-0.024278				
12	0.125080	-0.000054	-0.041657	-0.019338			
13	0.201276	0.073863	-0.016315	-0.037732	-0.015605		
14	0.244642	0.144688	0.038828	-0.024668	-0.033543	-0.012750	
15	0.257985	0.195122	0.100008	0.015434	-0.028455	-0.029522	-0.010537

Таблица П2.12.

Распределение Коши. Неизвестны оба параметра.
Коэффициенты масштабного параметра.

k	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7
3	-0.866101	0.866101					
4	-0.500000	0	0.500000				
5	-0.262874	-0.425301	0.425301	0.262874			
6	-0.144359	-0.433000	0	0.433000	0.144359		
7	-0.084109	-0.340573	-0.235501	0.235501	0.340573	0.084109	
8	-0.051778	-0.250008	-0.301759	0	0.301759	0.250008	0.051778
9	-0.033418	-0.180883	-0.288675	-0.147331	0.147331	0.288675	0.180883
10	-0.022447	-0.131427	-0.249032	-0.212590	0	0.212590	0.249032
11	-0.015590	-0.096691	-0.205584	-0.227389	-0.100494	0.100494	0.227389
12	-0.011168	-0.072177	-0.166688	-0.216527	-0.155371	0	0.155371
13	-0.008186	-0.054700	-0.134282	-0.194830	-0.178544	-0.072310	0.072310
14	-0.006141	-0.042072	-0.108315	-0.170289	-0.181275	-0.117732	0
15	-0.004690	-0.032807	-0.087472	-0.146571	-0.172799	-0.141745	-0.054838

Продолжение Табл. П2.12.

k	ν_8	ν_9	ν_{10}	ν_{11}	ν_{12}	ν_{13}	ν_{14}
3							
4							
5							
6							
7							
8							
9	0.033418						
10	0.131427	0.022447					
11	0.205584	0.096691	0.015590				
12	0.216527	0.166688	0.072177	0.011168			
13	0.178544	0.194830	0.134282	0.054700	0.008186		
15	0.117732	0.181275	0.170289	0.108315	0.042072	0.006141	
15	0.054838	0.141745	0.173154	0.146442	0.087636	0.032807	0.004690

Таблица П2.13.

Распределение минимального значения. Известен масштабный параметр.

Коэффициенты параметра сдвига.

k	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
2	-0.465996	1.000000								
3	-0.457864	0.532646	0.467354							
4	-0.451558	0.337327	0.401156	0.261517						
5	-0.446464	0.234433	0.319531	0.282886	0.163150					
6	-0.439500	0.173268	0.255595	0.260389	0.204166	0.106582				
7	-0.438789	0.132909	0.206402	0.227756	0.205867	0.150678	0.076389			
8	-0.435234	0.105473	0.170026	0.197786	0.193612	0.163122	0.115124	0.054857		
9	-0.434650	0.085691	0.141899	0.171209	0.176729	0.162204	0.130588	0.089212	0.042467	
10	-0.431147	0.071154	0.120352	0.149321	0.159955	0.154513	0.135559	0.106161	0.072205	0.030780

Таблица П2.14.

Распределение минимального значения. Известен параметр сдвига.

Коэффициенты масштабного параметра.

k	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
2	-0.800008	0.800008								
3	-0.215568	-0.236436	0.452003							
4	-0.282815	-0.217377	0.324412	0.175780						
5	-0.224455	-0.070208	-0.177386	0.304291	0.167758					
6	-0.247175	-0.069342	-0.173996	0.233460	0.172956	0.084097				
7	-0.227945	-0.030560	-0.083598	-0.141485	0.228320	0.171070	0.084197			
8	-0.238506	-0.030609	-0.083324	-0.140479	0.183618	0.156220	0.103241	0.049839		
9	-0.228374	-0.016560	-0.045192	-0.082831	-0.117161	0.182167	0.156321	0.103947	0.047683	
10	-0.233782	-0.016276	-0.045826	-0.083200	-0.117318	0.151883	0.139842	0.107086	0.069384	0.028207

Таблица П2.15.

Распределение минимального значения. Неизвестны оба параметра.

Коэффициенты параметра сдвига.

k	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7
3	0.424683	0.575317					
4	0.259011	0.567488	0.173501				
5	0.087506	0.370746	0.421752	0.119996			
6	0.059067	0.240645	0.409087	0.226108	0.065093		
7	0.034040	0.128765	0.314224	0.327084	0.150634	0.045253	
8	0.024398	0.088128	0.223156	0.321325	0.214830	0.098316	0.029848
9	0.016937	0.058429	0.144647	0.267738	0.266161	0.152971	0.071629
10	0.012575	0.043104	0.102543	0.204335	0.265207	0.195430	0.108795
11	0.009448	0.031650	0.073962	0.147302	0.231116	0.228465	0.144645
12	0.004967	0.026789	0.056002	0.109884	0.186937	0.225597	0.176844
13	0.006079	0.018848	0.042830	0.082698	0.143306	0.202875	0.195379
14	0.005025	0.016026	0.035664	0.124799	-0.036904	0.262761	0.197126
15	0.003854	0.012502	0.026878	0.050692	0.086782	0.136910	0.180218

Продолжение Табл. П2.15.

k	γ_8	γ_9	γ_{10}	γ_{11}	γ_{12}	γ_{13}	γ_{14}
3							
4							
5							
6							
7							
8							
9	0.021488						
10	0.052394	0.015617					
11	0.082857	0.038607	0.011948				
12	0.110986	0.063256	0.029678	0.009060			
13	0.138991	0.086621	0.050101	0.023869	0.008403		
14	0.158732	0.108465	0.066662	0.038618	0.018720	0.004305	
15	0.173365	0.130048	0.089562	0.055200	0.034471	0.015626	0.003891

Таблица П2.16.

Распределение минимального значения. Неизвестны оба параметра.

Коэффициенты масштабного параметра.

k	v_1	v_2	v_3	v_4	v_5	v_6	v_7
3	-0.443262	0.443262					
4	-0.362597	0.010676	0.351921				
5	-0.164819	-0.265862	0.194586	0.236094			
6	-0.116846	-0.249166	-0.024255	0.231207	0.159060		
7	-0.072171	-0.178331	-0.170090	0.107189	0.196781	0.116621	
8	-0.052789	-0.136821	-0.184766	-0.024164	0.155898	0.157513	0.085128
9	-0.037595	-0.100410	-0.159593	-0.116944	0.068569	0.151888	0.128660
10	-0.028437	-0.078267	-0.132102	-0.142130	-0.020502	0.111319	0.135342
11	-0.021709	-0.060751	-0.106995	-0.136589	-0.084165	0.046187	0.120032
12	-0.013695	-0.051398	-0.088032	-0.121692	-0.113227	-0.017017	0.083450
13	-0.013850	-0.038567	-0.071379	-0.103285	-0.115343	-0.065595	0.034214
14	-0.011017	-0.031711	-0.059539	-0.084847	-0.099454	-0.086942	-0.017490
15	-0.009104	-0.026590	-0.049557	-0.075451	-0.096666	-0.097958	-0.052212

Продолжение Табл. П2.16.

k	v_8	v_9	v_{10}	v_{11}	v_{12}	v_{13}	v_{14}
3							
4							
5							
6							
7							
8							
9	0.065425						
10	0.104056	0.050722					
11	0.118560	0.084909	0.040520				
12	0.114646	0.102898	0.071025	0.033041			
13	0.093916	0.104064	0.087906	0.059204	0.028715		
15	0.060535	0.092338	0.092034	0.075032	0.050203	0.020858	
15	0.026068	0.077015	0.091604	0.084358	0.067620	0.043148	0.017726

Таблица П2.17.

Распределение максимального значения. Известен масштабный параметр.

Коэффициенты параметра сдвига.

k	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
2	0.466000	1.000000								
3	0.457846	0.467366	0.532634							
4	0.451551	0.261502	0.401151	0.337347						
5	0.446446	0.163107	0.282950	0.319530	0.234412					
6	0.439447	0.106538	0.204170	0.260439	0.255570	0.173284				
7	0.438801	0.076396	0.150624	0.205966	0.227713	0.206407	0.132894			
8	0.435196	0.054833	0.115141	0.163062	0.193653	0.197802	0.170041	0.105469		
9	0.434102	0.042420	0.089228	0.130487	0.162259	0.176658	0.171181	0.141422	0.086345	
10	0.431129	0.030780	0.072225	0.106010	0.135783	0.154368	0.159998	0.149319	0.120359	0.071157

Таблица П2.18.

Распределение максимального значения. Известен параметр сдвига.

Коэффициенты масштабного параметра.

k	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
2	0.800000	-0.800000								
3	0.215640	-0.452038	0.236398							
4	0.282847	-0.175751	-0.324447	0.217351						
5	0.224370	-0.167758	-0.304212	0.177338	0.070262					
6	0.247618	-0.084962	-0.172411	-0.233338	0.173928	0.069165				
7	0.227362	-0.082969	-0.171188	-0.229064	0.141544	0.083647	0.030668			
8	0.239557	-0.049378	-0.104800	-0.155945	-0.183386	0.140280	0.083182	0.030489		
9	0.228374	-0.047729	-0.104272	-0.155873	-0.182392	0.117264	0.082857	0.045594	0.016178	
10	0.233517	-0.028196	-0.069328	-0.107331	-0.139477	-0.151871	0.117331	0.083271	0.045744	0.016340

Таблица П2.19.

Распределение максимального значения. Неизвестны оба параметра.

Коэффициенты параметра сдвига.

k	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7
3	0.575291	0.424709					
4	0.173521	0.567426	0.259052				
5	0.120076	0.421664	0.370677	0.087583			
6	0.065096	0.226173	0.409017	0.240575	0.059140		
7	0.045393	0.150584	0.326990	0.314076	0.128907	0.034050	
8	0.029839	0.098336	0.214838	0.321288	0.223260	0.087985	0.024454
9	0.021566	0.071175	0.153266	0.266189	0.267807	0.144652	0.058529
10	0.015625	0.052488	0.108575	0.195528	0.265200	0.204361	0.102587
11	0.010606	0.035637	0.077872	0.136593	0.219031	0.177120	0.256522
12	0.009273	0.030094	0.063836	0.111789	0.177120	0.225086	0.186016
13	0.008419	0.023822	0.050164	0.086647	0.139064	0.195342	0.202864
14	0.004876	0.020009	0.040625	0.068983	0.110637	0.159960	0.196076
15	0.003892	0.015612	0.034470	0.055220	0.089524	0.129984	0.173410

Продолжение Табл. П2.19.

k	γ_8	γ_9	γ_{10}	γ_{11}	γ_{12}	γ_{13}	γ_{14}
3							
4							
5							
6							
7							
8							
9	0.016816						
10	0.043083	0.012553					
11	0.045453	0.031558	0.009610				
12	0.108941	0.055241	0.025406	0.007196			
13	0.143174	0.082790	0.042654	0.018953	0.006107		
14	0.170246	0.110870	0.063887	0.033886	0.015181	0.004761	
15	0.180334	0.137064	0.087013	0.050281	0.026874	0.012305	0.004016

Таблица П2.20.

Распределение максимального значения. Неизвестны оба параметра.

Коэффициенты масштабного параметра.

k	v_1	v_2	v_3	v_4	v_5	v_6	v_7
3	-0.443282	0.443282					
4	-0.351904	-0.010733	0.362637				
5	-0.236074	-0.194577	0.265752	0.164899			
6	-0.159023	-0.231198	0.024209	0.249065	0.116946		
7	-0.116742	-0.196654	-0.107178	0.170028	0.178352	0.072194	
8	-0.085122	-0.157524	-0.155873	0.024169	0.184789	0.136716	0.052846
9	-0.065490	-0.128455	-0.152073	-0.068643	0.117014	0.159652	0.100547
10	-0.050761	-0.104111	-0.135300	-0.111312	0.020548	0.142141	0.132144
11	-0.041868	-0.087939	-0.122325	-0.123398	-0.047017	0.121113	0.087018
12	-0.032743	-0.070381	-0.102042	-0.113745	-0.082907	0.016227	0.111896
13	-0.028719	-0.059143	-0.087926	-0.104051	-0.093872	-0.034199	0.065508
14	-0.021210	-0.051141	-0.076712	-0.094397	-0.095515	-0.064268	0.013982
15	-0.017731	-0.043137	-0.067618	-0.084341	-0.091560	-0.077010	-0.026040

Продолжение Табл. П2.20.

k	v_8	v_9	v_{10}	v_{11}	v_{12}	v_{13}	v_{14}
3							
4							
5							
6							
7							
8							
9	0.037447						
10	0.078250	0.028400					
11	0.130226	0.061918	0.022272				
12	0.120328	0.086999	0.049424	0.016944			
13	0.115339	0.103269	0.071232	0.038662	0.013898		
15	0.090474	0.107864	0.088380	0.059537	0.031866	0.011140	
15	0.052158	0.098110	0.096826	0.075233	0.049476	0.026311	0.009324

Таблица П2.21.

Экспоненциальное распределение. Известен масштабный параметр.

Коэффициенты параметра сдвига.

k	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
2	-1.593600	1.000000								
3	-1.017438	1.000102	-0.000101							
4	-0.754284	0.999914	0.000025	0.000061						
5	-0.600472	1.000031	-0.000034	-0.000055	0.000058					
6	-0.499093	1.000006	0.000049	0.000004	0.000004	-0.000064				
7	-0.427654	1.000006	-0.000045	-0.000003	0.000037	0.000006	-0.000001			
8	-0.373854	0.999966	0.000056	-0.000004	-0.000004	-0.000022	0.000030	-0.000022		
9	-0.332335	1.000018	-0.000052	0.000015	-0.000007	0.000034	-0.000019	0.000012	-0.000001	
10	-0.298922	1.000029	0.000002	0.000010	-0.000024	-0.000011	-0.000009	-0.000008	0.000035	-0.000025

Таблица П2.22.

Экспоненциальное распределение. Известен параметр сдвига.

Коэффициенты масштабного параметра.

k	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
2	-0.627510	0.627510								
3	-0.702414	0.523434	0.178981							
4	-0.751473	0.447324	0.226437	0.077712						
5	-0.786766	0.390462	0.235904	0.119260	0.041140					
6	-0.814880	0.347021	0.232423	0.140462	0.071103	0.023870				
7	-0.834158	0.310826	0.222679	0.149123	0.090241	0.045675	0.015613			
8	-0.851302	0.282088	0.212135	0.152049	0.101837	0.061432	0.031344	0.010416		
9	-0.864189	0.257873	0.200905	0.151005	0.108217	0.072620	0.043710	0.022274	0.007585	
10	-0.876537	0.237973	0.190612	0.148533	0.111631	0.079967	0.053526	0.032301	0.016775	0.005219

Таблица П2.23.

Экспоненциальное распределение. Неизвестны оба параметра.
Коэффициенты параметра сдвига.

k	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9
3	1.638554	-0.638554							
4	1.529414	-0.394234	-0.135179						
5	1.451127	-0.268568	-0.135844	-0.046715					
6	1.393581	-0.195435	-0.118152	-0.059806	-0.020188				
7	1.347928	-0.148110	-0.099144	-0.059935	-0.030356	-0.010382			
8	1.312099	-0.116249	-0.083388	-0.055851	-0.033721	-0.017145	-0.005744		
9	1.282714	-0.093743	-0.070383	-0.050465	-0.033810	-0.020407	-0.010368	-0.003538	
10	1.258692	-0.077205	-0.060151	-0.045249	-0.032405	-0.021693	-0.013095	-0.006747	-0.002147

Таблица П2.24.

Экспоненциальное распределение. Неизвестны оба параметра.
Коэффициенты масштабного параметра.

k	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8	ν_9
3	-0.627510	0.627510							
4	-0.701990	0.522694	0.179296						
5	-0.751236	0.447204	0.226138	0.077894					
6	-0.788580	0.391679	0.236742	0.119838	0.040321				
7	-0.813559	0.346226	0.231825	0.140236	0.070997	0.024274			
8	-0.834905	0.311098	0.223038	0.149383	0.090140	0.045940	0.015307		
9	-0.850634	0.281917	0.211828	0.151828	0.101839	0.061345	0.031234	0.010642	
10	-0.865318	0.258285	0.201258	0.151293	0.108372	0.072540	0.043779	0.022689	0.007101

Таблица П2.25.

Полунормальное распределение (модуля многомерного нормального вектора $\mathbf{m}=1$). Известен масштабный параметр.
Коэффициенты параметра сдвига.

k	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
2	-1.482100	1.000000								
3	-1.227631	0.897007	0.102993							
4	-1.081004	0.844696	0.119280	0.036024						
5	-0.980236	0.815760	0.116454	0.051892	0.015894					
6	-0.905831	0.798161	0.108885	0.057728	0.026811	0.008414				
7	-0.847462	0.787421	0.100687	0.058840	0.032634	0.015498	0.004920			
8	-0.795816	0.780196	0.092560	0.058012	0.035894	0.020097	0.009806	0.003435		
9	-0.760019	0.776119	0.086147	0.055968	0.036836	0.023025	0.013578	0.006235	0.002091	
10	-0.726530	0.774120	0.079968	0.053882	0.037061	0.024659	0.015764	0.008925	0.004263	0.001358

Таблица П2.26.

Полунормальное распределение (модуля многомерного нормального вектора $\mathbf{m}=1$). Известен параметр сдвига.
Коэффициенты масштабного параметра.

k	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
2	-0.674718	0.674718								
3	-0.704142	0.473184	0.230958							
4	-0.722413	0.363462	0.248570	0.110382						
5	-0.735740	0.294563	0.233732	0.146226	0.061220					
6	-0.744850	0.246291	0.212493	0.155010	0.093037	0.038019				
7	-0.751695	0.211079	0.192007	0.152944	0.107565	0.062886	0.025213			
8	-0.757243	0.181841	0.173142	0.146711	0.113487	0.077846	0.045357	0.018858		
9	-0.760482	0.162097	0.158149	0.138578	0.113130	0.085305	0.058454	0.032049	0.012720	
10	-0.763981	0.145625	0.145153	0.130952	0.111389	0.088885	0.065966	0.043083	0.023880	0.009048

Таблица П2.27.

Полунормальное распределение (модуля многомерного нормального вектора $\mathbf{m}=1$). Неизвестны оба параметра.
Коэффициенты параметра сдвига.

k	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9
3	2.331657	-1.331658							
4	2.062342	-0.682096	-0.380245						
5	1.890308	-0.404082	-0.327990	-0.158235					
6	1.767838	-0.256992	-0.254186	-0.176657	-0.080004				
7	1.676566	-0.170900	-0.194988	-0.161235	-0.104130	-0.045314			
8	1.599207	-0.113820	-0.147829	-0.136952	-0.105326	-0.066159	-0.029122		
9	1.547135	-0.080680	-0.116946	-0.116452	-0.099067	-0.073097	-0.042943	-0.017951	
10	1.502026	-0.057287	-0.092729	-0.098588	-0.089717	-0.072285	-0.050289	-0.029411	-0.011722

Таблица П2.28.

Полунормальное распределение (модуля многомерного нормального вектора $\mathbf{m}=1$). Неизвестны оба параметра.
Коэффициенты масштабного параметра.

k	υ_1	υ_2	υ_3	υ_4	υ_5	υ_6	υ_7	υ_8	υ_9
3	-1.168633	1.168633							
4	-1.126402	0.741325	0.385076						
5	-1.096213	0.531031	0.387542	0.177640					
6	-1.070483	0.403913	0.344340	0.224620	0.097610				
7	-1.049187	0.320471	0.299515	0.228764	0.141160	0.059275			
8	-1.029147	0.259331	0.258654	0.217193	0.157603	0.095455	0.040911		
9	-1.014471	0.219505	0.227513	0.201689	0.160643	0.114043	0.064707	0.026372	
10	-1.001894	0.188919	0.201795	0.186708	0.157427	0.121190	0.081502	0.046349	0.018003

Таблица П2.29.

Распределение Рэля. Известен масштабный параметр.
Коэффициенты параметра сдвига.

k	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
2	-1.785300	1.000000								
3	-1.553687	0.852000	0.148000							
4	-1.422130	0.762609	0.182889	0.054501						
5	-1.333035	0.703070	0.188847	0.081988	0.026095					
6	-1.267274	0.659900	0.186468	0.095575	0.044327	0.013731				
7	-1.217782	0.626298	0.181093	0.101443	0.055937	0.026611	0.008619			
8	-1.176753	0.600017	0.175329	0.103670	0.062798	0.035425	0.017393	0.005368		
9	-1.143445	0.578335	0.169358	0.104005	0.066779	0.041930	0.023958	0.011791	0.003844	
10	-1.114384	0.560558	0.163911	0.103050	0.069250	0.045989	0.029213	0.016946	0.008697	0.002387

Таблица П2.30.

Распределение Рэля. Известен параметр сдвига.
Коэффициенты масштабного параметра.

k	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
2	-0.560130	0.560130								
3	-0.577885	0.373401	0.204483							
4	-0.588607	0.274654	0.213152	0.100801						
5	-0.595888	0.213808	0.194262	0.129885	0.057933					
6	-0.601818	0.173390	0.172322	0.135237	0.085201	0.035668				
7	-0.605079	0.143669	0.151606	0.130333	0.096385	0.058668	0.024418			
8	-0.608350	0.122011	0.134286	0.122610	0.099289	0.070765	0.042496	0.016894		
9	-0.610375	0.105118	0.119343	0.113767	0.097769	0.076721	0.053463	0.031503	0.012691	
10	-0.612741	0.092122	0.107007	0.105288	0.094530	0.078586	0.060214	0.041423	0.024598	0.008974

Таблица П2.31.

Распределение Рэлея. Неизвестны оба параметра.
Коэффициенты параметра сдвига.

k	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9
3	2.661347	-1.661348							
4	2.283366	-0.738014	-0.545353						
5	2.032754	-0.340908	-0.443225	-0.248621					
6	1.854648	-0.134459	-0.319414	-0.268175	-0.132599				
7	1.715171	-0.013414	-0.217654	-0.233479	-0.170377	-0.080246			
8	1.606495	0.060914	-0.142937	-0.190203	-0.168405	-0.114788	-0.051076		
9	1.516663	0.108900	-0.086342	-0.149020	-0.151609	-0.123065	-0.080214	-0.035315	
10	1.443694	0.140822	-0.045027	-0.113795	-0.131115	-0.119458	-0.092112	-0.059004	-0.024005

Таблица П2.32.

Распределение Рэлея. Неизвестны оба параметра.
Коэффициенты масштабного параметра.

k	υ_1	υ_2	υ_3	υ_4	υ_5	υ_6	υ_7	υ_8	υ_9
3	-1.164550	1.164550							
4	-1.069352	0.647552	0.421800						
5	-0.997485	0.397405	0.393997	0.206083					
6	-0.942770	0.253242	0.327466	0.246594	0.115468				
7	-0.894145	0.159722	0.262032	0.237659	0.161759	0.072973			
8	-0.855301	0.097229	0.209566	0.214999	0.173214	0.112327	0.047966		
9	-0.820615	0.052873	0.166468	0.188727	0.169259	0.128579	0.080463	0.034246	
10	-0.792488	0.020719	0.132878	0.164257	0.158925	0.133411	0.097864	0.060752	0.023683

Таблица П2.33.

Распределение Максвелла. Известен масштабный параметр.
Коэффициенты параметра сдвига.

k	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
2	-2.045100	1.000000								
3	-1.832302	0.818571	0.181429							
4	-1.713140	0.705480	0.224522	0.069998						
5	-1.633745	0.627780	0.231934	0.106264	0.034022					
6	-1.576747	0.570246	0.227929	0.123635	0.059058	0.019132				
7	-1.534117	0.525447	0.220090	0.131216	0.074684	0.036546	0.012016			
8	-1.499676	0.489580	0.211538	0.134013	0.084183	0.048637	0.024022	0.008027		
9	-1.472579	0.459795	0.202827	0.133917	0.089489	0.057481	0.033793	0.016748	0.005950	
10	-1.448280	0.434739	0.194684	0.132557	0.092735	0.063400	0.041253	0.024378	0.012070	0.004184

Таблица П2.34.

Распределение Максвелла. Известен параметр сдвига.
Коэффициенты масштабного параметра.

k	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
2	-0.488974	0.488974								
3	-0.501083	0.315026	0.186058							
4	-0.508340	0.225477	0.189297	0.093566						
5	-0.513416	0.171572	0.169164	0.118712	0.053967					
6	-0.516893	0.135790	0.147065	0.121084	0.078741	0.034214				
7	-0.519125	0.110515	0.127344	0.115003	0.087872	0.055030	0.023360			
8	-0.521039	0.092084	0.111005	0.106610	0.089403	0.065389	0.039856	0.016692		
9	-0.522084	0.077975	0.097134	0.097481	0.086834	0.069903	0.050045	0.029938	0.012773	
10	-0.523556	0.067099	0.085820	0.089131	0.083034	0.070974	0.055739	0.039149	0.023026	0.009583

Таблица П2.35.

Распределение Максвелла. Неизвестны оба параметра.
Коэффициенты параметра сдвига.

k	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9
3	2.948165	-1.948165							
4	2.471730	-0.772553	-0.699178						
5	2.155425	-0.275646	-0.543897	-0.335881					
6	1.925171	-0.021384	-0.363707	-0.351887	-0.188193				
7	1.748046	0.121461	-0.222060	-0.295297	-0.235147	-0.117003			
8	1.607809	0.206152	-0.118326	-0.228225	-0.226090	-0.163530	-0.077790		
9	1.492147	0.258618	-0.041663	-0.166015	-0.196621	-0.172592	-0.118251	-0.055623	
10	1.396352	0.291189	0.014351	-0.113848	-0.162943	-0.163283	-0.133699	-0.088015	-0.040105

Таблица П2.36.

Распределение Максвелла. Неизвестны оба параметра.
Коэффициенты масштабного параметра.

k	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8	ν_9
3	-1.162251	1.162251							
4	-1.031002	0.582016	0.448986						
5	-0.935057	0.310685	0.397958	0.226414					
6	-0.859317	0.158119	0.309081	0.260628	0.131489				
7	-0.796939	0.064290	0.230280	0.241169	0.177101	0.084100			
8	-0.745647	0.003592	0.168263	0.208317	0.183190	0.125062	0.057223		
9	-0.701050	-0.037887	0.119233	0.173508	0.172556	0.140152	0.091675	0.041813	
10	-0.663969	-0.066634	0.081618	0.142640	0.156284	0.141227	0.109148	0.069107	0.030580

Таблица П2.37.

Распределение модуля многомерного нормального вектора $\mathbf{m}=4$. Известен масштабный параметр.

Коэффициенты параметра сдвига.

k	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
2	-2.275900	1.000000								
3	-2.079245	0.793096	0.206905							
4	-1.969454	0.664061	0.253514	0.082425						
5	-1.897094	0.573286	0.258796	0.125967	0.041951					
6	-1.846131	0.507296	0.251132	0.144789	0.072391	0.024391				
7	-1.806249	0.454964	0.239105	0.153042	0.091075	0.046222	0.015591			
8	-1.778109	0.415517	0.227011	0.154243	0.101402	0.060965	0.030364	0.010498		
9	-1.762589	0.390168	0.213688	0.149549	0.104567	0.070595	0.042599	0.021323	0.007513	
10	-1.745787	0.366985	0.205367	0.147642	0.106566	0.074620	0.049350	0.029640	0.014908	0.004921

Таблица П2.38.

Распределение модуля многомерного нормального вектора $\mathbf{m}=4$. Известен параметр сдвига.

Коэффициенты масштабного параметра.

k	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
2	-0.439387	0.439387								
3	-0.448175	0.276311	0.171864							
4	-0.453622	0.194246	0.171946	0.087430						
5	-0.457339	0.144435	0.150906	0.110297	0.051701					
6	-0.459793	0.112319	0.129212	0.110423	0.074442	0.033397				
7	-0.461838	0.089556	0.109943	0.103841	0.082099	0.053230	0.023169			
8	-0.462865	0.073889	0.094804	0.094938	0.082295	0.062122	0.038342	0.016476		
9	-0.463164	0.064637	0.082987	0.084795	0.077421	0.064884	0.047522	0.028674	0.012245	
10	-0.463799	0.056671	0.074680	0.078549	0.073874	0.063712	0.050561	0.035900	0.021265	0.008587

Таблица П2.39.

Распределение модуля многомерного нормального вектора $\mathbf{m}=4$. Неизвестны оба параметра.
Коэффициенты параметра сдвига.

k	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9
3	3.208013	-2.208014							
4	2.640431	-0.798451	-0.841980						
5	2.260663	-0.207623	-0.629040	-0.423999					
6	1.984236	0.083283	-0.390752	-0.430251	-0.246516				
7	1.768356	0.244385	-0.208197	-0.345077	-0.301100	-0.158367			
8	1.605500	0.330209	-0.082308	-0.253860	-0.279672	-0.213653	-0.106217		
9	1.504315	0.367129	0.000492	-0.173692	-0.238350	-0.224160	-0.159110	-0.076624	
10	1.408522	0.394064	0.055239	-0.117720	-0.192360	-0.204507	-0.173583	-0.116742	-0.052913

Таблица П2.40.

Распределение модуля многомерного нормального вектора $\mathbf{m}=4$. Неизвестны оба параметра.
Коэффициенты масштабного параметра.

k	υ_1	υ_2	υ_3	υ_4	υ_5	υ_6	υ_7	υ_8	υ_9
3	-1.161440	1.161440							
4	-1.003512	0.534141	0.469371						
5	-0.889453	0.245860	0.397981	0.245612					
6	-0.800019	0.090919	0.290089	0.272268	0.146744				
7	-0.727138	-0.002923	0.199994	0.241469	0.192289	0.096309			
8	-0.669241	-0.058038	0.133035	0.199797	0.191573	0.137234	0.065640		
9	-0.632108	-0.087054	0.084567	0.157869	0.175279	0.151345	0.102368	0.047734	
10	-0.596600	-0.108087	0.052929	0.128473	0.152928	0.145412	0.116408	0.075410	0.033128

Таблица П2.41.

Распределение модуля многомерного нормального вектора $\mathbf{m}=5$. Известен масштабный параметр.

Коэффициенты параметра сдвига.

k	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
2	-2.485100	1.000000								
3	-2.300468	0.772891	0.227109							
4	-2.198605	0.631399	0.275211	0.093391						
5	-2.134295	0.535411	0.277013	0.139695	0.047881					
6	-2.088460	0.465014	0.266221	0.159932	0.081293	0.027540				
7	-2.056818	0.412611	0.250599	0.166571	0.101071	0.051540	0.017608			
8	-2.027428	0.365771	0.233580	0.167282	0.114075	0.070076	0.036385	0.012832		
9	-2.009192	0.336157	0.222346	0.165331	0.117443	0.078976	0.047627	0.023958	0.008162	
10	-1.989013	0.303763	0.207996	0.161449	0.121758	0.088132	0.058475	0.035287	0.017676	0.005463

Таблица П2.42.

Распределение модуля многомерного нормального вектора $\mathbf{m}=5$. Известен параметр сдвига.

Коэффициенты масштабного параметра.

k	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
2	-0.402334	0.402334								
3	-0.409478	0.246911	0.162567							
4	-0.413313	0.170656	0.159267	0.083390						
5	-0.416283	0.126919	0.139023	0.101898	0.048443					
6	-0.418665	0.096308	0.120929	0.105386	0.065885	0.030157				
7	-0.419907	0.074246	0.101784	0.101234	0.078549	0.043504	0.020590			
8	-0.419328	0.062137	0.082904	0.085419	0.075710	0.056226	0.040790	0.016141		
9	-0.423326	0.052049	0.073398	0.080790	0.078806	0.064745	0.041721	0.021729	0.010087	
10	-0.424170	0.045923	0.063263	0.070189	0.071398	0.064225	0.052255	0.031881	0.016984	0.008052

Таблица П2.43.

Распределение модуля многомерного нормального вектора $\mathbf{m}=5$. Неизвестны оба параметра.

Коэффициенты параметра сдвига.

k	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9
3	3.428556	-2.428555							
4	2.783602	-0.803275	-0.980326						
5	2.355620	-0.151390	-0.705993	-0.498237					
6	2.043555	0.164766	-0.435469	-0.490363	-0.282489				
7	1.782190	0.373138	-0.227676	-0.414154	-0.334708	-0.178790			
8	1.607576	0.434028	-0.045189	-0.267969	-0.318367	-0.272797	-0.137282		
9	1.463982	0.478547	0.071690	-0.194371	-0.298178	-0.269695	-0.169329	-0.082646	
10	1.346885	0.493328	0.134505	-0.093007	-0.223633	-0.256662	-0.210475	-0.128142	-0.062799

Таблица П2.44.

Распределение модуля многомерного нормального вектора $\mathbf{m}=5$. Неизвестны оба параметра.

Коэффициенты масштабного параметра.

k	υ_1	υ_2	υ_3	υ_4	υ_5	υ_6	υ_7	υ_8	υ_9
3	-1.157006	1.157006							
4	-0.979843	0.491271	0.488572						
5	-0.853685	0.202044	0.395791	0.255850					
6	-0.759257	0.051947	0.287701	0.271183	0.148425				
7	-0.674108	-0.054899	0.196837	0.252455	0.184050	0.095665			
8	-0.611964	-0.099096	0.104368	0.188077	0.189726	0.155181	0.073708		
9	-0.567693	-0.129183	0.050442	0.161089	0.190971	0.155890	0.093411	0.045073	
10	-0.525385	-0.145992	0.013136	0.110849	0.159084	0.161123	0.121158	0.071338	0.034689

Таблица П2.45.

Распределение модуля многомерного нормального вектора $\mathbf{m}=6$. Известен масштабный параметр.

Коэффициенты параметра сдвига.

k	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
2	-2.678800	1.000000								
3	-2.504070	0.756093	0.243907							
4	-2.407327	0.603953	0.292330	0.103716						
5	-2.349550	0.504136	0.290268	0.152131	0.053465					
6	-2.307831	0.431536	0.276404	0.172904	0.088584	0.030571				
7	-2.273124	0.371933	0.260207	0.181441	0.111390	0.055441	0.019589			
8	-2.253708	0.332680	0.238885	0.178164	0.123086	0.075590	0.038441	0.013154		
9	-2.229712	0.290285	0.222084	0.178534	0.132917	0.089294	0.052712	0.025963	0.008210	
10	-2.218104	0.266342	0.206559	0.169610	0.131952	0.096028	0.064964	0.038944	0.018902	0.006699

Таблица П2.46.

Распределение модуля многомерного нормального вектора $\mathbf{m}=6$. Известен параметр сдвига.

Коэффициенты масштабного параметра.

k	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
2	-0.373301	0.373301								
3	-0.378908	0.227142	0.151766							
4	-0.382431	0.154543	0.148387	0.079501						
5	-0.384266	0.113395	0.127634	0.096563	0.046674					
6	-0.385703	0.086889	0.108214	0.096249	0.064970	0.029381				
7	-0.386902	0.067287	0.091623	0.090440	0.072185	0.045287	0.020081			
8	-0.387581	0.055629	0.077252	0.080742	0.071583	0.054390	0.033664	0.014321		
9	-0.388468	0.043996	0.065364	0.073979	0.070844	0.058677	0.041646	0.024342	0.009621	
10	-0.388706	0.037958	0.056986	0.065437	0.064996	0.057698	0.046346	0.032678	0.018743	0.007865

Таблица П2.47.

Распределение модуля многомерного нормального вектора $\mathbf{m}=6$. Неизвестны оба параметра.
Коэффициенты параметра сдвига.

k	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9
3	3.659356	-2.659357							
4	2.922188	-0.817554	-1.104634						
5	2.446899	-0.098970	-0.769450	-0.578480					
6	2.102723	0.242718	-0.448040	-0.558472	-0.338930				
7	1.816912	0.430926	-0.200293	-0.437227	-0.394119	-0.216198			
8	1.638730	0.512081	-0.030070	-0.302286	-0.371445	-0.295865	-0.151146		
9	1.436076	0.570453	0.101498	-0.187145	-0.310383	-0.299983	-0.211564	-0.098952	
10	1.321730	0.581665	0.177527	-0.088641	-0.231857	-0.274560	-0.243369	-0.164519	-0.077975

Таблица П2.48.

Распределение модуля многомерного нормального вектора $\mathbf{m}=6$. Неизвестны оба параметра.
Коэффициенты масштабного параметра.

k	υ_1	υ_2	υ_3	υ_4	υ_5	υ_6	υ_7	υ_8	υ_9
3	-1.159418	1.159418							
4	-0.962991	0.461044	0.501947						
5	-0.826866	0.165665	0.392237	0.268964					
6	-0.724137	0.014596	0.269060	0.280374	0.160107				
7	-0.635680	-0.075103	0.167934	0.241349	0.197772	0.103728			
8	-0.579512	-0.121221	0.092397	0.188743	0.198355	0.148336	0.072902		
9	-0.513874	-0.156239	0.034550	0.143544	0.179251	0.158179	0.106528	0.048061	
10	-0.475806	-0.169111	-0.003569	0.099451	0.147822	0.153070	0.127277	0.082693	0.038174

Таблица П2.49.

Распределение модуля многомерного нормального вектора $\mathbf{m}=7$. Известен масштабный параметр.

Коэффициенты параметра сдвига.

k	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
2	-2.859200	1.000000								
3	-2.692969	0.741957	0.258043							
4	-2.601920	0.582543	0.305334	0.112123						
5	-2.546329	0.477678	0.299610	0.163804	0.058909					
6	-2.447666	0.426197	0.364707	0.081822	0.094079	0.033195				
7	-2.476063	0.344170	0.264487	0.191800	0.119698	0.059657	0.020189			
8	-2.459507	0.304208	0.238013	0.184314	0.131065	0.083075	0.043851	0.015474		
9	-2.438386	0.262153	0.219773	0.184243	0.141231	0.097046	0.057580	0.028113	0.009860	
10	-2.433102	0.231930	0.200089	0.174337	0.140724	0.081104	0.128340	0.016117	0.020574	0.006784

Таблица П2.50.

Распределение модуля многомерного нормального вектора $\mathbf{m}=7$. Известен параметр сдвига.

Коэффициенты масштабного параметра.

k	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
2	-0.349748	0.349748								
3	-0.354431	0.210235	0.144196							
4	-0.357313	0.141525	0.139620	0.076167						
5	-0.359002	0.102526	0.118920	0.092275	0.045281					
6	-0.365230	0.099219	0.091696	0.081969	0.063340	0.029005				
7	-0.361113	0.060279	0.084923	0.085733	0.068935	0.042872	0.018371			
8	-0.361578	0.049440	0.070259	0.074587	0.067219	0.052174	0.033348	0.014550		
9	-0.362132	0.038852	0.059325	0.068513	0.066749	0.056019	0.039850	0.023096	0.009728	
10	-0.361971	0.031963	0.050120	0.059982	0.061270	0.047191	0.067379	0.019439	0.017593	0.007034

Таблица П2.51.

Распределение модуля многомерного нормального вектора $\mathbf{m}=7$. Неизвестны оба параметра.
Коэффициенты параметра сдвига.

k	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9
3	3.861346	-2.861345							
4	3.048389	-0.824273	-1.224115						
5	2.522790	-0.037267	-0.828744	-0.656778					
6	1.729003	1.322780	-1.120450	-0.574857	-0.356476				
7	1.841224	0.512109	-0.193464	-0.481650	-0.439231	-0.238989			
8	1.649642	0.589102	0.007825	-0.309505	-0.408755	-0.344812	-0.183497		
9	1.431118	0.642106	0.146870	-0.184021	-0.338081	-0.338415	-0.241093	-0.118483	
10	1.292343	0.655069	0.238034	-0.070008	-0.282641	-0.298439	-0.261377	-0.186375	-0.086607

Таблица П2.52.

Распределение модуля многомерного нормального вектора $\mathbf{m}=7$. Неизвестны оба параметра.
Коэффициенты масштабного параметра.

k	υ_1	υ_2	υ_3	υ_4	υ_5	υ_6	υ_7	υ_8	υ_9
3	-1.158346	1.158345							
4	-0.947702	0.434144	0.513559						
5	-0.803161	0.132299	0.389795	0.281066					
6	-0.532265	-0.391423	0.491191	0.273295	0.159201				
7	-0.604611	-0.100007	0.155595	0.242864	0.201485	0.104673			
8	-0.547034	-0.142748	0.071758	0.179129	0.199971	0.158025	0.080899		
9	-0.479401	-0.173202	0.015327	0.133389	0.178449	0.162400	0.110403	0.052634	
10	-0.435828	-0.186996	-0.026179	0.086611	0.149498	0.175405	0.114049	0.085056	0.038384

Таблица П2.53.

Распределение модуля многомерного нормального вектора $\mathbf{m}=8$. Известен масштабный параметр.

Коэффициенты параметра сдвига.

k	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
2	-3.028900	1.000000								
3	-2.868365	0.729542	0.270458							
4	-2.785714	0.566605	0.314663	0.118731						
5	-2.733808	0.459868	0.306510	0.170574	0.063048					
6	-2.695790	0.381429	0.286223	0.188151	0.104579	0.039618				
7	-2.669615	0.322863	0.260739	0.196621	0.128079	0.067392	0.024306			
8	-2.650425	0.279572	0.237666	0.191931	0.140007	0.088620	0.045762	0.016441		
9	-2.634598	0.244288	0.212002	0.180628	0.143953	0.105031	0.066781	0.035157	0.012159	
10	-2.619368	0.218427	0.194892	0.167614	0.174833	0.084040	0.077932	0.048520	0.024326	0.009414

Таблица П2.54.

Распределение модуля многомерного нормального вектора $\mathbf{m}=8$. Известен параметр сдвига.

Коэффициенты масштабного параметра.

k	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
2	-0.330153	0.330153								
3	-0.334295	0.196328	0.137967							
4	-0.336413	0.131726	0.131893	0.072794						
5	-0.337778	0.095115	0.112123	0.087129	0.043411					
6	-0.338888	0.071186	0.093597	0.084278	0.060454	0.029373				
7	-0.339557	0.055019	0.077352	0.079310	0.065724	0.042863	0.019289			
8	-0.339994	0.044059	0.064917	0.070836	0.065044	0.050140	0.031365	0.013632		
9	-0.340608	0.035845	0.053571	0.061071	0.060685	0.053402	0.040326	0.025097	0.010611	
10	-0.341435	0.030383	0.046560	0.057523	0.058450	0.047088	0.043263	0.031391	0.018497	0.008281

Таблица П2.55.

Распределение модуля многомерного нормального вектора $\mathbf{m}=8$. Неизвестны оба параметра.

Коэффициенты параметра сдвига.

k	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9
3	4.046716	-3.046718							
4	3.176717	-0.839354	-1.337362						
5	2.609637	-0.000179	-0.883030	-0.726428					
6	2.192860	0.392291	-0.451772	-0.675612	-0.457767				
7	1.881918	0.580025	-0.161547	-0.506653	-0.502989	-0.290753			
8	1.646565	0.663589	0.042334	-0.327583	-0.447786	-0.377964	-0.199155		
9	1.460026	0.690438	0.192252	-0.155180	-0.347461	-0.384486	-0.301682	-0.153908	
10	1.314122	0.690312	0.160334	0.205687	-0.371968	-0.334950	-0.319007	-0.228338	-0.116189

Таблица П2.56.

Распределение модуля многомерного нормального вектора $\mathbf{m}=8$. Неизвестны оба параметра.

Коэффициенты масштабного параметра.

k	υ_1	υ_2	υ_3	υ_4	υ_5	υ_6	υ_7	υ_8	υ_9
3	-1.156469	1.156469							
4	-0.936963	0.414263	0.522700						
5	-0.786364	0.112184	0.385398	0.288783					
6	-0.671948	-0.039346	0.237379	0.289411	0.184504				
7	-0.584000	-0.119600	0.134165	0.237762	0.213657	0.118017			
8	-0.515763	-0.160700	0.056443	0.176421	0.202385	0.159871	0.081344		
9	-0.461451	-0.181597	-0.004412	0.113540	0.171750	0.171285	0.127852	0.063033	
10	-0.418305	-0.189137	0.002779	-0.011779	0.174091	0.157627	0.140312	0.096460	0.047952

Таблица П2.57.

Распределение модуля многомерного нормального вектора $\mathbf{m}=9$. Известен масштабный параметр.

Коэффициенты параметра сдвига.

k	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
2	-3.189600	1.000000								
3	-2.817506	0.966994	0.033006							
4	-2.626830	0.917130	0.078817	0.004054						
5	-2.500927	0.870043	0.112901	0.016540	0.000515					
6	-2.392547	0.818473	0.142571	0.033552	0.005449	-0.000045				
7	-2.329834	0.783888	0.157308	0.045708	0.011501	0.001762	-0.000167			
8	-2.272209	0.748379	0.169912	0.057663	0.018735	0.004862	0.000562	-0.000113		
9	-2.221608	0.715909	0.180274	0.067517	0.025252	0.008602	0.002337	0.000211	-0.000102	
10	-2.187397	0.689109	0.184025	0.075979	0.032274	0.013073	0.004485	0.001134	0.000012	-0.000091

Таблица П2.58.

Распределение модуля многомерного нормального вектора $\mathbf{m}=9$. Известен параметр сдвига.

Коэффициенты масштабного параметра.

k	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
2	-0.313519	0.313519								
3	-0.317299	0.184515	0.132784							
4	-0.318861	0.123458	0.125531	0.069871						
5	-0.319886	0.089384	0.106453	0.082926	0.041123					
6	-0.321349	0.064414	0.086391	0.082210	0.059743	0.028591				
7	-0.321631	0.051658	0.072519	0.073975	0.061929	0.042199	0.019352			
8	-0.322007	0.041234	0.060533	0.065858	0.060594	0.048081	0.031323	0.014383		
9	-0.322398	0.033420	0.051517	0.058426	0.056153	0.048505	0.037835	0.024815	0.011728	
10	-0.322666	0.028238	0.043536	0.051002	0.052282	0.048367	0.040476	0.030443	0.019449	0.008873

Таблица П2.59.

Распределение модуля многомерного нормального вектора $\mathbf{m}=9$. Неизвестны оба параметра.

Коэффициенты параметра сдвига.

k	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9
3	4.217833	-3.217836							
4	4.099474	-2.422484	-0.676987						
5	3.887614	-1.791502	-0.874541	-0.221571					
6	3.650745	-1.268228	-0.889839	-0.392981	-0.099699				
7	3.473038	-0.954817	-0.815399	-0.458305	-0.195128	-0.049386			
8	3.298503	-0.694257	-0.727332	-0.481441	-0.258806	-0.108231	-0.028436		
9	3.130699	-0.496636	-0.633413	-0.465903	-0.289816	-0.157502	-0.068386	-0.019043	
10	3.013535	-0.341902	-0.549523	-0.456375	-0.315485	-0.192015	-0.102198	-0.043999	-0.012037

Таблица П2.60.

Распределение модуля многомерного нормального вектора $\mathbf{m}=9$. Неизвестны оба параметра.

Коэффициенты масштабного параметра.

k	υ_1	υ_2	υ_3	υ_4	υ_5	υ_6	υ_7	υ_8	υ_9
3	-1.153799	1.153799							
4	-0.927007	0.396600	0.530408						
5	-0.773495	0.097285	0.383923	0.292286					
6	-0.648693	-0.067170	0.227166	0.296605	0.192093				
7	-0.570524	-0.131720	0.121440	0.231293	0.221666	0.127845			
8	-0.501396	-0.170335	0.043559	0.165280	0.201752	0.168518	0.092622		
9	-0.444080	-0.190964	-0.009850	0.106906	0.161914	0.168016	0.134040	0.074018	
10	-0.402716	-0.196149	-0.047436	0.062075	0.127568	0.152036	0.141946	0.106829	0.055847

Таблица П2.61.

Распределение Лапласа. Известен масштабный параметр.

Коэффициенты параметра сдвига.

k	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
2	-1.841600	1.000000								
3	0.000000	0.500000	0.500000							
4	0.153723	-0.000141	0.635880	0.364261						
5	0.000000	0.000059	0.499941	0.499941	0.000059					
6	-0.042959	-0.000046	0.428192	0.571756	0.000125	-0.000027				
7	0.000000	-0.000076	0.000242	0.499834	0.499834	0.000242	-0.000076			
8	-0.018034	-0.000119	0.000041	0.451693	0.548434	-0.000020	0.000039	-0.000069		
9	0.000000	0.000156	-0.000060	0.000043	0.499861	0.499861	0.000043	-0.000060	0.000156	
10	-0.009764	-0.000067	-0.000055	0.000086	0.464625	0.535177	0.000010	0.000160	0.000028	0.000036

Таблица П2.62.

Распределение Лапласа. Известен параметр сдвига.

Коэффициенты масштабного параметра.

k	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
2	-0.543006	0.543006								
3	0.000000	-0.313755	0.313755							
4	0.058556	-0.091442	-0.267173	0.300059						
5	0.000000	-0.089579	-0.261490	0.261490	0.089579					
6	-0.020830	-0.089715	-0.262001	0.221856	0.112329	0.038361				
7	0.000000	-0.038686	-0.113417	-0.223865	0.223865	0.113417	0.038686			
8	-0.010112	-0.039114	-0.114626	-0.225531	0.192772	0.116919	0.059435	0.020257		
9	0.000000	-0.020751	-0.059528	-0.117747	-0.194904	0.194904	0.117747	0.059528	0.020751	
10	-0.006010	-0.020473	-0.060411	-0.118934	-0.196739	0.171596	0.114761	0.069361	0.035042	0.011807

Таблица П2.63.

Распределение Лапласа. Неизвестны оба параметра.
Коэффициенты параметра сдвига.

k	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9
3	0.500000	0.500000							
4	-0.014327	0.600212	0.414115						
5	0.000059	0.499941	0.499941	0.000059					
6	0.003811	0.439841	0.562729	-0.004705	-0.001676				
7	-0.000076	0.000242	0.499834	0.499834	0.000242	-0.000076			
8	0.000587	0.002108	0.455844	0.545057	-0.002129	-0.001033	-0.000434		
9	0.000156	-0.000060	0.000043	0.499861	0.499861	0.000043	-0.000060	0.000156	
10	0.000133	0.000535	0.001247	0.466574	0.533533	-0.001111	-0.000517	-0.000315	-0.000079

Таблица П2.64.

Распределение Лапласа. Неизвестны оба параметра.
Коэффициенты масштабного параметра.

k	υ_1	υ_2	υ_3	υ_4	υ_5	υ_6	υ_7	υ_8	υ_9
3	-0.313755	0.313755							
4	-0.092281	-0.232027	0.324308						
5	-0.089579	-0.261490	0.261490	0.089579					
6	-0.089794	-0.271163	0.210135	0.112427	0.038395				
7	-0.038686	-0.113417	-0.223865	0.223865	0.113417	0.038686			
8	-0.039120	-0.114648	-0.230140	0.187260	0.116940	0.059446	0.020261		
9	-0.020751	-0.059528	-0.117747	-0.194904	0.194904	0.117747	0.059528	0.020751	
10	-0.020474	-0.060414	-0.118942	-0.199543	0.168389	0.114767	0.069364	0.035044	0.011807

ПРИЛОЖЕНИЕ 3. Акты внедрения результатов