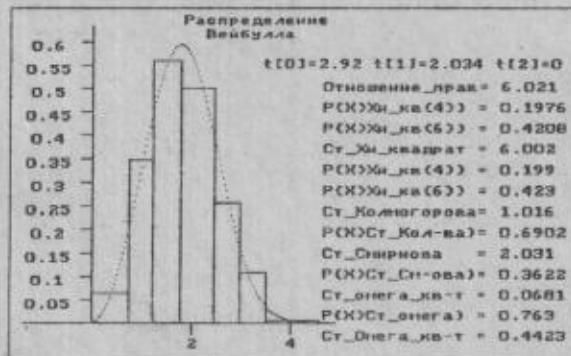




Б. Ю. ЛЕМЕШКО

*Статистический анализ
одномерных наблюдений случайных
величин*

Программная система



Новосибирск - 1995

Государственный комитет Российской Федерации
по высшему образованию

Новосибирский государственный технический университет

51
Л442

Б.Ю. Лемешко

*Статистический анализ
одномерных наблюдений случайных
величин*

Программная система

Новосибирск - 1995

УДК 519.2 (075.8)

Б.Ю. Лемешко. Статистический анализ одномерных наблюдений случайных величин: Программная система / Новосиб. гос. техн. ун-т. - Новосибирск, 1995. - 125 с.

ISBN 5-230-12098-3

Программная система предназначена для решения совокупности задач статистического анализа, решаемых при первичной обработке экспериментальных наблюдений.

В книге рассмотрены методы оценивания параметров непрерывных распределений по негруппированным, группированным, частично группированным и цензурированным выборкам, применение асимптотически оптимального группирования при оценивании параметров распределений и в критериях согласия, повышающее качество статистических выводов, кратко изложены основы алгоритмического обеспечения программной системы, приведены рекомендации и инструкции по использованию системы.

Книга и программная система предназначены для студентов, аспирантов, научных и инженерных работников, а также специалистов различных отраслей, использующих аппарат теории вероятностей и математической статистики для описания результатов наблюдений; рекомендуются в качестве учебного пособия для студентов, обучающихся по специальности "Прикладная математика".

Ил. 37, библ. 30 назв.

Рецензенты: В.В. Губарев, д-р техн. наук, проф.

В.И. Котюков, д-р техн. наук, проф.

А.А. Спектор, д-р техн. наук, проф.

Работа подготовлена на кафедре прикладной математики

ISBN 5-230-12098-3

© Новосибирский государственный
технический университет, 1995 г.

Предисловие

Предлагаемая работа охватывает достаточно узкую область математической статистики, связанную с первичной обработкой экспериментальных наблюдений, но с необходимостью которой сталкиваются многие исследователи.

Редкие исследования, в которых оперируют массовыми результатами наблюдений или моделирования, обходятся без статистической обработки методами, которым посвящена данная система.

Опыт её эксплуатации уже позволяет сделать вывод о том, что она существенно меняет взгляды пользователей на давно знакомые методы оценивания параметров, на использование критериев согласия и степень доверия к результатам их проверки, формирует мнение о предпочтительности тех или иных критериев согласия, показывает, что во многих случаях не так уж страшно наличие аномальных наблюдений. Некоторых пользователей применение программной системы и обескураживает, когда они убеждаются, что выводы, сделанные ими ранее на основании статистического анализа, оказываются не столь уж надёжными. Более того, совсем не бесспорными и неоднозначными оказываются и многие выводы, встречающиеся в литературных источниках, о наблюдаемых законах распределения. Работа с программной системой "Статистический анализ одномерных наблюдений случайных величин" даёт возможность закрепить теоретические знания, связать воедино все задачи, сопровождающие процесс выбора закона распределения вероятностей, отметить тонкие моменты анализа, ознакомиться с новыми для себя понятиями, оценить тот положительный эффект, который даёт применение группирования для получения рабочих оценок и асимптотически оптимального группирования в критериях согласия.

Программная система представляет собой достаточно надёжный инструмент, который со многих спишет проблему статистической обработки результатов их исследований, где-то позволит поднять качество и надёжность выводов, где-то сэкономить ресурсы, а где-то и изменить устоявшиеся стереотипы. И, главное, на мой взгляд, параллельно с полученными результатами анализа у пользователя будет формиро-

ваться та степень уверенности или уровень достоверия, с которыми он в дальнейшем будет использовать эти результаты в своей практике.

Теоретические аспекты, определяющие основные достоинства программной системы, излагаются во втором разделе. При первом знакомстве с материалом этого раздела следует иметь в виду, что законы распределения, которые в упоминаются в примерах, подробно описаны в третьем разделе.

Данное описание призвано, с одной стороны, служить руководством при эксплуатации программной системы, с другой - очерчивает фундамент, на котором она базируется и который может быть использован при решении соответствующих задач вне рамок данного программного обеспечения. Система продолжает совершенствоваться, будут появляться новые версии, намечены пути её дальнейшего развития, но изложенная здесь основа останется.

Работа носит достаточно специфический характер и предназначена служить описанием применения программной системы. Автор неставил свой целью отразить все вопросы, так или иначе связанные с решаемыми задачами. В тексте сделаны ссылки лишь на работы, предшествующие её созданию, на те источники, которые существенным образом повлияли на разработанное алгоритмическое и программное обеспечение или будут полезны пользователям системы для более глубокого понимания результатов её применения. Автор заранее приносит свои извинения и просит учесть, что отсутствие ссылок на многочисленные работы, связанные с излагаемым материалом, объясняется лишь характером работы и необходимой краткостью изложения.

Автор приносит свою благодарность профессору В.И. Денисову, под руководством и совместно с которым решались многие задачи, результаты которых в дальнейшем использовались при реализации программной системы, доценту Г.Г. Зачепе, с которым связано начало работ по обработке группированных данных в НГТУ, профессору Е.Б. Цою, также продолжающему заниматься вопросами анализа группированных данных. Автор признателен профессору В.В. Губареву, профессору А.А. Спектору за внимание к рукописи и ряд ценных замечаний.

Б.Ю. Лемешко

1. Введение

Назначение. Программная система предназначена для решения совокупности задач статистического анализа, решаемых при первичной обработке экспериментальных наблюдений. Она является мощным интеллектуальным средством автоматизации научных исследований, обеспечивающим решение таких нетривиальных задач, как выбор закона распределения, наиболее хорошо описывающего выборочные данные, определение вероятностных характеристик исследуемого объекта.

По ряду заложенных возможностей программное обеспечение системы не имеет аналогов среди программного обеспечения задач статистического анализа, использует совокупность оригинальных результатов. **Во-первых**, исходная выборка может быть негруппированной, группированной или частично группированной. **Во-вторых**, при проверке гипотез по критериям, предусматривающим группировку данных, используются полученные таблицы асимптотически оптимального группирования данных, обеспечивающие максимальную мощность критериев при близких альтернативных гипотезах. **В-третьих**, возможность группирования исходной выборки и дальнейшее оценивание параметров распределений по сгруппированным данным позволяют получать робастные оценки, менее чувствительные к наличию аномальных измерений. **В-четвертых**, применение программного комплекса при обработке результатов наблюдений даёт возможность группирования данных экспериментов в месте проведения исследований, что позволяет резко сократить объемы хранимых и передаваемых по каналам связи данных от удаленных источников без существенной потери информации о наблюдаемом законе распределения случайной величины.

Система позволяет находить оценки максимального правдоподобия параметров для большинства наиболее часто используемых в приложениях распределений, проверять гипотезы о согласии по критериям отношения правдоподобия, χ^2 Пирсона, Колмогорова, Смирнова и ω^2 Мизеса и принимать решение по их совокупности, вычислять вероятности для соответствующих распределений, определять различные вероятностные характеристики случайных

величин, строить графики функций плотности и гистограммы распределений, теоретические и эмпирические функции распределения, осуществлять печать графической информации, моделировать выборки в соответствии с законами распределений, включенными в комплекс.

Область применения. Система может использоваться в любой прикладной области, связанной с регистрацией наблюдений, а значит, и с проявлением случайных закономерностей: при метрологической аттестации измерительных приборов и обработке измерительной информации, при испытаниях на надежность, при контроле качества, при анализе результатов социологических обследований, при обработке наблюдений медицинских и биологических исследований, при математическом и имитационном моделировании сложных систем, анализе результатов моделирования и т.д.

Программное обеспечение может широко использоваться в учебном процессе при освоении дисциплин, связанных с обработкой информации и информационно-измерительными системами, теорией измерений, при освоении курсов теории вероятностей и математической статистики, теории случайных процессов, исследования операций, методов моделирования, статистической физики и т.д.

2. Теоретическое обоснование

2.1. Структура данных и методы оценивания

Система предназначена для обработки самого общего случая частично группированных выборок. Понятие частично группированной выборки объединяет по существу 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 [1], 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 -й интервал принадлежит к первому типу, если число n_i известно, но индивидуальные значения

x_j , $j = \overline{1, n_i}$, неизвестны;

б) i -й интервал принадлежит ко второму типу, если известно не только число n_i , но и все индивидуальные

значения x_j , $j = \overline{1, n_i}$.

В дальнейшем, когда речь будет идти о частично группированной выборке суммирование по интервалам первого и второго типов будем

обозначать соответственно через $(\sum_{(1)})$ и $(\sum_{(2)})$.

Наиболее часто на практике используются следующие методы вычисления оценок.

Метод моментов. Предполагается, что исходная выборка негруппированная. 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

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

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

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

Метод моментов приводит при определенных 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 [2]. 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 [3]

$$m_1 = m_1^*,$$

$$m_2 = m_2^* - \frac{1}{12} h^2,$$

$$m_3 = m_3^* - \frac{1}{4} m_1^* h^2,$$

$$m_4 = m_4^* - \frac{1}{2} m_1^* h^2 + \frac{7}{240} h^4,$$

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.

Целый ряд методов предполагает использование группированных выборок.

Метод минимума χ^2 [2]. 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 -й интервал.

Модифицированный метод минимума χ^2 . 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. 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. Минимизируемая статистика имеет вид

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

Мера расхождения Холдейна. 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 С.Р. Рао, дают состоятельные и асимптотически эффективные оценки [2]. Однако имеются и различия между этими методами, возникающие при учете введённой С.Р. Рао эффективности второго порядка [4]. 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.

В связи с вышесказанным в качестве основного метода оценивания параметров распределений выбран метод максимального правдоподобия.

Естественно, что описанные в данном пункте методы вычисления оценок параметров распределений не представляют собой полный перечень всех возможных методов. Вообще говоря, любая статистика, в том числе и из используемых в критериях согласия, может быть положена в основу метода оценивания.

2.2. Метод максимального правдоподобия

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

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

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 предполагается, что для всех $\theta \in \Omega$ $P_i(\theta) > 0$, $i = \overline{1, k}$. Для вычисления оценки макси-

мального правдоподобия (ОМП) дифференцируют функцию правдоподобия по θ и, приравнивая производные нулю, получают систему уравнений правдоподобия

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

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

Функция правдоподобия для частично группированной выборки имеет вид

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

система уравнений правдоподобия

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

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.

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:

a) $\lim_{N \rightarrow \infty} E[\sqrt{N}(\hat{\theta}_N - \theta)] = 0;$

б) $\lim_{N \rightarrow \infty} E\{\langle \sqrt{N}(\hat{\theta}_N - \theta) \rangle' [\sqrt{N}(\hat{\theta}_N - \theta)]\} = M^{-1}$.

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

в) 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} .

В случае достаточно больших N при определенных условиях регулярности для функции плотности ОМП существует практически всегда, состоятельна и асимптотически эффективна [2]. Условия существования, асимптотической эффективности и состоятельности по группированным и цензурированным выборкам рассматривались в работах Г. Куллэрфа [1] и Н.А. Бодина [5]. Г.Г. Зачепой [6] были получены условия существования ОМП основных параметров распределения Вейбулла и гамма-распределения. В наших работах [7-9] получены условия существования и единственности ОМП для параметров ряда непрерывных законов распределения случайных величин, а в [10] приведены сводные результаты.

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 группированных данных, ограничивали использование метода максимального правдоподобия.

2.3. Асимптотически оптимальное группирование

До настоящего времени 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. При этом различаются рекомендуемые значения минимального числа наблюдений, попавших в интервал: 5, 8, а то и не менее 10. Грубое правило, которым обычно пользуются, состоит в том, что ожидаемые частоты NP_i , где P_i — гипотетическая вероятность попадания в интервал, должны быть не меньше 5. Хотя в [11] говорится, что проведенные исследования показали: в случае унимодального распределения допускается уменьшение ожидаемых частот для одного или двух интервалов до 1 и даже ниже.

Всякая группировка данных по сравнению с негруппированной выборкой ведет к потере информации, понимаемой в общем широком смысле. Асимптотическая дисперсионная матрица ОМП по группированным наблюдениям определяется соотношением

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

где

$$M_{\Gamma}(\hat{\theta}) = \sum_{i=1}^s \frac{\nabla P_i(\hat{\theta}) \nabla^T P_i(\hat{\theta})}{P_i(\hat{\theta})},$$

— информационная матрица Фишера по группированным данным. 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. Эта задача сводится к максимизации информационного количества

Фишера о параметре по группированной выборке, т.е. к решению задачи

$$\max_{x_0 < x_1 < \dots < x_{k-1} < x_k} \sum_{i=1}^k \left(\frac{\partial \ln P_i(\theta)}{\partial \theta} \right)^2 \cdot P_i(\theta).$$

При оценивании вектора параметров мы имеем дело с информационной матрицей. В этом случае в качестве критерия оптимальности могут быть выбраны различные функционалы от асимптотической дисперсионной матрицы, но наиболее естественно минимизировать обобщенную асимптотическую дисперсию, т.е. решать задачу

$$\max_{x_0 < x_1 < \dots < x_{k-1} < x_k} \det M_\Gamma(\theta).$$

2.4. Связь критерия согласия с оптимальным группированием

В [11] 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, однако метода такого пока нет. Правило использования интервалов равной вероятности предложено Манном и Вальдом в 1942 г. При таком разбиении максимизируется энтропия, равная $-\sum_{i=1}^k P_i(\theta) \ln P_i(\theta)$ [12]. В этом случае максимизируется асимптотическая мощность критерия, если не требуется рассматривать какую-либо конкретную альтернативу.

Фишеровская информация служит мерой внутренней близости распределений случайных величин, и этот внутренний характер связан с мощностью различия между близкими значениями параметра [2]. Статистика редуцирует выборочные данные, и поэтому мощность различия с помощью статистики не больше, чем с помощью всей выборки. А это значит, если нужно выбирать между несколькими статистиками, следует предпочесть ту, для которой потери фишеровской информации минимальны.

Статистика критерия согласия χ^2 Пирсона вычисляется в соответствии с соотношением

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

и в пределе подчиняется χ^2 -распределению с $k-1$ степенью свободы, если верна нулевая гипотеза, и подчиняется нецентральному χ^2 -распределению с тем же числом степеней свободы и параметром нецентральности

$$\lambda = N \sum_{i=1}^k \frac{(P_i(\theta_1) - P_i(\theta))^2}{P_i(\theta)}.$$

если верна конкурирующая гипотеза и выборка соответствует распределению того же типа, но с параметром θ_1 (в общем случае векторным). Разлагая $P_i(\theta_1)$ в ряд Тейлора при малых $\Delta\theta = \theta_1 - \theta$ и пренебрегая членами высшего порядка, получаем

$$\begin{aligned} \lambda &\approx N \sum_{i=1}^k \frac{[P_i(\theta) + \nabla^T P_i(\theta) \delta\theta - P_i(\theta)]^2}{P_i(\theta)} = N \sum_{i=1}^k \frac{\delta\theta^T \nabla P_i(\theta) \nabla^T P_i(\theta) \delta\theta}{P_i(\theta)} = \\ &= N \delta\theta^T \left(\sum_{i=1}^k \frac{\nabla P_i(\theta) \nabla^T P_i(\theta)}{P_i(\theta)} \right) \delta\theta = N \delta\theta^T M_\Gamma(\theta) \delta\theta. \end{aligned}$$

Мощность критерия χ^2 Пирсона является неубывающей функцией 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 определенной, 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. Эти потери можно уменьшить, подбирая граничные точки так, чтобы $M_\Gamma(\theta)$ стремилась к информационной матрице по негруппированным данным $M(\theta)$, т.е. в данном случае приходим к

той же самой задаче асимптотически оптимального группирования, что и при оценивании параметров.

Мощность критерия согласия, использующих статистики, упомянутые выше, также зависит от информационной матрицы Фишера по группированным данным.

Так для отношения правдоподобия, 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. Обозначив $\theta_1 = \theta + \delta\theta$, разложим $P_i(\theta_1)$ в ряд Тейлора. Пренебрегая членами высшего порядка, имеем

$$-2 \ln l \approx 2N \sum_{i=1}^k (P_i(\theta) + \nabla^T P_i(\theta) \delta\theta) \ln \left(1 - \frac{\nabla^T P_i(\theta) \delta\theta}{P_i(\theta)} \right).$$

Далее, раскладывая $\ln(1+x)$ по формуле Тейлора и пренебрегая членами выше второго порядка, получаем

$$\begin{aligned} 2 \ln l &\approx 2N \sum_{i=1}^k (P_i(\theta) + \delta\theta^T \nabla P_i(\theta)) \left[\frac{\nabla^T P_i(\theta) \delta\theta}{P_i(\theta)} - \frac{\delta\theta^T \nabla P_i(\theta) \nabla^T P_i(\theta) \delta\theta}{2P_i^2(\theta)} \right] = \\ &\approx 2N \sum_{i=1}^k \left(\delta\theta^T \nabla P_i(\theta) + \frac{\delta\theta^T \nabla P_i(\theta) \nabla^T P_i(\theta) \delta\theta}{2P_i(\theta)} \right) = \\ &= N \delta\theta^T \left[\sum_{i=1}^k \frac{\nabla P_i(\theta) \nabla^T P_i(\theta)}{P_i(\theta)} \right] \delta\theta = N \delta\theta^T M_T(\theta) \delta\theta. \end{aligned}$$

Аналогично, в случае использования 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

$$\begin{aligned}\text{mod} \chi^2 &\approx N \sum_{i=1}^k \frac{[P_i(\theta) + \delta\theta^T \nabla P_i(\theta) - P_i(\theta)]^2}{P_i(\theta) + \delta\theta^T \nabla P_i(\theta)} = \\ &= N \sum_{i=1}^k \frac{\delta\theta^T \nabla P_i(\theta) \nabla^T P_i(\theta) \delta\theta}{P_i(\theta) [1 + \delta\theta^T \nabla P_i(\theta) / P_i(\theta)]}\end{aligned}$$

Используя разложение в ряд $(1+x)^{-1}$, имеем

$$\text{mod} \chi^2 \approx N \delta\theta^T \left[\sum_{i=1}^k \frac{\nabla P_i(\theta) \nabla^T P_i(\theta)}{P_i(\theta)} \right] \delta\theta = N \delta\theta^T M_r(\theta) \delta\theta.$$

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

$$\begin{aligned}H_D &\approx \arccos \sum_{i=1}^k P_i(\theta) \left[1 + \frac{1}{2} \frac{\delta\theta^T \nabla P_i(\theta)}{P_i(\theta)} - \frac{1}{8} \frac{\delta\theta^T \nabla P_i(\theta) \nabla^T P_i(\theta) \delta\theta}{P_i(\theta)} \right] = \\ &= \arccos \left[1 - \frac{1}{8} \delta\theta^T M_r(\theta) \delta\theta \right].\end{aligned}$$

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

$$H_{KL} \approx \sum_{i=1}^k P_i(\theta) \left[\frac{\nabla^T P_i(\theta) \delta\theta}{P_i(\theta)} - \frac{1}{2} \frac{\delta\theta^T \nabla P_i(\theta) \nabla^T P_i(\theta) \delta\theta}{P_i^2(\theta)} \right] = \frac{1}{2} \delta\theta^T M_\Gamma(\theta) \delta\theta.$$

Приемы группирования, в основном применяемые на практике: разбиение области, в которую попали выборочные значения случайной величины, на интервалы равной длины, или на интервалы равной вероятности, в общем случае весьма далеки от оптимального. Проведенные нами численные исследования по сравнению способов группировки данных показали несравненные преимущества асимптотически оптимального группирования.

Следует отметить, что информационная матрица Фишера зависит от неизвестных оцениваемых параметров. Однако для широкого ряда распределений при решении задач асимптотически оптимального группирования нам удалось получить граничные точки интервалов в виде, инвариантном относительно параметров распределений, и на их основе сформировать таблицы асимптотически оптимального группирования. Ранее в литературе были представлены только фрагмент таблицы для экспоненциального закона распределения и отдельно таблицы для математического ожидания и среднего квадратичного отклонения нормального распределения [1,13]. Нами эти результаты были уточнены и в дальнейшем в совокупности были получены таблицы для распределений экспоненциального, полунармального, Рэлея, Максвелла, модуля многомерного нормального вектора, Парето, Эрланга, Лапласа, нормального, логарифмически-нормальных (In и Ig), Коши, Вейбулла, распределений минимального и максимального значения, двойного показательного, гамма-распределения [14-26]. В общей сложности получено 54 таблицы оптимальных граничных точек и соответствующих вероятностей. Полностью они приведены в [10]. Эти таблицы могут использоваться как при оценивании, так и при проверке гипотез. Полученные таблицы используются в реализованном

программном обеспечении при проверке согласия по критерию χ^2 Пирсона и группировании данных при вычислении робастных оценок.

Для ряда 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 [14,23], распределения Джонсона. В этом случае формирование таблиц асимптотически оптимального группирования теряет смысл. Однако прогресс вычислительной техники настолько стремителен, что становится возможным решение задачи асимптотически оптимального группирования при конкретных значениях параметров в процессе проверки гипотез о согласии. Это является одним из направлений развития данного программного обеспечения.

2.5. Робастность

В статистике под робастностью понимают нечувствительность к малым отклонениям от предположений [27].

При решении задач статистического анализа и, в частности, при вычислении оценок параметров распределений чрезвычайно важное значение приобретает проблема наличия в выборке аномальных измерений. В практике решения таких задач широко известно, что наличие даже одного аномального наблюдения приводит к оценкам, которые совершенно не вяжутся с выборочными данными. Вообще говоря, наличие выбросов отражается на качестве всех оценок.

Понятно желание каждого исследователя, чтобы найденные оценки были как можно менее чувствительны к аномальным наблюдениям. Так как в противном случае, прежде чем переходить к оцениванию, приходится использовать процедуры исключения грубых ошибок измерений, что выливается в не совсем простую задачу. Наиболее ясна ситуация в случае нормального распределения, когда для отсева аномальных наблюдений используют правило "трех сигм". Но в случае произвольных распределений готовое правило найти труднее.

В данном случае следует подчеркнуть достоинство оценок, использующих группирование исходных выборочных данных, так как очевидно, что они менее чувствительны к случайным выбросам. Это было убедительно показано нами при решении задач оценивания параметров на моделируемых и реальных выборочных данных. Группирование выборки позволяет резко снизить влияние аномальных наблюдений, а иногда и совсем исключить влияние случайных выбросов, что бывает чрезвычайно важно, например, при передаче информации по каналам связи.

Чтобы показать, насколько существенно влияние аномальных измерений, приведем пример с нормальным распределением. Была сгенерирована выборка по нормальному закону с нулевым математическим ожиданием и единичной дисперсией, состоящая из 500 наблюдений. На рис. 2.1 представлены результаты оценивания параметров по этой выборке и проверки гипотез о согласии.

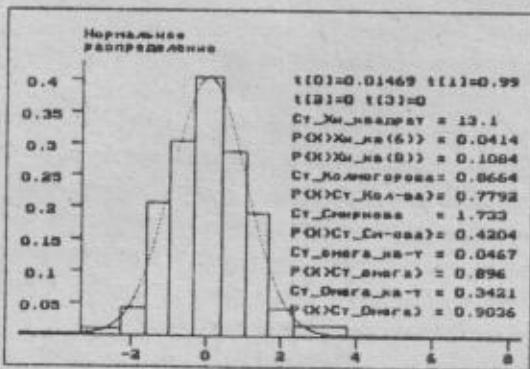


Рис. 2.1. Результаты анализа на исходной выборке

Затем в данной выборке увеличили первое наблюдение на 20 и снова проведен соответствующий анализ. Его результаты приведены на рис. 2.2 и 2.3. Как и следовало ожидать, наиболее существенно изменилась оценка среднеквадратичного отклонения. Согласие по всем критериям отвергается.

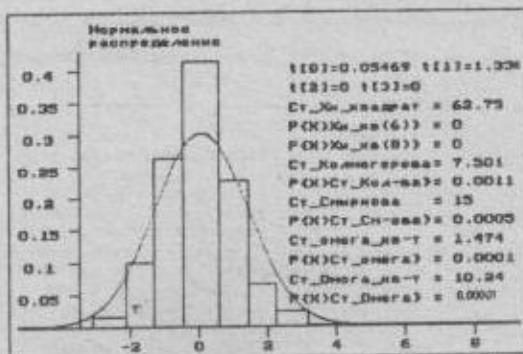


Рис. 2.2. Результаты анализа при наличии "аномального" наблюдения

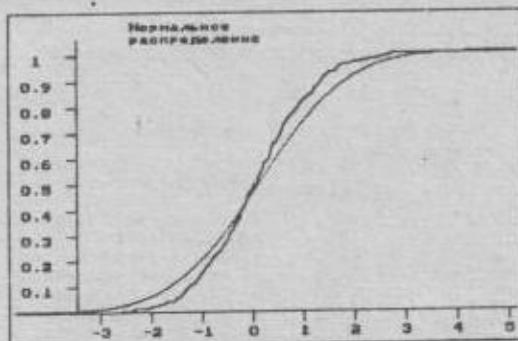


Рис. 2.3. Теоретическая и эмпирическая функции распределения при наличии "аномального" наблюдения

А далее осуществили группирование выборки с "аномальным" наблюдением, провели оценивание по группированной выборке и проверили гипотезы о согласии. Результаты представлены на рис. 2.4. Как видим, "случайная" ошибка в данных практически не повлияла на оценки параметров.

Посмотрим, как поведут себя оценки максимального правдоподобия параметров нормального распределения, когда на самом деле выборка принадлежит другому распределению, например распределению Коши. Распределение Коши - это распределение с "тяжёлыми"

хвостами. Смоделируем выборку по закону Коши с параметрами $\theta_0 = 1$, $\theta_1 = 0$. На рис. 2.5 представлены результаты моделирования.

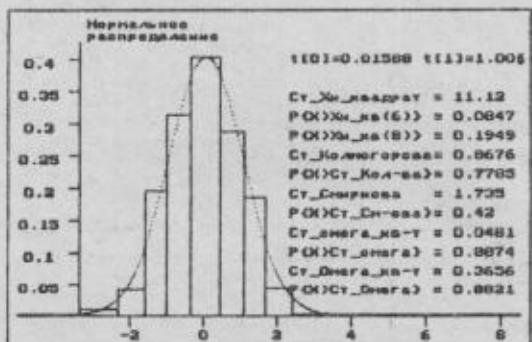


Рис. 2.4. Результаты робастного оценивания

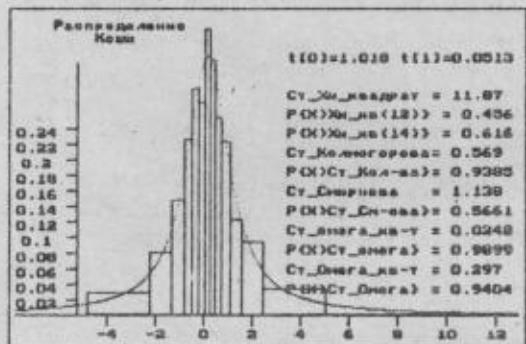


Рис. 2.5. Результаты моделирования распределения Коши

Предполагая, что на самом деле выборка принадлежит нормальному распределению, оценим его параметры. Результаты оценивания и проверки гипотез о согласии представлены на рис. 2.6, где пологая кривая - это теоретическая функция распределения, а ступенчатая - эмпирическая.

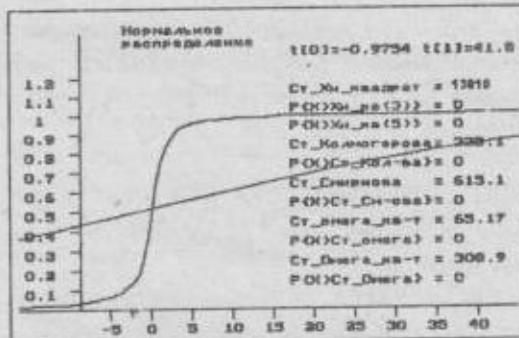


Рис. 2.6. Эмпирическая и теоретическая функции распределения нормального закона при вычислении оценок по негруппированной выборке

Теперь найдём робастные оценки. Для этого предварительно сгруппируем выборку, разбив её на интервалы равной частоты (равной вероятности), а затем по группированной выборке оценим параметры нормального распределения. В данном случае выборка разбивалась на 6 интервалов. Результаты оценивания и анализа представлены на рис. 2.7.

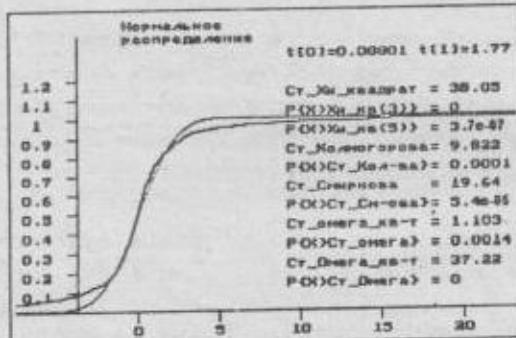


Рис. 2.7. Эмпирическая и теоретическая функции распределения нормального закона при вычислении оценок по группированной выборке

Естественно, что хотя о согласии найденного нормального закона с выборочными данными говорить не приходится, так как модели-

- ровался всё-таки закон распределения Коши, преимущество оценок, найденных по сгруппированной выборке, очевидно. Из рисунка видна близость эмпирической и теоретической функций распределения в середине и расхождение на хвостах.

Относительно использования группирования для получения робастных оценок необходимо обратить внимание исследователей на два существенных момента.

1. Выбор меньшего количества интервалов группирования позволяет получать более робастные оценки.
2. Разбиение выборки на интервалы равной вероятности (равной частоты) представляет собой также более робастную процедуру, т.е. является более предпочтительным, чем использование асимптотически оптимального группирования, если истинная функция распределения значительно отличается от предполагаемой.

Однако следует иметь в виду, что при меньшем количестве интервалов группирования мы проигрываем в качестве оценок. То же самое можно сказать об отказе от использования асимптотически оптимального группирования.

Таким образом, уменьшение количества интервалов при оценивании может быть оправдано лишь существенным засорением выборки аномальными измерениями, а отдача предпочтения разбиению на интервалы равной частоты может быть обоснована возможным резким отличием функции распределения от предполагаемой.

В любом случае с использованием данного программного обеспечения несложно совершить серию расчетов без использования группирования, с использованием группирования, изменяя количество интервалов и варианты группирования. Сравнение результатов позволит судить о степени засоренности выборки аномальными измерениями и предпочтительности выбора того или иного варианта группирования.

2.6. Проверка гипотез

Целью первичной обработки экспериментальных наблюдений обычно является выбор закона распределения, который наиболее хорошо описывает 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.

При проверке гипотез о согласии проверяется гипотеза вида H_0 : $f(x, \theta) = f(x, \hat{\theta})$. Проверка 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 распределения $g(S)$ соответствующих статистик при условии истинности гипотезы H_0 .

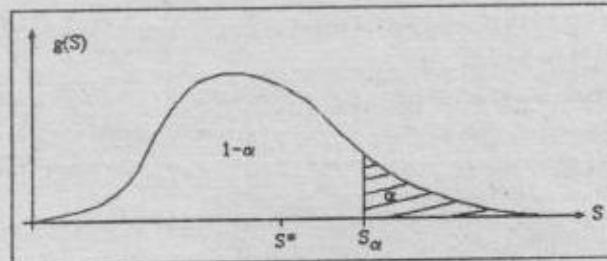


Рис. 2.8. Распределение статистики при истинной гипотезе H_0

Далее в принятой практике статистического анализа обычно сравнивают полученное значение статистики S^* с критическим для данного уровня значимости S_α и нулевую гипотезу отвергают, если $S^* > S_\alpha$ (см. рис. 2.8). Критическое значение S_α , определяемое из уравнения

$$\alpha = \int_{S_a}^{\infty} g(s) ds,$$

где α - задаваемый уровень значимости, обычно берётся из соответствующей статистической таблицы.

В такой ситуации остается за кадром величина вероятности превышения полученного значения статистики при истинности нулевой гипотезы, хотя именно она и позволяет судить о степени согласия, так как по существу представляет собой вероятность истинности нулевой гипотезы (рис. 2.9).

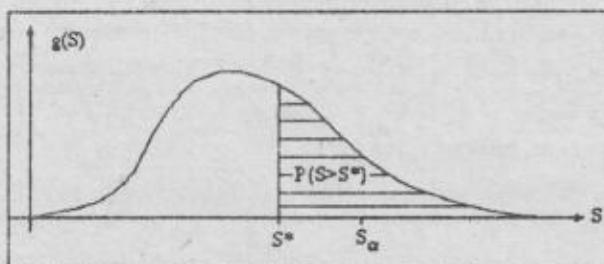


Рис. 2.9. Распределение статистики при истинной гипотезе H_0

В этой связи в программное обеспечение включён ряд критерев, при проверке которых в обязательном порядке отображается и соответствующая вероятность превышения полученного значения статистики при истинности нулевой гипотезы

$$P\{S > S^*\} = \int_{S^*}^{\infty} g(s) ds.$$

Гипотеза о согласии не отвергается, если $P\{S > S^*\} > \alpha$.

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. 2.10 $g(s|H_0)$ отображает плотность распределения статистики S при истинности гипотезы H_0 , а $g(s|H_1)$ - плотность распределения при справедливости гипотезы H_1 .

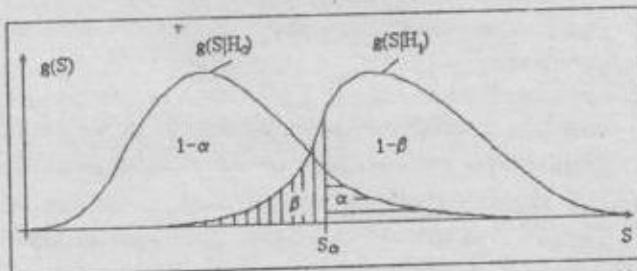


Рис. 2.10. Распределения статистик при справедливости гипотез H_0 и H_1

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 .

Когда мы говорим о том, что асимптотически оптимальное группирование обеспечивает максимальную мощность критерия согласия отношения правдоподобия и χ^2 Пирсона при близких конкурирующих гипотезах H_0 и H_1 , графически это означает, что на рис. 2.10 плотности $g(s|H_0)$ и $g(s|H_1)$ будут максимально "раздвинуты".

Ясно, что при проверке гипотез желательно пользоваться наиболее мощными критериями, а ещё лучше равномерно наиболее мощными критериями, которые являются наиболее мощными на всём множестве альтернативных гипотез. Но в критериях согласия альтернативная гипотеза обычно не выделяется. И, кроме того, совсем не очевидно, какой критерий согласия окажется наиболее мощным для конкретной ситуации (конкретной выборки, конкретной гипотезы H_0).

Большинство исследователей при решении задачи выбора распределения, наиболее хорошо описывающего их результаты, обычно пользуются одним из критериев согласия, опираясь на заданный уровень значимости (обычно 0.01, 0.05, 0.1), принимая гипотезу о согласии, если значение соответствующей статистики не превышает критического значения. При этом не учитывается, что различные критерии используют различные меры близости распределений, и принятие гипотезы о согласии по одному критерию совсем не означает, что распределения окажутся близкими в соответствии с другими критериями.

Так критерии Колмогорова и Смирнова измеряют расстояние между теоретическим (гипотетическим) и эмпирическим интегральными законами распределения в равномерной метрике, в критериях ω^2 и Ω^2 это расстояние измеряется в квадратичной метрике. В критерии χ^2 Пирсона в квадратичной метрике измеряется расстояние между теоретическим и эмпирическим дифференциальными законами распределения. Близкая мера используется в критерии отношения правдоподобия.

Наш опыт показывает, что в обязательном порядке необходимо использование ряда критериев, основанных на различных мерах близости. Это позволяет принимать решение по их совокупности.

2.7. Критерии проверки гипотез

При проверке гипотез о согласии используются 6 критериев: критерий χ^2 Пирсона, критерий отношения правдоподобия, критерий Колмогорова, критерий Смирнова и 2 критерия ω^2 Мизеса. Непараметрические критерии Колмогорова, Смирнова и ω^2 Мизеса применяются лишь в тех случаях, когда исходная выборка состоит из негруппированных наблюдений.

При использовании критерия χ^2 Пирсона, если предварительно оценивались все параметры распределения, то проверяется гипотеза о согласии с выбранным распределением, если же оценивался один из

параметров, то по существу проверяется гипотеза о равенстве данного параметра найденному значению.

Статистика χ^2 при истинной гипотезе H_0 в пределе подчиняется χ^2 -распределению с числом степеней свободы $k-1$, если по выборке не оценивались параметры, и с $k-p-1$, если по ней оценивалось p параметров закона распределения. Функция плотности χ^2 -распределения с числом степеней свободы k имеет вид

$$g_k(s) = \frac{1}{2^{k/2} \Gamma(k/2)} s^{k/2-1} e^{-s/2}.$$

В качестве результата проверки гипотезы о согласии выдается вычисленное значение статистики χ^2 Пирсона и вероятность того, что значение статистики может превысить данное значение при справедливости гипотезы о согласии $P\{S > S'\}$. Такая вероятность для распределения χ^2 вычисляется при числе степеней свободы $k-p-1$ и при $k-1$.

Гипотеза о согласии не должна приниматься, если задаваемый исследователем уровень значимости превышает найденную вероятность при соответствующем числе степеней свободы.

При построении статистики χ^2 используются таблицы асимптотически оптимального группирования, включенные в программное обеспечение. Эти таблицы получены при минимизации потерь информации Фишера, связанных с группированием данных, и обеспечивают максимальную мощность критерия χ^2 Пирсона при близких альтернативных гипотезах, т.е. при группировании данных в соответствии с атами таблицами критерий лучше улавливает небольшие отклонения от гипотетического распределения, согласие с которым проверяется.

Примечание. Для проверки согласия по критерию χ^2 Пирсона и по критерию отношения правдоподобия в случае частично группированной выборки Вы должны в исходной выборке задать все граничные точки для интервалов. Если в группированной части в эти интервалы не попало наблюдений, то соответствующие значения n_i должны быть заданы нулевыми.

Вторым критерием согласия, использующим группирование данных, является критерий отношения правдоподобия, для которого также справедливы все выводы, касающиеся асимптотически оптимального группирования. Отношение правдоподобия имеет вид

$$I = \prod_{i=1}^k \left(\frac{P_i(\theta)}{n_i / N} \right)^{n_i}.$$

В соответствующем критерии согласия используется статистика вида [11]

$$-2 \ln I = -2 \sum_{i=1}^k n_i \ln \left(\frac{P_i(\theta)}{n_i / N} \right).$$

При верной нулевой гипотезе эта статистика асимптотически распределена как χ^2 с $(k-1)$ -й степенью свободы. Если по выборке оценивались параметры распределения, то число степеней свободы уменьшается на количество оцененных параметров.

В случае негруппированной выборки вычисляются также статистики критериев Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса [28].

Статистика Колмогорова имеет вид

$$S_k = \frac{(6ND_N + 1)^2}{18N},$$

статистика Смирнова

$$S_m = \frac{(6ND_N^* + 1)^2}{9N},$$

где $D_N = \max(D_N^*, D_N^-)$, $D_N^* = \max_{1 \leq i \leq N} \left\{ \frac{i}{N} - F(x_i) \right\}$, $D_N^- = \max_{1 \leq i \leq N} \left\{ F(x_i) - \frac{i-1}{N} \right\}$,

N - объем выборки, x_1, x_2, \dots, x_N - упорядоченные по возрастанию выборочные значения, $F(x)$ - функция распределения, согласие с которой проверяется.

Распределение статистики Колмогорова S_k определяется законом Колмогорова с функцией распределения

$$K(y) = \sum_{k=0}^{\infty} (-1)^k e^{-2k^2y^2}.$$

Гипотеза о согласии не отвергается, если

$$P\{S_k > S'_k\} = 1 - K\left(\sqrt{\frac{S'_k}{2}}\right) > \alpha.$$

Статистика Смирнова S_n подчиняется в пределе распределению χ^2 с числом степеней свободы, равным 2. Гипотеза о согласии не отвергается, если

$$P\{S_n > S'_n\} = \int_{\zeta_0}^{\infty} \frac{1}{2} e^{-x^2/2} dx = 1 - e^{-\zeta_0^2/2} > \alpha.$$

Статистики Мизеса имеют вид

$$S_n = N\omega_n^2 = \frac{1}{12N} + \sum_{i=1}^N \left\{ F(x_i) - \frac{2i-1}{2N} \right\}^2$$

и

$$S_\Omega = N\Omega_n^2 = -N - 2 \sum_{i=1}^N \left\{ \frac{2i-1}{2N} \ln F(x_i) + \left(1 - \frac{2i-1}{2N}\right) \ln(1-F(x_i)) \right\}.$$

Для этих статистик известны предельные распределения вероятностей

$$\lim_{n \rightarrow \infty} P\{N\omega_n^2 < x\} = a1(x),$$

$$\lim_{n \rightarrow \infty} P\{N\Omega_n^2 < x\} = a2(x),$$

$$a1(x) = \frac{1}{\sqrt{2x}} \sum_{j=0}^{\infty} \frac{\Gamma(j+1/2)\sqrt{4j+1}}{\Gamma(1/2)\Gamma(j+1)} \exp\left\{-\frac{(4j+1)^2}{16x}\right\} \times \\ \times \left\{ I_{-\frac{1}{4}}\left[\frac{(4j+1)^2}{16x}\right] - I_{\frac{1}{4}}\left[\frac{(4j+1)^2}{16x}\right] \right\},$$

где $I_{\frac{v}{2}}(\cdot)$, $I_{\frac{1}{2}}(\cdot)$ - модифицированные функции Бесселя,

$$I_v(z) = \sum_{j=0}^{\infty} \frac{\left(\frac{z}{2}\right)^{v+2j}}{j! \Gamma(j+1)\Gamma(v+1)}, |z| < \infty, |\arg z| < \pi;$$

$$\alpha_2(x) = \frac{\sqrt{2\pi}}{x} \sum_{j=0}^{\infty} (-1)^j \frac{\Gamma(j+1/2)(4j+1)}{\Gamma(1/2)\Gamma(j+1)} \exp\left\{-\frac{(4j+1)^2 \pi^2}{8x}\right\} \times$$

$$\times \int_0^x \exp\left\{-\frac{x}{8(y^2+1)} - \frac{(4j+1)^2 \pi^2 y^2}{8x}\right\} dy.$$

Для этих статистик гипотезы о согласии не отвергаются, если выполняются неравенства

$$P\{S_n > S_n^*\} = 1 - \alpha_1(S_n^*) > \alpha$$

и

$$P\{S_Q > S_Q^*\} = 1 - \alpha_2(S_Q^*) > \alpha.$$

2.8. Многокритериальность задачи проверки гипотез о согласии

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 оценка по методу минимума χ^2 соответствует минимум статистики χ^2 , оценка 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 .

Замечание, изложенное в предыдущем абзаце, ещё раз подчёркивает необходимость принятия решения при подборе распределения, наиболее хорошо описывающего выборочные данные, по совокупности критериев проверки гипотез о согласии.

Когда перед нами возникает задача выбора распределения, с которым наиболее хорошо согласуются данные экспериментов, то последовательность наших действий описывается следующим алгоритмом.

- 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.
- Далее для выбранных распределений оцениваем параметры и проверяем гипотезы о согласии.
- 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, согласнее с которым наиболее хорошее.

В описываемом программном обеспечении при проверке гипотез о согласии для каждой используемой статистики S_i , $i = \overline{1, m}$, вычисляются

вероятности вида $P\{S_i > S'_i\} = \int_{S'_i}^{\infty} g_i(s)ds$, где S'_i - найденное по выборке значение соответствующей статистики, $g_i(s)$ - функция плотности распределения статистики S_i при условии, что гипотеза H_0 является истинной.

ной. S_i^* является функционалом, зависящим от конкретных выборки и закона распределения, т.е. $S_i^* = S_i^*(\bar{X}, f_i(x, \hat{\theta}))$, где через \bar{X} обозначена выборка случайной величины. Допустим, что на основании первичных предположений мы выделили множество законов распределений, для которых имеет смысла 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 $S_y^* = S_y^*(\bar{X}, f_j(x, \hat{\theta}))$ и вероятности $P(S_y > S_y^*) = \alpha_y$. Тогда при проверке гипотезы о согласии с j -м распределением по i -му критерию, если $\alpha_y > \alpha$, 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 $\forall i$

$$\alpha_d = \max_{j \in R_i} \alpha_j.$$

Обычно такой вывод можно сделать однозначно. 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. Такая "несогласованность" объясняется различием мер, используемых в критериях. Но вопрос, какой же закон следует предпочесть для описания случайной величины, выборка которой нами исследовалась, остаётся.

Таким образом, мы имеем естественную многокритеральную задачу принятия решения. Так как все критерии измеряются в единой шкале, то решить её можно, сформировав простой компромиссный критерий вида

$$\max_{j \in J} \sum_{i=1}^n \omega_i \alpha_{ij},$$

где ω_i - весовой коэффициент i -го критерия, $\sum_{i=1}^n \omega_i = 1$.

Некоторой неопределенностью здесь обладает лишь процедура выбора весовых коэффициентов. Определению их будет способствовать опыт эксплуатации системы программ и экспертные оценки. Наш опыт подсказывает, что примерно одинаковыми должны выбираться весовые коэффициенты для критериев Колмогорова, Смирнова, χ^2 и Ω^2 , для критериев χ^2 и отношения правдоподобия также примерно равными, но большими, чем четыре первых.

2.9. Алгоритмы вычисления оценок

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. 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 в модификации Флетчера-Ривса. Работа этого алгоритма 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 $\bar{S}^k = \nabla f(\bar{x}^k)$, при $k \neq 0$

$$\bar{S}^0 = \nabla f(\bar{x}^0) + \omega^0 \bar{S}^{k-1},$$

$$\omega^k = \frac{\|\nabla f(\bar{x}^k)\|^2}{\|\nabla f(\bar{x}^{k-1})\|^2},$$

а величина λ^k вычисляется в результате решения задачи $\max f(\bar{x}^k + \lambda \bar{S}^k)$. После $(n+1)$ -й итерации процедура циклически повторяется с заменой \bar{x}^0 на \bar{x}^{k+1} . Алгоритм заканчивается, как только выполнится соотношение $\|\nabla f(\bar{x}^k)\| < \epsilon$, где ϵ -заданная точность.

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.

3. Алгоритмическое обеспечение программной системы

Материал данного раздела охватывает распределения, включенные в программную систему. Изложение связано с реализацией программного обеспечения вычисления оценок максимального правдоподобия и использованием критерии 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 методом Ньютона, 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_{ij} определяются соотношением

$$J_{ij} = \sum_m n_i \times \sum_m \left(\frac{\partial \ln P_j(\hat{\theta})}{\partial \theta_i} \frac{\partial \ln P_i(\hat{\theta})}{\partial \theta_j} \right) P_i(\hat{\theta}) + \sum_m n_i \times \int \left(\frac{\partial \ln f(x, \hat{\theta})}{\partial \theta_i} \frac{\partial \ln f(x, \hat{\theta})}{\partial \theta_j} \right) f(x, \hat{\theta}) dx,$$

то приводятся выражения и для элементов этой матрицы.

При изложении материала приводятся лишь сами выражения, используемые при вычислении оценок, их дисперсий, при решении задач асимптотически оптимального группирования. Моменты, связанные с численным интегрированием и вычислением специальных функций, в тексте не затрагиваются и остаются за рамками изложения. Но, естественно, они решаются в описываемом программном обеспечении.

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 свodka представлена в работе [29].

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 с использованием порядковой статистики. Уравнение правдоподобия по θ_0 имеет вид

$$\frac{1}{\theta_0} \left\{ \sum_{(1)} n_i \frac{t_i e^{-t_i} - t_{i-1} e^{-t_{i-1}}}{e^{-t_{i-1}} - e^{-t_i}} + \sum_{(2)} \sum_{j=1}^n (1 - t_j) \right\} = 0,$$

где $t_i = \theta_0(x_i - \theta_1)$, $t_0 = \theta_0(x_0 - \theta_1)$. 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

$$J_0 = J(\theta_0) = \frac{1}{\theta_0} \left\{ \sum_{(1)} n_i \times \sum_{(0)} \frac{(t_i e^{-t_i} - t_{i-1} e^{-t_{i-1}})^2}{e^{-t_{i-1}} - e^{-t_i}} + \sum_{(2)} n_i \right\}.$$

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. 3.1.

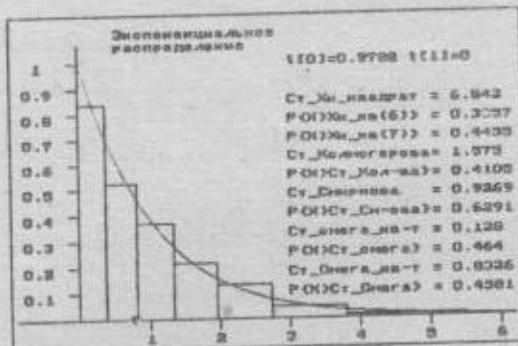


Рис. 3.1. Результаты оценивания параметров показательного распределения

3.2. Полунормальное распределение

Распределение модуля нормальной величины (полунормальное) имеет функцию плотности

$$f(x) = \frac{2}{\theta_0 \sqrt{2\pi}} e^{-x^2/2\theta_0^2},$$

где параметр $\theta_0 \in (0, \infty)$, и аргумент $x \in (0, \infty)$.

Параметр θ_0 находится как решение уравнения правдоподобия

$$\frac{1}{\theta_0} \left\{ \sum_{(i)} n_i \frac{t_{i+} e^{-t_{i+}^2/2} - t_i e^{-t_i^2/2}}{\sqrt{\frac{\pi}{2} (\Phi_0(t_i) - \Phi_0(t_{i-1}))}} + \sum_{(i)} \sum_{j=1}^n (t_{ij} - 1) \right\} = 0,$$

где $t_i = x_i / \theta_0$, $t_{ij} = x_{ij} / \theta_0$, $\Phi_0(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-d^2/2} dt$, и функция распределения $F(x) = \Phi_0(t)$. Вторая производная по θ_0 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{1}{\theta_0^2} \left[\sum_{(i)} n_i \left[\frac{(t_{i-1}^2 - 2)t_{i-1}e^{-t_{i-1}^2/2} - (t_i^2 - 2)t_i e^{-t_i^2/2}}{\sqrt{\frac{\pi}{2} \{\Phi_0(t_i) - \Phi_0(t_{i-1})\}}} - \left(\frac{t_{i-1}e^{-t_{i-1}^2/2} - t_i e^{-t_i^2/2}}{\sqrt{\frac{\pi}{2} \{\Phi_0(t_i) - \Phi_0(t_{i-1})\}}} \right)^2 \right] + \right. \\ \left. + \sum_{(2)} \sum_{j=1}^5 (1 - 3t_j^2) \right].$$

Количество информации Фишера о параметре θ_0 в частично группированной выборке

$$J_0 = J(\theta_0) = \frac{2}{\theta_0^2} \left\{ \sum_{(i)} n_i \times \sum_{(i)} \frac{(t_{i-1}e^{-t_{i-1}^2/2} - t_i e^{-t_i^2/2})^2}{\pi \{\Phi_0(t_i) - \Phi_0(t_{i-1})\}} + \sum_{(2)} n_i \right\}.$$

Асимптотическая дисперсия оценки θ_0 равна $D(\theta_0) = J^{-1}(\theta_0)$.

На рис. 3.2а приведен пример с результатами моделирования и последующего оценивания параметра θ_0 по негруппированным данным. Моделировалась выборка объемом 1000 наблюдений со значением $\theta_0 = 2$.

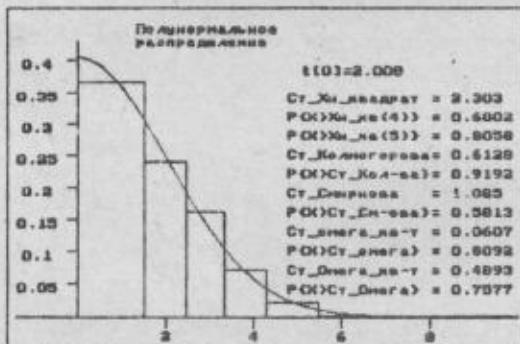


Рис. 3.2а. Результаты анализа по негруппированным данным

А на рис. 3.26 результаты анализа по той же выборке, но предварительно выборка группировалась. Как видно, существенной разницы в оценках не наблюдается.

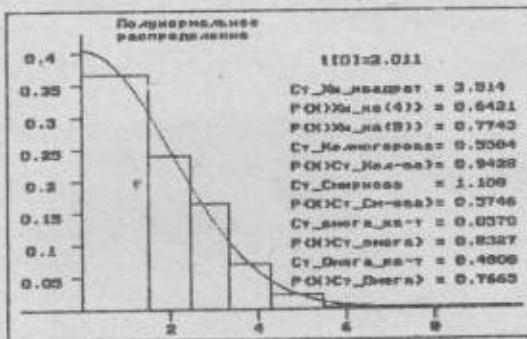


Рис. 3.26. Результаты анализа по группированным данным

3.3. Распределение Рэлея

Распределение Рэлея имеет функцию плотности

$$f(x) = \frac{x}{\theta_0^2} e^{-x^2/2\theta_0^2},$$

где параметр распределения $\theta_0 \in (0, \infty)$, аргумент $x \in [0, \infty)$, а функция распределения $F(x) = 1 - e^{-x^2/2\theta_0^2}$.

Оценка максимального правдоподобия параметра определяется как решение уравнения правдоподобия

$$\frac{1}{\theta_0} \left\{ \sum_i n_i \frac{t_i e^{-t_i^2/2\theta_0^2} - t_i e^{-t_i^2/\theta_0^2}}{e^{-t_i^2/2\theta_0^2} - e^{-t_i^2/\theta_0^2}} + \sum_{(j)} \sum_{i=1}^n (t_{ij}^2 - 2) \right\} = 0,$$

где $t_i = x_i / \theta_0$, $t_{ij} = x_{ij} / \theta_0$. Вторая производная по θ_0 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \left[\frac{(t_{i-1}^2 - 3)t_{i-1}^2 e^{-t_{i-1}^2/2} - (t_i^2 - 3)t_i^2 e^{-t_i^2/2}}{e^{-t_{i-1}^2/2} - e^{-t_i^2/2}} - \left(\frac{t_{i-1}^2 e^{-t_{i-1}^2/2} - t_i^2 e^{-t_i^2/2}}{e^{-t_{i-1}^2/2} - e^{-t_i^2/2}} \right)^2 \right] + \right.$$

$$\left. + \sum_{(2)} \sum_{j=1}^n (2 - 3t_j^2) \right\}.$$

Количество информации Фишера о параметре θ_0 в частично группированной выборке

$$J_0 = J(\theta_0) = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{(t_{i-1} e^{-t_{i-1}^2/2} - t_i e^{-t_i^2/2})^2}{e^{-t_{i-1}^2/2} - e^{-t_i^2/2}} + 4 \sum_{(2)} n_i \right\}.$$

Асимптотическая дисперсия оценки θ_0 равна $D(\theta_0) = J^{-1}(\theta_0)$.

На рис. 3.3 представлены результаты анализа выборки объемом 1000 наблюдений, смоделированной по закону Рэлея с параметром $\theta_0 = 3$.

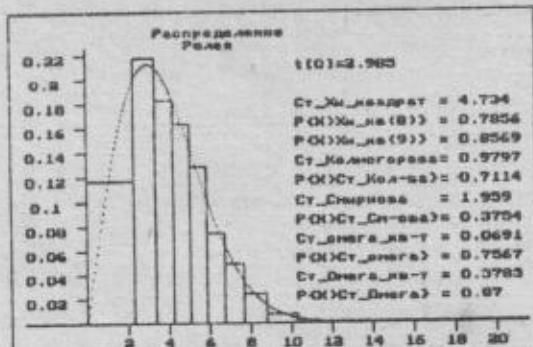


Рис.3.3. Плотность и гистограмма по результатам анализа

3.4. Распределение Максвелла

Распределение Максвелла имеет функцию плотности

$$f(x) = \frac{2x^2}{\theta_0^3 \sqrt{2\pi}} e^{-\frac{x^2}{2\theta_0^2}},$$

где параметр $\theta_0 > 0$ и аргумент $x \in [0, \infty)$. Функция распределения определяется выражением $F(x) = \Phi_0(t) - 2\Phi(t)$, где $t = x / \theta_0$.

$$\Phi_0(t) = 2 \int_0^t \phi(y) dy, \quad \phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$

Оценивается параметр θ_0 . Оценка получается в качестве единственного корня уравнения правдоподобия

$$\frac{1}{\theta_0} \left\{ \sum_{(i)} n_i \frac{2t_{i-1}^3 \phi(t_{i-1}) - 2t_i^3 \phi(t_{i-1})}{\Phi_0(t_i) - 2t_i \phi(t_i) - \Phi_0(t_{i-1}) + 2t_{i-1} \phi(t_{i-1})} + \sum_{(2)} \sum_{j=1}^n (t_j^2 - 3) \right\} = 0,$$

где $t_i = x_i / \theta_0$, $t_j = x_j / \theta_0$. Вторая производная по θ_0 определяется выражением

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_0^2} = & \frac{1}{\theta_0^2} \left\{ \sum_{(i)} n_i \left[\frac{(t_{i-1}^2 - 4)t_{i-1}^3 \phi(t_{i-1}) - (t_i^2 - 4)t_i^3 \phi(t_i)}{\Phi_0(t_i) - 2t_i \phi(t_i) - \Phi_0(t_{i-1}) + 2t_{i-1} \phi(t_{i-1})} - \right. \right. \\ & \left. \left. - 4 \left(\frac{t_{i-1}^3 \phi(t_{i-1}) - t_i^3 \phi(t_i)}{\Phi_0(t_i) - 2t_i \phi(t_i) - \Phi_0(t_{i-1}) + 2t_{i-1} \phi(t_{i-1})} \right)^2 \right] + 3 \sum_{(2)} \sum_{j=1}^n (1 - t_j^2) \right\}. \end{aligned}$$

Количество информации Фишера о параметре θ_0 в частично группированной выборке

$$J_0 = J(\theta_0) = \frac{1}{\theta_0^2} \left\{ 4 \sum_{(i)} n_i \times \sum_{(i)} \frac{[t_{i-1}^3 \phi(t_{i-1}) - t_i^3 \phi(t_i)]^2}{\Phi_0(t_i) - 2t_i \phi(t_i) - \Phi_0(t_{i-1}) + 2t_{i-1} \phi(t_{i-1})} + 6 \sum_{(2)} n_i \right\}.$$

Асимптотическая дисперсия оценки θ_0 равна $D(\theta_0) = J^{-1}(\theta_0)$.

На рис. 3.4 приведены результаты оценивания параметра θ_0 по смоделированной для данного распределения выборке со значением параметра $\theta_0 = 2$.

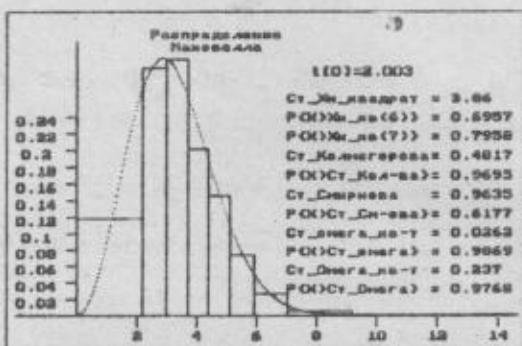


Рис. 3.4. Результаты оценивания параметра распределения Максвелла.

3.5. Распределение модуля многомерной нормальной величины

Распределение модуля многомерной нормальной случайной величины порядка n имеет функцию плотности

$$f(x) = \frac{2x^{n-1}}{(2\theta_0^2)^{n/2} \Gamma(n/2)} e^{-x^2/2\theta_0^2},$$

где $\theta_0 \in (0, \infty)$, $x \in [0, \infty)$, $\Gamma(\cdot)$ - гамма-функция Эйлера. Частными случаями этого распределения являются полунормальное распределение ($n=1$), распределение Рэлея ($n=2$), распределение Максвелла ($n=3$). Оценка параметра θ_0 получается в качестве единственного решения уравнения правдоподобия

$$\frac{1}{\theta_0} \left\{ \sum_{(1)} n_i \frac{t_{i-1} e^{-t_{i-1}^2/2} - t_i e^{-t_i^2/2}}{2^{n/2-1} \Gamma(n/2) [F(t_i) - F(t_{i-1})]} + \sum_{(2)} \sum_{j=1}^n (t_j^2 - n) \right\} = 0,$$

где $t_i = x_i / \theta_0$, $t_y = x_y / \theta_0$, а функция распределения при четном n определяется выражением

$$F(t) = 1 - e^{-t^2/2} \left(1 + \frac{t^2}{2} + \frac{t^4}{8} + \dots + \frac{t^{n-2}}{2^{n/2-1} \Gamma(n/2)} \right),$$

а при нечетном n :

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

и $\Phi_0(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-t'^2/2} dt$. Вторая производная по θ_0 определяется выражением

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_0^2} &= \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \left[\frac{(t_{i-1}^2 - n - 1)t_{i-1}''e^{-t_{i-1}^2/2} - (t_i^2 - n - 1)t_i''e^{-t_i^2/2}}{\xi [F(t_i) - F(t_{i-1})]} \right. \right. \\ &\quad \left. \left. - \left(\frac{t_{i-1}''e^{-t_{i-1}^2/2} - t_i''e^{-t_i^2/2}}{\xi [F(t_i) - F(t_{i-1})]} \right)^2 \right] + \sum_{(2)} \sum_{j=1}^n (n - 3t_j^2) \right\}, \end{aligned}$$

где $\xi = 2^{n/2-1} \Gamma(n/2)$. Количество информации Фишера о параметре θ_0 в частично группированной выборке

$$J_0 = J(\theta_0) = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} n_i \frac{(t_{i-1}''e^{-t_{i-1}^2/2} - t_i''e^{-t_i^2/2})^2}{\xi^2 [F(t_i) - F(t_{i-1})]} + 2n \sum_{(2)} n_i \right\}.$$

Асимптотическая дисперсия оценки θ_0 равна $D(\theta_0) = J^{-1}(\theta_0)$.

На рис. 3.5 приведены результаты оценивания параметра θ_0 по смоделированной для данного распределения выборке со значением параметра $\theta_0 = 2$ при $n = 6$.

3.6. Распределение Парето

Распределение Парето имеет функцию плотности

$$f(x) = \theta_0 \theta_1 x^{-\theta_0-1},$$

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.

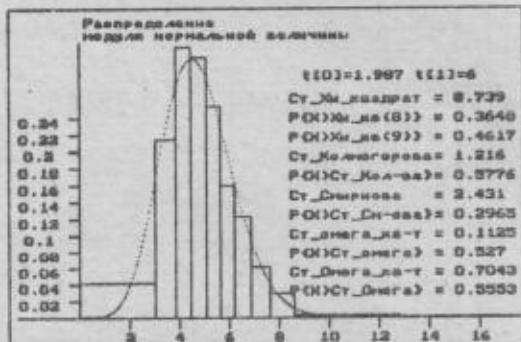


Рис. 3.5. Результаты оценивания параметра распределения модуля 5-мерного нормального вектора

Оценка θ_0 получается в качестве единственного корня уравнения правдоподобия

$$\frac{1}{\theta_0} \left\{ \sum_{(1)} n_i \frac{t_{i-1} \ln t_{i-1} - t_i \ln t_i}{t_{i-1} - t_i} + \sum_{(2)} \sum_{j=1}^n (1 + \ln t_j) \right\} = 0,$$

где $t_i = (x_i / \theta_1)^{-\theta_0}$, а $t_j = (x_j / \theta_1)^{-\theta_0}$. 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

$$J(\theta_0) = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{t_{i-1} \ln t_{i-1} - t_i \ln t_i}{t_{i-1} - t_i} + \sum_{(2)} n_i \right\}.$$

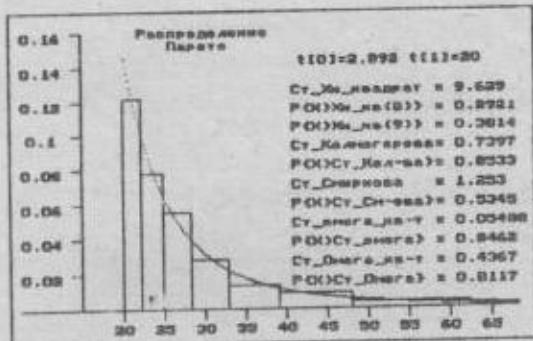


Рис. 3.6. Результаты анализа выборки, смоделированной по закону Парето.

На рис. 3.6 приведены результаты анализа выборки объёмом 1000 наблюдений, смоделированной по распределению Парето с параметрами $\theta_0 = 3$, $\theta_1 = 20$.

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 $x \in (0, \infty)$, θ_2 - параметр сдвига. Порядок m задается. В файле *ParDisGr.txt* порядок распределения m записывается как θ_0 . Параметр θ_1 при необходимости оценивается с использованием порядковых статистик. Распределение Эрланга является частным случаем гамма-распределения.

Оценка параметра θ_1 определяется как решение уравнения правдоподобия

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

где $t_i = \theta_1(x_i - \theta_2)$, $t_j = \theta_1(x_j - \theta_1)$. Вторая производная по θ_1 определяется выражением

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_1^2} &= \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{\int_{t_{i-1}}^{t_i} (m - 1 - t_i) e^{-t_i} - \int_{t_{i-1}}^{t_i} (m - 1 - t_i) e^{-t_{i-1}}}{\int_{t_{i-1}}^{t_i} t^{m-1} e^{-t} dt} - \right. \right. \\ &\quad \left. \left. \left(\frac{\int_{t_{i-1}}^{t_i} t_i^m e^{-t_i} - t_{i-1}^m e^{-t_{i-1}}}{\int_{t_{i-1}}^{t_i} t^{m-1} e^{-t} dt} \right)^2 \right] - m \sum_{(2)} n_i \right\}. \end{aligned}$$

Количество информации Фишера о параметре θ_1 в частично группированной выборке

$$J(\theta_1) = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{\left(\int_{t_{i-1}}^{t_i} t_i^m e^{-t_i} - \int_{t_{i-1}}^{t_i} t_{i-1}^m e^{-t_{i-1}} \right)^2}{\Gamma(m) \int_{t_{i-1}}^{t_i} t^{m-1} e^{-t} dt} + m \sum_{(2)} n_i \right\}.$$

На рис. 3.7 приведены результаты оценивания параметра распределения Эрланга. Исследуемая выборка моделировалась с параметром $\theta_1 = 2$ и $m = 4$.

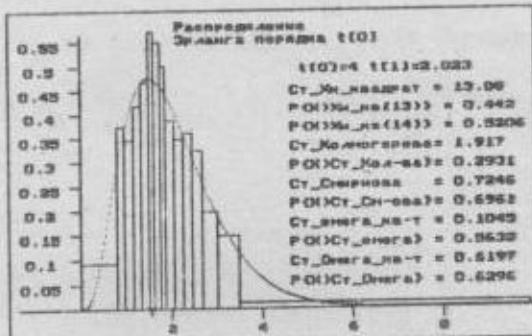


Рис. 3.7. Результаты моделирования и анализ распределения Эрланга

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 обычно задается. При необходимости его приближенная оценка находится по методу моментов. Оценка параметра θ_0 определяется как единственное решение уравнения приведения

$$\sum_i n_i \frac{\partial P_i(\theta)/\partial \theta_0}{P_i(\theta)} + \frac{1}{\theta_0} \sum_i \sum_{j=1}^n (1 - |t_j|) = 0,$$

где $t_i = \theta_0(x_i - \theta_1)$, $t_j = \theta_0(x_j - \theta_1)$.

$$P_i(\theta) = \begin{cases} \frac{1}{2}(e^{t_i} - e^{t_{i-1}}), & t_i < 0; \\ \frac{1}{2}(e^{-t_{i-1}} - e^{-t_i}), & t_{i-1} > 0; \\ 1 - \frac{1}{2}(e^{-t_i} + e^{t_{i-1}}), & t_{i-1} < 0 < t_i; \end{cases}$$

$$\frac{1}{\theta_0} \frac{\partial P_i(\theta)}{\partial \theta_0} = \begin{cases} \frac{1}{2}(t_i e^{t_i} - t_{i-1} e^{t_{i-1}}), & t_i < 0; \\ \frac{1}{2}(-t_{i-1} e^{-t_{i-1}} + t_i e^{-t_i}), & t_{i-1} > 0; \\ \frac{1}{2}(t_i e^{-t_i} - t_{i-1} e^{t_{i-1}}), & t_{i-1} < 0 < t_i. \end{cases}$$

Вторая производная по θ_0 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \sum_{(i)} n_i \left[\frac{\partial^2 P_i(\theta)}{\partial \theta_0^2} - \left(\frac{\partial P_i(\theta)}{\partial \theta_0} \right)^2 \right] / P_i(\theta) - \frac{1}{\theta_0^2} \sum_{(i)} n_i,$$

где

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \begin{cases} \frac{1}{2} (t_i^2 e^{t_i} - t_{i-1}^2 e^{t_{i-1}}), & t_i < 0, \\ \frac{1}{2} (t_{i-1}^2 e^{-t_{i-1}} - t_i^2 e^{-t_i}), & t_{i-1} > 0; \\ \frac{1}{2} (-t_i^2 e^{-t_i} - t_{i-1}^2 e^{t_{i-1}}), & t_{i-1} < 0 < t_i. \end{cases}$$

Количество информации Фишера в частично группированной выборке о параметре θ_0 имеет вид

$$J(\theta_0) = \sum_{(i)} n_i \times \left(\frac{\partial P_i(\theta)}{\partial \theta_0} \right)^2 / P_i(\theta) + \frac{1}{\theta_0^2} \sum_{(i)} n_i,$$

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. 3.8. Выборка объёмом 500 наблюдений моделировалась с параметрами $\theta_0 = 2$, $\theta_1 = 0$.

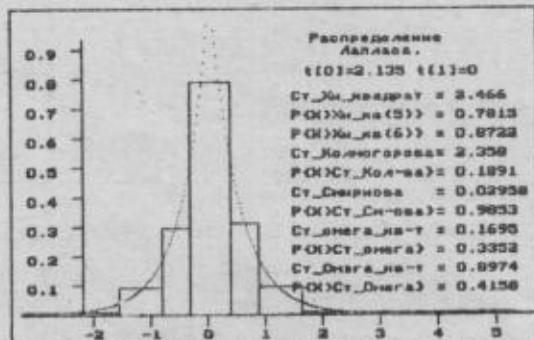


Рис.3.8. Анализ выборки, распределённой по закону Лапласа

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

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{\Phi(t_{i-1}) - \Phi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} \sum_{j=1}^n t_j \right\} = 0,$$

где $\Phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$, $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-t'^2/2} dt$, $t_i = \frac{x_i - \theta_0}{\theta_1}$, $t_j = \frac{x_j - \theta_0}{\theta_1}$. Вторая

производная логарифма функции правдоподобия по параметру θ_0 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{t_{i-1}\Phi(t_{i-1}) - t_i\Phi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} - \left(\frac{\Phi(t_{i-1}) - \Phi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} \right)^2 \right] - \sum_{(2)} n_i \right\}.$$

Оценка параметра θ_1 находится из уравнения правдоподобия

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{t_{i-1}\Phi(t_{i-1}) - t_i\Phi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} \sum_{j=1}^n (t_j^2 - 1) \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_1^2} = & \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{t_i(1-t_i^2)\Phi(t_i) - t_{i-1}(1-t_{i-1}^2)\Phi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} - \left(\frac{t_{i-1}\Phi(t_{i-1}) - t_i\Phi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} \right)^2 \right] \right. \\ & \left. + \sum_{(2)} \sum_{j=1}^n (1 - 3t_j^2) \right\}. \end{aligned}$$

При одновременном оценивании обоих параметров максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(i)} n_i \ln [\Phi(t_i) - \Phi(t_{i-1})] - \sum_{(i)} \sum_{j=1}^2 \left(\frac{t_j^2}{2} + \ln \theta_j + \ln \sqrt{2\pi} \right)$$

и оценки являются решением системы уравнений правдоподобия.

Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \frac{1}{\theta_0^2} \left\{ \sum_{(i)} n_i \times \sum_{(i)} \frac{[\phi(t_{i-1}) - \phi(t_i)]^2}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(i)} n_i \right\},$$

$$J_{11} = J(\theta_1) = \frac{1}{\theta_1^2} \left\{ \sum_{(i)} n_i \times \sum_{(i)} \frac{[t_{i-1}\phi(t_{i-1}) - t_i\phi(t_i)]^2}{\Phi(t_i) - \Phi(t_{i-1})} + 2 \sum_{(i)} n_i \right\},$$

$$J_{01} = J(\theta_0, \theta_1) = \frac{1}{\theta_1^2} \sum_{(i)} n_i \times \sum_{(i)} \frac{[\phi(t_i) - \phi(t_{i-1})][t_{i-1}\phi(t_{i-1}) - t_i\phi(t_i)]}{\Phi(t_i) - \Phi(t_{i-1})},$$

$$\text{и } J_{01} = J_{10}.$$

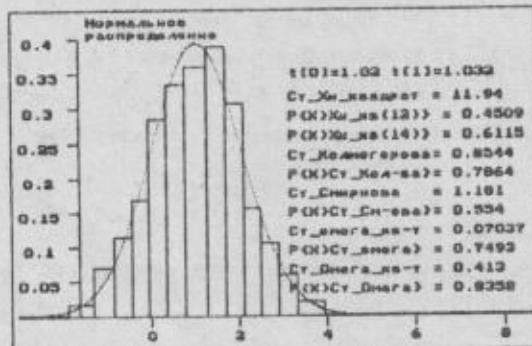


Рис. 3.9. Гистограмма и плотность нормального распределения

На рис. 3.9 приведены результаты оценивания параметров нормального распределения и последующего анализа с использованием оптимального группирования при пропуске согласия по критерию χ^2

Пирсона. Выборка объёмом 1000 наблюдений моделировалась с параметрами $\theta_0 = 1$, $\theta_1 = 1$.

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 $t_i, t_{\bar{x}}$ на $t_i = \frac{\ln x_i - \theta_0}{\theta_1}$, $t_{\bar{x}} = \frac{\ln \bar{x} - \theta_0}{\theta_1}$. На рис. 3.10 представлен результат анализа выборки объёмом 1000 измерений, смоделированной по данному закону с параметрами $\theta_0 = 1$, $\theta_1 = 0.4$.

На рисунке приведены результаты оценки параметров и проверки гипотезы о нормальности распределения. Проверка проводится с помощью критерия Колмогорова-Смирнова (КС), критерия Шапиро-Уилкса (ШУ) и критерия Пирсона (П).

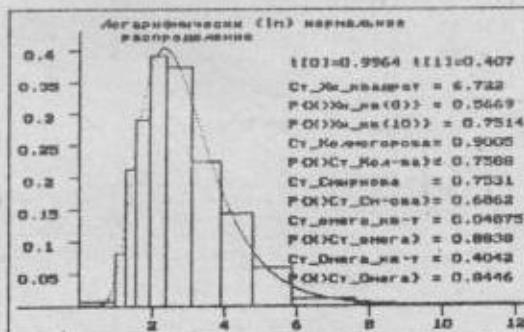


Рис. 3.10. Результаты анализа и оценивания логарифмически (ln) нормального распределения

3.11. Логарифмически (\lg) нормальное распределение

Логарифмически (\lg) нормальное распределение имеет функцию плотности

$$f(x) = \frac{1}{x\theta_1 \lg 10 \sqrt{2\pi}} e^{-(\lg x - \theta_0)^2 / 2\theta_1^2},$$

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 $t_i = \frac{\lg x_i - \theta_0}{\theta_1}$, $t_s = \frac{\lg x_s - \theta_0}{\theta_1}$.

На рис. 3.11 представлен результат анализа группированной выборки объёмом 1000 измерений, смоделированной по данному закону с параметрами $\theta_0 = 1$, $\theta_1 = 0.25$.

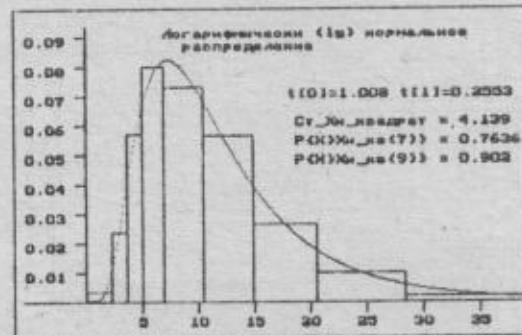


Рис. 3.11. Результаты анализа и оценивания логарифмически (\lg) нормального распределения

3.12. Распределение Коши

Распределение Коши, или распределение Брэйта-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 "патологического" 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

$$\frac{1}{\theta_0} \left\{ \sum_{(1)} n_i \frac{\frac{t_{i-1}}{1+t_{i-1}^2} - \frac{t_i}{1+t_i^2}}{\arctg t_i - \arctg t_{i-1}} + \sum_{(2)} \sum_{j=1}^n \left(1 - \frac{2}{1+t_j^2} \right) \right\} = 0,$$

где $t_i = (x_i - \theta_1)/\theta_0$, $t_\theta = (x_\theta - \theta_1)/\theta_0$. Вторая производная логарифма функции правдоподобия по параметру θ_0 определяется выражением

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_0^2} = & \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \left[\frac{\frac{2t_i}{(1+t_i^2)^2} - \frac{2t_{i-1}}{(1+t_{i-1}^2)^2}}{\arctg t_i - \arctg t_{i-1}} - \left(\frac{\frac{t_{i-1}}{1+t_{i-1}^2} - \frac{t_i}{1+t_i^2}}{\arctg t_i - \arctg t_{i-1}} \right)^2 \right] \right. \\ & \left. - \sum_{(2)} \sum_{j=1}^n \left(1 - \frac{2(1-t_j^2)}{(1+t_j^2)^2} \right) \right\}. \end{aligned}$$

Оценка параметра θ_1 находится как решение уравнения правдоподобия

$$\frac{1}{\theta_0} \left\{ \sum_{(1)} n_i \frac{\frac{1}{1+t_{i-1}^2} - \frac{1}{1+t_i^2}}{\arctg t_i - \arctg t_{i-1}} + 2 \sum_{(2)} \sum_{j=1}^n \frac{t_j}{1+t_j^2} \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \left[\frac{\frac{2t_{i-1}}{(1+t_{i-1}^2)^2} - \frac{2t_i}{(1+t_i^2)^2}}{\arctg t_i - \arctg t_{i-1}} - \left(\frac{\frac{1}{1+t_{i-1}^2} - \frac{1}{1+t_i^2}}{\arctg t_i - \arctg t_{i-1}} \right)^2 \right] - 2 \sum_{(2)} \sum_{j=1}^n \frac{1-t_j^2}{(1+t_j^2)^2} \right\}.$$

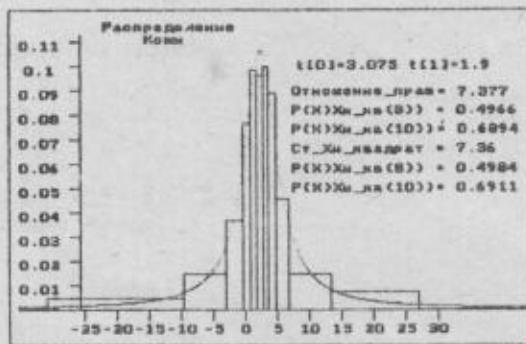


Рис. 3.12. Гистограмма и плотность распределения Коши

При одновременном оценивании обоих параметров максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln (\arctg t_i - \arctg t_{i-1}) - \sum_{(2)} \sum_{j=1}^n \left[\ln \theta_0 + \ln (1+t_j^2) \right] - \ln \pi \left(\sum_{(1)} n_i + \sum_{(2)} n_j \right).$$

и оценка векторного параметра определяется решением системы уравнений правдоподобия.

Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{1}{\pi} \frac{\left[\frac{t_{i-1}}{1+t_{i-1}^2} - \frac{t_i}{1+t_i^2} \right]^2}{\arctg t_i - \arctg t_{i-1}} + \frac{1}{2} \sum_{(2)} n_j \right\},$$

$$J_{11} = J(\theta_1) = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{1}{\pi} \frac{\left[\frac{1}{1+t_{i-1}^2} - \frac{1}{1+t_i^2} \right]^2}{\arctg t_i - \arctg t_{i-1}} + \frac{1}{2} \sum_{(2)} n_j \right\},$$

$$J_{01} = J(\theta_0, \theta_1) = \frac{1}{\pi \theta_0^2} \sum_{(1)} n_i \times \sum_{(1)} \frac{\left(\frac{t_{i-1}}{1+t_{i-1}^2} - \frac{t_i}{1+t_i^2} \right) \left(\frac{1}{1+t_{i-1}^2} - \frac{1}{1+t_i^2} \right)}{\arctg t_i - \arctg t_{i-1}}$$

и $J_{01} = J_{10}$. Некоторая "патология" данного распределения заключается 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. 3.12 приведены результаты оценивания параметров распределения и последующего анализа с использованием оптимального группирования при проверке согласия по критерию χ^2 Пирсона.

$\theta_0 = 3$, $\theta_1 = 2$. Оценивание осуществлялось по группированным данным.

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.

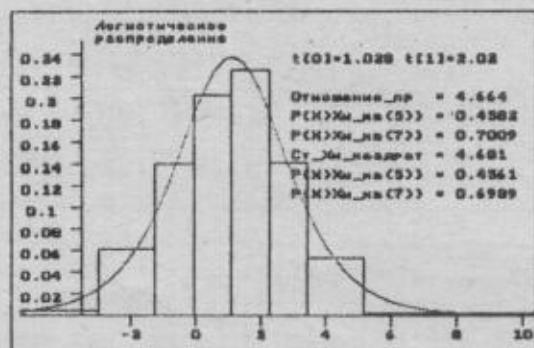


Рис. 3.13. Гистограмма и плотность логистического распределения

Оценка максимального правдоподобия параметра сдвига θ_0 получается в качестве единственного решения уравнения правдоподобия

$$\frac{\pi}{\theta_1 \sqrt{3}} \left(\sum_{(1)} n_i \frac{\frac{e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2} - \frac{e^{-t_i}}{(1+e^{-t_i})^2}}{\frac{(1+e^{-t_{i-1}})^{-1} - (1+e^{-t_{i-1}})^{-1}}{(1+e^{-t_i})^{-1} - (1+e^{-t_i})^{-1}}} + \sum_{(2)} \sum_{j=1}^n \left(1 - 2 \frac{e^{-t_j}}{1+e^{-t_j}} \right) \right) = 0,$$

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 даётся выражением

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{\pi^2}{3\theta_1^2} \left(\sum_{(1)} n_i \left[\frac{\frac{e^{-t_i}(e^{-t_i}-1)}{(1+e^{-t_i})^3} - \frac{e^{-t_{i-1}}(e^{-t_{i-1}}-1)}{(1+e^{-t_{i-1}})^3}}{\frac{(1+e^{-t_i})^{-1} - (1+e^{-t_i})^{-1}}{(1+e^{-t_{i-1}})^{-1} - (1+e^{-t_{i-1}})^{-1}}} - \right. \right.$$

$$\left. \left. - \left(\frac{\frac{e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2} - \frac{e^{-t_i}}{(1+e^{-t_i})^2}}{\frac{(1+e^{-t_{i-1}})^{-1} - (1+e^{-t_{i-1}})^{-1}}{(1+e^{-t_i})^{-1} - (1+e^{-t_i})^{-1}}} \right)^2 \right] - \sum_{(2)} \sum_{j=1}^n \frac{2e^{-t_j}}{(1+e^{-t_j})^2} \right).$$

Оценка масштабного параметра θ_1 находится как решение уравнения правдоподобия

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

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \frac{\pi^2}{30t_1^2} \left\{ \sum_{(1)} n_i \left[\frac{\frac{t_i e^{-t_i} (2 - t_i + 2e^{-t_i} + t_i e^{-t_i})}{(1+e^{-t_i})^2} - \frac{t_{i-1} e^{-t_{i-1}} (2 - t_{i-1} + 2e^{-t_{i-1}} + t_{i-1} e^{-t_{i-1}})}{(1+e^{-t_{i-1}})^2}}{(1+e^{-t_i})^{-1} - (1+e^{-t_{i-1}})^{-1}} \right. \right.$$

$$\left. \left. - \left(\frac{\frac{t_{i-1} e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2} - \frac{t_i e^{-t_i}}{(1+e^{-t_i})^2}}{(1+e^{-t_i})^{-1} - (1+e^{-t_{i-1}})^{-1}} \right)^2 \right] + \sum_{(2)} \sum_{j=1}^n \left(1 - \frac{2t_j (1+t_j e^{-t_j} - e^{-2t_j})}{(1+e^{-t_j})^2} \right) \right\}.$$

При одновременном оценивании обоих параметров максимизируется логарифм функции правдоподобия по частично группированной выборке, описываемый следующим выражением:

$$\ln L = \sum_{(1)} n_i \ln \left(\frac{1}{1+e^{-t_i}} - \frac{1}{1+e^{-t_{i-1}}} \right) + \sum_{(2)} \sum_{j=1}^n \left[\ln \frac{\pi}{\theta_1 \sqrt{3}} - t_j - 2 \ln (1+e^{-t_j}) \right],$$

и оценка является решением системы уравнений правдоподобия.

Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \frac{\pi}{\theta_1^2 \sqrt{3}} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \left[\frac{\frac{e^{-t_i}}{(1+e^{-t_i})^2} - \frac{e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2}}{(1+e^{-t_i})^{-1} - (1+e^{-t_{i-1}})^{-1}} \right]^2 + \frac{1}{3} \sum_{(2)} n_j \right\},$$

$$J_{11} = J(\theta_1) = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(0)} \left[\frac{\frac{t_i e^{-t_i}}{(1+e^{-t_i})^2} - \frac{t_{i-1} e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2}}{\frac{(1+e^{-t_i})^{-1}}{(1+e^{-t_i})^{-1}} - \frac{(1+e^{-t_{i-1}})^{-1}}{(1+e^{-t_{i-1}})^{-1}}} \right]^2 + \frac{\pi^2 + 3}{9} \sum_{(2)} n_i \right\},$$

$$J_{01} = J(\theta_0, \theta_1) = \frac{\pi^2}{\theta_1^2 \sqrt{3}} \sum_{(1)} n_i \times \sum_{(0)} \left[\frac{\frac{t_i e^{-t_i}}{(1+e^{-t_i})^2} - \frac{t_{i-1} e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2}}{\frac{(1+e^{-t_i})^{-1}}{(1+e^{-t_i})^{-1}} - \frac{(1+e^{-t_{i-1}})^{-1}}{(1+e^{-t_{i-1}})^{-1}}} \right] \times \\ \times \left[\frac{\frac{e^{-t_i}}{(1+e^{-t_i})^2} - \frac{e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2}}{\frac{(1+e^{-t_i})^{-1}}{(1+e^{-t_i})^{-1}} - \frac{(1+e^{-t_{i-1}})^{-1}}{(1+e^{-t_{i-1}})^{-1}}} \right],$$

$$\text{и } J_{01} = J_{10}.$$

На рис. 3.13 приведены результаты оценивания параметров логистического распределения и последующего анализа с использованием оптимального группирования при проверке согласия по критерию χ^2 Пирсона. Выборка объемом 1000 наблюдений моделировалась с параметрами $\theta_0 = 1$, $\theta_1 = 2$.

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

$$\frac{1}{\theta_0} \left\{ \sum_{(1)} n_i \frac{t_i e^{-t_i} \ln t_i - t_{i-1} e^{-t_{i-1}} \ln t_{i-1}}{e^{-t_{i-1}} - e^{-t_i}} + \sum_{(2)} \sum_{j=1}^n [(1 - t_j) \ln t_j + 1] \right\} = 0,$$

где $t_i = [(x_i - \theta_2)/\theta_1]^{b_0}$, $t_j = [(x_j - \theta_2)/\theta_1]^{b_0}$. Вторая производная логарифма функции правдоподобия по параметру θ_0 имеет вид

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_0^2} = & \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \left[\frac{e^{-t_i} t_i (1 - t_i) \ln^2 t_i - e^{-t_{i-1}} t_{i-1} (1 - t_{i-1}) \ln^2 t_{i-1}}{e^{-t_{i-1}} - e^{-t_i}} \right. \right. \\ & \left. \left. - \left(\frac{e^{-t_i} t_i \ln t_i - e^{-t_{i-1}} t_{i-1} \ln t_{i-1}}{e^{-t_{i-1}} - e^{-t_i}} \right)^2 \right] - \sum_{(2)} \sum_{j=1}^n (t_j \ln^2 t_j + 1) \right\}. \end{aligned}$$

Оценка параметра θ_1 находится как единственный корень уравнения правдоподобия

$$\frac{\theta_0}{\theta_1} \left\{ \sum_{(1)} n_i \frac{t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i}}{e^{-t_{i-1}} - e^{-t_i}} - \sum_{(2)} \sum_{j=1}^n (1 - t_j) \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\begin{aligned} \checkmark \quad \frac{\partial^2 \ln L}{\partial \theta_1^2} = & \frac{\theta_0^2}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{e^{-t_i} t_i (1 - t_i + 1/\theta_0) - e^{-t_{i-1}} t_{i-1} (1 - t_{i-1} + 1/\theta_0)}{e^{-t_{i-1}} - e^{-t_i}} \right. \right. \\ & \left. \left. - \left(\frac{e^{-t_{i-1}} t_{i-1} - e^{-t_i} t_i}{e^{-t_{i-1}} - e^{-t_i}} \right)^2 \right] + \frac{1}{\theta_0} \sum_{(2)} \sum_{j=1}^n [1 - (\theta_0 + 1)t_j] \right\}. \end{aligned}$$

При одновременном оценивании двух параметров θ_0 и θ_1 максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln [e^{-t_{i-1}} - e^{-t_i}] - \sum_{(2)} \sum_{j=1}^n (t_{ij} - \ln t_{ij} + \ln x_{ij} - \ln \theta_0)$$

и оценки являются решением системы приведенных выше уравнений правдоподобия. Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[t_i \ln t_i e^{-t_i} - t_{i-1} \ln t_{i-1} e^{-t_{i-1}}]^2}{e^{-t_{i-1}} - e^{-t_i}} + \sum_{(2)} n_i \times \left(1 + \frac{\pi^2}{6} + C^2 - 2C \right) \right\},$$

$$J_{11} = J(\theta_1) = \frac{\theta_0^2}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i}]^2}{e^{-t_{i-1}} - e^{-t_i}} + \sum_{(2)} n_i \right\},$$

$$J_{01} = J(\theta_0, \theta_1) = \frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{(t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i})(t_i \ln t_i e^{-t_i} - t_{i-1} \ln t_{i-1} e^{-t_{i-1}})}{e^{-t_{i-1}} - e^{-t_i}} + \sum_{(2)} n_i \times (C - 1) \right\},$$

где $C = 0.577215664$ - постоянная Эйлера, и $J_{01} = J_{10}$.

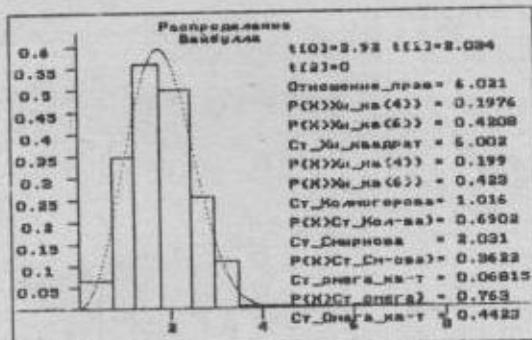


Рис. 3.14. Гистограмма и плотность распределения Вейбулла

На рис. 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 объемом 1000 наблюдений моделировалась с параметрами $\theta_0 = 3$, $\theta_1 = 2$, $\theta_2 = 0$.

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

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

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

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_0^2} &= \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{e^{-t_i} t_i (1-t_i) - e^{-t_{i-1}} t_{i-1} (1-t_{i-1})}{e^{-t_{i-1}} - e^{-t_i}} - \right. \right. \\ &\quad \left. \left. - \left(\frac{e^{-t_{i-1}} t_{i-1} - e^{-t_i} t_i}{e^{-t_{i-1}} - e^{-t_i}} \right)^2 \right] - \sum_{(2)} \sum_{j=1}^n t_{ij} \right\}. \end{aligned}$$

Оценка параметра θ_1 находится как единственный корень уравнения правдоподобия

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{s_{i-1} t_{i-1} e^{-t_{i-1}} - s_i t_i e^{-t_i}}{e^{-t_{i-1}} - e^{-t_i}} - \sum_{(2)} \sum_{j=1}^n [s_j (t_j - 1) - 1] \right\} = 0,$$

где $s_i = \frac{x_i - \theta_0}{\theta_1}$, $s_j = \frac{x_j - \theta_0}{\theta_1}$. Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_1^2} &= \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{s_{i-1} t_{i-1} e^{-t_{i-1}} (s_{i-1} t_{i-1} - s_{i-1} - 2) - s_i t_i e^{-t_i} (s_i t_i - s_i - 2)}{e^{-t_{i-1}} - e^{-t_i}} \right. \right. \\ &\quad \left. \left. - \left(\frac{s_{i-1} t_{i-1} e^{-t_{i-1}} - s_i t_i e^{-t_i}}{e^{-t_{i-1}} - e^{-t_i}} \right)^2 \right] - \sum_{(2)} \sum_{j=1}^n [2s_j (t_j - 1) + s_j^2 t_j - 1] \right\}. \end{aligned}$$

При одновременном оценивании двух параметров θ_0 и θ_1 максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln [e^{-t_{i-1}} - e^{-t_i}] - \sum_{(2)} \sum_{j=1}^n (t_j - s_j + \ln \theta_1)$$

и оценки являются решением системы приведенных выше уравнений правдоподобия. Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i}]^2}{e^{-t_{i-1}} - e^{-t_i}} + \sum_{(2)} n_i \right\},$$

$$J_{11} = J(\theta_1) = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[t_{i-1} s_{i-1} e^{-t_{i-1}} - t_i s_i e^{-t_i}]^2}{e^{-t_{i-1}} - e^{-t_i}} + \sum_{(2)} n_i \times \left(1 + \frac{\pi^2}{6} + C^2 - 2C \right) \right\},$$

$$J_{01} = J(\theta_0, \theta_1) = \frac{1}{\theta_1^2} \left\{ \sum_i n_i \times \sum_{(1)} \frac{(t_{i-1} e^{-t_i} - t_i e^{-t_i})(t_{i-1} s_{i-1} e^{-t_{i-1}} - t_i s_i e^{-t_i})}{e^{-t_{i-1}} - e^{-t_i}} + \right.$$

$$\left. + \sum_{(2)} n_i \times (C - 1) \right\},$$

где $C = 0.577215664$ — постоянная Эйлера, и $J_{01} = J_{10}$.

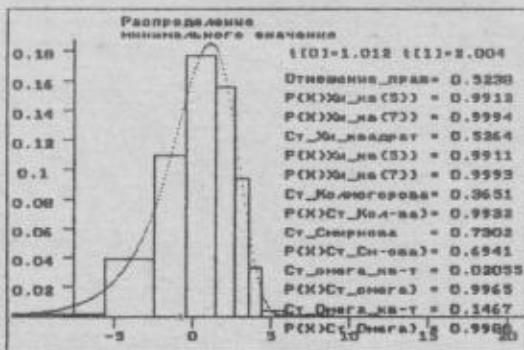


Рис. 3.15. Гистограмма и плотность распределения минимального значения

На рис. 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 $\theta_0 = 1$, $\theta_1 = 2$.

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

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

где $t_i = \exp\left(-\frac{x_i - \theta_0}{\theta_1}\right)$, $t_j = \exp\left(-\frac{x_j - \theta_0}{\theta_1}\right)$. Вторая производная логарифма функции правдоподобия по параметру θ_0 имеет вид

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{1}{\theta_1^2} & \left\{ \sum_{(1)} n_i \left[\frac{e^{-\theta_1} t_{i-1} (1 - t_{i-1}) - e^{-\theta_1} t_i (1 - t_i)}{e^{-\theta_1} - e^{-\theta_1}} - \right. \right. \\ & \left. \left. - \left(\frac{e^{-\theta_1} t_{i-1} - e^{-\theta_1} t_i}{e^{-\theta_1} - e^{-\theta_1}} \right)^2 \right] - \sum_{(2)} \sum_{j=1}^n t_j \right\}. \end{aligned}$$

Оценка параметра θ_1 находится как корень уравнения правдоподобия

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{s_i t_i e^{-\theta_1} - s_{i-1} t_{i-1} e^{-\theta_1}}{e^{-\theta_1} - e^{-\theta_1}} + \sum_{(2)} \sum_{j=1}^n [s_j (t_j - 1) - 1] \right\} = 0,$$

$$\text{где } s_i = -\frac{x_i - \theta_0}{\theta_1}, \quad s_j = -\frac{x_j - \theta_0}{\theta_1}.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{s_i t_i e^{-\theta_1} (s_i t_i - s_i - 2) - s_{i-1} t_{i-1} e^{-\theta_1} (s_{i-1} t_{i-1} - s_{i-1} - 2)}{e^{-\theta_1} - e^{-\theta_1}} - \right. \right. \\$$

$$\left. \left(\frac{s_i t_i e^{-t_i} - s_{i-1} t_{i-1} e^{-t_{i-1}}}{e^{-t_i} - e^{-t_{i-1}}} \right)^2 \right] + \sum_{(2)} \sum_{j=1}^n \left[1 + 2s_j (1-t_j) - s_j^2 t_j \right].$$

При одновременном оценивании двух параметров θ_0 и θ_1 максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln [e^{-t_i} - e^{-t_{i-1}}] + \sum_{(2)} \sum_{j=1}^n (s_j - t_j - \ln \theta_1)$$

и оценки являются решением системы приведенных выше уравнений правдоподобия.

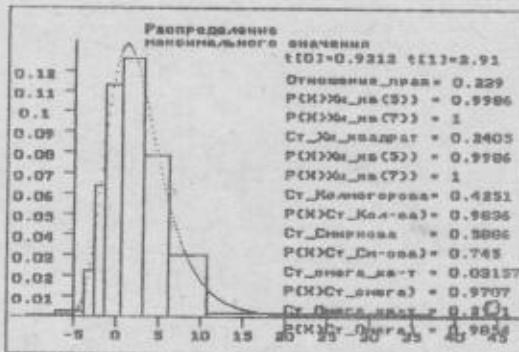


Рис. 3.16. Гистограмма и плотность распределения максимального значения

Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i}]^2}{e^{-t_i} - e^{-t_{i-1}}} + \sum_{(2)} n_i \right\},$$

$$J_{11} = J(\theta_1) = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{\left[t_i s_i e^{-t_i} - t_{i-1} s_{i-1} e^{-t_{i-1}} \right]^2}{e^{-t_i} - e^{-t_{i-1}}} + \sum_{(2)} n_i \times \left(1 + \frac{\pi^2}{6} + C^2 - 2C \right) \right\},$$

$$J_{01} = J(\theta_0, \theta_1) = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{(t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i})(t_i s_i e^{-t_i} - t_{i-1} s_{i-1} e^{-t_{i-1}})}{e^{-t_i} - e^{-t_{i-1}}} + \sum_{(2)} n_i \times (C - 1) \right\},$$

где $C = 0.577215664$ - постоянная Эйлера, и $J_{01} = J_{10}$.

На рис. 3.16 приведены результаты оценивания параметров распределения максимального значения по сгруппированным данным и последующего анализа с использованием оптимального группирования. Выборка объёмом 1000 наблюдений моделировалась с параметрами $\theta_0 = 1$, $\theta_1 = 3$.

3.17. Двойное показательное распределение

Двойное показательное распределение имеет функцию плотности

$$f(x) = \theta_0 \theta_1 \exp\{-x\theta_0 - \theta_1 \exp(-x\theta_0)\},$$

где масштабный параметр $\theta_0 \in (0, \infty)$, основной параметр $\theta_1 \in (0, \infty)$, и аргумент $x \in (-\infty, \infty)$. Оцениваться могут либо оба параметра, либо по одному. Оценка максимального правдоподобия масштабного параметра θ_0 определяется как решение уравнения правдоподобия

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{s_i t_i e^{-t_i} - s_{i-1} t_{i-1} e^{-t_{i-1}}}{e^{-t_i} - e^{-t_{i-1}}} + \sum_{(2)} \sum_{j=1}^n \left[s_j (t_j - 1) + \frac{\theta_1}{\theta_0} \right] \right\} = 0,$$

где $t_i = \theta_1 \exp(-\theta_0 x_i)$, $t_j = \theta_1 \exp(-\theta_0 x_j)$, $s_i = \theta_0 x_i$, $s_j = \theta_0 x_j$. Вторая производная логарифма функции правдоподобия по параметру θ_0 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \left[\frac{s_i^2 t_i e^{-t_i} (t_i - 1) - s_{i-1}^2 t_{i-1} e^{-t_{i-1}} (t_{i-1} - 1)}{e^{-t_i} - e^{-t_{i-1}}} - \right. \right.$$

$$\left. \left. - \left(\frac{s_i t_i e^{-t_i} - s_{i-1} t_{i-1} e^{-t_{i-1}}}{e^{-t_i} - e^{-t_{i-1}}} \right)^2 \right] + \sum_{(2)} \sum_{j=1}^n [1 - s_j^2 t_j] \right\}.$$

Оценка параметра θ_1 находится как корень уравнения правдоподобия

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i}}{e^{-t_i} - e^{-t_{i-1}}} + \sum_{(2)} \sum_{j=1}^n (1 - t_j) \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{t_i^2 e^{-t_i} - t_{i-1}^2 e^{-t_{i-1}}}{e^{-t_i} - e^{-t_{i-1}}} - \right. \right.$$

$$\left. \left. - \left(\frac{t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i}}{e^{-t_i} - e^{-t_{i-1}}} \right)^2 \right] - \sum_{(2)} n_i \right\}.$$

При одновременном оценивании двух параметров θ_0 и θ_1 максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln [e^{-t_i} - e^{-t_{i-1}}] + \sum_{(2)} \sum_{j=1}^n (\ln \theta_0 \theta_1 - s_j - t_j)$$

и оценки являются решением системы приведенных выше уравнений правдоподобия. Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \frac{1}{\theta_0^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[t_i s_i e^{-t_i} - t_{i-1} s_{i-1} e^{-t_{i-1}}]^2}{e^{-t_i} - e^{-t_{i-1}}} + \right.$$

$$+ \sum_{(2)} n_i \times \left(\ln^2 \theta_1 + 2(C-1) \ln \theta_1 + 1 + \frac{\pi^2}{6} + C^2 - 2C \right) \Bigg\},$$

$$J_{11} = J(\theta_1) = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i}]^2}{e^{-t_i} - e^{-t_{i-1}}} + \sum_{(2)} n_i \right\},$$

$$J_{01} = J(\theta_0, \theta_1) = \frac{1}{\theta_0 \theta_1} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{(t_i s_i e^{-t_i} - t_{i-1} s_{i-1} e^{-t_{i-1}})(t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i})}{e^{-t_i} - e^{-t_{i-1}}} + \sum_{(2)} n_i \times (1 - C - \ln \theta_2) \right\},$$

где $C = 0.577215664$ - постоянная Эйлера, и $J_{01} = J_{10}$.

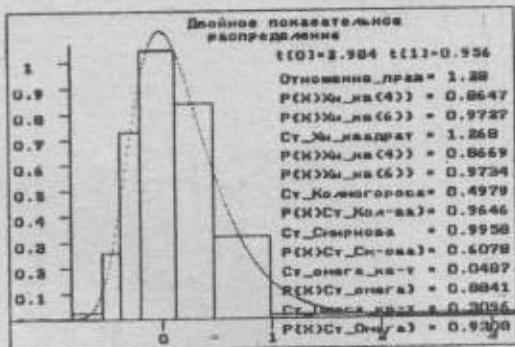


Рис. 3.17. Гистограмма и плотность двойного показательного распределения

На рис. 3.17 приведены результаты оценивания параметров двойного показательного распределения и последующего анализа. Выборка объёмом 1000 наблюдений моделировалась с параметрами $\theta_0 = 3$, $\theta_1 = 1$.

3.18. Распределение Накагами

Распределение Накагами имеет функцию плотности

$$f(x) = \frac{2}{\Gamma(\theta_1)} \left(\frac{\theta_1}{\theta_0^2} \right)^{\theta_1} (x - \theta_2)^{\theta_1 - 1} \exp \left\{ -\frac{\theta_1(x - \theta_2)^2}{\theta_0^2} \right\},$$

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.

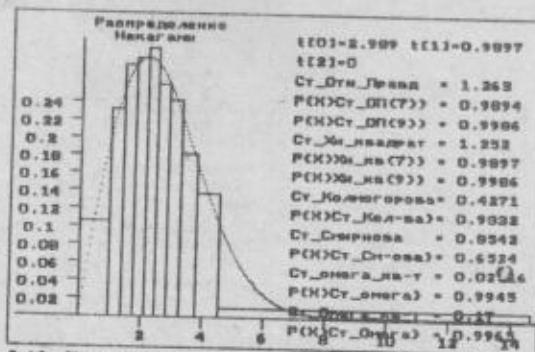


Рис. 3.18. Гистограмма и плотность распределения Накагами

Оценка максимального правдоподобия масштабного параметра θ_0 определяется как решение уравнения правдоподобия

$$\frac{2}{\theta_1} \left\{ \sum_{(1)} n_i \left[\int_{t_{i-1}}^{t_i} e^{-t/\theta_1} - \int_{t_i}^{t_{i+1}} e^{-t/\theta_1} \right] + \sum_{(2)} \sum_{j=1}^n (t_j - \theta_1) \right\} = 0,$$

где $t_i = \frac{\theta_1(x_i - \theta_2)^2}{\theta_0^2}$, $t_y = \frac{\theta_1(x_y - \theta_2)^2}{\theta_0^2}$. Вторая производная логарифма функции правдоподобия по параметру θ_0 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{2}{\theta_0^2} \left\{ \sum_{(1)} n_i \left[\frac{t_{i-1}^{\theta_1} e^{-t_{i-1}} (2t_{i-1} - 2\theta_1 - 1) - t_i^{\theta_1} e^{-t_i} (2t_i - 2\theta_1 - 1)}{\int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} dt} \right] + \right.$$

$$\left. \frac{2(t_{i-1}^{\theta_1} e^{-t_{i-1}} - t_i^{\theta_1} e^{-t_i})^2}{\int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} dt} \right] + \sum_{(2)} \sum_{j=1}^n (\theta_1 - 3t_j) \right\}.$$

Оценка параметра θ_1 находится как корень уравнения правдоподобия

$$\sum_{(1)} n_i \left\{ \frac{\int_{t_{i-1}}^t t^{\theta_1-1} e^{-t} \ln t dt + \frac{1}{\theta_1} (t_i^{\theta_1} e^{-t_i} - t_{i-1}^{\theta_1} e^{-t_{i-1}})}{\int_{t_{i-1}}^t t^{\theta_1-1} e^{-t} dt} - \psi(\theta_1) \right\} +$$

$$+ \sum_{(2)} \sum_{j=1}^n \left(\ln t_j - \frac{t_j}{\theta_1} \right) + (1 - \psi(\theta_1)) \sum_{(2)} n_i = 0,$$

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

$$\frac{\ln L}{\ln t_0} = \sum_{(1)} n_i \left\{ \frac{\int_{t_{i-1}}^t t^{\theta_0-1} \ln t dt + \frac{1}{\theta_0} \left[t_i^{\theta_0} e^{-\theta_0} \left[2 \ln t_i + 1 - \frac{1+t_i}{\theta_0} \right] - t_{i-1}^{\theta_0} e^{-\theta_0} \left[2 \ln t_{i-1} + 1 - \frac{1+t_{i-1}}{\theta_0} \right] \right]}{\int_{t_{i-1}}^t t^{\theta_0-1} e^{-\theta_0} dt} \right\}$$

$$\left(\frac{\int_{t_{i-1}}^t t^{\theta_0} e^{-\theta_0} \ln t dt + \frac{1}{\theta_0} (t_i^{\theta_0} e^{-\theta_0} - t_{i-1}^{\theta_0} e^{-\theta_0})}{\int_{t_{i-1}}^t t^{\theta_0-1} e^{-\theta_0} dt} \right)^2 + \frac{1}{\theta_0} \sum_{(2)} n_j - \psi'(\theta_0) \left(\sum_{(1)} n_i + \sum_{(2)} n_j \right)$$

При одновременном оценивании двух параметров θ_0 и θ_1 максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln \left(\int_{t_{i-1}}^t t^{\theta_0-1} e^{-\theta_0} dt \right) + \sum_{(2)} n_j (\ln 2 + \theta_0 \ln t_j - t_j - \ln x_j) - \ln \Gamma(\theta_0) \left(\sum_{(1)} n_i + \sum_{(2)} n_j \right)$$

и оценки являются решением системы приведенных уравнений правдоподобия. Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \frac{4}{\theta_0^2} \left\{ \sum_{(1)} n_i \times \frac{1}{\Gamma(\theta_0)} \sum_{(1)} \frac{\left[t_{i-1}^{\theta_0} e^{-\theta_0} - t_i^{\theta_0} e^{-\theta_0} \right]^2}{\int_{t_{i-1}}^t t^{\theta_0-1} e^{-\theta_0} dt} + \theta_0 \sum_{(2)} n_j \right\},$$

$$J_{11} = J(\theta_1) = \sum_{(1)} n_i \times \frac{1}{\Gamma(\theta_1)} \sum_{(1)} \left\{ \int_{t_{i-1}}^t t^{\theta_1-1} e^{-\theta_1} \ln t dt + \frac{1}{\theta_1} (t_i^{\theta_1} e^{-\theta_1} - t_{i-1}^{\theta_1} e^{-\theta_1}) \right\} -$$

$$-\psi(\theta_1) \left| \int_{t_{i-1}}^t t^{\theta_1-1} e^{-t} dt \right|^2 \Bigg/ \int_{t_{i-1}}^t t^{\theta_1-1} e^{-t} dt + \left(\psi'(\theta_1) - \frac{1}{\theta_1} \right) \sum_{(2)} n_i.$$

$$J_{01} = J(\theta_0, \theta_1) = \sum_{(1)} n_i \times \frac{2}{\theta_0 \Gamma(\theta_1)} \sum_{(1)} \left\{ \int_{t_{i-1}}^t t^{\theta_1-1} e^{-t} \ln t dt + \frac{1}{\theta_1} (t_i^{\theta_1} e^{-t_i} - t_{i-1}^{\theta_1} e^{-t_{i-1}}) - \right. \\ \left. - \psi(\theta_1) \int_{t_{i-1}}^t t^{\theta_1-1} e^{-t} dt \right\} (t_{i-1}^{\theta_1} e^{-t_{i-1}} - t_i^{\theta_1} e^{-t_i}) \Bigg/ \int_{t_{i-1}}^t t^{\theta_1-1} e^{-t} dt,$$

и $J_{01} = J_{10}$. На рис. 3.18 приведены результаты оценивания параметров распределения Накагами по сгруппированным данным и последующего анализа. Выборка объёмом 1000 наблюдений моделировалась с параметрами $\theta_0 = 3$, $\theta_1 = 1$, $\theta_2 = 0$. При группировании она разбивалась на интервалы равной вероятности (частоты).

3.19. Гамма-распределение

Гамма-распределение имеет функцию плотности

$$f(x) = \frac{\theta_0^{\theta_0}}{\Gamma(\theta_0)} (x - \theta_2)^{\theta_0-1} e^{-\theta_1(x-\theta_2)},$$

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

$$\sum_{(1)} n_i \left\{ \frac{\int_{t_{i-1}}^t t^{\theta_0-1} e^{-t} \ln t dt}{\int_{t_{i-1}}^t t^{\theta_0-1} e^{-t} dt} - \psi(\theta_0) \right\} + \sum_{(2)} \sum_{j=1}^n (\ln t_j - \psi(\theta_0)) = 0.$$

де $t_i = \theta_1(x_i - \theta_2)$, $t_{ij} = \theta_1(x_{ij} - \theta_2)$, $\psi(\theta_0) = \Gamma'(\theta_0)/\Gamma(\theta_0)$ - 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

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{t_i^{\theta_0} e^{-t_i} - t_{i-1}^{\theta_0} e^{-t_{i-1}}}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} dt} + \sum_{(2)} \sum_{j=1}^n (\theta_0 - t_{ij}) \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{t_i^{\theta_0} e^{-t_i} (\theta_0 - t_i - 1) - t_{i-1}^{\theta_0} e^{-t_{i-1}} (\theta_0 - t_{i-1} - 1)}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} dt} - \right. \right.$$

$$\left. \left. - \left(\frac{t_i^{\theta_0} e^{-t_i} - t_{i-1}^{\theta_0} e^{-t_{i-1}}}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} dt} \right)^2 \right] - \theta_0 \sum_{(2)} n_i \right\}$$

При одновременном оценивании двух параметров θ_0 и θ_1 максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln \left(\int_{t_{i-1}}^t t^{\theta_0-1} e^{-t} dt \right) + \sum_{(2)} \sum_{j=1}^n \left(\ln \theta_1 + (\theta_0 - 1) \ln t_j - t_j \right) - \\ - \ln \Gamma(\theta_0) \left(\sum_{(1)} n_i + \sum_{(2)} n_j \right)$$

и оценки являются решением системы приведенных выше уравнений правдоподобия.

Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \sum_{(1)} n_i \times \frac{1}{\Gamma(\theta_0)} \sum_{(1)} \left\{ \int_{t_{i-1}}^t t^{\theta_0-1} e^{-t} \ln t dt - \right. \\ \left. - \psi(\theta_0) \int_{t_{i-1}}^t t^{\theta_0-1} e^{-t} dt \right\}^2 \Bigg/ \int_{t_{i-1}}^t t^{\theta_0-1} e^{-t} dt + \psi'(\theta_0) \sum_{(2)} n_j,$$

$$J_{11} = J(\theta_1) = \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \frac{1}{\Gamma(\theta_0)} \sum_{(1)} \left[\int_{t_{i-1}}^t t^{\theta_0-1} e^{-t} dt - t_{i-1}^{\theta_0} e^{-t_{i-1}} \right]^2 + \theta_0 \sum_{(2)} n_j \right\},$$

$$J_{01} = J(\theta_0, \theta_1) = \sum_{(1)} n_i \times \frac{1}{\theta_1 \Gamma(\theta_0)} \sum_{(1)} n_j \left\{ \int_{t_{i-1}}^t t^{\theta_0-1} e^{-t} \ln t dt - \right. \\ \left. - \psi(\theta_0) \int_{t_{i-1}}^t t^{\theta_0-1} e^{-t} dt \right\} \left(t_i^{\theta_0} e^{-t_i} - t_{i-1}^{\theta_0} e^{-t_{i-1}} \right) \Bigg/ \int_{t_{i-1}}^t t^{\theta_0-1} e^{-t} dt - \frac{1}{\theta_1} \sum_{(2)} n_j,$$

$$\text{и } J_{01} = J_{10}.$$

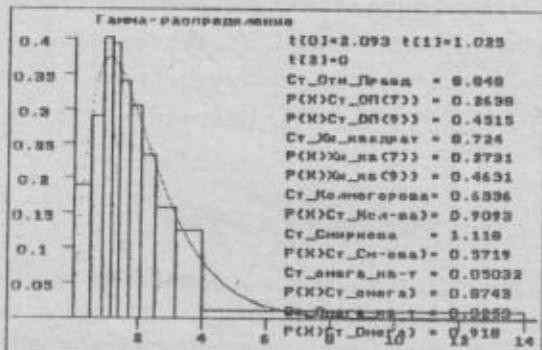


Рис. 3.19. Гистограмма и плотность гамма-распределения

Результаты оценивания параметров гамма-распределения и последующего анализа приведены на рис. 3.19. Выборка объёмом 1000 наблюдений моделировалась с параметрами $\theta_0 = 2$, $\theta_1 = 1$, $\theta_2 = 0$. При проверке согласия по критериям отношения правдоподобия и χ^2 Пирсона она разбивалась на интервалы равной вероятности (частоты).

3.20. Бета-распределение I рода

Бета-распределение I 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

$$\sum_{(1)} n_i \frac{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt}{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt} + \sum_{(2)} \sum_{j=1}^n \ln t_j - [\psi(\theta_0) - \psi(\theta_0 + \theta_1)] \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right) = 0,$$

где $t_i = (x_i - \theta_2) / \theta_3$, $t_j = (x_j - \theta_2) / \theta_3$, $\psi(\theta_0) = \Gamma'(\theta_0) / \Gamma(\theta_0)$ - логарифмическая производная гамма-функции Эйлера. Вторая производная логарифма функции правдоподобия по параметру θ_0 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \sum_{(1)} n_i \left\{ \frac{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} \ln^2 t dt}{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt} - \left(\frac{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt}{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt} \right)^2 \right\} -$$

$$- [\psi'(\theta_0) - \psi'(\theta_0 + \theta_1)] \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right).$$

Оценка параметра θ_1 находится как корень уравнения правдоподобия

$$\sum_{(1)} n_i \frac{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} \ln(1-t) dt}{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt} + \sum_{(2)} \sum_{j=1}^n \ln(1-t_j) -$$

$$- [\psi(\theta_1) - \psi(\theta_0 + \theta_1)] \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right) = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_i^2} = \sum_{(1)} n_i \left\{ \frac{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} \ln^2 (1-t) dt}{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt} - \left(\frac{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} \ln (1-t) dt}{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt} \right)^2 \right\} -$$

$$- [\psi'(\theta_1) - \psi'(\theta_0 + \theta_1)] \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right).$$

При одновременном оценивании двух параметров θ_0 и θ_1 максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln \left(\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt \right) + \sum_{(2)} \sum_{j=1}^n [(\theta_0 - 1) \ln t_j + (\theta_1 - 1) \ln (1 - t_j)] -$$

$$- \ln B(\theta_0, \theta_1) \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right)$$

и оценки являются решением представленной системы уравнений правдоподобия. Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \sum_{(1)} n_i \times \frac{1}{B(\theta_0, \theta_1)} \sum_{(1)} \left\{ \int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt \right\} -$$

$$- [\psi(\theta_0) - \psi(\theta_0 + \theta_1)] \int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt \Bigg\}^2 / \int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt +$$

$$+ [\psi'(\theta_0) - \psi'(\theta_0 + \theta_1)] \sum_{(2)} n_i,$$

$$J_{11} = J(\theta_1) = \sum_{(1)} n_i \times \frac{1}{B(\theta_0, \theta_1)} \sum_{(1)} \left\{ \int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} \ln (1-t) dt -$$

$$-\left[\psi(\theta_1) - \psi(\theta_0 + \theta_1)\right] \int_{t=1}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt\Bigg]^2 \Bigg/ \int_{t=1}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt +$$

$$+\left[\psi(\theta_1) - \psi(\theta_0 + \theta_1)\right] \sum_{(2)} n_i.$$

$$J_{01} = J(\theta_0, \theta_1) = \sum_{(1)} n_i \times \frac{1}{B(\theta_0, \theta_1)} \sum_{(1)} \left\{ \int_{t=1}^t t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt - \right.$$

$$\left. - [\psi(\theta_0) - \psi(\theta_0 + \theta_1)] \int_{t=1}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt \right\} \times \left\{ \int_{t=1}^t t^{\theta_0-1} (1-t)^{\theta_1-1} \ln(1-t) dt - \right.$$

$$\left. - [\psi(\theta_1) - \psi(\theta_0 + \theta_1)] \int_{t=1}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt \right\} \Bigg/ \int_{t=1}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt -$$

$$-\psi'(\theta_0 + \theta_1) \sum_{(2)} n_i,$$

и $J_{01} = J_{10}$.

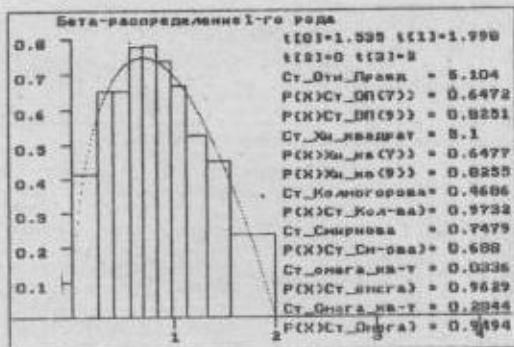


Рис. 3.20. Гистограмма и плотность бета-распределения 1 рода

Результаты оценивания параметров бета-распределения и последующего анализа приведены на рис. 3.20. Выборка объёмом 1000 наблюдений моделировалась с параметрами $\theta_0 = 1.5$, $\theta_1 = 2$, $\theta_2 = 0$.

$\theta_1 = 2$. При проверке согласия по критериям отношения правдоподобия и χ^2 Пирсона она разбивалась на интервалы равной вероятности (частоты).

3.21. Стандартное бета-распределение II рода

Стандартное бета-распределение II 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_g в виде $t_i = (x_i - \theta_2) / (1 + x_i - \theta_2)$ и $t_g = (x_g - \theta_2) / (1 + x_g - \theta_2)$.

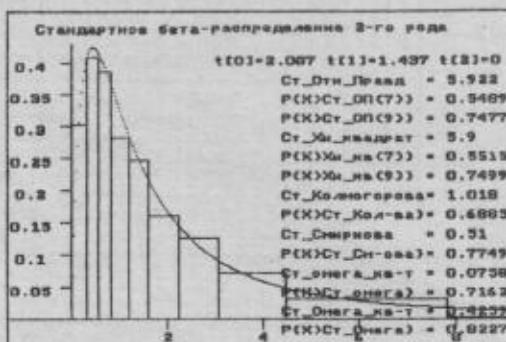


Рис. 3.21. Гистограмма и плотность бета-распределения II рода

Результаты оценивания параметров бета-распределения II рода и последующего анализа приведены на рис. 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).

3.22. Бета-распределение II рода

Функция плотности бета-распределения II 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

$$\sum_{(1)} n_i \frac{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt} + \sum_{(2)} \sum_{j=1}^n \ln t_{ij} - [\psi(\theta_0) - \psi(\theta_0 + \theta_1)] \left(\sum_{(1)} n_i + \sum_{(2)} n_j \right) = 0,$$

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

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \sum_{(1)} n_i \left\{ \frac{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} \ln^2 t dt}{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt} - \left(\frac{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt}{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt} \right)^2 \right\} -$$

$$- [\psi'(\theta_0) - \psi'(\theta_0 + \theta_1)] \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right).$$

Оценка параметра θ_1 находится как корень уравнения правдоподобия

$$\sum_{(1)} n_i \frac{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} \ln(1-t) dt}{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt} + \sum_{(2)} \sum_{j=1}^n \ln(1-t_j) -$$

$$- [\psi(\theta_1) - \psi(\theta_0 + \theta_1)] \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right) = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \sum_{(1)} n_i \left\{ \frac{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} \ln^2(1-t) dt}{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt} - \left(\frac{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} \ln(1-t) dt}{\int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt} \right)^2 \right\} -$$

$$- [\psi'(\theta_1) - \psi'(\theta_0 + \theta_1)] \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right).$$

Оценка параметра θ_2 находится как корень уравнения правдоподобия

$$\frac{\theta_0}{\theta_2} \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right) - \frac{\theta_0 + \theta_1}{\theta_2} \left[\sum_{(1)} n_i \frac{\int_{t=1}^t t^{\theta_0} (1-t)^{\theta_1-1} dt}{\int_{t=1}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt} + \sum_{(2)} \sum_{j=1}^n t_j \right] = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_2 имеет вид

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_2^2} &= \frac{\theta_0 + \theta_1}{\theta_2^2} \left[\sum_{(1)} n_i \left[\frac{(\theta_0 + \theta_1 + 1) \int_{t=1}^t t^{\theta_0+1} (1-t)^{\theta_1-1} dt}{\int_{t=1}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt} - \right. \right. \\ &\quad \left. \left. - (\theta_0 + \theta_1) \left(\frac{\int_{t=1}^t t^{\theta_0} (1-t)^{\theta_1-1} dt}{\int_{t=1}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt} \right)^2 \right] - \sum_{(2)} \sum_{j=1}^n t_j^2 \right] - \frac{\theta_0}{\theta_2^2} \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right) \end{aligned}$$

При одновременном оценивании трёх параметров θ_0 , θ_1 и θ_2 максимизируется логарифм функции правдоподобия

$$\begin{aligned} \ln L &= \sum_{(1)} n_i \ln \left(\int_{t=1}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt \right) + \sum_{(2)} \sum_{j=1}^n [(\theta_0 - 1) \ln t_j + (\theta_1 - 1) \ln (1 - t_j)] - \\ &\quad - \ln B(\theta_0, \theta_1) \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right) + \ln \theta_2 \sum_{(2)} n_i \end{aligned}$$

и оценки являются решением системы уравнений правдоподобия.

Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \sum_{(1)} n_i \times \frac{1}{B(\theta_0, \theta_1)} \sum_{(1)} \left\{ \int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt - \right. \\ \left. - [\psi(\theta_0) - \psi(\theta_0 + \theta_1)] \int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt \right\}^2 \Bigg/ \int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt + \\ + [\psi'(\theta_0) - \psi'(\theta_0 + \theta_1)] \sum_{(2)} n_i,$$

$$J_{11} = J(\theta_1) = \sum_{(1)} n_i \times \frac{1}{B(\theta_0, \theta_1)} \sum_{(1)} \left\{ \int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} \ln(1-t) dt - \right. \\ \left. - [\psi(\theta_1) - \psi(\theta_0 + \theta_1)] \int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt \right\}^2 \Bigg/ \int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt + \\ + [\psi'(\theta_1) - \psi'(\theta_0 + \theta_1)] \sum_{(2)} n_i,$$

$$J_{01} = J(\theta_0, \theta_1) = \sum_{(1)} n_i \times \frac{1}{B(\theta_0, \theta_1)} \sum_{(1)} \left\{ \int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt - \right. \\ \left. - [\psi(\theta_0) - \psi(\theta_0 + \theta_1)] \int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt \right\} \times \left\{ \int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} \ln(1-t) dt - \right. \\ \left. - [\psi(\theta_1) - \psi(\theta_0 + \theta_1)] \int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt \right\} \Bigg/ \int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt - \\ - \psi'(\theta_0 + \theta_1) \sum_{(2)} n_i,$$

$$J_{22} = J(\theta_2) = \sum_{(1)} n_i \times \frac{1}{\theta_2^2 B(\theta_0, \theta_1)} \sum_{(1)} \left\{ \theta_0 \int_{t_{i-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt - \right.$$

$$-(\theta_0 + \theta_1) \int_{t_{k-1}}^t t^{\theta_0} (1-t)^{\theta_1-1} dt \Bigg)^2 \left/ \int_{t_{k-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt + \frac{\theta_0 \theta_1}{\theta_2^2 (\theta_0 + \theta_1 + 1)} \sum_{(2)} n_i \right.,$$

$$J_{02} = J(\theta_0, \theta_2) = \sum_{(0)} n_i \times \frac{1}{\theta_2 B(\theta_0, \theta_1)} \sum_{(0)} \left\{ \frac{\int_{t_{k-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt}{\int_{t_{k-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt} - \right.$$

$$\left. - (\psi(\theta_0) - \psi(\theta_0 + \theta_1)) \right\} \left(\theta_0 \int_{t_{k-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt - (\theta_0 + \theta_1) \int_{t_{k-1}}^t t^{\theta_0} (1-t)^{\theta_1-1} dt \right) - \\ - \frac{\theta_1}{\theta_2 (\theta_0 + \theta_1)} \sum_{(2)} n_i,$$

$$J_{12} = J(\theta_1, \theta_2) = \sum_{(0)} n_i \times \frac{1}{\theta_2 B(\theta_0, \theta_1)} \sum_{(0)} \left\{ \frac{\int_{t_{k-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} \ln(1-t) dt}{\int_{t_{k-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt} - \right.$$

$$\left. - (\psi(\theta_1) - \psi(\theta_0 + \theta_1)) \right\} \left(\theta_0 \int_{t_{k-1}}^t t^{\theta_0-1} (1-t)^{\theta_1-1} dt - (\theta_0 + \theta_1) \int_{t_{k-1}}^t t^{\theta_0} (1-t)^{\theta_1-1} dt \right) - \\ + \frac{\theta_0}{\theta_2 (\theta_0 + \theta_1)} \sum_{(2)} n_i,$$

и $J_{01} = J_{10}$, $J_{02} = J_{20}$, $J_{12} = J_{21}$.

Результаты оценивания параметров бета-распределения II рода и последующего анализа приведены на рис. 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 струпин-

рованной выборке. При группировании она разбивалась на интервалы равной вероятности (частоты).

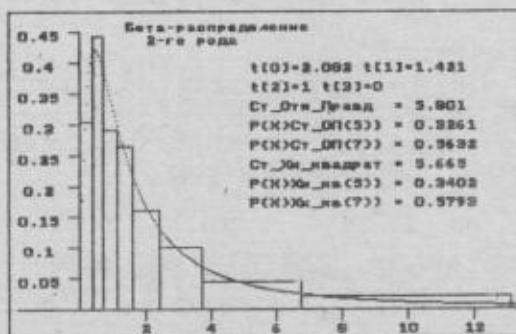


Рис. 3.22. Гистограмма и плотность бета-распределения II рода

3.23. Распределение Sb-Джонсона

Распределение Sb-Джонсона имеет функцию плотности вида

$$f(x) = \frac{\theta_0 \theta_1}{(x - \theta_3)(\theta_2 + \theta_3 - x)} \exp \left\{ -\frac{1}{2} \left[\theta_0 - \theta_1 \ln \frac{x - \theta_3}{\theta_2 + \theta_3 - x} \right]^2 \right\},$$

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

$$\sum_{(1)} n_i \frac{\phi(t_i) - \phi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} - \sum_{(2)} \sum_{j=1}^n t_j = 0,$$

где

$$t_i = \theta_0 + \theta_1 \ln \frac{x_i - \theta_3}{\theta_2 + \theta_3 - x_i}, \quad t_y = \theta_0 + \theta_1 \ln \frac{x_y - \theta_3}{\theta_2 + \theta_3 - x_y}, \quad \Phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$

$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-t^2/2} dt$. Вторая производная логарифма функции правдоподобия по параметру θ_0 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \sum_{(1)} n_i \left[\frac{t_{i-1}\phi(t_{i-1}) - t_i\phi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} - \left(\frac{\phi(t_i) - \phi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} \right)^2 \right] - \sum_{(2)} n_i.$$

Оценка параметра θ_1 находится в качестве решения уравнения правдоподобия

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{(t_i - \theta_0)\phi(t_i) - (t_{i-1} - \theta_0)\phi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} n_i - \sum_{(2)} \sum_{j=1}^n t_{ij} (t_{ij} - \theta_0) \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_1^2} = & \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{t_{i-1}(t_{i-1} - \theta_0)^2 \phi(t_{i-1}) - t_i(t_i - \theta_0)^2 \phi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} \right. \right. \\ & \left. \left. - \left(\frac{(t_i - \theta_0)\phi(t_i) - (t_{i-1} - \theta_0)\phi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} \right)^2 \right] - \sum_{(2)} \sum_{j=1}^n \left[1 + (t_{ij} - \theta_0)^2 \right] \right\}. \end{aligned}$$

При одновременном оценивании параметров θ_0 и θ_1 максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln [\Phi(t_i) - \Phi(t_{i-1})] + \ln \frac{\theta_1 \theta_2}{\sqrt{2\pi}} \sum_{(2)} n_i -$$

$$-\sum_{(2)} \sum_{j=1}^n \left\{ \frac{t_j^2}{2} + \ln[(x_j - \theta_3)(\theta_2 + \theta_3 - x_j)] \right\}$$

и оценки являются решением системы уравнений правдоподобия. Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \sum_{(1)} n_i \times \sum_{(1)} \frac{[\varphi(t_i) - \varphi(t_{i-1})]^2}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} n_i,$$

$$J_{11} = J(\theta_1) = \sum_{(1)} n_i \times \frac{1}{\theta_1^2} \sum_{(1)} \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]^2}{\Phi(t_i) - \Phi(t_{i-1})} + \frac{2 + \theta_0^2}{\theta_1^2} \sum_{(2)} n_i,$$

$$J_{01} = J(\theta_0, \theta_1) = \sum_{(1)} n_i \times \frac{1}{\theta_1} \sum_{(1)} \left\{ \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]}{\Phi(t_i) - \Phi(t_{i-1})} \times \right. \\ \left. \times [\varphi(t_i) - \varphi(t_{i-1})] \right\} - \frac{\theta_0}{\theta_1} \sum_{(2)} n_i,$$

$$\text{и } J_{10} = J_{01}.$$

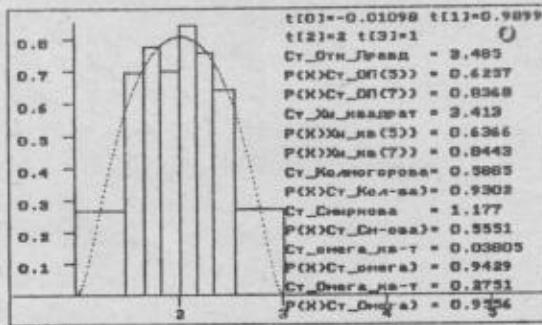


Рис. 3.23. Гистограмма и плотность распределения Sb-Джонсона

На рис. 3.23 приведены результаты оценивания параметров распределения Sb-Джонсона и последующей проверки критериев согласия.

Выборка объёмом 1000 наблюдений моделировалась с параметрами $\theta_0 = 0$, $\theta_1 = 1$, $\theta_2 = 2$, $\theta_3 = 1$.

3.24. Распределение СЛ-Джонсона

Распределение СЛ-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

$$\sum_{(1)} n_i \frac{\varphi(t_i) - \varphi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} - \sum_{(2)} \sum_{j=1}^n t_j = 0,$$

где $t_i = \theta_0 + \theta_1 \ln \frac{x_i - \theta_3}{\theta_2}$, $t_j = \theta_0 + \theta_1 \ln \frac{x_j - \theta_3}{\theta_2}$, $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$.

$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-t^2/2} dt$. 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

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{(t_i - \theta_0) \varphi(t_i) - (t_{i-1} - \theta_0) \varphi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} n_i - \sum_{(2)} \sum_{j=1}^n t_j (t_j - \theta_0) \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_1^2} &= \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{t_{i-1} (t_{i-1} - \theta_0)^2 \varphi(t_{i-1}) - t_i (t_i - \theta_0)^2 \varphi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} - \right. \right. \\ &\quad \left. \left. - \left(\frac{(t_i - \theta_0) \varphi(t_i) - (t_{i-1} - \theta_0) \varphi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} \right)^2 \right] - \sum_{(2)} \sum_{j=1}^n \left[1 + (t_j - \theta_0)^2 \right] \right\}. \end{aligned}$$

Оценка параметра θ_1 находится в качестве решения уравнения правдоподобия

$$\frac{\theta_1}{\theta_2} \left\{ \sum_{(1)} n_i \frac{\varphi(t_{i-1}) - \varphi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} \sum_{j=1}^n t_j \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_2 определяется выражением

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_2^2} &= \frac{\theta_1^2}{\theta_2^2} \left\{ \sum_{(1)} n_i \left[\frac{(t_{i-1} - 1/\theta_1) \varphi(t_{i-1}) - (t_i - 1/\theta_1) \varphi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} - \right. \right. \\ &\quad \left. \left. - \left(\frac{\varphi(t_{i-1}) - \varphi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} \right)^2 \right] - \sum_{(2)} n_i - \frac{1}{\theta_1} \sum_{(2)} \sum_{j=1}^n t_j \right\}. \end{aligned}$$

При одновременном оценивании параметров θ_0 , θ_1 и θ_2 максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln [\Phi(t_i) - \Phi(t_{i-1})] + \ln \frac{\theta_1}{\sqrt{2\pi}} \sum_{(2)} n_i -$$

$$-\sum_{(2)}^n \sum_{j=1}^n \left\{ \frac{t_j^2}{2} + \ln(x_j - \theta_1) \right\}$$

и оценки являются решением системы уравнений правдоподобии. Элементы информационной матрицы Фишера определяются соотношениями

$$\omega_{00} = J(\theta_0) = \sum_{(1)} n_i \times \sum_{(1)} \frac{[\varphi(t_i) - \varphi(t_{i-1})]^2}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} n_i,$$

$$J_{11} = J(\theta_1) = \sum_{(1)} n_i \times \frac{1}{\theta_1^2} \sum_{(1)} \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]^2}{\Phi(t_i) - \Phi(t_{i-1})} + \frac{2 + \theta_1^2}{\theta_1^2} \sum_{(2)} n_i,$$

$$J_{01} = J(\theta_0, \theta_1) = \sum_{(1)} n_i \times \frac{1}{\theta_1} \sum_{(1)} \left\{ \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]}{\Phi(t_i) - \Phi(t_{i-1})} \times \right.$$

$$\left. \times [\varphi(t_i) - \varphi(t_{i-1})] \right\} - \frac{\theta_0}{\theta_1} \sum_{(2)} n_i,$$

$$J_{22} = J(\theta_2) = \frac{\theta_2^2}{\theta_1^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[\varphi(t_{i-1}) - \varphi(t_i)]^2}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} n_i \right\},$$

$$J_{02} = J(\theta_0, \theta_2) = \frac{\theta_1}{\theta_2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[\varphi(t_i) - \varphi(t_{i-1})][\varphi(t_{i-1}) - \varphi(t_i)]}{\Phi(t_i) - \Phi(t_{i-1})} - \sum_{(2)} n_i \right\},$$

$$J_{12} = J(\theta_1, \theta_2) = \sum_{(1)} n_i \times \frac{1}{\theta_2} \sum_{(1)} \left\{ \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]}{\Phi(t_i) - \Phi(t_{i-1})} \times \right.$$

$$\left. \times [\varphi(t_{i-1}) - \varphi(t_i)] \right\} + \frac{\theta_0}{\theta_2} \sum_{(2)} n_i,$$

$$\text{и } J_{01} = J_{10}, \quad J_{02} = J_{20}, \quad J_{12} = J_{21}.$$

Результаты оценивания параметров распределения Sl-Джонсона и последующей проверки критериев согласия приведены на рис. 3.24. Выборка объемом 1000 наблюдений моделировалась с параметрами $\theta_0 = 0$, $\theta_1 = 1$, $\theta_2 = 2$, $\theta_3 = 1$.

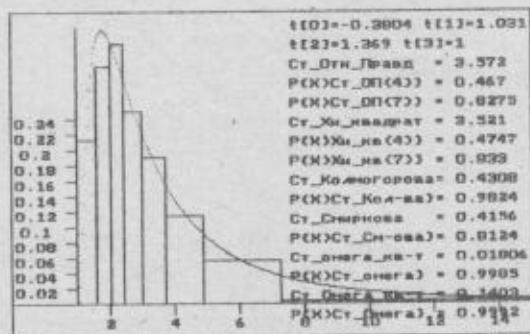


Рис. 3.24. Гистограмма и плотность распределения Sl-Джонсона

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

$$\sum_{(1)} n_i \frac{\phi(t_i) - \phi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} - \sum_{(2)} \sum_{j=1}^n t_{ij} = 0,$$

где

$$t_i = \theta_0 + \theta_1 \ln \left\{ \frac{x_i - \theta_3}{\theta_2} + \sqrt{\left(\frac{x_i - \theta_3}{\theta_2} \right)^2 + 1} \right\}, \quad \Phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$

$$t_y = \theta_0 + \theta_1 \ln \left\{ \frac{x_y - \theta_3}{\theta_2} + \sqrt{\left(\frac{x_y - \theta_3}{\theta_2} \right)^2 + 1} \right\}, \quad \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-u^2/2} du.$$

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

$$\frac{1}{\theta_1} \left\{ \sum_{(1)} n_i \frac{(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} n_i - \sum_{(2)} \sum_{j=1}^n t_y (t_y - \theta_0) \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \theta_1^2} &= \frac{1}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{t_{i-1}(t_{i-1} - \theta_0)^2 \varphi(t_{i-1}) - t_i(t_i - \theta_0)^2 \varphi(t_i)}{\Phi(t_i) - \Phi(t_{i-1})} \right. \right. \\ &\quad \left. \left. - \left(\frac{(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})}{\Phi(t_i) - \Phi(t_{i-1})} \right)^2 \right] - \sum_{(2)} \sum_{j=1}^n \left[1 + (t_y - \theta_0)^2 \right] \right\}. \end{aligned}$$

Оценка параметра θ_2 находится в качестве решения уравнения правдоподобия

$$\frac{\theta_1}{\theta_2} \left\{ \sum_{(1)} n_i \frac{\frac{s_{i-1}\Phi(t_{i-1})}{\sqrt{s_{i-1}^2+1}} - \frac{s_i\Phi(t_i)}{\sqrt{s_i^2+1}}}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} \sum_{j=1}^n \left[\frac{t_y s_y}{\sqrt{s_y^2+1}} - \frac{1}{\theta_1(s_y^2+1)} \right] \right\} = 0,$$

где $s_i = (x_i - \theta_3)/\theta_2$, $s_y = (x_y - \theta_3)/\theta_2$. Вторая производная логарифма функции правдоподобия по параметру θ_2 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_2^2} = \frac{\theta_1^2}{\theta_2^2} \left\{ \sum_{(1)} n_i \left[\frac{\Phi(t_{i-1}) \left(\frac{t_{i-1}s_{i-1}^2}{s_{i-1}^2+1} - \frac{s_{i-1}(s_{i-1}^2+2)}{(s_{i-1}^2+1)^{3/2}} \right) - \Phi(t_i) \left(\frac{t_i s_i^2}{s_i^2+1} - \frac{s_i(s_i^2+2)}{(s_i^2+1)^{3/2}} \right)}{\Phi(t_i) - \Phi(t_{i-1})} - \right. \right.$$

$$\left. \left. - \left(\frac{s_{i-1}\Phi(t_{i-1}) - s_i\Phi(t_i)}{\sqrt{s_{i-1}^2+1} - \sqrt{s_i^2+1}} \right)^2 \right] - \sum_{(2)} \sum_{j=1}^n \left[\frac{t_y s_y (s_y^2+2)}{\theta_1(s_y^2+1)^{3/2}} + \frac{s_y^2 - 1}{\theta_1^2(s_y^2+1)^2} + \frac{s_y^2}{s_y^2+1} \right] \right\}$$

Уравнение правдоподобия по параметру θ_3 имеет вид

$$\frac{\theta_1}{\theta_2} \left\{ \sum_{(1)} n_i \frac{\frac{\Phi(t_{i-1})}{\sqrt{s_{i-1}^2+1}} - \frac{\Phi(t_i)}{\sqrt{s_i^2+1}}}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} \sum_{j=1}^n \left[\frac{t_y}{\sqrt{s_y^2+1}} + \frac{s_y}{\theta_1(s_y^2+1)} \right] \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_3 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \frac{\theta_1^2}{\theta_2^2} \left[\sum_{(1)} n_i \left[\frac{\varphi(t_{i-1}) \left(\frac{t_{i-1}}{s_{i-1}^2 + 1} + \frac{s_{i-1}}{\theta_1(s_{i-1}^2 + 1)^{3/2}} \right) - \varphi(t_i) \left(\frac{t_i}{s_i^2 + 1} + \frac{s_i}{\theta_1(s_i^2 + 1)^{3/2}} \right)}{\Phi(t_i) - \Phi(t_{i-1})} \right] - \right.$$

$$\left. \left[\frac{\frac{\varphi(t_{i-1})}{\sqrt{s_{i-1}^2 + 1}} - \frac{\varphi(t_i)}{\sqrt{s_i^2 + 1}}}{\Phi(t_i) - \Phi(t_{i-1})} \right]^2 \right] - \sum_{(2)} \sum_{j=1}^n \left[\frac{t_j s_j}{\theta_1(s_j^2 + 1)^{3/2}} + \frac{s_j^2 - 1}{\theta_1^2(s_j^2 + 1)^2} - \frac{1}{s_j^2 + 1} \right]$$

При одновременном оценивании нескольких параметров максимизируется логарифм функции правдоподобия

$$\begin{aligned} \ln L = & \sum_{(1)} n_i \ln [\Phi(t_i) - \Phi(t_{i-1})] + \ln \frac{\theta_1}{\theta_2 \sqrt{2\pi}} \sum_{(2)} n_i - \\ & - \sum_{(3)} \sum_{j=1}^n \left[\frac{t_j^2}{2} + \frac{\ln(s_j^2 + 1)}{2} \right] \end{aligned}$$

и оценки являются решением соответствующей системы уравнений правдоподобия.

Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \sum_{(1)} n_i \times \sum_{(1)} \frac{[\varphi(t_i) - \varphi(t_{i-1})]^2}{\Phi(t_i) - \Phi(t_{i-1})} + \sum_{(2)} n_i,$$

$$J_{11} = J(\theta_1) = \sum_{(1)} n_i \times \frac{1}{\theta_1^2} \sum_{(1)} \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]^2}{\Phi(t_i) - \Phi(t_{i-1})} + \frac{2 + \theta_0^2}{\theta_1^2} \sum_{(2)} n_i,$$

$$J_{01} = J(\theta_0, \theta_1) = \sum_{(1)} n_i \times \frac{1}{\theta_1} \sum_{(1)} \left\{ \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]}{\Phi(t_i) - \Phi(t_{i-1})} \times \right.$$

$$\left. \times [\varphi(t_i) - \varphi(t_{i-1})] \right\} - \frac{\theta_0}{\theta_1} \sum_{(2)} n_i,$$

$$J_{12} = J(\theta_1, \theta_2) = \sum_{(1)} n_i \times \frac{\theta_1^2}{\theta_2^2} \sum_{(1)} \frac{\left(\frac{s_{i-1}\varphi(t_{i-1})}{\sqrt{s_{i-1}^2 + 1}} - \frac{s_i\varphi(t_i)}{\sqrt{s_i^2 + 1}} \right)^2}{\Phi(t_i) - \Phi(t_{i-1})} +$$

$$+ \frac{1}{\theta_2^2} \left(1 + \theta_1^2 - \theta_1^3 \int_{-\infty}^{\infty} \frac{\varphi(t)}{(s^2 + 1)^{3/2}} ds \right) \sum_{(2)} n_i.$$

$$\text{ где } t = \theta_0 + \theta_1 \ln(s + \sqrt{s^2 + 1}).$$

$$J_{02} = J(\theta_0, \theta_2) = \frac{\theta_1}{\theta_2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[\varphi(t_i) - \varphi(t_{i-1})] \left[\frac{\varphi(t_{i-1})s_{i-1}}{\sqrt{s_{i-1}^2 + 1}} - \frac{\varphi(t_i)s_i}{\sqrt{s_i^2 + 1}} \right]}{\Phi(t_i) - \Phi(t_{i-1})} - \right.$$

$$\left. - \sum_{(2)} n_i \times \theta_1 \int_{-\infty}^{\infty} \frac{\varphi(t)}{(s^2 + 1)^{3/2}} ds \right\},$$

$$J_{12} = J(\theta_1, \theta_2) = \frac{1}{\theta_2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]}{\Phi(t_i) - \Phi(t_{i-1})} \times \right.$$

$$\left. \times \left[\frac{\varphi(t_{i-1})s_{i-1}}{\sqrt{s_{i-1}^2 + 1}} - \frac{\varphi(t_i)s_i}{\sqrt{s_i^2 + 1}} \right] \right\} + \int_{-\infty}^{\infty} \frac{s\varphi(t)}{s^2 + 1} \left(\theta_0\theta_1 - \frac{2}{\sqrt{s^2 + 1}} \right) ds \sum_{(2)} n_i,$$

$$\begin{aligned}
J_{33} = J(\theta_3) &= \sum_{(1)} n_i \times \frac{\theta_1^2}{\theta_2^2} \sum_{(1)} \frac{\left(\frac{\varphi(t_{i-1})}{\sqrt{s_{i-1}^2 + 1}} - \frac{\varphi(t_i)}{\sqrt{s_i^2 + 1}} \right)^2}{\Phi(t_i) - \Phi(t_{i-1})} + \\
&\quad + \frac{\theta_1^3}{\theta_2^2} \int_{-\infty}^{\infty} \frac{\varphi(t)}{(s^2 + 1)^{3/2}} ds \sum_{(2)} n_i, \\
J_{03} = J(\theta_0, \theta_3) &= \frac{\theta_1}{\theta_2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[\varphi(t_i) - \varphi(t_{i-1})] \left[\frac{\varphi(t_{i-1})}{\sqrt{s_{i-1}^2 + 1}} - \frac{\varphi(t_i)}{\sqrt{s_i^2 + 1}} \right]}{\Phi(t_i) - \Phi(t_{i-1})} \right\} - \\
&\quad - \frac{\theta_1^2}{\theta_2} \int_{-\infty}^{\infty} \frac{\varphi(t)}{s^2 + 1} ds \sum_{(2)} n_i, \\
J_{13} = J(\theta_1, \theta_3) &= \frac{1}{\theta_2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]}{\Phi(t_i) - \Phi(t_{i-1})} \times \right. \\
&\quad \left. \times \left[\frac{\varphi(t_{i-1})}{\sqrt{s_{i-1}^2 + 1}} - \frac{\varphi(t_i)}{\sqrt{s_i^2 + 1}} \right] + \theta_0 \theta_1 \int_{-\infty}^{\infty} \frac{\varphi(t)}{s^2 + 1} ds \sum_{(2)} n_i \right\}, \\
J_{23} = J(\theta_2, \theta_3) &= \frac{\theta_1^2}{\theta_2^2} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{\left[\frac{\varphi(t_{i-1}) s_{i-1}}{\sqrt{s_{i-1}^2 + 1}} - \frac{\varphi(t_i) s_i}{\sqrt{s_i^2 + 1}} \right]}{\Phi(t_i) - \Phi(t_{i-1})} \times \right.
\end{aligned}$$

$$\times \left[\frac{\varphi(t_{i-1})}{\sqrt{s_{i-1}^2 + 1}} - \frac{\varphi(t_i)}{\sqrt{s_i^2 + 1}} \right] + \int_{-\infty}^{\infty} \frac{s\varphi(t)}{(s^2 + 1)^{3/2}} \left(\theta_1 + \frac{5}{\theta_1(s^2 + 1)} \right) ds \sum_{j=1}^{1000} n_j \Bigg\},$$

$$\text{и } J_{01} = J_{10}, J_{02} = J_{20}, J_{12} = J_{21}, J_{03} = J_{30}, J_{13} = J_{31}, J_{23} = J_{32}.$$

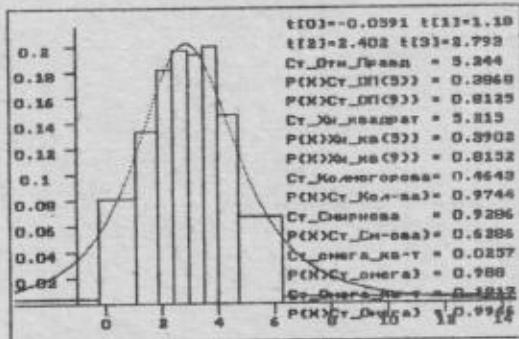


Рис. 3.25. Гистограмма и плотность распределения Su-Джонсона

Результаты оценивания параметров распределения Su-Джонсона и последующей проверки критериев согласия приведены на рис. 3.25. Выборка объемом 1000 наблюдений моделировалась с параметрами $\theta_0 = 0$, $\theta_1 = 1$, $\theta_2 = 2$, $\theta_3 = 3$.

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 $\theta_1, \theta_2 \in (0, \infty)$, 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

$$\frac{\theta_2}{\theta_1} \left\{ \sum_{(1)} n_i \frac{e^{-|t_{i-1}|^{\theta_2}} - e^{-|t_i|^{\theta_2}}}{\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}} + \sum_{(2)} \sum_{j=1}^{n_i} \frac{|t_j|^{\theta_2}}{t_j} \right\} = 0,$$

где $t_i = \frac{x_i - \theta_0}{\theta_1}$, $t_j = \frac{x_j - \theta_0}{\theta_1}$, $\gamma_i = \gamma(1/\theta_2, |t_i|^{\theta_2}) = \int_0^{|t_i|^{\theta_2}} y^{1/\theta_2 - 1} e^{-y} dy$ - значения неполной гамма-функции. Вторая производная логарифма функции правдоподобия по параметру θ_0 определяется выражением

$$\frac{\partial^2 \ln L}{\partial \theta_0^2} = \frac{\theta_2^2}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{\frac{|t_{i-1}|^{\theta_2}}{t_{i-1}} e^{-|t_{i-1}|^{\theta_2}} - \frac{|t_{i-1}|^{\theta_2}}{t_{i-1}} e^{-|t_i|^{\theta_2}}}{\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}} - \right. \right. \\ \left. \left. - \left(\frac{e^{-|t_{i-1}|^{\theta_2}} - e^{-|t_i|^{\theta_2}}}{\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}} \right)^2 \right] + \left(\frac{1}{\theta_2} - 1 \right) \sum_{(2)} \sum_{j=1}^{n_i} |t_j|^{\theta_2-2} \right\}.$$

Оценка параметра θ_1 находится в качестве решения уравнения правдоподобия

$$\frac{\theta_1}{\theta_0} \left\{ \sum_{(1)} n_i \frac{|t_{i-1}| e^{-|t_{i-1}|^{\theta_2}} - |t_i| e^{-|t_i|^{\theta_2}}}{\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}} + \sum_{(2)} \sum_{j=1}^{n_i} \left(|t_j|^{\theta_2} - \frac{1}{\theta_2} \right) \right\} = 0.$$

Вторая производная логарифма функции правдоподобия по параметру θ_1 имеет вид

$$\frac{\partial^2 \ln L}{\partial \theta_1^2} = \frac{\theta_2^2}{\theta_1^2} \left\{ \sum_{(1)} n_i \left[\frac{t_{i-1} e^{-|t_{i-1}|^{\theta_2}} \left(|t_{i-1}|^{\theta_2} - \frac{2}{\theta_2} \right) - t_i e^{-|t_i|^{\theta_2}} \left(|t_i|^{\theta_2} - \frac{2}{\theta_2} \right)}{\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}} - \right. \right.$$

$$\left. \left. - \left(\frac{t_{i-1} e^{-|t_{i-1}|^{\theta_2}} - t_i e^{-|t_i|^{\theta_2}}}{\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}} \right)^2 \right] + \sum_{(2)} \sum_{j=1}^n \left[\frac{1}{\theta_2^2} - \left(\frac{1}{\theta_2} + 1 \right) |t_j|^{\theta_2} \right] \right\}.$$

Оценка параметра θ_2 находится в качестве решения уравнения правдоподобия

$$\sum_{(1)} n_i \frac{\frac{1}{\theta_2^2} (\gamma_{i-1} \cdot \text{sign } t_{i-1} - \gamma_i \cdot \text{sign } t_i) + t_i e^{-|t_i|^{\theta_2}} \ln |t_i| - t_{i-1} e^{-|t_{i-1}|^{\theta_2}} \ln |t_{i-1}|}{\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}} +$$

$$+ \frac{1}{\theta_2^2} \psi \left(\frac{1}{\theta_2} \right) \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right) + \sum_{(2)} \sum_{j=1}^n \left(\frac{1}{\theta_2} - |t_j|^{\theta_2} \ln |t_j| \right) = 0,$$

где $\gamma_i = \gamma \left(1 / \theta_2, |t_i|^{\theta_2} \right) = \int_0^{|t_i|^{\theta_2}} y^{\theta_2-1} e^{-y} \ln y dy$ - 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

$$\frac{\partial^2 \ln L}{\partial \theta_2^2} = \sum_{(1)} n_i \left\{ \left[-\frac{2}{\theta_2^3} (\gamma_{i-1} \cdot \text{sign } t_{i-1} - \gamma_i \cdot \text{sign } t_i) + \right. \right.$$

$$\left. \left. + \frac{1}{\theta_2^2} \left[\text{sign } t_{i-1} \left(-\frac{\gamma_{i-1}}{\theta_2^2} + \theta_2 t_{i-1} e^{-|t_{i-1}|^{\theta_2}} \ln^2 |t_{i-1}| \right) - \text{sign } t_i \left(-\frac{\gamma_i}{\theta_2^2} + \theta_2 t_i e^{-|t_i|^{\theta_2}} \ln^2 |t_i| \right) \right] \right\} +$$

$$+ t_{i-1} |t_{i-1}|^{\theta_2} e^{-\gamma_{i-1}\theta_2} \ln^2 |t_{i-1}| - t_i |t_i|^{\theta_2} e^{-\gamma_i\theta_2} \ln^2 |t_i| \Big] / (\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}) +$$

$$+ \left(\frac{\frac{1}{\theta_2^2} (\gamma_{i-1} \cdot \text{sign } t_{i-1} - \gamma_i \cdot \text{sign } t_i) + t_i e^{-\gamma_i\theta_2} \ln |t_i| - t_{i-1} e^{-\gamma_{i-1}\theta_2} \ln |t_{i-1}|}{\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}} \right)^2 \Bigg)$$

$$- \frac{1}{\theta_2^2} \left[2\psi\left(\frac{1}{\theta_2}\right) + \frac{1}{\theta_2} \psi'\left(\frac{1}{\theta_2}\right) \right] \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right) - \sum_{(2)} \sum_{j=1}^n \left(\frac{1}{\theta_2^2} + |t_j|^{\theta_2} \ln^2 |t_j| \right).$$

$$\text{где } \gamma_i^* = \gamma^* \left(1/\theta_2, |t_i|^{\theta_2} \right) = \int_0^{\theta_2} y^{1/\theta_2 - 1} e^{-y} \ln^2 y dy.$$

При одновременном оценивании нескольких параметров максимизируется логарифм функции правдоподобия

$$\ln L = \sum_{(1)} n_i \ln (\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}) +$$

$$+ [\ln \theta_2 - \ln(2\theta_2 \Gamma(1/\theta_2))] \left(\sum_{(1)} n_i + \sum_{(2)} n_i \right) - \sum_{(2)} \sum_{j=1}^n |t_j|^{\theta_2}$$

и оценки являются решением соответствующей системы уравнений правдоподобия.

Элементы информационной матрицы Фишера определяются соотношениями

$$J_{00} = J(\theta_0) = \frac{\theta_0^2}{\theta_0^2 \Gamma(1/\theta_0)} \left\{ \sum_{(1)} n_i \times \sum_{(1)} \frac{2(e^{-\gamma_{i-1}\theta_0} - e^{-\gamma_i\theta_0})^2}{\gamma_i \cdot \text{sign } t_i - \gamma_{i-1} \cdot \text{sign } t_{i-1}} + \Gamma\left(2 - \frac{1}{\theta_0}\right) \sum_{(1)} n_i \right\}.$$

мирования 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 распределением, моделируется по формуле

$$X = \theta_1 - \frac{1}{\theta_0} \ln Y.$$

- Для полунармального распределения -

$$X = \theta_0 |Z|.$$

- Псевдослучайная величина, распределенная по закону Рэлея,

$$X = \theta_0 \sqrt{-2 \ln Y}.$$

- Для распределения Максвелла -

$$X = \theta_0 \sqrt{\sum_{i=1}^3 Z_i^2}.$$

- Псевдослучайные величины, распределенные в соответствии с законом модуля многомерной нормальной величины порядка m , генерируются в соответствии с соотношением

$$X = \theta_0 \sqrt{\sum_{i=1}^m Z_i^2}.$$

- Для распределения Парето -

$$X = \frac{\theta_1}{(1-Y)^{1/\theta_0}}.$$

- Распределение Эрланга порядка m моделируется в соответствии

с соотношением

$$X = \theta_1 - \frac{1}{\theta_0} \ln \left(\prod_{i=1}^n Y_i \right).$$

- Псевдослучайные числа по закону Лапласа моделируются по формуле

$$X = \begin{cases} \theta_0 + \frac{1}{\theta_0} \ln(2Y), & Y < 0.5; \\ \theta_0 - \frac{1}{\theta_0} \ln(2(1-Y)), & Y > 0.5; \\ \theta_0, & Y = 0. \end{cases}$$

- 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

$$X = \theta_1 + \theta_0 \lg[\pi(Y - 0.5)]$$

- Для логистического распределения

$$X = \theta_0 + \theta_1 \ln \left[\frac{1-Y}{Y} \right].$$

• Для распределения *Вейбулла*

$$X = \theta_2 + \theta_1 (-\ln Y)^{\theta_0}.$$

• Псевдослучайные величины в соответствии с распределением минимального значения генерируются по формуле

$$X = \theta_0 + \theta_1 \ln[-\ln(1-Y)].$$

• По закону максимального значения

$$X = \theta_0 - \theta_1 \ln[-\ln Y].$$

• Выборочные значения по двойному показательному распределению имитируются в соответствии с соотношением

$$X = -\frac{1}{\theta_0} \ln \left[-\frac{1}{\theta_1} \ln Y \right].$$

• Выборка для распределения *Накагами* моделируется по формулам

$$X = \theta_2 + \theta_0 \sqrt{\frac{1}{\theta_1} \left[(\beta - 1) \ln(Y^{1/(1-C)}) - \ln \left(\prod_{i=1}^{\lceil \theta_0 \rceil} Y_i \right) \right]},$$

где $C = \theta_1 - [\theta_1]$, $[\theta_1]$ - целая часть θ_1 , и $\beta = Y_1^{1/\theta_0} / (Y_1^{1/\theta_0} + Y_2^{1/\theta_0})$ - случайная величина, отдельно генерируемая в соответствии с бета-распределением.

• Для гамма-распределения -

$$X = \theta_2 + \frac{1}{\theta_1} \left[(\beta - 1) \ln(Y^{1/(1-C)}) - \ln \left(\prod_{i=1}^{\lceil \theta_0 \rceil} Y_i \right) \right],$$

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 $S = Y_1^{1/\theta_2} / (Y_1^{1/\theta_2} + Y_2^{1/\theta_1})$.

- Аналогично, для стандартного бета-распределение II рода имитируемые значения получаются в виде

$$X = \theta_2 + \frac{S}{1-S},$$

- Так же для бета-распределения II рода

$$X = \theta_1 + \frac{S}{\theta_1(1-S)}.$$

- 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

$$X = \theta_3 + \theta_2 e^S.$$

- А 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 определения её значений $X = F^{-1}(Y)$ численно решается уравнение $Y = F(X)$.

4. Меню системы

4.1. Выборка

4.1.1. Файл

В данном пункте меню задается имя существующего входного файла, содержащего выборку случайной величины, подвергаемой анализу.

Информация во входном файле представляется в следующем виде:

- Первая строка должна содержать некоторую текстовую информацию о выборке.
- Во второй строке указывается тип выборки:
 - 0 - негруппированная;
 - 1 - частично группированная;
 - 3 - группированная.
- Далее идет число интервалов (для группированной или частично группированной выборки) и число индивидуальных наблюдений (для негруппированной или частично группированной выборки).
- Далее, если выборка группированная или частично группированная, через разделитель должны содержаться количества наблюдений n_i , попавших в интервалы группирования наблюдений, затем ее граничные точки интервалов x_i , $i = 1, (k-1)$, где k - количество интервалов группирования.
- Далее негруппированная часть выборки - индивидуальные значения измерений x_i (если выборка негруппированная или частично группированная).

Примечание. В перечень граничных точек должны входить только значения, разделяющие интервалы: левая x_0 и правая x_k границы области определения случайной величины во входном файле не должны задаваться. Т.е., если, например, для показательного распределения левый интервал значений [0, 0.209], то первая граничная точка в файле должна быть 0.209.

В случае частично группированной выборки с числом интервалов группирования 5 и объемом негруппированной части выборки в 50 индивидуальных измерений входной файл будет, например, иметь вид:

Частично группированная выборка: Эксперимент 3

1	5	50					
751	2247	3998	2254	744			
0.209	0.4470	0.5530	0.791				
0.01	0.01	0.01	0.01	0.04	0.17	0.48	0.22
0.22	0.25	0.25	0.29	0.42	0.46	0.47	0.56
0.59	0.67	0.68	0.70	0.72	0.76	0.78	0.83
0.85	0.87	0.93	0.99	0.01	0.01	0.02	0.03
0.05	0.33	0.34	0.37	0.47	0.50	0.52	0.54
0.59	0.71	0.90	0.10	0.35	0.46	0.50	0.73
0.07	0.03						

4.1.2. Новая

Задаются параметры новой выборки:

- тип выборки;
- число индивидуальных наблюдений;
- число интервалов группирования;
- имя файла.

Если нажать кнопку *Моделирование*, то новая выборка будет смоделирована в соответствии с текущим распределением и текущими параметрами.

Если нажать кнопку *Ввод*, то выборочные значения будут запрашиваться в режиме диалога.

4.1.3. Сортировка

В случае негруппированной выборки осуществляется запись в файл отсортированной по возрастанию выборки.

4.1.4. Группирование

Если возникла необходимость сгруппировать выборку по интервалам с записью группированной выборки в другой файл для последующей обработки, это можно выполнить, используя данный пункт меню. Количество интервалов и вид (способ) группирования задаются в меню *Опции/Тип группирования*. Следует указать имя файла, в который будет записана сгруппированная выборка. По умолчанию предлагается файл с тем же именем. Вы можете задать любое другое имя. В случае *асимптотически оптимального группирования* используется текущее распределение с текущими параметрами.

Подвергать группированию желательно выборки большого или очень большого объема. В этом случае, хотя, с одной стороны, происходит определенная потеря информации, зато, с другой стороны, в некоторых ситуациях могут быть снижены вычислительные затраты при оценивании и, кроме того, получаемые по группированной выборке оценки будут менее зависимы от аномальных наблюдений, т.е. получим *робастные* оценки.

4.1.5. Характеристики

В окно на экран выводятся выборочные характеристики.

4.2. Распределение

4.2.1. Закон

Осуществляется выбор текущего распределения. В комплекс включены следующие распределения.

1. Экспоненциальное распределение (показательное).
2. Полунормальное распределение.
3. Распределение Рэлея.
4. Распределение Максвелла.
5. Распределение модуля многомерной нормальной величины.
6. Распределение Парето.
7. Распределение Эрланга.

8. Распределение Лапласа.
9. Нормальное распределение.
10. Логарифмически (ln) нормальное распределение.
11. Логарифмически (lg) нормальное распределение.
12. Распределение Коши.
13. Логистическое распределение.
14. Распределение Вейбулла.
15. Распределение минимального значения.
16. Распределение максимального значения.
17. Двойное показательное распределение.
18. Распределение Накагами.
19. Гамма-распределение.
20. Бета-распределение I рода.
21. Стандартное бета-распределение II рода.
22. Бета-распределение II рода.
23. Распределение Sb-Джонсона.
24. Распределение Sl-Джонсона.
25. Распределение Su-Джонсона.
26. Класс экспоненциальных распределений (двустороннее экспоненциальное).

4.2.2. Параметры

В этом пункте меню можно задать параметры текущего распределения, а также отметить оцениваемые параметры.

4.2.3. Оценивание

Осуществляется оценивание параметров текущего распределения по методу максимального правдоподобия.

4.2.4. Согласие

Осуществляется проверка согласия по критериям, выбранным в меню *Опции/Критерии*.

4.2.5. Вероятности

Для текущего распределения вычисляются вероятности попадания в интервалы значений, т.е. вероятности вида:

- $P\{X < a\}$;
- $P\{a < X < b\}$;
- $P\{X > b\}$.

Значения левой и правой границ этого интервала устанавливаются в соответствующих полях меню.

4.2.6. Идентификация

В соответствии с набором законов, указанных в меню *Опции/распределения*, осуществляется перебор распределений и выбирается закон, наиболее хорошо согласующийся с данной выборкой по совокупности заданных критерии. В файл результатов выдается таблица, в которой по каждому критерию указывается наиболее предпочтительный закон распределения из рассматриваемого набора. Если, что бывает достаточно часто, результаты анализа по разным критериям указывают на предпочтительность различных законов, Вы можете остановить свой выбор на распределении, на которое указывает компромиссный критерий.



4.3. Опции

4.3.1. Файл результатов

При выборе этого пункта необходимо задать имя файла, в который будут выводиться результаты. По умолчанию предлагается выбрать файл с тем же именем, что и выборка, но с расширением *res*.

При вычислении оценок и проверке гипотез о согласии, если не менять имя выходного файла, новые результаты будут дописываться в конец старого файла.

При идентификации результаты записываются в начало файла результатов, старое содержимое при этом теряется.

4.3.2. Распределения

В данном пункте меню указываются распределения, учитываемые при идентификации, т.е., для каждого отмеченного закона по данной выборке будут оценены его параметры и проверены гипотезы о согласии.

4.3.3. Критерии

В этом пункте меню вы можете задать критерии, в соответствии с которыми будет осуществляться проверка согласия введенной выборки с текущим распределением.

Система предоставляет возможность проверять гипотезы о согласии по следующим критериям:

- отношения правдоподобия;
- χ^2 -Пирсона;
- Колмогорова;
- Смирнова;
- ω^2 и Ω^2 - Мизеса.

Критерии Колмогорова, Смирнова, ω^2 и Ω^2 - Мизеса проверяются только по негруппированной исходной выборке.

При проверке согласия по критериям отношения правдоподобия и χ^2 -Пирсона негруппированная выборка группируется. Тип группирования и число интервалов задаются в меню *Опции/Тип группирования*.

Для каждого критерия вычисляется вероятность того, что значение статистики может превысить данное значение при справедливости гипотезы о согласии. Такая вероятность для критериев отношения правдоподобия и Пирсона вычисляется для двух чисел степеней свободы χ^2 -распределения: $k-p-1$, это соответствует случаю, когда p параметров оценивалось по данной выборке, где k - количество интервалов группирования; $k-1$, когда по данной выборке не оценивались параметры.

Гипотеза о согласии не будет приниматься, если задаваемый Вами (исследователем) уровень значимости α превысит вычисленные вероятности. При негруппированной исходной выборке Вы можете

использовать все критерии и остановить свой выбор на том распределении, для которого будет наблюдаться наилучшее согласие по совокупности критериев.

4.3.4. Тип группирования

Задается способ группирования и число интервалов группирования. Возможные варианты:

- оптимальное группирование (осуществляется с использованием таблиц асимптотически оптимального группирования, если они есть для данного распределения, в противном случае соответствующая оптимизационная задача решается непосредственно перед проверкой гипотез о согласии);
- разбиение выборки на интервалы равной частоты (вероятности);
- разбиение размаха выборки на интервалы равной длины;
- разбиение выборки на интервалы заданной равной длины.

Здесь же можно задать опцию *Группировать при оценивании*. В этом случае перед оцениванием выборка группируется. В результате получаются оценки, более устойчивые к аномальным наблюдениям.

4.3.5. Параметры графика

К параметрам графика относятся :

- цвет изображения;
- размер масштабной сетки;
- название графика.

4.4. Просмотр

4.4.1. Выборка

Осуществляется просмотр файла с выборкой.

4.4.2. Результат

Осуществляется просмотр файла результатов.

4.4.3. График

Осуществляется построение графика функции плотности и гистограммы распределения или теоретической и эмпирической функций распределения для закона, для которого последний раз вычислялись оценки. На график выдаются результаты проверки гипотез о согласии по критериям отношения правдоподобия, χ^2 Пирсона, Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса. По каждому из критериев на графике отображаются значения вычисленных статистик и вероятности вида $P\{X > \text{Значение_статистики}\}$. Гипотеза о согласии по выбранному критерию не будет отвергаться, если задаваемый исследователем уровень значимости α не превысит значения соответствующей вероятности.

Если Вы собираетесь выводить график с результатами анализа на принтер, перед входом в систему запустите на выполнение из каталога DOS программу *graphics.com*.

Для печати графика приведите его к нужному масштабу и положению и, если Ваш принтер готов к работе, нажмите клавишу *(PrintScreen)*.

При выполнении программы в среде Windows по клавише *(PrintScreen)* Вы можете сохранить изображение и, в дальнейшем отредактировав, экспортить его в соответствующий редактор.

Примечание 1. В данной версии системы при частично группированных выборках для построения гистограммы и проверки согласия по критериям отношения правдоподобия и χ^2 Пирсона Вы должны в исходной выборке сами задать граничные точки интервалов и количества наблюдений в них для группированной части выборки (в том числе нулевые значения).

Примечание 2. Данная версия графического обеспечения предназначена для работы на персональной ЭВМ только при наличии EGA и VGA драйверов.

Заключение

В книге рассмотрены методы оценивания параметров непрерывных распределений по негруппированным, группированным, частично группированным и цензурированным выборкам, применение асимптотически оптимального группирования при оценивании параметров распределений и в критериях согласия, повышающее качество статистических выводов, кратко изложены основы алгоритмического обеспечения программной системы, приведены рекомендации и инструкции по использованию системы.

Описание меню и рекомендации, изложенные начиная с 4-го раздела, соответствуют последней версии программной системы, которая продолжает развиваться и, следовательно, будут определённые отличия, связанные с частными вопросами использования или перечнем решаемых задач.

Дальнейшее развитие программной системы связано с расширением перечня решаемых задач. В настоящий момент в систему включено программное обеспечение для решения задач асимптотически оптимального группирования в темпе обработки экспериментальных данных для распределений Джонсона, гамма- и бета-, реализуется для класса экспоненциальных распределений. Начата работа по расширению системы за счёт обработки по частично группированным выборкам усечённых законов и смесей распределений. Это оказывается на порядок более сложным, так как существенно возрастают размерности задач. Но эти трудности будут оправданы увеличением круга приложений, для которых могут быть получены положительные результаты. Намечено объединение с программной системой корреляционного анализа многомерных случайных величин, что также несёт ряд положительных моментов.

Естественно, что на программной системе, особенно её сервисных функциях, будет отражаться развитие средств вычислительной техники и системного программного обеспечения.

Рекомендуемая литература

1. Куллдорф Г. Введение в теорию оценивания по группированным и частично группированным выборкам. - М.: Наука, 1966. - 176 с.
2. Рао С.Р. Линейные статистические методы и их применения. - М.: Наука, 1968. - 548 с.
3. Крамер Г. Математические методы статистики. - М.: Мир, 1975. 648 с.
4. Rao C.R. Criteria of estimation in large samples // Sankhya, 1962. - Vol. 25. - P. 189-206.
5. Бодин Н.А. Оценка параметров распределений по группированным выборкам. // Тр. ин-та им. В.А. Стеклова АН СССР. - 1970. - Т. 111. - С. 110-154.
6. 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.
7. Денисов В.И. Математическое обеспечение системы "ЭВМ-экспериментатор". - М.: Наука, 1977. - 251 с.
8. Лемешко Б.Ю. Об оценивании параметров распределений по группированным наблюдениям // Вопросы кибернетики. - М., 1977. - Вып. 30. - С. 80-96.
9. Лемешко Б.Ю. Об оценивании по частично группированным выборкам параметров распределения Эрланга // Применение ЭВМ в оптимальном планировании и проектировании. - Новосибирск, 1982. - С. 90-93
10. Денисов В.И., Зачепа Г.Г., Лемешко Б.Ю. Оптимальное группирование, оценка параметров и планирование регрессионных экспериментов: В 2 ч. / Новосиб. гос. техн. ун-т. - Новосибирск, 1993. - 347 с.
11. Кедалл М., Стьюарт А. Статистические выводы и связи. - М.: Наука, 1973. - 900 с.
12. Кокс Д., Хинкли Д. Теоретическая статистика. - М.: Мир, 1978. - 560 с.

13. Cox D.R. Note on grouping // J. of the American Statistical Association, 1957. - Vol. 52, № 280. - P. 543-547.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. Методические рекомендации по планированию экспериментов и обработке экспериментальных данных при исследовании надежности и качества функционирования систем "человек-техника" / В.И. Денисов, Б.Ю. Лемешко, Е.Б. Цой и др. - Л., ЛЭТИ, 1978. - 46 с.
20. Денисов В.И., Лемешко Б.Ю. Оптимальное группирование при обработке экспериментальных данных // Измерительные информационные системы. - Новосибирск, 1979. - С. 5-14.
21. Денисов В.И., Лемешко Б.Ю. Вычисление оценок параметров распределений с использованием таблиц асимптотически оптималь-

- ного группирования // Применение ЭВМ в оптимальном планировании и проектировании. - Новосибирск, 1981. - С. 3-17.
22. 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.
23. 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.
24. 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. - № 10/1989. - "Statistics for grouped Observations". - S. 63-81.
25. Denisov V.I., Lemeshko B.Yu., Tsöi E.B. Estimation of unknown parameters of onedimensional distributions with partially grouped data // Wissenschaftliche Schriftenreihe der Technischen universitat Karl-Marx-Stadt. - № 10/1989. - "Statistics for grouped Observations". - S. 6-21.
26. Информационно-управляющие человеко-машинные системы. Исследование, проектирование, испытания: Справочник / Под общ. ред. А.И. Губинского и В.Г. Енграфова. -М.: Машиностроение, 1993. - 528 с.
27. Хьюбер П. Робастность в статистике. - М.: Мир, 1984. - 303 с.
28. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. - М.: Наука, 1965. - 464 с.
29. Губарев В.В. Вероятностные модели: Справочник: В 2 ч. / Новосиб. электротехн. ин-т. - Новосибирск, 1992. - 422 с.
30. Ермаков С.М., Михайлов Г.А. Статистическое моделирование. - М.: Наука, 1982. - 296 с.

С о д е р ж а н и е

Предисловие	3
1. Введение	5
2. Теоретическое обоснование	7
2.1. Структура данных и методы оценивания	7
2.2. Метод максимального правдоподобия	12
2.3. Асимптотически оптимальное группирование	14
2.4. Связь критерииов согласия с оптимальным группированием	16
2.5. Робастность	21
2.6. Проверка гипотез	27
2.7. Критерии проверки гипотез	30
2.8. Многокритериальность задачи проверки гипотез о согласии	34
2.9. Алгоритмы вычисления оценок	37
3. Алгоритмическое обеспечение программной системы	39
3.1. Экспоненциальное распределение	40
3.2. Полунормальное распределение	41
3.3. Распределение Рэлея	43
3.4. Распределение Максвелла	44
3.5. Распределение модуля многомерного нормального вектора	46
3.6. Распределение Парето	47
3.7. Распределение Эрланга	49
3.8. Распределение Лапласа	51
3.9. Нормальное распределение	53
3.10. Логарифмически (ln) нормальное распределение	55
3.11. Логарифмически (lg) нормальное распределение	56
3.12. Распределение Коши	57
3.13. Логистическое распределение	60
3.14. Распределение Вейбулла	63
3.15. Распределение минимального значения	66
3.16. Распределение максимального значения	68
3.17. Двойное показательное распределение	71
3.18. Распределение Накагами	74

3.19. Гамма-распределение	77
3.20. Бета-распределение I рода	80
3.21. Стандартное бета-распределение II рода	84
3.22. Бета-распределение II рода	85
3.23. Распределение Sb-Джонсона	90
3.24. Распределение Sl-Джонсона	93
3.25. Распределение Su-Джонсона	96
3.26. Класс экспоненциальных распределений	102
3.27. Имитация выборочных значений	107
4. Меню системы	112
4.1. Выборка	112
4.1.1. Файл	112
4.1.2. Новая	113
4.1.3. Сортировка	113
4.1.4. Группирование	114
4.1.5. Характеристики	114
4.2. Распределение	114
4.2.1. Закон	114
4.2.2. Параметры	115
4.2.3. Оценивание	115
4.2.4. Согласие	115
4.2.5. Вероятности	116
4.2.6. Идентификация	116
4.3. Опции	116
4.3.1. Файл результатов	116
4.3.2. Распределения	117
4.3.3. Критерии	117
4.3.4. Тип группирования	118
4.3.5. Параметры графика	118
4.4. Просмотр	118
4.4.1. Выборка	118
4.4.2. Результат	118
4.4.3. График	119
Заключение	120
Рекомендуемая литература	121

Борис Юрьевич Лемешко

СТАТИСТИЧЕСКИЙ АНАЛИЗ
ОДНОМЕРНЫХ НАБЛЮДЕНИЙ
СЛУЧАЙНЫХ ВЕЛИЧИН

Программная система

Редактор И. А. Кескевич
Технический редактор Г. Е. Теляпникова

Лицензия № 020356 от 27.12.91. Подписано в печать 12.05.95.
Формат 60×84 1/16. Бумага оберточная. Тираж 200 экз. Уч.-изд. л. 7,3
Печ. л. 8,0. Изд. № 68. Заказ № 220 Цена договорная.

Отпечатано в типографии
Новосибирского государственного технического университета
630092, г. Новосибирск, пр. К. Маркса, 20.