

2. Применение критериев согласия типа χ^2

2.1. Введение

1

Проверка статистических гипотез о согласии эмпирических данных с теоретическим законом распределения с применением критериев согласия типа χ^2 обусловлена рядом условий, которые обеспечивают корректное решение задачи. К сожалению, не в каждом источнике, который используется исследователем в качестве руководства, находят отражение эти условия. Поэтому, несмотря на кажущуюся простоту, практика использования критериев согласия типа χ^2 изобилует примерами его некорректного или неэффективного применения, особенно при проверке сложных гипотез.

Анализ примеров «неудачного» применения критериев типа χ^2 позволяет выделить две группы причин, которые могут приводить к неверным статистическим выводам. Во-первых, это часто совершаемые принципиальные ошибки, при которых использование χ^2_{k-m-1} -распределения в качестве предельного оказывается неправомерным. Во-вторых, действия, использующие возможности критерия не наилучшим образом. В первом случае возрастает вероятность ошибки первого рода α (отклонить верную проверяемую гипотезу), во втором – вероятность ошибки второго рода β (принять проверяемую гипотезу при справедливости конкурирующей).

Число моделей непрерывных законов распределений, используемых в задачах статистического анализа (при контроле качества, исследованиях надежности и т. д.), немногим превышает 100, а для описания наблюдаемых случайных величин в прикладных исследованиях в основном применяют около 30 параметрических законов и семейств распределений.

Это не покрывает многообразия случайных величин, встречаемых на практике. Корректное приме-

нение критериев согласия часто приводит (и должно приводить) к отклонению гипотез о принадлежности выборки удобному и привычному закону распределения, например нормальному, так как законы реальных случайных величин, являющиеся следствием многочисленных причин, сложнее тех моделей, которые зачастую используют для их описания. Следовательно, и модели должны быть более сложными.

Целью первичной обработки экспериментальных наблюдений обычно является выбор закона распределения, наиболее хорошо описывающего случайную величину, выборку которой наблюдают. Насколько хорошо наблюдаемая выборка описывается теоретическим законом, проверяют с помощью различных критериев согласия. Цель проверки гипотезы о согласии опытного распределения с теоретическим – это стремление удостовериться в том, что данная модель теоретического закона не противоречит наблюдаемым данным и использование ее не приведет к существенным ошибкам при вероятностных расчетах. Некорректное использование критериев согласия может приводить к необоснованному принятию (чаще всего) или необоснованному отклонению проверяемой гипотезы.

Различают простые и сложные гипотезы о согласии. Простая проверяемая гипотеза имеет вид $H_0: f(x, \theta) = f(x, \theta_0)$, где $f(\cdot)$ – функция плотности, θ_0 – известный скалярный или векторный параметр теоретического распределения, с которым проверяют согласие. Сложная гипотеза имеет вид $H_0: f(x) \in \{f(x, \theta), \theta \in \Theta\}$, где Θ – пространство параметров. Если процесс вычисления оценки $\hat{\theta}$ скалярного или векторного параметра закона не опирается на ту же выборку, по которой проверяют гипотезу о согласии, то алгоритм применения критерия согласия при проверке сложной гипотезы не отличается от проверки простой гипотезы. В противном случае применение критерия согласия при проверке сложной гипотезы имеет существенные отличия.

Схема процедуры проверки гипотезы следующая. В соответствии с применяемым критерием согласия вычисляют значение S^* статистики критерия S как некоторой функции от выборки и теоретического закона распределения с плотностью $f(x, \theta_0)$ [или $f(x, \hat{\theta})$ при сложной гипотезе]. Для исполь-

зуемых на практике критериев асимптотические (предельные) распределения $G(S|H_0)$ соответствующих статистик при условии истинности гипотезы H_0 обычно известны. В общем случае для простых и сложных гипотез эти распределения различаются. Далее в принятой практике статистического анализа обычно полученное значение статистики S^* сравнивают с критическим значением S_α при заданном уровне значимости α . Нулевую гипотезу отвергают, если $S^* > S_\alpha$ (рис. 2.1).

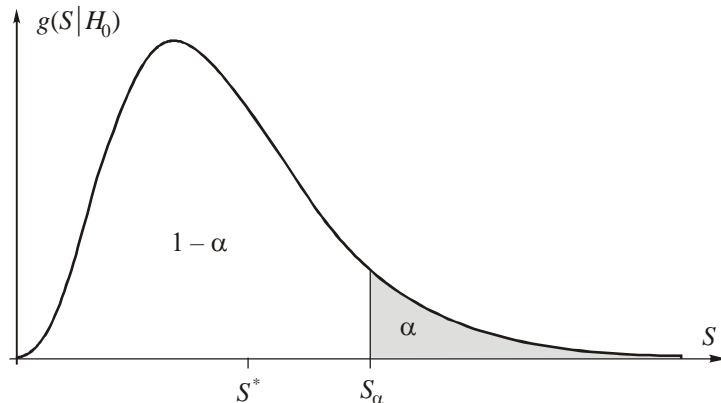


Рис. 2.1. Плотность распределения статистики при истинной гипотезе H_0

Критическое значение S_α , определяемое в случае одномерной статистики из уравнения

$$\alpha = \int_{S_\alpha}^{\infty} g(s|H_0)ds,$$

где $g(s|H_0)$ – условная плотность распределения статистики, обычно берут из соответствующей статистической таблицы или вычисляют.

Больше информации о степени согласия можно почерпнуть из «достигаемого уровня значимости»: величины вероятности возможного превышения полученного значения статистики при истинности ну-

левой гипотезы $P\{S > S^*\} = \int_{S^*}^{\infty} g(s|H_0) ds$. Именно эта вероятность позволяет судить о том, насколько

хорошо выборка согласуется с теоретическим распределением, так как по существу представляет собой вероятность истинности нулевой гипотезы (рис. 2.2). Гипотезу о согласии не отвергают, если $P\{S > S^*\} > \alpha$.

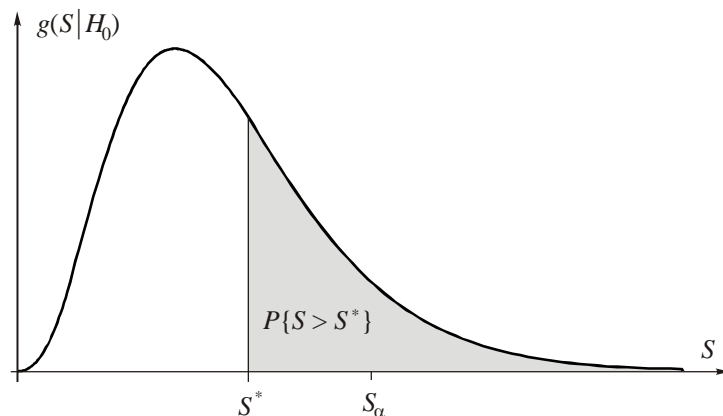


Рис. 2.2. Плотность распределения статистики при истинной гипотезе H_0 и достигаемый уровень значимости

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 двух видов: ошибка первого рода состоит в том, что отклоняют гипотезу H_0 , когда она верна; ошибка второго рода состоит в том, что принимают гипотезу H_0 , в то время как справедлива конкурирующая гипотеза H_1 . Величина α задает вероятность ошибки первого рода. Обычно в критериях согласия не рассматривают конкретную конкурирующую гипотезу. Тогда можно считать, что конкурирующая гипотеза имеет вид $H_1: f(x, \theta) \neq f(x, \theta_0)$. Если же гипотеза H_1 задана и имеет, например, вид $H_1: f(x, \theta) = f_1(x, \theta_1)$, то задание величины α определяет для используемого критерия проверки гипотез и вероятность ошибки второго рода β . На рис. 2.3 $g(s|H_0)$ отображает плотность распределения статистики S при истинности гипотезы H_0 , а $g(s|H_1)$ – плотность распределения при справедливости H_1 .

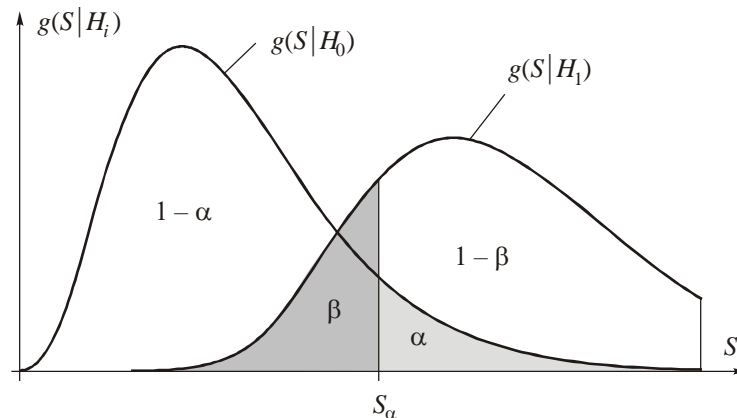


Рис. 2.3. Плотности распределения статистик при справедливости гипотез H_0 и H_1

Мощность критерия представляет собой величину $1-\beta$. Очевидно, что чем выше мощность используемого критерия при заданном значении α , тем лучше он различает гипотезы H_0 и H_1 . Особенно важно, чтобы этот критерий хорошо различал близкие конкурирующие гипотезы. Графически требование максимальной мощности критерия означает, что на рис. 2.3 плотности $g(s|H_0)$ и $g(s|H_1)$ должны быть максимально «раздвинуты».

2.2. Критерии типа χ^2 при простых гипотезах

Предположим, что $\xi_1, \xi_2, \dots, \xi_n$ – выборка значений наблюдаемой случайной величины объема n . Процедура проверки гипотез с использованием критериев типа χ^2 предусматривает группирование наблюдений. Область определения случайной величины разбивают на k непересекающихся интервалов граничными точками

$$x_0 < x_1 < \dots < x_{k-1} < x_k,$$

где x_0 – нижняя грань области определения случайной величины; x_k – верхняя грань. В соответствии с заданным разбиением подсчитывают число n_i выборочных значений, попавших в i -й интервал, и

вероятности попадания в интервал $P_i(\theta) = \int_{x_{i-1}}^{x_i} f(x, \theta) dx$, соответствующие теоретическому закону с

функцией плотности $f(x, \theta)$. При проверке простой гипотезы известны как вид функции плотности,

так и все параметры закона (известен скалярный или векторный параметр θ). При этом $n = \sum_{i=1}^k n_i$,

$\sum_{i=1}^k P_i(\theta) = 1$. В основе статистик, используемых в критериях согласия типа χ^2 , лежит измерение отклонений n_i / n от $P_i(\theta)$.

К критериям такого рода, в частности, относят критерий χ^2 Пирсона, критерий отношения правдоподобия [205] и критерии типа χ^2 [287, 288, 284, 1].

Статистику критерия согласия χ^2 Пирсона вычисляют по формуле

$$X_n^2 = n \sum_{i=1}^k \frac{(n_i / n - P_i(\theta))^2}{P_i(\theta)}. \quad (2.1)$$

8 В случае проверки простой гипотезы в пределе при $n \rightarrow \infty$ эта статистика подчиняется χ_r^2 -распределению с $r = k - 1$ степенями свободы, если верна нулевая гипотеза. Плотность χ_r^2 -распределения описывается формулой

$$g(s) = \frac{1}{2^{r/2} \Gamma(r/2)} s^{r/2-1} e^{-s/2}. \quad (2.2)$$

Если верна конкурирующая гипотеза H_1 и выборка соответствует распределению с плотностью $f_1(x, \theta_1)$ с параметром θ_1 , то эта же статистика в пределе подчиняется нецентральному χ_r^2 -распределению с тем же числом степеней свободы $r = k - 1$ и параметром нецентральности

$$v = n \sum_{i=1}^k \frac{(P_i^1(\theta_1) - P_i(\theta))^2}{P_i(\theta)}, \quad (2.3)$$

где $P_i^1(\theta_1)$ – вероятность попадания в интервал при справедливой гипотезе H_1 . Плотность нецентрального χ_r^2 -распределения имеет вид [205]

$$g(s, v) = \frac{e^{-(s+v)/2} s^{(r-2)/2}}{2^{r/2} \Gamma[(r-1)/2] \Gamma(1/2)} \sum_{k=0}^{\infty} \frac{v^k s^k}{(2k)!} B\left\{\frac{1}{2}(r-1), \frac{1}{2} + k\right\}, \quad (2.4)$$

где $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta) / \Gamma(\alpha + \beta)$ – бета-функция.

При заданном уровне значимости α нулевая гипотеза о согласии не должна быть отвергнута, если

9

$$P\left\{X_n^2 > X_n^{2*}\right\} = \frac{1}{2^{r/2} \Gamma(r/2)} \int_{X_n^{2*}}^{\infty} s^{r/2-1} e^{-s/2} ds > \alpha, \quad (2.5)$$

где X_n^{2*} – вычисленное в соответствии с формулой (2.1) значение статистики.

В критерии отношения правдоподобия использована статистика [205]

$$S_{\text{ОП}} = -2 \ln l = -2 \sum_{i=1}^k n_i \ln \left(\frac{P_i(\theta)}{n_i / n} \right), \quad (2.6)$$

которая при верной нулевой гипотезе также асимптотически рас-пределена как χ_r^2 с $r = k - 1$ степенями свободы. Если верна конку-рирующая гипотеза H_1 и выборка соответствует распределению с плотностью $f_1(x, \theta_1)$ и параметром θ_1 , мерой близости сравниваемых законов является величина

$$v = 2n \sum_{i=1}^k P_i^1(\theta_1) \ln \left(\frac{P_i^1(\theta_1)}{P_i(\theta)} \right). \quad (2.7)$$

2.3. Критерии типа χ^2 при сложных гипотезах

При справедливости H_0 в случае проверки сложной гипотезы и при условии, что оценки параметров находятся в результате минимизации статистики X_n^2 по этой же самой выборке, статистика X_n^2 асимптотически распределена как χ_r^2 с числом степеней свободы $r = k - m - 1$, где m – число оцененных параметров. Статистика X_n^2 имеет это же распределение, если в качестве метода оценивания выбирают метод максимального правдоподобия и оценки вычисляют по сгруппированным данным в результате максимизации по θ функции правдоподобия

$$L(\theta) = \gamma \prod_{i=1}^k P_i^{n_i}(\theta), \quad (2.8)$$

где γ – некоторая константа и $P(\theta) = \int_{x_{i-1}}^{x_i} f(x, \theta) dx$ – вероятность попадания наблюдения в i -й интервал значений, зависящая от θ .

При вычислении оценок максимального правдоподобия (ОМП) по негруппированным данным эта же статистика распределена как сумма независимых слагаемых $\chi_{k-m-1}^2 + \sum_{j=1}^m \lambda_j \xi_j^2$, где ξ_1, \dots, ξ_m – стандартные нормальные случайные величины, независимые одна от другой и от χ_{k-m-1}^2 , а $\lambda_1, \dots, \lambda_m$ – некоторые числа между 0 и 1 [287, 19, 341], представляющие собой корни уравнения

$$|(1-\lambda)\mathbf{J}(\theta) - \mathbf{J}_\Gamma(\theta)| = 0.$$

В данном уравнении $\mathbf{J}(\theta)$ – информационная матрица Фишера по негруппированным наблюдениям с элементами, определяемыми соотношением

$$J(\theta_l, \theta_j) = \int \left(\frac{\partial f(x, \theta)}{\partial \theta_l} \frac{\partial f(x, \theta)}{\partial \theta_j} \right) f(x, \theta) dx, \quad (2.9)$$

а $\mathbf{J}_\Gamma(\theta)$ – информационная матрица по группированным наблюдениям

$$\mathbf{J}_\Gamma(\theta) = \sum_{i=1}^k \frac{\nabla P_i(\theta) \nabla^T P_i(\theta)}{P_i(\theta)}. \quad (2.10)$$

Функция распределения статистики лежит между χ_{k-1}^2 и χ_{k-m-1}^2 -распределениями. В этом случае, принимая нулевую гипотезу, следует удостовериться, что статистика X_n^2 не превышает критических значений $\chi_{k-m-1, \alpha}^2$ и $\chi_{k-1, \alpha}^2$, где α – задаваемый уровень значимости. И если $\chi_{k-m-1, \alpha}^2 < X_n^2 < \chi_{k-1, \alpha}^2$, то, принимая или отклоняя гипотезу о согласии, можно с одинаковым риском совершить ошибку.

Все сказанное относится и к критерию отношения правдоподобия.

Влияние способа группирования на распределения этих статистик при использовании оценок максимального правдоподобия по негруппированным данным рассмотрено в разделе 2.7.

2.4. Статистика типа χ^2 Никулина

В работах [287, 288, 284, 1] предложено видоизменение стандартной статистики X_n^2 , при котором предельное распределение есть обычное χ_{k-1}^2 -распределение (число степеней свободы не зависит от числа оцениваемых параметров). Неизвестные параметры распределения $F(x, \theta)$ в этом случае следует оценивать по негруппированным данным методом максимального правдоподобия. При этом вектор $\mathbf{P} = (P_1, \dots, P_k)^T$ предполагают заданным и граничные точки интервалов определяют по соотношениям $x_i(\theta) = F^{-1}(P_1 + \dots + P_i)$, $i = \overline{1, (k-1)}$. Предложенная статистика имеет вид [291]

$$Y_n^2(\theta) = X_n^2 + n^{-1} a^T(\theta) \Lambda(\theta) a(\theta), \quad (2.11)$$

12

где X_n^2 вычислена по формуле (2.1); матрица

$$\Lambda(\theta) = \left\| J(\theta_l, \theta_j) - \sum_{i=1}^k \frac{w_{\theta_l i} w_{\theta_j i}}{P_i} \right\|^{-1},$$

элементы и размерность которой определяются оцениваемыми ком-понентами вектора параметров θ ; $J(\theta_l, \theta_j)$ – элементы инфор-мационной матрицы $\mathbf{J}(\theta)$ по негруппированным данным (2.9);

$a_{\theta_l} = w_{\theta_l 1} n_1 / P_1 + \dots + w_{\theta_l k} n_k / P_k$ – элементы вектора $a(\theta)$, и

$$w_{\theta_l i} = -f[x_i(\theta), \theta] \frac{\partial x_i(\theta)}{\partial \theta_l} + f[x_{i-1}(\theta), \theta] \frac{\partial x_{i-1}(\theta)}{\partial \theta_l}. \quad (2.12)$$

Для распределений, которые полностью определяются только параметрами сдвига и масштаба, справедливо соотношение

$$\left\| \sum_{i=1}^k \frac{w_{\theta_i} w_{\theta_{j^i}}}{P_i} \right\| = \sum_{i=1}^k \frac{\nabla P_i(\theta) \nabla^T P_i(\theta)}{P_i(\theta)} = \mathbf{J}_\Gamma(\theta) \quad (2.13)$$

и, следовательно,

$$\mathbf{\Lambda}(\theta) = [\mathbf{J}(\theta) - \mathbf{J}_\Gamma(\theta)]^{-1}. \quad (2.14)$$

13 Действительно, для законов с параметром сдвига θ_1 и масштаба θ_2 с функцией распределения $F((x - \theta_1)/\theta_2)$ и плотностью $\frac{1}{\theta_2} f((x - \theta_1)/\theta_2)$ элементы информационной матрицы $\mathbf{J}_\Gamma(\theta)$ имеют вид

$$J_\Gamma(\theta_1, \theta_1) = \sum_{i=1}^k \frac{1}{\theta_2^2 P_i(\theta)} (-f(t_i) + f(t_{i-1}))^2,$$

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

$$J_\Gamma(\theta_1, \theta_2) = \sum_{i=1}^k \frac{1}{\theta_2^2 P_i(\theta)} (-f(t_i) + f(t_{i-1}))(-t_i f(t_i) + t_{i-1} f(t_{i-1})),$$

где $t_i = (x_i - \theta_1) / \theta_2$. Тогда

$$w_{\theta_1 i} = \frac{1}{\theta_2} (-f(t_i) + f(t_{i-1})),$$

$$w_{\theta_2 i} = \frac{1}{\theta_2} (-t_i f(t_i) + t_{i-1} f(t_{i-1})).$$

14 Если проверяемая гипотеза H_0 о принадлежности наблюдаемого закона параметрическому семейству $f(x, \theta)$ неверна и на самом деле справедлива конкурирующая гипотеза H_1 , которой соответствует распределение с плотностью $f_1(x, \theta) = f(x, \theta) + \delta(x, \theta) / \sqrt{n}$, статистика $Y_n^2(\theta)$ в пределе подчиняется нецентральному χ_{k-1}^2 -распределению с параметром нецентральности [288]

$$v(\theta) = \sum_{i=1}^k \frac{c_i^2(\theta)}{P_i(\theta)} + d^T(\theta) \Lambda(\theta) d(\theta), \quad (2.15)$$

где $c_i(\theta) = \int_{x_{i-1}}^{x_i} \delta(x, \theta) dx$, $d_{\theta_l} = w_{\theta_l 1} c_1(\theta) / P_1 + \dots + w_{\theta_l k} c_k(\theta) / P_k$ — элементы вектора $d(\theta)$, соответствующие оцениваемым компонентам вектора θ , а размерность вектора равна числу оцениваемых параметров.

Другим примером модифицированного критерия типа χ^2 является критерий **Никулина–Джапаридзе** [192, 37]. Статистика критерия Никулина–Джапаридзе отличается от статистики критерия Пирсона также только при проверке сложных гипотез и имеет вид [37]

$$U^2(\theta) = X_n^2 - nL^T(\theta)\mathbf{J}^{-1}(\theta)L(\theta),$$

$$L(\theta) = (l_1(\theta), \dots, l_s(\theta)),$$

$$l_j(\theta) = \sum_{i=1}^k \frac{n_i}{nP_i(\theta)} \frac{\partial P_i(\theta)}{\partial \theta_j},$$

15

где m – число параметров закона, оцениваемых методом максимального правдоподобия по негруппированным данным; $\mathbf{J}(\theta)$ – информационная матрица Фишера. При справедливости H_0 статистика Никулина–Джапаридзе подчиняется χ^2 -распределению с числом степеней свободы $k - m - 1$.

Определенная трудность в широком применении модифицированных критериев согласия типа χ^2 заключается в отсутствии специализированного программного обеспечения.

О применении критериев Никулина (Никулина–Рао–Робсона [142]) и Никулина–Джапаридзе в приложениях упоминаний в публикациях практически не содержится. Дело в том, что их применение требует определенной подготовки и наличия программного обеспечения. В отсутствие последнего в приложениях идут по легкому пути, используя не совсем корректно более известные, но менее эффективные критерии.

2.5. Связь мощности критериев со способом группирования наблюдений

Очевидно, что группирование наблюдений приводит к потере информации и эти потери зависят от выбора способа группирования данных. Следуя рекомендациям различных литературных источников, на практике обычно строят интервалы равной длины или в лучшем случае интервалы равной вероятности. Потери информации о законе распределения в этих ситуациях различны, различна и способность критериев распознавать близкие гипотезы.

Мерой внутренней близости распределений случайных величин служит фишеровская информация, что связано с мощностью различения между близкими значениями параметра. Так как в любой статистике не больше информации, чем в исходной выборке, то мощность различения с помощью статистики не больше, чем с помощью всей выборки. Следовательно, если нужно выбирать между несколькими статистиками, надо предпочесть ту, для которой потери фишеровской информации минимальны [303].

Предположим, что конкурирующей гипотезе H_1 соответствует распределение того же типа, что и H_0 , но с параметром θ_1 . Можно показать, разлагая $P_i(\theta_1)$ в соотношении (2.3) в ряд Тейлора при малых $\delta\theta = \theta_1 - \theta$ и пренебрегая членами высшего порядка, что

$$\begin{aligned} v &\approx n \sum_{i=1}^k \frac{\left[P_i(\theta) + \nabla^T P_i(\theta) \delta\theta - P_i(\theta) \right]^2}{P_i(\theta)} = n \sum_{i=1}^k \frac{\delta\theta^T \nabla P_i(\theta) \nabla^T P_i(\theta) \delta\theta}{P_i(\theta)} = \\ &= n \delta\theta^T \left(\sum_{i=1}^k \frac{\nabla P_i(\theta) \nabla^T P_i(\theta)}{P_i(\theta)} \right) \delta\theta = n \delta\theta^T \mathbf{J}_\Gamma(\theta) \delta\theta, \end{aligned} \quad (2.16)$$

где

$$\mathbf{J}_\Gamma(\theta) = \sum_{i=1}^k \frac{\nabla P_i(\theta) \nabla^T P_i(\theta)}{P_i(\theta)}$$

17 – информационная матрица Фишера по группированным данным. Мощность критерия χ^2 Пирсона представляет собой неубывающую функцию от v . Матрица потерь информации, вызванных группированием, $\Delta \mathbf{J} = \mathbf{J}(\theta) - \mathbf{J}_\Gamma(\theta)$, где $\mathbf{J}(\theta)$ – информационная матрица Фишера по негруппированным наблюдениям, является неотрицательно определенной, и, следовательно, $\delta \theta^T \Delta \mathbf{J} \delta \theta \geq 0$. Так как $\delta \theta^T \mathbf{J}_\Gamma(\theta) \delta \theta = \delta \theta^T \mathbf{J}(\theta) \delta \theta - \delta \theta^T \Delta \mathbf{J} \delta \theta$, то очевидно, что с ростом потерь информации падает и мощность критерия при близких конкурирующих гипотезах.

Аналогично с ростом правой части соотношения (2.7) увеличивается мощность **критерия отношения правдоподобия**. Действуя, как и в предыдущем случае, и пренебрегая членами высшего порядка, можно будет иметь

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

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

18

$$\begin{aligned} v &\approx 2n \sum_{i=1}^k \left(P_i(\theta) + \delta \theta^T \nabla P_i(\theta) \right) \left[\frac{\nabla^T P_i(\theta) \delta \theta}{P_i(\theta)} - \frac{\delta \theta^T \nabla P_i(\theta) \nabla^T P_i(\theta) \delta \theta}{2 P_i^2(\theta)} \right] \approx \\ &\approx 2n \sum_{i=1}^k \left(\delta \theta^T \nabla P_i(\theta) + \frac{\delta \theta^T \nabla P_i(\theta) \nabla^T P_i(\theta) \delta \theta}{2 P_i(\theta)} \right) = \\ &= n \delta \theta^T \left[\sum_{i=1}^k \frac{\nabla P_i(\theta) \nabla^T P_i(\theta)}{P_i(\theta)} \right] \delta \theta = N \delta \theta^T \mathbf{J}_\Gamma(\theta) \delta \theta. \end{aligned} \quad (2.17)$$

Это соотношение аналогично соотношению (2.16).

Выражение (2.14) показывает, что свойства критерия, задаваемого статистикой (2.11), также зависят от потерь информации при группировании.

2.6. Асимптотически оптимальное группирование

На основании соотношений (2.16), (2.17) можно утверждать, что чем меньше потери информации, связанные с группированием наблюдений, тем выше мощность соответствующих критериев согласия при близких конкурирующих гипотезах.

19 Потери от группирования можно уменьшить, решая задачу асимптотически оптимального группирования и подбирая граничные точки так, чтобы $\mathbf{J}_\Gamma(\theta)$ стремилась к информационной матрице по негруппированным данным $\mathbf{J}(\theta)$. В случае скалярного параметра эта задача сводится к максимизации количества информации Фишера о параметре по группированной выборке

$$\max_{x_0 < x_1 < \dots < x_{k-1} < x_k} \sum_{i=1}^k \left(\frac{\partial \ln P_i(\theta)}{\partial \theta} \right)^2 P_i(\theta), \quad (2.18)$$

а в случае вектора параметров за критерии оптимальности могут быть приняты различные функционалы от информационной матрицы Фишера. Наиболее естественно максимизировать определитель информационной матрицы, т. е. решать задачу

$$\max_{x_0 < x_1 < \dots < x_{k-1} < x_k} \det \mathbf{J}_\Gamma(\theta). \quad (2.19)$$

Применяя на практике критерии типа χ^2 , наиболее часто используют интервалы равной длины или в лучшем случае интервалы равной вероятности. Выбор равновероятного группирования обоснован определенностью этой процедуры разбиения и ее оптимальностью при отсутствии конкретных альтернатив [207]. Однако при использовании и равновероятного, и равномерного группирования мощность критериев χ^2 Пирсона и отношения правдоподобия обычно далека от максимально возможной.

В общем случае информационная матрица Фишера зависит не только от граничных точек x_i , но и от параметров исследуемого распределения. Однако для достаточно широкого ряда распределений при решении задач асимптотически оптимального группирования граничные точки интервалов удастся получить в виде, инвариантном относительно параметров распределений, и на их основе формировать таблицы асимптотически оптимального группирования.

Применение асимптотически оптимального группирования в критериях согласия типа χ^2 впервые было предложено в работе [188]. Совокупность таблиц асимптотически оптимального группирования, построенная в [183, 220, 219, 242, 217, 191] в результате решения задач (2.18) и (2.19) для распределений экспоненциального, полунормального, Рэлея, Максвелла, модуля многомерного нормального вектора, Парето, Эрланга, Лапласа, нормального, логарифмически-нормальных (\ln и \lg), Коши, Вейбулла, распределений минимального и максимального значения, двойного показательного, гамма-распределения, представлена в приложении А. Таблицы приложения А.1–А.58 могут быть использованы как при проверке гипотез, так и при оценивании. Полученные таблицы используют в программной системе [224] при проверке согласия по критериям χ^2 Пирсона и отношения правдоподобия и при вычислении робастных оценок.

Для многих законов распределений граничные точки интервалов не могут быть выражены в виде, инвариантном относительно параметров распределений, т. е. они остаются функциями этих параметров. Это касается, например, таких законов, как гамма- и бета-распределения [183, 217], экспоненциального семейства распределений. В этом случае формирование таблиц асимптотически оптимального группирования теряет смысл. Однако можно решать задачу асимптотически оптимального группирования при конкретных значениях параметров в процессе проверки гипотез о согласии, как это реализуется в таких ситуациях в программной системе [224].

Положительный эффект применения асимптотически оптимального группирования на результатах статистического анализа проявляется при малых отклонениях выборки от предположений.

При проверке простых гипотез и использовании асимптотически оптимального группирования критерии χ^2 Пирсона и отношения правдоподобия оказываются мощнее непараметрических критериев Колмогорова, Смирнова, ω^2 Крамера–Мизеса–Смирнова и Ω^2 Андерсона–Дарлингa против близких конкурирующих гипотез, лучше улавливают малые отклонения от предположений в наблюдаемых данных [215, 216, 74].

2.7. Характер влияния способов группирования и метода оценивания на распределения статистик типа χ^2

Вычисленные по конкретной выборке значения статистик типа χ^2 очень сильно зависят от того, как сгруппированы данные. При выборе интервалов группирования одним способом нулевая гипотеза H_0 о согласии может быть отвергнута, другим – принята. При практическом использовании критериев согласия выбирают либо интервалы равной длины, либо интервалы равной вероятности (равной частоты), либо асимптотически оптимальные интервалы [188, 191, 224, 215, 216, 74, 191]. Использование интервалов равной вероятности было предложено еще в работе [104]. Разбиение области определения случайной величины (размаха выборки) на интервалы равной длины неоднозначно. Более определенными способами являются равновероятное и асимптотически оптимальное группирования. При асимптотически оптимальном группировании мощность этих критериев относительно близких конкурирующих гипотез максимальна [188, 215, 216, 74, 190].

Для статистики X_n^2 Пирсона, вычисляемой в соответствии с формулой (2.1), при справедливой простой гипотезе H_0 предельное распределение $G(X_n^2|H_0)$ есть χ^2 -распределение с числом степеней свободы $k-1$. При сложной гипотезе, если по выборке оценивали m параметров закона, статистика подчиняется χ_r^2 -распределению с $k-m-1$ степенями свободы. При справедливой альтернативной гипотезе H_1 предельное распределение $G(X_n^2|H_1)$ представляет собой нецентральное χ^2 -распределение с тем же числом степеней свободы и параметром нецентральности, определяемым соотношением (2.3).

В случае проверки сложных гипотез и оценивания по выборке параметров распределений использование в качестве предельных χ^2_{k-m-1} -распределений справедливо лишь при определении оценок параметров по сгруппированным данным [205].

Все сказанное относится и к критерию отношения правдоподобия, статистика которого определяется формулой (2.6).

Выбор способа группирования, в том числе равновероятного или асимптотически оптимального, отражается на предельных распределениях статистик критериев $G(S|H_1)$ (при верной конкурирующей гипотезе). При использовании асимптотически оптимального группирования распределения $G(S|H_0)$ и $G(S|H_1)$ в случае близких альтернатив максимально «раздвинуты» относительно друг друга [215, 216, 74, 190]. Это справедливо как при проверке простых, так и при проверке сложных гипотез.

В случае проверки простых гипотез распределения $G(S|H_0)$ статистик критериев χ^2 Пирсона и отношения правдоподобия при любом разумном способе группирования [равновероятном (РВГ), асимптотически оптимальном (АОГ), равной длины] хорошо согласуются с соответствующим χ^2_{k-1} -распределением [260, 255, 97].

При проверке сложных гипотез предельные распределения статистик критериев χ^2 Пирсона и отношения правдоподобия могут зависеть не только от числа оцененных параметров, но и от способа группирования, вида наблюдаемого закона распределения $f(x, \theta)$, метода оценивания и типа оцениваемого параметра.

В частности, при проверке сложных гипотез и использовании ОМП по негруппированным наблюдениям распределения $G(S|H_0)$ статистик данных критериев существенно зависят от способа группирования.

В работах [260, 255, 97] методами статистического моделирования были проведены исследования законов распределения статистик типа χ^2 при простых и различных сложных гипотезах, при справедливости гипотезы H_0 и справедливости конкурирующей гипотезы H_1 , при равновероятном и асимптотически оптимальном группировании.

Например, на рис. 2.4–2.7 сопоставлены полученные в результате моделирования эмпирические распределения статистик (при вычислении ОМП по негруппированным наблюдениям) с соответствующими χ^2_{k-m-1} -распределениями. Цифрой 1 на указанных рисунках отмечено распределение статистики при асимптотически оптимальном группировании, цифрой 2 – при равновероятном группировании.

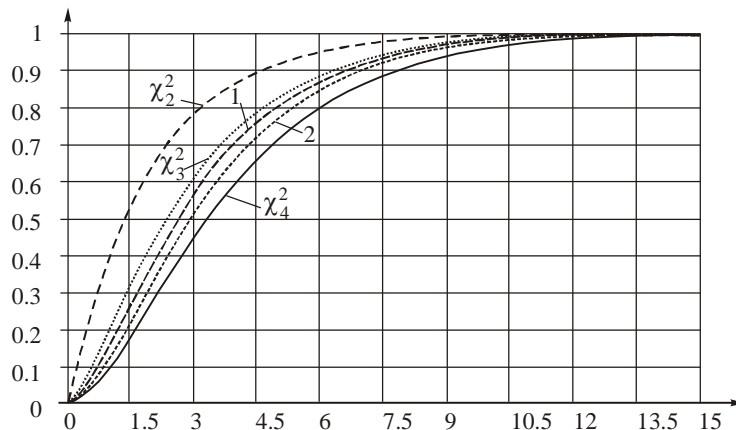


Рис. 2.4. Распределение статистики X_n^2 при пяти интервалах группирования и оценивании масштабного параметра нормального распределения

На рис. 2.4–2.6 в качестве примера представлены распределения статистики X_n^2 при справедливой гипотезе H_0 для пяти интервалов группирования и оценивании по негруппированным данным параметров нормального распределения.

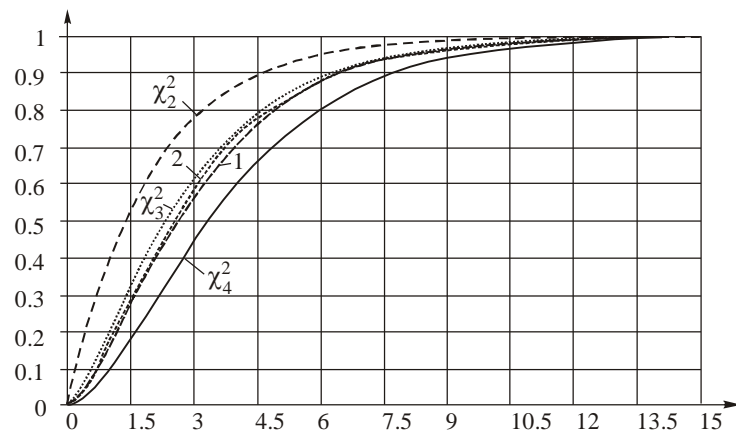


Рис. 2.5. Распределение статистики X_n^2 при пяти интервалах группирования и оценивании параметра сдвига нормального распределения

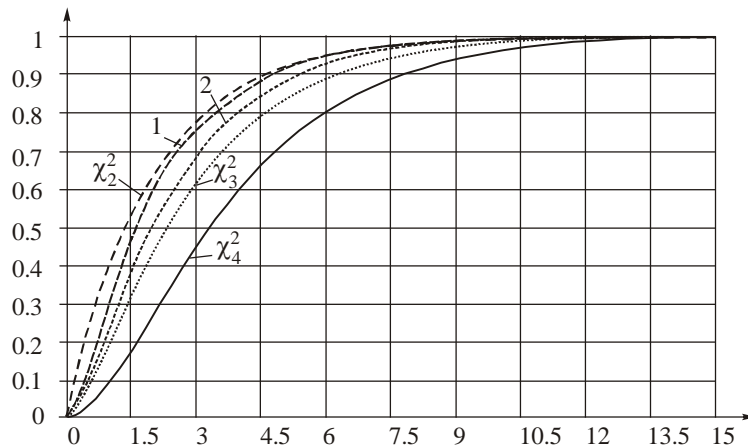


Рис. 2.6. Распределение статистики X_n^2 при пяти интервалах группирования и оценивании двух параметров нормального распределения

Для сравнения на рис. 2.7 показаны распределения статистики $S_{\text{ОП}}$ при пяти интервалах группирования и оценивании обоих параметров нормального распределения при асимптотически оптимальном и равновероятном группировании.

Результаты моделирования и анализа показали [260, 255, 97], что распределения статистик критериев χ^2 Пирсона и отношения правдоподобия при проверке сложных гипотез и вычислении оценок по негруппированным данным существенно отличаются при различных способах группирования. Эмпирический закон распределения статистики при асимптотически оптимальном группировании ближе к χ_{k-m-1}^2 -распределению, чем при равновероятном группировании. Более того, распределения статистик

зависят не только от числа оцененных по выборке параметров, но и от того, какой параметр оценивали. Например, оценивание параметра сдвига приводит к более значительному изменению распределения статистики, чем оценивание масштабного параметра. Распределения статистик зависят и от вида наблюдаемого закона. Но эта зависимость не проявляется так резко, как в случае применения для проверки сложных гипотез непараметрических критериев, например типа Колмогорова.

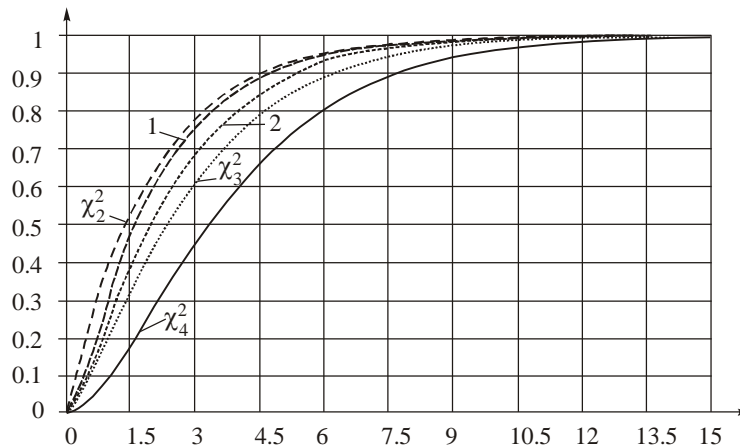


Рис. 2.7. Распределение статистики S_{OP} при пяти интервалах группирования и оценивании двух параметров нормального распределения

Эмпирические распределения статистики отношения правдоподобия оказываются ближе к χ^2_{k-m-1} -распределению, чем соответствующие эмпирические распределения статистики χ^2 Пирсона.

χ_r^2 -распределение с числом степеней свободы r является частным случаем гамма-распределения с основным параметром (формы), равным $r/2$, и с масштабным – 2. Оценивание одного из параметров учитывается уменьшением числа степеней свободы на 1. Так, если измерять изменение предельного закона «в степенях свободы», то оценивание по негруппированным наблюдениям даже параметра сдвига обычно приводит к изменению «числа степеней» на значение, меньшее 1, еще к меньшему изменению в степенях свободы приводит оценивание масштабного параметра. При оценивании по негруппированным наблюдениям двух параметров закона различие между χ_r^2 -распределением и действительным распределением статистики еще более значимо. Таким образом, при вычислении оценок параметров по негруппированным наблюдениям использование в критериях согласия χ_{k-m-1}^2 -распределения чревато занижением вероятности вида $P\{S > S^*\}$ и определенным риском отвергнуть верную гипотезу H_0 . Занижение $P\{S > S^*\}$ более существенно при малом числе интервалов группирования. В целом этот риск существенно меньше в случае применения асимптотически оптимального группирования.

Полученные в результате моделирования эмпирические законы распределения статистик рассматриваемых критериев практически всегда с достаточно высокой точностью описываются гамма-распределением.

Использование в критериях согласия χ^2 -распределения для вычисления вероятности вида $P\{S > S^*\}$ связано с риском отвергнуть верную гипотезу H_0 . В случае применения асимптотически оптимального группирования этот риск существенно меньше. В качестве примера в табл. 2.1 приведены значения вероятностей $P\{S > S^*\}$ для различных значений статистики S^* , вычисленные по χ_4^2 -распределению и в соответствии с «действительными» распределениями статистик при семи интервалах группирования и оценивании двух параметров нормального распределения. Эти данные вместе с другими результатами

позволяют судить о том, что в случае применения асимптотически оптимального группирования и использования χ^2_{k-m-1} -распределения ошибки при вычислении вероятности $P\{S > S^*\}$ практически незначимы.

Т а б л и ц а 2.1

Значения вероятностей вида $P\{S > S^*\}$

Значение S^*	χ^2_4 -распределение	Распределение статистики $S_{\text{ОП}}$		Распределение статистики X^2_n	
		При АОГ	При РВГ	При АОГ	При РВГ
3	0.5578	0.5876	0.6293	0.5914	0.6304
4	0.4060	0.4303	0.4700	0.4356	0.4716
5	0.2873	0.3049	0.3384	0.3108	0.3403
6	0.1991	0.2108	0.2370	0.2165	0.2389
7	0.1359	0.1430	0.1626	0.1481	0.1643
8	0.0916	0.0956	0.1096	0.0999	0.1111
9	0.0611	0.0632	0.0729	0.0666	0.0741
10	0.0404	0.0413	0.0480	0.0440	0.0489
11	0.0266	0.0268	0.0313	0.0288	0.0320
12	0.0173	0.0173	0.0202	0.0187	0.0207
13	0.0113	0.0110	0.0130	0.0120	0.0133

Исследование распределений статистик при вычислении ОМП по негруппированным данным и дальнейшем увеличении числа интервалов показало возрастающую близость распределений статистик к χ^2_{k-m-1} -распределениям. Следовательно, при использовании χ^2_{k-m-1} -распределения для вычисления вероятностей вида $P\{S > S^*\}$ с ростом числа интервалов (при достаточном объеме выборки) ошибки будут уменьшаться.

Как отражается способ группирования на предельных распределениях статистики X_n^2 при справедливости конкурирующей гипотезы H_1 , показано на рис. 2.8 и 2.9, где приведены полученные в результате моделирования функции распределения $G(S|H_1)$ при использовании асимптотически оптимального (АОГ) и равновероятного (РВГ) группирования, когда гипотеза H_0 соответствует нормальному закону распределения, а гипотеза H_1 – логистическому. Эти два закона достаточно близки и трудно-различимы с помощью критериев согласия. На рис. 2.8 приведены эмпирические распределения статистики при простой гипотезе H_0 , а на рис. 2.9 – при сложной. Число интервалов группирования – 7. Из рис. 2.8, 2.9 видно, насколько выше мощность критерия при асимптотически оптимальном группировании по сравнению с равновероятным.

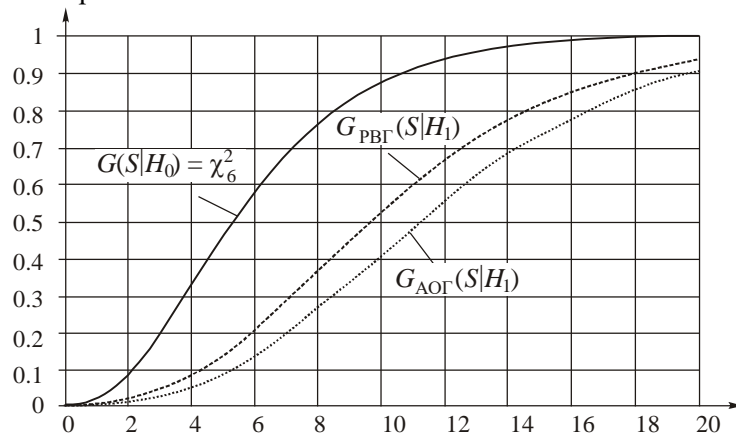


Рис. 2.8. Распределения статистики X_n^2 при проверке простой гипотезы H_0

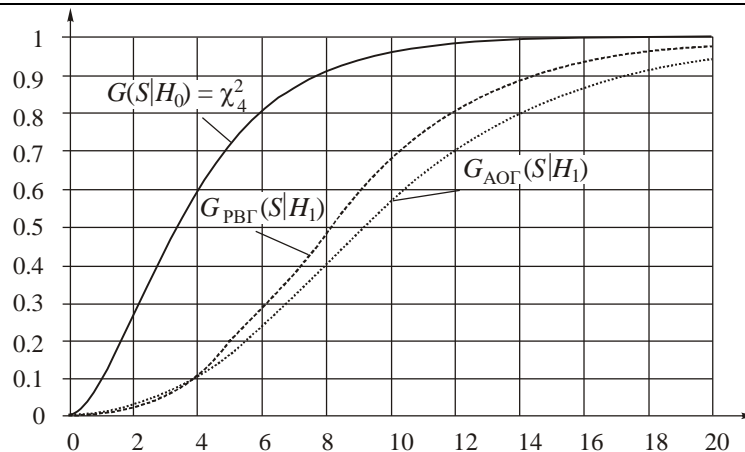


Рис. 2.9. Распределения статистики X_n^2 при проверке сложной гипотезы H_0

Распределения $G(S|H_0)$ и $G(S|H_1)$ статистики $Y_n^2(\theta)$, определяемой соотношением (2.11) при использовании оценок максимального правдоподобия параметров по негруппированным данным, несущественно зависят от способа группирования. Исследование этих распределений методами статистического моделирования показало хорошее согласие $G(S|H_0)$ с χ_{k-1}^2 -распределениями и при равновероятном и при асимптотически оптимальном группировании. При выборе различных способов группирования нет большой разницы между соответствующими распределениями статистики $G(S|H_1)$. При малом числе интервалов критерий оказывается несколько мощнее в случае равновероятного группирования.

Таким образом, по поводу распределений статистик критериев типа χ^2 можно сделать следующие выводы.

Если по наблюдаемой выборке не оценивали параметры закона (простая гипотеза), то распределения статистик типа χ^2 при справедливой гипотезе H_0 достаточно хорошо согласуются с χ^2_{k-1} как при равновероятном, так и при асимптотически оптимальном группировании. Различия между распределениями статистик при равновероятном и асимптотически оптимальном группировании заметны, но незначимы. Распределения статистик при равновероятном группировании в целом оказываются ближе к χ^2_{k-1} -распределению.

В условиях, когда оценки параметров определяют по негруппированным наблюдениям, распределения этих статистик хорошо согласуются с гамма-распределениями.

а) Распределения статистик критериев отношения правдоподобия и χ^2 Пирсона существенно зависят от способа группирования, особенно при малом числе интервалов.

б) Распределения статистик зависят не только от числа оцененных по выборке параметров, но и от того, какой параметр оценивали. В частности, оценивание параметра сдвига приводит к более значительному изменению распределений статистик, чем оценивание масштабного параметра. Картина аналогична той, что наблюдается для распределений статистик типа Колмогорова, ω^2 Крамера–Мизеса–Смирнова и Ω^2 Андерсона–Дарлинга при проверке сложных гипотез [258, 98].

в) Эмпирические распределения статистики отношения правдоподобия оказываются ближе к предельному теоретическому χ^2_{k-m-1} -распределению, чем соответствующие эмпирические распределения статистики χ^2 Пирсона.

г) В целом при малом числе интервалов и оценивании m параметров число степеней свободы предельного распределения уменьшается на «число степеней свободы», меньшее m . При этом эмпириче-

ский закон распределения статистики при асимптотически оптимальном группировании ближе к теоретическому χ^2_{k-m-1} -распределению, чем при равновероятном группировании.

д) С ростом числа интервалов k разность между функцией распределения χ^2_{k-m-1} и действительными функциями распределения статистик отношения правдоподобия и χ^2 Пирсона в случае асимптотически оптимального группирования убывает существенно быстрее.

е) Анализ показал, что при использовании для вычисления вероятностей $P\{S > S^*\}$ функции распределения χ^2_{k-m-1} при асимптотически оптимальном группировании и малом числе интервалов ($k < 10$) при малых значениях $P\{S > S^*\}$ (именно при малых вероятностях принимают решение отклонить гипотезу H_0 или нет) погрешность имеет значения, которые несущественны для практических задач, т. е. в этом случае, используя χ^2_{k-m-1} -распределение, не совершают большой ошибки. В то же время ошибка в определении критического значения S_α может быть достаточно большой. Поэтому целесообразно принимать решение по достигнутому уровню значимости: значению вероятности $P\{S > S^*\}$.

ж) С ростом числа интервалов (при соответствующем объеме выборки) отличие распределений статистик X_n^2 и $S_{\text{ОП}}$ от распределений χ^2_{k-m-1} и при асимптотически оптимальном, и при равновероятном группировании становится несущественным.

При справедливости конкурирующей гипотезы H_1 распределения статистик X_n^2 и отношения правдоподобия и при простых, и при сложных гипотезах сильно зависят от способа группирования.

При близких альтернативах разность $G(S|H_0) - G(S|H_1)$ для этих статистик максимальна в случае использования асимптотически оптимального группирования, следовательно, максимальна мощность критерия.

Зависимость распределений $G(S|H_1)$ статистики Y_n^2 Никулина от выбранного способа группирования выражена менее ярко, причем при малом числе интервалов предпочтительней может оказаться равновероятное группирование, а с ростом числа интервалов – асимптотически оптимальное.

2.8. Выбор числа интервалов

34 Число интервалов группирования, используемое (и рекомендуемое) при вычислении оценок параметров, построении гистограмм, вычислении статистик типа отношения правдоподобия или χ^2 Пирсона, колеблется в очень широких пределах. Большинство рекомендуемых формул для оценки числа интервалов k носит эмпирический характер и обычно дает завышенные значения.

Определение числа интервалов связано с объемом выборки. Целый ряд рекомендаций из различных источников по выбору числа интервалов k приведен в [290].

При выборе интервалов равной длины определяющим является требование, чтобы число наблюдений, попавших в интервалы, было не слишком малым и сравнимым. Такое требование выдвигают в связи с опасением, что в противном случае распределение статистики типа χ^2 не будет являться χ_r^2 -распределением. При этом наиболее часто рекомендуют, чтобы число наблюдений, попавших в интервал, было не менее десяти. В [20] отмечено, что на практике допустимо, чтобы число наблюдений в крайних интервалах было менее пяти. В работах [104, 105], посвященных изучению мощности критерия χ^2 Пирсона, в случае унимодального распределения допускается уменьшать ожидаемые частоты

попадания наблюдений для одного или двух интервалов до единицы и даже ниже. Статистическое моделирование подтверждает, что и в такой ситуации распределения статистик типа χ^2 хорошо согласуются с соответствующими χ_r^2 -распределениями.

Во многих источниках, например в [138], можно найти упоминание эвристической формулы Старджесса для определения «оптимального» числа интервалов

$$k = \log_2 n + 1 = 3,3 \lg n + 1.$$

В [350] для определения «оптимального» числа интервалов рекомендуют формулу Брукса и Каррузера

$$k = 5 \lg n.$$

В [59] рекомендуют соотношение

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

В [105] для равновероятных интервалов их число устанавливают порядка

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

где t – квантиль стандартного нормального распределения для заданного уровня значимости. В ряде работ приводят модификации данной формулы. В [321] предлагают значение

$$k = 4 \lg n,$$

а в [329] – дальнейшее развитие этого соотношения

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

В исследовании [152] получено соотношение

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

где $\aleph = 1 / \sqrt{\mu_4 / \sigma^4}$ – значение контрэксцесса; μ_4 – четвертый центральный момент; σ – среднее квадратичное отклонение.

При больших объемах выборок n разброс значений k , задаваемых различными формулами, достаточно велик. Поэтому на практике при выборе числа интервалов больше руководствуются тем, чтобы в интервалы попадало число наблюдений не менее 5...10. Так, например, в рекомендациях ВНИИМетрологии [168] в зависимости от n предлагают следующие значения k :

n	k
40...100	7...9
100...500	8...12
500...1000	10...16
1000...10000	12...22

Все перечисленные рекомендации опирались на предположение, что k следует выбирать таким образом, чтобы вид гистограммы был как можно ближе к плавной кривой плотности распределения генеральной совокупности. В [338] показано, что уклонение гистограммы от плотности распределения в лучшем случае имеет порядок $1/\sqrt[3]{n}$, достигаемый при числе интервалов k порядка $\sqrt[3]{n}$.

Очевидно, что «оптимальное» значение k зависит не только от объема выборки, но и от вида закона распределения и от способа группирования.

При асимптотически оптимальном группировании относительно скалярного параметра при 10 или 11 интервалах в группированной выборке сохраняется около 98 % информации, при оптимальном группировании относительно вектора параметров (два параметра) для 15 интервалов – около 95 %. Дальнейшее увеличение числа интервалов существенного значения не имеет.

Конкретное число интервалов при асимптотически оптимальном группировании можно выбирать исходя из следующих соображений. При оптимальном группировании вероятности попадания в интервалы в общем случае не равны. Обычно минимальны вероятности попадания в крайние интервалы. Поэтому k желательно выбирать из условия $nP_i(\theta) \geq 5 \dots 10$ для любого интервала при оптимальном группировании. По крайней мере, минимальная ожидаемая частота должна быть больше 1. В случае использования равновероятного группирования порядок k должен быть примерно таким же, как и при асимптотически оптимальном группировании.

Все наиболее разумные рекомендации по выбору числа интервалов, в том числе в случае асимптотически оптимального группирования, исходят из того, чтобы при данном n описать плотность распределения ее непараметрической оценкой (гистограммой) как можно лучше. Но ни одна из рекомендаций, за исключением [104, 105], не подходит к выбору k с позиций мощности критерия согласия! Не опираются на требование построения наиболее мощного критерия при близких конкурирующих гипотезах, хотя известно, что при заданном объеме выборки, заданных конкурирующих гипотезах H_0 и H_1 , выбранном способе группирования и фиксированном уровне значимости α с ростом числа интервалов k мощность критериев χ^2 Пирсона и отношения правдоподобия падает [344, 166].

Об изменении мощности критерия χ^2 Пирсона с ростом числа интервалов при проверке простой гипотезы можно судить по рис. 2.10. Через X_n^2 обозначена статистика, вычисляемая в соответствии с формулой (2.1). На рис. 2.10 представлены полученные экспериментально распределения статистики $G^k(X_n^2|H_0)$ и $G^k(X_n^2|H_1)$ при числе интервалов $k = 7, 10, 15, 20$ и объеме выборки $n = 500$, когда

гипотеза H_0 соответствует нормальному закону, а H_1 – логистическому (два очень близких закона). На рисунке приведены распределения статистики при равновероятном группировании. Ордината нижнего конца соответствующей вертикальной черты определяет значение β (вероятность ошибки второго рода) при уровне значимости $\alpha=0,1$ для соответствующего числа интервалов. Мощность равна $1-\beta$. Как видно, в полном соответствии с результатами работ [344, 166] при увеличении числа интервалов мощность критерия падает. При асимптотически оптимальном группировании и таких числах интервалов для данной пары конкурирующих гипотез соответствующая мощность выше.

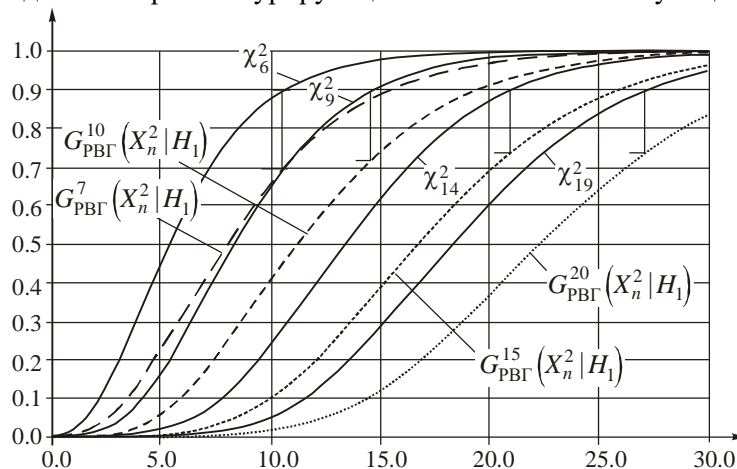


Рис. 2.10. Распределения статистики X_n^2 при проверке простой гипотезы: H_0 – нормальный закон, H_1 – логистический закон, $k = 7, 10, 15, 20$

Аналогичные изменения мощности критерия для статистики X_n^2 в зависимости от числа интервалов при проверке сложной гипотезы иллюстрирует рис. 2.11. Здесь также с ростом k мощность критерия падает.

Мощность критерия Никулина с использованием статистики Y_n^2 с ростом k уменьшается существенно медленней (рис. 2.12) и она выше, чем мощность критерия χ^2 Пирсона.

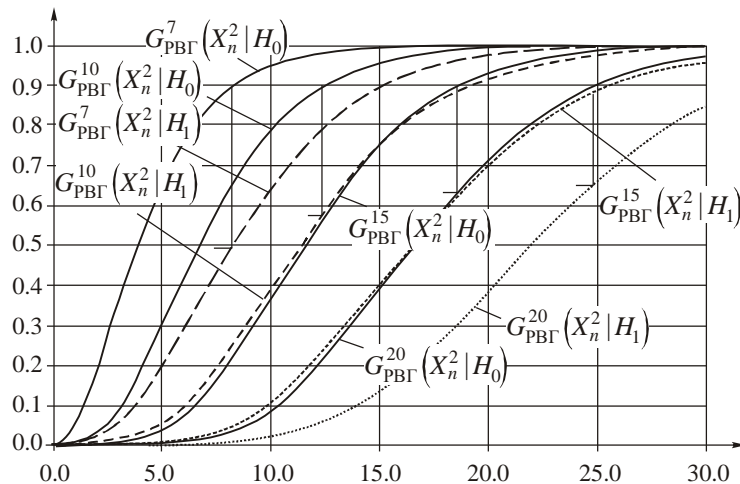


Рис. 2.11. Распределения статистики X_n^2 при проверке сложной гипотезы: H_0 – нормальный закон, H_1 – логистический закон, $k = 7, 10, 15, 20$

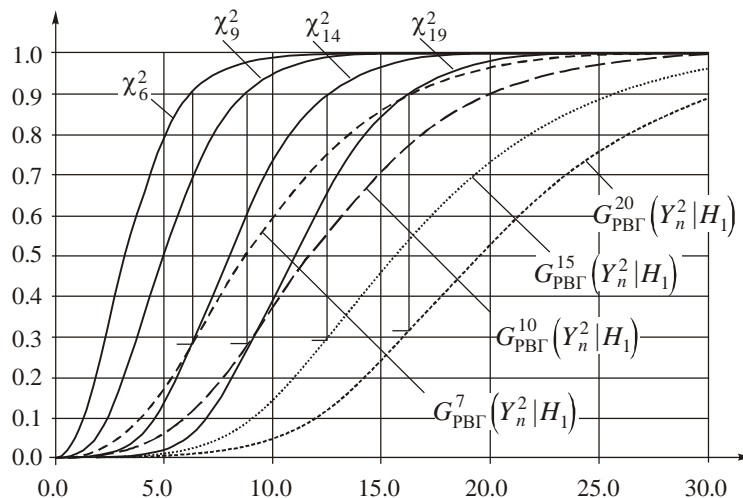


Рис. 2.12. Распределения статистики Y_n^2 при проверке сложной гипотезы: H_0 – нормальный закон, H_1 – логистический закон, $k = 7, 10, 15, 20$

В табл. 2.2 представлены значения мощности критериев χ^2 Пирсона и типа χ^2 Никулина при рассматриваемой альтернативе (H_0 – нормальный закон, H_1 – логистический закон) при различном числе интервалов k в случае простых и сложных гипотез. В колонке 2 даны значения мощности критерия χ^2 Пирсона при проверке простой гипотезы, полученные по результатам моделирования (экспериментально). В колонке 3 приведены теоретические значения мощности, вычисленные при условии, что $G(X_n^2 | H_0)$ представляет собой χ^2_{k-1} -распределение, а $G(X_n^2 | H_1)$ – соответствующее нецентральное

распределение. Колонка 4 содержит значения мощности критерия χ^2 Пирсона при проверке сложной гипотезы, полученные по результатам моделирования при использовании ОМП по негруппированным наблюдениям. В колонке 5 приведены теоретические значения мощности критерия χ^2 Пирсона при проверке сложной гипотезы, вычисленные при условии, что $G(X_n^2|H_0)$ представляет собой χ_{k-3}^2 -распределение (оценки параметров вычисляются при минимизации статистики X_n^2), а $G(X_n^2|H_1)$ – соответствующее нецентральное распределение. В колонке 6 даны значения мощности критерия Никулина, полученные в результате моделирования распределений статистики Y_n^2 , а в колонке 7 – расчетные значения мощности этой статистики при рассматриваемой паре гипотез H_0 и H_1 .

В табл. 2.2 приведены значения мощности критериев для k от 6 до 30. Проследив изменение мощности критериев при $k < 6$, можно убедиться, что если мощность критерия χ^2 Пирсона с уменьшением числа интервалов продолжает возрастать, то мощность критерия Никулина со статистикой Y_n^2 при $k \leq 6$ начинает падать. Это свидетельствует о том, что для критерия Никулина существует оптимальное число интервалов, при котором его мощность максимальна.

Таблица 2.2

Значения мощности критериев χ^2 Пирсона и типа χ^2 Никулина при уровне значимости $\alpha = 0.1$ (H_0 – нормальный закон, H_1 – логистический закон)

k	В случае простой гипотезы		В случае сложной гипотезы			
	Для статистики $X_n^2 = Y_n^2$		Для статистики X_n^2		Для статистики Y_n^2	
	по результатам моделирования	теоретическая	по результатам моделирования (при использовании ОМП)	теоретическая (при использовании оценок $\min X_n^2$)	по результатам моделирования	теоретическая
1	2	3	4	5	6	7
6	0.46	0.449	0.53	0.526	0.70	0.739
7	0.43	0.427	0.49	0.488	0.71	0.750
8	0.42	0.409	0.45	0.459	0.71	0.755
9	0.38	0.395	0.43	0.436	0.71	0.756
10	0.38	0.383	0.43	0.418	0.71	0.756
11	0.37	0.373	0.41	0.403	0.71	0.754
12	0.35	0.364	0.41	0.391	0.70	0.752
13	0.35	0.357	0.38	0.381	0.70	0.749
14	0.34	0.351	0.38	0.373	0.69	0.746
15	0.33	0.345	0.38	0.365	0.69	0.742
16	0.33	0.340	0.38	0.359	0.69	0.738

17	0.32	0.336	0.37	0.353	0.69	0.734
18	0.32	0.332	0.37	0.348	0.69	0.730
19	0.32	0.328	0.35	0.343	0.68	0.726
20	0.31	0.325	0.35	0.339	0.68	0.722
21	0.31	0.322	0.34	0.335	0.67	0.718
22	0.31	0.319	0.34	0.331	0.67	0.714
23	0.31	0.317	0.34	0.328	0.66	0.710
24	0.30	0.314	0.34	0.325	0.66	0.707
25	0.30	0.312	0.34	0.322	0.66	0.703
26	0.30	0.310	0.33	0.320	0.65	0.699
27	0.30	0.308	0.33	0.317	0.65	0.695
28	0.30	0.306	0.32	0.315	0.65	0.692
29	0.30	0.304	0.32	0.313	0.65	0.688
30	0.30	0.302	0.32	0.310	0.65	0.684

В [268] мощность критериев типа χ^2 в зависимости от выбираемого числа интервалов k была исследована при различных проверяемых гипотезах H_0 и различных конкурирующих гипотезах H_1 при разных объемах выборок. Мощность для критериев типа χ^2 может быть вычислена в соответствии с формулой [163]:

$$1 - \beta = P(v|r, \alpha) = e^{-v/2} \sum_{j=0}^{\infty} \frac{v^j}{j! 2^{2j-1+r/2} \Gamma(j+r/2)} \int_{\sqrt{\chi_{r,\alpha}^2}}^{\infty} y^{2j-1+r} e^{-y^2/2} dy, \quad (2.20)$$

где v – параметр нецентральности, определяемый формулами (2.3), (2.7) или (2.15); $\chi^2_{r,\alpha}$ представляет собой $(1-\alpha)$ -процентную точку χ^2_r -распределения с r степенями свободы (α – заданная вероятность ошибки первого рода, β – вероятность ошибки второго рода).

Максимальную мощность критерии χ^2 Пирсона и отношения правдоподобия зачастую имеют или при минимально возможном числе интервалов, определяемом условием $r = k - m - 1 \geq 1$, или при оптимальном числе интервалов, близком к минимально возможному. Оптимальное число интервалов для критерия Никулина обычно больше, чем для критериев χ^2 Пирсона и отношения правдоподобия, и не превышает значения $\sqrt[3]{n}$.

44 В работе [269] исследование мощности критериев Пирсона и Никулина как функции от объемов выборки n и числа интервалов k было проведено и аналитически, и методами статистического моделирования. Причем результаты аналитических вычислений мощности в соответствии с соотношением (2.20) полностью подтвердились оценками мощности, полученными на основании моделирования.

Все приводимые ниже функции мощности строились при уровне значимости $\alpha = 0.1$. Объемы выборок на рисунках обозначены n .

На рис. 2.13 и в табл. 2.3 в зависимости от числа k равновероятных интервалов при различных n представлены функции мощности критерия χ^2 Пирсона при проверке простой гипотезы о согласии с экспоненциальным законом ($H_0: f(x) = \theta \exp\{-\theta x\}$ при $\theta = 1$; $H_1: f(x) = \theta \exp\{-\theta x\}$ при $\theta = 1.05$).

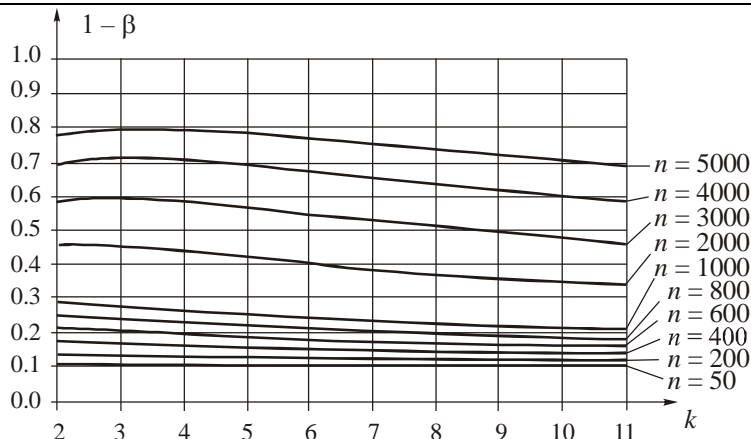


Рис. 2.13. Функции мощности критерия χ^2 Пирсона при проверке простой гипотезы о согласии с экспоненциальным законом при равновероятном группировании

На рис. 2.14 и в табл. 2.4 приведены аналогичные функции при использовании асимптотически оптимального группирования. Асимптотически оптимальные граничные точки, минимизирующие потери в информации Фишера, представлены в табл. А.1 приложения А. И в том и в другом случае с ростом k мощность падает, но в случае асимптотически оптимального группирования она выше, чем при равновероятном.

На рис. 2.15 и в табл. 2.5 приведены функции мощности критерия χ^2 Пирсона в случае разбиения области определения случайной величины на интервалы равной вероятности при проверке простой гипотезы о согласии с нормальным законом

$$H_0: f(x) = \frac{1}{\theta_1 \sqrt{2\pi}} \exp \left\{ -\frac{(x - \theta_0)^2}{2\theta_1^2} \right\}$$

при $\theta_0 = 0$, $\theta_1 = 1$ против H_1 : нормальный закон при $\theta_0 = 0.05$, $\theta_1 = 1.05$.

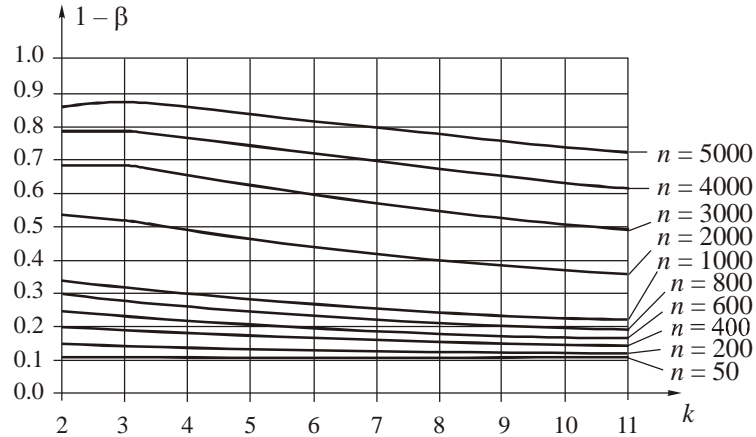


Рис. 2.14. Функции мощности критерия χ^2 Пирсона при проверке простой гипотезы о согласии с экспоненциальным законом при асимптотически оптимальном группировании

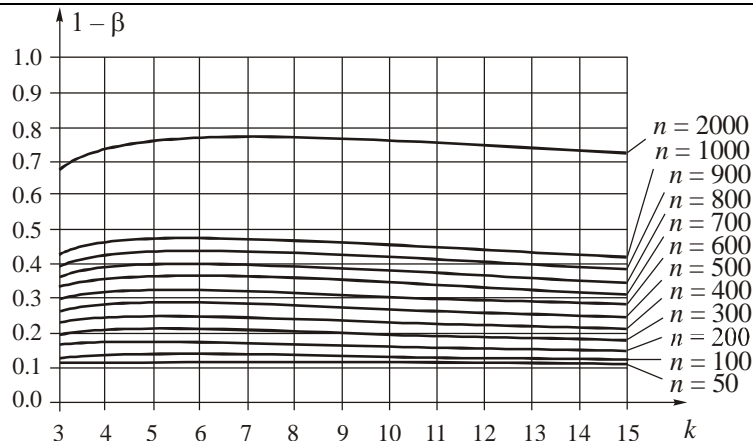


Рис. 2.15. Функции мощности критерия χ^2 Пирсона при проверке простой гипотезы о согласии с нормальным законом при равновероятном группировании

На рис. 2.16 и в табл. 2.6 представлены аналогичные функции мощности в случае использования асимптотически оптимального группирования. Асимптотически оптимальные граничные точки, минимизирующие потери в информации Фишера, представлены в табл. А.28 приложения А.

На рис. 2.17–2.20 представлены функции мощности критерия χ^2 Пирсона при проверке простых и сложных гипотез о согласии с нормальным законом

$$H_0: f(x) = \frac{1}{\theta_1 \sqrt{2\pi}} \exp \left\{ -\frac{(x - \theta_0)^2}{2\theta_1^2} \right\}$$

при логистическом законе в качестве конкурирующей гипотезы

$$H_1: f(x) = \frac{\pi}{\theta_1 \sqrt{3}} \exp \left\{ -\frac{\pi(x - \theta_0)}{\theta_1 \sqrt{3}} \right\} / \left[1 + \exp \left\{ -\frac{\pi(x - \theta_0)}{\theta_1 \sqrt{3}} \right\} \right]^2$$

при значениях параметров $\theta_0 = 0$, $\theta_1 = 1$.

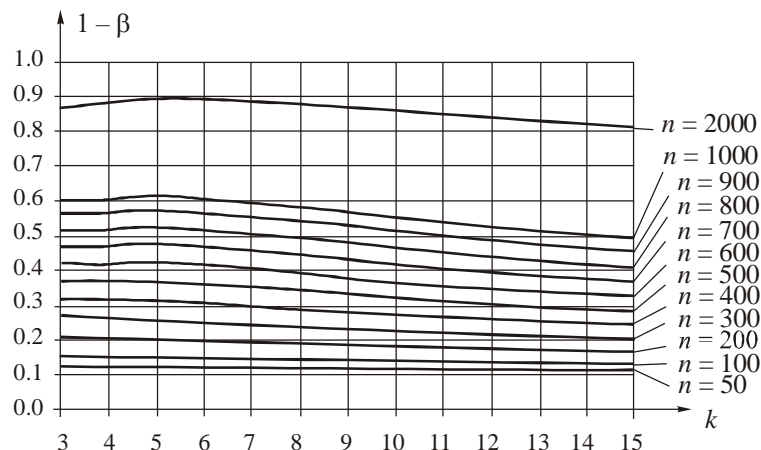


Рис. 2.16. Функции мощности критерия χ^2 Пирсона при проверке простой гипотезы о согласии с нормальным законом при асимптотически оптимальном группировании

На рис. 2.17 представлены функции мощности для случая равновероятного группирования и проверки простой гипотезы, соответствующей нормальному закону с параметрами $\theta_0 = 0$, $\theta_1 = 1$. На рис.

2.18 – то же, но при проверке сложной гипотезы. Как видим, и здесь с ростом k мощность падает. При проверке простой гипотезы функция мощности критерия Пирсона для $n > 100$ принимает максимальное значение при $k = 4$, и при дальнейшем увеличении объема выборки это оптимальное число интервалов не изменяется.

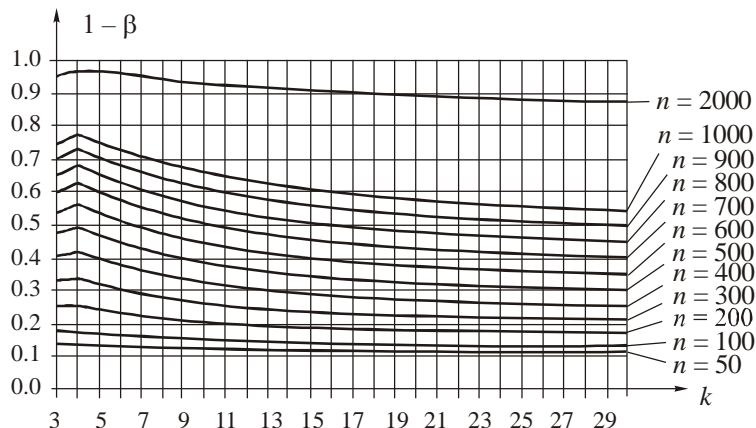


Рис. 2.17. Функции мощности критерия χ^2 Пирсона при проверке простой гипотезы о согласии с нормальным законом при равновероятном группировании и альтернативе, соответствующей логистическому закону

В случае проверки сложной гипотезы и оценивания по выборке параметров гипотетического распределения функция мощности критерия Пирсона принимает наибольшее значение при минимально возможном числе интервалов $k = 4$ и далее монотонно убывает с ростом k (рис. 2.18).

На рис. 2.19 отражены функции мощности критерия χ^2 Пирсона для приведенной выше альтернатив при проверке простой гипотезы, а на рис. 2.20 – при проверке сложной в случае применения асимптотически оптимального группирования. Здесь функции мощности представляют собой более интересную картину с провалами при $k=4$ и $k=6$. Эти провалы свидетельствуют о том, что при таких комбинациях граничных точек, несмотря на минимальные потери в информации Фишера, две рассматриваемые конкурирующие гипотезы плохо различаются. При дальнейшем росте k мощность критерия увеличивается.

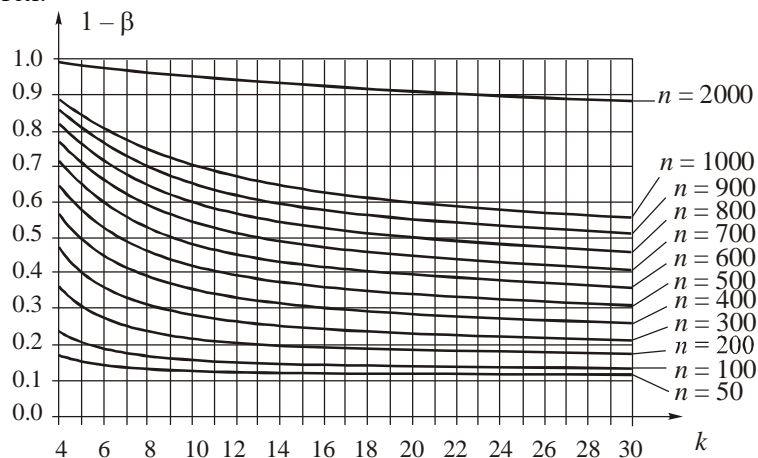


Рис. 2.18. Функции мощности критерия χ^2 Пирсона при проверке сложной гипотезы о согласии с нормальным законом при равновероятном группировании и альтернативе, соответствующей логистическому закону

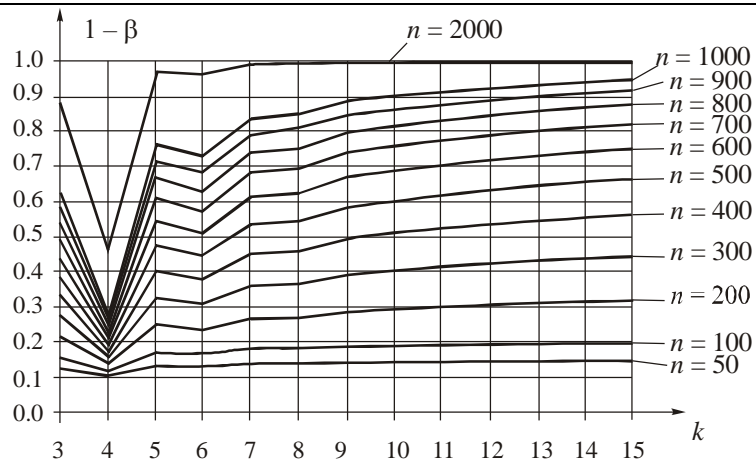


Рис. 2.19. Функции мощности критерия χ^2 Пирсона при проверке простой гипотезы о согласии с нормальным законом при асимптотически оптимальном группировании и альтернативе, соответствующей логистическому закону

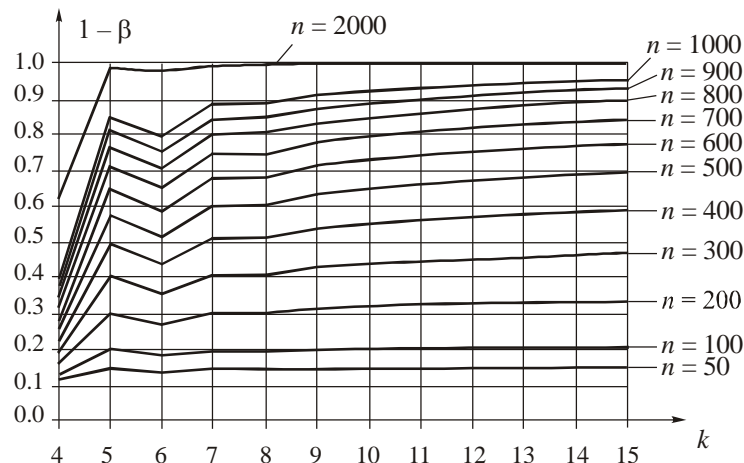


Рис. 2.20. Функции мощности критерия χ^2 Пирсона при проверке сложной гипотезы о согласии с нормальным законом при асимптотически оптимальном группировании и альтернативе, соответствующей логистическому закону

Функция мощности критерия типа χ^2 Никулина, как следует из рис. 2.21 и табл. 2.7, на области значений k , содержащей максимальное значение мощности, является выпуклой вверх функцией.

Отметим, что статистическое моделирование распределений статистик, которое проводилось нами в целях контроля результатов, полностью подтверждает аналитические расчеты, проиллюстрированные на приводимых рисунках.

Таким образом, выбирая число интервалов в критериях типа χ^2 , следует осознавать, что увеличение их числа не приводит к росту мощности критерия. Рекомендуется выбирать число интервалов k

так, чтобы для любого интервала при оптимальном группировании выполнялось условие $nP_i(\theta) \geq 5 \dots 10$. По крайней мере, минимальная ожидаемая частота должна быть больше 1. При этом следует опираться на соответствующую таблицу асимптотически оптимального группирования из приложения А. Указания на соответствующую таблицу асимптотически оптимального группирования даны в табл. 2.3. Снизу значение k ограничено условием $k \geq m + 2$.

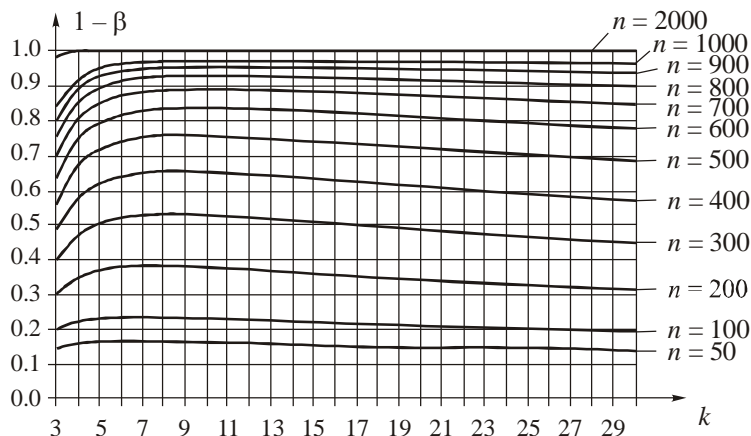


Рис. 2.21. Функции мощности критерия типа χ^2 Никулина при проверке сложной гипотезы о согласии с нормальным законом при равновероятном группировании и альтернативе, соответствующей логистическому закону

Если заданы конкретная конкурирующая гипотеза и объем выборки n , можно выбрать оптимальное число интервалов k так, чтобы максимизировать соотношение (2.20).

Некоторые выводы. При проверке согласия опытного распределения с теоретической моделью нас в

первую очередь должна интересоваться возможность с помощью критерия уловить отклонения в наблюдаемых данных, которые говорят о предпочтительности некоторой другой достаточно близкой к проверяемой модели, т. е. возможность различать близкие гипотезы. Это означает, что мы должны использовать критерии, которые обладают наибольшей мощностью относительно близких конкурирующих гипотез.

Максимизировать мощность критериев χ^2 Пирсона и отношения правдоподобия можно двумя путями: выбирая в качестве способа разбиения на интервалы асимптотически оптимальное группирование и подбирая «оптимальное» число интервалов k .

Мощность критерия Никулина можно максимизировать только за счет выбора «оптимального» числа интервалов и предпочтительного способа группирования, так как при малом числе интервалов предпочтительнее разбиение на интервалы равной вероятности, а при возрастании числа интервалов преимущество оказывается за асимптотически оптимальным.

Таблица 2.3

Теоретические значения мощности критерия χ^2 Пирсона в случае проверки простой гипотезы о согласии с экспоненциальным законом при равновероятном группировании

k	$n = 50$	$n = 200$	$n = 400$	$n = 600$	$n = 800$	$n = 1000$	$n = 2000$	$n = 3000$	$n = 4000$	$n = 5000$
2	0.1098	0.1391	0.1776	0.2154	0.2524	0.2884	0.4524	0.5876	0.6949	0.7775
3	0.1089	0.1360	0.1725	0.2092	0.2459	0.2823	0.4538	0.5993	0.7145	0.8015
4	0.1080	0.1324	0.1658	0.2000	0.2345	0.2692	0.4376	0.5853	0.7047	0.7957
5	0.1072	0.1294	0.1601	0.1917	0.2240	0.2568	0.4193	0.5663	0.6879	0.7824
6	0.1066	0.1270	0.1554	0.1848	0.2151	0.2460	0.4019	0.5468	0.6695	0.7668
7	0.1061	0.1251	0.1515	0.1790	0.2075	0.2366	0.3861	0.5282	0.6511	0.7505
8	0.1057	0.1234	0.1482	0.1741	0.2010	0.2286	0.3719	0.5108	0.6334	0.7343
9	0.1054	0.1221	0.1454	0.1700	0.1954	0.2217	0.3592	0.4948	0.6166	0.7185
10	0.1051	0.1209	0.1431	0.1663	0.1906	0.2156	0.3478	0.4801	0.6007	0.7033
11	0.1048	0.1199	0.1410	0.1632	0.1863	0.2102	0.3375	0.4665	0.5858	0.6888

Таблица 2.4

Теоретические значения мощности критерия χ^2 Пирсона в случае проверки простой гипотезы о согласии с экспоненциальным законом при асимптотически оптимальном группировании

k	$n = 50$	$n = 200$	$n = 400$	$n = 600$	$n = 800$	$n = 1000$	$n = 2000$	$n = 3000$	$n = 4000$	$n = 5000$
2	0.1127	0.1503	0.1996	0.2475	0.2939	0.3386	0.5342	0.6822	0.7885	0.8621
3	0.1109	0.1438	0.1883	0.2329	0.2772	0.3208	0.5200	0.6771	0.7909	0.8688
4	0.1094	0.1380	0.1774	0.2176	0.2582	0.2987	0.4910	0.6501	0.7697	0.8537
5	0.1082	0.1337	0.1689	0.2053	0.2425	0.2800	0.4630	0.6211	0.7445	0.8340
6	0.1074	0.1304	0.1624	0.1956	0.2298	0.2647	0.4383	0.5938	0.7194	0.8134
7	0.1068	0.1278	0.1571	0.1879	0.2196	0.2521	0.4169	0.5690	0.6955	0.7929
8	0.1063	0.1257	0.1529	0.1815	0.2111	0.2416	0.3983	0.5467	0.6732	0.7732
9	0.1058	0.1240	0.1494	0.1762	0.2040	0.2327	0.3821	0.5265	0.6525	0.7543
10	0.1055	0.1225	0.1465	0.1717	0.1979	0.2251	0.3678	0.5083	0.6333	0.7364
11	0.1052	0.1213	0.1439	0.1678	0.1927	0.2185	0.3552	0.4919	0.6156	0.7195

Таблица 2.5

Теоретические значения мощности критерия χ^2 Пирсона в случае проверки простой гипотезы о согласии с нормальным законом при равновероятном группировании

k	$n = 50$	$n = 100$	$n = 200$	$n = 300$	$n = 400$	$n = 500$	$n = 600$	$n = 700$	$n = 800$	$n = 900$	$n = 1000$	$n = 2000$
3	0.1167	0.1335	0.1675	0.2017	0.2359	0.2699	0.3035	0.3366	0.3691	0.4008	0.4317	0.6860
4	0.1175	0.1353	0.1719	0.2092	0.2469	0.2847	0.3223	0.3594	0.3958	0.4314	0.4660	0.7404
5	0.1173	0.1351	0.1720	0.2100	0.2488	0.2879	0.3270	0.3658	0.4040	0.4413	0.4775	0.7619
6	0.1168	0.1343	0.1705	0.2083	0.2471	0.2865	0.3260	0.3654	0.4043	0.4423	0.4794	0.7700
7	0.1163	0.1332	0.1686	0.2057	0.2440	0.2831	0.3225	0.3619	0.4009	0.4392	0.4766	0.7715
8	0.1157	0.1321	0.1664	0.2027	0.2402	0.2787	0.3177	0.3568	0.3957	0.4340	0.4714	0.7695
9	0.1152	0.1310	0.1643	0.1996	0.2363	0.2741	0.3125	0.3511	0.3896	0.4277	0.4651	0.7654
10	0.1146	0.1299	0.1623	0.1966	0.2325	0.2694	0.3071	0.3452	0.3832	0.4210	0.4581	0.7600
11	0.1142	0.1290	0.1604	0.1937	0.2287	0.2649	0.3018	0.3393	0.3768	0.4141	0.4509	0.7538
12	0.1137	0.1281	0.1586	0.1910	0.2251	0.2605	0.2967	0.3334	0.3704	0.4072	0.4437	0.7472
13	0.1133	0.1273	0.1569	0.1885	0.2217	0.2563	0.2917	0.3278	0.3642	0.4005	0.4365	0.7402
14	0.1129	0.1265	0.1553	0.1861	0.2185	0.2523	0.2870	0.3224	0.3582	0.3940	0.4296	0.7330
15	0.1126	0.1258	0.1538	0.1838	0.2155	0.2485	0.2825	0.3172	0.3524	0.3877	0.4228	0.7258

Таблица 2.6

Теоретические значения мощности критерия χ^2 Пирсона в случае проверки простой гипотезы о согласии с нормальным законом при асимптотически оптимальном группировании

k	$n = 50$	$n = 100$	$n = 200$	$n = 300$	$n = 400$	$n = 500$	$n = 600$	$n = 700$	$n = 800$	$n = 900$	$n = 1000$	$n = 2000$
3	0.1273	0.1550	0.2108	0.2664	0.3212	0.3743	0.4255	0.4744	0.5206	0.5641	0.6048	0.8693
4	0.1254	0.1515	0.2050	0.2596	0.3140	0.3676	0.4197	0.4698	0.5175	0.5626	0.6049	0.8778
5	0.1249	0.1507	0.2044	0.2597	0.3156	0.3709	0.4249	0.4770	0.5266	0.5735	0.6173	0.8929
6	0.1233	0.1476	0.1987	0.2519	0.3060	0.3602	0.4135	0.4653	0.5149	0.5620	0.6063	0.8893
7	0.1220	0.1452	0.1940	0.2453	0.2980	0.3510	0.4036	0.4548	0.5043	0.5515	0.5961	0.8854
8	0.1208	0.1427	0.1891	0.2382	0.2890	0.3404	0.3916	0.4419	0.4908	0.5376	0.5822	0.8778
9	0.1197	0.1405	0.1848	0.2319	0.2808	0.3305	0.3804	0.4297	0.4778	0.5242	0.5686	0.8697
10	0.1188	0.1386	0.1808	0.2259	0.2729	0.3211	0.3696	0.4177	0.4649	0.5107	0.5547	0.8606
11	0.1179	0.1368	0.1773	0.2205	0.2658	0.3124	0.3595	0.4065	0.4528	0.4979	0.5414	0.8511
12	0.1171	0.1352	0.1738	0.2153	0.2589	0.3038	0.3495	0.3952	0.4405	0.4848	0.5277	0.8406
13	0.1164	0.1338	0.1711	0.2111	0.2532	0.2968	0.3412	0.3859	0.4303	0.4739	0.5164	0.8316
14	0.1158	0.1325	0.1684	0.2070	0.2477	0.2899	0.3331	0.3767	0.4201	0.4629	0.5048	0.8218
15	0.1153	0.1314	0.1659	0.2032	0.2426	0.2835	0.3255	0.3680	0.4105	0.4526	0.4938	0.8121

Таблица 2.7

Теоретические значения мощности критерия типа χ^2 Никулина при проверке сложной гипотезы о согласии с нормальным законом при равновероятном группировании и альтернативе, соответствующей логистическому закону

k	$n = 50$	$n = 100$	$n = 200$	$n = 300$	$n = 400$	$n = 500$	$n = 600$	$n = 700$	$n = 800$	$n = 900$	$n = 1000$	$n = 2000$
3	0.15009	0.20093	0.30197	0.39866	0.48818	0.56898	0.64044	0.70258	0.75585	0.80096	0.83876	0.984880
4	0.16129	0.22521	0.35389	0.47530	0.58339	0.67556	0.75158	0.81261	0.86051	0.89740	0.92534	0.997981
5	0.16473	0.23371	0.37447	0.50700	0.62283	0.71871	0.79483	0.85323	0.89681	0.92856	0.95123	0.999382
6	0.16544	0.23635	0.38291	0.52137	0.64148	0.73937	0.81540	0.87221	0.91332	0.94225	0.96216	0.999710
7	0.16501	0.23638	0.38559	0.52745	0.65030	0.74968	0.82594	0.88201	0.92181	0.94922	0.96762	0.999825
8	0.16407	0.23514	0.38528	0.52909	0.65386	0.75452	0.83126	0.88717	0.92638	0.95301	0.97059	0.999876
9	0.16291	0.23328	0.38333	0.52821	0.65437	0.75618	0.83360	0.88971	0.92878	0.95507	0.97225	0.999903
10	0.16168	0.23113	0.38046	0.52582	0.65301	0.75588	0.83409	0.89064	0.92986	0.95611	0.97313	0.999917
11	0.16043	0.22886	0.37709	0.52252	0.65049	0.75432	0.83338	0.89053	0.93009	0.95649	0.97353	0.999924
12	0.15921	0.22656	0.37346	0.51868	0.64721	0.75193	0.83186	0.88971	0.92975	0.95642	0.97360	0.999927
13	0.15803	0.22430	0.36971	0.51452	0.64344	0.74897	0.82979	0.88840	0.92900	0.95604	0.97344	0.999924
14	0.15690	0.22210	0.36594	0.51018	0.63937	0.74564	0.82732	0.88673	0.92795	0.95543	0.97311	0.999916
15	0.15583	0.21997	0.36221	0.50576	0.63511	0.74204	0.82458	0.88479	0.92668	0.95464	0.97266	0.999902

Окончание табл. 2.7

k	$n = 50$	$n = 100$	$n = 200$	$n = 300$	$n = 400$	$n = 500$	$n = 600$	$n = 700$	$n = 800$	$n = 900$	$n = 1000$	$n = 2000$
16	0.15481	0.21793	0.35854	0.50133	0.63075	0.73828	0.82164	0.88267	0.92524	0.95372	0.97210	0.999880
17	0.15384	0.21598	0.35496	0.49693	0.62633	0.73440	0.81855	0.88040	0.92367	0.95270	0.97147	0.999850
18	0.15293	0.21411	0.35148	0.49259	0.62190	0.73045	0.81537	0.87802	0.92201	0.95160	0.97077	0.999809
19	0.15206	0.21233	0.34811	0.48832	0.61750	0.72648	0.81212	0.87556	0.92026	0.95042	0.97001	0.999756
20	0.15123	0.21062	0.34485	0.48414	0.61313	0.72249	0.80882	0.87304	0.91845	0.94920	0.96922	0.999690
21	0.15045	0.20899	0.34169	0.48005	0.60881	0.71851	0.80549	0.87047	0.91660	0.94792	0.96839	0.999609
22	0.14971	0.20744	0.33865	0.47606	0.60456	0.71455	0.80216	0.86787	0.91470	0.94661	0.96752	0.999512
23	0.14899	0.20595	0.33571	0.47218	0.60038	0.71062	0.79882	0.86525	0.91277	0.94527	0.96663	0.999397
24	0.14832	0.20452	0.33287	0.46839	0.59627	0.70673	0.79549	0.86262	0.91082	0.94390	0.96571	0.999263
25	0.14767	0.20315	0.33013	0.46471	0.59224	0.70288	0.79217	0.85998	0.90885	0.94251	0.96478	0.999108
26	0.14705	0.20185	0.32748	0.46113	0.58829	0.69909	0.78887	0.85733	0.90686	0.94110	0.96382	0.998931
27	0.14646	0.20059	0.32493	0.45764	0.58443	0.69534	0.78560	0.85469	0.90486	0.93967	0.96285	0.998732
28	0.14590	0.19938	0.32246	0.45425	0.58064	0.69165	0.78235	0.85205	0.90286	0.93823	0.96186	0.998508
29	0.14536	0.19822	0.32007	0.45095	0.57693	0.68802	0.77913	0.84942	0.90085	0.93678	0.96087	0.998259
30	0.14483	0.19710	0.31776	0.44774	0.57330	0.68443	0.77594	0.84680	0.89884	0.93532	0.95986	0.997984

Найти «оптимальное» число интервалов k для соответствующего критерия можно, максимизировав по k соотношение (2.20). При этом фиксируются пара конкурирующих гипотез, способ группирования, объем выборки n и вероятность ошибки первого рода α .

Мощность критериев χ^2 Пирсона и отношения правдоподобия часто оказывается максимальной относительно близких конкурирующих гипотез, если выборку разбивать на минимально возможное число интервалов группирования. С ростом же числа интервалов мощность критериев падает. Этот факт ускользает от внимания большинства исследователей, использующих данный критерий, и совсем не упоминается в рекомендациях различного уровня.

В некоторых ситуациях, т. е. при конкретных парах конкурирующих гипотез, функции мощности критериев χ^2 Пирсона и отношения правдоподобия оказываются выпуклыми вверх по k и существует «оптимальное» значение числа интервалов. Однако это «оптимальное» значение обычно достаточно мало отличается от минимально возможного и незначительно изменяется в сторону увеличения при значительном росте объема выборки n . Прослеживается зависимость «оптимальной» величины k не только от объема выборки, пары конкурирующих гипотез, но и от способа группирования.

Функция мощности критерия Никулина обычно оказывается выпуклой вверх по k . Поэтому «оптимальное» число интервалов обычно существует. Это «оптимальное» значение растет с увеличением объема выборки. Но его величина также меньше значений, рекомендуемых любыми действующими регламентирующими документами и справочными источниками.

И еще одно замечание. В последние годы в сознании многих исследователей сформировалось устойчивое мнение о критериях типа χ^2 как о плохих критериях согласия. Поэтому естествен вопрос: зачем использовать критерии типа χ^2 , если есть критерии Колмогорова и ω^2 Крамера–Мизеса–Смирнова? Имеется как минимум две причины, обосновывающие рекомендации применять критерии типа χ^2 . Во-первых, в случае проверки простых гипотез и использовании в критериях типа χ^2 асим-

птотически оптимального группирования с позиций наибольшей мощности относительно близких альтернатив критерии типа χ^2 оказываются предпочтительнее непараметрических критериев согласия Колмогорова, Крамера–Мизеса–Смирнова, Андерсона–Дарлинга [237, 79].

Во-вторых, непараметрические критерии при проверке сложных гипотез теряют «свободу от распределения», в то время как для критериев Никулина и χ^2 Пирсона предельные распределения статистик при проверке сложных гипотез известны. Конечно, для некоторых частных случаев проверки сложных гипотез различными способами получены модели предельных распределений статистик критериев типа Колмогорова, Крамера–Мизеса–Смирнова, Андерсона–Дарлинга. Однако это далеко не снимает остроту проблемы. Есть и третий довод. Критерии согласия используют различные меры близости распределений, по-разному улавливают различные отклонения в наблюдаемых данных от предполагаемых законов распределений. Поэтому не следует обеднять методику статистического анализа, отказываясь от какого-либо из зарекомендовавших себя критериев. Лучше всего использовать их совокупность.

Более подробная информация о сравнительном анализе мощности критериев согласия приведена в разделе 4.

2.9. Рекомендации по использованию асимптотически оптимального группирования в критериях согласия

Применяя критерии согласия типа χ^2 , можно по-разному разбивать область определения случайной величины на интервалы: равной длины, равных вероятностей или асимптотически оптимальные. Использование асимптотически оптимальных интервалов обеспечивает максимальную

мощность применяемого критерия против близких альтернатив, снижает риск принятия неверной нулевой гипотезы H_0 .

Если гипотеза H_0 справедлива или, наоборот, совершенно не соответствует характеру выборки, то выводы по критериям согласия при равновероятном, равноотстоящем и асимптотически оптимальном группировании обычно оказываются очень близкими по вычисляемому значению вероятности $P\{S > S^*\} > \alpha$. И гипотезу H_0 при любом группировании однозначно принимают или однозначно отвергают.

В случае равновероятного и равноотстоящего группирования наличие в выборке малых отклонений от предположений (от гипотезы H_0) оказывается незамеченным критерием, а в случае асимптотически оптимального группирования эти отклонения будут заметны, что в принципе снижает риск принятия неверной нулевой гипотезы H_0 .

Применяя критерии согласия типа χ^2 , не обязательно использовать только асимптотически оптимальное группирование наблюдений. Можно использовать и равновероятное группирование, и разбиение на интервалы равной длины. Но тогда следует помнить, что в этом случае критерии типа χ^2 будут хуже различать близкие гипотезы (близкие альтернативы).

В приложении А приведены таблицы асимптотически оптимального группирования (АОГ) для ряда распределений. Представлены таблицы двух видов. В таблицах первого вида содержатся граничные точки, инвариантные относительно параметров распределения, в таблицах второго – вероятности попадания в соответствующие интервалы.

Порядок использования асимптотически оптимального группирования в связи с конкретным законом распределения определяется табл. 2.8. В строках таблицы отражены функция плотности закона распределения, количество информации Фишера о соответствующем параметре по группированным и негруппированным данным, указаны номера таблиц в приложении А, в которых содержатся оптималь-

ные для данной ситуации граничные точки в виде $t_i = \varphi(x_i, \theta)$, инвариантном относительно θ , и соответствующие значения вероятностей P_i попадания наблюдений в i -й интервал.

Для тех законов распределения, для которых таблицы асимптотически оптимального группирования отсутствуют, поскольку решение задачи асимптотически оптимального группирования нельзя получить в виде, инвариантном относительно параметров распределений (например, для распределений типа Накагами, бета-распределений), в соответствующих колонках табл. 2.8 стоят прочерки (тире). В таких случаях, чтобы применить асимптотически оптимальное группирование, можно рекомендовать решение задач вида (2.18), (2.19) в процессе проверки согласия при определенных значениях параметров, используя содержащиеся в табл. 2.8 выражения для элементов информационной матрицы Фишера по группированным данным.

Для законов распределения, определяемых одним параметром (экспоненциального, Парето, Рэлея и т. п.), одну и ту же пару таблиц асимптотически оптимального группирования (t_i и P_i) используют при проверке простых и сложных гипотез.

Для законов распределения, определяемых двумя параметрами, в случае сложной гипотезы и оценивания только одного параметра требуемая пара таблиц асимптотически оптимального группирования указана в строке для этого параметра. В случае сложной гипотезы и оценивания двух параметров или в случае простой гипотезы используют пару таблиц, указанных в строке с двумя параметрами. В колонках для количества информации Фишера по группированным и негруппированным данным в такой строке даны выражения для недиагональных элементов соответствующих информационных матриц $J(\theta_i, \theta_j)$, а $J(\theta_i, \theta_j) = J(\theta_j, \theta_i)$. Например, при оценивании по выборке только параметра сдвига нормального закона следует использовать пару табл. А.24–А.25, при оценивании только параметра масштаба – пару табл. А.26–А.27, а при проверке простых гипотез и сложных гипотез, сопровождаемых оцениванием по данной выборке двух параметров нормального закона, – пару табл. А.28–А.29.

В работах [169, 170, 171] задача асимптотически оптимального группирования была решена для критериев А- и Е-оптимальности:

$$\max_{x_0 < x_1 < \dots < x_{k-1} < x_k} Sp\mathbf{J}_\Gamma(\theta),$$

$$\max_{x_0 < x_1 < \dots < x_{k-1} < x_k} \min_{i=\overline{1,r}} \lambda_i(\mathbf{J}_\Gamma(\theta)).$$

В разделе 4.9 будет показано, что при A -оптимальном группировании мощность критерия Пирсона, хотя и незначительно, но выше, чем при D -оптимальном. В табл. А.59 – А.66 приложения **А** представлены граничные точки интервалов и соответствующие вероятности попадания в интервал при A -оптимальном группировании для некоторых распределений.

Таблица 2.8

Функции плотности распределений, элементы информационных матриц Фишера по группированным и негруппированным наблюдениям, соответствующие таблицы асимптотически оптимального группирования

Функция плотности распределения	Параметр распределения	Количество информации Фишера для параметра по наблюдениям		Номер таблицы АОГ	
		группированным	негруппированным	t_i	P_i
1. Экспоненциальное $\frac{1}{\theta_0} e^{-(x-\theta_1)/\theta_0}$ $x \in (\theta_1, \infty)$ $t_i = (x - \theta_1) / \theta_0$	θ_0	$\frac{1}{\theta_0^2} \sum_{i=1}^k \frac{\left(t_i e^{-t_i} - t_{i-1} e^{-t_{i-1}}\right)^2}{e^{-t_{i-1}} - e^{-t_i}}$	$\frac{1}{\theta_0^2}$	А.1	А.2
2. Полунормальное $\frac{2}{\theta_0 \sqrt{2\pi}} e^{-(x-\theta_1)^2/2\theta_0^2}$ $x \in (\theta_1, \infty)$ $t_i = (x_i - \theta_1) / \theta_0$	θ_0	$\frac{2}{\theta_0^2} \sum_{i=1}^k \frac{\left(t_{i-1} e^{-t_{i-1}^2/2} - t_i e^{-t_i^2/2}\right)^2}{\pi\{\Phi_0(t_i) - \Phi_0(t_{i-1})\}}$ $\Phi_0(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-t^2/2} dt$	$\frac{2}{\theta_0^2}$	А.6	А.7
3. Рэлея $\frac{(x-\theta_1)}{\theta_0^2} e^{-(x-\theta_1)^2/2\theta_0^2}$ $x \in (\theta_1, \infty)$ $t_i = (x_i - \theta_1) / \theta_0$	θ_0	$\frac{1}{\theta_0^2} \sum_{i=1}^k \frac{\left(t_{i-1}^2 e^{-t_{i-1}^2/2} - t_i^2 e^{-t_i^2/2}\right)^2}{e^{-t_{i-1}^2/2} - e^{-t_i^2/2}}$	$\frac{4}{\theta_0^2}$	А.3	А.2

Функция плотности распределения	Параметр распределения	Количество информации Фишера для параметра по наблюдениям		Номер таблицы АОГ	
		группированным	негруппированным	t_i	P_i
4. Максвелла $\frac{2(x-\theta_1)^2}{\theta_0^3 \sqrt{2\pi}} e^{-(x-\theta_1)^2/2\theta_0^2}$ $x \in (\theta_1, \infty)$ $t_i = (x_i - \theta_1) / \theta_0$	θ_0	$\frac{4}{\theta_0^2} \sum_{i=1}^k \frac{\left[t_{i-1}^3 \varphi(t_{i-1}) - t_i^3 \varphi(t_i) \right]^2}{\Phi_0(t_i) - 2t_i \varphi(t_i) - \Phi_0(t_{i-1}) + 2t_{i-1} \varphi(t_{i-1})}$	$\frac{6}{\theta_0^2}$	A.4	A.5
5. Модуля нормального вектора $\frac{2(x-\theta_1)^{m-1}}{(2\theta_0^2)^{m/2} \Gamma(m/2)} e^{-(x-\theta_1)^2/2\theta_0^2}$ $x \in (\theta_1, \infty)$ $t_i = (x_i - \theta_1) / \theta_0$ $\xi = 2^{m/2-1} \Gamma(m/2)$	θ_0	$\frac{1}{\theta_0^2} \sum_{i=1}^k \frac{\left(t_{i-1}^m e^{-t_{i-1}^2/2} - t_i^m e^{-t_i^2/2} \right)^2}{\xi^2 [F(t_i) - F(t_{i-1})]},$ <p>при четном m</p> $F(t) = 1 - e^{-t^2/2} \left(1 + \frac{t^2}{2} + \frac{t^4}{8} + \dots + \frac{t^{m-2}}{2^{m/2-1} \Gamma(m/2)} \right),$ <p>при нечетном</p> $F(t) = \Phi_0(t) - e^{-t^2/2} \left(\frac{t}{2^{1/2} \Gamma(3/2)} + \frac{t^3}{2^{3/2} \Gamma(5/2)} + \dots + \frac{t^{m-2}}{2^{m/2-1} \Gamma(m/2)} \right)$	$\frac{2m}{\theta_0^2}$ $m = 4$ $m = 5$ $m = 6$ $m = 7$ $m = 8$ $m = 9$	A.8 A.10 A.12 A.14 A.16 A.18	A.9 A.11 A.13 A.15 A.17 A.19

6. Парето $\theta_0 \theta_1^{\theta_0} x^{-\theta_0-1}$ $x \in (\theta_1, \infty)$ $t_i = -\ln s_i = \theta_0 \ln(x_i / \theta_1)$	θ_0	$\frac{1}{\theta_0^2} \sum_{i=1}^k \frac{(s_{i-1} \ln s_{i-1} - s_i \ln s_i)^2}{s_{i-1} - s_i} =$ $= \frac{1}{\theta_0^2} \sum_{i=1}^k \frac{(t_i e^{-t_i} - t_{i-1} e^{-t_{i-1}})^2}{e^{-t_{i-1}} - e^{-t_i}}$	$\frac{1}{\theta_0^2}$	A.1	A.2
7. Эрланга порядка m $\frac{\theta_1^m}{\Gamma(m)} (x - \theta_2)^{m-1} e^{-\theta_1(x - \theta_2)}$ $x \in (\theta_1, \infty)$ $t_i = \theta_1(x_i - \theta_2)$	θ_1	$\frac{1}{\theta_1^2} \sum_{i=1}^k \frac{(t_i^m e^{-t_i} - t_{i-1}^m e^{-t_{i-1}})^2}{\Gamma(m) \int_{t_{i-1}}^{t_i} t^{m-1} e^{-t} dt}$	$\frac{m}{\theta_1^2}$	A.54	—
8. Лапласа $\frac{\theta_0}{2} e^{-\theta_0 x - \theta_1 }$ $x \in (-\infty, \infty)$ $t_i = \theta_0(x_i - \theta_1)$	θ_0	$\sum_{i=1}^k \left(\frac{\partial P_i(\theta)}{\partial \theta_0} \right)^2 \bigg/ P_i(\theta)$ $P_i(\theta) = \begin{cases} \frac{1}{2} (e^{t_i} - e^{t_{i-1}}), & t_i < 0; \\ \frac{1}{2} (e^{-t_{i-1}} - e^{-t_i}), & t_{i-1} > 0; \\ 1 - \frac{1}{2} (e^{-t_i} + e^{t_{i-1}}), & t_{i-1} < 0 < t_i \end{cases}$	$\frac{1}{\theta_0^2}$	A.46 A.47	A.48 A.49

Функция плотности распределения	Пара- метр распре- деления	Количество информации Фишера для параметра по наблюдениям		Номер таб- лицы АОГ	
		группированным	негруппи- рованным	t_i	P_i
9. Нормальное $\frac{1}{\theta_1 \sqrt{2\pi}} e^{-\frac{(x-\theta_0)^2}{2\theta_1^2}}$ $x \in (-\infty, \infty)$ $t_i = \frac{x_i - \theta_0}{\theta_1}$	θ_0	$\frac{1}{\theta_1^2} \sum_{i=1}^k \frac{[\varphi(t_{i-1}) - \varphi(t_i)]^2}{\Phi(t_i) - \Phi(t_{i-1})}$ $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}, \quad \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-t^2/2} dt$	$\frac{1}{\theta_1^2}$	A.24	A.25
	θ_1	$\frac{1}{\theta_1^2} \sum_{i=1}^k \frac{[t_{i-1}\varphi(t_{i-1}) - t_i\varphi(t_i)]^2}{\Phi(t_i) - \Phi(t_{i-1})}$	$\frac{2}{\theta_1^2}$	A.26	A.27
	θ_0, θ_1	$\frac{1}{\theta_1^2} \sum_{i=1}^k \frac{[\varphi(t_i) - \varphi(t_{i-1})][t_{i-1}\varphi(t_{i-1}) - t_i\varphi(t_i)]}{\Phi(t_i) - \Phi(t_{i-1})}$	0	A.28	A.29
10. Логарифмически нормальное (ln) $\frac{1}{x\theta_1 \sqrt{2\pi}} e^{-(\ln x - \theta_0)^2 / 2\theta_1^2}$ $x \in (0, \infty)$ $t_i = \frac{\ln x_i - \theta_0}{\theta_1}$	θ_0	$\frac{1}{\theta_1^2} \sum_{i=1}^k \frac{[\varphi(t_{i-1}) - \varphi(t_i)]^2}{\Phi(t_i) - \Phi(t_{i-1})}$	$\frac{1}{\theta_1^2}$	A.24	A.25
	θ_1	$\frac{1}{\theta_1^2} \sum_{i=1}^k \frac{[t_{i-1}\varphi(t_{i-1}) - t_i\varphi(t_i)]^2}{\Phi(t_i) - \Phi(t_{i-1})}$	$\frac{2}{\theta_1^2}$	A.26	A.27
	θ_0, θ_1	$\frac{1}{\theta_1^2} \sum_{i=1}^k \frac{[\varphi(t_i) - \varphi(t_{i-1})][t_{i-1}\varphi(t_{i-1}) - t_i\varphi(t_i)]}{\Phi(t_i) - \Phi(t_{i-1})}$	0	A.28	A.29

11. Логарифмически нормальное (lg) $\frac{1}{x\theta_1 \ln 10 \sqrt{2\pi}} e^{-(\lg x - \theta_0)^2 / 2\theta_1^2}$ $x \in (0, \infty)$ $t_i = \frac{\lg x_i - \theta_0}{\theta_1}$	θ_0	$\frac{1}{\theta_1^2} \sum_{i=1}^k \frac{[\varphi(t_{i-1}) - \varphi(t_i)]^2}{\Phi(t_i) - \Phi(t_{i-1})}$	$\frac{1}{\theta_1^2}$	A.24	A.25
	θ_1	$\frac{1}{\theta_1^2} \sum_{i=1}^k \frac{[t_{i-1}\varphi(t_{i-1}) - t_i\varphi(t_i)]^2}{\Phi(t_i) - \Phi(t_{i-1})}$	$\frac{2}{\theta_1^2}$	A.26	A.27
	θ_0, θ_1	$\frac{1}{\theta_1^2} \sum_{i=1}^k \frac{[\varphi(t_i) - \varphi(t_{i-1})][t_{i-1}\varphi(t_{i-1}) - t_i\varphi(t_i)]}{\Phi(t_i) - \Phi(t_{i-1})}$	0	A.28	A.29
12. Коши $\frac{\theta_0}{\pi [\theta_0^2 + (x - \theta_1)^2]}$ $x \in (-\infty, \infty)$ $t_i = (x_i - \theta_1) / \theta_0$	θ_0	$\frac{1}{\theta_0^2} \sum_{i=1}^k \frac{1}{\pi} \frac{\left[\frac{t_{i-1}}{1+t_{i-1}^2} - \frac{t_i}{1+t_i^2} \right]^2}{\arctg t_i - \arctg t_{i-1}}$	$\frac{1}{2\theta_0^2}$	A.36	A.37
	θ_1	$\frac{1}{\theta_0^2} \sum_{i=1}^k \frac{1}{\pi} \frac{\left[\frac{1}{1+t_{i-1}^2} - \frac{1}{1+t_i^2} \right]^2}{\arctg t_i - \arctg t_{i-1}}$	$\frac{1}{2\theta_0^2}$	A.38	A.39
	θ_0, θ_1	$\frac{1}{\pi \theta_0^2} \sum_{i=1}^k \frac{\left(\frac{t_{i-1}}{1+t_{i-1}^2} - \frac{t_i}{1+t_i^2} \right) \left(\frac{1}{1+t_{i-1}^2} - \frac{1}{1+t_i^2} \right)}{\arctg t_i - \arctg t_{i-1}}$	0	A.40	$P_i = \frac{1}{k}$

Функция плотности распределения	Параметр распределения	Количество информации Фишера для параметра по наблюдениям		Номер таблицы АОГ	
		группированным	негруппированным	t_i	P_i
13. Логистическое $\frac{\pi}{\theta_1 \sqrt{3}} \times \frac{\exp \left\{ -\frac{\pi(x - \theta_0)}{\theta_1 \sqrt{3}} \right\}}{\left[1 + \exp \left\{ -\frac{\pi(x - \theta_0)}{\theta_1 \sqrt{3}} \right\} \right]^2}$ $x \in (-\infty, \infty)$ $t_i = \pi(x_i - \theta_0) / \theta_1 \sqrt{3}$	θ_0	$\frac{\pi}{\theta_1^2 3} \sum_{i=1}^k \frac{\left[\frac{e^{-t_i}}{(1+e^{-t_i})^2} - \frac{e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2} \right]^2}{(1+e^{-t_i})^{-1} - (1+e^{-t_{i-1}})^{-1}}$	$\frac{\pi^2}{\theta_1^2 9}$	A.41	$P_i = \frac{1}{k}$
	θ_1	$\frac{1}{\theta_1^2} \sum_{i=1}^k \frac{\left[\frac{t_i e^{-t_i}}{(1+e^{-t_i})^2} - \frac{t_{i-1} e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2} \right]^2}{(1+e^{-t_i})^{-1} - (1+e^{-t_{i-1}})^{-1}}$	$\frac{1}{\theta_1^2} \frac{\pi^2 + 3}{9}$	A.42	A.43

	θ_0, θ_1	$\frac{\pi^2}{\theta_1^2 \sqrt{3}} \sum_{i=1}^k \left[\frac{\frac{t_i e^{-t_i}}{(1+e^{-t_i})^2} - \frac{t_{i-1} e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2}}{(1+e^{-t_i})^{-1} - (1+e^{-t_{i-1}})^{-1}} \right] \times$ $\times \left[\frac{e^{-t_i}}{(1+e^{-t_i})^2} - \frac{e^{-t_{i-1}}}{(1+e^{-t_{i-1}})^2} \right]$		0	A.44 A.45
14. Вейбулла	θ_0	$\frac{1}{\theta_0^2} \sum_{i=1}^k \left[\frac{t_i \ln t_i e^{-t_i} - t_{i-1} \ln t_{i-1} e^{-t_{i-1}}}{e^{-t_{i-1}} - e^{-t_i}} \right]^2$		$\frac{1}{\theta_0^2} \times$ $\times \left(1 + \frac{\pi^2}{6} + C^2 - 2C \right)$	A.20 A.21
	θ_1	$\frac{\theta_0^2}{\theta_1^2} \sum_{i=1}^k \frac{[t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i}]^2}{e^{-t_{i-1}} - e^{-t_i}}$		$\frac{\theta_0^2}{\theta_1^2}$	A.1 A.2
	θ_0, θ_1	$\frac{1}{\theta_1} \sum_{i=1}^k \frac{(t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i})(t_i \ln t_i e^{-t_i} - t_{i-1} \ln t_{i-1} e^{-t_{i-1}})}{e^{-t_{i-1}} - e^{-t_i}}$		$\frac{1}{\theta_1} (C-1)$	A.22 A.23

$\frac{\theta_0 (x - \theta_2)^{\theta_0 - 1}}{\theta_1^{\theta_0}} \times$
 $\times \exp \left\{ - \left(\frac{x - \theta_2}{\theta_1} \right)^{\theta_0} \right\}$
 $x \in (\theta_1, \infty)$
 $t_i = [(x_i - \theta_2) / \theta_1]^{\theta_0}$
 $C = 0.577215664$ –
 постоянная Эйлера

Функция плотности распределения	Параметр распределения	Количество информации Фишера для параметра по наблюдениям		Номер таблицы АОГ	
		группированным	негруппированным	t_i	P_i
15. Минимального значения $\frac{1}{\theta_1} \exp \left\{ \frac{x - \theta_0}{\theta_1} - \exp \left(\frac{x - \theta_0}{\theta_1} \right) \right\}$ $x \in (-\infty, \infty),$ $t_i = \exp(s_i),$ $s_i = \frac{x_i - \theta_0}{\theta_1}$	θ_0	$\frac{1}{\theta_1^2} \sum_{i=1}^k \frac{[t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i}]^2}{e^{-t_{i-1}} - e^{-t_i}}$	$\frac{1}{\theta_1^2}$	A.1	A.2
	θ_1	$\frac{1}{\theta_1^2} \sum_{i=1}^k \frac{[t_{i-1} s_{i-1} e^{-t_{i-1}} - t_i s_i e^{-t_i}]^2}{e^{-t_{i-1}} - e^{-t_i}}$	$\frac{1}{\theta_1^2} \left(1 + \frac{\pi^2}{6} + C^2 - 2C \right)$	A.20	A.21
	θ_0, θ_1	$\frac{1}{\theta_1^2} \sum_{i=1}^k \frac{(t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i})(t_{i-1} s_{i-1} e^{-t_{i-1}} - t_i s_i e^{-t_i})}{e^{-t_{i-1}} - e^{-t_i}}$	$\frac{1}{\theta_1^2} (C - 1)$	A.22	A.23
16. Максимального значения $\frac{1}{\theta_1} \exp \left\{ -\frac{x - \theta_0}{\theta_1} - \exp \left(-\frac{x - \theta_0}{\theta_1} \right) \right\}$ $x \in (-\infty, \infty)$ $t_i = \exp(s_i)$ $s_i = \frac{x_i - \theta_0}{\theta_1}$	θ_0	$\frac{1}{\theta_1^2} \sum_{i=1}^k \frac{[t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i}]^2}{e^{-t_i} - e^{-t_{i-1}}}$	$\frac{1}{\theta_1^2}$	A.30	A.31
	θ_1	$\frac{1}{\theta_1^2} \sum_{i=1}^k \frac{[t_i s_i e^{-t_i} - t_{i-1} s_{i-1} e^{-t_{i-1}}]^2}{e^{-t_i} - e^{-t_{i-1}}}$	$\frac{1}{\theta_1^2} \left(1 + \frac{\pi^2}{6} + C^2 - 2C \right)$	A.32	A.33

	θ_0, θ_1	$\frac{1}{\theta_1^2} \sum_{i=1}^k \frac{(t_{i-1}e^{-t_{i-1}} - t_i e^{-t_i})(t_i s_i e^{-t_i} - t_{i-1} s_{i-1} e^{-t_{i-1}})}{e^{-t_i} - e^{-t_{i-1}}}$	$\frac{1}{\theta_1^2} (C-1)$	A.34	A.35
17. Двойное показательное $f(x) = \theta_0 \theta_1 \exp\{-(x - \theta_2) \theta_0 -$ $\quad - \theta_1 \exp[-(x - \theta_2) \theta_0]\}$ $x \in (-\infty, \infty)$ $t_i = \theta_1 \exp\{-s_i\}$ $s_i = \theta_0 (x_i - \theta_2)$	θ_0	$\frac{1}{\theta_0^2} \sum_{i=1}^k \frac{[t_i s_i e^{-t_i} - t_{i-1} s_{i-1} e^{-t_{i-1}}]^2}{e^{-t_i} - e^{-t_{i-1}}}$	$\frac{1}{\theta_0^2} \left[\ln^2 \theta_1 + \right.$ $+ 2(C-1) \ln \theta_1 +$ $\left. + 1 + \frac{\pi}{6} + C^2 - 2C \right]$	A.50	A.51
	θ_1	$\frac{1}{\theta_1^2} \sum_{i=1}^k \frac{[t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i}]^2}{e^{-t_i} - e^{-t_{i-1}}}$	$\frac{1}{\theta_1^2}$	A.1	A.31
	θ_0, θ_1	$\frac{1}{\theta_0 \theta_1} \times$ $\times \sum_{i=1}^k \frac{(t_i s_i e^{-t_i} - t_{i-1} s_{i-1} e^{-t_{i-1}})(t_{i-1} e^{-t_{i-1}} - t_i e^{-t_i})}{e^{-t_i} - e^{-t_{i-1}}}$	$\frac{1}{\theta_0 \theta_1} \times$ $\times (1 - C - \ln \theta_2)$	A.52	A.35
18. Накагами $\frac{2}{\Gamma(\theta_1)} \left(\frac{\theta_1}{\theta_0^2} \right)^{\theta_1} (x - \theta_2)^{2\theta_1 - 1} \times$ $\times \exp \left\{ -\frac{\theta_1 (x - \theta_2)^2}{\theta_0^2} \right\}$	θ_0	$\frac{4}{\theta_0^2 \Gamma(\theta_1)} \sum_{i=1}^k \frac{[t_{i-1}^{\theta_1} e^{-t_{i-1}} - t_i^{\theta_1} e^{-t_i}]^2}{\int_{t_{i-1}}^{t_i} t^{\theta_1 - 1} e^{-t} dt}$	$\frac{4\theta_1}{\theta_0^2}$	—	—

Функция плотности распределения	Параметр распределения	Количество информации Фишера для параметра по наблюдениям		Номер таблицы АОГ	
		группированным	негруппированным	t_i	P_i
$x \in (\theta_2, \infty)$ $t_i = \frac{\theta_1(x_i - \theta_2)^2}{\theta_0^2}$ $\psi(\theta_1) = \Gamma'(\theta_1)/\Gamma(\theta_1)$	θ_1	$\frac{1}{\Gamma(\theta_1)} \sum_{i=1}^k \left\{ \int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} \ln t dt + \frac{1}{\theta_1} \left(t_i^{\theta_1} e^{-t_i} - t_{i-1}^{\theta_1} e^{-t_{i-1}} \right) - \right.$ $\left. - \psi(\theta_1) \int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} dt \right\}^2 \bigg/ \int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} dt$	$\psi'(\theta_1) - \frac{1}{\theta_1}$	—	—
	θ_0, θ_1	$\frac{2}{\Gamma(\theta_1)} \sum_{i=1}^k \left\{ \int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} \ln t dt + \frac{1}{\theta_1} \left(t_i^{\theta_1} e^{-t_i} - t_{i-1}^{\theta_1} e^{-t_{i-1}} \right) - \right.$ $\left. - \psi(\theta_1) \int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} dt \right\} \left(t_{i-1}^{\theta_1} e^{-t_{i-1}} - t_i^{\theta_1} e^{-t_i} \right) \bigg/ \int_{t_{i-1}}^{t_i} t^{\theta_1-1} e^{-t} dt$	0	—	—
19. Гамма-распределение $\frac{\theta_1^{\theta_0}}{\Gamma(\theta_0)} (x - \theta_2)^{\theta_0-1} \times$ $\times e^{-\theta_1(x-\theta_2)}$	θ_0	$\frac{1}{\Gamma(\theta_0)} \sum_{i=1}^k \left\{ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} \ln t dt - \right.$	$\psi'(\theta_0)$	A.53	A.54

$x \in (\theta_2, \infty)$ $t_i = \theta_1 (x_i - \theta_2)$		$\left. -\psi(\theta_0) \int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} dt \right\}^2 \bigg/ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} dt$			
	θ_1	$\frac{1}{\theta_1^2 \Gamma(\theta_0)} \sum_{i=1}^k \frac{\left[t_i^{\theta_0} e^{-t_i} - t_{i-1}^{\theta_0} e^{-t_{i-1}} \right]^2}{\int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} dt}$	$\frac{\theta_0}{\theta_1^2}$	A.55	A.56
	θ_0, θ_1	$\frac{1}{\theta_1 \Gamma(\theta_0)} \sum_{i=1}^k \left\{ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} \ln t dt - \right.$ $\left. -\psi(\theta_0) \int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} dt \right\} \left(t_i^{\theta_0} e^{-t_i} - t_{i-1}^{\theta_0} e^{-t_{i-1}} \right) \bigg/ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} e^{-t} dt$	$-\frac{1}{\theta_1}$	A.57	A.58
20. Бета-распределение первого рода $\frac{1}{\theta_3 B(\theta_0, \theta_1)} \times$ $\times \left(\frac{x - \theta_2}{\theta_3} \right)^{\theta_0-1} \times$ $\times \left(1 - \frac{x - \theta_2}{\theta_3} \right)^{\theta_1-1}$	θ_0	$\frac{1}{B(\theta_0, \theta_1)} \sum_{i=1}^k \left\{ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt - \right.$ $\left. -[\psi(\theta_0) - \psi(\theta_0 + \theta_1)] \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt \right\}^2 \bigg/$ $\int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt$	$\psi'(\theta_0) -$ $-\psi'(\theta_0 + \theta_1)$	—	—

Функция плотности распределения	Параметр распределения	Количество информации Фишера для параметра по наблюдениям		Номер таблицы АОГ	
		группированным	негруппированным	t_i	P_i
$x \in [\theta_2, \theta_2 + \theta_3]$ $B(\theta_0, \theta_1) =$ $= \Gamma(\theta_0)\Gamma(\theta_1) / \Gamma(\theta_0 + \theta_1)$ $t_i = (x_i - \theta_2) / \theta_3$	θ_1	$\frac{1}{B(\theta_0, \theta_1)} \sum_{i=1}^k \left\{ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln(1-t) dt - \right.$ $\left. - [\psi(\theta_1) - \psi(\theta_0 + \theta_1)] \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt \right\}^2 /$ $/ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt$	$\psi'(\theta_1) - \psi'(\theta_0 + \theta_1)$	—	—
	θ_0, θ_1	$\frac{1}{B(\theta_0, \theta_1)} \sum_{i=1}^k \left\{ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt - \right.$ $\left. - [\psi(\theta_0) - \psi(\theta_0 + \theta_1)] \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt \right\} \times$ $\times \left\{ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln(1-t) dt - \right.$	$-\psi'(\theta_0 + \theta_1)$	—	—

		$\left. -[\psi(\theta_1) - \psi(\theta_0 + \theta_1)] \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt \right\} /$ $/ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt$			
21. Стандартное бета-распределение второго рода $\frac{1}{B(\theta_0, \theta_1)} \frac{(x - \theta_2)^{\theta_0-1}}{(1 + x - \theta_2)^{\theta_0 + \theta_1}}$ $x \in [\theta_2, \infty)$ $t_i = \frac{x_i - \theta_2}{1 + x_i - \theta_2}$	θ_0	См. соответствующие соотношения для бета-распределения первого рода	$\psi'(\theta_0) -$ $-\psi'(\theta_0 + \theta_1)$	—	—
	θ_1	См. соответствующие соотношения для бета-распределения первого рода	$\psi'(\theta_1) -$ $-\psi'(\theta_0 + \theta_1)$	—	—
	θ_0, θ_1	См. соответствующие соотношения для бета-распределения первого рода	$-\psi'(\theta_0 + \theta_1)$	—	—
22. Бета-распределение второго рода $\frac{\theta_2}{B(\theta_0, \theta_1)} \times$ $\times \frac{[\theta_2(x - \theta_3)]^{\theta_0-1}}{[1 + \theta_2(x - \theta_3)]^{\theta_0 + \theta_1}}$ $x \in [\theta_3, \infty)$	θ_0	См. соответствующие соотношения для бета-распределения первого рода	—»—	—	—
	θ_1	См. соответствующие соотношения для бета-распределения первого рода	—»—	—	—

Функция плотности распределения	Пара- метр распре- деления	Количество информации Фишера для параметра по наблюдениям		Номер таб- лицы АОГ	
		группированным	негруппи- рованным	t_i	P_i
$t_i = \frac{\theta_2 (x_i - \theta_3)}{1 + \theta_2 (x_i - \theta_3)}$	θ_0, θ_1	См. соответствующие соотношения для бета- распределения первого рода	—»—	—	—
	θ_2	$\frac{1}{\theta_2^2 B(\theta_0, \theta_1)} \sum_{i=1}^k \left\{ \theta_0 \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt - \right. \\ \left. - (\theta_0 + \theta_1) \int_{t_{i-1}}^{t_i} t^{\theta_0} (1-t)^{\theta_1-1} dt \right\}^2 /$ $/ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt$	$\frac{\theta_0 \theta_1}{\theta_2^2 (\theta_0 + \theta_1 + 1)}$	—	—
	θ_0, θ_2	$\frac{1}{\theta_2 B(\theta_0, \theta_1)} \sum_{i=1}^k \left\{ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln t dt / \right. \\ \left. / \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt - (\psi(\theta_0) - \psi(\theta_0 + \theta_1)) \right\} \times$	$- \frac{\theta_1}{\theta_2 (\theta_0 + \theta_1)}$	—	—

			$\times \left(\theta_0 \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt - (\theta_0 + \theta_1) \int_{t_{i-1}}^{t_i} t^{\theta_0} (1-t)^{\theta_1-1} dt \right)$			
	θ_1, θ_2		$\frac{1}{\theta_2 B(\theta_0, \theta_1)} \sum_{i=1}^k \left\{ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} \ln(1-t) dt \right. /$ $/ \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt - (\psi(\theta_1) - \psi(\theta_0 + \theta_1)) \Big\} \times$ $\times \left(\theta_0 \int_{t_{i-1}}^{t_i} t^{\theta_0-1} (1-t)^{\theta_1-1} dt - (\theta_0 + \theta_1) \int_{t_{i-1}}^{t_i} t^{\theta_0} (1-t)^{\theta_1-1} dt \right)$	$\frac{\theta_0}{\theta_2 (\theta_0 + \theta_1)}$	—	—
23. Распределение Sb-Джонсона $\frac{\theta_1 \theta_2}{\sqrt{2\pi}(x - \theta_3)(\theta_2 + \theta_3 - x)} \times$ $\times \exp \left\{ -\frac{1}{2} \left[\theta_0 + \theta_1 \ln \frac{x - \theta_3}{\theta_2 + \theta_3 - x} \right]^2 \right\}$ $x \in [\theta_3, \theta_2 + \theta_3]$		θ_0	$\sum_{i=1}^k \frac{[\varphi(t_i) - \varphi(t_{i-1})]^2}{\Phi(t_i) - \Phi(t_{i-1})}$	1	A.24	A.25
$t_i = \theta_0 + \theta_1 \ln \frac{x_i - \theta_3}{\theta_2 + \theta_3 - x_i}$	θ_1		$\frac{1}{\theta_1^2} \sum_{i=1}^k \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]^2}{\Phi(t_i) - \Phi(t_{i-1})}$	$\frac{2 + \theta_0^2}{\theta_1^2}$	A.26	A.27

Функция плотности распределения	Пара- метр распре- деления	Количество информации Фишера для параметра по наблюдениям		Номер таб- лицы АОГ	
		группированным	негруппи- рованным	t_i	P_i
	θ_0, θ_1	$\frac{1}{\theta_1} \sum_{i=1}^k \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})][\varphi(t_i) - \varphi(t_{i-1})]}{\Phi(t_i) - \Phi(t_{i-1})}$	$-\frac{\theta_0}{\theta_1}$	A.28	A.29
24. Распределение SI-Джонсона $\frac{\theta_1}{\sqrt{2\pi}(x - \theta_3)} \times$ $\times \exp \left\{ -\frac{1}{2} \left[\theta_0 + \theta_1 \ln \frac{x - \theta_3}{\theta_2} \right]^2 \right\}$ $x \in [\theta_3, \infty)$ $t_i = \theta_0 + \theta_1 \ln \frac{x_i - \theta_3}{\theta_2}$		θ_0	$\sum_{i=1}^k \frac{[\varphi(t_i) - \varphi(t_{i-1})]^2}{\Phi(t_i) - \Phi(t_{i-1})}$	1	A.24 A.25
	θ_1	$\frac{1}{\theta_1^2} \sum_{i=1}^k \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]^2}{\Phi(t_i) - \Phi(t_{i-1})}$	$\frac{2 + \theta_0^2}{\theta_1^2}$	A.26	A.27
	θ_0, θ_1	$\frac{1}{\theta_1} \sum_{i=1}^k \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})][\varphi(t_i) - \varphi(t_{i-1})]}{\Phi(t_i) - \Phi(t_{i-1})}$	$-\frac{\theta_0}{\theta_1}$	A.28	A.29
	θ_2	$\frac{\theta_1^2}{\theta_2^2} \sum_{i=1}^k \frac{[\varphi(t_{i-1}) - \varphi(t_i)]^2}{\Phi(t_i) - \Phi(t_{i-1})}$	$\frac{\theta_1^2}{\theta_2^2}$	A.24	A.25

	θ_0, θ_2	$\frac{\theta_1}{\theta_2} \sum_{i=1}^k \frac{[\varphi(t_i) - \varphi(t_{i-1})][\varphi(t_{i-1}) - \varphi(t_i)]}{\Phi(t_i) - \Phi(t_{i-1})}$	$-\frac{\theta_1}{\theta_2}$	—	—
	θ_1, θ_2	$\frac{1}{\theta_2} \sum_{i=1}^k \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})][\varphi(t_{i-1}) - \varphi(t_i)]}{\Phi(t_i) - \Phi(t_{i-1})}$	$\frac{\theta_0}{\theta_2}$	—	—
25. Распределение Су-Джонсона	$\frac{\theta_1}{\sqrt{2\pi}\sqrt{(x-\theta_3)^2 + \theta_2^2}} \times$ $\times \exp \left\{ -\frac{1}{2} \left[\theta_0 + \theta_1 \ln \left\{ \frac{x-\theta_3}{\theta_2} + \sqrt{\left(\frac{x-\theta_3}{\theta_2} \right)^2 + 1} \right\} \right]^2 \right\}$ $x \in (-\infty, \infty)$ $t_i = \theta_0 + \theta_1 \ln \left\{ s_i + \sqrt{s_i^2 + 1} \right\}$ $s_i = (x_i - \theta_3) / \theta_2$				
	θ_0	$\sum_{i=1}^k \frac{[\varphi(t_i) - \varphi(t_{i-1})]^2}{\Phi(t_i) - \Phi(t_{i-1})}$	1	A.24	A.25
	θ_1	$\frac{1}{\theta_1^2} \sum_{i=1}^k \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]^2}{\Phi(t_i) - \Phi(t_{i-1})}$	$\frac{2 + \theta_0^2}{\theta_1^2}$	—	—
	θ_0, θ_1	$\frac{1}{\theta_1} \sum_{i=1}^k \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})][\varphi(t_i) - \varphi(t_{i-1})]}{\Phi(t_i) - \Phi(t_{i-1})}$	$-\frac{\theta_0}{\theta_1}$	A.28	A.29

Функция плотности распределения	Пара- метр распре- деления	Количество информации Фишера для параметра по наблюдениям		Номер таб- лицы АОГ	
		группированным	негруппированным	t_i	P_i
	θ_2	$\frac{\theta_1^2}{\theta_2^2} \sum_{i=1}^k \frac{\left(\frac{s_{i-1}\varphi(t_{i-1})}{\sqrt{s_{i-1}^2+1}} - \frac{s_i\varphi(t_i)}{\sqrt{s_i^2+1}} \right)^2}{\Phi(t_i) - \Phi(t_{i-1})}$	$\frac{1+\theta_1^2}{\theta_2^2} - \frac{\theta_1^3}{\theta_2^2} \int_{-\infty}^{\infty} \frac{\varphi(t)}{(s^2+1)^{3/2}} ds$	—	—
	θ_0, θ_2	$\frac{\theta_1}{\theta_2} \sum_{i=1}^k \frac{[\varphi(t_i) - \varphi(t_{i-1})] \left[\frac{\varphi(t_{i-1})s_{i-1}}{\sqrt{s_{i-1}^2+1}} - \frac{\varphi(t_i)s_i}{\sqrt{s_i^2+1}} \right]}{\Phi(t_i) - \Phi(t_{i-1})}$	$-\theta_1 \int_{-\infty}^{\infty} \frac{\varphi(t)}{(s^2+1)^{3/2}} ds \Big\}$	—	—
	θ_1, θ_2	$\frac{1}{\theta_2} \sum_{i=1}^k \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]}{\Phi(t_i) - \Phi(t_{i-1})} \times$ $\times \left[\frac{\varphi(t_{i-1})s_{i-1}}{\sqrt{s_{i-1}^2+1}} - \frac{\varphi(t_i)s_i}{\sqrt{s_i^2+1}} \right] \Big\}$	$\theta_0\theta_1 \int_{-\infty}^{\infty} \frac{s\varphi(t)}{s^2+1} ds -$ $-2 \int_{-\infty}^{\infty} \frac{s\varphi(t)}{(s^2+1)^{3/2}} ds$	—	—

		θ_3	$\frac{\theta_1^2}{\theta_2^2} \sum_{i=1}^k \frac{\left(\frac{\varphi(t_{i-1})}{\sqrt{s_{i-1}^2+1}} - \frac{\varphi(t_i)}{\sqrt{s_i^2+1}} \right)^2}{\Phi(t_i) - \Phi(t_{i-1})}$	$\frac{\theta_1^3}{\theta_2^2} \int_{-\infty}^{\infty} \frac{\varphi(t)}{(s^2+1)^{3/2}} ds$	—	—
		θ_0, θ_3	$\frac{\theta_1}{\theta_2} \sum_{i=1}^k \frac{[\varphi(t_i) - \varphi(t_{i-1})] \left[\frac{\varphi(t_{i-1})}{\sqrt{s_{i-1}^2+1}} - \frac{\varphi(t_i)}{\sqrt{s_i^2+1}} \right]}{\Phi(t_i) - \Phi(t_{i-1})}$	$-\frac{\theta_1^2}{\theta_2} \int_{-\infty}^{\infty} \frac{\varphi(t)}{s^2+1} ds$	—	—
		θ_1, θ_3	$\frac{1}{\theta_2} \sum_{i=1}^k \frac{[(t_i - \theta_0)\varphi(t_i) - (t_{i-1} - \theta_0)\varphi(t_{i-1})]}{\Phi(t_i) - \Phi(t_{i-1})} \times$ $\times \left[\frac{\varphi(t_{i-1})}{\sqrt{s_{i-1}^2+1}} - \frac{\varphi(t_i)}{\sqrt{s_i^2+1}} \right]$	$\frac{\theta_0 \theta_1}{\theta_2} \int_{-\infty}^{\infty} \frac{\varphi(t)}{s^2+1} ds$	—	—
	θ_2, θ_3		$\frac{\theta_1^2}{\theta_2^2} \sum_{i=1}^k \frac{\left[\frac{\varphi(t_{i-1})s_{i-1}}{\sqrt{s_{i-1}^2+1}} - \frac{\varphi(t_i)s_i}{\sqrt{s_i^2+1}} \right] \left[\frac{\varphi(t_{i-1})}{\sqrt{s_{i-1}^2+1}} - \frac{\varphi(t_i)}{\sqrt{s_i^2+1}} \right]}{\Phi(t_i) - \Phi(t_{i-1})}$	$\frac{\theta_1^3}{\theta_2^2} \int_{-\infty}^{\infty} \frac{s\varphi(t)}{(s^2+1)^{3/2}} ds +$ $+ \frac{5\theta_1}{\theta_2^2} \int_{-\infty}^{\infty} \frac{s\varphi(t)}{(s^2+1)^{5/2}} ds$	—	—

Функция плотности распределения	Параметр распределения	Количество информации Фишера для параметра по наблюдениям		Номер таблицы АОГ	
		группированным	негруппированным	t_i	P_i
26. Класс экспоненциальных распределений $\frac{\theta_2}{2\theta_1\Gamma(1/\theta_2)} \times \exp\left\{-\left(\frac{ x-\theta_0 }{\theta_1}\right)^{\theta_2}\right\}$ $x \in (-\infty, \infty)$ $t_i = \frac{x_i - \theta_0}{\theta_1}$ $\gamma_i = \int_0^{ t_i ^{\theta_2}} y^{1/\theta_2-1} e^{-y} dy$	θ_0	$\frac{\theta_2^2}{\theta_1^2\Gamma(1/\theta_2)} \sum_{i=1}^k \frac{2\left(e^{- t_{i-1} ^{\theta_2}} - e^{- t_i ^{\theta_2}}\right)^2}{\gamma_i \operatorname{sign} t_i - \gamma_{i-1} \operatorname{sign} t_{i-1}}$	$\frac{\theta_2^2\Gamma\left(2-\frac{1}{\theta_2}\right)}{\theta_1^2\Gamma(1/\theta_2)}$	—	—
	θ_1	$\frac{\theta_2^2}{2\theta_1^2\Gamma(1/\theta_2)} \sum_{i=1}^k \frac{\left(t_{i-1}e^{- t_{i-1} ^{\theta_2}} - t_i e^{- t_i ^{\theta_2}}\right)^2}{\gamma_i \operatorname{sign} t_i - \gamma_{i-1} \operatorname{sign} t_{i-1}}$	$\frac{\theta_2}{\theta_1^2}$	—	—
	θ_0, θ_1	$\frac{\theta_2^2}{2\theta_1^2\Gamma(1/\theta_2)} \times \sum_{i=1}^k \frac{\left(e^{- t_{i-1} ^{\theta_2}} - e^{- t_i ^{\theta_2}}\right)\left(t_{i-1}e^{- t_{i-1} ^{\theta_2}} - t_i e^{- t_i ^{\theta_2}}\right)}{\gamma_i \operatorname{sign} t_i - \gamma_{i-1} \operatorname{sign} t_{i-1}}$	$-\frac{\theta_2^2}{\theta_1^2\Gamma\left(-\frac{\theta_2-1}{\theta_2}\right)}$	—	—

$\gamma'_i = \int_0^{ t_i ^{\theta_2}} y^{1/\theta_2-1} e^{-y} \ln y dy$	θ_2	$\frac{1}{2\Gamma(1/\theta_2)} \sum_{i=1}^k \left\{ \frac{1}{\theta_2^2} \left[\gamma'_i \operatorname{sign} t_i - \gamma'_{i-1} \operatorname{sign} t_{i-1} + \right. \right. \\ \left. \left. + \psi(1/\theta_2) \left(\gamma_i \operatorname{sign} t_i - \gamma_{i-1} \operatorname{sign} t_{i-1} \right) \right] + \right. \\ \left. + t_i e^{- t_i ^{\theta_2}} \ln t_i - t_{i-1} e^{- t_{i-1} ^{\theta_2}} \ln t_{i-1} \right\}^2 \Bigg/ \\ \Bigg/ \left(\gamma_i \operatorname{sign} t_i - \gamma_i \operatorname{sign} t_{i-1} \right)$	$\frac{(1+\theta_2)\psi'(1/\theta_2)}{\theta_2^4} + \\ + \frac{\psi(1/\theta_2) + 2\theta_2 + 2}{\theta_2^3} \times \\ \times \psi(1/\theta_2) + \frac{1}{\theta_2^3}$	$-$	$-$
	θ_0, θ_2	$\frac{\theta_2}{2\theta_1\Gamma(1/\theta_2)} \sum_{i=1}^k \left(e^{- t_{i-1} ^{\theta_2}} - e^{- t_i ^{\theta_2}} \right) \times \\ \times \left\{ \frac{1}{\theta_2^2} \left[\gamma'_i \operatorname{sign} t_i - \gamma'_{i-1} \operatorname{sign} t_{i-1} + \right. \right. \\ \left. \left. + \psi(1/\theta_2) \left(\gamma_i \operatorname{sign} t_i - \gamma_{i-1} \operatorname{sign} t_{i-1} \right) \right] + \right. \\ \left. + t_i e^{- t_i ^{\theta_2}} \ln t_i - t_{i-1} e^{- t_{i-1} ^{\theta_2}} \ln t_{i-1} \right\} \Bigg/ \\ \Bigg \Bigg/ \left(\gamma_i \operatorname{sign} t_i - \gamma_i \operatorname{sign} t_{i-1} \right)$	$\frac{\psi(1/\theta_2) + \theta_2 C}{\theta_1 \theta_2 \Gamma(1/\theta_2)}$	$-$	$-$

Функция плотности распределения	Пара- метр распре- деления	Количество информации Фишера для параметра по наблюдениям		Номер таб- лицы АОГ	
		группированным	негруппиро- ванным	t_i	P_i
	θ_1, θ_2	$\frac{\theta_2}{2\theta_1\Gamma(1/\theta_2)} \sum_{i=1}^k \left(t_{i-1} e^{- t_{i-1} ^{\theta_2}} - t_i e^{- t_i ^{\theta_2}} \right) \times$ $\times \left\{ \frac{1}{\theta_2^2} \left[\gamma'_i \operatorname{sign} t_i - \gamma'_{i-1} \operatorname{sign} t_{i-1} + \right. \right.$ $\left. - \psi(1/\theta_2) (\gamma_i \operatorname{sign} t_i - \gamma_{i-1} \operatorname{sign} t_{i-1}) \right] +$ $\left. + t_i e^{- t_i ^{\theta_2}} \ln t_i - t_{i-1} e^{- t_{i-1} ^{\theta_2}} \ln t_{i-1} \right\} /$ $\left \left(\gamma_i \operatorname{sign} t_i - \gamma_i \operatorname{sign} t_{i-1} \right) \right $	$-\frac{1 + \theta_2 + \psi(1/\theta_2)}{\theta_1 \theta_2}$	—	—

2.10. Примеры использования таблиц асимптотически оптимального группирования

Ниже на примерах рассмотрено применение асимптотически оптимального группирования в случае простых и сложных гипотез.

Пример 2.1. Проверяют простую гипотезу о принадлежности выборки экспоненциальному закону распределения. Упорядоченная выборка объемом 200 наблюдений имеет вид

0.0005	0.0095	0.0189	0.0200	0.0225	0.0261	0.0339	0.0343	0.0463	0.0509
0.0625	0.0748	0.0797	0.0802	0.0857	0.0887	0.1049	0.1065	0.1276	0.1395
0.1465	0.1539	0.1565	0.1582	0.1584	0.1648	0.1679	0.1681	0.1689	0.1818
0.1940	0.1944	0.1950	0.1976	0.1992	0.2064	0.2260	0.2277	0.2311	0.2433
0.2441	0.2442	0.2509	0.2522	0.2566	0.2683	0.2840	0.3006	0.3025	0.3052
0.3056	0.3098	0.3158	0.3286	0.3288	0.3297	0.3396	0.3490	0.3578	0.3650
0.3728	0.3755	0.3823	0.3834	0.4017	0.4093	0.4135	0.4259	0.4419	0.4486
0.4487	0.4503	0.4587	0.4699	0.4711	0.4802	0.4814	0.4818	0.4981	0.4984
0.4989	0.5005	0.5085	0.5191	0.5279	0.5301	0.5328	0.5331	0.5389	0.5392
0.5432	0.5455	0.5548	0.5801	0.6001	0.6009	0.6694	0.6803	0.6817	0.6838
0.6844	0.6995	0.7127	0.7358	0.7515	0.7566	0.7699	0.7953	0.8048	0.8072
0.8116	0.8232	0.8271	0.8323	0.8771	0.8908	0.8949	0.9272	0.9571	0.9593
0.9682	0.9708	0.9761	0.9881	0.9931	1.0110	1.0188	1.0364	1.0447	1.0612
1.0683	1.0909	1.0984	1.1030	1.1068	1.1076	1.1100	1.1493	1.1562	1.1634
1.1702	1.1866	1.1964	1.2006	1.2103	1.2325	1.2607	1.2945	1.3185	1.3437
1.3918	1.4287	1.4398	1.4505	1.5068	1.5126	1.5491	1.5501	1.5681	1.5933
1.6323	1.6446	1.6686	1.6782	1.6816	1.6839	1.7054	1.7335	1.7560	1.7629
1.7677	1.7681	1.9750	1.9840	1.9873	2.0031	2.0870	2.0877	2.0957	2.1226
2.1382	2.1925	2.1950	2.2029	2.2315	2.2318	2.3487	2.4474	2.5498	2.8392
2.9899	3.0846	3.1179	3.3316	3.5926	3.5937	3.6218	4.0604	4.8403	6.0534

Проверяемая гипотеза имеет вид

$$H_0 : f(x, \theta) = \frac{1}{\theta_0} e^{-(x-\theta_1)/\theta_0}$$

при $\theta_0 = 1$, $\theta_1 = 0$.

Оптимальные граничные точки в виде $t_i = (x_i - \theta_1) / \theta_0$ берут из табл. А.1. Выбирают число интервалов $k = 5$, так как $P_k n = 0.0189 \cdot 200 > 3$, а при большем числе интервалов ожидаемые частоты попадания в крайний интервал будут еще меньше. При $k = 5$ значения t_i соответственно: 0.6004; 1.3545; 2.3720; 3.9657. При $\theta_0 = 1$ и $\theta_1 = 0$ значения x_i совпадают со значениями t_i . Соответствующие такому разбиению вероятности попадания наблюдений в интервалы берут из табл. А.2: 0.4514; 0.2905; 0.1648; 0.0744; 0.0189. По выборке находят числа наблюдений n_i , попавших в интервалы: 95, 55, 37, 10, 3. В соответствии с формулой (2.1) вычисляют значение статистики $X_n^2 = 2.6672$. При справедливости H_0 статистика подчиняется χ_4^2 -распределению и при значениях уровня значимости

$$\alpha < P\{X_n^2 > 2.6672\} = \frac{1}{2^{r/2} \Gamma(r/2)} \int_{2.8694}^{\infty} s^{r/2-1} e^{-s/2} ds = 0.6150$$

гипотеза H_0 будет принята. Вычисляемое в соответствии с (2.3) значение статистики отношения правдоподобия $S_{\text{ОП}} = 2.8694$; $P\{S_{\text{ОП}} > 2.8694\} = 0.5799$. И таким образом, в соответствии с данным критерием также нет оснований для отклонения гипотезы H_0 . На рис. 2.22 представлены плотность, гистограмма и результаты проверки согласия.

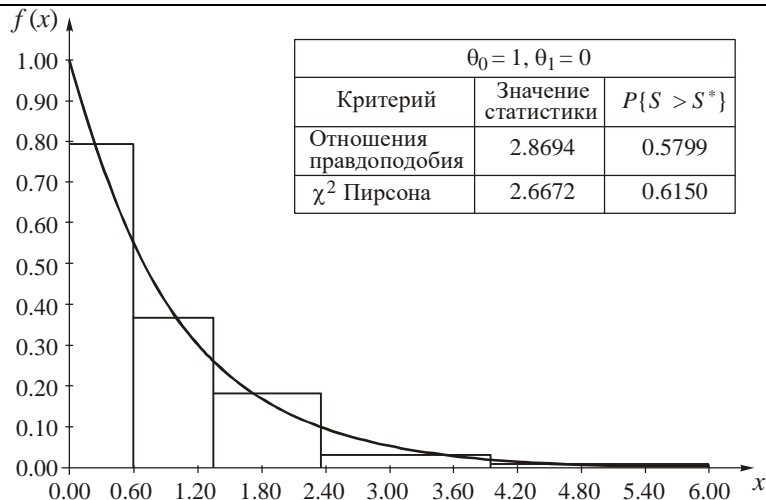


Рис. 2.22. Результаты проверки согласия к примеру 2.1

Пример 2.2. Проверяют сложную гипотезу о принадлежности выборки экспоненциальному закону распределения. Упорядоченная выборка объемом 300 наблюдений имеет вид

0.0204	0.0242	0.0336	0.0342	0.0410	0.0511	0.0539	0.0859	0.0888	0.0925
0.0970	0.0996	0.1006	0.1039	0.1065	0.1067	0.1158	0.1170	0.1185	0.1188
0.1206	0.1207	0.1215	0.1260	0.1336	0.1336	0.1389	0.1462	0.1518	0.1565
0.1641	0.1664	0.1677	0.1738	0.1785	0.1812	0.1823	0.1889	0.2002	0.2032
0.2042	0.2074	0.2143	0.2221	0.2223	0.2265	0.2299	0.2335	0.2351	0.2408
0.2413	0.2424	0.2566	0.2576	0.2676	0.2697	0.2745	0.2749	0.2766	0.2780
0.2785	0.2826	0.2839	0.2909	0.2936	0.3029	0.3097	0.3251	0.3280	0.3352
0.3593	0.3597	0.3675	0.3688	0.3751	0.3816	0.3831	0.3875	0.3934	0.4049

0.4057	0.4057	0.4127	0.4133	0.4146	0.4212	0.4225	0.4231	0.4236	0.4363
0.4381	0.4414	0.4493	0.4553	0.4649	0.4712	0.4733	0.4798	0.4816	0.4817
0.4837	0.4880	0.4890	0.4917	0.4927	0.4982	0.5044	0.5081	0.5092	0.5135
0.5245	0.5301	0.5350	0.5351	0.5430	0.5432	0.5534	0.5609	0.5636	0.5637
0.5747	0.5806	0.5891	0.5898	0.5926	0.5962	0.6099	0.6137	0.6423	0.6670
0.6696	0.6713	0.6759	0.6885	0.6910	0.6913	0.6990	0.7043	0.7089	0.7150
0.7180	0.7189	0.7197	0.7226	0.7343	0.7530	0.7627	0.7635	0.7639	0.7822
0.7883	0.7889	0.7930	0.7943	0.7972	0.7979	0.8190	0.8191	0.8214	0.8234
0.8289	0.8291	0.8327	0.8382	0.8400	0.8408	0.8466	0.8571	0.8655	0.8679
0.8770	0.8895	0.8898	0.8925	0.8971	0.9041	0.9106	0.9134	0.9216	0.9234
0.9279	0.9395	0.9430	0.9458	0.9550	0.9625	0.9693	0.9792	0.9855	0.9877
1.0074	1.0127	1.0130	1.0172	1.0225	1.0227	1.0254	1.0520	1.0537	1.0720
1.0761	1.1009	1.1041	1.1071	1.1183	1.1586	1.1630	1.1719	1.1766	1.1909
1.1980	1.1986	1.2061	1.2169	1.2240	1.2298	1.2633	1.2698	1.2722	1.2749
1.2990	1.3150	1.3262	1.3345	1.3366	1.3369	1.3409	1.3423	1.3456	1.3501
1.3507	1.3518	1.3765	1.3791	1.3836	1.4011	1.4236	1.4339	1.4583	1.4630
1.4872	1.4935	1.5021	1.5291	1.5326	1.5347	1.5360	1.5503	1.5638	1.5642
1.5727	1.5934	1.6162	1.6165	1.6488	1.6543	1.6668	1.6916	1.7138	1.7300
1.7316	1.7339	1.7905	1.8173	1.8480	1.8725	1.8863	1.9178	1.9426	1.9720
1.9804	2.0209	2.0388	2.0816	2.1159	2.1169	2.1566	2.1732	2.1754	2.2319
2.2391	2.2620	2.3139	2.3222	2.3830	2.4445	2.4560	2.4864	2.5580	2.5796
2.7214	2.7843	3.0117	3.1078	3.1312	3.1805	3.2445	3.4491	3.6123	4.9320

Проверяемая гипотеза

$$H_0 : f(x) \in \left\{ f(x, \theta) = \frac{1}{\theta_0} e^{-(x-\theta_1)/\theta_0}, \theta_0 \in (0, \infty), \theta_1 = 0 \right\}.$$

При этом θ_0 оценивают по выборке.

Оценка максимального правдоподобия $\hat{\theta}_0 = 0.937338$. Оптимальные граничные точки в виде $t_i = (x_i - \theta_1) / \theta_0$, как и в предыдущем случае, берут из табл. А.1 при $k=5$: 0.6004; 1.3545; 2.3720; 3.9657. При $\theta_0 = 0.937338$ и $\theta_1 = 0$ вычисляют значения $x_i = t_i \theta_0 + \theta_1$: 0.5628; 1.2696; 2.2234; 3.7172. Вероятности попадания наблюдений в интервалы из табл. А.2: 0.4514; 0.2905; 0.1648; 0.0744; 0.0189. По выборке находят числа наблюдений n_i , попавших в соответствующие интервалы (118, 99, 62, 20, 1). В соответствии с соотношением (2.1) вычисляют значение статистики $X_n^2 = 11.145$. При справедливости гипотезы H_0 статистика приближенно подчиняется χ_3^2 -распределению. Вероятность

91

$$P\{X_n^2 > 11,145\} = \frac{1}{2^{r/2} \Gamma(r/2)} \int_{11,145}^{\infty} s^{r/2-1} e^{-s/2} ds = 0.010967.$$

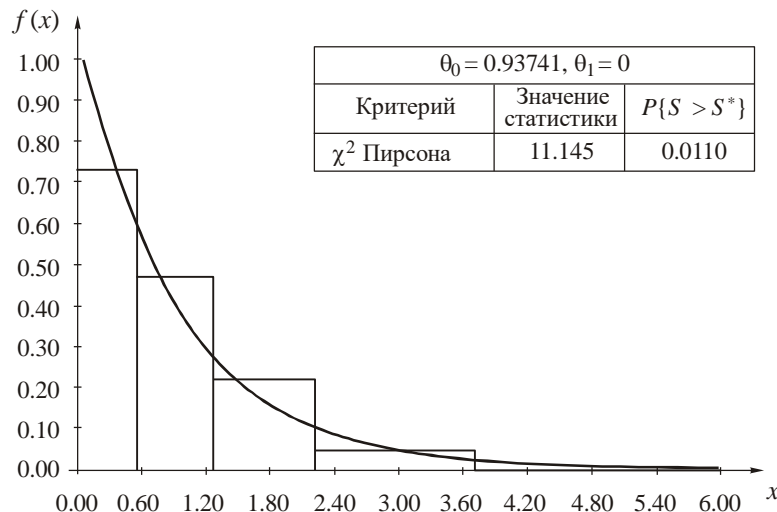


Рис. 2.23. Результаты проверки согласия к примеру 2.2

Уже при значениях $\alpha = 0.05$ нет оснований для принятия гипотезы H_0 , но есть основание проверить и другие распределения. Результаты проверки гипотезы о согласии иллюстрирует рис. 2.23.

Пример 2.3. Проверяют простую гипотезу о принадлежности выборки нормальному закону распределения. Упорядоченная выборка объемом 300 наблюдений имеет вид

-5.1852 -3.7318 -3.2899 -3.1038 -2.9031 -2.7243 -2.7011 -2.6448 -2.5168 -2.4975
 -2.4858 -2.4268 -2.1459 -1.9894 -1.9467 -1.9348 -1.9273 -1.8873 -1.8383 -1.8217
 -1.6802 -1.6713 -1.6704 -1.6420 -1.6322 -1.6315 -1.6005 -1.5491 -1.5238 -1.5204
 -1.4727 -1.4546 -1.3420 -1.3233 -1.3225 -1.3150 -1.3097 -1.2917 -1.2767 -1.2046

-1.1606	-1.1334	-1.1259	-1.0769	-1.0659	-1.0446	-1.0258	-1.0138	-1.0077	-0.9914
-0.9818	-0.9546	-0.9511	-0.9167	-0.8927	-0.8758	-0.8311	-0.8302	-0.8171	-0.7794
-0.7571	-0.7228	-0.7211	-0.7064	-0.6991	-0.6873	-0.6743	-0.6442	-0.6440	-0.5798
-0.5421	-0.5308	-0.5108	-0.4845	-0.4491	-0.4294	-0.4212	-0.3943	-0.3251	-0.2607
-0.2506	-0.2351	-0.2201	-0.1698	-0.1326	-0.1217	-0.0368	-0.0194	0.0014	0.0108
0.0470	0.0748	0.0822	0.1281	0.1284	0.1455	0.1498	0.1539	0.1564	0.1895
0.1949	0.2079	0.2402	0.2643	0.2669	0.3104	0.3616	0.3723	0.4415	0.4929
0.5315	0.5318	0.5350	0.5359	0.5396	0.5843	0.5877	0.5944	0.6205	0.6279
0.6345	0.6391	0.6416	0.6445	0.6592	0.6630	0.6990	0.7361	0.7419	0.7426
0.7472	0.7511	0.7658	0.8359	0.8499	0.8681	0.9283	0.9368	0.9504	0.9657
0.9782	1.0158	1.0381	1.0505	1.0702	1.0873	1.0913	1.1313	1.1750	1.1908
1.2197	1.2347	1.2620	1.2799	1.3040	1.3160	1.3186	1.3245	1.3529	1.3729
1.3857	1.3979	1.4047	1.4101	1.4194	1.4485	1.4515	1.4782	1.4850	1.5759
1.6056	1.6076	1.6220	1.6223	1.6311	1.6395	1.6476	1.6610	1.6793	1.6800
1.6908	1.7208	1.7356	1.7358	1.7359	1.7550	1.8296	1.8573	1.8757	1.9448
1.9451	1.9924	2.0045	2.0153	2.0338	2.0400	2.0557	2.1055	2.1161	2.1224
2.1237	2.2229	2.2289	2.2432	2.2600	2.2989	2.3003	2.3088	2.3315	2.3659
2.4024	2.4152	2.4187	2.4600	2.4914	2.5100	2.5145	2.5311	2.5515	2.5706
2.5916	2.6982	2.7127	2.7282	2.7553	2.7778	2.8131	2.8296	2.8364	2.8445
2.8552	2.9012	2.9126	2.9517	3.0283	3.0510	3.1031	3.1199	3.1451	3.1453
3.1539	3.1633	3.2561	3.2995	3.3040	3.3056	3.3152	3.3159	3.3208	3.3567
3.4532	3.5665	3.5775	3.5782	3.5858	3.6054	3.6368	3.6435	3.6630	3.6788
3.6798	3.6985	3.6993	3.7203	3.7421	3.8047	3.8366	3.8389	3.9597	3.9682
3.9784	3.9828	3.9894	4.1512	4.1674	4.1887	4.2089	4.4316	4.4451	4.5262
4.5282	4.5397	4.5459	4.5898	4.6170	4.6560	4.6902	4.7263	4.7912	5.0490
5.1871	5.2678	5.4110	5.5609	5.6245	5.6484	5.6736	5.7743	6.1626	6.2185

Проверяемая гипотеза имеет вид

$$H_0 : f(x, \theta) = \frac{1}{\theta_1 \sqrt{2\pi}} e^{-(x-\theta_0)^2 / 2\theta_1^2}$$

при $\theta_0 = 1$, $\theta_1 = 2$.

Оптимальные граничные точки в виде $t_i = (x_i - \theta_0) / \theta_1$ берут из табл. А.28 при $k = 7$ (−2.0600; −1.2647; −0.4918; 0.4918; 1.2647; 2.0600), так как из табл. А.29 $P_k n = 0,0197 \cdot 300 > 5$. При $\theta_0 = 1$, $\theta_1 = 2$ x_i принимают значения (−3.12; −1.5294; 0.0164; 1.9836; 3.5294; 5.12). Вероятности попадания наблюдений в интервалы из табл. А.29: 0.0197; 0.0833; 0.2084; 0.3772; 0.2084; 0.0833; 0.0197. По выборке находим количества наблюдений n_i , попавших в соответствующие интервалы (3, 25, 62, 101, 60, 39, 10). В соответствии с соотношением (2.1) вычисляют значение статистики $X_n^2 = 13.532$. При справедливости гипотезы H_0 статистика подчиняется χ_6^2 -распределению. Вероятность $P\{X_n^2 > 13.532\} = 0.0353$. При значениях уровня значимости $\alpha > 0.0353$ гипотеза H_0 будет отклонена. Результаты проверки гипотезы о согласии показаны на рис. 2.24.

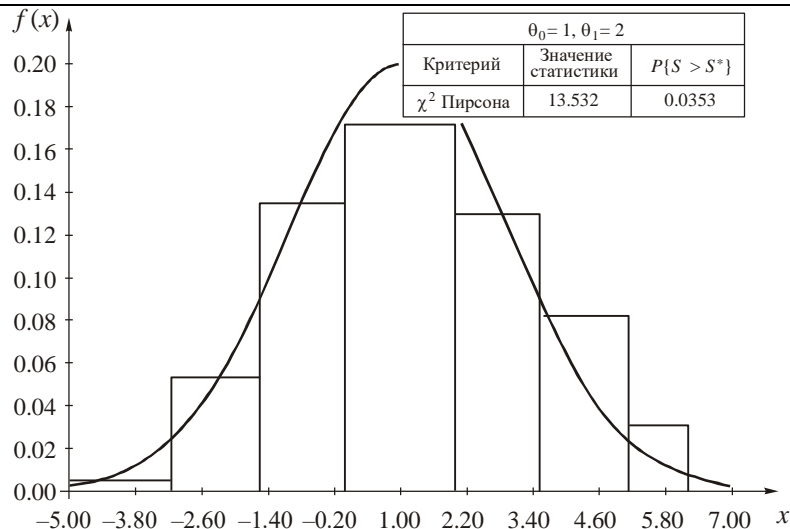


Рис. 2.24. Результаты проверки согласия к примеру 2.3

Пример 2.4. Проверяют сложную гипотезу о принадлежности выборки из примера 2.3 нормальному закону. По выборке оценивают параметр сдвига. Параметр масштаба $\theta_1 = 2$.

Вычисленная по выборке ОМП параметра сдвига $\hat{\theta}_0 = 1.2097$. Оставляют $k = 7$, так как из табл. А.25 $P_k n = 0.0536 \cdot 300 > 15$. Оптимальные граничные точки в виде $t_i = (x_i - \theta_0) / \theta_1$ берут из табл. А.24.

$(-1.6108; -0.8744; -0.2803; 0.2803; 0.8744; 1.6108)$. При $\theta_0 = 1.2097$; $\theta_1 = 2$ x_i принимают значения $-2.0119; -0.5391; 0.6491; 1.7703; 2.9585; 4.4313$. Вероятности попадания наблюдений в интервалы из

табл. А.25: 0.0536; 0.1373; 0.1987; 0.2208; 0.1987; 0.1373; 0.0536. По выборке находят числа наблюдений n_i , попавших в соответствующие интервалы (13, 58, 53, 62, 48, 33, 23). В соответствии с соотношением (2.1) вычисляют значение статистики $X_n^2 = 13.752$. При справедливости H_0 статистика приближенно подчиняется χ^2_5 -распределению. Вероятность $P\{X_n^2 > 13.752\} = 0.0173$. При значениях уровня значимости $\alpha > 0,0173$ гипотеза H_0 будет отклонена. Результаты проверки гипотезы о согласии иллюстрирует рис. 2.25.

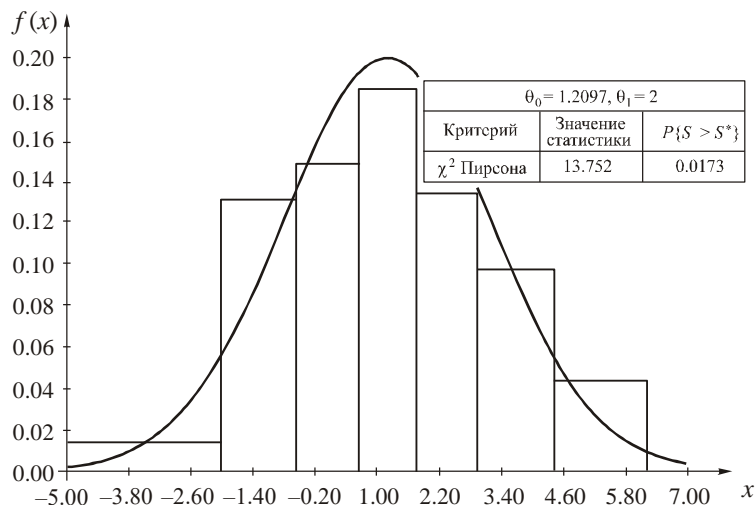


Рис. 2.25. Результаты проверки согласия к примеру 2.4

Пример 2.5. Проверяют сложную гипотезу о принадлежности выборки из примера 2.3 нормальному закону распределения. По выборке оценивают масштабный параметр (стандартное отклонение). Параметр сдвига $\theta_0 = 1$.

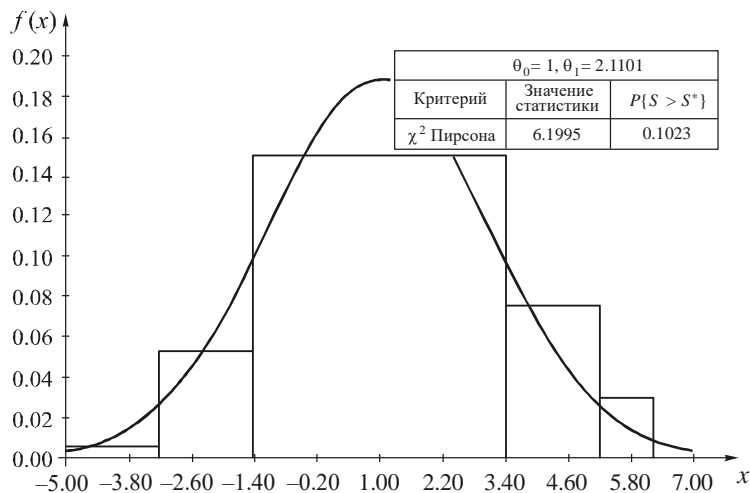


Рис. 2.26. Результаты проверки согласия к примеру 2.5

По выборке вычислена ОМП параметра масштаба $\hat{\theta}_0 = 2.1101$. Выбирают $k=5$, так как из табл. А.27 $P_k n = 0,023 \cdot 300 > 6$ и при большем числе интервалов ожидаемое число попаданий наблюдений в крайние интервалы не превышает 3. Оптимальные граничные точки в виде $t_i = (x_i - \theta_0) / \theta_1$ берут из табл. А.26: -1.9956 ; -1.1401 ; 1.1401 ; 1.9958 . При $\theta_0 = 1$, $\theta_1 = 2.1101$ x_i принимают значения –

3.21092; -1.40573; 3.405725; 5.211338. Вероятности попадания наблюдений в интервалы из табл. А.27: 0.0230; 0.1041; 0.7458; 0.1041; 0.0230. По выборке находят числа наблюдений n_i , попавших в соответствующие интервалы

(3, 29, 221, 41, 9). В соответствии с соотношением (2.1) вычисляют значение статистики $X_n^2 = 6.1995$.

При справедливости H_0 статистика приближенно подчиняется χ_3^2 -распределению. Вероятность $P\{X_n^2 > 6.1995\} = 0.1023$. При значениях уровня значимости $\alpha < 0.1$ нет оснований для отклонения гипотезы H_0 . Результаты проверки гипотезы о согласии показаны на рис. 2.26.

Пример 2.6. Проверяют сложную гипотезу о принадлежности выборки из примера 2.3 нормальному закону распределения. Проверяемая гипотеза имеет вид H_0 :

$$f(x) \in \left\{ f(x, \theta) = \frac{1}{\theta_1 \sqrt{2\pi}} e^{-(x-\theta_0)^2 / 2\theta_1^2}, \theta_0 \in (-\infty, \infty), \theta_1 \in (0, \infty) \right\}.$$

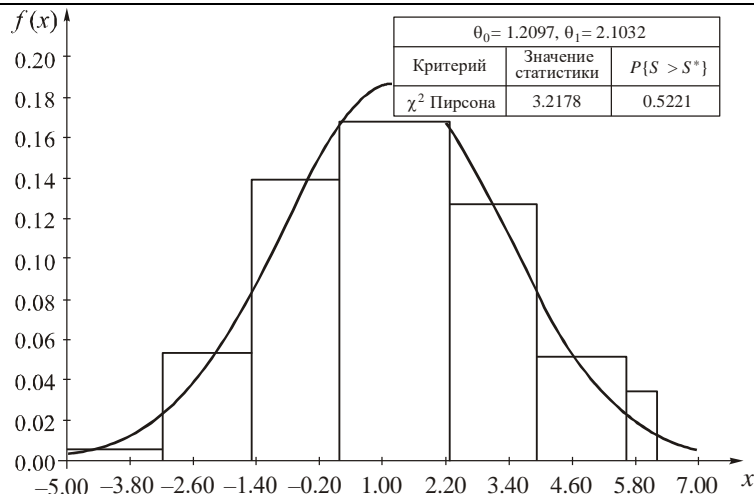


Рис. 2.27. Результаты проверки согласия к примеру 2.6

По выборке оценивают оба параметра нормального распределения. Полученные ОМП параметров $\hat{\theta}_0 = 1.2097$; $\hat{\theta}_1 = 2.1032$. Как и в примере 2.3, оптимальные граничные точки в виде $t_i = (x_i - \theta_0) / \theta_1$ берут из табл. А.28 при $k = 7$ (−2.0600; −1.2647; −0.4918; 0.4918; 1.2647; 2.0600), так как из табл. А.29 $P_k n = 0.0197 \cdot 300 > 5$. При $\theta_0 = 1.2097$; $\theta_1 = 2.1032$ x_i принимают значения −3.1435; −1.4629; 0.1704; 2.2490; 3.8823; 5.5629. Вероятности попадания наблюдений в интервал из табл. А.29: 0.0197; 0.0833; 0.2084; 0.3772; 0.2084; 0.0833; 0.0197. По выборке находят числа наблюдений n_i , попавших в соответствующие интервалы (3, 29, 67, 105, 64, 26, 6). В соответствии с соотношением (2.1) вычисляют значение статистики $X_n^2 = 3.2178$. При справедливости гипотезы H_0 статистика

приближенно подчиняется χ_4^2 -распределению. Вычисленное значение вероятности $P\{X_n^2 > 3.2178\} = 0.5221$ подтверждает хорошее согласие экспериментальных данных с теоретическим распределением по данному критерию. При значениях уровня значимости $\alpha < 0.5221$ гипотеза H_0 будет принята. Результаты проверки гипотезы о согласии иллюстрирует рис. 2.27.

Пример 2.7. Проверяют сложную гипотезу о принадлежности выборки, представленной в примере 2.2, распределению Вейбулла. Проверяют гипотезу

$$H_0: f(x) \in \left\{ f(x, \theta) = \frac{\theta_0(x - \theta_2)^{\theta_0 - 1}}{\theta_1^{\theta_0}} \exp \left[- \left(\frac{x - \theta_2}{\theta_1} \right)^{\theta_0} \right], \right. \\ \left. \theta_0, \theta_1 \in (0, \infty), \theta_2 = 0 \right\}$$

при оценивании по выборке параметров θ_0 и θ_1 методом максимального правдоподобия.

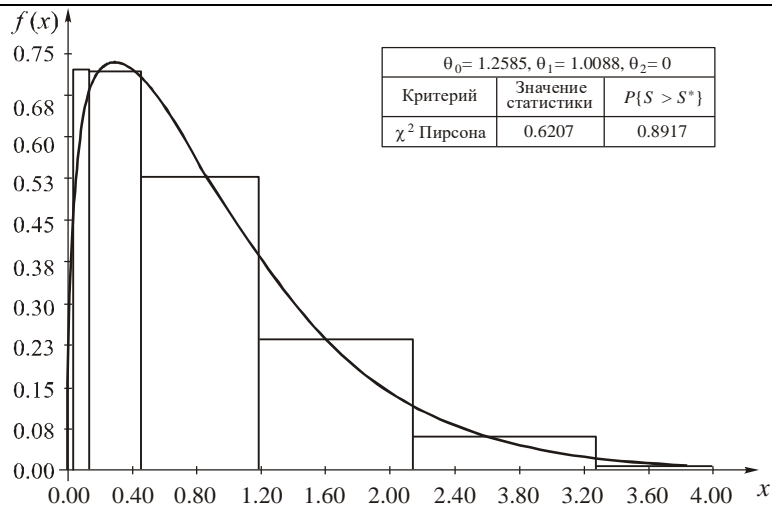


Рис. 2.28. Результаты проверки согласия к примеру 2.7

Вычисленные ОМП $\hat{\theta}_0 = 1.2585$, $\hat{\theta}_1 = 1.0088$. Для того чтобы ожидаемое число наблюдений, попавших в интервал с наименьшей вероятностью, было не очень малым, по табл. А.23 выбирают $k = 6$. В этом случае $P_k n = 0,0122 \cdot 300 > 3$. Из табл. А.22 в виде $t_i = [(x_i - \theta_2)/\theta_1]^{\theta_0}$ берут оптимальные граничные точки: 0.0772; 0.3649; 1.2269; 2.5726; 4.4096. При значениях параметров $\theta_0 = 1.2585$; $\theta_1 = 1.0088$; $\theta_2 = 0$ по значениям t_i определяем граничные точки интервалов

$$x_i = \theta_2 + \exp\left(\frac{1}{\theta_0} \ln t_i + \ln \theta_1\right):$$

0.1318; 0.4528; 1.1868; 2.1374; 3.2798. Вероятности попадания наблюдений в интервал из табл. А.23: 0.0743; 0.2314; 0.4011; 0.2169; 0.0641; 0.0122. По выборке находят числа наблюдений n_i , попавших в соответствующие интервалы (24, 69, 116, 67, 21, 3). Значение статистики $X_n^2 = 0.6207$. При справедливости гипотезы H_0 статистика приближенно подчиняется χ_3^2 -распределению. Так как вычисленная вероятность $P\{X_n^2 > 0.6207\} = 0.8917$, то это свидетельствует об очень хорошем согласии данных с теоретическим распределением. Результаты проверки гипотезы иллюстрирует рис. 2.28.

2 2.11. Порядок применения критериев типа χ^2 при проверке простой гипотезы

При проверке простой гипотезы о согласии опытного распределения с теоретическим распределением случайной величины X действуют следующим образом.

1. Формулируют проверяемую гипотезу, выбирая теоретическое распределение случайной величины $F(x, \theta)$, согласие которого с опытным распределением этой величины следует проверить.
2. Из совокупности отбирают случайную выборку объемом n .
3. В зависимости от объема выборки n в соответствии с рекомендациями раздела 2.8 выбирают число интервалов k .
4. Выбирают граничные точки интервалов группирования:

а) интервалы можно выбирать равной длины, в этом случае необходимо подсчитать число n_i и определить значения вероятностей $P_i(\theta)$;

б) интервалы можно выбирать равновероятными, тогда $P_i(\theta) = 1/k$, но необходимо вычислить значения граничных точек x_i , разделяющих интервалы;

в) можно разбивать выборку на интервалы в соответствии с асимптотически оптимальным группированием для данного закона распределения (см. раздел 2.9, табл. 2.8). Такое разбиение предпочтительно, поскольку повышает способность критерия различать близкие гипотезы.

5. После подсчета n_i и $P_i(\theta)$ в соответствии с выбранным критерием проверки вычисляют значение статистики S^* критерия по формуле (2.1) или (2.6).

6. В соответствии с χ^2_{k-1} -распределением по формуле (2.5) вычисляют значение $P\{S > S^*\}$. Если $P\{S > S^*\} > \alpha$, где α – задаваемый уровень значимости, то нет оснований для отклонения проверяемой гипотезы. В противном случае проверяемую гипотезу H_0 отвергают.

Можно вычисленное значение статистики S^* сравнить с критическим значением $S_{r,\alpha}$, где $r = k - 1$ – число степеней свободы, определяемым из уравнения

$$g_1(s) = \frac{1}{2^{k/2} \Gamma(k/2)} s^{k/2-1} e^{-s/2}.$$

Значения $S_{r,\alpha}$ приведены в приложении А. Гипотезу о согласии отвергают, если значение статистики попадает в критическую область, т. е. при $S^* > S_{r,\alpha}$.

2.12. Порядок применения критериев типа χ^2 при проверке сложной гипотезы

1. При проверке сложной гипотезы и вычислении оценок параметров по группированным данным в результате минимизации статистик, задаваемых формулами (2.1) и (2.6), или в результате максимизации функции правдоподобия (2.8) порядок проверки аналогичен случаю простой гипотезы (см. разд. 2.11) с заданием числа степеней свободы $r = k - m - 1$, где m – число оцененных по данной выборке параметров. При этом остаются в силе рекомендации по способам группирования.

2. При проверке сложной гипотезы и вычислении оценок максимального правдоподобия по исходным негруппированным выборочным данным и использовании для проверки гипотезы статистики (2.11) число степеней свободы предельного распределения $r = k - 1$.

3. При вычислении оценок максимального правдоподобия по исходным негруппированным выборочным данным и применении для проверки гипотезы статистики, определяемой формулой (2.1) или (2.6), руководствуются следующими соображениями:

а) значения $S_{r,\alpha}$ процентных точек χ_r^2 -распределения при числе степеней свободы $r = k - m - 1$ меньше, чем значения процентных точек действительных предельных распределений статистик (2.1) и (2.6) для такой ситуации (см. раздел 2.7). Поэтому если $S^* < S_{r,\alpha}$, то гипотеза H_0 о согласии при заданном уровне значимости α заведомо не должна быть отвергнута;

б) значения процентных точек действительных распределений статистик (2.1) и (2.6) существенно отличаются от процентных точек χ_r^2 -распределения при числе степеней свободы $r = k - m - 1$.

При больших значениях S^* вероятности вида $P\{S > S^*\}$, вычисляемые в соответствии с χ_r^2 -распределением по формуле (2.5), достаточно близки к вероятностям $P\{S > S^*\}$, соответствующим действительным распределениям статистик.

При малых значениях S^* вероятности, вычисляемые по формуле (2.5), много меньше действительных значений $P\{S > S^*\}$ (см. разд. 2.7). Поэтому если приближенное значение, вычисленное по формуле (2.5), больше значения α , то заведомо $P\{S > S^*\} > \alpha$ и проверяемая гипотеза о согласии не должна быть отвергнута.

2.13. Выводы

Таким образом, если вы, применяя критерии типа χ^2 , стремитесь обеспечить корректность статистических выводов при обработке измерительной информации, необходимо обратить внимание на следующие три момента.

Во-первых, на то, по каким данным вычисляются оценки при проверке сложных гипотез. Предельными χ_{k-m-1}^2 -распределениями для критериев χ^2 Пирсона и отношения правдоподобия можно пользоваться только при оценивании параметров по группированным наблюдениям. Если вы в силу понятных причин отдаете предпочтение ОМП по точечным наблюдениям, целесообразней воспользоваться критерием Никулина (Рао–Робсона–Никулина). Используя в данной ситуации критерии χ^2 Пирсона и

отношения правдоподобия, следует помнить, что величина вероятности $P\{\chi^2 > X_n^2\}$, вычисленная в соответствии с χ_{k-m-1}^2 -распределением, оказывается заниженной по сравнению с истинной.

Во-вторых, на то, каким образом разбить область определения случайной величины на интервалы. Использование асимптотически оптимального группирования максимизирует мощность критериев χ^2 Пирсона и отношения правдоподобия по отношению к близким гипотезам в случае простых и сложных гипотез. Кроме того, применение таблиц асимптотически оптимального группирования благодаря тому, что они содержат значения вероятностей попадания в интервал, облегчает и процесс вычислений. В случае критерия Никулина можно воспользоваться или асимптотически оптимальным группированием, или разбиением на интервалы равных вероятностей.

В-третьих, на выбор числа интервалов. Выбор слишком большого числа интервалов приводит к падению мощности. Оптимальное число интервалов k зависит от объема выборки n и от конкретной пары конкурирующих гипотез H_0 и H_1 . Чаще всего оптимальное k оказывается существенно меньше значений, рекомендуемых различными регламентирующими документами и задаваемых множеством эмпирических формул, представленных в [290]. Максимальная мощность критериев при заданном объеме выборки n часто достигается при минимально возможном или достаточно малом числе интервалов k . Если вас интересует конкретная пара конкурирующих гипотез, относительно которых часто приходится принимать решение, воспользуйтесь соотношением (2.20) для подбора оптимального числа интервалов k при заданном объеме выборки n . Если это окажется затруднительным, можно при выборе числа интервалов опираться на таблицы асимптотически оптимального группирования, выбирая k таким образом, чтобы ожидаемое число наблюдений, попадающих в любой интервал при асимптотически оптимальном группировании, было не очень малым: $nP_i \geq 5 \dots 10$. Как показывает практика, в этом случае число k обычно оказывается близким к оптимальному.

Выполнив первое условие, мы будем иметь возможность точно вычислять значение статистики критерия, соответствующее задаваемой величине вероятности ошибки первого рода α (или вычислять достигаемый уровень значимости по предельному распределению статистики $G(S|H_0)$). Выбрав же оптимальное число интервалов и подобрав оптимальное разбиение области определения случайной величины на интервалы, мы получим критерий максимальной мощности, наилучшим образом различающий конкретные конкурирующие гипотезы (обеспечивающий минимальную вероятность ошибки второго рода β при заданной вероятности ошибки первого рода α).

Результаты исследований критериев типа χ^2 были использованы нами при подготовке рекомендаций по стандартизации Р 50.1.033–2001 [301]. Анализ ошибок, наиболее часто совершаемых при применении критериев согласия типа χ^2 , посвящена работа [221].

Следует отметить, что материал, изложенный в данном разделе, не охватывает всех тонкостей применения критериев типа χ^2 . Много полезного относительно их использования и обширную библиографию можно найти в книге [52]. Но и это множество работ пока не даст ответа на все вопросы, которые могут возникнуть и возникают в связи с применением критериев данного типа