Math-Net.Ru

Общероссийский математический портал

Б. И. Селиванов, О предельных распределениях статистики χ^2 К. Пирсона в схеме независимых испытаний, *Матем. заметки*, 2008, том 83, выпуск 6, 899–911

DOI: <https://doi.org/10.4213/mzm4050>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением <http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 54.198.187.58

26 апреля 2023 г., 16:10:44

О предельных распределениях статистики χ^2 К. Пирсона в схеме независимых испытаний

Б. И. Селиванов

Изучается статистика хи-квадрат Пирсона в схеме независимых и в общем случае неоднородных испытаний с фиксированным числом исходов. Предполагается, что вероятности появления исходов испытаний удовлетворяют некоторым условиям. На рассматриваемую схему переносятся результаты, известные для статистики хи-квадрат в случае полиномиальных испытаний. Получены явные выражения и оценки для математического ожидания и дисперсии статистики хи-квадрат. Для статистики хи-квадрат, центрированной и нормированной подходящим образом, найдены предельные распределения (нормальное, распределение суммы квадратов нормальных случайных величин и, в частности, распределение хи-квадрат). Указаны условия сходимости к соответствующим предельным распределениям.

Библиография: 11 названий.

1. Введение. Рассмотрим последовательность n независимых случайных величин (с.в.)

$$X_1, X_2, \dots, X_n, \quad (1)$$

которые принимают значения в множестве $\{1, \dots, N\}$ (N предполагается фиксированным). В дальнейшем будем называть (1) *последовательностью независимых испытаний*, а значения с.в. (1) – *исходами испытаний*. При этом в каждом из испытаний (1) реализуется некоторая полиномиальная схема с указанным множеством исходов. Независимые испытания называются *полиномиальными*, если они однородны, т.е. в них реализуется одна и та же полиномиальная схема.

Условимся о том, что в дальнейшем изложении, если не оговорено противное, индексы i и j принимают значения $1, \dots, N$.

Согласно основной гипотезе (обозначим ее $\mathbf{H}_0$) (1) является последовательностью полиномиальных испытаний и вероятность появления исхода i в ней равна $p_i > 0$, $\sum_{i=1}^N p_i = 1$. Для построения критерия согласия с гипотезой $\mathbf{H}_0$ часто используют статистику хи-квадрат Пирсона (см., например, [1; гл. 30])

$$\chi^2 = \sum_{i=1}^N \frac{1}{np_i} (h_i - np_i)^2 = \sum_{i=1}^N \frac{h_i^2}{np_i} - n, \quad (2)$$

где $h_i = h_i(n)$ – частота исхода i в испытаниях (1), $\sum_{i=1}^N h_i = n$.

Работа выполнена при поддержке программы Отделения математических наук РАН “Современные проблемы теоретической математики” и программы “Ведущие научные школы” (грант № НШ-4129.2006.1).

Предельные распределения статистики χ^2 как при гипотезе $\mathbf{H}_0$, так и при простых конкурирующих с $\mathbf{H}_0$ гипотезах (альтернативах $\mathbf{H}_0$), в соответствии с которыми испытания (1) являются полиномиальными, известны (см. [1; § 30.1], [2; раздел 7.12], [3], [4]). Но на практике предположение об однородности испытаний (1) не всегда выполняется. Поэтому представляет интерес вопрос об асимптотическом поведении распределения статистики (2), когда испытания независимы, но, возможно, не являются полиномиальными. Цель настоящей работы – перенести известные из указанных работ результаты, относящиеся к статистике χ^2 , на общий случай независимых испытаний.

Предположим, что испытания (1), возможно, неоднородны и вероятности появления исходов $1, \dots, N$ в t -м испытании равны соответственно

$$p_i(t) \geq 0, \quad i = 1, \dots, N, \quad \sum_{i=1}^N p_i(t) = 1, \quad t = 1, 2, \dots, n. \quad (3)$$

Набор вероятностей (3) определяет простую фиксированную альтернативу, конкурирующую с основной гипотезой; обозначим ее $\mathbf{H}(p)$. Ниже на вероятности $p_i(t)$ мы наложим некоторые условия, обеспечивающие справедливость многомерной центральной предельной теоремы для вектора частот исходов $(h_1, \dots, h_N)$.

Будем рассматривать также асимптотическое поведение распределения статистики χ^2 в случае альтернатив $\mathbf{H}(\tilde{p})$ гипотезе $\mathbf{H}_0$, “близких” к $\mathbf{H}(p)$ в том смысле, что при $n \rightarrow \infty$ вероятности исходов испытаний (1) имеют вид

$$\tilde{p}_i(t) = p_i(t) + \frac{d_i(t)}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right), \quad i = 1, \dots, N, \quad t = 1, \dots, n, \quad (4)$$

где оценка $o(1/\sqrt{n})$ равномерна по t , величины $d_i(t)$ не зависят от n и удовлетворяют условиям

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n d_i(t) = d_i, \quad \frac{1}{n^2} \sum_{t=1}^n d_i(t) d_j(t) = o(1), \quad i, j = 1, \dots, N. \quad (5)$$

Заметим, что из (3), (4) и первого из соотношений (5) следуют равенства

$$\sum_{t=1}^N d_i(t) = 0, \quad t = 1, 2, \dots, \quad \text{и} \quad \sum_{t=1}^N d_i = 0.$$

Ясно, что второе из соотношений (5) выполняется, если имеют место неравенства $|d_i(t)| \leq C$, где C – константа. Будем считать, что множество альтернатив (4) не содержит $\mathbf{H}(p)$.

Работа состоит из трех пунктов. В п. 2 приведены формулировки вспомогательных и основных результатов, полученных в работе. Их доказательства даны в п. 3.

В лемме 2, имеющей и самостоятельное значение, выписаны явные выражения для математического (мат.) ожидания и дисперсии статистики χ^2 в общем случае независимых испытаний. Предельные теоремы для статистики χ^2 получены в предположении справедливости альтернатив $\mathbf{H}(p)$ или $\mathbf{H}(\tilde{p})$. При доказательстве теорем 1 и 2, по существу, используется “ δ -метод”, изложение которого можно найти, например, в [5; § 6а.2] В теореме 1 указаны достаточные условия асимптотической

нормальной статистики χ^2 . В теореме 2 показано, что некоторое выражение, содержащее χ^2 , сходится по распределению к сумме квадратов нормальных с.в. Вид предельного распределения в теоремах 1 и 2 определяется условиями, найденными в лемме 4 с использованием неравенства Коши–Буняковского. Наконец, в теореме 3 получены простые достаточные условия, при выполнении которых упомянутое выше выражение имеет асимптотическое распределение хи-квадрат (центральное или нецентральное).

2. Формулировка результатов. Введем с.в.

$$\xi_i = \frac{1}{n} h_i$$

и, используя вероятности (3), положим

$$u_{ij}^{(k,l)}(n) = \sum_{t=1}^n p_i^k(t) p_j^l(t), \quad i, j = 1, \dots, N, \quad k, l = 0, 1, \dots, \quad (6)$$

$$u_i^{(k)}(n) = u_{ii}^{(k,0)}(n), \quad i = 1, \dots, N, \quad k = 0, 1, \dots, \quad (7)$$

$$a_i(n) = \frac{1}{n} u_i^{(1)}(n) = \frac{1}{n} \sum_{t=1}^n p_i(t), \quad i = 1, \dots, N, \quad (8)$$

$$c_{ij}(n) = \frac{1}{n} u_{ij}^{(1,1)}(n) = \frac{1}{n} \sum_{t=1}^n p_i(t) p_j(t), \quad i, j = 1, \dots, N. \quad (9)$$

ЛЕММА 1. 1) Если при проведении испытаний (1) справедлива альтернатива $\mathbf{H}(p)$, то для мат. ожиданий, дисперсий и ковариаций с.в. ξ_i выполняются равенства

$$\begin{aligned} E\xi_i &= a_i(n), \quad D\xi_i = \frac{1}{n} [a_i(n) - c_{ii}(n)], \quad i = 1, \dots, N, \\ \text{Cov}(\xi_i, \xi_j) &= -\frac{1}{n} c_{ij}(n), \quad i, j = 1, \dots, N, \quad i \neq j, \end{aligned}$$

где величины $a_i(n)$ и $c_{ij}(n)$ определяются формулами (8) и (9).

2) Пусть в испытаниях (1) выполняется либо альтернатива $\mathbf{H}(p)$, либо альтернатива вида $\mathbf{H}(\tilde{p})$ и существуют пределы

$$\lim_{n \rightarrow \infty} a_i(n) = a_i, \quad \lim_{n \rightarrow \infty} c_{ij}(n) = c_{ij}, \quad i, j = 1, \dots, N. \quad (10)$$

Тогда при $n \rightarrow \infty$ для случайного вектора

$$\bar{\xi}_n = (\sqrt{n}(\xi_1 - a_1(n)), \dots, \sqrt{n}(\xi_N - a_N(n)))$$

имеет место сходимость

$$\bar{\xi}_n \Rightarrow \begin{cases} N(\bar{0}, \mathbf{A}), & \text{если верна гипотеза } \mathbf{H}(p), \\ N(\bar{d}, \mathbf{A}), & \text{если справедлива альтернатива } \mathbf{H}(\tilde{p}), \end{cases}$$

где значок “ $\Rightarrow$ ” обозначает сходимость с.в. по распределению, $\bar{0}$ – нулевой вектор, компоненты вектора $\bar{d} = (d_1, \dots, d_N)$ составляют числа d_i (см. первое из соотношений (5)), $N(\bar{0}, \mathbf{A})$ и $N(\bar{d}, \mathbf{A})$ – нормальные случайные векторы, имеющие векторы

мат. ожиданий $\bar{0}$ и $\bar{d}$ соответственно и ковариационную матрицу

$$\mathbf{A} = \|a_{ij}\|_{i,j=1}^N, \quad a_{ij} = \delta_{ij}a_i - c_{ij}, \quad i, j = 1, \dots, N, \quad (11)$$

δ_{ij} – символ Кронекера.

ЛЕММА 2. Если в испытаниях (1) выполняется альтернатива $\mathbf{H}(p)$, то для мат. ожидания и дисперсии статистики χ^2 имеют место равенства

$$E\chi^2 = \frac{1}{n} \sum_{i=1}^N \frac{1}{p_i} [u_i^{(1)}(n) - u_i^{(2)}(n) + (u_i^{(1)}(n))^2] - n, \quad (12)$$

$$\begin{aligned} D\chi^2 = & \frac{1}{n^2} \sum_{i=1}^N \frac{1}{p_i^2} [u_i^{(1)}(n) - 6u_i^{(2)}(n) + 8u_i^{(3)}(n) \\ & + 6(u_i^{(1)}(n))^2 - 12u_i^{(1)}(n)u_i^{(2)}(n) + 4(u_i^{(1)}(n))^3] \\ & + \frac{1}{n^2} \sum_{i,j=1}^N \frac{1}{p_i p_j} [-6u_{ij}^{(2,2)}(n) + 2u_{ij}^{(2,1)}(n) + 2u_{ij}^{(1,2)}(n) - u_{ij}^{(1,1)}(n) \\ & - 2u_i^{(1)}(n)u_{ij}^{(1,1)}(n) + 2(u_{ij}^{(1,1)}(n))^2 - 2u_j^{(1)}(n)u_{ij}^{(1,1)}(n) \\ & + 4u_{ij}^{(1,2)}(n)u_i^{(1)}(n) + 4u_{ij}^{(2,1)}(n)u_j^{(1)}(n) - 4u_i^{(1)}(n)u_{ij}^{(1,1)}(n)u_j^{(1)}(n)]. \end{aligned} \quad (13)$$

Если вероятности (3) имеют вид

$$p_i(t) = q_i \geq 0, \quad i = 1, \dots, N, \quad t = 1, \dots, n, \quad \sum_{i=1}^N q_i = 1, \quad (14)$$

то формулы (12) и (13) сводятся к известным выражениям для мат. ожидания и дисперсии статистики (2) как при гипотезе $\mathbf{H}_0$, так и при фиксированных полиномиальных альтернативах гипотезе $\mathbf{H}_0$ (см., например, [1; с. 454] и [6; с. 619] соответственно).

СЛЕДСТВИЕ 1. Пусть в испытаниях (1) выполняется альтернатива $\mathbf{H}(p)$ и $n \rightarrow \infty$. Тогда справедливы оценки

$$E\chi^2 = (b_n - 1)n + O(1), \quad D\chi^2 = 4V_n n + O(1), \quad (15)$$

где

$$b_n = \sum_{i=1}^N \frac{a_i^2(n)}{p_i}, \quad V_n = \sum_{i=1}^N \frac{a_i^3(n)}{p_i^2} - \sum_{i,j=1}^N c_{ij}(n) \frac{a_i(n)a_j(n)}{p_i p_j}. \quad (16)$$

Если в случае $n \rightarrow \infty$ вместо (10) выполняются соотношения

$$a_i(n) = a_i + O\left(\frac{1}{\sqrt{n}}\right), \quad c_{ij}(n) = c_{ij} + O\left(\frac{1}{\sqrt{n}}\right), \quad i, j = 1, \dots, N, \quad (17)$$

то оценки (15) принимают вид

$$E\chi^2 = (b - 1)n + O(1), \quad D\chi^2 = 4Vn + O(1), \quad (18)$$

где

$$b = \sum_{i=1}^N \frac{a_i^2}{p_i}, \quad V = \sum_{i=1}^N \frac{a_i^3}{p_i^2} - \sum_{i,j=1}^N c_{ij} \frac{a_i a_j}{p_i p_j}. \quad (19)$$

ЛЕММА 3. Если существуют пределы (10), то для величин a_i и c_{ij} выполняются равенства

$$\sum_{i=1}^N a_i = 1, \quad \sum_{i,j=1}^N c_{ij} = 1, \quad (20)$$

$$\sum_{i=1}^N c_{ij} = a_j, \quad \sum_{j=1}^N c_{ij} = a_i, \quad i, j = 1, \dots, N. \quad (21)$$

Нас интересуют условия равенства нулю величины V , входящей во вторую из оценок (18) и определенной второй из формул (19). Чтобы сформулировать лемму 4, в которой представлены эти условия, нам потребуется следующее построение. Предположим, что пределы (10) существуют и координаты вектора $a = (a_1, \dots, a_N)$ определяются первыми из соотношений (10). Множество номеров ненулевых координат вектора a обозначим M . Заметим, что если i -я координата вектора a равна нулю, то вследствие второго из равенств (21) имеем $c_{ij} = 0$ для всех $j = 1, \dots, N$. Построим граф G с множеством вершин M , соединив две его вершины $i, j \in M$ (неориентированным) ребром, если $c_{ij} > 0$. Предположим, что граф G имеет компоненты связности $G_1, \dots, G_K$, $K \geq 1$, с множествами вершин $M_1, \dots, M_K$, $\bigcup_{k=1}^K M_k = M$. Теперь мы готовы представить условия, обеспечивающие равенство $V = 0$.

ЛЕММА 4. В предположениях леммы 1 величина V неотрицательна и равна нулю тогда и только тогда, когда для ненулевых координат вектора a выполняются равенства

$$a_i = \frac{\alpha_k}{\beta_k} p_i, \quad \text{если } i \in M_k, \quad (22)$$

где

$$\alpha_k = \sum_{i \in M_k} a_i, \quad \beta_k = \sum_{i \in M_k} p_i, \quad k = 1, \dots, K. \quad (23)$$

В частном случае, когда вероятности (3) имеют вид (14), утверждение леммы 4 можно найти в работе [4].

Теперь представим предложение, в котором указаны условия, достаточные для асимптотической нормальности статистики χ^2 .

ТЕОРЕМА 1. Предположим, что для независимых испытаний (1) верна либо альтернатива $\mathbf{H}(p)$, либо альтернатива вида $\mathbf{H}(\tilde{p})$, существуют пределы (10), выполняется неравенство $V > 0$, и пусть $n \rightarrow \infty$. Тогда с.в.

$$\frac{1}{2\sqrt{Vn}}(\chi^2 - (b_n - 1)n) \quad (24)$$

асимптотически нормальна с параметрами $(0, 1)$ в случае гипотезы $\mathbf{H}(p)$ и с параметрами $((1/V) \sum_{i=1}^N d_i(a_i/p_i), 1)$ в случае альтернативы $\mathbf{H}(\tilde{p})$, где V задается второй из формул (19), b_n определяется первым из равенств (16), а числа d_i – первыми из соотношений (5).

ЗАМЕЧАНИЕ 1. Если вместо (10) выполняются более точные соотношения (17), то ввиду оценок (18) с.в. (24) в формулировке теоремы 1 можно заменить на $1/(2\sqrt{Vn}) \times (\chi^2 - (b-1)n)$. Вид величин b и V указан в (19).

Далее рассмотрим асимптотическое поведение распределения статистики χ^2 в случае $V = 0$. Напомним, что условия, обеспечивающие это равенство, описаны в лемме 4.

ТЕОРЕМА 2. *Предположим, что при проведении случайных независимых испытаний (1) выполняется либо альтернатива $\mathbf{H}(p)$, либо альтернатива вида $\mathbf{H}(\tilde{p})$ и $n \rightarrow \infty$. Пусть при этом справедливы соотношения (10) и $V = 0$. Тогда имеет место сходимость*

$$\chi^2 + \left(1 + \sum_{i=1}^N \frac{a_i^2(n)}{p_i}\right)n - 2 \sum_{i=1}^N \frac{a_i(n)}{p_i} h_i \Rightarrow \sum_{i=1}^N \eta_i^2, \quad (25)$$

где $\eta_1, \dots, \eta_N$ – нормальные с.в. с ковариационной матрицей

$$A^* = \|a_{ij}^*\|_{i,j=1}^N, \quad a_{ij}^* = \frac{1}{\sqrt{p_i p_j}} (\delta_{ij} a_i - c_{ij}), \quad i, j = 1, \dots, N, \quad (26)$$

из мат. ожидания определяются равенствами

$$E\eta_i = \begin{cases} 0, & \text{если верна альтернатива } \mathbf{H}(p), \\ \frac{d_i}{\sqrt{p_i}}, & \text{если верна альтернатива } \mathbf{H}(\tilde{p}), \end{cases} \quad (27)$$

и числа d_i определяются первыми из соотношений (5).

ЗАМЕЧАНИЕ 2. Если вместо (10) в теореме 2 потребовать выполнение условий (17), то соотношение (25) упрощается и приводится к виду

$$\chi^2 - (b-1)n \Rightarrow \eta_1^2 + \dots + \eta_N^2.$$

ЗАМЕЧАНИЕ 3. Выражение из левой части соотношения (25), используя обозначения (7), можно компактнее записать в виде суммы $\sum_{i=1}^N (1/(np_i))(h_i - u_i^{(1)}(n))^2$. В этой связи упомянем публикацию [7], в которой в схеме независимых испытаний другими методами исследуется статистика

$$\sum_{i=1}^N \frac{1}{u_i^{(1)}(n)} (h_i - u_i^{(1)}(n))^2.$$

ЗАМЕЧАНИЕ 4. Относительно распределения суммы квадратов нормальных с.в. η_i , для которых $E\eta_i = 0$, известен достаточно просто формулируемый результат Чернова и Лемана [8]. Если же вектор мат. ожиданий с.в. η_i отличен от нулевого, то распределение суммы $\sum_{i=1}^N \eta_i^2$ в общем случае устроено более сложно (см., например, [9]).

Наконец, укажем достаточные условия, при выполнении которых выражение из левой части соотношения (25) имеет в качестве предельного распределение хи-квадрат.

ТЕОРЕМА 3. Предположим, что в испытаниях (1) $n \rightarrow \infty$ и выполняется либо альтернатива $\mathbf{H}(p)$, либо альтернатива вида $\mathbf{H}(\tilde{p})$. Пусть, далее, для вероятностей (3) справедливы соотношения

$$\lim_{t \rightarrow \infty} p_i(t) = q_i \geq 0, \quad \sum_{i=1}^N q_i = 1,$$

и при этом

$$V_0 = \sum_{i=1}^N \frac{q_i^3}{p_i^2} \left(\sum_{i=1}^N \frac{q_i^2}{p_i} \right)^2 = 0, \quad (28)$$

а альтернатива $\mathbf{H}(\tilde{p})$ такова, что в первых соотношениях (5)

$$d_i = 0, \quad \text{если } i \in \overline{M} = \{1, \dots, N\} \setminus M, \quad M = \{j : q_j > 0, j = 1, \dots, N\}.$$

Тогда

$$\begin{aligned} & \beta \left[\chi^2 + \left(1 + \sum_{i=1}^N \frac{a_i^2(n)}{p_i} \right) n - 2 \sum_{i=1}^N \frac{a_i(n)}{p_i} h_i \right] \\ & \Rightarrow \begin{cases} \chi_{|M|-1}^2, & \text{если справедлива альтернатива } \mathbf{H}(p), \\ \chi_{|M|-1}^2 \left(\beta \sum_{i=1}^N \frac{d_i^2}{p_i} \right), & \text{если верна альтернатива } \mathbf{H}(\tilde{p}), \end{cases} \end{aligned} \quad (29)$$

где $\beta = \sum_{i \in M} p_i$, $|M|$ – число элементов множества M , $\chi_k^2(\lambda)$ – с.в., имеющая нецентральный распределение хи-квадрат с k степенями свободы и параметром нецентральности λ , $\chi_k^2 = \chi_k^2(0)$.

СЛЕДСТВИЕ 2. Пусть q_i – вероятности такие, что

$$q_i = \frac{1}{\beta} p_i, \quad \text{если } i \in M, \quad \text{и} \quad q_i = 0, \quad \text{если } i \in \overline{M}, \quad (30)$$

где p_i – вероятности, определяющие гипотезу $\mathbf{H}_0$, β , M и $\overline{M}$ заданы в формулировке теоремы 3. Предположим, что испытания (1) являются полиномиальными и при их проведении выполняется либо альтернатива $\mathbf{H}$ с вероятностями исходов (30), либо альтернатива $\tilde{\mathbf{H}}$ с вероятностями исходов

$$\tilde{p}_i(t) = q_i + \frac{d_i(t)}{\sqrt{n}} + O\left(\frac{1}{\sqrt{n}}\right), \quad i = 1, \dots, N, \quad t = 1, \dots, n,$$

где величина $d_i(t)$, $t = 1, \dots, n$, не зависят от n и удовлетворяют условиям (5). Тогда при $n \rightarrow \infty$ имеет место сходимость

$$\beta \left[\chi^2 - \left(\frac{1}{\beta} - 1 \right) n \right] \Rightarrow \begin{cases} \chi_{|M|-1}^2, & \text{если верна альтернатива } \mathbf{H}, \\ \chi_{|M|-1}^2(\beta\delta), & \text{если справедлива альтернатива } \tilde{\mathbf{H}}, \end{cases}$$

где $\delta = \sum_{i \in M} d_i^2/p_i$ и числа d_i задаются первыми из соотношений (5).

Следствие 2 содержит все известные предельные теоремы (см. [1; § 30.1], [2; раздел 7.12] и [4]) для статистики χ^2 в случае полиномиальных испытаний с фиксированным числом исходов, если для вероятностей p_i и q_i выполняется условие (28). Более того, утверждение следствия 2 для альтернатив вида $\tilde{\mathbf{H}}$, “близких” $\mathbf{H} \neq \mathbf{H}_0$, является новым.

3. Доказательства. Приведем доказательства утверждений, сформулированных в п. 2.

ДОКАЗАТЕЛЬСТВО ЛЕММЫ 1. Равенства, составляющие первое утверждение леммы, известны и в доказательстве не нуждаются. Второе утверждение леммы вытекает из первого, условий (10), определения альтернатив $\mathbf{H}(\tilde{p})$ и многомерной центральной предельной теоремы.

ДОКАЗАТЕЛЬСТВО ЛЕММЫ 2. Вторая из формул (2) позволяет написать равенства

$$E\chi^2 = \frac{1}{n} \sum_{i=1}^N \frac{\alpha_i^{(2)}}{p_i} - n, \quad (31)$$

$$\begin{aligned} D\chi^2 &= \frac{1}{n^2} \sum_{i=1}^N \frac{1}{p_i^2} D h_i^2 + \frac{1}{n^2} \sum_{\substack{i,j=1 \\ i \neq j}}^N \frac{1}{p_i p_j} \text{Cov}(h_i^2, h_j^2) \\ &= \frac{1}{n^2} \sum_{i=1}^N \frac{1}{p_i^2} (\alpha_i^{(4)} - (\alpha_i^{(2)})^2) + \frac{1}{n^2} \sum_{\substack{i,j=1 \\ i \neq j}}^N \frac{1}{p_i p_j} (\alpha_{ij}^{(2,2)} - \alpha_i^{(2)} \alpha_j^{(2)}), \end{aligned} \quad (32)$$

где

$$\alpha_i^{(2)} = E h_i^2, \quad \alpha_i^{(4)} = E h_i^4, \quad \alpha_{ij}^{(2,2)} = E h_i^2 h_j^2, \quad i \neq j, \quad i, j = 1, \dots, N, \quad (33)$$

– моменты случайного вектора $h = (h_1, \dots, h_N)$.

Теперь воспользуемся известными соотношениями между моментами и семиинвариантами случайных векторов, которые можно найти, например, в [10; с. 931], и формулами для семиинвариантов вектора $\bar{h}$. Последние можно сразу написать с помощью леммы 2.1 из работы [10] или достаточно просто вывести самостоятельно. В результате, учитывая равенства (31) и (32) и используя обозначения (6), (7) и (33), получим формулу (12) для $E\chi^2$ и следующее выражение для $D\chi^2$:

$$\begin{aligned} D\chi^2 &= \frac{1}{n^2} \sum_{i=1}^N \frac{1}{p_i^2} [u_i^{(1)}(n) - 7u_i^{(2)}(n) + 12u_i^{(3)}(n) \\ &\quad - 6u_i^{(4)}(n) + 6(u_i^{(1)}(n))^2 - 16u_i^{(1)}(n)u_i^{(2)}(n) + 8u_i^{(1)}(n)u_i^{(3)}(n) \\ &\quad + 2(u_i^{(2)}(n))^2 + 4(u_i^{(1)}(n))^3 - 4(u_i^{(1)}(n))^2 u_i^{(2)}(n)] \\ &\quad + \frac{1}{n^2} \sum_{\substack{i,j=1 \\ i \neq j}}^N \frac{1}{p_i p_j} [-6u_{ij}^{(2,2)}(n) + 2u_{ij}^{(2,1)}(n) + 2u_{ij}^{(1,2)}(n) - u_{ij}^{(1,1)}(n) \\ &\quad - 2u_i^{(1)}(n)u_{ij}^{(1,1)}(n) + 2(u_{ij}^{(1,1)}(n))^2 - 2u_j^{(1)}(n)u_{ij}^{(1,1)}(n) \\ &\quad + 4u_{ij}^{(1,2)}(n)u_i^{(1)}(n) + u_{ij}^{(2,1)}(n)u_j^{(1)}(n) - 4u_i^{(1)}(n)u_{ij}^{(1,1)}(n)u_j^{(1)}(n)]. \end{aligned} \quad (34)$$

Наконец, заметив, что

$$\sum_{\substack{i,j=1 \\ i \neq j}}^N g_{ij} = \sum_{i,j=1}^N g_{ij} - \sum_{i=1}^N g_{ii},$$

приведем равенство (34) к виду, указанному в формулировке леммы.

ДОКАЗАТЕЛЬСТВО СЛЕДСТВИЯ 1. Главный член в оценке мат. ожидания статистики χ^2 дает слагаемые

$$\frac{1}{n} \sum_{i=1}^N \frac{1}{p_i} (u_i^{(1)}(n))^2 \quad \text{и} \quad -n$$

суммы (12), а в оценке дисперсии χ^2 — члены

$$\frac{4}{n^2} \sum_{i=1}^N \frac{1}{p_i^2} (u_i^{(1)}(n))^3 \quad \text{и} \quad -\frac{4}{n^2} \sum_{i,j=1}^N \frac{1}{p_i p_j} u_i^{(1)}(n) u_i^{(1,1)}(n) u_j^{(1)}$$

суммы (13). Остальные слагаемые в правых частях формул (12) и (13) ввиду оценок $u_i^{(k)} \leq n$, $u_{ij}^{(k,l)} \leq n$, $k, l = 1, 2, \dots$, имеют порядок, самое большее $O(1)$. С учетом обозначений (8) и (9) эти наблюдения приводят к оценкам (15), в которых величины b_n и V_n определяются равенствами (16).

ДОКАЗАТЕЛЬСТВО ЛЕММЫ 3. Из равенств (3), (8) и (9) следует, что

$$\sum_{i=1}^N a_i(n) = 1, \quad \sum_{i,j=1}^N c_{ij}(n) = 1, \quad (35)$$

$$\sum_{i=1}^N c_{ij}(n) = a_j(n), \quad \sum_{j=1}^N c_{ij}(n) = a_i(n), \quad i, j = 1, \dots, N. \quad (36)$$

Устремим n в суммах (35) и (36) к ∞ и, учитывая (10), убедимся в справедливости равенств (20) и (21).

ДОКАЗАТЕЛЬСТВО ЛЕММЫ 4. Применим к сумме

$$\sum_{i,j=1}^N c_{ij} \frac{a_i a_j}{p_i p_j}$$

неравенство Коши–Буняковского. Получим

$$\sum_{i,j=1}^N c_{ij} \frac{a_i a_j}{p_i p_j} \leq \sqrt{\sum_{i,j=1}^N c_{ij} \frac{a_i^2}{p_i^2} \sum_{i,j=1}^N c_{ij} \frac{a_j^2}{p_j^2}}$$

и, принимая во внимание равенства (21), будем иметь

$$\sum_{i,j=1}^N c_{ij} \frac{a_i a_j}{p_i p_j} \leq \sum_{i=1}^N \frac{a_i^3}{p_i^2}. \quad (37)$$

Следовательно, $V \geq 0$. Более того, воспользуемся условиями равенства в неравенстве Коши–Буняковского и выясним, когда $V = 0$. А именно, равенство в (37) выполняется тогда и только тогда, когда существуют числа λ и μ , которые не зависят от i и j и не равны одновременно нулю, такие, что

$$\lambda \sqrt{c_{ij}} \frac{a_i}{p_i} = \mu \sqrt{c_{ij}} \frac{a_j}{p_j}, \quad i, j = 1, \dots, N. \quad (38)$$

Зафиксируем $i \in M$ и найдем в множестве M_k , содержащем i , элемент j (возможно, совпадающий с i), для которого $c_{ij} > 0$. Такое j по построению графа G всегда существует. Для выбранной пары i, j соответствующее равенство (38) принимает вид

$$\lambda \frac{a_i}{p_i} = \mu \frac{a_j}{p_j}, \quad (39)$$

где $a_i > 0$ и $a_j > 0$. Так как $c_{ji} = c_{ij}$, то среди равенств (38) имеется также равенство

$$\lambda \frac{a_j}{p_j} = \mu \frac{a_i}{p_i}. \quad (40)$$

Из (39) и (40) вытекает, что $\lambda = \mu \neq 0$, и поэтому система (38) приводится к виду

$$\sqrt{c_{ij}} \frac{a_i}{p_i} = \sqrt{c_{ij}} \frac{a_j}{p_j}, \quad i, j = 1, \dots, N. \quad (41)$$

Нетривиальные равенства (41), для которых $c_{ij} > 0$, соответствуют ребрам графа G , построенного в п. 2. Сокращая в них $\sqrt{c_{ij}}$, получаем систему

$$\frac{a_i}{p_i} = \frac{a_j}{p_j}, \quad i, j \in M, \quad c_{ij} > 0,$$

которая распадается на K независимых подсистем. Следовательно, для каждой связной компоненты G_k графа G имеют место равенства

$$\frac{a_i}{p_i} = \frac{a_j}{p_j} = \dots = \frac{a_h}{p_h} \quad (i, j, \dots, h - \text{вершины } G_k), \quad k = 1, 2, \dots, K.$$

Отсюда с учетом обозначений (23) вытекает справедливость равенств (22) и леммы 4.

ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ 1. Воспользуемся с.в. ξ_i , определенными в п. 2, и, используя второе выражение в правой части (2), запишем

$$\chi^2 = (b_n - 1)n + 2\sqrt{n}\eta_1(n) + \eta_2(n), \quad (42)$$

где

$$\eta_1(n) = \sum_{i=1}^N \frac{a_i(n)}{p_i} [\sqrt{n}(\xi_i - a_i(n))], \quad (43)$$

$$\eta_2(n) = \sum_{i=1}^N \frac{1}{p_i} [\sqrt{n}(\xi_i - a_i(n))]^2. \quad (44)$$

С помощью (42), учитывая положительность V и равенства (43) и (44), получаем

$$\frac{1}{2\sqrt{Vn}}(\chi^2 - (b_n - 1)n) = \frac{\eta_1(n) - E\eta_1(n)}{\sqrt{D\eta_1(n)}} \cdot \frac{\sqrt{D\eta_1(n)}}{\sqrt{V}} + \frac{E\eta_1(n)}{\sqrt{V}} + \frac{\eta_2(n)}{2\sqrt{Vn}}. \quad (45)$$

Из второго утверждения леммы 1 и векторной версии предложения 12 на с. 119 в [5] вытекает, что с.в. $\eta_1(n)$ и $\eta_2(n)$ при $n \rightarrow \infty$ сходятся по распределению. Более того, поскольку с.в. $\eta_1(n)$ является линейной комбинацией асимптотически нормальных с.в. $\sqrt{n}(\xi_i - a_i(n))$, то $\eta_1(n)$ также асимптотически нормальна с параметрами

$(E\eta_1(n), D\eta_1(n))$. Далее, согласно первому утверждению леммы 1 $E\eta_1(n) = 0$, если справедлива альтернатива $\mathbf{H}(p)$, и

$$E\eta_1(n) = \frac{1}{n} \sum_{i=1}^N \frac{a_i(n)}{p_i} \sum_{t=1}^n d_i(t) + O(1),$$

если верна альтернатива $\mathbf{H}(\tilde{p})$. При этом в случае справедливости $\mathbf{H}(p)$ ввиду первого утверждения леммы 1 имеем равенство $D\eta_1(n) = V_n$, где V_n определяется второй из формул (16). Если же в испытаниях (1) выполняется альтернатива $\mathbf{H}(\tilde{p})$, то аналогично находим, что $D\eta_1(n) = V_n + O(1)$, $n \rightarrow \infty$. Отсюда и из соотношений (5), (10) и (16) следует, что если $n \rightarrow \infty$, то

$$\frac{\eta_1(n) - E\eta_1(n)}{\sqrt{D\eta_1(n)}} \Rightarrow N(0, 1), \quad \frac{\sqrt{D\eta_1(n)}}{\sqrt{V}} \rightarrow 1 \quad (46)$$

($N(0, 1)$ – стандартная с.в.) и

$$\frac{1}{V} E\eta_1(n) \rightarrow \begin{cases} 0, & \text{если справедлива альтернатива } \mathbf{H}(p), \\ \frac{1}{V} \sum_{i=1}^N \frac{a_i}{p_i} d_i, & \text{если верна альтернатива } \mathbf{H}(\tilde{p}). \end{cases} \quad (47)$$

Кроме того, с.в. $(1/(2\sqrt{Vn}))\eta_2(n)$ в случае $n \rightarrow \infty$ сходится по вероятности к нулю, так как $\eta_2(n)$ имеет некоторое предельное распределение. Поэтому из соотношений (46), (47) и предложения 2с.4(X) в книге [5] следует, что при $n \rightarrow \infty$ правая часть (45) действительно асимптотически нормальна с указанными в формулировке теоремы параметрами. Теорема доказана.

ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ 2. Снова используя с.в. ξ_i и формулы (2), получим равенство

$$\chi^2 + \left(\sum_{i=1}^N \frac{a_i^2(n)}{p_i} + 1 \right) n - 2 \sum_{i=1}^N \frac{a_i(n)}{p_i} h_i = \sum_{i=1}^N \left[\frac{\sqrt{n}}{\sqrt{p_i}} (\xi_i - a_i(n)) \right]^2. \quad (48)$$

Если $n \rightarrow \infty$ и справедлива либо $\mathbf{H}(p)$, либо $\mathbf{H}(\tilde{p})$, то, учитывая (11), аналогично второму утверждению леммы 1 имеем сходимость

$$\left(\frac{\sqrt{n}}{\sqrt{p_1}} (\xi_1 - a_1(n)), \dots, \frac{\sqrt{n}}{\sqrt{p_N}} (\xi_N - a_N(n)) \right) \Rightarrow (\eta_1, \dots, \eta_N),$$

где $\eta_1, \dots, \eta_N$ – с.в., определенные в формулировке теоремы 2. Наконец, применим к сумме из правой части равенства (48) векторную версию теоремы 2с.4(XII) из книги [5] и завершим доказательство.

ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ 3. Заметим, что кроме пределов $\lim_{t \rightarrow \infty} p_i(t)$ существуют также пределы $\lim_{t \rightarrow \infty} p_i(t)p_j(t) = q_i q_j$. Поэтому из одного результата Коши (см., например, [11; с. 83] вытекает, что

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n p_i(t) = q_i, \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n p_i(t)p_j(t) = q_i q_j, \quad i, j = 1, \dots, N. \quad (49)$$

Учитывая обозначения (8), (9) и равенства (49), мы заключаем, что выполняется предположение теоремы 2 о справедливости условий (10) с $a_i = q_i$, $c_{ij} = q_i q_j$. Отсюда и из (28) вытекает равенство $V = V_0$ для второй из величин (19), а для статистики χ^2 согласно теореме 2 имеет место соотношение (25). Перепишем его следующим образом:

$$\beta \left[\chi^2 + \left(1 + \sum_{i=1}^N \frac{a_i^2(n)}{p_i} \right) n - 2 \sum_{i=1}^N \frac{a_i(n)}{p_i} h_i \right] \Rightarrow \sum_{i=1}^N (\sqrt{\beta} \eta_i)^2. \quad (50)$$

Нетрудно понять, что в данном случае граф G , который определен построением, предшествующем формулировке леммы 4, состоит из одной компоненты с множеством вершин M , заданным в формулировке теоремы 3. Поэтому с помощью леммы 4 мы получаем равенства

$$\frac{p_i}{q_i} = \beta, \quad i \in M. \quad (51)$$

Из (26), (51) и равенств $a_i = q_i$, $c_{ij} = q_i q_j$, указанных выше, следует, что элементы ковариационной матрицы $\mathbf{A}_1 = \|a_{ij}^{(1)}\|_{i,j=1}^N$ нормальных с.в. $\sqrt{\beta} \eta_i$ имеют вид

$$a_{ij}^{(1)} = \sqrt{\frac{\beta q_i}{p_j}} \delta_{ij} - \sqrt{q_i q_j}, \quad i, j = 1, \dots, N. \quad (52)$$

Покажем, что сумма $\sum_{i=1}^N (\sqrt{\beta} \eta_i)^2$ из правой части (50) в случае справедливости $\mathbf{H}(p)$ и $\mathbf{H}(\tilde{p})$ имеет распределения, указанные соотношениями (29). Для этого сначала, используя равенства (51) и (52), проверим идемпотентность матрицы $\mathbf{A}_1$, т.е. справедливость равенства $\mathbf{A}_1^2 = \mathbf{A}_1$. Затем воспользуемся теоремой 3.1 из работы [9], согласно которой квадратичная форма $\sum_{i=1}^N (\sqrt{\beta} \eta_i)^2$ имеет нецентральное распределение хи-квадрат тогда и только тогда, когда выполняются равенства

$$\mathbf{A}_1^3 = \mathbf{A}_1^2, \quad \mathbf{A}_1^2 \bar{\mu} = \mathbf{A}_1 \bar{\mu}, \quad \bar{\mu}^T \mathbf{A}_1 \bar{\mu} = \bar{\mu}^T \bar{\mu}, \quad (53)$$

где $\bar{\mu}^T = (E(\sqrt{\beta} \eta_1), \dots, E(\sqrt{\beta} \eta_N))$. (Мы полагаем, что $\bar{\mu}$ – вектор-столбец, значок “ T ” означает транспонирование.) Кроме того, число степеней свободы указанного распределения хи-квадрат есть след $\text{tr}(\mathbf{A}_1)$ матрицы $\mathbf{A}_1$ и его параметр нецентральности равен $\bar{\mu}^T \mathbf{A}_1 \bar{\mu}$.

Найдем след матрицы $\mathbf{A}_1$. Учитывая (51), (52) и равенство $\sum_{i \in M} q_i = 1$, получаем

$$\text{tr}(\mathbf{A}_1) = \sum_{i \in M} \left(\sqrt{\frac{\beta q_i}{p_i}} - q_i \right) = \sum_{i \in M} (1 - q_i) = |M| - 1.$$

Если в испытаниях (1) справедлива альтернатива $\mathbf{H}(p)$, то $\bar{\mu} = \bar{0}$, второе и третье из равенств (53) выполняются тривиальным образом, а первое из этих равенств вытекает из идемпотентности $\mathbf{A}_1$. Таким образом, для этого случая утверждение теоремы 3 верно.

Предположим теперь, что в испытаниях (1) выполняется альтернатива $\mathbf{H}(\tilde{p})$, указанная в формулировке теоремы 3. Идемпотентность $\mathbf{A}_1$ имеет следствием первое

и второе из равенств (53), и нам остается лишь проверить справедливость третьего из этих равенств. Согласно (27) и определению вектора $\bar{\mu}$ имеем

$$\bar{\mu}^T = \left(\frac{\sqrt{\beta}d_1}{\sqrt{p_1}}, \dots, \frac{\sqrt{\beta}d_N}{\sqrt{p_N}} \right).$$

Поэтому, снова используя равенства (51), (52) и учитывая, что для рассматриваемых в теореме альтернатив $d_i = 0$, если $i \in \bar{M}$, и $\sum_{i \in M} d_i = 0$, получаем

$$\begin{aligned} \bar{\mu}^T \mathbf{A}_1 \bar{\mu} &= \sum_{j \in M} \sum_{i \in M} \sqrt{\beta} \frac{d_i}{\sqrt{p_i}} \left[\sqrt{\frac{\beta q_i}{p_j}} \delta_{ij} - \sqrt{q_i q_j} \right] \sqrt{\beta} \frac{d_j}{\sqrt{p_j}} \\ &= \sum_{j \in M} \beta \frac{d_j^2}{p_j} - \left(\sum_{i \in M} d_i \right)^2 = \sum_{j \in M} \beta \frac{d_j^2}{p_j} = \bar{\mu}^T \bar{\mu}. \end{aligned}$$

Теорема доказана.

СПИСОК ЦИТИРОВАННОЙ ЛИТЕРАТУРЫ

- [1] Г. Крамер, *Математические методы статистики*, ГИИЛ, М., 1948.
- [2] Э. Леман, *Проверка статистических гипотез*, Наука, М., 1964.
- [3] J. D. Broffitt, R. H. Randles, "A power approximation for the chi-square goodness-of-fit test: simple hypothesis case", *J. Amer. Statist. Assoc.*, **72**:359 (1977), 604–607.
- [4] Б. И. Селиванов, "О предельных распределениях статистики χ^2 ", *Теория вероятн. и ее примен.*, **29**:1 (1984), 132–134.
- [5] С. Р. Рао, *Линейные статистические методы и их применения*, Наука, М., 1968.
- [6] М. Кендалл, Ф. Стьюарт, *Статистические выводы и связи*, Наука, М., 1973.
- [7] B. Györfi, "On the asymptotic behaviour of the generalized multinomial distribution", *Recent Developments in Statistics* (Proc. European Meeting Statisticians, Grenoble, 1976), North-Holland, Amsterdam, 1977, 461–470; *Publ. Math. Debrecen*, **24**:1–2 (1977), 163–171.
- [8] H. Chernoff, E. L. Lehmann, "The use of maximum likelihood estimates in χ^2 tests for goodness of fit", *Ann. Math. Statist.*, **25**:3 (1954), 579–586; рус. пер.: Г. Чернов, Е. Л. Леман, "Использование оценок максимума правдоподобия в критериях согласия хи-квадрат", *Математика. Сб. пер.*, **7**:6 (1963), 131–138.
- [9] J. J. Dik, M. C. M. de Gunst, "The distribution of general quadratic form in normal variables", *Statist. Neerlandica*, **39**:1 (1985), 14–26.
- [10] Б. И. Селиванов, "Об одном классе статистик типа хи-квадрат", *Обзорные приклад. и промышл. матем.*, **2**:6 (1995), 926–966.
- [11] Г. М. Фихтенгольц, *Курс дифференциального и интегрального исчисления*, т. 1, ГИТТЛ, М.–Л., 1951.

Б. И. Селиванов

Математический институт им. В. А. Стеклова РАН, Москва

Поступило

23.06.2006

Исправленный вариант

20.11.2006