Math-Net.Ru

Общероссийский математический портал

Б. И. Селиванов, Семейство многомерных статистик типа хи-квадрат, *Дискрет. матем.*, 2004, том 16, выпуск 4, 110–116

DOI: <https://doi.org/10.4213/dm179>

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: 54.224.187.69

26 апреля 2023 г., 12:49:08

Семейство многомерных статистик типа хи-квадрат

© 2004 г. Б. И. Селиванов

Рассматривается последовательность независимых испытаний, каждое из которых согласно гипотезе H_0 проводится по одной и той же полиномиальной схеме. Вводится семейство многомерных статистик типа хи-квадрат. При этом исходы снабжаются весами, зависящими от номера испытания. На это семейство переносятся результаты, полученные ранее для семейства многомерных статистик хи-квадрат.

Работа выполнена при поддержке грантом НШ-1758.2003.1 Президента Российской Федерации поддержки ведущих научных школ.

1. Введение

Пусть дана последовательность независимых случайных величин

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

значениями которых являются элементы множества $\{1, \dots, N\}$, предполагается, что N фиксировано. В дальнейшем, рассматривая случайные величины (1), будем использовать терминологию А. Н. Колмогорова [1] и говорить о независимых испытаниях с возможными исходами $1, \dots, N$. Согласно нулевой гипотезе H_0 каждое из испытаний (1) проводится по одной и той же полиномиальной схеме с вероятностями исходов, равными соответственно $p_1, \dots, p_N$, где $0 < p_i < 1$, $i = 1, \dots, N$, $p_1 + \dots + p_N = 1$.

Зададим r последовательностей действительных чисел

$$a_k(1), a_k(2), \dots, a_k(n), \dots, \quad k = 1, \dots, r. \quad (2)$$

Число $a_k(t)$ назовем k -м весом исходов в испытании t . Введем случайные величины

$$R_{ki}(n) = \sum_{t=1}^n a_k(t) I\{X_t = i\}, \quad k = 1, \dots, r, \quad i = 1, \dots, N,$$

где $I\{X_t = i\}$ — индикатор события $\{X_t = i\}$, и

$$\xi_{ki} = \frac{1}{\sqrt{A_k^{(2)}(n)p_i}} (R_{ki}(n) - A_k^{(1)}(n)p_i), \quad k = 1, \dots, r, \quad i = 1, \dots, N, \quad (3)$$

где

$$A_k^{(1)}(n) = \sum_{t=1}^n a_k(t), \quad A_k^{(2)}(n) = \sum_{t=1}^N a_k^2(t). \quad (4)$$

С использованием случайных величин (3) определим r -мерную статистику типа хи-квадрат

$$\chi^2 = (\chi_1^2, \dots, \chi_r^2) \quad (5)$$

с компонентами

$$\chi_k^2 = \sum_{i=1}^N \xi_{ki}^2 = \sum_{i=1}^N \frac{1}{A_k^{(2)}(n)p_i} (R_{ki}(n) - A_k^{(1)}(n)p_i)^2, \quad k = 1, \dots, r, \quad (6)$$

зависящими от выбора весов (2). Задавая различным образом последовательности (2), получим семейство r -мерных статистик типа хи-квадрат.

Построенное семейство включает в себя семейство r -мерных статистик хи-квадрат, определенное автором в [2]. Действительно, пусть $A_1, \dots, A_r$ — конечные непустые множества натуральных чисел. Определим веса исходов, для $k = 1, \dots, r$ и $t = 1, \dots, n$ положив $a_k(t) = 1$, если $t \in A_k$, и $a_k(t) = 0$ в противном случае. Тогда статистика (5) с компонентами (6) есть как раз r -мерная статистика χ^2 из [2], примерами которой являются статистики из статей [3] и [4].

В настоящей работе результаты, полученные в [2] для обычных статистик хи-квадрат, переносятся на обобщенные статистики типа хи-квадрат (относительно одномерных статистик типа хи-квадрат см. [5]). В разделе 2 приведены вспомогательные результаты, используемые при доказательстве двух предельных теорем для статистик (5), которые сформулированы в разделе 3. В них для статистик (5) утверждается существование предельных распределений как при гипотезе H_0 , так и при альтернативах, континуальных к H_0 . Получены также представления для преобразования Лапласа предельных распределений, принадлежащих классу многомерных распределений хи-квадрат, определение которого можно найти в [2]. Техника, использованная в [2], работает и в случае, рассматриваемом здесь, что позволило нам ограничиться лишь формулировками полученных результатов. К сожалению, многочисленные ошибки, появившиеся в процессе редактирования и компьютерного набора, затрудняют понимание работы [2].

2. Вспомогательные результаты

Рассмотрим последовательность векторов

$$\mathbf{p} = (\mathbf{p}_1, \dots, \mathbf{p}_n), \quad (7)$$

где для $t = 1, \dots, n$

$$\mathbf{p}_t = (p_1(t), \dots, p_N(t)), \quad 0 \leq p_i(t) < 1, \quad i = 1, \dots, N, \quad \sum_{i=1}^N p_i(t) = 1. \quad (8)$$

Будем говорить, что в испытаниях (1) выполняется (простая) гипотеза $H(\mathbf{p})$, если в t -м из них, $t = 1, \dots, n$, реализуется полиномиальная схема (8). В случае $\mathbf{p}_1 = \dots = \mathbf{p}_n = \mathbf{p}$,

где $p = (p_1, \dots, p_N)$, выполняется гипотеза H_0 . Если же в (7) $p_t \neq p$ для некоторого $t = 1, \dots, n$, то гипотезу $H(p)$ будем называть альтернативой к H_0 .

Следующие три леммы обобщают соответствующие леммы 1, 2 и 3 из [2]. Вначале с использованием определенных в разделе 1 случайных величин $R_{ki}(n)$ доказывается следующее утверждение.

Лемма 1. Если испытания (1) проведены при гипотезе $H(p)$, задаваемой последовательностью (7) полиномиальных схем (8), то для случайных величин (3) выполняются равенства

$$\mathbf{E} \xi_{ki} = \frac{1}{\sqrt{A_k^{(2)}(n) p_i}} \sum_{t=1}^n a_k(t)(p_i(t) - p_i), \quad (9)$$

$$\text{cov}(\xi_{ki}, \xi_{lj}) = \frac{\sum_{t=1}^n a_k(t) a_l(t) \sqrt{p_i(t) p_j(t)}}{\sqrt{A_k^{(2)}(n) A_l^{(2)}(n) p_i p_j}} (\delta_{ij} - \sqrt{p_i(t) p_j(t)}), \quad (10)$$

где δ_{ij} — символ Кронекера, $k, l = 1, \dots, r$, $i, j = 1, \dots, N$.

Следствие 1. Если верна гипотеза H_0 , то имеют место равенства

$$\mathbf{E} \xi_{ki} = 0, \quad \text{cov}(\xi_{ki}, \xi_{lj}) = \Delta_{kl}(n)(\delta_{ij} - \sqrt{p_i p_j}), \quad (11)$$

где

$$\Delta_{kl}(n) = \frac{\sum_{t=1}^n a_k(t) a_l(t)}{\sqrt{A_k^{(2)}(n) A_l^{(2)}(n)}}, \quad k, l = 1, \dots, r, \quad i, j = 1, \dots, N. \quad (12)$$

Частными случаями формул (9)–(12) являются соответствующие утверждения леммы 1 и следствия 1 из [2].

Будем предполагать, что длина n последовательности (1) неограниченно растет. Последовательность $\{a(t)\}_{t=1}^n$ действительных чисел, как и в [5], назовем регулярной, если при $n \rightarrow \infty$ выполняется соотношение

$$\max_{1 \leq t \leq n} a^2(t) \bigg/ \sum_{t=1}^n a^2(t) = o(1). \quad (13)$$

Частным случаем этого условия является условие (F_1) из [2].

Прежде чем сформулировать следующую лемму введем rN -мерный вектор

$$\xi = (\xi_{11}, \dots, \xi_{1N}; \dots; \xi_{r1}, \dots, \xi_{rN}), \quad (14)$$

компонентами которого являются случайные величины (3).

Лемма 2. Предположим, что испытания (1) проводятся при гипотезе H_0 , все r последовательностей весов исходов (2) регулярны и величины (12) имеют пределы

$$\lim_{n \rightarrow \infty} \Delta_{kl}(n) = \rho_{kl}, \quad k, l = 1, \dots, r. \quad (15)$$

Тогда случайный вектор (14) при $n \rightarrow \infty$ сходится по распределению к rN -мерному нормальному случайному вектору

$$\xi^{(1)} = (\xi_{11}^{(1)}, \dots, \xi_{1N}^{(1)}; \dots; \xi_{r1}^{(1)}, \dots, \xi_{rN}^{(1)}), \quad (16)$$

с нулевым вектором средних значений и ковариационной матрицей $D = \|d_{ki,lj}\|$ порядка rN , элементы которой определяются равенствами

$$d_{ki,lj} = \rho_{kl}(\delta_{ij} - \sqrt{p_i p_j}), \quad k, l = 1, \dots, r, \quad i, j = 1, \dots, N.$$

Настоящая лемма представляет собой обобщение леммы 2 из [2] и теоремы 3.2 из [5]. Ее доказательство, как и указанных утверждений из [2] и [5], непосредственно вытекает из многомерной центральной предельной теоремы. Вместо леммы 2 можно было бы сформулировать более сильное утверждение, аналогичное лемме 3.3 из [5]. Заметим также, что условие (F_2) из [2], представляет собой частный случай условия (15).

В дальнейшем в качестве альтернатив $H(\mathbf{p})$ нулевой гипотезе, как и в [2], будем рассматривать лишь те, которые контигуальны H_0 в смысле Ле Кама ([6], гл. 6). Известно (см. [5]), что альтернатива $H(\mathbf{p})$ будет контигуальна H_0 тогда и только тогда, когда выполняется условие

$$\lim_{n \rightarrow \infty} \sup \sum_{t=1}^n \sum_{i=1}^N \frac{(p_i(t) - p_i)^2}{p_i} < \infty. \quad (17)$$

В [2] это есть условие (F_3) .

Следующая лемма является обобщением леммы 3 из [2], и доказывается она аналогичным образом.

Лемма 3. *Предположим, что испытания (1) проводятся в предположении, что верна некоторая альтернатива $H(\mathbf{p})$ гипотезе H_0 , все r последовательностей весов исходов $\{a(t)\}_{t=1}^n$ регулярны и выполняются условия (15), (17). Тогда при $n \rightarrow \infty$ случайный rN -мерный вектор*

$$\xi^{(2)} = (\xi_{11}^{(2)}, \dots, \xi_{1N}^{(2)}; \dots; \xi_{r1}^{(2)}, \dots, \xi_{rN}^{(2)}),$$

с компонентами

$$\xi_{kl}^{(2)} = \frac{1}{\sqrt{A_k^{(2)}(n)p_i}} (R_{ki}(n) - \sum_{t=1}^n a_{ki}(t)p_i(t)), \quad k = 1, \dots, r, \quad i = 1, \dots, N,$$

сходится по распределению к нормальному случайному вектору (16), определенному в формулировке леммы 2.

3. Предельные теоремы

Введем несколько обозначений. Пусть I — единичная матрица порядка r , $T = \text{diag}\{t_1, \dots, t_r\}$ и $R = \|\rho_{kl}\|_{k,l=1}^r$ — матрица порядка r , элементы которой задаются соотношениями (15). Определитель произвольной матрицы B будем обозначать $|B|$, штрих у вектора будет означать его транспонирование.

Выделим в отдельную лемму следующее утверждение, которое использовалось в [3], [4] и [2].

Лемма 4. *Рассмотрим случайные векторы*

$$\eta_i = (\eta_{i1}, \dots, \eta_{ir}), \quad i = 1, \dots, N,$$

где

$$\eta_{ik} = \sum_{j=1}^N c_{ij} \xi_{kj}^{(1)}, \quad k = 1, \dots, r,$$

$\xi_{kj}^{(1)}$ — компоненты вектора (16), c_{ij} — элементы ортогональной матрицы $C = \|c_{ij}\|_{i,j=1}^N$ порядка N , элементы последней строки которой имеют вид $c_{Nj} = \sqrt{p_i}$, $j = 1, \dots, N$. Тогда выполняются равенства

$$\text{cov}(\eta_{ik}, \eta_{jl}) = \rho_{kl}(\delta_{ij} - \delta_{iN}\delta_{jN}), \quad k, l = 1, \dots, r, \quad i, j = 1, \dots, N,$$

в которых величины ρ_{kl} определяются соотношениями (15).

Следующая теорема является естественным обобщением теоремы 1 из [2].

Теорема 1. Пусть испытания (1) проводятся в предположении, что верна гипотеза H_0 , последовательности весов (2) регулярны и выполняются условия (15) и (17). Тогда при $n \rightarrow \infty$ статистика (5) имеет предельное распределение, а его преобразование Лапласа есть

$$L_r(t_1, \dots, t_r) = |I + 2TR|^{-(N-1)/2}, \quad (18)$$

где $t_k \geq 0$, $k = 1, \dots, r$.

Доказательство теоремы 1 аналогично доказательству теоремы 1 из [2] и основано на леммах 2 и 4, приведенных выше, и лемме 4 из [2].

Теорема 2. Пусть для каждого $n = 1, 2, \dots$ испытания (1) проводятся в предположении, что справедлива альтернатива $H(\mathbf{p}) = H(\mathbf{p}_n)$ (элементы последовательности (7) зависят от n), выполнены условия (15), (17) и условие регулярности (13) последовательностей (2). Предположим, что существуют пределы

$$\lim_{n \rightarrow \infty} \sum_{i=1}^N \lambda_{ki}(n) \lambda_{lj}(n) = \alpha_{kl}, \quad k, l = 1, \dots, r, \quad k \leq l, \quad (19)$$

где

$$\lambda_{ki}(n) = \frac{1}{\sqrt{A_k^{(2)}(n)}} \sum_{t=1}^n a_k(t)(p_i(t) - p_i), \quad k = 1, \dots, r, \quad i = 1, \dots, N. \quad (20)$$

Тогда при $n \rightarrow \infty$ статистика (5) сходится по распределению, и преобразование Лапласа ее предельного распределения есть

$$L_r(t_1, \dots, t_r) = |I + 2TR|^{-(N-1)/2} \exp \left\{ - \sum_{k,l=1}^r \alpha_{kl} q_{kl} \right\}. \quad (21)$$

Здесь $t_k \geq 0$, $k = 1, \dots, r$, $\alpha_{lk} = \alpha_{kl}$ для $k, l = 1, \dots, r$, $k < l$, и q_{kl} — элементы матрицы

$$Q = \|q_{kl}\|_{k,l=1}^r = (I + 2TR)^{-1} T. \quad (22)$$

Доказательство теоремы 2 проводится по той же схеме, что и теоремы 2 в [2]. При этом используются условия регулярности последовательностей (2), леммы 3 и 4, сформулированные выше, а также лемма 4 из [2].

Из (18) и леммы 4 из [2] следует, что предельное распределение статистики (5) в условиях теоремы 1 есть многомерное центральное распределение хи-квадрат с параметрами r , $N - 1$ и ρ_{kl} , $1 \leq k \leq l \leq r$. Если же не все величины α_{kl} , определяемые соотношениями (19), равны нулю, то в условиях теоремы 2 указанное предельное распределение есть многомерное нецентральное распределение хи-квадрат, имеющее параметры r , $N - 1$ и ρ_{kl} , α_{kl} , $1 \leq k \leq l \leq r$.

Укажем два случая, когда выполняется условие (19) теоремы 2, а утверждение этой теоремы несколько упрощается.

Предположим, что для полиномиальных схем (7), определяющих альтернативу $H(p)$, выполняются равенства $\mathbf{p}_t = \mathbf{q}$, $t = 1, \dots, n$, где

$$\mathbf{q} = \mathbf{q}(n) = (q_1, \dots, q_N), \quad 0 \leq q_i \leq 1, \quad i = 1, \dots, N, \quad \sum_{i=1}^N q_i = 1.$$

Соответствующую альтернативу гипотезе H_0 обозначим H_q .

Следствие 2. Пусть испытания (1) проводятся в предположении, что верна альтернатива H_q гипотезе H_0 , последовательности (2) регулярны, выполнено условие (15) и существуют пределы

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{[A_k^{(1)}(n)]^2}{n A_k^{(2)}(n)} &= \beta_k^2, \quad k = 1, \dots, r, \\ \lim_{n \rightarrow \infty} n \sum_{i=1}^N \frac{(q_i - p_i)^2}{p_i} &= \gamma^2 < \infty. \end{aligned} \quad (23)$$

где величины $A_k^{(1)}(n)$ и $A_k^{(2)}(n)$ определяются равенствами (4).

Тогда при $n \rightarrow \infty$ статистика (5) имеет предельное распределение, преобразование Лапласа которого есть

$$L_r(t_1, \dots, t_r) = |I + 2TR| \exp\{-\gamma^2 b Q b'\},$$

где $t_k \geq 0$, $k = 1, \dots, r$, $b = (\beta_1, \dots, \beta_r)$, Q — матрица (22) и матрица R определена в начале раздела 3.

Следствие 3. Пусть выполнены предположения теоремы 2, но вместо пределов (19) существуют пределы (23) и

$$\lim_{n \rightarrow \infty} \frac{\sqrt{n A_k^{(2)}(n)}}{A_k^{(1)}(n)} \lambda_{ki}(n) = \lim_{n \rightarrow \infty} \frac{\sqrt{n}}{A_k^{(1)}(n)} \sum_{t=1}^n a_k(t) (p_i(t) - p_i) = \gamma_{ki},$$

где $A_k^{(1)}(n)$ и $A_k^{(2)}(n)$ заданы формулами (4), а величины $\lambda_{ki}(n)$ определены равенствами (20).

Тогда при $n \rightarrow \infty$ статистика (5) имеет предельное распределение, а его преобразование Лапласа есть

$$\begin{aligned} L_r(t_1, \dots, t_r) &= |I + 2TTR|^{-(N-1)/2} \exp \left\{ - \sum_{i=1}^N g_i Q g_i' \right\} \\ &= |I + 2TR|^{-(N-1)/2} \exp \left\{ - \sum_{k,l=1}^r \left(\sum_{i=1}^N \gamma_{ki} \gamma_{li} \right) \beta_k \beta_l q_{kl} \right\}, \end{aligned}$$

где $t_k \geq 0$, $k = 1, \dots, r$, $g_i = (\beta_1 \gamma_{1i}, \dots, \beta_r \gamma_{ri})$, $i = 1, \dots, N$, и $Q = \|q_{kl}\|_{k,l=1}^r$ — матрица (22).

Список литературы

1. Колмогоров А. Н., *Основные понятия теории вероятностей*. Наука, Москва, 1974.
2. Селиванов Б. И., Семейство многомерных статистик хи-квадрат. *Дискретная математика* (2002) **14**, №3, 130–142.
3. Захаров В. К., Сарманов О. В., Севастьянов Б. А., Последовательный критерий χ^2 . *Матем. сб.* (1969) **79**, №3, 444–460.
4. Тихомирова М. И., Чистяков В. П., Скользящий хи-квадрат. *Дискретная математика* (2000) **12**, №4, 46–52.
5. Селиванов Б. И., Об одном классе статистик типа хи-квадрат. *Обзорные прикладной и промышленной матем.* (1995) **2**, №6, 926–966.
6. Гаек Я., Шидак З., *Теория ранговых критериев*. Наука, Москва, 1971.

Статья поступила 12.11.2003.