Math-Net.Ru

Общероссийский математический портал

Б. И. Селиванов, В. П. Чистяков, Два замечания о многомерной статистике χ^2 , *Дискрет. матем.*, 1999, том 11, выпуск 4, 145–151

DOI: <https://doi.org/10.4213/dm404>

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: 52.90.164.192

26 апреля 2023 г., 14:33:23

УДК 519.2

Два замечания о многомерной статистике χ^2

© 1999 г. Б. И. Селиванов, В. П. Чистяков

Рассматривается последовательность полиномиальных испытаний с N исходами. По ней строится многомерная статистика χ^2 . При этом используются выборки нарастающих объемов $n_1, \dots, n_r$, $1 \leq n_1 < \dots < n_r$, $r \geq 2$, и каждая последующая выборка содержит предыдущую. Предполагается, как и в [1], что N постоянно, $n_i \rightarrow \infty$ и $n_i/n_{i+1} \rightarrow \rho_i^2$, $0 < \rho_i < 1$, $i = 1, \dots, r-1$.

Для фиксированных (неблизких) альтернатив к проверяемой простой гипотезе доказана слабая сходимость распределения векторной статистики χ^2 , компоненты которой подходящим образом центрируются и нормируются, к многомерным нормальному и хи-квадрат распределениям. Показано также, что в случае нормальной сходимости компоненты предельного нормального случайного вектора образуют неоднородную цепь Маркова и найдены плотности вероятностей перехода этой цепи.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проекты 96-01-00531, 96-15-96092.

1. Рассматривается последовательность полиномиальных испытаний с множеством возможных исходов $I = \{1, \dots, N\}$, $N \geq 2$. Проверяется гипотеза H_0 , согласно которой вероятность реализации исхода j равна p_j и

$$0 < p_j < 1, \quad j = 1, \dots, N, \quad \sum_{j=1}^N p_j = 1. \quad (1)$$

Альтернативы (простые конкурирующие гипотезы) H к гипотезе H_0 будем задавать векторами вероятностей исходов $\pi = (\pi_1, \dots, \pi_N)$,

$$0 \leq \pi_j < 1, \quad j = 1, \dots, N, \quad \sum_{j=1}^N \pi_j = 1, \quad (2)$$

отличными от вектора $(p_1, \dots, p_N)$, так что всегда $H \neq H_0$. Согласно (2) из рассмотрения исключаются тривиальные альтернативы, при которых не равна нулю только одна из вероятностей π_j , $j = 1, \dots, N$.

Пусть $1 \leq n_1 < \dots < n_r$, $r \geq 2$, — нарастающие объемы выборок из исходной последовательности испытаний, которая получена или при гипотезе H_0 , или при некоторой альтернативе H . При этом каждая последующая выборка содержит предыдущую. Обозначив ν_{ij} частоту исхода j в i -й выборке, $i = 1, \dots, r$, $j = 1, \dots, N$,

положим

$$\chi_i^2 = \sum_{j=1}^N \frac{1}{n_i p_j} (\nu_{ij} - n_i p_j)^2, \quad i = 1, \dots, r, \quad (3)$$

и $\chi^2 = (\chi_1^2, \dots, \chi_r^2)$.

В настоящей работе, как и в статье [1], будем предполагать, что выполняются условия: N фиксировано,

$$n_1 \rightarrow \infty, \quad \frac{n_i}{n_{i+1}} \rightarrow \rho_i^2, \quad 0 < \rho_i < 1, \quad i = 1, \dots, r-1. \quad (4)$$

В [1] изучалось предельное распределение случайного вектора χ^2 в случае, когда справедливы условия (4) и испытания проводятся или при гипотезе H_0 , или при альтернативах, сближающихся с H_0 . В данной работе, дополняющей результаты, полученные в [1], указаны предельные распределения векторной статистики χ^2 для неблизких (фиксированных) альтернатив. В разделе 2 обобщаются на многомерный случай результаты работ [2, 3], относящиеся к обычной одномерной статистике χ^2 Пирсона (3). Доказана слабая сходимость распределения векторной статистики χ^2 , компоненты которой подходящим образом центрированы и нормированы, к многомерным нормальному и хи-квадрат распределениям. В разделе 3 показано, что в случае нормальной сходимости компоненты предельного нормального вектора образуют цепь Маркова и найдены плотности вероятностей перехода этой цепи.

2. Нам потребуется введенное в [3] конечное множество альтернатив к гипотезе H_0 , определяемых следующим образом. Выберем подмножество

$$A \subset I, \quad 2 \leq |A| < N$$

($|A|$ — число элементов A), и положим $\pi_j = 0$, если $j \in I \setminus A$, и $\pi_j = \alpha p_j$, если $j \in A$, где

$$\alpha = \left(\sum_{j \in A} p_j \right)^{-1}, \quad (5)$$

и ввиду условий (1) α вполне определено. Альтернативы этого вида будем обозначать H_A , имея в виду, что A — собственное подмножество I и $|A| > 1$.

Для произвольной альтернативы H с вероятностями исходов $\pi_1, \dots, \pi_N$ положим

$$a = \sum_{j=1}^N \frac{\pi_j^2}{p_j} - 1, \quad \sigma^2 = \sum_{j=1}^N \frac{\pi_j^3}{p_j^2} - \left(\sum_{j=1}^N \frac{\pi_j^2}{p_j} \right)^2. \quad (6)$$

Ясно, что $\sigma^2 \geq 0$ для всех рассматриваемых альтернатив к гипотезе H_0 и, как замечено в [3], равенство $\sigma^2 = 0$ имеет место лишь для альтернатив H_A , определенных выше.

Если $H \neq H_A$, то из известных выражений для математического ожидания и дисперсии статистики χ^2 Пирсона, которые цитируются в [2, 3], следует, что выполняются соотношения

$$\begin{aligned} \mathbf{M}(\chi_i^2 | H) &= an_i \left(1 + O\left(\frac{1}{n_i}\right) \right), \\ \mathbf{D}(\chi_i^2 | H) &= 4\sigma^2 n_i \left(1 + O\left(\frac{1}{n_i}\right) \right), \end{aligned}$$

где $i = 1, \dots, r$ и a, σ^2 определяются формулами (6). Если же $H = H_A$, то согласно [3]

$$\begin{aligned} \mathbf{M}(\chi_i^2 | H_A) &= (\alpha - 1)n_i \left(1 + O\left(\frac{1}{n_i}\right) \right), \\ \mathbf{D}(\chi_i^2 | H_A) &= 2\alpha^2(|A| - 1) \left(1 + O\left(\frac{1}{n_i}\right) \right), \end{aligned}$$

где $i = 1, \dots, r$ и α задается формулой (5).

Цель этого раздела — доказать следующее утверждение.

Теорема 1. *Предположим, что выполнены условия (4) и полиномиальные испытания проводятся при некоторой фиксированной альтернативе H . Тогда, если $H \neq H_A$, распределение случайного вектора*

$$\tilde{\eta} = (\tilde{\eta}_1, \dots, \tilde{\eta}_r), \quad \tilde{\eta}_i = \frac{\chi_i^2 - an_i}{2\sqrt{n_i}}, \quad i = 1, \dots, r, \quad (7)$$

слабо сходится к r -мерному невырожденному нормальному распределению с нулевым вектором математических ожиданий и ковариационной матрицей $B = \|\sigma^2 \rho_{ij}\|_{i,j=1}^r$, где

$$\rho_{ij} = \rho_{ji} = \begin{cases} 1, & \text{если } i = j = 1, \dots, r, \\ \rho_i \rho_{i+1} \dots \rho_{j-1}, & \text{если } i \neq j, 1 \leq i < j \leq r. \end{cases} \quad (8)$$

Если же $H = H_A$, то распределение r -мерной статистики

$$\eta^* = (\eta_1^*, \dots, \eta_r^*), \quad \eta_i^* = \frac{1}{\alpha}(\chi_i^2 - (\alpha - 1)n_i), \quad i = 1, \dots, r, \quad (9)$$

слабо сходится к многомерному центральному распределению хи-квадрат с параметрами $r, t = |A|$, определенному в [1].

Доказательство. Пусть верна альтернатива $H \neq H_A$. Из (1) и (7) находим, что

$$\tilde{\eta}_i = \frac{1}{2\sqrt{n_i}} \sum_{j=1}^N \frac{\pi_j}{p_j} \xi_{ij}^2 + \sum_{j=1}^N (\pi_j - p_j) \frac{\sqrt{\pi_i}}{p_j} \xi_{ij}, \quad i = 1, \dots, r, \quad (10)$$

где

$$\xi_{ij} = \frac{\nu_{ij} - n_i \pi_j}{\sqrt{n_i \pi_j}}, \quad i = 1, \dots, r, \quad j = 1, \dots, N.$$

Если выполняются условия (4), то согласно [1] rN -мерный случайный вектор

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

асимптотически нормален с нулевым вектором математических ожиданий и ковариационной матрицей, элементы которой задаются равенствами

$$\text{cov}(\xi_{ih_1}, \xi_{jh_2}) = \rho_{ij} \sqrt{\frac{\pi_{h_1}}{\pi_{h_2}}} (\delta(h_1, h_2) - \pi_{h_2}),$$

где $\delta(h_1, h_2)$ — символ Кронекера, для величин ρ_{ij} имеют место равенства (8), $i, j = 1, \dots, r$, $h_1, h_2 = 1, \dots, N$. Поэтому в условиях (4) случайный вектор

$$\zeta_1 = \left(\sum_{j=1}^N \frac{1}{2\sqrt{n_1}} \frac{\pi_j}{p_j} \xi_{1j}^2, \dots, \sum_{j=1}^r \frac{1}{2\sqrt{n_r}} \frac{\pi_j}{p_j} \xi_{rj}^2 \right) \quad (11)$$

сходится по вероятности к нулю, а вектор

$$\zeta_2 = \left(\sum_{j=1}^N (\pi_j - p_j) \frac{\sqrt{\pi_j}}{p_j} \xi_{1j}, \dots, \sum_{j=1}^r (\pi_j - p_j) \frac{\sqrt{\pi_j}}{p_j} \xi_{rj} \right) \quad (12)$$

асимптотически нормален с математическими ожиданиями компонент, равными нулю, и ковариационной матрицей B , определенной в формулировке теоремы. Учитывая (10), (11) и (12), имеем равенство $\tilde{\eta} = \zeta_1 + \zeta_2$, откуда вытекает справедливость утверждения теоремы о нормальной сходимости вектора $\tilde{\eta}$. Известно (см. [1]), что определитель $|R|$ матрицы $R = \|\rho_{ij}\|_{i,j=1}^r$ равен

$$|R| = \prod_{i=1}^{r-1} (1 - \rho_i^2) \quad (13)$$

и в силу условий (4) положителен. Отсюда получаем, что и определитель $|B|$ положителен, вследствие чего предельное нормальное распределение вектора $\tilde{\eta}$ является невырожденным. Первое утверждение теоремы полностью доказано.

Пусть теперь $H = H_A$. В этом случае, как указано в [3], случайные величины η_i^* , определенные равенствами (9), приводятся к виду

$$\eta_i^* = \sum_{j \in A} \frac{\nu_{ij}^2}{n_i \pi_j} - n_i, \quad i = 1, \dots, r. \quad (14)$$

Из (14) следует, что к вектору η^* применима теорема 1 из работы [1], что завершает доказательство нашей теоремы.

Замечание. Преобразование Лапласа и плотность предельного многомерного распределения вектора (9) можно найти, используя теорему 1 и формулу (4.10) из [1]. При этом вероятности p_j следует заменить на αp_j , $j \in A$.

3. Обозначим $\eta = (\eta_1, \dots, \eta_r)$ нормально распределенный случайный вектор с нулевыми математическими ожиданиями компонент и ковариационной матрицей B , вид которой указан в формулировке теоремы 1. В работе [1] показано, что компоненты многомерного нормального распределения хи-квадрат, определенного там же, образуют неоднородную цепь Маркова. Это утверждение относится и к частному случаю указанного распределения, о котором идет речь во второй части теоремы 1. Покажем, что аналогичным свойством обладает и случайный вектор η .

Теорема 2. Компоненты случайного вектора $\eta = (\eta_1, \dots, \eta_r)$ образуют цепь Маркова и их совместная плотность распределения вероятностей может быть представлена в виде

$$p_\eta(x_1, \dots, x_r) = p^{(0)}(x_1) \prod_{k=2}^r p_{x_{k-1}}^{(k-1)}(x_k). \quad (15)$$

Здесь

$$p^{(0)}(x_1) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{x_1^2}{2\sigma^2} \right\} \quad (16)$$

и для $k = 2, \dots, r$

$$p_{x_{k-1}}^{(k-1)}(x_k) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp \left\{ -\frac{(x_k - b_k)^2}{2\sigma_k^2} \right\}, \quad (17)$$

где

$$b_k = \rho_{k-1}x_{k-1}, \quad \sigma_k^2 = \sigma^2(1 - \rho_{k-1}^2), \quad (18)$$

величина σ^2 определена второй из формул (6).

Доказательство. Запишем плотность $p_r(x_1, \dots, x_r)$ случайного вектора η в виде

$$p_\eta(x_1, \dots, x_r) = p^{(0)}(x_1) \prod_{k=2}^r p^{(k-1)}(x_k | \eta_1 = x_1, \dots, \eta_{k-1} = x_{k-1}), \quad (19)$$

где $p^{(0)}(x_1)$ — плотность маргинального распределения первой компоненты вектора η и $p^{(k-1)}(x_k | \eta_1 = x_1, \dots, \eta_{k-1} = x_{k-1})$ — условная плотность распределения компоненты η_k при условии $\{\eta_1 = x_1, \dots, \eta_{k-1} = x_{k-1}\}$, $k = 2, \dots, r$. Равенство (16) достаточно очевидно. Для того, чтобы найти условную плотность $p^{(k-1)}(x_k | \eta_1 = x_1, \dots, \eta_{k-1} = x_{k-1})$, заметим вначале, что ковариационная матрица вектора $(\eta_1, \dots, \eta_k)$ есть

$$B_k = \|\sigma^2 \rho_{ij}\|_{i,j=1}^k, \quad k = 1, \dots, r,$$

и при этом $B_r = B$, $B_1 = \sigma^2$. Представим B_k для $k = 2, \dots, r$ в блочном виде

$$B_k = \begin{pmatrix} C_{11}^{(k)} & C_{12}^{(k)} \\ C_{21}^{(k)} & C_{22}^{(k)} \end{pmatrix},$$

где $C_{11}^{(k)}, C_{12}^{(k)}, C_{21}^{(k)}, C_{22}^{(k)}$ — блоки размеров $(k-1) \times (k-1), (k-1) \times 1, 1 \times (k-1)$ и 1×1 , соответственно. Тогда

$$C_{11}^{(k)} = B_{k-1}, \quad C_{22}^{(k)} = \sigma^2, \quad C_{12}^{(k)} = (C_{21}^{(k)})^\top \quad (20)$$

(значок $\top$ обозначает транспонирование) и ввиду (8)

$$C_{21}^{(k)} = (\sigma^2 \rho_{k1}, \dots, \sigma^2 \rho_{k,k-1}) = \rho_{k-1} (\sigma^2 \rho_{k-1,1}, \dots, \sigma^2) = \rho_{k-1} B_{k-1}^{(k-1)}, \quad (21)$$

где $B_{k-1}^{(k-1)}$ обозначает $(k-1)$ -ю строку матрицы B_{k-1} . (В формулах (20) и (21) везде $k = 2, \dots, r$.)

С помощью формулы для условной плотности многомерного нормального распределения (см., например, [4, стр. 139]) получаем, что

$$\begin{aligned} p^{(k-1)}(x_k | \eta_1 = x_1, \dots, \eta_{k-1} = x_{k-1}) \\ = \frac{1}{(\sqrt{2\pi})^k} \sqrt{\frac{|C_{11}^{(k)}|}{|B_k|}} \exp \left\{ -\frac{1}{2} (x_k - c_k) (C_{22}^{(k)} - C_{21}^{(k)} (C_{11}^{(k)})^{-1} C_{12}^{(k)})^{-1} (x_k - c_k) \right\}, \end{aligned} \quad (22)$$

и

$$c_k = C_{21}^{(k)} (C_{11}^{(k)})^{-1} (x_1, \dots, x_{k-1})^\top, \quad k = 2, \dots, r. \quad (23)$$

Аналогично равенству (13) вычислим определители

$$|B_k| = \sigma^{2k} \prod_{i=1}^{k-1} (1 - \rho_i^2), \quad k = 2, \dots, r, \quad (24)$$

и, учитывая первое из равенств (20), находим, что

$$|C_{11}^{(k)}| = \begin{cases} \sigma^2, & \text{если } k = 2, \\ \sigma^{2(k-1)} \prod_{i=1}^{k-1} (1 - \rho_i^2), & \text{если } k = 3, \dots, r. \end{cases} \quad (25)$$

Далее, используя (21) и первое из равенств (20), найдем величины (23):

$$c_2 = \rho_{k-1} B_{k-1}^{(k-1)} B_{k-1}^{-1} (x_1, \dots, x_{k-1})^\top = \rho_{k-1} x_{k-1}, \quad k = 2, \dots, r, \quad (26)$$

так как произведение $B_{k-1}^{(k-1)} B_{k-1}^{-1}$ равно $(k-1)$ -мерному вектору $(0, \dots, 0, 1)$. Аналогично, снова используя (8), (20) и (21), получаем, что

$$C_{21}^{(k)} (C_{11}^{(k)})^{-1} C_{12}^{(k)} = \sigma^2 \rho_{k-1} \rho_{k-1,k} = \sigma^2 \rho_{k-1}^2, \quad k = 2, \dots, r,$$

откуда следует, что

$$(C_{22}^{(k)} - C_{21}^{(k)} (C_{11}^{(k)})^{-1} C_{12}^{(k)})^{-1} = \frac{1}{\sigma^2 (1 - \rho_{k-1}^2)}, \quad k = 2, \dots, r. \quad (27)$$

Из (22), (24)–(27) вытекает справедливость формулы (17), в которой величины b_k и σ_k^2 имеют вид (18). Наконец, из (16), (17) и (19) следует равенство (15), что завершает доказательство теоремы.

Список литературы

1. Захаров В.К., Сарманов О.В., Севастьянов Б.А., Последовательный критерий χ^2 . *Матем. сб.* (1969) **79** (121), №3 (7), 444–460.
2. Broffitt J.D., Randles R.H., A power approximation for the chi-square goodness-of-fit: simple hypothesis case. *J. Amer. Statist. Assoc.* (1977) **72**, №359, 604–607.
3. Селиванов Б.И., О предельных распределениях статистики χ^2 . *Теория вероятностей и ее применения* (1984) **29**, №1, 132–134.
4. Королук В.С., Портенко Н.И., Скороход А.В., Турбин А.Ф., *Справочник по теории вероятностей и математической статистике*. Наука, Москва, 1985.

Статья поступила 21.01.1999.