Math-Net.Ru

Общероссийский математический портал

М. И. Тихомирова, В. П. Чистяков, О некоторых статистиках типа хи-квадрат, функционально зависящих от оценок неизвестных параметров, *Дискрет. матем.*, 2005, том 17, выпуск 1, 22–34

DOI: <https://doi.org/10.4213/dm85>

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: 54.84.234.179

26 апреля 2023 г., 16:57:05

О некоторых статистиках типа хи-квадрат, функционально зависящих от оценок неизвестных параметров

© 2005 г. М. И. Тихомирова, В. П. Чистяков

Известны теоремы о сходимости к предельному распределению распределений статистик типа χ^2 , построенных по частотам исходов в последовательности полиномиальных испытаний. Для несплошных цепочек исходов аналогичные результаты получены только для случая равновероятных исходов. В данной работе эти результаты переносятся на полиномиальные испытания с произвольными положительными вероятностями исходов; исследуются также статистики, получающиеся заменой вероятностей исходов в рассматривавшихся ранее статистиках типа χ^2 их оценками. Для сплошных цепочек получены предельные распределения этих статистик в общем случае, а для несплошных цепочек — в некоторых частных случаях. Приводится также предельная теорема в случае, когда исследуемая последовательность является цепью Маркова, сближающейся с полиномиальными испытаниями.

Работа выполнена при поддержке программой Президента Российской Федерации поддержки ведущих научных школ, грант НШ-1758.2003.1.

1. Введение

Пусть

$$x_1, x_2, \dots, x_t, \dots \quad (1)$$

— последовательность независимых случайных величин, для которых

$$\mathbf{P}\{x_t = k\} = p_k > 0, \quad k = 1, \dots, N, \quad \sum_{k=1}^N p_k = 1, \quad t = 1, 2, \dots$$

Положим

$$v_i^{(L)}(n) = \sum_{t=1}^n \chi(x_{t+r_1-1} = i_1, \dots, x_{t+r_L-1} = i_L), \quad L \geq 1. \quad (2)$$

где $\chi(C)$ — индикатор события C , r_l — последовательность целых чисел, удовлетворяющих условию

$$1 = r_1 < r_2 < \dots < r_L < r_{L+1} < \dots, \\ i = (i_1, \dots, i_L), \quad i_l \in \{1, \dots, N\}, \quad l = 1, \dots, L.$$

Случайные величины $v_i^{(L)}(n)$ являются частотами цепочек $i = (i_1, \dots, i_L)$ в выборке L -цепочек

$$(x_{t+r_1-1}, \dots, x_{t+r_L-1}), \quad t = 1, \dots, n. \quad (3)$$

Определим статистику $X_L(n)$ равенством

$$X_L(n) = \sum_i (\xi_i^{(L)}(n))^2, \quad i = (i_1, \dots, i_L), \quad (4)$$

где

$$\xi_i^{(L)}(n) = \frac{v_i^{(L)}(n) - na_i^{(L)}}{\sqrt{na_i^{(L)}}}, \quad a_i^{(L)} = p_{i_1} \dots p_{i_L}, \quad (5)$$

суммирование проводится по всем указанным значениям i . Далее, если не указаны пределы суммирования, будем также считать, что суммирование проводится по всем возможным значениям указанной переменной.

В случае

$$L = s, \quad r_{l+1} = r_l + 1, \quad l = 1, 2, \dots, \quad (6)$$

выборка (3) является выборкой перекрывающихся сплошных s -цепочек. Выборку (3) будем называть выборкой несплошных цепочек, если условие (6) не выполнено.

Из результатов В. Е. Степанова [1] следует, что распределение квадратичной формы

$$X(n) = X_s(n) - X_{s-1}(n) \quad (7)$$

при $n \rightarrow \infty$ в случае сплошных s -цепочек (условие (6)) сходится к центральному χ^2 -распределению с $N^s - N^{s-1}$ степенями свободы. Для несплошных цепочек аналогичное утверждение получено в [2], когда $p_k = 1/N$, $k = 1, \dots, N$. В данной работе это утверждение переносится на произвольные положительные вероятности $\{p_k\}$.

В обычной статистике χ^2 , построенной по частотам исходов полиномиальных испытаний, замена неизвестных параметров их оценками (см. [3]) приводит к тому, что предельное распределение остается χ^2 -распределением, но число степеней свободы уменьшается на число параметров, замененных оценками. В данной работе это утверждение переносится на статистики $\bar{X}(n)$, полученные из статистик (7) заменой вероятностей $\{p_k\}$ их оценками.

Критерии, основанные на статистике $\bar{X}(n)$, могут использоваться для проверки независимости близких соседних элементов последовательности. Использование этих статистик для проверки независимости далеких друг от друга членов последовательности затруднительно из-за необходимости использовать длинные цепочки, содержащие проверяемые на независимость члены последовательности. В таких случаях можно использовать статистики, построенные по частотам несплошных цепочек исходов. В данной работе получены утверждения о предельных распределениях этих статистик.

В работе рассматривается также случай, когда исходная последовательность является эргодической цепью Маркова, а при предельном переходе в схеме серий каждая строка матрицы вероятностей переходов сближается с вектором вероятностей значений величин x_t последовательности (1).

2. Полугруппа матриц

Рассматриваемые в данной работе статистики являются квадратичными формами от нормированных частот вида (5). Коэффициенты этих квадратичных форм образуют различные симметрические матрицы A . Распределение вектора нормированных частот при $n \rightarrow \infty$ сходится к нормальному распределению. Квадратичная форма с матрицей коэффициентов A от любого нормально распределенного вектора с ковариационной матрицей B и вектором математических ожиданий a имеет распределение χ^2 в том и только в том случае, когда выполняются условия (см. [4])

$$BABAB = BAB, \quad a'ABAa = a'Aa, \quad BABAa = BAa. \quad (8)$$

Число степеней свободы этого χ^2 -распределения равно следу матрицы AB , а параметр нецентральности равен $a'Aa$. При центрировке частот (5) вектор a равен 0.

В рассматриваемых далее случаях матрицы A и B являются линейными комбинациями $N^L \times N^L$ матриц из некоторого множества $D(p_1, \dots, p_N)$, где $p_l > 0$, $l = 1, \dots, N$, $p_1 + \dots + p_N = 1$.

Определим это множество. Назовем неполными подстановками степени L таблицы

$$\pi_k = \begin{pmatrix} i_1 & \dots & i_k \\ j_1 & \dots & j_k \end{pmatrix}, \quad 1 \leq k \leq L, \quad \pi_0 = \begin{pmatrix} \emptyset \\ \emptyset \end{pmatrix}, \quad (9)$$

образованные двумя упорядоченными подмножествами $\{i_1, \dots, i_k\}$, $i_1 < \dots < i_k$, и $\{j_1, \dots, j_k\}$ множества $\{1, 2, \dots, L\}$. Неполную подстановку π_k будем рассматривать также как взаимно однозначное отображение $\begin{pmatrix} i \\ j \end{pmatrix}$ верхней строки таблицы в нижнюю:

$$i_l \rightarrow j_l, \quad l = 1, \dots, k.$$

Условимся элемент любой матрицы C , стоящий в строке с номером u и в столбце с номером v , обозначать $(C)_{uv}$.

Множество $D_L(p_1, \dots, p_N)$ образовано матрицами $D_L(\pi_k)$ с элементами следующего вида:

$$(D_L(\pi_0))_{uv} = \sqrt{a_u^{(L)} a_v^{(L)}},$$

$$(D_L(\pi_k))_{uv} = \delta_{u_{i_1} v_{j_1}} \dots \delta_{u_{i_k} v_{j_k}} \sqrt{\frac{a_u^{(L)} a_v^{(L)}}{p_{u_{i_1}} \dots p_{u_{i_L}} p_{v_{j_1}} \dots p_{v_{j_L}}}}, \quad (10)$$

где $1 \leq k \leq L$, $u = (u_1, \dots, u_L)$, $v = (v_1, \dots, v_L)$, δ_{kl} — символ Кронекера, $a_i^{(L)}$ определены вторым равенством (5), координаты векторов u и v принимают значения из множества $\{1, \dots, N\}$.

Произведением $\begin{pmatrix} i \\ j \end{pmatrix} \begin{pmatrix} f \\ g \end{pmatrix}$ двух неполных подстановок $\begin{pmatrix} i \\ j \end{pmatrix} \begin{pmatrix} f \\ g \end{pmatrix}$, где $i = (i_1, \dots, i_k)$, $j = (j_1, \dots, j_k)$, $f = (f_1, \dots, f_l)$, $g = (g_1, \dots, g_l)$, назовем неполную подстановку, в которой верхняя строка является прообразом множества

$$j \cap f = (j_1, \dots, j_k) \cap (f_1, \dots, f_l)$$

при отображении $\begin{pmatrix} i \\ j \end{pmatrix}$, а нижняя строка — образом $j \cap f$ при отображении $\begin{pmatrix} f \\ g \end{pmatrix}$.

Например, при $L = 5$

$$\begin{pmatrix} 1 & 3 & 4 \\ 5 & 2 & 3 \end{pmatrix} \begin{pmatrix} 2 & 4 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 5 \\ 3 & 1 & 4 \end{pmatrix} \begin{pmatrix} 2 & 5 \\ 4 & 1 \end{pmatrix} = \begin{pmatrix} \emptyset \\ \emptyset \end{pmatrix}.$$

Единичной подстановкой e_L является подстановка

$$e_L = \begin{pmatrix} 1 & 2 & \dots & L \\ 1 & 2 & \dots & L \end{pmatrix}.$$

Для любой неполной подстановки π_k

$$e_L \pi_k = \pi_k e_L = \pi_k, \quad \pi_0 \pi_k = \pi_k \pi_0 = \pi_0.$$

Непосредственно из определения произведения матриц и приведенного определения произведения неполных подстановок нетрудно получить следующее утверждение.

Лемма 1. Для любых матриц $D_L(\pi_k)$ и $D_L(\tilde{\pi}_l)$ вида (10)

$$D_L(\pi_k) D_L(\tilde{\pi}_l) = D_L(\pi_k \tilde{\pi}_l).$$

Это утверждение будет использоваться при проверке условий (8).

Введем обозначения для двух часто встречающихся матриц:

$$F_L = D_L(\pi_0), \quad G_L = D_L \begin{pmatrix} 1 & 2 & \dots & L-1 \\ 1 & 2 & \dots & L-1 \end{pmatrix}. \quad (11)$$

Из леммы 1 и равенств (11) следует, что

$$\begin{aligned} E_L D_L(\pi_k) &= D_L(\pi_k) E_L = D_L(\pi_k), \\ F_L D_L(\pi_k) &= D_L(\pi_k) F_L = F_L, \end{aligned} \quad (12)$$

где E_L — единичная $N^L \times N^L$ матрица.

3. Квадратичные формы частот

Статистику $X(n) = X_M(n) - X_{M-1}(n)$, определенную равенствами (4) и (5), будем рассматривать как квадратичную форму от нормированных частот $\xi_u^{(M)}(n)$. Для исследования предельных распределений удобно представить $X(n)$ в виде

$$X(n) = \sum_{u,v} (A_M)_{uv} \xi_u^{(M)}(n) \xi_v^{(M)}(n), \quad (13)$$

где $u = (u_1, \dots, u_M)$, $v = (v_1, \dots, v_M)$, $u_l, v_l \in \{1, \dots, N\}$, $l = 1, \dots, M$, $\xi_u^{(M)}(n)$ определены равенством (5), A_M — некоторая симметричная $N^M \times N^M$ матрица, суммирование проводится по всем значениям векторов u и v , координаты которых принимают значения из множества $\{1, \dots, N\}$.

Лемма 2. Матрица A_M квадратичной формы (13) имеет вид

$$A_M = A_M^{(M)} - A_{M-1}^{(M)}, \quad (14)$$

где $A_M^{(M)} = E_M$ — единичная $N^L \times N^L$ матрица,

$$(A_{M-1}^{(M)})_{uv} = (G_M)_{uv} = \delta_{u_1 v_1} \dots \delta_{u_{M-1} v_{M-1}} \sqrt{p_{u_M} p_{v_M}},$$

δ_{kl} — символ Кронекера, матрица G_M определена равенством (11).

Доказательство. Случайную величину $\xi_i^{(M-1)}(n)$, определенную равенством (5), можно представить в виде

$$\xi_i^{(M-1)}(n) = \sum_{l=1}^N \xi_{(i,l)}^{(M)}(n) \sqrt{p_l}, \quad i = (i_1, \dots, i_{M-1}), \quad (i, l) = (i_1, \dots, i_{M-1}, l). \quad (15)$$

Используя это равенство, для квадратичной формы $X_{M-1}(n)$ получим выражение

$$\begin{aligned} X_{M-1}(n) &= \sum_i \sum_{l_1, l_2}^N \xi_{(i, l_1)}^{(M)}(n) \xi_{(i, l_2)}^{(M)}(n) \sqrt{p_{l_1} p_{l_2}} \\ &= \sum_{u, v} \delta_{u_1 v_1} \dots \delta_{u_{M-1} v_{M-1}} \sqrt{p_{u_M} p_{v_M}} \xi_u^{(M)}(n) \xi_v^{(M)}(n), \end{aligned}$$

где $i = (i_1, \dots, i_{M-1})$, $u = (u_1, \dots, u_M)$, $v = (v_1, \dots, v_M)$, следовательно,

$$X_{M-1}(n) = \sum_{u, v} (A_{M-1}^{(M)})_{uv} \xi_u^{(M)}(n) \xi_v^{(M)}(n).$$

Отсюда и из очевидного равенства $A_M^{(M)} = E_M$ следует утверждение леммы.

Определим множества $N_l(k)$, полагая

$$N_l(k) = \{i = (i_1, \dots, i_M) : i_m \in \{1, \dots, N\}, \quad m \neq l, \quad i_l = k\}. \quad (16)$$

Положим

$$\bar{p}_k^{(l)} = \frac{\bar{v}_k^{(l)}}{n}, \quad \bar{v}_k^{(l)}(n) = \sum_{i \in N_l(k)} v_i^{(M)}(n). \quad (17)$$

Заменяя в $X_M(n)$, $X_{M-1}(n)$ параметры $p_{i_1}, \dots, p_{i_M}$ их оценками $\bar{p}_{i_1}^{(1)}, \dots, \bar{p}_{i_M}^{(M)}$, получим статистики

$$\bar{X}_L(n) = \sum_i (\eta_i^{(L)}(n))^2, \quad i = (i_1, \dots, i_L), \quad L = M, M-1, \quad (18)$$

где

$$\eta_i^{(L)}(n) = \frac{v_i^{(L)}(n) - n \bar{a}_i^{(L)}}{\sqrt{n \bar{a}_i^{(L)}}}, \quad \bar{a}_i^{(L)} = \bar{p}_{i_1}^{(1)} \dots \bar{p}_{i_L}^{(L)}. \quad (19)$$

Статистики $\bar{X}_M$, $\bar{X}_{M-1}$ будем рассматривать, как функции величин $\{\xi_u^{(M)}(n)\}$, $u = (u_1, \dots, u_M)$. При $n \rightarrow \infty$ вектор $\{\xi_u^{(M)}(n)\}$ асимптотически нормален, и следовательно, совместное распределение величин $\bar{X}_M(n)$, $\bar{X}_{M-1}(n)$ сходится к некоторому предельному распределению. Такое же предельное распределение сохранится, если в знаменателях величин (19) оценки $\bar{a}_i^{(L)}$ заменим на $a_i^{(L)}$. Величины, получившиеся из величин (19) после такой замены, обозначим $\hat{\eta}_i^{(L)}(n)$. Определим величины $\bar{\xi}_k^{(l)}(n)$, полагая

$$\bar{\xi}_k^{(l)}(n) = \frac{\bar{v}_k^{(l)} - n p_k}{\sqrt{n p_k}}, \quad k = 1, \dots, N.$$

Используя равенства

$$\bar{p}_k^{(l)} = \bar{v}_k^{(l)}(n)/n = p_k + \bar{\xi}_k^{(l)}(n)\sqrt{p_k/n}, \quad k = 1, \dots, N,$$

представим величины $\hat{\eta}_i^{(L)}(n)$ в виде

$$\hat{\eta}_i^{(L)}(n) = \frac{v_i^{(L)}(n) - na_i^{(L)}(1 + \bar{\xi}_{i1}^{(l)}(n)/\sqrt{np_{i1}}) \dots (1 + \bar{\xi}_{iL}^{(L)}(n)/\sqrt{np_{iL}})}{\sqrt{na_i^{(L)}}}.$$

Отсюда следует, что

$$\hat{\eta}_i^{(L)}(n) = \xi_i^{(L)}(n) - \sqrt{a_i^{(L)}} \sum_{l=1}^L \bar{\xi}_{il}^{(l)}(n)/\sqrt{p_{il}} + \varepsilon_i^{(L)}(n), \quad L = M, M-1, \quad (20)$$

где величины $\varepsilon_i^{(L)}(n)$ при $n \rightarrow \infty$ сходятся по вероятности к нулю.

Из равенств (18), (19), (20) получаем следующее утверждение.

Лемма 3. *Предельное при $n \rightarrow \infty$ совместное распределение квадратичных форм*

$$\hat{X}_L(n) = \sum_i \left(\xi_i^{(L)}(n) - \sqrt{a_i^{(L)}} \sum_{l=1}^L \bar{\xi}_{il}^{(l)}(n)/\sqrt{p_{il}} \right)^2, \quad L = M, M-1, \quad (21)$$

совпадает с предельным распределением статистик $\bar{X}_L(n)$, $L = M, M-1$, определенных равенством (18).

Представим $\hat{X}_L(n)$, $L = M, M-1$, в виде

$$\hat{X}_L(n) = \sum_{u,v} (\bar{A}_l^{(M)})_{uv} \xi_u^{(M)} \xi_v^{(M)}, \quad (22)$$

где $L = M, M-1$, $u = (u_1, \dots, u_M)$, $v = (v_1, \dots, v_M)$.

Лемма 4. *Матрицы $\bar{A}_L^{(M)}$, $L = M, M-1$ имеют вид*

$$\bar{A}_M^{(M)} = E_M - \sum_{l=1}^M D_M \begin{pmatrix} l \\ l \end{pmatrix} + 2C_M^2 F_M, \quad (23)$$

$$\bar{A}_{M-1}^{(M)} = G_M - \sum_{l=1}^M D_M \begin{pmatrix} l \\ l \end{pmatrix} + 2C_{M-1}^2 F_M, \quad (24)$$

где E_M — единичная $N^M \times N^M$ матрица, остальные матрицы определены равенствами (10), (11), $C_L^2 = L!/(2!(L-2)!)$, $L = M, M-1$.

Доказательство. Просуммировав по i каждое слагаемое, получившееся после возведения в квадрат выражения в круглых скобках суммы (21) с $L = M$, получим равенство

$$\hat{X}_M(n) = \sum_{l=0}^M Q_{ll} - 2 \sum_{l=1}^M Q_{0l} + 2 \sum_{1 \leq l_1 < l_2 \leq L} Q_{l_1 l_2},$$

где

$$\begin{aligned}
 Q_{00} &= \sum_i (\xi_i^{(M)}(n))^2, \\
 Q_{ll} &= \sum_i a_i^{(M)} (\bar{\xi}_{i_l}^{(M)}(n))^2 / p_{i_l}, \quad l = 1, \dots, M, \\
 Q_{0l} &= \sum_i \sqrt{a_i^{(M)} / p_{i_l}} \xi_i^{(M)}(n) \bar{\xi}_{i_l}^{(l)}(n), \\
 Q_{l_1 l_2} &= \sum_i (a_i^{(M)} / \sqrt{p_{i_{l_1}} p_{i_{l_2}}}) \bar{\xi}_{i_{l_1}}^{(l_1)}(n) \bar{\xi}_{i_{l_2}}^{(l_2)}(n), \quad i = (i_1, \dots, i_M).
 \end{aligned}$$

Используя равенство

$$\bar{\xi}_k^{(l)}(n) = \sum_{i \in \mathcal{N}_l(k)} \sqrt{a_i^{(M)} / p_k} \xi_i^{(M)}(n), \quad i = (i_1, \dots, i_M),$$

например, для Q_{11} получим выражение

$$\sum_i p_{i_1} \dots p_{i_M} \sum_{u_2, \dots, u_M=1}^N \sum_{v_2, \dots, v_M=1}^N \sqrt{p_{u_2} \dots p_{u_M}} \sqrt{p_{v_2} \dots p_{v_M}} \xi_{i_1, u_2, \dots, u_M}^{(M)}(n) \xi_{i_1, v_2, \dots, v_M}^{(M)}(n).$$

Отсюда

$$\begin{aligned}
 Q_{11} &= \sum_{u, v} \sqrt{p_{u_2} \dots p_{u_M} p_{v_2} \dots p_{v_M}} \delta_{u_1 v_1} \xi_u^{(M)}(n) \xi_v^{(M)}(n) \\
 &= \sum_{u, v} \left(D_M \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right)_{u, v} \xi_u^{(M)}(n) \xi_v^{(M)}(n),
 \end{aligned} \tag{25}$$

где $u = (u_1, \dots, u_M)$, $v = (v_1, \dots, v_M)$. Так же проверяется, что матрица коэффициентов, соответствующая Q_{ll} , равна $D_M \begin{pmatrix} l \\ l \end{pmatrix}$ при всех l . Очевидно, что Q_{00} соответствует E_M . Нетрудно проверить, что $Q_{0l} = Q_{ll}$ и что всем $Q_{l_1 l_2}$ соответствует матрица $D_M(\pi_0)$. Отсюда следует (23). Аналогично проверяется (24).

Квадратичную форму

$$\hat{X}(n) = \hat{X}_M(n) - X_{M-1}(n)$$

можно представить в виде

$$\hat{X}(n) = \sum_{u, v} (\bar{A}_M)_{u, v} \xi_u^{(M)}(n) \xi_v^{(M)}(n),$$

где $\bar{A}_M = \bar{A}_M^{(M)} - \bar{A}_{M-1}^{(M-1)}$. Отсюда и из леммы 4 следует, что

$$\bar{A}_M = E_M - D_M \begin{pmatrix} M \\ M \end{pmatrix} - G_M + 2(M-1)F_M. \tag{26}$$

Из доказанных лемм и сделанных замечаний получаем следующее утверждение.

Теорема 1. При $n \rightarrow \infty$ распределение статистики

$$\bar{X}(n) = \bar{X}_M(n) - \bar{X}_{M-1}(n), \quad (27)$$

где $\bar{X}_M(n)$, $\bar{X}_{M-1}(n)$ определены формулой (18), сходится к распределению квадратичной формы

$$X^* = \sum_{u,v} (\bar{A}_M)_{u,v} \xi_u^{(M)}(n) \xi_v^{(M)}(n), \quad u = (u_1, \dots, u_M), \quad v = (v_1, \dots, v_M), \quad (28)$$

где матрица $\bar{A}_M$ определяется равенством (26), а распределение величин $\{\xi_u\}$ является нормальным с нулевым вектором математических ожиданий и ковариационной матрицей B с

$$(B)_{uv} = \lim_{n \rightarrow \infty} \text{cov}(\xi_u^{(M)}(n), \xi_v^{(M)}(n)). \quad (29)$$

Замечание 1. Предельное распределение квадратичной формы, получающейся из $\bar{X}(n)$ заменой всех $\bar{p}_k^{(l)}$, $l = 1, \dots, M$, на $\bar{p}_k^{(M)}$, совпадает с распределением квадратичной формы (28).

Это утверждение следует из равенств

$$\bar{v}_k^{(l)}(n) = \bar{v}_k^{(M)}(n) + \varepsilon_{l,k}(n), \quad k = 1, \dots, N,$$

где все случайные величины $\varepsilon_{l,k}(n)$ ограничены с вероятностью 1 одной и той же постоянной.

4. Предельные распределения статистик, функционально зависящих от известных параметров

Для сплошных цепочек (условие (6)) матрицу B можно представить (см. [5]) в виде

$$B = E_s - (2s - 1)F_s + \sum_{m=1}^{s-1} (B_m + B'_m), \quad (30)$$

где

$$\begin{aligned} B_m &= D_s \begin{pmatrix} m+1 & m+2 & \dots & s \\ 1 & 2 & \dots & s-m \end{pmatrix}, \\ B'_m &= D_s \begin{pmatrix} 1 & 2 & \dots & s-m \\ m+1 & m+2 & \dots & s \end{pmatrix} \end{aligned} \quad (31)$$

Сходимость к χ^2 -распределению распределения статистики $X(n)$, определенной формулой (7), в случае сплошных цепочек отмечалась во введении. Формулы для ковариационной матрицы B и сходимость распределения статистики

$$X(n) = X_M(n) - X_{M-1}(n),$$

определенной формулами (18), для несплошных цепочек в равновероятном случае ($p_k = 1/N$, $k = 1, \dots, N$) были получены в [2]. Эти результаты с очевидными изменениями в вычислениях и доказательствах, приведенных в [2], переносятся на случай произвольных положительных $\{p_k\}$. Приведем утверждение о сходимости распределений статистики $X(n)$.

Теорема 2. Если $p_k > 0$, $k = 1, \dots, N$, то распределение статистики

$$X(n) = X_M(n) - X_{M-1}(n),$$

определенной формулами (4) и (5), для любых $r_1, r_2, \dots, r_M$, удовлетворяющих условию $1 = r_1 < r_2 < \dots < r_M$, при $n \rightarrow \infty$ сходится к центральному χ^2 -распределению с $N^M - N^{M-1}$ степенями свободы.

Приведем формулы для ковариационной матрицы B нормированных частот (5) при $M = 3$, $r_1 = 1$, $r_2 = r$, $r_3 = s$.

Если $2r \neq s + 1$, то матрица B ковариаций нормированных частот $\xi_u^{(3)}(n)$, $u = (u_1, u_2, u_3)$ имеет вид

$$B = E_3 - 7F_3 + D_3 \binom{2}{1} + D_3 \binom{1}{2} + D_3 \binom{3}{1} + D_3 \binom{1}{3} + D_3 \binom{3}{2} + D_3 \binom{2}{3}. \quad (32)$$

Входящие в правую часть этого равенства матрицы определены в разделе 2.

Если $2r = s + 1$, то матрица B при любом r совпадает с матрицей B для сплошных цепочек с $s = 3$. В этом случае можно воспользоваться формулами (30), (31).

Это утверждение сохраняется при любом M . Если числа $r_1, r_2, \dots, r_M$ образуют арифметическую прогрессию, то ковариационная матрица для несплошных цепочек совпадает с ковариационной матрицей сплошных цепочек с $s = M$.

5. Предельные распределения статистик, зависящих от оценок неизвестных параметров

Найдем предельные распределения статистик

$$\bar{X}(n) = \bar{X}_s(n) - \bar{X}_{s-1}(n), \quad (33)$$

где $\bar{X}_L(n)$, $L = s, s-1$, определены равенствами (18), для случая сплошных цепочек (условие (6)). По теореме 1 при условии (6) предельное распределение статистики (33) совпадает с распределением квадратичной формы (28) с матрицей коэффициентов

$$\bar{A}_s = E_s - D_s \binom{s}{s} - C_s + 2(s-1)F_s \quad (34)$$

и ковариационной матрицей B , определенной формулами (30), (31).

Проверим для этих матриц $\bar{A}_s$ и B первое условие (8). Остальные условия (8) выполняются, так как предельные математические ожидания нормированных частот равны 0.

Найдем сначала произведение $B\bar{A}_s$.

Лемма 5. Произведение $B\bar{A}_s$ можно представить в виде

$$B\bar{A}_s = E_s - D_s \binom{s}{s} - G_s + (s-1)F_s + \sum_{m=1}^{s-1} B_m - \sum_{m=1}^{s-2} D_s \begin{pmatrix} 1 & \dots & s-m-1 \\ m+1 & \dots & s-1 \end{pmatrix} - \sum_{m=1}^{s-1} D_s \binom{s-m}{s},$$

где матрицы B_m определены равенством (31).

Доказательство. Из равенств (12) и формул (30), (31) следует, что

$$BF_s = O_s, \quad (35)$$

где O_s является $N^s \times N^s$ матрицей, все элементы которой равны 0, матрица F_s определена равенством (11).

Используя лемму 1, нетрудно проверить, что

$$\begin{aligned} B_m G_s &= B_m, & m &= 1, \dots, s-1, \\ B'_m G_s &= D_s \begin{pmatrix} 1 & \dots & s-m-1 \\ m+1 & \dots & s-1 \end{pmatrix}, & m &= 1, \dots, s-2, \\ D_s \binom{s}{s} G_s &= F_s, \\ B'_m D_s \binom{s}{s} &= D_s \binom{s-m}{s}, & m &= 1, \dots, s-1, \\ B_m D_s \binom{s}{s} &= F_s, & m &= 1, \dots, s. \end{aligned} \quad (36)$$

Из равенств (35), (36) и (12) следует утверждение леммы 5.

Лемма 6. Произведение $\bar{A}_s B \bar{A}_s$ можно представить в виде

$$\bar{A}_s B \bar{A}_s = \bar{A}_s - (2s-3)F_s. \quad (37)$$

Доказательство. Используя лемму 1 и равенства (12), нетрудно проверить, что

$$\begin{aligned} \bar{A}_s B \bar{A}_s &= \bar{A}_s D_s \binom{s-m}{s} = \bar{A}_s F_s = (2s-3)F_s, & m &= 1, \dots, s-1, \\ \bar{A}_s D_s \begin{pmatrix} 1 & \dots & s-m-1 \\ m+1 & \dots & s-1 \end{pmatrix} &= (2s-3)F_s, & m &= 1, \dots, s-2. \end{aligned}$$

Отсюда и из леммы 5 следует утверждение леммы.

Умножив обе части равенства (37) слева на B и воспользовавшись равенством (35), получим, что

$$B \bar{A}_s B \bar{A}_s = B \bar{A}_s. \quad (38)$$

Отсюда следует первое равенство условия (8). Таким образом, предельное распределение статистики (33) является центральным χ^2 -распределением, число степеней свободы которого равно следу матрицы $\bar{A}_s B$. Нетрудно проверить, что

$$\text{tr}(\bar{A}_s B) = \text{tr}(B \bar{A}_s) = N^s - N^{s-1} - N + 1.$$

Получена следующая теорема.

Теорема 3. Если $p_k > 0$, $k = 1, \dots, N$, то распределение статистики (33), определенной по частотам сплошных цепочек (условие (6)), сходится при $n \rightarrow \infty$ к центральному χ^2 -распределению с $(N^s - N^{s-1}) - (N - 1)$ степенями свободы.

Матрицы коэффициентов $\bar{A}_M$ квадратичных форм, использованных для исследования статистик, одинаковы в случаях сплошных и несплошных цепочек. Таким образом, из теоремы 3 следует аналогичное утверждение для несплошных цепочек в тех случаях, когда ковариационные матрицы совпадают.

Теорема 4. Если $p_k > 0$, $k = 1, \dots, N$, и числа $r_1, \dots, r_M$ образуют арифметическую прогрессию, то распределение статистики $\bar{X}(n) = \bar{X}_M(n) - \bar{X}_{M-1}(n)$, определенной по частотам M -цепочек выборки (3), сходится при $n \rightarrow \infty$ к центральному χ^2 -распределению с $(N^s - N^{s-1}) - (N - 1)$ степенями свободы.

Пусть теперь $r_1, \dots, r_M$ не образуют арифметическую прогрессию. Рассмотрим случай $M = 3$, $r_1 = 1$, $r_2 = r$, $r_3 = s$, $2r \neq s + 1$.

Матрица $\bar{A}_3$ в этом случае имеет вид

$$\bar{A}_3 = E_3 - D_3 \begin{pmatrix} 3 \\ 3 \end{pmatrix} - G_3 + 4F_3, \quad (39)$$

а ковариационная матрица B определяется равенством (32).

Используя лемму 1 и равенство $B D_3(\pi_0) = O_3$, нетрудно проверить, что

$$\begin{aligned} B D_3 \begin{pmatrix} 3 \\ 3 \end{pmatrix} &= D_3 \begin{pmatrix} 3 \\ 3 \end{pmatrix} - 3F_3, \\ B G_3 &= G_3 - 5F_3 + D_3 \begin{pmatrix} 2 \\ 1 \end{pmatrix} + D_3 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + D_3 \begin{pmatrix} 3 \\ 1 \end{pmatrix} + D_3 \begin{pmatrix} 3 \\ 2 \end{pmatrix}. \end{aligned}$$

Отсюда и из равенства (39) получим, что

$$B \bar{A}_3 = E_3 + D_3(\pi_0) - G_3 - D \begin{pmatrix} 3 \\ 3 \end{pmatrix}. \quad (40)$$

Нетрудно проверить, что

$$\bar{A}_3 F_3 = \bar{A}_3 G_3 = \bar{A}_3 D_3 \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 3F_3.$$

Из этих равенств и равенства (40) следует, что

$$\bar{A}_3 B \bar{A}_3 = \bar{A}_3 - 3F_3.$$

Умножив обе части этого равенства слева на B , получим, что $B \bar{A}_3 B \bar{A}_3 = B \bar{A}_3$, так как $B F_3 = O_3$.

Получено следующее утверждение.

Теорема 5. Если $p_k > 0$, $k = 1, \dots, N$, $M = 3$, $r_1 = 1$, $r_2 = r$, $r_3 = s$, $2r \neq s + 1$, то распределение статистики $\bar{X}(n) = \bar{X}_3(n) - \bar{X}_2(n)$, где $\bar{X}_3(n)$, $\bar{X}_2(n)$ определены равенствами (18), (19), сходится при $n \rightarrow \infty$ к центральному χ^2 -распределению с $(N^3 - N^2) - (N - 1)$ степенями свободы.

6. Предельные распределения статистик в случае марковской зависимости

Будем предполагать, что последовательность (1) является эргодической цепью Маркова с $N \times N$ матрицей вероятностей переходов (q_{kl}) и стационарными вероятностями (q_l) .

В случае сплошных цепочек предельное распределение $X(n) = X_s(n) - X_{s-1}(n)$ было получено (см. [5]) в схеме серий, когда выборка (3) образована по марковской последовательности с вероятностями переходов $q_{kl} = q_{kl}(n)$, и при $n \rightarrow \infty$

$$q_{kl} = p_l(1 + \varepsilon_{kl}(n)/\sqrt{n}), \quad \varepsilon_{kl}(n) \rightarrow \varepsilon_{kl}, \quad p_l > 0, \quad k, l = 1, \dots, N. \quad (41)$$

В этом случае

$$q_l = p_l(1 + \varepsilon_l(n)/\sqrt{n}), \quad \lim_{n \rightarrow \infty} \mathbf{E} \xi_i^{(s)} = \hat{a}_i^{(s)}, \quad i = (i_1, \dots, i_s), \quad (42)$$

где

$$\hat{a}_i^{(s)} = (\varepsilon_{i_1} + \varepsilon_{i_1 i_2} + \dots + \varepsilon_{i_{s-1} i_s}) \sqrt{a_i^{(s)}}, \quad (43)$$

$a_i^{(s)}$ определено равенством (5),

$$\sum_{l=1}^N p_l \varepsilon_{kl} = \sum_{l=1}^N p_l \varepsilon_{kl} = 0, \quad k = 1, \dots, N, \quad (44)$$

предельное распределение нормированных частот (5) является нормальным.

Ковариационная матрица предельного распределения остается такой же, как в случае независимых испытаний. Таким образом, квадратичная форма нормально распределенных случайных величин, соответствующая предельному распределению, отличается от случая независимых испытаний только вектором математических ожиданий. В случае цепей Маркова вектор математических ожиданий $a = (a_i^{(s)})$ определяется формулой (43). Первое и третье условия (8) следуют из (38). Остается проверить только второе условие (8).

Нетрудно проверить, что

$$a' F_s a = 0. \quad (45)$$

Отсюда и из (37) следует, что

$$a' \bar{A}_s B \bar{A}_s a = a' (\bar{A}_s - (2s - 3) F_s) a' = a' \bar{A}_s a,$$

где матрица $\bar{A}_s$ определена равенством (34).

Таким образом, условие (8) полностью проверено. Параметр нецентральности b^2 предельного распределения определяется равенством $b^2 = a' \bar{A}_s a$. Из равенств

$$\begin{aligned} a' E_s a &= \sum_{l=1}^N \varepsilon_l^2 p_l + (s-1) \sum_{l=1}^N (\Delta_2(l))^2 p_l, \\ a' D_s \binom{s}{s} a &= \sum_{l=1}^N (\Delta_1(l))^2 p_l, \\ a' C_s a &= \sum_{l=1}^N \varepsilon_l^2 p_l + (s-2) \sum_{l=1}^N (\Delta_2(l))^2 p_l, \end{aligned}$$

где

$$\Delta_m(l) = \sum_{k=1}^N \varepsilon_{kl}^m p_k, \quad m = 1, 2, \quad (46)$$

и из равенств (34) следует, что

$$b^2 = \sum_{l=1}^N (\Delta_2(l) - (\Delta_1(l))^2) p_l. \quad (47)$$

Получено следующее утверждение.

Теорема 6. Если последовательность (1) является цепью Маркова и в схеме серий при $n \rightarrow \infty$ матрица вероятностей переходов удовлетворяет условиям (41), то распределение статистики $\bar{X}(n)$, определенной равенством (33) по частотам сплошных цепочек состояний (условие (6)), сходится к нецентральному χ^2 -распределению с $(N^s - N^{s-1}) - (N - 1)$ степенями свободы и параметром нецентральности b^2 , определенным равенствами (46), (47).

Список литературы

1. Степанов В. Е., Некоторые статистические критерии для цепей Маркова. *Теория вероятностей и ее применения* (1957) **2**, №1, 143–144.
2. Чистяков В. П., Статистики хи-квадрат, построенные по частотам неполных цепочек исходов независимых испытаний. *Дискретная математика* (2001) **13**, №4, 52–59.
3. Крамер Г., *Математические методы статистики*. Мир, Москва, 1976.
4. Dik J. J., de Gunst M. C. M., The distribution of general quadratic form in normal variables. *Stat. Neerl.* (1985) **39**, №1, 14–26.
5. Селиванов Б. И., Чистяков В. П., Последовательный χ^2 -критерий, построенный по s -цепочкам состояний цепи Маркова. *Дискретная математика* (1997) **9**, №4, 127–136.

Статья поступила 03.08.2004.

Переработанный вариант поступил 02.11.2004.