Math-Net.Ru

Общероссийский математический портал

М. И. Тихомирова, В. П. Чистяков, Скользящий хиквадрат, *Дискрет. матем.*, 2000, том 12, выпуск 4, 46–52

DOI: <https://doi.org/10.4213/dm349>

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: 18.208.226.222

26 апреля 2023 г., 13:04:28

Скольльзящий хи-квадрат

© 2000 г. М. И. Тихомирова, В. П. Чистяков

Последовательность независимых одинаково распределенных величин, принимающих значения из множества $\{1, 2, \dots, N\}$ разбивается на непересекающиеся отрезки длины n , и из s последовательных отрезков, начиная с отрезка с номером t , образуют выборку объема ns с номером t . Доказано, что при $n \rightarrow \infty$ и фиксированных N и r совместное r -мерное распределение χ^2 -статистик, построенных по выборкам объемов ns с номерами $t_1 < \dots < t_r$, сходится к некоторому предельному распределению. Для этого предельного распределения найдено преобразование Лапласа и указано нормальное приближение.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проект 00-15-96136.

1. Введение

Рассмотрим последовательность независимых одинаково распределенных случайных величин

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

принимающих значения из множества $\{1, 2, \dots, N\}$ с вероятностями

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

В [1] исследовалось совместное распределение χ^2 -статистик

$$\tilde{X}_{n_1}, \tilde{X}_{n_2}, \dots, \tilde{X}_{n_r},$$

построенных по нарастающим выборкам

$$(x_1, x_2, \dots, x_{n_k}), \quad k = 1, \dots, r; \quad n_1 < n_2 < \dots < n_r.$$

В данной работе выборки, по которым строятся χ^2 -статистики, образуются следующим образом. Последовательность (1) разбивается на непересекающиеся отрезки объема n

$$x(v, n) = \{x_{n(v-1)+1}, x_{n(v-1)+2}, \dots, x_{nv}\}, \quad v = 1, 2, \dots \quad (3)$$

Из последовательности $\{x(v, n)\}$ образуется последовательность s -цепочек

$$(x(1, n), \dots, x(s, n)), (x(2, n), \dots, x(s+1, n)), \dots \quad (4)$$

По t -й s -цепочке объема ns

$$x(t, n), x(t+1, n), \dots, x(t+s-1, n) \quad (5)$$

строится χ^2 -статистика

$$X_t(n) = \sum_{k=1}^N \frac{(\nu_k^{(t)}(ns) - nsp_k)^2}{nsp_k}, \quad (6)$$

где

$$\nu_k^{(t)}(ns) = \tilde{\nu}_k^{(t)}(n) + \tilde{\nu}_k^{(t+1)}(n) + \dots + \tilde{\nu}_k^{(t+s-1)}(n),$$

$\tilde{\nu}_k^{(v)}(n)$ — частота исхода k в v -й выборке $x(v, n)$, определенной равенством (3). Ниже показано, что распределение вектора

$$(X_{t_1}(n), \dots, X_{t_r}(n)), \quad 1 \leq t_1 < t_2 < \dots < t_r,$$

при $n \rightarrow \infty$ и фиксированном N сходится к предельному распределению, найдено преобразование Лапласа предельного распределения и получено нормальное приближение этого распределения.

2. Сходимость к предельному распределению

Статистики (6) можно представить в виде

$$X_t(n) = \sum_{k=1}^N \xi_{t,k}^2(n), \quad (7)$$

где

$$\xi_{t,k}(n) = \frac{\nu_k^{(t)}(ns) - nsp_k}{\sqrt{nsp_k}}. \quad (8)$$

Математические ожидания случайных величин (8) равны нулю. Для вторых моментов этих величин справедливо следующее утверждение.

Лемма 1. При любых целых $n, s, t_1, t_2 \geq 1$ и $k, l \in \{1, 2, \dots, N\}$

$$\text{cov}(\xi_{t_1,k}(n), \xi_{t_2,l}(n)) = \rho_{t_1 t_2}(\delta_{kl} - \sqrt{p_k p_l}), \quad (9)$$

где $\delta_{kl} = 0$ при $k \neq l$, $\delta_{kk} = 1$,

$$\rho_{t_1 t_2} = 1 - \frac{|t_1 - t_2|}{s}, \quad 0 \leq |t_1 - t_2| \leq s-1,$$

и $\rho_{t_1 t_2} = 0$ при $|t_1 - t_2| \geq s$.

Доказательство. В случае $|t_1 - t_2| \geq s$ величины $\xi_{t_1,k}(n)$ и $\xi_{t_2,l}(n)$ определяются по независимым выборкам и их ковариация равна нулю.

Пусть теперь $t_1 \leq t_2 < t_1 + s$. Тогда

$$\begin{aligned}\nu_k^{(t_1)}(ns) &= \nu_{1k} + \nu_{2k}, \\ \nu_l^{(t_2)}(ns) &= \nu_{2l} + \nu_{3l},\end{aligned}$$

где ν_{2k} и ν_{2l} — частоты исходов k и l в общей части двух выборок (5) с $t = t_1$ и $t = t_2$, а ν_{1k} и ν_{3l} — соответствующие частоты, подсчитанные по оставшимся частям выборок. Объем общей части выборок равен $(s - |t_1 - t_2|)n$. Таким образом,

$$\begin{aligned}\text{cov}(\nu_k^{(t_1)}(ns), \nu_l^{(t_2)}(ns)) &= \text{cov}(\nu_{2k}, \nu_{2l}) \\ &= (s - |t_1 - t_2|)n(\delta_{kl}p_k - p_k p_l).\end{aligned}$$

Отсюда и из (8) следует утверждение леммы.

При $n \rightarrow \infty$ совместное распределение TN величин

$$\xi_{t,k}(n), \quad t = 1, 2, \dots, T, \quad k = 1, 2, \dots, N,$$

определенных формулой (8), сходится к многомерному нормальному распределению с нулевым вектором математических ожиданий и ковариационной матрицей, определенной формулой (9).

Обозначим

$$\xi_{11}, \dots, \xi_{1N}, \xi_{21}, \dots, \xi_{2N}, \dots, \xi_{T1}, \dots, \xi_{TN} \quad (10)$$

случайный вектор, имеющий нормальное распределение с нулевым математическим ожиданиям и ковариационной матрицей, определенной формулой (9).

Из асимптотической нормальности $\{\xi_{t,k}(n)\}$ следует (см. [2]) сходимость распределений непрерывных функций от этих величин. Сформулируем полученное утверждение в виде теоремы.

Теорема 1. При $n \rightarrow \infty$ и постоянных $r, N, t_1 < t_2 < \dots < t_r$, распределение вектора

$$X_{t_1}(n), X_{t_2}(n), \dots, X_{t_r}(n) \quad (11)$$

сходится к распределению вектора

$$X_{t_1}, X_{t_2}, \dots, X_{t_r},$$

где

$$X_{t_v} = \sum_{k=1}^N \xi_{t_v,k}^2, \quad v = 1, \dots, r, \quad (12)$$

$\xi_{t_v,k}$ — координаты вектора (10).

3. Преобразование Лапласа

Квадратичные формы (12) удобно выразить через случайные величины

$$\zeta_{t_v,k} = \sum_{i=1}^N c_{ki} \xi_{t_v,i}, \quad k = 1, \dots, N, \quad (13)$$

где (c_{ij}) — ортонормированная $N \times N$ матрица. Эта матрица не зависит от t_v , и ее последняя строка определяется равенствами

$$c_{Ni} = \sqrt{p_i}, \quad i = 1, 2, \dots, N.$$

Лемма 2. Ковариационная матрица величин $\{\zeta_{t_v,k}\}$ определяется формулой

$$\text{cov}(\zeta_{t_v,k}, \zeta_{t_u,l}) = \rho_{t_v,t_u} (\delta_{kl} - \delta_{lN} \delta_{kN}), \quad (14)$$

где ρ_{t_v,t_u} определены в лемме 1.

Утверждение этой леммы следует из равенства (9) и следующих равенств:

$$\begin{aligned} \text{cov}(\zeta_{t_v,k}, \zeta_{t_u,l}) &= \sum_{i,j=1}^N c_{ki} c_{lj} \text{cov}(\xi_{t_v,i}, \xi_{t_u,j}) \\ &= \rho_{t_v,t_u} \sum_{i,j=1}^N c_{ki} c_{lj} (\delta_{ij} - \sqrt{p_i p_j}) \\ &= \rho_{t_v,t_u} \left(\sum_{i=1}^N c_{li} c_{ki} - \sum_{i=1}^N c_{li} \sqrt{p_i} \sum_{i=1}^N c_{ki} \sqrt{p_i} \right), \end{aligned}$$

так как в силу ортонормированности (c_{ij})

$$\sum_{i=1}^N c_{li} c_{ki} = \delta_{kl}, \quad \sum_{i=1}^N c_{li} c_{Ni} = \delta_{lN},$$

а $c_{Ni} = \sqrt{p_i}$, $i = 1, \dots, N$.

Из формулы (14) и равенства $\mathbf{E} \zeta_{t_v,B} = 0$ нетрудно получить следующие утверждения.

Следствие 1. При любом $v = 1, \dots, r$ дисперсия $\mathbf{D} \zeta_{t_v,N} = 0$ и $\zeta_{t_v,N} = 0$ с вероятностью единица.

Следствие 2. Случайные векторы

$$(\zeta_{t_1,k}, \zeta_{t_2,k}, \dots, \zeta_{t_r,k}), \quad k = 1, \dots, N-1,$$

независимы и одинаково распределены.

Следствие 3. Случайный вектор $(X_{t_1}, \dots, X_{t_r})$ представим в виде суммы $N-1$ независимых одинаково распределенных векторов:

$$(X_{t_1}, \dots, X_{t_r}) = \sum_{k=1}^{N-1} (\zeta_{t_1,k}^2, \dots, \zeta_{t_r,k}^2). \quad (15)$$

Отметим, что вектор $(\zeta_{t_1,k}, \dots, \zeta_{t_r,k})$ при любом $k, k = 1, \dots, N-1$, имеет нормальное распределение и

$$\mathbf{E}\zeta_{t_i,k} = 0, \quad \text{cov}(\zeta_{t_i,k}, \zeta_{t_j,k}) = \rho_{t_i,t_j}. \quad (16)$$

Покажем, что определитель матрицы (ρ_{t_i,t_j}) не равен нулю. Дисперсия случайной величины

$$\zeta = A_1 \zeta_{t_1,1}(n) + \dots + A_r \zeta_{t_r,1}(n),$$

где

$$\zeta_{t_i,1}(n) = \sum_{k=1}^N c_{ik} \xi_{t_i,k}(n), \quad i = 1, \dots, r,$$

является неотрицательно определенной квадратичной формой:

$$\mathbf{D}\zeta = \sum_{i,j=1}^r \rho_{t_i,t_j} A_i A_j.$$

Предположим, что определитель равен нулю. Тогда найдутся $A_1, \dots, A_r$, при которых

$$\mathbf{D}\zeta = 0, \quad \sum_{i=1}^r A_i^2 > 0.$$

Среди $A_1, \dots, A_r$ выберем первый отличный от нуля коэффициент. Пусть, например, $A_1 \neq 0$. Случайную величину $\zeta_{t_1,1}(n)$ представим в виде

$$\zeta_{t_1,1}(n) = \zeta'_{t_1,1}(n) + \zeta''_{t_1,1}(n),$$

где

$$\begin{aligned} \zeta'_{t_1,1}(n) &= \sum_{i=1}^N c_{1i} \xi'_{t_1,i}(n), \\ \zeta''_{t_1,1}(n) &= \sum_{i=1}^N c_{1i} \xi''_{t_1,i}(n), \\ \xi'_{t_1,1}(n) &= \frac{\tilde{\nu}_i^{(t_1)}(n) - nsp_i}{\sqrt{nsp_i}}, \\ \xi''_{t_1,1}(n) &= \frac{\tilde{\nu}_i^{(t_1+1)}(n) + \dots + \tilde{\nu}_i^{(t_1+s-1)}(n) - nsp_i}{\sqrt{nsp_i}}, \end{aligned}$$

$\tilde{\nu}_i^{(v)}(n)$ — частота исхода i в v -й выборке $x(v, n)$, определенной равенством (3). Частоты исходов, входящие в определение величин

$$\xi''_{t_1,i}, \quad \xi_{t_l,i}, \quad l = 2, \dots, r, i = 1, \dots, N,$$

определяются по выборкам, не пересекающимися с выборкой $x(t_1, n)$. Таким образом,

$$\zeta = A_1 \zeta'_{t_1,1}(n) + A_1 \zeta''_{t_1,1}(n) + \sum_{l=2}^r A_l \zeta_{t_l,1}(n)$$

и величины $A_1 \zeta'_{t_1,1}(n)$ и $\zeta - A_1 \zeta'_{t_1,1}(n)$ независимы. Следовательно,

$$D\zeta \geq A_1^2 D\zeta'_{t_1,1}(n).$$

Отсюда и из равенств

$$\begin{aligned} D\zeta'_{t_1,1}(n) &= \sum_{i,j=1}^N c_{1i} c_{1j} \text{cov}(\xi'_{t_1,i}, \xi'_{t_1,j}) \\ &= \frac{1}{s} \sum_{i,j=1}^N c_{1i} c_{1j} (\delta_{ij} - \sqrt{p_i p_j}) = \frac{1}{s} \end{aligned}$$

получаем, что $D\zeta > 0$. Получаемое противоречие доказывает, что определитель не равен нулю.

Преобразование Лапласа

$$l(s_1, \dots, s_r) = \mathbf{E} \exp\{-s_1 \zeta_{t_1,k} - \dots - s_r \zeta_{t_r,k}\}$$

можно представить в виде

$$l(s_1, \dots, s_r) = \frac{1}{(\sqrt{2\pi})^2 |(\rho_{t_i, t_j})|} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp \left\{ -\sum_{i=1}^r s_i x_i - \frac{1}{2} \sum_{i,j=1}^r \rho_{t_i, t_j}^* x_i x_j \right\} dx_1 \dots dx_r, \quad (17)$$

где $|(\rho_{t_i, t_j})|$ — определитель матрицы (ρ_{t_i, t_j}) и $(\rho_{t_i, t_j}^*) = (\rho_{t_i, t_j})^{-1}$. Интеграл (17) вычисляется так же, как соответствующий интеграл в [1], с использованием формулы (П.12, 1а) из книги Г. Крамера [3].

Воспользовавшись результатом этого вычисления и представлением (15), получаем следующее утверждение.

Теорема 2. Преобразование Лапласа $L(s_1, \dots, s_r)$ вектора $(X_{t_1}, \dots, X_{t_r})$ имеет вид

$$L(s_1, \dots, s_r) = (|(\rho_{t_i, t_j})| |(\rho_{t_i, t_j}^* + 2s_i \delta_{ij})|)^{-(N-1)/2}, \quad (18)$$

где $|(\rho_{t_i, t_j})|$, $|(\rho_{t_i, t_j}^* + 2s_i \delta_{ij})|$ — определители матриц (ρ_{t_i, t_j}) , $(\rho_{t_i, t_j}^* + 2s_i \delta_{ij})$ и $(\rho_{t_i, t_j}^*) = (\rho_{t_i, t_j})^{-1}$.

4. Нормальное приближение

Найдем моменты первых двух порядков вектора $(X_{t_1}, \dots, X_{t_r})$. Это можно сделать, дифференцируя преобразование Лапласа или непосредственно, используя представление (15).

Во втором случае нужно воспользоваться известными формулами (см. [3]) для моментов нормального распределения с ковариационной матрицей (16):

$$\begin{aligned} \mathbf{E} \zeta_{t_i, k}^2 &= 1, \quad \mathbf{E} \zeta_{t_i, k}^4 = 3, \quad \mathbf{E}(\zeta_{t_i, k}, \zeta_{t_j, k}) = \rho_{t_i, t_j}, \\ \mathbf{E}(\zeta_{t_i, k}^2, \zeta_{t_j, k}^2) &= 2(\mathbf{E} \zeta_{t_i, k} \zeta_{t_j, k})^2 + \mathbf{E} \zeta_{t_i, k}^2 \mathbf{E} \zeta_{t_j, k}^2. \end{aligned}$$

Отсюда нетрудно получить, что

$$\begin{aligned} \mathbf{E}X_{t_i} &= N - 1, & \mathbf{D}X_{t_i} &= 2(N - 1), \\ \text{cov}(X_{t_i}, X_{t_j}) &= 2(N - 1)\rho_{t_i, t_j}^2. \end{aligned}$$

Применив центральную предельную теорему к представлению вектора

$$(X_{t_1}, \dots, X_{t_r})$$

в виде (15), получим следующее утверждение.

Теорема 3. *Распределение вектора $(\eta_{t_1}, \dots, \eta_{t_r})$, где*

$$\eta_{t_v} = \frac{X_{t_v} - (N - 1)}{\sqrt{2(N - 1)}}, \quad v = 1, \dots, r,$$

и X_{t_v} , $v = 1, \dots, r$, определены равенствами (10), (12), сходится при $N \rightarrow \infty$ к нормальному распределению с нулевым вектором математических ожиданий и ковариационной $r \times r$ матрицей (ρ_{t_i, t_j}^2) , где ρ_{t_i, t_j} определены в лемме 1.

Список литературы

1. Захаров В. К., Сарманов О. В., Севастьянов Б. А., Последовательный критерий χ^2 . *Матем. сб.* (1969) **79**, №3, 444–460.
2. Биллингсли П., *Сходимость вероятностных мер*. Наука, Москва, 1977.
3. Крамер Г., *Математические методы статистики*. Мир, Москва, 1975.

Статья поступила 03.05.2000.