Math-Net.Ru

Общероссийский математический портал

В. М. Круглов, Сходимость вполне статистики Пирсона,
Матем. заметки, 1999, том 66, выпуск 4, 626–629

DOI: <https://doi.org/10.4213/mzm1204>

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: 54.209.52.79

26 апреля 2023 г., 13:36:05

СХОДИМОСТЬ ВПОЛНЕ СТАТИСТИКИ ПИРСОНА

В. М. Круглов

Статистика Пирсона, ее также называют статистикой хи-квадрат, была введена К. Пирсоном в связи с решением ряда задач математической статистики. С многочисленными применениями этой статистики можно познакомиться, например, по учебнику [1]. Нас интересуют асимптотические свойства статистики Пирсона. Имеется большое число публикаций, посвященных исследованию асимптотических свойств самой статистики Пирсона и многочисленных ее модификаций (см. [1]–[10] и литературу, указанную там). Кажется, что свойства статистики Пирсона, выполняющиеся почти всюду, до сих пор систематически не обсуждались. В предлагаемой заметке приведены достаточные условия сходимости вполне и, следовательно, с вероятностью единица статистики Пирсона.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, гранты № 99-01-00846 и № 99-01-00847.

Пусть на вероятностном пространстве $(\Omega, \mathcal{F}, P)$ определен случайный вектор $\bar{\nu} = (\nu_1, \dots, \nu_s)$ с полиномиальным распределением. В соответствии с определением полиномиального распределения мы имеем

$$P\{\nu_1 = m_1, \nu_2 = m_2, \dots, \nu_s = m_s\} = \frac{n!}{m_1! m_2! \dots m_s!} p_1^{m_1} p_2^{m_2} \dots p_s^{m_s}$$

для любых неотрицательных целых чисел $m_1, m_2, \dots, m_s$, сумма которых равна n , где $p_k > 0$, $k = 1, 2, \dots, s$, и сумма этих чисел равна 1.

Статистикой Пирсона называется случайная величина

$$\chi_n^2 = \sum_{k=1}^s \frac{n}{p_k} \left(\frac{\nu_k}{n} - p_k \right)^2. \quad (1)$$

Легко видеть, что случайная величина ν_k имеет биномиальное распределение с параметрами n и p_k . Положим $q_k = 1 - p_k$. Стандартные вычисления приводят к следующим выражениям для математических ожиданий, дисперсий и ковариаций слагаемых суммы (1):

$$\begin{aligned} E\left(\frac{n}{p_k} \left(\frac{\nu_k}{n} - p_k\right)^2\right) &= q_k, \quad \text{Var}\left(\frac{n}{p_k} \left(\frac{\nu_k}{n} - p_k\right)^2\right) = 2q_k^2 + \frac{q_k}{np_k}(1 - 6p_k q_k), \\ \text{Cov}\left(\frac{n}{p_k} \left(\frac{\nu_k}{n} - p_k\right)^2, \frac{n}{p_j} \left(\frac{\nu_j}{n} - p_j\right)^2\right) &= 2p_k p_j + \frac{1}{n}(2p_k + 2p_j - 6p_k p_j) - \frac{1}{n}, \quad k \neq j. \end{aligned}$$

С помощью этих формул нетрудно найти известные выражения для математического ожидания и дисперсии статистики Пирсона:

$$E\chi_n^2 = s - 1, \quad \text{Var}\chi_n^2 = 2(s - 1) + \frac{1}{n} \left(\sum_{k=1}^s \frac{1}{p_k} - s^2 - 2s + 2 \right).$$

ТЕОРЕМА. Если $s = s_n$ и p_k , $k = 1, 2, \dots, s_n$, зависят от n таким образом, что

$$\sum_{n=1}^{\infty} \frac{1}{s_n^2} < \infty, \quad \inf_n \left\{ n \min_{1 \leq k \leq s_n} p_k \right\} = \alpha > 0, \quad \sup_n \left\{ \frac{s}{n} \right\} = \beta < \infty,$$

то для любого $\varepsilon > 0$

$$\sum_{n=1}^{\infty} P\left\{ \max_{1 \leq m \leq s_n} \left| \sum_{k=1}^m \left(\frac{n}{p_k} \left(\frac{\nu_k}{n} - p_k \right)^2 - q_k \right) \right| \geq \varepsilon s_n \right\} < \infty \quad (2)$$

и, следовательно,

$$P\left\{ \lim_{n \rightarrow \infty} \frac{1}{s_n} \sum_{k=1}^{s_n} \left(\frac{n}{p_k} \left(\frac{\nu_k}{n} - p_k \right)^2 - q_k \right) = 0 \right\} = 1. \quad (3)$$

ДОКАЗАТЕЛЬСТВО. Нетрудно видеть, что (3) является следствием (2). В свою очередь (2) вытекает из неравенства

$$P\left\{ \max_{1 \leq m \leq s_n} \left| \sum_{k=1}^m \left(\frac{n}{p_k} \left(\frac{\nu_k}{n} - p_k \right)^2 - q_k \right) \right| \geq \lambda \right\} \leq \frac{C}{\lambda^4} s_n^2, \quad \lambda > 0, \quad (4)$$

где C – некоторая положительная постоянная, не зависящая от s и n .

Неравенство (4) будет выведено из неравенства Биллингсли (см. [11, с. 134]). Напомним его. Пусть $X_1, X_2, \dots, X_m$ – случайные величины, которые не обязаны быть независимыми. Положим $S_0 = 0$, $S_k = X_1 + X_2 + \dots + X_k$, $k = 1, 2, \dots, m$.

Предположим, что существуют положительные постоянные $u_1, u_2, \dots, u_m$ такие, что

$$E|S_l - S_{r-1}|^\gamma \leq \left(\sum_{i=r}^l u_i \right)^\alpha, \quad 1 \leq r < l \leq m,$$

для некоторых $\gamma > 0$ и $\alpha > 1$. Тогда для всех положительных λ

$$P\left\{ \max_{1 \leq k \leq m} |S_k| \geq \lambda \right\} \leq \frac{K_{\gamma, \alpha}}{\lambda^\gamma} \left(\sum_{i=1}^m u_i \right)^\alpha, \quad (5)$$

где постоянная $K_{\gamma, \alpha}$ зависит только от α и γ .

Докажем (4). В [12] приведены вычисления ряда центральных моментов статистики Пирсона и, в частности, четвертого центрального момента

$$\begin{aligned} E(\chi_n^2 - E\chi_n^2)^4 &= 12(s-1)(s+3) + 12[(s+31)R_1 - (s^3 + 25s^2 + 44s - 38)]n^{-1} \\ &+ [112R_2 + 3R_1^2 - 2(3s^2 + 118s + 658)R_1 + 3(s^4 + 44s^3 + 328s^2 + 488s - 380)]n^{-2} \\ &+ [R_3 - 4(s+28)R_2 - 3R_1^2 + 4(3s^2 + 56s + 236)R_1 - 6(s^4 + 20s^3 + 116s^2 + 160s - 120)]n^{-3}, \end{aligned}$$

где

$$R_1 = \sum_{k=1}^s \frac{1}{p_k}, \quad R_2 = \sum_{k=1}^s \frac{1}{p_k^2}, \quad R_3 = \sum_{k=1}^s \frac{1}{p_k^3}.$$

Из этого выражения следует, что

$$E(\chi_n^2 - E\chi_n^2)^4 \leq As^2, \quad (6)$$

$$A = \sup_n \left\{ s^{-2} E(\chi_n^2 - E\chi_n^2)^4 \right\} < 1224 + 1464\beta + 1119\beta^2 + \frac{236\beta^2 + 944\beta + 384}{\alpha} + \frac{115}{\alpha^2} + \frac{1}{\alpha^3}.$$

Пусть даны натуральные числа r и l , $1 \leq r < l < s_n$, $l - r + 1 \leq s_n$. Положим

$$\begin{aligned} p_{r,k} &= p_k, \quad \nu_{r,k} = \nu_k, \quad k = r, r+1, \dots, l, \\ p_{r,l+1} &= 1 - p_r - \dots - p_l, \quad \nu_{r,l+1} = n - \nu_r - \nu_{r+1} - \dots - \nu_r. \end{aligned}$$

Случайный вектор $(\nu_{r,r}, \dots, \nu_{r,l+1})$ имеет полиномиальное распределение, так как для любых неотрицательных целых чисел $m_r, \dots, m_{l+1}$, $m_r + m_{r+1} + \dots + m_{l+1} = n$, выполнено

$$P\{\nu_{r,r} = m_r, \dots, \nu_{r,l+1} = m_{l+1}\} = \frac{n!}{m_r! \dots m_{l+1}!} p_{r,r}^{m_r} \dots p_{r,l+1}^{m_{l+1}}.$$

Запишем соответствующую статистику Пирсона в виде суммы двух слагаемых

$$\sum_{k=r}^{l+1} \frac{n}{p_{r,k}} \left(\frac{\nu_{r,k}}{n} - p_{r,k} \right)^2 = \chi_n^2(r, l) + \frac{n}{p_{r,l+1}} \left(\frac{\nu_{r,l+1}}{n} - p_{r,l+1} \right)^2, \quad (7)$$

где

$$\chi_n^2(r, l) = \sum_{k=r}^l \frac{n}{p_k} \left(\frac{\nu_k}{n} - p_k \right)^2.$$

Математическое ожидание k -го слагаемого в сумме, стоящей слева в формуле (7), равно $q_{r,k} = 1 - p_{r,k}$. Вычитая их сумму из обеих частей равенства (7), мы получим

$$\sum_{k=r}^{l+1} \left(\frac{n}{p_{r,k}} \left(\frac{\nu_{r,k}}{n} - p_{r,k} \right)^2 - q_{r,k} \right) = \chi_n^2(r, l) - E\chi_n^2(r, l) + \frac{n}{p_{r,l+1}} \left(\frac{\nu_{r,l+1}}{n} - p_{r,l+1} \right)^2 - q_{r,l+1}.$$

Применение элементарного неравенства $(a + b)^4 \leq 16(|a|^4 + |b|^4)$, справедливого для любых вещественных чисел, дает

$$\begin{aligned} E(\chi_n^2(r, l) - E\chi_n^2(r, l))^4 &\leq 16E\left(\left|\sum_{k=r}^{l+1}\left(\frac{n}{p_{r,k}}\left(\frac{\nu_{r,k}}{n} - p_{r,k}\right)^2 - q_{r,k}\right)\right|^4\right. \\ &\quad \left.+ \left|\frac{n}{p_{r,l+1}}\left(\frac{\nu_{r,l+1}}{n} - p_{r,l+1}\right)^2 - q_{r,l+1}\right|^4\right). \end{aligned} \quad (8)$$

Оценим слагаемые в правой части (8). К первому из них можно применить неравенство (6), согласно которому

$$E\left|\sum_{k=r}^{l+1}\left(\frac{n}{p_{r,k}}\left(\frac{\nu_{r,k}}{n} - p_{r,k}\right)^2 - q_{r,k}\right)\right|^4 \leq A(l+2-r)^2 \leq 4A(l+1-r)^2.$$

Далее, заметим, что случайная величина $\nu_{r,l+1} = n - \nu_r - \dots - \nu_l$ имеет биномиальное распределение с параметрами n и $p_{r,l+1} = 1 - p_r - \dots - p_l$. Обозначим временно $p_{r,l+1} = p$, $q_{r,l+1} = 1 - p = q$ и воспользуемся вычислениями из [2], согласно которым

$$\begin{aligned} E(\nu_{r,l+1} - np_{r,l+1})^8 &= 105n^4 p^4 q^4 + 70n^3 p^3 q^3 (7 - 34pq) + 7n^2 p^2 q^2 (17 - 308pq + 1044p^2 q^2) \\ &\quad + npq(1 - 126pq + 1680p^2 q^2 - 5040p^3 q^3). \end{aligned}$$

Из этого выражения следует, что

$$\sup_n E\left(\frac{n}{p_{r,l+1}}\left(\frac{\nu_{r,l+1}}{n} - p_{r,l+1}\right)^2\right)^4 = B < 105 + \frac{490}{\alpha} + \frac{7427}{\alpha^2} + \frac{1681}{\alpha^3}.$$

Привлекая еще раз упомянутое выше элементарное неравенство, мы получим

$$\begin{aligned} E\left|\frac{n}{p_{r,l+1}}\left(\frac{\nu_{r,l+1}}{n} - p_{r,l+1}\right)^2 - q_{r,l+1}\right|^4 &\leq 16E\left(\left(\frac{n}{p_{r,l+1}}\left(\frac{\nu_{r,l+1}}{n} - p_{r,l+1}\right)^2\right)^4 + q_{r,l+1}^4\right) \\ &\leq 16(B+1). \end{aligned}$$

Замена слагаемых в правой части (8) их оценками дает

$$E(\chi_n^2(r, l) - E\chi_n^2(r, l))^4 \leq 16(4A(l+1-r)^2 + 16(B+1)) \leq (64A + 256B + 256)(l+1-r)^2.$$

Теперь (4) вытекает из (5), где следует положить $\alpha = 2$, $\gamma = 4$, $m = s_n$,

$$X_k = \frac{n}{p_k} \left(\frac{\nu_k}{n} - p_k \right)^2 - q_k, \quad k = 1, 2, \dots, s_n,$$

$u_1 = u_2 = \dots = u_{s_n} = 2\sqrt{64A + 256B + 256}$. В качестве постоянной C , фигурирующей в (4), можно взять $C = K_{4,2}(64A + 256B + 256)$, где $K_{4,2}$ – постоянная, о которой говорится в неравенстве Биллингсли. Теорема доказана.

СПИСОК ЦИТИРОВАННОЙ ЛИТЕРАТУРЫ

1. Крамер Г. Математические методы статистики. М.: Мир, 1975.
2. Туманян С. Х. // Теория вероятн. и ее примен. 1956. Т. 1. № 1. С. 131–145.
3. Гихман И. И. // Теория вероятн. и ее примен. 1956. Т. 1. № 3. С. 344–348.
4. Ивченко Г. И., Левин В. В. // Обзорение прикл. промышл. матем. Сер. дискретн. матем. 1995. Т. 2. № 6. С. 871–887.
5. Ермаков М. С. // Теория вероятн. и ее примен. 1997. Т. 42. № 4. С. 668–695.
6. Morris C. // Ann. Statist. 1975. V. 3. № 1. P. 165–188.
7. Медведев Ю. И. // Докл. АН СССР. 1970. Т. 192. № 5. С. 987–989.
8. Медведев Ю. И. // Теория вероятн. и ее примен. 1977. Т. 22. № 1. С. 3–17.
9. Медведев Ю. И. // Теория вероятн. и ее примен. 1977. Т. 22. № 3. С. 623–631.
10. Захаров В. К., Сарманов О. В., Севастьянов Б. А. // Матем. сб. 1969. Т. 79. № 3. С. 444–460.
11. Биллингсли П. Сходимость вероятностных мер. М.: Наука, 1977.
12. Haldane J. B. S. // Biometrics. 1937. V. 29. P. 133–143.