УДК 519.234.33

Предельные распределения статистики Пирсона для неоднородной полиномиальной схемы

М. П. Савелов[*†]

[*] *Московский государственный университет им. М.В.Ломоносова, Ленинские Горы 1, Москва, Россия, 119992*

[†] *Московский физико-технический институт (государственный университет), Институтский пер., д.9, г. Долгопрудный, Московская область, Россия, 141701*

Аннотация. Для неоднородной полиномиальной схемы найдены условия, при которых распределение статистики Пирсона сходится к распределению неотрицательно определенной квадратичной формы от независимых случайных величин со стандартным нормальным распределением.

Ключевые слова: критерий хи-квадрат, статистика Пирсона, предельные распределения.

1. Введение

Рассмотрим схему независимых испытаний с N исходами. В соответствии с нулевой гипотезой H_0 будем считать, что вероятность j-го исхода в t-м испытании равна $p_j^t, j = 1, \ldots, N$. Для $t \geq 1$ и $1 \leq j \leq N$ введем индикаторы $I_j(t)$: $I_j(t) = 1$, если при t-м испытании осуществился j-й исход, и $I_j(t) = 0$ в противном случае. Тогда $\nu_j(n) := \sum_{t=1}^{n} I_j(t)$ — частота исхода j в первых n испытаниях. Пусть $p_1, ..., p_N$ - фиксированные положительные числа. Статистика К. Пирсона

$$X(n) := \sum_{j=1}^{N} \frac{1}{np_j} \left(\nu_j(n) - np_j\right)^2$$

используется, например, в критерии согласия (см., например, [1]; подробный обзор есть в [4]) с гипотезой, которая в наших обозначениях имеет вид: «$p_j^t = p_j, 1 \leq j \leq N, t \geq 1$». Мы рассмотрим ситуацию, когда статистика $X(n)$ используется для построения критерия согласия с гипотезой H_0 в более общем случае.

2. Основная часть

Пусть выполнены следующие условия:

1. $\sqrt{n}\left(\frac{p_i^1+\ldots+p_i^n}{n} - p_i\right) \to 0$ при $n \to \infty$ для каждого фиксированного $1 \leq i \leq N$.

2. Существует предел $\lim_{n\to\infty}\left\|\frac{1}{n}\sum_{t=1}^{n}p_i^t p_j^t\right\|_{i,j=1}^{N}$, где предел понимается как поэлементный предел матриц.

Положим по определению $A = \left\|\delta_{ij} - \lim_{n\to\infty}\frac{\sum_{t=1}^{n}p_i^t p_j^t}{n\sqrt{p_i p_j}}\right\|_{i,j=1}^{N}$. Через $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ обозначим собственные значения матрицы A (вещественные в силу симметричности A), расположенные в порядке неубывания. Установлена следующая теорема.

Теорема. *Если выполнены условия 1 и 2, то* $\lambda_N = 0$ *и*

$$X(n) \xrightarrow{d} \sum_{i=1}^{N-1}\lambda_i\xi_i^2, \quad n\to\infty,$$

где $\xi_i \sim N(0,1)$ *— независимые стандартные нормальные величины, причем* $\lambda_1 \leq 1$.

Заметим, что величина $\sum_{i=1}^{N-1}\lambda_i\xi_i^2$, фигурирующая в условии теоремы 1, стохастически меньше (точнее, не больше) величины χ^2_{N-1}, так как $1 \geq \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N = 0$.

Пример. Пусть при каждом $1 \leq j \leq N$ величины p_j^t меняются периодическим образом с периодом L_j, т.е. $p_j^t = p_j^{(t \mod L_j)}$, тогда, как несложно видеть, условия 1 и 2 выполнены (при $p_j = \frac{p_j^1+\ldots+p_j^{L_j}}{L_j}$) и, следовательно, утверждение теоремы 1 верно. Отметим, что наличие периода L_j для каждого j равносильно наличию «общего» периода L, не зависящего от j.

3. Заключение

Теорема 1 означает, что, грубо говоря, «слишком большое» значение статистики $X(n)$, которое предписывает отвергнуть гипотезу вида «$p_j^t = p_j, 1 \leq j \leq N$», одновременно предписывает отвергнуть целый класс гипотез, которые однозначно определяются числами p_j^t, удовлетворяющими условиям 1–2.

Отметим, что очень похожие утверждения о поведении статистики хи-квадрат критерия можно получить и в ситуации, когда вместо условий 1–2 требуется строгая стационарность последовательности наблюдений (см. [3]). Более того, схожие эффекты (причем вместо неравенства в смысле стохастического порядка выполняется неравенство между дисперсиями) в другой ситуации отмечены в [2] на стр.301 (перед теоремой 1). Наконец, в работе [5] в задаче, близкой к рассматриваемой, установлено поведение статистики Пирсона в случае, когда частоты центрируются их математическими ожиданиями.

Благодарности

Автор благодарит А.М. Зубкова за постановку задачи и постоянное внимание к работе.

Литература

1. *Крамер Г.* Математические методы статистики. — М.: Мир, 1975.
2. *Samuels S.M.*, On the Number of Successes in Independent Trials // Ann. Math. Statist. — 1965. — Vol. 36, no. 4, — P. 1272–1278.
3. *Chanda K. C.* Chi-Square Goodness-of-Fit Tests Based on Dependent Observations // Statistical Distributions in Scientific Work, NATO Advanced study Institutes Series. — 1980. — Vol. 79, — P. 35–49.
4. *Balakrishnan N., Voinov N., Nikulin M.S.*, Chi-Squared Goodness of Fit Tests with Applications, 1st Edition. — Academic Press, 2013.
5. *Selivanov B.I.* Limit distributions of the χ^2 statistic of K. Pearson in a sequence of independent trials // Mathematical Notes. — 2008. — Vol. 83, no. 5–6. — P. 821–832.

UDC 519.234.33

Limit distributions of the Pearson statistics for nonhomogeneous polynomial scheme

M. P. Savelov[*][†]

[*] *Lomonosov Moscow State University*
[†] *Moscow Institute of Physics and Technology*

We consider nonhomogeneous polynomial scheme and find conditions on the probabilities of outcomes under which it is possible to describe limit distributions of the Pearson statistics.

Keywords: chi-square test, Pearson statistics, limit distribution.