Math-Net.Ru

Общероссийский математический портал

Е. Л. Емельянова, Л. П. Усольцев, Теоретико-числовая модификация статистики Пирсона, *Вестн. Сам. гос. техн. ун-та. Сер. Физ.-мат. науки*, 2005, выпуск 38, 15–18

DOI: <https://doi.org/10.14498/vsgtu364>

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: 35.173.219.12

26 апреля 2023 г., 13:19:00

Теория вероятностей

УДК 511.2+519.2

Е. Л. Емельянова, Л. П. Усольцев

ТЕОРЕТИКО – ЧИСЛОВАЯ МОДИФИКАЦИЯ СТАТИСТИКИ ПИРСОНА

Предлагается модификация статистики К. Пирсона на базе последовательностей, порождаемых дробными долями показательных функций с целым основанием. При таком подходе в процесс обработки наблюдений случайной выборки привносится эффект «увеличения» числа независимых серий наблюдений, что существенно, в частности, для малых выборок.

В основе наших построений лежат нижеследующие два фундаментальных свойства последовательностей вида $\{\alpha q^x\}$, $x = 0, 1, 2, \dots$, где $q \geq 2$ — целое число и $\alpha \in (0, 1)$.

1. При каждом целом $q \geq 2$ для почти всех (по мере Лебега) чисел $\alpha \in (0, 1)$ дробные доли $\{\alpha q^x\}$, $x = 0, 1, 2, \dots$ равномерно распределены в промежутке $[0, 1)$ (см. [1]), т. е. для любого промежутка $\Delta \subset [0, 1)$ длины $|\Delta| < 1$ при $P \rightarrow \infty$ выполняется соотношение

$$N_p(\Delta) = |\Delta| \cdot P + o(P),$$

где $N_p(\Delta)$ — число попаданий дробных долей $\{\alpha q^x\}$, $x = 0, 1, 2, \dots, P-1$ в промежуток Δ .

При этом в случае, когда $q = p^m$ есть натуральная степень простого числа, известна методика построения чисел $\alpha \in (0, 1)$ (в q -ичном представлении) с очень быстрым равномерным распределением дробных долей $\{\alpha q^x\}$, $x = 0, 1, 2, \dots$ в промежутке $[0, 1)$, конкретно, с

$$N_p(\Delta) = |\Delta| \cdot P + O(P^{1/3} \ln^{4/3} P) \quad (1)$$

при $P \rightarrow \infty$ (случай, когда $q = p$ — простое число, рассмотрен в [2], а случай, когда $q = p^m$ есть произвольная степень простого числа — в [3]; в [3] содержится также корректировка работы [2]).

2. Нарастающие суммы

$$\sum_{x=0}^{P-1} f(\alpha q^x) \quad (0 \leq \alpha < 1)$$

значений вещественнозначных периодических функций с периодом 1 (и определенными ограничениями на величину их коэффициентов Фурье) при $P \rightarrow \infty$ имеют асимптотически нормальное распределение (см. [4, 5]), а именно, если

$$a_0 = \int_0^1 f(\alpha) d\alpha \quad \text{и} \quad \sigma^2 = \lim_{P \rightarrow \infty} \int_0^1 \left(\frac{1}{\sqrt{P}} \sum_{x=0}^{P-1} (f(\alpha q^x) - a_0) \right)^2 d\alpha \quad (\sigma \geq 0),$$

то в случае $\sigma \neq 0$ при $P \rightarrow \infty$

$$\text{mes} \left\{ \alpha \in [0, 1): \frac{1}{\sigma \sqrt{P}} \sum_{x=0}^{P-1} (f(\alpha q^x) - a_0) < \lambda \right\} \rightarrow \Phi(\lambda), \quad (2)$$

где $\Phi(\lambda)$ — нормальная функция распределения с параметрами $(0, 1)$. При этом в наиболее важном частном случае, когда

$$f(t) = \chi_{\Delta}(t) = \begin{cases} 1 & \text{при } \{t\} \in \Delta, \\ 0 & \text{при } \{t\} \notin \Delta, \end{cases} \quad (0 \leq t < \infty), \quad (3)$$

т. е. когда

$$\sum_{x=0}^{P-1} f(\alpha q^x) = N_p(\Delta)$$

и $\sigma \neq 0$, скорость сходимости в предельном соотношении (2) такая же, как в центральной предельной теореме для нарастающих сумм независимых случайных величин (см. [6, 7]), т. е. при $P \rightarrow \infty$

$$\text{mes} \left\{ \alpha \in [0,1): \frac{1}{\sigma \sqrt{P}} (N_p(\Delta) - |\Delta| \cdot P) < \lambda \right\} = \Phi(\lambda) + O\left(\frac{1}{\sqrt{P}}\right).$$

Это означает, что последовательности $\{\alpha q^x\}$, $x=0,1,2,\dots$ $i=1,2,\dots,r$ можно трактовать как последовательности почти независимых случайных величин (в соответствующем вероятностном пространстве).

Все это создает предпосылки для использования конкретных последовательностей дробных долей $\{\alpha q^x\}$, $x=0,1,2,\dots$ (а также последовательностей вычетов $\alpha q^x \bmod m$, $x=0,1,2,\dots$ с достаточно большим m) в качестве материала для построения последовательностей случайных чисел (см., например, [8]) и теоретико-числовых модификаций различных алгоритмов, прямо или косвенно связанных с нормальным законом распределения.

Рассмотрим следующую статистическую задачу. Пусть наблюдается некоторая случайная величина X . Разобьем числовую прямую на r непересекающихся промежутков

$$\Gamma_i (i=1,2,\dots,r): \bigcup_{i=1}^r \Gamma_i = (-\infty, +\infty)$$

и при $i \neq j$ $\Gamma_i \cap \Gamma_j = \emptyset$. Пусть нас интересует не функция распределения случайной величины X , а только набор неизвестных вероятностей $p_i = P(X \in \Gamma_i)$ ($i=1,2,\dots,r$). Пусть при этом задан набор чисел p_i^0 ($i=1,2,\dots,r$) таких, что $\sum_{i=1}^r p_i^0 = 1$, и X^0 — случайная величина с каким-

либо известным законом распределения, причем $P(X^0 \in \Gamma_i) = p_i^0$ ($i=1,2,\dots,r$). Пусть, далее, $\bar{X} = (X_1, X_2, \dots, X_n)$ с $n \leq r$ — некоторая случайная выборка, отвечающая случайной величине X , и мы хотим знать, совпадают ли вероятности p_i , отвечающие наблюдениям выборки $\bar{X}$, с вероятностями p_i^0 (т. е. можно ли отождествить X с X^0), или нет.

Обозначим через H_0 гипотезу, состоящую в том, что при всех $i=1,2,\dots,r$ выполняется равенство $p_i = p_i^0$, через H_1 — противоположную гипотезу, состоящую в том, что хотя бы при одном $i \in \{1,2,\dots,r\}$ справедливо неравенство $p_i \neq p_i^0$, а через m_i — количество элементов выборки $\bar{X}$, наблюдения которых попали в промежуток Γ_i ($i=1,2,\dots,r$). Когда объем n выборки $\bar{X}$ растет, то частота $\frac{m_i}{n}$ сходится по вероятности к p_i . Поэтому гипотеза H_0 принимается,

если все частоты $\frac{m_i}{n}$ мало отличаются от p_i^0 . В соответствии с этим, в классическом варианте для проверки гипотезы H_0 используют статистику

$$\eta_{r,n} = \sum_{i=1}^r \frac{n}{p_i^0} \left(\frac{m_i}{n} - p_i^0 \right)^2 = \sum_{i=1}^r \frac{(m_i - np_i^0)^2}{np_i^0}, \quad (4)$$

называемую статистикой Пирсона (или статистикой χ^2) с $r-1$ степенями свободы. Мы же поступим несколько иначе. Будем модифицировать статистику (4), привнося в процесс обработки наблюдений случайной выборки $\bar{X}$ эффект «увеличения» числа независимых серий наблюдений.

Возьмем какое-либо целое число q , являющееся натуральной степенью простого числа, и, используя методику работ [2, 3], построим конкретное число $\alpha \in (0,1)$ (в q -ичном представлении) с быстрым равномерным распределением последовательности $\{\alpha q^x\}$, $x=0,1,2,\dots$ в промежутке $[0,1)$, возьмем промежуток $\Delta \subset [0,1)$ длины $|\Delta|=1/2$ и определим функцию $f(t)$ равенствами (3). Сопоставим каждому элементу X_j выборки $\bar{X}$ последовательность дробных

долей $\{\alpha q^{j+x-1}\}$, $x=0, 1, 2, \dots$ и, взяв произвольно целое число $s \geq 2$, образуем s подвыборок $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_s$ выборки $\bar{X}$, относя к $\bar{X}_k$ ($k=1, 2, \dots, s$) те элементы X_j ($j=1, 2, \dots, n$), для которых $\{\alpha q^{j+k-2}\} \in \Delta$. Так как при каждом $k=1, 2, \dots, s$ объем b_k подвыборки $\bar{X}_k$ равен $\sum_{j=1}^n f(\alpha q^{j+k-2})$, то при $n \rightarrow \infty$, вследствие (1), будет:

$$b_k \sim |\Delta| \cdot n = \frac{n}{2}. \quad (5)$$

Обозначим через a_{ik} ($i=1, 2, \dots, r$; $k=1, 2, \dots, s$) количество элементов подвыборки $\bar{X}_k$, наблюдения которых попали в Γ_i , и положим

$$a_i = \frac{2}{s} \sum_{k=1}^s a_{ik}. \quad (6)$$

Очевидно, что

$$a_{ik} = \sum_{j=1}^n \chi_{\Delta_i \cap \Delta}(\alpha q^{j+k-2}).$$

Когда $n \rightarrow \infty$, то, в силу закона больших чисел и того, что последовательность $\{\alpha q^x\}$ $x=1, 2, \dots$ можно трактовать как последовательность почти независимых величин, частота

$$\frac{1}{s} \sum_{k=1}^s \frac{m_{ik}}{b_k}$$

сходится по вероятности к p_i ($i=1, 2, \dots, r$). А так как, вследствие (5) и (6),

$$\frac{a_i}{n} = \frac{1}{s} \sum_{k=1}^s \frac{m_{ik}}{n/2} \sim \frac{1}{s} \sum_{k=1}^s \frac{m_{ik}}{b_k},$$

то и $\frac{a_i}{n}$ сходится по вероятности к p_i ($i=1, 2, \dots, r$).

Возьмем произвольно l промежутков $\Delta_u \subset [0, 1]$ с длинами $|\Delta_u| = 1/2$ ($u=1, 2, \dots, l$) и повторим наши рассуждения, беря вместо промежутка Δ промежутки $\Delta_1, \Delta_2, \dots, \Delta_l$; вместо величин $a_{i,k}$ и a_i получим соответственно два числовых набора $a_{i,k}^{(u)}$ и $a_i^{(u)}$ ($u=1, 2, \dots, l$) с ($i=1, 2, \dots, r$; $k=1, 2, \dots, s$). Для $u=1, 2, \dots, l$ введем величины

$$a^{(u)} = \frac{1}{r} \sum_{i=1}^r a_{ik}^{(u)}, \quad D^{(u)} = \sum_{i=1}^r (a_{ik}^{(u)} - a^{(u)})^2,$$

найдем $D = \min_{1 \leq u \leq l} D^{(u)}$ и определим число $\tilde{u}$ из условия $\sum_{u=1}^l (a_{ik}^{(\tilde{u})} - a^{(\tilde{u})})^2 = D$. Положим $\tilde{m}_i = a_i^{(\tilde{u})}$

($i=1, 2, \dots, r$). Так как при всех $i=1, 2, \dots, r$ частота $\frac{\tilde{m}_i}{n}$ сходится по вероятности к p_i , то для проверки гипотезы H_0 , наряду с классической статистикой (4), можно использовать и модифицированную статистику

$$\tilde{\eta}_{r,n,s} = \sum_{i=1}^r \frac{n}{p_i^0} \left(\frac{\tilde{m}_i}{n} - p_i^0 \right)^2 = \sum_{i=1}^r \frac{(\tilde{m}_i - np_i^0)^2}{np_i^0}, \quad (7)$$

с которой работаем далее точно также, как и со статистикой (4).

Заметим, что если бы мы взяли $\Delta = [0, 1]$ и $s=1$, то вместо статистики (7) получили бы классическую статистику (4). С увеличением s вычисления, естественно, усложняются, но зато возрастает и эффект «увеличения» числа независимых серий наблюдений, что существенно для выборок небольшого объема.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Weyl H. Über die Gleichverteilung von Zahlen mod Eins // Math. Ann., 1916. B. 17. S. 313–352.
2. Коробов Н. М. О распределении дробных долей показательных функций // Вестн/ МГУ: Серия матем., мех., 1966. №. 4. С. 42–46.

3. *Усольцев Л. П.* Задача на построение, связанная с равномерным распределением дробных долей показательных функций // Известия ВУЗов: Серия матем., 1967. № 12(67). С. 75–83.
4. *Fortet R.* Sur une suite également répartie // *Studia math.*, 1940. B. 1. P. 54–69.
5. *Кас М.* On the distribution of values of sums of type $\sum f(2^k t)$ // *Ann. Math.: Sec. ser.*, 1946. Vol. 47, No. 1. P. 33–49.
6. *Усольцев Л. П.* О больших отклонениях для распределения дробных долей показательной функции // *Обозрение прикл. и промышл. матем.: Серия вероятн. и статист.*, 1999. Т. 6, вып. 1. С. 209–210.
7. *Усольцев Л. П.* О больших отклонениях для классических распределений, порождаемых дробными долями показательной функции с целым основанием // *Вестн. Сам. гос. техн. ун-та. Сер.: Физ.–мат. науки*, 2004. № 16. С. 62–73.
8. *Таусворт Р.* Случайные числа, порождаемые линейными рекуррентными соотношениями по модулю 2 // *Киберн. сб.: нов. серия*, 1979. Вып. 16. С. 62–73.

Поступила 24.08.2005 г.