ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 30 МАТЕМАТИКИ Выпуск 1
2023

М. С. Тихов (Нижний Новгород, ННГУ). **Оценивание распределений в зависимости «доза-эффект» по выборкам случайного объема.**

УДК 519.2 DOI https://doi.org/10.52513/08698325_2023_30_1_1

Резюме: Рассматривается асимптотическое поведение оценок функции распределения в зависимости «доза-эффект» для случайного объема выборки когда для неслучайного объема выборки предельное распределение является нормальным. Показано, что в зависимости от распределения объема выборки в качестве предельных могут быть распределение Лапласа, а также t_2-распределение Стьюдента с двумя степенями свободы.

Ключевые слова: Зависимость «доза-эффект», центральная предельная теорема, случайные суммы, распределение Лапласа, t_2-распределение.

В задачах математической статистики объем выборки обычно считается детерминированным, фиксированным заранее. В то же время в немалом числе ситуаций встречаются задачи, где объем выборки τ является случайной величиной. Это: последовательная проверка гипотез [1], последовательное оценивание неизвестного параметра распределения [2], [3], где τ есть момент остановки, т. е. зависит от наблюдаемых случайных величин. В работах [3]–[6] случайный объем не зависит от наблюдаемых случайных величин. Случайность объема выборки приводит к тому, что предельные нормированные суммы (обычно предполагается, что $\mathbf{E}(\tau) \leqslant n$) могут уже не иметь нормального распределения — они могут иметь распределение Лапласа, или распределение Стьюдента [6], поэтому доверительный интервал для функции распределения или параметра (при одинаковой надежности интервала) надо брать шире, чем при фиксированном объеме выборки. В докладе мы будем рассматривать суммы случайного объема, когда τ не зависит от наблюдаемых величин. Обычно в качестве распределения для с.в. τ обычно рассматриваются либо геометрическое распределение, либо отрицательное биномиальное (**NB**). Заметим, что при последовательном оценивании параметра сдвига [3] равномерного на интервале $(\theta-1/2, \theta+1/2)$ оптимальный момент остановки имеет **NB**-распределение.

Пусть $T_m = T_m(X_1, \ldots, X_m), m \in \mathbf{N}$ — статистика, построенная по выборке $\mathcal{X}^{(m)} = (X_1, X_2, \ldots, X_m)$ неслучайного объема $m \in \mathbf{N}$. Рассмотрим последовательность дискретных случайных величин $\tau_1, \tau_n, \ldots, \tau_n, \ldots,$ зависящих от натурального n и принимающих натуральные значения. Натуральное значение $\tau_n = k \geqslant 1$ есть объем выборки $\mathcal{X}^{(k)}$. Будем предполагать, что при каждом n с. в. τ_n не зависит от $(X_1, X_2, X_3, \ldots)$. Определим T_{τ_n} полагая $T_{\tau_n} = T_k$ на множестве $(\tau_n = k), k \in \mathbf{N}$.

Предположения.

A1. *Существует функция распределения* $F(x)$, *такая, что для некоторого* $\gamma > 0$,
$$\sup_x |\, \mathbf{P}(m^\gamma T_m < x) - F(x) \,| \underset{m\to\infty}{\longrightarrow} 0.$$

A2. *Существует функция распределения* $H(x)$, $H(+0) = 0$ *и последовательность чисел* $0 < g_n \uparrow \infty$, *такие, что* $\sup_y |\, \mathbf{P}(g_n^{-1}\tau_n < y) - H(y) \,| \underset{n\to\infty}{\longrightarrow} 0.$

Как и в [6], можно показать, что имеет место следующий результат.

Теорема 1. *Пусть даны:* $\gamma = 2/5$, *статистика* T_m, *и случайный объем выборки* τ_n *и выполнены предположения* **A1** *и* **A2**. *Тогда*

$$\sup_{x\in\mathbf{R}} |\,\mathbf{P}(g_n^{\gamma}T_{\tau_n} < x) - G_n(x, 1/g_n)\,| \underset{n\to\infty}{\longrightarrow} 0, \text{ где } \quad G_n(x, 1/g_n) = \int_{1/g_n}^{\infty} F(xy^{\gamma}\, dH(y).$$

Если [6] рассматривались значения $\gamma \in \{-1, -1/2, 0, 1/2, 1\}$, то здесь $\gamma = 2/5$.

Пусть с. в. σ_n имеет геометрическое распределение $\mathbf{P}(\sigma_n \geqslant k) = q = \left(1 - \frac{1}{n^{\beta}}\right)^k$, $k = 1, 2, \ldots,$ а $\tau_n = \sigma_n^{1/\beta}, \beta > 0$. Тогда $\mathbf{P}(\tau_n \geqslant k) = \mathbf{P}(\sigma_n \geqslant k^{\beta}) = q^{k^{\beta}}$, т. е. τ_n имеет дискретное распределение Вейбулла и поэтому $\mathbf{P}(\tau_n/n \geqslant y) \underset{n\to\infty}{\longrightarrow} e^{-y^{\beta}} = 1 - H(y)$, $y > 0$, — непрерывное распределение Вейбулла. Возьмем $\beta = 2\gamma = 4/5$. Пусть $g_n = n$. Тогда [6]

$$\mathbf{P}(g_n^{\gamma}T_{\tau_n} \leqslant x) = \mathbf{P}(\tau_n^{\gamma}T_{\tau_n} \leqslant x(\tau_n/n)^{\gamma}) = \sum_{m=1}^{\infty} \mathbf{P}(m^{\gamma}T_m \leqslant x(m/n)^{\gamma})\,\mathbf{P}(\tau_n = m)$$

$$\approx \mathbf{E}(F(x(\tau_n/n)^{\gamma}) = \int_{1/g_n}^{\infty} F(xy^{\gamma}) d\mathbf{P}(\tau_n/n < y) \approx \int_{1/g_n}^{\infty} F(xy^{\gamma})\, dH(y).$$

Пусть $F(x) = \Phi(x)$ и $H(y) = 1 - \exp(-y^{\gamma})$, $y > 0$. В этом случае при указанном выборе $\mathbf{P}(g_n^{\gamma}T_{\tau_n} < x) \approx \int_0^{\infty} \Phi(xy^{\gamma})\, dH(y)$ с соответствующей предельной плотностью t_2 -распределения Стьюдента:

$$w_{\gamma}(x; s) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} y^{\gamma} e^{-x^2y^{2\gamma}/2}\, d(1 - \exp(-y^{\gamma})) = \frac{1}{2\sqrt{2}} \left(1 + \frac{x^2}{2}\right)^{-3/2}, \qquad x \in \mathbf{R}.$$

Рассмотрим еще следующее дискретное распределение с. в. τ_n :

$$\mathbf{P}(\tau_n \leqslant k) = \exp\left(-\frac{sn}{k^{\alpha}}\right), \qquad k, n \in \mathbf{N},\ \alpha > 0,\ s > 0. \tag{1}$$

Здесь $\mathbf{P}(\tau_n/n^{1/\alpha} \leqslant x) = \mathbf{P}(\tau_n \leqslant x/n^{1/\alpha}) \underset{m\to\infty}{\longrightarrow} e^{-s/x^{\alpha}} = H(y)$ и $g_n = n^{1/\alpha}$.
Положим $\alpha = 2\gamma$. Тогда

$$\mathbf{P}(\tau_n^{\gamma}T_{\tau_n} \leqslant x((\tau_n/g_n)^{\gamma} \approx \int_0^{\infty} \Phi(xy^{\gamma})\, d(e^{-s/y^{2\gamma}}),$$

с соответствующей предельной плотностью распределения (Лапласа):

$$w_{\gamma}(x; s) = \frac{s}{\sqrt{2\pi}} \int_0^{\infty} y^{\gamma}\, e^{-x^2y^{2\gamma}/2 - s/y^{2\gamma}}\, d\left(\frac{1}{y^{2\gamma}}\right) = \frac{s}{\sqrt{2\pi}} \int_0^{\infty} t^{-3/2}\, e^{-x^2t/2 - s/t}\, dt$$

$$= \frac{\sqrt{2s}}{2}\, e^{-\sqrt{2s}\,|\,x\,|}, \qquad x \in \mathbf{R}.$$

Пусть $\mathcal{X}^{(n)} = \{(X_i, U_i), 1 \leqslant i \leqslant n\}$ — последовательность независимых пар копий двумерных случайных величин (X, U) с совместной функцией распределения $F(x)Q(y), (x, y) \in \mathbf{R}^2$ и плотностью распределения $f(x)q(y)$. Мы наблюдаем выборку $\mathcal{U}^{(n)} = \{(U_i, W_i), 1 \leqslant i \leqslant n\}$, где $W_i = I(X_i < U_i)$ — индикатор события $(X_i < U_i)$. Требуется оценить функцию распределения $F_1(x)$ по выборке $\mathcal{U}^{(n)}$. Рассмотрим статистики

$$S_{1n}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - U_i}{h}\right), \quad S_{2n}(x) = \frac{1}{nh} \sum_{i=1}^{n} W_i K\left(\frac{x - U_i}{h}\right) \tag{2}$$

и оценку $\widehat{F}_n(x) = S_{2n}(x)/S_{1n}(x)$, полагая $\widehat{F}_n(x) = 0$, когда $S_{1n}(x) = 0$.

Здесь а) $K(x) \geqslant 0$; б) $\int_{-\infty}^{\infty} K(x)\,dx = 1$; в) $K(x) = K(-x)$; г) $||\,K\,||^2 = \int_{-\infty}^{\infty} K^2(x)\,dx$. При некоторых условиях регулярности (см. [7], [8]) и $h = n^{1/5}$ асимптотическое распределение нормированной разности $n^{2/5}(\widehat{F}_n(x) - \mathbf{E}(\widehat{F}_n(x)))/\sigma(x)$, где $\sigma^2(x) = \sigma_1^2(x)/q(x)$, $\sigma_1^2(x) = F(x)(1-F(x))||\,K\,||^2$, является нормальным с ф. р. $\Phi(x)$, и здесь мы имеем $\gamma = 2/5$.

Определим интервал $(x-\widetilde{h}, x+\widetilde{h})$ так, чтобы в него попало k случайных величин U_i. Тогда $\widetilde{h}$ будет случайным и при его подстановке в (2), мы получим kNN-оценки $\widetilde{F}_n(x)$ функции распределения $F(x)$. При выборе $k = n^{4/5}$ асимптотическое распределение разности $\sqrt{k}(\widetilde{F}_n(x) - \mathbf{E}(\widetilde{F}_n(x)))$ также является нормальным с ожиданием 0 и дисперсией $\sigma_1^2(x)$. Значит, если объем выборки τ_n случаен и имеет дискретное распределение Вейбулла, то из предыдущих рассуждений получаем, что предельным распределением нормированных разностей оценок имеет распределение Стьюдента t_2 с ф.р. $T_2(x) = \dfrac{1}{2} + \dfrac{x}{2\sqrt{2}}\,{}_2F_1\left(\dfrac{1}{2}, \dfrac{3}{2}; \dfrac{3}{2}; -\dfrac{x^2}{2}\right)$, а при выборе распределения (1) получим в качестве предельного распределение Лапласа, у которых более тяжелые хвосты, чем у нормального распределения. В подтверждение этого рассмотрим плотности:

$$\omega_\gamma(x;s) = \frac{s}{\sqrt{2\pi}} \int_0^\infty y^{\gamma-2}\, e^{-(x^2y^{2\gamma}/2 + s/y)}\,dy, \quad \rho_\lambda(x) = \frac{1}{\sqrt{2\pi}} \int_0^\infty y^\gamma\, e^{-(x^2y^{2\gamma}/2 + y)}\,dy,$$

и функции распределения: $\Omega_\gamma(x;s) = \int_{-\infty}^x \omega_\gamma(z;s)\,dz$, $L(x) = \frac{1}{\sqrt{2}} \int_{-\infty}^x e^{-\sqrt{2}\,|z|}\,dz$, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2}\,dz$, $R_\gamma(x) = \int_{-\infty}^x \rho_\lambda(z)\,dz$.

Тогда

$\Omega_{2/5}(2,01424) = 0,975$, $\Phi(1,96) = 0,975$, $L(2,19) = 0,75$, $R_{2/5}(3,54) = 0,75$.

СПИСОК ЛИТЕРАТУРЫ

1. *Wald A.* Sequential Test of Statistical Hypotheses. — Ann. Math. Statist., 1945, v. 16, № 2, p. 117–186.
2. *Линник Ю. В.*, *Романовский И. В.* К теории последовательного оценивания. — ДАН СССР, 1970, т. 194, № 2, с. 270–272.
3. *Тихов М. С.* Последовательное оценивание параметра сдвига равномерного распределения по цензурированным типа II выборкам. — Записки научн. семин. ЛОМИ, 1984, т. 136, с. 182–192.
4. *Renyi A.* On the central limit theorem for the sum of a random number of independent random variables. — Acta math. Acad, Sci. Hung., 1960, v. 11, p. 97–102.
5. *Blum J.*, *Hanson D.*, *Rosenblatt J.* On the Central Limit Theorem for the Sum of a Random Number of Independent Random Variables. — Z. Wahrscheinlichkeitsttheories 1, 1963, p. 389–393.
6. *Christoph G.*, *Ulyanov V.* Second Order Chebyshev-Edgeworth-Type Approximations for Statistics Based on Random Size Samples, — Mathematics, 2023, 11, 1848.
7. *Tikhov M.*, *Ivkin M.* Multivariate k-Nearest Neighbors Distribution Function Estimates in Dose-effect Relationship. — Proc. of the 2014 Intern. Conf. on Math. Models and Meth. in Appl. Sci. (MMAS'2014), 2014, p. 325–329.
8. *Tikhov M.* Statistical Estimation based on Interval Censored Data. — In: Param. and Semiparam. Models with Appl. to Rel., Surv. Analisys, and Qual. of Life: Springer-Verlag: Theor. & Meth., 2004, XLIV, p. 209–215.

UDC 519.2 DOI https://doi.org/10.52513/08698325_2023_30_1/00

Tikhov M. S. (Nizhny Novgorod, National Research Lobachevsky State University of Nizhny Novgorod) **Estimating distributions in «dose-effect» relationships on random size samples.**

Abstract: The asymptotic behavior of the distribution function estimates in as a function of «dose-effect» relationships for a random sample, when the asymptotical distribution for a non-random sample size is normal. It is shown that, depending on the distribution of the sample size, the Laplace distribution and the Student's t_2-distribution with two degrees of freedom can be used as limiting ones.

Keywords: Central limit theorem, «dose-effect» relationships, the sums of a random number, Laplace distribution, t_2-distribution.