ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 27 МАТЕМАТИКИ Выпуск 1
2020

М. С. Тихов (Нижний Новгород, ННГУ). **kNN-оценки функции распределения в модифицированном методе Рида и Менча.**

УДК 519.2

Резюме: Мы рассматриваем модифицированный метод k-ближайших соседей Рида-Менча для оценки функции распределения в зависимости «доза-эффект». Мы также обсуждаем связь между теоремой Фиеллера для отношения зависимых нормальных случайных величин и условными распределениями, используемыми для получения доверительных областей.

Ключевые слова: зависимость доза-эффект, kNN-метод Рида и Менча, теорема Фиеллера.

Пусть $\mathcal{U}^{(n)} = \{(U_i, W_i)_{i=1}^n\}$ есть независимые и одинаково распределенные случайные векторы, где случайные величины U_i имеют ограниченную плотность распределения $g(u) > 0$, $W_i = I(X_i < U_i)$ есть индикатор события $(X_i < U_i)$, а случайные величины X_i имеют функцию распределения $F(x)$ и плотность $f(x) > 0$. Задача состоит в том, чтобы оценить функцию распределения $F(x)$ по выборке $\mathcal{U}^{(n)}$. В работах [1], [2] для этих целей использовалась модификация метода Рида и Менча [3] (ядерное оценивание), приводящая к состоятельным и асимптотически нормальным оценкам функции распределения $F(x)$ с дисперсией, зависящей от плотности $g(x)$, поэтому точность оценивания на краях распределения, как правило, становится хуже. В настоящем докладе мы рассматриваем вариант этой модификации, когда параметр сглаживания является случайной величиной, а именно, мы рассматриваем kNN-оценки. Дисперсия этой оценки уже не зависит от плотности $g(x)$. Более того, если асимптотическое (по объему выборки) распределение является нормальным, то при конечном объеме выборки оно может отличаться от нормального распределения, поэтому в этой ситуации мы предлагаем использовать Fieller-метод.

Пусть $H_i = H((u_i - x)/h)$, $S_{1n}(x) = k^{-1}\sum_{i=1}^n w_i H_i$, $S_{2n}(x) = k^{-1}\sum_{i=1}^n (1 - w_i)(1 - H_i)$, $S_{3n}(x) = S_{1n}(x) + S_{2n}(x)$, где $H(x) = \int_{-1}^x K(t)\,dt$, для $-1 \leqslant x \leqslant 1$, $H(x) = 0$, если $x \notin [-1,1]$, $K(x)$ — четная неотрицательная функция, т. е. $K(x) = 0$, если $x \notin [-1,1]$, $\int_{-\infty}^{+\infty} K(x)\,dx = 1$, $||H||^2 = \int_{-\infty}^{+\infty} H^2(x)\,dx < \infty$. Мы определим также последовательность $k = k(n) = [n^{4/5}]$ натуральных чисел и пусть $h = h(n) = \xi_n^{(k)}$ — k-я порядковая статистика для $\{\xi_i = |\,U_i - x\,|\}_{i=1}^n$.

Рассмотрим оценку, определенную как $\hat{F}_n(x) = S_{1n}(x)/S_{3n}(x)$.

Теорема. *При некоторых условиях регулярности* (*см.* [4]),

$$\sqrt{k}\,(S_{1n}(x) - \mathbf{E}(S_{1n}(x))) \xrightarrow[n\to\infty]{d} N(0, g^2(x)F(x)\,||\,H\,||^2),$$

$$\sqrt{k}\,(S_{3n}(x) - \mathbf{E}(S_{3n}(x))) \xrightarrow[n\to\infty]{d} N(0, g^2(x)\,||\,H\,||^2).$$

Пусть даны две нормальных случайных величины Z_1 и Z_2 со средними a_1 и a_2, дисперсиями σ_1^2 и σ_2^2 соответственно, и коэффициентом корреляции ρ. Известно,

что если совместное распределение пары $(Z_1, Z_2)^T$ имеет плотность $f(x_1, x_2)$ по мере Лебега на R^2, то плотность распределения отношения
$\varsigma = Z_1/Z_2$ находится как $f_\zeta(z) = \int_{-\infty}^{+\infty} |y| f(yz, y)\, dy$ (см. также [5],[6])

$$f_\zeta(z) = \frac{\sqrt{1-\rho^2}}{\pi\sigma_1\sigma_2\gamma^2} \exp\left(-\frac{\tau^2}{2\,(1-\rho^2)}\right) + \frac{\eta\lambda}{\sqrt{2\pi}\sigma_1\sigma_2\gamma^3}\left(2\Phi\left(\frac{\eta}{\gamma\sqrt{1-\rho^2}}\right) - 1\right),$$

$$\gamma^2 = \frac{z^2}{\sigma_2^2} - 2\rho\frac{z}{\sigma_1\sigma_2} + \frac{1}{\sigma_1^2}, \quad \eta = \frac{z}{\sigma_2}\left(\frac{a_2}{\sigma_2} - \rho\frac{a_1}{\sigma_1}\right) + \frac{1}{\sigma_1}\left(\frac{a_1}{\sigma_1} - \rho\frac{a_2}{\sigma_2}\right),$$

$$\tau^2 = \frac{a_2^2}{\sigma_2^2} - 2\rho\frac{a_1 a_2}{\sigma_1\sigma_2} + \frac{a_1^2}{\sigma_1^2}, \quad \lambda = \exp\left(-\frac{1}{2\,(1-\rho^2)}\left(\frac{\eta^2}{\gamma^2} - \tau^2\right)\right),$$

$$\Phi\,(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-u^2/2} du.$$

Чтобы лучше объяснить нашу идею, рассмотрим специальный случай $\sigma_1^2 = \sigma_2^2 = 1$, $\rho = 0$ and $\alpha = \exp(-(a_1^2 + a_2^2)/2)$. Плотность $f_\varsigma(z)$ является унимодальной если точка (a_1, a_2) находится в высокой узкой области положительного квадранта, в противном случае $f_\varsigma(z)$ бимодальна. Причину этого можно понять заметив, что плотность $f_\varsigma(z)$ представляет собой смесь двух плотностей, одна из которых является плотностью Коши, а другая бимодальной плотностью. Мы имеем:

$$f_\zeta(z) = \alpha f_1(z) + (1-\alpha) f_2(z), \quad f_1(z) = 1/(\pi(1+z^2)),$$
$$f_2(z) = \frac{z a_2 + a_1}{\sqrt{2\pi}(1+z^2)^{3/2}} \frac{2\Phi((z a_2 + a_1)/\sqrt{1+z^2}}{\exp((a_1^2 + a_2^2)/2) - 1}.$$

Если мы возьмем $Z_1 \in N(\mathbf{E}(S_{1n}(x)), F(x) g^2(x) ||\, H\,||^2/k)$,
$Z_2 \in N(\mathbf{E}(S_{3n}(x)), g^2(x) ||\, H\,||^2/k)$, и $\zeta_n = (Z_1/Z_2 - \mathbf{E}(S_{1n}(x))/\mathbf{E}(S_{3n}(x))\sqrt{k}$, то получим

$$f_{\zeta_n}(y) \underset{n\to\infty}{\longrightarrow} f_0(y) = \frac{1}{\sqrt{2\pi F(x)(1-F(x))\,||\,H\,||^2}} \exp\left(-\frac{y^2}{2F(x)(1-F(x))\,||\,H\,||^2}\right).$$

Плотность $f_\zeta(y)$ и оценка $\hat{F}_n(x)$ используются при построении доверительных множеств для функции распределения $F(x)$.

СПИСОК ЛИТЕРАТУРЫ

1. *Тихов М. С.*, *Шкилева К. Н.* Модифицированный метод оценивания Рида и Менча в зависимости доза-эффект. — Вестник ТвГУ. Серия: Прикладная математика, 2019, в. 4, с. 5–26.
2. *Тихов М. С.*, *Шкилева К. Н.* Непараметрическое оценивание квантилей в модели бинарной регрессии. — Вестник ТвГУ. Серия: Прикладная математика, 2020, № 1, с. 5–19.
3. *Reed L.*, *Muench H.* A simple method of estimating fifty per cent endpoints. — The Amer. Journal of Hygiene, 1938, v. 27, № 3, p. 493–497.
4. *Тихов М. С.*, *Ярощук М. В.* Асимптотическая нормальность kNN-оценок в зависимости доза-эффект. — Вестник Нижегородского ун-та. Серия: Математика, 2006, № 1(4), с. 129–137.
5. *Fieller E.* The distribution of the Index in a Normal Bivariate Population. — Biometrika, 1932, v. 24 (3/4), p. 428–440.
6. *Hinkley D.* On the Ratio of Two Correlated Normal Random Variables. Biometrika, 1969, v. 56 (3), p.635–639.

УДК 519.2

Tikhov M. S. (Nizhny Novgorod, NNSU). **kNN-estimates of the distribution function in the modified Reed-Muench method.**

Abstract: We consider a modified k-nearest neighbors Reed-Muench method for estimating the distribution function in dose-effect relationships. We also discuss the relationship between the Filler theorem for the ratio of dependent normal random variables and conditional distributions used to obtain confidence regions.

Keywords: dose-effect relationship, kNN-method of Reed and Muench, Fieller-theorem.