ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 28 МАТЕМАТИКИ Выпуск 2
2021

М. С. Тихов, **Н. В. Капкаев** (Нижний Новгород, ИИТММ, ННГУ). **Оценивание функции распределения положительных случайных величин в зависимости доза-эффект.**

УДК 519.2

Резюме: Для задачи оценивания функции распределения ширина окна ядра работает не очень хорошо, когда базовая функция распределения требует разного сглаживания в разных точках. В этом сообщении мы предлагаем преобразовывать данные с той целью, чтобы глобальная ширина окна была более подходящей для функции распределения преобразованных данных. Оценка функции распределения исходных данных представляет собой «обратное преобразование» путем изменения переменных оценки функции распределения этих преобразованных данных.

Ключевые слова: преобразования Бокса-Кокса; оценки ядра; непараметрическая оценка функции распределения; выбор ширины окна.

Рассматривается модель распределения случайной величины X, заданная неизвестной непрерывной функцией распределения $F(x)$ с негладкой плотностью $f(x)$. Однако сама случайная величина X ненаблюдаема и вместо нее мы имеем повторную выборку пар $(U_1, W_1), (U_2, W_2), \ldots, (U_N, W_N)$, где $W_i = I(U_i > X_i)$ есть индикатор события $(U_i > X_i)$, а случайная величина U независима от X и имеет плотность распределения $g(u) > 0$ на носителе этого распределения. В качестве оценок функции распределения $F(x)$ предлагаются оценки типа Надарая-Ватсона (NW), именно,

$$F_N(x) = \frac{S_{2N}(x+\varepsilon)}{S_{1N}(x+\varepsilon)}, \quad \text{где} \quad S_{2N}(x) = \sum_{i=1}^{N} \frac{W_i}{N} K_h(U_i - x),\ S_{1N}(x) = \frac{1}{N}\sum_{i=1}^{N} K_h(U_i - x),$$

$K_h(x) = (1/h)K(x/h)$, $h = h(N)$ — ширина окна просмотра данных, $K(x)$ — неотрицательная, четная, финитная с носителем на $[-1,1]$ плотность (ядро), $\varepsilon = \varepsilon(N)$ — некоторый сдвиг ядра в оценке функции распределения.

Обычно в непараметрическом оценивании плотность распределения $f(x)$ предполагается гладкой. Нарушение предположения о гладкости $f(x)$ может иметь такие последствия как неправильный выбор ширины окна и асимптотическое смещение оценок. В предлагаемом подходе мы рассматриваем *положительные* случайные величины X и U, а вместо величин X и U мы будем рассматривать величины $Z = \ln X$ и $Y = \ln U$. При таком выборе преобразования мы используем постоянную ширину окна. Модели с положительными случайными величинами являются важными в актуарных науках, финансах, а также при изучении распределения времени жизни (см. [1]-[3]). Оценки распределений с помощью преобразования случайных величин впервые исследовались в [4]. В настоящем сообщении мы рассматриваем цензурированные случайной точкой выборки, или зависимость «доза-эффект» [5].

В качестве примера рассмотрим $F(x) = 1 - \exp(-x), x > 0$. Ниже приведены (для сравнения) оценки обычной NW-оценки (рис. 1) и NW-оценки после преобразования (рис. 2).

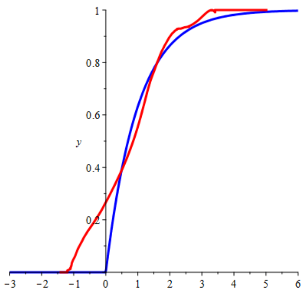

Рис. 1. Ядерная NW-оценка.

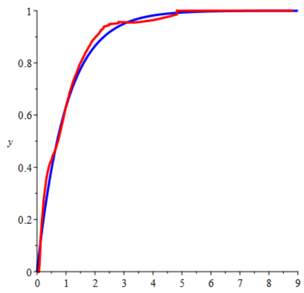

Рис. 2. NW-оценка после преобразовании.

При логарифмическом преобразовании имеем:

$$H_h\left(\frac{x}{u}\right) = K_h\left(\frac{\ln u - \ln x}{h}\right) = \frac{1}{h}K\left(\ln\left(\frac{u}{x}\right)^{1/h}\right), \quad u, x > 0.$$

В таком случае $S_{2N}(x) = \frac{1}{N}\sum_{i=1}^{N} W_i H_h\left(\frac{U_i}{x}\right)$, $\quad S_{1N}(x) = \frac{1}{N}\sum_{i=1}^{N} H_h\left(\frac{U_i}{x}\right)$.

Пусть $\quad \| K \|^2 = \int_{-1}^{+1} K^2(x)\,dx, \quad \mu_2(K) = \int_{-1}^{+1} x^2 K(x)\,dx,$

$$r_1(x) = \frac{xh^2}{2}\,(g(x) + 4xg'(x) + x^2 g''(x))\mu_2(K), \quad \sigma_1^2(x) = xg(x)\,\| K \|^2,$$

$$r_2(x) = \frac{xh^2}{2}\,(g(x)F(x) + 4x(g(x)F(x))' + x^2(g(x)F(x))'')\mu_2(K),$$

$$\sigma_2^2(x) = xg(x)F(x)\,\| K \|^2, \quad \sigma^2(x) = \frac{F(x)(1-F(x))\,\| K \|^2}{xg(x)}, \quad x > 0.$$

Теорема 1. *Если:*
($\mathbf{H}_1$) $h \to 0$, $\varepsilon \to 0$ *при* $N \to \infty$,
($\mathbf{F}_2$) $f(x)$ *и* $g(x)$ *имеют ограниченные непрерывные производные до второго порядка включительно на* $[0, \infty)$, *тогда*

$$\sup_{x>0} |\, F_N(x) - F(x)\,| \xrightarrow[N\to\infty]{p} 0.$$

Теорема 2. *Предположим, что выполнены следующие условия*:
($\mathbf{H}_2$) $Nh \to \infty$, $Nh\varepsilon \to \infty$, $Nh^2 \to 0$, $Nh\varepsilon^2 \to 0$, *при* $N \to \infty$,
($\mathbf{F}_3$) $f(x)$ *и* $g(x)$ *имеют третьи ограниченные производные на* $[0, \infty)$.
Тогда
(i) $\quad \sqrt{Nh}(S_{1N}(x) - xg(x)) \xrightarrow[N\to\infty]{d} \mathcal{N}(r_1(x), \sigma_1^2(x))$,
(ii) $\quad \sqrt{Nh}(S_{2N}(x) - xg(x)F(x)) \xrightarrow[N\to\infty]{d} \mathcal{N}(r_2(x), \sigma_2^2(x))$,
(iii) $\quad \sqrt{Nh}(F_N(x) - \mathbf{E}(F_N(x))) \xrightarrow[N\to\infty]{d} \mathcal{N}(0, \sigma^2(x))$.

Что касается оптимального выбора параметра h, то он осуществляется используя процедуру cross validation. Рассматривались также и другие преобразования, например, Box-Kox'а преобразование $(x^\lambda - 1)/\lambda$, когда $\lambda \neq 0$.

СПИСОК ЛИТЕРАТУРЫ

1. *Chaubey Y. P.*, *Li J.*, *Sen A.*, *Sen P. K.* A new smooth density estimator for non-negative random variables. — Journal of the Indian Statistical Association, 2007, v. 50, p. 83–104.

2. *Chaubey Y. P.*, *Sen P. K.* On nonparametric estimation of the density of a non-negative function of observations. — Calcutta Statistical Association Bulletin, 2013, v. 65, p. 75–101.
3. *Chaubey Y. P.* A new smooth density estimator for non-negative random variables. — Proceedings of the International Conference on Statistics: Theory and Applications (ICSTA'19), Lisbon, Portugal — August 13–14, 2019, Paper № 23.
4. *Wand M. P., Marron J. S., Ruppert D.* Transformations in Density Estimation. — Journal of the American Statistical Association, Jun. 1991, v. 86, № 414, p. 343–353.
5. *Криштопенко С. В.*, *Тихов М. С.*, *Попова Е. Б.* Доза-эффект. М.: Медицина, 2008, 285 с.

UDC 519.2

Tikhov M. S., ***Kapkaev N. V.*** (Nizhny Novgorod, Institute for Information Technology, Mathematics and Mechanics of Lobachevsky University). **Estimation of the distribution function of a non-negative random variables.**

Abstract: For the distribution function estimation problem window width kernel of the estimator does not perform well when the underlying distribution function has features that require different amounts of smoothing at different locations. In this report we propose to transform the data with the intention that a global window width is more appropriate for the distribution function of the transformed data. The distribution function estimate of the original data is the "back-transform" by change of variables of the global window width estimate of the transformed data's distribution function.

Keywords: Box-Kox transformations; kernel estimators; nonparametric distribution function estimation; window width selection.