ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 24 МАТЕМАТИКИ Выпуск 3
2017

М. С. Ивкин, **М. С. Тихов** (Нижний Новгород, ННГУ). **Фурье-метод для оценивания многомерной функции распределения по данным бинарных откликов с ошибками в переменных.**

Пусть $\mathcal{U}^{(n)} = \{(\mathbf{U}_1, W_1), \dots, (\mathbf{U}_n, W_n)\}$ обозначает случайную выборку, а $\mathcal{K}(\mathbf{x})$ есть ядерная функция, т. е. финитная симметричная плотность распределения. Пусть $\mathbf{U} \in \mathbf{R}^m$ — наблюдаемый, а $\mathbf{X}$ — ненаблюдаемый случайный вектор и $W = I(\mathbf{X} < \mathbf{U})$, $F(\mathbf{x}) = \mathbf{P}(\mathbf{X} < \mathbf{x})$. Задача исследования: по выборке $\mathcal{U}^{(n)}$ построить состоятельную оценку неизвестной функции распределения $F(\mathbf{x})$. Если $\mathbf{X}, \mathbf{U}$ независимы, то $\mathbf{E}(W|\mathbf{U} = \mathbf{x}) = F(\mathbf{x})$, поэтому в качестве оценки $F(\mathbf{x})$ возьмем $\widehat{F}_n(\mathbf{x}) = (n|\,\mathbf{H}\,|)^{-1} \sum_{j=1}^n W_j\, \mathcal{K}(\mathbf{H}^{-1}(\mathbf{U}_j - \mathbf{x}))/\widehat{g}_n(\mathbf{x})$, где

$$\widehat{g}_n(\mathbf{x}) = \frac{1}{n|\,\mathbf{H}\,|} \sum_{j=1}^n \mathcal{K}(\mathbf{H}^{-1}(\mathbf{U}_j - \mathbf{x}))$$

есть ядерная оценка плотности ковариат $\mathbf{U}$ с весовой матрицей $\mathbf{H} = (h_{ij})_{m\times m}$. Предположим [1], что вместо величин $\mathbf{U}_1, \dots, \mathbf{U}_n$ наблюдаются величины $\mathbf{Y}_j = \mathbf{U}_j + \mathbf{e}_j$, $j = 1, \dots, n$, поэтому оценку плотности $\widehat{g}_n(\mathbf{x})$ будем строить по наблюдениям $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ Пусть $F_{\mathbf{e}}(\mathbf{y})$ обозначает функцию распределения ошибок наблюдения $\mathbf{e}$, а $f_{\mathbf{e}}(\mathbf{y})$ ее плотность. Тогда $f_{\mathbf{Y}}(\mathbf{y}) = \int f_{\mathbf{U}}(\mathbf{y} - \mathbf{x})\, f_{\mathbf{e}}(\mathbf{x})\, d\mathbf{x}$.

Используя ядерную функцию $\mathcal{K}(\mathbf{x})$ рассмотрим следующую оценку

$$\widetilde{g}_n(\mathbf{x}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{-i\mathbf{t}^T\mathbf{x}\}\, \varphi_{\mathcal{K}}(\mathbf{H}\mathbf{t}) \frac{\widehat{\varphi}_n(\mathbf{t})}{\varphi_{\mathbf{e}}(\mathbf{t})}\, d\,\mathbf{t},$$

где $\varphi_{\mathcal{K}}(\mathbf{t})$ есть Фурье-преобразование ядра $\mathcal{K}(\mathbf{x})$, $\varphi_{\mathbf{e}}(\mathbf{t})$ есть характеристическая функция (х.ф.) ошибки $\mathbf{e}$ и $\widehat{\varphi}_n(\mathbf{t})$ — эмпирическая х. ф.: $\widehat{\varphi}_n(\mathbf{t}) = \sum_{j=1}^n \exp\{i\mathbf{t}^T\mathbf{Y}_j\}$, а $\widehat{\phi}_n(\mathbf{t}) = \sum_{j=1}^n W_j \exp\{i\mathbf{t}^T\mathbf{Y}_j\}$. Заметим, что оценку $\widetilde{g}_n(\mathbf{x})$ можно переписать в виде

$$\widetilde{g}_n(x) = \frac{1}{n\,|\,\mathbf{H}\,|} \sum_{1}^{n} \mathcal{K}_n(\mathbf{H}^{-1}(\mathbf{Y}_j - \mathbf{x})),\ \mathcal{K}_n(\mathbf{x}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\{-i\mathbf{t}^T\mathbf{x}\}\, \frac{\varphi_{\mathcal{K}}(\mathbf{t})}{\varphi_{\mathbf{e}}(\mathbf{H}^{-1}\mathbf{t})}\, d\,\mathbf{t}.$$

Учитывая это, мы используем следующую ядерную оценку функции распределения $F(\mathbf{x})$ по данным бинарных откликов, учитывающую ошибку в измерении доз, считая, что распределение ошибки известно:

$$\widetilde{F}_n(x) = \sum_{j=1}^n W_j \mathcal{K}_n(\mathbf{H}^{-1}(\mathbf{x} - \mathbf{Y}_j))/\widetilde{g}_n(\mathbf{x}).$$

Рассмотрим одномерный случай. Известно, что если $\varphi(t) = \mathcal{F}(f(x))(t)$, то $f(x) = \mathcal{F}^{-1}(\varphi(t))(x) = \mathcal{F}(\varphi(-t))(x)$, поэтому для наших целей достаточно иметь прямое преобразование Фурье. Для примера в качестве Фурье-преобразования ядра мы брали (с точностью до константы) $\varphi_{\mathcal{K}}(t) = \operatorname{sinc}^4(t))$. В таком случае оценку типа Надарая–Ватсона можно найти следующим образом:

$$F_n(x) = \mathcal{F}\left(\varphi_{\mathcal{K}}(th)\frac{\widehat{\phi}_n(t)}{\varphi_e(t)}\right)(x) \Big/ \mathcal{F}\left(\varphi_{\mathcal{K}}(th)\frac{\widehat{\varphi}_n(t)}{\varphi_e(t)}\right)(x). \tag{1}$$

Что касается скорости сходимости оценок (1), то для ошибки, имеющей распределение логарифмически-хи-квадрат распределение с одной степенью свободы нормирующий множитель есть $\dfrac{\sqrt{n}}{\exp\{2\pi/h\}}$ для оценки $\mathcal{F}\Big(\varphi_{\mathcal{K}}(th)\dfrac{\widehat{\varphi}_n(t)}{\varphi_e(t)}\Big)(x)$ (см. [2]–[4]). Мы показываем, что такой же нормирующий множитель будет и для числителя $\mathcal{F}\Big(\varphi_{\mathcal{K}}(th)\dfrac{\widehat{\phi}_n(t)}{\varphi_e(t)}\Big)(x)$ Отсюда получаем, что $\dfrac{\sqrt{n}}{\exp\{2\pi/h\}}(F_n(x)-F(x))$ будет асимптотически нормальна при $n\to\infty$ Если ошибка будет иметь нормальное распределение, то скорость сходимости оценок функции распределения $(\ln n)^{-1/2}$, а для двойного показательного распределения она равна $n^{-1/5}$.

СПИСОК ЛИТЕРАТУРЫ

1. *Fan J.* Multivariate regression estimation with errors in variables: asymptotic normality for mixing processes. — J. Multivar. Analysis, 1992, v. 43, p. 237–271.
2. *Es A.*, *Uh H.* Asymptotic normality of nonparametric kernel type deconvolution density estimators: crossing the Cauchy boundary. — J. Nonparametric Statist., 2004, v. 16, № 1–2, p. 261–277.
3. *Zhang C.-H.* Fourier methods for estimating mixing densities and distributions. — Ann. Stat., 1990, v. 18, № 2, p. 806–831.
4. *Zu Y.* A note on the asymptotic of the kernel deconvolution density estimator with logarithmic chi-square noise. — Econometrics, 2015, v. 3, № 3, p. 561–576.