УДК 519.2

ФУРЬЕ-МЕТОД РЕКУРСИВНОГО ОЦЕНИВАНИЯ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ В ЗАВИСИМОСТИ ДОЗА-ЭФФЕКТ

Тихов М.С.
Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского, г. Нижний Новгород

Поступила в редакцию 06.08.2018, после переработки 19.09.2018.

Предлагается ядерный Фурье-метод рекурсивного оценивания функции распределения в модели доза-эффект, когда вводимые дозы измеряются с ошибкой. Доказана асимптотическая нормальность предложенных оценок. Обсуждается возможность повышения точности оценивания за счет повторных измерений вводимых доз.

Ключевые слова: зависимость доза-эффект, сверточные ядерные оценки функции распределения, асимптотическая нормальность.

Вестник ТвГУ. Серия: Прикладная математика. 2018. № 4. С. 31–49.
https://doi.org/10.26456/vtpmk516

1. Введение

Рассматривается зависимость «доза-эффект», которая описывается следующим образом [1]. Имеется ненаблюдаемая (латентная) случайная величина (с.в.) X с функцией распределения $F(x) = \mathbf{P}(X < x)$ и плотностью $f(x)$, которые неизвестны. Имеется также случайная величина U, плотность распределения которой равна $g(x) > 0$ и которую мы также считаем неизвестной. Наблюдению доступны пары $\{(u_i, w_i), 1 \le i \le n\}$ – повторная выборка, где w_i есть значение случайной величины W_i – индикатора события $(X_i < U_i)$ (эффект), т.е. $W_i = I(X_i < U_i)$, а u_i – введенная в организм «доза». Случайная величина X – это нижняя граница, с которой начинается реакция организма: если $U > X$, то $W = 1$ (есть реакция организма), а если $U \leqslant X$, то $W = 0$ (реакция отсутствует). Требуется по наблюдениям $\{(u_i, w_i), 1 \le i \le n\}$ оценить функцию распределения $F(x)$. Если величины X и U – независимы, то условное ожидание $\mathbf{E}(W|U = x) = \mathbf{E}(I(U > X)|U = x) = \mathbf{E}(I(X < x)|U = x) = \mathbf{P}(X < x) = F(x)$ является регрессией и для оценки $F(x)$ можно использовать ядерную оценку регрессии вида $F_n(x) = S^*_{2n}(x)/S^*_{1n}(x)$, где $S^*_{1n}(x) = \frac{1}{n}\sum\limits_{i=1}^{n} K_b(u_i - x)$, $S^*_{2n}(x) = \frac{1}{n}\sum\limits_{i=1}^{n} w_i K_b(u_i - x)$, $K_b(x) = K(x/b)/b$, $K(x)$ есть ядерная функция, т.е. четная финитная плотность распределения с носителем распределения на отрезке $[-1, 1]$, а b есть параметр сглаживания, $b = b(n), b \to 0$ при $n \to \infty$, но $nb \to \infty$. В работах [2–4] показано, что для $b = n^{-1/5}$ разность $\zeta_n = n^{2/5}\,(F_n(x) - F(x))$

при $n \to \infty$ и некоторых условиях регулярности состоятельна и асимптотически нормальна $N(a(x), \rho^2(x))$ с ожиданием $a(x) = f'(x) + 2f(x)g'(x)/g(x)$ и дисперсией $\rho^2(x) = F(x)(1 - F(x)) \parallel K \parallel^2 /g(x), \parallel K \parallel^2 = \int K^2(x)\,dx$. Повысить скорость сходимости оценки функции распределения $F(x)$ с $n^{2/5}$ до $n^{4/9}$ можно, используя двухшаговую процедуру, предложенную и изученную в работе [5].

В работах [2–5] предполагалось, что вводимые дозы измеряются без ошибок. Однако реальные данные, которые мы наблюдаем, измеряются с ошибкой. Измерения с ошибкой возникают во многих научных областях. Игнорирование ошибки измерения может приводить к большому смещению оценки и к ошибочным заключениям при анализе данных. В данной работе мы будем предполагать, что измерения сделаны с ошибкой, которая является случайной величиной ε с неизвестной непрерывной плотностью распределения $f_\varepsilon(x)$. Иными словами, вместо выборки $(u_i, w_i), 1 \le i \leqslant n$ (*прямые наблюдения*) мы наблюдаем выборку $(y_i, w_i), 1 \le i \le n$, где $y_i = u_i + \varepsilon_i$, $W_i = I(X_i < U_i)$ (*непрямые наблюдения*). Такая модель наблюдений и статистика $\hat{F}_n(x) = \hat{S}_{2n}(x)/\hat{S}_{1n}(x)$, где $\hat{S}_{\nu n}(x) = (n)^{-1}\sum_{j=1}^{n} w_j^{2-\nu} K_b(y_j - x), \nu = 1, 2$, полагая $0^0 = 0$, изучались в работе [6]. Как при этом может измениться предельное распределение оценки $\hat{F}_n(x)$?
В [7] было показано, что оценка $\hat{F}_n(x)$ сходится по вероятности при $n \to \infty$ к свертке $R(x) = (F * f_\varepsilon)(x)$. К тому же при $b = n^{-1/5}$ последовательность $n^{2/5}(\hat{F}_n(x) - F(x))$ асимптотически нормальна с дисперсией $\hat{\rho}^2(x) = R(x)(1 - R(x)) \| K \|^2/(g * f_\varepsilon)(x)$. В [8] предлагались методы уменьшения ошибки измерения с помощью поправленных значений доз. Было показано, что оценки, построенные по поправленным значениям, являются состоятельными оценками $F(x)$ и асимптотически нормальными. Многомерный случай изучался в работах [8,9]. В работах Fan J. [10] и Zu Y. [11] исследовался Фурье-метод ядерного оценивания *плотности* в модели свертки. Там была доказана асимптотическая нормальность рассмотренных оценок.

В настоящей работе изучаются непараметрические *рекурсивные* оценки функции распределения ядерного Фурье-метода *в зависимости доза-эффект* для оценки вида $F_n = \hat{S}_{2n}(x)/\hat{S}_{1n}(x)$, $\hat{S}_{\nu n}(x) = (n)^{-1}\sum_{j=1}^{n} w_j^{2-\nu} K_{b_j}(y_j - x), \nu = 1, 2$, (полагая $0^0 = 0$), где $\{b_j\}$ есть последовательность сглаживающих параметров, и доказана их асимптотическая нормальность. Однако скорость сходимости оценок, построенных Фурье-методом, оказывается несколько ниже, чем при наблюдениях без ошибок, поэтому в данной работе для уменьшения погрешности измерения предложено при каждом из значений доз $u_j, j = 1, 2, ..., n$, производить по m повторных наблюдений.

2. Постановка задачи

Пусть $\mathscr{U}^{(n)} = (U_1, W_1),\ ..., (U_n, W_n)$ обозначает случайную выборку объема n и $K(x)$ есть ядерная функция. Случайная величина U наблюдаема, а X – ненаблюдаема, величины U, X – независимы. В качестве оценки $F(x)$ по $\mathscr{U}^{(n)}$ возьмем статистику

$$F_n(x) = \sum_{j=1}^{n} W_j K(b_j^{-1}(U_j - x)) / \sum_{j=1}^{n} K(b_j^{-1}(U_j - x)) = \tag{1}$$

$$= \frac{1}{n}\sum_{j=1}^{n} b_j^{-1} W_j K(b_j^{-1}(U_j - x))/S_{1n}(x) = S_{2n}(x)/S_{1n}(x),$$

где $S_{1n}(x) = n^{-1}\sum\limits_{j=1}^{n} K_{b_j}(U_j - x)$ есть ядерная оценка плотности распределения $g(x)$ случайной дозы U_j.

Рекурсивное вычисление оценки $F_n(x)$ может быть выполнено посредством соотношений: $S_{1,0}(x) = S_{2,0}(x) = F_0(x) = 0$ и для $n \geqslant 1$,

$$S_{1,n}(x) = \frac{n-1}{n} S_{1,n-1}(x) + \frac{1}{nb_n} K\left(\frac{U_n - x}{b_n}\right),$$

$$S_{2,n}(x) = \frac{n-1}{n} S_{2,n-1}(x) + \frac{1}{nb_n} W_n K\left(\frac{U_n - x}{b_n}\right),$$

так что

$$F_n(x) = \frac{(n-1)b_n S_{1,n-1}(x) F_{n-1}(x) + W_n K((U_n - x)/b_n)}{(n-1)b_n S_{1,n-1}(x) + K((U_n - x)/b_n)}.$$

Это рекурсивное свойство особенно полезно, когда выборка большого объема, так как $F_n(x)$ может быть легко обновлено каждым дополнительным наблюдением. Кроме того (см. [12]), при определенных обстоятельствах рекурсивная оценка будет более эффективна, чем нерекурсивный вариант.

Предположим теперь, что вместо величин $U_1, ..., U_n$ наблюдаются величины $Y_j = U_j + \varepsilon_j$, $j = 1, ..., n,$, где ε_j и U_j – независимы. Оценку плотности $\hat{g}_n(x)$ будем строить по наблюдениям $\mathscr{Y}^{(n)} = \{(Y_j, W_j)_{j=1}^n\}$. Пусть $f_\varepsilon(x)$ обозначает плотность распределения величины ε – погрешности измерения дозы. Тогда плотность распределения величины Y равна $f_Y(y) = (g * f_\varepsilon)(x) = \int g(y-x) f_\varepsilon(x)\, dx$, а соответствующие характеристические функции связаны соотношением $\varphi_{Y}(t) = \varphi_{U}(t)\,\varphi_\varepsilon(t)$. Если $\varphi_\varepsilon(t) \neq 0$, то $\varphi_{U}(t) = \varphi_{Y}(t)/\,\varphi_\varepsilon(t)$.

Пусть $\varphi_{K}(t)$ есть характеристическая функция (х.ф.) плотности $K(x)$. Тогда х.ф. для $b_j^{-1}K(x/b_j)$ равна $\varphi_{K}(b_j t)$, а х.ф. свертки функций $b_j^{-1}K(x/b_j)$ и $f_{Y}(x)$ равна $\varphi_{K}(b_j t)\varphi_{Y}(t)/\,\varphi_\varepsilon(t)$. В качестве оценки х.ф. $\varphi_{Y}(t)$ возьмем $\hat{\varphi}_n(t) = \hat{\varphi}_{nY}(t) = n^{-1}\sum_{j=1}^n \exp(itY_j)$. Если функция $\varphi_{K}(b_j t)\,\varphi_{Y_j}(t)/\varphi_\varepsilon(t)$ интегрируема, то получим следующую оценку плотности $g(u)$ (см. [10,11]):

$$\tilde{S}_{1n}(x) = = \frac{1}{2\pi n}\sum_{j=1}^{n}\int_{-\infty}^{+\infty} \exp\left(-it(x - Y_j)\right) \varphi_{K}(b_j t)\frac{\varphi_{Y_j}(t)}{\varphi_\varepsilon(t)}\, d\,t. \tag{2}$$

Оценку (2) можно переписать в виде

$$\tilde{S}_{1n}(x) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{b_j} K_n\left(\frac{x - Y_j}{b_j})\right), \tag{3}$$

с ядром

$$K_n(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp\left(-itx\right) \frac{\varphi_{K}(t)}{\varphi_\varepsilon(t/b_j)}\, dt. \tag{4}$$

Теперь рассмотрим следующую ядерную оценку Фурье-метода функции распределения $F(x)$ в зависимости доза-эффект, учитывающую ошибку в измерении доз:

$$\tilde{F}_n(x) = \tilde{S}_{2n}(x)/\tilde{S}_{1n}(x), \quad \text{где} \quad \tilde{S}_{2n}(x) = \frac{1}{n}\sum_{j=1}^{n} W_j b_j^{-1} K_n(b_j^{-1}(Y_j - x)). \qquad (5)$$

3. Условия

Условие 1.
(i) Характеристическая функция ошибки равна $\varphi_\varepsilon(t) = 2^{it}\Gamma(1/2+it)/\sqrt{\pi}$.
(ii) Плотность распределения $g(x) > 0$ есть ограниченная функция и имеет ограниченные и непрерывные производные до третьего порядка включительно.
(iii) Плотность распределения $f(x)$ ограничена и имеет ограниченные и непрерывные производные до второго порядка включительно.

Условие 2. Ядерная функция $K(x)$ – плотность распределения равна

$$K(x) = \operatorname{sinc}^{2m}(x)/c_m, m \in N, \quad \text{где} \quad c_m = \frac{\pi}{(2m-1)!}A(2m-1,m),$$

$A(n,k),\ 0 \le k \le n$ – числа Эйлера (см. [13], с.298).

Пусть $B_n = \sum_{j=1}^{n} b_j^{-1}$.
Условие 3. Последовательность b_j такова, что при $n \to \infty$

$$\text{(i) } b_j \to 0, j = 1,2,...,n; \quad \text{(ii) } n^2 B_n^{-1} \to \infty;$$

$$\text{(iii) найдется } \delta > 0 \text{ такое, что} \quad \frac{\sum_{j=1}^{n} b_j^{-\delta/2}\exp((2+\delta)/b_j)}{(\sum_{j=1}^{n}\exp((2/b_j))^{1+\delta/2}} \underset{n\to\infty}{\longrightarrow} 0.$$

Замечание к условию 1. Представленная характеристическая функция есть х.ф. величины, равной $\ln\chi_1^2$ (логарифм хи-квадрат с.в. с одной степенью свободы). В общем случае $\varepsilon = \ln\chi_\nu^2$ имеет плотность $(\nu = 1,2,...)$

$$f_\varepsilon(x) = \exp(\frac{1}{2}\nu(x - e^x))/(2^{\nu/2}\Gamma(\nu/2)), \quad -\infty < x < \infty,$$

и характеристическую функцию $\varphi_\varepsilon(t) = \dfrac{2^{\,it}\,\Gamma(\nu/2+it)}{\Gamma(\nu/2)\,\nu^{\,it+\nu/2}}$.

В [11] показано, что для $\nu = 1$ (с точностью до константы и при произвольном ν) и при $|\,t\,| \to \infty$,

$$|\,\varphi_\varepsilon(t)\,| = \sqrt{2}\exp\left(-\frac{1}{2}\pi|\,t\,|\right)\left(1 + O\left(\frac{1}{|\,t\,|}\right)\right). \qquad (6)$$

Кроме того, вместо величины ε будем рассматривать с.в. $\sigma_0\varepsilon, \sigma_0 > 0$.

Замечание к условию 2(i). Рассмотрим преобразование Фурье для некоторой плотности $f(x)$. Тогда

$$\mathcal{F}(f)(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{itx} f(x)\, dx = \varphi(t),$$

поэтому для обратного преобразования Фурье имеем:

$$\mathcal{F}^{-1}(\varphi(t))(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-itx} f(x)\, dx = \mathcal{F}(\varphi(-t))(x).$$

Рассмотрим непрерывное равномерное распределение $f(x) = 1/2$, для $|\,x\,| \le 1$. Ее характеристическая функция равна $\mathrm{sinc}(t)/2$ и ее четная степень также будет характеристической функцией свертки нескольких равномерных распределений, т.е. будет иметь распределение Ирвина-Хала $f_n(x) = (2(n-1)!)^{-1} \sum_{j=1}^{n} (-1)^k C_n^k \cdot$ $\cdot (x-k)^{n-1} \mathrm{sgn}(x-k)$. Тогда обратное преобразование для $\mathrm{sinc}^{2m}(x)/c_m$ будет равно нулю вне конечного интервала $[-A, A]$ (будем считать, что $\varphi_{_K}(t) = 0$ для $|\,t\,| > 1$.
Замечание к условию 3. Если все b_j равны h, то это условие сводится к требованию $h \to 0$; $nh \to \infty$ при $n \to \infty$.

4. Вспомогательные результаты

Лемма 1. *Если выполнены условия 1 – 3, то при* $n \to \infty$,

$$\mathbf{E}(\tilde{S}_{2n}(x)) = F(x)g(x) + \frac{1}{2n}\mu_2(K)(Fg)''(x) \sum_{j=1}^{n} b_j^2 (1 + o(1)), \tag{7}$$

$$\mathbf{E}(\tilde{S}_{1n}(x)) = g(x) + \frac{1}{2n}\mu_2(K) g''(x) \sum_{j=1}^{n} b_j^2 (1 + o(1)),$$

где

$$\mu_2(K) = \int_{-\infty}^{+\infty} u^2 K(u)\, du.$$

Доказательство. Рассмотрим сначала статистику $\tilde{S}_{2n}(x)$. Имеем:

$$\mathbf{E}(\tilde{S}_{2n}(x)) = n^{-1} \mathbf{E} \left(\sum_{j=1}^{n} b_j^{-1} K_n(b_j^{-1}(Y_j - x)) W_j \right) =$$

$$= (2\pi n)^{-1} \sum_{j=1}^{n} \mathbf{E} \left(W_j \int_{-\infty}^{+\infty} \exp(-itx)\, \varphi_K(b_j t) \frac{\hat{\varphi}_{Y_j}(t)}{\varphi_\varepsilon(t)}\, dt \right) =$$

$$= (2\pi n)^{-1} \int_{-\infty}^{+\infty} \sum_{j=1}^{n} \exp(-itx)\, \varphi_K(b_j t) \frac{\mathbf{E}(W_j \hat{\varphi}_{Y_j}(t))}{\varphi_\varepsilon(t)}\, dt =$$

$$= (2\pi n)^{-1} \sum_{j=1}^{n} \int_{R^2} \exp(-itx)\, \varphi_K(b_j t) \frac{\mathbf{E}(W_j \hat{\varphi}_{Y_j}(t) |\, U_j = u)}{\varphi_\varepsilon(t)}\, g(u)\, du\, dt =$$

$$= (2\pi n)^{-1} \sum_{j=1}^{n} \int_{R^2} \exp(-it(x-u)\,\varphi_K(b_j t)\,F(u)g(u)\,du\,dt =$$

$$= n^{-1} \sum_{j=1}^{n} h_j^{-1} \int_{-\infty}^{+\infty} K(b_j^{-1}(u-x))F(u)g(u)\,du.$$

Кроме того,

$$\mathbf{E}(\tilde{S}_{2n}(x)) = n^{-1} \sum_{j=1}^{n} \int_{-\infty}^{+\infty} K(u)F(x+b_j u)g(x+b_j u)\,du =$$

$$= F(x)g(x) + \frac{1}{2n}\mu_2(K)(Fg)''(x) \sum_{j=1}^{n} b_j^2(1+o(1)).$$

Значит, имеем представление (8).

Аналогично,

$$\mathbf{E}(\tilde{S}_{1n}(x)) = n^{-1} \sum_{j=1}^{n} b_j^{-1} \mathbf{E} \left(\sum_{j=1}^{n} K_n \left(b_j^{-1}(Y_j - x) \right) \right) =$$

$$= n^{-1} \sum_{j=1}^{n} b_j^{-1} \int_{-\infty}^{+\infty} K(b_j^{-1}(u-x))g(u)\,du = n^{-1} \sum_{j=1}^{n} \int_{-\infty}^{+\infty} K(u)g(x+b_j u)\,du =$$

$$= g(x) + \frac{1}{2n}\mu_2(K)g''(x) \sum_{j=1}^{n} b_j^2(1+o(1)).$$

Лемма доказана. □

Лемма 2. *При условиях 1 – 3 и при* $n \to \infty$,

(i) $\dfrac{n^2 \mathbf{Var}(\tilde{S}_{2n}(x))}{B_n \parallel K_n \parallel^2} = (Fg * f_\varepsilon)(x)\,(1+o(1))$; (ii) $\dfrac{n^2 \mathbf{Var}(\tilde{S}_{1n}(x))}{B_n \parallel K_n \parallel^2} = (g * f_\varepsilon)(x)\,(1+o(1))$.

Доказательство. Рассмотрим дисперсии представленных статистик. Имеем:

$$\mathbf{Var}(\tilde{S}_{2n}(x)) = n^{-2} \sum_{j=1}^{n} \mathbf{Var}(b_j^{-1} K_n(b_j^{-1}(Y_j - x))W)) =$$

$$= n^{-2} \sum_{j=1}^{n} b_j^{-2} \mathbf{E}(W_j^2 K_n^2(b_j^{-1}(Y_j - x))) - n^{-2} \sum_{j=1}^{n} \mathbf{E}^2(b_j^{-1} K_n(b_j^{-1}(Y_j - x))W_j) =$$

$$= n^{-2} \sum_{j=1}^{n} b_j^{-2} \mathbf{E}(W_j K_n^2(b_j^{-1}(Y_j - x)))(1+o(1)).$$

Теперь

$$n^{-2} \sum_{j=1}^{n} b_j^{-2} \mathbf{E}(W_j K_n^2(b_j^{-1}(Y_j - x))) =$$

$$= n^{-2}\sum_{j=1}^{n} b_j^{-2}\int_{R^2} F(u)K_n^2(b_j^{-1}(x-u-v))\,g(u)\,f_\varepsilon(v)\,dvdu =$$

$$= n^{-2}\sum_{j=1}^{n} b_j^{-1}\int_{R^2} K_n^2(z)F(x-v-b_jz)\,g(x-v-b_jz)\,f_\varepsilon(v)\,dvdz =$$

$$= n^{-2}\sum_{j=1}^{n} b_j^{-1}\parallel K_n\parallel^2\int_{-\infty}^{+\infty} F(x-v)\,g(x-v)\,f_\varepsilon(v)\,dv\,(1+o(1)) =$$

$$= n^{-2}\sum_{j=1}^{n} b_j^{-1}\parallel K_n\parallel^2(Fg*f_\varepsilon)(x)(1+o(1)).$$

Аналогично,

$$\mathbf{Var}(\tilde S_{1n}(x)) = n^{-2}\sum_{j=1}^{n}\mathbf{Var}(b_j^{-1}K_n(h_j^{-1}(Y_j-x))) =$$

$$= n^{-2}\sum_{j=1}^{n} b_j^{-2}\mathbf{E}(K_n^2(b_j^{-1}(Y_j-x)))(1+o(1)).$$

Кроме того,

$$\frac{1}{n^2}\sum_{j=1}^{n}\frac{1}{b_j^2}\mathbf{E}(K_n^2\left(\frac{Y_j-x}{b_j}\right)) = \frac{1}{n^2}\sum_{j=1}^{n}\frac{1}{b_j^2}\int\limits_{R^2} K_n^2\left(\frac{x-u-v}{b_j}\right)g(u)\,f_\varepsilon(v)\,dv =$$

$$= n^{-2}\sum_{j=1}^{n} b_j^{-1}\int_{R^2} K_n^2(z)\,g(x-v-b_jz)\,dF_\varepsilon(v)dz =$$

$$= n^{-2}\sum_{j=1}^{n} b_j^{-1}\parallel K_n\parallel^2\int_{-\infty}^{+\infty} g(x-v)\,f_\varepsilon(v)\,dv\,(1+o(1)) =$$

$$= n^{-2}\sum_{j=1}^{n} b_j^{-1}\parallel K_n\parallel^2(g*f_\varepsilon)(x)(1+o(1)),$$

что заканчивает доказательство леммы 2. □

Пусть теперь $\parallel f\parallel_p = \left(\int_a^b f^p(x)\,dx\right)^{1/p}$, $\parallel f\parallel_\infty = \sup\limits_{x\in[a,b]} f(x) = M,\ p>2.$

Тогда $\parallel f\parallel_p \le \parallel f\parallel_\infty^{1-2/p}\parallel f\parallel_2^{2/p}$. Действительно,

$$\int_a^b f^p(x)\,dx = \int_a^b f^{p-2}(x)\cdot f^2(x)\,dx \le M^{p-2}\int_a^b f^2(x)\,dx.$$

Отсюда $\parallel f\parallel_p \le M^{(p-2)/p}\left(\left(\int_a^b f^2(x)\right)^{1/2}\right)^{2/p} = \parallel f\parallel_\infty^{1-2/p}\parallel f\parallel_2^{2/p}.$

Для $p>1$ нам понадобится следующее неравенство:

$$|\,x+y\,|^p \le 2^{p-1}(\,|\,x\,|^p + \,|\,y\,|^p). \tag{8}$$

Действительно, не нарушая общности можно считать, что $x, y > 0$. Для $p>1$ функция x^p выпукла, поэтому $\left(\dfrac{x+y}{2}\right)^p \leqslant \dfrac{x^p+y^p}{2}$, отсюда получаем (8).

5. Асимптотическая нормальность статистик

Рассмотрим $\mathbf{E}(|\,\xi_j-\mathbf{E}(\xi_j)\,|^{2+\delta})$, где $\xi_j = \dfrac{1}{b_j}K_n\left(\dfrac{Y_j-x}{b_j}\right)$, тогда $S_1 = \tilde S_{1n} = = \dfrac{1}{n}\sum\limits_{j=1}^{n}\xi_j$, $S_2 = \tilde S_{2n} = \dfrac{1}{n}\sum\limits_{j=1}^{n} W_i\xi_j$.

Используя неравенство (8), имеем:

$$|\,\xi_j-\mathbf{E}(\xi_j)\,|^{2+\delta} \le 2^{1+\delta}(\,|\,\xi_j\,|^{2+\delta} + |\,\mathbf{E}(\xi_j)\,|^{2+\delta}).$$

Беря от обеих частей математическое ожидание, получаем

$$\mathbf{E}(|\,\xi_j-\mathbf{E}(\xi_j)\,|^{2+\delta}) \leqslant 2^{2+\delta}\mathbf{E}(|\,\xi_j\,|^{2+\delta}) =$$

$$= (2/b_j)^{2+\delta}\int_{-\infty}^{+\infty}\left|\,K_n\left(b_j^{-1}(x-y)\right)\right|^{2+\delta} f_Y(y)\,dy =$$

$$= \int_{-\infty}^{+\infty}|K_n(u)|^{2+\delta} f_Y(x+ub_j)\,du = \frac{2^{2+\delta}}{b_j^{1+\delta}} f_Y(x)\int_{-\infty}^{+\infty}|\,K_n(u)\,|^{2+\delta}\,du(1+o(1)),$$

так как $|f'_Y(x)| \le C_1$ и $|f''_Y(\xi)| \le C_2$.

Далее,

$$\int_{-\infty}^{+\infty}|\,K_n(u)\,|^{2+\delta}\,du \le \|\,K_n\,\|_2^{2+\delta},$$

$$\|\,K_n\,\|_2^2 = \int_{-\infty}^{+\infty}|\,K_n(x)\,|^2\,dx = \frac{1}{4\pi^2}\int_{-\infty}^{+\infty}\left|\int_{-\infty}^{+\infty} e^{-itu}\frac{\varphi_K(t)}{\varphi_\varepsilon(t/b_j)}\,dt\right|^2 du \le$$

$$\le \frac{1}{4\pi^2}\int_{-\infty}^{+\infty}\left|\int_{-1}^{1}\frac{1}{\varphi_\varepsilon(t/b_j)}\,dt\right|^2 du \le \frac{1}{2\pi^2 b_j}\exp\left(\frac{2\sigma_0}{b_j}\right), \tag{9}$$

так как $\varphi_K(t)$ равна нулю вне интервала $[-1,1]$ и мы использовали (7).

Значит,

$$\|\,K_n\,\|_2^{2+\delta} \le \frac{1}{(2\pi^2 b_j)^{1+\delta/2}}\exp(\sigma_0(2+\delta)/b_j)(1+o(1)). \tag{10}$$

Теорема 1. *Пусть выполнены условия 1 – 3. Тогда*

$$\frac{\sqrt{2}\pi n(\tilde F_n(x) - \boldsymbol{E}(\tilde F_n(x)))}{c_n} \xrightarrow[n\to\infty]{d} N(0,\Sigma^2(x)),$$

где

$$\Sigma^2(x) = \frac{1}{g^2(x)}[(1-2F(x))(Fg*f_\varepsilon)(x) + F^2(x)(g*f_\varepsilon)(x)], \quad c_n^2 = \sum_{j=1}^{n} b_j^{-1}\exp(2\sigma_0/b_j).$$

Доказательство. Рассмотрим $\mathbf{Var}(\xi_j)$. Нетрудно показать, что

$$\mathbf{Var}(\xi_j) = \mathbf{E}(\xi_j^2) - \mathbf{E}^2(\xi_j) \sim \mathbf{E}(\xi_j^2) \sim \frac{f_Y(x)}{b_j} \int_{-\infty}^{+\infty} |\, K_n(u)\,|^2\, du =$$

$$= \frac{g * f_\varepsilon(x)}{4\pi^2} \int_{-\infty}^{+\infty} \left| \int_{-\infty}^{+\infty} e^{-itu} \frac{\varphi_K(t)}{\varphi_\varepsilon(t/b_j)}\, dt \right|^2 du \sim \frac{1}{2\pi^2 b_j} \exp\left(\frac{2\sigma_0}{b_j}\right).$$

В таком случае

$$\mathbf{Var}(\tilde{S}_{1n}(x)) = \frac{g * f_\varepsilon(x)}{2\pi^2 n^2} \sum_{j=1}^{n} b_j^{-1} \exp(2\sigma_0/b_j) = \frac{g * f_\varepsilon(x)}{2\pi^2 n^2} c_n^2$$

(если $b_j = h, j = 1, 2, ..., n$, то $\mathbf{Var}(S_{1n}(x)) = \dfrac{g * f_\varepsilon(x)}{2\pi^2 nh} \exp(2\sigma_0/h)$).

Значит, для дроби Ляпунова (с точностью до константы) имеем представление:

$$L_n = \frac{\sum_{j=1}^{n} \mathbf{E}(|\, \xi_j - \mathbf{E}(\xi_j)\,|^{2+\delta})}{(\sum_{j=1}^{n} \mathbf{Var}(\xi_j))^{1+\delta/2}} \sim \frac{\sum_{j=1}^{n} b_j^{-\delta/2} \exp((2+\delta)/b_j)}{(\sum_{j=1}^{n} \exp((2/b_j))^{1+\delta/2}} \underset{n\to\infty}{\longrightarrow} 0,$$

т.е. условие центральной предельной теоремы Ляпунова [14] выполнено и

$$\frac{\tilde{S}_{1n}(x) - \mathbf{E}(\tilde{S}_{1n}(x))}{\sqrt{\mathbf{Var}(\tilde{S}_{1n}(x))}} \underset{n\to\infty}{\overset{d}{\longrightarrow}} N(0,1).$$

Подобным же образом, выводим, что

$$\frac{\tilde{S}_{2n}(x) - \mathbf{E}(\tilde{S}_{2n}(x))}{\sqrt{\mathbf{Var}(\tilde{S}_{2n}(x))}} \underset{n\to\infty}{\overset{d}{\longrightarrow}} N(0,1).$$

Далее будем считать что $b_1 = b_2 = ... = b_n = h$, то $L_n \sim Cn^{-\delta/2}$ и для $\delta = 1$, $L_n \sim C/\sqrt{n}$.

Теперь положим $a_2 = \mathbf{E}(S_2)$, $a_1 = \mathbf{E}(S_1)$ и установим асимптотическую нормальность отношения S_2/S_1.

Учитывая неравенство $\left| \dfrac{1}{1+x} - 1 + x \right| \le 2x^2$, $|\, x\,| \le \dfrac{1}{2}$, имеем:

$$\frac{S_2}{S_1} = \frac{S_2 - a_2 + a_2}{S_1 - a_1 + a_1} = \frac{S_2 - a_2}{a_1\left(1 + \frac{S_1 - a_1}{a_1}\right)} + \frac{a_2}{a_1\left(1 + \frac{S_1 - a_1}{a_1}\right)} =$$

$$= \frac{S_2 - a_2}{a_1} \cdot \left(1 - \frac{S_1 - a_1}{a_1}\right) + \frac{a_2}{a_1} - \frac{a_2}{a_1} \cdot \frac{S_1 - a_1}{a_1} + q_{1n},$$

$$|\, q_{1n}\,| \le 2\frac{(S_1 - a_1)^2}{a_1^2} \cdot \frac{|\, S_2 - a_2\,|}{a_1} + 2\frac{a_2}{a_1} \cdot \frac{(S_1 - a_1)^2}{a_1^2}.$$

Тогда

$$e^{-\sigma_0/h}\left(\frac{S_2}{S_1}-\frac{a_2}{a_1}\right)=\exp(-\sigma_0/h)\left(\frac{S_2-a_2}{a_1}-\frac{S_1-a_1}{a_1}-q_{2n}\right),$$

где $\quad q_{2n}=(S_2-a_2)(S_1-a_1)/a_1^2.$

Имеем: $e^{-\sigma_0/h}(S_2-a_2)\xrightarrow[n\to\infty]{a.s.}0$. Действительно, достаточно показать (см. [15], с.215), что $\mathbf{P}\left(\bigcup\limits_{m=n}^{\infty}\{e^{-\sigma_0/h}|\,S_{2m}-\mathbf{E}(S_{2m})\,|\geq\varepsilon\}\right)\to 0$ *для любого* $\varepsilon>0$.

Так как

$$\mathbf{P}\left(\bigcup_{m=n}^{\infty}\{e^{-\sigma_0/h}|\,S_{2m}-\mathbf{E}(S_{2m})\,|\geq\varepsilon\}\right)\leq\sum_{m=n}^{\infty}\mathbf{P}\left(e^{-\sigma_0/h}|\,S_{2m}-\mathbf{E}(S_{2m})\,|\geq\varepsilon\right)$$

и по неравенству Чебышева

$$\mathbf{P}\left(e^{-\sigma_0/h}|\,S_{2m}-\mathbf{E}(S_{2m})\,|\geq\varepsilon\right)\leq\mathbf{E}((S_{2m}-\mathbf{E}(S_{2m}))^4)/(\varepsilon^4e^{4\sigma_0/h}),$$

то рассмотрим $\mathbf{E}((S_{2m}-\mathbf{E}(S_{2m}))^4)$.

Используя неравенство (10), имеем:

$$\mathbf{E}((S_{2n}-a_2)^4)=\mathbf{E}\left(\left(n^{-1}\sum_{j=1}^{n}(W_j\xi_j-a_2)\right)^4\right)\leq$$

$$\leq\frac{1}{n^3}\mathbf{E}(\xi_1-\mathbf{E}(\xi_1)^4)+\frac{6(n-1)}{n^3}\mathbf{E}^2(\xi_1))^2\leq$$

$$\leq e^{4\sigma_0/h}\left(\frac{C_1}{n^3h^2}+\frac{C_2}{n^2h^2}\right)\sim e^{4\sigma_0/h}\frac{C_2}{n^2h^2}.$$

Если $h=n^{-1/5}$, то $n^2h^2=n^{8/5}$ и так как ряд $\sum\limits_{n=1}^{\infty}\dfrac{1}{n^{8/5}}$ сходится, то $\sum\limits_{n=m}^{\infty}\dfrac{1}{n^{8/5}}\underset{m\to\infty}{\longrightarrow}0$, откуда $e^{4\sigma_0/h}(S_{2n}-a_2)\xrightarrow[n\to\infty]{a.s.}0$.

Кроме того, применяя метод Колмогорова (см.[16], с.106), можно показать, что

$$\mathbf{P}\left(\sqrt{nh}\max_{1\leq k\leq n}e^{2\sigma_0/h}(S_{1k}-a_1)^2\geq\lambda\right)\leq\frac{nhe^{-2\sigma_0/h}\mathbf{E}((S_{1n})-a_1)^4}{\lambda^2}\sim Cn^{-4/5}\underset{n\to\infty}{\longrightarrow}0.$$

Действительно, обозначим

$A=\{\max\limits_{1\leq k\leq n}(S_{1k}-a_1)^2\geq\lambda\}$,

$A_k=\{(S_{1j}-a_1)^2<\lambda, j=1,2,...,k-1,(S_{1k}-a_1)^2\geq\lambda\}, 1\leq k\leq n$.

Тогда $A=\bigcup\limits_{k=1}^{n}A_k$ и, полагая для простоты обозначений $a_1=0$, имеем:

$$\mathbf{E}(S_{1n}^4)\geq\mathbf{E}((S_{1n}^4)I_A)=\sum_{k=1}^{n}\mathbf{E}((S_{1n}^4)I_{A_k}).$$

Но

$$\sum_{k=1}^{n}\mathbf{E}((S_{1n}^4)I_{A_k})=\mathbf{E}((S_{1k}I_{A_k}+(1/n)(\sum_{j=k+1}^{n}\xi_j)I_{A_k})^4)=\mathbf{E}(S_{1k}^4I_{A_k})+$$

$$+6\mathbf{E}(S_{1k}^2((1/n)(\sum_{j=k+1}^{n}\xi_j)^2)I_{A_k})+\mathbf{E}(((1/n)(\sum_{j=k+1}^{n}\xi_j)^4)I_{A_k})\geq\mathbf{E}(S_{1k}^4I_{A_k})\geq\lambda^2\mathbf{P}(A_k).$$

Значит,

$$\mathbf{E}(S_{1n}^4)\geq\lambda^2\mathbf{P}(A).$$

Кроме того, при $n\to\infty$,

$$nhe^{-4\sigma_0/h}\mathbf{E}(S_{1n}^4)\sim\frac{C_1}{n^2h}+\frac{6C_2(n-1)}{n^2}=\frac{1}{n}\left(\frac{C_1}{nh}+C_2\right).$$

Окончательно получаем:

$$\sqrt{nh}e^{-2\sigma_0/h}\max_{1\leq k\leq n}(S_{1k}-a_1)^2\xrightarrow[n\to\infty]{p}0.$$

Далее,

$$\mathbf{E}\left(\frac{S_2}{S_1}-F(x)\right)\sim\mathbf{E}\left(\frac{S_2-F(x)g(x)}{g(x)}-F(x)\cdot\frac{S_1-g(x)}{g(x)}\right)=h^2\beta_2(x)+o(h^2),$$

где

$$\beta_2(x)=\frac{1}{2g(x)}((F(x)g(x))''-F(x)g''(x))=\frac{1}{2g(x)}\left(f'(x)+2\frac{f(x)}{g(x)}g'(x)\right).$$

Кроме того,

$$e^{-2\sigma_0/h}nh\left(\frac{\mathbf{D}(S_2)}{a_1^2}+\frac{a_2^2}{a_1^4}\cdot\mathbf{D}(S_1)-2\cdot\frac{a_2}{a_1^3}\cdot\mathbf{Cov}(S_2,S_1)\right)\sim$$

$$\sim\frac{(Fg*f_\varepsilon)(x)}{g^2(x)}+\frac{F^2(x)g^2(x)}{g^4(x)}\cdot(g*f_\varepsilon)(x)-2\cdot\frac{F(x)g(x)}{g^3(x)}(Fg*f_\varepsilon)(x)=$$

$$=\frac{1}{g^2(x)}[(1-2F(x))(Fg*f_\varepsilon)(x)+F^2(x)(g*f_\varepsilon)(x)]=\Sigma^2(x),$$

поскольку $e^{-2\sigma_0/h}nh\,\mathbf{Cov}(S_2,S_1)\sim(Fg*f_\varepsilon)(x)\|\,K_n\,\|^2$.

Теперь из наших рассуждений выводим, что при $n\to\infty$ величина $\dfrac{\hat{F}_n(x)-\mathbf{E}(F_n(x))}{\sqrt{\mathbf{Var}(F_n(x))}}$ асимптотически нормальна с параметрами $(0,\Sigma^2(x))$. □

С практической точки зрения интерес представляет оценка скорости сходимости в этой теореме. Известно, что если $\zeta_1,\zeta_2,...,\zeta_n$ независимы и одинаково распределены, $S_n^*=\zeta_1+\zeta_2+...+\zeta_n$, $\rho^2=\mathbf{Var}(\zeta_j)$, то в равномерной метрике

$$\sup_{x\in R}|\,\mathbf{P}(S_n^*\leqslant x)-\Phi(x)\,|\leqslant\frac{C}{\sqrt{n}\sigma^3}\mathbf{E}(\,|\,\zeta_1\,|^3\,).$$

Применительно к нашему исследованию получаем, что как $\tilde{S}_{1n}$, так и $\tilde{S}_{2n}$ (а, значит, и $\hat{F}_n$) сходятся к нормальному распределению в равномерной метрике со скоростью $C_1/\sqrt{n}$. Это следует как из неравенства Берри-Ессеена, так и из метода Стейна (см. [17] – [19]).

Вообще-то, метод Стейна в первоначальном виде [17] предназначался для оценивания точности аппроксимации распределений сумм зависимых случайных величин, в частности, m-зависимых величин, для которых граница имеет вид [19]:

$$d_W\left(\frac{S_n^*}{\mathbf{Var}(S_n^*)}, \zeta_0\right) \leqslant \frac{7(2m+1)^2 \cdot \max_j \mathbf{E}(|\,\zeta_j\,|)}{\sqrt{n}\,\mathbf{E}(|\,S_n^*\,|)},$$

где $\zeta_0 \in N(0,1)$, а $d_W(\mathbf{P},\mathbf{Q}) = \sup\left\{\left|\int \omega\, d\mathbf{P} - \int \omega\, d\mathbf{Q}\right|, \omega \in \mathrm{Lip}(1)\right\}$ есть метрика Вассерштейна [20] (по-сути, это метрика Канторовича-Рубинштейна [21], [22]). Иными словами, с помощью метода Стейна оценку скорости сходимости можно расширить до m-зависимых случайных величин.

6. Повторные наблюдения

Отметим, что в теореме 1 мы рассматривали характеристическую функцию ядра $\varphi_K(t)$ вида (7). Такие же выводы мы получим, если возьмем $\varphi_K(t) = \exp(-\gamma|\,t\,|)$.

Предположим, что при каждом значении u_j наблюдается n_j значений y_{jl}, $l = 1, 2, ..., n_j$ [23] – [25]. Примем все n_i равными m. В таком случае мы имеем $N = nm$ наблюдений. Предположим, что $B = [n^\alpha]$ (для простоты будем считать $m = n^\alpha$). Тогда $N = n^{1+\alpha}$.

Вместо y_{jl}, $l = 1, 2, ..., n_j$ рассмотрим $\bar{y}_j = \dfrac{1}{m}\sum_{l=1}^m y_{jl}$. Пусть $f_\varepsilon^{(*n)}(x)$ – n–я свертка $f_\varepsilon(x)$ – четной плотности распределения ошибки ε. Тогда $m\, f_\varepsilon^{(*n)}((\bar{y}_j - u_j))$ есть плотность распределения $\bar{\varepsilon}_m$. Будем считать, что плотность распределения $f_\varepsilon(x)$ известна, положим $m = [1/h]$, а в ядре $K_n(x)$ возьмем $f_\varepsilon^{(*n)}(x)$. Тогда мы будем иметь ядерную оценку с нормировкой $\sqrt{nh}$.

7. Оптимальная скорость сходимости оценок

Какова оптимальная скорость сходимости рассматриваемых оценок?

Определим $\mathbf{MISE}(\tilde{S}_{l,n})$ для статистики $\tilde{S}_{l,n}(x), l = 1, 2$, как оценки функции $g(x)$: $\mathbf{MISE}(\tilde{S}_{l,n}) = \int \mathbf{E}(\tilde{S}_{l,n}(x) - g(x))^2\, dx$.

Определим класс функций $\mathcal{F}_{\alpha,C}$ (см. [21]):

$g \in \mathcal{F}_{\alpha,C} = \{$плотности $g \in \mathcal{C}_\alpha, \|\, g^{(\alpha)}\,\|_\infty < C$ и $\int (g^{(\alpha)})^2 < C\}$,

где $\mathcal{C}_\alpha$ есть класс α раз непрерывно дифференцируемых функций.

Пусть ошибка ε имеет характеристическую функцию $\varphi_\varepsilon(t)$ такую, что

(Ф1) $d_1|\,t\,|^{-\beta} \leqslant |\,\varphi_\varepsilon(t)\,| \leqslant d_2|\,t\,|^{-\beta}$ для любого $|\,t\,| > M$;

(Ф2) $d_1|\,t\,|^{\gamma_1}\exp(d_3|\,t\,|^{-\beta}) \leqslant |\,\varphi_\varepsilon(t)\,| \leqslant d_2|\,t\,|^{\gamma_2}\exp(d_3|\,t\,|^{-\beta})$ для любого $|\,t\,| > M$,

где $M, d_1, d_2, d_3, \gamma_1, \gamma_2$ и β – положительные константы.

Рассмотрим модель

$$Y_j = U_j + \sigma\varepsilon_j.$$

Теорема 2. *Пусть $f_\varepsilon(x)$ – гладкая функция порядка β, $K(x)$ – гладкая функция порядка α, выполнено условие $\int |\,t\,|^{2\beta} |\,\varphi_K(t)\,|\,dt < \infty$, то имеем:*

(i) если $\sigma = O(n^{-1/(2\alpha+1)})$, а $h \sim n^{-1/(2\alpha+1)}$, то

$$\sup_{g\in\mathcal{F}_{\alpha,C}} \mathbf{MISE}(\tilde{S}_{l,n}) = O(n^{-2\alpha/(2\alpha+1)}), \quad l = 1, 2;$$

(ii) если $\sigma \gg n^{-1/(2\alpha+1)}$, а $h \sim \sigma^{2\beta/(2\alpha+2\beta+1)} n^{-1/(2\alpha+2\beta+1)}$, то

$$\sup_{g\in\mathcal{F}_{\alpha,C}} \mathbf{MISE}(\tilde{S}_{l,n}) = O(\sigma^{4\alpha\beta/(2\alpha+2\beta+1)} n^{-2\alpha/(2\alpha+1)}), \quad l = 1, 2.$$

Случай $l = 1$ установлен в [26]. Нетрудно показать, что аналогичные соотношения имеют место и для $\tilde{S}_{2n}(x)$. Именно, как и выше нетрудно показать, что

$$\int_{-\infty}^{+\infty} [\mathbf{Bias}\{\tilde{S}_{2n}(x)\}]^2\,dx = \frac{h^{2\alpha}\mu_{K,\alpha}^2}{(\alpha!)^2} \int_{-\infty}^{+\infty} ((Fg)^{(\alpha)})^2\,dx + o(h^{2\alpha}), \tag{11}$$

$$\int_{-\infty}^{+\infty} \mathbf{Var}\{\tilde{S}_{2n}(x)\}\,dx = \frac{1}{2\pi nh} \int_{-\infty}^{+\infty} |\,\varphi_K(t)\,|^2 |\,\varphi_\varepsilon(\sigma t/h)\,|^{-2}\,dt + O(n^{-1}), \tag{12}$$

$$\mu_{K,\alpha} = \int_{-\infty}^{+\infty} x^\alpha K(x)\,dx.$$

Тогда имеет место следующая теорема.

Теорема 3. *Пусть $f_\varepsilon(x)$ – гладкая функция порядка β, $K(x)$ – гладкая функция порядка α, выполнено условие $\int |\,t\,|^{2\beta} |\,\varphi_K(t)\,|\,dt < \infty$. Тогда*

(i) если $\sigma = O(n^{-1/(2\alpha+1)})$, а $h \sim n^{-1/(2\alpha+1)}$, то

$$\sup_{g\in\mathcal{F}_{\alpha,C}} \mathbf{MISE}(\tilde{S}_{l,n}) = O(n^{-2\alpha/(2\alpha+1)});$$

(ii) если $\sigma \gg n^{-1/(2\alpha+1)}$, а $h \sim \sigma^{2\beta/(2\alpha+2\beta+1)} n^{-1/(2\alpha+2\beta+1)}$, то

$$\sup_{g\in\mathcal{F}_{\alpha,C}} \mathbf{MISE}(\tilde{S}_{l,n}) = O(\sigma^{4\alpha\beta/(2\alpha+2\beta+1)} n^{-2\alpha/(2\alpha+1)}).$$

Доказательство. Из (7) и (8) мы можем записать:

$$\mathbf{MISE}(S_{\delta n}) \sim c_1 h^{2\alpha} + I, \quad \text{где} \quad I = (2\pi nh)^{-1} \int_{-\infty}^{+\infty} |\,\varphi_K(t)\,|^2 |\,\varphi_\varepsilon(\sigma t/h)\,|^{-2}\,dt,$$

и c_1 – положительная константа. Из условия (Ф1) мы находим

$$I \leqslant \frac{c}{2\pi nh} \int_{|\,t\,|\leqslant Mh/\sigma} |\,\varphi_K(t)\,|^2 dt + \frac{d_1^{-2}\sigma^{2\beta}}{2\pi nh^{2\beta+1}} \int_{|\,t\,|>Mh/\sigma} |\,\varphi_K(t)\,|^2 |\,t\,|^{2\beta} dt = \mathcal{I}_2. \tag{13}$$

Если $\sigma = O(h)$, то $\mathcal{I}_2 \leqslant c_2/(nh)$, а $I \geqslant c_3/(nh)$. Отсюда следует результат теоремы. Аналогично доказывается часть (ii). □

Теперь уже нетрудно вывести, что оптимальная скорость сходимости, например, в случае (i) будет $n^{-2\alpha/(2\alpha+1)}$. Для $\alpha = 2$ это будет $1/(nh) = n^{-4/5}$.

Теорема 4. *Если $f_\varepsilon(x)$ – гладкая функция порядка β и $K(x)$ – гладкая функция порядка α, $\varphi_K(t)$ имеет носитель на $[-1,1]$, выполнено условие $\int [|\,t\,|^{-2\gamma_1} + |\,t\,|^{-2\gamma_2}|\,\varphi_K(t)\,|^2\,dt < \infty$, то имеем:*

(i) если $\sigma = O(n^{-1/(2\alpha+1)})$, а $h \sim n^{-1/(2\alpha+1)}$, то

$$\sup_{g\in\mathcal{F}_{\alpha,C}} \mathbf{MISE}(\tilde{S}_{l,n}) = O(n^{-2\alpha/(2\alpha+1)});$$

(ii) если $\sigma = n^{-1/(2\alpha+1)}a(n)$,
где $1 \ll a(n)a \ll n^{-1/(2\alpha+1)}$ $h = (2d_3/D)^{1/\beta}\sigma\{\ln a(n)\}^{-1\beta}$, то

$$\sup_{g\in\mathcal{F}_{\alpha,C}} \mathbf{MISE}(\tilde{S}_{l,n}) = O(\sigma^{2\alpha}\{\ln a(n)\}^{-2\alpha/\beta}).$$

Заключение

В работе исследована оценка функции распределения в зависимости «доза-эффект» в случае, когда ковариаты наблюдаются с ошибкой. Доказана асимптотическая нормальность построенной оценки. Установлена оптимальная скорость сходимости оценок в случае, когда ошибка имеет нормальное распределение, распределение Коши, устойчивое распределение (или близкие к ним). Однако сходимость этой оценки к «истинной» функции распределения оказывается медленной. Показано, что точность оценивания можно увеличить, если при каждом значении ковариат проводить повторные измерения.

Список литературы

[1] Криштопенко С.В., Тихов М.С., Попова Е.Б. Доза-Эффект. М.: Медицина, 2008. 228 с.

[2] Тихов М.С. Статистическое оценивание на основе интервальных цензурированных данных // Статистические методы оценивания и проверки гипотез: межвуз. сб. Пермь: Пермский университет, 2000. С. 49–70.

[3] Tikhov M.S. Statistical estimation based on interval censored data // Parametric and Semiparametric Models with Applications to Reliability, Survival Analysis, and Quality of Life. Eds. by N. Balakrishnan, M.S. Nikulin, M. Mesbah, N. Limnios. Series: Statistics for Industry and Technology. Boston, MA: Birkhauser, 2004. Pp. 211–218. https://doi.org/10.1007/978-0-8176-8206-4_14

[4] Tikhov M.S. Statistical estimation based on interval censored data // Journal of Mathematical Sciences. 2004. Vol. 119, № 3. Pp. 321–335.

[5] Tikhov M.S., Dolgih I.S. Asymptotically unbiased estimates of a distribution function in dose-effect relationships // Journal of Mathematical Sciences. 2015. Vol. 205, № 1. Pp. 113–120.

[6] Тихов М.С., Криштопенко С.В. Статистическое оценивание эффективной дозы зависимости доза-эффект с использованием как прямых, так и непрямых наблюдений // 2-я Всероссийская школа-коллоквиум по стохастическим методам. М.: ТВП, 1995. С. 81–82.

[7] Тихов М.С., Криштопенко Д.С. Асимптотические распределения интегрированных квадратичных ошибок непараметрических оценок по непрямым наблюдениям в зависимости доза-эффект // Статистические методы оценивания и проверки гипотез: межвуз. сб. Пермь: Пермский университет, 2007. С. 82–97.

[8] Tikhov M., Borodina T., Ivkin M. On reduction of measurement errors at estimation of distribution in dose-effect relationships // Advances in Mathematics and Statistical Sciences: Proc. of the 3rd International Conference on Mathematical, Computational and Statistical Sciences, MCSS'15. 2015. Pp. 19–27.

[9] Tikhov M., Ivkin M. Goodness of fit tests on the basis of the kernel quantile estimators in dose-effect relationship // International Journal of Mathematics and Computers in Simulation. 2015. Vol. 9. Pp. 127–135.

[10] Fan J. On optimal rates of convergence for nonparametric deconvolution problemes // Annals of Statistics. 1991. Vol. 19, № 3. Pp. 1257–1272.

[11] Zu Y. A note on the asymptotic normality of the kernel deconvolution density estimator with logarithmic chi-square noise // Econometrics. 2015. Vol. 3, № 3. Pp. 561–576. https://doi.org/10.3390/econometrics3030561

[12] Roussas G.G., Tran L.T. Asymptotic normality of the recursive kernel regression estimate under dependence conditions // Annals of Statistics. 1992. Vol. 20, № 1. Pp. 98–120.

[13] Грэхем Р., Кнут Д., Паташник О. Конкретная математика. М.: Мир, 1998. 703 с.

[14] Гнеденко Б.В. Курс теории вероятностей. М.: Эдиториал УРСС, 2005. 448 с.

[15] Петров В.В. Предельные теоремы для сумм независимых случайных величин. М.: Наука, 1987. 318 с.

[16] Колмогоров А.Н. Основные понятия теории вероятностей. М.: Наука, 1974. 120 с.

[17] Stein C. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables // Proc. of the Sixth Berkeley Symposium on Mathematical Statistics and Probability. Vol. 2. Berkeley: University of California Press, 1972. Pp. 583–602.

[18] Stein C. Approximate Computation of Expectations. Lecture Notes-Monograph Series. Institute of Mathematical Statistics, 1986. 164 p.

[19] Dey P. Stein's Method for Normal Approximation [Electronic resource]. URL: http://math.uiuc.edu/ math561sp16/lec25.pdf.

[20] Васерштейн Л.Н. Марковские процессы на счетном произведении пространств, описывающие большие системы автоматов // Проблемы передачи информации. 1969. Т. 5, № 3. С. 64–72.

[21] Канторович Л.В., Рубинштейн Г.Ш. Об одном пространстве вполне аддитивных функций // Вестник Ленинградского университета. Серия 1: Математика, механика, астрономия. 1958. Т. 7, № 2. С. 52–59.

[22] Канторович Л.В., Акилов Г.П. Функциональный анализ в нормированных пространствах. М.: ГИФМЛ, 1959. 684 с.

[23] Delaigle A., Hall P., Meister A. On deconvolution with repeated measurements // Annals of Statistics. 2008. Vol. 36, № 2. Pp. 665–685.

[24] Delaigle A., Hall P. Estimation of observation-error variance in errors-in-variables regression // Statistica Sinica. 2011. Vol. 21, № 3. Pp. 1023–1063.

[25] Comte F., Kappus J. Density deconvolution from repeated measurements without symmetry assumption on the errors // Journal of Multivariate Analysis. 2015. Vol. 140. Pp. 31–46.

[26] Delaigle A. An alternative view of the deconvolution problem // Statistica Sinica. 2008. Vol. 18, № 3. Pp. 1025–1045.

Образец цитирования

Тихов М.С. Фурье-метод рекурсивного оценивания функции распределения в зависимости доза-эффект // Вестник ТвГУ. Серия: Прикладная математика. 2018. № 4. С. 31–49. https://doi.org/10.26456/vtpmk516

Сведения об авторах

1. **Тихов Михаил Семенович**
профессор лаборатории прикладной теории вероятностей кафедры программной инженерии института информационных технологий, математики и механики Национального исследовательского Нижегородского государственного университета им. Н.И. Лобачевского.

Россия, 603950, г. Нижний Новгород, пр. Гагарина, д. 23, ННГУ им. Н.И. Лобачевского. E-mail: tikhovm@mail.ru

FOURIER METHODS FOR RECURSIVE ESTIMATING OF DISTRIBUTION FUNCTION IN DOSE-EFFECT RELATIONSHIP

Tikhov Mikhail Semenovich
Professor of Applied Probability Theory laboratory at Software Engineering department, Institute of Information Technology, Mathematics and Mechanics, Lobachevsky State University of Nizhniy Novgorod
Russia, 603950, Nyzhniy Novgorod, 23 Gagarin av., UNN.
E-mail: tikhovm@mail.ru

Received 06.08.2018, revised 19.09.2018.

It is proposed a kernel Fourier methods for recursive estimating of the distribution function in dose-effect relationship where the entered doses are observed with errors. The asymptotic normality of the offered estimates is proved. The possibility for increasing the accuracy of the estimators through repeated measurements of the entered doses is discussed.

Keywords: dose-effect relationship, deconvolving kernel distribution function estimators, asymptotic normality.

Citation

Tikhov M.S., "Fourier methods for recursive estimating of distribution function in dose-effect relationship", *Vestnik TvGU. Seriya: Prikladnaya Matematika* [Herald of Tver State University. Series: Applied Mathematics], 2018, no. 4, 31–49. (in Russian) https://doi.org/10.26456/vtpmk516

References

[1] Krishtopenko S.V., Tikhov M.S., Popova E.B., *Dosa-Effect [Dose-Effect]*, Medicina Publ., Moscow, 2008 (in Russian), 228 pp.

[2] Tikhov M.S., "Statistical Estimation on the Basis of Interval-Censored Data", *Statisticheskie Metody Ozenivaniya i Proverki Gipotez: Mezhvuz. Sb. [Interuniversity Transactions on Statistical Method of Estimation and Testing Hypotheses]*, Perm University, Perm, 2000, 49–70 (in Russian).

[3] Tikhov M.S., "Statistical estimation based on interval censored data", *Parametric and Semiparametric Models with Applications to Reliability, Survival Analysis, and Quality of Life*, Statistics for Industry and Technology, eds. N. Balakrishnan, M.S. Nikulin, M. Mesbah, N. Limnios, Birkhauser, Boston, MA, 2004, 211–218, https://doi.org/10.1007/978-0-8176-8206-4_14.

[4] Tikhov M.S., "Statistical estimation based on interval censored data", *Journal of Mathematical Sciences*, **119**:3 (2004), 321–335.

[5] Tikhov M.S., Dolgih I.S., "Asymptotically unbiased estimates of a distribution function in dose-effect relationships", *Journal of Mathematical Sciences*, **205**:1 (2015), 113–120.

[6] Tikhov M.S., Krishtopenko S.V., "Asymptotically unbiased estimates of a distribution function in dose-effect relationships", *2-ya Vserossijskaya shkola-kollokvium po stohasticheskim metodam [2nd All-Russian Colloquium on Stochastic Methods]*, TVP Publ., Moscow, 1995, 81–82 (in Russian).

[7] Tikhov M.S., Krishtopenko D.S., "Asymptotic distributions of integrated square errors of nonparametric estimation based on indirect observations under dose-effect dependence", *Statisticheskie Metody Ozenivaniya i Proverki Gipotez: Mezhvuz. Sb. [Interuniversity Transactions on Statistical Method of Estimation and Testing Hypotheses]*, Perm University, Perm, 2007, 82–97 (in Russian).

[8] Tikhov M., Borodina T., Ivkin M., "On reduction of measurement errors at estimation of distribution in dose-effect relationships", *Advances in Mathematics and Statistical Sciences: Proc. of the 3rd International Conference on Mathematical, Computational and Statistical Sciences*, MCSS'15, 2015, 19–27.

[9] Tikhov M., Ivkin M., "Goodness of fit tests on the basis of the kernel quantile estimators in dose-effect relationship", *International Journal of Mathematics and Computers in Simulation*, **9** (2015), 127–135.

[10] Fan J., "On optimal rates of convergence for nonparametric deconvolution problemes", *Annals of Statistics*, **19**:3 (1991), 1257–1272.

[11] Zu Y., "A note on the asymptotic normality of the kernel deconvolution density estimator with logarithmic chi-square noise", *Econometrics*, **3**:3 (2015), 561–576, https://doi.org/10.3390/econometrics3030561.

[12] Roussas G.G., Tran L.T., "Asymptotic normality of the recursive kernel regression estimate under dependence conditions", *Annals of Statistics*, **20**:1 (1992), 98–120.

[13] Graham R., Knuth D., Patashnik O., *Concrete Mathematics*, Addison-Wesley Publishing Company, New York, 1994, 657 pp.

[14] Gnedenko B.V., *The Theory of Probability*, AMS Chelsea Publishing, Providence, Rhode Island, 2005, 529 pp.

[15] Petrov V.V., *Predelnye teoremy dlya summ nezavisimykh sluchainykh velichin [Limit theorems for sums of independent random variables]*, Nauka Publ., Moscow, 1987 (in Russian), 318 pp.

[16] Kolmogorov A.N., *Osnovnye Ponyatiya Teorii Veroyatnostei [Foundations of The Theory of Probability]*, Nauka Publ., Moscow, 1974 (in Russian), 120 pp.

[17] Stein C., "A bound for the error in the normal approximation to the distribution of a sum of dependent random variables", *Proc. of the Sixth Berkeley Symposium on Mathematical Statistics and Probability.* V. 2, University of California Press, Berkeley, 1972, 583–602.

[18] Stein C., *Approximate Computation of Expectations*, Lecture Notes-Monograph Series, Institute of Mathematical Statistics, 1986, 164 pp.

[19] Dey P., *Stein's Method for Normal Approximation*, URL: http://math.uiuc.edu/math561sp16/lec25.pdf.

[20] Vaserstein L.N., "Markov processes on countable space products describing large systems of automata", *Problems of Information Transmission*, **5**:3 (1969), 47–52.

[21] Kantorovich L.V., Rubinshtein S.G., "On a space of totally additive functions", *Vestnik of the St. Petersburg University: Mathematics*, **13**:7 (1958), 52–59.

[22] Kantorovich L.V., Akilov G.P., *Functsionalniy Analiz v Normirovannykh Prostranstvakh [Functional Analysis in Normed Spaces]*, GIFML Publ., Moscow, 1959 (in Russian), 684 pp.

[23] Delaigle A., Hall P., Meister A., "On deconvolution with repeated measurements", *Annals of Statistics*, **36**:2 (2008), 665–685.

[24] Delaigle A., Hall P., "Estimation of observation-error variance in errors-in-variables regression", *Statistica Sinica*, **21**:3 (2011), 1023–1063.

[25] Comte F., Kappus J., "Density deconvolution from repeated measurements without symmetry assumption on the errors", *Journal of Multivariate Analysis*, **140** (2015), 31–46.

[26] Delaigle A., "An alternative view of the deconvolution problem", *Statistica Sinica*, **18**:3 (2008), 1025–1045.