© 2005 г. Н.М.Маркович, канд.тех.наук

(Институт проблем управления им. В.А.Трапезникова РАН, Москва)

ТОЧНОСТЬ ТРАНСФОРМИРОВАННЫХ ЯДЕРНЫХ ОЦЕНОК ПЛОТНОСТЕЙ РАСПРЕДЕЛЕНИЯ С ТЯЖЕЛЫМИ ХВОСТАМИ

Рассматривается непараметрическое оценивание плотности распределения вероятностей (ПР) с тяжелыми хвостами с использованием трансформирования исходных наблюдений. Исследуется точность трансформированных ядерных оценок с постоянной и переменной шириной окна в смысле интегральной среднеквадратичной ошибки для различных трансформаций. Для улучшения оценивания на хвостах ПР предлагаются пограничные ядра. Доказывается, что выбор в ядерной оценке с переменной шириной окна методом невязки обеспечивает близкую к оптимальной интегральную среднеквадратичную ошибку оценивания.

1. Введение

Пусть $X^n = \{X_1, \ldots, X_n\}$ — независимые одинаково распределенные неотрицательные случайные величины (сл.в.) с ПР $f(x)$ и функцией распределения (ФР) $F(x)$. Будем рассматривать оценивание тяжело-хвостовых ПР.

О п р е д е л е н и е. Распределение F (или сл.в. X) называется тяжело-хвостовым, если для его хвоста — функции $\overline{F}(x) = 1 - F(x) > 0$, $x \geq 0$, выполняется следующее свойство:

$$\lim_{x\to\infty} P\{X > x + y | X > x\} = \lim_{x\to\infty} \overline{F}(x+y)/\overline{F}(x) = 1,$$

для любых $y \geq 0$ [1].

Среди известных непараметрических оценок ПР, таких как гистограмма, проекционные, ядерные оценки, только ядерные оценки могут быть адекватно использованы на неограниченном интервале. Но и они плохо оценивают ПР, если распределение имеет тяжелые хвосты. Ввиду того, что параметр сглаживания (ширина окна) ядерных оценок фиксирован, они могут давать на пики в хвостовой области распределения, соответствующей редким событиям, или переглаживать основную часть ПР. Пример такого поведения приведен в [2, с. 18] для данных по самоубийствам.

Для улучшения оценивания тяжело-хвостовых ПР будем использовать трансформации данных. Для этого наблюдаемые данные преобразуются с помощью некоторой монотонно возрастающей непрерывно дифференцируемой взаимно-однозначной функции трансформации $T(x)$ к новым сл.в. $Y_1, \ldots, Y_n$ $(Y_i = T(X_i))$. Далее оценивается ПР $g_0(x)$ сл.в. Y_i, и после обратного преобразования получается ПР сл.в. X_i по формуле

$$\widehat{f}(x) = \widehat{g_0}(T(x))T^{'}(x).$$

В основе идеи трансформации лежит необходимость различного сглаживания в разных областях тяжело-хвостовой ПР. Тогда трансформированные оценки с фиксированными параметрами сглаживания работают как локально-адаптивные оценки.

Альтернативный подход заключается в использовании ядерной оценки с переменной шириной окна [2-4]

$$\widehat{f}^A(x|h) = \frac{1}{nh}\sum_{i=1}^{n} f(X_i)^{1/2} K\left((x - X_i)f(X_i)^{1/2}/h\right), \tag{1}$$

ее модификаций [5-6] и обобщений [7]. Так как $f(X_i)$ неизвестна, то используется оценка

$$\widetilde{f}^A(x|h_1, h) = \frac{1}{nh}\sum_{i=1}^{n} \widehat{f}_{h_1}(X_i)^{1/2} K\left((x - X_i)\widehat{f}_{h_1}(X_i)^{1/2}/h\right),$$

где $\widehat{f}_{h_1}(X)$ - некоторая предварительная, обычно ядерная оценка с постоянным ядром

$$\widehat{f}_h(x) = (nh)^{-1}\sum_{i=1}^{n} K((x - X_i)/h), \tag{2}$$

$K(x)$ - ограниченная борелевская функция (ядро). Однако оценки $\widehat{f}^A$ и $\widetilde{f}^A$ не могут быть применены для оценивания хвостовой области ПР, по меньшей мере, для компактных ядер, так как они определены на интервале, грубо говоря, чуть большем размаха выборки.

В схеме с трансформациями ПР $g_0(x)$ может быть оценена различными непараметрическими методами: ядерными, проекционными оценками или гистограммой. Выбор $T(x)$ также принципиален. Так как ФР сл.в. Y_i задается формулой

$$G(x) = P\{Y_i < x\} = P\{T(X_i) < x\} = P\{X_i < T^{-1}(x)\} = F(T^{-1}(x)),$$

то трансформация $T(x)$ полностью определяется ФР $G(x)$ и $F(x)$. Если "целевая" ФР $G(x)$ может быть задана, то ФР $F(x)$ неизвестна. В [8] доказывается, что трансформации к ограниченному интервалу, т.е. $T : R_+ \to [0,1]$, где в качестве $G(x)$ берется треугольная на интервале [0,1] ФР, обеспечивают оптимальное оценивание ядерными оценками (2) в смысле наименьшей средней ошибки в пространстве L_1 $E\int|\widehat{f}_h(x) - f(x)|dx$. При этом $F(x)$ предлагается оценивать каким-то параметрическим методом [1]. Однако, как конкретно оценивать $F(x)$, и каков при этом будет порядок убывания на бесконечности трансформированных оценок ПР, не рассматривается. В [9-10] предлагается выбирать фиксированную трансформацию (т.е. не зависящую от $F(x)$) $T_\lambda(x)$ из некоторого параметрического семейства, например, из

$$(3) \qquad T_\lambda(x) = \begin{cases} x^\lambda sign(\lambda), & \text{если } \lambda \neq 0, \\ \ln x, & \text{если } \lambda = 0, \end{cases}$$

где λ — параметр, минимизирующий функционал $\int_R (g''(y))^2 dy$, $g(x)$ — неизвестная ПР трансформированной сл.в. $Y_1 = T_\lambda(X_1)$, требующая предварительной оценки. Так как $\int_R (g''(y))^2 dy$ характеризует "кривизну" ПР, то такие трансформации направлены на улучшение оценивания многомодальных, необязательно тяжело-хвостовых ПР. В [11] рассматривается фиксированная трансформация $T(x) = (2/\pi)\operatorname{arctg} x$, обеспечивающая хорошую точность на ряде тяжело-хвостовых распределений. Однако без предположений о типе хвоста любая трансформация может привести к непредсказуемой ПР и не обеспечить хорошего оценивания в хвостовой области. Преодолеть это позволяют адаптивные трансформации. В [12-13] предложена трансформация $T_{\widehat{\gamma}}(x) : R_+ \to [0,1]$, адаптирующаяся к данным через использование оценки $\widehat{\gamma}$ параметра γ, определяющего форму хвоста так, чтобы обеспечить

[1] Эмпирическая ФР не может быть использована в качестве оценки $F(x)$ ввиду ее недифференцируемости в конечном числе точек.

наилучшую точность оценивания хвостовой области ПР. Для построения такой оценки $T_{\widehat{\gamma}}(x)$ для $T(x)$ в качестве "целевой" ФР $G(x)$ предлагается ФР треугольного распределения $\Phi^{+tri}(x) = (2x - x^2)\mathbf{1}\{x \in [0,1]\} + \mathbf{1}\{x > 1\}$, а в качестве "предполагаемой" ФР (вместо $F(x)$) — ФР Парето

$$\Psi_{\widehat{\gamma}}(x) = \begin{cases} 1 - (1 + \widehat{\gamma}x)^{-1/\widehat{\gamma}}, & \text{если} \quad x \geq 0, \\ 0, & \text{если} \quad x < 0. \end{cases}$$

Тогда трансформация от $\Psi_{\widehat{\gamma}}(x)$ к $\Phi^{+tri}(x)$ определяется по формуле

$$T_{\widehat{\gamma}}(x) = (\Phi^{+tri})^{-1}(\Psi_{\widehat{\gamma}}(x)) = 1 - \sqrt{1 - \Psi_{\widehat{\gamma}}(x)} = 1 - (1 + \widehat{\gamma}x)^{-1/(2\widehat{\gamma})}, \tag{4}$$

$$T'_{\widehat{\gamma}}(x) = 0,5(1 + \widehat{\gamma}x)^{-1/(2\widehat{\gamma})-1},$$

$$T^{-1}_{\widehat{\gamma}}(x) = ((1-x)^{-2\widehat{\gamma}} - 1)/\widehat{\gamma}, \qquad (T^{-1}_{\widehat{\gamma}}(x))' = 2(1-x)^{-2\widehat{\gamma}-1}.$$

Цель работы — исследование точности ядерных оценок в рамках схемы с трансформациями и сравнение этих оценок для различных трансформаций. Хотя в [8] была доказана оптимальность в пространстве L_1 трансформированных ядерных оценок с трансформациями к треугольно распределенным сл.в., вопрос о точности в пространстве L_2 остается открытым. В качестве точности оценивания ПР $f(x)$ будем рассматривать интегральную среднеквадратичную ошибку (the mean integrated squared error ($MISE$))

$$MISE^h(\widehat{\gamma}, \Omega) = E\int_{\Omega}(\widehat{f}(x) - f(x))^2 dx \tag{5}$$

$$\begin{aligned} &= E\int_{\Omega}(\widehat{g}_h(T_{\widehat{\gamma}}(x)) - g(T_{\widehat{\gamma}}(x)))^2 T'_{\widehat{\gamma}}(x) dT_{\widehat{\gamma}}(x) \\ &= E\int_{\Omega^*}(\widehat{g}_h(y) - g(y))^2 T'_{\widehat{\gamma}}(T^{-1}_{\widehat{\gamma}}(y)) dy, \end{aligned}$$

где $\Omega^* = T_{\widehat{\gamma}}(\Omega)$, $g(x)$ — ПР, которая в действительности оценивается вместо $g_0(x) = f(T^{-1}_{\gamma}(x))(T^{-1}_{\gamma}(x))'$, так как $\widehat{\gamma} \neq \gamma$:

$$g(x) = f(T^{-1}_{\widehat{\gamma}}(x))(T^{-1}_{\widehat{\gamma}}(x))', \tag{6}$$

а $\widehat{g}_h(x)$ — ядерная оценка $g(x)$ со сглаживающим параметром h.

Главное преимущество ядерных оценок типа $\widehat{f}^A(x|h)$, $\widetilde{f}^A(x|h_1,h)$ состоит в том, что они обеспечивают порядок $MISE = E\int_{-\infty}^{\infty}\left(\widehat{f}^A(x|h)-f(x)\right)^2dx \sim n^{-8/9}$ с сохранением положительности, если производная $f^{(4)}$ непрерывна, в случае использования ядер таких, что $\int t^4K(t)dt < \infty$, $h \sim n^{-1/9}$ [2, 14]. Для обычных ядерных оценок (2) $MISE \sim n^{-4/5}$ в случае использования ядер второго порядка и $MISE \sim n^{-8/9}$ для ядер четвертого порядка, которые неположительны. Как замечено в [5], величина $MISE$ для оценок $\widehat{f}^A(x|h)$ зависит от тяжести хвоста ПР, т.е. $MISE$ дрейфует от $\sim n^{-8/9}$ к $\sim n^{-4/5}$ для более легких хвостов.

Оптимальный параметр $h = c \cdot n^{-1/9}$, при котором достигается $MISE \sim n^{-8/9}$ зависит от производных неизвестной ПР вплоть до четвертого порядка. Оценивание же производных от ПР само по себе является трудной задачей. Так, при оценивании второй производной ПР оптимальный порядок $AMISE \sim n^{-4/9}$ (асимптотической $MISE$) такой же, как и у оптимальной $AMISE$ полигона частот при оценивании пятимерной ПР [15]. Отсюда для практических расчетов лучше применять адаптивные методы выбора h по выборкам ограниченных объемов: метод скользящего контроля [16] или метод невязки [17]. Сходимость ядерных оценок в случае использования метода скользящего контроля доказана лишь для ПР с компактным носителем [16]. Для тяжело-хвостовых ПР такие оценки расходятся, поэтому приходится применять трансформацию наблюдений к ограниченному интервалу.

В статье решаются следующие задачи. В разделе 2 показано, что для таких трансформаций, при которых функция $T'(T^{-1}(y))$ ограничена, $MISE$ оценки $\widehat{f}(x)$ на неограниченном интервале Ω определяется среднеквадратичной ошибкой (MSE) оценки $\widehat{g}_h(x)$ на интервале Ω^*, т.е. $E\int_{\Omega^*}(\widehat{g}_h(y)-g(y))^2dy$. Таким образом, может быть достигнут оптимальный порядок $MISE$ на $[0,\infty)$ как для фиксированных трансформаций, так и для адаптивных в случае достаточно гладких ПР $g(x)$. Сравниваются $MISE$ трансформированных ядерных оценок для трансформации $T_{\widehat{\gamma}}(x)$ и двух фиксированных трансформаций $\ln x$ и $(2/\pi)\operatorname{arctg} x$. Интерес представляет трансформированная ядерная оценка (2) с $MISE \sim n^{-4/5}$ и хорошей точностью в хвостовой области ПР. Для лучшего оценивания на хвостах ПР в разделе 3 предлагаются пограничные ядра, которые используются для оценивания

ПР трансформированной сл.в. $g(x)$ в точках, близких к граничным. Рассмотрение можно легко распространить на ядерную оценку с переменной шириной окна. В разделе 4 рассматривается метод невязки для нахождения параметра сглаживания h_2 в оценке $\widetilde{f}^A(x|h_1,h)$ по выборке. Доказано, что этот метод обеспечивает дисперсию $\widetilde{f}^A(x|h_1,h)$ порядка $n^{-8/9}$ и $MSE \sim n^{-8/9}(\ln n)^2$, если четвертые производные от ПР $g(x)$ и от функции $1/g(x)$ непрерывны на ограниченном интервале. Схема с трансформациями дает возможность получить ПР с указанными свойствами и тем самым оценить ПР оптимальным образом. $T_{\widehat{\gamma}}(x)$ дает пример такой трансформации в классе распределений с правильно меняющимися хвостами (14). В Приложении приведено доказательство теоремы.

2. $MISE$ на R_+ трансформированных ядерных оценок

Для фиксированных трансформаций, как например (3), и неслучайных интервалов Ω^* $MISE$ имеет более простой вид по сравнению с (5),

$$MISE^h(\Omega) = \int_{\Omega^*} T'(T^{-1}(y))E(\widehat{g}_h(y) - g(y))^2 dy. \tag{7}$$

Если для некоторой трансформации $T(x)$ (необязательно фиксированной) функция $T'(T^{-1}(x))$ ограничена на интервале Ω^*, т.е.

$$0 < T'(T^{-1}(x)) \le c, \tag{8}$$

где c — некоторая константа, то

$$MISE^h(\Omega) \le c \int_{\Omega^*} E(\widehat{g}_h(y) - g(y))^2 dy. \tag{9}$$

Таким образом, интегральная среднеквадратичная ошибка на Ω для трансформированных оценок, по-меньшей мере, не хуже порядка среднеквадратичной ошибки оценки ПР трансформированной сл.в.

2.1. MSE ядерной оценки с постоянной шириной окна

Рассмотрим сначала ядерную оценку (2) с ядром 2-го порядка, т.е. $K(x)$ — функция, удовлетворяющая условиям:

$$\int K(x)dx = 1, \qquad \int x^i K(x)dx = 0, i = 1,...,p-1, \qquad \int x^p K(x)dx = K_{p-1} \neq 0 \tag{10}$$

при $p = 2$ [2]. Предположим, что неизвестная ПР $g(x)$ имеет непрерывную вторую производную и $h \to 0$, $nh \to \infty$ при $n \to \infty$. Согласно стандартным выкладкам, приведенным в [2] и [15], смещение оценки $\widehat{g}_h(x)$ равно в силу (10)

$$b(\widehat{g}_h(x)) = E\widehat{g}_h(x) - g(x) = \frac{1}{h}\int K\left(\frac{x-y}{h}\right) g(y)dy - g(x)\int K(u)du \tag{11}$$

$$= -hg'(x)\int uK(u)du + h^2 g''(x)/2 \int u^2 K(u)du + O(h^2)$$

$$= h^2 g''(x) K_1/2 + O(h^2).$$

Обозначим $K_2 = \int K^2(t)dt$. Дисперсия оценки $\widehat{g}_h(x)$ равна

$$v(\widehat{g}_h(x)) = E\widehat{g}_h(x)^2 - (E\widehat{g}_h(x))^2$$

$$= (nh^2)^{-1}\int K^2\left((x-y)/h\right) g(y)dy - n^{-1}\{g(x) + bias\widehat{g}_h(x)\}^2$$

$$= (nh)^{-1}\int K^2(u)\{g(x) - hug'(x) + (hu)^2 g''(x)/2 + \ldots\}du$$

$$-\frac{1}{n}\{g(x) + h^2 g''(x)K_1/2 + O(h^2)\}^2$$

$$= (nh)^{-1} g(x) K_2 + O(n^{-1}).$$

Отсюда и из (11) выполнено

$$MSE(\widehat{g}_h) = E(\widehat{g}_h(x) - g(x))^2 = h^4 (g''(x))^2 K_1^2/4 + (nh)^{-1} g(x) K_2 + o\{(nh)^{-1} + h^4\}. \tag{12}$$

Если рассматривается трансформация к $[0,1]$, удовлетворяющая (8), $g(x)$ — финитная ПР, сосредоточенная на $[0,1]$, и $\int_0^1 g(x)dx = 1$, то из (9) следует, что

$$MISE^h(\Omega) \le c\left(\frac{h^4}{4}K_1^2 J_1 + \frac{K_2}{nh}\right),$$

где

$$J_1 = \int_0^1 (g''(x))^2 dx.$$

Минимум правой части достигается для

$$h^{opt} = \left(\frac{K_2}{K_1^2 J_1}\right)^{1/5} n^{-1/5}.$$

[2] Наиболее простой пример таких ядер — симметричные ядра, сосредоточенные на ограниченном носителе.

Для таких h^{opt} получим

$$MISE^{h^{opt}}(\Omega) \le c(5/4)(K_1K_2^2)^{2/5}n^{-4/5}J_1^{1/5}.$$

Оптимальным в смысле наименьшей величины $K_1K_2^2$ ядром 2-го порядка для оценки (2) является ядро Епанечникова $K(x) = (3/4)(1-x^2)1\{|x| \le 1\}$ [2, 15].

Однако идеальная оценка недостижима, поскольку h^{opt} зависит от неизвестной производной $g''(y)$. На практике можно заменить $g(y)$ предварительной ядерной оценкой

$$\widehat{g}_{h_1}(y, \widehat{\gamma}) = \frac{1}{nh_1}\sum_{i=1}^{n} K(\frac{y-Y_i}{h_1}),$$

вычисляемой по трансформированной выборке $Y_i = T_{\widehat{\gamma}}(X_i), i = \overline{1,n}$. Тогда производная $g''(y)$ оценивается второй производной от $\widehat{g}_{h_1}(x, \widehat{\gamma})$. Выберем ядро $K(x)$, имеющее две ограниченные производные и удовлетворяющее (10), например ядро Епанечникова. Отсюда, так как $g''(x)$ предполагается непрерывной, то $\widehat{g}''_{h_1}(x, \widehat{\gamma})$ оценивает $g''(x)$ [18]. Параметр сглаживания h_1 для $\widehat{g}_{h_1}(x, \widehat{\gamma})$ может быть выбран равным $\sigma_Y\left(84\sqrt{\pi}/(5n^2)\right)^{1/13}$, что соответствует параметру ядерной оценки с гауссовым ядром, обеспечивающим оптимальное оценивание $\int \left(g''(x)\right)^2 dx$ [9], либо методом адаптивного выбора, как кросс-валидация [16] или метод невязки [17].

2.2. MSE ядерной оценки с переменной шириной окна

Рассмотрим среднеквадратичную ошибку оценки (1). Если $g^{(4)}(x)$ существует и симметричное ядро $K(x)$ удовлетворяет условиям $\int t^4K(t)dt < \infty$, $\int K(t)dt = 1$, $\sup|K| < \infty$, то согласно [4]

$$MSE(\widehat{g}^A(x)) = E(\widehat{g}^A(x|h) - g(x))^2 = h^8\left(\frac{K_4}{24}\right)^2\left(\left(\frac{d}{dx}\right)^4\frac{1}{g(x)}\right)^2 + g(x)^{3/2}\frac{K_2}{nh} + o\{\frac{1}{nh} + h^8\} \tag{13}$$

при $h \to 0$. Если рассматривается трансформация к $[0,1]$ и $\int_0^1\left(\left(\frac{d}{dx}\right)^4 g(x)^{-1}\right)^2 dx \ne 0$, то минимум $\int_0^1 E(\widehat{g}^A(x|h) - g(x))^2dx$ достигается на

$$h^{opt} = \left(\frac{F_2}{8F_1}\right)^{1/9} n^{-1/9},$$

где $F_1 = (K_4/24)^2 \int_0^1 \left(\left(\frac{d}{dx}\right)^4 g(x)^{-1} \right)^2 dx$, $F_2 = K_2 \int_0^1 g(x)^{3/2} dx$. При таком h^{opt} имеем, если выполнено (8), что

$$MISE^{h^{opt}}(\widehat{\gamma}, \Omega) \leq 9 \cdot 8^{-8/9} c F_1^{1/9} F_2^{8/9} n^{-8/9} + o(n^{-8/9})$$

при $n \to \infty$. Для $\widehat{g}^A(x|h)$ в [7] рекомендовано ядро

$$K(x) = (1 - x^4)^4 \mathbf{1}\{|x| \leq 1\}/2a_4,$$

где $a_k = \int_0^1 (1 - x^4)^k dx = 4^k k!/(4k + 1)!!!!$, $m!!!! = m(m - 4)...1$.

2.3. Адаптивная трансформация

Рассмотрим трансформацию (4). Пусть истинная ПР $f(x)$ принадлежит классу ПР

$$f(x) = \begin{cases} \ell(x)(1 + \gamma x)^{-(1/\gamma+1)}, & 0 < c \leq x, \\ 0, & x < c, \end{cases} \tag{14}$$

где $\ell(x)$ — медленно меняющаяся функция. Это означает, что $\lim_{x\to\infty} \ell(tx)/\ell(x) = 1$, $\forall t > 0$. Примерами таких функций служат положительные константы, функции, сходящиеся к положительным константам, логарифмы. К таким распределениям относятся распределения с правильно меняющимися хвостами: Парето, Коши, Бурра (Burr), Фреше (Fréchet) и другие.

Тогда оцениваемая ПР трансформированной сл.в. $Y_1 = T_{\widehat{\gamma}}(X_1)$ выражается из (6) как

$$g(x) = 2\ell\left(\frac{(1 - x)^{-2\widehat{\gamma}} - 1}{\widehat{\gamma}}\right) \left(1 + (\gamma/\widehat{\gamma})((1 - x)^{-2\widehat{\gamma}} - 1)\right)^{-(1+1/\gamma)} (1 - x)^{-2\widehat{\gamma}-1} \tag{15}$$

$$\simeq 2\ell\left(\frac{(1 - x)^{-2\widehat{\gamma}} - 1}{\widehat{\gamma}}\right) \left(\frac{\widehat{\gamma}}{\gamma}\right)^{1+1/\gamma} (1 - x)^{2\widehat{\gamma}/\gamma-1}.$$

Все производные функции $g(x)$ существуют на $\Omega^* = [0, 1 - \varepsilon]$ для любых $0 < \varepsilon < 1$. Так как функция $T'_{\widehat{\gamma}}(T_{\widehat{\gamma}}^{-1}(x)) = 0,5(1-x)^{1+2\widehat{\gamma}}$ ограничена на $[0, 1]$, то из (9) и (12) (или (13)) следует $MISE^h(\widehat{\gamma}, \Omega) \sim n^{-4/5}$ и $\sim n^{-8/9}$ для оценок (2) и (1) соответственно при соответствующем выборе ядер, ширины окна и $\Omega = T_{\widehat{\gamma}}^{-1}(\Omega^*)$.

2.4. $MISE(R_+)$ для трансформаций $T(x) = \ln x$ и $T(x) = (2/\pi) \operatorname{arctg} x$

Предположим, что для оценивания неизвестной ПР $f(x)$ используется оценка (2). Рассмотрим часто применяемую на практике трансформацию $T(x) = \ln x : [0, \infty) \to$

$(-\infty, \infty)$. Так как $T'(T^{-1}(x)) = e^{-x}$, то из (7) и (12) при $h \to 0$, $\Omega = R_+$ и если ПР трансформированной сл.в. $Y_1 = \ln X_1$ имеет непрерывную вторую производную, получаем

$$(16) \qquad MISE^h(\Omega) = \int_{-\infty}^{\infty} e^{-x} \left(h^4 (g''(x))^2 K_1^2/4 + (nh)^{-1} g(x) K_2 + o\{(nh)^{-1} + h^4\}\right) dx.$$

Если $f(x)$ — логнормальная ПР, то $g(x)$ — нормальная ПР, $g''(x)$ непрерывна и все интегралы в (16) сходятся. Из (16) следует, что $MISE^h(\Omega) \sim n^{-4/5}$, если $h \sim n^{-1/5}$. Однако если $f(x)$ принадлежит классу ПР с правильно меняющимися хвостами (14), то $MISE^h(\Omega)$ может быть не определена. Действительно, согласно (6) в этом случае

$$(17) \qquad g(x) = \ell(e^x)\left(1 + \gamma e^x\right)^{-(1/\gamma+1)} e^x.$$

Предположим, что $\ell(x)$ такая, что $(\ell(e^x))''$ непрерывна на Ω^* (условие (A)).
Тогда

$$g''(x) = (1 + \gamma e^x)^{-(1/\gamma+1)} e^x [(\ell(e^x))'' + \ell(e^x) + 2(\ell(e^x))'$$

$$+\ell(e^x) e^x (1+\gamma)(1+\gamma e^x)^{-1} \left(-3 + (1+2\gamma)(1+\gamma e^x)^{-1} e^x - \gamma\right)$$

$$-2(\ell(e^x))' e^x (1+\gamma)(1+\gamma e^x)^{-1}]$$

непрерывна и верно (12).

В отличие от $\ell^*(x) = \exp\left((\ln(1+x))^{1/2} \cos((\ln(1+x))^{1/2})\right)$ [3] функции $\ell(x)$, такие как $\lim_{x\to\infty} \ell(x) = \ell_\infty, 0 < \ell_\infty < \infty$, (это, например, положительные константы или $\ell(x) = (1 + 1/x)^{-1}$) и $\ell(x) = \ln x$, удовлетворяют условию (A) на $\Omega^* = (-\infty, \infty)$, $\Omega^* = [a, \infty)$, где $a > 0$ - некоторая константа.

Рассмотрим $\Omega^* = (-\infty, \infty)$ и $\ell(x)$, удовлетворяющие (A). Тогда для (17) интеграл

$$\int_{-\infty}^{\infty} g(x) e^{-x} dx = \int_{-\infty}^{\infty} \ell(e^x)\left(1 + \gamma e^x\right)^{-(1/\gamma+1)} dx = \int_0^{\infty} \frac{\ell(u/\gamma) du}{u(1+u)^{1/\gamma+1}}$$

в (16) не сходится, и поэтому $MISE^h(\Omega)$ не определена.

Пусть в (7) $\Omega^* = [a, \infty)$, $a > 0$. Для монотонных и ограниченных функций $\ell(x)$ ($|\ell(x)| \le L$, $0 < a \le x < \infty$) или таких $\ell(x)$, что $\ell(x) \sim x^{1/\gamma+2-\omega}$, $(x \to \infty)$, ω — некоторая константа такая, что $\omega > 3/2$, интегралы $\int_{\Omega^*} g(x) e^{-x} dx$ и $\int_{\Omega^*} g''(x)^2 e^{-x} dx$

[3] Эта функция $\ell(x)$ дает пример медленно меняющейся функции, осциллирующей на бесконечности, т.е. $\lim_{x\to\infty} \inf \ell(x) = 0$ и $\lim_{x\to\infty} \sup \ell(x) = \infty$ [19].

сходятся. Примерами таких $\ell(x)$ являются $\ell(x) = \ln x$ и $\ell(x)$ такие, что $\lim_{x\to\infty} \ell(x) = \ell_\infty, 0 < \ell_\infty < \infty$. Так как такие $\ell(x)$ удовлетворяют условию (A), то $g''(x)$ непрерывна и для $h \sim n^{-1/5}$ получим

$$MISE_x^h(\Omega) = \frac{h^4 K_1^2}{4} \int_a^\infty g''(x)^2 e^{-x} dx + \frac{c}{nh} \int_a^\infty g(x) e^{-x} dx + o_a\{\frac{1}{nh} + h^4\} \sim n^{-4/5}.$$

Теперь рассмотрим трансформацию $T(x) = (2/\pi) \operatorname{arctg} x : [0, \infty) \to [0, 1]$. Для нее

$$T'(T^{-1}(x)) = 2/(\pi(1 + \operatorname{tg}^2(\pi x/2)))$$

ограничена на $[0, 1]$. Если $f(x)$ из класса ПР (14), то из (6) ПР трансформированной сл.в. имеет вид

$$g(x) = \ell\left(\operatorname{tg}\left(\frac{\pi x}{2}\right)\right) \frac{\pi}{2} \frac{1 + \operatorname{tg}^2((\pi/2)x)}{(1 + \gamma \operatorname{tg}((\pi/2)x))^{1/\gamma+1}}. \tag{18}$$

Производная $\left(\ell\left(\operatorname{tg}\left(\frac{\pi x}{2}\right)\right)\right)''$ не непрерывна для $\ell^*(x)$ в отличие от $\lim_{x\to\infty} \ell(x) = \ell_\infty, 0 < \ell_\infty < \infty$, и $\ell(x) = \ln x$. Для таких $\ell(x)$ и для $\gamma \leq 1/3$ $g''(x)$ непрерывна на Ω^*. Поэтому для нее применимы результаты п. 2.1. Согласно (9) $MISE_x^h(\Omega) \sim n^{-4/5}$.

3. Пограничные ядра

Если форма хвоста известна, независимо от трансформации, порядки убывания трансформированных оценок на хвостах могут быть близкими к таковым для истинной ПР $f(x)$ при правильном выборе ядер и величины сглаживающего параметра h вблизи границ изменения трансформированной сл.в. $Y_1 = T(X_1)$ [4]. Далее будем предполагать, что ПР сл.в. X_1 принадлежит классу (14).

Для трансформации (4) после обратного преобразования получим оценку ПР сл.в. X_1 в виде

$$\widehat{f}(x) = \widehat{g}_h(T_{\widehat{\gamma}}(x)) T'_{\widehat{\gamma}}(x) = 0,5 \widehat{g}_h(T_{\widehat{\gamma}}(x))(1 + \widehat{\gamma} x)^{-1/(2\widehat{\gamma})-1}. \tag{19}$$

Нам необходимо получить $\widehat{f}(x) \sim (1 + \widehat{\gamma} x)^{-1/\widehat{\gamma}-1}$, что близко к (14) для достаточно точной оценки $\widehat{\gamma}$. Точность хвоста оценки $\widehat{f}(x)$ зависит от поведения оценки $\widehat{g}_h(x)$ вблизи 1, так как $T_{\widehat{\gamma}}(x) \to 1$ при $x \to \infty$.

Как известно, ядерные оценки (2), применяемые к ПР, сосредоточенным на ограниченных носителях, страдают краевыми эффектами ввиду усечения ядер на

[4] Класс хвоста можно установить с помощью непараметрического теста [20].

границе носителя [21-23]. Для того чтобы преодолеть этот эффект, применяются пограничные ядра. Такие ядра могут быть полезными для улучшения оценивания ПР, принимающей нулевое значение или терпящей разрыв (ограниченный скачок) на границе. Например, в [21] предложены модифицированные ядра типа $K_b(x) = (15/16)(1-x^2)_+^2$. В [22] рассмотрена техника зеркального отражения выборки на соответствующие отрицательные значения. Для успешного применения этих методов требуется точная информация о расположении разрыва или нулевого значения ПР. Вообще говоря, такая информация может быть получена с помощью гистограммы. Пограничное ядро строится при определенных ограничениях на ядра. Например, что ядро и его производная равны нулю на границе, когда ПР принимает там нулевое значение. Так как такое пограничное ядро не зависит от ПР, то гарантировать хорошее восстановление нельзя.

Следуя [23], смещение оценки на границе может быть уменьшено путем использования линейной комбинации B_1 или B_2 двух ядер $K(x)$ и $L(x)$ ($K(x)$ — обычное ядро, используемое для восстановления ПР во внутренних точках интервала и определенное на $[0,1]$, а $L(x)$ — некоторое ядро, отличное от $K(x)$, но каким-то образом относящееся к нему):

$$B_1(x) = \frac{c_1(p)K(x) - a_1(p)L(x)}{c_1(p)a_0(p) - a_1(p)c_0(p)},$$

где

$$a_l(p) = \int_{1-p}^{1} u^l K(u)du, \qquad c_l(p) = \int_{1-p}^{1} u^l L(u)du, \qquad 0 < p < 1,$$

или

$$B_2(x) = \frac{(a_2(p) - a_1(p)x)\, K(x)}{a_0(p)a_2(p) - a_1^2(p)},$$

если $L(x) = xK(x)$. Это возможно только тогда, когда вторая производная оцениваемой ПР непрерывна.

Смещение ядерной оценки с ядром B_1 или B_2 в пограничной области имеет порядок $O(h^2)$, а дисперсия $O((nh)^{-1})$ (такие же, как и во внутренних точках интервала).

Следует найти такие пограничные ядра, которые позволят получить для трансформированных оценок такое же убывание на бесконечности, как у истинной ПР. Разумеется, этого невозможно достигнуть, не располагая информацией о форме хвоста распределения. Поскольку упомянутые выше подходы не используют такой информации, то они не всегда полезны для этой цели. В частности, порядок убывания хвоста оценки (2) с ядром $B_2(x)$ зависит от ядра $K(x)$ для класса (14) и трансформации (4) (см. ниже).

Из (19) ясно, что для получения порядка убывания на бесконечности, близкого к таковому у (14), необходимо выбрать ядро такое, чтобы обеспечить

$$\widehat{g}_h(T_{\widehat{\gamma}}(x)) \sim (1+\widehat{\gamma}x)^{-1/(2\widehat{\gamma})} \qquad \text{при} \qquad x \to \infty. \tag{20}$$

Это можно сделать двумя путями:

1. Выбрав ядро в граничных точках $(Y_{(n)}, 1]$ в ядерной оценке $\widehat{g}_h(x)$ совпадающим с "целевой" ПР $g(x)$;

2. Выбрав h для произвольного ядра из уравнения:

$$\frac{1}{h}K\left(\frac{T(x)-Y_{(n)}}{h}\right)T'(x) = \widehat{f}(x),$$

   где $Y_{(n)}$ — максимальная порядковая статистика, соответствующая трансформированной выборке.

Для (2) получим в граничных точках $y \in (Y_{(n)}, 1]$, что

$$\widehat{g}_h(y) \simeq \frac{1}{h}K\left(\frac{y-Y_{(n)}}{h}\right),$$

т.е. $\widehat{g}_h(y)$ аппроксимируется в пограничных точках последним слагаемым в сумме (2). Получим из (4) для треугольного ядра $K(x) = 2(1-x)\mathbf{1}\{x \in [0,1]$, совпадающего с "целевой" ПР, и для ядра Епанечникова, что

$$K\left(\frac{T_{\widehat{\gamma}}(x)-Y_{(n)}}{h}\right) \sim (1+\widehat{\gamma}x)^{-1/(2\widehat{\gamma})} \qquad \text{при} \qquad x \to \infty,$$

когда взято $h = 1 - Y_{(n)}$. Использование $B_2(x)$ в $\widehat{g}_h(y)$ дает (20) для $h = 1 - Y_{(n)}$, т.е.

$$\widehat{g}_h(T_{\widehat{\gamma}}(x)) \simeq \frac{1}{h}B_2\left(\frac{T_{\widehat{\gamma}}(x)-Y_{(n)}}{h}\right)$$

$$= \frac{1}{h}\frac{\left(a_2(p)-a_1(p)\left(T_{\widehat{\gamma}}(x)-Y_{(n)}\right)/h\right)K\left((T_{\widehat{\gamma}}(x)-Y_{(n)})/h\right)}{a_0(p)a_2(p)-a_1^2(p)} \sim (1+\widehat{\gamma}x)^{-1/(2\widehat{\gamma})}.$$

Тогда из (19) следует, что

$$\widehat{f}(x) \simeq (1+\widehat{\gamma}x)^{-(1/\widehat{\gamma}+1)}.$$

Отсюда, для обоих ядер хвост оценки $\widehat{f}(x)$ совпадает с хвостом (14), когда $\widehat{\gamma}$ близко к γ [5]. Однако для ядра K_b верно

$$K\left(\frac{T_{\widehat{\gamma}}(x)-Y_{(n)}}{h}\right) \sim (1+\widehat{\gamma}x)^{-1/\widehat{\gamma}} \qquad \text{при} \qquad x \to \infty$$

и (20) не выполняется. С другой стороны, h на границе можно найти из равенства

$$\frac{1}{h}B_2\left(\frac{T_{\widehat{\gamma}}(x)-Y_{(n)}}{h}\right)(1+\widehat{\gamma}x)^{-1/(2\widehat{\gamma})-1} = (1+\widehat{\gamma}x)^{-(1/\widehat{\gamma}+1)}.$$

Рассмотрим трансформацию $T(x) = (2/\pi)\operatorname{arctg} x$. Тогда ПР трансформированной сл.в. $g(x)$ непрерывна на $[0,1]$, если $\gamma \leq 1$ (см. (18)). Для $\gamma > 1$ $g(x) \to \infty$ при $x \to 1$ и оценивание $g(x)$ вблизи 1 проблематично. Более того, $g''(x)$ непрерывна на $[0,1]$ для некоторого $\ell(x)$, когда $\gamma \leq 1/3$ (см. п. 2.4). В последнем случае оценка с ядром B_1 или B_2 могла бы быть применена к $g(x)$ в пограничных точках для сокращения смещения. В других случаях можно использовать пограничные ядра, подобные рекомендованным в [21]. В любом случае, так как $T(x)$ не зависит от γ, оценка $\widehat{f}(x) = 2\widehat{g}_h\left(\frac{2}{\pi}\operatorname{arctg} x\right)/(\pi(1+x^2))$ не зависит от γ, когда $K(x)$ или h выбираются независимо от γ. Тогда порядок убывания, достижимый для класса (14) не обеспечивается.

Выбирая пограничное ядро

$$K(y) = \frac{\pi}{2}\frac{1+\operatorname{tg}^2(\pi y/2)}{(1+\widehat{\gamma}\operatorname{tg}(\pi y/2))^{1/\widehat{\gamma}+1}},$$

совпадающее с $g(x)$, для любых $\gamma > 0$ получим

$$\widehat{g}_h(y) \simeq \frac{\pi}{2h}\frac{1+\operatorname{tg}^2(\pi(y-Y_{(n)})/(2h))}{(1+\widehat{\gamma}\operatorname{tg}(\pi(y-Y_{(n)})/(2h))^{1/\widehat{\gamma}+1}}.$$

Возьмем $h = 1-\pi Y_{(n)}/(2\operatorname{arctg} x)$ для $x > \operatorname{tg}((\pi/2)Y_{(n)})$.

Тогда $\widehat{f}(x) = 2\widehat{g}_h\left((2/\pi)\operatorname{arctg} x\right)/(\pi(1+x^2))$ стремится к хвосту (14) при $x \to \infty$. Здесь ширина окна h зависит от x.

Конечно, последнее ядро выглядит довольно искусственно, так как оно подобно (18). Вместо целевой ПР в качестве пограничного ядра, можно использовать γ—зависимую ширину окна h для обычного ядра $K(x)$, чтобы обеспечить правильный порядок

[5] Параметр γ может быть оценен методом Хилла [24] или каким-либо другим методом по выборкам ограниченного объема (см. например, [25—26]).

убывания на бесконечности. Например, если $\gamma \leq 1/3$, то h на границе можно найти из равенства

$$\widehat{f}(x) \simeq \frac{1}{h} B_2 \left(\frac{(2/\pi)\operatorname{arctg}(x) - Y_{(n)}}{h} \right) \frac{2}{\pi(1+x^2)} = (1+\widehat{\gamma}x)^{-(1/\widehat{\gamma}+1)}.$$

4. Метод сглаживания

На практике для выборок ограниченных объемов адаптивный выбор сглаживающего параметра h по эмпирическим данным обычно оказывается лучше теоретически рекомендуемых значений, например $h = n^{-1/9}$, следующих из асимптотических результатов.

Найдем порядок сглаживающего параметра h для ядерной оценки с переменной шириной окна $\widetilde{f}^A(x) = \widetilde{f}^A(x|h_1, h)$ в случае, если h определяется из уравнения невязки

$$\rho(F_h^A, F_n) = \delta,$$

где $F_h^A(x) = \int_0^x \widetilde{f}^A(t)dt$, δ — известная неточность оценивания ФР $F(x)$ эмпирической ФР $F_n(t)$, т.е. $\delta = \rho(F, F_n)$; $\rho(\cdot,\cdot)$ — метрика в пространстве ФР. Предполагается, что h_1 для (2) оптимально, т.е. порядка $n^{-1/5}$. Методы невязки ранее рассматривались в [17, 27-28] для сглаживания непараметрических оценок ПР, причем неизвестная на практике δ определялась как некоторая квантиль предельного распределения статистики Мизеса—Смирнова или статистики Колмогорова—Смирнова [6]

$$D_n = \sqrt{n} \sup_{-\infty<x<\infty} |F(x) - F_n(x)|.$$

Пусть теперь h — решение уравнения

$$\sup_{x\in\Omega^*} |F_n(x) - F_h^A(x)| = n^{-4/9} \left(\ln n + (\ln n)^{-9/8}/2\right), \tag{22}$$

где $\Omega^* \subseteq [0, \infty)$ - некоторый конечный интервал.

Т е о р е м а 1. *Пусть $X^n = \{X_1, ..., X_n\}$ — независимые одинаково распределенные сл.в. с ПР $f(x)$, которая определена на Ω^*. Производные $f(x)^{(4)}$ и $\varphi(x) = (d/dx)^4 (1/f(x))$ непрерывны на Ω^*. Пусть $f(x)$ оценивается ядерной оценкой $\widetilde{f}^A(t)$ с неотрицательным непрерывным ядром $K(t)$ таким, что $K_3 = \int x^4 K(x)dx < \infty$,*

[6] Распределения этих статистик не зависят от $F(x)$.

$\sup_x K(x) < \infty$, $\int K(x)dx = 1$ *и* K', K'' *ограничены. Кроме того,* $cov(Z, \widehat{f}^A(x|h)) = 0$, *где* Z — *стандартная нормальная сл.в., а* $c_\varphi(x) = \int_0^x \varphi(t)dt > 0$. *Обозначим* $\xi(t) = c^2 + f(t)^{3/2}K_2$, $K_2 = \int K^2(t)dt$, c — *некоторая константа. Пусть* h *решение уравнения (22). Тогда существует хотя бы одно* h, *содержащееся в интервале* $[n^{-1/9}, \infty)$. *Если существуют различные* h, *удовлетворяющие (22), то для максимального из них* h_{max} *верно неравенство:*

$$P\{h_{max} < n^{-1/9}\} < 2\exp(-2A^2n^{1/9}) + 2 - 2\Phi\left(\frac{\ln n - a(n)}{b}\right),$$

$$P\{h \notin [n^{-1/9}, (\ln n)^{1/4}n^{-1/9}]\} < 2\exp(-2A^2n^{1/9}) + 2 - 2\Phi\left(\frac{\ln n - a(n)}{b}\right) + \Phi\left(\frac{1}{b\ln n}\right),$$

где A — *константа,* $0 < A < (\ln n)^{-9/8}/2$, $a(n) = B_f(x^*, n)$, $b = \int_0^{x^*} \xi(t)^{1/2}dt$, $B_f(x,n) = (K_3/24)c_\varphi(x) + xo\left(n^{-4/9}\right)n^{4/9}$, x^* — *некоторая точка на интервале* Ω^*, *а* K *таково, что* $a(n) \geq 1$.

Т е о р е м а 2. *В предположениях теоремы 1 верно*

$$P\{\overline{\lim}_{n\to\infty} n^{8/9} v(\widetilde{f}^A(x)) \leq c_1^*\} = 1,$$

где $v(\widetilde{f}^A(x))$ — *дисперсия оценки* $\widetilde{f}^A(x)$,

$$P\{\overline{\lim}_{n\to\infty} n^{8/9}(\ln n)^{-2} MSE(\widetilde{f}^A(x)) \leq c_2^*\} = 1,$$

где c_1^*, c_2^* — *не зависящие от* n *величины.*

З а м е ч а н и е. ПР (15) является примером ПР, удовлетворяющей предположениям теоремы 1 для $\Omega^* = [0, 1-\varepsilon]$, $0 < \varepsilon < 1$, так как $g(x)^{(4)}$ и

$$\left(\frac{d}{dx}\right)^4 \frac{1}{g(x)} \simeq \frac{1}{2}\left(\frac{\gamma}{\widehat{\gamma}}\right)^{1+1/\gamma}\left(2\frac{\widehat{\gamma}}{\gamma} - 1\right)2\frac{\widehat{\gamma}}{\gamma}\left(2\frac{\widehat{\gamma}}{\gamma} + 1\right)\left(2\frac{\widehat{\gamma}}{\gamma} + 2\right)(1-x)^{-2\widehat{\gamma}/\gamma - 3}$$

непрерывны на Ω^*, а $c_\varphi(x) > 0$, $\forall x \in \Omega^*$, при состоятельных оценках $\widehat{\gamma}$. Условие $a(n) \geq 1$ достигается, например, для ядра Епанечникова при соответствующем выборе ε.

5. Заключение

Известно, что ядерные оценки обеспечивают хорошие асимптотические $MISE$ и MSE при достаточно гладких ПР. Это демонстрируют ядерные оценки с переменной шириной окна, которые и без предварительной трансформации имеют $MISE \sim n^{-8/9}$

($MSE \sim n^{-8/9}$), если четвертая производная ПР существует. Однако для выборок ограниченного объема трансформации помогают улучшить оценивание в случае многомодальных, "скошенных" ПР (см. [9-10]) или тяжело-хвостовых ПР [11-13].

Для практиков более привлекательными являются простые преобразования, не требующие знания вида распределения, как например $\ln x$ или $(2/\pi)\,\mathrm{arctg}\, x$. Тем не менее они могут привести к разрывным ПР трансформированных сл.в., трудным для оценивания. Если класс распределений известен (здесь мы рассматриваем распределения с правильно меняющимися хвостами), то оптимальный порядок $MISE$ может быть получен при помощи фиксированных трансформаций не для всех ПР из класса. Трансформация arctg имеет ограничения на γ и на медленно меняющуюся функцию $\ell(x)$, а логарифмическое преобразование — на $\ell(x)$ и область определения Ω оптимальной $MISE$. Адаптивная трансформация (4) также имеет ограничение на Ω (см. п. 2.3). Очевидно, что для других классов ПР также могут быть выявлены ограничения на применимость фиксированных преобразований.

Вообще говоря, без предположений о форме хвоста восстановить точно хвост ПР непараметрическими методами нельзя, так как в этом случае невозможно предсказать форму ПР трансформированной сл.в. и правильно подобрать пограничное ядро. Порядок убывания трансформированных оценок на хвостах ПР может быть близким к таковому для истинной ПР при определенных граничных ядрах и параметрах сглаживания для них.

Принципиально новой идеей в статье является выбор параметра сглаживания h методом невязки в ядерной оценке с переменной шириной окна. Метод обеспечивает оптимальную дисперсию $O\left(n^{-8/9}\right)$ для оценки $\widetilde{f}^A(x)$ и $MSE(\widetilde{f}^A) \sim n^{-8/9}\,(\ln n)^2$, когда четвертые производные ПР $f(x)$ и функции $1/f(x)$ непрерывны, а распределение таково, что оценка (2) некоррелирована со стандартной нормальной сл.в.. Хотя этот порядок $MSE(\widetilde{f}^A)$ больше оптимального $\sim n^{-8/9}$, метод невязки позволяет оценить h по конечной выборке без оценивания производных неизвестной ПР, как требует теория для достижения оптимальной MSE. Использование адаптивных трансформаций позволяет получить сл.в. с ПР, удовлетворяющими указанным свойствам. Это принципиально невозможно для фиксированных трансформаций.

Приложение

Д о к а з а т е л ь с т в о   т е о р е м ы 1.

Обозначим $\Delta^h(n) = \sup_{x\in\Omega^*} |F(x) - F_h^A(x)|$. Очевидно, что

$$\sup_{x\in\Omega^*} |F_n(x) - F_h^A(x)| \le \sup_{x\in\Omega^*} |F_n(x) - F(x)| + \Delta^h(n), \tag{П.1}$$

$$\Delta^h(n) \le \sup_{x\in\Omega^*} |F_n(x) - F(x)| + \sup_{x\in\Omega^*} |F_n(x) - F_h^A(x)|. \tag{П.2}$$

Заметим, что выполнено

$$\widetilde{f}^A(x) - f(x) = \widetilde{f}^A(x) - E\widetilde{f}^A(x) + E\widetilde{f}^A(x) - f(x). \tag{П.3}$$

В [14] показано, что если $h_1 \cdot h^{-1} \to 0$ при $n \to \infty$, то

$$\widetilde{f}^A(x) = \widehat{f}^A(x|h) + cZ(nh)^{-1/2} + o((nh)^{-1/2}),$$

где c — константа, Z — стандартная нормальная сл.в. Отсюда $\widetilde{f}^A(x)$ имеет тоже асимптотическое смещение

$$E\widetilde{f}^A(x) - f(x) = \frac{K_3}{24} h^4 \varphi(x) + o(h^4), \tag{П.4}$$

что и $\widehat{f}^A(x|h)$ [14], а дисперсия, в силу $cov(Z, \widehat{f}^A(x|h)) = 0$, равна

$$\begin{aligned} v\left(\widetilde{f}^A(x)\right) &= v\left(\widehat{f}^A(x|h)\right) + c^2(nh)^{-1} + o((nh)^{-1}) \\ &= (nh)^{-1}\left(c^2 + f(x)^{3/2}K_2\right) + o((nh)^{-1}) = (nh)^{-1}\xi(x) + o((nh)^{-1}). \end{aligned} \tag{П.5}$$

Аналогично [14, с. 41], центральная предельная теорема для $\widetilde{f}^A(x)$ формулируется так: $\widetilde{f}^A(x) - E\widetilde{f}^A(x)$ асимптотически нормально распределена со средним 0 и дисперсией (П.5), т.е. равномерно по $t \in (-\infty, \infty)$

$$P\{(\widetilde{f}^A(x) - E\widetilde{f}^A(x))/(v\widetilde{f}^A(x))^{1/2} < t\} \to \Phi(t),$$

$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy$. Это означает, что для достаточно больших n $\widetilde{f}^A(x) - E\widetilde{f}^A(x)$ может быть представлена для фиксированного x как

$$\widetilde{f}^A(x) - E\widetilde{f}^A(x) = N(x)(nh)^{-1/2}\xi(x)^{1/2} + o((nh)^{-1/2}), \tag{П.6}$$

где $N(x)$ — асимптотическая стандартная нормальная сл.в. Для различных величин $x_1, ..., x_m$ сл.в. $N(x)$ асимптотически независимы. Из непрерывности и ограниченности $\widetilde{f}^A(x) - E\widetilde{f}^A(x)$ на ограниченном интервале Ω^*, непрерывности ядра следует, что $N(x)$ непрерывна и ограниченна на Ω^*.

В предположениях теоремы $f(x)^{3/4}$ и $\varphi(x)$ непрерывны. Отсюда имеем из (П.3), (П.4), (П.6), что

$$\text{(П.7)} \qquad \Delta^h(n) = \sup_{x\in\Omega^*} |\int_0^x (f(t) - \widetilde{f}^A(t))dt|$$

$$= \sup_{x\in\Omega^*} |\int_0^x \left(N(t)(nh)^{-1/2}\xi(t)^{1/2} + \frac{K_3}{24}h^4\varphi(t) + o(h^4)\right) dt|.$$

Рассмотрим два значения h: $h_1 = n^{-1/9}$ и $h_2 = (\ln n)^{1/4} n^{-1/9}$. Пусть

$$\text{(П.8)} \qquad \sup_{x\in\Omega^*} |\int_0^x N(t)\xi(t)^{1/2}dt + B_f(x,n)| < \ln n,$$

$$\text{(П.9)} \qquad \inf_{x\in\Omega^*} \int_0^x N(t)\xi(t)^{1/2}dt > 1/(\ln n).$$

Отсюда и из (П.7) имеем, что

$$\Delta^{h_1}(n) = n^{-4/9} \sup_{x\in\Omega^*} |\int_0^x N(t)\xi(t)^{1/2}dt + B_f(x,n)| < n^{-4/9}\ln n,$$

$$\Delta^{h_2}(n) = n^{-4/9} \sup_{x\in\Omega^*} |\int_0^x N(t)\xi(t)^{1/2}dt\,(\ln n)^{-1/8} + B_f(x,n)\ln n|$$

$$\geq n^{-4/9}\left((\ln n)^{-9/8} + B_f(x^*,n)\ln n\right) > n^{-4/9}\left((\ln n)^{-9/8} + \ln n\right),$$

в силу предположения $B_f(x^*,n) \geq 1$. Предположим, что

$$\text{(П.10)} \qquad \sup_x |F_n(x) - F(x)| \leq n^{-4/9}A,$$

где $0 < A < (\ln n)^{-9/8}/2$. Тогда очевидно следует из (П.1) и для h_1, что

$$\text{(П.11)} \qquad \sup_{x\in\Omega^*} |F_n(x) - F_h^A(x)| \leq n^{-4/9}(\ln n + A),$$

из (П.2) и для h_2, что

$$\sup_{x\in\Omega^*} |F_n(x) - F_h^A(x)| \geq n^{-4/9}((\ln n)^{-9/8} + \ln n - A).$$

Заметим, что для выбранного A верно $\ln n + A < (\ln n)^{-9/8} + \ln n - A$. Это означает, что вследствие непрерывности $\sup_{x\in\Omega^*} |F_n(x) - F_h^A(x)|$ по h решение уравнения (22) расположено между h_1 и h_2 в предположениях (П.8)—(П.10). Отсюда

$$\begin{aligned} P\{h \notin [h_1, h_2]\} &< P\{\sup_x |F_n(x) - F(x)| > n^{-4/9}A\} \\ &+ P\{\sup_{x\in\Omega^*} |\int_0^x N(t)\xi(t)^{1/2}dt + B_f(x,n)| \geq \ln n\} \\ &+ P\{\inf_{x\in\Omega^*} \int_0^x N(t)\xi(t)^{1/2}dt \leq 1/(\ln n)\}. \end{aligned}$$

Первое слагаемое правой части может быть оценено, используя асимптотическую оценку скорости сходимости эмпирической ФР к истинной ФР [18]:

$$P\{\sup_x |F_n(x) - F(x)| > \eta\} \leq 2\exp(-2n\eta^2).$$

В силу непрерывности $N(t)$ на Ω^* и по теореме о среднем имеем, что

$$P\{\sup_{x\in\Omega^*} |\int_0^x N(t)\xi(t)^{1/2}dt + B_f(x,n)| \geq \ln n\} = 2 - 2\Phi\left(\frac{\ln n - a(n)}{b}\right).$$

Аналогично имеем, что

$$P\{\inf_{x\in\Omega^*} \int_0^x N(t)\xi(t)^{1/2}dt \leq 1/(\ln n)\} = \Phi\left(\frac{1}{b\ln n}\right).$$

Отсюда вероятность того, что решение (22) расположено вне интервала $[h_1, h_2]$, меньше, чем $2\exp(-2A^2n^{1/9}) + 2 - 2\Phi\left((\ln n - a(n))/b\right) + \Phi\left(1/(b\ln n)\right)$, где $a(n)$ убывает по n, т.е. выполнено второе утверждение теоремы.

С другой стороны, из (П.7) следует, что $\Delta^h(n) \to \infty$ при $h \to \infty$. Тогда из (П.2) получим, что и $\sup_{x\in\Omega^*} |F_n(x) - F_h^A(x)| \to \infty$. Вместе с (П.11) это означает, что существует решение (22) h такое, что $h \geq n^{-1/9}$, если выполнены (П.8) и (П.10). Тогда если h_{max} - максимальное решение (22), то

$$\begin{aligned} P\{h_{max} < n^{-1/9}\} &< P\{\sup_x |F_n(x) - F(x)| > n^{-4/9}A\} \\ &+ P\{\sup_{x\in\Omega^*} |\int_0^x N(t)\xi(t)^{1/2}dt + B_f(x,n)| \geq \ln n\}. \end{aligned}$$

Отсюда следует первое утверждение теоремы.

Доказательство теоремы 2.

Предположим, что $h \geq n^{-1/9}$. Тогда из (П.5) получим, что

$$v(\widetilde{f}^A(x)) \leq c_1^* n^{-8/9},$$

где $c_1^* = \xi(x) + 1$. Отсюда имеем

$$P\{v(\widetilde{f}^A(x)) > c_1^* n^{-8/9}\} < P\{h < n^{-1/9}\}$$

$$< 2\exp(-2A^2 n^{1/9}) + 2 - 2\Phi\left(\frac{\ln n - a(n)}{b}\right) = \psi_1(n).$$

Пусть $n^{-1/9} \leq h \leq (\ln n)^{1/4} n^{-1/9}$. Отсюда и из (П.4), (П.5) следует:

$$MSE(\widetilde{f}^A(x)) = (K_3/24)^2 h^8 (\varphi(x))^2 + (nh)^{-1}\xi(x) + o(h^8) \leq c_2^* n^{-8/9} (\ln n)^2,$$

$$c_2^* = (K_3/24)^2 (\varphi(x))^2 + \xi(x) + 1.$$

Тогда из теоремы 1 следует, что

$$P\{MSE(\widetilde{f}^A(x)) > c_2^* n^{-8/9} (\ln n)^2\} < P\{h \notin [n^{-1/9}, (\ln n)^{1/4} n^{-1/9}]\}$$

$$< 2\exp(-2A^2 n^{1/9}) + 2 - 2\Phi\left(\frac{\ln n - a(n)}{b}\right) + \Phi\left(\frac{1}{b \ln n}\right) = \psi_2(n).$$

Так как ряды $\sum_{n=1}^{\infty} \psi_1(n)$, $\sum_{n=1}^{\infty} \psi_2(n)$ сходятся, то согласно лемме Бореля-Кантелли справедливы утверждения теоремы.

СПИСОК ЛИТЕРАТУРЫ

1. *Asmussen S., Klüppelberg C., Sigman K.* Sampling in subexponential times, with queueing applications //Stochastic Proc. Appl. 1999. No. 79. P. 265–286.
2. *Silverman B.W.* Density Estimation for Statistics and Data Analysis. N.Y., Chapman&Hall, 1986.
3. *Breiman L., Meisel W., Purcell E.* Variable kernel estimates of multivariate densities //Technometrics. 1977. No. 19. P. 135–144.
4. *Abramson I.S.* On bandwidth estimation in kernel estimators - A square root law // Ann. Statist. 1982. No. 10. P. 1217–1223.
5. *Hall P.* On global properties of variable bandwidth density estimators // Ann. Statist. 1992. No. 20 (2). P. 762–778.
6. *Naito K.* On a certain class of nonparametric density estimators with reduced bias // Statist. Prob. Lett. 2001. No. 51. P. 71-78.
7. *Novak S.Y.* Generalised kernel density estimator // Theory Probab. Appl. 1999. No. 44 (3). P. 570–583.
8. *Devroye L., Györfi L.* Nonparametric density estimation: The L_1 view. N.Y.: Wiley, 1985.
9. *Wand M.P., Marron J.S., Ruppert, D.* Transformations in density estimation // J. American Statist. Assoc. Theory and Methods. 1991. No. 86(414). P. 343–353.
10. *Yang L., Marron J.S.* Iterated transformation-kernel density estimation // J. American Statist. Assoc. 1999. No. 94(446). P. 580–589.
11. *Markovitch N.M., Krieger U.R.* Nonparametric estimation of long-tailed density functions and its application to the analysis of World Wide Web traffic // Performance Evaluation. 2000. No. 42(2-3). P. 205-222.
12. *Markovich N.M.* Transformed Estimates of Densities of Heavy-Tailed Distributions and Classification // Autom. Remote Control. 2002. No. 63(4). P. 627-640.
13. *Maiboroda R.E., Markovich N.M.* Estimation of heavy-tailed probability density function with application to Web data // Comput. Stat. 2004. No. 4. P. 43 - 67.
14. *Hall P., Marron J.S.* Variable window width kernel estimates of probability densities // Prob. Theory Related Fields. 1988. No. 80. P. 37 - 49.
15. *Scott D.W.* Multivariate Density Estimation Theory, Practice and Visualization. N.Y.: Wiley, 1992.
16. *Chow Y.-S., Geman S., Wu, L.-D.* Consistent cross-validated density estimation // Ann. Statist. 1983. No. 11. P. 25–38.

17. *Markovich N.M.* Experimental analysis of nonparametric probability density estimates and of methods for smoothing them // Autom. Remote Control. 1989. No. 50. P. 941-948.
18. *Prakasa Rao B.L.S.* Nonparametric Functional Estimation. N.Y.: Academic Press, 1983.
19. *Mikosch T.* Regular Variation, Subexponentiality and Their Applications in Probability Theory // Report 99-013. University of Groningen. 1999.
20. *Jurečková J., Picek J.* A class of tests on the tail index // Extremes. 2001. No. 4. P. 165-183.
21. *Scott D.W.* Multivariate Density Estimation. Theory, Practice, and Visualization. N.Y.: John Wiley& Sons, Inc., 1992.
22. *Schuster J.* Incorporating support constraints into nonparametric estimators of densities //Commun. Stat. Theory Meth. 1985. No. 14 (5). P. 1123–1136.
23. *Simonoff J.S.* Smoothing methods in statistics. N.Y.: Springer, 1996.
24. *Hill B.M.* A simple general approach to inference about the tail of a distribution // Ann.Statist. 1975. No. 3. P. 1163-1174.
25. *Resnick S., Stărică C.* Smoothing the Moment Estimate of the Extreme Value Parameter // Extremes. 1999. No. 1(3). P. 263-294.
26. *Dekkers A.L., Einmahl J.H.J., de Haan L.* A moment estimator for the index of an extreme-value distribution // Ann. Statis. 1989. No. 17(4). P. 1833-1855.
27. *Айду Ф.А., Вапник В.Н.* Оценивание плотности вероятностей на основе метода стохастической регуляризации // АиТ. 1989. No. 4. С. 84–97.
28. *Vapnik V.N., Markovich N.M., Stephanyuk A.R.* Rate of convergence in L_2 of the projection estimator of the distribution density // Autom. Remote Control. 1992. No. 53. P. 677–686.