

Гамма-ядерные оценки многомерной плотности и её частной производной по зависимым данным

Л. А. МАРКОВИЧ

Институт проблем управления им. Трапезникова РАН,
Институт проблем передачи информации им. Харкевича РАН
e-mail: markovichliubov@gmail.com

УДК 519.213

Ключевые слова: гамма-ядро, многомерная плотность, непараметрические оценки, производная плотности распределения.

Аннотация

Рассматриваются непараметрические оценки многомерной плотности распределения вероятности и её частной производной с носителем на положительной полуоси по зависимым данным. Использован класс ядерных оценок с асимметричным гамма-ядром. Гамма-ядра неотрицательны, меняют свою форму в зависимости от положения на полуоси и обладают хорошими граничными свойствами для широкого класса плотностей. Получены асимптотические оценки многомерной плотности и её частных производных, такие, как смещения, дисперсии и ковариации. Оптимальный параметр сглаживания получен для обеих оценок как минимум средней квадратичной ошибки (MISE) по зависимым данным с сильным перемешиванием. Найдены оптимальные скорости сходимости MISE как для плотности, так и для её производной.

Abstract

L. A. Markovich, Nonparametric estimation of multivariate density and its derivative by dependent data using gamma kernels, Fundamentalnaya i prikladnaya matematika, vol. 22 (2018), no. 3, pp. 145–177.

We consider the nonparametric estimation of the multivariate probability density function and its partial derivative with a support on nonnegative axis by dependent data. We use the class of kernel estimators with asymmetric gamma kernel functions. The gamma kernels are nonnegative, they may change their shape depending on the position on the semi-axis and possess good boundary properties for a wide class of densities. Asymptotic estimates of the multivariate density and of its partial derivatives such as biases, variances, and covariances are derived. The optimal bandwidth of both estimates is obtained as a minimum of the mean integrated squared error (MISE) by dependent data with a strong mixing. Optimal convergence rates of the MISE both for the density and its derivative are found.

1. Введение

Плотность вероятности с носителем на положительной полуоси используется для моделирования широкого спектра приложений в области инженерии,

при обработке сигналов, в медицинских исследованиях, при контроле качества, в актуарных науках и климатологии. Так, для оптимальной фильтрации при обработке сигналов и управлении нелинейными процессами рассматривается экспоненциальный класс плотностей (см. [13]). В задачах, связанных со страховыми случаями, часто встречаются гамма-плотности (см. [17]), которые определены на положительной полуоси, смещены вправо и унимодальны. Гамма-распределения используются для моделирования осадков (см. [1]), а распределения Эрланга и χ^2 — при моделировании страховых портфелей (см. [19]).

Широкое использование этих плотностей на практике приводит к необходимости их оценки с помощью конечных выборок. Непараметрические ядерные оценки в этом случае — один из самых популярных методов. Ядерные оценки были первоначально введены для одномерных независимых одинаково распределённых наблюдаемых случайных величин и для симметричных ядер в [31, 32]. Если плотность распределения определена на неограниченном носителе, то данный подход показывает хорошие результаты. Для плотности, определённой на носителе $[0, \infty)$, использование классических ядерных оценок с симметричными ядрами даст существенное смещение на границе вблизи нуля, что ухудшает качество такой оценки (см. [40]). Очевидно, что граничное смещение для многомерных ядер играет ещё более существенную роль (см. [5]). Существует множество методов по сокращению граничного смещения, например: метод отражений [36], метод граничных ядер [27], гибридный метод [18], метод локальной линейной оценки [21, 23] и др. Другой подход заключается в использовании несимметричных ядер. В случае одинаковых положительных независимо распределённых случайных величин оценка плотности с помощью гамма-ядра была предложена в [10]. В [6] гамма-ядерная оценка была распространена на случай одномерных зависимых случайных величин, а в [5] — на случай многомерных независимых случайных величин.

Гамма ядра неотрицательны и меняют свою форму в зависимости от положения на полуоси. Это позволяет хорошо оценивать многомодальные плотности и их производные. Оценки с гамма-ядром не имеют граничного смещения в случае $f''(0) = 0$ («условие плеча»), т. е. когда плотность $f(x)$ имеет выпуклость в точке $x = 0$ (см. [41, формула (4.3)]). Многие распределения экспоненциального семейства удовлетворяют этому условию. Ряд методов, устраняющих граничное смещение без условия на плечо, приведён в [20] для одномерных одинаково распределённых случайных величин, а в [16] — для многомерных одинаково распределённых случайных величин. Существуют другие асимметричные ядерные оценки, например на основе обратного гауссова и взаимно-обратного гауссова ядер [34]. Сравнение этих ядер с гамма-ядром приведено в [7].

Кроме плотности, часто нужно оценить и её производную. В [14, 15] была предложена гамма-ядерная оценка по одномерным зависимым данным, а в [25] — для зависимых данных с сильным перемешиванием. Порядок сходи-

мости таких оценок $n^{-4/7}$ с параметром размытости $n^{-2/7}$. В [40] оптимальная средняя интегральная ошибка (MISE) ядерной оценки для первой производной плотности имеет порядок сходимости $n^{-4/7}$, что соответствует параметру размытости порядка $n^{-1/7}$ для симметричных ядер.

1.1. Результаты статьи

Впервые введена мультипликативная гамма-ядерная оценка для многомерной плотности с неотрицательным носителем и её частных производных по многомерным зависимым данным с сильным перемешиванием. Получена асимптотика оценок и оптимальные значения параметров сглаживания, минимизирующие MISE. Заметим, что оценка производной требует определённого выбора параметра сглаживания, отличающегося от оптимального параметра для плотности.

1.2. Практическая значимость

На практике чаще всего встречаются последовательности зависимых наблюдений. Чтобы ослабить условие независимости случайных величин, можно рассматривать стационарные процессы, удовлетворяющие условию сильного перемешивания. В качестве примера таких процессов можно привести авторегрессионные процессы. Наряду с оценкой функции плотности и её производной по зависимым выборкам оценка логарифмической производной плотности является актуальной проблемой (см. уравнение для оптимальной фильтрации в [24], формулы (2.4), (2.8)). Логарифмическая производная плотности — это отношение производной плотности к самой плотности. Оценка производной плотности имеет практическое применение для оптимальной фильтрации при обработке сигналов и управлении нелинейными процессами, где используется только экспоненциальный класс плотностей (см. [11, 13]). Подробный обзор см. в разделе 4.3. Производная плотности может быть полезна для нахождения наклона кривой плотности, её локальных экстремумов, седловых точек, существенных особенностей в данных, а также в регрессионном анализе (см. [8]). Производная плотности также играет ключевую роль в кластеризации данных посредством поиска мод (см. [33]).

Содержание статьи следующее. В разделе 2 определены гамма-ядерные оценки для многомерной совместной плотности распределения и её частных производных. В разделе 3 получены смещение, дисперсия и ковариация оценки совместной плотности. С использованием этих результатов получен оптимальный параметр сглаживания и соответствующая скорость сходимости оптимального MISE. В разделе 3.2 то же получено для оценки частной производной плотности. В разделе 4 проведено моделирование. Раздел 5 содержит заключение. Доказательства теорем приведены в приложении.

2. Гамма-ядерное оценивание

Пусть $\{\mathbf{X}_i = (X_{i1}, \dots, X_{id})^T\}_{i=1}^n$ — сильно стационарная последовательность d -мерных случайных величин с неизвестной плотностью распределения $f(x_1^d)$, определённой на $x_1^d = (x_1, \dots, x_d)^T \in \mathbf{R}^{+d}$. Последовательность $\{\mathbf{X}_s\}$ удовлетворяет условию сильного перемешивания (α -перемешивания) с коэффициентом $\alpha(k)$, если

$$\alpha(k) = \sup_t \sup_{\substack{A \in \sigma\{X_s, s \leq t\} \\ B \in \sigma\{X_s, s \geq t+k\}}} |P(A \cap B) - P(A)P(B)|, \quad k \geq 1. \quad (1)$$

Коэффициент $\alpha(k) \rightarrow 0$ при $k \rightarrow \infty$. Условию α -перемешивания, или сильно-го перемешивания, удовлетворяют многие стохастические процессы, например авторегрессионный процесс первого порядка (см. [2]). В дальнейшем такую последовательность будем обозначать $\{\mathbf{X}_s\} \in \mathcal{M}(\alpha)$. Для оценки неизвестной многомерной плотности распределения было предложено мультипликативное ядро [5]

$$\hat{f}(x_1^d) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d K_{\rho(x_j, b_j), b_j}(X_{ij}), \quad (2)$$

где $b \equiv \{b_{j,n}\}_{j=1}^d$ — параметры размытости, такие что $b \rightarrow 0$ при $n \rightarrow \infty$. В качестве ядра для каждой переменной выбрано гамма-ядро (см. [10])

$$K_{\rho(x,b), b}(t) = \frac{t^{\rho(x,b)-1} \exp(-t/b)}{b^{\rho(x,b)} \Gamma(\rho(x,b))} \quad (3)$$

где $\Gamma(\cdot)$ — гамма-функция. Чтобы улучшить свойства гамма-ядра, в [10] было предложено использовать параметр $\rho_b(x)$, определённый следующим образом:

$$\rho(x, b) = \begin{cases} \rho_1(x, b) = x/b & \text{при } x \geq 2b, \\ \rho_2(x, b) = (x/(2b))^2 + 1 & \text{при } x \in [0, 2b). \end{cases} \quad (4)$$

Поскольку гамма-ядро неотрицательно, оно не имеет веса на отрицательных полуосях в отличие от симметричных ядерных оценок. Следовательно, использование гамма-ядер более естественно для оценки плотностей с положительным носителем.

С использованием (4) оценка для плотности (2) может быть записана как

$$\hat{f}(x_1^d) = \begin{cases} (1/n) \sum_{i=1}^n \prod_{j=1}^d K_{\rho_1(x_j, b_j), b_j}(X_{ij}) & \text{при } x \geq 2b, \\ (1/n) \sum_{i=1}^n \prod_{j=1}^d K_{\rho_2(x_j, b_j), b_j}(X_{ij}) & \text{при } x \in [0, 2b). \end{cases} \quad (5)$$

В [15, 25] производная одномерной плотности оценивалась точно так же, как и производная гамма-ядерной оценки. Для многомерного случая можно аналогично оценить любые частные производные $f(x_1^d)$. Например, частная производная

$f(x_1^d)$ по x_k , $1 \leq k \leq d$, может быть оценена как

$$\hat{f}'_{x_k}(x_1^d) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d K_{\rho(x_j, b_j), b_j}(X_{ij}) (K_{\rho(x_k, b_k), b_k}(X_{ik}))'_{x_k}, \quad (6)$$

где

$$\begin{aligned} & (K_{\rho(x, b), b}(t))'_x = \\ & = \begin{cases} K'_{\rho_1(x, b), b}(t) = (1/b)K_{\rho_1(x, b), b}(t)L_1(t, x, b) & \text{при } x \geq 2b, \\ K'_{\rho_2(x, b), b}(t) = (x/2b^2)K_{\rho_2(x, b), b}(t)L_2(t, x, b) & \text{при } x \in [0, 2b), \end{cases} \quad (7) \end{aligned}$$

частная производная (3) и

$$L_i(t, x, b) = \ln t - \ln b - \Psi(\rho_i(x, b)), \quad i = 1, 2.$$

Здесь $\Psi(\rho_i(x, b))$ — дигамма-функция, логарифмическая производная гамма-функции.

Замечание 1. Математический аппарат, применяемый для оценки производной, аналогичен аппарату, применяемому для плотности. Однако формулы становятся намного сложнее, потому что появляются специальные функции дигамма, включающие вектор параметра сглаживания b . Следовательно, нужно вычлнить порядок по b из сложных выражений, содержащих логарифмы и специальные функции.

С использованием (7) формула (6) может быть записана как

$$\begin{aligned} \hat{f}'_{x_k}(x_1^d) &= \\ &= \begin{cases} (1/n) \sum_{i=1}^n (1/b_k)L_1(X_{ik}, x_k, b_k) \prod_{j=1}^d K_{\rho_1(x_j, b_j), b_j}(X_{ij}) & \text{при } x \geq 2b, \\ (1/n) \sum_{i=1}^n (x_k/2b_k^2)L_2(X_{ik}, x_k, b_k) \prod_{j=1}^d K_{\rho_2(x_j, b_j), b_j}(X_{ij}) & \text{при } x \in [0, 2b). \end{cases} \quad (8) \end{aligned}$$

Интегральная средняя квадратическая ошибка для зависимых случайных величин определяется как

$$\begin{aligned} \text{MISE}(\hat{f}(x_1^d)) &= \mathbb{E} \int_0^\infty (f(x_1^d) - \hat{f}(x_1^d))^2 dx, \\ \text{MISE}(\hat{f}'_{x_k}(x_1^d)) &= \mathbb{E} \int_0^\infty (f'_{x_k}(x_1^d) - \hat{f}'_{x_k}(x_1^d))^2 dx \end{aligned} \quad (9)$$

Используем её как меру ошибки предлагаемых оценок (5) и (8). Неизвестные параметры, сглаживающие (5) и (8), получаются как минимумы (9).

Замечание 2. Интегралы (9) могут быть разбиты на два: \int_0^{2b} , \int_{2b}^{∞} . В случае когда $x \geq 2b$, интеграл \int_0^{2b} стремится к нулю при $b \rightarrow 0$. Следовательно, можно его не рассматривать, в отличие от [41]. Первый интеграл имеет тот же порядок по b , что и второй интеграл, и, очевидно, не влияет на выбор оптимального параметра сглаживания. Поэтому мы будем опускать индексы в $\rho(x_j, b_j)$ и $L(X_{ik}, x_k, b_k)$.

Известно, что средняя квадратическая ошибка для зависимых случайных величин определяется как

$$\begin{aligned} \text{MSE}(\hat{f}(x_1^d)) &= \text{Bias}(\hat{f}(x_1^d))^2 + \text{Var}(\hat{f}(x_1^d)), \\ \text{MSE}(\hat{f}'_{x_k}(x_1^d)) &= \text{Bias}(\hat{f}'_{x_k}(x_1^d))^2 + \text{Var}(\hat{f}'_{x_k}(x_1^d)), \end{aligned} \quad (10)$$

где дисперсия для оценки плотности имеет вид

$$\begin{aligned} \text{Var}(\hat{f}(x_1^d)) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d K_{\rho(x_j, b_j), b_j}(X_{ij})\right) = \\ &= \frac{1}{n^2} \left(\sum_{i=1}^n \text{Var}(\tilde{K}(X_i, x, b)) + \sum_{i, j=1, i \neq j}^n \text{Cov}(\tilde{K}(X_i, x, b), \tilde{K}(X_j, x, b)) \right) = \\ &= \frac{1}{n} \overline{\text{Var}}(\hat{f}(x_1^d)) + \frac{2}{n} \sum_{i=1}^n \left(1 - \frac{i}{n}\right) \text{Cov}(\tilde{K}(X_i, x, b), \tilde{K}(X_{1+i}, x, b)), \end{aligned} \quad (11)$$

а для оценки частной производной имеем

$$\begin{aligned} \text{Var}(\hat{f}'_{x_k}(x_1^d)) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n \frac{1}{b_n} L(X_{in}, x_n, b_n) \prod_{j=1}^d K_{\rho(x_j, b_j), b_j}(X_{ij})\right) = \\ &= \frac{1}{n^2 b_k^2} \left(\sum_{i=1}^n \text{Var}(\tilde{K}(X_i, x, b)) + \sum_{i, j=1, i \neq j}^n \text{Cov}(\tilde{K}(X_i, x, b), \tilde{K}(X_j, x, b)) \right) = \\ &= \frac{1}{n b_k^2} \overline{\text{Var}}(\hat{f}'_{x_k}(x_1^d)) + \frac{2}{n b_k^2} \sum_{i=1}^n \left(1 - \frac{i}{n}\right) \text{Cov}(\tilde{K}(X_i, x, b), \tilde{K}(X_{1+i}, x, b)). \end{aligned} \quad (12)$$

Здесь мы использовали следующие обозначения:

$$\tilde{K}(X_i, x, b) = \prod_{j=1}^d K_{\rho(x_j, b_j), b_j}(X_{ij}), \quad \tilde{K}(X_i, x, b) = L(X_{ik}, x_k, b_k) \tilde{K}(X_i, x, b).$$

3. Основные результаты

3.1. Асимптотические свойства оценки плотности

В этом разделе исследуются асимптотические свойства оценки (5). Для этого мы находим смещение, дисперсию и ковариацию (10).

В следующих леммах, посвящённых смещению и дисперсии, предполагается, что $f(x_1^d)$ является дважды непрерывно дифференцируемой. Все доказательства проводились в предположении, что параметры размытости для каждого одномерного ядра, входящего в мультипликативное ядро, различны. Однако для практического использования целесообразно считать их одинаковыми $b_1 = b_2 = \dots = b_n = b$, что и было сделано в финальных утверждениях лемм.

Лемма 1. Пусть $b \rightarrow 0$ и все параметры сглаживания равны: $b_1 = b_2 = \dots = b_n = b$. Тогда смещение оценки плотности (5) имеет вид

$$\text{Bias}(\hat{f}(x_1^d)) = \frac{b}{2} \sum_{j=1}^d x_j \frac{\partial^2 f(x_1^d)}{\partial x_j^2} + o(b). \quad (13)$$

Следующая лемма посвящена дисперсии непараметрической оценки плотности.

Лемма 2. Пусть $b \rightarrow 0$, $nb^{\tau+1/2} \rightarrow \infty$ и все параметры размытости равны $b_1 = b_2 = \dots = b_n = b$. Тогда дисперсия оценки плотности (5) имеет вид

$$\begin{aligned} \text{Var}(\hat{f}(x_1^d)) = & \frac{b^{-d/2}}{n} \left(\prod_{j=1}^d \frac{x_j^{-1/2}}{2\sqrt{\pi}} \right) (f(x_1^d) + bv_1(x_1^d) + b^2v_2(x_1^d)) - \\ & - \frac{1}{n} \left(f(x_1^d) + \frac{1}{2} \sum_{j=1}^d x_j b \frac{\partial^2 f(x_1^d)}{\partial x_j^2} \right)^2 + o(b^2), \quad (14) \end{aligned}$$

где

$$\begin{aligned} v_1(x_1^d) = & \sum_{j=1}^d \left(-\frac{1}{2} \frac{\partial f(x_1^d)}{\partial x_j} + \frac{x_j}{4} \frac{\partial^2 f(x_1^d)}{\partial x_j^2} \right), \\ v_2(x_1^d) = & - \sum_{j=1}^d \sum_{i=1}^d \frac{x_j}{8} \frac{\partial^3 f(x_1^d)}{\partial x_j^2 \partial x_i}. \end{aligned}$$

Перейдём к ковариации, определённой в (11).

Лемма 3. Пусть

- 1) $\{X_j\}_{j \geq 1} \in \mathcal{S}(\alpha)$ и $\int_1^{\infty} \alpha(\tau)^v d\tau < \infty$, $0 < v < 1$,
- 2) $f(x_1^d)$ — дважды непрерывно дифференцируемая функция,
- 3) $b \rightarrow 0$ и $nb^{d(v+1)^2} \rightarrow \infty$ при $n \rightarrow \infty$.

Тогда ковариация $C(x)$ ограничена сверху:

$$|\text{Cov}(\hat{f}(x_1^d))| \leq \frac{b^{-d(v+1)/2}}{n} \times \\ \times \int_1^\infty \alpha(\tau)^v d\tau D(v, x_1^d) \left(bS(v, x_1^d) + f(x_1^d) \frac{3v-1}{2(v-1)} \right)^{1-v} + o(b^2),$$

где

$$S(v, x_1^d) = \sum_{i=1}^d \frac{v+1}{(v-1)^2 x_i} f(x_1^d) + \frac{v+1}{v-1} \frac{\partial f(x_1^d)}{\partial x_i} + \frac{x_i}{2} \frac{\partial^2 f(x_1^d)}{\partial x_i^2}, \\ D(v, x_1^d) = 2(2\pi)^{-(d(v+1)-2)/2} \left(\prod_{j=1}^d x_j^{-(v+1)/2} \right).$$

Используя приведённые результаты, можно получить верхнюю границу MISE (9) и найти выражение оптимального параметра сглаживания b как минимума последней. Имеет место следующая теорема.

Теорема 1. Если условия лемм 1–3 выполняются, то оптимальный параметр сглаживания, минимизирующий MISE, имеет порядок сходимости

$$b = \left(\frac{d(v+1)(3v-1)^{1-v}}{n(2v-2)^{1-v}} \left(\int D(v, x_1^d) f(x_1^d)^{1-v} dx_1^d / \right. \right. \\ \left. \left. / \int \left(\sum_{j=1}^d x_j \frac{\partial f(x_1^d)}{\partial x_j^2} \right)^2 dx_1^d \right) \int_1^\infty \alpha(\tau)^v d\tau \right)^{(d(v+1)+4)/2}, \quad (15)$$

где кратные интегралы обозначены как $\int \dots \int dx_1 \dots dx_d$ при $\int dx_1^d$.

Очевидно, что пользоваться оценкой из теоремы 1 нельзя, так как она содержит неизвестный коэффициент перемешивания. Чтобы устранить этот недостаток, применим результат следующей леммы.

Лемма 4. Если дополнительно к условиям леммы 3 выполняется, что $f_\tau(x, y)$ равномерно ограничена, т. е. существует $M > 0$, такое что для всех $\tau > 1$ и для всех $(x, y) \in \mathbb{R}^2$ справедливо $|f_\tau(x, y) - f(x)f(y)| \leq M$, то справедливо следующее утверждение о ковариации оценки многомерной плотности:

$$|\text{Cov}(\hat{f}(x_1^d))| \leq \frac{2}{n} \sum_{\tau=2}^{c(n)} \text{Cov}(\tilde{K}(X_1, x, b), \tilde{K}(X_\tau, x, b)) + \\ + \frac{2}{n} \sum_{\tau=c(n)+1}^\infty \text{Cov}(\tilde{K}(X_1, x, b), \tilde{K}(X_\tau, x, b)) = I_1 + I_2,$$

$$I_1 \leq \frac{2M}{nb^{d/8}},$$

$$I_2 \leq \frac{D(\kappa, x_1^d)}{nb^{d/16}} \left(f(x_1^d) \frac{6\kappa - 1}{2(2\kappa - 1)} + bS(\kappa, x_1^d) \right)^{1-2\kappa} \sum_{\tau=2}^{\infty} \tau \alpha(\tau)^{2\kappa} + o(b^2). \quad (16)$$

По лемме 4 справедливо следующее утверждение о скорости сходимости ковариации:

$$\text{Cov}(\hat{f}_n(x_1^d)) \sim n^{-1}b^{-d/8}.$$

Теперь мы можем пренебречь ковариацией оценки плотности в сравнении с её дисперсией $\text{var}(\hat{f}(x_1^d)) \sim n^{-1}b^{-d/2}$. Тогда справедлива следующая теорема.

Теорема 2. Если условия лемм 1–4 выполняются, то оптимальный параметр сглаживания, минимизирующий MISE, имеет порядок сходимости

$$b = \left(d \int_0^{\infty} \left(\prod_{j=1}^d \frac{x_j^{-1/2}}{2\sqrt{\pi}} \right) f(x_1^d) dx_1^d / \int_0^{\infty} \left(\sum_{j=1}^d x_j \frac{\partial^2 f(x_1^d)}{\partial x_j^2} \right)^2 dx_1^d \right)^{2/(4+d)} n^{-2/(4+d)}. \quad (17)$$

Следствие 1. Результаты теоремы 2 согласуются с результатами [10] для одномерной плотности и независимых данных.

3.2. Асимптотические свойства оценки частной производной плотности

Аналогично разделу 3 исследуем асимптотические свойства оценки частной производной плотности (8). Смещение, дисперсия и ковариация, определённые в (10), были найдены в следующих леммах.

Лемма 5. Пусть $b \rightarrow 0$ и все параметры размытости равны: $b_1 = b_2 = \dots = b_n = b$. Тогда смещение оценки производной плотности (8) имеет вид

$$\text{Bias}(\hat{f}'_{x_k}(x_1^d)) = bB_1(x_1^d) + b^2B_2(x_1^d) + o(b), \quad (18)$$

где были использованы следующие обозначения:

$$B_1(x_1^d) = \frac{f(x_1^d)}{12x_k^2} + \frac{1}{4x_k} \sum_{j=1}^d x_j \frac{\partial^2 f(x_1^d)}{\partial x_j^2}, \quad B_2(x_1^d) = \frac{1}{24x_k^2} \sum_{j=1}^d x_j \frac{\partial^2 f(x_1^d)}{\partial x_j^2}.$$

Лемма 6. Пусть $b \rightarrow 0$, $nb^{(\tau+3)/2} \rightarrow \infty$ и все параметры размытости равны: $b_1 = b_2 = \dots = b_n = b$. Тогда дисперсия оценки производной плотности (8) имеет вид

$$\text{Var}(\hat{f}'_{x_k}(x_1^d)) = \frac{b^{-d/2}}{n} \left(\prod_{j=1}^d \frac{x_j^{-1/2}}{2\sqrt{\pi}} \right) \left(bV_1(x_1^d) + b^2V_2(x_1^d) + \frac{1}{b}V_3(x_1^d) + V_4(x_1^d) \right) - \frac{1}{n} \left(b^2B_1^2(x_1^d) + \left(\frac{\partial f(x_1^d)}{\partial x_k} \right)^2 + 2 \frac{\partial f(x_1^d)}{\partial x_k} (bB_1(x_1^d) + b^2B_2(x_1^d)) \right) + o(b^2), \quad (19)$$

где

$$\begin{aligned} V_1(x_1^d) &= -\frac{1}{24x_k^2} \frac{\partial f(x_1^d)}{\partial x_k} + \sum_{j=1}^d \left(\frac{1}{8x_k} \frac{\partial^2 f(x_1^d)}{\partial x_j^2} - \frac{1}{8x_k^2} \frac{\partial f(x_1^d)}{\partial x_j} \right) + \frac{7}{48x_k^3} f(x_1^d), \\ V_2(x_1^d) &= \frac{7}{576x_k^4} f(x_1^d) + \sum_{j=1}^d \left(\frac{1}{16x_k^2} \frac{\partial^2 f(x_1^d)}{\partial x_j^2} - \frac{7}{96x_k^3} \frac{\partial f(x_1^d)}{\partial x_j} + \frac{1}{48x_k^2} \frac{\partial^2 f(x_1^d)}{\partial x_k \partial x_j} \right), \\ V_3(x_1^d) &= \frac{f(x_1^d)}{2x_k}, \\ V_4(x) &= \frac{f(x_1^d)}{4x_k^2} - \sum_{j=1}^d \frac{1}{4x_k} \frac{\partial f(x_1^d)}{\partial x_j}. \end{aligned}$$

Наконец, найдём ковариацию, определённую в (12), и её скорость сходимости.

Лемма 7. Пусть

- 1) $\{X_j\}_{j \geq 1} \in \mathcal{S}(\alpha)$ и $\int_1^\infty \alpha(\tau)^v d\tau < \infty$, $0 < v < 1$,
- 2) $b \rightarrow 0$ и $nb^{d(v+1)/2} \rightarrow \infty$ при $n \rightarrow \infty$.

Тогда ковариация оценки производной плотности $C(x)$ ограничена сверху:

$$\begin{aligned} & |\text{Cov}(\hat{f}'_{x_k}(x_1^d))| \leq \\ & \leq \frac{R(v, x_1^d)}{nb^{d(v+1)/2}} (b^2 V(v, x_1^d) + bW(v, x_1^d) + L(v, x_1^d))^{1-v} \int_1^\infty \alpha(\tau)^v d\tau + o(b^2), \quad (20) \end{aligned}$$

где были использованы следующие обозначения:

$$\begin{aligned} V(v, x_1^d) &= \sum_{i=1}^d \left(\left(\frac{(v+1)(3v-1)}{72(v-1)^3 x_k^2} + \frac{v+1}{(v-1)^2 x_i} - \frac{v(v+1)}{9(v-1)^4 x_k^3} \right) f(x_1^d) + \right. \\ & \quad \left. + \frac{v+1}{v-1} \frac{\partial f(x_1^d)}{\partial x_i} - \frac{v(v+1)}{9(v-1)^3 x_k^2} \frac{\partial f(x_1^d)}{\partial x_k} + \frac{x_i}{2} \frac{\partial^2 f(x_1^d)}{\partial x_i^2} \right), \\ W(v, x_1^d) &= \sum_{i=1}^d \left(\left(\frac{3v-1}{4(v-1)x_k} + \frac{v+1}{(v-1)^2 x_i} + \frac{2x_i(v+1)}{3(v-1)^3 x_k^2} \right) f(x_1^d) + \right. \\ & \quad \left. + \frac{v+1}{v-1} \frac{\partial f(x_1^d)}{\partial x_i} + \frac{2(v+1)x_i}{3(v-1)^2 x_k} \frac{\partial f(x_1^d)}{\partial x_k} + \frac{x_i}{2} \frac{\partial^2 f(x_1^d)}{\partial x_i^2} \right), \\ L(v, x_1^d) &= \sum_{i=1}^d \left(f(x_1^d) \frac{3v-1}{2(v-1)} + x_i \left(-\frac{4}{v-1} \frac{\partial f(x_1^d)}{\partial x_k} - \frac{4}{(v-1)^2 x_k} f(x_1^d) \right) \right), \\ R(v, x_1^d) &= \left(\prod_{j=1}^d x_j^{-(v+1)/2} \right) \frac{(2\pi)^{-(d(v-3)+2)/2}}{2x_k^2}. \end{aligned}$$

Используя верхнюю границу ковариации $|C(\hat{f}'_{x_k}(x_1^d))|$, получаем верхнюю границу MISE, а также находим оптимальный параметр размытости b .

Теорема 3. Если условия лемм 5–7 выполняются, то оптимальный параметр сглаживания, минимизирующий MISE оценки производной плотности по x_k , $1 \leq k \leq d$, имеет порядок сходимости

$$b_k^* = \left(\frac{d+2}{2^{d-1}\pi^{d/2}} \left(\int_0^\infty \frac{f(x_1^d)}{x_k} \prod_{j=1}^d x_j^{-1/2} dx_1^d / \int_0^\infty \left(\frac{f(x_1^d)}{3x_k^2} + \frac{1}{x_k} \sum_{i=1}^d x_i \frac{\partial^2 f(x_1^d)}{\partial x_i^2} \right)^2 dx_1^d \right)^{2/(d+6)} n^{-2/(d+6)} \right) \quad (21)$$

а оптимальный MISE имеет порядок сходимости $\text{MISE} \sim n^{-4/(d+6)}$.

Следствие 2. Результаты теоремы 3 согласуются с результатами [25] для производной одномерной плотности.

4. Применение

Параметры сглаживания (17) и (21), оптимальные для оценки плотности и её производной, зависят от неизвестной истинной плотности $f(x_1^d)$ и её производных.

В [37] было предложено использовать параметр сглаживания, полученный заменой неизвестной плотности $f(x_1^d)$ в оптимальной формуле на $g(x_1^d)$ — подставную плотность, параметры которой рассчитаны по известной выборке. Этот метод имеет название «метод большого пальца».

Другими словами, получаем оценку оптимального параметра размытости по выборке в виде

$$\hat{b} = C_1 R_1(\hat{\mu}, \hat{\sigma}^2) n^{-2/(d+4)}, \quad \hat{b}^* = C_2 R_2(\hat{\mu}, \hat{\sigma}^2) n^{-2/(d+6)},$$

где $R_1(\hat{\mu}, \hat{\sigma}^2)$ и $R_2(\hat{\mu}, \hat{\sigma}^2)$ являются функциями, зависящими от оценок среднего значения выборки и дисперсии.

Обычно для симметричных ядер в качестве такой плотности используют гауссовскую плотность. В этой работе использована асимметричная гамма-плотность.

4.1. Двумерная гамма-плотность

На практике часто приходится оценивать двумерные плотности распределения. Построим метод большого пальца для такой оценки.

Существует множество двумерных гамма-плотностей. Некоторые из них можно найти в [28, 29]. Двумерная гамма-плотность Маккея определяется следующим образом:

$$f(x, y) = \frac{x^{\alpha-1}(y-x)^{\beta-1} \exp(-y/\mu)}{\mu^{\alpha+\beta} \Gamma(\alpha) \Gamma(\beta)}, \quad x \leq y, \quad y > 0, \quad (22)$$

где $\alpha, \beta > 0$, $0 < \mu < \infty$ и $X \sim \Gamma(\alpha, \mu)$, $Y \sim \Gamma(\alpha + \beta, \mu)$. Возьмём её в качестве подставной плотности $g(x, y)$. Используя метод моментов, получим

$$\bar{m}_X = \alpha\mu, \quad D_X = \alpha\mu^2, \quad \bar{m}_Y = (\alpha + \beta)\mu, \quad D_Y = (\alpha + \beta)\mu^2,$$

где

$$\bar{m}_X = \frac{1}{n} \sum_{i=1}^n X_i, \quad \bar{m}_Y = \frac{1}{n} \sum_{i=1}^n Y_i, \\ \bar{D}_X = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{m}_X)^2, \quad \bar{D}_Y = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{m}_Y)^2.$$

Тогда параметры определяются по формулам

$$\hat{\mu} = \frac{\bar{D}_X}{\bar{m}_X}, \quad \hat{\alpha} = \frac{\bar{m}_X^2}{\bar{D}_X}, \quad \hat{\beta} = \frac{\bar{m}_Y \bar{m}_X}{\bar{D}_X} - \frac{\bar{m}_X^2}{\bar{D}_X}.$$

Параметр сглаживания (17) содержит вторую производную плотности, которую так же заменяем на подставную:

$$\frac{\partial^2 g(x, y)}{\partial x^2} = \frac{x^{\alpha-3} e^{-y/\mu} (y-x)^{\beta-3}}{\Gamma(\alpha) \Gamma(\beta) \mu^{\alpha+\beta}} (x^2(\alpha + \beta - 2)(\alpha + \beta - 3) + \\ + y^2(\alpha - 1)(\alpha - 2) - 2yx(\alpha - 1)(\alpha + \beta - 3)), \\ \frac{\partial^2 g(x, y)}{\partial y^2} = \frac{x^{\alpha-1} e^{-y/\mu} (y-x)^{\beta-3}}{\Gamma(\alpha) \Gamma(\beta) \mu^{\alpha+\beta+2}} (x^2 + y^2 + 2x(\mu\beta - y - \mu) + \\ + 2y\mu(1 - \beta) + \mu^2(\beta^2 - 3\beta + 2)).$$

Следовательно, можно взять оценку

$$\hat{b} = \left(2 \int_0^\infty \int_0^\infty \left(\frac{(x_1 x_2)^{-1/2}}{4\pi} \right) \hat{g}(x_1, x_2) dx_1 dx_2 / \right. \\ \left. / \int_0^\infty \int_0^\infty \left(x_1 \frac{\partial^2 \hat{g}(x_1, x_2)}{\partial x_1^2} + x_2 \frac{\partial^2 \hat{g}(x_1, x_2)}{\partial x_2^2} \right)^2 dx_1 dx_2 \right)^{2/7} n^{-2/7} \quad (23)$$

как оптимальный параметр размытости (17) и подставить его в двумерную гамма-ядерную оценку:

$$\hat{f}(X_1, X_2) = \frac{1}{n} \sum_{i=1}^n \frac{X_{1i}^{\hat{b}-1} \exp(-X_{1i}/\hat{b})}{\hat{b}^{X_{1i}/\hat{b}} \Gamma(X_{1i}/\hat{b})} \frac{X_{2i}^{\hat{b}-1} \exp(-X_{2i}/\hat{b})}{\hat{b}^{X_{2i}/\hat{b}} \Gamma(X_{2i}/\hat{b})}. \quad (24)$$

4.2. Многомерная гамма-плотность

Пусть $Y = (Y_0, Y_1, \dots, Y_n)^T$ — $(n+1)$ -мерный вектор независимых случайных величин, $Y_i \sim \Gamma(x_i/b_i, 1/b_i)$, $i = 0, 1, \dots, n$. Пусть задана матрица

$$A = \begin{pmatrix} \alpha_0/\alpha_1 & 1 & 0 & 0 & \dots & 0 \\ \alpha_0/\alpha_2 & \alpha_1/\alpha_2 & 1 & 0 & \dots & 0 \\ \alpha_0/\alpha_3 & \alpha_1/\alpha_3 & \alpha_2/\alpha_3 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_0/\alpha_n & \alpha_1/\alpha_n & \alpha_2/\alpha_n & \alpha_3/\alpha_n & \dots & 1 \end{pmatrix},$$

где $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)^T = (1/b_1, 1/b_2, \dots, 1/b_n)$. На практике часто полагают $b_i \equiv b$. Тогда совместное распределение случайных величин $X = AY$ — многомерное гамма-распределение лестничного типа, предложенное в [17]. Возьмём

$$g(x_1^d) = \frac{x_1^{x_1/b} e^{-x_d/b}}{\Gamma(x_1/b + 1)} \prod_{i=2}^d \frac{(x_i - x_{i-1})^{x_i/b-1}}{b^{\bar{\gamma}_i} \Gamma(x_i/b)}$$

в качестве подставной функции. Здесь n -мерный вектор

$$\bar{\gamma} = \left(1 + \frac{x_1}{b}, 1 + \frac{x_1}{b} + \frac{x_2}{b}, \dots, 1 + \frac{x_1}{b} + \frac{x_2}{b} + \dots + \frac{x_d}{b}\right)$$

задаёт параметр формы. Неизвестные параметры так же, как и в двумерном случае, определим по методу моментов.

Построенный таким образом метод большого пальца позволяет производить оценку оптимальных параметров сглаживания для оценок многомерной плотности и её производной по зависимым данным.

4.2.1. Результаты моделирования

В этом разделе изучены свойства полученных оценок для конечных выборок двумерных данных с плотностью, определённой на неотрицательном носителе. С этой целью мы используем многомерную гамма-плотность, введённую в [26]. Предположим, что V_1, \dots, V_k , где $V_i \sim \Gamma(\alpha_i, \beta, \gamma_i)$, $i = 1, \dots, k$, взаимно независимы. Пусть

$$Z_1 = V_1, Z_2 = V_1 + V_2, \dots, Z_k = V_1 + \dots + V_k. \quad (25)$$

Тогда совместная плотность распределения $Z = (Z_1, \dots, Z_k)$ — многомерная гамма-плотность (см. [26, теорема 1.1]). Можно переписать (25) как

$$Z_t = Z_{t-1} + V_t, \quad t = 1, \dots, k,$$

что есть не что иное, как $AR(1)$ -процесс. Согласно [2] он удовлетворяет условию сильного перемешивания.

Сгенерируем две одномерные одинаково распределённые последовательности $V_i \sim \Gamma(\alpha_i, \mu, 0)$, $i = 1, 2$, длины $n \in \{100, 500, 1000, 2000\}$, используя

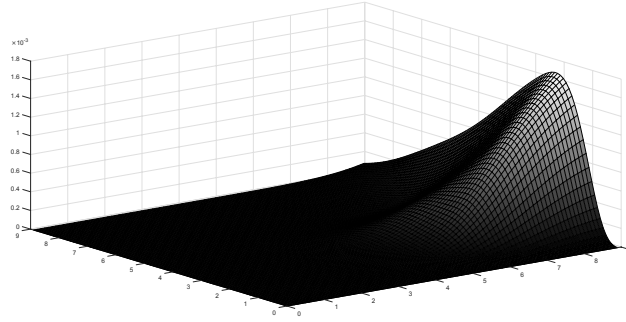
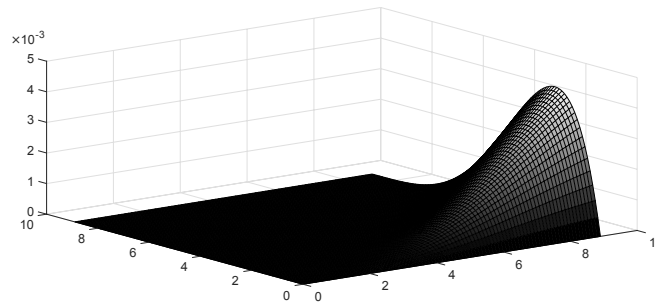
Рис. 1. Оценка плотности Маккея для выборки длиной $n = 1000$ 

Рис. 2. Плотность Маккея

стандартные генераторы MATLAB. Тогда две случайные величины $Z_1 = V_1$ и $Z_2 = Z_1 + V_2$ зависимы, а их совместная плотность — плотность Маккея.

Гамма-ядерная оценка (24) с оптимальным параметром сглаживания показана на рис. 1 для $\alpha = 3$, $\beta = 5$, $\mu = 3$. Оптимальный параметр сглаживания считается для каждого повторения симуляции с использованием метода большого пальца, где (22) берётся в качестве подставной плотности. Ошибка оценки плотности вычисляется по следующей формуле:

$$m = \int_0^{\infty} (f(x, y) - \hat{f}(x, y))^2 dx dy,$$

где $f(x, y)$ — истинная плотность, а $\hat{f}(x, y)$ — её оценка. Значение m усреднено по 500 симуляциям, и стандартное отклонение для искомой плотности приведено в таблице 1.

Таблица 1. Средние ошибки и стандартные отклонения в скобках

n	100	500	1000	2000
Маккей	$1,7925 \cdot 10^{-4}$ ($8,1002 \cdot 10^{-5}$)	$7,3154 \cdot 10^{-5}$ ($3,8101 \cdot 10^{-5}$)	$6,3605 \cdot 10^{-5}$ ($1,8349 \cdot 10^{-5}$)	$2,7118 \cdot 10^{-5}$ ($1,2023 \cdot 10^{-5}$)

Как и ожидалось, средняя ошибка и стандартное отклонение уменьшаются при увеличении длины выборки.

4.3. Практическое использование гамма-ядерной оценки в оптимальной фильтрации

Результаты, полученные в работе, имеют практическое применение в обработке сигналов и управлении линейными и нелинейными системами. Обработка стационарных случайных последовательностей при непараметрической неопределённости определяется проблемой фильтрации, когда распределение сигнала неизвестно. Полезный сигнал $(S_n)_{n \geq 1}$ считается марковским. Это предположение позволяет оценить неизвестный сигнал (S_n) , используя только наблюдаемую случайную последовательность $(X_n)_{n \geq 1}$. В [24] уравнение оптимальной фильтрации для экспоненциального семейства условных плотностей даётся в следующем виде:

$$E(Q(S_n)|x_1^n) \cdot T'_{x_n}(x_n) = \frac{f'_{x_n}(x_n|x_1^{n-1})}{f(x_n|x_1^{n-1})} - \frac{h'_{x_n}(x_n)}{h(x_n)} = \ln \left(\frac{f(x_n|x_1^{n-1})}{h(x_n)} \right)'_{x_n}. \quad (26)$$

Это уравнение содержит логарифмическую производную неизвестной условной плотности, которая характеризует сигнал. Последняя плотность и её производная могут быть записаны как

$$f(x_n|x_1^{n-1}) = \frac{f(x_1^n)}{f(x_1^{n-1})}, \quad f'_{x_n}(x_n|x_1^{n-1}) = \frac{f'_{x_n}(x_1^n)}{f(x_1^{n-1})}.$$

Следовательно, их отношение определяется как

$$\frac{f'_{x_n}(x_n|x_1^{n-1})}{f(x_n|x_1^{n-1})} = \frac{f'_{x_n}(x_1^n)}{f(x_1^n)}.$$

Поэтому, используя результаты предыдущих разделов, можно записать

$$\frac{\hat{f}'_{x_n}(x_1^n)}{\hat{f}(x_1^n)} = \begin{cases} \left(\frac{(1/b'_n) \sum_{i=1}^n L_1(X_{in}, x_n, b'_n) \prod_{j=1}^n K_{\rho_1(x_j, b'_j), b'_j}(X_{ij})}{\sum_{i=1}^n \prod_{j=1}^n K_{\rho_1(x_j, b_j), b_j}(X_{ij})} \right) / & \text{при } x_1^n \geq 2b, \\ \left(\frac{(x_n/(2(b'_n)^2)) \sum_{i=1}^n L_2(X_{in}, x_n, b'_n) \prod_{j=1}^n K_{\rho_2(x_j, b'_j), b'_j}(X_{ij})}{\sum_{i=1}^n \prod_{j=1}^n K_{\rho_2(x_j, b_j), b_j}(X_{ij})} \right) / & \text{при } x_1^n \in [0, 2b). \end{cases}$$

В случае когда параметры сглаживания являются оптимальными, последняя оценка имеет следующий вид:

$$\frac{\hat{f}'_{x_n}(x_1^n)}{\hat{f}(x_1^n)} = \begin{cases} \left((1/b') \sum_{i=1}^n L_1(X_{in}, x_n, b') \prod_{j=1}^n K_{\rho_1(x_j, b'), b'}(X_{ij}) \right) / \\ \left(\sum_{i=1}^n \prod_{j=1}^n K_{\rho_1(x_j, b), b}(X_{ij}) \right) & \text{при } x_1^n \geq 2b, \\ \left((x_n/(2b'^2)) \sum_{i=1}^n L_2(X_{in}, x_n, b') \prod_{j=1}^n K_{\rho_2(x_j, b'), b'}(X_{ij}) \right) / \\ \left(\sum_{i=1}^n \prod_{j=1}^n K_{\rho_2(x_j, b), b}(X_{ij}) \right) & \text{при } x_1^n \in [0, 2b). \end{cases}$$

Если подставить последнюю оценку в уравнение оптимальной фильтрации (26), можно оценить условное среднее $E(Q(S_n)|x_1^n)$, зная только наблюдаемую часть сигнала.

5. Заключение

Предложены непараметрические оценки плотности вероятности и её частных производных на положительной полуоси с помощью многомерных зависимых данных. Оценки основаны на мультипликативных гамма-ядрах с неотрицательным носителем. Исследованы асимптотические свойства оценок, т. е. оптимальные скорости сходимости их среднеквадратичных ошибок. Выведены явные формулы для оптимальных параметров сглаживания как для плотности, так и для частных производных по выборкам зависимых случайных величин. Кроме того, предложенные гамма-ядерные оценки имеют ту же скорость сходимости, что и гауссовские ядерные оценки для симметричного случая. Поскольку оптимальные параметры сглаживания зависят от неизвестной плотности, необходимо построить их оценку по зависимым данным. Для этого предлагается известный метод большого пальца. Приведены результаты моделирования на конечных выборках. Дальнейшее развитие может касаться исследования альтернативных методов выбора параметра сглаживания. Результаты также могут быть расширены для выборок с другими условиями перемешивания.

А. Доказательства

А.1. Доказательство леммы 1

Чтобы найти смещение оценки $\hat{f}(x_1^d)$, запишем математическое ожидание:

$$E_X(\hat{f}(x_1^d)) = \int f(t_1^d) \prod_{j=1}^d K_{\rho(x_j, b_j)}(t_j) dt_1^d = E_\xi(f(\xi_1^d)), \quad (27)$$

где случайные величины ξ_j — независимые одинаково гамма-распределённые случайные величины с математическим ожиданием $\mu_j = x_j$ и дисперсией $\sigma_j^2 = x_j b_j$. Мультипликативное ядро $\prod_{j=1}^d K_{\rho(x_j, b_j)}(t_j)$ использовано в (27) в качестве плотности. Чтобы найти $E_\xi(f(\xi_1^d))$, воспользуемся разложением по Тейлору $f(\xi_1^d)$ в точке μ_j :

$$f(\xi_1^d) = f(\mu_1^d) + \sum_{j=1}^d (\xi_j - \mu_j) \frac{\partial f(\xi_1^d)}{\partial x_j} \Big|_{\xi=\mu} + \frac{1}{2} \sum_{j=1}^d (\xi_j - \mu_j)^2 \frac{\partial^2 f(\xi_1^d)}{\partial x_j^2} \Big|_{\xi=\mu} + \\ + \sum_{j \neq l} (\xi_j - \mu_j)(\xi_l - \mu_l) \frac{\partial^2 f(\xi_1^d)}{\partial x_j \partial x_l} \Big|_{\xi=\mu} + o\left(\sum_{j=1}^d b_j\right).$$

Взяв математическое ожидание обеих сторон последнего уравнения, можно написать

$$E_\xi(f(\xi_1^d)) = f(x_1^d) + \frac{1}{2} \sum_{j=1}^d x_j b_j \frac{\partial^2 f(x_1^d)}{\partial x_j^2} + o\left(\sum_{j=1}^d b_j\right). \quad (28)$$

Следовательно, смещение многомерной оценки плотности задаётся формулой

$$\text{Bias}(\hat{f}(x_1^d)) = \frac{1}{2} \sum_{j=1}^d x_j b_j \frac{\partial^2 f(x_1^d)}{\partial x_j^2} + o\left(\sum_{j=1}^d b_j\right).$$

Когда $b_1 = b_2 = \dots = b_d = b$, получаем (13).

А.2. Доказательство леммы 2

По определению дисперсия оценки $\hat{f}(x_1^d)$ задаётся

$$\text{Var}(\hat{f}(x_1^d)) = \frac{1}{n} \left(E \left(\left(\prod_{j=1}^d K_{\rho(x_j, b_j)}(X_j) \right)^2 \right) - E^2 \left(\prod_{j=1}^d K_{\rho(x_j, b_j)}(X_j) \right) \right). \quad (29)$$

С использованием математического ожидания (28) второй член в (29) может быть записан как

$$\frac{1}{n} E_X^2(\hat{f}(x_1^d)) = \frac{1}{n} \left(f(x_1^d) + \frac{1}{2} \sum_{j=1}^d x_j b_j \frac{\partial^2 f(x_1^d)}{\partial x_j^2} + o\left(\sum_{j=1}^d b_j\right) \right)^2. \quad (30)$$

Первый член в (29) имеет вид

$$\frac{1}{n} E \left(\left(\prod_{j=1}^d K_{\rho(x_j, b_j)}(X_j) \right)^2 \right) = \frac{1}{n} \int \left(\prod_{j=1}^d K_{\rho(x_j, b_j)}^2(X_j) \right) f(t_1^d) dt_1^d = \\ = \frac{1}{n} \int \left(\prod_{j=1}^d \frac{t_j^{2x_j/b_j - 2} \exp(-2t_j/b_j)}{b_j^{2x_j/b_j} \Gamma^2(x_j/b_j)} \right) f(t_1^d) dt_1^d = \frac{1}{n} \left(\prod_{j=1}^d B(x_j, b_j) \right) E_\eta(f(\eta_1^d)), \quad (31)$$

где $\{\eta_j\}$ — независимые одинаково гамма-распределённые случайные величины с математическим ожиданием $\mu_j = x_j - b_j/2$ и дисперсией $\sigma_j^2 = x_j b_j/2 - b_j^2/4$. Обозначим

$$B(x_j, b_j) = \frac{b_j^{-3} x_j^2 \Gamma(2x_j/b_j - 1)}{2^{2x_j/b_j - 1} \Gamma^2(x_j/b_j + 1)}. \quad (32)$$

Воспользуемся обозначениями [9], где $R(z) = \sqrt{2\pi} \exp(-z) z^{z+1/2} / \Gamma(z+1)$ для $z \geq 0$. Следовательно, можно выразить гамма-функции в (32) как

$$\begin{aligned} \Gamma^2\left(\frac{x}{b} + 1\right) &= \left(\frac{\sqrt{2\pi} \exp(-x/b) (x/b)^{x/b+1/2}}{R(x/b)}\right)^2, \\ \Gamma\left(\frac{2x}{b} - 1\right) &= \frac{\Gamma(2x/b + 1)}{(2x/b)(2x/b - 1)} = \frac{\sqrt{2\pi} \exp(-2x/b) (2x/b)^{2x/b+1/2}}{(2x/b)(2x/b - 1)R(2x/b)}. \end{aligned}$$

Подставляя последние выражения в выражение (32), можно переписать его как

$$B(x_j, b_j) = \frac{b_j^{-1/2} x_j^{-1/2} R^2(x_j/b_j)}{2\sqrt{\pi} R(2x_j/b_j) (1 - b_j/(2x_j))}.$$

Согласно [9, лемма 3], $R(z)$ — возрастающая функция, которая сходится к 1 при $z \rightarrow \infty$, и $R(z) < 1$ для любых $z > 0$. Тогда верно, что

$$B(x_j, b_j) = \begin{cases} (b_j^{-1/2} x_j^{-1/2}) / (2\sqrt{\pi}), & \text{если } x_j/b_j \rightarrow \infty, \\ (b^{-1} k^2 \Gamma(2k - 1)) / (2^{2k-1} \Gamma^2(k + 1)), & \text{если } x_j/b_j \rightarrow k. \end{cases} \quad (33)$$

Математическое ожидание в (31) может быть записано аналогично (28) как

$$E_\eta(f(\eta_1^d)) = f(\mu_1^d) + \frac{1}{2} \sum_{j=1}^d \left(\frac{x_j b_j}{2} - \frac{b_j^2}{4} \right) \frac{\partial^2 f(\mu_1^d)}{\partial x_j^2} + o\left(\sum_{j=1}^d b_j^2\right).$$

Поэтому, применяя разложение по Тейлору аргументов функций $f(\mu_1^d)$

$$f(\mu_1^d) = f(x_1^d) - \sum_{j=1}^d \frac{b_j}{2} \frac{\partial f(x_1^d)}{\partial x_j} + \sum_{j=1}^d \frac{b_j^2}{8} \frac{\partial^2 f(x_1^d)}{\partial x_j^2} + o\left(\sum_{j=1}^d b_j^2\right),$$

получаем, что (31) может быть переписано как

$$\begin{aligned} \frac{1}{n} \left(\prod_{j=1}^d B(x_j, b_j) \right) &\left(f(x_1^d) - \sum_{j=1}^d \frac{b_j}{2} \frac{\partial f(x_1^d)}{\partial x_j} + \frac{b_j^2}{8} \frac{\partial^2 f(x_1^d)}{\partial x_j^2} + \right. \\ &\left. + \frac{1}{2} \sum_{j=1}^d \left(\frac{x_j b_j}{2} - \frac{b_j^2}{4} \right) \left(\frac{\partial^2 f(x_1^d)}{\partial x_j^2} - \sum_{i=1}^d \frac{b_i}{2} \frac{\partial^3 f(x_1^d)}{\partial x_j^2 \partial x_i} \right) \right) + o\left(\sum_{j=1}^d b_j^2\right). \end{aligned}$$

Объединив (30) и (33), можно записать дисперсию оценки плотности как

$$\begin{aligned} \text{Var}(\hat{f}(x_1^d)) &= \frac{1}{n} \left(\prod_{j=1}^d \frac{b_j^{-1/2} x_j^{-1/2}}{2\sqrt{\pi}} \right) \left(f(x_1^d) - \sum_{j=1}^d \frac{b_j}{2} \frac{\partial f(x_1^d)}{\partial x_j} + \right. \\ &+ \sum_{j=1}^d \frac{b_j^2}{8} \frac{\partial^2 f(x_1^d)}{\partial x_j^2} + \frac{1}{2} \left(\frac{x_j b_j}{2} - \frac{b_j^2}{4} \right) \left(\frac{\partial^2 f(x_1^d)}{\partial x_j^2} - \sum_{i=1}^d \frac{b_i}{2} \frac{\partial^3 f(x_1^d)}{\partial x_j^2 \partial x_i} \right) \left. \right) - \\ &- \frac{1}{n} \left(f(x_1^d) + \frac{1}{2} \sum_{j=1}^d x_j b_j \frac{\partial^2 f(x_1^d)}{\partial x_j^2} \right)^2 + o\left(\sum_{j=1}^d b_j^2 \right). \end{aligned}$$

В случае когда $b_1 = b_2 = \dots = b_d = b$, дисперсия имеет вид (14).

А.3. Доказательство леммы 3

Для оценки ковариации $\hat{f}(x_1^d)$ используем неравенство Давыдова

$$|\text{Cov}(\tilde{K}(X_1, x, b), \tilde{K}(X_{1+i}, x, b))| \leq 2\pi\alpha(i)^{1/r} \|\tilde{K}(X_1, x, b)\|_q \|\tilde{K}(X_{1+i}, x, b)\|_p,$$

где $p^{-1} + q^{-1} + r^{-1} = 1$, $1 \leq p, q, r \leq \infty$ (см. [4]), $\alpha(i)$ — коэффициент перемешивания (1). Норма \mathcal{L}_p по определению равна

$$\begin{aligned} \|\tilde{K}(X_1, x, b)\|_q &= \left(\int (\tilde{K}(y, x, b))^q f(y) dy \right)^{1/q} = \\ &= \left(\int \left(\prod_{j=1}^d K_{\rho(x_j, b_j)}(t_j) \right)^q f(t_1^d) dt_1^d \right)^{1/q} = \\ &= \left(\mathbb{E} \left(f(\xi_1^d) \prod_{j=1}^d K_{\rho(x_j, b_j)}^{q-1}(\xi_j) \right) \right)^{1/q}, \end{aligned} \quad (34)$$

где мультипликативное ядро $\prod_{j=1}^d K_{\rho(x_j, b_j)}(\xi_j)$ использовано в качестве плотности. $\{\xi_j\}$ — независимые одинаково гамма-распределённые случайные величины с математическим ожиданием $\mu_j = x_j$ и дисперсией $\text{Var} = x_j b_j$. Аналогично предыдущим доказательствам используем разложение по Тейлору, чтобы найти математическое ожидание (34):

$$\begin{aligned} \mathbb{E} \left(f(\xi_1^d) \prod_{j=1}^d K_{\rho(x_j, b_j)}^{q-1}(\xi_j) \right) &= f(\mu_1^d) \prod_{j=1}^d K_{\rho(x_j, b_j)}^{q-1}(\mu_j) + \\ &+ \sum_{i=1}^d \frac{\text{Var}}{2} \frac{\partial^2}{\partial \xi_i^2} \left(f(\xi_1^d) \prod_{j=1}^d K_{\rho(x_j, b_j)}^{q-1}(\xi_j) \right) \Big|_{\xi=\mu} + o\left(\sum_{j=1}^d b_j^2 \right) = \end{aligned}$$

$$\begin{aligned}
&= \left(\prod_{j=1}^d K_{\rho(x_j, b_j)}^{q-1}(x_j) \right) \left(f(x_1^d) \frac{(3-q)}{2} + \sum_{i=1}^d \frac{q(q-1)b_i}{2x_i} f(x_1^d) - \right. \\
&\quad \left. - (q-1)b_i \frac{\partial f(x_1^d)}{\partial x_i} + \frac{b_i x_i}{2} \frac{\partial^2 f(x_1^d)}{\partial x_i^2} \right) + o\left(\sum_{j=1}^d b_j^2 \right). \quad (35)
\end{aligned}$$

Используя формулу Стирлинга

$$\Gamma(z) = \sqrt{\frac{2\pi}{z}} \left(\frac{z}{e} \right)^z \left(1 + O\left(\frac{1}{z} \right) \right),$$

перепишем ядерную функцию как

$$\prod_{j=1}^d K_{\rho(x_j, b_j)}(x_j) = \prod_{j=1}^d \frac{x_j^{x_j/b_j - 1} \exp(-x_j/b_j)}{b_j^{x_j/b_j} \Gamma(x_j/b_j)} = \prod_{j=1}^d \frac{x_j^{-1/2} b_j^{-1/2}}{\sqrt{2\pi} (1 + O(b_j/x_j))}.$$

Её верхняя граница задаётся выражением

$$\prod_{j=1}^d K_{\rho(x_j, b_j)}^{q-1}(x_j) \leq \prod_{j=1}^d \frac{1}{(2\pi x_j b_j)^{(q-1)/2}}. \quad (36)$$

Следовательно, подставляя (36) в (35), можно переписать (34) как

$$\begin{aligned}
\mathbb{E} \left(f(\xi_1^d) \prod_{j=1}^d K_{\rho(x_j, b_j)}^{q-1}(\xi_j) \right) &= \left(\prod_{j=1}^d (2\pi x_j b_j)^{(1-q)/2} \right) \times \\
&\times \left(f(x_1^d) \frac{(3-q)}{2} + \sum_{i=1}^d \frac{q(q-1)b_i}{2x_i} f(x_1^d) - (q-1)b_i \frac{\partial f(x_1^d)}{\partial x_i} + \frac{b_i x_i}{2} \frac{\partial^2 f(x_1^d)}{\partial x_i^2} \right) + \\
&+ o\left(\sum_{j=1}^d b_j^2 \right).
\end{aligned}$$

В случае когда $b_1 = b_2 = \dots = b_d = b$, получаем

$$\begin{aligned}
\|\tilde{K}(X_1, x, b)\|_q &= b^{(\tau+1)(1-q)/2q} \left(\prod_{j=1}^d (2\pi x_j)^{(1-q)/(2q)} \right) \times \\
&\times \left(f(x_1^d) \frac{(3-q)}{2} + b \sum_{i=1}^d \frac{q(q-1)}{2x_i} f(x_1^d) - (q-1) \frac{\partial f(x_1^d)}{\partial x_i} + \frac{x_i}{2} \frac{\partial^2 f(x_1^d)}{\partial x_i^2} \right)^{1/q} + o(b^2).
\end{aligned}$$

Пусть $p = q$. Подставляя последнее выражение в неравенство Давыдова, получаем следующее выражение:

$$|\text{Cov}(\tilde{K}(X_1, x, b), \tilde{K}(X_{1+k}, x, b))| \leq 2\pi\alpha(k)^{1/r} \left(b^{d(1-q)/q} \left(\prod_{j=1}^d (2\pi x_j)^{(1-q)/q} \right) \times \right. \\ \left. \times \left(f(x_1^d) \frac{(3-q)}{2} + b \sum_{i=1}^d (q-1) \left(\frac{q}{2x_i} f(x_1^d) - \frac{\partial f(x_1^d)}{\partial x_i} \right) + \frac{x_i}{2} \frac{\partial^2 f(x_1^d)}{\partial x_i^2} \right)^{2/q} \right) + o(b^2).$$

Принимая $p = q = 2 + \delta$, $r = (2 + \delta)/\delta$, можно показать, что ковариация оценки многомерной плотности задаётся как

$$\text{Cov}(\hat{f}(x_1^d)) = \frac{2}{n} \sum_{k=1}^n \left(1 - \frac{k}{n} \right) \text{Cov}(\tilde{K}(X_1, x, b), \tilde{K}_1(X_{1+k}, x, b)) \leq \\ \leq \frac{2b^{-d(\delta+1)/(\delta+2)}}{n} (2\pi)^{1-d(\delta+1)/(\delta+2)} \times \\ \times \left(\prod_{j=1}^d x_j^{-(\delta+1)/(\delta+2)} \right) \sum_{k=1}^n \left(1 - \frac{k}{n} \right) \alpha(k)^{\delta/(2+\delta)} \times \\ \times \left(f(x_1^d) \frac{1-\delta}{2} + b \sum_{i=1}^d \frac{(\delta+1)(\delta+2)}{2x_i} f(x_1^d) - (\delta+1) \frac{\partial f(x_1^d)}{\partial x_i} + \frac{x_i}{2} \frac{\partial^2 f(x_1^d)}{\partial x_i^2} \right)^{2/q} + \\ + o(b^2).$$

Введём следующие обозначения:

$$S(\delta, x_1^d) = \sum_{i=1}^d \frac{(\delta+1)(\delta+2)}{2x_i} f(x_1^d) - (\delta+1) \frac{\partial f(x_1^d)}{\partial x_i} + \frac{x_i}{2} \frac{\partial^2 f(x_1^d)}{\partial x_i^2}, \quad (37)$$

$$D(\delta, x_1^d) = 2(2\pi)^{1-d(\delta+1)/(\delta+2)} \left(\prod_{j=1}^d x_j^{-(\delta+1)/(\delta+2)} \right). \quad (38)$$

В этих обозначениях ковариация может быть ограничена следующим выражением:

$$\text{Cov}(\hat{f}(x_1^d)) \leq \frac{D(\delta, x_1^d)}{n} b^{-d(\delta+1)/(\delta+2)} \times \\ \times \left(f(x_1^d) \frac{1-\delta}{2} + bS(\delta, x_1^d) \right)^{2/(2+\delta)} \int_1^\infty \alpha(\tau)^{\delta/(2+\delta)} d\tau + o(b^2).$$

Обозначим $\delta/(2 + \delta) = v$, $0 < v < 1$. Тогда верхняя граница ковариации имеет вид (26).

А.4. Доказательство теоремы 1

Объединив смещение, дисперсию и ковариацию из лемм 1–3, можно записать среднеквадратическую ошибку оценки плотности как

$$\begin{aligned}
\text{MSE}(\hat{f}(x_1^d)) &= \text{Bias}(\hat{f}(x_1^d))^2 + \text{Var}(\hat{f}(x_1^d)) + \text{Cov}(\hat{f}(x_1^d)) = \\
&= \left(\frac{b}{2} \sum_{j=1}^d x_j \frac{\partial^2 f(x_1^d)}{\partial x_j^2} \right)^2 - \frac{1}{n} \left(f(x_1^d) + \frac{b}{2} \sum_{j=1}^d x_j \frac{\partial^2 f(x_1^d)}{\partial x_j^2} \right)^2 + \\
&+ \frac{1}{n} \left(\prod_{j=1}^d \frac{b^{-d/2} x_j^{-1/2}}{2\sqrt{\pi}} \right) (f(x_1^d) + bv_1(x_1^d) + b^2 v_2(x_1^d)) + \\
&+ \frac{D(v, x_1^d)}{n} b^{-d(v+1)/2} \left(bS(v, x_1^d) + f(x_1^d) \frac{3v-1}{2(v-1)} \right)^{1-v} \int_1^\infty \alpha(\tau)^v d\tau + o(b^2).
\end{aligned}$$

Так как скорость сходимости по b ковариации больше, чем дисперсии для $b \rightarrow 0$ при $n \rightarrow \infty$, т. е.

$$b^{-d/2} < b^{-d(v+1)/2},$$

мы можем пренебречь дисперсионным членом. Беря интеграл от MSE и минимизируя приведённую MISE по b , получаем оптимальный параметр сглаживания (15).

А.5. Доказательство леммы 4

Представим оценку сверху ковариации оценки плотности в виде двух сумм:

$$\begin{aligned}
\text{Cov}(\hat{f}(x_1^d)) &\leq \frac{2}{n} \sum_{\tau=2}^{c(n)} \text{Cov}(\tilde{K}(X_1, x, b), \tilde{K}(X_\tau, x, b)) + \\
&+ \frac{2}{n} \sum_{\tau=c(n)+1}^\infty \text{Cov}(\tilde{K}(X_1, x, b), \tilde{K}(X_\tau, x, b)) = I_1 + I_2. \quad (39)
\end{aligned}$$

Используя лемму 7, получаем, что вторая сумма в (39) может быть ограничена сверху:

$$\begin{aligned}
I_2 &\leq \frac{D(\delta, x_1^d)}{n} b^{-d(\delta+1)/(\delta+2)} \times \\
&\times \left(f(x_1^d) \frac{1-\delta}{2} + bS(\delta, x_1^d) \right)^{2/(2+\delta)} \sum_{\tau=c(n)}^\infty \alpha(\tau)^{\delta/(2+\delta)} + o(b^2),
\end{aligned}$$

где были использованы обозначения (37) и (38). Более того, можно переписать последнюю формулу как

$$\begin{aligned}
I_2 &\leq \frac{D(\delta, x_1^d)}{c(n)n} b^{-d(\delta+1)(\delta+2)} \times \\
&\times \left(f(x_1^d) \frac{1-\delta}{2} + bS(\delta, x_1^d) \right)^{2/(2+\delta)} \sum_{\tau=2}^\infty \tau \alpha(\tau)^{\delta/(2+\delta)} + o(b^2).
\end{aligned}$$

Обозначим $\delta = 4\kappa/(1 - 2\kappa)$, $0 < \kappa < 1/2$. В этих обозначениях

$$I_2 \leq \frac{D(\kappa, x_1^d)}{c(n)n} b^{-d(2\kappa+1)/2} \left(f(x_1^d) \frac{6\kappa - 1}{2(2\kappa - 1)} + bS(\kappa, x_1^d) \right)^{1-2\kappa} \sum_{\tau=2}^{\infty} \tau \alpha(\tau)^{2\kappa} + o(b^2).$$

Первая сумма (39) имеет вид

$$\begin{aligned} I_1 &= \frac{2}{n} \sum_{\tau=2}^{c(n)} |\text{Cov}(\tilde{K}(X_1, x, b), \tilde{K}(X_\tau, x, b))| = \\ &= \frac{2}{n} \sum_{\tau=2}^{c(n)} |\mathbb{E}(\tilde{K}(X_1, x, b) \cdot \tilde{K}(X_\tau, x, b)) - \mathbb{E}(\tilde{K}(X_1, x, b)) \cdot \mathbb{E}(\tilde{K}(X_\tau, x, b))| = \\ &= \frac{2}{n} \sum_{\tau=2}^{c(n)} \left| \int_0^\infty \int_0^\infty \tilde{K}(u, x, b) \tilde{K}(v, x, b) f_\tau(u, v) du dv - \right. \\ &\quad \left. - \int_0^\infty \tilde{K}(u, x, b) f(u) du \int_0^\infty \tilde{K}(v, x, b) f(v) dv \right| = \\ &= \frac{2}{n} \sum_{\tau=2}^{c(n)} \left| \int_0^\infty \int_0^\infty \tilde{K}(u, x, b) \tilde{K}(v, x, b) (f_\tau(u, v) - f(u)f(v)) du dv \right| \leq \\ &\leq \frac{2}{n} \sum_{\tau=2}^{c(n)} \int_0^\infty \int_0^\infty |\tilde{K}(u, x, b) \tilde{K}(v, x, b)| |f_\tau(u, v) - f(u)f(v)| du dv. \end{aligned}$$

При условии леммы $|f_\tau(x, y) - f(x)f(y)| \leq M$ получим

$$I_1 \leq \frac{2M}{n} \sum_{\tau=2}^{c(n)} \left[\int_0^\infty \left| \prod_{j=1}^d K_{\rho(x_j, b_j), b_j}(u) \right| du \right]^2 = \frac{2Mc(n)}{n}.$$

Мы стремимся сделать скорость сходимости I_1 по b больше, чем скорость I_2 . Тогда должно выполняться условие

$$c(n) < b^{-d(2\kappa+1)/4}.$$

Выберем, например, $c(n) = b^{-d/8}$, $\kappa = 1/4$. Тогда скорости сходимости I_1 и I_2 задаются выражением (16).

А.6. Доказательство теоремы 2

Используя смещение, дисперсию и ковариацию из лемм 1, 2 и 4, напомним среднеквадратическую ошибку оценки плотности $\hat{f}(x_1^d)$ и найдём оптимальный параметр сглаживания b , который минимизирует MISE. MSE представим как

$$\begin{aligned} \text{MSE}(\hat{f}(x_1^d)) &= \left(\frac{b}{2} \sum_{j=1}^d x_j \frac{\partial^2 f(x_1^d)}{\partial x_j^2} \right)^2 + \frac{1}{n} \left(b^{-d/2} \prod_{j=1}^d \frac{x_j^{-1/2}}{2\sqrt{\pi}} \right) \times \\ &\times (f(x_1^d) + bv_1(x_1^d) + b^2v_2(x_1^d)) - \frac{1}{n} \left(f(x_1^d) + \frac{b}{2} \sum_{j=1}^d x_j \frac{\partial^2 f(x_1^d)}{\partial x_j^2} \right)^2 + \\ &+ \frac{2M}{nb^{d/8}} + \frac{D(\kappa, x_1^d)}{nb^{d/16}} \left(f(x_1^d) \frac{6\kappa - 1}{2(2\kappa - 1)} + bS(\kappa, x_1^d) \right)^{1-2\kappa} \sum_{\tau=2}^{\infty} \tau \alpha(\tau)^{2\kappa} + o(b^2). \end{aligned}$$

Так как скорость сходимости дисперсии по b больше, чем скорость сходимости ковариации, мы можем пренебречь последней. Взяв интеграл от MSE и минимизируя его по b , найдём оптимальный параметр сглаживания (17).

А.7. Доказательство леммы 5

Чтобы найти смещение оценки $\hat{f}'_{x_k}(x_1^d)$, используем тот же приём, что и в А.1. Математическое ожидание равно

$$\begin{aligned} \mathbb{E}_X(\hat{f}'_{x_k}(x_1^d)) &= \frac{1}{b_k} \mathbb{E} \left(\prod_{j=1}^d K_{\rho(x_j, b_j)}(X_j) L(X_k, x_k, b_k) \right) = \\ &= \frac{1}{b_k} \int L(t_k, x_k, b_k) f(t_1^d) \prod_{j=1}^d K_{\rho(x_j, b_j)}(t_j) dt_1^d = \frac{1}{b_k} \mathbb{E}_{\xi} (f(\xi_1^d) L(\xi_k, x_k, b_k)) = \\ &= \frac{1}{b_k} \mathbb{E}_{\xi} (f(\xi_1^d) \ln(\xi_n)) - \frac{1}{b_k} \mathbb{E}_{\xi} (f(\xi_1^d) (\ln b_k + \Psi(\rho(x_k, b_k)))) + o\left(\sum_{j=1}^d b_j\right), \quad (40) \end{aligned}$$

где $\{\xi_j\}$ — независимые одинаково гамма-распределённые случайные величины с математическим ожиданием $\mu_j = x_j$ и дисперсией $x_j b_j$. Используя разложение по Тейлору μ_1^d , можно записать

$$\begin{aligned} f(\xi_1^d) \ln(\xi_n) &= f(\mu_1^d) \ln(\mu_n) + \sum_{j=1}^d (\xi_j - \mu_j) \frac{\partial (f(\xi_1^d) \ln(\xi_n))}{\partial x_j} \Big|_{\mu} + \\ &+ \frac{1}{2} \sum_{j=1}^d (\xi_j - \mu_j)^2 \frac{\partial^2 (f(\xi_1^d) \ln(\xi_n))}{\partial x_j^2} \Big|_{\mu} + \\ &+ \sum_{j \neq l} (\xi_j - \mu_j)(\xi_l - \mu_l) \frac{\partial^2 (f(\xi_1^d) \ln(\xi_n))}{\partial x_j \partial x_l} \Big|_{\mu} + o\left(\sum_{j=1}^d b_j\right). \end{aligned}$$

Беря математическое ожидание от правой и левой частей последнего выражения, получаем первое слагаемое в (40):

$$\begin{aligned} \mathbb{E}_\xi(f(\xi_1^d) \ln(\xi_n)) &= f(x_1^d) \ln(x_k) + \frac{1}{2} \sum_{j=1}^d x_j b_j \frac{\partial^2 f}{\partial x_j^2} \ln(x_k) + \\ &+ \frac{1}{2} x_k b_k \left(\frac{2}{x_k} \frac{\partial f(x_1^d)}{\partial x_k} - \frac{f(x_1^d)}{x_k^2} \right) + o\left(\sum_{j=1}^d b_j\right). \end{aligned}$$

Для второго слагаемого в (40) используем аппроксимацию дигамма-функции при $\rho \rightarrow \infty$:

$$\Psi(\rho) = \ln \rho - \frac{1}{2\rho} - \frac{1}{12\rho^2} + \frac{1}{120\rho^4} - \frac{1}{252\rho^6} + O\left(\frac{1}{\rho^8}\right).$$

Следовательно, мы можем написать, что

$$\ln b_k + \Psi(\rho(x_k, b_k)) = \ln x_k - \frac{b_k}{2x_k} - \frac{b_k^2}{12x_k^2} + o(b_k^2).$$

Тогда (40) можно переписать следующим образом:

$$\begin{aligned} \mathbb{E}_X(\hat{f}'_{x_k}(x_1^d)) &= f(x_1^d) \frac{b_k}{12x_k^2} + \frac{\partial f(x_1^d)}{\partial x_k} + \\ &+ \frac{1}{2} \sum_{j=1}^d x_j b_j \frac{\partial^2 f(x_1^d)}{\partial x_j^2} \left(\frac{1}{2x_k} + \frac{b_k}{12x_k^2} \right) + o\left(\sum_{j=1}^d b_j\right). \quad (41) \end{aligned}$$

Многомерное математическое ожидание (41) согласуется с его одномерной версией в [15]. Из (41) получаем смещение оценки производной плотности:

$$\text{Bias}(\hat{f}'_{x_k}(x_1^d)) = f(x_1^d) \frac{b_k}{12x_k^2} + \frac{1}{2} \sum_{j=1}^d x_j b_j \frac{\partial^2 f(x_1^d)}{\partial x_j^2} \left(\frac{1}{2x_k} + \frac{b_k}{12x_k^2} \right) + o\left(\sum_{j=1}^d b_j\right).$$

В случае $b_1 = b_2 = \dots = b_d = b$ смещение оценки производной плотности задаётся формулой (18).

А.8. Доказательство леммы 6

По определению дисперсия для оценки $\hat{f}'_{x_k}(x_1^d)$ задаётся как

$$\begin{aligned} \text{Var}(\hat{f}'_{x_k}(x_1^d)) &= \frac{1}{n} \text{Var} \left(\frac{1}{b_k} L(X_k, x_k, b_k) \prod_{j=1}^d K_{\rho(x_j, b_j)}(X_j) \right) = \\ &= \frac{1}{nb_k^2} \mathbb{E} \left(L^2(X_k, x_k, b_k) \prod_{j=1}^d K_{\rho(x_j, b_j)}^2(X_j) \right) - \\ &- \frac{1}{nb_k^2} \mathbb{E}^2 \left(L(X_k, x_k, b_k) \prod_{j=1}^d K_{\rho(x_j, b_j)}(X_j) \right). \quad (42) \end{aligned}$$

Второй член правой части (42) является квадратом (40). Следовательно, используя (41), мы можем написать, что

$$\begin{aligned} & \frac{1}{n} \mathbf{E}_X^2(\hat{f}'_{x_k}(x_1^d)) = \\ & = \frac{1}{n} \left(\text{Bias}^2(\hat{f}'_{x_k}(x_1^d)) + \left(\frac{\partial f(x_1^d)}{\partial x_k} \right)^2 + 2 \frac{\partial f(x_1^d)}{\partial x_k} \text{Bias}(\hat{f}'_{x_k}(x_1^d)) + o\left(\sum_{j=1}^d b_j^2 \right) \right) \end{aligned} \quad (43)$$

Первый член правой части (42) может быть представлен как

$$\begin{aligned} & \frac{1}{nb_k^2} \mathbf{E} \left(L^2(X_k, x_k, b_k) \prod_{j=1}^d K_{\rho(x_j, b_j)}^2(X_j) \right) = \\ & = \frac{1}{nb_k^2} \int L^2(t_k, x_k, b_k) f(t_1^d) \prod_{j=1}^d K_{\rho(x_j, b_j)}^2(t_j) dt_1^d = \\ & = \frac{1}{nb_k^2} \int \left(\prod_{j=1}^d \frac{t_j^{2x_j/b_j-2} \exp(-2t_j/b_j)}{b_j^{2x_j/b_j} \Gamma^2(x_j/b_j)} \right) L^2(t_k, x_k, b_k) f(t_1^d) dt_1^d. \end{aligned} \quad (44)$$

Используя свойство гамма-функции

$$\Gamma^2\left(\frac{x}{b} + 1\right) = \left(\frac{x}{b}\right)^2 \Gamma^2\left(\frac{x}{b}\right),$$

получаем, что (44) может быть записана как

$$\begin{aligned} & \frac{1}{nb_k^2} \int \left(\prod_{j=1}^d \frac{b_j^{-3} x_j^2 \Gamma(2x_j/b_j - 1)}{2^{2x_j/b_j-1} \Gamma^2(x_j/b_j + 1)} \frac{t_j^{2x_j/b_j-2} \exp(-2t_j/b_j)}{(b_j/2)^{2x_j/b_j-1} \Gamma(2x_j/b_j - 1)} \right) \times \\ & \times L^2(t_k, x_k, b_k) f(t_1^d) dt_1^d = \\ & = \frac{1}{nb_k^2} \left(\prod_{j=1}^d B(x_j, b_j) \right) \mathbf{E}_\eta(L^2(\eta_k, x_k, b_k) f(\eta_1^d)) = \\ & = \frac{1}{nb_k^2} \left(\prod_{j=1}^d B(x_j, b_j) \right) \left(\mathbf{E}_\eta(f(\eta_1^d) \ln^2 \eta_k) + \left(\ln b_k + \Psi\left(\frac{x_k}{b_k}\right) \right)^2 \mathbf{E}_\eta(f(\eta_1^d)) - \right. \\ & \left. - 2 \left(\ln b_k + \Psi\left(\frac{x_k}{b_k}\right) \right) \mathbf{E}_\eta(f(\eta_1^d) \ln \eta_k) \right), \end{aligned} \quad (45)$$

где $\{\eta_j\}$ — независимые одинаково гамма-распределённые случайные величины с математическим ожиданием $\mu_j = x_j - b_j/2$ и дисперсией $x_j b_j/2 - b_j^2/4$, а $B(x_j, b_j)$ определена в (33). Чтобы определить математическое ожидания в (45), используем тот же метод, что и в предыдущих доказательствах. А именно,

$$\begin{aligned}
E_{\eta}(f(\eta_1^d) \ln(\eta_k)) &= f(\mu_1^d) \left(\ln \mu_k - \frac{b_k}{4\mu_k} \right) + \sum_{j=1}^d \left(\frac{x_j b_j}{2} - \frac{b_j^2}{4} \right) \frac{\partial^2 f(\mu_1^d)}{\partial x_j^2} \frac{\ln \mu_k}{2} + \\
&+ \frac{b_k}{2} \frac{\partial f(\mu_1^d)}{\partial x_k} + o\left(\sum_{j=1}^d b_j^2\right), \\
E_{\eta}(f(\eta_1^d)) &= f(\mu_1^d) + \frac{1}{2} \sum_{j=1}^d \left(\frac{x_j b_j}{2} - \frac{b_j^2}{4} \right) \frac{\partial^2 f(\mu_1^d)}{\partial x_j^2} + o\left(\sum_{j=1}^d b_j^2\right), \\
E_{\eta}(f(\eta_1^d) \ln^2(\eta_k)) &= f(\mu_1^d) \left(\ln^2(\mu_k) + \frac{b_k}{2\mu_k} (1 - \ln \mu_k) \right) + b_k \ln \mu_k \frac{\partial f(\mu_1^d)}{\partial x_k} + \\
&+ \frac{\ln^2 \mu_k}{2} \sum_{j=1}^d \left(\frac{x_j b_j}{2} - \frac{b_j^2}{4} \right) \frac{\partial^2 f(\mu_1^d)}{\partial x_j^2} + o\left(\sum_{j=1}^d b_j^2\right).
\end{aligned}$$

Используя известные формулы Тейлора

$$\ln \mu_k = \ln x_k - \frac{b_k}{2x_k} - \frac{b_k^2}{8x_k^2} + o(b_k^2), \quad \frac{1}{\mu_k} = \frac{1}{x_k} + \frac{b_k}{2x_k^2} + \frac{b_k^2}{4x_k^3} + o(b_k^2),$$

можно записать (45) как

$$\begin{aligned}
&\frac{1}{n} \left(\prod_{j=1}^d B(x_j, b_j) \right) \left(f(\mu_1^d) \left(\frac{1}{4x_k^2} + \frac{1}{2x_k b_k} + \frac{7b_k}{48x_k^3} + \frac{7b_n^2}{576x_k^4} + \frac{b_k^3}{192x_k^5} \right) + \right. \\
&\left. + \sum_{j=1}^d \left(\frac{x_j b_j}{2} - \frac{b_j^2}{4} \right) \frac{\partial^2 f(\mu_1^d)}{\partial x_j^2} \frac{b_k^2}{1152x_k^4} - \frac{b_k}{24x_k^2} \frac{\partial f(\mu_1^d)}{\partial x_k} + o\left(\sum_{j=1}^d b_j^2\right) \right).
\end{aligned}$$

Поэтому, используя разложение по Тейлору аргументов функций $f(\mu_1^d)$ в x_1^d , заменив его в последнем выражении и используя (43), мы можем записать дисперсию оценки производной плотности как

$$\begin{aligned}
\text{Var}(\hat{f}'_{x_k}(x_1^d)) &= \\
&= \frac{1}{n} \left(\prod_{j=1}^d \frac{b_j^{-1/2} x_j^{-1/2}}{2\sqrt{\pi}} \right) \left(-\frac{b_k}{24x_k^2} \left(\frac{\partial f(x_1^d)}{\partial x_k} - \frac{b_j}{2} \frac{\partial^2 f(x_1^d)}{\partial x_k \partial x_j} + \sum_{j=1}^d \frac{b_j^2}{4} \frac{\partial^3 f(x_1^d)}{\partial x_k \partial x_j^2} \right) + \right. \\
&+ \frac{b_k^2}{1152x_k^4} \sum_{j,i=n-\tau}^n \left(\frac{x_j b_j}{2} - \frac{b_j^2}{4} \right) \left(\frac{b_i^2}{4} \frac{\partial^4 f(x_1^d)}{\partial x_j^2 \partial x_i^2} - \frac{b_i}{2} \frac{\partial^3 f(x_1^d)}{\partial x_j^2 \partial x_i} \right) + \frac{\partial^2 f(x_1^d)}{\partial x_j^2} \times \\
&\times \left(\left(\frac{x_j b_j}{2} - \frac{b_j^2}{4} \right) \frac{b_k^2}{1152x_k^4} + \frac{b_j^2}{4} \left(\frac{1}{4x_k^2} + \frac{1}{2x_k b_k} + \frac{7b_k}{48x_k^3} + \frac{7b_n^2}{576x_k^4} + \frac{b_k^3}{192x_k^5} \right) \right) + \\
&+ \left(f(x_1^d) - \frac{b_j}{2} \frac{\partial f(x_1^d)}{\partial x_j} \right) \left(\frac{1}{4x_k^2} + \frac{1}{2x_k b_k} + \frac{7b_k}{48x_k^3} + \frac{7b_n^2}{576x_k^4} + \frac{b_k^3}{192x_k^5} \right) - \\
&- \frac{1}{n} \left(\text{Bias}^2(\hat{f}'_{x_k}(x_1^d)) + \left(\frac{\partial f(x_1^d)}{\partial x_k} \right)^2 + 2 \frac{\partial f(x_1^d)}{\partial x_k} \text{Bias}(\hat{f}'_{x_k}(x_1^d)) \right) + o\left(\sum_{j=1}^d b_j^2\right).
\end{aligned}$$

В случае когда $b_1 = b_2 = \dots = b_d = b$, получаем дисперсию оценки производной плотности (19).

А.9. Доказательство леммы 7

Применяем неравенство Давыдова для ковариации оценки производной плотности:

$$|\text{Cov}(\tilde{K}(X_1, x, b), \tilde{K}(X_{1+i}, x, b))| \leq 2\pi\alpha(i)^{1/r} \|\tilde{K}(X_1, x, b)\|_q \|\tilde{K}(X_{1+i}, x, b)\|_p,$$

где $p^{-1} + q^{-1} + r^{-1} = 1$, $1 \leq p, q, r \leq \infty$, $\alpha(i)$ — коэффициент перемешивания (1). Норма \mathcal{L}_p задаётся как

$$\begin{aligned} \|\tilde{K}(X_1, x, b)\|_q &= \left(\int (\tilde{K}(y, x, b))^q f(y) dy \right)^{1/q} = \\ &= \left(\int \left(L(t_k, x_k, b_k) \prod_{j=1}^d K_{\rho(x_j, b_j), b_j}(t_j) \right)^q f(t_1^d) dt_1^d \right)^{1/q} = \\ &= \left(\int \left(L^q(t_k, x_k, b_k) f(t_1^d) \prod_{j=1}^d K_{\rho(x_j, b_j), b_j}^{q-1}(t_j) \right) \prod_{j=1}^d K_{\rho(x_j, b_j), b_j}(t_j) dt_1^d \right)^{1/q} = \\ &= \left(\mathbb{E} \left(L(\xi_k, x_k, b_k)^q f(\xi_1^d) \prod_{j=1}^d K_{\rho(x_j, b_j), b_j}^{q-1}(\xi_j) \right) \right)^{1/q}, \end{aligned} \quad (46)$$

где использовано ядро $\prod_{j=1}^d K_{\rho_1(x_j), b_j}(\xi_j)$, а $\{\xi_j\}$ — независимые одинаково гамма-распределённые случайные величины с математическим ожиданием $\mu_j = x_j$ и дисперсией $x_j b_j$. Разложение по Тейлору (46) следующее:

$$\begin{aligned} L(\xi_n, x_k, b_k)^q f(\xi_1^d) \prod_{j=1}^d K_{\rho(x_j, b_j), b_j}^{q-1}(\xi_j) &= \\ &= \sum_{i=1}^d (\xi_i - \mu_i) \frac{\partial}{\partial \xi_i} \left(L(\xi_k, x_k, b_k)^q f(\xi_1^d) \prod_{j=1}^d K_{\rho(x_j, b_j), b_j}^{q-1}(\xi_j) \right) \Big|_{\xi=\mu} + \\ &+ \sum_{i \neq l} \frac{(\xi_i - \mu_i)(\xi_l - \mu_l)}{2} \frac{\partial^2}{\partial \xi_i^2} \left(L(\xi_k, x_k, b_k)^q f(\xi_1^d) \prod_{j=1}^d K_{\rho(x_j, b_j), b_j}^{q-1}(\xi_j) \right) \Big|_{\xi=\mu} + \\ &+ \sum_{i=1}^d \frac{(\xi_i - \mu_i)^2}{2} \frac{\partial^2}{\partial \xi_i^2} \left(L(\xi_k, x_k, b_k)^q f(\xi_1^d) \prod_{j=1}^d K_{\rho(x_j, b_j), b_j}^{q-1}(\xi_j) \right) \Big|_{\xi=\mu} + \\ &+ L(\mu_k, x_k, b_k)^q f(\mu_1^d) \prod_{j=1}^d K_{\rho(x_j, b_j), b_j}^{q-1}(\mu_j) + o\left(\sum_{j=1}^d b_j^2 \right). \end{aligned}$$

Взяв математическое ожидание от последнего выражения и используя известные формулы Тейлора

$$\begin{aligned} L(x_k, x_k, b_k) &= \ln x_k - \ln b_k - \Psi\left(\frac{x_k}{b_k}\right) = \frac{b_k}{2x_k} + \frac{b_k^2}{12x_k^2} + o(b_k^2), \\ L^q(x_k, x_k, b_k) &= \left(\frac{b_k}{2x_k}\right)^q \left(1 + q\frac{b_k}{6x_k} + \frac{q(q-1)}{2}\left(\frac{b_k}{6x_k}\right)^2\right), \\ \frac{\partial}{\partial \xi_k} L(\xi_n, x_k, b_k) &= \frac{1}{\xi_k} \Big|_{\xi_k=x_k} = \frac{1}{x_k}, \\ \frac{\partial^2}{\partial \xi_k^2} L(\xi_k, x_k, b_k) &= -\frac{1}{\xi_n} \Big|_{\xi_k=x_k} = -\frac{1}{x_k^2}, \end{aligned}$$

математическое ожидание (46) можно записать как

$$\begin{aligned} &\mathbb{E}\left(\prod_{j=1}^d K_{\rho_1(x_j), b_j}^{q-1}(\xi_j) L_1(\xi_n, x_k, b_k)^q f(\xi_1^d)\right) = \\ &= \left(\prod_{j=1}^d K_{\rho_1(x_j), b_j}^{q-1}(x_j)\right) \left(\left(\frac{b_k}{2x_k}\right)^q \left(1 + q\frac{b_k}{6x_k} + \frac{q(q-1)}{2}\left(\frac{b_k}{6x_k}\right)^2\right) f(x_1^d) + \right. \\ &+ \left(\frac{b_k}{2x_k}\right)^{q-1} \sum_{i=1}^d \frac{x_i b_i}{2} \left(\left(\frac{b_k}{2x_k} + \frac{qb_k^2}{12x_k^2} + \frac{q(q-1)b_k^3}{144x_k^3}\right) \times \right. \\ &\times \left. \left(f(x_1^d) \left(\frac{1-q}{b_i x_i} - \frac{q(1-q)}{x_i^2}\right) + \frac{\partial f(x_1^d)}{\partial x_i} \frac{2(1-q)}{x_i} + \frac{\partial^2 f(x_1^d)}{\partial x_i^2}\right) - \right. \\ &\left. - \left(1 + \frac{(q-1)b_k}{6x_k} + \frac{(q-1)(q-2)b_k^2}{72x_k^2}\right) \left(f(x_1^d) \frac{q^2}{x_k^2} - \frac{2q}{x_k} \frac{\partial f(x_1^d)}{\partial x_k}\right)\right) + o\left(\sum_{j=1}^d b_j^2\right). \end{aligned}$$

Если $b_1 = b_2 = \dots = b_d = b$, то, используя (36), получаем

$$\begin{aligned} \|\tilde{K}(X_1, x, b)\|_q &\leq \left(\prod_{j=1}^d (2\pi x_j)^{(1-q)/2q}\right) \frac{b^{d(1-q)/(2q)+1}}{2x_k} \times \\ &\times \left(\sum_{i=1}^d \left(1 + \frac{qb}{6x_k} + \frac{q(q-1)b^2}{72x_k^2}\right) \times \right. \\ &\times \left. \left(f(x_1^d) \frac{3-q}{2} + b(1-q) \frac{\partial f(x_1^d)}{\partial x_i} + \frac{x_i b}{2} \frac{\partial^2 f(x_1^d)}{\partial x_i^2} - \frac{q(1-q)b}{2x_i} f(x_1^d)\right) + \right. \\ &\left. + x_i \left(1 + \frac{(q-1)b}{6x_k} + \frac{(q-1)(q-2)b^2}{72x_k^2}\right) \left(2q \frac{\partial f(x_1^d)}{\partial x_k} - \frac{q^2}{x_k} f(x_1^d)\right)\right)^{1/q} + o(b^2). \end{aligned}$$

То же верно для $\|\tilde{K}(X_{1+i}, x, b)\|_p$. Выберем $p = q$. Тогда ковариация оценки производной плотности определяется формулой

$$\begin{aligned}
& |\text{Cov}(\tilde{K}(X_1, x, b), \tilde{K}(X_{1+k}, x, b))| \leq \\
& \leq 2\pi\alpha(k)^{1/r} \left(\prod_{j=1}^d (2\pi x_j)^{(1-q)q} \right) \frac{b^{d(1-q)/q+2}}{4x_k^2} \left(\sum_{i=1}^d \left(1 + \frac{qb}{6x_k} + \frac{q(q-1)b^2}{72x_k^2} \right) \times \right. \\
& \times \left(f(x_1^d) \frac{3-q}{2} + b(1-q) \frac{\partial f(x_1^d)}{\partial x_i} + \frac{x_i b}{2} \frac{\partial^2 f(x_1^d)}{\partial x_i^2} - \frac{q(1-q)b}{2x_i} f(x_1^d) \right) + \\
& \left. + x_i \left(1 + \frac{(q-1)b}{6x_k} + \frac{(q-1)(q-2)b^2}{72x_k^2} \right) \left(2q \frac{\partial f(x_1^d)}{\partial x_k} - \frac{q^2}{x_k} f(x_1^d) \right) \right)^{2/q} + o(b^2).
\end{aligned}$$

Выберем $p = q = 2 + \delta$, $r = (2 + \delta)/\delta$. Используя обозначения

$$\begin{aligned}
V(\delta, x_1^d) &= \sum_{i=1}^d \left(-(\delta+1) \frac{\partial f(x_1^d)}{\partial x_i} + \frac{\delta(\delta+1)(\delta+2)}{36x_k^2} \frac{\partial f(x_1^d)}{\partial x_k} + \frac{x_i}{2} \frac{\partial^2 f(x_1^d)}{\partial x_i^2} - \right. \\
&\quad \left. - \left(\frac{(\delta-1)(\delta+1)(\delta+2)}{144x_k^2} - \frac{(\delta+1)(\delta+2)}{2x_i} + \frac{\delta(\delta+1)(\delta+2)^2}{72x_k^3} \right) f(x_1^d) \right), \\
W(\delta, x_1^d) &= \sum_{i=1}^d \left(\left(\frac{1-\delta}{4x_k} + \frac{(\delta+1)(\delta+2)}{2x_i} - \frac{x_i(\delta+1)(\delta+2)^2}{6x_k^2} \right) f(x_1^d) - \right. \\
&\quad \left. - (\delta+1) \frac{\partial f(x_1^d)}{\partial x_i} + \frac{(\delta+1)(\delta+2)x_i}{3x_k} \frac{\partial f(x_1^d)}{\partial x_k} + \frac{x_i}{2} \frac{\partial^2 f(x_1^d)}{\partial x_i^2} \right), \\
L(\delta, x_1^d) &= \sum_{i=1}^d \left(f(x_1^d) \frac{(1-\delta)}{2} + x_i \left(2(2+\delta) \frac{\partial f(x_1^d)}{\partial x_k} - \frac{(2+\delta)^2}{x_k} f(x_1^d) \right) \right), \\
R(\delta, x_1^d) &= \left(\prod_{j=1}^d x_j^{-(\delta+1)/(\delta+2)} \right) \frac{(2\pi)^{(\delta+2-d(\delta+1))/(\delta+2)}}{2x_k^2},
\end{aligned}$$

получаем, что ковариация оценки производной плотности задаётся выражением

$$\begin{aligned}
\text{Cov}(\hat{f}'_{x_k}(x_1^d)) &= \frac{2}{nb^2} \sum_{i=1}^n \left(1 - \frac{i}{n} \right) \text{Cov}(\tilde{K}(X_1, x, b), \tilde{K}(X_{1+i}, x, b)) \leq \\
&\leq \frac{R(\delta, x_1^d)}{b^{d(\delta+1)/(\delta+2)n}} (b^2 V(\delta, x_1^d) + bW(\delta, x_1^d) + L(\delta, x_1^d))^{2/(\delta+2)} \int_1^\infty \alpha(\tau)^{\delta/(2+\delta)} d\tau + o(b^2).
\end{aligned}$$

Обозначим $\delta/(2 + \delta) = v$, $0 < v < 1$. Тогда получаем (20).

А.10. Доказательство теоремы 3

Используя смещение, дисперсию и ковариацию лемм 5–7, запишем средне-квадратическую ошибку для оценки производной плотности:

$$\begin{aligned}
\text{MSE}(\hat{f}'_{x_k}(x_1^d)) &= \text{Bias}(\hat{f}'_{x_k}(x_1^d))^2 + \text{Var}(\hat{f}'_{x_k}(x_1^d)) + \text{Cov}(\hat{f}'_{x_k}(x_1^d)) = \\
&= (bB_1(x_1^d) + b^2B_2(x_1^d))^2 \left(1 - \frac{1}{n}\right) + \\
&+ \frac{b^{-d/2}}{n} \left(\prod_{j=1}^d \frac{x_j^{-1/2}}{2\sqrt{\pi}}\right) (bV_1(x_1^d) + b^2V_2(x_1^d) + \frac{1}{b}V_3(x_1^d) + V_4(x_1^d)) - \\
&- \frac{1}{n} \left(\left(\frac{\partial f(x_1^d)}{\partial x_k}\right)^2 + 2\frac{\partial f(x_1^d)}{\partial x_k} (bB_1(x_1^d) + b^2B_2(x_1^d))\right) + \frac{R(v, x_1^d)}{n} b^{-d(v+1)/2} \times \\
&\times (b^2V(v, x_1^d) + bW(v, x_1^d) + L(v, x_1^d))^{1-v} \int_1^\infty \alpha(\tau)^v d\tau + o(b^2).
\end{aligned}$$

Так как скорость сходимости ковариации меньше скорости сходимости дисперсии, мы можем пренебречь ковариационным членом. Беря интеграл от MSE и минимизируя MISE по b , получаем оптимальный параметр сглаживания (21).

Работа выполнена при поддержке гранта № 14-50-00150 Российского научного фонда.

Литература

- [1] Aksoy H. Use of gamma distribution in hydrological analysis // Turkish J. Eng. Environ. Sci. — 2000. — Vol. 24. — P. 419–428.
- [2] Andrews D. W. K. First order autoregressive processes and strong mixing. — Yale University, Cowles Foundation Discussion Paper 66. — 1983.
- [3] Bhattacharya P. K. Estimation of a probability density function and its derivatives // Indian J. Statist. — 1967. — Vol. A 29. — P. 373–382.
- [4] Bosq S. Nonparametric Statistics for Stochastic Processes. Estimation and Prediction. — New York: Springer, 1996.
- [5] Bouezmarni T., Rombouts J. V. K. Nonparametric density estimation for multivariate bounded data // J. Statist. Planning Inference. — 2007. — Vol. 140, no. 1. — P. 139–152.
- [6] Bouezmarni T., Rombouts J. V. K. Nonparametric density estimation for positive times series // Comput. Statist. Data Analysis. — 2010. — Vol. 54, no. 2. — P. 245–261.
- [7] Bouezmarni T., Scaillet O. Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data // Econometric Theory. — 2005. — Vol. 21, no. 2. — P. 390–412.
- [8] De Brabanter K., de Brabanter J., de Moor B. Nonparametric derivative estimation // Proc. of the 23rd Benelux Conf. on Artificial Intelligence (BNAIC), Gent, Belgium. — 2011. — P. 75–81.
- [9] Brown B. M., Chen S. X. Beta-Bernstein smoothing for regression curves with compact support // Scand. J. Statist. — 1999. — Vol. 26. — P. 47–59.
- [10] Chen S. X. Probability density function estimation using gamma kernels // Ann. Inst. Statist. Math. — 2000. — Vol. 54. — P. 471–480.

- [11] Dobrovidov A. V., Koshkin G. M. Data-based nonparametric signal filtration // *Austr. J. Statist.* — 2011. — Vol. 40, no. 1. — P. 15–24.
- [12] Dobrovidov A. V., Koshkin G. M. Regularized data-based nonparametric filtration of stochastic signals // *Proc. of the World Congress on Engineering 2011. Vol. 1.* — 2011. — P. 333–337.
- [13] Dobrovidov A. V., Koshkin G. M., Vasiliev V. A. *Non-Parametric State Space Models.* — Kendrick Press, 2012.
- [14] Dobrovidov A. V., Markovich L. A. Data-driven bandwidth choice for gamma kernel estimates of density derivatives on the positive semi-axis // *IFAC Proc. Volumes.* — 2013. — Vol. 46, no. 11. — P. 500–505.
- [15] Dobrovidov A. V., Markovich L. A. Nonparametric gamma kernel estimators of density derivatives on positive semi-axis // *IFAC Proc. Volumes.* — 2013. — Vol. 46, no. 9. — P. 910–915.
- [16] Funke B., Kawka R. Nonparametric density estimation for multivariate bounded data using two non-negative multiplicative bias correction methods // *Comput. Statist. Data Analysis.* — 2015. — Vol. 92. — P. 148–162.
- [17] Furman E. On a multivariate gamma distribution // *Statist. Probab. Lett.* — 2008. — Vol. 78. — P. 2353–2360.
- [18] Hall P., Wehrly T. E. A geometrical method for removing edge effects from kernel-type nonparametric regression estimators // *J. Am. Statist. Assoc.* — 1991. — Vol. 86. — P. 665–672.
- [19] Hürlimann W. Analytical evaluation of economic risk capital for portfolios of gamma risks // *Turkish J. Eng. Environ. Sci.* — 2001. — Vol. 31. — P. 107–122.
- [20] Igarashi G., Kakizawa Y. Bias corrections for some asymmetric kernel estimators // *J. Statist. Planning Inference.* — 2015. — Vol. 159. — P. 37–63.
- [21] Jones M. C. Simple boundary correction for density estimation kernel // *Statist. Comput.* — 1993. — Vol. 3. — P. 135–146.
- [22] Kushnir A. F. Asymptotically optimal tests for a regression problem of testing hypotheses // *Theory Probab. Its Appl.* — 1967. — Vol. 13, no. 4. — P. 647–666.
- [23] Lejeune M., Sarda P. Smooth estimators of distribution and density functions // *Comput. Statist. Data Analysis.* — 1992. — Vol. 14. — P. 457–471.
- [24] Markovich L. A. Inferences from optimal filtering equation // *Lithuan. Math. J.* — 2015. — Vol. 55, no. 3. — P. 413–432.
- [25] Markovich L. A. Gamma kernel estimation of the density derivative on the positive semi-axis by dependent data // *REVSTAT-Statist. J. Statist. Portugal.* — 2016. — Vol. 14, no. 3. — P. 327–348.
- [26] Mathal A. M., Moschopoulos P. G. A form of multivariate gamma distribution // *Ann. Inst. Statist. Math.* — 1992. — Vol. 44. — P. 97–106.
- [27] Müller H. G. Smooth optimum kernel estimators near endpoints // *Biometrika.* — 1991. — Vol. 78, no. 3. — P. 521–530.
- [28] Nadarajah S. Reliability for some bivariate gamma distributions // *Math. Probab. Eng.* — 2005. — Vol. 2. — P. 151–163.
- [29] Nadarajah S., Gupta A. K. Some bivariate gamma distributions // *Appl. Math. Lett.* — 2006. — Vol. 19. — P. 767–774.

- [30] Nadaraya É. A. On non-parametric estimates of density functions and regression curves // *Theory Probab. Its Appl.* — 1964. — Vol. 10, no. 1. — P. 186–190.
- [31] Parzen E. On estimation of a probability density function and mode // *Ann. Math. Statist.* — 1962. — Vol. 33, no. 3. — P. 1065–1076.
- [32] Rosenblatt M. Remarks on some nonparametric estimates of a density function // *Ann. Math. Statist.* — 1956. — Vol. 27, no. 3. — P. 832–837.
- [33] Sasaki H., Hyvärinen A., Sugiyama M. Clustering via mode seeking by direct estimation of the gradient of a log-density // *Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2014 / Calders T., Esposito F., Hüllermeier E., Meo R., eds.* — Berlin: Springer, 2014. — (Lect. Notes Comput. Sci.; Vol. 8726).
- [34] Scaillet O. Density estimation using inverse and reciprocal inverse Gaussian kernels // *J. Nonparam. Statist.* — 2004. — Vol. 16. — P. 217–226.
- [35] Schuster E. F. Estimation of a probability function and its derivatives // *Ann. Math. Statist.* — 1969. — Vol. 40. — P. 1187–1195.
- [36] Schuster E. F. Incorporating support constraints into nonparametric estimators of densities // *Commun. Statist. Theory Methods.* — 1985. — Vol. 14. — P. 1123–1136.
- [37] Silverman B. W. *Density Estimation for Statistics and Data Analysis.* — London: Chapman and Hall, 1986.
- [38] Tsytkin Ya. Z. Optimality in adaptive control systems // *Uncertainty and Control / J. Ackermann, ed.* — Berlin: Springer, 1985. — (Lect. Notes Control Information Sci.; Vol. 70). — P. 153–214.
- [39] Turlach B. A. Bandwidth Selection in Kernel Density Estimation: A Review. — CORE and Inst. de Statistique, 1993.
- [40] Wand M. P., Jones M. C. *Kernel Smoothing.* — London: Chapman and Hall, 1995.
- [41] Zhang S. A note on the performance of the gamma kernel estimators at the boundary // *Statist. Probab. Lett.* — 2010. — Vol. 80. — P. 548–557.

