УДК 519.2

Ядерные оценки плотности в пространствах произвольной природы

А.И. Орлов[1]

[1] Московский государственный технический университет им. Н.Э. Баумана, Институт высоких статистических технологий и эконометрики; Московский физико-технический институт; Москва, Россия; Центральный научно-исследовательский институт машиностроения; г. Королев Московской обл., Россия; prof-orlov@mail.ru; http://orlovs.pp.ru; +7(916)8305117; 123104, Москва, Сытинский пер., д.7/14, кв.14.

Аннотация. *Изучение линейных оценки плотности распределения вероятностей в пространствах произвольной природы начато в статье автора в предыдущем вып.25. Частный случай таких оценок - ядерные оценки плотности. Настоящая статья посвящена основам общей теории ядерных оценок плотности. При различных условиях доказана состоятельность и асимптотическая нормальность ядерных оценок плотности, изучена равномерная сходимость. Найдена асимптотика дисперсии ядерной оценки плотности. Рассмотрены примеры, в том числе в пространстве интегрируемых с квадратом функций.*

Ключевые слова: *нечисловая статистика, плотность распределения вероятностей, пространства произвольной природы, линейные оценки плотности, ядерные оценки плотности, предельные теоремы, состоятельные оценки, асимптотическая нормальность, равномерная сходимость.*

1. Введение

Оценки плотности распределения вероятностей в пространствах произвольной природы используют для решения различных задач нечисловой статистики [1], называемой также статистикой объектов нечисловой природы или статистикой нечисловых данных. Систематическое изложение теории таких оценок начато в статье [2], непосредственным продолжением которой является настоящая статья. Регулярно используются ссылки на условия и утверждения из статьи [2].

Пусть (Z, A) – измеримое пространство, p и q – сигма-конечные меры на (Z, A), причем p абсолютно непрерывна относительно q, т.е. из $q(B) = 0$ следует $p(B) = 0$ для любого множества B из сигма-алгебры A. В этом случае на (Z, A) существует неотрицательная измеримая функция $f(x)$ такая, что

$$q(C) = \int\limits_C f(x)dp \qquad (1)$$

для любого множества C из сигма-алгебры измеримых множеств A. Функция $f(x)$ называется производной Радона – Никодима меры q по мере p, а в случае, когда q – вероятностная мера, также плотностью вероятности q по отношению к мере p [3, с.460].

Пусть $X_1, X_2, ..., X_n$ – независимые одинаково распределенные случайные элементы (величины), распределение которых задается вероятностной мерой q. В статье [2] введено несколько видов непараметрических оценок плотности вероятности q по выборке $X_1, X_2, ..., X_n$. Подробнее изучены линейные оценки. В настоящей статье рассмотрим их частные случаи – ядерные оценки плотности в пространствах произвольной природы.

2. Ядерные оценки плотности

В предположении непрерывности неизвестной плотности $f(x)$ представляется целесообразным "размазать" каждый атом эмпирической меры, т.е. рассмотреть линейные оценки, введенные в нашей первой работе по нечисловой статистике [4, с.24]:

$$f_n(x) = \frac{1}{n} \sum_{1 \le i \le n} g_n(x, X_i), \qquad g_n : Z^2 \to R^1, \tag{2}$$

в которых действительнозначные функции g_n удовлетворяют некоторым условиям регулярности, обсуждаемым ниже. Ядерные оценки плотности – это линейные оценки (2), в которых слагаемые имеют специальный вид.

Пусть d – показатель различия (синоним – мера близости) на Z (см., например [1]). В наиболее важных частных случаях он представляет собой метрику на Z. В докладе [5] нами введены ядерные оценки плотности – оценки вида (2) с

$$g_n(x, X_i) = \frac{1}{b(h_n, x)} K\left(\frac{d(x, X_i)}{h_n}\right), \qquad K : [0, +\infty) \to R^1, \tag{3}$$

где $K = K(u)$ – ядро (ядерная функция), h_n– последовательность положительных чисел (показателей размытости), $b(h_n, x)$ – нормировочный множитель. В статье [6] линейные оценки (2) с функциями g_n из (3) названы „обобщенными оценками типа Парзена – Розенблатта“, т.к. в частном случае $Z = R^1, d(x, X_i) = |x - X_i|, b(h_n, x) = h_n$ они переходят в известные оценки, введенные Розенблаттом [7] и Парзеном [8]. Поскольку Парзен и Розенблатт никогда не рассматривали оценки вида (2) для пространств произвольной природы, далее будем называть оценки (2) – (3) ядерными.

Согласно [2] линейные оценки (2) удовлетворяют условию нормировки:

$$\int_Z g_n(x, y) p(dy) = 1. \tag{4}$$

Из условия нормировки следует, что нормировочный множитель $b(h_n, x)$ имеет вид

$$b(h_n, x) = \int\limits_{Z} K\left(\frac{d(x,y)}{h_n}\right) p(dy). \tag{5}$$

Введем шар радиуса t с центром в точке x:

$$L_t(x) = \{y : d(x, y) < t\} \tag{6}$$

(в соответствии с замечанием после условия (IV) статьи [2] шар $L_t(x)$ – измеримое множество, т.е. $L_t(x) \in A$). Зависимость меры шара от радиуса зададим выражением

$$F_x(t) = p(L_t(x)). \tag{7}$$

Справедливо полезное для дальнейших рассуждений равенство

$$\int\limits_{L_t(x)} K\left(\frac{d(x,y)}{h_n}\right) p(dy) = \int\limits_0^t K\left(\frac{t}{h_n}\right) dF_x(t). \tag{8}$$

Для ядерных оценок (3) рассмотрим условия (I) - (VII), введенные в статье [2] для линейных оценок.

С учетом (8) условие (I) переходит в следующее:

$$\frac{1}{b(h_n, x)} \int\limits_0^\infty \left| K\left(\frac{t}{h_n}\right) \right| dF_x(t) < \infty. \tag{9}$$

При рассмотрении условия (VI) естественно принять, что для любой окрестности $U(x)$ точки x найдется положительный радиус $t > 0$ такой, что шар $L_t(x)$ содержится в $U(x)$, т.е. топология в Z порождена системой шаров $L_t(x)$, $x \in Z, t > 0$. Тогда вместо (VI) получаем условие:

$$\lim_{n\to\infty} \frac{1}{b(h_n, x)} \int\limits_t^\infty \left| K\left(\frac{g}{h_n}\right) \right| dF_x(g) = 0 \tag{10}$$

для любого $t > 0$.

Замечание. Симметричность меры близости d, т.е. свойство $d(x, y) = d(y, x)$, в настоящей статье не используется. Примером несимметричного показателя различия $d : R^2 \to [0, +\infty)$ является

$$d(x, y) = \left| \frac{y}{x} - 1 \right|. \tag{11}$$

Имеем взамен условия (VII)

$$d_n = d_n(x) = \frac{1}{b^2(h_n, x)} \int\limits_0^\infty K^2\left(\frac{g}{h_n}\right) dF_x(g) < \infty. \tag{12}$$

Формулировки условий (II), (IV), (V) менять нет оснований.

Из теорем 1 и 2 статьи [2] вытекает следующее утверждение.

Теорема 1. *Пусть выполнены соотношения (5), (9), (10), а также условия* $(II), (IV), (V)$ *статьи [2]. Тогда ядерная оценка плотности (2) – (3) является состоятельной, причем*

$$\lim_{n\to\infty} \mathbf{M} f_n(x) = f(x). \tag{13}$$

Если, кроме того, выполнено (12) и условия

$$\mathbf{D}K\left(\frac{d(x, X_1)}{h_n}\right) \neq 0, \quad \lim_{n\to\infty} \frac{d_n}{n} = 0, \tag{14}$$

то $f_n(x)$ *– состоятельная и асимптотически нормальная оценка плотности* f *в точке* x*.*

Для доказательства достаточно сослаться на результаты статьи [2] о линейных оценках и доказать соотношение (18) этой статьи, т.е.

$$\lim_{n\to\infty} \frac{1}{n} \mathbf{D} g_n(x, X_1) = 0. \tag{15}$$

Имеем с учетом условия (II) статьи [2]

$$\frac{1}{n} \mathbf{D} g_n(x, X_1) \leq \frac{c_2 d_n}{n}, \tag{16}$$

откуда в соответствии с (14) следует (15). Теорема 1 доказана.

Если ядерная функция $K(u)$ является финитной, т.е. $K(u) = 0$ при $u > u_0$, то значение оценки плотности $f_n(x)$ определяется только по элементам выборки X_i, для которых $d(x, X_i) < u_0 h_n$. Поскольку предполагаем, что $h_n \to 0$ при $n \to \infty$, то $f_n(x)$ определяется по сужающейся окрестности, а верхние пределы интегрирования в формулах (8) – (10), (12) можно заменить на $u_0 h_n$. Поскольку $u_0 h_n \to 0$ при $n \to \infty$, то соотношение (10) выполнено автоматически. Если $K(u)$ – ограниченная функция, то (9) и (12) также выполнены. Вместо условия (II) статьи [2] достаточно ограниченности f в некоторой окрестности x, а это вытекает из непрерывности f. Согласно сказанному получаем следующую теорему.

Теорема 2. *Пусть K – ограниченная финитная функция, $h_n \to 0$ при $n \to \infty$, справедливы условия (IV), (V) статьи [2] и следствие условия нормировки (5). Тогда выполнено (13). Если, кроме того, справедливы условия (14), то $f_n(x)$ – состоятельная и асимптотически нормальная оценка плотности f в точке x.*

3. Предпочтительный показатель различия

Чтобы продвинуться дальше, введем более сильное ограничение:

(VIII) Функция $F_x(t)$ – непрерывная функция по t.

Это ограничение сильное, поскольку исключает распределения, имеющие атомы, в частности, все конечные пространства Z. (В дальнейших исследованиях мы рассмотрим последовательность конечных Z и заменим условие (VIII) на некоторую аппроксимацию.)

Поскольку статистик может сам выбирать меру близости d, введем новый показатель (назовем его "предпочтительный показатель различия")

$$d_1(x, y) = F_x(d(x, y)). \tag{17}$$

Тогда

$$F_x^1(t) = p(L_t^1(x)) = p\{y : d_1(x, y) < t\} = p\{y : F_x(d(x, y)) < t\}. \tag{18}$$

Поскольку $F_x(t)$ – непрерывная по t функция, то определена обратная функция

$$F_x^{-1}(\alpha) = \min_\beta\{\beta : F_\alpha(\beta) = \alpha\}. \tag{19}$$

Она возрастает по α, т.к. функция $F_x(t)$ не убывает по t. Поэтому неравенство $F_x(d(x, y)) < t$ эквивалентно неравенству $F_x^{-1}(F_x(d(x, y)) < F_x^{-1}(t)$. Тогда

$$F_x^{-1}(F_x(d(x, y)) = \min_\beta\{\beta : F_x(\beta) = F_x(d(x, y))\} \tag{20}$$

по определению обратной функции. Чтобы не усложнять излишне изложение, несколько усилим условие (VIII).

(VIII′) $F_x(t)$ – строго возрастающая непрерывная по t функция, причем $F_x(0) = 0$.

Если выполнено условие (VIII′), то $F_x^{-1}(\alpha)$ – тоже строго возрастающая функция, непрерывная на области своего определения, и

$$F_x^{-1}(F_x(t)) = t, \qquad F_x(F_x^{-1}(\alpha)) = \alpha.$$

Если t входит в область определения функции F_x^{-1}, то

$$F_x^1(t) = p\{y : F_x(d(x, y)) < t\} = p\{y : d(x, y) < F_x^{-1}(t)\} =$$

$$= F_x(F_x^{-1}(t)) = t. \tag{21}$$

Если $p(Z) = \infty$, то цепочка равенств (21) верна при любом t. Если же $p(Z) < \infty$, то $F_x^1(t) = p(Z)$ при $t \geq p(Z)$. Итак,

$$F_x^1(t) = \begin{cases} t, & 0 \leq t \leq p(Z), \\ p(Z), & t \geq p(Z). \end{cases} \tag{22}$$

Пример 1. Пусть $Z = R^k$, d – евклидово расстояние, p – мера Лебега. Тогда $F_x(t)$ не зависит от x и $F_x(t) = F(t) = p\{y : d(x,y) < t\} = c_k t^k$, где c_k – объем шара единичного радиуса в R^k, естественный показатель различия имеет вид $d_1(x,y) = c_k d^k(x,y)$, поскольку

$$p\{y : c_k d^k(x,y) < t\} = p\left\{y : d(x,y) < \left(\frac{t}{c_k}\right)^{\frac{1}{k}}\right\} = c_k\left(\left(\frac{t}{c_k}\right)^{\frac{1}{k}}\right)^k = t. \tag{23}$$

Пример 2. При непараметрическом оценивании плотности в R_k часто используют ядра вида

$$K\left(\frac{x_1 - X_1(i)}{h_n}, \frac{x_2 - X_2(i)}{h_n}, ..., \frac{x_k - X_k(i)}{h_n}\right), \tag{24}$$

причем добавляют различные требования симметричности. В виде ядерных функций (3) такие ядра, как правило, представить нельзя, т.к. в (3) $K : [0, +\infty) \to R^1$ – функция одного числового аргумента. Если $d(x,y) = \max_j |x_j - y_j|$, то функция $K = K(d)$ постоянна на поверхности куба $\max_j |x_j - y_j| = const$. Тогда

$$F(t) = p\{y : d(x,y) < t\} = 2^k t^k, \qquad d_1(x,y) = 2^k d^k(x,y). \tag{25}$$

Поскольку для любого расстояния в R^k, полученного с помощью некоторой нормы в R^k, функция $F(t)$ имеет порядок t^k, то всегда $d_1 = Cd^k$, где C – константа. В R^k в качестве p обычно используют меру Лебега, хотя можно рассматривать, например, меру, порожденную нормально распределенным случайным вектором. При изучении скорости сходимости ядерных оценок можно увидеть некоторые преимущества меры Лебега [9].

4. Ядерные оценки плотности на основе предпочтительного показателя различия

Далее будем использовать предпочтительный показатель различия. Пусть $F_x(t)$ имеет вид (22). Рассмотрим сначала случай $p(Z) = \infty$. С помощью замены переменных $u = t/h_n$ из (5) и (8) получаем, что

$$b(h_n, x) = \int_0^\infty K\left(\frac{t}{h_n}\right) dt = h_n \int_0^\infty K(u) du. \tag{26}$$

Поскольку согласно (3) можно одновременно умножить ядерную функцию K и нормировочный множитель b на одну и ту же константу и $b(h_n, x) \neq 0$, то можно положить

$$\int\limits_0^\infty K(u)du = 1. \tag{27}$$

Тогда

$$b(h_n, x) = h_n \tag{28}$$

и

$$g_n(x, X) = \frac{1}{h_n} K\left(\frac{d(x, X)}{h_n}\right). \tag{29}$$

Если же $p(Z) < \infty$, то при справедливости (27) вместо (29) имеем

$$b(h_n, x) = h_n \left(1 - \int\limits_{p(Z)/h_n}^\infty K(u)du\right). \tag{30}$$

Если ядерная функция $K(u)$ финитна, то в силу $h_n \to 0$ при достаточно большом n справедливы равенства (28) и (29).

Соотношение (9) переходит в условие

$$\frac{1}{h_n} \int\limits_0^\infty \left|K\left(\frac{t}{h_n}\right)\right| dt = \int\limits_0^\infty |K(u)|\, du < \infty. \tag{31}$$

Если неравенство (31) выполнено, то для любого $t > 0$ при $n \to \infty$

$$\frac{1}{h_n} \int\limits_t^\infty \left|K\left(\frac{g}{h_n}\right)\right| dg = \int\limits_{t/h_n}^\infty |K(u)|\, du \to 0, \tag{32}$$

поскольку $h_n \to 0$. Следовательно, соотношение (10) выполнено.

Справедливость условия (V) вытекает из (31). Далее имеем

$$d_n = d_n(x) = \frac{1}{h_n^2} \int\limits_0^\infty K^2\left(\frac{g}{h_n}\right) dg = \frac{1}{h_n} \int\limits_0^\infty K^2\,(u)\, du, \tag{33}$$

и (12) выполнено, если

$$\int\limits_0^\infty K^2\,(u)\, du < \infty. \tag{34}$$

Если при этом $nh_n \to \infty$ при $n \to \infty$, то выполнены условия (14). Подведем итоги проведенных рассуждений в виде теорем.

Теорема 3. *Пусть выполнены условия $(VIII')$, (27), (31), (34), условия (II) и (IV) статьи [2], причем $h_n \to 0$ и $nh_n \to \infty$ при $n \to \infty$. Тогда справедливы заключения теорем 1 и 2.*

Определение 1. Пусть X и Y – пространства с мерами близости ρ_1 и ρ_2 соответственно. Функция $f : X \to Y$ называется равномерно непрерывной в X в смысле мер близости ρ_1 и ρ_2, если для любого $\varepsilon > 0$ найдется $\delta > 0$ такое, что для любых x, x' таких, что $\rho_1(x, x') < \delta$, имеем $\rho_2(f(x), f(x')) < \varepsilon$.

Теорема 4. *Пусть выполнены условия теоремы 3 и, кроме того, плотность f равномерно непрерывна в Z в смысле меры близости d и евклидова расстояния в R^1 (см. определение 1). Тогда*

$$\lim_{n\to\infty} \sup_{x\in Z} |\mathbf{M} f_n(x) - f(x)| = 0. \tag{35}$$

Все условия теоремы 3 выполнены равномерно по Z, за исключением условия (IV) статьи [2]. Это условие используется при выборе окрестности точки x в доказательстве теоремы 1 той же статьи (см. формулу (10) статьи [2]). Равномерная непрерывность позволяет использовать окрестность одного и того же радиуса для всех $x \in Z$, а это обеспечивает равномерность всех оценок, т.е. (35).

5. Естественные меры близости

Определение 2. Мера близости ρ в пространстве X называется естественной, если из $\rho(x_n, x) \to 0$ и $\rho(y_n, x) \to 0$ вытекает, что $\rho(x_n, y_n) \to 0$, а из $\rho(x_n, x) \to 0$ и $\rho(x_n, y_n) \to 0$ вытекает, что $\rho(y_n, x) \to 0$ при $n \to \infty$ для любых $x, x_n, y_n \in X$, $\quad n = 1, 2, ...$

Лемма 1. *Пусть X счетно-компактно [10], топология в X порождена естественной мерой близости ρ_1, в пространстве Y задана естественная мера близости ρ_2, функция $f : X \to Y$ непрерывна. Тогда функция f равномерно непрерывна.*

Доказательство. Предположим, что отображение f не является равномерно непрерывным. Это означает, что для некоторого $\varepsilon > 0$ и каждого натурального n найдутся в X такие точки x_{1n} и x_{2n}, что

$$\rho_1(x_{1n}, x_{2n}) < \frac{1}{n}, \quad \rho_2(f(x_{1n}), f(x_{2n})) \geq \varepsilon. \tag{36}$$

Бесконечное подмножество $\{x_{1n}, n = 1, 2, ...\}$ имеет хотя бы одну предельную точку. Значит, существуют точка $x_0 \in X$ и последовательность

$n(k)$, $\quad k = 1, 2, ...$, такие, что $x_{1n(k)} \to x_0$ при $k \to \infty$. Тогда в силу первого неравенства в (36) и определения естественной меры близости имеем $x_{2n(k)} \to x_0$. Из определения непрерывности f следует, что при $k \to \infty$

$$\rho_2(f(x_{1n(k)}), f(x_0)) \to 0, \quad \rho_2(f(x_{2n(k)}), f(x_0)) \to 0, \tag{37}$$

а это в силу определения естественной меры близости несовместимо со вторым неравенством в (36). Лемма 1 доказана.

Теорема 5. *Пусть выполнены условия теоремы 3, пространство Z счетно-компактно, мера близости d является естественной (определение 2). Тогда справедливо соотношение (35).*

Для доказательства достаточно сослаться на теорему 4 и лемму 1.

Замечание. В теоремах 4 и 5 условие (VIII$'$) предполагается выполненным при всех $x \in Z$. Требовать, чтобы $nh_n \to \infty$ и было справедливо (34), нет необходимости.

6. Асимптотическое поведение дисперсий ядерных оценок плотности

Теорема 6. *Пусть выполнены условия теоремы 3. Тогда*

$$\lim_{n\to\infty} nh_n \mathbf{D} f_n(x) = f(x) \int_0^\infty K^2(u)\, du. \tag{38}$$

Доказательство проводится по той же схеме, что и доказательство теоремы 1 в статье [2]. Согласно формуле (17) статьи [2] и (28) имеем

$$\mathbf{D} f_n(x) = \frac{1}{nh_n^2}\left\{ \int_Z K^2\left(\frac{d(x,y)}{h_n}\right) f(y)p(dy) - \left[\int_Z K\left(\frac{d(x,y)}{h_n}\right) f(y)p(dy)\right]^2 \right\}. \tag{39}$$

Возьмем $\varepsilon > 0$. По условию (IV) статьи [2] найдется $\delta > 0$ такое, что

$$|f(x) - f(y)| < \varepsilon \tag{40}$$

для всех $y \in Z$ таких, что $d(x,y) < \delta$. Имеем для $i = 1, 2$

$$\mathbf{M} K^i\left(\frac{d(x,X)}{h_n}\right) = \int_{L_\delta(x)} K^i\left(\frac{d(x,y)}{h_n}\right) f(y)p(dy) +$$

$$+ \int_{Z\setminus L_\delta(x)} K^i\left(\frac{d(x,y)}{h_n}\right) f(y)p(dy). \tag{41}$$

Второе слагаемое в (41) не превосходит в силу условий (II) статьи [2] и (VIII′)

$$c_2 \int\limits_{\delta}^{\infty} \left| K^i \left(\frac{t}{h_n} \right) \right| dt = c_2 h_n \int\limits_{\delta/h_n}^{\infty} \left| K^i (t) \right| dt. \tag{42}$$

Рассмотрим первое слагаемое в формуле (41). В силу (40)

$$\begin{gathered} \left| \int\limits_{L_\delta(x)} K^i \left(\tfrac{d(x,y)}{h_n} \right) f(y) p(dy) - f(x) \int\limits_{L_\delta(x)} K^i \left(\tfrac{d(x,y)}{h_n} \right) p(dy) \right| \leq \\ \leq \varepsilon \int\limits_{L_\delta(x)} \left| K^i \left(\tfrac{d(x,y)}{h_n} \right) \right| p(dy) \leq \varepsilon \int\limits_0^\infty \left| K^i \left(\tfrac{t}{h_n} \right) \right| dt = \varepsilon h_n \int\limits_0^\infty \left| K^i (u) \right| du. \end{gathered} \tag{43}$$

Кроме того,

$$\begin{gathered} \int\limits_{L_\delta(x)} K^i \left(\tfrac{d(x,y)}{h_n} \right) p(dy) = \int\limits_0^\delta K^i \left(\tfrac{t}{h_n} \right) dt = \\ = h_n \int\limits_0^{\delta/h_n} K^i (u)\, du = h_n \left(\int\limits_0^\infty K^i (u)\, du - \int\limits_{\delta/h_n}^\infty K^i (u)\, du \right). \end{gathered} \tag{44}$$

Из (43) и (44) вытекает, что

$$\begin{gathered} \left| \int\limits_{L_\delta(x)} K^i \left(\tfrac{d(x,y)}{h_n} \right) f(y) p(dy) - f(x) h_n \int\limits_0^\infty K^i (u)\, du \right| \leq \\ \leq \varepsilon h_n \int\limits_0^\infty \left| K^i (u) \right| du + h_n f(x) \int\limits_{\delta/h_n}^\infty \left| K^i (u) \right| du. \end{gathered} \tag{45}$$

Из (41), (42) и (45) следует, что для $i = 1, 2$ справедливы оценки

$$\begin{gathered} \left| \mathbf{M} K^i \left(\tfrac{d(x,X)}{h_n} \right) - h_n f(x) \int\limits_0^\infty K^i(u) du \right| \leq \\ \leq h_n \left\{ \varepsilon \int\limits_0^\infty \left| K^i(u) \right| du + (c_2 + f(x)) \int\limits_{\delta/h_n}^\infty \left| K^i(u) \right| du \right\} = h_n \beta_i. \end{gathered} \tag{46}$$

(Заметим, что из неравенства (46) вытекает существование рассматриваемых математических ожиданий). С помощью (39), (46) и (27) получаем, что

$$\mathbf{D} f_n(x) =$$

$$= \frac{1}{n h_n^2} \left\{ h_n f(x) \int\limits_0^\infty K^2(u) du + h_n \beta(2) - \left[h_n f(x) \int\limits_0^\infty K(u) du + h_n \beta(1) \right]^2 \right\} =$$

$$= \frac{f(x)}{nh_n}\int\limits_0^\infty K^2(u)du + \frac{\beta(2)}{nh_n} - \frac{f^2(x)+2\beta(1)f(x)+\beta^2(1)}{n}, \tag{47}$$

$$|\beta(i)| \le \beta_i, i = 1, 2.$$

Выбирая $\varepsilon > 0$ достаточно малым, а затем n достаточно большим, можно в силу (46), (27), (31) и (34) сделать β_1 и β_2 сколь угодно малыми, откуда и следует (38). Теорема 6 доказана.

Теорема 7. *Пусть выполнены условия теоремы 6 и, кроме того, плотность f равномерно непрерывна. Тогда*

$$\lim_{n\to\infty}\left\{nh_n \sup_{x\in Z}\left|\mathbf{D}f_n(x) - \frac{f(x)}{nh_n}\int\limits_0^\infty K^2(u)\,du\right|\right\} = 0. \tag{48}$$

Следствие. В условиях теоремы 7

$$\lim_{n\to\infty}\sup_{x\in Z}\mathbf{D}f_n(x) = 0. \tag{49}$$

Теорема 8. *Пусть выполнены условия теоремы 6, пространство Z счетно-компактно, мера близости d является естественной. Тогда справедливы соотношения (48) и (49).*

Доказательства теорем 7 и 8 аналогичны доказательствам теорем 4 и 5. Условие (VIII$'$) предполагается выполненным при всех $x \in Z$.

7. Ядерные оценки плотности в пространстве интегрируемых с квадратом функций

Пример 3. Пусть пространство Z – пространство интегрируемых с квадратом функций $g : [0,1] \to R^1$ таких, что $g(0) = g(1) = 0$. Пусть мера p порождается броуновским мостом – гауссовским процессом $\xi(t)$ на $[0,1]$ таким, что

$$\mathbf{M}\xi(t) = 0, \quad \mathbf{M}\xi(t)\xi(s) = \min(s,t) - st, \tag{50}$$

а мера q порождается процессом $\xi(t) + a(t)$, где $a(t)$ – детерминированная функция вида

$$a(t) = \sum_{1\le i\le m} a_i\phi_i(t). \tag{51}$$

В (51) $\phi_k(t)$ – собственные функции ядра $\min(s,t) - st$, т.е. функции (см. [11, с.248])

$$\phi_k(t) = \sqrt{2}\sin(k\pi t), \quad k = 1, 2, ..., \tag{52}$$

соответствующие собственным числам

$$\lambda_k = \frac{1}{k^2\pi^2}, \quad k = 1, 2, \dots . \tag{53}$$

Согласно теореме 2 из [12, стр.115] плотность q по p существует и имеет вид
$$f(x) = \frac{dq}{dp}(x) = \exp\left\{(b, A^{-1/2}x) - \frac{1}{2}(b,b)\right\}, \tag{54}$$
где A – корреляционный оператор процесса $\xi(t)$, $\quad b = A^{-1/2}a$, т.е.
$$b = \sum_{1\le k\le m} a_k\phi_k(t)\frac{1}{\sqrt{\lambda_k}} = \sum_{1\le k\le m} a_k k\pi\phi_k(t). \tag{55}$$
Далее,
$$(b, A^{-1/2}x) = \sum_{1\le k\le m} (x,\phi_k)a_k k^2\pi^2, \tag{56}$$
где
$$(x,\phi_k) = \int_0^1 x(t)\sqrt{2}\sin(k\pi t)dt. \tag{57}$$
Значит,
$$(b, A^{-1/2}x) = \int_0^1 x(t)\phi(t)dt, \tag{58}$$
где
$$\phi(t) = \sum_{1\le k\le m} a_k k^2\pi^2\sqrt{2}\sin(k\pi t). \tag{59}$$
Поскольку
$$(b,b) = \sum_{1\le k\le m} a_k^2 k^2\pi^2 = 2C \tag{60}$$
при некотором C, то согласно (54), (58) и (60)
$$f(x) = \exp\left\{\int_0^1 x(t)\phi(t)dt - C\right\}. \tag{61}$$

Рассмотрим ядерные оценки плотности. Для плотности $f(x)$, заданной формулой (61), условие (II) статьи [2] не выполнено – плотность не является ограниченной. Поэтому рассмотрим финитные ядра $K(u)$ (см. теорему 2). В качестве меры близости рассмотрим
$$\rho_2(x,y) = \sup_{0\le t\le 1}|x(t) - y(t)|. \tag{62}$$
Ясно, что в топологии, порожденной ρ_2, плотность $f(x)$ непрерывна. Имеем:
$$F_x(q) = P\left\{\sup_{0\le t\le 1}|\xi(t) - x(t)| < q\right\}. \tag{63}$$

В частности, при $x = x(t) \equiv 0$

$$F_x(q) = F_{(1)}(q) = \sum_{-\infty<k<\infty} (-1)^k e^{-2k^2q^2} - \tag{64}$$

– функция распределения Колмогорова. Значит, можно сделать преобразование (17), перейдя к предпочтительному показателю различия.

Если в качестве меры близости взять

$$\rho_3(x,y) = \int\limits_0^1 (x(t) - y(t))^2 dt, \tag{65}$$

то в топологии, порожденной ρ_3, плотность $f(x)$ непрерывна (в силу неравенства Коши – Буняковского – Шварца). Тогда

$$F_x(q) = P\left\{ \int\limits_0^1 (\xi(t) - x(t))^2 dt < q \right\}. \tag{66}$$

При $x = x(t) \equiv 0$ имеем согласно [13, 14]

$$F_x(q) = F_{(2)}(q) = 1 + \frac{2}{\pi} \sum_{k=1}^{\infty} (-1)^k \int\limits_{(2k-1)\pi}^{2k\pi} \frac{e^{-\lambda^2 q/2}}{\sqrt{-\lambda \sin \lambda}} d\lambda, \tag{67}$$

т.е. $F_x(q)$ есть функция распределения, асимптотическая для функции распределения статистики Крамера – Мизеса – Смирнова ω_n^2.

В обоих случаях справедливы приведенные выше теоремы об асимптотическом поведении непараметрических оценок плотности. Естественно, что разные способы оценивания приводят к одному и тому же результату – значению плотности в рассматриваемой точке.

Рассмотрим подробнее оценивание плотности в точке $x = x(t) \equiv 0$. Тогда для предпочтительного показателя различия (см. (17)), построенного исходя из ρ_2 (см. (62)), имеем оценку значения плотности

$$f_{1n}(0) = \frac{1}{nh_n} \sum_{1 \le i \le n} K\left(\frac{F_{(1)}\left(\sup\limits_{0 \le t \le 1} |X_i(t)| \right)}{h_n} \right), \tag{68}$$

а для предпочтительного показателя различия (см. (17)), построенного исходя из ρ_3 (см. (65)), имеем оценку значения плотности

$$f_{2n}(0) = \frac{1}{nh_n} \sum_{1 \le i \le n} K\left(\frac{F_{(2)}\left(\int_0^1 X_i^2(t) dt \right)}{h_n} \right), \tag{69}$$

где $X_1(t), X_2(t), ..., X_n(t)$ – независимые реализации процесса $\xi(t) + a(t)$. Если ядро K удовлетворяет условиям (27), (31), (34) и финитно, $h_n \to 0$ и $nh_n \to \infty$ при $n \to \infty$, то согласно теореме 3

$$f_{1n}(0) \to \exp(-C), \quad f_{2n}(0) \to \exp(-C) \tag{70}$$

при $n \to \infty$, где C определено в (60).

Пример 3 показывает, что развиваемая в настоящей статье теория применима, в частности, к функциональным пространствам Z и позволяет получать отдельные новые результаты для случайных процессов. Однако эта теория развита для нужд статистики объектов нечисловой природы (статистики нечисловых данных, нечисловой статистики), в которой основной интерес представляют конечные пространства Z. Для них приведенные выше результаты нельзя применять непосредственно, поскольку не выполнено условие (VIII$'$), функция $F_x(t)$ – функция дискретного распределения (а не непрерывного), а потому "не проходят" доказательства теорем $3-8$. Нами развита теория, охватывающая случай конечных пространств Z. Она будет изложена в дальнейших публикациях.

Библиографический список

1. *Орлов А.И.* Организационно-экономическое моделирование: учебник : в 3 ч. Часть 1: Нечисловая статистика. – М.: Изд-во МГТУ им. Н.Э. Баумана. 2009. – 541 с.

2. *Орлов А.И.* Оценки плотности в пространствах произвольной природы // Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. / Перм. гос. нац. иссл. ун-т. – Пермь, 2013. – Вып. 25. – С.21–33.

3. Вероятность и математическая статистика: Энциклопедия / Гл. ред. Ю.В. Прохоров. – М.: Большая Российская Энциклопедия, 1999. – 910 с.

4. *Орлов А.И.* Статистика объектов нечисловой природы и экспертные оценки // Экспертные оценки / Вопросы кибернетики. Вып.58. – М.: Научный Совет АН СССР по комплексной проблеме "Кибернетика 1979. – С.17-33.

5. *Орлов А.И.* Статистика объектов нечисловой природы // Теория вероятностей и ее применения. – 1980. – Т.XXV. – № 3. – С.655-656.

6. *Орлов А.И.* Непараметрические оценки плотности в топологических пространствах // Прикладная статистика. Ученые записки по статистике, т.45. – М.: Наука, 1983. – С. 12-40.

7. *Rosenblatt M.* Remarks on some nonparametric estimates of a density function // Ann. Math. Statist. – 1956. – V.27. – № 5. – P.832 – 837.

8. *Parzen E.* On estimation of a probability density function and mode // Ann. Math. Statist. – 1962. – V.33. – № 6. – P. 1065-1076.

9. *Орлов А.И.* Ядерные оценки плотности в пространствах произвольной природы // Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. / Перм. гос. нац. иссл. ун-т. – Пермь, 1996. – Вып. 11. – С.68-75.

10. *Колмогоров А.Н., Фомин С.В.* Элементы теории функций и функционального анализа. – М.: Наука, 1972. – 496 с.

11. *Гихман И.И., Скороход А.В.* Введение в теорию случайных процессов. – М.Наука, 1965. – 656 с.

12. *Скороход А.В.* Интегрирование в гильбертовом пространстве. – М.: Наука, 1975. – 232 с.

13. *Смирнов Н.В.* О распределении ω^2 – критерия Мизеса // Математический сборник. – 1937. – Т.2. – Вып.1. – С.973 – 993.

14. *Орлов А.И.* Скорость сходимости распределения статистики Мизеса – Смирнова // Теория вероятностей и ее применения. – 1974. – Т.19. – Вып.4. – С.766 – 786.

Nuclear density estimates in spaces of arbitrary nature

A.I. Orlov[1]

[1] Bauman Moscow State Technical University, Institute of high statistical technologies and econometrics; Moscow Physics-Technical Institute; Moscow, Russia;
prof-orlov@mail.ru; http://orlovs.pp.ru; +7(916)8305117; 123104, Moscow, the Sytinsky lane, house 7/14, apartment 14.

Abstract. *The study of linear estimates of the density of probability distribution in spaces of arbitrary nature is initiated by the author in the previous vyp.25. A special case of such estimates - nuclear density estimates. This article covers the basics of the general theory of kernel density estimates. Under different conditions are proved the consistency and asymptotic normality of kernel density estimates, is studied uniform convergence. Is found the asymptotic behavior of the dispersion of nuclear density estimates. Examples are considered, including in the space of square-integrable functions.*

Key words: *non-numeric statistics, probability density function, the space of an arbitrary nature, linear density estimates, nuclear density estimates, limit theorems, consistent estimates, asymptotic normality, uniform convergence.*