УДК 519.2

Ядерные оценки плотности в конечных пространствах

А.И. Орлов

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

Аннотация. *Ядерные оценки плотности рассмотрены для последовательностей пространств с мерами. Найдены условия, при которых разность плотностей распределений вероятностей и математических ожиданий их ядерных оценок равномерно стремится к 0. Установлена равномерная сходимость для дисперсий. Выявлены условия на ядерные функции, при которых имеют место указанные равномерные сходимости. В качестве примеров рассмотрены пространства нечетких подмножеств конечных множеств и пространства всех подмножеств конечных множеств. Найдено условие, обеспечивающее возможность применения ядерных оценок плотности в конечных пространствах. Приведен контрпример пространства ранжировок, в котором применение ядерных оценок плотности нельзя признать корректным.*
Ключевые слова: *нечисловая статистика, плотность распределения вероятностей, пространства произвольной природы, ядерные оценки плотности, предельные теоремы, состоятельные оценки, конечные пространства, равномерная сходимость.*

1. Введение

Непараметрические оценки плотности распределения вероятностей в пространствах произвольной природы – один из основных инструментов нечисловой статистики [1, 2], называемой также статистикой объектов нечисловой природы или статистикой нечисловых данных. Систематическое изложение теории таких оценок начато в статьях [3, 4], непосредственным продолжением которых является настоящая статья. Регулярно используются ссылки на условия и утверждения из статей [3, 4].

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

$$q(C) = \int_C f(x)dp$$

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 [5, с.460].

Пусть $X_1, X_2, \ldots, X_n$ – независимые одинаково распределенные случайные элементы (величины), распределение которых задается вероятностной мерой q. В статье [3] введено несколько видов непараметрических оценок плотности вероятности q по выборке $X_1, X_2, \ldots, X_n$. Подробнее изучены линейные оценки. В статье [4] рассмотрены их частные случаи – ядерные оценки плотности в пространствах произвольной природы. Однако асимптотическая теория оценок плотности развита прежде всего для нужд статистики конкретных видов объектов нечисловой природы, в которой основной интерес представляют конечные пространства Z. Мера p при этом не непрерывная, а дискретная, например, считающая. *Таким образом, в рамках единого подхода рассматриваем оценки плотностей и оценки вероятностей.*

Для конечных пространств Z полученные в [3, 4] результаты нельзя применять непосредственно, поскольку, в частности, не выполнено условие (*VIII'*) статьи [4], функция $F_x(t)$ – функция дискретного распределения (а не непрерывного), а потому "не проходят" приведенные в [4] доказательства теорем 3 - 8. Нами развита теория, охватывающая случай конечных пространств Z(пространств бинарных отношений, подмножеств конечных множеств и др.). Этой теории посвящена настоящая статья.

2. Последовательности моделей оценивания плотностей

Будем изучать асимптотику последовательностей пространств с мерами (с целью в дальнейшем рассмотреть последовательности конечных пространств). Введем новый параметр m и рассмотрим последовательность пространств с мерами (Z_m, p_m) и соответствующих функций $F_m(x,t)$, задающих зависимость мер шаров с центром в точке x из Z_m от радиуса t,

$$F_m(x,t) = p_m\{y : d_m(x,y) < t\}, \ \ m = 1, 2, ..., \tag{1}$$

где d_m - мера близости в Z_m(здесь мы несколько модернизируем обозначения, использованные в формулах (6) и (7) статьи [4]). Предположим, что

$$\lim_{m\to\infty} F_m(x,t) = F_x(t) = \begin{cases} t, \ 0 \le t \le C \le \infty, \\ C, \ t \ge C. \end{cases} \tag{2}$$

Рассмотрим также плотности f_m в пространствах (Z_m, p_m), задающие непараметрические оценки плотности с ядрами K_m, $m = 1$, 2, ... Укажем условия, при которых полученные в статьях [3, 4] результаты оказываются асимптотически (при $m \to \infty$ и $n \to \infty$) справедливыми для последовательности вероятностных моделей оценивания плотности, задаваемых кортежами $(Z_m, p_m, d_m, f_m, K_m)$.

Поскольку вместо одной плотности f появляется последовательность плотностей f_m, то условия на плотность, в частности, условие (IV) статьи [3], необходимо изменить. Пусть выполнено следующее условие.

(IX) Для любого $\varepsilon > 0$ существует $\delta > 0$ такое, что $|f_m(x) - f_m(y)| < \varepsilon$, если $d_m(x, y) < \delta$.

Для упрощения рассуждений наложим на ядра K_m условие равномерной финитности и ограниченности.

(X) Существуют константы D и E такие, что $|K_m(t)| < D$ при всех $t \geq 0$ и $K_m(t) = 0$ при $T > E$.

Рассмотрим ядерные оценки $f_{nm}(x)$ плотностей $f_m(x)$, которые будем изучать,

$$f_{nm}(x) = \frac{1}{nb_m(h_n, x)} \sum_{1 \leq i \leq n} K_m\left(\frac{d_m(x, X_i)}{h_n}\right), \qquad K_m : [0, +\infty] \to R^1 \quad (3)$$

где $K_m = K_m(u)$ – ядра (ядерные функции), h_n – последовательность положительных чисел (показателей размытости), $b_m(h_n, x)$ – нормировочные множители. Согласно формулам (1) и (4) статьи [3]

$$b_m(h_n, x) = \int\limits_{Z_m} K_m\left(\frac{d_m(x, y)}{h_n}\right) p_m(dy) = \int\limits_0^\infty K_m\left(\frac{t}{h_n}\right) dF_m(x, t). \quad (4)$$

Повторим проведенные ранее в статьях [3, 4] рассуждения, отмечая новые моменты, связанные с введением параметра m.

Согласно условию (X) в определении $f_{nm}(x)$ участвуют лишь те элементы выборки X_i, для которых

$$d_m(x, X_i) \leq Eh_n. \quad (5)$$

Правая часть неравенства (5) задает радиус окрестности $U(x)$ точки x пространства Z_m (в смысле меры близости d_m), рассмотрением которого достаточно ограничиться.

Примем, как и в [3, 4], что при $n \to \infty$ показатель размытости

$$h_n \to \infty. \quad (6)$$

Тогда радиус рассматриваемой окрестности стремится к 0. В силу (2) при $0 \leq t \leq Eh_n \leq C$ предельная функция имеет вид $F_x(t) = t$. Положим

$$F_m(x, t) = F_x(t) + H_m(x, t). \quad (7)$$

Тогда

$$\int_0^{Eh_n} g\left(\frac{t}{h_n}\right) dF_m(x,t) =$$

$$= \int_0^{Eh_n} g\left(\frac{t}{h_n}\right) dF_x(t) + \int_0^{Eh_n} g\left(\frac{t}{h_n}\right) dH_m(x,t) = \int_0^{E} g(u)du + \alpha_{mn}(g). \quad (8)$$

Нам понадобится соотношение (8) для $g = K_m$, $g = |K_m|$, $g = |K_m|^2$, а качество аппроксимации будет определяться скоростью сходимости $\alpha_{mn}(g)$ к 0 при $m \to \infty$, $n \to \infty$.

Имеем согласно (4), условию (X) и (8):

$$b_m(h_n,x) = \int_0^{Eh_n} K_m\left(\frac{t}{h_n}\right) dF_m(x,t) = h_n \int_0^{E} K_m(u)du + \alpha_{mn}(K_m). \quad (9)$$

Примем

$$\int_0^{E} K_m(u)du = 1, \quad \alpha_{mn}(K_m) = o(h_n). \quad (10)$$

Аналогично (9) имеем

$$\int_0^{Eh_n} \left|K_m\left(\frac{t}{h_n}\right)\right| dF_m(x,t) = h_n \int_0^{E} |K_m(u)|\, du + \alpha_{mn}(|K_m|). \quad (11)$$

Для справедливости формулы (9) статьи [4] и условия (V) статьи [3] достаточно, чтобы

$$\alpha_{mn}(|K_m|) = o(h_n), \quad (12)$$

поскольку согласно условию (X)

$$\int_0^{E} |K_m(u)|\, du \le D \int_0^{E} du = DE. \quad (13)$$

Теорема 1. *Если выполнены условия $(IX), (X)$ и справедливы соотношения (6), (10), (12), то разность математических ожиданий оценок $Mf_{nm}(x)$ и плотностей $f_m(x)$ равномерно стремится к 0 при $m \to \infty$, $n \to \infty$:*

$$\lim_{\substack{n\to\infty \\ m\to\infty}} \sup_{x\in Z_m} |Mf_{mn}(x) - f_m(x)| = 0. \quad (14)$$

Доказательство. Возьмем $\varepsilon > 0$ и согласно условию (IX) выберем $\delta > 0$, обладающее указанным в этом условии свойством. Пусть $Eh_n < \delta$ при $n > n_0$. Согласно условию (X) и соотношению (11) статьи [3]

$$Mf_{nm}(x) - f_m(x) = \int\limits_{U(x)} g_{nm}(x,y)(f_m(y) - f_m(x))p_m(dy). \tag{15}$$

где функции $g_{nm}(x)$ определены формулой (4) статьи [3]. При $n > n_0$ согласно соотношению (12) статьи [3], формулам (9) и (11)

$$|Mf_{nm}(x) - f_m(x)| \leq \varepsilon \int\limits_{U(x)} |g_{nm}(x,y)|p_m(dy) =$$

$$= \varepsilon(h_n + \alpha_{mn}(K_m))^{-1} \left\{ h_n \int\limits_0^E |K(u)|du + \alpha_{mn}(|K_m|) \right\}. \tag{16}$$

Из (10) и (12) следует, что правая часть (16) не превосходит $F\varepsilon$, где $F = const$, равномерно по всем x из Z_m и $n > n_0$, откуда и вытекает (14). Теорема 1 доказана.

Согласно соотношению (12) статьи [4] для существования дисперсии у оценки $f_{nm}(x)$ достаточно справедливости условия

$$A_n = \frac{1}{b_m^2(h_n,x)} \int\limits_0^\infty K_m^2\left(\frac{t}{h_n}\right) dF_m(x,t) < \infty. \tag{17}$$

Учитывая (8) - (10), получаем

$$A_n = \frac{1}{(h_n + \alpha_{nm}(K_m))^2} \left(h_n \int\limits_0^E K_m^2(u)du + \alpha_{mn}(K_m^2) \right). \tag{18}$$

Теорема 2. *Пусть дополнительно к условиям теоремы 1*

$$\alpha_{mn}(K_m^2) = o(h_n), \quad nh_n \to \infty. \tag{19}$$

Тогда

$$\lim_{\substack{n\to\infty \\ m\to\infty}} \sup_{x\in Z_m} \left| nh_n Df_{nm}(x) - f_m(x) \int\limits_0^E K_m^2(u)du \right| = 0. \tag{20}$$

Доказательство проводится аналогично доказательству теорем 6 и 7 статьи [4].

Как видно из проведенных выше рассуждений, при рассмотрении последовательностей моделей оценивания плотностей в пространствах с мерой принципиально новыми являются условия

$$\int_0^{Eh_n} g_m\left(\frac{t}{h_n}\right) d(F_m(x,t)-t) = o(h_n), \qquad (21)$$

где $g_m = K_m, \quad g_m = |K_m|, \quad g_m = |K_m|^2$ (ср. (8)). С помощью замены переменных $u = t/h_n$ от (21) перейдем к условиям

$$\int_0^E g_m(u) d(F_m(x,h_n u)-h_n u) = h_n \int_0^E g_m(u) d\left(\frac{F_m(x,h_n u)}{h_n} - u\right) = o(h_n), \qquad (22)$$

т.е. к условиям

$$\lim_{\substack{n\to\infty\\ m\to\infty}} \int_0^E g_m(u) d\left(\frac{F_m(x,h_n u)}{h_n} - u\right) = 0. \qquad (23)$$

Введем условие

$$\lim_{\substack{n\to\infty\\ m\to\infty}} \frac{F_m(x,h_n u)}{h_n} = u, \quad u \in (0,E]. \qquad (24)$$

В [1, с.230] показано, что для вывода (23) из (24) необходимо и достаточно (в указанном там смысле), чтобы функции $g_m(u)$ были равностепенно (по m) интегрируемы по Риману. В частности, достаточно, чтобы они были равностепенно непрерывными.

Теорема 3. *Соотношения (10), (12), (19) выполнены, если ядерные функции K_m равностепенно непрерывны и справедливо (24).*

Требование равностепенной непрерывности связано с тем, что ядра K_m могут зависеть от параметра m. В приложениях обычно достаточно положить $K_m = K$ и вместо условия (X) принять условие (X'):

(X') Ядро $K : [0,+\infty] \to R^1$ – непрерывная финитная функция.

Рассмотрим примеры применения развитой выше теории для построения ядерных оценок плотности в конкретных дискретных пространствах.

3. Нечеткие подмножества конечных множеств

Пример 1. Рассмотрим последовательность $Z_m, m = (r,q)$, пространств нечетких множеств, являющихся подмножествами конечных множеств

$$Y_r = \left\{\frac{1}{r}, \frac{2}{r}, ..., \frac{i}{r}, ..., \frac{r-1}{r}, 1\right\}, \quad r = 1, 2, ..., \qquad (25)$$

функции принадлежности которых принимают значения из конечных множеств

$$W_q = \left\{0, \frac{1}{q}, \frac{2}{q}, ..., \frac{j}{q}, ..., \frac{q-1}{q}, 1\right\}, \quad q = 1, 2, ... \tag{26}$$

Очевидно, число элементов пространства Z_m равно

$$Card(Z_m) = (q+1)^r. \tag{27}$$

Пусть f и g – функции принадлежности нечетких множеств, т.е. функции, отображающие Y_r в W_q. Естественно рассмотреть меру близости

$$d(f,g) = \sup_{y \in Y_r} |f(y) - g(y)| \tag{28}$$

и окрестности

$$L_{t/q}(f) = \left\{g : \sup_{y \in Y_r} |f(y) - g(y)| \le t/q\right\}, \quad t = 0, 1, 2, ... \quad . \tag{29}$$

Если

$$\frac{t}{q} \le \min_{y \in Y_r} f(y) \le \max_{y \in Y_r} f(y) \le 1 - \frac{t}{q}, \tag{30}$$

то, как нетрудно видеть, число элементов в рассматриваемой окрестности равно

$$Card\left(L_{t/q}(f)\right) = (2t+1)^r. \tag{31}$$

Пусть p_m – вероятностная мера на Z_m, приписывающая всем элементам Z_m одну и ту же вероятность. Тогда при справедливости (30)

$$p_m\left(L_{t/q}(f)\right) = \frac{Card\left(L_{t/q}(f)\right)}{(q+1)^r} = \left(\frac{2t+1}{q+1}\right)^r = \left[\frac{2\left(\frac{t}{q}\right) + \frac{1}{q}}{1 + \frac{1}{q}}\right]^r. \tag{32}$$

Положим

$$d_1(f,g) = \left[2 \sup_{y \in Y_r} |f(y) - g(y)| + \frac{1}{q}\right]^r \left[1 + \frac{1}{q}\right]^{-r}. \tag{33}$$

Тогда

$$F_m(f,t) = p_m\{g : d_1(f,g) \le t\} = t \tag{34}$$

в точках, в которых

$$t^{1/r} = \frac{2s+1}{q+1} \tag{35}$$

при некотором целом неотрицательном s. Функция $F_m(f,t)$ имеет скачки в точках

$$t_s = \left(\frac{2s+1}{q+1}\right)^r, \quad s = 0, 1, 2, ..., \tag{36}$$

и кусочно постоянна между ними. Значит, при любых t, таких, что $t_{s-1} \le t \le t_s$, имеем

$$0 \le t - F_m(f,t) \le t_s - t_{s-1}. \quad (37)$$

Элементарные выкладки дают

$$t_s - t_{s-1} = \frac{2r}{q} t_s^{1-1/r}. \quad (38)$$

Следовательно,

$$\frac{F_m(f, h_n u)}{h_n} = u + O\left(\frac{2r}{q} h_n^{-1/r} u^{1-1/r}\right). \quad (39)$$

Для справедливости (24) и, следовательно, для применимости развитой выше теории ядерных оценок плотности достаточно, чтобы двумерный параметр $m = (r, q)$ был связан с объемом выборки n таким образом, чтобы

$$\frac{r}{q} h_n^{-1/r} \to 0. \quad (40)$$

Если $r = n^\alpha, q = n^\beta$, то

$$h_n^{-1/r} \le n^{1/n^\alpha}, \quad (41)$$

и для справедливости (40) необходимо и достаточно, чтобы $\beta > \alpha$.

Ядерная оценка плотности имеет вид

$$f_{nm}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K_m \left(\frac{\left[2 \sup_{y \in Y_r} |x(y) - X_i(y)| + \frac{1}{q}\right]^r}{h_n \left(1 + \frac{1}{q}\right)^r} \right). \quad (42)$$

4. Пространства всех подмножеств конечных множеств

Пример 2. Рассмотрим последовательность Z_m пространств всех подмножеств множеств $D(m)$ из m элементов, т.е. $Z_m = 2^{D(m)}$. Пространство Z_m состоит из 2^m элементов. Пусть p_m – вероятностная мера, соответствующая равномерному распределению на Z_m. Известно [1], что из некоторой естественной системы аксиом вытекает, что в качестве расстояния между двумя множествами A, B следует использовать меру симметрической разности этих множеств:

$$d_m(A, B) = p_m(A \Delta B) = \frac{Card(A \Delta B)}{2^m}. \quad (43)$$

Вычислим $p_m\{B\colon d_m(A,B) \le t\}$. В силу изотропности пространства Z_m (см. [1]) достаточно рассмотреть случай $A = \emptyset$. Тогда условие $Card(A\Delta B) \le s$ переходит в условие $Card(B) \le s$, а потому

$$p_m\left\{B : d_m(A,B) \le \frac{s}{2^m}\right\} = \frac{\sum\limits_{0\le i\le s}\binom{m}{i}}{2^m}. \tag{44}$$

Поскольку в силу интегральной теоремы Муавра-Лапласа

$$\lim_{m\to\infty}\sup_s\left|\frac{\sum\limits_{0\le i\le s}\binom{m}{i}}{2^m} - \Phi\left(\frac{2s-m}{\sqrt{m}}\right)\right| = 0, \tag{45}$$

то естественно рассмотреть меру близости

$$d_{1m}(A,B) = \Phi\left(\frac{2Card(A\Delta B)-m}{\sqrt{m}}\right) \tag{46}$$

и ядерную оценку плотности

$$f_{nm}(A) = \frac{1}{nh_n}\sum_{0\le i\le n} K_m\left(\frac{\Phi\left(\frac{2Card(A\Delta X_i)-m}{\sqrt{m}}\right)}{h_n}\right). \tag{47}$$

Таким образом, функция $F_m(A,t)$ кусочно постоянна, имеет скачки в точках

$$t_s = \Phi\left(\frac{2s-m}{\sqrt{m}}\right), \quad s = 0,1,2 \tag{48}$$

и в этих точках принимает значения

$$F_m(A,t_s) = 2^{-m}\sum_{0\le i\le s}\binom{m}{i}. \tag{49}$$

Следовательно,

$$F_m(A,h_n u) = P\left(X \le \frac{\sqrt{m}\Phi^{-1}(h_n u)+m}{2}\right). \tag{50}$$

где X имеет биномиальное распределение с параметрами m и $p = 0,5$. Для справедливости (24) необходимо, чтобы

$$\frac{P_m\left(\Phi^{-1}(h_n u)\right)}{h_n u} \to 1, \tag{51}$$

где
$$P_m\left(\Phi^{-1}(h_n u)\right) = P\left(\frac{2X-m}{\sqrt{m}} \le \Phi^{-1}(h_n u)\right). \tag{52}$$
Обозначим
$$c_n = c_n(u) = \Phi^{-1}(h_n u). \tag{53}$$
Тогда $c_n \to -\infty$ при $n \to \infty$, и (51) переходит в условие
$$\frac{P_m(c_n)}{\Phi(c_n)} \to 1. \tag{54}$$
По известной теореме [6, с.207] для справедливости (54) достаточно, чтобы
$$\frac{c_n^3}{\sqrt{m}} \to 0. \tag{55}$$
Согласно известной формуле [, с.11] об асимптотике нормального распределения
$$c_n \sim \sqrt{-2(\ln h_n + \ln u)}. \tag{56}$$
Следовательно, если
$$m \ge G[\ln h_n]^{6+\gamma}, G > 0, \gamma > 0, \tag{57}$$
то выполнены соотношения (55), (54), а потому (51) и (24), и использование ядерной оценки плотности (47) корректно.

5. Широкая распространенность условия (24)

В примерах 1 и 2 использовались свойства конкретных дискретных пространств нечисловой природы. Продемонстрируем, что справедливость условия (24) широко распространена, а не является исключением.

Пусть Z_m – последовательность конечных пространств нечисловой природы,
$$p_m(A) = \frac{Card\{A\}}{Card\{Z_m\}}, \quad A \subseteq Z_m, \tag{58}$$
и d_m - расстояние (мера близости) в Z_m. Тогда
$$p_m(L_t(x)) = \frac{Card\{y : d_m(y,x) \le t\}}{Card\{Z_m\}} = F_m^1(x,t) \tag{59}$$
Как обычно, положим
$$d_{1m}(y,x) = F_m^1(x, d_m(y,x)) \tag{60}$$
и рассмотрим
$$F_m(x,t) = p_m\{y : d_{1m}(y,x) < t\}. \tag{61}$$

Тогда $F_m(x,t)$ кусочно постоянна и имеет скачки в точках $t_i, i = 1, 2, ...,$ причем

$$F_m(x, t_i) = t_i. \tag{62}$$

Кроме того, при $t_{i-1} < t < t_i$ имеем

$$0 < t - F_m(x,t) < t_i - t_{i-1}. \tag{63}$$

Чтобы можно было приблизить дискретную модель непрерывной, необходимо потребовать, чтобы при $m \to \infty$

$$\max(t_i - t_{i-1}) \to 0. \tag{64}$$

Из (64) согласно (63) следует, что

$$\lim_{m\to\infty} F_m(x,t) = t \tag{65}$$

для любого $t > 0$.

Если зафиксировать n, то (24) эквивалентно условию

$$\lim_{m\to\infty} \frac{F_m(x,t)}{t} = 1, \quad t > 0. \tag{66}$$

Пусть выполнено (65). Покажем, что можно построить последовательность $m_n = m(h_n)$, $n = 1$, 2, ..., так, чтобы было справедливо соотношение (24).

Рассмотрим убывающую u_n и возрастающую m_n последовательности такие, что

$$h_n u_n = t_n, \quad u_n \to 0, \quad \sup_{t\geq t_n} \left| \frac{F_m(x,t)}{t} - 1 \right| < \frac{1}{n}, \quad m \geq m_n. \tag{67}$$

Покажем, что удовлетворяющая (67) последовательность m_n существует. Как известно, их поточечной сходимости (65) вытекает равномерная сходимость функций распределения:

$$\sup_t |F_m(x,t) - t| < \varepsilon \tag{68}$$

при $m > m(\varepsilon)$. Следовательно, при любом t

$$\left| \frac{F_m(x,t)}{t} - 1 \right| < \frac{\varepsilon}{t}. \tag{69}$$

Поскольку $t \geq t_n$, то из (69) следует, что

$$\sup_{t\geq t_n} \left| \frac{F_m(x,t)}{t} - 1 \right| < \frac{\varepsilon}{t_n}. \tag{70}$$

Следовательно, положив $\varepsilon = t_n/n$, получим, что при $m_n \geq m\,(t_n/n)$ справедливо (67).

Рассмотрим последовательность $q_n = h_n u$, $n = 1, 2, \ldots$, Тогда при $n > n_0$ имеем $u_n < u$. В соответствии с (67)

$$\left|\frac{F_m(x, h_n u)}{h_n} - u\right| < \frac{u}{n} \tag{71}$$

при $m > m_n$. Следовательно, если при $n \to \infty$ параметр m меняется так, что $m > m_n$, то справедливо (24).

Таким образом, условие (64) оказывается достаточным для обоснования корректности применения ядерных оценок плотности.

6. Контрпример

Приведем пример пространства, в котором применение ядерных оценок плотности нельзя признать корректным.

В пространствах объектов нечисловой природы расстояния могут вводиться с помощью понятия "соседства". Некоторые пары объектов объявляют "соседями"и расстояние между ними принимают равным 1. Чтобы измерить расстояние между объектами a и b, строят всевозможные последовательности объектов из этого пространства $a_1 = a$, a_2, a_3, ..., $a_k = b$ такие, что a_i и a_{i+1} являются соседями, $i = 1, 2, \ldots, k-1$. Расстоянием $d(a, b)$ между a и b объявляют наименьшую из длин последовательностей рассматриваемого вида [8, 9].

Пример 3. В качестве примера рассмотрим пространство Z_m ранжировок (без связей) m объектов (в терминах комбинаторики – пространство перестановок m объектов). Тогда

$$Card(Z_m) = m! \tag{72}$$

"Соседями"назовем ранжировки, отличающиеся одной инверсией. Тогда у ранжировки R имеется $\binom{m}{2}$ соседей. Условию

$$d(R, X) = 2 \tag{73}$$

удовлетворяют

$$\binom{m}{2}\left[\binom{m}{2} - 1\right] \tag{74}$$

ранжировок, поскольку для каждой из $\binom{m}{2}$ ранжировок Y, удовлетворяющих условию $d(R,Y)=1$, одна из $\binom{m}{2}$ инверсий приводит опять к R, а остальные $\binom{m}{2}-1$ инверсий – к ранжировкам X, удовлетворяющих условию (73). Можно показать, что при малых (по сравнению с m) значениях t справедлива аппроксимация

$$Card(L_t(R)) \sim \binom{m}{2}^t \tag{75}$$

Пусть p_m - вероятностная мера, соответствующая равномерному распределению на Z_m. Тогда согласно (75) естественно рассмотреть ядерную оценку плотности

$$f_{nm}(R) = \frac{1}{nh_n} \sum_{1\le i\le n} K_m\left(\frac{\binom{m}{2}^{d(R,X_i)}}{h_n m!}\right). \tag{76}$$

Проверим справедливость соотношения (64). В силу (75)

$$t_i = \frac{\binom{m}{2}^i}{m!}, \quad \frac{t_i}{t_{i-1}} = \binom{m}{2}. \tag{77}$$

Значит, соотношение (64) не может быть выполнено, и полученные выше в настоящей статье результаты не применимы к статистике (76). Причина состоит в слишком быстром росте $Card(L_t(R))$ согласно (75) (ср. (77) с ростом аналогичных величин в других пространствах объектов нечисловой природы - пространстве нечетких множеств (см. (31)) и пространстве подмножеств конечного множества (см. (44)).

Таким образом, примеры 1 и 2 показывают, что результаты настоящей статьи позволяют строить ядерные оценки плотности в конкретных дискретных (конечных) пространствах нечисловой природы. В следующем разделе 5 сформулировано общее свойство рассматриваемых пространств, позволяющее обосновать корректность ядерных оценок. Пример 3 является контрпримером, демонстрирующим, что проведенные в настоящей статье рассмотрения применимы не ко всем пространствам объектов нечисловой природы.

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.* О развитии статистики объектов нечисловой природы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. №93. С. 41-50.

3. *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.

4. *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, 2015. – Вып. 26. – С. 43-57.

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

6. *Феллер В.* Введение в теорию вероятностей и ее приложения: Том 1. – М.: Мир, 1984. – 528 с.

7. *Большев Л.Н., Смирнов Н.В.* Таблицы математической статистики / 3-е изд. – М.: Наука, 1983. – 416 с.

8. *Куликов С.М.* Структурные меры близости в пространстве классификаций и разбиений // Прикладная статистика. – М.: Наука, 1983. – С. 282-286.

9. *Тюрин Ю.Н.* Экспертная классификация // Экспертные методы в системных исследованиях / Сборник трудов ВНИИСИ. – 1979. – Вып.4. – С.5-15.

Nuclear density estimates in finite spaces

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. *Kernel density estimates are considered for sequences of spaces with measures. We give the conditions under which the difference between the densities of probability distributions and of the mathematical expectations of their nuclear estimates uniformly tends to 0. Is established the uniform convergence of the variances. We find the conditions on the kernel functions, in which take place these theorems about uniform convergence. As examples, are considered the spaces of fuzzy subsets of finite sets and the spaces of all subsets of finite sets. We give the condition to support the use of kernel density estimation in finite spaces. We discuss the counterexample of space of rankings in which the application of kernel density estimators can not be correct.*

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