

УДК 519.2

UDC 519.2

01.00.00 Физико-математические науки

Physics and mathematical sciences

НЕПАРАМЕТРИЧЕСКИЕ ЯДЕРНЫЕ ОЦЕНКИ ПЛОТНОСТИ ВЕРОЯТНОСТИ В ДИСКРЕТНЫХ ПРОСТРАНСТВАХ**NONPARAMETRIC KERNEL ESTIMATORS OF PROBABILITY DENSITY IN THE DISCRETE SPACES**

Орлов Александр Иванович

д.э.н., д.т.н., к.ф.-м.н., профессор

РИНЦ SPIN-код: 4342-4994

*Московский государственный технический**университет им. Н.Э. Баумана, Россия, 105005,**Москва, 2-я Бауманская ул., 5, prof-orlov@mail.ru**Московский физико-технический институт, 141700,**Моск. обл., г. Долгопрудный, Институтский пер., 9*

Orlov Alexander Ivanovich

Dr.Sci.Econ., Dr.Sci.Tech., Cand.Phys-Math.Sci.,
professor*Bauman Moscow State Technical University,
Moscow, Russia**Moscow Physics-Technical Institute; Moscow
region, Dolgoprudny, Russia*

Оценки плотности распределения вероятностей в пространствах произвольной природы используют для решения различных задач нечисловой статистики. Систематическое изложение теории таких оценок начато в наших статьях [3, 4], непосредственным продолжением которых является настоящая статья. Регулярно используются ссылки на условия и утверждения из статей [3, 4], в которой введено несколько видов непараметрических оценок плотности вероятности по выборке. Подробнее изучены линейные оценки. В настоящей статье рассмотрим их частные случаи – ядерные оценки плотности в дискретных пространствах. При оценивании плотности числовой случайной величины ядерные оценки переходят в оценки Парзена-Розенблатта. При различных условиях доказана состоятельность и асимптотическая нормальность ядерных оценок плотности. Введено понятие "предпочтительный показатель различия" и изучены ядерные оценки плотности на его основе. Введены и изучены естественные меры близости, используемые при анализе асимптотического поведения ядерных оценок плотности. 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 сходимости 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: СТАТИСТИЧЕСКИЕ МЕТОДЫ, МАТЕМАТИЧЕСКАЯ СТАТИСТИКА,

Some estimators of the probability density function in spaces of arbitrary nature are used for various tasks in statistics of non-numerical data. Systematic exposition of the theory of such estimators has been started in our articles [3, 4]. This article is a direct continuation of these works [3, 4]. We will regularly use references to conditions and theorems of the articles [3, 4], in which introduced several types of nonparametric estimators of the probability density. We have studied linear estimators. In this article, we consider particular cases - kernel density estimates in discrete spaces. When estimating the density of the one-dimensional random variable, kernel estimators become the Parzen-Rosenblatt estimators. Under different conditions, we prove the consistency and asymptotic normality of kernel density estimators. We have introduced the concept of "preferred rate differences" and are studied nuclear density estimators based on it. We have introduced and studied natural affinity measures which are used in the analysis of the asymptotic behavior of kernel density estimators. 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, there 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

Keywords: STATISTICAL METHODS, MATHEMATICAL STATISTICS,

НЕПАРАМЕТРИЧЕСКАЯ СТАТИСТИКА,
НЕЧИСЛОВАЯ СТАТИСТИКА, ПЛОТНОСТЬ
РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ,
ДИСКРЕТНЫЕ ПРОСТРАНСТВА, ЯДЕРНЫЕ
ОЦЕНКИ ПЛОТНОСТИ, ПРЕДЕЛЬНЫЕ
ТЕОРЕМЫ, СОСТОЯТЕЛЬНЫЕ ОЦЕНКИ,
АСИМПТОТИЧЕСКАЯ НОРМАЛЬНОСТЬ,
РАВНОМЕРНАЯ СХОДИМОСТЬ

NONPARAMETRIC STATISTICS, NON-
NUMERIC STATISTICS, PROBABILITY
DENSITY FUNCTION, THE DISCRETE SPACES,
KERNEL DENSITY ESTIMATORS, LIMIT
THEOREMS, CONSISTENT ESTIMATORS,
ASYMPTOTIC NORMALITY, UNIFORM
CONVERGENCE

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

$$q(C) = \int_C f(x) dp \quad (1)$$

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 [5, с.460].

Пусть X_1, X_2, \dots, 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, \dots, 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 ,

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

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

$$\lim_{m \rightarrow \infty} F_m(x, t) = F_x(t) = \begin{cases} t, & 0 \leq t \leq C \leq \infty, \\ C, & t \geq C. \end{cases} \quad (2)$$

Рассмотрим также плотности f_m в пространствах (Z_m, p_m) , задающие непараметрические оценки плотности с ядрами K_m , $m = 1, 2, \dots$. Укажем условия, при которых полученные в статьях [3, 4] результаты оказываются асимптотически (при $m \rightarrow \infty$ и $n \rightarrow \infty$) 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,

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

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]

$$b_m(h_n, x) = \int_{Z_m} K_m \left(\frac{d_m(x, y)}{h_n} \right) p_m(dy) = \int_0^{\infty} K_m \left(\frac{t}{h_n} \right) dF_m(x, t). \quad (4)$$

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 $n \rightarrow \infty$ показатель размытости

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

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 $m \rightarrow \infty$, $n \rightarrow \infty$.

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)

$$\int_0^E |K_m(u)| du \leq D \int_0^E du = DE. \quad (13)$$

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 $m \rightarrow \infty$, $n \rightarrow \infty$:

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

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]

$$Mf_{nm}(x) - f_m(x) = \int_{U(x)} g_{nm}(x, y)(f_m(y) - f_m(x))p_m(dy), \quad (15)$$

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

$$\begin{aligned} |Mf_{nm}(x) - f_m(x)| &\leq \varepsilon \int_{U(x)} |g_{nm}(x, y)| p_m(dy) = \\ &= \varepsilon (h_n + \alpha_{mn}(K_m))^{-1} \left\{ h_n \int_0^E |K(u)| du + \alpha_{mn}(|K_m|) \right\}. \quad (16) \end{aligned}$$

Из (10) и (12) следует, что правая часть (16) не превосходит $F\varepsilon$, где $F = \text{const}$, 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

$$A_n = \frac{1}{b_m^2(h_n, x)} \int_0^\infty K_m^2\left(\frac{t}{h_n}\right) dF_m(x, t) < \infty. \quad (17)$$

Учитывая (8) - (10), получаем

$$A_n = \frac{1}{(h_n + \alpha_{mn}(K_m))^2} \left(h_n \int_0^E K_m^2(u) du + \alpha_{mn}(K_m^2) \right). \quad (18)$$

Теорема 2. Пусть дополнительно к условиям теоремы 1

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

Тогда

$$\limsup_{\substack{n \rightarrow \infty \\ m \rightarrow \infty \\ x \in Z_m}} \left| nh_n Df_{nm}(x) - f_m(x) \int_0^E K_m^2(u) du \right| = 0. \quad (20)$$

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

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

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

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

т.е. к условиям

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

Введем условие

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

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 $K: [0, +\infty] \rightarrow R^1$ - 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

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

функции принадлежности которых принимают значения из конечных множеств

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

Очевидно, число элементов пространства Z_m равно

$$Card(Z_m) = (q+1)^r. \quad (27)$$

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

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

и окрестности

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

Если

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

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

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

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

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

Положим

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

Тогда

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

в точках , в которых

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

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

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

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

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

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

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

Если $r = n^\alpha, q = n^\beta$, то

$$h_n^{-\gamma/r} \leq n^{\gamma/n^\alpha}, \tag{41}$$

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

Ядерная оценка плотности имеет вид

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

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:

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

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

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

Поскольку в силу интегральной теоремы Муавра-Лапласа

$$\lim_{m \rightarrow \infty} \sup \left| \frac{\sum_{0 \leq i \leq s} \binom{m}{i}}{2^m} - \Phi \left(\frac{2s - m}{\sqrt{m}} \right) \right| = 0, \tag{45}$$

то естественно рассмотреть меру близости

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

и ядерную оценку плотности

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

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

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

и в этих точках принимает значения

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

Следовательно,

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

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

$$\frac{P_m(\Phi^{-1}(h_n u))}{h_n u} \rightarrow 1, \quad (51)$$

где

$$P_m(\Phi^{-1}(h_n u)) = P\left(\frac{2X - m}{\sqrt{m}} \leq \Phi^{-1}(h_n u)\right). \quad (52)$$

Обозначим

$$c_n = c_n(u) = \Phi^{-1}(h_n u). \quad (53)$$

Тогда $c_n \rightarrow -\infty$ при $n \rightarrow \infty$, и (51) переходит в условие

$$\frac{P_m(c_n)}{\Phi(c_n)} \rightarrow 1. \quad (54)$$

По известной теореме [6, с.207] для справедливости (54) достаточно, чтобы

$$\frac{c_n^3}{\sqrt{m}} \rightarrow 0. \quad (55)$$

Согласно известной формуле [1, с.11] об асимптотике нормального распределения

$$c_n \sim \sqrt{-2(\ln h_n + \ln u)}. \quad (56)$$

Следовательно, если

$$m \geq G[\ln h_n]^{6+\gamma}, \quad G > 0, \quad \gamma > 0, \quad (57)$$

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,

$$p_m(A) = \frac{\text{Card}\{A\}}{\text{Card}\{Z_m\}}, \quad A \subseteq Z_m, \quad (58)$$

и d_m - расстояние (мера близости) в Z_m . Тогда

$$p_m(L_t(x)) = \frac{\text{Card}\{y : d_m(y, x) \leq t\}}{\text{Card}\{Z_m\}} = F_m^1(x, t). \quad (59)$$

Как обычно, положим

$$d_{1m}(y, x) = F_m^1(x, d_m(y, x)) \quad (60)$$

и рассмотрим

$$F_m(x, t) = p_m\{y : d_{1m}(y, x) < t\}. \quad (61)$$

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

$$F_m(x, t_i) = t_i. \quad (62)$$

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

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

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

$$\max(t_i - t_{i-1}) \rightarrow 0. \quad (64)$$

Из (64) согласно (63) следует, что

$$\lim_{m \rightarrow \infty} F_m(x, t) = t \quad (65)$$

для любого $t > 0$.

Если зафиксировать n , то (24) эквивалентно условию

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

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

Рассмотрим убывающую u_n и возрастающую m_n последовательности такие, что

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

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

$$\sup_t |F_m(x, t) - t| < \varepsilon \quad (68)$$

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

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

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

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

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

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

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

при $m > m_n$. Следовательно, если при $n \rightarrow \infty$ 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 $a_1 = a, a_2, a_3, \dots, a_k = b$ такие, что a_i и a_{i+1} являются соседями, $i = 1, 2, \dots, k - 1$. 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

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

"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

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

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.

7. Роль ядерных оценок плотности в статистике в пространствах произвольной природы

Статистика в пространствах произвольной природы – сердцевина статистики объектов нечисловой природы (статистики нечисловых данных, нечисловой статистики) [1, 2]. В этой научной области есть ряд ключевых понятий. К ним относятся прежде всего понятия эмпирических и теоретических средних.

Вначале эмпирические и теоретические средние в пространствах произвольной природы были введены и законы больших чисел для них получены на языке теории случайных множеств в работах 1978 г. [10, 11]. В первой научной монографии автора настоящей статьи (написана летом 1977 г., вышла в 1979 г. в серии «Проблемы советской экономики») были подведены итоги исследований в конкретных направлениях статистики объектов нечисловой природы (теории измерений, теории нечеткости, теории толерантностей, теории конечных случайных множеств и др.), установлены связи между отдельными видами объектов нечисловой

природы, появился (в предисловии) сам термин «статистика объектов нечисловой природы» [12]. Однако в ней еще не была выявлена роль статистики в пространствах общей (т.е. произвольной) природы. Например, законы больших чисел формулировались для случайных множеств, а не для эмпирических и теоретических средних в пространствах общей природы.

Надо отметить, что термин «устойчивость» в массовом сознании специалистов до сих пор часто ассоциируется с устойчивостью по Ляпунову и иной тематикой динамических систем, описываемых дифференциальными уравнениями. Поэтому название нашей первой научной монографии [12], хотя и верное по существу дела, при поверхностном восприятии порождало неверные ассоциации.

После выхода первой научной монографии автор настоящей статьи сузил область исследований, сосредоточившись на статистике объектов нечисловой природы. Название «статистика объектов нечисловой природы» ассоциируется правильно – с прикладной (математической) статистикой. Неправильные ассоциации с официальной государственной статистикой (ЦСУ, Госкомстат, Росстат) имеются на более высоком уровне иерархии понятий. Заслуживает обсуждения предложение заменить термин «прикладная статистика» на иной термин для ликвидации ложных ассоциаций, например, на термин «анализ данных», чтобы окончательно отмежеваться от официальной государственной статистики.

Следующий принципиально важный шаг был сделан в написанной годом позже (в 1978 г.) работе [13]. В ней была разработана программа развития нового направления прикладной статистики - статистики объектов нечисловой природы, реализованная позже, в основном в 80-х годах. Сформулированы все основные постановки и результаты. Показано, что стимулом к развитию статистики объектов нечисловой природы является теория и практика экспертных оценок. Статистика объектов

нечисловой природы является частью общей теории устойчивости, выделенной нами в качестве наиболее актуальной и перспективной.

Важным этапом в становлении и развитии статистики в пространствах общей природы является небольшая книга (брошюра) [14]. Она представляет собой «выжимку» моих работ 70-х годов, т.е. теории устойчивости и в особенности статистики объектов нечисловой природы, с уклоном в методологию. Эта книга включает в себя основные результаты по теории нечеткости и ее сведению к теории случайных множеств, а также новые результаты (первая публикация!) по статистике нечетких и случайных множеств. Задачи оптимизации увязывались с медианой Кемени, эмпирическими и теоретическими средними в пространствах произвольной природы. Именно с этой небольшой книги можно посоветовать начинать знакомство с рассматриваемым научным направлением. С брошюрой [14] тоже связано недоразумение - она вышла в научно-популярной серии "Математика, кибернетика" издательства "Знание", хотя по существу была научной монографией. Поэтому, несмотря на солидный тираж (40 тыс. экз.), эта брошюра сравнительно редко упоминается в научных изданиях.

Принципиально важной является работа [15]. В ней получены наиболее общие результаты, касающиеся законов больших чисел и асимптотики решений экстремальных (т.е. оптимизационных) статистических задач в пространствах общей природы. Доказательства нигде больше не публиковались. Работу целесообразно переиздать, чтобы сделать формулировки и доказательства теорем доступными современным читателям.

Краткое описание полученных в статье 1982 г. [15] результатов дано в заметке, посвященной их применению в конкретных задачах прикладной статистики [16]. На основе общих результатов об асимптотическом

поведении решения экстремальных статистических задач изучены, в частности, свойства нового метода экспертных оценок [17].

Важный частный случай асимптотики решений экстремальных статистических задач – оптимизационный подход к определению средних величин и законы больших чисел в пространствах общей природы – рассмотрен в серии статей [18 - 21].

Непараметрические оценки плотности распределения вероятностей в произвольных пространствах были введены в программной статье 1979 г. [13], "породившей" новую область прикладной статистики - статистику объектов нечисловой природы. Весьма важна новаторская фундаментальная работа [22], полностью посвященная непараметрическим оценкам плотности в топологических пространствах. В ней впервые введен ряд классов непараметрических оценок плотности в пространствах произвольной природы и доказана их состоятельность. Доказательства нигде больше не публиковались. Работу целесообразно переиздать, чтобы сделать формулировки и доказательства теорем доступными современным читателям.

Дальнейшее развитие теории непараметрических оценок плотности в пространствах произвольной природы дано в статьях [23 - 25]. Показано, что оптимальная скорость сходимости – та же, что и для непараметрических оценок плотности для числовых случайных величин. Поэтому можно сказать, что рассматриваемая теория доведена до ее естественных границ.

В этих статьях указано также на использование ядерных (и иных) оценок плотности для решения задач классификации [26, 27], прежде всего дискриминации (диагностики) [28, 29].

Современный этап углубленного изучения непараметрических оценок плотности в пространствах произвольной природы, в том числе

ядерных оценок, отражен в работах [3, 4], а также [30, 31]. Нужны дальнейшие исследования.

Литература

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 распределения вероятностей в пространствах произвольной природы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 99. С. 15-32.
4. Орлов А.И. Предельные теоремы для ядерных оценок плотности в пространствах произвольной природы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 108. С. 316 – 333.
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 её 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.
10. Орлов А.И. Элементы теории конечных случайных множеств // Прикладной многомерный статистический анализ. Ученые записки по статистике, т.33. - М.: Наука, 1978. С.301-307.
11. Орлов А.И. Случайные множества: законы больших чисел, проверка статистических гипотез // Теория вероятностей и ее применения. 1978. Т. XXIII. №2. С. 462-464.
12. Орлов А.И. Устойчивость в социально-экономических моделях. - М.: Наука, 1979. - 296 с.
13. 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.
14. Орлов А.И. Задачи оптимизации и нечеткие переменные. - М.: Знание, 1980. - 64 с.
15. Орлов А.И. Асимптотика решений экстремальных статистических задач // Анализ нечисловых данных в системных исследованиях. Сборник трудов. Вып.10. - М.: Всесоюзный научно-исследовательский институт системных исследований, 1982. С. 4-12.
16. Орлов А.И. Асимптотическое поведение решений экстремальных статистических задач // Заводская лаборатория. Диагностика материалов. 1996. Т.62. №10. С. 45-46.

17. Орлов А.И. Теоретическое обоснование «турнирного» метода ранжирования вариантов // Заводская лаборатория. Диагностика материалов. 2005. Т.71. №7. С.60-61.

18. Орлов А.И., Жихарев В.Н. Законы больших чисел и состоятельность статистических оценок в пространствах произвольной природы // Статистические методы оценивания и проверки гипотез. Межвузовский сборник научных трудов. – Пермь: Изд-во Пермского государственного университета, 1998. С.65-84.

19. Орлов А.И. Средние величины и законы больших чисел в пространствах произвольной природы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 89. С. 175-200.

20. Орлов А.И. Средние величины и законы больших чисел в топологических пространствах с показателями различия // Статистика и её применения. Материалы республиканской научно-практической конференции (Ташкент, 17-18 октября 2013 г.). – Ташкент: Изд-во «Университет», 2013. – С.30-37.

21. Орлов А. И. О средних величинах // Управление большими системами. Выпуск 46. М.: ИПУ РАН, 2013. С.88-117.

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

23. Орлов А.И. Классификация объектов нечисловой природы на основе непараметрических оценок плотности // Проблемы компьютерного анализа данных и моделирования: Сборник научных статей. - Минск: Изд-во Белорусского государственного университета, 1991. С.141-148.

24. Орлов А.И. Ядерные оценки плотности в пространствах произвольной природы // Статистические методы оценивания и проверки гипотез. Межвузовский сборник научных трудов. - Пермь: Пермский госуниверситет, 1996. С.68-75.

25. Орлов А.И. Математические методы исследования и диагностика материалов (Обобщающая статья) // Заводская лаборатория. Диагностика материалов. 2003. Т.69. №3. С.53-64.

26. Орлов А.И. О развитии математических методов теории классификации // Заводская лаборатория. Диагностика материалов. 2009. Т.75. №7. С.51-63.

27. Орлов А.И. Математические методы теории классификации // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 95. С. 23 – 45.

28. Орлов А.И. Базовые результаты математической теории классификации // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 110. С. 219–239.

29. Орлов А.И. Три основных результата математической теории классификации // Заводская лаборатория. Диагностика материалов. 2016. Т.82. №5. С. 63-70.

30. 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.

31. 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.

References

1. Orlov A.I. Organizacionno-jekonomicheskoe modelirovanie: uchebnik : v 3 ch. Chast' 1: Nechislovaja statistika. – M.: Izd-vo MGTU im. N.Je. Baumana. 2009. – 541 s.
2. Orlov A.I. O razvitii statistiki ob#ektov nechislovoj prirody // Politematicheskij setевой jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2013. № 93. S. 41-50.
3. Orlov A.I. Ocenki plotnosti raspredelenija verojatnostej v prostranstvah proizvod'noj prirody // Politematicheskij setевой jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 99. S. 15-32.
4. Orlov A.I. Predel'nye teoremy dlja jadernyh ocenok plotnosti v prostranstvah proizvod'noj prirody // Politematicheskij setевой jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2015. № 108. S. 316 – 333.
5. Verojatnost' i matematicheskaja statistika: Jenciklopedija / Gl. red. Ju.V. Prohorov. – M.: Bol'shaja Rossijskaja Jenciklopedija, 1999. – 910 s.
6. Feller V. Vvedenie v teoriju verojatnostej i ejo prilozhenija: Tom 1. - M.: Mir, 1984. - 528 s.
7. Bol'shev L.N., Smirnov N.V. Tablicy matematicheskoi statistiki / 3-e izd. - M.: Nauka, 1983. - 416 s.
8. Kulikov S.M. Strukturnye mery blizosti v prostranstve klassifikacij i razbienij // Prikladnaja statistika. - M.: Nauka, 1983. - S. 282-286.
9. Tjurin Ju.N. Jekspertnaja klassifikacija // Jekspertnye metody v sistemnyh issledovanijah / Sbornik trudov VNIISI. - 1979. - Vyp.4. - S.5-15.
10. Orlov A.I. Jelementy teorii konechnyh sluchajnyh mnozhestv // Prikladnoj mnogomernyj statisticheskij analiz. Uchenye zapiski po statistike, t.33. - M.: Nauka, 1978. S.301-307.
11. Orlov A.I. Sluchajnye mnozhestva: zakony bol'shih chisel, proverka statisticheskikh gipotez // Teorija verojatnostej i ee primenenija. 1978. T. XXIII. №2. S. 462-464.
12. Orlov A.I. Ustojchivost' v social'no-jekonomicheskikh modeljah. - M.: Nauka, 1979. - 296 s.
13. Orlov A.I. Statistika ob#ektov nechislovoj prirody i jekspertnye ocenki // Jekspertnye ocenki / Voprosy kibernetiki. Vyp.58. - M.: Nauchnyj Sovet AN SSSR po kompleksnoj probleme «Kibernetika», 1979. S.17-33.
14. Orlov A.I. Zadachi optimizacii i nechetkie peremennye. - M.: Znanie, 1980. - 64 s.
15. Orlov A.I. Asimptotika reshenij jekstremal'nyh statisticheskikh zadach // Analiz nechislovyh dannyh v sistemnyh issledovanijah. Sbornik trudov. Vyp.10. - M.: Vsesojuznyj nauchno-issledovatel'skij institut sistemnyh issledovanij, 1982. S. 4-12.
16. Orlov A.I. Asimptoticheskoe povedenie reshenij jekstremal'nyh statisticheskikh zadach // Zavodskaja laboratorija. Diagnostika materialov. 1996. T.62. №10. S. 45-46.
17. Orlov A.I. Teoreticheskoe obosnovanie «turnirnogo» metoda ranzhirovanija variantov // Zavodskaja laboratorija. Diagnostika materialov. 2005. T.71. №7. S.60-61.
18. Orlov A.I., Zhiharev V.N. Zakony bol'shih chisel i sostojatel'nost' statisticheskikh ocenok v prostranstvah proizvod'noj prirody // Statisticheskie metody ocenivaniya i proverki gipotez. Mezhvuzovskij sbornik nauchnyh trudov. – Perm': Izd-vo Permskogo gosudarstvennogo universiteta, 1998. S.65-84.
19. Orlov A.I. Srednie velichiny i zakony bol'shih chisel v prostranstvah proizvod'noj prirody // Politematicheskij setевой jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2013. № 89. S. 175-200.

20. Orlov A.I. Srednie velichiny i zakony bol'shih chisel v topologicheskikh prostranstvakh s pokazateljami razlichija // Statistika i ejo primenenija. Materialy respublikanskoj nauchno-prakticheskoy konferencii (Tashkent, 17-18 oktjabrja 2013 g.). – Tashkent: Izd-vo «Universitet», 2013. – S.30-37.
21. Orlov A. I. O srednih velichinah // Upravlenie bol'shimi sistemami. Vypusk 46. M.: IPU RAN, 2013. S.88-117.
22. Orlov A.I. Neparametricheskie ocenki plotnosti v topologicheskikh prostranstvakh // Prikladnaja statistika. Uchenye zapiski po statistike, t.45. - M.: Nauka, 1983. S. 12-40.
23. Orlov A.I. Klassifikacija ob#ektov nechislovoj prirody na osnove neparametricheskikh ocenok plotnosti // Problemy komp'juternogo analiza dannyh i modelirovanija: Sbornik nauchnyh statej. - Minsk: Izd-vo Belorusskogo gosudarstvennogo universiteta, 1991. S.141-148.
24. Orlov A.I. Jadernye ocenki plotnosti v prostranstvakh proizvol'noj prirody // Statisticheskie metody ocenivaniya i proverki gipotez. Mezhvuzovskij sbornik nauchnyh trudov. - Perm': Permskij gosuniversitet, 1996. S.68-75.
25. Orlov A.I. Matematicheskie metody issledovanija i diagnostika materialov (Obobshhajushhaja stat'ja) // Zavodskaja laboratorija. Diagnostika materialov. 2003. T.69. №3. S.53-64.
26. Orlov A.I. O razvitii matematicheskikh metodov teorii klassifikacii // Zavodskaja laboratorija. Diagnostika materialov. 2009. T.75. №7. S.51-63.
27. Orlov A.I. Matematicheskie metody teorii klassifikacii // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 95. S. 23 – 45.
28. Orlov A.I. Bazovye rezul'taty matematicheskoy teorii klassifikacii // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2015. № 110. S. 219–239.
29. Orlov A.I. Tri osnovnyh rezul'tata matematicheskoy teorii klassifikacii // Zavodskaja laboratorija. Diagnostika materialov. 2016. T.82. №5. S. 63-70.
30. Orlov A.I. Ocenki plotnosti v prostranstvakh proizvol'noj prirody // Statisticheskie metody ocenivaniya i proverki gipotez: mezhvuz. sb. nauch. tr. / Perm. gos. nac. issl. un-t. – Perm', 2013. – Vyp. 25. – S.21-33.
31. Orlov A.I. Jadernye ocenki plotnosti v prostranstvakh proizvol'noj prirody // Statisticheskie metody ocenivaniya i proverki gipotez: mezhvuz. sb. nauch. tr. / Perm. gos. nac. issl. un-t. – Perm', 2015. – Vyp. 26. – S. 43-57.