

УДК 303.732.4 : 519.2

08.00.13 Математические и инструментальные методы экономики (экономические науки)

СТАТИСТИКА НЕЧИСЛОВЫХ ДАННЫХ - ЦЕНТРАЛЬНАЯ ЧАСТЬ СОВРЕМЕННОЙ ПРИКЛАДНОЙ СТАТИСТИКИ

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

В 1979 г. статистика нечисловых данных была выделена как самостоятельная область прикладной статистики. Первоначально для обозначения этой области математических методов экономики использовался термин "статистика объектов нечисловой природы". Наш базовый учебник по статистике нечисловых данных называется "Нечисловая статистика". Статистика нечисловых данных - одна из четырех основных областей прикладной статистики (наряду со статистикой чисел, многомерным статистическим анализом, статистикой временных рядов и случайных процессов). Статистика нечисловых данных делится на статистику в пространствах общей природы и разделы, посвященные конкретным типам нечисловых данных (статистика интервальных данных, статистика нечетких множеств, статистика бинарных отношений и др.). В настоящее время статистика в пространствах общей природы - центральная часть прикладной статистики, а включающая ее статистика нечисловых данных - основная область прикладной статистики. Это утверждение подтверждается, в частности, анализом публикаций в разделе "Математические методы исследования" журнала "Заводская лаборатория. Диагностика материалов" - основном месте публикаций отечественных исследований по прикладной статистике. Настоящая статья посвящена анализу основных идей статистики нечисловых данных на фоне развития прикладной статистики с позиций новой парадигмы математических методов исследования. Описаны различные виды нечисловых данных. Проанализирован исторический путь статистической науки. Рассказано о развитии статистики нечисловых данных. Разобраны основные идеи статистики в пространствах общей природы: средние величины, законы больших чисел, экстремальные статистические задачи, непараметрические оценки плотности распределения вероятностей, методы классификации (диагностики и кластер-анализа), статистики интегрального типа. Кратко

UDC 303.732.4 : 519.2

08.00.13 Mathematical and instrumental methods of Economics (economic sciences)

NON-NUMERICAL DATA STATISTICS IS A CENTRAL PART OF MODERN APPLIED STATISTICS

Orlov Alexander Ivanovich
Dr.Sci.Econ., Dr.Sci.Tech., Cand.Phys-Math.Sci., professor
Bauman Moscow State Technical University, Moscow, Russia

In 1979, non-numerical data statistics was singled out as an independent area of applied statistics. Initially, the term "statistics of objects of non-numerical nature" was used to denote this area of mathematical methods of economics. Our basic non-numeric statistics textbook is called "Non-Numeric Statistics". Non-numerical data statistics is one of the four main areas of applied statistics (along with number statistics, multidimensional statistical analysis, statistics of time series and random processes). Statistics of non-numerical data are divided into statistics in spaces of a general nature and sections devoted to specific types of non-numerical data (statistics of interval data, statistics of fuzzy sets, statistics of binary relations, etc.). Currently, statistics in spaces of a general nature is the central part of applied statistics, and non-numeric data statistics including it is the main area of applied statistics. This statement is confirmed, in particular, by the analysis of publications in the section "Mathematical Research Methods" of the journal "Industrial Laboratory. Diagnostics of Materials" - the main place of publication of russian studies on applied statistics. This article is devoted to the analysis of the basic ideas of non-numerical data statistics against the background of the development of applied statistics from the perspective of a new paradigm of mathematical research methods. Various types of non-numeric data are described. The historical path of statistical science is analyzed. We have discussed the development of statistics of non-numerical data. The article analyzes basic ideas of statistics in spaces of a general nature: average values, laws of large numbers, extreme statistical problems, nonparametric estimates of the probability density, classification methods (diagnostics and cluster analysis), statistics of the integral type. Some statistical methods for analyzing data lying in specific spaces of non-numeric nature are briefly considered: non-parametric statistics (real distributions usually differ significantly from normal), statistics of fuzzy sets, theory of expert estimates (the Kemeny median is a sample average

рассмотрены некоторые статистические методы анализа данных, лежащих в конкретных пространствах нечисловой природы: непараметрическая статистика (реальные распределения обычно существенно отличаются от нормальных), статистика нечетких множеств, теория экспертных оценок (медиана Кемени - это выборочное среднее экспертных упорядочений) и др. Обсуждаются некоторые нерешенные задачи статистики нечисловых данных

of expert orderings), etc. We have also discussed some unsolved problems in statistics of nonnumeric data

Ключевые слова: МАТЕМАТИЧЕСКИЕ МЕТОДЫ ЭКОНОМИКИ, ПРИКЛАДНАЯ СТАТИСТИКА, НЕЧИСЛОВЫЕ ДАННЫЕ, СТАТИСТИКА В ПРОСТРАНСТВАХ ОБЩЕЙ ПРИРОДЫ, ЗАДАЧИ ОПТИМИЗАЦИИ, СРЕДНИЕ ВЕЛИЧИНЫ, ЗАКОНЫ БОЛЬШИХ ЧИСЕЛ, НЕПАРАМЕТРИЧЕСКИЕ ОЦЕНКИ ПЛОТНОСТИ, СТАТИСТИКИ ИНТЕГРАЛЬНОГО ТИПА, НЕПАРАМЕТРИЧЕСКАЯ СТАТИСТИКА, ЭКСПЕРТНЫЕ ОЦЕНКИ, НЕРЕШЕННЫЕ ЗАДАЧИ

Keywords: MATHEMATICAL METHODS OF ECONOMY, APPLIED STATISTICS, NONNUMERICAL DATA, STATISTICS IN SPACES OF GENERAL NATURE, OPTIMIZATION PROBLEMS, THE AVERAGE, LAWS OF LARGE NUMBERS, NONPARAMETRIC DENSITY ESTIMATES, INTEGRAL TYPE STATISTICS, NONPARAMETRIC STATISTICS, EXPERT ESTIMATION, UNSOLVED PROBLEMS

DOI: <http://dx.doi.org/10.21515/1990-4665-156-007>

1. Введение

В настоящее время статистика нечисловых данных - одна из четырех основных областей прикладной статистики. Остальные три - статистика чисел, многомерный статистический анализ, статистика временных рядов и случайных процессов. В свою очередь статистика нечисловых данных делится на статистику в (нелинейных) пространствах общей природы и разделы, посвященные конкретным типам нечисловых данных, такие, как статистика интервальных данных, статистика нечетких множеств, статистика бинарных отношений и др. Естественно, что научные результаты, полученные в рамках статистики в пространствах общей природы, могут быть использованы для конкретных видов данных (например, теория непараметрических оценок плотности распределения вероятностей). Следовательно, статистика в пространствах общей природы - центральная часть прикладной статистики, а включающая ее статистика нечисловых данных - основная область прикладной статистики. Это утверждение подтверждается, в частности, анализом публикаций в разделе

"Математические методы исследования" журнала "Заводская лаборатория. Диагностика материалов" - основного издания по прикладной статистике в России. По данным [1], на первое место по числу публикаций вышла именно статистика нечисловых данных. Так, за десять лет (2006 - 2015) ей посвящены 27,6% всех публикаций раздела "Математические методы исследования", т.е. 63,0% статей по прикладной статистике [1].

Первоначально для новой области прикладной статистики использовался термин "статистика объектов нечисловой природы". Он впервые появился в 1979 г. в нашей монографии [2] для обозначения совокупности некоторых полученных в ней научных результатов. В том же году в статье [3] нами была развернута программа построения этой новой области статистических методов, приведены первоначальные формулировки большинства основных теорем. Через год в «Заводской лаборатории» (так тогда назывался этот журнал) появилась обобщающая статья [4] пяти наиболее активных авторов среди занимавшихся различными аспектами статистики нечисловых данных. Итоги первых десяти лет развития новой области прикладной статистики были подведены в нашем обстоятельном обзоре [5] (120 литературных ссылок). Дальнейшее развитие было не менее плодотворным. Обзор [6] за тридцать лет содержал 150 литературных ссылок. К тридцатилетию вышел и первый учебник по статистике нечисловых данных [7]. В названии учебника использован термин "нечисловая статистика". Он представляется слишком кратким, в то время как исходный термин "статистика объектов нечисловой природы" - слишком тяжеловесным. Далее будем называть рассматриваемую область прикладной статистики "статистикой нечисловых данных". Такое название в наилучшей степени отражает ее содержание. Все три термина (статистика объектов нечисловой природы, статистика нечисловых данных, нечисловая статистика) - синонимы.

В настоящей статье обсудим содержание, развитие и основные идеи

статистики нечисловых данных.

2. Новая парадигма математических методов исследования - основа статистики нечисловых данных

Появление и развитие статистики нечисловых данных соответствуют переход к новой парадигме математических методов исследования.

Как известно, научная парадигма (от греч. *paradeigma* — пример, образец) — совокупность научных достижений (подходов, представлений, результатов), признаваемых научным сообществом в тот или иной период времени и служащих основой и образцом новых научных исследований. Понятие парадигмы получило широкое распространение после выхода в свет книги [8] американского историка науки Т. Куна «Структура научных революций» (1962).

Математические методы исследования используются для решения практических задач с давних времен. В Ветхом Завете рассказано о весьма квалифицированно проведенной переписи военнообязанных (Четвертая книга Моисеева "Числа"). В первой половине XX в. была сформирована классическая парадигма методов обработки данных, полученных в результате измерений (наблюдений, испытаний, анализов, опытов, обследований). Математические методы исследования, соответствующие классической парадигме, широко используются и в настоящее время. Со стороны может показаться, что в этой области основное давно сделано, современные работы направлены на мелкие усовершенствования. Однако это совсем не так. Новая парадигма математических методов исследования принципиально меняет прежние представления. Она зародилась в 1980-х гг., но была развита в серии наших статей, монографий и учебников уже в XXI в. (см. [9, 10] и др.).

Типовые исходные данные в новой парадигме – объекты нечисловой природы (элементы нелинейных пространств, которые нельзя складывать и

умножать на число, например, множества, бинарные отношения, графы, тексты), а в старой – числа, конечномерные векторы, функции [11]. Ранее (в классической старой парадигме) для расчетов использовались разнообразные суммы и функции от них, однако объекты нечисловой природы нельзя складывать, поэтому в новой парадигме применяется другой математический аппарат, основанный на расстояниях между объектами нечисловой природы и решении задач оптимизации.

Изменились постановки задач анализа статистических данных. Старая (классическая) парадигма исходит из идей начала XX в., когда К. Пирсон предложил четырехпараметрическое семейство распределений для описания распределений реальных данных. В это семейство как частные случаи входят, в частности, подсемейства нормальных, экспоненциальных, Вейбулла-Гнеденко, гамма-распределений. Сразу было ясно, что распределения реальных данных, как правило, не входят в семейство распределений Пирсона (об этом говорил, например, академик С.Н. Бернштейн в 1927 г. в докладе на Всероссийском съезде математиков). Однако математическая теория параметрических семейств распределений (методы оценивания параметров и проверки гипотез) оказалась достаточно интересной, и именно на ней до сих пор основано преподавание во многих вузах. Итак, в старой парадигме основной подход к описанию данных – распределения из параметрических семейств, а оцениваемые величины – их параметры, в новой парадигме рассматривают произвольные распределения, а оценивают – характеристики и плотности распределений, зависимости, правила диагностики и др. Центральная часть теории – уже не статистика числовых случайных величин, а статистика в пространствах произвольной природы.

В старой парадигме источники постановок новых задач – традиции, сформировавшиеся к середине XX века, а в новой – современные потребности математического моделирования и анализа данных (XXI век),

т.е. запросы практики. Конкретизируем это общее различие. В старой парадигме типовые результаты - предельные теоремы, в новой - рекомендации для конкретных значений параметров, в частности, для конкретных объемов выборок. Изменилась роль информационных технологий – ранее они использовались в основном для расчета таблиц, теперь же они - инструменты получения выводов (имитационное моделирование, датчики псевдослучайных чисел, методы размножения выборок, в т.ч. бутстреп, и др.). Вид постановок задач приблизился к потребностям практики – при анализе данных от отдельных задач оценивания и проверки гипотез перешли к статистическим технологиям (технологическим процессам анализа данных). Выявилась важность проблемы «стыковки алгоритмов» - влияния выполнения предыдущих алгоритмов в технологической цепочке на условия применимости последующих алгоритмов. В старой парадигме эта проблема не рассматривалась, для новой – весьма важна.

Если в старой парадигме вопросы методологии моделирования практически не обсуждались, достаточными признавались схемы начала XX в., то в новой парадигме роль методологии (учения об организации деятельности) является основополагающей. Резко повысилась роль моделирования – от отдельных систем аксиом произошел переход к системам моделей. Сама возможность применения вероятностного подхода теперь – не «наличие повторяющегося комплекса условий» (это требование - реликт "физического" определения вероятности, использовавшегося до аксиоматизации теории вероятностей А.Н. Колмогоровым в 1930-х гг.), а наличие обоснованной вероятностно-статистической модели. Если раньше данные считались полностью известными, то для новой парадигмы характерен учет свойств неопределенности данных, в частности, интервальных и нечетких. Изменилось отношение к вопросам устойчивости выводов – в старой

парадигме практически отсутствовал интерес к этой тематике, в новой разработана развитая теория устойчивости (робастности) выводов по отношению к допустимым отклонениям исходных данных и предпосылок моделей.

Статистика нечисловых данных развивается в соответствии с новой парадигмой математических методов исследования.

3. Различные виды нечисловых данных

Типичный исходный объект в прикладной статистике - это выборка, т.е. совокупность независимых одинаково распределенных случайных элементов. Какова природа этих элементов? В классической математической статистике элементы выборки - это числа. В многомерном статистическом анализе - вектора. А в статистике нечисловых данных элементы выборки - это объекты нечисловой природы, которые нельзя складывать и умножать на числа. Другими словами, объекты нечисловой природы лежат в пространствах, не имеющих векторной (линейной) структуры.

Примерами объектов нечисловой природы являются:

- значения качественных признаков (измеренных в шкалах наименований и порядковых), в том числе результаты кодировки объектов с помощью заданного перечня категорий (градаций);

- упорядочения (ранжировки) экспертами объектов экспертизы - образцов продукции (при оценке её технического уровня, качества, конкурентоспособности), ее характеристик; заявок на проведение научных работ (при проведении конкурсов на выделение грантов) и т.п.;

- классификации, т.е. разбиения объектов на группы сходных между собой (кластеры);

- толерантности, т.е. бинарные отношения, описывающие сходство объектов между собой, например, сходства тематики научных работ,

оцениваемого экспертами с целью рационального формирования экспертных советов внутри определенной области науки;

- результаты парных сравнений или контроля качества продукции по альтернативному признаку («годен» - «брак»), т.е. последовательности из 0 и 1;

- множества (обычные или нечеткие), например, зоны, пораженные коррозией, или перечни возможных причин аварии, составленные экспертами независимо друг от друга;

- слова, предложения, тексты;

- графы;

- вектора, координаты которых - совокупность значений разнотипных признаков, например, результат составления статистического отчета о научно-технической деятельности организации или анкета эксперта, в которой ответы на часть вопросов носят качественный характер, а на часть - количественный;

- ответы на вопросы экспертной, медицинской, маркетинговой или социологической анкеты, часть из которых носит количественный характер (возможно, интервальный), часть сводится к выбору одной из нескольких подсказок, а часть представляет собой тексты; и т.д.

Все средства измерения имеют погрешности. Однако до недавнего времени это очевидное обстоятельство никак не учитывалось в статистических процедурах. Только с конца 1970-х годов начала развиваться статистика интервальных данных, в которой предполагается, что исходные данные (элементы выборки) - это не числа, а интервалы. Статистику интервальных данных можно рассматривать как часть интервальной математики. Выводы в ней часто принципиально отличны от классических.

Различным подходам к статистическому анализу интервальных данных посвящена принципиально важная дискуссия [12]. Работают две

основные научные школы - А.П. Воцинина и наша. В первой из них изучены проблемы регрессионного анализа, планирования эксперимента, сравнения альтернатив и принятия решений в условиях интервальной неопределенности. К этой школе относится недавняя работа Н.В. Скибицкого [13]. В разработанной нами асимптотической статистике интервальных данных на значения случайных величин наложены малые интервальные неопределенности. Основные результаты этого направления подробно изложены в обширных главах учебников [7, 14, 15], монографии [16], в недавнем обзоре [17].

Интервальные данные можно рассматривать как частный случай нечетких множеств. Действительно, если характеристическая функция нечеткого множества равна 1 на некотором интервале и равна 0 вне этого интервала, то задание такого нечеткого множества эквивалентно заданию интервала. С методологической точки зрения важно, что *теория нечетких множеств в определенном смысле сводится к теории случайных множеств*. Цикл соответствующих теорем приведен в монографии [2], а также в учебниках [7, 14, 15, 18], монографии [16], недавней статье [19]. Казалось бы, много публикаций. Но приходится констатировать, что отнюдь не все специалисты знакомы с теоремами о сведении теории нечетких множеств к теории вероятностей.

4. Об истории и структуре статистической науки

Развитие статистических методов в нашей стране проанализировано в главе 2 монографии [20]. Дадим здесь краткую сводку, позволяющую выявить роль статистики нечисловых данных.

К 60-м годам XX в. в стране и мире сформировалась научно-практическая дисциплина, которую называем классической математической статистикой. Новое поколение отечественных исследователей училось теории по фундаментальной монографии

шведского математика Г. Крамера [21], написанной в военные годы и впервые изданной на русском языке в 1948 г. Из прикладных руководств назовем учебник [22] и таблицы с комментариями [23].

Затем внимание многих специалистов сосредоточилось на изучении математических конструкций, используемых в статистике. Примером таких работ является монография [24]. В ней получены продвинутое математические результаты, но трудно выделить рекомендации для статистика, анализирующего конкретные данные.

Как реакция на уход в математику выделилась новая научная дисциплина - прикладная статистика. В базовом учебнике по прикладной статистике [14] в качестве рубежа, когда это стало очевидным, указан 1981 г. – дату выхода массовым тиражом (33 940 экз.) сборника [25], в названии которого использован термин «прикладная статистика». С этого времени линии развития математической статистики и прикладной статистики разошлись. Первая из этих дисциплин полностью ушла в математику, перестав интересоваться практическими делами. Вторая позиционировала себя в качестве науки об обработке данных – результатов наблюдений, измерений, испытаний, анализов, опытов, обследований [14].

Вполне естественно, что в прикладной статистике стали создаваться свои математические методы и модели. Необходимость их развития вытекает из потребностей конкретных прикладных исследований. Это математизированное ядро прикладной статистики назовем *теоретической статистикой*. Тогда под собственно *прикладной статистикой* следует понимать обширную промежуточную область между теоретической статистикой и применением статистических методов в конкретных областях. Таким образом, общая схема современной статистической науки выглядит следующим образом (от абстрактного к конкретному):

1. Математическая статистика (математические методы в статистике) – часть математики, изучающая статистические структуры. Сама по себе

не дает рецептов анализа статистических данных, однако разрабатывает методы, полезные для использования в теоретической статистике.

2. Теоретическая статистика – наука, посвященная моделям и методам анализа конкретных статистических данных.

3. Прикладная статистика (в узком смысле) посвящена статистическим технологиям сбора и обработки данных. В нее входят, в частности, вопросы формирования вероятностно-статистических моделей и выбора конкретных методов анализа данных (т.е. методология прикладной статистики и других статистических методов), проблемы разработки и применения информационных статистических технологий, организации выборочных исследований, сбора данных и использования статистических программных продуктов.

4. Применение статистических методов в конкретных областях (в экономике и управлении (менеджменте) – эконометрика, в биологии – биометрика, в химии – хемометрия, в технических исследованиях – технометрика, в геологии, демографии, социологии, медицине, психологии, истории, и т.д.).

Часто позиции 2 и 3 вместе называют прикладной статистикой. Иногда позицию 1 именуют теоретической статистикой. Эти терминологические расхождения связаны с тем, что описанное выше развитие рассматриваемой научно-прикладной области не сразу, не полностью и не всегда адекватно отражается в сознании специалистов. Так, до сих пор выпускают учебники, соответствующие старой парадигме - уровню представлений середины XX века.

Примечание. Здесь мы уточнили схему внутреннего деления статистической теории, предложенную ранее в [26]. Естественный смысл приобрели термины «теоретическая статистика» и «прикладная статистика» (в узком смысле). Однако необходимо иметь в виду, что в базовом учебнике [14] прикладная статистика понимается в широком

смысле, т.е. как объединение позиций 2 и 3. К сожалению, в настоящее время невозможно отождествить теоретическую статистику с математической, поскольку последняя (как часть математики - научной специальности «теория вероятностей и математическая статистика») заметно оторвалась от задач практики.

Отметим, что математическая статистика, как и теоретическая с прикладной, заметно отличается от ведомственной науки органов официальной государственной статистики. ЦСУ, Госкомстат, Росстат применяли и применяют лишь проверенные временем приемы XIX века. Возможно, нам следовало бы от этого ведомства полностью отмежеваться и сменить название научной дисциплины, например, на «Анализ данных». В настоящее время компромиссным самоназванием является термин «статистические методы».

Во второй половине 1980-х годов в нашей стране развернулось общественное движение, имеющее целью создание профессионального объединения статистиков. Аналогами являются британское Королевское статистическое общество (основано в 1834 г.) и Американская статистическая ассоциация (создана в 1839 г.). К сожалению, деятельность учрежденной в 1990 г. Всесоюзной статистической ассоциации оказалась парализованной в результате развала СССР. Среди стран СНГ наибольшую активность в настоящее время проявляют узбекские исследователи, регулярно проводящие на высоком научном уровне представительные конференции по статистике и ее применениям.

5. О развитии статистики нечисловых данных

С 70-х годов XX в. в основном на основе запросов теории экспертных оценок (а также технических исследований, экономики, социологии и медицины) развивались различные направления статистики нечисловых данных. Были установлены основные связи между

конкретными видами таких объектов, разработаны для них базовые вероятностные модели [27]. Сводка полученных результатов дана в монографии [2], обзоре [4]. Это - предыстория статистики нечисловых данных. А история начинается с осмысления созданного, констатации [2, 3] в 1979 г. появления новой области прикладной статистики.

Следующий этап (1980-е годы) - развитие статистики нечисловых данных в качестве самостоятельного научного направления в рамках математических методов исследования, ядром которого являются методы статистического анализа данных произвольной природы. Для работ этого периода характерна сосредоточенность на внутренних (внутриматематических) проблемах статистики нечисловых данных. Проводились всесоюзные конференции, выпускались монографии, сборники трудов, защищались диссертации (Орлов А.И., Пярна К.А., Рыданова Г.В., Сатаров Г.А., Трофимов В.А., Шер А.П., Шмерлинг Д.С. и др.). Наиболее представительным является сборник [28], подготовленный совместно комиссией «Статистика объектов нечисловой природы» Научного Совета АН СССР по комплексной проблеме «Кибернетика» и Институтом социологических исследований АН СССР. Конкретная информация по работам 80-х годов имеется в обзорах [5, 6].

В настоящее время в связи с активным использованием наукометрических показателей разнообразными администраторами в области научной деятельности распространилась преувеличенная оценка роли журналов в развитии науки. Опыт статистики нечисловых данных показывает, что естественная цепочка развития научного результата такова: тезисы доклада — тематический сборник — монография — учебник — широкое использование [29]. Для развития нового направления публикации в научных журналах, вообще говоря, не обязательны. Ясно, что издание собственных журналов или завоевание позиций в уже существующих возможно лишь на этапе зрелости нового направления, но

не на этапе его создания.

К 1990-м годам статистика нечисловых данных с теоретической точки зрения была достаточно хорошо развита, основные идеи, подходы и методы были разработаны и изучены математически, в частности, доказано достаточно много теорем. Однако она оставалась недостаточно апробированной на практике. И в 90-е годы наступило время перейти от теоретико-статистических исследований к применению полученных результатов на практике, а также включить их в учебный процесс, что и было сделано. В 90-е годы в «Заводской лаборатории» опубликованы обзоры [5, 11, 27] по статистике объектов нечисловой природы и многочисленные конкретные исследования, рассмотренные позже в [6]. Серия работ была выполнена по статистике интервальных данных.

В 2000-е годы наиболее заметное явление - развернутые изложения основных результатов статистики нечисловых данных в учебниках по прикладной статистике, теории принятия решений, эконометрике [14, 15, 18]. Выпущен первый учебник по статистике нечисловых данных [7].

В 2010-е годы представленная научной общественности новая парадигма математических методов исследования закрепила положение статистики нечисловых данных как центральной быстро растущей части современной прикладной статистики (ср. обзор [1]). Опубликована серия работ по непараметрическим оценкам плотности распределения.

6. Основные идеи статистики в пространствах общей природы

В чем принципиальная новизна статистики нечисловых данных? Для классической математической статистики характерна операция сложения. При расчете выборочных характеристик распределения (выборочное среднее арифметическое, выборочная дисперсия и др.), в регрессионном анализе и других областях этой научной дисциплины постоянно используются суммы. Математический аппарат - законы больших чисел,

Центральная предельная теорема и другие теоремы - нацелены на изучение сумм. Принципиально важно, что в статистике нечисловых данных нельзя использовать операцию сложения, поскольку элементы выборки лежат в пространствах, где нет операции сложения. Методы обработки нечисловых данных основаны на принципиально ином математическом аппарате - на применении различных расстояний в пространствах объектов нечисловой природы и решении задач оптимизации.

Следует отметить, что в статистике нечисловых данных одна и та же математическая схема может с успехом применяться во многих прикладных областях, для анализа данных различных типов, а потому ее лучше всего формулировать и изучать в наиболее общем виде, для объектов произвольной природы.

Кратко рассмотрим несколько идей, развиваемых в статистике нечисловых данных для элементов выборок, лежащих в пространствах произвольного вида. Они нацелены на решение классических задач описания данных, оценивания, проверки гипотез - но для неклассических данных, а потому неклассическими методами.

Первой обсудим проблему определения средних величин. В рамках теории измерений удастся указать вид средних величин, соответствующих тем или иным шкалам измерения [30]. Теория измерений [31, 32], в середине XX в. рассматривавшаяся как часть математического обеспечения психологии, к настоящему времени признана общенаучной дисциплиной. Проблемы теории измерений постоянно рассматриваются в разделе "Математические методы исследования" журнала "Заводская лаборатория. Диагностика материалов" (см. соответствующий раздел [33]).

В классической математической статистике средние величины вводят с помощью операций сложения (выборочное среднее арифметическое, математическое ожидание) или упорядочения (выборочная и теоретическая медианы). В пространствах произвольной

природы средние значения нельзя определить с помощью операций сложения или упорядочения. Теоретические и эмпирические средние приходится вводить как решения экстремальных задач. Теоретическое среднее определяется как решение задачи минимизации математического ожидания (в классическом смысле) расстояния от случайного элемента со значениями в рассматриваемом пространстве до фиксированной точки этого пространства (минимизируется указанная функция от этой точки). Для получения эмпирического среднего математическое ожидание берется по эмпирическому распределению, т.е. берется сумма расстояний от некоторой точки до элементов выборки и затем минимизируется по этой точке (примером является медиана Кемени [34], методам нахождения которой посвящены недавние работы М.С. Жукова и его диссертация [35]). При этом как эмпирическое, так и теоретическое средние как решения экстремальных задач могут быть не единственными элементами рассматриваемого пространства, а являться некоторыми множествами таких элементов. Они могут оказаться и пустыми. Тем не менее удалось сформулировать и доказать законы больших чисел для та определенных средних величин, т.е. установить сходимость (в специально определенном смысле) эмпирических средних к теоретическим [7, 36].

Оказалось, что методы доказательства законов больших чисел допускают существенно более широкую область применения, чем та, для которой они были разработаны. А именно, удалось изучить асимптотику решений экстремальных статистических задач [7, 37], к которым, как известно, сводится большинство постановок прикладной статистики. В частности, дополнительно к законам больших чисел установлена состоятельность оценок минимального контраста, в том числе оценок максимального правдоподобия и робастных оценок. К настоящему времени подобные оценки изучены также и в статистике интервальных данных. Полученные результаты относительно асимптотики решений

экстремальных статистических задач применяются в ряде работ.

В статистической теории в пространствах общей природы большую роль играют непараметрические оценки плотности распределения вероятностей, используемые, в частности, в различных алгоритмах регрессионного, дискриминантного, кластерного анализ. В статистике нечисловых данных предложен и изучен ряд типов непараметрических оценок плотности в пространствах произвольной природы, в том числе в дискретных пространствах [38]. В частности, доказана их состоятельность, изучена скорость сходимости и установлен (для ядерных оценок плотности) примечательный факт совпадения наилучшей скорости сходимости в произвольном пространстве с той, которая имеет быть в классической теории для числовых случайных величин [39].

Введем обозначения. 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

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

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 [40, с.460].

Пусть X_1, X_2, \dots, X_n – независимые одинаково распределенные случайные элементы (величины), распределение которых задается вероятностной мерой q . Нами введено несколько видов непараметрических оценок плотности вероятности q по выборке X_1, X_2, \dots, X_n . Подробнее изучены линейные оценки. Подробнее рассмотрены их частные случаи – ядерные оценки плотности в пространствах произвольной природы.

Асимптотическая теория ядерных 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 удастся рассмотреть оценки плотностей и оценки вероятностей.

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 статистике нечисловых данных [3, с.24]:

$$f_n(x) = \frac{1}{n} \sum_{1 \leq i \leq n} g_n(x, X_i), \quad g_n : Z^2 \rightarrow R^1, \quad (2)$$

в которых действительнзначные функции g_n удовлетворяют некоторым условиям регулярности.

Пусть d – показатель различия (синоним - мера близости) на Z [7] (в наиболее важных частных случаях – метрика на Z). Нами введены ядерные оценки плотности – оценки вида (2) с

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

где $K = K(u)$ – ядро (ядерная функция), h_n – последовательность положительных чисел (показателей размытости), $b(h_n, x)$ – нормировочный множитель. Линейные оценки (2) с функциями g_n из (3) названы нами «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 и Парзеном.

Цель статей [38, 41] - завершение цикла работ, посвященного математическому изучению асимптотических свойств различных видов непараметрических оценок плотности распределения вероятности в пространствах общей природы. Изучен средний квадрат ошибки ядерной

оценки плотности. С целью максимизации порядка его убывания обоснован выбор ядерной функции и последовательности показателей размытости. Основные понятия - круговая функция распределения и круговая плотность. Порядок сходимости в общем случае тот же, что и при оценивании плотности числовой случайной величины [39], но основные условия наложены не на плотность случайной величины, а на круговую плотность. Далее рассматриваем другие виды непараметрических оценок плотности - гистограммные оценки и оценки типа Фикс-Ходжеса. Затем изучаем непараметрические оценки регрессии и их применение для решения задач дискриминантного анализа в пространстве общей природы

Дискриминантный, кластерный, регрессионный анализы в пространствах произвольной природы основаны либо на параметрической теории - и тогда применяется подход, связанный с асимптотикой решения экстремальных статистических задач - либо на непараметрической теории - и тогда используются алгоритмы на основе непараметрических оценок плотности.

Для анализа нечисловых, в частности, экспертных данных весьма важны методы классификации [42]. Интересно движение мысли в обратном направлении - наиболее естественно ставить и решать задачи классификации, основанные на использовании расстояний или показателей различия, именно в рамках статистики объектов нечисловой природы (а не, скажем, многомерного статистического анализа). Это касается как распознавания образов с учителем (другими словами, дискриминантного анализа), так и распознавания образов без учителя (т.е. кластерного анализа). Аналогичным образом задачи многомерного шкалирования, т.е. визуализации данных, также естественно отнести к статистике объектов нечисловой природы. Важны методы оценки истинной размерности признаков пространства [43].

Отметим несколько конкретных научных результатов

математической теории классификации. В задачах диагностики (дискриминантного анализа), как следует из леммы Неймана-Пирсона, целесообразно строить алгоритмы на основе отношения непараметрических оценок плотностей распределения вероятностей, соответствующих классам. Установлено, что наилучшим показателем качества алгоритма диагностики является прогностическая сила [44]. Устойчивость классификации относительно выбора метода кластер-анализа обосновывает вывод о реальности кластеров. И т.д. (см. соответствующий раздел в обзоре [1]).

Для проверки гипотез в пространствах нечисловой природы могут быть использованы статистики интегрального типа [3, 45], в частности, типа омега-квадрат. Отметим, что предельная теория таких статистик, построенная первоначально в классической постановке, приобрела естественный (завершенный, изящный) вид именно для пространств произвольного вида [46], поскольку при этом удалось провести рассуждения, опираясь на базовые математические соотношения, а не на те частные (с общей точки зрения), что были связаны с конечномерным пространством.

7. О некоторых областях статистики конкретных нечисловых данных

Кратко рассмотрим некоторые статистические методы анализа данных, лежащих в конкретных пространствах нечисловой природы.

Непараметрическая статистика – это прежде всего ранговая статистика, т.е. основанная на рангах – номерах элементов выборок в вариационных рядах. Ранги измерены в порядковых шкалах, а значения ранговых статистик инвариантны относительно любых строго возрастающих преобразований - допустимых преобразований в таких шкалах. Непараметрическая статистика позволяет делать статистические

выводы, оценивать характеристики и плотность распределения, проверять статистические гипотезы без слабо обоснованных предположений о том, что функция распределения элементов выборки входит в то или иное параметрическое семейство. Например, широко распространена вера в то, что статистические данные часто подчиняются нормальному распределению. Математики думают, что это - экспериментальный факт, установленный в прикладных исследованиях. Прикладники уверены, что математики доказали нормальность результатов наблюдений. Между тем анализ конкретных результатов наблюдений, в частности, погрешностей измерений, приводит всегда к одному и тому же выводу - в подавляющем большинстве случаев реальные распределения существенно отличаются от нормальных. На этот объективный факт обращал внимание В.В. Налимов в своей классической монографии [47]. Научная школа метролога П.В. Новицкого многочисленными экспериментами подтвердила отсутствие нормальности погрешностей измерений [48]. В сводке [49], в частности, установлено, что по выборкам объемов $n = 6 - 50$, как правило, не удастся отличить нормальное распределение от других видов распределений.

Некритическое использование гипотезы нормальности часто приводит к значительным ошибкам, например, при отбраковке резко выделяющихся результатов наблюдений (выбросов), при статистическом контроле качества и в других случаях [14]. Поэтому целесообразно использовать непараметрические методы, в которых на функции распределения результатов наблюдений наложены лишь весьма слабые требования. Обычно предполагается лишь их непрерывность. К настоящему времени с помощью непараметрических методов можно решать практически тот же круг задач, что ранее решался параметрическими методами. Примеры - оценивание характеристик распределения и проверка гипотезы однородности для независимых и связанных выборок [14]. Однако эта информация еще не вошла в массовое

сознание. До сих пор тупиковой тематике параметрической статистики посвящены обширные разделы учебников и программных продуктов. Современное состояние непараметрической статистики проанализировано в [50]. Эта область исследований продолжает активно развиваться.

Представляют практический интерес результаты, связанные с конкретными областями статистики объектов нечисловой природы, в частности, со статистикой нечетких множеств [51] и со статистикой случайных множеств (напомним, что теория нечетких множеств в определенном смысле сводится к теории случайных множеств [19]), с непараметрической теорией парных сравнений и люсианов (бернуллиевских бинарных векторов) [52], с аксиоматическим введением метрик в конкретных пространствах объектов нечисловой природы [7], а также с рядом других конкретных постановок.

Результаты контроля штучной продукции по альтернативному признаку представляют собой последовательности из 0 и 1, т.е. объекты нечисловой природы (люсианы), а потому теорию статистического контроля относят к статистике нечисловых данных [5, 6]. Постоянно публикуем работы по этой тематике, предназначенные для специалистов по статистическим методам управления качеством продукции (см. [15, 53] и др.).

Статистика нечисловых данных порождена потребностями практики, прежде всего в области экспертных оценок. Можно констатировать, что анализ экспертных оценок [54] - это прикладное «зеркало» общей теории. Решения задач теории экспертных оценок обобщались в статистике нечисловых данных. При движении мысли в обратном направлении результаты статистики в пространствах общей природы интерпретировались для анализа экспертных оценок. Как и для статистики нечисловых данных в целом, публикации шли по траектории: тезисы доклада — тематический сборник — монография — учебник — широкое

использование [29]. Вполне естественно, что названия сборников трудов неформального научного коллектива, развивающего статистику нечисловых данных, начинались со слов «Экспертные оценки» [55 - 58]. Отметим, что публикации в журналах не сыграли значительной роли в развитии рассматриваемых научных направлений. Обзор развития экспертных технологий в нашей стране дан в [59].

Вопросы внедрения математических методов исследования всегда были в центре внимания специалистов по статистике нечисловых данных [60]. Подчеркивалось большое теоретическое и прикладное значение статистики нечисловых данных, необходимость перехода от отдельных методов анализа данных к разработке высоких статистических технологий [61] и использования современных систем внедрения математических методов, таких как система «Шесть сигм» и ее аналоги. Обсуждались проблемы программного обеспечения [35, 62]. Однако приходится констатировать, что создание линейки современных программных продуктов по статистике нечисловых данных – пока дело будущего.

8. Некоторые нерешенные задачи статистики нечисловых данных

Начнем с обсуждения влияния отклонений от традиционных предпосылок. В вероятностной теории статистических методов выборка обычно моделируется как конечная последовательность независимых одинаково распределенных случайных величин или векторов. В устаревшей парадигме середины XX в. часто предполагают, что эти величины (вектора) имеют нормальное распределение.

При внимательном взгляде совершенно ясна нереалистичность приведенных классических предпосылок. Независимость результатов измерений обычно принимается «из общих предположений», между тем во многих случаях очевидна их коррелированность. Одинаковая

распределенность также вызывает сомнения из-за изменения во времени свойств измеряемых образцов, средств измерения и психофизического состояния специалистов, проводящих измерения (испытания, анализы, опыты). Даже обоснованность самого применения вероятностных моделей иногда вызывает сомнения, например, при моделировании уникальных измерений (согласно классическим воззрениям, теорию вероятностей обычно привлекают при изучении массовых явлений). И уж совсем редко распределения результатов измерений можно считать нормальными.

Итак, методы классической математической статистики обычно используют вне сферы их обоснованной применимости. Каково влияние отклонений от традиционных предпосылок на статистические выводы? В настоящее время об этом имеются лишь отрывочные сведения. Так, три примера в статье [6] показывают весь спектр возможных свойств классических расчетных методов в случае отклонения от нормальности. Так, методы построения доверительного интервала для математического ожидания оказываются вполне пригодными при таких отклонениях. Методы проверки однородности двух независимых выборок с помощью двухвыборочного критерия Стьюдента пригодны в некоторых случаях. В задаче отбраковки (исключения) резко выделяющихся наблюдений (выбросов) расчетные методы, основанные на нормальности, оказались полностью непригодными.

Очевидно, имеется ***необходимость изучения свойств расчетных методов классической математической статистики, опирающихся на предположение нормальности, в ситуациях, когда это предположение не выполнено.*** Аппаратом для такого изучения наряду с методом Монте-Карло могут послужить предельные теоремы теории вероятностей, прежде всего Центральная Предельная Теорема, поскольку интересующие нас расчетные методы обычно используют разнообразные суммы. Пока подобное изучение не проведено, остается неясной научная ценность,

например, применения основанного на предположении многомерной нормальности факторного анализа к векторам из переменных, принимающих небольшое число градаций и к тому же измеренных в порядковой шкале.

Нерешенным проблемам статистики посвящены статьи [132, 133]. Одна из важных проблем - использование асимптотических результатов при конечных объемах выборок. Конечно, естественно изучить свойства алгоритма с помощью метода Монте-Карло. Однако из какого конкретного распределения брать выборки при моделировании? От выбора распределения зависит результат. Кроме того, датчики псевдослучайных чисел лишь имитируют случайность. До сих пор неизвестно, каким датчиком целесообразно пользоваться в случае возможного безграничного роста размерности пространства.

Другая проблема – обоснование выбор одного из многих критериев для проверки конкретной гипотезы. Например, для проверки однородности двух независимых выборок можно использовать критерии Стьюдента, Крамера-Уэлча, Лорда, хи-квадрат, Вилкоксона (Манна-Уитни), Ван-дер-Вардена, Сэвиджа, Н.В. Смирнова, типа омега-квадрат (Лемана-Розенблатта), Реньи, Г.В.Мартынова и др. Какой выбрать?

Критерии однородности проанализированы в [65]. Естественных подходов к сравнению критериев несколько - на основе асимптотической относительной эффективности по Бахадуру, Ходжесу-Леману, Питмену. И каждый из перечисленных критериев является оптимальным при соответствующей альтернативе или подходящем распределении на множестве альтернатив. При этом математические выкладки обычно используют альтернативу сдвига, сравнительно редко встречающуюся в практике анализа реальных статистических данных. Итог печален - блестящая математическая техника, продемонстрированная в [65], не позволяет дать рекомендации для выбора критерия проверки однородности

при анализе реальных данных.

Проблемы разработки высоких статистических технологий поставлены в [61] (см. также одноименный сайт <http://orlovs.pp.ru>). Используемые при обработке реальных данных статистические технологии состоят из последовательности операций, каждая из которых, как правило, хорошо изучена, поскольку сводится к оцениванию (параметров, характеристик, распределений) или проверке той или иной гипотезы. Однако статистические свойства результатов обработки, полученных в результате последовательного применения таких операций, мало изучены [66]. Необходима теория, позволяющая изучать свойства статистических технологий и так их конструировать, чтобы обеспечить высокое качество обработки данных.

В заключение отметим, что развернутое описание статистики нечисловых данных дано в монографиях [7, 14, 18]. При дальнейшем развитии исследований важно опираться на современную методологию статистических методов [67].

Литература

1. Орлов А.И. Развитие математических методов исследования (2006 – 2015 гг.) / Заводская лаборатория. Диагностика материалов. 2017. Т.83. №1. Ч.1. С. 78-86.
2. Орлов А.И. Устойчивость в социально-экономических моделях. – М. : Наука, 1979. – 296 с.
3. 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.
4. Тюрин Ю.Н., Литвак Б.Г., Орлов А.И., Сатаров Г.А., Шмерлинг Д.С. Анализ нечисловой информации / Заводская лаборатория. 1980. Т.46. №10. С. 931-935.
5. Орлов А.И. Статистика объектов нечисловой природы (Обзор) / Заводская лаборатория. 1990. Т.56. №3. С. 76-83.
6. Орлов А.И. Тридцать лет статистики объектов нечисловой природы (обзор) / Заводская лаборатория. Диагностика материалов. 2009. Т.75. №5. С. 55-64.
7. Орлов А.И. Организационно-экономическое моделирование. Часть 1. Нечисловая статистика. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2009. – 544 с.
8. Кун Т. Структура научных революций. М.: АСТ, 2003. — 605 с.
9. Орлов А. И. Новая парадигма прикладной статистики / Заводская лаборатория. Диагностика материалов. 2012. Т.78. №1, часть I. С. 87-93.

10. Орлов А. И. О новой парадигме математических методов исследования / Научный журнал КубГАУ. 2016. №122. С. 807–832.
11. Орлов А.И. Объекты нечисловой природы / Заводская лаборатория. Диагностика материалов. 1995. Т.61. №3. С.43-52.
12. Дискуссия по анализу интервальных данных / Заводская лаборатория. 1990. Т.56. №7. С.75-95.
13. Скибицкий Н.В. Решение задачи аналитического описания статических характеристик в условиях интервальной неопределенности / Заводская лаборатория. Диагностика материалов. 2019. Т.85. № 3. С. 64-74.
14. Орлов А.И. Прикладная статистика. - М.: Экзамен, 2006. - 671 с.
15. Орлов А.И. Теория принятия решений.– М.: Экзамен, 2006. – 576 с.
16. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. – Краснодар, КубГАУ. 2014. – 600 с.
17. Орлов А.И. Статистика интервальных данных (обобщающая статья) / Заводская лаборатория. Диагностика материалов. 2015. Т. 81. № 3. С. 61 - 69.
18. Орлов А. И. Эконометрика. - М.: Экзамен, 2002. – 576 с.
19. Орлов А.И. Теория нечетких множеств – часть теории вероятностей / Научный журнал КубГАУ. 2013. № 92. С. 51-60.
20. Орлов А.И., Луценко Е.В., Лойко В.И. Организационно-экономическое, математическое и программное обеспечение контроллинга, инноваций и менеджмента: монография / под общ. ред. С. Г. Фалько. – Краснодар : КубГАУ, 2016. – 600 с.
21. Крамер Г. Математические методы статистики. - М.: Мир, 1975. - 648 с.
22. Смирнов Н. В., Дунин-Барковский И. В. Курс теории вероятностей и математической статистики для технических приложений. Изд. 3-е, стереотипное. – М.: Наука, 1969. – 512 с.
23. Большев Л. Н., Смирнов Н. В. Таблицы математической статистики / 3-е изд.- М.: Наука, 1983. - 416 с. (1-е изд. – 1965).
24. Каган А. М., Линник Ю. В., Рао С. Р. Характеризационные задачи математической статистики. - М.: Наука, 1972. - 656 с.
25. Современные проблемы кибернетики (прикладная статистика). - М.: Знание, 1981. – 64 с.
26. Орлов А.И. О перестройке статистической науки и её применений / Вестник статистики. 1990. № 1. С.65 – 71.
27. Орлов А.И. Вероятностные модели конкретных видов объектов нечисловой природы / Заводская лаборатория. Диагностика материалов. 1995. Т.61. №5. С.43-51.
28. Андреенков В.Г., Орлов А.И., Толстова Ю.Н. (отв. ред.). Анализ нечисловой информации в социологических исследованиях. - М.: Наука, 1985. - 220 с.
29. Лойко В.И., Луценко Е.В., Орлов А.И. Современные подходы в наукометрии. – Краснодар: КубГАУ, 2017. – 532 с.
30. Орлов А.И. Характеризация средних величин шкалами измерения / Научный журнал КубГАУ. 2017. №134. С. 877 – 907.
31. Психологические измерения. Сб. статей. - М.: Мир, 1967. - 196 с.
32. Пфанцагель И. Теория измерений. - М.: Мир, 1976. - 248 с.
33. Орлов А.И. Статистика нечисловых данных за сорок лет (обзор) / Заводская лаборатория. Диагностика материалов. 2019. Т.85. №11. С. 69-84.
34. Кемени Дж., Снелл Дж. Кибернетическое моделирование: Некоторые приложения. – М.: Советское радио, 1972. – 192 с.

35. Жуков М. С., Орлов А. И. Задача исследования итогового ранжирования мнений группы экспертов с помощью медианы Кемени / Научный журнал КубГАУ. 2016. № 122. С. 785 – 806.
36. Орлов А.И. Средние величины и законы больших чисел в пространствах произвольной природы / Научный журнал КубГАУ. 2013. № 89. С. 556 – 586.
37. Орлов А.И. Предельная теория решений экстремальных статистических задач / Научный журнал КубГАУ. 2017. №133. С. 579 – 600.
38. Орлов А.И. Асимптотика оценок плотности распределения вероятностей / Научный журнал КубГАУ. 2017. № 131. С. 845 – 873.
39. Ибрагимов И.А., Хасьминский Р.З. Асимптотическая теория оценивания. – М.: Наука, 1979. – 528 с.
40. Вероятность и математическая статистика: Энциклопедия / Гл. ред. Ю.В. Прохоров. – М.: Большая Российская Энциклопедия, 1999. – 910 с.
41. Орлов А.И. Скорость сходимости ядерных оценок плотности в пространствах произвольной природы / Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. / Перм. гос. нац. иссл. ун-т. - Пермь, 2018. - Вып.28. - С. 35-45.
42. Орлов А.И. Математические методы теории классификации / Научный журнал КубГАУ. 2014. № 95. С. 23 – 45.
43. Луценко Е.В., Орлов А.И. Методы снижения размерности пространства статистических данных / Научный журнал КубГАУ. 2016. № 119. С. 92–107.
44. Орлов А.И. Прогностическая сила – наилучший показатель качества алгоритма диагностики / Научный журнал КубГАУ. 2014. № 99. С. 33–49.
45. Орлов А.И. Асимптотическое поведение статистик интегрального типа / Доклады АН СССР. 1974. Т.219. №4. С. 808-811.
46. Орлов А.И. Предельная теория непараметрических статистик / Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 100. С. 31-52.
47. Налимов В.В. Применение математической статистики при анализе вещества. – М.: Физматгиз, 1960. – 430 с.
48. Новицкий П.В., Зограф И.А. Оценка погрешностей результатов измерений. – Л.: Энергоатомиздат, 1985. – 248 с.
49. Орлов А.И. Распределения реальных статистических данных не являются нормальными / Научный журнал КубГАУ. 2016. № 117. С. 71–90.
50. Орлов А.И. Современное состояние непараметрической статистики / Научный журнал КубГАУ. 2015. № 106. С. 239 – 269.
51. Орлов А.И. Статистика нечетких данных / Научный журнал КубГАУ. 2016. №119. С. 75–91.
52. Орлов А.И. Теория люсианов / Научный журнал КубГАУ. 2014. № 101. С. 275 – 304.
53. Орлов А.И. Предельные теоремы в статистическом контроле / Научный журнал КубГАУ. 2016. № 116. С. 462 – 483.
54. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч.2. Экспертные оценки. — М.: Изд-во МГТУ им. Н. Э. Баумана, 2011. — 486 с.
55. Статистические методы анализа экспертных оценок / Ученые записки по статистике, т. 29. - М.: Наука, 1977. - 385 с.
56. Экспертные оценки / Вопросы кибернетики. - Вып.58. - М.: Научный Совет АН СССР по комплексной проблеме "Кибернетика". 1979. - 200 с.
57. Экспертные оценки в системных исследованиях / Сборник трудов. - Вып.4. - М.: ВНИИСИ, 1979. - 120 с.

58. Экспертные оценки в задачах управления / Сборник трудов. - М.: Институт проблем управления. 1982. - 106 с.
59. Орлов А.И. Теория экспертных оценок в нашей стране / Научный журнал КубГАУ. 2013. № 93. С. 1-11.
60. Гнеденко Б.В., Орлов А.И. Роль математических методов исследования в кардинальном ускорении научно-технического прогресса / Заводская лаборатория. 1988. Т.54. №1. С.1 - 4.
61. Орлов А.И. О высоких статистических технологиях / Научный журнал КубГАУ. 2015. № 105. С. 14 – 38.
62. Орлов А.И. Компьютерно-статистические методы: состояние и перспективы / Научный журнал КубГАУ. 2014. № 103. С. 163 – 195.
63. Загоруйко Н.Г., Орлов А.И. Некоторые нерешенные математические задачи прикладной статистики / Современные проблемы кибернетики (прикладная статистика). - М.: Знание, 1981. - С.53-63.
64. Орлов А.И. Некоторые нерешенные вопросы в области математических методов исследования / Заводская лаборатория. Диагностика материалов. 2002. Т.68. №3. С.52-56.
65. Никитин Я.Ю. Асимптотическая эффективность непараметрических критериев. - М.: Наука, 1995. - 240 с.
66. Орлов А.И. Проблема множественных проверок статистических гипотез / Заводская лаборатория. Диагностика материалов. 1996. Т.62. №5. С.51-54.
67. Орлов А.И. О методологии статистических методов / Научный журнал КубГАУ. 2014. № 104. С. 53–80.

References

1. Orlov A.I. Razvitie matematicheskikh metodov issledovaniya (2006 – 2015 gg.) / Zavodskaya laboratoriya. Diagnostika materialov. 2017. T.83. №1. CH.1. S. 78-86.
2. Orlov A.I. Ustojchivost' v social'no-ekonomicheskikh modelyah. – М. : Nauka, 1979. – 296 s.
3. Orlov A.I. Statistika ob"ektov nechislovoj prirody i ekspertnye ocenki / Ekspertnye ocenki. Voprosy kibernetiki. Vyp.58. - М.: Nauchnyj Sovet AN SSSR po kompleksnoj probleme «Kibernetika», 1979. S. 17-33.
4. Tyurin YU.N., Litvak B.G., Orlov A.I., Satarov G.A., SHmerling D.S. Analiz nechislovoj informacii / Zavodskaya laboratoriya. 1980. T.46. №10. S. 931-935.
5. Orlov A.I. Statistika ob"ektov nechislovoj prirody (Obzor) / Zavodskaya laboratoriya. 1990. T.56. №3. S. 76-83.
6. Orlov A.I. Tridcat' let statistiki ob"ektov nechislovoj prirody (obzor) / Zavodskaya laboratoriya. Diagnostika materialov. 2009. T.75. №5. S. 55-64.
7. Orlov A.I. Organizacionno-ekonomicheskoe modelirovanie. CHast' 1. Nechislovaya statistika. – М.: Izd-vo MGTU im. N.E. Baumana, 2009. – 544 s.
8. Kun T. Struktura nauchnyh revolyucij. М.: AST, 2003. — 605 s.
9. Orlov A. I. Novaya paradigma prikladnoj statistiki / Zavodskaya laboratoriya. Diagnostika materialov. 2012. T.78. №1, chast' I. S. 87-93.
10. Orlov A. I. O novej paradigme matematicheskikh metodov issledovaniya / Nauchnyj zhurnal KubGAU. 2016. №122. S. 807–832.
11. Orlov A.I. Ob"ekty nechislovoj prirody / Zavodskaya laboratoriya. Diagnostika materialov. 1995. T.61. №3. S.43-52.
12. Diskussiya po analizu interval'nyh dannyh / Zavodskaya laboratoriya. 1990. T.56. №7. S.75-95.

13. Skibickij N.V. Reshenie zadachi analiticheskogo opisaniya staticheskikh karakteristik v usloviyah interval'noj neopredelennosti / Zavodskaya laboratoriya. Diagnostika materialov. 2019. T.85. № 3. S. 64-74.
14. Orlov A.I. Prikladnaya statistika. - M.: Ekzamen, 2006. - 671 s.
15. Orlov A.I. Teoriya prinyatiya reshenij.– M.: Ekzamen, 2006. – 576 s.
16. Orlov A.I., Lucenko E.V. Sistemnaya nechetskaya interval'naya matematika. – Krasnodar, KubGAU. 2014. – 600 s.
17. Orlov A.I. Statistika interval'nyh dannyh (obobshchayushchaya stat'ya) / Zavodskaya laboratoriya. Diagnostika materialov. 2015. T. 81. № 3. S. 61 - 69.
18. Orlov A. I. Ekonometrika. - M.: Ekzamen, 2002. – 576 s.
19. Orlov A.I. Teoriya nechetkih mnozhestv – chast' teorii veroyatnostej / Nauchnyj zhurnal KubGAU. 2013. № 92. S. 51-60.
20. Orlov A.I., Lucenko E.V., Lojko V.I. Organizacionno-ekonomicheskoe, matematicheskoe i programmnoe obespechenie kontrollinga, innovacij i menedzhmenta: monografiya / pod obshch. red. S. G. Fal'ko. – Krasnodar : KubGAU, 2016. – 600 s.
21. Kramer G. Matematicheskie metody statistiki. - M.: Mir, 1975. - 648 s.
22. Smirnov N. V., Dunin-Barkovskij I. V. Kurs teorii veroyatnostej i matematicheskoy statistiki dlya tekhnicheskikh prilozhenij. Izd. 3-e, stereotipnoe. – M.: Nauka, 1969. – 512 s.
23. Bol'shev L. N., Smirnov N. V. Tablicy matematicheskoy statistiki / 3-e izd.- M.: Nauka, 1983. - 416 s. (1-e izd. – 1965).
24. Kagan A. M., Linnik YU. V., Rao S. R. Harakterizacionnye zadachi matematicheskoy statistiki. - M.: Nauka, 1972. - 656 s.
25. Sovremennye problemy kibernetiki (prikladnaya statistika). - M.: Znanie, 1981. – 64 s.
26. Orlov A.I. O perestrojke statisticheskoy nauki i eyo primenenij / Vestnik statistiki. 1990. № 1. S.65 – 71.
27. Orlov A.I. Veroyatnostnye modeli konkretnyh vidov ob"ektov nechislovoj prirody / Zavodskaya laboratoriya. Diagnostika materialov. 1995. T.61. №5. S.43-51.
28. Andreenkov V.G., Orlov A.I., Tolstova YU.N. (otv. red.). Analiz nechislovoj informacii v sociologicheskikh issledovaniyah. - M.: Nauka, 1985. - 220 s.
29. Lojko V.I., Lucenko E.V., Orlov A.I. Sovremennye podhody v naukometrii. – Krasnodar: KubGAU, 2017. – 532 s.
30. Orlov A.I. Harakterizaciya srednih velichin shkalami izmereniya / Nauchnyj zhurnal KubGAU. 2017. №134. S. 877 – 907.
31. Psihologicheskie izmereniya. Sb. statej. - M.: Mir, 1967. - 196 s.
32. Pfancagl' I. Teoriya izmerenij. - M.: Mir, 1976. - 248 s.
33. Orlov A.I. Statistika nechislovyh dannyh za sorok let (obzor) / Zavodskaya laboratoriya. Diagnostika materialov. 2019. T.85. №11. S. 69-84.
34. Kemeni Dzh., Snell Dzh. Kiberneticheskoe modelirovanie: Nekotorye prilozheniya. – M.: Sovetskoe radio, 1972. – 192 s.
35. ZHukov M. S., Orlov A. I. Zadacha issledovaniya itogovogo ranzhirovaniya mnenij gruppy ekspertov s pomoshch'yu mediany Kemeni / Nauchnyj zhurnal KubGAU. 2016. № 122. S. 785 – 806.
36. Orlov A.I. Srednie velichiny i zakony bol'shikh chisel v prostranstvah proizvod'noj prirody / Nauchnyj zhurnal KubGAU. 2013. № 89. S. 556 – 586.
37. Orlov A.I. Predel'naya teoriya reshenij ekstremal'nyh statisticheskikh zadach / Nauchnyj zhurnal KubGAU. 2017. №133. S. 579 – 600.
38. Orlov A.I. Asimptotika ocenok plotnosti raspredeleniya veroyatnostej / Nauchnyj zhurnal KubGAU. 2017. № 131. S. 845 – 873.

39. Ibragimov I.A., Has'minskij R.Z. Asimptoticheskaya teoriya ocenivaniya. – M.: Nauka, 1979. – 528 s.
40. Veroyatnost' i matematicheskaya statistika: Enciklopediya / Gl. red. YU.V. Prohorov. – M.: Bol'shaya Rossijskaya Enciklopediya, 1999. – 910 s.
41. Orlov A.I. Skorost' skhodimosti yadernyh ocenok plotnosti v prostranstvah proizvol'noj prirody / Statisticheskie metody ocenivaniya i proverki gipotez: mezhvuz. sb. nauch. tr. / Perm. gos. nac. issl. un-t. - Perm', 2018. - Vyp.28. - S. 35-45.
42. Orlov A.I. Matematicheskie metody teorii klassifikacii / Nauchnyj zhurnal KubGAU. 2014. № 95. S. 23 – 45.
43. Lucenko E.V., Orlov A.I. Metody snizheniya razmernosti prostranstva statisticheskikh dannyh / Nauchnyj zhurnal KubGAU. 2016. № 119. S. 92–107.
44. Orlov A.I. Prognosticheskaya sila – nailuchshij pokazatel' kachestva algoritma diagnostiki / Nauchnyj zhurnal KubGAU. 2014. № 99. S. 33–49.
45. Orlov A.I. Asimptoticheskoe povedenie statistik integral'nogo tipa / Doklady AN SSSR. 1974. T.219. №4. S. 808-811.
46. Orlov A.I. Predel'naya teoriya neparametricheskikh statistik / Politematicheskij setевой elektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 100. S. 31-52.
47. Nalimov V.V. Primenenie matematicheskoy statistiki pri analize veshchestva. – M.: Fizmatgiz, 1960. – 430 s.
48. Novickij P.V., Zograf I.A. Ocenka pogreshnostej rezul'tatov izmerenij. – L.: Energoatomizdat, 1985. – 248 s.
49. Orlov A.I. Raspredeleniya real'nyh statisticheskikh dannyh ne yavlyayutsya normal'nymi / Nauchnyj zhurnal KubGAU. 2016. № 117. S. 71–90.
50. Orlov A.I. Sovremennoe sostoyanie neparametricheskoy statistiki / Nauchnyj zhurnal KubGAU. 2015. № 106. S. 239 – 269.
51. Orlov A.I. Statistika nechetkih dannyh / Nauchnyj zhurnal KubGAU. 2016. №119. S. 75–91.
52. Orlov A.I. Teoriya lyusianov / Nauchnyj zhurnal KubGAU. 2014. № 101. S. 275 – 304.
53. Orlov A.I. Predel'nye teoremy v statisticheskom kontrole / Nauchnyj zhurnal KubGAU. 2016. № 116. S. 462 – 483.
54. Orlov A.I. Organizacionno-ekonomicheskoe modelirovanie : uchebnik : v 3 ch. CH.2. Ekspertnye ocenki. — M.: Izd-vo MGTU im. N. E. Baumana, 2011. — 486 s.
55. Statisticheskie metody analiza ekspertnyh ocenok / Uchenye zapiski po statistike, t. 29. - M.: Nauka, 1977. - 385 s.
56. Ekspertnye ocenki / Voprosy kibernetiki. - Vyp.58. - M.: Nauchnyj Sovet AN SSSR po kompleksnoj probleme "Kibernetika". 1979. - 200 s.
57. Ekspertnye ocenki v sistemnyh issledovaniyah / Sbornik trudov. - Vyp.4. - M.: VNIISI, 1979. - 120 s.
58. Ekspertnye ocenki v zadachah upravleniya / Sbornik trudov. - M.: Institut problem upravleniya. 1982. - 106 s.
59. Orlov A.I. Teoriya ekspertnyh ocenok v nashej strane / Nauchnyj zhurnal KubGAU. 2013. № 93. S. 1-11.
60. Gnedenko B.V., Orlov A.I. Rol' matematicheskikh metodov issledovaniya v kardinal'nom uskorenii nauchno-tehnicheskogo progressa / Zavodskaya laboratoriya. 1988. T.54. №1. S.1 - 4.
61. Orlov A.I. O vysokih statisticheskikh tekhnologiyah / Nauchnyj zhurnal KubGAU. 2015. № 105. S. 14 – 38.
62. Orlov A.I. Komp'yuterno-statisticheskie metody: sostoyanie i perspektivy /

Nauchnyj zhurnal KubGAU. 2014. № 103. S. 163 – 195.

63. Zagorujko N.G., Orlov A.I. Nekotorye nereshennye matematicheskie zadachi prikladnoj statistiki / Sovremennye problemy kibernetiki (prikladnaya statistika). - M.: Znanie, 1981. - S.53-63.

64. Orlov A.I. Nekotorye nereshennye voprosy v oblasti matematicheskikh metodov issledovaniya / Zavodskaya laboratoriya. Diagnostika materialov. 2002. T.68. №3. S.52-56.

65. Nikitin YA.YU. Asimptoticheskaya effektivnost' neparametricheskikh kriteriev. - M.: Nauka, 1995. - 240 s.

66. Orlov A.I. Problema mnozhestvennykh proverok statisticheskikh gipotez / Zavodskaya laboratoriya. Diagnostika materialov. 1996. T.62. №5. S.51-54.

67. Orlov A.I. O metodologii statisticheskikh metodov / Nauchnyj zhurnal KubGAU. 2014. № 104. S. 53–80.