

УДК 519.2:303.732.4

UDC 519.2:303.732.4

01.00.00 Физико-математические науки

Physics and mathematical sciences

СОСТОЯНИЕ И ПЕРСПЕКТИВЫ РАЗВИТИЯ ПРИКЛАДНОЙ И ТЕОРЕТИЧЕСКОЙ СТАТИСТИКИ**STATE AND PROSPECTS OF APPLIED AND THEORETICAL STATISTICS**

Орлов Александр Иванович
д.э.н., д.т.н., к.ф.-м.н., профессор
РИНЦ SPIN-код: 4342-4994

Orlov Alexander Ivanovich
Dr.Sci.Econ., Dr.Sci.Tech., Cand.Phys-Math.Sci.,
professor

Московский государственный технический университет им. Н.Э. Баумана, Россия, 105005, Москва, 2-я Бауманская ул., 5, prof-orlov@mail.ru

Bauman Moscow State Technical University, Moscow, Russia

Общая схема современной статистической науки такова. Математическая статистика – часть математики, изучающая статистические структуры (сама по себе не дает рецептов анализа статистических данных, однако разрабатывает методы, полезные для использования в теоретической статистике). Теоретическая статистика – наука, посвященная моделям и методам анализа конкретных статистических данных. Прикладная статистика (в узком смысле) посвящена статистическим технологиям сбора и обработки данных (она включает в себя методологию статистических методов, вопросы организации выборочных исследований, разработки статистических технологий, создания и использования статистических программных продуктов). Применения статистических методов в конкретных областях (в экономике и менеджменте – эконометрика, в биологии – биометрика, в химии – хемометрия, в технических исследованиях – технометрика, в геологии, демографии, социологии, медицине, 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 уровню представлений середины XX века. В статье проведен анализ послевоенного развития отечественной статистики. Выделены пять «точек роста»: непараметрика, робастность, бутстреп, статистика интервальных данных, статистика нечисловых данных. Обсуждается содержание, развитие и основные идеи статистики объектов нечисловой природы. Рассмотрен ряд нерешенных проблем теоретической и прикладной статистики

The general scheme of modern statistical science is just like this. Mathematical Statistics is a part of mathematics that studies the statistical structure (it itself does not give recipes analysis of statistical data, however, it is developing methods that are useful for use in theoretical statistics). Theoretical Statistics - the science dedicated to the models and methods of analysis of concrete statistical data. Applied Statistics (in the narrow sense) is devoted to the statistical techniques of data collection and processing (it includes the methodology of statistical methods, the organization of sample surveys, the development of statistical techniques, the creation and use of statistical software). Applications of statistical methods in concrete fields (in economics and management - Econometrics, in biology - Biometrics, in chemistry - Chemometrics, in technical research - Technometric, in geology, demography, sociology, medicine, history, etc.). Often positions 2 and 3 together are called Applied Statistics. Sometimes position 1 is called Theoretical Statistics. These terminological differences are related to the fact that the above-described development of the considered scientific and applied field not once, not completely and not always adequately reflected in the minds of experts. Meanwhile, there are still textbooks of appropriate level of representation of the mid-twentieth century. The article analyzes the post-war development of the national statistics. We have identified five "growth points": nonparametrics, robustness, bootstrap, statistics of interval data, and statistics of non-numeric data. We have discussed content, development and the basic ideas of statistics of non-numeric data. We have given a number of unresolved problems of theoretical and applied statistics

Ключевые слова: ПРИКЛАДНАЯ СТАТИСТИКА, ТЕОРЕТИЧЕСКАЯ СТАТИСТИКА, МАТЕМАТИЧЕСКАЯ СТАТИСТИКА, СТАТИСТИЧЕСКИЕ МЕТОДЫ, ТОЧКИ РОСТА,

Keywords: APPLIED STATISTICS, THEORETICAL STATISTICS, MATHEMATICAL STATISTICS, STATISTICAL METHODS, GROWTH POINTS,

НЕПАРАМЕТРИЧЕСКАЯ СТАТИСТИКА,
РОБАСТНОСТЬ, БУТСТРЕП, СТАТИСТИКА
ИНТЕРВАЛЬНЫХ ДАННЫХ, СТАТИСТИКА
НЕЧИСЛОВЫХ ДАННЫХ

NONPARAMETRIC STATISTICS,
ROBUSTNESS, BOOTSTRAP, STATISTICS OF
INTERVAL DATA, STATISTICS OF NON-
NUMERICAL DATA

1. Введение

Математические результаты объективны. Теорема либо доказана, либо нет. А вот о значении тех или иных результатов споры возникают. Тем более трудно настаивать на полной объективности выводов, рассуждая о развитии всей статистической науки. Поэтому жанр этой статьи – субъективный анализ состояния и перспектив развития нашей научной области.

Интерес к тематике возник в 1970-х гг., проявившись в создании нами подкомиссии "Статистика объектов нечисловой природы" Научного Совета АН СССР по комплексной проблеме "Кибернетика". Позже на Учредительном съезде Всесоюзной статистической ассоциации (ВСА) в октябре 1990 г. коллеги выбрали меня вице-президентом ВСА (по секции статистических методов).

С 1970-х годов приходилось обдумывать ситуацию в статистике. Конкретные результаты собраны в монографиях [1-3] и др., которые можно рассматривать и как учебники (см. также книги и статьи на сайте «Высокие статистические технологии» <http://orlovs.pp.ru>). Доказательства многих конкретных результатов опубликованы прежде всего в научных журналах "Заводская лаборатория. Диагностика материалов", "Научном журнале КубГАУ", периодических сборниках научных трудов "Статистические методы оценивания и проверки гипотез", "Ученые записки по статистике". А общие соображения – методологическая основа ведущихся работ – рассмотрены в этой статье.

2. Послевоенное развитие отечественной статистики

К 1960-м гг. в нашей стране сформировалась научно-практическая дисциплина, которую называем *классической математической статистикой*. Статистики моего поколения учились теории по книге Г. Крамера [4], написанной в военные годы и впервые изданной у нас в 1948 г. Из прикладных руководств назовем учебник [5] и таблицы с комментариями [6].

Затем внимание многих специалистов сосредоточилось на изучении математических конструкций, используемых в статистике. Примером таких работ является монография [7]. В ней получены продвинутое математические результаты, но нет никаких рекомендаций, которые мог бы использовать статистик, анализирующий конкретные данные наблюдений, измерений, испытаний, анализов, опытов, обследований.

Как реакция на уход в чистую математику значительной части специалистов была выделена новая научная дисциплина - *прикладная статистика*. В учебнике [1] в качестве рубежа, когда это стало очевидным, мы указали 1981 г. – дату выхода массовым тиражом (33 940 экз.) сборника [8], 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 чистую математику, перестав интересоваться практическими делами. Вторая позиционировала себя в качестве науки об обработке данных – результатов наблюдений, измерений, испытаний, анализов, опытов.

Вполне естественно, что в прикладной статистике стали развиваться свои новые математические методы и модели. Необходимость их развития вытекает из потребностей конкретных прикладных исследований. Это математизированное ядро прикладной статистики целесообразно назвать *теоретической статистикой*. Тогда под собственно прикладной статистикой следует понимать обширную промежуточную область между

теоретической статистикой и применением статистических методов в конкретных областях. 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 сбора и анализа данных, т.е. разработки статистических технологий.

Таким образом, общая схема современной статистической науки выглядит следующим образом (от абстрактного к конкретному):

1. Математическая статистика – 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. Она включает в себя методологию статистических методов, вопросы организации выборочных исследований, разработки статистических технологий, создания и использования статистических программных продуктов.

4. Применения статистических методов в конкретных областях (в экономике и менеджменте – эконометрика, в биологии – биометрика, в химии – хеометрия, в технических исследованиях – технометрика, в геологии, демографии, социологии, медицине, 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 уровню представлений середины XX века.

Примечание. Здесь мы уточнили схему внутреннего деления статистической теории, предложенную в [9]. Естественный смысл приобрели термины «теоретическая статистика» и «прикладная статистика» (в узком смысле). Однако необходимо иметь в виду, что в недавнем учебнике [1] 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») полностью оторвалась от задач практики.

Отметим, что математическая статистика, как и теоретическая с прикладной, не имеет ничего общего с ведомственной наукой органов официальной государственной статистики. Эти органы, именовавшиеся последовательно ЦСУ, Госкомстат, Росстат, применяли и применяют лишь проверенные временем приемы позапрошлого века. Возможно, следовало бы от этого ведомства отмежеваться полностью и сменить название дисциплины, например, на «Анализ данных». Такое кардинальное решение, к сожалению, сделано бы неясным положение сотен тысяч публикаций, в названиях которых есть слова "статистика" или "статистический". В настоящее время компромиссным самоназванием нашей научно-практической дисциплины является термин «статистические методы».

Во второй половине 80-х годов развернулось общественное движение, имеющее целью создание профессионального объединения статистиков (об этом движении подробнее рассказано в статье [10]). 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. Некоторую активность проявили созданные на базе ВСА Российская ассоциация статистических методов, Российская академия статистических методов, Белорусская статистическая ассоциация. Пришло время оживить их деятельность. В последние годы большие ежегодные конференции проводятся в Ташкенте (см., например, материалы III научно-практической конференции "Статистика и ее применения" [11]).

В ходе создания ВСА было проанализировано состояние и перспективы развития теоретической и прикладной статистики. Обсудим их.

3. Новые идеи последних десятилетий: точки роста

В работе [12] впервые выделены пять актуальных направлений, в которых развивается современная прикладная статистика, т.е. пять «точек роста» статистической науки: непараметрика, робастность, бутстреп, интервальная статистика, статистика объектов нечисловой природы (см. также [13, 14]). Кратко обсудим эти актуальные направления.

Непараметрика, или 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 [2]. 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 [1]. 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 [15, 16]. Однако эта информация еще не вошла в массовое сознание. До сих пор тупиковой тематике параметрической статистики посвящены обширные разделы учебников и программных продуктов.

Основная идея работ по *робастности*, или *устойчивости*, состоит в том, что выводы, полученные на основе математических методов исследования, должны мало меняться при небольших изменениях исходных данных и отклонениях от предпосылок модели [17, 18]. Здесь есть два круга задач [19, 20]. Один - это изучение устойчивости распространенных алгоритмов анализа данных. Второй - поиск робастных алгоритмов для решения тех или иных задач. Отметим, что сам по себе термин "робастность" не имеет точно определенного смысла. Всегда необходимо указывать конкретную вероятностно-статистическую модель. При этом модель «засорения» Тьюки - Хубера - Хампеля обычно не является практически полезной. Дело в том, что она ориентирована на «утяжеление хвостов», а в реальных ситуациях «хвосты» обрезаются априорными ограничениями на результаты наблюдений, связанными, например, с ограниченностью шкал используемых средств измерения.

Бутстреп - направление непараметрической статистики, опирающееся на интенсивное использование информационных технологий [21]. Основная идея состоит в «размножении выборок», т.е. в получении набора из многих выборок, напоминающих полученную в эксперименте. По такому набору можно непосредственно оценить свойства различных статистических процедур, не прибегая к излишне обременительным (а чаще и необоснованным) семействам вероятностно-статистических моделей. Простейший способ «размножения выборки» состоит в исключении из нее одного результата наблюдения. Исключаем первое наблюдение, получаем выборку, похожую на исходную, но с объемом, уменьшенным на 1. Затем возвращаем исключенный результат первого наблюдения, но исключаем второе наблюдение. Получаем вторую выборку, похожую на исходную. Затем возвращаем результат второго наблюдения, и т.д. Есть и иные способы «размножения выборок». Например, можно по исходной выборке построить ту или иную оценку функции распределения, а затем методом статистических испытаний смоделировать ряд выборок из элементов, функция распределения которых совпадает с этой оценкой. Обобщая, можно сказать, что к настоящему времени в дополнение к классическим инструментам прикладной статистики – предельным теоремам теории вероятностей – добавились новые, основанный на интенсивном использовании компьютеров [22]. Бутстреп – лишь один из таких инструментов. Метод статистических испытаний (Монте-Карло) – вот партнер и конкурент асимптотическим методам математической статистики.

Интервальная статистика - это совокупность методов анализа интервальных статистических данных. Вполне очевидно, что все средства измерения имеют погрешности. Однако до недавнего времени это очевидное обстоятельство никак не учитывалось в статистических процедурах. Только недавно начала развиваться теория интервальной

статистики, в которой предполагается, что исходные данные - это не числа, а интервалы. Интервальную статистику можно рассматривать как часть интервальной математики. Выводы в ней часто принципиально отличны от классических. Основные результаты статистики интервальных данных рассмотрены в статьях [23, 24], а подробные изложения включены в учебники [1, 3, 25], а также в монографию [26].

4. Статистика объектов нечисловой природы

Перейдем к статистике объектов нечисловой природы (она же - статистика нечисловых данных, или нечисловая статистика) [25, 27]. Сначала напомним, что 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 нечисловой статистике элементы выборки - это объекты нечисловой природы, которые нельзя складывать и умножать на числа. Другими словами, объекты нечисловой природы лежат в пространствах, не имеющих векторной структуры.

Примерами объектов нечисловой природы являются:

- значения качественных признаков, в том числе результаты кодировки объектов с помощью заданного перечня категорий (градаций);
- упорядочения (ранжировки) экспертами образцов продукции (при оценке её технического уровня, качества и конкурентоспособности) или заявок на проведение научных работ (при проведении конкурсов на выделение грантов);
- 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;

- графы различных видов (неориентированные, ориентированные, с весами, иерархические структуры и т.п.);

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

- слова, предложения, тексты;

- 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.

Рассмотренные выше интервальные данные тоже можно рассматривать как пример объектов нечисловой природы, а именно, как частный случай нечетких множеств. Если характеристическая функция нечеткого множества равна 1 на некотором интервале и равна 0 вне этого интервала, то задание такого нечеткого множества эквивалентно заданию интервала. Напомним, что *теория нечетких множеств в определенном смысле сводится к теории случайных множеств*. Цикл соответствующих

теорем приведен в работах [17, 28], а также в учебниках [1 - 3], монографии [26].

С 70-х годов в основном на основе запросов теории экспертных оценок [29 - 31] (а также технических исследований, экономики, социологии и медицины) развивались различные направления статистики объектов нечисловой природы. Были установлены основные связи между конкретными видами таких объектов, разработаны для них базовые вероятностные модели. Сводка дана в монографии [17].

Следующий этап (80-е годы) - выделение статистики объектов нечисловой природы в качестве самостоятельной дисциплины в рамках математических методов исследования, ядром которого являются методы статистического анализа данных произвольной природы. Для работ этого периода характерна сосредоточенность на внутренних проблемах нечисловой статистики.

К 90-м годам статистика объектов нечисловой природы 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 и включить их в учебный процесс, что и было сделано (см., например, учебники [1 - 3]).

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

5. Основные идеи статистики объектов нечисловой природы

В чем принципиальная новизна нечисловой статистики? 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. В нечисловой же статистике 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.

Кратко рассмотрим несколько идей, развиваемых в статистике объектов нечисловой природы для данных, 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 [17]. 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 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 [32]. 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, которые могут оказаться и пустыми. Тем не менее удалось сформулировать и доказать законы больших чисел для средних величин, определенных указанным образом, т.е. установить сходимость (в специально определенном смысле) эмпирических средних к теоретическим [1, 2].

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, к которым, как известно, сводится большинство постановок прикладной статистики. В частности, кроме законов больших чисел установлена и состоятельность оценок минимального контраста, в том числе оценок максимального правдоподобия и робастных оценок [25]. К настоящему времени подобные оценки изучены также и в интервальной статистике.

В статистике в пространствах произвольной природы большую роль играют непараметрические оценки плотности, используемые, в частности, в различных алгоритмах регрессионного, дискриминантного, кластерного анализов. В нечисловой статистике предложен и изучен ряд типов непараметрических оценок плотности в пространствах произвольной природы, в том числе в дискретных пространствах. В частности, доказана их состоятельность, изучена скорость сходимости и установлен примечательный факт совпадения наилучшей скорости сходимости в

произвольном пространстве с той, которая имеет быть в классической теории для числовых случайных величин [33, 34].

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 проверки гипотез могут быть использованы статистики интегрального типа, в частности, типа омега-квадрат. Любопытно, что предельная теория таких статистик, построенная первоначально в классической постановке, приобрела естественный (завершенный, изящный) вид именно для пространств произвольного вида, 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 [35].

Представляют практический интерес результаты, связанные с конкретными областями статистики объектов нечисловой природы [36] и соответствующими математико-статистическими моделями порождения нечисловых данных [37], в частности, со статистикой нечетких множеств и со статистикой случайных множеств (напомним, что теория нечетких множеств в определенном смысле сводится к теории случайных множеств), с непараметрической теорией парных сравнений и люсианов (бернуллевских бинарных векторов) [38], с аксиоматическим введением метрик в конкретных пространствах объектов нечисловой природы [39], а также с рядом других конкретных постановок.

Для анализа нечисловых, в частности, экспертных данных весьма важны методы классификации [40]. Констатируем, что наиболее естественно ставить и решать задачи классификации, основанные на

использовании расстояний или показателей различия, в рамках статистики объектов нечисловой природы. Это касается как распознавания образов с учителем (другими словами, дискриминантного анализа), так и распознавания образов без учителя (т.е. кластерного анализа) [41, 42].

6. О нерешенных проблемах теоретической и прикладной статистики

За каждым новым научным результатом открывается многообразие неизвестного. Рассмотрим несколько конкретных постановок.

В статистике в пространствах общей природы получены аналоги классического закона больших чисел. Но нет аналога центральной предельной теоремы. Какова скорость сходимости эмпирических средних к теоретическим? Как сравнить различные способы усреднения? В частности, что лучше применять для усреднения упорядочений – медиану Кемени или среднее по Кемени (среднее отличается от медианы тем, что в качестве показателя различия берется не расстояние Кемени, а его квадрат)? Какие конкретные представители различных классов непараметрических оценок плотности достойны рекомендации для использования в нацеленных на практическое применение алгоритмах анализа нечисловых данных?

До сих пор не проведена классификация классических статистических методов с точки зрения теории измерений. Законченные результаты получены только для теории средних [1 - 3]. Установлено, что для измерений в порядковой шкале в качестве средних можно использовать только порядковые статистики, например, медиану (при нечетном объеме выборки). Среднее арифметическое, столь любимое профанами, применять нельзя. Однако многочисленные эксперименты показывают, что упорядочения объектов по средним арифметическим

рангов и по медианам рангов в подавляющем большинстве случаев совпадают. Нужна теория, объясняющая этот экспериментальный факт.

Все более широкое распространение получает теория нечеткости. Давно установлено, что она в определенном смысле сводится к теории случайных множеств [1 - 3]. Требуется на основе этого сведения проанализировать различные теоретические и прикладные постановки теории нечеткости и рассмотреть их в рамках вероятностно-статистического моделирования.

Перейдем к классическим областям статистики. 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 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 [1, 2].

Итак, методы классической математической статистики обычно используют вне сферы их обоснованной применимости. Какова влияние

отклонений от традиционных предпосылок на статистические выводы? В настоящее время об этом имеются лишь отрывочные сведения. Приведем три примера.

Пример 1. Построение доверительного интервала для математического ожидания обычно проводят с использованием распределения Стьюдента (при справедливости гипотезы нормальности). Как следует из Центральной предельной теоремы (ЦПТ) теории вероятностей, в асимптотике (при большом объеме выборки) такие расчетные методы дают правильные результаты (из ЦПТ вытекает использование квантилей нормального распределения, а из классической теории - квантилей распределения Стьюдента, но при росте объема выборки квантили распределения Стьюдента стремятся к соответствующим квантилям нормального распределения).

Пример 2. Для проверки однородности двух независимых выборок (на самом деле - для проверки равенства математических ожиданий) обычно рекомендуют использовать двухвыборочный критерий Стьюдента. Предпосылки его использования – это нормальность распределений, соответствующих выборкам, и равенство их дисперсий. Что будет при отклонении от нормальности распределений, из которых взяты выборки, от нормальности? Если объемы выборок равны или если дисперсии совпадают, то в асимптотике (когда объемы выборок безгранично возрастают) классический метод является корректным. Если же объемы выборок существенно отличаются или дисперсии различны, то критерий Стьюдента проверки гипотезы однородности применять нельзя, поскольку распределение двухвыборочной статистики Стьюдента будет существенно отличаться от классического. Поскольку проверка равенства дисперсий - более сложная задача, чем проверка равенства математических ожиданий, то для выборок разного объема использовать двухвыборочную статистику

Стьюдента не следует, целесообразно применять критерий Крамера-Уэлча [1, 2, 43].

Пример 3. В задаче отбраковки (исключения) резко выделяющихся наблюдений (выбросов) расчетные методы, основанные на нормальности, являются крайне неустойчивыми по отношению к отклонениям от нормальности, что полностью лишает эти методы научной обоснованности [1, 2].

Примеры 1 - 3 показывают весь спектр возможных свойств классических расчетных методов в случае отклонения от нормальности. Методы примера 1 оказываются вполне пригодными при таких отклонениях, примера 2 - пригодными в некоторых случаях, примера 3 - полностью непригодными.

Итак, 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 ЦПТ, 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.

Почему необходимо изучение классических алгоритмов, а не построение новых, специально предназначенных для работы в условиях отклонения от классических предпосылок?

Во-первых, потому, что классические алгоритмы в настоящее время наиболее распространены (благодаря сложившейся системе образования прикладников). Например, в научных медицинских исследованиях для <http://ej.kubagro.ru/2016/01/pdf/13.pdf>

проверки однородности двух независимых выборок традиционно используют критерий Стьюдента, при этом условия его применимости не проверяют. Насколько обоснованными являются выводы? Как следует из примера 2, во многих случаях выводы нет оснований подвергать сомнению, хотя они получены с помощью некорректной процедуры.

Во-вторых, более новые подходы зачастую методологически уязвимы. Так, известная робастная модель засорения Тьюки - Хубера нацелена на борьбу с большими выбросами, которые зачастую физически невозможны из-за ограниченности интервала значений измеряемой характеристики, в котором работает конкретное средство измерения. Следовательно, модель Тьюки - Хубера - Хампеля имеет скорее теоретическое значение, чем практическое. Сказанное, конечно, не обозначает, что следует прекратить разработку, изучение и внедрение непараметрических и устойчивых методов, выделенных выше как «точки роста» современной прикладной статистики.

Нерешенным проблемам теоретической и прикладной статистики посвящены статьи [44, 45]. 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 предложить 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 [46]. Естественных подходов к сравнению критериев несколько - на основе асимптотической относительной эффективности по Бахадуру, Ходжесу - Леману, Питмену. И каждый критерий 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 [46], не позволяет дать рекомендации для выбора критерия проверки однородности при анализе реальных данных.

Проблемы разработки высоких статистических технологий поставлены в [47, 48] (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. Необходима теория, позволяющая изучать свойства статистических технологий и так их конструировать, чтобы обеспечить высокое качество обработки данных.

Литература

1. Орлов А.И. Прикладная статистика. - М.: Экзамен, 2006. - 672 с.
2. Орлов А.И. Эконометрика. Изд. 3-е, переработанное и дополненное. - М.: Экзамен, 2004. - 576 с.
3. Орлов А.И. Теория принятия решений. - М.: Экзамен, 2006. - 576 с.
4. Крамер Г. Математические методы статистики. - М.: Мир, 1975. - 648 с.

5. 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.
6. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики / 3-е изд.- М.: Наука, 1983. - 416 с. (1-е изд. – 1965).
7. Каган А.М., Линник Ю.В., Рао С.Р. Характеризационные задачи математической статистики. - М.: Наука, 1972. - 656 с.
8. Современные проблемы кибернетики (прикладная статистика). - М.: Знание, 1981. – 64 с.
9. Орлов А.И. О перестройке статистической науки и её применений. – Журнал «Вестник статистики». 1990. №1. С.65 – 71.
10. Орлов А.И. Непараметрическая и прикладная статистика в нашей стране // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 101. С. 197–226.
11. Материалы республиканской научно-практической конференции "Статистика и ее применения - 2015". Под редакцией профессора А.А. Абдушукурова. - Ташкент: НУУз, 2015. - 500 с.
12. Орлов А.И. Современная прикладная статистика // Заводская лаборатория. Диагностика материалов. 1998. Т.64. № 3. С. 52-60.
13. Орлов А.И. Точки роста статистических методов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 103. С. 136–162.
14. Орлов А.И., Луценко Е.В., Лойко В.И. Перспективные математические и инструментальные методы контроллинга. Под научной ред. проф. С.Г. Фалько. Монография (научное издание). – Краснодар, КубГАУ. 2015. – 600 с.
15. Орлов А.И. Современное состояние непараметрической статистики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 106. С. 239 – 269.
16. Орлов А.И. Структура непараметрической статистики (обобщающая статья) // Заводская лаборатория. Диагностика материалов. 2015. Т.81. №7. С. 62-72.
17. Орлов А.И. Устойчивость в социально-экономических моделях. - М.: Наука, 1979. - 296 с.
18. Орлов А.И. Устойчивые экономико-математические методы и модели. - Saarbrücken (Germany), Lambert Academic Publishing, 2011. - 436 с.
19. Орлов А.И. Устойчивые математические методы и модели // Заводская лаборатория. Диагностика материалов. 2010. Т.76. №3. С.59-67.
20. Орлов А.И. Новый подход к изучению устойчивости выводов в математических моделях // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 100. С. 146-176.
21. Орлов А.И. О реальных возможностях бутстрепа как статистического метода // Заводская лаборатория. Диагностика материалов. 1987. Т.53. №10. С.82-85.
22. Орлов А.И. Компьютерно-статистические методы: состояние и перспективы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 103. С. 163 – 195.
23. Орлов А.И. Основные идеи статистики интервальных данных // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 94. С. 55-70.
24. Орлов А.И. Статистика интервальных данных (обобщающая статья) // Заводская лаборатория. Диагностика материалов. 2015. Т.81. №3. С. 61 - 69.

25. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч. 1. Нечисловая статистика. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2009. — 541 с.
26. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с.
27. Орлов А.И. О развитии статистики объектов нечисловой природы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 93. С. 41-50.
28. Орлов А.И. Теория нечетких множеств – часть теории вероятностей // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 92. С. 51-60.
29. Орлов А.И. Экспертные оценки // Заводская лаборатория. Диагностика материалов. 1996. Т.62. №1. С.54-60.
30. Орлов А.И. Организационно-экономическое моделирование: учеб. Ч.2. Экспертные оценки. - М.: Изд-во МГТУ им. Н.Э. Баумана, 2011. - 486 с.
31. Орлов А.И. Теория экспертных оценок в нашей стране // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 93. С. 1-11.
32. 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.
33. 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.
34. 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.
35. Орлов А.И. Предельная теория непараметрических статистик // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 100. С. 31-52.
36. Орлов А.И. Многообразие объектов нечисловой природы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 102. С. 32 – 63.
37. Орлов А.И. Вероятностные модели порождения нечисловых данных // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 105. С. 39–66.
38. Орлов А.И. Теория люсианов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 101. С. 275 – 304.
39. Орлов А.И. Расстояния в пространствах статистических данных // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета, 2014. № 101. С. 227 – 252.
40. Орлов А.И. Математические методы теории классификации // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 95. С. 23 – 45.
41. Орлов А.И. Прогностическая сила – наилучший показатель качества алгоритма диагностики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 99. С. 33–49.

42. Орлов А.И. Базовые результаты математической теории классификации // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 110. С. 219–239.
43. Орлов А.И. Проверка статистической гипотезы однородности математических ожиданий двух независимых выборок: критерий Крамера-Уэлча вместо критерия Стьюдента // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 110. С. 197–218.
44. Загоруйко Н.Г., Орлов А.И. Некоторые нерешенные математические задачи прикладной статистики // Современные проблемы кибернетики (прикладная статистика). - М.: Знание, 1981. - С.53-63.
45. Орлов А.И. Некоторые нерешенные вопросы в области математических методов исследования // Заводская лаборатория. Диагностика материалов. 2002. Т.68. №3. С. 52-56.
46. Никитин Я.Ю. Асимптотическая эффективность непараметрических критериев. - М.: Наука, 1995. - 240 с.
47. Орлов А.И. Высокие статистические технологии // Заводская лаборатория. Диагностика материалов. 2003. Т.69. №11. С. 55-60.
48. Орлов А.И. О высоких статистических технологиях // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 105. С. 14–38.

References

1. Orlov A.I. Prikladnaja statistika. - M.: Jekzamen, 2006. - 672 s.
2. Orlov A.I. Jekonometrika. Izd. 3-e, pererabotannoe i dopolnennoe. - M.: Jekzamen, 2004. – 576 s.
3. Orlov A.I. Teorija prinjatija reshenij.– M.: Jekzamen, 2006. – 576 s.
4. Kramer G. Matematicheskie metody statistiki. - M.: Mir, 1975. - 648 s.
5. Smirnov N.V., Dunin-Barkovskij I.V. Kurs teorii verojatnostej i matematicheskoj statistiki dlja tehniceskix prilozhenij. Izd. 3-e, stereotipnoe. – M.: Nauka, 1969. – 512 s.
6. Bol'shev L.N., Smirnov N.V. Tablicy matematicheskoj statistiki / 3-e izd.- M.: Nauka, 1983. - 416 s. (1-e izd. – 1965).
7. Kagan A.M., Linnik Ju.V., Rao S.R. Harakterizacionnye zadachi matematicheskoj statistiki. - M.: Nauka, 1972. - 656 s.
8. Sovremennye problemy kibernetiki (prikladnaja statistika). - M.: Znanie, 1981. – 64 s.
9. Orlov A.I. O perestrojke statisticheskoj nauki i ejo primenenij. – Zhurnal «Vestnik statistiki». 1990. No.1. S.65 – 71.
10. Orlov A.I. Neparametricheskaja i prikladnaja statistika v nashej strane // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 101. S. 197–226.
11. Materialy respublikanskoj nauchno-prakticheskoy konferencii "Statistika i ee primeneniya - 2015". Pod redakciej professora A.A. Abdushukurova. - Tashkent: NUUZ, 2015. - 500 s.
12. Orlov A.I. Sovremennaja prikladnaja statistika // Zavodskaja laboratorija. Diagnostika materialov. 1998. T.64. № 3. S. 52-60.

13. Orlov A.I. Tochki rosta statisticheskikh metodov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 103. S. 136–162.
14. Orlov A.I., Lucenko E.V., Lojko V.I. Perspektivnye matematicheskie i instrumental'nye metody kontrollinga. Pod nauchnoj red. prof. S.G. Fal'ko. Monografija (nauchnoe izdanie). – Krasnodar, KubGAU. 2015. – 600 s.
15. Orlov A.I. Sovremennoe sostojanie neparametricheskoj statistiki // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2015. № 106. S. 239 – 269.
16. Orlov A.I. Struktura neparametricheskoj statistiki (obobshhajushhaja stat'ja) // Zavodskaja laboratorija. Diagnostika materialov. 2015. T.81. №7. S. 62-72.
17. Orlov A.I. Ustojchivost' v social'no-jekonomicheskikh modeljah. - M.: Nauka, 1979. - 296 s.
18. Orlov A.I. Ustojchivye jekonomiko-matematicheskie metody i modeli. - Saarbrücken (Germany), Lambert Academic Publishing, 2011. - 436 s.
19. Orlov A.I. Ustojchivye matematicheskie metody i modeli // Zavodskaja laboratorija. Diagnostika materialov. 2010. T.76. №3. S.59-67.
20. Orlov A.I. Novyj podhod k izucheniju ustojchivosti vyvodov v matematicheskikh modeljah // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 100. S. 146-176.
21. Orlov A.I. O real'nyh vozmozhnostjah butstropa kak statisticheskogo metoda // Zavodskaja laboratorija. Diagnostika materialov. 1987. T.53. №10. S.82-85.
22. Orlov A.I. Komp'juterno-statisticheskije metody: sostojanie i perspektivy // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 103. S. 163 – 195.
23. Orlov A.I. Osnovnye idei statistiki interval'nyh dannyh // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2013. № 94. S. 55-70.
24. Orlov A.I. Statistika interval'nyh dannyh (obobshhajushhaja stat'ja) // Zavodskaja laboratorija. Diagnostika materialov. 2015. T.81. №3. S. 61 - 69.
25. Orlov A.I. Organizacionno-jekonomicheskoe modelirovanie : uchebnik : v 3 ch. Ch. 1. Nechislovaja statistika. – M.: Izd-vo MGTU im. N.Je. Baumana, 2009. — 541 s.
26. Orlov A.I., Lucenko E.V. Sistemnaja nechetkaja interval'naja matematika. Monografija (nauchnoe izdanie). – Krasnodar, KubGAU. 2014. – 600 s.
27. Orlov A.I. O razvitii statistiki ob#ektov nechislovoj prirody // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2013. № 93. S. 41-50.
28. Orlov A.I. Teorija nechetkih mnozhestv – chast' teorii verojatnostej // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2013. № 92. S. 51-60.
29. Orlov A.I. Jekspertnye ocenki // Zavodskaja laboratorija. Diagnostika materialov. 1996. T.62. №1. S.54-60.
30. Orlov A.I. Organizacionno-jekonomicheskoe modelirovanie: ucheb. Ch.2. Jekspertnye ocenki. - M.: Izd-vo MGTU im. N.Je. Baumana, 2011. - 486 s.
31. Orlov A.I. Teorija jekspertnyh ocenok v nashej strane // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2013. № 93. S. 1-11.
32. Orlov A.I. Srednie velichiny i zakony bol'shix chisel v prostranstvah proizvol'noj prirody // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2013. № 89. S. 175-200.

33. Orlov A.I. Ocenki plotnosti raspredelenija verojattnostej v prostranstvah proizvol'noj prirody // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 99. S. 15-32.
34. Orlov A.I. Predel'nye teoremy dlja jadernyh ocenok plotnosti v prostranstvah proizvol'noj prirody // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2015. № 108. S. 316 – 333.
35. Orlov A.I. Predel'naja teorija neparametricheskikh statistik // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 100. S. 31-52.
36. Orlov A.I. Mnogoobrazie ob#ektov nechislovoj prirody // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 102. S. 32 – 63.
37. Orlov A.I. Verojattnostnye modeli porozhdenija nechislovyh dannyh // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2015. № 105. S. 39–66.
38. Orlov A.I. Teorija l'jusjanov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 101. S. 275 – 304.
39. Orlov A.I. Rasstojanija v prostranstvah statisticheskikh dannyh // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta, 2014. № 101. S. 227 – 252.
40. Orlov A.I. Matematicheskie metody teorii klassifikacii // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 95. S. 23 – 45.
41. Orlov A.I. Prognosticheskaja sila – nailuchshij pokazatel' kachestva algoritma diagnostiki // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 99. S. 33–49.
42. Orlov A.I. Bazovye rezul'taty matematicheskij teorii klassifikacii // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2015. № 110. S. 219–239.
43. Orlov A.I. Proverka statisticheskij gipotezy odnorodnosti matematicheskikh ozhidaniy dvuh nezavisimyh vyborok: kriterij Kramera-Ujelcha vmesto kriterija St'judenta // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2015. № 110. S. 197–218.
44. Zagorujko N.G., Orlov A.I. Nekotorye nereshennye matematicheskie zadachi prikladnoj statistiki // Sovremennye problemy kibernetiki (prikladnaja statistika). - M.: Znanie, 1981. - S.53-63.
45. Orlov A.I. Nekotorye nereshennye voprosy v oblasti matematicheskikh metodov issledovanija // Zavodskaja laboratorija. Diagnostika materialov. 2002. T.68. №3. S. 52-56.
46. Nikitin Ja.Ju. Asimptoticheskaja jeffektivnost' neparametricheskikh kriteriev. - M.: Nauka, 1995. - 240 s.
47. Orlov A.I. Vysokie statisticheskie tehnologii // Zavodskaja laboratorija. Diagnostika materialov. 2003. T.69. №11. S. 55-60.
48. Orlov A.I. O vysokih statisticheskikh tehnologijah // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2015. № 105. S. 14–38.