

Математические методы исследования

УДК 519.28

НОВАЯ ПАРАДИГМА ПРИКЛАДНОЙ СТАТИСТИКИ

© А. И. Орлов¹

Статья поступила 16 марта 2011 г.

Показана необходимость смены старой парадигмы математической статистики на новую. Эта смена основана на переходе от параметрических статистических методов к непараметрическим, от числовых данных — к нечисловым, на интенсивном использовании информационных технологий. Показаны отличительные черты и преимущества новой парадигмы прикладной статистики перед старой парадигмой математической статистики, широко используемой с середины XX века.

Ключевые слова: прикладная статистика; развитие статистики; сравнение новой и старой парадигм; последствия перехода на новую парадигму.

Прикладная статистика — это наука о том, как обрабатывать данные [1]. В середине XX в. наука об обработке данных называлась математической статистикой. За сменой названий стоит смена парадигм. Под парадигмой понимаем систему идей, взглядов и понятий, различных моделей решения проблем, а также методов исследования.

В данной работе приведено описание и дано сравнение двух парадигм в области анализа данных — старой и новой.

Послевоенное развитие отечественной статистики

К середине XX в. в нашей стране, как и за рубежом, сформировалась научно-практическая дисциплина, называемая классической *математической статистикой*. Специалисты-статистики учились теории по книге Г. Крамера [2], написанной в военные годы и впервые изданной у нас в 1948 г. Из прикладных руководств стоит отметить учебник [3] и таблицы с комментариями [4].

Затем внимание многих специалистов сосредоточилось на изучении математических конструкций, используемых в статистике. Например, в монографии [5] получены интересные математические результаты, но из них трудно (почти невозможно) выделить рекомендации для статистика, анализирующего конкретные данные. Причиной такого сдвига интересов послужили некоторые особенности исторического развития нашей страны.

Большой вред развитию статистической науки и практики нанесло Всесоюзное совещание статистиков 1954 г. На нем было принято, что статистика — это одна из экономических наук, фактически — ведомственная наука ЦСУ — Госкомстата — Росстата (Федеральной службы государственной статистики). При этом организаторы совещания не посмели покуситься на само существование математической статистики, но отнесли ее исключительно к математике, в которой была выделена специальность «теория вероятностей и математическая статистика». Все остальные области применения статистических методов перестали замечаться официальными структурами, т.е. стали полуполевыми. Конечно, специалисты нашли способы противодействия. Например, статистические методы в химии относились к «химической кибернетике», в медицине — к «математическому моделированию в медицине».

В результате решений Всесоюзного совещания статистиков 1954 г. работы по математической статистике стали рассматриваться исключительно с позиций математики. Начали цениться теоремы (типа приведенных в монографии [5]), никак не связанные с анализом реальных данных. В то же время вопросы практики применения статистических методов стали отодвигаться на задний план и даже подвергаться гонениям. Типичным примером является провал при защите на мехмате МГУ им. М. В. Ломоносова в 1971 г. докторской диссертации В. В. Федорова, в которой были получены базовые результаты в области планирования эксперимента — одного из важнейших направлений статистических методов.

Как реакция на уход в математику выделилась новая научная дисциплина — *прикладная статистика*.

¹ Институт высоких статистических технологий и эконометрики Московского государственного технического университета им. Н. Э. Баумана, Москва, Россия;
e-mail: prof-orlov@mail.ru

В качестве рубежа, когда это стало очевидным, мы указали 1981 г. [1] — дату выхода массовым тиражом (33940 экз.) сборника [6], в названии которого использован термин «прикладная статистика» [полное название: «Современные проблемы кибернетики (прикладная статистика)»]. С этого времени пути развития математической статистики и прикладной статистики окончательно разошлись. Первая из этих дисциплин полностью ушла в математику, потеряв интерес к практическим делам. Вторая позиционировала себя в качестве науки об обработке данных — результатов наблюдений, измерений, испытаний, анализов, опытов.

Вполне естественно, что в прикладной статистике стали развиваться свои математические методы и модели. Необходимость их появления следовала из потребностей конкретных прикладных исследований. Это математизированное ядро прикладной статистики вполне естественно 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. Именно прикладной статистике посвящены многие публикации в разделе «Математические методы исследования» нашего журнала.

Таким образом, общая схема современной статистики выглядит, по нашему мнению, 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. *Применение статистических методов* в конкретных областях: в экономике и менеджменте (эконометрика); биологии (биометрика); химии (хеометрия); технических исследованиях (технометрика); 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 ранее [7]. 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») заметно оторвалась от задач практики.

Отметим, что математическая статистика, как и теоретическая с прикладной, заметно отличается от ведомственной науки органов официальной государственной статистики. Центральное статистическое управление (ЦСУ), Госкомстат, Росстат применяли и применяют лишь проверенные временем приемы позапрошлого (XIX) века. Приходится с сожалением констатировать, что большинство учебных курсов в экономических вузах и учебников с названиями «Статистика» или «Общая теория статистики» истолковывают эти понятия в смысле Росстата. Возможно, следовало бы от этого ведомства полностью отмежеваться и сменить название дисциплины, например на «Анализ данных». В настоящее время компромиссным самоназванием рассматриваемой научно-практической дисциплины является термин «*статистические методы*».

Во второй половине 80-х годов 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 (ВСА) [8] оказалась парализованной в результате развала СССР. Некоторую активность проявили Российская ассоциация статистических методов, Российская академия статистических методов, Белорусская статистическая ассоциация, созданные на базе ВСА.

В ходе создания ВСА проанализированы состояние и перспективы развития статистических методов. Коллективными усилиями создана *новая парадигма прикладной статистики*, основанная, в частности, на переходе от параметрической статистики, использу-

ющей параметрические семейства распределений, к непараметрической и нечисловой статистике. Выявлена необходимость создания нового поколения учебной литературы, которая должна сменить издания на основе идей середины XX в.

Какова ситуация в других странах, в мире в целом? Отставание от англо-саксонских стран более чем на 150 лет при создании профессионального объединения статистиков отнюдь не случайно. За эти годы был пройден долгий путь организационного строительства, найдены способы эффективного сочетания теории и практики, объединения статистиков различных направлений — от теоретиков-вероятностников до работников официальной государственной статистики. Центральным ядром была и остается математическая статистика. Есть и развитие: основной статистический журнал «The Annals of Mathematical Statistics», созданный в 1930 г., в 1973 г. был разделен на два — «The Annals of Statistics» и «The Annals of Probability». Общее число научных журналов, публикующих работы по теории вероятностей, математической и прикладной статистике, составляет около 1500 [9].

Статистические методы основаны на теории вероятностей. В нашей стране была создана самая мощная в мире научная школа по теории вероятностей. К сожалению, это стратегическое преимущество не удалось реализовать в виде адекватного развития статистических методов. Число специалистов (на миллион граждан страны) у нас на порядок меньше, чем в

других развитых странах. Если бы удалось ввести в средней школе полноценный курс вероятности и статистики, который имеется в Японии, США, Швейцарии, Кении и Ботсване, почти во всех странах, кроме стран бывшего социалистического лагеря [10], то ситуация могла бы резко улучшиться. Такой курс нужно строить на высоких статистических технологиях [11], а не на устаревших, т.е. он должен отражать современные достижения, а не концепции пятидесятилетней или столетней давности.

Новая парадигма прикладной статистики

Демонстрацией необходимости новой парадигмы является само появление новой научной области — прикладной статистики. Сравнение парадигм удобно провести с помощью таблицы, в которой выделены 17 основных характеристик систем идей, взглядов и понятий.

Основные составляющие новой парадигмы подробно разбирались на страницах нашего журнала. Проанализировано развитие нечисловой статистики за 30 лет [12]. Рассмотрены пять актуальных направлений, в которых развивается современная прикладная статистика, т.е. пять «точек роста» — непараметрика, робастность, бутстреп, статистика интервальных данных, статистика объектов нечисловой природы [13], проблемы устойчивости (робастности) [14]. При постановках задач, построении систем моделей большую роль стала играть методология [15, 16]. Стало

Характеристики новой и старой парадигм

Характеристика	Парадигма	
	старая	новая
Типовые исходные данные	Числа, вектора, функции	Объекты нечисловой природы
Основной подход к описанию данных	Распределения из параметрических семейств	Произвольные (непрерывные) функции распределения
Математический аппарат	Суммы	Расстояния и алгоритмы оптимизации
Источник постановок задач	Математические традиции, сформировавшиеся к середине XX века	Современные потребности анализа данных (XXI век)
Отношение к вопросам устойчивости выводов	Отсутствие интереса	Развитая теория устойчивости (робастности)
Оцениваемые величины	Параметры распределений	Характеристики и плотности распределений, зависимости и др.
Возможность применения	Наличие повторяющегося комплекса условий	Наличие обоснованной вероятностно-статистической модели
Центральная часть теории	Статистика числовых случайных величин	Статистика в пространствах произвольной природы
Роль информационных технологий	Только для расчета таблиц	Инструмент получения выводов (датчики псевдослучайных чисел, размножение выборок, в том числе бутстреп, и др.)
Отношение к информатике	Находится вне статистики	Учет возможностей проведения расчетов
Точность данных	Данные полностью известны	Учет свойств данных, в частности интервальных
Типовые результаты	Предельные теоремы	Рекомендации для конкретных объемов выборок
Вид постановок задач	Отдельные задачи	Статистические технологии
Стыковка алгоритмов	Не рассматривается	Весьма важна
Роль моделирования	Отдельные системы аксиом	Системы моделей
Анализ экспертных оценок	Отдельные алгоритмы	Прикладное «зеркало» общей теории
Роль методологии	Практически отсутствует	Основополагающая

ясно, что надо разделять четыре уровня работ: *методологический* — рассмотрение основных идей метода, а также решение вопросов, как ставить задачу (как организовать моделирование); *теоретический* — проработка основных идей, доказательство теорем; *методический* — разработка методик, алгоритмов, программных продуктов, рекомендаций по практическому использованию; *прикладной* — практическое использование (адаптация и применение разработанного метода при решении конкретных практических задач).

Особенно важным представляется разделение методологического и теоретического уровней. Например, на методологическом уровне формулируется утверждение о том, что распределение суммы независимых случайных величин близко к нормальному. На теоретическом уровне этому утверждению соответствует цепь теорем, полученных с начала XVII в. по 30-е годы XX в. (речь идет об истории Центральной предельной теоремы).

Была определена роль статистики в пространствах произвольной природы [12, 17], позволяющей единообразно анализировать как результаты измерений, наблюдений, испытаний, анализов, опытов, так и экспертные оценки разных видов [18, 19]. В частности, оказалось, что задачи классификации [20] наиболее естественно ставить и решать в рамках статистики в пространствах произвольной природы и тем самым относить их к нечисловой статистике [17], а не к многомерному статистическому анализу.

Принципиально важным является понятие «высокие статистические технологии» [11], вводящее в прикладную статистику производственный процесс анализа данных, состоящий из последовательностей операций, каждая из которых в отдельности (оценивание, проверка гипотезы и пр. операции) используется и в математической статистике. Важной оказалась проблема «стыковки алгоритмов», т.е. проверки условий применимости последующей статистической операции (алгоритма) после окончания выполнения предыдущей.

Например, перед проведением регрессионного анализа иногда рекомендуют разбить совокупность данных на однородные части, т.е. провести классификацию. Однако после классификации выделенные подсовокупности сосредоточены в отдельных частях исходного пространства, т.е. моделирующие их плотности распределения равны нулю вне соответствующих частей, а потому не могут иметь нормальное распределение. Следовательно, к полученным подсовокупностям результатов измерений (наблюдений, испытаний, анализов, опытов) принципиально нельзя применять методы регрессионного анализа, предполагающие нормальность погрешностей (ошибок, невязок). Следует использовать непараметрические методы регрессии, не опирающиеся на заведомо невыполненную в рассматриваемом случае гипотезу нормальности погрешностей.

За рубежом по каждому из перечисленных узких направлений ведутся научные исследования. Однако в единое целое — в новую научную парадигму — они интегрированы именно в нашей стране. Важно эффективно использовать это конкурентное преимущество — общее достояние российского научного сообщества.

Последствия перехода на новую парадигму

В качестве примера рассмотрим подготовку специалистов, поскольку именно она определяет уровень выполняемых ими в дальнейшем научных и прикладных работ. При переходе на преподавание согласно новой парадигме прикладной статистики необходимо существенно изменить содержание традиционного для технических вузов курса «Теория вероятностей и математическая статистика». В частности, необходимо изучать: случайные величины (точнее, случайные элементы) со значениями в произвольных пространствах (в пространстве бинарных отношений, конечных множеств, других нелинейных пространствах); центральные предельные теоремы в полном объеме — для разнораспределенных слагаемых, в многомерном пространстве; средние величины в произвольных пространствах как решения оптимизационных задач; законы больших чисел в произвольных пространствах; непараметрический подход ко всем основным задачам прикладной математической статистики. В то же время почти отпадает необходимость в изучении таких традиционных тем, как геометрические вероятности, параметрические семейства распределений (за исключением нормального распределения, появляющегося в центральных предельных теоремах), параметрические постановки в математической статистике, достаточные статистики, неравенство Рао-Крамера, метод максимального правдоподобия, метод одношаговых оценок, проверка параметрических гипотез с использованием распределений Стьюдента и Фишера.

Ряд проблем связан с использованием при преподавании распространенных программных продуктов. Очевидно, что математические методы исследования, в том числе методы статистического анализа данных, требуют больших вычислений и зачастую невозможны без компьютеров. Применение новой парадигмы прикладной статистики, а также высоких статистических технологий [11] предполагает использование соответствующих программных продуктов. Статистические пакеты — постоянно используемые интеллектуальные инструменты исследователей, инженеров, управленцев, занимающихся анализом больших массивов данных. Более 20 статистических пакетов, разработанных под нашим руководством Всесоюзным центром статистических методов и информатики, в том числе пакеты СПК, АТСТАТ-ПРП, СТАТКОН, АВРОРА-РС, ЭКСПЛАН, ПАСЭК, НАДИС, проанализированы в работах [21, 22]. Сравнительному анализу четырех диалоговых систем по статистическому

контролю посвящена статья [23]. Однако наряду с очевидной пользой статистические пакеты могут принести вред неискушенному пользователю. Например, в них зачастую пропагандируется применение двухвыборочного критерия Стьюдента, когда условия его применимости не проверены, а зачастую и не выполнены. Между тем хорошо известно, каковы отрицательные последствия использования критерия Стьюдента вне сферы его применимости, а также и то, что применять его нет необходимости, поскольку разработаны более адекватные критерии [24].

Другой пример: малограмотность переводчиков в русифицированной версии MS Excel (по крайней мере в разделе «Анализ данных»), где «объем выборки» именуется «счет». С сожалением приходится констатировать, что не соответствует современным требованиям и электронный учебник — обзор методов, реализованных в пакете STATISTICA-6. Анализ допущенных в документации к пакету недочетов занял бы не меньше места, чем сама документация. В работе [21] продемонстрировано, насколько трудоемким оказался критический анализ всего лишь нескольких десятков стандартов по статистическим методам управления качеством.

Замечания касаются, конечно, не только пакетов. Из одной публикации в другую кочуют одни и те же ошибки. Например, распространенная ошибка при использовании критериев Колмогорова и омега-квадрат рассмотрена в статье [25], ошибочные утверждения о том, какие гипотезы можно проверять с помощью двухвыборочного критерия Вилкоксона, разобраны в работе [26].

Основное противоречие в области разработки статистических пакетов на настоящий момент таково. Те, кто программирует, не являются специалистами по прикладной статистике, поскольку это не входит в их профессиональные обязанности. Специалисты же по статистическим методам не берутся реализовывать их в пакетах, поскольку это весьма трудоемкая и ответственная работа, обычно не соответствующая их профессиональным устремлениям. Судя по опыту Всесоюзного центра статистических методов и информатики, стоимость разработки (на профессиональном уровне) пакета среднего уровня сложности составляет порядка 70 тыс. руб. (в ценах 1990 г.), что соответствует 10,5 млн руб. в ценах 2011 г. (индекс инфляции за 21 год равен 150 при расчете по методике [27]). Это означает, что разработкой, распространением и сопровождением статистических пакетов должны заниматься специализирующиеся на этом организации или подразделения, в том числе коммерческие, как это и происходит во всем мире.

В нашей стране активная работа по созданию развернутой системы отечественных статистических пакетов развернулась в 80-х годах. Как уже отмечалось, только Всесоюзным центром статистических методов и информатики было разработано более 20 программ-

ных продуктов по прикладной статистике и другим статистическим методам. Эта работа проводилась в рамках более широкого проекта, нацеленного на объединение усилий специалистов по статистическим методам с целью повышения эффективности теоретических и прикладных исследований. Важным промежуточным итогом было создание в 1990 г. Всесоюзной организации по статистическим методам и их применениям и Всесоюзной статистической ассоциации [8]. Планы тех лет отражены в статье [28]. Итогом виделось создание (развертывание, организационное оформление) новой отрасли прикладной науки по образцу метрологии.

Развал СССР, либерализация цен и гиперинфляция начала 90-х положили конец рассматриваемому проекту. Из плана работ реализована только подготовка серии современных учебников [1, 17, 19, 29, 30 и др.], составленных на основе статей, опубликованных в журнале «Заводская лаборатория» (учебники выложены в свободном доступе на сайте «Высокие статистические технологии» <http://orlovs.pp.ru> и странице «Лаборатории экономико-математических методов в контроллинге МГТУ им. Н. Э. Баумана» <http://ibm.bmstu.ru/nil/biblio.html>). Предприятия и организации, лишившись оборотных средств из-за инфляции, перестали покупать статистические программные продукты, коллективы разработчиков распались, перестали поддерживать статистические пакеты в условиях быстрого обновления технических средств и базового программного обеспечения. В результате многообразие продуктов на отечественном рынке статистических пакетов резко сократилось и монополистами оказались SPSS, STATISTICA, STATGRAPHICS (и немногие другие). При анализе программных продуктов целесообразно сказать несколько слов об аналитических надстройках над такими распространенными системами широкого назначения, как, например, Oracle и т.п. Приходится констатировать, что в них примитивная статистика сочетается с хорошей визуализацией, облегчающей практическое использование подобных систем и при этом, увы, дающей импульс распространению устаревших статистических методов.

На опасность бездумного применения статистических пакетов В. В. Налимов обращал внимание еще около 40 лет назад [31]. Он имел в виду прежде всего склонность к проведению расчетов без знакомства с сутью применяемых методов. Необходимо обратить внимание также на научно-технический уровень самих пакетов и сопровождающей документации. Дополнительно к сказанному ранее приходится добавить, что в популярных в настоящее время в России статистических пакетах нет примерно половины того, что разработано представителями отечественной вероятностно-статистической научной школы и включено в современные учебники [1, 17, 27], подготовленные в соответствии с рекомендациями Все-

союзной статистической ассоциации, а позже — Российской ассоциации статистических методов. Сказанное легко проверить, сопоставив содержание указанных учебников и перечень методов, включенных в распространенные пакеты. Поэтому в научно-учебном комплексе «Инженерный бизнес и менеджмент» МГТУ им. Н. Э. Баумана мы сознательно избегаем использования в учебном процессе популярных в настоящее время пакетов, чтобы не приучать студентов к статистике 60 – 70-х годов прошлого века. Однако поскольку нет современных пакетов, приходится для практических расчетов использовать устаревшие программные продукты.

Тиражи пакетов и учебников сопоставимы. Пакет STATGRAPHICS имеет более 40 тыс. зарегистрированных пользователей, учебник [1] выпущен суммарным тиражом 3 тыс. экземпляров, его электронную версию только с сайта «Высокие статистические технологии» на 01.05.2011 скачали 35 тыс. пользователей (<http://forum.orlovs.pp.ru/viewtopic.php?p=4319#4319>). Поэтому состав пакетов и качество документации имеют большое значение. Они во многом определяют качество прикладных научных работ и обоснованность хозяйственных решений.

По сравнению с 80-ми годами к настоящему времени возросло внимание не только к конкретным методам обработки данных, но и к статистическим технологиям [11]. В этом суть популярного ныне подхода Data Mining («добыча данных», «интеллектуальный анализ данных»). Термин Data Mining введен эмигрантом из СССР Г. Пятецким-Шапиро в 1989 г. Задачи, решаемые Data Mining, — классификация, кластеризация, регрессия, ассоциация (поиск повторяющихся паттернов, например, поиск устойчивых связей) — это типичные задачи прикладной статистики. Новизна состоит в разработке технологий добычи данных путем решения не одной, а ряда таких задач.

Итак, статистические пакеты — интеллектуальные инструменты, необходимые широким кругам научных работников, инженеров, менеджеров. Однако распространенные в настоящее время статистические программные продукты отстают от современного уровня научных исследований примерно на 30 лет. Весьма актуальна задача разработки статистических пакетов нового поколения, соответствующих современному научному уровню и одновременно обеспечивающих удобства пользователей, достигнутые в популярных ныне пакетах. Эта задача должна решаться одновременно с созданием систем обучения, сопровождения и внедрения современных пакетов, в частности, в соответствии с технологиями типа «Шесть сигм» [32].

Подчеркнем, что прикладная статистика является общенаучной дисциплиной. В США, например, статистиков больше, чем математиков [31]. У прикладной статистики должен быть такой же статус и такая же инфраструктура, как у математики, в частности, сис-

тема научно-исследовательских институтов, статистические факультеты в ведущих вузах, отделение в составе РАН и т.д.

Должны быть организованы подготовка студентов всех специальностей по современным методам прикладной статистики и смежным дисциплинам, а также переподготовка специалистов. Важно от старой парадигмы 1950-х годов, распространенной при обучении теории вероятностей и математической статистики, перейти к новой парадигме, выраженной, например, в учебниках [1, 17, 19, 27, 29, 30]. Новая парадигма основана, в частности, на непараметрической и нечисловой статистике, в то время как старая — на нереалистических предположениях о возможности применения семейств параметрических распределений данных (результатов измерений, наблюдений, испытаний, анализов, опытов) в инженерных, управленческих, экономических, медицинских и других практических задачах, встающих перед выпускниками вузов, научными и практическими работниками. Для модернизации народного хозяйства и эффективного внедрения инноваций необходима смена старой парадигмы анализа данных на новую, современную.

ЛИТЕРАТУРА

1. Орлов А. И. Прикладная статистика. — М.: Экзамен, 2006. — 671 с.
2. Крамер Г. Математические методы статистики. — М.: Мир, 1975. — 648 с.
3. Смирнов Н. В., Дунин-Барковский И. В. Курс теории вероятностей и математической статистики для технических приложений. Изд. 3-е, стереотип. — М.: Наука, 1969. — 512 с.
4. Большев Л. Н., Смирнов Н. В. Таблицы математической статистики. 3-е изд. — М.: Наука, 1983. — 416 с.
5. Каган А. М., Линник Ю. В., Рао С. Р. Характеризационные задачи математической статистики. — М.: Наука, 1972. — 656 с.
6. Современные проблемы кибернетики (прикладная статистика). — М.: Знание, 1981. — 64 с.
7. Орлов А. И. О перестройке статистической науки и ее применений / Вестник статистики. 1990. № 1. С. 65 – 71.
8. Орлов А. И. Создана единая статистическая ассоциация / Вестник АН СССР. 1991. № 7. С. 152 – 153.
9. Шмерлинг Д. С. Журналы по теории вероятностей, математической статистике и их применениям / Вероятность и математическая статистика: Энциклопедия / Гл. ред. Ю. В. Прохоров. — М.: Большая Российская энциклопедия, 1999. С. 893 – 910.
10. The teaching of statistics / Studies in mathematics education. Vol. 7. — Paris: UNESCO, 1989. — 258 p.
11. Орлов А. И. Высокие статистические технологии / Заводская лаборатория. Диагностика материалов. 2003. Т. 69. № 11. С. 55 – 60.
12. Орлов А. И. Тридцать лет статистики объектов нечисловой природы (обзор) / Заводская лаборатория. Диагностика материалов. 2009. Т. 75. № 5. С. 55 – 64.
13. Горский В. Г., Орлов А. И. Математические методы исследования: итоги и перспективы / Заводская лаборатория. Диагностика материалов. 2002. Т. 68. № 1. С. 108 – 112.
14. Орлов А. И. Устойчивые математические методы и модели / Заводская лаборатория. Диагностика материалов. 2010. Т. 76. № 3. С. 59 – 67.

15. Новиков А. М., Новиков Д. А. Методология. — М.: СИНТЕГ, 2007. — 668 с.
16. Орлов А. И. О развитии методологии статистических методов / Статистические методы оценивания и проверки гипотез: Межвуз. сб. науч. тр. — Пермь: Изд-во Пермского государственного университета, 2001. С. 118 – 131.
17. 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.
18. Орлов А. И. О развитии экспертных технологий в нашей стране / Заводская лаборатория. Диагностика материалов. 2010. Т. 76. № 11. С. 64 – 70.
19. 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.
20. Орлов А. И. О развитии математических методов теории классификации / Заводская лаборатория. Диагностика материалов. 2009. Т. 75. № 7. С. 51 – 63.
21. Орлов А. И. Сертификация и статистические методы (обобщающая статья) / Заводская лаборатория. Диагностика материалов. 1997. Т. 63. № 3. С. 55 – 62.
22. Орлов А. И. Внедрение современных статистических методов с помощью персональных компьютеров. — В сб.: Качество и надежность изделий. — М.: Знание, 1992, № 5(21). С. 51 – 78.
23. Орлов А. И. Математическое обеспечение сертификации: сравнительный анализ диалоговых систем по статистическому контролю / Заводская лаборатория. Диагностика материалов. 1996. Т. 62. № 7. С. 46 – 49.
24. Орлов А. И. О проверке однородности двух независимых выборок / Заводская лаборатория. Диагностика материалов. 2003. Т. 69. № 1. С. 55 – 60.
25. Орлов А. И. Распространенная ошибка при использовании критериев Колмогорова и омега-квадрат / Заводская лаборатория. Диагностика материалов. 1985. Т. 51. № 1. С. 60 – 62.
26. Орлов А. И. Какие гипотезы можно проверять с помощью двухвыборочного критерия Вилкоксона? / Заводская лаборатория. 1999. Т. 65. № 1. С. 51 – 55.
27. Орлов А. И. Эконометрика. Изд. 4-е. — Ростов-на-Дону: Феникс, 2009. — 572 с.
28. Орлов А. И. О современных проблемах внедрения прикладной статистики и других статистических методов / Заводская лаборатория. 1992. Т. 58. № 1. С. 67 – 74.
29. Орлов А. И. Теория принятия решений. — М.: Экзамен, 2006. — 576 с.
30. Орлов А. И. Организационно-экономическое моделирование: теория принятия решений. — М.: КноРус, 2011. — 568 с.
31. Налимов В. В. О преподавании математики экспериментаторам // О преподавании математической статистики экспериментаторам / Препринт № 17. — М.: Межфакультетская лаборатория статистических методов, 1971. С. 5 – 39.
32. Орлов А. И. «Шесть сигм» — новая система внедрения математических методов исследования / Заводская лаборатория. Диагностика материалов. 2006. Т. 72. № 5. С. 50 – 53.