

А.И. Орлов

Вероятность и прикладная статистика: основные факты

Справочник



МОСКВА
2010

УДК 519.2(075.8)
ББК 22.172я73
О-66

Орлов А.И.

О-66 Вероятность и прикладная статистика: основные факты : справочник / А.И. Орлов. — М. : КНОРУС, 2010. — 192 с.

ISBN 978-5-406-00173-8

Сжато, но строго рассмотрены вероятностно-статистические основы современных статистических методов и основные понятия. Изложение доведено до переднего края научных исследований и практических разработок. Рассмотрены все основные понятия, используемые при применении современных статистических методов. Особое внимание уделено непараметрическим подходам, статистике нечисловых данных и другим перспективным элементам высоких статистических технологий. Справочник составлен в соответствии с рекомендациями Российской ассоциации статистических методов.

Для студентов и преподавателей вузов, слушателей структур послевузовской подготовки и повышения квалификации, инженеров, менеджеров, экономистов, специалистов различных отраслей народного хозяйства, научных работников, аспирантов, для всех, кому необходимо в сжатые сроки овладеть понятийной базой статистических методов, в том числе при освоении методик анализа данных или подготовке к экзаменам.

УДК 519.2(075.8)
ББК 22.172я73

Орлов Александр Иванович

**ВЕРОЯТНОСТЬ И ПРИКЛАДНАЯ СТАТИСТИКА:
ОСНОВНЫЕ ФАКТЫ**

Санитарно-эпидемиологическое заключение
№ 77.99.60.953.Д.003365.04.09 от 01.04.2009 г.

Изд. № 1719. Подписано в печать 22.04.09. Формат 60×90/16.

Печать офсетная. Усл. печ. л. 12,0. Уч.-изд. л. 6,16.

Тираж 1000 экз. Заказ №

ООО «Издательство КноРус».

129110, Москва, ул. Большая Переяславская, 46, стр. 7.

Тел.: (495) 680-7254, 680-0671, 680-1278.

E-mail: office@knorus.ru <http://www.knorus.ru>

Отпечатано в ГУП «Брянское областное полиграфическое объединение».
241019, г. Брянск, пр-т Ст. Димитрова, 40.

ISBN 978-5-406-00173-8

© Орлов А.И., 2010
© ЗАО «МЦФЭР», 2010
© ООО «Издательство КноРус», 2010

ОГЛАВЛЕНИЕ

Предисловие	7
Глава 1. Вероятность и статистика нужны всем	14
Примеры применения теории вероятностей и математической статистики	15
Задачи оценивания	19
Современное представление о математической статистике	20
Коротко об истории математической статистики	22
Вероятностно-статистические методы и оптимизация	24
Глава 2. Основы теории вероятностей	26
События и множества	26
Вероятность события	29
Независимые события	31
Независимые испытания	34
Условные вероятности	35
Формула полной вероятности	36
Формулы Байеса	37
Случайные величины	38
Математическое ожидание	39
Независимость случайных величин	43
Дисперсия случайной величины	47
Биномиальное распределение	50
Неравенства Чебышёва	52
Закон больших чисел	55
Сходимость частот к вероятностям	57
О проверке статистических гипотез	58

Глава 3. Суть вероятностно-статистических методов	65
Глава 4. Случайные величины и их распределения	69
Распределения случайных величин и функции распределения.....	69
Характеристики случайных величин	72
Квантили.....	72
Характеристики положения	75
Характеристики разброса	78
Преобразования случайных величин	80
Моменты случайных величин	82
Стандартное нормальное распределение и центральная предельная теорема	83
Семейство нормальных распределений	87
Распределения Пирсона (хи-квадрат), Стьюдента и Фишера.....	88
Центральная предельная теорема (общий случай)	90
Непрерывные распределения, используемые в вероятностно-статистических методах	92
Логарифмически нормальные распределения	92
Экспоненциальные распределения	93
Распределения Вейбулла—Гнеденко.....	95
Гамма-распределения	98
Дискретные распределения, используемые в вероятностно-статистических методах	101
Подробнее о биномиальном распределении	101
Гипергеометрическое распределение	103
Распределение Пуассона	105
Глава 5. Основные проблемы прикладной статистики — описание данных, оценивание и проверка гипотез	108
Основные понятия, используемые при описании данных.....	108
Виды выборок.....	109
Частоты.....	110
Эмпирическая функция распределения	111

Выборочные характеристики распределения	113
Основные понятия, используемые при оценивании.....	117
Точечное оценивание	119
Состоятельность, несмещенность и эффективность оценок	120
Наилучшие асимптотически нормальные оценки	124
Доверительное оценивание	125
Доверительное оценивание для дискретных распределений	131
Основные понятия, используемые при проверке гипотез.....	133
Параметрические и непараметрические гипотезы	139
Статистические критерии	140
Уровень значимости и мощность	141
Состоятельность и несмещенность критериев	144
Глава 6. Некоторые типовые задачи прикладной статистики и методы их решения	145
Статистические данные и прикладная статистика	145
Статистический анализ точности и стабильности технологических процессов и качества продукции.....	147
Задачи одномерной статистики (статистики случайных величин)	148
Непараметрическое оценивание математического ожидания.....	151
Непараметрическое оценивание функции распределения	152
Проблема исключения промахов.....	154
Многомерный статистический анализ.....	156
Корреляция и регрессия	157
Дисперсионный анализ	159
Методы классификации	161
Снижение размерности	163
Статистика случайных процессов и временных рядов.....	164
Статистика объектов нечисловой природы	165
Цитированная литература	168

Контрольные вопросы и задачи	170
Темы докладов, рефератов, исследовательских работ.....	172
<i>Приложение. Некоторые постановки задач прикладной статистики</i>	173
Предметный указатель	180
Об авторе	187

ПРЕДИСЛОВИЕ

Прикладная статистика — наука о том, как обрабатывать данные. Статистические методы, основанные на вероятностных моделях, активно применяются в технических исследованиях, экономике, теории и практике управления (менеджмента), в социологии, медицине, геологии, истории и т.д. С обработкой результатов наблюдений, измерений, испытаний, опытов, анализов имеют дело специалисты во всех отраслях практической деятельности, почти во всех областях научных исследований.

Наукоемкие технологии основаны на применении высоких статистических технологий организации и управления производством и активно используются в высокотехнологичных отраслях промышленности. Без вероятностно-статистических методов немыслимы оценка и анализ риска, страхование, финансовая деятельность. Инженеры, менеджеры, экономисты, социологи, врачи, психологи, историки успешно применяют интеллектуальные инструменты принятия решений, основанные на вероятности и статистике.

Статистические методы и модели и их база — теория вероятностей — активно развиваются во всем мире. Американская статистическая ассоциация насчитывает более двадцати тысяч членов; Королевское статистическое общество (Великобритания) — более десяти тысяч. Статьи по вероятности и статистике публикуются более чем в пятистах научных журналах. В университетах США статистических факультетов больше, чем математических и физических. Восемь Нобелевских премий по экономике получены эконометриками — специалистами по статистическим методам в экономике.

Современная теория вероятностей основана на аксиоматике академика АН СССР А.Н. Колмогорова. Однако в нашей стране

специалисты и научные работники, студенты и преподаватели еще недостаточно знакомы с последними достижениями в области вероятностно-статистических методов, хотя ссылки на них постоянно встречаются в научно-технической, деловой и учебной литературе.

Цель этой книги — на современном научном уровне рассказать об основных вероятностно-статистических понятиях и фактах. Те, кто еще не знаком с этой ведущей областью современной науки, смогут быстро добраться до переднего фронта исследований. Уже изучавшие вводные курсы теории вероятностей и математической статистики быстро восстановят свои знания и расширят их до уровня, позволяющего квалифицированно использовать статистические методы в своей научной и практической работе, успешно применять профессиональные статистические программные продукты, нормативно-техническую и инструктивно-методическую документацию.

Кому нужна эта книга?

Специалисту. В своей профессиональной деятельности инженеру, менеджеру, экономисту, научному работнику, практически любому специалисту приходится сталкиваться с необходимостью осознанно и квалифицированно применять методы, основанные на теории вероятностей и статистике. Но почти у всех при столкновении с такими методами возникают проблемы. Вероятностно-статистические термины и подходы плохо понятны. Но освоить их надо.

В вузовском курсе высшей математики разбирались основы теории вероятностей и математической статистики. Казалось бы, надо взять учебники и изучить заново. Но в этих книгах нет многих понятий и концепций для практического использования вероятностно-статистических методов, поскольку вузовский курс — только введение в предмет.

Необходима книга — справочное пособие, позволяющая быстро выйти на современный уровень развития статистических методов, краткая, но содержащая разбор всех необходимых понятий. Она перед Вами.

Студенту. В специальных дисциплинах часто используются вероятностно-статистические методы и модели. Значит, надо уметь в них разобраться. То, что сдано годы назад, забыто, да и недостаточно для решения новых задач.

Не стоит искать старые конспекты и заново читать толстые учебники. Сейчас надо быстро освежить свои знания или заново познакомиться с основными фактами теории вероятностей и статистики. Эта книга для Вас!

Профессионалу. Вы постоянно обрабатываете данные с помощью статистических методов. Но вероятностно-статистические методы и модели — очень быстро развивающаяся область. Отслеживаете ли Вы эти изменения? Вы знаете, что критерий Стьюдента остался в прошлом, применять его нецелесообразно? Вам известно, какие методы надо использовать вместо критерия Стьюдента? Вы хорошо знакомы со статистикой нечисловых данных? Если Ваш ответ «да», эта книга для Вас слишком элементарна. Если «нет» — познакомьтесь с современным взглядом на теорию вероятностей и статистику!

Сравнение с аналогами

Как познакомиться с терминологией незнакомой области? Естественная мысль — обратиться к энциклопедии, например, к наиболее солидной под названием «Вероятность и математическая статистика» (см. ссылку [2] в списке цитированной литературы в конце книги). Однако размеры энциклопедии впечатляют, а большинство статей в ней доступны лишь математикам-профессионалам.

Делались попытки составлять более или менее полные сводки терминов, определений и обозначений. Например, в учебник [12] по статистическим методам в экономике (т.е. по эконометрике) нами включена такая сводка в качестве приложения. Однако получить целостное представление о необходимой для освоения учебника базовой области знания таким способом невозможно.

Конечно, аналогами являются учебники и учебные пособия по теории вероятностей и математической статистике (как части типового курса высшей математики) и по общей теории статистики (как части экономического образования). Однако эти издания содержат много информации, в дальнейшем не используемой в практической работе (хотя и полезной при первоначальном изучении предмета). В них отсутствуют необходимые сведения о современных статистических методах. Например, типовые учебники и учебные пособия по теории вероятностей и математической статистике не содержат информации о методах, которым посвящена существенная часть распространенных программных продуктов по статистическим методам, например SPSS или Statistica.

Замысел книги

Первоначальный вариант этой книги подготовлен с целью преодолеть разрыв между типовыми курсами по теории вероятностей и математической статистике и государственными стандартами по статистическим методам управления качеством промышленной продукции. Эти стандарты содержали широко распространенные методы, тем не менее не существовало (и не существует) учебно-методической литературы, заполняющей разрыв между вводными курсами и практически используемыми в технических исследованиях статистическими методами.

Похожие проблемы имеются и в других направлениях, в которых работал автор — в социально-экономической области (в экономике, менеджменте, социологии), в научных медицинских исследованиях.

Необходимость создания нового типа книг для информационной поддержки современных разработок с использованием статистических методов очевидна. Такие книги должны на современном научном уровне давать введение в используемые в настоящее время статистические методы.

Структура книги

Подробное оглавление по существу — сводка основных понятий в области теории вероятностей и статистики. По ходу изложения постоянно отмечаются возможности применения рассматриваемых концепций при решении практических задач. Конкретные методы обработки данных почти не разбираются, но дается вся необходимая база для восприятия описаний таких методов — это и есть основная задача книги.

Содержание книги исчерпывающе представлено в оглавлении. В соответствии с направленностью книги доказательства теорем не приводятся. Исключение — гл. 2, посвященная опытам с конечным числом исходов. В этом случае доказательства проводятся элементарно. Автор неоднократно проводил занятия для школьников и студентов по материалам этой главы.

Замечание для математиков-профессионалов. В изложении удалось обойти ряд математических сложностей. Математические основы теории вероятностей предполагают использование σ -алгебр событий (измеримых множеств) и интеграла Лебега, но прикладникам эти понятия вряд ли нужны, поэтому этим вопросам внимания не уделяется. Также не акцентируется внимание на условиях справедливости Центральной Предельной Теоремы и т.д.

Нумерация формул, теорем, примеров, рисунков, таблиц — своя в каждой главе. Список литературы содержит только процитированные источники (всего же по теории вероятностей и статистике напечатано больше миллиона статей и книг). Для облегчения труда преподавателей и обучающихся приведены контрольные вопросы и задачи, примерные темы докладов, рефератов и исследовательских работ.

Приложение — краткий перечень основных типов постановок задач прикладной статистики, широко используемых в практической деятельности и в научных исследованиях. Обширность этого перечня показывает, что конкретным статистическим методам должны быть посвящены отдельные издания достаточно большого объема.

Включенные в книгу материалы прошли многолетнюю и всестороннюю проверку, использовались во многих отечественных и зарубежных образовательных структурах, а также организациях, занимающихся научной и практической деятельностью. Автор благодарен своим многочисленным коллегам, слушателям и студентам прежде всего различных образовательных структур Московского государственного технического университета им. Н.Э. Баумана, за полезные обсуждения.

Автор благодарен научному редактору В.И. Осипову, всем сотрудникам издательства за поддержку нашего научного направления и большую работу по подготовке рукописи к изданию.

С текущей научной информацией по статистическим методам можно познакомиться на сайте «Высокие статистические технологии» <http://orlovs.pp.ru> (и его версиях www.antorlov.nm.ru, www.antorlov.chat.ru, www.newtech.ru/~orlov, www.antorlov.euro.ru), его форуме <http://forum.orlovs.pp.ru/>, а также на странице Лаборатории экономико-математических методов в контроллин-

ге <http://www.ibm.bmstu.ru/nil/lab.html> (на сайте научно-учебного комплекса «Инженерный бизнес и менеджмент» Московского государственного технического университета им. Н.Э. Баумана). Достаточно большой объем информации содержит еженедельник «Эконометрика», выпускаемый с июля 2000 г. (<http://subscribe.ru/catalog/science.humanity.econometrika>). Автор искренне благодарен разработчику сайтов и редактору электронного еженедельника А.А. Орлову за многолетний энтузиазм.

В книге раскрыто представление о случае, вероятности и статистике, соответствующее общепринятому в мире. Сделана попытка довести рассказ до современного уровня научных исследований в этой области. Конечно, возможны различные точки зрения по тем или иным частным вопросам. Автор будет благодарен читателям, если они сообщат свои вопросы и замечания по адресу издательства или непосредственно автору на форуме сайта <http://orlovs.pp.ru> или по электронной почте: prof-orlov@mail.ru.