ПАРАДОКСЫ

В ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

Gábor J.Székely

PARADOXES IN PROBABILITY THEORY AND MATHEMATICAL STATISTICS

Akadémiai Kiadó Budapest 1986

Г. СЕКЕЙ

ПАРАДОКСЫ

В ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

Перевод с английского В. В. Ульянова
под редакцией В. В. Сазонова

ИЗДАТЕЛЬСТВО МИР

МОСКВА «МИР» 1990

ББК 22.17
С 28
УДК 519.21

Секей Г.

С 28 Парадоксы в теории вероятностей и математической статистике: Пер. с англ. — М.: Мир, 1990. — 240 с., ил.

ISBN 5-03-001293-1

Книга венгерского математика, содержащая собрание неожиданных выводов и утверждений из теории вероятностей, математической статистики и теории случайных процессов. Она написана живо и увлекательно, представленный в ней материал можно использовать для иллюстрации в вузовских лекциях по теории вероятностей, а некоторые разделы — в работе школьных математических кружков.

Для математиков разной квалификации, для всех изучающих теорию вероятностей и математическую статистику.

С $\frac{1602090000-055}{041(01)-90}$ 18—90 **ББК 22.17**

Редакция литературы по математическим наукам

Учебное издание

Габор Секей

ПАРАДОКСЫ В ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

Заведующий редакцией
чл.-корр. АН СССР В. И. Арнольд
Зам. зав. редакцией А. С. Попов
Научн. редактор М. В. Хатунцева
Мл. научн. редактор Л. А. Королева
Художник А. Я. Коршунов
Художественный редактор В. И. Шаповалов
Технический редактор Л. П. Бирюкова
Корректор Л. А. Королева

ИБ № 7095

Сдано в набор 14.03.88. Подписано в печать 31.10.89. Формат 60×90$^1/_{16}$. Бумага кн.-журн. имп. Литературная гарнитура. Высокая печать. Объем 7,5 бум. л. Усл. печ. л. 15. Усл. кр.-отт. 15. Уч.-изд. л. 14,92. Тираж 24000 экз. Зак. 2277. Цена 1 р. 40 к. Изд. № 1/6474.

Издательство «Мир»
В/О «Совэкспорткнига» Государственного комитета СССР по печати. 129820, ГСП, Москва, 1-й Рижский пер., 2.

Набрано в Ленинградской типографии № 2 головного предприятия ордена Трудового Красного Знамени Ленинградского объединения «Техническая книга» им. Евгении Соколовой Государственного комитета СССР по печати. 198052, г. Ленинград, Л-52, Измайловский проспект, 29.
Отпечатано в Ленинградской типографии № 4 ордена Трудового Красного Знамени Ленинградского объединения «Техническая книга» им. Евгении Соколовой Государственного комитета СССР по печати. 191126. Ленинград, Социалистическая ул., 14.

ISBN 5-03-001293-1 (русск.)
ISBN 963-05-4151 (англ.)

ПРЕДИСЛОВИЕ РЕДАКТОРА ПЕРЕВОДА И ПЕРЕВОДЧИКА

> Любопытный отыскивает редкости только затем, чтобы им удивляться; любознательный же — затем, чтобы узнать их и перестать удивляться.
>
> *Р. Декарт*

Настоящая книга представляет собой замечательное собрание неожиданных, противоречащих «здравому смыслу» выводов и утверждений, которых немало в теории вероятностей, математической статистике, теории случайных процессов, и анализ которых ведет к более глубокому пониманию этих предметов. Автор показывает, как разрешение различных парадоксов, связанных со случайностью, способствовало возникновению и развитию теории вероятностей и ее приложений. В книге рассматриваются как парадоксы, ставшие уже классическими, так и неожиданные результаты последних лет, разбросанные по журнальным статьям. Детально разбирается около 50 парадоксов. Для каждого приводится история его возникновения и рассказывается о первых, не всегда удачных попытках его разрешения, затем формулируется сам парадокс и дается его объяснение с точки зрения современных представлений, рассматриваются различные обобщения парадокса, важнейшие понятия и факты из теории вероятностей и ее применений, на установление которых повлияло разрешение данного парадокса. Кроме детально разобранных в книге приводятся также более 50 парадоксов, подробный анализ которых предлагается провести читателям.

Глава I «Классические парадоксы теории вероятностей» содержит разбор таких широко известных задач, как парадокс де Мере, задача о справедливом разделе ставки, парадокс Бертрана, петербургская игра и т. п. Для читателей, которые только начинают изучение теории вероятностей, эта глава интересна тем, что в ней в увлекательной доступной форме излагаются фундаментальные понятия теории вероятностей: классическая и геометрическая вероятности, условные вероятности, функция распределения, плотность, моментные характеристики, пуассоновское и нормальное распределения, закон больших чисел и центральная предельная теорема. Основная часть главы проста для понимания и может быть использована в работе школьных математических кружков. Для тех, кто уже знаком с теорией вероятностей, будет интересна история возникновения и решения классических парадоксов. Любопытны также задачи, подробный анализ которых оставлен читателям, например парадокс Бореля о случайной точке на поверхности шара, задача о страховании имущества и др. Заканчивается глава двумя софизмами Льюиса Кэрролла, связанными с понятием вероятности.

В главе II «Парадоксы в математической статистике» рассматриваются вопросы, касающиеся оценок по методам наименьших квадратов и максимального правдоподобия, регрессии, достаточных статистик, доверительных интервалов, проверки гипотез и т. д. Иными словами, она содержит богатый материал, который можно использовать в курсах лекций по математической статистике, иллюстрируя границы применимости статистических методов и обращая внимание на то, что статистические выводы справедливы лишь при выполнении определенных вероятностных условий. Некоторые утверждения просты, например: не все случайные величины с нулевым коэффициентом корреляции независимы. Однако большая часть парадоксов менее очевидна, например парадокс о существовании оценок более эффективных, чем оценки максимального правдоподобия.

Главы III и IV посвящены парадоксам, обнаруженным сравнительно недавно, например парадоксам случайных процессов, моделирующих броуновское движение, избирательные кампании, деятельность фондовой биржи и т. п. В главе IV шире, чем в других главах, представлены связи теории вероятностей с другими науками. Здесь и теория чисел, и квантовая физика, и моделирование псевдослучайных чисел на ЭВМ, и нестандартный анализ, и криптография. Обсуждаются и чисто вероятностные неожиданные факты, такие, как существование небезгранично делимых компонент у безгранично делимых распределений.

В последней главе на нескольких примерах иллюстрируется мысль автора о том, что математика, как и большинство разделов науки,— это история парадоксов.

Книга написана живо и увлекательно. Ее материал можно использовать в вузовских курсах лекций по вероятности и статистике и при самостоятельном изучении этих дисциплин для более глубокого и активного овладения предметом.

При работе над переводом с автором поддерживался постоянный контакт, что способствовало быстрому решению возникавших вопросов. Русский перевод несколько отличается от английского оригинала, в него внесены изменения и уточнения, предложенные автором для русского издания.

Москва, февраль 1989 г.

В. Сазонов
В. Ульянов

ПРЕДИСЛОВИЕ К РУССКОМУ ИЗДАНИЮ

Я очень рад переводу моей книги на русский язык, это большая честь для меня, так как на мои исследования по теории вероятностей оказали влияние работы многих советских математиков. Отмечу, что совместная с И. Ружей монография по алгебраической теории вероятностей (Wiley, 1988) была вдохновлена в основном трудами А. Я. Хинчина, Ю. В. Линника и И. В. Островского. Я никогда не видел Хинчина и Островского, но встречался с Ю. В. Линником, а также с А. Н. Колмогоровым и Б. В. Гнеденко. Все они стимулировали мои дальнейшие исследования и обещали содействие в публикации книги по парадоксам теории вероятностей. К сожалению, ни Линнику, ни Колмогорову не пришлось прочитать эту книгу, но в 1987 г. в Москве я смог подарить экземпляр книги на английском языке профессору Гнеденко, который был со мной очень любезен. Здесь я еще раз хочу поблагодарить его за гостеприимство.

Надеюсь, что моя книга понравится советским читателям, и они прочтут ее с интересом.

Будапешт, май 1988 г. *Габор Секей*

ПРЕДИСЛОВИЕ РЕДАКТОРА СЕРИИ

> Приступай к решению своих проблем правильно и начинай с ответов. Тогда, возможно, однажды ты найдешь главный вопрос.
>
> *«Отшельник в журавлиных перьях» в книге Р. ван Гулика «Убийства в китайском лабиринте»*

> Дело не в том, что они не видят решения. Дело в том, что они не видят проблемы.
>
> *Г. К. Честертон. «Острие булавки» из сборника «Скандальное происшествие с отцом Брауном»*

Все возрастающая специализация и растущее многообразие направлений в математике привели к появлению множества монографий и учебников по различным специальным вопросам, число которых постоянно увеличивается. Однако «дерево» математических знаний и связанных с математикой наук растет не только благодаря появлению новых ветвей. Достаточно часто случается так, что направления, считавшиеся совершенно различными, оказываются вдруг связанными.

Математические знания и уровень их применения в различных науках за последние годы в значительной степени изменялись: теория меры (нетривиально) используется в экономической географии и теоретической экономике; алгебраическая геометрия взаимодействует с физикой; лемма Минковского, теория кодирования и структура воды встречаются в теории упаковки и покрытия; теория гомотопий оказывается полезной в математическом программировании, при изучении квантовых полей и дефектов кристаллов; алгебры Ли связаны с фильтрацией; при прогнозировании и в электротехнике применимы пространства Штейна. И в дополнение к этому возникли такие новые дисциплины, как «экспериментальная математика», «вычислительная гидродинамика», «вполне интегрируемые системы», «хаос, синергетика и дальний порядок», которые почти невозможно подогнать под уже существующие схемы классификации. Эти дисциплины опираются на различные разделы математики. Наша программа «Математика и ее применения» посвящена новым направлениям и таким (новым) взаимосвязям, как:

— центральное понятие, играющее важную роль в нескольких различных областях математики и/или других науках;

— новые применения результатов и идей из одной области научных исследований в другой области;

— влияние, которое оказывают результаты, проблемы и понятия одной области исследований на развитие другой области.

В рамках программы «Математика и ее применения» делается попытка издания книг, тщательно отобранных в соответствии с изложенными выше принципами. С помощью таких книг, которые скорее заинтересовывают читателей, чем дают полную информацию о предмете, и являются более интригующими, чем энциклопедическими, мы надеемся способствовать лучшему взаимопониманию специалистов в разных областях науки.

Поскольку объем научных исследований, проводимых в Советском Союзе, Восточной Европе и Японии, очень велик, было решено обратить особое внимание на работы, выполненные в этих регионах. Таким образом, было решено начать издание трех региональных серий в рамках главной программы «Математика и ее применения».

Прогресс в математике, как и в других науках, во многом определяется такими вопросами и/или результатами, которые заставляют, так сказать, встряхнуть наши прежние представления и увидеть, что новые решения и естественны, и прекрасны. Парадоксы, т. е. противоречащие интуиции, но верные результаты, возможно, являются лучшим примером проблем такого вида. Теория вероятностей — наука о случайных событиях — всегда была особенно богата парадоксами, много их и сейчас.

Изучение и понимание какой-либо области науки с помощью парадоксов есть, видимо, один из лучших способов развить в себе настоящую интуицию. Для теории вероятностей подобная книга была бы идеальной.

Ничем не объяснимая эффективность математики в науках...

Эуген Вигнер

Что ж, если вам известна лучшая роль, займитесь ею.

Брюс Бейрнсфатер

То, что сейчас доказано, когда-то существовало лишь в воображении.

Уильям Блейк

До тех пор, пока алгебра и геометрия двигались различными путями, их развитие было медленным, а применение ограниченным.

Но когда эти науки объединились, они почерпнули друг у друга свежие силы и в результате быстро двинулись вперед к совершенству.

Жозеф Луи Лагранж

Бассам, март 1986 г.

Михель Хазевинкель

ВВЕДЕНИЕ

> Самое прекрасное и глубокое переживание, выпадающее на долю человека, — это ощущение таинственности. Оно лежит в основе религии и всех наиболее глубоких тенденций в искусстве и науке. Тот, кто не испытал этого ощущения, кажется мне, если не мертвецом, то, во всяком случае, слепым.
>
> *Альберт Эйнштейн. Мое кредо, 1934* [1])

> Замечательно, что науке, начинавшейся с рассмотрения азартных игр, суждено было стать важнейшим объектом человеческого знания...
>
> *Пьер Симон Лаплас. Аналитическая теория вероятностей, 1812*

Как и любая другая область науки, математика отражает противоречия окружающего нас мира. Поэтому история математики, естественно, полна интересных парадоксов, и некоторые из них служили отправной точкой больших изменений. Особенно богата парадоксами математика случайного. По мнению Карла Пирсона, в математике нет другого такого раздела, в котором столь же легко допустить ошибку, как в теории вероятностей. Цель данной книги — показать, каким образом эта стремительно прогрессирующая и широко применяемая область знаний развивалась из парадоксов. В книге сделана попытка остановиться на тех волнующих моментах, которые предшествовали или сразу следовали за решением ряда знаменитых парадоксов, редко упоминаемых в монографиях. В книге рассматриваются не только интересные, но не очень значительные «жемчужины» теории вероятностей, лежащие в стороне от главных направлений ее развития; напротив, центральное место в книге занимают противоречия, разрешение которых в наибольшей степени способствовало преодолению фундаментального кризиса в математике случайного. В книге изложены также проблемы, которые первоначально не рассматривались как парадоксы. В книге о парадоксах, очевидно, должна отражаться история развития теории вероятностей, и поэтому в начале разбираются старейшие парадоксы этой науки.

Важно различать парадоксы и софизмы. Первые — это справедливые, хотя и неожиданные утверждения, в то время как вторые — ложные результаты, полученные с помощью рассуждений, формально кажущихся правильными. И парадоксы, и со-

---

[1]) Эйнштейн А. Собрание научных трудов. В 4-х томах. Мое кредо. Т. 4. — М.: Наука, 1967.

физмы очень интересны и поучительны, но эта книга главным образом посвящена парадоксам (за исключением, например, «парадоксов» в IV/1)[1]). При формулировке парадоксов я стремился к тому, чтобы каждый парадокс был ясен сам по себе. Хотя, очевидно, читателю, не играющему в бридж или незнакомому с нормальными распределениями, будет труднее в тех случаях, когда эти специальные понятия лежат в основе парадоксов. Однако, читая книгу последовательно с самого начала, читатель найдет определения важнейших понятий. (Правила игры в бридж в книге не излагаются, однако все необходимое для понимания соответствующего парадокса в ней можно найти.)

Книга состоит из четырех основных глав. Обсуждение каждого парадокса проходит по схеме, состоящей из пяти частей: история, формулировка, объяснение парадокса, замечания и, наконец, список литературы. В конце каждой главы содержится еще несколько парадоксов. Они разбираются кратко, но не потому, что они менее важны и интересны, а потому, что лежат несколько в стороне от главного направления книги.

Первый, кто вдохновил меня на создание книги о парадоксах теории вероятностей, был мой покойный учитель профессор Альфред Реньи. Эту идею поддержал А. Н. Колмогоров во время нашей встречи в Будапеште в 1972 г. В 1976 г. в течение семестра я работал в университете Амстердама, где профессор А. Балкема привлек мое внимание к нескольким интересным парадоксам. Вдохновляющими были и обсуждения моих лекций в университете Джонса Хопкинса, в Колумбийском и Йельском университетах, в Массачусетском технологическом институте. Мне посчастливилось также встречаться и обсуждать вероятностные проблемы с Джорджем Пойа в Станфордском университете и в Будапеште. Я признателен ему за его советы. Особую благодарность необходимо выразить некоторым моим коллегам из Будапештского университета им. Лоранда Этвеша и Математического института Венгерской академии наук. Их имена часто встречаются в книге.

Наконец, следует отметить, что английское издание книги является переработанным и дополненным вариантом венгерского издания.

---

[1]) То есть парадоксов в п. 1 гл. IV.

ГЛАВА I
КЛАССИЧЕСКИЕ ПАРАДОКСЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

Классика — это то, что все хотели бы прочитать, но никто читать не хочет.

Марк Твен

Опытом каждый называет свои ошибки.

Оскар Уайльд

...истинной логикой для этого мира является исчисление вероятностей, занимающееся нахождением величин вероятностей, которые учитывает или должен учитывать любой здравомыслящий человек.

Дж. Клерк Максвелл

Размышления о случайном (например, золотые правила игроков в азартные игры) были уже в древнейшие времена, но математические вычисления вероятностей и вероятностные парадоксы появляются в письменных источниках начиная лишь с XV века. Хотя сегодня теория вероятностей имеет столько же общего с азартными играми, как и геометрия — с измерениями площадей при земляных работах, тем не менее первые парадоксы возникли из популярных азартных игр.

1. Парадокс игры в кости. «Азартные игры» в мире физических частиц

а) *История парадокса*

Игра в кости была самой популярной азартной игрой до конца средних веков. Само слово «азарт» также относится к игре в кости, так как оно происходит от арабского слова “al-zar“, переводимого как «игральная кость». Карточные игры стали популярны в Европе лишь в XIV веке, в то время как игра в кости пользовалась успехом еще в Древнем Египте во времена 1-й династии и позднее в Греции, а также в Римской империи. (Согласно греческой легенде, игру в кости предложил *Паламедей* для развлечения греческих солдат, скучающих в ожидании битвы при Трое. *Павсаний*, писатель, живший во II веке, упоминает написанную в V веке до нашей эры картину *Полигнота*, на которой изображены Паламедей и Ферсит, играющие в кости.) Самой ранней книгой по теории вероятностей является «Книга об игре в кости» (“De Ludo Aleae“) *Джероламо Кардано* (1501—1576 гг.), которая в основном посвящена игре в кости. Эта небольшая книжка была опубликована лишь в 1663 г., спустя почти 100 лет как была написана. Видимо, поэтому *Галилей* стал заниматься той же самой задачей о костях, хотя она была уже решена в работе Кардано. Галилей также написал трактат на эту тему где-то между 1613 и 1624 гг. Первоначально он назывался «Об открытиях, совершенных при игре в кости» (“Sopra le Scoperte dei Dadi“), но в собрании сочинений Галилея, изданном в 1718 г., название изменили на следующее: «О выходе очков при игре в кости» (“Consideratione sopra il Giuoco dei Dadi“).

б) Парадокс

Правильная игральная кость при бросании с равными шансами падает на любую из граней 1, 2, 3, 4, 5 или 6[1]). В случае бросания двух костей сумма выпавших чисел заключена между 2 и 12. Как 9, так и 10 из чисел 1, 2, ..., 6 можно получить двумя разными способами: $9=3+6=4+5$ и $10=4+6=5+5$. В задаче с тремя костями и 9, и 10 получаются шестью способами. Почему тогда 9 появляется чаще, когда бросают две кости, а 10, когда бросают три?

в) Объяснение парадокса

Задача настолько проста, что кажется странным, что в свое время ее считали страшно трудной. И *Кардано*, и *Галилей* отмечали необходимость учета порядка выпадания чисел. (В противном случае не все исходы были бы равновозможными.) В случае двух костей 9 и 10 могут получаться следующим образом: $9=3+6=6+3=4+5=5+4$ и $10=4+6=6+4=5+5$. Это означает, что в задаче с двумя костями 9 можно «выбросить» четырьмя способами, а 10 — лишь тремя. Следовательно, шансы получить 9 предпочтительней. (Поскольку две кости дают $6\times 6=36$ различных равновозможных пар чисел, шансы получить 9 равны 4/36, а для 10 — лишь 3/36.) В случае трех костей ситуация меняется на противоположную: 9 можно «выбросить» 25 способами, а 10 — уже 26 способами. Так что 10 более вероятно, чем 9.

г) Замечания

(i) Несмотря на простоту задачи о костях, некоторым великим математикам не удавалось ее решить, так как они забывали о необходимости учета порядка выпадений костей. (Эту ошибку

---

[1]) На обычной игральной кости сумма очков на противоположных гранях равна 7, т. е. падение кости на грань 1 обозначает выпадение 6 и т. д. — *Прим. перев.*

часто допускают даже в наши дни.) Ошибались и Лейбниц, один из создателей дифференциального и интегрального исчисления, и Даламбер, один из величайших авторов знаменитой Французской энциклопедии. Однажды Даламберу задали следующий вопрос: с какой вероятностью монета, брошенная дважды, по крайней мере один раз выпадет гербом? Ответ ученого был 2/3, так как он считал, что есть лишь три возможных исхода (герб-герб, герб-решка, решка-решка) и среди них только один неблагоприятный, т. е. когда выпадают две решки. Даламбер пренебрегал тем, что три возможных исхода не равновероятны. Правильным ответом является 3/4, потому что из равновозможных исходов герб-герб, герб-решка, решка-герб и решка-решка только последний является неблагоприятным. Точка зрения Даламбера была даже опубликована в Энциклопедии в 1754 г. в статье «Герб и решетка» (“Croix on pile”).

(ii) Задача о костях некоторым образом связана с направлениями физики XIX и XX веков. Предположим, что вместо игральной кости мы имеем дело с физическими частицами. Каждая грань кости соответствует теперь фазовой ячейке, в которой частица оказывается случайным образом и которая характеризует состояние частицы. В этом случае игра в кости эквивалентна модели *Максвелла — Больцмана* для физических частиц. В этой модели, используемой обычно для молекул газа, каждая частица с равными шансами попадает в любую ячейку, так что при описании множества равновозможных исходов следует учитывать порядок так же, как и в задаче о костях. Существует другая модель, в которой частицы неразличимы, и поэтому при подсчете равновозможных исходов порядок не надо принимать во внимание. Эта модель названа именами *Бозе* и *Эйнштейна.* Используя эту терминологию, можно сказать, что суть нашего парадокса в том, что игра в кости описывается моделью Максвелла — Больцмана, а не Бозе — Эйнштейна. Следует отметить, что ни одна из этих моделей не корректна для связанных электронов, так как в этом случае в каждой ячейке может оказаться не более одной частицы. Используя терминологию игры в кости, можно сказать, что если на одной кости выпала 6, то ни на какой другой кости 6 уже выпасть не может. Это модель *Ферми — Дирака.* Возникает вопрос, применение какой модели корректно в конкретной ситуации. (Кроме указанных трех моделей, существует множество других.) Вообще говоря, мы не можем выбрать какую-либо модель, исходя лишь из логических соображений. В большинстве случаев решить этот вопрос позволяют опыт либо наблюдения. Однако в случае с игральной костью очевидно, что корректна модель Максвелла — Больцмана, а в данный момент это все, что нам нужно.

д) Литература

Классической монографией по истории классической теории вероятностей является книга

Todhunter I. *History of the Mathematical Theory of Probability*, которая первоначально была опубликована в 1865 г. и переиздана в 1949 г. издательством «Publishing House Chelsea».

Описание предыстории и ранних периодов развития теории вероятностей можно найти в книге [1])

David F. N. *Games, Gods and Gambling*. Griffin, London, 1962.

Следующие книги посвящены историческим и философским аспектам ранней теории вероятностей:

Hacking I. *The Emergence of Probability*. Cambridge University Press, 1973.

Майстров Л. Е. Развитие понятия вероятности. М.: Наука, 1980.

Читатели, которых интересует начальная история теории вероятностей, могут найти дополнительную информацию в журналах *Biometrica* и *Archive for the History of Exact Sciences*.

Английский перевод первой монографии о задачах, связанных с игрой в кости, а также подробную биографию автора монографии Джероламо Кардано можно найти в книге

Ore Ø. *Cardano, the Gambling Scholar*. Princeton University Press, Princeton, 1953.

2. Парадокс де Мере

а) История парадокса

Существует старая история, впервые рассказанная, видимо, Лейбницем, о том, как известный французский игрок XVII века шевалье де Мере по дороге в свое имение в Пуату встретил *Блеза Паскаля*, одного из знаменитейших ученых XVII века. Де Мере поставил перед Паскалем две задачи, обе связанные с азартными играми. Первой задачей был рассматриваемый парадокс (вторая — разбирается в следующем параграфе). Обе задачи Паскаль обсуждал в 1654 г. в своей переписке с *Пьером де Ферма*, другим высокоодаренным ученым, жившим в Тулузе. Оба ученых пришли к одинаковому результату, что весьма обрадовало Паскаля. В своем письме он отмечал: «Я вижу, что истина одинакова и в Тулузе, и в Париже». *Эйштейн Оре*, профессор Йельского университета, утверждал, что парадоксы, приписываемые де Мере, на самом деле были широко известны значительно раньше, дело лишь в том, что Паскаль о них не знал. Неверно также, что шевалье был страстным игроком. Парадоксы его интересовали скорее теоретически, чѐм практически, и поэтому его не удовлетворило, что Паскаль «всего лишь» решил

---

[1]) Полезно также прочитать «Очерк истории теории вероятностей» в книге Б. В. Гнеденко «Курс теории вероятностей», М.: Наука, 1988. — *Прим. перев.*

задачу, подтвердив правильный ответ, который де Мере уже знал. Шевалье не смог понять, каким образом разрешается парадокс.

б) Парадокс

При четырех бросаниях одной игральной кости вероятность того, что по крайней мере один раз выпадет 1, больше 1/2. В то же время при 24 бросаниях двух костей вероятность выпадения двух 1 одновременно (по крайней мере однажды) меньше 1/2. Это кажется удивительным, так как шансы получить одну 1 в шесть раз больше, чем шансы выпадения двух 1, а 24 как раз в 6 раз больше 4.

в) Объяснение парадокса

Если правильную игральную кость бросают k раз, то число возможных (и равновероятных) исходов равно 6^k. В 5^k случаях из этих 6^k кость не ляжет на шестерку, и, следовательно, вероятность выпадения по крайней мере один раз 1 при k бросаниях равна

$$(6^k - 5^k)/6^k = 1 - (5/6)^k,$$

что больше 1/2, если $k = 4$. С другой стороны, величина $1 - (35/36)^k$, которая получается аналогично, все еще меньше 1/2 для $k = 24$ и превосходит 1/2, начиная с $k = 25$. Так что «критическое значение» для одной кости равно 4, а для пары костей равно 25. Это безусловно правильное решение на самом деле не удовлетворило де Мере, так как сам ответ он уже знал, но из решения так и не понял, почему ответ не согласуется с «правилом пропорциональности критических значений», утверждающим, что если вероятность уменьшается в шесть раз, то критическое значение возрастает в шесть раз ($4 : 6 = 24 : 36$). *Абрахам де Муавр* (1667—1754 гг.) в своей книге «Доктрина шансов» ("Doctrine of Chances"), опубликованной в 1718 г., доказал, что «правило пропорциональности критических значений» недалеко от истины, так как, если p — вероятность некоторого события (например, вероятность «выбросить» единицу есть $p = 1/6$), то критическое значение k можно найти, решая уравнение

$$(1 - p)^x = 1/2$$

(это уравнение имеет решение, если p заключено строго между 0 и 1). Критическое значение k есть наименьшее целое число, превосходящее x. Решение приведенного выше уравнения дается формулой

$$x = -\ln 2/\ln(1 - p) = \ln 2/(p + p^2/2 + \ldots), \qquad (*)$$

где $\ln$ означает натуральный логарифм (по основанию $e = 2.71...$). Из вида решения ясно, что если p^2 пренебрежимо мало, то p убывает почти пропорционально возрастанию критического значения, как де Мере и предполагал. Де Муавр использовал приближенную формулу $x \approx \ln 2/p \approx 0.69/p$ для исследования вопросов, связанных с Лондонской лотереей. В этом случае значение p было $1/32$, а для $p = 1/32$ точное значение $x = 22.135...$, в то время как приведенная выше формула дает приближение 22.08, что достаточно близко к истинному значению. Парадокс де Мере возникает потому, что для $p = 1/6$ величина $p^2/2$ (и другие слагаемые знаменателя в формуле (*)) не настолько мала, чтобы ею можно было пренебречь. Таким образом, «правило пропорциональности критических значений» является правилом асимптотически верным, ошибка от его применения растет с ростом p. Это и есть настоящее решение данного парадокса.

г) Замечания

(i) Типично неверное решение задачи де Мере восходит к Кардано. Он рассуждал следующим образом: вероятность получения пары единиц равна 1/36, следовательно, для того, чтобы с вероятностью 1/2 получить пару единиц по крайней мере один раз, нужно бросить кость ровно 18 раз. Рассуждая таким образом, мы придем к тому, что если кость бросают более 36 раз, то вероятность выпадения пары единиц больше 1, что, конечно, невозможно.

(ii) Существуют некоторые «случайные величины», которые подчиняются «правилу пропорциональности». (Мы рассмотрим эти случайные величины в парадоксе 8.) Некоторые из этих случайных величин играют важную роль в ядерной физике, в которой критическое значение называется периодом полураспада. Период полураспада обратно пропорционален постоянной распада, которая соответствует p.

(iii) Число бросаний правильной игральной кости до первого выпадения единицы есть величина, зависящая от случая, иными словами *случайная величина*. Обозначим эту случайную величину через v. Возможными значениями v являются $1, 2, 3, 4, \ldots$. Вероятность того, что $v = k$ (где k — произвольное положительное целое число), равна $(5/6)^{k-1}/6$. Так что *среднее* или *математическое ожидание* величины v (определяемое как взвешенное среднее ее значений с весами, равными соответствующим вероятностям) равно

$$1 \cdot 1/6 + 2 \cdot 5/6 \cdot 1/6 + 3\,(5/6)^2/6 + 4\,(5/6)^3/6 + \ldots = 6.$$

В более общем случае, если p — вероятность появления некоторого события A и мы повторяем независимые испытания до тех пор, пока A не произойдет, математическое ожидание числа необходимых испытаний равно $1/p$. Таким образом, эти математические ожидания подчиняются «правилу пропорциональности»: для получения пары единиц в среднем потребуется в шесть раз больше бросаний, чем для получения одной единицы.

(iv) До сих пор мы использовали интуитивное представление о вероятностной независимости. К этому мы вернемся позднее.

(v) Объяснение парадокса де Мере не стало широко известным. В 1693 г., спустя почти сорок лет после того, как проблема

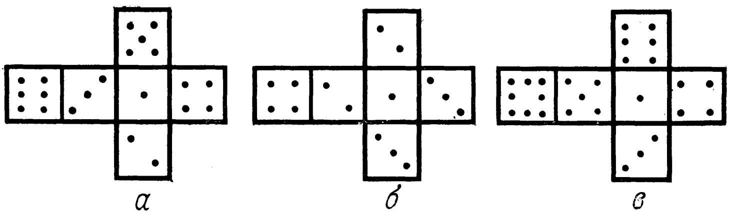

Рис. 1. Грани одной игральной кости, размеченной обычным способом, и грани двух костей, размеченных нестандартно. При бросании двух симметричных костей, размеченных нестандартным образом (*б*, и *в*), вероятности того, что в сумме получится 2, 3, ..., 12, совпадают с соответствующими вероятностями при бросании двух обычных игральных костей.

была решена Паскалем, *Самьюэл Пепис* (президент Королевского общества с 1684 г.) предложил Ньютону почти точно такую же задачу. Ньютон также нашел правильный ответ, который в свою очередь не удовлетворил Пеписа.

(vi) Дрейдель ("Dreydel" или "draydl") — это древняя игра, похожая на игру в кости. (Она также напоминает английскую игру «положи-и-возьми».) В дрейдель играют евреи во время праздника Ханука. Сравнительно недавно Фейнерман обнаружил (см. ссылку ниже), что эта игра несправедлива, если число играющих больше двух, хотя, как это ни парадоксально, никто этого не заметил в течение более 2000 лет!

Дрейдель — это четырехсторонний волчок, стороны которого обозначаются буквами N, G, H и S (соответствующими буквам иврита: нун, гимел, хей и шин). В игре участвует произвольное число игроков, каждый из которых делает ставку, равную некоторой денежной единице, чтобы начать игру. Игроки по очереди запускают волчок до тех пор, пока по взаимному согласию не прекращают игру. Выплаты (игроку, который запускал волчок) соответствуют каждому из четырех возможных исходов: N — игрок ничего не получает, H — получает половину общей ставки, G — забирает всю общую ставку, S — делает еще одну ставку. В том случае, когда при запуске волчка получается G и игрок забирает всю общую ставку, все игроки вновь делают ставки для продолжения игры. Если обозначить число игроков через m, то математическое ожидание выигрыша при n-м запуске волчка равно $E_n = m/4 + (5/8)^{n-1}(m-2)/8$. Таким образом, если $m > 2$, то E_n — строго убывающая последовательность. Следовательно, первый игрок (который запускает волчок при $n = 1,\ m + 1,\ 2m + 1, \ldots$) всякий раз имеет больший ожидаемый выигрыш, чем второй игрок...

д) Литература

Ore Ø. "Pascal and the invention of probability theory", *The American Math. Monthly*, **67**, 409—419, (1960). В этой статье сделан обзор результатов недавних исследований, связанных с ролью де Мере и Паскаля.

Rényi A. *Letters on Probability*, Akadémiai Kiadó, Budapest, 1972. [Имеется перевод в книге Реньи А. Трилогия о математике. — М.: Мир. 1980.] Эта книга знакомит читателя с ранним периодом развития теории вероятностей в форме четырех выдуманных писем (якобы написанных Паскалем Ферма).

Schnell E. D. "Samuel Pepys, Isaac Newton and the Probability", *The American Statistician*, **14**, 27—30, (1960). Это статья о переписке между Ньютоном и Пеписом.

За исключением первого письма Паскаля (которое исчезло), переписка между Паскалем и Ферма опубликована в книге Smith D. E. *A Source Book in Mathematics*, McGraw-Hill, New York, 1929, 546—565 (есть новое издание 1959 г.).

Feinerman R. "An ancient unfair game", *The American Math. Monthly*, **83**, 623—625 (1976).

3. Парадокс раздела ставки

а) История парадокса

Этот парадокс был впервые опубликован в Венеции в 1494 г. в обзоре средневековой математики. Автор *Фра Лука Пачоли* (1445—1509 гг.) назвал свою книгу «Сумма знаний по арифметике, геометрии, отношениям и пропорциональности» ("Summa de arithmetica, geometria, proportioni et proportionalita"). В этой книге впервые использовалось слово «миллион» и объяснялась двойная бухгалтерия. Интересно отметить, что в Милане Фра Лука близко подружился с *Леонардо да Винчи*, и благодаря этой дружбе Леонардо иллюстрировал работу Фра Лука «О божественных пропорциях» ("De Divina Proportione"), опубликованную в Венеции в 1509 г. Недавно Эйштейн Оре обнаружил итальянскую рукопись, датированную 1380 г., в которой также упоминается парадокс раздела ставки. Многое указывает на арабское происхождение задачи или по крайней мере на то, что в Италию задача попала вместе с арабским учением. Как бы ни стара была проблема, фактом остается, что для ее правильного решения потребовалось очень много времени. Сам Пачоли даже не видел связи этой задачи с теорией вероятностей; он рассматривал ее как задачу о пропорциях. Неверное решение дал *Никколо Тарталья* (1499—1557 гг.), хотя он был достаточно гениален, чтобы в математической дуэли за одну ночь открыть формулу корней кубического уравнения. После нескольких неудачных попыток Паскаль и Ферма в конце концов в 1654 г. независимо друг от друга нашли правильный ответ задачи. Это открытие было настолько важным, что многие считают этот год временем рождения теории вероятностей, а все предшествующие результаты относят к ее предыстории.

б) Парадокс

Два игрока играют в безобидную игру (т. е. у обоих шансы победить одинаковы), и они договорились, что тот, кто первым выиграет 6 партий, получит весь приз. Предположим, что на самом деле игра остановилась до того, как один из них выиграл приз (например, первый игрок выиграл 5 партий, а второй — 3). Как справедливо следует разделить приз? Хотя в действительности эта проблема не является парадоксом, безуспешные попытки некоторых величайших ученых решить ее, а также неверные противоречивые ответы создали легенду о парадоксе. Согласно одному ответу, приз следовало разделить пропорционально выигранным партиям, т. е. 5 : 3[1]. Тарталья предложил делить в отношении 2 : 1. (Наиболее вероятно, что он рассуждал следующим образом: так как первый игрок выиграл на две партии больше, что составляет третью часть от необходимых для победы 6 партий, то первый игрок должен получить одну треть от приза, а оставшуюся часть следует разделить пополам.) На самом деле справедливым является раздел в отношении 7 : 1, что сильно отличается от предыдущих результатов.

в) Объяснение парадокса

И Паскаль, и Ферма рассматривали эту проблему как задачу о вероятностях. Так что справедливым будет раздел, пропорциональный шансам первого игрока выиграть приз. Покажем, что в случае, когда первому игроку осталось выиграть только одну партию, а второму игроку для победы требуется выиграть три, справедливое отношение равно 7 : 1. Следуя идее Ферма, продолжим игру тремя фиктивными партиями, даже если некоторые из них окажутся лишними (т. е. когда один из игроков

---

[1]) Как утверждает Б. В. Гнеденко, этот ответ предложен Пачоли. — *Прим. перев.*

уже выиграл приз). Такое продолжение делает все $2 \cdot 2 \cdot 2 = 8$ возможных исходов равновероятными. Поскольку только при одном исходе второй игрок получает приз (т. е. когда он выигрывает все три партии), а в остальных случаях побеждает первый игрок, справедливым является отнощение 7 : 1.

г) Замечания

(i) Общее решение для случая, когда первому игроку для получения приза требуется выиграть еще n партий, а второму — m партий, также было найдено Паскалем и Ферма. Для первого игрока шансы получить приз равны

$$2^{-n-m+1} \sum_{j=n}^{n+m-1} \binom{n+m-1}{j}.$$

(Здесь число фиктивных партий равно $n+m-1$ и все возможные 2^{n+m-1} исходов равновероятны.) В 1654 г. весь Париж говорил о возникновении новой науки — теории вероятностей. Спустя несколько месяцев в Париж из Голландии приехал молодой гений *Христиан Гюйгенс*, чтобы обсудить с Паскалем или Ферма вероятностные проблемы, которыми он тоже интересовался. Однако случилось так, что он не встретился ни с одним из них. (Паскаль был поглощен религией и не принимал гостей, а Ферма жил вдали от Парижа.) Тем не менее о наиболее интересных результатах Гюйгенс узнал. Вскоре он вернулся в Голландию и начал писать книгу по теории вероятностей. Эта замечательная работа, которая, в частности, содержит решение задачи о разделе ставки для трех игроков, была опубликована в 1657 г. под названием «О расчете в азартных играх» (“De Ratiociniis in Aleae Ludo”) в виде части (пятой книги) труда *Схоутена* «Математические этюды» (“Exercitationes Mathematicarum”). В работе Гюйгенса 16 страниц; она начинается с предисловия и содержит решения 14 задач, связанных с азартными играми.

(ii) Прекрасную идею Ферма о продолжении игры в 1977 г. (!) использовал Андерсон (см. ссылку ниже). Он доказал следующую поразительную теорему: игрок, который подает первым, имеет одинаковые шансы выиграть в N партиях раньше своего соперника независимо от того, подают ли игроки поочередно или подает тот, кто выигрывает предыдущую партию [1]).

---

[1]) Эту теорему комбинаторными методами раньше доказал Кингстон (см. Kingston J. G. Comparison of Scoring systems in two-sided competitions, *J. Comb. Theory* (A), **20**, 357—362, (1976)). Андерсону принадлежит простое доказательство теоремы с использованием идеи Ферма. — *Прим. перев.*

д) Литература

Anderson C. L. "Note on the advantage of the first serve", *J. Combinatorical Theory* (A) **23**, 363, (1977).
Jordan K. *Chapters on the Classical Calculus of Probabilities*, Akadémiai Kiadó, Budapest, 1972.

4. Парадокс независимости

а) История парадокса

Сначала определим понятие независимости для двух случайных событий A и B. Обозначим их вероятности через $P(A)$ и $P(B)$, и пусть $P(AB)$ — вероятность того, что наступят и A, и B. (Символ P широко используется для обозначения вероятности события, поскольку не только в английском, но и во многих других языках начальной буквой слова «вероятность» является P — probabilitas в латинском, probabilité во французском, probabilidad в испанском, probabilità в итальянском языке и т. д.) Пусть A — произвольное событие и B — событие, имеющее положительную вероятность. Вероятность события A при условии, что событие B произошло, иными словами, условную вероятность A относительно B, будем обозначать $P(A|B)$ и определять отношением

$$P(A|B) = P(AB)/P(B).$$

Два события A и B называются независимыми, если справедливо равенство

$$P(A|B) = P(A),$$

т. е. если условная вероятность равна безусловной. Если записать предыдущее соотношение в виде

$$P(AB) = P(A) \cdot P(B), \qquad (*)$$

то мы получаем простое равенство, симметричное по A и B, в котором даже не надо предполагать, что вероятность $P(B)$ положительна. Следовательно, предпочтительнее начать со следующего определения: два события A и B независимы, если выполнено равенство $(*)$.

Математическое определение независимости, как правило, согласуется с нашим обычным представлением о независимости. Например, если мы бросаем две кости, то события «шестерка на первой кости» и «шестерка на второй кости», очевидно, независимы как в общепринятом смысле, так и математическом. Это согласование, однако, наблюдается не всегда. С. Н. Бернштейн обратил внимание на следующий парадокс.

б) Парадокс

Предположим, что бросают две правильные монеты. Пусть событие A — «на первой монете выпал герб», событие B — «на второй монете выпал герб» и событие C — «на одной (и только на одной) монете выпал герб». Тогда события A, B и C попарно независимы, но любые два из них однозначно определяют третье.

в) Объяснение парадокса

Во-первых, A и B, очевидно, независимы, поскольку результаты первого бросания не зависят от результатов второго. С другой стороны, события A и C (а также B и C) на первый взгляд кажутся зависимыми, но так как $P(AC) = P(A) \cdot P(C) = 1/4$ и аналогично $P(BC) = P(B) \cdot P(C)$, они на самом деле независимы. Справедливо также, что любые два события определяют третье, поскольку каждое событие (A, B и C) происходит тогда и только тогда, когда происходит ровно одно из двух других событий. Этот парадоксальный феномен показывает, что попарная независимость событий не означает их независимость в совокупности. Если мы хотим выразить последнее, то должны предполагать больше, чем попарную независимость. События из некоторой совокупности называются взаимно независимыми, если для любого конечного набора событий из этой совокупности $A_1, A_2, \ldots, A_n$ выполняется свойство мультипликативности

$$P(A_1 A_2 \ldots A_n) = P(A_1) \cdot P(A_2) \cdot \ldots \cdot P(A_n), \qquad (**)$$

т. е. если вероятность совместного события равна произведению индивидуальных (маргинальных) вероятностей.

г) Замечания

(i) Если события $A_1, A_2, \ldots, A_n$ не независимы, то можно утверждать лишь, что

$$-(1-1/n)^n \leqslant P(A_1 A_2 \ldots A_n) - P(A_1) P(A_2) \cdot \ldots \cdot P(A_n) \leqslant (n-1) n^{-n/(n-1)}.$$

(ii) Несколько простых парадоксов можно разрешить только с помощью понятия независимости. Рассмотрим следующую задачу. Юноша собирается сыграть три теннисных матча со своими родителями, и он должен победить два раза подряд. Порядок матчей может быть следующим: «отец — мать — отец» или «мать — отец — мать». Юноше нужно решить, какой порядок для него предпочтительнее, учитывая, что отец играет лучше матери. С первого взгляда кажется, что второй порядок предпочтительнее для юноши, так как в этом случае он дважды играет со своей матерью. Однако при этом юноша должен победить в единственном матче, который он играет против отца, в противном случае у него не будет двух побед подряд. Может быть, лучше выбрать первый вариант? Если юноша выигрывает у отца с вероятностью p и с вероятностью q — у матери, то $p < q$, так как отец играет лучше матери. Выбрав первый вариант, юноша должен выиграть либо первый и второй матчи — и вероятность этого pq, либо второй и третий матчи — и вероятность такого исхода qp. Таким образом, вероятность того, что произойдет одно из этих двух событий, равна $pq + qp - pqp$ (pqp необходимо вычесть, так как в противном случае дважды учитывается вероятность выигрыша юноши в трех матчах). Аналогично если юноша выбирает второй возможный вариант, то вероятность того, что он победит два раза подряд, равна $qp + pq - qpq$. Поскольку $p < q$, получаем $pq + qp - pqp > > qp + pq - qpq$, откуда следует, что для юноши лучше выбрать вариант «отец — мать — отец»!

(iii) Можно также определить независимость случайных величин. Пусть X_1, X_2, ... — произвольные случайные величины, принимающие действительные значения. Эти величины называются *взаимно независимыми* (или, кратко, независимыми), если для произвольных действительных значений x_1, x_2, ... события

$$A_{x_1} = (X_1 < x_1), \quad A_{x_2} = (X_2 < x_2), \ \dots$$

взаимно независимы. Функция $F(x) = P(X < x)$ называется *функцией распределения* случайной величины X, и функция $F(x, y, \dots, w) = P(X < x, Y < y, \dots, W < w)$ называется *совместной функцией распределения* случайных величин X, Y, ..., W. Теперь мы можем определить совместную независимость для произвольного (конечного или бесконечного) множества случайных величин следующим образом: множество случайных величин называется независимым, если для произвольного конечного подмножества S этого множества совместная функция распределения случайных величин из S равняется произведению их индивидуальных (маргинальных) функций распределения. Если функция распределения $F(x)$ и совместная функция распределения $F(x, y, \dots)$ могут быть записаны в виде

$$F(x) = \int_{-\infty}^{x} f(\bar{x})\, d\bar{x}$$

и

$$F(x, y, \dots, w) = \int_{-\infty}^{x} \int_{-\infty}^{y} \dots \int_{-\infty}^{w} f(\bar{x}, \bar{y}, \dots, \bar{w})\, d\bar{x}\, d\bar{y} \dots d\bar{w},$$

то функции $f(\bar{x})$ и $f(\bar{x}, \bar{y}, \dots, \bar{w})$ называются *плотностями вероятности*. Если плотности вероятности существуют, то независимость означает, что совместная плотность равна произведению индивидуальных плотностей.

(iii) Если плотность вероятности $f(x)$ случайной величины X существует, то математическое ожидание X равно

$$E(X)=\int_{-\infty}^{\infty} xf(x)\,dx.$$

Математическое ожидание случайной величины $(X-E(X))^2$ называется *дисперсией* величины X. Положительное значение квадратного корня из дисперсии называется стандартным отклонением и является мерой разброса случайной величины X вокруг ее среднего значения. (Существуют и другие меры разброса, но наиболее важной безусловно является стандартное отклонение. Впервые термины «стандартное отклонение» и «дисперсия» использовались К. Пирсоном (1895 г.) и Р. Фишером (1920 г.) соответственно.) Если $f(x)$ — плотность случайной величины X, то ее дисперсия

$$D^2(X)=\int_{-\infty}^{\infty}(x-E(X))^2 f(x)\,dx.$$

Если X и Y независимы, то $E(XY)=E(X)\cdot E(Y)$ и $D^2(X+Y)=D^2(X)+D^2(Y)$ (при условии, что у X и Y дисперсии существуют). Равенство $E(X+Y)=E(X)+E(Y)$ имеет место без предположения о независимости X и Y.

д) Литература

Joffe A. "On a set of almost deterministic k-dependent random variables", *Annals of Prob.* **2**, 161—162, (1974).

Wang Y. H. "Dependent random variables with independent subsets", *The American Math. Monthly*, **86**, 290—292, (1979).

Следующая статья посвящена довольно любопытному соотношению между независимостью и кривыми Пеано:

Holbrook J. A. R. "Stochastic independence and space-filling curves", *The American Math. Monthly*, **88**, 426—432, (1981).

5. Парадоксы бриджа и лотереи

а) История парадокса

История азартных игр начинается с древнейших времен. Они получили настолько широкое распространение, что некоторые государства и религии считали своим долгом азартные игры запретить. *Фридрих II*, император Священной Римской империи, в 1232 г. запретил игру в кости. (В то время это была, видимо, единственная популярная азартная игра.) *Луи IX*, король Франции, в 1255 г. издал закон, по которому игровые кости нельзя было даже изготавливать. В еврейском Талмуде игроков считали ворами, преследовала игроков и христианская церковь.

Среди современных азартных игр наиболее распространены, несомненно, карточные игры. Слово «карта» происходит от греческого слова χαρτηζ — бумага, однако карточные игры были известны еще до появления бумаги. Хотя неизвестно, где возникли карточные игры, в Европу они, видимо, попали через Венецию во время крестовых походов в XIII веке, а в Венецию — через Китай—Персию—Сирию—Палестину. Факты состоят в следующем. Согласно китайской энциклопедии XVII века игры, похожие на карточные, были известны в Китае уже в 1120 г. до нашей эры. Остатки арабской карты XIII века можно увидеть во дворце музее Топкапы в Стамбуле. Флорентийский указ 1376 г. запрещал карточную игру «наиби». Согласно рукописи 1377 г., хранящейся в Британском музее, приблизительно в это время карточные игры стали популярны в Швейцарии. В Национальной библиотеке в Париже находятся 17 карт Таро, изготовленных в 1392 г. для *Карла VI.* Иоганн Гутенберг отпечатал карты Таро в том же году, что и свою знаменитую Библию [1]. Современная колода возникла из карт Таро Гутенберга. В колоде Таро было 78 карт; 22 главные карты, известные под названием "atouts" (т. е. «главнее всех»; позднее эти "atouts" назвали козырями). Спустя несколько десятилетий французы убрали 22 карты "atouts" и 4 рыцарей ("Knights"). Оставшиеся 52 карты и образуют современную колоду. С того времени число популярных карточных игр достигло нескольких сотен, но возросло также и число шулеров. Этот факт отражает знаменитая картина *Караваджо* «Карточные шулеры», написанная в 1593 г. Чтобы ограничить влияние шулеров, в 1765 г. лейтенант французской полиции *Габриель де Сартин* ввел рулетку. Она стала самой увлекательной и самой старой игрой в казино и существует до сих пор. С XVII века игры типа лотерей, организуемые государством, становятся все более популярными. Первая публичная лотерея с денежными призами, Lotto de Firenze, состоялась во Флоренции в 1530 г. Другой вариант лотереи появился в 1620 г., когда в Совет Женевы потребовалось выбрать пять новых членов для заполнения вакантных мест. Эти члены выбирались среди 90 граждан, записки с фамилиями которых были положены в урну и затем вытаскивалось пять записок. Жителям Женевы разрешалось ставить на

---

[1]) 42-строчная «Библия» печаталась примерно в 1450—1455 гг. — *Прим. перев.*

пять счастливчиков. Даже сегодня карточные игры, рулетка, лотереи и другие азартные игры очень популярны. Иногда некоторые стратегии, приводящие к выигрышу, объявляются «абсолютно надежными», но в действительности также утверждения оказываются научно необоснованными. С другой стороны, точные научные теории известны лишь крайне небольшому числу математиков. Эти теории обычно подтверждают правила, используемые на практике. Однако математические теоремы могут противоречить здравому смыслу и становиться источником парадоксов. Здесь мы рассмотрим лишь два из них.

в) Парадоксы

(i) Парадокс бриджа

Предположим, что в коалиции двух игроков на руках 26 карт, среди которых 6 козырей. Тогда наиболее вероятное распределение козырей следующее: 4 — на одной руке и 2 — на другой. Вероятность такого распределения в точности равна 78/161, что немного меньше 1/2, а вероятность распределения 3—3 несколько больше 1/3, ее точное значение равно 286/805. Теперь предположим, что дважды ходили с козырей и оба игрока в коалиции дважды положили козырные карты. В этом случае у коалиции осталось лишь 2 козыря, причем либо обе карты на одной руке, либо у каждого игрока по одному козырю. Если между двумя игроками распределяются 2 козыря и 20 других карт, то шансы того, что у одного из игроков окажутся оба козыря, равны 10/21, вероятность второго варианта равна 11/21. Итак, второй вариант более вероятен, т. е. *более вероятное* распределение 1—1 получается из *менее вероятного* распределения 3—3. Нет ли здесь противоречия?

(ii) Парадокс лотереи

Большинство участников лотерей обычно не ставят на «слишком симметричные» комбинации, хотя все комбинации равновозможны. Причина этого очень проста. Игроки по опыту знают, что, как правило, выигрывают несимметричные комбинации. В действительности выгоднее ставить на наиболее симметричные комбинации именно потому, что большинство игроков их избегает.

в) Объяснение парадоксов

(i) Шансы получить распределение козырей 3—3 равны

$$\binom{6}{3}\binom{20}{10}\Big/\binom{26}{13}=\frac{286}{805}.$$

Аналогично для распределений 2—4 и 4—2 имеем

$$\binom{6}{2}\binom{20}{11}\Big/\binom{26}{13}+\binom{6}{4}\binom{20}{9}\Big/\binom{26}{13}=\frac{78}{161},$$

поэтому второе распределение более вероятно. Для двух козырей и 20 других карт распределение козырей 1 — 1 имеет шансы

$$\binom{2}{1}\binom{20}{10}\Big/\binom{22}{11}=\frac{11}{21}.$$

Вероятность противоположного события, очевидно, равна 10/21, что и утверждалось выше. Но тогда где же ошибка? Прежде всего покажем, где ее быть не может. Естественно считать, что после захода с козырей дважды (и после того, как оба игрока положили по 2 козыря) в силу полученной в результате этого информации вероятности изменились. Действительно, условные вероятности (при условии, что оба игрока имеют по меньшей мере по 2 козыря) отличаются от безусловных вероятностей, но получаются из них путем умножения на одно и то же число. Следовательно, их отношение не изменится, так что таким путем парадокс не разрешить. ($286/805+78/105=676/805$, и поэтому условные вероятности в 805/676 раз больше соответствующих безусловных.) Истинная причина ошибки в следующем. Если исходное распределение козырей было 3—3, то каждый игрок мог сбросить свои козыри $3\cdot 2=6$ различными способами, что всего дает $6\cdot 6=36$ возможных вариантов. Если распределение было 4—2 или 2—4, то они могли сбросить свои козыри лишь $4\cdot 3\cdot 2\cdot 1=24$ способами. Теперь мы видим, что очень важно знать, каким было исходное распределение козырей перед заходом с козырей. Учитывая начальную ситуацию, получим отношение, составляющее лишь $24/36=2/3$ отношения, вычисленного выше. Действительно, $286/805:78/161=11/15$ составляет 2/3 от отношения $11/21:10/21=11/10$. Теперь парадокс полностью разрешен.

(ii) Совсем неудивительно, что симметричные или регулярные комбинации выигрывают крайне редко. Если ставят на 5 чисел из 90, то общее число возможных комбинаций приблизительно составляет 44 миллиона (и в точности равно 43 929 268), среди которых регулярных пятерок всего лишь несколько тысяч. В случае выигрыша регулярной пятерки, на которую очень редко ставят другие игроки (хотя шансы выигрыша остаются теми же самыми), величина приза действительно возрастает. Если игрок будет участвовать в лотерее многократно, то спустя некоторое время он в этом убедится.

г) Замечания

(i) Пятерки чисел, на которые ставят редко, обнаружить просто, так как в газетах всегда сообщается, сколько участников лотереи угадали 2, 3, 4 или 5 чисел и сколько составил выигрыш. (Для пятерок чисел, на которые ставят чаще, выигрыш меньше.) В случае прогноза результатов футбольных матчей математический анализ немного сложнее, потому что здесь нет фиксированных комбинаций. Вычисления могут основываться на прогнозах, публикуемых в некоторых газетах, и на числе людей, принимающих во внимание эти прогнозы.

(ii) Публикации по азартным играм (от рулетки, где результаты неизбежно случайны, до бриджа, где влияние случайности сведено до минимума) могли бы заполнить несколько библиотек. В XX веке, во многом благодаря трудам *Джона фон Неймана*, была разработана общая теория игр. Мы вернемся к ней позднее.

(iii) Следующий парадокс появился в 1693 г.(!) в «Философских трудах Королевского общества» (*Philosophical Transaction of the Royal Society*, **17**,677—681) в работе «Арифметический парадокс, касающийся шансов выигрыша в лотереях» члена Королевского общества, почтенного Фрэнсиса Робертса, эсквайра.

«В то время как некоторые истины (подобно аксиомам в геометрии или метафизике) самоочевидны с первого взгляда, другие, не менее справедливые в своих основах, обладают совершенно другими чертами и без строгого и тщательного анализа кажутся неверными. Примеры подобных истин можно найти в большинстве наук... Я предложу еще один пример из арифметики, который, видимо, покажется столь же великим парадоксом, как и любой из известных ранее.

В каждой из двух лотерей игрок платит шиллинг за право тянуть жребий или сделать бросок. Верные исчисления показывают, что шансы на выигрыш у игрока в первой лотерее равны 1 к 3 во второй лотерее — 1 к 2, тем не менее игроку одинаково невыгодно (и не больше) участвовать ни в первой лотерее, ни во второй».

Пример, приводимый после этой проблемы, состоит в следующем (будем использовать современную терминологию). Пусть X обозначает наш выигрыш в игре, зависящей от случая. (При проигрыше величина X отрицательна.) Пусть $X^+ = X$, если X положительна, и 0 в противном случае; $X^- = X$, если X отрицательна, и 0 в противном случае. Обозначим через Y, Y^+, Y^- те же случайные величины для другой игры. Как показывает пример, хотя математические ожидания величин X и Y

совпадают, отношения математических ожиданий величин X^+ и X^- могут отличаться от соответствующего отношения для Y^+ и Y^-. Это означает, что из равенства $E(X)=E(Y)$ не следует равенство $E(X^+)/E(X^-)=E(Y^+)/E(Y^-)$. Вряд ли когонибудь удивит этот результат. (Очевидно, если математические ожидания и X, и Y равны 0, то оба предыдущих отношения равны 1). Если все-таки рассматривать эту проблему как парадокс, то ее скорее следует назвать парадоксом математического ожидания, а не «арифметическим парадоксом».

д) Литература

В следующих книгах рассказано о математике бриджа:

Borel E., Cheron A. *Théorie mathêmatique du bridge*, Gauthier-Villars, Paris, 1940.

Jacoby O. *How to Figure the Odds? Doubleday*, New York, 1947.

Автор следующих статьи и книги заработал миллионы при игре в очко и заставил изменить правила этой игры.

Thorp E. O. "A favourable strategy for twenty-one", *Proc. Nat. Acad. Sci.*, **47**, 110—112, (1961).

Thorp E. O. *Beat the Dealer: A Winning Strategy of the Game Twenty-one*, Blaisdell, New York, 1962.

Для любителей покера:

Findler N. V. "Computer poker", *Sci. Amer.*, **239**, 112—119, (July, 1978).

Работы по истории карточных игр:

Beal G. *Playing cards and Their Story*, London, 1975.

Schreiber W. L. Die altesten Spielkarten, Stuttgart, 1937.

6. Парадокс раздачи подарков; травмы, причиненные лошадьми; телефонные вызовы; опечатки

а) История парадокса

Классическая теория вероятностей занимается в основном комбинаторными задачами (связанными с азартными играми). В этих задачах случайные события обычно состоят из конечного числа возможных элементарных исходов и все элементарные исходы равновероятны. В таком простом случае вероятность события (A) равна отношению числа «благоприятных» исходов к «общему числу исходов». В первой подробной монографии по теории вероятностей также рассматривались такие вероятности. Это была книга *Ремона де Монмора*, опубликованная в Париже в 1708 г. «Парадокс раздачи подарков» — это вариант задачи, обсуждаемой в книге Монмора на примере карточных игр.

б) Парадокс

Несколько человек решили сделать друг другу подарки следующим образом. Каждый приносит подарок. Подарки складываются вместе, перемешиваются и случайно распределяются среди участников. Этот справедливый способ раздачи подарков применяется часто, так как считают, что для больших групп людей вероятность совпадения, т. е. получения кем-то собственного подарка, очень мала. Парадоксально, но вероятность по крайней мере одного совпадения намного больше вероятности того, что совпадений нет (кроме случая, когда группа состоит из двух человек, тогда вероятность отсутствия совпадений равна 50 %).

в) Объяснение парадокса

Рассмотрим компанию из n человек, тогда число подарков также равно n. Подарки могут быть распределены $n!$ различными способами. (Это общее число исходов.) Число исходов, в которых никто не получит свой собственный подарок, равно

$$\binom{n}{0} n! - \binom{n}{1}(n-1)! + \binom{n}{2}(n-2)! -$$
$$- \binom{n}{3}(n-3)! + \ldots (-1)^n 0!,$$

так что отношение числа благоприятных исходов к общему числу исходов вычисляется по формуле

$$p_n = 1/2! - 1/3! + \ldots + (-1)^n/n!$$

и p_n действительно меньше $1/2$ при $n > 2$.

Когда собирается, к примеру, по крайней мере 6 человек ($n \geqslant 6$), имеем $p_n \approx 1/e \approx 0.3679$ с точностью до 4 знаков после запятой. Вероятность определенного совпадения, т. е. вероятность того, что конкретный человек получит свой собственный подарок, очевидно, равна $1/n$, и $1/n$ стремится к 0 при увеличении n. Этот парадокс показывает «по капельке — море, по былинке — стог»: несмотря на малость вероятностей ($1/n$) определенных совпадений, вероятность того, что произойдет по крайней мере одно совпадение, приблизительно равна 2/3.

г) Замечания

(i) Вероятность p_n сходится к e^{-1} с ростом n. Если n не меньше 6, то $p_n = e^{-1}$ с точностью до четырех знаков после запятой. В более общем случае вероятность ровно k совпадений равна $e^{-1}/k!$ (в указанном выше смысле).

(ii) Исследуем другую проблему, связанную с парадоксом раздачи подарков. Опять рассмотрим совокупность из n человек и n подарков. Теперь подарки распределяются таким образом, что каждый человек может получить любой подарок с одинаковой вероятностью независимо от распределения других подарков. Следовательно, может так случиться, что кто-то получит более одного подарка, а другие не получат подарков совсем. Тогда подарки могут быть распределены n^n различными способами (n^n — общее число исходов). Пусть событие A заключается в том, что определенный человек не получит подарка. Тогда все n подарков распределяются среди остальных $(n-1)$ человек, и это можно сделать $(n-1)^n$ разными способами. Таким образом, вероятность события A равна

$$q_n = (n-1)^n/n^n = (1-1/n)^n.$$

Последовательность q_n так же, как и p_n, стремится к e^{-1}. Обобщая наш результат, получим: вероятность того, что определенный человек получит ровно k подарков, сходится к $e^{-1}/k!$ при

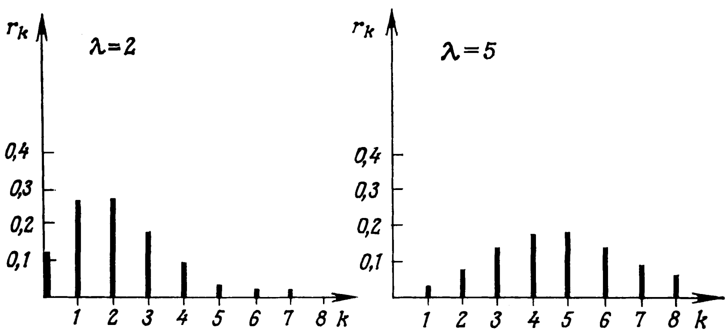

Рис. 2. Распределение Пуассона с параметрами $\lambda = 2$ и $\lambda = 5$.

$n \to \infty$. Рассмотрим теперь случай, когда число людей (n) не обязательно совпадает с числом подарков (m). В этом случае искомая вероятность равна $q_n = (1-1/n)^m$. Если отношение m/n стремится к параметру λ (т. е. если среднее число подарков, приходящееся на одного человека, равно λ или стремится к λ), то q_n сходится к $e^{-\lambda}$ (где λ может быть любым положительным числом). Наконец, вероятность того, что определенный человек получит ровно k подарков, сходится к

$$r_k = \lambda^k e^{-\lambda}/k!.$$

Говорят, что случайная величина, принимающая лишь неотрицательные целочисленные значения, имеет *распределение Пуассона*, если она принимает значение k с вероятностью r_k. Выше мы видели, что случайное число подарков, которые получит определенный человек, приблизительно подчиняется распределению Пуассона с параметром λ, если среднее число (математическое ожидание числа) подарков, приходящихся на одного человека, равно λ. Возвращаясь к «парадоксу раздачи подарков», получаем, что число людей, которым достанутся их собственные подарки, также подчиняется распределению Пуассона с параметром $\lambda = 1$. Это достаточно естественно, так как в среднем существует только один человек, получивший свой собственный подарок. (Вероятность того, что определенный человек получит свой собственный подарок, равна $1/n$, и для n человек сумма таких вероятностей даст 1, каково бы ни было n.)

(iii) Понятие распределения Пуассона впервые появилось в книге французского ученого *Симеона Дениса Пуассона* (1781—1840 гг.). (В работе «Исследования о вероятности судебных приговоров по уголовным и гражданским делам» ("Recherches sur la Frobabilité des Jugements en Matière Criminelle et en Matière Civile, Précédées des Gègles Génerales du Calcul des Probabilités"), опубликованной в 1837 г., разд. 81 посвящен приложениям теории вероятностей к судебной практике.) Пуассон обсуждает следующую проблему. Рассмотрим эксперимент, в котором один и тот же феномен наблюдается повторно. Предполагается, что испытания в этом эксперименте независимы, для каждого испытания есть только два возможных исхода и их вероятности остаются неизменными для всех испытаний. (Эксперимент такого типа называется последовательностью испытаний Бернулли.) Как правило, исход, происходящий с вероятностью p, называют «успехом», а другой исход — «неудачей». Пример испытаний Бернулли дают последовательные бросания несбалансированной монеты. Пусть b_k — вероятность того, что при n испытаниях Бернулли будет k успехов (например, вероятность того, что выпадет ровно k гербов при n бросаниях несимметричной монеты). Тогда

$$b_k = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0,\ 1,\ \ldots,\ n.$$

Таким образом, число успехов S_n в n испытаниях Бернулли есть случайная величина, которая принимает значение k с вероятностью b_k. Говорят, что случайная величина с возможными значениями $0, 1, 2, \ldots, n$ имеет биномиальное распределение, если она принимает значение k с вероятностью b_k. Прилагательное «биномиальное» указывает на тот факт, что b_k как раз является k-м членом в разложении бинома $(p + (1-p))^n$, поскольку по биномиальной формуле $(p + (1-p))^n = b_0 + b_1 + b_2 + \ldots + b_n$. Пуассон обнаружил, что если делать p все меньше и меньше и в то же время увеличивать n так, что произведение $np = \lambda$ остается постоянным, то b_k стремится к r_k. Следовательно, распределение Пуассона является приближением для биномиального распределения. Возможность широкого применения и большая важность пуассоновского распределения не были поняты в середине прошлого века; более того, о нем почти полностью забыли. Однако после 1894 г. это распределение использовали при изучении странного феномена. За 20 лет между 1875 и 1894 годами в 14 различных кавалерийских корпусах германской армии была собрана статистика трагических случаев, когда солдат был убит ударом копыта. Согласно 280 наблюдениям[1]), 196 солдат погибли таким образом, т. е. в среднем за год $\lambda = 0{,}7$. Если бы число трагических исходов подчинялось распределению Пуассона с параметром $\lambda = 0.7$, то можно было бы ожидать, что при 280 наблюдениях в 139 случаях смертей нет, в 97 случаях — 1 смерть, в 34 случаях — 2 смерти и т. д. А что показала статистика? В действительности данные были соответственно 140, 91, 32 и т. д.; практика и теория оказались в столь хорошем согласии, что вряд ли можно было ожидать большего.

Это сравнение появилось в 1898 г. в знаменитой монографии *Л. Борткевича.* Название его книги «Закон малых чисел» отражает тот факт, что при пуассоновской аппроксимации p стремится к 0 при увеличении n. (Название достаточно двусмысленно, так как оно наводит на ошибочную мысль о том, что пуассоновское приближение в каком-то смысле противоположно закону больших чисел, который обсуждается ниже.) Пуассоновское распределение стало широко применяться лишь в XX веке. Например, пуассоновским распределением приближенно описывается количество определенного вида товаров, проданное в течение конкретного дня, число молекул гемоглобина, видимое под микроскопом, число забастовок и войн в году, число опечаток в тексте или число телефонных соединений с определенным номером в течение фиксированного дня. Если среднее число молекул гемоглобина или опечаток, или телефонных соединений равно λ, то их число имеет распределение Пуассона с параметром λ. Если N — число телефонных линий, доступных для телефонного обмена, то число занятых линий приблизительно имеет распределение Пуассона. Проблемы такого вида изучались датским математиком *А. К. Эрлангом* (1878—1929 гг.). В 1906 г. он установил, что лучшее приближение можно получить, если использовать следующее усеченное пуассоновское распределение:

$$e_k = c\lambda^k/k!, \text{ где } k = 0, 1, \ldots, N$$

и

$$c = \left(\sum_{k=0}^{N} \lambda^k/k!\right)^{-1}.$$

---

[1]) 280 наблюдений = 20 лет $\times$ 14 корпусов. — *Прим. перев.*

С тех пор это распределение называется распределением Эрланга.

д) *Литература*

Haight F. A. *Handbook of the Poisson Distribution*, Wiley, New York, 1967.

Montmort P. R. *Essay d'analyse sur les jeux de hazard*, Paris, 1708.

Sheynin O. B. "S. D. Poisson's work in probability", *Archive for the History of Exact Sciences*, **18**, 245—300, (1978).

Stigler S. M. "Poisson on the Poisson distribution", *Statistics and Probabillity Letters*, **1**, 33—35, (1982).

Takás L. "The problem of coincidence", *Archive for the History of Exact Sciences*, **21**, 229—245, (1980).

7. Санкт-петербургский парадокс

а) *История парадокса*

Теория вероятностей, которая начиналась с исследования результатов азартных игр, развилась в универсальную теорию, нашедшую применения во многих областях жизни. Поэтому совсем неудивительно, что почти все крупные научные журналы последовали примеру английского журнала «Труды по философии» и регулярно публиковали статьи по теории вероятностей. Все больше и больше ученых считали, что теория вероятностей — это не что иное, как путеводитель по жизни, здравый смысл, выраженный в числах. Однако в начале XVII века Академия наук в Санкт-Петербурге опубликовала статью, математические вычисления в которой казалось противоречили здравому смыслу. Статью написал *Даниил Бернулли*, и благодаря ему петербургский парадокс стал известен. Однако впервые проблему поднял его двоюродный брат *Николай Бернулли* и упомянул о парадоксе в письме к Монмору в сентябре 1713 г. (Бернулли — знаменитая семья математиков, несколько членов которой занимались теорией вероятностей, особенно Якоб Бернулли, о котором еще пойдет речь ниже в связи с законами больших чисел.)

б) *Парадокс*

Единичное испытание в петербургской игре состоит в бросании правильной монеты до тех пор, пока не выпадет решка; если это произойдет при r-м бросании, игрок получает 2^r долларов из банка. Таким образом, с каждым бросанием выигрыш удваивается. Вопрос в следующем: сколько следует заплатить игроку за участие в игре, чтобы игра стала безобидной? Безобидность петербургской игры рассматривается в классическом смысле: среднее значение (или математическое ожидание) чистого выигрыша должно быть равно 0. Однако, как ни удивительно, это естественное требование невыполнимо, какую бы (конечную) сумму денег игрок ни заплатил.

в) Объяснение парадокса

Потери банка имеют бесконечное математическое ожидание, так как вероятность окончания игры при k-м бросании равна $1/2^k$, и в этом случае игрок получает 2^k долларов. Тогда банк в среднем должен заплатить

$$\frac{1}{2}\cdot 2+\frac{1}{4}\cdot 4+\frac{1}{8}\cdot 8+\ \ldots\ =1+1+1+\ \ldots$$

долларов, что составляет бесконечно большую сумму денег, так что игра стала бы безобидной при бесконечном взносе. Хотя все математические вычисления корректны, результат неприемлем, поэтому некоторые математики предположили реализуемые модификации.

(i) *Бюффон*, *Крамер* и другие предложили исходить из естественного предположения об ограниченности ресурсов (т. е. банк имеет лишь ограниченное количество денег). Пусть в банке есть миллион долларов. Тогда математическое ожидание выигрыша для игрока равно

$$\frac{1}{2}\cdot 2+\frac{1}{4}\cdot 4+\ \ldots\ +\frac{1}{2^{10}}\cdot 2^{10}+\left(\frac{1}{2^{20}}+\frac{1}{2^{21}}+\ \ldots\right)\cdot 10^6=$$
$$=19+1.90\ \ldots\ \approx 21$$

(мы учли, что $2^{20}>10^6$). Следовательно, при вступительном взносе игрока, равном 21 доллару, игра станет в некоторой степени выгодной для банка.

(ii) *В. Феллер* отметил, что можно так определить вступительный взнос, что петербургская игра станет безобидной. Обозначим через n число игр, в которых участвовал игрок. Игру можно считать безобидной, если отношение суммарного выигрыша N_n к суммарному вступительному взносу R_n сходится к 1 при n, стремящемся к бесконечности, точнее, если для любого $\varepsilon>0$

$$P\{|N_n/R_n-1|<\varepsilon\}\to 1 \quad \text{при} \quad n\to\infty. \qquad (*)$$

Феллер доказал, что петербургская игра становится безобидной, если положить $R_n=n\log_2 n$. Как следует из парадокса, игра не может быть безобидной для $R_n=cn$, где c — произвольная конечная постоянная. Однако, если вступительный взнос может зависеть от числа игр, в которых участвовал игрок, то (согласно теореме Феллера) петербургский парадокс разрешается.

г) Замечания

(i) Соотношение (*) выражает свойство устойчивости величины N_n. Аналогичные свойства наблюдаются и при $R_n = cn$ в «Парадоксе Бернулли для закона больших чисел».

(ii) По результатам 2084 игр *Бюффон* обнаружил, что игра становится безобидной при вступительном взносе, приближенно равном 10 долларам.

(iii) Следующий парадокс близок санкт-петербургскому парадоксу. (Я услышал о нем от Сэма Гутмана после моего выступления на семинаре Дадли в Массачусетском технологическом институте в 1983 г.) Предположим, что вам предоставлена возможность выиграть $(-2)^n$ долларов с вероятностью 2^{-n}, $n = 1, 2, 3, \ldots$. Это вас радует или огорчает? Ответ: вы и обрадованы, и огорчены. Вы обрадованы потому, что такая лотерея эквивалентна двухступенчатой лотерее, в которой на первом шаге определяется, в какой из лотерей, каждая из которых для вас выгодна (математическое ожидание выигрыша положительно), вы участвуете на втором шаге. Точнее, с вероятностью $(2^{-1} + 2^{-2} + 2^{-4})$ вы участвуете в лотерее, в которой выигрываете $(-2)^j$ долларов с вероятностью

$$2^{-j}/(2^{-1} + 2^{-2} + 2^{-4}) \quad (j = 1,\ 2,\ 4),$$

или с вероятностью $(2^{-3} + 2^{-6} + 2^{-8})$ вы участвуете в лотерее, в которой выигрываете $(-2)^k$ долларов с вероятностью

$$2^{-k}/(2^{-3} + 2^{-6} + 2^{-8}) \quad (k = 3,\ 6,\ 8),$$

и т. д. В каждой из лотерей на втором шаге возможны три выигрыша, и математическое ожидание выигрыша положительно. Так что вы рады. Однако исходную лотерею можно переделать и в лотереи с тремя выигрышами, в каждой из которых математическое ожидание отрицательно. Выигрышами в первой лотерее являются $(-2)^1$, $(-2)^2$, $(-2)^3$, во второй — $(-2)^4$, $(-2)^5$, $(-2)^7$ и т. д. Так что вы одновременно огорчены. [Для тех, кто знаком с понятием условного математического ожидания, переформулируем парадокс иначе. Представим себе, что условное математическое ожидание $E(X|Y)$ определено не как обычно, а как $\int xP(dx|Y)$, где $P(dx|Y)$ определено обычным образом. Тогда существуют случайные величины X, Y и Z, такие, что $E(X|Y) > 0 > E(X|Z)$ с вероятностью 1! Действительно, пусть X — конечный выигрыш в лотерее, т. е. $X = (-2)^n$ с вероятностью 2^{-s}. Положим $Y = 1$, если нам досталась первая «положительная» лотерея (т. е. $X = (-2)$, $(-2)^2$ или $(-2)^4$), положим $Y = 2$, если нам досталась вторая лотерея (т. е. $X = (-2)^3$,

$(-2)^6$ или $(-2)^8$) и т. д. Пусть $Z=1$, если нам досталась первая «отрицательная» лотерея (т. е. $X=(-2)$, $(-2)^2$, $(-8)^3$), и $Z=2$, если нам досталась вторая лотерея (т. е. $X=(-2)^4$, $(-2)^5$, $(-2)^7$) и т. д.]

д) Литература

Feller W. *An Introduction to Probability Theory and Its Applications*, New York, John Wiley, 1969. [Имеется перевод: Феллер В. Введение в теорию вероятностей и ее приложения. — М.: Мир, 1984.]

Hinčin A. Ja. "Su una legge dei grandi numeri generalizzata", *Giorn. Ist. Ital. d. Attuari*, 7, 365—377, (1936).

Martin-Löf A. "A limit theorem which clarifies the 'Petersburg paradox'", *J. Appl. Prob.*, 22, 634—643, (1985).

8. Парадокс смертности населения. Безвозрастный мир атомов и слов

а) История парадокса

Математические исследования по смертности населения и продолжительности жизни начались на раннем этапе развития капитализма благодаря потребностям страховых компаний. Вслед за результатами, полученными *Джоном Граунтом* (1662 г.), *ван Худденом* и *Джоном де Уиттом* (1671 г.), *Эдмунд Галлей* (открывший комету, названную его именем) опубликовал в 1693 г. статью о таблицах смертности, которая положила начало математической теории страхования жизни. Следующий парадокс (замеченный Даламбером) показывает одну из «зубодробильных» проблем новой теории.

б) Парадокс

По таблице Галлея средняя продолжительность жизни равна 26 годам и вместе с тем с равными шансами можно умереть до 8 лет и прожить больше 8 лет.

в) Объяснение парадокса

Действительно, согласно таблице Галлея, с равными шансами можно прожить больше 8 лет и умереть в возрасте до 8 лет, но если человек дожил до 8 лет, то он может прожить еще несколько десятилетий. Следовательно, неудивительно, что средняя продолжительность жизни намного больше 8 лет.

Предположим, что среди тысячи человек лишь один достигает возраста Мафусаила[1]. Тогда средний возраст значительно

[1]) Мафусаил — в ветхозаветных преданиях один из праотцов человечества, прославившийся своим долголетием («мафусаилов век»), он прожил 969 лет. — *Прим. перев.*

увеличится, но их вероятная продолжительность жизни (возраст, до которого они доживают с вероятностью 50 %) существенно не изменится.

г) Замечания

(i) Пусть $F(x)$ обозначает вероятность того, что продолжительность жизни человека, случайно выбранного из некоторой группы населения, меньше, чем x единиц времени. ($F(x)$ — функция распределения продолжительности жизни.) Предположим, что $F(x)$ имеет плотность $f(x)$. Средняя продолжительность жизни вычисляется по формуле $M = \int_0^\infty xf(x)\,dx$. С другой стороны, вероятная продолжительность жизни m определяется из уравнения $F(m) = 1/2$. Другими словами, за период времени m вымирает половина населения. Из этих формул ясно, что, вообще говоря, M и m принимают совершенно разные значения. M есть математическое ожидание продолжительности жизни, а m называется медианой этой продолжительности.

(ii) Понятие смертности нетрудно расширить. Если рассматривать амортизацию промышленных изделий или распад атомов как смерть, то получаем широко применимую математическую теорию, возникающую из исследований смертности населения. Однако в этой более широкой области также могут наблюдаться достаточно парадоксальные явления. В то время как человек не бессмертен, не вечно молод, в природе и обществе можно найти так называемые безвозрастные объекты. Определим понятие безвозрастности. Считаем объект безвозрастным, если вероятность существования объекта в течение определенного интервала времени не зависит от времени, которое объект уже существовал. Очевидно, человек таким свойством не обладает, так как, чем он старше, тем с большей вероятностью он умрет в течение заданного промежутка времени. Интересно, что не все объекты подобны нам. Например, радиоактивные атомы безвозрастны. Если средняя продолжительность существования безвозрастного объекта равна T, то вероятность того, что он не прекратит своего существования за следующий период времени x, равна $e^{-x/T}$, где x — положительное число. Свойство безвозрастности радиоактивных частиц вытекает из того факта, что их скорость распада пропорциональна числу нераспавшихся частиц. Коэффициент пропорциональности называется постоянной распада и обозначается λ. Если в момент времени $t = 0$ было N_0 нераспавшихся частиц, то (так как скорость распада равна постоянной, получаемой путем интегрирования) в момент времени x число нераспавшихся частиц составляет $N_x = N_0 e^{-\lambda x}$. Это означает, что вероятность выживания до момента x равна $e^{-\lambda x}$. Следовательно, радиоактивные частицы действительно безвозрастны, и их средняя продолжительность существования равна $T = 1/\lambda$. Другими словами, продолжительность существования радиоактивных частиц описывается *показательным распределением* с параметром λ, т. е. имеет плотность вероятности $\lambda e^{-\lambda x}$ [1]). Период полураспада безвозрастных объектов (период, в течение которого половина объектов прекращает существование) есть корень следующего уравнения:

$$e^{-\lambda x} = 1/2, \quad \text{т. е.} \quad x = \ln 2/\lambda.$$

(iii) Понятие периода полураспада безвозрастных объектов стало фундаментальным в некоторых областях науки. Радиоуглеродный метод, разработанный американским химиком *Уиллардом Фрэнком Либби*, до сих пор является самым распространенным методом датирования в области археологической хронологии. (В 1960 г. за это открытие ему была присуждена Нобелевская премия.) В 1950 г., следуя идеям Либби, *М. Свадеш* применил его метод в лингвистике, предполагая, что не только радиоактивные атомы, но и атомы речи, т. е. слова, можно считать безвозрастными. Предполагается, что период полураспада древнего базового словаря языков составляет 2000 лет. Используя эту идею, можно определить дату, когда два родственных языка (например, латинский язык и санскрит) разделились. Для нахождения даты разделения двух языков нужно только знать, какая часть базового словаря до сих пор существует в обоих языках. *А. Раун* и *Е. Кангсмаа-Минн* сравнили венгерский и финский языки. Они обнаружили, что идентичные элементы в этих языках составляют соответственно 21 % и 27 %. (Вычисления проводились разными методами.) Отсюда был сделан вывод, что венгерский и финский языки разделились приблизительно 4—5 тысяч лет назад. Метод Свадеша, разработанный 30 лет назад, применяется очень часто и известен как лексикостатистика или глоттохронология. (Оригинальная статья Свадеша опубликована в «Международном журнале американской лингвистики».)

(iv) Предположим, что λ — постоянная распада. Тогда вероятность распада ровно k частиц за время t равна

$$(\lambda t)^k e^{-\lambda t}/k!.$$

Это означает, что число распавшихся частиц есть случайная величина, имеющая распределение Пуассона, которое нам уже из-

---

[1]) Плотность вероятности показательного распределения с параметром λ равна $\lambda e^{-\lambda x}$, если $x > 0$, и 0 в противном случае. — *Прим. перев.*

вестно из парадокса раздачи подарков. Математическое ожидание этого распределения равно λt, что довольно естественно.

(v) Мы увидели, что существуют безвозрастные объекты. Что еще более удивительно, существуют объекты, которые со временем становятся «моложе», например механизмы, у которых в период их функционирования есть интервал, когда вероятность безотказной работы в течение определенного промежутка со временем растет. Легко видеть, что математически в обозначениях пункта (i) это означает убывание $f(x)/(1-F(x))$ (интенсивности выхода из строя) как функции от x. Изучение этой интенсивности очень важно в теории надежности и проблемах хранения.

(vi) Наконец, остановимся на одном любопытном вопросе, связанном со смертностью населения. Можно ли с помощью вероятностных методов оценить общее число людей, когда-либо живших на Земле? Обоснование следующего удивительного утверждения дано в книге Голдберга (см. ссылку ниже): «9 процентов когда-либо живших на Земле людей живут сейчас». Это утверждение было вынесено в заголовок статьи в «Нью-Йорк Таймс» (6 октября 1981 г., с. 61).

д) Литература

Barlow E. R., Proschan F. *Statistical Theory of Reliability and Life testing*, New York, Holt, Rinehart and Winston, Inc., 1975.

Беляев Ю. К., Гнеденко Б. В., Соловьев А. Д. Математические методы в теории надежности. — Москва: Наука, 1965.

Goldberg S. *Probability in Social Science*, Birkhäuser, Boston — Basel — Stuttgart, 1983.

Halley E. "An estimate of the degrees of mortality of mankind, drawn from curious tables of the births and funerals at the city of Breslau; with an attempt to ascertain the price of annuities upon lives", *Philosophical Transactions of the Roy. Soc.*, **17**, 596—610, 654—656, (1963).

(Интересно, что Галлей, будучи англичанином, тем не менее использовал данные по Бреслау (сейчас Вроцлав), который был тогда столицей Силезии, а не по Лондону или Дублину. Галлей использовал ежемесячные сведения о числе рождений и смертей в этом месте за период 1687—1691 гг., потому что считал, что влияние миграции населения, искажающее результаты, в Бреслау намного меньше.)

9. Парадокс закона больших чисел Бернулли

а) История парадокса

В математике найдется немного законов, которые столь же часто понимались превратно, как законы больших чисел. (Не слишком широко известно даже то, что существует несколько законов больших чисел.) Первый закон больших чисел был доказан Якобом Бернулли (1654—1705 гг.) в его книге, названной «Ars conjectandi» («Искусство предположений»), которая была опубликована только после смерти автора в 1713 г. Сам Бернулли не использовал понятия «закон больших чисел»; это название ввел Пуассон лишь в 1837 г. По закону Бернулли, если правильную монету бросают n раз и при этом k раз выпадает герб, то при увеличении числа бросаний (n) отношение k/n (относительная частота выпадения герба) стремится к 1/2. Точнее, для произвольных положительных чисел ε и δ и достаточно большого n (зависящего от ε и δ) величина $|k/n - 1/2|$ меньше ε с вероятностью, превосходящей $1 - \delta$. Эта теорема вовсе не столь сложна, как можно было бы ожидать из-за большого числа ее неправильных трактовок и парадоксов, к которым она привела. Наиболее типичный парадокс приведен ниже.

б) Парадокс

Игроки часто уверены, что если правильная монета много раз падает гербом, то, согласно закону больших чисел, вероятность выпадения решки с необходимостью возрастает. (В противном случае нарушалось бы то, что при очень большом числе бросаний выпадения герба и решки происходят приблизительно одинаково часто.) С другой стороны, у монет, очевидно, нет памяти, поэтому они не знают, сколько раз они уже выпадали гербом или решкой. По этой причине шансы выпадения герба при каждом бросании равны 1/2, даже если монета уже выпала гербом тысячу раз подряд. Не противоречит ли это закону Бернулли?

в) Объяснение парадокса

По закону Бернулли при очень большом числе бросаний герб выпадает приблизительно столько же раз, сколько и решка, но вся суть в том, что означает «приблизительно». Игрок, который полагает, что *разность* между числом выпадений герба и числом появлений решки должна быть очень мала, ошибается, так как закон Бернулли утверждает лишь, что отношение числа выпадений герба к общему числу бросаний приближенно равно 1/2 (с вероятностью, близкой к 1) или, что то же самое, отношение числа выпадений герба к числу появлений решки приблизительно равно 1. Другими словами, *разность логарифмов этих чисел* стремится к 0 (при увеличении числа бросаний). Если бы разность самих чисел была мала, то это противоречило бы отсутствию памяти у монет.

г) Замечания

(i) Теперь ясно, что, как бы долго мы ни наблюдали последовательное выпадение гербов, вероятность появления решки при следующем бросании все же не становится больше. Но возникает следующий вопрос. Предположим, что монету бросают n раз. Какой максимальной длины серию из гербов мы можем ожидать? При n бросаниях, если $n = 100$, можно ожидать серию в 6—7 гербов подряд, если $n = 1000$, можно ожидать 9—10 гербов подряд, и 19—20 для $n = 10^6$. Следующую теорему доказали *Паул Эрдеш* и *Альфред Реньи*. При бросании монеты n раз серия из гербов длины $\log_2 n$ наблюдается с вероятностью, стремящейся к 1 при $n \to \infty$. Этот факт очень полезен, когда нужно решить, описывает ли последовательность, составленная из двух символов, результаты бросания монеты или кто-то ее придумал, «тщательно» избегая включения длинных серий. Из-за широко распространенного неправильного понимания закона больших чисел Бернулли многие люди не будут повторять один и тот же знак 7 или более раз подряд в последовательности из 100 знаков.

(ii) В силу предыдущего замечания чистые серии (состоящие только из гербов или решек) могут быть достаточно длинными. С другой стороны, легко посчитать, что ожидаемая длина 1-й, 2-й, 3-й и т. д. чистой серии всегда равна 2. Если же монета несимметрична и вероятность выпадения герба равна $0 < p < 1$, то ожидаемая длина 1-й, 3-й,... (каждой нечетной) серии равна $p/q + q/p$, а ожидаемая длина 2-й, 4-й,... (каждой четной) чистой серии всегда равна 2 (независимо от p(!)). Сумма $p/q + q/p$ не может быть меньше 2; это означает, что нечетные серии по крайней мере не короче четных. Этот факт вовсе неудивителен, потому что при первом бросании монета скорее всего выпадет более вероятной стороной. Таким образом, первая серия имеет больше шансов быть длинной, чем короткой, так что она в среднем будет длинной или по крайней мере длиннее второй серии, ожидаемая длина которой равна всего лишь 2. Удивительно же то, что ожидаемая длина четных серий не зависит от p.

(iii) Закон больших чисел Бернулли можно кратко выразить с помощью понятия *сходимости по вероятности*. Говорят, что последовательность случайных величин $X_1, X_2, \ldots$ сходится по вероятности к случайной величине X, если вероятность события $|X_n - X| > \varepsilon$ сходится к нулю для любого положительного ε, т. е. если $P(|X_n - X| > \varepsilon) \to 0$. (Парадоксально, но может случиться так, что последовательность случайных величин $X_1, X_2, \ldots$ сходится по вероятности к 0, а последовательность $(X_1 + X_2 + \ldots + X_n)/n$ нет.) Закон Бернулли утверждает, что относительная частота k/n события сходится по вероятности к вероятности этого события. Для доказательства сходимости по вероятности обычно используют неравенство Чебышева—

Бьенеме, согласно которому, если E — математическое ожидание случайной величины X и D^2 — ее дисперсия, то

$$P(|X-E|>\varepsilon)\leqslant D^2/\varepsilon^2.$$

Интересно, что русский ученый *П. Л. Чебышев* и французский ученый *Ж. Бьенеме* опубликовали свои неравенства, которые они открыли независимо друг от друга, в одном и том же номере французского журнала (*J. Math. Pures et Appl.* XII. 1867). Из этого неравенства сразу следует, что если распределения независимых случайных величин $X_1, X_2, \ldots$ одинаковы и их дисперсия D^2 конечна, то арифметическое среднее

$$(X_1+X_2+\ldots+X_n)/n$$

сходится по вероятности к общему математическому ожиданию величин X_i (дисперсия арифметического среднего равна D^2/n и стремится к 0 при $n\to\infty$). Это есть один из общих (слабых) законов больших чисел. В слабых законах больших чисел речь идет о сходимости по вероятности, а в сильных законах больших чисел устанавливается сходимость с вероятностью 1. Следующее замечание относится к сильным законам больших чисел.

(iv) Среди сильных законов больших чисел наиболее известна теорема Колмогорова: если $X_1, X_2, \ldots$ — взаимно независимые случайные величины с одинаковой функцией распределения, имеющей (общее) конечное математическое ожидание E, то арифметическое среднее

$$(X_1+X_2+\ldots+X_n)/n$$

сходится к E при $n\to\infty$ с вероятностью 1. Если случайные величины положительны и $S_n^{(k)}$ обозначает элементарный симметричный многочлен

$$\sum_{1<j_1<j_2<\ldots<j_k<n} X_{j_1}X_{j_2}\ldots X_{j_k},$$

то

$$\sqrt[k]{S_n^{(k)}\Big/\binom{n}{k}}$$

также сходится к E (с вероятностью 1), где $k=k(n)$ — натуральное число такое, что $k/n\to 0$ при $n\to\infty$. С вероятностью 1 также существует предел даже в том случае, когда k/n стремится к некоторому положительному числу c. Тогда пределом будет постоянная, зависящая от c (если $0<c<1$, то достаточно предполагать, что конечно математическое ожидание величины $\log(1+X_i)$, и то же самое предполагаем относительно $|\log X_i|$, если $c=1$, см. статью Холаса и Секея). Проблема значительно усложняется, если случайные величины могут принимать как отрицательные, так и положительные значения.

(v) «Закон больших чисел» неверен в смысле категории (см. статью Мендеза).

e) Литература

Erdos P., Rényi A. "On a new law of large numbers", *Jour. Analyse Mathematique*, **23**, 103—111, (1970).

Erdös P., Révézs P. "On the length of the longest head-run", *Coll. Math. Soc. J. Bolyai*, **16** (ed. Csiszár, I., Ellias P.) (**1975**).

Halász G., Székely G. J. "On the elementary symmetric polynomials of independent random variables", *Acta Math. Acad. Sci. Hung.*, **28**, 397—400, (1976).

Komlós J. "A generalization of a problems of Steinhaus", *Acta Math. Acad. Sci. Hung.*, **18**, 217—229, 1967.

Mendez C. G. "On the law of large numbers, infinite games, and category", *The American Math. Monthly*, **88**, 40—41, (1981).

Móri T. F., Székely G. J. "Asymptotic behaviour of symmetric polinomial statistics", *Annals of Probability*, **10**, 124—131, 1982.

Móri T. F., Székely G. J. "Asymptotic independence of "pure-head" stopping times", *Statistics and Probability Letters*, **2**, 5—8, 1984.

Révész P. *The Laws of Large Numbers*, Akadémiai Kiadó, Budapest, 1967.

10. Парадокс де Муавра; экономия энергии

а) История парадокса

Одной из выдающихся фигур в теории вероятностей является *Абрахам де Муавр* (1667—1754 гг.). Этот математик, родившийся во Франции, после отмены Нантского эдикта (который предоставлял гугенотам свободу вероисповедания) переехал в Англию. Там в 1718 г. была опубликована его основная работа «Доктрина шансов» («The Doctrine of Chances»). В 3-м издании этой книги (1756 г.) де Муавр так писал о своем открытии мирового значения (о котором он сообщил некоторым друзьям еще в 1733 г.), содержащем в себе намного больше, чем закон больших чисел Бернулли: «...Возьму на себя смелость утверждать, что это труднейшая проблема о случайном...». Без сомнения, открытое де Муавром нормальное распределение, стало краеугольным камнем науки о случайном. (Де Муавр, как это ни странно, не включил свое открытие во 2-е издание книги в 1738 г.).

б) Парадокс

Согласно закону больших чисел Бернулли, вероятность того, что при бросании монеты число выпадений герба приблизительно равно числу появившихся решек, стремится к 1 при увеличении числа бросаний (приближенное равенство чисел означает, что их отношение стремится к 1). С другой стороны, вероятность того, что число гербов будет в точности равно числу решек, стремится к нулю. Например, при 6 бросаниях монеты вероятность выпадения 3 гербов равна 5/16; при 100 бросаниях вероятность выпадения 50 гербов равна 8 %; при 1000 бросаний вероятность выпадения 500 гербов составляет менее 2 %. В общем случае, когда монету бросают $2n$ раз, вероятность того, что герб выпадает ровно n раз, равна $p=\binom{2n}{n}2^{-2n}$, и для достаточно больших n вероятность p приближенно равняется $1/\sqrt{\pi n}$, что действительно стремится к нулю с ростом n. Суммируем сказанное: вероятность того, что число гербов приближенно равно числу решек, стремится к 1: в то же время вероятность того, что число гербов в точности совпадает с числом решек, стремится к 0. Разрыв между этими двумя фактами был окружен «атмосферой парадоксальности» до тех пор, пока де Муавр не построил над ним математический мост.

в) Объяснение парадокса

Пусть H_n и T_n обозначают число выпадений герба и решки соответственно при n бросаниях монеты. Согласно закону больших чисел Бернулли, вероятность того, что разность $H_n - T_n$ становится пренебрежимо малой по сравнению с n, стремится к 1 (что совсем неудивительно). Однако де Муавр заметил, что величина $|H_n - T_n|$ не является пренебрежимо малой по сравнению с $\sqrt{n}$. Он вычислил, например, что для $n = 3600$ вероятность того, что $|H_n - T_n|$ не превосходит 60, равна 0.682688... . Пусть x — произвольное положительное число, и пусть $A_n(x)$ обозначает вероятность того, что $|H_n - T_n| < x\sqrt{n}$. Согласно де Муавру, $A_n(x)$ стремится с ростом n к величине $A(x)$, заключенной между 0 и 1. При увеличении x от нуля до бесконечности $A(x)$ также постоянно возрастает от нуля до единицы (см. замечание (i)). Эта функция $A(x)$ и есть тот математический мост, о котором говорилось выше. Для нахождения $A(x)$ де Муавр использовал формулу Стирлинга, которую он открыл независимо от Стирлинга. (Формула *Джеймса Стирлинга* была доказана в 1730 г., она утверждает, что $n!$ приблизительно равняется $\sqrt{2\pi n}\,(n/e)^n$.)

г) Замечания

(i) Точный вид функции $A(x)$ следующий:

$$A(x) = \sqrt{2/\pi}\int_0^x e^{-u^2/2}\,du.$$

Используя эту формулу, теорему Муавра можно записать в следующем виде

$$\lim_{n\to\infty} P\left(|H_n - T_n| < x\sqrt{n}\right) = A(x), \quad \text{если} \quad x > 0,$$

или

$$\lim_{n\to\infty} P\left(H_n - T_n < x\sqrt{n}\right) = \Phi(x),$$

где

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du.$$

Говорят, что случайная величина, которая принимает значения, не превосходящие x, с вероятностью $\Phi(x)$ (где x — произвольное действительное число), подчиняется стандартному нормальному распределению. В силу результата де Муавра величина

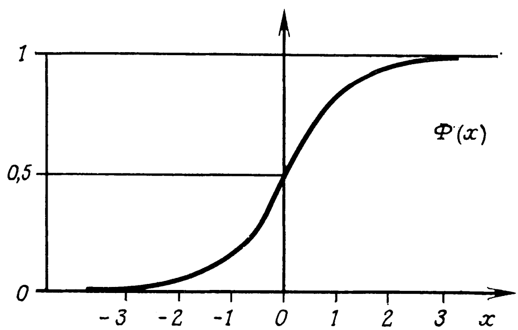

Рис. 3. Стандартная нормальная функция распределения.

$(H_n - T_n)/\sqrt{n}$ приблизительно подчиняется стандартному нормальному распределению (если n достаточно велико).

[В табл. 1 в конце книги приведены значения функции $\Phi(x)$.]

(ii) Поскольку $H_n + T_n = n$, указанный выше результат можно сформулировать следующим образом:

$$\lim_{n\to\infty} \left(H_n < n/2 + x\sqrt{n}/2\right) = \Phi(x).$$

Де Муавр также исследовал случай, когда у монеты смещен центр (монета несимметрична) и герб выпадает с вероятностью p, а решетка — с вероятностью $1 - p$. Тогда справедливо соотношение

$$\lim_{n\to\infty} P\left(H_n < np + x\sqrt{np(1-p)}\right) = \Phi(x),$$

известное под названием «предельная теорема Муавра — Лапласа». Эту теорему можно широко использовать в различных областях при планировании, например при планировании расхода энергии.

Пример. Рассмотрим 300 одинаковых станков на фабрике. Если в среднем 70 % станков работают и 30 % находятся в ремонте, то нужно обеспечить энергией в среднем 210 станков. Од-

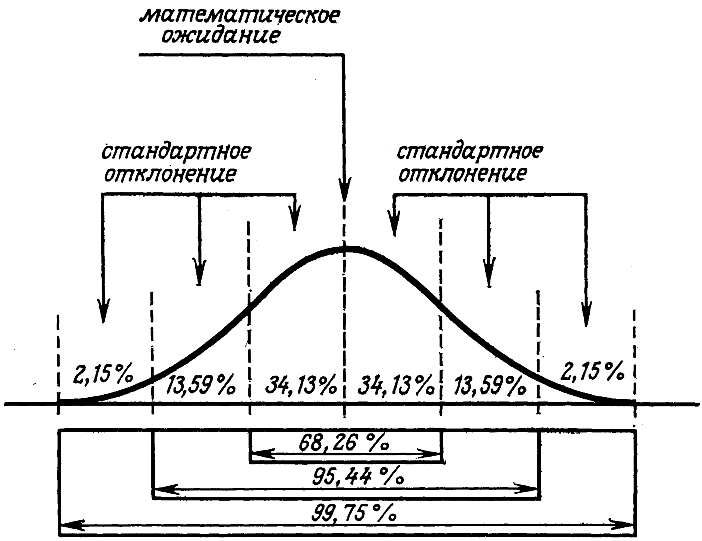

Рис. 4. Нормальная плотность вероятности.

нако иногда могут работать все 300 станков. Каким количеством энергии нужно обеспечить фабрику, чтобы с вероятностью 99.9 % все исправные станки могли работать? (Предполагается, что станки выходят из строя независимо друг от друга.) В приведенной выше формуле H_n теперь означает число работающих станков, $n = 300$ и $p = 0.7$. По табл. 1 для $x = 3$ находим $\Phi(x) \approx 0.999$. Для этих значений имеем $np + x\sqrt{np(1-p)} = 210 + 3\sqrt{63}$, так что достаточно принимать во внимание 234 станка. (Однако на практике учитывают практически все 300 станков, проявляя тем самым излишнюю осторожность.)

(iii) Теорему Муавра — Лапласа, которую мы обсуждали выше, можно обобщать в различных направлениях. Представители математической школы Санкт-Петербурга во главе с *П. Л. Чебышевым* (1821—1894 гг.), в особенности *А. М. Ляпунов* (1857—1918 гг.) и *А. А. Марков* (1856—1922 гг.), получили всеобщее признание за свои работы по обобщению теоремы Муавра — Лапласа. Пусть X_1, X_2, ... — взаимно независимые случайные величины с одним и тем же распределением (т. е. с общей функцией распределения). Предположим, что у случайных величин существуют конечные математические ожидания и стандартные отклонения.

Пусть M обозначает общее математическое ожидание и D — стандартное отклонение случайных величин X_1, X_2, ..., и пусть $S_n = X_1 + X_2 + \ldots + X_n$. Тогда

$$\lim_{n \to \infty} P\left(S_n < nM + xD\sqrt{n}\right) = \Phi(x).$$

Это — *центральная предельная теорема*, наиболее важная из всех предельных теорем (из-за своей значимости она и была названа «центральной», так ее впервые назвал Джордж Пойа). В предельных теоремах, как правило, рассматриваются асимптотические распределения различных функций (например, суммы, произведения, максимума и т. д.) от случайных аргументов. Центральная предельная теорема и ее обобщения объясняют, почему в природе нормальное распределение встречается так часто, особенно, в связи с величинами, которые составлены из многих («почти») одинаково распределенных («почти») независимых случайных компонент. Следует подчеркнуть, что в природе такие «композиции» из случайных величин не всегда образованы их суммой, поэтому изучение поведения других функций очень важно. *Пуанкаре* как-то заметил с сарказмом, что все верят в универсальность нормального распределения: физики верят потому, что думают, что математики доказали его логическую необходимость, а математики верят, так как считают, что физики проверили это лабораторными экспериментами.

д) Литература

Гнеденко Б. В., Колмогоров А. Н. Предельные распределения для сумм независимых случайных величин. — М.-Л.; Гос. изд. техн.-теоретич. литературы, 1949.

Ostrowski E. "On the remainder term of the de Moivre — Laplace formula (To the 70th birthday of Eugene Lukacs)", *Aequa. Math.*, **20**, 283—277, (1980).

Петров В. В. Суммы независимых случайных величин. — М.: Наука, 1972.

Székely G. J. "A limit theorem for elementary symmetric polynomials of independent random variables", *Z. Wahrscheinlichkeistheorie verw. Geb.*, **59**, 356—359, (1982).

11. Парадокс Бертрана

а) История парадокса

Жорж Бюффон (1707—1788 гг.), знаменитый французский ученый, в работе, написанной в 1733 г. (но опубликованной в 1777 г.), положил начало новому направлению в теории вероятностей. Решение знаменитой «задачи об игле», обсуждаемое в этой статье, потребовало использования скорее геометрического (а не комбинаторного) метода. В задачах такого типа предполагается, что случайные точки равномерно распределены в некоторой области. (Например, попадания пуль в мишень.) Вероятность попадания в произвольную часть данной области пропорциональна ее площади (длине или объему). Таким образом, для вычисления вероятности нам лишь нужно найти отношение «благоприятной» площади ко «всей» площади (длине или объему). Вероятности такого типа приводят к ряду парадоксов. Например, шанс поразить центр (или любую другую заданную точку) мишени равен, очевидно, 0. С другой стороны, попасть в эту точку можно и, следовательно, мы должны различать события, происходящие с вероятностью 0, и невозможные события (вероятность невозможного события равна 0, но обратное неверно). Довольно странно выглядит также следующий факт: события «поразить по крайней мере одну точку из конечного множества точек» и «поразить лишь одну точку» имеют одинаковые вероятности. (Обе вероятности равны 0. См. парадокс о нулевой вероятности.) Другая странность: взаимно однозначное преобразование может совершенно изменить шансы. Например, если мы случайным образом выбираем точку на интервале (0, 1), то шансы выбрать число, которое меньше 1/2, равны 50 %. Однако если все числа из (0, 1) возвести в квадрат и равномерно выбирать из этих квадратов, то шансы увеличатся до 65,6 %. Конечно, первый ответ, т. е. 50 %, более естественнен. Однако в других задачах выбор между естественным и неестественным ответами может оказаться более сложным. Выше уже отмечалось (в последнем замечании к первому парадоксу), что такой выбор не всегда возможен на основе лишь логических рассуждений без учета данных практики. Именно в этом суть следующего парадокса, опубликованного в книге «Исчисление вероятностей» ("Calcul des probabilites") (1889 г.) *Жозефа Луи Бертрана.*

б) Парадокс

Для некоторой окружности случайным образом выбирается хорда. Найти вероятность того, что эта хорда длиннее стороны правильного треугольника, вписанного в данную окружность.

Парадокс утверждает, что эта вероятность определяется неоднозначно, т. е. различные методы приводят к разным результатам.

Первый метод:

Случайным образом (равномерно) в данном круге выбирается точка. Эта случайная точка определяет единственную хорду, серединой которой она является. Эта хорда длиннее стороны нашего правильного треугольника тогда и только тогда, когда ее середина лежит внутри круга, вписанного в треугольник. Радиус этого круга равен половине радиуса исходного круга, следовательно площадь вписанного круга составляет 1/4 площади исходного. Таким образом, вероятность того, что случайно выбранная точка лежит внутри вписанного круга, равна 1/4. Так что этот метод дает ответ 1/4.

Второй метод:

Исходя из соображений симметрии, можем считать, что одним концом хорды является произвольная фиксированная точка на окружности. Пусть этой точкой является вершина вписанного

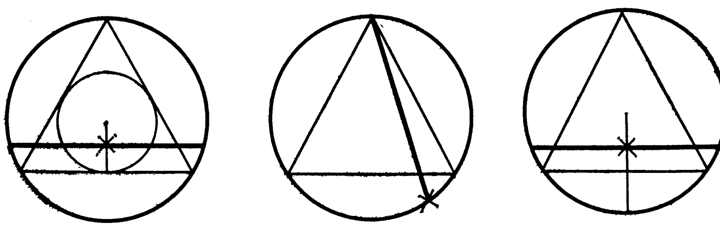

Рис. 5. Три способа выбора случайной хорды.

треугольника. Выберем другой конец случайно с равномерным распределением. Вершины треугольника делят окружность на три равные дуги, и случайная хорда длиннее стороны правильного треугольника, если она пересекает этот треугольник. Так что искомая вероятность теперь равна 1/3.

Третий метод:

Выберем точку случайным образом равномерно на радиусе окружности и возьмем хорду, которая перпендикулярна этому радиусу и проходит через выбранную точку. Тогда случайная хорда длиннее стороны вписанного правильного треугольника, если случайная точка лежит на той половине радиуса, которая ближе к центру. Исходя из соображений симметрии, неважно какой радиус был выбран для построения, поэтому искомая вероятность равна 1/2.

в) Объяснение парадокса

Получение разных результатов кажется парадоксальным, так как было убеждение, что слова «равномерный случайный выбор» однозначно определяют искомую вероятность. Парадокс показывает, что возможны различные способы выбора равномерным образом, причем каждый способ выглядит по-своему «естественным». Каждый из трех указанных выше методов использует равномерное распределение (в круге, на окружности и радиусе круга). По мнению Пуанкаре («Исчисление вероятностей» («Calcul des probabilites»). Париж, 1912 г.), если у нас нет никакой дополнительной информации, то следует воспользоваться третьим методом (где ответ был 1/2), так как в этом случае имеем: если два множества хорд геометрически конгруэнтны, то с равными вероятностями случайно выбранная хорда будет принадлежать любому из этих множеств. Изучение инвариантности такого типа привело к очень интересному разделу математики, который называется интегральной геометрией. (Этот термин был введен *Вильгельмом Бляшке* в 1934 г.)

Следующие требования инвариантности также приводят к вероятности 1/2 (см. Janes E. T. «The Well-posed problem», *Foundations of Physics*, 3, 477—493, (1973)). Пусть R — радиус круга. Положение хорды определяется заданием полярных координат (r, θ) ее центра. Найдем ответ на более сложный, чем у Бертрана, вопрос: какова должна быть плотность вероятности $f(r, \theta)dA = f(r, \theta)\ r dr\theta$ внутри круга? Поскольку распределение длины хорды зависит только от радиальной компоненты, $f(r, \theta) = f(r)$. Итак, проблема сведена к нахождению функции $f(r)$, нормированной согласно равенству

$$\int_0^{2\pi}\int_0^R f(r)\, r\, dr\, d\theta = 1, \quad \text{т. е.} \quad 2\pi\int_0^R f(r)\, r\, dr = 1.$$

Инвариантность масштаба (т. е. инвариантность при изменении масштаба) приводит к уравнению

$$a^2 f(ar) = 2\pi f(r)\int_0^{aR} f(u)\, u\, du, \quad 0 < a \leqslant 1, \ 0 \leqslant r \leqslant R.$$

Дифференцируя по a, полагая $a = 1$, и решая полученное дифференциальное уравнение, находим, что общее решение (удовлетворяющее перечисленным выше условиям нормировки) записывается в виде

$$f(r) = qr^{q-2}/(2\pi R^q),$$

где q — произвольная постоянная, которая не определяется точно условием инвариантности масштаба. Наконец, если мы сдвинем круг на расстояние b, то преобразование $(r, \theta) \Rightarrow (r', \theta')$ задается соотношениями $r' = |r - b\cos\theta|$ и

$$\theta' = \begin{cases} \theta, & \text{если } r > b\cos\theta, \\ \theta + \pi, & \text{если } r < b\cos\theta. \end{cases}$$

Инвариантность относительно сдвигов дает $q = 1$.

Таким образом, в соответствии с третьим методом мы получаем

$$f(r, \theta) = (2\pi R r)^{-1}, \quad 0 < r \leqslant R, \quad 0 \leqslant \theta \leqslant 2\pi.$$

Поскольку хорда, координаты середины которой (r, θ), имеет длину $L = 2(R^2 - r^2)^{1/2}$, функция плотности вероятности величины $X = L/(2R)$ равна $x/(1 - x^2)^{1/2}$, $0 \leqslant x < 1$, что согласуется с догадкой Бореля («Élements de la théorie des probabilités», Париж, 1909 г.).

г) Замечания

(i) При обсуждении парадокса Бертрана мы рассмотрели три метода выбора случайной хорды, однако существует множество других столь же естественных методов. Например, если мы случайно выберем точку в заданном круге, затем проведем через эту точку хорду в произвольном направлении (угол, определяющий направление, равномерно распределен во всей области изменений угла и не зависит от выбора точки), то искомая вероятность равна

$$1/3 + \sqrt{3}/(2\pi) = 0.609\ldots.$$

Неудивительно, что полученный результат больше 1/2, так как такой способ выбора чаще дает хорды большей длины. Еще большую вероятность (0.7449) мы получим, если случайная хорда получается соединением двух случайных точек на окружности. Другой, менее естественный, метод выбора заключается в следующем. Проведем окружность (радиуса r) с тем же центром, что и исходная окружность (радиуса R), и выберем случайную точку (равномерно распределенную) в круге радиуса r. Проведем прямую через эту точку в направлении, равномерно распределенном во всей области и не зависящем от выбранной точки. Вопрос заключается в следующем. Если прямая пересекает круг радиуса R, то чему равна вероятность того, что хорда, образованная при пересечении прямой с кругом радиуса R, длиннее уже не раз упоминавшейся стороны правильного вписанного треугольника? Ответ получается просто. Если r возрастает от $r = R/2$ до $r = \infty$, то искомая вероятность убывает от 1 до $1/2$.

(ii) Интегральная геометрия, возникшая в связи с геометрическими вероятностями, приобретает все возрастающее значение во многих областях науки, например, для восстановления трехмерных фигур по их двумерным сечениям или проекциям, а также в минералогии, металлургии и биологии (особенно в томографии для трехмерной реконструкции опухолей).

д) Литература

Ainley E. S. "A probable paradox", *Math. Gaz.*, **66**, 300—301, (1982).

Kendall M. G., Moran P. A. P. *Geometric Probability*, Griffin, London, 1963. [Имеется перевод: Кендалл М., Моран П. Геометрические вероятности. — М.: Наука, 1972.]

Matheron G. *Random Sets and Integral Geometry*, Wiley, New York, 1975. [Имеется перевод: Матерон Ж. Случайные множества и интегральная геометрия. — М.: Мир, 1978.]

Santaló L. A. *Integral Geometry and Geometric Probability*, Addison-Wesley, Reading, 1976. [Имеется перевод: Сантало Л. Интегральная геометрия и геометрические вероятности. — М.: Наука, 1983.]

de Temple D. W., Robertson J. M. "Constructing Buffon curves from their distributions", *The American Math. Monthly*, **87**, 779—784, (1980).

Zalcman L. "Offbeat integral geometry", *The American Math. Monthly*, **87**, 161—174, (1980).

Schuster E. F. "Buffon's needle experiment", *The American Math. Monthly*, **81**, 26—29, (1974).

12. Парадокс из теории игр. Парадокс гладиатора

а) История парадокса

Хотя азартные игры в различных формах существуют со времен палеолита и математические исследования разных игр восходят к эпохе Возрождения, общая теория игр возникла лишь в XX веке (и лишь тогда была установлена ее связь с другими науками, например, такими, как экономика). В 1921 г. *Эмиль Борель* попытался создать математическую теорию игровых стратегий, однако *принцип минимакса*, фундаментальную теорему в теории игр, в 1928 г. доказал основоположник теории игр *Джон фон Нейман.* (Ранее даже Борель сомневался в ее справедливости.)

Следующий парадокс поможет понять суть теоремы о минимаксе.

б) Парадокс

Двое детей R и Q играют в известную игру, которая состоит в следующем. Оба одновременно поднимают один или два пальца, если общее число поднятых пальцев четно, то Q платит R, и если оно нечетно, то R платит Q сумму, равную общему числу поднятых пальцев. Ниже в таблице (матрице выплат) указаны денежные суммы, которые Q должен заплатить R. Хотя многие считают эту игру справедливой (видимо потому, что числа в

R \ Q	1 палец	2 пальца
1 палец	\$ 2	\$-3
2 пальца	\$-3	\$ 4

Рис. 6.

таблице при сложении дают 0, действительно, $2+(-3)+$ $+(-3)+4=0$), она такой вовсе не является: эта игра выгодна для Q.

в) Объяснение парадокса

Очевидно, если один из игроков все время поднимает один палец или всегда поднимает два, то второй игрок, заметив это, будет вести себя так, чтобы все время выигрывать. Следовательно, выгодными могут быть только «смешанные стратегии», т. е. при каждом испытании игрок случайно, но с фиксированными вероятностями, выбирает одну из двух возможностей (поднять один или два пальца). Предположим, что для обоих игроков мы уже нашли оптимальные стратегии, т. е. мы знаем, что лучшей стратегией для игрока R является то, что нужно поднять один палец с вероятностью p_1 и поднять два пальца с вероятностью p_2 (очевидно, $p_1+p_2=1$), и аналогично для Q наиболее выгодно поднять один палец с вероятностью q_1 и два пальца с вероятностью $q_2 (q_1+q_2=1)$. Поскольку оба игрока принимают решения независимо друг от друга, сумма денег, которую Q в среднем выплатит R (если оба игрока применяют оптимальные стратегии), равна

$$V=2p_1q_1-3p_1q_2-3p_2q_1+4p_2q_2. \qquad (*)$$

Игра была бы справедливой, если бы $V=0$. Однако мы покажем, что $p_1=q_1=7/12$, $p_2=q_2=5/12$, и тогда $V=-1/12$, что означает, что Q выигрывает в среднем 1/12 доллара после каждой игры даже в том случае, когда R применяет свою оптимальную стратегию.

Подставим $q_1=1$, $q_2=0$ в $(*)$. Тогда $V=Q_1=2p_1-3p_2$. Аналогично, если $q_2=1$ и $q_1=0$, то $V=Q_2=-3p_1+4p_2$. В этих обозначениях имеем $V=q_1Q_1+q_2Q_2$. Так как V — это средний проигрыш игрока Q, если он использует свою оптимальную стратегию, $Q_1 \geqslant V$ и $Q_2 \geqslant V$, то $V=q_1Q_1+q_2Q_2 \geqslant q_1V+q_2V=(q_1+q_2)V=V$.

Поскольку ни q_1, ни q_2 не могут быть равны нулю, из предыдущего соотношения вытекает, что $V=Q_1=Q_2$, т. е. $2p_1-3p_2=-3p_1+4p_2$, поэтому (вспоминая, что $p_1+p_2=1$) имеем $p_1=7/12$, $p_2=5/12$ и $V=-1/12$. Аналогично, $2q_1-3q_2=-3q_1+4q_2 (q_1+q_2=1)$ и, следовательно, $q_1=7/12$, $q_2=5/12$. Таким образом, доказано, что игра справедливой не является, и найдены оптимальные стратегии. Для обоих игроков выгодно поднимать один палец с вероятностью 7/12.

Подставляя в формуле $(*)$ $1-p_1$ вместо p_2 и $1-q_1$ вместо q_2, получаем $V=12p_1q_1-7p_1-7q_1+4$. Независимо от значения q_1 при $p_1=7/12$ имеем $V=-1/12$. Аналогично, независимо от значения p_1 при $q_1=7/12$ имеем $V=-1/12$. Таким образом, игроку все равно, как играть, если он знает, что его противник использует оптимальную стратегию.

г) Замечания

(i) Главной целью исследований Неймана в теории игр был поиск оптимальной стратегии в игре с m игроками. Для простоты мы предполагаем, что $m=2$ (т. е. друг против друга выступают только два игрока), и игра является игрой с нулевой суммой (т. е. проигрыш первого игрока равен выигрышу второго). Пусть S_1 и S_2 обозначают множества чистых стратегий для первого и второго игроков соответственно (чистая стратегия — это правило, которое задает первый шаг первого игрока, а также ответы на всевозможные ходы второго игрока). Пусть $L(s_1, s_2)$ — функция двух переменных, которая определяет проигрыш второго игрока, когда он использует чистую стратегию $s_2 \in S_2$, и первый игрок применяет стратегию $s_1 \in S_2$ (таблица на стр. 55 дает пример такой функции). Для осторожного игрока лучшей является стратегия, которая минимизирует его максимальный проигрыш (что происходит при оптимальной защите). Первый игрок всегда может добиться выигрыша

$$V_1=\max_{s_1}(\min_{s_2} L(s_1, s_2)),$$

а второй игрок проигрыша

$$V_2=\min_{s_2}(\max_{s_1} L(s_1, s_2)).$$

(Естественно, величины V_1 и V_2 могут быть отрицательными.) В том случае, когда $V_1 = V_2$ и множество возможных стратегий конечно, обоим игрокам выгодно применять такие стратегии s_1^*, s_2^*, для которых $V_1 = V_2 = L(s_1^*, s_2^*)$. Такая пара стратегий (s_1^*, s_2^*) является седловой точкой игры, но она не всегда существует. Однако у Неймана была замечательная идея расширить множество возможных стратегий, он ввел «смешанные стратегии», которые соответствуют случайному выбору из множества чистых стратегий. Таким образом, смешанная стратегия есть вероятностное распределение на множестве чистых стратегий. (В примере с игрой детей смешанными стратегиями были пары p_1, p_2 и q_1, q_2 соответственно.) При смешанных стратегиях игрок не может «видеть противника насквозь», но зато появляется случайность, даже в тех играх, которые сами от случая не зависят. Если мы хотим найти оптимальную смешанную стратегию, то мы, естественно, должны определить функцию потерь на множестве пар (π_1, π_2) смешанных стратегий. Пусть $L(\pi_1, \pi_2)$ — потери в среднем, которые второй игрок выплачивает первому, когда игроки применяют смешанные стратегии $\pi_1 \in P_1$ и $\pi_2 \in P_2$. Теорема Неймана о минимаксе (фундаментальная теорема теории игр) утверждает, что если множества S_1 и S_2 конечны, то

$$\max_{\pi_1 \in P_1} \min_{\pi_2 \in P_2} L(\pi_1, \pi_2) = \min_{\pi_2 \in P_2} \max_{\pi_1 \in P_1} L(\pi_1, \pi_2),$$

т. е. в классе смешанных стратегий всегда существует седловая точка. Иными словами, для обоих игроков существуют оптимальные смешанные стратегии.

Общую модель теории игр можно также использовать для исследования конфликтов, возникающих в других сферах жизни. Например, с математической точки зрения коммерческую конкуренцию можно рассматривать как «игру», в которой оба игрока хотят найти свои оптимальные стратегии. Поскольку все менее и менее вероятно, что конкуренты могут постоянно надувать друг друга, компромиссы (соответствующие седловым точкам) становятся все более и более необходимыми во многих областях. Теория игр внесла новые аспекты и в математическую статистику, благодаря в основном трудам *Абрахама Вальда.* Ниже иллюстрируются некоторые применения теории игр в статистике.

(ii) Типичной задачей статистики является оценка неизвестного параметра $\theta \in \Theta$ вероятностного распределения F_θ, исходя из наблюдений X_1, X_2, ..., X_n, имеющих распределение F_θ (и обычно независимых), т. е. исходя из выборки. (Обычно, Θ — произвольное множество чисел или векторов.) Рассмотрим функцию двух переменных $L(\theta, c)$, значениями которой являются наши потери, когда в качестве оценки для неизвестного параметра θ берется c. Естественно предполагать, что потери тем больше, чем больше отклонение $|\theta - c|$. Таким образом, $L(\theta, c)$, как правило, является монотонно возрастающей функцией от величины $|\theta - c|$, например, $L(\theta, c) = |\theta - c|^d$, где $d > 0$. Оценка $\hat{\theta} = f(X_1, X_2, \ldots, X_n)$ хороша, если средние потери малы, т. е. если функция риска $R(\theta, \hat{\theta}) = E(L(\theta, \hat{\theta}))$ мала. Однако при сравнении двух оценок может оказаться, что значение функции риска для первой оценки при некоторых значениях параметра θ меньше, чем для второй оценки, а при других значениях θ ситуация противоположная. Для достаточно большого класса оценок существуют функции риска, которые убывают при одних значениях θ только тогда, когда при других значениях θ они возрастают. Оценки такого типа называются *допустимыми* оценками, т. е. оценка $\hat{\theta}_0$ является допустимой, если неравенство $R(\theta, \hat{\theta}) \leqslant R(\theta, \hat{\theta}_0)$ справедливо для всех $\theta \in \Theta$ тогда и только тогда, когда $R(\theta, \hat{\theta}) = R(\theta, \hat{\theta}_0)$ для всех $\theta \in \Theta$. Имеет смысл использовать только допустимые оценки, так как для оценки, которая не является допустимой, мы всегда можем найти другую оценку, функция риска которой нигде не больше, а в некоторых точках строго меньше функции риска недопустимой оценки. Если мы хотим найти допустимую оценку, которая минимизирует средние потери при «наихудшем» действительном значении параметра (т. е. когда функция риска достигает максимума), то мы получим минимаксную оценку. Эта «осторожная» оценка определяется следующим образом: оценка θ^* называется минимаксной, если

$$\sup_{\theta \in \Theta} R(\theta, \theta^*) = \inf_{\hat{\theta}} \sup_{\theta \in \Theta} R(\theta, \hat{\theta}),$$

где $\hat{\theta}$ пробегает множество всех возможных оценок. «Минимаксный аспект» математической статистики рассматривает поиск оценки как игру, в которой Природа «выбирает» параметр θ, а мы выбираем оценку $\hat{\theta}$. Цель игры — добиться наименьших возможных средних потерь. Средние потери можно уменьшить, допустив смешанные стратегии, когда Природа выбирает параметр θ случайно из множества Θ с распределением τ, и мы выбираем оценку случайно с распределением α из множества всевозможных оценок. В этом случае функцией риска является функция $r(\tau, \alpha) = E(R(T, A))$, где T — случайная величина с распределением τ на множестве параметров Θ, и A имеет распределение α на множестве всевозможных стратегий. Теорема о минимаксе остается справедливой для функций риска такого вида при достаточно общих условиях:

$$\sup_{\tau} \inf_{\alpha} r(\tau, \alpha) = \inf_{\alpha} \sup_{\tau} r(\tau, \alpha).$$

Поскольку распределение τ неизвестно, полезно выбирать минимаксную стратегию α^* как оценку, удовлетворяющую уравнению

$$\sup_{\tau} r(\tau, \alpha^*) = \inf_{\alpha} \sup_{\tau} r(\tau, \alpha).$$

(iii) Следующий парадокс гладиатора исходит от *К. Каминского, Е. Лакса* и *П. Нелсона.* Предположим, что в некотором соревновании, называемом игрой гладиаторов, участвуют две команды гладиаторов, которые на арене устраивают бой. В последовательных турах из команды $A = (A_1, A_2, \ldots, A_n)$ выбирается гладиатор для встречи с гладиатором из команды $B = (B_1, B_2, \ldots, B_n)$. Победитель возвращается в свою команду с неослабевшим желанием сразиться вновь, если потребуется. Предполагается, что проигравший теряет возможность сражаться и выбывает из турнира. Предполагается также, что индивидуальные встречи носят стохастический характер и представляют собой независимые испытания, в которых $0 < P(A_i, B_j) < 1$ обозначает вероятность того, что гладиатор A_i нанесет поражение гладиатору B_j. Турнир продолжается до тех пор, пока не уничтожается одна из команд. Рассмотрим вопрос о существовании стратегии S, которая оптимальна в том смысле, что она максимизирует $P_S(A, B)$, вероятность выигрыша команды A у команды B, когда используется стратегия S. Под стратегией здесь понимается правило, определяющее порядок выхода на арену гладиаторов обеих команд. (При указании стратегии можно исходить только из текущего состояния команд на каждом этапе игры.) В одном из вариантов игры гладиаторов предполагается, что игрокам $A_1, A_2, \ldots, A_n, B_1, B_2, \ldots, B_n$ ставятся в соответствие положительные числа $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n$ таким образом, что во всех встречах A_i против B_j имеем $P(A_i, B_j) = a_i/(a_i + b_j)$. Тогда вероятность $P_S(A, B)$ одинакова для всех S! В этом суть парадокса гладиаторов. Другой парадокс этой же игры состоит в следующем. Будем говорить, что команда A превосходит команду B, если $P(A, B) > 1/2$. Далее, если A превосходит B и B превосходит C, то A не обязательно превосходит C. Существуют примеры, показывающие, что $m = \min\{P(A, B), P(B, C), P(C, A)\} > 1/2$. Хотя открытым остается интригующий вопрос о точном верхнем пределе для m. (Аналогичную ситуацию см. в парадоксе 1/13е).

(iv) Еще одним парадоксом из теории игры является знаменитая «Дилемма заключенного». Здесь мы лишь сошлемся на работу Бремса, Штрафина и Хофштадтера.

д) Литература

Brams S. J., Straffin Jr. S. J. "Prisoners dilemma and professional sport drafts", *The Amer. Math. Monthly*, 86, 80—88, (1979).

Hofstadter D. R. "Computer tournaments of the Prisoner's dilemma suggest how cooperation evolves", *Sci. Amer.*, 16—26, (May 1983).
Joó I. "A simple proof for Neumann's minimax theorem", *Acta Sci. Math.* (Szeged), **42**, 91—94, (1980) и *Acta Math. Hung.*, **44**, 363—365, (1984).
Neumann J., Morgenstern O. *Theory of Games and Economic Behavior*, Princeton Univ. Press, Princeton, 1944. [Имеется перевод. Нейман Дж., Моргенштерн О. Теория игр и экономическое поведение. — М.: Наука, 1970.]
Savage L. J. *The Foundation of Statistics*, Wiley, New York, 1954.
Wald A. *Statistical Decision Functions*, Wiley, New York, 1954.
Williams J. D. *The Complete Strategyst Being a Primer on the Theory of Games of Strategy*, McCraw-Hill, Book Company, New York. [Имеется перевод: Вильямс Дж. Д. Совершенный стратег, или Букварь по теории стратегических игр. — М.: Советское радио, 1960.]

13. Еще несколько парадоксов

а) Парадокс событий, происходящих почти наверно

Рассмотрим события, происходящие с вероятностями 0.99 и 0.9999 соответственно. Можно сказать, что обе вероятности практически одинаковы, оба события происходят почти наверно. Тем не менее в некоторых случаях разница становится заметной. Рассмотрим, например, независимые события, которые могут происходить в любой день года с вероятностью $p = 0.99$; тогда вероятность того, что они будут происходить каждый день в течение года, меньше, чем $P = 0.03$, в то же время, если $p = 0.9999$, то $P = 0.97$.

б) Парадокс вероятности и относительной частоты

Следующая история, принадлежащая *Джорджу Пойа*, показывает, как не следует интерпретировать частотную концепцию вероятности. Д-р Тел (доктор телепатии), закончив осмотр пациента, покачал головой. «Вы очень серьезно больны», — сказал он, — «из десяти человек с такой болезнью выживает только один». Пациента эта информация изрядно испугала, и д-р Тел начал его успокаивать: «Но Вам очень повезло, что Вы пришли ко мне, сэр. У меня уже умерли от этой болезни девять пациентов, так что Вы выживете».

(Лит.: Polya G. Patterns of Plausible Inference, vol. II, Princeton Univ. Press, 1954, p. 101. [Имеется перевод: Пойа Д. Математика и правдоподобные рассуждения. — М.: Наука, 1975.])

в) Парадоксы, связанные с бросанием монеты

(i) Предположим, что мы бросаем правильную монету до тех пор, пока не выпадут последовательно два герба (ГГ) или герб и решка (ГР). Очевидно, вероятность того, что (ГГ) появится быстрее, чем (ГР), равна вероятности того, что (ГР) появится быстрее, так как после выпадения Г монета выпадет на Г или Р с равной вероятностью. Несмотря на это, для появления (ГГ) потребуется в среднем больше бросаний, чем для появления (ГР). Серия (ГГ) появится в среднем через 6 бросаний, а серия (ГР) через 4. [Пусть M обозначает математическое ожидание числа бросаний до появления (ГГ) в предположении, что при первом бросании выпал Г, и $M_Р$ обозначает математическое ожидание числа бросаний до появления (ГГ) при условии, что при первом бросании выпала Р. Тогда $M_Г = 1 + (1 + M_Р)/2$ и $M_Р = 1 + (M_Г + M_Р)/2$. Отсюда следует, что $(M_Г + M_Р)/2 = 6$, т. е. среднее число бросаний до появления (ГГ) действительно равно 6.] Если сравнить серии (ГРГР) и (РГРР), то различие будет больше. Вероятность того, что (ГРГР) появится быстрее, чем (РГРР), равна $9/14 > 1/2$, но среднее число бросаний, которое потребуется для получения серии (ГРГР) все же больше, чем среднее число бросаний до появления серии (РГРР). (Первое равно 20, а последнее лишь 18.) Следовательно, даже если вероятность того, что событие A появится быстрее, чем событие B, больше вероятности того, что B появится быстрее; может так случиться, что до появления A придется ждать дольше, чем до появления B.

Можно доказать, что среди всех серий Г — Р наибольшее среднее время ожидания имеют чистые серии, т. е. те, которые состоят только из Г или только из Р. В этом случае среднее число бросаний равно $2^{n+1} - 2$. Наименьшее возможное (среднее) число бросаний равно 2^n, что соответствует серии, состоящей из $n - 1$ герба и заканчивающейся одной решкой или из $n - 1$ решки с одним гербом в конце. (Таким образом, до появления серии, состоящей из одних гербов, нужно ждать почти вдвое больше времени, чем до появления серии из $n - 1$ Г и одной Р, хотя вероятность того, что первая появится быстрее, равна вероятности того, что раньше появится последняя.)

Нахождение времени, которое нужно ждать до появления заданной серии Г — Р длины n, при больших n обычно требует утомительных вычислений (решения многомерной системы линейных уравнений). Вычисления можно в значительной степени упростить, пользуясь «магическим» алгоритмом Конвея, который обсуждается в статье Ли, см. ссылку ниже. Приведем этот алгоритм для случая более общего, чем серии Г — Р при бросании правильной монеты. Пусть $X, X_1, X_2, \ldots$ — независимые одинаково распределенные случайные величины, принимающие конечное или счетное число значений с положительными вероятностями. Обозначим множество этих значений через V, и пусть $A = (a_1, a_2, \ldots, a_m)$ и $B = (b_1, b_2, \ldots, b_n)$ — две (конечные) последовательности элементов из V. Введем следующие обозначения:

$$d_{ij} = \begin{cases} 1/P(X=b_j), & \text{если} \quad 1 \leqslant i \leqslant m,\ 1 \leqslant j \leqslant n,\ a_i = b_j, \\ 0 & \text{в противном случае} \end{cases}$$

и $A \cdot B = d_{11}d_{22} \ldots d_{mm} + d_{21}d_{32} \ldots d_{mm-1} + \ldots + d_{m1}$. Пусть T_A и T_B обозначают число X-величин до первого появления последовательностей A и B в $X_1, X_2, \ldots$ соответственно. Тогда математическое ожидание величины T_A равно $A \cdot A$, а вероятность того, что T_A меньше, чем T_B (в предположении, что ни A, ни B не содержат в себе другой последовательности) вычисляется по формуле

$$(B \cdot B - A \cdot A)/(A \cdot A + B \cdot B - A \cdot B - B \cdot A).$$

Например, для серии Г — Р при бросании правильной монеты имеем $d_{ij} = 0$ или $d_{ij} = 2$. Если $A =$ (ГРГР) и $B =$ (РГРР), то $A \cdot A = 20$, $B \cdot B = 18$, $A \cdot B = 10$ и $B \cdot A = 0$, так что вероятность того, что A произойдет быстрее B, равна $18/28 = 9/14$, что и утверждалось выше.

Сформулируем теперь более изощренную теорему (до сих пор не опубликованную). Предположим, что мы ждем появления всех возможных серий Г — Р длины n, которые могут появиться при бросании правильной монеты (таких серий всего 2^n), и пусть τ_n означает это (случайное) время ожидания. Тогда

$$\lim_{n \to \infty} P(\tau_n/2^n - \log 2^n < x) = \exp(-e^{-x}).$$

(ii) Если при бросании правильной монеты мы хотим получить заданную последовательность длины 3 (например, ГГР), мы должны бросать монету в среднем не менее 8 раз. Число необходимых бросаний будет наименьшим для получения любой следующей последовательности: (ГГР), (РГГ), (РРГ), (ГРР). (В каждом таком случае среднее число необходимых бросаний равно 8, а во всех других случаях оно больше 8.) Сравним эти серии следующим образом:

α) вероятность того, что серия (ГГР) появится быстрее, чем (РГГ), равна 1/4.

β) (РГГ) появится раньше, чем (РРГ), с вероятностью 1/3,

γ) (РРГ) произойдет быстрее, чем (ГРР), с вероятностью 1/4, и наконец,

δ) вероятность того, что (ГРР) наступит раньше, чем (ГГР) равна 1/3.

Таким образом, начав с серии (ГГР), мы опять вернулись к ней, хотя на каждом шаге вероятности сравнения были строго меньше 1/2.

(Лит.: Li Shou-Yen R. "A martingale approach to the study of occurrence of sequence patterns in repeated experiments". *Annals of Probability*, 8, 1171—1176, (1980).)

г) Парадокс условной вероятности

Существуют события A, B и C такие, что

α) условная вероятность события A при условии, что событие B произошло, меньше, чем условная вероятность события A при условии, что B не произошло;

β) условная вероятность A при условии, что произошли оба события B и C, больше условной вероятности A при условии, что C наступило, а B нет;

γ) условная вероятность A относительно B и дополнения к C больше условной вероятности A при условии, что ни B, ни C не произошли.

Символически эти три утверждения можно записать следующим образом:

$$P(A\mid B) < P(A\mid \bar{B}), \quad P(A\mid BC) > P(A\mid \bar{B}C)$$

и

$$P(A\mid B\bar{C}) > P(A\mid \overline{B}\overline{C}).$$

Это представляется парадоксальным, потому что можно подумать, что $P(A|B)$ есть среднее величин $P(A|BC)$ и $P(A|B\bar{C})$ и, аналогично, $P(A|\bar{B})$ — среднее величин $P(A|\bar{B}C)$ и $P(A|\bar{B}\bar{C})$, а среднее двух меньших величин должно быть меньше среднего двух бóльших величин. Объяснение этого ошибочного заключения состоит в том, что $P(A|B)$ и $P(A|\bar{B})$ являются взвешенными средними указанных выше вероятностей, причем соответствующие веса в обоих случаях различны:

$$P(A\mid B) = P(C\mid B)\,P(A\mid BC) + P(\bar{C}\mid B)\,P(A\mid B\bar{C}),$$

в то время, как

$$P(A\mid \bar{B}) = P(C\mid \bar{B})\,P(A\mid \bar{B}C) + P(\bar{C}\mid \bar{B})\,P(A\mid \bar{B}\bar{C}).$$

Тем не менее, если события B и C независимы, то $P(C|B) = P(C|\bar{B})$ и $P(\bar{C}|B) = P(\bar{C}|\bar{B})$, так что в этом случае парадоксов не возникает.

(Лит.: Blyth C. R. "On Simpson's paradox and the sure thing principle". *J. Amer. Statist. Assoc.*, 67, 364—366, (1972).)

д) Парадокс случайных времен ожидания

Предположим, что два случайных события происходят через (случайное) время X и Y. Парадоксально, но может так случиться, что $X > Y$ с вероятностью по крайней мере 0.99, однако X стохастически меньше, чем Y[1]), т. е. вероятность того, что $X < t$, больше вероятности события $Y < t$ при любом фиксированном времени t (или, другими словами, функция распределения случайной величины X всюду больше, чем функция распределения Y). Например, это так, если Y равномерно распределена на интервале $[0, 1]$, $X = Y + (1 - Y)/1000$ с вероятностью 0.99 и $X = Y/1000$ с вероятностью 0.01. [Такая парадоксальная ситуация невозможна, если X и Y независимы: пусть F и G обозначают их функции распределения; для простоты предположим, что функция G непрерывна и существует обратная к ней функция G^{-1}. Тогда функция распределения случайной величины $Z = G^{-1}(F(X))$ совпадает с G. Поскольку $F > G$, имеем $Z > X$. Следовательно, $P(X > Y) \leqslant P(Z > Y) = 1/2$, так как Z и Y независимые одинаково распределенные случайные величины, т. е. $P(X > Y)$ должно быть намного меньше, чем 0.99. На самом деле эта вероятность не превосходит 0.5.]

Следующий парадокс аналогичен предыдущему. Пусть X и Y две независимые случайные величины, причем X стохастически меньше Y. Тогда можно подумать, что $\max(X, X + Y)$ также стохастически меньше $\max(Y, X + Y)$, но это неверно. Например, в случае, когда X и Y принимают только значения -1, 0, 1 с вероятностями 1/4, 1/2, 1/4 и 1/4, 1/4, 1/2 соответственно.

(Лит.: Blyth C. R. "Some probability paradoxes in choice from among random alternatives" (с комментариями Д. Линдли, И. Гуда, Р. Уинклера и Дж. Пратта), *J. Amer. Stat. Assoc.*, 67, 366—388, (1972). См. также SIAM Rev., April 1970.)

e) *Парадокс транзитивности*

Два игрока A и B играют в следующую игру. На первом шаге игрок A расставляет числа по своему вкусу на 3-х игральных костях, записывая по одному числу из набора 1, 2, 3, ..., 18 на каждой грани костей (каждое число должно быть использовано только один раз).

На втором шаге B, внимательно изучив эти 3 кости (занумерованные игроком A), выбирает одну из них.

На третьем шаге A выбирает одну из оставшихся 2-х игральных костей. На последнем шаге и A, и B бросают свои игральные кости и побеждает тот, у кого выпало большее число.

Можно подумать, что эта игра более выгодна игроку B, поскольку независимо от того, как A занумерует 3 кости, B всегда

---

[1]) Для ограниченных случайных величин X, Y считаем, что X стохастически меньше Y, если $F(t)$, функция распределения величины X, всюду не меньше $G(t)$, функции распределения Y, и найдется значение t_0 такое, что $F(t_0) > G(t_0)$. — *Прим. перев.*

может выбрать лучшую из них (или одну из лучших). Следовательно, шанс победить у B составляет по крайней мере 50 %. Однако справедливо именно противоположное: A может так занумеровать кости, что он побеждает с вероятностью 21/36 (что больше 50 %), независимо от того, какую кость выбирает B. Это происходит из-за системы нумерации «поражения по кругу», при которой каждая кость побеждает ровно одну из других двух костей, что означает, что среди костей нет «лучшей». Пусть I, II и III обозначают три кости, и предположим, что игрок A занумеровал кости следующим образом: он записал числа

18, 10, 9, 8, 7, 5 на гранях кости I
17, 16, 15, 4, 3, 2 на гранях кости II
14, 13, 12, 11, 6, 1 на гранях кости III

Нетрудно вычислить, что на кости I появится число, большее, чем на кости II, с вероятностью 21/36. Аналогично на кости II выпадет число, большее, чем на кости III, с вероятностью 21/36. И вероятность появления бóльшего числа при бросании кости III, чем при бросании кости I, также равна 21/36. Следовательно, вероятность «поражения по кругу» равна 21/36. Итак, если A занумеровал кости таким образом, то его позиция предпочтительней по сравнению с B. (Если B выберет кость I, II или III, и A возьмет соответственно кость III, I или II, то у A больше шансов победить.) Можно также доказать, что вероятность «поражения по кругу» не может быть больше 21/36. Суть этого парадокса заключается в том, что случайные величины не всегда можно упорядочить таким образом, что одна будет больше другой с вероятностью, превосходящей 50 %, потому что нет транзитивности. В случае, когда одно и то же число можно записать на нескольких гранях, например,

1, 4, 4, 4, 4, 4 на кости I
2, 2, 2, 5, 5, 5 на кости II
3, 3, 3, 3, 3, 6 на кости III

то вероятность «поражения по кругу» также равна 21/36. Сформулируем парадокс в более общем виде. Пусть $X_1, X_2, \ldots, X_n$ — произвольные числа, зависящие от случая (т. е. случайные величины). Обозначим вероятности событий $X_1 < X_2$; $X_2 < X_3$; ... ...; $X_n < X_1$ через $p_1, p_2, \ldots, p_n$ соответственно. Тогда $\min(p_1, p_2, \ldots, p_n)$ есть вероятность «поражения по кругу». Пусть k_n обозначает эту вероятность. Чем больше k_n, тем интереснее становится парадокс. Можно легко показать, что k_n никогда не превзойдет $(n-1)/n$, и эта величина является наименьшим верхним пределом. Вычисление наименьшего верхнего предела для

k_n значительно сложнее, если предполагается, что случайные величины $X_1, X_2, \ldots, X_n$ независимы как исходы при бросании кости. Пусть f_n обозначает наименьший верхний предел в этом случае. *Усыскин* вычислил (*Annals of Statist.*, **35**, 857—862, 1964), что $f_2 = 1/2$, $f_3 = (\sqrt{5} - 1)/2$ (соотношение золотого сечения), $f_4 = 2/3$ и т. д. Последовательность (f_n) монотонно возрастает и сходится к 3/4. Можно показать, что скорость сходимости имеет порядок n^{-2}.

ж) Парадокс измерения регулярности игральной кости

При бросании правильной игральной кости одна и та же грань появится два раза подряд в среднем при 7 бросаниях и три раза подряд при 43 бросаниях (см. конец этого парадокса). Если кость неправильная (т. е. разные грани выпадают с разными вероятностями), то среднее число бросаний до появления одной и той же грани дважды или трижды будет меньше. Будем говорить, что кость I регулярней кости II, если для появления одной и той же грани дважды (трижды) кость I нужно бросить большее число раз, чем кость II. Может оказаться, что для получения одной и той же грани два раза подряд потребуется в среднем больше бросаний кости I, чем кости II, однако для выпадения одной и той же грани трижды нужно большее число раз бросать кость II.

Следующий простой пример придуман Т. Мори. (См. также статью Шомоди А. «Об одном парадоксе, связанном с бросанием костей», Сб. Труды конференции молодых ученых Будапештского и Московского университетов. Будапешт, с. 76—83, 1982.) Предположим, что грани кости I выпадают с вероятностями 0.03; 0.03; 0.19; 0.19; 0.28; 0.28 и соответствующие вероятности для кости II равны 0.04; 0.04; 0.17; 0.17; 0.29; 0.29. Тогда кость I нужно бросить в среднем 5.41 раз и кость II — 5.47 раз для выпадения одной и той же грани два раза подряд. В то же время, если мы хотим получить одну и ту же грань три раза подряд, то кость I нужно в среднем бросать 22.54 раза, а кость II — 22.35. Этот парадокс показывает, что плохо определять «регулярность» кости так, как это сделали мы. (В общем случае можно доказать, что если некоторая грань появляется с положительной вероятностью p, то среднее число необходимых бросаний для получения этой грани k раз подряд равно $m_p = p^{-1} + p^{-2} + \ldots + p^{-k}$. Рассмотрим кость, грани которой появляются с вероятностями $p_1, p_2, \ldots,$ и пусть M_k обозначает среднее число бросаний, необходимых для получения одной и той же грани k раз подряд. Тогда $M_k^{-1} = m_{p_1}^{-1} + m_{p_2}^{-1} + \ldots$ Если мы положим $p_1 = p_2 = \ldots = 1/6$, то $M_2 = 7$ и $M_3 = 43$, как мы уже отмечали.)

з) Парадокс дня рождения

Если собираются вместе не более, чем 365 человек, то возможно, что все они имеют различные дни рождения. Однако среди 366 человек наверняка (100 %) найдутся по крайней мере два таких, у которых дни рождения приходятся на один и тот же день в году. (Предположим, что мы здесь не рассматриваем високосные года.) Однако, если мы зададимся целью найти, сколько должно быть людей, чтобы с надежностью 99 % два из них имели один и тот же день рождения, то с удивлением обнаружим, что достаточно 55(!) человек. В то же время среди 68 человек с вероятностью 99.9 % по крайней мере два имеют одинаковый день рождения. Почти неправдоподобно, что такая малая разница между вероятностями 99 % и 100 % может привести к столь большим различиям в числе людей. Этот парадоксальный случай иллюстрирует одну из главных причин, почему теория вероятностей применяется так широко. (Подобный феномен был рассмотрен в замечании (ii) к I. 10).

Обозначим через n число дней в году, и пусть x ($<n$) — число людей в группе. Тогда вероятность того, что никакие два человека в этой группе не имеют одинаковых дней рождения, равна

$$n(n-1)(n-2)\ \dots\ (n-x+1)/n^x.$$

Следовательно, если

$$n(n-1)(n-2)\ \dots\ (n-x+1)/n^x = 1-p,$$

то p — вероятность того, что среди x людей найдутся имеющие один и тот же день рождения. Приближенное решение этого уравнения (при условии, что $0<p<1$) равно $x \approx \approx \sqrt{2n \ln (1-p)^{-1}}$. Следовательно, порядок величины x есть $\sqrt{n}$ для любого значения p из открытого интервала $(0, 1)$. В то же время, если $p=1$, то $x=n+1$. Обобщение проблемы дня рождения состоит в следующем. Вычислить нижнюю границу x так, что с вероятностью p в группе из x человек по крайней мере у k дни рождения приходятся на один и тот же день в году. Ответом здесь является $x \approx cn^{(k-1)/k}$, где c — постоянная, зависящая лишь от p и k (точнее $c=(k! \ln(1-p)^{-1})^{1/k}$).

и) Парадокс гербов и решек

Предположим, что играя в гербы и решки, мы подбросили правильную монету 100 раз. Тогда, к нашему удивлению, вероятность события $A=\{$выпало ровно 50 гербов$\}$ больше вероятности события $B=\{$выпало по крайней мере 60 гербов$\}$.

(Как мы отмечали в I. 10, $P(A) \approx 8$ %, а по теореме Муавра — Лапласа $P(B) = 1-\Phi(2) \approx 3$ %. Вероятность выпадения по крайней мере 55 гербов возрастает до 16 % (приблизительно), что вдвое больше $P(A)$.]

к) Ребро монеты

Обычно событие, состоящее в том, что монета встанет на ребро, не рассматривается, так как оно почти никогда не происходит. Найдем теперь размеры монеты, при которых с равными вероятностями (1/3) монета падает на ребро, герб или решку. Для простоты рассмотрим монету в виде прямого цилиндра, у которого основания соответствуют гербу и решке, а боковая поверхность — ребру монеты. Если монета вращается вокруг оси, проходящей через центр монеты и параллельной основанию, то достаточно рассматривать сечение монеты, проходящее через центр и перпендикулярное обоим основаниям. Этим сечением является прямоугольник. Проведем окружность вокруг этого прямоугольника и точку приземления выберем случайным образом на этой окружности. Естественно считать, что монета упадет на ребро с вероятностью, равной вероятности того, что радиус, соединяющий центр и случайную точку на окружности, пересечет сторону прямоугольника, соответствующую боковой поверхности монеты. В этой модели монета становится на ребро с вероятностью 1/3, если отношение ее толщины к диаметру равно $\operatorname{tg} 30^\circ \approx 0.577$. Однако проблема не сводится к задаче на плоскости, если монета может вращаться свободно, точнее, если на поверхности сферы, описанной вокруг монеты, случайно выбирается точка и мы предполагаем, что монета становится на ребро, когда радиус, соединяющий эту случайную точку с центром, пересекает боковую поверхность монеты. В этой модели выпадения герба, решки или вставание на ребро равновероятны, если отношение толщины монеты к ее диаметру равно 0.354... . Безусловно, существуют и более реалистические модели. Самой поразительной из них является та, где отношение, о котором говорилось выше, наименьшее (т. е. когда монета наиболее плоская).

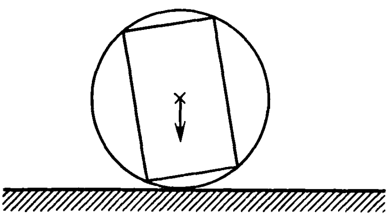

Рис. 7. Когда монета встанет на ребро?

л) Парадокс Бореля

Пусть случайная точка равномерно выбирается на поверхности шара (например, на Земле, предполагая, что она — шар).

Вообще говоря, положение точки задается ее широтой и долготой. При данной широте долгота равномерно распределена, но при фиксированной долготе распределение широты не является равномерным. (Плотность этого распределения пропорциональна косинусу долготы.) Следовательно, плотность случайной точки не одинакова, когда она находится на гринвичском меридиане или на экваторе, хотя и гринвичский меридиан, и экватор являются большими окружностями на сфере, и поэтому их роль представляется одинаковой.

Следующая проблема представляет собой аналогичный парадокс. Пусть X и Y — две независимые случайные величины с

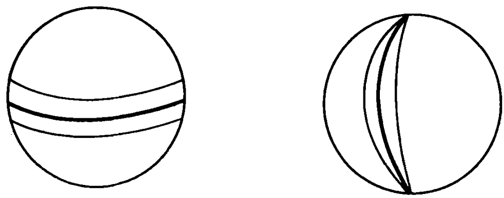

Рис. 8. Хотя и экватор, и меридиан являются самыми длинными окружностями на глобусе, при вычислении условных вероятностей необходимо учитывать, что экватор окружен сферическими зонами, а меридиан — двухугольными.

нормальным распределением. Вектор (X, Y) можно рассматривать как случайную точку на плоскости. Пусть R и φ — ее полярные координаты. Предполагая, что $X = Y$, получим, что распределение $R^2 = 2X^2$ совпадает с распределением квадрата стандартной случайной величины, умноженного на 2. В то же время, при условии, что $\varphi = \pi/4$ или $\varphi = 5\pi/4$, распределение случайной величины $R^2 = X^2 + Y^2$ то же, что и суммы квадратов двух независимых стандартных нормальных случайных величин (так как R и φ независимы). Следовательно, для R^2 мы получили совершенно различные распределения, когда $X = Y$ и когда $\varphi = \pi/4$ или $\varphi = 5\pi/4$, что кажется парадоксальным потому, что оба условия означают одно и то же, только в первом случае условие сформулировано в обычных координатах, а во втором случае — в полярных.

(Лит.: Billingsley P. *Probability and Measure*, Wiley, New York — Chichester — Brisbane — Toronto, 1979.)

м) Парадокс условных распределений

Пусть X и Y — случайные величины и $f(x, y)$ — функция двух переменных такая, что при любом фиксированном y величина $f(X, y)$ не зависит от Y. Справедливо ли, что в этом случае $f(X,$

Y) также не зависит от Y? Следующий простой пример показывает, что ответ отрицательный. Пусть $X = Y$ и X равномерно распределена на интервале $(0, 1)$. Предположим, что $f(x, y) = y$, если $x = y$ и $f(x, y) = 0$ в противном случае. Тогда $f(X, y)$ тождественно равна 0 (с вероятностью 1), следовательно она не зависит от Y, однако $f(X, Y) = Y$, очевидно, от Y зависит.

(Лит.: Perlman N. D., Wichura M. J. "A note on substitution in conditional distribution", *Annals of Statist.*, 3, 1175—1179, (1975).)

н) *Как играть в проигрышную игру*

Предположим, что в некоторой игре число испытаний (n) всегда четно. Первый игрок A выигрывает очко с вероятностью $p = 0.45$; для B эта вероятность $p = 0.55$. Чтобы выиграть игру, игрок должен набрать больше половины всех очков. Если у A есть возможность выбирать число n, то, как ни странно, $n = 2$ не является лучшим выбором. (Это будет лучшим выбором, когда p очень мало, точнее, когда p меньше $1/3$). Если $p = 0.45$ и $n = 2$, то вероятность выигрыша для A равна всего лишь $0.45^2 = 0.2025$. Если же испытаний будет больше, то A окажется в лучшей ситуации. Легко доказать, что оптимальным является выбор $n = 10$. Такой результат на первый взгляд противоречит общему «принципу»: чем раньше мы прекратим проигрышную игру, тем лучше. Предположим, например, что нам нужно 20 долларов, а у нас есть только 10. Мы собираемся получить недостающую сумму, сыграв в рулетку. Поскольку рулетка — это проигрышная игра, рекомендуется сделать наименьшее возможное число попыток, т. е. мы должны поставить сразу все наши деньги, например, на «красное». В этом случае шансы выиграть равны 18/38 (в американской рулетке есть два нуля: 0 и 00). С другой стороны, если мы каждый раз будем ставить лишь по одному доллару, то достигнем своей цели с вероятностью 0.11. Более подробную информацию можно найти в книге Dubins L. E., Savage L. J. How to Gamble if you Must, New York, McGraw-Hill, 1965.

(Лит.: Mosteller F. *Fifty Challenging. Problems in Probability with Solutions*, Reading, Addison-Wesley, 1965. [Имеется перевод: Мостеллер Ф. Пятьдесят занимательных вероятностных задач с решениями. — М.: Наука, 1985.])

о) *Парадокс страхования*

Клиент, который владеет собственностью V, хочет застраховать часть $bV (0 < b < 1)$ своей собственности от возможного неблагоприятного события, которое происходит ежегодно с вероятностью p. Ежегодный страховой взнос составляет $cV (0 <$

$< c < 1$). Страхование выгодно страховой компании лишь тогда, когда ожидаемая прибыль положительна, т. е. когда c больше pb. Почему все же клиенты страхуют имущество, если они знают, что страхование выгодно для компании, а не для них? Предположим, что клиент застраховал имущество и платил деньги в течение n лет, но страховой компании ни разу не пришлось выплачивать страховку. Тогда начальная собственность клиента (V) уменьшится до величины $V(1-c)^n$. А что было бы, если клиент не застраховался? Пусть X_k обозначает случайную величину, которая равна 1, если клиент понес убытки в k-м году, и $X_k = 0$ в противном случае. Тогда величина собственности в $(k+1)$-м году равна $V_{k+1} = V_k(1-bX_{k+1})$, следовательно, спустя n лет, имеем

$$V_n = V\prod_{k=1}^{n}(1-bX_k) = V\exp\left(\sum_{k=1}^{n}\ln(1-bX_k)\right).$$

Поскольку ожидаемое значение величины $\ln(1-bX_k)$ равно $p\ln(1-b)$, с большой вероятностью получаем

$$V_n \approx V\exp(np\ln(1-b)) = V(1-b)^{np}.$$

Таким образом, страхование выгодно для клиента, когда $V(1-b)^{np}$ меньше, чем $V(1-c)^n$, т. е. (используя разложение в степенной ряд), когда c меньше, чем

$$pb + \frac{p(1-p)}{2}b^2 + \frac{p(1-p)(2-p)}{6}b^3 + \ldots .$$

Это означает, что страхование выгодно, как для клиента, так и для компании, если c больше pb, но меньше, чем указанная выше сумма. Легко видеть, что, чем меньше b (т. е. чем меньшая часть собственности страхуется), тем меньше свободы в выборе величины c, т. е. возможность компромисса уменьшается. (В некотором смысле участие в лотерее также представляет собой вид страхования. Предположим, что некто всегда ставил на одни и те же числа, а спустя некоторое время перестал участвовать в лотерее и в этот раз «его» числа выиграли. Тогда этот некто возможно скончается от удара. С такой точки зрения цена лотерейного билета представляется недорогой. Совершенно другая ситуация в футбольных пулах, так как в них редко кто всегда ставит на одну и ту же комбинацию и поэтому неясно, что теряет такой человек, не участвуя в игре.).

п) Абсурдные результаты Льюис Кэррол

Закончим серию парадоксов абсурдными результатами и софизмами. Приведем эти результаты совместно с их ошибочными выводами, однако для того, чтобы найти ошибки потребуется поломать голову. Знаменитый писатель Льюис Кэррол был большим любителем нелепостей и в математике, и в литературе. (В работе «Абсурдная литература» («The Absurd Literature») Никола Балоте рассматривает Кэррола как главного предвестника современного абсурда.) В последние 10 лет жизни Кэррола привлекали абсурдные математические выводы (см. собрание «Curiosa Mathematica» 1888 г. или статью о Разуме, опубликованную в апреле 1895 г.). В работе Кэррола «Проблемы на подушке» («Pillow Problems», 1894 г.) можно найти следующий абсурдный результат.

В мешке находятся два шара, которые могут быть либо красными, либо белыми. Попробуем отгадать их цвет, не заглядывая в мешок. Согласно Кэрролу единственно правильный ответ заключается в том, что один из них красный, а другой белый. Он объясняет это следующим образом. Когда в мешке находятся 2 красных (R) шара и 1 белый (W), вероятность вытащить красный шар равна 2/3. С другой стороны, если в мешке было 3 шара и вероятность вынуть красный шар равнялась 2/3, то в мешке находились 2 шара R и 1W.

Теперь положим шар R в мешок, который первоначально содержал только два шара. В этом случае существуют четыре равновозможных (1/4) комбинации шаров: RRR, RWR, RRW и RWW. Если на самом деле имеет место первая комбинация, то вероятность вынуть шар R равна 1, для второй и третьей комбинаций эта вероятность равна 2/3 и для последней комбинации равна 1/4. Следовательно, вероятность вынуть шар R равна $1\cdot 1/4+2/3\cdot 1/4+2/3\cdot 1/4+1/3\cdot 1/4=2/3$. Таким образом, в мешке должны быть 2 шара R и 1 шар W, следовательно, перед тем, как мы положили в мешок шар R, в нем должен быть 1 шар R и 1 шар W. Этот результат, очевидно, абсурден, так что его вывод должен быть ошибочным. Но в чем ошибка?

Следующие рассуждения также приводят к абсурдным результатам. Двое из трех заключенных, обозначаемых A, B и C, будут казнены. Они это знают, но не могут догадаться, кому же из них повезет. A рассуждает: «Вероятность, что меня не казнят, равна 1/3. Если я попрошу охранника назвать имя (отличное от моего) одного из двух заключенных, которых казнят, то тогда останется только две возможности. Либо другой, кого казнят, это я, либо нет, и поэтому шансы, что я выживу, увеличатся до 1/2». Однако также справедливо, что уже перед тем, как A спросит охранника, он знает, что одного из его компаньонов наверняка казнят, так что охранник не сообщит A никакой новой информации относительно его судьбы. Почему тогда вероятность казни изменилась?

(Ответ очень прост: вероятность совсем не изменилась, она осталась равной 1/3. Заключенный упустил из виду, что охранник называет, например, B с вероятностью 1/2, если собираются казнить B и C, но эта вероятность равна 1, когда жертвами являются A и B. Следовательно, на самом деле шансы для A избежать казни равны отношению вероятности в последнем случае к сумме вероятностей в обоих случаях: $1/6(1/6 + 1/3) = 1/3$.)

ГЛАВА II

ПАРАДОКСЫ В МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

> Статистика — это физика чисел.
>
> *П. Диаконис*
>
> Все самое важное раньше сказал тот, кто этого не понял.
>
> *А. Н. Уайтхед*
>
> Если кто-то способен предсказать, чем закончатся его исследования, то эта проблема не очень глубока и, можно сказать, практически не существует.
>
> *А. Шильд*

Первоначально статистика была «государственной арифметикой». (Слово «статистика» происходит от латинского слова *status* — государство.) С древнейших времен статистику использовали для того, чтобы информировать правителей стран о величине налога, который можно собрать с их подданных, или о числе солдат, на которое можно рассчитывать в военное время. В Китае учет населения проводился более четырех тысяч лет назад. Согласно Библии, Моисей также подсчитывал всех мужчин своего народа старше 20 лет. Их оказалось 603 550 человек. Четвертая книга Моисея (Числа) содержит множество других результатов подсчета людей, однако они кажутся преувеличенными, так же как и данные *Афинея* о числе рабов в греческих полисах во времена Римской империи. Весьма сомнительно, чтобы в Афинах было 400 000 рабов, а в Коринфе — 460 000. Неясно, как получены эти данные, но точно известно, что согласно результатам учета населения Рим был первым городом с населением более миллиона жителей. Первый английский статистический документ «Книга сýдного дня», написанный в XI веке, также возник в связи с потребностями армии и налогообложения. По этой причине при переписи населения женщины не учитывались вплоть до недавнего времени. Статистика стала наукой лишь в XVII веке. Ее основоположниками являются *Джон Граунт* (1620—1674 гг.) и *сэр Уильям Петти* (1623—1687 гг.). В книге Граунта «Естественные и политические наблюдения, сделанные над бюллетенями смертности» (1662 г.) исследовались вопросы народонаселения. В 1669 г. Гюйгенс на основе данных Граунта опубликовал таблицы смертности. В книгах Петти «Трактат о налогах» (1662 г.) и «Наблюдения над дублинскими записями смертности» (1681 г.) также использовались результаты и идеи Граунта. В работе Петти «Политическая арифметика», опубликованной в 1689 г. после смерти автора, Англия, Голландия и Франция сравниваются по их населению, торговле и судоходству. Термин «политическая арифметика» можно считать предвестником слова «статистика». С развитием капитализма статистическими данными стали интересоваться не только государственные деятели, но и капиталисты. Для обработки данных использовались все более сложные математические методы, при этом увеличивалась и выгода от их применения, например, в страховом деле. Компания Ллойда, одна из крупнейших страховых компаний в мире, была основана в XVII веке и занимала в то время лишь кофейню на Тауэр-стрит в Лондоне. Успех в страховом деле определяется точностью данных и надлежащими математическими выводами. Математическая статистика, постепенно развиваясь с XVII века, превратилась сейчас в самостоятельную область математики. Ее основной целью является получение как можно более верной и полезной информации из данных результатов наблюдений и измерений, или кратко, *статистической выборки*. (Измерение количества информации независимо от ее конкретного содержания лишь в XX веке развилось в новую ветвь математики, которая теперь называется теорией информации. Она тесно связана с математической статистикой.) Трудно не писать сатирические произведения по крайней мере в духе Ювенала, но еще труднее не найти парадоксов в математической статистике. Согласно одной шутке в 1901 г. 33 % студенток Гарвардского университета вышли замуж за своих преподавателей. На самом же деле в то время в университете обучались только 3 девушки, одна из которых вышла замуж за своего профессора. Следовательно, утверждение верно, но вводит в заблуждение. Предположим, что в некоторой стране в университеты принято на 20 % юношей больше, чем девушек. Если все абитуриенты одинаково хорошо подготовлены и число абитуриентов-юношей совпадает с числом абитуриентов-девушек, то очевиден вывод, что приемные комиссии отдают предпочтение юношам. Однако, поскольку больше девушек, чем юношей, хотят учиться на более популярных факультетах, где доля непринятых выше, может оказаться, что несмотря на пропорциональный прием, в университете будет учиться больше юношей, чем девушек. Анализ текстов, проведенный в 1913 г. *Л. Эйресом*, аналогичным образом вводит в заблуждение или, по крайней мере, его легко можно неверно истолковать. Эйрес утверждал, что 50 наиболее часто употребляемых слов составляют приблизительно 50 % обычного текста, 300 наиболее часто встречающихся слов составляют 75 % текста, а 1000 наиболее часто употребляемых слов составляют 90 %. Несмотря на этот факт, нельзя сказать, что если нам известны 50 или 100 слов какого-то языка, то мы уже наполовину его понимаем, так как знание нескольких слов, даже если они часто используются, вряд ли поможет в понимании любого текста. Не удивительно, что многие люди считают, что существует три вида лжи: невинная ложь, наглое вранье и статистика. Надеемся, что объяснение парадоксов математической статистики поможет нам хорошо разбираться в статистических нелепостях, видеть пользу и необходимость статистических выводов, а также извлекать наиболее важную информацию из данных.

1. Парадокс Байеса

а) История парадокса

Томас Байес, ученик де Муавра, является одним из выдающихся основателей математической статистики. Его теорема, доказанная где-то около 1750 г. и опубликованная лишь после его смерти, стала источником некоторых разногласий в статистике. Жар споров до сих пор не утих. Более того, теоретическая пропасть между последователями байесовского и антибайесовского подходов продолжает увеличиваться. Простая формулировка теоремы Байеса заключается в следующем. Пусть A и B — произвольные события, имеющие вероятность $P(A) > 0$ и $P(B) \geqslant 0$ соответственно. Обозначим через $P(AB)$ вероятность совместного осуществления событий A и B, и пусть $P(A|B)$ есть условная вероятность A, если известно, что B уже произошло. Тогда

$$P(B \mid A) = \frac{P(AB)}{P(A)}, \text{ т. е. } P(B \mid A) = \frac{P(A \mid B)\, P(B)}{P(A)}.$$

Следовательно, если $B_0, B_1, \ldots$ — попарно непересекающиеся события, имеющие положительные вероятности, и одно из них происходит всегда (или по крайней мере с вероятностью 1), то

$$P(B_k \mid A) = \frac{P(A \mid B_k)\, P(B_k)}{P(A \mid B_0)\, P(B_0) + P(A \mid B_1)\, P(B_1) + \ldots}.$$

Это и есть формула Байеса. Она показывает, как по *априорным* вероятностям $P(B_k)$ (вероятностям событий B_k до того как событие A произошло) найти *апостериорные* вероятности (после того, как событие A произошло). Если рассматривать события B_k как причины, то формула Байеса представляет собой теорему о вероятностях причин. Сама по себе теорема бесспорна, но в большинстве ее применений вероятности $P(B_k)$ неизвестны. В этом случае, как правило, считают, что, поскольку отсутствует предварительная информация о причинах событий B_k, то все вероятности $P(B_k)$ равны, однако такой подход, вообще говоря, неприемлем. Байес использовал свою теорему в случаях, когда априорные вероятности были вероятностями непрерывных распределений, в частности, равномерного распределения на интервале (0, 1). По теореме Байеса, если в $n+m$ наблюдениях событие, имеющее неизвестную вероятность p, произошло n раз, то вероятность того, что p принадлежит подынтервалу (a, b) интервала (0, 1) равна

$$\int_{n}^{b} x^n (1-x)^m \, dx \Big/ \int_{0}^{1} x^n (1-x)^m \, dx.$$

Байес выдвинул идею о том, что, если у нас нет никакой предварительной информации о p, то априорная вероятностная плотность параметра p равномерна на всем интервале (0, 1). Например, если $n=1$, $m=0$, $a=1/2$ и $b=1$, то по приведенной выше формуле, шансы того, что искомая вероятность p больше 1/2, равны 3/4. До сих пор лишь немногие доверяют этому результату, в частности, потому что они сомневаются в равномерности априорного распределения.

Незнание априорного распределения оказалось столь разрушительным для обоснованности статистических выводов из теоремы Байеса, что эта теорема была почти исключена из статистических исследований. Однако во второй трети XX века байесовский подход вновь получил некоторое развитие, благодаря важной роли, которую он играет при поиске допустимых и минимаксных оценок (см. замечания в I. 12 и книгу *Фергюсона*). Все более распространялась мысль о том, что последовательное применение формулы Байеса (когда после каждого наблюдения апостериорные вероятности пересчитываются и на следующем шаге они используются как априорные вероятности) снижает роль исходного априорного распределения, так как после многократного пересчета исходное распределение вряд ли оказывает влияние на заключительное апостериорное распределение. (Очевидно, что некоторые вырожденные случаи не рассматриваются, например, когда значение p равно 1/10, а априорное распределение равномерно на отрезке [1/2, 1], не содержащем точку 1/10.)

б) Парадокс

Пусть возможными значениями случайной величины X являются целые числа, и предположим, что вероятностное распределение X зависит от параметра p, принадлежащего отрезку $[a, b]$. Если независимые наблюдения X_1, X_2, X_3, ... получены из неизвестного распределения X (т. е. распределения с неизвестным параметром p; X_i имеют то же распределение, что и X), то можно ожидать, что последовательность апостериорных распределений (вычисленных по исходному равномерному априорному распределению) все более и более концентрируется около истинного значения p. Парадоксально, но это не всегда верно. Например, истинное значение p может равняться $1/4$, а последовательность апостериорных распределений (при увеличении числа наблюдений) все более сосредотачивается, например, около $3/4$.

в) Объяснение парадокса

Парадоксальность ситуации состоит в том, что ожидается, что функция апостериорной плотности будет принимать наибольшие значения в окрестности истинного значения, т. е. вблизи $1/4$. Однако это соображение не противоречит тому, что функции апостериорных плотностей могут все более сосредотачиваться около $3/4$. Нужно только, чтобы функция плотности, имеющая пик в $1/4$, затем быстро убывала, но оставалась высокой около $3/4$. Если число возможных значений величины X конечно, то такая ситуация невозможна, но когда значениями X могут быть любые целые числа, парадоксальная ситуация может осуществиться. Пусть априорное распределение p равномерно на отрезке $[1/8, 7/8]$. Определим теперь функцию $f(p)$ на этом отрезке таким образом, что значениями $f(p)$ всегда являются натуральные числа, за исключением точек $p=1/4$ и $p=3/4$, где $f(1/4)=f(3/4)=+\infty$. Пусть распределение случайной величины X (зависящее от p) имеет следующий вид

$$P(X=i)=c(1-p)p^i, \quad i=0,\ 1,\ 2,\ \ldots,\ f(p),$$

где $c=c_p$ есть постоянная, для которой

$$\sum_{i=1}^{f(p)} c(1-p)p^i=1.$$

При соответствующем выборе $f(p)$ указанная выше парадоксальная ситуация осуществима. Дальнейшие детали см. в статье Фридмана [Freedman D. F. (1963)].

г) Замечания

(i) *С. Бернштейн* и *Р. Мизес* еще до 1920 г. указывали на то, что при некоторых условиях многократное применение теоремы Байеса дает последовательность апостериорных распределений, сходящихся к истинному распределению, каково бы ни было априорное распределение. Поэтому априорное распределение асимптотически не играет роли. Как показывает парадокс, такое утверждение невозможно без каких-либо ограничений.

(ii) Субъективный выбор априорных распределений порождает общий вопрос о том, можно ли вообще объективно определять неизвестные вероятности и вероятностные распределения независимо от наших наблюдений и измерений, или они имеют смысл только благодаря нашей субъективной информации. *Бруно де Финетти*, глава итальянской школы по теории вероятностей, в своей монографии утверждает, что вероятность как и флогистон [1]), не существует объективно в отличие от абсолютного пространства и времени или вселенной. «Объективная вероятность» является всего лишь попыткой выделить и материализовать наши вероятностные представления. По мнению Финетти любое событие (например, завтра пойдет дождь) либо произойдет, либо не произойдет (это объективно), и, опираясь на доступную информацию, мы можем посчитать «субъективную» вероятность события. Индивидуальная или субъективная вероятность отражает степень нашей уверенности в том, что событие произойдет. Мы можем говорить о субъективной вероятности, даже если «случайность» не объективна. Однако необходимо подчеркнуть, что значительно бо́льшая часть ученых утверждает, что объективная случайность и объективная вероятность существуют. Они убеждены в том, что объективные вероятности будущих событий заложены в современном состоянии мира. Так понимал объективное существование вероятности лауреат Нобелевской премии *Макс Борн*, который известен тем, что ввел объективную вероятность в квантовую физику [2]).

д) Литература

Bayes T. "An essay towards solving a problem in the doctrine of chances", 1763, Reprint: *Biometrika*, **45**, 293—315, (1958).

Berkson J. "My encounter with neo-Bayesianism", *Internat. Statist. Rev.*, **45**, 1—9, (1977).

Born M. *Natural Philosophy of Cause and Chance*, Dover, New York, 1964.

David A. P., Stone M., Zidek J. V. "Marginalization paradoxes in Bayesian and structural inference", *J. Roy. Statist. Soc., Ser. B.*, **35**, 189—233, (1973).

Ferguson T. S. *Mathematical Statistics, A Decision Theoretic Approach.*, Academic Press, New York — London, 1967.

de Finetti B. *Theorie Delle Probabilità*, Einaudi, Torino, 1970.

de Finetti B. "Bayesianism", *Internat. Statist Rev.*, **42**, 117—130, (1974).

Freedman D. F. "On the asymptotic behavior of Bayes' estimates in the discreete case", *Annals of Math. Statist.*, **34**, 1386—1403, (1963).

Holland J. D. "The reverend Thomas Bayes F. P. S. (1762—1761)", *J. Roy. Statist. Soc.* (A), **125**, 451—461, (1962).

Lindley D. V. "The use of prior probability distributions in statistical inference and decision", *Proc. 4th Berkeley Symp. on Math. Statistist. and Prob.*, **1**, 453—468, (1968).

---

[1]) Флогистон, по представлению химиков конца XVII — начала XVIII века, — гипотетическая составная часть веществ, которую они якобы теряют при горении и обжиге. Гипотеза флогистона опровергнута трудами А. Лавуазье. (Советский энциклопедический словарь). — *Прим. перев.*

[2]) См. также статью А. Н. Колмогорова «Вероятность» в «Математическом энциклопедическом словаре» (М.: Сов. энциклопедия, 1988). — *Прим. перев.*

Lindley D. V. "The future of statistics — a Bayesian 21th century", *Advances in Appl. Prob.*, 106—115, (1975).

Lindley D. V. "A problem in forensic science", *Biometrika*, **64**, 207—213, (1977).

Lindley D. V. "The Bayesian approach", *Scand. J. Statist.*, **5**, 1—26, (1978). (3rd point: Marginalization paradoxes.)

Pearson E. S. (ed.) *The History of Statistics in the 17th and 18th Centuries Against the Changing Background of Intellectual, Scientific and Religious Thought*, Lectures by K. Pearson given at University College London during the academic sessions 1921—1933. Griffin, London, (1978).

Pflug G. *Decision Theoretic Paradoxes in Decision Making under Uncertainty* (ed. R. W. Sholtz). Elsevier, 375—383, (1983).

Savage L. J. *The foundations of Statistics*, Dover, New York, (1972).

Stone M., Springer B. G. F. "A paradox involving quasi prior distributions", *Biometrika*, **52**, 623—627, (1965).

Shafer G. "Lindley's paradox", *J. Amer. Statist. Assoc.*, **77**, 325—334, (1982).

2. Парадокс оценок математического ожидания

а) История парадокса

Уравнивание противоположных значений и отклонений в «среднем», т. е. суммирование наблюдений в одно значение имеет давние традиции. Эсхил писал в трагедии «Эвмениды»: «Богу всегда середина любезна, и меру чтит божество»,[1] а последователи китайского философа Конфуция говорят, что «в неподвижности среднего (=Чжун Йюн) есть величайшее совершенство». Математически понятие «среднего» можно интерпретировать различными способами (арифметическое среднее, геометрическое среднее, медиана и т. д.). Однако в статистических применениях в течение долгого времени крайне важную роль играло арифметическое среднее. Уже в первых выдающихся результатах в теории вероятностей и математической статистике изучалось арифметическое среднее статистической выборки и росло понимание важности его использования.

Рассмотрим множество $\{F_\theta\}$, $\theta \in \Theta$, вероятностных распределений с конечным математическим ожиданием, где параметр θ как раз и является математическим ожиданием распределения F_θ. Мы хотим оценить значение независимого параметра θ, опираясь на наблюденные значения (т. е. выборку) $X_1, X_2, \ldots, X_n$ (предполагается, что элементы выборки X_i являются независимыми случайными величинами с распределением F_θ. Арифметическое среднее

$$\hat{\theta} = \overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

[1]) Эсхил, Трагедии. Пер. с древнегреческого С. Апта. — М.: Искусство, 1978. — *Прим. перев.*

как оценка параметра θ, обладает многими достоинствами, например, она всегда несмещена, т. е. $E(\hat{\theta})=\theta$ при всех $\theta\in\Theta$ (другими словами оценка колеблется около истинного значения). Законы больших чисел утверждают, что оценка $\hat{\theta}=\bar{X}$ состоятельна, т. е. для любого $\varepsilon>0$

$$\lim_{n\to\infty} P(|\hat{\theta}-\theta|<\varepsilon)=1 \text{ для всех } \theta\in\Theta,$$

поэтому ошибку оценивания можно сделать как угодно малой, взяв достаточно большую выборку. Однако может существовать много несмещенных состоятельных оценок параметра, тогда (среди них) целесообразно отдать предпочтение оценкам с меньшей дисперсией. Парадоксы здесь показывают, что (за исключением случая нормальных распределений) арифметическое среднее выборки не является несмещенной оценкой математического ожидания с наименьшей дисперсией. Более того, если мы не настаиваем на свойстве несмещенности, то даже в случае многомерных нормальных распределений не всегда полезно оценивать математическое ожидание выборочным средним, так как эта оценка не является допустимой для квадратичной функции потерь. [Определение допустимой оценки см. в замечании (ii) парадокса I. 12]. Аналогичный парадокс будет рассмотрен далее в этой главе (параграф 13 п. д).

б) Парадоксы

(i) (*Каган — Линник — Рао*) Пусть $F(x)$ — произвольная функция распределения с нулевым средним и конечным стандартным отклонением и пусть $F_\theta(x)=F(x-\theta)$, где параметр θ может принимать произвольные действительные значения. Если элементы выборки $X_1, X_2, \ldots, X_n$ являются случайными величинами с распределением F_θ, то выборочное среднее $\bar{X}$ дает состоятельную и несмещенную оценку неизвестного параметра θ (который, очевидно, равен математическому ожиданию распределения F_θ). Однако оценка $\bar{X}$ не очень эффективна (за исключением случая нормального распределения): для любого $n>2$ существует несмещенная оценка, стандартное отклонение которой меньше стандартного отклонения величины $\bar{X}$ (точнее, для всех θ ее стандартное отклонение по крайней мере столь же мало, как и у $\bar{X}$, и по крайней мере для одного θ оно строго меньше).

(ii) (*К. Стейн*) $\bar{X}$ дает «образцовую» оценку для математического ожидания нормального распределения: $\bar{X}$ является несмещенной состоятельной оценкой с минимальной дисперсией, допустимой при квадратичной функции потерь $L(\theta, c) = (\theta - c)^2$, а также — минимаксной оценкой. Именно поэтому открытие К. Стейна, сделанное более 20 лет назад и утверждающее, что в случае многомерных нормальных распределений оценка $\bar{X}$ не является допустимой, явилось сюрпризом. Точнее, рассмотрим вероятностные распределения, определенные на k-мерном евклидовом пространстве, координаты которых (для простоты) независимы и имеют нормальное распределение $N(\theta, \sigma)$, причем стандартное отклонение σ известно. Мы ищем допустимую оценку вектора $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$; при которой квадратичная функция потерь

$$L(\theta, \hat{\theta}) = \|\theta - \hat{\theta}\|^2 = \sum_{i=1}^{k} (\theta_i - \hat{\theta}_i)^2$$

минимальна в среднем. Тогда вектор $\hat{\theta} = \bar{X}$ (т. е. k-мерный вектор выборочного среднего) является допустимой оценкой только в одно- и двумерном пространствах и не является таковым в пространствах более высокой размерности (хотя минимаксное свойство $\bar{X}$ сохраняется). Открытие Стейна показывает, что даже тогда, когда рассматривается классическая проблема оценивания (т. е. оценка математического ожидания нормального распределения), $\bar{X}$ — это не единственная оценка, которую следует принимать во внимание.

в) Объяснение парадокса

(i) Интересный результат Кагана, Линника и Рао требует скорее доказательства, чем пояснений. Но вместо доказательства мы изложим метод построения асимптотически оптимальных оценок. Прежде всего, рассмотрим в качестве примера функцию равномерного распределения $F(x)$ на интервале $(-c, c)$ (где c — произвольное положительное число) и пусть $F_\theta(x) = F(x - \theta)$. Тогда $D^2(\bar{X}) = c^2/3n$. Если $X_1^* = \min X_i$ и $X_n^* = \max X_i$, т. е. X_1^* — наименьший и X_n^* — наибольший элементы выборки (они оба определяются однозначно с вероятностью 1, так как распределение непрерывно), то

$$D^2\left(\frac{X_1^* + X_n^*}{2}\right) = \frac{2c^2}{(n+1)(n+2)},$$

что при больших n намного меньше, чем $D^2(\bar{X})$. Поскольку полусумма $(X_1^* + X_2^*)/2$ также является несмещенной и состоятельной оценкой параметра θ, она предпочтительнее по сравнению с «обычной» оценкой $\bar{X}$. Перейдем к общему случаю. Пусть

$X_1^* \leqslant X_2^* \leqslant \ldots \leqslant X_n^*$ есть *упорядоченная выборка* (т. е. X_1^* является наименьшим элементом из $X_1, X_2, \ldots, X_n$ и т. д.) и

$$\tilde{X} = \sum_{i=1}^{n} a_{in} X_i^*,$$

где $a_{in} (i = 1, 2, \ldots, n; \ n = 1, 2, \ldots)$ суть действительные числа, зависящие от F. Можно показать, что при выполнении некоторых слабых условий выбор величин a_{in}, указанный ниже, приводит к несмещенной оценке параметра θ с минимальной (по крайней мере асимптотически при $n \to \infty$) дисперсией. Пусть $a(x)$ — действительная функция на $[0, 1]$ и $a_{in} = a(i/n)/n$. Если функция F трижды дифференцируема, то оптимальный выбор $a(x)$ определяется соотношением

$$a(F(x)) = -[(A + Bx)(\log f(x))']',$$

где

$$A = \frac{\mu_2}{\mu_0\mu_2 - \mu_1^2}, \quad B = \frac{\mu_1}{\mu_0\mu_2 - \mu_1^2},$$

$$\mu_0 = \int \frac{(f'(x))^2}{f(x)}\, dx, \quad \mu_1 = \int \frac{(f'(x))^2}{f(x)}\, x\, dx$$

и

$$\mu_2 = \int \frac{(f'(x))^2}{f(x)}\, x^2\, dx - 1$$

(штрихи обозначают производные, $F' = f$ есть плотность). Эта формула для $a(x)$ применима даже тогда, когда у F не существует математического ожидания и θ обозначает центр симметрии функции $f_\theta(x)$. Например, пусть $f_\theta(x) = 1/(\pi(1 + (x - \theta)^2))$ (плотность распределения Коши; см. историю парадокса II/4). Тогда как это не удивительно, $a(x) = -A \cos 2\pi x \sin^2 \pi x$ является оптимальным выбором, но $a(x)$ отрицательна (!) для x близких к 0 или 1. В этом случае «обычная» оценка $\bar{X}$ даже не будет состоятельной. (Более подробный анализ можно найти в статье Móri T. F., Székely G. J. «How to estimate location and scale parameters», Technical report, Eötvös L. Univ. (1986); см. также Chernoff H., Gastwirth J. L., Johns M. V. Jr. «Asymptotic distribution of linear combinations of order statistics with applications to estimation», *Annals of Math. Statistics*, 38, 52—72, (1967).)

(ii) После статьи Стейна, опубликованной в 1956 г., Джеймс и Стейн в 1961 г. предложили следующую простую оценку для математического ожидания многомерного нормального распределения

$$X^* = \left(1 - \frac{(k-2)\,\sigma^2}{\|\bar{X}\|^2}\right)\bar{X}, \quad \text{где} \quad k > 2.$$

Тогда $E\|X^*-\theta\|^2 < k$, но $E\|\bar{X}-\theta\|^2 = k$. Следовательно, оценка $\bar{X}$ действительно не является допустимой. Оценка X^* переводит вектор $\bar{X}$ ближе к началу координат, а так как начало координат можно выбрать произвольно, то оценка

$$Q+\left(1-\frac{(k-2)\,\sigma^2}{\|\bar{X}-Q\|^2}\right)(\bar{X}-Q)$$

также лучше, чем $\bar{X}$, при любом Q. Таким образом, оценка Джеймса — Стейна зависит от выбора начала координат Q, в то же время $\bar{X}$ от Q не зависит. (Можно показать, что оценка

$$\tilde{X}=\max\left\{1-\frac{(k-2)^2}{\|\bar{X}\|^2};\ 0\right\}\bar{X}$$

даже несколько лучше, чем X^*.)

Теперь попытаемся эвристически объяснить, почему оценка X^* лучше, чем $\bar{X}$. Рассмотрим в совокупности выборки в k независимых задачах оценки параметров. Разброс скалярных элементов выборки определяется частично (общим) стандартным отклонением σ каждого из k распределений, частично (вообще говоря) неравными математическими ожиданиями θ_i. Хотя эти неизвестные математические ожидания могут быть очень разными, разброс общей выборки может указывать на то, что значения θ_i различаются не сильно. Например, в случае, когда $\sigma = 1$ и приблизительно 16 % наблюдений превосходит 1, а 16 % наблюдений меньше -1, естественно считать, что все математические ожидания θ_i близки к нулю. В этом случае, если $X_i = 0.8$, то обычная оценка i-го параметра даст 0.8, в то же время согласно более «рациональной» концепции Джеймса — Стейна математическое ожидание θ_i близко к нулю. Хотя такое объяснение может убедить нас в «рациональности» оценок, предложенных Джеймсом и Стейном, однако их метод все же представляется странным, когда мы имеем дело с задачами, между которыми вряд ли существует какая-либо связь, например, когда нужно оценить математические ожидания (нормально распределенных) высоты тела, скорости света и цены продукта.

г) Замечания

(i) Следующее неравенство, принадлежащее *Крамеру* и *Рао*, дает полезную информацию, касающуюся обоих парадоксов. Пусть $f(x, \theta)$ — совместная плотность (зависящая от параметра θ) элементов выборки X_i, $i = 1, 2, \ldots, n$, и пусть $B(\theta)$ обозначает смещение $E(\hat{\theta}_n-\theta)$ оценки $\hat{\theta}_n$ от параметра θ. Тогда неравенство Крамера — Рао утверждает, что при выполнении некоторых условий регулярности справедливо соотношение

$$E(\hat{\theta}_n - \theta)^2 \geqslant B(\theta)^2 + \frac{(1 + B'(\theta))^2}{nI(\theta)},$$

где $B'(\theta)$ — производная функции $B(\theta)$ и

$$I(\theta) = E\,\frac{-d^2 \ln f(X_i, \theta)}{d\theta^2}$$

есть информация Фишера. Следовательно, скорость сходимости $E(\widehat{\theta}_n - \theta)^2$ к нулю не может быть быстрее, чем $1/n$. Однако в примере, рассмотренном в первом парадоксе (равномерное распределение, $B(\theta) = 0$ и поэтому $E(\widehat{\theta}_n - \theta)^2 = D^2(\widehat{\theta}_n)$), скорость сходимости составляет $1/n^2$. Противоречия здесь нет, так как пример представляет собой типичный случай, когда условия регулярности, которые упоминались выше, не выполнены. (Например, достаточным было бы следующее условие регулярности: множество чисел x, для которых функция $f(x, \theta)$ положительна, не зависит от θ.) Теперь относительно второго парадокса. Из неравенства Крамера — Рао вытекает, что если разрешено использовать смещенные оценки, т. е. мы отказываемся от условия $B(\theta) \equiv 0$, и если производная $B'(\theta)$ отрицательна, то $E(\widehat{\theta}_n - \theta)^2$ может убывать значительно быстрее, чем дисперсия несмещенной оценки с минимальной дисперсией.

(ii) Пусть $X_1^* < X_2^* < \ldots < X_n^*$ обозначают упорядоченную выборку и X' обозначает выборочную медиану, т. е. $X' = X^*_{(n+1)/2}$ для нечетных n и

$$X' = (X^*_{n/2} + X^*_{n/2+1})/2$$

для четных n. Если выборка взята из нормального распределения, то

$$D^2(\overline{X}) \approx \frac{2}{\pi} D^2(X') \approx 0.63 D^2(X'),$$

т. е. эффективность оценки X' составляет (асимптотически) всего лишь 63 % от эффективности $\bar{X}$ [1]). Однако ситуация изменится, если мы немного «расстроим» нормальное распределение: рассмотрим случайную величину, которая является смесью двух нормальных распределений, а именно, 91 % составляет нормальное распределение со средним θ и дисперсией 1 и 9 % — нормальное распределение с тем же средним θ и дисперсией 9.

---

[1]) Эффективность оценки $\widehat{\theta}$ пропорциональна $D^{-2}(\widehat{\theta})$. — *Прим. перев.*

В этом случае медиана X' является для θ оценкой лучшей, чем $\bar{X}$.

(iii) Следующий парадокс допустимой оценки принадлежит С. Масани. Пусть X_1, X_2 — две независимые случайные величины с математическими ожиданиями m_1, m_2. Масани строит два примера (один с биномиальным распределением, другой с нормальным), в которых оценка, зависящая только от X_2, является допустимой при оценивании m_1. Здесь мы рассмотрим только биномиальный случай. Пусть $X_1, X_2, \ldots, X_m$ — независимые случайные величины, имеющие биномиальные распределения с параметрами n_i, p_i. Можно доказать, что линейная оценка

$$P_1(X_1, X_2, \ldots, X_m) = \sum_{i=1}^{m} a_i X_i / n_i + c$$

является допустимой оценкой параметра p_1 при квадратичной функции потерь $L(p_1, p) = (p_1 - p)^2$ тогда и только тогда, когда или

$$0 \leqslant a_1 < 1, \quad 0 \leqslant c \leqslant 1$$

и

$$0 \leqslant \sum_{i=2}^{m} a_i + c \leqslant 1, \quad 0 \leqslant \sum_{i=1}^{m} a_i + c \leqslant 1,$$

или $a_1 = 1$ и $a_2 = a_3 = \ldots = a_m = c = 0$. Полагая $a_1 = 0$, получаем большой класс допустимых оценок параметра p_1, не зависящих от X_1.

д) Литература

Efron B. "Biased versus unbiased estimation", *Advances in Math.*, **16**, 259—277, (1975).

James W., Stein C. "Estimation with quadratic loss", *Proc. 4th Berkeley Symp. on Math. Statist. and Prob.*, **1**, 361—380. Univ. California Press, Berkeley, (1961).

Kagan A. M., Linnik Yu. V., Rao C. R. "On a characterization on the normal low based on a property of the sample average", *Sankhya, Ser. A*, **27**, 3—4, 405—406, (1965).

Masani S. M. "A paradox in admissibility", *Annals of Statist.*, **5**, 544—546, (1977).

Stein C. "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution", *Proc. 3rd Berkeley Symp. on Math. Statist. and Prob.*, **1**, 197—206, Univ. California Press, Berkeley, (1956).

Tukey J. W. "A survey of sampling from contaminated distributions", *Contrib. to Prob. and Statist.*, (Ed. I. Olkin) 448—485, Standford Univ. Press, (1960).

Zacks S. *The Theory of Statistical Inference*, Wiley, New York, (1971). [Имеется перевод: Закс Ш. Теория статистических выводов. — М.: Мир, 1975.]

3. Парадокс оценок дисперсии

а) История парадокса

Важнейшей характеристикой случайных величин и их распределений наряду с математическим ожиданием является дисперсия. Оценим неизвестную дисперсию D^2 случайной величины X по выборке $X_1, X_2, \ldots, X_n$ (где $X_1, X_2, \ldots, X_n$ являются независимыми наблюдениями, имеющими то же распределение, что и X). При известном математическом ожидании E оценка

$$\hat{D}_0^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - E)^2$$

— несмещенная. Ситуация меняется, когда E неизвестно и (в предыдущую формулу) вместо E ставится его несмещенная оценка $\bar{X}$. Тогда оценка

$$\hat{D}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

уже не является несмещенной. Поскольку несмещенность (со времен Гаусса) была одним из необходимых свойств, которыми должна обладать хорошая оценка, $\hat{D}^2$ изменяют так, чтобы получить несмещенную оценку. (Некоторые параметры вообще не имеют несмещенных оценок. В этих случаях ограничиваются требованием *асимптотической несмещенности*, т. е. $\lim\limits_{n\to\infty} E(\hat{\theta}_n) = \theta$ для всех $\theta \in \Theta$. Это требование для $\hat{D}^2$ выполнено.) Наряду со свойством несмещенности постепенно возникли и другие требования, которым должна удовлетворять хорошая оценка. Парадокс возникает тогда, когда различные требования к качеству оценок не приводят к одной и той же оценке.

б) Парадокс

Умножая $\hat{D}^2$ на множитель *Бесселя* $n/(n-1)$, получаем величину

$$D^{*2} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2,$$

являющуюся несмещенной оценкой дисперсии.

Предположим, что случайная величина X нормально распределена (с неизвестными математическим ожиданием и дисперсией), и нам нужны минимаксные оценки (см. замечания в I/12) с функцией потерь $L(\tilde{D}^2, D^2) = (\tilde{D}^2 - D^2)/D^4$. Тогда $\tilde{D}^2$ надо изменить следующим образом: умножим $\hat{D}^2$ на $n/(n+1)$ и получим минимаксную оценку

$$\frac{1}{n+1}\sum_{i=1}^{n}(X_i-\bar{X})^2.$$

(Ее риск равен $2/(n+1)$.) Итак, принцип минимакса и несмещенность приводят к различным оценкам.

в) Объяснение парадокса

Сумма $\sum_{i=1}^{n}(X_i-a)^2$ минимальна только при $a=\bar{X}$. Однако математическое ожидание E, вообще говоря, не равно величине $\bar{X}$ (лишь близко к ней). Следовательно, оценка $\hat{D}_0^2$, показывающая действительное отклонение, больше, чем $\hat{D}^2$. Вот почему необходима поправка Бесселя. С другой стороны нет причин, по которым минимаксная или допустимая оценки должны быть несмещенными. (В II/2 мы уже видели, что оценка Джеймса — Стейна для математического ожидания лучше, чем обычная несмещенная оценка $\bar{X}$.) Поскольку несмещенная и минимаксная оценки дисперсии нормального распределения не совпадают, в каждой конкретной задаче нужно решать, какую оценку предпочесть. К счастью, даже при малых значениях n различие между двумя оценками невелико. (Однако в других проблемах разница может быть существенной.)

г) Замечания

Как ни удивительно, но можно показать, что указанная выше минимаксная оценка не является допустимой. (См. статью Стейна или книгу Закса.) С другой стороны, если математическое ожидание нормального распределения известно, то оценка

$$\frac{1}{n+2}\sum_{i=1}^{n}(X_i-E)^2$$

является не только минимаксной (с риском $2/(n+2)$), но и допустимой при указанной выше функции потерь. (См. книгу Закса.)

д) Литература

Stein C. "Inadmissibility of the usual estimate for the variance of a normal distribution with unknown mean", *Annals Inst. Statist. Math.*, **16**, 155—160, (1964).

Zacks S. *The Theory of Statistical Inference*, Wiley, New York — London — Sydney — Toronto (1971). [Имеется перевод: Закс Ш. Теория статистических выводов. — М.: Мир, 1975.]

4. Парадокс метода наименьших квадратов

а) История парадокса

Из-за неизбежных ошибок измерений часто кажется, что теоретические формулы и эмпирические данные противоречат друг другу. В начале прошлого века *Лежандр*, *Гаусс* и *Лаплас* предложили эффективный метод, позволяющий уменьшить влияние ошибок измерений. (Например, Лежандр разработал и применил его в 1805 г. для нахождения орбит комет.) Основоположниками этой теории были *Галилей* (1632), *Ламберт* (1760), *Эйлер* (1778) и другие. Новый прием, названный методом наименьших квадратов, детально исследован Гауссом в его работе «Теория движения небесных тел» (1809). Именно Гаусс указал также на вероятностный характер этого метода. (Хотя Лежандр обвинял Гаусса в плагиате, он не мог представить для этого достаточных оснований. Гаусс претендовал на приоритет лишь в использовании метода, а не его публикации.) Лаплас опубликовал свой основной труд по теории вероятностей в 1812 г., посвятив его «великому Наполеону». На протяжении всей четвертой главы его книги излагается исчисление ошибок. С того времени метод наименьших квадратов развился в новый раздел математики. Возможности метода порой переоценивают и часто используют тогда, когда другие методы были бы более подходящими. На эту проблему обращал внимание еще Коши (Comptes Rendus, 1853) во время «дебатов» с *Бьенеме* (в ходе диспута Коши использовал плотность вероятности $1/(\pi(1+x^2))$, названную позднее его именем, хотя он и не был первым ученым, применившим «плотность Коши»).

б) Парадокс

Пусть $ae^{-b|x-\mu|}$ — плотность распределения наших наблюдений, подверженных случайным ошибкам измерений. Постоянные a и b известны, а μ нужно оценить. Проведем независимые наблюдения $X_1, X_2, \ldots, X_n$. По методу наименьших квадратов μ следует оценить величиной $\hat{\mu}$, которая минимизирует сумму

$$(X_1-\hat{\mu})^2+(X_2-\hat{\mu})^2+\ldots+(X_n-\hat{\mu})^2.$$

Легко посчитать, что эта сумма принимает наименьшее значение, когда $\hat{\mu}$ есть среднее арифметическое результатов наблюдений

$$\hat{\mu}=(X_1+X_2+\ldots+X_n)/n.$$

Однако, если нам нужна оценка $\bar{\mu}$, для которой максимальна вероятность (точнее, плотность вероятности) того, что n наблюдений будут именно $X_1, X_2, \ldots, X_n$, т. е. $\bar{\mu}$ максимизирует функцию

$$a^n e^{-b(|X_1-\mu|+\ldots+|X_n-\mu|)}$$

или, что эквивалентно, $\bar{\mu}$ минимизирует

$$|X_1-\bar{\mu}|+|X_2-\bar{\mu}|+\ldots+|X_n-\bar{\mu}|,$$

то мы приходим к противоречию, так как сумма квадратов и сумма абсолютных величин не достигают минимума при одном и том же значении μ, т. е. оценки $\hat{\mu}$ и $\bar{\mu}$ различны. Какая из них лучше?

в) Объяснение парадокса

Если ошибки измерения нормально распределены (т. е. если их плотность вероятности имеет вид $ae^{-b(x-\mu)^2}$, то указанного выше противоречия не будет, так как $\hat{\mu}$ максимизирует

$$a^n e^{-b((X_1-\mu)^2+\ldots+(X_n-\mu)^2)}.$$

В методе наименьших квадратов Гаусс исходил из предположения о нормальном распределении ошибок, что встречается на практике чаще всего. Однако, когда известно, что распределение ошибок отлично от нормального, использование метода наименьших квадратов не всегда выгодно. В указанном выше парадоксе применение оценки $\bar{\mu}$ более оправдано (см. также предыдущий раздел).

Используя стандартные понятия математической статистики, парадокс можно кратко сформулировать следующим образом: оценка по методу наименьших квадратов не всегда совпадает с оценкой максимального правдоподобия (об оценках максимального правдоподобия см. разд. 8). Действительно, если $f(x)$ — положительная плотность, полунепрерывная снизу в точке $x=0$; $f(x-\theta)$ — плотность распределения измерений и $\bar{X} = (X_1+X_2+\ldots+X_n)/n$ есть оценка максимального правдоподобия параметра θ для $n=2, 3$, то $f(x)$ является плотностью нормального распределения с нулевым средним. Это — закон Гаусса об ошибках, который можно доказать следующим образом: если предположить для простоты, что существует производная f', и произведение $\prod_{i=1}^{n} f(X_i-\theta)$ максимально при $\theta=\bar{X}$, то

$$\sum_{i=1}^{n} \frac{f'}{f}(X_i-\bar{X})=0,$$

т. е. (обозначая $\Delta_i = X_i - \bar{X}$) из равенства

$$\sum_{i=1}^{n} \Delta_i = 0 \quad \text{вытекает} \quad \sum_{i=1}^{n} \frac{f'}{f}(\Delta_i) = 0,$$

что возможно при $n = 2, 3$ (если отношение f'/f измеримо) только тогда, когда $\frac{f'}{f}(x) = cx$; откуда следует, что $f = de^{-cx^2}$, где c и d — положительные числа (в противном случае функция f не была бы плотностью). Таким образом, оценка параметра сдвига по методу наименьших квадратов совпадает с оценкой максимального правдоподобия только для нормальных распределений.

г) Замечание

Арифметическое среднее $\hat{\mu} = \bar{X}$ и медиана $\tilde{\mu}$ являются одними из немногих «простых» оценок максимального правдоподобия параметра сдвига μ, имеющих вид

$$L = \sum_{i=1}^{n} a_i X_i^*,$$

где $X_1^* \leqslant X_2^* \leqslant \ldots \leqslant X_n^*$ есть упорядоченная выборка и $a_i \geqslant 0$, $i = 1, 2, \ldots, n$, $\sum_{i=1}^{n} a_i = 1$. В статье, которая готовится к публикации, мы (с З. Буцоличем) доказываем, что если L отлично от арифметического среднего $\bar{X}$, то не более двух a_i могут быть отличны от 0. Если два коэффициента a_i отличны от нуля, то это либо a_1 и a_n, и в этом случае элементы выборки равномерно распределены на некотором интервале, либо ненулевыми являются два соседних коэффициента a_i и a_{i+1}. Возможно, что ненулевым является лишь один из коэффициентов a_i, например, когда L — медиана, или когда элементы выборки имеют показательное распределение, тогда $a_1 = 1$ $(L = X_1^*)$.

д) Литература

Berkson J. "Estimation by least squares and by maximum likelihood", *Proc. 3rd Berkeley Symp. on Math. Statist. and Prob.*, **1**, 1—11, (1956).

Bloomfield P. B., Steiger W. L. *Applications of Least Absolute Deviations*, Biskhäuser Verlag, Basel — Boston — Stuttgart, (1983).

Harter H. L. "The Method of least squares and some alternatives", Part. I—V. *Internat. Statist. Rev.* (1974—1975).

Линник Ю. В. Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений, изд. 2-е. — М.: Физматгиз, 1962.

Sheynin O. B. "C. F. Gauss and the theory of errors", *Archive for History of Exact Sciences*, **19**, 21—72, (1979).

Stigler S. M. "Cauchy and the witch of Agnesi", *Biometrika*, **61**, 375—380, (1974).

5. Парадоксы корреляции

а) История парадокса

К последней трети прошлого века некоторые науки (например, молекулярная физика) достигли такого уровня развития, что стало необходимым использование в них теории вероятностей и математической статистики. В 1859 г. книга *Дарвина* произвела революцию в биологии и вскоре после этого двоюродный брат Дарвина *Фрэнсис Гальтон* заложил основы генетики человека. (Исследования *Менделя* по генетике были заново «открыты» лишь на рубеже веков; слово «генетика» употребляется только с 1905 г., но результаты Гальтона привлекли всеобщее внимание уже в прошлом веке). Гальтон и его ученики (особенно *Карл Пирсон*) ввели такие важные понятия, как *корреляция* и *регрессия*, которые стали основными понятиями в теории вероятностей и математической статистике (а также в связанных с ними науках). Вес и рост человека, естественно, тесно связаны между собой, но они не определяют друг друга однозначно. Корреляция выражает эту связь одним числом, абсолютная величина которого не превосходит 1. Для двух случайных величин X и Y корреляция определяется следующим образом. Пусть E_x и D_x, E_y и D_y обозначают математическое ожидание и стандартное отклонение X и Y соответственно. Тогда коэффициент корреляции (или кратко: корреляция) для X и Y определяется формулой

$$r = r(X, Y) = \frac{E[(X - E_x)(Y - E_y)]}{D_x D_y}.$$

Абсолютное значение корреляции максимально (т. е. равно 1), когда между X и Y существует линейная зависимость, т. е. $Y = aX + b$ (где $a \neq 0$). Если X и Y независимы (и их дисперсии конечны), то их корреляция равна 0, другими словами, они некоррелированы. В математической статистике оценкой для корреляции r, как правило, является выборочный коэффициент корреляции, который строится по независимой выборке (X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n) следующим образом:

$$\hat{r} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2 \sum_{i=1}^{n} (Y_i - \overline{Y})^2}}.$$

В ряде случаев r хорошо характеризует связь между X и Y, но уже на рубеже веков вычислялись зависимости, лишенные смысла; например, корреляция между числом гнезд аистов и числом младенцев. Понятие корреляции постепенно мистифицировалось и некоторые «внутренние» (вообще говоря, случайные) связи стали считать существующими, если была велика корреляция (т. е. близка по абсолютной величине к 1). Вот почему возникли совершенно абсурдные результаты, и это чуть не дискредитировало всю статистику. Как правило, игнорировался тот факт, что большая корреляция для X и Y может быть результатом влияния какой-то третьей величины. Например, в Англии и Уэльсе заметили, что с увеличением числа радиослушателей возрастало число сумасшедших и умственно отсталых людей. Однако такая интерпретация совершенно ошибочна, так как нельзя психически заболеть от того, что слушаешь радио. Дело лишь в том, что с течением времени растет и число радиослушателей, и число случаев психических заболеваний, но между ними нет никакой причинной зависимости. К сожалению, неверные толкования не всегда столь очевидны, например, в технических или экономических приложениях. Сравнение вероисповедания и роста людей дает еще один пример надуманной зависимости, согласно которой при движении от Шотландии к Сицилии доля католиков в населении постепенно возрастает и в то же время средний рост людей убывает. Однако какая-либо причинная связь здесь совершенно невозможна. (В фашистской расовой теории еще более нелепые идеи провозглашались здравыми и даже научными.) Рассмотрим лишь некоторые из существующих парадоксов корреляции.

б) Парадоксы

(i) Пусть случайная величина X равномерно распределена на интервале $(-1, 1)$ и $Y = |X|$. Очевидно, что между X и Y существует тесная связь, однако их корреляция $r(X, Y) = 0$. (Корреляция для X и $Y = |X|$ всегда равна 0, когда X — случайная величина с конечной дисперсией и симметричным относительно 0 распределением.)

(ii) Пусть величины $X_1, X_2, \ldots, X_n$ обозначают температуру в комнате в n различных моментов времени и $Y_1, Y_2, \ldots, Y_n$ — количество топлива, использованное для обогрева в те же моменты времени (точнее, за данный промежуток времени; например, в течение часа до рассматриваемого момента). Логично считать, что чем больше топлива использовано, тем теплее будет в комнате. Это означает, что корреляция для X и Y строго положительна. Однако корреляция может оказаться отрицательной, что можно было бы интерпретировать так: чем больше топим, тем холоднее становится.

(iii) Пусть случайный вектор (X, Y) распределен нормально, т. е. его плотность имеет вид

$$f(x, y)=\frac{1}{2\pi D_x D_y \sqrt{1-r^2}}\exp\left\{\frac{1}{2(1-r^2)}\left[\left(\frac{X-E_x}{D_x}\right)^2-\right.\right.$$
$$\left.\left.-\frac{2r(X-E_x)(Y-E_y)}{D_x D_y}+\left(\frac{Y-E_y}{D_y}\right)^2\right]\right\},$$

где E_x, D_x, E_y и D_y — математические ожидания и дисперсии величин X и Y, а r — их корреляция. Предположим, что абсолютная величина корреляции строго меньше 1. При неизвестной корреляции r мы можем оценить ее через $\hat{r}$, используя n элементов выборки. Если E_x и E_y известны, то целесообразно в формуле для $\hat{r}$ заменить $\bar{X}$ и $\bar{Y}$ соответственно на E_x и E_y. Таким путем получим новую оценку $\bar{r}$. Поскольку $\bar{r}$ использует больше информации (а именно, знание величин E_x и E_y) можно было бы ожидать, что дисперсия у $\bar{r}$ меньше, чем у $\hat{r}$. Однако *А. Стюарт* вычислил, что

$$D^2(\hat{r})=\frac{1}{n}(1-r^2)^2, \text{ тогда как } D^2(\bar{r})=\frac{1}{n}(1+r^2);$$

таким образом, последняя дисперсия больше.

в) Объяснение парадоксов

(i) Если X и Y независимы, то $r(X, Y)=0$, но обратное утверждение неверно. Некоррелированные случайные величины могут быть сильно зависимы, как в указанном выше примере, когда $Y=|X|$. Поэтому «некоррелированность» не следует понимать как независимость. С другой стороны, можно доказать, что если X и Y некоррелированы и $x_1<X<x_2$, $y_1<$ $<Y<y_2$, то каковы бы ни были числа $x_1<x_2$ и $y_1<y_2$, величины X и Y независимы.

(ii) Нельзя забывать о влиянии температуры вне комнаты!

Корреляции часто получаются совершенно невероятными, потому что вычисляемый коэффициент корреляции для двух случайных величин искажается третьей «извне влияющей величиной». Как раз для того, чтобы избежать этих помех, было введено понятие *частной корреляции*. Если корреляция для X и Y вычисляется только после того, как влияние величины Z ликвидировано, то результат перестает быть парадоксальным. Пусть r_{12}, r_{13} и r_{23} обозначают корреляции $r(X, Y)$, $r(X, Z)$ и $r(Y, Z)$ соответственно. Тогда частная корреляция для X и Y без влияния Z равняется

$$\frac{r_{12}-r_{13}r_{23}}{\sqrt{\left(1-r_{13}^2\right)\left(1-r_{23}^2\right)}}.$$

В частном случае, когда $r_{13}=r_{23}=0$ частная корреляция для X и Y совпадает с корреляцией r_{12}. Когда r_{12}, r_{13}, r_{23} неизвестны, их можно оценить по выборке аналогично тому, как это происходило для r. С помощью этих оценок получим оценку коэффициента частной корреляции.

(iii) Парадокс Стюарта можно рассматривать с разных точек зрения. Главное заключается в том, что $\hat{r}$ и $\bar{r}$ не являются несмещенными оценками для r, т. е. тождества $E(\hat{r})\equiv r$ и

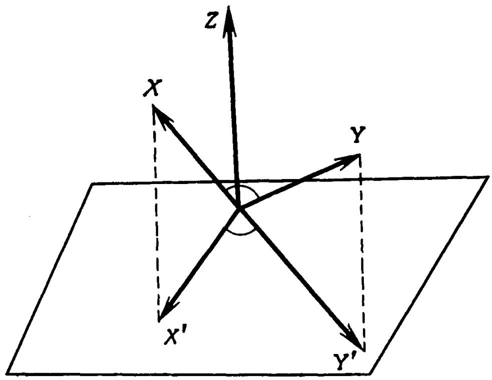

Рис. 9. Рассмотрим случайные величины X, Y и Z как векторы. Тогда корреляция для случайных величин X и Y равна косинусу угла между векторами X и Y, а их частная корреляция — косинусу угла между проекциями этих векторов на плоскость, перпендикулярную к вектору Z.

$E(\bar{r})\equiv r$ неверны, поэтому нецелесообразно считать лучшей ту оценку, у которой дисперсия меньше. В то же время обе оценки $\hat{r}$ и $\bar{r}$ смещены несильно (они являются асимптотически несмещенными), следовательно, для объяснения парадокса требуется дополнительный анализ. (См. замечания ниже и статью Стюарта.)

г) Замечания

(i) Смещение оценки r (в случае двумерного нормального распределения) равно

$$E(\hat{r}-r)=\frac{r^2-1}{n}+o(n^{-1}),$$

где $o(n^{-1})$ обозначает выражение, которое будучи умноженным на n, все же сходится к 0. Таким образом, смещение достаточно быстро стремится к 0 (при увеличении объема выборки n). С другой стороны, интересно отметить, что $\arcsin\hat{r}$

есть несмещенная оценка для $\arcsin r$ и, если $E(g(\hat{r}))=$ $=E(g(r))$ для некоторой функции g, не зависящей от n, то $g(r)=a\cdot\arcsin r+b$, где a, b — произвольные постоянные. В 1958 г. *И. Олкин* и *Дж. Пратт* доказали, что если оценка коэффициента корреляции r явным образом зависит от n, то можно указать несмещенную оценку для r, а именно

$$r^*=\hat{r}F\left(\frac{1}{2},\ \frac{1}{2},\ \frac{n-2}{2},\ 1-\hat{r}^2\right),$$

где F — гипергеометрическая функция, определяемая формулой $F(x,\ a,\ b,\ c)=$

$$=1+\sum_{k=1}^{\infty}\frac{a(a+1)\dots(a+k-1)\,b(b+1)\dots(b+k-1)}{k!c(c+1)\dots(c+k-1)}x^k,$$

где a, b, c ($c\neq 0,\ -1,\ -2,\ \dots$) являются параметрами. А среди несмещенных оценок уже следует предпочесть те, у которых дисперсия минимальна. Можно показать, что оценка r^* не только является несмещенной, но и имеет наименьшую дисперсию. Однако для практических применений оценка r^* достаточно сложна, поэтому рекомендуется использовать ее аппроксимацию $\hat{r}\left(1+\frac{1-\hat{r}^2}{2(n-4)}\right)$.

(ii) Следующий факт не является парадоксом, тем не менее удивительно, что при выборе наугад m чисел из множества $1, 2, \dots, n$ $\left(\text{выборка без возвращения, т. е. число равновозможных исходов равно } \binom{n}{m}\right)$ коэффициент корреляции для наименьшего и наибольшего среди выбранных чисел равен $1/m$, т. е. он не зависит от n ($m=1, 2, \dots, n-1$). Более того, если $X_1, X_2, \dots, X_m$ обозначают возрастающую последовательность из m выбранных значений, то

$$r(X_i,\ X_j)=\sqrt{\frac{i(m+1-j)}{j(m+1-i)}},$$

что также не зависит от n.

[Следующий результат принадлежит Т. Мори. Возьмем выборку с возвращением объема n из множества $1, 2, \dots, m+1$ и обозначим через Y_i число элементов выборки, величина которых не больше, чем i. Тогда $r(Y_i, Y_j)=r(X_i, X_j)$.] Другой задачей похожего типа является следующая. Если $Z_1\leqslant$ $\leqslant Z_2\leqslant\ \dots\ \leqslant Z_m$ есть возрастающая последовательность независимых равномерно распределенных случайных величин («упорядоченная выборка»), то $r(Z_i, Z_j) = r(X_i, X_j)$; если равномерное распределение заменить любым другим распределением, для которого $r(Z_i, Z_j)$ существует, то $r(Z_i, Z_j) \leqslant r(X_i, X_j)$, т. е. указанное свойство является экстремальным свойством равномерных распределений (см. статью Мори, Секея). В действительности можно доказать больше. Пусть максимальная корреляция для двух случайных величин U и V определяется как $\sup\limits_{f, g} r(f(U), g(V))$, где f и g пробегают множество интегрируемых в квадрате действительных функций от U и V соответственно:

$$\max \operatorname{corr}(U, V) = \sup_{f, g} r(f(U), g(V)).$$

Используя это обозначение, можно доказать, что

$$\max \operatorname{corr}(Z_i, Z_j) = r(X_i, X_j).$$

Равенство следует из того факта, что $\max \operatorname{corr}(U, V) = r(U, V)$, когда регрессия U на V (определение регрессии см. в следующем разделе) и регрессия V на U линейны (и не равны тождественно постоянной). Это как раз тот случай, когда коэффициент корреляции является хорошей мерой близости.

д) Литература

Kendal M. G., Stuart A. *The Advanced Theory of Statistics*, Vol. 2, Griffin, London, 1961. [Имеется перевод: Кендалл М. Дж., Стьюарт А. Статистические выводы и связи. — М.: Наука, 1973.]

Mori T. F., Székely G. J. "An extremal property of rectangular distributions", *Statistics and Probability Letters*, 3, 107—109, (1985).

Olkin I., Pratt J. W. "Unbiased estimation of certain correlation coefficients", *Annals of Math. Statist.*, 29, 201—211, (1958).

Stuart A. "A paradox in statistical estimation", *Biometrika*, 42, 527—529, (1955).

6. Парадоксы регрессии

а) История парадоксов

Коэффициент корреляции описывает зависимость между двумя случайными величинами одним числом, а регрессия выражает эту зависимость в виде функционального соотношения и поэтому дает более полную информацию. Например, регрессией является средний вес тела человека как функция от его роста. Понятие «регрессии» ввел *Гальтон*, который в конце прошлого века сравнивал рост родителей с ростом их детей. Он обнаружил, что рост детей у высоких (или низких) родителей обычно выше (или ниже) среднего, но не совпадает с ростом родителей. Линия, показывающая, в какой мере рост (и другие характеристики, к которым мы позже вернемся) регрессируют (возобновляются) в среднем в последующих поколениях, была названа Гальтоном *линией регрессии.* Позднее регрессией стали называть любую функциональную зависимость между случайными величинами. Вначале регрессионный анализ применялся в биологии и важнейшим научным журналом,

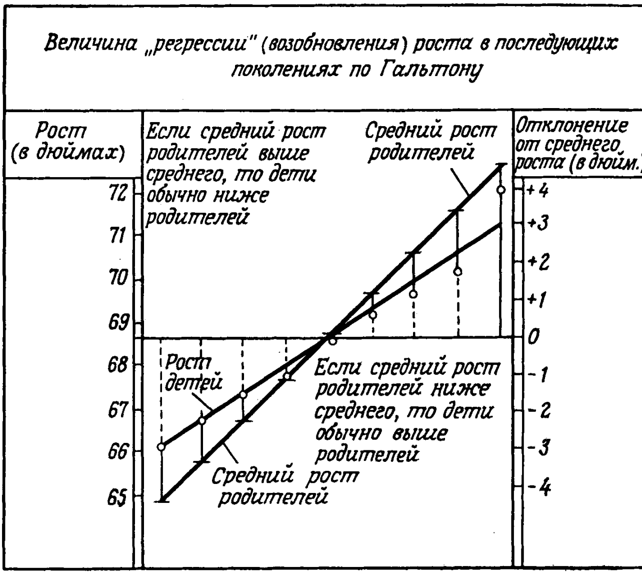

Рис. 10. Линия регрессии Гальтона.

в котором освещалась эта тема, был журнал «Биометрика» (“*Biometrika*”), выходящий с октября 1901 г. Между 1920 и 1930 гг. большое значение приобрело использование регрессионного анализа в экономике и возникла новая область науки: эконометрика (термин, принадлежащий *Р. Фришу* (1926), которому позднее была присуждена Нобелевская премия) со своим журналом «Эконометрика» (“*Econometrika*”), впервые вышедшим в 1933 г. От изучения частных регрессионных задач исследователи постепенно перешли к регрессионному анализу структуры, присущей глобальным экономическим системам (*Дж. Кейнс*, *Я. Тинберген* и другие, например, *Р. Клейн*, которому в 1980 г. присуждена Нобелевская премия по экономике). Журнал «Технометрика» (“*Technometrics*”) публикуется с 1959 г. и в основном посвящен техническим приложениям. Регрессионный анализ величины X, определяемой по другой величине Y, когда X измерить трудно, а Y достаточно легко, весьма важен. В настоящее время регрессионный анализ используется практически во всех областях науки, что само по себе неплохо, но, к сожалению, регрессионный анализ иногда является одним из главных средств для достижения «громких научных успехов», для проведения небрежных исследований и замазывания (научных) проблем. Регрессия никогда не подменяет научных концепций и теоретических обоснований, хотя и облегчает их поиск.

б) Парадоксы

Предположим, что зависимость двух величин выражается функцией следующего вида $y = f(x;\ a_1,\ a_2,\ \ldots,\ a_m)$, (например, $y = a_1x + a_2$), где неизвестны только параметры $a_1,\ a_2,\ \ldots$ $\ldots,\ a_m$ (а тип функции известен, например, линейный, квадратичный и т. д.). Если мы можем измерить значения y только со случайными ошибками наблюдений, т. е. вместо $y_i = f(x_i;$ $a_1,\ a_2,\ \ldots,\ a_m)$ мы наблюдаем значения Y_i, подверженные ошибкам, то согласно методу наименьших квадратов оценки неизвестных параметров a_i минимизируют сумму квадратов

$$\sum_{i=1}^{n} (Y_i - f(x_i;\ a_1,\ a_2,\ \ldots,\ a_m))^2.$$

(i) Если $f(x) = e^{ax}$, то оценка параметра a соответственно минимизирует сумму

$$\sum_{i=1}^{n} (Y_i - e^{ax_i})^2.$$

В этом случае задача вычисления регрессионной кривой f обычно упрощается, если вычислить логарифм от обоих членов разности, стоящих в скобках, и минимизировать величину

$$\sum_{i=1}^{n} (\ln Y_i - ax_i)^2,$$

что нетрудно сделать, находя минимум квадратичного многочлена. Однако эти два подхода к минимизации дают разные оценки. Какой выход из этой парадоксальной ситуации?

(ii) Предположим, что тип функции f можно выбрать различными способами, например, f_1 — это многочлен, а f_2 — экспоненциальная функция. Кажется естественным предпочесть тот тип, для которого указанная выше сумма квадратов меньше (при оптимальном выборе параметров). Хотя этому принципу часто следуют на практике, обычно он не оправдан (иногда следует установить хотя бы теоретическую возможность такого выбора).

(iii) Пусть $y = ax$ есть теоретическая линия регрессии и $Y_i = ax_i + \varepsilon_i$, где ε_i $(i = 1, 2, \ldots, n)$ являются независимыми нормально распределенными ошибками с математическим ожиданием 0 и дисперсией $D^2(\varepsilon_i) = cy$ (c — известная постоянная). Теперь предположим, что наблюдения идеально согласуются с линией регрессии, т. е. $Y_i = a_0 x_i$ для некоторого a_0 и

$$\sum_{i=1}^{n} (Y_i - a_0 x_i)^2 = 0.$$

Тогда оценкой параметра a по методу наименьших квадратов будет a_0, но, как ни парадоксально, она не является «лучшей» оценкой (в смысле максимального правдоподобия, определение см. в парадоксе 8).

в) Объяснение парадоксов

(i) Методу наименьших квадратов, несомненно, отвечает первая сумма, однако полезно разобраться не только в букве, но и в духе метода наименьших квадратов, сутью которого является минимизация суммарного влияния ошибок. Эта цель может быть достигнута путем минимизации суммы квадратов

$$\sum_{i=1}^{n} (h(Y_i) - h(f(x_i; a_1, a_2, \ldots, a_m)))^2,$$

где $h(x)$ — монотонно возрастающая функция (например, $h(x) = \ln x$). Хороший выбор h «линеаризует» задачу, т. е. делает выражение для $h(f(x_1; a_1, a_2, \ldots, a_m))$ линейной функцией от неизвестных параметров a_i (в этом случае легко находятся оптимальные значения для a_i). Если мы хотим определить неизвестные параметры, следуя духу метода наименьших квадратов, то, очевидно, лучше выбрать второй способ. Однако иногда нужно минимизировать все же исходную сумму, например, когда известно, что результатом ошибок являются финансовые потери, пропорциональные этой сумме, хотя такая возможность совершенно нетипична.

(ii) Первая часть вопроса очень проста: сумма квадратов может оказаться меньше для f_1, чем для f_2, но если взять чуть больше элементов выборки, то сумма квадратов становится меньше при выборе f_2. Математическая статистика старается избегать подобных неустойчивых ситуаций. Существует несколько методов принятия решений, которые применимы в ряде случаев и указывают выбор с заданной надежностью, например, 99 % (т. е. если функция f_1 отвергнута, то вероятность того, что правильным являлся выбор f_1, равна 1 %). В книге *Плакетта* обсуждается, например, метод, позволяющий определить правильную степень регрессионного многочлена (в случае независимых нормально распределенных ошибок наблюдений). К сожалению, многие из типичных задач по выбору вида регрессии невозможно решить должным образом. Например, *правило Вебера — Фехнера* утверждает, что между раздражителем и ощущением существует логарифмическая зависимость, в частности, между объемом и интенсивностью звука или между частотой и высотой звука. В настоящее время это правило теоретически и практически рассматривается лишь как первое приближение, потому что кажется, что ближе к истине является степенная зависимость. (В действительности, проблема сложнее, поскольку ощущение громкости зависит не только от интенсивности, но и от частоты и спектра звука, а также от продолжительности эксперимента.)

(iii) Оценка $\hat{a} = a_0$ не подходит, так как тогда оценка для $D^2(\varepsilon_i)$ равнялась бы нулю, что противоречит условию $D^2(\varepsilon_i) = cy$. Более оправданной будет оценка (максимального правдоподобия) $[(\sqrt{1+4c^2}-1)/(2c^2)]\,a_0$.

г) Замечания

(i) Очень типичной, в частности, в фармакологии и при изучении рынка сбыта, является логит-пробит альтернатива. В соответствии с методом наименьших квадратов в логит-анализе с данными согласуется функция

$$Y = e^{a_1+a_2x}/(1+e^{a_1+a_2x}),$$

минимизирующая сумму

$$\sum_{i=1}^{n}\left(\ln\frac{Y_i}{1-Y_i} - a_1 - a_2x_i\right)^2.$$

[Здесь преобразование, которое линеаризует задачу, дается функцией $h(x) = \ln(x/(1-x))$.] В пробит-анализе с данными согласуется функция нормального распределения (при соответствующем выборе параметров). Формы кривых этих двух типов могут быть очень похожи, поэтому не всегда легко решить, какую из них следует выбрать; в этом случае большую помощь может оказать теоретическое обоснование модели.

(ii) С увеличением числа параметров регрессии мы, очевидно, получим лучшее согласие наблюденных значений с моделью, однако при этом возрастут дисперсии оценок параметров, так что оценки станут менее устойчивыми и менее надежными.

(iii) «Парадокс двух регрессий» см. в статье Калмана (1982). В этой статье (вышедшей вслед за пионерскими работами Гини (1921) и Фриша (1934)) предполагается, что у обеих величин есть случайные (аддитивные) ошибки: $X = x + \hat{x}$ и $Y = y + \hat{y}$ ($\hat{x}$, $\hat{y}$ являются ошибками или «шумом»). Предполагая, что $y = ax$, можно дать «беспристрастную» оценку параметра a лишь в виде отрезка $a_1 \leqslant a \leqslant a_2$. Здесь одним из концов отрезка является классический коэффициент регрессии (когда ищут регрессию y по x), а другим концом — обратный коэффициент регрессии (когда ищут регрессию x по y). Выбор любого из концов отрезка a_1 или a_2 в качестве оценки означает, что предполагается отсутствие шума у регрессионной переменной. (Таким путем разрешается «парадокс двух регрессий».)

д) Литература

Berkson J. "Minimum chi-square, not maximum likelihood!" *Annals of Statist.*, **8**, 457—487, (1980).

Box G. E. P. "Use and abuse of regression", *Technometrics*, **8**, (1966).

Box G. E. P., Cox D. R. "An analysis of transformations", *J. R. Statist. Soc. Ser. B*, **26**, 211—243, (1964).

Cox D. R. *The Analysis of Binary Data*, Methuen, London, 1970.

Daniel C., Wood F. S. *Fitting Equation to Data*, Wiley, New York, 1971.

Draper N. R., Smith H. *Applied Regression Analysis*, Wiley, New York, 1966. [Имеется перевод: Дрейпер Н., Смит Г. Прикладной регрессионный анализ. — М.: Статистика, 1973.]

Durbin J. "Errors in variables", *Rev. Inst. Int. Statist.*, **22**, 23—32, (1954).

Frisch R. "Statistical confluence analysis by means of complete regression systems", *Publ. No. 5; Univ. Oslo Economic Inst.*, 192 pages, (1934).

Gini C. "Sull' interpolazione de una retta quando i valori della variabile indipendente son affetti da errori accidentali", *Metron*, **1**, 63—82, (1921).

Kalman R. E. "Identification from real data", In: *Current Developments in the Interface*: *Economics*, *Econometrics*, *Mathematics*, Reidel, (Eds. M. Hazewinkel and A. H. G. Rinnooy Kan), 161—196, 1982.

Plackett R. L. *Regression Analysis*, Oxford University Press, London, 1960.

Rao C. R. "Some thoughts on regression and prediction", *Proc. of the Symposium to Honour Jerzy Neyman*, Warsaw, 1974.

Sclove S. L. "(Y vs. X) or (log Y vs. X)?", *Technometrics*, **14**, (1972).

Stigler S. M. "Gergonne's 1815 paper on the design and analysis of polynomial regression experiments", *Historia Math.*, **1**, 431—477, (1974).

7. Парадоксы достаточности

а) История парадокса

Достаточность является одним из важнейших понятий в математической статистике. Ввел ее *Р. Фишер* в 20-е годы нашего века. Фишер выдвинул идею о том, что для статистического анализа, касающегося неизвестных параметров, не всегда нужно знать все элементы выборки в отдельности. Достаточно знать некоторые функции от выборки, называемые достаточными статистиками. Например, в случае одномерного нормального распределения вся информация о его математическом ожидании содержится в арифметическом среднем $\bar{X}$ элементов выборки $X_1, X_2, \ldots, X_n$. Это следует из того факта, что распределения случайного вектора $(X_1-\bar{X}, X_2-\bar{X}, \ldots, X_n-\bar{X})$ не зависит от неизвестного математического ожидания; и поэтому из знания случайных величин $X_1-\bar{X}, X_2-\bar{X}, \ldots, X_n-\bar{X}$ мы не получим о математическом ожидании никакой дополнительной информации. Математическое определение достаточности состоит в следующем. Функции $T_1 = T_1(X_1, X_2, \ldots, X_n)$, $T_2 = T_2(X_1, X_2, \ldots \ldots, X_n), \ldots, T_k = T_k(X_1, X_2, \ldots, X_n)$ называются достаточной статистикой для параметра θ распределения, общего для всех случайных величин X_i, если совместное распределение величин $X_1, X_2, \ldots, X_n$ при фиксированных $T_1, T_2, \ldots, T_k$ не зависит от θ. Возвращаясь к предыдущему примеру, получаем, что совместная условная плотность независимых случайных величин $X_1, X_2, \ldots, X_n$ при $\bar{X}=\bar{x}$ равна

$$\frac{1}{(\sqrt{2\pi}\,\sigma_0)^{n-1}\sqrt{n}}\,e^{-\frac{1}{2\sigma_0^2}\sum\limits_{i=1}^{n}(x_i-\bar{x})^2},$$

(где σ_0 обозначает стандартное отклонение величины X_i), и эта плотность не зависит от θ.

б) Парадокс

В 1934 г. Фишер указал на следующий парадокс. Он изучал двумерное нормальное распределение, координаты которого были (для простоты) независимыми случайными величинами с единичной дисперсией. Неизвестными были лишь их математические ожидания. Арифметическое среднее $\bar{X} = (\bar{X}_1, \bar{X}_2)$ двумерной выборки является достаточной статистикой для неизвестной пары математических ожиданий. Предположим, что известно расстояние между математическим ожиданием (рассматриваемым как вектор) и началом координат, т. е. пусть, например, $\sqrt{\theta_1^2+\theta_2^2}$ равно 3. Тогда $(\theta_1, \theta_2) = 3(\cos\theta, \sin\theta)$ где θ — единственный неизвестный параметр. Его можно оценить величиной $\hat{\theta} = \operatorname{arctg}(\bar{X}_2/\bar{X}_1)$. Эта оценка несмещенная: $E(\hat{\theta}) = \theta$, и ее дисперсия равна $E(\hat{\theta}-\theta)^2 = 0.12$. Легко доказать, что распределение величины $r=\sqrt{\bar{X}_1^2+\bar{X}_2^2}$ не зависит от θ (поскольку распределение вектора $(\bar{X}_1, \bar{X}_2)$ сферически симметрично относительно точки (θ_1, θ_2)), следовательно, если мы будем принимать во внимание r, то в силу достаточности не получим никакой информации о θ. Однако это совершенно неверно. Математическое ожидание случайной величины $(\hat{\theta}-\theta)^2$ (т. е. эффективность оценки) в значительной степени определяется r. Например, $E((\hat{\theta}-\theta)^2|r=1.5)=0.26$, $E=((\hat{\theta}-\theta)^2|r=3)=0.12$ и $E((\hat{\theta}-\theta)^2|r=4.5)=0.08$.

в) Объяснение парадокса

Парадокс Фишера указывает на то, что слова «обладать всей информацией» можно понимать по-разному. При вычислении эффективности вспомогательные статистики (аналогичные r) могут играть важную роль. К сожалению, далеко не всегда легко решить, что следует взять в качестве вспомогательной статистики. Очевидно не имеет смысла брать в качестве вспомогательной статистики всю выборку. Если рассмотреть проблему Фишера с точки зрения байесовского подхода и предположить, что случайная величина θ равномерно распределена на интервале $(-\pi, \pi)$, то

$$E(\hat{\theta}-\theta)^2|\bar{X}_1, \bar{X}_2)=E((\hat{\theta}-\theta)^2|r).$$

г) Замечания

Современная теория достаточности развита в работах *П. Халмоша* и *Л. Дж. Сэвиджа* (1949). В этой области также появился ряд интересных парадоксов. Например, Беркхолдер (см. список литературы) предложил несколько примеров, показывающих, что добавление к достаточным статистикам немного дополнительной информации может испортить достаточность. Такие примеры совершенно противоречат нашим представлениям о достаточности. За последнее десятилетие было опубликовано несколько глубоких статей, где вводятся некоторые «условия регулярности», обеспечивающие непарадоксальное поведение достаточных статистик.

д) Литература

Burkholder D. L. "Sufficiency in the undominated case", *Annals of Math. Statist.*, 32, 1191—1200, (1961).

Burkholder D. L. "On the order structure of the set of sufficient σ-fields", *Annals of Math. Statist.*, 33, 596—599, (1962).

Efron B. "Controversies in the foundations of statistics", *The American Math. Monthly*, 85, 231—246, (1978).

Fisher R. A. "On the mathematica foundations of theoretical statistics", *Phil. Trans. Roy. Soc. Ser. A*, 222, 309—368, (1922).

8. Парадоксы метода максимального правдоподобия

а) История парадоксов

Метод максимального правдоподобия является одним из наиболее эффективных методов оценки неизвестных параметров. Он получил распространение в 20-е годы нашего века благодаря работам английского статистика *Р. Фишера*. И хотя у Фишера были предшественники, именно его статья, написанная в 1912 г., сыграла решающую роль. Для знакомства с методом предположим для простоты, что у вероятностного распределения (зависящего от неизвестного параметра θ) существует плотность, которую обозначим через $f_\theta(u)$. Если элементы выборки $X_1, X_2, \ldots, X_n$ независимы, то их совместная плотность запишется в виде $\prod_{i=1}^{n} f_\theta(u_i)$. Пусть числа $x_1, x_2, \ldots, x_n$ — наблюденные значения выборки. Тогда $\hat\theta$ является оценкой максимального правдоподобия параметра θ, если $\hat\theta$ максимизирует произведение $\prod_{i=1}^{n} f_\theta(x_i)$ как функцию от θ (предполагаем, что максимум существует и единствен). В случае дискретных случайных величин X_i максимизируем совместную вероятность $P_\theta(X_1=x_1, X_2=x_2, \ldots, X_n=x_n)$. Если мы оцениваем θ по методу максимального правдоподобия, то вероятность (или плотность вероятности) того, что будут наблюдаться значения $x_1, x_2, \ldots, x_n$, становится максимальной. Оценка максимального правдоподобия обладает рядом хороших свойств, и поэтому соответствующий метод получил широкое распространение. Например, если $\hat\theta$ является оценкой максимального правдоподобия параметра θ, то $g(\hat\theta)$ — оценка максимального правдоподобия для $g(\theta)$. Можно также доказать, что при достаточно общих условиях оценка максимального правдоподобия $\hat\theta$ асимптотически ведет себя как нормально распределенная случайная величина со средним значением θ и дисперсией $1/(nI(\theta))$ (см. парадокс 2, замечание (i)), следовательно, $\hat\theta$ — состоятельная оценка, и ее дисперсия асимптотически минимальна (т. е. сама оценка $\hat\theta$ асимптотически эффективна). Более того, если достаточная статистика существует (сравните с «Парадоксами достаточности»), то метод максимального правдоподобия приведет к функции от этой достаточной статистики.

б) Парадоксы

(i) Пусть $X_1, X_2, \ldots, X_n$ — независимые случайные величины, равномерно распределенные на интервале $(\theta, 2\theta)$. Оценкой максимального правдоподобия неизвестного параметра θ является величина $\max(X_i/2)$. Немного изменив ее, получим

$$\hat\theta = \frac{2n+2}{2n+1} \max X_i/2$$

— несмещенную оценку для θ с дисперсией $D^2(\hat{\theta}) = 1/(4n^2)$. С другой стороны, дисперсия оценки

$$\frac{n+1}{5n+4}(\min X_i + 2\max X_i)$$

асимптотически эквивалентна $1/(5n^2)$, следовательно, эта оценка более эффективна, чем оценка максимального правдоподобия, обладающая наибольшей асимптотической эффективностью.

(ii) Можно привести очень простой пример, показывающий, что оценка максимального правдоподобия не всегда состоятельна. Пусть A — множество рациональных чисел между 0 и 1, а B — счетное множество иррациональных чисел между 0 и 1. Предположим, что значениями независимых элементов выборки $X_1, X_2, \ldots, X_n$ являются только 0 и 1, причем значение 1 принимается с вероятностью θ, если θ — элемент множества A, и с вероятностью $1-\theta$, если θ — элемент B. Тогда оценка максимального правдоподобия для θ несостоятельна. (Хотя несколько более сложная состоятельная оценка для θ все же существует.)

в) Объяснение парадоксов

(i) Статистики $\xi = \min X_i$ и $\eta = \max X_i$ в совокупности содержат всю информацию о параметре θ; точнее, при заданных ξ и η совместная плотность вероятности величин $X_1, X_2, \ldots, X_n$ не зависит от θ (т. е. ξ и η в совокупности образуют достаточную статистику). Таким образом, естественно считать, что как оценка максимального правдоподобия, так и оценка, которая оказалась лучше, зависят лишь от ξ и η. Поскольку оценка максимального правдоподобия зависит только от статистики η, которая не является достаточной (она не содержит всю информацию о θ), совсем неудивительно, что обнаружилась лучшая оценка. Это не противоречит асимптотической эффективности оценок максимального правдоподобия, так как в случае равномерного распределения «общие условия», обеспечивающие эффективность, не выполнены.

(ii) Объяснение достаточно просто: оценкой максимального правдоподобия для θ является относительная частота $\sum_{i=1}^{n} X_i/n$, которая стремится к $1-\theta$ для иррациональных θ.

Хотя эта задача в каком-то смысле патологическая, ее по крайней мере легко понять. (В статье *Д. Басу* дана состоятельная оценка для θ.) Существуют другие примеры несостоятельных оценок максимального правдоподобия, которые менее искусственны, но более сложны (сравните со статьями *Неймана — Скотта, Кифера — Волфовица и Фрегюсона*).

г) Замечания

(i) В статистической литературе можно обнаружить большое число оценок «максимального правдоподобия», при построении которых были найдены не точки максимума (а лишь седловые точки) или же рассматривался только один локальный максимум [1]. Хотя частое появление таких примеров достаточно любопытно, их нельзя считать парадоксами, а лишь «упущениями», даже если они опубликованы в первоклассных журналах в статьях лучших математиков.

(ii) Примеры *Дж. Ходжеса* и других поставили вопрос о парадоксе суперэффективности. Здесь мы лишь укажем диссертацию *Ле Кама* и статью *Г. Чернова*. (Оценка параметра θ является суперэффективной, если ее распределение асимптотически нормально со средним значением θ и дисперсией, которая не превосходит асимптотически минимальную дисперсию $1/(nI(\theta))$, и строго меньше ее по крайней мере для одного значения θ.)

(iii) В книге Питмена, указанной ниже, можно найти пример, когда оценки максимального правдоподобия для параметров сдвига и масштаба не существуют. А именно, если плотность вероятности f обладает следующим свойством: $|x|^{1+\varepsilon}f(x)\to\infty$ при $x\to\infty$ (или $-\infty$) для всех $\varepsilon>0$ (например, $f(x)=$ $=c/(x\log^2 x))$, $f_v(x)=\frac{1}{\sigma}f\left(\frac{x-\mu}{\sigma}\right)$, то для $v=(\mu,\sigma)$ оценки максимального правдоподобия не существует.

д) Литература

Bahadur R. R. "Examples of inconsistencies of maximum likelihood estimates", *Sankhya*, **20**, 207—210, (1958).

Barnett V. D. "Evalution of the maximum likelihood estimator when the likelihood equation has multiple roots, *Biometrika*, **53**, 151—165, (1966).

Basu D. "An inconsistency of the method of maximum likelihood", *Annals of Math. Statist.*, **26**, 144—145, (1955).

Berkson J. "Minimum chi-square, not maximum likelihood!" *Annals of Statist.*, **8**, 457—487, (1980).

Boyles R. A., Marschal A. W., Proschan R. "Inconsistency of distribution having increasing failure rate average", *Annals of Statist.*, **13**, 413—417, (1985).

Chernoff H. "Large sample theory: parametric case", *Annals of Math. Statist.*, **27**, 1—22, (1956).

Edwards A. V. T. "The history of likelihood", *Internat. Statist. Rev.*, **42**, 9—15, (1974).

Ferguson T. S. "An inconsistent maximum likelihood estimate", *J. Amer. Statist., Assoc.*, **77**, 831—834, (1982).

Fischer R. A. "On an absolute criterion for fitting frequency curves", *Messenger of Mathematics*, **41**, 155—160, (1912).

---

[1]) Одним из простейших и наиболее важных примеров, когда локальный максимум не единствен, является случай нормального распределения с неизвестным математическим ожиданием θ и дисперсией, пропорциональной θ^2.

Fischer R. A. "On mathematical foundations of theoretical statistics", *Phil. Trans. Roy, Soc. (London) Ser. A*, **222**, 309—368, (1922).

Kale B. K. "Inadmissibility of the maximum likelihood estimation in the presence of prior information", *Canad. Math. Bull.*, **13**, 391—393, (1970).

Kiefer J., Wolfowitz J. "Consistency of the maximum likelihood estimation in the presence of infinitely many incidental parameters", *Annals of Math. Statist.*, **27**, 887—906, (1956).

Konijin H. S. "Note on the nonexistence of a maximum likelihood estimation", *Aust. J. Statist.*, **5**, 143—146, (1963).

Kraft C., Le Cam L. "A remark on the roots of the likelihood equation", *Ann. Math. Statist.*, **27**, 1174—1177, (1956).

Le Cam L. "On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates", *Univ. California Publ. Stat.*, **1**, 277—330, (1953).

Neyman J., Scott E. L. "Consistent estimates based on partially consistent observations", *Econometrica*, **16**, 1—32, (1948).

Norden N. H. "A survey of maximum likelihood estimation", *Intern. Statist. Rev.*, **40**, 329—354, (1972).

Pitman E. J. G. *Some Basic Theory for Statistical Inference*, Wiley, New York, 1979. [Имеется перевод: Питмен Э. Основы теории статистических выводов. — М.: Мир, 1986.]

Rao C. R. "Apparent anomalies and irregularities in maximum likelihood estimation", *Sankhya*, **24**, 73—102, (1962).

Reeds J. A. "Asymptotic number of roots of Cauchy location equations", *Annals of Statist.*, **13**, 775—784, (1985).

9. Парадокс интервальных оценок

а) История парадокса

Теория интервального оценивания была разработана в основном *Г. Фишером* и *Е. Нейманом* между 1925 и 1935 гг. *Доверительный интервал* Неймана содержит неизвестный параметр 0 с заданной вероятностью α. Пусть $X_1, X_2, \ldots, X_n$ обозначают последовательность элементов выборки, и предположим, что $A = A(X_1, X_2, \ldots, X_n, \alpha)$ и $B = B(X_1, X_2, \ldots, X_n, \alpha)$ таковы, что $P(A < \theta < \mathrm{B}) = \alpha$. Тогда (A, B) называется α-доверительным интервалом для θ. Если θ обозначает неизвестное математическое ожидание нормального распределения со стандартным отклонением σ, то $P(\bar{X} - 2\sigma/\sqrt{n} < \theta < \bar{X} + 2\sigma/\sqrt{n}) \approx 0.95$, т. е. $(\bar{X} - 2\sigma/\sqrt{n},\ \bar{X} + 2\sigma/\sqrt{n})$ является 95 % доверительным интервалом для θ. При другом подходе к интервальному оцениванию случайным параметром считается не выборка, а неизвестный параметр θ. В этом случае интервал (A, B) не зависит от случая, и равенство $P(A < \theta < B) = \alpha$ просто означает, что θ попадает в интервал (A, B) с вероятностью α. Например, если θ обозначает неизвестное математическое ожидание нормального распределения, то из-за случайных ошибок измерений θ не определяется полностью выборочным средним $\bar{X}$. Такой параметр θ можно рассматривать как нормально распределенную случайную величину с математическим ожиданием X и стандартным отклонением $\sigma/\sqrt{n}$. Следовательно,

$$P(\bar{X} - 2\sigma/\sqrt{n} < \theta < \bar{X} + 2\sigma/\sqrt{n}) \approx 0.95.$$

Такой вид интервальных оценок, называемых *фидуциальными интервалами*, ввел *Фишер*. В случае нормального распределения, как мы видим, доверительные и фидуциальные интервалы формально совпадают; различается лишь их «философия». В течение некоторого времени считали, что эти два вида интервалов практически совпадают, и споры о различии между доверительными и фидуциальными интервалами являются чисто теоретическими. (Вначале Нейман поддерживал фидуциальную теорию Фишера главным образом потому, что Фишеру также не удалось использовать теорему Байеса.) Однако вскоре обнаружились парадоксы, имеющие практическое значение. Разные подходы Фишера и Неймана привели и к различным результатам в практических применениях. В 1959 г. *К. Стейн* указал на чрезвычайно парадоксальный случай. Для простоты он рассмотрел доверительные и фидуциальные интервалы, в которых $B=\infty$ или $A=-\infty$ потому, что такие интервалы определяются одним значением (вторым концом интервала).

б) Парадокс

Пусть $X_1, X_2, \ldots, X_k$ суть независимые нормально распределенные случайные величины с единичной дисперсией ($k \geqslant 2$). Обозначим через $\theta_1, \theta_2, \ldots, \theta_k$ их неизвестные математические ожидания. Пусть вектор $\theta=(\theta_1, \theta_2, \ldots, \theta_k)$ находится на расстоянии $|\theta|=\sqrt{\theta_1^2+\theta_2^2+\ldots+\theta_k^2}$ от начала координат. Стейн доказал, что доверительный и фидуциальный интервалы для $|\theta|$ могут сильно различаться, что и показано в следующем парадоксе. Будем оценивать каждое θ_i средним значением $\bar{X}_i$ выборки объема n. Пусть расстояние между началом координат и вектором выборочных средних $(\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k)$ равно $|\bar{X}|=\sqrt{\bar{X}_1^2+\bar{X}_2^2+\ldots+\bar{X}_k^2}$. Тогда $P(|\bar{X}|>|\theta|)>0.5$, когда $\bar{X}$ — случайная величина (строится доверительный интервал) и каково бы ни было значение неизвестного параметра θ. С другой стороны, если θ — случайная величина (строится фидуциальный интервал), то $P(|\theta|>|\bar{X}|)>0.5$ для любого значения выборочного среднего $\bar{X}$. Иными словами, вероятность того, что доверительный интервал $(-\infty, |\bar{X}|)$ содержит неизвестное значение $|\theta|$, больше 50 %; в то же время с вероятностью, большей 50 %, случайная величина $|\theta|$ попадает на (фидуциальный) интервал $(|\bar{X}|, +\infty)$. Таким образом, с точки зрения теории доверительных интервалов выгоднее ставить на неравенство $|\bar{X}| > |\theta|$, а при фидуциальном подходе ситуация прямо противоположная.

в) Объяснение парадокса

Невозможно показать все противоречия между фидуциальным подходом и теорией доверительных интервалов, возникающие в связи с задачей Стейна. Здесь мы ограничимся изложением решения, которое предложил сам Стейн. Если фидуциальный подход применяется не к элементам выборки, заданным своими координатами, а (в силу сферической симметрии нормального распределения) к суммам квадратов координат, то фидуциальные интервалы совпадают с доверительными интервалами (см. статью Стейна). Следовательно, выгоднее ставить на то, что «$|\bar{X}|$ больше, чем $|\theta|$».

г) Замечания

(i) Построим интервальную оценку для неизвестного математического ожидания θ нормального распределения с известным стандартным оклонением σ, используя априорную информацию о том, что величина θ нормально распределена с математическим ожиданием μ и стандартным отклонением s (величины μ и s известны). Если $\bar{X}$ — среднее значение выборки объема n, то по теореме Байеса *апостериорное* распределение величины θ также нормально с математическим ожиданием $\theta^* = \mu + C(\bar{X} - \mu)$ и стандартным отклонением D, где

$$C = \frac{n/\sigma^2}{1/s^2 + n/\sigma^2} \quad \text{и} \quad D^2 = \frac{1}{1/s^2 + n/\sigma^2}.$$

Следовательно, $(\theta^* - 2D,\ \theta^* + 2D)$ есть 95 % интервальная оценка для θ, так как $P(\theta^* - 2D < \theta < \theta^* + 2D) \approx 0.95$. Отсутствие априорной информации означает, что $s = \infty$, т. е. $C = 1$. Таким образом, $\theta^* = \bar{X}$ и $D^2 = \sigma^2/n$, а это и есть фидуциальный интервал. Следовательно, в случае многомерного нормального распределения байесовский подход приводит к тому же самому парадоксу, что и фидуциальный подход. Другой парадокс подобного типа (он исходит от московской статистической школы) заключается в следующем. Машина состоит из m деталей, соединенных последовательно. Поэтому, если i-я деталь исправна с вероятностью p_i $(i = 1, 2, \ldots, m)$, то машина исправна с вероятностью $p = p_1 \cdot p_2 \cdot \ldots \cdot p_m$. Возьмем теперь выборку из $n_1, n_2, \ldots, n_m$ элементов, и предположим, что все элементы исправно работают. Используя эту информацию, надо найти интервальную оценку в виде $P(p > p^*) = \alpha$. Как это ни странно, но доверительный интервал (т. е. величина p^*) не зависит от m, а зависит лишь от $\min\limits_{1 \leqslant i \leqslant m} n_i = n_0$ и соответствующей вероятности p_0. В то же самое время в рамках байесовского подхода интервальная оценка для p зависит от m.

(ii) Фишер (1890—1962 гг.) начал заниматься интервальными оценками немного раньше, чем Нейман (1894—1981 гг.). Фишер даже обвинил Неймана, который тогда работал в Польше, в присваивании и обобщении своих идей. В это время у Фишера уже были личные и профессиональные конфликты с другими выдающимися статистиками. Он ненавидел *К. Пирсона* (1857—1936 гг.) и поэтому не публиковался после 1920 г. в журнале «Биометрика» (ведущем периодическом журнале по статистике, среди основателей и редакторов которого был Пирсон). Фишер перенес свою антипатию, хотя и в ослабленном виде, на Э. Пирсона (1895—1980 гг.), сына К. Пирсона, и его друга Е. Неймана. Позднее Нейман стал одним из ведущих статистиков в США, и их личный спор перерос в спор англо-американский. Фишеру никогда не нравилась идея сведения статистических выводов к принятию решений с помощью функций потерь. (Это «американское» направление в статистике было разработано венгром *Абрахамом Вальдом* на основе теории игр Джона фон Неймана.) Главное противоречие выражалось в следующем: в Америке (в соответствии с прагматизмом *Пирса*) не важно, о чем вы думаете, а важно, что вы делаете. В Англии же — наоборот. Хотя рассуждения Фишера и не были всегда убедительными, он является одним из крупнейших (если не крупнейшим) статистиков, когда-либо живших. Поэтому странно, что он не стал профессором статистики. В действительности, в 1943 г. он-таки стал профессором Кембриджского университета, но по генетике. Между 1952 и 1954 г. он также был президентом Королевского общества.

(iii) Нам нужно оценить параметр сдвига θ по выборке $X_1, X_2, \ldots, X_n$, элементы которой имеют показательную плотность вероятности $e^{\theta - x}$ (если $x > \theta$ и 0 в противном случае). Оценка

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} (X_i - 1)$$

является несмещенной, и ее плотность вероятности пропорциональна $(x - \theta + 1)^{n-1} e^{-n(x-\theta+1)}$ для $x > \theta - 1$. С помощью этой плотности можно легко найти 90 % доверительный интервал наименьшей длины. В случае, когда $X_1 = 12$, $X_2 = 14$, $X_3 = 16$, этот доверительный интервал запишется в виде $12.1471 < \theta < 13.8264$. С другой стороны, θ, очевидно, меньше, чем $X_1^* =$

$= \min X_i = 12$. *Таким образом, 90 % доверительный интервал наименьшей длины лежит в области, в которой θ находиться не может!* Джейнес подчеркнул (см. ссылку ниже), что для построения интервальной оценки следует воспользоваться байесовским подходом. Если априорная плотность равна постоянной величине, то апостериорная плотность величины θ будет $ne^{n(\theta - X_1^*)}$, если $\theta < X_1^*$ и 0 в противном случае. Таким образом, интервал $X_1^* - q < \theta < X_1^*$, где $q = n^{-1}\log(1 - P)$, задает наименьшую апостериорную зону, содержащую апостериорную вероятность с вероятностью P. Для указанной выше выборки получим $11.23 < \theta < 12.0$. С точки зрения теории доверительных интервалов можно было бы сказать, что $\widehat{\theta}$ не является достаточной статистикой для θ, а статистика X_1^* — достаточная. Доверительный интервал наименьшей длины, построенный по достаточной статистике, совпадает с байесовским интервалом, указанным выше. Но даже если мы работаем с X_1^*, может оказаться, что 90 % доверительный интервал $(-\infty, f(X_1^*))$ лежит на отрицательной полуоси, а нам известно (априорная информация), что величина θ не может быть отрицательной.

д) Литература

Bartholomew D. J. "A comparison of some Bayesian and frequentist inference", *Biometrika*, **52**, 19—35, (1965).

Birnbaum A. "On the foundation of statistical inference (with discussion)", *J. Amer. Statist. Assoc.*, **57**, 269—326, (1962).

Box J. E., Fisher R. A. *The life of Scientist*, Wiley, New York, 1978.

Dempster A. P. "On a paradox concerning inference about a covariance matrix", *Annals of Math. Statist.*, **34**, 1414—1418, (1963).

Efron E. "Controversies in the foundations of statistics", *The American Math. Monthly*, **85**, 231—246, (1978).

Fischer R. A. "Inverse probabilities", *Proc. Cambridge Phil. Soc.*, **26**, 528—535, (1930).

Fisher R. A. *Statistical Methods and Scientific Inference*, Oliver and Boyd, London, 1956.

Jaynes E. T. "Confidence intervals vs Bayesian intervals", In: *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Sciences*, eds. by Harper W. L. and Hooker C. A.; D. Reidel Publishing Co., Dordrecht, Holland, 1976.

Jeffreys H. *Theory of Probability*, Clarendon Press, Oxford, 1967.

Kendall M. G. "R. A. Fischer 1890—1962", *Biometrika*, **50**, 1—16, (1963).

Neyman J. "On the two different aspects of representative method: The method of stratified sampling and the method of purposive selection", *Estadistica*, **17**, 587—651, (1959).

Robinson G. K. "Some counterexamples to the theory of confidence intervals", *Biometrika*, **62**, 155—162, (1975).

Stejn C. "An example of wide discrepancy between fiducial and confidence intervals", *Annals of Math. Statist.*, **30**, 877—880, (1959).

Stone M. "Strong inconsistency from uniform priors", *J. Amer. Statist. Assoc.*, **71**, 114—116, (1976).

10. Парадокс проверки гипотез

а) *История парадокса*

Трудно сказать что-то определенное о том, когда предпринимались первые попытки проверять статистические гипотезы. *Б. В. Гнеденко* в своей книге отмечает, что учет населения, проведенный в древнем Китае в 2238 г. до н. э., показал, что доля родившихся мальчиков составляла 50 %. *Джон Арбутнот* (1667—1735 гг.), английский математик, врач и писатель, был первым, кто (в 1710 г.) заметил, что гипотеза о равном соотношении родившихся мальчиков и девочек должна быть отвергнута, так как согласно демографическим данным за 82 года (доступным в то время) мальчиков каждый год рождалось больше, чем девочек. Если бы вероятность рождения мальчика была равна 1/2, то итог за 82 года был бы настолько маловероятен ($1/2^{82}$), что его можно было бы считать практически невозможным. Итак, Арбутнот был первым, кто отверг естественную статистическую гипотезу. Этот (нематематический) парадокс заинтересовал *Лапласа*. В 1784 г. он с удивлением обнаружил, что в нескольких различных районах доля родившихся мальчиков приблизительно равнялась 22/43, а в Париже это отношение было равно 25/49. Лаплас был заинтригован таким различием, но вскоре нашел для него разумное объяснение: в общее число родившихся в Париже включались также все подкидыши, а население из пригородов предпочитало подкидывать младенцев одного пола. Когда Лаплас исключил подкидышей из общего числа родившихся, доля новорожденных мальчиков стала близкой к 22/43.

В 1734 г. Французская академия присудила *Даниилу Бернулли* премию за исследование по орбитам планет. С помощью некоторого критерия проверки гипотез Бернулли пытался показать, что схожесть орбит планет является далеко не случайной. Из правила правой руки ясно, что каждая орбита соответствует некоторой точке на единичной сфере, и Бернулли проверял гипотезу о том, что распределение этих точек на единичной сфере равномерно. В 1812 г. Лаплас исследовал похожую проблему. Он пытался применить статистические методы для решения вопроса о том, какую из гипотез следует принять: являются ли кометы обычными элементами Солнечной системы или они всего лишь «незваные гости». В последнем случае углы между орбитами комет и эклиптикой были бы равномерно распределены на интервале от 0 до $\pi/2$, что как раз совпадает с математической записью предположения Лапласа. (Он обнаружил, что кометы не являются обычными элементами Солнечной системы.) Основоположниками современной теории проверки статистических гипотез были *К. Пирсон*, *Э. Пирсон*, *Р. Фишер* и *Е. Нейман.*

Предположим, что нужно проверить гипотезу о том, что распределение некоторой случайной величины равно F. (В проблеме Лапласа распределение F было равномерным на интервале $[0, \pi/2]$.) Для решения этой проблемы «степени согласия» *К. Пирсон*, *Х. Крамер*, *Р. фон Мизес*, *А. Н. Колмогоров*, *Н. В. Смирнов* и другие ученые, работавшие позднее, предложили несколько различных критериев, и возникла необходимость сравнивать их эффективности. Э. Пирсон и Е. Нейман сделали первые шаги на пути решения теоретических задач по нахождению лучших методов принятия решений. Во-первых, они ввели понятие *альтернативной гипотезы*, которая, вообще говоря, не является полным отрицанием основной, *нулевой гипотезы.* Рассмотрим, например, случайную величину, имеющую нормальное распределение с единичной дисперсией и неизвестным математическим ожиданием. Если *нулевая гипотеза* состоит в том, что «математическое ожидание равно -1», а альтернативная гипотеза заключается в том, что «математическое ожидание равно $+1$», то обе гипотезы, очевидно, не охватывают все возможные случаи. В 1933 г. Нейман и Пирсон показали, что для таких простых гипотез (когда как нулевая, так и альтернативная гипотезы, определяются по одному распределению) существует критерий, наиболее мощный в следующем смысле. При использовании статистических критериев возможны ошибки двух типов. Можно отвергнуть нулевую гипотезу, когда она верна, и допустить тем самым *ошибку 1-го рода* (или ошибку первого типа). С другой стороны, можно принять нулевую гипотезу, когда она неверна, и допустить *ошибку 2-го рода* (или ошибку второго типа). Метод принятия решений (критерий), основанный на выборке данного объема, называется наиболее мощным критерием, если для любой заданной вероятности ошибки 1-го рода вероятность ошибки 2-го рода мала настолько, насколько это возможно. (При фиксированном объеме выборки сумма вероятностей ошибок обоих типов не может быть сделана как угодно малой. Это своего рода *принцип неопределенности* при проверке гипотез.) Предположим для простоты, что оба распределения (в нулевой и альтернативной гипотезах) имеют плотности вероятности. Тогда по основной лемме Неймана — Пирсона существует наиболее мощный критерий следующего вида. Обозначим через f_0 и f_1 плотности вероятности выборки $X = (X_1, X_2, \ldots, X_n)$ при условии, что верна соответственно нулевая или альтернативная гипотеза. Нулевая гипотеза принимается тогда и только тогда, когда

$$\frac{f_1(X)}{f_0(X)} < c, \text{ где } c \text{ — соответствующая постоянная.}$$

(Для простоты предполагается, что вероятность того, что $f_1(X)/f_0(X) = c$ равна 0.) Теория Неймана и Пирсона стала основной при проверке гипотез, не лишенной, однако, парадоксов. В 1950 г. Герберт Роббинс показал, что существует критерий, который в некотором смысле является более мощным, чем наиболее мощный критерий Неймана — Пирсона.

б) Парадокс

Предположим, что случайная величина X нормально распределена с математическим ожиданием θ и дисперсией 1. Пусть нулевая гипотеза состоит в том, что $\theta = -1$, а альтернативная гипотеза заключается в том, что $\theta = +1$. На основе выборки из одного элемента X наиболее мощным критерием проверки нулевой гипотезы против альтернативной гипотезы является следующий: если $X \leqslant 0$, то нулевая гипотеза принимается, а альтернативная отвергается; в противном случае нулевая гипотеза отвергается, а альтернативная принимается. В этом случае вероятность ошибок обоих видов составляет приблизительно 16 %, так как

$$P(X > 0 \mid \theta = -1) = P(X < 0 \mid \theta = +1) =$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-(x+1)^2/2}\, dx = 0.1587 \ldots \approx 0.16.$$

Если воспользоваться этим критерием в N независимых случаях, то при больших N среднее число ошибочных решений приблизительно равно $0.16\, N$. Поскольку в каждом случае использовался наиболее мощный критерий, следовало бы ожидать, что среднее число ошибочных решений никогда не может быть меньше, чем $0.16N$. Как ни парадоксально, но следующий метод Роббинса показывает, что это не так.

Пусть $\bar{X}$ — арифметическое среднее наблюдений $X_1, X_2, \ldots$ $\ldots, X_N$. Критерий Роббинса состоит в следующем:

если $\bar{X} < -1$, то $\theta_i = -1$ для всех $i = 1, 2, \ldots, N$,

если $\bar{X} > +1$, то $\theta_i = +1$ для всех $i = 1, 2, \ldots, N$, и, наконец,

если $-1 \leqslant \bar{X} \leqslant +1$, то $\theta_i = -1$ или $\theta_i = +1$, в зависимости от того, выполнено неравенство

$$X_i \leqslant \frac{1}{2} \ln \frac{1 - \bar{X}}{1 + \bar{X}}$$

или нет. Этот метод удивителен тем, что он объединяет независимые друг от друга задачи. Если истинное отношение тех θ_i, для которых $\theta_i = +1$, к тем θ_i, для которых $\theta_i = -1$, равно 0 или 1, то при больших N (например, для $N = 100$) критерий Роббинса дает ответ со 100 % надежностью; для отношения 0.1 или 0.9 вероятность ошибки (обоих типов) составляет 7 %; для отношения 0.2 или 0.8 вероятность неверного решения равна 11 %; для отношения 0.3 или 0.7 она составляет 14 % и даже для отношения 0.4 или 0.6 процент ошибок меньше 16 % уровня наиболее мощного критерия. Метод Роббинса становится менее эффективным, чем наиболее мощный критерий, лишь в случае отношения, близкого к 0.5.

в) Объяснение парадокса

Парадокс Роббинса показывает, что даже тогда, когда нужно принять решение о приеме или отказе от продукции, поступающей с различных независимо работающих фабрик, общее число ошибочных решений будет в среднем меньше, если мы не будем принимать решения независимо одно от другого. Поскольку это по существу та же проблема, что и парадокс Стейна о допустимых оценках для математического ожидания, здесь мы лишь предложим посмотреть объяснение в разд. 2 и фундаментальную статью Роббинса.

г) Замечание

Другие парадоксы проверки гипотез будут обсуждаться в разд. 12 и 13.

д) Литература

Lehmann E. L. *Testing Statistical Hypotheses*, Wiley, New York, 1959. [Имеется перевод: Леман Э. Проверка статистических гипотез. — М.: Наука, 1979.]

Neyman J. "Two breakthrough in the theory of statistical decision making", *Internat. Statist. Rev.*, **30**, 11—27, (1962).

Neyman J., Egon S. "Pearson (August 11, 1895 — June 12, 1980)", *Annals of Statist.*, **9**, 1—2, (1981).

Neyman J., Pearson E. S. "On the problem of the most efficient test of statistical hypotheses", *Phil. Trans. Roy. Soc.*, **231**, 289—337, (1933).

Robbins H. "Asymptotically subminimax solution of the compound statistical decision problem", *Proc. 2nd Berkeley Symp. on Math. Statist. and Prob.*, 131—148, Univ. Calif. Press, Berkeley, (1950).

11. Парадокс Реньи из теории информации

а) История парадокса

Одной из основных задач теории информации является измерение количества информации. Основатели этого раздела математики (*К. Шеннон*, *Н. Винер* и другие) понимали, что количество информации измеряется числом, не зависящим от действительного значения и вида информации, подобно тому, как объем тела, которое может изменять свою форму, от формы не зависит. Единицей измерения информации служит информационное содержание ответа «да» или «нет». В двоичном коде эту информацию можно записать одной цифрой (например, 1 для «да», 0 для «нет»), которая называется *битом* (bit — сокращение от binary digit). Такое сокращение особенно удобно, так как слово «bit» означает «кусочек». Содержание информации измеряется средним количеством двоичных кодов, необходимых для записи информации. Если случайная величина может принимать только конечное или счетное число значений с положительными вероятностями $p_1, p_2, \ldots$, то согласно *формуле Шеннона* ее количество информации равно

$$H = H(p_1, p_2, \ldots) = \sum_i p_i \log p_i^{-1},$$

где log означает логарифм по основанию 2. Величина H называется *энтропией* вероятностного распределения $p_1, p_2, p_3, \ldots$. Это — средняя длина наиболее экономичных кодовых комбинаций, с помощью которых можно описать все исходы для событий, происходящих с вероятностями $p_1, p_2, \ldots$. Увеличение информации является другим важным понятием теории информации. Если наблюдение случайной величины (или события) изменяет вероятностное распределение с $p_1, p_2, p_3, \ldots$ на $q_1, q_2, q_3, \ldots$, то количество полученной информации равно

$$\sum_i q_i \log \frac{q_i}{p_i}.$$

Пусть теперь неизвестный параметр θ некоторого вероятностного распределения является случайной величиной (в соответствии с байесовским подходом в математической статистике). Для простоты предположим, что θ может принимать лишь конечное или счетное число значений с вероятностями $p_1, p_2, p_3, \ldots$. Таким образом, энтропия величины θ равна $H(\theta) = H(p_1, p_2, p_3, \ldots)$. Далее, предположим, что случайная выборка $X = (X_1, X_2, \ldots, X_n)$ также может состоять лишь из конечного или счетного числа различных значений с положительными вероятностями $q_1, q_2, q_3, \ldots$. Наконец, пусть r_{jk} обозначает вероятность того, что θ принимает j-е значение (с вероятностью p_j) и в то же время X принимает k-е значение (с вероятностью q_k). Тогда количество информации относительно θ, полученное в результате наблюдения X, равно

$$I(X, \theta) = \sum_{j,k} r_{jk} \log \frac{r_{jk}}{p_j q_k}.$$

Функция $f(X) = f(X_1, X_2, \ldots, X_n)$ от выборки X называется достаточной, если $I(f(X), \theta) = I(X, \theta)$, т. е. если $f(X)$ содержит столько же информации о θ, как исходная выборка X. Если функция f, вообще говоря, не является достаточной, то отношение $I(f(X), \theta)/I(X, \theta)$ показывает, во сколько раз изменяется количество информации о θ, если вместо полной выборки взять $f(X)$. Свойство, заключающееся в том, что при проведении все большего числа наблюдений можно в конце концов получить всю информацию о θ, на языке теории информации выражается следующим образом. Если наблюдения $X_1, X_2, \ldots$ являются независимыми одинаково распределенными случайными величинами, распределение которых F_θ различно для разных значений параметра θ, то $\lim\limits_{n\to\infty} I((X_1, X_2, \ldots, X_n), \theta) = H(\theta)$. Парадокс *А. Реньи*, обсуждаемый ниже, возник из применения теории информации для проверки гипотез.

б) Парадокс

Наблюдая случайную величину X, которая связана с событием A, мы хотели бы отгадать, произошло событие A или нет. Если событие A происходит с вероятностью $P(A) = p$, то количество информации, содержащееся в A, равно $H(p, 1-p)$. После наблюдения величины X мера оставшейся неопределенности запишется в виде $\bar{H}_X = E(H(P(A|X), 1 - P(A|X)))$, где $P(A|X)$ означает условную вероятность события A относительно X. Следовательно, количество информации относительно A при условии, что наблюдалась величина X, равно

$$I(X, A) = H(p, 1-p) - \bar{H}_X.$$

Наблюдая X, положим $d(X) = 1$, если мы решили, что A произошло, и $d(X) = 0$ в случае, если произошло событие, противоположное A, т. е. $\bar{A}$. Вероятность неверного решения (ошибки) равна

$$\delta = pP(d(X) = 0 \mid A) + (1-p)P(d(X) = 1 \mid \bar{A}).$$

Легко доказать (например, с помощью фундаментальной леммы Неймана — Пирсона, см. II/10), что никакое решение не может иметь ошибку δ меньше, чем следующее «стандартное решение»:

$$d_0(X) = \begin{cases} 1, & \text{если } P(A|X) > P(\bar{A}|X), \\ 0, & \text{если } P(A|X) < P(\bar{A}|X). \end{cases}$$

Если $P(A|X) = P(\bar{A}|X)$, то положим $d_0(X) = 1$ с вероятностью p и 0 с вероятностью $1 - p$. Парадокс, возникающий здесь, состоит в следующем. Пусть $Y = d_0(X)$. Тогда информация относительно A, содержащаяся в Y, определяется формулой

$$I(Y, A) = H(p, 1-p) - \overline{H}_Y.$$

Но Y есть функция от X, поэтому $I(Y, A) \leqslant I(X, A)$. Равенство имеет место тогда и только тогда, когда $P(A|X)$ может принимать только два различных значения, т. е. вообще говоря, X содержит об A информации больше, чем Y. Тем не менее, зная X, мы не можем выбрать решение относительно A лучше, чем если бы мы знали лишь $Y = d_0(X)$. Отсюда следует, что, хотя в X содержится, вообще говоря, больше информации об A, чем в Y, однако, использовать эту дополнительную информацию невозможно.

в) *Объяснение парадокса*

Дополнительную информацию можно использовать, наблюдая другую случайную величину. Пусть, например, $Z = X + U$, где U — индикатор события A. Это означает, что $U = 1$, если событие A произошло, и 0 в противном случае. Очевидно, наблюдая одновременно X и Z, мы получим полную информацию относительно A, т. е. дополнительная информация относительно A, скрытая в X, может быть использована в результате наблюдения вспомогательной величины Z.

г) *Замечания*

Теория информации тесно связана с рядом практических задач, например, с поиском оптимальных способов дистанционной передачи данных или с основаниями биологии (см. ссылки ниже).

д) *Литература*

Ash. B. B. *Information Theory*, Wiley, New York, 1966.

Brioullin L. *Science and Information Theory*, New York, Academic Press, 1956. [Имеется перевод: Бриллюэн Л. Наука и теория информации.— М.: Физматгиз, 1967].

Kullback S. *Information Theory and Statistics, Wiley, New York*, 1959. [Имеется перевод: Кульбак С. Теория информации и статистика. М.: Наука, 1967].

Quaster H. (ed.). *Information Theory in Biology*, Univ. Illinois Press, Urbana, 1953.

Rényi A. *Notes on Information Theory*, Akadémiai Kiadó, Budapest, 1984.

Shannon C., Weaver W, *The Mathematical Theory of Communication*, Univ. Illinois Press, Urbana. 1949.

Turán P. (ed.). *Selected Papers of Alfréd Rényi*, Akadémiai Kiadó, Budapest, Vol. III, 442, 1976.

Yaglom A. M., Yaglom I. M. Hinčin Ya. A. *Mathematical Foundation of Information Theory* (in Hungarian) Müszaki Könyvkiadó, Budapest, 1959.

12. Парадокс t-критерия Стьюдента

а) История парадокса

В классической теории математической статистики предполагается, что элементы выборки (наблюдения) заранее известны. В основе одного из важнейших направлений современной статистики лежит понимание того, что не нужно фиксировать заранее объем выборки, его следует определять в зависимости от результатов более ранних наблюдений. Таким образом, объем выборки случаен. Эта идея последовательного выбора постепенно развивалась в работах *Г. Доджа* и *Г. Ромига* (1929 г.), *П. Махаланобиса* (1940 г.), *Г. Хотеллинга* (1941 г.) и *У. Бертки* (1943 г.), но настоящим основателем теории последовательного анализа в математической статистике является *А. Вальд* (1902—1950 гг.). Его последовательный критерий отношения правдоподобия (1943 г.) стал важным открытием, позволившим (в типичных ситуациях) на 50 % уменьшить среднее число наблюдений (при тех же вероятностях ошибок). Неудивительно, что в годы второй мировой войны открытие Вальда было объявлено «секретным». Его основная книга «Последовательный анализ» опубликована лишь в 1947 г. Год спустя Вальд и *Дж. Волфовиц* доказали, что методы, отличные от последовательного критерия отношения правдоподобия, не дают такого уменьшения числа элементов выборки. Но и в этой области обнаружились парадоксы. Здесь мы обсудим парадокс, принадлежащий *К. Стейну*, хотя этот парадокс относится к двухшаговым критериям, а не к последовательным.

б) Парадокс

Пусть $X_1, X_2, \ldots, X_n$ — выборка из независимых нормально распределенных случайных величин с общим неизвестным математическим ожиданием θ и общим неизвестным стандартным отклонением σ. На основе этой выборки мы хотим различить следующие нулевую и альтернативную гипотезы. Нулевая гипотеза состоит в том, что $\theta = \theta_0$ (где θ_0 — некоторое заданное число), а альтернативная — в том, что $\theta \neq \theta_0$. Пусть

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad D^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

и

$$t_n = \frac{\bar{X} - \theta_0}{D^*/\sqrt{n}}.$$

Такие гипотезы различаются с помощью t-критерия Стьюдента. Согласно t-критерию нулевая гипотеза принимается или отвергается в зависимости от того, близко значение статистики t_n к 0 или нет. В 1940 г. *Г. Данциг* показал, что при заданной вероятности ошибки 1-го рода вероятность ошибки 2-го рода для любого решающего правила зависит от неизвестного стандартного отклонения σ. Парадоксально, но пятью годами позже К. Стейн доказал, что если объем выборки n не фиксировать заранее, а определять по уже полученным элементам выборки (как в последовательном анализе Вальда), то существует t-критерий, для которого (при заданной вероятности ошибки 1-го рода) вероятность ошибки 2-го рода не зависит от неизвестного стандартного отклонения σ (а зависит лишь от разности $\theta-\theta_0$).

в) Объяснение парадокса

На первом шаге возьмем выборку $X_1, X_2, \ldots, X_n$, где n_0 — некоторое фиксированное число. Выборочная дисперсия определяется формулой

$$s^2=\frac{1}{n_0-1}\left\{\sum_{i=1}^{n_0} X_i^2-\frac{1}{n_0}\left(\sum_{i=1}^{n_0} X_i\right)^2\right\}.$$

Предположим, что объем всей выборки n зависит от величины s и заранее фиксированного числа z следующим образом:

$$n=\max\left\{\left[\frac{s^2}{z}\right]+1,\ n_0+1\right\},$$

где скобки [ ] означают целую часть действительного числа. Выберем положительные числа $a_1, a_2, \ldots, a_n$ так, что $\sum\limits_{i=1}^{n} a_i=1$, $a_1=a_2=\ldots=a_{n_0}$ и $s^2\sum\limits_{i=1}^{n} a_i^2=z$, и попытаемся различить эти гипотезы с помощью статистики

$$t^*=\frac{\sum\limits_{i=1}^{n} a_i X_i-\theta_0}{\sqrt{z}}=t+\frac{\theta-\theta_0}{\sqrt{z}},$$

где

$$t=\frac{\sum\limits_{i=1}^{n} a_i\left(X_i-\theta\right)}{\sqrt{z}}.$$

Очевидно, при заданном s случайная величина t нормально распределена с математическим ожиданием 0 и дисперсией $\sigma^2\sum\limits_{i=1}^{n} a_i^2/z=\sigma^2/s^2$. С другой стороны, распределение величины

$(n_0-1)s^2/\sigma^2$ (для произвольного σ) совпадает с распределением суммы квадратов n_0-1 независимых стандартных нормальных случайных величин (т. е. с хи-квадрат распределением $\chi^2_{n_0-1}$), не зависящим от σ. Следовательно, распределение величины t также не зависит от σ, так что t^* зависит лишь от $\theta-\theta_0$, но не от σ.

г) Замечания

(i) Распределение случайной величины t_n не является нормальным, так как D^* не число, а случайная величина. (Если бы значение стандартного отклонения было известно, и мы подставили бы это значение вместо D^*, то распределение случайной величины t_n стало бы стандартным нормальным.) Это замечательное наблюдение и анализ случайной величины t_n в 1908 г. опубликовал *Стьюдент*, он же *Уильям Д. Госсет*. (С 1899 г. он работал в Дублине на пивоваренном заводе Гиннесса, и его начальник настоял на том, чтобы Госсет писал под псевдонимом.) В течение долгого времени никто не осознавал важности статьи Стьюдента. (Даже в 1922 г. *Р. Фишер* был единственным, как утверждал Стьюдент, кто использовал t-распределение. В действительности, именно Фишер впервые обозначил распределение Стьюдента через t в своей книге, вышедшей в 1925 г. Сам Стьюдент использовал символ z, но не для обозначения величины t_n, а для $(n-1)t_n$.)

(ii) Определение момента прекращения наблюдений в последовательном анализе является сутью современной теории оптимальных остановок для различных процессов. Рассматривая выборку как процесс, мы тем самым устанавливаем связь между математической статистикой и теорией стохастических процессов, которая будет обсуждаться в следующей главе. Эта взаимосвязь оказывается полезной для обоих направлений. В наши дни фундаментальные теоремы Вальда из последовательного анализа являются частным случаем общей теории стохастических процессов с остановкой (см. книгу Ширяева А. Н.).

д) Литература

Chow Y., Robbins H., Siegmund D. *Great Expectations: Theory of Optimal Stopping,* Houghton Miffin Co., Boston, 1971. [Имеется перевод: Роббинс Г., Сигмунд Д., Чао И. Теория оптимальных правил остановки. — М.: Наука, 1977.]

Fisher R. A. *Statistical Methods of Research Workers*, Oliver and Boyd, Edinburgh, 1925. [Имеется перевод: Фишер Р. А. Статистические методы для исследователей. — М.: Госстатиздат, 1958.]

Ширяев А. Н. Статистический последовательный анализ. — М.: Наука, 1969.

Stein C. "A two sample test for a linear hypothesis whose power is independent of the variance", *Annals of Math. Statist.*, **16**, 243—258, (1945).

Student "The probable error of a mean", *Biometrika*, 6, 1—24, (1908).
Wald A. "Sequential analysis of statistical data: Theory, *Restricted Report*, Sept., 1943.
Wald A. *Sequential Analysis*, Wiley, New York, 1947. [Имеется перевод: Вальд А. Последовательный анализ. М.: Физматгиз, 1960.]

13. Еще несколько парадоксов

а) Парадокс типичного и среднего

Понятие среднего, например средняя зарплата, часто используется как синоним типичного. На самом деле если в некоторой стране есть всего лишь несколько очень богатых семей и большое количество бедных, чьи доходы соответственно огромны и малы, то арифметическое среднее их доходов вовсе нетипично. Например, медиана доходов дает более реалистичную картину. (Медиана означает такой доход, что число людей с большим доходом, равно числу людей, имеющих меньший доход.) Кроме средней зарплаты есть и другие средние характеристики, вводящие в заблуждение. «Средний человек» (*l'homme moyen*) — одна из них. Неудивительно, что исследования бельгийского ученого *Л. А. Ж. Кетле* по этому вопросу стали источником горячих споров. Худшее в «среднем человеке» не его серость, а возникающие противоречия. Например, средний рост не соответствует среднему весу и т. д. Только по одной этой причине можно усомниться в справедливости слов Дж. Рейнольдса (первого президента Королевской академии художеств в Лондоне), когда он сказал, что в среднем источник прекрасного.

(Лит.: Quételet L. A. J. *Essai de Physique sociale*, (1835); *L'homme moyen*, Physique Sociale, Vol. 2, Bruxelles, 1869).

Несмотря на свою непоследовательность, книга Кетле (1835) рассматривается как веха, если вообще не как начало количественного анализа общественных явлений. *Ф. Гальтон*, *К. Пирсон* и *Ф. Эджворт* ценили Кетле как гениального первооткрывателя регрессионного подхода. Под влиянием его книги Гальтон занялся научными исследованиями. Однако у Кетле есть и другие заслуги перед наукой. В 1820 г. он основал Королевскую обсерваторию Бельгии и стал ее первым директором. Он был также великолепным организатором: в 1834 г. по его предложению было создано Статистическое общество в Лондоне, он был инициатором проведения Первого международного конгресса по статистике в Брюсселе в 1853 г.

б) Парадокс оценивания

Квадрат оценки, вообще говоря, не совпадает с оценкой квадрата. Например, если $\bar{X}$, т. е. среднее значение результатов наблюдений $X_1, X_2, \ldots, X_n$, является оценкой некоторого параметра, то очевидной оценкой для квадрата параметра будет $\bar{X}^2$, что, вообще говоря, отличается от арифметического среднего квадратов результатов наблюдений. Это же справедливо, если заменить квадрат любой другой нелинейной функцией.

(Лит.: Carnap R. *Logical Foundation of Probability*, Routledge and Kegan Paul Ltd. Broadway House, London, 1950.)

в) Парадокс точности измерения

Предположим, что нам нужно найти длину двух различных стержней с помощью двух измерений. Прибор, которым мы можем измерять длину, дает результат со случайной ошибкой, имеющей стандартное отклонение σ. Парадоксально, но измерение каждого стержня в отдельности не является лучшим способом. Стандартное отклонение результата будет меньше, если сначала измерить общую длину T стержней, приложив конец одного стержня к концу другого, а затем положить стержни рядом и найти разницу их длин D. Тогда приближенные длины стержней соответственно равны

$$\frac{T+D}{2} \text{ и } \frac{T-D}{2}.$$

Стандартное отклонение этих длин равно $\sigma/\sqrt{2}$, что действительно меньше, чем σ.

(Лит.: Hotelling H. "Some improvements in weighing and other experimental techniques", Annals of Math. Statist., 15, 297—306, (1944).)

г) Парадоксальное оценивание вероятности

Оценкой для неизвестной вероятности обычно служит относительная частота. Например, если при ста бросаниях монеты решка выпала 47 раз, то оценкой для вероятности выпадения решки будет 47/100. Однако, если при 10 бросаниях более или менее правильной монеты решка ни разу не появилась, то нет оснований считать вероятность выпадения решки равной 0. При наличии некоторой *априорной* информации (например, что монета более или менее правильная) оценивание через относительную частоту, вообще говоря, не является лучшим способом. Наша *априорная* информация хорошо выражается через бета-распределение, зависящее от двух параметров a и b. Плотность вероятности бета-распределения равна 0 вне интервала (0, 1) и пропорциональна $x^{a-1}(1-x)^{b-1}$ на (0, 1); ($a>0$, $b>0$). Математическое ожидание и дисперсия бета-распределения соответственно равны

$$m=\frac{a}{a+b} \text{ и } d=\frac{ab}{(a+b)^2(a+b+1)}.$$

Решая эту систему уравнений, получаем, что наша *априорная* информация относительно m и d может быть выражена через a и b (например, если монета правильная, то $m = 1/2$, следовательно, $a = b$). Если *априорное* распределение является бета-распределением с параметрами (a, b), то по теореме Байеса *апостериорное* распределение также будет бета-распределением. (Это свойство объясняет широкую применимость бета-распределения.) Если в n экспериментах событие, имеющее неизвестную вероятность, произошло k раз, то параметрами *апостериорного* распределения будут $(a+k,\ b+n-k)$, следовательно, *апостериорное* математическое ожидание запишется в виде

$$M = \frac{a+k}{a+b+n},$$

что дает более содержательную и лучшую оценку для неизвестной вероятности, чем относительная частота k/n. Очевидно, при достаточно больших n величина M практически не отличается от относительной частоты, однако, например, когда $n = 10$, $k = 0$ и $a = b = 100$, имеем $M = 100/210 \approx 0.48$; в то же время относительная частота равна 0, что совершенно нелепо.

(Лит.: Good I. J. The Estimation of Probability, MIT Press, Cambridge, 1965.)

д) Чем больше данных, тем хуже выводы

Довольно естественно считать, что большее количество данных позволяет получать лучшие результаты. Однако следующий парадокс, кажется, показывает противоположное. Пусть X_1, X_2 и X_3 обозначают независимые случайные величины, и предположим, что распределения величин X_1 и X_2 совпадают: и X_1 и X_2 равны либо 0, либо 2 с равными вероятностями, следовательно, математические ожидания у обеих случайных величин одинаковы и равны 1. Пусть случайная величина X_3 принимает значения 1 и 2.5 с равными вероятностями, так что ее математическое ожидание равно 1.75. Вся эта информация неизвестна математику, который сделал выборки из этих распределений, чтобы найти распределение с наибольшим математическим ожиданием. Очевидно, в качестве искомого распределения нужно взять распределение с наибольшим выборочным средним. Сначала возьмем выборку из одного элемента из каждого распределения. Тогда вероятность правильного выбора равна

$$P(X_1 < X_3 \text{ и } X_2 < X_3) =$$
$$P(X_3 = 2.5) + P(X_3 = 1)\,P(X_1 = 0)\,P(X_2 = 0) = 5/8.$$

Посмотрим теперь, что произойдет, если увеличить объем одной из выборок (например, для X_3) до 2 (остальные выборки оставить без изменений)? Тогда вероятность правильного выбора равна

$$P(X_1 < \bar{X}_3 \text{ и } X_2 < \bar{X}_3) = 7/16,$$

где $\bar{X}_3$ — арифметическое среднее двух элементов выборки. Таким образом, если раньше вероятность правильного выбора была больше 50 %, то теперь она уменьшилась и составляет менее 50 %.

(Лит.: Chius W. K., Lam K. The American Statistician, 1975.)

Знаменитый парадокс *Ф. Эджворта* (1883 г.) касается аналогичной проблемы: если X_1 и X_2 являются независимыми случайными величинами с одинаковыми плотностями вероятности $f(x-\theta)$, симметричными относительно θ, то может так случиться, что X_1 ближе к θ, чем $\bar{X} = (X_1 + X_2)/2$, в том смысле, что

$$P(|X_1 - \theta| \leqslant \varepsilon) > P(|\bar{X} - \theta| \leqslant \varepsilon)$$

для любого положительного ε. Например, так будет в случае, когда

$$f(x) = \frac{3}{2} \cdot \frac{1}{(1+|x|)^4},$$

потому что плотность вероятности случайной величины X_1 в точке θ больше, чем плотность величины $\bar{X}$.

(Лит.: Stigler S. M. "An Edgeworth curiosum", *Annals of Statist.*, **8**, 931—934, (1980).)

е) *Парадокс равенства математических ожиданий*

Предположим, что математические ожидания трех нормально распределенных случайных величин с одинаковыми дисперсиями равны m_1, m_2 и m_3. Может так случиться, что используя, например, t-критерий Стьюдента, мы примем гипотезы $m_1 = m_2$ и $m_2 = m_3$ (при некотором уровне значимости), но отвергнем $m_1 = m_3$! (Задача проверки гипотезы о равенстве математических ожиданий при наличии двух выборок объема n решается с помощью статистики

$$t = \frac{\bar{X} - \bar{Y}}{2D^*},$$

где $\bar{X}$ и $\bar{Y}$ — выборочные средние, а D^* — выборочное стандартное отклонение. Статистика t имеет распределение Стьюдента с параметром $2n-2$.) Этот парадокс положил начало многим исследованиям по одновременной проверке нескольких гипотез (дисперсионный анализ и т. д.).

(Лит.: Hodges J. L., Lehmann E. L. "The efficiency of some nonparametric competitors of the t-est", Annals of Math. Statist., **27**, 324—335, (1956).
Scheffé H. The Analysis of Variance, Wiley, New York, 1959. [Имеется русский перевод: Шеффе Г. Дисперсионный анализ. — М.: ГИФМЛ, 1963.])

ж) Парадоксальная оценка для математического ожидания нормального распределения

Мы хотим, используя выборку объема n, оценить неизвестное математическое ожидание (одномерного) нормального распределения с единичным стандартным отклонением. Известно, что арифметическое среднее $\bar{X}$ выборки дает оценку, обладающую многими хорошими свойствами. Например, эта оценка имеет наименьшую дисперсию, является несмещенной, допустимой и минимаксной при квадратичной функции потерь. Однако, несмотря на эти свойства, если нам нужно получить лишь оценку, которая близка к θ насколько это возможно, то существует оценка $\hat{\theta}$, лучшая в том смысле, что

$$P(|\hat{\theta}-\theta|<|\bar{X}-\theta|)>1/2$$

для любого возможного значения θ. Оценкой такого типа является следующая

$$\hat{\theta}=\bar{X}-\frac{1}{2\sqrt{n}}\min(\sqrt{n}\,\bar{X},\ \Phi(-\sqrt{n}\,\bar{X})),$$

если $\bar{X}\geqslant 0$, и

$$\hat{\theta}=\bar{X}+\frac{1}{2\sqrt{n}}\min(\sqrt{n}\,\bar{X},\ \Phi(-\sqrt{n}\,\bar{X})),$$

если $\bar{X}\leqslant 0$, где Φ обозначает функцию стандартного нормального распределения.

з) Парадокс проверки нормальности

Мы хотим проверить гипотезу о том, что данная выборка $X_1, X_2, \ldots, X_n$ получена из распределения с непрерывной функцией распределения $F(x)$. Пусть

$$F_n(x)=\frac{1}{n}\sum_{i:X_i<x}1$$

обозначает эмпирическую (кумулятивную) функцию распределения выборки. Если гипотеза верна, то по теореме Колмогорова

$$\lim_{n\to\infty} P\left(\sqrt{n}\,\sup|F_n(x)-F(x)|<z\right)=\sum_{j=-\infty}^{+\infty}(-1)^j e^{-j^2z^2}=K(z).$$

Используя этот результат, легко построить критерий с уровнем доверия α. (Если величина $\sqrt{n}\,\sup|F_n(x)-F(x)|$ больше критического значения z_0, для которого $K(z_0)=\alpha$, то гипотеза отвергается.) В случае, когда нужно проверить нормальность вероятностного распределения, сначала следует по выборке обычным образом оценить математическое ожидание и стандартное отклонение через $\bar{X}$ и D^*, затем применить указанный выше критерий Колмогорова для нормального распределения $F(x)$ с математическим ожиданием $\bar{X}$ и стандартным отклонением D^*. Можно было бы ожидать, что при достаточно больших n подстановка вместо неизвестных параметров величин $\bar{X}$ и D^* не повлечет за собой значительных изменений. Однако изменения произойдут существенные. Например, при 95 % уровне доверия критическое значение z_0 в критерии Колмогорова равно 1.36, в то же время точный анализ показывает, что истинное критическое значение равняется 0.9. Объяснение этого парадокса достаточно просто. Из-за подстановки функция $F(x)$ и эмпирическая функция $F_n(x)$ стали ближе друг к другу, поэтому целесообразно выбрать меньшее критическое значение.

(Лит.: Durbin S. "Some methods of constructing exact tests", *Biometrika*, **48**, 41—45, (1961).

Kac M., Kiefer J., Wofrowitz J. "On tests of normality and other tests of goodness of fit based on distance methods", *Annals of Math. Statist.*, **26**, 189—211, (1955).

Sarkadi K. "On testing for normality", *Proc. 5th Berkeley Symp. on Math. Statist. and Prob.*, I. 373—387, (1967).)

и) Парадокс линейной регрессии

Предположим, что случайная величина X измеряется с ошибкой ε, имеющей математическое ожидание 0. Другими словами, результат измерения $Y=X+\varepsilon$ является очень простой линейной регрессией по X. Есть ли для X «лучшая оценка», чем измеренная величина Y? Удивительно, но в некоторых частных случаях ответ оказывается утвердительным. По крайней мере существует оценка $\hat{X}$, для которой $E(\hat{X}-X)^2$ меньше, чем $E(Y-X)^2=E\varepsilon^2$. Предположим, например, что X и ε некоррелированы и регрессия X по Y также линейна. Тогда оценка

$$\hat{X}=\frac{D^2(\varepsilon)}{D^2(Y)}\cdot E(X)+\frac{D^2(X)}{D^2(Y)}\cdot Y$$

лучше, чем Y. (В вырожденном случае, когда $D^2(\varepsilon) = 0$, получаем $\hat{X} = Y$.)

к) Парадокс Сетурамана

Существуют статистические функции A и B такие, что несмещенная оценка неизвестного параметра θ, построенная по A, имеет дисперсию меньшую, чем оценка, построенная по B (при любом действительном значении параметра θ). С другой стороны, при проверке нулевой гипотезы $\theta = \theta_0$ (например, при альтернативной гипотезе $\theta > \theta_0$) критерий на базе функции A не обязательно будет лучше критерия, построенного по B. Последний может быть локально лучше (в некоторой окрестности нулевой гипотезы). Если, например, элементы выборки $X_1, X_2, \ldots, X_n$ равномерно распределены на интервале $(\theta; 2\theta)$, то оценка максимального правдоподобия для θ запишется в виде

$$U = \frac{1}{2} \max(X_1, X_2, \ldots, X_n),$$

и ее небольшое изменение приводит к несмещенной оценке

$$B = \frac{2n+2}{2n+1} U.$$

Следующая оценка A также является несмещенной, но имеет меньшую дисперсию

$$A = \frac{n+1}{5n+4}(4U + V), \text{ где } V = \min(X_1, X_2, \ldots, X_n).$$

Однако при проверке гипотезы $\theta = \theta_0$ критерий, основанный на B, является локально более мощным.

(Лит.: Sethuraman J. "Conflicting criteria of 'goodness' of statistics" Sankhya, 23, 187—190, (1961).)

л) Парадокс минимаксной оценки

Понятие минимаксной оценки было введено в замечании (ii) к парадоксу I/12. Минимаксные оценки, как правило, согласуются со здравым смыслом. Однако следующий пример *Х. Рубина* показывает обратное. Если функция потерь задается равенством $L(p, c) = \min((p - c)^2/p^2; 2)$, то единственной минимаксной оценкой неизвестной вероятности $p \neq 0$ является оценка, тождественно равная 0. Поэтому для любой выборки неизвестный параметр следует оценивать нулем (значением, которое заранее исключено из возможных значений параметра p).

Замечание. Для функции потерь $L(p, c) = (p - c)^2$ минимаксная оценка запишется в виде

$$p^* = \frac{\varkappa + \sqrt{n/2}}{n + \sqrt{n}},$$

где n — объем выборки, $\varkappa$ — частота события с неизвестной вероятностью.

м) Парадокс Роббинса

Хорошо известно, что при наличии выборки из одного наблюдения X «лучшей» оценкой параметра пуассоновского распределения является само X. (Это несмещенная оценка максимального правдоподобия с минимальной дисперсией.) Но как оценить параметры $\theta_1, \theta_2, \ldots, \theta_k$ для k независимых пуассоновских распределений на основе соответствующих наблюдений $X_1, X_2, \ldots, X_k$, если мы хотим, чтобы величина

$$E\left(\sum_{i=1}^{k} (\hat{\theta}_i - \theta_i)^2\right)$$

была минимальной? Существует ли оценка лучшая, чем $\hat{\theta}_i = X_i$? *Х. Роббинс* первым указал на то что, хотя k пуассоновских распределений независимы, все же можно найти лучшие оценки, учитывающие не только «свои» наблюдения, но и другие. Если k велико, и $N(X)$ обозначает число наблюдений, равных X, то в силу результата Роббинса оценка $\hat{\theta}_i = (X_i + 1) N(X_i + 1) / N(X_i)$ лучше, чем $\hat{\theta}_i = X_i$. Суть парадокса заключается в следующем: наблюдения, которые не связаны с оцениваемым параметром, могут влиять на качество его оценки (сравните с парадоксом II/2 (ii)).

(Лит.: Robbins H. "An empirical Bayes'approach to statistics", *Proc. 3rd Berkeley. Symp. on Math. Statist. and Prob.*, I, 157—164, 1956.)

н) Парадокс байесовской модели

Пусть плотность вероятности $f_p(x)$ случайной величины X является смесью двух положительных плотностей вероятности $f_0(x)$ и $f_1(x)$:

$$f_p(x) = p f_0(x) + (1 - p) f_1(x), \text{ где } 0 \leqslant p \leqslant 1.$$

Значение параметра p неизвестно, и при достаточно большом n мы надеемся определить его с любой заданной точностью по независимым наблюдениям $X_1, X_2, \ldots, X_n$ (распределения величин X_i и X совпадают). Для решения задачи мы хотим воспользоваться теоремой Байеса: выберем число p_0, $0 < p_0 < 1$, и предположим, что *априорная* плотность вероятности величины X

равна $p_0 f_0(x) + (1 - p_0) f_1(x)$. Тогда *апостериорная* плотность вероятности величины X (при наличии выборки $X_1, X_2, \ldots, X_n$) запишется в виде $p_n f_0(x) + (1 - p_n) f_1(x)$, где

$$\frac{p_n}{1-p_n} = \frac{p_0}{1-p_0} \prod_{i=1}^{n} \frac{f_0(X_i)}{f_1(X_i)}.$$

Используя элементы выборки, действительно можно найти значение параметра p с любой заданной точностью, если

$$\lim_{n\to\infty} \frac{p_n}{1-p_n} = \frac{p}{1-p} \text{ (с вероятностью 1).}$$

Однако это равенство не всегда выполнено. Например, если математическое ожидание величины $\log \frac{f_0(X)}{f_1(X)}$ равно 0, то по теореме Чжуна — Фукса (сравните с III/7)

$$\lim_{n\to\infty} \sup \frac{p_n}{1-p_n} = \infty \quad \text{и} \quad \lim_{n\to\infty} \inf \frac{p_n}{1-p_n} = 0,$$

следовательно, $\lim \frac{p_n}{1-p_n}$ в этом случае даже не существует. Парадокса не будет, если вместо $(p_0, 1 - p_0)$ взять *априорное* распределение, имеющее плотность вероятности, положительную на всем интервале $0 < p < 1$. Такая модель лучше, потому что в ней истинная плотность f_p учитывается с положительной плотностью.

(Лит.: Berk R. H. "Limiting behavior of the posterior distributions when the model is incorrect", Annals of Math. Statist., 37, 51—58 (1966).)

о) Парадокс доверительных интервалов

Пусть $X_1, X_2, X_3, \ldots$ — нормально распределенные случайные величины с общим математическим ожиданием m и единичной дисперсией, и через S_n обозначим следующую сумму

$$S_n = X_1 + X_2 + \ldots + X_n.$$

Вероятность того, что для *любого фиксированного* n

$$\frac{S_n - 2\sqrt{n}}{n} < m < \frac{S_n + 2\sqrt{n}}{n}$$

приблизительно равна 95 %, и в то же время вероятность того, что эти неравенства выполнены для *всех* n, равна 0. Последняя вероятность остается нулевой, если даже заменить 2 на как угодно большое число. (См. Robbins H. «Statistical methods related to the law of the iterated logarithm», *Annals of Math. Statist.*, 41, 1397—1409, (1970).)

п) Парадокс проверки независимости; являются ли эффективные лекарства эффективными?

Ниже приведены три таблицы, в которых показано действие некоторого лекарства только на мужчин, только на женщин и, наконец, на больных обоего пола (объединенные результаты).

МУЖЧИНЫ

	Принимавшие лекарство	*Не принимавшие лекарство*
Выздоровевшие	700	80
Не выздоровевшие	800	130

ЖЕНЩИНЫ

	Принимавшие лекарство	*Не принимавшие лекарство*
Выздоровевшие	150	400
Не выздоровевшие	70	280

ОБЪЕДИНЕННЫЕ РЕЗУЛЬТАТЫ

	Принимавшие лекарство	*Не принимавшие лекарство*
Выздоровевшие	850	480
Не выздоровевшие	870	410

Рис. 11.

Из таблиц видно, что после приема лекарства доля выздоровевших больше как среди мужчин, так и среди женщин. (Существенность различий может быть обоснована статистически с помощью критериев независимости.) С другой стороны, как это ни странно, из таблицы с объединенными результатами следует, что доля выздоровевших больше среди тех людей, которые лекарство не принимали. Следовательно, лекарство, показавшее свою эффективность как среди мужчин, так и среди женщин, дало отрицательный результат для смешанной группы мужчин и женщин. Аналогично, новое лекарство может оказаться эффективным в каждом из десяти различных госпиталей, но объединение результатов укажет на то, что это лекарство либо бесполезно, либо вредно.

(Лит.: Pflug G. "Paradoxien der Wahrscheinlichkeitstrechnung", in: Stochastik im Schulunterricht, Wein, Teubner, 155—163, 1981.)

p) Парадокс компьютерной статистики

Начиная с 50-х годов, в связи с появлением компьютеров лицо статистики изменилось. Раньше отсутствие компьютеров вынуждало ученых использовать слишком упрощенные модели, даже если эти модели были нереалистичными. Однако за последние тридцать лет любое статистическое решение, которое можно просчитать на компьютере за сравнительно небольшой период времени, стало «доступным». Таким образом, «устойчивые» («робастные») и многомерные методы с огромным числом операций вошли в повседневную статистическую практику. В это же время статистика, по крайней мере частично, стала эмпирической наукой: за несколько минут компьютеры могут порождать миллионы данных, и, используя их, можно «проверять» новейшие методы. Многие «эмпирические теоремы» вошли в практику без достаточных теоретических обоснований. С другой стороны теория робастных статистик (см., например, Huber P. J. *Robust Statistics*, *Wiley*, *New York*, 1981. [Имеется перевод: Хьюбер П. Робастность в статистике. — М.: Мир, 1984.]) дает теоретическое обоснование для многих эмпирических «хитрых трюков», используемых в статистической практике. О противоречиях и парадоксах нового периода можно прочитать в приведенных ниже статьях.

(Лит.: Efron B. "Bootstrap methods: another look at the jackknife", *Annals of Statist.*, 6, 1—26, (1979).
Efron B. "Computers and the theory of statistics: Thinking the unthinkable", *SIAM Review*, 1979 Okt.
Hampel F. R. "Robust estimation: A condensed partial survey", *Zeitsch, Wahrsch, theorie vrw. Geb.*, 27, 87—104, (1973).
Miller R. G. "The jackknife — a review", Biometrika, 61, 1—15, (1974).
Tukey J. W. "The future of data analysis", *Annals of Math. Statist.*, 33, 1—67, (1962).

ГЛАВА III

ПАРАДОКСЫ СЛУЧАЙНЫХ ПРОЦЕССОВ

Как же случилось то, что стеченье материи дало
Землю и своды небес, а также и моря глубины,
Солнца пути и луны, — разъясню я теперь по порядку
Первоначала вещей, разумеется, вовсе невольно
Все остроумно в таком разместилися стройном порядке
И о движеньях своих не условились раньше, конечно.
Если ж начала вещей во множестве, многоразлично
От бесконечных времен постоянным толчкам подвергаясь
Тяжестью также своей гнетомые, носятся вечно,
Всячески между собой сочетаясь и все испытуя,
Что только могут они породить из своих столкновений, —
То и случается тут, что они в этом странствии вечном,
Всякие виды пройдя сочетаний и разных движений,
Сходятся так, наконец, что взаимная их совокупность
Часто великих вещей собой образует зачатки:
Моря, земли и небес, и племени тварей живущих.

Лукреций, О природе вещей, кн. V, с. 416—431 [1])

Первые значительные результаты в теории случайных процессов или стохастических процессов, если воспользоваться словом греческого происхождения, появились лишь в прошлом веке. Благодаря успехам, достигнутым в классической механике, в XVII и XVIII веках главным образом изучались детерминистические процессы. В это же время в науке развился «механистический детерминизм», согласно которому случайность считалась чем-то несущественным и из основных наук ее следовало по возможности исключить. Однако со второй половины прошлого века математические исследования по случайным процессам получают все большее распространение в фундаментальных областях науки, в частности, в различных разделах физики: статистической физике, квантовой физике XX века, где случайные процессы играют главную роль. С углублением научного знания все очевиднее становилась необходимость в изучении стохастических процессов.

1. Парадокс ветвящихся процессов

а) История парадокса

В первой половине прошлого века было замечено следующее интересное явление: некоторые знаменитые аристократические и простые фамилии постепенно исчезали. Эту проблему с математической точки зрения изучали *И. Ж. Бьенеме* в 1845 г. и *де Кондолье* в 1873 г. В 1874 г. *Гальтон* и *Ватсон* опублико-

---

[1]) Лукреций Кар. О природе вещей. Перевод с лат. Ф. А. Петровского.— М.: Гос. изд. «Художественная литература», 1937.

вали важнейшую статью, посвященную этому вопросу. Ветвящиеся цепочки фамилий стали первым примером случайного ветвящегося процесса. Процессы такого типа появляются в химии, физике и некоторых других областях. Например, процесс деления ядер, или цепная реакция, в ядерной физике хорошо моделируется случайными процессами. Поколения нейтронов сменяют друг друга чаще, чем поколения людей, однако в обоих случаях главный вопрос остается одним и тем же: при каких условиях процесс затухнет (фамилия исчезнет) или разовьется до бесконечности (бомба взорвется). Понятие ветвящегося процесса введено в 1947 г. *А. Н. Колмогоровым* и *Н. А. Дмитриевым.*

б) Парадокс

Пусть с вероятностями $p_0, p_1, p_2, \ldots$ взрослый мужчина имеет $0, 1, 2, \ldots$ сыновей. Вычислим вероятность q того, что спустя несколько поколений потомков мужского пола не останется (фамилия исчезнет). Производящая функция вероятностного распределения $p_0, p_1, p_2, \ldots$ определяется формулой

$$g(z)=\sum_{k=0}^{\infty} p_k z^k,$$

где $|z| \leqslant 1$. Обозначим аналогичную производящую функцию n-го поколения через $g_n(z)$ $(g_1(z)=g(z).)$ Легко видеть, что $g_{n+1}(z)=g(g_n(z))$, т. е. эта производящая функция может быть получена путем последовательных итераций функции $g(x)$. Вероятность исчезновения всех потомков мужского пола в n-м поколении равна $q_n=g_n(0)$. Поскольку последовательность q_n монотонно возрастает, предел $\lim q_n=q$ существует, поэтому из равенства $q_{n+1}=g(q_n)$ следует, что $q=g(q)$. Таким образом, вероятность q находится из этого уравнения. Поскольку $q=1$ является решением этого уравнения, Ватсон (ошибочно) предположил, что вероятность вырождения всегда равна 1 и, следовательно, неизбежна. Хотя результат Ватсона совершенно невероятен, лишь в 20-е годы нашего века *Р. Фишер*, *Дж. Халдейн*, *Дж. Стеффенсен* и другие показали, что уравнение имеет и другой корень, который меньше 1, при условии, что среднее число сыновей

$$m=\sum_k kp_k$$

больше 1. В этом случае истинную вероятность вырождения дает меньший корень. С другой стороны, неудивительно, что, когда m меньше 1, вероятность вырождения равна 1. Парадокс может возникнуть лишь в случае $m=1$. Если предположить, что у каждого мужчины в среднем по одному сыну ($m = 1$), то вероятность вырождения все же останется равной 1 (за исключением частного случая $p_1 = 1$). Следовательно, несмотря на то, что среднее число потомков мужского пола во всех поколениях одинаково (всегда равно 1), вырождение неизбежно (точнее, оно происходит с вероятностью 1), хотя можно показать, что среднее время до момента вырождения бесконечно.

в) Объяснение парадокса

Равенства

$$q = \lim_{n\to\infty} q_n = 1 \quad \text{и} \quad m = 1$$

друг другу не противоречат. Первое соотношение означает, что вероятность рождения мальчиков в n-м поколении близка к 0, но если они все-таки есть, то их число может быть бóльшим, поэтому среднее значение может равняться 1.

г) Замечания

Модель Гальтона — Ватсона обычно используется в случае, когда $p_k = ab^{k-1}$ ($k = 1, 2, \ldots$) и $p_0 = 1 - p_1 - p_2 - \ldots$, где a и b — положительные числа и a меньше, чем $1 - b$. Тогда $g_n(z)$ есть просто отношение линейных функций. В 1931 г. *А. Лотка* посчитал значения для a, b и p_0, относящиеся к США. Он получил, что $a = 0.2126$, $b = 0.5893$ и $p_0 = 0.4825$, поэтому вероятность исчезновения мужской линии равна $q = 0.819$. Красивые старые фамилии постепенно исчезают, и их место занимают более заурядные, например, Смит и т. д. Использование комбинаций из двух или трех фамилий не всегда позволяет избежать совпадения фамилий, иногда даже в одном учреждении. Предлагаемый ниже порядок присваивания фамилий представляется необычным, но разумным и не зависящим от пола. Каждый ребенок наследует две фамилии, одну от матери и другую от отца. Поскольку у каждого из родителей тоже по две фамлии, в качестве фамилий для ребенка можно взять более редкие (или более привлекательные). Кроме таких двойных фамилий у людей также будут и имена. При таком порядке мир фамилий стал бы более красочным и индивидуальным.

д) Литература

Bienaymé I. J. "De la loi de multiplication et la durée des familles", *Soc. Philomath*, Paris, 37—39, (1845).

Harris T. E. *The Theory of Branching Processes*, Springer, Berlin — Göttingen — Heildelberg, 1963. [Имеется перевод: Харрис Т. Теория ветвящихся случайных процессов. — М.: Мир, 1966.]

Jagers P. *Branching Processes with Biological Applications*, London, Wiley, 1975.

Lotka A. J. "The extinction of families I—II", *J. Wash. Acad. Sci.*, **21**, 377—380 and 453—549, (1931).
Schrödinger E. "Probability problems in nuclear chemistry", *Proc. Roy. Irish Acad.*, **51**, (1945).
Watson H. W., Galton F. "On the probability of the extinction of families", *J. Antropol.*, Inst. Great Britain and Ireland, **4**, 138—144, (1874).

2. Марковские цепи и физический парадокс

а) История парадокса

Понятие марковской цепи принадлежит русскому математику *А. А. Маркову*, чьи первые статьи по этому вопросу были опубликованы в 1906—1908 гг. (см. список литературы ниже). Марков использовал новое понятие для статистического анализа распределения букв в знаменитой поэме Пушкина «Евгений Онегин». «Цепь Маркова» (это название было предложено А. Я. Хинчиным) — важнейшее математическое понятие, возникшее (по крайней мере частично) при решении лингвистических проблем. Последовательность (цепь) дискретных случайных величин X_1, X_2, ..., X_t, ... образует цепь Маркова (по определению), если для любого момента времени t «поведение» последовательности в будущем (после t) зависит от ее «поведения» в прошлом (до t) лишь через значение величины X_t, т. е. равенство

$$P(X_{t+1}=i_{t+1}\,|\,X_t=i_t,\ X_{t-1}=i_{t-1},\ \ldots)=$$
$$=P(X_{t+1}=i_{t+1}\,|\,X_t=i_t)$$

справедливо для всех возможных значений i_{t+1}, i_t, ... случайных величин, иными словами, для всех возможных состояний цепи. Последовательности такого типа появляются во многих областях, например, в классической физике, где будущее развитие системы полностью определяется ее настоящим состоянием (например, мгновенной скоростью и местонахождением) и не зависит от того, как система оказалась в настоящем состоянии. Если $\{X_t\}$ — цепь Маркова и условные вероятности $P(X_{t+1}= i_{t+1}|X_t=i_t)$, ***вероятности перехода***, не зависят от t, то цепь Маркова называется однородной. Совокупность вероятностей перехода однородной цепи Маркова можно записать в виде матрицы $A=(p_{ij})$, где

$$p_{ij}=P(X_{t+1}=j\,|\,X_t=i).$$

Тогда для n-кратной степени матрицы переходных вероятностей имеем $A^n=(p_{ij}^{(n)})$, где $p_{ij}^{(n)}=P(X_{t+n}=j\,|\,X_t=i)$. Это соотношение дает возможность использовать в теории марковских цепей факты из теории матриц. В наши дни цепи Маркова (и их обобщение на случай непрерывного времени и непрерывного фазового пространства — марковские процессы) играют в естественных и технических науках намного бо́льшую роль, чем в лингвистике, где они первоначально применялись.

Проблема обратимости-необратимости — это интересный парадокс классической механики и термодинамики, и марковские цепи являются эффективным средством его анализа. Суть проблемы заключается в том, что законы классической механики обратимы и поэтому не могут объяснить, почему кусок сахара растворяется в чашке кофе, но мы никогда не наблюдаем обратный процесс. Необратимость нашего мира отражает второй закон термодинамики (впервые сформулированный *Л. С. Карно*). (Первый закон термодинамики — это закон сохранения энергии.) Спустя сорок лет *Р. Клаузиус* ввел математическое понятие энтропии, ставшее основным в теории необратимых процессов. (Согласно Клаузиусу [Memoir read at the Philos. Soc. Zürich, April 24. Pogg. Ann. 125: 353, 1865] слово «энтропия» происходит от греческого τροπη, означающего «поворот», «превращение». Клаузиус утверждает, что он добавил «эн», чтобы слово звучало аналогично «энергии», однако греческое слово εντροπη имеет самостоятельное значение — «повернуть голову в сторону».) Используя понятие энтропии, второй закон термодинамики можно сформулировать следующим образом: в изолированной системе энтропия не может уменьшиться, обычно она возрастает. *Л. Больцман* пытался проверить этот закон с помощью кинематики атомов и молекул. (В то время идея Больцмана вовсе не выглядела естественной, так как многие физики сомневались в самом существовании атомов, например, *М. Фарадей*, *Э. Мах* или основатель «энергетизма» *В. Ф. Оствальд*.) Огромное влияние на Больцмана оказала работа Максвелла по динамической теории газов. В 70-е годы прошлого века Больцман обнаружил связь между энтропией и термодинамической вероятностью (сравните с замечанием (i)). Он показал, что необратимость не противоречит обратимой механике Ньютона: применение последней к большому числу частиц с необходимостью приведет к необратимости, так как системы, состоящие из миллионов молекул, стремятся перейти в состояние, имеющее бо́льшую термодинамическую вероятность. Это и есть «основная причина» распада, износа, старения (и, как утверждают некоторые, упадка нравов или цивилизации).

В 1907 г. *П.* и *Т. Эренфесты* создали модель, разъясняющую парадокс обратимости-необратимости с помощью цепей Маркова.

б) Парадокс

Предположим, что есть система из N молекул, каждая из которых может находиться в одном из двух энергетических состояний. Если молекула находится в первом состоянии, то за один шаг она переходит в другое состояние с вероятностью p (и остается в первом состоянии с вероятностью $1-p$). Если же молекула находится во втором состоянии, то (за один шаг) она переходит в первое состояние с вероятностью q (и остается во втором с вероятностью $1-q$). Так как каждая молекула «выбирает» одно из двух возможных состояний, вся система из N молекул может находиться в 2^N различных состояниях. Предположив, что молекулы неразличимы, получим лишь $N+1$ возможных различных состояний системы: состояние системы определяется числом молекул, находящихся в первом состоянии. Пусть X_t обозначает (случайное) число молекул, находящихся в первом состоянии в момент времени t. Тогда последовательность $X_1, X_2, X_3, \ldots$, очевидно, образует цепь Маркова, описывающую развитие системы. Как эта модель увязывает обратимость классической механики (симметрия во времени) с необратимостью термодинамики (асимметрия во времени)?

в) Объяснение парадокса

Можно показать, что если $|1-p-q|<1$, то предел производящей функции распределения $P(X_t=j,\ X_{t+s}=k)$ равен

$$\lim_{t\to\infty} E\left(z^{X_t}w^{X_{t+s}}\right)=$$
$$=\left[\frac{(p+qz)(p+qw)+pq(1-p-q)^s(1-z)(1-w)}{(p+q)^2}\right]^N,$$

и эта функция симметрична по z и w, следовательно, имеет место равновесие

$$P(X_t=j,\ X_{t+s}=k)=P(X_t=k,\ X_{t+s}=j).$$

Это равенство означает симметрию между прошлым и будущим процесса (обратимость), в то же время следующее соотношение отражает необратимость

$$P(X_{t+s}=k\,|\,X_t=j)\neq P(X_{t+s}=j\,|\,X_t=k).$$

Например, если $p=q=1/2$, то

$$\lim_{t\to\infty} P(X_t=j)=\binom{N}{j}2^{-N},$$

поэтому вероятность того, что цепь Маркова близка к состоянию $N/2$ больше, чем вероятность того, что она от него далека.

г) Замечания

(i) Пусть $f(v, t)$ обозначает распределение случайных скоростей молекул газа в момент времени t (для простоты предположим, что это распределение не зависит от местонахождения молекул). В 1872 г. Больцман сформулировал свою теорему в следующем виде: производная функции

$$H(t)=\int f(v, t)\log_a f(v, t)\,dv, \ a>1,$$

не может быть положительной, т. е. H не может возрастать с ростом t. (Величина $-H$ соответствует термодинамической энтропии, которая не может убывать, она обычно возрастает.) В 1876 г. австрийский физик *Й. Лошмидт* поставил вопрос об обратимости-необратимости таким образом: законы классической физики инвариантны относительно преобразования $t\to -t$ (они содержат вторые производные по t), в то же время преобразование $t\to -t$ превращает теорему Больцмана в теорему ей противоположную: функция $H(-t)$ не может убывать. В результате анализа этого парадокса оказалось, что для доказательства теоремы Больцмана необходимо предполагать совершенную однородность молекулярных столкновений, но такая идеализация чрезмерна. Теорема Больцмана справедлива лишь статистически: вероятность роста $H(t)$ со временем очень мала.

Другой парадокс вытекает из теоремы *А. Пуанкаре*. Он показал, что в замкнутой и конечной газовой системе фазовая точка, описывающая состояние этой системы (и двигающаяся по эквипотенциальной поверхности в многомерном евклидовом пространстве) возвращается в произвольно малую окрестность своего начального местонахождения за ограниченное время. Но это, как в 1896 г. установил *Э. Цермело*, противоречит теореме Больцмана: если процесс необратим (его энтропия возрастает), то фазовая точка вернуться не может. Однако статистическая формулировка теоремы Больцмана позволяет решить и эту проблему: последовательность событий, происходящих с очень малыми вероятностями, может привести к возвращению фазовой точки, но согласно Больцману для этого потребуется $10^{10^{19}}$ лет, так что это событие практически ненаблюдаемо, в то время как необратимость проявляется наглядно.

Парадоксы Лошмидта — Цермело показывают, что теория вероятностей играет ключевую роль в основаниях молекулярной физики, в становление которой важный вклад внесли венгерские ученые. Например, в 1926 г. *Лео Силард* похоронил заманчивую идею демона Максвелла, способного создать вечный двигатель. (Максвелл утверждал, что если энтропия возрастает только статистически, то демон, способный управлять движением каждой молекулы, мог бы построить вечный двигатель. Но, как показал Силард, существование такого «хорошо информированного» демона потребовало бы огромной энтропии, поэтому вечный двигатель, управляемый демоном Максвелла, невозможен.)

(ii) Статистический анализ текста «Евгения Онегина» не был единичным исследованием подобного типа. В конце прошлого века стало модным изучать частотное распределение слов в различных текстах (чтобы помочь в обучении иностранным языкам и стенографии). В 1898 г. *Ф. Каэдиг* опубликовал первый частотный словарь (Häufigkeitswörterbuch der Deutschen Sprache) немецкого языка, который был основан на текстах, содержащих 11 миллионов слов. Во многом благодаря работам американского ученого *Дж. Ципфа* (1902—1950 гг.) применения математической статистики в лингвистике переросли в отдельное научное направление. Его книга «Человеческое поведение и принцип наименьшего усилия» («Human behavior and the Principle of Least Effort») содержит очень глубокие идеи.

д) Литература

Boltzmann L. *Lectures on Gas Theory*, Univ. California Press, Berkeley, 1964. [Имеется перевод: Больцман Л. Лекции по теории газов. — М.: Гостехиздат, 1956.]

Chung K. L. *Markov Chains with Stationary Transition Probabilities*, Springer, Berlin — Göttingen — Heidelberg, 1960. [Имеется перевод: Чжун Кайлай, Однородные цепи Маркова. — М.: Мир, 1964.]

Дынкин Е. Б. Марковские процессы. — М.: Физматгиз, 1963.

Ehrenfest P., Ehrenfest T. "Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem", *Physik. Zeit.*, **8**, 311—314, (1907).

Feller W. *An Introduction to Probability Theory and Its Applications*, New York, John Wiley, 1969. [Имеется перевод: Феллер В. Введение в теорию вероятностей и ее приложения. В 2-х томах. — М.: Мир, 1984.]

Kempermann J. H. B. *The Passage Problem for a Stationary Markov Chain*, Univ. Chicago Press, 1961.

Марков А. А. Распространение закона больших чисел на величины, зависящие друг от друга. — "Изв. Физ.-мат. общества при Казанском университете", 2-я серия, 1906, т. 15, № 4.

Марков А. А. Распространение предельных теорем исчисления вероятностей на сумму величин, связанных в цепь. СПБ, 1908. Записки Акад. Наук по Физ.-мат. отделению, VIII серия, т. 22, № 9.

Takács L. "On an urn problem of Paul and Tatiana Ehrenfest", *Math. Proc. Camb. Phil. Soc.*, **86**, 127—130, (1979).

Truesdell C. *The Tragicomic History of Thermodynamics 1822—1854*, Springer, New York, 1980.

3. Парадокс броуновского движения

а) История парадокса

Во время экспериментов с использованием микроскопа английский ботаник *Р. Броун* (1773—1858 гг.) обнаружил не только существование ядер у клеток, но и наблюдал интересное, хотя в то время необъяснимое, явление: случайное движение взвешенных частиц, известное ныне как броуновское движение.

Поскольку Броун проводил свои эксперименты с цветочной пыльцой (в июне — августе 1827 г.), предположили, что движение имеет биологические причины. Великой заслугой Броуна является экспериментальное доказательство исключительно физической природы этого явления. Степень развития микрофизики тогда не позволяла объяснить это явление научно. Неудивительно, что даже в 1879 г. немецкий ботаник *К. Негели* отказался поверить, что причина броуновского движения в тепловой диффузии частиц. С другой стороны, *А. Пуанкаре* на одной из лекций (в Париже в 1904 г.) утверждал, что когда большие частицы размером приблизительно 0.1 мм со всех сторон много раз ударяются движущимися атомами, то эти частицы остаются на месте, так как по закону больших чисел случайные столкновения нейтрализуют друг друга, однако в случае более мелких частиц воздействие толчков недостаточно для их общей нейтрализации, поэтому частицы двигаются зигзагообразно. Количественное объяснение явления было дано в 1905 г. *Эйнштейном* и польским ученым *Смолуховским* независимо друг от друга. По теореме Эйнштейна средняя длина пути частиц пропорциональна корню квадратному из времени движения t. Следовательно, их средняя скорость пропорциональна $1/\sqrt{t}$. Отсюда вытекает, что мгновенная скорость частиц в любой момент времени должна равняться бесконечности, следовательно, при определении мгновенной скорости для броуновского движения возникают проблемы. Для их решения требовался более глубокий математический анализ. Однако прошло более десяти лет прежде, чем он был проведен *Н. Винером*. В знак признания его заслуг в этой области математическая модель броуновского движения была названа его именем — винеровский процесс. Винеровский процесс — это движение с непрерывной траекторией (его реализации непрерывны), нигде недифференцируемой с вероятностью 1. Это означает, что мгновенную скорость нельзя определить ни в одной точке.

Непрерывные, но нигде недифференцируемые функции были известны задолго до Винера. Однако такие патологические функции рассматривались лишь как курьез. В 1806 г. знаменитый физик *А. М. Ампер* даже собирался доказать, что любая непрерывная функция дифференцируема, за исключением быть может нескольких изолированных точек. Понятие функции стало в значительной степени более общим в результате исследований по рядам Фурье, проведенных *Б. Больцано* (1834 г.), *Б. Риманом* (1854 г.) и *К. Вейерштрассом* (1872 г.). В 1875 г. *П. Дюбуа-Реймон* впервые опубликовал пример непрерывной, но нигде недифференцируемой функции Вейерштрасса. Многие выдающиеся математики встретили эти нововведения без большого энтузиазма. Как считал Пуанкаре (Science et Méthode, 1909. [Имеется перевод: см. «Наука и метод» в книге Пуанкаре Анри. О науке. — М.: Наука, 1983.]): «Некогда при нахождении новых функций имелась в виду какая-нибудь практическая цель. Теперь функции изобретаются специально для того, чтобы обнаружить недостаточность рассуждений наших отцов, никакого иного вывода, кроме этого, из них нельзя извлечь». *Ш. Эрмит* в письме к *Т. Стилтьесу* высказывался аналогичным образом: «Я в ужасе отворачиваюсь от этой страшной чумы: функций, не имеющих производных». Винеровский процесс, очевидно, опровергал приведенные выше обвинения, так как никто не мог утверждать, что броуновское движение введено лишь для придумывания патологических контрпримеров. Исследования, проведенные в XX веке, со всей очевидностью показали, что среди непрерывных функций именно недифференцируемые типичны в том смысле, что они образуют подавляющее большинство. (Oxtoby L. J. C. *Measure and Category*, Springer, New York, 1971. [Имеется перевод: Окстоби Дж. Мера и категория.— М.: Мир, 1974.]) Однако почти все непрерывные функции, применяемые в практических приложениях, дифференцируемы. Это напоминает ситуацию с иррациональными числами. Несмотря на их большинство среди действительных чисел (случайное число иррационально с вероятностью 1), на практике обычно используются рациональные числа.

б) Парадокс

Траектории (реализации) броуновского движения достаточно нерегулярны (т. е. они нигде не дифференцируемы). Обычно любую нерегулярную кривую такую, как траекторию броуновского движения на плоскости, рассматривают как одномерную. В то же время можно показать, что траектория броуновского движения на плоскости в действительности заполняет всю плоскость (в любую заданную окрестность произвольной точки траектория попадает с вероятностью 1). Следовательно, траектории можно рассматривать и как двумерные кривые. Какой из подходов предпочесть?

в) Объяснение парадокса

Понятие размерности в обычном повседневном смысле использовалось уже в начале века. Кривые, поверхности и тела рассматривались соответственно как одномерные, двумерные и трехмерные объекты. Обычно говорят, что фигура имеет размерность k, если для «характеризации» точек фигуры необходимо

k параметров (координат). Используя интуитивные идеи Пуанкаре, *Л. Э. Я. Брауэр* в 1913 г. определил топологическую размерность. Позднее, в 1922 г., *К. Менгер* и *П. С. Урысон*, работая независимо, также пришли к этому понятию. (Дальнейшие детали см. в книге Гуревича и Волмэна.) В силу определения топологической размерности броуновское движение одномерно. С другой стороны, в 1919 г. *Ф. Хаусдорф* ввел следующее понятие размерности, в соответствии с которым броуновское движение двумерно. В d-мерном евклидовом пространстве объем единичного шара равен $v(d) = \Gamma(1/2)^d/\Gamma(1+d/2)$, где Γ обозначает обычную гамма-функцию (см. «Обозначения» в конце книги). Это выражение имеет смысл и для нецелых $d \geqslant 0$. Пусть в n-мерном евклидовом пространстве взято множество E, которое покрывается конечным числом n-мерных шаров с радиусами $r_1, r_2, \ldots$. Тогда d-мера Хаусдорфа множества E равна

$$\lim_{r\to\infty} \inf_{r_i<r} \sum_i v(d) r_i^d.$$

А. Бесикович доказал, что всегда существует (действительное) число D такое, что если $d < D$, то d-мера множества E бесконечна, но в случае $d > D$ она равна 0. Это число D называется размерностью Хаусдорфа или Хаусдорфа — Бесиковича множества E. При таком определении значение размерности необязательно будет целым числом. Например, обе координаты броуновского движения на плоскости как функции времени (т. е. кривые «одномерного броуновского движения») имеют размерность Хаусдорфа 3/2. Следовательно, эти кривые находятся где-то между «настоящими» кривыми и «настоящими» поверхностями. Размерность кривой броуновского движения на плоскости равна 2, как и у «настоящих» поверхностей.

г) Замечания

(i) За последние несколько лет опубликовано много работ о фигурах, у которых топологическая размерность отличается от размерности Хаусдорфа. *Б. Мандельбройт* назвал их *фракталами*. Фракталы, например, винеровские процессы, играют важную роль при описании неправильных фигур, встречающихся в природе. Хотя евклидова линия — это «буква», которая наиболее часто употребляется при описании природных объектов правильной формы, в случае неправильных форм (облака, морское побережье) в этой роли выступает винеровский процесс. В действительности ни «настоящие» линии (которые можно продолжать только в длину), ни «настоящие» винеровские процессы (нигде недифференцируемые) не существуют в природе, но с их помощью можно получить довольно хорошее описание «настоящих» объектов. Фракталы по-новому осветили знаменитый парадокс *Олберса*. В соответствии с этим парадоксом не понятно, почему ночью небо не освещено равномерно, в то время как равномерно распределение звезд во Вселенной. (См. книгу Мандельбройта.)

(ii) Мандельбройт в своей книге упоминает и другие понятия размерности, например, размерность *Фурье*. По поводу алгебраической размерности см. статью *Секея*.

(iii) Нерегулярность винеровского процесса привела к развитию нового направления на стыке теории вероятностей и анализа — теории стохастических дифференциальных уравнений. В рамках этой теории получены результаты, значительно отличающиеся от фактов из обычного дифференциального и интегрального исчислений. Например, если функция $f(t)$ дифференцируема, то

$$\int_0^1 f(t)\, df(t) = \int_0^1 f(t) f'(t)\, dt = \frac{f^2(1) - f^2(0)}{2}.$$

В теории стохастических интегралов это выражение тоже имеет смысл, если даже $f(t)$ — (нигде недифференцируемый) винеровский процесс. Но значение интеграла будет меньше, чем в приведенном выше (при наличии производной) случае. Разность в точности равна $1/2$.

д) Литература

Bachelier L. *Théorie de la spéculation*, Thesis, 1900.

Brown R. "A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen on plants; and on the general existence of active molecules in organic and inorganic bodies", *Edinburgh New Phil. Journal*, **5**, 358—371, (1828).

Гихман И. И., Скороход А. В. Введение в теорию случайных процессов. — М.: Наука, 1965.

Hausdorff F. "Dimension und äusseres Mass", *Math. Annalen*, **79**, 157—179, (1919).

Hurewitz W., Wallmann H. *Dimension Theory*, Princeton Univ. Press, 1941. [Имеется перевод: Гуревич В., Волмэн Г. Теория размерности. — М.: Гос. изд-во иностранной литературы, 1948.]

McKean H. P., Jr. *Stochastic Integrals*, Academic Press, New York — London, 1969. [Имеется перевод: Маккин Г. Стохастические интегралы. — М.: Мир, 1972.]

Mandelbroit B. B. *Fractals, From, Chance and Dimension*, W. H. Freeman and Co., San Francisco, 1977.

Székely G. J. "Algebraic dimension of semigroups with application to invariant measures", *Semigroup Forum*, 17, 185—187, (1979).

Wiener N. *Collected Works*, Cambridge, Mass. M. I. T. Press, 1976.

4. Парадокс времени ожидания (Ходят ли автобусы чаще в обратном направлении?)

а) История парадокса

Хотя современная технология постоянно уменьшает потери на время ожидания, они все еще существуют и во многом определяют нашу повседневную нервозность. Поэтому за попытками математиков и инженеров сократить время ожидания следят с большим интересом. *А. Эрланг* исследовал проблему времени ожидания для телефонных станций (см. I/6, замечание (iii)). В 30-е годы нашего века *В. Феллер* ввел понятие процессов гибели и размножения, что придало новый импульс математическому анализу времени ожидания и во многом способствовало возникновению теории исследования операций. Изучение систем с очередями превратилось в независимую область науки на границе между теорией вероятностей и исследованием операций.

б) Парадокс

На автобусных остановках обычно указывается интервал движения автобуса, т. е. среднее время между двумя последовательными прибытиями автобусов. Предположим, что на некоторой автобусной остановке интервал движения составляет 10 мин. Тогда естественно считать, что люди ждут автобус в среднем 5 мин. Однако оказывается, что среднее время ожидания может не только превосходить 5 мин, но и быть бесконечным! (Опыт показывает, что в повседневной жизни ситуация не столь безнадежна.)

в) Объяснение парадокса

Если бы автобусы приходили на автобусную остановку не только в среднем, но в точности каждые 10 мин, то среднее время ожидания в действительности равнялось 5 мин. Однако на самом деле автобусы ходят «партиями» (за исключением случая, когда мы находимся недалеко от автобусной станции, откуда автобусы отправляются). Следовательно, время ожидания имеет большой разброс относительно среднего значения. Предположим, что интервалы времени между последовательными прибытиями автобусов являются независимыми одинаково распределенными случайными величинами с математическим ожиданием m и стандартным отклонением s. Тогда можно показать, что среднее время ожидания равно $T=(m^2+s^2)/2m$. Пусть $F(t)$ — функция распределения и $f(t)$ — плотность вероятности для интервалов времени между последовательными прибытиями автобусов. (Сейчас мы предположили существование плотности вероятности, однако от этого условия ценой некоторых изменений в рассуждениях можно отказаться.) Пусть время t измеряется от момента отправления последнего автобуса перед нашим приходом. Тогда плотность вероятности для случайного интервала времени до прибытия следующего автобуса равняется не $f(t)$, а другой функции, пропорциональной $tf(t)$, т. е. $tf(t)/m$, так как вероятность нашего появления в течение некоторого интервала времени пропорциональна его длине t. Таким образом, среднее время ожидания T вычисляется по формуле

$$T = \frac{1}{2m} \int_0^\infty t^2 f(t)\,dt = \frac{m^2 + s^2}{2m}.$$

(Плотность вероятности для нашего времени ожидания равна $(1 - F(t))/m$.) Следовательно, $T = m/2$ только в случае $s = 0$, но если $s = \infty$, то и $T = \infty$. Эти крайние случаи, безусловно, далеки от реальности. В действительности интервалы между прибытиями автобусов имеют почти показательное («безвозрастное») распределение с некоторым параметром λ. Тогда $m = s = 1/\lambda$, т. е. $T = m$. Это означает, что если частота движения составляет 10 мин, то среднее время ожидания также равно 10 мин, а не 5 мин.

Эвристическое объяснение этого парадокса довольно просто. Когда кто-то приходит на автобусную остановку в случайный момент времени, то имеет большие шансы ждать долго. Его время ожидания будет коротким, если он попадает на автобус из «партии», но автобусы в «партии» прибывают через малые интервалы, поэтому шансов успеть на один из них немного. Следовательно, если интервалы времени между последовательными прибытиями автобусов имеют большую дисперсию, то лишь немногие люди будут ждать мало, а большинство — в течение долгого времени. Это означает, что среднее время ожидания T велико.

г) Замечания

(i) У нас часто возникает иллюзия, что куда бы нам не надо было ехать, автобусы и трамваи чаще идут в противоположном направлении. В действительности это, естественно, невозможно. Объяснение очень простое. Мы видим только один автобус (на который мы сели), едущий в нужном нам направлении, и в то же время положительна вероятность того, что, пока мы ждем, в противоположную сторону пройдут два или три автобуса. Их математическое ожидание равно

$$\frac{m^2 + s^2}{2m} : m = \frac{1}{2} + \frac{s^2}{2m^2},$$

что действительно больше 1/2, если s положительно. Отсюда вытекает асимметрия между двумя направлениями. Однако на самом деле это не так. Симметрия между двумя направлениями заключается в том, что вероятность того, что ни один автобус не проедет в противоположном направлении, пока мы ждем свой автобус, в точности равна 1/2 (но если один автобус пройдет в противоположном направлении, то могут пройти и несколько, поэтому возможно, что математическое ожидание будет как угодно большим). Пусть p_k обозначает вероятность того, что, пока мы ждем, ровно k автобусов пройдут в противоположном направлении. Если интервалы между последовательными прибытиями автобусов имеют показательное распределение, то

$$p_k = \left(\frac{1}{2}\right)^{k+1};$$

в случае равномерного распределения на интервале (0, 1) имеем

$$p_k = 4\left(\frac{1}{(k+2)!} - \frac{2}{(k+3)!} + \frac{1}{(k+4)!}\right),$$

где $k=1, 2, \ldots$ (p_0 всегда равно 1/2, как мы уже отмечали).

(ii) В том случае, когда число автобусов на линии велико, автобусное обслуживание можно сделать более устойчивым, если разрешить автобусам дольше стоять на остановках. В результате время ожидания будет сокращено. (В действительности я никогда не видел автобусов, ожидающих кого-то на остановках, чтобы сделать движение более устойчивым. С другой стороны, иногда задерживают лифты с тем, чтобы подождать людей, которые скоро подойдут. В результате движение лифтов замедляется, но вместе с тем уменьшается среднее время ожидания!) Пусть $t_1, t_2, t_3, \ldots$ обозначают моменты времени, когда автобусы прибывают на некоторую остановку. Положим $X_1 = t_1$, $X_i = t_i - t_{i-1}$, $(i = 2, 3, \ldots)$. Если F — функция распределения каждой из случайных величин $X_1, X_2, \ldots$, то, как уже отмечалось, плотность вероятности величины X_1 равна $[1 - F(t)]/m$ и математическое ожидание $T = (m^2 + s^2)/2m$. Замедление движения означает, что величины X_i возрастают и становятся равными $X_i + g(X_i)$ $(i = 2, 3, \ldots)$, где g — неотрицательная функция. Можно показать, что время ожидания уменьшится в наибольшей степени, если, выбирая среди интегрируемых функций, взять функцию $g(x) = \max(0, (c - x))$, где c — единственное решение уравнения

$$cE(X) + \int_0^c (c - x) F(x)\, dx = E(X^2)/2$$

(здесь X — случайная величина, имеющая то же распределение, что и X_2, X_3, ...). Например, если X имеет показательное распределение, точнее, если $F(x) = 1 - e^{-x}$ $(x > 0)$, то и математическое ожидание, и дисперсия времени ожидания равны 1. При выборе оптимальной функции $g(x) = \max(0, (0.901 - x))$ (с точностью до трех знаков после запятой) математическое ожидание времени ожидания станет равным 0.901 (и дисперсия — 0.691).

(iii) Следующий парадокс также связан с движением транспорта. (*Г. Шэй* привлек мое внимание к этой проблеме после моего выступления в Массачусетском технологическом институте в 1983 г.) Суть парадокса в следующем: неверно представление о том, что чем выше скорость машин, тем большее их число проезжает на зеленый свет, так как при более высокой скорости водители должны сохранять и большее расстояние между машинами. Для нахождения оптимальной скорости начнем со следующей модели. Предположим, что машины движутся с одинаковыми скоростями v; пусть X_i обозначает (случайное) время между началом движения i-й и $(i+1)$-й машины. Величины X_i независимы и одинаково распределены (для простоты предположим, что их распределение показательно с параметром $\lambda > 0$). Машины проезжают светофор с интервалами времени Y_1, Y_2, ... , которые, вообще говоря, не равны X_1, X_2, ... , так как машины должны двигаться друг за другом на некотором расстоянии. Пусть l_i обозначает длину i-го автомобиля и a_i — уменьшение скорости при торможении; величины l_i и a_i, как правило, зависимы, но мы предположим, что вектора (l_i, a_i), $i = 1, 2, \ldots$, независимы и одинаково распределены. Тормозное расстояние равно $v^2/2a_i$, следовательно, машины должны следовать друг за другом на расстоянии $l_i + v^2/2a_i$. Время между проездом первой и $(n+1)$-й машинами равно

$$\sum_{i=1}^{n} Y_i = \max\left\{ \sum_{i=1}^{n} X_i,\ \sum_{i=1}^{n-1} Y_i + Z_{n-1} \right\},$$

где

$$Z_i = \left(l_i + \frac{v^2}{2a_i}\right)\Big/ v.$$

Если

$$M(t) = \max\left\{ n\colon \sum_{i=1}^{n} Y_i \leqslant t \right\},$$

то число машин, которые могут проскочить на зеленый свет за время от t до $(t+h)$ равно $M(t+h) - M(t)$. Известно (из теории очередей), что

$$\lim_{t\to\infty}\frac{M(t)}{t}=\begin{cases} E(Z_1)^{-1}, \text{ если } E(X_1)\leqslant E(Z_1) \\ \text{(это соответствует движению в часы пик)}, \\ E(X_1)^{-1}, \text{ если } E(X_1)>E(Z_1). \end{cases}$$

Пусть t — случайное время внутри интервала $[0, T]$. Тогда среднее число машин, проезжающих на зеленый свет, вычисляется по формуле

$$\frac{1}{T}\int_0^T [M(t+h)-M(t)]\,dt=\frac{1}{T}\left[\int_T^{T+h} M(t)\,dt-\int_0^h M(t)\,dt\right]=$$

$$=h\min\left\{E(Z_1)^{-1},\ E(X_1)^{-1}\right\}+o(1), \text{ если } T\to\infty;$$

($o(1)$ обозначает величину, стремящуюся к 0 при $T\to\infty$.) Следовательно, ищем максимум величины

$$\min\left\{\lambda,\ \left[E(l_1)/v+E(a_1^{-1})\,v/2\right]^{-1}\right\}$$

как функции от v. Второй член максимален при $v=\sqrt{2E(l_1)/E(a_1^{-1})}$.

д) Литература

Гаджиев А. Г. Минимизация среднего времени ожидания в системах с рекуррентным обслуживанием. — Вестник МГУ, сер. 1, № 3, 1980 с. 19—24.

Kleinrock L. *Queuing Systems*, Wiley, New York, 1975. [Имеется перевод: Клейнрок Л. Вычислительные системы с очередями. — М.: Мир, 1979.]

Takács L. *Introduction to the Theory of Queues*, Oxford Univ. Press, New York, 1962.

5. Парадокс случайных блужданий

а) История парадокса

Приблизительно 60 лет назад американский математик венгерского происхождения *Джордж Пойа*, прогуливаясь в парке, обычно встречался с одной и той же парой. В то время он не осознавал, насколько случайны были эти встречи, т. е. насколько мала была их вероятность. Вскоре после этого он вычислил вероятность встреч в модели, когда 2 человека случайно блуждают по квадратной решетке независимо друг от друга (в каждом узле решетки вероятность выбора любого из четырех возможных направлений одна и та же). Пойа установил, что вероятность встречи равна 1. (Следовательно, если время блуждания не ограничено, то люди могут встретиться бесконечное число раз с вероятностью 1.) Однако в случае кубической решетки вероятность встречи строго меньше 1 (поэтому вероятность бесконечного числа встреч теперь равна 0). Из этого интересного открытия за последние 60 лет в теории вероятностей развилось совершенно новое направление. Его изложению посвящена хорошая книга *Ф. Спицера*, вышедшая в 1964 г.

б) Парадокс

Из теоремы Пойа вытекает, что, рассматривая случайное блуждание по целым точкам на прямой с началом в 0 и движением за один шаг вправо или влево на 1 с равными вероятностями 1/2 (независимо от предыдущих шагов), мы будем возвращаться в 0 с вероятностью 1. Возникает вопрос: сколько раз до (первого) возвращения в 0 мы будем проходить через фиксированное целое число k? Естественно предположить, что, чем больше $|k|$, т. е. чем дальше случайное блуждание уходит от 0, тем реже в среднем это будет происходить. Удивительно, но случайное блуждание до первого возвращения в 0 будет всегда проходить через k в среднем одно и то же количество раз, а именно один, как бы ни было велико $|k|$.

в) Объяснение парадокса

Парадокс можно разъяснить очень просто. Среднее число шагов, необходимых для возвращения в 0 (т. е. математическое ожидание времени возвращения) бесконечно, следовательно, имеется достаточно времени для того, чтобы в среднем один раз достичь любой точки на прямой.

г) Замечания

(i) Пусть выполнены указанные выше условия, но теперь перемещение вправо происходит на 2 единицы, а влево на одну. Тогда случайное блуждание несимметрично, и легко видеть, что, начав движение из 0, мы с меньшей вероятностью попадем в -1, чем в 1. Удивительно, но эта вероятность в точности равна $(\sqrt{5}-1)/2$, т. е. отношению золотого сечения.

(ii) Для практики очень важны случайные блуждания диффузионного типа, когда вероятность перемещения влево или вправо зависит от настоящего местонахождения (k). Пусть p_k обозначает вероятность движения вправо и $(1-p_k)$ — вероятность движения влево. Предположим. что

$$p_k=\frac{1}{2}\left(1+\frac{c}{k}\right),$$

(по крайней мере для больших значений $|k|$), где c — произвольная постоянная. При таком случайном блуждании возвращение в 0 происходит с вероятностью 1 (т. е. блуждание возвратно), если $c \leqslant 1/2$. Для $c < -1/2$ математическое ожидание времени возвращения конечно, следовательно, парадоксальная ситуация, как в случае $c = 0$, не возникает.

(iii) Исследования по случайным блужданиям можно распространить с квадратных решеток на более общие структуры, называемые графами. Такие обобщения имеют интересные приложения в теории электрических сетей. См. глубокую статью *К. Неш-Вильямса*, написанную в 1959 г. Множество других применений (в физике, химии и биологии) обсуждаются в работе Вейса.

д) Литература

Bass R. F., Griffin P. S. "The most visited site of Brownian motion and simple random walk", *Zeitsch 'Wahrch' theorie verw. Geb.*, **70**, 417—436, (1985).

Nash-Williams C. St. J. A. "Random walk and electronic currents in networks", *Proc. Camb. Phil. Soc.*, **55**, 181—194, (1959).

Polya G. "Über eine Aufgabe der Wahrscheinlichkeitsrechnung und das Momentproblem", *Math. Annalen*, **84**, 149—160, (1921).

Spitzer F. *Principles of Random Walk*, Van Nostrand, New York — Toronto — London, 1964. [Имеется перевод: Спицер Ф. Принципы случайного блуждания. — М.: Мир, 1969.]

Weiss G. H. "Random walks and their applications", *American Scientist*, **71**, 65—71, (1983).

6. Биржевый парадокс; мартингалы

а) История парадокса

Математические исследования, связанные с фондовой биржей, насчитывают почти столько же лет, сколько и сама биржа. Математический подход, видимо, использовался уже на бирже *Грешема* в XVI веке, однако основные методы теории вероятностей не применялись довольно долго. Характерно, что даже в 1900 г., когда в Париже *Луи Башелье* защищал докторскую диссертацию о связи между колебанием цен на фондовой бирже и броуновским движением (еще до того, как броуновским движением занялись физики), комиссия с трудом воспринимала его по существу новые идеи. Башелье создал общую математическую модель безобидных игр, так называемый *мартингал*, который позднее, после исследований *Ж. Вилле*, *П. Леви*, *Дж. Дуба* и других, стал одним из важнейших стохастических процессов. Последовательность случайных величин X_1, X_2, X_3, ... называется мартингалом, если для любого n условное математическое ожидание разности $X_{n+1} - X_n$ («прибыль, полученная за время n») относительно случайных величин X_n, X_{n-1}, ... равно 0 с вероятностью 1, т. е.

$$E(X_{n+1} - X_n \mid X_n, X_{n-1}, \ldots) = 0$$

с вероятностью 1. Последовательность X_1, X_2, X_3, ... является *супермартингалом* (или субмартингалом), если указанное выше математическое ожидание неположительно (или неотрицательно). Мартингал представляет собой общую модель безобидных игр, модель «количественной справедливости», которая применима во многих случаях, в частности, при анализе парадоксов, связанных с фондовой биржей.

б) Парадокс

Если ожидается, что какие-то акции принесут прибыль, то решение об их покупке кажется естественным, если же прибыли не будет, то надо их продавать. Столь же разумной представляется покупка на все деньги акций, от которых ожидается наибольшая прибыль. Все это верно, но на практике поступают иначе, так как хотя наша ожидаемая прибыль может возрастать (общий ожидаемый капитал будет стремиться к бесконечности), одновременно наше состояние будет убывать до нуля с вероятностностью 1. Так что действовать на фондовой бирже надо осторожно: акции, которые обещают принести прибыль, иногда следует продать.

в) Объяснение парадокса

Предположим, что мы собираемся купить акции и можем выбрать среди k различных акций. За год i-я акция ($i = 1, 2, \ldots, k$) увеличивает наш капитал, который был в начале года, в $X^{(i)}$ раз. (Очевидно, $X^{(i)} \geqslant -1$.) Для простоты предположим, что величина $X^{(i)}$ ограничена, хотя от этого условия можно отказаться, если немного изменить приводимые ниже рассуждения. Случайный вектор $X = (X^{(1)}, \ldots, X^{(k)})$ описывает котировку акций. Считаем, что векторы X_j ($j = 1, 2, \ldots$), описывающие котировку в j-м году, независимы и распределены так же, как и X. Пусть T_0 — наш начальный капитал, и $a_j^{(i)}$ означает долю нашего капитала, который мы потратили на покупку акций i-го типа в j-м году. Величина $a_j^{(i)} \geqslant 0$ может зависеть от случайных векторов $X_1, X_2, \ldots, X_{j-1}$. Вектор $a_j = (a_j^{(1)}, a_j^{(2)}, \ldots, a_j^{(k)})$ описывает нашу закупочную стратегию в j-м году. Очевидно, $\sum_{i=1}^{k} a_j^{(i)} \leqslant 1$. Через a_jX_j обозначим следующую сумму:

$$\sum_{i=1}^{k} a_j^{(i)} X_j^{(i)}.$$

С помощью этого обозначения наш общий капитал в конце n-го года запишется в виде

$$T_n = T_0 \prod_{j=1}^{n} (1 + a_j X_j).$$

Математическое ожидание величины T_n будет наибольшим, если каждый год покупать на все деньги наиболее прибыльные акции. (Предполагаем, что по крайней мере одна акция прибыльная.) В этом случае математическое ожидание величины T_n стремится к бесконечности (так что мы, кажется, богатеем), и в то же время наш общий капитал T_n может стремиться к нулю с вероятностью 1! Проанализируем эту парадоксальную ситуацию подробно. Ясно, что

$$\log T_n - \log T_0 = \sum_{j=1}^{n} \log (1 + a_j X_j).$$

Предположим, что $a_j = a$ есть постоянный вектор (не зависящий от j; это предположение довольно естественно, так как в нашем случае распределения котировок не меняются). Тогда правая часть приведенного выше равенства принимает наибольшее значение (с вероятностью 1 при больших n это следует из закона больших чисел), если величина $E(\log(1 + a_j X_j))$ максимальна (при условии, что $a_j^{(i)} \geqslant 0$ и $\sum_{i=1}^{k} a_j^{(i)} \leqslant 1$). Пусть a^* обозначает стратегию, максимизирующую приведенное выше равенство. Пусть также T_n^* и T_n обозначают наш общий капитал, если мы выберем соответственно стратегию a^* или произвольную стратегию a. Тогда можно показать, что последовательность T_n/T_n^* $(n = 1, 2, \ldots)$ всегда является неотрицательным супермартингалом (более того, если каждая координата $a^{*(i)}$ вектора a^* положительна и

$$\sum_{i=1}^{k} a^{*(i)} < 1,$$

то последовательность будет мартингалом). Следовательно, в силу одной хорошо известной теоремы из теории мартингалов получаем, что предел

$$\lim_{n \to \infty} T_n/T_n^* = T$$

всегда существует с вероятностью 1 и его максимальное математическое ожидание равно 1. Итак, a^* является оптимальной стратегией (в этом смысле) на долгосрочный период. Таким образом, удобнее максимизировать математическое ожидание величины $\log T_n$, а не величины T_n. Эвристическое объяснение этого факта достаточно просто: для любой разумной стратегии величина T_n возрастает экспоненциально и для максимизации скорости этого роста можно максимизировать как раз математическое ожидание величины $\log T_n$.

Рассмотрим теперь простой (но крайний) случай. Предположим, что можно выбирать из акций двух типов. Пусть с вероятностью $p_{11} = 10\ \%$ стоимость обеих акций удвоится, с вероятностью $p_{00} = 5\ \%$ обе акции обесценятся, с вероятностью $p_{10} = 50\ \%$ стоимость первой акции удвоится и вторая обесценится и, наконец, с вероятностью $p_{01} = 35\ \%$ произойдет то же самое, но акции поменяются местами. Тогда первая акция прибыльна (с вероятностью 60 %), а вторая — убыточна (она прибыльна с вероятностью 45 %), однако имеет смысл купить акции обоих типов, точнее каждый год тратить треть денег на покупку акций в отношении 13:4. В общем случае разумно тратить часть денег, равную

$$\frac{p_{11} - p_{00}}{p_{11} + p_{00}}$$

на покупку обоих типов акций в отношении

$$(p_{11}p_{10} - p_{01}p_{00})/(p_{11}p_{01} - p_{10}p_{00})$$

(предполагаем, что все разности положительны). Хотя проблемы, возникающие в практике фондовой биржы намного сложнее, чем предыдущий пример, парадокс, который мы рассматриваем, входит в число этих сложных проблем.

г) Замечания

(i) Мартингал как система игры был хорошо известен задолго до появления математической теории мартингалов. (Приведем цитату из статьи Снелла (Snell J. L. «Gambling, probability and martingales», *The Mathematical Intelligencer*, 4, 118—124, (1982)): «Основная идея мартингальной системы заключается в удваивании ставки при проигрыше. Предположим, что мы играем в рулетку и всегда ставим на красное. Сначала поставим 1 доллар. Если выигрываем, то прекращаем игру; при проигрыше ставим в следующий раз 2 доллара. При выигрыше наш капитал увеличится на 1 доллар, и мы прекратим игру; при проигрыше мы теряем уже $1 + 2 = 3$ доллара и ставим 4 доллара. При выигрыше наш капитал увеличится на 1 доллар, и мы прекратим игру; при проигрыше ставим в следующий раз 8 долларов и т. д. Если колесо рулетки хотя бы раз остановиться на красном числе, то при такой мартингальной системе игры мы покинем казино, став на 1 доллар богаче, чем когда в него вошли. Поскольку красное в конце концов должно появиться, кажется, что эта система гарантирует успех. Однако предположим, что мы вошли в казино со 100 долларами, и 6 раз подряд выигрывало черное. Тогда мы проиграем $2^6 - 1 = 63$ доллара и не сможем сделать следующую ставку в нужном размере — 64 доллара».

В семейной хронике «Ньюкомы» Теккерей замечает: «Еще не ставили? И не ставьте! А если уж начнете, то никогда не удваивайте ставки при проигрыше» [1]).

Это хороший совет для игроков, однако математики ему не следуют и в итоге в теории вероятностей появляются многие важные результаты.

(ii) *Томас Грешем* (1519—1579 гг.), основатель Лондонской фондовой биржи, видимо, догадывался, что математика играет важную роль при анализе биржевых операций и экономической жизни. В завещании Грешема содержался план создания колледжа, в котором при изучении экономики математика была бы одним из основных предметов. *Генри Бригс*, опубликовавший в 1617 г. первые таблицы логарифмов, был профессором в Грешем Колледже, который во многих отношениях предопределил появление Лондонского королевского общества.

(iii) Мартингалы с успехом применяются в генетике, теории потенциала, стохастических интегралах и т. д. Выдающимися монографиями в этой области являются труды *Ж. Неве*, *П. Мейера*, *К. Хейди* и *П. Холла*.

д) Литература

Breiman L. "Optimal gambling systems for favorable games", *Proc. 4th Berkeley Symp. on Math. Statist. and Prob.*, 65—78, (1961).

Doob J. L. *Stochastic Processes*, Wiley, New York, 1953. [Имеется перевод: Дуб Дж. Л. Вероятностные процессы. — М.: ИЛ, 1956.]

Móri T. F., Székely G. J. "How to win if you can", *Coll. Math. Sci. Bolyai*, **36** (ed. P. Révész), 791—806, (1982).

Ville J. *Étude critique de la notion de collectif*, Gauthier-Villars, Paris, 1939.

7. Еще несколько парадоксов

а) Парадокс Иакова и Лавана

Согласно библейской легенде о Ивакове и Лаване, в награду за свою службу Иаков получал от Лавана скот с пятнами. Хотя доля скота с пятнами во всем стаде была очень мала, Иаков постепенно разбогател и стал намного богаче Лавана.

---

[1]) Теккерей У. Собрание сочинений. В 12-ти томах. Т. 8. Ньюкомы. Кн. 1. Пер. с англ. Р. Померанцевой. — М.: «Худож. лит.», 1978.

Существует множество мистических объяснений этого парадокса (один содержится в самой библии, этой загадкой занимался и Томас Манн), однако, как однажды отметил Альфред Реньи, в этом парадоксе нет ничего мистического, его можно понять исходя из простых математических рассуждений, основанных на том факте, что Иаков никогда не возвращал скотину Лавану, а Лаван всегда отдавал Иакову часть своего скота.

Обозначим среднюю численность стада Иакова и стада Лавана в n-м году соответственно через J_n и L_n (в начале, в 0-м году $J_0 = 0$ и величина L_0 положительна). Предположим, что ежегодно у каждой овцы рождается в среднем U ягнят. Пусть q обозначает долю овец в стаде Лавана, которую он отдает Иакову ($p = 1 - q$ — доля, остающаяся у Лавана). Тогда $L_{n+1} - L_n = UpL_n$ и $J_{n+1} - J_n = UJ_n + UqL_n$, следовательно, $L_n = L_0(1 + Up)^n$ и $J_n = L_0(1 + U)^n - L_0(1 + Up)^n$. Таким образом

$$\frac{J_n}{L_n} = \left(\frac{1+U}{1+Up}\right)^n - 1,$$

что стремиться к бесконечности с ростом n, поэтому действительно Иаков со временем станет богаче Лавана. Например, для $q = 10\ \%$, $U = 2$, $n = 20$ отношение для J_n/L_n приблизительно равно 3.

б) Парадокс процессов с независимыми приращениями

Процессы с независимыми приращениями и их дискретные варианты, частичные суммы независимых случайных величин, являются классическими объектами исследования в теории вероятностей. Пусть X_1, X_2, ... — независимые (не равные тождественно нулю) случайные величины с нулевым математическим ожиданием. Тогда суммы $S_n = X_1 + X_2 + \ldots + X_n$, $n = 1, 2, \ldots$ колеблются около нуля, т. е. если величины X_i одинаково распределены, то (согласно теореме *К. Чжуна* и *В. Фукса*, доказанной в 1951 г.)

$$P(\limsup_{n\to\infty} S_n = +\infty) = P(\liminf_{n\to\infty} S_n = -\infty) = 1.$$

Однако это свойство колебаний может не выполняться, если случайные величины X_i распределены неодинаково. Например, положим $X_i = Y_i/\sqrt{1 - i^{-2}}$, где $P(Y_i = i^{-1}) = 1 - i^{-2}$ и $P(Y_i = -i + i^{-1}) = i^{-2}$. Если случайные величины Y_i независимы, то величины X_i также независимы и имеют нулевое математическое ожидание и единичную дисперсию. По *лемме Бореля — Кантелли*, если A_1, A_2, A_3, ... — произвольные события и сумма их вероятностей сходится, то с вероятностью 1 происходит только конечное число событий A_k. Следовательно, событие $Y_i = -i + i^{-1}$ также происходит лишь конечное число раз (так как $\sum_{i=1}^{\infty} i^{-2} < \infty$), поэтому для достаточно больших n с вероятностью 1 имеем $Y_i = i^{-1}$, т. е. $X_i = 1/\sqrt{i^2 - 1}$. Таким образом,

$$P(\lim_{n\to\infty} S_n = \infty) = 1.$$

в) Парадокс забитых голов

Две команды A и B играют друг с другом в футбол. Предположим, что силы команд равны (т. е. в любой момент матча обе команды забивают следующий гол с вероятностью 1/2). Если продолжительность интервала времени между двумя последовательными голами постоянна, то кажется естественным считать, что в течение 50 % игрового времени впереди была команда A и в течение других 50 % — выигрывала команда B. Удивительно, но верно как раз противоположное: наименее вероятен тот случай, когда A (или B) ведет в игре в течение половины игрового времени (если общий счет равный, то считается, что вела в игре та команда, которая выигрывала перед последним голом). Если в игре было забито $n = 20$ мячей, то вероятность того, что после 10 голов впереди была команда A и после других 10 голов выигрывала команда B, равна всего лишь 6 %. А вероятность того, что одна из команд выигрывала в течение всей игры, приблизительно составляет 35 %. Удивительно также, что вероятность того, что одна из команд впереди во время всей второй половины игры, равна 50 %, независимо от величины n.

Ситуация сильно изменится, если «способность забивать голы» у команд зависит от счета в игре. Пусть

$$p_k = \frac{1}{2}\left(1 + \frac{c}{k}\right)$$

есть вероятность того, что команда A забивает следующий гол, если A впереди на k мячей и $k \neq 0$; $p_0 = 1/2$. При больших c и малых k неравенства $0 < p_k \leqslant 1$ могут нарушаться, тогда положим $p_k = 1/2$. (Если $c = 0$, то $p_k = 1/2$ для всех k и мы приходим к простой модели, которую только что рассмотрели.) При положительных значениях c выигрывающая команда имеет больше шансов забить следующий гол. Если $c > 1/2$, то спустя некоторое время одна из команд «ломается», т. е. при большом числе забитых мячей только одна команда (какая именно — зависит от случая) впереди почти все 100 % игрового времени. С другой стороны, при отрицательных значениях c проигрывающая команда забивает гол с большей вероятностью. При $c < -1/2$ ход матча переменчив и интересен: половина матча впереди одна команда, во второй половине — другая команда.

Можно показать, что при $c = 0$ вероятность того, что доля игрового времени, в течение которого выигрывает команда A, не превышает x $(0 < x < 1)$, сходится к

$$F(x) = \frac{2}{\pi} \arcsin \sqrt{x} \quad \text{при} \quad n \to \infty.$$

Соответствующая плотность вероятности для $0 < x < 1$ есть функция

$$f(x) = \frac{1}{\pi \sqrt{x(1-x)}},$$

которая принимает наименьшее значение при $x = 1/2$. Таким образом, плотность вероятности минимальна в точке, соответствующей тому, что A выигрывает в точности в течение 50 % игрового времени. Это *закон арксинуса Поля Леви* (1939 г.). (Я предполагаю, что в общем случае плотность вероятности пропорциональна $(2c+1)$-й степени функции $f(x)$, если $c < 1/2$.)

Наконец, еще один удивительный факт: пусть в случае $c = 0$ игра закончилась вничью $(n:n)$. Мы хотим узнать, в течение какого времени выигрывала команда A, и в качестве единицы времени берем интервал между двумя последовательными голами. Тогда вероятность того, что A выигрывала в течение $2k$ $(k = 0, 1, 2, \ldots, n)$ единиц времени не зависит от k!

(Лит.: Feller W. *An Introduction to Probability Theory and its Applications*, John Wiley, New York, 1969. [Имеется перевод: Феллер В. Введение в теорию вероятностей и ее приложения. В 2-х томах. — М.: Мир, 1984.]
Lamperti J. "Criteria for recurrence or transience of stochastic process I", *J. Math. Anal. Appl.*, 1, 314—330, (1960).)

г) Парадокс ожидаемого времени разорения

Пусть A и B играют в орлянку. Если выпадает герб, то A платит B, если решка, то B платит A 1 доллар. Начальный капитал у A составляет 1 доллар, у B — 999 долларов; они играют до тех пор, пока один из них не разорится. У A, конечно, больше шансов первому остаться без денег. Если при первом бросании монеты выпадает герб, то A уже разорен. Как это ни удивительно, но ожидаемая продолжительность игры довольно велика: в среднем лишь после 999 подбрасываний монеты один из игроков разорится. (Не является ли такая продолжительность намного больше того, что мы ожидали? В общем случае можно доказать, что если A имеет a долларов, и у его противника B есть b долларов, то средняя продолжительность игры составляет ab испытаний, в частности при $a = b$ ожидаемая продолжительность игры равна a^2.) *Ф. Штерн* исследовал случай, когда монета может быть несимметричной, и в 1975 г. привлек внимание ученых к следующему удивительному явлению (см. *Math. Mag.* 48, 286—288). Предположим, что при каждом бросании монеты игрок A выигрывает с вероятностью p, B — с вероятностью $1-p$ $(0 < p < 1)$, и оба игрока имеют по a долларов в начале игры. Кажется очевидным, что при $p \neq 1/2$ условное математическое ожидание продолжительности игры при условии, что разорился игрок A, совершенно отличается от условного математического ожидания при условии, что к концу игры разорился игрок B. Однако можно показать, что независимо от предположения, кто именно, A или B, разорился, средние продолжительности игр, а также их распределения совпадают. Доказательство простое: вероятность разорения B после $(2k+a)$-го испытания запишется в виде $p_{2k+a} = c_{k,a}p^{k+a}(1-p)^k$ $(k = 0, 1, 2, \ldots)$, и, аналогично, вероятность разорения A после $(2k+a)$-го испытания равна $q_{2k+a} = c_{k,a}p^k(1-p)^{k+a}$, где $c_{k,a}$ — общее число игр, в которых ровно k раз выпадал герб и $k+a$ раз — решка. Поскольку отношение $p_{2k+a} : q_{2k+a}$ не зависит от k, условные распределения $p_{2k+a} \Big/ \sum_k p_{2k+a}$ и $q_{2k+a} \Big/ \sum_k q_{2k+a}$ совпадают, как мы уже отмечали. Объяснение этого явления заключается в следующем факте: при $p = 0.99$ продолжительная игра с большой вероятностью закончится разорением B, поэтому ожидаемая продолжительность игры при условии, что разорился A, будет столь же короткой, как и в случае, когда известно о разорении B. (Другое объяснение см. в статье Seneta E. «Another look at independence of hitting place and time for simple random work», *Stoch. Proc. and their Appl.*, 10, 101—104, (1980).) Мы уже отмечали, что если игроки A и B имеют по a долларов и монета симметричная, то ожидаемая продолжительность игры равна a^2. Но что произойдет, если игра проводится с помощью двух различных монет: вероятность выигрыша для A, когда бросается первая монета, равна $p_1 = 1/2 + \varepsilon$, и при бросании второй монеты — $p_2 = 1/2 - \varepsilon$ $(0 < \varepsilon < 1/2)$. При каждом испытании выбор p_1 или p_2 игроками случайным образом зависит от капитала k $(k = 1, 2, \ldots, 2a-1)$ игрока A, а именно, перед началом игры каждому значению, независимо одно от другого, мы приписываем p_1 или p_2 с равными вероятностями. Интуитивно кажется, что эта игра с ее усложненной формулировкой практически идентична, по крайней мере при больших a, игре, в которой $p_1 = p_2 = 1/2$ для всех k (т. е. классической игре, связанной с бросанием монеты), так как большое число членов $\pm\varepsilon$ при больших a уравновешивают друг друга. Но это не так. *Я. Г. Синай* недавно доказал, что средняя продолжительность такой усложненной игры значительно возрастает. Даже логарифм среднего числа необходимых бросаний имеет порядок $\sqrt{a}$ (в отличие от упоминавшегося выше a^2). Этот удивительный факт можно объяснить, используя замечание (i) в I/9. В последовательности длины a, состоящей из независимых и равновероятных p_1 и p_2, с большой вероятностью найдется серия из p_1 или p_2 длины $\log a$, что приближает текущий капитал игроков к начальному капиталу, следовательно, момент разорения откладывается. Пробиться сквозь эти «толстые стены» очень трудно и поэтому средняя продолжительность игры возрастает. Проблемы подобного типа (т. е. случайные блуждания при меняющихся случайным образом условиях) тесно связаны с теорией случайных полей, о которой идет речь в последнем парадоксе этого раздела.

д) Парадокс оптимальных правил остановки

Мы играем в орлянку с помощью симметричной монеты и прекращаем игру после n-го бросания. В этом случае мы либо выигрываем $\frac{n}{n+1}2^n$ долларов, либо не выигрываем ничего в зависимости от того, всегда ли выпадали решки или нет. Когда рекомендуется остановиться? Пусть I_n обозначает наш выигрыш (зависящий от случая) после n-й игры:

$$I_n = \frac{n}{n+1}2^n \quad \text{или} \quad I_n = 0.$$

Предположив, что $I_n \neq 0$, запишем ожидаемое значение выигрыша I_{n+1}

$$E(I_{n+1} \mid I_n \neq 0) = \frac{n+1}{n+2}2^{n+1}\frac{1}{2},$$

которое больше, чем $\frac{n}{n+1}2^n$, следовательно, всегда стоит продолжать игру. Однако вероятность того, что $I_n = 0$ при некотором (возможно, большом) значении n, равна 1. Действительно ли стоит играть до тех пор, пока мы все не потеряем?

(Лит.: Chow Y. S., Robbins H. and Seigmund D. *Great Expectations: The Theory of Optimal Stopping*, Houghton Miffin, Boston, 1971. [Имеется перевод: Роббинс Г., Сигмунд Д., Чао И. Теория оптимальных правил остановки. — М.: Наука, 1977.]

Ширяев А. Н. Статистический последовательный анализ. — М.: Наука, 1976.)

е) Парадокс выбора

Нам часто нужно выбрать лучшее (с какой-то точки зрения) из некоторой совокупности людей или объектов (например, при покупке товаров или выборе будущего супруга). Для анализа этой проблемы предположим, что людей или объекты можно упорядочить по их достоинствам, т. е. сравнивая любые два из них, можно всегда сказать, какой из них лучше. Выбор лучшего не представляет трудностей, когда мы видим все объекты. Однако в большинстве случаев объекты или людей рассматривают последовательно и, раз что-то или кого-то отвергнув, мы к этому вернуться не можем. В дальнейшем будем предполагать, что если «кандидат» не выбран, когда подошла его очередь, то позднее мы не можем изменить наше решение. Но и в этом случае проблема не описана однозначно. Мы можем даже не знать общего числа объектов, из которых должны выбирать. (Как правило, нет такой информации при выборе будущего мужа или жены.) Предположим, что всего имеется n возможностей, точнее n лиц или объектов, проходящих мимо нас в произвольной последовательности (эти последовательности считаются равновероятными). Вопрос состоит в следующем. Исходя из какого метода выбирать лучшего кандидата, если любого из них можно сравнивать, естественно, только с предыдущими? Если всегда выбирать, например, третьего, то шансы выбрать лучшего равны $1/n$. С ростом n величина $1/n$ стремится к 0 и поэтому при большом числе предложений вероятность выбора лучшего близка к 0. Удивительно, но есть метод, позволяющий выбирать лучшего кандидата с вероятностью близкой к 30 % даже при больших значениях n. Метод состоит в следующем. После того, как пройдут первые 37 % (точнее $100/e$ %) кандидатов, выбираем первого, кто окажется лучше всех предыдущих (если такого нет, то выбираем последнего). В этом случае шансы выбрать лучшего приблизительно равны $1/e$, т. е. ≈ 37 %, как бы ни было велико значение n.

Если можно выбрать два, три, ... или, в общем случае, k кандидатов, и задача заключается в том, чтобы лучший оказался среди этих k отобранных кандидатов, то оптимальная вероятность p_k такого события вычисляется следующим образом. Пусть числа c_j удовлетворяют тождеству

$$\sum_{j=1}^{\infty} x^{j-1} e^{-c_j x} \equiv 1.$$

Тогда

$$p_k = \sum_{j=1}^{k} e^{-c_j},$$

например,

$$p_2 = \frac{1}{e} + \frac{1}{e^{3/2}},$$

что больше, чем 1/2! Можно также показать, что

$$\left(1 - \frac{1}{e}\right)^k \leqslant 1 - p_k \leqslant e^{-k/e},$$

следовательно, p_k сходится к 1 при стремлении k к бесконечности.

Если число кандидатов N случайно, то шансы выбрать лучшего кандидата могут уменьшиться. Предположим, что распределение N_m/m сходится к распределению случайной величины X. Тогда оптимальная вероятность выбора лучшего кандидата (точнее, ее предел при $m \to \infty$) запишется в виде

$$p_X = \max_x E\,(f\,(x/X)),$$

где $f(x) = \max(0, x \ln x)$. Вероятность p_X может оказаться очень маленькой, так как $\inf\limits_X p_X = 0$.

(Лит.: Chow Y., Robbins H., Seigmund D. *Great Expectations: The Theory of Optimal Stopping*, Houghton Miffin Co., Boston, 1971. [Имеется перевод: Роббинс Г., Сигмунд Д., Чао И. Теория оптимальных правил остановки. — М.: Наука, 1977.]
Freeman P. R. "The secretary problem and its extensions: A review", *Internat. Statist. Review*, 51, 189—206, (1983).
Березовский Б. А., Гнедин А. В. Задача наилучшего выбора проблем. — М.: Наука, 1984.)

ж) Парадокс Пинскера о стационарных процессах

Последовательность случайных величин X_n ($n = \ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$) называется стационарной (точнее, стационарной в широком смысле), если, во-первых, математическое ожидание случайной величины X_n не зависит от n (следовательно, не теряя общности, можно предполагать, что общее математическое ожидание равно 0) и, во-вторых, ковариации $E(X_n X_m) = r_{n-m}$ (предполагаем, что они существуют) зависят только от разности $n - m$ (в частности, при $n = m$ дисперсии не зависят от n). Последовательность случайных векторов X_n ($n = \ldots, -2, -1, 0, 1, 2, \ldots$) стационарна, если $E(X_n)$ тождественно равно нулевому вектору, и математическое ожидание произведения i-й координаты вектора X_n и j-й координаты вектора X_m зависит только от i, j и $n - m$. Двумя основными типами стационарных процессов являются сингулярные и регулярные процессы. Первые детерминированы (т. е для любого значения n величина X_{n+1} не содержит какой-либо «информации», не связанной со случайными величинами, предшествующими X_{n+1}), у регулярных же процессов нет детерминированной составляющей (т. е. при отбрасывании величин X_n, X_{n-1}, X_{n-2} и т. д. постепенно теряется вся информация). Таким образом, мир сингулярных процессов предсказуем, и с течением времени новой информации не возникает, регулярные же процессы создают новый мир из ничего, т. е. далекое будущее почти не зависит от настоящего. В гильбертовом пространстве случайных величин, интегрируемых в квадрате, т. е. имеющих конечные дисперсии, приведенное выше утверждение можно сформулировать следующим образом. Если H_n обозначает подпространство, порожденное случайными величинами, предшествующими X_n, то в сингулярном случае $H_n = H_{n-1}$ при всех n, а в регулярном случае $\bigcap\limits_n H_n = 0$. Важность сингулярных и регулярных процессов показана в теореме Вольда, которая утверждает, что любой стационарный процесс может быть представлен и, при том единственным образом, в виде суммы регулярного и сингулярного процессов. Достаточно очевидно, что если процесс X_n сингулярен, то X_{-n} также сингулярен, и если X_n регулярен, то и X_{-n} регулярен. Иными словами, считая n переменной времени, можно сказать, что сингулярность и регулярность остаются неизменными, когда прошлое и будущее меняются местами. Удивительно, но это верно только тогда, когда процесс X_n скалярен. Пинскер построил двумерный регулярный стационарный процесс, обратный к которому (когда $-n$ заменяется на n) уже сингулярен. Таким образом, сингулярность может превратиться в регулярность и наоборот, если прошлое и будущее поменяются местами.

з) Парадоксы голосования и выборов; случайные поля

При голосовании или выборах исход, как правило, случаен, поэтому неудивительно, что и в этой области получены важные вероятностные результаты. В 1878 г. *В. Уитворт* доказал следующую знаменитую теорему о баллотировке. Предположим, что было два кандидата, скажем A и B, A получил n голосов, B — m голосов и $n > m$ (т. е. A победил). Пусть p обозначает вероятность того, что при последовательном подсчете голосов A все время был впереди (при условии, что любой порядок подсчета голосов равновозможен с любым другим). Тогда

$$p = \frac{n-m}{n+m}.$$

Таким образом, если $n = 2m$, то $p = 1/3$, т. е. если A получил вдвое больше голосов, чем B, то вероятность того, что во время подсчета голосов были моменты, когда у B было столько же голосов, сколько у A, вдвое больше вероятности того, что A все время был впереди. (См. Feller W. *Probability Theory and Its*

Applications (2nd ed.), Willey, New York, 1965, p. 66. [Имеется перевод: Феллер В. Введение в теорию вероятностей и ее приложения. Т. 1. — М.: Мир, 1984, с. 87.]) Этот факт может показаться странным, но он все-таки не является парадоксом. Однако в этой области встречаются и парадоксы. Маркиз Кондорсе (один из друзей Вольтера) в 1758 г. привел следующий пример. (Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix.) Предположим, что в выборах участвовали три кандидата A, B и C, и они набрали соответственно 23, 19 и 18 голосов. Тогда кандидата A, как набравшего наибольшее число голосов, следует объявить победителем, но в действительности все 19 избирателей, голосовавших за B, возможно, предпочли бы C вместо A. В 1950 г. Кеннет Арроу (в 1972 г. ему присуждена Нобелевская премия по экономике) использовал этот пример, чтобы показать, что логически невозможно создать совершенно справедливую систему выборов. Таким образом, неудивительно, что нет единой системы выборов, принятой во всем мире. (О вероятностной противоречивости системы выборов в США см. статью Грофмана.) Следующий парадокс относится к специфическому виду голосования — судебному. Предположим, что есть пять присяжных заседателей A, B, C, D и E. Большинством голосов они решают вопрос о виновности или невиновности подсудимого. С вероятностями 5 % заседатели A и B выносят неверное решение, для C и D эти вероятности составляют 10 %, и E ошибается с вероятностью 20 %. (Заседатели ошибаются независимо друг от друга.) В этом случае вероятность вынесения неверного приговора приблизительно равна 0.7 %. Парадоксально, но эта вероятность возрастает и становится приблизительно равной 1.15 %, если заседатель E (который ошибается чаще всего) перестанет рассуждать самостоятельно, а будет всегда повторять решение заседателя A (который ошибается наиболее редко). Следующий парадокс также иллюстрирует те удивительные ситуации, которые возникают, когда избиратели перестают рассуждать самостоятельно. Предположим, что в каждом узле квадратной решетки на плоскости находится человек, который может голосовать независимо от других людей за или против с соответствующими вероятностями p и $1-p$. Однако при следующем голосовании каждый человек выбирает одного из своих четырех соседей и голосует так, как этот сосед голосовал в предыдущий раз. Третье, четвертое и т. д. голосования проводятся аналогично. (При n-м голосовании каждый голосует так, как выбранный сосед голосовал при $(n-1)$-м голосовании.) Вопрос заключается в следующем: что произойдет при $n \to \infty$? Можно показать, что в конце концов все будут голосовать одинаково, иными словами, наступит «полная гармония». (Каждый голосует за или против с соответствующими вероятностями p и $1-p$.) Следует отметить, что если избиратели находятся в узлах трехмерной кубической решетки (и у каждого по шесть соседей), то такая экстремальная ситуация не наступит, т. е. различные мнения могут гармонировать друг с другом (точнее получится эргодическое предельное распределение). То же самое справедливо в случае, когда размерность больше трех. Это основное различие между двумерным и трехмерным случаями тесно связано с тем фактом (см. III/5а), что в случае двумерной квадратной решетки при симметричных случайных блужданиях каждый узел достигается с вероятностью 1, а в трехмерном случае это неверно. (См. Bramson M., Griffeath D. «Renormalizing the 3-dimensional voter model», *Annals of Prob.*, 418—432, (1972).)

За последние годы приведенная выше математическая модель избирателей, находящихся в узлах квадратной или кубической решетки, стала играть очень важную роль в математической физике. Избиратели заменяются «величинами» с двумя возможными значениями (например, спинами электронов в ферромагнитных материалах). Такие *случайные поля* являются обобщениями стохастических процессов, в которых переменная времени t заменяется на элемент многомерного пространства, например, если t означает узлы d-мерной кубической решетки, и $X(t)$ — случайная величина при любом фиксированном t (в модели голосования $X(t)$ принимает только два значения), то $X(t)$ — случайное поле. Так же, как выше мы предполагали, что на мнение избирателя влияют его соседи, в физике мы можем считать (в виде первого приближения), что на каждую частицу влияют ее соседи. Такой вид случайного поля называется марковским полем (это аналог цепи Маркова). Благодаря исследованиям норвежского физика и химика *Л. Онсагера*, проведенным в 1944 г., при изучении ферромагнетизма важную роль стало играть марковское поле специального вида — модель Изинга. За последние несколько лет марковские поля и, в особенности, модель Изинга, применялись для решения проблемы фазовых переходов. Хотя строгое определение марковского поля было дано лишь в 1968 г. советским математиком *Р. Л. Добрушиным*, первые описания понятия фазы и некоторых случайных полей появились значительно раньше в связи с потенциальными функциями в книге *Дж. В. Гиббса*, опубликованной в 1902 г. (Gibbs J. W. Elementary Principles of Statistical Mechanics, Yale Univ. Press). [Имеется перевод: Гиббс Дж. В. Термодинамика. Статистическая механика. — М.: Наука, 1982.]) Описание марковских полей через потенциальные функции особенно важно потому, что фазовые переходы происходят как раз тогда, когда потенциал не определяет однозначно марковское поле. В физических терминах это означает, что при одной и той же температуре могут быть разные фазовые состояния. Теория также объясняет, почему фазовые переходы невозможны, когда температура выше критической (критическую температуру удалось определить уже Онсагеру). Интересно отметить, что хотя в одномерной модели фазовый переход произойти не может, в случае двумерной модели (на квадратной решетке) это уже возможно. В последнем случае, несмотря на симметричность потенциальной функции (ее значение не меняется, когда все состояния «да» меняются на «нет» и наоборот), само марковское поле несимметрично. В силу этого парадокса (называемого парадоксом нарушения симметрии) ферромагнитные материалы не утрачивают свой магнетизм при температуре ниже критической.

(Лит.: Grofman B. "Fair appointment and the Banzhaf index", *The Amer. Math. Monthly*, **88**, 1—5, (1981).
Kindermann R., Snell J. L. *Markov Random Fields*, Contemporary Math. Vol. 1, AMS, Providence RI, 1980.
Preston C. J. *Gibbs States on Countable Sets*, Cambridge Univ. Press, 1974.
Синай Я. Г. Теория фазовых переходов: Строгие результаты. — М.: Наука, 1980.)

ГЛАВА IV

ПАРАДОКСЫ В ОСНОВАНИЯХ ТЕОРИИ ВЕРОЯТНОСТЕЙ. РАЗНЫЕ ПАРАДОКСЫ

> Природе разума свойственно рассматривать вещи не как случайные, но как необходимые.
>
> *Б. Спиноза, Этика, часть 2, теорема XLIV*[1])

> Вероятность — это важнейшее понятие в современной науке особенно потому, что никто совершенно не представляет, что оно означает.
>
> *Бертран Рассел, Из лекции, 1929 г.*

> Исчисление вероятностей. Первая лекция. 1. Едва ли можно дать удовлетворительное определение вероятности...
>
> *А. Пуанкаре, Исчисление вероятностей, 1896 г., с. 1*

> Мой тезис, парадоксальный и немного провокационный, но вместе с тем искренний, заключается всего лишь в следующем: ВЕРОЯТНОСТЬ НЕ СУЩЕСТВУЕТ.
>
> *Б. де Финетти, Теория вероятностей, 1974 г.*

В 1900 г. на Международном математическом конгрессе в Париже *Давид Гильберт* среди 23 важнейших нерешенных проблем в математике назвал проблему построения оснований теории вероятностей. Хотя на рубеже веков теория вероятностей уже дала много выдающихся результатов, из-за отсутствия математических оснований эта теория в целом еще не была одним из разделов математики. В этом заключалась, видимо, главная причина, почему профессор университета в Геттингене *К. Клейн* даже не упомянул теорию вероятностей в своей работе «Лекции о развитии математики в XIX столетии». Используя результаты работ многих математиков, в особенности *Э. Бореля*, *А. Ломницкого*, *Х. Штейнгауса*, а также теорию множеств и теорию меры, *А. Н. Колмогоров* в 1933 г. построил математически строгую теорию вероятностей. (Более подробно см. в *Archive for Hist. of Exact Sci.*, **18**, 123—190, 1978.) В основе теории, разработанной Колмогоровым, лежит тот факт, что любое событие, вероятность которого хотим найти (такие события называются наблюдаемыми), может быть представлено в виде некоторого подмножества множества элементарных событий (т. е. в виде подмножества *фазового пространства*). Например, при бросании

---

[1]) Спиноза Б. Избранные произведения. Т. 1. Этика. Пер. с лат. Н. А. Иванцова. — М.: Гос. изд-во политич. литературы, 1957.

игральной кости исходами могут быть числа 1, 2, ..., 6. Совокупность этих элементарных событий образует фазовое пространство и событие, заключающееся в выпадении четного числа, представимо в виде подмножества фазового пространства, состоящего из четных чисел $\{2, 4, 6\}$. Достоверное событие представимо всем фазовым пространством, которое традиционно обозначается через Ω. В теории Колмогорова предполагается, что наблюдаемые события образуют сигма-алгебру (здесь «сигма» указывает на бесконечность), т. е. совместное наступление двух произвольных наблюдаемых событий; наступление по крайней мере одного из конечного или счетного числа наблюдаемых событий и дополнение к любому наблюдаемому событию также являются наблюдаемыми событиями. Каждому наблюдаемому событию приписывается некоторое неотрицательное число, называемое вероятностью этого события, таким образом, чтобы вероятность достоверного события (т. е. всего фазового пространства) была равна 1, и выполнялось свойство сигма-аддитивности, т. е. в случае попарно исключающих друг друга событий вероятность наступления по крайней мере одного (и, следовательно, в силу попарной несовместимости, ровно одного) наблюдаемого события в конечной или счетной совокупности наблюдаемых событий совпадала с суммой вероятностей наблюдаемых событий из этой совокупности.

Возникает вопрос: почему при определении вероятности нам потребовались сигма-алгебры, а не множество всех подмножеств фазового пространства Ω? Ответ очень простой: в общем случае вероятность нельзя определить на множестве всех подмножеств Ω, точнее, если вероятность определена на сигма-алгебре, состоящей из некоторых подмножеств Ω, то эту вероятность нельзя продолжить на остальные подмножества Ω так, чтобы сохранялось свойство сигма-аддитивности (если только Ω не состоит из конечного или счетного числа элементов). *Дж. Витали* знал этот результат еще в 1905 г.

Пусть фазовым пространством будет интервал $(0, 1)$, и попробуем определить вероятность на всех его подмножествах так, чтобы получить «равномерное распределение». Очевидно, вероятность $b-a$ следует приписать подынтервалу (a, b). Таким образом, в силу сигма-аддитивности вероятность автоматически определена на наименьшей сигма-алгебре, содержащей интервалы. Эту вероятность можно продолжить на некоторые другие множества, но существуют множества, на которые продолжение невозможно, т. е. на которых вероятность нельзя определить в соответствии с «равномерным распределением».

Пример такого «патологического» множества был построен *Э. Цермело* следующим образом. Он разбил точки интервала

(0, 1) на непересекающиеся классы так, что точки, расстояние между которыми равно какому-то рациональному числу, принадлежат одному и тому же классу. Тогда, используя аксиому выбора, Цермело определил множество H, которое содержит в точности по одной точке из каждого такого класса. Можно доказать, что на этом множестве H нельзя определить вероятность, соответствующую «равномерному распределению».

Можно также показать, что если мы откажемся от «равномерности», но потребуем, чтобы каждое подмножество Ω имело вероятность, и вероятность каждой точки из Ω равнялась 0, то даже такое определение вероятности невозможно в случае, когда фазовое пространство Ω счетно или имеет мощность континуума при условии, что справедлива континуум-гипотеза (см. Birkhoff G. *Lattice Theory*, Amer. Math. Soc., Providence, 1967, p. 266). До сих пор неизвестно, существует ли пространство Ω (достаточно большой мощности), для которого можно определить вероятность, удовлетворяющую указанным выше условиям. В этом состоит так называемая *проблема измеримых мощностей.* Ситуация сильно изменится, если отказаться от аксиомы выбора: см. Jech T. *Set Theory*, Acad. Press, New York, 1978 и Solovay R. M. A model of set theory in which every set of reals is Lebesgue measurable, *Annals of Math.*, 1—56, (1970).

Хотя в теории Колмогорова вероятность всегда неотрицательна, некоторые теоремы в теории вероятностей можно обобщить на случай, когда отрицательные числа выступают как вероятности. Например К. Хохберг (Hochberg K. J. *Proc. Amer. Math. Soc.*, **79**, 298—302, 1980) доказал, что в теоремах, полученных в результате такого обобщения центральной предельной теоремы, возникают действительные (положительные и отрицательные) «плотности вероятности» $u_n(t, x)$, которые можно вывести из фундаментальных решений следующего обобщения дифференциального уравнения теплопроводности

$$\frac{\partial u}{\partial t} = (-1)^{n+1} \frac{\partial^{2n} u}{\partial x^{2n}},$$

$n = 2, 3, \ldots$ (при $n = 1$ имеем обычное дифференциальное уравнение теплопроводности). Однако в настоящей книге не рассматриваются отрицательные и комплекснозначные вероятностные меры и не обсуждаются другие обобщения вероятности.

1. Парадоксы случайных натуральных чисел

а) История парадоксов

В теории вероятностей Колмогорова невозможно выбрать отдельное натуральное (целое положительное) число из множества натуральных чисел случайным образом в смысле равномерного распределения, так как если вероятность выбора, например, единицы равна 0, то в силу равномерности вероятность выбора любого другого натурального числа также равна 0. Таким образом, свойство сигма-аддитивности приводит к противоречию, поскольку вероятность выбора натурального числа равна 1, а не 0. С другой стороны, если вероятность выбора единицы положительна, то сигма-аддитивность снова приводит к противоречию (тогда вероятность достоверного события бесконечна). Несмотря на этот факт, естественно ожидать, что вероятность выбора нечетного или четного числа составляет 1/2. Следующее определение (в котором нет требования сигма-аддитивности) приводит именно к этой вероятности. Пусть K — произвольное подмножество натуральных чисел, и k_n обозначает число элементов в K, которые не больше n. Относительная частота k_n/n дает вероятность выбора числа из K при условии, что выбор производится из первых n чисел случайно и равномерно. Если существует предел относительных частот k_n/n при n, стремящемся к бесконечности, то этот предел называется вероятностью множества K. По этому определению вероятности выбора целого числа, которое делится на 2, 3 и т. д., равны соответственно 1/2, 1/3, ... Можно также легко вычислить вероятность того, что два случайных целых числа (распределенных равномерно и выбранных независимо друг от друга) взаимно просты. Сначала предположим, что оба этих числа не больше n. Тогда вычисляется соответствующая вероятность (зависящая от n) и затем находится ее предел при $n \to \infty$. Еще в прошлом веке *Чебышев* показал, что этот предел равен $6/\pi^2 (\approx 2/3)$. Следовательно, если числитель и знаменатель дроби являются случайными натуральными числами, то дробь несократима с вероятностью $6/\pi^2$. Следующие парадоксы также связаны со случайными натуральными числами. По утверждению *Дж. Литлвуда* первый парадокс принадлежит знаменитому физику *Э. Шредингеру*. В статье *Ф. Кантелли*, опубликованной в 1935 г., второй парадокс приписывается *П. Леви*. (П. Леви был одним из самых выдающихся специалистов по теории вероятностей. В Парижской академии наук он занял место Пуанкаре и Адамара.)

б) Парадоксы

(i) На лбу у каждого из игроков A и B пишутся по одному из двух последовательных случайных натуральных чисел. Игрок с меньшим числом проигрывает и должен заплатить другому сумму в долларах, равную числу, написанному на его собственном лбу. Оба игрока имеют право вето, т. е. посчитав, что число на лбу соперника слишком велико, любой из них может потребовать проведения новой игры. (Игроку, естественно, неизвестно то число, которое написано на его собственном лбу.) Однако никто из игроков не хочет пользоваться правом вето по следующей причине. Каждый из них рассуждает: «На лбу соперника я вижу число k. Следовательно, у меня на лбу либо $k-1$, либо $k+1$. Каждое из чисел равновозможно, но при проигрыше я заплачу $k-1$ долларов, а при выигрыше получу k долларов, поэтому лучше правом вето не пользоваться». Поскольку ожидаемое значение выигрыша положительно, игра представляется выгодной для обоих игроков, что, конечно, невозможно.

(ii) Выберем два независимых случайных равномерно распределенных натуральных числа X и Y. Для любого фиксированного (неслучайного) числа x вероятность того, что $Y \leqslant x$, равна 0. Аналогично, для любого фиксированного y вероятность события $X \leqslant y$ равна 0. Следовательно, также с нулевой вероятностью одновременно выполняются неравенства $Y \leqslant X$ и $X \leqslant Y$, что невозможно, так как одно из них всегда справедливо.

в) Объяснение парадоксов

(i) Парадокс иллюстрирует тот факт, что на множестве натуральных чисел нельзя определить равномерное распределение. Если числа, которые пишутся у игроков на лбу, меньше тысячи, то на таких числах уже можно задать равномерное распределение, но тогда рассуждения, которые были выше и привели к парадоксу, станут совершенно ложными.

(ii) Вероятность события $Y \leqslant x$ при любом фиксированном x, несомненно, равна 0 (в силу определения, о котором говорилось в истории парадоксов), но отсюда не следует, что нулевой является также вероятность того, что $Y \leqslant X$. Это было бы так только в том случае, когда вероятность сигма-аддитивна, но вероятность, о которой сейчас идет речь (как мы уже отмечали), не сигма-аддитивна.

г) Замечания

(i) Теория чисел и теория вероятностей тесно взаимосвязаны. Для иллюстрации того, как вероятностные идеи можно использовать в теории чисел, сначала вспомним, что относительная частота простых чисел среди целых чисел, не превосходящих n, приблизительно составляет $1/\ln n$ (при достаточно больших n). Предположим, что простые числа распределены случайно и независимо друг от друга среди первых n чисел. Тогда вероятность выбора двух простых чисел близких к n приблизительно равна $1/(\ln n)^2$ (в силу их независимости). Рассмотрим интервал длины c, содержащий n (c мало по сравнению с n, но достаточно велико для получения статистических выводов). Из предыдущего результата следует, что число близнецов (простых чисел, разность между которыми равна 2), принадлежащих этому интервалу, примерно равно $c/(\ln n)^2$. Более тщательный анализ (принимающий во внимание, например, тот факт, что целое число, отличающееся от простого числа ($\neq 2$) на 2, очевидно, нечетно и поэтому скорее всего само является простым) показывает, что ожидаемое число близнецов приблизительно на 32 % больше, чем $c/(\ln n)^2$. Исходя из этого, *М. Джонс*, *М. Лал* и *У. Бландон* в 1967 г. опубликовали в «Вычислительной математике» таблицу, которая показывает, например, что среди первых 150 тысяч чисел после 100 миллионов ожидаемое число близнецов равно 584. В действительности их 601. Разница сравнительно мала. Аналогично, при рассмотрении первых 150 тысяч чисел после 10^{14} ожидается встретить среди них 191 близнеца, на самом деле их 186. Такой «статистический» подход к изучению простых чисел дает относительно хорошие результаты и представляет особый интерес, так как до сих пор неизвестно конечно или бесконечно множество близнецов. (Наибольшее простое число, известное к настоящему моменту, равно $2^{86243}-1$.) Простые числа следуют друг за другом по очень сложному правилу, которое кажется случайным. Поэтому вероятностный подход здесь наиболее приемлем. Мы вернемся к связи между сложностью и случайностью в «парадоксе метода Монте-Карло».

(ii) *Для конечно-аддитивного равномерного распределения* вероятность любого конечного подмножества A (натуральных чисел) равна 0. Предположим теперь, что распределение не равномерное, но обладает тем свойством, что для произвольного заданного положительного числа ε найдется конечное множество A, имеющее вероятность $P(A) > 1-\varepsilon$. Тогда различие между аддитивностью и сигма-аддитивностью исчезает. Точнее, если вероятность P аддитивна на конечных подмножествах натуральных чисел (или на произвольном счетном множестве Ω), то эту вероятность можно продолжить на любое подмножество таким образом, что продолжение будет сигма-аддитивным на сигма-алгебре всех подмножеств натуральных чисел.

Если множество Ω — несчетное (например, Ω — весь интервал $(0, 1)$), то могут возникать довольно странные аддитивные вероятности. Они могут принимать только значения 0 и 1 и определяться на всех подмножествах интервала $(0, 1)$ (в этом случае требуется аксиома выбора). Эти вероятности являются странными, так как счетное число событий, имеющих вероятность 1, вряд ли произойдет одновременно, т. е. вероятность этого события может равняться 0. Аналогично, по крайней мере одно из счетного числа событий, имеющих нулевую вероятность, может происходить с вероятностью 1.

д) Литература

Elliot P. *Probabilistic Number Theory*, Springer, New York, 1980.

Kac M. "Statistical independence in probability, analysis and number theory", Carus, *Math. Monographs*, **12**, Wiley, New York, 1959. [Имеется перевод: Кац М. Статистическая независимость в теории вероятностей, анализе и теории чисел. — М.: ИЛ, 1963.]

Rényi A. "On a new axiomatic theory of probability", *Acta Math. Acad. Sci. Hung.*, **6**, 285—335, (1959).

В следующей статье содержится парадоксальный подход к одной знаменитой нерешенной проблеме. В статье утверждается, что гипотеза *Римана* (в которой случайности нет!) справедлива с вероятностью единица:

Good I. J., Churchhouse R. F. "The Riemann hypothesis and pseudorandom features of the Möbius sequence", *Mathematics of Computation*, **22**, 857—864, (1968).

Вероятностный аналог известной теоремы Дирихле о существовании бесконечно большого числа простых чисел во всякой арифметической прогрессии из целых чисел обсуждается в статье

Ruzsa I. Z., Székely G. J. "Intersections of traces of random walks with fixed sets", *Annals of Probability*, **10**, 132—136, (1982).

2. Парадокс Банаха — Тарского

а) История парадокса

Равномерное распределение, которое соответствует длине, площади и объему в одно-, дву- и трехмерных пространствах, нельзя определить на всех множествах, если требуется *сигма-аддитивность меры*. Однако польский математик *С. Банах* показал, что если предполагать только *аддитивность* (т. е. мера объединения двух несовместных событий равна сумме мер этих событий), то в одно- и двумерном пространствах ограниченное множество становится измеримым (имеет длину или площадь). Таким образом, в одно- и двумерном случаях равномерное распределение можно определить на любом (ограниченном) множестве, если от вероятностей требовать только аддитивность. Вместе с тем в 1914 г. *Хаусдорф* доказал, что такое продолжение мер в трехмерном пространстве невозможно. В 1924 г. *С. Банах* и *А. Тарский* доказали парадоксальную теорему, в которой красочно показано, что ни аддитивная мера (объем), ни соответствующее равномерное распределение не могут быть определены на всех ограниченных множествах трехмерного пространства.

б) Парадокс

Рассмотрим шар радиуса $r = 1$ см. Его можно разделить на некоторое конечное число частей и затем, перегруппировав эти части и собрав вновь, образовать шар радиуса $R = 1$ км. В общем случае, если A и B — ограниченные подмножества в R^3, имеющие непустую внутренность, то существуют натуральное число n и разбиения $\{A_j: 1 \leqslant j \leqslant n\}$ и $\{B_j: 1 \leqslant j \leqslant n\}$ на n частей соответственно множеств A и B такие, что A_j конгруэнтно B_j для всех j. (Подмножество X в R^3 ограничено, если оно содержится в некотором шаре, и X имеет непустую внутренность, если оно содержит некоторый шар. Разбиение множества X означает совокупность попарно не пересекающихся подмножеств X, объединение которых совпадает с X.)

в) Объяснение парадокса

Разделив шар радиуса $r = 1$ см на конечное число частей, мы интуитивно ожидаем, что, складывая эти части вместе, можно получить только сплошные фигуры, объем которых равен объему исходного шара радиуса в 1 см. Однако это справедливо только в случае, когда шар делится на части, имеющие объем. Суть парадокса заключается в том, что в трехмерном пространстве существуют неизмеримые множества, которые не имеют объема, если под объемом мы понимаем то, что обладает свойством аддитивности, и предполагаем, что объемы двух конгруэнтных множеств совпадают. (В доказательстве теоремы Банаха — Тарского используется аксиома выбора Цермело.)

г) Замечания

Некоторые выдающиеся математики (например, итальянский ученый де Финетти) считают, что сигма-аддитивность вероятности — это слишком сильное ограничение, но признают аддитивность. Парадокс Банаха — Тарского показывает, что при замене сигма-аддитивности на аддитивность не только не решаются все проблемы, но возникают новые. В теории автоматов дилемма — принимать или не принимать сигма-аддитивность — стала настолько критической, что даже Британская энциклопедия приняла участие в ее обсуждении. Электронные вычислительные машины часто используются для получения (теоретически) бесконечных последовательностей случайных чисел (см. следующий парадокс). Вероятность каждой последовательности нулевая, но вероятность их объединения равна 1. Таким образом, принятие сигма-аддитивности основано на молчаливом предположении, что автомат не может породить случайное явление, т. е. случайные и неслучайные последовательности разделяются так же, как и в греческой мифологии: там есть различные богини — Тюхэ (богиня случая, удачи) и Мойры (богини судьбы).

д) Литература

Banach S., Tarski A. "Sur la décomposition des ensembles de points en parties respectivement congruentes", *Fund. Math.*, **6**, 244—277, (1924).

Stromberg K. "The Banach — Tarski paradox", *The American Math. Monthly*, **86**, 151—160, (1979).

3. Парадокс метода Монте-Карло

а) История парадокса

Метод Монте-Карло — численный метод, основанный на случайной выборке. При решении вычислительных задач часто можно найти подходящую вероятностную модель, в которую входит искомое неизвестное число. Затем для решения задачи много раз наблюдаются исходы случайных экспериментов, включенных в вероятностную модель, с тем чтобы с заданной точностью (на основе наблюденных значений) можно было оценить искомое число. Хотя идея этого метода довольно стара, его настоящее применение началось лишь с появлением компьютеров, когда *Е. Нейман*, *С. Улам* и *Э. Ферми* использовали метод Монте-Карло для приближенного решения трудных вычислительных задач, связанных с ядерными реакциями. Название метода объясняется тем, что в нем применяются последовательности случайных чисел, в качестве которых могли бы выступать регулярно объявляемые результаты игр, проводимых в казино, например, в Монте-Карло. Однако на практике случайные числа, необходимые для метода, выдает сам компьютер. Следовательно, симпатичное название (его впервые использовали в 1949 г. *Н. Метрополис* и *С. Улам*) вводит в заблуждение (метод вряд ли поможет выиграть в Монте-Карло). Идея метода Монте-Карло впервые появилась в 1777 г. в работе *Бюффона* (см. I. 11), где излагался метод оценки числа π путем бросания иголки наугад. Предположим, что на столе проведены параллельные прямые на единичном расстоянии друг от друга, и на стол наугад бросается иголка длиной $L < 1$, при этом угол между прямыми и иглой и расстояние от середины иглы до ближайшей прямой являются независимыми случайными величинами, равномерно распределенными соответственно на $(0, 2\pi)$ и $(-1/2,\ 1/2)$. Тогда игла пересечет какую-нибудь прямую с вероятностью $2L/\pi$. Если проводить эксперимент много раз, то относительная частота пересечений будет очень близка к теоретической вероятности $2L/\pi$, и таким путем можно вычислить значение π. Этот метод нахождения приближенного значения π имеет чисто теоретическое значение, так как для получения двух точных знаков после запятой нужно совершить несколько тысяч бросаний. (С помощью другого метода можно определить миллион знаков числа π, см. статью *Г. Мила*.) Задача Бюффона об игле показывает, что метод Монте-Карло не подходит для очень точных вычислений. Даже для получения результатов с точностью до двух или трех знаков требуется проведение тысяч или миллионов экспериментов. Следовательно, метод Монте-Карло применим только тогда, когда проведение экспериментов моделируется компьютером. Вместо бросания иглы выдаются два независимых случайных числа, которые определяют положение (предполагаемой) иглы и произошло ли ее пересечение с (предполагаемыми) прямыми. Поскольку компьютер способен выдавать несколько миллионов чисел в минуту, моделирование миллионов экспериментов не займет слишком много времени; без компьютера для этого потребовалась бы вся жизнь.

Теория построения случайных чисел на компьютерах превратилась в важное направление в математике. Вместо настоящих случайных чисел (которые возникают в ходе случайных физических процессов, например, в ходе радиоактивного распада) популярными становятся псевдослучайные числа, конструируемые с помощью детерминированных вычислительных алгоритмов.

В связи с псевдослучайными числами возникает следующий вопрос. В каком смысле их можно считать случайными, если они получены с помощью детерминированных (неслучайных) алгоритмов? После статьи *фон Мизеса*, вышедшей в 1919 г., некоторые выдающиеся математики исследовали эту проблему. (Философскими аспектами проблемы занимались *П. Киршенманн*, *П. Макшейн* и другие.)

б) Парадокс

В 1965—1966 гг. *Колмогоров* и *Мартин-Лёф* представили понятие случайности в новом свете. Они определили, когда последовательность, состоящую из 0 и 1, можно считать случайной. Основная идея состоит в следующем. Чем сложнее описать последовательность (т. е. чем длиннее «самая короткая» программа, конструирующая эту последовательность), тем более случайной ее можно считать. Длина «самой короткой» программы, естественно, различна для разных компьютеров. По этой причине выбирают стандартную машину, называемую машиной Тьюринга. Мерой сложности последовательности является длина наиболее короткой программы на машине Тьюринга, которая генерирует эту последовательность. Сложность — мера иррегулярности. Последовательности, длина которых N, называются случайными, если их сложность близка к максимальной. (Можно показать, что большинство последовательностей именно таковы.) Мартин-Лёф доказал, что эти последовательности можно считать случайными, так как они удовлетворяют всем статистическим тестам на случайность. Таким образом, сложность и случайность тесно взаимосвязаны. Если программист собирается получить «настоящие» случайные числа, то в силу результатов Колмогорова и Мартин-Лёфа он может это сделать только с помощью достаточно длинной программы. В то же время на практике генераторы случайных чисел очень короткие. Как совместить эти два факта?

в) Объяснение парадокса

Последовательности, генерируемые короткими программами и используемые в качестве случайных, в действительности удовлетворяют лишь некоторым, а не всем тестам на случайность. Однако в приложениях это почти не создает проблем. Например, для численного интегрирования достаточно иметь псевдослучайные числа, равномерно распределенные на некотором интервале.

Предположим, что надо проинтегрировать функцию ограниченной вариации на интервале (0, 1). Величина

$$I = \int_0^1 f(x)\,dx$$

приближается арифметическим средним

$$I_N = \frac{1}{N}\sum_{i=1}^{N} f(x_i)$$

не только тогда, когда последовательность $x_1, x_2, \ldots, x_N$ случайна и равномерно распределена на интервале (0, 1). Достаточно потребовать равномерную распределенность этой последовательности. Это означает, что при $N \to \infty$

$$D_N = \sup_{0<x<1} |c(x, N) - x| \to 0,$$

где $c(x, N)$ — отношение числа элементов последовательности $x_1, x_2, \ldots, x_N$, принадлежащих $(0, x)$, к N, т. е. относительная частота.

Можно показать, что

$$|I - I_N| \leqslant V_f D_N,$$

где V_f — постоянная, зависящая от функции f (полная вариация функции f). Отсюда следует, что приближение для I тем точнее, чем меньше D_N. Однако значение D_N минимально не тогда, когда берут случайную последовательность. Для случайной последовательности порядок приближения равен $N^{-1/2}$, а для неслучайной последовательности можно получить точность порядка $N^{-1}\log N$.

Часто вместо того, чтобы пытаться справиться с «неосязаемым понятием случайности», следует использовать детерминированные последовательности, которые лучше подходят для решения конкретной задачи. В этом суть квазиметода Монте-Карло.

г) Замечания

(i) Взаимосвязь между случайностью и сложностью недавно привела к нескольким интересным открытиям. В течение долгого времени в математике существует общая практика обращения с очень сложными структурами как со случайными (например, поведение сложных последовательностей простых чисел часто описывается вероятностными законами). Идея о том, что случайность и сложность неразличимы, является настолько революционной, что она очень важна и с философской точки зрения. Используя эту идею, девиз Спинозы можно сформулировать следующим образом: люди предпочитают простые вещи сложным — это, безусловно, верно. В то же время очевидно, что чем глубже мы пытаемся понять законы природы, тем яснее осознаем, что далеко не все они простые.

(ii) Последовательности случайных чисел применяются достаточно широко. Они нужны для численного интегрирования, для численного решения дифференциальных уравнений, для моделирования на ЭВМ физических, химических, биологических, технических и экономических задач и т. д. Они помогают при решении задач уличного движения, транспортных и других оптимизационных задач, а также при создании астрономических моделей. С помощью случайных чисел можно проверять эффективность различных программ на ЭВМ.

В заключение следует отметить совершенно иную область применения — компьютерное искусство, где последовательности случайных чисел предлагают миллионы вариаций (последовательности случайных чисел могут, очевидно, соответствовать последовательностям звуков, цветов, букв и т. д.). Из этих последовательностей компьютер оставляет только те, которые удовлетворяют установленным правилам. Если компьютер работает с достаточно большим числом выборок, то художественное произведение будет очень хорошим. Например, греческий композитор Ксенакис использовал в своих работах звуки, полученные случайным образом с помощью ЭВМ. Уже проведено несколько выставок компьютерной графики. (Необходимо отметить, что не всякая компьютерная графика использует случайные последовательности.) В 1970 г. образована международная организация компьютерных художников.

д) Литература

Kirschenmann P. "Concept of randomness", *J. Phil. Logic*, **1**, 395—414, (1972).

Knuth D. E. The Art of Computer Programming, (Chapter 3 — Random numbers), Addison-Wesley, 1969. [Имеется перевод: Кнут Д. Искусство программирования для ЭВМ. Т. 1 (Гл. 3 — Случайные числа). — М.: Мир, 1976.]

Martin-Löf P. "The definition of random sequences", *Information and Control*, **9**, 602—619, (1966).

McShane P. *Randomness, Statistics and Emergence*, Univ. Notre Dame, 1970.

Metropolis N., Ulam S. M. "The Monte Carlo method", *J. Amer. Statist. Assoc.*, **44**, 335—341, (1949).

Miel G. "An algorithm for calculation of π", *The American Math. Monthly*, **86**, (1979).

Niederreiter H. "Quasi-Monte Carlo methods and pseudorandom numbers", *Bull. Amer. Math. Soc.*, **84**, 957—1041, (1978).

Schnorr C. P. *Zufälligkeit und Wahrscheinlichkeit*, Lecture Notes in Math., **218**, Springer, Berlin — New York, 1971.

Соболь И. М. Метод Монте-Карло. — М.: Наука, 1978.

Székely G. J., Tusnády G. "On the philosophical concept of randomness from a mathematical point of view" (на венгерском языке), *препринт*.

4. Парадокс неинтересных чисел; невычислимая вероятность

а) История парадокса

Мнение о том, является ли некоторое число интересным или неинтересным, полностью субъективно, но, опираясь на предыдущий парадокс, можно дать объективное определение интересного числа. Будем называть число интересным, если его сложность (которая определена в предыдущем разделе) мала. Таким образом, рациональные числа интересны, так как в их записи в десятичной системе знаки повторяются периодически. Среди иррациональных чисел числа π и e также интересны, поскольку цифры в их записи можно получить с помощью довольно простой программы на ЭВМ. Однако существуют менее регулярные иррациональные числа. Например, нормальные числа обладают следующим свойством: в их записи в виде бесконечной десятичной дроби каждый знак (более того, любая группа цифр фиксированного размера) встречается с одной и той же вероятностью. Большинство иррациональных чисел нормальны, но трудно решить, является ли конкретное число нормальным или нет. Поэтому, например, неизвестно, нормально число π (первый миллион знаков которого опубликован в 1974 г.) или e или нет. В то же время существует очень простой (но искусственно построенный) пример нормального числа. В начале тридцатых годов *Д. Шамперноун* показал, что следующее число нормально:

0.123 456 789 101 112 131 415 161 718 192 021 222 324 252 6...

(десятичные знаки являются последовательными натуральными числами). Более ста лет назад похожая ситуация была в арифметике, когда в 1844 г. *Лиувилль* впервые построил трансцендентное число, т. е. число, не являющееся решением никакого алгебраического уравнения с целыми коэффициентами. Трансцендентность чисел π и e доказана позднее, соответственно в 1882 г. *Ф. Линдеманом* и в 1873 г. *Ш. Эрмитом*. Изучение чисел в связи с их нормальностью началось лишь на рубеже веков, во многом благодаря исследованиям *Э. Бореля*. С этого времени, особенно после работ *А. Н. Колмогорова*, *П. Мартин-Лёфа*, *Р. Соломонова* и *Г. Шайтина* исследования по регулярности и иррегулярности последовательностей цифр развились в интересную теорию. Следующий парадокс — один из многих парадоксов в этой области.

б) Парадокс

У большинства чисел цифры следуют друг за другом случайно, т. е. большинство чисел неинтересны в следующем смысле: программы на ЭВМ, которые генерируют эти числа, ненамного короче самих чисел. Несмотря на обилие неинтересных чисел, для большинства из них нельзя доказать, что они неинтересны (в любой системе аксиом, лишенной противоречий). Существует бесконечно много неинтересных чисел, но «неинтересность» можно доказать лишь для конечного числа из них.

в) Объяснение парадокса

Сначала может показаться удивительной возможность существования чего-то такого, что нельзя проверить, однако с подобным явлением мы встречаемся не только в мире математики. Например, если все сто тысяч мест на стадионе заняты, но было продано только девяносто девять тысяч билетов, то ясно, что тысяча человек проникли на стадион без билетов. Однако поиск этих людей безнадежен (особенно, если у каждого на входе билеты отбираются). Таким образом, мы знаем наверняка, что на стадионе присутствует тысяча человек, попавших туда без билета, но мы не можем доказать, что любой конкретный человек прошел на стадион без билета. В математике подобные явления встречаются часто.

Вовсе неудивительно, что большинство чисел неинтересны.

Достаточно вспомнить, как трудно бывает даже в семизначном телефонном номере «обнаружить» какую-либо регулярность, чтобы легче его запомнить. Еще сложнее найти регулярность для большей доли чисел, записываемых сотней или тысячью знаками. Так что более удивительна вторая часть парадокса, особенно для тех, кто недостаточно хорошо знаком с парадоксами логики XX века. Среди этих парадоксов наиболее близок к нашему парадокс *Г. Берри.* (Этот парадокс впервые был опубликован 70 лет назад в работе Б. Рассела и А. Уайтхеда «Principia Matematica».) «Компьютеризированный вариант» парадокса Берри принадлежит *Э. Беккенбаху.* В нем утверждается, что неинтересные натуральные числа не могут существовать, потому что наименьшее из них было бы интересным. Иными словами: наименьшее из чисел, которые строятся только с помощью длинных программ на ЭВМ, можно также генерировать с помощью короткой программы, что, безусловно, является противоречием. Необходимо признать, что некоторые числа являются неинтересными, даже если мы не можем этого доказать. Если система аксиом и правила вывода содержат n бит информации, то можно показать, что свойство «неинтересности» числа нельзя доказать в том случае, когда его количество информации намного больше, чем n бит.

г) Замечание

Важным критерием случайности знаков в записи числа является невозможность их экстраполировать, или предсказывать. Возникает вопрос, существует ли какое-нибудь (неслучайное) число, которое может быть точно определено, но цифры которого нельзя предсказать. Утвердительный ответ дан в примере *Шайтина.* Пусть случайная последовательность из «гербов» и «решек», возникающая при бросании монеты, или соответствующая ей последовательность из нулей и единиц, вводится в некоторую машину, а именно в машину Тьюринга. Вероятность того, что машина Тьюринга остановится в какой-то момент ввода случайной последовательности, и определяет число Шайтина. (Теоретически машина может работать как угодно долго, потому что она не получает команд, заставляющих ее остановиться.) Можно доказать, что число Шайтина «неинтересно» и цифры этого числа предсказать нельзя. В то же время это «неинтересное» число Шайтина обладает рядом интересных свойств. Например, если бы было известно несколько первых тысяч цифр в его десятичной записи, то тем самым мы получили бы решения некоторых классических нерешенных проблем в математике таких, как *великая теорема Ферма* и *проблема Гольдбаха.* Великая теорема Ферма (которая утверждает, что уравнение $x^n + y^n = z^n$ не имеет решений в натуральных числах x, y, z при $n > 2$)[1]) теоретически может быть доказана или опровергнута с помощью компьютерной программы — «программы Ферма», которая для данных значений n и z проверяет, существуют ли целые числа x и y, являющиеся решением уравнения Ферма. При постепенном увеличении значений n и z будут рассмотрены все случаи. Когда решение будет найдено, вычислительная машина остановится. Если машина когда-нибудь остановится, то теорема Ферма будет опровергнута, в противном случае доказана. «Единственная» проблема состоит в следующем: как бы долго уже не работала машина, мы никогда не можем быть уверены, что она не остановится на следующем шаге. Эту проблему можно было бы обойти, если знать постоянную Шайтина. Рассмотрим всевозможные вводимые последовательности в двоичном коде конечной длины и попробуем определить те программы (вводимые последовательности), которые останавливают машину. Сначала посмотрим, останавливает ли машину первая программа на первом шаге, затем — останавливает ли машину вторая программа на первом шаге. Далее, позволим первой программе выполняться до второго шага, третьей программе до первого шага, второй программе до второго шага и первой программе до третьего шага и т. д. Если машина остановится на некоторой входной последовательности, которая записана в двоичных кодах, и ее длина k, то положим мысленно $1/2^k$ долю единицы веса в мешок. Постепенно мешок будет становиться все тяжелей и его вес будет сходиться к постоянной Шайтина (так как с вероятностью $1/2^k$ любая конкретная двоичная последовательность длины k появляется среди последовательностей, соответствующих бросанию симметричной монеты). Пусть m — длина двоичной «программы Ферма». Предположим, что программы выполняются до тех пор, пока разность между постоянной Шайтина и весом, накопленным в мешке, не станет меньше, чем $1/2^m$. Если к этому моменту великая теорема Ферма не оказалась опровергнутой, то она должна быть справедливой, так как если «программа Ферма» остановит машину позднее, мы должны будем положить в мешок $1/2^m$ единицы веса, что противоречит тому факту, что мы нашли значение постоянной Шайтина с точностью лучшей, чем $1/2^m$. Постоянная Шайтина позволяет получить решения (или дает теоретическую возможность решений) всех проблем, которые можно свести к задаче об остановке, подобной той, которую мы только что обсудили.

---

[1]) Недавно немецкий математик *Г. Фалтингс* доказал очень глубокую теорему, из которой следует, что возможные (по существу различные) решения уравнения Ферма конечны. Это большой шаг на пути решения проблемы Ферма.

д) Литература

Chaitin G. J. "Randomness and mathematical proof", *Sci. Amer.*, **232**, 47—52, (1975).

Gardner M. "The random number omega bids fair to hold the mysteries of the universe", *Sci. Amer.*, **241**, 22—31, (1979).

Guilloud J., Bouyer M. *Un million de decimals de* π. *Paris*, 1974.

Miel G. "An algorithm for calculation of π". *The Amer. Math. Monthly*, **86**, (8), (1979).

Shanks D., Wrench J. W. "Calculation of π to 100 000 decimals", *Mathematics of Computation*, **16**, 76—99, (1962).

5. Парадокс случайных графов

а) История парадокса

Структурные проблемы в ряде областей науки (например, проблему электронных схем) можно легко промоделировать и решить с помощью графов, т. е. точек и линий, их соединяющих. Точки называются вершинами, а линии — ребрами графа. Ребра могут также отражать связь, носящую случайный характер. Поэтому очень важно изучение структуры случайных графов. Теория случайных графов была разработана в основном в работах *Паула Эрдеша* и *Альфреда Реньи.*

Предположим, что у графа n вершин, и все ребра проводятся с вероятностью p независимо от существования других ребер. Пусть $\varepsilon>0$ — произвольное число. В 1920 г. Эрдеш и Реньи доказали, что если

$$p\leqslant \frac{(1-\varepsilon)\log_2 n}{n},$$

то вероятность того, что граф является связным, стремится к 0 с ростом n. С другой стороны, если

$$p\geqslant \frac{(1+\varepsilon)\log_2 n}{n},$$

то эта вероятность стремится к 1. (Говорят, что граф является связным, если из каждой вершины можно по ребрам попасть в любую другую вершину.) Таким образом, вероятность $(\log_2 n)/n$ играет роль «водораздела». За последние два десятилетия теория случайных графов была обобщена на графы с бесконечным числом вершин. В связи с такими бесконечными графами Эрдеш и Реньи обратили внимание на следующий парадокс.

б) Парадокс

Говорят, что два графа G_1 и G_2 изоморфны, если существует взаимно однозначное соответствие между множествами вершин графов G_1 и G_2 ; при этом две вершины в G_1 соединены тогда и только тогда, когда соответствующие вершины в G_2 также соединены.

У изоморфных графов мощности множеств их вершин одинаковы, но лишь одного этого недостаточно для изоморфизма. Однако, если множества вершин бесконечны, точнее, их мощность совпадает с мощностью множества натуральных чисел, и если любые две вершины в каждом графе соединяются с вероятностью 1/2 независимо от других ребер, то эти графы изоморфны с вероятностью 1. Следовательно, в этом смысле все бесконечные случайные графы одинаковы!

в) Объяснение парадокса

Говорят, что граф универсален, если для произвольных последовательностей вершин $u_1, u_2, \ldots, u_n$ и $v_1, v_2, \ldots, v_n$ (отличных друг от друга) существует вершина w, отличная от u_i и v_i и такая, что w соединена с каждой вершиной u_i и ни с одной v_i. Легко показать, что если графы G_1 и G_2 универсальны, то они и изоморфны. Вероятность того, что случайные графы, о которых говорится в парадоксе, не универсальны, равна 0 (т.е. w существует с вероятностью 1).

г) Замечания

Помимо исследований по случайным графам за последние несколько лет был проведен анализ и других случайных структур (случайных матриц, случайных алгебраических уравнений, случайных степенных рядов и т. д.), который также дал интересные результаты. Например, *Н. Б. Маслова* доказала следующую теорему: если коэффициенты X_j случайного алгебраического уравнения

$$\sum_{j=1}^{n} X_j z^j = 0$$

являются независимыми одинаково распределенными случайными величинами с нулевыми математическими ожиданиями (но сами величины не равны тождественно 0) и

$$E\left(|X_j|^{2+\varepsilon}\right) < \infty$$

для некоторого положительного ε, то число действительных корней этого уравнения распределено нормально с математическим ожиданием $\frac{2}{\pi}\ln n$ и стандартным отклонением

$$2\sqrt{\pi^{-1}(1-2\pi^{-1})\ln n}\,.$$

д) Литература

Erdös P., Spencer J. *Probabilistic Methods in Combinatorics*, Akadêmiai Kiadó, Budapest, 1974. [Имеется перевод: Эрдёш П., Спенсер Дж. Вероятностные методы в комбинаторике. — М.: Наука, 1978.]

Маслова Н. Б. "О распределении числа вещественных корней случайных полиномов", Теория вероятн. и ее примен., т. XIX, вып. 3, 488—500, 1974.

Mehta M. L. *Random Matrices*, Academic Press, New York, 1967.

6. Парадокс математического ожидания

а) История парадокса

Одна известная теорема в теории вероятностей утверждает, что если X и Y — случайные величины с конечными математическими ожиданиями, то математическое ожидание их суммы существует и равно сумме их математических ожиданий $E(X+Y)=E(X)+E(Y)$. Легко показать, что даже если математические ожидания $E(X)$ и $E(Y)$ не существуют, но существует $E(X+Y)$, то $E(X+Y)$ зависит только от распределений случайных величин X и Y, т. е. $E(X+Y)$ можно полностью определить, не зная совместного распределения X и Y. Удивительно, но для трех случайных величин это неверно.

б) Парадокс

Если X, Y и Z — произвольные случайные величины, для которых $E(X+Y+Z)$ существует, то это математическое ожидание не всегда определяется лишь индивидуальными распределениями X, Y и Z.

в) Объяснение парадокса

Определим случайные величины X, Y и Z двумя различными способами. В обоих случаях их распределения будут одинаковыми, но математические ожидания $E(X+Y+Z)$ различными.

Пусть U — случайная величина, равномерно распределенная на интервале $(0, 1)$. Тогда ясно, что $1-U$ и $V=(2U-1)$ так же равномерно распределены на $(0, 1)$. Если

$$X=Y=\operatorname{tg}\left(\frac{\pi}{2}U\right)$$

и $Z=-2X$, то $X+Y+Z=0$, и поэтому $E(X+Y+Z)=0$. С другой стороны, если

$$X=\operatorname{tg}\left(\frac{\pi}{2}U\right),\quad Y=\operatorname{tg}\left(\frac{\pi}{2}(1-U)\right)$$

и

$$Z=-2\operatorname{tg}\left(\frac{\pi}{2}V\right),$$

то неравенство $X+Y+Z>0$ справедливо с вероятностью 1, следовательно, математическое ожидание $E(X+Y+Z)$ также положительно, точнее

$$E(X+Y+Z)=\frac{4}{\pi}\ln 2.$$

г) Замечания

(i) Поскольку $E(X+Y+Z)=E((X+Y)+Z)$ и $E(X+Y+Z+W)=E((X+Y)+(Z+W))$, математические ожидания сумм трех и четырех случайных величин однозначно определяются двумерными распределениями. Однако неизвестно, остается ли это справедливым, когда число слагаемых в сумме больше четырех.

(ii) *Ружа* и *Секей* показали, что для каждой случайной величины X можно так определить действительное число $E(X)$, что это число равно математическому ожиданию этой случайной величины, если оно существует и конечно, при этом для независимых X и Y всегда справедливо равенство

$$E(X+Y)=E(X)+E(Y).$$

Наш парадокс показывает, что такое обобщенное математическое ожидание не существует у сумм случайных величин, которые не обязательно являются независимыми [для случайных величин, определенных в части *в*), величина $E(X+Y+Z)$ равнялась бы нулю, так как $E(X)=E(Y)$ и $E(Z)=-2E(X)$].

д) Литература

Ruzsa I. Z., Székely G. J. "An extension of expectation", *Zeitsch' Wahrsch' theorie verw. Geb.*, **53**, 17—20, (1980).

Simons G. "An unexpected expectation", *Annals of Prob.*, **5**, 157—158, (1977).

7. Парадокс первой цифры

а) История парадокса

Приблизительно сто лет назад, в 1881 г., *Саймон Ньюкомб* в «Американском математическом журнале» (Amer. J. Math.) обратил внимание читателей на один интересный эмпирический факт. Однако вскоре это открытие было забыто и сделано вновь спустя 60 лет физиком *Фрэнком Бенфордом*, работавшим в компании «Дженерал Электрик». Закон получил имя Бенфорда. (Ньюкомб — не единственный ученый, с которым обошлись несправедливо. Закон эпонимии саркастически утверждает, что ни одна теорема, ни одно научное открытие не были названы именем первооткрывателя.) У. Уивер в «Леди Удача» изложил историю Бенфорда: «Мне рассказали, что приблизительно двадцать пять лет назад один инженер, работавший в компании «Дженерал Электрик», по дороге на службу нес книгу, содержавшую подробную таблицу логарифмов. Он держал книгу сбоку корешком вниз. Взглянув на книгу, он заметил, что наиболее загрязнены края страниц в начале книги, затем они становятся чище — как будто чаще всего смотрят первые страницы, реже — страницы в середине книги и совсем редко — последние страницы. «Это странно,— подумал он.— Это означает, что людям чаще всего приходится искать логарифмы чисел, которые начинаются с 1, чуть реже — чисел, начинающихся с 2 и так далее и, наконец, реже всего — логарифмы чисел, начинающихся с 9. Но это совершенно невозможно, так как людям нужны значения логарифмов самых разных чисел, поэтому различные цифры должны быть представлены одинаково.»

б) Парадокс

Рассмотрим какую-нибудь таблицу, например таблицу целых степеней двойки или любую таблицу физических постоянных или таблицы демографической статистики. Как правило, окажется, что первая цифра ($\neq 0$) чисел в таблице не будет равномерно распределена на множестве 1, 2, 3, ..., 9. Цифра 1 встречается чаще всего, затем идет 2 и так далее, 9 будет самой редкой цифрой. Согласно закону Бенфорда относительная частота первых цифр, не превосходящих k, равна не $k/9$ (что означало бы равномерность распределения), а $\lg(k+1)$ (где lg означает $\log_{10}$). Следовательно, относительные частоты для 1, 2, ..., 9 приблизительно равны 30 %, 17 %, ..., 5 %. (Закон Бенфорда можно сформулировать иным образом, а именно, мантиссы логарифмов распределены почти равномерно на интервале (0, 1).) Закон Бенфорда не утверждает, что 1 — наиболее часто встречающаяся первая цифра во всех таблицах (каждый может придумать таблицу, в которой единиц вообще не будет), но все-таки единица как первая цифра в таблицах появляется обычно чаще, чем, например, девятка.

в) Объяснение парадокса

Существуют несколько вероятностных и невероятностных подходов к объяснению закона Бенфорда. Сначала рассмотрим невероятностный подход.

Проанализируем таблицу степеней двойки. Первой цифрой числа 2^n является 1, если существует такое целое число s, что $10^s \leqslant 2^n < 2 \cdot 10^s$. Если n (и, следовательно, s) достаточно велико, то s/n приблизительно равно $\lg 2$. Это означает, что среди первых n степеней двойки каждая $\lg 2$-я начинается с 1. Аналогично, по закону Бенфорда доля степеней двойки, которые начинаются с цифры, не превосходящей k, приблизительно равна $\lg(k+1)$.

Вероятностный подход немного сложнее. Опять нам нужно начать с того факта, что первая ненулевая цифра положительного случайного числа X не превосходит k, если существует целое число s, для которого

$$10^s \leqslant X < (k+1) \cdot 10^s.$$

Следовательно, закон Бенфорда имеет место только тогда, когда вероятность того, что дробная часть $\{\lg X\}$ не превосходит $\lg(k+1)$, в точности равна $\lg(k+1)$. Для этого достаточно, чтобы дробная часть $\{\lg X\}$ была равномерно распределена на интервале $(0, 1)$. Теперь рассмотрим первый вопрос: при каких условиях на распределение случайной величины X распределение $\{\lg X\}$ приблизительно равномерно? Во-вторых, почему таблицы очень часто обладают этим свойством? Хотя первый вопрос изучался несколькими математиками (например, *Р. Пинкхэмом* и *Дж. Кемперманом*) и получены достаточно хорошие результаты, на второй вопрос удовлетворительных ответов нет. Они часто приводят к путаной «философии» и даже к числовому мистицизму. По закону Бенфорда человек, например, считает арифметически: 1, 2, 3, ..., природа же автоматически берет логарифм от чисел и считает e^0, e^x, e^{2x}, ... Бенфорд утверждает, что числовые характеристики в природе складываются из геометрических прогрессий, для которых (как и для степеней двойки) закон Бенфорда справедлив. Бенфорд приводит несколько примеров из различных областей науки и производства, которые иллюстрируют закон Вебера — Фехнера, открытый в XIX веке. Согласно этому закону зависимость между раздражителем и ощущением — логарифмическая. К сожалению, аналогии, приводимые Бенфордом, удовлетворительного ответа на второй вопрос тоже не дают. Дальнейшие подробности можно найти в обзорной статье *Р. Райми*, в которой содержится обширная библиография.

г) Замечания

(i) Если нам нужно найти вероятность p_k того, что k ($k = 1, 2, \ldots, 9$) есть первая цифра числа, наугад взятого из числовой таблицы, и мы предполагаем существование определенного решения и его инвариантность относительно изменения масштаба (о масштабе единиц данных в таблице ничего не говорится), то мы приходим к логравномерному распределению $p_k = \lg(k+1) - \lg k$.

(ii) Анализируя вторые или третьи и т. д. цифры, обнаружим, что влияние «эффекта Бенфорда» практически отсутствует, если вообще существует, т. е. вторые, третьи и другие цифры распределены почти равномерно.

д) Литература

Benford F. "The law of anomalous numbers", *Proc. Amer. Phil. Soc.*, **78**, 551—572, (1938).

Newcomb S. "Note on the frequency of use of the different digits in natural numbers", *Amer. J. Math.*, **4**, 39—40, (1881).

Raimi R. A. "The first digit problem", *The American Math. Monthly*, **83**, 521—538, (1976).

8. Парадокс нулевой вероятности (Можно ли из ничего получить что-то?)

а) История парадокса

Вероятность невозможного события есть нуль, обратное же неверно: вероятность попадания в центр мишени равна нулю, но это событие не является невозможным. Вероятность попадания в любую из тысячи фиксированных точек также нулевая, но кажется, что это событие более правдоподобно, чем попадание в центр. Следовательно, возникает вопрос: можно ли сравнить «шансы» событий, имеющих нулевую вероятность, или нет? Другая проблема состоит в том, что вероятность попадания в какую-то конкретную точку мишени равна нулю, но хороший стрелок в мишень наверняка попадет, поэтому объединение событий, имеющих нулевую вероятность, может дать событие, имеющее вероятность 1, т. е. из ничего, когда его много, действительно, что-то получается. Возможно ли это на самом деле? Этот парадокс подобен известному уже две с половиной тысячи лет парадоксу Зенона о невозможности движения. Зенон утверждал, что летящая стрела в каждое мгновение неподвижна (или, иными словами, перемещение стрелы за нулевой интервал времени также должно быть равно нулю), поэтому невероятно, что стрела вообще движется. Вопрос здесь тот же самый: как может случиться, что сложение многих «ничто» дает в результате «нечто»? Таким образом, нашему парадоксу по существу несколько тысяч лет, однако его удовлетворительное объяснение появилось лишь в последние десятилетия, благодаря исследованиям *Абрахама Робинсона* (1918—1974 гг.).

б) Парадокс

Возьмем наугад точку из интервала (0, 1). Тогда вероятность того, что мы взяли именно точку 1/2, равна нулю так же, как и вероятность выбора любой из точек 1/100, 2/100, 3/100, ..., хотя последнее событие кажется более правдоподобным. Действительно ли невозможно различить вероятности этих двух событий?

в) Объяснение парадокса

В ходе развития арифметики вводились все более сложные типы чисел: после натуральных чисел и дробей появились нуль, отрицательные, действительные (= рациональные + иррациональные) и комплексные числа. В 1960-х годах множество чисел пополнилось в результате введения бесконечно малых чисел, или просто бесконечно малых. Со времен Ньютона и Лейбница в дифференциальном и интегральном исчислениях использовалось словосочетание «бесконечно малые», однако оно применялось чисто символически без четкого определения или обоснования. Именно из-за отсутствия обоснования бесконечно малые были изгнаны в прошлом веке из строгой математики, однако не исчезли совсем (так как физики использовали их постоянно). Математики перешли к «эпсилон-дельта» анализу, который до сих пор соответствует духу университетского образования. Теория Робинсона закладывает прочный логический фундамент «под» применение бесконечно малых, и в будущем веке, видимо, возобновят обучение студентов оригинальным эвристическим идеям Ньютона и Лейбница. (В университетах штата Висконсин и в Массачусетском технологическом институте студенты уже сейчас могут, если захотят, выбрать вместо эпсилон-дельта теории Вейерштрасса теорию Робинсона.) Бесконечно малые можно обычно использовать в вычислениях подобно другим числам. Хотя деление на нуль запрещено, деление на бесконечно малое строго определено: величиной, обратной к бесконечно малой, является бесконечно большое число, и наоборот, величина, обратная к бесконечно большому числу, всегда есть бесконечно малая. До появления теории Робинсона считали, что рациональные и иррациональные (т. е. действительные) числа заполняют всю числовую прямую. Исследуя отдельную точку на числовой прямой под «математическим микроскопом» Робинсона, мы видим не только эту точку, но и множество бесконечно малых, которые бесконечно близки к ней. Из уважения к Лейбницу этот образ называется «монадой». С помощью бесконечно малых можно разрешить многие парадоксы, в частности парадокс Зенона и парадокс нулевой вероятности. Суть в том, что нужно делать различие между нулем и бесконечно малыми числами. Например, каждому подмножеству некоторого интервала можно приписать вероятность так, что эта вероятность равна нулю только для пустого подмножества, соответствующего невозможному событию, а для любого другого события вероятность положительна, хотя, возможно, и бесконечно мала. Далее, для множества A, имеющего вероятность $P(A)$ в традиционном смысле, получим вероятность, отличающуюся от $P(A)$ самое большее на бесконечно малое. (Эта новая вероятность не сигма-аддитивна, а лишь аддитивна.) Теперь мы действительно можем сказать, что вероятность выбора одной точки, например центра интервала, меньше вероятности выбора одной из двух точек: разница есть бесконечно малое число.

г) Замечание

Ньютон попытался сформулировать законы природы в математической форме, таким образом, он приблизился к границе между конечными и бесконечными величинами. А Робинсон воспринял саму бесконечность (следуя примеру Г. Кантора и других) и сделал ее привычной для всех математиков.

д) Литература

Luxemburg W. A. J. (ed.) *Applications of Model Theory to Algebra, Analysis and Probability*, Holt, Rinehart and Winston, New York, 1969.

Robinson A. "Non-standard analysis", *Proc. Nederl. Akad. Wetensch*, **64**, 432—440, (1961).

Robinson A. *Non-standard Analysis*, North-Holland, Amsterdam, 1966.

9. Парадокс безгранично делимых распределений

а) История парадокса

Понятие безгранично делимого распределения ввел в 1929 г. *Б. де Финетти*. Распределение F называется безгранично делимым, если для любого положительного целого числа n найдутся n независимых одинаково распределенных случайных величин таких, что функция распределения их суммы совпадает с F. Пусть F_1 и F_2 — функции распределения двух независимых случайных величин. Обозначим функцию распределения их суммы через $F_1{*}F_2$. Операция $*$ называется сверткой. Очевидно,

$$(F_1 * F_2) * F_3 = F_1 * (F_2 * F_3), \quad F_1 * F_2 = F_2 * F_1$$

(это означает, что алгебраически функции распределения с операцией свертки образуют коммутативную полугруппу). Функция распределения F является $*$-безгранично делимой (по указанному выше определению), если для всех натуральных чисел n существует функция распределения F_n, для которой

$$\underbrace{F_n * F_n * \ldots * F_n}_{n \text{ раз}} = F.$$

Например, нормальное, пуассоновское и показательное распределения безгранично делимы. Основное значение безгранично делимых распределений состоит в том, что они выступают как предельные распределения для сумм независимых случайных величин. В 1936 г. *Крамер* доказал, что если свертка двух распределений нормальна, то оба распределения обязаны быть нормальными и, следовательно, безгранично делимыми. Два года спустя *Райков* получил аналогичный результат для пуассоновских распределений. Эти результаты удивительны тем, что из них следует, что нормальные распределения можно разложить только на нормальные, и аналогичный факт верен для пуассоновских распределений. Но еще более удивительно то, что безгранично делимые распределения можно разложить на компоненты, не являющиеся безгранично делимыми.

б) Парадокс

Существуют функции распределения, которые не безгранично делимы, но их свертка безгранично делима.

в) Объяснение парадокса

Покажем, что показательное распределение можно представить в виде свертки распределений, которые не являются безгранично делимыми. Рассмотрим показательное распределение с параметром $\lambda = 1$, его плотность вероятности равна e^{-x}, если $x > 0$, и 0 для $x < 0$. Это распределение, действительно, безгранично делимо, так как n-кратная свертка гамма-распределения порядка $1/n$ (с плотностью вероятности

$$x^{1/n-1}e^{-x}/\Gamma(1/n), \quad \text{если} \quad x > 0,$$

и 0 для $x < 0$) совпадает с нашим показательным распределением. Однако его можно разложить не только на гамма-распределения (которые сами безгранично делимы), но и представить в виде свертки двух распределений, одно из которых имеет значения $k = 0, 1, 2, \ldots$ с вероятностями $2^{-(k+1)}$, а другое сосредоточено на интервале $(0, 1)$, т. е. все его значения принадлежат интервалу $(0, 1)$. Последнее распределение не является безгранично делимым. Согласно замечанию (i) распределение любой ограниченной случайной величины не может быть безгранично делимым. Итак, мы уже знаем, что безгранично делимое распределение может не иметь безгранично делимых компонент в смысле свертки. Далее, обе компоненты показательного распределения, полученные выше, можно разложить дальше так, что каждая компонента сосредоточена в двух точках, точнее, в 0 и в целой степени двойки. Эти распределения (как и любое распределение, сосредоточенное в двух точках) не только не безгранично делимы, но и далее неразложимы (т. е. их нельзя представить в виде свертки невырожденных распределений, т. е. распределений, не сосредоточенных в одном числе).

г) Замечания

(i) Покажем, что функция распределения любой ограниченной случайной величины не может быть безгранично делимой, если только случайная величина невырождена (случайная величина вырождена, если она принимает только одно значение с вероятностью 1; в этом случае ее дисперсия равна 0). Для ограниченной случайной величины X найдется число K такое, что $|X| < K$. Если функция распределения случайной величины X безгранично делима, то существуют независимые случайные величины $X_1, X_2, \ldots, X_n$, одинаково распределенные и такие, что функции распределения суммы $X_1 + X_2 + \ldots + X_n$ и X совпадают. Поскольку супремум и дисперсия суммы независимых случайных величин есть сумма супремумов и дисперсий слагаемых, имеем

$$|X_i| < K/n \quad \text{и} \quad D(X_i) = D(X)/\sqrt{n}.$$

Следовательно, если $D(X) \neq 0$ и n — достаточно велико, то дисперсия случайной величины X_i станет больше, чем супремум величины $|X_i|$, что невозможно. Таким образом, если X ограничена и безгранично делима, то $D(X) = 0$, т. е. случайная величина X вырождена.

(ii) Кроме нормального, пуассоновского и гамма-распределений, безгранично делимо и логнормальное распределение. (Логнормальным называется распределение положительной случайной величины, логарифм которой нормально распределен, см. статью Торина). Безгранично делимы также t-распределение Стьюдента и распределение Коши (распределение отношения двух независимых случайных величин, имеющих стандартное нормальное распределение), см. книгу Лукача и статьи Гроссвальда, Эпштейна и Бондессона.

(iii) Показательное распределение служит примером безгранично делимого распределения, представимого в виде бесконечной (счетной) свертки неразложимых распределений. Еще более удивительно то, что существуют безгранично делимые распределения, представимые в виде свертки лишь двух неразложимых распределений (см. статью Леви).

(iv) Характеризация безгранично делимых распределений была найдена в 1930-е годы *Колмогоровым*, *Леви* и *Хинчиным*. Легко показать, что функция распределения суммы $X_1 + X_2 + \ldots + X_N$ всегда безгранично делима, если $X_1, X_2, \ldots$ —

произвольные независимые одинаково распределенные случайные величины, принимающие неотрицательные целочисленные значения, и N — случайная величина с пуассоновским распределением, не зависящая от всех X_i. В то же время из теоремы Леви — Хинчина следует, что всякое безгранично делимое распределение, сосредоточенное на неотрицательных целых числах, представимо именно таким образом. Несмотря на то, что характеризационные теоремы для безгранично делимых распределений известны уже 50 лет, до сих пор не решена проблема характеризации безгранично делимого распределения, имеющего в представлении в виде свертки распределений только безгранично делимые компоненты (нормальное и пуассоновское распределения принадлежат этому классу распределений, а показательное распределение, как мы видели, не принадлежит).

(v) В теории вероятностей понятие безграничной делимости появляется не только в связи со свертками. На множестве функций распределения можно определить и другие важные операции. Например, функция распределения максимума из двух независимых случайных величин, имеющих функции распределения $F_1(x)$ и $F_2(x)$, равна произведению $F_1(x)\cdot F_2(x)$. Произведение функций распределения часто появляется во многих вероятностных задачах, например, в теории надежности, когда мы хотим получить вероятностное распределение времени безотказной работы параллельных соединений. Для любого натурального числа n и произвольной функции распределения $F(x)$ величина $\sqrt[n]{F(x)}$ также, очевидно, является функцией распределения, таким образом, каждая (одномерная) функция распределения безгранично делима. В случае большей размерности характеризация безгранично делимых распределений менее тривиальна (см. статью Балкема и Ресника). Третья операция состоит в следующем.

Пусть $F_1 \circ F_2$ обозначает мультипликативную свертку, т. е. функцию распределения произведения независимых случайных величин с функциями распределения F_1 и F_2. Пуассоновское распределение безгранично делимо относительно операции свертки $*$, но не является таковым для операции $\circ$. Более того, если X и Y обозначают независимые случайные величины, и произведение XY имеет пуассоновское распределение, то либо величина X, либо Y с вероятностью 1 сосредоточена на множестве из двух элементов $\{0, 1\}$. Это означает, что пуассоновское распределение $\circ$-неразложимо (см. статью Секея и Земплени). В то же время стандартное нормальное распределение $\circ$-безгранично делимо. (Однако пока неизвестно, является ли $\circ$-безгранично делимым нормальное распределение с положительным математическим ожиданием; если математическое ожидание отрицательно, то нормальное распределение, очевидно, не $\circ$-безгранично делимо.)

д) Литература

Balkama A. A., Resnick S. I. "Max-infinite divisibility", *J. Appl. Prob.*, **14**, 309—319, (1977).
Bondesson L. "A general result of infinite divisibility", *Annals of Probability*, **7**, 965—979, (1979).
Epstein B. "Infinite divisibility of Student's *t*-distribution", *Sankhya, Ser. B.*, **39**, 103—120, (1977).
Fisz M. "Infinitely divisible distributons: recent results and applications", *Annals of Math. Statist.*, **33**, 68—84, (1962).
Göndőcs F., Michaletzky G., Móri T., Székely G. J. "A characterization of infinitely divisible Markov chains with finite state space", *Ann. Univ. Sci. Budapest Sect. Math.*, **27**, 137—141, (1985).
Grosswald E. "The Student *t*-distribution of any degree of freedom is infinite divisible", *Zeitsch 'Wahrsch' theorie verv. Geb.*, **36**, 103—109, (1976).
Lévy P. "Sur les exponentielles de polinómes", *Ann. Sci. Ècole Normale Supérieure*, **54**, 231—292, (1937).
Lukacs E. Characteristic Functions, Griffin, London, 1960. [Имеется перевод: Лукач Е. Характеристические функции. — М.: Наука, 1979.]
Steutel F. W. "Infinite divisibility in theory and practice", *Scand. J. Statist.*, **6**, 57—64, (1979).
Székely G. J. "Multiplicative infinite divisibility of the normal distribution", *Proc. 7th Brasov Conf. on Probab. Theory*, Acad. Publ. Bucureşti, 579—582, 1984.
Székely G. J., Zemplémi A. "Advanced problem 6431", *The American Math. Monthly*, **90**, 402, (1983).
Székely G. J. "Problem 180" *Statistica Neerlandica*, **39**, 324, (1985).
Thorin O. "On the infinite divisibility of the lognormal distribution", *Scand. Acturial J.*, 121—148, (1977).
Золотарев В. М. "Общая теория перемножения независимых случайных величин". — ДАН СССР, т. 142, № 4, 788—791, 1962.

10. Парадоксы характеризации

а) История парадоксов

Следующую проблему впервые поставил *Джордж Пойа*. Рассмотрим две независимые одинаково распределенные случайные величины X и Y. Может ли сумма $aX + bY$ иметь то же распределение, что X и Y, если a и b — положительные числа? Пойа исследовал этот вопрос в статье, опубликованной в 1923 г. Следующий замечательный результат появился спустя много лет, лишь в 1936 г., когда *Е. Гири* начал описание распределения F, удовлетворяющего свойству: если величины $X_1, X_2, \ldots, X_n$ независимы и имеют распределение F, то

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \quad \text{и} \quad S = \sum_{i=1}^{n} (X_i - \overline{X})^2$$

также независимы. В 1939 г. *М. Кац* и в 1941 г. С. Н. Бернштейн ответили на такой вопрос: если X и Y — независимые одинаково распределенные случайные величины, то при каких условиях величины $X+Y$ и $X-Y$ независимы? С сороковых годов после работ таких выдающихся математиков как *Ю. В. Линник*, *Е. Лукач*, *А. А. Зингер*, *С. Р. Рао* и *А. М. Каган* направление, связанное с характеризационными задачами, развилось в очень важную как теоретически, так и практически, часть теории вероятностей.

б) Парадоксы

Пусть $X_1, X_2, \ldots, X_n$ — независимые одинаково распределенные случайные величины. Могут ли величины $Y=f(X_1, X_2, \ldots, X_n)$ и $Z=g(X_1, X_2, \ldots, X_n)$ быть одинаково распределенными или независимыми, если f и g — различные, например, линейные функции? В некоторых случаях, например, когда f (и, следовательно, Y) тождественно равна постоянной, Y и Z, очевидно, независимы для любой функции g, однако в общем случае ожидается, что Y и Z не будут ни независимыми, ни одинаково распределенными. Удивительно, но исключения возникают как раз в наиболее важных случаях, когда величины X_i нормально распределены. Если, например, X_1 и X_2 представляют собой координаты вектора скорости точки, случайно двигающейся на плоскости, и X_1, X_2 — независимые случайные величины со стандартным нормальным распределением, то сумма $Y=X_1^2+X_2^2$ (пропорциональная кинетической энергии) и отношение $Z=X_1/X_2$ (определяющее направление движения) независимы. Свойства такого типа часто характеризуют нормальные (или другие важные) распределения.

в) Объяснение парадоксов

Пусть обе функции f и g линейны:

$$Y=\sum_{i=1}^{n} a_i X_i \quad \text{и} \quad Z=\sum_{i=1}^{n} b_i X_i.$$

Если найдутся числа a_i и b_i такие, что произведение $a_i b_i$ не всегда равно нулю, и равенство $a_i=b_i$ не выполняется для всех i, но Y и Z все же одинаково распределены, и все моменты величин X_i конечны, т. е. $E(|X_i|^k)$ конечно для каждого k ($k=0, 1, 2, \ldots$), то величины X_i нормально распределены. Эта теорема *Марцинкевича* обобщает теорему Пойа. (Дальнейшие обобщения можно найти в книге Кагана, Линника и Рао.)

Теорема *Г. Дармуа* и *В. Р. Скитовича* утверждает, что если

$$Y = \sum_{i=1}^{n} a_i X_i \quad \text{и} \quad Z = \sum_{i=1}^{n} b_i X_i$$

независимы и $a_i b_i$ не равно нулю для всех i, то величины X_i нормально распределены.

Следующее обобщение теоремы Гири (включающее в себя также случай нелинейных функций) играет очень важную роль в математической статистике. Утверждается, что если $\bar{X}$ и S независимы и $n \geqslant 2$, то величины X_i нормально распределены.

г) Замечания

(i) Условие независимости величины $\bar{X}$ и S является очень сильным, но коэффициент корреляции $r(\bar{X}, S) = 0$, например, для всех симметричных случайных величин $X_1, X_2, \ldots, X_n$, для которых корреляция существует. Хотя $r(\bar{X}, S)$ всегда меньше 1, его супремум равен 1. Для одновершинных распределений точная верхняя грань равна $\sqrt{15/16}$. Для доказательства этого результата можно применить точное неравенство $m_3^2 \leqslant (m_4 - m_2^2) m_2$, где $m_k = E(X - E(X))^k$, $k = 2, 3, 4$.

(ii) Для характеризации семейства распределений существует несколько интересных и естественных способов. Например, показательные распределения можно характеризовать следующим свойством: среди всех распределений, сосредоточенных на интервале $(0, \infty)$ и имеющих заданное математическое ожидание, энтропия

$$-\int f(x) \log_2 f(x)\, dx$$

[$f(x)$ обозначает плотность вероятности] максимальна для показательных распределений. Среди распределений на интервале $(-\infty, \infty)$ с заданными математическим ожиданием и дисперсией нормальные распределения имеют максимальную энтропию. На конечном интервале энтропия максимизируется равномерными распределениями (без каких-либо других предложений).

д) Литература

Galambos J., Kotz S. *Characterization of Probability Distributions*, Springer, Berlin — Heidelberg — New York, 1978.

Каган А. М., Линник Ю. В., Рао С. Р. Характеризационные задачи математической статистики. — М.: Наука, 1972.

Mathai A. M., Pederzoli G. *Characterizations of the Normal Probability Law*, Wiley E. L., New Delhi — Bangalore — Bombay, 1977.

Polya G. "Herleitung des Gauss'schen Fehlergesetzes aus einer Funktionalgleichung", *Math. Zeit.*, **18**, 96—108, (1923).

11. Парадоксы факторизации

а) История парадоксов

Основные теоремы в классической теории вероятностей (такие, как законы больших чисел, теоремы о предельном распределении) связаны с распределением суммы независимых случайных величин и опираются на свойства слагаемых этих сумм. «Обратными» к этим теоремам о «композиции» являются теоремы о «декомпозиции», или «факторизационные» теоремы, в которых распределение суммы известно, и мы хотим получить какую-либо информацию о возможных слагаемых или «факторах» (делителях). Теорема *Крамера*, которая уже упоминалась, представляет собой именно такой результат о декомпозиции. Она утверждает, что все делители нормального распределения также нормальны. Как в теоремах о композиции, так и в теоремах о декомпозиции, важную «техническую роль» играют характеристические функции случайных величин. Характеристическая функция случайной величины X определяется как математическое ожидание комплексной случайной величины e^{itX} ($i=\sqrt{-1}$ и t — действительное число), т. е. $\varphi_X(t)=E(e^{itX})$. Каждая случайная величина имеет характеристическую функцию, которая однозначно определяет функцию распределения этой величины. Характеристическая функция суммы независимых случайных величин равна произведению характеристических функций слагаемых. Из этих свойств ясно, почему характеристические функции столь важны при решении проблем, связанных с композицией и факторизацией. Характеристические функции использовали еще в 1853 г. *А. Коши* и на рубеже XX века *А. М. Ляпунов*. С 1920-х годов под сильным влиянием работ *Дж. Пойа* и *П. Леви* характеристические функции очень часто применяются для решения проблем о композициях. С 1930-х годов благодаря теоремам *Крамера*, *Хинчина* и *Райкова* возникает теория декомпозиции. В этой области нет недостатка в удивительных результатах и парадоксах (некоторые из них см. ниже).

б) Парадоксы

(i) Существуют случайные величины X, Y и Z такие, что вероятностные распределения сумм $X+Y$ и $X+Z$ совпадают, но распределения величин Y и Z различны. Этот факт, впервые отмеченный *Хинчиным* в 1937 г., достаточно удивителен, так как если X — ограниченная случайная величина, или если ее характеристическая функция нигде не обращается в нуль (например, когда она безгранично делима), то распределения величин Y и

Z должны совпадать. В силу парадокса Хинчина в общем случае бессмысленно говорить об «остатке» вероятностного распределения после его сокращения на один из делителей, так как оставшаяся часть не определяется однозначно. По этой причине в алгебре вероятностных распределений возникает очень много трудностей. В то же время, так как характеристическая функция нормального распределения нигде не равна 0 (характеристическая функция стандартного нормального распределения есть $e^{-t^2/2}$) естественно поставить вопрос о том, что останется, если удалить у вероятностного распределения нормальную составляющую. Однако определенная осторожность необходима и в этом случае. А именно, существуют независимые одинаково распределенные случайные величины X и Y, у которых нет нормально распределенных делителей (с положительной дисперсией), но она есть у их суммы $X+Y$. Впервые этот результат опубликовали в 1948 г. *Д. Дюге* и *Р. Фишер* (см. ниже их статью).

(ii) Пусть $X_1, X_2, \ldots, X_n$ — независимые случайные величины с неизвестными (и, вообще говоря, различными) распределениями. Но предположим, что известны распределения линейных комбинаций

$$Y_j = \sum_{k=1}^{n} c_{jk} X_k, \quad j = 1, 2, \ldots, n;$$

(c_{jk} — произвольное число). Если существуют величины $X_1, X_2, \ldots, X_n$, удовлетворяющие этой системе уравнений, и детерминант матрицы (c_{jk}) отличен от 0, то (поскольку в этом случае $Y_1, Y_2, \ldots, Y_n$ определяют $X_1, X_2, \ldots, X_n$ однозначно) можно ожидать, что распределения величин $X_1, X_2, \ldots, X_n$ также определяются однозначно. Однако, как показал в 1950 г. *А. Реньи*, это не так.

в) Объяснение парадоксов

(i) Можно показать, что если значение функции $\varphi(t)$ неотрицательно при любом действительном t, $\varphi(0)=1$, $\varphi(-t) = \varphi(t)$, и для всех положительных значений t функция $\varphi(t)$ выпукла и $\lim \varphi(t) = 0$ при $t \to \infty$, то найдется случайная величина, для которой $\varphi(t)$ — ее характеристическая функция. Следовательно, существует случайная величина X с характеристической функцией, равной $1-|t|$, если $|t| \leqslant 1$, и 0 в противном случае. Существуют также случайные величины Y и Z, характеристические функции которых одинаковы на интервале $|t| \leqslant 1$, но различны вне этого интервала. Таким образом, для независимых X, Y и Z имеем

$$\varphi_{X+Y}(t) = \varphi_{X+Z}(t),$$

т. е. суммы $X+Y$ и $X+Z$ распределены одинаково, но распределения величин Y и Z различны. (Другой пример см. в п. 13*а*).

Можно доказать нечто большее, чем то, что утверждается в парадоксе. Можно показать, что если $\psi(t)$ — периодическая функция с периодом 2 и $\psi(t)=1-|t|$ при $|t|\leqslant 1$, то найдется случайная величина Y, характеристическая функция которой как раз равна $\psi(t)$. Отсюда следует, что $\varphi(t)\psi(t)\equiv\varphi(t)^2$, т. е. существуют независимые случайные величины X, Y и Z такие, что распределения сумм $X+Y$ и $X+Z$ совпадают, одинаковы также распределения величин X и Z, но распределения величин Y и Z различны.

(ii) Если случайные величины Y_j заданы и определитель $\|c_{jk}\|\neq 0$, то случайные величины X_j определяются однозначно. Однако случайные величины Y_j можно задать разными способами так, что их распределения остаются неизменными. Следовательно, если предполагать только, что $\|c_{jk}\|\neq 0$, то совсем не очевидно, что распределения величин Y_j однозначно определят распределения X_j.

г) Замечания

Реньи доказал, что, вообще говоря, необходимо также предполагать отличие от нуля определителя $\|c_{jk}^2\|$. Если известно, что

$$\|c_{jk}\|\neq 0 \text{ и } \|c_{jk}^2\|\neq 0,$$

то при достаточно общих условиях гарантирована однозначность распределений X_j (например, если характеристическая функция величины Y_j — целая функция порядка $\leqslant 2$, и если вообще есть решение). Этот факт имеет очень важные практические последствия. В II/13*в* мы видели, что с помощью двух измерений можно найти более точные оценки (их дисперсии меньше) двух неизвестных значений, если измерять эти значения не отдельно друг от друга, а сначала измерить их сумму и затем — их разность. Аналогичная ситуация возникает, когда мы хотим найти n различных неизвестных величин $X_1, X_2, \ldots, X_n$. При измерении определенных линейных комбинаций $Y_1, Y_2, \ldots, Y_n$ можно получить более точные результаты. В общем случае целесообразно использовать матрицу (c_{jk}), элементами которой являются только $+1$ и -1. Тогда, очевидно, $\|c_{jk}^2\|=0$ и, следовательно, распределения величин $Y_1, Y_2, \ldots, Y_n$ не определяют однозначно распределения $X_1, X_2, \ldots, X_n$. Пусть, например, каждая из величин $Y_1=X_1+X_2$ и $Y_2=X_1-X_2$ имеет стандартное нормальное распределение. Тогда теорема Крамера утверждает, что X_1 и X_2 также распределены нормально с нулевым математическим ожиданием. Однако их дисперсии не определяются однозначно. Они лишь удовлетворяют соотношению $D^2(X_1) + D^2(X_2) = 1$.

д) Литература

Dugué D., Fischer R. A. “Un résultat assez inattendu d’arithmétique des lois des probabilités”, *C. r. Acad. Sci.*, **227**, 1205—1207, (1948).

Feller W. *An Introduction to Probability Theory and its Applications*, Vol. II, Wiley, New York, 1966 (Chapter XV). [Имеется перевод: Феллер В. Введение в теорию вероятностей и ее приложения. В 2-х томах. Т. 2. Гл. XV. — М.: Мир, 1984.]

Lukacs E. *Characteristic Functions*, Griffin, London, 1960. [Имеется перевод: Лукач Е. Характеристические функции. — М.: Наука, 1979.]

Renyi A. “On the algebra of distributions”, *Publ. Math.*, **1**, 135—149, (1950).

12. Парадокс неразложимых и простых распределений

а) История парадокса

Неразложимые числа, т. е. числа (больше 1), имеющие в качестве делителей только единицу и самих себя, играют в арифметике фундаментальную роль. Эти числа: 2, 3, 5, 7, 11, 13, 17, 19... являются также *простыми числами*, т. е. числами, которые, будучи делителями произведения двух натуральных чисел, являются делителями по крайней мере одного из этих сомножителей. Среди натуральных чисел множества неразложимых и простых чисел совпадают, и числа 2, 3, 5, 7, ... всегда называются простыми числами. Фундаментальная теорема арифметики утверждает, что для каждого целого числа, которое больше единицы, существует единственное разложение на простые сомножители (порядок сомножителей не учитывается). Таким образом, простые числа в арифметике подобны детским кубикам или атомам в физическом мире. Наиболее естественный способ получения информации о какой-либо сложной структуре состоит в разбиении ее на атомы, поэтому вполне объяснимо, что (начиная с прошлого века) понятия неразложимости и простоты были распространены на общие алгебраические структуры. Эти понятия можно интерпретировать и для вероятностных распределений: здесь роль натуральных чисел играют вероятностные распределения и роль умножения — свертка (определение свертки см. в «Парадоксе безгранично делимых распределений»).

Распределение F *неразложимо*, если из равенства $F = G * H$ следует, что одно из распределений G или H вырождено (т. е. сосредоточено в одной точке с вероятностью 1; эти распределения играют роль единиц). Распределение F называется *простым распределением*, если оно является делителем свертки $G * H$

только тогда, когда оно также является делителем G или H. В 1937 г. Хинчин доказал, что каждое распределение представляет собой свертку безгранично делимого распределения с конечной или счетной сверткой неразложимых распределений, т. е. каждую характеристическую функцию $\varphi(t)$ можно представить в следующем виде $\varphi(t) = \psi(t) \prod_i \varphi_i(t)$, где $\psi(t)$ и $\varphi_i(t)$ — характеристические функции соответственно безгранично делимого и неразложимых распределений. Это похоже на фундаментальную теорему арифметики, но есть одно важное отличие: факторизация распределений не единственна. Например, если распределение F имеет значения 0, 1, 2, 3, 4, 5 с одинаковыми вероятностями 1/6, то F представимо в виде свертки неразложимых распределений двумя различными способами: при первом разложении первый делитель принимает значения 0 и 1, второй делитель принимает значения 0, 2 и 4 с одинаковыми вероятностями; при втором разложении первый делитель принимает значения 0 и 3, второй делитель принимает значения 0, 1 и 2 с равными вероятностями. Эта неопределенность показывает, что сходство между «арифметиками» чисел и вероятностных распределений неполное. Следующий парадокс демонстрирует более существенные различия.

б) Парадокс

Вероятностные распределения с операцией свертки образуют алгебраическую структуру, содержащую множество неразложимых распределений (например, каждое распределение, сосредоточенное в двух точках, неразложимо), но в ней нет ни одного простого распределения. Следовательно, если мы действительно считаем простые распределения «атомами», то здесь вообще нет атомов. Этот факт впервые отметили в 1979 г. *И. Ружа* и *Г. Секей.*

в) Объяснение парадокса

Совсем неудивительно, что простота и неразложимость — это обычно не одно и то же. Этот факт представляется неожиданным только потому, что в самой знакомой и важной структуре — множестве натуральных чисел — эти оба понятия эквивалентны. Однако в общем случае можно лишь утверждать, что простой элемент всегда является неразложимым, а обратное не обязательно верно. Совпадение обоих понятий (грубо говоря) означает, что факторизация на неразложимые элементы единственна. Уже в 1937 г. Хинчин заметил, что представление вероятностных распределений в виде свертки неразложимых распределений не единственно, т. е. получается, что не каждое неразложимое распределение простое. В своей статье Ружа и Секей показывают, что в этой структуре вообще нет простых элементов, и поэтому, что касается связи между понятиями неразложимости и простоты, структура натуральных чисел с умножением и структура вероятностных распределений со сверткой полярно противоположны.

г) Замечание

Два распределения F и G называются *взаимно простыми*, если F и G могут быть делителями распределения H только тогда, когда $F * G$ также является делителем H. В статье, о которой говорилось выше, доказано, что ограниченные (и невырожденные) распределения не могут быть взаимно простыми, и мы думаем, что вообще нет взаимно простых распределений; но эта проблема еще не решена.

д) Литература

Kendall D. G., Harding E. F. (eds.) *Stochastic Analysis*, Wiley, New York, 1973.

Линник Ю. В., Островский И. В. Разложения случайных величин и векторов. — М.: Наука, 1972.

Ruzsa I. Z., Szekely G. J. "No distribution is prime", *Zeitsch' Wahrsch' theorie verw. Geb.*, **70**, 263—270, (1985).

Ruzsa I. Z., Székely G. J. Algebraic Probability Theory, Wiley, New York, 1988.

Ruzsa I. Z., Székely G. J. "Theory of Decomposition in Semigroups", *Advances in Math.*, **56**, 9—27, (1985).

Ruzsa I. Z., Székely G. J. "How to eliminate probabilities from probability theory?", (в печати).

Ruzsa I. Z., Székely G. J. "A note on our paper 'Theory of Decomposition im Semigropes'" *Advances in Math.*, **60**, 235—236 (1986).

Székely G. J., Zempleni A. "Multiplicative arithmetic of distribution functions", (в печати).

13. Еще несколько парадоксов

а) Парадокс деления распределений пополам

Пусть X и Y — независимые одинаково распределенные случайные величины. Распределение суммы $X+Y$ обычно однозначно определяет общее распределение величин X и Y, однако, как это не парадоксально, так бывает не всегда. Это удивительно потому, что на практике распределения, как правило, делятся однозначно, т. е. их n-я часть (если она существует) определяется единственным образом, как, например, в случае ограниченных или безгранично делимых распределений. Рассмотрим теперь парадоксальный пример.

Если случайная величина принимает значения $2k+1$ ($k=0, \pm 1, \pm 2, \ldots$) с соответствующими вероятностями

$$\frac{4}{\pi^2 (2k+1)^2},$$

то ее характеристическая функция $\varphi(t)$ является периодической с периодом 2π и на интервале $-\pi \leqslant t \leqslant \pi$

$$\varphi(t) = 1 - 2|t|/\pi.$$

Определим сейчас другую случайную величину, которая равна нулю с вероятностью $1/2$ и принимает значения $4k+2$ с вероятностями

$$\frac{2}{\pi^2 (2k+1)^2}.$$

Характеристическая функция $\psi(t)$ этой случайной величины также периодическая. Ее период равен π и на интервале $-\pi/2 \leqslant t \leqslant \pi/2$

$$\psi(t) = 1 - 2|t|/\pi.$$

Очевидно, $\psi(t) = |\varphi(t)|$, следовательно, $\psi(t)^2 = \varphi(t)^2$. Таким образом, если характеристическая функция суммы $X+Y$ есть $\psi(t)^2 = \varphi(t)^2$, то общей характеристической функцией величин X и Y может быть либо $\varphi(t)$, либо $\psi(t)$, поэтому однозначной определенности нет.

Заметим, что $(\psi(t) + \varphi(t))/2$ также является характеристической функцией, следовательно,

$$\frac{\varphi(t) + \psi(t)}{2}\varphi(t) = \frac{\varphi(t) + \psi(t)}{2}\psi(t)$$

что дает другой пример первого парадокса факторизации, так как приведенное выше равенство нельзя сокращать на $(\varphi(t) + \psi(t))/2$.

Можно также указать такие характеристические функции, что их значения не всегда действительны, но квадраты функций действительны всегда. Следовательно, существуют вероятностные распределения, симметричные относительно нуля, но «их половины» несимметричны в том смысле, что существуют независимые одинаково распределенные случайные величины X и Y, сумма которых $X+Y$ распределена симметрично, но X и Y несимметричны. (Если случайные величины ограничены, то такая ситуация невозможна.) Удивительным также представляется тот факт, что если плотность вероятности $f(x)$ случайной величины X симметрична ($f(-x) = f(x)$), при этом $0 < a < f(t) < b < \infty$ для $|t| \leqslant c < \infty$ и $f(t) = 0$ при $|t| > c$, то у X нет половины, т. е. не существует характеристической функции φ, для которой $\varphi_X = \varphi^2$. (Лит.: Проблема 10, *Mat Lapok*, **30**, 1982, с. 272. (На венгерском языке.) Предложили Т. Мори и Г. Секей.)

б) Патологические вероятностные распределения

(i) Пусть функция распределения F некоторой случайной величины обладает следующими свойствами: $F(0) = 0$, $F(1) = 1$ и для

$$x = \sum_{r=0}^{\infty} 2^{-a_r} \quad (a_0 < a_1 < a_2 < \ldots \text{ — положительные числа})$$

имеем

$$F(x) = \sum_{r=0}^{\infty} a^r (1+a)^{-a_r},$$

где a — произвольное положительное число. *Л. Такач* показал (*The American Math. Monthly*, **85**, 35—37, 1978), что $F(x)$ — строго возрастающая и непрерывная функция на интервале $(0, 1)$. При $a = 1$ получим $F(x) = x$ на интервале $(0, 1)$, т. е. случайная величина равномерно распределена, ее плотность вероятности равна нулю вне интервала $(0, 1)$ и равна 1 на самом интервале. Удивительно, но если $a \neq 1$, то F (и соответствующая случайная величина) вообще не имеют плотности, т. е. не существует функции f, для которой

$$F(x) = \int_{-\infty}^{x} f(u)\, du.$$

Хотя наиболее часто встречающиеся непрерывные распределения всегда имеют плотности вероятности, нельзя забывать о патологических случайных величинах, которые мы только что рассмотрели. Например, интересно отметить, что если равномерно распределенная случайная величина представляется в виде суммы двух независимых случайных величин с непрерывными функциями распределения, то по крайней мере одно из распределений патологическое, т. е. не имеет плотности вероятности. (История патологических и очень патологических, т. е. «сингулярных», функций началась в 1904 г., когда *А. Лебег* опубликовал свою книгу по интегрированию функций. Один из последних результатов по сингулярным функциям принадлежит также Cater F. S. "Most monotone functions are not singular", *Т. Замрифеску*, см. Zamrifescu T. "Most monotone functions are singular", *The American Math. Monthly*, **88**, 47—49, (1981). См. *The American Math. Monthly*, **89**, 466—469, (1982).)

(ii) Пусть совместная плотность вероятности случайных величин X и Y задана равенством

$$h(x,\ y) = \frac{|x|}{2\sqrt{2\pi}} e^{-(|x|+x^2y^2/2)}.$$

Тогда плотность вероятности величины X есть

$$f(x) = \int_{-\infty}^{+\infty} h(x,\ y)\, dy = \begin{cases} 0, & x = 0, \\ \frac{1}{2} e^{-|x|}, & x \neq 0. \end{cases}$$

Ясно, что функция h непрерывна, а функция f в нуле разрывна. *Л. Кларк* построил пример (*The American Math. Monthly*, **82**, 845—846, (1975), когда h непрерывна везде, а f не является непрерывной нигде! Легко показать, что f всегда полунепрерывна снизу, т. е.

$$\lim_{x' \to x} f(x') \geqslant f(x).$$

Более того, если интеграл по всей прямой от неотрицательной полунепрерывной функции равен единице, то существует непрерывная плотность вероятности $h(x,\ y)$, для которой

$$f(x) = \int_{-\infty}^{\infty} h(x,\ y)\, dy.$$

(Лит.: Pelling M. J., Verbeek A. "On marginal density functions of continuous densities II", *The American Math. Monthly*, **84**, 364—365, (1977).)

в) Парадокс продавца газет

Продавец газет заказывает ежедневно N газет. С каждой проданной газеты он получает прибыль в b долларов и теряет c долларов на каждой газете, оставшейся непроданной. Каким нужно выбрать N, чтобы максимизировать ожидаемую прибыль? Число покупателей, очевидно, случайно. Предположим, что оно подчиняется распределению Пуассона с некоторым параметром λ, т. е. вероятность того, что число покупателей в точности равно n, есть $\lambda^n e^{-\lambda}/n!$. Положим $b = 1$, $c = 2$ и $\lambda = 10$, т. е. среднее число покупателей равно 10. Тогда можно показать, что следует заказывать в среднем 9 газет. Однако, очевидно, если продавец заказывает лишь 9 газет ежедневно, то среднее число покупателей уменьшится с 10 до 9; но в этом случае оптимальное число газет будет равно 8 и т. д. Объяснение этой парадоксальной ситуации состоит в том, что необходимо принимать во внимание убытки, возникающие из-за потери «потенциального покупателя», который уходит неудовлетворенным. Пусть d долларов составляют потери продавца, если он не может предложить какому-либо покупателю газету. (Значение d нельзя определить столь же просто, как значения b и c, по этой причине его, к сожалению, часто вообще не учитывают. Например, если $d = 1\$$, то для $\lambda = 10$ оптимальное значение N равно 10.)

В общем случае обозначим через X (случайное) число покупателей (теперь мы не предполагаем, что X имеет распределение Пуассона). Можно показать, что оптимальное значение N есть решение уравнения $P(X > N) \approx \frac{c}{b+c+d}$, т. е. если продавец газет запасается N экземплярами газет, где N — решение этого уравнения, то его ожидаемая прибыль будет максимальна.

(Лит.: Morse P. M., Kimball G. E. *Methods of Operations Research*, Wiley, New York, 1951.
DeGroot M. H. *Optimal Statistical Decisions*, McGraw Hill, New York, 1970. [Имеется перевод: Де Гроот М. Оптимальные статистические решения. — М.: Мир, 1974.])

г) Парадокс Кестена

Согласно усиленному закону больших чисел Колмогорова для независимых одинаково распределенных случайных величин $X_1, X_2, X_3, \ldots$ последовательность

$$\bar{X}_n = (X_1 + X_2 + \ldots + X_n)/n$$

сходится к некоторой постоянной M с вероятностью 1 тогда и только тогда, когда у величин X_i существуют математические ожидания; эти математические ожидания равны как раз M. Итак, если M существует, то с вероятностью 1 последовательность $\bar{X}_n$ очень «регулярна»: ее единственной предельной точкой является M. До какой степени последовательность $\bar{X}_n$ «иррегулярна», когда M не существует? В 1970 г. *Гарри Кестен* доказал, что множество предельных точек последовательности $\bar{X}_n$ может быть произвольным (неслучайным) замкнутым множеством, содержащим $-\infty$ и ∞ с вероятностью 1. Поэтому множество предельных точек может совпадать со всей числовой прямой, хотя пока неизвестно, какими свойствами в этом случае должна обладать функция распределения величин X_i.

(Лит.: Kesten H. "The limit points of a normalized random walk", *Annals of Math. Statist.*, **41**, 1173—1205, (1970).)

д) Парадокс стохастического гейзера

В 1962 г. Альфред Реньи поставил следующий вопрос. Рассмотрим гейзер, выбрасывающий фонтан с интервалами X_1, $X_2, \ldots$. Предположим, что эти интервалы являются независимыми одинаково распределенными случайными величинами.

Найдем моменты времени, когда происходили выбросы, $S_n = X_1 + X_2 + \ldots + X_n$. Как велики могут быть ошибки при определении с вероятностью 1 неизвестных распределений случайных величин X_i на основе полученных данных? Этот вопрос тесно связан со следующей проблемой. Пусть $S_n = X_1 + X_2 + \ldots + X_n$ и $T_n = Y_1 + Y_2 + \ldots + Y_n$ — частичные суммы независимых одинаково распределенных случайных величин (величины X_i, не обязательно не зависят от величин Y_i). Насколько близко S_n может быть к T_n, если распределения X_i и Y_i различны? Если $(S_n - T_n)/n$ стремится к нулю, то по усиленному закону больших чисел математические ожидания величин X_i и Y_i (при условии, что они существуют) не могут различаться (с вероятностью 1). Известный закон повторного логарифма устанавливает, что если у X существует стандартное отклонение $D(X)$, то

$$\lim_{n\to\infty} \sup \frac{S_n - E(S_n)}{\sqrt{2n \ln \ln n}} = D(X).$$

Следовательно, если даже $(S_n - T_n)/\sqrt{n}$ стремится к нулю, то стандартные отклонения у X и Y должны совпадать. Исследования, проведенные Скороходом и Штрассеном на основе теории броуновского движения, привели к предположению, что если, например, величина X ограничена, а Y нормально распределена, то $|S_n - T_n|$ не меньше $\sqrt[4]{n}$ (при достаточно больших n). Таким образом, $(S_n - T_n)/\sqrt[4]{n}$ не может сходиться к нулю (с вероятностью 1). В силу этих результатов считали, что моменты выброса у гейзера достаточно измерять с ошибкой, не превосходящей $\sqrt[4]{n}$. Большой неожиданностью явилось то, что после первых результатов *П. Ревеса* и *М. Черге* в 1974 г. *Я. Комлош*, *П. Майор* и *Г. Тушнади* показали, что S_n можно приблизить T_n настолько хорошо, что даже величина $(S_n - T_n)/\ln n$ остается ограниченной. Итак, при записи моментов выброса ошибки измерений должны быть в пределах $\ln n$. *П. Бартфай* доказал, что если ошибки измерения, деленные на $\ln n$, сходятся к нулю, то распределения интервалов между последовательными выбросами могут быть определены с вероятностью 1.

(Лит.: Csörgő M., Révész P. *Strong Approximations in Probability and Statistics*, Akadémiai Kiadó, Budapest, 1981.)

e) Парадокс вероятности в квантовой физике

Методы теории вероятностей широко использовались в физике уже в прошлом веке. Классическая статистическая физика начиналась с идеи о том, что равновесие системы (состоящей из большого числа частиц) есть наиболее вероятное состояние системы. Считалось, что методы статистической физики лишь приближенно описывают макроскопическое поведение системы. Однако в результате вероятностной интерпретации квантовой физики случайность и вероятность стали неотъемлемой частью всей физики. Вероятность стала столь же фундаментальным понятием как, например, энергия, а не чем-то используемым лишь в качестве приближения и без чего в принципе можно обойтись. Даже Эйнштейну, хотя он вовсе не был консервативен, не нравились эти радикальные изменения в основаниях физики. В письме к *Максу Борну* (получившему Нобелевскую премию за вероятностную интерпретацию волновой функции в квантовой механике) он написал, что верит в существование совершенных законов Природы: «Бог не играет в кости». В своем ответе Борн объяснил, что вместо решения большого числа дифференциальных уравнений в некоторых случаях можно получить приемлемые результаты, бросая игральную кость. С тех пор идеи Борна стали доминирующими. Случайность и вероятность — уже признанные понятия в физике. Эти изменения повлияли и на философию: механистический детерминизм потерял свою главенствующую роль. Современное состояние мира не полностью определяет его будущее состояние. На основе настоящих знаний можно определить лишь вероятность событий, которые произойдут в будущем. Однако это не означает признание агностицизма, так как законы проявления случайности познаваемы (именно этим и занимается теория вероятностей). Парадоксально, но физическая концепция вероятности не является простым применением математической вероятности в физике. Мотивы и дух обеих концепций различны. Согласно *Р. Фейнману*, которому в 1965 г. присуждена Нобелевская премия по физике, законы квантовой физики можно понять, опираясь на теорию вероятностей, возникающую из теории азартных игр, если применить законы теории вероятностей к большому числу частиц, однако эти законы не объясняют поведение *отдельного* электрона или протона. Волновая теория *де Бройля* и *Шрёдингера* и принцип неопределенности *Гейзенберга* привели, во многом благодаря работам *Борна*, к созданию между 1926 и 1929 гг. новой квантовой теории вероятностей. Математическая теория вероятностей Колмогорова была построена также приблизительно в это время. Выяснение связи между двумя типами теории вероятностей началось значительно позже, почти двадцать лет спустя; большую роль здесь сыграла работа *Г. Макки*, основанная на более ранних исследованиях фон Неймана. В конце концов была разработана общая единая теория вероятностей, которая включала в себя и классическую, и квантовую теории вероятностей (см. книгу Гаддера). Это разрешает противоречие и позволяет описать теорию вероятностей, основанную на общей структуре случайных событий.

(Лит.: Born M. *Natural Philosophy of Cause and Chance*, Dover Pub., New York, 1964.
Gudder S. P. *Stochastic Methods in Quantum Mechanics*, North-Holland, New York, 1979.)

ж) Парадокс криптографии

На протяжении нескольких тысячелетий истории криптографии специалисты по тайнописи придумывали все более и более изощренные шифры, а их противники соответственно находили способы их перехитрить с помощью все более и более эффективной техники разгадки шифров. *Эдгар Аллан По*, считавший себя мастером тайнописи, был убежден, что «...человеческая изобретательность не может придумать такой шифр, который бы человеческая находчивость не смогла разгадать». Первый перелом в истории криптографии наступил в двадцатые годы, когда были открыты «одноразовые шифры». Такие одноразовые шифры были впервые использованы в Германии и употреблялись повсюду на протяжении полувека. Различные типы одноразовых шифров считаются очень эффективными и сегодня постоянно используются во многих странах для специальных посланий. В знаменитой «горячей линии» между Вашингтоном и Москвой также применяются одноразовые шифры. Эти шифры действительно неразгадываемы в принципе, так как различные шифры-сдвиги кодируют каждый символ в исходном тексте, при этом всякий раз сдвиг выбирается случайным образом. Если букву «е» всегда кодировать как «t», то это был бы простой подстановочный шифр, который легко разгадать с помощью статистического анализа (так как «е» — наиболее часто употребляемая буква во многих языках). Однако, если «е» кодировать иногда как «а», иногда как «с», а порой как «w», и выбор замены осуществлять случайно, то такой одноразовый шифр нельзя разгадать даже в принципе. Шифрованный текст ничем не обнаруживает закодированное содержание. Недостаток подобной процедуры заключается в том, что одноразовые шифры можно использовать лишь однажды для единственного послания. Замечательное открытие специалистов по электронике из Станфордского университета *Уитфилда Диффи* и *Мартина Хеллмана* революционизировало всю область секретной связи. Под влиянием математической теории сложности они предложили в 1975 г. новый тип шифра, который разгадываем в принципе, но абсолютно не разгадываем на практике. Точнее, эти новые шифры можно разгадать, но лишь с помощью программ на ЭВМ, которые будут выполняться миллионы лет. Интересно, что процедура шифровки и расшифровки, предложенная Диффи и Хеллманом, несимметрична, т. е. если известен лишь метод шифровки, то с помощью вычислений нереально обнаружить способ расшифровки, а это обеспечивает абсолютную секретность. (Такой метод передачи шифрованных сообщений позволяет добиться того, что Диффи и Хеллман назвали функцией-ловушкой с односторонним входом.) Секрет может быть закодирован и раскрыт разными способами (и для расшифровки требуется более тонкий метод). Основная идея одностороннего шифрования очень проста: два числа легко перемножить, например, произведение чисел 101 и 211 вычисляется просто и оно равно 21311, но если нам нужно найти два числа, которые больше единицы и их произведение равно 21311, то для того, чтобы установить, что единственно возможное решение — это 101 и 211, потребуется значительно больше времени. Естественно, существуют вычислительные алгоритмы для разложения чисел на множители, но в случае, когда число состоит из 40—50 цифр, время, необходимое для нахождения множителей, составило бы миллионы лет. На основе теории простых чисел была найдена простая функция-ловушка: ключ для шифровки зависит лишь от произведения двух простых чисел, в то время как для расшифровки текста необходимо знать сами простые числа. Остановимся подробнее на этой функции-ловушке!

Пусть p и q — два больших случайных простых числа. Произведение n этих двух чисел и еще одно случайное число E являются ключом шифра (E, n) для пользователя, при этом ключ не обязательно держать в секрете, его можно поместить в какой-либо общий файл, например, в телефонный справочник. Для использования ключа отправитель сначала преобразует свое послание в цепочку цифр, которую он затем разбивает на блоки B_1, B_2, ... Каждое число B_i в исходном тексте должно быть между 0 и $n-1$. Для каждого числа в исходном тексте B_i отправитель вычисляет число в шифрованном тексте $C_i = B_i^E$ по модулю n (т. е. C_i — остаток при делении E-й степени числа B_i на n). Такой общедоступный способ шифровки основан на том факте, что хотя нахождение больших простых чисел (p и q) с вычислительной точки зрения просто, однако разложение произведения двух таких чисел на множители с помощью вычислений в настоящее время нереально, так что, зная лишь (E, n) и C_i, безнадежно пытаться найти B_i. Для расшифровки текста C_1, C_2, ... пользователю нужны n и секретный ключ для расшифровки D, получаемый из простых сомножителей p и q числа n. D есть мультипликативно обратное к E число по модулю $(p-1)(q-1)$, т. е. произведение ED по модулю $(p-1)(q-1)$

равно 1. (Произведение $(p-1)(q-1)$ дает количество целых чисел между 1 и n, не имеющих с n общих делителей.) После всего этого получатель легко находит B_i:

$$C_i^D = (B_i^E)^D = B_i \text{ по модулю } n.$$

Этот метод разработали *Ривест*, *Шамир* и *Адлеман* и он называется системой RSA.

(Лит.: Hellman M. E. "The mathematics of public-key cryptography", *Sci. Amer.*, 241, 130—139, (1979).
Simmons G. J. "Cryptology, the mathematics of secure communication", *The Math. Intelligencer*, 1, 233—246, (1979).
Shamir A. "A polinomial time algorithm for breaking Merkle — Hellman cryptosystems", *Research Announcement*, 1982.)

з) *Парадокс поэзии и теории информации*

В качестве последнего парадокса в этой книге приведу слова моего покойного учителя профессора *Альфреда Реньи:*

«С тех пор, как я начал заниматься теорией информации, я часто размышляю над краткостью стихотворений; почему одна строка стихотворения содержит значительно больше «информации», чем очень короткая телеграмма такой же длины. Удивительное богатство значений в литературных трудах кажется противоречит законам теории информации. Ключом к этому парадоксу, я думаю, является понятие «резонанса». Писатель не только сообщает нам информацию, но и играет на струнах языка с таким мастерством, что наш разум и даже само подсознание резонируют. Поэт с помощью удачного слова может вызвать цепочку идей, эмоций и воспоминаний. В этом смысле труд писателя — волшебство.»

ГЛАВА V

ПАРАДОКСОЛОГИЯ

> В глубине души мы верим в реальность математики, но когда философы указывают нам на парадоксы в математике, мы, конечно, спешим спрятаться за абстракциями и говорим: математика — это лишь комбинация бессмысленных символов...
>
> *Ж. А. Э. Дьедонне, 1970*

Как и большинство разделов науки, математика — это также и история парадоксов. Величайшие открытия, как правило, разрешали величайшие парадоксы (вспомните Дарвина и Эйнштейна) и в то же время они в свою очередь были источником новых парадоксов. Метод обучения Сократа, по которому о новых идеях надо узнавать через парадоксы, является самым фундаментальным, так как процесс научного познания сам опирается на парадоксы.

Для развития дедуктивной математики огромную роль сыграло то, что (несмотря на мнение пифагорийцев, что «все есть число», т. е. целое число) существуют отрезки (например, диагональ и сторона квадрата), отношение длин которых не является отношением целых чисел, а это означает, что такое отношение с точки зрения пифагорийцев — не число. (В соответствии с современной терминологией оно не является рациональным числом.) Этот парадокс «несоизмеримости» привел к распаду пифагорийской школы и ослаблению числового мистицизма, а также к евклидовой геометрии (где роль чисел стали играть геометрические фигуры) и к «математическому идеализму» Платона (на практике «несоизмеримость» нельзя проверить непосредственно, поэтому, согласно Платону, опыт не может привести к истинному знанию). Величайший парадокс математики средних веков состоял в том, что «ничто», т. е. нуль, следовало рассматривать как нечто и как-то его обозначать. В результате, благодаря индоарабскому способу записи чисел, вычисления в значительной степени облегчились. Позднее возникли парадоксы, связанные с отрицательными и затем комплексными числами. Например, один из парадоксов утверждал, что равенство $(-1):1=1:(-1)$ невозможно, так как отношение меньшего числа к большему не может равняться отношению большего числа к меньшему. В наше время во всех разделах математики появилось несколько новых парадоксов, начиная от разрешимости алгебраического уравнения и кончая геометрией Бойаи. Интересно, что уже в первой половине прошлого века чешский математик из Праги Б. Больцано посвятил целую книгу парадоксам бесконечного («Paradoxen des Unendlichen»), хотя наиболее любопытные парадоксы бесконечного появились лишь после опубликования в 1872 г. труда Г. Кантора по теории множеств. Ведущие математики прошлого века такие, как Гаусс, Коши, Кронекер, Пуанкаре и другие, отвергали понятие актуальной бесконечности и приписывали бесконечности лишь символическое значение. Однако теория Кантора, в которой используется понятие актуальной бесконечности, является краеугольным камнем современной математики, хотя следует подчеркнуть, что «ужас перед бесконечным» все еще не исчез. На самом деле новые парадоксы увеличивают число приверженцев «финитизма». Похожий страх перед случайностью существует до сих пор. Математические парадоксы бесконечного и случайного чрезвычайно важны, так как оба эти понятия оказывают фундаментальное влияние на наши взгляды и философские представления. Теория вероятностей развивалась как отражение в символах случайной вселенной, поэтому есть надежда, что парадоксы, приведенные в этой книге, помогут читателю действовать наилучшим образом в нашем случайном мире.

> Все постоянство в зыбкости одной,
> Все смутно пред глазами очевидца,
> Я сложность вижу в истине простой,
>
> . . .
>
> Наукой правит случай — не закон...
>
> *Вийон Ф. Баллада, написанная для состязания в Блуа* [1])

---

[1]) Вийон Ф. Стихи: сборник/Сост. Г. К. Косиков. — М.: Радуга, 1984. Отрывок из стихотворения дан в переводе Ю. А. Кожевникова.

ОБОЗНАЧЕНИЯ

$n!$	$= 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$
$\binom{n}{k}$	$= \dfrac{n(n-1)(n-2)\ldots(n-k+1)}{1 \cdot 2 \cdot \ldots \cdot k}$
$P(A)$	вероятность события A
$P(\bar{A})$	вероятность события, противоположного A ($\bar{A}$)
$P(AB)$	вероятность совместного осуществления событий A и B
$P(A \mid B)$	вероятность события A при условии, что произошло событие B
$\bar{X}$	$= (X_1 + X_2 + \ldots + X_n)/n$
$\hat{\theta}$	оценка параметра θ
$E(X)$	или EX — математическое ожидание (или среднее значение) случайной величины X
$D(X)$	стандартное отклонение случайной величины X
$\Gamma(z)$	$= \int_0^\infty e^{-t} t^{z-1}\, dt$, если действительная часть комплексного числа z положительна; $\Gamma(z+1) = z\Gamma(z)$; $\Gamma(n+1) = n!$; $\Gamma(1/2) = \sqrt{\pi}$
$\Phi(x)$	$= \dfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du$

Таблица 1. **Стандартная нормальная функция распределения**

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du, \quad [\Phi(-x) = 1 - \Phi(x)]$$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,5000	0,51	0,6950	1,02	0,8461	1,53	0,9370	2,08	0,9812
0,01	0,5040	0,52	0,6985	1,03	0,8485	1,54	0,9382	2,10	0,9821
0,02	0,5080	0,53	0,7019	1,04	0,8508	1,55	0,9394	2,12	0,9830
0,03	0,5120	0,54	0,7054	1,05	0,8531	1,56	0,9406	2,14	0,9838
0,04	0,5160	0,55	0,7088	1,06	0,8554	1,57	0,9418	2,16	0,9846
0,05	0,5199	0,56	0,7123	1,07	0,8577	1,58	0,9429	2,18	0,9854
0,06	0,5239	0,57	0,7157	1,08	0,8599	1,59	0,9441	2,20	0,9861
0,07	0,5279	0,58	0,7190	1,09	0,8621	1,60	0,9452	2,22	0,9868
0,08	0,5319	0,59	0,7224	1,10	0,8643	1,61	0,9463	2,24	0,9875
0,09	0,5359	0,60	0,7257	1,11	0,8665	1,62	0,9474	2,26	0,9881
0,10	0,5398	0,61	0,7291	1,12	0,8686	1,63	0,9484	2,28	0,9887
0,11	0,5438	0,62	0,7324	1,13	0,8707	1,64	0,9495	2,30	0,9893
0,12	0,5478	0,63	0,7352	1,14	0,8729	1,65	0,9505	2,32	0,9898
0,13	0,5517	0,64	0,7389	1,15	0,8749	1,66	0,9515	2,34	0,9904
0,14	0,5557	0,65	0,7422	1,16	0,8770	1,67	0,9525	2,36	0,9909
0,15	0,5596	0,66	0,7454	1,17	0,8790	1,68	0,9535	2,38	0,9913
0,16	0,5636	0,67	0,7486	1,18	0,8810	1,69	0,9545	2,40	0,9918
0,17	0,5675	0,68	0,7517	1,19	0,8830	1,70	0,9554	2,42	0,9922
0,18	0,5714	0,69	0,7549	1,20	0,8849	1,71	0,9564	2,44	0,9927
0,19	0,5753	0,70	0,7580	1,21	0,8869	1,72	0,9572	2,46	0,9931
0,20	0,5793	0,71	0,7611	1,22	0,8888	1,73	0,9582	2,48	0,9934
0,21	0,5832	0,72	0,7642	1,23	0,8907	1,74	0,9591	2,50	0,9938
0,22	0,5871	0,73	0,7673	1,24	0,8925	1,75	0,9599	2,52	0,9941
0,23	0,5910	0,74	0,7703	1,25	0,8944	1,76	0,9608	2,54	0,9945
0,24	0,5948	0,75	0,7734	1,26	0,8962	1,77	0,9616	2,56	0,9948
0,25	0,5987	0,76	0,7764	1,27	0,8980	1,78	0,9625	2,58	0,9951
0,26	0,6026	0,77	0,7794	1,28	0,8997	1,79	0,9633	2,60	0,9953
0,27	0,6064	0,78	0,7823	1,29	0,9015	1,80	0,9641	2,62	0,9956
0,28	0,6103	0,79	0,7853	1,30	0,9032	1,81	0,9649	2,64	0,9959
0,29	0,6141	0,80	0,7881	1,31	0,9049	1,82	0,9656	2,66	0,9961
0,30	0,6179	0,81	0,7910	1,32	0,9066	1,83	0,9664	2,68	0,9963
0,31	0,6217	0,82	0,7939	1,33	0,9082	1,84	0,9671	2,70	0,9965
0,32	0,6255	0,83	0,7967	1,34	0,9099	1,85	0,9678	2,72	0,9967
0,33	0,6293	0,84	0,7995	1,35	0,9115	1,86	0,9686	2,74	0,9969
0,34	0,6331	0,85	0,8023	1,36	0,9131	1,87	0,9693	2,76	0,9971
0,35	0,6368	0,86	0,8051	1,37	0,9147	1,88	0,9699	2,78	0,9973
0,36	0,6406	0,87	0,8078	1,38	0,9162	1,89	0,9706	2,80	0,9974
0,37	0,6443	0,88	0,8106	1,39	0,9177	1,90	0,9713	2,82	0,9976
0,38	0,6480	0,89	0,8133	1,40	0,9192	1,91	0,9719	2,84	0,9977
0,39	0,6517	0,90	0,8159	1,41	0,9207	1,92	0,9726	2,86	0,9979
0,40	0,6554	0,91	0,8186	1,42	0,9222	1,93	0,9732	2,88	0,9980
0,41	0,6591	0,92	0,8212	1,43	0,9236	1,94	0,9738	2,90	0,9981
0,42	0,6628	0,93	0,8238	1,44	0,9251	1,95	0,9744	2,92	0,9982
0,43	0,6664	0,94	0,8264	1,45	0,9265	1,96	0,9750	2,94	0,9984
0,44	0,6700	0,95	0,8289	1,46	0,9279	1,97	0,9756	2,96	0,9985
0,45	0,6736	0,96	0,8315	1,47	0,9292	1,98	0,9761	2,98	0,9986
0,46	0,6722	0,97	0,8340	1,48	0,9306	1,99	0,9767	3,00	0,9987
0,47	0,6808	0,98	0,8365	1,49	0,9319	2,00	0,9772	3,20	0,9993
0,48	0,6844	0,99	0,8389	1,50	0,9332	2,02	0,9783	3,40	0,9996
0,49	0,6879	1,00	0,8413	1,51	0,9345	2,04	0,9793	3,60	0,9998
0,50	0,6915	1,01	0,8438	1,52	0,9367	2,06	0,9803	3,80	0,9999

Таблица 2. **Первые 2 000 знаков числа π. Действительно ли случайно появление этих десятичных знаков или оно подчиняется некоторому закону?**

$\pi = 3,\ +$

1415926535	8979323846	2643383279	5028841971	6939937510
8214808641	3282306647	0938446095	5058223172	5359408128
4428810975	6659334461	2847564823	3786783165	2712019091
7245870066	0631558817	4881520920	9628292540	9171536436
3305727036	5759591953	0921861173	8193261179	3105118548
9833673362	4406566430	8602139494	6395224737	1907021798
0005681271	4526356082	7785771342	7577896091	7363717872
4201995611	2129021960	8640344181	5981362977	4771309960
5024459455	3469083026	4252230825	3344685035	2619311881
5982534904	2875546873	1159562863	8823537875	9375195778
3809525720	1065485863	2788659361	5338182796	8230301952
5574857242	4541506959	5082953311	6861727855	8890750983
8583616035	6370766010	4710181942	9555961989	4676783744
9331367702	8989152104	7521620569	6602405803	8150193511
6782354781	6360093417	2164121992	4586315030	2861829745
3211653449	8720275596	0236480665	4991198818	3479775356
8164706001	6145249192	1732172147	7235014144	1973568548
4547762416	8625189835	6948556209	9219222184	2725502542
8279679766	8145410095	3883786360	9506800642	2512520511
0674427862	2039194945	0471237137	8696095636	4371917287
9465764078	9512694683	9835259570	9825822620	5224894077
4962524517	4939965143	1429809190	6592509372	2169646151
6868386894	2774155991	8559252459	5395943104	9972524680
4390451244	1365497627	8079771569	1435997700	1296160894
0168427394	5226746767	8895252138	5225499546	6672782398
1507606947	9451096596	0940252288	7971089314	5669136867
9009714909	6759852613	6554978189	3129784821	6829989487
5428584447	9526586782	1051141354	7357395231	1342716610
0374200731	0578539062	1983874478	0847848968	3321445713
8191197939	9520614196	6342875444	0643745123	7181921799
5679452080	9514655022	5231603881	9301420937	6213785595
0306803844	7734549202	6054146659	2520149744	2850732518
1005508106	6587969981	6357473638	4052571459	1028970641
2305587631	7635942187	3125147120	5329281918	2618612586
9229109816	9091528017	3506712748	5832228718	3520935396
6711136990	8658516398	3150197016	5151168517	1437657618
8932261854	8963213293	3089857064	2046752590	7091548141
2332609729	9712084433	5732654893	8239119325	9746366730
1809377344	4030707469	2112019130	2033038019	7621101100
2131449576	8572624334	4189303968	6426243410	7732269780
6655730925	4711055785	3763466820	6531098965	2691862056
3348850346	1136576867	5324944166	8039626579	7877185560
7002378776	5913440171	2749470420	5622305389	9456131407
6343285878	5698305235	8089330657	5740679545	7163775254
0990796547	3761255176	5675135751	7829666454	7791754011
9389713111	7904297828	5647503203	1986915140	2870808599
8530614228	8137585043	0633217518	2979866223	7172159160
9769265672	1463853067	3609657120	9180763832	7166416274
6171196377	9213375751	1495950156	6049631862	9472654736
6222247715	8915049530	9844489333	0963408780	7693259939

$\pi=3$, +

5820974944	5923078164	0628620899	8628034825	3421170679
4811174502	8410270193	8521105559	6446229489	5493038196
4564856692	3460348610	4543266482	1339360726	0249141273
7892590360	0113305305	4882046652	1384146951	9415116094
0744623799	6274956735	1885752724	8912279381	8301194912
6094370277	0539217176	2931767523	8467481846	7669405132
1468440901	2249534301	4654958537	1050792279	6892589235
5187072113	4999999837	2978049951	0597317328	1609631859
7101000313	7838752886	5875332083	8142061717	7669147303
1857780532	1712268066	1300192787	6611195909	2164201989
0353018529	6899577362	2599413891	2497217752	8347913151
8175463746	4939319255	0604009277	0167113900	9848824012
9448255379	7747268471	0404753464	6208046684	2590694912
2533824300	3558764024	7496473263	9141992726	0426992279
5570674983	8505494588	5869269956	9092721079	7509302955
6369807426	5425278625	5181841757	4672890977	7727938000
1613611573	5255213347	5741849468	4385233239	0739414333
5688767179	0494601653	4668049886	2723279178	6085784383
7392984896	0841284886	2694560424	1965285022	2106611863
4677646575	7396241389	0865832645	9958133904	7802759009
2671947826	8482601476	9909026401	3639443745	5305068203
5709858387	4105978859	5977297549	8930161753	9284681382
8459872736	4469584865	3836736222	6260991246	0805124388
4169486855	5848406353	4220722258	2848864815	8456028506
6456596116	3548862305	7745649803	5593634568	1743241125
2287489405	6010150330	8617928680	9208747609	1782493858
2265880485	7564014270	4775551323	7964145152	3746234364
2135969536	2314429524	8493718711	0145765403	5902799344
8687519435	0643021845	3191048481	0053706146	8067491927
9839101591	9561814675	1426912397	4894090718	6494231961
6638937787	0830390697	9207734672	2182562599	6615014215
6660021324	3408819071	0486331734	6496514539	0579626856
4011097120	6280439039	7595156771	5770042033	7869936007
7321579198	4148488291	6447060957	5270695722	0917567116
5725121083	5791513698	8209144421	0067510334	6711031412
3515565088	4909989859	9823873455	2833163550	7647918535
6549859461	6371802709	8199430992	4488957571	2828905923
5836041428	1388303203	8249037589	8524374417	0291327656
4492932151	6084244485	9637669838	9522868478	3123552658
2807318915	4411010446	8232527162	0105265227	2111660396
4769312570	5863566201	8558100729	3606598764	8611791045
8455296541	2665408530	6143444318	5867697514	5661406800
1127000407	8547332699	3908145466	4645880797	2708266830
2021149557	6158140025	0126228594	1302164715	5097925923
2996148903	0463994713	2962107340	4375189573	5961458901
0480109412	1472213179	4764777262	2414254854	5403321571
7716692547	4873898665	4949450114	6540628433	6639379003
8888007869	2560290228	4721040317	2118608204	1900042296
4252308177	0367515906	7350235072	8354056704	0386743513
7805419341	4473774418	4263129860	8099888687	4132604721

Продолжение табл. 2.

$\pi=3$, +				
5695162396	5864573021	6315981931	9516735381	2974167729
4037014163	1496589794	0924323789	6907069779	4223625082
5578297352	3344604281	5126272037	3431465319	7777416031
3162499341	9131814809	2777710386	3877343177	2075456545
3166636528	6193266863	3606273567	6303544776	2803504507
9456127531	8134078330	3362542327	8394497538	2437205835
0408591337	4641442822	7726346594	7047458784	7787201927
8350493163	1284042512	1925651798	0694113528	0131470130
9562586586	5570552690	4965209858	0338507224	2648293972
4803048029	0058760758	2510474709	1643961362	6760449256
2901618766	7952406163	4252257719	5429162991	9306455377
2540790914	5135711136	9410919139	2351910760	2082520261
2784768472	6860849003	3770242429	1651300500	5168323364
1960121228	5993716231	3017114448	4640903890	6449544400
2283345085	0486082503	9302133219	7155184306	3545500766
4611996653	8581538420	5685338621	8672523340	2830871123
2184564622	0134967151	8819097303	8119800497	3407239610
9567302292	1913933918	5680344903	9820595510	0226353536
2711172364	3435439478	2218185286	2608514006	6604433258
0516553790	6866273337	9958511562	5784322988	2737231989
4460477464	9159950549	7374256269	0104903778	1986835938
3634655379	4986419270	5638728317	4872332083	7601123029
0126901475	4668476535	7616477379	4675200490	7571555278
2772190055	6148425551	8792530343	5139844253	2234157623
2276930624	7435363256	9160781547	8181152843	6679570611
3776700961	2071512491	4043027253	8607648236	3414334623
9164219399	4907236234	6468441173	9403265918	4044378051
1266830240	2929525220	1187267675	6220415420	5161841634
9104140792	8862150784	2451670908	7000699282	1206604183
9598470356	2226293486	0034158722	9805349896	5022629174
5771028402	7998066365	8254889264	8802545661	0172967026
5178609040	7086671149	6558343434	7693385781	7113864558
6161528813	8437909904	2317473363	9480457593	1493140529
9203767192	2033229094	3346768514	2214477379	3937517034
8244625759	1633303910	7225383742	1821408835	0865739177
3408005355	9849175417	3818839994	4697486762	6551658276
4043523117	6006651012	4120065975	5851276178	5838292041
7267507981	2554709589	0455635792	1221033346	6974992356
7429958180	7247162591	6685451333	1239480494	7079119153
9289647669	7583183271	3142517029	6923488962	7668440323
9588970695	3653494060	3402166544	3755890045	6328822505
6909411303	1509526179	3780029741	2076651479	3942590298
6922210327	4889218654	3648022967	8070576561	5144632064
2248261177	1858963814	0918390367	3672220888	3215137556
2543709069	7939612257	1429894671	5435784687	8861444581
5510500801	9086996033	0276347870	8108175450	1193071412
1596131854	3475464955	6978103829	3097164651	4384070070
3084076118	3013052793	2054274628	6540360367	4532865105
5020141020	6723585020	0724522563	2651341055	9240190274
2645600162	3742880210	9276457931	0657922955	2498872758

$\pi=3$, +				
4526379628	9612691572	3579866205	7340837576	6873884266
1443545419	5762584735	6421619813	4073468541	1176688311
9467300244	3450054499	5399742372	3287124948	3470604406
3583446283	9476304775	6450150085	0757894954	8931393944
7643800125	4365023714	1278346792	6101995585	2247172201
7235467456	4239058585	0216719031	3952629445	5439131663
9187497519	1011472315	2893267725	3391814660	7300089027
1054967973	1001678708	5069420809	2232908070	3832634534
7550493412	6787643674	6384902063	9640197666	8559232565
3845534372	9141446513	4749407848	8442377217	5154334260
1370924353	0136776310	8491351615	6422698475	0743032971
9637637076	1799191920	3579582007	5956053023	4626775794
8135789400	5599500183	5425118417	2136055727	5221035268
3889710770	8229310027	9766593583	8758909395	6881485602
3624078250	5270487581	6470324581	2908783952	3245323789
8127810217	9912317416	3058105545	9880130048	4562997651
6165356024	7338078430	2865525722	2753049998	8370153487
3938352293	4588832255	0887064507	5394739520	4396807906
5306184548	5903798217	9945996811	5441974253	6344399602
0072629339	5505482395	5713725684	0232268213	0124767945
4555711266	0396503439	7896278250	0161101532	3516051965
0135295661	3978884860	5097860859	5701773129	8155314951
4466283998	8060061622	9848616935	3373865787	7359833616
4837011709	6721253533	8758621582	3101331038	7766827211
7547570354	6994148929	0413018638	6119439196	2838870543
3055861598	9991401707	6983854831	8875014293	8908995068
1651716677	6672793035	4851542040	2381746089	2328291703
4453184040	4185540104	3513483895	3120132637	8369283580
4490453627	5600112501	8433560736	1222765949	2783937064
6202215368	4941092605	3876887148	3798955999	9112099164
8158686952	5069499364	6101756850	6016714535	4315814801
0556211671	5468484778	0394475697	9804263180	9917564228
0259627304	3067931653	1149401764	7376938735	1409336183
5955595546	5196394401	8218409984	1248982623	6737714672
1431885988	2769449011	9321296827	1588841338	6943468285
4165654797	3670548558	0445865720	2276378404	6682337895
4502285157	9795797647	4670228409	9956160156	9108903845
8365620945	5543461351	3415257006	5974881916	3413595567
7948282769	3895352178	3621850796	2977851461	8843271922
8442253547	2653098171	5784478342	1582232702	0690287232
6282727632	1779088400	8786148022	1475376578	1058197022
8307332335	6034846531	8730293026	6596450137	1837542889
2679332107	2686870768	0626399193	6196504409	9542167627
1591845821	5625181921	5523370960	7483329234	9210345146
3996122816	1521931488	8769388022	2810830019	8601654941
8052187292	4610867639	9589161458	5505839727	4209809097
4627134946	3685318154	6969046696	8693925472	5194139929
5888676950	1612497228	2040303995	4632788306	9597624936
3547891129	2547696176	0050479749	2806072126	8039226911
2968106203	7765788371	6690910941	8074487814	0490755178

$\pi=3$, +				
5667227966	1988578279	4848855834	3975187445	4551296563
7634180703	9476994159	7915945300	6975214829	3366555661
4599041608	7532018683	7937023488	8689479151	0716378529
1384987137	5704710178	7957310422	9690666702	1449863746
1963403911	4732023380	7150952220	1068256342	7471646024
0273900749	7297363549	6453328886	9844061196	4961627734
4810688732	0685990754	0792342402	3009259007	0173196036
4495397907	0903023460	4614709616	9688688501	4083470405
2431974404	7718556789	3482308934	1068287027	2280973624
0630792615	9599546262	4629707062	5948455690	3471197299
3831591256	8989295196	4272875739	4691427253	4366941532
0298865925	7866285612	4966552353	3829428785	4253404830
0598135220	5117336585	6407826484	9427644113	7639386692
6977631279	5722672655	5625962825	4276531830	0134070922
7099558561	1349802524	9906698423	3017350358	0440811685
2669178352	5870785951	2983441729	5351953788	5534573742
1080485485	2635722825	7682034160	5048466277	5045003126
2432227188	5159740547	0214828971	1177792376	1225788734
2778944236	2167411918	6269439650	6715157795	8675648239
8783419689	6861181558	1587360629	3860381017	1215855272
6435495561	8689641122	8214075330	2655100424	1048967835
3100735477	0549815968	0772009474	6961343609	2861484941
0558642219	6483491512	6390128038	3200109773	8680662877
8936793008	1699805365	2027600727	7496745840	0283624053
6640075094	2608788573	5796037324	5141467867	0368809880
7391253835	5915031003	3303251117	4915696917	4502714943
2342548326	1119128009	2825256190	2052630163	9114772473
1112635835	5387136101	1023267987	7564102468	2403226483
9630878154	3221166912	2464159117	7673225326	4335686146
8850280054	1436131462	3082102594	1737562389	9420751367
4725470316	5661399199	9682628247	2706413362	2217892390
7386480584	7268954624	3882343751	7885201439	5600571048
3604103182	3507365027	7859089757	8272731305	0488939890
9077271307	0686917092	6462548423	2407485503	6608013604
5823626729	3264537382	1049387249	9669933942	4685516483
5220935701	6263846485	2851490362	9320199199	6882851718
5982311225	0628905854	9145097157	5539002439	3153519090
7806101371	5004489917	2100222013	3501310601	6391541589
0594741234	1933984202	1874564925	6443462392	5319531351
9149419394	0608572486	3968836903	2655643642	1664425760
8153740996	1545598798	2598910937	1712621828	3025848112
2994972574	5303328389	6381843944	7707794022	8435988341
1000162076	9363684677	6413017819	6593799715	5746854194
6778113949	8616478747	1407932638	5873862473	2889645643
2961273998	4134427260	8606187245	5452360643	1537101127
5829486191	9667091895	8089833201	2103184303	4012849511
7253558119	5840149180	9692533950	7577840006	7465526031
0425845988	4199076112	8725805911	3935689601	4316682831
7903688786	8789493054	6955703022	6190095020	7643349335
1786227363	7169757741	8302398600	6591481616	4049449650

$\pi=3,\ +$				
4434803966	4205579829	3680435220	2770984294	2325330225
5678736400	5366656416	5473217043	9035213295	4352916941
0234529244	9773659495	6305100742	1087142613	4974595615
4595280824	3694457897	7233004876	4765241339	0759204340
3354400515	2126693249	3419672977	0415956837	5355516673
4951827369	5588220757	3551766515	9895519098	6665393549
2254756478	9406475483	4664776041	1463233905	6513433068
4607429586	9913829668	2468185710	3188790652	8703665083
8093996270	6074726455	3992539944	2808113736	9433887294
6409089418	0595343932	5123623550	8134949004	3642785271
3610045373	0488198551	7065941217	3524625895	4873016760
8330701653	7228563559	1525347844	5981831341	1290019992
4803118364	4536985891	7544264739	9882284621	8449008777
3343657791	6012809317	9401718598	5999338492	3549564005
5265311709	9570899427	3287092584	8789443646	0050410892
6085902908	1765155780	3905946408	7350612322	6112009373
2008007998	0492548534	6941469775	1649327095	0493463938
7718819682	5462981268	6858170507	4027255026	3329044976
9391760426	0176338704	5499017614	3641204692	1823707648
6683008238	3404656475	8804051380	8016336388	7421637140
2858829024	3670904887	1181900904	9453314421	8287661810
7850171807	7930681085	4690009445	8995279424	3981392135
9239718014	6134324457	2640097374	2570073592	1003154150
4603726341	6554259027	6018348403	0681138185	5105979705
6097164258	4975951380	6930944940	1515422221	9432913021
3151558854	0392216409	7229101129	0355218157	6282328318
3148573910	7775874425	3876117465	7867116941	4776421441
4641766369	8066378576	8134920453	0224081972	7856471983
1865452226	8126887268	4459684424	1610785401	6768142080
2751674573	1891894562	8352570441	3354375857	5342698699
3176085428	9437339356	1889165125	0424404008	9527198378
1194988423	9060613695	7342315590	7967034614	9143447886
0992391350	3373250855	9826558670	8924261242	9473670193
6689511840	0936686095	4632500214	5852930950	0009071510
2611341461	1068026744	6637334375	3407642940	2668297386
3953669134	5222444708	0459239660	2817156551	5656661113
2107119457	3002438801	7661503527	0862602537	8817975194
5780371177	9277522597	8742891917	9155224171	8958536168
0331147639	4911995072	8584306583	6193536932	9699289837
7914710869	9843157337	4964883529	2796328220	7629472823
3890119682	2142945766	7580718653	8065064870	2612289282
0035838542	3897354243	9564755568	4095224844	5541392394
6334893748	4391297423	9143365936	0410035234	3777065888
5987746676	3847946650	4074111825	6583788784	5485814896
4680977870	4464094758	2803487697	5894822824	1239292960
6203534280	1441276172	8583024355	9830032042	0245120728
4461670508	2768277222	3534191102	6341631571	4740612385
7632356732	5417073420	8173322304	6298799280	4908514094
9106024545	0864536289	3545686295	8531315337	1838682656
1173213138	9574706208	8474802365	3710311508	9842799275

Продолжение табл. 2.

$\pi=3$, +

4426853277	9743113951	4357417221	9759799359	6852522857
4059909935	0500081337	5432454635	9675048442	3528487470
8654489377	6979566517	2796623267	1481033864	3913751865
3471606325	8306498297	9551010954	1836235030	3094530973
8992161255	2559770143	6858943585	8775263796	2559708167
7772370041	7808419423	9487254068	0155603599	8390548985
1345308939	0620467843	8778505423	9390524731	3620129476
7689631148	1090220972	4520759167	2970078505	8071718638
5203802786	0990556900	1341371823	6837099194	9516489600
4639138363	1857456981	4719620184	1080961884	6054560390
3066988317	6833100113	3108690421	9390310801	4378433415
6746964066	6531527035	3254671126	6722246055	1199581831
3936307463	0569010801	1494271410	0939136913	8107258137
0373572652	7922417373	6057511278	8721819084	4900617801
6322439372	6562472776	0378908144	5883785501	9702843779
6029841669	2254896497	1560698119	2186584926	7704039564
1212415363	7451500563	5070127815	9267142413	4210330156
9300806260	1809623815	1613669033	4111138653	8510919367
7086806445	0969865488	0168287434	3786126543	8158342807
9025100158	8827216474	5006820704	1937615845	4712318346
2264482091	0235647752	7230820810	6351889915	2692889108
5904211844	9499077899	9200732947	6805868577	8787209829
6814671769	5976099421	0036183559	1387778176	9845875810
1338413385	3684211978	9389001852	9569196780	4554482858
5726949518	1795897546	9399264219	7915523385	7662316762
6777432242	7680913236	5449485366	7680000010	6526248547
5453076511	6803337322	2651756622	0752695179	1442252808
2754257508	6765511785	9395002793	3895920576	6827896776
8271937831	2654961745	9970567450	7183320650	3455664403
7842645676	3388188075	6561216896	0504161139	0390639601
6464411918	5682770045	7424343402	1672276445	5893301277
0545886056	4550133203	5786454858	4032402987	1709348091
0987399876	6973237665	7370158080	6822904599	2123661689
3216142802	1497633991	8983548487	5625298752	4238730775
2606163364	3296406335	7281070788	7581640438	1485018841
9006664080	6314077757	7257056307	2940049294	0302420498
2827105784	3197535417	9501134727	3625774080	2134768260
8245026792	6594205550	3958792298	1852648007	0683765041
1964965403	2187271602	6485930490	3978748958	9066127250
3223810158	7444505286	6523802253	2843891375	2738458923
3300538621	6347988509	4695472004	7952311201	5043293226
2630971749	5072127248	4794781695	7296142365	8595782090
7557971449	9246540386	8179921389	3469244741	9850973346
8409146698	5692571507	4315740793	8053239252	3947755744
2643744980	5596103307	9941453477	8457469999	2128599999
6542616968	5867883726	0958774567	6182507275	9929508931
8172932393	0106766386	8240401113	0402470073	5085782872
1465242385	7762550047	4852964768	1479546700	7050347999
1510102436	5553522306	9061294938	8599015734	6610237122
0277722610	2544149221	5765045081	2067717357	1202718024

Продолжение табл. 2.

$\pi=3$, +				
4786724229	2465436680	0980676928	2382806899	6400482435
2168895738	3798623001	5937764716	5122893578	6015881617
9906655418	7639792933	4419521541	3418994854	4473456738
3220777092	1201905166	0962804909	2636019759	8828161332
7723554710	5859548702	7908143562	4014517180	6246436267
3114771199	2606381334	6776879695	9703098339	1307710987
7152807317	6790770715	7213444730	6057007334	9243693113
4781643788	5185290928	5452011658	3934196562	1349142415
8584783163	0577775606	8887644624	8246857926	0395352773
2742042083	2085661190	6254543372	1315359584	5068772460
9914037340	4328752628	8896399587	9475729174	6426357455
8798531887	7058429725	9167781314	9699009019	2116971737
3503895170	2989392233	4517220138	1280696501	1784408745
6198690754	8516026327	5052983491	8740786680	8818338510
8282949304	1377655279	3975175461	3953984683	3936383047
2827892125	0771262946	3229563989	8989358211	6745627010
3685406643	1939509790	1906996395	5255300545	0580685501
1920419947	4553859381	0234395544	9597782779	0237421617
8856986705	4315470696	5747458550	3323233421	0730154594
8757141595	7811196358	3300594087	3068121602	8764962867
1465741268	0492564879	8556145372	3478673303	9046883834
9113679386	2708943879	9362016295	1541337142	4892830722
1965362132	3926406160	1363581559	0742202020	3187277605
3610642506	3904975008	6562710953	5919465897	5141310348
0861533150	4452127473	9245449454	2368288606	1340841486
5189757664	5216413767	9690314950	1910857598	4423919862
3338945257	4239950829	6591228508	5558215725	0310712570
8475651699	9811614101	0029960783	8690929160	3028840026
7180653556	7252532567	5328612910	4248776182	5829765157
8788202734	2092222453	3985626476	6914905562	8425039127
6407655904	2909945681	5065265305	3718294127	0336931378
7367812301	4587687126	6034891390	9562009939	3610310291
7634757481	1935670911	0137751721	0080315590	2485309066
4366199104	0337511173	5471918550	4644902636	5512816228
1509682887	4782656995	9957449066	1758344137	5223970968
5848358845	3142775687	9002909517	0283529716	3445621296
9748442360	8007193045	7618932349	2292796501	9875187212
3025494780	2490114195	2123828153	0911407907	3860251522
2673430282	4418604142	6363954800	0448002670	4952482017
2609275249	6035799646	9256504936	8183609003	2380929345
4525564056	4482465151	8754711962	1844396582	5337543885
9695946995	5657612186	5619673378	6236256125	2163208628
9279068212	0738837781	4233562823	6089632080	6822246801
0037279839	4004152970	0287830766	7094447456	0134556417
2314593571	9849225284	7160504922	1242470141	2147805734
2339086639	3833952942	5786905076	4310063835	1983438934
7360411237	3599843452	2516105070	2705623526	6012764848
7065874882	2596815793	6789766974	2205700596	8344086973
2162484391	4035998953	5394590944	0704691209	1409387001
4610126483	6999892256	9596881592	0560010165	5256375678

Литература по теории вероятностей и математической статистике, выпущенная издательством «Мир» [1])

1970

Хеннан Э. Представления групп и прикладная теория вероятностей.
Холл М. Комбинаторика.

1971

Карлин С. Основы теории случайных процессов.
Рюэль Д. Статистическая механика.
Такач Л. Комбинаторные методы в теории случайных процессов.

1972

Маккин Г. Стохастические интегралы.
Соле Ж.-Л. Основы структуры математической статистики.

1973

Джейсуол Н. Очереди с приоритетами.
Кахан Ж.-П. Случайные функциональные ряды.
Мейер П.-А. Вероятность и потенциал.
Сеа Ж. Оптимизация. Теория и алгоритмы.

1974

Барра Ж. Р. Основные понятия математической статистики.
Де Гроот М. Оптимальные статистические решения.
Хеннан Э. Многомерные временные ряды.

1975

Делашери К. Емкости и случайные процессы.
Закс Ш. Теория статистических выводов.
Крамер Г. Математические методы статистики. 2-е издание.
Русас Дж. Континуальность вероятностных мер.

1976

Андерсон Т. Статистический анализ временных рядов.
Дуда Р., Харт П. Распознавание образов и анализ сцен.
Эрдёш П., Спенсер Дж. Вероятностные методы в комбинаторике.

1977

Математика в социологии: моделирование и обработка информации. Под ред. А. Аганбегяна и др.
Фу К. Структурные методы в распознавании образов.
Харари Ф., Палмер Э. Перечисление графов.

1978

Гиббсовские состояния в статистической физике. Сб. статей.
Евклидова квантовая теория поля. Марковский подход. Сб. статей.
Кокс Д., Хинкли Д. Теоретическая статистика.
Матерон Ж. Случайные множества и интегральная геометрия.
Орнстей Д. Эргодическая теория, случайность и динамические системы.
Случайные процессы. Выборочные функции и пересечения. Сб. статей.
Ту Дж., Гонсалес Р. Принципы распознавания образов.

1979

Гренандер У. Лекции по теории образов. В 2-х томах. Т. 1. Синтез образов.

1980

Бриллинджер Д. Временные ряды. Обработка данных и теория.
Классификация и кластер. Под ред. Дж. ван Рицина.

1981

Гренандер У. Лекции по теории образов. В 2-х томах. Т. 2. Анализ образов.

[1]) Настоящий список литературы приводится для удобства читателей и содержит книги, выпущенные издательством «Мир» в 1970—1979 гг. В этих книгах можно найти дополнительную информацию по вопросам, изложенным в данной книге. — *Прим. ред*

Кокс Д., Хинкли Д. Задачи по теоретической статистике с решениями.
Хейер Х. Вероятностные меры на локально-компактных группах.

1982

Афифи А., Эйзен С. Статистический анализ. Подход с использованием ЭВМ.
Браттели У., Робинсон Д. Операторные алгебры и квантовая статистическая механика.
Куммер Б. Игры на графах.
Отнес Р., Эноксон Л. Прикладной анализ временных рядов.

1983

Гренандер У. Лекции по теории образов. В 2-х томах. Т. 3. Регулярные структуры.
Маршалл А., Олкин Н. Неравенства. Теория мажоризации и ее приложения.
Партасарати К. Введение в теорию вероятностей и теорию меры.
Сато М., Дзимбо М., Мива Т. Голономные квантовые поля.

1984

Грехем Р. Начала теории Рамсея.
Кокс Д., Снелл Э. Прикладная статистика. Принципы и примеры.
Прабху Н. Стохастические процессы теории запасов.
Феллер В. Введение в теорию вероятностей и ее приложения. В двух томах.

1985

Мулен Э. Теория игр.
Хьюбер Дж. Робастность в статистике.
Чисар П., Кёрнер Я. Теория информации: теоремы кодирования для дисперсионных систем.

1986

Кестен Х. Теория просачивания для математиков.
Питмен Э. Основы теории статистических выводов.
Эллиотт Р. Стохастический анализ и его приложения.

1987

Беллман Р. Математические методы в медицине.
Грин Д., Кнут Д. Математические методы анализа алгоритмов.
Коэн Дж., Боксма О. Граничные задачи в теории массового обслуживания.
Чжун К., Уильямс Р. Введение в стохастическое интегрирование.

1988

Балакришнан А. В. Теория фильтрации Калмана.
Деврой Л., Дьёрфи Л. Непараметрическое оценивание плотности. L_1-подход.
Квантовые случайные процессы и открытые системы. Сб. статей.
Линский В. Комбинаторика для программистов.
Теннант-Смит Дж. Бейсик для статистиков.

1989

Лиггетт Т. Марковские процессы с локальным взаимодействием.
Лидбеттер М., Линдгрен Г., Ротсен Х. Экстремумы случайных последовательностей и процессов.
Нуммелин Э. Общие неприводимые цепи Маркова и неотрицательные операторы.
Хампель Ф., Рончетти Э., Рауссеу П. и др. Робастная статистика. Подход на основе функций влияния.

1990 (план)

Альбеверио С., Фенстад Й., Хеэг-Крон Р., Линдстрем Т. Нестандартные методы в стохастическом анализе и математической физике.
Секей Г. Парадоксы в теории вероятностей и математической статистике.

ИМЕННОЙ УКАЗАТЕЛЬ

(Римские цифры указывают номер главы, арабские — номер параграфа)

Адамар (Hadamard J.) IV/1
Адлеман (Adleman) IV/13ж
Ампер (Ampère A. M.) III/3
Андерсон (Anderson C. L.) I/3
Арбутнот (Arbuthnot J.) II/10
Арроу (Arrow K.) III/7з
Афиней (Athenaios) II/Введ.

Байес (Bayes Th.) II/1
Балкема (Balkama A. A.) Введ., IV/9
Балоте (Balote N.) I/13п
Банах (Banach S.) IV/2
Бартфай (Bártfai P.) IV/13д
Басу (Basu D.) II/8
Башелье (Bachelier L.) III/3, 6
Баккенбах (Beckenbach E. F.) IV/4
Беляев Ю. К. I/8
Бенфорд (Benford F.) IV/7
Березовский Б. А. III/7е
Беркхолдер (Burkholder D. L.) II/7
Бернулли Д. (Bernoulli D.) I/7, II/10
Бернулли Н. (Bernoulli N.) I/7
Бернулли Я. (Bernoulli J.) I/7, 9, 10
Бернштейн С. Н. I/4, II/1
Берри (Berry G. G.) IV/4
Бертки (Bertky W.) II/12
Бертран (Bertrand J. L.) I/11
Бесикович (Besicovitch A. S.) III/3
Бессель (Bessel F. W.) II/3
Бландон (Blundon W. J.) IV/1
Бляшке (Blaschke W.) I/11
Бозе (Bose S.) I/1
Бойаи (Bolyai) V
Больцано (Bolzano B.) III/3, V
Больцман (Boltzman L. E.) I/1, III/2
Бондессон (Bondesson L.) IV/9
Борель (Borel E.) I/5, 11, 12, 13л, III/7б, IV/Введ., 4
Борн (Born M.) II/1, IV/13е
Борткевич (Bortkiewicz L.) 1/6
Брауэр (Brouwer L. E. J.) III/3
Бремс (Brams S. J.) I/12
Бригс (Briggs H.) III/6
Бриллюэн (Brioullin L.) II/11
Бройль (Broglie L. V. de) IV/13е
Броун (Brown R.) III/3
Буцолич (Buczolich Z.) II/4
Бьенеме (Bienayme J.) I/9, II/4, III/1
Бюффон (Buffon G. L. L.) I/7, 11, IV/3

Вальд (Wald A.) I/12, II/9, 12
Ватсон (Watson H. W.) III/1
Вебер (Weber E.) II/6, IV/7
Вейерштрасс (Weierstrass K.) III/3, IV/8
Вейс (Weiss G. H.) III/5
Вийон (Villon F.) V
Вилле (Ville J.) III/6
Вильямс (Willams J. D.) I/12
Винер (Winer N.) II/11, III/3
Витали (Vitali G.) IV/Введ.
Волмэн (Wallmann H.) III/3
Волфовиц (Wolfowitz J.) II/8, 12, 12з
Вольтер (Voltaire F. M. A.) III/7з

Гаддер (Gudder S. P.) IV/13е
Гаджиев А. Г. III/4
Галилей (Galilei G.) I/1, II/4
Галлей (Halley E.) I/8
Гальтон (Galton F.) II/5, 6, 13а, III/1
Гаусс (Gauss C. F.) II/4
Гейзенберг (Heisenberg W.) IV/13е
Гиббс (Gibbs J. W.) III/7з
Гильберт (Hilbert D.) IV/Введ.
Гини (Gini C.) II/6
Гири (Geary E. C.) IV/10
Гихман И. И. III/3
Гнеденко Б. В. I/1, 3, 8, 10, II/10
Гнедин А. В. III/7е
Голдберг (Goldberg S.) I/8
Гольдбах (Goldbach C.) IV/4
Граунт (Graunt J.) I/8, II/Введ.
Грешем (Gresham Th.) III/6
Гроссвальд (Grosswald E.) IV/9
Грофман (Grofman B.) III/7з
Гуд (Good I. J.) I/13д, II/13г, IV/1
Гуревич (Hurewitz W.) III/3
Гутенберг (Gutenberg J.) I/5
Гутман (Gutman S.) I/7
Гюйгенс (Huygens C.) I/3, II/Введ.

Дадли (Dudley R. M.) I/7
Даламбер (D'Alambert J. R.) L/1,8
Данциг (Dantzig G.) II/12
Дарвин (Darwin Ch.) II/5, V
Дармуа (Darmois G.) IV/10
Де Гроот (DeGroot M. H.) IV/13в
Джеймс (James W.) II/2,3
Джейнес (Jaynes E. T.) II/9
Джонс (Jones M. F.) IV/1
Диаконис (Diaconis P.) II/Введ.
Дирак (Dirac P. M. A.) I/1
Дирихле (Dirichlet P. G. L.) IV/1
Диффи (Diffie W.) IV/13ж
Дмитриев Н. А. III/1
Добрушин Р. Л. III/7з
Додж (Dodge H. F.) II/12
Дрейпер (Draper N. R.) II/6
Дуб (Doob D. L.) III/6
Дынкин Е. Б. III/2
Дюбуа-Реймон (Bois-Reimond P. du) III/3
Дюге (Dugué D.) IV/11
Дьедонне (Dieudonne J. A.) V

Закс (Zacks S.) II/2,3
Замрифеску (Zamrifecu T.) IV/13б
Земплени (Zempléni A.) IV/9, 12
Зенон (Zenon) IV/8
Зингер А. А. IV/10
Золотарев В. М. IV/9

Иаков (Jacob) III/7а

Каган А. М. II/2, IV/10
Калман (Kalman R. E.) II/6
Каминский (Kaminsky K. S.) I/12
Кангсмаа-Минн (Kangsmaa-Minn E.) I/8
Кантелли (Cantelli F. P.) III/7б, IV/1
Кантор (Cantor G.) IV/8, V
Караваджо (Caravaggio) I/5
Кардано (Cardano G.) I/1,2
Карл VI (Charles VI) I/5
Карно (Carnot L. S.) III/2
Кац (Kac M.) II/13з, IV/1,10
Каэдиг (Kaedig F. W.) III/2
Кейнс (Keynes J. M.) II/6
Кемперман (Kempermann J. H. B.) IV/7
Кендалл (Kendall M. G.) I/11, II/5
Кестен (Kesten H.) IV/13г
Кетле (Quételet L. A. J.) II/13а
Кингстон (Kingston J. G.) I/3
Киршенманн (Kirschenmann P.) IV/3
Кифер (Kiefer J.) II/8, 13з
Кларк (Clarke L. E.) IV/13б
Клаузиус (Clausius R. E.) III/2
Клейн К. (Klein K.) IV/Введ.
Клейн Р. (Klein R. L.) II/6
Клейнрок (Kleinrock L.) III/4
Кнут (Knuth D. E.) IV/3
Колмогоров А. Н. Введ., I/9, 10, II/1, 10, 13з, III/1, IV/Введ., 1, 3, 4, 9, 13г, е
Комлош (Komlós J.) I/9, IV/13д
Кондолье (Condolle de) III/1
Кондорсе (Condorcet M.) III/7з
Конфуций (Confucius) II/2
Коши (Cauchy A. L.) II/2, 4, IV/11, V
Крамер (Cramér H.) I/7, II/2
Кронекер (Kronecker L.) V
Кульбак (Kullback S.) II/11
Кэррол (Carroll L.) I/13п

Лаван (Laban) III/7а
Лавуазье (Lavoisier A. L.) II/1
Лакс (Luks E. M.) I/12
Лал (Lal M.) IV/1
Ламберт (Lambert J. H.) II/4
Лаплас (Laplace P. S.) Введ., II/4, 10
Лебег (Lebesgue H.) IV/13б
Леви (Lévy P.) III/6, 7в, IV/1, 9, 11
Лежандр (Legendre A. M.) II/4
Лейбниц (Leibniz G. W.) I/1, 2, IV/8
Ле Кам (Le Cam L.) II/8
Леман (Lehmann E. L.) II/10, 13д, е
Леонардо да Винчи (Leonardo da Vinci) I/3
Ли (Li S.-Y. R.) I/13в
Либби (Libby W. F.) I/8
Линдеман (Lindemann F.) IV/4
Линдли (Lindley D. V.) I/13д, II/1
Линник Ю. В. II/2, 4, IV/10, 12
Литлвуд (Littlewood J. E.) IV/1
Лиувилль (Lioville J.) IV/4
Ломницкий (Lomnitzky A.) IV/Введ.
Лотка (Lotka A. J.) III/1
Лошмидт (Loschmidt J.) III/2
Луи IX (Louis IX) I/5
Лукач (Lukacs E.) IV/9, 10, 11
Лукреций (Lucretius) III/Введ.
Ляпунов А. М. I/10, IV/11

Майор (Major P). IV/13д
Майстров Л. Е. I/1
Макки (Mackey G.) IV/13е
Маккин (McKean H. P. Jr.) III/3
Максвелл (Maxwell J. C.) I/Введ., III/2
Макшейн (McShane P.) IV/3

Мандельбройт (Mandelbroit B. B.) III/3
Манн (Mann Th.) III/7а
Марков А. А. I/10, III/2, 7з
Мартин-Лёф (Martin-Löf) IV/3, 4
Марцинкевич (Marcinkiewicz J.) IV/10
Масани (Masani S. M.) II/2
Маслова Н. Б. IV/5
Матерон (Matheron G.) I/11
Мафусаил (Methuselah) I/8
Мах (Mach E.) III/2
Махалонобис (Mahalonobis P. C.) II/12
Мейер (Meyer P. A.) III/6
Менгер (Menger K.) III/3
Мендез (Mendez C. G.) I/9
Мендель (Mendel G. J.) II/5
Мере де (de Méré) I/2
Метрополис (Metropolis N.) IV/3
Мизес (Mises R. von) II/1, 10, IV/3
Мил (Miel G.) IV/3
Моисей (Moses) II/Введ.
Мойры (Moira) IV/2
Монмор (Monmort P. R. de) I/6, 7
Моран (Moran P. A. P.) I/11
Моргенштерн (Morgenstern O.) I/12
Мори (Móri T. F.) I/9, 13ж, II/2, 5, III/6, IV/6, IV/9, 13а
Мостеллер (Mosteller F.) I/13н
Муавр (Moivre A. de) I/2, 10, II/1

Наполеон (Napoleon B.) II/4
Неве (Neveu J.) III/6
Негели (Nägeli C. W.) III/3
Нейман Джон фон (Neumann J.) I/5, 12, II/9, IV/13е
Нейман Ежи (Neyman J.) II/8, 9, 10, IV/3
Нелсон (Nelson P. I.) I/12
Неш-Вильямс (Nash-Williams C.) III/5
Ньюкомб (Newcomb S.) IV/6
Ньютон (Newton I.) 1/2, III/2, IV/8

Окстоби (Oxtoby L. J. C.) III/3
Олкин (Olkin I.) II/5
Онсагер (Onsager L.) III/7з
Оре (Ore Ø.) I/1, 2, 3
Оствальд (Ostwald W. F.) III/2
Островский Е. И. I/10
Островский И. В. IV/12

Павсаний (Pausanias) I/1
Паламедей (Palamedeo) I/1
Паскаль (Pascal B.) I/2, 3
Пачоли (Paccioli F. L.) I/3
Пеано (Peano G.) I/4
Пепис (Pepys S.) I/2
Петров В. В. I/10
Петти (Petty W.) II/Введ.
Пинкхэм (Pinkham R. S.) IV/7
Пинскер (Pinsker M. S.) III/7ж
Пирс (Peirce Ch. S.) II/9
Пирсон К. (Pearson K.) I/4, II/5, 9, 10, 13а
Пирсон Э. (Pearson E. S.) II/1, 9, 10
Питмен (Pitman E. J. G.) II/8
Плакетт (Plackett R. L.) II/6
Платон (Platon) V
По (Poe E. A.) IV/13ж
Пойа (Polya G.) Введ., I/10, 13б, III/5, IV/10, 11
Полигнот (Polygnotos) I/1
Пратт (Pratt J. W.) I/13д, II/5
Пуанкаре (Poincaré J. H.) I/10, III/2, 3, IV/Введ., 1, V
Пуассон (Poisson S. D.) I/6, 9, IV/9
Пушкин А. С. III/2

Райков Д. А. IV/9, 11
Райми (Raimi R. A.) IV/7
Рао (Rao C. R.) II/2, 6, 8, IV/10
Рассел (Russel B.) IV/Введ., 4
Раун (Raun A.) I/8
Ревес (Révész P.) I/9, IV/13д
Рейнольдс (Reynolds J.) II/13а
Реньи (Rényi A.) Введ., 1/2, 9, II/11, III/7а, IV/1, 5, 11, 13д, з
Ресник (Resnick S. I.) IV/9
Ривест (Rivest) IV/13ж
Риман (Riemann G. F. B.) III/3, IV/1
Роббинс (Robbins H.) II/10, 12, 13м, о, III/7д, е
Робертс (Roberts F.) I/5
Робинсон (Robinson A.) IV/8
Ромиг (Romig H. G.) II/12
Рубин (Rubin H.) II/13л
Ружа (Ruzsa I. Z.) IV/1, 6, 12

Сантало (Santaló L. A.) I/11
Сартин (Sartine G. de) I/5
Свадеш (Swadesh M.) I/8
Секей (Székely G. J.) I/9, 10, II/2, 5, III/3, 6, IV/1, 3, 6, 9, 12, 13а
Сетураман (Sethuraman J.) II/13k
Сигмунд (Siegmund D.) II/12, III/7д, е
Силард (Szilard L.) III/2

Эджворт (Edgeworth F. Y.) II/13а, д
Эйлер (Euler L.) II/4
Эйнштейн (Einstein A. B.) Введ., III/3, IV/13е, V
Эйрес (Ayres L. P.) II/Введ.
Эпштейн (Epstein B.) IV/9
Эрдеш (Erdös P.) I/9, IV/5
Эренфест П. (Ehrenfest P.) III/2
Эренфест Т. (Ehrenfest T.) III/2
Эрланг (Erlang A. K.) I/6, III/4
Эрмит (Hermite Ch.) III/3, IV/4
Эсхил (Aeschylus) I/2

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

(Римские цифры указывают номер главы, арабские — номер параграфа)

Абсурдные результаты (absurdities) I/13п
Акция (share) III/6
Алгоритм *Конвея* (Convay algorithm) I/13в

Безвозрастность (agelesness) I/8
Биржа фондовая (Stock Exchange) III/6
Бит (bit) II/11
Близнецы, простые числа (twin primes) IV/1
Бридж (bridge) I/5
Броуновское движение (Brownian motion) III/3

Вероятность апостериорная (a posteriori probability) II/1
— априорная (a priori probability) II/1
— в квантовой физике (quantum physical probability) IV/13е
— вырождения (probability of extinction) III/1
— геометрическая (geometrical probability) I/11
— объективная (objective probability) II/1
— отрицательная (negative probability) IV/Введ.
— перехода (transition probability) III/2
— «поражения по кругу» («round defeat» probability) I/13е
— субъективная (subjective probability) II/1
— условная (conditional probability) I/4, II/1 и др.
Время ожидания (waiting time) III/4
— остановки оптимальные (optimal stopping) III/7д
— разорения (ruin time) III/7г
Выборка (sample) I/12, II/2
— упорядоченная (ordered sample) II/2, 4 и др.
Вырождение фамилий (extinction of family names) III/1

Гипотеза альтернативная (alternative hypothesis) II/10
— нулевая (null hypothesis) II/10
— статистическая (statistical hypothesis) II/10
Глоттохронология (glottochronology) I/8
Голосование (voting) III/7з
Граф (graph) IV/5
— связный (connected graph) IV/5
— случайный (random) graph) IV/5
— универсальный (universal graph) IV/5
Графы изоморфные (isomorthic graphs) IV/5

Делитель вероятностного распределения (factor of a probability distribution) IV/11
Демон *Максвелла* (Maxwell demon) III/2
Дисперсия (variance) I/4, II/3 и др.
Дифференциальное уравнение теплопроводности (differential equation of heat conduction) IV/Введ.
Достаточная статистика (sufficient statistic) II/7
Дрейдель (dreydel или draydl) I/2

Задача *Бюффона* об игле (Buffon's needle problem) IV/3
Закон арксинуса (arcsine law) III/7в
— *Бенфорда* (Benford's law) IV/7
— больших чисел (law of large numbers) I/6, 9, 10
— — — усиленный (strong law of large numbers) IV/13г, д
— *Вебера—Фехнера* (Weber—Fechner's law) IV/7
— *Гаусса* (Gaussian law) II/4
— малых чисел (law of small numbers) I/6

— повторного логарифма (law of the iterated logarithm) IV/13д
— термодинамики первый и второй (first and second laws of thermodynamics) III/2
— эпонимии (law of Eponymy) IV/7

Игра безобидная (fair game) I/3, 7 и др.
— в кости (dicing) I/1, 2 и др.
— в «очко» (game twenty one) I/5
Игры карточные (card games) I/5
Инвариантность масштаба (scale invariance) I/11
— относительно сдвигов (translational invariance) I/11
Интегральная геометрия (integral geometry) I/11
Интервал доверительный (confidence interval) II/9, 13о
— фидуциальный (fiducial interval) II/9
Информация *Фишера* (Fisher information) II/2
Испытания *Бернулли* (Bernoulli trials) I/6

Квазиметод Монте-Карло (quasi-Monte-Carlo-method) IV/3
Количество информации (amount of information) II/11
Комета (comet) II/10
Компьютерная статистика (computer statistics) II/13р
Компьютерное искусство (computer art) IV/3
Корреляция (correlation) II/5
— максимальная (maximal correlation) II/5
— частная (partial correlation) II/5
Коэффициент корреляции (correlation coefficient) II/5
— — выборочный (sample correlation coefficient) II/5
Криптография (cryptograthy) IV/13ж
Критерий для проверки гипотез (test of hypotheses) II/10
— *Колмогорова* (Kolmogorov test) II/13з
— наиболее мощный (most powerful test) II/10

Лексикостатистика (lexicostatistics) I/8
Лемма *Бореля—Кантелли* (Borel—Cantelli lemma) III/7б
— *Неймана—Пирсона* (Neyman—Pearson's lemma) II/10
Линия регрессии (regression line) II/6
Логит-анализ (logit-analysis) II/6
Лотерея (lottery) I/5, 7

Марковское поле (Markov field) III/7з
Мартингал (martingale) III/6
Математическое ожидание (expectation или expected value) I/2, 4, II/13а и др.
— — условное (conditional expectation) I/7
Матрица переходных вероятностей (transition matrix) III/2
Машина *Тьюринга* (Turing machine) IV/4
Медиана (median) I/8, II/4, 13а
Мера аддитивная (additive measure) IV/1, 2
— сигма-аддитивная (sigma-additive measure) IV/1, 2
Меридиан (meridian) I/13л
Метод максимального правдоподобия (maximum-likelihood method) II/8
— Монте-Карло (Monte-Carlo method) IV/3
— наименьших квадратов (least squares method) II/4
— радиоуглеродный (radiocarbon method) I/8
Множитель *Бесселя* (Bessel factor) II/3
Модель байесовская (Bayes model) II/13н
— *Бозе—Эйнштейна* (Bose—Einstein model) I/1
— *Гальтона—Ватсона* (Galton—Watson model) III/1
— *Изинга* (Ising model) III/7з
— *Максвелла—Больцмана* (Maxwell—Boltzman model) I/1
— *Ферми—Дирака* (Fermi—Dirac model) I/1
Монада (monad) IV/8

Независимые события и случайные величины (independent events and random variables) I/4
Неравенство *Чебышева—Бьенеме* (Cebyshev—Biename inequality) I/9
— *Крамера—Рао* (Cramer—Rao inequality) II/2

Оценка асимптотически несмещенная (asymptotic unbiased estimator) II/3
— *Джеймса—Стейна* (James—Stein estimator) II/2, 3
— дисперсии (estimation of variance) I/3
— допустимая (admissible estimator) I/12, II/1, 2 и др.
— коэффициента корреляции (estimation of correlation) II/2
— максимального правдоподобия (maximum likelihood estimator) II/4, 8 и др.
— математического ожидания (estimation of expectation) II/2
— минимаксная (minimax estimator) I/12, II/1, 2
— неизвестного параметра (estimation of unknown parameter) I/12, II/2 и др.
— несмещенная (unbiased estimator) II/2 и др.
— по методу наименьших квадратов (least squares estimator) II/4, 6
— с минимальной дисперсией (minimum variance estimator) II/2
— состоятельная (consistent estimator) II/2
— суперэффективная (superefficient estimator) I/8
— эффективная (efficient estimator) II/2
Ошибки I-го и 2-го рода (type 1 and 2 errors) II/10

Парадокс *Берри* (Berry's paradox) IV/4
Период полураспада (half-life) I/2, 8
Плотность (density function) I/4
Последовательный анализ (sequential analysis) II/12
Правило *Вебера—Фехнера* (Weber—Fechner rule) II/6
— пропорциональности критических значений (proportionality rule of critical values) I/2
Принцип минимакса (minimax principle) I/12, II/3
— неопределенности (uncertainty principle) II/10
Пробит-анализ (probit-analysis) II/6
Проблема выбора (choice problem) III/7е
— *Гольдбаха* (Goldbach-confecture) IV/4
— измеримых мощностей (problem of measurable cardinalities) IV/Введ.
Проверка гипотез (hypothesis testing) II/10
— независимости (independence testing) II/13п
Продолжительность жизни (life span) I/8
— — вероятная (probable life span) I/8
Процесс ветвящийся (branching process) III/1
— винеровский (Wiener process) III/3
— с независимыми приращениями (process with independent (increments) III/7б
— стационарный (stationary process) III/7ж
— — регулярный (regular stationary process) III/7ж
— — сингулярный (singular stationary process) III/7ж

Равновероятные (равновозможные) исходы (equally probable outcomes) I/1, 3, 6 и др.
Раздел ставки (division of the prize) I/3
Размерность топологическая (topological dimension) III/3
— *Фурье* (Fourier dimension) III/3
— *Хаусдорфа—Бесиковича* (Hausdorf—Besicovitch dimension) III/3
Распределение апостериорное (a posteriori distribution) II/1, 9
— априорное (a priory distribution) II/1, 9
— безгранично делимое (infinitely divisible distribution) IV/9
— бета (beta distribution) II/13в
— биномиальное (binomial distribution) I/6, II/2
— *Коши* (Cauchy distribution) II/2
— логнормальное (lognormal distribution) IV/9
— неразложимое (irreductible distribution) IV/12
— нормальное (normal distribution) I/10, II/2 и др.
— — многомерное (multivariate normal distribution) II/2, 5 и др.
— — стандартное (standard normal distribution) I/10
— показательное (exponentional distribution) I/8, II/9 и др.

— простое (prime distribution) IV/12
— *Пуассона* (Poisson distribution) I/6, 8
— равномерное (uniform distribution) I/11, II/1, 5, IV/1 и др.
— хи-квадрат (chi-square distribution) II/12
— *Эрланга* (Erlange distribution) I/6
Распределения взаимно простые (relatively prime distributions) IV/12
— функция (distribution function) I/4
— — сингулярная (singular distribution function) IV/13б
— — совместная (joint distribution function) I/4
Регрессия (regression) II/5, 6
— линейная (linear regression) II/13и
Рулетка (roulette) I/5

Свертка (convolution) IV/9
— мультипликативная (multiplicative convolution) IV/9
Седловая точка игры (saddle point of the game) I/12
Серии из гербов или решек (run of heads or tails) I/9, 13в
Сигма-алгебра (sigma-algebra) IV/ /Введ.
Случайная величина (random variable) I/2
Случайное алгебраическое уравнение (random algebraic equation) IV/5
— блуждание (random walk) III/5
— поле (random field) III/7з
Случайность и сложность (randomness and complexity) IV/3
Среднее значение случайной величины (mean of the random variable) *см.* Математическое ожидание
Стандартное отклонение (standard deviation) I/4, II/2 и др.
Статистическое общество в Лондоне (London Statistical Society) II/13а
Стохастический гейзер (stochastic geyser) IV/13д
— интеграл (stochastic integral) III/3
Стохастическое дифференциальное уравнение (stochastic differential equation) III/3
Стратегия (strategy) I/12
— оптимальная (optimal strategy) I/12
— смешанная (mixed strategy) I/12
— чистая (pure strategy) I/12
Сходимость по вероятности (convergence in probability) I/9
— с вероятностью 1 (convergence with probability 1) I/9

Таблицы смертности (mortality tables) I/8
Теорема *Байеса* (Bayes' theorem) II/1, 9
— *Банаха—Тарского* (Banach—Tarski theorem) IV/2
— *Больцмана* (Boltzmann theorem) III/2
— *Вольда* (Wold's theorem) III/7ж
— *Дармуа—Скитовича* (Darmois—Skitovic theorem) IV/10
— *Дирихле* (Dirichlet's theorem) IV/1
— *Колмогорова* (Kolmogorov's theorem) I/9
— *Крамера* (Cramer's theorem) IV/11
— *Леви—Хинчина* (Levy—Hincin's theorem) IV/9
— *Марцинкевича* (Marcinkiewicz theorem) IV/10
— *Неймана* о минимаксе (Neumann's minimax theorem) I/12
— *Пойа* (Polya's theorem) III/5
— предельная *Муавра—Лапласа* de Moivre—Laplace limit theorem) I/10, 13и
— — центральная (central limit theorem) I/10
— *Ферма* (Fermat—conjecture) IV/4
— *Чжуна—Фукса* (Chung—Fuchs theorem) II/13н, III/7б
— *Эйнштейна* (Einstein's theorem) III/3
Теория *Колмогорова* (Kolmogorov's theory) IV/Введ.
— *Робинсона* (Robinson's theory) IV/8

Фазовое пространство (phase space) IV/Введ.
Факторизация (factorization) IV/11
Формула *Стирлинга* (Stirling's formula) I/10
— *Шеннона* (Shannon's formula) II/11
Фракталы (fractals) III/3
Функция гипергеометрическая (hypergeometric function) II/5
— нигде не дифференцируемая (nowhere differentiable function) III/3
— потерь (loss function) I/12
— — квадратичная (quadratic loss function) I/12

— производящая (generating function) III/1, 2
— риска (risk function) I/12
— характеристическая (characteristic function) IV/11

Цепь *Маркова* (Markov chain) III/2
— — однородная (homogeneous Markov chain) III/2

Число бесконечно малое (infinitely small number) IV/8
— интересное (interesting number) IV/4
— нормальное (normal number) IV/4
— простое (prime) IV/1
— псевдослучайное (pseudo-random number) IV/3
— случайное (random number) IV/3
— трансцендентное (transcendental number) IV/4
— *Шайтина* (Chaitin number) IV/4

Шифр одноразовый (one-time cipher) IV/13ж
— сдвиг (shift cipher) IV/13ж

Энтропия (entropy) II/11, III/2, IV/10

t-критерий *Стьюдента* (Student's *t*-test) II/12, 13е

ОГЛАВЛЕНИЕ

Предисловие редактора перевода и переводчика 5
Предисловие к русскому изданию 7
Предисловие редактора серии 8
Введение 10

ГЛАВА I. КЛАССИЧЕСКИЕ ПАРАДОКСЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ . . . 12

1. Парадокс игры в кости. «Азартные игры» в мире физических частиц 12
2. Парадокс де Мере 15
3. Парадокс раздела ставки 19
4. Парадокс независимости 22
5. Парадоксы бриджа и лотереи 25
6. Парадокс раздачи подарков; травмы, причиненные лошадьми; телефонные вызовы; опечатки 30
7. Санкт-петербургский парадокс 35
8. Парадокс смертности населения. Безвозрастный мир атомов и слов 38
9. Парадокс закона больших чисел Бернулли 41
10. Парадокс де Муавра; экономия энергии 45
11. Парадокс Бертрана 50
12. Парадокс из теории игр. Парадокс гладиатора 54
13. Еще несколько парадоксов 60
а) Парадокс событий, происходящих почти наверно 60
б) Парадокс вероятности и относительной частоты 60
в) Парадоксы, связанные с бросанием монеты 60
г) Парадокс условной вероятности 63
д) Парадокс случайных времен ожидания 63
е) Парадокс транзитивности 64
ж) Парадокс измерения регулярности игральной кости 66
з) Парадокс дня рождения 67
и) Парадокс гербов и решек 67
к) Ребро монеты 68
л) Парадокс Бореля 68
м) Парадокс условных распределений 69
н) Как играть в проигрышную игру 70
о) Парадокс страхования 70
п) Абсурдные результаты, Льюис Кэррол 71

ГЛАВА II. ПАРАДОКСЫ В МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ 74

1. Парадокс Байеса 76
2. Парадокс оценок математического ожидания 80
3. Парадокс оценок дисперсии 87
4. Парадокс метода наименьших квадратов 89
5. Парадоксы корреляции 92
6. Парадоксы регрессии 97
7. Парадоксы достаточности 102
8. Парадоксы метода максимального правдоподобия 104

9. Парадокс интервальных оценок . . . 108
10. Парадокс проверки гипотез . . . 113
11. Парадокс Реньи из теории информации . . . 116
12. Парадокс t-критерия Стьюдента . . . 120
13. Еще несколько парадоксов . . . 123
а) Парадокс типичного и среднего . . . 123
б) Парадокс оценивания . . . 123
в) Парадокс точности измерения . . . 124
г) Парадоксальное оценивание вероятности . . . 124
д) Чем больше данных, тем хуже выводы . . . 125
е) Парадокс равенства математических ожиданий . . . 126
ж) Парадоксальная оценка для математического ожидания нормального распределения . . . 127
з) Парадокс проверки нормальности . . . 127
и) Парадокс линейной регрессии . . . 128
к) Парадокс Сетурамана . . . 129
л) Парадокс минимаксной оценки . . . 129
м) Парадокс Роббинса . . . 130
н) Парадокс байесовской модели . . . 130
о) Парадокс доверительных интервалов . . . 131
п) Парадокс проверки независимости; являются ли эффективные лекарства эффективными? . . . 132
р) Парадокс компьютерной статистики . . . 133

ГЛАВА III. ПАРАДОКСЫ СЛУЧАЙНЫХ ПРОЦЕССОВ . . . 134

1. Парадокс ветвящихся процессов . . . 134
2. Марковские цепи и физический парадокс . . . 137
3. Парадокс броуновского движения . . . 141
4. Парадокс времени ожидания (Ходят ли автобусы чаще в обратном направлении?) . . . 146
5. Парадокс случайных блужданий . . . 150
6. Биржевый парадокс; мартингалы . . . 152
7. Еще несколько парадоксов . . . 156
а) Парадокс Иакова и Лавана . . . 156
б) Парадокс процессов с независимыми приращениями . . . 157
в) Парадокс забитых голов . . . 158
г) Парадокс ожидаемого времени разорения . . . 159
д) Парадокс оптимальных правил остановки . . . 161
е) Парадокс выбора . . . 162
ж) Парадокс Пинскера о стационарных процессах . . . 163
з) Парадоксы голосования и выборов; случайные поля . . . 164

ГЛАВА IV. ПАРАДОКСЫ В ОСНОВАНИЯХ ТЕОРИИ ВЕРОЯТНОСТЕЙ. РАЗНЫЕ ПАРАДОКСЫ . . . 168

1. Парадоксы случайных натуральных чисел . . . 170
2. Парадокс Банаха — Тарского . . . 174
3. Парадокс метода Монте-Карло . . . 176
4. Парадокс неинтересных чисел; невычислимая вероятность . . . 180
5. Парадокс случайных графов . . . 184
6. Парадокс математического ожидания . . . 186
7. Парадокс первой цифры . . . 187
8. Парадокс нулевой вероятности (Можно ли из ничего получить что-то?) . . . 190
9. Парадокс безгранично делимых распределений . . . 192
10. Парадоксы характеризации . . . 196

11. Парадоксы факторизации 199
12. Парадокс неразложимых и простых распределений 202
13. Еще несколько парадоксов 204
а) Парадокс деления распределений пополам 204
б) Патологические вероятностные распределения 206
в) Парадокс продавца газет 207
г) Парадокс Кестена 208
д) Парадокс стохастического гейзера 208
е) Парадокс вероятности в квантовой физике 209
ж) Парадокс криптографии 211
з) Парадокс поэзии и теории информации 213

ГЛАВА V. ПАРАДОКСОЛОГИЯ 214

Обозначения 216
Таблицы 217
Литература по теории вероятностей 226
Именной указатель 228
Предметный указатель 233