ДВАДЦАТЬ ПЯТАЯ

ХРЕСТОМАТИЯ

По истории теории вероятностей и статистики
составитель О. Б. Шейнин

Берлин
2019

Содержание

oscar.sheynin@gmail.com

От составителя

Ниже мы приводим общие соображения об отдельных статьях, которые мы обозначили римскими цифрами в соответствии с Содержанием. Библиографические сведения включены в надлежащие пристатейные библиографии.

Вот обозначение, принятое во всём сборнике:

S, G, i: скачиваемый документ i на русском языке есть на нашем сайте www.sheynin.de. Сайт перепечатывает Google, см. Oscar Sheynin.

[i] В примечаниях мы отметили некоторые недостатки статьи, и особо (прим. 1.12) неверное представление об объективной вероятности. Воображаемые разговоры с читателем слишком подробны, но нет вторжений в примыкающие соображения и понятия. Так, нет никаких правил приближённых вычислений, и, хуже того, допущены ошибки в подобных расчётах (прим. 2.11).

Отсутствие *вторжений* затруднит читателям более глубокое знакомство с теорией вероятностей. Так, автор широко пользуется схемой случаев, но ведь существуют учебники и руководства, которые обходятся без неё. Пояснение случайного события в самом начале брошюры явно недостаточное.

Мы литературно отредактировали текст; в скверной отечественной традиции автор не обращала должного внимания на свой стиль. Мы старались ссылаться только на свои сочинения, что целесообразно в подобных случаях, и исходили из иных методических принципов: описали намного более широкий круг вопросов, *вторглись* где только можно и привели большое число сведений из истории теории вероятностей и статистики.

[ii] Автор неизменно вводит эксперимент, забывая о независимой объективной реальности (прим. 9 и 12). Вообще же мы убеждены, что в популярной литературе на первом месте должно быть не умение решать конкретные задачи, как у автора, а понимание общих принципов. Так, многие понятия остались за рамками брошюры (прим. 17 и др.), и при дальнейшем изучении теории вероятностей читатель вполне может смутиться. Описанию на должном математическом уровне следовало предпосылать краткое, чёткое описание на уровне естествознания (прим. 31). Сведения из истории теории вероятностей нередко ошибочны (например, Паскаль будто бы ошибся).

Никак нельзя оправдать умолчание о введённых Фишером понятий о состоятельных, достаточных и эффективных оценках. Введённые им же латинские квадраты упомянуты фактически, но сам термин не назван. Порядок источников в библиографии непонятен, некоторые из них не упомянуты в основном тексте и их появление в библиографии не разъяснено.

Мы обязаны заключить, что методически брошюра автора неудовлетворительна.

[iii] Автор не комментировал точку зрения Мизеса (§ 3.3). См. о нём Хинчин (1961). Переводчик без всяких оснований чуть ли не на каждом шагу применил пассивные обороты, а порядок слов в предложениях был нередко негодным. Во многих случаях мы исправили эти недостатки.

I

Е. С. Вентцель

Теория вероятностей (первые шаги)

Москва, 1977

1. Что такое теория вероятностей и чем она занимается

В семье математических наук особое место принадлежит теории вероятностей. Эта наука изучает особого рода законы, управляющие случайными явлениями. Глубокое понимание этих законов придёт к вам позже, когда (если) вы будете изучать теорию вероятностей специально. Моя задача куда скромнее: дать читателю кое-какие элементарные представления о теории вероятностей, о её задачах и методах, возможностях и ограничениях.

Современный период развития науки характеризуется широким наступлением вероятностных (статистических) методов на всех фронтах, во всех областях знаний[1.1]. Элементарное знакомство с теорией вероятностей в наше время требуется от каждого инженера, научного работника, организатора производства. Однако, опыт учит, что теория вероятностей вообще довольно трудно даётся начинающим, что к её специфическим особенностям нелегко приноровиться тому, кто привык совсем к другим, традиционным научным приёмам. Именно первые шаги в понимании и применении вероятностных методов оказываются наиболее трудными. Чем раньше преодолён этот своеобразный психологический барьер, тем лучше.

Не претендуя на систематическое изложение теории вероятностей, мы попытаемся облегчить начинающему читателю именно эти первые шаги. Несмотря на непринуждённую (подчас шуточную) форму изложения, внимательному читателю придётся иной раз серьёзно задуматься.

Прежде всего, несколько слов о *случайных явлениях*. Что это такое? Представьте себе, что производится некоторый *опыт* (или *испытание*), исход которого не может быть заранее предсказан. Например, мы подбрасываем монету. Заранее нельзя сказать, как она упадёт: гербом или решкой[1.2]. Другой пример: вынимаем наугад карту из колоды. Заранее нельзя сказать, какой она будет масти. Ещё пример: мы выходим к автобусной остановке в момент времени, не связанный с расписанием движения (мы этого расписания не знаем). Сколько времени придётся нам ждать нужного номера? Заранее сказать нельзя. Это, как говорится, *зависит от случая*. Сколько изделий завода забракует отдел технического контроля? Тоже заранее сказать нельзя.

Все эти примеры относятся к области случайных явлений. В каждом из них исход опыта заранее непредсказуем. Если такие опыты с неопределённым исходом повторять раз за разом, то от раза к разу результат будет меняться. Например, взвешивая несколько раз подряд одно и то же тело на точных весах, мы будем получать, вообще говоря, различные значения веса. В чём причина этих различий? В том, что условия опыта, которые нам представляются одинаковыми, собственно говоря, различны: на исход каждого из них влияет множество малых, трудно уловимых факторов, обуславливающих в своей совокупности неопределённость исхода.

Итак, пусть производится некоторый опыт, результат которого заранее неизвестен. *Случайным событием* (или просто *событием*) мы назовём всякий факт, который в результате опыта может произойти или не произойти[1.3]. Например, при подбрасывании монеты может произойти (а может и не произойти) событие A – *выпадение герба*. В другом опыте *подбрасывание двух монет* может произойти (или не произойти) событие B – выпадение двух гербов. Ещё пример: некто покупает билет спортлото и отмечает в нём 6 номеров из 49. Через некоторое время публикуется список *выигрышных* номеров. В этом опыте может произойти (или не произойти) каждое из событий

A – три номера из отмеченных на билете совпали с опубликованными (*угадано три номера*);

B – *угадано четыре номера*; C – *угадано пять номеров*;
и, наконец, самое счастливое (и самое неправдоподобное!) событие D – *угаданы все шесть номеров*.

Так вот, теория вероятностей позволяет нам измерять количественно степень *правдоподобия* (или вероятности) различных событий. Сравнивать их друг с другом по вероятности, но главное – на основе оценки вероятностей *предсказывать* исходы случайных явлений.

Ничего не понимаю! – может быть с раздражением подумаете вы. – *Только что было сказано, что случайные явления как раз отличаются непредсказуемостью. И вдруг – предсказывать*!

Подождите, имейте терпение. Предсказывать можно только те случайные события, которые обладают высокой степенью правдоподобия, или, как говорят, большой вероятностью. А определить, какие события относятся к категории весьма правдоподобных, как раз и позволяет нам знание теории вероятностей.

Поговорим о вероятностях событий. Заранее ясно, что не все случайные события одинаково вероятны, что среди них бывают более или менее вероятные. Возьмём опыт *бросание игральной кости*. Как по вашему, какое событие в этом опыте более вероятно:

A – выпадение шести очков, или

B – выпадение чётного числа очков?

Если вы затрудняетесь ответить, закройте книжку и прекратите читать. Но подобное событие в высшей степени неправдоподобно (маловероятно!). Напротив, можно почти наверняка поручиться, что читатель сразу же скажет:

Ну, что за вопрос? Конечно же, событие B!

И будет совершенно прав. Потому что элементарное понимание того, что такое *вероятность события*, присуще любому человеку со здравым смыслом. Мы со всех сторон окружены случайными явлениями, случайными событиями и с детства привыкли, обдумывая свои действия, как-то оценивать вероятности событий, различать среди них вероятные, маловероятные и совсем неправдоподобные.

Если вероятность какого-то события весьма мала, здравый смысл подсказывает нам не рассчитывать всерьёз на его появление[1.4]. Пусть, например, в книге 500 страниц, и на какой-то из них напечатана интересующая нас формула. Можно ли всерьёз рассчитывать на то, что, раскрыв книгу наугад, мы наткнёмся как раз на эту страницу? Очевидно, нельзя. Это событие возможно, но маловероятно.

А теперь договоримся, как мы будем вычислять (оценивать) вероятности случайных событий. Для этого надо прежде всего выбрать единицу измерения. За такую единицу в теории вероятностей принято считать *вероятность достоверного события*. Достоверным называется такое событие, которое в данном опыте непременно произойдёт. Например, *выпадение не более шести очков при броске игральной кости*. Вероятность достоверного события будем считать равной единице, а нулевую вероятность припишем *невозможному событию*, т. е. такому, которое в данном опыте вовсе не может произойти, например, *выпадение отрицательного числа очков при броске игральной кости*.

Условимся обозначать вероятность случайного события *A* через $P(A)$. Очевидно, она всегда должна быть заключена между нулём и единицей:

$$0 \le P(A) \le 1. \qquad (1.1)$$

Запомните это важнейшее свойство вероятности! И если, решая задачи, вы когда-нибудь получите вероятность больше единицы (или, того хуже, отрицательную), будьте совершенно уверены, что ваш расчёт содержит ошибку. Этого не понимал один из моих учеников: получив в контрольной работе по теории вероятностей $P(A) = 4$, он приписал в качестве пояснения: *что показывает, что событие более чем вероятно*[1.5].

Но вернёмся к нашим рассуждениям. Вероятность невозможного события равна нулю; вероятность достоверного события равна единице; вероятность $P(A)$ любого случайного события A есть некоторое число, заключённое между нулём и единицей. Это число показывает, какую *часть* или долю вероятности достоверного события составляет вероятность события A.

В дальнейшем вы сами научитесь (в некоторых простых задачах) вычислять вероятности случайных событий. Но оставим это *на потом*, а пока поразмыслим о некоторых принципиальных вопросах, связанных с теорией вероятностей и её применениями.

Прежде всего, зададим себе вопрос: а зачем, собственно, нам нужно уметь вычислять вероятности? Разумеется, это само по себе достаточно интересно: уметь оценивать числами степени правдоподобия разных событий, сравнивать их друг с другом. Но наша конечная цель иная: пользуясь вычисленными вероятностями, *предсказывать* исходы опытов в области случайных явлений.

Существуют такие опыты, где, несмотря на наличие случайности, исход оказывается предсказуемым, если не точно, то приближённо. Если не с полной достоверностью, то *почти с полной*. С так называемой *практической достоверностью*. К этому и сводится одна из важнейших задач теории вероятностей: выявлять особого рода события *практически достоверные* и *практически невозможные*.

Событие A называется практически достоверным, если его вероятность не в точности равна единице, но очень близка к ней:

$$P(A) \approx 1.$$

Аналогично, практически невозможным называется событие A, вероятность которого близка к нулю:

$$P(A) \approx 0.$$

Рассмотрим пример: опыт состоит в том, что 100 человек подбрасывают каждый свою монету. Событие A состоит в том, что у всех выпадет герб. Возможно ли такое событие?

Теоретически возможно, можно представить себе такую *игру случая*. Но вероятность этого события ничтожно мала; вскоре мы вычислим её и убедимся, что она равна $(1/2)^{100}$.

Вывод: событие A можно считать практически невозможным. А противоположное ему событие $\overline{A}$, состоящее в том, что A не произойдёт (хотя бы у одного человека выпадет решка), будет практически достоверно.

В задачах теории вероятностей практически невозможное и практически достоверное события всегда появляются парами. Если какое-то событие в данном опыте практически достоверно, то противоположное событие $\overline{A}$ практически невозможно и обратно.

Допустим, по нашему расчёту в данном опыте событие A практически достоверно. Что из этого следует? Мы можем *предсказать* появление события A в данном опыте! Правда, предсказать не совсем наверняка, а *почти наверняка*, но и это уже большой успех: ведь речь идёт о случайных явлениях!

Пользуясь методами теории вероятностей, мы можем предсказать, например, максимальную погрешность, которую почти наверняка не превзойдёт ошибка расчёта на компьютере[**1.6**]; максимальное и минимальное практически возможное число запасных частей, которое понадобится в автохозяйстве за год; границы, в которых с практической достоверностью будет заключено число попаданий в цель; максимальное возможное число бракованных деталей на производстве и т. п.

Отметим, что такие предсказания, как правило, оказываются возможными, если речь идёт не о единичном, отдельном случайном явлении, а о *массе* однородных случайных явлений. Например, нельзя предсказать заранее, выпадет ли подброшенная монета гербом или решкой, и никакая теория вероятностей нам в этом не поможет. Но если производится большое число, скажем, 500 или 1000 бросаний, то мы уже можем предсказать, в каких пределах будет лежать число выпавших гербов. Примеры подобных предсказаний встретятся в дальнейшем. Все они даются не наверняка, а *почти наверняка* и осуществляются не буквально всегда, а в большинстве случаев.

Часто спрашивают: а как велика должна быть вероятность события, чтобы считать его практически достоверным? Должна ли она быть равна, скажем, 0,99? Или 0,995? Или ещё больше, 0,999? Однозначно ответить на этот вопрос нельзя: всё зависит от того, насколько важен для нас успех предсказания, и что нам угрожает, если оно не сбудется.

Пусть, например, мы с вероятностью 0,99 предсказываем, что, пользуясь таким-то видом транспорта, мы не опоздаем на работу больше, чем на 10 минут. Можно ли считать, что это событие практически достоверно? Пожалуй, да. Но можно ли считать практически достоверным событие *благополучная посадка космического корабля*, если оно гарантируется с той же вероятностью 0,99? Очевидно, нет! Запомним, что в любом предсказании, которое даётся методами теории вероятностей, всегда присутствуют две особенности:

1. Предсказание делается не наверняка, а *почти наверняка*, т. е. большой вероятностью.

2. Численное значение этой большой вероятности (иначе, *уровня доверия*) назначается самим исследователем более или менее произвольно, но в согласии со здравым смыслом, с учётом важности, которую имеет для нас успех предсказания[1.7].

Если после всех этих оговорок вы ещё не окончательно разочаровались в теории вероятностей, читайте дальше и познакомьтесь с некоторыми элементарными приёмами расчёта вероятностей случайных событий. Кое-какое представление об этих приёмах у вас, по-видимому, уже есть. Чему, например, равна вероятность выпадения герба при подбрасывании монеты? Почти наверняка (с очень большой вероятностью!) вы, не задумываясь, ответите: 1/2 . И этот ответ верен, если монета симметрична, правильной формы [и её плотность постоянна], а исход *монета встанет на ребро*, практически невозможен[1.8].

Так же легко вы ответите на вопрос: какова вероятность появления шести очков при броске игральной кости? Почти наверное вы ответите: 1/6 (разумеется, с теми же оговорками). Но как вы получили этот ответ? Очевидно, вы подсчитали число возможных исходов опыта (их шесть) и в силу симметричности они равновероятны. Естественно поэтому приписать каждому вероятность 1/6, что вы и сделали. И были совершенно правы.

А теперь ещё один вопрос: чему равна вероятность появления более четырёх очков? Поразмыслив немного, вы, вероятно, ответите: 1/3 и опять будете совершенно правы. Из шести равновозможных исходов опыта два (пятёрка и шестёрка) дают более четырёх очков (*благоприятны* этому событию). Разделив 2 на 6, вы и получили правильный ответ, 1/3.

Браво! Сами того не подозревая, вы применили здесь *классический способ вычисления вероятностей в схеме случаев*. А что такое *схема случаев*? Объясняем. Сначала введём несколько терминов. (В теории вероятностей, как и во многих других науках, терминология играет большую роль.)

Пусть производится опыт, который имеет ряд возможных исходов: A_1, A_2, … A_n. Эти исходы (события) *несовместимы*, если они взаимно исключают друг друга, т. е. никакие два из них не могут появиться совместно.

Они образуют *полную группу*, если исчерпывают собой все возможные исходы, т. е. если не может быть так, чтобы в результате опыта ни одно из них не произошло.

Они, события $A_1, A_2, \dots A_n$, называются *равновозможными*, если условия опыта обеспечивают одинаковую возможность (вероятность) появления каждого из них.

Если эти события обладают всеми тремя свойствами, т. е. если они несовместны, образуют полную группу и равновозможны, то они называются *случаями*[1.9], а про опыт говорят, что он сводится к *схеме случаев*. К примеру, опыт *подбрасывание монеты* сводится к схеме случаев, потому что события A_1 – выпадение герба, и A_2 – выпадение решки, несовместны, образуют полную группу и равновозможны.

К той же схеме сводится опыт бросание игральной кости. В нём шесть несовместных, образующих полную группу и равновозможных исходов, которые можно пронумеровать по числу выпадающих очков: $A_1, A_2, \dots A_6$.

Рассмотрим теперь опыт *подбрасывание двух монет* и попробуем перечислить случаи. Если вы легкомысленны и торопливы, то поспешите назвать три события:

B_1, выпадение двух гербов; B_2, выпадение двух решек;

B_3, выпадение одного герба и одной решки

Нет, эти три события – *не случаи*, так как они не равновозможны[1.9]. Событие B_3 вдвое более вероятно, чем каждое из остальных. Перечислим действительные *случаи* для нашего опыта:

A_1, герб на первой монете и герб на второй

A_2, решка на первой монете и решка на второй

A_3, герб на первой монете и решка на второй

A_4, решка на первой монете и герб на второй

События B_1 и B_2 совпадают с A_1 и A_2, но вот событие B_3 распадается на два варианта, A_3 и A_4 и поэтому оно вдвое вероятнее каждого из остальных.

В следующем примере мы впервые употребим традиционный для теории вероятностей образ *урны* с шарами. Урна, это, попросту говоря, сосуд, в котором находится то или иное число шаров различного цвета. Они тщательно перемешаны и одинаковы наощупь, что обеспечивает равновероятный выбор каждого из них. Эти условия подразумеваются во всех задачах, которые мы будем рассматривать.

Каждую задачу теории вероятностей, в которой опыт сводится к схеме случаев, можно свести к той или иной задаче, где речь пойдёт о вынимании шаров из урн. *Задачи из урны* являются своего рода единым языком, на котором можно излагать самые разнообразные по внешней форме задачи.

Так вот, пусть имеется урна, в которой лежат семь шаров, три белых и четыре чёрных. Из урны наугад выбирается один шар. Требуется перечислить соответствующие случаи. При ответе можно снова ошибиться и легкомысленно назвать два события: B_1 – появление белого шара и B_2 – появление чёрного шара.

Если у Вас появилось такое искушение, вы не созданы для теории вероятностей[1.10]. Но, всего вернее, вы так уже не ответите: вы поняли, что события B_1 и B_2 не равновозможны: белых шаров три, а чёрных – четыре. В данном опыте не два случая, а семь (по числу шаров), которые можно обозначить так: $Б_1$, $Б_2$, $Б_3$, $Ч_1$, $Ч_2$, $Ч_3$, $Ч_4$, т. е. белый первый, белый второй, …, чёрный четвёртый. Эти события несовместны, образуют полную группу и равновозможны, а потому представляют собой случаи.

Для всякого ли опыта можно построить группу случаев? Нет, далеко не для всякого. Пусть опыт состоит в подбрасывании несимметричной (погнутой) монеты, тогда события *выпадение герба* и *выпадение решки* уже не будут случаями, так как они не равновозможны; можно так изогнуть монету, что одно из них станет вообще невозможным.

Этот опыт не сводится к схеме случаев, а чтобы сводился, опыт должен обладать некоторой симметрией, которая обеспечит равновозможность исходов. Эта симметрия иногда достигается за счёт физической симметрии предметов, применяемых в опыте (монета, игральная кость), а иногда за счёт перемешивания, *тасовки* элементов, обеспечивающей равновозможный выбор любого из них (урна с шарами, колода карт, барабан с лотерейными билетами и т. д.).

Чаще всего такая симметрия наблюдается в искусственно организованных опытах, где приняты специальные меры для её обеспечения. Типичными примерами являются азартные игры (*орлянка*, *кости*, некоторые карточные игры). Заметим, что именно с анализа таких игр началось когда-то развитие теории вероятностей.

Если опыт сводится к схеме случаев, то вероятность любого события A в нём может быть подсчитана как *отношение числа случаев, благоприятных событию A, к общему числу случаев*:

$$P(A) = \frac{m_A}{n}, \qquad (1.2)$$

где n – общее число случаев, m_A – число случаев, благоприятных событию A (обеспечивающих его появление).

Формула (1.2), так называемая *классическая формула*[1.11], применялась для вычисления вероятностей событий с самого начала возникновения науки о случайных явлениях. Долгое время она даже рассматривалась как *определение вероятности*. Те опыты, которые не обладали симметрией возможных исходов, искусственно подгонялись под схему случаев.

В наше время точка зрения на вероятность и приёмы изложения теории вероятностей изменились: для нас с вами формула (1.2) не будет универсальной, но позволит вычислять вероятности событий в некоторых наиболее простых опытах. В последующих главах вы увидите, как вычисляются вероятности событий, когда опыт не сводится к схеме случаев.

Теперь мы рассмотрим примеры на вычисление вероятностей случайных событий по формуле (1.2). Некоторые из них очень просты, другие – не очень.

Пример 1. Подбрасывание двух монет. Найти вероятность A того, что появляется хотя бы один герб.

Решение. Случаев здесь 4, см. выше. Три из них благоприятны событию A, значит $m_A = 3$, $n = 4$ и по формуле (1.2) $P(A) = 3/4$.

Пример 2. В урне 3 белых и 4 чёрных шара. Вынимается один шар. Найти вероятность A того, что этот шар – белый.

Решение. $m_A = 3$, $n = 7$, $P(A) = 3/7$.

Пример 3. Урна с тем же составом шаров. Мы вынимаем один шар и, не глядя, убираем его в сторону. Затем вынимаем второй шар. Найти вероятность того, что этот шар будет белым (событие A).

Решение. Если хорошенько подумать, можно убедиться в том, что предварительное вынимание какого-то неизвестного шара никак не влияет на вероятность появления белого шара при следующем вынимании[1.12]. Она останется той же, что в предыдущем примере: 3/7. С первого взгляда этот результат может показаться неверным: ведь при втором вынимании в урне стало 6, а не 7 шаров. Не означает ли это, что и число случаев стало тоже 6? Нет, не значит! Пока мы не знаем, какого цвета был первый шар, число случаев остаётся равным семи. Чтобы убедиться в этом, изменим условия опыта ещё резче: предварительно, не глядя, вынем из урны все шары кроме одного, Какова вероятность, что он белый? Очевидно, она равна 3/7, так как решительно всё равно: выбрать ли шар для того, чтобы *вынуть* его из урны, или, чтобы *оставить* в ней.

Если вы всё ещё не убеждены, что вероятность появления белого шара остаётся равной 3/7, сколько шаров мы ни вынимали бы не глядя предварительно, представьте себе такой опыт. Четыре из всех семи вынутых шаров не глядя кладём на шкаф, остальные разбрасываем по полу и наступаем на один из последних. Какова вероятность, что этот шар белый? Если вам и теперь неясно, что она равна 3/7, то вам уже ничего не поможет: все наши доводы исчерпаны.

Пример 4. Состав урны тот же. Из неё вынимается сразу два шара. Какова вероятность того, что оба шара белые (событие А)?

Решение. Задача немного труднее двух предыдущих: здесь не так просто подсчитать общее число случаев и их благоприятное число. Придётся задуматься о том, сколькими способами можно выбрать из урны 2 шара? Выбрать два белых шара из имеющихся?

Подобными вопросами занимается специальная наука, комбинаторика, входящая в состав элементарной алгебры[1.13]. Из формул комбинаторики нам понадобится только одна, формула для числа сочетаний. Число сочетаний из k элементов по s это число способов, которыми можно выбрать s различных элементов из k, причём комбинации различаются только составом элементов, но не их порядком. Это число вычисляется по формуле

$$C_k^s = \frac{k(k-1)...(k-s+1)}{1 \cdot 2 \cdot ... \cdot s}, \text{ причём } C_k^s = C_k^{k-s}. \qquad (1.3;\ 1.4)$$

По формуле (1.3) всех возможных случаев выбора двух шаров из семи будет

$$C_7^2 = \frac{7 \cdot 6}{1 \cdot 2} = 21.$$

Подсчитаем теперь число способов, которыми можно выбрать два шара из имеющихся трёх. По формулам (1.3) и (1.4)

$$m_A = C_3^2 = C_3^1 = 3,$$

и по формуле (1.2)

$$P(A) = \frac{C_3^1}{C_7^2} = \frac{3}{21} = \frac{1}{7}.$$

Пример 5. Урна всё с тем же составом (потерпите немножко, скоро мы с ней расстанемся) Из неё сразу же вынимают три шара, Найти вероятность того, что два из них чёрные и один белый (событие A).

Решение. Общее число случаев здесь равно

$$n = C_7^3 = \frac{7 \cdot 6 \cdot 5}{1 \cdot 2 \cdot 3} = 35.$$

Подсчитаем число благоприятных случаев. Сколькими способами можно выбрать два из четырёх шаров? Очевидно, шестью:

$$C_4^2 = \frac{4 \cdot 3}{1 \cdot 2} = 6.$$

Но это ещё не всё: к каждой комбинации чёрных шаров можно тремя способами присоединить один белый из трёх. Каждая комбинация чёрных шаров может сочетаться с каждым из белых и общее число благоприятных случаев поэтому равно

$$m_A = C_4^2 \cdot C_3^1 = 6 \cdot 3 = 18,$$

и по формуле (1.2) $P(A) = 18/35$.

Теперь мы созрели для того, чтобы решить следующую задачу.

Задача. В урне a белых и b чёрных шаров. Из неё наугад вынимают k шаров. Найти вероятность того, что среди них будет l белых и, стало быть, $k - l$ черных $l \le a$, $k - l \le b$.

Решение. Подсчитаем число благоприятных случаев m_A. Число способов для выбора l белых шаров из a и число способов, которыми к ним можно присоединить $k - l$ чёрных шаров, равно, соответственно, C_a^l и C_b^{k-l}, так что

$$P(A) = \frac{C_a^l \cdot C_b^{k-l}}{C_{a+b}^k}. \qquad (1.5)$$

Знаменатель здесь равен числу всех случаев.

Эту формулу можно применять в разных областях, например, в задачах о выборочном контроле продукции. *Урной* в таких задачах служит партия изделий, среди которых какое-то количество доброкачественных (*белые шары*) и какое-то количество дефектных (*чёрные шары*), а роль вынимаемых k шаров играет контрольная партия изделий.

Теперь рассмотрим ещё один пример, который, быть может, покажется вам интересным.

Пример 6. Некто купил карточку Спортлото и наугад отметил в ней 6 из имеющихся 49 номеров. Найти вероятность того, что он правильно угадал 3 из тех шести номеров, которые выиграли.

Решение. Рассмотрим происшедшее событие A. Так ведь это наша предыдущая задача. Действительно, 49 номеров, в числе которых 6 выигрышных это как бы урна, в которой 6 белых шаров и 43 чёрных, надо найти вероятность того, что при выборе наугад шести шаров 3 окажутся белыми, и 3 – чёрными. Это мы уже умеем! В формуле (1.5) положим $a = 6$, $b = 43$, $k = 6$ и $l = 3$. Тогда

$$P(A) = \frac{C_6^3 \cdot C_{43}^3}{C_{a+b}^k} = \frac{6 \cdot 5 \cdot 4 \cdot 43 \cdot 42 \cdot 41 \cdot 4 \cdot 5}{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44}.$$

Если не поленится и вычислить значение этой дроби, скажем, при помощи логарифмической линейки, то получим

$$P(A) \approx 0{,}0176.$$

Итак, вероятность правильно угадать три номера из шести [выбранных из 49] очень невелика, около 1,8%. Ясно, что угадать вероятность угадать 4, 5 и (о, чудо!) все 6 ещё меньше, значительно меньше. Если у вас повышенная склонность к арифметическим расчётам, можете попробовать и вычислить все эти вероятности. Значит, кое-что вы уже умеете …

2. Вероятность и частота

Вы познакомились с предметом теории вероятностей, с некоторыми её основными понятиями (событие, вероятность события) и научились вычислять вероятности событий по так называемой классической формуле

$$P(A) = \frac{m_A}{n}, \qquad (1.2) = (2.1)$$

где n – общее число случаев и m_A – число случаев, благоприятных событию A.

Из этого не следует, что вы хорошо вооружены для применения теории вероятностей на практике. К сожалению, сфера применения классической формулы (2.1) не так обширна, как нам хотелось бы. Она пригодна только в тех опытах, которые обладают симметрией возможных исходов (сводятся к схеме случаев). А что это за опыты? Они главным образом относятся к области азартных игр, где симметрия достигается специальными мерами. В настоящее время профессия игрока не так уже распространена[2.1] (в давние времена это было иначе), практическое значение формулы (2.1) для вычисления вероятностей очень ограничено. Большинство опытов со случайным исходом, с которыми нам приходится иметь дело на практике, не сводится к схеме случаев. Как же быть с вероятностями событий? Существуют ли они для таких опытов? А если существуют, как их вычислять?

Этим мы и займёмся. Нам придётся ввести новое основное понятие теории вероятностей, понятие *частоты события*.

Подойдём к нему несколько издалека. Представим себе какой-то опыт, не сводящийся к схеме случаев, например, подбрасывание *неправильной*, несимметричной игральной кости[2.2].

Рассмотрим в этом опыте событие A – выпадение 6 очков. Опыт не сводится к схеме случаев и формула (2.1) неприменима, нельзя считать, что $P(A) = 1/6$. Как же найти эту вероятность, хотя бы приближённо? На этот вопрос легко ответит каждый разумный человек: надо попробовать, побросать кость много раз подряд и посмотреть, насколько часто (в какой доле всех опытов) появится событие A. Эту долю (или *процент*) появлений события A и можно принять за его вероятность.

Разумный человек безусловно прав. Сам того не зная, он воспользовался понятием *частоты события*. В серии из N опытов ей называется *отношение числа опытов, в которых это событие произошло, к общему числу произведённых опытов*.

Эту частоту иначе называют *статистической вероятностью*. В опытах, не обладающих симметрией возможных исходов, именно статистика массовых случайных явлений служит базой для определения вероятностей события.

Частоту события A будем обозначать $P^*(A)$, отличая её от родственной ей вероятности $P(A)$. По определению

$$P^*(A) = \frac{M_A}{N}, \qquad (2.2)$$

где N – общее число опытов, M_A – число опытов, в которых появилось событие A; короче, число появления события A.

Несмотря на внешнее сходство формул (2.1) и (2.2), они совершенно различны по существу. Формула (2.2) служит для *экспериментального определения* частоты события. Чтобы ей воспользоваться, нужен опытный статистический материал.

Поразмышляем немного над природой частоты. Ясно, что между частотой события и его вероятностью существует некоторая связь: более вероятные события, в общем, происходят чаще, чем маловероятные. Тем не менее, понятия *частота* и *вероятность* отнюдь не тождественны. Сходство, *родство* между частотой и вероятностью появляется тем заметнее, чем большее число опытов произведено. При малом числе опытов частота события в значительной степени случайна и может значительно отличаться от вероятности. Например, при 10 подбрасываниях монеты герб вполне может выпасть 3 раза, частота его появления будет 0,3, что сильно отличается от вероятности 0,5. Однако, при увеличении числа опытов частота события постепенно теряет свой случайный характер.

Случайные обстоятельства, сопровождающие каждый отдельный опыт, в массе опытов взаимно погашаются[2.3], и частота постепенно стабилизируется, приближаясь с незначительными колебаниями к некоторой средней, постоянной величине. Естественно предположить, что эта постоянная величина и есть *вероятность* события.

Проверить это утверждение мы можем только для тех событий, вероятности которых могут быть вычислены по формуле (2.1), т. е. для опытов, сводящихся к схеме случаев[2.4]. И что же? Утверждение оказывается верным. Если не полениться, вы сами можете проверить его, хотя бы на простеньком примере: при увеличении числа подбрасываний монеты частота появления герба будет приближаться к вероятности этого события, к 0,5.

Подбрасывайте монету 10, 20, … раз, пока хватит терпения, и подсчитайте частоту появления герба в зависимости от числа подбрасываний. Для экономии сил и времени можете применить *маленькую хитрость*: подбрасывайте монеты десятками (разумеется, предварительно встряхнув их хорошенько). Потом нанесите полученные числа на график. В таких опытах вы не будете одиноки: ими не брезговали и крупные учёные, которые интересовались природой случайности. Карл Пирсон подбросил монету 24 тысячи раз[2.5] и получил 12 012 гербов, что даёт частоту, чрезвычайно близкую к 0,5. Неоднократно производились опыты и с подбрасыванием игральной кости и аналогично получали частоты, близкие к 1/6. Таким образом, факт приближения частот к вероятностям можно считать экспериментально проверенным.

Свойство *устойчивости частот* при большом числе однородных опытов – одна из наиболее характерных закономерностей, наблюдаемых в массовых случайных явлениях. Если мы повторяем (воспроизводим) один и тот же опыт много раз, причём обеспечиваем *независимость* исходов отдельных опытов, частота события становится всё менее и менее случайной, выравнивается, приближается к постоянной.

Для опытов, сводящихся к схеме случаев, можно непосредственно убедиться, что эта постоянная – не что иное, как *вероятность события*. А если опыт не сводится к схеме случаев, а частота обнаруживает устойчивость, приближается к постоянной? Ну, так примем, что правило остаётся в силе, и назовём эту постоянную вероятностью события. Таким образом, мы ввели вероятности не только для событий и опытов, сводящихся к схеме случаев, но и для тех, которые к этой схеме не сводятся, лишь бы в них наблюдалось свойство *устойчивости частот*.

Как же обстоит дело с этой устойчивостью? Все ли случайные явления её обнаруживают? Не все, но многие. Поясним эту далеко не простую мысль рассуждениями. Постарайтесь хорошенько в них вникнуть, это избавит вас от возможных ошибок в применении вероятностных методов.

Мы предполагали, что можем неограниченно повторять один и тот же опыт (подбрасывание монеты, игральной кости и т. п,). Действительно, *такой* опыт ничто не мешает нам повторять сколько угодно раз, было бы время. Но ведь бывает так, что *опыт* производим не мы сами, а только наблюдаем результаты тех *опытов*, которые *ставит* за нас природа. В подобных случаях нельзя заранее гарантировать устойчивость частот: в ней приходится каждый раз убеждаться.

Пусть, например, *опыт* состоит в рождении ребёнка, и нас интересует вероятность события *рождение мальчика* (природа ежегодно ставит огромное множество *опытов* подобного рода). Имеется ли в таких случайных явлениях устойчивость частот? Да, имеется. Экспериментально установлено, что частота рождения мальчика очень устойчива, почти не зависит от географического положения страны, национальности родителей, их возраста и т. п. Эта частота несколько выше, чем 0,5 (приближённо равна 0,51)[2.6].

Свойством устойчивости частот, по крайней мере, не в течение слишком большого времени, обладают такие, например, случайные явления, как отказы технических устройств, брак на производстве; ошибки механизмов; заболеваемость и смертность населения; метеорологические и многие биологические явления. Эта устойчивость и позволяет с успехом применять вероятностные методы для изучения таких случайных явлений, для прогнозирования и управления ими.

Но существуют и такие случайные явления, для которых устойчивость частот сомнительна, а то и просто не существует. Это такие явления, где бессмысленно говорить о большом числе *однородных опытов*, где не существует (или в принципе не может быть получен) достаточно обширный массив статистических данных. В подобных явлениях тоже могут наблюдаться те или иные события, вообще представляющиеся нам более или менее правдоподобными, хотя у нас нет возможности приписать им какие-либо определённые вероятности. Так, вряд ли возможно (и вряд ли целесообразно) вычислять вероятность того, что через три года женщины будут носить длинные юбки или мужчины – длинные усы. Соответствующего массива статистических данных не существует. Если рассматривать следующие друг за другом годы как *опыты*, то их ни в каком смысле нельзя считать *однородными*.

Другой пример, где ещё менее осмысленно судить о вероятности события. Пусть кто-то задался целью вычислить вероятность того, что на Марсе существует органическая жизнь. По-видимому, этот вопрос будет решён в ближайшие годы; многие учёные считает наличие органической жизни на Марсе вполне правдоподобным. Но *степень правдоподобия* ещё не есть вероятность! Оценивая вероятность *на-глаз*, мы неизбежно оказываемся в мире туманных домыслов. Если мы хотим оперировать подлинными вероятностями, то должны опираться на достаточно обширную статистику. А можно ли говорить об обширном массиве статистических данных в этом случае? Разумеется, нет. Ведь Марс-то один!

Итак, о вероятностях событий в опытах, не сводящихся к схеме случаев, мы будем говорить только, если они относятся к категории массовых случайных явлений, обладающих свойством устойчивости частот[2.7]. Но обладают они этим свойством или нет, обычно решается на уровне здравого смысла. Можно ли достаточно много раз воспроизвести опыт, не меняя существенно его условия? Есть ли надежда собрать необходимую статистику? На эти вопросы должен ответить сам исследователь, который собирается применить вероятностные методы в какой-либо области.

Остановимся ещё на одном вопросе, непосредственно связанном с предыдущим. Говоря о вероятности события в каком-то опыте, необходимо прежде всего тщательно оговорить основные условия опыта, которые предполагаются фиксированными и не меняются при его повторении. Одна из самых частых ошибок в практическом применении теории вероятностей, особенно у начинающих, это обсуждение вероятности события без уточнения условий опыта, о котором идёт речь, или того *статистического массива* случайных явлений, в котором эта вероятность могла бы проявиться в форме частоты.

Например, совершенно бессмысленно говорить о вероятности опоздания поезда. Какого поезда? Откуда и куда он идёт? Товарный он или пассажирский? На какой железной дороге? Только после уточнения всех этих подробностей можно позволить себе рассмотрение вероятности данного события как определённого числа. Это предупреждение о *подводных камнях*, угрожающих тем, кто в теории вероятностей интересуется не только забавными задачками на *монеты* и *игральные карты*[2.8], а хочет применить её методы на практике для достижения реальных целей.

Теперь предположим, что все эти условия выполнены: есть возможность произвести достаточно много однородных опытов и устойчивость частот налицо. Тогда для приближённого определения вероятности события можно воспользоваться его частотой в длинной серии опытов. Ведь мы договорились, что при увеличении числа опытов частота события приближается к его вероятности. Как будто очень просто, а на деле – не очень. Соотношение между частотой и вероятностью довольно тонко.

Поразмыслим немного над термином *приближается*. Что это означает? *Что за странный вопрос*, быть может, подумаете вы. *Это значит: становится всё ближе и ближе, так о чём тут размышлять*? Есть, есть о чём. Ведь речь идёт о случайных явлениях, где всё по-особому, *не как у людей*. Когда мы говорим, что сумма членов геометрической прогрессии

$$1+\frac{1}{2}+\frac{1}{2^2}+\ldots+\frac{1}{2^n}$$

при увеличении n стремится (неограниченно приближается) к двум, это значит, что, чем больше членов мы возьмём, тем ближе сумма будет к своему пределу, то это совершенно достоверно.

В области случайных явлений таких категорических утверждений делать нельзя. Да, частота события, вообще говоря, приближается к вероятности, но по-своему, не совсем наверняка, а *почти наверняка*, с очень большой вероятностью. Может оказаться, что частота события даже при очень большом числе опытов будет сильно отличаться от вероятности. Но вероятность такого события очень мала, тем меньше, чем большее число опытов произведено.

Пусть, например, мы подбросили монету 100 раз. Спросим себя, может ли случиться, что частота появления герба будет сильно отличаться от вероятности 0,5 и, например, будет равна нулю. Такое событие теоретически возможно (не противоречит законам природы), но вероятность его очень мала. Вычислим её (мы уже умеем решать такие простые задачи).

Подсчитаем общее число случаев. Каждая монета может выпасть двумя способами, и любой из них может сочетаться с любым способом выпадения каждой из остальных монет. Обще число случаев равно 2^{100}. Благоприятен нашему событию (не появлению ни одного герба) только один случай, и его вероятность окажется равной $1/2^{100}$. Эта величина чрезвычайно мала, порядка числа с 30 нулями после запятой. Событие с такой вероятностью можно смело считать практически невозможным. В действительности и гораздо меньшие отклонения частоты от вероятности тоже будут практически невозможными.

Но какие же отклонения частоты от вероятности при большом числе опытов практически возможны? Мы сейчас напишем формулу, которая позволит ответить на этот вопрос. Доказать её мы, к сожалению, не можем; впрочем, некоторое обоснование будет дано ниже. Пока что вам только остаётся принять её на веру[2.9].

Пусть производится N опытов, в каждом из которых событие A появляется с вероятностью p. Тогда частота этого события с вероятностью 0,95 укладывается на участке

$$p \pm 2\sqrt{\frac{p(1-p)}{N}}. \qquad (2.3)$$

Участок, определяемый формулой (2.3), мы будем называть *доверительным интервалом* для частоты события, который соответствует *уровню доверия* 0,95. Это значит, что наше предсказание, состоящее в том, что частота не выйдет за пределы этого участка, будет выполняться почти всегда, а точнее, в 95% случаев. Разумеется, в 5% случаев мы будем ошибаться, но … *волков бояться – в лес не ходить*. Другими словами, если бояться ошибок, то не надо браться за предсказания в области случайных явлений: все они осуществляются не наверняка, а почти наверняка.

Но если вероятность ошибки 0,05 покажется слишком высокой[2.10], то можно перестраховаться и построить немного более широкий доверительный интервал: в формуле (2.3) заменить коэффициент 2 на 3. Новому интервалу соответствует очень высокий уровень доверия, 0,997.

Ну, а если потребовать полной достоверности предсказания? Доверительной вероятности, равной единице? Тогда мы сможем только утверждать, что частота события не выйдет за пределы участка от 0 до 1. Но это утверждение довольно тривиально, ясное и без всяких вычислений.

В гл. 1 мы отмечали, что назначение той вероятности, при которой мы считаем событие практически достоверным, в какой-то мере произвольно. Уговоримся в дальнейшем довольствоваться скромным уровнем доверия 0,95 и пользоваться формулой (2.3). В конце концов, если мы иной раз и ошибёмся, ничего катастрофического не произойдёт.

Оценим по этой формуле практически возможный диапазон значений (доверительный интервал) для частоты появления герба при 100 подбрасываниях монеты. Для нашего опыта $p = 0{,}5$, $1 - p = 0{,}5$ и

$$0,5 \pm 2\sqrt{\frac{0,25}{100}} = 0,5 \pm 0,1.$$

Таким образом, с вероятностью (с уровнем доверия) 0,95 можно предсказать, что при ста подбрасываниях монеты частота появления герба не отклонится от вероятности больше, чем на 0,1. Гм … Ошибка, прямо скажем, не малая. Как её уменьшить? Очевидно, увеличить число опытов N. Тогда ширина доверительного интервала уменьшится, к сожалению, не так быстро, как нам хотелось бы, а обратно пропорционально $\sqrt{N}$. Например, при $N = 10\,000$ формула (2.3) даст $0,5 \pm 0,01$.

Итак, связь между частотой и вероятностью события можно сформулировать следующим образом:

При достаточно большом числе независимых опытов частота события с практической достоверностью будет сколь угодно близка к его вероятности.

Это положение называется теоремой Якоба Бернулли, а иначе *простейшей формой закона больших чисел*. Мы привели его без доказательства, однако вряд ли у вас возникло серьёзное сомнение в его справедливости.

Итак, мы разобрались в вопросе о том, что означает *частота приближается к вероятности*. Осталось сделать ещё один шаг: приближённо найти вероятность события по его частоте и оценить ошибку этого приближения. В последнем нам поможет всё та же формула (2.3), если хотите, с коэффициентом 3.

Пусть произведено большое число опытов N, найдена частота $P^*(A) = p^*$ события A и желательно приближённо найти его вероятность. Приближённо положим, что искомая вероятность равна частоте

$$p \approx p^*. \qquad (2.4)$$

Теперь оценим практически возможную максимальную ошибку этого приближённого равенства. Воспользуемся формулой (2.3). Она покажет, насколько при уровне доверия 0,95 частота может отличаться от вероятности.

Но, спросите вы, ведь в формулу (2.3) входит неизвестная нам вероятность p, а её-то мы и хотим определить.

Совершенно законный вопрос. Вы правы! Но дело в том, что формула (2.3) служит только для *приближённой прикидки* доверительного интервала. Чтобы *грубо* оценить ошибку в вероятности можно в этой формуле вместо неизвестной вероятности p подставить известную нам и приближённо равную ей частоту p^*.

Так и сделаем. Решим, к примеру, такую задачу. В серии из 400 опытов получена частота события $p^* = 0{,}25$. Определить при уровне доверия 0,95 с какой максимальной практически возможной ошибкой мы найдём вероятность события, полагая её равной 0,25?

По формуле (2.3) с приближённой заменой p на $p^* = 0{,}25$ мы получим

$$0{,}25 \pm 2\sqrt{\frac{0{,}25 \cdot 0{,}75}{400}} \approx 0{,}25 \pm 0{,}043.$$

Итак, максимальная практически возможная ошибка равна 0,043[2.11]. А как быть, если такая точность нас не устраивает? Если нам нужно знать вероятность с ошибкой, скажем, не больше 0,01? Разумеется, надо увеличить число опытов. Но до каких пор? Воспользуемся снова нашей любимой формулой (2.3). Полагая в ней вероятность p приближённо равной частоте $p^* = 0{,}25$ в уже произведённой серии опытов, получим по этой формуле приближённо максимально практически возможную ошибку и положим её равной заданному значению 0,01:

$$2\sqrt{\frac{0{,}25 \cdot 0{,}75}{N}} = 0{,}01.$$

Следовательно, $N = 7500$. Итак, для вычисления вероятности порядка 0,25 по частоте с ошибкой не более 0,01 при уровне доверия 0,95 требуется произвести 7500 опытов. Страшно подумать!

Формула (2.3) быть может с коэффициентом 3 может помочь решению ещё одного вопроса: можно ли объяснить полученное в опыте отклонение частоты от вероятности *случайными причинами* или же оно указывает, что *вероятность не такова, какой мы её считаем*?

Пусть, подбросив монету 800 раз, мы получили частоту появления герба равную 0,52. У нас возникло подозрение, что монета *неправильная* и чаще выпадает гербом, а не решкой. Обосновано ли наше подозрение? Будем исходить из предположения, что всё в порядке: монета правильная, вероятность выпадения герба 0, 5, нормальная[2.12].

Найдём доверительный интервал при уровне доверия 0,95 для частоты появления герба. Если полученное в опыте значение 0,52 укладывается в этот интервал, то всё в норме. Если нет – надо подозревать, что монета *неправильная*. Формула (2.3) приближённо даёт для частоты появления герба интервал 0,5 ±

0,035. Полученное значение частоты укладывается в этот интервал, и монету придётся *очистить* от подозрений.

Аналогичными методами пользуются для того, чтобы судить: случайны или *значимы* различные отклонения от среднего, наблюдаемые в случайных явлениях. Например, случайно был получен некоторый *недовес* в нескольких образцах расфасованных товаров, или же он указывает на систематический обман покупателей. Случайно ли повысился процент выздоровлений у больных, применявших данный препарат, или же это связано с его применением.

Таким образом, в опытах, не сводящихся к схеме случаев, вы научились приближённо находить вероятности событий по статистическим данным и даже как-то оценивать получающуюся при этом ошибку!

Ну и наука, эта теория вероятностей, может быть подумаете вы. *Если вероятность события нельзя найти по формуле* (2.1)*, приходится ставить опыты, и ставить, и ставить, пока хватит терпения и сил, потом подсчитать частоту события, положить её равной вероятности, и ещё может быть оценить ошибку. Экая скучища!*

Подумаете и будете совершенно неправы. Потому что такое *статистическое* нахождение вероятности – не единственный и далеко не главный приём. Гораздо большее значение имеют не прямые, а косвенные методы вычисления вероятностей. Они позволяют выражать вероятности интересующих нас событий через вероятности других событий, с ними связанных, вероятности сложных событий по вероятностям простых.

Те, в свою очередь, – через вероятности ещё более простых и т. д., пока не дойдём до самых простых, уже *неразложимых* событий[2.13]. Их вероятности либо вычисляются по формуле (2.1), либо находятся экспериментально, через частоты. Последнее, конечно, требует постановки опытов или сбора статистики. Следует стремиться к тому, чтобы цепочка событий была как можно длинней, а необходимые опыты – как можно проще, дешевле. Для достижения этого нужно как можно больше сведений получить расчётным путём и как можно меньше – экспериментальным. Ведь из всех материалов, из которых делаются сведения, дешевле всего бумага и время учёного[2.14]. О том, как вычисляются вероятности сложных событий через вероятности простых, мы расскажем в следующей главе.

3. Основные правила теории вероятностей

В гл. 2 мы подчеркнули, что главное содержание теории вероятностей составляют не *прямые*, а *косвенные* методы, которые позволяют выражать вероятности одних событий через вероятности других, более простых. Здесь мы займёмся такими методами. Все они *стоят на двух китах*, т. е. основаны на двух главнейших принципах, *правилах* теории вероятностей, правилах сложения и умножения вероятностей.

Правило сложения вероятностей. *Вероятность того, что произойдёт одно из двух несовместных событий (всё равно какое) равна сумме вероятностей этих событий.*

Пусть A и B – два несовместных события. Тогда

$$P(A \text{ или } B) = P(A) + P(B). \quad (3.1)$$

Вы спросите: а откуда взялось это правило? Теорема это или аксиома? И то, и другое. Оно может быть строго доказано для опытов, сводящихся к схеме случаев. Действительно, если A и B – два несовместных события, то число случаев, благоприятных сложному событию *A или B,* равно $m_A + m_B$, значит

$$P(A \text{ или } B) = \frac{m_A + m_B}{n} = \frac{m_A}{n} + \frac{m_B}{n} = P(A) + P(B).$$

Поэтому часто правило сложения вероятностей называют *теоремой сложения*. Но нельзя забывать, что теоремой оно является только для схемы случаев. В опытах, не сводящихся к этой схеме, оно принимается без доказательства, как *принцип* или *аксиома*. Для частот (статистических вероятностей) правило сложения справедливо, в чём вы можете убедиться самостоятельно.

Правило сложения вероятностей легко обобщается на любое число событий:

Вероятность того, что произойдёт какое-либо из нескольких несовместных событий (всё равно, какое именно) равно сумме вероятностей этих событий:

$$P(A_1 \text{ или } A_2 \text{ или } \ldots \text{ или } A_n) = P(A_1) + P(A_2) + \ldots + P(A_n). \quad (3.2)$$

Из правила сложения вероятностей вытекают некоторые важные следствия. Во-первых,

Если события A_1, $A_2, \ldots, A_n$ несовместны и образуют полную группу, то сумма их вероятностей равна единице.

Попробуйте сами доказать это следствие. Во-вторых (следствие следствия), если A – какое-то событие, а $\bar{A}$ – противоположное ему (состоящее в не появлении A), то

$$P(A) + P(\bar{A}) = 1, \quad (3.3)$$

т. е. сумма вероятностей противоположных событий равна единице.

На формуле (3.3) основан очень распространённый в теории вероятностей приём перехода к *противоположному событию*. Часто бывает, что трудно вычислить вероятность интересующего нас события, а вероятность противоположного ему – легко.

Правило умножения вероятностей *Вероятность совмещения двух событий (т. е. совместного появления того и другого) равна вероятности одного из них, умноженной на вероятность другого, вычисленную при условии что первое произошло.*

В виде формулы это правило запишется так:

$$P(A \text{ и } B) = P(A) + P(B/A). \qquad (3.4)$$

Здесь $P(B/A)$ – так называемая *условная вероятность* события B, вычисленная при условии (в предположении), что событие A произошло.

Правило умножения тоже является теоремой и может быть строго доказано для схемы случаев, в других же условиях оно принимается без доказательства как *принцип* или *аксиома*. Для частот (статистических вероятностей) оно также справедливо.

При пользовании правилом умножения совершенно безразлично, какое событие считать *первым*, а какое *вторым* и правило умножения можно записать и в виде

$$P(A \text{ и } B) = P(B)P(A/B). \qquad (3.5)$$

Пример 1. В урне находятся 3 белых шара и 4 чёрных. Два шара вынимаются один за другим. Найти вероятность того, что оба они белые.

Решение. Рассмотрим два события: A – первый шар белый, B – второй шар белый. Требуется найти вероятность совмещения этих событий. Применяем правило умножения вероятностей (3.4).Очевидно, $P(A) = 3/7$. Вычислим $P(B/A)$. Если первый шар белый, то второй выбирается из 6 оставшихся шаров. Из них 2 белых, и потому $P(B/A) = 2/6 = 1/3$. Поэтому

$$P(A \text{ и } B) = 3/7 \cdot 1/3 = 1/7.$$

Этот пример мы уже решили непосредственным подсчётом числа случаев (пример 4 из гл. 1).

Теперь подумайте: изменится ли это решение, если шары вынимаются не последовательно, один за другим, а *сразу*. С первого взгляда может показаться: да, изменится. Но стОит немного подумать, и станет ясно: нет, не изменится.

В самом деле, пусть мы вынимаем два шара одновременно, но двумя руками. Условно назовём *первым* тот, который в правой руке, а *вторым* – тот, который в левой. Изменится ли ход наших рассуждений по сравнению с приведёнными в примере 1? Нисколько! Вероятность двух белых шаров будет всё та же, 1/7. *А если вынимаем одной рукой*, может придраться кто-нибудь. Тогда назовём первым тот, который ближе к большому пальцу, а вторым – тот, который ближе к мизинцу. *А если* … не унимается недоверчивый читатель. Предложим ему замолчать, ведь в душе он уже всё понял ...

Правило умножения становится особенно простым для особого рода событий, которые называются *независимыми*.

События A и B называются независимыми, если появление одного из них никак не влияет на вероятность появления другого. Условная вероятность события A в предположении, что B произошло, совершенно такая же, как и без этого предположения:

$$P(A/B) = P(A). \qquad (3.6)$$

В противном случае эти события называются *зависимыми*[3.1].

В нашем примере 1 события А и В были зависимыми: вероятность появления белого шара при втором тираже *зависела* от того, белым или чёрным был первый шар.

А теперь изменим условия опыта. Пусть после выемки первого шара его вкладывают обратно в урну и смешивают с остальными, после чего снова вынимают шар. В *таком* опыте события A – первый шар белый, B – второй шар белый будут уже независимы:

$$P(B/A) = P(B) = 3/7.$$

Понятие о зависимости и независимости событий очень важно. Неполное владение этим понятием нередко приводит к ошибкам, особенно у начинающих, склонных забывать о зависимости событий, когда она имеет место, и, обратно, приписывать какую-то зависимость событиям, которые на самом деле независимы.

Вот, спросим у кого-нибудь, неискушённого в теории вероятностей: мы подбросили монету 10 раз и каждый раз получили герб. Что вероятнее получить при 11-м броске? Почти наверное он ответит: *конечно, решку*! *Ведь герб уже появился 10 раз, должно же это когда-нибудь скомпенсироваться.* Скажет, и будет совершенно неправ. Вероятность появления герба при каждом очередном подбрасывании совершенно не зависит от того, сколько раз перед этим появился герб[3.2]. Конечно, если мы подбрасываем её так, как надо, например, кладём на ноготь большого пальца и подкидываем щелчком, и если монета *правильная*. Но появление герба 10 раз подряд может вызвать подозрение в правильности монеты и склонность считать, что вероятность появления герба вообще превышает 1/2.

Что-то не верится, может быть, скажете вы. *Не может быть, чтобы на 11-й раз не было вероятнее выпадение решки*. Так давайте поспорим. Пусть год назад вы подбросили монету 10 раз и все 10 раз выпал герб. Сегодня вы вспомнили этот любопытный результат и подбросили монету ещё раз. И вы по-прежнему думаете, что решка вероятнее, чем герб? Пожалуй, вы начали в этом сомневаться.

Добьём вас окончательно. Много лет назад статистик Карл Пирсон, подбросив монету 24 000 раз, в каких-то 10 случаях получил 10 гербов подряд[3.3]. Сегодня вы вспомнили об этом и решили продолжить этот опыт: подбросить монету ещё один раз. Что будет вероятнее, герб или решётка? Думаю, вы уже сдались и признали, что они равновероятны.

Распространим теперь правило умножения вероятностей на несколько событий $A_1, A_2, \ldots, A_n$. В общем случае, когда события зависимы, вероятность их совмещения вычисляется так. Вероятность одного события множится на вероятность другого при условии, что первое произошло, затем на условную вероятность третьего при условии, что оба предыдущих произошли и т. д. Записывать это правило в виде формулы мы не будем: его гораздо удобнее запомнить в словесном виде[3.4].

Пример 2. В урне 5 перенумерованных шаров. Все шары вынимаются один за другим. Найти вероятность того, что номера шаров будут следовать в возрастающем порядке.

Решение[3.5]. По правилу умножения вероятностей

$$P(1, 2, 3, 4, 5) = 1/5 \cdot 1/4 \cdot 1/3 \cdot 1/2 \cdot 1 = 1/120.$$

Пример 3. В ранце школьника лежат 8 букв разрезной азбуки: две буквы *а*, три буквы *к* и три буквы *м*. Мы вынимаем три карточки (?) одну за другой и кладём их на стол в порядке появления. Найти вероятность получения слова *мак*.

Решение. По правилу умножения вероятностей

$$P(\text{мак}) = 3/8 \cdot 2/7 \cdot 3/6 = 3/56.$$

Попробуем теперь решить тот же пример в изменённом виде. Найти вероятность того, что из вынутых букв можно составить слово *мак*. *Нет, дудки, меня не проведёшь*, может быть, подумаете вы. *Я-то знаю, что всё равно, вынимать ли карточки вместе или раздельно. Вероятность будет по-прежнему* 3/56.

Если вы на самом деле так подумали, то ошиблись. Изменены не только условия выемки букв, но и само событие. В примере 3 требовалось, чтобы буква *м* стояла именно на первом месте, *а* – на втором и *к* – на третьем. Теперь же порядок букв безразличен: будет ли это *мак* или *кам* или ещё какое-нибудь. Вот событие A, о котором теперь идёт речь:

A – из вынутых букв можно составить *мак.* Это событие распадается на несколько вариантов, а именно

A = (*мак* или *амк* или *кма* или …)

Таких вариантов будет столько, сколько имеется перестановок из трёх элементов: $P_3 = 1 \cdot 2 \cdot 3 = 6$.

Следует вычислить вероятности всех этих вариантов и сложить их в соответствии с правилом сложения. Но все эти вероятности одинаковы:

$$P(\text{амк}) = 2/8 \cdot 3/7 \cdot 3/6 = 3/56;\ P(\text{кам}) = 3/8 \cdot 2/7 \cdot 3/6 = 3/56 \text{ и т. д.}$$

Их сумма равна $3/56 \cdot 6 = 9/28$.

Особенно простой вид принимает правило умножения вероятностей для независимых событий[3.6]: надо перемножать не условные вероятности, а просто вероятности событий:

$$P(A_1 \text{ и } A_2 \text{ и } \dots \text{ и } A_n) = P(A_1)\, P(A_2)\ \dots \cdot P(A_n), \qquad (3.8)$$

т. е. вероятность совмещения независимых событий равна произведению их вероятностей.

Пример 4. Стрелок производит четыре выстрела по цели. Попадание или промах при каждом выстреле не зависит от остальных результатов (выстрелы независимы). Вероятность попадания при каждом выстреле равна 0,3. Найти вероятность того, что первые три раза стрелок промахнулся, но попал в цель при последнем выстреле.

Решение. Попадание обозначим значком +, промах, значком –. По правилу умножения вероятностей независимых событий

$$P(- - - +) = 0{,}7\ 0{,}7\ 0{,}7\ 0{,}3 = 0{,}1029.$$

А теперь решим пример чуточку посложнее.

Пример 5. Условия те же. Найти вероятность в точности двух попаданий.

Решение. Исследуемое событие A может осуществляться в нескольких вариантах. Подсчитаем их число. Оно равно числу способов, которыми можно из четырёх выстрелов выбрать два:

$$C_4^2 = \frac{4 \cdot 3}{1 \cdot 2} = \frac{4 \cdot 3}{1 \cdot 2} = 6.$$

Событие A может, стало быть, произойти в шести вариантах. Вероятность каждого из них равна

$$P = 0{,}3 \cdot 0{,}3 \cdot 0{,}7 \cdot 0{,}7 = 0{,}0441.$$

Складывая эти вероятности по правилу сложения, получим

$$0{,}0441 \cdot 6 = 0{,}2646 \approx 0{,}265.$$

Сформулируем общее правило, которым пользуются при решении задач: если требуется найти вероятность события, надо прежде всего выяснить, как оно может произойти, затем вычислить и сложить вероятности каждого варианта.

В следующем примере вероятности вариантов не будут одинаковыми.

Пример 6. Три стрелка производят по одному выстрелу по общей мишени. Вероятности попаданий равны 0,4, 0,5 и 0,7. Найти вероятность того, что число попаданий окажется равным двум.

Решение. Искомое событие A = (+ + – или + – + или – + +) распадается на $C_3^2 = C_3^1 = 3$ варианта с вероятностями соответственно

$$0{,}4 \cdot 0{,}5 \cdot 0{,}3;\ 0{,}4 \cdot 0{,}5 \cdot 0{,}7;\ 0{,}6 \cdot 0.5 \cdot 0{,}7,$$

равными 0,060; 0,140 и 0,210.

Их сумма равна $P(A) = 0{,}410$.

Вас, наверное, удивляет обилие примеров с *выстрелами* и *попаданиями* в цель. Но эти примеры опытов, не сводящихся к схеме случаев, столь же неизбежны и традиционны в теории вероятностей, как классические примеры с монетами, игральными картами и т. п. в схеме случаев. Последние не свидетельствуют о какой-либо особой склонности к азартным играм у лиц, занимающихся теорией вероятностей, а первое не является признаком их особой кровожадности. Всё дело в том, что такие примеры – самые простые. Вытерпите ещё один пример.

Пример 7. Условия те же, что в Примере 4. Найти вероятность события C, попадания в цель хотя бы один раз.

Решение. Как это событие может произойти? У него множество вариантов, например C = (+ + + + или + + + –). Можно найти вероятность каждого варианта и сложить эти вероятности, но это было бы глупо. Гораздо проще перейти от события C к противоположному: ни одного попадания. Появится только один вариант (– – – –) с вероятностью $0{,}7^4 \approx 0{,}240$. Вероятность искомого события C будет равна примерно $1 - 0{,}240 = 0{,}760$.

Если противоположное событие распадается на меньшее число вариантов, чем прямое, то его и следует вычислять.

Один из признаков, по которому можно почти безошибочно заключить, что стОит перейти к противоположному событию, это наличие в формулировке искомого события указание *хотя бы* или *по крайней мере*.

Пример 8. Собралось n незнакомых друг другу людей. Найти вероятность того события C, что хотя бы у двух из них совпадают дни и месяцы рождений.

Решение. Допустим, что рождения во все дни года равновероятны**[3.7]**. Формулировка *хотя бы* должна сразу насторожить вас: а не лучше ли перейти к противоположному событию? И впрямь событие C очень сложно и распадается на такое громадное число вариантов, что при одной мысли об их переборе мурашки начинают идти по коже. Противоположное событие гораздо скромнее, и его вероятность может быть найдена очень просто.

Выберем кого-то из собравшихся и условно назовём его *первым*. Ему можно родиться в любой день года, и вероятность этого равна единице. Произвольно выберем *второго*, который может родиться в любой день, кроме дня рождения *первого*. Вероятность этого равна 364/365. Для рождения третьему остаётся 363 дня и т. д. По правилу умножения вероятностей получаем вероятность противоположного события

$$P(\overline{C}) = 1 \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \ldots \cdot \frac{365-(n-1)}{365}, \qquad (9)$$

и искомая вероятность легко находится**[3.8]**.

Отметим любопытную особенность этой задачи: при увеличении n, даже довольно скромном, событие C быстро становится практически достоверным. Уже при $n = 50$ формула (9) даёт $P(\overline{C}) \approx 0{,}03$, $P(C) \approx 0{,}97$. С высоким уровнем доверия 0,97 событие C можно считать практически достоверным.

Если пожелаете, этот нехитрый расчёт может помочь вам представиться кудесником-предсказателем. Пусть где-то собралось много людей порядка 50 или немного больше. Вы берётесь утверждать, что среди присутствующих есть люди, дни рождения которых совпадают. Вы почти наверняка окажитесь правым.

А теперь решим одну важную общую задачу, которая часто встречается в самых различных формах.

Задача 1. Производится n независимых опытов, в каждом из которых событие A появляется с вероятностью p. Найти вероятность события C, появления в этих опытах события A хотя бы один раз.

Решение. Волшебное выражение *хотя бы* отсылает нас к противоположному событию (ни одного появления события C), которое происходит только в одном варианте: (– – – …). По правилу умножения вероятностей для независимых событий

$$P(\bar{C}) = (1 - p)^n,$$

$$P(C) = 1 - (1 - p)^n. \qquad (10)$$

Обратите внимание на общую формулу (10), потому что по ней решаются многие практически важные задачи.

Пример 9. Вероятность обнаружения космического объекта при одном цикле обзора радиолокационной станции равна p = 0,14; обнаружения при различных циклах независимы. Найти вероятность того, что после 10 циклов обзора объект будет обнаружен.

Решение. По формуле (10)

$$P(C) = 1 - (1 - 0{,}1)^{10} = 1 - 0{,}9^{10} = 1 - 0{,}348 = 0{,}652.$$

Пример 10. Техническое устройство состоит из семи узлов (элементов). Каждый из них может независимо от других оказаться неисправным с вероятностью 0,05. При неисправности хотя бы одного узла происходит авария (событие C). Какова её вероятность?

Решение. По формуле (10)

$$P(C) = 1 - (1 - 0{,}05)^7 = 1 - 0{,}95^7 \approx 1 - 0{,}695 = 0{,}305.$$

Вот так штука! Вероятность аварии превышает 30%. Аварийное положение! Нужно спешно повышать надёжность, т. е. вероятность исправного состояния каждого узла. Обратите внимание: неисправность каждого узла довольно маловероятна. Легкомысленно осмысливая этот факт можно *махнуть на него рукой* и объявить её практически невозможной; мы именно так и поступали при предсказании результатов подбрасывания монеты. Но сейчас совсем другое дело. Узлов у нас не один, а семь, и отказ *хотя бы одного из них* совсем не маловероятен. А последствия нашего легкомыслия (авария!) отнюдь не так невинны, как неуспешное предсказание. Иной раз расчёты с помощью теории вероятностей приводят к неожиданным результатам, как бы противоречащим здравому смыслу. Вот пример, по форме шуточный.

Пример 12. Охотники Семён и Жорик отправились на охоту, увидели медведя и одновременно выстрелили по нему. Медведь был убит, и в его шкуре обнаружилась одна пробоина, неизвестно от чьего выстрела. Более правдоподобно, что Семёну. Он охотник старый, опытный, и в цель размером с медведя попадает с того расстояния, с какого был сделан выстрел, с вероятностью 0,8.

Жорик – молодой и менее опытный охотник, и для него та же вероятность составляет всего 0,4. Шкуру медведя продали за 50 руб. Как справедливо разделить эту сумму?

Решение. Вам, вероятно, захотелось разделить эти деньги пропорционально вероятностям 0,8 и 0,4, т .е. дать Семёну 2/3 суммы, а Жорику – 1/3. И, вообразите себе, вы были неправы.

Изменим чуточку условия задачи. Пусть Семён попадает в медведя сразу же с вероятностью 1, а Жорик – только с вероятностью 0,5. Кому принадлежит пробоина (и шкура)? Вы разделили бы в том же соотношении 2:1. Что-то неладно в ваших рассуждениях, но что?

Вы делили сумму пропорционально вероятностям попадания при одном выстреле, но один стрелок попал, другой промахнулся. И этого промаха вы не учли. Событие A, одно попадание, могло произойти двумя способами:

A_1: Семён попал, Жорик промахнулся

A_2: Жорик попал, Семён промахнулся

Вероятности этих вариантов найдём по правилу умножения:

$$P(A_1) = 0{,}8 \cdot 0{,}6 = 0{,}48;\ P(A_2) = 0{,}4 \cdot 0{,}2 = 0{,}08.$$

Вот пропорционально этим вероятностям и надо по справедливости разделить полученную сумму.

4. Случайные величины

Здесь вы ознакомитесь с новым и очень важным понятием *случайной величины*. В первой главе вы познакомились с простейшим способом вычисления вероятности события – непосредственным подсчётом доли благоприятных случаев. Только-только вы освоились с этим способом, как вас постигло разочарование. Оказалось, что этот подсчёт применим лишь в тех сравнительно редких задачах, где опыт сводится к схеме случаев, т. е. обладает симметрией возможных исходов.

Зато в следующей главе вы научились приближённо находить вероятность события по его частоте. Никаких ограничений на опыт уже не накладывалось. Только вы немного освоились и с этим способом, как вас снова постигло разочарование. Оказалось, что не он является основным в теории вероятностей. В третьей главе вы, наконец-то, дошли до методов, которые являются в этой науке основными.

Наконец-то, думаете вы*, удалось мне добраться до самых основ, до самой сути. Больше разочарований не будет.*

Увы! Вас ждёт ещё одно, на этот раз последнее, разочарование. В современной теории вероятностей тот аппарат событий, с которым вы имели дело, не является основным.

На кой чёрт вы заставляли меня знакомиться с не основным аппаратом?

Без того, что вы уже изучили, невозможно даже подступиться к современному аппарату, аппарату случайных величин. Эта глава и будет посвящена основному понятию современной теории вероятностей, понятию случайной величины, разновидностям случайных величин, способам их описания и действий с ними. О них мы будем говорить не так подробно, как о событиях, а более в *описательном* духе. Здесь математический арсенал оказался бы более сложным и мог бы оттолкнуть начинающего. А мы как раз хотим облегчить начинающему его *первые шаги*.

Итак, случайной величиной называется величина, которая в результате опыта может принять то или иное значение, неизвестно заранее – какое именно[4.1].

Как всегда в теории вероятностей определение смутноватое, упоминает о какой-то неопределённости, неизвестности … Чтобы овладеть понятием случайной величины, надо к нему попросту привыкнуть. Для этого рассмотрим примеры случайных величин.

1. Подбрасываем две монеты. Число появившихся гербов – случайная величина (СВ), её возможные значения: 0, 1, 2. Какое из них она примет – заранее неизвестно.

2. Ученик сдаёт экзамен. Его отметка –СВ с возможными значениями 2, 3, 4, 5.

3. В группе студентов 28 человек. Число неявившихся на занятия в некоторый день – СВ с возможными значениями 0, 1, …, 28. Но ведь все 28 *никак заболеть* (*прогулять*) *не могут*! Да, это практически невозможное событие. Но кто вам сказал, что все значения случайной величины равновероятны?

Все примеры относились к так называемым *дискретным* СВ. Они дискретны, если их возможные значения отделены друг от друга какими-то интервалами. На оси абсцисс эти значения изобразятся *отдельными* точками.

Бывают СВ другого типа, *непрерывные*. Их значения сплошь заполняют какой-то участок числовой оси. Границы этого участка иногда чётки, а иногда расплывчаты, неопределённы. Рассмотрим примеры.

1. Промежуток времени между двумя отказами (сбоями) вычислительной машины. Значения этой СВ сплошь заполняют какой-то участок числовой оси. Его нижняя граница вполне чётка (нуль), а верхняя – расплывчата, неопределённа. Она может быть найдена только в результате опыта.

2. Вес товарного поезда, который подаётся на станцию для разгрузки.

3. Высота подъёма воды в половодье.

4. Ошибка взвешивания тела на аналитических весах. Эта СВ, в отличие от предыдущих, может принимать и отрицательные значения.

5. Удельный вес пробы молока.

6. Время, проводимое учеником восьмого класса, у телевизора в течение дня.

Чтобы говорить о СВ в смысле, который ей приписывается в теории вероятностей, необходимо уточнение: *в чём состоит опыт*, при котором она принимает то или иное значение. Так, в примере 1 о непрерывной СВ следует указать тип машины, её возраст и условия применения. В примере 4 следует уточнить, на каких весах производится взвешивание, какими разновесками. Эти подробности мы не всегда будем оговаривать.

Далее, все СВ, которые мы назвали непрерывными, могут быть измерены только в каких-то единицах (минутах, сантиметрах, тоннах) и, строго говоря, являются дискретными. Так, СВ, рост человека, нет смысла измерять точнее, чем до сантиметра, и она в сущности оказывается дискретной. Но если таких значений очень много и расположены они очень *тесно*, удобнее рассматривать такую величину как непрерывную.

СВ будем обозначать заглавными буквами латинского алфавита, а их возможные значения – теми же строчными буквами (X и x_1, x_2, …). Естественно, не все эти значения равновероятны.

Законом распределения СВ называется такая функция, которая описывает его, т. е. описывает распределение вероятностей между её значениями. Мы познакомим вас только с некоторыми, самыми простыми формами таких законов.

Закон распределения дискретной СВ может быть записан в виде *ряда распределения*, т. е. таблицы из двух строк: строки её возможных значений и строки соответствующих вероятностей. Сумма всех этих вероятностей очевидно равна единице.

Пример 1. Производится три независимых выстрела по мишени, вероятность попадания при каждом из них равна $p = 0{,}4$. Дискретная СВ X – число попаданий. Требуется построить её ряд распределения.

Решение. Возможные значения X: 0, 1, 2, 3, соответствующие вероятности p_0, p_1, p_2, p_3. Вычислять их мы умеем (гл. 3):

$$p_0 = P(-\ -\ -) = 0{,}6^3 = 0{,}216$$
$$p_1 = P[(+\ -\ -) \text{ или } (-\ -\ +) \text{ или } (-\ +\ -)] = 3\cdot 0{,}6^2\cdot 0{,}4 = 0{,}432$$
$$p_2 = P[(+\ +\ -) \text{ или } (+\ -\ +) \text{ или } (-\ +\ +)] = 3\cdot 0{,}4^2\cdot 0{,}6 = 0{,}288$$
$$p_3 = P(+\ +\ +) = 0{,}4^3 = 0{,}064$$

Сумма вероятностей действительно равна единице.

Пример 2. Спортсмен пытается забросить мяч в кольцо. При каждой независимой попытке вероятность успеха равна p. Сколько попыток придётся сделать? Их число – дискретная СВ, X. Найти её ряд распределения.

Решение. Возможные значения X: 1, 2, …, теоретически – до бесконечности. Их вероятности: вероятность успеха при первой попытке $p_1 = p$. Для успеха при двух попытках должны совместиться два события: промах при первой попытке, вероятность $(1 - p)$, успех при второй попытке, вероятность p. Вероятность совмещения $p(1 - p)$. Если удалась только третья попытка, то $p_3 = p(1 - p)^2$, и вообще $p_i = p(1 - p)^{i-1}$. Вероятности p_i образуют геометрическую прогрессию со знаменателем $(1 - p)$, и поэтому указанное распределение вероятностей называется геометрическим[4.2].

Но чем можно характеризовать распределение вероятностей для непрерывной СВ? Ряда распределения для неё построить нельзя, нельзя даже записать её возможные значения. Какую бы пару значений мы бы не поставили рядом, между ними непременно найдутся другие значения[4.3]. И встретится ещё одна трудность: вероятность каждого отдельного значения непрерывной СВ равна нулю! Нет, мы не обмолвились.

Представьте себе пляж, усыпанной галькой. Вес отдельного камня – СВ X. Взвешиваем камни с точностью в 1 *г*. Получим какую-то частоту, допустим, веса 30 *г*. М её не знаем, да это и неважно. Теперь взвешиваем с точностью 0,1 *г*. Некоторые камни, вес которых считался равным 30 *г, отпадут* и прежняя частота события $X = 30$ *г* станет примерно в 10 раз меньше. Теперь взвешиваем с точностью в 1 *мг*, та же частота уменьшится ещё в сто раз …

Частота – родная сестра вероятности и приближается к ней при большом числе опытов, камней же на пляже достаточно. Так какую же вероятность следует приписать событию, весу камня, равному в точности 30 *г*? Нулевую, деться некуда.

Вы недоумеваете, быть может, возмущаетесь: ведь нулевой вероятностью обладают *невозможные события*, это же событие возможно. Да, действительно, *вероятность невозможного события равна нулю*, но мы не утверждали, что любое событие с нулевой вероятностью невозможно. И нам пришлось познакомиться с возможными событиями, вероятности которых равны нулю.

Поразмышляем немного. Забудьте на минуту теорию вероятностей и представьте себе некоторую фигуру площади *S* на плоскости и произвольную точку внутри фигуры. Площадь этой точки равна нулю[4.4], но фигура с ненулевой площадью состоит из точек! К этому парадоксу вы привыкли, и так же надо привыкнуть к тому, что вероятность попадания в каждую отдельную точку для непрерывной СВ в точности равна нулю[4.5].

Но как же тогда можно говорить о распределении вероятностей для таких СВ? ведь все её вероятности равны нулю.

Говорить о распределении вероятностей между *отдельными значениями* непрерывной СВ не имеет смысла, но распределения существуют и для них. Так, рост человека 170 *см* вероятнее роста 210 *см*, хотя оба эти значения возможны.

Введём новое важное понятие: *плотность вероятности*. В физике плотностью вещества называется вес единицы его объёма [прежний термин, который автор применила выше, *удельный вес*]. А если вещество неоднородно? Приходится рассматривать его *местную* плотность. И в теории вероятностей рассматривается местная плотность, т. е. вероятность единицы длины в данной точке *x*.

Плотностью вероятности непрерывной СВ называется предел отношения вероятности попадания этой СВ на малый участок, примыкающий к точке x, к длине этого участка, когда последняя стремится к нулю.

Понятие *плотность вероятности* легко выводится из родственного ему понятия *плотность частоты*. Рассмотрим непрерывную СВ, например, рост человека или вес камня на пляже. Произведём над ней ряд опытов, в каждом из которых она принимает какое-то значение Разобьём весь диапазон значений СВ на какие-то части (*разряды*), например. 150 – 155. 155 – 160, …, 195 – 200. Подсчитаем, сколько значений СВ попало в каждую[4.6], разделим на общее число опытов, получим частоту части (разряда). Сумма частот должна быть равна единице. Разделим частоту на длину разряда; эти длины могут быть и различными.

Если массив статистических данных достаточно большой (скажем, несколько сотен, лучше – ещё больше), то, построив плотность частоты, можем получить достаточное представление о распределении этой СВ, о её плотности вероятности. При обработке такого материала удобно прежде всего построить специальный график, гистограмму. На каждом разряде как на основании строится прямоугольник, площадь которого равна частоте разряда (а, значит, высота – плотности частоты). Площадь, ограниченная гистограммой, очевидно равна единице.

По мере увеличения числа опытов можно брать разряды всё меньшими и меньшими. Ступенчатый характер гистограммы будет сглаживаться, и она начнёт приближаться к некоторой плавной кривой, *кривой распределения*. По оси ординат будут откладываться уже не плотность частоты, а плотность вероятности. Полная площадь, ограниченная кривой распределения, как и площадь её *сестры*, гистограммы, будет равна единице.

Вероятность попадания СВ на участок от a до b будет равна площади, опирающейся на него. Обозначим плотность вероятности $f(x)$, тогда вероятность попадания этой СВ на указанный участок выразится интегралом

$$P(a,b)=\int_a^b f(x)dx. \qquad (4.1)$$

И если не пожалеть времени и средств, плотность вероятности может быть определена как угодно точно по опытным данным. Но стОит ли овчинка выделки? Так ли уж нам нужно *совершенно точно* знать плотность? Очень часто этого не нужно, достаточно иметь лишь приближённое представление о законе распределения СВ, ведь все вероятностные расчёты по своей природе имеют приближённый, *прикидочный* характер.

Чтобы знать приближённо, *примерно*, каков закон распределения СВ, вовсе не нужно какого-то огромного числа опытов, можно ограничиться довольно скромным числом 300 – 400, а иной раз и меньшим. Построив гистограмму, можно затем *выровнять* её той или иной плавной кривой (ограничивающей единичную площадь).

Теория вероятностей располагает целым набором кривых, удовлетворяющих этому условию. Некоторые из них обладают известными преимуществами перед другими. Например, интеграл (4.1) для них легко вычисляется или для него составлены таблицы. Или же условия возникновения нашей СВ подсказывают определённый тип распределения, вытекающий из теоретических соображений. Вдаваться здесь в такие подробности мы не будем, это дело специальное. Но подчеркнём, что при нахождении законов распределения СВ основное значение имеют не *прямые*, а *косвенные* методы, позволяющие находить распределения одних CD не непосредственно из опыта, а через имеющиеся данные о других СВ, связанных с ними.

В таких косвенных методах большую роль играет аппарат так называемых *числовых характеристик* СВ. Это – некоторые числа, характеризующие те или другие свойства, отличительные признаки СВ, например, среднее значение, вокруг которого происходит случайный разброс, степень этого разброса (как бы *степень случайности* этой СВ) и ряд других признаков. Многие задачи теории вероятностей можно решать, не прибегая, или почти не прибегая, к законам распределения, а пользуясь только числовыми характеристиками. Мы познакомим вас только с двумя, но зато важнейшими характеристиками, *ожиданием*[4.7] и *дисперсией*.

Ожиданием EX дискретной СВ X называется сумма произведений всех её возможных значений на соответствующие вероятности

$$\mathrm{E}X = x_1p_1 + x_2p_2 + \ldots + x_np_n = \sum_{i=1}^{n} x_i p_i, \qquad (4.2;\ 4.3)$$

где $x_1, x_2, \ldots, x_n$ – возможные значения СВ X и p_i – вероятность значения x_i.

Как видно из формулы (4.3), ожидание СВ X есть не что иное, как *среднее взвешенное* всех её возможных значений. причём *веса* равны соответствующим вероятностям. Если число возможных значений СВ, как в примере 2, бесконечно, то сумма (4.3) состоит из бесконечного числа слагаемых.

Ожидание или среднее значение СВ является как бы её *представителем*, которым можно её заменить при грубо ориентировочных расчётах. В сущности, мы так всегда и поступаем, когда в каких-нибудь задачах не учитываем случайности.

Пример 3. Найти ожидание X, рассмотренной в примере 1.

Решение. По формуле (4.3)

$$\mathrm{E}X = 0\cdot 0{,}216 + 1\cdot 0{,}432 + 2\cdot 0{,}288 + 3\cdot 0{,}064 = 1{,}2.$$

Понятие ожидания вводится также и для непрерывной СВ X, но, естественно, в формуле (4.3) сумма заменяется интегралом:

$$\mathrm{E}X = \int_{-\infty}^{\infty} x f(x)\,dx. \qquad (4.4)$$

Здесь $f(x)$ – плотность вероятности непрерывной СВ X.

Поговорим немного об ожидании, его физическом смысле и *генеалогии*. Аналогично тому, как у вероятности есть *родная сестра* – частота, у ожидания есть родной брат (родная сестра? родственник?) – среднее арифметическое результатов наблюдений. Так же, как частота приближается к вероятности при большом числе опытов, с увеличением того же числа среднее арифметическое значений СВ приближается к её ожиданию.

Докажем это для дискретных СВ; думаем, что для непрерывных СВ вы охотно примете это на веру. Итак, пусть имеется дискретная СВ X с рядом распределений

$$x_1,\ x_2,\ \ldots,\ x_n;\ p_1,\ p_2,\ \ldots,\ p_n.$$

Пусть произведено N опытов, в которых x_1 появилось M_1 раз, x_2 появилось M_2 раз и т. д. Среднее арифметическое значений X равно

$$\bar{X} = \frac{x_1M_1 + x_2M_2 + \ldots + x_nM_n}{N} = \sum \frac{x_iM_i}{N}.$$

Но M_i/N есть не что иное, как частота события $X = x_i$: $M_i/N = p^*_i$. Следовательно,

$$\bar{X} = \sum x_i p^*_i .$$

Мы знаем, что при увеличении числа опытов N частота события p^*_i с практической достоверностью приближается к его вероятности p_i. Значит,

При увеличении числа опытов среднее арифметическое наблюдённых значений случайной величины с практической достоверностью будет сколь угодно близко к её ожиданию.

Это положение представляет собой одну из форм *закона больших чисел* (так называемую теорему Чебышёва), играющую очень большую роль в практических приложениях теории вероятностей. Действительно, так же, как неизвестную вероятность события можно приближённо определить по его частоте в длинной серии опытов, можно приближённо найти ожидание СВ X как среднее арифметическое её наблюдённых значений:

$$\mathrm{EX} \approx \bar{X} . \qquad (4.5)$$

Для того, чтобы вычислить по опытным данным ожидание интересующей нас СВ, вовсе нет надобности знать её закон распределения, просто надо вычислить среднее из всех результатов наблюдений.

Замечание: для определения ожидания СВ с удовлетворительной точностью совсем не нужно столько опытов (порядка нескольких сот), как для построения гистограммы, достаточно значительно меньшего числа, порядка десятков.

Введём теперь вторую важнейшую числовую характеристику СВ, её *дисперсию*. В переводе это слово означает *рассеивание*, и дисперсия СВ действительно характеризует разброс (рассеивание) её значений вокруг среднего. Чем больше дисперсия, тем *случайнее* СВ.

Дисперсия дискретной СВ X вычисляется так: из каждого её возможного значения вычитается среднее (ожидание). Полученное отклонение возводится в квадрат, множится на вероятность соответствующего значения, и все такие произведения суммируются:

$$DX = \sum_{i=1}^{n} (x_i - \mathrm{E}x_i)^2 p_i. \tag{4.6}$$

Может возникнуть вопрос: для чего потребовалось возводить отклонение от среднего в квадрат? Для того, чтобы избавиться от знака (плюс или минус). Можно было бы просто отбросить знак и взять абсолютную величину отклонения, но полученная таким образом характеристика гораздо менее удобна для вычисления и действий с ней, чем дисперсия.

Пример 4. Найти дисперсию числа попаданий X в условиях примера 1.

Решение. По формуле (4.6) имеем

$$\mathrm{D}X = (0 - 1{,}2)^2 \cdot 0{,}216 + (1 - 1{,}2)^2 \cdot 0{,}432 +$$
$$(2 - 1{,}2)^2 \cdot 0{,}288 + (3 - 1{,}2)^2 \cdot 0{,}064 = 0{,}72.$$

Формула (4.6) не самая удобная для вычисления дисперсии. Её можно и обычно удобно вычислять по формуле

$$\mathrm{D}X = \mathrm{E}(X^2) - (\mathrm{E}X)^2. \tag{4.7}$$

Дисперсия СВ равна ожиданию её квадрата без квадрата её ожидания.

Эту формулу легко вывести из (4.6) с помощью тождественных преобразований, но мы не задерживаемся на этом. Лучше проверим её справедливость на предыдущем примере:

$$\mathrm{D}X = 0^2\cdot 0{,}216 + 1^2\cdot 0{,}432 + 2^2\cdot 0{,}288 + 3^2\cdot 0{,}064 - (1{,}2)^2 = 0{,}72.$$

Для непрерывных СВ дисперсия вычисляется по формуле, аналогичной (4.6), но сумма, естественно, заменяется интегралом:

$$DX = \int_{-\infty}^{\infty} (x - \mathrm{E}x)^2 f(x)dx. \tag{4.8}$$

Обычно удобнее находить её по формуле (4.7), которая в непрерывном случае имеет вид

$$DX = \int_{-\infty}^{\infty} x^2 f(x) - (\mathrm{E}x)^2 dx. \tag{4.9}$$

Так же, как для приближённого нахождения ожидания нет надобности обязательно знать закон распределения, можно приближённо найти и дисперсию непосредственно по результатам опытов, осредняя квадраты отклонений наблюдённых значений СВ от их среднего арифметического:

$$DX \approx \frac{1}{N}\sum_{k=1}^{N}(x_k - \bar{X})^2. \tag{4.10}$$

Здесь k – номер опыта x_k, значение СВ в этом опыте и N, число опытов.

Опять-таки удобнее вычислять дисперсию как *средний квадрат минус квадрат среднего*:

$$DX \approx \frac{1}{N}\sum_{k=1}^{N}x_k^2 - \bar{X}^2. \tag{4.11}$$

Формулы (4.10; 4.11) пригодны для грубой приближённой оценки дисперсии и при не очень большом числе опытов. Лучше что-нибудь, чем ничего! В этом случае в математической статистике принято вводить *поправку за малое число опытов*, умножая результат на поправочный множитель $N/(N-1)$. Но не следует придавать этой поправке слишком большого значения, потому что при малом числе опытов всё равно из их обработки ничего хорошего получить нельзя, а при большом N поправочный множитель близок к единице[4.8].

Дисперсия СВ как характеристика рассеивания имеет неприятную особенность: её размерность, как видно из формулы (4.6), равна квадрату размерности СВ X. Если СВ X выражена, скажем, в минутах, то её дисперсия – в *квадратных минутах*, что не очень наглядно. Чтобы избежать этого, из дисперсии извлекают квадратный корень, получая новую характеристику рассеивания, так называемое *среднее квадратическое отклонение* или *стандарт*:

$$\sigma_x = \sqrt{DX}. \qquad (4.12)$$

Это очень наглядная и удобная характеристика рассеивания. Она сразу же даёт понятие о размахе колебаний СВ около среднего значения. Для большинства встречающихся на практике СВ

С практической достоверностью можно утверждать, что они не отклоняются от своего ожидания больше, чем на $3\sigma_x$.

Уровень доверия зависит от закона распределения СВ, но во всех не искусственно придуманных случаях он довольно высок. Приведённое выше правило называется *правилом трёх сигма*. И если нам тем или иным способом удалось найти две числовые характеристики СВ, её ожидание и среднее квадратическое отклонение, мы сразу же получаем ориентировочное представление о том, в каких пределах могут лежать её возможные значения.

Вы можете спросить: *если мы находили эти характеристики из опыта, кто мог бы помешать из того же опыта найти и пределы возможных значений*? Да, верно, если эти характеристики находятся непосредственно из опыта. Но не эти прямые методы нахождения числовых характеристик являются в теории вероятностей основными. Снова, в который уже раз мы скажем, что основными являются косвенные методы, позволяющие находить числовые характеристики интересующих нас СВ по числовым характеристикам других СВ, с ними связанных. При этом применяются *основные правила* действий с числовыми характеристиками. Некоторые из них, разумеется, без доказательства, мы приведём.

1. *Сложение ожиданий*. Ожидание суммы СВ равно сумме ожиданий слагаемых.

2. *Сложение дисперсий*. Дисперсия суммы независимых СВ равна сумме дисперсий слагаемых.

3. *Вынесение неслучайного множителя из-под знака ожидания*. Неслучайный множитель c можно выносить из-под знака ожидания: $\mathrm{E}(cX) = c\mathrm{E}X$.

4. *Вынесение неслучайного множителя из-под знака дисперсии*. Неслучайный множитель c можно выносить из-под знака дисперсии, возводя его в квадрат: $\mathrm{D}(cX) = c^2\mathrm{D}X$.

Эти правила, кроме, быть может, последнего, представляются довольно естественными. Чтобы убедить вас в его справедливости, приведём такой пример. Мы удвоили СВ X. Её ожидание, естественно, тоже удвоилось. Отклонение отдельного значения от среднего тоже удвоилось, а квадрат этого отклонения учетверился!

Уже такого небольшого набора правил достаточно, чтобы решать некоторые интересные задачи.

Задача 1. Производится N независимых опытов, в каждом из которых событие A появляется с вероятностью p. Рассматривается СВ X, число опытов, в которых появится A или число его появлений. Найти ожидание и дисперсию X.

Решение. Представим X в виде суммы N СВ

$$X = X_1 + X_2 + \ldots + X_N. \qquad (4.13)$$

Здесь $X_k = 1$, если событие A произошло в k-м опыте и $X_k = 0$ в противном случае. По правилу сложения ожиданий

$$\mathrm{E}X = \mathrm{E}X_1 + \mathrm{E}X_2 + \ldots + \mathrm{E}X_N. \qquad (4.14)$$

Опыты независимы и СВ в формуле (4.14) также независимы. По правилу сложения дисперсий

$$\mathrm{D}X = \mathrm{D}X_1 + \mathrm{D}X_2 + \ldots + \mathrm{D}X_N. \qquad (4.15)$$

Найдём теперь ожидание и дисперсию каждой X_k. Любая из них дискретна с возможными значениями 0 и 1 с вероятностями $(1 - p)$ и p. Поэтому

$$\mathrm{E}X_k = 0\cdot(1 - p) + 1\cdot p = p.$$

Дисперсию этой СВ найдём по формуле (4.7):

$$\mathrm{D}X_k = 0^2\cdot(1 - p) + 1^2\cdot p - (\mathrm{E}X)^2 = p - p^2 = p(1 - p).$$

С учётом формул (4.14) и (4.15) мы получим

$$\mathrm{E}X = Np,\ \mathrm{D}X = Np(1 - p).$$

Задача 2. Производится N независимых опытов, в каждом из которых событие A происходит с вероятностью p. Рассматривается СВ P^*, частота события A. Найти приближённо диапазон практически возможных значений этой СВ.

Решение. По определению, частота равна числу X появления этого события, делённому на число опытов N, т. е. $P^* = X/N$. По правилам 3 и 4 числовые характеристики этой СВ равны

$\mathrm{E}P^* = \mathrm{E}(X/N) = (1/N)\mathrm{E}X = Np/N = p$;
$\mathrm{D}P^* = \mathrm{D}(X/N) = (1/N^2)\mathrm{D}X = Np(1-p)/N^2 = p(1-p)/N$.
$\sigma_{P*} = \sqrt{p(1-p)/N}$.

По правилу трёх сигма искомый диапазон равен $p \pm 3\sigma_{P*}$. *Так ведь это – наша старая знакомая. Эту самую формулу нам демонстрировали для доверительного интервала, в который с вероятностью 0,997 уложится значение частоты события при большом числе опытов. Наряду с этим, и даже предпочтительно, нам рекомендовали другую, с двойкой, а не тройкой перед корнем, которая выполнялась с вероятностью 0,95. Но почему эта вероятность*?

Нужно познакомиться с очень важным законом распределения, так называемым *нормальным*. Рассмотрим СВ X. Она распределена по нормальному закону, если плотность её вероятности выражена формулой

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp[-\frac{(x-m)^2}{2\sigma^2}]. \qquad (4.16)$$

Кривая нормального закона симметрична, имеет колоколообразный вид и достигает максимума в точке m. С удалением от m плотность вероятности падает и в пределе стремится к нулю. Закон (4.16) зависит от двух параметров[4.9], m и σ. Первый из них есть не что иное, как ожидание X, а второй – её среднее квадратическое отклонение.

При изменении m кривая, не меняя своей формы, *ездит* туда-сюда вдоль оси абсцисс. Если же менять σ, кривая будет менять форму, *распластываться* при возрастании σ, а при его уменьшении вытягиваться вверх, становиться иглообразной.

Особая роль, которую играет нормальный закон в теории вероятностей. связана с его замечательным свойством: если складывать большое число независимых или слабо зависимых СВ, сравнимых друг с другом по порядку своих дисперсий, то,

Каковы бы не были законы распределения слагаемых, закон распределения суммы будет близок к нормальному,
тем ближе, чем больше было слагаемых.

Это – грубая формулировка так называемой центральной предельной теоремы (ЦПТ), играющей очень большую роль в теории вероятностей. У этой теоремы много различных форм, различающихся друг от друга условиями, которым должны удовлетворять СВ.

Очень многие СВ образуются по принципу суммы и поэтому распределены нормально или почти нормально. Так, ошибки всевозможных измерений, представляющие собой сумму многих *элементарных* и практически независимых ошибок, вызванными отдельными причинами[4.10]. По тому же закону распределяются как правило ошибки стрельбы, наведения, совмещения. Отклонения напряжения в сети от номинала также вызваны суммарным действием многих независимых причин. Нормальному или близкому к нему закону подчиняются такие СВ, как суммарная выплата страхового общества за большой период, суммарное время простоя компьютера за год и т. д. Покажем, что такая интересная СВ как *частота события* при большом числе опытов N тоже распределена приближённо нормально. Действительно,

$$P^* = (X_1 + X_2 + \ldots + X_N)/N$$

где X_k – *СВ,* равная единице если в k-м опыте событие A появилось и нулю в противном случае. Частота P^* при большом N есть сумма большого числа независимых слагаемых, каждое из которых имеет одну и ту же дисперсию

$$\mathrm{D}(X_k/N) = (1/N^2)p(1 - p).$$

Поэтому частота события A распределена примерно нормально.

Нередко приходится вычислять вероятность попадания СВ X, распределённой нормально, в пределы участка (a, b). Интеграл (4.16) не выражается через элементарные функции (*не берётся*) и для его вычисления пользуются таблицами специальной функции, так называемой функции *Лапласа*[4.11]

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x \exp(-t^2/2)dt.$$

При $x \geq 4$ с точностью до четвёртого знака после запятой, можно принимать что $\Phi(x) = 0{,}5000$. Следует учесть, что $\Phi(x)$ является нечётной функцией, $\Phi(-x) = -\Phi(x)$.

Вероятность попадания СВ, имеющей нормальное распределение с параметрами a и σ, в участок (a, b) выражается формулой

$$P(a, b) = \Phi[(b - m)/\sigma] - \Phi[(a - m)/\sigma]. \qquad (4.17)$$

Пример 5. Найти вероятность того, что СВ X, имеющей нормальное распределение с параметрами m и σ, отклонится от него не больше, чем на 2σ; на 3σ.

Решение. По формуле (4.17) и таблицей [выпущена]

$$P(m-2\sigma,\ m+2\sigma)=\Phi[\frac{m+2\sigma-m}{\sigma}]-\Phi[\frac{m-2\sigma-m}{\sigma}]=$$

$$\Phi(2)-\Phi(-2)=\Phi(2)+\Phi(2)=[\ldots]\approx 0{,}95.$$

Второй вопрос решается аналогично. Ответ: вероятность приближённо равна 0,997.

Так вот они, эти доверительные вероятности из гл. 2. Теперь мы можем решать некоторые поучительные задачи.

Пример 6. Поезд состоит из N = 100 вагонов. Вес каждого вагона, СВ с ожиданием m_q = 65 тонн и средним квадратическим отклонением σ_q = 9 тонн. Локомотив может вести поезд весом не более 6600 тонн, в противном же случае приходится прицеплять второй локомотив. Найти вероятность того, что этого делать не придётся.

Решение. Вес поезда X можно представить как сумму весов отдельных вагонов с одним и тем же ожиданием и одной и той же дисперсией σ_q^2 = 81. По правилу сложения ожиданий $\mathrm{E}X = 100\cdot 65 = 6500$, а по правилу сложения дисперсий $\mathrm{D}X = 100\cdot 81 = 8100$, а среднее квадратическое отклонение оказывается равным $\sigma_X = 90$.

X можно считать нормально распределённым. По формуле (4.17)

$$P(0,\ 6600) = \Phi[(6600-6500)/90] - \Phi[(0-6500)/90] = [\ldots] \approx 0{,}887.$$

Итак, одним локомотивом можно обойтись с вероятностью 0,887. Уменьшим на два число вагонов в поезде. Подсчитайте сами вероятность того же результата. Вас удивит, что эта вероятность приближённо равна 0,99, т. е. что исследуемое событие практически достоверно. И всего-то пришлось убрать два вагона.

Вот какие любопытные задачки можно решать, если иметь дело с суммами большого числа СВ. Естественно возникает вопрос: сколько это *много*? Сколько нужно сложить СВ, чтобы закон распределения суммы уже *нормализовался*? Это зависит от того, каковы законы распределения слагаемых. Бывают такие замысловатые законы, что требуется очень, очень много слагаемых. *Чего только не придумают математики*! Но природа не устраивает нам таких пакостей специально. Обычно для того, чтобы пользоваться нормальным законом бывает достаточно

5 – 6, ну 10; ну, от силы 20 слагаемых, особенно если они имеют одно и то же распределение.

Быстроту, с которой нормализуется закон распределения суммы одинаково распределённых СВ, можно иллюстрировать на примере. Тут вам придётся верить нам на слово, ничего не поделаешь. Но до сих пор мы вас не обманывали.

Пусть имеется непрерывная СВ, распределённая с постоянной плотностью на участке (0, 1). Кривая распределения имеет вид прямоугольника. Казалось бы, уж как это непохоже на нормальный закон! Но сложим две такие (независимые) случайные величины, получим новую СВ, распределённую по так называемому закону Симпсона[4.12]. Тоже *непохоже* на нормальный закон, но уже лучше. А если сложить три такие *равномерно* распределённые СВ, то получится СВ, кривая распределения которой состоит из трёх дуг парабол и по виду страшно напоминает нормальный закон. Сложим шесть – и никто не сможет отличить полученную кривую распределения от нормальной.

На подобном процессе основан широко применяемый метод получения нормально распределённой СВ при моделировании случайных явлений на компьютере. Достаточно сложить шесть независимых СВ, равномерно распределённых на отрезке (0, 1). Кстати, датчиками подобных случайных чисел (!) оснащено в настоящее время большинство компьютеров.

И всё-таки нельзя слишком увлекаться и сразу провозглашать нормальным закон распределения суммы нескольких СВ. Нужно сначала посмотреть, хотя бы в первом приближении, каковы их распределения. Если они, например, очень асимметричны, может понадобиться большое число слагаемых. Нужно с некоторой осторожностью считать, что *при большом числе опытов частота события распределяется по нормальному закону*.

Если вероятность события p в одном опыте очень мала, или, напротив, очень близка к единице, для этого может понадобиться огромное число опытов[4.13]. Вот практический способ проверки того, можно ли для частоты события пользоваться нормальным законом. Нужно построить доверительный интервал для частоты события с уровнем доверия 0.997, см. формулу (2.3) с коэффициентом 3,

$$p \pm 3\sqrt{p(1-p)/N}.$$

Если он весь (если оба его конца) не выходит за разумные границы, в которых может лежать частота и вероятность, можно пользоваться нормальным законом, в противном же случае нельзя. Но тогда для приближённого решения задачи можно воспользоваться так называемым распределением Пуассона. Но и это, и очень многие другие распределения выходят за пределы наших *первых шагов*.

Изучив нашу немудрёную книжку, вы, по-видимому, получили некоторое понятие о том, что такое теория вероятностей и чем она занимается. Возможно, что вы решительно отвратились от неё, и тогда первое знакомство с ней станет для вас и последним. Жалко, но ничего не поделаешь. Даже некоторые математики органически не переносят теории вероятностей.

Но возможно (в сущности, для этого и написана книжка), что тематика, методы и возможности теории вероятностей вас заинтересовали. Тогда познакомьтесь с нашей наукой поподробнее, см. список литературы. Не будем скрывать, что углублённое знакомство с теорией вероятностей потребует от вас гораздо более серьёзных умственных усилий, чем до сих пор. Будет труднее, но интереснее. *Корень учения горек, а плоды сладки*. Желаем вам самых сладких плодов!

Литература

1. Румшиский Л. З. (1976), *Элементы теории вероятностей*. М.

2. Гутер Р. С., Овчинский Б. В. (1967), *Основы теории вероятностей*. М.

3. Вентцель Е. С. (1964), *Теория вероятностей*. М.

4. Вентцель Е. С., Овчаров Л. А. (1973), *Теория вероятностей.* (*Задачи и упражнения*). М.

5. Мостеллер Ф., Рурке Р., Томас Дж. (1969). *Вероятность.* М.

6. Гнеденко Б. В., Хинчин А. Я. (1970), *Элементарное введение в теорию вероятностей*. М.

7. Мостеллер Ф. (1971), *Пятьдесят занимательных вероятностных задач*. М.

8. Пугачёв В. С. (1968), *Введение в теорию вероятностей*. М.

9. Яглом А. М., Яглом И. М. (1973), *Вероятность и информация*. М.

10. Гмурман В. Е. (1970), *Руководство к решению задач по теории вероятностей и математической статистике*. М.

Примечания

1.1. Статистический метод ворвался в естествознание уже в XIX веке (Шейнин 2019, § 11.8).

1.2. Решка: решётка, которая была видна на российских монетах, возможно до 1917 г.

1.3. Случайные события (равно как достоверные и невозможные) существует и независимо от опытов.

1.4. Указание автора неверно. С низкой вероятностью могут происходить вредные и даже пагубные события, ожидания которых поэтому вовсе не пренебрегаемы.

1.5. Он исправился и подобных ошибок более не допускал. Е. В.

А вот великий логик Де Морган допускал: в 1864 г. он рассуждал о вероятностях превышающих единицу и даже отрицательных, а ранее, в 1842 г., заявил, что тангенс и котангенс бесконечности равны ± мнимой единице (Шейнин 2019, конец § 5.4).

1.6. Устаревшее *ЭЦВМ* (электронная цифровая вычислительная машина) автора мы заменяли *компьютером*.

1.7. При серьёзных исследованиях выбор уровня доверия регламентируется.

1.8 Те немногие, кто не сразу ответили 1/2, являются *придирами.* Это может быть признаком глубины мышления, но чаще вызывается дурным характером. Е. В.

1.9. Автор сузила значение общеупотребительного слова *случай*.

1.10. Гораздо точнее: не развито логическое мышление.

1.11. Следовало сообщить, во-первых, что классическое определение вероятности случайного события (которое, кстати, ввёл Муавр), содержит порочный круг, и, во-вторых, вообще является только формулой для вычисления, а не определением, которого не существует. Это относится ко многим математическим понятиям, например к прямой линии, см. также прим. 4/4. И обязательно следовало упомянуть подход Мзеса.

1.12. Автор жестоко ошиблась: её рассуждение относилось лишь к субъективной, а не действительной, реальной вероятности (Шейнин 2002).

1.13. Комбинаторика связана со многими ветвями математики.

2.1. Симметрия в азартных играх существует далеко не всегда, а *специальными мерами* служат их условия. Профессиональные игроки это просто жулики, шулеры, от которых следует отличать игроманов. Ср. прим. 2.2.

2.2. Неправильность достигается, скажем, заделкой свинцового грузика внутрь кости. Подобной практикой занимались профессиональные игроки, что могло окончиться их беспощадным избиением. Е. В.

2.3. Компенсируются в какой-то степени случайные погрешности, а систематические ошибки требуют особого внимания.

2.4. Плохо известно, что Якоб Бернулли (гл. 4 четвёртой части *Искусства предположений*) допускал статистическое определение неизвестной вероятности и даже введение статистической вероятности взамен несуществующей теоретической.

2.5. Пирсон (1900) только описал опыт Р. Уэлдона, который подбросил 12 костей 26 306 раз и вывел статистическую вероятность выпадения пяти ил шести очков, равную 0,33770. В четырёх бросках эти исходы появились 11 раз. Об известном аналогичном опыте Бюффона см. Шейнин (2019, конец § 4.3.4).

2.6. При некоторых особых условиях (например, во время и после войны) эта частота по невыясненным причинам может отклониться от многолетней устойчивой средней. Е. В.

2.7. Помимо массовых явлений статистике приходится заниматься и малыми выборками, которые ввёл Стьюдент (Госсет), см. Box (1987), и вообще обработкой небольшого числа наблюдений, например, при выборочных обследованиях.

2.8. Задачи на азартные игры исключительно важны, притом не только с методической точки зрения. Достаточно упомянуть проблему разорения игрока.

2.9. Рассказывают, что знаменитый Даламбер, обучая тупого и знатного ученика, в отчаянии воскликнул: *Ну, честное слово, эта теорема верна. – Почему Вы сразу так не сказали? Вы – дворянин, и я – дворянин. Вашего слова для меня достаточно*. Е. В.

2.10. *Высокая*, а не *большая* вероятность. Единственный раз автор применил более подходящее прилагательное.

2.11. Здесь и во многих случаях ниже количество значащих цифр указано либо неверно, либо бесполезно.

2.12. Автор фактически ввёл начальную гипотезу.

2.13. Вот подходящее мнение (Boole 1851, с. 251): *Цель теории вероятностей можно выразить так: по вероятностям любого предложения определить вероятность другого*. То же по существу заявили Прохоров и Севастьянов (1999, с. 77) относительно случайных событий.

2.14. Время учёного следовало бы ценить исключительно высоко.

3.1. Зависимость и независимость событий всегда взаимны. Если $P(A/B) = P(A)$, то и $P(B/A) = P(B)$. Докажите это, воспользовавшись записью правила умножения (3.4) и (3.5). Е. В.

3.2. Bertrand (1888, с. XXII) доходчиво заметил: *Рулетка не имеет ни воли, ни памяти*.

3.3. См. прим. 2.5.

3.4. Автор описала формулу включения и выключения.

3.5. Событие после символа вероятности P мы будем часто обозначать его соответствующей понятной записью. Е. В.

3.6. Несколько событий называются независимыми, если ни одно из них не зависит от любой комбинации (совмещения) остальных. Попарная независимость событий для этого недостаточна. *Чего только не придумают математики*! Е. В.

Пример на недостаточность попарной независимости в указанном случае см. Бернштейн (1946, с. 47). О. Ш.

3.7. Это не совсем так, но в первом приближении указанное допущение можно принять. Е. В.

3.8. При *n* намного большем 50 опыт становится малоэффективным. Е. В.

4.1. Это определение случайной величины неполно и косвенно уточняется в дальнейшем. Мы полагаем, что такой порядок изложения вряд ли приемлем.

4.2. Легко проверить, что сумма указанной бесконечно убывающей геометрической прогрессии равна единице.

4.3. Для читателей, знакомых с теорией множеств, добавим: число этих значений несчётно. Е. В.

4.4. В геометрии Евклида точка определялась как нечто, не имеющее ни длины, ни ширины, ни высоты (как безразмерная величина). Современная логика, а потому и современная математика, не признаёт отрицательных определений, и точка вводится без всякого определения.

4.5. Существуют и СВ *смешанного* типа: кроме сплошного участка возможных значений с нулевыми вероятностями, они имеют и отдельные значения с ненулевыми. Таких *хитрых* СВ мы не рассматриваем. Е. В.

4.6. Если значение *X* пришлось в точности на границу двух разрядов, мы приписываем половину к каждому. Е. В.

4.7. Прилагательное *математическое* к термину *ожидание* добавил Лаплас (1812/1886, с. 189), чтобы отличить его от вошедшего в то время в моду *морального* ожидания. Это добавление напрасно сохранилось по меньшей мере во французской и русской литературе, и мы не применяем его.

4.8. Эту поправку ввёл Гаусс (1823/1957, с. 146), руководствуясь теоретической необходимостью и *отчасти* достоинством науки.

4.9. Впервые автор применила термин *параметр* вместо *числовая характеристика*. Так обычно и происходит у авторов популярных сочинений: рано или поздно приходится вводить сравнительно простые термины.

4.10. Необходимо также как-то исключать систематические ошибки. См. также прим. 2.3.

4.11. Нормальное распределение следовало бы приписывать и Лапласу, и Гауссу. Пирсон где-то указал, что *нормальным* он назвал это распределение, чтобы не сравнивать заслуги этих классиков науки.

4.12. Употребительное название: *треугольное* распределение (которое действительно ввёл Т. Симпсон).

4.13. Случай низких (и высоких) вероятностей описывается распределением Пуассона.

Литература составителя

Бернштейн С. Н. (1946), *Теория вероятностей*. М. – Л., 1946. Четвёртое издание.

Прохоров Ю. В., Севастьянов Б. А. (1999), Вероятностей теория. В книге Прохоров Ю. В., *Вероятность и математическая статистика. Энциклопедия.* М., с. 46.

Шейнин О. Б., Sheynin O. (2002), Sampling without replacement. History and applications. *NTM, Intern. Z. f. Geschichte u. Ethik Naturwiss., Technik, Med.*, Bd. 10, pp. 181 – 187.

--- (2018), *Элементы теории вероятностей. Элементарное руководство с привлечением сведений из истории этой науки*. М.

--- (2019), *Теория вероятностей. Исторический очерк*. Берлин. **S, G,** 11.

Bertrand J. (1888), *Calcul des probabilités*. Paris. Второе изд., 1907. перепечатки New York, 1970, 1972. Ссылка на перепечатки.

Boole G. (1851), On the theory of probability. В книге автора *Studies in Logic and Probability*, vol. 1. London, 1952, pp. 247 – 259.

Box Joan Fisher (1987), Guinness, Gossett, Fisher and small samples. *Stat. Science*, vol. 2, pp. 45 – 52.

Gauss C. F. (1823, латин.), Теория комбинаций наблюдений. *Избр. геод. соч.*, т. 1. М., 1957, с. 17 – 57.

Laplace P. S. (1812), *Théorie analytique des probabilités*. *Oeuvr. Compl.*, t. 7. Paris, 1886.

Pearson K. (1900), On a criterion that a given system of deviations … can be reasonably supposed to have arisen from random sampling. *London, Edinb., Dublin Phil. Mag.*, ser. 5, vol. 50, pp. 157 – 175.

II

Ю. Н. Тюрин

Что такое математическая статистика

Москва, 1975

Введение

Общепринятой является точка зрения на математическую статистику как на науку об общих способах обработки результатов эксперимента. Одновременно статистика устанавливает и качества, которые должен обладать эксперимент, чтобы сделанные на его основании выводы были правильными. Поэтому она часто даёт рекомендации по проведению эксперимента, становится отчасти наукой о его планировании. В настоящее время эта сторона статистики развивается особенно бурно.

Но так было далеко не всегда (Hodges, Jr & Lehmann [9, p. 209])[1]:

Значение этого слова [статистика] *за последние два столетия значительно изменилось. Оно имеет тот же корень со словом государство* (*state*) *и первоначально означало искусство и науку управления. Первые преподаватели статистики университетов Германии XVIII века сегодня назывались бы специалистами по общественным* (*политическим*) наукам. *Поскольку решения правительства до некоторой степени основаны на данных о населении, мануфактурах* (*ремёслах*)*, сельском хозяйстве и пр., статистики естественно стали интересоваться и такими данными, и постепенно слово статистика стало означать сбор данных о государстве, а затем и вообще сбор и обработку данных. Это значение всё ещё широко употребляется, но наблюдается и дальнейшее изменение значения. Нет смысла извлекать данные, если от них нет какой-то пользы, и статистики естественно начинают заниматься истолкованием данных. Современный статистик изучает методы, при помощи которых можно сделать выводы о популяции на основе данных, которые обычно получают из выборки*.

Отмеченные здесь задачи современной статистики относятся к одной части науки, которая сейчас называется описательной статистикой. Но ещё со времён Гаусса, который применял статистические приёмы для вычисления элементов планетных орбит по несовершенным данным[2], задачей этой науки стала обработка количественных научных экспериментов, и сейчас она составляет её ядро.

Я начал с определения статистики не случайно. Почти каждая книга, посвящённая этой науке, начинается с определения, что такое математическая статистика. В серьёзных статьях собраны коллекции из сотен таких дефиниций. Краткую выборку на русском языке см. Налимов [5, с. 242]. Но мало кого беспокоит отсутствие определений математики или физики. Дело здесь, я думаю, в том, что математическая статистика и теория вероятностей в силу неполно понятых причин, которые лучше назвать историческими[3], довольно поздно вошли в духовную культуру человечества.

Ещё древние халдеи были астрономами, а греки – геометрами[4]. Во времена Рима (?) разрабатывалось учение о государстве. А давно ли человечество занялось обдумыванием, что такое случай и случайность? Даже Паскаль ошибался, решая простейшие задачи об азартных играх. Идеи и представления медленно входят в обиход. Вспомним хотя бы драматическую судьбу законов Менделя о наследственности[5]. На наших глазах статистические методы начинают применяться в химии и геологии, лингвистике и истории.

В какой-то мере наше изложение соответствует истории развития науки. Поэтому первые разделы будут содержать сведения, стоящие ближе к истокам и требующие меньшей подготовки, а последние коснутся ныне развивающихся понятий и идей.

1. Генеральная совокупность и выборка

Значительная часть математической статистики связана с необходимостью описать большую совокупность объектов. Её обычно называют генеральной. Такой совокупностью, например, могут быть все жители Москвы, взятые вместе, месячная продукция завода, производящего телевизоры, популяция рыб в данном заливе. Но генеральная совокупность – не просто множество. Этот термин обычно употребляется в тех случаях, когда множество изучается выборочным методом.

Если интересующая нас совокупность объектов слишком многочисленна, или если её объекты труднодоступны, или по другим причинам все объекты не могут быть изучены, изучают какую-то их часть. Эта часть, выбранная для полного исследования, называется выборкой. Желание выбрать её так, чтобы она лучшим образом представляла целое, была *репрезентативной*, естественно.

Как добиться этого? Если целое, т. е. генеральная совокупность, мало или совсем неизвестна, нет ничего лучшего чисто случайного выбора. Большая осведомлённость позволяет действовать лучше, но всё равно на некоторой стадии наступает незнание, и, как результат, случайный выбор. Приведём примеры.

Допустим, что на большом поле мы хотим оценить урожай прежде, чем он собран. На различных участках растения развивались по-разному. На возвышениях им не хватало влаги, во впадинах влага была избыточна, удобрения были рассеяны тоже не вполне равномерно, по краям поля сорняки сильнее угнетали их. Теоретически было возможно разделить всё поле на ряд однородных участков и изучить каждый отдельно, но такой подход [невозможен]. И всё равно с учётом всего этого останутся менее важные, но заметные источники неоднородности, – случайные колебания числа растений на единицы площади, степень полегания растений, различное действие заморозков и многое другое. Разбиение на совершенно однородные участки невозможно.

Выборочный метод предлагает иной подход. Мысленно разделим всё поле на квадраты, площадью, скажем, 1 *кв. м*. Если площадь поля 12 тысяч *кв. м*. то генеральная совокупность будет состоять из стольких же малых участков. Допустим, что мы обследуем 50 из них. Сколько именно надо обследовать и как наш выбор повлияет на точность результата, мы разберём в §§ 2 и 6, а пока спросим: как их выбрать?

Пронумеруем все участки пятизначными числами, безразлично по какому принципу. Важно только, чтобы каждый участок имел номер и чтобы номера не повторялись. Выбирать 50 из них будем чисто случайно (см. § 2), например, с помощью таблицы случайных чисел. К примеру, участки 13 и 137 записываются в виде чисел 00013 и 00137. На произвольном месте такой таблицы выписываем те пятёрки подряд идущих цифр, которые не превосходят 12 тысяч, но повторяющиеся номера пропускаем. Действуем так, пока не наберём 50 различных номеров, которые и образуют выборку.

Определим урожай в отобранных участках и вычислим его среднее значение, возможные границы ошибки которого также можно будет указать (§ 6).

Сейчас с различными целями проводится немало социальных обследований, но неизменно требуется распределение в исследуемом обществе каких-либо признаков, будь то политические симпатии или использование свободного времени. В зависимости от поставленной задачи генеральной совокупностью может быть население города или страны, служащие определённого предприятия или студенты вузов.

Основной метод исследования – выборочный, ведь на сплошное обследование не хватает ни времени, ни сил. Но как осуществить чисто случайный выбор? Гораздо легче сказать, как этого не следует делать. Неразумно, например, спрашивать прохожих на улице. Они спешат и, возможно не захотят разговаривать, но важнее, что при этом можно не получить представления о тех, кто в это время находится на работе или вообще редко появляется на улице. Не стОит обходить подряд встретившиеся дома и квартиры, так мы получим представление только о жителях данной улицы или района, которые, как правило, принадлежат определённому социальному слою и т. д.

Социологи придумали и придумывают много методик, имитирующих чисто случайный выбор. Как правило, выбираются легко наблюдаемые признаки, которые, как мы уверены, не связаны с интересующим нас признаком. Например, из списка всех сотрудников данного предприятия, можно выделить каждого десятого. Это похоже на имитацию случайности посредством псевдослучайных последовательностей, см. § 2.

Нарушение принципов случайного выбора порой приводило к серьёзным ошибкам. Самым знаменитым своей неудачей стал опрос, который произвёл журнал *Literary Digest*, об исходе президентских выборов 1936 г. в США. Кандидатами были Ф. Д.

Рузвельт и А. М. Ландон. В качестве генеральной совокупности журнал использовал телефонные книги. Отобрав случайно 4 миллиона адресов, он разослал по всей стране открытки с вопросами об отношении к кандидатам. Затратив большую сумму на рассылку и обработку открыток, журнал объявил, что с большим перевесом голосов будет избран Ландон, однако результат оказался противоположным.

Журнал совершил две ошибки. Во-первых, телефонные книги не дают репрезентативную выборку из населения страны, хотя бы потому, что абонентами являются как правило главы зажиточных семейств. Во-вторых, отвечали только достаточно уверенные в себе люди, притом привыкшие отвечать на письма, т. е. в значительной части представители делового мира, которые и поддерживали Ландона. Если бы редакция журнала критически отнеслась к своей работе, она поняла бы, что открыла всем известное.

Подобное явление, когда выборка представляет не всю генеральную совокупность, а только какой-то её слой, какую-то её часть, называется смещением выборки. Оно – один из основных источников ошибок при использовании выборочного метода. В то же время социологи Дж. Галлап и Э. Роупер правильно предсказали победу Рузвельта, основываясь только на четырёх тысячах анкет. Причину этого успеха, который прославил этих авторов, было правильное составление выборки. Они учли, что общество распадается на социальные группы, которые более однородны по отношению к кандидатам в президенты. По этой причине выборка из слоя с той же точностью результатов может быть относительно небольшой. Получив данные обследования по слоям, можно характеризовать общество в целом. Сейчас такая методика является общепринятой, а её количественную сторону мы обсудим в § 2.

Для оценки роли статистического описания генеральной совокупности мы обратимся к примеру начала XX века, который я излагаю по Фишеру, одного из создателей современной математической статистики.

Иоханнес Шмидт из Карлберговской лаборатории в Копенгагене был не только ихтиологом, но и неутомимым биостатистиком. Он развивал идею, что рыбы одного вида распадаются на относительно изолированные сообщества, группы. Между ними он находил статистические различия по числу позвонков или лучей плавников. Причиной различия было то, что эти сообщества не смешивались при размножении. Каждое метала икру в своём месте. И часто эти различия были заметны даже между стаями рыб одного вида, обитавшие в одном и том же фиорде.

Но для угря Шмидт не смог найти никаких статистических различий между выборками, выловленными даже в очень далёких друг от друга местах, будь то различные части Европейского материка, Азорские острова, Нил или Исландия, хотя число позвонков у угря весьма изменчиво. Шмидт решил, что угри всех различных речных систем составляют одно сообщество, а потому должны иметь общее место размножения.

Через некоторое время его предположение подтвердила экспедиция исследовательского судна *Дана.* Одним из её главных успехов была поимка личинок угря в ограниченном районе западной Атлантики (*Journal* 1953)[6].

2. Некоторые понятия теории вероятностей

Начальным понятием теории вероятностей является случайный эксперимент[7]. Это идеализация, абстракция реально проводимых экспериментов, вроде подбрасывания монеты или игральной кости, когда до проведения опыта нельзя сказать, чем он кончится. Впрочем, мысленно можно представить себе его возможные окончания: монета может упасть либо гербом, либо цифрой кверху. При бросании кости мыслимых способов падения шесть. Реальный опыт может закончиться иначе: монета затеряется, кость разобьётся.

Математический случайный эксперимент имеет чётко оговорённое множество так называемых элементарных исходов, не делимых на более простые способы окончания. Каждому элементарному событию приписывают его неотрицательную вероятность, сумма которых равна единице.

Надо представлять себе дело так, словно полная, единичная вероятность распределяется между всеми элементарными исходами, отсюда и термин *распределение вероятностей*. Пример: при бросании игральной кости считают вероятности выпадения её различных граней одинаковыми. Ведь если кость имеет правильную форму куба и изготовлена из однородного материала, все грани равноправны и имеют равные возможности появления при бросании.

Приведённая здесь аргументация была типична для времён Ферма и Паскаля, которые первыми, в XVII веке, начинали теорию случайного. В ней явно проступает сомнение: можно ли исходам неопределённого опыта приписывать какие-то вероятности, которые управляют случайностью, и откуда вообще в природе может взяться случайное? Дискуссии на эту тему, изменяясь с развитием науки, идут постоянно. Сейчас предпочитают отделить математическую абстракцию случайного эксперимента от реально проводимых экспериментов и спорить о приложимости математической модели.

Из элементарных событий составляются комбинации, которые называются просто событиями или исходами. Так, можно говорить о событии *выпало чётное число очков*. Вероятность составного события равна сумме вероятностей составляющих его элементарных событий, так что вероятность выпадения чётного числа очков при бросании игральной кости равна 1/2.

Значительная часть задач математической статистики связана с чисто случайным выбором, и именно с него мы начали рассказ об этой науке. Это – эксперимент, состоящий в выборе одного элемента из генеральной совокупности, при котором все элементы имеют равные вероятности быть выбранными. Иными словами, если N – число элементов (объём) генеральной совокупности, то при чисто случайном выборе каждый элемент может быть выбран с вероятностью $1/N$. Однократным случайным выбором обычно не ограничиваются, и путём подобных многократных выборов получают выборку объёма n. Во многих случаях этот объём определяется результатами предыдущих испытаний, но мы предположим, что величина n задана заранее.

Различают случайные выборы без возвращения и с возвратом. В последнем случае выбранный элемент возвращается в генеральную совокупность, так что при последующих извлечениях он может появиться снова. В практике обычно устраивают выбор без возвращения. При больших объёмах генеральной совокупности N и относительно малом объёме выборки n, т. е. при малом n/N, различие в формулах, описывающих оба варианта выбора, невелико. Но выбор с возвращением изучать легче, потому что он происходит в неизменных условиях при любом извлечении. Впрочем, следует озаботиться тем, чтобы результаты предыдущих извлечений не влияли на последующие результаты.

Так, при извлечении карты из колоды наудачу, после каждого её возвращения необходимо тщательно перетасовывать колоду. Но вот при выборке без возврата при втором извлечении генеральная совокупность уже не та, которая была перед ним.

Как ни странно, осуществить реально независимый и случайный выбор не так-то просто. Нелегко создать устройство, которое осуществляло бы независимые случайные испытания с заданными вероятностями исходов. Впрочем, здесь требуется разъяснение. Как можно судить о вероятностях исходов случайного эксперимента? Как реально проявляются эти вероятности? Вес предмета можно узнать, используя весы, температуру – термометром, но как измерить вероятность? Мы обсудим это после ознакомления с законом больших чисел (ЗБЧ).

Можно, конечно, бросать монету или кость. При малых сериях бросаний результаты окажутся неплохими[8], но при больших сериях станут заметны незначительные отклонения от идеала, и это может оказаться существенным. Например, станет заметным небольшое неравенство вероятностей исходов или влияние предшествовавшего результата на последующий.

Вероятно все слышали об азартных играх, т. е. таких, исход которых полностью определяется случайным испытанием[9]. Самая известная азартная игра, видимо, рулетка. В подобных играх всякое отклонение от чистой случайности игроки могут использовать к своей выгоде против владельца рулетки. По этой причине организаторы игр прилагают много усилий для устройства генератора случайного, как можно более близкого к идеалу. И всё же временами появлялись игроки, которым удавалось подметить несовершенство какого-то конкретного рулеточного колеса и за счёт этого выигрывать.

В общем, чистая случайность – примерно такая же научная абстракция как и другие научные понятия (абсолютно чёрное тело, абсолютно упругое соударение или тело весом 1 *кг*.).

Кроме того, скорость генерирования случайности механизмом вроде рулетки или бросанием кости невысока, а длинные случайные серии бывают необходимы при специальной работе быстродействующих компьютеров[10], то потому необходимы и работающие с той же скоростью источники идеальной случайности. В данном случае речь идёт о чисто случайном и многократном независимом выборе одной цифры из десяти. Выход находят в том, что самим компьютером запускают процесс вычисления какого-нибудь иррационального числа, скажем $\sqrt{2}$, и используют получающуюся последовательность цифр вместо чисто случайных. Если начать десятичную запись этого числа не с первых, хорошо известных цифр, то никакой закономерности не будет заметно[11]. Эта очень любопытная замена понятия случайности понятием отсутствия закономерности может служить темой отдельного разговора.

При случайном выборе небольшого объёма из небольшой генеральной совокупности названия объектов последней иногда записывают на карточках, так что каждому объекту соответствует в точности одна карточка, и из полученной колоды наудачу извлекают нужное число карточек. Исследуются те объекты генеральной совокупности, которые соответствуют извлечённым карточкам.

Этот метод непосредственного случайного выбора применять тем труднее, чем больше объём генеральной совокупности: трудно изготовить большое число карточек, да и сам случайный выбор становится неудобным.

Ввиду перечисленных трудностей появляется мысль проводить выбор заранее в тщательно поставленном случайном эксперименте. При реальном выборе из реальной совокупности можно будет применить полученные впрок результаты. На этой идее основаны таблицы случайных чисел, этот необходимый инструмент для проведения статистических обследований.

Представим некое устройство, которое может независимо и чисто случайно выбирать цифры, так что вероятность появления каждой из них при однократном выборе равна 0,1. Такое устройство можно запустить и записывать появляющуюся последовательность цифр, например 10, 09, 73, … Она и будет записана в качестве таблицы случайных чисел.

Пусть теперь в генеральной совокупности объёма N объекты обладают каким-то количественным признаком. Так, каждый человек имеет определённый возраст, каждый камень на пляже – определённый вес и т. д. Обозначим значения признака через x_1, x_2, …, x_n и соответствующие количества объектов через N_1, N_2, …, N_n. При случайном выборе вероятность появления объекта с признаком x_i равна $N_i/N = p_i$. Итак, при указанных условиях величина признака X принимает значения x_1, x_2, …, x_n с вероятностями p_1, p_2, …, p_n. Сумма этих вероятностей, разумеется, равна единице.

Величину X обычно называют случайной с возможными значениями x_1, x_2, …, x_n, между которыми распределяется полная, единичная вероятность: p_1, p_2, …, p_n. Случайные величины (СВ) появляются и иначе, не обязательно путём чисто случайного выбора. Они служат математическим образом многих реальных объектов, – числа вызовов, поступающих на станцию обслуживания в единицу времени, числа судов, посетивших данный порт за данный месяц, количества заболеваний гриппом в данном районе за неделю и т. п[12].

Ясно, что СВ являются основным объектом теории вероятностей и математической статистики. Описанный здесь тип СВ, принимающих отдельно стоящие значения, отличается от непрерывного типа, см. § 5.

Изучение СВ это изучение распределения их вероятностей. Удобными числовыми характеристиками подобных распределений служат ожидание[13] и дисперсия.
Ожиданием СВ ($\mathrm{E}X$) называют величину

$$\mathrm{E}X = x_1p_1 + x_2p_2 + \ldots + x_np_n.$$

Если считать, что возможные значения СВ суть координаты точек на числовой прямой с массами, пропорциональными величинам $p_1, p_2, \ldots, p_n$, то $\mathrm{E}X$ окажется координатой центра этих масс, как бы центром вероятностей. Термин *ожидание* имеет не менее употребительный синоним, *среднее значение.* Важнейшее свойство ожидания выражается формулой

$$\mathrm{E}(X + У) = \mathrm{E}X + \mathrm{E}У.$$

Так, ожидание суммы очков, выпадающих при броске двух игральных костей, равно 7, потому что среднее значение числа очков, выпадающих при броске одной кости равно одной шестой суммы 1 + 2 + … + 6.

Дисперсия СВ X,

$\mathrm{D}X$ или $\sigma^2(X)$ или $\mathrm{Var}X$ определяется формулой $\mathrm{E}(X - \mathrm{E}X)^2$

и служит некоей мерой разброса вероятностей около центра $\mathrm{E}X$. Механическим аналогом этого понятия является момент инерции системы масс. В сущности, разброс СВ измеряется корнем из её дисперсии, что, кстати, подсказывается и соображениями размерности. Этот корень называется среднеквадратическим или стандартным отклонением. Дисперсия завоёвывает право на существование благодаря формуле[14]

$$\mathrm{D}(X + Y) = \mathrm{D}X + \mathrm{D}Y,$$

которая имеет место для независимых X и $У$. Мы не приводим формального определения независимости и ограничиваемся её описанием: X и $У$ независимы, если сведения о значении, которое приняла одна из них, не меняют распределения вероятностей другой.

Пример. В урне находится три шара с номерами 1, 2 и 3. Двое по очереди наудачу извлекают из неё по одному шару и выигрывают соответственно 1, 2 или 3 рубля. Их выигрыши обозначим через X и $У$. Можно доказать, что распределения X и $У$ совпадают, так что, к примеру, события $X = 2$ и $У = 2$ равновероятны. Но вероятность того, что в одном эксперименте $У = 2$ при условии, что $X = 2$, равна нулю, потому что в урне только один шар с этим номером, и он уже был извлечён. Величины X и $У$ зависимы. Если после извлечения шара первым участником опыта шар возвращается в урну и указанная условная вероятность совпадает с безусловной, то случайные величины независимы. В этом и состоит удобство выбора с возвращением, которое, как мы отметили, упрощает его изучение.

Пусть теперь в урне N_1 шаров с номером 1, N_2 шаров с номером 2, и N_3 шаров с номером 3. Вероятности выигрыша для игроков будет равны $p_1 = N_1/N$, $p_2 = N_2/N$ и $p_3 = N_3/N$, где N – общее число шаров. При выборе без возвращения условные вероятности для второго игрока будут отличаться от указанных безусловных в зависимости от результата первого извлечения. Так, вероятность второму выиграть 2 рубля, если первый выиграл столько же, равна $(N_2 - 1)/(N - 1)$, что отличается от N_2/N, т. е. от условной вероятности при выборе с возвращением. Но ясно, что при больших величинах N_1, N_2 и N_3 подобные различия невелики, так что между двумя способами выбора почти не будет различия.

Одним из важнейших результатов теории вероятностей как науки является ЗБЧ. Это название покрывает многочисленные результаты общего типа, ясное представление о которых даёт следующий.

Пусть $X_1, X_2, \ldots$ независимые, одинаково распределённые СВ, ожидание которых обозначим через a. Тогда среднее арифметическое

$$\overline{X} = \frac{1}{n}(X_1 + \ldots + X_n)$$

при неограниченном возрастании n сходится к a. Более формально: для любой границы $\varepsilon > 0$ вероятность того, что разность $|\overline{X} - a|$ превзойдёт её, стремится к нулю при возрастающем n, т. е.

$$P(|\overline{X} - a| > \varepsilon) \to 0.$$

Таким образом, ЗБЧ утверждает, что при большом n СВ $\overline{X}$ приближённо равна a. О точности этого приближения см. § 6. Каждый знаком с ЗБЧ не как с математической теоремой, а как с разумным правилом действия. Каждый раз, когда хотят возможно точнее измерить неизвестную величину недостаточно совершенным прибором, проводят несколько измерений и усредняют результат[15].

В указанной форме ЗБЧ сейчас доказывают с помощью неравенства Чебышёва[16]. Пусть X – СВ, $\mathrm{E}X = a$, $\mathrm{D}X = \sigma^2$ и ε произвольное положительное число. Тогда это знаменитое неравенство выглядит так:

$$P(|X-a|\geq\varepsilon)\leq\frac{\sigma^2}{\varepsilon^2}.$$

До Чебышёва ЗБЧ получался как следствие центральной предельной теоремы (ЦПТ), см. § 5, которая была известна в очень частном случае и доказывалась весьма сложно.

Теперь доказательство ЗБЧ занимает несколько строк. По свойствам ожидания и дисперсии E $\overline{X} = a$, D $\overline{X} = (1/n)\sigma^2$ (DX = σ^2). По неравенству Чебышёва

$$P(|\overline{X}-a|\geq\varepsilon)\leq\frac{\sigma^2}{n\varepsilon^2}.$$

При $n\to\infty$ правая часть стремится к нулю, что и завершает доказательство.

Неравенство Чебышёва и его различные обобщения применяются при теоретических исследованиях. Иногда встречается и его неразумное использование для оценки точности приближения $\overline{X}$ к a. Правильное представление о характере приближения даёт ЦПТ, см. ниже.

Вернёмся к чисто случайному выбору с возвращением. Пусть только часть объектов генеральной совокупности обладает неким свойством А. Например, только часть шаров в урне красные, только часть людей – курильщики. Пусть p – вероятность выбрать объект, обладающий этим свойством, при случайном выборе. Проведено n таких извлечений с возвратом и получено случайное количество X объектов с указанным свойством.

Возможны значения X, равные 0, 1, …, n с вероятностями

$$P\,|\,X=k\,|=C_n^k(1-p)^{n-k}.$$

Можно доказать, что E$X = np$ и D$X = np(1-p)$. Исторически первый вариант ЗБЧ, так называемая теорема Бернулли, утверждает, что $X/n\to p$. Иначе говоря, доля объектов А в выборке приближённо равна их доле в генеральной совокупности. Следовательно, теорема Бернулли (совместно с указанной связью выборов с возвращением и без него) утверждает, что чисто случайный выбор с большой вероятностью обеспечивает репрезентативную выборку.

Поговорим подробнее. Пусть в генеральной совокупности большого объёма N имеются объекты типа $A_1, A_2, \ldots, A_m$, составляющие доли $p_1, p_2, \ldots, p_m$. Проведён случайный выбор без возвращения намного меньшего объёма n. В ней окажется n_1 объектов типа A_1, n_2 объектов типа A_2 и т. д. По теореме Бернулли $n_1/n\approx p_1$, $n_2/n\approx p_2$ и т. д.

Объекты каждого типа представлены в выборке большого объёма примерно в тех же пропорциях, что и в генеральной совокупности, т. е. случайная выборка репрезентативна. Этот факт объясняет роль случайного выбора в статистических исследованиях. И теперь ясно, как можно измерить вероятность события, разумеется, применяя ЗБЧ. Частота появления события в длинной серии независимых испытаний примерно равна его вероятности[17]. К сожалению, измерение является лишь приближённым, и о его точности см. § 6.

До сих пор мы обсуждали случайный выбор из всей генеральной совокупности, теперь же обратимся к затронутому ранее вопросу об изучении этой совокупности, разделённой на слои $H_1, \ldots, H_m$. Пусть доля слоя H_i в общей численности равна p_i, т. е. вероятности получить при чисто случайном выборе элементы из слоя H_i. Будем считать, что эти величины известны. В социальных исследованиях группы $H_1, \ldots, H_m$ являются социальными слоями, на которые можно разделить общество, и их численности известны достаточно точно.

Допустим по-прежнему, что каждый элемент генеральной совокупности обладает каким-то значением количественного признака X и что мы интересуемся величиной a, средним значением X по всей генеральной совокупности. Для оценивания a можно применять выборочный метод двояко: сделать чисто случайный выбор из генеральной совокупности либо в целом, либо для каждого слоя в отдельности. Общий объём выборки, n, считаем постоянным.

Ясно, что среднее значение признака по слою H_i, вообще говоря, отличается от a. Нетрудно понять, что

$$a = a_1p_1 + a_2p_2 + \ldots + a_mp_m.$$

Проводя случайный выбор по каждому слою отдельно, можем оценить величины $a_1, a_{2, \ldots,} a_m$, а потому и a. Пусть объём выборки из слоя H_i равен n_i. Как говорилось ранее,

$$n = n_1 + \ldots + n_m.$$

Получив n_i значений признака X, вычислим их среднее арифметическое y_i. Теперь в качестве оценки a выбираем сумму

$$y_1p_1 + \ldots + y_mp_m,$$

ожидание которой совпадает с a. Другую оценку получим, если проведём чисто случайный выбор объёма n из всей генеральной совокупности. Ей окажется среднее арифметическое $\overline{X}$ из полученных значений признака. Ясно, что $\mathrm{E}\,\overline{X} = a$.

Для сравнения двух различных оценок a надо обратиться к их дисперсиям по всей генеральной совокупности σ^2 и по слою H_i, σ^2_i. Они связаны соотношением

$$\sigma^2 = \sum_{i=1}^{m} \sigma_i^2 p_i + \sum_{i=1}^{m} (a_i - a)^2 p_i.$$

Далее,

$$D\overline{X} = \frac{1}{n}\sigma^2, \; D(y_1 p_1 + \ldots + y_m p_m) = \sum_{i=1}^{m} \frac{\sigma_i^2}{n_i} p_i^2 .$$

Обыкновенно численности $n_1, \ldots, n_m$ выбирают пропорционально численности самих слоёв, т. е. в конечном счёте пропорционально вероятностям $p_1, \ldots, p_m$. И тогда последняя формула несколько упрощается:

$$D(y_1 p_1 + \ldots + y_m p_m) = \frac{1}{n} \sum_{i=1}^{m} \sigma_i^2 p_i .$$

Видно, что за счёт второго слагаемого (?) левая часть этого равенства превосходит дисперсию оценки по слоям. Это слагаемое неотрицательно и выбор по слоям является беспроигрышным.

3. Принцип практической достоверности

Вспомним основные предыдущие результаты. При больших объёмах случайная выборка репрезентативна. Впрочем, точное утверждение предоставляет ЗБЧ: при больших объёмах выборки и любых $\varepsilon > 0$ и $\delta > 0$

$$P\left(\left|\frac{n_1}{n} - p_1\right| < \varepsilon, \; \left|\frac{n_2}{n} - p_2\right| < \varepsilon, \ldots, \; \left|\frac{n_m}{n} - p_m\right| < \varepsilon\right) > 1 - \delta .$$

Иными словами, выборка повторяет основные черты генеральной совокупности лишь с близкой 1 достоверностью. Существует положительная, хоть и малая вероятность, что это соответствие нарушится действием случая.

Указанное положение типично для приложений теории вероятностей: при определённых условиях её утверждения выполняются с высокой вероятностью, *как правило*. Всегда остаётся маловероятная возможность исключения из правила. Ниже нам встретятся другие примеры подобных утверждений. Так как же быть с практической реализацией вероятностных рекомендаций?

Связь теории вероятностей и математической статистики с практикой определяется так называемым *принципом практической достоверности*. В однократном эксперименте маловероятное событие не наступает. Сразу же отметим, что вероятность подобных событий всё же положительна, но, тем не менее, этот принцип имеет право на существование. Оно доказывается успешностью его применения.

Иногда для доказательства указывают, что при большом числе случайных испытаний, хотя бы и различных, мы будем ошибаться с низкой вероятностью (будут происходить события, которые мы сочли невозможными). По ЗБЧ процент неверных решений будет мал, но отличен от нуля. Как же быть? ЗБЧ снова привёл нас к положению, из которого мы надеялись выбраться с его помощью. Принцип практической достоверности приходится принимать как оправдывающийся на практике.

По существу, он является обычным рассуждением по непрерывности: небольшое изменение в причинах приводит к небольшому изменению следствий**[18]**, и, в частности, считаем, что низкую вероятность можно полагать равной нулю.

Остаётся неясность: что такое маловероятное событие? Обычно ими считают вероятности ниже 0,05, но на каком основании? Полагают, что решение должно приниматься конкретно в каждом отдельном случае в зависимости от того, какими последствиями угрожает ошибка. Логически более гармонична концепция Вальда, см. § 10. Там же мы укажем причины её малой известности.

Один мой знакомый объясняет студентам суть маловероятного события следующим образом. Он заявляет, что знает признак делимости на 23 и предлагает слушателям проверить его, назвав десятизначное число. Такое число называют, знакомый же заявляет, что оно не делится. Проверка чаще всего оправдывает его утверждение, потому что только примерно 4,3% десятизначных чисел делится на 23. Ошибка, правда, ему ничем не грозит.

В реальных задачах ошибка всегда влечёт за собой неприятные последствия, и, казалось бы, чем более низкую вероятность считать невозможной, тем лучше. Но это неверно: чем выше вероятность события, тем обычно менее полезна уверенность в том, что оно произойдёт. Как правило, стремясь сделать наши утверждения более содержательными, приходится повышать вероятность событий, которые объявляются невозможными**[19]**.

4. Проверка статистических гипотез

Подобная проверка предоставляет хороший пример применения принципа практической достоверности. Представим себе случайный эксперимент с неизменным множеством исходов, вероятность которых может реализоваться несколькими способами. Обычно имеется параметрическое семейство распределений, и гипотеза формулируется относительно истинного значения нужного нам параметра.

Итак, статистическая гипотеза это гипотеза о том, как распределяется вероятность между исходами данного эксперимента. Она утверждает и то, что может быть, т. е. указывает возможные способы её нарушения. Иначе говоря, кроме тех распределений, которые гипотеза допускает, следует указывать и те *конкурирующие распределения,* которые могут встретиться, если она окажется неверной.

Как проверяют гипотезы? Заранее, до опыта, или вне зависимости от его результатов, надо выбрать событие S, маловероятное при верной гипотезе и возможно более вероятное в противном случае. Это событие, как сказано, следует выбирать так, чтобы $P(S)$, вычисленная по каждому закону распределения, допускаемого гипотезой, была низкой, а событие S практически невозможным. Тогда, стало быть, при верной гипотезе событие S окажется практически невозможным.

С другой стороны при нарушении гипотезы такой выбор S при котором $P(S)$ по возможности высока, увеличивает чувствительность проверки: если эта вероятность высока, высока и вероятность обнаружить нарушение гипотезы. Итак, гипотеза отвергается, если в эксперименте происходит событие *S,* в противном же случае не отвергается. Последнее не означает, что гипотеза доказана, можно только утверждать, что результат эксперимента ей не противоречит.

Вот пример (который сообщил мне М. Д. Каргер) в несколько упрощённом виде. В определённом районе площадью несколько сот квадратных километров Большого Кавказского хребта имеются многочисленные рудные объекты разной величины. Как правило, они сопровождаются разломами земной коры разной глубины, которые можно выявлять геофизическими методами. Имеются основания предполагать, что рудные объекты связаны с этими разломами, причём чем глубже разлом, тем большего размера могут быть сопровождающие их рудные объекты. Конечно, если эта закономерность реальна, то лишь как тенденция. Было бы очень полезно подтвердить такую тенденцию, потому что она сильно облегчит поиски месторождений. Попытаемся применить для этого статистические методы.

Выберем 10 крупнейших рудных объектов и пронумеруем их в порядке убывания размера и убывания глубины разломов. Вторая нумерация такова: 3, 1, 2, 6, 4, 5, 9, 8, 10, 7. Наше предположение чисто качественное и довольно трудно сказать, насколько оно подтверждается наблюдением. Поэтому попробуем испытать предположение, что никакой связи не существует[2ß], так что вторая нумерация не имеет никакой связи с первой, случайна по отношению к ней.

Статистическая гипотеза будет означать, что вторая последовательность есть результат случайного (в смысле теории вероятностей) эксперимента, в котором каждая нумерация имеет равную вероятность осуществления. Мы имеем в виду выбор перестановки чисел 1, 2, …, 10, числом 10! = 3 628 800. Гипотеза относится к одному закону распределения, равномерному, т. е. в равной возможности всех нумераций.

Надо искать событие маловероятное, если гипотеза верна, и правдоподобное, если она нарушена. Нарушение может быть более или менее сильное и вероятность события *S* должна быть тем выше, чем резче нарушена гипотеза. Мы ожидаем нарушения гипотезы об относительной случайности второй нумерации. При совпадении обоих последовательностей мы, пожалуй, не стали бы сомневаться в указанном нарушении; впрочем, мы ниже обсудим насколько это верно. Поэтому в множество [уже не отдельное событие] должна войти первая последовательность и близкие к ней.

Близость между двумя перестановками $x_1, \ldots, x_n$ и $y_1, \ldots, y_n$ можно измерять различными способами. Часто для этого применяют так называемый ранговый коэффициент корреляции по Спирмену:

$$\rho = 1 - \frac{6}{n^3 - n} \sum_{i=1}^{n} (x_i - y_i)^2.$$

Здесь $(n^3 - n)/3$ – наибольшее возможное значение суммы квадратов, и поэтому ρ изменяется от 1, если нумерации совпадают, до – 1, если они противоположны. В нашем примере изменяться может только вторая последовательность. Так как она случайна, то случайно и значение ρ. Если вторая последовательность чисто случайна, то близкие к 1 по абсолютному значению ρ имеют низкую вероятность, типичные же значения ρ близки к нулю.

Распределение ρ при разных значениях n рассчитано, и его можно отыскать в специальных сборниках статистических таблиц, например в таблицах Большева и Смирнова [6, с. 153]. Мы надеемся на проявление связи между двумя признаками и поэтому должны взять событие вида $| \rho > \rho_0|$. Величина ρ_0 выбирается так, чтобы событие S при верной гипотезе имело низкую вероятность, например, 0,01 или 0,005.

В нашем случае $\rho \approx 0{,}84$, сумма квадратов равна 26. В таблицах находим, что значения этой суммы, меньшие или равные 32, имеют вероятность 0,0036. Значение $\rho \geq 0{,}084$ могло появиться лишь с вероятностью 0,0036, что практически невозможно. Это заставляет нас считать, что гипотеза независимости признаков опровергается; доказана, стало быть, положительная связь глубины разлома и размером месторождения.

Попробуем теперь уменьшить объём наших наблюдений и ограничиться пятью месторождениями. Мы получили последовательности 1, 2, 3, 4, 5 и 3, 1, 2, 5, 4. Предположим, что вторая нумерация чисто случайна. Тогда каждая из 5! перестановок имеет вероятность $1/120 \approx 0{,}0083$. В качестве события S можно предложить множество, состоящее только из одной последовательности 1, 2, 3,4, 5 (полное совпадение). Вероятность S достаточно низка, примерно 0,0083. Впрочем, при меньших значениях n, например, при $n = 3$, даже полное совпадение нумераций не доказывает наличия закономерности. Если же при $n = 5$ попытаемся увеличить S (чтобы повысить чувствительность к нарушениям гипотезы, конечно, за счёт некоторого повышения вероятности отвергнуть верную гипотезу), то в S надо будет включить последовательности, наиболее похожие на 1, 2, 3, 4, 5. Таковы последовательности, содержащие одну перестановку номеров, т. е. 2, 1, 3, 4, 5; 1, 3, 2, 4, 5; 1. 2, 4, 3, 5 и 1, 2, 3, 5, 4. Теперь S будет состоять из пяти последовательностей и его вероятность повысилась до $5/120 \approx 0{,}017$. Если бы в нашем примере второй последовательностью оказалась одна из перечисленных, можно было бы с осторожностью решить, что связь существует, и рассмотрели бы дополнительный материал. Впрочем, связь при $n = 5$ не подтвердилась. Это общая черта статистических методов: для их применения требуется достаточный объём наблюдений.

Могло бы статься, что и при $n = 20$ статистически значимая связь не обнаруживается. Это означало бы, что связь слаба, и тем слабее, чем больше нужно наблюдений, чтобы её выявить. Конечно, нельзя точно сказать, что здесь означает *сила связи*, и поэтому нам пришлось сформулировать задачу без упоминания этого термина. Но уж если мы не собираемся использовать связь признаков, то нет смысла чрезмерно увеличивать объём наблюдений, особенно если связь слаба, т. е. если признаки почти независимы.

Наш пример довольно типично описывает применение статистики в естествознании. Как правило, естественнонаучные гипотезы не имеют статистического характера, и исследователь должен выразить их в подходящей статистической форме. Из многих возможных способов следует выбрать тот, который позволит либо использовать имеющийся материал, либо спланировать опыт, который его доставит.

Проверки статистических и иных гипотез сходны. Обычные естественнонаучные гипотезы не допускают прямой проверки, и их стараются подтвердить или опровергнуть, проверяя по возможности их логические следствия. Так, в момент его открытия Ньютоном, закон всемирного тяготения нельзя было проверить непосредственным опытом. Для самогО Ньютона и его последователей решающей проверкой был вывод закономерностей движения Луны, и он объяснил многие особенности этого движения[21].

Создание полной теории движения нашего спутника оказалось труднейшей задачей, которая потребовала всей силы созданной в XIX веке небесной механики. Положение Луны надо было суметь очень точно рассчитать на десятилетия вперёд и проверять наблюдениями, а вычисленные на 25 веков назад моменты солнечных затмений должны были совпадать с отмеченными в хрониках и летописях.

Статистическая проверка гипотез также исследует их следствия. Но даже следствия правильных гипотез оказываются только практически достоверными. Если такого вывода сделать нельзя, гипотеза отбрасывается, хотя остаётся маловероятная, как правило, возможность ошибочно забраковать верную гипотезу. Подобную вероятность желательно по возможности снизить, но при этом легко перестараться: получаемое следствие получит высокую вероятность и при многих других гипотезах. Критерий окажется нечувствительным к нарушению гипотезы: не будет отвергать ошибочной[22].

5. Непрерывные распределения

Изучение конечных генеральных совокупностей выборочным методом является важной, но не главной задачей математической статистики, важнейшей служат эксперименты с непрерывным распределением. Вот его пример.

Перед нами новая электрическая лампочка. Каков срок её службы? Предвидеть этого мы не можем, он случаен. В математическом смысле это не просто неопределённость, это означает, что можно обсуждать вероятности событий, связанных с этим сроком службы, X. Случайность X означает, что мы признаём существование вероятностей событий вида $X < c$ или

$X > c$ или $a < X < b$ и т. д. при произвольных a, b, c. Например, мы можем думать о вероятности события *купленная мной лампочка не прослужит и трёх дней*.

Вероятности событий вычисляются здесь через плотность вероятности. Понятие плотности имеет общий характер и может относиться к различным величинам. Пусть мы хотим вычислить давление кучи песка на опору. Оно пропорционально количеству песка, т. е. пропорционально его объёму. Объёмы и площади математически выражаются интегралами.

Подобным образом можно рассуждать о распределении массы по длине неоднородного стержня (т. е. о плотности массы), о годовом распределении осадков по поверхности Земли, о плотности радиоактивного излучения и т. д. Плотность вероятности выделяется лишь тем, что общая плотность всегда считается равной единице.

Поговорим о случайных величинах и распределениях вероятностей аккуратнее. Итак, пусть X – случайная величина, к примеру, срок службы лампочки. Она принимает действительные значения, т. е. распределяет вероятность на действительной числовой оси. Плотность вероятности в точке x обозначим через $p(x)$. Здесь x обозначает любую точку на числовой оси. Срок службы лампочки не может быть отрицательным, значит вероятность распределяется только среди положительных чисел и для $x < 0$ $p(x) = 0$. Вероятность любого события вида $a < X < b$ определяется через плотность интегралом

$$P(a < X < b) = \int_a^b p(x)dx.$$

Неплохим приближением для плотности в нашем примере служит формула

$$p(x) = ae^{-ax} \text{ при } x \geq 0. \qquad (5.1)$$

Это так называемый показательный закон распределения. Случайные величины с примерно показательным распределением встречаются в природе довольно широко. Этому закону часто подчиняются длительность обслуживания (в системах массового обслуживания), в частности длительность телефонных разговоров, а также величины типа время жизни изделий, живых организмов, человека. Возможно, это объясняется характеристическим свойством, которое выделяет показательное распределение: прожив t лет, вероятность прожить ещё s лет не будет зависеть от t, она зависит только от s. Процесс как бы возобновляется каждое мгновение, забывая своё прошлое.

Это удобно выразить в других терминах. Для любой случайной величины с плотностью $p(x)$ определим риск в момент x формулой

$$r(x) = \frac{p(x)}{\int_x^{\infty} p(u)du}.$$

Условная вероятность умереть в течение малого интервала времени $(x, x + dx)$ при условии дожить до момента x приближённо равна $r(x)\mathrm{d}x$. Для показательного распределения с плотностью ae^{-ax} риск постоянен и равен a. В реальных процессах риск изменяется. Вскоре после рождения высокий риск, который отражает повышенную детскую смертность, резко падает, затем следует участок примерного постоянства риска, после которого он в старости возрастает.

Понятие риска уже несколько столетий используется при страховании жизни. По наблюдениям составлены специальные таблицы вероятностей прожить ещё ряд лет для лиц разного возраста с учётом состояния здоровья, рода занятий и пр. [и пола!]. Конечно, ввиду изменения средней продолжительности жизни постоянно приходится вносить поправки.

Но напомним, что мы говорим о случайной величине X с плотностью вероятности $p(x)$. Как и для дискретных случайных величин, для X определяется ожидание $\mathrm{E}X$ и дисперсия $\mathrm{D}X = (X - \mathrm{E}X)^2$, причём

$$\mathrm{E}X = \int_{-\infty}^{+\infty} xp(x)dx, \mathrm{D}X = \int_{-\infty}^{+\infty} (x - \mathrm{E}X)^2 p(x)dx.$$

Для показательного распределения с параметром a ожидание равно $1/a$. Свойства ожидания и дисперсии одни и те же как для непрерывных, так и для дискретных распределений. И вообще различие между ними имеет место лишь в математическом аппарате, которым приходится пользоваться при их описании.

В частности, для последовательности независимых и одинаково распределённых случайных величин с общим ожиданием действует ЗБЧ, см. § 2.

Независимость случайных величин понимается так же, как в дискретном случае: при последовательном экспериментировании значение, принятое X_1 не влияет на то, какое значение примет X_2, ни одно из этих значений не влияет на значение, которое примет X_3 и т. д. Время горения первой лампочки никак не влияет на срок службы какой-либо другой.

Впрочем, независимость иногда сомнительна. Пусть срок службы первой лампочки оказался равным 400 часам. Тогда мы ожидаем, что тот же срок для второй лампочки окажется ближе к 400, чем к 40 или к четырём, и это можно принять за проявление зависимости между X_1 и X_2. На самом деле результат первого испытания представляет сведения не о X_2, а о параметре a. Не будь a известно, никакого уточнения X_2 сведения об X_1 нам бы не дали, так что величины X_1 и X_2 независимы.

Важнейшим для математической статистики является распределение с несколькими наименованиями: гауссовское, лапласовское, и нормальное в знак того, что для случайной величины именно оно находится в порядке вещей. Ниже мы скажем, какими путями его нашли Гаусс и Лаплас. Нормальное распределение задаётся плотностью, которая зависит от двух параметров, традиционно обозначаемых *a* и σ (σ > 0)

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-a)^2}{2\sigma^2}).$$

Случайная величина с такой плотностью имеет *a* своим ожиданием и σ^2 – дисперсией. Важную роль играет интеграл от стандартной нормальной плотности, который иногда называют функцией Лапласа

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

Этой функцией либо обратной ей приходится пользоваться во многих задачах. Для этой функции составлены подробные таблицы, такие, как для синуса и косинуса.

Мы уже говорили, что само название нормального распределения свидетельствует о его широкой распространённости. По-видимому, один из лучших и исторически первых примеров так распределённых случайных величин дают астрономические и геодезические измерения, которые проводились очень тщательно. И всё же повторные измерения одного и того же угла дают несколько отличающиеся результаты. Вот что об этом писал Гаусс[23].

Этот взгляд принят и сейчас. Считается, что аккуратно проведённые измерения одной той же величины представляют последовательность независимых случайных величин. Для достижения независимости наблюдений и устранения систематических ошибок принимаются специальные меры. В таких случаях ошибки наблюдения обычно распределены нормально с нулевым средним и дисперсией, зависящей от качества измерений.

Наиболее важным в сказанном является предположение о случайности ошибок. Математическая случайность отличается от просто неопределённости тем, что при большом числе повторений даёт закономерность ввиду ЗБЧ.

Теория вероятностей не доказывает и не может доказать, что исходные наблюдения являются независимыми реализациями случайной величины. Она принимает это как предпосылку и на ней основывает свои выводы. Многочисленные и тщательно поставленные опыты, например, с бросанием монеты, подтверждают это предположение. По существу, подчинение ряда явлений статистическим закономерностям составляет фундаментальный закон природы.

Не всякие измерения являются случайными. Неряшливые наблюдения содержат грубые ошибки, подгонки[24] и т. п. и никакое искусство статистики не сможет извлечь из них полезных сведений. Не всегда это легко понять. Довольно часто приходится сталкиваться со статистической обработкой плохих данных. Например, для нахождения параметров технологических процессов используют журналы операторов, в которых через определённые интервалы времени записываются показания приборов. Эти записи недостаточно точны (недостаточно число знаков), да они и недостоверны: ведь оператор должен вести процесс, чтобы удерживать эти показания в установленных границах. Поэтому встречающиеся изредка нарушения режима, самые интересные для технолога, в журналы не попадают.

Гауссовскому распределению подчиняется изменчивость многих признаков в природе. Вот что пишет Ван дер Варден (1960, с. 84):

Я до сих пор живо помню, как однажды, когда я ещё был ребёнком, отец привёл меня на край города, где на берегу стояли ивы, и велел мне сорвать наугад сотню новых листочков. После отбора листьев с повреждёнными кончиками у нас осталось 89 целых листиков. Вернувшись домой, мы расположили их в ряд по росту, как солдат. Затем отец провёл через кончики листьев кривую и сказал: Это и есть кривая Кетле. Глядя на неё, ты видишь, что посредственности всегда составляют подавляющее большинство, и что лишь немногие поднимаются выше, или так и остаются внизу.

Кривая Кетле это повёрнутая горизонтально функция Лапласа $\Phi(x/\sigma)$ с подходящим параметром σ.

С гауссовским распределением связан один из наиболее важных и глубоких результатов теории вероятностей, так называемая центральная предельная теорема (ЦПТ). В вольной передаче она состоит в том, что сумма большого числа независимых случайных величин [с дисперсиями одного и того же порядка] распределена приблизительно нормально, независимо от распределения отдельных слагаемых. (От этого распределения зависит точность соответствия нормальному закону. Так, сумма равномерно распределённых слагаемых нормализуется очень быстро, на чём основан метод искусственно получения в компьютере нормального распределения случайных чисел. Обычно берут сумму 10 равномерно распределённых чисел[25].)

Открытие этого фундаментального закона принадлежит Лапласу[26], хотя у него были предшественники. Вот одна из форм ЦПТ, которую сейчас называют теоремой Муавра – Лапласа. Она говорит о распределении уже знакомого нам случайного числа успехов S в n испытаниях Бернулли, когда n велико. Теорема утверждает, что для фиксированных произвольных значений a, b, $a < b$,

$$P\left(a < \frac{S - np}{\sqrt{np(1-p)}} < b\right) \to \Phi(b) - \Phi(a),\ n \to \infty. \qquad (5.2)$$

Здесь p – вероятность успеха в одном испытании и Φ – упомянутая функция Лапласа. Мы уже заметили, что $\mathrm{E}S = np$, и $DS = np(1 - p)$, так что дробь в левой части формулы (5.2) имеет нулевое среднее и единичную дисперсию. Оказывается, что при больших n её закон распределения похож на нормальный, также с нулевым средним и единичной дисперсией.

Теперь понятно, как Лаплас пришёл к нормальному закону. ЦПТ отчасти объясняет широкую распространённость гауссовского распределения. Оно возникает каждый раз, когда на результат суммарно действуют много малых независимых причин. Ввиду их малости их даже чаще всего трудно выделить.

Важнейший после Лапласа шаг сделал Ляпунов[27]. Многочисленные последующие исследования привели к уверенности, что при сложении большого числа слабо зависящих друг от друга величин, каждая из которых вносит в суммарный результат небольшой вклад, закон распределения сумм оказывается приблизительно нормальным.

6. Доверительные интервалы

Мы уже сталкивались с оцениванием неизвестного параметра, когда (§ 2) говорили об испытаниях Бернулли. Напомню, что его схема достаточно хорошо описывала выбор большого объёма из генеральной совокупности. Вообще, установить оценку неизвестной величины по результатам наблюдений означает найти её приближённое значение. Мы все хорошо знакомы с приближёнными вычислениями и с понятием приближения. Надо ясно представлять себе и точность приближения, границу возможной ошибки. К примеру, число 1000 может считаться приближённым значением и для 920, и для 993 и для 999,3. Точность такого приближения составит 100, 10 и 1 соответственно. Без указания точности приближения приближённое значение само по себе мало что даёт.

Естественно, что такая важная общая идея проявляется и в статистике. Здесь понятие точности приближения реализуется в виде доверительного интервала. Пусть неизвестная величина a измерена n раз с помощью прибора заданной точности и получены значения

$$x_1, x_2, \ldots, x_n. \tag{6.1}$$

Допустим, что мы почти устранили все источники систематических ошибок, так что остаточные ошибки случайны. На основании § 5 мы принимаем обычную математическую модель ситуации: указанные результаты наблюдений суть реализации независимых случайных, нормально распределённых величин. Их ожидание a неизвестно, и его надо оценить.

Качество прибора естественно выражается дисперсией случайных ошибок наблюдений. Чем больше дисперсия, тем менее устойчивы результаты. Это качество, т. е. $Dx_i = \sigma^2$, можно считать известным[28]. Основываясь на ЗБЧ, примем за оценку a среднее арифметическое наблюдений (6.1), т. е.

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Можно показать, что $\bar{x}$ также распределено нормально со средним a и дисперсией σ^2/n. Из свойств ожидания и дисперсии следует, что дробь $\sqrt{n}(\bar{x} - a)/\sigma$ будет иметь стандартное нормальное распределение. Иными словами, вероятность этой дроби быть меньше z будет равна $\Phi(z)$ или $2\Phi(z) - 1$.

Зададимся какой-либо вероятностью Q, которую целесообразно принять близкой к 1. Чаще всего принимают значения 0,9; 0,95; 0,99. Выберем теперь z так, чтобы $2\Phi(z) - 1 = Q$. Это уравнение можно решить пользуясь таблицей обратной функции Лапласа, упомянутой в § 5.

Событие $|\sqrt{n}(\bar{x} - a)/\sigma < z|$ можно записать в эквивалентной форме $|\bar{x} - a| < |z\sigma/\sqrt{n}|$. Это неравенство, правая часть которого известна, даёт представление о точности приближения $\bar{x}$ к неизвестному a: разность не превосходит $z\sigma/\sqrt{n}$. Отличие от нестатистических приближений лишь в том, что эта точность гарантируется только с определённой вероятностью Q, меньшей единицы. И понятно теперь, почему следует принимать Q близко к единице. При массовых вычислениях подобного рода точность приближения будет не менее $z\sigma/\sqrt{n}$ примерно в $100Q$% случаев. В оставшейся доле случаев приближение $\bar{x}$ к a окажется хуже.

Мы, конечно же, хотели бы гарантировать определённую точность вычислений с возможно более высокой вероятностью[29]; более того, мы хотели бы иметь высокую точность с высокой вероятностью. К сожалению, это невозможно: с увеличением Q быстро увеличивается и z, так что с высокой вероятностью можно гарантировать относительно низкую точность. Чем надёжнее гарантия, тем меньше она гарантирует.

Ясен смысл и σ и n. Чем меньше дисперсия, т. е. разброс в измерениях, тем большей точности в определении неизвестного a можно достичь. Ещё раз подчеркнём, что мерой разброса служит квадратный корень из дисперсии, который и входит в формулу. Также ясно, что увеличение числа наблюдений увеличивает точность. Беда всей математической статистики в том, что увеличение точности пропорционально не n, а $\sqrt{n}$. Для достижения высокой точности необходимы очень большие выборки.

В статистике делают ещё один шаг в преобразовании исходного выражения $|\sqrt{n}(\bar{x} - a)/\sigma| < |z|$. Его записывают в виде $|\bar{x} - z\sigma|/\sqrt{n} < a < |\bar{x} + z\sigma|/\sqrt{n}$. Оказывается, что с вероятностью Q интервал со случайными концами $|\bar{x} - z\sigma|/\sqrt{n}$, $|\bar{x} + z\sigma|/\sqrt{n}$ накрывает неизвестное неслучайное значение a. Саму вероятность Q называют уровнем доверия, а указанный интервал – доверительным для неизвестного a.

Наши предположения о нормальности распределения наблюдений чаще всего выполняются лишь приближённо. Часты также случаи, когда дисперсия неизвестна. Простейший и одновременно наиболее принципиальный способ преодолеть эти осложнения таков. Прежде мы делали вывод о нормальном характере распределения $\sqrt{n}(\bar{x} - a)/\sigma$ на основании свойств нормального закона. Теперь этого нет, но на основании ЦПТ нормальное распределение с параметрами 0 и 1 является хорошим приближением, так что равенство

$$P|\sqrt{n}(\overline{x} - a)/\sigma| < z) = 2\Phi(z) - 1$$

становится приближённым. Точность этого приближения остаётся, правда, неопределённой, так как она зависит от неизвестного закона распределения наблюдений. Будем считать, что число n достаточно велико для того, чтобы точность интересующего нас приближения оказалась удовлетворительной. Но в выражение доверительного интервала входит теперь неизвестное значение σ. Мы заменим его оценкой по той же выборке. Вспомним, что $\mathrm{D}x_i = \mathrm{E}(x_i - \mathrm{E}x_i)^2$. По ЗБЧ среднее арифметическое большого числа наблюдений приблизительно равно ожиданию. Поэтому мы получим приближённое значение дисперсии, если заменим ожидание усреднением по выборке, т. е. в качестве оценки σ^2 примем выборочную дисперсию

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

При больших n это значение s^2 близко к σ^2. (Более точно, разность $\sigma^2 - s^2$ есть величина порядка $1/\sqrt{n}$.) Таким образом, мы получаем доверительный интервал $\overline{x} - zs/\sqrt{n},\ \overline{x} + zs/\sqrt{n}$ с уровнем доверия примерно равным Q. Это приближение тем точнее, чем больше объём выборки.

Вернёмся к случайному выбору без возвращения, о котором см. §§ 1 и 2. Мы показали там, что чисто случайный выбор даёт репрезентативную выборку в том смысле, что доля объектов с определённым свойством А в выборке приближённо равна их доле в генеральной совокупности. Теперь мы можем указать точность этого приближения.

В обозначениях § 5 проведено n испытаний Бернулли с вероятностью успеха p, S – число успехов. По теореме Муавра – Лапласа дробь

$$\frac{S - np}{\sqrt{np(1-p)}}$$

распределена приблизительно нормально с параметрами 0 и 1. Из прежних соображений выбираем Q, по нему находим z и строим приближённый доверительный интервал с уровнем доверия Q

$$|\frac{S}{n} - p| < z\frac{1}{\sqrt{n}}\sqrt{\frac{S}{n}(1 - \frac{S}{n})}.$$

Оговоримся, что это не единственный и не лучший способ построения доверительного интервала для p. Да и в других затронутых случаях существуют более точные и более сложные рекомендации. Мне хотелось только указать принципиальные возможности. И, конечно, строить доверительные интервалы приходится не только с помощью нормального распределения.

7. Оценивание параметров

Каждое встретившееся распределение (биномиальное, показательное, нормальное) было, собственно, целым семейством распределений, зависящим от одного или нескольких параметров. Из условий реальной задачи обычно бывает относительно ясно, с каким семейством мы имеем дело. Но неизвестными остаются конкретные значения параметров распределения, которые входят в любую интересующую нас характеристику. Поэтому необходимо знать хотя бы приближённое значение этих величин.

Примером послужит показательное распределение. Было сказано, что оно может служить приближением для распределения случайного срока службы лампочки[30]. Важной характеристикой для приборов такого рода является вероятность безотказной работы в течение в течение произвольного интервала времени T. Она равна e^{-aT}, т. е. содержит неизвестный параметр. В люстре с несколькими лампочками часть их обычно не горит. Допустим, что в люстре 5 лампочек. Какое время они все будут гореть с вероятностью хотя бы 0,99? Это время можно назвать гарантированным временем работы, если в духе § 3 пренебрегать вероятностью 0,01. Простой подсчёт показывает, что это время в 5 раз меньше, чем аналогичное время службы одной лампочки. В его выражение снова входит a.

Определение a возможно многими методами. Можно, например, включить все 5 и отмечать моменты выхода из строя. Мы получим n (?) значений независимых случайных величин с одним и тем же законом распределения. Но для малых a и большого среднего времени службы $1/a$ придётся долго ждать выхода из строя последнего элемента. Поэтому в некоторых случаях заранее ограничивают длительность эксперимента некоторым T. Его результатом оказывается число вышедших из строя элементов и продолжительность их работы.

При обилии возможных оценок параметра a возникает вопрос: какой же из них следует пользоваться? Что подразумевать под качеством оценки? Оценки являются функциями случайных наблюдений и сами тоже случайны. Теоретически можно установить закон их распределения. В некоторых случаях мерой достоинства оценки можно принять ожидание квадрата разности между ней и истинным значением параметра. Соответствующая ветвь математической статистики называется теорией оценивания и активно развивается.

Фундаментальным следствием ЦПТ является то, что при больших объёмах выборки в типических условиях все разумные оценки имеют приблизительно нормальное распределение. Среднее значение обычно пренебрегаемо мало отличается от неизвестного параметра, а дисперсию можно указать достаточно точно.

Точность оценивания неизвестного параметра можно указывать в духе § 6. Кроме того, упрощается сравнение оценок.

Действительно, вместо сравнения истинных распределений различных оценок можно сравнивать соответствующие нормальные, которые определяются всего двумя параметрами. Обычно приходится сравнивать только дисперсии, потому что средние мало отличаются от истинного значения. Ясно, что из двух оценок лучше та, которой соответствует нормальное распределение с меньшей дисперсией. Более того, среди оценок находятся даже наилучшие в указанном смысле.

Одну из таких асимптотически наилучших оценок позволяет получить метод, который был известен ещё Гауссу[31], метод наибольшего правдоподобия. В 1920-е годы его систематически исследовал Фишер. Мы изложим его применительно к выборке, хотя рамки его целесообразного применения гораздо шире.

Итак, пусть $X_1, X_2, \ldots, X_n$ – выборка, причём плотность отдельного наблюдения $p(x|a)$ зависит от неизвестного параметра a. Природа этого параметра не очень важна; он может быть и числом, и вектором. Ввиду независимости случайных наблюдений плотность вектора $(X_1, X_2, \ldots, X_n)$ в n-мерном пространстве наблюдений в произвольной точке $(x_1, x_2, \ldots, x_n)$ выражается произведением $p_1(x|a)p_2(x|a) \ldots p_n(x|a)$.

Ясно, что плотность вероятности в зависимости от параметра a может быть в данной точке большей или меньшей. Метод наибольшего правдоподобия рекомендует выбирать a так, чтобы плотность приняла наибольшее возможное значение в наблюдённой точке $(x_1, x_2, \ldots, x_n) = (X_1, X_2, \ldots, X_n)$. Ясно, что такой выбор a происходит в зависимости от $(X_1, X_2, \ldots, X_n)$, так что получаемый результат есть функция от наблюдённых значений.

Но совершенно непонятно, почему этот метод приводит к хорошей оценке истинного a. Некоторое время господствовавшие в литературе ссылки на выбор вероятнейшей возможности в таком контексте, который принят сейчас в теории, изложен нами здесь, но они ничего не объясняют. Для Гаусса они значили совсем другое (§ 9). Сейчас, однако, мы попытаемся выяснить, почему этот метод действует.

Фишер доказал, что в определённом смысле оценки наибольшего правдоподобия наилучшим образом используют информацию о параметрах, содержащихся в наблюдениях, и метод наибольшего правдоподобия стал очень популярным. Во многих задачах он даёт хорошие результаты. Но эти задачи столь разнородны, что не покрываются единой теорией, которая описывала бы свойства метода и указывала границы его применимости. Идея наилучшей обработки результатов эксперимента привлекательна логически. Исследования завершились созданной Вальдом в 1940-х годах теорией статистических решений, см. § 10.

Однако, будучи применима последовательно, идея наилучшей обработки результатов эксперимента может привести к неправильным заключениям: предположение о принадлежности неизвестной плотности распределения определённому параметрическому семейству (нормальному, показательному или иному) выполняется лишь приближённо.

Применяя это предположение безоговорочно можно получить результаты, не являющиеся даже приблизительно верными. Это может происходить при определённых, хоть и небольших отклонениях от начальных предположений. Особенно чувствительны к таким нарушениям должны быть оптимальные методы. Вольно выражаясь, они используют всю информацию, ничего не оставляя в качестве запаса прочности.

8. Метод наименьших квадратов

Посмотрим, как Гаусс пришёл к гауссовскому распределению, используя известный ему метод наибольшего правдоподобия[32].

9. Планирование экспериментов

При всяком статистическом исследовании, особенно если оно потребует многих сил, времени, затрат, возникает необходимость в его предварительном планировании. План должен предусматривать получение необходимой информации с наименьшими усилиями и в удобной форме. Впервые потребность в систематическом планировании возникла в научных исследованиях по сельскому хозяйству[33].

Допустим, что мы собираемся испытать r сортов и s способов их возделывания. Пусть m_{ij} обозначает урожай сорта i при способе обработки j с участка единичной площади, $i = 1, 2, \ldots, r$, а $j = 1, 2, \ldots, s$.

Опытное поле делят на единичные участки, на каждом из которых выращивается определённый сорт с применением определённой агротехники. Ввиду случайных причин фактический урожай X_{ij} сорта i при способе возделывания j будет отличаться от m_{ij}. При аккуратно поставленных опытах случайная величина X_{ij} имеет приблизительно нормальное распределение со средним m_{ij}. Дисперсии всех этих случайных величин можно считать равными.

Далее, предположим, что теоретический урожай m_{ij} получается за счёт независимых выборов сорта и способа возделывания, так что $m_{ij}= a_i + b_j$ где a_i – влияние выбора сорта и b_j – влияние способа обработки. Это предположение выполняется далеко не всегда. Разные орта могут требовать различной агротехники. Так, современные высокие нормы внесения искусственных удобрений требуют специальных низкостебельных сортов зерновых, устойчивых к полеганию. Оба фактора поэтому действуют совместно и их сумма не подходит. Впрочем, при испытании относительно схожих сортов и способов возделывания взаимодействие факторов проявляется незначительно и им можно пренебречь.

Целью эксперимента является определение всех величин a_i и b_j. Оценив их, можно, например, выделить лучшие сорта и приёмы обработки. Указанную оценку получают методом наименьших квадратов из условия

$$\min \sum_{ij} (X_{ij} - a_i - b_j)^2.$$

Теперь можно говорить о плане эксперимента в целом. Отдельный опыт состоит в выращивании какого-то сорта i при условиях обработки j. Здесь могут использоваться не все комбинации, притом необязательно, чтобы сочетания не повторялись. Наиболее употребительным и удобным является план, при котором каждая пара (i, j) появлялась ровно один раз. Этот план требует проведения rs опытов, оценки величин очень просты, для них нетрудно вычислить и доверительные интервалы.

В некоторых случаях, в зависимости от исходного материала, можно заметить преимущества тех или иных сортов или методов обработки, или группы методов или сортов. Впрочем, обычно предварительно проверяют, есть ли основания для подобных выводов. Например, проверяют гипотезу об отсутствии преимуществ одного сорта перед другим, т. е. проверяют равенство $a_1 = a_2 = \ldots = a_n$. Если статистический материал не отвергает эту гипотезу, то все наблюдённые различия урожайности можно объяснить одним только действием случайности. Для проверки указанной гипотезы и ей подобных применяют приём, похожий на упомянутый в § 8.

При планировании подобных экспериментов приходится обращать внимание, как на общее число необходимых опытов, так и на простоту обработки численных результатов. Мы рассмотрели действие только двух факторов, но и они не должны быть слишком разнообразны, иначе произведение rs станет чрезмерно большим.

Но весьма правдоподобно, что сама система агротехнических мероприятий, которую мы хотим испытать, распадается на независимые части: сроки внесения и состав минеральных удобрений, сроки сева и нормы высева и т. д. Каждый из этих факторов может принимать несколько значений, и общее число комбинаций окажется очень большим. Поэтому следует перейти к учёту воздействия на урожайность каждого фактора в отдельности, сохраняя предположение о независимости их действия. Существуют планы эксперимента, которые не требуют проведения опытов при всех комбинациях условий и тем не менее дают возможность удобно оценить действие каждого фактора.

Впрочем, ныне планирование эксперимента чаще связывают с другой задачей. Представим себе, что мы хотим найти экстремум функции двух переменных $y = f(x_1, x_2)$. Мы можем только выбирать значения переменных или факторов x_1 и x_2 и узнавать соответствующее значение переменной или отклика y. Например, y – процент выхода продукта в химическом производстве, x_1 и x_2 – условия ведения процесса, скажем, температура и давление. Желателен максимальный процент выхода. В реальных процессах количество управляемых факторов может быть довольно большим, но и в таких случаях оперируют только несколькими главными.

Задача становится статистической из-за того, что наблюдаемые значения отклика отклоняются от $f(x_1, x_2)$ ввиду действия ряда неучтённых факторов. В разных опытах эти случайные отклонения независимы.

Но представим, что этих отклонений нет. Тогда наибольшее значение функции можно было бы искать методом крутого восхождения, как его иногда называют. Допустим, что график функции $f(x_1, x_2)$ есть поверхность в виде холма. Его вершина и будет искомой точкой максимума. Теперь представим, что у подножья холма стоит альпинист, который хочет забраться на самый верх, но не знает дороги, да к тому же и туман. Он видит только небольшое пространство вокруг себя, но заведомо достигнет цели, если в каждой точке своего маршрута будет подниматься по самому крутому пути, а вершина только одна.

Чтобы узнать направление самого крутого подъёма, надо вычислить значение частных производных от функции. В нашей задаче это приходится делать приближённо и вместо, скажем, первой частной производной вычислять

$$\frac{\Delta f}{\Delta x_1} = \frac{f(x_1 + \Delta x_1, x_2) - f(x_1, x_2)}{\Delta x_1}$$

при малом Δx_1. Выбрав таким образом направление, делают по нему несколько шагов пока функция не начнёт убывать. В точке остановки всё повторяют и т. д.

К сожалению, при случайных отклонениях пользоваться указанной формулой нельзя: числитель окажется ошибочным и при малом знаменателе полученные дроби будут сильно отличаться от частных производных. Для поиска направления наиболее крутого подъёма приходится использовать другие соображения.

Вспомним, что малые участки поверхности кажутся почти плоскими. Их поэтому можно заменить участками касательной плоскости, и замечательно, что направления скорейшего подъёма по поверхности и по плоскости совпадают.

Задача планирования опыта возникает при выборе точек измерения, их числа и области их размещения. Область изменения факторов должна быть довольно малой, чтобы различиями между поверхностью и касательной плоскостью можно было бы пренебрегать. С другой стороны, при увеличении этой области уменьшается относительное влияние отклонений функции y от $f(x_1, x_2)$. К сожалению, выбрать компромиссное решение можно только с учётом свойств функции, которые, как правило, неизвестны. Чаще всего интервал изменения каждого фактора выбирается на основании интуиции и опыта исследователя. На плоскости факторов (иксов) это порождает прямоугольник.

Планирование экстремальных экспериментов начало развиваться в 1950-е годы. Применение его идей оказалось исключительно плодотворным, известны случаи увеличения отклика в десятки и даже сотни раз[34].

10. Бейесовский подход

На протяжении всего изложения математической статистики мы считали, что перед исследователем стоИт вполне определённая задача, которую он должен решить по имеющимся статистическим материалам. Конечно же, этот материал следовало использовать как можно лучше.

В некоторых случаях приходится, однако, решать не одну-единственную, а много подобных задач, которые возможно различаются друг от друга значениями параметра. Сам параметр становится переменной величиной, и это следует учитывать. В большинстве подобных случаев переменный параметр обнаруживает определённую статистическую устойчивость, и это позволяет считать его случайной величиной и принимать в расчёт при получении выводов из статистического материала. Учёт подобной массовости при решении задач может в среднем дать выигрыш по сравнению со стратегией наилучших действий в каждом отдельном случае.

Концепцию массового подхода можно распространить и на менее бесспорные случаи, а именно для решения единственной задачи. К ней примысливают статистический ансамбль подобных задач, выбирают способ решения, хорошо действующий в среднем, см. выше. Бывают случаи, когда такой способ найти легче, и его применяют для решения индивидуальной задачи, хоть он и не является для неё оптимальным.

Вот конкретный пример. В течение планового периода (скажем, года) предприятие должно направить другим предприятиям определённые количества продукции в заданные моменты времени. Потребители указывают эти моменты, исходя из своих потребностей, а при нарушении сроков поставщик уплачивает штрафы.

Планирование производства поставщика (которому, в свою очередь, поставляют продукцию иные предприятия) становится очень сложным, притом ему придётся работать по новому плану и устойчивость производства нарушается.

Но к заданной последовательности моментов поставок можно отнестись как к реализации случайного выбора этих моментов, т. е. считать эту задачу представителем статистического ансамбля. Обычно считают, что заявленные моменты поставок образуют чисто случайную и в определённом смысле наименее закономерную последовательность с постоянной средней интенсивностью, так называемый пуассоновский поток. Статистические исследования не отвергают подобной гипотезы, и в первом приближении её можно использовать.

При этом поставщику оказывается гораздо легче выработать оптимальный в среднем план. Удобно, что он уже не зависит от конкретной реализации потока требований. В одном звене затраты могут несколько возрасти (придётся создать определённый резерв), но экономия за счёт устойчивости работы самогО предприятия и его поставщиков перекрывает затраты.

Не следует забывать и об условности первоначальной постановки задачи, т. е. о точной фиксации моментов поставок.

Как правило, потребитель просто не в состоянии обоснованно указать такие моменты за год вперёд, и их выбор весьма произволен. Статистический подход ослабляет эту неприятность.

Сейчас такой подход, при котором неизвестные параметры трактуются как случайные величины, связываются с именем Томаса Бейеса, английского мыслителя XVIII века. Формула пересчёта вероятностей, названная его именем, входит ныне во все учебники[35]. Обсудим её.

Представим себе, что случайный эксперимент может проходить в n различных условиях $\mathrm{H}_1, \ldots, \mathrm{H}_n$. По традиции говорят, что имеется n этих самых гипотез об условиях проведения эксперимента, и, следовательно, о распределениях вероятностей на совокупности его исходов. Пусть A – событие, относящееся к исходу эксперимента, а его вероятность при осуществлении гипотезы H_i записывается в виде $P(\mathrm{A}|\mathrm{H}_i)$. При проведении эксперимента мы не знаем, какие условия действительно осуществляются, но предполагаем, что выбор гипотезы случаен с известными нам вероятностями $P(\mathrm{H}_1), \ldots$ $P(\mathrm{H}_n)$.

Проведём эксперимент и допустим, что он окончился элементарным исходом w. В этом случае условные вероятности гипотез $\mathrm{H}_1, \ldots, \mathrm{H}_n$ будут отличаться от первоначальных вероятностей $P(\mathrm{H}_1), \ldots P(\mathrm{H}_n)$. Формула Бейеса даёт правило пересчёта

$$P(\mathrm{H}_i \mid w) = \frac{P(w \mid \mathrm{H}_i) P(\mathrm{H}_i)}{\sum_{i=1}^{n} P(w \mid \mathrm{H}_i) P(\mathrm{H}_i)}.$$

Подчеркнём ещё раз, что выбор условий проведения эксперимента должен быть случаен в смысле теории вероятностей, а вероятности должны быть известны нам. Такие эксперименты можно найти только в задачниках по теории вероятностей.

В то же время идея пересчёта вероятностей долгое время была господствующей в науке и, ввиду недостаточной чёткости понятий того времени, служила основой математической статистики. Под шансами гипотез понимали не объективно существующие вероятности (которых чаще всего просто нет), а собственное представление об их правдоподобии. В частности, полное отсутствие сведений истолковывалось как равенство шансов, как равновозможность всех гипотез, т. е., по-нашему, как равномерное распределение вероятностей между ними.

Так рассуждал и Гаусс, когда применял математическую статистику для целей астрономии. Для него слова *наиболее правдоподобное значение параметра* имели буквальный смысл. Действительно, если при исходе опыта w вычислить условную вероятность гипотезы H_i

$$P(\mathrm{H}_i \mid w) = \frac{P(w \mid \mathrm{H}_i)P(\mathrm{H}_i)}{P(w)}$$

и считать все вероятности $P(\mathrm{H}_i)$ равными, то наиболее правдоподобной окажется та гипотеза, для которой $P(w|\mathrm{H}_i)$ имеет наибольшее значение. Простоты ради мы приняли конечное число гипотез, на самом же деле параметр принимает непрерывное множество значений. Принципиальную сторону дела это не меняет.

С развитием более чёткого представления о вероятности и её связи с частотами представление о неизвестном как об определяемом случаем становилось неприемлемым. Естествоиспытатели этой концепции не применяли. Никто из них не задавался вероятностями справедливости того или иного закона природы и не перевычислял их после своих экспериментов[36]. Этот метод был широко распространён в литературе и преподавался в университетах, но он не применялся.

Однако, в наше время бейесовский подход переживает вторую молодость. Отчасти это связано с теорией статистических решений, предложенной крупным американским учёным А. Вальдом.

Он пытался проанализировать деятельность естествоиспытателя при обработке статистической информации. Назовём его на английский манер статистиком, хотя в обиходе это слово имеет иной смысл. Работая с материалом, статистик должен сделать какой-то вывод, например, принять или отвергнуть гипотезу, указать приближённое значение параметра. Как говорит Вальд, статистик должен принять решение.

Обозначим наличный статистический материал через X, решение через $d = d(X)$, а реальная ситуация пусть описывается параметром a. Ясно, что при разных a последствия решения могут быть разными. Ущерб от решения d при состоянии a есть функция от d и a, $W(d, a)$. Она называется функцией потерь. При материале X будет принято решение $d(X)$ и потому потери составят случайную величину $W(d(X), a)$. Будем считать, что подобная задача решается многократно, а суммарные потери будут в первую очередь определяться ожиданием величины $W(d(X), a)$. Эти средние потери называются риском и зависят они от правила принятия решения и параметра a.

Естественно искать такое решающее правило $d(X)$, которое даёт наименьший риск при всяком a. К сожалению, таких правил просто не бывает. Большинство правил даже нельзя сравнивать друг с другом: одни лучше при одних значениях a, другие – при других a. Проблема неразрешима, оптимум не существует.

Не задерживаясь на многих интересных мыслях и теориях, высказанных в связи с этим, сделаем следующий шаг. Если предположить теперь, что состояние среды, т. е. параметр a, сам является случайным, то можно будет произвести осреднение по a и вместо несравнимых функций от a получить числа и найти наименьшее из них. Решающее правило, породившее это число, и будет наилучшим, а Вальд назвал его бейесовским решением.

К сожалению, как уже было сказано, неизвестный параметр далеко не всегда разумно представлять реализацией случайной величины. Но теория Вальда неожиданно оказалась полезной при несколько ином толковании.

В последнее время принятие решений в условиях неопределённости интенсивно исследуется. Из ряда возможных способов действия надо выбрать один, но условия, в которых проявятся последствия выбора, не вполне ясны. Мыслить и действовать более чётко помогает концепция персональных вероятностей. Для каждой возможной ситуации мы определяем свою меру правдоподобия, т. е. нашу личную меру её вероятности. Они, как и объективно существующие, будут меняться по формуле Бейеса после получения новой информации и вообще обладать всеми свойствами вероятности, если только они с самого начала были заданы непротиворечиво. Лицу, принимающему решение, предлагается принимать оптимальное бейесовское решение относительно его системы вероятностей.

Здесь нельзя не сказать, что Вальд первым систематически исследовал одну, ранее не упомянутую возможность воздержаться от принятия решения, но дополнительно исследовать проблему. Раздел математической статистики, изучающий эту стратегию, называется последовательным анализом. Исследования в этой области весьма интенсивны, это одна из *точек роста* статистической науки. Накопление информации для принятия достаточно обоснованного решения предполагает более активную роль статистика в эксперименте. Мы указывали во Введении, что статистика отчасти становится наукой о проведении эксперимента. Сейчас мы указали одно из подтверждений этого указания.

11. О возможностях математической статистики

Мы видели, что большая часть утверждений математической статистики обсуждает поведение применяемых процедур обработки данных при (неограниченном) возрастании числа наблюдений. При этом действуют закономерности больших чисел: интересующие нас параметры оцениваются всё более точно, доверительные интервалы для них сужаются, проверяемые гипотезы могут быть всё более тонко отделены от конкурирующих и т. д.

Это может создать иллюзию неограниченных возможностей статистических методов, стОит лишь озаботиться об увеличении статистического материала. К сожалению, всё не так просто. Прежде всего, обратим ещё раз внимание на то, что во все выражения, описывающие точность наших выводов, число наблюдений входит под знаком квадратного корня. Для увеличения точности выводов, например, в 10 раз, объём информации надо увеличить в 100 раз. Но есть и более серьёзные принципиальные трудности. Самое важное: наша математическая модель независимых однородных наблюдений (или какая-либо иная модель) есть лишь приближённое, упрощённое описание реального явления.

Степень соответствия модели реальному материалу может быть удовлетворительной при умеренных объёмах, но при больших объёмах могут проявляться отклонения от этой простой закономерности, вспомните хотя бы трудности случайного выбора. Чаще всего нарушается однородность материала. Опыты порой требуют много времени, в них должны участвовать разные группы исследователей и т. п., они идут в различных условиях. Собранный материал может оказаться неоднородным, наблюдённые значения нельзя считать реализациями одинаково распределённых величин. Наконец, наиболее важным чаще всего оказывается изменение изучаемого явления во времени. Измерения, сделанные в разное время, в сущности, измеряют разные вещи. Для небольших выборок нужно немного времени, изменения самогО явления не очень значительны и пренебрегаемы.

Эти обстоятельства серьёзно препятствуют применению статистических методов. И нельзя надеяться поправить плохие измерения с большой дисперсией увеличением числа наблюдений. Статистик должен это помнить и уделять много времени улучшению самих измерений.

В последнее время заметили, что реальные статистические данные в большей или меньшей степени непременно содержат грубо ошибочные значения[37]. Их обычный источник – ошибки регистрации результатов, переписывания и т. п. Подобное *засорение* данных не так опасно при их обработке вручную: ошибки можно заметить и отбросить, а в сомнительных случаях руководствоваться статистическими критериями[1]. Но, скажем, при машинной обработке наблюдений надо принимать специальные меры.

Приведём пример нахождения центра распределения по выборке. Пусть наблюдения $X_1, \ldots, X_n$ распределены нормально с неизвестными параметрами a и σ^2, и мы хотим определить a. При принятых предположениях наилучший результат – среднее арифметическое наблюдений. Но если несколько наблюдений далеко отстоят от a, среднее также будет сильно отличаться от истинного значения. Впрочем, для выборки из нормального распределения такая возможность маловероятна. Если среди $X_1, \ldots, X_n$ имеются грубо ошибочные или не относящиеся к делу результаты, они могут увести среднее весьма далеко от a, даже если процент засорения выборки невысок.

Имеются, однако, другие оценки, которые даже в этих условиях дают хорошее приближение к истинному центру распределения. Так, выборочная медиана почти не изменяется при произвольном изменении небольшой доли выборки.

Можно сказать, что, в отличие от среднего арифметического, эта оценка устойчива по отношении к засорению выборки.

Современные рекомендации по определению центра советуют отбрасывать 10 – 15% крайних с каждой стороны значений измерений. С остатними можно поступить как угодно, например, взять их среднее арифметическое. Если засорения нет, и принятые предосторожности излишни, качество оценки, разумеется, несколько ухудшится по сравнению с наилучшей. Удивительно, что это ухудшение очень незначительно, даже если отбрасывается существенная доля наблюдений.

Многие известные статистические процедуры чувствительны к тем или иным небольшим нарушениям основных предположений. Поэтому, если имеется подобная опасность, их надо заменить устойчивыми процедурами. Терминология здесь не сложилась, иногда такие процедуры называют крепкими[38].

Любопытно, что как бы возрождается спор Гаусса и Лапласа. Гаусс использовал метод наименьших квадратов, Лаплас же предлагал метод наименьших абсолютных ожиданий. В применении к определению центра нормальной выборки метод Лапласа сводится к выбору выборочной медианы.

Библиография автора

1. А. Вальд. *Последовательный анализ.* М., 1960.

2. Ю. Нейман. *Вводный курс теории вероятностей и математической статистики.* М., 1968.

3. Н. В. Смирнов, И. В. Дунин-Барковский. *Курс теории вероятностей и математической статистики для технических приложений*. М., 1968.

4. А. Хальд. *Математическая статистика с техническими приложениями*. М., 1956.

5. В. В. Налимов. *Вероятностная модель языка.* М., 1974.

6. Л. Н. Большев, Н. В. Смирнов. *Таблицы математической статистики*. М., 1965.

7. К. Ф. Гаусс. *Способ наименьших квадратов. Мемуар о соединении наблюдений*. М., 1859.

8. Б. Л. Ван дер Варден. *Математическая статистика*. М., 1960.

9. J. L. Hodges, Jr, E. L. Lehmann. *Basic Concepts of Probability and Statistics*. Holden Day, 1964.

10. *Journal of the Royal Statistical Society* vol. A116, pt. 1, 1953.

11. В. В. Налимов, Н. А. Чернова. *Статистические методы планирования экстремальных экспериментов*. М., 1965.

12. В. Г. Горский, Ю. П. Адлер. *Планирование промышленных экспериментов.* М., 1974.

13. Р. Фишер. *Статистические методы для исследователей.* М., 1958.

Примечания

1. Автору следовало сослаться на Ахенваля или Шлёцера, см. Шейнин (2019, §§ 7.2.1 и 7.2.2). Этот источник мы будем записывать сокращённо (Ш 2019).

2. Автор несколько раз упоминает несовершенные данные. Следует понимать: данные, искажённые неизбежными погрешностями. Статистические приёмы для обработки применял ещё Кеплер (Ш 2019, § 2.2.4), и, в своеобразной форме, древние астрономы.

3. Неплохой принцип для определения той или иной науки предложил Пирсон (1892/1911, с. 15):

Единство всей науки состоит лишь в её методе, но не материале.

Статистика – не наука, у неё нет своего предмета! Эти неоднократные утверждения можно забыть. Медицинская статистика это приложение статистического метода к медицине, теория ошибок (которую статистики не жалуют) – его же приложение к обработке наблюдений. Математический метод состоит в изучении всё более сложных систем, которые могут и не существовать в природе. Пример такой системы: неименованные натуральные числа.

Чуть выше автор упомянул коллекции определений статистики. Помнится, что авторами такое коллекции были Никитина и др., видимо, под редакцией Налимова (1974). Напрасно автор не сослался на определение математической статистики Колмогорова и Прохорова (1974, с. 480).

4. Автор как-то забыл Гиппарха и Птолемея.

5. Хорошо известно, что драматическая судьба законов Менделя произошла в СССР.

6. Пример слишком сложен и неубедителен. Влияние географической изоляции исследовал ещё Дарвин и быть может не впервые.

7. Основным понятием теории вероятностей является случайная величина, ср. прим. 12.

8. При малом числе испытаний вычисленная частота появления исследуемого события может значительно отличаться от соответствующей вероятности.

9. Автор неизменно ссылается на эксперимент. По нашему мнению, следовало ссылаться на объективную реальность.

10. *ЭВМ* автора мы заменяем *компьютером*.

11. Закономерность отсутствует с самого начала.

12. Следовало привести примеры случайности в природе. Здесь же основным объектом теории вероятностей справедливо названа случайная величина, ср. прим. 7.

13. К *ожиданию* мы не добавляем никакого прилагательного, см. [i, прим. 4.7].

14. Эту формулу, как и соответствующую формулу для ожиданий, можно было бы назвать важной.

15. Но скверные наблюдения никак нельзя улучшить. Сам автор так и говорит в § 11.

16. Это неравенство следует называть по именам Бьенеме и Чебышева.

17. Мы не можем согласиться с этим пояснением. Следовало ввести и обсудить *классическое* определение вероятности (которое, кстати, впервые появилось у Муавра, а не у Лапласа) и обязательно упомянуть Мизеса. Подобных умолчаний в брошюре немало.

18. Снова умолчание. Следовало упомянуть устойчивость.

19. Нужен был пример.

20. Автор проверял *начальную гипотезу*.

21. Крупные естествоиспытатели подделывали свои наблюдения. В 1992 г. о Ньютоне мне написал известнейший американский учёный, ныне покойный К. Труcделл:

Ньютон действительно обманывал, ошибался, пользовался неверными данными.

Но, *Что дозволено Юпитеру, то не дозволено быку*.

22. Следовало ввести понятие об ошибках первого и второго рода.

23. Автор цитирует самое начало мемуара Гаусса 1823 г. по переводу 1859 г., которого мы не смогли установить. Следовало сослаться на перевод 1957 г., см. Гаусс (1823/1957). В прежнем переводе (или в его передаче автором) пропущена оговорка Гаусса, которая касалась погрешностей нанесения штрихов лимба.

24. Не следовало объединять подгонку с грубыми ошибками.

25. Вентцель, в самом конце своей брошюры [i] упоминала 6, а не 10 чисел.

26. Предшественники Лапласа (Якоб Бернулли, Муавр, Бейес) относили возникающую теорию вероятностей к чистой математике, однако он решительно перенёс её в прикладную математику. Это позволило ему добиться фундаментальных достижений и в естествознании, и в математике. По поводу его эвристического доказательства предельных теорем см. Гнеденко и Шейнин (1978, с. 194 – 195).

27. Ляпунов действительно сделал важнейший шаг в доказательстве ЦПТ, но нельзя же было забывать Чебышёва и Маркова, а если судить о теории вероятностей в целом, то обязательно надо было упомянуть Пуассона.

28. Точность реального прибора едва ли может быть известна абсолютно правильно. В таком случае наше предположение означает, что дисперсия известна так точно, что при наших объёмах наблюдений её возможная погрешность не будет иметь значения. Ю. Т.

29. Как и у Вентцель [i, прим. 2.10] мимолётно появилась *высокая* вероятность.

30. На более высоком уровне таким приближением не удовлетворяются и привлекают более богатые семейства плотностей, содержащие большее число параметров, например, распределение Вейбулла. Ю. Т.

31. Принцип наибольшего правдоподобия ввёл Ламберт, и им воспользовался Даниил Бернулли (Ш 2019, § 7.3.1), а затем Гаусс (там же, § 10А.2-2). Каждый поступал независимо от предшествовавших ему. С методической целью следовало бы вводить новые понятия на естественнонаучном языке и только затем переходить к математическому описанию. Это было особенно желательно в § 10 (бейесовский подход).

32. Мы не стали повторять известный материал (Гаусс 1957; Ш 2019, § 10.А.2.), притом автор здесь упустил из вида основной мемуар Гаусса 1823 г.

33. Впервые планирование экспериментов было фактически введено в практической астрономии и геодезии, например, для прокладки рядов триангуляции (Шейнин 2007, § 9).

34. Следовало бы привести пример.

35. Эту формулу автор привёл ниже. Бейес указал только её частный случай, общий же случай ввёл Лаплас (1814, седьмой принцип).

36. Автор ошибся, см. классическую задачу о вероятности последующего восхода Солнца хотя бы в комментариях Прайса к мемуару Бейеса (1764 – 1765).

37. О грубых ошибках и методах их предотвращения статистики и математики могли бы поучиться у геодезистов. По своему опыту второй половины XX в. мне известно, что при прокладке высокоточной триангуляции карандашные записи и зачёркивания запрещались, количество наблюдений, которые отбрасывались в процессе работы, сильно ограничивалось и пр.

Критерии отбраковки наблюдений оказались почти бесполезными, см. Barnett & Lewis, 1978, с. 360):

Основная проблема при изучении отклоняющихся наблюдений остаётся без изменения со времени самых первых исследователей: какое наблюдение считать отклоняющимся, и как поступать с ними.

38. Вначале эти оценки называли робастными (рабский перевод с английского), ныне же их величают устойчивыми.

Библиография составителя

Гнеденко Б. В., Шейнин О. Б. (1978), Теория вероятностей. Глава в книге *Математика XIX века*. М., с. 184 – 240. Редакторы А. Н. Колмогоров, А. П. Юшкевич.

Колмогоров А. Н., Прохоров Ю. В. (1974), Математическая статистика. БСЭ, третье издание, т. 15, с. 480 – 484.

Е. П. Никитина и др. *Коллекция определений термина статистика*. М. В книге Налимов (1974).

Шейнин О. Б. (2007), *История теории ошибок.* Берлин. **S, G,** 25.

--- (2019), *Теория вероятностей. Исторический очерк*. Берлин. **S, G,** 11.

Barnett V., Lewis T. (1978), *Outliers in Statistical Data*. Chichester, 1984.

Bayes T. (1764 – 1765), *Essay towards Solving a Problem in the Doctrine of Chances*. *Phil. Trans. Roy. Soc*., vols. 53 – 54 за 1763 – 1764, pp. 360 – 418; 296 – 325. **S, G,** 14. Первая часть мемуара перепечатана: *Biometrika*, vol 45, 1958, pp. 293 – 315.

Gauss C. F., Гаусс К. Ф. (1823, латин.), Теория комбинаций наблюдений. В книге автора *Избр. геод. соч*,, т. 1. М, 1957, с. 17 – 57.

Laplace P. S., Лаплас П. С. (1814, франц.), Опыт философии теории вероятностей. В книге Ю. В. Прохоров, редактор, *Вероятность и математическая статистика. Энциклопедия*. М., 1999, с. 834 – 863.

Pearson K., Пирсон К. (1892, англ.), *Грамматика науки*. СПБ, 1911.

III

Харальд Крамер

Полвека с теорией вероятностей: наброски воспоминаний

Москва, 1979. Переводчик И. Б. Гуревич

Harald Cramér, Half a century with probability theory. Some personal reflections. *Annals of Prob*., vol. 4, No. 4, 1976, pp. 509 – 546

1. Введение

Предыдущий редактор этого журнала, профессор Роналд Пайк, любезно предложил мне написать воспоминания о моей деятельности за полстолетия, с 1920 по 1970 г., когда я активно работал в области математической теории вероятностей. Реализуя это предложение, я не имел в виду писать историю теории вероятностей. Я остановлюсь лишь на собственных впечатлениях о выполненных работах и о некоторых из тех, кто их выполнял, и в основном лишь на тех задачах, которые были интересны мне лично, но едва упоминаю многие другие направления, даже, возможно, в общем, более интересные.

Статью столь субъективного характера вероятно удобнее начать с кратких сведений о себе и о причинах моего интереса к теории вероятностей. Я родился 25 сентября 1893 г. в Стокгольме и был вторым сыном у своих родителей, состоявших в двоюродном родстве. В течение нескольких веков наши предки проживали на острове Готланд в старинном городке Висбю, но мои родители жили в Стокгольме. Отец был банкиром.

Когда я приступил в 1912 г. к занятиям в Стокгольмском университете, меня почти в равной степени интересовали химия и математика. Моя первая должность в университете, ассистент-исследователь, была связана с биохимией. Она продолжалась в течение года непосредственно перед началом первой мировой войны, и моя первая научная публикация также относилась к этой области.

Вскоре, однако, я понял, что моё настоящее дело это математика. Мне повезло, что у меня был такой учитель математики и друг, как Марсел Рис, молодой венгр, приехавший работать а Институт Миттаг-Леффлера и оставшийся затем в Швеции. Позже он стал профессором Лундского университета. Благодаря ему, я получил образование, соответствовавшее тогдашним *стандартам* математики[1] и вошёл в круг проблем, связанных с мерой Лебега и теорией интегрирования.

Мои собственные исследования были посвящены аналитической теории чисел. Я познакомился с интегралами типа интегралов Фурье, которые очень схожи с теми, что встретились мне позже при изучении связей между распределением вероятностей и его характеристической функции. В 1917 г. я получил степень доктора философии за диссертацию о ряде Дирихле, одному из основных аналитических инструментов теории простых чисел.

Для молодого шведского математика моего поколения, искавшему работу, которая позволила бы ему обеспечить свою семью, было вполне естественно обратиться к страховому делу. Шведские страховые компании традиционно использовали в качестве актуариев высококвалифицированных математиков, и несколько моих университетских друзей получили такую работу. Начав в 1918 г. работу в этой области, я к 1920 г. получил должность актуария в компании, специализировавшейся в области страхования жизни.

Именно обязанности актуария столкнули меня с вероятностными задачами и дали моим математическим интересам новое направление. В 1918 г. я начал знакомиться с доступной мне литературой по теории вероятностей и занялся решением отдельных задач, связанных с математической стороной определения страхового риска. Эти задачи, хоть и имели весьма специфический характер, были очень тесно связаны с теми разделами теории вероятностей, которые впоследствии заняли в ней центральное место, так что несколько вводных замечаний о них я сделаю прямо сейчас.

В качестве результатов деятельности страховой компании на период, скажем, одного года можно рассматривать сумму результатов операций по всем страховым делам. Если допустить, что они статистически независимы, то становится вполне очевидной связь этой задачи с классической центральной предельной теоремой (ЦПТ) теории вероятностей.

Можно, однако, всё страховое дело рассматривать как экономическую систему, развивающуюся (изменяющуюся) во времени и подвергающуюся во все моменты времени случайным возмущениям. Такие системы изучались в некоторых ранних работах, которые теперь считаются предшественниками современной теории случайных процессов.

Обе эти задачи заинтересовали меня уже на этой ранней стадии. В следующем разделе я попытаюсь кратко набросать положение, существовавшее в теории вероятностей до 1920 г., и остановлюсь на истории некоторых конкретных задач.

2. Теория вероятностей до 1920 г.

2.1. Общие замечания, основания. Глазам молодого человека, образованного в области чистой математики и воспитанного согласно тогдашним стандартам математической строгости, представлялась довольно противоречивая картина. Существовал большой классический курс теории вероятностей, написанный Лапласом и впервые опубликованный в 1812 г. Его было интересно и полезно читать, но для *современного* математика он был совершенно нестрогим и, кроме того, на удивление некритическим в том, что касалось оснований и приложений теории. Кстати, быть может интересно вспомнить, что император Наполеон, при котором Лаплас был министром и сенатором, позже отозвался о нём (?) неодобрительно, заметив, что Лаплас внёс в практическую административную деятельность l'esprit des infiniments petits (дух бесконечно малых).

Французские последователи Лапласа, даже математики столь высокого класса как Пуанкаре и Борель также, казалось, не имели намерения строить единую, хорошо организованную теорию на удовлетворительной основе. Книги и статьи, посвящённые вероятностным задачам, за незначительными исключениями страдали очень существенным недостатком: отсутствием математической строгости[2].

За пределами России работы русской школы, в частности, Чебышёва, Маркова и Ляпунова, весьма строгие и очень высокого уровня, были в то время мало известны. Существовавшее тогда положение можно охарактеризовать несколькими цитатами. Английский экономист Кейнс [61] говорил по поводу классической теории вероятностей, что

Для учёных она имела привкус астрологии или алхимии.

Мизес [98] утверждал, что

В настоящее время теория вероятностей не является математической дисциплиной,

тогда как французский вероятностник Поль Леви [85, с. 71] рассказывал в своей автобиографии о первом знакомстве с ней в годы юности:

В определённом смысле этой теории вообще не существовало, её предстояло создать.

Согласно Лапласу и его последователям классическое определение вероятности события через пресловутые *равновозможные случаи* рассматривалось как универсальное и пригодное для использования даже тогда, когда оказывалось невозможным дать чёткое толкование природы этих *случаев*[3]. Предпринимались отдельные попытки преодолеть эти затруднения, но они оказывались неубедительными.

Некоторые пытались построить определение вероятности, исходя из свойств статистической частоты. Мизес [98; 99] ввёл подобное определение на аксиоматической основе. Он рассмотрел последовательность независимых испытаний, выполненных в одинаковых условиях, и постулировал существование предельных значений для частот различных наблюдаемых событий, а также инвариантность этих значений для любой должным образом выбранной подпоследовательности испытаний. У него были и верные последователи, и суровые критики. Поль Леви [85, с. 79] позже выразил своё отношение к этому, заметив, что подобное получение удовлетворительного определения *столь же невозможно, как квадратура круга*.

Меня лично работа Мизеса заинтересовала, хотя и в критическом плане, поскольку я рассчитывал на нечто более приемлемое. Оно должно было появиться, однако в 1920 г. время для этого ещё не пришло.

2.2. ЦПТ. Понятие ЦПТ ввёл Пойа [103]; я ещё вернусь к этой теореме. Согласно нынешней терминологии, эта теорема утверждает, что распределение вероятностей суммы большого числа независимых случайных величин (СВ) при выполнении соответствующих условий будет близка к нормальному (гауссовому). Для очень частного случая теорему доказал Муавр [100][4]. В общем виде её сформулировал Лаплас, однако его доказательство было неполным. Если считать члены суммы малыми *элементарными ошибками*, то эту теорему можно было рассматривать как объяснение возникновения нормального распределения для ошибок наблюдения.

После ряда неудачных попыток, предпринятыми многими с целью найти правильное доказательство общей теоремы, которую сформулировал Лаплас, перспективный подход предложил Чебышёв [115; 116], воспользовавшийся методом моментов. Но первое полное доказательство дал в 1901 г. Ляпунов, который оперировал аналитическим инструментом, ныне известным под именем характеристических функций. Его работа была мало известна за пределами России, но мне очень повезло в этом отношении: я смог познакомиться с заметками Хаусдорфа по поводу его работы[5], и они оказали колоссальное влияние на мою дальнейшую работу в этой области.

2.3. Первые работы по случайным процессам. В первое десятилетие XX века было опубликовано несколько работ, которые для современного читателя являются предтечами теории случайных процессов, созданной в 1930-е годы. Все они были посвящены развитию во времени некоторых переменных процессов, подверженных случайным воздействиям.

Анализируя колебания цен на фондовой бирже, Башелье [2] пришёл к важному частному случаю случайного процесса. С этим же процессом встретился Эйнштейн в своей известной работе [38], посвящённой изучению броуновского движения. Этот процесс был предложен в качестве математического описания непрерывных, хотя и исключительно нерегулярных путей частиц, находящихся в жидкости во взвешенном состоянии и подвергающихся случайным соударениям с молекулами.

В промежутке между появлением работ Башелье и Эйнштейна Филипп Лундберг закончил в Упсале диссертацию [92], (которая была написана на шведском языке). В ней он рассмотрел сугубо дискретные изменения накопленного количества заявлений о выплате страхового возмещения, поданных в страховую компанию. Он пришёл к случайному процессу совершенно иного типа, чем непрерывный процесс, описывающий броуновское движение. *Пуассоновский процесс*, известный в настоящее время благодаря своим многочисленным приложениям, является частным случаем процесса Лундберга при равенстве всех сумм выплат по заявлениям о страховых возмещениях.

При рассмотрении общего случая, развитого также и в ряде последующих работ, Лундберг воспользовался функциональным уравнением, которое представляет собой частный случай известного уравнения для процесса без *последействия*, предложенного Колмогоровым [73] для общего класса непрерывных процессов. Феллер [43] изучил соответствующий класс дискретных процессов, который включал в себя процесс, изучавшийся Лундбергом в качестве частного случая.

В связи с изучением задач телефонной связи Эрланг [39] ввёл пуассоновские процессы, и то же самое сделали Резерфорд и Гейгер [113] применительно к изучению радиоактивного распада.

Все эти пионеры теории случайных процессов использовали математические методы с большей или меньшей степенями нестрогости, но они обладали чудесной способностью интуитивно оперировать понятиями и методами, которые должны были дождаться до 1930-х годов, чтобы оказаться строго обоснованными.

3. Подготовительное десятилетие, 1920 – 1929

3.1. Случайные процессы и предельные теоремы. Летом 1920 г. я провёл некоторое время в Кембридже, занимаясь работой в области теории чисел под руководством крупнейшего математика Г. Х. Харди. Там я встретился с одним американцем, моим сверстником Норбертом Винером. В своей автобиографии [122, с. 59] он отметил как интересное совпадение встречу во Франции с Полем Леви и со мной на протяжении одной поездки в Европу, поскольку наши работы всегда были тесно связаны с тем, чем он сам занимался. Уже в том юном возрасте Норберт производил впечатление *большой личности*. Во время той первой встречи нам не удалось обсудить вероятностные проблемы, но всего лишь три года спустя, в 1923 г., Винер [118] опубликовал свою знаменитую работу, в которой за несколько лет до появления основных работ Колмогорова ввёл вероятностную меру в функциональном пространстве, создав таким образом строгую теорию случайных процессов, используемых для описания броуновского движения. Сейчас эти процессы известны под именем винеровских. Среди других результатов ему удалось показать, что почти все рассматриваемые траектории процесса являются непрерывными функциями, которые не имеют производной ни в одной точке. К сожалению, чтение статьи вызывало очень существенные трудности, и многие, интересовавшиеся этой проблемой, и я в том числе, не смогли оценить её подлинного значения до тех пор, пока выдающиеся работы 1930-х годов не подготовили для этого почву.

Кстати замечу, что моей первой вероятностной работой (написанной по-шведски), была короткая заметка [12], посвящённая пуассоновскому процессу. На основе простого набора необходимых и достаточных условий в ней выводится ныне хорошо известное выражение для соответствующего распределения вероятностей. В этой связи я рассмотрел процесс Лундберга, описывающий страховой риск, но моя работа, посвящённая той же проблеме, была опубликована позже.

Что касается ЦПТ, то меня очень заинтересовали работы Пойа [103] и Линдберга [87]. Пойа, венгерский друг Риса, часто посещал Швецию. Как я уже отмечал, именно он ввёл этот термин. Он ссылался на работу Ляпунова, предложил доказательство [ЦПТ?], основанное на применении характеристических функций, и подчеркнул их аналогию с методами, которые использовались в теории простых чисел[6]. Кроме того, он обсуждал метод моментов, который применил Чебышёв и, в более общем виде, Стильтьес.

Линдберг [87] дал полное доказательство ЦПТ при более общих, чем у Ляпунова, условиях. Он ввёл своё известное условие, см. ниже, и мне очень хотелось лично познакомиться с ним. Такая возможность осуществилась на математическом конгрессе в Хельсинки в 1922 г.

Линдберг был профессором Хельсинкского университета, а кроме того у него была чудесная ферма. Когда его упрекали за недостаточно активное занятие научной работой, он отвечал *На самом-то деле я – фермер*. Если же кому-то случалось заметить, что его ферма управляется не лучшим образом, он отвечал:

Естественно, ведь моё настоящее дело – быть профессором[7].

Он очень нравился мне, и в последующие годы мы часто встречались.

Для задач страхового риска, которыми я занимался, было, однако, недостаточно знать, что некоторое распределение вероятностей приблизительно нормально, необходимо было представление о величине ошибки приближения. Ляпунов установил её верхний предел, и я изучил его метод, упростил его доказательство и немного улучшил полученный им результат [13]. Для особенно важного случая тождественно распределённых переменных мой вариант теоремы Ляпунова выглядел следующим образом.

Пусть $x_1, x_2, \ldots, x_n$ – независимые и одинаково распределённые СВ; все x_i имеют нулевое среднее значение, среднее квадратическое отклонение σ и третий абсолютный момент β_3. Пусть, далее, G_n – функция распределения суммы

$$z_n = \frac{x_1 + x_2 + \ldots + x_n}{\sigma\sqrt{n}}. \qquad (3.1.1)$$

Тогда для всех $n > (\beta_3/\sigma^2)^2$ верно неравенство

$$|G_n(x) - \Phi(x)| < \frac{3\beta_3}{\sigma^3}\frac{\log n}{\sqrt{n}}, \qquad (3.1.2)$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \exp(-t^2/2)dt.$$

Несколько более сложная теорема имеет место для разно распределённых x_i.

Вскоре, однако, выяснилось, что оценка разности, которая определялась выражением (3.1.2), не удовлетворяла численным приложениям к задачам страхового риска, которые я имел в виду. Значительная часть моей работы в 1920-е годы была посвящена попыткам улучшить оценку, предпочтительно в форме асимптотического разложения остатка для больших значений (?).

В доступной мне в то время литературе по математической статистике рассматривались два вида разложения в ряд разности типа (3.1.2). Гнеденко и Колмогоров [48] отметили, что оба предложил Чебышёв, но их часто называют по имени Эджуорта и Шарлье. Я приведу лишь выражение в явном виде, которое Эджуорт [37] рассмотрел сугубо формально. Часто оказывается, что проще иметь дело с рядом Шарлье, который является перегруппировкой членов ряда Эджуорта.

Шарлье [11] утверждал, что его ряд обладает определёнными асимптотическими свойствами, но его доказательство было совершенно неверным, поскольку в своей основе содержало неправильное использование того, что сейчас называется формулой обращения для характеристических функций. Более того, утверждение, которое Шарлье считал доказанным, было справедливо лишь в изменённом виде и при более жёстких, чем у него, условиях. Ср. Крамер [33].

На самом деле асимптотические свойства таких рядов строго не рассматривались, и проблема всё ещё оставалась открытой. В предварительной работе [14] и более полной работе [16] я занялся этой проблемой и получил решение, справедливое при некоторых условиях (ср. [19] и [48, гл. 8]). Для частного случая одинаково распределённых СВ основная теорема в [16] такова[8].

Пусть переменные x_i, входящие в выражение (3.1.1), независимы и одинаково распределены с нулевыми средними значениями, средними квадратическими отклонениями σ и конечными моментами $\alpha_1 = 0$, $\alpha_2 = \sigma^2$, α_3, … , α_k, где $k \geq 3$. Пусть для всех x_i заданы распределение $F(x)$ и характеристическая функция

$$f(t) = \int_{-\infty}^{\infty} e^{itx} dF(x). \qquad (3.1.3)$$

Если

$$\lim\sup |f(t)| < 1, \ t \to \infty, \qquad (3.1.4)$$

то

$$G_n(x) = \Phi(x) + \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) \sum_{j=1}^{k-2} \frac{p_j(x)}{\sqrt{n}} + o(\frac{1}{\sqrt{n^{k-2}}}), \qquad (3.1.5)$$

равномерно по x. Здесь $p_j(x)$ – многочлен степени $3j - 1$, коэффициенты которого зависят только от моментов α_3, …, α_{j+2}. В частности,

$$p_1(x) = \frac{\alpha_3}{3!}(1 - x^2).$$

Поскольку $k \geq 3$, можно убедиться в том, что условие (3.1.4) позволяет исключить коэффициент $\log n$ из верхней оценки ошибки в теореме Ляпунова. Важный частный случай, для которого указанное условие выполняется, возникает, когда функция распределения F содержит абсолютно непрерывную составляющую, не равную тождественно нулю.

Я [16] рассмотрел также случай неодинаково распределённых величин и соответствующее разложение для плотности вероятности $G_n(x)$. Мне также удалось показать, что при невыполнении условия (3.1.4) проинтегрированное среднее ряда

(3.1.5) на малом интервале, содержащем точку x, всё ещё сохраняет необходимые асимптотические свойства.

В связи с этим отмечу, что много лет спустя я [30] рассмотрел асимптотическое разложение $G_n(x)$ в случае, когда предельное распределение является устойчивым и отличным от нормального. При выполнении определённых условий разложение, соответствующее (3.1.5), всё ещё существует, однако его форма сложнее.

3.2. Книга Леви [80]. Планы собственной книги. В то время, когда я был погружён в свою работу над асимптотическими разложениями, появилась книга Леви [80]. Я не мог полностью согласиться с его точкой зрения на основания теории вероятностей, но сразу понял, что появление этой книги стало важнейшей вехой в развитии математической теории вероятностей. Было очевидно, что эта первая попытка, пользуясь математически строгими методами, представить эту теорию как единое целое. Леви впервые систематически изложил теорию СВ, распределений их вероятностей и их характеристических функций. Я уже несколько лет применял эти понятия в своих работах, но представленный им вариант теории позволил мне узнать много нового. Он обсуждал ЦПТ и устойчивые распределения вероятностей[9] и включил чрезвычайно интересную главу о кинетической теории газов.

В 1927 г., во время посещения Англии, я встретился со своим учителем и старым другом Г. Х. Харди. Когда я рассказал ему, что заинтересовался теорией вероятностей, он заметил, что по этому предмету на английском языке нет ни единой удовлетворительной книги, и посоветовал мне таковую написать. Я был более чем готов последовать его совету, хоть и сознавал, что такая работа займёт очень много времени. И действительно, лишь через десять лет мой труд был готов [19].

3.3. Основания. Подходы Мизеса и Леви к основаниям теории вероятностей были различны в своей основе, но я не мог согласиться ни с одним из них. Но у Мизеса [98] содержалось одно общее положение, которое я разделял всем сердцем, хоть мне и казалось, что он сам, когда строил свою общую теорию, не руководствовался вытекающими из него следствиями.

На с. 58 Мизес высказал точку зрения, согласно которой теория вероятностей является

Одной из естественных наук точно так же, как геометрия или теоретическая механика.

Смысл его теории состоит в описании определённых наблюдаемых явлений, причём

Не точно, а с некоторой абстракцией и идеализацией.
Другими словами, теорию вероятностей следует рассматривать

как математическую модель некоторого класса наблюдаемых явлений.

В статье [15], написанной по-шведски, я обратился к этому положению Мизеса. Согласившись с ним, я сделал несколько дополнительных замечаний, из которых мне хотелось бы привести следующее:

Понятие вероятности следует вводить посредством чисто математического определения, из которого математические свойства вероятности и классические теоремы могут быть выведены при помощи чисто математических операций […]. *Никакие возражения против такой теории, кроме чисто математически основанных, не могут быть справедливы. Но следует подчеркнуть, что такая математическая теория не говорит чего-либо о тех реальных событиях, которые будут происходить. Вероятностные формулы не могут определять характера реальных событий точно так же, как формулы классической механики не могут предписывать звёздам осуществлять взаимное притяжение согласно закону Ньютона. Лишь опыт может направлять нас в этом отношении и оценивать приемлемость аппроксимации результатов наблюдений выбранной нами математической моделью*.

Эти замечания всё ещё представляются мне вполне обоснованными[10], и очень приятно, что я опубликовал их за семь лет до того, как Колмогоров формализовал теорию вероятностей.

3.4. Новая русская школа. В конце 1920-х годов стало ясно, что в Советском Союзе интенсивно развиваются работы по теории вероятностей. В замечательной работе Бернштейн [7] рассмотрел распространение ЦПТ на суммы необязательно независимых СВ. Он предложил интересный метод, который позволил работать с подобными случаями, см. ниже.

Основные русские работы того периода были, однако, сделаны двумя совсем молодыми математиками, А. Я. Хинчиным и А. Н. Колмогоровым, которым было суждено стать во главе последующего развития этой области. В совместной работе [70] они доказали знаменитую теорему *трёх рядов*, определяющую необходимые и достаточные условия сходимости ряда, членами которого являются независимые СВ. Вероятность сходимости такого ряда может быть равна только нулю или единице, что представляет собой частный случай так называемого *закона нуля или единицы*, открытого тогда же.

Колмогоров [71] доказал известное неравенство для сумм независимых СВ, которое является существенно улучшенным обобщением хорошо известного элементарного неравенства Чебышёва[11]. Пусть $x_1, \dots x_n$, – независимые СВ с нулевыми средними значениями и конечными (не обязательно равными) средними квадратическими отклонениями. В этом случае неравенство Колмогорова устанавливает, что

$$P(\max \sum_{i=1}^{j} x_i \geq k) \leq \frac{\mathrm{E}(x_1 + \ldots + x_n)^2}{k^2}, \qquad (3.4.1)$$

причём максимум понимается для $j = 1, 2, \ldots, n$.

Доказательство (3.4.1) и других неравенств подобного типа основывалось на квалифицированном использовании условных вероятностей и ожиданий и служило провозвестником общей теории этого раздела теории вероятностей, которую Колмогоров вскоре разработал. Неравенство (3.4.1) служит ценным инструментом во всех исследованиях, посвящённых изучению сумм независимых СВ.

Далее, Колмогоров [72] опубликовал доказательство так называемого закона повторного логарифма, который Хинчин [62] до этого установил для частного случая. Из ЦПТ следует, что для всякой функции $h(n)$, стремящейся к бесконечности по мере возрастания n, вероятность выполнения неравенства $z_n > h(n)$ будет стремиться к нулю по мере стремления n к бесконечности. Здесь z_n – нормированная сумма независимых и тождественно распределённых переменных (3.1.1).

И всё же, рассматривая бесконечную последовательность $z_1, z_2, \ldots$, можно рано или поздно ожидать появления очень больших значений. Закон повторного логарифма даёт точное выражение для этого неопределённого утверждения, которое устанавливает, что соотношение

$$\lim\sup \frac{z_n}{\sqrt{2\log\log n}} = 1, \ n \to \infty \qquad (3.4.2)$$

выполняется с вероятностью 1. Интересно отметить, что частный случай этого утверждения доказал Хинчин [62], который исследовал частоту появления различных цифр в двоичных дробях. Эта задача рассматривалась как исключительно относящаяся к теории меры, которая заинтересовала меня до того, как я начал заниматься теорией вероятностей. Общее утверждение, которое доказал Колмогоров, произвело громадное впечатление и расчистило путь для отождествления вероятности с мерой, которое он должен был вскоре осуществить.

4. Великие перемены: 1930 – 1939

4.1. Стокгольмская группа. Здесь мне ещё в меньшей степени, чем раньше, удастся строго придерживаться хронологического порядка. Среди огромного числа новых идей, выдвинутых в 1930-е годы, можно выделить несколько основных групп, развитие которых будет рассмотрено по-отдельности, неизменно с моей индивидуальной точки зрения. Я начну с нескольких замечаний о собственной деятельности в начале этого *героического периода* математической теории вероятностей.

По инициативе шведских страховых компаний в Стокгольмском университете была учреждена должность профессора по курсу *Актуарная математика и математическая статистика*, и я первым занял её осенью 1929 г. С самого начала и на протяжении нескольких лет мне везло со студентами: я работал с группой честолюбивых и весьма хорошо подготовленных молодых людей.

С глубоким интересом мы следили за новыми работами, которые появлялись за рубежом, и старались внести собственный вклад в развитие *нашей* области математики. В первые годы этого периода мы в основном продолжали работы 1920-х годов, связанные с предельными теоремами и процессами страхового риска. Здесь получение новых интересных результатов казалось вполне реальным.

Осенью 1934 г. наша группа с радостью приняла нового сотрудника, приехавшего к нам из-за рубежа. То были чёрные дни нацистского господства в Германии, когда так много выдающихся учёных покидали эту страну. Уилл Феллер, выгнанный из Кильского университета, приехал, чтобы присоединиться к нашей группе, и остался в Стокгольме (на пять лет). Работая вместе с экономистами, биологами, равно как и с сотрудниками нашей группы вероятностников, он приобрёл в Швеции массу друзей.

Феллер учился в Гёттингене и глубоко проникся добрыми традициями этого математического центра. Мы очень старались найти для него постоянную работу в Швеции, но в те предвоенные годы это было почти невозможно, и с крайним сожалением мы наблюдали за его отъездом в США. Там его ждала выдающаяся карьера. В следующих разделах я остановлюсь на его работах среди нас.

4.2. Основания. Если вернуться к истокам новой эры в математической теории вероятностей, станет ясно, что на самом деле переворот в этой области произошёл после появления книги Колмогорова [75]. В ней он обосновал абстрактную теорию для её использования в качестве математической модели некоторых классов наблюдаемых событий. Основное понятие теории ныне широко известно, это классическая концепция вероятностного пространства (Ω, *A, P*), где Ω – пространство, образованное точками ω, которые рассматриваются как элементарные события, *A* это σ-алгебра множеств в Ω и *P* – вероятностная мера, определённая для всех A-измеримых событий, т. е. для всех множеств S, принадлежащих A.

Сегодня хорошо известен способ введения СВ $x = x(\omega)$ и случайного процесса $x(t) = x(t, \omega)$, где t принадлежит некоторому параметрическому пространству T, но в 1933 г. он был существенно новым. Стало ясно, что при подобном подходе случайный процесс определяет распределение вероятностей в пространстве X всех функций $x(t)$ переменной t. Конечномерным распределением процесса $x(t)$ называется n-мерное совместное распределение случайных переменных $x(t_1), \ldots, x(t_n)$, где $t_1, \ldots, t_n$ – произвольное конечное множество точек. Семейство всех распределений такого рода удовлетворяет некоторым очевидным условиям совместимости.

Одна из основных теорем в книге Колмогорова утверждает, что если задано семейство конечномерных распределений, которое удовлетворяет условиям согласованности, то существует случайный процесс, соответствующий заданным распределениям. Более того, вероятность принадлежности функции $x(t)$ множеству S из пространства функций X однозначно определяется конечномерными распределениями для всех борелевских множеств в X. Иначе говоря, для всех множеств S, принадлежащих наименьшей σ-алгебре множеств в X, содержащей все множества функций, удовлетворяющих конечному множеству неравенств вида $a_i < x(t_i) < b_i$.Это относится к действительным функциям $x(t)$, но расширение на комплексный случай очевидно.

В результате было установлено строгое основание для изучения случайных процессов и начался быстрый количественный и качественный рост работ этого направления. Вскоре, однако, выяснилось, что часто встречаются множества функций, интересных с прикладной точки зрения, но не являющихся борелевскими. Так, для винеровского процесса и процесса страхового риска Лундберга (см. §§ 2 и 3) следовало бы определить вероятность множества всех функций $x(t)$, таких, что $x(t) < a$ при $0 < t < b$, но эти множества не борелевские и конечномерные распределения не определяют однозначно соответствующих вероятностей. Подобные случаи требуют изменения общих определений.

В книге [120] приведён один из вариантов видоизменения для винеровских процессов, причём аналогичный метод можно применить для процесса страхового риска. В серии глубоких работ, результаты которых он объединил в своей выдающейся книге [35], Дуб проанализировал общий случай. Все эти работы, однако, покоятся на основаниях, которые дал Колмогоров.

В его книге содержится глава, посвящённая условным вероятностям и ожиданиям, в которой эти понятия вводятся и анализируются при помощи принципиально нового метода.

Его книга всё ещё остаётся основным документом современной теории вероятностей. Если в 1920 г. можно было говорить о том, что эта теория не является разделом математики (ср. § 2), то после выхода этой книги в 1933 г. выступать с такой точкой зрения было уже невозможно[12].

4.3. Марковские процессы. За два года до этого Колмогоров [73] изложил результаты изучения одного общего класса случайных процессов, которые позже стали известны как марковские. Действительно, они являются естественным обобщением классического понятия марковских цепей. Рассмотрим процесс $x(t)$, где t действительный параметр, представляющий время. Если при любом $t_0 < t_1$ условное распределение $x(t_1)$, связанное с гипотезой $x(t_0) = a$, не зависит ни от какой дополнительной информации о значениях, которые принимает $x(t)$ при $t < t_0$, то $x(t)$ называется марковским процессом.

Колмогоров показал, что распределения вероятностей, связанные с марковским процессом, удовлетворяют некоторым функциональным уравнениям. При выполнении определённых условий непрерывности эти уравнения сводятся к дифференциальным уравнениям в частных производных параболического типа, последние же однозначно определяют соответствующие распределения.

В нашей стокгольмской группе именно Феллер посвятил развитию этой общей теории Колмогорова две работы [42; 43] (вторая была опубликована в первый год его пребывания в США). В них рассматривались дифференциальные уравнения в частных производных для непрерывного процесса и интегро-дифференциальные уравнения, возникающие в дискретном случае. Хорошо известно, что с тех пор марковские процессы стали областью интенсивных исследований.

Работа Колмогорова [73] и её продолжение Феллером глубоко впечатлили меня, но сам я занимался исследованиями в области теории универсальных марковских процессов быть может потому, что не считал себя достаточно знакомым с дифференциальными уравнениями в частных производных. Существует однако, класс марковских процессов, который уже в начале 1930-х годов представлял непосредственный интерес для тех из нас, кто занимался случайными процессами, связанными с теорией страхового риска. Я имею в виду процессы с независимыми приращениями (§ 4.4).

4.4. Процессы с независимыми приращениями. Колмогоров [74] предложил выражение для характеристической функции СВ $x(t)$, соответствующей случайному процессу, который удовлетворяет следующим условиям: **1.** $x(t)$ имеет нулевое среднее значение и конечный момент второго порядка; $x(0) = 0$. **2.** При $0 \le t_0 < t_1 \ldots < t_n$ разности $x(t_i) - x(t_{i-1})$ являются независимыми СВ. **3.** Распределение вероятностей $x(t_i) - x(t_{i-1})$ зависит только от разности $t_i - t_{i-1}$.

Процессы такого типа известны как процессы со стационарными независимыми приращениями. Их исследовал в своей превосходной работе Леви [81] при более общих условиях, которые не требовали существования конечных моментов. Он получил классическое общее выражение для характеристической функции $x(t)$. Другое выражение, иногда более удобное, предложил Хинчин [67].

Мы, в нашей стокгольмской группе, восприняли эти работы как откровение и восторженно их изучали. После появления формулы Колмогорова [74] стало ясно, что винеровский и лундберговский процессы страхового риска характеризуют противоположные частные случаи одного и того же общего выражения. Действительно, при выполнении трёх перечисленных выше условий характеристическая функция

$$f(z,\, t) = \mathrm{E}(e^{izx(t})$$

определяется по формуле Колмогорова:

$$\log f(z,t) = t[-\frac{\sigma^2 z^2}{2} + \int_{-\infty}^{\infty} \frac{e^{izu} - 1 - izu}{u^2} dK(u)].$$

Здесь $\sigma^2 \ge 0$ – константа, а $K(u)$, ограниченная и неубывающая функция, непрерывная при $u = 0$.

Если $K(u)$ тождественно равна нулю, то, очевидно, имеет место винеровский процесс. Леви [81] сделал интересное замечание: если $x(t)$ является почти наверное непрерывной функцией t, то указанный случай неизбежен. Но если $\sigma^2 = 0$ и

$$K(u) = \lambda \int_{-\infty}^{u} v^2 dG(v),$$

где λ положительная константа, а $G(v)$ функция распределения, то имеет место процесс страхового риска, причём заявления о выплате страхового возмещения поступают в соответствии с пуассоновским процессом с параметром λ, а суммы страховых возмещений являются независимыми СВ, каждая из которых характеризуется функцией распределения *G.* В данном случае изменения $x(t)$ сугубо дискретны.

Мы в Стокгольме изучали именно этот случай, и $x(t)$ представляло сумму выплат страховых возмещений, накопленную к моменту t. Задача о разорении, имеющая особое значение для страхового дела, связана с вероятностью того, что на протяжении всего периода $0 < t < T$ окажется $x(t) < a + bt$. Член нашей группы Сегердаль изучил эту проблему в своей диссертации [114] и доказал ряд важных неравенств, связанных с вероятностью разорения.

4.5. Безгранично делимые распределения. Арифметика распределений. Леви предложил считать, что распределение СВ $x = x_1 + x_2,$ где слагаемые независимы, является произведением их распределений и содержит каждое слагаемое в качестве множителя. Если подобное представление только тривиально, то говорят, что распределение неприводимо.

Для $x(t)$ – СВ, связанной с процессом со стационарными независимыми приращениями, Леви [81] показал, что её можно представить суммой произвольного числа независимых СВ с одним и тем же распределением. В таком случае говорят, что распределение $x(t)$ безгранично делимо. Формула Леви определяет общее выражение для распределений такого рода. Нормальное и пуассоновское распределения, равно как и устойчивые распределения вообще принадлежат к этому классу.

Леви предположил, что любой из указанных выше сомножителей нормального распределения должен быть нормальным. Он повторил своё предположение в нескольких последующих работах, указав, что оно верно с весьма высокой вероятностью, но что доказать это он не может. Мне, однако, удалось доказать это [18], а Леви [85, с. 111] высказал сожаление в том, что он сам не сумел этого сделать. Ведь доказательство было основано на непосредственном использовании теории характеристических функций, которой он систематически пользовался. Вскоре Райков [105] доказал соответствующее утверждение для распределения Пуассона.

Хинчин [68] доказал общую теорему, которая утверждала, что любое распределение можно представить в виде произведения безгранично делимого распределения на, самое большее, счётное число неприводимых распределений. В рассмотренном общем случае эта факторизация не единственная.

4.6. Предельные теоремы. Ляпунов и Линдеберг (§ 3) установили достаточные условия ЦПТ для сумм независимых СВ. В интересной книге Хинчин [64] дал сводку основных результатов, полученных к тому времени в этой области. Но открытой ещё оставалась проблема установления необходимых и достаточных условий [ЦПТ], равно как и случая, при котором не предполагается существования конечных моментов случайных слагаемых в указанной сумме. Общую задачу, которую в 1930-е годы пытались решить несколько исследователей, можно сформулировать так.

Пусть $x_1, x_2, \ldots$ независимые СВ. Требуется найти условия, при которых существуют такие константы a_n и b_n, что распределения вероятностей нормированных сумм

$$(x_1 + \ldots + x_n - a_n)/b_n$$

приближаются к нормальному распределению по мере стремления n к бесконечности.

Важный вклад в изучение этой проблемы принадлежит Леви и Хинчину, но лишь Феллеру [41] удалось впервые получить полное решение. Его работа состоит из двух частей и была написана во время его пребывания в нашей стокгольмской вероятностной группе.

Феллер показал, что искомые необходимые и достаточные условия можно получить подходящим видоизменением достаточного условия Линдеберга [87]. В частном случае, когда x_i имеют функции распределения F_i с нулевыми средними значениями и такими конечными средними квадратическими отклонениями σ_i, что $\sigma_i/s_n \to 0$ при $n \to \infty$ и $i = 1, 2, \ldots, n,$ можно положить $a_n = 0$ и $b_n = s_n$. Здесь, как обычно,

$$s_n^2 = \sigma_1^2 + \ldots + \sigma_n^2.$$

Феллер доказал, что необходимым и достаточным для сходимости к нормальному распределению является исходное условие Линдеберга

$$\lim \frac{1}{s_n^2} \sum_{i=1}^{n} \int x^2 dF_i(x) = 0.$$

Здесь $n \to \infty$ и область интегрирования определяется неравенством $|x| > \varepsilon s_n$ при любом заданном $\varepsilon > 0$.

Феллер получил также более сложные условия для случая, при котором x_i не имеют конечных моментов.

Выдающийся итальянский математик Ф. П. Кантелли работал нал математическими проблемами актуариев, аналогичным тем, которыми занимались некоторые члены стокгольмской группы. Он [10] предложил улучшенный вариант закона повторного логарифма (§ 3). В переписке с ним я [17] рассмотрел изменённый вариант этой задачи, результат которой мог иметь *предельно возможную точность*.

Утверждение подобного типа для классического закона повторного логарифма в виде (3.4.2) доказал Феллер [44]. Приведу для сравнения теорему Феллера [44], а затем свою [17], причём в обоих случаях без подробного перечня соответствующих условий.

В приведённых обозначениях Феллер доказал, что для всех достаточно больших n вероятность

$$P(|x_1 + \ldots + x_n| < \mu_n s_n)$$

равна 1 или 0 в зависимости от того, сходится или расходится ряд

$$\sum_{n=1}^{\infty} \frac{\sigma_k^2}{s_k^2} \mu_n \exp(-\frac{\mu_n^2}{2}) .$$

В моей работе эта задача ставилась несколько иначе. Вместо простой последовательности независимых СВ $x_1, x_2, \ldots$ я рассмотрел двойную последовательность $x_{11}, x_{21}, x_{22}, \ldots, x_{n1}, x_{n2}, \ldots, x_{nn}$, все члены которой независимы, имеют нулевые средние значения и конечные средние квадратические отклонения σ_{ij}. Обозначив

$$s_n^2 = \sigma_{n1}^2 + \ldots + \sigma_{nn}^2 ,$$

я показал, что для всех достаточно больших n вероятность

$$P(|x_{n1} + \ldots + x_{nn}| < \lambda_n s_n)$$

равна 1 или 0 в зависимости от сходимости или расходимости ряда

$$\sum_{n=1}^{\infty} \frac{1}{\lambda^n} \exp(-\frac{\lambda_n^2}{2}).$$

Хинчин [67] привёл аналогичное, но значительно более существенное обобщение задачи предельных распределений для суммы независимых СВ в целом (?), затем последовали работы Гнеденко [47] и других русских авторов. Я остановлюсь лишь на основных направлениях их деятельности, подробности которой можно найти в интересной книге Гнеденко и Колмогорова [48].

Рассмотрим двойную последовательность СВ

$x_{11}, x_{12}, \ldots, x_{1k_1}, \ \ldots,$

$x_{n1}, x_{n2}, \ldots, x_{nk_n}.$

Все СВ, находящиеся в одной и той же строке предполагаются независимыми. Пусть z_n обозначает сумму СВ строки n. Допустим, что для любого $\varepsilon > 0$

$$\lim P(|x_{nk}| > \varepsilon, n \to \infty, k = 1, \ldots, k_n.$$

Для таких случаев Хинчин [67] получил следующий фундаментальный результат. Для того, чтобы F являлась предельной функцией распределения СВ $z_n - b_n$, полученных из указанной двойной последовательности, где b_n произвольная константа, необходимо и достаточно, чтобы эта функция была безгранично делимой. Класс всех возможных предельных законов, связанных с суммами рассматриваемого типа, таким образом совпадает с классом всех безгранично делимых функций распределения. Едва ли стОит добавлять, что мы в нашей стокгольмской группе с огромным интересом следили за развитием новых идей нашими русскими коллегами.

4.7. Характеристические функции. Использование аналитического инструмента, более или менее эквивалентного известному нам под названием характеристической функции, восходит к временам Лагранжа, Лапласа и Коши. Ляпунов пользовался им при доказательстве ЦПТ.

Только Леви [80], см. § 3, первым систематически изложил теорию характеристических функций. Я пользовался этим методом в своей работе по асимптотическим разложениям, и на протяжении 1930-х годов мы в стокгольмской группе пытались развивать это направление. Результатом было полученное мной доказательство предложения Леви, связанное с нормальным распределением, см. § 4.4. В совместной работе [35], посвящённой многомерным распределениям, мы рассмотрели обобщения характеристических функций и ЦПТ.

В том же 1936 г. я завершил работу над книгой по математической теории вероятностей [19], которую в 1927 г. побудил меня написать Харди. В её основу я положил аксиоматику Колмогорова, и главным образом книга была посвящена подробному изложению теории распределений вероятностей в конечномерных пространствах и их характеристических функций. Материал был изложен применительно к таким проблемам, как ЦПТ, соответствующие асимптотические разложения и случайные процессы с независимыми приращениями.

В общей теории характеристических функций мне удалось несколько дополнить результаты Леви. К сожалению, я допустил досадную ошибку, которую удалось исправить только в последующих изданиях книги (1963 и 1970 гг.). Я остановлюсь на этом несколько подробнее.

Известная *теорема непрерывности* для характеристических функций утверждает, что последовательность функций распределения F_n сходится к функции распределения F во всех точках непрерывности последней в том и только в том случае, если соответствующие характеристические функции $f_n(t)$ сходятся при всех t к предельной функции, непрерывной при $t = 0$. В утверждении теоремы в первом издании книги требовалась лишь сходимость $f_n(t)$ в некотором конечном интервале $|t| < c$. Об этой ошибке мне сообщил Хинчин в письме. Достаточно указать, что можно найти две характеристические функции $f_1(t)$ и $f_2(t)$, равные при $|t| < 1$, но не равные тождественно.

Действительно, рассмотрим

$$f_1(t) = f_2(t) = 1 - |t|$$

при $|t| \le 1$ и пусть $f_1(t)$ – периодическая функция с периодом 2 и $f_2(t) = 0$ при $|t| > 1$. Легко убедиться в том, что эти функции характеристические. Последовательность, члены которой поочерёдно равны f_1 и f_2, сходятся при $|t| < 1$, но не при любых t, откуда следует ложность утверждения первого издания книги.

Добавлю, что я [21] привёл ряд теорем о представлении функций интегралом Фурье. Помимо других результатов, там содержалось упрощение теоремы Бохнера, которая определяла необходимые и достаточные условия для того, чтобы данная функция являлась характеристической для некоторого распределения вероятностей.

4.8. Стационарные процессы. Мне придётся отступить на несколько лет назад и остановиться на новом направлении развития нашей науки, возникшем в 1930-е годы. Хинчин [65] опубликовал основополагающую работу, в которой ввёл класс стационарных случайных процессов. Он указал, что марковскими процессами нельзя пользоваться, если предыстория рассматриваемой системы оказывает существенное влияние на прогноз её развития, что, например, имеет место в статистической механике. В качестве подходящего инструмента для изучения подобных систем он предложил класс стационарных процессов.

Хинчин определил как класс стационарных в узком смысле процессов, так и класс процессов, которые я буду просто называть стационарными. Процесс $x(t)$ с непрерывным временнЫм параметром t называется стационарным в узком смысле, если соответствующие конечномерные распределения инвариантны относительно сдвигов во времени. Процесс называется стационарным, если эта инвариантность сохраняется для его моментов первого и второго порядка.

Изучая результаты основополагающей статьи Хинчина [65] и дальнейшие достижения в этой области, я впоследствии пришёл к рассмотрению общего случая комплексного процесса $x(t)$, непрерывного в среднеквадратическом смысле и такого, что

$$\mathrm{E}x(t) = 0,\ \mathrm{E}x(t)\overline{x}(\overline{u}) = r(t-u).$$

Ковариационная функция $r(t)$ непрерывна для всех действительных t, и Хинчин показал, что она допускает спектральное представление

$$r(t) = \int_{-\infty}^{\infty} e^{itu} dF(u), \qquad (4.8.1)$$

где спектральная функция $F(u)$ действительная, неубывающая и ограниченная. На основе такого представления он вывел ряд свойств функции $r(t)$ и доказал, что среднее по времени

$$\frac{1}{T}\int_{0}^{T} x(t)dt$$

сходится в среднеквадратическом смысле при $T \to \infty$, что эквивалентно эргодической теореме фон Неймана.

Несколько раньше Хинчин [63] рассмотрел стационарный в узком смысле процесс с дискретным временем, т. е. последовательность СВ $\ldots, x_{-1}, x_0, x_1, \ldots$, удовлетворяющую условиям стационарности в узком смысле. Он доказал, что даже среднее по времени

$$\frac{1}{N}\sum_{n=1}^{N} x_n$$

сходится с вероятностью 1, что для такого процесса эквивалентно эргодической теореме Биркгофа. Соответствующий результат для процессов с непрерывным временем несколько позже доказал Колмогоров [76].

Существуют интересные связи между теорией стационарных процессов Хинчина и более ранней работой Норберта Винера по обобщённому гармоническому анализу. В обширной работе Винер [119] рассмотрел комплексные функции f, предел которых

$$\lim \frac{1}{2T}\int_{-T}^{T} f(u+t)\overline{f}(\overline{u})du = r(t)\,,\ T \to \infty \tag{4.8.2}$$

существует при всех действительных t. Свойства предельной функции r очень близки свойствам ковариационной функции стационарного процесса. В частности, она непрерывна при любых t, если непрерывна при $t = 0$, и в этом случае допускает спектральное представление вида (4.8.1). Есть основания ожидать, что винеровское соотношение (4.8.2) будет в принципе справедливо для реализаций стационарного процесса.

Эту работу Винера как и статью [118] о дифференциальном пространстве понять нелегко, но Маслин [95] очень доступно её изложил с современных позиций. Он показал, в частности, что Винер рассмотрел пример объекта, который был позже назван нормальным стационарным процессом. Любая его реализация с вероятностью единица удовлетворяет соотношению Винера (4.8.2).

Хотел бы, однако, заметить, что нетрудно найти пример, в котором это свойство отсутствует. В самом деле, пусть нормальный действительный стационарный процесс имеет периодическую ковариационную функцию с периодом 2 и пусть $r(t) = 1 - |t|$ при $|t| \le 1$. В этом случае можно показать, что для любого заданного t существует множество реализаций рассматриваемого процесса с положительной мерой и таких, что первый член винеровского соотношения (4.8.2) не сходится к конечному пределу при $T \to \infty$.

Работа Хинчина [65] по стационарным процессам оказала очень существенное влияние на последующее развитие. Было очевидно, что этот новый тип случайных процессов является подходящей математической моделью не только в статистической механике, но и, например, в метеорологии и экономике. Введение таких процессов открыло, в частности, новые возможности для исследования явлений, имеющих склонность к периодичности поведения и для работ в области теории информации.

В нашей стокгольмской группе этой областью занялся Херман Волд, который в своей диссертации [123] рассмотрел стационарные процессы с дискретным временем, т. е. стационарные последовательности СВ. Он изучил их ковариационные и спектральные свойства и показал возможности их приложения к ряду статистических задач.

Наиболее значительным из его результатов явилось доказательство важной теоремы разложения для классов процессов такого типа. Позже выяснилось, что с некоторыми изменениями она справедлива для процессов более общего типа. Он доказал, что для стационарной последовательности x_n существует единственное разложение

$$x_n = u_n + v_n,$$

где u_n и v_n такие взаимно независимые стационарные последовательности, что, на современном языке, u_n абсолютно недетерминированна, а v_n детерминированна. Кроме того, последовательность u_n можно представить как

$$u_n = \sum_{i=-\infty}^{n} c_{n-i} z_i,$$

где z_i взаимно независимые СВ, а c_i неслучайные константы.

Я [22] исследовал ковариационные свойства стационарного векторного процесса $\mathbf{x}(t) = (x_1(t);\ \dots,\ x_n(t))$ с непрерывным временем t и доказал, что спектральное представление смешанных ковариационных функций имеет вид

$$r_{mn}(t) = \mathrm{E}x_m(t+u)\overline{x}_n(\overline{u}) = \int_{-\infty}^{\infty} e^{iu} dF_{mn}(u), \qquad (4.8.3)$$

где F_{mn} такие комплексные функции с ограниченной областью изменения, что приращения ΔF_{mn} в любом интервале образуют неотрицательную эрмитову матрицу.

Это означает, что функция матрицы $\mathbf{F}$ с элементами F_{mn} не убывает. Она представляет собой сумму трёх составляющих: абсолютно непрерывной, чисто дискретной и сингулярной. Я также доказал наличие соответствующих свойств у векторного процесса с дискретным временем.

4.9. Париж, Лондон и Женева, 1937 – 1939. Весной 1937 г. меня пригласили в Париж, чтобы прочесть несколько лекций в Сорбонне. Из французских вероятностников я был знаком с Фреше, но это была моя первая встреча с Полем Леви и рядом представителей более молодого поколения, в частности, с Дёблиным, Дюге, Форте и Лоэвом. В молодые годы Фреше был выдающимся математиком, одним из пионеров функционального анализа. Теорией вероятностей он занялся уже в довольно преклонном возрасте, и я вынужден отметить, что его результаты не произвели на меня глубокого впечатления.

С другой стороны, в 1937 г. было совершенно очевидно, что Леви является одним из лидеров в области теории вероятностей, особенно после появления его книги [83]. Кстати, в предисловии к ней он отметил, что решил написать её после того, как в начале 1936 г. получил моё письмо с доказательством его предположения о нормальном распределении. Среди молодых учёных выделялся Дёблин, который уже тогда получил выдающиеся результаты. Его гибель в первые месяцы войны была большой потерей для науки[13].

Позже в том же 1937 г. в Женеве состоялась конференция по теории вероятностей. Из Стокгольма на ней присутствовали Феллер и я. Зрелище такой большой группы собравшихся именитых вероятностников произвело на меня очень сильное впечатление. Среди моих новых знакомых были Штейнгауз из Польши, Хопф из Германии и Ежи Нейман, ещё работавший в Англии. Несколько коллег из Советского Союза также приняли приглашение и сообщили нам темы докладов, с которыми они собирались выступить, но к нашему величайшему сожалению никто из них не приехал[14].

Нейман прочёл доклад, посвящённый разработанный им теории доверительных интервалов, бывших в то время совершенно новым понятием. На той ранней стадии его идеи ещё не были приведены к окончательному виду, однако (?) и Фреше, и Леви отнеслись к ним весьма критически. Я уже успел прочесть его работу [101], посвящённую этой проблеме, и заключил, что его основные идеи были совершенно справедливы, что затем и подтвердилось.

Феллер рассказал об аксиоматике, а я [20] доложил о *больших отклонениях*, связанных с ЦПТ. Если в обозначениях § 3 допустить, что $x \to \infty$ при $n \to \infty$, то и $G_n(x)$ и $\Phi(x) \to 1$, так что утверждение теоремы Ляпунова становится очевидным. В моей работе были рассмотрены отношения

$$\frac{1-G_n(x)}{1-\Phi(x)} \text{ и } \frac{G_n(x)}{\Phi(x)},$$

где x и n стремятся к ∞ в первом случае и к $-\infty$ во втором.

Я показал, что при выполнении определённых условий в каждом случае можно найти асимптотическое разложение.

Первое обобщение этих результатов получил Феллер [44], который использовал их для улучшения закона повторного логарифма (§ 3.4). Дальнейшие весьма существенные обобщения получил Ю. В. Линник и его ленинградская группа, см. обзор их работ в прекрасной книге [55].

В октябре 1938 г. я гостил в Англии. Это было вскоре после мюнхенской конференции, и вопрос *мир или война* уже занимал все умы. В Кембридже я встретился со своим старым учителем и другом Харди, который занимал квартиру Ньютона в Тринити-колледже. Он одобрил учебник, который я написал по его предложению. В Лондоне меня приняли Фишер, Уильям Элдертон и Эгон Пирсон, и я познакомился с некоторыми работами по математической статистике, которыми они занимались. Нейман к тому времени находился уже в Калифорнии.

Летом 1939 г. в Женеве снова состоялась конференция, на этот раз посвящённая математической статистике. Я был счастлив познакомиться с Сэмом Уилксом и Морисом Бартлеттом, которые прочли интересные доклады. Нейман не приехал, но представил основной доклад [102] по теории проверки статистических гипотез, которую он сформулировал незадолго до конференции в сотрудничестве с Эгоном Пирсоном.

Среди событий этой конференции мне вспоминается беседа с Фишером. Я выразил ему своё восхищение по поводу геометрической интуиции, которую он проявил в связи с распределениями вероятностей в многомерных пространствах. Он довольно кисло ответил: *Меня иногда обвиняют в интуиции как в преступлении*!

Когда в июле 1939 г. я покидал Женеву, было совершенно ясно, что война стремительно надвигается.

5. Военные годы: 1940 – 1945

5.1. Изоляция в Швеции. Во время войны наши международные контакты были сведены к минимуму. Швеция оставалась нейтральной, но вокруг шла война. Нацисты оккупировали Данию и Норвегию, а Финляндия воевала с Россией. Было крайне мало возможностей обмениваться литературой или письмами с коллегами в Англии, Франции и США, и совсем никаких – в Советском Союзе. Тем не менее, мы пытались развивать свои исследования настолько, насколько это было возможно.

В своей диссертации [93] член нашей стокгольмской группы Уве Лундберг исследовал новый класс случайных процессов и применил его при решении ряда статистических задач, связанных со страхованием, не относившихся к страхованию жизни. Основываясь на выдвинутых им принципах, актуарии во многих странах интенсивно развивали эту тему.

В феврале 1941 г. я организовал в Стокгольме конференцию по математической теории вероятностей. Мы с радостью приветствовали нескольких гостей из Дании и Финляндии. В Норвегии нацистский оккупационный режим был суровее, и никто из наших коллег не смог оттуда приехать. Гаральд Бор и Берге Иессен из Дании прочитали доклад об основаниях теории вероятностей и её связях с математическим анализом, а Густав Эльфвинг из Финляндии о марковских процессах.

Несколько членов нашей стокгольмской группы доложили о результатах своих диссертаций (см. также выше), а я сообщил о теореме, связанной со спектральным представлением стационарного случайного процесса, которую доказал непосредственно перед этой конференцией. Если $x(t)$ стационарный процесс (§ 4) с ковариационной функцией, имеющей спектральное представление вида (4.8.1), то существует такой процесс с ортогональными приращениями $z(u)$, что при соответствующем задании стохастического интеграла

$$x(t) = \int_{-\infty}^{\infty} e^{itu} dz(u), \qquad (5.1.1)$$

у которого, в стандартных обозначениях, оказывается

$$\mathrm{E}dz(u) = 0,\ \mathrm{E}[dz(u)]^2 = dF(u).$$

Это показывает, каким образом процесс $x(t)$ аддитивно строится из элементарных гармонических колебаний $e^{itu}dz(u)$, каждое из которых имеет угловую частоту u, тогда как амплитуда и фаза являются СВ, которые определяются по $dz(u)$.

Я [23] опубликовал работу по спектральному представлению (5.1.1), в которой определённо указал его связь с винеровским обобщённым гармоническим анализом. Мне, однако, ещё не было ясно, что этот результат на самом деле означает установление вероятностного варианта теоремы Стоуна о спектральном представлении унитарной группы в гильбертовом пространстве. Метод, которым я пользовался, предусматривал применение стохастических интегралов Фурье, но не теории этого пространства. Фундаментальное значение этой теории стало понятно нам лишь в послевоенные годы.

Арлей, молодой датский физик, написал диссертацию [1], посвящённую приложению случайных процессов к теории космического излучения. К моему величайшему удивлению весной 1943 г. я получил разрешение приехать в Копенгаген в качестве члена экзаменационного комитета. Диссертация была хорошей, и мне было очень приятно снова увидеться с датскими коллегами, но больно было наблюдать немецкие войска, марширующие по улицам Копенгагена.

Эссеен подготовил в Упсале диссертацию [40] по Фурье-анализу функций распределения. То была очень существенная работа, основанная на глубоком изучении свойств характеристических функций. Он, в частности, показал, что коэффициент $\log n$ в верхней оценке (3.1.2) остатка в теореме Ляпунова можно всегда опускать и таким образом обобщил мой результат, справедливый лишь при выполнении условия (3.1.4), который упоминался в связи с (3.1.5). Эссеен получил также оценки остатка, зависящие и от n, и от x. И он улучшил моё [16] асимптотическое разложение.

Казалось, что конец войны ещё далёк, и я решил использовать годы вынужденной изоляции, чтобы написать книгу [24]. Она должна была быть закончена в 1946 г. Процитирую следующие строки из Предисловия[15].

Таким образом, эту книгу не следует считать вкладом в математическую теорию вероятностей, скорее она является трудом, посвящённым использованию теории вероятностей в современных статистических методах. Книга посвящена моей жене, которая, как всегда, поддерживала и ободряла меня на протяжении всей работы. Когда я ещё писал эту книгу, я как-то сказал ей, что надеюсь, что эта книга станет моим входным билетом в новый послевоенный мир. Вероятно, в какой-то степени так оно и было: сегодня она издана на английском, русском, испанском, польском и японском языках.

5.2. Развитие исследований в других странах в военные годы. В тех редких случаях, когда из США приходила почта, мы получали письма и оттиски от Феллера и могли следить за его работой над марковскими процессами и усовершенствованием закона повторного логарифма (§ 4).

Многие математики в воюющих странах были связаны с работами по управлению зенитным огнём и фильтрацией помех радиолокаторов. Оказалось, что стационарные случайные процессы являются эффективным инструментом для решения этих задач. В частности, возможность предсказания развития такого процесса на основе прошлых наблюдений имела решающее значение. Независимо друг от друга Колмогоров в СССР и Винер в США получили в этой области важные результаты. Судя по всему, до послевоенных лет они не имели ни малейшего понятия о работах друг друга.

Колмогоров [77; 78] рассмотрел стационарные случайные процессы с дискретным временем и показал, что класс всех СВ с конечными моментами второго порядка образует гильбертово пространство, если скалярное произведение в двух точках определено в виде ковариации соответствующих СВ. В таком случае случайный процесс с дискретным временем можно рассматривать как последовательность точек гильбертова пространства, теория которого привлечена для изучения таких процессов.

Колмогоров показал также, что применение теории гильбертова пространства к стационарным последовательностям СВ позволяет без затруднений получить все известные результаты как разложение Волда и ковариационные характеристики векторного процесса с дискретным временем. Я [22], см. § 4.8, описал эти характеристики.

Но Колмогоров кроме того воспользовался мощными методами теории функций комплексных переменных, чтобы впервые получить необходимое и достаточное условие полной недетерминированности случайной стационарной последовательности (регулярности на его языке). Наконец, Колмогоров полностью решил задачу линейного предсказания по методу наименьших квадратов. Засухин [127] развил работу Колмогорова, рассмотрев стационарный векторный процесс с дискретным временем, и получил ряд важных результатов.

Фундаментальное значение работы Колмогорова определяется тем, что он показал, каким образом абстрактная теория гильбертова пространства (как, разумеется, и пространств других типов) может быть применена в теории СВ и случайных процессов. Это оказало чрезвычайно сильное влияние на дальнейшее развитие последней.

Работа, выполненная в годы войны Винером, была связана с линейным предсказанием и фильтрацией для стационарных процессов и с дискретным, и с непрерывным временем. Для предсказания значения стационарного процесса $x(t)$ с непрерывным временем в момент $t = h > 0$, исходя из наблюдений предыдущих значений процесса вплоть до $t = 0$, Винер ввёл формулу предсказания вида

$$x^*(h) = \int_{-\infty}^{0} x(t)dK(t),$$

где K – функция с ограниченной областью изменения. Он показал, как определить K, чтобы минимизировать среднеквадратическую ошибку предсказания

$$\mathrm{E}[x^*(h) - x(h)]^2.$$

Это неполное решение математической проблемы линейного предсказания, которое, однако, имеет важнейшие приложения в ряде технических задач. В 1942 г. Винер закончил работу над этой задачей и связанных с ней задач. Несколько последующих лет она распространялась в литографических копиях и стала известной под именем *жёлтая опасность*[16] ввиду сложности использованного в ней математического аппарата. Она была опубликована лишь в 1949 г. [121].

6. Послевоенные годы: 1946 – 1970

6.1. Вводные замечания. После нарушения всех связей снова постепенно стала появляться возможность научно-исследовательской работы на международной основе и возобновления контактов с коллегами в других странах. Теперь всем, кто имел отношение к математической теории вероятностей, было ясно, что за 25 лет между 1920 и 1945 гг. она радикально изменилась. В 1920 г. она едва ли заслуживала название математической теории, но в 1945 г. она вступила в послевоенный мир как хорошо организованный раздел чистой математики с собственными задачами и методами, с постоянно расширяющимися сферами приложения в других науках, равно как и в различных видах практической деятельности. Существовали глубокие взаимосвязи между чисто математической теорией и приложениями, и отдельный исследователь уже вряд ли мог охватить всю эту область целиком.

Послевоенные годы ознаменовались дальнейшим интенсивным развитием в самых различных направлениях. Ясно, что в воспоминаниях, подобно моим нынешним, ещё в меньшей степени, чем для периода до 1945 г., можно полностью обозреть развитие этого послевоенного периода. Я вынужденно ограничусь только теми разделами, с которыми был более или менее тесно связан и лишь теми людьми, с которыми поддерживал личные контакты.

Начну с краткого перечня своих контактов с коллегами непосредственно после войны, затем перейду к обзору последующего развития теории вероятностей, основываясь, как всегда, на своих личных впечатлениях.

6.2. Париж, Принстон, Йельский университет, Беркли, 1946 – 1947. Вскоре после войны я получил приглашения в перечисленные университеты. Весной 1946 г. в Париже я должен был прочесть цикл из пяти лекций: две по статистическому оцениванию и три по стационарным процессам. Затем я занял должность приглашённого профессора в Принстоне (осенний семестр 1946 г.), Йеле[17] (весенний семестр 1947 г.) и Беркли (очередной летний семестр).

В Париже я с радостью снова встретился с теми, с кем подружился в 1937 г. за исключением юного Дёблина, убитого в начале войны[18]. Квартиру Леви нацисты разграбили, уничтожили его книги и рукописи, но он уже приступил к новой работе по случайным процессам на той основе, на которой вскоре появилась его интересная книга [84].

Мои парижские лекции по статистическому оцениванию основывались на моей книге [24], которая уже печаталась. Я предполагал остановиться на спорных проблемах фишеровской фидуциальной вероятности и неймановских доверительных интервалах; я был всецело на стороне Неймана. Не вполне приятным сюрпризом для меня оказалось присутствие Фишера в Париже и его посещение моей лекции. Позже мы встретились вдвоём для обсуждения этих вопросов, которое закончилось благополучнее, чем я предполагал, частично, быть может, потому, что я случайно знал место, где можно было хорошо поесть: год спустя после окончания войны таких мест в Париже было не так много.

В парижских лекциях по стационарным процессам я, кроме прочих проблем, остановился на спектральном представлении ковариационной функции по Хинчину (4.8.1) и на моём собственном представлении самогО изменяющегося процесса (5.1.1). Оказалось, что Лоэв независимо от меня также нашёл последнее. В приложении [90] к книге Леви [84] Лоэв обозрел результаты важной работы, связанной с этой и рядом родственных задач. На основе этого обзора появилась глава написанной им замечательной книги [91].

В сентябре 1946 г. я впервые приехал в США. В Принстоне меня принимал Сэм Уилкс. Я прочёл лекции по случайным процессам, и среди моих слушателей были К. Л. Чжун, Тед Харрис, Дж. А. Хант и Сэм Карлин. Было очень приятно работать с этими умными молодыми людьми. В том году Принстонскому университету исполнялось двести лет, и в научные конференции юбилейной программы входила одна, названная *Проблемы математики*. На ней среди множества выдающихся математиков присутствовали мои старые друзья Эйнар Хилл, Уилл Феллер, Ежи Нейман и Норберт Винер, которых мне было очень приятно увидеть снова. Я приобрёл и много новых знакомых, среди которых были такие выдающиеся вероятностники и статистики как Дж. Л. Дуб, Харолд Хотеллинг и Марк Кац.

На конференции работала секция по математической теории вероятностей, на которой все эти выдающиеся математики сделали интересные сообщения. Я был особенно рад встретиться с Дубом, чья работа по случайным процессам была мне известна и восхищала. Я [25] обозрел старые и новые задачи этой области.

По приглашению Гертруды Кокс и Харолда Хотеллинга я впервые посетил Чапел-Хилл, провёл там всего лишь несколько дней, но встретился с П. Л. Су и Хербертом Роббинсом. На протяжении весеннего семестра в Йеле и летнего в Беркли я продолжал свою работу и чтение лекций по случайным процессам. Нейман там подготавливал знаменитую серию Берклиевских симпозиумов по теории вероятностей и математической статистике. Мне доставило большое удовольствие знакомство с группой, занятой этой работой и в частности с Дж. Л. Ходжисом, Эрихом Леманом, Хенри Шеффе и Бетти Скотт. Все они теперь хорошо известны своими значительными научными достижениями.

6.3. Работа в стокгольмской группе. Во время моего отсутствия в Швеции в 1946 и 1947 гг. моё место в Стокгольмском университете занимал Густав Эльфвинг из Хельсинки. Через него и его земляка Кари Карунена наша группа знакомилась с вероятностными исследованиями, проводимыми в Финляндии. В своей диссертации, представленной в 1947 г. в Хельсинки, Карунен [59] систематически изучил применение теории гильбертова пространства к анализу случайных процессов, продолжив тем самым работу Колмогорова (§ 5). В диссертации содержались важные новые результаты по стохастическим интегралам и стационарным процессам, а после её завершения Карунен некоторое время работал в нашей группе.

Он [60] исследовал стационарный процесс с непрерывным временем и получил результаты, соответствующие колмогоровским для случая дискретного времени, в том числе разложение Волда и необходимое и достаточное условие абсолютной недетерминированности подобного процесса. Результаты такого же рода содержались и у О. Ханнера [53], другого сотрудника нашей группы.

В работе, написанной к Берклиевскому симпозиуму 1950 г., также опираясь на теорию гильбертова пространства, я [26] рассмотрел ещё несколько более общих классов случайных процессов и вывел общий вид спектрального представления. Для стационарных процессов он оказался простейшим из известных к тому времени, ср. Дуб [35].

Диссертация члена нашей группы Ульфа Гренандера [49] явилась подлинным новым словом в этой трудной и важной области. Известно, что он продолжал работу в этом направлении, и, например, выпустил в соавторстве прекрасную книгу [50]. В другой книге, также в соавторстве, он [51] распространил свои замечательные результаты на смежные области, и продолжил свои усилия в ряде других работ.

В 1958 г. именно Гренандер предложил издать *Сборник Харальда Крамера* [52] со статьями по теории вероятностей и математической статистике, написанными многими моими друзьями. Я был счастлив получить этот символ дружбы. Гренандер стал моим преемником на посту профессора Стокгольмского университета. Позже он, однако, покинул Стокгольм и сейчас работает в Университете Брауна, Провиденс, штат Род-Айленд, США.

С 1950 г. я вёл в университете административную работу, которая занимала значительную часть моего времени вплоть до того, как мне удалось в 1961 г. освободиться от неё. И всё же я смог написать монографию по теории риска [27], наибольшую часть которой составляет изучение задачи о разорении применительно к лундсбергскому процессу страхового риска (ср. § 4). В самом общем случае эта задача приводит к интегральному уравнению Винера – Хопфа. Его обсуждение может привести к ряду асимптотических результатов, ценных для приложений. Новые методы, которые позже предложил Феллер, позволяют получить часть этих результатов, не прибегая к уравнению Винера – Хопфа.

6.4. Москва, 1955. В мае 1955 г. Московский университет отмечал своё двухсотлетие, и я получил приглашение представлять на торжествах Стокгольмский университет. Это было крупное событие, мне же оно дало возможность лично познакомиться с советскими математиками, работы которых так сильно способствовали развитию теории вероятностей. К сожалению, Хинчин был болен и вскоре умер, но я встретился с Колмогоровым.

Он произвёл на меня впечатление значительной личности и очень крупного учёного, и я несколько раз с интересом беседовал с ним. Мне было также приятно познакомиться с другими членами их вероятностной группы. Среди них были Дынкин, приступавший к своей значительной работе по марковским процессам, Гнеденко, написавший вместе с Колмогоровым книгу по предельным задачам (см. выше), Линник, начинавший изучать большие уклонения, т. е. тему, столь близкую к моей собственной работе [19], Яглом и Розанов, получившие выдающиеся результаты по стационарным процессам, и многие другие. Научная деятельность этой группы была превосходна, и они подготавливали издание нового журнала, *Теория вероятностей и её применения*. Вскоре он завоевал ведущее положение в этой области в международном масштабе.

В связи с двухсотлетием состоялась и математическая конференция, на которой я прочёл лекцию о своей недавней работе, задаче о разорении. Позже в том же году Колмогоров провёл некоторое время в Стокгольме в качестве гостя нашего университета и прочёл цикл лекций для нашей группы. Они были посвящены предельным теоремам теории вероятностей.

6.5. Книги по теории вероятностей. До войны имелось очень мало книг по математической теории вероятностей, основанных на современной аксиоматике. Некоторые из них упоминались выше. После войны положение кардинально изменилось. Учебники и монографии по отдельным разделам теории образовали настоящий поток. Я приведу очень краткий обзор, строго ограниченный теми книгами, которые непосредственно повлияли на мои собственные исследования. За пределами этого обзора естественно окажутся многие значительные работы. В частности, обширная литература по марковским процессам и мартингалам будет отсутствовать, хоть я и полностью сознаю огромную важность этих проблем. Я неизменно указываю лишь год первого издания книги на языке оригинала.

Первый универсальный послевоенный учебник по теории вероятностей написал Феллер [45]. Для молодого [научного] поколения 1950-х годов эта книга являлась прекрасным введением в новую важную область научных исследований. В блестящем стиле, присущем её автору, она описывала основную теорию и значительную часть приложений. Второй том содержал ряд новых, и упрощённые доказательства некоторых известных результатов.

Учебник Лоэва [91] отличается проникновением в глубины математического анализа. В него вошли главы о мере и интегрировании. Он подробно рассмотрел классические предельные теоремы и их современные расширения на взаимозависимые переменные, а также важные классы случайных процессов.

Русский учебник Прохорова и Розанова [104] хорошо написан и полезен и учитывает последние результаты. Их книга характеризует высокий класс работ в нашей области, ведущихся в Советском Союзе.

Мы уже говорили о появившихся вскоре после окончания войны монографиях Гнеденко и Колмогорова по предельным задачам, Винера по предсказанию и фильтрации временнЫх рядов и книг Гренандера – Розенблатта и Гренандера – Сеге.

В 1948 и 1953 гг. появились три монографии об общей теории случайных процессов. Две из них, написанные Леви и Дубом, упоминались выше. Третью написали Блан-Лапьер и Форте. Все они являются классическими в этой области. Книга Леви, в частности, подробно рассматривает броуновское движение по прямой и в плоскости. Дуб дал полный обзор теории, в том числе сложных проблем, связанных с теорией меры, равно как и обширный материал по наиболее важным классам процессов. Последняя заслуживает особенного внимания ввиду подробного обсуждения приложений случайных процессов ко многим важным физическим задачам.

Яглом [124] написал хорошее введение в теорию стационарных процессов. Подробную монографию об этом важном классе процессов опубликовал Розанов [110]. Первая часть книги Линника и Ибрагимова [55] исчерпывающе описала работы Линника и его группы по большим уклонениям для сумм независимых случайных величин, вторая же часть была посвящена стационарным процессам. И книга Розанова [110], и это совместное сочинение содержали новые важные результаты, например, обобщение ЦПТ на стационарные процессы.

Другая монография Розанова [111] о бесконечномерных гауссовых распределениях содержала прекрасную главу о задаче эквивалентности или перпендикулярности нормальных распределений в функциональном пространстве. Четвёртый том великолепной работы Гельфанда и Виленкина [46] по обобщённым функциям включал интересную главу об обобщённых случайных процессах.

Наконец, следует упомянуть автобиографии двух ведущих вероятностников: Норберта Винера [122] и Поля Леви [85]. Каждая содержала массу научного материала, представлявшего значительный интерес.

6.6. Стационарные и связанные с ними случайные процессы. Мои собственные исследования в теории вероятностей развивались после 1960 г. по двум различным направлениям. Я сейчас кратко обозреваю свои собственные и иные работы по этой тематике, но строго не придерживаюсь хронологического порядка. Начну с работ по стационарным и родственным им случайным процессам 1950-х и 1960-х годов.

Это направление, начало которому положили работы Колмогорова и Засухина военного времени, развивали в конце 1940-х годов Карунен, Ханнер и другие и вскоре были получены существенные результаты. При помощи теории гильбертова пространства свойства одномерного стационарного процесса были обобщены на случаи векторных процессов с дискретным и непрерывным временем и однородных случайных полей. Иначе говоря, на процессы с несколькими независимыми параметрами, т. е. такими, как две координаты плоскости, четыре координаты пространства – времени, удовлетворяющие некоторому условию, аналогичному стационарности для случая одного параметра.

Русские исследователи, книги которых упоминались выше, а также Винер и Масани [94 – 96] изучали спектральные свойства стационарного векторного процесса. Для случая недетерминированного одномерного процесса со спектральной функцией F Колмогоров и Карунен показали, что эти функции чисто недетерминированной и детерминированной составляющих идентичны соответственно абсолютно непрерывной и *скачкообразной* частям F. При изучении соответствующих свойств векторного процесса возникают трудности, связанные с рангом неубывающей спектральной матрицы **F** (см. § 4.8). Если ранг этой матрицы (определённой соответствующим образом) максимален, то разложение, соответствующее одномерному случаю, можно непосредственно обобщить (Масани [94]), в противном же случае положение существенно осложняется.

Сотрудник нашу стокгольмской группы Матерн [97] представил диссертацию по *пространственным вариациям*, посвящённую однородным случайным полям в двумерной плоскости и их приложениям к задаче оценивания в лесоводстве.

Начало ещё одной важной линии исследований положила попытка обобщить ЦПТ на стационарные процессы. Я (§ 3.4) сослался на Бернштейна [7]. Он показал, что ЦПТ можно обобщить на суммы слабо зависящих друг от друга СВ и ввёл метод, позволяющий справляться с такими задачами. Розенблатт [109] дал хорошее определение слабой зависимости, которое привело к новой концепции сильного перемешивания для случайных процессов.

Рассмотрим процесс x_n с дискретным временем и пусть $\mathbf{M}_{a,b}$ – наименьшая полная σ-алгебра событий (ω-множества), относительно которой все x_n, $a \le n \le b$ измеримы. В качестве меры зависимости между $\mathbf{M}_{-\infty,m}$ и $\mathbf{M}_{m+n,\infty}$ Розенблатт ввёл верхнюю грань $P|(A \cap B - P(A)P(B)|$ для всех событий $A \in \mathbf{M}_{-\infty,m}$ и $B \in \mathbf{M}_{m+n,\infty}$. Если верхняя грань такой величины при всех m стремится к нулю при n, стремящемся к бесконечности, то зависимость между двумя любыми частями последовательности x_n, далеко отстоящими друг от друга, неизменно слаба. В этом случае говорят о сильно перемешанном процессе.

Данное условие является усиленным вариантом обычного условия перемешивания, который применяется в эргодической теории. Обобщение на случай непрерывного времени следует непосредственно. Если указанное условие выполнено и вводятся дополнительные условия, связанные со свойствами моментов вида

$$\mathrm{E}[\sum_m^{m+n} x_i]^2 \text{ и } \mathrm{E}\,|\sum_m^{m+n} x_i\,|^3$$

для больших значений n, то можно сказать, что сумма $\sum_{m}^{m+n} x_i$ асимптотически нормальна для любого фиксированного значения m при $n \to \infty$. Для последовательности, стационарной в узком смысле, оказалось целесообразным, следуя Ибрагимову, ввести условие *равномерно сильного перемешивания*, что упрощает доказательство ЦПТ и позволяет обобщить закон повторного логарифма. Эти проблемы исследуются в книге Линника и Ибрагимова [55] и статьях Волконского и Розанова [117] и Резника [106]. В лекциях, прочитанных в Университете Архуса (Дания) в 1967 г. и Копенгагенском университете в 1969 г., я обозрел эти работы, и в некоторых случаях мне удалось получить окончательные результаты.

Все они были получены по методу Бернштейна [7]. Для частного случая нормальных процессов можно вывести более точные результаты, как, например, показали в своей важной работе Колмогоров и Розанов [79].

В 1962 г., работая в Research Triangle Inst. of North Carolina, США, я занимался задачами, возникшими в навигации космических летательных аппаратов и связанными с экстремальными значениями. Моим помощником был молодой новозеландец Росс Лидбеттер, сотрудничество с которым оказалось очень продуктивным. Мы получили многообещающие результаты и решили, что работу стОит продолжить и написать совместную монографию о наших задачах.

Составляя план книги, мы вскоре пришли к выводу, что в неё целесообразно включить достаточно подробное изложение общей теории стационарных процессов с упором на свойства их реализаций. В частности, мы хотели рассмотреть такие аналитические свойства реализаций, как непрерывность и дифференцируемость и исследовать СВ, представляющие число пересечений реализацией некоторого фиксированного уровня или кривой на протяжении заданного интервала времени.

В основу своей работы мы собирались положить ряд более ранних исследований, среди которых могу назвать работы Беляева [3 – 6], Булинской [9], Каца и Слепяна [56], Райса [107; 108], Волконского и Розанова [117] и Айлвизейкера [125; 126]. Основополагающим для всей этой области была работа Райса. Для нормального стационарного процесса, удовлетворяющего сильному условию перемешивания, Волконский и Розанов доказали важную теорему о том, что пересечения реализаций с очень высоким уровнем аппроксимируются пуассоновским процессом.

Мне удалось доказать, что это свойство выполняется и при более слабых ограничениях и что из него следуют интересные заключения о величинах и распределениях экстремальных значений реализаций [34]. Многие наши результаты были улучшены другими, в частности Беляевым, который написал предисловие к русскому изданию этой книги; С. М. Берманом, который выступил с серией значительных работ по экстремальным значениям; и Линдгреном [88], который привёл ряд результатов о форме и распределении максимумов и минимумов реализаций.

6.7. Структурные задачи для общего класса случайных процессов. Начиная с 1958 г. я пытался обобщить некоторые результаты, полученные для стационарных векторных процессов Засухиным, Винером – Масани и другими (см. выше). Было очевидно, что возможности обобщения результатов о спектральных представлениях очень ограничены. Но казалось, что та часть этой проблематики, которую Винер и Масани назвали анализом процессов *по временнОй области*, вполне допускает широкие обобщения для нестационарных случаев.

В докладе [28] я рассмотрел эти задачи для случаев временнОй области и спектрального анализа. В последнем мне удалось получить лишь предварительные результаты для специального класса процессов, но в анализе временнОй области я продвинулся значительно дальше. Оказалось, что в процессах с дискретным и непрерывным временем положения существенно различались.

Рассмотрим произвольный векторный процесс с дискретным временем, например, $\mathbf{x}_n = (x_{n1}, \ldots, x_{nq})$, где n принимает все отрицательные и положительные целые значения. Пусть составляющие имеют нулевые средние значения и конечные моменты второго порядка. Определим детерминированный и абсолютно недетерминированный процессы обычным образом. В этом случае можно без всяких дополнительных условий распространить на них разложение Волда, получив

$$\mathbf{x}_n = \mathbf{u}_n + \mathbf{v}_n,$$

где $\mathbf{u}_n$ абсолютно недетерминированный, а $\mathbf{v}_n$ детерминированный процессы, причём $\mathbf{u}_n$ и $\mathbf{v}_n$ взаимно ортогональны. Более того, недетерминированные составляющие $\mathbf{u}_n$ можно представить в виде линейной формы

$$\mathbf{u}_n = \sum_{i=-\infty}^{n} \mathbf{c}_{ni}\mathbf{z}_i,$$

где $\mathbf{z}_i = (z_{i1}, \ldots, z_{ir_i})$ вектор-столбец порядка $r_i \le q$ и $\mathbf{c}_{ni}$ матрица порядка $q \times r_i$. Составляющие z_{ij} взаимно ортогональны для всех i и j.

Для случая векторного процесса $\mathbf{x}(t) = (x_1(t), \ldots, x_q(t))$ с непрерывным временем t существует аналогичное разложение

$$\mathbf{x}(t) = \mathbf{u}(t) + \mathbf{v}(t)$$

с абсолютно недетерминированной $\mathbf{u}(t)$ и детерминированной $\mathbf{v}(t)$ и притом взаимно ортогональными составляющими. В этом случае представление $\mathbf{u}(t)$ оказывается, однако, более сложным.

Если существуют среднеквадратические пределы $\mathbf{u}(t + 0)$ и $\mathbf{u}(t - 0)$ для всех t, то можно показать, что гильбертово пространство, натянутое на все составляющие $\mathbf{u}(t)$, сепарабельно. В этом случае, опираясь на геометрию гильбертова пространства, можно показать, что для всех t

$$\mathbf{u}(t) = \int_{-\infty}^{t} \mathbf{G}(t, v) d\mathbf{z}(v),$$

где $\mathbf{z}(v) = (z_1(v), \ldots, z_q(v))$ вектор-столбец порядка N, составляющими которого служат взаимно ортогональные случайные процессы с ортогональными приращениями, и $\mathbf{G}(t, v)$ неслучайная матрица порядка $q \times N$. Число N однозначно определяется заданным процессом $\mathbf{x}(t)$ и может принимать любое целое положительное значение либо быть бесконечно большим.

Таким образом, в случае дискретного времени порядок векторов $\mathbf{z}_i$ не больше заданного векторного процесса q, но процесс $\mathbf{z}(v)$, возникающий в непрерывном случае, может иметь любой и даже бесконечный порядок.

И при дискретном, и при непрерывном времени представление недетерминированной составляющей $\mathbf{u}_n$ или $\mathbf{u}(t)$ немедленно обеспечивает явное линейное предсказание методом наименьших квадратов. Свойства этого представления я исследовал ещё в нескольких статьях [29; 31; 32]. Важно было бы научиться определять многообразие N, соответствующее заданному процессу $\mathbf{x}(t)$, и мне удалось частично решить эту задачу. Её полное решение как будто ещё неизвестно.

Представление $\mathbf{u}(t)$ для нормального процесса предложил Хид [54] в то же время, когда я выступил с докладом [28]. В дальнейшем интересные результаты получили Каллианпур и Мандрекар [57; 58] и совсем недавно Розанов [112]. Ряд работ об этой и родственным задачам содержится в сборнике Эфримайдиса и Томаса [39].

6.9. Путешествия и работа: 1961 – 1970. В начале лета 1961 г. я подал в отставку со своего поста в университетской администрации и перешёл на положение независимого учёного. На Парижском конгрессе Международного статистического института я встретился с Гертрудой Кокс, которая пригласила меня приехать поработать в Research Triangle Inst. of North Carolina, США. В 1962, 1963 и 1965 гг. я провёл там по нескольку месяцев. Кроме того, в 1963 г. я был приглашённым профессором в Колумбийском университете, в 1965 г., в Йеле, и в 1966 г., в Беркли. Выше я уже говорил о работах по случайным процессам, которыми я занимался в те годы. Во всех научных центрах я работал вместе с друзьями, старыми и новыми.

Летом 1962 г. в Стокгольме проходил международный математический конгресс, и мы с женой устроили у себя в доме *вероятностный* завтрак, на котором присутствовали мои выдающиеся друзья, и среди них Дуб, Хант, Ито, Каппос. Колмогоров, Линник, Масани, Реньи, Розенблатт, Такач и Урбаник[19].

На Всесоюзную конференцию по теории вероятностей и математической статистике в Тбилиси в октябре 1963 г. я приехал прямо из США. Программа конференции была прекрасно составлена. Она включала лекции русских вероятностников, работы которых я упомянул выше, и доклады молодых учёных, высокий научный уровень и актуальность которых произвели очень сильное впечатление.

Колмогоров прочёл лекцию о применении вероятностных методов к анализу поэзии, Яглом, Ибрагимов и Розанов доложили о различных аспектах теории случайных процессов, а Линник – о статистических критериях. Мой доклад был посвящён случайным процессам как кривым в гильбертовом пространстве и содержал некоторые результаты о многообразиях, см. выше. После конференции мне предоставили возможность прочесть лекции в Москве и Ленинграде.

В 1967 г. я присутствовал на конференции по статистическим экстремумам в Фару (Португалия). В её программе были хорошо представлены и теоретические, и прикладные аспекты проблемы. В дальнейшем я был приглашённым профессором в Университете Архуса (Дания), а в 1969 г., в Копенгагенском университете.

В 1970 г. Джон Тьюки пригласил меня прочесть первую лекцию С. С. Уилкса [32] на торжественном открытии нового корпуса в Принстонском университете. Во время этой поездки мы с женой посетили Чапел-Хилл, Принстон, Провиденс и Сторрс и с удовольствием увидели работавших там друзей. И на этом моменте, на 70-м году жизни, я заканчиваю свои воспоминания о развитии математической теории вероятностей на протяжении половины столетия.

Библиография

Теория вероятностей и её применения = *ТВ и её прим.*
Успехи математич. наук = УМН
Giorn. Ist. Ital. Attuari = GIIA
Nordisk Stat. Tidskr. = NST
Proc. Berkeley Symp. Math. Stat. Prob. = *Berkeley*
Skand. Aktuarietidskr. = SAT

1. Arley N. (1943), *Stochastic Processes and Cosmic Radiation*. Copenhagen. Диссертация.

2. Bachelier L. (1900), Théorie de la speculation. *Ann. Ecole Norm. Sup*., t. 17, pp. 21 – 86. Paris, 1995

3. Беляев Ю. К., Beliaev Yu. K. (1959), Аналитические случайные процессы. *ТВ и её прим*., т. 4, № 4, с. 437 – 444.

4. --- (1960), Локальные свойства выборочных функций стационарных гауссовских процессов. Там же, т. 5, № 1, с. 128 – 131.

5. --- (1960), Continuiity and Hölder's conditions for sample functions of stationary Gaussian processes. *Fourth Berkeley*, 23 – 33.

6. --- (1966), О числе пересечений уровня гауссовским случайным процессом. *ТВ и её прим*., т. 11, № 1, с. 120 – 128.

7. Бернштейн С. Н., Bernstein S. N. (1927, франц.), Распространение предельной теоремы теории вероятностей на суммы зависимых величин. УМН, вып. 10, 1944, с. 65 – 114.

8. Brockmeyer E., Halstrom H. L., Jensen A. (1948), *Life and Works of A. K. Erlang*. *Trans. Danish Acad. Tech. Sci*., No. 2.

9. Булинская Е. В. (1961), О среднем числе пересечений некоторого уровня стационарным гауссовским процессом, *ТВ и её прим*., т. 6, № 4, с. 474 – 478.

10. Cantelli F. P. (1933), Considerazione sulla legge uniforme dei grandi numeri. GIIA, 4, No. 3.

11. Charlier C. V. L. (1905), Über das Fehlergesetz. *Ark. Mat. Astr. Fys.* 2, No. 8.

12. Cramér H., Крамер Г. (или **Х.)** (1919), Bidrag till utjämningsförsäkringens teori. *Insur. Inspect.* Stockholm.

13. --- (1923), Das Gesetz von Gauss und die Theorie des Risikos. SAT, pp. 209 – 237.

14. --- (1925), On some classes of series used in mathematical statistics. *Sixth Scand. Math. Congr. Copenhagen*, pp. 399 – 425.

15. --- (1926), Sannolikhetskalkylen i den vetenskapliga litteraturen. NST, t. 5, pp. 1 – 32.

16. --- (1928), On the composition of elementary errors. SAT, pp. 13 – 74, 141 – 180.

17. --- (1934), Su un teorema relativo alla legge uiforme dei grandi numeri. GIIA, 5, No. 1.

18. --- (1936), Über eine Eigenschaft der normalen Verteilungsfunktion. *Math. Z*., Bd. 41, pp. 405 – 414.

19. --- (1937, англ.), *Случайные величины и распределения вероятностей*. М., 1947.

20. --- (1938, франц.), Об одной новой предельной теореме теории вероятностей. УМН, вып. 10, 1944, с. 166 – 178.

21. --- (1939), On the representation of a function by certain Fourier integrals. *Trans. Amer. Math. Soc*., vol. 46, pp. 191 – 201.

22. --- (1940), On the theory of stationary random processes. *Ann. of Math.*, vol. 41, pp. 215 – 230.

23. --- (1942), On harmonic analysis in certain functional spaces. *Ark. Mat. Astr. Fys.* 28B, No. 12.

24. --- (1945, англ.), *Математические методы статистики*. М., 1948, 1975.

25. --- (1947), Problems in probability theory. *Ann. Math. Stat.*, vol. 18, pp. 165 – 193.

26. --- (1950), Contribution to the theory of stochastic processes. *Second Berkeley*, 329 – 339.

27. --- (1955), *Collective Risk Theory*. Stockholm.

28. --- (1960), On some classes of non-stationary stochastic processes. *Fourth Berkeley*, 2, 57 – 77.

29. --- (1961), On the structure of purely non-deterministic stochastic processes. *Ark. Mat.* 4, pp. 249 – 266.

30. --- (1963), On asymptotic expansions for sums of independent random variables with a limiting stable distribution. *Sankhya*, vol. A25, pp. 13 – 24.

31. --- (1964), Случайные процессы как кривые в гильбертовом пространстве. *ТВ и её прим.*, т. 9, № 2, с. 193 – 204.

32. --- (1971), *Structural and Statistical Problems for a Class of Stochastic Processes.* S. S. Wilks Lecture. Princeton.

33. --- (1972), On the history of a certain expansion used in mathematical statistics. *Biometrika*, vol. 59, pp. 205 – 207.

34. Крамер Г., Линдбеттер М. (1967, англ.), *Стационарные случайные процессы. Свойства выборочных функций и их приложения.* М., 1969.

35. Cramér H., Wold H. (1936), Some theorems on distribution functions. *J. London Math. Soc.*, vol. 11, pp. 290 – 294.

36. Дуб Дж. Л. (1953, англ.), *Вероятностные процессы.* М., 1956.

37. Edgeworth F. Y. (1905), The law of error. *Proc. Cambr. Phil. Soc.*, vol. 20, pp. 36 – 141.

38. Эйнштейн А. (1906, нем.), К теории броуновского движения. *Собр. научн. тр.*, т. 3. М., 1966, с. 118 – 127.

39. Ephremides A., Thomas J. B., редакторы (1973), *Random Processes*. Stroudsburg.

40. Esseen C. G. (1945), Fourier analysis of distríbution functions. *Acta Math.*, vol. 77, pp. 1 – 125.

41. Feller W. Феллер В. (1935, 1937), Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung. *Math. Z.*, Bd. 40, pp. 521 – 559; Bd. 42, pp. 301 – 312.

42. --- (1936, нем.), К теории стохастических процессов. Теоремы существования и единственности. УМН, вып. 5, 1938, с. 57 – 96.

43. --- (1940), On the integro-differential equations of purely discontinuous Markov processes. *Trans. Amer. Math. Soc.*, vol. 48, pp. 488 – 575.

44. --- (1943), The general form of the so-called law of the iterated logarithm. Там же, vol. 54, pp. 373 – 402.

45. --- (1950, 1966, англ.), *Введение в теорию вероятностей и её применения*. М., 1952, 1964. Второе издание М., 1967.

46. Гельфанд И. М., Виленкин Н. Я. (1961), *Некоторые применения гармонического анализа. Оснащённые гильбертовы пространства.* М.

47. Гнеденко Б. В. (1939), К теории предельных теорем для сумм независимых случайных величин. *Изв. АН СССР*, сер. математич. № 2, с. 181 – 232.

48. Гнеденко Б. В., Колмогоров А. Н. (1949), *Предельные распределения для сумм независимых случайных величин*. М. – Л.

49. Гренандер У. (1950, англ.), *Случайные процессы и статистические выводы*. М., 1961.
50. Grenander U., Rosenblatt M. (1956), *Statistical Analysis of Stationary Time Series.* Stockholm.
51. Гренандер У., Сеге Г. (1958, англ.), *Теплицевы формы и их приложения*. М., 1961.
52. Grenander U., редактор (1959), *Probability and Statistics. Harald Cramér Volume*. Stockholm – New York.
53. Hanner O. (1949), Deterministic and non-deterministic stationary random processes. *Ark. Mat.*, t. 1, pp. 161 – 177.
54. Hida T. (1960), Canonical representations of Gaussian processes and their applications. *Mem. Coll. Sci. Univ. Kyoto*, vol. A33, pp. 109 – 155.
55. Ибрагимов И. А., Линник Ю. В. (1965), *Независимые и стационарно связанные величины*. М.
56. Kac M., Slepian D. (1959), Large excursions of Gaussian processes. *Ann. Math. Stat.*, vol. 30, pp. 1215 – 1228.
57. Kallianpur G., Mandrekar V. Каллианпур Г., Мандрекар В. (1970), On the connections between multiplicity theory and O. Hanner's time domain analysis of weakly stationary stochastic processes. *Essays in Probability and Statistics*. Univ. of North Carolina, pp. 385 – 396.
58. --- (1965), Теория кратности и теория представления чисто недетерминистских случайных процессов. *ТВ и её прим.*, т. 10, № 4, с. 614 – 644.
59. Karhunen K. (1947), Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fenn*, t. A37, pp. 1 – 79.
60. --- (1949), Über die Struktur stationärer zufälliger Funktionen. *Ark. Mat.*, t. 1, pp. 141 – 160.
61. Keynes J. M. (1921), *Treatise on Probability*. *Coll. Writings*, vol. 8. London, 1973.
62. Хинчин А. Я., Khinchine A. (1923), Über dyadische Brüche. *Math. Z.*, Bd. 18, pp. 109 – 116.
63. --- (1933), Zur mathematischen Begründung der statistischen Mechanik. *Z. angew. Math. Mech*., Bd. 13, pp. 101 – 103.
64. --- (1933, нем.), *Асимптотические законы теории вероятностей*. М. – Л., 1936.
65. --- (1934, нем.), Теория корреляции стационарных стохастических процессов. УМН, вып. 5, 1938, с. 42 – 51.
66. --- (1936), Sul dominio di attrazione della legge di Gauss. GIIA, t, 7, pp. 3 – 18.
67, 68. --- (1937), Новый вывод одной формулы П. Леви. Об арифметике законов распределения. *Бюлл. МГУ*, секция А, т. 1, № 1.
69. --- (1937), Zur Theorie der unbeschränkt teilbaren Verteilungsgesetze. *Математич. сб*., т. 44, с. 79 – 120.
70. Хинчин А. Я., Колмогоров А. Н. (1925), Über Konvergenz von Reihen, deren Glieder durch den Zufall bestimmt werden. Там же, т. 32, с. 668 – 677. Также О сходимости рядов, члены которых определяются случаем. В книге Колмогорова *Теория вероятностей и математическая статистика*. М., 1986, с. 7 – 17. Также, без пояснений:
71. Колмогоров А. Н., Kolmogorov A. N. (1928), Über die Summen durch den Zufall bestimmter unabhängiger Größen. *Math. Ann.*, Bd. 99, pp. 309 – 319.
72. --- (1929, нем.), О законе повторного логарифма. В книге Колмогорова *Теория вероятностей и математическая статистика*. М., 1986, с. 34 – 43.
73. --- (1931, нем.), Об аналитических методах в теории вероятностей. Там же, с. 60 – 105.

74.--- (1932, итал.), Об общей форме стохастического однородного процесса. Проблема Бруно де Финетти. Там же, с. 117 – 123.

75. --- (1933, нем.), *Основные понятия теории вероятностей*. М., 1936, 1974.

76. --- (1938), Упрощённое доказательство эргодической теоремы Биркгофа – Хинчина. УМН, вып. 5, с. 51 – 56.

77. --- (1941), Стационарные последовательности в гильбертовом пространстве. *Бюлл. МГУ*, секция А, т. 2, № 6, с. 1 – 40.

78. --- (1941), Интерполирование и экстраполирование стационарных случайных последовательностей. *Изв. АН СССР*, серия математич., вып. 5, с. 3 – 14.

79. Колмогоров А. Н., Розанов Ю. А. (1960), Об условиях сильного перемешивания гауссовского стационарного процесса. *ТВ и её прим.*, т. 5, № 2, с. 222 – 227.

80. Lévy P., Леви П. (1925), *Calcul des probabilités*. Paris.

81. --- (1934), Sur les intégrales dont les éléments sont des variables aléatoires indépendantes. *Ann. Ecole Norm. Sup. Pisa*, t. 3, pp. 337 – 366.

82. --- (1935), Propriétés asymptotiques des sommes des variables aléatoires indépendantes ou enchainées. *J. math. pures appl.*, t. 14, pp. 347 – 402.

83. --- (1937), *Théorie de l'addition des variables aléatoires*. Paris.

84. --- (1948, франц.), *Статистические процессы и броуновское движение*. М., 1972.

85. --- (1970), *Quelques aspects de la pensée d'un mathèmaticien*. Paris.

86. Ляпунов А. М. (1901, франц.), Новая форма теоремы о пределе вероятности. *Собр. соч*., т. 1. М., 1954, с. 157 – 176.

87. Lindeberg J. W. (1922), Eine neue Herleitung des Exponentialgesetz in der Wahrscheinlichkeitsrechnung. *Math. Z.*, Bd. 15, pp. 211 – 225.

88. Lindgren G. (1972), *On Wave-Forms in Normal Random Processes*. Lund. Диссертация.

89. Линник Ю. В. См. № 55.

90. Loève M., Лоэв М. (1948), Fonctions aléatoires du second ordre. Note à Lévy. *Processus stochastiques et mouvement Brownien*. 84 (?), pp. 299 – 352. Paris.

91. --- (1955, англ.), *Теория вероятностей*. М., 1962.

92. Lundberg F. (1903), *Approximerad framställning av sannolikhetsfunktionen.* Uppsala. Диссертация.

93. Lundberg O. (1940), *On Random Processes and Their Application to Sickness and Accident Statistics*. Stockholm. Диссертация.

94. Masani P. (1959), Cramér's theorem on monotone matrix-valued functions and the Wold decomposition. *Probability and Statistics. Harald Cramér Volume*. Stockholm – New York, pp. 175 – 189.

95. --- (1966), Wiener's contributions to generalized harmonic analysis, prediction theory and filter theory. Norbert Wiener 1894 – 1964. *Bull. Amer. Math. Soc.*, vol. 72, pp. 73 – 125.

96. Masani P., Wiener N. (1957 – 1958), Prediction theory of multivariate stochastic processes. *Acta Math.*, vol. 98, pp. 111 – 150; vol. 99, pp. 93 – 137.

97. Matérn B. (1960), Spatial variation. *Swed. Forestry Res. Inst*. t. 49, pp. 1 – 144.

98. von Mises R., Мизес Р. (1919), Grundlagen der Wahrscheinlichkeitsrechnung. *Math. Z.*, Bd. 5, pp. 52 – 99. Перепечатка: *Sel. Papers*, vol. 2. Providence, RI, pp. 57 – 105.

99. --- (1928, нем.), *Вероятность и статистика.* М., 1930; М., 2009, пятое издание.

100. De Moivre A. (1733), *Miscellanea analytica*. Второе дополнение. London.

101. Neyman J. (1937), Outline of a theory of statistical estimation based on the classical theory of probability. *Phil. Trans. Roy. Soc. Lond.*, vol. 236, pp. 333 – 380.

102. --- (1942), Basic ideas and some recent results of the theory of testing statistical hypotheses. *J. Roy. Stat. Soc.*, vol. 105, pp. 292 – 327.

103. Pólya G. (1920), Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem. *Math. Z.*, Bd. 8, pp. 171 – 181.

104. Прохоров Ю. В., Розанов Ю. А. (1967, 1973), *Теория вероятностей. Основные понятия, предельные теоремы, случайные процессы.* М.

105. Райков Д. А. (1938), О разложении законов Гаусса и Пуассона. *Изв. АН СССР*, серия математич., т. 2, с. 91 – 124.

106. Резник М. Х. (1968), Закон повторного логарифма для некоторых классов стационарных процессов. *ТВ и её прим.*, т. 13, № 4, с. 642 – 656.

107. Rice S. O. (1945), Mathematical analysis of random noise. *Bell System Tech. J.*, vol. 24, pp. 46 – 156.

108. --- (1958), Distribution of the duration of fades in radio transmission. Там же, vol. 37, pp. 581 – 635.

109. Rosenblatt M. (1956), A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci.*, vol. 42, pp. 43 – 47.

110. Розанов Ю. А. (1963), *Стационарные случайные процессы.* М.

111. --- (1968), *Гауссовские бесконечномерные распределения.* М.

112. --- (1974), *Теория обновляющих процессов.* М.

113. Резерфорд Э., Гейгер Г. (1908, англ.), Электрический метод счёта α-частиц, испускаемых радиоактивными веществами. В книге Резерфорд Э., *Избр. научн. тр*. *Строение атома и искусственное превращение элементов*. М., 1972, с. 123 – 142.

114. Segerdahl C. O. (1939), *On Homogeneous Random Processes and Collective Risk Theory*. Stockholm. Диссертация.

115. Чебышёв П. Л. (1867), О средних величинах. *Полн. собр. соч*., т. 2. М., 1947, с. 431 – 437.

116. --- (1887), О двух теоремах относительно вероятностей. Там же, т. 3. М., 1948, с. 229 – 239.

117. Волконский В. А., Розанов Ю. А. (1959 – 1961), Некоторые предельные теоремы для случайных функций. *ТВ и её прим*., т. 4, № 2, с. 186 – 207; т. 6, № 2, с. 200 – 215.

118. Wiener N. Винер Н. (1923), Differential space. *J. Math. Phys.*, vol. 2, pp. 131 – 174.

119. --- (1930), Generalized harmonic analysis. *Acta Math.*, vol. 55, pp. 117 – 258.

120. --- (1934, англ.), *Преобразование Фурье в комплексной области*. М., 1964. Соавтор Пэли Раймонд.

121.--- (1949), *Extrapolation, Interpolation and Smoothing of Stationary Time Series. With Engineering Applications*. New York.

122. --- (1956, англ.), *Я – математик*. М., 1964. Сокращённый перевод.

123. Wold H. (1938), *Study in the Analysis of Stationary Time Series*. Stockholm. Диссертация.

124. Яглом А. М. (1952), Введение в теорию стационарных случайных функций. УМН, т. 7, № 5 (51), с. 3 – 168.

125. Ylvisaker N. D. (1965), The expected number of zeros of a stationary Gaussian process. *Ann. Math. Stat.*, vol. 36, pp. 1043 – 1046.

126. --- (1966), On a theorem of Cramér and Leadbetter. vol. 37, pp. 682 – 685. Ср. № 34.

127. Засухин В. Н. (1941), К теории многомерных стационарных случайных процессов. *Докл. АН СССР*, т. 33, с. 435 – 437.

Примечания

1. Где же Рис мог обучать Крамера?

2. См.[ii, прим. 26]. Эпоха теории вероятностей как прикладной дисциплины закончилась в 1920-е годы.

3. Следует вспомнить Якоба Бернулли [i, прим. 2.4].

4. В указанном сочинении [100] Муавр приблизился к доказательству своей теоремы, окончательно же доказал её в 1733 г. (Шейнин 2019, § 5.4).

5. Ляпунов опубликовал свои вероятностные сочинения в Петербурге на французском языке. Кроме того, он опубликовал две заметки (естественно, также по-французски) в *C. r. Acad. Sci. Paris*, т. 132, 1901.

6. Фраза отвратительна: характеристические функции аналогичны методам.

7. Напрашивается сравнение с композитором и химиком А. П. Бородиным.

8. При определении порядка малости члена, характеризующего ошибку, я заменил в первоначальном варианте символ *O* на *o* в соответствии с результатом Эссеена [40]. Автор.

9. Леви [80, с. vii] заявил, что без теории ошибок его сочинение об устойчивых законах было бы бесцельным. Нет, это его сочинение оказалось бесцельным для теории ошибок, см. Sheynin (1995).

10. Длинное и плохо понятное рассуждение.

11. Правильно: неравенство Бьенеме – Чебышёва.

12. Не следует забывать книгу Леви [80]. Основными объектами теории вероятностей оказались у него плотности вероятностей и характеристические функции. Эпоха Лапласа закончилась. Ср. прим. 2.

13. В первую мировую войну Дёблин был солдатом французской армии. Будучи евреем, он застрелился, чтобы избежать неминуемого плена. См. о нём Bru (1993).

14. Участнику конференции Максу Борну как раз исполнилось 55 лет, и присутствовавшие поднесли ему адрес, который хранится в отделе рукописей Берлинского университета. На конференции не было ни одного участника из социалистического рая. Надо же было по возможности скрывать Большой террор!

15. Эти строки можно легко найти в начале предисловия к русскому изданию книги Крамера.

16. Книга имела жёлтую суперобложку. Переводчик

В конце XIX и начале XX века жёлтой опасностью (yellow peril) называли опасность, грозящую (или будто бы грозящую) Западу от народов Востока.

17. Йельский университет находится в городе Нью-Хейвен, и потому *Йель* можно считать просто обычным словоупотреблением.

18. См. прим. 13.

19. Фамилии перечислены в порядке английского (латинского) алфавита.

Дополнительная библиография

Хинчин А. Я. (1961), Частотная теория Мизеса и современные идеи теории вероятностей. *Вопр. философии*, № 1, с. 91 – 102; № 2, с. 77 – 89.

Шейнин О. Б., Sheynin O.(1995), Density curves in the theory of errors. *Arch. Hist. Ex. Sci*., vol. 49, pp. 163 – 196.

--- (2019), *Теория вероятностей. Исторический очерк*. Берлин. **S, G,** 11.

Bru B. (1993), Doeblin’s life and work from his correspondence. *Contemporary Math.*, vol. 149, pp. 1 – 64.