О. Б. Шейнин

Теория вероятностей
Элементарное руководство с привлечением сведений из истории этой науки

Берлин, 2008

ISBN 3-938417-78-1

Оглавление

Рисунки

Рисунок 1.1.2 относится к § 1.1 и т. д.

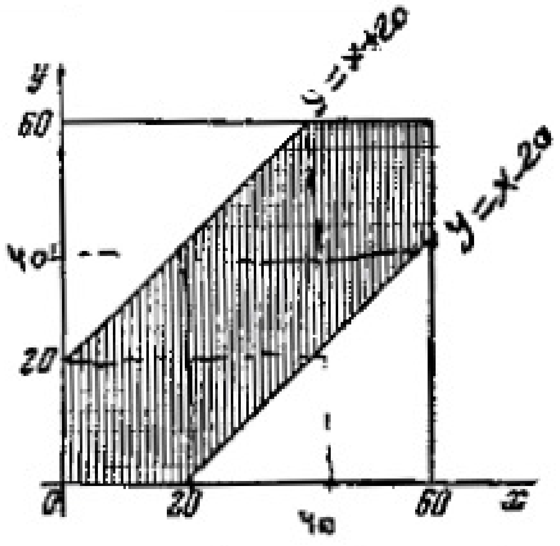

Рис. 1.1.2 (Гнеденко 1954, с. 36). Задача о встрече.

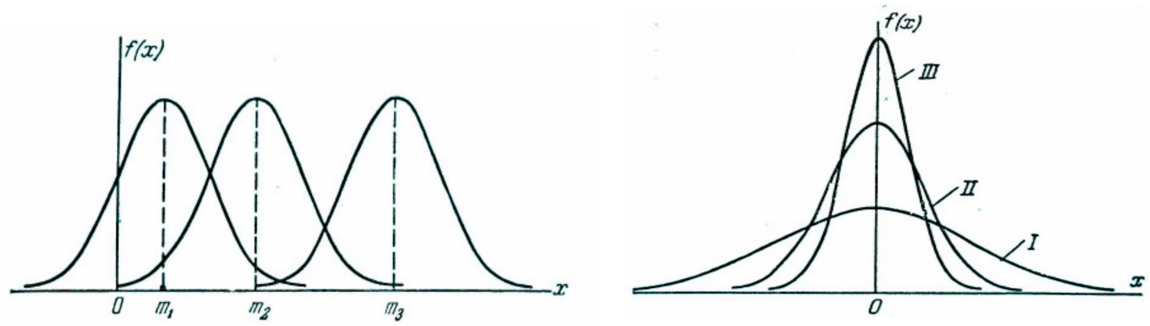

Рис. 2.1 (Вентцель 1969, с. 119). Плотности нормального распределения при различных значениях ее параметров (см. п. 2.2.4), а именно дисперсии (Рис. 2a) и параметра сдвига (Рис. 2b).

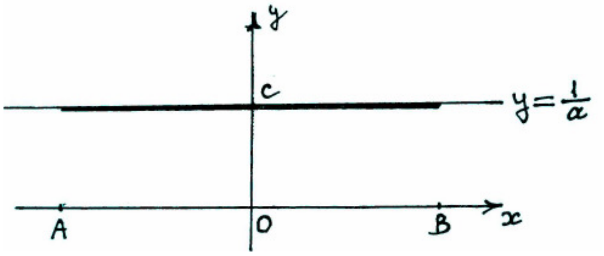

Рис. 2.2.1. Плотность непрерывного равномерного распределения. Точки A(– a; 0) и B(a; 0) не обязательно симметричны относительно начала координат. Ломаная ACB (на рисунке не проведена) есть плотность непрерывного треугольного распределения (см. п. 2.2.2). Для наглядности масштабы вдоль осей координат различны.

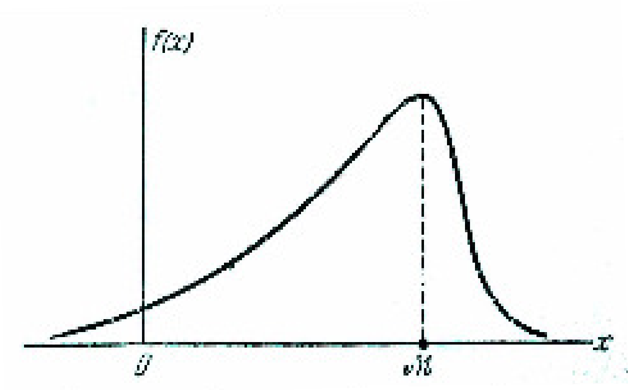

Рис. 2.5.1-2 (Вентцель 1969, с. 90). Мода или точка максимума плотности распределения.

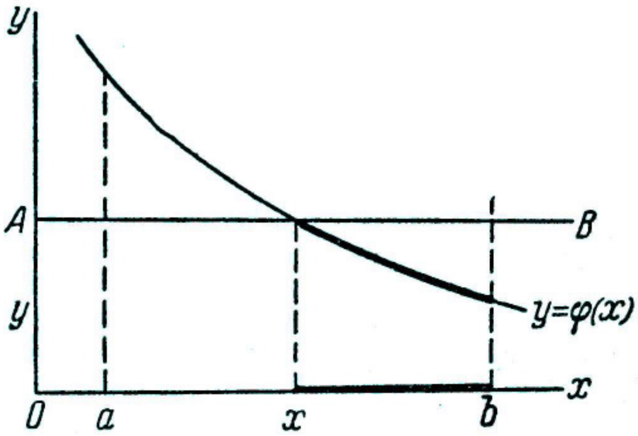

Рис. 2.8 (Вентцель 1969, с. 265). Функцию, обратную заданной, можно увидеть, если повернуть чертеж так, как сказано в нашем пояснении. Для согласования рисунка с текстом следует заменить $\varphi(x)$ на $f(x)$, а x и y – на x_0 и y_0.

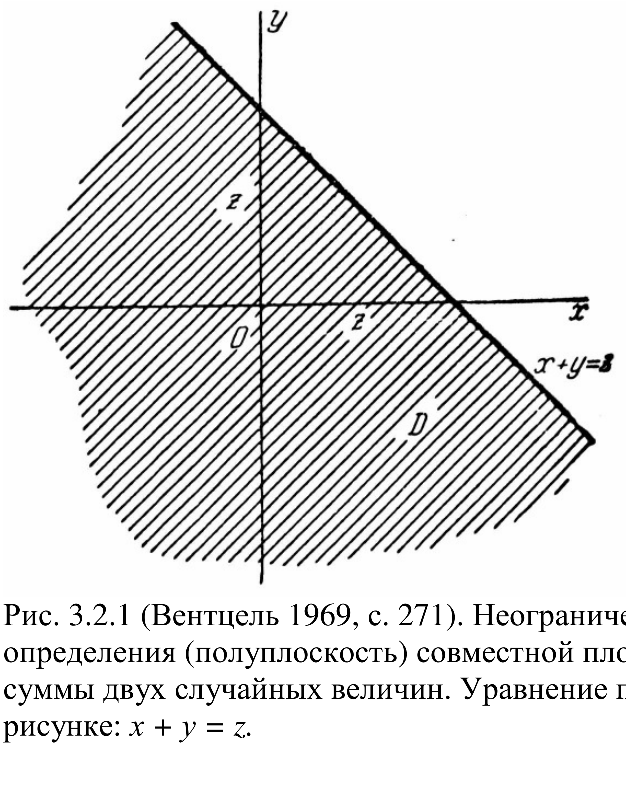

Рис. 3.2.1 (Вентцель 1969, с. 271). Неограниченная область определения (полуплоскость) совместной плотности распределения суммы двух случайных величин. Уравнение прямой, показанной на рисунке: $x + y = z$.

0. Введение

0.1. Некоторые пояснения

Эта брошюра написана по возможности элементарно, и в более сложных случаях окончательные формулы приведены без вывода. И всё-таки без интегралов обойтись было нельзя, а в одном случае пришлось дифференцировать интеграл по параметру.

Изложение сопровождается многочисленными примерами из истории теории вероятностей (с непременным указанием соответствующих источников), и мы надеемся, что это оживило содержание. Мало того, считаем уместным привести здесь покаянное признание Пирсона (1978, с. 1):

Я действительно чувствую, как неосмотрительно было работать так долго в области статистики и пренебрегать ее историей.

Английское *wrongful* могло означать и *неверно, нехорошо* и т. д., а статистика в английском языке включает и математическую статистику и даже иногда теорию вероятностей. Пирсон, кстати, совершенно неверно оценил заслуги Якоба Бернулли (п. 4.1.1), но вот его историческое исследование (1978) заслуживает серьезного внимания.

Мы уделили много внимания понятию *вероятность*, что вполне соответствует утверждению Ланжевена (Langevin 1913/1914, с. 3):

Во всех вопросах [кинетической теории газов] *основная трудность* […] *– дать точное и ясное определение вероятности.*

Учебников и руководств по теории вероятностей, написанных на более высоком уровне, отыскать нетрудно; мы сослались на один такой источник (Гнеденко 1954), но гораздо труднее найти подходящие источники по ее истории.

0.2.Чем занимается теория вероятностей?

Уравнения, встречающиеся в математике, обычно имеют решение или решения, которые, правда, не всегда удается получить в точном виде. *Обычно* здесь означает: не в теории вероятностей; можно, например, вспомнить о квадратных или тригонометрических уравнениях. Но вот вероятностная, или, как говорят, *стохастическая* задача: монета мысленно подбрасывается 10 раз. Сколько раз она выпадет гербом вверх? Никаких уравнений составить не удастся, и произойти это может либо 1 раз, либо 2 раза, …, либо все 10 раз, либо ни одного раза. Больше ничего определенного пока сказать нельзя, но одно существенное обстоятельство следует подчеркнуть сразу: *более четкий ответ к этой задаче и вообще ко всем вероятностным задачам в принципе возможен только в вероятностном смысле*. В данном случае можно будет уточнить только так: монета выпадет гербом вверх 1 раз с вероятностью такой-то, 2 раза – с какой-то иной вероятностью и т. д.

Пример несерьезный? Ну, во-первых, теория вероятностей провела свою молодость у игорного стола, потому что *азартные* игры (результат которых зависит только от случая, но не от искусства участников) доставляли интересные задачи, методически важные и сегодня. Во-вторых, есть у нее и действительно существенные приложения. Как только изучаются массовые случайные события, так приходится иметь дело с вероятностями, например, в статистике населения (рождения, смертность, продолжительность жизни), при обработке точных наблюдений (искаженных неизбежными случайными ошибками), выборочном контроле качества массовой продукции (с его вероятностной оценкой), а также в различных областях науки, особенно естествознания (например, в кинетической теории газов, поскольку следить за движением каждой молекулы невозможно, да и бессмысленно).

Итак, *теория вероятностей изучает массовые случайные события, а точнее – их закономерности*, которые, как показывают статистические исследования, действительно существуют. *Случайное в единичном – закономерное в массе, вот диалектика*. Аристотель не ошибался, но бил мимо цели, когда заявил, что *ни одна традиционная наука не интересуется случайным* (например, *Метафиз*. 1064b15).

Вот разумное, но неопределенное высказывание Лапласа (1814/1999, с. 863, правый столбец): "Теория вероятностей есть в сущности не что иное, как здравый смысл, сведенный к исчислению". Впрочем, в те времена так можно было отзываться о математике в целом. Сам Лаплас применял теорию вероятностей к исследованию законов природы (особо, в астрономии) и статистики населения, и являлась она прикладной математической дисциплиной (а самого себя Лаплас фактически считал прикладным математиком).

Но времена изменились, и новую и скучную, казалось бы, цель поставил перед теорией вероятностей Буль (1851/1952, с. 251): "По данным вероятностям любого [логического] предложения определить вероятность другого". То же по существу заявил в своих лекциях Чебышев (1879 – 1880/1936, с. 148) и даже много раньше (1845/1951, с. 29), но только относительно событий. Вот его более позднее определение:

Теория вероятностей имеет целью определить шансы для совершения известного [некоторого] *события, причем под словом <u>событие</u> разумеется вообще всё то, чего вероятность определяется.* […] *Вероятность служит в математике для означения некоторой величины, подлежащей измерению.*

Та же цель сохранилась и поныне (Прохоров и Севастьянов 1999, с. 77). А после своей аксиоматизации стала теория вероятностей относиться к чистой математике.

Глава 1. Основные понятия, теоремы и формулы

1.1. Вероятность

Геометрия Евклида начиналась с определений точки, прямой и плоскости; точка, к примеру, не имела ни одного измерения (ни длины, ни ширины, ни толщины). В наше время эти определения можно считать лишь эвристическими, основанными только на здравом смысле. Скажем больше: исходные понятия просто не на чем основывать, их приходится вводить без всякого определения.

Действительно, что такое, к примеру, длина? Вот подходящее, но не совсем четкое утверждение Маркова (1908, с. 2; 1924, с. 2):

Различные понятия определяются не столько словами, каждое из которых может в свою очередь потребовать определения, как нашим отношением к ним, которое выясняется постепенно.

В школьной геометрии подобные соображения вряд ли приводят к каким-либо методическим осложнениям, но вот в теории вероятностей необходимо сразу же всерьез озаботиться если не определениями в строгом смысле слова, то пояснениями.

1.1.1. Теоретическая вероятность

Пусть результат испытания зависит от n несовместимых и равновозможных случаев, из которых m и только m *благоприятных* приводят к появлению некоторого события А. Тогда его вероятность принимается равной

$$p = \frac{m}{n} \tag{1.1}$$

и она также может изменяться от 0 до 1, т. е. характеризовать события от невозможного до достоверного. Остальные ($n - m$) случаев можно назвать *неблагоприятными* по отношению к событию А. Это – так называемое классическое определение вероятности, восходящее к Муавру (De Moivre 1712/1984, с. 237). Сразу заметим, что оно логически несовершенно (вероятность зависит от равновозможных случаев, равновозможность же фактически означает равную вероятность) и ограничено по своим приложениям: редко когда существуют равновозможные случаи.

Первое обстоятельство обусловило справедливые и неоднократные нарекания, и вот одно свидетельство (Камке 1933, с. 14): в 1910 г. в гёттингенском университете было в ходу изречение: вероятность – число, лежащее между 0 и 1, *про которое ничего больше не известно*. Гёттинген, надо сказать, считался математической столицей мира, но вот в 1934 г. работавший там Гильберт заявил, что после изгнания из него еврейских профессоров университет перестал существовать.

Подобных свидетельств немало. Вот Марков (1911, с. 162):

Я не буду защищать […] *основных теорем, связанных с основными понятиями исчисления вероятностей, о равновозможности, о независимости событий и т. д., так как знаю, что можно спорить без конца даже об основных положениях такой признаваемой всеми точной науки, как геометрии.*

Действительно, до своей аксиоматизации *защищать* теорию вероятностей было трудно. Недаром Хинчин (1961, с. 94) подметил, что

Каждый автор, начиная свой трактат, неизменно говорил о равновозможных и благоприятных случаях, стараясь, впрочем, как можно скорее оставить эту неприятную тему.

Впервые определение (1.1) предложил Муавр в 1712 г., см. выше, но сформулировал его на языке шансов и даже в 1756 г. не отказался от них. Впрочем, Якоб Бернулли (посм. публ. 1713/1986, с. 24) еще раньше ввел вероятность в духе определения (1.1), хоть и не вполне формально и не использовал ее в своем дальнейшем изложении.

В смысле того же определения можно истолковать правила Талмуда о запрещенной пище. Так, Маймонид (Рабинович 1973, с. 41) признавал семь уровней запрета в зависимости от соотношений запрещенного и разрешенного (например, зерна двух видов).

И всё-таки вместо классического определения оказалось возможным отправляться лишь от статистической вероятности (см. ниже), что, однако, было связано с иными существенными затруднениями. Существует, правда, и аксиоматическое определение вероятности, – и вообще, в соответствии с пожеланием Гильберта (1901, Проблема № 6), современная теория вероятностей является аксиоматической, – но для практических целей она явно не подходит.

Пример на теоретическую вероятность. В заметке Галилея, опубликованной лишь в 1718 г., он подсчитывал количества случаев, приводящих к различным исходам при броске трех игральных костей. Правильная игральная кость – это кубик, каждая грань которого выпадает с одной и той же вероятностью, которая, конечно же, равна 1/6.

Возможные исходы – 3, 4, …, 17, 18 очков. Он прежде всего заметил, что исходы, равно удаленные от средних, равновозможны. Так, одинаково возможны 3 и 18 очков, 4 и 17, …, 10 и 11. Простое доказательство этого факта (который подтверждается прямыми подсчетами), нам неизвестно, но во всяком случае его можно найти (Эйлер 1748/1961, глава О разбиении чисел на слагаемые). Далее, Галилей засвидетельствовал, что игрокам было известно, что 9 или 12 очков менее благоприятны, чем 10 или 11.

Выпишем сочетания очков, приводящие к этим результатам.

9 очков: 1, 2, 6; 1, 3, 5; 1, 4, 4; 2, 2, 5; 2, 3, 4; 2, 3, 4; 3, 3, 3 очка
10 очков: 1, 3, 6; 1, 4, 5; 2, 2, 6; 2, 3, 5; 2, 4, 4; 3, 3, 4 очков

Казалось бы, число возможностей одно и то же, но Галилей указывает: три одинаковых числа могут появиться только в одном варианте; два одинаковых и третье, отличное от них, – в трех случаях, а три различных исхода – в шести вариантах. Так, числа 1, 3, 6 *размещаются* 3! = 6 способами. Всего же оказывается 25 возможностей для 9 очков и 27 – для 10. Всех исходов при броске трех костей – 216 (можно сказать и так: 108 исходов для трех, четырех, …, 10 очков и столько же для 11, 12, …, 18).

Вывод такой: игроки сумели отличить вероятности 27/216 и 25/216, т. е. подметить разность вероятностей, равную 1/108. Но очень возможно, что они поступили иначе: обращали внимание только на интересующие их броски, т. е. только на 27 и 25 и, стало быть, на вероятности 27/52 и 25/52, разность между которыми была равна 1/26.

Теперь необходимо добавление. Если наступили два события, А и В, то говорят, что произошло их произведение АВ; если произошло хотя бы одно из них, то имела место их сумма, А + В, а если появилось А, но не наступило В, то говорят, что наступила их разность А – В. И теперь следует сказать, что определение (1.1) применяется к полной группе несовместимых друг с другом событий, удовлетворяющих следующим условиям: она должна включать в себя невозможное и достоверное события, а если содержит события А и В, то должна содержать и АВ, и А + В, и А – В.

Пример. Предстоят два шахматных турнира, и вероятности некоему шахматисту занять в них первые места считаются известными. Какова вероятность, что он займет первое место в обоих случаях? Если турниры будет проходить одновременно, т. е.

если произведение соответствующих событий, А и В, не имеет смысла, то и говорить не о чем. Можно, конечно, сказать, что появилось невозможное событие, но ведь его нельзя получить из заданных вероятностей. Произведение АВ не входит в рассматриваемую группу событий, которая поэтому не отвечает поставленным выше требованиям.

1.1.1-1. Теорема сложения вероятностей. Вероятность обычно обозначается буквой *P* или *p* от французского *probabilité*, да и в соответствии с английским словом *probability*. Для несовместимых событий А и В

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

Пусть в урне находятся *n* шаров, из них *a* красных, *b* синих и $[n - (a + b)]$ белых. Какова вероятность извлечь цветной шар (Румшиский 1966)? Ясно, что она равна $a/n + b/n$.

Но вот, казалось бы, аналогичный пример. Найти вероятность выпадения одной шестерки в двух бросках игральной кости.

Решение. Событие А: выпадение шестерки в первом броске.

Событие В: ее выпадение во втором броске.

$$P(\text{A}) = 1/6,\ P(\text{B}) = 1/6,\ \text{P}(\text{A} + \text{B}) = 2/6 = 1/3.$$

Стоп, что-то не то: в шести бросках вероятность окажется равной 1, т. е. шестерка появится достоверно, но ведь не должно быть этого! А если, упаси боже, 7 бросков?

Дело в том, что события А и В здесь не являются несовместимыми, шестерка может появиться при каждом броске. Верное решение см. ниже.

Теоремы сложения и умножения (см. ниже) интуитивно понимаемых вероятностей фактически и также интуитивно применялись издавна. Аристотель (*О небе* 292а30) заявил, что 10 000 бросков *коан* (что бы это ни означало) подряд при игре в кости невозможны. Он использовал этот довод как обоснование того, что звезды не могут случайно вращаться заодно около полюса мира.

1.1.1-2. Обобщение теоремы сложения: формула включения и выключения. Для двух событий общая формула теоремы сложения имеет вид

$$P(\text{A} + \text{B}) = P(\text{A}) + P(\text{B}) - P(\text{AB}).$$

В нашем примере $P(\text{AB}) = 1/36$ и

$$P(\text{A} + \text{B}) = 1/6 + 1/6 - 1/36 = 11/36.$$

Ответ можно проверить, выписав все исходы двух бросков. Из 36 возможных исходов благоприятных окажется 11, а не 12.

Для трех событий соответствующая формула имеет вид

$$P(\text{A} + \text{B} + \text{C}) = P(\text{A}) + P(\text{B}) + P(\text{C}) - P(\text{AB}) - P(\text{AC}) - P(\text{BC}) + P(\text{ABC}),$$

а в общем случае

$$P(\mathrm{A}_1\mathrm{A}_2 \ldots \mathrm{A}_n) = P(\sum_i A_i) - P(\sum_{i<j} A_iA_j) + P(\sum_{i<j<k} A_iA_jA_k) - \ldots$$

Эту формулу включения и исключения применял уже Монмор в 1708 г. Вообще же она является частным случаем положения о взаимном расположении произвольных множеств.
Условия $i < j$, $i < j < k$, … обеспечивают перебор всех индексов без повторений. Так, для четырех событий условие $i < j$ означает, что учитываются события с индексами 1, 2; 1, 3; 1, 4; 2, 3; 2, 4 и 3, 4, а всего 6 вариантов, что равно числу сочетаний из четырех элементов по 2.

1.1.1-3. Теорема умножения. Предварительно введем обозначение P(B/A), означающее вероятность события B при условии, что произошло событие A. Теперь сама теорема:

$$P(\mathrm{AB}) = P(\mathrm{A})P(\mathrm{B/A}). \qquad (1.2)$$

Если поменять местами A и B, то

$$\mathrm{P(AB) = P(B)P(A/B)}. \qquad (1.3)$$

Пример 1 (Румшиский 1966). В некоторой партии изделий 4% брака, а из остальных – 75% первосортных. Какова вероятность, что отобранное наудачу изделие окажется первосортным? Обозначим извлечение стандартного изделия буквой A, а первосортного – буквой B.

$$P(\mathrm{A}) = 1 - 0.04 = 0.96;\ P(\mathrm{B/A}) = 0.75.\ P(\mathrm{AB}) = 0.96\cdot 0.75 = 0.72.$$

Пример 2. Какое количество очков более вероятно при броске двух костей, 11 или 12? Вот рассуждение Лейбница (Тодхантер 1865, с. 48): эти исходы равновероятны, потому что каждый может произойти только одним образом, а именно 1) при выпадении 5 и 6 очков 2) при выпадении 6 и 6 очков. Такова элементарная ошибка, допущенная великим человеком! Обозначим эти исходы при броске одной кости через A и B соответственно. Имеем P(A) = 1/6, P(B) = 1/6. Да, действительно, в обоих случаях оказывается, что P(AB) = P(A)P(B) = 1/36, но следует еще учесть, что первый случай может произойти в двух вариантах (а потому будет вдвое вероятнее): либо 5 и 6, либо 6 и 5.

Вообще же при P(B/A) = P(B) теорема умножения принимает вид

$$P(\mathrm{AB}) = P(\mathrm{A})P(\mathrm{B}),$$

а *события A и B называются независимыми*.

Пример 3. A и B, имея по 12 жетонов, играют с тремя костями при условии, что при каждом выкидывании 11 очков A должен отдать один жетон B, а B должен отдать жетон A при каждом выкидывании 14 очков, и выигрывает тот, кто первым завладеет всеми жетонами. Требуется определить соотношение шансов игроков на победу.

Это – Дополнительная задача № 5 из трактата Гюйгенса (1657), который привел только ответ, но не решение, сформулировал же ее Паскаль. Сразу скажем, что существуют 216 возможных бросков трех костей, из которых 15 благоприятны для выхода 14 очков, и 27 – для выхода 11 очков. Иначе, вероятности или шансы выигрыша относятся как 15/27 = 5/9. Выиграть требуется 12 раз, так что шансы победы относятся друг к другу как $5^{12}/9^{12}$.

Такова была первая из серии задач на *разорение игрока*, которые оказались исключительно интересными и исследовались, в числе прочих, Муавром и Лапласом. В частном случае предполагалось, что капитал одного игрока не ограничен. Вообще же серию азартных игр можно рассматривать как процесс случайного блуждания, а обобщение подобных задач переводят ее в случайный процесс (п. 5.2).

Почти очевидно обобщение теоремы умножения на большее число событий (скажем, на четыре):

$$P(A_1A_2A_3A_4) = P(A_1)P(A_2/A_1)P(A_3/A_1A_2)P(A_4/A_1A_2A_3).$$

Последний сомножитель, к примеру, обозначает вероятность события A_1 при условии, что произошли все три остальных.

Читатель, потерпи еще немного! Нужны еще два утверждения, каждое из них быть может не очень изящное – как на чей вкус.

1.1.1-4. Более существенное обобщение теоремы умножения. Пусть событие А может произойти совместно с одним и только одним из нескольких несовместимых событий $B_1, B_2, \ldots, B_n$. Из условия следует, что то, что раньше обозначалось P(AB), можно обозначить просто P(A), так что вместо формулы (1.3) можно будет написать

$$P(A) = P(B_1)\,P(A/B_1) + P(B_2)\,P(A/B_2) + \ldots + P(B_n)\,P(A/B_n) =$$

$$\sum_{i=1}^{n} P(B_i)P(A/B_i)\,. \qquad (1.4)$$

Формула (1.4) называется *формулой полной вероятности*. События B_i можно считать *причинами* появления события А; каждая из них приводит к его появлению, но только с некоторой присущей ей вероятностью.

Пусть, например, имеются три урны, содержащие 1 белый и 2 черных шара; 2 белых и 1 черный; 3 белых и 5 черных шаров. По жребию извлекаем шар из какой-то урны; какова вероятность, что он окажется белым?

Вероятности извлечь белый шар из этих урн равны $P(A/B_i)$ = 1/3, 2/3 и 3/8, а вероятности выбора каждой урны одни и те же и равны $P(B_i)$ = 1/3. Поэтому

$$P(A) = (1/3)x(1/3) + (1/3)x(2/3) + (1/3)x(3/8) = 0.458 < 1.$$

Иными словами, очень возможно, что шар окажется черным. Но не может ли оказаться, что P(A) > 1? Пусть

$P(A/B_1) = P(A/B_2) = P(A/B_3) = 0.99$
и, конечно же, $P(B_1) + P(B_2) + P(B_3) = 1$.

Но даже при таких высоких значениях первых трех вероятностей этого не произойдет. А может ли случиться, что $P(A) = 1$?

1.1.1-5. Формула Бейеса. Левые части равенств (1.2) и (1.3) совпадают, поэтому их правые части равны друг другу:

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

или, в предыдущих обозначениях,

$$P(A)P(B_i/A) = P(B_i)P(A/B_i),$$

откуда

$$P(B_i / A) = \frac{P(B_i)P(A / B_i)}{P(A)}. \quad (1.5)$$

Наконец, заменим $P(A)$ по формуле (1.4):

$$P(B_i / A) = \frac{P(B_i)P(A / B_i)}{\sum_{i=1}^{n} P(B_i)P(A / B_i)}. \quad (1.6)$$

Пора оглядеться. Причинам $B_1, B_2, \ldots, B_n$ мы приписали вероятности $P(B_1), P(B_2), \ldots, P(B_n)$, – и они входят в правую часть полученной формулы. Но вероятности эти были до-опытными, априорными, а затем произошел опыт: случилось-таки событие А! И теперь эти вероятности можно подправить, заменить апостериорными, после-опытными вероятностями $P(B_1/A), P(B_2/A), \ldots, P(B_n/A)$.

Формулу (1.6) можно найти в мемуаре Бейеса (1764) только в частном случае $n = 1$ с оставлением в знаменателе $P(A)$, но по укоренившейся традиции она всё же называется его именем. Точнее, с 1830 г. теоремой Бейеса стали называть его основную формулу (4.5), см. Хальд (1998, с. 141), но Cournot (Курно 1843, § 88), хоть вначале и с оговоркой, назвал так формулу (1.6). Появилась она не у Бейеса, а у Лапласа (1812, § 26).

А кто такой Бейес? Талантливый математик, оставивший миру посмертно опубликованный мемуар (1764 – 1765), о котором в начале XX в. оживленно спорили несколько десятилетий в связи с тем, что априорные вероятности редко бывают известны; можно ли, к примеру, считать, что они равны друг другу? Этот вопрос напоминает о подходе Лапласа (1814/1999, с. 861 левый столбец) к гипотезам вообще: их можно и нужно принимать, но непременно исправлять в соответствии с новыми данными.

Споры и сейчас не утихли, а вообще с его именем связано много современных терминов (Бейесовская оценка, Бейесовский принцип и т. д.), равно как и *теорема Бейеса*, см. выше.

И теперь пример. Для тех же трех урн, которые мы рассматривали выше, дроби в правой части формулы (1.5) будут отличаться друг от друга только сомножителями P(A/B$_i$), которые находятся в соотношении (1/3):(2/3):(3/8) = 8:16:9. В этом же соотношении, стало быть, окажутся и апостериорные вероятности P(B$_i$/A). Можно, конечно, учесть ранее установленное P(A) = 0. 458, и вычислить их:

$$P(B_1 / A) = \frac{1}{3\cdot 3\cdot 0.458},\ P(B_2 / A) = \frac{2}{3\cdot 3\cdot 0.458},\ P(B_3 / A) = \frac{3}{3\cdot 8\cdot 0.458}.$$

Неудивительно, что наивысшей оказалась вероятность P(B$_2$/A): именно во второй урне была самая большая доля белых шаров.

Стиглер (1983/1999) применил "теорему Бейеса" в указанном выше частном случае для утверждения о том, что мемуар Бейеса на самом деле написал другой английский математик, Сондерсон. Он (с. 300) назначил субъективные вероятности трем различным предположениям (например, поддерживал ли каждый из них отношения с Муавром) и перемножил их по отдельности для Сондерсона и Бейеса. Их отношение оказалось равным 3:1 в пользу первого, и, молчаливо допустив равенство соответствующих априорных вероятностей, Стиглер (с. 301) заключил, что вероятность авторства Сондерсона втрое выше.

Мы считаем, что допущение Стиглера было совершенно неприемлемо и что его (давно забытый) вывод следует решительно отвергнуть. Этот же Стиглер позволил себе выблевать хулу на Эйлера (п.6. 2) и Гаусса (Шейнин 1999b, с. 463 – 466).

1.1.1-6. Субъективная вероятность. Она, конечно же, является априорной и несколько дополняет теоретическую (1.1). В косвенном виде субъективная вероятность применяется очень часто, и особенно тогда, когда "нет основания" считать, что отдельные случаи не равновозможны. Так, вероятность выпадения каждой грани игральной кости принимается равной 1/6, хотя каждая данная кость не совсем *правильна*.

Впервые о субъективной вероятности четко заявили Пуассон (1837, с. 30 – 31) и Курно (1843/1970, с. 28 – 29), которые даже назвали ее и объективную вероятность различными терминами, – *шансом* и *вероятностью*. Вот поучительная задача Пуассона:

Урна содержит n шаров, белых и черных, число которых по отдельности неизвестно. Требуется определить вероятность извлечения белого шара. Число белых шаров может быть равно 0, 1, 2, …, n, – всего (n + 1) равновозможных (будто бы!) случаев. Искомая вероятность поэтому равна среднему из всех возможных вероятностей

$$\frac{1}{n+1}(\frac{n}{n}+\frac{n-1}{n}+...+\frac{1}{n}+\frac{0}{n}) = \frac{1}{2}.$$

Но как можно вывести какое-то определенное следствие из полного незнания? Очень просто: можно установить *субъективное* следствие. Кстати, полученный результат не противоречит теории информации: половинная вероятность соответствует наименьшему количеству информации (о составе урны).

Пуассон применил субъективную вероятность и при исследовании одной азартной игры, в которой две серии карт извлекались из одной и той же пачки, состоящей из шести колод, и молчаливо посчитал, что вторая серия карт как бы извлекается из исходной нетронутой колоды. На самом деле это совсем не так, но игроки не имели возможности поступать иначе и исходили из принципа: неизвестное изменение состава пачки не изменяет вероятности извлечения того или иного количества очков.

Именно так, кажется, обязаны поступать банкомёты при игре в очко: после каждого кона игра продолжается с оставшейся частью пачки, но во избежание перебора они, как и в начале игры, останавливаются на 17 очках. Игрок, обладающий острой памятью, может, однако, изменять свое поведение от кона к кону.

Вот дальнейшие примеры. В Ветхом завете весьма кратко, а в Иерусалимском Талмуде (Санхедрин 1^4) подробнее описаны жеребьевка при выкупе за перворожденных. Моисей подготовил 22 000 билетиков, освобождающих от выкупа, и еще 273, требующих выкупа, однако затем добавил еще 273 билетика первого типа, потому что участники жеребьевки решили, что последним из них может и не остаться освобождающих билетиков. Ясно, что Моисей подверг себя (незначительному) риску недобора ожидаемой суммы, но этого не произошло. Для нас важнее другое: можно доказать (Тутубалин 1972, с. 12), что с самого начала жеребьевка была справедливой.

Второй пример (Рабинович 1973, с. 40) можно считать вероятностным софизмом. В XIII в. или в самом начале XIV в. один из комментаторов Талмуда рассуждал о запрещенной (не кошерной) пище. Вот мое собственное толкование. Если, помимо нескольких кусков кошерного мяса (например, пяти), на тарелке лежит один кусок запрещенного, то можно съесть все куски подряд. И ведь действительно, вероятность не нарушить запрет остается неизменной и равной соотношению кошерных кусков к тому же числу, увеличенному всего на единицу (в нашем примере 5/6)!

Комментатор же разъяснил иначе: вероятность (как мы сказали бы) съесть запрещенный кусок вначале не слишком высока, а когда кусков останется всего два, можно будет сказать: запрещенный кусок я уж, наверное, съел, так что эти – кошерные. Тоже софизм!

Субъективные *мнения* являются предметом математических исследований, и примерами могут служить экспертные оценки и системы голосования. В обоих случаях требуется расположить по порядку достоинства те или иные, скажем, экономические проекты и кандидатов на выборные должности, см. п. 5.1.

1.1.2. Геометрическая вероятность

Классическое определение вероятности можно обобщить, и первым так фактически и поступил Ньютон. В рукописи 1664 – 1666 гг., увидевшей свет лишь в 1967 г., он рассмотрел мысленный опыт: падение мяча на центр круга, разделенного на два сектора, соотношение площадей которых было равно $2:\sqrt{5}$. Он указал, хоть и не прямо, что вероятности падения мяча в них будут находиться в том же соотношении, т. е. будут равны $2/(2 + \sqrt{5})$ и $\sqrt{5}/(2 + \sqrt{5})$.

В общем случае оказывается, что классическое определение остается в силе, однако m и n можно считать действительными, а не только натуральными числами. Именно так, но опять же только фактически, *геометрическую вероятность* применяли многие последующие математики. Бюффон (1777, § 23) окончательно ввел ее в обиход своей знаменитой задачей: Игла длиной $2r$ падает “случайным образом” на пучок параллельных прямых, расположенных на расстоянии $a > 2r$ друг от друга. Требуется определить вероятность того, что она пересечет одну из них. Ясно, что вероятность должна зависеть от соотношения r/a, а подсчет показывает, что она равна

$$P = \frac{4r}{\pi a}.$$

Измерим теперь r и a, заметим, сколько пересечений произойдет после, скажем, 100 падений иглы и вычислим по этой формуле число π, считая его неизвестным. Таков был первый пример, придуманный Лапласом (1812, гл. 5), на применение *метода Монте-Карло или статистических испытаний*, как его сейчас называют.

Эту задачу неоднократно обобщали, заменяя, например, параллельные прямые сеткой прямоугольников или квадратов.

Бюффон пояснил, что имел в виду “ввести геометрию в свои права” в теории вероятностей. Но лишь Курно (1843, § 18) посчитал нужным формально определить новое понятие. Точнее, он предложил общее определение и для дискретного, и для непрерывного вариантов случайной величины и назвал вероятностью отношение *протяженности* благоприятных случаев к их общей *протяженности*. Сейчас протяженность заменили бы *мерой* (в частности – площадью).

Геометрической вероятностью фактически пользовались все авторы, вводившие непрерывные законы распределения случайных величин (см. п. 2.1), т. е. начиная с Николая Бернулли (1709/1975, с. 296 – 297), см. также Тодхантер (1865, с. 195 – 196). Ее же, можно сказать, применил Больцман (1868/1984, с. 30), который определял вероятность скорости молекулы быть заключенной в бесконечно малом промежутке как отношение времени, когда это событие имело место, ко всему периоду наблюдений. Здесь можно было бы вспомнить эргодическую гипотезу, но это выходит за рамки нашей темы.

Знаменитым стал парадокс Бертрана. Вероятность того, что длина случайной хорды данного круга окажется меньше стороны правильного вписанного в круг треугольника, как оказалось, зависит от того, как именно проведена эта хорда. Сам он предложил три решения, Чубер (Czuber 1903/1968, с. 107 – 108) – еще три, а De Montessus (1903), хоть его выкладки и содержали непростительные ошибки, доказал, что множество решений несчетно.

Мы ограничимся здесь указанием, что понятие *случайно* нельзя считать вполне определенным. Вот, к примеру, два истолкования: хорда должна пройти через *случайно* заданную точку круга; она должна иметь *случайно* заданное направление.

Закончим решением задачи о встрече (Laurent 1873, с. 67 – 69): двое договорились встретиться в определенном месте, куда каждый приходит независимо от другого в любое время от 12 до 13 часов. Пришедший первым ожидает 20 минут, после чего уходит. Какова вероятность встречи (Гнеденко 1954, с. 36)? Условие, да и решение можно представить графически (Рис. 1.1). Если обозначить моменты прихода обоих на встречу через x и y, то по условию $|x - y| \le 20$, или, что то же, $|y - x| \le 20$, т. е. $- 20 \le y - x \le 20$, $x - 20 \le y \le x + 20$.

На Рис. 1 показан квадрат со стороной 60 (минут), а его площадь соответствует всем возможным случаям прихода обоих на встречу. Возможность встречи соответствует площади заштрихованной фигуры и искомая вероятность равна отношению этой площади к площади квадрата. Текущую точку этого квадрата (x; y) можно назвать значением двумерной случайной величины.

Здесь же удобно показать, как истолковать физические понятия “выше” и “ниже” прямой, а точнее: что можно сказать взамен?

Находится ли точка B (0; 60) выше “верхней” прямой AC? Этого мы не знаем. Но можем твердо сказать: точки B и O (начало координат) находятся по разные стороны от этой прямой. Подставим координаты B в уравнение прямой; окажется, что 60 > 0 + 20 или иначе: для этой точки $y > x + 20$. Для начала координат, напротив, окажется, что $y < x + 20$, ч. т. д.

1.1.3. Статистическая вероятность

Если при ν испытаниях случайное событие произошло μ раз, то его частость или статистическая вероятность равна

$$\hat{p} = \frac{\mu}{\nu} \tag{1.7}$$

и изменяется, как это очевидно, от 0 до 1.

Заканчивая описания своего мысленного опыта (см. п. 1.1.2) Ньютон добавил:

Если игральная кость является не правильным телом, а параллелепипедом или имеет неравные грани каким-то иным образом, можно определить насколько легче добиться одного исхода, нежели другого.

Можно думать, что он имел в виду статистическое определение вероятностей, Граунт же в 1662 г. все свои выводы о статистике населения и медицинской статистике основывал на статистических данных и статистических вероятностях.

Статистическая вероятность применялась фактически издавна, хотя и основывалась она не на числовых данных, а на общих представлениях. Так, Celsus (1935, с. 19) указал, что “Внимательные люди замечали, что именно в общем подходит, и стали назначать то же самое своим пациентам. Так возникло искусство врачевания”. Видимо тогда же было замечено, что одно и то же средство лечения может совсем не “подходить” некоторым пациентам.

Но указанное выше определение имеет смысл только, если испытания взаимно независимы, а вычисленная вероятность остается почти без изменения в новой серии испытаний, проведенных в тех же условиях. Опять-таки только в этом случае событие можно считать случайным. Пусть, например, исследуется относительное количество дорожных происшествий. Если оно существенно изменяется от одного дня недели к другому, то следовало бы изучать каждый день недели по отдельности. А какие испытания независимы друг от друга? Пока скажем так: те, результаты которых не влияют друг на друга.

Несовершенство теоретической вероятности и недостаточно широкое поле ее приложений привели к появлению статистической вероятности как основного исходного понятия, и ввел ее в этом качестве Рихард Мизес в 1920-е годы. Воплощение его простой идеи в строгом виде оказалось исключительно трудным, и споры по поводу его *частотной теории* так и не закончились.

Вот эта идея: будем многократно подбрасывать монету и время от времени вычислять частость (1.7) появления герба. После достаточно большого ν эта величина будет изменяться лишь незначительно и при $\nu \to \infty$ достигнет некоторого предельного значения. Его-то Мизес и предложил называть статистической вероятностью соответствующего случайного события.

Бесконечного количества испытаний быть не может; Мизес, однако, добавляет, что подобные предельные переходы известны в физике и механике. Так, мгновенная скорость есть предел скорости при стремлении к нулю соответствующего отрезка времени. Он также указал, что последовательность рассматриваемых испытаний (*коллектив*, по его выражению) должна быть иррегулярна; иначе говоря, любая бесконечная подпоследовательность, извлеченная из нее, должна приводить к тому же значению вероятности $\hat{p}$.

Последнее требование слишком неопределенно. Сколько подпоследовательностей необходимо проверить, чтобы убедиться в иррегулярности? И нельзя ли "случайно" выбрать слишком своеобразную? Даже такие поверхностные замечания указывают на громадные трудности теории Мизеса, но ей по необходимости, за неимением лучшей, широко пользуются естествоиспытатели.
Да, она теоретически несовершенна; хотя отношение математиков к ней смягчилось (Колмогоров 1963, с. 369), определить случайность по Мизесу не удается (Успенский и др. 1990, § 1.3.4).

1.1.4. Независимость событий и наблюдений

Понятие о независимости событий мы ввели в п. 1.1.1-3. События A и B независимы, если

$$P(\text{AB}) = P(\text{A})P(\text{B}), \quad (1.2)$$

в противном же случае

$$P(\text{AB}) = P(\text{A})P(\text{B/A}).$$

Если же считать событие B первым, а A – вторым, то

$$P(\text{AB}) = P(\text{B})P(\text{A/B}). \quad (1.3)$$

Из этих простых равенств можно вывести примечательное следствие: Если A не зависит от B, то и B не зависит от A; *свойство независимости является взаимным, и два события независимы, если появление одного из них не влияет на вероятность появления другого*. Доказательство несложно. По условию $P(\text{A/B}) = P(\text{A})$. Изменим соответственно формулу (1.3):

$$P(\text{AB}) = P(\text{B})P(\text{A})$$

и сравним полученное с формулой (1.2):

$$P(\text{B/A}) = P(\text{B}), \text{ ч. т. д.}$$

Но вот вроде бы противоречащий пример: если считать случайной летнюю погоду в данном городе в течение, скажем, недели, то можно будет сказать, что от нее зависит также случайный объем продажи прохладительных напитков. Обратной зависимости, однако, никак не существует. И всё-таки следствие остается в силе: погода и торговля не относятся к одной и той же группе событий.

Независимость событий впервые упомянул Муавр (1712, предисловие), а во Введении к своему основному сочинению, *Учению о шансах*, определил ее: Два события независимы, "если они не связаны друг с другом, и появление одного из них не способствует, и не препятствует появлению другого". Два события зависимы, “если вероятность появления одного из них изменяется после появления другого”. Независимость таким образом оказалась взаимной, ср. выше!

После Муавра независимость событий упоминали и другие классики теории вероятностей, особо см. ниже, и всё-таки, имея дело с независимыми событиями, авторы забывали про нее. Положение резко изменилось после того, как Марков специально исследовал *цепи Маркова* (п. 5.2) и тем самым добавил теории вероятностей новое направление, – изучение зависимых случайных событий и величин.

Гаусс (1823b, § 18) считал, что если некоторое наблюдение является общим для двух функций результатов наблюдений, то погрешности этих функций не будут независимыми друг от друга. Зависимыми он (1828, § 3) считал и результаты уравнивания. Так, если в измеренные углы треугольника введены поправки, и их сумма стала точно равна теоретической, то эти *уравненные* углы уже не являются независимыми (потому что в них остались неизбежные погрешности, которые как-то связывают их).

Таким образом, Гаусс рассматривал независимость функций случайных величин (погрешностей наблюдений), см. п. 1.2.3.

Мнение Гаусса всегда разделялось геодезистами. В триангуляции Советского Союза отдельные звенья триангуляции начинались и заканчивались базисами и астрономически измеренными направлениями (азимутами), так что после их предварительного уравнивания можно было считать звенья не зависимыми друг от друга и вводить их в общее уравнивание *четырехугольников* (*полигонов*), составленных из них, в качестве достаточно независимых элементов. Базисы и азимуты всё-таки оказывались общими, но они измерялись точнее, чем углы треугольников. Скажем более четко по поводу базисов: влияние относительной погрешности их измерения было меньше влияния ошибок угловых измерений.

1.2. Случайность и случайная величина

1.2.1. Случайность

Вначале была случайность философским понятием, затем в известном смысле перешла в математику. Обратимся поэтому снова к Аристотелю, который включил ее в свое учение о причинах, и приведем три из его многочисленных примеров случайности.

1) Роя яму для дерева, некто отыскивает вклад; *Метафиз*. 1025a.

2) Двое знакомых встречаются, – сталкиваются, – хоть и не имели такой цели; *Физ*. 196b30.

3) У самки животного рождается самка, а не самец (у женщины – девочка, а не мальчик); *Физ*. 199b1 и *О возникновении животных* 767b5.

В первых двух примерах есть нечто общее: будь яма вырыта чуть в стороне, клад остался бы в земле; запоздай (или поспеши) хоть один из двоих на минуту, встреча не состоялась бы. Суть этого второго примера в том, что “случайно” пересекаются две не зависящие друг от друга цепи событий. Многие древние ученые представляли себе случай именно таким образом, и это пояснение подхватил Курно (1843, § 40). Малые причины (выход хотя бы одного из двоих чуть раньше или позже) приведут к существенному последствию (встречи не будет). В начале XX в. такой схемой Пуанкаре (1896/1987, с. 4 – 6; с. 11 – 12 в переводе) объяснил возникновение случая, добавив, однако: *при неустойчивом равновесии*. Вот его пример, который использовал и Курно (1843, § 43): поставьте вертикально прямой круговой конус вершиной на горизонтальную плоскость и отпустите руку. Конус упадет в *случайном* направлении.

Подобный пример есть и у Галена (1951, с. 202), римского врача и естествоиспытателя II в.: молодой человек может безнаказанно довольно серьезно нарушить свой режим, но у пожилого человека уже небольшое его нарушение может привести к серьезным последствиям. Пожилой живет в неустойчивом равновесии!

Но вспомним третий пример Аристотеля. Рождение самки (девочки) – это нарушение цели природы, первая степень уродства, которое тем не менее необходимо. Таково было первое, притом неудачное диалектическое противопоставление случайности и необходимости.

Нарушение цели природы не было забыто. Кеплер, открывший законы движения планет и, в частности, установивший, что они движутся по эллиптическим орбитам, считал, что целью природы были круговые, как бы более совершенные орбиты, эллиптичность же появилась потому, что эта цель не была достигнута (1618 – 1621/1952, с. 932). Только Ньютон доказал, что эллиптичность следует из его закона всемирного тяготения, и что мера уклонения эллиптической орбиты данной планеты от круговой была определена ее начальной скоростью.

Но не может математическая теория основываться на встречах знакомых и целях природы. Оставим в стороне весьма интересные, но намного опередившие свое время, а потому безуспешные попытки Ламберта (1771, §§ 323 – 324; 1772 – 1775), см. также Шейнин (1971, с. 245 – 246), определить случайность в математическом смысле и сразу скажем: задана бесконечная числовая последовательность из нулей и единиц, например

0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 0, …

Случайна она или нет? Вопрос оказался фундаментальным, к его решению подходили с разных сторон, но он еще далеко не закрыт и еще сложнее оказался тот же вопрос для конечной последовательности. Во всяком случае требуется, чтобы начальный отрезок бесконечной последовательности был иррегулярным, не законосообразным. Иррегулярность таким образом является существенным свойством случайности (как, впрочем, полагал и Мизес).

Случайность в философии противопоставляется необходимости, и в естествознании эти понятия сопоставил Пуанкаре (1896/1987, с. 1, с. 9 в переводе): “Ни в одной области точные законы не определяли всего, они лишь очерчивали пределы, в которых дозволялось пребывать случаю”. Уместно также вспомнить знаменитое высказывание Лапласа (1814/1999, с. 835 левый столбец), будто бы доказывающее его отрицание случайности: для всеведущего ума, способного на любые вычисления, случайности не существовало бы и “будущее как и прошлое предстало бы перед его взором”.

Но раз такого ума нет, не о чем и говорить. Кроме того, существуют неустойчивые движения, чувствительные к малым погрешностям начальных условий (п. 5.1). Наконец, *лапласов детерминизм* до него высказывали Мопертюи и Бошкович (Шейнин 2005, с. 129).

Мы сказали *будто бы доказывающие* … Возможно, что все астрономические исследования Лапласа и наверняка его работы по статистике населения опровергают его (действительно имевшие место) высказывания, отрицавшие случайность.

1.2.2. Причина или случай?

Что можно подумать, если монета выпадет одной и той же стороной 10 раз кряду? 20 раз? Здравый смысл подскажет: монета жульническая! И всё-таки мы обсудим этот пример. В конце концов, после появления в середине XIX в. неевклидовой геометрии здравому смыслу в математике можно доверять лишь в первом приближении.

Стороны монеты обозначим символами + и –. После двух бросков результаты могут быть + +, + –, – + и – –, и все они равновероятны. При третьем броске исход + + изменится либо на + + +, либо на + + –. Иначе говоря, исход + + + не окажется менее вероятным, чем каждый из семи остальных, и легко видеть, что то же заключение останется в силе при любом другом числе бросков. 10 плюсов подряд, конечно же, маловероятны, но точно такую же низкую вероятность будут иметь все остальные возможные исходы.

Каков же вывод? Послушаем Лапласа (1776, с. 152; 1814/1999, с. 837 левый столбец), который обсуждал так называемую с тех пор задачу Даламбера – Лапласа:

На столе находятся печатные буквы, расположенные в следующем порядке: ИНФИНИТЕЗИМАЛЬНЫЙ. Причина, которая заставляет нас думать, что это расположение не вызвано случаем, нельзя обосновать тем, что […] *оно менее вероятно, чем остальные*.

Здесь нет никаких формул, к которым можно было бы обратиться, и Лаплас сразу же заключает, что кто-то специально расположил эти буквы, потому что они обозначают осмысленное слово. В нашем примере, как можно думать, кто-то специально постарался, чтобы монета всегда выпадала одной и той же стороной, так что здравый смысл нас не обманул.

Лаплас позднее заменил выбранное им слово на *Константинополь*, которое с самого начала принял Даламбер (1767, с. 245 – 255). Впрочем, его рассуждение было менее разумным.

Вообще же сформулированный в заглавии вопрос заставляет подумать о разделении равновероятных случаев (если таковые существуют) на обычные и примечательные; расположение букв, приводящее к слову *Константинополь*, было поистине примечательным. Кеплер, как известно, был и астрологом (и считал себя основателем научной, якобы, астрологии о мягком влиянии неба на человека) и рассуждал об *аспектах*, т. е. о примечательных расположениях Солнца, Луны и планет. Он (1601, § 40/1979, с. 97) указал, что *добавил* три аспекта к прежним пяти, признаваемым древними. Он же (1604/1977, с. 337) заявил, что появление Новой звезды было не обычным событием, “как при игре игральной костью, а великим чудом”. Если и не чудом, то наверняка примечательным событием.

Другая примыкающая тема – суеверия и самообман игроков, да и не только их. Вдоль окружности колеса рулетки катится шарик, остановка которого в 37 положениях равновероятна. Игроки ставят на любой *номер*, – 0, 1, 2, …, 35, 36, – и угадавший забирает ставки; впрочем, при остановке шарика на нуле выигрывает банкомёт, а проще – владелец игорного заведения. Такова схема простейшего варианта игры в рулетку.

Три раза подряд шарик остановился на номере 18; благоразумно ли иметь это в виду (а если да – то как именно) при следующей игре? Вот замечание Лапласа (1814/1999, с. 855 левый столбец):

Когда во французской лотерее какой-либо номер уже долгое время не выходил, толпа […] *спешит покрыть его ставками. Она считает, что номер, долгое время не выходивший, должен при ближайшем тираже выйти преимущественно перед другими.* […] *Благодаря иллюзии, противоположной предыдущ*[ей], *отыскивают в предшествующих тиражах* […] *наиболее часто выходившие номера* […]

Самый убедительный довод против подобного самообмана привел Бертран (1888, с. XXII): *рулетка не имеет ни воли, ни памяти*. Вновь вспомним Лапласа (1814/1999, с. 855 левый столбец):

Я видел, как люди, страстно желавшие иметь сына, с прискорбием узнавали о рождениях мальчиков в том месяце, когда они должны были стать отцами. Воображая, что отношение этих рождений к рождениям девочек должно быть одним и тем же к концу каждого месяца, они считали, что уже рожденные мальчики делают более верным ближайшие рождения девочек.

Подобных примеров можно привести много, но общее в них одно: не следует считать явно *не зависящие друг от друга события* зависимыми; напротив, следует помнить изречение Бертрана.

Но нельзя ли перехитрить рулетку? Убедиться в ее правильной работе, т. е. в примерно одной и той же частости выхода всех номеров, трудно, но допустим, что это условие соблюдено. Поставим некоторую сумму на какой-то номер. Проиграли? Поставим на него же вдвое больше (если это допускается). Снова проиграли? Поставим на него же вчетверо больше и т. д. Когда-нибудь выиграем (а не только вернем потери), но хватит ли денег? Не выест ли роса очи, пока Солнце взойдет? Есть хорошая поговорка для тех, кто может позволить себе небольшой проигрыш: *Играй, да не отыгрывайся*! Пушкин выбрал ее в качестве эпиграфа к одной из глав своей *Капитанской дочки*, и лучше не скажешь.

1.2.3. Случайная величина

Непосредственно в теории вероятностей применяется понятие случайной величины (более выразителен английский термин *случайная переменная*). Вот ее простейшее определение в *дискретном* варианте: *переменная, принимающая различные значения, каждое из них с некоторой вероятностью*. Обозначим эти значения $x_1, x_2, \ldots, x_n$, а соответствующие вероятности – $p_1, p_2, \ldots, p_n$, сумма которых должна равняться единице. Совместно рассматриваемые значения случайной величины и их вероятностей называются ее *законом распределения*. Случайное событие можно рассматривать как частный случай случайной величины при $n = 2$.

Вариант $n \to \infty$ не исключается; осуществляться он может либо дискретно, например $x_1 = 1, x_2 = 2, x_3 = 3, \ldots$, при *счетном* числе значений, либо *непрерывно*, если, например, случайная величина принимает все (*несчетные*) значения в интервале [0; 1]. В обоих этих случаях появляется нечто новое: может наступить событие, вероятность которого равна нулю. Назначим, например, наугад точку в указанном интервале. Вероятность выбора любой наперед заданной точки несомненно равна нулю, но ведь какую-то точку мы выбрали! Здесь можно вспомнить геометрическую вероятность (п. 1.1.2).

Во всех задачах теории вероятностей должна участвовать случайная величина (или ее обобщение, о чем мы говорить не будем, или случайное событие). Так, при броске игральной кости случайной величиной является его исход; он принимает шесть различных значений, каждое со своей вероятностью (в данном примере все они совпадают друг с другом).

Можно привести множество интересных примеров случайной величины. Так, в начале XVII в. участники знаменитой генуэзской лотереи (точнее: числового лото) могли по желанию отгадывать один, два, …, пять номеров из 90. Выигрыши соответственно возрастали, хотя никак не пропорционально трудности угадывания (Бирман 1957). Здесь случайная величина (случайный выигрыш) имела 5 значений со вполне определенными вероятностями, вычислять которые, однако, могли лишь считанные единицы. С 1662 г. начала оцениваться случайная продолжительность человеческой жизни (Граунт).

Наконец, Симпсон (1756 – 1757) ввел случайную величину в будущую теорию ошибок, – в теорию математической обработки наблюдений, – которая примерно с того времени и вплоть до 1930-х годов оказалась важнейшей областью приложения теории вероятностей. Он принял, что шансы различных ошибок, которыми может быть искажено каждое наблюдение (данной серии), выражаются определенными числами. Иначе говоря, каждое наблюдение явилось возможным значением некоторой случайной величины, а все наблюдения вместе опять же представляли собой (малую часть) возможных значений той же величины, на этот раз с указанием вполне определенных шансов (а потому и вероятностей).

Понятие случайной величины оказалось исключительно важным, но вот формально ее ввел только Пуассон (1837, с. 140 – 141 и 254), да и то неуверенно, назвав ее *вещью А*. Сам термин *случайная величина* вошел в употребление далеко не сразу. Быть может последним его противником был Марков (Ондар 1977, с. 71, письмо А. А. Чупрову 1912 г.), который еще в начале ХХ в. считал его неопределенным. Ничего лучшего он, однако, не придумал, и нередко так и писал: *неопределенная величина*, что никак не было более понятно.

Можно считать, что в известном смысле всё развитие теории вероятностей сводилось ко всё более общему пониманию случайной величины. Вначале изучался случай вообще (фактически – случайная величина с равномерным распределением, см. п. 2.2.1), который противопоставлялся необходимости, затем стали изучаться случайные величины, обладающие различными распределениями, и зависимые величины и случайные функции, ср. п. 5.2.

Скажем и иначе: в теории вероятностей постепенно повышался уровень абстракции, что вообще характерно для развития математики в целом. Хорошо известно, что чем выше оказывался этот уровень, т. е. чем дальше математика отходила от природы, тем полезнее она становилась. Вот комплексные числа и функции комплексных переменных: насколько они чужды природе, настолько же они полезны и самой математике, и ее приложениям.

Глава 2. Законы распределения случайной величины, их характеристики и параметры

2.1. Функция и плотность распределения

Для характеристики непрерывной случайной величины (назовем ее ξ) нам потребуется установить ее закон распределения, как это и было сделано в п. 1.2.3 для дискретных величин. Вероятность того, что она меньше некоторого x, мы обозначим через $F(x)$:

$P(\xi < x) = F(x).$

Эта $F(x)$ называется *функцией* (или *интегральной функцией*) *распределения* величины ξ. Если ξ может принимать любые значения от $-\infty$ до ∞, то

$$P(\xi < -\infty) = F(-\infty) = 0,\ P(\xi < \infty) = F(\infty) = 1.$$

Выберем теперь две произвольные точки, x_1 и x_2, $x_2 > x_1$, тогда

$$P(\xi < x_2) \geq P(\xi < x_1) \text{ или } F(x_2) \geq F(x_1).$$

Действительно, $P(-\infty < \xi < x_2)$ не может оказаться ниже, чем $P(-\infty < \xi < x_1)$. В крайнем случае, если на отрезке $[x_1; x_2]$ случайная величина не принимает никаких значений (но остается непрерывной вне его), то

$$P(\xi < x_2) = P(\xi < x_1) \text{ или } F(x_2) = F(x_1). \tag{2.1}$$

Итак, функция $F(x)$ во всяком случае не убывает, а если (2.1) не имеет места, то возрастает. И заметим еще, что

$$F(x_2) - F(x_1) = P(\xi < x_2) - P(\xi < x_1). \tag{2.2}$$

Интегральные функции распределения вошли в употребление в XX в., хотя мимолетно они появлялись с 1669 г. В том году Гюйгенс (1669/1895, между с. 530 и 531) с методической целью нарисовал график функции, уравнение которой можно записать в виде

$$y = 1 - F(x),\ 0 \leq x \leq 100.$$

Кривая характеризовала продолжительность человеческой жизни (ξ), а именно вероятность $P(\xi \geq x)$, но не была основана на количественных данных. Ту же вероятность исследовал Муавр в 1725 г. и, аналогично, вероятность свободного пробега молекулы не меньшего x, определял Клаузиус (1858/1867, с. 268).

До внедрения функций распределения для характеристики непрерывной случайной величины применяли *плотность ее распределения* (иначе: *ее вероятности*) $\varphi(x)$, см. Рис. 2. Рассмотрим бесконечно малый промежуток $[x_1; x_1 + dx_1]$ на числовой оси. Интересующая нас случайная величина принимает на нем значение, зависящее от x_1; мы можем сказать – принимает одно и то же значение $\varphi(x_1)$, потому что длина отрезка бесконечно мала. На примыкающем справа отрезке той же длины значение этой функции можно принять равным $\varphi(x_2)$, где $x_2 = x_1 + dx_1$. И таким образом мы получим ряд значений $\varphi(x_1)$, $\varphi(x_2)$, … и теперь сможем соотнести эту функцию, – плотность, – и интегральную функцию распределения $F(x)$:

$$F(x_n) = \int_{-\infty}^{x_n} \varphi(x)dx,\ F(x_1) = \int_{-\infty}^{x_1} \varphi(x)dx,\ F(x_n) - F(x_1) = \int_{x_1}^{x_n} \varphi(x)dx\,.$$

Эта формула дополнительно разъясняет равенство (2.2). Строго говоря, по определению полагают, что

$$F'(x) = \varphi(x),$$

но приведенное выше разъяснение сути $\varphi(x)$ конечно же остается в силе. В более простых примерах плотность оказывается непрерывной функцией на всем конечном или бесконечном интервале своей области определения. В соответствии со своим определением площадь "под" кривой плотности должна равняться единице.

Термины *под, над, выше, ниже, влево, вправо* – не являются математическими, потому-то мы употребили один из них в кавычках. Ограничимся этим пояснением и позволим себе в дальнейшем применять их без угрызения совести и без кавычек. В более *высоких* математических сочинениях так, однако, не поступают.

Сразу скажем, что *вместо случайных величин теория вероятностей изучает их функции распределения или плотности*, как, например, в алгебре изучают квадратные трехчлены

$$f(x) = ax^2 + bx + c,\ a \neq 0.$$

Зная параметры трехчлена a, b и c, мы можем установить, будут ли корни трехчлена действительны или комплексны, а в первом случае – совпадающими или нет, и т. д. Таким же образом, изучая указанные функции, мы устанавливаем поведение соответствующей случайной величины.

Стоп! А где же параметры плотности или функции распределения? Функцию $f(x)$ мы можем записать в виде $f(x; a; b; c)$, чтобы указать: ее аргументом является переменная x, но ее поведение определяют еще и параметры, постоянные для каждой данной функции, т. е. для каждого данного трехчлена. И плотность распределения тоже имеет параметры, характерные для каждой случайной величины.

Как правило, статистики оценивают параметры функций или плотностей распределения. Пусть, например, известно, что имеет место непрерывное треугольное распределение (предположения такого рода далеко не всегда оправданы) с неизвестным параметром a (см. п. 2.2.2). Требуется *оценить* его, т. е. установить для него некоторое значение $\hat{a}$, что возможно только по результатам наблюдений, и определить его возможную погрешность. Если закон распределения имеет, к примеру, два параметра, то, естественно, требуется оценить их оба. Опираться можно здесь только на наблюдения изучаемой случайной величины, так что значение $\hat{a}$ окажется выборочным.

2.2. Некоторые распределения

2.2.1. Равномерное распределение

Случайная величина, обладающая этим распределением, принимает все свои значения с одной и той же вероятностью. Так, при броске игральной кости исходы 1, 2, …, 5, 6 равновероятны (и вероятность каждого равна 1/6).

В непрерывном варианте случайная величина принимает одно и то же значение на некотором интервале, например, на отрезке [– a; a], см. Рис. 3. И так как площадь под этим отрезком должна равняться единице, то находиться он должен будет выше оси Ох на $1/a$, так что

$$\varphi(x) = 1/a = \text{Const}.$$

Параметром равномерного распределения можно считать число a.

2.2.2. Непрерывное треугольное распределение

Плотность φ (x) здесь является ломаной АСВ.

Можно установить уравнения прямых, которым принадлежат отрезки АС и ВС:

$$y = (1/a^2)x + (1/a) \text{ и } y = -(1/a^2)x + (1/a).$$

Если для проверки подставить координаты точек А и С в первое уравнение и координаты С и В – во второе, то в обоих случаях уравнения превратятся в числовые равенства, ч. т. д. Эти же уравнения окажутся уравнениями функции плотности, если только добавить к ним необходимые ограничения: к первому уравнению – интервал $-a \le x \le 0$, – т. е. от точки А до точки О, – а ко второму $0 \le x \le a$.

Величина а является (единственным) *параметром непрерывного треугольного распределения.* Впрочем, об этом можно было догадаться и раньше, поскольку она и только она определяла координаты точек А, В и С.

Треугольное распределение мы выбрали главным образом потому, что было легко установить смысл его параметра. Ввел его Симпсон (п. 1.2.3).

2.2.3. Биномиальное распределение

Все мы помним формулу квадрата суммы двух чисел, а некоторые ухитрились не забыть и куб суммы двух чисел. Но есть и общая формула для натурального показателя $n = 1, 2, 3, \ldots$:

$$(p + q)^n = p^n + C_n^1 p^{n-1}q + C_n^2 p^{n-2}q^2 + \ldots + C_n^{n-1} pq^{n-1} + q^n. \quad (2.3)$$

Впрочем, мы ограничимся подобным разложением для случая $p + q = 1$, т. е. для тех величин, которые могут описывать вероятности противоположных событий. Величинами C_n^k здесь обозначены количества сочетаний из n элементов по k:

$$C_n^k = \frac{n\ (n-1)\ \ldots\ (n-k+1)}{k!},\ C_n^k = C_n^{n-k}.$$

В числителе столько же сомножителей, сколько в знаменателе. К примеру,

$C_5^3 = \frac{5 \cdot 4 \cdot 3}{3!}$, $3! = 1 \cdot 2 \cdot 3$.

Пусть теперь спрашивается: какова вероятность дважды выкинуть единицу при броске четырех костей (или четырехкратном броске одной)? При одном броске вероятность *единицы* равна p = 1/6, а вероятность всех остальных исходов, взятых совместно, равна q = 5/6. Выпишем разложение бинома [(1/6) + (5/6)] в четвертой степени:

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

Искомой вероятности будет соответствовать член, содержащий $1^2 \cdot 5^2$, ибо в нем и только в нем содержится сомножитель 1^2, т. е. выпадение единицы дважды (и двойное же выпадение иного исхода). Точнее, если еще учесть общий сомножитель перед скобкой, эта вероятность равна $6 \cdot (1/6)^2 \cdot (5/6)^2 = 25/6^3 = 25/216$. Мы таким образом учли число возможных расположений этих удачных бросков, т. е. число сочетаний из четырех по два, равное шести. Без этой шестерки произведение остальных двух сомножителей соответствовало бы искомой вероятности при некотором *заданном* расположении этих бросков (например, при выпадении единицы в первом и третьем бросках).

За параметры биномиального распределения (2.3) можно принять число испытаний n и отношение p/q; считать параметрами и p, и q нет смысла, потому что одно из них определяется через другое: $p + q = 1$. Как понятно из примера, каждый член биномиального ряда (2.3) вида $C_n^k p^{n-k} q^k$ равен вероятности того, что интересующее нас случайное событие произойдет k раз в n испытаниях (безразлично в каких именно). Существенна, правда, и частота появлений события в данной серии испытаний, см. ниже и п. 2.4.1.

Биномиальное распределение дискретно, и вероятности значений соответствующей случайной величины равны

$$p(x) = C_n^k p^{n-k} q^k, x = 0, 1, 2, \dots, n.$$

Относятся эти вероятности, однако, не к единичному испытанию, а к частоте появлений события в серии из n испытаний в целом.

Гораздо более интересные примеры биномиального распределения появились, например, при изучении соотношения рождаемости девочек и мальчиков (ср. п. 4.2).

Обобщением биномиального распределения является полиномиальное, когда каждое испытание относится уже не к событию, а к случайной величине, которая может принимать несколько значений. Соответственно, оно описывается полиномом

$$(a + b + c + \dots)^n.$$

Качественные ссылки на это распределение без упоминания вероятностей можно найти у Маймонида (Рабинович 1973, с. 74) и Мопертюи (1745/1756, т. 2, с. 120 – 121 и 109). Маймонид убеждал, что “среди случайных вещей некоторые весьма правдоподобны, другие возможности [возможности других вещей?] весьма маловероятны, а еще некоторые промежуточны". Мопертюи отметил, что ребенок обычно похож на родителей, но изредка – на дальнего предка.

2.2.4. Нормальное распределение

Начнем издалека. Примерно в 1874 г. Гальтон (1877) изобрел прибор, наглядно показывающий это распределение. Дробь высыпается из воронки в сосуд и падает внутри него на ряд горизонтально расположенных и параллельных друг другу стержней. Каждая дробинка отклоняется при этом либо влево, либо вправо и продолжает падать. Таких рядов стержней около 20, и столько же раз каждая дробинка снова отклоняется то туда, то сюда. Наконец, дробинки оседают на дне сосуда, образуя кучку, и можно вообразить, что она имеет форму, подобную показанным на Рис. 2.1а.

Положение каждой дробинки определилось как бы 20-кратным равномерным распределением с возможными значениями уклонений в – 1 и 1, а форма кучки дробинок соответствует плотности того распределения, которое называется нормальным. Нормальное распределение появилось из исходных равномерных в качестве практически предельного суммирования. Соответствующая *центральная предельная теорема* (ЦПТ) утверждает, притом не только для равномерного распределения, что *сумма большого числа независимых случайных величин, каждая из которых лишь незначительно влияет на свою сумму, распределена нормально*. Сам термин ЦПТ впервые употребил Полиа (1920).

В простейшем случае ее доказательство восходит к Муавру, который (п. 5.2) рассматривал появление нормального распределения из биномиального. В дальнейшем ее неоднократно доказывали в общем случае, но строгое обоснование впервые предложили Марков и Ляпунов; вполне строгое доказательство не удалось даже их учителю, Чебышеву.

Вот пример замены биномиального распределения нормальным. Требуется подсчитать вероятность семикратного появления единицы при 100 бросках кости. Она равна $C_{100}^{93}\, p^7 q^{93}$ при $p = 1/6$, см. п. 3.4.2. Мы подсчитали ее, пользуясь нормальным распределением, поскольку практически 100 – это почти предельный случай $n \to \infty$. Если вероятности рождения мальчика и девочки обозначить через p и q (т. е. пренебречь всеми остальными редкими рождениями), аналогичный вопрос можно задать относительно вероятности того или иного числа мужских рождений. Именно эти вероятности и рассматривал Муавр при выводе нормального распределения, а точнее – вероятности, что число этих рождений заключено между некоторыми заданными пределами. С 1712 г. параметр p/q в подобных задачах, или, иначе, соотношение мужских рождений к женским, оказался (оказалось) предметом изучения статистиков и математиков (п. 5.2).

После столь длинного предисловия пора сообщить, что плотностью нормального распределения называется функция

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{(x-a)^2}{2\sigma^2}],\ -\infty < x < \infty, \qquad (2.4)$$

обладающая двумя параметрами, a и σ > 0, вероятностный смысл которых мы рассмотрим в п. 3.4.2. Не менее важна и соответствующая функция распределения

$$F(z) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{z} \exp[-\frac{(x-a)^2}{2\sigma^2}]dx.$$

Если принять $a = 0$ и σ = 1, – если *стандартизировать* ее, – то

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp[-\frac{x^2}{2}]dx. \qquad (2.5)$$

Удобнее, впрочем, табулировать для различных значений z функцию

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{0}^{z} \exp[-\frac{x^2}{2}]dx. \qquad (2.6)$$

Действительно, подынтегральная функция в формуле (2.5) четная, поэтому интеграл (2.5) в пределах (– ∞; 0] равен тому же интегралу в пределах [0; + ∞), т. е. 1/2, а в пределах, скажем, (– ∞; – 1] – разности между 1/2 и интегралом (2.6) при $z = 1$.

Функция (2.6) равна 0.499 уже при $z \approx 3$, а удвоенный интеграл при $z \to + \infty$ (или, что то же самое, тот же интеграл в бесконечных пределах) равна, как и должно быть, единице.

Термин *нормальное распределение* (окончательно) ввел Пирсон в 1893 г., чтобы не думать, указывать ли его по имени Гаусса (1809 г.) или Лапласа, который широко применял его, поскольку (нестрого) доказал несколько вариантов ЦПТ. До Пирсона этот термин употреблял, например, Гальтон, а впервые его предложил Пирс (1873, с. 206).

2.2.5. Распределение Пуассона

Это – дискретное распределение (Poisson 1837, p. 205), закон которого можно записать в виде

$$P(x) = \frac{a^x}{x!}e^{-a}, \ x = 0, 1, 2, \ldots$$

Можно проверить, что сумма вероятностей $P(x)$ по всему бесконечному множеству значений x равна 1, как и должно быть. Действительно, суммируются только дроби, а e^{-a} является общим множителем:

$$\sum_{x=0}^{\infty} \frac{a^x}{x!}e^{-a} = e^{-a}(1 + \frac{a}{1!} + \frac{a^2}{2!} + \frac{a^3}{3!} + \ldots)$$

и сумма бесконечного ряда равна e^a. Умножая ее на общий множитель, получим 1, ч. т. д.

Весьма интересная схема, приводящая к распределению Пуассона, такова: точки вводятся на отрезок в соответствии с равномерным распределением, по отдельности и независимо одна от другой. Число точек, попавших на некоторый интервал отрезка, как оказывается, подчиняется распределению Пуассона.

Примером этой схемы служит поступление вызовов на телефонную станцию, и поэтому загрузка станции может исследоваться вероятностными методами.

Пусть, например, телефонная подстанция обслуживает 300 абонентов и вероятность разговора абонента в течение часа равна $p = 0.01$. Какова будет тогда вероятность четырех или более (независимых друг от друга) звонков в течение часа?

Предпосылки применения распределения Пуассона выполнены, и мы имеем $a = pn = 3$. Далее,

$$P(\xi \geq 4) = \sum_{x=0}^{\infty} \frac{a^x}{x!} e^{-a} - P(\xi = 0) - P(\xi = 1) - P(\xi = 2) - P(\xi = 3).$$

Сумма равна единице (см. выше), а остальные слагаемые можно легко вычислить.

Другой пример: распределение звезд на небесной сфере (Мичелл, 1767): если звезды распределены равномерно (не на отрезке, а на сфере), то некоторые из них окажутся весьма близкими друг к другу (двойные, тройные, … звезды).Уже в то время было известно немалое количество таких близких звезд, и возник вопрос: случайно ли это или нет? Какова же вероятность, спрашивал Мичелл, что две звезды из общего их числа находятся на расстоянии не более 1° друг от друга? Ньюком (1860, с. 427 – 429) применил распределение Пуассона для вычисления вероятности того, что некоторая малая часть небесной сферы будет содержать s звезд из их общего числа n, равномерно распределенных на небесной сфере.

В известном смысле именно распределение Пуассона лучше всего характеризует *случайное* расположение множества точек. Его параметром, естественно, является a.

В 1898 г. Борткевич ввел в обиход свой *закон малых чисел*, который в течение нескольких десятилетий считался чуть ли не основным законом статистики. На самом деле единственной новинкой была популяризация еще малоизвестного в то время распределения Пуассона, и в настоящее время закон Борткевича заслуженно забыт, см. Шейнин (2008).

2.2.6. Гипергеометрическое распределение

Оно важно для статистического контроля качества массовой продукции, что выяснится ниже. Рассмотрим Дополнительную задачу № 4 Гюйгенса (1657). Дано 12 жетонов, 4 из них белые (как бы бракованные). Извлечено без возврата 7 жетонов; какова вероятность, что 3 из них окажутся белыми (бракованными)?

Ясно, что все 12 жетонов следовало бы сразу выбросить вон, и всё-таки решим эту задачу вслед за Якобом Бернулли (1713, часть 3-я, задача № 6), хотя уже с применением гипергеометрического распределения. Обозначим исходные данные следующим образом: $N = 12$, $M = 4$, $n = 7$, $m = 3$. Простые комбинаторные соображения приводят к формуле, которая и называется формулой указанного дискретного распределения:

$$P(\xi = m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}.$$

2.3. Основные характеристики распределений

2.3.1. Математическое ожидание

Для дискретной случайной величины оно равно сумме произведений всех ее значений на их вероятности. Пусть случайная величина ξ принимает n значений $x_1, x_2, \ldots, x_n$ с вероятностями $p_1, p_2, \ldots, p_n$. Тогда ее математическое ожидание равно

$$\mathrm{E}\xi = \frac{p_1x_1 + p_2x_2 + \ldots + p_nx_n}{p_1 + p_2 + \ldots + p_n} \tag{2.7}$$

(знаменатель, конечно же, равен единице).

Буква E соответствует французскому ésperance и английскому expectation, а прилагательное *математическое* ввел Лаплас (1812/1886, с. 189), чтобы отличить прежнее *ожидание* от ставшим в его время модным, но давно уже забытого морального ожидания (см. ниже). Пора бы забыть и о предложении Лапласа, которого до сих пор придерживаются во французской и русской литературе, и мы явочным порядком будем обходиться без него. Это сократит письмо и позволит сказать *ожидаемое значение* случайной величины.

Ожидание можно считать естественным мерилом случайной величины, как бы ее средним значением, а в теории ошибок ему соответствует обобщенное среднее арифметическое. Обозначим результаты наблюдений $x_1, x_2, \ldots, x_n$, а их веса (достоинства) $p_1, p_2, \ldots, p_n$. По определению обобщенное среднее арифметическое равно

$$\overline{x} = \frac{p_1x_1 + p_2x_2 + \ldots + p_nx_n}{p_1 + p_2 + \ldots + p_n}, \tag{2.8a}$$

однако сумма весов конечно же не равна единице.

Можно легко убедиться, что при равенстве всех весов друг другу

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}. \tag{2.8b}$$

Существует, однако, принципиальное отличие между ожиданием (2.7) и средним арифметическим (2.8). Ожидание – число, постоянная величина, потому что предполагается, что в его состав вошли все возможные значения случайной величины, среднее арифметическое же составлено из результатов измерений, неизбежно искаженных случайными ошибками (также и систематическими, но сейчас не до них), а потому само является случайной величиной, как бы выборочным значением неизвестного ожидания, и его погрешность полагается оценивать. *Аналогичное замечание будет относиться и к другим характеристикам случайной величины*. И в то же время среднее арифметическое принимается равным соответствующей, измеряемой во многих отраслях науки константе, см. п. 6.2.

Для непрерывной случайной величины ожиданием является интеграл

$$E\xi = \int_a^b x\varphi(x)dx. \qquad (2.9)$$

Пределами интеграла являются крайние точки области определения φ(x), т. е. плотности этой величины; если плотность определена на всей числовой оси, то пределы интегрирования бесконечны: $a = -\infty$, $b = \infty$.

Ожидание начало применяться в теории вероятностей до вероятности. Впервые, вполне возможно основанное на интуитивных и субъективных шансах, притом на житейском уровне, оно появилось у Маймонида (Рабинович 1973, с. 164): брачный контракт, обеспечивающий вдову или разведенную жену, в 1000 зуз [денежных единиц] "можно продать за 100, но контракт в 100 зуз – меньше, чем за 10". Большие (но не более благоприятные) возможные выигрыши считались предпочтительнее, и та же субъективная склонность существует и сегодня (и нещадно используется устроителями лотерей).

Аналогичные идеи, также не вполне определенные, возникли в Европе на несколько столетий позже в связи со страхованием жизни (Шейнин 1977, с. 206 – 209).

Теория вероятностей родилась в 1654 г., в переписке Паскаля и Ферма по поводу азартных игр (Pascal 1654/1998), в которой появились первые вероятностные вычисления, основанные на новых представлениях. Вот одна из основных задач, которую они решили независимо друг от друга, фактически исходя из понятия ожидания:

Игроки A и B договариваются играть до тех пор, пока кто-нибудь из них не выиграет пяти партий (не обязательно подряд) и не заберет все поставленные на кон деньги. По какой-то причине им, однако, пришлось прерваться при счете, скажем, 4:3 в пользу A. Как им разделить ставки, каждая из которых равнялась a?

Эта задача уже в то время имела почтенный возраст, и имеются указания, что один математик решил ее по меньшей мере в одном частном случае. Но заметим сразу же, что раздел ставки в том же отношении 4:3 был бы справедлив при игре, например, в шахматы, при которой решающее значение имеет мастерство игроков. В азартных играх, напротив, всё зависит от случая, а потому прошлое не должно влиять на раздел.

Вот решение. Игрок A имеет вероятность $p_1 = 1/2$ (эти ученые применяли шансы, а не вероятности) выиграть следующую партию и победить, т. е. получить $2a$. Но с той же вероятностью $p_2 = 1/2$ он может проиграть эту партию, и счет тогда сравняется, так что общую ставку надо будет поделить пополам. Его ожидание поэтому будет равно

$$E\xi = (1/2)\cdot 2a + (1/2)\cdot a = 3a/2.$$

Ожидание второго игрока должно быть поэтому равно $a/2$, но его можно было бы и подсчитать независимо.

Внимание Паскаля к азартным играм обратил светский человек, Де Мере (Паскаль 1654/1998, конец Письма 29 июля 1654 г.), который не понял, почему выпадение шестерки в четырех бросках кости не равновероятно выпадению двух шестерок в 24 бросках двух костей. Равенство вероятностей этих событий он обосновывал старинной приближенной формулой; вот, однако, эти вероятности:

$$P_1 = 1 - (5/6)^4 = 0.518, \; P_2 = 1 - (35/36)^{24} = 0.492.$$

Де Мере, стало быть, знал, что игроки подметили разность вероятностей, равную 0.026. Ср. аналогичное замечание Галилея в п.1.1.1.

Уже в 1657 г. Гюйгенс опубликовал трактат о расчетах в азартных играх, в котором ввел ожидание формально, чтобы обосновать и раздел ставки, и решение других задач. Целесообразность ожидания в качестве мерила случайной величины (случайного выигрыша) он пояснял разумными, но в некоторой степени сложными экономическими соображениями.

Гораздо проще ее пояснил Якоб Бернулли (1713, ч. 1), в своем классическом, правда, совсем по другой причине, сочинении, см. п. 4.1.1. Он начал изложение с перепечатки комментированного трактата Гюйгенса и вот его соображение:

Предложение 3 (Гюйгенса). Имея p шансов получить a и q шансов получить b и полагая все эти шансы равными друг другу, я получаю

$$\frac{pa+qb}{p+q}. \qquad (2.10)$$

Раз p и q не вероятности, а шансы, то их сумма не равна единице, как в формуле (2.7).

И вот комментарий Бернулли. Пусть число игроков ($p + q$), и в каждой из p коробок находится сумма a, а в каждой из q коробок – сумма b. Каждый игрок забирает по коробке, и все вместе они получают ($pa + qb$). Но раз они находятся в равном положении, то все должны получить поровну, т. е. по сумме (2.10).

Но не может математическая теория основываться на коробках и игроках, и уже Муавр ввел ожидание без всякого обоснования, аксиоматически. Так оно понимается и сейчас, хотя Лаплас (1814/1999, с. 837 правый столбец) всё-таки без особого пояснения назвал его “единственной справедливой” мерой.

Практика нескольких веков подтвердила значимость ожидания, хотя в том же 1713 г. Николай Бернулли (письмо Монмору, см. Монмор 1713, с. 402) придумал азартную игру, к которой оно совсем не подходило. Игрок А подбрасывает кость … Впрочем, очень скоро кость заменили монетой и тем самым упростили игру. Итак, А подбрасывает монету. Если орел выпадет сразу, В уплачивает ему 1 экю; если это произойдет только при втором броске, то он получит 2 экю, если только при третьем, то 4 экю и т. д. Спрашивается, сколько А должен будет заранее уплатить игроку В, чтобы тот согласился с указанными условиями?

Игрок А получит 1 экю с вероятностью 1/2, 2 экю – с вероятностью 1/4, 4 экю – с вероятностью 1/8 и т. д., и ожидание его выигрыша будет равно

$$1\cdot(1/2) + 2\cdot(1/4) + 4\cdot(1/8) + \ldots = (1/2) + (1/2) + (1/2) + \ldots = \infty. \qquad (2.11)$$

Но кто же согласится уплатить вперед хоть сколько-нибудь значительную сумму в обмен на будто бы бесконечное благо? Следует просто решить, что орел должен будет впервые выпасть если не до пятого, то до шестого или седьмого броска, и что поэтому уплатить вперед следует не более, чем пять – семь раз по пол-экю. Иначе говоря, все последующие члены ряда (2.11) надо заменить нулями.

Бюффон (1777, § 18) сообщил о серии из 2048 таких игр, в которой средний выигрыш А составил 4.9 единиц, а максимальная продолжительность игры оказалась равной девяти броскам, притом лишь в шести случаях. Таково было первое статистическое исследование азартных игр. Аналогичное, но гораздо более обширное и компьютерное исследование проделал Дутка (1988).

Теоретически этот парадокс продолжал интересовать математиков вплоть до последних лет. Быть может самым интересным оказалось замечание Кондорсе (1784, с. 714) о том, что одна игра, пусть даже теоретически бесконечная, всё же является одним-единственным испытанием, и для включения ее в теоретико-вероятностную схему следует рассматривать серию таких игр.

Именно это предложение без ссылки на своего предшественника повторил Фрейденталь (1951), который также ввел распределение ролей игроков в каждой игре по жребию.

Аналогичное соображение о пренебрежении низкими вероятностями можно высказать по поводу всех азартных игр. Если, скажем, в лотерее разыгрываются и весьма крупные выигрыши, но вероятность их получения очень низка (о чем устроители лотерей непременно позаботятся), то о них лучше всего просто забыть, точно так же, как и об отбрасываемых членах ряда (2.11), хотя бы их ожидание и не было пренебрегаемо. И вообще лотереи как правило настолько невыгодны, что уже Петти (1662/1899, с. 64) прямо заявил, что они являлись налогом на несчастливых самонадеянных дураков.

И всё-таки: какой же вероятностью следует пренебрегать? Бюффон (§ 8) тогда же заявил, что ничтожными являются вероятности ниже 1/10 000. Такова в соответствии с составленной им таблицей смертности была вероятность здоровому человеку в возрасте 56 лет умереть в течение ближайших 24 часов. Подсчитаем, сколько членов в ряду (2.11) следует принять во внимание в соответствии с этой рекомендацией. Имеем

$$1/2^n = 1/10\,000,\ 2^n = 10\,000,\ n\lg 2 = 4 \text{ и } n \approx 13.3.$$

Тринадцать или четырнадцать членов, хотя тот же Бюффон (см. выше, § 18) установил, что ни в одном из 2048 случаев не потребовалось бы учитывать более девяти. Это, кстати, означает, что иногда можно пренебрегать и более высокими вероятностями.

Но в принципе нельзя устанавливать единую для всех случаев меру ничтожности вероятности. Падение крупного астероида с катастрофическими последствиями может иметь ничтожную вероятность, но учитывать ее надо (не совсем, правда, ясно, как именно).

Читатель! А какой вероятностью несчастного случая ты пренебрегаешь, когда переходишь улицу? А когда надеешься на авось?

2.3.1-1. Свойства ожидания. 1) Пусть ξ – постоянная величина, равная c. Тогда

$$\mathrm{E}c = \int_a^b c\varphi(x)dx = c\int_a^b \varphi(x)dx = c,$$

т. е. *ожидание постоянной равно ей самой*.

2) Пусть задана случайная величина $a\xi$. Ее ожидание равно

$$\mathrm{E}a\xi = a\int_a^b x\varphi(x)dx = a\mathrm{E}\xi.$$

При умножении случайной величины на постоянный множитель ее ожидание умножается на тот же множитель.

3) Пусть отыскивается ожидание суммы двух (в принципе – любого числа) случайных величин, $\xi + \eta$ и плотность второй величины равна $\psi(y)$. Искомое ожидание равно двойному интегралу

$$E(\xi + \eta) = \int_a^b \int_c^d (x + y)\varphi(x)\psi(y)dxdy.$$

Здесь c и d – крайние точки области определения второй плотности. Этот двойной интеграл легко разбивается в сумму двух таких же интегралов:

$$E(\xi + \eta) = \int_a^b \int_c^d x\varphi(x)\psi(y)dydx + \int_a^b \int_c^d y\varphi(x)\psi(y)dxdy.$$

Первый интеграл можно представить в виде

$$\int_c^d \psi(y)dy \int_a^b x\varphi(x)dx = E\xi,$$

поскольку интеграл по y равен единице. Точно таким же образом можно показать, что второй интеграл равен $E\eta$ и поэтому

$$E(\xi + \eta) = E\xi + E\eta.$$

Ожидание суммы случайных величин равно сумме ожиданий слагаемых.

Таким же образом ожидание разности случайных величин равно разности их ожиданий. Но следует добавить, что в подобных теоремах (и не только в теории вероятностей) разность не упоминается, потому что операция вычитания по своему определению равнозначна операции сложения противоположных величин; так, $a - c \equiv a + (-c)$.

4) Без доказательства заметим, что *ожидание произведения двух <u>независимых</u> случайных величин равно произведению их ожиданий:*

$$E(\xi\eta) = E\xi \cdot E\eta.$$

Это свойство сразу же обобщается на большее число случайных величин.

Все эти следствия имеют место и для ожиданий дискретных случайных величин.

2.3.2. Дисперсия

Второе основное понятие, характеризующее распределения случайных величин, описывает их рассеивание. Вот пример: на спичечной коробке надпись “приблизительно 50 спичек”. Если в одной коробке 30 спичек, а в другой – 70, то в среднем 50, но не слишком ли велик разброс? И что значит *приблизительно*?

Пусть известны лишь некоторые значения $x_1, x_2, \ldots, x_n$ случайной величины ξ (известна выборка объемом n). Тогда выборочная дисперсия величины ξ будет равна

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}. \qquad (2.12)$$

Ее также называют эмпирической, поскольку значения x_i известны из какого-то опыта, испытания.

Почему за меру рассеивания принимается именно функция (2.12), и почему ее знаменатель $(n - 1)$, а не n? Постараемся объяснить, но вначале укажем, что просто дисперсией (не выборочной) той же величины ξ, дискретной и непрерывной, называются соответственно величины

$$\sum_{i=1}^{n} p_i (x_i - \mathrm{E}\xi)^2, \ \mathrm{D}\xi = {\sigma_\xi}^2 = \int_a^b (x - \mathrm{E}\xi)^2 \varphi(x) dx, \qquad (2.13)$$

где смысл a, b и $\varphi(x)$ тот же, что в формуле (2.9).

Дисперсию в качестве меры рассеивания наблюдений ввел Гаусс. Он указал, что ее выбор более или менее произволен, однако, во-первых, эта мера должна была быть особо чувствительна к крупным ошибкам, т. е. содержать $(x - \mathrm{E}\xi)$ в некоторой натуральной, но не в первой степени. Во-вторых, эта мера должна была оставаться положительной, а потому степень упомянутой разности должна быть четной. И, наконец, мера должна была быть по возможности проста, так что показатель степени следовало полагать равным двум. Практически, однако, Гауссу (1823b, § 37 – 38) пришлось определять только выборочную дисперсию и вместо Eξ применять среднее арифметическое. К преимуществам дисперсии мы еще вернемся.

Пусть величины x_i являются ошибками наблюдения, тогда выборочная дисперсия выражается функцией

$$\frac{\sum_{i=1}^{n} {x_i}^2}{n}.$$

Но эти ошибки не известны, их приходится заменять уклонениями наблюдений от среднего арифметического, а тогда, как доказал Гаусс в указанном месте, выборочная дисперсия выражается формулой (2.12). Он (1823a/1957, с. 146) заметил, что замена знаменателя на $(n - 1)$ “отчасти” требовалась и “достоинством науки”.

Если серия наблюдений искажена примерно одной и той же систематической ошибкой, эта формула не отразит их и может сильно приукрасить действительность: разброс окажется небольшим, хоть наблюдения и уклоняются от измеряемой константы. Сам Гаусс, который несколько лет непосредственно проводил геодезические измерения, не доверял своей формуле (безусловно ввиду неизбежных систематических ошибок) и измерял каждый угол до тех пор, пока не убеждался в ненужности дальнейших усилий; выписки из его полевых журналов опубликованы в т. 9 его *Трудов*, с. 278 – 281.

Практически широко используется не только выборочная дисперсия, но и корень квадратный из нее, – не только s^2, но и s. Величина s называется *стандартным отклонением*; впрочем, в теории ошибок принят термин *средняя квадратическая ошибка* (наблюдения).

Теперь мы можем уточнить утверждения типа “в коробке примерно 50 спичек”. Займемся неблагодарным трудом и подсчитаем число спичек $x_1, x_2, \ldots, x_{10}$ в 10 коробках, вычислим их среднее (т. е. выборочное среднее, поскольку коробок существует несметное количество) $\bar{x}$ и уклонения от него $(x_1 - \bar{x})$, $(x_2 - \bar{x})$, …, $(x_{10} - \bar{x})$, а затем и выборочную дисперсию (или стандартное отклонение). Ее можно сравнить с гарантийным значением, которое в более серьезных случаях следовало бы указывать. Уклонение, превышающее две средние квадратические ошибки, уже можно считать значительным.

Мы видели, что ожидание случайной величины может оказаться бесконечным, а сейчас добавим: то же самое имеет место для дисперсии. Вот пример: непрерывная случайная величина распределена по закону Коши

$$\varphi(x) = \frac{2}{\pi(1+x^2)},\ 0 \le x < \infty. \tag{2.14}$$

Заметим, что равенств типа $x = \infty$ следует избегать, потому что бесконечность – это не число, а переменная величина. Кроме того, распределение (2.14) встречалось у Пуассона (1824, с. 278) раньше, чем у Коши.

Так вот, в этом случае дисперсия равна

$$\mathrm{D}\xi = \frac{2}{\pi}\int_0^\infty x^2 \cdot \frac{1}{1+x^2}\,dx.$$

Представим подынтегральную функцию в виде

$$\frac{(x^2+1)-1}{1+x^2} = 1 - \frac{1}{1+x^2}$$

и тогда

$$\mathrm{D}\xi = \frac{2}{\pi}\left[\int_0^\infty 1 \cdot dx - \int_0^\infty \frac{1}{1+x^2}\,dx\right].$$

Второй интеграл равен

$$\operatorname{arctg} x \Big]_0^\infty = \frac{\pi}{2},$$

а вот первого не существует:

$$\int_0^\infty dx = x\Big]_0^\infty \to \infty.$$

Среднее арифметическое из наблюдений, если они настолько негодные, что их ошибки подчиняются распределению Коши, будет поэтому не лучше, чем отдельное наблюдение. По формуле (2.16) п. 2.3.2-1 оказывается, что дисперсия среднего из n наблюдений в n раз меньше дисперсии одного наблюдения, т. е. в n раз меньше бесконечности, а потому также бесконечна.

2.3.2-1. Второе определение дисперсии. Выпишем определение (2.13b) в развернутом виде:

$$\mathrm{D}\xi = \int_a^b x^2\varphi(x)dx - 2\int_a^b x\mathrm{E}\xi\varphi(x)dx + \int_a^b (\mathrm{E}\xi)^2\varphi(x)dx$$

и вспомним, что Eξ не переменная, а постоянная величина, которую можно вынести за знак интеграла:

$$\mathrm{D}\xi = \int_a^b x^2\varphi(x)dx - 2\mathrm{E}\xi\int_a^b x\varphi(x)dx + (\mathrm{E}\xi)^2\int_a^b \varphi(x)dx.$$

Первый интеграл по определению равен $\mathrm{E}\xi^2$, второй по той же причине равен Eξ, а интеграл от $\varphi(x)dx$ равен единице по свойству плотности распределения. Таким образом

$$\mathrm{D}\xi = \mathrm{E}\xi^2 - 2(\mathrm{E}\xi)^2 + (\mathrm{E}\xi)^2 = \mathrm{E}\xi^2 - (\mathrm{E}\xi)^2. \qquad (2.15)$$

Формула (2.15) обычно и принимается за основное определение дисперсии. Возможно, однако, что наше описание более доходчиво.

2.3.2-2. Свойства дисперсии

1) Дисперсия суммы независимых случайных величин. В соответствии со вторым определением дисперсии сразу напишем:

$$\mathrm{D}(\xi + \eta) = \mathrm{E}(\xi + \eta)^2 - [\mathrm{E}(\xi + \eta)]^2 =$$
$$\mathrm{E}\xi^2 + 2\mathrm{E}(\xi\eta) + \mathrm{E}\eta^2 - [(\mathrm{E}\xi)^2 + 2\mathrm{E}\xi\mathrm{E}\eta + \mathrm{E}\eta^2].$$

По четвертому свойству ожидания для независимых случайных величин (см. выше п. 2.3.1-1)

$$\mathrm{E}(\xi\eta) = \mathrm{E}\xi\cdot\mathrm{E}\eta$$

и таким образом

$$\mathrm{D}(\xi + \eta) = [\mathrm{E}\xi^2 - (\mathrm{E}\xi)^2] + [\mathrm{E}\eta^2 - (\mathrm{E}\eta)^2] = \mathrm{D}\xi + \mathrm{D}\eta.$$

Дисперсия суммы независимых случайных величин равна сумме их дисперсий.

2) Следствие (дисперсия среднего арифметического)

Нам даны наблюдения $x_1, x_2, \ldots, x_n$ и вычислено их среднее арифметическое (2.8b)

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

Формула (2.12) устанавливает выборочную дисперсию наблюдения x_i, но какова будет дисперсия среднего? Используя теоремы о дисперсии суммы случайных величин (а результаты наблюдения случайны!) и произведения случайной величины на постоянный сомножитель (который здесь равен $1/n$), сразу получим простую, но важную формулу

$$\mathrm{D}\,\overline{x} = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n(n-1)}. \qquad (2.16)$$

Дисперсия среднего арифметического из n наблюдений в n раз меньше, чем дисперсия каждого из них. Здесь, как и в формуле (2.12), конечно же подразумевается, что наблюдения являются возможными значениями одной и той же случайной величины.

3) Дисперсия линейной функции случайной величины

Пусть $\eta = a + b\xi$ – линейная функция случайной величины ξ (и потому также случайна как всякая функция, зависящая от случайной величины). Дисперсия ξ известна и равна $\mathrm{D}\xi$ и требуется определить $\mathrm{D}\eta$. Такова задача, которая встречается достаточно часто. Имеем по формуле (2.15)

$$\mathrm{D}\eta = \mathrm{E}\eta^2 - (\mathrm{E}\eta)^2 = \mathrm{E}(a + b\xi)^2 - [\mathrm{E}(a + b\xi)]^2.$$

Первое слагаемое равно

$$\mathrm{E}(a^2 + 2ab\xi + b^2\xi^2) = a^2 + 2ab\mathrm{E}\xi + b^2\mathrm{E}\xi^2.$$

Второе слагаемое

$$[\mathrm{E}a + \mathrm{E}(b\xi)]^2 = (\mathrm{E}a)^2 + 2\mathrm{E}a\mathrm{E}(b\xi) + (\mathrm{E}b\xi)^2 = a^2 + 2ab\mathrm{E}\xi + b^2(\mathrm{E}\xi)^2$$

и разность слагаемых равна $b^2[\mathrm{E}\xi^2 - (\mathrm{E}\xi)^2]$.
В соответствии с формулой (2.15) оказывается, что $\mathrm{D}\eta = b^2\mathrm{D}\xi$.

Итак, добавление постоянной величины к случайной не изменяет дисперсии, а умножение случайной величины на постоянный коэффициент умножает ее дисперсию на квадрат постоянной:

$$\mathrm{D}(a + \xi) = \mathrm{D}\xi, \ \mathrm{D}(b\xi) = b^2\mathrm{D}\xi.$$

2.4. Параметры некоторых распределений

Параметры нескольких распределений мы установили в п. 2.2, но остались биномиальное и нормальное распределения.

2.4.1. Биномиальное распределение

Пусть μ_k – случайное число наступления события в k-м испытании, равное 0 или 1. Если вероятность этого наступления равна p, то

$$E\mu_k = 1 \cdot p + 0 \cdot q = p.$$

В серии из n испытаний событие происходит

$$(\mu_1 + \mu_2 + \ldots + \mu_n) = \mu \text{ раз}, \ E\mu = E\mu_1 + E\mu_2 + \ldots + E\mu_n = pn.$$

Далее, см. формулу (2.15),

$$D\mu_k = E\mu_k^2 - (E\mu_k)^2.$$

Но μ_k^2 принимает те же значения 0 и 1, что и μ_k, притом с теми же вероятностями p и q, так что

$$D\mu_k = p - p^2 = p(1 - p) = pq, \ D\mu = D\mu_1 + D\mu_2 + \ldots + D\mu_n = pqn.$$

Величины $E\mu$ и $D\mu$ характеризуют частоту μ. Напомним, что параметры самого биномиального распределения мы рассматривали в п. 2.2.3.

2.4.2. Нормальное распределение

Геометрический смысл параметра σ понятен из Рис. 2.1a, из которого усматривается, что от значения σ зависит форма кривой. Чем σ меньше, тем больше площадь под кривой сосредоточивается в ее центральной части. Те значения случайной величины, которые близки к абсциссе вершины кривой, становятся более вероятными, случайная величина как бы съеживается.

При $a = 0$ график плотности нормального распределения окажется симметричным относительно оси Oy (Рис. 2.1b), так что a – *параметр сдвига* плотности. Заметим, что этот термин относится к любым плотностям, формула которых содержит разность $x - a$.

Аналитический смысл обоих параметров весьма прост:

$$a = E\xi, \ \sigma = D\xi. \qquad (2.17a, 2.17b)$$

Докажем (2.17a) и наметим доказательство (2.17b). Имеем

$$E\xi = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp[-\frac{(x-a)^2}{2\sigma^2}]dx.$$

Сомножитель x в подынтегральной функции равен $[(x - a) + a]$ и интеграл можно поэтому записать в виде

$$\frac{1}{\sigma\sqrt{2\pi}}\{\int_{-\infty}^{\infty}(x-a)\exp[-\frac{(x-a)^2}{2\sigma^2}]dx + a\int_{-\infty}^{\infty}\exp[-\frac{(x-a)^2}{2\sigma^2}dx\}.$$

Подынтегральная функция в первом интеграле – нечетная функция аргумента ($x - a$), который, как и x, изменяется в пределах ($-\infty$; ∞) ибо замена переменной $(x - a) = z$ не изменит пределов интегрирования. Это означает, что первый интеграл равен нулю ("отрицательная" площадь под осью Ох, расположенная левее Оу, равна положительной площади над осью Ох, расположенной правее Оу).

Во втором интеграле положим

$$\frac{x-a}{\sigma\sqrt{2}} = z,\ dx = \sigma\sqrt{2}dz, \qquad (2.18)$$

так что этот интеграл без сомножителя a равен

$$\int_{-\infty}^{\infty} \exp(-z^2)dz \cdot \sigma\sqrt{2}.$$

Интеграл (без сомножителя $\sigma\sqrt{2}$) в свое время вычислил Эйлер, он равен $\sqrt{\pi}$. Окончательно, с учетом всех трех сомножителей a, $\sigma\sqrt{2}$ и $1/\sigma\sqrt{2\pi}$, второй интеграл равен a и формула доказана.

Теперь формула (2.17b):

$$D\xi = E(\xi - E\xi)^2 = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-a)^2 \exp[-\frac{(x-a)^2}{2\sigma^2}]dx.$$

В правой части мы уже имели право применить только что выведенную формулу (2.17a). Теперь следует перейти к новой переменной, см. (2.18), и интегрировать по частям.

2.5. Другие характеристики распределений

2.5.1. Характеристики, заменяющие ожидание

В своем выборочном варианте (иногда: в своем единственном выборочном варианте) эти характеристики заменяют и среднее арифметическое или как-то иначе оценивают положение измеряемой константы.

2.5.1-1. Медиана. Расположим наблюдения $x_1, x_2, \ldots, x_n$ случайной величины по порядку, т. е. будем считать, что $x_1 \le x_2 \le \ldots \le x_n$. Заметим, что полученная таким образом последовательность называется *вариационным рядом.*
Их медианой называется среднее наблюдение, которое вполне определено при нечетном n. Так, при $n = 7$ медианой окажется x_4. При четном числе наблюдений за медиану принимают полусумму двух средних; например, при $n = 12$ ей окажется полусумма x_6 и x_7.

Для непрерывной случайной величины с плотностью распределения $\varphi(x)$ медианой называется точка x_0 на числовой оси, которая делит пополам площадь под кривой $\varphi(x)$:

$$\int_{a}^{x_0} \varphi(x)dx = \int_{x_0}^{b} \varphi(x)dx = 1/2.$$

Иначе, медиана соответствует равенству $F(x) = 1/2$.
Напомним: в соответствии с определением плотности вся эта площадь равна единице, а a и b – крайние точки области определения $\varphi(x)$.

При некоторых плотностях распределения, равно как и при неизвестной плотности, медиана характеризует соответствующую случайную величину надежнее, чем выборочное среднее арифметическое. Тот же вывод следует принять, если крайние наблюдения быть может грубо ошибочны: они могут сильно сместить это среднее, но они меньше влияют на медиану.

В свое время Менделеев (1877/1949, с. 156), который был также выдающимся метрологом, ошибочно считал, что при неизвестной плотности распределения ошибок наблюдения следует принимать среднее арифметическое.

Для непрерывных распределений вводятся также *квантили*, соответствующие вероятности p (или порядка p), т. е. точки $x = x_p$, для которых $F(x) = p$, так что медиана является квантилем порядка $p = 1/2$. Положение квантилей, как и медианы, может оказаться неопределенным.

2.5.1-2. Мода. Так называют точку (или точки) максимума плотности распределения (Рис. 4). Может показаться, что плотность (как на этом рисунке) имеет только одну точку максимума, притом иногда совпадающую со средним арифметическим, но бывает, что ни одно из этих свойств не выполняется. В соответствии с количеством мод плотность называют одно-, дву-, … модальной; впрочем, давно уже предпочитают говорить одно-, дву-, … вершинная. Есть даже термин *антимодальная*, как бы антивершинная плотность, – плотность, обладающая точкой минимума. По отношению к дискретным случайным величинам мода малоупотребительна.

2.5.1-3. Полусумма крайних наблюдений. Это – очень простая, но и ненадежная мера, потому что крайние наблюдения могут быть грубо ошибочными, а большинство наблюдений вообще не принимается в расчет. Ее широко применяли в XVIII в. в метеорологии для оценки среднемесячных значений метеорологических элементов, например, температуры воздуха. Действительно, вместо вывода среднего из, скажем, 30 чисел гораздо проще вычислить среднее из двух из них. Даниил Бернулли (1778, § 10) указал, правда, без всяких пояснений, что эта мера точнее, чем можно было бы предполагать.

2.5.2. Характеристики, заменяющие дисперсию

2.5.2-1. Размах варьирования. (Выборочным) размахом ряда значений случайной величины называется разность между наибольшим и наименьшим значениями. Иногда учитываются и вообще-то неравные друг другу “полуразмахи” $(x_n - \bar{x})$ и $(\bar{x} - x_1)$. Эти разности весьма ненадежны. Во-первых, крайние наблюдения (как мы уже отметили по отношению к полусумме крайних значений) могут быть грубо ошибочны. Во-вторых, они вполне могут возрастать с ростом числа наблюдений. Так, может появиться результат, меньший по величине, чем x_1 или больший, чем x_n.

Можно, конечно, принимать взамен указанных мер дроби $(x_n - x_1)/n$, $(x_n - \bar{x})/n$ и $(\bar{x} - x_1)/n$. Число наблюдений здесь совпадает со знаменателем и возрастает соответственно, про числитель же сказать то же самое нельзя наверняка, а первое отрицательное свойство прежних мер сохраняется и здесь.

Все эти меры характеризуют не отдельное наблюдение, а весь ряд наблюдений в целом.

2.5.2-2. Средняя абсолютная ошибка. Эта мера, как и вероятная ошибка (см. ниже) характеризует уже именно ошибки наблюдения. Если обозначить их $x_1, x_2, \ldots, x_n$, то средняя абсолютная ошибка будет равна

$$\frac{\sum_{i=1}^{n} | x_i |}{n}.$$

В свое время она получила некоторое распространение при обработке наблюдений, быть может потому, что в вычислительном смысле она предпочтительнее.

2.5.2-3. Вероятная ошибка. Вероятную ошибку наблюдения ввел Бессель (1816, с. 141 – 142) как мерило точности наблюдений, но идею о вероятном значении случайной величины высказал еще Гюйгенс в письме брату (1669/1895, письмо 28 ноября 1669 г.). Обсуждая случайный срок жизни человека, он указал на различие между ожидаемым сроком (т. е. средним арифметическим, вычисленным для многих лиц) и тем вероятным возрастом, *достичь или не достичь которого равно возможно*.

И тот, и другой срок жизни должен подсчитываться отдельно для мужчин и женщин, но в те времена об этом не знали. *В среднем* женщины живут добрых несколько лет дольше, что быть может вполне возмещает трудности их более тяжелой (и в биологическом, и в социальном смысле) жизни. Об этом обстоятельстве они, кстати, редко вспоминают.

Идею вероятного значения, которую с тех пор неоднократно использовали и в статистике населения, и, скажем, при исследовании малых изменений в периоде колебания маятника (Даниил Бернулли 1780), Бессель и применил. *Вероятной он назвал такую ошибку, которая с равной вероятностью может оказаться либо большей, либо меньшей, чем реально допущенная.*

Для симметричных распределений вероятная ошибка принимается численно равной длине отрезка между медианой и квантилем порядка p = 1/4 или 3/4. Чтобы это определение соответствовало предложенному Бесселем, следует считать, что она равна вероятности уклонения наблюдения в любую сторону от медианы на длину отрезка от нее до квантиля любого из указанных порядков. Для нормального распределения этот отрезок равен 0.6745σ, и многие авторы принимали соотношение 1 вероятная ошибка = 0.6745σ, молчаливо полагая, что либо имели дело с нормальным распределением, либо что указанное соотношение универсально.

Скажем больше: мы не уверены, что существует общепринятое определение, пригодное для асимметричных распределений, при которых расстояния между медианой и квантилями порядков p = 1/4 и 3/4 не совпадают. Видимо, позволительно в таких случаях считать, что вероятная ошибка (если она еще имеет смысл) равна половине длины отрезка между этими квантилями.

Идея вероятной ошибки настолько естественна, что эта мера стала всеобще признанной, а о средней квадратической ошибке почти забыли. Так продолжалось быть может до второй половины XX в. Бомфорд (1971, с. 610 – 611), автор серьезного руководства по геодезии, только в *третьем* его издании "неохотно" отказался от вероятной ошибки в пользу средней квадратической.

Так чем же последняя лучше? Можно принять во внимание, что вероятная ошибка связана с медианой, которая лишь иногда предпочтительнее среднего арифметического. Но есть и вторая причина, по которой средняя квадратическая ошибка (точнее, дисперсия) надежнее всех остальных мер. Дисперсия (можно говорить лишь о выборочной дисперсии) – случайная величина, а потому обладает своей собственной выборочной дисперсией. Аналогичное замечание, как мы выше упомянули, можно отнести к любым выборочным мерам, и мы особо указали по этому поводу среднее арифметическое.

И тут надо добавить: дисперсия дисперсии, в отличие от дисперсии остальных мер разброса, известна, ее вычислил уже Гаусс (1823b, § 40). Он, правда, допустил при этом элементарную ошибку, которую исправил Гельмерт (1904), а затем независимо – Колмогоров с соавторами (1947). С другой стороны, есть особое обстоятельство: на практике пользуются не дисперсией, а корнем из нее, т. е. стандартным отклонением (= средней квадратической ошибкой). И если дисперсия дисперсии равна некоему числу a, то это никак не означает, что дисперсия стандартного отклонения равна $\sqrt{a}$; она вообще известна лишь для частного случая нормальной случайной величины (см. ниже). Мало того: выборочная дисперсия является *несмещенной* оценкой *генеральной* дисперсии, т. е. ее ожидание равно этой дисперсии, и Гаусс (п. 2.3.2) особо подчеркивал существенное значение несмещенности (которое в настоящее время признается гораздо меньше). Но вот выборочное стандартное отклонение этим свойством не обладает: его ожидание не равно генеральному стандартному отклонению.

2.5.2-4. Неопределенное указание разброса. Иногда встречаются указания типа "Эта величина равна $a \pm c$". По смыслу это должно означать "равна любому значению между $a - c$ и $a + c$", но возможно также, что c – не предельная погрешность, а, например, вероятная. И кроме того: как именно было получено c? Здесь мы подошли к важному вопросу доверительного оценивания.

2.6. Доверительное оценивание

Обозначим произвольный параметр функции или плотности распределения через λ. Пусть по результатам наблюдения получено его выборочное значение $\hat{\lambda}$ и требуется оценить разность $|\hat{\lambda} - \lambda|$. Ее *доверительное* оценивание означает, что указываются такие значения α и δ, что

$$P(|\hat{\lambda} - \lambda| < \delta) > 1 - \alpha.$$

После этого можно будет сказать, что *доверительный интервал* $[\hat{\lambda} - \delta; \hat{\lambda} + \delta]$ покроет остающееся неизвестным значение λ с *доверительной вероятностью* (или *коэффициентом доверия*) (1 – α).

Это оценивание имеет смысл, если удается выбрать небольшое α (равное, например, 0.01 или 0.05, но уж наверное намного большее, чем бюффоново значение 1/10 000), при котором значение δ тоже окажется достаточно малым; в противном же случае доверительное оценивание покажет, что наблюдений было либо слишком мало, либо, что они не были достаточно точными. Следует также иметь в виду, что в любом случае другой набор наблюдений может привести к иным значениям $\hat{\lambda}$ и δ.

Если по результатам наблюдения устанавливается константа А, то в простейшем предположении (Берви, 1899) можно полагать, что *размах варьирования*, т. е. интервал между крайними наблюдениями $[x_1; x_n]$ “покрывает” ее с вероятностью

$$P(x_1 \le A \le x_n) = 1 - (1/2^{n-1}).$$

Недостаток этого приема тот же, что указан в п. 2.5.1-3. Подобные заключения встречались уже в древней астрономии (Шейнин 1993b, § 2.1). Именно, на основании всех соответствующих наблюдений (а не только своих собственных) выбирались какие-то границы (a и b) и утверждалось, что $a \le A \le b$. Вероятность при этом вообще не упоминалась, но по смыслу оказывалось, что сделанный вывод считался почти достоверным.

При установлении константы любую меру рассеивания можно истолковать как равнозначную доверительной. Пусть известен соответствующий закон распределения и по результатам наблюдения вычислены среднее арифметическое $\overline{x}$ и его средняя квадратическая ошибка m. Тогда вероятность $P(\overline{x} - m \le \overline{x} \le \overline{x} + m)$ можно будет найти по статистическим таблицам закона распределения как разность $P(0 \le \overline{x} \le \overline{x} + m) - P(0 \le \overline{x} \le \overline{x} - m)$; различием между строгими и нестрогими неравенствами можно пренебречь. Она и будет доверительной вероятностью, а $[\overline{x} - m; \overline{x} + m]$ – доверительным интервалом.

2.7. Моменты случайной величины

Эта тема вполне вписывается в предыдущий п. 6, но заслуживает отдельного рассмотрения. Моменты характеризуют плотность, а иногда могут устанавливать ее.

Начальным моментом порядка s дискретной или непрерывной случайной величины ξ называются соответственно

$$\alpha_s(\xi) = \sum_x x^s p(x),\ \nu_s = \int x^s \varphi(x)dx. \qquad (2.19)$$

В первом варианте суммирование распространяется на все значения x, обладающие вероятностями $p(x)$, а пределами интегрирования являются крайние точки области определения известной или неизвестной плотности распределения $\varphi(x)$ непрерывной величины.

Применяются и центральные моменты

$$\mu_s(\xi) = \sum_i (x_i - \mathrm{E}\xi)^s p(x_i),\ \mu_s(\xi) = \int (x - \mathrm{E}\xi)^s \varphi(x)dx. \qquad (2.20)$$

Эти же формулы могут быть записаны совместно в виде

$$\mu_s(\xi) = \mathrm{E}(\xi - E\xi)^s.$$

Выборочные или эмпирические начальные моменты для дискретной и непрерывной случайной величины конечно же совпадают и равны

$$m_s(\xi) = \frac{1}{n}\sum_i x_i^s, \qquad (2.21)$$

где n – число измеренных (наблюденных) значений величины ξ.

Центральные выборочные моменты будут равны

$$m_s(\xi - \bar{x}) = \frac{1}{n}\sum (x_i - \bar{x})^s.$$

Часто указанные выше наблюденные значения объединяют в интервалы. Вот пример (Смирнов и Дунин-Барковский 1959, § 1 в гл. 3). Для контроля размеров изготовленных деталей было отобрано 70 *проб* по 5 деталей в каждой. Оказалось, что в 55 пробах размеры всех пяти деталей были стандартными, а в 12 и трех пробах эти размеры были нестандартны у одной и двух деталей соответственно. Или иначе:

Частоты	55	12	3
Количества дефектных деталей	0	1	2
Частости дефектных деталей	0.786	0.171	0.043

Заметим, что если в серии n испытаний изучаемое событие произошло m раз, то m называется *частотой* событий, а m/n – его *частостью* (эмпирической вероятностью). Так, 0.786 = 55/70 и сумма всех частостей равна единице.

На протяжении более трех столетий было предложено немало определений математической статистики, но лишь одно **определение статистических данных** (Колмогоров и Прохоров 1982, с. 576). Ими называются *сведения о числе объектов в какой-либо более или менее обширной совокупности, обладающих теми или иными признаками.*

Исходные данные в приведенном примере как раз и являются статистическими данными, а интервалами, на которые разбита совокупность, естественно служили количества дефектных деталей в пробах. Часто, однако, разбивка на интервалы может быть неоднозначной, но во всяком случае, если весь промежуток значений случайной величины достаточно широк (здесь случайная величина – это количество дефектных деталей, возможно от нуля до пяти, но фактически только от нуля до двух, т. е. промежуток оказался очень узким), то интервалов не должно быть слишком мало. Но и слишком большого их числа допускать не следует: это – "роскошь в цифрах, научное шарлатанство". Таково было общее утверждение Кетле (1846, с. 278).

Так вот, при объединении в интервалы формула (2.21) переходит в

$$m_s(\xi) = \frac{1}{n}\sum_i n_{x_i} x_i^s, \qquad (2.22)$$

где n_x – число значений случайной величины в интервале x. В приведенном примере

$$m_s(\xi) = \frac{1}{70}(55 \cdot 0^s + 12 \cdot 1^s + 3 \cdot 2^s) = 0.786 \cdot 0^s + 0.171 \cdot 1^s + 0.043 \cdot 2^s.$$

Случаи $s = 1$ и 2 совпадают с ожиданием и дисперсией случайной величины и формулы (2.20) соответствуют формулам (2.13). *Первым моментом является ожидание, а вторым – дисперсия*. Но можно ли и нужно ли что-либо устанавливать про остальное, чуть ли не бесконечное количество моментов? Достаточно будет рассмотреть третий и четвертый моменты.

Допустим, что плотность распределения φ(x) симметрична относительно оси Оу. Тогда при нечетном s моменты

$$\nu_s = \int_{-\infty}^{\infty} x^s \varphi(x)dx = 0.$$

Действительно, в этом случае вся подынтегральная функция, т. е. произведение нечетной функции на четную, окажется нечетной, а интегрирование происходит в симметричных пределах.

Если же какой-нибудь нечетный момент отличен от нуля, то плотность не может быть симметричной (т. е. четной), и этот момент будет характеризовать отклонение φ(x) от симметрии. Но какой же момент выбрать в качестве меры асимметрии?

Все выборочные моменты подсчитываются по наблюденным значениям соответствующей случайной величины, сами поэтому также являются случайными величинами и стало быть обладают дисперсией. Известно к тому же, что *у старших моментов дисперсия больше, чем у первых*, т. е. что *они ненадежны*. За меру асимметрии плотности распределения φ(x) случайной величины естественно поэтому принять ее третий момент, а практически – третий выборочный момент, потому что кроме как на наблюдениях случайной величины основываться не на чем:

$$m_3(\xi) = \frac{1}{n-1}\sum (x_i - \bar{x})^3. \qquad (2.23)$$

Но это еще не всё. Третий момент имеет размерность, равную кубу размерности $(x_i - \bar{x})$ и для его перевода в безразмерную величину его следует разделить на s^3, см. формулу (2.12). Окончательно мерой асимметрии φ(x) является

$$s_k = \frac{m_3}{s^3}.$$

Поясняя формулу (2.12), мы указали причину, по которой ее знаменатель равен (n – 1), а не n, и то же относится к формуле (2.23).

Важен и четвертый момент. Для нормальной случайной величины он, как оказывается, равен $3\sigma^4$, а второй момент равен σ^2, см. п. 3.3.2. Таким образом, для этого распределения

$$\frac{\nu_4}{\sigma^4} = 3.$$

Если теперь для неизвестного распределения вычислить так называемый эксцесс (точнее, выборочный эксцесс)

$$\varepsilon_k = \frac{m^4}{s^4} - 3,$$

то его уклонение от нуля может служить мерой уклонения неизвестной плотности распределения от нормального закона. Вычисление эксцесса полезно потому, что нормальное распределение “обычно” является в том или ином смысле наилучшим. Эксцесс ввел Пирсон (1905, с. 181) в связи со своими исследованиями асимметричных распределений.

Вообще же при неизвестной плотности распределения знание первых четырех моментов существенно: приравнивая их к соответствующим теоретическим моментам этой плотности, можно будет представить ее вид, а потому вычислить ее параметры, которых вряд ли окажется более четырех. Повторим, что у нормального распределения их только два. Поэтому, если вычисленный эксцесс окажется достаточно близким к нулю, то распределение определится по первым двум моментам. Здесь, правда, возможно затруднение: что значит *достаточно близко*? Этим вопросом мы уже не будем заниматься.

2.8. Распределение функции случайной величины

Пусть случайные величины ξ и η обладают плотностями распределения $\varphi_1(x)$ и $\varphi_2(y)$ соответственно и вторая величина функционально зависит от первой: $\eta = f(\xi)$, причем функция f непрерывна и дифференцируема. Плотность $\varphi_1(x)$ известна, и требуется определить $\varphi_2(y)$. Эта задача важна, и ее часто приходится решать. Прежде, чем приступить к ней, напомним сведения об обратной функции.

Мы ограничимся случаем строго монотонной (возрастающей или убывающей) функции; область определения произвольной функции можно разделить на интервалы ее возрастания и убывания и рассматривать те и другие по отдельности. Пусть функция $y = f(x)$ строго убывает на интервале $[a; b]$, см. Рис. 5. Если повернуть рисунок так, чтобы ось Оу оказалась горизонтальной и направленной влево, то график заданной функции можно будет считать графиком обратной функции $x = \psi(y)$, также однозначной ввиду монотонности прямой функции. Вот только положительное направление оси Оу, а потому и аргумента y (да, уже y, а не x)

необычно, но и это можно будет устранить, если мысленно заглянуть на график с другой стороны плоскости чертежа.

Теперь вернемся к нашей задаче. При изменении ξ на интервале $[a; b]$ случайная точка $(\xi; \eta)$ перемещается вдоль кривой $y = f(x)$. Пусть, например, $\xi = x_0$, тогда $\eta = f(x_0) = y_0$. Можно заметить, что функция распределения (не плотность) $F(y)$ величины η, равная $P(\eta < y)$, равна

$$P(\eta < y) = P(x < \xi < b) = \int_x^b \varphi_1(x)dx = \int_x^b \varphi_1(z)dz,$$

где $(x; y)$ – текущая точка на кривой $y = f(x)$.

В появившемся интеграле мы с методической целью заменили переменную, но, конечно же, оставили без изменения нижний предел, который можно теперь выразить через y: $x = \psi(y)$.
И теперь

$$F(y) = P(\eta < y) = \int_{\psi(y)}^b \varphi_1(z)dz.$$

Продифференцируем обе части равенства по y, чтобы в левой его части перейти к плотности:

$$F'(y) = \varphi_2(y) = -\varphi_1[\psi(y)]\cdot\psi'(y).$$

Если функция $f(x)$ строго возрастает, то все рассуждения остаются теми же, но переменным пределом интеграла окажется верхний, и знак минус исчезнет. Оба случая можно объединить, записав

$$\varphi_2(y) = \varphi_1[\psi(y)]\cdot|\psi'(y)|.$$

Пример (Вентцель 1969, с. 265).

$$\eta = 1 - \xi^3,\ \varphi_1(x) = \frac{1}{\pi(1+x^2)},\ -\infty < x < \infty.$$

График функции $\varphi_1(x)$, т. е. распределения Коши (которое мы упоминали в п. 2.3.2 чуть в ином виде), на некотором интервале $[a; b]$, как можно считать, показан на том же рисунке. Имеем

$$x = \psi(y) = \sqrt[3]{1-y},$$

$$\psi'(y) = -\frac{1}{3\sqrt[3]{(1-y)^2}},\ \varphi_1[\psi(y)] = f[\sqrt[3]{1-y}] = \frac{1}{\pi[1+\sqrt[3]{(1-y)^2}]},$$

$$\varphi_2(y) = \frac{1}{\pi[1+\sqrt[3]{(1-y)^2}\,]}\,\frac{1}{3\sqrt[3]{(1-y)^2}}\,.$$

Вот такая простая функция … Можно, конечно, в обеих частях этой формулы заменить *y* на *x*.

2.9. Неравенство Бьенеме – Чебышева

Так называется неравенство

$$P(|\xi - \mathrm{E}\,\xi| < \beta) > 1 - \frac{\sigma^2}{\beta^2},\ \beta > 0 \qquad (2.24)$$

или

$$\mathrm{P}(|\xi - \mathrm{E}\,\xi| \ge \beta) < \frac{\sigma^2}{\beta^2}.$$

Второе неравенство уже очевидно, поскольку сумма левых частей обоих выражений равна единице.

Неравенство (2.24) имеет место для любых случайных величин, обладающих ожиданием и дисперсией, а потому исключительно интересно теоретически (о втором неравенстве говорить особо нет смысла). Практически, однако, его универсальность означает, что оно довольно грубо. Оно как бы объединяет эти характеристики, притом не нуждаясь ни в каких иных сведениях.

Установил указанное неравенство Бьенеме в 1853 г., но, в отличие от позднейших работ Чебышева (1867 и позже), не обратил на него особого внимания, потому что тема его мемуара не была непосредственно связана с ним.

У. Гершель(1817/1912, с. 579) решил, что звезда, выбранная “наудачу” из 14 тысяч более ярких звезд, “вряд ли будет намного отличаться по своим размерам от их среднего размера”. Он еще не знал, что размеры звезд чудовищно отличались друг от друга, так что их средний размер не имел смысла. По сути, как стало ясно позднее, они не составляют единой совокупности, ср. наше замечание о дорожных происшествиях в разные дни недели (п. 1.1.3). Вспомнив о неравенстве (2.24), можно еще добавить: дисперсия неизвестного размера звезды была также чудовищна.

Аналогичный пример доставляет нам утверждение английского врача Дж. Симпсона (1847 – 1848/1871, с. 102). Он объединил данные о смертности после ампутаций конечностей по английским больницам за 1794 – 1839 гг. и заявил, что именно это обстоятельство приводит к достоверности его выводов. Как раз напротив! Следовало отбросить более старые данные и вообще подумать, можно ли было объединять данные по различным больницам.

Глава 3. Системы случайных величин. Корреляция

3.1. Корреляция

В первом приближении можно сказать, что переменная величина *y* является функцией аргумента *x* на некотором отрезке или на всей числовой прямой, если на нем (на ней) каждому значению *x* соответствует одно и только одно значение *y*. Такого же рода зависимость может существовать между случайными величинами. Пример Бесселя (1838, §§ 1 – 2): погрешность некоторого вида измерений равна $\eta = a\xi^2$.

Но между случайными величинами могут существовать и менее тесные связи, как рост детей в зависимости от роста родителей. Их изучение является предметом важной главы математической статистики, теории корреляции, а само слово *корреляция* означает *сопоставление*. Точнее, корреляция рассматривает изменение закона распределения одной случайной величины в зависимости от изменения самой другой случайной величины (или самих других случайных величин) и, как правило, в зависимости от сопутствующих обстоятельств.

Без подобного уточнения и без количественного изучения соответствующих явлений корреляция была известна в древности, и мы назвали бы ее качественной. Вот пример (утверждение Гиппократа): очень полные люди *склонны* умирать в более раннем возрасте, чем худощавые. Вся древняя наука была качественной. Так, климатические пояса были выделены уже в древности, но лишь Гумбольдт (1817) связал их определение со среднегодовой температурой воздуха. Добавим: у древних ученых не было количественного понятия о температуре.

Впервые корреляцию количественно исследовал Зейдель (1865 – 1866), немецкий астроном и математик. Заинтересовавшись заболеваемостью брюшным тифом, он исследовал влияние на нее уровня грунтовых вод, а затем – и этого уровня, и количества осадков.

Собственно теорию корреляции начал развивать Гальтон (1889), а несколько позднее – Пирсон. И всё-таки ее далеко не сразу удалось в достаточной мере усовершенствовать.

Марков (1916/1951, с. 533) отозвался о ней пренебрежительно (хотя, пожалуй, не совсем справедливо): ее

Положительная часть невелика и состоит в простом применении способа наименьших квадратов к разысканию линейных зависимостей. Но […], *не довольствуясь приближенным определением различных коэффициентов*, [она] *указывает еще их вероятные погрешности и здесь она вступает в область фантазии, гипноза и веры в математические формулы, которые в действительности не имеют твердого научного основания*.

Гораздо вернее положение оценил Слуцкий в письме Маркову 1912 г. (Шейнин 1999а, с. 132):

Недостатки <u>изложения</u> теории корреляции у Пирсона временные, такого же порядка, как […] *недостатки математики 17 и 18 века* [веков]. *Строгий фундамент под работу гениев был подведен только post factum, то же будет с Пирсоном. Я* [1912] *взял на себя изложение того, что сделано. А. А. Чупров изложит когда-нибудь вопрос о корреляции с философско-логической стороны, осветит его как метод исследования. Зрелому математическому уму чистого математика предоставлено будет усовершенствовать математический фундамент теории.*

Опишем понятие о коэффициенте корреляции. Даны две случайные величины, ξ и η. Вычислим момент

$$\mu_{\xi\,\eta} = E[(\xi - E\xi)(\eta - E\eta)] = E(\xi\eta) - 2E\xi\, E\eta + E\xi\, E\eta = E(\xi\eta) - E\xi\, E\eta.$$

После его деления на стандартные отклонения σ_ξ и σ_η

$$r_{\xi\eta} = \frac{\mu_{\xi\eta}}{\sigma_\xi \sigma_\eta}$$

он станет безразмерным. Полученная величина называется коэффициентом корреляции. Ясно, что $\mu_{\xi\eta}$, а потому и $r_{\xi\eta}$ равно нулю при независимых ξ и η, однако обратное утверждение неверно. Пример: воробей – птица, обратное же утверждение (птица – воробей) верно не всегда. В данном случае также достаточно одного примера, чтобы опровергнуть обратное утверждение, и вот он. Пусть плотность величины ξ – четная функция, тогда $E\xi = 0$ и $E\xi^3 = 0$; далее, выберем $\eta = \xi^2$. Тогда $\mu_{\xi\eta} = E\xi^3 - 0 = 0$, ч. т. д. Зависимость, следовательно, может существовать при нулевом коэффициенте корреляции и даже быть функциональной.

Интересно, что коэффициент корреляции принимает значения от – 1 до 1. В частности, это означает, что корреляция может быть и отрицательной (что можно было заметить и раньше). Пример: корреляция (зависимость) между ростом человека и его весом положительна, а зависимость между расстоянием от источника света и его яркостью отрицательна; статистики в таких случаях говорят *прямая* и *обратная*.

Не удовлетворяясь развивавшейся теорией корреляции, Каптейн (1912) предложил свой собственный коэффициент корреляции. На Гаусса (п. 1.1.4) он не сослался, но по существу исходил из его соображений, и новый коэффициент корреляции оказался равным отношению числа общих аргументов двух функций к полному числу аргументов. Серьезным ограничением, правда, было предположение о равенстве абсолютных значений ошибок аргументов.

Предложение Каптейна не было замечено, но нам хорошо известно, что оно интуитивно применялось (без ссылок на коэффициент корреляции) в геодезии для приближенных оценок зависимости.

3.2. Распределение систем случайных величин

Рассмотрим вероятность $P(\xi < x, \eta < y)$. На плоскости эти неравенства соответствуют бесконечной области $-\infty < \xi < x$, $-\infty < \eta < y$, аналитически же можно считать, что указанная вероятность выражается функцией распределения:

$$F(x; y) = P(\xi < x, \eta < y).$$

Отдельно же случайные величины ξ и η имеют функции распределения $F_1(x)$ и $F_2(y)$. К примеру, поскольку при бесконечно большом значении x равенство $\xi < x$ выполняется тождественно, то

$$F(+\infty; y) = F_2(y).$$

Аналогично одномерному случаю здесь вводится и плотность:

$$P[(\xi; \eta) \text{ принадлежит области } D] = \iint_D f(x; y)dxdy.$$

Плотностью и является функция $f(x; y)$, и для независимых ξ и η имеет место равенство

$$f(x; y) = f_1(x)\, f_2(y).$$

Приведем теперь формулу плотности $f(x; y)$ двумерного нормального закона:

$$\frac{1}{2\pi\sigma_x\sigma_y}\exp\{-\frac{1}{2(1-r^2)}[\frac{(x-\mathrm{E}\xi)^2}{\sigma_\xi^2}-\frac{2r(x-\mathrm{E}\xi)(y-\mathrm{E}\eta)}{\sigma_\xi\sigma_\eta}+\frac{(y-\mathrm{E}\eta)^2}{\sigma_\eta^2}]\}.$$

Здесь, помимо прежних обозначений, r – коэффициент корреляции величин ξ и η.

3.2.1. Распределение суммы случайных величин

Заданы случайные величины ξ и η с плотностями распределения $\varphi_1(x)$ и $\varphi_2(y)$. Требуется определить закон распределения их суммы $\omega = \xi + \eta$. Для функции распределения их системы имеем

$$F(x; y) = \iint \varphi(x; y)\, dxdy.$$

В случае бесконечных областей определения обеих функций интегрирование происходит по бесконечной полуплоскости, показанной на Рис. 6, так что

$$F(x; y) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\omega-x} \varphi(x; y) dy\,.$$

Дифференцируя этот интеграл по ω, получим

$$F'_\omega(x; y) = \int_{-\infty}^{\infty} \varphi(x; \omega - x) dx = f(\omega)\,,$$

или, при перемене мест x и y,

$$F'_\omega(x; y) = \int_{-\infty}^{\infty} \varphi(\omega - y; y) dy = f(\omega)\,.$$

Особо рассмотрим случай независимости исходных случайных величин, при которой искомый закон распределения называется *композицией* их плотностей. Из полученных формул следует, что

$$f(\omega) = \int_{-\infty}^{\infty} \varphi_1(x)\varphi_2(\omega - x) dx = \int_{-\infty}^{\infty} \varphi_1(\omega - y)\varphi_2(y) dy\,.$$

Иногда вычисления значительно упрощаются при помощи геометрических построений, см. Вентцель (1969, §§ 12.5 – 12.6), мы же заметим, что задачу о встрече (п. 1.1.2) можно истолковать поновому, пользуясь уже понятиями случайной величины и плотности распределения. Действительно, речь там шла о том, что случайная точка (ξ; η) должна была находиться в некотором квадрате, а сумма ($\xi + \eta$) – между двумя параллельными прямыми. Ее распределение могло бы ответить на вопрос о вероятности встречи; напомним, что моменты прихода на встречу, ξ и η, считались независимыми.

Глава 4. Предельные теоремы

4.1. Законы больших чисел

По результатам независимых испытаний можно определить статистическую вероятность $\hat{p}$ появления события, см. формулу (1.7), теоретическая же его вероятность p определяется по формуле (1.1). Из недостатков последней, указанных нами в п. 1.1.1, повторим, что ее далеко не всегда можно применить: нет равновозможных случаев, и приходится прибегать к статистической вероятности.

4.1.1. Якоб Бернулли

“Самому ограниченному человеку” ясно (Бернулли 1713/1986, гл. 4 из ч. 4-й, с. 42), что чем больше число испытаний n, “тем менее опасность не достичь цели”.

Но он продолжал: “остается исследовать, будет ли при […] увеличении числа наблюдений” разность $|p - \hat{p}|$ неограниченно убывать. Если нет, то эта разность останется больше некоторого положительного числа, и статистическая вероятность окажется недостаточно хорошей оценкой p. Таков был вопрос, который Бернулли задал сам себе, и иначе его можно поставить так: не окажется ли индукция (т. е. испытания, опыты) принципиально хуже дедукции (теории)? На этот вопрос он смог ответить вполне определенно: при $n \to \infty$ разность $|p - \hat{p}|$ стремится к нулю, индукция в этом смысле не хуже дедукции. *Его исследование открыло для теории вероятностей широкий простор*, которого раньше у нее не было.

Но продолжим. Точнее, к нулю стремится не сама разность, а ее вероятность: при $n \to \infty$ и сколь угодно малом положительном ε она имеет предел

$$\lim P(|p - \hat{p}| < \varepsilon) = 1. \qquad (4.1)$$

Этот предел равен именно 1, а не какому-нибудь меньшему положительному числу, что и означает: индукция равнозначна дедукции. Есть тут, правда, одна тонкость, присущая вообще пределу некоторой вероятности: хорошо, пусть вероятность в пределе равняется 1, и всё же в редких случаях отклонение статистической вероятности $\hat{p}$ от теоретической, см. формулу (4.1), может оказаться значительным. Не верите? Вспомните пример (п. 1.2.3), в котором происходит событие с нулевой вероятностью, а именно выбор определенной точки, расположенной на заданном интервале. В данном случае подобным является событие $|p - \hat{p}| \geq \varepsilon$.

Здесь видно принципиальное различие предела вероятности от "обычного" предела, применяемого в других отраслях математики. Вот там-то ничего подобного произойти не может.

Формула (4.1) со времени Пуассона называется законом больших чисел. Есть в теории вероятностей и усиленный закон больших чисел, который при определенных условиях позволяет избежать описанной неприятности, но мы оставляем его в стороне.

Самой формулы (4.1) у Бернулли не было, но зато он доказал, что неравенство

$$P(|p - \hat{p}| \le \varepsilon) \ge 1 - \delta, \delta > 0$$

начнет выполняться при заданных ε и δ как только число n превысит некоторое N, зависящее от этих величин; при уменьшении одной из них, т. е. при ужесточении условий, N, естественно, определенным образом (который Бернулли и установил) увеличивается.

Бернулли и сейчас считается классиком теории вероятностей, хоть его подсчеты и не были удачны: значения N можно было значительно снизить (Марков 1924, с. 46 и далее). Пирсон (1924) также улучшил оценки Якоба Бернулли, но он при этом воспользовался не известной в те времена формулой Стирлинга. И во всяком случае сравнивать закон Бернулли с неверной птолемеевой системой мира, ограничиваясь при этом утверждением, что применяющие его разорятся и не замечая его громадного принципиального значения (Пирсон 1925), недопустимо.

В 1703 – 1705 гг., т. е., конечно же, до посмертной публикации *Искусства предположений*, Бернулли обменялся несколькими письмами с Лейбницем; латинские тексты их переписки частично переведены на немецкий язык (Gini 1946). Лейбниц так и не согласился с тем, что наблюдения могут обеспечивать практическую достоверность и заявил (там же, с. 405), что учет всех обстоятельств важнее утонченных вычислений. Впоследствии ту же мысль высказал Mill (1843/1914, с. 490), но можно полагать, что учет обстоятельств и вычисления не противоречат друг другу, важно и одно, и другое.

4.1.2. Пуассон

Но что же сам Пуассон назвал законом больших чисел? Вот его качественное определение (1837, с. 7):

Вещи любой природы подвержены универсальному закону, который можно назвать <u>законом больших чисел</u>. Он состоит в том, что, если наблюдать весьма значительное число событий одной и той же природы, зависящие и от постоянных причин, и от беспорядочно изменяющихся то в одном направлении, то в другом, то среди этих чисел обнаружатся почти постоянные соотношения.

Современное качественное определение таково (Гнеденко 1954, с. 185): этим законом называется

Вся совокупность предложений, утверждающих с вероятностью, сколь угодно близкой к единице, что наступит некоторое событие, зависящее от неограниченно увеличивающегося числа случайных событий, каждое из которых оказывает на него лишь незначительное влияние.

Сейчас теоремой Пуассона принято называть равенство

$$\lim P(|\frac{\mu}{n} - \overline{p}| < \varepsilon) = 1,\ n \to \infty, \tag{4.2}$$

где μ/n – частость события в n независимых испытаниях, имеющего вероятность наступления p_k в испытании k.

4.1.3. Дальнейшая история

Чебышев (1867) доказал подобную теорему в более общем случае, а в 1928 г. Хинчин (1927) обобщил и теорему Чебышева. Наконец, приведем еще не совсем общую формулу закона: если существует такое a, что

$$\lim P(|\overline{\xi}_n - a| < \varepsilon) = 1,\ n \to \infty,$$

то последовательность величин ξ_k подчиняется этому закону.

4.2. Теорема Муавра – Лапласа

Пусть в каждом испытании изучаемое событие появляется с вероятностью p и не наступает с вероятностью q, $p + q = 1$ и в серии n таких независимых испытаний оно произошло μ раз. Тогда при $n \to \infty$

$$\lim P(a \le \frac{\mu - np}{\sqrt{npq}} \le b) = \frac{1}{\sqrt{2\pi}} \int_a^b \exp(-\frac{z^2}{2}) dz\,. \tag{4.3}$$

Грубо говоря, биномиальное распределение в пределе становится нормальным. Эту теорему Муавр, правда, установил для частного случая $p = q = 1/2$ (в его обозначениях $a = b = 1/2$), но в заключение своего рассуждения он справедливо указал, что переход к общему случаю не представляет труда, да и название его мемуара включало слова *бином* $(a + b)^n$.

Напомним, что $np = E\mu$ и $npq = D\mu$ и заметим, что формула (4.3) представляет собой частный случай центральной предельной теоремы (п. 2.2.2-4). При выводе своей формулы Муавр, как и другие математики того времени, использовал разложения в расходящиеся ряды, суммируя их только до тех пор, пока их члены не начинали возрастать, – пока ряд не начинал действительно расходиться.

Впоследствии Лаплас (1812, гл. 3-я) вывел формулу (4.3), применив новое средство, – формулу суммирования Маклорена – Эйлера, – и добавив к правой части поправочный член, учитывающий неточность, возникающую ввиду неизбежной конечности n. Марков (1914/1951, с. 511) назвал формулу (4.3) теоремой Муавра – Лапласа; впрочем, в англоязычной литературе этот термин не привился.

Более точно, эта формула описывает *интегральную* теорему Муавра – Лапласа, и следует сформулировать и соответствующую *локальную* теорему:

$$P(\mu) \approx \frac{1}{\sqrt{2\pi npq}} \exp[-\frac{(\mu - np)^2}{2npq}]. \qquad (4.4)$$

Назначая некоторое μ в правой части формулы и подставляя соответствующие величины *n, p, q*, мы вычислим приближенную вероятность этого значения μ. Экспоненциальные функции, входящие в формулы (4.3) и (4.4), табулированы во многих учебниках и статистических таблицах.

Несколько дополнительных замечаний. Формула (4.3) имеет место равномерно относительно a и b, но понятие равномерной сходимости не было еще известно ни Муавру, ни Лапласу. Во-вторых, в те времена строгие неравенства в подобных формулах не отличались от нестрогих, но сейчас следовало бы во втором случае указывать строгое неравенство. Далее, сходимость к нормальному распределению ухудшается с уменьшением p или q от значения 1/2. Муавр этого не указал, быть может потому, что всё-таки не обратил должного внимания на общий случай $p \neq q$, но это можно заметить по современному доказательству (Гнеденко 1954, § 13).

Муавр опубликовал свой результат в 1733 г. на латинском языке, разослав некоторое число отпечатанных экземпляров составленного им мемуара своим коллегам, а в 1738 г. включил его английский перевод во второе издание своего *Учения о шансе*. Но английский язык был мало распространен на континенте Европы, доказательство теоремы было плохо понятно, потому что английские математики, следуя Ньютону, не пользовались символом интеграла, а Тодхантер (1865, с. 192 – 193), крупнейший историк теории вероятностей XIX в., весьма неудачно описал вывод формулы (4.3) и не заметил ее большого значения. Так, он чисто формально указал, что Муавр доказал ее только для частного случая. Теорема стала широко известной только в конце следующего столетия.

В более раннем сочинении 1730 г. Муавр независимо вывел формулу Стирлинга; последний лишь сообщил ему значение константы, $\sqrt{2\pi}$. И Пирсон (1924), и Марков (1924, с. 55 прим.) справедливо считали, что эту формулу следовало бы называть по имени обоих ученых, но кроме того Муавр в том же более раннем сочинении составил таблицу lg n! для n = 10(10)900 с 14-ю знаками; верны 11 – 12 знаков и допущена одна опечатка.

Опишем теперь вычисление, о котором было упомянуто в п. 2.2.4. Требовалось определить вероятность выпадения шестерки в семи бросках кости из 100. Имеем: p = 1/6 и n = 100, тогда np = 16.7 и $\sqrt{npq}$ = 13.9 и в соответствии с формулой (4.4),

$$P(\mu = 7) \approx \frac{1}{\sqrt{2\pi}\sqrt{npq}} \exp[-\frac{(7 - np)^2}{2npq}].$$

Сомножитель $\sqrt{2\pi}$ в знаменателе мы выделили потому, что экспоненциальная функция табулируется вместе с сомножителем $1/\sqrt{2\pi}$.

В п. 2.2.4 мы указали, что Муавр изучал соотношение мужских и женских рождений. Скажем более определенно: что и послужило непосредственной причиной его исследования. Об этой статистической задаче следует упомянуть особо.

Арбутнот (1712) заметил, что 82 года подряд в Лондоне крестили больше мальчиков, чем девочек. Он заявил, что случайно этого произойти не могло, поскольку вероятность подобного факта равнялась $(1/2)^{-82}$, и что статистика свидетельствовала о божественной заботе о роде человеческом, ибо смертность мальчиков и мужчин была выше, чем девочек и женщин.

Его рассуждения нельзя признать удовлетворительными (но сама задача оказалась исключительно плодотворной). Он приравнял крещения и рождения, сведения имел только о Лондоне, притом только о христианах, а о сравнительной смертности мог только догадываться (но угадал верно). Далее, ничтожная вероятность еще ничего не доказывала, ср. п. 1.2.2, и, наконец, было бы разумнее объяснить обнаруженный факт биномиальным распределением рождений.

В письме 1713 г. Николай Бернулли (Монмор 1713, с. 280 – 285) ввел именно это распределение. Обозначим ежегодное число рождений через n, из которых μ мальчиков, а неизвестное соотношение мужских и женских рождений через m/f и $p = m/(m + f)$. Заметим, что p безразмерно, т. е. что знания m и f по отдельности не требовалось.

Фактически Бернулли получил

$$P(\frac{|\mu - np|}{\sqrt{npq}} \le s) \approx 1 - \exp[-\frac{s^2}{2}],$$

где s имело порядок $\sqrt{n}$. Мы сами, как указано в этой формуле, приближенно представили правую часть в функции нормального распределения, и оказалось, что в неявной форме Бернулли ввел его гораздо раньше, чем Муавр.

Юшкевич (1986) сообщил, что по его просьбе три математика заключили, что Бернулли близко подошел к локальной теореме (4.4), однако мы не убеждены в этом. Сам факт, что исследование Бернулли изучали несколько человек, свидетельствует о трудности его истолкования.

В п. 2.2.4 мы по существу указали, что в начале своего развития теория вероятностей имела целью отделение случайного от необходимого, – от предначертания. Именно эту цель преследовал и Арбутнот, и Николай Бернулли, и Муавр, который посвятил первое издание своего *Учения о шансе* Ньютону. Это посвящение он перепечатал в последнем издании книги (с. 329). В нем Муавр четко заявил, что видел свою цель в выработке метода вычисления влияния шансов и тем самым в установлении

Определенных правил для оценки того, в какой степени некоторые виды событий могут быть вызваны предначертанием, а не шансом [что позволит], *исходя из Вашей* [Ньютона] *философии,* [изучать] *способ сбора* […] *свидетельств утонченной мудрости и предначертания, которые выявляются в явлениях природы по всей вселенной*.

Подчеркнем, что Муавр написал это до того, как доказал свою предельную теорему.

4.3. Предельная теорема Бейеса

Основная формула Бейеса (1764) имела вид

$$P(b \le r \le c) = \int_b^c u^p (1-u)^q dx \div \int_0^1 u^p (1-p)^q dx. \qquad (4.5)$$

В отличие от Бейеса, который вывел ее путем сложных логических построений, мы истолкуем ее так: в результате $n = p + q$ испытаний точка r, все положения которой на единичном отрезке равновероятны ввиду “полного незнания” (Поучение к

Предложению 9), оказалась p раз внутри отрезка [b; c], расположенного внутри единичного, и q раз вне его.

Иначе говоря, Бейес отыскивал апостериорное распределение случайной величины с априорным равномерным распределением. Напомним (п. 1.1.1-5), что непризнание указанной предпосылки послужило основанием для споров вокруг мемуара Бейеса. Но кроме того положение точки r вовсе не случайно, а неизвестно. Нельзя было бы, например, применять формулу (4.5) для отыскания вероятности какой-нибудь далекой цифре в разложении π быть равной четырем или пяти (Нейман 1938/1944, с. 212).

Четкого понятия о плотности распределения во времена Бейеса не было, но сейчас можно сказать, что формула (4.5) не противоречит ее определению. Знаменателем формулы является бета-функция Эйлера, и Бейес смог установить его. И это вычисление, и последующие выкладки были затруднительны, проследить за ними нелегко. Прояснил их Тимердинг, редактор немецкого издания мемуара Бейеса. Он же остроумным приемом записал результат Бейеса для больших p и q в виде предельной теоремы. Мы приведем ее в виде

$$\lim P\left(a \le \frac{\overline{p} - p/n}{\sqrt{pq/n^3}} \le b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{z^2}{2}\right) dz, \; n \to \infty. \qquad (4.6)$$

Здесь $\overline{p}$ – статистическая оценка неизвестной вероятности p попадания точки r в отрезок [b; c],

$$\frac{p}{n} = \mathrm{E}\overline{p}, \;\; \frac{pq}{n^3} = \mathrm{D}\overline{p}.$$

Сравнение формул Муавра (4.3) и Бейеса (4.5) убеждает нас, что обе описывают предельное поведение двух различных случайных величин

$$\frac{\xi_i - \mathrm{E}\xi_i}{\sqrt{\mathrm{D}\xi_i}}, \; i = 1, 2.$$

Исключительно интересно, что, не имея представления о понятии дисперсии, Бейес понимал, что формула Муавра недостаточно хорошо описывает определение теоретической вероятности по статистическим данным. Его (посмертный) мемуар комментировал известный статистик Прайс, который и упомянул это обстоятельство.

Но почему сам Бейес не представил свой результат в виде предельной теоремы? Одну из причин косвенно указал Прайс: предельная теорема Муавра, как он заметил, недостаточна точна при конечных значениях n (и при малых значениях p или q, о чем см. наш комментарий к формуле Муавра). Но была и другая причина. В посмертной заметке того же 1764 г., опубликованной в том же источнике, Бейес первым четко указал на опасность использования расходящихся рядов (в частности, Муавром, которого он не назвал по имени). Заметим, что Тимердинг, конечно же, обошелся без них.

Мы полагаем, что Бейес завершил построение первого варианта теории вероятностей, который таким образом включал закон больших чисел Якоба Бернулли, теоремы Муавра и его самого.

Глава 5. Случайные процессы. Цепи Маркова

5.1. Случайные функции или процессы

Случайными функциями называются случайные величины, изменяющиеся дискретно или непрерывно во времени, и примером могут служить неизбежные помехи, возникающие при работе многих приборов. Фиксируя некоторый момент времени, мы получим соответствующую случайную величину, – *сечение случайной функции*.

Естественно, что законом распределения случайной функции служит функция двух аргументов, одним из которых является время. Поэтому ожиданием случайной функции является не число, а функция (обычная, а не случайная). При фиксировании момента времени ожидание будет относиться к соответствующему сечению случайной функции, и аналогично обстоит дело для дисперсии. При сложении случайных функций и учете их зависимости потребуется обобщение понятия корреляции.

Случайными функциями без последействия называются такие случайные функции, для которых существует определенная вероятность перейти за время от одного момента до другого из заданного состояния в одно из некоторых состояний, притом дополнительные сведения о более раннем не изменяют указанной вероятности. К такому типу относится броуновское движение мельчайших частиц в жидкости под действием молекулярных сил. Открыл его в 1827 г. английский ботаник Броун.

В последнее время вошел в употребление термин *хаос*, под которым понимается квазислучайное поведение траекторий движения частиц при сильной неустойчивости. Малая и неизбежная погрешность в начальных данных приводит к невозможности предсказания движения на сколько-нибудь длительное время, к неопределенности и к запутанности и *расплыванию* траекторий, к необходимости неформальных описаний соответствующих явлений.

Изучая субъективные мнения (конец п. 1.1.1-6) Лаплас (1781; 1812, § 15) ввел случайные функции. Пусть некоторый отрезок разделен на равные или неравные интервалы, из концов которых восставлены перпендикуляры числом i. Сумма длин перпендикуляров равна единице, и его концы образуют невозрастающую последовательность. Такие перпендикуляры и ломаные, соединяющие концы перпендикуляров, а также и соответствующие кривые строятся n раз, после чего вычисляется среднее значение текущей ординаты.

По мысли Лапласа это построение, если длины перпендикуляров назначаются n различными лицами, позволило бы расположить по порядку убывания значимости i причин или достоинств стольких же кандидатов на выборные должности. Вычисления, правда, оказались весьма сложными. Каждую кривую можно считать случайной функцией, а всё их множество – случайным процессом, средняя же кривая окажется ожиданием случайной функции.

Обратимся теперь к Дарвину и рассмотрим множество особей (точнее, особей одного и того же пола) какого-либо вида. Каждую особь можно в принципе характеризовать размерами тела и его частей и органов, причем невообразимое число таких размеров n не имеет значения. Введем также обычное определение расстояния в n-мерном пространстве, и тогда каждую особь можно будет считать точкой в этом уже евклидовом пространстве.

В этом же пространстве будет находиться точка или подпространство U оптимальных для данного вида размеров, а среди потомков любой пары родителей в следующем поколении лучше других приспособленными к жизни окажутся те, кто будет ближе к U. Само U также движется в зависимости от изменений во внешнем мире, и если это изменение происходит слишком быстро, вид может исчезнуть, вообще же погибают или оставляют меньше потомства те особи, которые расположены сравнительно далеко от U. Всё это можно представить себе как дискретный случайный процесс, сечениями которого являются точки-особи каждого поколения. Наша схема, однако, является чисто качественной, потому что никаких соответствующих вероятностей мы не знаем, – ни, например, вероятностей спаривания двух данных особей, ни вероятностей размеров их потомства, неизвестны и необходимые исходные размеры. Мало того: Дарвин разумно придавал большое значение повадкам животных, про которые мы также ничего не можем сказать. Наконец, между различными размерами тела одной и той же особи существует та или иная корреляция.

Сам Дарвин (1859/1958, с. 77) заметил, что гораздо легче было бы исследовать падение подброшенной горстки перьев (тоже случайный процесс!), чем эволюцию.

Противники Дарвина в основном основывали свои доводы на невозможности эволюции в результате случая, – в результате действия случайности с равномерным распределением. И это лишний раз свидетельствует о том, что именно это распределение столь долго рассматривалось как единственное, характеризующее случайность! Бэр (1873, с. 6) и Данилевский (1885, ч. 1, с. 194) упомянули по этому поводу горе-философа, изображенного в *Путешествиях Гулливера*, но взятого Свифтом у Раймунда Луллия (XIII – XIV вв.). Этот “изобретатель”, надеясь познать все истины, записывал каждую осмысленную цепочку слов, которая появлялась при их “равномерно случайном” поиске. Впрочем, сказать что-либо более определенное стало возможным лишь в начале XX в., после обнаружения незамеченных в свое время законов Менделя.

5.2. Цепи Маркова (процессы с дискретным временем)

Пусть при испытании s осуществляется одно и только одно из событий $A_1^{(s)}, A_2^{(s)}, \ldots, A_k^{(s)}$, а в следующем испытании (условная) вероятность события $A_i^{(s+1)}$ зависит от того, что произошло в испытании s, но не от предшествовавших испытаний. Указанные условия, если они выполняются при любом испытании, определяют *однородную цепь Маркова*.

Обозначим условную вероятность события $A_j^{(s+1)}$ от $A_i^{(s)}$ через p_{ij}, тогда процесс, описываемой такой цепью Маркова, определяется квадратной матрицей (таблицей) подобных вероятностей. Ее первую строку можно обозначить последовательностью $p_{11}, p_{12}, \ldots, p_{1k}$, вторую строку – последовательностью $p_{21}, p_{22}, \ldots, p_{2k}, \ldots$, и последнюю – $p_{k1}, p_{k2}, \ldots, p_{kk}$, притом сумма вероятностей в каждой строке равна единице. Матрица в целом называется *матрицей перехода*, и состоит она из *вероятностей перехода*.

Можно аналитически составить матрицу перехода сразу через n испытаний и предельную матрицу, которая существует (т. е. существуют соответствующие предельные вероятности), если только при некотором s все элементы матрицы положительны. Марков вывел и этот результат, и несколько других, которые стали называться эргодическими теоремами. В частности, оказалось, что при определенных условиях все предельные вероятности равны друг другу.

Это замечательное свойство может пояснить, например, равномерное распределение малых планет вдоль эклиптики: ссылка на эти предельные вероятности, которые не зависят от исходных вероятностей, была бы уже достаточна. На самом деле, правда, малые планеты движутся по эллиптическим орбитам и в несколько различных плоскостях.

Собственно, Пуанкаре (1896/1987, с. 150, с. 123 в переводе), который не был знаком с цепями Маркова (он не ссылался не только на Чебышева, Маркова и Ляпунова, но даже и на Лапласа и Пуассона), уже обосновал этот факт, хоть и сложным путем. Также сложно, введя гиперкомплексные числа, он (там же, с. 301; в переводе с. 245) доказал, что при длительном тасовании расположение карт в колоде станет подчиняться равномерному распределению.

Сам Марков применил свои результаты только к исследованию чередования гласных и согласных в русском языке. Возможно, что он следовал своему принципу: "Я ни на шаг не выйду из той области, где компетентность моя не может подлежать сомнению" (Ондар 1977, с. 59; письмо А. А. Чупрову 1910 г.).

Термин *цепь Маркова* (на французском языке) впервые применил Бернштейн (1926/1964, с. 158) и ввел их Марков в серии своих исследований 1906 – 1913гг. Некоторые примыкающие темы предшествовавших авторов таковы: броуновское движение, вымирание семей, спекуляции на бирже, случайное блуждание.

Для последнего термина нет общепринятого определения, но во всяком случае процесс в описанной ниже урновой задаче можно считать (одномерным) случайным блужданием, т. е., как можно истолковать ее, дискретным перемещением частицы, движущейся вдоль некоторой прямой, влево или вправо с соответствующими вероятностями, p и q, зависящими от происшедшего в предшествовавший дискретный момент. Таким же процессом, но многомерным, является диффузия, но вот случайное блуждание с постоянными p и q, как, например, "блуждание" количества выигрышей одного из двух игроков в серии игр, уже нельзя отнести к цепям Маркова.

Итак, мы остановимся на урновой задаче Даниила Бернулли – Лапласа, равнозначной знаменитой модели Эренфестов (1907), с которой принято начинать историю дискретных случайных процессов, т. е. цепей Маркова.

В первой урне находится n белых шаров, во второй – столько же черных. Требуется найти (ожидаемое!) число белых шаров в первой урне после r циклических перекладок шаров из урны в урну (Даниил Бернулли 1770). Интереснее была его вторая задача, а именно обобщение первой на случай трех урн и шаров трех цветов. Бернулли сумел изящно решить ее, и самым важным в его исследовании было обнаружение предельного состояния, – равного (ожидаемого) количества шаров каждого цвета в каждой урне.

Этот результат проще всего подтвердить ссылкой на соответствующую эргодическую теорему для однородных цепей Маркова, но следует еще доказать, что задача Бернулли подходит под их схему. Но это просто. Действительно, для случая, например, двух урн, на каждом ($k + 1$)- м шаге перекладок мы имеем четыре события, вероятность каждого из которых изменяется от результатов шага k. Вот эти события: вынуты белые шары из каждой урны; вынут белый шар из первой урны и черный – из второй; и т. д.

Лаплас (1812, гл. 3) обобщил задачу Бернулли (не сославшись на него), допустив произвольный исходный состав урн, а затем (1814/1999, с. 843 левый столбец) указал, что в процессе перекладок можно даже добавлять новые урны, опять-таки с произвольным распределением шаров. И он заключил, видимо, слишком оптимистически, что

Можно распространить этот результат на все сочетания в природе, в которых постоянные силы […] *устанавливают правильный образ действий, способный даже из недр хаоса вызвать системы, управляемые удивительными законами.*

Глава 6. Теория ошибок и метод наименьших квадратов

6.1. Теория ошибок

Термин *теория ошибок* (на немецком языке) ввел Ламберт (1765, § 321). Он вошел во всеобщее употребление лишь в середине следующего века; в отличие от Бесселя, ни Лаплас, ни Гаусс его не применяли. Теория ошибок имеет целью исследование не только ошибок измерений, но и их обработки и к ней поэтому следует отнести метод наименьших квадратов (МНКв). Мы тем не менее выделили эту тему ввиду ее важности.

Теорию ошибок можно подразделить на вероятностную и детерминированную части. Первая изучает случайные ошибки и их влияние на результаты измерений (также и влияние ошибок округления и некоторой зависимости между получаемыми соотношениями). Вторая исследует целесообразность организации измерений при заданном порядке ошибок и методы исключения систематических ошибок.

Если обозначить случайную ошибку буквой ξ, то ее ожидание будет равно $E\xi = 0$. Если же это равенство не соблюдается (как это и бывает на самом деле), т. е. если ξ уже не является чисто случайной ошибкой, то $E\xi = a$ называется систематической ошибкой. Она сдвигает четный график плотности случайных ошибок либо вправо (при $a > 0$), либо влево (при $a < 0$).

С 1756 г. (Симпсон, п. 1.2.3) вероятностная теория ошибок стала важнейшей областью приложения теории вероятностей и оставалась весьма важной вплоть до 1920-х годов. В посмертной публикации Пуанкаре (1921/1980, с. 343) указал, что “теория ошибок естественно была […] основной целью” его трактата 1896 г., а Lévy (1925, с. vii) прямо заявил, что без приложения к теории ошибок его исследование устойчивых законов распределения было бы “бесцельным”.

Устойчивые законы стали существенным понятием теории вероятностей, но вот по отношению к теории ошибок книга Леви оказалась мертворожденной. Из формального определения устойчивого закона следует, что сумма $\sum\xi_i$ и среднее арифметическое $\bar{\xi}$ имеют то же распределение, что и независимые и одинаково распределенные случайные величины ξ_i, и Леви доказал, что реальная оценка точности указанных функций случайных величин будет весьма затруднена, если соответствующее распределение не было устойчивым.

Но наблюдатель никак не может установить, подчиняются ли погрешности его измерений устойчивому закону или нет. Более того: закон Коши тоже устойчив, но вообще не обладает дисперсией (п. 2.3.2).

В свою очередь, математическая статистика восприняла от вероятностной теории ошибок принципы наибольшего правдоподобия и наименьшей дисперсии (см. п. 3 ниже).

Теперь о детерминированной теории ошибок. О существовании систематических ошибок (в первую очередь вызванных вертикальной рефракцией) не могли не знать Гиппарх и Птолемей, хотя первым, кто четко подразделил ошибки на случайные и систематические, правда, лишь в ограниченном смысле, был Даниил Бернулли (1780).

Также в древности астрономы весьма успешно наблюдали при наилучших условиях. Хорошим примером могут служить наблюдения планет в периоды их *стояния*, т. е. изменения направления видимого движения, когда ошибка регистрации некоторого момента менее всего сказывается на результатах последующих вычислений (Neugebauer 1950/1983, с. 250).

По существу, однако, детерминированная теория ошибок возникла вместе с дифференциальным исчислением. Вот простейшая геодезическая задача: В треугольнике измерены два угла, α и β, и сторона *a* и требуется определить порядок ошибки любой из двух других вычисленных сторон по известному порядку ошибок измеренных элементов. Обозначим длину любой из этих двух сторон через *W*. Она является функцией измерений

$$W = f(a; \alpha; \beta),$$

и ее дифференциал, примерно равный ее погрешности, вычисляется по стандартным формулам.

Сразу скажем, что от рассмотрения отдельного треугольника детерминированная теория ошибок постепенно перешла к изучению рядов и даже сетей триангуляции. И вот особая задача, которая доказывает широту детерминированной теории ошибок (Бессель 1839).

Мерный жезл длиной несколько футов опирается на подставки в двух своих точках, расположенных, естественно, симметрично относительно его середины. Жезл изгибается под действием своего собственного веса, и при помощи составленных им дифференциальных уравнений Бессель определил, где именно должны быть расположены опорные точки, чтобы жезл укоротился меньше всего.

С Гаусса и Бесселя началась, мы бы сказали, новая стадия экспериментальной науки. Стало обязательным тщательно поверять каждый прибор и обдумывать всевозможные приемы исключения систематических ошибок наблюдения (и уменьшать случайные погрешности). Разумеется, очень многое было достигнуто в этом направлении и раньше и, например, усилия Тихо Браге в этом направлении известны явно недостаточно.

6.2. Истинное значение измеряемой константы

Многие науки и научные дисциплины имеют по меньшей мере одной из своих целей измерение констант, и с ними же имеет дело МНКв; метрология, например, без подобных измерений просто не существовала бы. Но что, собственно, означает часто употребляемое выражение, которым мы озаглавили этот подпункт? Оно ведь скорее относится не к математике, а к философии. В 1826 г. Фурье предложил ранее *фактически* признаваемое определение этого понятия: истинным значением константы он назвал предел, к которому стремится при $n \to \infty$ среднее арифметическое из n ее отдельных измерений (проведенных, разумеется, при одних и тех же условиях, при помощи одного и того же прибора или равноценных приборов и т. д.).

Его предложение было забыто, но многие авторы независимо друг от друга повторяли его определение. Нетрудно видеть, что оно сродни мыслям Мизеса об определении понятия вероятности события (см. п. 1.1.3). В 1963 г. один из таких авторов (Ч. Эйзенхарт) пояснил порядок применения этого ставшим обычным определения к метрологии. Масса эталона массы, как он заметил, принимается равной массе его металла плюс масса пузырьков воздуха, “приклеивающихся” к эталону. Иными словами, неизбежную остаточную систематическую погрешность измерения приходится включать в *истинное значение* константы.

Статистики редко имеют дело с константами, хотя истинное значение всё-таки иногда упоминают, а Гаусс (1816, §§ 3 и 4), как, впрочем, и некоторые статистики, даже рассуждал об истинном значении меры погрешности измерений, т. е. о том, чего в природе не существует.

По поводу этой темы см. Шейнин (2007b).

6.3. Метод наименьших квадратов

МНКв является стандартным методом математической обработки наблюдений; его обычно относят не к теории вероятностей, а к математической статистике.

Пусть неизвестные константы x и y связаны с результатами наблюдений $w_1, w_2, \ldots, w_n$ линейными уравнениями

$$a_ix + b_iy + \ldots + w_i = 0,\ i = 1, 2, \ldots, n. \qquad (6.1)$$

Число неизвестных k произвольно, но если $k > n$, то решить систему (6.1) не удастся, а если $k = n$, то никакой особой *обработки* наблюдений не потребуется. Итак, $k < n$. Коэффициенты a_i и b_i должны быть заданы соответствующей теорией, а линейность системы (6.1) можно оправдать в случае небольших x и y. Действительно, член c_ix^2, если включить его в систему, окажется очень небольшим. Если, скажем, $x = 0.1$, то $x^2 = 0.01$. Значения неизвестных невелики потому, что практически они более или менее известны (например, по прикидкам или по прежним данным). Пусть $x \approx x_0$ и $y \approx y_0$. Обозначим $x = x_0 + \Delta x$, $y = y_0 + \Delta y$ и подставим эти равенства в любое из уравнений (6.1):

$$a_i(x_0 + \Delta x) + b_i(y_0 + \Delta y) + \ldots + w_i = 0,$$
$$a_i\Delta x + b_i\Delta y + (a_ix_0 + b_iy_0 + w_i) = 0.$$

Сумма в последней скобке известна и заменяет прежнее w_i и значения новых неизвестных Δx и Δy невелики.

Но как решать систему (6.1) при $k < n$? Результаты наблюдений полагаются независимыми (или почти независимыми) друг от друга, и отбрасывание “лишних” уравнений (притом: каких именно?) было бы равносильно отбрасыванию доброкачественных наблюдений. Решение прямоугольной системы уравнений ($k < n$) в строгом смысле невозможно. При любом подборе x и y останутся какие-то *невязки*, какие-то остаточные свободные члены (назовем их v_i).

Приходится накладывать на них то или иное дополнительное условие; в любом случае они должны будут оказаться небольшими, притом положительными и отрицательными примерно в одном и том же количестве, иначе найденные x и y будут систематически искажены. Есть и обязательное *организационное* требование: два вычислителя, решая одну и ту же систему, должны получить одно и то же решение.

Самым распространенным оказалось дополнительное требование

$$v_1^2 + v_2^2 + \ldots + v_n^2 = \min, \tag{6.2}$$

откуда и произошло название МНКв. Проследим, что именно следует из него. Имеем

$$v_i^2 = (a_i x + b_i y + \ldots + w_i)^2.$$

Нам следует установить значения x, y, …, которые привели бы к условию (6.2). Они, эти “неизвестные”, являются здесь поэтому переменными величинами. Имеем

$$\frac{\partial v_i^2}{\partial x} = 2a_i(a_i x + b_i y + \ldots + w_i).$$

Теперь, как и полагается при отыскании экстремумов, приравняем нулю сумму левых частей этого уравнения по всем i:

$$\frac{\partial v_1^2}{\partial x} + \frac{\partial v_2^2}{\partial x} + \ldots + \frac{\partial v_n^2}{\partial x} = 2\sum(a_i a_i x + a_i b_i y + \ldots + a_i w_i) = 0,$$

или, в изящных обозначениях Гаусса,

$$[aa]x + [ab]y + \ldots + [aw] = 0.$$

Таким же образом, дифференцируя v_i^2 по y, и снова складывая полученные производные по всем i, получим

$$[ab]x + [bb]y + \ldots + [bw] = 0.$$

Выведенные уравнения называются *нормальными*, и их число, как сразу же усматривается, совпадает с числом неизвестных (да, теперь уже неизвестных). Решение нормальных уравнений означает оценивание исходных неизвестных, и вычисленные значения неизвестных (и даже неизвестные, входящие в нормальные уравнения) являются оценками этих исходных неизвестных, соответствующими МНКв. Обозначать такие оценки принято по-иному: $\hat{x}, \hat{y}, \ldots$ Необходимо также оценить погрешность полученных результатов, т. е. этих $\hat{x}, \hat{y}, \ldots$, но мы на этом останавливаться не будем.

Требование (6.2), как было установлено, обладает ценными качествами (Петров 1954), о которых мы говорить не будем, но заметим, что оно соответствует минимуму дисперсии

$$\frac{v_1^{\ 2}+v_2^{\ 2}+\ldots+v_n^{\ 2}}{n-k}. \qquad (6.3)$$

В знаменателе здесь число избыточных наблюдений, и точно тот же смысл имеет знаменатель формулы (2.12), которая соответствует случаю одного-единственного неизвестного. В этом случае система (6.1) записалась бы в виде

$$a_i x + w_i = 0.$$

и, как можно легко проверить, приводит к выбору обобщенного среднего арифметического.

Классическим примером систем (6.1) оказались системы с двумя неизвестными, параметрами земного эллипсоида вращения, наилучшим образом представлявшим фигуру Земли. По геодезическим измерениям (по *градусным* измерениям) определялась длина дуги меридиана протяжением в 1° (например, между 60 и 61° сев. широты). Широты, естественно, устанавливались по астрономическим наблюдениям. Разность длин таких дуг в различных широтах указывала на отклонение фигуры Земли от окружности, и для вычисления двух неизвестных (точнее, лишь примерно известных) параметров требовались два градусных измерения, практически же – много больше и для контроля измерений, и для повышения точности вывода.

Лежандр (1805, с. 72 – 73; 1814) первым рекомендовал МНКв, обосновав его просто разумными соображениями, притом ошибочно назвал величины v_i погрешностями измерений. Кроме того, по контексту явно следовало, что МНКв обеспечивал сведение к минимуму наибольшего $|v_i|$, хотя на самом деле это условие достигается методом *минимакса*, см. ниже п. 4.

Гаусс применял МНКв с 1794 или 1795 г., полагал, что он был известен издавна, а с его публикацией не спешил еще и потому, что вообще связывал приоритет не со статьями в журналах (или в книгах), а с фактическим открытием. Он (1809, § 186) назвал МНКв "своим принципом", что глубоко оскорбило Лежандра. Гаусс, впрочем, *обосновал* МНКв, но не удовлетворился своими рассуждениями, в частности потому, что нормальный закон оказывался единственным распределением ошибок наблюдений.

Много позже он (письмо Бесселю 1839 г.; *Труды*, т. 8, с. 146 – 147), впрочем, назвал основной причиной отказа от своего первого обоснования предпочтительность интегральной меры надежности (дисперсии) перед принципом наибольшего правдоподобия, который он применил в 1809 г. при выводе нормального распределения.

Тогда, в 1809 г., он не сослался на Ламберта (1760, § 295), который впервые и предложил этот принцип для плотности распределения, показанной им только на рисунке. Она соответствовала более или менее симметричной одновершинной кривой вида $\varphi(x - \hat{x})$, где $\hat{x}$ можно понимать как параметр сдвига. При наблюдениях $x_1, x_2, \ldots, x_n$ Ламберт предложил отыскивать $\hat{x}$ под условием

$$\varphi(x_1 - \hat{x})\ \varphi(x_2 - \hat{x})\ \ldots\ \varphi(x_n - \hat{x}) = \max.$$

В настоящее время оно (т. е. указанный принцип) находит применение в математической статистике.

В 1823 г. Гаусс опубликовал второе, окончательное обоснование, исходящее из принципа минимума дисперсии (6.3). Впрочем, Колмогоров (1946, § 6, с. 64), как можно понять, предложил принимать минимум дисперсии в качестве исходного положения.

Рассуждения Гаусса, которые привели его к условию (6.2), были весьма сложными, а нормальное распределение более или менее соответствовало погрешностям наблюдений и считалось почтенным ввиду центральной предельной теоремы. Кроме того, в 1860 г. Максвелл (нестрого) доказал, что в состоянии равновесия распределение скоростей газовых молекул нормально. И неудивительно, что авторы учебников многие десятилетия вообще не упоминали второго обоснования МНКв.

Первым действенным противником нормального распределения оказался Ньюком (1886), который заметил, что отклонения от нормальности в длинных рядах наблюдений вызывались тем, что им нельзя было приписать одну и ту же точность. Взамен он предложил смесь нормальных законов, но соответствующие вычисления были чрезвычайно сложны, а Эддингтон (1933, § 5) доказал, что подобная смесь не являлась нормальной.

Сомнения в применимости нормального распределения вполне мог высказать Бессель (1818) при обсуждении ряда 300 наблюдений Брадлея. Он заметил, что крупные ошибки появились в нем "несколько чаще", чем следовало бы при этом распределении, но непонятным образом объяснил это недостаточно большим числом наблюдений.

Мы (2000) также обнаружили много других ошибок и даже нелепостей в его сочинениях (и в упомянутом нами в п. 3.1 мемуаре 1838 г.).

Укажем еще, что результаты Маркова в МНКв описывались ошибочно. В 1934 г. Нейман приписал ему второе гауссово обоснование метода, после чего оно стало называться *теоремой Гаусса – Маркова*. Этот надуманный термин сохранился до сих пор (Dodge 2003, с. 161), хотя тот же Нейман (1952, с. 228) позже признал свою ошибку.

Далее, Линник и др. (1951, с. 637) заявили, что Марков "по существу" ввел понятия, "эквивалентные" нынешним понятиям несмещенности и эффективности (т. е. обладанием наименьшими дисперсиями) статистик для оценки параметров законов распределений. С таким же успехом эти авторы могли бы назвать Гаусса!

Марков (1899) поддержал второе обоснование МНКв, быть может намного решительнее многих своих предшественников, первый из которых появился в 1825 г. Приходится, однако, добавить, что он (1899/1951, с. 246) обесценил свою поддержку:

Мы не приписываем ему [МНКв] *наивероятнейшие или наиблагонадежнейшие результаты, а рассматриваем его только как общий прием, дающий приближенные величины неизвестных с условной оценкой результатов*.

Подобный "общий прием" вообще не требует никакого обоснования! См. Шейнин (2006b).

6.4. Метод минимакса

Существуют и иные методы решения систем (6.1), которые уже не обладают качествами оценок по МНКв, но могут быть полезны в некоторых частных случаях. Есть и особый метод, приводящий к наименьшей абсолютной величине наибольшую из величин v_1, v_2, …, v_n:

$$|v_{\max}| = \min. \qquad (6.4)$$

Наименьший здесь означает *из всех мыслимых*, а в простейшем случае – из нескольких разумных вариантов. Метод минимакса не имеет отношения к теории вероятностей и не приводит ни к каким "наилучшим" решениям, но он позволяет делать определенные выводы.

Напомним, что коэффициенты a_i, b_i, … в системе (6.1) задаются соответствующей теорией и спросим теперь: подтверждают ли ее наблюдения w_i ? Определив $\hat{x}, \hat{y}, \ldots$, мы можем вычислить остаточные свободные члены v_i и установить, не слишком ли они велики по сравнению с известным порядком ошибок. Иначе: не следует ли считать, что теория неверна (или что наблюдения хуже, чем они должны были бы быть)? И если даже условие (6.4) приводит к слишком крупному $|v_i|$, то наши опасения окажутся обоснованными.

Именно методом минимакса воспользовался и Эйлер, и Лаплас (который предварительно установил соответствующий алгоритм), чтобы определить, не противоречат ли проведенные градусные измерения (см. п. 3 выше) эллиптичности фигуры Земли. И элементами метода минимакса вполне мог Кеплер (1609/1992, с. 334/143) обосновать опровержение системы Птолемея: достаточно точные наблюдения Тихо Браге не согласовывались с ней. В астрономии, правда, соответствующие уравнения не являются ни линейными, ни даже алгебраическими, и Кеплеру наверняка пришлось преодолевать немалые дополнительные трудности.

Условие (6.4) равносильно требованию

$$\lim(v_1^{2k} + v_2^{2k} + \ldots + v_n^{2k}) = \min,\ k \to \infty,$$

что почти очевидно: пусть наибольшим является $|v_1|$, тогда при $k \to \infty$ можно пренебречь всеми слагаемыми суммы кроме первого. Следовательно, для достижения минимума v_1^{2k}, а потому и $|v_1|$ должно быть возможно меньшим.

Не зная броду, сунулся Стиглер (1986, с. 28) в воду и самоуверенно заявил, что Эйлер (см. выше)

Не доверял сочетаниям уравнений, считая вместе с [другими] *математиками, что при этом влияние погрешностей возрастает, а не как статистики, что случайные ошибки компенсируют друг друга.*

Интересно еще, что в конце XVIII в. математики, включая даже Лапласа и Лежандра, действительно опасались “сочетания уравнений” и отказались уравнивать звено триангуляции между двумя базисами. Взамен, они решили вычислить каждую половину звена от “своего” базиса (и, очевидно, уравнять только одну общую сторону обеих половин). Примерно в 1819 г., в Третьем дополнении к своей *Аналитической теории*, Лаплас (1812/1886, с. 590 – 591) оправдал этот отказ тем, что в то время не было еще “истинной” теории уравнивания, которую именно он (а будто бы вовсе не Гаусс, на которого он не сослался!) с тех пор установил. Намного подробнее это описано в статье Шейнин (1993а, с. 50).

В п. 1.1.4 мы указали, что при уравнивании триангуляции Советского Союза ее отдельные звенья предварительно обрабатывались и только после этого вводились в общее уравнивание. Иначе поступать было нельзя хотя бы ввиду громадного объема работы, но было и дополнительное соображение: влияние систематических ошибок следовало ограничить пределами каждого данного звена.

Глава 7. Теория вероятностей, статистика, теория ошибок

7.1. Аксиоматизация теории вероятностей

В п. 1.1.1 мы указали, возможно в тысячный раз, что классическое определение вероятности нельзя считать удовлетворительным и упомянули, что Гильберт посчитал целесообразным аксиоматизировать теорию вероятностей. Вот его высказывание (1901/1969, с. 34):

С исследованиями по основаниям геометрии близко связана задача об аксиоматическом построении по этому же образцу тех физических дисциплин, в которых уже теперь математика играет выдающуюся роль: это в первую очередь теория вероятностей и механика.

Теория вероятностей в то время была прикладной математической дисциплиной, но относить ее к физике вряд ли следовало.

Буль (1854/1952, с. 288) косвенно опередил Гильберта:

Притязания [теории вероятностей] *принадлежать чистой науке должны основываться на степени, в которой она удовлетворяет следующим условиям. Первое, принципы, на которых основаны ее методы, должны быть по своей сути аксиоматическими.*

Буль выставил еще два общенаучных условия.

Попытки аксиоматизации начались почти сразу же после рекомендации Гильберта, но по всеобщему признанию вполне удовлетворительного успеха добился только Колмогоров. Нисколько не вдаваясь в суть дела (см., например, Гнеденко 1954, § 8 в гл. 1), мы полагаем полезным привести две выдержки из его брошюры (1933/1974, Предисловие и с. 9):

Ведущей мыслью автора было при этом естественное включение основ теории вероятностей, считавшихся еще недавно совершенно своеобразными, в ряд общих понятий современной математики.

Теория вероятностей как математическая дисциплина может и должна быть аксиоматизирована совершенно в том же смысле, как геометрия и алгебра.

Результаты Колмогорова вовсе не были сразу же восприняты (Doob 1989 и 1994/1996, с. 593):

Для большинства математиков математическая вероятность относилась к математике так же, как черный рынок к маркетингу.[…] *Долгие годы* [значение монографии Колмогорова] *не признавалось, и некоторые математики насмешливо заявляли,* *что*[…] *вероятность, возможно, нуждается в строгости, но не в трупном окоченении.*

Мысль о том, что математическая случайная величина – это просто функция без какого-либо романтического сопутствующего значения, некоторым из [специалистов по теории вероятностей] *представлялась довольно унизительной*.

В аксиоматизации теории вероятностей определенную роль сыграл Хаусдорф. Его трактат по теории множеств (1914, с. 416 – 417) содержал ссылки на вероятность, но гораздо больше оказалось в его рукописях, см. Girlich (1996) и Hausdorff (2006).

Упомянем и Маркова (1924). На с. 10 он ввел странную аксиому и заявил (с. 24), что “теоремы о сложении и умножении вероятностей” служат вместе с ней “незыблемой основой для исчисления вероятностей как отдела чистой математики”, см. Шейнин (2006b). Это заявление, равно как и его аксиома были благополучно забыты.

Недавно Вовк и Шафер (2001, 2003) опубликовали свою собственную аксиоматизацию, возможно весьма интересную. Истоию нашей темы описали Barone & Novikoff (1978) и Hochkirchen (1999), который очень высоко оценил неопубликованную лекцию Хаусдорфа 1923 г.

7.2. Определения и пояснения

В нынешнем понимании статистика произошла от политической арифметики, которая возникла в середине XVII в. (У. Петти, Дж. Граунт). Она изучала население, экономику, торговлю при помощи чисел (а не качественных характеристик) и обсуждала соответствующие причины и связи, привлекая для этого простейшие соображения вероятностного характера.

Вот подходящее мнение (Kendall 1960):

Статистика, как мы сейчас понимаем этот термин, не зародилась до XVII в., да и тогда не внутри <u>статистики</u> [государствоведения], *а в политической арифметике. Феодальное государство Средневековья просто не было заинтересовано в статистике (в нашем смысле).*

Постепенно, и особенно со второй половины XIX в., статистика начала проникать в различные отрасли естествознания, что привело к возникновению термина *статистический метод*. Впрочем, мы бы выделили три периода в его развитии. На первом выводы формулировались на основе подмеченных закономерностей, см., например, утверждения Гиппократа в п. 2.1 или Цельса в п. 1.1.3. При желании к первому периоду можно отнести и библейское задание Моисея (Числа 13: 18 – 21):

И послал их Моисей высмотреть землю Ханаанскую […]. *И осмотреть землю, какова она, и народ живущий на ней, силен ли он или слаб, малочислен ли он или многочислен? И какова земля, на которой он живет* […]

Во втором периоде выводы основывались на статистических данных почти без привлечения вероятностных идей или методов (Граунт 1662). Примечательным примером может служить и выявление пути распространения холерных эпидемий (английский врач Snow 1855/1965, с. 58 – 59). На примере населения примерно в 500 тысяч человек Сноу установил, что, у тех, кто пил неочищенную воду, смертность от холеры была примерно в восемь раз выше, чем у тех, в чьи дома подавали очищенную воду. В третьем периоде, с конца XIX в., выводы начали проверяться количественными критериями.

Внутри самой статистики выделился этап предварительного исследования исходных данных, элементы которого содержались уже у Кетле (1846). Полное признание этот этап получил, пожалуй, лишь во второй половине XX в. Подобные исследования носят общенаучный, а не математический характер, а потому не признаются *математической статистикой*. Именно его включение в *теоретическую статистику* в основном и привело к различию в этих двух понятиях; интересно, что некоторые авторы признают только либо математическую, либо теоретическую статистику. В 1953 г. Колмогоров (Обзор 1954, с. 47), например, полагал, что общая теория статистики "по существу" сводилась к математической и "некоторым техническим приемам собирания и обработки статистических данных". Вот эти-то приемы и составляют исследование исходных данных.

Колмогоров и Прохоров (1982, с. 576) определили математическую статистику как раздел математики, посвященный "систематизации, обработке и использованию статистических данных"; напомним, что они определили и понятие *статистические данные* (п. 2.7). Примечательно, что схожее определение появилось чуть ли не двести лет до этого (Butte 1808, с. XI): "Теория статистики есть наука о познании и оценивании статистических данных, об их сборе и систематизации".

Термин *математическая статистика*, понимаемой в смысле развития политической арифметики, появился в середине XIX в. (Knies 1850, с. 163; Вернадский 1852, с. 237), а *теорию статистики* еще до Бутте упомянул в заглавии своей книги Schlözer (1804). Он же (с. 86) в качестве иллюстрации статистики придумал крылатую фразу "История – это текущая статистика, а статистика – стоящая неподвижно история". Многие статистики восприняли это изречение как *определение* своей науки, хотя, пожалуй, с таким же успехом можно было бы определить, например, самолет: *Автомобиль – это приземлившийся самолет, а самолет – взлетевший вверх автомобиль*.

Детерминированная теория ошибок (п. 6.1) имеет много общего и с предварительным исследованием данных, и с детищем Р. А. Фишера, планированием эксперимента (т. е. рациональной организацией измерений, подверженных случайным ошибкам). В целом же всю эту теорию мы склонны считать приложением статистического метода ко всему процессу измерений и наблюдений в экспериментальной науке, но не главой математической статистики, как это обычно утверждают. Свое мнение мы можем подкрепить аналогией: звездная статистика является приложением статистического метода к астрономии; медицинская статистика является … и т. д. Более того: в отличие от математической статистики теория ошибок не может отказаться от понятия истинного значения измеряемой константы, см. п. 6.2.

7.3. Внедрение теории вероятностей в статистику

Вряд ли кто-либо сейчас станет отрицать, что статистика основана на теории вероятностей, но так было далеко не всегда, хотя уже Якоб Бернулли (п. 4.1.1) прочно обосновал возможность приложения последней. Статистики вовсе не поспешили воспользоваться представившейся возможностью, и частично это объясняется тем, что статистические данные были в то время слишком ненадежны; основная задача состояла в их общей обработке. Далее, статистические исследования не сводятся к математическим вычислениям, очень важны и обстоятельства, сопутствовавшие изучаемому явлению, ср. мнение Лейбница в п. 4.1.1.

Наконец, по своему образованию статистики никак не были подготовлены к восприятию математических идей, и чуть ли не до 1870-х годов упорно держались за *равновозможные случаи*, т. е. за теоретическую вероятность. При отсутствии таковых считалось, что теорию вероятностей применять нельзя. Да что там 1870-е годы! Вот А. А. Кауфман в 1916 г. (Плошко и Елисеева 1990, с. 133): теория вероятностей применима лишь к независимым испытаниям с постоянной вероятностью “успеха”и при наличии равновозможных случаев!

Кетле, правда, внедрил средние наклонности к преступлению и женитьбе (впрочем, напрасно не по отдельным группам населения), но недостаточно сильно подчеркнул, что его нововведение нельзя было относить к отдельному человеку. В результате, в конце его жизни и после его смерти в 1874 г. математически малограмотные статистики ополчились против и этих наклонностей, и теории вероятностей вообще (Rümelin 1867/1875, с. 25):

Если статистика говорит мне, что я в течение ближайшего года умру с вероятностью 1:49, […] *то я должен буду смиренно склониться перед этой суровой правдой. Но если она, опираясь на подобные средние числа, захочет сказать мне, что с вероятностью 1 к такому-то числу я* [совершу преступление], *то я смогу* […] *ответить: сапожник, не суди превыше сапога*!

Физически здоровый человек мог бы с тем же основанием не соглашаться с выводами из таблицы смертности (Чупров 1909/1959, с. 211– 212).

Новую струю в статистику (в статистику населения) внес В. Лексис, которому последовали Борткевич, Чупров, Больман, Марков, которые и основали так называемое *континентальное направление* статистики. В Англии несколько позже возникла биометрическая школа (Гальтон, Пирсон), поставившая своей целью статистическое изучение дарвинизма.

Первый номер ее ставшим знаменитым журнала *Биометрика* с подзаголовком *Журнал для статистического изучения биологических проблем* уведомлял в редакционной статье:

Проблема эволюции является статистической проблемой. Дарвин основал теорию происхождения, не прибегая к математическим идеям, [но] *каждое его понятие, – вариация, естественный отбор,* […] *– сразу же представляется приспособленным к математическому определению и требующим статистического анализа.*[…] *До сего времени области работы биологов, математиков и статистиков были значительно отделены друг от друга* .[…] *Придет день* […] *когда некоторые математики станут полноправными биологами, а биологи – компетентными математиками.*

Тем не менее, биометрической школе были присущи крупные недостатки (Колмогоров 1948, с. 143): в ней

Оставались на уровнеXVIII в. представления о логической структуре теории вероятностей […] *строгие результаты относительно близости эмпирических выборочных характеристик к теоретическим относились только к случаю независимых испытаний,* […] *вспомогательный аппарат таблиц, употребляемых при статистическом исследовании, несмотря на огромную* […] *работу,*[…] *оказался весьма несовершенным в отношении охвата переходных от "малых" к "большим" выборкам случаев.*

Чупров приложил немалые усилия для сближения континентального направления и биометрической школы, но, видимо, большого успеха не достиг. Даже в 1955 г. Бауер (1955/1968, с. 238) заключил, что работы обеих школ шли рядом, но не смыкались. Впрочем, уже задолго до того можно было говорить о единой математической статистике, возникновение которой очень многим обязано Фишеру (о котором Бауер ничего не сказал).

Материал этой главы см. в статье Шейнин (2002).

Библиография

Отдельно укажем основные источники по теории вероятностей и ее истории, включенные в наш основной список:

Вентцель, Гнеденко, Колмогоров и Юшкевич, Прохоров, Румшиский, Свешников, Шейнин (2005; 2006а), Юшкевич (1970 и 1972), Arley & Buch (Арлей и Бух), Cournot (Курно), Dodge, Hald (1990; 1998), Kendall & Plackett, Kotz, Pearson (1978), Pearson & Kendall, Todhunter.

Переводы многих иностранных источников на русский язык см. в основном списке. Наши собственные переводы и сборники переводов мы указываем здесь, потому что они были опубликованы в малом числе экземпляров.

Бернулли Я. (1713, латин.), *Искусство предположений*, части 1 – 3. Берлин, 2006.

Граунт Дж., Галлей Э. (2005), *Начала статистики населения, медицинской статистики и математики страхового дела.* Берлин.

Шесть *Хрестоматий по истории теории вероятностей и статистики.* Берлин, 2006 – 2008.

Статьи, включенные в эти *Хрестоматии*, помечены в основном списке обозначениями **Х-1, Х-2** и т. д.

Все наши берлинские книги выпустило издательство NG и все они размещены в Интернете (www.sheynin.de).

Сокращение

AHES = *Archive for History of Exact Sciences*

Аристотель, Мы ссылаемся на его собрание сочинений под редакцией Д. Росса (1908 – 1930, 11 томов, и 1954, т. 12-й, Оксфорд). Есть позднейшее двухтомное издание (Принстон, 1984), а на русском языке имеются его *Сочинения*, тт. 1 – 4, 1975 – 1983.

Берви Н. В. (1899), Определение вероятнейшего значения измеренного объекта помимо постулата Гаусса. *Изв. Имп. моск. общества любителей естествознания, антропологии и этнографии при Моск. Унив.*, Отд. физич. наук, т. 10, вып. 1, с. 41 – 45.

Бернштейн С. Н. (1926, франц.), Распространение предельной теоремы теории вероятностей на суммы зависимых величин. *Собр. соч.*, т. 4. М., 1964, с. 121 – 176.

Бирман К.-Р. (1957), Задачи генуэзского лото в работах классиков теории вероятностей. *Историко-математич. исследования*, т. 10, с. 649 – 670.

Бэр К. М., Baer K. (1873), *Zum Streit über den Darwinismus*. Dorpat [Tartu].

Вентцель Е. С. (1969), *Теория вероятностей*, 4-е изд. М. Не менее 10 изданий.

Вернадский В. И. (1852), Задачи статистики. В книге Дружинин Н. К. (1963), *Хрестоматия по истории русской статистики*. М., с. 221 – 238.

Вовк В. Г., Шафер Г. Р., Shafer G., Vovk V. (2001), *Probability and Finance. It's Only a Game*! New York.

(2003), Вклад А. Н. Колмогорова в основания теории вероятностей. *Проблемы передачи информации*, т. 39, с. 24 – 35.

Гнеденко Б. В. (1954), *Курс теории вероятностей*, изд. 2-е. Большое число последующих изданий на англ. и нем. языках.

Данилевский Н. Я. (1885), *Дарвинизм*, т. 1, ч. 1 – 2. СПб.

Колмогоров А. Н. (1933, нем.), *Основные положения теории вероятностей*. М., 1936 и 1974.

--- (1946), К обоснованию метода наименьших квадратов. *Успехи математических наук*, т. 1, № 1, с. 57 – 71.

--- (1948), Е. Е. Слуцкий. Там же, т. 3, № 4, с. 143 – 151.

--- (1963), On tables of random numbers. *Sankhya, Indian J. Statistics*, vol. A25, pp. 369 – 376.

Колмогоров А. Н., Петров А. А., Смирнов Ю. М. (1947), Одна формула Гаусса из теории метода наименьших квадратов. *Изв. АН СССР*, сер. математич., т. 11, с. 561 – 566.

Колмогоров А. Н., Прохоров Ю. В. (1982), Математическая статистика. *Математическая энциклопедия*, т. 3, с. 576 – 581.

Колмогоров А. Н., Юшкевич А. П., ред. (1978), *Математика XIX века*, т. 1. М.

Линник Ю. В., Сапогов Н. А., Тимофеев В. Н. (1951), Очерк работ А. А. Маркова по теории чисел и теории вероятностей. В книге Марков (1951, с. 614 – 640).

Марков А. А. (1899), Закон больших чисел и метод наименьших квадратов. В книге автора (1951, с. 231 – 251).

--- (1900), Исчисление вероятностей. СПб. Дальнейшие издания: 1908, 1913, 1924.
--- (1911), Об основных положениях исчисления вероятностей и о законе больших чисел. В книге Ондар (1977, с. 161 – 166).
--- (1914), О задаче Якова Бернулли. В книге автора (1951, с. 511 – 521).
--- (1916), О коэффициенте дисперсии. Там же, с. 523 – 535.
--- (1951), *Избранные труды*. М.
Менделеев Д. И. (1877), Нефтяная промышленность в Северо-Американском штате Пенсильвания и на Кавказе. *Соч.*, Л. – М., т. 10, 1949, с. 17 – 244.
Нейман Ю., Neyman J. (1938, франц.), Статистические оценки как проблема классической теории вероятностей. *Успехи математических наук*, т. 10, 1944, с. 207 – 229.
--- (1952), *Lectures and Conferences on Mathematical Statistics and Probability*. Washington.
Обзор (1954), Обзор научного совещания по вопросам статистики. *Вестник статистики*, № 5, с. 39 – 95.
Ондар Х. О., ред. (1977), *О теории вероятностей и математической статистике. Переписка А. А. Маркова и А. А. Чупрова*. М.
Петров В. В. (1954), О методе наименьших квадратов и его экстремальных свойствах. *Успехи математических наук*, т. 9, с. 41 – 62.
Плошко Б. Г., Елисеева И. И. (1990), *История статистики.* М.
Прохоров Ю. В., ред., (1999), *Вероятность и математическая статистика. Энциклопедия.* М.
Прохоров Ю. В., Севастьянов Б. А. (1999), Вероятностей теория. В книге Прохоров (1999, с. 77 – 81).
Румшиский Л. З. (1966), *Элементы теории вероятностей*. М.
Свешников А. А., ред. (1970), *Сборник задач по теории вероятностей, математической статистике и теории случайных функций.* М. 2-е изд.
Слуцкий Е. Е. (1912), *Теория корреляции*. Киев.
Смирнов Н. В., Дунин-Барковский И. В. (1959), *Математическая статистика для технических приложений*. М. 2-е изд., 1965.
Тутубалин В. Н. (1972), Теория вероятностей в естествознании. М.
Успенский В. А., Семенов А. Л., Шень А. Х. (1990). Может ли (индивидуальная) последовательность нулей и единиц быть случайной? *Успехи математич. наук*, т. 45, с. 105 – 162.
Хинчин А. Я., Khinchin A. Ya. (1927), Über das Gesetz der großen Zahlen. *Math. Ann.*, Bd. 96, pp. 152 – 168.
--- (1961), Частотная теория Мизеса и современные идеи теории вероятностей. *Вопр. философии*, № 1, с. 91 – 102; № 2, с. 77 – 89.
Чебышев П. Л. (1845), Опыт элементарного анализа теории вероятностей. *Полн. собр. соч.*, т. 5. М. – Л., 1951, с. 26 – 87.
--- (1867), О средних величинах. Там же, т. 2. М. – Л., 1947, с. 431 – 437.
--- (1879 – 1880, лекции), *Теория вероятностей*. М. – Л., 1936.

Чупров А. А. (1909), *Очерки по теории статистики*. М. Последующие издания: 1910, 1959.

Шейнин О. Б., Sheynin O. B. (1971), Lambert's work in probability. AHES, vol. 7, pp. 244 – 256.

--- (1973), Finite random sums. AHES, vol. 9, pp. 275 – 305.

--- (1977), Early history of the theory of probability. AHES, vol. 17, pp. 201 – 259.

--- (1993a), On the history of the principle of least squares. AHES, vol. 46, pp. 39 – 54.

--- (1993b), Treatment of observations in early astronomy. AHES, vol. 46, pp. 153 – 192.

--- (1999a), Е. Е. Слуцкий: к 50-летию со дня смерти. *Историко-математические исследования*, вып. 3 (38), с. 128 – 137.

--- (1999b), Gauss and the method of least squares. *Jahrbücher f. Nationalökonomie und Statistik*, Bd. 219, pp. 458 – 467.

--- (2000), Bessel: some remarks on his work. *Hist. Scientiarum*, vol. 10, pp. 77 – 83.

--- (2002), Теория статистики: исторический обзор. *Вопр. статистики*, № 9, с. 64 – 69.

--- (2005), *Теория вероятностей. Исторический очерк.* Берлин.

--- (2006a), *История теории вероятностей и статистики в кратких высказываниях.* Берлин.

--- (2006b), Markov's work on the treatment of observations. *Hist. Scientiarum*, vol. 16, pp. 80 – 95. Русский перевод будет опубликован в *Историко-математич. исследованиях* в 2009 г.

--- (2007a), *История теории ошибок.* Берлин.

--- (2007b), The true value of a measured constant and the theory of errors. *Hist. Scientiarum*, vol. 17, pp. 38 – 48.

--- (2008), Bortkiewicz' alleged discovery: the law of small numbers. *Hist. Scientiarum*, vol. 18,

Юшкевич А. П., ред. (1970, 1972), *История математики с древнейших времен до начала XIX века*, тт. 2 и 3. М.

--- (1986), Николай Бернулли и издание "Искусства предположений" Якоба Бернулли. *Теория вероятностей и ее применения*, т. 31, с. 333 – 352.

Arbuthnot J. (1712), An argument for Divine Providence taken from the constant regularity observed in the birth of both sexes. В книге Kendall & Plackett (1977, pp. 30 – 34). **X-1**

Arley N., Buch K. R., Арлей Н., Бух К. Р. (1949, англ.), *Введение в теорию вероятностей и математическую статистику*. М., 1951.

Barone J., Novikoff A. (1978), A history of the axiomatic formulation of probability from Borel to Kolmogorov. AHES, vol. 18, pp. 123 – 190.

Bauer R. K., Бауер Р. К. (1955, нем.), Теория дисперсии Лексиса в ее отношениях к новым течениям статистической методологии. В книге Четвериков Н. С., ред. (1968), *О теории дисперсии*. М., 1968, с. 225 – 238.

Bayes T. (1764 – 1765), An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Roy. Soc.*, vol. 53 за 1763, pp. 360 – 418; vol. 54 за 1764, pp. 296 – 325. Перепечатка 1-й части: *Biometrika*, vol. 45, 1958, pp. 293 – 315; также Pearson & Kendall

(1970, pp. 131 – 153). Нем. перевод обеих частей: Лейпциг, 1908. **X-1.**

Bernoulli D. (1770), Disquisitiones analyticae de nouo problemata coniecturale. В книге автора (1982, pp. 306 – 324).

--- (1778, латин.), The most probable choice between several discrepant observations and the formation therefrom of the most likely induction. *Biometrika*, vol. 48, 1961, pp. 3 – 13. Также в книге Pearson & Kendall (1970, pp. 155 – 172). **X-1.**

--- (1780), Specimen philosophicum de compensationibus horologicis et veriori mensura temporis. В книге автора (1982, pp. 376 – 390).

--- (1982), *Werke*, Bd. 2. Basel.

Bernoulli J., Бернулли Я. (1713, латин.), *Wahrscheinlichkeits-rechnung*. Leipzig, 1899; Thun – Frankfurt/Main, 1999. Русск. перевод части 4-й с комментариями в книге автора *О законе больших чисел*. М., 1986.

--- (1975), *Werke*, Bd. 3. Basel.

Bernoulli N. (1709), De usu artis conjectandi in jure. В книге Bernoulli J. (1975, pp. 289 – 326).

--- (1713), Письма Монмору. В книге Montmort (1708/1713).

Bertrand J. (1888), *Calcul des probabilités*. Paris. 2-е изд., 1907. Перепечатка: New York, 1970, 1972. **X-1 (переведено аннотированное автором содержание).**

Bessel F. W., Бессель Ф. В. (1816), Untersuchungen über die Bahn des Olbersschen Kometen. *Abh. Preuss. Akad. Wiss. Berlin*, math. Kl., 1812 – 1813, pp. 119 – 160.

--- (1818), *Fundamenta astronomiae*. Königsberg. Нем. перевод отрывка в книге Schneider I., ред. (1988), *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfangen bis 1933*. Darmstadt, pp. 277 – 279.

--- (1838, нем.), Исследование о вероятности ошибок наблюдений. В книге автора *Избр. геодезич. соч*. М., 1961, с. 226 – 258.

--- (1839, нем.), Влияние силы тяжести на фигуру жезла. Там же, с. 187 – 199.

Bienaymé I. J. (1853), Considérations à l'appui de la découverte de Laplace sur la loi des probabilités dans la méthode des moindres carrés. *C. r. Acad. Sci. Paris*, t. 37, pp. 309 – 324. Также *J. math. pures et appl*., t. 12, 1867, pp. 158 – 176.

Boltzmann L., Больцман Л. (1868, нем.), Исследование равновесия живой силы между двумя движущимися материальными точками. *Избр. произв.* М., 1984, с. 30 – 66.

Bomford G., Бомфорд Г. (1952), *Geodesy*. Oxford. Дальнейшие издания: 1962, 1971. Русский перевод: *Геодезия*. М., 1958.

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

--- (1854), On the conditions by which the solution of questions in the theory of probabilities are limited. Там же, с. 280 – 288.

Bortkiewicz L. (1898), *Das Gesetz der kleinen Zahlen*. Leipzig. **X-5.**

Buffon G. L. L. (1777), Essai d'arithmétique morale. *Oeuvr. philosophiques*. Paris, 1954, pp. 456 – 488. **X-3 (частичный перевод).**

Butte W. (1808), *Die Statistik als Wissenschaft*. Landshut.

Celsus, Цельс (1935), *De medicina*, vol. 1. London. На англ. яз. Написано в I в. н. э. Русский перевод: *О медицине*. М., 1959.
Clausius R. (1858), Über die mittlere Länge der Wege […] bei der Molekularbewegung. *Abhandlungen über die mechanische Wärmetheorie*, Abt. 2. Braunschweig, 1867, pp. 260 – 276.
Condorcet M. J. A. N. Caritat (1784), Sur le calcul des probabilités. *Hist. Acad. Roy. Sci. Paris* за *1781 avec Mém. math. et phys. pour la même année*, pp. 707 – 728.
Cournot A. A., Курно О. (1843, франц.), *Основы теории шансов и вероятностей.* М., 1970.
Czuber E. (1903), *Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fehlerausgleichung, Statistik und Lebensversicherung*, Bd. 1. New York, 1968.
D'Alembert Le Rond J. (1767), Doutes et questions sur le calcul des probabilités. В книге автора *Mélanges de literature, d'histoire et de philosophie*, t. 5. Paris, pp. 239 – 264.
Darwin C., Дарвин Ч. (1859), *Origin of Species*. London – New York, 1958. *Происхождение видов.* М., 1952.
De Moivre A. (1712, латин.), De mensura sortis, or, the measurement of chance. *Intern. Stat. Rev.*, vol. 52, 1984, pp. 236 – 262. Комментарий (А. Хальд): с. 229 – 236. **Х-6 (переведено Предисловие, Посвящение Ньютону и Введение).**
--- (1718), *Doctrine of Chances*. Дальнейшие издания: 1738, 1756. Перепечатка последнего издания: New York, 1967. Х-1 (
--- (1725), *Treatise of Annuities on Lives*. Включено в *Doctrine of Chances*, 1756, pp. 261 – 328.
--- (1733, латин.), A method of approximating the sum of the terms of the binomial $(a + b)^n$ expanded into a series. Авторский перевод включен в *Doctrine of Chances*, 1738 и 1756. В 1756 г. (с дополнениями) с. 243 – 254.
De Montessus R. (1903), Un paradoxe du calcul des probabilités. *Nouvelles annales math.*, sér. 4, t. 3, pp. 21 – 31.
Dodge Y., ред. (2003), *Oxford Dictionary of Statistical Terms*. Oxford.
Doob J. L. (1989), Commentary on probability. В книге Duren P., ред. *Centenary of Mathematics in America*, pt. 2. Providence, Rhode Island, pp. 353 – 354.
--- (1994), The development of rigor in mathematical probability, 1900 – 1950. *Amer. Math. Monthly*, vol. 103, 1996, pp. 586 – 595.
Dutka J. (1988), On the St. Petersburg paradox. AHES, vol. 39, pp. 13 – 39.
Eddington A. S. (1933), Notes on the method of least squares. *Proc. Phys. Soc.*, vol. 45, pp. 271 – 287. **Х-6.**
Ehrenfest P., Ehrenfest T.; Эренфест П., Эренфест Т. (1907, нем.). О двух известных возражениях против Н-теоремы Больцмана. В книге П. Эренфест (1972), *Сборник статей*. М., с. 89 – 97.
Eisenhart C. (1963), Realistic evaluation of the precision and accuracy of instrument calibration. В книге Ku H. H., ред. (1969), *Precision Measurement and Calibration*. Nat. Bureau of Standards. *Sp. Publ. 300*, vol. 1. Washington, pp. 21 – 47.

Euler L., Эйлер Л. (1748, латин.), *Введение в анализ бесконечных*, т. 1. М., 1961.
Fourier J. B. J. (1826), Sur les résultats moyens déduits d'un grande nombre d'observations. *Oeuvr.*, t. 2. Paris, 1890, pp. 525 – 545.
Freudenthal H. (1951), Das Petersburger Problem in Hinblick auf Grenzwertsätze der Wahrscheinlichkeitsrechnung. *Math. Nachr.*, Bd. 4, pp. 184 – 192.
Galen C. (1951), *Hygiene*. Springfield, Illinois.
Galilei G. (опубл. 1718), Sopra le scoperte dei dadi. В книге David F. N. (1962), *Games, Gods and Gambling.* London, pp. 192 – 195. **X-2.** Англ. перевод, но без англ. названия.
Galton F. (1877), Typical laws of heredity. *Nature*, vol. 15, pp. 492 – 495, 512 – 514, 532 – 533. Также *Roy. Institution of Gr. Britain*, 1879, vol 8, pp. 282 – 301.
--- (1889), *Natural Inheritance*. London – New York. [New York, 1973.]
Gauss C. F. (1809, латин.), Теория движения небесных тел, кн. 2, раздел 3. В книге автора (1957, с. 89 – 109).
--- (1816, нем.), Определение точности наблюдений. Там же, с. 121 – 128.
--- (1823a, нем.), Авторское сообщение о второй части мемуара Гаусс (1823b). Там же, с. 144 – 147.
--- (1823b, латин.), Теория комбинаций наблюдений, подверженных наименьшим ошибкам, части 1 – 2. Там же, с. 17 – 57.
--- (1828, латин.), Теория комбинаций …, Дополнение. Там же, с. 59 – 88.
--- (1900, 1903), *Werke*, Bde. 8, 9. Göttingen – Leipzig.
--- (1957), *Избранные геодезические сочинения*, т. 1. М.
Gini C. (1946), Gedanken von Theorem von Bernoulli. *Z. f. Volkswirtschaft u. Statistik*, 82. Jg, pp. 401 – 413.
Girlich H.-J. (1996), Hausdorffs Beiträge zur Wahrscheinlichkeitstheorie. В книге Brieskorn E., ред. *Felix Hausdorff zum Gedächtnis*, Bd. 1. Braunschweig, pp. 31 – 70.
Graunt J. (1662), *Natural and Political Observations Made upon the Bills of Mortality*. Baltimore, 1939.
Hald A. (1990), *History of Probability and Statistics and Their Applications before 1750*. New York.
--- (1998), *History of Mathematical Statistics from 1750 to 1930*. New York.
Hausdorff F. (1914), *Grundlehren der Mengenlehre*. Leipzig.
--- (2006), *Ges. Werke*, Bd. 5. Berlin.
Helmert F. R. (1904), Zur Ableitung der Formel von Gauss für die mittleren Beobachtungsfehler und ihrer Genauigkeit. *Z. f. Vermessungswesen*, Bd. 33, pp. 577 – 587. Также в *Sitz.-Ber. Kgl. Preuss. Akad. Wiss. Berlin*, 1904, Hlbbd 1, pp. 950 – 964.
Herschel W. (1817), Astronomical observations and experiments tending to investigate the local arrangement of celestial bodies in space. *Scient. Papers*, vol. 2. London, pp. 575 – 591. [Bristol, 2003.]
Hilbert D., Гильберт Д. (1901, нем.), *Проблемы Гильберта.* М., 1969.

Hochkirchen T. (1999), *Die Axiomatisierung der Wahrscheinlichkeitsrechnung*. Göttingen.
Humboldt A. (1817), Des lignes isothermes. *Mém. Phys. Chim. Soc. D'Arcueil*, t. 3, pp. 462 – 602.
Huygens C. (1657, латин., 1660, голл.), De calcul dans les jeux de hazard. *Oeuvr. Compl.*, t. 14. La Haye, 1920, pp. 49 – 91. Франц. и голл. **Х-1.**
--- (1669, франц.), Переписка. *Oeuvr. Compl.*, t. 6. La Haye, 1895, 515 – 518, 524 – 532, 537 – 539. **Х-2.**
Kamke E. (1933), Über neuere Begründungen der Wahrscheinlichkeitsrechnung. *Jahresber. Deutsche Mathematiker-Vereinigung*, Bd. 42, pp. 14 – 27.
Kapteyn J. C. (1912), Definition of the correlation-coefficient. *Monthly Notices Roy. Astron. Soc.*, vol. 72, pp. 518 – 525.
Kendall M. G. (1960), Where shall the history of statistics begin? *Biometrika*, vol. 47, pp. 447 – 449.
Kendall M. G., Plackett R. L., ред. (1977), *Studies in the History of Statistics and Probability*, vol. 2. London. Сборник перепечаток.
Kepler J. (1601, латин.), On the most certain fundamentals of astrology. *Proc. Amer. Phil. Soc.*, vol. 123, 1979, pp. 85 – 116.
--- (1604, нем.), Thorough description of an extraordinary new star. *Vistas in Astronomy*, vol. 20, 1977, pp. 333 – 339.
--- (1609, латин.), *New Astronomy*. Cambridge, 1992.
--- (1618 – 1621, латин.), *Epitome of Copernican Astronomy*, books 4 – 5. В книге *Great Books of the Western World*, vol. 16. Chicago, 1952, pp. 845 – 1004.
Knies C. G. A. (1850), *Die Statistik als selbstständige Wissenschaft*. Kassel.
Kotz S., ред. (2006), *Encyclopedia of Statistical Sciences*, vols 1 – 16. Hobokan, New Jersey. 2-е изд.
Lambert J. H. (1760), *Photometria*. Augsburg.
--- (1765), Anmerkungen und Zusätze zur praktischen Geometrie. В книге автора *Beiträge zum Gebrauche der Mathematik und deren Anwendung*, Tl. 1. Berlin, pp. 1 – 313.
--- (1771), *Anlage zur Architektonik*, Bd. 1. Hildesheim, 1965.
--- (1772 – 1775), Essai de taxéometrie ou sur la mesure de l'ordre. *Phil. Schriften*, Bd. 8, Tl. 1. Hildesheim, 2007, pp. 423 – 460.
Langevin P. (1913), La physique du discontinu. В книге *Les progrès de la physique moléculaire*. Paris, 1914, pp. 1 – 46.
Laplace P. S., Лаплас П. С. (1776), Recherches sur l'intégration des equations différentielles. *Oeuvr. Compl.*, t. 8. Paris, 1891, pp. 69 – 197.
--- (1781), Sur les probabilités. *Oeuvr. Compl.*, t. 9. Paris, 1893, pp. 383 – 485.
--- (1812), *Théorie analytique des probabilités*. *Oeuvr. Compl.*, t. 7. Paris, 1886. **Х-1 (переведено Предисловие к 1-му изд); Х-2 (переведена гл. 4).**
--- (1814, франц.), Опыт философии теории вероятностей. В книге Прохоров (1999, с. 834 – 863).
Laurent P. H. (1873), *Traité du calcul des probabilités*. Paris.
Legendre A. M. (1805), *Nouvelles méthodes pour le détermination des orbites des comètes.* Paris.

--- (1814), Méthode des moindres quarrés. *Mém. Acad. Sci. Paris*, Cl. sci. math. et phys., t. 11, pt. 2, année 1910, pp. 149 – 154. **X-2.**

Lévy P. (1925), *Calcul des probabilités*. Paris.

Maupertuis P. L. M. (1745), Venus physique. *Oeuvr.*, t. 2. Lyon, 1756, 1768, pp. 1 – 133. Англ. перевод New York – London, 1966.

Michell J. (1767), An inquiry into the probable parallax and magnitude of the fixed stars. *Phil. Trans. Roy. Soc. Abridged*, vol. 12, 1809, pp. 423 – 438.

Mill J. S., Милль Дж. С. (1843, англ.), *Система логики*. СПб, 1914.

Montmort P. R. (1708), *Essay d' analyse sur les jeux de hazard*. Paris, 1713. New york, 1980.

Neugebauer O. (1950), The alleged Babylonian discovery of the precession of the equinoxes. В книге автора *Astronomy and History. Sel. Essays*. New York, 1983, pp. 247 – 254.

Newcomb S. (1860), [Discussion of the principles of probability theory]. *Proc. Amer. Acad. Arts and Sciences*, vol. 4 за 1857 – 1860, pp. 433 – 440.

--- (1886), A generalized theory of the combination of observations. *Amer. J. Math.*, vol. 8, pp. 343 – 366. **X-6.**

Newton I. (1664 – 1666), Рукопись без названия. *Math. Papers*, vol. 1. Cambridge, 1967, pp. 58 – 61. **X-1.**

Pascal B. (1654) [Correspondance]. *Oeuvr. Compl.*, t. 1. Paris, 1998, pp. 145 – 166. **X-1.**

Pearson K. (1893), Asymptotical frequency curves. *Nature*, vol. 15, pp. 615 – 616.

--- (1905), Das Fehlergesetz und seine Verallgemeinerungen durch Fechner und Pearson. A rejoinder. *Biometrika*, vol. 4, pp. 169 – 212.

--- (1924), Historical note on the origin of the normal curve of errors. *Biometrika*, vol. 16, pp. 402 – 404.

--- (1925), James Bernoulli's theorem. *Biometrika*, vol. 17, pp. 201 – 210.

--- (1978), *History of Statistics in the 17th and 18th Centuries against the Changing Background of Intellectual, Scientific and Religious Thought.* Лекции 1921 – 1933 гг. Ред. E. S. Pearson. London.

Pearson K., Kendall M. G., ред. (1970), *Studies in the History of Statistics and Probability*. London. Сборник перепечаток.

Peirce B. (1873), On the errors of observations. Appendix 2 to *Report of the Superintendent of US Coast Survey*. Washington, pp. 200 – 224.

Petty W., Петти В. (1662), A Treatise on Taxes and Contributions. В книге автора *Economic Writings*, vol. 1. Cambridge, pp. 1 – 97. [London, 1997.] Трактат о налогах и сборах. В книге автора *Экономические и статистические работы.* М., 1940, с. 154 – 205.

Poincaré H., Пуанкаре А. (1896), *Calcul des probabilités*. Paris, 1912. Перепечатки 1923; Sceaux, 1987. *Теория вероятностей*. Ижевск, 1999.

--- (1921), Résumé analytique [собственных работ]. В книге Browder F. E., ред. (1980), *Mathematical Heritage of H. Poincaré*. Providence, Rhode Island, pp. 257 – 357.

Poisson S. D. (1824), Sur la probabilité des resultats moyens des observations, pt. 1. *Conn. des Temps* за 1827, pp. 273 – 302. **X-2.**

--- (1837), *Recherches sur la probabilité des jugements*. Paris. [Paris, 2003.] **X-1 (переведено аннотированное автором содержание).**

Polya G. (1920), Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem. *Math. Z.*, Bd. 8, pp. 171 – 181.

Quetelet A. (1846), *Lettres sur la théorie des probabilités*. Bruxelles.

Rabinovitch N. L. (1973), *Probability and Statistical Inference in Ancient and Medieval Jewish Literature*. Toronto.

Rümelin G. (1867), Über den Begriff eines sozialen Gesetzes. В книге автора *Reden und Aufsätze*. Freiburg i/B – Tübingen, 1875, pp. 1 – 31.

Schlözer A. L. (1804), *Theorie der Statistik.* Göttingen.

Seidel L. (1865), Über den […] Zusammenhang zwischen der Häufigkeit der Typhus-Erkrankungen und dem Stande des Grundwassers. *Z. Biol.*, Bd. 1, pp. 221 – 236.

--- (1866), Vergleichung der Schwankungen der Regenmengen mit der Schwankungen in der Häufigkeit des Typhus. *Z. Biol.*, Bd. 2, pp. 145 – 177.

Simpson J. Y. (1847 – 1848), Anaesthesia. *Works*, vol. 2. Edinburgh, 1871, pp. 1 – 288.

Simpson T. (1756), On the advantage of taking the mean of a number of observations. *Phil. Trans. Roy. Soc.*, vol 49, pp. 82 – 93. **X-1.**

--- (1757), То же название. В книге автора *Misc. Tracts on Some Curious* […] *Subjects*. London, pp. 64 – 75. **X-1.**

Snow J. (1855), On the mode of communication of cholera. В книге автора *On Cholera*. New York, 1965, pp. 1 – 139.

Stigler S. M. (1983), Who discovered Bayes's theorem? В книге автора *Statistics on the Table*. Cambridge, Mass. – London, 1999, pp. 291 – 301.

--- (1986), *The History of Statistics*. Cambridge, Mass. – London.

Todhunter I. (1865), *History of the Mathematical Theory of Probability.* New York, 1949, 1965.

Именной указатель