ПОСОБИЯ ДЛЯ ВЫСШЕЙ ШКОЛЫ

С. Н. БЕРНШТЕЙН

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

ГОСУДАРСТВЕННОЕ ИЗДАТЕЛЬСТВО

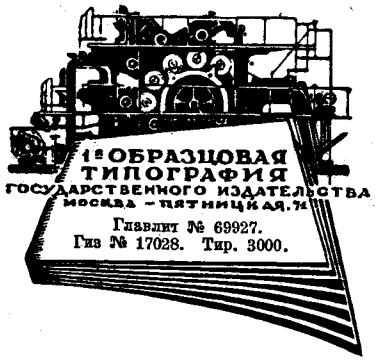
ПОСОБИЯ ДЛЯ ВЫСШЕЙ ШКОЛЫ

С. Н. БЕРНШТЕЙН

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

Научно-технической секцией Государственного Ученого Совета рекомендовано в качестве руководства для физматов ов и втузов

ГОСУДАРСТВЕННОЕ ИЗДАТЕЛЬСТВО
МОСКВА ★ 1927 ★ ЛЕНИНГРАД



Главлит № 69927.
Гиз № 17028. Тир. 3000.

Зак. № 1681.

ОГЛАВЛЕНИЕ

ЧАСТЬ ЧЕТВЕРТАЯ

Закон нормального распределения вероятностей

ЧАСТЬ ПЯТАЯ

Основы теории кривых и поверхностей распределения

Приложение

ПРЕДИСЛОВИЕ

Целью настоящей книги является изложение принципов и общих методов теории вероятностей, лежащих в основе ее многообразных приложений к физике, биологии, экономике и пр. Поскольку это было возможно без значительного ущерба для намеченной цели, я старался довести до минимума применяемый мною математический аппарат; лишь некоторые набранные мелким шрифтом места книги, пропуск которых не помешает пониманию ее в целом, предполагают у читателя сравнительно более углубленное знание высшего анализа, не выходящее однако за пределы минимальных программ математических факультетов. Первые три части книги вовсе не требуют знакомства с высшей математикой, за исключением V главы (Метод пределов и графическое изображение вероятностей) второй части, где впервые вводится и тут же разъясняется понятие интеграла, а также второй главы (Уточнение неравенств Чебышева) третьей части, которая менее подготовленным читателем может быть опущена при первом чтении. Таким образом усвоение основных идей и приемов исчисления вероятностей, в частности, закона больших чисел, которое должно быть достигнуто проработкой первой половины книги, не усложняется особыми трудностями непривычной математической техники. Но приступая к последним двум частям, наиболее важным для приложений, читателю необходимо будет предварительно пройти сокращенный вводный курс математического анализа, без которого вообще нельзя обойтись ни статистику, ни биологу.

Считаю приятным долгом выразить признательность Научному отделу Госиздата, в лице его уважаемого руководителя

проф. В. Ф. Кагана, которому принадлежит инициатива издания этой книги, за чрезвычайно внимательное отношение при ее печатании и ценные указания редакционного характера. Кроме того выражаю искреннюю благодарность аспиранту Л. Я. Гиршвальду за помощь при исправлении корректур и проверку некоторых вычислений.

С. БЕРНШТЕЙН.

20/II—1927 г.

ЧАСТЬ ПЕРВАЯ
ОСНОВОПОЛОЖЕНИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ.

ГЛАВА ПЕРВАЯ
ОСНОВНЫЕ СВОЙСТВА И ОПРЕДЕЛЕНИЕ МАТЕМАТИЧЕСКОЙ ВЕРОЯТНОСТИ

1. Наиболее обычная и важная теоретическая схема, которой мы пользуемся для познания внешнего мира и для предвидения отдельных фактов, состоит в том, что на основании предшествующего опыта утверждается достоверность наступления события известного класса A, если осуществлен некоторый определенный комплекс условий α, каковы бы ни были прочие обстоятельства. Поскольку в данном конкретном опыте соблюдены условия α, наступление факта A неизбежно.

В дальнейшем нам придется еще рассмотреть в свете теории вероятностей, на чем покоится и насколько обоснована эта уверенность, и мы увидим, что абсолютной достоверностью наши предвидения о ходе реальных явлений никогда не могут обладать; но пока нам нет надобности на этом останавливаться. Заметим лишь, что предполагаемый нами закон, связывающий комплекс α с фактом A, только тогда будет практически полезен, если совокупность условий α не слишком сложна и не слишком трудно поддается наблюдению; если, напротив, неуловимое изменение условий опыта влечет за собой противоположный результат, если, даже при самой совершенной технике, время, нужное для исследования того, соблюдены ли условия α или нет, превышает продолжительность опыта, — то такой закон оказывается непригодным для предвидения факта A, который мы называем тогда случайным. Хотя опыты этого рода мало распространены в лабораториях, но в практической жизни почти ни один наш поступок не мог бы быть совершон, если бы мы хотели знать с безусловной достоверностью, что соблюден весь комплекс условий, обеспечивающий наступление нужного нам факта: мы обычно не выскакиваем на улицу из окна, но предпочитаем спуститься по лестнице, хотя и здесь можно упасть

и разбиться на-смерть, между тем как прыжок из окна не всегда смертелен. Поступая так или иначе, мы руководствуемся не уверенностью в наступлении определенного результата, а сознательной или инстинктивной оценкой его вероятности. Таким образом, говоря, что некоторый факт A случаен, т.-е. не может быть предсказан с безусловной уверенностью, мы на этом не успокаиваемся, а стремимся ввести вместо слишком усложнившегося комплекса α, из которого A вытекает всегда, такой более простой комплекс условий β, при наличии которого в данном опыте наступление факта A имеет определенную вероятность. При этом, пока речь идет только об отдельном конкретном опыте, особенно важны те случаи, когда, как в вышеуказанном примере, наступлению рассматриваемого события приписывается очень большая или очень малая вероятность, подразумевая под этим, что факты очень вероятные происходят почти всегда, а маловероятные — почти никогда.

Теория вероятностей, дающая правила для вычисления вероятностей одних событий, когда известны вероятности некоторых других, имеет одной из главных своих целей установление таких фактов, вероятности которых очень велики; поэтому в том обстоятельстве, что эти последние факты в действительности почти всегда осуществляются, можно видеть экспериментальное подтверждение объективной правильности общих предпосылок, лежащих в основе теории вероятностей. Вместе с тем и всякая специальная схема теории вероятностей, служащая для описания или объяснения определенной группы явлений, может считаться объективно соответствующей действительности тогда и только тогда, когда факты, теоретически обладающие очень большой вероятностью, на самом деле осуществляются почти всегда.

2. Основное допущение теории вероятностей (постулат существования математической вероятности) состоит в том, что существуют такие комплексы условий β, которые (теоретически по крайней мере) могут быть реализованы неограниченное число раз, при наличности которых в данном опыте наступление факта A имеет определенную вероятность, выражающуюся математическим числом. Таким образом, если факт B также обладает при соответствующих условиях определенной вероятностью, то всегда имеет место одно из трех соотношений:

$$вер.\ A = вер.\ B;\ вер.\ A > вер.\ B;\ вер.\ A < вер.\ B.$$

Так же, как и в случае закономерных связей, о которых мы говорили вначале, опыт имеет решающий голос в вопросе о том, возможно ли при осуществлении данного комплекса условий β и полной неопределенности прочих обстоятельств приписать факту A определенную вероятность. Одной из важнейших целей прикладной теории вероятностей, или статистики, именно и является установление a posteriori, при помощи применения соответствующих математических теорем, совместимо ли с наблюденными статистическими данными допущение, что в ряде опытов, где были осуществлены известные одинаковые условия β, факт A имел определенную постоянную вероятность. Таким образом можно, например, считать довольно точно проверенным экспериментально, что при бросании с достаточной высоты симметрично сделанной игральной кости появление на верхней ее грани определенного числа очков имеет во всех опытах ту же самую вероятность. С более специальными статистическими примерами этого рода мы встретимся в дальнейшем, и на них нам пока незачем останавливаться. Но следует заметить, что во многих простых опытах — между прочим в обычных играх в карты, кости, лото и т. д. — мы без всяких статистических наблюдений усматриваем a priori, что некоторые результаты A и B данного опыта следует признать равновероятными (равновозможными), т.-е. имеющими равные вероятности. Основанием для этого суждения служит в большинстве случаев симметрия, а именно допущение, что при самом внимательном анализе фактических условий каждого опыта нельзя указать ни одного обстоятельства, которое нужно было бы изменить в постановке опыта, если бы мы взаимно заменили названия фактов A и B. Например, тщательным и добросовестным перемешиванием карт полной колоды мы достигаем — или по крайней мере одинаково перед каждой игрой стремимся достигнуть — полной независимости в расположении карт от их значения или масти, так, чтобы объективно учитываемые условия предстоящей игры не изменились от того, назовем ли мы козырем (т.-е. привилегированной в каком-нибудь отношении мастью) бубновую или какую-нибудь другую масть; благодаря указанной симметрии при вынимании карты из колоды считают одинаково вероятным появление бубновой масти и какой-нибудь другой, например трефовой. Точно так же, если в урне имеется несколько перенумерованных шаров, не различимых на ощупь и тщательно перемешанных, то физические условия опыта вынимания шара сим-

метричны по отношению к любым номерам, и появлению шара с одним номером мы приписываем ту же вероятность, что и появлению шара со всяким другим.

Указанные примеры, по отношению к которым все следствия теории можно считать экспериментально проверенными, являются типичными образцами опытов, при которых соответствующие факты имеют определенную вероятность. Таким образом утверждение, что некоторый факт A (при опытах, где соблюден определенный комплекс условий β) имеет известную вероятность, а именно ту же самую, что и появление данного шара в определенном опыте с урной вышеуказанного образца, означает, что объективные условия для возникновения в рассматриваемых опытах факта A равноценны по существу условиям, осуществляемым в соответствующем схематическом опыте с шарами.

Например, если мы говорим, что при некоторых определенного рода опытах вероятность наступления события A равна вероятности извлечения данного шара из урны, в которой находятся 100 одинаковых шаров, то мы этим желаем выразить, что в каждом отдельном опыте мы так же мало можем рассчитывать на наступление факта A, как на появление указанного шара из рассматриваемой урны; в частности, при многократном повторении опыта нет никаких объективных оснований ожидать, что факт A будет появляться чаще (или реже), чем данный шар при повторении извлечения из нашей урны (полагая, что вынутый шар всегда кладется обратно и тщательно смешивается с остальными).

3. В дальнейшем, опыты с шарами и картами будут служить нам иллюстрациями для пояснения общих правил, которыми нужно руководиться при установлении соотношений между вероятностями одних фактов, когда известны определенные соотношения между вероятностями некоторых других связанных с ними фактов. Правила эти, из которых дедуктивным путем выводятся все положения теории вероятностей, сводятся к трем аксиомам: 1) аксиома сравнения вероятностей, 2) аксиома о несовместимых событиях, 3) аксиома о совмещении событий.

Следует заметить, что первые две аксиомы, говоря о вероятности какого-нибудь факта A, всегда имеют в виду существование некоторого неизменного комплекса условий β опыта; напротив, последняя аксиома (о совмещении) связывает вероятность факта A при одних данных условиях α с вероятностью того же факта при

другом соответствующем комплексе условий β. Для того, чтобы лучше уяснить самостоятельное значение последней аксиомы, мы отложим ее формулировку до второй главы, ограничившись пока рассмотрением первых двух аксиом и простых, но важных следствий, которые вытекают из них.

Введем предварительно несколько основных понятий и терминов, которыми нам неоднократно придется пользоваться. Положим, что из колоды карт вынимается одна карта, и интересующее нас событие A заключается в появлении туза безразлично какой масти; в таком случае каждый из фактов a_1, a_2, a_3, a_4 появления туза соответственно бубновой, трефовой, пиковой и червонной масти называется частным случаем события A. Таким образом всякий частный случай a события A характеризуется тем, что наступление a означает также и наступление A. Понятно, что появление A_1 туза черных мастей a_2 или a_3 также является частным случаем появления A туза вообще.

Часто бывает полезно делать различие между частными случаями в узком и широком смысле слова: если мы знаем, например, что в колоде кроме тузов черных мастей есть и красные тузы, так что событие A возможно и без наступления его частного случая A_1, то A_1 называется частным случаем A в узком смысле слова (или, короче, видом события A); напротив, если бы наличность красных тузов нам не была известна, т.-е. если бы мы не были уверены в возможности появления A помимо A_1, мы ограничились бы утверждением, что A_1 есть частный случай A (в широком смысле слова), не исключая, таким образом, допущения, что A_1 вполне тождественно с событием A. Итак, утверждая, что a частный случай в узком смысле слова (или вид) события A, мы выражаем не только то, что A несомненно осуществляется, коль скоро наступило a, но также и то, что осуществление A возможно и без появления a.

Очевидно, что, применяя вышеуказанным образом термин вид, мы должны в соответствии с нашим интуитивным представлением о „вероятности" принять, что всякое событие A всегда более вероятно, чем некоторый его вид (частный случай в узком смысле слова) a, т.-е. *вер.* $A >$ *вер.* a. Естественно поэтому принять следующую аксиому:

Аксиома сравнения вероятностей. Если a есть вид (частный случай в узком смысле слова) события A, то *вер.* $a <$ *вер.* A;

обратно, если между вероятностями фактов a_1 и A существует неравенство *вер.* $a_1 <$ *вер.* A, то оно означает, что *вер.* $a_1 =$ *вер.* a, где a есть некоторый вид события A.

Согласно первой части нашей аксиомы мы принимаем, например, что в некоторых определенных условиях вероятность рождения ребенка вообще больше, чем вероятность рождения мальчика.

Из первой части этой аксиомы вытекают два следствия, которые и непосредственно очевидны:

1. Если факт A достоверен (т.-е. необходимо должен наступить), а факт B только возможен, то наша уверенность в наступлении A больше, чем в наступлении B (*вер.* $A >$ *вер.* B), ибо B является лишь одним из видов осуществления A.

2. Если факт B возможен, а событие C невозможно, то мы больше рассчитываем на появление B, чем на появление C (*вер.* $B >$ *вер.* C), так как событие B может произойти и без C.

Иначе говоря, достоверность есть максимальная вероятность, а невозможность есть минимальная вероятность. Таким образом все достоверные факты имеют одну и ту же максимальную, т.-е. наибольшую вероятность, а невозможные факты имеют минимальную, т.-е. наименьшую вероятность.

Поясним теперь вторую часть высказанной аксиомы. Когда мы говорим, что вероятность смерти в течение года больше, чем вероятность смерти в течение месяца, то последний факт есть частный случай первого. Но что выражает утверждение, что вероятность смерти A в течение года для лиц *одной категории* (например для стариков 70 лет) больше, чем вероятность смерти a_1 в течение года для лиц другой категории (например для юношей 20 лет)? Очевидно, что здесь не может быть речи о том, что a_1 является частным случаем A; но если мы введем частный случай a события A, заключающийся в смерти старика в течение некоторого более или менее короткого промежутка времени, то, согласно первой части аксиомы, по мере уменьшения рассматриваемого промежутка времени уменьшается и вероятность факта a: в согласии со второй частью аксиомы утверждение, что *вер.* $A >$ *вер.* a_1 означает существование такого срока (примерно три недели), для

которого *вер.* $a =$ *вер.* a_1, т.-е. вероятность старику умереть в течение некоторого определенного срока, меньшего, чем год, та же, что вероятность юноше 20 лет умереть в продолжение года.

Разумеется, фактическое указание такого вида a события A, что *вер.* $a =$ *вер.* a_1, может представлять более или менее значительные затруднения, но тем не менее теоретическая возможность этого принципиально необходима для того, чтобы утверждение *вер.* $A >$ *вер.* a_1 имело смысл.

4. Переходим к следующей аксиоме.

Аксиома о несовместимых событиях. Если известно, что события A и A_1 несовместимы между собой, и, с другой стороны, события B и B_1 также между собой несовместимы, при чем *вер.* $A =$ *вер.* B и *вер.* $A_1 =$ *вер.* B_1, то вероятность факта C, заключающегося в наступлении события A или события A_1, равна вероятности факта C_1, заключающегося в наступлении B или B_1, т.-е. *вер.* (A или A_1) $=$ *вер.* (B или B_1).

Понятие несовместимости двух фактов A и A_1 вряд ли нуждается в особом пояснении: оно означает, что наступление в данном опыте одного из них невозможно в случае наступления другого.

Высказанную аксиому можно формулировать еще иначе: вероятность наступления какого-нибудь одного из двух несовместимых событий A или A_1 есть величина, которая вполне определяется вероятностями каждого из них в отдельности, т.-е. представляет собой определенную функцию *вер.* A и *вер.* A_1, вовсе не зависящую от природы самих фактов A и A_1; иными словами,

$$\text{вер. } (A \text{ или } A_1) = f\,(\text{вер. } A,\ \text{вер. } A_1). \qquad (1)$$

Например, если по тем или иным основаниям мы знаем, что для лица данной категории, вступающего в брак, вероятность овдоветь в течение трех лет равна вероятности, что при вынимании шара из данной урны появится белый шар, а вероятность овдоветь после трех лет равна вероятности появления черного шара, то вероятность вообще овдоветь для лица указанной группы равна вероятности появления из урны шара белого либо черного цвета.

Высказанную аксиому нетрудно распространить на какое угодно число несовместимых событий. А именно:

Следствие 1. Если A_1, A_2, $A_3, \ldots$, A_n между собой несовместимы и факты B_1, B_2, $B_3, \ldots$, B_n также

между собой несовместимы, при чем *вер.* $A_1 =$ *вер.* B_1, *вер.* $A_2 =$ *вер.* $B_2, \ldots$, *вер.* $A_n =$ *вер.* B_n, то *вер.* (A_1 или $A_2 \ldots$ или A_n) $=$ *вер.* (B_1 или $B_2 \ldots$ или B_n).

В самом деле, наше утверждение для $n = 2$ соответствует выше формулированной аксиоме; поэтому для доказательства его в общем виде достаточно показать, что если утверждение справедливо для некоторого значения n, то оно верно также и для числа $n+1$. Итак, дано, что A_{n+1} несовместимо ни с одним из событий $A_1, A_2, \ldots, A_n$; следовательно A_{n+1} несовместимо также и с наступлением факта α, состоящего в появлении A_1 или $A_2 \ldots$ или A_n, точно так же B_{n+1} несовместимо с фактом β, состоящим в наступлении B_1 или $B_2 \ldots$ или B_n. Но мы полагаем уже известным, что *вер.* (A_1 или $A_2 \ldots$ или A_n) $=$ *вер.* (B_1 или $B_2 \ldots$ или B_n), т.-е. *вер.* $\alpha =$ *вер.* β, и кроме того, по условию, *вер.* $A_{n+1} =$ *вер.* B_{n+1}. Следовательно, в силу нашей аксиомы

$$\text{вер. } (\alpha \text{ или } A_{n+1}) = \text{вер. } (\beta \text{ или } B_{n+1}),$$

т.-е. *вер.* (A_1 или $A_2 \ldots$ или A_{n+1}) $=$ *вер.* (B_1 или $B_2 \ldots B_{n+1}$), что и требовалось доказать.

С другой стороны, сопоставляя обе аксиомы, получаем

Следствие 2. Если A и A_1 несовместимы между собой, а B и B_1 несовместимы между собой, при чем *вер.* $A =$ *вер.* B и *вер.* $A_1 >$ *вер.* B_1, то *вер.* (A или A_1) $>$ *вер.* (B или B_1).

В самом деле, согласно аксиоме сравнения вероятностей, можно указать такой вид a_1 факта A_1, что *вер.* $a_1 =$ *вер.* B_1. В таком случае наступление A или a_1 будет частным случаем в узком смысле слова факта (A или A_1), и следовательно *вер.* (A или A_1) $>$ $>$ *вер.* (A или a_1) $=$ *вер.* (B или B_1).

Следствие второе, очевидно, равнозначно утверждению, что функция f (*вер.* A, *вер.* A_1), которая стоит в формуле (1), возрастает при возрастании одной из двух вероятностей, от которых она зависит, если другая из них остается неизменной. Тем более, конечно, функция f возрастает, если обе вероятности увеличиваются. Таким образом получаем:

Следствие 3. Если (при сохранении прежних условий несовместимости) *вер.* $A >$ *вер.* B и *вер.* $A_1 >$ *вер.* B_1, то *вер.* (A или A_1) $>$ *вер.* (B или B_1).

Применяя затем тот же способ математической индукции, которым мы воспользовались для доказательства следствия (1), нетрудно

распространить следствия (2) и (3) на произвольное число фактов, и получаем

Следствие 4. Если события A_1, A_2, ..., A_n между собой несовместимы, а, с другой стороны, события B_1, B_2, ..., B_n несовместимы между собой, при чем *вер.* $A_1 \geqslant$ *вер.* B_1, ..., *вер.* $A_{n-1} \geqslant$ *вер.* B_{n-1}, *вер.* $A_n >$ *вер.* B_n, то *вер.* (A_1 или A_2... или A_n) $>$ *вер.* (B_1 или B_2... или B_n).

Предположим теперь, в частности, что рассматриваемые нами несовместимые между собой случаи A_1, A_2, ..., A_n имеют равные вероятности и, кроме того, единственно возможны, т.-е. один из фактов A_1, A_2, ..., A_n непременно должен наступить; предположим также, что и события B_1, B_2, ..., B_n несовместимы между собой, равновероятны и единственно возможны. Нетрудно понять, что при таких условиях вероятность каждого из событий A должна быть равна вероятности события B; действительно, если бы этого не было, т.-е. если бы мы допустили, например, что *вер.* $A_1 >$ *вер.* B_1, то мы имели бы также *вер.* $A_2 >$ *вер.* B_2, ..., *вер.* $A_n >$ *вер.* B_n, и благодаря следствию (4) пришлось бы признать, что *вер.* (A_1 или A_2... или A_n) $>$ *вер.* (B_1 или B_2... или B_n), что невозможно, так как наступление A_1 или A_2... или A_n достоверно, равно как и наступление B_1 или B_2... или B_n. Аналогичным образом убеждаемся в недопустимости предположения, что *вер.* $A_1 <$ *вер.* B_1; а потому, какова бы ни была физическая природа фактов A и фактов B, мы вынуждены им приписать одну и ту же вероятность.

Принимая во внимание следствие (1), мы выводим отсюда

Следствие 5. Если событие X происходит в некотором опыте, допускающем n равновероятных, несовместимых и единственно возможных исходов A_1, A_2, ..., A_n, тогда и только тогда, когда наступает какое-нибудь из m ($m < n$) определенных исходов A_1, A_2, ..., A_m (иначе говоря, если факту X благоприятствуют только случаи A_1, A_2, ..., A_m); если, с другой стороны, и событию Y благоприятствуют m определенных случаев B_1, ..., B_m из каких-нибудь n равновероятных, несовместимых и единственно возможных случаев B_1, B_2, ..., B_n, — то событие X и событие Y имеют равные вероятности.

Последнему утверждению, очевидно, можно дать и такую формулировку:

событию X благоприятствуют m случаев из общего числа всех n единственно возможных несовместимых и равновероятных случаев, то вероятность события X зависит только от чисел m и n (а не от природы рассматриваемого опыта), т.-е.

$$вер.\ X = F(m, n), \tag{2}$$

где $F(m, n)$ есть некоторая определенная функция, которая должна быть установлена раз навсегда.

Какова же эта функция $F(m, n)$? Можем ли мы ее выбрать по своему усмотрению, или же, напротив, наш выбор в той или иной мере предопределен принятыми нами аксиомами? Ответ на этот вопрос дает следующая основная теорема.

5. *Теорема.* Если событию X благоприятствуют m случаев из общего числа n несовместимых, равновероятных и единственно возможных случаев $A_1, A_2, \ldots, A_n$, а событию Y благоприятствуют m_1 случаев из n_1 несовместимых, равновероятных и единственно возможных случаев $B_1, B_2, \ldots, B_{n_1}$, то *вер.* $X =$ *вер.* Y, если

$$\frac{m}{n} = \frac{m_1}{n_1};$$

напротив, *вер.* $X >$ *вер.* Y, если

$$\frac{m}{n} > \frac{m_1}{n_1}.$$

В самом деле, предположим сначала, что $\frac{m}{n} = \frac{m_1}{n_1}$; общее значение этих обеих дробей можно представить в виде несократимой дроби $\frac{M}{N}$. Иными словами, обозначая соответственно через k и k_1 множители, на которые нужно сократить дроби $\frac{m}{n}$ и $\frac{m_1}{n_1}$, чтобы привести их к несократимому виду $\frac{M}{N}$, можем написать:

$$m = kM, \quad n = kN, \quad m_1 = k_1 M, \quad n_1 = k_1 N.$$

Располагая все случаи $A_1, A_2, \ldots, A_n$ в виде таблицы, содержащей k строк и N столбцов ($kN = n$)

$$\left|\begin{array}{cccc} \overbrace{\qquad\qquad}^{M} & & \overbrace{\qquad\qquad}^{N-M} & \\ A_1 & A_{k+1} \ldots\ldots & A_{m+1} & \ldots\ldots \\ A_2 & A_{k+2} \ldots\ldots & A_{m+2} & \ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ A_k & A_{2k} \ldots\ldots A_m & A_{m+k} & \ldots A_n \end{array}\right|$$

можем поместить в первых M столбцах все $m = kM$ благоприятные событию X случаи, а в остальных столбцах тогда будут все неблагоприятные случаи.

Таким образом, если появление любого из случаев 1-го столбца назовем событием a_1, появление любого из случаев 2-го столбца назовем a_2 и т. д., то мы будем иметь всего N несовместимых и единственно возможных событий, $a_1, a_2, \ldots, a_M, \ldots, a_N$, которые, на основании следствия (1), будут равновероятны, при чем первые M из них благоприятствуют нашему факту X. Аналогичным образом убеждаемся в том, что и событие Y можно рассматривать как наступающее при одном из M случаев из общего числа соответствующих N единственно возможных, несовместимых и равновероятных случаев $b_1, b_2, \ldots, b_N$, которые получаем, соединяя, подобно предыдущему, случаи $B_1, \ldots, Bn_1$, в N групп по k_1 случаев в каждой группе. Следовательно, согласно следствию (5),

$$вер.\ X = вер.\ Y,$$

и первая часть нашей теоремы таким образом доказана.

Предположим далее, что $\frac{m}{n} > \frac{m_1}{n_1}$. Ввиду того, что на основании только что полученного вывода вероятность события X и события Y не изменяется от того, что числа m и n будут умножены на одно и то же целое произвольное число h, а числа m_1 и n_1 будут помножены на какое-нибудь другое число h_1, мы можем, не нарушая общности, предположить, что наши дроби $\frac{m}{n}$ и $\frac{m_1}{n_1}$ имеют равных знаменателей, т.-е. $n = n_1$, а $m > m_1$. В таком случае вид X_1 события X, который состоит в появлении

только одного из m_1 случаев, благоприятствующих событию X, будет иметь ту же вероятность, что событие Y, а потому, согласно аксиоме сравнения вероятностей, *вер.* $X >$ *вер.* Y, что и требовалось доказать.

Таким образом мы видим, что вероятность события X зависит только от отношения $\frac{m}{n}$ числа благоприятных этому событию случаев к общему числу всех единственно возможных, несовместимых и равновероятных случаев, при чем с возрастанием отношения $\frac{m}{n}$ возрастает и вероятность события X. Другими словами функция, фигурирующая в формуле (2), в действительности зависит только от одной переменной $\frac{m}{n}$ и возрастает одновременно с этой последней, т.-е. можем написать, что

$$\text{вер. } X = F\left(\frac{m}{n}\right), \tag{3}$$

где $F\left(\frac{m}{n}\right)$ есть возрастающая функция дроби $\frac{m}{n}$.

Нетрудно убедиться, с другой стороны, что всякая возрастающая функция $F\left(\frac{m}{n}\right)$ действительно удовлетворяет обеим принятым нами аксиомам.

В самом деле, если X_1 есть вид события X, то число случаев m_1, благоприятствующих X_1, представляет лишь часть числа m случаев, которые благоприятствуют событию X при том же числе n всех возможных случаев; поэтому $\frac{m}{n} > \frac{m_1}{n}$, и аксиома сравнения удовлетворяется благодаря тому, что $F\left(\frac{m}{n}\right) > F\left(\frac{m_1}{n}\right)$. С другой стороны, аксиома о несовместимых событиях тоже удовлетворена. Действительно, если A и A_1 — два несовместимых события, то благоприятствующие им соответственно m и m_1 случаев различны между собой, так что факту a_1, состоящему в наступлении A или A_1, благоприятствуют $m + m_1$ случаев из общего числа n несовместимых, равновероятных и единственно возможных случаев; следовательно

$$\text{вер. } (A \text{ или } A_1) = F\left(\frac{m + m_1}{n}\right).$$

Точно так же, если двум другим несовместимым между собой событиям B и B_1 благоприятствуют соответственно m' и m_1' случаев из общего числа

некоторых n' несовместимых, равновероятных и единственно возможных случаев, то

$$вер.\ (B \text{ или } B_1) = F\left(\frac{m' + m_1'}{n'}\right)$$

Таким образом, замечая, что по доказанной теореме условия

$$вер.\ A = вер.\ B,\ вер.\ A_1 = вер.\ B_1$$

означают, что

$$\frac{m}{n} = \frac{m'}{n'},\quad \frac{m_1}{n} = \frac{m_1'}{n'},$$

заключаем, что

$$\frac{m + m_1}{n} = \frac{m' + m_1'}{n'},$$

а потому

$$вер.\ (A \text{ или } A_1) = F\left(\frac{m + m_1}{n}\right) = F\left(\frac{m' + m_1'}{n'}\right) = вер.\ (B \text{ или } B_1),$$

т.-е. вторая аксиома также удовлетворена.

6. Только что полученный вывод показывает нам, что на основании принятых допущений мы должны назвать вероятностью события A, которому благоприятствуют m случаев из общего числа n всех единственно возможных, несовместимых и равновероятных случаев, некоторую определенную раз навсегда, возрастающую функцию $F\left(\frac{m}{n}\right)$. Таким образом мы с одинаковым правом могли бы, в согласии с нашими аксиомами, условиться называть математической вероятностью как само отношение $\frac{m}{n}$, так и $c\frac{m}{n}$, где c — определенная постоянная, или, например, выражение

$$\frac{m}{n-m} = \frac{\frac{m}{n}}{1 - \frac{m}{n}}$$

(т.-е. отношение числа благоприятных случаев к числу неблагоприятных)[1]), или еще $\log \frac{cm}{n}$ и т. д. Напротив, мы не можем, как было доказано выше, назвать вероятностью, например, отношение $\frac{m}{n^2}$, если не хотим нарушить принятых аксиом, диктуемых нам повседневным опытом и здравым смыслом.

---

[1]) При последнем соглашении невозможное событие имело бы вероятность равную 0, а вероятность достоверного события была бы бесконечна; вероятностью события могло бы быть любое положительное число.

Выбор той или иной возрастающей $F\left(\frac{m}{n}\right)$ всецело зависит от нашего усмотрения и определяется исключительно техническими удобствами. Выводы теории вероятностей совершенно не зависят по существу от этого выбора и сохранят свою силу так же, как законы физики, которые не нарушаются от того, назовем ли мы температурой абсолютную температуру или какую-нибудь определенную функцию ее, как, например, число градусов по Реомюру или Фаренгейту. Изменяется лишь определенным образом словесная формулировка законов и теорем и численные значения интересующих нас величин (так, например, при $m=\frac{n}{2}$, т.-е. если событию A благоприятствует половина всех равновозможных случаев, то вероятность будет равна $\frac{1}{2}$ или 1, в зависимости от того, условимся ли мы называть вероятностью дробь $\frac{m}{n}$ или же $\frac{m}{n-m}$); во всяком случае достаточно будет указать, какова выбранная нами функция $F\left(\frac{m}{n}\right)$, чтобы без всяких затруднений от одной системы теорем и значений вероятностей перейти к другой, подобно тому, как при помощи идеально точного словаря каждое предложение можно перевести с одного языка на другой, не нарушая его смысла.

Следуя общепринятому соглашению, приводящему к наиболее простой формулировке большинства теорем теории вероятностей, мы условимся раз навсегда называть математической вероятностью события A отношение $\frac{m}{n}$ числа m случаев, благоприятствующих A, к числу n всех единственно и равновозможных несовместимых случаев, т.-е. остановим наш выбор на функции $F\left(\frac{m}{n}\right)=\frac{m}{n}$.

Таким образом, в частности, невозможное событие имеет вероятность 0, а вероятность достоверного события равна $\frac{n}{n}=1$. Всякое возможное, но не достоверное событие имеет вероятность, заключенную между 0 и 1.

Примеры. 1) При бросании игральной кости вероятность выпадения четного числа очков равна $\frac{3}{6}=\frac{1}{2}$, если мы считаем появления всех граней равновероятными.

2) Если в урне 20 одинаковых шаров, отличающихся только цветом, из коих 4 шара красных, 6 голубых и 10 белых, и мы считаем, что каждый шар имеет одинаковую вероятность быть вынутым, то вероятность вынуть красный шар равна $\frac{4}{20}=\frac{1}{5}$, вероятность вынуть голубой шар равна $\frac{6}{20}=\frac{3}{10}$, а вероятность вынуть вообще цветной шар равна $\frac{4+6}{20}=\frac{10}{20}=\frac{1}{2}$, т.-е. столь же велика, как и вероятность вынуть белый шар [1]). Очевидно, неправильно было бы сказать, что вероятность появления красного шара равна $\frac{1}{3}$, основываясь на том, что мы имеем единственно возможных 3 исхода опыта (красный, голубой или белый шар), так как следует принять во внимание, что ввиду различного числа шаров каждого цвета указанные 3 случая нельзя считать равновероятными.

7. Сделанное нами техническое соглашение относительно величины, которую мы, в согласии с принятыми аксиомами, условились называть математической вероятностью события A, когда ему в известном опыте благоприятствуют m из n единственно и равновозможных несовместимых случаев, позволяет нам дать определенное выражение функции f в формуле (1), при помощи которой, на основании аксиомы о несовместимых событиях, должен быть указан совершенно общий прием для вычисления вероятности наступления любого из двух несовместимых событий A и B, если вероятности каждого из них в отдельности известны. А именно, легко доказать следующую теорему, известную под названием теоремы сложения вероятностей:

Теорема. Если событие A состоит в наступлении одного из двух несовместимых фактов α или β, при чем *вер.* $\alpha=p$, *вер.* $\beta=p_1$, то

$$\textit{вер. } A=p+p_1.$$

В самом деле, предположим, что p и p_1 — рациональные числа (т.-е. правильные дроби), так что, по приведении их к одному знаменателю, можем написать

$$p=\frac{m}{n},\quad p_1=\frac{m_1}{n},$$

---

[1]) Заметим, что утверждение равенства этих двух вероятностей, как и вообще все утверждения о равенствах и неравенствах вероятностей, вовсе не зависят от принятого нами определения математической вероятности.

где m, m_1, n—целые числа. Допустим, что $m+m_1 \leqslant n$; мы увидим дальше, что противоположное предположение: $m+m_1 > n$, соответствующее $p+p_1 > 1$, недопустимо. Возьмем урну с n одинаковыми шарами, так чтобы появление каждого из них было равновероятным; в таком случае, если m шаров будет красных, а m_1 — голубых, то событие a появления красного шара будет иметь ту же вероятность $\frac{m}{n}=p$, что и наше событие α; точно так же событие b, состоящее в появлении шара голубого цвета, будет иметь ту же вероятность $p_1=\frac{m_1}{n}$, что и данное событие β. Таким образом, согласно аксиоме о несовместимых событиях,

$$вер.\ A = вер.\ (\alpha \text{ или } \beta) = вер.\ (a \text{ или } b) = \frac{m+m_1}{n} = \frac{m}{n} + \frac{m_1}{n} = \\ = p + p_1,$$

так как событие (a или b) есть появление красного или синего шара, которому благоприятствуют $m+m_1$ шаров из общего числа n равновозможных шаров.

Из нашего рассуждения видно, что если $m+m_1=n$, то

$$вер.\ A = \frac{m+m_1}{n} = 1, \text{ т.-е. } A \text{ достоверно.}$$

Таким образом, если два несовместимых события единственно возможны (иначе говоря, если второе событие соответствует отсутствию, т.-е. ненаступлению первого, и наоборот), то сумма их вероятностей равна единице.

Отсюда ясно, что если два события α и β несовместимы, то сумма их вероятностей $p+p_1$ не может быть больше 1, так как в противном случае, согласно следствию (2), вероятность наступления α или β была бы больше единицы, т.-е. больше достоверности.

Во многих случаях, как мы увидим дальше, события могут иметь определенные вероятности, хотя нельзя указать непосредственно всех равновозможных исходов соответствующего опыта; при этом теоретически вероятность может оказаться иррациональным числом.

Поэтому для полноты доказательства теоремы следует рассмотреть случай, когда одно или оба числа p и p_1 иррациональны. Пусть

$$p = \underset{n=\infty}{\text{пред.}}\ \rho_n = \underset{n=\infty}{\text{пред.}}\ \rho_n',$$

где ρ_n и ρ_n' правильные дроби, при чем $\rho_n > p > \rho_n'$, и, с другой стороны, пусть

$$p_1 = \underset{n=\infty}{\text{пред.}}\ \sigma_n = \underset{n=\infty}{\text{пред.}}\ \sigma_n',$$

где σ_n и σ_n' также рациональные числа, при чем $\sigma_n > p_1 > \sigma_n'$.

Согласно сделанному выше замечанию достаточно ограничиться предположением, что $p + p_1 < 1$; следовательно, при n достаточно большом имеем также $\rho_n + \sigma_n < 1$ (и тем более $\rho_n' + \sigma_n' < 1$). Поэтому можем построить два несовместимых факта α_n и β_n, имеющих соответственно вероятностями рациональные числа ρ_n и σ_n; в таком случае, по доказанному, вероятностью (α_n или β_n) будет сумма $\rho_n + \sigma_n$, которая на основании следствия (3) должна быть больше, чем *вер.* A; таким же образом убеждаемся, что $\rho_n' + \sigma_n'$ меньше, чем *вер.* A; но так как $\rho_n + \sigma_n$, так же, как и $\rho_n' + \sigma_n'$, стремится с увеличением n к тому же самому пределу $p + p_1$, то *вер.* A не может быть равна никакому иному числу, кроме $p + p_1$.

8. Распространим теорему сложения на произвольное число событий [1]).

Если событие A заключается в появлении любого из несовместимых фактов $\alpha_1, \alpha_2, \ldots, \alpha_k$, имеющих соответственно вероятности $p_1, p_2, \ldots, p_k$, то

$$\text{вер. } A = p_1 + p_2 + \ldots + p_k. \tag{4}$$

Действительно, в случае $k = 2$ наше утверждение уже содержится в доказанной теореме; следовательно, достаточно показать, что если оно правильно для данного значения $k = n$, то оно справедливо и для $n + 1$. С этой целью замечаем, что событие A, состоящее в появлении одного из $n + 1$ фактов $\alpha_1, \ldots, \alpha_n, \alpha_{n+1}$, можно рассматривать как состоящее в появлении одного из двух фактов: (α_1 или $\alpha_2 \ldots$ или α_n или α_{n+1}). Применяя теорему сложения вероятностей к последним двум фактам, находим, что *вер.* $A =$ *вер.* (α_1 или $\alpha_2 \ldots$ или α_n) + *вер.* $\alpha_{n+1} = (p_1 + p_2 + \ldots + p_n) + p_{n+1}$;

---

[1]) Нетрудно проверить, что если бы мы условились называть вероятностью вместо $\frac{m}{n}$ некоторую определенную возрастающую функцию $F\left(\frac{m}{n}\right)$, то формулу (4) пришлось бы заменить такою:

$$\text{вер. } A = F(\varphi(p_1) + \varphi(p_2) + \ldots + \varphi(p_k)),$$

где φ есть функция обратная F: так, например, если $F\left(\frac{m}{n}\right) = \frac{m^2}{n^2}$, то мы имели бы:

$$\text{вер. } A = (\sqrt{p_1} + \sqrt{p_2} + \ldots + \sqrt{p_k})^2.$$

следовательно, если для $k=n$ правильность нашего утверждения была уже установлена, то оно справедливо и для $k=n+1$.

Если несовместимые события $a_1, a_2, \ldots, a_k$ единственно возможны (т.-е. если одно из них обязательно должно наступить), то

$$p_1+p_2+\ldots+p_k=1. \tag{5}$$

Например, если мы обозначим через p_1 вероятность для новорожденного ребенка умереть на первом году жизни, через p_2 — вероятность умереть на втором году жизни и т. д., то указанные вероятности необходимо удовлетворяют равенству (5), где k есть число лет, более которого индивиды рассматриваемого класса новорожденных не могут прожить.

В дальнейшем мы будем обозначать через $\bar{A}$ факт ненаступления события A, т.-е. отрицание события A; таким образом, в частности,

$$вер.\ A+вер.\ \bar{A}=1. \tag{5 bis}$$

При применении теоремы сложения необходимо помнить, что все факты, вероятности которых складываются, должны быть несовместимы. Например, вероятность, что одно из двух данных лиц умрет в течение года, не равна сумме вероятностей для каждого из них умереть в течение года, так как не исключена возможность их совместной смерти; напротив, мы в праве утверждать, согласно теореме сложения, что для каждого лица вероятность умереть в течение 2 лет равна сумме вероятностей смерти в течение первого года и вероятности умереть в течение второго года.

Теорема сложения вероятностей позволяет нам исчерпывающим образом резюмировать всю математическую часть I главы следующим положением:

Основной принцип теории вероятностей. Условие, необходимое и достаточное для того, чтобы при данном определенном комплексе условий вероятности фактов $A_1, A_2, \ldots$ могли быть соответственно равны числам $p_1, p_2, \ldots$, состоит в том, чтобы числа эти были не отрицательны и чтобы вероятность события, заключающегося в наступлении одного или другого из нескольких несовместимых фактов, была

всегда равна сумме вероятностей каждого из них, при чем вероятности достоверных событий должны быть равны 1, а невозможные факты должны иметь вероятность, равную 0.

Далеко не всегда бывает возможно в конкретных задачах установить с достаточным основанием, какие вероятности следует приписать каждому из интересующих нас событий; единственное формальное требование, которому при данном неизменном комплексе условий эти вероятности должны удовлетворять, выражено в указанном основном принципе, который применяется не только для проверки логической непротиворечивости принятых значений вероятностей, но и для установления неравенств, которым неизвестные вероятности должны удовлетворять.

ГЛАВА ВТОРАЯ

ТЕОРЕМА УМНОЖЕНИЯ ВЕРОЯТНОСТЕЙ

1. Читатель, который дал себе труд внимательно прочесть всю первую главу, должен испытать некоторое недоумение, что мы формулировали только что основной принцип теории вероятностей, как будто забыв о сделанном вначале указании, что теория вероятностей опирается еще на одну аксиому (о совмещении).

Дело в том, что вышеуказанный принцип говорит о свойствах вероятностей событий при предположении, что наличный комплекс условий, при которых рассматриваются наши события, остается неизменным. Между тем мы указали вначале, что вероятность каждого факта существенно зависит от того, каков тот комплекс условий, который предполагается реализованным. Поэтому весьма важным является выяснение вопроса о том, возможно ли определенным образом связать вероятности фактов, при одном комплексе условий, с вероятностями тех же фактов, если данные условия будут изменены и, в частности, если станет известным, что некоторый факт, который раньше мы считали только вероятным, в действительности осуществился. Понятно, что принцип, формулированный в конце главы I, не может служить мостом для перехода от значений вероятностей при одном комплексе S условий к системе значений, соответствующей другому комплексу S_1; эту роль выполняет аксиома о совмещении событий.

Мы называем факт α совмещением событий A и B, если наступление факта α означает совместное осуществление обоих событий A и B. В частности, если событие A есть частный случай факта B, то α тождественно с A, так как наступление A влечет за собой осуществление B[1]), и, следовательно, *вер.* $\alpha =$ *вер.* A (например, при одном бросании игральной кости совмещение факта A появления 6 очков с фактом B появления четного числа очков — равнозначно факту A, так как 6 есть четное число); с другой стороны, если факты A

[1]) См. стр. 5.

и B несовместимы между собой, то их совмещение a невозможно — *вер.* $a = 0$ (например, если факт A есть появление 6 очков, а факт B состоит в появлении нечетного числа очков). Таким образом, если нам известны вероятности двух фактов A и B, то этого недостаточно для того, чтобы определить вероятность их совмещения; в частности, если при наступлении A событие B несомненно, то *вер.* $a =$ *вер.* A; если же, в случае наступления A, событие B невозможно, то *вер.* $a = 0$. Но, в общем случае, после наступления факта A событие B может быть и не достоверным и не невозможным, а только получать ту или иную вероятность, — тогда и совмещение их a в зависимости от этого будет получать ту или иную вероятность. Следует отметить как особо важный случай, когда при наступлении факта A событие B сохраняет ту же вероятность, какую оно имело до того, как стало известным наступление факта A; мы говорим тогда, что событие B независимо от A (как будет показано дальше, в таком случае событие A также независимо от B, т.-е. осуществление факта B также не изменяет значения вероятности A). В противном случае события A и B называются зависимыми.

Примером двух независимых событий может служить появление 6 очков при бросании игральной кости и появление туза при вынимании карты из некоторой колоды карт. В примерах, рассмотренных выше, когда A есть частный случай B или несовместимо с B, факты A и B зависимы, и даже весьма значительно; зависимыми являются также факты A и B заболевания одною и тою же заразной болезнью двух лиц, живущих в одной квартире.

2. Если нам известны только вероятности обоих фактов A и B и мы по той или иной причине не можем принять в расчет, какую вероятность получает один из фактов, когда становится известным осуществление другого, — то для вероятности совмещения A и B нельзя указать определенного значения, но можно на основании принципа (стр. 18) предыдущей главы установить некоторые неравенства, которым эта вероятность должна удовлетворять. Действительно, мы можем утверждать только, что

$$\left.\begin{aligned}
&\textit{вер.}\ (A \text{ и } B) + \textit{вер.}\ (A \text{ и } \bar{B}) = \textit{вер.}\ A,\\
&\textit{вер.}\ (A \text{ и } B) + \textit{вер.}\ (\bar{A} \text{ и } B) = \textit{вер.}\ B,\\
&\textit{вер.}\ (\bar{A} \text{ и } B) + \textit{вер.}\ (\bar{A} \text{ и } \bar{B}) = \textit{вер.}\ \bar{A},\\
&\textit{вер.}\ (A \text{ и } \bar{B}) + \textit{вер.}\ (\bar{A} \text{ и } \bar{B}) = \textit{вер.}\ \bar{B},
\end{aligned}\right\} \qquad (6)$$

где $\bar{A}$ есть отрицание (т.-е. ненаступление) A, и $\bar{B}$ есть отрицание B, так как первое из равенств (6) означает, что наступление факта A возможно либо совместно с фактом B либо с $\bar{B}$ и аналогичные значения имеют и остальные равенства. На первый взгляд может показаться, что из четырех уравнений (6) можно определить неизвестные вероятности четырех совмещений: A и B, A и $\bar{B}$, $\bar{A}$ и B, $\bar{A}$ и $\bar{B}$, если даны вероятности[1]) факта A и факта B; но в действительности одно из этих уравнений является следствием из трех других, так как оба последние уравнения (6) равнозначны одному уравнению

$$вер.\ (A \text{ и } B) + вер.\ (A \text{ и } \bar{B}) + вер.\ (\bar{A} \text{ и } B) + вер.\ (\bar{A} \text{ и } \bar{B}) = 1 \quad (6\ \text{bis})$$

(выражающему, что все указанные четыре совмещения единственно возможны), из которого они могут быть соответственно получены вычитанием из (6 bis) каждого из первых двух уравнений (6). Поэтому, принимая во внимание, что мы имеем только 3 независимых уравнения, одному из рассматриваемых совмещений можем приписать произвольную вероятность, и тогда вероятности прочих совмещений получат вполне определенные значения; необходимо только, чтобы ни одна из этих вероятностей не оказалась отрицательной. Таким образом, придавая *вер.* (A и B) то или иное значение, получим:

$$вер.\ (A \text{ и } \bar{B}) = вер.\ A - вер.\ (A \text{ и } B),$$
$$вер.\ (\bar{A} \text{ и } B) = вер.\ B - вер.\ (A \text{ и } B),$$
$$вер.\ (\bar{A} \text{ и } \bar{B}) = 1 - вер.\ A - вер.\ B + вер.\ (A \text{ и } B).$$

Для того, чтобы все совмещения имели неотрицательные вероятности, следовательно, необходимо и достаточно, чтобы *вер.* (A и B) удовлетворяла неравенствам:

$$вер.\ (A \text{ и } B) \geqslant 0,\ вер.\ (A \text{ и } B) \leqslant вер.\ A,\ вер.\ (A \text{ и } B) \leqslant вер.\ B,$$
$$вер.\ (A \text{ и } B) \geqslant вер.\ A + вер.\ B - 1. \quad (7)$$

Например, если *вер.* $A = 0{,}5$, *вер.* $B = 0{,}9$, то необходимо и достаточно соблюдение неравенства $0{,}4 \leqslant$ *вер.* (A и B) $\leqslant 0{,}5$ для того, чтобы все вероятности получили неотрицательные значения в соответствии с основным принципом.

3. *Аксиома совмещения событий.* Если α есть частный случай факта A, то вероятность α при данных усло-

[1]) Напоминаем, что *вер.* $\bar{A} = 1 -$ *вер.* A и *вер.* $\bar{B} = 1 -$ *вер.* B.

виях зависит только от вероятности факта A при тех же условиях и от вероятности, которую приобретает α в случае осуществления факта A.

Иными словами, если α_1 есть частный случай какого-нибудь другого факта A_1, то вероятности событий α и α_1 при данных условиях равны, если *вер.* $A =$ *вер.* A_1 при данных первоначальных условиях и если вероятность, которую получает α после осуществления A, равна вероятности, которую получает α_1 в случае осуществления факта A_1.

При этом следует заметить, что если α есть совмещение факта A с некоторым другим фактом B, то при осуществлении факта A наступление события α (т.-е. совмещения A и B) становится равнозначным наступлению одного только факта B; поэтому нашу аксиому можно формулировать и таким образом: вероятность совмещения A и B (при данных условиях) зависит исключительно от вероятности A (при тех же условиях) и от вероятности, которую приобретает факт B после осуществления A. В частности, в случае независимых событий, т.-е. когда вероятность факта B, после осуществления события A, остается та же, что и при первоначальных условиях, наша аксиома означает следующее: вероятность совмещения A и B зависит только от первоначальных вероятностей этих фактов, если события A и B независимы.

Разъясним сначала аксиому совмещения для случая независимых событий. Рассмотрим два явления, которые протекают вне всякой связи между собой, так что исход одного из них совершенно не влияет на исход другого; иными словами, тот или другой результат первого опыта не зависит от исхода второго опыта, и наоборот, т.-е. вероятность определенного исхода одного опыта не изменяется от того, что становится известным результат другого опыта. Оба явления или опыта называются тогда независимыми Примером двух таких независимых опытов может служить бросание игральной кости, с одной стороны, и вынимание шара из урны, в которой имеется одинаковое число белых и черных шаров, — с другой. Наша аксиома утверждает в данном примере очевидную истину, что если мы считаем равновозможными все шесть граней кости и, с другой стороны, столь же вероятным появление черного шара, как и белого, то вероятность совмещения факта появления

черного шара с появлением 6 очков равна вероятности совмещения белого шара с 5 очками и вообще каждое из 12 возможных совмещений (шара белого цвета с любой гранью или черного шара с любой гранью) имеет одну и ту же вероятность, так как каждое из них представляет совмещение двух независимых фактов с соответственно равными вероятностями. Вероятность каждого из этих совмещений (ввиду принятого нами соглашения о том, чтобы называть математической вероятностью события отношение числа благоприятствующих ему равновероятных, независимых и единственновозможных случаев к общему числу случаев) равна $\frac{1}{12}$.

Перейдем теперь к рассмотрению примеров, поясняющих общую аксиому совмещения. Допустим, что Иван покупает по одному билету в двух лотереях, а Петр покупает билет только в первой лотерее с тем, чтобы купить один билет во второй лотерее лишь в том случае, если он выиграет в первой. (Мы допускаем, что по окончании первой лотереи билеты второй еще не будут все распроданы.) Вероятность Ивана и Петра выиграть в первой лотерее одна и та же, но первоначальная вероятность у Ивана, для которого обе игры независимы, выиграть во второй лотерее, очевидно, больше, чем у Петра, который участвует в последней лотерее только в случае выигрыша в первой. Тем не менее, полагая, что более поздняя покупка билета не меняет шансов выигрыша, т.-е. что вероятность для Петра, когда он купит билет после своего первого выигрыша, выиграть во второй лотерее будет та же, что и для Ивана, мы утверждаем, что вероятность выиграть в обеих лотереях одна и та же у Ивана и у Петра; наше последнее утверждение представляет применение аксиомы совмещения к данному примеру, так как оно выражает, что вероятность совмещения обоих выигрышей зависит лишь от вероятности первого выигрыша (которая, по предположению, у Ивана и Петра одинакова) и от вероятности второго выигрыша, после осуществления первого (которая также одинакова у обоих игроков).

Возьмем другой пример. Положим, что вероятность для здорового десятилетнего ребенка умереть в течение года равна P, а вероятность заболеть скарлатиной равна Q. Эти два числа не дают нам возможности судить о том, какова вероятность для здорового десятилетнего ребенка заболеть и умереть в течение года от скарлатины. Искомая вероятность зависит от вероятности Q заболеть скарлатиной

и от вероятности R для всякого десятилетнего ребенка, заболевшего скарлатиной, умереть; иными словами, вероятность совмещения у одного и того же ребенка этих двух зависимых фактов — заболевания скарлатиной и смерти — та же, что вероятность совмещения двух независимых фактов: заболевания скарлатиной одного ребенка десяти лет и смерти другого такого же ребенка, больного скарлатиной. Из теоремы, которую мы сейчас докажем, будет вытекать, что вероятность смерти от скарлатины для здорового десятилетнего ребенка равна произведению $Q \cdot R$ и, таким образом, вовсе не зависит от P.

В дальнейшем, для упрощения письма, мы будем обозначать через (A) вероятность факта A при данных первоначальных условиях. Если же дополнительно к этим первоначальным условиям прибавляется сведение, что произошел некоторый факт B, то вероятность A после осуществления B мы будем обозначать через $(A)_B$. Вероятность же совмещения событий A и B мы будем обозначать символом $(A, B) = (B, A)$.

Таким образом аксиома совмещения математически выражается так:

$$(A, B) = \Phi[(A), (B)_A] = \Phi[(B), (A)_B], \tag{8}$$

где Φ есть некоторая раз навсегда определенная функция. Теорема умножения вероятностей, которую мы сейчас докажем, устанавливает форму этой функции.

4. *Теорема умножения вероятностей* [1]). Вероятность совмещения событий A и B равна произведению вероятности события A на вероятность, которую приобретает событие B, когда становится известным осуществление факта A; иными словами,

$$(A, B) = (A) \cdot (B)_A. \tag{9}$$

[1]) Функция Φ получила бы другую форму, отличную от (9), если бы нами было сделано другое соглашение относительно того, что мы будем называть математической вероятностью. Так, например если бы называли вероятностью $\left(\text{вместо } \frac{m}{n}\right)$ отношение $\frac{m}{n-m}$, где m — число благоприятных случаев, а n — общее число случаев, то мы должны были бы заменить формулу (9) формулой $(A, B) = \frac{(A) \cdot (B)_A}{1 + (A) + (B)_A}$.

В самом деле, положим, что вероятности (A) и $(B)_A$ равны некоторым рациональным числам

$$(A)=\frac{m}{n}, \quad (B)_A=\frac{m_1}{n_1}.$$

Рассмотрим 2 урны с шарами: в первой m белых шаров при общем числе шаров, равном n, а во второй m_1 белых шаров при общем числе шаров, равном n_1. Положим, что из каждой урны вынимают по одному шару, при чем появление белого шара из первой урны назовем событием A_1, а появление белого шара из второй урны назовем событием B_1. Условия вынимания из второй урны полагаем независимыми от результата вынимания из первой урны, так что $(B_1)_{A_1}=(B_1)$. Кроме того, считая все шары равновероятными в каждой урне, имеем:

$$(A_1)=\frac{m}{n}, \quad (B_1)=(B_1)_{A_1}=\frac{m_1}{n_1}.$$

В таком случае, согласно аксиоме совмещения, одинаково вероятно совместное появление любого определенного шара из первой урны и любого определенного шара из второй. Таким образом все единственно возможные и несовместимые результаты обоих вынимаий, соответствующие совместному появлению каждого шара первой урны с каждым шаром второй урны, число которых равно произведению $n\cdot n_1$, являются равновероятными. Поэтому, замечая, что совмещению событий A_1 и B_1 благоприятствуют только пары вынутых белых шаров, число которых равно $m\cdot m_1$ (так как каждый белый шар первой урны может соединиться с любым белым шаром из второй урны), и применяя определение математической вероятности, заключаем, что вероятность совмещения A_1 и B_1 равна

$$(A_1, B_1)=\frac{m\cdot m_1}{n\cdot n_1}=\frac{m}{n}\cdot\frac{m_1}{n_1}=(A_1)\cdot(B_1)_{A_1}. \tag{10}$$

Но так как, согласно нашей аксиоме, мы должны признать, что

$$(A, B)=(A_1, B_1), \tag{11}$$

коль скоро $(A)=(A_1)$ и $(B)_A=(B_1)_{A_1}$, то, принимая во внимание (10) и (11), мы приходим к заключению, что, каковы бы ни были события A и B,

$$(A, B)=(A)\cdot(B)_A. \tag{9}$$

Таким образом теорема доказана для случая, когда вероятности (A) и $(B)_A$ рациональны Теорема распространяется так же, как и теорема сложения, посредством перехода к пределу (см. стр. 16), на случай иррациональных вероятностей.

Пример. Даны две урны с шарами: в одной 3 белых, 7 красных и 15 черных шаров; во второй урне 10 белых, 6 красных и 9 черных шаров. Вынимается по одному шару из каждой урны, при чем шары предполагаются одинаковыми наощупь и тщательно размешанными; спрашивается, какова вероятность, что вынутая пара шаров будет одинакового цвета.

Вероятности появления белого, красного и черного шара из первой урны соответственно равны $\frac{3}{25}$, $\frac{7}{25}$ и $\frac{15}{25}$; для второй урны соответствующие вероятности равны $\frac{10}{25}$, $\frac{6}{25}$ и $\frac{9}{25}$. Поэтому вероятность появления пары белых шаров равна $\frac{3}{25}\cdot\frac{10}{25}=\frac{30}{625}$, вероятность появления красной пары равна $\frac{7}{25}\cdot\frac{6}{25}=\frac{42}{625}$, вероятность появления черной пары равна $\frac{15}{25}\cdot\frac{9}{25}=\frac{135}{625}$. Следовательно, по теореме сложения, вероятность появления одноцветной пары шаров равна

$$\frac{30}{625}+\frac{42}{625}+\frac{135}{625}=\frac{207}{625},$$

т.-е. немного менее $\frac{1}{3}$. Вероятность, что вынутые шары будут разных цветов, равна $1-\frac{207}{625}=\frac{418}{625}$.

5. Из теоремы умножения вероятностей вытекает, что если нам известны вероятность (A) факта A и также вероятность (A,B) совмещения факта A с некоторым другим событием B, то мы можем определить, какова вероятность $(B)_A$ факта B после осуществления события A. Действительно, из равенства

$$(A,B)=(A)\cdot(B)_A \tag{9}$$

находим:

$$(B)_A=\frac{(A,B)}{(A)}. \tag{12}$$

Таким же образом можно получить вероятность A, когда событие B предполагается осуществленным:

$$(A)_B=\frac{(A,B)}{(B)}. \tag{12 bis}$$

Отсюда заключаем, в частности, что если B независимо от A, то и A независимо от B. Действительно, если

$(B)_A = (B)$, то $(A,B) = (A)\cdot(B)$, и, подставляя это значение в (12 bis), получим:

$$(A)_B = \frac{(A)\cdot(B)}{(B)} = (A).$$

Заменяя в формуле (9) A и B взаимно, мы получим:

$$(B,A) = (B)\cdot(A)_B.$$

Отсюда заключаем, что

$$(A)\cdot(B)_A = (B)\cdot(A)_B. \tag{13}$$

Таким образом, зная первоначальные вероятности событий A и B и зная вероятность второго из них B в случае наступления первого A, мы можем определить, какова вероятность события A, когда становится известным, что осуществилось событие B; а именно из (13) выводим:

$$(A)_B = \frac{(A)\cdot(B)_A}{(B)}. \tag{14}$$

Предположим, например, что для лиц данной группы вероятность смерти в течение года равна $(B) = 0{,}006$, вероятность же быть помещенным в больницу $(A) = 0{,}02$. Допустим далее, что для лица данной группы, которое находилось в больнице, вероятность смерти $(B)_A = 0{,}06$; какова вероятность $(A)_B$, что умершее лицо данной группы подвергалось больничному лечению? Применяя формулу (14), находим:

$$(A)_B = \frac{0{,}02\cdot 0{,}06}{0{,}006} = 0{,}2.$$

Формулу (14) можно еще написать в виде пропорции

$$\frac{(A)_B}{(A)} = \frac{(B)_A}{(B)} = \lambda, \tag{15}$$

которая выражает, что вероятность факта A после наступления B во столько же раз больше (меньше) его первоначальной вероятности, во сколько вероятность события B в случае наступления A больше (меньше) его первоначальной вероятности. Величину λ рассматриваемого отношения назовем коэффициентом совместимости между событиями A и B. Если $\lambda = 1$

события независимы; если $\lambda < 1$, то после появления одного из событий вероятность другого уменьшается, и, в частности, при $\lambda = 0$ события A и B несовместимы; напротив, если $\lambda > 1$, то в случае наступления одного из событий другое становится более вероятным, чем было первоначально.

Формулу (9) можно, введя коэффициент совместимости, написать в виде

$$(A, B) = \lambda (A) \cdot (B). \qquad (9 \text{ bis})$$

В следующей главе мы рассмотрим некоторые другие коэффициенты, которыми пользуются для характеристики зависимости между двумя явлениями, когда хотят дать сразу представление не только о связи между исходами A и B двух опытов, но между всеми четырьмя комбинациями исходов: A и B, A и $\overline{B}$, $\overline{A}$ и B, $\overline{A}$ и $\overline{B}$.

Упражнения. 1) Доказать, что если вероятность факта A равна $(A) = 0{,}99$, а вероятность факта B равна $(B) = \frac{1}{2}$, то $(A)_B \geqslant 0{,}98$.

Ответ. Это вытекает из неравенства $(A)_B \geqslant 1 - \frac{(\overline{A})}{(B)}$, которое получается из $(A, B) \geqslant (A) + (B) - 1$.

2) Дано, что вероятность для новорожденного дожить до 5 лет равна $p_5 = \frac{2}{3}$, а вероятность дожить до 50 лет равна $p_{50} = \frac{1}{2}$; требуется определить вероятность x, что ребенок, достигший 5 лет, проживет до 50 лет.

Ответ. $x = \dfrac{\frac{1}{2}}{\frac{2}{3}} = \frac{3}{4}$.

3) Доказать тождество:

$$\begin{vmatrix} \dfrac{1}{(A)_B} & \dfrac{1}{(A)_{\overline{B}}} & 1 \\ \dfrac{1}{(B)_A} & \dfrac{1}{(B)_{\overline{A}}} & 1 \\ 1 & 1 & 1 \end{vmatrix}$$

Ответ. Нужно исключить (A) и (B) из равенств:

$$(A)(B)_A = (B)(A)_B, \quad [1 - (A)](B)_{\overline{A}} = (B)[1 - (A)_B],$$
$$[1 - (B)](A)_{\overline{B}} = (A)[1 - (B)_A].$$

ГЛАВА ТРЕТЬЯ

КОЭФФИЦИЕНТЫ РЕГРЕССИИ, КОРРЕЛЯЦИИ И СВЯЗИ МЕЖДУ ДВУМЯ СОБЫТИЯМИ

1. Пусть A и B будут два как угодно связанные события, вероятность совмещения которых равна (A, B). Если бы эти события были независимы, то вероятность их совмещения была бы равна $(A)\cdot(B)$. В противном случае, разность

$$(A, B) - (A)(B) = \delta \tag{16}$$

будет отлична от нуля; эту разность δ мы назовем связью между событиями A и B. Связь между A и B положительна, если вероятность одного факта возрастает после наступления другого; напротив, связь между событиями A и B отрицательна, если наступление одного из них уменьшает вероятность другого. Это видно из того, что благодаря (9) равенство (16) можем написать в виде

$$(A)[(B)_A - (B)] = \delta, \text{ или } (B)_A = (B) + \frac{\delta}{(A)}. \tag{17}$$

В частности, условие необходимое и достаточное для того, чтобы события A и B были независимы, заключается в том, что связь δ между ними должна быть равна 0.

Легко показать, что связь между событием A и событием $\bar{B}$ равна $-\delta$. Действительно,

$$(A, B) + (A, \bar{B}) = (A), \tag{18}$$

так как событие A может произойти либо при осуществлении факта B либо при его ненаступлении; с другой стороны, имеем также:

$$(A)\cdot(B) + (A)\cdot(\bar{B}) = (A)[(B) + (\bar{B})] = (A). \tag{19}$$

Поэтому, вычитая (19) из (18), получаем:

$$(A, B) - (A) \cdot (B) + (A, \bar{B}) - (A) \cdot (\bar{B}) = 0,$$

откуда, замечая, что

$$(A, B) - (A) \cdot (B) = \delta, \tag{16}$$

заключаем, что

$$(A, \bar{B}) - (A) \cdot (\bar{B}) = -\delta. \tag{16 bis}$$

Таким образом знак связи меняется, если одно из событий B заменить его отрицанием $\bar{B}$, но абсолютная величина связи сохраняется. Поэтому находим также, что

$$(\bar{A}, B) - (\bar{A}) \cdot (B) = -\delta;$$

$$(\bar{A}, \bar{B}) - (\bar{A})\,(\bar{B}) = \delta. \tag{16 ter}$$

Можем указать еще другое выражение для связи, которое получим, если заметим, на основании (16), (16 bis) и (16 ter), что

$$(A, B)\,(\bar{A}, \bar{B}) = [(A)\,(B) + \delta]\,[(\bar{A})\,(\bar{B}) + \delta]$$

и

$$(A, \bar{B})\,(\bar{A}, B) = [(A)\,(\bar{B}) - \delta]\,[(B)\,(\bar{A}) - \delta]. \tag{20}$$

Поэтому, вычитая второе равенство из первого, получим:

$$(A, B)\,(\bar{A}, \bar{B}) - (A, \bar{B})\,(\bar{A}\,B) =$$

$$= \delta\,[(A)\,(B) + (A)\,(\bar{B}) + (\bar{A})\,(B) + (\bar{B})\,(\bar{A})] = \delta. \tag{21}$$

Следует заметить, что сама по себе абсолютная величина связи δ является мало характерной, так как она существенным образом зависит от индивидуальных вероятностей фактов A и B. Очевидно, что мы должны считать два явления максимально связанными между собой, если факты A и B тождественны, т.-е. если они всегда происходят совместно, иными словами, если $(A) = (B)$ и $(A, \bar{B}) = (\bar{A}, B) = 0$; тогда

$$\delta = (A, B)\,(\bar{A}, \bar{B}) = (A) - (\bar{A}) = (A)\,[1 - (A)].$$

Таким образом, если бы факт A имел вероятность 0,1, то мы получили бы, что связь его с самим собой была бы равна $\delta = 0{,}1 \cdot 0{,}9 = 0{,}09$; с другой стороны, если $(A) = (B) = \frac{1}{2}$ и $(B)_A = \frac{3}{4}$, то

по формуле (17) мы получили бы $\delta = \frac{1}{2}\left(\frac{3}{4} - \frac{1}{2}\right) = \frac{1}{8} = 0{,}125$. Следовательно, хотя в последнем случае факт B не всегда наступает вслед за фактом A, мы получаем, что его связь с фактом A больше, чем при предшествующих данных. Вообще нельзя вполне охарактеризовать зависимость между двумя событиями при помощи одного только числа, и, смотря по тому, в каком отношении мы эту зависимость хотим охарактеризовать, можно пользоваться тем или иным коэффициентом, но для полной характеристики необходимо знать 3 числа, которые позволили бы определить вероятности всех четырех совмещений $(A, B), (A, \bar{B}), (\bar{A}, B), (\bar{A}, \bar{B})$, сумма которых (как единственно возможных несовместимых событий) равна единице.

2. Для уяснения дальнейшего рассмотрим пример. Предположим, что при лечении дифтерита применяется новое средство — противодифтеритная сыворотка; применение этого средства назовем событием A, а факт выздоровления больного назовем B. Пусть вероятность выздоровления при наличности A равна $(B)_A = \frac{29}{30}$, а вероятность выздоровления без применения сыворотки равна $(B)_{\bar{A}} = \frac{13}{15}$, так что в первом случае вероятность смерти $(\bar{B})_A = \frac{1}{30}$, а без сыворотки вероятность смерти $(\bar{B})_{\bar{A}} = \frac{2}{15}$. Следовательно оба числа $(B)_A$ и $(B)_{\bar{A}}$ вполне достаточны для суждения о том, насколько опасна болезнь и в какой мере она поддается сывороточному лечению. Одно только число $(B)_A$ давало бы нам лишь представление, насколько дифтерит поддается лечению сывороткою, но понятно, что, не зная о том, какова смертность при отсутствии указанного способа лечения, мы не могли бы судить о степени его целесообразности. Напротив, поскольку применение сыворотки (событие A) зависит от нашего усмотрения, вероятность (A, B) нас мало интересует; поэтому, характеризуя зависимость между нашими фактами A и B при помощи одного коэффициента, мы должны во всяком случае выбрать его так, чтобы он не изменял своего значения, если (A, B) будет изменяться, при неизменных значениях $(B)_A$ и $(B)_{\bar{A}}$.

Связь δ, о которой мы говорили выше, этому условию не удовлетворяет. Действительно, применяя к каждому из четырех совме-

щений, входящих в формулу (21), теорему умножения вероятностей, мы находим, что

$$\delta = (A)(B)_A(\bar{A})(\bar{B})_{\bar{A}} - (A)(\bar{B})_A(\bar{A})(B)_{\bar{A}} =$$

$$= (A)(\bar{A})\left[(B)_A(\bar{B})_{\bar{A}} - (\bar{B})_A(B)_{\bar{A}}\right]. \qquad (21\ \text{bis})$$

Таким образом для рассмотренного выше примера мы находим:

$$\delta = (A)(\bar{A})\left[\frac{29}{30}\cdot\frac{2}{15} - \frac{1}{30}\cdot\frac{13}{15}\right] = \frac{(A)(\bar{A})}{10}.$$

Поэтому значение δ существенно зависит от (A) и $(\bar{A}) = 1 - (A)$. Вообще, принимая во внимание, что $(\bar{B})_{\bar{A}} = 1 - (B)_{\bar{A}}$ и $(\bar{B})_A = 1 - (B)_A$, можем написать формулу (21 bis) в виде

$$\delta = (A)(\bar{A})\left[(B)_A - (B)_{\bar{A}}\right]; \qquad (22)$$

откуда заключаем, что

$$\rho_B = \frac{\delta}{(A)(\bar{A})} = (B)_A - (B)_{\bar{A}}. \qquad (23)$$

Величина ρ_B $(B)_A - (B)_{\bar{A}}$, которая в нашем примере равна $\frac{1}{10}$, представляет наиболее важную величину, характеризующую эффективность применения сыворотки: она показывает, насколько увеличивается вероятность выздоровления больного, если вместо того, чтобы отвергнуть прививку, эту прививку произведут. Можно сказать, зная только число $(B)_A - (B)_{\bar{A}} = \frac{1}{10}$, что если вероятность выздоровления больного, не подвергшегося прививке, равна вероятности вынуть белый шар из урны, в которой имеется x белых шаров при общем числе шаров, равном 100, то для больного, к которому применили сыворотку, вероятность выздоровления изменилась на столько же, как если бы в нашей урне 10 не белых шаров заменили белыми, т.-е. увеличилась на $\frac{10}{100} = \frac{1}{10}$.

Величину $\rho_B = (B)_A - (B)_{\bar{A}}$ называют коэффициентом регрессии события B относительно события A. Коэффициент регрессии, так же как и связь δ, обращается в 0 только, если факты A и B независимы, и, как видно из (23), имеет

всегда тот же знак, что и δ. Кроме того, вообще $|\rho_B| \leqslant 1$, так как ρ_B есть разность между двумя положительными числами, не превышающими 1; при этом $\rho_B = 1$, когда факты A и B тождественны, т.-е. происходят лишь совместно $[(B)_A = 1,\ (B)_{\bar{A}} = 0]$, и $\rho_B = -1$, если B тождественно с отрицанием A.

Можно совершенно так же рассматривать и коэффициент регрессии

$$\rho_A = (A)_B - (A)_{\bar{B}} = \frac{\delta}{(B)(\bar{B})} \qquad (23\ \text{bis})$$

события A относительно события B. Однако в данном примере этот коэффициент менее интересен, и, кроме того, даже индивидуальные значения $(B)_A = \frac{29}{30}$ и $(B)_{\bar{A}} = \frac{13}{15}$ недостаточны для того, чтобы его определить.

Действительно, из формулы (23) и (23 bis) находим:

$$\rho_A = \rho_B \frac{(A)(\bar{A})}{(B)(\bar{B})}, \qquad (24)$$

где $(B) = (A)(B)_A + (\bar{A})(B)_{\bar{A}}$, потому что B происходит либо вместе с A, либо вместе с $\bar{A}$; так что множитель при ρ_B зависит кроме $(B)_A$ и $(B)_{\bar{A}}$ еще и от (A). Так, в рассматриваемом примере ρ_A может оказаться сколь угодно малым, если вероятность прививки (A) весьма мала [1]) или, наоборот, достаточна близка к единице; между тем, если положить $(A) = \frac{3}{4}$, то $(\bar{A}) = \frac{1}{4}$, $(B) = \frac{3}{4} \cdot \frac{29}{30} + \frac{1}{4} \cdot \frac{13}{15} = \frac{113}{120}$, $(\bar{B}) = \frac{7}{120}$; поэтому $\rho_A = \frac{2700}{791} \rho_B = \frac{270}{791} \neq 0{,}34$.

Из (24) мы видим, что a priori коэффициенты регрессии ρ_A и ρ_B связаны только общностью знака, по численному же значению один из коэффициентов может быть близок к 0, когда другой близок к 1. Во всяком случае, знание обоих коэффициентов регрессии дает довольно существенное понятие о взаимоотношениях между событиями A и B, так как каждый из них выражает соответственно разность между вероятностями одного из событий, в случае наступления и в случае ненаступления другого.

---

[1]) Полагая $(A) = 0{,}001$, мы получили бы $\rho_A = \frac{0{,}001 \cdot 0{,}999}{\frac{2{,}6003}{3} \cdot \frac{0{,}3997}{3}} \neq 0{,}009$; значение того же порядка мы получим, если $(A) = 0{,}999$.

Таким образом вообще коэффициенты регрессии ρ_A и ρ_B между собой; для того, чтобы они были равны ($\rho_A = \rho_B$), необходимо и достаточно, чтобы имело место одно из двух условий: либо $(A) = (B)$, либо $(\bar{A}) = 1 - (A) = (B)$, как это видно из равенства (24).

Из равенства (16), (16 bis) и (16 ter) немедленно можно также усмотреть, что условие $(A) = (B)$ означает, что вероятности совмещений $(A, \bar{B})$ и $(\bar{A}, B)$ равны между собой, а условие $(\bar{A}) = (B)$ равнозначно условию, что $(A, B) = (\bar{A}, \bar{B})$.

В некоторых случаях для характеристики взаимной зависимости между двумя событиями A и B ограничиваются тем, что указывают лишь среднюю геометрическую из обоих коэффициентов регрессии

$$R = \pm\sqrt{\rho_A \cdot \rho_B} = \rho_A \sqrt{\frac{(B)(\bar{B})}{(A)(\bar{A})}} = \rho_B \sqrt{\frac{(A)(\bar{A})}{(B)(\bar{B})}}, \qquad (25)$$

при чем R должен иметь тот же знак, каким обладают оба коэффициента регрессии, которые, конечно, всегда имеют одинаковый знак, а именно знак δ. Число R, определенное таким образом, называется коэффициентом корреляции между фактами A и B. Коэффициент корреляции имеет наиболее существенное значение в том случае, когда коэффициенты регрессии ρ_A и ρ_B равны [т.-е. когда $(A) = (B)$ или $(A) = (\bar{B})$]; тогда

$$R = \rho_A = \rho_B. \qquad (26)$$

Предположим, например, что вероятность обоим новорожденным близнецам быть мужского пола равна $(A, B) = 0{,}32$, а вероятность близнецам быть женского пола равна $(\bar{A}, \bar{B}) = 0{,}28$. В таком случае вероятность, что новорожденные близнецы будут разного пола, равна $(A, \bar{B}) + (\bar{A}, B) = 1 - 0{,}32 - 0{,}28 = 0{,}40$, при чем, очевидно, $(A, \bar{B}) = (A, B) = 0{,}20$, так как безразлично, назовем ли мы фактом A принадлежность к мужскому полу первого из рассматриваемых нами младенцев или второго; поэтому

$$(A) = (B) = (A, B) + (A, \bar{B}) = 0{,}32 + 0{,}20 = 0{,}52;$$

$$(\bar{A}) = (\bar{B}) = 0{,}48,$$

т.-е. вероятность взятому на удачу из близнецов быть мальчиком равна 0,52, а вероятность быть девочкой равна 0,48. Заметим, что вероятность хоть одному из двух близнецов быть мальчиком [1]) равна $(A, B) + (A, \bar{B}) + (\bar{A}, B) = 0{,}72$.

В данном случае, по формуле (16)

$$\delta = 0{,}32 - (0{,}52)^2 = 0{,}0496,$$

а потому $R = \rho_A = \frac{0{,}0496}{0{,}52 \,.\, 0{,}48} = \frac{496}{2496} = 0{,}1987.$

Вообще, если известен коэффициент корреляции R и кроме того известны вероятности (A) и (B) [не предполагая более, что $(A) = (B)$] событий A и B, то можно легко вычислить вероятности $(B)_A$, $(B)_{\bar{A}}$ и т. д. В самом деле, из формул (23) и (25) получаем:

$$R = \frac{\delta}{\sqrt{(A)\,(\bar{A})\,(B)\,(\bar{B})}}; \tag{27}$$

поэтому, замечая, что $\delta = (A)\,[(B)_A - (B)]$, находим:

$$R = [(B)_A - (B)] \sqrt{\frac{(A)}{(\bar{A})\,(B)\,(\bar{B})}}, \tag{27 bis}$$

откуда [2])

$$(B)_A = (B) + R \sqrt{\frac{(\bar{A})\;(B)\;(\bar{B})}{(A)}}, \tag{28}$$

и аналогичным образом

$$(B)_{\bar{A}} = (B) - R \sqrt{\frac{(A)\,(B)\,(\bar{B})}{(\bar{A})}}. \tag{28 bis}$$

Отметим основные свойства коэффициента корреляции.

---

[1]) Разумеется, ошибочно было бы применять здесь непосредственно теорему сложения вероятностей к обоим событиям A и B, т.-е. считать, что искомая вероятность равна $(A) + (B)$, так как это было бы допустимо лишь, если бы A и B были несовместимы, между тем как в действительности $(A, B) > 0$.

[2]) Ту же формулу можно получить, исходя из равенства $(B)_A = (B)_{\bar{A}} + \rho_B = (B) + [(B)_A - (B)] + \rho_B = (B) - (A)\,\rho_B + \rho_B = (B) + \rho_B(\bar{A}) = (B) + R \sqrt{\frac{(\bar{A})\,(B)\,(\bar{B})}{(A)}}$.

Коэффициент корреляции между A и $\bar{B}$ равен по численному значению и противоположен по знаку коэффициенту корреляции между A и B. Коэффициент корреляции между $\bar{A}$ и $\bar{B}$ тот же, что между A и B.

2) Коэффициент корреляции обращается в нуль тогда и только тогда, когда события A и B независимы.

3) Если (A) и (B) даны, то $(B)_A$ [а также $(A)_B$] возрастает вместе с алгебраическим возрастанием R; напротив, $(B)_{\bar{A}}$ уменьшается с возрастанием R.

Все это вытекает непосредственно из формулы (28).

4) Коэффициент корреляции R никогда по численному значению не бывает больше единицы. Коэффициент корреляции $R=1$ тогда и только тогда, когда $(A)=(B)=(A,B)$, т.-е. если факты A и B совпадают; напротив $R=-1$ тогда и только тогда, когда A и B являются взаимными отрицаниями (т.-е. когда факты A и $\bar{B}$ совпадают), т.-е. $(A)=(\bar{B})=(A, \bar{B})$.

Действительно, из формулы (16) заключаем, во-первых, что если $(A)=(B)=(A,B)$, то $\delta=(A)-(A)^2=(A)\cdot(\bar{A})$; поэтому из (27) вытекает, что $R=1$. Обратно, пусть будет дано, что $R=1$, и допустим, что $(A)\gtrless(B)$; пусть для определенности $(A)>(B)$, так что $(\bar{A})<(\bar{B})$. Тогда, написавши равенство (28 bis) в виде

$$(B)_{\bar{A}}=(B)\left[1-R\sqrt{\frac{(A)(\bar{B})}{(\bar{A})(B)}}\right], \tag{29}$$

мы получили бы под корнем величину больше единицы, и так как $R=1$, то $(B)_{\bar{A}}$ оказалось бы отрицательным числом. Следовательно наше допущение невозможно, и необходимо принять, что $(A)=(B)$; тогда равенство (29) дает $(B)_{\bar{A}}=0$, т.-е. $(\bar{A},B)=0$, а потому из равенства $(\bar{A},B)+(A,B)=(B)$ заключаем, что $(A,B)=(B)=(A)$. Очевидно, что если бы мы предположили $R>1$, то в равенстве (29) при корне [который сам не меньше единицы, так как $(A)\geqslant(B)$] стоял бы еще множитель $R>1$, и мы пришли бы снова к нелепому заключению, что $(B)_{\bar{A}}<0$; следовательно, всегда $R\leqslant 1$. Так как, заменяя событие B его отрицанием $\bar{B}$, мы мо-

жем всегда случай, когда $R<0$, привести к рассмотренному только что случаю $R>0$, то аналогичным образом убеждаемся, что всегда $R\geqslant -1$, а $R=-1$ означает, что факты A и $\bar{B}$ совпадают.

Вообще из равенства

$$(B)_{\bar{A}}=(B)\left[1-R\sqrt{\frac{(A)(\bar{B})}{(B)(\bar{A})}}\right] \tag{29}$$

видно, что

$$R\leqslant\sqrt{\frac{(\bar{A})(B)}{(A)(\bar{B})}}, \tag{30}$$

при чем знак равенства в (30) соответствует [(из-за (29)] равенству $(B)_{\bar{A}}=0$, т.-е. $(A,B)=(B)$, что означает, что событие A является необходимым следствием факта B; таким образом коэффициент корреляции может оказаться значительно меньше 1, и, несмотря на это, факт A может быть следствием факта B. Найдем, например, коэффициент корреляции между выпадением 6 очков B и появлением A четного числа очков при бросании данной игральной кости. В данном случае $R=\sqrt{\frac{1}{5}}<\frac{1}{2}$, хотя факт A является обязательным следствием события B; только в том случае, если бы и B также обязательно имело место при наступлении факта A, лишь тогда мы получили бы $R=1$.

Таким образом коэффициент корреляции дает весьма неточное представление о характере зависимости между фактами A и B, если мы не знаем кроме того индивидуальных вероятностей этих фактов или по крайней мере значения $\frac{(\bar{A})(B)}{(\bar{B})(\bar{A})}$, являющегося наибольшим значением коэффициента корреляции при данных значениях (A) и $(B)\leqslant(A)$, которое дает также возможность (25) указать оба коэффициента регрессии.

4. На-ряду с коэффициентами регрессии и корреляции представляет интерес рассмотрение величины

$$Q=\frac{\delta}{(A,B)(\bar{A},\bar{B})+(A,\bar{B})(\bar{A},B)}=$$

$$=\frac{(A,B)(\bar{A},\bar{B})-(A,\bar{B})(A,\bar{B})}{(A,B)(\bar{A},\bar{B})+(A,\bar{B})(\bar{A},B)}, \tag{31}$$

которая называется коэффициентом связи между событиями A и B. Коэффициент связи Q, подобно коэффициенту корреляции R, является симметричным по отношению к обоим явлениям, но преимущество его перед коэффициентом корреляции заключается в том, что коэффициент Q, подобно ρ_B, всецело определяется вероятностями $(B)_A$ и $(B)_{\bar{A}}$ события B в случае осуществления и неосуществления события A, но не зависит от вероятности (A); ввиду симметричности коэффициента Q очевидно, что он обладает (как и ρ_A) аналогичным свойством по отношению к событию A. Действительно, равенство (31) можем записать в виде

$$Q = \frac{(A)(\bar{A})[(B)_A(\bar{B})_{\bar{A}} - (\bar{B})_A(B)_{\bar{A}}]}{(A)(\bar{A})[(B)_A(\bar{B})_{\bar{A}} + (\bar{B})_A(B)_{\bar{A}}]} =$$

$$= \frac{\rho_B}{1 - [(B)_A(B)_{\bar{A}} + (\bar{B})_A(\bar{B})_{\bar{A}}]}, \qquad (32)$$

а также, заменяя взаимно A и B, в виде

$$Q = \frac{\rho_A}{1 - [(A)_B(A)_{\bar{B}} + (\bar{A})_B(\bar{A})_{\bar{B}}]} \qquad (32\ \text{bis})$$

Из формул (32) и (32 bis) видим, что Q имеет всегда тот же знак, что и коэффициенты регрессии ρ_A и ρ_B, и кроме того всегда $|Q| \geqslant |\rho_A|$, $|Q| \geqslant |\rho_B|$; при этом, за исключением случая независимости событий $(Q=0)$, равенство $Q=\rho_B$ имеет место только тогда, когда $(B)_A(B)_{\bar{A}} + (\bar{B})_A(\bar{B})_{\bar{A}} = 0$, т.-е. при условии, что факты A и B совпадают $[(B)_{\bar{A}} = (\bar{B})_A = 0]$ или противоположны $[(B)_A = (\bar{B})_{\bar{A}} = 0]$; в таком случае $Q = \rho_B = \rho_A = \pm 1$.

Принимая во внимание, что $R = \pm\sqrt{\rho_A \rho_B}$, заключаем также, что

$$|Q| > |R|, \qquad (33)$$

как больше и коэффициентов регрессии, если только $|R|$ не равно ни нулю, ни единице: в последнем случае $Q = R = \rho_A = \rho_B = \pm 1$.

Из вышесказанного следует, что коэффициент связи Q является, так сказать, более чувствительным показателем связи между событиями, нежели коэффициенты регрессии и корреляции.

В частности, из формулы (31) легко вывести, что $Q=1$ означает, что один по крайней мере из фактов A и B является необходимым следствием другого (именно тот, вероятность которого больше), т.-е. $(A, B)=(A)$, при $(A)\leqslant(B)$; условие $Q=-1$ означает, что факты A и B несовместимы, т.-е. $(A, B)=0$, либо их отрицания $\bar{A}$ и $\bar{B}$ несовместимы.

Действительно, так как числитель выражения (31) является разностью тех же количеств, сумма которых входит в знаменатель, то условие, необходимое и достаточное для того, чтобы $Q=1$, состоит в том, что

$$(A, \bar{B})\cdot(\bar{A}\, B)=0,$$

т.-е. имеет место по крайней мере одно из равенств $(A, \bar{B})=0$ или $(\bar{A}, B)=0$; но, имея в виду, что $(A, \bar{B})+(A, B)=(A)$ и $(\bar{A}, B)+(A, B)=(B)$, это значит, что либо $(A, B)=(A)$, либо $(A, B)=(B)$; иными словами, по крайней мере один из двух фактов A и B является частным случаем другого, и, полагая для определенности $(A)\leqslant(B)$, видим, что A есть частный случай события B, т.-е. что осуществление факта A влечет за собой наступление B.

Аналогичным образом убеждаемся, что $Q=-1$ означает равенство

$$(A, B)\cdot(\bar{A}, \bar{B})=0,$$

т.-е. по крайней мере одна из вероятностей (A, B) или $(\bar{A}, \bar{B})$ равна нулю, т.-е. несовместимы будут сами факты A и B или их отрицания в зависимости от того, будет ли $(A)+(B)<1$ или же $(A)+(B)>1$; если же $(A)+(B)=1$, то B будет отрицанием A.

В примере с антидифтеритной сывороткой, где $(B)_A=\frac{29}{30}$, $(B)_{\bar{A}}=\frac{13}{15}$, находим по формуле (32):

$$Q=\frac{\frac{29}{30}\cdot\frac{2}{15}-\frac{1}{30}\cdot\frac{13}{15}}{\frac{29}{30}\cdot\frac{2}{15}+\frac{1}{30}\cdot\frac{13}{15}}=\frac{45}{71}=0{,}634;$$

таким образом коэффициент связи значительно больше коэффициента регрессии $\rho_B=0{,}1$.

Легко видеть, что если бы сыворотка всегда излечивала дифтерит, т.-е. $(B)_A = 1$, то мы получили бы $Q = 1$; а между тем коэффициент регрессии ρ_B был бы равен $1 - \frac{13}{15} = \frac{2}{15}$; что же касается коэффициента корреляции R, то он зависит существенным образом от того (как было указано выше), как часто применяется сывороточный метод лечения: таким образом в последнем предположении $\left((B)_A = 1,\ (B)_{\bar{A}} = \frac{2}{15}\right)$ мы видим (25), что

$$R = \rho_B \sqrt{\frac{(A)(\bar{A})}{\left[(A) + \frac{13}{15}(\bar{A})\right]\frac{2}{15}(\bar{A})}} =$$

$$= \frac{2}{15}\sqrt{\frac{(A)}{\frac{2}{15}\left[\frac{13}{15} + \frac{2}{15}(A)\right]}},$$

при изменении (A) от 0 до 1 получаем для R всевозможные значения от 0 до $\sqrt{\frac{2}{15}}$.

В примере же с близнецами коэффициент связи Q между их полами равен $Q = \frac{0{,}0496}{0{,}32 \cdot 0{,}28 + (0{,}20)^2} = \frac{0{,}0496}{0{,}1296} = 0{,}382$, т.-е. почти вдвое больше коэффициента корреляции $R = 0{,}1987$.

Резюмируя полученные выводы, можно сказать, что **знание обоих коэффициентов регрессии ρ_A, ρ_B и коэффициента связи Q между событиями A и B дает исчерпывающую картину взаимоотношения между этими событиями, позволяя вычислить вероятности как самих фактов, так и всех совмещений (A, B), $(A, \bar{B})$ $(\bar{A}, B)$, $(\bar{A}, \bar{B})$. Если известно, что $(A) = (B)$, тогда роль обоих коэффициентов регрессии всецело выполняется коэффициентом корреляции R, так как в этом случае $R = \rho_A = \rho_B$. В общем же случае, когда $(A) \lesseqgtr (B)$ и коэффициенты ρ_A и ρ_B не даны индивидуально, а даны только коэффициент корреляции $R = \sqrt{\rho_A \rho_B}$ и коэффициент связи Q, эти два числа**

дают также довольно полную картину взаимоотношения между событиями, так как близость к единице коэффициента Q свидетельствует о том, что один по крайней мере из фактов приобретает очень большую вероятность при осуществлении другого, и, смотря по тому, насколько велик коэффициент корреляции, можно судить о том, в какой мере эта зависимость обратима.

4. *Упражнения.*

1) Взято 1 000 пар отцов с сыновьями. Среди них 625 светлоглазых отцов и столько же (625) светлоглазых сыновей, так что при выборе отца из указанной группы наудачу вероятность (A) (считая каждого индивида равновероятным) выбрать светлоглазого отца равна $(A)=\frac{625}{1\,000}=\frac{5}{8}$ и $(\overline{A})=\frac{3}{8}$; точно так же при выборе индивида из поколения сыновей наудачу вероятность выбрать светлоглазого $(B)=\frac{5}{8}$ и $(\overline{B})=\frac{3}{8}$. Таким образом, если бы взять наудачу одно лицо из отцовского поколения и независимо от него одно лицо из другого поколения, то вероятность, что они оба окажутся светлоглазыми равна $(A)\,(B)=\left(\frac{5}{8}\right)^2=\frac{25}{64}$; однако, если после выбора отца мы индивида из поколения сыновей выбираем не произвольно, а именно как раз сына данного отца, то мы не можем заранее знать, чему равна вероятность (A, B), что и отец и его сын будут светлоглазыми: то или иное значение для (A, B) мы получим в зависимости от того, сколько среди взятых пар состояло из отца и сына с светлыми глазами; положим, что таких пар оказалось 475, тогда $(A, B)=\frac{475}{1\,000}=\frac{19}{40}$, так как каждую пару мы считаем равновероятной. Требуется определить вероятности $(B)_A$, $(B)_{\overline{A}}$ и т. д., а также коэффициенты корреляции (регрессии) между одинаковым цветом глаз у отца и сына данной группы и коэффициент связи между ними.

Ответ. $(B)_A=(A)_B=\frac{(A,\ B)}{A}=\frac{\frac{19}{40}}{\frac{5}{8}}=\frac{19}{25}=0{,}76$; коэффициент совместимости между A и B равен $\lambda=\frac{(A,\ B)}{(A)\,(B)}=\frac{\frac{19}{40}}{\frac{25}{64}}=\frac{152}{125}=1{,}216$; связь $\delta=(A,\ B)-(A)(B)=\frac{19}{40}-\frac{25}{64}=\frac{27}{320}$; $(\overline{A},\ B)=0{,}15$; $(B)_{\overline{A}}=0{,}4$; коэффициент корреляции $R=\rho_A=\rho_B=\frac{\delta}{(A)\,(\overline{A})}=\frac{\frac{27}{320}}{\frac{5}{8}\cdot\frac{3}{8}}=\frac{27}{75}=0{,}36$; коэффи-

циент связи $Q = \frac{R}{(B)_A + (B)_{\bar{A}} - 2(B)_A(B)_{\bar{A}}} = \frac{0{,}36}{0{,}76 + 0{,}4 - 2 \cdot 0{,}76 \cdot 0{,}4} = \frac{15}{23} = 0{,}652.$

2) В данном населении вероятность, что мужчина, вступающий в брак, моложе 30 лет, равна $(A) = 0{,}4$; вероятность, что женщина, вступающая в брак, моложе 30 лет, равна $(B) = 0{,}75$. Кроме того дано, что коэффициент R корреляции между A и B равен 0,25. Требуется определить вероятность $(B)_A$, что мужчина до 30 лет выбирает себе жену до 30 лет, а также вероятность $(B)_{\bar{A}}$ и коэффициент связи Q.

Ответ. $(B)_A = (B) + R\sqrt{\frac{(A)(B)(\bar{B})}{(A)}} = 0{,}75 + 0{,}25\sqrt{\frac{0{,}6 \cdot 0{,}75 \cdot 0{,}25}{0{,}4}} =$

$= 0{,}75 + 0{,}0625\sqrt{4{,}5} \neq 0{,}871$; $(B)_{\bar{A}} = (B) - R\sqrt{\frac{(A)(B)(\bar{B})}{(\bar{A})}} \neq 0{,}663$;

$$Q = \frac{(B)_A - (B)_{\bar{A}}}{(B)_A + (B)_{\bar{A}} - 2(B)_A(B)_{\bar{A}}} \neq 0{,}54.$$

3) Определить вероятности $(B)_A$ и $(B)_{\bar{A}}$, если известны коэффициент регрессии ρ_B и коэффициент связи Q.

Ответ. Для определения $(B)_A = x$ и $(B)_{\bar{A}} = y$ имеем два уравнения $\rho_B = \rho = x - y$ и $Q = \frac{x - y}{x + y - 2xy}$; исключая из них y, получаем уравнение $x^2 - (1 + \rho)x + \frac{\rho}{2Q}(1 + Q) = 0$; откуда

$$x = \frac{1 + \rho \pm \sqrt{1 + \rho^2 - \frac{2\rho}{Q}}}{2}; \quad y = \frac{1 - \rho \pm \sqrt{1 + \rho^2 - \frac{2\rho}{Q}}}{2}. \tag{34}$$

Таким образом получаем две пары решений, беря одновременно знак + или — перед корнем в обеих формулах. Естественно, что мы получили две пары решений, так как коэффициент регрессии события B относительно A тот же, что коэффициент регрессии события $\bar{B}$ относительно $\bar{A}$; поэтому, если $x_1 = (B)_A$ и $y_1 = (B)_{\bar{A}}$ есть одна пара решений, то $x_2 = (\bar{B})_{\bar{A}} = 1 - y_2$ и $y_2 = (\bar{B})_A = 1 - x_1$ является другой парой решений, а из полученных формул видно, что обе пары решений (получающиеся изменением знака перед корнем) связаны равенствами $x_1 + y_2 = y_1 + x_2 = 1$.

Принимая во внимание, что x и y не могут быть мнимыми числами, получаем, что подкоренное количество в формулах (34) не может быть отрицательно; поэтому всегда должны быть соблюдены такие неравенства:

$$1 + \rho_B^2 \geqslant \frac{2\rho_B}{Q}, \quad 1 + \rho_A^2 \geqslant \frac{2\rho_A}{Q}, \tag{35}$$

каждое из которых обращается в равенство соответственно при $(B)_A = (\overline{B})_{\overline{A}}$ и при $(A)_B = (\overline{A})_{\overline{B}}$. Неравенства (35) равнозначны неравенствам

$$|\rho_A| \leqslant \frac{|Q|}{1+\sqrt{1+Q^2}}, \quad |\rho_B| \leqslant \frac{Q}{1+\sqrt{1+Q^2}}. \tag{35 bis}$$

Перемножая эти неравенства, легко проверить, что коэффициент корреляции удовлетворяет такому же неравенству

$$|R| \leqslant \frac{|Q|}{1+\sqrt{1+Q^2}}, \text{ т.-е. } |Q| \geqslant \frac{2|R|}{1+R^2}. \tag{36}$$

Кроме того можно проверить, что последнее неравенство обращается в равенство только при $(A) = (B) = \frac{1}{2}$. Если бы были известны оба коэффициента регрессии, то указанный прием позволил бы определить также $(A)_B$ и $(A)_{\overline{B}}$. Пусть, например, $\rho_A = \rho_B = R = 0{,}36$ и $Q = \frac{15}{23}$; в таком случае из (34) находим $(A)_B = (B)_{\overline{A}} = \frac{1{,}36 \pm 0{,}16}{2}$, $(A)_{\overline{B}} = (B)_{\overline{A}} = \frac{0{,}64 \pm 0{,}16}{2}$, т.-е. либо $(A)_B = (B)_A = 0{,}76$ и $(A)_{\overline{B}} = (B)_{\overline{A}} = 0{,}4$, что соответствует как раз данным предыдущей задачи, либо $(A)_B = (B)_A = 0{,}6$ и $(A)_{\overline{B}} = (B)_A = 0{,}24$, что соответствовало бы также предыдущей задаче, если бы мы событиями A и B называли соответственно не светлые глаза у отца и сына, а темные.

После того как определены значения $(A)_B$, $(B)_A$ $(A)_{\overline{B}}$ и $(B)_{\overline{A}}$, для вычисления (A) и (B) пользуемся уравнениями:

$$(A) = (B)\,(A)_B + (\overline{B})\,(A)_{\overline{B}}, \quad (B) = (A)\,(B)_A + (\overline{A})\,(B)_{\overline{A}},$$

которые легко преобразовать к виду

$$(A) = (B)\,[(A)_B - (A)_{\overline{B}}] + (A)_{\overline{B}}, \quad (B) = (A)\,[(B)_A - (B)_{\overline{A}}] + (B)_{\overline{A}},$$

откуда

$$(A) = \frac{(A)_{\overline{B}} + \rho_A (B)_{\overline{A}}}{1-R^2}, \quad (B) = \frac{(B)_{\overline{A}} + \rho_B (A)_{\overline{B}}}{1-R^2}.$$

В частности, в случае, когда $(A) = (B)$, находим сразу

$$(A) = (B) = \frac{(A)_{\overline{B}}}{1+(A)_{\overline{B}} - (A)_B} = \frac{(A)_{\overline{B}}}{1-R} \tag{37}$$

и видим, что в рассмотренном примере $(A) = (B) = \frac{0{,}4}{1+0{,}4-0{,}76} = \frac{5}{8}$. Таким образом коэффициенты корреляции R и связи Q совместно позволяют определить все вероятности при условии, что $(A) = (B)$.

4) Проверить, что если $R = \frac{1}{2}$, то $Q \geqslant \frac{4}{5}$, а если $Q = \frac{3}{5}$, то $R \leqslant \frac{1}{3}$.

Ответ. Это вытекает из (36)

5) Полагая $(A)=\frac{1}{2}+\alpha$, $(B)=\frac{1}{2}+\beta$, где α и β весьма малые величины, показать, что приближенное равенство

$$Q=\frac{2R}{1+R^2}$$

имеет погрешность порядка $\alpha^2+\beta^2$.

6) Проверить, что оба коэффициента регрессии ρ_A и ρ_B удовлетворяют неравенству:

$$\frac{1}{1+\alpha}\leqslant\frac{\rho}{R}\leqslant 1+\alpha,$$

где

$$1+\alpha=\frac{Q}{R[1+\sqrt{1-Q}]},$$

при чем одно из неравенств обращается в равенство, если

$$(A)_B+(A)_{\bar{B}}=1 \text{ или } (B)_A+(B)_{\bar{A}}=1,$$

и оба неравенства обращаются в равенства, когда $\alpha=0$, т.-е. $(A)=(B)=\frac{1}{2}$.

7) Положим $(A, B)=\lambda(A)(B)$, $(A,\bar{B})=\mu(A)(\bar{B})$, $(\bar{A}, B)=\mu'(\bar{A})(B)$, $(\bar{A},\bar{B})=\lambda'(\bar{A})(\bar{B})$.

Доказать, что если $\lambda>1$, то и $\lambda'>1$, а $\mu<1$ и $\mu'<1$; если же $\lambda=1$ (независимость), то $\lambda'=\mu=\mu'=1$. Доказать, что

$$Q=\frac{\lambda\lambda'-\mu\mu'}{\lambda\lambda'+\mu\mu'}, \text{ т.-е. } \frac{\lambda\lambda'}{\mu\mu'}=\frac{1+Q}{1-Q}; \tag{38}$$

вывести отсюда: что равенство $Q=1$ означает, что либо $\mu=0$, либо $\mu'=0$; а $Q=-1$ означает, что либо $\lambda=0$, либо $\lambda'=0$. Показать, что $(A)=(B)$ равнозначно $\mu=\mu'$; и $(A)=(B)$ равнозначно $\lambda=\lambda'$.

Показать, что можно задать произвольно 3 коэффициента совместимости $\lambda>1$, $\mu<1$, $\mu'<1$, и вероятности (A) и (B) определяются тогда однозначно, а именно:

$$(A)=\frac{1-\mu'}{\lambda-\mu'},\ (B)=\frac{1-\mu}{\lambda-\mu},\ (A,B)=\frac{\lambda(1-\mu')(1-\mu)}{(\lambda-\mu')(\lambda-\mu)}. \tag{39}$$

8) Доказать, что

$$\lambda=1+R\sqrt{\frac{(\bar{A})(\bar{B})}{(A)(B)}},\quad \lambda'=1+R\sqrt{\frac{(A)(B)}{(\bar{A})(\bar{B})}},$$

$$\mu=1-R\sqrt{\frac{(\bar{A})(B)}{(A)(\bar{B})}},\quad \mu'=1-R\sqrt{\frac{(A)(\bar{B})}{(B)(\bar{A})}};$$

вывести отсюда, что

$$R^2 = (\lambda - 1)(\lambda' - 1) = (\mu - 1)(\mu' - 1) \tag{40}$$

и что, при несовместимости фактов A и B, $\lambda' = 1 - R^2$, а также, что (при $R \geqslant 0$)

$$\lambda\lambda' = 1 + R\left[\sqrt{\frac{(\bar{A})(\bar{B})}{(A)(B)}} + \sqrt{\frac{(A)(B)}{(\bar{A})(\bar{B})}}\right] + R^2 \geqslant (1 + R)^2,$$

при чем знак равенства имеет место лишь при условии, что $(A)(B) = (\bar{A})(\bar{B})$, т.-е. при $\lambda = \lambda'$; тогда $\lambda = 1 + R = 1 + \rho_A = 1 + \rho_B$.

Показать, что всегда $\frac{1}{\lambda} + \frac{1}{\lambda'} \geqslant 1$, $\frac{1}{\mu} + \frac{1}{\mu'} \geqslant 1$, при чем знак равенства означает, что события тождественны или противоположны.

9) Доказать, что, если Q и $R \geqslant 0$ даны, то

$$\frac{1+Q}{1-Q}(1-R)^2 \geqslant \lambda\lambda' \geqslant (1+R)^2, \quad (1-R)^2 \geqslant \mu\mu' \geqslant \frac{1-Q}{1+Q}(1+R)^2, \tag{41}$$

при чем оба левые неравенства обращаются в равенства, когда $\mu = \mu'$, а правые неравенства обращаются в равенства, когда $\lambda = \lambda'$; поэтому, когда $(A) = (B)$, то λ и λ' являются корнями уравнения $Z^2 - \frac{2(1-R)(1-RQ)}{1-Q}Z + \frac{1+Q}{1-Q}(1-R)^2 = 0$, а $\mu = \mu' = 1 - R$; напротив, когда $(A) = (\bar{B})$, то $\lambda = \lambda' = 1 + R$, а μ и μ' являются корнями уравнения:

$$Z^2 - \frac{2(1+R)(1-RQ)}{1+Q}Z + \frac{1-Q}{1+Q}(1+R)^2 = 0.$$

Если одновременно $\lambda = \lambda'$ и $\mu = \mu'$, то $\lambda = \lambda' = 1 + R$, а $\mu = \mu' = 1 - R$; но при $R \lessgtr 0$ это имеет место лишь при $(A) = (B) = (\bar{B}) = \frac{1}{2}$ [тогда и неравенство (36) обращается в равенство $Q = \frac{2R}{1+R^2}$].

Вывести из (40) и (41), что λ и λ' всегда удовлетворяют неравенству:

$$\frac{1-R}{1-Q}\left[1 - QR - \sqrt{Q[Q(1+R^2) - 2R]}\right] \leqslant \lambda \leqslant$$
$$\leqslant \frac{1-R}{1-Q}\left[1 - QR + \sqrt{Q[Q(1+R^2) - 2R]}\right].$$

ГЛАВА ЧЕТВЕРТАЯ

РАСПРОСТРАНЕНИЕ ТЕОРЕМЫ УМНОЖЕНИЯ ВЕРОЯТНОСТЕЙ НА ПРОИЗВОЛЬНОЕ ЧИСЛО СОБЫТИЙ

1. Теорема умножения вероятностей может быть распространена на совмещения какого угодно числа событий. Рассмотрим сначала случай независимых событий $A_1, A_2, \ldots, A_k$. Мы говорим, что события $A_1, A_2, \ldots, A_k$ независимы между собой, если вероятность наступления каждого из них не меняет своего значения после того, как одно или несколько из остальных событий осуществились. Образцами таких событий могут служить появления белого шара из каждой из k данных урн, из которых вынимается по одному шару. В таком случае вероятность (A_1, A_2, A_3) совмещения трех событий равна

$$(A_1, A_2, A_3) = (A_1, A_2) \cdot (A_3) = (A_1)(A_2)(A_3).$$

Точно так же, чтобы получить вероятность совмещения четырех событий A_1, A_2, A_3, A_4 замечаем, что

$$(A_1, A_2, A_3, A_4) = (A_1, A_2, A_3) \cdot (A_4) = (A_1)(A_2)(A_3)(A_4),$$

так как, по предположению, вероятность события A_4 сохраняет свое первоначальное значение и после того, как становится известным, что факты A_1, A_2, A_3 осуществились. Итак, повторяя то же рассуждение сколько угодно раз, находим, что вероятность совмещения k независимых событий равна произведению вероятностей всех этих событий.

Например: в трех ящиках находится по 10 билетиков с номерами от 1 до 10, из каждого ящика вынимается по одному билетику; какова вероятность, что все три вынутые билета будут иметь четные номера? Так как вероятность появления четного номера из одного

ящика равна $\frac{5}{10}=\frac{1}{2}$ и все три вынимания независимы, то появления трех четных номеров равно $\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)=\left(\frac{1}{2}\right)^3=\frac{1}{8}$.

Предположим, что, как в рассмотренном примере, производится одновременно или повторяется n однородных независимых опытов, при которых вероятности $(A_1)=(A_2)=\ldots=(A_n)=p$; тогда вероятность, что все события $A_1, A_2, \ldots, A_n$ осуществятся, равна произведению $(A_1)\ldots(A_n)=p\cdot p\ldots p=p^n$

Таким образом, если в отдельном опыте вероятность события A равна p, то при повторении опыта n раз (полагая, что результаты предшествующих опытов ни в какой мере не влияют на результат последующих) вероятность, что события произойдут при всех n опытах, равна p^n.

Ввиду того, что $p<1$, если событие не достоверно, мы видим что p^n уменьшается с возрастанием n, а потому, чем n больше, тем менее вероятно, что событие A будет повторяться при всех n опытах; кроме того, при неограниченном возрастании числа n, вероятность p^n повторения A при n опытах стремится к нулю.

Предположим, например, что вероятность в данном населении новорожденному быть мальчиком [1]) $p=\frac{1}{2}$; следовательно, вероятность, что из 10 новорожденных все десять будут мальчиками, равна $p^{10}=\frac{1}{1024}\neq 0{,}001$; таким образом в городе, где в течение недели регистрируется около 10 рождений, маловероятно, но возможно, что в течение недели не родится ни одной девочки. Однако в продолжение года, когда число рождений в данном городе достигнет 500, вероятность, что все новорожденные окажутся мальчиками, равна $\left(\frac{1}{2}\right)^{500}\neq\frac{1}{10^{50}}$, и мы можем считать этот столь маловероятный факт практически невозможным, если только не будет найдено средство искусственно влиять на пол новорожденного.

---

[1]) В зависимости от групп населения p может немного колебаться, но обычно p немного больше половины (около 0,51).

Возьмем другой пример. Ребенку, не знающему азбуки, предлагают разложить одну за другой карточки с буквами, которые он вынимает из ящика, где имеются поровну в весьма большом количестве одинаковые карточки с одной из 30 букв русского алфавита, так что вероятность вынуть ту или иную определенную букву равна $\frac{1}{30}$ (для того, чтобы вероятность появления каждой буквы все время оставалась равна $\frac{1}{30}$, можно предположить, что кто-нибудь сейчас же после появления данной буквы вкладывает в ящик карточку с той же буквой, тщательно размешивая все карточки). Тогда вероятность, что сложивши первые 5 вынутых карточек ребенок составит слово Париж равна $\left(\frac{1}{30}\right)^5 = \frac{1}{24\,300\,000}$; это весьма мало вероятно, но не невозможно. Однако вероятность, что, сложивши 20 000 букв, он составит определенное сочинение, например лекцию Эйнштейна „Эфир и принцип относительности", равна $\frac{1}{30^{29\,000}} < \frac{1}{10^{29000}}$ (т.-е. меньше десятичной дроби с 29 000 нулями перед первой значащей цифрой), и, хотя теоретически такая комбинация букв не является безусловно невозможной, однако практически мы считаем это столь же невозможным, как то, что вода в кастрюле, стоящей на горячей плите, обратится в лед, а вода в сосуде, поставленном в снег, закипит.

Таким образом факты, имеющие чрезвычайно малую математическую вероятность, можно считать практически невозможными.

Однако этому утверждению нельзя придавать абсолютного значения во избежание логических противоречий. Действительно, пусть лотерея содержит 1 000 000 билетов при одном выигрыше. Если у меня есть только один билет, то вероятность выигрыша 0,000 001 ничтожна; однако если бы я считал свой выигрыш невозможным, то пришел бы к противоречию, так как один из миллиона всех держателей билетов, которые находятся в таком же положении как и я, несомнено выиграет. Поэтому утверждение, что факт A, вероятность которого равна $\frac{1}{N}$, где N — чрезвычайно большое число, практически невозможен, означает лишь следующее: одновременно

с фактом A мы должны рассматривать еще $N-1$ физически эквивалентных ему фактов, и из всей этой огромной совокупности N фактов в действительности происходит только один, вся же остальная масса фактов не осуществляется.

Точно также, если вероятность p факта A очень близка к единице, то для отдельных опытов появление A почти достоверно; однако и тогда вероятность p^n, что факт повторится при всех n данных опытах, становится весьма малой, когда n велико. Пусть, например, вероятность A равна $p=0{,}999=1-\frac{1}{1\,000}$; тогда при $n=1\,000$ опытах вероятность, что A будет повторяться постоянно, равна $p^n=\left(1-\frac{1}{1\,000}\right)^{1\,000}\neq e^{-1}\neq 0{,}367\,9$; вероятность же, что A хоть раз не произойдет, равна $1-p^n\neq 0{,}632\,1$, т.-е. почти вдвое больше; а если мы произведем $n=10\,000$ опытов, то вероятность, что A всегда будет происходить, равна всего $\left(1-\frac{1}{1\,000}\right)^{10\,000}\neq e^{-10}\neq 0{,}000\,04=\frac{1}{25\,000}$, и при 100 000 опытах вероятность постоянного повторения A равна $\left(1-\frac{1}{1\,000}\right)^{100\,000}\neq$ $\neq e^{-100}\neq\frac{1}{10^{43}}$, т.-е. практически невозможна.

Вообще, как бы близка к 1 ни была вероятность $p=1-\frac{1}{N}$ факта A в отдельном из однородных опытов, вероятность его постоянного появления при n указанных опытах будет значительна только тогда, когда $\frac{n}{N}$ весьма малая дробь; но, по мере увеличения $\frac{n}{N}$, эта вероятность быстро убывает, становясь практически равной нулю, когда $\frac{n}{N}$ достигает нескольких десятков.

2. Если события A_1, A_2, A_3 зависимы между собой, то во всяком случае, применяя дважды теорему умножения вероятностей, мы находим, что вероятность (A_1, A_2, A_3) совмещения их определится формулой

$$(A_1, A_2, A_3)=(A_1, A_2)\cdot(A_3)_{A_1A_2}=(A_1)(A_2)_{A_1}(A_3)_{A_1A_2},$$

где $(A_3)_{A_1A_2}$ означает вероятность события A_3 после того, как оба события A_1 и A_2 осуществились. Точно также, рассматривая совме-

ние четырех событий A_1, A_2, A_3, A_4 как осуществление четвертого из них при совместном наступлении первых трех, находим, что

$$(A_1, A_2, A_3, A_4) = (A_1, A_2, A_3)(A_4)_{A_1A_2A_3} =$$
$$= (A_1)(A_2)_{A_1}(A_3)_{A_1A_2}(A_4)_{A_1A_2A_3},$$

и аналогичным образом получим формулу для вероятности совмещения любого числа событий в виде произведения вероятностей всех этих событий, при чем входящая множителем вероятность каждого последующего события A_k берется не первоначальная, а та, которая соответствует предположению, что все предыдущие события $A_1, A_2, \ldots, A_k$ осуществлены. Условие независимости событий, которое было указано выше, означает, что

$$(A_i)_{A_k A_l \ldots A_s} = (A_i),$$

каковы бы ни были различные между собой числа $i, k, \ldots, s$, принимающие всевозможные значения от 1 до n, где n есть число всех событий: так, в случае $n = 2$,

$$(A_1)_{A_2} = (A_1),\ (A_2)_{A_1} = (A_2);$$

в случае $n = 3$

$$\begin{aligned} (A_1)_{A_2A_3} &= (A_1)_{A_3} = (A_1)_{A_2} = (A_1), \\ (A_2)_{A_1A_3} &= (A_2)_{A_1} = (A_2)_{A_3} = (A_2), \\ (A_3)_{A_1A_2} &= (A_3)_{A_1} = (A_3)_{A_2} = (A_3) \end{aligned} \tag{42}$$

и т. п.

Вообще, подсчитывая, сколько различных вероятностей можно было бы приписать каждому определенному событию A_i в зависимости от того, какие из прочих событий считаются осуществившимися, видим, что [кроме первоначальной вероятности (A_i)] будет $n - 1 = C^1_{n-1}$ значение, соответствующее $(A_i)_{A_k}$ при всевозможных k, $C^2_{n-1} = \frac{(n-1)(n-2)}{2}$ значений вида $(A_i)_{A_kA_l}$ и т. д., т.-е. всего

$$C^1_{n-1} + C^2_{n-1} + \ldots + C^{n-1}_{n-1} = 2^{n-1} - 1$$

значений, которые в случае независимости должны быть равны (A_i). Но так как i может получать n различных значений, то условие независимости выражается $n(2^{n-1} - 1)$ равенствами.

Однако не все эти равенства независимы; можно доказать, что многие из них являются следствиями из других. Мы в этом уже убедились для $n=2$, когда показали, что $(A_1)_{A_2}=(A_1)$ есть следствие из равенства $(A_2)_{A_1}=(A_2)$. Точно также из 9 равенств (42) только 4, например, $(A_2)_{A_1}=(A_2)$ и $(A_3)_{A_1A_2}=(A_3)_{A_1}=(A_3)_{A_2}=(A_3)$ независимы, остальные же 5 равенств являются необходимыми следствиями из предыдущих: действительно, ввиду упомянутого результата для $n=2$, мы видим сразу, что $(A_1)_{A_2}=(A_1)$ $(A_1)_{A_3}=(A_1)$, $(A_2)_{A_3}=(A_2)$; поэтому $(A_1, A_2, A_3)=(A_1)\ (A_2)$, $(A_3)=$ $=(A_1)\ (A_3)\ (A_2)_{A_1A_3}=(A_2)\ (A_3)\ (A_1)_{A_2A_3}$, откуда

$$(A_2)_{A_1A_3}=(A_2) \text{ и } (A_3)_{A_1A_2}=(A_3).$$

Вообще аналогичным образом можно проверить, что условие, необходимое и достаточное для того, чтобы n событий $A_1, A_2, \ldots, A_n$ были независимые между собой, состоит в том, чтобы были соблюдены равенства

$$(A_i, A_k)=(A_i)\ (A_k);\ (A_i, A_k, A_l)=(A_i)\ (A_k)\ (A_l) \text{ и т. д.},$$

т.-е. чтобы вероятности совмещений любых нескольких из рассматриваемых фактов были равны произведениям первоначальных вероятностей этих фактов.

Так как число попарных совмещений есть число сочетаний из n по 2, т.-е. C_n^2, а совмещений по 3 может быть C_n^3 и т. д., то общее число указанных независимых равенств составляет

$$C_n^2+C_n^3+\ldots+C_n^n=2^n-n-1.$$

События называют только попарно независимыми, если соблюдены лишь условия $(A_iA_k)=(A_i)\ (A_k)$ при любых i, k; события называют независимыми только по 3, если $(A_i, A_k, A_l)=(A_i)\ (A_k)\ (A_l)$ и т. д.

Небесполезно уяснить на примере, что попарная независимость между тремя, например, событиями отнюдь не означает, что эти события совершенно независимы, т.-е. возможно, что

$$(A_1, A_2)=(A_1)\ (A_2),\ (A_1, A_3)=(A_1)\ (A_3),\ (A_2, A_3)=(A_2)\ (A_3),$$

а между тем $(A_1, A_2, A_3)\gtrless(A_1)\ (A_2)\ (A_3)$. А именно, рассмотрим пример, в котором $(A_1)=(A_2)=(A_3)=\frac{1}{2}$, $(A_1, A_2)=(A_1, A_3)=(A_2, A_3)=\left(\frac{1}{2}\right)^2=\frac{1}{4}$ (так что события попарно независимы), в то время как совмещение всех трех событий вместе невозможно, т.-е. $(A_1, A_2, A_3)=0$ $\left(\text{вместо } \frac{1}{8}\right)$.

Положим, что в ящике находятся 4 одинаковых билетика, так что вероятность вынуть каждый из них одна и та же; пусть номера этих билетов будут соответственно 112, 121, 211 и 222. Назовем событием A_1 цифру 1 на месте сотен, событием A_2 цифру 1 на месте десятков и событием A_3 цифру 1 на месте единиц в вынутом номере; в таком случае $(A_1)=\frac{2}{4}=\frac{1}{2}$,

так как событие A_1 произойдет, если будет вынут первый или второй билетик, точно также $(A_2)=(A_3)=\frac{1}{2}$, так как событие A_2 наступит, если будет вынут первый или третий билет, а A_3 наступит, если будет вынут второй или третий билет. С другой стороны, $(A_1, A_2)=\frac{1}{4}$, так как цифры сотен и десятков обе равны 1 только в первом числе, и аналогичным образом $(A_1, A_3)=(A_2, A_3)=\frac{1}{4}$; но, очевидно, $(A_1, A_2, A_3)=0$, так, как нет среди наших чисел ни одного, в котором все три цифры были бы единицами.

3. Упражнения.

1) Какова вероятность P при бросании трех игральных костей, что на одной по крайней мере окажется 6 очков?

Ответ. $P=1-\left(\frac{5}{6}\right)^3=\frac{91}{216}$.

2) Доказать, что условие, необходимое и достаточное для того, чтобы неотрицательные числа $p_1, p_2, p_3, p_{12}, p_{13}, p_{23}, p_{123}$ могли быть соответственно вероятностями некоторых событий A_1, A_2, A_3 и их совмещений по 2 и по 3, является соблюдение неравенств

$$p_{ik} \geqslant p_{123},\ p_i + p_{123} \geqslant p_{ik} + p_{il},$$

$$1 + p_{12} + p_{13} + p_{23} \geqslant p_1 + p_2 + p_3 + p_{123}, \tag{43}$$

где i, k могут быть равны 1, 2, 3 (так что имеется по три неравенства первого и второго вида).

Ответ. Если обозначим через $x_0=(A_1, A_2, A_3)$, $x_1=(\overline{A}_1, A_2, A_3)$, $x_2=(A_1, \overline{A}_2, A_3)$, $x_3=(A_1, A_2, \overline{A}_3)$, $x_{12}=(\overline{A}_1, \overline{A}_2, A_3)$, $x_{13}=(\overline{A}_1, A_2, \overline{A}_3)$, $x_{23}=(A_1, \overline{A}_2, \overline{A}_3)$, $x_{123}=(\overline{A}_1, \overline{A}_2, \overline{A}_3)$, то $x_0=p_{123}$, $x_l=p_{ik}-p_{123}$, $x_{kl}=p_i-p_{ik}-p_{il}+p_{123}$, $x_{123}=1-p_1-p_2-p_3+p_{12}+p_{13}+p_{23}-p_{123}$, откуда получаем искомые неравенства, выражая, согласно основному принципу первой главы, что все $x \geqslant 0$.

3) Какова вероятность p_{123} трем лицам A_1, A_2, A_3 встретиться всем троим в клубе, если вероятность каждому из них пойти в клуб $p_1=p_2=p_3=\frac{2}{3}$, вероятность же встречи попарно A_1 и A_2 равна $p_{12}=\frac{1}{2}$, а вероятности $p_{13}=\frac{5}{9}$, $p_{23}=\frac{4}{9}$?

Ответ. Из неравенств (43) получаем $\frac{4}{9} \geqslant p_{123} \geqslant \frac{7}{18}$. Наибольшее значение $p_{123}=\frac{4}{9}$ соответствует $(\overline{A}_1, A_2, A_3)=0$, т.-е. случаю, когда присутствие последних двух лиц всегда влечет за собой появление первого; напротив, наименьшее значение $p_{123}=\frac{7}{18}$ соответствует $(A_1, \overline{A}_2, \overline{A}_3)=0$, т.-е. невозможности появления первого, когда отсутствуют оба последние лица.

4) Доказать, что условия, необходимые и достаточные для того, чтобы неотрицательные числа p_1, p_2, p_3, p_{12}, p_{13}, p_{23} могли быть соответствен-

но вероятностями некоторых событий A_1, A_2, A_3 и их попарных совмещений, состоит в выполнении 13 неравенств:

$$p_i \geqslant p_{ik};\ 1 + p_{ik} \geqslant p_i + p_k;\ p_i + p_{kl} \geqslant p_{il} + p_{ik};$$
$$1 + p_{12} + p_{13} + p_{23} \geqslant p_1 + p_2 + p_3. \tag{44}$$

Ответ. Первая группа неравенств получается сложением неравенств (43) $p_{ik} \geqslant p_{123}$ и $p_i + p_{123} \geqslant p_{ik} + p_{il}$; вторая группа получается сложением неравенств $p_i + p_{123} \geqslant p_{ik} + p_{il}$ и $1 + p_{12} + p_{13} + p_{23} \geqslant p_1 + p_2 + p_3 + p_{123}$; третья группа получается сложением $p_{lk} \geqslant p_{123}$ и $p_i + p_{123} \geqslant p_{ik} + p_{li}$; наконец, последнее неравенство получается, если в последнем из неравенств (43) положить $p_{123} = 0$. Таким образом неравенства (44) являются необходимым следствием неравенств (43). Наоборот, если неравенства (44) осуществлен, то соответствующим выбором $p_{123} \geqslant 0$ возможно удовлетворить неравенствам (43), так как достаточно положить p_{123} равным наименьшему из неотрицательных чисел p_{ik} и $1 - p_1 - p_2 - p_3 + p_{12} + p_{13} + p_{23}$. Таким образом, если бы в предыдущем примере мы взяли $p_{23} = \frac{1}{3}$ вместо $\frac{4}{9}$, то мы получили бы неприемлемые данные, так как неравенство $p_1 + p_{23} \geqslant p_{12} + p_{13}$ не было бы соблюдено (если при $p_1 = p_2 = p_3 = \frac{2}{3}$ вероятность встречи A_1 и A_2 равна $p_{12} = \frac{1}{2}$ и $p_{13} = \frac{5}{9}$, то вероятность встречи A_2 и A_3 не может быть меньше, чем $\frac{7}{18}$ и не может быть больше, чем $\frac{11}{18}\Big)$.

5) Вывести из (43), что если события A_1, A_2, A_3 имеют вероятности p_1, p_2, p_3 и попарно независимы, то можно утверждать лишь, что $p_1 p_2 p_3 + q_1 q_2 q_3 \geqslant (A_1,\ A_2,\ A_3) \geqslant p_1 p_2 p_3 - p_i q_k q_l$, $p_i p_k \geqslant (A_1,\ A_2,\ A_3)$, где $q_i = 1 - p_i$, (i, k, l — любые из трех различных чисел 1, 2, 3).

Ответ. Последняя группа неравенств соответствует первой группе неравенств (43), а двойные неравенства равнозначны второму и третьему неравенствам. Отсюда видно, что $(A_1,\ A_2,\ A_3)$ может быть равно нулю только при $p_k p_l \leqslant q_k q_l$, так что, если бы в рассмотренном на странице 52 примере одно из значений $(A_k) = p_k$ было больше $\frac{1}{2}$, мы бы не могли получить $(A_1,\ A_2,\ A_3) = 0$.

6) Доказать, что 2^n чисел p_i, p_{ik}, p_{ikl}, ... могут представлять соответственно вероятности событий $A_1, A_2, \ldots, A_n$ и их совмещений по 2, по 3 и т. д. в том случае и только в том случае, если они удовлетворяют неравенствам вида:

$$\left.\begin{array}{ll}
1 - p_1 - p_2 \ldots - p_n + p_{12} + \ldots + p_{(n-1)n} - p_{123} - \ldots \geqslant 0 & \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad (1\ \text{неравенство}) & \\
p_i - p_{i1} - p_{i2} - \ldots - p_{in} + p_{i12} + \ldots \ldots \geqslant 0\ (n\ \text{неравенств}) & \\
p_{ik} - p_{ik1} - p_{ik2} \ldots\ldots\ldots \geqslant 0 \left(\frac{n(n-1)}{2}\ \text{неравенств}\right) & \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\
p_{12\ldots n} \geqslant 0 \qquad\qquad\qquad (1\ \text{неравенство}). &
\end{array}\right\} \begin{array}{c}(45)\\ \text{всего}\\ 2^n\ \text{неравенств.}\end{array}$$

Ответ. Это вытекает из того, что правая часть первого неравенства представляет собой вероятность ненаступления ни одного из данных событий; правая часть каждого из n следующих неравенств выражает соответственно вероятность наступления событий A_i при отсутствии остальных; следующая группа неравенств соответствует вероятности наступления двух событий A_i и A_k при отсутствии остальных и т. д.

7) Проверить, что неравенства (45) соблюдаются, если события $A_1, \ldots, A_n$ независимы, т.-е. если $p_{ikl\ldots} = p_i\, p_k\, p_l \ldots$

8) Доказать, что события $A_1, A_2, \ldots, A_n$ могут иметь попарно один тот же коэффициент совместимости λ, т.-е. при всяком i и k $(A_i, A_k) = \lambda p_i\, p_k$, если λ удовлетворяет неравенствам

$$1 \leqslant \lambda \leqslant \frac{1}{p_i}.$$

Ответ. Действительно, неравенствам (45) можно удовлетворить, положив $p_{ik} = \lambda\, p_i p_k$, $p_{ikl} = \lambda^2 p_i p_k p_l$ и т. д., так как все они (кроме первого) получат вид:

$$p_i(1-\lambda p_1)(1-\lambda p_2)\ldots(1-\lambda p_n) \geqslant 0,$$

$$\lambda p_i p_k\,(1-\lambda p_1)\,(1-\lambda p_2)\ldots(1-\lambda p_n) \geqslant 0.$$

Первое, после умножения на λ, можно написать в виде:

$$\lambda - 1 + (1-\lambda p_1)(1-\lambda p_2)\ \ldots\ (1-\lambda p_n) \geqslant 0.$$

Очевидно, что условие $\lambda \leqslant \frac{1}{p_i}$ является также необходимым, так как λp_i есть условная вероятность A_i; напротив, условие $\lambda \geqslant 1$ не является, вообще, необходимым, однако, если n бесконечно возрастает (а числа p_n не стремятся к нулю), то λ не может быть меньше 1. Действительно, умножая первое неравенство (45) на $\left(\sum p_i\right)^2$, неравенства второй группы на $\left(\sum p_i - 1\right)^2$, неравенства третьей группы на $\left(\sum p_i - 2\right)^2$ и т. д. и складывая полученные результаты почленно, находим

$$\left(\sum p_i\right)^2 - 2\left(\sum p_i\right)^2 + \sum p_i - 2\sum p_{ik} + 4\sum p_{ik} =$$
$$= \sum p_i + 2\sum p_{ik} - \left(\sum p_i\right)^2 \geqslant 0, \qquad (45\ \text{bis})$$

необходимое, вообще, следствие из (45), ибо все члены, содержащие p_{ikl}, p_{ik} и т. д. сокращаются благодаря тождествам

$$1 - (m-1) + \frac{(m-1)(m-2)}{1\cdot 2} - \cdots = 0,$$

$m > 1$, и

$$m - (m-1)^2 + \frac{(m-1)(m-2)^2}{1\cdot 2} - \cdots = 0,$$

$m > 2$.

Полагая в неравенстве (45 bis) $p_{ik}=\lambda p_i p_k$, получим

$$\sum p_i+2\lambda\sum p_i p_k-\left(\sum p_i\right)^2=\sum p_i+\lambda\left[\left(\sum p_i\right)^2-\sum p_i^2\right]-\left(\sum p_i\right)^2\geqslant 0,$$

откуда

$$\lambda\geqslant\frac{\left(\sum p_i\right)^2-\sum p_i}{\left(\sum p_i\right)^2-\sum p_i^2}>1-\frac{1}{\sum p_i}.$$

9) Доказать, что неотрицательные числа $p_1, p_2, \ldots, p_n$ тогда и только тогда могут быть вероятностями несовместимых по три и независимых попарно событий, когда соблюдены n неравенств:

$$p_1+p_2+\cdots+p_{n-1}\leqslant 1,$$

$$p_1+p_2+\cdots+p_{n-2}+p_n\leqslant 1,\ \ldots,\ p_2+p_3+\cdots+p_n\leqslant 1.$$

Ответ. Несовместимость по три означает, что $p_{ikl}=p_{iklm}=\cdots=0$, так что в данном случае уравнения (45) обращаются в n уравнений вида $p_i(1-p_2-\cdots-p_n)\geqslant 0$ и одно уравнение

$$1-(p_1+\cdots+p_n)+\sum p_i p_k\geqslant 0,$$

которое вытекает из предыдущих, так как ему можно дать форму

$$(1-p_n)(1-p_1-p_2-\cdots-p_{n-1})+\sum{}' p_i p_k\geqslant 0,$$

где $\sum'$ не содержит p_n.

10) Доказать, что для того, чтобы неотрицательные числа $p_1, p_2, \ldots, p_n$ могли быть вероятностями фактов несовместимых по три (т.-е. $p_{ikl}=0$), необходимо и достаточно, чтобы (кроме $p_i\leqslant 1$)

$$p_1+p_2+\ldots+p_n\leqslant 2.$$

Ответ. Необходимость этого условия доказывается сложением всех неравенств (45) второй группы с удвоенным первым неравенством. Обратно, если даны числа $p_1\geqslant p_2\geqslant\cdots\geqslant p_n$, удовлетворяющие данному условию можно построить ряд событий $A_1, A_2, \ldots, A_n$, несовместимых по три, чтобы $(A_i)=p_i$. Очевидно, достаточно ограничиться рассмотрением случая, когда $p_1+p_2+\ldots+p_n=2$; отсюда следует, что $2p_i\leqslant p_1+p_2+\ldots+p_n$; нужно показать, что при этом последнем условии можно подыскать неотрицательные числа p_{ik}, удовлетворяющие системе уравнений

$$p_1=\sum^n p_{1i},\ p_2=\sum^n p_{2i},\ \ldots,\ p_n=\sum^n p_{ni}.$$

Находим, полагая $p_{1,n}=p_n$, $p_{2,n}=0, \ldots, p_{n-1,n}=0$, что написанные n уравнений равноценны $(n-1)$ уравнениям: $p_1-p_n=\sum^{n-1} p_{1i}$, $p_2=\sum^{n-1} p_{2i}, \ldots$

той же формы, при чем $2(p_1 - p_n) \leqslant p_1 + p_2 + \ldots + p_{n-1} - p_n$ и $2p_i \leqslant p_1 + p_2 + \ldots + p_{n-1} - p_n$, где $i > 1$; поэтому, понижая последовательно число n, из правильности нашего утверждения для $n = 3$, заключаем о его правильности и для любого n. Пусть, например,

$$p_1 = \frac{3}{5},\ p_2 = \frac{1}{2},\ p_3 = \frac{2}{5},\ p_4 = \frac{3}{10},\ p_5 = \frac{1}{5};$$

полагая

$$p_{1,5} = p_5 = \frac{1}{5},\ p_{2,5} = p_{3,5} = p_{4,5} = 0,$$

получаем уравнения

$$p_1 - p_5 = \frac{2}{5} = p_{12} + p_{13} + p_{14},\ p_2 = \frac{1}{2} = p_{12} + p_{23} + p_{24},$$

$$p_3 = \frac{2}{5} = p_{13} + p_{23} + p_{34},\ p_4 = \frac{3}{10} = p_{14} + p_{24} + p_{34};$$

полагая затем

$$p_{24} = p_4 = \frac{3}{10},\ p_{14} = p_{34} = 0,$$

имеем

$$\frac{2}{5} = p_{12} + p_{13},\ \frac{1}{5} = p_{12} + p_{23},\ \frac{2}{5} = p_{13} + p_{23},$$

откуда

$$p_{13} = \frac{3}{10},\ p_{12} = p_{23} = \frac{1}{10}.$$

ЧАСТЬ ВТОРАЯ

ГЛАВНЕЙШИЕ МЕТОДЫ ВЫЧИСЛЕНИЯ МАТЕМАТИЧЕСКИХ ВЕРОЯТНОСТЕЙ

ГЛАВА ПЕРВАЯ

ПРИЕМ ПОДСЧЕТА РАВНОВОЗМОЖНЫХ СЛУЧАЕВ

1. В предшествующих главах мы установили теоретические основы для вычисления вероятностей одних фактов, когда известны вероятности некоторых других связанных с ними событий. Теперь мы займемся рассмотрением главнейших методов, которые применяются в тех или иных случаях, при чем возможно, конечно, для решения той же задачи пользоваться различными способами и, если только исходные вероятности были оценены одинаково и во время вычисления не было сделано неявно некоторых новых допущений, то все приемы, основанные исключительно на вышеустановленных положениях, приводят к одинаковым результатам.

Начнем с простейшего приема, который заключается в непосредственном применении определения математической вероятности события A, как отношения $\frac{m}{n}$ числа благоприятных событию A случаев, к общему числу всех единственно и равновозможных несовместимых случаев. Применение этого приема требует непосредственного подсчета обоих чисел m и n, при чем должно быть обращено особое внимание на то, чтобы вполне отчетливо указать, какие случаи считаются равновозможными.

Первый пример. В ящике имеется M белых и N черных шаров; вынимается n шаров; требуется определить, какова вероятность, что среди вынутых шаров будет l шаров белого цвета.

Полагая, что шары в ящике тщательно смешаны и не образуют скоплений из шаров определенного цвета, мы допускаем, что любая комбинация из n шаров одинаково вероятна. Принимая во внимание, что общее число шаров в ящике равно $M+N$, мы видим, что, если бы мы их мысленно перенумеровали, то появление любых n

нумеров было бы одинаково вероятно; таким образом, если бы вместо цвета мы различали шары нумерами, то число всех единственно и равновозможных несовместимых случаев при вынимании n шаров было бы равно

$$C^n_{M+N} = \frac{(M+N)!}{n!\,(M+N-n)!}.$$

Если мы вообразим себе все белые шары перенумерованными первыми M номерами, а черные — следующими N номерами, то благоприятствовать нашему событию A будут те и только те из рассмотренных случаев, которые соответствовали бы присутствию среди вынутых n шаров l таких, которые были бы нумерованными какими-нибудь из первых M номеров; это могли бы быть, например, нумера: $1,2,\ldots, l$ или $2,3,\ldots, l+1$ (полагая, что $l<M$) и т. п. Таким образом событию A благоприятствуют не только определенные l нумеров, а всевозможные сочетания по l из числа M, т.-е.

$$C^l_M = \frac{M!}{l!\,(M-l)!}.$$

Но, если мы возьмем определенную благоприятную для нас группу номеров (например, номера $1,2,\ldots, l$), то она встретится не в одном только из вышеупомянутых единственно возможных C^n_{M+N} случаев, так как одновременно с этими l шарами могут быть вынуты какие угодно $n-l$ черных шаров, т.-е. $n-l$ номеров из последующих N номеров: значит, каждой группе благоприятной событию A соответствует C^{n-l}_N различных случаев. Следовательно общее число всех благоприятных A случаев равно произведению

$$C^l_M \cdot C^{n-l}_N = \frac{M!\,N!}{l!\,(M-l)!\,(n-l)!\,(N-n+l)!};$$

поэтому искомая вероятность

$$(A) = \frac{C^l_M \cdot C^{n-l}_N}{C^n_{M+N}} = \frac{M!\,N!\,n!\,(M+N-n)!}{(M+N!)\,l!\,(n-l)!\,(M-l)!\,(N-n+l)!}. \qquad (46)$$

Пусть, например, $M=14$, $N=7$, $n=4$, $l=2$; тогда

$$(A) = \frac{14!\;7!\;4!\;17!}{21!\;2!\;2!\;12!\;5!} = \frac{13\cdot 14\cdot 6\cdot 7\cdot 6}{18\cdot 19\cdot 20\cdot 21} = \frac{91}{285}.$$

Второй пример. Дано n одинаковых ящичков, в которые бросают наудачу N дробинок, так что вероятность каждой дробинке попасть в каждый ящик одна и та же, независимо от того, сколько дробинок уже попало в данный ящичек. Нужно определить, какова вероятность, что в данный ящик попадает h дробинок.

Общее число всех равновозможных случаев равно здесь n^N, так как каждая дробинка может попасть в любой из ящичков (поскольку мы индивидуализируем каждую дробинку, давая ей мысленно определенное название или номер).

Благоприятными из них будут те из этих случаев, при которых в данный ящик попало h каких-нибудь дробинок; пусть это будет h определенных дробинок; тогда одновременно остальные $N-h$ дробинок могут произвольно занять оставшиеся $n-1$ место, и, следовательно, этому обстоятельству благоприятствует

$$(n-1)^{N-h},$$

из рассмотренных выше случаев. Но так как для нас не важно, чтобы в выбранный нами ящик попали определенные h дробинок, а это могут быть любые h из данных N дробинок, то подбор дробинок в указанный ящик может быть произведен C_N^h способами. Поэтому общее число благоприятных нашему событию случаев равно

$$C_N^h \cdot (n-1)^{N-h},$$

и следовательно искомая вероятность

$$P_h = \frac{C_N^h \cdot (n-1)^{N-h}}{n^N}. \tag{47}$$

Легко проверить, что, согласно требованию теоремы сложения,

$$\sum_{h=0}^{h=N} P_h = \left(\frac{n-1}{n}\right)^N \left[1 + C_N^1 \cdot \frac{1}{n-1} + C_N^2 \cdot \frac{1}{(n-1)^2} + \ldots\right] =$$

$$= \left(\frac{n-1}{n}\right)^N \left[1 + \frac{1}{n-1}\right]^N = 1.$$

Во многих приложениях формулы (47) (например, в теории разреженных газов) числа n и N весьма велики, при чем $\frac{N}{n}=k$ сохраняет определенное конечное значение, так что дробь $\frac{k}{n}$ весьма мала. Во всяком случае формулу (47) можем записать в виде:

$$P_h=\frac{kn(kn-1)\ldots(kn-h+1)}{h!\,(n-1)^h}\left(1-\frac{1}{n}\right)^{kn}.$$

Полагая затем, что n безгранично возрастает и замечая, что

$$\underset{n=\infty}{\text{пред.}}\left(1-\frac{1}{n}\right)^{kn}=e^{-k},$$

находим, для данного конечного числа h,

$$P_h'=\text{пред. } P_h=\frac{e^{-k}}{h!}\ \text{пред.}\ \frac{k\left(k-\frac{1}{n}\right)\ldots\left(k-\frac{h-1}{n}\right)}{\left(1-\frac{1}{n}\right)^h}=e^{-k}\cdot\frac{k^h}{h!}. \quad (47\ \text{bis})$$

Таким образом

$$P_0'=e^{-k},\ P_1'=ke^{-k},\ P_2'=\frac{k^2}{2}e^{-k},\ P_3'=\frac{k^3}{6}e^{-k}$$

и т. д.

Заметим, что

$$\sum_0^\infty P_h'=e^{-k}\left[1+k+\frac{k^2}{2!}+\ldots+\frac{k^n}{n!}+\ldots\right]=e^{-k}\cdot e^k=1.$$

Возвращаясь к случаю, когда N и n конечные числа, положим, например, что $N=n=8$; тогда [1])

$$P_0=\left(\frac{7}{8}\right)^8\neq 0{,}344,\quad P_1=\left(\frac{7}{8}\right)^7\neq 0{,}393,\quad P_2=\frac{1}{2}\left(\frac{7}{8}\right)^7\neq 0{,}197,$$

$$P_3\neq 0{,}056 \text{ и т. д.}$$

Таким образом в данном случае, взявши наудачу ящик, из всех чисел наиболее вероятное число дробинок, которое мы в нем найдем, это 1; немного менее вероятно, что мы не найдем ни одной дробинки (вероятность, что мы найдем более 1 дробинки равна только $1-P_0-P_1\neq 0{,}263$).

Аналогичным образом можем получить вероятность, что в первом ящике окажется h_1 дробинок, а во втором ящике — h_2 дробинок. Действительно, общее число всех равновозможных случаев

[1]) $h!$ при $h=0$ равен 1.

остается n^N. Если в первый ящик попали определенные h_1 дробинок, а во второй h_2 другие определенные дробинки, то остальные $N-h_1-h_2$ дробинок могли занять оставшиеся $n-2$ ящика произвольным образом, т.-е. $(n-2)^{N-h_1-h_2}$ способами. Но так как нам безразлично, какие дробинки попали в первые два ящика, то первые h_1 дробинок могут быть выбраны $C_N^{h_1}$ способами, и тогда останется (при каждом подборе первых h_1 дробинок) произвольный выбор h_2 дробинок из $N-h_1$ свободных еще дробинок. Поэтому число всех благоприятных случаев равно

$$C_N^{h_1} C_{N-h_1}^{h_2} (n-2)^{N-h_1-h_2},$$

а потому искомая вероятность

$$P_{h_1h_2} = C_N^{h_1} C_{N-h_1}^{h_2} \frac{(n-2)^{N-h_1-h_2}}{n^N} = \frac{N!\,(n-2)^{N-h_1-h_2}}{h_1!\,h_2!\,(N-h_1-h_2)!\,n^N};$$

например, если $n=N=8$ и $h_1=h_2=1$, то

$$P_{1,1} = 56\,\frac{6^6}{8^8} = \frac{5\,103}{32\,768} \neq 0{,}156.$$

Применяя то же рассуждение, видим, что вероятность $P_{h_1h_2h_3}$, что в первом ящике h_1 дробинки, во втором — h_2 и в третьем — h_3, равна

$$P_{h_1h_2h_3} = C_N^{h_1} C_{N-h_1}^{h_2} C_{N-h_1-h_2}^{h_3} \frac{(n-3)^{N-h_1-h_2-h_3}}{n^N} =$$

$$= \frac{N!\,(n-3)^{N-h_1-h_2-h_3}}{h_1!\,h_2!\,h_3!\,(N-h_1-h_2-h_3)!\,n^N}.$$

Продолжая тот же прием, находим, что вероятность того, что в первом ящике будет h_1, во втором — h_2 ... и наконец в n-ом ящике $h_n = N-h_1-h_2-\ldots-h_{n-1}$ дробинок, равна

$$P_{h_1h_2\ldots h_n} = \frac{N!}{h_1!\,h_2!\ldots h_n!} \cdot \frac{1}{n^N}$$

(числа $h_1, h_2, \ldots, h_n$ могут быть и не различны между собой).

Например, при $N=n=8$ вероятность, что во всех ящиках окажется по одной дробинке, равна

$$P=\frac{8!}{8^8}\neq 0{,}002\,403,$$

вероятность же, что в первых двух ящиках не будет ни одной дробинки, в следующих четырех ящиках по одной и в последних двух по две, равна

$$P'=\frac{8!}{4\,.\,8^8}\neq 0{,}000\,601;$$

поэтому вероятность, что вообще найдутся 2 пустых ящика, 4 ящика с 1 дробинкой и 2 ящика с двумя дробинками равна

$$\frac{8\cdot 7}{1\cdot 2}\cdot\frac{6\cdot 5}{1\cdot 2}\cdot P'\neq 0{,}252\,3,$$

так как два пустые ящика могут быть выбраны $C_8^2=\frac{8\cdot 7}{1\cdot 2}$ способами и после этого еще $C_6^4=\frac{6\cdot 5}{1\cdot 2}$ способами могут быть взяты ящики с одной дробинкой. Вообще, вероятность, что в каких-нибудь α_1 ящиках будет h_1 дробинок, в каких-нибудь других α_2 ящиках будет h_2 дробинок и т. д., равна

$$R=\frac{N!\,n!}{(h_1!)^{\alpha_1}(h_2!)^{\alpha_2}\ldots(h_k)^{\alpha_k}\,\alpha_1!\,\alpha_2!\ldots\alpha_k!}\,\frac{1}{n^N},$$

где $\alpha_1+\alpha_2+\ldots+\alpha_k=n$, $h_1\alpha_1+h_2\alpha_2+\ldots+h_k\alpha_k=N$ (числа $h_1, h_2, \ldots, h_k$ различны между собою).

При $n=2$

$$P_{h,\,N-h}=P_{N-h,\,h}=\frac{N!}{h!\,(N-h)!}\,\frac{1}{2^N},\quad R=\frac{N!}{h!\,(N-h)!}\,\frac{1}{2^{N-1}},$$

если $h\gtrless\frac{N}{2}$

Упражнения. 1) За круглым столом сидят 5 мужчин и 5 женщин; какова вероятность, что никакие два лица одинакового пола не сидят рядом, если

места занимаются случайно (т.-е. для каждого лица каждое место одинаково вероятно)?

Ответ. $P=2\frac{(5!)^2}{10!}=\frac{1}{126}$.

2) В три ящика бросают наудачу 12 шаров; требуется определить наиболее вероятное распределение шаров по ящикам.

Ответ. Наиболее вероятно, что в ящиках (независимо от их порядка) окажется в одном 3, в другом 4, в третьем 5 шаров; вероятность этого $R=\frac{12!}{4!\,5!}\cdot\frac{1}{3^{12}}=\frac{6\,160}{19\,683}\neq 0{,}313$; напротив, вероятность, что во всех ящиках будет поровну (по четыре шара) равна только $P=\frac{12!}{(4!)^3}\cdot\frac{1}{3^{12}}=\frac{3\,850}{59\,049}\neq 0{,}065$; из распределений, в которых в двух ящиках будет поровну, наиболее вероятно то, в котором два ящика содержат по 3 шара, а в третьем 6 шаров; вероятность этого равна $Q=3\frac{12!}{(3!)^2 6!}\cdot\frac{1}{3^{12}}=\frac{6\,160}{59\,049}\neq 0{,}104$; таким образом, даже $Q>P$ (следующей за R по величине является вероятность S, что в каком-нибудь одном ящике будет два шара, в другом 4 и в третьем 6: $S=\frac{1}{2}R\neq 0{,}156$).

3) Из колоды, состоящей из 52 карт, вынимают вместе 10 карт. Какова вероятность P_1, что среди вынутых карт будет хоть один туз? Какова вероятность P_2, что среди вынутых карт окажется не менее 2 тузов?

Ответ. $P_1=1-\frac{C_{48}^{10}}{C_{52}^{10}}=1-\frac{42\cdot 41\cdot 40\cdot 39}{52\cdot 51\cdot 50\cdot 49}=1-\frac{246}{595}=\frac{349}{595}$, так как всего равновозможных случаев столько, сколько существует сочетаний из 52 карт по 10; из них не содержат ни одного туза все сочетания по 10 из 48 карт колоды, лишенной тузов; поэтому вероятность, что не будет вынуто ни одного туза равна $\frac{C_{48}^{10}}{C_{52}^{10}}$, и, вычитая из 1 последнее число, находим P_1. Вероятность же, что будет вынут только 1 туз бубен, например, равна $\frac{C_{48}^{9}}{C_{52}^{10}}$, так как в группе из 10 карт, где есть туз бубен, имеется еще 9 каких-нибудь карт из 48 прочих карт (без тузов); то же самое относится к тузу любой масти, и поэтому вероятность p_1, что вообще будет вынут только 1 туз равна $p_1=4\frac{C_{48}^{9}}{C_{52}^{10}}=\frac{656}{1\,547}$. Следовательно вероятность, что будет не менее двух тузов, равна

$$P_2=P_1-p_1=\frac{349}{595}-\frac{656}{1\,547}=\frac{1\,257}{7\,735}.$$

4) Из ящика, содержащего n белых и m черных шаров ($n < m$) вынимают наудачу все шары под-ряд. Какова вероятность P, что будет момент, когда в ряду вынутых шаров будет поровну белых и черных?

Ответ. Число всех возможных рядов равно C^m_{n+m}; благоприятными будут, во-первых, все те ряды, которые начинаются белыми шарами, так как число черных шаров (которых больше чем белых) сравняется с числом белых после некоторого числа $2k$ выниманий; число таких рядов равно C^m_{n+m-1}. Но благоприятными будут также все те ряды, в которых указанная группа из $2k$ шаров, заканчивающаяся черным шаром и содержащая одинаковое число белых и черных шаров, будет перевернута. Поэтому

$$P = 2\frac{C^m_{n+m-1}}{C^m_{n+m}} = 2\frac{n}{n+m}.$$

ГЛАВА ВТОРАЯ

НЕПОСРЕДСТВЕННОЕ ПРИМЕНЕНИЕ ТЕОРЕМ СЛОЖЕНИЯ И УМНОЖЕНИЯ ВЕРОЯТНОСТЕЙ

1. Во многих вопросах нельзя просто связать данные события с некоторой совокупностью единственно и равновозможных несовместимых фактов, чтобы воспользоваться приемом простого подсчета, который был указан выше. В таком случае для определения вероятности данного события A стараются установить различные несовместимые благоприятствующие ему случаи $a_1, a_2, \ldots, a_k$ так, чтобы эти последние не были обязательно равновозможными, но только, чтобы вероятности их было бы возможно легко вычислить: тогда $(A) = (a_1) + (a_2 \ldots) + \ldots + (a_k)$; после этого каждый из случаев a_1, a_2 и т. д. рассматривают как совмещения некоторых других фактов, так что вероятности $(a_1), \ldots, (a_k)$ могут быть определены при помощи теоремы умножения вероятностей; если факты, совмещением которых являются соответственно события $a_1, a_2, \ldots$, имеют известные вероятности, то задача решена; в противном случае приходится продолжать тот же прием попеременного применения теорем сложения и умножения вероятностей.

Первый пример. Даны три урны; в первой m_1 белых шаров и n_1 черных, во второй m_2 белых и n_2 черных шаров, в третьей m_3 белых и n_3 черных шаров; вероятности, что шар будет выниматься соответственно из первой, второй или третьей урны равны α_1, α_2, α_3. Какова вероятность, что вынутый шар будет белым?

Искомая вероятность $P = (a_1) + (a_2) + (a_3)$, где (a_1) есть вероятность того, что белый шар будет вынут именно из первой урны, (a_2) вероятность того, что белый шар будет вынут из второй урны и (a_3) вероятность, что белый шар будет вынут из третьей урны. Но для того, чтобы белый шар был вынут из первой урны, необходимо совмещение двух фактов: вынимания из этой урны и чтобы

результатом вынимания оказалось появление белого шара; поэтому $(a_1) = a_1 \cdot \frac{m_1}{m_1 + n_1}$, так как, если известно, что вынимание произошло из первой урны, то вероятность появления белого шара становится равной отношению $\frac{m_1}{m_1 + n_1}$ числа белых шаров к общему числу шаров данной урны; таким же образом получаем

$$(a_2) = a_2 \frac{m_2}{m_2 + n_2} \text{ и } (a_3) = a_3 \frac{m_3}{m_3 + n_3}.$$

Следовательно

$$P = a_1 \frac{m_1}{m_1 + n_1} + a_2 \frac{m_2}{m_2 + n_2} + a_3 \frac{m_3}{m_3 + n_3}. \qquad (48)$$

Если бы, например, было дано три одинаковых урны, так что $a_1 = a_2 = a_3 = \frac{1}{3}$, при чем в первой только 1 белый шар, во второй 2 белых и 1 черный шар, а в третьей 1 белый и 4 черных шара, то

$$P = \frac{1}{3}\left[1 + \frac{2}{3} + \frac{1}{5}\right] = \frac{28}{45};$$

вероятность же появления черного шара $Q = \frac{17}{45}$. Таким образом появление белого шара более вероятно, чем появление черного, хотя белых шаров во всех урнах всего 4, а черных 5: это объясняется тем, что не все шары имеют одинаковую вероятность быть вынутыми, а именно шар первой урны имеет вероятность $\frac{1}{3}$, в то время как каждый шар второй урны имеет вероятность $\frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$, а каждый шар последней урны (где преобладают черные шары) имеет только вероятность $\frac{1}{3} \cdot \frac{1}{5} = \frac{1}{15}$.

2. *Второй пример.* В теории наследственности очень часто находит применение следующий закон, открытый ботаником Менделем. Пусть даны 2 чистые (по отношению к некоторому признаку) расы индивидов, т.-е. такие, что скрещивание между собой индивидов одной расы всегда приводит к индивидам той же расы;

согласно закону Менделя, если индивид чистой расы A скрещивается с индивидом чистой расы B, то получается всегда индивид нового класса (гибрид) C; при скрещивании гибридов C между собой могут получаться индивиды всех трех типов, при чем вероятность рождения индивида гибридного класса C равна $\frac{1}{2}$, вероятности же новорожденному принадлежать к каждой из чистых рас A или B соответственно равны $\frac{1}{4}$; наконец, при скрещивании индивида A с индивидом C одинаково вероятно рождение индивида A и индивида C, а индивид B появиться не может; точно так же при скрещивании другой чистой расы B с гибридом C вероятности рождения индивида расы B и C равны, а появление индивида класса противоположной чистой расы A невозможно. Предположим, что в данном населении имеются индивиды всех трех классов, при чем вероятности, что взятый наудачу индивид (как мужского, так и женского пола) принадлежит к классу A, B или C соответственно равны α, β, γ. Спрашивается, каковы соответственные вероятности, что, скрещивая двух взятых наудачу индивидов мужского и женского пола, мы получим индивид, принадлежащий к классу A, к классу B или к классу C.

На основании теоремы умножения вероятностей, вероятность (A, A), что оба взятые индивида принадлежат к классу A равна $\alpha . \alpha = \alpha^2$; вообще, $(A, A) = \alpha^2$, $(B, B) = \beta^2$, $(C, C) = \gamma^2$. Вероятность же, что мужской индивид принадлежит к классу A, а женский к классу B равна $(A, B) = \alpha\beta$; точно так же $(B, A) = \alpha\beta$, и таким же образом находим, что

$$(A, C) = (C, A) = \alpha\gamma \text{ и } (B, C) = (C, B) = \beta\gamma.$$

Какова же в таком случае вероятность α_1, что новорожденный принадлежит к классу A? Это может произойти только в одном из четырех несовместимых случаев: 1) оба родителя относятся к классу A, 2) отец класса A, а мать класса C, 3) отец класса C, мать класса A, 4) оба родителя класса C. Если оба родителя класса A, то новорожденный всегда принадлежит к классу A; поэтому вероятность совмещения рождения индивида расы A с фак-

что оба родителя принадлежат к той же расе равна α^2. Далее, так как вероятность, что отец класса A, а мать класса C, равна $\alpha\gamma$, а после того как природа родительской пары таким образом определена, вероятность появления индивида класса A по закону Менделя становится равной $\frac{1}{2}$, заключаем, что совмещение такой родительской пары с появлением особи класса A равна произведению $\alpha\gamma \cdot \frac{1}{2} = \frac{1}{2}\alpha\gamma$; совершенно так же убеждаемся, что совмещение родительской пары, в которой отец класса C, а мать класса A, с фактом рождения индивида класса A тоже равна $\frac{1}{2}\alpha\gamma$. Наконец, если становится известным, что оба родителя относятся к классу гибридов C, то по закону Менделя вероятность новорожденного принадлежать к классу A равна $\frac{1}{4}$; следовательно вероятность последнего совмещения этих двух обстоятельств равна $\frac{1}{4}\gamma^2$.

Беря сумму всех этих вероятностей

$$\alpha^2 + \frac{1}{2}\alpha\gamma + \frac{1}{2}\alpha\gamma + \frac{1}{4}\gamma^2,$$

получим искомую вероятность α_1, что новорожденный (от неизвестных родителей) будет принадлежать к классу A:

$$\alpha_1 = \alpha^2 + \alpha\gamma + \frac{\gamma^2}{4} = \left(\alpha + \frac{\gamma}{2}\right)^2; \tag{49}$$

так как закон Менделя симметричен по отношению к обеим чистым расам, то достаточно в найденной формуле (49) заменить α через β, чтобы получить вероятность β_1 новорожденному принадлежать к расе B:

$$\beta_1 = \beta^2 + \beta\gamma + \frac{\gamma^2}{4} = \left(\beta + \frac{\gamma}{2}\right)^2. \tag{49 bis}$$

Можно было бы аналогичным образом, рассмотревши все родительские пары, которые могут родить гибрида, получить значение γ_1 вероятности, что новорожденный будет принадлежать к классу гибридов C. То же значение γ_1 можем получить и другим способом, заметивши, что

$$\gamma_1 = 1 - \alpha_1 - \beta_1, \tag{50}$$

так как новорожденный может принадлежать только к одному из трех упомянутых классов. Точно так же и

$$\alpha+\beta+\gamma=(\alpha+\beta+\gamma)^2=1.$$

Следовательно формула (50) преобразуется в

$$\gamma_1=(\alpha+\beta+\gamma)^2-\left(\alpha+\frac{\gamma}{2}\right)^2-\left(\beta+\frac{\gamma}{2}\right)^2=$$

$$=2\alpha\beta+\alpha\gamma+\beta\gamma+\frac{\gamma^2}{2}=2\left(\alpha+\frac{\gamma}{2}\right)\left(\beta+\frac{\gamma}{2}\right). \quad (50\ \text{bis})$$

Таким образом вероятности индивиду второго поколения (любого пола) принадлежать соответственно классам A, B и C определяются значениями α_1, β_1, γ_1, полученными из формул (49), (49 bis) и (50 bis).

Весьма замечательно, что, скрещивая наудачу индивидов второго поколения, мы находим, что и в третьем поколении вероятности α_2, β_2, γ_2 индивиду принадлежать к каждому из классов A, B, C остаются те же, что и во втором поколении, т.-е. $\alpha_2=\alpha_1$, $\beta_2=\beta_1$, $\gamma_2=\gamma_1$.

Действительно, мы имеем, очевидно,

$$\alpha_2=\left(\alpha_1+\frac{\gamma_1}{2}\right)^2,\ \beta_2=\left(\beta_1+\frac{\gamma_1}{2}\right)^2,$$

$$\gamma_2=2\left(\alpha_1+\frac{\gamma_1}{2}\right)\left(\beta_1+\frac{\gamma_1}{2}\right); \quad (51)$$

подставляя в первую из формул (51) значения α_1 и γ_1 из формул (49) и (50 bis), получаем

$$\alpha_2=\left[\left(\alpha+\frac{\gamma}{2}\right)^2+\left(\alpha+\frac{\gamma}{2}\right)\left(\beta+\frac{\gamma}{2}\right)\right]^2=$$

$$=\left[\left(\alpha+\frac{\gamma}{2}\right)\left(\alpha+\beta+\gamma\right)\right]^2=\left(\alpha+\frac{\gamma}{2}\right)^2=\alpha_1.$$

Таким же образом легко проверить, что и $\beta_2=\beta_1$, $\gamma_2=\gamma_1$. Очевидно, что найденные вероятности сохраняются во всех последующих поколениях. Например, если в первом поколении $\alpha=\beta=\gamma=\frac{1}{3}$, то во всех последующих поколениях

$$\alpha_1=\left(\frac{1}{3}+\frac{1}{6}\right)^2=\frac{1}{4},\quad \beta_1=\frac{1}{4},\quad \gamma_1=\frac{1}{2}.$$

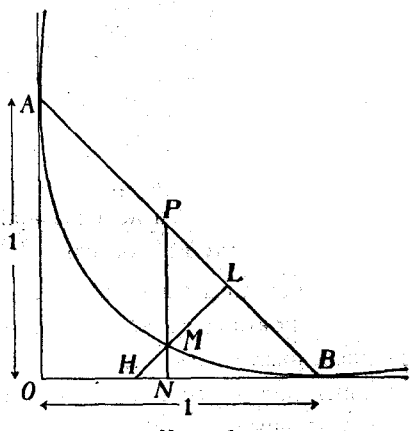
Из равенств (51) видно, что при устанавливающемся со 2-го поколения распределении вероятностей величины α_n, β_n, γ_n не могут быть произвольными (не считая соотношения $\alpha_n + \beta_n + \gamma_n = 1$), но должны удовлетворять равенству

$$4\,\alpha_n \beta_n = \gamma_n^2; \tag{51 bis}$$

исключая из (51 bis) $\gamma_n = 1 - \alpha_n - \beta_n$, находим (извлекая корень из обеих частей равенства) после простых преобразований, что

$$\sqrt{\alpha_n} + \sqrt{\beta_n} = 1. \tag{52}$$

Можно указать следующий простой графический прием (черт. 1) для нахождения вероятностей $\alpha_n = \alpha_1$, $\beta_n = \beta_1$, $\gamma_n = \gamma_1$, если дано первоначальное распределение вероятностей α, β, γ. Уравнение (52), в котором α_n и β_n будем рассматривать как абсциссу и ординату точки M, представляет дугу параболы, касательной к обеим осям координат в точках A и B, отстоящих от начала координат на расстоянии 1. Таким образом всякому установившемуся (со 2-го поколения) распределению вероятностей соответствует некоторая точка M этой дуги, при чем абсцисса $ON = \alpha_n$, ордината $MN = \beta_n$, и отрезок между прямой линией AB и параболой, $MP = \gamma_n$, так как $\gamma_n = 1 - \alpha_n - \beta_n$. Чтобы получить точку M, соответствующую данному первоначальному распределению (α, β, γ), замечаем из (49) и (49 bis), что



Черт. 1.

$$\alpha_1 - \beta_1 = \left(\alpha + \frac{\gamma}{2}\right)^2 - \left(\beta + \frac{\gamma}{2}\right)^2 = (\alpha - \beta)\,(\alpha + \beta + \gamma) = \alpha - \beta;$$

следовательно разность $\alpha_n - \beta_n$ сохраняет во всех поколениях свое первоначальное значение; поэтому, откладывая на оси абсцисс отрезок $OH = \alpha - \beta$ и проводя через H прямую HL перпендикулярно к AB, получим искомую точку M на пересечении этой прямой с нашей параболой (очевидно, что та же точка M соответствует всякой первоначальной паре значений α, β, представляющих координаты любой точки прямой HL).

3. *Третий пример.* Производится n независимых опытов; при каждом опыте вероятность появления события A равна p. Требуется определить вероятность $P_{m,n}$, что событие A повторится при этом ровно m раз.

Для случая $m=n$ эта задача нами была уже раньше решена: мы нашли $P_{n,n}=p^n$; точно так же, если положим $q=1-p$, где q есть вероятность ненаступления события A в отдельном опыте, то вероятность $P_{0,n}$, что событие A ни разу не произойдет, равна q^n. Если же m какое угодно число $(0<m<n)$, то можем рассуждать следующим образом. Пусть $A\bar{A}\bar{A}A\ldots A$ будет некоторая определенная последовательность появлений и непоявлений события A ($\bar{A}$ означает непоявление события); каждая такая последовательность представляет собой совмещение n независимых фактов, вероятности которых нам известны: $(A)=p$, $(\bar{A})=q$. Чтобы получить вероятность рассматриваемого совмещения, мы должны перемножить вероятности всех составляющих фактов, и так как рассматриваемая последовательность будет относиться к числу тех, при которых событие A осуществляется ровно m раз, если символ A встречается в ней m раз, а символ $\bar{A}$ остальные $n-m$ раз, то, следовательно, вероятность такой вполне определенной последовательности появлений и непоявлений нашего события равна $p^m q^{n-m}$. Но для того, чтобы событие A повторилось m раз, безразлично в какой последовательности, нужно только, чтобы в ряду $A\bar{A}\bar{A}\ldots A$ символ A повторялся m раз; таким образом, если мы перенумеруем все последовательные места этого ряда: $1, 2, 3, \ldots, n$, нужно предоставить для A какие-нибудь m номеров; это можно сделать столькими различными способами (согласно определению понятия сочетания), сколько существует различных сочетаний C_n^m из n элементов по m. Так как мы нашли, что вероятность каждой такой последовательности равна $p^m q^{n-m}$, то, применяя теорему сложения вероятностей, находим

$$P_{m,n}=C_n^m p^m q^{n-m}=\frac{n!}{m!\,(n-m)!}\,p^m q^{n-m}. \qquad (53)$$

Заметим, что, если мы по формуле бинома Ньютона составим разложение

$$(p+q)^n=p^n+np^{n-1}q+\frac{n(n-1)}{2}p^{n-2}q_m^2+$$

$$+\ldots+C_n^m p^m q^{n-m}+\ldots+q^n,$$

то член этого разложения, содержащий p^m, представляет как раз найденную нами вероятность

$P_{m,n}$, что при n опытах событие A, имеющее вероятность p, повторится ровно m раз. Отсюда видно между прочим, что

$$\sum_{m=0}^{m=n} P_{m,n} = (p+q)^n = 1.$$

Если, например, $p = q = \frac{1}{2}$, $n = 6$, то

$$P_{0,6} = P_{6,6} = \frac{1}{64};\ P_{1,6} = P_{5,6} = \frac{6}{64};\ P_{2,6} = P_{4,6} = \frac{15}{64};\ P_{3,3} = \frac{20}{64};$$

таким образом, наиболее вероятным определенным числом появления события A является 3; однако появление события какое-нибудь другое число раз все же оказывается более вероятным, так как сумма прочих вероятностей равна $\frac{44}{64}$.

Если $p = \frac{2}{3}$, $q = \frac{1}{3}$, $n = 5$, то $P_{0,5} = \frac{1}{243}$; $P_{1,5} = \frac{10}{243}$; $P_{2,5} = \frac{40}{243}$; $P_{3,5} = \frac{80}{243}$; $P_{4,5} = \frac{80}{243}$; $P_{5,5} = \frac{32}{243}$. В данном случае числа $m = 3$ и $m = 4$ имеют наибольшие равные между собою вероятности.

4. Пусть, вообще, p $(0 < p < 1)$ и n будут некоторые данные числа (n целое число); давая в формуле (53) m различные целые значения, можем всегда определить, для каких значений m вероятность $P_{m,n}$ будет наибольшею. С этой целью напишем формулу (53) в виде

$$P_{m,n} = \frac{n(n-1)\ldots(n-m+1)}{m!} p^m q^{n-m} \qquad (53\ \text{bis})$$

и, заменяя в ней m через $m+1$, получим

$$P_{m+1,n} = \frac{n(n-1)\ldots(n-m)}{(m+1)!} p^{m+1} q^{n-m-1},$$

откуда

$$\frac{P_{m+1,n}}{P_{m,n}} = \frac{n-m}{m+1} \cdot \frac{p}{q}. \qquad (54)$$

Ввиду того, что с увеличением m числитель дроби (54) уменьшается, а знаменатель увеличивается, значение этой дроби уменьшается с возрастанием m; таким образом

$$\frac{np}{q} = \frac{P_{1,n}}{P_{0,n}} > \frac{P_{2,n}}{P_{1,n}} > \frac{P_{3,n}}{P_{2,n}} > \ldots > \frac{P_{n,n}}{P_{n-1,n}} = \frac{p}{nq}. \qquad (55)$$

Если n достаточно велико, так что $np > q$ и $nq > p$, тогда крайний член слева в написанной системе неравенств больше 1, а крайний член справа — меньше 1; поэтому убывающие (с возрастанием m) отношения $\frac{P_{m+1,n}}{P_{m,n}}$ будут сначала больше 1, пока мы не дойдем до $m = \mu$, когда $\frac{P_{\mu+1,n}}{P_{\mu,n}} \leqslant 1$; после этого, для всех значений $m > \mu$, уже $\frac{P_{m+1,n}}{P_{m,n}} < 1$. Иначе говоря,

$$P_{m,n} < P_{m+1,n} \quad \text{пока } m < \mu$$
$$P_{m,n} \geqslant P_{m+1,n} \quad \text{при } m = \mu$$
$$P_{m,n} > P_{m+1,n} \quad \text{при } m > \mu.$$

Следовательно с увеличением m вероятность $P_{m,n}$ сначала возрастает до тех пор, пока m не достигнет значения μ; после этого вероятность $P_{m,n}$ сразу начинает убывать, если $P_{\mu+1,n} < P_{\mu,n}$, и тогда $P_{\mu,n}$ есть наибольшее из всех значений $P_{m,n}$, а μ является наиболее вероятным числом появлений нашего события при n опытах; если же $P_{\mu+1,n} = P_{\mu,n}$, тогда максимальное значение вероятности достигается как для $m = \mu$, так и $m = \mu + 1$: оба эти числа являются в этом случае наиболее вероятными. В последнем случае μ определяется из уравнения

$$\frac{P_{\mu+1,n}}{P_{\mu,n}} = \frac{n-\mu}{\mu+1} \cdot \frac{p}{q} = 1,$$

откуда $np - \mu p = \mu q + q$, и замечая, что $p + q = 1$,

$$\mu = np - q.$$

Этот случай представится, если $np - q$ будет целым числом, тогда наиболее вероятных чисел появления события будет два: $\mu = np - q$ и $\mu + 1 = np - q + 1 = (n+1)p$. (Это имело место в рассмотренном выше примере, где $p = \frac{2}{3}$, $n = 5$.) Если $np - q$ не целое число, тогда отношение $\frac{P_{m+1,n}}{P_{m,n}}$ не может быть равно 1, и переход этого отношения от значений бо́льших 1 к значениям меньшим 1 совершается сразу:

говоря, μ есть тогда наименьшее из целых чисел m для

$$\frac{P_{m+1,n}}{P_{m,n}}=\frac{n-m}{m+1}\cdot\frac{p}{q}<1,$$

т.-е.

$$np-mp<mq+q;$$

следовательно μ есть такое целое число, которое удовлетворяет неравенству

$$np-q<\mu,$$

между тем как $\mu-1$ удовлетворяет противоположному неравенству; поэтому

$$\mu-1<np-q<\mu,$$

и искомое наиболее вероятное целое число μ определяется неравенством

$$np-q<\mu<np-q+1=(n+1)p. \qquad (56)$$

Заметим, что во всяком случае наиболее вероятное число появлений события при n опытах (будет ли оно единственным или их окажется два) отличается от произведения np меньше, чем на 1; и, в частности, если np есть целое число, то $\mu=np$, так как число np удовлетворяет неравенству (56).

Остается еще рассмотреть случай, когда $np<q$ или $np=q$. В первом случае все отношения $\frac{P_{m+1,n}}{P_{m,n}}<1$; поэтому наиболее вероятным будет число $\mu=0$, т.-е. что событие ни разу не наступит; если же $np=q$, то $P_{1,n}=P_{0,n}$; поэтому наиболее вероятными будут числа 0 и 1. Аналогичный результат получим, если $nq\leqslant p$. Если, например, $p=\frac{1}{5}$ и $n=4$, то наиболее вероятными числами появлений события будут 0 и 1:

$$P_{0,4}=P_{1,4}=\frac{256}{625};\ P_{2,4}=\frac{96}{625};\ P_{3,4}=\frac{16}{625};\ P_{4,4}=\frac{1}{625};$$

если $p=\frac{4}{5}$ и $n=4$, то наиболее вероятными числами будут 3 и 4.

5. Формулу (53) можно получить другим способом, применяя алгебраический прием, который применим и к более сложным задачам.

Положим, что события $A_1, A_2, \ldots, A_n$ в ряде независимых опытов имеют соответственно вероятности $p_1, p_2, \ldots, p_n$. Вероятности $P_{m,n}$ того, что при указанных n опытах осуществится ровно m из рассматриваемых событий, равна коэффициенту при x^m многочлена

$$S(x) = (p_1x + q_1)(p_2x + q_2)\ldots(p_nx + q_n) = \\ = P_{n,n}x^n + P_{n-1,n}x^{n-1} + \ldots + P_{m,n}x^m + \ldots + P_{0,n} \quad (57)$$

(здесь, по обыкновению, положено $q_i = 1 - p_i$).

Наше утверждение очевидно для $P_{n,n}$ и $P_{0,n}$, так как вероятность наступления всех n событий равна произведению их вероятностей $p_1, p_2, \ldots, p_n$, и, с другой стороны, производя умножение двучленов по правилам алгебраического умножения, получаем коэффициент при x^n, перемножая коэффициенты при x во всех двучленах; точно так же свободный член получается от перемножения свободных членов $q_1, q_2, \ldots, q_n$, произведение которых как раз представляет вероятность совмещения $\bar{A}_1, \bar{A}_2, \ldots, \bar{A}_n$, т.-е. ненаступления ни одного из наших событий.

Что же касается члена, содержащего x^m в произведении $S(x)$, то, согласно правилам элементарной алгебры, для получения его нужно в m из данных двучленов взять первые члены, а в остальных $n - m$ взять вторые члены и перемножить; это даст произведение вида $p_1 q_2 \ldots p_n x^m$ (в котором буква p будет фигурировать m раз, а буква q входит $n - m$ раз); сделавши приведение всех подобных членов указанного вида, содержащих x в m-й степени, мы получаем соответствующий член произведения $S(x)$, так что коэффициент при x^m равен $\Sigma p_1 q_2 \ldots p_n$, т.-е. сумме всех произведений n множителей, среди которых буква p (с любыми значками) встречается m раз. Но произведение вида $p_1 q_2 \ldots p_n$ представляет собой вероятность совмещения событий $A_1, A_2, \ldots, A_n$, т.-е. совмещения наступления m событий A, значки которых совпадают с значками p, с ненаступлением остальных $n - m$ событий, которые соответствуют значкам буквы q в рассматриваемом произведении. Поэтому сумма всех произведений такого вида, где буква p будет входить со всевозможными m (из общего числа n) значками, которая, как мы только что заметили, есть коэффициент при x^m в $S(x)$, по теореме сложения вероятностей, представит вероятность того, что какие-нибудь m из рассматриваемых событий осуществятся при ненаступлении остальных. Ч. и т. д.

Например, если $p_1=\frac{1}{4}$, $p_2=\frac{1}{2}$, $p_3=\frac{3}{4}$, то

$$S(x)=\left(\frac{1}{4}x+\frac{3}{4}\right)\left(\frac{1}{2}x+\frac{1}{2}\right)\left(\frac{3}{4}x+\frac{1}{4}\right)=$$

$$=\frac{1}{32}[3x^3+13x^2+13x+3],$$

так, что

$$P_{0,3}=P_{3,3}=\frac{3}{32},\quad P_{1,3}=P_{2,3}=\frac{13}{32}.$$

В частности, если все вероятности $p_i=p$, то формула (57) превращается в формулу бинома Ньютона

$$(px+q)^n=p^nx^n+\ldots+C_n^m p^m q^{n-m}x^m+\ldots+q^n,$$

и коэффициент при x^m дает нам как раз то же самое значение для вероятности $P_{m,n}$, которое выражено формулой (53).

Покажем, что, каковы бы ни были значения $p_1, p_2, \ldots, p_n$, — остаются в силе неравенства

$$\frac{P_{1,n}}{P_{0,n}}>\frac{P_{1,n}}{P_{1,n}}>\ldots>\frac{P_{n,n}}{P_{n-1,n}}, \tag{55bis}$$

из которых вытекает, что $P_{m,n}$ с возрастанием m сначала возрастает и, достигнув наибольшей величины для некоторого одного или двух смежных значений μ, затем убывает (разумеется, как и в случае $p_i=p$, наиболее вероятное значение μ может соответствовать и крайним значениям $m=0$, если $\frac{P_{1,n}}{P_{0,n}}<1$, или $m=n$, если $\frac{P_{n,n}}{P_{n-1,n}}>1$). Для доказательства неравенств (55 bis) применяем метод математической индукции, замечая, что для $n=2$

$$P_{0,2}=q_1q_2,\ P_{1,2}=p_1q_2+p_2q_1,\ P_{2,2}=p_1p_2,$$

а потому неравенство

$$\frac{P_{1,2}}{P_{0,2}}>\frac{P_{2,2}}{P_{1,2}},$$

равнозначное $p_1^2q_2^2+p_2^2q_1^2+p_1p_2q_1q_2>0$, очевидно, соблюдено при любых положительных числах p_1, p_2, q_1, q_2. Итак, допустивши, что неравенства (55 bis) соблюдены для некоторого числа n опытов, мы должны проверить, что они справедливы и для $n+1$.

Умножая равенство (57) на $px+q$, чтобы получить вероятности $P_{m,n+1}$, что при $n+1$ опыте (в каждом из которых вероятность появления

события A равна соответственно $p_1, p_2, \ldots, p_n$, p) событие A произойдет m раз, находим, применяя правила алгебраического умножения,

$$P_{m,n+1} = P_{m-1,n}\, p + P_{m,n}\, q.$$

Эту формулу можно получить также, заметив, что событие произойдет m раз при $n+1$ опытах в одном из двух случаев: либо при n опытах событие A произошло $m-1$ раз и затем наступило в $(n+1)$ опыте, либо оно произошло m раз уже при n первых опытах и в последнем опыте не наступило; вероятность первого совмещения есть $P_{m-1,n}\, p$, а вероятность второго случая равна $P_{m,n}\, q$, а потому, складывая эти две величины, находим по теореме сложения вероятность $P_{m,n+1}$. Следовательно надо проверить, что из неравенств (55 bis) вытекают неравенства

$$\frac{P_{0,n}\, p + P_n\, q}{P_{0,n}\, q} > \frac{P_{1,n}\, p + P_{2,n}\, q}{P_{0,n}\, p + P_{1,n}\, q} > \ldots > \frac{P_{m,n}\, p + P_{m+1,n}\, q}{P_{m-1,n}\, p + P_{m,n}\, q} >$$

$$> \frac{P_{m+1,n}\, p + P_{m+2,n}\, q}{P_{m,n}\, p + P_{m+1}\, q} > \ldots > \frac{P_{n,n}\, p}{P_{n-1,n}\, p + P_{n,n}\, q};$$

но для этого достаточно заметить, что

$$(P_{m,n}\, p + P_{m+1,n}\, q)^2 - (P_{m+1,n}\, p + P_{m+2,n}\, q)\,(P_{m-1,n}\, p + P_{m,n}\, q) =$$

$$= (P^2_{m,n} - P_{m+1,n} P_{m-1,n})\, p^2 + (P_{m,n} P_{m+1,n} - P_{m+2,n} P_{m-1,n})\, pq +$$

$$+ (P^2_{m+1,n} - P_{m+2,n} P_{m,n})\, q^2 > 0,$$

так как коэффициенты при p^2, pq и q^2 положительны благодаря тому, что

$$\frac{P_{m,n}}{P_{m-1,n}} > \frac{P_{m+1,n}}{P_{m,n}} > \frac{P_{m+2,n}}{P_{m+1,n}}.$$

Пусть, например, $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{3}$, $p_3 = \frac{1}{4}$, $p_4 = \frac{1}{5}$, $p_5 = \frac{1}{6}$; тогда

$$P_{0,5} = \frac{120}{720},\ P_{1,5} = \frac{274}{720},\ P_{2,5} = \frac{225}{720},\ P_{3,5} = \frac{85}{720},\ P_{4,5} = \frac{15}{720},\ P_{5,5} = \frac{1}{720};$$

наиболее вероятно, что событие произойдет 1 раз.

Упражнения. 1) Показать, что если вероятность $p = \frac{1}{2}$ и число опытов $n = 2s$, то наиболее вероятное число появлений события есть s, и вероятность этого числа $P_{s,2s} = \frac{1}{2} \cdot \frac{3}{4} \cdots \frac{2s-1}{2s}$. Показать, что

$$\frac{1}{2\sqrt{s}} < P_{s,2s} < \frac{1}{\sqrt{2s+1}},$$

так что $P_{s,2s}$ стремится к нулю с возрастанием s.

Ответ. Для установления правого неравенства, замечаем, что

$$P_{s,2s} < \frac{2}{3} \cdot \frac{4}{5} \cdots \frac{2s}{2s+1} = \frac{2}{1} \cdot \frac{4}{3} \cdots \frac{2s}{2s-1} \cdot \frac{1}{2s+1} = \frac{1}{P_{s,2s}} \cdot \frac{1}{2s+1};$$

поэтому $P^2_{s,2s} < \frac{1}{2s+1}$.

Для установления левого неравенства, замечаем, что

$$2P_{s,2s} > \frac{2}{3} \cdot \frac{4}{5} \cdots \frac{2s-2}{2s-1} = \frac{1}{2s} \cdot \frac{1}{P_{s,2s}}, \text{ откуда } P_{s,2s} > \frac{1}{2\sqrt{s}}.$$

Например, если $n = 80$, то вероятность, что число m появлений события будет равно как раз $\frac{n}{2} = 40$, менее, чем $\frac{1}{\sqrt{81}} = \frac{1}{9}$.

2) Два игрока A и B повторяют некоторую партию, при которой A имеет вероятность p выиграть и q проиграть; выигрыш B соответствует проигрышу A, и наоборот. Игра продолжается до тех пор, пока A выиграет m партий или проиграет n партий. Какова вероятность P для игрока A оказаться победителем?

Ответ. $P = p^m \left[1 + C^1_m q + C^2_{m+1} q^2 + \ldots + C^{n-1}_{m+n-2} q^{n-1} \right]$.

Это значение для P получаем, замечая, что игра может закончаться выигрышем A либо после m партий, либо после $(m+1)$ партий и т. д., либо, наконец, после $(m+n-1)$ партий. Вероятность первого случая равна p^m, так как это произойдет лишь при удачном исходе всех первых m партий; для того, чтобы игра закончилась выигрышем A после $(m+1)$ партии, нужно, чтобы этот игрок выиграл $(m-1)$ из первых m партий и затем выиграл также $(m+1)$ партию: вероятность такого совмещения равна $C^1_m \; p^{m-1} q \cdot p = C^1_m \; p^m q$. Точно также победа достанется A после $(m+2)$ партии, если он выиграет $(m-1)$ партий из первых $(m+1)$ партий и затем выиграет $(m+2)$-ю партию: совмещение этих двух обстоятельств имеет вероятность $C^2_{m+1} \, p^m \, q^2$ и т. д.

Аналогичным образом получаем вероятность

$$Q = q^n \left[1 + C^1_n \, p + C^2_{n+1} \, p^2 + \ldots + C^{m-1}_{n+m-2} p^{m-1} \right]$$

выигрыша второго игрока (т. е. проигрыша игрока A).

Пусть, например, $p = \frac{1}{4}$, $q = \frac{3}{4}$, $m = 2$, $n = 6$; тогда $P = \frac{4447}{8192}$, $Q = \frac{3645}{8192}$. Каковы бы ни были данные числа $p > 0$ и m, вероятность Q стремится к 0, если n безгранично возрастает.

3) Проверить тождество

$$1 + \frac{A-a}{A-1} + \frac{(A-a)(A-a-1)}{(A-1)(A-2)} + \ldots + \frac{(A-a)\ldots 2.1}{(A-1)\ldots(a+1)a} = \frac{A}{a},$$

где $A \geqslant a > 0$ целые числа, применяя теоремы сложения и умножения вероятностей.

Ответ. Левая часть равенства, помноженная на $\frac{a}{A}$, представляет вероятность, что при вынимании последовательно шаров из урны, содержащей a белых шаров при общем числе A шаров, будет наконец вынут белый шар.

4) Какова вероятность при бросании n игральных костей, что сумма очков на верхних гранях этих костей будет равна данному числу l? Какова вероятность, что сумма будет не больше l?

Ответ. Для решения задачи строим многочлен

$$S(x) = \left[\frac{x + x^2 + x^3 + x^4 + x^5 + x^6}{6}\right]^n = \sum_{l=n}^{l=6n} P_l x^l;$$

коэффициент P_l при x^l представит первую из искомых вероятностей. Многочлен $S(x)$ может быть преобразован следующим образом:

$$S(x) = \frac{x^n}{6^n}\left[\frac{1 - x^6}{1 - x}\right]^n =$$

$$= \frac{x^n}{6^n}\left[1 - nx^6 + \frac{n(n-1)}{2}x^{12} - \ldots\right] \cdot \left[1 + nx + \frac{n(n+1)}{2}x^2 + \ldots\right];$$

поэтому

$$P_l = \frac{1}{6^n}\left[C_{l-1}^{n-1} - n\,C_{l-7}^{n-1} + \frac{n(n-1)}{2}C_{l-13}^{n-1} - \ldots\right]$$

(сумма в скобках естественно обрывается, как только нижний значок у C становится меньше верхнего). Например, при $n = 10$ вероятность, что $l = 20$, равна

$$P_{20} = \frac{1}{6^{10}}\left[C_{19}^{9} - 10\,C_{13}^{9}\right] = \frac{92\,378 - 7\,150}{6^{10}} = \frac{21\,307}{15\,116\,544} \neq 0{,}001\,41.$$

Вероятность R_l, что сумма очков будет не больше l, равна

$$R_l = P_n + P_{n+1} + \ldots + P_l = \frac{1}{6^n}\left[C_l^n - n\,C_{l-6}^n + \frac{n(n-1)}{2}C_{l-12}^n - \ldots\right],$$

так как вообще $C_{n-1}^{n-1} + C_n^{n-1} + \ldots + C_{N-1}^{n-1} = C_N^n$. Например, при $n = 10$, $l = 20$, $R_{20} = \frac{1}{6^{10}}\left[C_{20}^{10} - 10\,C_{14}^{10}\right] = \frac{184\,756 - 10\,010}{6^{10}} = \frac{87\,373}{30\,233\,088} \neq 0{,}002\,89.$

ГЛАВА ТРЕТЬЯ

ВЫЧИСЛЕНИЕ ВЕРОЯТНОСТЕЙ A POSTERIORI

1. Как было указано вначале, мы всегда должны иметь в виду, говоря о вероятности какого-нибудь факта, что эта вероятность существенным образом зависит от того, к какому классу фактов мы его относим, или, иначе говоря, какие обстоятельства опыта, сопровождающегося возможным появлением рассматриваемого факта, считаются определенными, т.-е. известными. Если мы рассматриваем, например, некоторый опыт, протекающий во времени, результатом которого может быть появление известного факта A, то, по мере выясняющегося течения опыта, вероятность события A обычно меняется, получая часто самые разнообразные значения в зависимости от тех или иных перипетий опыта, при чем в каждый данный момент вероятность A имеет вполне определенный объективный смысл, поскольку факт A всегда рассматривается нами как произвольный представитель известного класса фактов, могущих наступить или не наступить при наличности данных точно указанных обстоятельств.

В начале опыта, в момент t_1 факт A имеет некоторую вероятность (A), соответствующую первоначально известным условиям опыта, которую называют априорной, или первоначальною, вероятностью факта A; если в какой-нибудь последующий момент t_2 мы изменяем наше суждение о вероятности A, то это происходит потому, что мы видим, что, кроме данных условий, осуществились еще некоторые обстоятельства B; таким образом, согласно принятым раньше обозначениям, в момент t_2 вероятность факта A становится равной $(A)_B$, при чем никакой общеобязательной математической зависимости между этими двумя величинами не существует, если не считать того, что $(A)_B = (A)$ при $(A) = 1$ и при $(A) = 0$ (т.-е., что факт, бывший достоверным до наступления события B, остается достоверным и после наступления B, а факт невозможный не может ни при каких обстоятельствах сделаться возможным).

Вероятность $(A)_B$ (соответствующую какому-нибудь последующему моменту) называют апостериорной вероятностью. Ясно, что первоначальная вероятность (A) носит лишь относительный характер, поскольку можно вообразить, что в какой-нибудь предшествующий t_1 момент t_0 комплекс известных нам обстоятельств был еще меньше, чем в момент t_1. Поэтому логически более удовлетворительным является (не вводя элемента времени) определить первоначальную, априорную вероятность факта A как вероятность (A), соответствующую минимуму (наименьшему числу) обстоятельств, которые в данной задаче считаются известными. Если же вероятность факта A рассматривается при предположении дополнительного осуществления того или иного обстоятельства B, то каждому B соответствует апостериорная или условная вероятность $(A)_B$ факта A. Не делая никакого принципиального различия между последними двумя терминами, первым пользуются по преимуществу тогда, когда хотят отметить эмпирический характер вероятности $(A)_B$, соответствующей фактическому наступлению события B, между тем как термин условной вероятности обычно применяется, когда осуществление обстоятельства B носит условный предположительный характер и в ходе рассматриваемых опытов непосредственно не наблюдается.

Предположим, что результатом некоторого конкретного опыта является наступление или ненаступление события C. Пусть, например, перед нами находится ряд урн с шарами, и событие C заключается в появлении белого шара при извлечении одного шара, вынимаемого из выбранной наудачу урны. Допустим, что по составу находящихся в них шаров, наши урны делятся на 3 группы: — A_1, A_2, A_3, при чем, благодаря известной нам пропорции белых шаров в урне каждой из указанных групп, мы знаем, что если извлечение происходит из урны A_1 то вероятность появления белого шара (условная вероятность события C при осуществлении обстоятельства A_1) равна $(C)_{A_1}=p_1$, точно так же, если извлечение происходит из урны A_2, то $(C)_{A_2}=p_2$, и при извлечении шара из урны A_3 вероятность появления белого

$(C)_{A_3}=p_3$. Кроме того даны первоначальные вероятности $\alpha_1=(A_1)$, $\alpha_2=(A_2)$, $\alpha_3=(A_3)$ того, что извлечение будет, соответственно, происходить из урны первой, второй или третьей категории. Например, если бы все урны были одинаковы с виду, вероятности α_1, α_2, α_3 были бы соответственно пропорциональны числам урн каждой категории; во всяком случае $\alpha_1+\alpha_2+\alpha_3=1$, так как, по условию, извлечение из какой-нибудь урны обязательно должно произойти. Допустим, что, извлекая при данных условиях шар, мы фактически вынули белый шар C; требуется определить, каковы, после выяснения этого нового обстоятельства, вероятности (a posteriori)

$$x_1=(A_1)_C,\quad x_2=(A_2)_C,\quad x_3=(A_3)_C$$

того, что появившийся белый шар вынут, соответственно, из урны 1-й, 2-й или 3-й группы.

Для определения x_1 достаточно заметить, что по формуле (13)

$$(C)(A_1)_C=(A_1)(C)_{A_1},\ \text{т.-е.}\ (A_1)_C=\frac{(A_1)(C)_{A_1}}{(C)};$$

следовательно, так как на основании (48) первоначальная вероятность факта C появления белого шара равна

$$(C)=\alpha_1p_1+\alpha_2p_2+\alpha_3p_3,$$

то

$$(A_1)_C=x_1=\frac{\alpha_1p_1}{\alpha_1p_1+\alpha_2p_2+\alpha_3p_3},\tag{58}$$

и точно так же

$$x_2=\frac{\alpha_2p_2}{\alpha_1p_1+\alpha_2p_2+\alpha_3p_3},\quad x_3=\frac{\alpha_3p_3}{\alpha_1p_1+\alpha_2p_2+\alpha_3p_3}.$$

Предположим, например, что у нас было только 3 одинаковых урны, при чем в 1-й урне 1 белый шар и 3 черных, во 2-й — 2 белых и 2 черных, в 3-й — 3 белых шара и 1 черный, так что

$$\alpha_1=\alpha_2=\alpha_3=\frac{1}{3},\ \text{и}\ p_1=\frac{1}{4},\quad p_2=\frac{1}{2},\quad p_3=\frac{3}{4};$$

тогда

$$x_1=\frac{\frac{1}{3}\cdot\frac{1}{4}}{\frac{1}{3}\cdot\frac{1}{4}+\frac{1}{3}\cdot\frac{1}{2}+\frac{1}{3}\cdot\frac{3}{4}}=\frac{1}{6},\quad x_2=\frac{1}{3},\quad x_3=\frac{1}{2}.$$

Вообще, если $a_1 = a_2 = a_3 = a$, формулы (58) после сокращения на a принимают более простой вид:

$$x_1 = \frac{p_1}{p_1 + p_2 + p_3}, \quad x_2 = \frac{p_2}{p_1 + p_2 + p_3}, \quad x_3 = \frac{p_3}{p_1 + p_2 + p_3}. \qquad (58\ \text{bis})$$

Очевидно, что ничего в нашем рассуждении не изменилось бы, если бы вместо трех категорий урн у нас их было n, при чем априорные вероятности извлечения из урн каждой категории были бы соответственно равны $a_1, a_2, \ldots, a_n$, а $p_1, p_2, \ldots, p_n$ были бы соответственно условные вероятности появления белого шара, когда шар вынимается из урны 1-й, 2-й, ..., n-й группы.

Поэтому, замечая, что в рассматриваемом более общем случае

$$(C) = a_1 p_1 + a_2 p_2 + \ldots + a_n p_n = \sum_{i=1}^{i=n} a_i p_i,$$

мы получим для вероятности x_1 a posteriori (после того как в действительности факт C произошел), что вынутый белый шар был извлечен из урны 1-й категории

$$x_1 = \frac{a_1 p_1}{a_1 p_1 + \ldots + a_n p_n}$$

и вообще, при всяком k,

$$x_k = \frac{a_k p_k}{a_1 p_1 + \ldots + a_n p_n}. \qquad (58)$$

Формулы (58) носят название формул Байесса. Очевидно, что эти формулы применимы не только к специальному опыту с урнами, но соответствуют всякому опыту, обладающему аналогичной формальной структурой. Если рассматриваются n единственно возможных и несовместимых обстоятельств или гипотез $A_1, A_2, \ldots, A_n$, имеющих соответственно априорные вероятности $a_1, a_2, \ldots, a_n$, при чем в случае правильности гипотезы A_k вероятность некоторого факта C равна p_k $(k = 1, \ldots, n)$, то вероятность $x_k = (A_k)_C$ a posteriori каждой гипотезы A_k определяется формулой (58), если становится известно, что факт C действительно произошел. Наоборот, если бы в действительности факт C не произошел, то

апостериорная вероятность $y_k = (A_k)_{\overline{C}}$ гипотезы A_k была бы дана формулой

$$y_k = \frac{a_k q_k}{\sum_{i=1}^{i=n} a_i q_i}, \qquad (58\ \text{ter})$$

где $q_k = 1 - p_k$.

Из формулы (58) видно, что вероятность x_k a posteriori каждой из гипотез A_k пропорциональна ее первоначальной вероятности a_k и кроме того пропорциональна условной вероятности p_k, т.-е. вероятности наступления происшедшего в действительности факта C при условии, что справедлива именно гипотеза A_k. Формулы Байесса непосредственно вытекают из утверждения этой двойной пропорциональности, так как, написавши, что

$$x_1 = \lambda a_1 p_1, \quad x_2 = \lambda a_2 p_2, \ldots, \quad x_n = \lambda a_n p_n,$$

определяем коэффициент пропорциональности λ из равенства

$$\lambda (a_1 p_1 + a_2 p_2 + \ldots + a_n p_n) = 1,$$

выражающего, что несовместимые между собой гипотезы $A_1, \ldots, A_n$ единственно возможны.

Из формул (58), также, как из данного выше словесного выражения их содержания, нетрудно заключить, что если априорные вероятности a_k всех гипотез равны между собой, то наиболее вероятной после наступления факта C становится та гипотеза, при осуществлении которой факт C получил бы вероятность бо́льшую, чем при осуществлении одной из прочих гипотез, так как в данном случае

$$x_k = \frac{p_k}{\sum_{i=1}^{i=n} p_i}.$$

Если же, напротив, все p_i равны между собой, т.-е. все рассматриваемые нами гипотезы одинаково благоприятны (неблагоприятны) факту C, то наступление или ненаступление C не дает никакого основания для изменения нашего первоначального суждения о вероятностях каждой из гипотез: в этом случае $x_k = a_k$. (Вообще, если среди всех гипотез имеются две A_1 и A_2, при которых вероятности события C равны между собой, $p_1 = p_2$, то $\frac{x_1}{x_2} = \frac{a_1}{a_2}$.)

2. Формула (58) Байесса (или, как ее иногда называют, формула вероятностей гипотез) является, как мы видим, простым и вполне строго обоснованным логическим следствием из установленных вначале принципов теории вероятностей. Однако практическое применение ее нередко бывает затруднительно за недостатком достаточно обоснованных данных для того, чтобы придать те или иные определенные значения $\alpha_1, \alpha_2, \ldots, \alpha_n$ первоначальным вероятностям гипотез $A_1, A_2, \ldots, A_n$. Предположим, что перед нами имеется урна с 3 шарами; из нее вынимают 1 шар, который оказывается белым. Какова вероятность x, что все шары в этой урне были белые? Согласно формуле (58), мы можем только написать, что

$$x = \frac{\alpha_1}{\alpha_1 + \frac{2}{3}\alpha_2 + \frac{1}{3}\alpha_3},$$

так как кроме гипотезы, что все шары белые, возможно еще два предположения: в урне мог быть только 1 белый шар или же 2 белых шара. В зависимости от того, какие значения мы придаем первоначальным вероятностям $\alpha_1, \alpha_2, \alpha_3$, мы получим различные величины для x. Часто, за отсутствием определенных оснований для иного суждения, полагают $\alpha_1 = \alpha_2 = \alpha_3$; тогда

$$x = \frac{1}{1 + \frac{2}{3} + \frac{1}{3}} = \frac{1}{2}.$$

Но последнему результату нельзя придавать особого значения, и правильнее было бы сказать, что данные о постановке опыта недостаточны для того, чтобы искомой вероятности можно было приписать ту или иную определенную величину. Действительно, учитывая то обстоятельство, что шары могут быть самого разнородного цвета (и неодинаковые), следовало бы признать весьма маловероятным a priori, что в данную урну были положены как раз только шары одинакового цвета; поэтому не без основания можно было бы предположить, что α_1 гораздо меньше, чем α_2, и в особенности, чем α_3. Эта неопределенность значений $\alpha_1, \alpha_2, \alpha_3$ не играла бы столь существенной роли, если бы извлечение шара было произведено не один раз, а много раз при одинаковых условиях: если, вынувши шар, мы положим его обратно в урну и, пере-

мешавши шары, произведем второе извлечение, затем снова возвратим его в урну и повторим наше извлечение шара n раз. Если, поступая таким образом, мы все время будем вынимать белый шар, например $n = 10$ раз под-ряд, то без всякого вычисления мы склонны будем считать довольно вероятным, что все 3 шара в нашей урне белые. Вычисление же приведет нас к следующему результату: совершившийся факт C (10-кратное появление белого шара при 10 опытах) имел бы вероятность равную 1, если бы все шары в урне были белые; он имел бы вероятность $\left(\frac{2}{3}\right)^{10}$, при предположении, что в нашей урне 2 белых шара, и условная вероятность факта C равна $\left(\frac{1}{3}\right)^{10}$, если в урне только 1 белый шар; поэтому, применяя формулу Байесса для апостериорной вероятности x первой гипотезы (что все шары белые), получаем:

$$x = \frac{\alpha_1}{\alpha_1 + \left(\frac{2}{3}\right)^{10}\alpha_2 + \left(\frac{1}{3}\right)^{10}\alpha_3} = \frac{\alpha_1}{\alpha_1 + \frac{1\,024}{59\,049}\alpha_2 + \frac{1}{59\,049}\alpha_3} =$$

$$= \frac{1}{1 + \frac{1\,024}{59\,049}\frac{\alpha_2}{\alpha_1} + \frac{1}{59\,049}\frac{\alpha_3}{\alpha_1}}. \qquad (59\text{ bis})$$

В частности, если $\alpha_1 = \alpha_2 = \alpha_3$, то $x = \frac{59\,049}{60\,074} \doteq 0.983$; но, как было замечено выше, последнее допущение необоснованно; принимая во внимание высказанные при этом соображения, мы не можем указать определенных значений для отношений $\frac{\alpha_2}{\alpha_1}$ и $\frac{\alpha_3}{\alpha_1}$, и если первоначальное сомнение наше в возможности того, чтобы все шары в урне были белые, очень велико, то рассматриваемые отношения $\frac{\alpha_2}{\alpha_1}$ и $\frac{\alpha_3}{\alpha_1}$ могут выражаться настолько большими числами, что, несмотря на малость множителей, сопровождающих их в формуле (59 bis), величина апостериорной вероятности x могла бы оказаться все-таки незначительной, но во всяком случае гораздо большею, чем априорная вероятность α_1. Например, если $\frac{\alpha_2}{\alpha_1} = 2000$

и $\frac{\alpha_3}{\alpha_1} = 1\,000\,000$, то $x = \frac{59\,049}{3\,107\,049} \neq 0{,}019$, между тем как $\alpha_1 = \frac{1}{1\,002\,001} \neq 0{,}000\,001$. Если бы наш скептицизм был не так велик и мы с самого начала допускали бы, например, что $\frac{\alpha_2}{\alpha_1} \leqslant 10$, $\frac{\alpha_3}{\alpha_1} \leqslant 100 \left(\alpha_1 \geqslant \frac{1}{111}\right)$, то мы получили бы, что вероятность x близка к единице:

$$x \geqslant \frac{1}{1 + \frac{10\,240}{59\,049} + \frac{100}{59\,049}} = \frac{59\,049}{69\,389} \neq 0{,}85.$$

Ясно, что чем n будет больше, тем меньшую роль будет играть наша (субъективная) оценка априорных вероятностей α, и опыт, постоянно приводящий к неизменному появлению белого шара, убедит в конце концов самого крайнего скептика в том, что иных шаров, кроме белых, в нашей урне нет: вероятность x этой гипотезы, при неограниченном возрастании n стремится к достоверности.

Предположим, вообще, что при данных неизменно повторяющихся условиях отдельного опыта некоторое событие C наступало в каждом из n рассматриваемых опытов. Сохраняя прежние обозначения, полагаем, что с появлением события C совместимы три гипотезы A_1, A_2, A_3, имеющие вероятности a priori равные $\alpha_1, \alpha_2, \alpha_3$, при которых вероятности появления C в каждом из рассматриваемых опытов равны соответственно определенным числам p_1, p_2, p_3; следовательно при каждой гипотезе вероятности факта C_n, заключающегося в появлении C во всех n опытах, соответственно равны p_1^n, p_2^n, p_3^n, а потому, на основании формул Байесса, находим для апостериорных вероятностей наших трех гипотез, после наступления факта C_n, соответственно значения

$$x_1^{(n)} = \frac{\alpha_1 p_1^n}{\alpha_1 p_1^n + \alpha_2 p_2^n + \alpha_3 p_3^n}, \quad x_2^{(n)} = \frac{\alpha_2 p_2^n}{\alpha_1 p_1^n + \alpha_2 p_2^n + \alpha_3 p_3^n},$$

и

$$x_3^{(n)} = \frac{\alpha_3 p_3^n}{\alpha_1 p_1^n + \alpha_2 p_2^n + \alpha_3 p_3^n}.$$

Из этих формул заключаем, что если, при гипотезе A факт C достоверен ($p_1 = 1$), а при прочих гипотезах

факт C недостоверен, а только более или менее вероятен ($p_2 < 1$, $p_3 < 1$), то вероятность $x_1^{(n)}$ гипотезы A_1 стремится к достоверности, когда число n опытов, при которых факт C неизменно повторялся, неограниченно возрастает.

Действительно,

$$x_1^{(n)} = \frac{a_1}{a_1 + a_2 p_2^n + a_3 p_3^n} = \frac{1}{1 + \frac{a_2}{a_1} p_2^n + \frac{a_3}{a_1} p_3^n}; \tag{59}$$

поэтому, замечая, что с возрастанием n, p_2^n и p_3^n стремятся к 0, видим, что, каковы бы ни были постоянные $\frac{a_2}{a_1}$, $\frac{a_3}{a_1}$,

$$\underset{n=\infty}{\text{пред.}}\, x_1^{(n)} = 1.$$

Полученный вывод совершенно не зависит от первоначальных вероятностей a_1, a_2, a_3 наших гипотез; так, в рассмотренном выше примере $\left(\frac{a_2}{a_1} = 2\,000, \frac{a_3}{a_1} = 1\,000\,000, p_2 = \frac{2}{3}, p_3 = \frac{1}{3}\right)$, если $n = 100$, то

$$x_1^{(100)} = \frac{1}{1 + 2\,000\left(\frac{2}{3}\right)^{100} + 1\,000\,000\left(\frac{1}{3}\right)^{100}} > 1 - \frac{1}{10^{14}}.$$

Только тогда из постоянной повторяемости события C мы не могли бы заключить об истинности гипотезы A_1, если бы мы a priori отрицали ее возможность ($a_1 = 0$); тогда, конечно, никакой опыт не мог бы нас убедить в правильности этой гипотезы (при всяком n, мы получали бы $x_1^{(n)} = 0$).

Когда из формул (59) определены вероятности всех трех гипотез A_1, A_2, A_3, нетрудно вычислить, какою вероятностью обладает факт C в рассматриваемой серии однородных опытов, после того как обнаруживается, что имело место C_n (т.-е. n-кратное появление события C в n опытах); искомая вероятность равна, согласно (48),

$$(C)_{C_n} = x_1^{(n)} p_1 + x_2^{(n)} p_2 + x_3^{(n)} p_3 =$$

$$= \frac{a_1 p_1^{n+1} + a_2 p_2^{n+1} + a_3 p_3^{n+1}}{a_1 p_1^n + a_2 p_2^n + a_3 p_3^n} \tag{60}$$

Очевидно, что при $p_1 = 1$ ($a_1 > 0$), $(C)_{C_n}$ также стремится к 1, когда n неограниченно возрастает, так как

$(C)_{C_n} \geqslant x_1^{(n)}$. Последний вывод является математическим выражением основного принципа индуктивной науки, сводящегося в конечном счете к утверждению, что явление, которое при наличности определенной совокупности условий всегда осуществлялось, необходимо должно осуществляться при указанных условиях [1]). Таким образом следует признать, что в сущности нет принципиальной разницы между предсказанием будущего появления некоторого события, опирающимся на установленные эмпирические законы, и утверждением, основанным на правильном применении принципов теории вероятностей, что рассматриваемое событие имеет вероятность, неизмеримо мало отличающуюся от 1.

Упражнения. 1) Даны 2 ящика с шарами одинаковых по виду $\left(\alpha_1 = \alpha_2 = \frac{1}{2}\right)$; в первом 3 белых шара и 1 черный шар, во втором — 2 белых и 2 черных. Из одного ящика извлечен шар, который оказался белым; спрашивается, какова вероятность, что при вторичном вынимании из того же ящика появится белый в том случае, если шар возвращен в свой ящик, а также в том случае, когда шар не возвращен.

Ответ. В первом случае $x = \dfrac{\frac{1}{2}\left(\frac{3}{4}\right)^2 + \frac{1}{2}\left(\frac{1}{2}\right)^2}{\frac{1}{2}\cdot\frac{3}{4} + \frac{1}{2}\cdot\frac{1}{2}} = \frac{13}{20}$;

во втором случае $x = \dfrac{\frac{1}{2}\cdot\frac{3}{4}\cdot\frac{2}{3} + \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{3}}{\frac{1}{2}\cdot\frac{3}{4} + \frac{1}{2}\cdot\frac{1}{2}} = \frac{8}{15}$.

2) Доказать, что если опыт повторяется при одинаковых условиях, то вероятность появления события C в $(n+1)$-м опыте, после того как известно, что оно произошло в n предшествующих опытах, вообще, больше

---

[1]) Допущение $\alpha_1 = 0$, которое в рассмотренном выше примере с 3 шарами лишено всякого разумного основания, вообще, не так абсурдно, если значение p_2 очень близко к 1. Современная физика рассматривает теперь многие явления C, которые всегда наблюдаются при данных условиях (например, одинаковые показания температуры двух одинаковых термометров, вставленных в изолированный баллон, наполненный газом) лишь как обладающие вероятностью p_2, неизмеримо мало отличающейся от 1. Если в формуле (60) $\alpha_1 = 0$ и $p_2 > p_3$, то, с возрастанием n, $(C)_{C_n}$ стремится к p_2; но значения p_2 в вышеупомянутых физических опытах обычно так мало отличаются от 1, что потребовались бы многие миллионы лет непрерывного наблюдения, чтобы непоявление C хоть один раз имело некоторые реальные шансы.

вероятности наступления того же события C в n-м опыте, когда известно, что оно произошло в предшествующих $(n-1)$ опытах; исследовать, в каких случаях обе указанные вероятности могут быть равны.

Ответ. Нужно доказать (60), что

$$\frac{\alpha_1 p_1^n + \alpha_2 p_2^n + \ldots + \alpha_k p_k^n}{\alpha_1 p_1^{n-1} + \ldots + \alpha_k p_k^{n-1}} < \frac{\alpha_1 p_1^{n+1} + \alpha_2 p_2^{n+1} + \ldots + \alpha_k p_k^{n+1}}{\alpha_1 p_1^n + \ldots + \alpha_k p_k^n},$$

т.-е.

$$\left[\alpha_1 p_1^{n+1} + \alpha_2 p_2^{n+1} + \ldots + \alpha_k p_k^{n+1}\right]\left[\alpha_1 p_1^{n-1} + \ldots + \alpha_k p_k^{n-1}\right] >$$
$$> \left[\alpha_1 p_1^n + \ldots + \alpha_k p_k^n\right]^2.$$

Члены, содержащие $\alpha_1^2, \alpha_2^2, \ldots, \alpha_k^2$, равны в обоих частях неравенства и взаимно сокращаются, члены же, содержащие произведения $\alpha_1\alpha_2$, в первой части сводятся к $\alpha_1\alpha_2\left[p_1^{n+1}p_2^{n-1} + p_1^{n-1}p_2^{n+1}\right]$, а во второй — равны $2\alpha_1\alpha_2 p_1^n p_2^n$; поэтому, переносия все в первую часть неравенства, находим неравенство

$$\alpha_1 \alpha_2 p_1^{n-1} p_2^{n-1} (p_1 - p_2)^2 + \ldots > 0,$$

которое очевидно, так как все слагаемые в первой его части положительны или равны 0. Доказанное неравенство заменится равенством, только если все члены вида $\alpha_1 \alpha_2 p_1^{n-1} p_2^{n-1} (p_1 - p_2)^2 = 0$, или, полагая вообще $\alpha_i > 0$, необходимо, чтобы все p_i (отличные от 0) были равны между собой, т.-е. чтобы независимо от правильности той или иной гипотезы, совместимой с появлением C, вероятность факта C была равна вполне определенному числу p.

3) Повторяя опыт n раз при одинаковых условиях, найдено, что событие C произошло m раз. Требуется определить вероятность a posteriori каждой гипотезы.

Ответ. Применяя формулу Байесса и замечая, что условная вероятность m-кратного появления C в n опытах равна $C_n^m p_1^m q_1^{n-m}$ при первой гипотезе, $C_n^m p_2^m q_2^{n-m}$ при второй гипотезе и т. д., получаем для искомых вероятностей

$$x_1 = \frac{C_n^m p_1^m q_1^{n-m}}{C_n^m p_1^m q_1^{n-m} + C_n^m p_2^m q_2^{n-m} + \ldots} =$$

$$= \frac{p_1^m q_1^{n-m}}{p_1^m q_1^{n-m} + p_2^m q_2^{n-m} + \ldots}$$

и т. п.

ГЛАВА ЧЕТВЕРТАЯ

СПОСОБ МАТЕМАТИЧЕСКИХ ОЖИДАНИЙ

1. Во многих случаях вычисление вероятностей облегчается рассмотрением некоторых связанных с вероятностями вспомогательных величин, так называемых математических ожиданий, которые играют существенную роль в теории вероятностей и ее приложениях. С основными свойствами этого нового понятия нам необходимо теперь ознакомиться.

Положим, что тот или иной результат рассматриваемого при данных условиях опыта связан для данного лица с тем или иным выигрышем, т.-е. с получением некоторой суммы денег. В таком случае мы говорим, что в данных условиях математическое ожидание выигрыша для рассматриваемого лица равно сумме произведений каждого возможного выигрыша на вероятность этого выигрыша. Если, например, при бросании игральной кости игрок выигрывает столько рублей, сколько окажется очков на верхней грани кости, то математическое ожидание его выигрыша равно

$$1\cdot\frac{1}{6}+2\cdot\frac{1}{6}+3\cdot\frac{1}{6}+4\cdot\frac{1}{6}+5\cdot\frac{1}{6}+6\cdot\frac{1}{6}=\frac{21}{6}=3^1/_2 \text{ руб.}$$

(Математическое ожидание выигрыша является всегда именованным числом, выраженным в тех же денежных единицах, в каких выражен каждый возможный выигрыш.)

В предыдущем примере все исходы игры были более или менее благоприятны для игрока, но большей частью некоторые результаты игры могут быть связаны с проигрышем; в таком случае проигрыш рассматривается как отрицательный выигрыш, и математическим ожиданием выигрыша, вообще, называется (алгебраическая) сумма произведений каждого выигрыша на его вероятность.

Если, например, при появлении 6 очков на данной кости игрок выплачивает 5 рублей, а при всяком другом числе очков выигры-

вает 1 рубль, то математическое ожидание выигрыша равно $1 \cdot \frac{5}{6} - 5 \cdot \frac{1}{6} = 0$.

Когда математическое ожидание выигрыша для данного игрока равно нулю, то игра называется безобидной для него; когда математическое ожидание выигрыша положительно, игра называется выгодной; и наконец, если математическое ожидание выигрыша отрицательно, то игра называется невыгодной. Игра, выгодная для данного игрока, — невыгодна для его противника, и наоборот, если выигрыш одного соответствует равному проигрышу другого. Вопрос о том, являются ли математически выгодные (согласно определению) игры таковыми в практическом смысле слова, будет нами исследован в дальнейшем на основании закона больших чисел.

Если выигрыш игрока равен A, а возможный проигрыш есть B, при чем вероятности выигрыша и проигрыша равны соответственно p и q, то для безобидности игры нужно, чтобы $Ap - Bq = 0$, т.-е. $\frac{A}{q} = \frac{B}{p}$. В таком случае игра безобидна и для противника нашего игрока.

Совершенно так же, как мы определили математическое ожидание выигрыша, можно определить математическое ожидание всякой величины x, если известны все значения $x_1, x_2, \ldots, x_n$, которые x может получить, и известны соответственные вероятности: p_1 того что $x = x_1$, p_2 того, что $x = x_2$ и т. д. Тогда мы говорим, что *математическое ожидание* $x = p_1 x_1 + p_2 x_2 + \ldots + p_n x_n$ (61), где $x_1, x_2, \ldots, x_n$ могут быть и отрицательными числами (даже комплексными), т.-е. математическим ожиданием величины x называется алгебрическая сумма произведений каждого возможного значения x на его вероятность. Разумеется, числа p отрицательными не бывают, и кроме того $p_1 + p_2 + \ldots + p_n = 1$, так как, по условию, никакое из возможных для x значений не должно быть пропущено. (Фактически, если одно из возможных для x значений есть $x_n = 0$, то соответствующее слагаемое $p_n x_n = 0$ в формуле (61) можно не писать, имея всегда в виду, что если в членах, фигурирующих в формуле (61), сумма вероятностей $P < 1$, то следует подразумевать, что одно из возможных значений для x есть нуль, и вероятность этого значения равна $1 - P$.) Например, если вероят-

ность события A в данном ряде n опытов равна p, то мы можем говорить о математическом ожидании числа m появлений этого события в рассматриваемых n опытах. Число m может получать значения $0, 1, 2, \ldots, n$, и, согласно формуле (53), вероятности этих значений равны соответственно: q^n, npq^{n-1}, $\frac{n(n-1)}{2!}p^2q^{n-2}, \ldots, p^n$.

Поэтому

$$\text{мат. ожид. } m = 0 \cdot q^n + 1 \cdot npq^{n-1} + 2 \cdot \frac{n(n-1)}{2!}p^2q^{n-2} + \ldots +$$

$$+ m \cdot \frac{n(n-1)\ldots(n-m+1)}{m!}p^m q^{n-m} + \ldots =$$

$$= np\left[q^{n-1} + (n-1)pq^{n-2} + \ldots +\right.$$

$$\left. + \frac{(n-1)\ldots(n-m+1)}{(m-1)!}p^{m-1}q^{n-m} + \ldots + p^{n-1}\right] =$$

$$= np[q+p]^{n-1} = np. \qquad (64 \text{ bis})$$

Следовательно математическое ожидание числа появлений события при n опытах равно np, где p есть вероятность этого события в каждом опыте.

Заметим, что в частности, если $n = 1$, то математическое ожидание m равно самой вероятности p, так как в данном случае математическое ожидание $m = p \cdot 1 + q \cdot 0 = p$. Таким образом вероятность является весьма частным случаем математического ожидания числа x, когда это последнее получает лишь два значения 1 или 0 в зависимости от того, произошло ли рассматриваемое событие или нет.

2. Вряд ли нуждаются в пояснениях следующие два положения: математическое ожидание достоверной величины a равно a; математическое ожидание ax, где a — данный численный множитель, равно произведению a *математ. ожид.* x.

Не столь очевидна, и во всяком случае нуждается в точном доказательстве, следующая чрезвычайно общая теорема сложения математических ожиданий.

Теорема. — Если при данных обстоятельствах *математическое ожидание* $x = A$, *математическое ожидание* $y = B$, то при тех же обстоятельствах *математическое ожидание* $(x + y) = A + B$; другими словами, математическое

ожидание суммы двух величин равно сумме математических ожиданий этих величин.

Действительно, по условию

$$A = p_1 x_1 + p_2 x_2 + \ldots + p_n x_n,$$
$$B = P_1 y_1 + P_2 y_2 + \ldots + P_m y_m,$$

где p_i есть вероятность равенства $x = x_i$, а P_k есть вероятность равенства $y = y_k$.

Предположим сначала, для большей наглядности, что $m = n = 2$, т.-е. что x может получить только два значения x_1 и x_2, а y также может получить лишь значения y_1 и y_2. В таком случае сумма $(x + y)$ имеет только четыре возможных значения:

$$x_1 + y_1, \quad x_1 + y_2, \quad x_2 + y_1, \quad x_2 + y_2,$$

вероятности которых p_{11}, p_{12}, p_{21}, p_{22} являются соответственно вероятностями совмещений равенств $x = x_1$ с равенством $y = y_1$, равенства $x = x_1$ с равенством $y = y_2$, равенства $x = x_2$ с равенством $y = y_1$, равенства $x = x_2$ с равенством $y = y_2$. Согласно определению математического ожидания всякой величины, пишем, таким образом,

$$\begin{aligned} \textit{мат. ож.}\ (x + y) &= p_{11}(x_1 + y_1) + p_{12}(x_1 + y_2) + \\ &+ p_{21}(x_2 + y_1) + p_{22}(x_2 + y_1) = \\ &= (p_{11} + p_{12}) x_1 + (p_{21} + p_{22}) x_2 + (p_{11} + p_{21}) y_1 + \\ &+ (p_{12} + p_{22}) y_2. \end{aligned}$$

Принимая наконец во внимание, что $p_{11} + p_{12} = p_1$ и $p_{21} + p_{22} = p_2$ представляют соответственно вероятности равенств $x = x_1$ и $x = x_2$ (так как каждое из этих равенств имеет место либо совместно с $y = y_1$ либо совместно с $y = y_2$) и что $p_{11} + p_{21} = P_1$ и $p_{12} + p_{22} = P_2$ являются соответственно вероятностями равенств $y = y_1$ и $y = y_2$, убеждаемся, что

$$\textit{мат. ож.}\ (x + y) = p_1 x_1 + p_2 x_2 + P_1 y_1 + P_2 y_2 = A + B.$$

То же рассуждение применим теперь к общему случаю, когда числа n и m возможных значений для x и y какие угодно.

Если обозначим через p_{ik} вероятность совместного существования равенств

$$x = x_i \text{ и } y = y_k,$$

то

$$p_i = \sum_{k=1}^{k=m} p_{ik} = p_{i1} + p_{i2} + p_{im}, \quad P_k = \sum_{i=1}^{i=n} p_{ik}. \tag{62}$$

Но, с другой стороны, чтобы получить математическое ожидание $(x+y)$, надо взять всевозможные значения суммы $(x+y)$, умноженные на соответствующие им вероятности, и сложить. Поэтому

$$\textit{мат. ож.}\ (x+y) = \sum_{i=1}^{i=n} \sum_{k=1}^{k=m} p_{ik}(x_i + y_k) =$$

$$= \sum_{i=1}^{i=n} \sum_{k=1}^{k=m} p_{ik} x_i + \sum_{i=1}^{i=n} \sum_{k=1}^{k=m} p_{ik} y_k.$$

Благодаря (62)

$$\sum_{i=1}^{i=n} \sum_{k=1}^{k=m} p_{ik} x_i = \sum_{i=1}^{i=n} x_i \sum_{k=1}^{k=m} p_{ik} = \sum_{i=1}^{i=n} x_i p_i \text{ и } \sum_{i=1}^{i=n} \sum_{k=1}^{k=m} p_{ik} y_k =$$

$$= \sum_{k=1}^{k=m} y_k \sum_{i=1}^{i=n} p_{ik} = \sum_{k=1}^{k=m} P_k y_k.$$

Следовательно

$$\text{М. О. } (x+y) = \sum_{i=1}^{i=n} p_i x_i + \sum_{k=1}^{k=m} P_k y_k = A + B.$$

Ч. и т. д.

Следствие. — Математическое ожидание суммы любого числа величин $x+y+z+\ldots$ равно сумме математических ожиданий этих величин:

$$\text{М. О. } (x+y+z+\ldots) = \text{М. О. } x + \text{М. О. } y + \text{М. О. } z + \ldots \tag{63}$$

Действительно, М. О. $(x+y+z)$ = М. О. $(x+y)$ + М. О. z = = М. О. x + М. О. y + М. О. z.

Аналогичным образом переходом от n к $(n+1)$ формула (63) распространяется на любое число слагаемых.

Предположим, в частности, что производится n опытов, при которых вероятности появления события A соответственно равны $p_1, p_2, \ldots, p_n$. В таком случае математическое ожидание числа m появлений события в рассматриваемых n опытах равно

$$\text{М. О. } m = p_1 + p_2 + \ldots + p_n. \tag{64}$$

(При этом опыты могут и не быть независимыми.)

Действительно, если каждое появление события в соответствующем i-ом опыте отмечается числом 1, а непоявление — числом 0, то математическое ожидание числа x_i, которым будет отмечен результат i-го опыта, М. О. $x_i = p_i \cdot 1 + q_i \cdot 0 = p_i$. Поэтому, согласно формуле (63),

$$\text{М. О.}(x_1 + x_2 + \ldots + x_n) = \text{М. О.}\, x_1 + \text{М. О.}\, x_2 + \ldots + \text{М. О.}\, x_n =$$
$$= p_1 + p_2 + \ldots + p_n;$$

но отмеченная нами сумма $x_1 + x_2 + \ldots + x_n$ будет как раз равна числу m появлений события A, так как каждое слагаемое этой суммы равно единице, если событие произошло, и равно 0, если A не произошло, следовательно,

$$\text{М. О. } m = p_1 + p_2 + \ldots + p_n. \tag{64}$$

Заметим, что формула (64) содержит как весьма частный случай формулу (64 bis), в которую она превращается, когда все $p_i = p$ и опыты независимы.

Рассмотрим теперь некоторые простейшие примеры применения математического ожидания к вычислению вероятностей.

3. *Первый пример.* Два игрока Иван и Петр повторяют некоторую безобидную партию (например, один из них бросает игральную кость, и первый получает от второго 5 рублей в случае появления 6 очков и выплачивает второму 1 рубль в случае какого-нибудь другого числа очков). Игра продолжается до тех пор, пока один из игроков не проиграет всех своих наличных денег, которых имеется у Ивана a рублей, а у Петра b рублей. Требуется определить вероятность P разорения Ивана.

Ввиду того, что каждая партия безобидна, то в силу теоремы сложения математических ожиданий, т.-е. формулы (63), безобидной будет и вся игра, слагающаяся из совокупности этих партий. Для Ивана эта совокупная безобидная игра закончится либо выигрышем всего имущества b Петра, либо проигрышем всего своего капитала a. Обозначая вероятность первого случая через Q,

а вероятность последнего случая через P, получаем, как условие безобидности [1]),

$$bQ - aP = 0, \text{ откуда } \frac{P}{b} = \frac{Q}{a} = \frac{1}{a+b}.$$

Например, если у Ивана $a = 10$ рублям, а у Петра $b = 40$ рублям, то $40Q - 10P = 0$, т.-е. $\frac{P}{4} = Q = \frac{1}{5}$: в этом примере вероятность P разорения Ивана равна $\frac{4}{5}$ и таким образом в четыре раза больше вероятности Q разорения Петра. Вообще, для игроков, повторяющих безобидную игру до полного разорения одного из них, вероятность разорения каждого обратно пропорциональна его капиталу. В частности, если Иван вступает в безобидную игру со всяким партнером, то вероятность его разорения чрезвычайно близка к единице, т.-е. к достоверности, так как совокупный капитал всех его противников, вместе взятых, практически неограниченно велик.

[1]) Сделанный нами вывод опирается на допущение, что игра непременно должна когда-нибудь окончиться. Однако это не очевидно, и, наоборот, было бы не вполне правильно, если бы повторяемые партии не были тождественны; если бы игроки условились, например, уменьшать пропорционально ставки каждой партии так, чтобы максимальный возможный проигрыш игрока в отдельной партии никогда не превышал половины имеющегося у него в наличности капитала, то игра, понятно, не закончилась бы никогда, так как для ее окончания требуется полное разорение одного из игроков. Но наш вывод нетрудно сделать вполне строгим, если все партии тождественны, принимая во внимание (как мы увидим дальше), что тогда, как бы мало ни было ε, возможно указать достаточно большое n, чтобы $1 - (P_n + Q_n) < \varepsilon$, где P_n и Q_n — вероятности разорения каждого из игроков, если число партий не превышает n. Поэтому, замечая, что игра стала бы выгодной для Ивана, если бы, в случае неопределенного результата n партий, он получал весь капитал b своего противника, и, наоборот, стала бы невыгодной, если бы он при указанном неопределенном результате должен был отдать Петру все свое имущество, заключаем, что

$$-\varepsilon b < bQ_n - aP_n < \varepsilon a,$$

откуда $-\varepsilon b < b(1 - P_n) - aP_n$ и, следовательно, $P_n < \frac{b(1+\varepsilon)}{a+b}$; с другой стороны, $b(1 - P_n - \varepsilon) - aP_n < \varepsilon a$, следовательно, $P_n > \frac{b}{a+b} - \varepsilon$. Итак, при n достаточно большом,

$$\left| P_n - \frac{b}{a+b} \right| < \varepsilon, \quad \left| Q_n - \frac{a}{a+b} \right| < \varepsilon.$$

Таким образом безобидность игры отнюдь не является указанием того, что эту игру безопасно повторять. Напротив, предприятия, деятельность которых, вроде страховых обществ, независимо от их социальной пользы или вреда, представляет математически полную аналогию с азартной игрой, не могли бы длительно существовать, в силу вышесказанного, если бы „игра" для них была безобидна. Такого рода предприятия должны себе обеспечить финансовые условия математически „выгодной игры".-Более полно этот вопрос будет освещен в дальнейшем.

4. *Второй пример.* Даны 2 ящика с шарами: в первом находится $3M$ белых шаров и M черных шаров; во втором имеется M белых и $3M$ черных шаров. Из первого ящика вынимают наудачу $2M$ шаров и перекладывают во второй ящик; размешавши после этого все шары во втором ящике, из него вынимают шар. Спрашивается, какова вероятность, что этот шар будет белый?

Так как вероятность для каждого шара, вынутого из 1-го ящика, быть белым равна $\frac{3}{4}$, то математическое ожидание числа вынутых белых шаров равно $\frac{3}{4}\cdot 2M=\frac{3M}{2}$, математическое ожидание числа белых шаров, оказавшихся во втором ящике, равно $M+\frac{3M}{2}=$ $=\frac{5M}{2}$; общее же число шаров во втором ящике становится равным $6M$. Поэтому математическое ожидание вероятности P появления белого шара из второго ящика равно $\frac{5M}{12M}=\frac{5}{12}$. Но вероятность P некоторого события и математическое ожидание этой вероятности [1]) представляют собой одно и то же, так как, по определению математического ожидания,

$$\text{М. О. } P=a_1P_1+a_2P_2+\ldots+a_nP_n,$$

где $P_1, P_2, \ldots, P_n$ означают все возможные значения, получаемые вероятностью P данного события, если наступает один из несовме-

[1]) Можно заметить также, что вообще, поскольку вероятность при условиях есть определенное число, ее математическое ожидание с этим определенным значением.

стимых случаев, имеющих соответственные вероятности $\alpha_1, \alpha_2, \ldots, \alpha_n$; таким образом, написанное выражение представляет собой не что иное, как вероятность P нашего события, если бы ее вычислять согласно формуле (48), имея в виду все возможные комбинации перекладывания шаров; но такой прием вычисления был бы гораздо сложнее. Итак, искомая вероятность $P = \frac{5}{12}$.

Аналогичным образом получим, что если переложить обратно $2M$ шаров из второго ящика в первый, то после этого второго перекладывания вероятность появления белого шара из первого ящика равна $\frac{7}{12}$.

5. Способ математических ожиданий в применении к более сложным задачам приводит к системам линейных уравнений со многими неизвестными и даже с бесконечным множеством неизвестных. Однако очень часто, не решая полностью всех этих уравнений, требующих довольно тонких математических исследований, оказывается возможным указать простые неравенства, которым удовлетворяют искомые вероятности, и таким образом получить чрезвычайно ценные для теории и практики приближенные значения последних.

Пример такого рода весьма важного неравенства содержится в следующем простом предложении, которое лежит в основе знаменитой статьи П. Л. Чебышева „О средних величинах".

Лемма. — Если *мат. ож.* $x = A$, причем x может получать различные положительные значения (или нуль), то вероятность Q, что x получит значение большее, чем At^2, меньше, чем $\frac{1}{t^2}$, каково бы ни было число t.

Действительно, по определению:

$$A = \text{М. О. } x = p_1 x_1 + p_2 x_2 + \ldots + p_n x_n. \tag{61}$$

Если из всех возможных значений x только $x_1, x_2, \ldots, x_h$ превышают At^2, то сумма $p_1 x_1 + \ldots + p_h x_h > (p_1 + p_2 + \ldots + p_h) At^2$, а потому (так как в (61) нет отрицательных членов) и подавно $A > (p_1 + p_2 + \ldots + p_h) At^2$; а следовательно,

$$Q = p_1 + p_2 + \ldots + p_h < \frac{1}{t^2},$$

где Q есть вероятность, что x получит какое-нибудь значение, превышающее At^2.

Следствие. Вероятность P (при тех же данных), что $x \leqslant At^2$, больше, чем $1 - \frac{1}{t^2}$.

В самом деле, $P + Q = 1$; поэтому, если, в силу только что доказанного, слагаемое $Q < \frac{1}{t^2}$, то другое слагаемое P должно быть больше, чем $1 - \frac{1}{t^2}$.

Предшествующую лемму вместе с выведенным из нее следствием мы будем называть леммой Чебышева.

Из леммы Чебышева мы выведем в дальнейшем самые общие формы закона больших чисел, а пока применим ее к выводу теоремы Якова Бернулли, являющейся простейшим частным случаем закона больших чисел.

6. *Теорема Якова Бернулли.* Пусть при каждом из n независимых опытов вероятность наступления события A равна p; тогда вероятность того, что число m появлений события A удовлетворит неравенству

$$\left| \frac{m}{n} - p \right| < \varepsilon, \tag{69}$$

где ε — данное произвольно малое число, становится сколь угодно близкой к единице (достоверности), если число n опытов достаточно велико.

Для доказательства нужно прежде всего вычислить

$$H = \text{М. О. } (m - np)^2,$$

т.-е. математическое ожидание квадрата отклонения числа m появлений события A от числа np, которое, согласно (64 bis), есть математическое ожидание m. Полагая $m = x_1 + x_2 + \ldots + x_n$, где x_1 получает значения 1 или 0, смотря по тому, происходит ли событие A в первом опыте или нет, и, вообще, $x_i = 1$ или $x_i = 0$ в зависимости от того, наступает ли событие A в i-ом опыте ($i = 1, 2, \ldots, n$) или нет, — можем написать:

$$H = \text{М. О. } (x_1 + x_2 + \ldots + x_n - np)^2$$

и, применяя теорему сложения математических ожиданий,

$$H = \text{М. О. } x_1^2 + \text{М. О. } x_2^2 + \ldots + \text{М. О. } x_n^2 + 2 \text{ М. О. } x_1 x_2 + \\ + 2 \text{ М. О. } x_1 x_3 + \ldots + 2 \text{ М. О. } x_{n-1} x_n - 2np \text{ М. О. } (x_1 + x_2 + \\ + \ldots + x_n) + n^2 p^2. \qquad (65)$$

Но математическое ожидание $x_1^2 = p \cdot 1 + q \cdot 0 = p$; следовательно, вообще, М. О. $x_i^2 = p$. С другой стороны, М. О. $x_1 x_2 = p^2$, так как произведение $x_1 x_2$ может быть равно либо 1, либо 0, при чем $x_1 x_2 = 1$ только в том случае, когда $x_1 = 1$ и $x_2 = 1$, т.-е. когда событие A осуществляется и в первом и во втором опыте, так что вероятность равенства $x_1 x_2 = 1$ равна p^2. По той же причине, вообще, М. О. $x_i x_k = p^2$. Следовательно, замечая, что членов вида 2 М. О. $x_i x_k$ в формуле (65) имеется $\frac{n(n-1)}{2}$, получаем:

$$H = np + n(n-1)p^2 - 2np \text{ М. О. } (x_1 + x_2 + \ldots + x_n) + n^2 p^2 = \\ = np + n(n-1)p^2 - 2n^2p^2 + n^2p^2 = np - np^2 = npq,$$

так как М. О. $(x_1 + x_2 + \ldots + x_n) = np$, согласно формуле (64 bis).

Итак, мы установили, что

$$H = \text{М. О. } (m - np)^2 = npq. \qquad (66)$$

Следовательно, применяя к неотрицательному числу $(m - np)^2$ лемму Чебышева, мы заключаем, что вероятность Q неравенства

$$(m - np)^2 > t^2 npq$$

меньше, чем $\frac{1}{t^2}$, а вероятность P противоположного неравенства

$$(m - np)^2 \leqslant t^2 npq \qquad (67)$$

больше, чем $1 - \frac{1}{t^2}$. Но неравенство (67) равнозначно неравенству

$$\left(\frac{m}{n} - p\right)^2 \leqslant \frac{t^2 pq}{n}, \qquad (68)$$

которое получается от деления обеих частей (67) на n^2; извлекая корень квадратный из обеих частей (68), получаем новое равнозначное неравенство

$$\left|\frac{m}{n} - p\right| \leqslant t\sqrt{\frac{pq}{n}}, \qquad (69 \text{ bis})$$

вероятность которого, следовательно, также равна $P > 1 - \frac{1}{t^2}$. Полагая $t\sqrt{\frac{pq}{n}} = \varepsilon$, откуда $\frac{1}{t^2} = \frac{pq}{\varepsilon^2 n}$, можем наконец придать последнему утверждению такую форму: вероятность P неравенства

$$\left|\frac{m}{n} - p\right| \leqslant \varepsilon \tag{69}$$

больше, чем $1 - \frac{pq}{\varepsilon^2 n}$. Следовательно, если при данном ε (как бы мало оно ни было) число n опытов будет неограниченно возрастать, то вероятность $P > 1 - \frac{pq}{\varepsilon^2 n}$ сделается сколь угодно близкой к единице, благодаря тому, что $\frac{pq}{\varepsilon^2 n}$ будет стремиться к 0 с увеличением n, что и требовалось доказать.

Так, например, при бросании монеты n раз, находим, полагая вероятность появления „орла" $p = \frac{1}{2}$, что вероятность P неравенства

$$\left|\frac{m}{n} - \frac{1}{2}\right| \leqslant \frac{1}{20},$$

т.-е.

$$\frac{9}{20} \leqslant \frac{m}{n} \leqslant \frac{11}{20} \tag{69 ter}$$

больше, чем $1 - \frac{100}{n}$; таким образом, если $n = 1000$, то $P > 0,9$, если $n = 10\,000$, то $P > 0,99$ и т. д.

В дальнейшем мы найдем более точные значения для вероятности P, из которых видно будет, что P на самом деле еще гораздо ближе к 1, так что для $n = 10\,000$ осуществление неравенства (69 ter) является практически достоверным $\left(P > 1 - \frac{1}{10^{20}}\right)$, и даже для $n = 1000$ вероятность $P > 0,998$.

Неравенство

$$P > 1 - \frac{pq}{\varepsilon^2 n} \tag{70}$$

дает некоторую нижнюю границу для вероятности P, когда даны число опытов n и величина ε максимального уклонения $\left|\frac{m}{n} - p\right|$. Если, напротив, мы хотим указать, какое число опытов

n окажется достаточно большим для того, чтобы вероятность P осуществления неравенства (69) была во всяком случае больше $1-\eta$, где η некоторое данное весьма малое число, то, как видно из (70), для этого достаточно, чтобы $1-\frac{pq}{\varepsilon^2 n}>1-\eta$, т.-е.

$$\eta>\frac{pq}{\varepsilon^2 n},$$

откуда

$$n>\frac{pq}{\varepsilon^2\eta}. \qquad (70\ \text{bis})$$

Если, например, при повторном бросании игральной кости, мы хотим утверждать с вероятностью большею, чем $0{,}99=1-0{,}01$, что число m появлений 6 очков удовлетворит неравенству

$$\left|\frac{m}{n}-\frac{1}{6}\right|\leqslant\frac{1}{10},\ \text{т.-е.}\ \frac{1}{15}\leqslant\frac{m}{n}\leqslant\frac{4}{15},$$

то достаточно для этого, чтобы

$$n>\frac{5}{36}\cdot 10\,000>1\,388.$$

Из неравенства (69 bis) получаем также, что если n и $\eta=\frac{1}{t^2}$ даны, то с вероятностью большею, чем $1-\eta$, следует ожидать осуществления неравенства

$$\left|\frac{m}{n}-p\right|\leqslant\sqrt{\frac{pq}{n\eta}}.$$

Таким образом при 50 000 бросаниях игральной кости мы можем гарантировать с вероятностью большею, чем 0,99, что

$$\left|\frac{m}{n}-\frac{1}{6}\right|\leqslant\sqrt{\frac{\frac{1}{6}\cdot\frac{5}{6}\cdot 100}{50\,000}}=\frac{1}{60},\ \text{т.-е.}\ \frac{9}{60}\leqslant\frac{m}{n}\leqslant\frac{11}{60}.$$

Принимая во внимание, что $n=50\,000$, мы видим, что хотя отношение $\frac{m}{n}$, которое мы должны ожидать с большею вероятностью мало отличается от $\frac{1}{6}$, тем не менее пределы, в которых

заключаться само число m появлений 6 очков, довольно широки, именно:

$$50\,000 \cdot \frac{9}{60} \leqslant m \leqslant \frac{11}{60} \cdot 50\,000,$$

т.-е.

$$7\,500 \leqslant m \leqslant 9\,167.$$

Правда, мы не имеем основания утверждать, что при более точной оценке вероятностей не окажется в действительности возможным значительно сузить пределы полученного неравенства. Этот вопрос будет в полной мере разъяснен в дальнейшем. Но пока мы должны лишь заметить, что из того обстоятельства, что отношение $\frac{m}{n}$ практически неограниченно приближается к вероятности p с увеличением числа опытов n, отнюдь не следует, что и само число m становится весьма близким к своему математическому ожиданию np, так как хотя разность $\frac{m}{n} - p = \delta$ очень мала, но, умножая ее на большое число n, мы получаем разность $m - np = n\delta$, которая может быть велика.

Для случая $p = \frac{1}{2}$ мы нашли (стр. 79), что вероятность $P_{s,2s}$ того, что при $n = 2s$ событие произойдет s раз, удовлетворяет неравенству $P_{s,2s} < \frac{1}{\sqrt{2s+1}}$; но так как s есть наиболее вероятное число появлений нашего события, то, тем более, вообще $P_{m,2s} < \frac{1}{\sqrt{2s+1}}$; поэтому вероятность, что m равно одному из $2k+1$ значений:

$$s-k,\ s-k+1, \ldots, s,\ s+1, \ldots, s+k,$$

т.-е. вероятность P, что

$$\left| m - \frac{n}{2} \right| \leqslant k,$$

менее, чем $\frac{2k+1}{\sqrt{n+1}}$; таким образом, как бы велико ни было данное число k, вероятность P стремится к нулю, когда n неограниченно возрастает: например, если $n = 1\,000\,000$, $k = 15$, то $P < 0{,}031$.

Упражнения. 1) Иван играет с Петром на следующих условиях: из полной колоды карт вынимается одна карта за другой до тех пор, пока не появится туз; Иван в начале партии выплачивает Петру 10 рублей с тем, что Петр возвратит ему столько рублей, сколько карт будет предшествовать тузу. Для кого из игроков игра выгодна?

Ответ. Игра выгодна для Петра и невыгодна для Ивана: математическое ожидание выигрыша для последнего равно $\frac{48}{5} - 10 = -0{,}4 = -40$ ко-

пеек. Действительно, если все 52 карты колоды будут вынуты одна за другой, то первому тузу будет предшествовать некоторое число k карт, между первым и вторым тузом будет l карт, m будет число карт между вторым и третьим тузом, n — между третьим и четвертым, и p карт последует за четвертым тузом. Очевидно, $k+l+m+n+p=48$; но, так как положение туза на любом месте одинаково вероятно (имеет вероятность равную $\frac{1}{13}$), то указанное расположение столь же вероятно, как и такое, при котором числа k, l, m, n, p расположатся в другом порядке; иными словами: М. О. $x_1 =$ $=$ М. О. $x_2 =$ М. О. $x_3 =$ М. О. $x_4 =$ М. О. x_5, где x_1 — число карт, предшествующих первому тузу, x_2, x_3, x_4 — числа карт, соответствующие промежуткам между двумя последовательными тузами, и x_5 — число карт, следующих за последним тузом; следовательно, М. О. $x_1 = \frac{48}{5}$.

Вообще, если $A+1$ есть число карт и a число тузов, то М. О. $x_1 =$ $= \frac{A-a+1}{a+1}$; вычисляя М. О. x_1 другим способом, находим

$$\text{М. О. } x_1 = \frac{(A-a+1)a}{(A+1)A} + 2\frac{(A-a+1)(A-a)a}{(A+1)A(A-1)} + \ldots,$$

откуда между прочим получаем тождество

$$1 + 2\frac{A-a}{A-1} + 3\frac{(A-a)(A-a-1)}{(A-1)(A-2)} + \ldots = \frac{A(A+1)}{a(a+1)}$$

для любых целых чисел $A \geqslant a > 0$.

2) Иван играет с Петром на прежних условиях, только вынутая карта каждый раз после появления кладется обратно. Для кого игра выгодна?

Ответ. Игра выгодна для Ивана. Математическое ожидание его выигрыша равно $12-10=2$ руб. Действительно, если $p=\frac{1}{13}$ есть вероятность появления туза, а $q=\frac{12}{13}$ — вероятность появления другой карты, то вероятность, что тузу предшествует h карт, равна $q^h p$; поэтому число x предшествующих тузу карт имеет математическое ожидание, равное

$$\text{М. О. } x = qp + 2q^2p + \ldots + hq^hp + \ldots = \frac{qp}{(1-q)^2} = \frac{q}{p} = 12.$$

3) Определить, каковы соответственные вероятности p_1, p_2, p_3, p_4, p_5 числу x получить значения $-2, -1, 0, 1, 2$, если дано, что иных значений x не получает и что М. О. $x=0$, М. О. $x^2=2$, М. О. $x^3=0$, М. О. $x^4=6$

Ответ. $p_1=p_5=\frac{1}{6}$, $p_2=p_4=\frac{1}{3}$, $p_3=0$.

4) Сколько раз следует повторить опыт, при котором вероятность появления события $p=\frac{1}{4}$, для того, чтобы с вероятностью не меньшею, чем 0,999, можно было ожидать, что $\frac{39}{160} \leqslant \frac{m}{n} \leqslant \frac{41}{160}$?

Ответ. Достаточно, вследствие (70 bis), чтобы $n > 4\,800\,000$.

ГЛАВА ПЯТАЯ

МЕТОД ПРЕДЕЛОВ И ГРАФИЧЕСКОЕ ИЗОБРАЖЕНИЕ ВЕРОЯТНОСТЕЙ

1. Во многих задачах нельзя указать никакого логического основания для того, чтобы ограничить тем или иным числом совокупность фактов, вероятности которых могут представить практический интерес, или совокупность значений, которые может принимать та или иная неизвестная величина. Даже и тогда, когда мы знаем, что число возможных исходов опыта конечно, но только чрезвычайно велико, для вычисления вероятностей отдельных фактов (с достаточной для практики точностью) часто оказывается проще считать число возможных исходов бесконечным и пользоваться, благодаря этому, методом пределов, т.-е. принципами дифференциального и интегрального исчислений.

Предположим, например, что циферблат хронометра имеет 100 равноотстоящих делений, и положим, что стрелка его механически отмечает момент наступления какого-нибудь явления A, независимого от хода хронометра, так что мы можем считать равновозможным, что отмеченный момент A окажется в любом из 100 указанных промежутков [1]), и вероятность этого для каждого промежутка равна $\frac{1}{100}$; в таком случае, согласно определению вероятности, если обозначим через s длину окружности циферблата, то вероятность, что A упадет, например, на дуге длины $\frac{15}{100}s$, между делением 10 и делением 25, равна $\frac{15}{100}$. Если бы каждый промежуток цифер-

---

[1]) Возможны, конечно, сомнительные случаи, когда A будет казаться данному наблюдателю точно совпавшим с точкой деления циферблата; относительно этих случаев следует принять некоторое единообразное соглашение: можно, например, условиться, что если A совпадает с делением a, то мы его относим к промежутку $(a, a+1)$.

блата хронометра был разделен еще на 10 равных частей, т.-е. если бы вместо 100 делений было 1000 равноотстоящих делений, и мы могли бы различать попадание точки A в эти более мелкие промежутки, то мы аналогичным образом получили бы для вероятности того, что точка A попадет, например, на дугу между делениями 102 и 256, длина которой $\frac{154}{1000}s$, значение $\frac{154}{1000}$, так как такому положению точки A благоприятствуют 154 случая из 1000 единственно и равно возможных несовместимых случаев. Продолжая тот же процесс деления и допуская, что, как бы малы ни были дальнейшие подразделения, наблюдатель имеет возможность решить, в какой промежуток попала точка A, мы приходим к заключению, что во всяком случае вероятность точке оказаться на дуге $\widehat{mn}$, концы которой совпадают с подразделениями циферблата, равна отношению числа промежутков, помещающихся на дуге $\widehat{mn}$, к общему числу промежутков, т.-е. отношению длины дуги $\widehat{mn}$ к длине всей окружности циферблата.

Окружность циферблата можно мысленно развернуть на прямолинейный отрезок длины s. В таком случае вероятность того, что точка A окажется на части этого отрезка, длина которой x, равняется $\frac{x}{s}$.

Вообще, когда отрезок LM, длина которого s, не очень велик, то часто условия постановки опыта таковы, что не может быть указано никаких объективных оснований для того, чтобы некоторая точка A, отмечаемая на отрезке, попала скорее в одну часть отрезка, чем в другую, если только размеры этих двух частей отрезка одинаковы; иными словами, во многих случаях следует принять, что вероятность P точке A отрезка LM оказаться в данном промежутке, длина которого x, есть функция $P(x)$, зависящая только от длины x, а не от положения промежутка на отрезке LM. В таком случае вероятность того, что точка A окажется в промежутке, длина которого $x+y$, есть $P(x+y)$, и, применяя теорему сложения вероятностей, мы находим, что

$$P(x+y)=P(x)+P(y), \tag{71}$$

откуда

$$P(x+y+z)=P(x)+P(y)+P(z)$$

и т. д.

Следовательно, в частности, каково бы ни было целое число n, имеем:

$$P(nx) = nP(x), \tag{72}$$

лишь бы $nx \leqslant s$ и, кроме того, $P(s) = 1$, так как, по условию, точка A несомненно находится на данном отрезке LM, длина которого s. Поэтому, полагая $nx = s$, получаем из (72), что $1 = nP\left(\frac{s}{n}\right)$, т.-е. $P\left(\frac{s}{n}\right) = \frac{1}{n}$ при всяком целом n.

Если теперь положим в равенстве (72) $x = \frac{s}{m}$, где целое число $m > n$, то получим:

$$P\left(\frac{n}{m}s\right) = nP\left(\frac{s}{m}\right) = \frac{n}{m}. \tag{73}$$

С другой стороны, из равенства (72) заключаем также, что если $y < \frac{s}{m}$, то $mP(y) = P(my) \leqslant P(s) = 1$, т.-е. $0 \leqslant P(y) \leqslant \frac{1}{m}$ (так как вероятность не может быть отрицательной).

Следовательно, если на отрезке LM дан определенный промежуток длины $h = \frac{n}{m}s + y$, где $0 \leqslant y \leqslant \frac{s}{m}$, то на основании равенств (71) и (73)

$$\frac{n}{m} \leqslant P(h) = P\left(\frac{n}{m}s\right) + P(y) \leqslant \frac{n+1}{m}; \tag{74}$$

но, неограниченно увеличивая m, имеем

$$\text{пред.}\ \frac{n}{m} = \text{пред.}\ \frac{n+1}{m} = \frac{h}{s}$$

и из (74) получаем, наконец, что

$$P(h) = \frac{h}{s}, \tag{75}$$

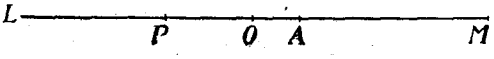
какова бы ни была длина $h \leqslant s$.

Таким образом, при сделанном выше предположении, вероятность P, что точка A упадет в определенный промежуток длины h на отрезке $LM = s$, равняется отношению $\frac{h}{s}$ длины промежутка h к длине всего отрезка s.

Формулу (75) называют законом равномерного распределения вероятностей. На основании теоремы Бернулли, доказанной в предыдущей главе, если число n точек A на отрезке LM, вероятности которых подчиняются указанному закону, весьма велико, то следует считать практически достоверным, что отношение числа m этих точек, находящихся в данном промежутке длины h отрезка LM, к общему их числу n очень мало отличается от отношения $\frac{h}{LM}$; иными словами, при неограниченном возрастании числа n, распределение точек A на отрезке LM стремится стать равномерным.

Мы не будем останавливаться здесь на рассмотрении некоторых парадоксов, к которым приводит закон равномерного распределения вероятностей; заметим лишь, что этот закон требует, чтобы мы приписали вероятность равную нулю, т.-е. признали невозможным идеальное совпадение точки A с определенной геометрической точкой; экспериментально такое совпадение действительно не может быть осуществлено, но если бы при постановке той или иной теоретической задачи указанное совпадение считалось возможным, то это означало бы, что точка A не подчиняется закону равномерного распределения вероятностей.

Пример. Дан отрезок LM и на нем произвольная точка A. Какова вероятность того, что возможно построить треугольник, имеющий сторонами LA, AM и $\frac{1}{2}LM = LO = OM$ (см. черт. 2)?



Черт. 2.

Для того, чтобы такой треугольник был возможен, необходимо и достаточно существование неравенств

$$AM + \frac{1}{2}LM > LA, \quad LA + \frac{1}{2}LM > AM,$$

выражающих, что сумма двух сторон треугольника больше третьей (неравенство $LA + AM > \frac{1}{2}LM$ всегда существует). Следовательно, полагая

$$LA = \frac{1}{2}LM + x \text{ и } AM = \frac{1}{2}LM - x,$$

неравенства наши можем записать в виде $LM - x > \frac{1}{2} LM + x$ и $LM + x > \frac{1}{2} LM - x$, т.-е. $\frac{1}{4} LM > x > -\frac{1}{4} LM$.

Иначе говоря, построение указанного треугольника будет возможно тогда и только тогда, когда точка A будет находиться вблизи середины O отрезка по ту или другую сторону от O на расстоянии, не превышающем $\frac{1}{4} LM$; следовательно, из всех положений точки A на отрезке LM благоприятными для существования треугольника будут лишь те, при которых она окажется внутри названного промежутка, длина которого равна $\frac{1}{2} LM$. Поэтому, считая все положения точки A на отрезке LM одинаково вероятными (т.-е. считая распределение вероятностей равномерным), получаем, что искомая вероятность $P = \frac{{}^1/_2 LM}{LM} = \frac{1}{2}$.

2. Вообще, равномерное распределение вероятностей на отрезке не является единственно допустимым предположением. Если, например, в середине отрезка помещается цель, в которую метит опытный стрелок, то для двух равных промежутков, из которых один ближе к середине, а другой дальше, более вероятно, что он попадет в первый промежуток, нежели во второй.

Самый общий закон распределения вероятностей на данной прямой, которую можем принять за ось абсцисс, совместимый с основным принципом теории вероятностей (глава первая, стр. 18), определяется произвольно заданной неубывающей (монотонной) функцией $F(x)$, где $F(x_0)$ представляет вероятность того, что абсцисса[1]) x точки A удовлетворяет неравенству

$$x < x_0,$$

при чем $F(-\infty) = 0$ и $F(+\infty) = 1$.

Если точка A может находиться только на данном определенном

[1]) Согласно общему приему графического изображения величин, абсцисса x может соответствовать любой величине (времени, температуре, весу, числу людей и т. п.).

отрезке LM, концы которого L и M имеют соответственно абсциссы a и b $(a < b)$, то

$$\left.\begin{aligned} F(x) &= 0, \quad \text{когда } x \leqslant a, \\ F(x) &= 1, \quad \text{когда } x > b, \end{aligned}\right\} \tag{76}$$

так как в этом случае величина x не может быть менее a и должна быть менее или равна b. В частности, при равномерном распределении вероятностей на отрезке LM, длина которого $s = b - a$, функция $F(x)$ (кроме условий (76)) имеет вид

$$F(x) = \frac{x - a}{b - a},$$

когда $a \leqslant x \leqslant b$; эта формула выражает, как и формула (75), что вероятность точке A оказаться в определенной части $LP = h = x - a$ отрезка LM равна отношению

$$\frac{LP}{LM} = \frac{h}{s} = \frac{x - a}{b - a}.$$

Если дана функция $F(x)$, которая называется **интегральной функцией распределения вероятностей**, то вероятность того, что точка A с абсциссой x находится в промежутке, концы которого имеют абсциссами α и β $(\alpha < \beta)$, или, точнее, вероятность P_α^β того, что $\alpha \leqslant x < \beta$, равна

$$P_\alpha^\beta = F(\beta) - F(\alpha). \tag{77}$$

Действительно, по определению, $F(\beta)$ есть вероятность неравенства $x < \beta$; но это неравенство имеет место при одном из двух несовместимых обстоятельств: либо $x < \alpha$, либо $x \geqslant \alpha$ (при $x < \beta$); вероятность первого случая есть $F(\alpha)$, а вероятность второго случая равна искомой вероятности P_α^β. Следовательно, по теореме сложения,

$$F(\beta) = F(\alpha) + P_\alpha^\beta,$$

откуда $P_\alpha^\beta = F(\beta) - F(\alpha)$.

В случае равномерного распределения вероятностей на отрезке LM

$$P_\alpha^\beta = \frac{\beta - a}{b - a} - \frac{\alpha - a}{b - a} = \frac{\beta - \alpha}{b - a} \tag{78}$$

в согласии с формулой (75).

Из равенства (77) видно, что если интегральная функция распределения $F(x)$ непрерывна, то P_α^β стремится к 0, когда β приближается к α, а потому вероятность точного равенства $x=\alpha$ равна нулю; иными словами, допущение, что функция $F(x)$ непрерывна, означает, что равенство $x=\alpha$ не может быть экспериментально установлено. Следовательно, наоборот, если равенство $x=\alpha$ может быть математически точно осуществлено, то значение α является точкой разрыва непрерывности для функции $F(x)$.

3. В дальнейшем мы будем почти исключительно иметь дело с задачами, в которых функция $F(x)$ непрерывна.

Из случаев, когда функция $F(x)$ прерывна, т.-е. изменяется скачками, мы остановимся лишь на одном, который имеет большое практическое значение. Пусть, например,

$$F(x)=0 \quad \text{при } x<0,$$

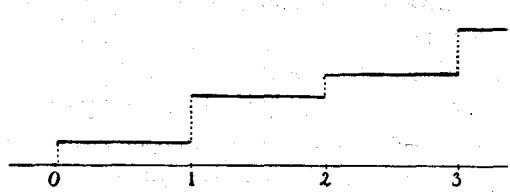
$$F(x)=\frac{1}{6} \quad \text{при } 0\leqslant x<1,$$

$$F(x)=\frac{1}{2} \quad \text{при } 1\leqslant x<2,$$

$$F(x)=\frac{2}{3} \quad \text{при } 2\leqslant x<3,$$

$$F(x)=1 \quad \text{при } 3\leqslant x.$$

Графически функция $y=F(x)$ представлена на фиг. 3 ступенчатой ломаной линией; при этом замечаем, что $P_\alpha^\beta=0$, если обе



Черт. 3.

абсциссы α и β соответствуют точкам нашей линии, находящимся на одной и той же горизонтальной ступени; напротив, если между α и β находится одна из точек разрыва (0, 1, 2, 3), то, как бы мал ни был промежуток $\overline{\alpha\beta}$, вероятность P_α^β точке A находиться в этом промежутке имеет вполне определенное значение, равное

скачку функции $F(x)$ в соответствующей точке разрыва; поэтому точка A может занять только 4 положения: 0, 1, 2, 3, и вероятность каждого из этих положений равна соответственному скачку функции $F(x)$ в этой точке, а именно: $\frac{1}{6}, \frac{1}{3}, \frac{1}{6}, \frac{1}{3}$.

Подобную же ступенчатую ломаную линию (у которой верхняя ступень находится на высоте, равной 1) мы получим каждый раз, как число x может приобретать лишь конечное число значений.

Если, например, x есть число появлений события A при n опытах, то функция $F(x)$ имеет $n+1$ точек разрыва непрерывности: $0, 1, \ldots, n$, соответствующих всем возможным значениям величины x, при чем скачок вверх в точке разрыва должен быть равен вероятности (53) равенства $x=m$, т.-е. $C_n^m p^m q^{n-m}$, где p есть вероятность события A в каждом опыте. Аналогичными свойствами прерывности обладают также функции распределения $F(x)$, соответствующие, например, вероятностям дереву данного рода в данных климатических условиях принести то или иное число x плодов, или животному данной породы родить в течение года число x детенышей и т. п.

Если, как в примере с числом появлений события при повторении опыта, число n точек разрыва непрерывности становится велико, то, принимая во внимание, что высота нашей ломаной линии не превышает 1, заключаем, что с увеличением n скачки становятся вообще весьма малыми. Так как, с другой стороны, при высоте линии, равной единице, естественно, для наглядности чертежа, не растягивать его слишком в ширину, а выбрать масштабы для измерения абсцисс и ординат так, чтобы ширина и высота чертежа были более или менее одинаковы, то с увеличением n не только высота, но и ширина каждой ступени соответствующей ломаной линии будет становиться очень малой, и таким образом, практически, наша ступенчатая линия будет сливаться с некоторой непрерывной, плавно поднимающейся кривой. Благодаря этому изучение наиболее интересного случая, когда число точек разрыва непрерывности прерывной функции $F(x)$ велико, сводится обыкновенно к исследованию предельного случая, когда функция $F(x)$ непрерывна. В дальнейшем мы неоднократно встретимся с подобными примерами.

4. Допустим теперь, что интегральная функция распределения вероятностей $F(x)$ непрерывна. Кроме того мы предположим, что внутри весьма малого промежутка распределение вероятностей очень мало отличается от равномерного. Иными словами, полагаем, что

$$P_{\alpha}^{\beta} = F(\beta) - F(\alpha) = (\beta - \alpha)[f(\alpha) + \varepsilon], \tag{80}$$

где ε может стать произвольно малым, если $\beta - \alpha < \delta$ взять достаточно малым, т.-е. допускаем, что существует

$$\underset{\beta=\alpha}{\text{пред.}} \frac{F(\beta) - F(\alpha)}{\beta - \alpha} = F'(\alpha) = f(\alpha). \tag{79}$$

Функция $f(x)$, которая является производною функции $F(x)$ и, очевидно, не может быть отрицательна, называется плотностью распределения вероятностей, или дифференциальной функцией распределения вероятностей. Равномерное распределение вероятностей на данном отрезке характеризуется, как видно из (78), постоянством плотности распределения вероятностей на этом отрезке.

Следует заметить, что дифференциальная функция распределения вероятностей $f(x)$ (в противоположность интегральной функции $F(x)$, которая никогда не превышает единицы) может получать сколь угодно большие положительные значения и даже не исключена возможность того, что в некоторых точках функция $f(x)$ бесконечно возрастает. Если, например, на отрезке $\overline{01}$ $F(x) = \sqrt{x}$, то $f(x) = F'(x) = \frac{1}{2\sqrt{x}}$ становится бесконечной при $x = 0$.

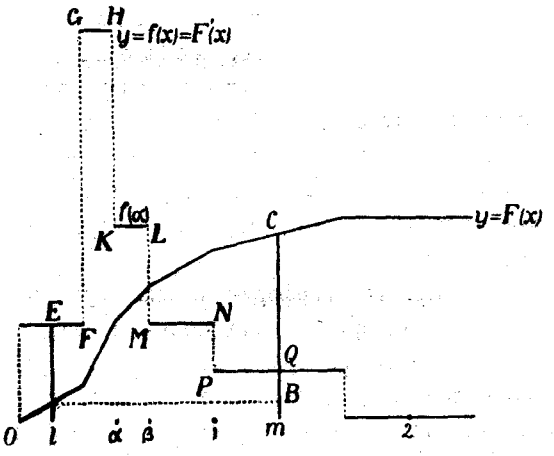
Поэтому графическое изображение плотности распределения вероятностей $f(x)$ приводит к кривым гораздо более разнообразных видов, нежели график функции $F(x)$; в частности, в только что указанном примере, где $f(x) = \frac{1}{2\sqrt{x}}$ бесконечно возрастает при $x = 0$, кривая $y = f(x)$ имеет прямую $x = 0$ асимптотою. Вследствие этого на практике обычно пользуются графиком функции $f(x)$, который гораздо рельефнее, чем график интегральной функции $F(x)$, характеризует распределение вероятностей величины x.

Обе линии $y=f(x)$ и $y=F(x)$ тесно связаны между собой, и по одной из них легко восстановить другую. Действительно, из того обстоятельства, что

$$F'(x)=f(x), \qquad (79\ \text{bis})$$

заключаем, что точки обеих кривых M_1 и M с одинаковыми абсциссами характеризуются тем, что ордината первой равна тангенсу угла φ, который образует касательная в точке M ко второй (интегральной) кривой с осью абсцисс.

Если, например, графиком интегральной функции распределения является ломаная непрерывная линия $y=F(x)$ (черт. 4), то плот-



Черт. 4.

ность $f(x)$ распределения вероятностей постоянна в каждом из промежутков между двумя ее соседними вершинами, изменяя свои значения скачками при переходе из одного промежутка в следующий.

В этом случае формула (80) приобретает вид

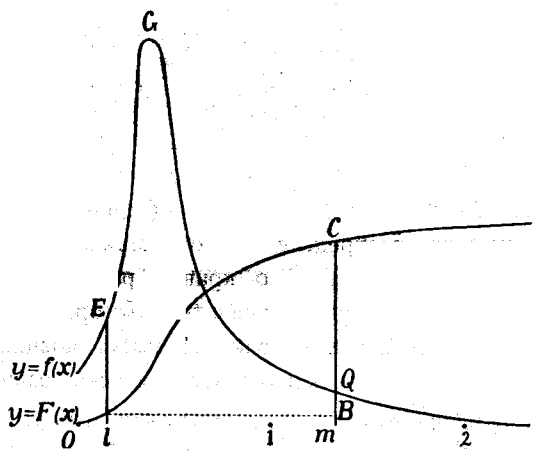
$$P_{\alpha}^{\beta}=(\beta-\alpha)\cdot f(\alpha)=F(\beta)-F(\alpha), \qquad (80\ \text{bis})$$

если только $f(x)$ сохраняет постоянное значение в интервале от α до β, т.-е. вероятность P_{α}^{β} точке A с абсциссой x оказаться внутри промежутка $\overline{\alpha\beta}$ измеряется площадью прямоугольника, имеющего основанием этот промежуток, а высотой $f(\alpha)$; этой же величине равен и соответствующий подъем $F(\beta)-F(\alpha)$ ломаной линии

$y = F(x)$ при переходе от абсциссы α к абсциссе β. Так как в данном случае любой промежуток $\overline{lm}$ состоит из конечного числа интервалов, в каждом из которых плотность $f(x)$ сохраняет постоянное значение, то мы находим что вообще вероятность того, что $l < x < m$, равна

$$P_l^m = BC = \text{площ.}\,[lEFGHKLMNPQm],$$

т.-е. измеряется суммой площадей всех прямоугольников, построенных на основании $\overline{lm}$ и ограниченных сверху ломаной линией $y = f(x)$; этою же площадью измеряется и увеличение $F(m) - F(l) = BC$ ординаты ломаной линии $y = F(x)$.



Черт. 5.

К тому же заключению приходим, какова бы ни была функция $f(x)$ и соответствующая ей интегральная функция распределения вероятностей $F(x)$, которые можем рассматривать как пределы только что рассмотренных ломаных линий, когда интервалы между двумя вершинами стремятся к 0, так что число их неограниченно возрастает. Действительно, согласно основной теореме интегрального исчисления, равенство (79) равнозначно равенству

$$F(m) - F(l) = \int_l^m f(x)\,dx, \tag{81}$$

де вторая часть равенства (называемая определенным интегралом функции $f(x)$ между пределами l и m) представляет площадь, ограниченную слева и справа прямыми $x = l$ и $x = m$, снизу осью абсцисс и сверху линией $y = f(x)$ (черт. 5). В частности, площадь

бесконечно тонкого прямоугольника, имеющего основанием dx (дифференциал x) и высотой $f(x)$ (плотность вероятности), равная произведению $f(x)\,dx$, представляет вероятность

$$P_x^{x+dx} = F(x+dx) - F(x) = f(x)\,dx$$

того, что точка A (подчиняющаяся указанному закону распределения вероятностей) окажется в бесконечно малом промежутке $(x, x+dx)$.

Обычно закон распределения вероятностей, которому подчиняется точка A, характеризуют функцией $f(x)$, которая, согласно вышеизложенному, не может быть отрицательной и должна удовлетворять лишь одному условию, что

$$\int_{-\infty}^{\infty} f(x)\,dx = \underset{l=-\infty,\, m=\infty}{\text{пред.}} \int_l^m f(x)\,dx = 1, \tag{82}$$

так как, по предположению, точка должна находиться где-то на оси абсцисс $(-\infty < x < \infty)$, т.-е. $F(\infty) - F(-\infty) = 1$. Кривую, изображающую распределение плотностей вероятностей $y = f(x)$, называют для краткости просто кривой распределения вероятностей. Но необходимо хорошо усвоить, что ординаты $f(x)$ отнюдь не представляют вероятности точке A иметь абсциссу x (вероятность такого точного совпадения теоретически равна нулю): для получения графического изображения вероятности P_l^m точке A с абсциссой x находиться в промежутке $\overline{lm}$ нужно измерить площадь, заключенную между кривой $y = f(x)$, осью абсцисс и перпендикулярами к последней, восставленными в точках l и m, равную

$$P_l^m = \int_l^m f(x)\,dx. \tag{81 bis}$$

Если точка A не может находиться вне отрезка (a, b), то $f(x) = 0$ при $x < a$ и при $x > b$. В этом случае условие (82) превращается в условие

$$\int_a^b f(x)\,dx = 1, \tag{82 bis}$$

так как имеем тождественно $\int_{-\infty}^{a} f(x)\,dx = \int_b^{\infty} f(x)\,dx = 0$. Предполо-

жим, например, что точка A должна непременно находиться на отрезке $\overline{01}$ и плотность вероятности на этом отрезке определяется функцией $f(x)=2x$. Условие (82 bis), которое в данном случае получает форму

$$\int_0^1 f(x)\,dx=1,$$

соблюдено, так как $\int_0^1 2x\,dx=1$. Вероятность точке A оказаться в промежутке (α, β) равна площади трапеции

$$\int_\alpha^\beta 2x\,dx=\beta^2-\alpha^2=(\beta-\alpha)(\beta+\alpha),$$

которую легко получить при помощи элементарной геометрии, не прибегая к интегральному исчислению. (В данном случае $F(x)=x^2$ на отрезке 01.)

Вопрос о том, какова в каждом частном случае функция $f(x)$ распределения вероятностей, является одной из важнейших задач теории вероятностей и ее приложений, которую мы впоследствии исследуем более подробно. Сейчас заметим лишь, что если нельзя указать определенного конечного отрезка, внутри которого должна находиться точка A, так что абсцисса ее x может получать сколь угодно большие как положительные, так и отрицательные значения, то все же необходимо, чтобы функция $f(x)$ стремилась к 0, когда x безгранично возрастает, ибо в противном случае мы имели бы

$$\int_{-\infty}^{\infty} f(x)\,dx=\infty,$$

и, следовательно, условие (81) было бы нарушено. В частности, равномерное распределение вероятностей на всей бесконечной оси, очевидно, невозможно.

Пусть, например, из некоторой точки B, находящейся на расстоянии R от оси абсцисс, проводится в произвольном направлении прямая BCA (можно предположить, например, что точка B является центром циферблата хронометра, а прямая BC определяется положением стрелки хронометра, соответствующим моменту C какого-нибудь случайного явления), которую мы продолжаем в ту или

другую сторону до ее пересечения с осью абсцисс в точке A. В таком случае, принимая за ось ординат перпендикуляр, опущенный из B на ось абсцисс, и обозначая через θ угол, образуемый прямою BCA с осью ординат BO, видим, что угол θ всегда получает значения от $-\frac{\pi}{2}$ до $+\frac{\pi}{2}$; при этом, ввиду равноценности всех направлений, вероятность того, что $\theta_0 < \theta < \theta_1$, равна $\frac{\theta_1 - \theta_0}{\pi}$. Но, когда θ удовлетворяет указанному условию, абсцисса x точки A удовлетворяет неравенствам

$$R \operatorname{tg} \theta_0 < x < R \operatorname{tg} \theta_1;$$

поэтому, в частности, полагая $\theta_0 = -\frac{\pi}{2}$, находим, что вероятность неравенства

$$-\infty < x < R \operatorname{tg} \theta_1$$

равна $\frac{\theta_1}{\pi} + \frac{1}{2}$. Следовательно, полагая $R \operatorname{tg} \theta_1 = x_1$ (откуда $\theta_1 = \operatorname{arc\,tg} \frac{x_1}{R}$), находим, что вероятность $F(x_1)$ неравенства $x < x_1$ равна $F(x_1) = \frac{1}{\pi} \operatorname{arc\,tg} \frac{x_1}{R} + \frac{1}{2}$; поэтому, пользуясь формулой (79), заключаем, что плотность вероятностей для точки A

$$f(x) = F'(x) = \frac{1}{\pi} \cdot \frac{R}{x^2 + R^2}, \tag{83}$$

очевидно, стремится к 0 с возрастанием x и удовлетворяет условию (82).

Функция $f(x)$, имеющая наиболее важное теоретическое значение и многочисленные практические применения, соответствующая так называемому **нормальному** распределению вероятностей, имеет вид

$$f(x) = \frac{e^{-\frac{x^2}{k^2}}}{k\sqrt{\pi}},$$

где k — некоторая постоянная величина. В этом случае

$$\int_{-\infty}^{\infty} f(x)\, dx = 1,$$

так как известно, что

$$\int e^{-\frac{x^2}{k^2}}\frac{dx}{k}=\int_{-\infty}^{\infty} e^{-x^2}dx=\sqrt{\pi}\,^{1)},$$

и вероятность P_{α}^{β} того, что x заключен между α и β, равна

$$P_{\alpha}^{\beta}=F(\beta)-F(\alpha)=\frac{1}{k\sqrt{\pi}}\int_{\alpha}^{\beta} e^{-\frac{x^2}{k^2}}dx.$$

5. Понятие математического ожидания без особых затруднений распространяется на случаи, когда неизвестная величина может получать бесчисленное множество различных значений. Для большей отчетливости мы ограничимся предположением, что плотность $f(x)$ распределения вероятностей различных значений x конечна и непрерывна; кроме того допустим сначала, что $f(x)$ отлична от нуля только на некотором данном конечном отрезке (a, b), т.-е. что x должен удовлетворять неравенствам $a \leqslant x \leqslant b$. В таком случае математическое ожидание x определяется как предел, к которому стремится сумма [2])

$$I_n=\xi_1 P_a^{x_1}+\xi_2 P_{x_1}^{x_2}+\cdots+\xi_i P_{x_i}^{x_i+1}+\cdots+\xi_n P_{x_n}^{b}=$$

$$=\sum_a^b \xi_i P_{x_i}^{x_i+1}, \tag{84}$$

[1]) Действительно, полагая $I=\int_{-\infty}^{\infty} e^{-x^2}dx$, получаем

$$I^2=\int_{-\infty}^{\infty} e^{-x^2}dx\,.\int_{-\infty}^{\infty} e^{-y^2}dy=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}dx\,dy=$$

$$=\int_0^{2\pi}\int_0^{\infty} e^{-\rho^2}\rho\,d\rho\,d\theta=2\pi\int_0^{\infty} e^{-\rho^2}\rho\,d\rho=\pi.$$

Следовательно, $I=\sqrt{\pi}$.

[2]) I_n представляет собой математическое ожидание, которое имел бы x, если бы он мог получить только одно из n значений: $\xi_1, \xi_2, \cdots, \xi_i, \cdots, \xi_n$, при чем вероятность значения ξ_i равна $P_{x_i}^{x_i+1}$.

где $a = x_0 < x_1 < x_2 < \cdots < x_i < x_{i+1} < \cdots < x_n < x_{n+1} = b$ представляют собой точки деления отрезка (a, b), $P_{x_i}^{x_{i+1}}$ — вероятность того, что x заключен в промежутке (x_i, x_{i+1}), и, наконец, ξ_i — произвольное число, удовлетворяющее неравенству $x_i \leqslant \xi_i \leqslant x_{i+1}$, когда все разности $(x_{i+1} - x_i)$ равномерно стремятся к нулю. Принимая во внимание, что в силу равенства (80) и непрерывности функции $f(x)$

$$P_{x_i}^{x_{i+1}} = (x_{i+1} - x_i)[f(\xi_i) + \varepsilon(\xi_i)],$$

где $\varepsilon(\xi_i)$ равномерно стремится к нулю, когда равномерно стремятся к нулю все $(x_{i+1} - x_i)$, можем написать:

$$I_n = \xi_0 f(\xi_0)(x_1 - a) + \xi_1 f(\xi_1)(x_2 - x_1) + \cdots$$
$$\cdots + \xi_i f(\xi_i)(x_{i+1} - x_i) + \cdots + \xi_n f(\xi_n)(b - x_n) + \varepsilon_n,$$

полагая

$$\varepsilon_n = \xi_0 \varepsilon(\xi_0)(x_1 - a) + \xi_1 \varepsilon(\xi_1)(x_2 - x_1) + \cdots$$
$$\cdots + \xi_i \varepsilon(\xi_i)(x_{i+1} - x_i) + \cdots + \xi_n \varepsilon(\xi_n)(b - x_n).$$

Следовательно,

$$\textit{мат. ожид. } x = \text{пред. } I_n = \int_a^b x f(x)\, dx, \tag{84 bis}$$

так как пред. $\varepsilon_n = 0$.

Определяя аналогичным образом математическое ожидание любой непрерывной функции $\varphi(x)$ неизвестной величины x как пред. $\sum \varphi(\xi_i) P_{x_i}^{x_{i+1}}$, находим

$$\text{М. О. } \varphi(x) = \int_a^b \varphi(x) f(x)\, dx, \tag{85}$$

где $f(x)$ есть плотность распределения вероятностей величины x. В частности, при равномерном распределении вероятностей на отрезке (a, b), т.-е., если $f(x) = \frac{1}{b-a}$,

$$\text{М. О. } x = \frac{1}{b-a}\int_a^b x\, dx = \frac{a+b}{2},$$

$$\text{М. О. } x^2 = \frac{1}{b-a}\int_a^b x^2\, dx = \frac{1}{b-a} \cdot \frac{b^3 - a^3}{3} = \frac{b^2 + ab + a^2}{3}$$

Предположим, например, что вероятность новорожденному умереть, не достигнув возраста x, равна $F(x)$; тогда вероятность умереть в возрасте x, где $x_0 \leqslant x < x_1$ равна

$$F(x_1) - F(x_0) = \int_{x_0}^{x_1} f(x)\,dx,$$

где $F'(x) = f(x)$. Математическое ожидание продолжительности жизни, или средняя продолжительность жизни, для рассматриваемой группы новорожденных равно

$$\text{М. О. } x = M = \int_0^{\Omega} x f(x)\,dx,$$

где Ω — какое-нибудь число (например 200 лет) достаточно большое, чтобы быть уверенным, что никто этого возраста не переживет, так что

$$\int_0^{\Omega} f(x)\,dx = 1.$$

(Понятно, что без ущерба для правильности написанных формул можно поставить ∞ на место Ω, так как, по предположению, $f(x) = 0$ при $x > \Omega$.)

В согласии с данным выше общим определением математического ожидания величины x, которая изменяется непрерывно, можно дать приближенное значение числа M, если известны вероятности $p_k = F(k+1) - F(k) = \int_k^{k+1} f(x)\,dx$ новорожденному дожить до целого числа k лет и умереть, не достигнув $(k+1)$ лет. Действительно,

$$kp_k = k\int_k^{k+1} f(x)\,dx < \int_k^{k+1} x f(x)\,dx <$$
$$< (k+1)\int_k^{k+1} f(x)\,dx = (k+1)\,p_k;$$

поэтому

$$M_0 = p_1 + 2p_2 + \cdots + kp_k + \cdots < M <$$
$$< p_0 + 2p_1 + \cdots + (k+1)p_k + \cdots = M_1.$$

Величина M_0 (представляющая математическое ожидание целого числа лет жизни индивида рассматриваемого класса) носит название

укороченного математического ожидания продолжительности жизни новорожденного, при чем, очевидно,

$$M - M_0 < 1,$$

так как

$$M_1 - M_0 = p_0 + p_1 + \cdots + p_k + \cdots = 1.$$

Обычно за приближенное значение M принимают величину $M_0 + \frac{1}{2}$, которая соответствует допущению, что плотность вероятностей $f(x)$ сохраняет постоянное значение p_k при $k < x < k+1$.

Нетрудно проверить, что

$$M(h) = \frac{\int_h^{\Omega} x f(x)\, dx}{\int_h^{\Omega} f(x)\, dx}$$

представит математическое ожидание продолжительности жизни для индивида, который уже достиг возраста h.

Формулу (85) можно записать и в таком виде

$$\text{M. O. } \varphi(x) = \frac{\int_a^b \varphi(x) f(x)\, dx}{\int_a^b f(x)\, dx}, \qquad (85 \text{ bis})$$

принимая во внимание, что знаменатель последнего выражения равен 1. Заметим, что формула (85 bis) остается в силе, если заменить в ней $f(x)$ через $\lambda f(x)$, где λ — произвольная постоянная величина, так как при такой замене числитель и знаменатель окажутся умноженными на одно и то же число λ; поэтому формулой (85 bis) можно пользоваться и тогда, когда функция распределения $f(x)$ известна только с точностью до постоянного множителя.

Предположим теперь, что x может получать какие угодно значения от $-\infty$ до $+\infty$, при чем

$$\int_{-\infty}^{\infty} f(x)\, dx = 1. \qquad (82)$$

Если $I_n = \sum_a^b \xi_i P_{x_i}^{x_i+1}$ стремится к определенному пределу, когда a и b стремятся соответственно к $-\infty$ и $+\infty$, а интервалы $(x_{i+1} - x_i)$ стремятся к нулю, то

$$\text{М. О. } x = \text{пред.} \sum_a^b \xi_i P_{x_i}^{x_{i+1}} = \text{пред.} \int_a^b x f(x)\, dx =$$

$$= \int_{-\infty}^{\infty} x f(x)\, dx. \qquad (84\ \text{ter})$$

Таким образом условие, необходимое и достаточное (полагая функцию $f(x)$ непрерывной) для того, чтобы *мат. ожид.* x имело определенное значение, состоит в том, чтобы $\int_a^b x f(x)\, dx$ стремился к пределу при $a = -\infty$, $b = \infty$, т.-е. чтобы $\int_{-\infty}^{\infty} x f(x)\, dx$ имел смысл.

Если, например, $f(x) = \frac{e^{-x^2}}{\sqrt{\pi}}$, то

$$\text{М. О. } x = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} x e^{-x^2} dx = \underset{\substack{a=-\infty \\ b=\infty}}{\text{пред.}} \left[\frac{e^{-b^2} - e^{-a^2}}{2\sqrt{\pi}} \right] = 0.$$

Напротив, если, как в ранее рассмотренном примере (83),

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

то

$$\text{М. О. } x = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x\, dx}{1+x^2} = \underset{\substack{a=-\infty \\ b=\infty}}{\text{пред.}} \frac{1}{2\pi} [\log(1+b^2) - \log(1+a^2)],$$

и мы видим, что, в зависимости от того, какая из величин b^2 или a^2 будет быстрее возрастать,

$$\text{пред. } [\log(1+b^2) - \log(1+a^2)] =$$

$$= \text{пред. } \log \frac{1+b^2}{1+a^2} = \text{пред. } \log \frac{b^2}{a^2}$$

может быть равен любому положительному или отрицательному числу. Поэтому в данном случае понятие *мат. ожид.* x не имеет смысла.

Аналогичным образом получаем, вообще,

$$\text{М. О. } \varphi(x) = \int_{-\infty}^{\infty} \varphi(x) f(x)\, dx, \qquad (85\ \text{ter})$$

где $f(x)$ есть плотность распределения вероятностей величины x. Разумеется, выражение (85 ter) может [при данном $f(x)$] иметь смысл для одних функций $\varphi(x)$ и не иметь смысла для других.

Например, при $f(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$, применяя интегрирование по частям, находим

$$\text{М. О. } x^2 = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2}\, dx = \frac{-1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} x\, de^{-x^2} =$$

$$= \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-x^2}\, dx = \frac{1}{2}.$$

Читатель проверит без труда, что при $f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}$ М. О. x^2 бесконечно велико, т.-е. не имеет смысла.

Заметим, что свойства математических ожиданий, установленные в главе III, распространяются и на рассматриваемые здесь случаи, когда неизвестные величины могут получать бесчисленное множество различных значений. Это утверждение очевидно по отношению к лемме Чебышева [1]), где в нашем рассуждении никакой роли не играет число значений, принимаемых величиной x; обобщение же теоремы сложения математических ожиданий получается из замечания, что, вообще, согласно данному нами определению М. О. $x =$ $=$ пред. М. О. ξ, если ξ имеет пределом x. Следовательно, если теорема сложения справедлива для величин ξ и η, имеющих соответственно пределами x и y, то

$$\text{М. О. } (x+y) = \text{пред. М. О. } (\xi+\eta) =$$
$$= \text{пред. } [\text{М. О. } \xi + \text{М. О. } \eta] = \text{М. О. } x + \text{М. О. } y.$$

[1]) См. стр. 100.

Упражнения. 1) На гладком горизонтальном полу начерчены отстоящие друг от друга на расстоянии $2a$ параллельные прямые. Требуется определить вероятность P того, что монета радиуса $R < a$, брошенная на пол, заденет одну из указанных прямых.

Ответ: $P = \frac{R}{a}$. Действительно, обозначая через x расстояние от центра монеты до ближайшей к ней прямой, замечаем, что $0 \leqslant x \leqslant a$, при чем все значения в этом промежутке можно считать одинаково вероятными. Благоприятствовать пересечению монеты с прямой будут все значения $x \leqslant R$; следовательно, $P = \frac{R}{a}$.

2) На тот же пол бросают проволочную конвексную фигуру (многоугольник) H, периметр которой равен $2s$, при чем наибольшая ее ширина менее $2a$ (так что H не может одновременно пересечь двух параллельных прямых). Какова вероятность P того, что H пересечет одну из данных параллельных прямых?

Ответ: $P = \frac{s}{\pi a}$. Для доказательства полагаем сначала, что H есть многоугольник с n сторонами, длины которых обозначим через $2l_1, \cdots, 2l_n$; в таком случае вероятность x_i, что сторона, длина которой равна $2l_i$, пересечет одну из данных прямых, пропорциональна длине этой стороны, так как, деля какой-нибудь отрезок AB на равные части, можно считать одинаково вероятным, что точка пересечения AB с данной прямой будет находиться на любой из равных частей. Следовательно, $x_i = 2\alpha l_i$, где α есть одна и та же (пока неизвестная) постоянная величина, не зависящая от l_i. С другой стороны, замечая, что благодаря конвексности многоугольника H пересечение его с данной прямой обусловлено тем, что две из его сторон $2l_i$ и $2l_k$ одновременно пересекаются с этой прямой, находим обозначая через $x_{ik} = x_{ki}$ вероятность такого совмещения,

$$P = \frac{1}{2}[(x_{12} + x_{13} + \cdots + x_{1n}) + (x_{21} + x_{23} + \cdots + x_{2n}) + \cdots] =$$
$$= \frac{1}{2}(x_1 + x_2 + \cdots + x_n) = \alpha\ (l_1 + l_2 + \cdots + l_n) = \alpha s,$$

так как $x_{i1} + x_{i2} + \cdots + x_{in} = x_i$. Так как равенство $P = \alpha s$ не зависит от числа и размеров сторон, то оно справедливо и для криволинейного конвексного контура, который всегда можно рассматривать как предел многоугольника с бесконечно возрастающим числом сторон. В частности, эта формула применима и к окружности, для которой мы нашли выше $P = \frac{R}{a} = \frac{2\pi R}{2\pi a}$; но так как периметр окружности $2s = 2\pi R$, то $P = \frac{s}{\pi a}$; отсюда следует, что $\alpha = \frac{1}{\pi a}$. Поэтому, во-первых, доказана для любого контура формула $P = \frac{s}{\pi a}$ и, во-вторых, установлено, что вероятность x отрезку длины $2l$ (не превышающей $2a$) пересечь одну из рассматриваемых параллельных

прямых равна $x=\frac{2l}{\pi a}$ (последний результат дает решение задачи, известной под именем задачи Бюффона).

3) Берутся произвольно 2 точки на окружности; какова вероятность P того, что хорда, соединяющая их, меньше радиуса R окружности?

Ответ: $P=\frac{1}{3}$. Действительно, без ущерба для общности можно считать одну точку определенной; полагая все положения другой точки на окружности равновероятными и замечая, что благоприятными будут ее положения на соответствующей дуге, равной $\frac{1}{3}$ окружности, находим $P=\frac{1}{3}$.

4) Определить математическое ожидание расстояния d между двумя точками A и M окружности радиуса R, полагая распределение вероятностей на окружности равномерным [1]).

Ответ: *мат. ожид.* $d=\frac{4R}{\pi}$. Действительно, считая одну точку A определенной, находим, что М. О. $d=\frac{R}{\pi}\int_{-\pi}^{\pi}\cos\frac{\theta}{2}\,d\theta=\frac{4R}{\pi}$. Точно так же получим: М. О. $d^2=\frac{2R^2}{\pi}\int_{-\pi}^{\pi}\cos^2\frac{\theta}{2}\,d\theta=2R^2$.

5) Вычислить М. О. x^k, если вероятность, что $x<x_1$, равна

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x_1}e^{-\frac{x^2}{2\sigma^2}}dx.$$

---

[1]) Это допущение соответствовало бы, например, случаю, когда обе точки представляли бы собой моменты наступления двух независимых событий, отмечаемые хронометром. Но в других опытах может оказаться, что это допущение не осуществляется. Действительно, положим, что (как в первой задаче) бросают монету или круглый диск радиуса R на пол, на котором начерчены параллельные прямые, отстоящие друг от друга на расстоянии $2R$, так что диск пересекает, вообще, одну и только одну из наших прямых, отрезая таким образом на этой прямой хорду определенной длины (от 0 до $2R$). Здесь, как и в первой задаче, естественно принять, что любые положения центра диска между двумя прямыми равновероятны, а потому, принимая во внимание, что хорда будет менее радиуса R, если расстояние центра до ближайшей прямой больше, чем $\frac{R\sqrt{3}}{2}$, заключаем, что при указанной постановке опыта вероятность, что хорда меньше радиуса, равна $1-\frac{\sqrt{3}}{2}$, т.-е. значительно меньше, чем при сделанном выше допущении, соответствующем опыту с хронометром.

Ответ. М. О. $x^k=0$, если k — нечетное; М. О. $x^{2m}=1\cdot 3\cdots(2m-1)\cdot\sigma^{2m}$. Первое утверждение вытекает из того, что плотность распределения вероятностей есть четная функция. С другой стороны;

$$\text{М. О. } x^{2m}=\frac{1}{\sigma\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty}x^{2m}e^{-\frac{x^2}{2\sigma^2}}dx=\frac{\sigma}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty}x^{2m-1}d\left(-e^{-\frac{x^2}{2\sigma^2}}\right)=$$

$$=(2m-1)\,\sigma^2\text{ М. О. }x^{2m-2},$$

откуда последовательной подстановкой получаем указанную формулу.

6) Закон распределения вероятностей, одинаковый для точек A и A_1, на отрезке LM симметричен по отношению к середине O отрезка. Пусть P будет вероятность точке A (как и точке A_1) оказаться внутри промежутка BOC, при чем $BO=OC=\frac{1}{4}LM$; полагая $a=\frac{P}{1-P}$, $x=\frac{R}{1-R}$, где R есть вероятность, что середина N отрезка AA_1 упадет внутри промежутка BOC, показать, что

$$x>a+\frac{1}{2}.$$

Отв.: Действительно, N окажется между B и C, если A и A_1 находятся одновременно на BO или OC, а также и во всех случаях, когда точка O находится между A и A_1; поэтому

$$R>\frac{1}{2}(P^2+1),\text{ откуда } x>\frac{1+P^2}{1-P^2}>\frac{P}{1-P}+\frac{1}{2}=a+\frac{1}{2}.$$

Отсюда следует, что, обозначая через R_m вероятность средней арифметической 2^m точек, подчиняющихся тому же самому закону распределения, оказаться внутри BOC, имеем $\frac{R_m}{1-R_m}>\frac{m}{2}$, т.-е., при возрастании m, пред. $R_m=1$.

ГЛАВА ШЕСТАЯ

СПОСОБ КОНЕЧНЫХ РАЗНОСТЕЙ И МЕТОД ПРОИЗВОДЯЩИХ ФУНКЦИЙ

1. Одним из наиболее важных аналитических методов вычисления вероятностей и математических ожиданий является способ конечных разностей, который может быть уяснен на следующих примерах.

Пример первый. Два игрока A и B повторяют некоторую партию до тех пор, пока один из них не проиграет весь свой капитал. Пусть капитал первого равен a рублям, капитал второго равен b рублям; вероятности выигрыша отдельной партии для каждого из игроков соответственно равны p и q $(p+q=1)$, при чем выигрыш одного (и одновременный проигрыш другого) равен 1 рублю. Требуется определить вероятность P разорения игрока B, т. е. перехода всего его капитала в руки первого игрока A.

Очевидно, что сумма обоих капиталов во время всей игры неизменно остается равной $(a+b)$ рублям. Обозначим через P_y вероятность разорения игрока B, когда из указанной суммы у игрока A имеется y рублей. В частности, следовательно,

$$P_{a+b}=1, \qquad P_0=0,$$

так как, если игрок A овладел всей суммой $(a+b)$, то игрок B разорен; напротив, когда у игрока A ничего не осталось, то игра закончена в пользу B; искомая же вероятность $P=P_a$. Вообще

$$P_y=p\cdot P_{y+1}+(1-p)P_{y-1}, \tag{86}$$

так как, если перед некоторой партией у игрока A было y рублей, то его окончательный выигрыш (разорение B), вероятность которого равна P_y, наступит в одном из двух несовместимых случаев:

либо он выиграет наступающую партию, и совместно с этим обстоятельством будет разорен игрок B, либо он проиграет наступающую партию, и, все же, в конечном итоге будет разорен B. Первое из этих совмещений, согласно нашим обозначениям, имеет вероятность pP_{y+1} [так как, в случае выигрыша партии, у игрока A окажется $(y+1)$ рублей], второе же совмещение имеет вероятность $qP_{y-1}=(1-p)P_{y-1}$ [так как при проигрыше наступающей партии у игрока A останется $(y-1)$ рублей].

Уравнение (86) называется уравнением в конечных разностях (или рекурентной формулой). Для решения его напишем уравнение (86), отняв из обеих частей равенства $P_{y-1}+(1-p)P_y$, в виде

$$(1-p)(P_y-P_{y-1})=p(P_{y+1}-P_y);$$

полагая затем

$$Z_y=P_y-P_{y-1},$$

получим

$$(1-p)Z_y=pZ_{y+1}. \tag{87}$$

Таким образом, если $p=\frac{1}{2}$, то

$$Z_y=Z_{y+1}=h,$$

где h — некоторая постоянная, не зависящая от y. Следовательно, в этом частном случае $P_{y+1}=P_y+h$; поэтому числа P_y являются членами арифметической прогрессии, при чем, так как $P_0=0$, то $P_1=h$, и вообще $P_y=yh$; но, замечая, что $P_{a+b}=1$, заключаем, что $1=(a+b)h$, откуда $h=\frac{1}{a+b}$. Следовательно,

$$P=P_a=\frac{a}{a+b},$$

т.-е. вероятность разорения игрока B в данном случае равна $\frac{a}{a+b}$, и вероятность разорения A равна $\frac{b}{a+b}$; мы получили таким образом результат, найденный раньше (стр. 98) способом математических ожиданий, так как при $p=\frac{1}{2}$ рассматриваемая игра безобидна.

Переходим теперь к общему случаю, когда $p \gtrless \frac{1}{2}$. Из (87) видим, что Z_y представляет собой член геометрической прогрессии с знаменателем $\frac{1-p}{p} = \frac{q}{p}$; поэтому, складывая равенства

$$Z_1 = P_1,\ Z_2 = P_2 - P_1, \ldots, Z_y = P_y - P_{y-1},$$

находим:

$$P_y = Z_1 + Z_2 + \cdots + Z_y = Z_1 \left[1 + \frac{q}{p} + \cdots + \left(\frac{q}{p}\right)^{y-1} \right] =$$

$$= Z_1 \frac{1 - \left(\frac{q}{p}\right)^y}{1 - \frac{q}{p}}, \tag{88}$$

и, в частности,

$$1 = P_{a+b} = Z_1 \frac{1 - \left(\frac{q}{p}\right)^{a+b}}{1 - \frac{q}{p}}. \tag{88 bis}$$

Деля (88) на (88 bis), получаем наконец

$$P_y = \frac{1 - \left(\frac{q}{p}\right)^y}{1 - \left(\frac{q}{p}\right)^{a+b}};$$

поэтому

$$P = P_a = \frac{1 - \left(\frac{q}{p}\right)^a}{1 - \left(\frac{q}{p}\right)^{a+b}}. \tag{89}$$

Аналогичным образом для вероятности разорения A получим:

$$Q = \frac{1 - \left(\frac{p}{q}\right)^b}{1 - \left(\frac{p}{q}\right)^{a+b}}.$$

Нетрудно проверить, что $P + Q = 1$.

Положим, для определенности, $p > q$ и допустим, что капитал b игрока B чрезвычайно велик, так что теоретически можем поло-

жить $b=\infty$. В таком случае $\left(\frac{q}{p}\right)^{a+b}$ имеет пределом 0, и формула (89) принимает вид

$$\text{пред. } P=1-\left(\frac{q}{p}\right)^{a},$$

откуда

$$\text{пред. } Q=\left(\frac{q}{p}\right)^{a}.$$

Следовательно (в противоположность безобидной игре), если игра выгодна для A, благодаря тому, что $p>q$, то, при сколь угодно большом капитале у его противника, разорение A не достоверно; напротив, если его собственный капитал достаточно велик, то вероятность Q его разорения ничтожна, и, продолжая игру достаточно долго, игрок A имеет много шансов выиграть сколь угодно большую сумму денег b.

Если, например, $p=\frac{3}{4}$ (так что матем. ожид. выигрыша для A равно $\frac{3}{4}\cdot 1-\frac{1}{4}\cdot 1=\frac{1}{2}$ рубля $=50$ коп.), то при капитале $a=10$ рублям вероятность Q его разорения равна $\left(\frac{1}{3}\right)^{10}\neq 0{,}000\,017$, между тем как вероятность выиграть произвольно большую сумму денег равна $P\neq 0{,}999\,983$.

Сохраняя предположение $p\geqslant q$ (т.-е. полагая игру выгодной или безобидной), рассмотрим практически важный вопрос, какова вероятность Q разорения игрока A, имеющего капитал a, при игре с безгранично богатым противником, если число партий не может превысить данного значения n. Разорение A может наступить после a партий, если все партии были для него неудачны, после $(a+2)$ партий при одной удачной и, вообще, после $(a+2i)$ партий при i удачных партиях, где $a+2i\leqslant n$. Вероятность, что из $(a+2i)$ партий $(a+i)$ неудачны и i партий удачны для A, равна

$$C^{i}_{a+2i}p^{i}q^{a+i};$$

однако не всякая серия из таких $(a+2i)$ партий может осуществиться, а только такая, в которой излишек проигранных партий над выигранными ни разу до конца не достигнет a (так как иначе разорение наступило бы еще раньше). Поэтому, изображая серию из $(a+2i)$ партий в виде ряда букв $A\bar{A}\bar{A}\ldots A\bar{A}\bar{A}$, где A соответствует выигрышу, а $\bar{A}$ — проигрышу,

замечаем, что лишь те серии из $(a+2i)$ букв осуществятся, в которых, считая буквы справа налево, число $\bar{A}$ всегда больше, чем число A (ибо, если бы число $\bar{A}$ оказалось равным числу A, то, отбрасывая всю эту последнюю группу букв, мы получили бы слева меньшую серию, где излишек $\bar{A}$ над A также был бы равен a); как было показано (стр. 65), число таких серий относится к общему числу серий, как a к $(a+2i)$. Таким образом вероятность разориться как раз после $(a+2i)$-й партии равна

$$C^{i}_{a+2i}p^{i}q^{a+i}\frac{a}{a+2i}=\frac{a\cdot(a+i+1)\cdots(a+2i-1)}{i!}p^{i}q^{a+i}.$$

Поэтому вероятность Q разорения игрока A в течение n партий равна сумме членов указанного вида

$$Q=q^{a}\left[1+apq+\frac{a(a+3)}{2!}p^{2}q^{2}+\ldots+\right.$$
$$\left.+\frac{a(a+i+1)\cdots(a+2i-1)}{i!}p^{i}q^{i}+\ldots\right],$$

где наибольшее значение i равно целому числу $\frac{n-a}{2}$ или $\frac{n-a-1}{2}$.

Например, если $n=10$, $a=3$, $p=q=\frac{1}{2}$, то в этой безобидной игре вероятность разорения игрока A равна

$$Q=\left(\frac{1}{2}\right)^{3}\left[1+3\cdot\frac{1}{4}+\frac{3\cdot6}{2!}\left(\frac{1}{4}\right)^{2}+\frac{3\cdot7\cdot8}{3!}\left(\frac{1}{4}\right)^{3}\right]=\frac{11}{32}.$$

Если $n=10$, $a=2$, $p=\frac{2}{3}$, $q=\frac{1}{3}$, то в этой выгодной игре вероятность разорения

$$Q=\left(\frac{1}{3}\right)^{2}\left[1+2\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)+\frac{2\cdot5}{2!}\left(\frac{2}{9}\right)^{2}+\frac{2\cdot6\cdot7}{3!}\left(\frac{2}{9}\right)^{3}+\frac{2\cdot7\cdot8\cdot9}{4!}\left(\frac{2}{9}\right)^{4}\right]=\frac{4\,259}{19\,683}.$$

Но, между тем как в первом примере с возрастанием n вероятность разорения стремится к достоверности, во втором она приближается лишь к $\frac{1}{4}$.

2. *Пример второй.* Даны два ящика с белыми и черными шарами; в первом ящике, при общем числе шаров N, находится M белых шаров, а во втором ящике имеется M_1 белых шаров при общем числе N_1 шаров. Производится ряд опытов, связанных между собой следующим образом: из обоих ящиков вынимается одновременно наудачу по 1 шару, который кладется в другой ящик; перемешавши шары в обоих ящиках, повторяют тот же процесс двойного

перекладывания l раз. Требуется определить математическое ожидание числа белых шаров в первом ящике по окончании указанных l опытов.

Если X_k есть *мат. ожид.* числа белых шаров в первом ящике после k опытов, а Y_k — *мат. ожид.* числа белых шаров во втором ящике, то, во-первых,

$$X_k + Y_k = M + M_1$$

(при чем $X_0 = M$, $Y_0 = M_1$) и, во-вторых, при $(k+1)$-м опыте вероятность p_{k+1}, что из первого ящика будет вынут белый шар, равна $p_{k+1} = \frac{X_k}{N}$, а вероятность P_{k+1}, что из второго ящика будет вынут белый шар, равна $P_{k+1} = \frac{Y_k}{N_1}$. Но, с другой стороны, *мат. ожид.* числа белых шаров, вынутых из первого ящика в течение k опытов, по формуле (64) равно $p_1 + p_2 + \ldots + p_k = \sum_{i=1}^{i=k} p_i$, а *мат. ожид.* числа белых шаров, перешедших из второго ящика в первый, равно $P_1 + P_2 + \cdots + P_k = \sum_{i=1}^{i=k} P_i$; следовательно,

$$X_k = M - \sum_{i=1}^{i=k} p_i + \sum_{i=1}^{i=k} P_i = M + \sum_{i=1}^{i=k} \left(\frac{Y_{i-1}}{N_1} - \frac{X_{i-1}}{N} \right) =$$

$$= M + \sum_{i=1}^{i=k} \left(\frac{M + M_1 - X_{i-1}}{N_1} - \frac{X_{i-1}}{N} \right).$$

Если заменим в этом равенстве k через $k+1$, то к сумме, стоящей во второй части, прибавится еще одно слагаемое $\frac{M + M_1 - X_k}{N_1} - \frac{X_k}{N}$; поэтому, вычитая из этого нового равенства предыдущее, получим:

$$X_{k+1} - X_k = \frac{M + M_1 - X_k}{N_1} - \frac{X_k}{N}. \tag{90}$$

Из этого уравнения в конечных разностях можем последовательно определить $X_1, X_2, \ldots, X_l$, так как нам известно, что $X_0 = M$

таким образом, полагая в (90) $k=0$, получаем сначала $X_1=M+\frac{M_1}{N_1}-\frac{M}{N}$; полагая затем $k=1$, находим:

$$X_2=\left(M+\frac{M_1}{N_1}-\frac{M}{N}\right)\left(1-\frac{1}{N_1}-\frac{1}{N}\right)+\frac{M+M_1}{N_1}$$

и т. д. Но удобнее сразу дать формулу для X_k, каково бы ни было число k. Прием, который мы для этого сейчас применим, пригоден для решения всякого линейного уравнения в конечных разностях первого порядка с постоянными коэффициентами вида

$$X_{k+1}-X_k=a-bX_k. \tag{91}$$

Замечаем сначала, что уравнению (91) удовлетворяет, в частности, постоянная величина $X=\frac{a}{b}$; поэтому, если бы было дано, что $X_0=\frac{a}{b}$, то мы имели бы также $X_1=\frac{a}{b}$, $X_2=\frac{a}{b}, \ldots, X_k=\frac{a}{b}$. Но это решение есть частное решение, соответствующее лишь предположению, что $X_0=\frac{a}{b}$; положим, вообще, $X_k=\frac{a}{b}+x_k$; тогда, подставляя это значение (и соответствующее значение $X_{k+1}=\frac{a}{b}+x_{k+1}$) в уравнение (91), получим:

$$x_{k+1}-x_k=a-b\left(\frac{a}{b}+x_k\right)=-bx_k,$$

т.-е.

$$x_{k+1}=(1-b)x_k,$$

откуда заключаем, что x_k есть член геометрической прогрессии с знаменателем $(1-b)$; следовательно,

$$x_k=(1-b)^k x_0 \text{ и } X_k=\frac{a}{b}+(1-b)^k x_0,$$

где x_0 — произвольная постоянная величина. Но, так как, в частности, $X_0=\frac{a}{b}+x_0$, то получаем общую формулу

$$X_k=\frac{a}{b}+(1-b)^k\left(X_0-\frac{a}{b}\right), \tag{92}$$

дающую решение уравнения (91), каково бы ни было данное значение X_0. Заметим, что если $0 < b < 2$, то X_k имеет всегда пределом $\frac{a}{b}$, когда k неограниченно возрастает, так как $(1-b)^k$ стремится к 0.

В рассматриваемом нами уравнении (90) $a = \frac{M+M_1}{N_1}$, $b = \frac{1}{N} + \frac{1}{N_1}$; поэтому $\frac{a}{b} = \frac{M+M_1}{N+N_1} N$, $1 - b = 1 - \frac{1}{N} - \frac{1}{N_1}$, $X_0 = M$, $X_0 - \frac{a}{b} = M - \frac{M+M_1}{N+N_1} N = \frac{MN_1 - NM_1}{N+N_1}$; следовательно,

$$X_k = \frac{M+M_1}{N+N_1} N + \frac{MN_1 - NM_1}{N+N_1} \left(1 - \frac{1}{N} - \frac{1}{N_1}\right)^k. \quad (92\ \text{bis})$$

В частности, если $MN_1 - NM_1 = 0$, т.-е. если $\frac{M}{N} = \frac{M_1}{N_1}$, что означает, что пропорция белых шаров первоначально одна и та же в обоих ящиках, то, каково бы ни было число k перекладываний, *мат. ожид.* числа белых шаров

$$X_k = \frac{M+M_1}{N+N_1} N = M$$

остается неизменным. Во всяком случае, $\underset{k=\infty}{\text{пред.}}\ X_k = \frac{M+M_1}{N+N_1} N$, а следовательно, $\underset{k=\infty}{\text{пред.}}\ Y_k = (M+M_1)\left(1 - \frac{N}{N+N_1}\right) = \frac{M+M_1}{N+N_1} N_1$, следовательно, $\frac{\text{пред. } X_k}{N} = \frac{\text{пред. } Y_k}{N_1}$, так что вероятности появления белого шара из обоих ящиков стремятся стать одинаковыми при возрастании k.

Допустим, например, что в первом ящике сначала были только белые шары ($M = N$), а во втором ящике — только черные ($M_1 = 0$); тогда

$$X_k = \frac{N^2}{N+N_1} + \frac{NN_1}{N+N_1}\left(1 - \frac{1}{N} - \frac{1}{N_1}\right)^k.$$

Если бы, кроме того, N_1 было бесконечно велико, т.-е. если бы взамен любого шара, вынутого из первого ящика, на его место клался всегда черный шар[1]), то мы получили бы

$$X_k = N\left(1 - \frac{1}{N}\right)^k, \tag{92 bis}$$

так как $\underset{N_1=\infty}{\text{пред.}} \frac{N^2}{N+N_1} = 0, \quad \underset{N_1=\infty}{\text{пред.}} \frac{NN_1}{N+N_1} = N.$

В последнем случае, если N и k велики, формула (92 bis) может быть заменена приближенно формулой

$$X_k \neq Ne^{-\frac{k}{N}}.$$

3. *Пример третий.* Производится ряд опытов, причем известна вероятность p_k наступления события A в m-м опыте в случае k ($k < m$) появлений события A в предшествующих опытах. Требуется определить: 1) вероятность B_m^k, что при m опытах событие A произойдет ровно k раз; 2) вероятность P_m^k, что для появления события A k раз потребуется m опытов; 3) математическое ожидание M_k числа опытов m, которое понадобится для того, чтобы A произошло k раз.

Замечаем, что

$$B_{m+1}^{k+1} = B_m^k p_k + B_m^{k+1} q_{k+1} \quad (\text{где } q_{k+1} = 1 - p_{k+1}), \tag{93}$$

так как для того, чтобы при $(m+1)$-м опыте событие A произошло $k+1$ раз, нужно одно из двух: либо, чтобы оно произошло k раз при m опытах и снова осуществилось в $(m+1)$-м опыте, либо, чтобы событие A произошло $(k+1)$ раз при m опытах, но не наступило в $(m+1)$-м опыте. Для решения уравнения (93) применим способ, называемый методом производящих функций. Положим

$$f_k(y) = B_0^k + B_1^k y + \ldots + B_m^k y^m + \ldots;$$

в таком случае, умножая обе части равенства (93) на y^{m+1}, получим:

$$B_{m+1}^{k+1} y^{m+1} = y\left[y^m B_m^k p_k + y^m B_m^{k+1} q_{k+1}\right].$$

---

[1]) Рекомендуем читателю в виде упражнения непосредственно решить указанную задачу, которая, в этом частном более простом случае, приводит к уравнению $X_{k+1} - X_k = -\frac{X_k}{N}$ вместо уравнения (90).

Придавая m все целые значения (0, 1, 2, ...) и складывая все таким образом полученные равенства, получим:

$$f_{k+1}(y) = y\left[p_k f_k(y) + q_{k+1} f_{k+1}(y)\right],$$

так как $B_0^{k+1} = 0$. Поэтому, при $k \geqslant 0$,

$$f_{k+1}(y) = \frac{p_k y}{1 - q_{k+1} y} f_k(y),$$

откуда последовательной подстановкой ($k = 0, 1, \ldots$) получаем:

$$f_{k+1}(y) = \frac{p_0 p_1 p_2 \cdots p_k y^{k+1} f_0(y)}{(1 - q_1 y)(1 - q_2 y) \cdots (1 - q_{k+1} y)}.$$

Но $f_0(y) = B_0^{(0)} + B_1^{(0)} y + \ldots = 1 + q_0 y + q_0^2 y^2 + \ldots = \frac{1}{1 - q_0 y}$;

поэтому производящая функция вероятностей B_m^{k+1} равна

$$f_{k+1}(y) = \frac{p_0 p_1 \cdots p_k y^{k+1}}{(1 - q_0 y)(1 - q_1 y) \cdots (1 - q_{k+1} y)}, \tag{94}$$

а именно, если разложить эту функцию по степеням y, то B_m^{k+1} будет коэффициентом при y^m.

Замечаем далее, что $P_{m+1}^{k+1} = B_m^k p_k$, так как для того, чтобы событие A произошло в $(k+1)$-й раз, именно при $(m+1)$-м опыте, нужно, чтобы при m опытах оно произошло k раз и наступило в последний раз при $(m+1)$-м опыте.

Поэтому, полагая $F_k(y) = P_0^k + P_1^k y + \ldots + P_m^k y^m + \cdots$, находим:

$$F_{k+1}(y) = p_k y f_k(y) = \frac{p_0 p_1 \cdots p_k y^{k+1}}{(1 - q_0 y) \cdots (1 - q_k y)}. \tag{94 bis}$$

В частности, если p_k сохраняет постоянное значение, т.-е. $p_k = p$, то

$$F_k(y) = \left(\frac{py}{1 - qy}\right)^k = p^k y^k \left[1 + kqy + \frac{k(k+1)}{2} q^2 y^2 + \ldots\right];$$

поэтому коэффициент при y^m равен

$$P_m^k = \frac{k(k+1) \cdots (m-1)}{(m-k)!} p^k q^{m-k} = C_{m-1}^{k-1} p^k q^{m-k};$$

точно так же из

$$f_k(y) = \frac{(py)^k}{(1 - qy)^{k+1}} = p^k y^k \left[1 + (k+1) qy + \frac{(k+1)(k+2)}{2} q^2 y^2 + \ldots\right]$$

находим известное нам уже значение

$$B_m^k = \frac{(k+1) \cdots m}{(m-k)!} p^k q^{m-k} = C_m^k p^k q^{m-k}. \tag{52}$$

Возвращаясь к общему случаю, получим математическое ожидание числа m опытов, необходимых для появления события A k раз, замечая, что

$$M_k = \text{мат. ожид. } m = k\,P_k^k + (k+1)\,P_{k+1}^k + \ldots + m\,P_m^k + \ldots;$$

поэтому из

$$F_k'(y) = P_1^k + \ldots + k\,P_k^k\,y^{k-1} + \ldots + m\,P_m^k\,y^{m-1} + \ldots,$$

где $P_i^k = 0$ при $i < k$, заключаем, что

$$M_k = F_k'(1).$$

Но дифференцирование формулы (94 bis) (если заменить в ней $k+1$ через k) дает

$$F_k'(y) = \frac{p_0 p_1 \cdots p_{k-1} y^k}{(1-q_0 y)\cdots(1-q_{k-1} y)}\left[\frac{k}{y} + \frac{q_0}{1-q_0 y} + \frac{q_1}{1-q_1 y} + \ldots\right].$$

Поэтому

$$M_k = F_k'(1) = k + \frac{q_0}{p_0} + \frac{q_1}{p_1} + \ldots + \frac{q_{k-1}}{p_{k-1}} = \frac{1}{p_0} + \frac{1}{p_1} + \ldots + \frac{1}{p_{k-1}}.$$

Если, например (частный случай примера второго), $p_k = \frac{m-k}{n}$, где $m \leqslant n$, то математическое ожидание числа выниманий, необходимых для извлечения всех m белых шаров из ящика с n шарами (когда после каждого извлечения вместо вынутого шара обратно кладется черный шар), равно

$$M_m = \frac{n}{m} + \frac{n}{m-1} + \frac{n}{m-2} + \cdots + n,$$

и при m весьма большом пред. $\frac{M_m}{n \log m} =$ пред. $\frac{1 + \frac{1}{2} + \ldots + \frac{1}{m}}{\log m} = 1.$

Упражнения. 1) Вычислить математическое ожидание m^3 и m^4, где m есть число появлений события A при n независимых опытах, если при каждом опыте вероятность A равна p.

Ответ: *М. О.* $m^3 = n^3 p^3 + 3n^2 p^2 q + npq\,(q-p)$; М. О. $m^4 = n^4 p^4 + 6n^3 p^3 q + n^2 p^2 q\,(7q - 4p) + npq\,(1 - 6pq)$.

Для вывода этих формул полагаем $f(x) = (px+q)^n$, тогда

$$f'(1) = np = \text{М. О. } m;\ f''(1) = n(n-1)\,p^2 = \text{М. О. } m\,(m-1);$$
$$f'''(1) = n(n-1)(n-2)\,p^3 = \text{М. О. } m\,(m-1)(m-2);$$

$$f^{(IV)}(1) = n(n-1)(n-2)(n-3)\,p^4 = \text{М. О. } m\,(m-1)(m-2)(m-3).$$

Требуемые формулы получаются благодаря тождествам

$$m^3 = m\,(m-1)(m-2) + 3m\,(m-1) + m;\ m^4 = m\,(m-1)(m-2)(m-3) + 6\,m\,(m-1)(m-2) + 7\,m\,(m-1) + m.$$

2) Игрок повторяет большое число n раз невыгодную игру, при которой a есть вероятность проиграть 1 рубль, $c < a$ есть вероятность выиграть 1 рубль и $b = 1 - a - c$ есть вероятность безразличного исхода. Невыгодность игры компенсируется тем, что, когда игрок разоряется, он освобождается от платежа при проигрыше, так что тогда $a + b$ является вероятностью безразличного результата партии, при чем c попрежнему есть вероятность выиграть 1 рубль. Допускаем, что P_h есть предел, к которому стремится вероятность $P_{n,h}$ для игрока иметь h рублей после n партий, если n неограниченно возрастает; требуется определить P_h.

Ответ: $P_h = \left(1 - \frac{c}{a}\right)\left(\frac{c}{a}\right)^h$. Действительно, если $h \geqslant 1$, то

$$P_{n+1,h} = a P_{n,h+1} + b P_{n,h} + c P_{n,h-1},$$

и, кроме того, $P_{n+1,0} = (a+b) P_{n,0} + a P_{n,1}$. Следовательно (при $h > 0$)

$$(a+c) P_h = a P_{h+1} + c P_{h-1},$$

откуда

$$P_h = A + B\left(\frac{c}{a}\right)^h,$$

где A и B — постоянные коэффициенты; при этом $A = 0$, так как $P_1 = \left(\frac{c}{a}\right) P_0$, а B определяется из условия, что $\sum_0^\infty P_h = 1$.

3) Пусть A и B некоторые 2 из несовместимых исходов опыта E, имеющие вероятности соответственно равные a и b, где $a > b$; пусть $x > 0$ данное целое число. Требуется определить вероятность P_x, что, при неограниченном повторении опыта E, когда-нибудь осуществится равенство $h - k = x$, где h есть число появлений событий B, а k — соответствующее число появлений события A.

Ответ: $P_x = \left(\frac{b}{a}\right)^x$. Действительно, для осуществления равенства $h - k = x$ нужно сначала осуществить $h - k = x - 1$, а затем еще $h - k = 1$. Поэтому $P_x = P_1 \cdot P_{x-1}$; следовательно, $P_x = \lambda^x$, где $\lambda < 1$, — постоянное положительное число. Но, с другой стороны, $P_x = a P_{x+1} + b P_{x-1} + (1-a-b) P_x$; поэтому λ должно удовлетворять уравнению $(a+b)\lambda^x = a\lambda^{x+1} + b\lambda^{x-1}$, откуда $(a+b)\lambda = a\lambda^2 + b$, и, отбрасывая корень $\lambda = 1$, находим $\lambda = \frac{b}{a}$.

ЧАСТЬ ТРЕТЬЯ

ЗАКОН БОЛЬШИХ ЧИСЕЛ

ГЛАВА ПЕРВАЯ

НЕРАВЕНСТВО ЧЕБЫШЕВА И ЕГО СЛЕДСТВИЯ

1. Одной из наиболее важных задач теории вероятностей является установление фактов, вероятности которых весьма велики (т. е. близки к единице), так что практически их можно считать достоверными. Под общим названием закона больших чисел можно было бы объединить все предложения теории вероятностей, утверждающие наступление некоторого факта A_n с вероятностью, сколь угодно близкой к достоверности, когда число n случайных событий, с которыми связан факт A_n, достаточно велико. Однако обычно в понятие закона больших чисел вносится более определенное содержание. Предположим, что рассматривается весьма большое число n величин $x_1, x_2, \ldots, x_n$, математические ожидания которых: М. О. $x_1 = a_1$, М. О. $x_2 = a_2, \ldots,$ М. О. $x_n = a_n$ известны. Мы будем говорить, что к величинам $x_1, x_2, \ldots, x_n$ применим закон больших чисел, если вероятность того, что средняя арифметическая из значений, фактически получаемых этими величинами, будет сколь угодно мало отличаться от средней арифметической из их математических ожиданий, стремясь к достоверности при безграничном увеличении числа n.

Иными словами, применимость закона больших чисел к весьма многочисленной совокупности случайных величин $x_1, x_2, \ldots, x_n$ означает, что индивидуальные отклонения каждой из них от соответствующего математического ожидания, происходящие под влиянием не подлежащих учету причин, существенным образом в массе взаимно компенсируются, так что практически можно быть уверенным в чрезвычайно точном (тем более точном, чем больше число n) совпа-

дении средней арифметической этих величин с вполне определенным значением средней арифметической из их математических ожиданий.

С наибольшею точностью нивелирующее влияние закона больших чисел обнаруживается в физических явлениях, благодаря невообразимо огромному числу молекул, находящихся даже в небольших объемах. Закономерности, являющиеся результатом закона больших чисел, здесь настолько поразительно точны, что можно было бы усомниться в методологической допустимости применения к физике понятия вероятности, если бы после постановки соответствующих опытов (движение Броуна) не удалось изолировать сравнительно небольшие количества молекул так, чтобы не дать возможности закону больших чисел совершенно уничтожить эффект индивидуальных случайных колебаний отдельной молекулы.

В явлениях общественных и биологических статистический метод во многих случаях оказывает незаменимые услуги. Между тем возможность применения к ним закона больших чисел гораздо более ограниченна, чем в физике, и самые условия применения этого закона также носят более сложный характер. Для выявления таких закономерностей необходимо поэтому прежде всего установить возможно более общие формы закона больших чисел, которые могли бы охватить весьма разнообразные совокупности величин $x_1, x_2, \ldots, x_n$, встречающиеся на практике.

Одно из простейших условий применимости закона больших чисел дает доказанная нами раньше (гл. IV, часть 2-я) теорема Бернулли, в которой рассматриваются величины $x_1, x_2, \ldots, x_n$, получающие каждая значение 1 или 0, в зависимости от того, наступает ли или нет событие A в ряде независимых n опытов, при чем вероятность события A сохраняет во всех опытах постоянное значение p; тогда М. О. $x_i = p$, и теорема Бернулли утверждает, что к величинам $x_1, x_2, \ldots, x_n$ применим закон больших чисел: действительно, в данном случае средняя арифметическая $\frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{m}{n}$, где m есть общее число появлений события A при n опытах, а средняя арифметическая из их *мат. ожид.* равна $\frac{\overbrace{p + p + \ldots + p}^{n}}{n} = p$; по теореме Бернулли, при n достаточно большом

следует ожидать с вероятностью сколь угодно близкой к достоверности, что разность между p и $\frac{m}{n}$ будет произвольно мала.

Однако область применения теоремы Бернулли на практике довольно ограниченна, так как независимость опытов и постоянство вероятностей, лежащие в основе схемы Бернулли, в действительности осуществляются очень редко.

Весьма общую форму закона больших чисел, которая охватывает как частный случай и вышеупомянутую теорему Бернулли, дал Чебышев.

Для получения результатов Чебышева нам нужно предварительно установить одно свойство математических ожиданий двух независимых между собой величин. Мы говорим, что две величины x и y независимы между собой, если вероятности любого равенства $y=c$ или неравенства $\alpha < y < \beta$ остаются неизменны, каковы бы ни были сведения о значении, которое получил x [1]).

Лемма. Если величины x и y независимы, то математическое ожидание произведения этих величин равно произведению математических ожиданий этих величин.

В самом деле, пусть М. О. $x=a$, М. О. $y=A$; допустим, кроме того, что x, равно как y, может получать только конечное число значений, при чем вероятности равенств $x=x_1$, $x=x_2, \ldots$, $x=x_n$ равны соответственно $p_1, p_2, \ldots, p_n$, и, с другой стороны, вероятности равенств $y=y_1$, $y=y_2, \ldots, y=y_N$ соответственно равны $P_1, P_2, \ldots, P_N$.

В таком случае, согласно определению математического ожидания,

$$\text{М. О. } x=a=p_1x_1+p_2x_2+\ldots+p_nx_n=\sum_{i=1}^{i=n} p_ix_i,$$

$$\text{М. О. } y=A=P_1y_1+P_2y_2+\ldots+P_Ny_N=\sum_{k=1}^{k=N} P_ky_k;$$

[1]) Можно доказать, что в таком случае, и наоборот, распределение вероятностей различных значений величины x не зависит от значения, полученного величиной y.

но, принимая во внимание независимость величин x и y, мы знаем, что вероятность совмещения равенств $x=x_1$ и $y=y_1$ равна p_1P_1 и, вообще, вероятность совмещения равенств $x=x_i$ и $y=y_k$ равна p_iP_k. Поэтому

$$\text{М. О. } xy=p_1P_1x_1y_1+p_1P_2x_1y_2+\ldots+p_nP_Nx_ny_N=$$

$$=\sum_{\substack{i=1\\k=1}}^{\substack{k=N\\i=n}} p_iP_kx_iy_k=\sum_{i=1}^{i=n} p_ix_i\sum_{k=1}^{k=N} P_ky_k=a\cdot A,$$

т.-е. *мат. ожид.* произведения xy равно произведению М. О. x на М. О. y. Применяя способ пределов, можно освободиться от ограничения, что число возможных значений x и y конечно.

Следствие. Если величины $x, y, z, u, \ldots$ независимы[1]) между собой, то математическое ожидание произведения всех этих величин равно произведению их математических ожиданий.

Для доказательства достаточно применить способ математической индукции.

2. При применении доказанной только что леммы нужно твердо помнить, что величины x и y предполагаются независимыми. Например, было бы ошибочно считать, что $b=\text{М. О. } x^2$ должно быть равно $a\cdot a=a^2$, как это имело бы место, если бы лемма была применима к произведению $x\cdot x=x^2$. В действительности же из равенств

$$\beta=\text{М. О. }[x-a]^2=\text{М. О. }[x^2-2ax+a^2]=$$
$$=b-2a^2+a^2=b-a^2 \tag{95}$$

заключаем, что $b>a^2$, так как *мат. ожид.* положительной величины $(x-a)^2$ всегда должно быть положительно; только тогда $b=a^2$, т.-е. М. О. $(x-a)^2=0$, когда имеем тождественно $x-a=0$; другими словами, равенство $b=a^2$ означало бы, что величина x вполне определена и равна в таком случае своему математическому ожиданию.

[1]) В данном случае величины $x, y, z, u, \ldots$ предполагаются совершенно независимыми, т.-е. распределение вероятностей каждой из величин остается неизменным, каковы бы ни были значения, полученные теми или иными из прочих величин. См. следующую выноску.

Разность $x-a$ (между величиной x и ее математическим ожиданием) называется отклонением величины x (от ее математического ожидания). Величина $\beta = \text{М. О.}\,(x-a)^2$, т.-е. математическое ожидание квадрата отклонения x, называется дисперсией величины x. Итак, дисперсия всякой величины x положительна и лишь тогда равна нулю, когда $x=a$ есть величина вполне определенная.

Напротив, согласно доказанной лемме, математическое ожидание произведения отклонений двух независимых величин равно 0.

Действительно,

$$\text{М. О.}\,(x-a)(y-A) = \text{М. О.}\,(x-a)\cdot\text{М. О.}\,(y-A) = 0,$$

так как $\text{М. О.}\,(x-a) = a-a = 0$.

Допустим теперь, что рассматривается n независимых[1]) (попарно) величин $z_1, z_2, \ldots, z_n$, математические ожидания которых равны 0 (так что отклонение каждой величины z_i от ее математического ожидания равно самой величине z_i); пусть $\text{М. О.}\,z_i^2 = \beta_i$. В таком случае

$$\text{М. О.}\,(z_1+z_2+\ldots+z_n)^2 =$$
$$= \text{М. О.}\,z_1^2 + \text{М. О.}\,z_2^2 + \ldots + \text{М. О.}\,z_n^2 + 2\,\text{М. О.}\,z_1z_2 + \ldots +$$
$$+ 2\,\text{М. О.}\,z_{n-1}z_n.$$

Но, так как $\text{М. О.}\,z_1z_k = \text{М. О.}\,z_i\cdot\text{М. О.}\,z_k = 0$, то мы получаем

$$\text{М. О.}\,(z_1+z_2+\ldots+z_n)^2 = \beta_1+\beta_2+\ldots+\beta_n. \qquad (96)$$

[1]) Мы подагаем, что каждая пара величин z_i и z_k ($i \gtrless k$) независимы между собой в указанном выше смысле; но для законности всех последующих рассуждений отнюдь нет необходимости требовать, чтобы все n величин были совершенно независимы: все выводы остаются в силе и тогда, когда распределение вероятностей z_i могло бы измениться после того, как станут известны значения, полученные, по крайней мере, двумя из величин z_i и z_k, так как в этой главе нам нигде не придется рассматривать математические ожидания произведения более чем двух величин. Укажем пример попарной независимости 3 величин при отсутствии полной независимости: я задумываю одно из четырех данных чисел: 112, 121, 211, 222 и предлагаю своему собеседнику угадать цифру сотен моего числа; если я назову цифру единиц или цифру десятков, то это не облегчит его задачу (останется равно возможным, что моя цифра 1 или 2); но, если я скажу цифру и единиц и десятков, то ответ будет очевиден.

Положим, вообще, что дано n попарно независимых величин $x_1, x_2, \ldots, x_n$, математические ожидания которых соответственно равны: М. О. $x_1 = a_1$, М. О. $x_2 = a_2, \ldots$, М. О. $x_n = a_n$, и пусть, кроме того, М. О. $x_1^2 = b_1$, М. О. $x_2^2 = b_2, \ldots$, М. О. $x_n^2 = b_n$.

Обозначим отклонения $x_1, x_2, \ldots, x_n$ от их математических ожиданий соответственно через $z_1, z_2, \ldots, z_n$, т.-е. положим $z_1 = x_1 - a_1$, $z_2 = x_2 - a_2, \ldots, z_n = x_n - a_n$; тогда отклонение суммы $x_1 + x_2 + \ldots + x_n$ (от ее математического ожидания) будет равно

$$z_1 + z_2 + \ldots + z_n,$$

так как

$$(x_1 + x_2 + \ldots + x_n) - (a_1 + a_2 + \ldots + a_n) =$$
$$= (x_1 - a_1) + (x_2 - a_2) + \ldots + (x_n - a_n) = z_1 + z_2 + \ldots + z_n.$$

Поэтому

$$B = \text{М. О. } [(x_1 + x_2 + \ldots + x_n) - (a_1 + a_2 + \ldots + a_n)]^2 =$$
$$= \text{М. О. } (z_1 + z_2 + \ldots + z_n)^2 = \beta_1 + \beta_2 + \ldots + \beta_n, \qquad (96\ \text{bis})$$

где, согласно (95),

$$\beta_1 = b_1 - a_1^2, \ \beta_2 = b_2 - a_2^2, \ldots, \ \beta_n = b_n - a_n^2.$$

На основании данного выше определения дисперсии, $\beta_1, \beta_2, \ldots, \beta_n$ представляют соответственно дисперсии величин $x_1, x_2, \ldots, x_n$, а B есть дисперсия суммы $(x_1 + x_2 + \cdots + x_n)$. Поэтому формула (96 bis) означает, что *дисперсия суммы любого числа попарно независимых величин равна сумме дисперсий рассматриваемых величин, и, в частности, если дисперсии последних равны, то дисперсия суммы пропорциональна числу слагаемых* n.

Необходимо помнить, что все слагаемые $x_1, x_2, \ldots, x_n$ предполагаются здесь *независимыми*. Напротив, если λ определенное число, то М. О. $\lambda x = \lambda a$, М. О. $(\lambda x)^2 = \lambda^2 b$, поэтому

$$\text{М. О. } (\lambda x - \lambda a)^2 = \lambda^2 (b - a^2) = \lambda^2 \beta,$$

так что, *при увеличении величины* x *в* λ *раз, ее дисперсия увеличится в* λ^2 *раз, а при уменьшении рассматриваемой величины в* λ *раз ее дисперсия уменьшается в* λ^2 *раз.*

В частности, в то время как дисперсия суммы

$(x_1+x_2+\ldots+x_n)$ равна $B=\beta_1+\beta_2+\ldots+\beta_n$,

дисперсия средней арифметической из величин $x_1, x_2, \ldots, x_n$, т.-е. дисперсия величины $\frac{x_1+x_2+\ldots+x_n}{n}$ равна $\frac{B}{n^2}$.

Таким образом, применяя лемму Чебышева (стр. 101), мы получаем, что вероятность P осуществления неравенства

$$[(x_1+x_2+\ldots+x_n)-(a_1+a_2+\ldots+a_n)]^2 \leqslant Bt^2 \quad (97\ \text{bis})$$

более, чем $1-\frac{1}{t^2}$; но, так как неравенство (97 bis) эквивалентно неравенству

$$|(x_1+x_2+\ldots+x_n)-(a_1+a_2+\ldots+a_n)| \leqslant t\sqrt{B}, \quad (97)$$

где

$$B=\beta_1+\beta_2+\ldots+\beta_n=b_1-a_1^2+b_2-a_2^2+\ldots+b_n-a_n^2,$$

то вероятность P неравенства (97) более, чем $1-\frac{1}{t^2}$.

Это последнее утверждение известно в науке под названием неравенства Чебышева.

Деля неравенство (97) на n, мы видим, что и неравенство

$$\left|\frac{x_1+x_2+\ldots+x_n}{n}-\frac{a_1+a_2+\ldots+a_n}{n}\right| \leqslant \frac{t}{n}\sqrt{B} \quad (98)$$

имеет ту же самую вероятность $P>1-\frac{1}{t^2}$.

Полученный результат приводит нас к теореме Чебышева, являющейся одной из наиболее общих и важных форм закона больших чисел.

3. *Теорема Чебышева.* Если $x_1, x_2, \ldots, x_n$ представляют собой какие-нибудь попарно независимые величины, при чем М. О. $x_i=a_i$ и М. О. $x_i^2=b_i$; если существует такое число L, что дисперсия каждой из рассматриваемых величин не более L, т.-е. $\beta_i=b_i-a_i^2 \leqslant L$, то к величинам $x_1, x_2, \ldots, x_n$ применим закон боль-

ших чисел, а именно, как бы мало ни было заданное число ε, вероятность R неравенства

$$\left|\frac{x_1+x_2+\ldots+x_n}{n}-\frac{a_1+a_2+\ldots+a_n}{n}\right|\leqslant\varepsilon \qquad (99)$$

становится сколь угодно близкой к достоверности, если число n достаточно велико.

Действительно, из условия $\beta_i\leqslant L$ заключаем, что

$$B=\beta_1+\beta_2+\ldots+\beta_n\leqslant nL;$$

поэтому, если мы в неравенстве (98) заменим B через nL, т.-е. напишем во второй его части $\frac{t}{n}\sqrt{nL}=t\sqrt{\frac{L}{n}}$ вместо $\frac{t}{n}\sqrt{B}$, то полученное таким образом неравенство

$$\left|\frac{x_1+x_2+\ldots+x_n}{n}-\frac{a_1+a_2+\ldots+a_n}{n}\right|\leqslant t\sqrt{\frac{L}{n}} \qquad (98\text{ bis})$$

будет, конечно, всегда иметь место (как более широкое), если только осуществится неравенство (98); следовательно, $R\geqslant P$, где R есть вероятность неравенства (98 bis), а P вероятность (98), и тем более $R>1-\frac{1}{t^2}$. Таким образом достаточно положить $t\sqrt{\frac{L}{n}}=\varepsilon$, чтобы убедиться, что вероятность неравенства

$$\left|\frac{x_1+x_2+\ldots+x_n}{n}-\frac{a_1+a_2+\ldots+a_n}{n}\right|\leqslant\varepsilon, \qquad (99)$$

равная R, более, чем $1-\frac{1}{t^2}=1-\frac{L}{n\varepsilon^2}$, т.-е. становится сколь угодно близкой к 1, если (при данном ε) взять число n достаточно большим, что и требовалось доказать.

Из теоремы Чебышева можно вывести некоторые более частные формы закона больших чисел, представляющие особый интерес, а потому нам необходимо здесь на них остановиться.

4. Предположим, что рассматривается ряд n независимых опытов; пусть p_i будет вероятность наступления некоторого факта A при i-м опыте, а m — общее

число появлений факта A при рассматриваемых n опытах (вообразим, например, ряд урн с шарами, в которых отношения числа белых шаров в каждой урне к общему числу шаров той же урны не одинаковы, но известны и равны p_i для i-й урны; каждый опыт заключается в извлечении одного шара из соответствующей урны, а под событием A будем подразумевать появление при каждом опыте белого шара). Обозначим через P среднюю арифметическую всех вероятностей p_i, которую можем назвать, для краткости, среднею вероятностью события A в рассматриваемом ряде n опытов; иными словами, положим[1]):

$$P = \frac{p_1 + p_2 + \ldots + p_n}{n}.$$

В таком случае, как бы мало ни было данное число ε, вероятность неравенства

$$\left| \frac{m}{n} - P \right| < \varepsilon \qquad (99\ \text{bis})$$

стремится к достоверности, когда число опытов n неограниченно возрастает.

Действительно, фактическое число m появлений события A можно рассматривать как сумму $m = x_1 + x_2 + \ldots + x_n$, где каждое слагаемое x_i получает значение 1, если событие A произошло при i-м опыте, и получает значение 0, если оно не произошло. Таким образом

$$a_i = \text{М. О. } x_i = p_i \cdot 1 + q_i \cdot 0 = p_i, \quad b_i = \text{М. О. } x^2 = p_i \cdot 1 + q_i \cdot 0 = p_i;$$

поэтому дисперсия β_i каждой из величин x_i равна

$$\beta_i = p_i - p_i^2 = p_i q_i < 1,$$

и, следовательно, величины x_i удовлетворяют всем условиям, требуемым теоремой Чебышева, при чем неравенство (99) обращается в неравенство (99 bis), так как

$$\frac{m}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n} \text{ и } P = \frac{a_1 + a_2 + \ldots + a_n}{n}.$$

[1]) Следует заметить, что P вообще зависит от числа n.

Доказанное только что следствие из теоремы Чебышева, являющееся обобщением теоремы Бернулли (соответствующей случаю, когда все $p_i = P$), носит название теоремы Пуассона.

Хотя найденная выше нижняя граница $1 - \frac{L}{n\varepsilon^2}$, указанная Чебышевым для вероятности R неравенства (99 bis), не имеет особого практического значения, так как в действительности R обычно еще гораздо ближе к единице, однако не бесполезно заметить, что каковы бы ни были вероятности p_i события A, в рассматриваемом случае теоремы Пуассона

$$\beta_i = p_i q_i \leqslant \frac{1}{4} = l$$

(так как, полагая $p_i = \frac{1}{2} + \alpha$, $q_i = \frac{1}{2} - \alpha$, получаем $p_i q_i = \frac{1}{4} - \alpha^2 \leqslant \frac{1}{4}$) а потому вероятность R неравенства (99 bis) всегда удовлетворяет неравенству

$$R > 1 - \frac{1}{4n\varepsilon^2}. \tag{100}$$

Пусть будет дано, например, $n = 5\,000$ урн с шарами, по 100 шаров в каждой, при чем в первых 50 урнах имеется по 1 белому шару, в следующих 50 урнах по 2 белых шара и т. д., и, наконец, в последних 50 урнах все шары белые. В таком случае, если вынимать из всех урн по одному шару, то средняя вероятность появления белого шара

$$P = \frac{50 \cdot \frac{1}{100} + 50 \cdot \frac{2}{100} + \ldots + 50 \cdot \frac{100}{100}}{5\,000} =$$

$$= \frac{1 + 2 + \ldots + 100}{10\,000} = \frac{101}{200},$$

и вероятность осуществления неравенства

$$\left| \frac{m}{5\,000} - \frac{101}{200} \right| \leqslant \frac{1}{20}, \text{ т.-е. } \frac{91}{200} \leqslant \frac{m}{5\,000} \leqslant \frac{111}{200}$$

больше, чем $1 - \frac{400}{4 \cdot 5\,000} = 1 - \frac{1}{50} = 0{,}98$.

Как мы видим, постоянство вероятности во всех опытах не является необходимым условием применимости закона больших чисел; более существенную роль играет требование независимости опытов, которое лежит в основе доказательства теоремы Чебышева, но и оно, как мы увидим дальше, может быть заменено менее ограничительными условиями.

5. Одним из важных следствий теоремы Чебышева является применение ее к случаю, когда все независимые величины x_i имеют одно и то же математическое ожидание, т.-е. когда все $a_i = a$ и, кроме того, все М. О. $x_i^2 = b_i = b$.

Предположим, для наглядности, что x_i представляет собой выигрыш (или проигрыш) игрока A, наступающий при i-й партии из общего числа n повторяемых игроком A аналогичных (независимых) игр. В таком случае из теоремы Чебышева вытекает, что если число n партий будет достаточно велико, то вероятность R неравенства

$$\left|\frac{x_1 + x_2 + \ldots + x_n}{n} - a\right| \leqslant \varepsilon \qquad (101)$$

(при всяком данном ε) становится сколь угодно близкой к достоверности: $R > 1 - \frac{b - a^2}{n\varepsilon^2}$.

Таким образом, обозначая через $x = \frac{x_1 + x_2 + \ldots + x_n}{n}$ средний выигрыш игрока A в течение указанных n партий, видим, что с увеличением n этот средний выигрыш x стремится к математическому ожиданию выигрыша; поэтому (предполагая, что нет технических препятствий к повторению игры произвольно большое число n раз) игрок A, для которого отдельная партия выгодна ($a > 0$), может быть уверен в выигрыше сколь угодно большой суммы, если будет сыграно достаточно большое число партий; напротив, если $a < 0$ (невыгодная игра), то при повторении игры достаточное число раз практически несомненно, что игрок A будет разорен.

Действительно, вся сумма выигрыша игрока A равна $S = nx = x_1 + x_2 + \ldots + x_n$; поэтому, умножая неравенство (101) на n, заменяем его равнозначным неравенством

$$-\varepsilon n \leqslant S - an \leqslant \varepsilon n,$$

т.-е.

$$(a - \varepsilon)n \leqslant S \leqslant (a + \varepsilon)n.$$

Полагая, в частности, $\varepsilon = \frac{a}{2}$, видим таким образом, что вероятность R неравенства

$$\frac{an}{2} \leqslant S \leqslant \frac{3an}{2} \tag{102}$$

больше, чем $1 - \frac{4}{n}\left(\frac{b-a^2}{a^2}\right) > 1 - \frac{4b}{na^2}$. Следовательно, если $a > 0$, осуществление неравенства (102) соответствует, при n достаточно большом, выигрышу сколь угодно большой наперед назначенной суммы денег, при чем вероятность этого становится сколь угодно близкой к единице.

Если, например, при бросании игральной кости игрок A выигрывает 8 рублей при появлении 6 очков и уплачивает 1 рубль при появлении меньшего числа очков, то игра для него выгодна, так как математическое ожидание его выигрыша в отдельной игре

$$a = \text{М. О. } x = 8 \cdot \frac{1}{6} - 1 \cdot \frac{5}{6} = 0{,}5.$$

При этом $b = \text{М. О. } x^2 = 64 \cdot \frac{1}{6} + 1 \cdot \frac{5}{6} = \frac{69}{6} = 11{,}5$; следовательно, применяя (102), находим, что в данном случае вероятность того, что игрок выиграет после n партий сумму денег S, заключенную между $\frac{n}{4}$ и $\frac{3n}{4}$ рублями, больше, чем $1 - \frac{180}{n}$, так как $\frac{b-a^2}{a^2} = 45$. Например, если число сыгранных партий $n = 1\,000$, то вероятность, что наш игрок в результате всей игры приобретет не менее 250 рублей, больше, чем 0,82 (игрок A имеет больше оснований рассчитывать на получение указанной суммы, чем если бы выигрыш этой суммы был связан с появлением белого шара из урны, содержащей 100 шаров, из которых 82 белых.)

Таким образом, несомненно, наиболее важной характеристикой игры[1]), с точки зрения коммерческой целесообразности ее многократного повторения, является величина a математического ожидания

[1]) Под игрой мы можем подразумевать здесь любую финансовую операцию, вероятности различных исходов которой имеют более или менее точное математическое значение; указанным свойством обладают различные виды страхования, основанного на соответствующих статистических данных.

связанного с ней выигрыша и, прежде всего, знак этой величины: полученный вывод в достаточной мере оправдывает термины выгодной и невыгодной игры в зависимости от того, будет ли $a > 0$ или $a < 0$. При этом, однако, необходимо твердо помнить, что практическое значение понятия математически выгодной и невыгодной игры обусловлено предположением о возможности многократного ее повторения, необходимого для того, чтобы, в силу закона больших чисел, выравнивающего случайности отдельных партий, средний выигрыш партии мало отличался от его математического ожидания.

Если же по той или иной причине игра не может быть повторена (или же может быть повторена лишь небольшое число раз), то математическая выгодность ее или невыгодность имеет мало практического значения. Миллионер, который для увеличения своего капитала охотно согласится играть в орлянку, если при появлении „орла" он выигрывает 200 рублей, а в противном случае уплачивает 100 руб. ($a = 200 \cdot \frac{1}{2} - 100 \cdot \frac{1}{2} = 50$ руб.), вряд ли будет настолько легкомыслен, что станет рисковать потерей всего своего капитала в 1 000 000 руб. в случае непоявления „орла", в расчете на то, что при появлении „орла" он выиграет 1 000 100 руб., хотя последние условия игры также соответствуют математически выгодной игре, где $a = 50$ руб. Если же какие-нибудь особые причины побуждали бы все-таки нашего миллионера согласиться на подобные условия, ввиду исключительного значения, придаваемого им немедленному удвоению своего капитала, то, конечно, он не задумался бы предпринять игру и в том случае, если бы вместо 1 000 100 руб. он выигрывал бы 999 900 руб., т.-е. если бы игра была невыгодна ($a = -50$ руб.).

6. При доказательстве закона больших чисел в общей форме, данной ему Чебышевым, кроме независимости величин $x_1, x_2, \ldots, x_n$, существенную роль играет предположение о том, что дисперсия этих величин ограничена (нужно, чтобы $\beta_i = b_i - a_i^2 < L$). Мы укажем пример, из которого видно будет, что если бы это последнее условие могло быть отброшено, то во всяком случае его нужно было бы заменить каким-нибудь аналогичным ограничением, так как одного условия независимости величин недостаточно для применимости к ним закона больших чисел. Действительно, положим, что

x_1 может получать значения 1 и -1, x_2 получает значения 2 или -2 и т. д., вообще, x_k может получать значения k и $-k$, при чем положительное и отрицательное значение каждого x_k одинаково вероятно, так что М. О. $x_k = 0$ и М. О. $x_k^2 = k^2$ (условие ограниченности дисперсии не соблюдено). Следовательно, если бы закон больших чисел в данном случае был применим, вероятность неравенства

$$\left|\frac{x_1 + x_2 + \ldots + x_n}{n}\right| < \varepsilon \qquad (103)$$

стремилась бы к 1 с возрастанием n; но этого не может быть, потому что все возможные значения суммы $S_n = x_1 + x_2 + \ldots + x_n$ распадаются на две группы $S_n = S_{n-1} + n$ и $S_n' = S_{n-1} - n$, которые равновероятны; поэтому, если $\left|\frac{S_n}{n}\right| < \varepsilon$, то не может быть $\left|\frac{S_n'}{n}\right| = \left|\frac{S_n}{n} - 2\right| < \varepsilon$; отсюда заключаем, что вероятность неравенства (103) (при $\varepsilon < 1$) не может быть больше $\frac{1}{2}$.

Однако покойный академик А. А. Марков показал, что для применимости к независимым величинам $x_1, x_2, \ldots, x_n$ закона больших чисел достаточно, чтобы существовало число $\delta > 0$, такое, что М. О. $|x_i|^{1+\delta} < L$, где L — не зависящая от i величина (условие Чебышева соответствует частному случаю $\delta = 1$). Доказательство этого положения читатель найдет в курсе А. А. Маркова.

Мы ограничимся здесь лишь рассмотрением примера, уясняющего, каким образом закон больших чисел может оставаться в силе, несмотря на бесконечное возрастание дисперсии величин $x_1, x_2, \ldots, x_n$. Представим себе, что каждая из величин x_i может получать любые целые положительные и отрицательные значения (не 0), и пусть вероятность для всякого x_i получить целое значение $\pm h$ равна

$$P_h = P_{-h} = \frac{2}{h(h+1)(h+2)},$$

так что

$$\sum_{h=-\infty}^{h=\infty} P_h = 4\sum_{1}^{\infty} \frac{1}{h(h+1)(h+2)} =$$

$$= 2\sum_{1}^{\infty}\left[\frac{1}{h(h+1)} - \frac{1}{(h+1)(h+2)}\right] = 1, \text{ и М. О. } x_i = 0, \text{ но М. О. } x_i^2 = \infty,$$

так как М. О. $x^2 = \sum_1^\infty \frac{4h}{(h+1)(h+2)} > \sum_1^\infty \frac{1}{h+1}$, а последний ряд, как известно, расходящийся. Покажем тем не менее, что (хотя дисперсия каждой из величин x бесконечно велика) вероятность неравенства

$$\left|\frac{x_1 + x_2 + \ldots + x_n}{n}\right| < \varepsilon \tag{103}$$

при любом ε стремится к достоверности, когда n неограниченно возрастает.

Для этого, полагая $l \leqslant n$, будем отбрасывать, как непригодные, все наблюденные значения x_i, которые не меньше n по численному значению; таким образом будем требовать не только соблюдения неравенства (103), но, кроме того, чтобы все $|x_i| < n$.

Согласно теореме умножения вероятностей находим, что, после того как последнее условие соблюдено, вероятность P'_h, что $x_i = h < n$, равна

$$P'_h = \frac{P_h}{\sum\limits_{-n<i<n} P_i} = \frac{\frac{1}{h(h+1)(h+2)}}{2\sum\limits_{i=1}^{i=n-1} \frac{1}{i(i+1)(i+2)}} =$$

$$= \frac{\frac{1}{h(h+1)(h+2)}}{\sum\limits_{i=1}^{i=n-1}\left(\frac{1}{i(i+1)} - \frac{1}{(i+1)(i+2)}\right)} = \frac{2}{h(h+1)(h+2)} \cdot \frac{1}{1 - \frac{2}{n(n+1)}}. \tag{104}$$

Поэтому, если вместо первоначальных величин x_i, которые могут быть безгранично велики, мы рассматриваем независимые величины x'_i, для которых вероятности получить значения $\pm h$ даны формулой (104), то

$$\text{М. О. } x'_i = 0 \text{ и М. О. } x'^2_i =$$

$$= \frac{4}{1 - \frac{2}{n(n+1)}} \sum_{h=1}^{h=n-1} \frac{h}{(h+1)(h+2)} < \frac{4 \log n}{1 - \frac{2}{n(n+1)}}$$

[так как $\frac{h}{(h+1)(h+2)} < \log(h+1) - \log h$]. Следовательно, вероятность неравенства

$$\left|\frac{x'_1 + x'_2 + \ldots + x'_n}{n}\right| < \varepsilon$$

больше

$$1 - \frac{4 \log n}{\left(n - \frac{2}{n+1}\right)\varepsilon^2}.$$

Но вероятность, что $x_i < h < n$ равна $\sum\limits_{-n<i<n} P_i = \left[1 - \frac{2}{n(n+1)}\right]$; поэтому вероятность, что все $x_i < n$ $(i = 1, 2, \ldots, n)$, равна $\left[1 - \frac{2}{n(n+1)}\right]^n$.

Таким образом вероятность, что все $x_i < n$ и что, кроме того, осуществляется неравенство

$$\left|\frac{x_1+x_2+\ldots+x_n}{n}\right| < \varepsilon, \tag{103}$$

больше, чем

$$\left[1-\frac{2}{n(n+1)}\right]^n\left(1-\frac{4\log n}{(n-1)\varepsilon^2}\right) > \left(1-\frac{2}{n+1}\right)\left(1-\frac{4\log n}{(n-1)\varepsilon^2}\right),$$

так как, применяя бином Ньютона и группируя попарно члены, видим, что

$$\left[1-\frac{2}{n(n+1)}\right]^n = \left[1-\frac{2}{n+1}\right]+$$
$$+\left[\frac{(n-1)}{2n}\cdot\left(\frac{2}{n+1}\right)^2-\frac{n-1}{2n}\cdot\frac{n-2}{3n}\cdot\left(\frac{2}{n+1}\right)^3\right]+\ldots$$

представляет сумму положительных слагаемых (каждая разность в прямоугольных скобках положительна).

Тем более осуществление одного только неравенства (103) (которое возможно было бы и без соблюдения условия, что $x_i < n$) имеет вероятность

$$P > \left(1-\frac{2}{n+1}\right)\left(1-\frac{4\log n}{(n-1)\varepsilon^2}\right) > 1-\frac{4\log n}{(n-1)\varepsilon^2}-\frac{2}{n+1},$$

которая (при данном ε) стремится к 1, когда n неограниченно возрастает, так как при этом $\frac{4\log n}{(n-1)\varepsilon^2}$ стремится к 0. Как видим, то обстоятельство, что М. О. x_i^2 бесконечно, не мешает в данном случае применимости к величинам $x_1, x_2, \ldots, x_n$ закона больших чисел (предоставляем читателю проверить, что в нашем примере указанное выше условие А. А. Маркова соблюдено для всякого $\delta < 1$).

Упражнения. 1) Производится n независимых опытов, при которых вероятности появления события A соответственно равны: $p_1 = \frac{3}{4}$, $p_2 = \left(\frac{3}{4}\right)^2, \ldots, p_n = \left(\frac{3}{4}\right)^n$; требуется установить, применяя неравенство Чебышева, нижнюю границу вероятности P, что $m < 6$.

Ответ: $P > \frac{17}{21} \neq 0{,}81$. Действительно,

$$p_1+p_2+\cdots+p_n = 3\left[1-\left(\frac{3}{4}\right)^n\right], \quad p_1q_1+\cdots+p_nq_n =$$
$$= \frac{12}{7}-3\left(\frac{3}{4}\right)^n+\frac{9}{7}\left(\frac{9}{16}\right)^n < \frac{12}{7};$$

поэтому неравенство

$$\left|m-3\left[1-\left(\frac{3}{4}\right)^n\right]\right| < t\sqrt{\frac{12}{7}}$$

имеет вероятность бо́льшую, чем $1-\frac{1}{t^2}$; при $t\sqrt{\frac{12}{7}} = 3$ наше неравенство

обращается в неравенство

$$-3\left(\frac{3}{4}\right)^n < m < 6 - 3\left(\frac{3}{4}\right)^n \qquad 1)$$

которое (при любом n) является частным случаем неравенства $m < 6$; поэтому тем более $P > 1 - \frac{1}{t^2} = 1 - \frac{12}{63} = \frac{17}{21}$.

2) При бросании двух игральных костей игрок A выигрывает число рублей, равное разности между числом очков, выпавших на одной и другой кости, если последние числа неравны. В случае же одинакового числа очков на обеих костях игрок A уплачивает число рублей, равное сумме очков на обеих костях. Какой выигрыш S с вероятностью большею, чем 0,98, можно гарантировать игроку A при повторении партии $n = 3\,000$ раз?

Ответ: $S \geqslant 815$ рублей.

Действительно

$$a = \frac{5}{18} + \frac{4}{9} + \frac{3}{6} + \frac{4}{9} + \frac{5}{18} = \frac{12}{36} + \frac{10}{36} + \frac{8}{36} + \frac{6}{36} + \frac{4}{36} + \frac{2}{36} = \frac{7}{9}$$

$$b = \frac{287}{18};$$

поэтому вероятность, что $\left| S - 3\,000 \cdot \frac{7}{9} \right| \leqslant 3\,000\,\varepsilon$, больше, чем $1 - \frac{b - a^2}{n\varepsilon^2} = 1 - \frac{2\,485}{162 \cdot 3\,000\,\varepsilon^2}$; полагая $\frac{2\,485}{162 \cdot 3\,000\,\varepsilon^2} = 0{,}02$, получаем $\varepsilon = 0{,}5055$; откуда $S \geqslant 815$.

ГЛАВА ВТОРАЯ

УТОЧНЕНИЕ НЕРАВЕНСТВА ЧЕБЫШЕВА

1. Доказанная нами выше теорема Чебышева особенно важна в теоретическом отношении, как устанавливающая чрезвычайно общую форму закона больших чисел. Однако практическое непосредственное применение этой теоремы требует, чтобы число n величин $x_1, x_2, \ldots, x_n$ было чрезвычайно велико, гораздо больше, чем это необходимо для того, чтобы закон больших чисел фактически вступил в силу, так как нижняя граница вероятности соответствующего неравенства, данная на стр. 148, значительно меньше истинного значения этой вероятности, которая в действительности гораздо быстрее приближается к 1 с возрастанием n. Поэтому, подчиняя некоторым ограничениям, на практике обычно осуществляемым, рассматриваемые независимые величины $x_1, x_2, \ldots, x_n$, полезно указать значение, представляющее во многих случаях более точную нижнюю границу рассматриваемой вероятности, которая должна, согласно закону больших чисел, стремиться к 1 при неограниченном увеличении n.

Мы предположим, что рассматриваемые нами независимые величины $x_1, x_2, \ldots, x_n$ имеют математические ожидания соответственно равные $a_1, a_2, \ldots, a_n$ и, обозначая вообще отклонения $x_i - a_i = z_i$ и М. О. $z_i^2 = \beta_i$, допустим, что соблюдено условие

$$\left| \text{М. О. } z_i^k \right| \leqslant \frac{\beta_i}{2} H^{k-2} k! \tag{105}$$

при всех значениях $k \geqslant 2$, где H — некоторое положительное число.

Теорема. При соблюдении условия (105) вероятности каждого из неравенств

$$(x_1 + x_2 + \ldots + x_n) - (a_1 + a_2 + \ldots + a_n) \geqslant 2t\sqrt{B}$$

и

$$(x_1 + x_2 + \ldots + x_n) - (a_1 + a_2 + \ldots + a_n) \leqslant -2t\sqrt{B}, \tag{106}$$

где $B = \beta_1 + \beta_2 + \ldots + \beta_n$, меньше, чем e^{-t^2}, а веро-

ятность P неравенства

$$|(x_1+x_2+\ldots+x_n)-(a_1+a_2+\ldots+a_n)|<2t\sqrt{B} \quad (107)$$

больше, чем $1-2e^{-t^2}$, если только $0<t\leqslant\frac{\sqrt{B}}{2H}$.

Для доказательства заметим, что второе из неравенств (106) равнозначно неравенству

$$(-x_1-x_2-\ldots-x_n)-(-a_1-a_2-\ldots-a_n)\geqslant 2t\sqrt{B},$$

в которое обращается первое из неравенств (106), если его применять к величинам $x_i'=-x_i$. Поэтому достаточно ограничиться рассмотрением лишь первого из этих неравенств.

Полагаем

$$I=\text{M. O. } e^{\varepsilon(z_1+z_2+\ldots+z_n)},$$

где ε — некоторое положительное число, подчиненное условию, что

$$\varepsilon H\leqslant\frac{1}{2}. \quad (108)$$

В таком случае, применяя лемму Чебышева к неотрицательной величине $e^{\varepsilon(z_1+z_2+\ldots+z_n)}$, заключаем, что вероятность неравенства

$$\varepsilon(z_1+z_2+\ldots+z_n)>t^2+\log I,$$

равнозначного неравенству

$$e^{\varepsilon(z_1+z_2+\ldots+z_n)}>e^{t^2}I,$$

должна быть меньше, чем e^{-t^2}.

Но, на основании свойства математических ожиданий произведений независимых величин[1]),

$$I=\text{M. O. } e^{\varepsilon z_1}\cdot e^{\varepsilon z_2}\ldots e^{\varepsilon z_n}=$$

$$=\text{M. O. } e^{\varepsilon z_1}\cdot\text{M. O. } e^{\varepsilon z_2}\ldots\text{M. O. } e^{\varepsilon z_n}. \quad (109)$$

---

[1]) Здесь величины $z_1, z_2, \ldots, z_n$ предполагаются совершенно независимыми, т.-е. каковы бы ни были известные уже значения нескольких из них, это не влияет на вероятности получения прочими из рассматриваемых величин тех или иных значений.

Полагая

$$R_i = \text{М. О. } e^{\varepsilon z_i} = \text{М. О. } \left[1 + \varepsilon z_i + \frac{\varepsilon^2 z_i^2}{2} + \frac{\varepsilon^3 z_i^3}{3!} + \ldots \right]$$

и применяя теорему сложения математических ожиданий, находим

$$R_i = 1 + \text{М. О. } \varepsilon z_i + \text{М. О. } \frac{\varepsilon^2 z_i^2}{2!} + \text{М. О. } \frac{\varepsilon^3 z_i^3}{3!} + \ldots =$$

$$= 1 + \frac{\varepsilon^2 \beta_i}{2!} + \frac{\varepsilon^3}{3!} \text{М. О. } z_i^3 + \ldots + \frac{\varepsilon^k}{k!} \text{М. О. } z_i^k + \ldots,$$

так как М. О. $z_i = 0$. Поэтому, вследствие условия (105),

$$R_i \leqslant 1 + \frac{\varepsilon^2 \beta_i}{2} \left[1 + \varepsilon H + \ldots + (\varepsilon H)^{k-2} + \ldots \right],$$

а следовательно, благодаря (108),

$$R_i \leqslant 1 + \varepsilon^2 \beta_i < e^{\varepsilon^2 \beta_i}, \tag{110}$$

так как ряд $e^{\varepsilon^2 \beta_i} = 1 + \varepsilon^2 \beta_i + \frac{\varepsilon^4 \beta_i^2}{2} + \cdots$ содержит только положительные слагаемые. Из (109) и (110) заключаем, что

$$I < e^{\varepsilon^2 \beta_1} \cdot e^{\varepsilon^2 \beta_2} \ldots e^{\varepsilon^2 \beta_n} = e^{\varepsilon^2 (\beta_1 + \beta_2 + \ldots + \beta_n)} = e^{\varepsilon^2 B},$$

а потому

$$\log I < \varepsilon^2 B,$$

и, следовательно, неравенство

$$\varepsilon(z_1 + z_2 + \ldots + z_n) \geqslant t^2 + \varepsilon^2 B \tag{112}$$

еще менее вероятно, чем неравенство

$$\varepsilon(z_1 + z_2 + \ldots + z_n) > t^2 + \log I, \tag{111}$$

которое, как было замечено выше, имеет вероятность меньшую, чем e^{-t^2}. Тем более, значит, и вероятность неравенства (112), при любом положительном ε, удовлетворяющем (108), должна быть менее, чем e^{-t^2}.

Полагая, в частности, $\varepsilon = \frac{t}{\sqrt{B}}$ [что мы в праве сделать, не нарушая (108), лишь бы только $t \leqslant \frac{\sqrt{B}}{2H}$] и деля обе части нера-

венства (112) на ε, получаем равнозначное ему неравенство

$$z_1 + z_2 + \ldots + z_n \geqslant t\sqrt{B} + t\sqrt{B},$$

которое представляет собой не что иное, как

$$(x_1 + x_2 + \ldots + x_n) - (a_1 + a_2 + \ldots + a_n) \geqslant 2t\sqrt{B}; \quad (106)$$

отсюда заключаем, наконец, что вероятность неравенства (106) действительно менее, чем e^{-t^2}.

Ввиду сделанного в начале доказательства замечания, видим, что и неравенство

$$(x_1 + x_2 + \ldots + x_n) - (a_1 + a_2 + \ldots + a_n) \leqslant -2t\sqrt{B} \quad (106 \text{ bis})$$

имеет вероятность меньшую, чем e^{-t^2}. Следовательно, вероятность Q осуществления одного из двух несовместимых неравенств (106) менее, чем $2e^{-t^2}$, а потому вероятность $P = 1 - Q$ ненаступления ни одного из них, т.-е. осуществления неравенства

$$|x_1 + x_2 + \ldots + x_n - (a_1 + a_2 + \ldots + a_n)| < 2t\sqrt{B}, \quad (107)$$

больше, чем $1 - 2e^{-t^2}$.

2. Остановимся на применении полученного результата к случаю, когда $x_1, x_2, \ldots, x_n$ представляют собой частные значения некоторой величины x, получаемой при повторении n независимых однородных опытов, так что, вообще,

$$\text{М. О. } x_i = a_i = a \text{ и М. О. } (x_i - a_i)^2 = \beta_i = \beta, \text{ а } B = n\beta.$$

Полагая $X = \dfrac{x_1 + x_2 + \ldots + x_n}{n}$ и деля на n неравенство (107), находим, что вероятность P неравенства

$$|X - a| < 2t\sqrt{\frac{\beta}{n}} \quad (101 \text{ bis})$$

больше, чем $1 - 2e^{-t^2}$ $\left(\text{ограничиваясь}^{1)} \text{ значениями } t \leqslant \dfrac{\sqrt{n\beta}}{2H}\right)$.

---

[1]) Это ограничение практически не существенно, если только $\dfrac{H}{\beta}$ не слишком велико, так как оно сводится к тому, что правая часть неравенства не должна превышать $\dfrac{\beta}{H}$; но, когда это значение достигается, то вероятность

Разумеется, для применимости полученного результата необходимо знать, что математические ожидания всех степеней величины x удовлетворяют неравенству вида

$$|\text{М. О. } (x-a)^k| \leqslant \frac{\beta}{2} H^{k-2} \cdot k! \qquad (105 \text{ bis})$$

Но в большинстве задач, которые встречаются на практике, величина x вообще ограничена; поэтому обычно можно без всяких вычислений указать более или менее грубо такое число λ, относительно которого нет сомнения, что $\lambda > |x-a|$. Очевидно, что тогда

$$|\text{М. О. } (x-a)^k| < \text{М. О. } (x-a)^2 \lambda^{k-2} = \beta\lambda^{k-2}.$$

Поэтому, если возьмем $H = \frac{\lambda}{3}$, то неравенство (105 bis) будет конечно удовлетворено, так как, при всяком $k > 2$, имеем

$$\beta\lambda^{k-2} \leqslant \frac{\beta}{2}\left(\frac{\lambda}{3}\right)^{k-2} k!,$$

в чем легко убедиться, умножая обе части на $\frac{1}{\beta}\left(\frac{3}{\lambda}\right)^{k-2}$.

В частности, в серии опытов Бернулли или Пуассона можно взять $H = \frac{1}{3}$.

Рассмотрим для примера 2-е упражнение предыдущей главы: при бросании двух игральных костей игрок A проигрывает сумму рублей, равную сумме очков на обеих костях, если на них оказывается одинаковое число очков; напротив, он выигрывает сумму рублей, равную разности между числами очков, если числа эти различны. В этом случае

$$a = \text{М. О. } x = \frac{7}{9}, \quad b = \text{М. О. } x^2 = \frac{287}{18},$$

$$\beta = b - a^2 = \frac{2\,485}{162} \text{ и } |x-a| \leqslant 12 + \frac{7}{9} = \frac{115}{9} = \lambda,$$

так как наибольшее абсолютное значение разности между x и a соответствует максимальному проигрышу $x = -12$.

соответствующего неравенства уже превышает $1 - 2e^{-\frac{n\beta}{2H}}$, т.-е. при n достаточно большом уже чрезвычайно близка к 1: например, если $\frac{H}{\beta} = 5$, то, при $n = 40$, $1 - 2e^{-\frac{\beta}{2H}} > 0{,}963$, а при $n = 250$, $1 - 2e^{-23} \neq$ $\neq 0{,}999\,999\,999\,97$ (с точностью до последнего знака).

Таким образом, при $n = 3\,000$, вероятность неравенства

$$\left| x - \frac{7}{9} \right| < 2t \sqrt{\frac{2\,485}{162 \cdot 3\,000}}$$

больше, чем $1 - 2e^{-t^2}$ $\left(\text{лишь бы } t \leqslant \frac{27}{230} \sqrt{3\,000 \cdot \frac{2\,485}{162}} \neq 25{,}1\right)$.

Полагая, например, $t = 3$, можем с вероятностью большею, чем 0,999 749, утверждать, что

$$\left| x - \frac{7}{9} \right| < 6 \sqrt{\frac{2\,485}{162 \cdot 3\,000}},$$

т.-е. что весь выигрыш $S = nx$ будет заключен в пределах

$$2\,333 - \sqrt{\frac{4\,970\,000}{3}} < S < 2\,333 + \sqrt{\frac{4\,970\,000}{3}},$$

т.-е.

$$1\,045 < S < 3\,621.$$

Вероятность же только того, что выигрыш $S > 1\,045$ рублей, больше, чем $1 - e^{-9} = 0{,}999\,874$. Между тем, применяя теорему Чебышева, мы для того же неравенства получили бы нижнюю границу $1 - \frac{1}{4 \cdot 9} = 0{,}972\,22$.

Вообще, для тех же неравенств, для которых теорема Чебышева дает нижнюю границу $1 - \frac{1}{4t^2}$, выведенная только что формула дает $1 - 2e^{-t^2}$; преимущество последней начинает обнаруживаться лишь при $t > 2$, приобретая все большее значение с увеличением t. Для сравнения и для практического применения укажем таблицу:

$t =$	1,5	2	2,5	3	3,5	4	5
$2e^{-t^2} =$	0,210 8	0,036 93	0,003 86	0,000 251	0,000 009 6	0,000 000 23	0,000 000 000 03
$\frac{1}{4t^2} =$	0,111 1	0,062 5	0,04	0,027 778	0,020 408	0,015 625	0,01

3. Данную формулу можно было бы еще уточнить [1]), но, не останавливаясь на общем случае, где задача несколько усложняется вследствие несим-

[1]) На практике обычно пользуются предельной формулой Лапласа (см. четвертую часть) и ее обобщениями, которая фактически для значений n, превышающих несколько сотен, обладает большой точностью и дает часто лучшее приближение для вероятности, чем указанные нами формулы, однако строгое доказательство этого довольно затруднительно. См. мою статью „Об одном видоизменении неравенства Чебышева". Ученые записки Н. И. Кафедр Украины Отдел математический. Вып. 1.

метричного распределения вероятностей рассматриваемых величин x_i, мы ограничимся предположением, что М. О. $(x_i - a_i)^{2k+1} = 0$ для любого целого значения k; это соответствует допущению, что равные между собой по абсолютной величине положительные и отрицательные отклонения равновероятны [ибо при этом допущении М. О. $(x_i - a_i)^{2k+1}$ представляет собой алгебраическую сумму членов, попарно равных по величине и противоположных по знаку и потому взаимно уничтожающихся].

Заменяя условие (105) несколько более ограничительным условием [1])

$$\text{М. О. } (x_i - a_i)^{2k} \leqslant \left(\frac{\beta_i}{2}\right)^k \cdot \frac{(2k)!}{k!} \tag{113}$$

и применяя рассуждение аналогичное предшествующему мы найдем, что в данном случае вероятность неравенства

$$|x_1 + x_2 + \ldots + x_n - (a_1 + a_2 + \ldots + a_n)| \leqslant t\sqrt{2B} \tag{114}$$

больше [2]), чем $1 - 2e^{-t^2}$.

В самом деле, составляя, подобно предыдущему, выражение $R_i = \text{М. О. } e^{\varepsilon z_i}$, где $z_i = x_i - a_i$, находим:

$$R_i = 1 + \frac{\varepsilon^2}{2}\beta_i + \frac{\varepsilon^3}{3!}\,\text{М. О. } z_i^3 + \ldots + \frac{\varepsilon^h}{h!}\,\text{М. О. } z_i^h + \ldots,$$

и, так как, по условию, при h нечетном М. О. $z_i^h = 0$, а при $h = 2k$ соблюдается (113), то

$$R_i \leqslant 1 + \frac{\varepsilon^2\beta_i}{2} + \ldots + \frac{\left(\varepsilon^2\frac{\beta_i}{2}\right)^k}{k!} + \ldots = e^{\frac{\varepsilon^2\beta_i}{2}}.$$

следовательно, в силу (109),

$$I \leqslant e^{\frac{\varepsilon^2 B}{2}} \quad \text{и} \quad \log I \leqslant \frac{\varepsilon^2 B}{2}.$$

Отсюда заключаем, что, при всяком ε, неравенство

$$\varepsilon\,[(x_1 + x_2 + \ldots + x_n) - (a_1 + a_2 + \ldots + a_n)] > t^2 + \frac{\varepsilon^2 B}{2}$$

---

[1]) Знак равенства в (113) соответствовал бы так называемому нормальному распределению вероятностей величин x_i.

[2]) На основании полученного выше результата мы могли бы лишь утверждать, что эта вероятность больше, чем $1 - 2e^{-\frac{t^2}{2}}$. Заметим, кроме того, что в данном случае значения t могут быть произвольно велики.

имеет вероятность меньшую, чем e^{-t^2}. Поэтому, полагая последовательно $\varepsilon = t\sqrt{\frac{2}{B}}$ и $\varepsilon = -t\sqrt{\frac{2}{B}}$, где t — положительное число, находим соответственно, что каждое из неравенств

$$(x_1 + x_2 + \ldots + x_n) - (a_1 + a_2 + \ldots + a_n) > t\sqrt{2B}$$

и

$$(x_1 + x_2 + \ldots + x_n) - (a_1 + a_2 + \ldots + a_n) < -t\sqrt{2B}$$

имеет вероятность [1]) меньшую, чем e^{-t^2}. Отсюда, как и раньше, вытекает, что неравенство (114) имеет вероятность большую, чем $1 - 2e^{-t^2}$.

4. Предположим, например, что каждая из величин x_h может получать только два значения $+h^\delta$ и $-h^\delta$, вероятности которых равны $\frac{1}{2}$, где δ — некоторое определенное положительное число.

В таком случае М. О. $x_h^{2k+1} = 0$ и М. О. $x_h^{2k} = h^{2\delta k}$; по этому условию (113) удовлетворяют все x_h, так как неравенство

$$h^{2\delta k} \leqslant \frac{h^{2\delta k}}{2^k} \cdot \frac{(2k)!}{k!}$$

равнозначно

$$1 \leqslant 1 \cdot 3 \cdots (2k-1).$$

Посмотрим, при каких значениях δ закон больших чисел будет приложим, т.-е. можно будет при данном произвольно малом ε утверждать, что вероятность неравенства

$$\left|\frac{x_1 + x_2 + \ldots + x_n}{n}\right| < \varepsilon$$

стремится к достоверности с возрастанием n. В данном случае

$$B = 1 + 2^{2\delta} + \ldots + h^{2\delta} + \ldots + n^{2\delta} =$$
$$= n^{2\delta}\left[\left(\frac{1}{n}\right)^{2\delta} + \left(\frac{2}{n}\right)^{2\delta} + \ldots + \left(\frac{n}{n}\right)^{2\delta}\right].$$

Поэтому, замечая, что, согласно определению интеграла,

$$\int_{\frac{1}{n}}^{\frac{n+1}{n}} x^{2\delta}\,dx > \frac{1}{n}\left[\left(\frac{1}{n}\right)^{2\delta} + \left(\frac{2}{n}\right)^{2\delta} + \cdots + \left(\frac{n}{n}\right)^{2\delta}\right] > \int_0^1 x^{2\delta}\,dx,$$

[1]) Заметим, что если бы мы знали, что М. О. $(x_i - a_i)^{2k+1} \leqslant 0$, то наше утверждение оставалось бы верным по отношению к первому из этих неравенств (при обратном знаке, оно было бы верно для второго).

заключаем, что

$$\frac{(n+1)^{1+2\delta}}{1+2\delta} > B > \frac{n^{1+2\delta}}{1+2\delta}.$$

Следовательно, неравенство

$$(x_1 + x_2 + \ldots + x_n) < t\sqrt{\frac{2(n+1)^{1+2\delta}}{1+2\delta}}, \tag{115}$$

которое еще более вероятно, чем неравенство (114), имеет вероятность бо́льшую, чем $1 - 2e^{-t^2}$, а потому и неравенство

$$\left|\frac{x_1 + x_2 + \ldots + x_n}{n}\right| < \varepsilon = t\sqrt{\frac{2\left(1+\frac{1}{n}\right)^{1+2\delta}}{1+2\delta}} \cdot n^{\delta - \frac{1}{2}}$$

имеет вероятность бо́льшую, чем $1 - 2e^{-t^2}$.

Если $\delta < \frac{1}{2}$, то, при данном произвольно большом t (обеспечивающем вероятность, произвольно близкую к достоверности), мы можем, достаточно увеличивши число n, сделать ε сколь угодно близким к 0, так как $\delta - \frac{1}{2} < 0$; следовательно, *при* $\delta < \frac{1}{2}$ *закон больших чисел применим, и, беря среднюю арифметическую из наблюдений, которых уклонения от 0 все более и более возрастают, мы все же можем быть практически уверены, в этом случае, что получим среднюю, сколь угодно мало отличающуюся от 0.* Если, например, извлекая корни 4-й степени из всех целых чисел до миллиона, мы будем брать перед ними знак + или — в зависимости от того, выпадает ли или нет „орел“ при соответствующем бросании монеты, то вероятность P того, что средняя арифметическая из всех этих корней по абсолютному значению меньше, чем $\varepsilon = \frac{1}{10}$, больше [1]), чем $1 - 2e^{-t^2}$, где $t^2 = 7{,}5$, т.-е. $P > 0{,}998\,89$.

Если $\delta = \frac{1}{2}$, то $B = \frac{n(n+1)}{2}$; таким образом мы можем лишь утверждать, что вероятность неравенства

$$\left|\frac{\pm 1 \pm \sqrt{2} \pm \sqrt{3} \ldots \pm \sqrt{n}}{n}\right| < t\sqrt{1+\frac{1}{n}} = \varepsilon \tag{116}$$

больше, чем $1 - 2e^{-t^2}$. Следовательно, мы не можем утверждать, что с воз-

---

[1]) Заметим, что применимость закона больших чисел к данному случаю могла бы быть выведена также из основного неравенства Чебышева (но не из его теоремы, где предполагается, что $\beta_i < L$); однако мы получили бы только, что $P > 1 - \frac{1}{2t^2} = 0{,}933\,33$.

растанием n при малых значениях ε вероятность неравенства (116) становится велика; однако мы не в праве непосредственно заключить из этого, что в данном случае закон больших чисел неприменим. Но, как будет сейчас показано, закон больших чисел при $\delta \geqslant \frac{1}{2}$ действительно, неприменим. Достаточно рассмотреть для определенности случай $\delta = \frac{1}{2}$, так как очевидно, что при $\delta > \frac{1}{2}$ применим тот же способ рассуждений. Итак, положим $\delta = \frac{1}{2}$ и допустим, что, как бы мало ни было η, при неограниченном возрастании n вероятность неравенства (116), где ε весьма малая величина, равна $1 - \eta$.

Разобьем совокупность всех значений, принимаемых абсолютным значением величины $X = \frac{x_1 + x_2 + \cdots + x_n}{n}$, на смежные промежутки

$$(0, \varepsilon),\ (\varepsilon, \gamma_0),\ (\gamma_0, \gamma_1), \ldots, (\gamma_k, \gamma_{k+1}),$$

где γ_0 определим условием, что $2e^{-\frac{n\gamma_0^2}{n+1}} = \eta < 2e^{-\frac{n}{n+1}\varepsilon^2}$, прочие же промежутки будут произвольным образом стремиться к 0, а $\gamma_{k+1} = \sqrt{n}$. При бесконечном возрастании n

$$\text{пред. М. О. } X^2 = \text{пред. } \frac{B}{n^2} = \frac{1}{2}.$$

С другой стороны,

$$\text{М. О. } X^2 = B_1 < \varepsilon^2(1 - \eta) + P_0\gamma_1^2 + P_1\gamma_2^2 + \ldots + P_k\gamma_{k+1}^2,$$

где P_0 — вероятность, что $|X|$ находится в промежутке (ε, γ_1), P_1 есть вероятность того, что $|X|$ находится в промежутке (γ_1, γ_2), и вообще, P_k есть вероятность, что $|X|$ находится в промежутке (γ_k, γ_{k+1}), при чем $P_0 + P_1 + \ldots + P_k = \eta$. Очевидно, что мы вторую часть неравенства еще увеличим, если P_k заменим бо́льшим числом $2e^{-\frac{\gamma_k^2 n}{n+1}}$; после этого наибольшее значение, какое можем придать P_{k-1}, будет

$$2e^{-\frac{\gamma_{k-1}^2 n}{n+1}} - 2e^{-\frac{\gamma_k^2 n}{n+1}}$$

и т. д.

Таким образом

$$B_1 < \varepsilon^2(1 - \eta) + 2\left(e^{-\frac{\gamma_0^2 n}{n+1}} - e^{-\frac{\gamma_1^2 n}{n+1}}\right)\gamma_1^2 + \ldots$$

$$\ldots + 2\left(e^{-\frac{\gamma_{k-1}^2 n}{n+1}} - e^{-\frac{\gamma_k^2 n}{n+1}}\right)\gamma_k^2 + 2e^{-\frac{\gamma_k^2 n}{n+1}}\gamma_{k+1}^2.$$

Переходя к пределу $n=\infty$, т.-е ставя $\frac{1}{2}$ вместо B_1 и пред. $e^{-\frac{\gamma^2 n}{n+1}} = e^{-\gamma^2}$ вместо $e^{-\frac{\gamma^2 n}{n+1}}$ и делая простую перегруппировку слагаемых, получаем:

$$\frac{1}{2} \leqslant \varepsilon^2(1-\eta) + \eta\gamma_1^2 + 2\left(\gamma_2^2 - \gamma_1^2\right)e^{-\gamma_1^2} + \ldots + 2\left(\gamma_{k+1}^2 - \gamma_k^2\right)e^{-\gamma_k^2},$$

или, полагая, что промежутки (γ_1, γ_0), (γ_2, γ_1), ... стремятся к 0,

$$\frac{1}{2} \leqslant \varepsilon^2(1-\eta) + \eta\gamma_0^2 + 4\int_{\gamma_0}^{\infty} \gamma e^{-\gamma^2}\, d\gamma = \varepsilon^2(1-\eta) + \eta\gamma_0^2 + 2e^{-\gamma_0^2}.$$

Отсюда, с помощью равенства $\eta = 2e^{-\gamma_0^2}$, выводим

$$\frac{\frac{1}{2} - \eta\left(1 + \log\frac{2}{\eta}\right)}{1-\eta} \leqslant \varepsilon^2.$$

Таким образом, если хотим, чтобы η стремилось к 0, то необходимо, чтобы $\varepsilon^2 \geqslant \frac{1}{2}$. Следовательно, как бы велико ни было n, нельзя сделать сколь угодно близкой к достоверности вероятность неравенства (116), если $\varepsilon < \frac{1}{\sqrt{2}}$. См. упражнения гл III, ч. 4-я.

Упражнения. 1) Производится 600 бросаний игральной кости; требуется определить с вероятностью большею, чем 0,999 749, в каких пределах будет заключено число появлений 6 очков.

Ответ. $46 \leqslant m \leqslant 154$. Из таблички значений $2e^{-t^2}$ находим, что $2e^{-t^2} = 1 - 0{,}999\,749 = 0{,}000\,251$ соответствует $t=3$; поэтому с требуемой вероятностью можем утверждать неравенство

$$-6\sqrt{\frac{5\cdot 600}{36}} < m - 100 < 6\sqrt{\frac{5\cdot 600}{36}}, \text{ т.-е. } 46 \leqslant m \leqslant 154.$$

2) Полагая, что появление 6 очков связано с выигрышем 8 рублей, а непоявление — с проигрышем 1 рубля, установить нижний предел вероятности P, что при n бросаниях общий выигрыш превзойдет $\frac{n}{4}$ рублей.

Ответ. $P > 1 - e^{-\frac{n}{720}}$ $\left(\text{а по теореме Чебышева } P > 1 - \frac{180}{n}\right)$; например, при 6 480 партиях вероятность выиграть более, чем 1 620 рублей, больше чем 0,999 875.

3) Дано, что распределение вероятностей величин x на отрезке $(-l, +l)$ равномерно. Требуется определить, какое число n наблюдений достаточно для того, чтобы с вероятностью не меньшею, чем 0,999 999,

можно было утверждать, что средняя арифметическая из этих наблюдени окажется в промежутке $\left(-\frac{l}{10}, +\frac{l}{10}\right)$.

Ответ. $n \geqslant 967$. Для применения указанного выше приема приравниваем $0{,}000\,001 = 2e^{-t^2}$, откуда $t^2 = \frac{6+\log 2}{\log e} \neq \frac{6{,}301\,0}{0{,}434\,3} \neq 14{,}5$. В данном случае М. О. $x^{2k+1} = 0$, а М. О. $x^{2k} = \frac{1}{2l}\int_{-l}^{l} x^{2k}\,dx = \frac{l^{2k}}{2k+1}$, и, в частности, $\beta =$ М. О. $x^2 = \frac{l^2}{3}$; поэтому неравенство (113), получающее вид

$$\frac{l^{2k}}{2k+1} \leqslant \left(\frac{l^2}{6}\right)^k \cdot \frac{(2k)!}{k!},$$

соблюдено, так что применимо (114), и для определения n имеем неравенство $t\sqrt{\frac{2}{3n}} \leqslant \frac{1}{10}$, т.-е. $n \geqslant \frac{200t^2}{3} \neq 967$. Применяя неравенство Чебышева, мы получили бы $t^2 = 500\,000$, поэтому $n \geqslant 33\,333\,333$, — разумеется, в таком огромном числе наблюдений нет никакой надобности.

4) При условиях игры, данных во 2-м упражнении главы I, требуется указать верхнюю границу вероятности Q, что игрок A, располагая капиталом в M рублей, когда-нибудь разорится при повторении указанной выгодной игры. Рассмотреть случай, когда $M = 500$ руб.

Ответ. $Q < \dfrac{e^{-\frac{223M}{8\,520}}}{1 - e^{-\frac{7}{710}}}$; при $M = 500$, $Q < 0{,}000\,139$.

Пусть a будет *мат. ожид.* выигрыша x одной партии и $\beta =$ М. О. $(x-a)^2$; в таком случае, обозначая через S всю сумму выигрыша, вероятность, что $S - na \leqslant -2t\sqrt{n\beta}$, менее, чем e^{-t^2}, при условии, что $t \leqslant \frac{\sqrt{n\beta}}{2H}$; если же $t > \frac{\sqrt{n\beta}}{2H}$, то, возвращаясь к неравенству (112), замечаем, что неравенство

$$S - na \leqslant -\frac{t^2}{\varepsilon} - \varepsilon n\beta$$

имеет при всяком t и n вероятность меньшую, чем e^{-t^2}, если только $\varepsilon H \leqslant \frac{1}{2}$. Поэтому, полагая $\frac{1}{2\varepsilon} = H_1 \geqslant H$, заключаем, что вероятность неравенства

$$S - na \leqslant -2H_1 t^2 - \frac{n\beta}{2H_1}$$

всегда меньше, чем e^{-t^2}.

При наличности капитала M для разорения нужно, чтобы $S - na \leqslant -M - na$; поэтому, если из равенства

$$M + na = 2t\sqrt{n\beta}$$

находим значение $t=\frac{M+na}{2\sqrt{n\beta}}\leqslant\frac{\sqrt{n\beta}}{2H}$, то вероятность, что после n-й партии у игрока не будет положительной суммы денег, оказывается меньше [1]), чем

$$e^{-t^2}=e^{-\frac{(M+na)^2}{4n\beta}}.$$

Если бы расчет совершался только после данного числа n партий, то найденное значение было бы верхнею границей вероятности, что наш игрок разорится или окажется банкротом, т.-е. не сможет оплатить свой долг после указанных n партий.

Однако, если капитал не очень мал, то, по крайней мере, для небольших значений n нужное для получения вывода условие, что

$$M+na\leqslant\frac{n\beta}{H},$$

не будет соблюдено; если бы $a>\frac{\beta}{H}$, то это условие, вообще, никогда бы не выполнялось. Поэтому, вообще, мы должны приравнять

$$M+na=2H_1t^2+\frac{n\beta}{2H_1},$$

и, следовательно, вероятность, что после n-й партии $S+M\leqslant 0$, всегда менее, чем e^{-t^2}, где $t^2=\frac{n\left(a-\frac{\beta}{2H_1}\right)+M}{2H_1}$, лишь бы $H_1\geqslant H$. Полагая сначала $H\leqslant\frac{\beta}{a}$, можем взять $H_1=\frac{\beta}{a}$; в таком случае $t^2=\frac{\frac{na}{2}+M}{\frac{2\beta}{a}}=$ $=\frac{na^2+2Ma}{4\beta}$.

---

[1]) Если бы были соблюдены условия, соответствующие неравенству (113), тогда, при всяких значениях M, n, a, β, рассматриваемая вероятность была бы меньше, чем $e^{-\frac{(M+na)^2}{2n\beta}}$, а потому вероятность Q, что разорение вообще когда-нибудь наступит, во всяком случае меньше, чем

$$\sum_{n=1}^{\infty}e^{-\frac{(M+na)^2}{2n\beta}}<\sum_{1}^{\infty}e^{-\frac{2Ma+na^2}{2\beta}}=\frac{e^{-\frac{Ma}{\beta}}}{1-e^{-\frac{a^2}{2\beta}}}.$$

сли известно, что максимальный проигрыш не превышает $\frac{M}{k}$, то сумму ожно брать, начиная только от $n=k$, поэтому

$$Q<e^{-\frac{ka^2}{2\beta}}\frac{e^{-\frac{Ma}{\beta}}}{1-e^{\frac{a^2}{2\beta}}}.$$

Если же $H > \frac{\beta}{a}$, то можем взять $H_1 = H$, и тогда

$$t^2 = \frac{n\left(a - \frac{\beta}{2H}\right) + M}{2H} > \frac{\frac{na}{2} + M}{2H}.$$

Таким образом в первом случае, который наиболее обычен, вероятность, что после n партий $S + M \leqslant 0$, меньше, чем $e^{-\frac{na^2+2Ma}{4\beta}}$; во втором случае эта вероятность меньше, чем $e^{-\frac{na+2M}{4H}}$

Поэтому в первом случае вероятность Q, что неравенство $S + M \leqslant 0$ (т.-е. разорение) вообще когда-нибудь наступит, меньше, чем

$$\sum_{n=k}^{n=\infty} e^{-\frac{na^2+2Ma}{4\beta}} = \frac{e^{-\frac{ka^2+2Ma}{4\beta}}}{1 - e^{-\frac{a^2}{4\beta}}}, \tag{117}$$

где k есть наименьшее число партий, при котором разорение возможно [т.-е. $(k-1)\mu < M \leqslant k\mu$, обозначая через μ максимальный проигрыш, возможный в одной партии]; точно так же во втором случае

$$Q < \frac{e^{-\frac{ka^2+2M}{4H}}}{1 - e^{-\frac{a}{4H}}}. \tag{117 bis}$$

Рассматриваемый пример подходит под первый случай (117), так как

$$a = \frac{7}{9},\ \beta = \frac{2\,485}{162} \text{ и } H = \frac{\lambda}{3} = \frac{115}{27} < \frac{\beta}{a} = \frac{355}{18};$$

поэтому

$$Q < \frac{e^{-\frac{18M+7k}{710}}}{1 - e^{-\frac{7}{710}}} = \frac{e^{-\frac{223}{8\,520}M}}{1 - e^{-\frac{7}{710}}} < \frac{710}{7} \cdot \frac{1\,420}{1\,413} e^{-\frac{223M}{8\,520}} < 102 e^{-\frac{223M}{8\,520}}$$

ибо максимальный проигрыш партии $\mu = 12$, так что $k \geqslant \frac{M}{12}$.

Полагая $M = 500$, находим

$$Q < 102 e^{-\frac{11\,150}{852}} \neq 0{,}000\,139.$$

Полагая $M = 1\,000$, получили бы

$$Q < 102 e^{-\frac{22\,300}{852}} = 0{,}000\,000\,000\,2.$$

Таким образом, даже при $M = 500$, игрок A почти не рискует оказаться лишенным возможности продолжать игру за отсутствием средств, и, на

основании найденного раньше, закон больших чисел обеспечивает ему произвольно большой барыш.

Из формул (117) и (117 bis) видно, что **риск** Q **разорения при возрастании капитала** M **в арифметической прогрессии уменьшается быстрее, чем члены геометрической прогрессии со знаменателем** $e^{-\frac{a}{2\beta}}$ **или** $e^{-\frac{1}{2H}}$. (В действительности, точный закон убывания риска сложнее и лишь в редких случаях выражается более или менее простой формулой для значений M, не превышающих во много раз ставок отдельной партии.)

Если игрок, обладающий капиталом M, может в единицу времени по своему выбору предпринять одну игру или 2 аналогичные независимые игры, но все ставки которых вдвое меньше, то для него всегда благоразумнее предпочесть второе. Действительно, если разорение не помешает использовать для игры n единиц времени (т.-е. сыграть n крупных или $2n$ вдвое менее крупных партий), то в обоих случаях, по закону больших чисел, его средний выигрыш в единицу времени будет мало отличаться от *мат. ожид.* выигрыша a одной крупной партии. Но дисперсия $\beta = \text{М. О. } (x-a)^2$ одной крупной партии вдвое больше дисперсии $\beta' = \text{М. О. } (x'+y'-a)^2 = \text{М. О. } \left(x'-\frac{a}{2}+y'-\frac{a}{2}\right)^2 = \frac{\beta}{4}+\frac{\beta}{4} = \frac{\beta}{2}$ двух маленьких партий; поэтому риск его разорения при малой игре будет, приблизительно таков же, как если бы основной капитал M был вдвое больше.

Применяя вышесказанное к торговым и финансовым организациям, мы можем формулировать, как общее правило, допускающее лишь незначительные уклонения, что солидность коммерческого предприятия определяется отношением среднего размера его отдельных (независимых) операций к величине резервного капитала: чем крупнее отдельная операция, тем должен быть больше резервный капитал, независимо от суммы годового оборота.

Вероятность Q разорения игрока A, имеющего в начале игры капитал M, как было замечено, может быть в некоторых случаях вычислена точно; простой пример этого был дан на стр. 133. Обобщим немного рассмотренные там данные. Допустим, что отдельная игра допускает несколько выигрышных ставок, скажем, для определенности, две: a_1 рублей и a_2 рублей, вероятности которых в каждой партии соответственно равны p_1 и p_2, а проигрыш может быть равен только 1 рублю. Пусть $f(M) = Q$ есть вероятность разорения игрока A при неограниченном повторении игры, когда он обладает капиталом M, и, вообще, вероятность его разорения равна $f(x)$, если его капитал равен x. В таком случае, очевидно,

$$f(x+1) = f(x) \cdot f(1),$$

так как потеря $x+1$ руб. всегда происходит после потери x рублей, за которой должна еще наступить утрата 1 рубля. Следовательно,

$$f(x) = [f(1)]^x. \tag{118}$$

Таким образом риск разорения при наличности капитала x равен x-му члену геометрической прогрессии, первый член которой, также как и знаменатель, равен $f(1)$.

С другой стороны,

$$f(x) = p_1 f(x + a_1) + p_2 f(x + a_2) + (1 - p_1 - p_2) f(x - 1),$$

так как разорение может наступить (предполагая, что у игрока в начале игры было x рублей), если он при первой партии выиграет a_1 или a_2 рублей или же проиграет 1 рубль; но в каждом из указанных случаев вероятность разорения становится соответственно равной $f(x + a_1)$, $f(x + a_2)$ и $f(x - 1)$, принимая во внимание изменение капитала нашего игрока после первой партии. Следовательно, подставляя значение $f(x)$ из формулы (118), находим, что $f(1) = z$ удовлетворяет уравнению

$$z^x = p_1 z^{x+a_1} + p_2 z^{x+a_2} + (1 - p_1 - p_2) z^{x-1},$$

которое по разделении на z^{x-1} получает вид

$$z = p_1 z^{1+a_1} + p_2 z^{1+a_2} + (1 - p_1 - p_2). \qquad (119)$$

Полагая игру выгодной, мы знаем, что $f(x) < 1$, а потому и $f(1) = z < 1$; ввиду этого решение z уравнения (119), которое может соответствовать значению $f(1)$, должно обязательно быть меньше 1. Поэтому, придав уравнению (119) форму

$$z - 1 = p_1 (z^{1+a_1} - 1) + p_2 (z^{1+a_2} - 1),$$

из которой видно, что оно имеет также решение $z = 1$, можем разделить обе части равенства на $z - 1$, откуда получаем уравнение

$$1 = p_1 (z^{a_1} + z^{a_1 - 1} + \ldots + 1) + p_2 (z^{a_2} + \ldots + 1), \qquad (119\ \text{bis})$$

которое всегда имеет одно и только одно решение z, заключенное между 0 и 1, так как вторая часть равенства с увеличением z от 0 до 1 возрастает от $p_1 + p_2$ до $p_1(a_1 + 1) + p_2(a_2 + 1)$, при чем $p_1 + p_2 < 1$, а $p_1(a_1 + 1) + p_2(a_2 + 1) > 1$, если математическое ожидание выигрыша отдельной игры положительно. Следовательно, $f(1)$ равно единственному положительному корню уравнения (119 bis).

Пусть, например, $p_1 = \frac{1}{5}$, $a_1 = 3$, $p_2 = \frac{5}{14}$, $a_2 = 2$. В этом случае математическое ожидание выигрыша равно $\frac{3}{5} + \frac{5}{7} - \frac{31}{70} = \frac{61}{70}$ рубля; уравнение (119 bis) приобретает вид

$$1 = \frac{1}{5}(z^3 + z^2 + z + 1) + \frac{5}{14}(z^2 + z + 1),$$

которое имеет решение $z = \frac{1}{2}$, так что $f(x) = \left(\frac{1}{2}\right)^x$ и, при $M = 10$, например, риск разорения равен $\frac{1}{1\,024}$. Обычно z гораздо ближе к 1, и риск

разорения больше. Ограничимся случаем, когда $p_2 = 0$; тогда легко проверить, что $z = f(1) > 1 - p_1 = q_1$. Действительно, подставляя это значение в уравнение, видим, что вторая его часть получит значение $p_1 \frac{1 - q_1^{a_1+1}}{1 - q_1} = 1 - q_1^{a_1+1} < 1$; так как мы знаем, что уравнение имеет лишь одно положительное решение, то нетрудно вообще его приближенно решить. Укажем, что, если М. О. $x = a_1 p_1 - q_1 \geqslant 1$, то $z < 1 - \frac{p_1}{2}$; действительно, подстановка последнего значения в наше уравнение дает во второй части $2\left[1 - \left(1 - \frac{p_1}{2}\right)^{a+1}\right]$, и нужно лишь проверить, что эта величина > 1, т.-е. $1 > 2\left(1 - \frac{p}{2}\right)^{a+1}$, или, логарифмируя, что $\log \frac{1}{2} = -\left[\frac{1}{2} + \frac{1}{2}\left(\frac{1}{2}\right)^2 + \ldots\right] > -(a+1)\left[\frac{p_1}{2} + \frac{1}{2}\left(\frac{p_1}{2}\right)^2 + \ldots\right]$; но это неравенство соблюдено, так как, по предположению, $(a+1)p_1 \geqslant 2$, вследствие чего оно приводится к неравенству

$$1 + \frac{1}{2}\left(\frac{p_1}{2}\right) + \frac{1}{3}\left(\frac{p_1}{2}\right)^2 + \cdots > \frac{1}{2} + \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^2 + \frac{1}{3}\left(\frac{1}{2}\right)^3 + \ldots,$$

которое имеет место даже при $p_1 = 0$.

Пусть, например, $a_1 = 10\,000$, $p_1 = \frac{1}{5\,000}$; тогда М. О. $x = \frac{10\,000}{5\,000} - \frac{4\,999}{5\,000} = 1{,}000\,2$, поэтому $0{,}999\,8 < f(1) < 0{,}999\,9$. Следовательно, $(1 - 0{,}000\,2)^x < f(x) < (1 - 0{,}000\,1)^x$; игрок, у которого было бы всего 100 рублей, имел бы вероятность разориться (не получив возможности использовать преимущества многократного повторения этой выгодной игры) бо́льшую, чем $\left(1 - \frac{2}{10\,000}\right)^{100} \neq e^{-\frac{1}{50}} \neq 0{,}98$. Напротив, капиталист, у которого капитал $M = 200\,000$ рублям, наверное обогатится, повторяя эту игру с риском разорения меньшим

$$\left(1 - \frac{1}{10\,000}\right)^{200\,000} \neq e^{-20} \neq 0{,}000\,000\,002.$$

ГЛАВА ТРЕТЬЯ

РАСПРОСТРАНЕНИЕ ЗАКОНА БОЛЬШИХ ЧИСЕЛ НА ЗАВИСИМЫЕ ВЕЛИЧИНЫ

1. Предположим теперь, что величины $x_1, x_2, \ldots, x_n$ могут так или иначе зависеть друг от друга, и пусть М. О. $x_i = a_i$ и, попрежнему, М. О. $x_i^2 = b_i$. То, что отличает рассматриваемый нами теперь общий случай от того, которым мы занимались до сих пор, это — **неприменимость**, вообще говоря, теоремы умножения математических ожиданий, установленной лишь для независимых величин, благодаря которой мы имели право утверждать, что М. О. $(x_i - a_i)(x_k - a_k) = 0$.

Таким образом в общем случае формула для дисперсии B суммы $x_1 + x_2 + \cdots + x_n$ уже не будет иметь вида

$$B = \beta_1 + \beta_2 + \cdots + \beta_n, \qquad (96\ \text{bis})$$

где $\beta_i = b_i - a_i^2$, но принимая во внимание, что

$$\text{М. О. } (x_i - a_i)(x_k - a_k) = \beta_{ik}$$

вообще не равно 0, мы будем иметь

$$B = \text{М. О. } [(x_1 - a_1) + (x_2 - a_2) + \ldots + (x_n - a_n)]^2 =$$
$$= \beta_1 + \beta_2 + \ldots + \beta_n + 2\beta_{12} + 2\beta_{13} + \ldots + 2\beta_{n-1,n}, \qquad (120)$$

где члены β_{ik} соответствуют всем парам различных значков i, k, которые мы можем выбрать среди n чисел: $1, 2, \ldots, n$, и, следовательно, число этих членов равно $C_n^2 = \frac{n(n-1)}{2}$.

Однако, если мы будем пользоваться выражением (120) для B вместо выражения (96 bis), то, применяя ту же лемму Чебышева, которой мы пользовались для вывода неравенства Чебышева, мы получим точно также, что вероятность P неравенства

$$|x_1 + x_2 + \ldots + x_n - (a_1 + a_2 + \ldots + a_n)| \leqslant t\sqrt{B} \qquad (97\ \text{ter}$$

больше, чем $1 - \frac{1}{t^2}$.

Деля затем неравенство (97 ter) на n, мы получим равнозначное неравенство

$$\left|\frac{x_1+x_2+\ldots+x_n}{n}-\frac{a_1+a_2+\ldots+a_n}{n}\right|\leqslant t\sqrt{\frac{B}{n^2}}, \quad (98\ \text{bis})$$

откуда заключаем, что

закон больших чисел применим к зависимым величинам $x_1, x_2, \ldots, x_n$, если, при возрастании n, $\frac{B}{n^2}$ стремится к 0 (теорема А. А. Маркова).

Действительно, пусть η и ε будут два наперед заданные произвольно малые числа. Положим $\frac{1}{t^2}=\eta$, т.-е. $t=\frac{1}{\sqrt{\eta}}$, и возьмем затем n настолько большим, чтобы $\frac{B}{n^2}<\varepsilon^2\eta$, что несомненно осуществимо, ибо, по предположению, $\frac{B}{n^2}$ стремится к 0 с возрастанием n. Но в таком случае, заменяя в неравенстве (98 bis) t через $\frac{1}{\sqrt{\eta}}$, а $\frac{B}{n^2}$ через $\varepsilon^2\eta$, мы получим более широкое неравенство

$$\left|\frac{x_1+x_2+\ldots+x_n}{n}-\frac{a_1+a_2+\ldots+a_n}{n}\right|<\varepsilon, \quad (121)$$

так как $t\sqrt{\frac{B}{n^2}}<\frac{1}{\sqrt{\eta}}\sqrt{\varepsilon^2\eta}=\varepsilon$, а потому тем более вероятность неравенства (121) должна превысить $1-\frac{1}{t^2}=1-\eta$.

Таким образом, как бы мало ни было ε, вероятность P неравенства (121) при возрастании n становится сколь угодно близкой к достоверности.

Этой теоремой А. А. Маркова исчерпываются все важнейшие случаи приложимости закона больших чисел, хотя, как мы уже раньше видели (стр. 155), теоретически возможно указать примеры, когда закон больших чисел остается в силе, несмотря на то, что все величины $\beta_i=b_i-a_i^2$, а следовательно и $\frac{B}{n^2}$, при любом n бесконечно велики (т.-е., собственно говоря, лишены смысла). Отбра-

сывая эти исключительные случаи и вводя допущение, что средняя арифметическая

$$X = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

ни при каких частных значениях x_i не может превзойти по абсолютной величине некоторого сколь угодно большого, но определенного предела L, нетрудно убедиться, что теорема А. А. Маркова становится обратимой, т.-е. (при сделанном только что ограничении) закон больших чисел не может быть применим к величинам $x_1, x_2, \ldots, x_n$, если $\frac{B}{n^2}$ не стремится к 0.

В самом деле, предположим обратное: пусть, действительно, неравенство

$$|X - A| = \left|\frac{x_1 + x_2 + \ldots + x_n}{n} - \frac{a_1 + a_2 + \ldots + a_n}{n}\right| < \varepsilon \qquad (121)$$

при произвольно малом ε получает вероятность $1 - \eta$ сколь угодно близкую к 1, когда n весьма велико.

В таком случае

$$\frac{B}{n^2} = \text{М. О. } [X - A]^2 < (1 - \eta)\varepsilon^2 + \eta 4L^2; \qquad (122)$$

действительно, все значения $X - A$ разбиваются на 2 класса: в первом классе значения $|X - A| < \varepsilon$, ко второму классу относятся остальные значения, которые, во всяком случае, меньше по абсолютной величине, чем $2L$ (ибо из условия, что $|X| < L$ следует, что $|A| < L$); поэтому, если в сумме, составляющей М. О. $[X - A]^2$, заменим $X - A$ через ε для всех слагаемых, соответствующих первому классу, и через $2L$ для слагаемых второго класса, то эта сумма будет увеличена и окажется как раз равной выражению, стоящему во второй части неравенства (122), потому что, по предположению, вероятности, что $X - A$ принадлежит к первому и ко второму классу, соответственно равны $1 - \eta$ и η.

Но если бы ε и η могли быть выбраны произвольно малыми, то и сумма

$$(1 - \eta)\varepsilon^2 + 4\eta L^2$$

также должна была бы сделаться сколь угодно малой, а следовательно и $\frac{B}{n^2}$, которое меньше указанной суммы, стремилось бы к 0.

Поэтому, раз по условию $\frac{B}{n^2}$ к нулю не стремится, то сделанное предположение должно быть отвергнуто.

2. Возьмем, например, серию $n+1$ независимых опытов, при которых появление события A имеет постоянную вероятность p, и будем каждый раз рассматривать результаты двух смежных опытов: i-го и $(i+1)$-го, при чем будем давать числу x_i значение 2, если в обоих опытах событие A произошло, значение 0, если A произошло лишь в одном опыте, и значение -2, если событие не наступило ни в одном из рассматриваемых опытов.

В таком случае $a_i = \text{М. О. } x_i = 2p^2 - 2q^2 = 2(p-q)(p+q) = 2(p-q)$, и $\text{М. О. } (x_1 + x_2 + \ldots + x_n) = 2n(p-q)$.

Очевидно, что смежные значения x_i и x_{i+1} не будут независимы, так как, если $x_i = 2$, это значит, что факт A наступил в i-м и в $(i+1)$-м опыте, а поэтому x_{i+1} может уже только получить значение 2 или 0, при чем первое из этих значений имеет вероятность p, соответствующую вероятности появления A в $(i+2)$-м опыте, а вероятность второго равна q; следовательно,

$$\text{М. О. } x_i x_{i+1} = 2 \cdot 2p^3 + (-2)(-2)q^3 = 4(p^3 + q^3) = 4(p^2 - pq + q^2)(p+q) = 4(p^2 - pq + q^2),$$

так как $x_i x_{i+1}$ отлично от 0 лишь в том случае, когда все 3 последовательных опыта имеют одинаковые исходы. Напротив, x_i и x_k независимы, если $|i-k| > 1$, и в этом случае

$$\text{М. О. } x_i x_k = \text{М. О. } x_i \cdot \text{М. О. } x_k = 4(p-q)^2;$$

наконец, $\text{М. О. } x_i^2 = 4p^2 + 4q^2$. Следовательно,

$$\beta_i = \text{М. О. } (x_i - a_i)^2 = 4p^2 + 4q^2 - 4(p-q)^2 = 8pq,$$

$$\beta_{ik} = \text{М. О. } (x_i - a_i)(x_k - a_k) = 0 \qquad (\text{если } |i-k| > 1),$$

$$\beta_{ik} = \text{М. О. } (x_i - a_i)(x_k - a_k) = \text{М. О. } (x_i x_k - a_i x_k - a_k x_i + a_i a_k) = \text{М. О. } x_i x_k - a_i a_k = 4[p^2 - pq + q^2 - (p-q)^2] = 4pq$$

$$(\text{если } i - k = 1).$$

Поэтому

$$B = \sum \beta_i + 2\sum \beta_{ik} = 8npq + 2[\beta_{12} + \beta_{23} + \ldots + \beta_{n-1,n}] = 8npq + 8(n-1)pq = (16n - 8)pq,$$

так что $\frac{B}{n^2}$ стремится к 0 с возрастанием n. Следовательно, при не-

ограниченном возрастании n, вероятность неравенства

$$\left|\frac{x_1+x_2+\ldots+x_n}{n}-2(p-q)\right|<\varepsilon$$

будет стремиться к достоверности, как бы мало ни было ε.

Та же серия опытов может быть использована для получения ряда зависимых величин, которые все связаны между собой и к которым закон больших чисел неприменим, хотя дисперсии β_i каждой из них ограничены.

Обозначим через $m_1, m_2, \ldots, m_n$ последовательно наблюдаемые числа появлений события A при $1, 2, \ldots, n$ опытах и будем рассматривать числа

$$x_i=\frac{m_i-ip}{\sqrt{ipq}}.$$

Очевидно, М. О. $x_i=0$. Можем ли мы в данном случае утверждать, что неравенство

$$\left|\frac{x_1+x_2+\ldots+x_n}{n}\right|<\varepsilon$$

(при всяком $\varepsilon>0$) обладает вероятностью, стремящеюся к достоверности, при увеличении n?

Находим

$$\beta_i=\text{М. О. } x_i^2=1.$$

$$\beta_{ik}=\frac{1}{pq\sqrt{ik}}\text{ М. О. }(m_i-ip)(m_k-kp)=$$

$$=\frac{1}{pq\sqrt{ik}}\text{ М. О. }(m_i-ip)[m_i-ip+m_k-m_i+(i-k)p]=$$

$$=\frac{1}{pq\sqrt{ik}}\cdot\text{М.О. }(m_i-ip)^2=\sqrt{\frac{i}{k}},$$

так как, полагая $k>i$, замечаем, что ввиду независимости результатов последующих $(k-i)$ опытов от результатов первых i опытов М. О. $(m_i-ip)[m_k-m_i+(i-k)p]=0$.

Следовательно,

$$B=n+2\left[\frac{1}{\sqrt{2}}+\frac{1}{\sqrt{3}}+\ldots+\frac{1}{\sqrt{n}}\right]+2\left[\sqrt{\frac{2}{3}}+\sqrt{\frac{2}{4}}+\ldots\right.$$

$$\left.\ldots+\sqrt{\frac{2}{n}}\right]+\ldots+2\sqrt{\frac{n-1}{n}}=n+2\left[\frac{1}{\sqrt{2}}+\frac{1+\sqrt{2}}{\sqrt{3}}+\right.$$

$$\left.+\frac{1+\sqrt{2}+\sqrt{3}}{\sqrt{4}}+\ldots+\frac{1+\sqrt{2}+\ldots+\sqrt{n-1}}{\sqrt{n}}\right]>$$

$$>n+2\left[\frac{1}{2}+\frac{2}{2}+\frac{3}{2}+\ldots+\frac{n-1}{2}\right]=n+\frac{n(n-1)}{2}=\frac{n(n+1)}{2}.$$

так как вообще

$$\sqrt{a}+\sqrt{n-a}>\sqrt{n}, \text{ а потому } 1+\sqrt{2}+\ldots+\sqrt{n-1}>\frac{n-1}{2}\sqrt{n}.$$

Из того обстоятельства, что $B>\frac{n(n+1)}{2}$, заключаем, что $\frac{B}{n^2}$ к 0 не стремится. Мы отсюда не в праве непосредственно заключить, что в данном случае закон больших чисел неприменим, так как величины $x_i=\frac{m_i-ip}{\sqrt{ipq}}$ могут неограниченно возрастать. Для доказательства этого нужно было бы сделать дополнительное рассуждение, аналогичное тому, которое было применено в конце главы II; не останавливаясь на этом, заметим лишь, что указанное заключение является следствием следующего общего положения: если для достаточно больших значений L вероятность неравенства

$$\left|\frac{x_1+x_2+\ldots+x_n}{n}-\frac{a_1+a_2+\ldots+a_n}{n}\right|>L$$

меньше, чем $\frac{1}{L^{2+\alpha}}$, где $\alpha>0$, то закон больших чисел неприменим к величинам $x_1, x_2, \ldots, x_n$, когда $\frac{B}{n^2}$ не стремится к 0 с возрастанием n.

3. Не следует, однако, думать, что закон больших чисел будет обязательно нарушен, если связь существует между всеми величинами $x_1, x_2, \ldots, x_n$. Необходимо только, чтобы эта связь была не слишком велика, чтобы она, например, достаточно быстро убывала при удалении величин x_i и x_k друг от друга, т.-е. когда $|i-k|$ достаточно велико. Мы можем придать высказанному замечанию более определенную форму, введя несколько новых понятий, которые нам понадобятся и впоследствии.

Пусть x_1 и x_2 будут две какие-нибудь величины, имеющие М. О. $x_1=a_1$, М. О. $x_2=a_2$. Мы уже раньше назвали дисперсией β величины x *мат. ожид.* $(x-a)^2$, т.-е. положили $\beta_1=$ М. О. $(x-a_1)^2$, $\beta_2=$ М. О. $(x-a_2)^2$ и т. д. Назовем теперь штандартом σ величины x значение $\sigma=\sqrt{\beta}$; там, где это не будет вызывать неясностей, мы будем также приписывать σ тот же значок, что и x, как мы это делали для математических ожиданий и для дисперсии; но в иных случаях придется писать $\sigma(x)$, $\beta(x)$, $a(x)$ для соответствующих постоянных, так что, например:

$$\sigma_1=\sigma(x_1),\ \beta_1=\beta(x_1),\ a_1=a(x_1),$$
$$\sigma_2=\sigma(x_2),\ \beta_2=\beta(x_2),\ a_2=a(x_2).$$

Согласно данному определению, вообще

$$\beta = \sigma^2. \tag{123}$$

Зависимость между двумя величинами в значительной мере характеризуется[1]) величиной математического ожидания произведения отклонений рассматриваемых величин, и если мы хотим отвлечься от абсолютного значения каждой из этих величин и характеризовать эту связь отвлеченным (а не именованным) числом, то для этой цели служит коэффициент корреляции $R_{1,2} = R(x_1, x_2)$ между данными числами, который определяется формулой

$$R_{1,2} = \frac{\text{М. О.}\,[x_1 - a_1]\,[x_2 - a_2]}{\sigma_1 \sigma_2} = \frac{\text{М. О.}\,x_1 x_2 - a_1 a_2}{\sigma_1 \sigma_2}. \tag{124}$$

Очевидно, в случае, когда величины x_1 и x_2 независимы, коэффициент корреляции между ними $R_{1,2} = 0$, однако, обратное заключение, разумеется, вообще несправедливо: например, если при $x_1 = 2$, $x_2 = 0$, а при $x_1 = -1$ имеем либо $x_2 = 1$, либо $x_2 = -1$, где все пары значений $(2, 0), (-1, +1), (-1, -1)$ равновероятны, то М. О. $x_1 = \frac{1}{3} \cdot 2 - \frac{2}{3} \cdot 1 = 0$, М. О. $x_2 = \frac{1}{3} \cdot 1 - \frac{1}{3} \cdot 1 = 0$, М. О. $x_1 x_2 = \frac{1}{3} \cdot 1 - \frac{1}{3} \cdot 1 = 0$, а между тем зависимость между x_1 и x_2 настолько велика, что, если дано значение x_2, мы сразу можем указать значение x_1, и если известно, что $x_1 = 2$, можем утверждать, что $x_2 = 0$.

Следует заметить, что определение коэффициента корреляции между двумя событиями, данное в гл. II, ч. 1-я, является частным случаем общего определения, которое мы только что ввели. Действительно, если (A_1) и (A_2) суть соответственные вероятности фактов A_1 и A_2, а (A_1, A_2) — вероятность их совмещения, то М. О. $(x_1) = (A_1)$, М. О. $(x_2) = (A_2)$, М. О. $x_1^2 = (A_1)$, М. О. $x_2^2 = (A_2)$, М. О. $x_1 x_2 = (A_1, A_2)$

[1]) Более подробно мы на этом остановимся впоследствии.

(x_1 и x_2 означают соответственно число (1 или 0) появлений события A_1 и A_2 в одном опыте). Тогда

$$\beta_1 = \sigma_1^2 = (A_1) - (A_1)^2 = (A_1)(\overline{A_1}),$$
$$\beta_2 = \sigma_2^2 = (A_2) - (A_2)^2 = (A_2)(\overline{A_2}),$$

где $(\overline{A})$ есть вероятность непоявления факта A. Отсюда видим, что формула (124) равнозначна в этом случае формуле

$$R_{1,2} = \frac{(A_1, A_2) - (A_1)(A_2)}{\sqrt{(A_1)(A_2)(\overline{A_1})(\overline{A_2})}}, \tag{27}$$

выведенной в указанном месте.

Проверим, что так же, как и в указанном частном случае, вообще

$$-1 < R_{1,2} < 1, \tag{125}$$

при чем $R_{1,2}$ достигает крайних значений $+1$ и -1 тогда и только тогда, когда между уклонениями $x_1 - a_1$ и $x_2 - a_2$ существует строгая пропорциональность[1]), т.-е.

$$\frac{x_1 - a_1}{x_2 - a_2} = \pm \frac{\sigma_1}{\sigma_2}, \tag{126}$$

где знак $+$ соответствует $R_{1,2} = 1$, а знак $-$ имеет место при $R_{1,2} = -1$.

Действительно, принимая во внимание (124),

$$\text{М. О.}\left[\frac{x_1 - a_1}{\sigma_1} \pm \frac{x_2 - a_2}{\sigma_2}\right]^2 = 1 \pm 2\frac{\text{М. О.}\,(x_1 - a_1)(x_2 - a_2)}{\sigma_1 \sigma_2} + 1 =$$
$$= 2 \pm 2R_{1,2}, \tag{127}$$

и так как первая часть равенства не отрицательна, то

$$1 \pm R_{1,2} \geqslant 0, \text{ т.-е. } -1 \leqslant R_{1,2} \leqslant 1.$$

При этом, если, выбирая определенный знак (например $+$), мы получаем во второй части (127) нуль, т.-е. если $R_{1,2} = -1$, то это происходит тогда и только тогда, когда имеем тождественно

$$\frac{x_1 - a_1}{\sigma_1} + \frac{x_2 - a_2}{\sigma_2} = 0, \text{ т.-е. когда } \frac{x_1 - a_1}{x_2 - a_2} = -\frac{\sigma_1}{\sigma_2}.$$

---

[1]) Следует заметить, что иная функциональная зависимость между x_1 и x_2 совместима с любыми значениями коэффициента корреляции $|R| < 1$. (См. стр. 358.)

4. Полагая теперь, что мы имеем любое число n величин, находим, пользуясь вновь введенными обозначениями,

$$B = \text{М. О. } [(x_1 + x_2 + \ldots + x_n) - (a_1 + a_2 + \ldots + a_n)]^2 =$$
$$= \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2 + 2\,[R_{1,2}\,\sigma_1\,\sigma_2 + R_{1,3}\,\sigma_1\,\sigma_3 + \ldots + R_{n-1,n}\,\sigma_{n-1}\,\sigma_n]. \qquad (120\text{ bis})$$

Прежде всего можем заметить, что благодаря (125), во всяком случае,

$$B \leqslant \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2 + 2\,\sigma_1\,\sigma_2 + 2\,\sigma_1\,\sigma_3 + \ldots + 2\sigma_{n-1}\,\sigma_n = (\sigma_1 + \sigma_2 + \ldots + \sigma_n)^2.$$

Поэтому, как бы велика ни была корреляция между величинами $x_1, x_2, \ldots, x_n$ (доходящая до точной пропорциональности), закон больших чисел к ним применим, если $\frac{\sigma_1 + \sigma_2 + \ldots + \sigma_n}{n}$ стремится к 0 с увеличением n, так как тогда

$$\frac{B}{n^2} \leqslant \left(\frac{\sigma_1 + \sigma_2 + \ldots + \sigma_n}{n}\right)^2$$

и подавно стремится к 0. Однако этот случай находит мало применений, так как он, очевидно, может представиться лишь при условии, что большинство величин σ_n сами становятся произвольно малы, когда n достаточно велико [1]).

Более интересен тот случай, когда изменчивость величин $x_1, x_2, \ldots, x_n$, измеряемая их дисперсией (или штандартом), более или менее одинакова, так что вообще можно указать два определенных положительных числа таких, что $l < \sigma_i < L$ при всех значениях i (за исключением, может быть, немногих отдельных значков i, наличности которых мы можем, для упрощения рассуждений, вовсе не предполагать).

---

[1]) Это условие было бы соблюдено в примере, рассмотренном на стр. 180, если бы мы положили $x_i = \frac{m_i - ip}{(ipq)^{\frac{1}{2}+\varepsilon}}$, где $\varepsilon > 0$, вместо $\frac{m_i - ip}{\sqrt{ipq}}$.

Тогда, очевидно,

$$ln < \sigma_1 + \sigma_2 + \ldots + \sigma_n < Ln,$$

а потому вообще можем лишь утверждать, что

$$B < L^2 n^2$$

и

$$\frac{B}{n^2} < L^2.$$

Кроме того, если бы при этом все величины x_i и x_k были связаны между собой так, что все коэффициенты $R_{i,k}$ имели бы некоторую положительную нижнюю границу ρ, т.-е.

$$R_{i,k} > \rho > 0,$$

то

$$B = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2 + 2(R_{1,2}\sigma_1\sigma_2 + \ldots + R_{n-1,n}\sigma_{n-1}\sigma_n) >$$
$$> nl^2 + \rho l^2 n(n-1) > n^2 \rho l^2,$$

и

$$\frac{B}{n^2} > \rho l^2;$$

следовательно, когда коэффициенты корреляции между всеми величинами положительны и не стремятся к нулю, то закон больших чисел к этим величинам вообще [1]) неприменим.

Предположим, что n кандидатам дается одна и та же задача из числа 100 задач, которые имеют равные шансы быть предложенными. Допустим, что каждый кандидат умеет решить только 75 из этих задач. Очевидно, что если бы факт правильного решения задачи одним из кандидатов не изменял вероятности того, что эта задача будет также удачно решена и другими кандидатами, т.-е. если бы благоприятные исходы экзамена для двух кандидатов были независимыми событиями, то в данном случае был бы применим закон больших чисел в форме теоремы Бернулли, а именно: если число кандидатов достаточно велико, то можно быть уверенным, что приблизительно $\frac{3}{4}$ из них решат задачу правильно; так что, полагая, например, что экзамен состоит лишь в том, чтобы ответить „да“ или „нет“ на поставленный вопрос, следовало бы

---

[1]) Предполагая, конечно, что величины σ_i не стремятся к 0.

всегда признать правильным то решение предложенной задачи, которое дано большинством кандидатов.

Приблизительно так рассуждали некоторые философы и математики конца XVIII и начала XIX веков, которые утверждали, что судебный трибунал, состоящий из достаточно большого числа судей, решая вопросы большинством голосов, практически был бы непогрешим в своих приговорах; это утверждение, очевидно, неверно, потому что здесь не принимается во внимание, что все судьи судят на основании тех же самых свидетельских показаний и вещественных доказательств, так что в простом деле все они более или менее одинаково разберутся, а если запутанные обстоятельства вводят в заблуждение одних, то и для других судей ошибка становится более вероятной: иначе говоря, в случае судебного приговора отсутствует условие независимости между суждениями отдельных судей, и это коренным образом изменяет положение вещей. Действительно, допустим, возвращаясь к нашим экзаменационным задачам, что 15 определенных задач могут решить все кандидаты, 5 задач не умеет решить ни один и лишь отношение к остальным 80 задачам независимо у всех кандидатов (так что если данная задача относится к последней категории, то вероятность ее решения для одного кандидата равна $\frac{3}{4}$, независимо от того, известно или нет, что другой ее решил). В таком случае вероятность $p_{1,2}$, что задача будет решена двумя кандидатами, равна

$$p_{1,2} = 0{,}15 + 0{,}8 \cdot \left(\frac{3}{4}\right)^2 = 0{,}6,$$

и, следовательно, по формуле (27), коэффициент корреляции

$$R_{1,2} = \frac{0{,}6 - \left(\frac{3}{4}\right)^2}{\frac{3}{4} \cdot \frac{1}{4}} = 0{,}2.$$

Таким образом, на основании вышесказанного общего положения, закон больших чисел здесь неприменим, что очевидно и само собой, так как имеется 15 шансов на 100, что задача будет решена всеми, и 5 шансов на 100, что задача не будет решена правильно никем.

5. Докажем, напротив, следующую теорему:

Теорема. Если дисперсии β_i всех рассматриваемых величин x_i ограничены (т.-е. $\beta_i \leqslant L^2$, $\sigma_i \leqslant L$) и зависимость между x_i и x_k настолько ослабевает, когда $|i-k|$ становится достаточно большим, что коэффициент корреляции $R_{i,k}$ между ними равномерно стремится при этом к нулю, то вероятность неравенства

$$\left|\frac{x_1+x_2+\ldots+x_n}{n}-\frac{a_1+a_2+\ldots+a_n}{n}\right|<\varepsilon, \qquad (121)$$

при произвольно малом ε, стремится к достоверности, когда n безгранично возрастает (иными словами, к величинам $x_1, x_2, \ldots, x_n$ применим закон больших чисел).

Действительно, по условию, как бы мало ни было наперед заданное число α, всегда можно указать такое достаточно большое число h (независимое от n), что, при $|i-k| \geqslant h$, $|R_{i,k}| < \alpha$. Но для доказательства теоремы достаточно установить, что

$$\frac{B}{n^2}=\frac{1}{n^2}\left[S_1+S_2+\ldots+S_n\right]$$

с возрастанием n стремится к 0, где

$$S_1=\sigma_1^2+R_{1,2}\sigma_1\sigma_2+R_{1,3}\sigma_1\sigma_3+\ldots+R_{1,n}\sigma_1\sigma_n,$$
$$S_2=R_{2,1}\sigma_1\sigma_2+\sigma_2^2+R_{2,3}\sigma_2\sigma_3+\ldots+R_{2,n}\sigma_2\sigma_n,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$S_n=R_{n,1}\sigma_1\sigma_n+R_{n,2}\sigma_2\sigma_n+\ldots+\sigma_n^2.$$

Для этой цели замечаем, что каждая из сумм S_i удовлетворяет неравенству

$$|S_i|<L^2[2h+n\alpha], \qquad (128)$$

так как $\sigma_i^2 < L^2$ и $|R_{i,k}\sigma_i\sigma_k| < L^2$ для любых i и k, но, кроме того, при $|i-k| \geqslant h$, $|R_{i,k}\sigma_i\sigma_k| < L^2\alpha$.

Отсюда следует, что

$$\frac{B}{n^2}<\frac{nL^2(2h+n\alpha)}{n^2}=2L^2\frac{h}{n}+L^2\alpha.$$

Задавая теперь произвольно малую величину δ, можем выбрать α настолько малым, чтобы $L^2\alpha < \frac{\delta}{2}$, а затем (принимая во внимание, что значение h вполне определилось после выбора α) подобрать n настолько большим, чтобы $2L^2\frac{h}{n} < \frac{\delta}{2}$. Поэтому $\frac{B}{n^2} < \delta$, что и требовалось доказать.

В примере, рассмотренном выше, когда $x_i = \frac{m_i - ip}{\sqrt{ipq}}$, где закон больших чисел оказался неприменим, условия нашей теоремы, разумеется, не соблюдены. В самом деле, хотя первое условие имеет место, так как $\beta_i = \sigma_i^2 = \text{М. О.} \left[\frac{m_i - ip}{\sqrt{ipq}}\right]^2 = 1$, так что $\sigma_i = L = 1$, но мы нашли, полагая $k > i$, что

$$R_{i,k} = \text{М. О. } x_i x_k = \sqrt{\frac{i}{k}}$$

(ведь $a_i = 0$). Поэтому, здесь $R_{i,k}$ лишь тогда стремится к 0, когда $\frac{i}{k}$ стремится к 0, так что, как бы велика ни была разность $k - i = h$, не только нельзя утверждать, что коэффициент корреляции $R_{i,k} = \sqrt{\frac{k-h}{k}}$ весьма мал, но, при k достаточно большом, $R_{i,k}$ становится сколь угодно близким к 1.

Заметим, что при доказательстве данной выше теоремы существенную роль играло лишь требование, чтобы положительные коэффициенты корреляции $R_{i,k}$ стремились к 0 при $|i - k|$ достаточно большом; если же среди наших коэффициентов $R_{i,k}$ имеются отрицательные, то они лишь способствуют уменьшению B, ибо соответствующие им члены $R_{i,k}\sigma_i\sigma_k$ всегда отрицательны, так что $R_{i,k}$ может даже быть произвольно близким к -1, не нарушая применимости закона больших чисел (опасаться же, что отрицательные члены могут взять перевес над положительными, не приходится, так как величина B никогда не может сделаться отрицательной).

Однако можно заметить, что если все $R_{i,k}$ отрицательны, то не может существовать такого определенного числа ρ, чтобы все $|R_{i,k}| > \rho > 0$ при n весьма большом, так как легко проверить, что в этом случае необходимо должны быть значения $R_{i,k}$, удовлетворяющие неравенству[1] $|R_{i,k}| \leqslant \frac{L}{l} \cdot \frac{1}{n-1}$ (полагая $l \leqslant \sigma_i \leqslant L$).

[1] Иначе, все суммы S_i, рассматриваемые на предыдущей странице, были бы отрицательны.

6. Применим вышеизложенные общие соображения к ряду опытов, образующих, по терминологии А. А. Маркова, простую цепь. Допустим, что производится ряд опытов, в которых отмечается появление некоторого события A. Пусть вероятность этого события в $(n+1)$-м опыте существенным образом зависит от того, произошло ли событие A в n-м опыте или нет: положим, что в первом случае эта вероятность равна α, а во втором случае равна β, причем результаты всех предшествующих $(n-1)$ опытов не оказывают никакого влияния на появление A в $(n+1)$-м опыте, после того как выяснился исход n-го опыта. Мы говорим тогда, что рассматриваемые опыты образуют простую цепь. Предположим, например, что мы имеем ряд урн с двумя шарами в каждой, одним черным и одним белым; положим, что из этих урн по порядку вынимают один шар, называя фактом A появление белого шара, и затем вынутый шар перекладывают в следующую урну, так что (за исключением первого извлечения) перед каждым $(n+1)$-м вниманием в урне находятся 3 шара, из которых 2 белых, если событие A произошло в n-м опыте, и только 1 белый шар, если A в n-м опыте не наступило. Таким образом в этом примере $\alpha=\frac{2}{3}$ и $\beta=\frac{1}{3}$, каковы бы ни были результаты предшествующих $(n-1)$ опытов.

Определим, какова первоначальная вероятность p_n события A в n-м опыте, пока нам ничего не известно ни об одном из результатов предшествующих опытов, а также, какова вероятность $p_n^{(k)}$ факта A в n-м опыте, если известно, что он произошел в одном определенном, k-м, из предшествующих опытов.

Нетрудно видеть, что и p_n и $p_n^{(k)}$ удовлетворяют одному и тому же уравнению в конечных разностях

$$p_{n+1}=\alpha p_n+\beta(1-p_n), \tag{129}$$

так как наступление A в $(n+1)$-м опыте возможно при двух обстоятельствах: либо совместно с его наступлением в n-м опыте, либо с его ненаступлением в том же опыте.

Применяя указанный раньше метод решения подобных уравнений (стр. 136) в конечных разностях, мы находим частное решение уравнения $P=\frac{\beta}{1+\beta-\alpha}$, которое показывает, что в случае, когда

вероятность p_1 в первом опыте равна $p_1 = P = \frac{\beta}{1+\beta-\alpha}$, мы имеем также, для всякого n, $p_n = \frac{\beta}{1+\beta-\alpha}$. В частности, это имеет место в нашем примере с шарами, где первая урна, подобно последующим, содержит 1 белый и 1 черный шар; тогда $p_1 = p_n = \frac{1}{2}$.

Если же $p_1 - P \gtrless 0$, то, по формуле (92),

$$p_n = P + (p_1 - P)(\alpha - \beta)^{n-1} = \frac{\beta}{1+\beta-\alpha} + \left(p_1 - \frac{\beta}{1+\beta-\alpha}\right)(\alpha-\beta)^{n-1}. \tag{130}$$

Точно так же находим вероятность $p_n^{(k)}$ появления A в n-м опыте, если известно, что факт A произошел в k-м опыте; при $k=1$ рассматриваемая вероятность $p_n^{(1)}$ должна получиться из формулы (130), в которой $p_1 = 1$, так как факт A был бы в первом опыте достоверен, т.-е.

$$p_n^{(1)} = \frac{\beta}{1+\beta-\alpha} + \frac{1-\alpha}{1+\beta-\alpha}(\alpha-\beta)^{n-1}; \tag{130 bis}$$

поэтому, принимая во внимание, что k-й опыт играет такую же роль по отношению к n-му опыту, как 1-й опыт по отношению к $(n-k+1)$-му опыту, заключаем, что вообще

$$p_n^{(k)} = \frac{\beta}{1+\beta-\alpha} + \frac{1-\alpha}{1+\beta-\alpha}(\alpha-\beta)^{n-k}. \tag{131}$$

Таким же образом, обозначая через $p_n^{(\bar{k})}$ вероятность наступления A в n-м опыте, если известно, что в k-м опыте факт A не произошел, мы найдем (заменяя во второй части (130) p_1 через 0, а $n-1$ через $n-k$)

$$p_n^{(\bar{k})} = \frac{\beta}{1+\beta-\alpha} - \frac{\beta}{1+\beta-\alpha}(\alpha-\beta)^{n-k}. \tag{131 bis}$$

В указанном выше примере мы имеем

$$p_n^{(k)} = \frac{1}{2}\left[1 + \left(\frac{1}{3}\right)^{n-k}\right], \quad p_n^{(\bar{k})} = \frac{1}{2}\left[1 - \left(\frac{1}{3}\right)^{n-k}\right],$$

так что, например, если мы знаем, что из первого ящика был вынут белый шар (и переложен во второй), то вероятность, что из второго ящика будет извлечен белый шар, равна $\frac{1}{2}\cdot\frac{4}{3}=\frac{2}{3}$ (как нам и раньше было известно); вероятность же, что белый шар появится из 3-го ящика, равна уже только $\frac{1}{2}\left(\frac{10}{9}\right)=\frac{5}{9}$, и т. д., при чем влияние результата 1-го опыта идет все ослабевая и, так как при n очень большом $\left(\frac{1}{3}\right)^{n-1}$ стремится к 0, то во всяком случае для очень удаленных ящиков вероятность появления белого шара остается весьма близкой к $\frac{1}{2}$.

Тот же результат имеет место и в общем случае [1]), когда $|\alpha-\beta|<1$, так как независимо от исхода k-го опыта вероятность наступления A в n-м опыте при $n-k$ весьма большом, стремится к $P=\frac{\beta}{1+\beta-\alpha}$, ибо $(\beta-\alpha)^n$ стремится к 0.

Вместе с тем, нетрудно видеть, что в данном случае закон больших чисел соблюден, т.-е. можно утверждать с вероятностью, приближающейся к достоверности, что

$$\left|\frac{m}{n}-\frac{p_1+p_2+\ldots+p_n}{n}\right|<\varepsilon, \qquad (99\ \text{bis})$$

---

[1]) $\alpha-\beta=1$ означает, что $\alpha=1$ и $\beta=0$, в таком случае $p_n=p_1$: если событие A произошло в первом опыте, оно будет повторяться неизменно; напротив, если оно не произошло, то оно никогда и не наступит; ясно, что тогда не может быть речи о законе больших чисел. Этот случай представится, например, если мы имеем две урны с шарами, из которых в одной только белые, а в другой только черные шары, при чем шары во всех опытах извлекаются из той же самой урны. При $\alpha-\beta=-1$, имеем $\alpha=0$ и $\beta=1$; этот случай представится, если, имея те же две урны, мы будем поочередно вынимать шары из обеих урн. Тогда $p_n=\frac{1}{2}\mp\left(p_1-\frac{1}{2}\right)$, т.-е. $p_2=q_1$, и вообще $p_{2k+1}=p_1$, а $p_{2k}=q_1$, поэтому $\frac{m}{n}$ может быть равно только одному из трех значений $\frac{1}{2}-\frac{1}{2n}$, $\frac{1}{2}$, $\frac{1}{2}+\frac{1}{2n}$; таким образом в этом случае закон больших чисел соблюдается с исключительной точностью.

где m есть число появлений события A при n опытах, а ε — произвольно малая величина. Для этого достаточно, пользуясь формулой (131), заметить, что коэффициенты корреляции $R_{k,\eta}$ стремятся к 0, когда $n-k$ неограниченно возрастает.

7. Имея в виду дальнейшие приложения, мы вычислим предел, к которому стремится средняя вероятность $\frac{p_1+p_2+\ldots+p_n}{n}$ события A при возрастании n, а также соответствующее значение дисперсии B и связанного с нею коэффициента дисперсии, который мы определяем, следуя А. А. Маркову.

Пусть

$$B = \text{М. О. } [m-(p_1+p_2+\ldots+p_n)]^2;$$

полагая $np = p_1+p_2+\ldots+p_n$, видим, что, если бы мы рассматривали ряд n независимых опытов, где вероятность появления события A' была бы равна p, то, согласно теореме Бернулли, число m' появлений события A' будет удовлетворять тому же неравенству (99 bis), что и число m. Одним из существенных отличий обеих серий опытов (которое сможет быть обнаружено статистически) является обычно различная величина дисперсии B. В случае Бернулли мы знаем, что

$$\text{М. О. } (m-np)^2 = npq. \tag{66}$$

Поэтому существенное значение имеет для каждого данного ряда опытов рассмотрение величины

$$E = \frac{B}{npq}, \tag{132}$$

называемой **теоретическим коэффициентом дисперсии** данной схемы опытов, который представляет собой отношение дисперсии B числа m в данном ряде n опытов к дисперсии m' числа появлений события A' в ряде n независимых опытов с постоянною вероятностью $p = \frac{p_1+p_2+\ldots+p_n}{n}$.

Заметим, что на существенную роль коэффициента дисперсии для статистики обратил впервые внимание Лексис (Lexis), который систематически применял его к исследованию различных демо-

графических явлений; однако, в отличие от А. А. Маркова, школа Лексиса называет коэффициентом дисперсии

$$\sqrt{E}=\sqrt{\frac{B}{npq}}.$$

Итак, вычислим сначала среднюю вероятность для серии опытов, образующих простую цепь. Применяя формулу (130), находим, что

$$p_1+p_2+\ldots+p_n=nP+(p_1-P)[1+(\alpha-\beta)+(\alpha-\beta)^2+\ldots$$
$$\ldots+(\alpha-\beta)^{n-1}]=nP+(p_1-P)\frac{(\alpha-\beta)^n-1}{\alpha-\beta-1},$$

откуда

$$p=\frac{p_1+p_2+\ldots+p_n}{n}=P+\frac{1}{n}(p_1-P)\frac{1-(\alpha-\beta)^n}{1-\alpha+\beta}.$$

Поэтому, при возрастании n, во всяком случае пред. $p=P$; но, кроме того, если $p_1=P$, то и $p=P$ при всяком n; в последнем случае в нашей серии опытов средняя вероятность постоянна, хотя опыты не независимы. Следовательно, вообще при всяком p_1 неравенство

$$\left|\frac{m}{n}-P\right|<\varepsilon, \tag{133}$$

произвольно мало отличающееся от (99 bis) для достаточно больших значений n, также стремится к достоверности.

Для некоторого упрощения вычислений мы в дальнейшим ограничимся предположением, что $p=p_1=P$, предоставляя читателю проверить, что наши выводы остаются в силе и в общем случае.

Вообще, согласно (120), имеем

$$B=\text{M. O.}[m-(p_1+p_2+\ldots+p_n)]^2=p_1q_1+p_2q_2+\ldots$$
$$\ldots+p_nq_n+2\left\{p_1[(p_2^{(1)}-p_2)+(p_3^{(1)}-p_3)+\ldots]+\right.$$
$$\left.+p_2[(p_3^{(2)}-p_3)+(p_4^{(2)}-p_4)+\ldots]+\ldots\right\},$$

так так в данном случае

$$\beta_{ih}=\text{M. O.}(x_i-p_i)(x_h-p_h)=\text{M. O. }x_ix_h-p_ip_h=$$
$$=p_ip_h^{(i)}-p_ip_h=p_i(p_h^{(i)}-p_h).$$

Ввиду сделанного допущения, что $p_1 = P$, находим, что

$$p_i q_i = PQ, \; p_i p_h = P^2,$$

$$p_i p_h^{(i)} = P\left[P + \frac{1-\alpha}{1+\beta-\alpha}(\alpha-\beta)^{h-i}\right],$$

где $h > i$ [см. (131)]; поэтому

$$B = nPQ + 2P\frac{1-\alpha}{1+\beta-\alpha}\Big\{[(\alpha-\beta)+(\alpha-\beta)^2+\ldots$$

$$\ldots+(\alpha-\beta)^{n-1}]+[(\alpha-\beta)+\cdots+(\alpha-\beta)^{n-2}]+\ldots\Big\} =$$

$$= nPQ + 2P\frac{1-\alpha}{1+\beta-\alpha}\left\{\frac{1-(\alpha-\beta)^{n-1}}{1+\beta-\alpha}+\right.$$

$$\left.+\frac{1-(\alpha-\beta)^{n-2}}{1+\beta-\alpha}+\ldots\right\}(\alpha-\beta) =$$

$$= nPQ + 2PQ(\alpha-\beta)\left[\frac{n}{1+\beta-\alpha}+\frac{(\alpha-\beta)^n-1}{(1+\beta-\alpha)^2}\right],$$

так как

$$\frac{1-\alpha}{1+\beta-\alpha} = 1-P = Q.$$

Следовательно,

$$B = \frac{nPQ(1+\alpha-\beta)}{1+\beta-\alpha}+\frac{2PQ(\alpha-\beta)}{(1+\beta-\alpha)^2}[(\alpha-\beta)^n-1]$$

и

$$E = \frac{B}{nPQ} = \frac{1+\alpha-\beta}{1+\beta-\alpha}\left[1-\frac{2(\alpha-\beta)}{n}\cdot\frac{1-(\alpha-\beta)^n}{(1+\beta-\alpha)^2}\right],$$

откуда видим [1]), что при возрастании n

$$\text{пред. } E = \frac{1+\alpha-\beta}{1+\beta-\alpha} = \frac{1+\rho}{1-\rho}. \tag{134}$$

В примере с шарами $E = \dfrac{1+\frac{1}{3}}{1-\frac{1}{3}} = 2.$

А. А. Марков исследовал также более общий случай, когда α и β зависят от n и, кроме того, распространил свои выводы на так назы-

[1]) По определению коэффициента регрессии, данному в гл. III, ч. 1, $\alpha - \beta = \rho$ есть коэффициент регрессии (корреляции) между наступлением события в $(n+1)$-м и в n-м опыте.

ваемые сложные цепи, когда вероятность появления события A в $(n+1)$-м опыте делается вполне определенной, если известны результаты предшествующих ему i опытов, сохраняя то же значение, каковы бы ни были результаты всех первых $n-i$ опытов.

Идеи Маркова представляют не только большой математический интерес, но особенно важны в методологическом отношении, давая теоретические схемы, пригодные для интерпретации многих явлений. Мы рассмотрим немного позднее один из примеров, тщательно разработанных самим А. А. Марковым, и укажем здесь на приложение схемы опытов, образующих цепь, к наследственности.

8. В случае однополого размножения можно принять, что если индивид обладает одним из нескольких s (для простоты положим $s=2$) несовместимых признаков A_1 или A_2, то, каковы бы ни были его предки, вероятность его потомству обладать признаком A_1 равна соответственно α или β. В таком случае, обозначая через p_n вероятность индивиду n-го поколения обладать признаком A_1, мы имеем характерное для простой цепи соотношение:

$$p_{n+1}=\alpha p_n+\beta(1-p_n).$$

Если принимать во внимание те или иные данные изменения окружающей среды, α и β могут быть определенными функциями n.

Более интересный случай двуполого размножения приводит к несколько более сложной схеме (которая Марковым рассмотрена не была), представляющей собой соединение нескольких переплетающихся цепей. Предполагая опять, для простоты, что $s=2$, мы можем допустить (в соответствии с общепринятыми биологическими воззрениями), что вероятность принадлежности потомка к классу A_1 или A_2 всецело определяется в данной среде характером его обоих родителей [1]) (независимо от более отдаленных предков). Поэтому, полагая, для определенности, что вероятность матери обладать признаком A_1 не зависит от того, обладает ли указанным признаком отец (допущение такой независимости соответствует

[1]) Это имеет место лишь в том случае, если все индивиды класса A_1 (и соответственно класса A_2) генотипически однородны; поэтому случай $s=2$ практически имеет мало приложений и обычно $s=3^a$, где a есть число генов, которым определяется данный признак.

отсутствию полового отбора по данному признаку), находим, что

$$p_{n+1}=\alpha p_n p'_n+\beta p_n\left(1-p'_n\right)+\gamma\left(1-p_n\right)p'_n+$$
$$+\delta\left(1-p_n\right)\left(1-p'_n\right),$$
$$p'_{n+1}=\alpha' p_n p'_n+\beta' p_n\left(1-p'_n\right)+\gamma'\left(1-p_n\right)p'_n+$$
$$+\delta'\left(1-p_n\right)\left(1-p'_n\right), \qquad (135)$$

где p_n и p'_n — вероятности обладать признаком A_1 индивидам n-го поколения соответственно мужского и женского пола, а $(\alpha, \beta, \gamma, \delta)$ и $(\alpha', \beta', \gamma', \delta')$ — вероятности, что особь мужского и женского пола, рождающаяся у соответствующей пары родителей, будет обладать указанным признаком. В большинстве случаев можно принять, что $p'_n=p_n$, $\alpha'=\alpha$, $\beta'=\gamma'=\beta=\gamma$ и $\delta=\delta'$; тогда мы получим лишь одно уравнение

$$p_{n+1}=\alpha p^2+2\beta p_n(1-p_n)+\delta(1-p_n)^2, \qquad (136)$$

характеризующее простейшую переплетающуюся цепь. Мы получили бы простую цепь А. А. Маркова лишь в том случае, если бы предположили, что p'_n дано, например $p'_n=1$, что соответствовало бы допущению, что женская особь всегда обладает признаком A_1, т.-е. что самцы всегда скрещиваются с самками, обладающими признаком A_1; в этом предположении уравнения (135) приведутся к одному уравнению

$$p_{n+1}=\alpha p_n+\gamma(1-p_n).$$

Не останавливаясь на математическом анализе полученных уравнений, заметим лишь, что они обнаруживают чрезвычайную плодотворность пути исследований, намеченного А. А. Марковым.

Заканчивая эту главу, полезно будет указать общую теорему, дающую во всех случаях условие необходимое и достаточное для приложимости к данному ряду зависимых опытов закона больших чисел.

Теорема. Пусть p_h представляет первоначальную вероятность факта A в h-м опыте, $p_h^{(k)}$ — его вероятность, когда известно, что факт A произошел в k-м опыте,

а $p_h^{(\bar{k})}$ — вероятность A в h-м опыте, если известно, что в k-м опыте он не наступил; пусть далее

$$P_n = \frac{p_1 + p_2 + \ldots + p_n}{n}, \quad P_n^{(k)} = \frac{p_1^{(k)} + p_2^{(k)} + \ldots + p_n^{(k)}}{n},$$

$$P_n^{(\bar{k})} = \frac{p_1^{(\bar{k})} + p_2^{(\bar{k})} + \ldots + p_n^{(\bar{k})}}{n}$$

представляют: первая — первоначальную среднюю вероятность события A в наших n опытах, а $P_n^{(k)}$ и $P_n^{(\bar{k})}$ — соответственно условные значения средней вероятности [1]) после того, как становится известным положительный или отрицательный результат одного только k-го опыта. В таком случае для применимости закона больших чисел необходимо и достаточно, чтобы произведение

$$p_k q_k \left| P_n^{(k)} - P_n^{(\bar{k})} \right|$$

стремилось равномерно (т.-е. независимо от k) к нулю при возрастании n. Иначе говоря, рассматривая любой определенный k-й опыт из нашего ряда, нужно, чтобы разность между условными средними $P_n^{(k)}$ и $P_n^{(\bar{k})}$ вообще стремилась к 0, но эта разность не должна непременно стремиться к 0 для тех значений k, для которых $p_k q_k$ само стремится к 0, т.-е. для тех опытов, исход которых почти несомненен.

Упражнения. 1) Даны 2 урны с шарами; вероятности появления белого шара из каждой урны все время остаются равны соответственно $\frac{3}{4}$ и $\frac{1}{4}$. В случае появления белого шара, извлечение снова производится из той же

---

[1]) Теорема остается в силе, если вместо $P_n^{(k)}$ и $P_n^{(\bar{k})}$ рассматривать

$$R_n^{(k)} = \frac{p_{k+1}^{(k)} + p_{k+2}^{(k)} + \ldots + p_n^{(k)}}{n} \text{ и } R_n^{(\bar{k})} = \frac{p_{k+1}^{(\bar{k})} + p_{k+2}^{(\bar{k})} + \ldots + p_n^{(\bar{k})}}{n};$$

иными словами, значения $p_i^{(k)}$ и $p_i^{(\bar{k})}$ не играют роли при $i < k$, т.-е. влияние k-го опыта на предшествующие опыты может не быть учитываемо. См. мою статью „О законе больших чисел". Сообщ. Харьк. Мат. О-ва 1918 г.

урны; в противном случае выбирается другая урна. При первом вынимании выбор обеих урн равновероятен. Требуется определить среднюю вероятность появления белого шара и дисперсию числа m этих появлений.

Отв. Пусть α_n, β_n, $(\alpha_n+\beta_n=1)$ — вероятности, что в n-м опыте извлечение производится соответственно из 1-й или 2-й урны; p_n — вероятность появления белого шара в n-м опыте. Имеем уравнения

$$p_n=\frac{3}{4}\alpha_n+\frac{1}{4}\beta_n,$$

$$\alpha_{n+1}=\frac{3}{4}\alpha_n+\frac{3}{4}\beta_n=\frac{3}{4},$$

$$\beta_{n+1}=\frac{1}{4}\alpha_n+\frac{1}{4}\beta_n=\frac{1}{4},$$

которым удовлетворяют как первоначальные вероятности, так и условные вероятности $p_n^{(k)}$, $\alpha_n^{(k)}$, $\beta_n^{(k)}$, соответствующие предположению, что при k-м опыте вынут белый шар. Поэтому $\alpha_{n+1}=\frac{3}{4}$, $\beta_{n+1}=\frac{1}{4}$, $p_{n+1}=\frac{5}{8}$ при всяком $n>0$, и точно так же $\alpha_{k+h}^{(k)}=\frac{3}{4}$, $\beta_{k+h}^{(k)}=\frac{1}{4}$, $p_{k+h}^{(k)}=\frac{5}{8}$ при $h>1$, а

$$\alpha_{k+1}^{(k)}=\frac{\frac{3}{4}\cdot\frac{3}{4}}{\frac{3}{4}\cdot\frac{3}{4}+\frac{1}{4}\cdot\frac{1}{4}}=\frac{9}{10},\quad \beta_{k+1}^{(k)}=\frac{1}{10},\quad p_{k+1}^{(k)}=\frac{7}{10}$$

при $k>1$, и $\alpha_2^{(1)}=\frac{3}{4}$, $\beta_2^{(1)}=\frac{1}{4}$, $p_2^{(1)}=\frac{5}{8}$.

Следовательно,

$$P=\frac{p_1+p_2+\ldots+p_n}{n}=\frac{5}{8}-\frac{1}{8n},$$

а

$$B=\frac{1}{4}+(n-1)\frac{15}{64}+2(n-2)\frac{3}{64}=\frac{21}{64}n-\frac{11}{64};$$

коэффициент дисперсии

$$E=\frac{B}{\frac{15}{64}n+\frac{1}{32}-\frac{1}{64n}}=\frac{21n-11}{15n+2-\frac{1}{n}}\neq\frac{7}{5}$$

Очевидно, что здесь мы не имеем цепи А. А. Маркова, так как если бы белый шар появился k раз под-ряд, вероятность его появления в $(k+1)$-м опыте имела бы пределом $\frac{3}{4}$, при неограниченном возрастании k, потому что с увеличением k приближается к достоверности вероятность, что извлечение происходит из первой урны.

2) Дана урна с N шарами, из которых M белых; после каждого извлечения на место вынутого шара кладется черный шар. Вычислить дисперсию B_n числа m вынутых белых шаров при n извлечениях.

Отв. $B_n = M\left[\left(1-\frac{1}{N}\right)^n - \left(1-\frac{2}{N}\right)^n\right] + M^2\left[\left(1-\frac{2}{N}\right)^n - \left(1-\frac{1}{N}\right)^{2n}\right]$.

Для получения этой формулы можно воспользоваться методом конечных разностей (стр. 136). Находим сначала

$$\text{М. О. } m = M\left[1-\left(1-\frac{1}{N}\right)^n\right];$$

замечаем затем, что

$$B_n = \text{М. О. } \left\{m - \left[1-\left(1-\frac{1}{N}\right)^n\right]M\right\}^2$$

удовлетворяет уравнению в конечных разностях

$$B_{n+1} = \left(1-\frac{2}{N}\right)B_n + \frac{M}{N}\left(1-\frac{1}{N}\right)^n\left[1-\frac{M}{N}\left(1-\frac{1}{N}\right)^n\right],$$

искомым решением которого является написанное выше выражение.

Коэффициент дисперсии

$$E = \frac{\left(1-\frac{1}{N}\right)^n - \left(1-\frac{2}{N}\right)^n + M\left[\left(1-\frac{2}{N}\right)^n - \left(1-\frac{1}{N}\right)^{2n}\right]}{1-\left(1-\frac{1}{N}\right)^n - \frac{M}{n}\left[1-\left(1-\frac{1}{N}\right)^n\right]^2}.$$

При больших значениях N, полагая $n = \alpha N$, получаем (приближенно)

$$E = \frac{e^{-\alpha} - e^{-2\alpha}\left(1+\frac{\alpha M}{N}\right)}{\left(1-e^{-\alpha}\right)\left(1-\frac{M}{n}e^{-\alpha}\right)} = e^{-\alpha}\left[1+\frac{\frac{M}{N}e^{-\alpha}\left(\frac{1-e^{-\alpha}}{\alpha}-\alpha\right)}{\left(1-e^{-\alpha}\right)\left(1-\frac{M}{\alpha N}e^{-\alpha}\right)}\right].$$

Таким образом, при α малом, E близко к единице; но при больших α E быстро убывает, как $e^{-\alpha}$, естественно приближаясь к 0, когда число вынутых белых шаров настолько велико, что дальнейшее появление их становится практически невозможным.

ГЛАВА ЧЕТВЕРТАЯ

СТАТИСТИЧЕСКИЕ ВЕРОЯТНОСТИ, СРЕДНИЕ ВЕЛИЧИНЫ И КОЭФФИЦИЕНТ ДИСПЕРСИИ

1. Предположим, что рассматривается некоторая весьма обширная или неограниченная совокупность объектов, из которых одни обладают данным признаком A, а другие этим признаком не обладают; мы можем также смотреть на совокупность этих объектов как на совокупность опытов, при реализации которых данное событие A иногда наступает, а иногда не наступает. Разумеется, этого вообще недостаточно для того, чтобы утверждать, что факту или признаку A можно приписать определенную, характеризующую всю совокупность опытов или объектов, вероятность. Только в том случае, когда постановка опытов такова, что неизбежное различие двух конкретных опытов, наблюдаемое нами, не может влиять на различие их исходов, проистекающее от не подлежащих учету причин, можно принять, что вероятность A во всех опытах одинакова. Это имело бы место, например, если бы событие A заключалось в появлении белого шара из урны с неизменяющимся составом шаров, при условии, что, вкладывая обратно после каждого извлечения вынутый шар, мы тщательно перемешиваем все шары. В большинстве практических примеров требуемое условие не бывает соблюдено; тем не менее, произведя большое число n опытов, пытаются охарактеризовать всю совокупность отношением $\frac{m}{n}$ числа m появившихся событий к общему числу n опытов, или, что то же самое, отношением числа объектов, обладающих признаком A к общему числу наблюденных объектов. Это отношение $\frac{m}{n}$ называется статистической, или эмпирической, вероятностью события или признака A.

Значение этого понятия и пределы законности его применения выясняются следующей теоремой, которая является некоторым образом обратной теореме Бернулли.

Если известно, что в рассматриваемой совокупности независимых опытов появление события A имеет одну и ту же вероятность p, то после осуществления достаточно большого числа n опытов, при которых событие A произошло m раз, можно утверждать с вероятностью, сколь угодно близкой к достоверности, что имеет место неравенство

$$\left|p-\frac{m}{n}\right|\leqslant\varepsilon, \tag{137}$$

где ε—произвольно малая величина, лишь бы a priori такое неравенство (при всяком достаточно малом $\varepsilon>0$) было возможно.

Наше основное допущение состоит в том, что постоянная вероятность p существует. Пусть отрезок теоретически возможных значений p от 0 до 1 разделен был первоначально на интервалы так, чтобы длина каждого интервала не превышала $\frac{\varepsilon}{2}$; тогда, обозначая через M_0 один из концов интервала, где находится $\frac{m}{n}$, будем иметь

$$\left|\frac{m}{n}-M_0\right|<\frac{\varepsilon}{2}. \tag{138}$$

Мы допускаем, согласно условию, что неравенство

$$|p-M_0|<\frac{\varepsilon}{4} \tag{139}$$

a priori возможно [1]) и имеет вероятность [2]) $f(\varepsilon)>0$; но так как для осуществления неравенства (138) достаточно осуществление (139) совместно с неравенством

$$\left|p-\frac{m}{n}\right|<\frac{\varepsilon}{4},$$

[1]) Это допущение может показаться настолько очевидным, что о нем излишне упоминать, но нужно иметь в виду, что p могло бы при соответствующей постановке опытов допускать лишь ограниченное число значений: например, если в урне только 3 шара, то невозможно неравенство $\left|p-\frac{23}{50}\right|<\frac{1}{10}$, так как p может быть только $\frac{1}{3}$ или $\frac{2}{3}$ (кроме 0 и 1).

[2]) Наиболее естественно допущение, что $f(\varepsilon)=\frac{\varepsilon}{2}$, но в этом ограничении нет необходимости.

которое, согласно (70), имеет вероятность бо́льшую, чем $1-\frac{16pq}{\varepsilon^2 n}$, то вероятность a priori неравенства (138) больше, чем $f(\varepsilon)\left[1-\frac{16pq}{\varepsilon^2 n}\right] \geqslant$ $\geqslant f(\varepsilon)\left[1-\frac{4}{\varepsilon^2 n}\right] > \frac{1}{2} f(\varepsilon)$, если n достаточно велико.

С другой стороны, вероятность осуществления неравенства (138) и нарушения (137) меньше, чем $\frac{pq}{\varepsilon^2 n} \leqslant \frac{1}{4\varepsilon^2 n}$, но вероятность этого совмещения равна вероятности a priori неравенства (138), умноженной на вероятность Q того, что

$$\left| p-\frac{m}{n} \right| > \varepsilon$$

после осуществления неравенства (138); поэтому

$$\frac{1}{2} Qf(\varepsilon) < \frac{1}{4\varepsilon^2 n},$$

и, следовательно, при данном ε,

$$Q < \frac{1}{2f(\varepsilon)\varepsilon^2 n}$$

стремится к 0 с возрастанием n.

Поэтому, после того как известно, что событие произошло m раз, вероятность P неравенства (137) приближается к пределу 1 с возрастанием n.

Пользуясь неравенством (106) вместо (70), можно было бы таким же образом установить, что

$$Q < \frac{4}{f(\varepsilon)} e^{-n\varepsilon^2}. \tag{140}$$

Полагая, например, $f(\varepsilon)=\frac{\varepsilon}{2}$, находим, что если при $n=600$ $m=300$, то вероятность, что $\left| p-\frac{1}{2} \right| < \frac{1}{10}$, больше, чем $1-\frac{8}{\varepsilon} e^{-n\varepsilon^2} = 1-80e^{-6} \neq 0{,}8$, а при $n=6\,000$ и $m=3\,000$, эта вероятность больше, чем $1-80e^{-60} > 1-\frac{1}{10^{24}}$. Если бы было

известно, например, что в урне 6 шаров, то мы могли бы с указанной степенью вероятности утверждать, что число белых шаров в ней равно 3, т.-е. $p=\frac{1}{2}$.

Это же заключение осталось бы в силе и в том случае, если бы при 6 000 извлечений было извлечено не менее 2 700 и не более 3 300 белых шаров, так как при $m=2\,700$ указанную вероятность, граничащую с достоверностью, имело бы неравенство [1])

$$\left|p-\frac{27}{60}\right|<\frac{1}{10},$$

т.-е.

$$\frac{21}{60}<p<\frac{33}{60},$$

но из всех возможных значений p этому неравенству удовлетворяет лишь $p=\frac{1}{2}$.

Доказанная теорема устанавливает таким образом, что вообще статистическая вероятность $\frac{m}{n}$ дает приближенное значение неизвестной (постоянной, по предположению) вероятности p существования признака A у объектов данной совокупности, точность которого безгранично возрастает с возрастанием n.

2. Аналогичный результат можем получить для обоснования теоретического значения средней арифметической из ряда наблюденных величин $x_1, x_2, \ldots, x_n$.

Основное допущение, которое здесь опять нужно сделать, это предположение, что все рассматриваемые величины $x_1, x_2, \ldots, x_n$ имеют одно и то же математическое ожидание a. Кроме того необходимо, чтобы к величинам $x_1, x_2, \ldots, x_n$ был применим закон больших чисел; для определенности вводим ограничение Чебышева, что величины x_i независимы и имеют ограниченную дисперсию $\beta_i=\text{М. О. }(x_i-a)^2<L$.

В таком случае, после того как определятся величины $x_1, x_2, \ldots, x_n$, можно утверждать с вероят-

[1]) Величина M_0, входящая в данное выше доказательство, может быть взята равной $\frac{29}{60}$.

ностью, приближающейся при возрастании n к достоверности, что неизвестное нам значение a математического ожидания, общего, по предположению, всем величинам x_i, удовлетворяет неравенству

$$\left| a - \frac{x_1 + x_2 + \ldots + x_n}{n} \right| < \varepsilon,$$

где ε — данная произвольно малая величина (лишь бы только a priori возможность такого неравенства не была исключена).

Доказательство совершенно аналогично данному выше для случая, когда a представляет собой неизвестную вероятность, и поэтому нет надобности его воспроизводить.

Таким образом, если мы знаем, что все наблюдаемые нами значения $x_1, x_2, \ldots, x_n$ имеют одно и то же математическое ожидание a, то средняя арифметическая $\frac{x_1 + x_2 + \ldots + x_n}{n} = X$ дает приближенное значение a, точность которого безгранично возрастает с увеличением n. Более полно вопрос о погрешности этого приближения будет разобран в дальнейшем, а пока полученные результаты могут служить нам достаточным основанием для того, чтобы пользоваться практически приближенными равенствами

$$a \neq X \text{ и } p \neq \frac{m}{n}.$$

Следует заметить, что, если вместо простой средней арифметической X взять взвешенную среднюю $X_\rho = \frac{\rho_1 x_1 + \rho_2 x_2 + \ldots + \rho_n x_n}{\rho_1 + \rho_2 + \ldots + \rho_n}$, где $\rho_1, \rho_2, \ldots, \rho_n$ — какие-нибудь данные, ограниченные сверху и снизу ($\lambda < \rho_i < \mu$ при всяком i) положительные числа, то, применяя аналогичное рассуждение, можно убедиться в том, что и X_ρ с возрастанием n сколь угодно приближается к a, так что справедливо также приближенное равенство $a \neq X_\rho$.

Напротив, пользование другими средними для приближенного нахождения $a =$ М. О. x_i вообще недопустимо; и если мы возьмем, например, геометрическую среднюю $\xi = \sqrt[n]{x_1 x_2 \ldots x_n}$, то понятно, что ξ не может служить приближенным значением для a, так как достаточно, например, чтобы хоть одна из величин x_i была равна нулю, чтобы, независимо от всех прочих, оказалось и $\xi = 0$.

Правило средней арифметической обычно применяется, когда, на основании ряда наблюдений неизвестной величины, хотят получить возможно точное ее значение. Предыдущие замечания вполне оправдывают этот способ, если неизвестной величиной является математическое ожидание a наблюденных величин, при чем никакого другого физического смысла величина a не имеет. Но тем же способом пользуются часто, когда неизвестной является некоторая физическая величина x, которая, вообще говоря, отлична от математического ожидания a наблюденных величин $x_1, x_2, \ldots, x_n$. Предположим, например, что наблюдатель производит определенный физический опыт для определения скорости звука (или света) в пустоте. Повторяя тот же опыт при одинаковых условиях, он получает, благодаря сложности опыта, не вполне одинаковые численные результаты $x_1, x_2, \ldots, x_n$ для искомой скорости x. Беря среднюю арифметическую $X = \frac{x_1 + x_2 + \ldots + x_n}{n}$, он рассчитывает получить, произведя достаточно большое число опытов, сколь угодно точное значение скорости x. Правильно ли это?

В действительности, средняя арифметическая X даст ему лишь приближенное значение величины $a = \text{М. О.}\, x_i$, и если сама схема его опыта была неудовлетворительна или приборы плохо проверены (например, измерительная линейка вместо 1 *м* равна 0,999 *мм*), то, как бы точно наш наблюдатель ни нашел значение a, у него нет оснований считать, что X или a соответствуют истинному значению скорости звука, которая может быть наблюдаема в других самых разнообразных опытах. Основное допущение, которое должно было бы оправдать применение способа средней арифметической к физическим измерениям такого рода, состоит в предположении, что неизвестная величина $a = \text{М. О.}\, x_i$ или, другими словами, что измерение (или вычисление) производится без систематической ошибки.

Не останавливаясь более подробно на этом вопросе, который не имеет прямого отношения к интересующей нас теме, заметим еще, что и при отсутствии систематической ошибки невозможно при помощи измерительных приборов, обладающих ограниченною точностью, получить благодаря многократному измерению значение искомой величины с произвольно большою точностью. Если, например, я возьму из таблицы пятизначных логарифмов мантиссу какого-нибудь числа, заканчивающуюся цифрой 5, и предложу

100 слушателям моей аудитории отбросить последнюю цифру, по своему усмотрению сохранивши или увеличивши на 1 единицу, согласно правилам приближенного вычисления, 4-ю цифру мантиссы, то было бы очевидною нелепостью предполагать, что средняя арифметическая полученных таким образом 100 мантисс, которая выразится шестизначной дробью, имеет больше шансов соответствовать действительному значению шестизначного логарифма рассматриваемого числа, чем если я сам наудачу припишу к пятизначной мантиссе шестую цифру: полученный результат будет зависеть исключительно от фантазии моих слушателей, которая ни в какой мере не направляется истинным значением логарифма, и мы получили бы тот же результат, если бы предложенное число было взято из любой другой таблицы.

Вообще, способ средней арифметической, в силу закона больших чисел, сглаживает уклонения отдельных наблюдений x_i от их математического ожидания a, но если самим наблюдениям x_i присуща некоторая неопределенность $\pm\delta$, то, согласно определению математического ожидания, той же неопределенностью обладает также и a; поэтому, поскольку при составлении средней арифметической мы можем (независимо от каких бы то ни было объективных оснований) заменять x_i через $x_i+\theta_i\delta$ $(-1<\theta_i<1)$, задача определения a (или физической величины x, приравниваемой $a=\text{М. О. } x_i$) с погрешностью меньшею, чем δ, лишена смысла.

3. В силу установленных выше положений вычисление статистической вероятности $\frac{m}{n}$ приобретает особое значение, когда мы знаем, что опыты рассматриваемой совокупности независимы и при каждом из них появление факта (признака) A имеет одну и ту же вероятность p. Важно поэтому указать прием для проверки того, имеет ли место последнее предположение.

В данном случае, как и во всех вопросах теоретической схематизации явлений природы, опыт обычно не доказывает, что та или иная гипотеза справедлива, но может лишь доказать ее неправильность, если следствия, вытекающие из теории, не подтверждаются в действительности. Правда, поскольку утверждения теории вероятностей не обладают в применении к практике абсолютною катего-

ричностью, опыт часто дает лишь уклончивый ответ на вопрос о том, совместима ли наша гипотеза с наблюдениями или нет, но если осуществляются факты, которые, согласно допущенной гипотезе, имеют очень малую вероятность, и, напротив, очень вероятные факты не происходят, то это является более или менее определенным указанием того, что гипотезу нужно отвергнуть. До тех же пор, пока особо резких несоответствий между теорией и действительностью не наблюдается, схемы теории вероятностей, обладающие достаточной простотой, могут служить рабочими гипотезами, дающими возможность при помощи нескольких численных показателей характеризовать более или менее точно данную статистическую совокупность.

Итак, остановимся на важнейших следствиях, вытекающих из предположения, что при осуществлении весьма большого числа N независимых опытов вероятность факта A все время была равна p.

Разобьем наши N опытов на s равных групп (s не должно быть мало, примерно не менее 15—20) по n опытов в каждой группе, так что $N=sn$. Положим, что в каждой группе отмечено, соответственно $m_1, m_2, \ldots, m_s$ появлений события A ($m_1+m_2+\ldots+m_s=M$). В таком случае для каждой группы дисперсия β_i числа m_i появлений факта A равна

$$\beta_i = \text{М. О. } (m_i - np)^2 = npq.$$

Поэтому, полагая

$$x_i = \frac{(m_i - np)^2}{npq},$$

видим, что М. О. $x_i = 1$, и следовательно, допуская, что к величинам x_i применим закон больших чисел, можем утверждать, что при s достаточно большом вероятность неравенства

$$\left| \frac{(m_1 - np)^2}{snpq} + \frac{(m_2 - np)^2}{snpq} + \ldots + \frac{(m_s - np)^2}{snpq} - 1 \right| < \varepsilon,$$

т.-е.

$$\left| \frac{\sum_{i=1}^{i=s} (m_i - np)^2}{Npq} - 1 \right| < \varepsilon, \qquad (141)$$

при произвольно малом ε, сколь угодно близка к достоверности,

Для того чтобы проверить, что закон больших чисел в данном случае применим, достаточно убедиться, что М. О. $(x_i - 1)^2$ остается ограниченным.

С этой целью вычисляем М. О. $x_i^2 = \frac{\text{М. О. } (m_i - np)^4}{n^2p^2q^2}$; пользуясь формулами, данными в упражнении в конце главы VI, ч. II, находим

$$\text{М. О. } x_i^2 = 3 + \frac{1 - 6pq}{npq}, \tag{142}$$

откуда

$$\text{М. О. } (x_i - 1)^2 = 2 + \frac{1}{npq}(1 - 6pq) \neq 2$$

так как n обычно довольно велико).

Применяя теорему Чебышева, видим, что вероятность неравенства

$$\left| \frac{\sum (m_i - np)^2}{Npq} - 1 \right| < \varepsilon \tag{141}$$

больше, чем $1 - \frac{2}{s\varepsilon^2}$; если $\varepsilon = 1$, то эта вероятность больше, чем $1 - \frac{2}{s}$, и, например, при $s = 20$ она больше 0,9.

Можно доказать [1]), что, если $\varepsilon \leqslant 1$, то применимо неравенство (106); поэтому вероятность, что

$$D = \sum \frac{(m_i - np)^2}{Npq} > 2 \tag{143}$$

---

[1]) Полагая n достаточно большим, убеждаемся, что $|\text{М. О. } (x_i - 1)^h| < 2^h \cdot h!$, так что в неравенстве (105) можно положить $H = 2$. Для этой цели, пользуясь результатами следующей главы, достаточно заметить, что

$$\text{М. О. } (x_i - 1)^h = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (z^2 - 1)^h e^{-\frac{z^2}{2}} dz =$$

$$= \frac{2(h-1)}{\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} (z^2 - 1)^{h-1} e^{-\frac{z^2}{2}} dz + \int_{-\infty}^{\infty} (z^2 - 1)^{h-2} e^{-\frac{z^2}{2}} dz \right] =$$

$$= 2(h-1)[\text{М. О. } (x_i - 1)^{h-1} + \text{М. О. } (x_i - 1)^{h-2}].$$

Поэтому допуская, что

$|\text{М. О. } (x_i - 1)^{h-1}| < 2^{h-1}(h-1)!$ и $|\text{М. О. } (x_i - 1)^{h-2}| < 2^{h-2}(h-2)!$, находим также, что

$$|\text{М. О. } (x_i - 1)^h| < 2^{h-1}(h-1)![2(h-1) + 1] < 2^h \cdot h!;$$

но так как для $h = 2$ требуемое неравенство осуществлено, то оно осуществляется и для всякого h. Для малых значений n неравенство (105) также осуществлено для значения H близкого к 2.

меньше, чем $e^{-\frac{s}{8}}$ $\left(\text{так как } \varepsilon = 1 \text{ соответствует } t^2 = \frac{s}{4 \cdot 2}\right)$; следовательно, например, при $s = 40$, вероятность (143) меньше, чем $e^{-5} \neq 0{,}006\,75$; для $s = 100$ — меньше $e^{-12{,}5} \neq 0{,}000\,001\,5$.

Рассмотрим пример, численные данные которого мы заимствуем у Charlier „Grundzüge der Mathematischen Statistik".

Было произведено 10 000 выниманий карты из полной колоды карт с возвращением обратно извлеченной карты, так что вероятность появления карты черной масти $p = \frac{1}{2}$ во всех опытах; все опыты были разделены на $s = 20$ групп по 500 опытов в каждой; оказалось, что

	$m_i - 250$	$(m_i - 250)^2$
$m_1 = 252$	2	4
$m_2 = 235$	— 15	225
$m_3 = 248$	— 2	4
$m_4 = 271$	21	441
$m_5 = 260$	10	100
$m_6 = 246$	— 4	16
$m_7 = 228$	— 22	484
$m_8 = 229$	— 21	441
$m_9 = 234$	— 16	256
$m_{10} = 250$	0	0
$m_{11} = 271$	21	441
$m_{12} = 234$	— 16	256
$m_{13} = 258$	8	64
$m_{14} = 233$	— 17	289
$m_{15} = 273$	23	529
$m_{16} = 244$	— 6	36
$m_{17} = 249$	— 1	1
$m_{18} = 241$	— 9	81
$m_{19} = 231$	— 19	361
$m_{20} = 246$	— 4	16

$$\sum (m_i - np)^2 = 4\,045; \quad Npq = 10\,000 \cdot \frac{1}{4} = 2\,500;$$

$$D = \frac{4\,045}{2\,500} = 1{,}618.$$

Величину

$$D = \sum \frac{(m_i - np)^2}{Npq} = \frac{1}{s} \sum \frac{(m_i - np)^2}{npq} \tag{144}$$

мы будем называть эмпирическим коэффициентом дисперсии данной группы опытов, соответствующим предположению, что во всех опытах вероятность равна числу p. Теоретическое значение D в случае, если вероятность p постоянна, равно 1; если бы s неограниченно возрастало, то, согласно закону больших чисел, D должно было бы мало отличаться от своего теоретического значения 1, которое является теоретическим коэффициентом дисперсии m в каждой группе.

Действительно, располагая предыдущие 10 000 извлечений карты в 1 000 групп по 10 в каждой, Charlier находит, что

$m = 0$	в	3	группах
$m = 1$	„	10	„
$m = 2$	„	43	„
$m = 3$	„	116	„
$m = 4$	„	221	„
$m = 5$	„	247	„
$m = 6$	„	202	„
$m = 7$	„	115	„
$m = 8$	„	34	„
$m = 9$	„	9	„
$m = 10$	„	0	„

$$\sum (m_i - np)^2 = 3 \cdot 25 + 10 \cdot 16 + 43 \cdot 9 + 116 \cdot 4 + 221 \cdot 1 + 202 \cdot 1 + 115 \cdot 4 + 34 \cdot 9 + 9 \cdot 16 = 2\,419.$$

$$D = \frac{\Sigma(m_i - np)^2}{N\ pq} = \frac{2\,419}{2\,500} = 0{,}9676.$$

Таким образом указанные численные результаты не противоречат сделанной a priori гипотезе, что появление черной карты имело вероятность равную $\frac{1}{2}$ во всех опытах и что эти опыты были независимы между собой.

Для вычислений удобнее разбивать всю совокупность опытов на равные группы, но тот же прием может быть приложен и в том случае, если статистические данные естественно оказались распределенными между неравными группами, так что в каждой из s групп имеется соответственно по $n_1, n_2, \ldots, n_s$ опытов (или объектов).

В таком случае опять составляем для каждой группы

$$x_i = \frac{(m_i - n_i p)^2}{n_i\, pq},$$

так что теоретический коэффициент дисперсии m_i, т.-е. М. О. $x_i = 1$; поэтому, при s весьма большом, вероятность неравенства вида

$$\left| \sum \frac{x_i}{s} - 1 \right| < \varepsilon \qquad (145)$$

стремится к достоверности. $D=\frac{1}{s}\sum\frac{(m_i-n_ip)^2}{n_ipq}$ является, как и раньше, эмпирическим коэффициентом дисперсии, теоретическое значение которого равно 1. Заметим, что школа Лексиса называет коэффициентом дисперсии не D, а $\sqrt{D}$. Будем ли мы придерживаться терминологии Лексиса или принятой нами терминологии (Маркова), во всяком случае для того, чтобы гипотеза о постоянстве вероятности была допустима, нужно, чтобы эмпирический коэффициент дисперсии D (или $\sqrt{D}$) мало отличался от 1 (коэффициент дисперсии А. А. Маркова, очевидно, более чувствителен, чем коэффициент Лексиса).

4. В примере, который был рассмотрен выше, a priori была известна вероятность $p\left(=\frac{1}{2}\right)$. Но в действительности у нас часто нет теоретических оснований, чтобы a priori приписать определенную вероятность рассматриваемому факту или признаку A. В таком случае мы принимаем за предполагаемую вероятность наблюденную статистическую вероятность $\frac{M}{N}$, где $N=n_1+n_2+\ldots+n_s$ есть число всех опытов, а $M=m_1+m_2+\ldots+m_s$ — число всех появлений нашего события A, и проверяем, насколько эта гипотеза допустима.

Таким образом, в случае равных групп ($N=ns$), следует положить

$$p=\frac{M}{N}, \tag{146}$$

откуда

$$x_i=\frac{\left(m_i-n\frac{M}{N}\right)^2}{n\frac{M}{N}\left(1-\frac{M}{N}\right)}=\frac{\left(m_i-\frac{M}{s}\right)^2}{\frac{M}{s}\left(1-\frac{M}{N}\right)}$$

и

$$D=\frac{1}{s}\sum x_i=\frac{\sum\left(m_i-\frac{M}{s}\right)^2}{M\left(1-\frac{M}{N}\right)}. \tag{147}$$

Эмпирический коэффициент дисперсии D попрежнему должен быть близок к 1 для больших значений s.

Как было указано выше, практически $D < 2$, при значениях s, не превышающих несколько десятков, можно считать удовлетворительно согласующимся с предположением постоянной вероятности; если же s достигает 100 и больше, то неравенство (106) или формула Лапласа (Гаусса), о которой речь будет впереди, позволяет более точно установить допустимые границы уклонения эмпирического коэффициента дисперсии от 1.

Если группы не равны, то

$$x_i = \frac{\left(m_i - n_i \frac{M}{N}\right)^2}{n_i \frac{M}{N}\left(1 - \frac{M}{N}\right)} \quad \text{и} \quad D = \frac{\sum \frac{1}{n_i}\left(m_i - n_i \frac{M}{N}\right)^2}{\frac{sM}{N}\left(1 - \frac{M}{N}\right)}. \qquad (147\,\text{bis})$$

Однако, во избежание сложных вычислений, если группы, соответствующие, например, различным областям страны, не очень различны по размерам, предпочтительно сделать равные механические выборки из более обширных групп так, чтобы все элементы группы имели равные шансы быть выбранными. Сводя таким образом число новорожденных в 24 провинциях Швеции в мае 1883 года, Шарлье (Charlier) находит, что из 12 000 новорожденных 6 186 было мальчиков, так что $\frac{M}{N} = 0{,}516$; сумма квадратичных уклонений $\sum\left(m_i - \frac{M}{s}\right)^2 = \sum(m_i - 258)^2$ оказывается равной 4 600. Поэтому, по формуле (147) $D = \frac{4\,600}{12\,000 \cdot 0{,}516 \cdot 0{,}484} \neq 1{,}53$, и гипотеза постоянной вероятности довольно правдоподобна.

Если опыты рассматриваемой совокупности не независимы или же вероятность факта A не постоянна, то вообще теоретический коэффициент дисперсии не равен 1, а следовательно и эмпирическое его значение D более значительно уклоняется от 1. Нужно помнить, однако, что только при больших значениях s эмпирическое и теоретическое значения коэффициента дисперсии друг от друга мало отличаются.

5. Предположим, что в каждой группе[1]) вероятность p_i постоянна, но значения p_i в разных группах раз-

[1]) Мы ограничиваемся, для простоты, случаем, когда все группы численно равны $n_i = \frac{N}{s}$.

личны; тогда средняя вероятность для всех групп

$$P = \frac{p_1 + p_2 + \ldots + p_s}{s}.$$

Очевидно, по закону больших чисел, мы нашли бы, что для больших значений N отношение $\frac{M}{N} = \frac{m_1 + m_2 + \ldots + m_s}{ns}$ мало отличается от P; таким образом та же самая статистическая вероятность могла бы быть получена в данном предположении, как и в случае, когда все $p_i = P$. Однако рассматриваемый теперь случай существенно отличается от предыдущего тем, что ему соответствует коэффициент дисперсии D больший, чем 1, или, как принято говорить, дисперсия в такой серии опытов сверхнормальна.

Действительно, мы получили бы нормальную дисперсию, т.-е. коэффициент дисперсии равный 1, если бы брали отдельно для каждой группы

$$\frac{\text{М. О. } (m_i - np_i)^2}{np_iq_i};$$

но если мы рассматриваем величины

$$x_i = \frac{(m_i - nP)^2}{nPQ},$$

то

$$\text{М. О. } x_i = \frac{\text{М. О. } [m_i - np_i - n(P - p_i)]^2}{nPQ} = \frac{np_iq_i + n^2(P - p_i)^2}{nPQ}$$

$$= \frac{p_iq_i + n(P - p_i)^2}{PQ}.$$

Поэтому

$$\sum \text{М. О. } x_i = \frac{\sum p_iq_i + n\sum (P - p_i)^2}{PQ} =$$

$$= \frac{\sum (p_i - p_i^2) + \sum (P^2 - 2Pp_i + p_i^2) + (n-1)\sum (P - p_i)^2}{PQ} =$$

$$= \frac{sP + sP^2 - 2sP^2 + (n-1)\sum (P - p_i)^2}{PQ} =$$

$$= s + \frac{n-1}{PQ}\sum (P - p_i)^2.$$

Отсюда теоретическое значение коэффициента дисперсии

$$E = \sum \text{М. О.} \frac{x_i}{s} = \text{М. О. } D = 1 + \frac{n-1}{PQs} \sum (P - p_i)^2. \quad (148)$$

Таким образом, обозначая среднее квадратичное уклонение отдельных вероятностей p_i от их общей средней P через δ, а именно, полагая $\delta = \frac{1}{s} \sum (P - p_i)^2$, находим

$$E = \text{М. О. } D = 1 + (n-1) \frac{\delta}{PQ}. \quad (148 \text{ bis})$$

Отсюда заключаем, что если n значительно, то коэффициент дисперсии становится большим и способен дать некоторые указания относительно величины δ.

Предположим, например, что произведено 100 групп опытов по 100 в каждой, при чем в 50 группах вероятность появления данного события $\frac{2}{5}$, и $\frac{3}{5}$ — в других 50. Здесь $P = \frac{1}{2}$, $\delta = \left(\frac{1}{10}\right)^2 = \frac{1}{100}$, поэтому $E = 1 + 0{,}99 \cdot 4 = 4{,}96$, и число $s = 100$ достаточно велико, чтобы и по значению эмпирического коэффициента дисперсии D (который, без сомнения, будет не менее 3) можно было видеть, что средняя вероятность $P = \frac{1}{2}$ не есть постоянная вероятность, общая всем опытам.

Отметим также случай независимых опытов с поднормальной дисперсией, когда $E < 1$, хотя на практике он имеет мало значения.

Предположим, что вероятности в различных опытах вообще различны, но средняя арифметическая вероятность в каждой группе одинакова, так что в различных n опытах каждой группы вероятности соответственно равны $p_1, p_2, \ldots, p_n$, и их средняя равна $P = \frac{p_1 + p_2 + \ldots + p_n}{n}$, которая служит таким образом общей средней вероятностью и для всей совокупности.

Тогда

$$\text{М. О. } x_i = \frac{\text{М. О. } (m_i - nP)^2}{nPQ} = \frac{\sum p_k q_k}{nPQ} =$$

$$= \frac{\sum (p_k - p_k^2) - \sum (P - p_k)^2 - nP^2 + \sum p_k^2}{nPQ} =$$

$$= \frac{nP - nP^2 - \sum (P - p_k)^2}{nPQ} = \frac{nPQ - \sum (P - p_k)^2}{nPQ} =$$

$$= 1 - \frac{\sum (P - p_k)^2}{nPQ}, \qquad (149)$$

откуда, полагая $\delta' = \frac{1}{n} \sum (P - p_k)^2$, находим

$$E = \frac{\sum \text{М. О. } x_i}{s} = 1 - \frac{\sum (P - p_k)^2}{nPQ} = 1 - \frac{\delta'}{PQ}. \qquad (150)$$

Здесь $E < 1$, но, вообще говоря, незначительно отличается от 1. Если в предыдущей совокупности мы соединим вместе по две группы различного характера, чтобы составить 50 одинаковых групп по 200 опытов (вместо прежних 100 групп по 100 опытов), так что в каждой группе теперь оказывается поровну опытов с вероятностями $\frac{2}{5}$ и $\frac{3}{5}$, то $P = \frac{1}{2}$, но $E = 1 - \frac{4}{100} = 0{,}96$, и весьма трудно было бы статистически отличить этот случай от случая постоянной вероятности.

В общем случае независимых опытов с различными вероятностями, когда в каждой группе средние арифметические вероятности $P_1, P_2, \ldots, P_s$ не равны и $P = \frac{\sum_1^s P_i}{s}$, при помощи аналогичных вычислений приходим к формуле

$$\text{М. О. } \sum \frac{x_i}{s} = 1 + (n-1)\frac{\delta}{PQ} - \frac{\delta'}{PQ}, \qquad (151)$$

где $x_i = \frac{(m_i - nP)^2}{nPQ}$; здесь δ есть попрежнему среднее квадратичное уклонение $\delta = \frac{\sum (P_i - P)^2}{s}$ (средней) вероятности каждой

группы от общей средней P, а δ' — среднее квадратичное уклонение каждой из вероятностей $p_k^{(h)}$ от средней P_h той группы, к которой $p_k^{(h)}$ принадлежит, т.-е. $\delta' = \frac{1}{N}\sum_k\sum_h\left(p_k^{(h)} - P_h\right)^2$.

В значении $E = 1 + (n-1)\frac{\delta}{PQ} - \frac{\delta'}{PQ}$ теоретического коэффициента дисперсии преобладающую роль обычно играет второй (положительный) член, так как, при сколько-нибудь значительном n, $(n-1)\delta > \delta'$. Это имеет место в том наиболее распространенном случае, когда разбивка на группы приводит к такому распределению объектов, что в одних группах скопляются объекты, в среднем более предрасположенные обладать признаком A, чем в других ($P_k \gtrless P_h$). Поэтому, когда коэффициент дисперсии больше 1 (практически > 2), то это свидетельствует о некотором симптоматическом (а не чисто случайном) различии между группами.

6. Если мы заранее предполагаем, каковы должны быть вероятности $p_1, p_2, \ldots, p_s$ в каждой группе, то способ дисперсии позволяет сделать проверку этого, так как каждая из величин

$$x_i = \frac{(m_i - np_i)^2}{np_iq_i}$$

имеет в этом случае математическое ожидание равное 1 и, следовательно, по закону больших чисел средний эмпирический коэффициент дисперсии

$$D = \frac{1}{s}\sum\frac{(m_i - np_i)^2}{np_iq_i}$$

должен с возрастанием s приближаться к 1.

Пусть, например, вся совокупность N объектов распадается на группы, обладающие признаками $A_1, A_2, \ldots, A_s$, и мы, по тем или иным соображениям, полагаем, что вероятность данному объекту обладать признаком A_i равна p_i $\left(\sum p_i = 1\right)$. Допустим, что фактически в каждой группе оказалось $m_1, m_2, \ldots, m_s$ объектов; в таком случае $x_i = \frac{(m_i - Np_i)^2}{Np_iq_i}$ имеет математическое ожидание равное 1.

Поэтому средний коэффициент дисперсии

$$D = \frac{1}{s} \sum x_i \qquad (144\ \text{bis})$$

должен и в данном случае приближаться к единице с возрастанием s, если только закон больших чисел применим к величинам x_i.

Чтобы убедиться в применимости закона больших чисел к величинам x_i, которые теперь не вполне независимы, так как $\sum m_i = N$, необходимо вычислить

$$B = \text{М. О.} \left(\sum x_i - s\right)^2$$

и убедиться, что $\frac{B}{s^2}$ стремится к 0 с возрастанием s.

Для этой цели замечаем [1]), что (при $i \gtrless k$)

$$\text{М. О. } (m_i - Np_i)^2 (m_k - Np_k)^2 =$$
$$= N^2 [3p_i^2 p_k^2 + p_i p_k (1 - p_i - p_k)] - Np_i p_k (4p_i p_k + 2q_i q_k - 1).$$

Поэтому

$$\text{М. О. } [(x_i - 1)(x_k - 1)] = \frac{3p_i p_k + 1 - p_i - p_k}{q_i q_k} - 1 +$$
$$+ \frac{1 - 4p_i p_k - 2q_i q_k}{Nq_i q_k}, \qquad (152)$$

и, пренебрегая последним членом ввиду его малости при больших значениях N, получаем приближенно

$$\text{М. О. } (x_i - 1)(x_k - 1) \neq \frac{2p_i p_k}{q_i q_k}. \qquad (152\ \text{bis})$$

Отсюда, принимая во внимание (142) и (152 bis),

$$B \neq 2s + 4 \sum_{i,k} \frac{p_i p_k}{q_i q_k} < 2 \left[s + \left(\sum_i \frac{p_i}{q_i} \right)^2 \right] < 4s,$$

если предположить, что $q_i > \frac{1}{\sqrt{s}}$. Следовательно, пред. $\frac{B}{s^2} = 0$.

---

[1]) Я не привожу подробно вычислений, которые читатель может воспроизвести в виде упражнения.

В таблице функции Лапласа $F(z)$, имеющейся в конце книги А. А. Маркова „Исчисление вероятностей", мною рассмотрены значения z от 0,001 до 0,500 и отмечена 6-я десятичная цифра функции $F(z)$; оказалось, что эта цифра имела значения i в числе m_i случаев, указанных в таблице

$i=$	0	1	2	3	4	5	6	7	8	9
$m=$	60	46	50	60	39	44	46	51	57	47
$Np=$	50	50	50	50	50	50	50	50	50	50
$(m_i-Np)^2=$	100	16	0	100	121	36	16	1	49	9

Таким образом средний эмпирический коэффициент дисперсии

$$D=\frac{1}{10}\frac{\sum(m_i-50)^2}{500\cdot\frac{9}{10}\cdot\frac{1}{10}}=\frac{448}{450}\neq 0{,}996$$

настолько близок к 1, что гипотеза $p_i=\frac{1}{10}$ может считаться вполне удовлетворительной.

7. До сих пор мы полагали, что все опыты (объекты) независимы. В некоторых случаях, как мы видели в главе III, опыты могут быть зависимы, и для интерпретации действительных явлений целесообразно бывает ввести ту или иную гипотезу о характере этой зависимости. Мы упомянем здесь лишь об опытах, образующих, по терминологии А. А. Маркова, простую цепь. При разбивке совокупности из $N=ns$ опытов на равные группы по n элементов мы нашли для $x_i=\frac{(m_i-nP)^2}{nPQ}$ (полагая n достаточно большим)

$$\text{М. О. } x_i=\frac{1+\alpha-\beta}{1+\beta-\alpha}=\frac{1+\rho}{1-\rho}, \qquad (134\text{ bis})$$

где $\alpha-\beta=\rho$ — коэффициент корреляции между двумя смежными событиями цепи. Таким образом теоретический коэффициент дисперсии

$$D'=\frac{1}{s}\sum x_i=\frac{1+\rho}{1-\rho}$$

будет больше или меньше 1 (сверхнормальная или поднормальная дисперсия) в зависимости от знака коэффициента корреляции ρ, и при $\rho=0$ дисперсия становится нормальной (опыты тогда становятся независимыми).

Упражнения. 1) В вышеупомянутой таблице функции $F(z)$ рассматривается 6-я цифра, и событием A считается случай, когда эта цифра четная (0, 2, 4, 6, 8). 2 500 чисел от 0,000 до 2,499 этой таблицы распределены в 25 групп по 100 в каждой; следующая таблица показывает, в скольких группах событие A повторилось m_i раз:

$m_i =$	41	45	46	47	48	49	50	51	52	53	54	55	56	57	58
Число групп	1	3	2	2	4	3	1	0	2	1	2	2	0	0	2

Нужно определить коэффициент дисперсии при условии, что $p = \frac{1}{2}$, а также при условии, что a priori p неизвестно.

Отв. В первом случае $D = \frac{452}{625} = 0{,}7232$. Вычисление дает

$$\sum (m_i - 50)^2 = 452, \quad Npq = 2\,500 \cdot \frac{1}{4}.$$

Во втором случае статистическая вероятность $p = \frac{1}{2} + (2 \cdot 8 + 2 \cdot 5 + 2 \cdot 4 + 3 + 2 \cdot 2 - 3 - 4 \cdot 2 - 2 \cdot 3 - 2 \cdot 4 - 3 \cdot 5 - 9) \cdot \frac{1}{2500} = \frac{1}{2} - \frac{8}{2500} = 0{,}4968$, поэтому $\sum (m_i - 49{,}68)^2 = \sum (m_i - 50 + 0{,}32)^2 = 452 - 25(0{,}32)^2 = 449{,}44$; $Npq = 2\,500 \cdot 0{,}4968 \cdot 0{,}5032 = 624{,}9744$; $D = \frac{449{,}44}{624{,}9744} \neq 0{,}7194$.

2) Доказать, что $\sum_{i=1}^{i=s} \left(m_i - \frac{M}{s}\right)^2 = \sum \left(m_i - a\right)^2 - s\left(\frac{M}{s} - a\right)^2$, каково бы ни было a, где $M = m_1 + \cdots + m_s$.

3) Доказать, что если $B(x_i) = \sigma^2$, то М. О. $\left[x_i - \frac{x_1 + x_2 + \ldots + x_s}{s}\right]^2 = \frac{s-1}{s}\sigma^2$.

4) Статистика рождений в Германии за 10 лет (1882—1891) дает, соответственно, для последовательных лет следующие статистические вероятности для рождения мальчиков: 0,5150; 0,5147; 0,5151; 0,5148; 0,5146; 0,5140; 0,5146; 0,5142; 0,5150; 0,5151; при этом среднее число рождений в год $n = 1\,814$ тысяч (ввиду того, что число рождений в год колеблется не особенно значительно, от 1 750 до 1 903 тысяч, этих чисел можно не принимать в расчет, приравнивая $n = 1\,814\,000$ ежегодное число рождений). Требуется определить, совместимы ли данные с гипотезой, что вероятность рождения мальчика есть величина постоянная.

Отв. Да, так как полагая $p = 0{,}5147$, находим

$$D = \frac{\sum \left(\frac{m}{n} - p\right)^2}{10\,pq} n \neq 0{,}92.$$

5) Из таблицы логарифмов берут 300 рядов последовательных чисел по 50 в каждом ряду, и отмечается число h мантиссе имеющих на 7-м месте

цифру менее 5 (0, 1, 2, 3, 4). Полагая все цифры равновероятными, находим для вероятности, что $h=i$, значение $p_i = C_{50}^i \left(\frac{1}{2}\right)^{50}$.

В нижеследующей табличке даи результат наблюдений

Число рядов =	1	0	3	2	3	7	9	18	26	21	32	42	36	30	28	15	16	5	2	2	1	1	0
в которых $h=$	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36

Требуется определить средний эмпирический коэффициент дисперсии, соответствующий указанным теоретическим вероятностям.

Отв. $D \neq 0{,}8$ получаем на основании следующей таблицы:

i	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
$300C_{50}^i\left(\frac{1}{2}\right)^{50}$	0,2	0,6	1,4	2,6	4,8	8,1	12,5	17,9	23,6	28,8	32,5	33,7	32,5	28,8	23,6	17,9	12,5	8,1	4,8	2,6	1,4	0,6	0,2
$\lvert m_i - 300p_i \rvert$	0,8	0,6	1,6	0,6	1,8	1,1	3,5	0,1	2,4	7,8	0,5	8,3	3,5	1,2	4,4	2,9	3,5	3,1	2,8	0,6	0,4	0,4	0,2
$\frac{(m_i-300p_i)^2}{300p_iq_i}$	3,2	0,6	1,8	0,1	0,7	0,1	0,9	0,0	0,2	2,1	0,0	2,2	0,3	0,1	0,8	0,5	0,8	1,2	1,6	0,1	0,1	0,3	0,2

6) В биографическом словаре Poggendorff'a 450 страниц разбиты на 15 последовательных групп по 15 пар страниц в каждой. На каждой паре страниц вообще встречается от 0 до 17 имен. Каждую пару страниц, на которой менее 8 имен, характеризующуюся, таким образом, скоплением более длинных биографий, назовем событием A. Оказывается, что событие A произошло

	3	5	6	7	8	9	10	11 раз
в	1	1	1	2	4	2	1	3 группах,

т.-е. всего 121 раз, так что статистическая вероятность наступления события A на произвольно взятой паре страниц равна $p = \frac{121}{225}$. Допустимо ли предположение, что A имеет одну и ту же постоянную вероятность p во всех 15 группах?

Отв.

$$D = \frac{\sum (m-15p)^2}{225\,pq} =$$

$$= \frac{\left(\frac{76}{15}\right)^2 + \left(\frac{46}{15}\right)^2 + \left(\frac{31}{15}\right)^2 + 2\left(\frac{16}{15}\right)^2 + 4\left(\frac{1}{15}\right)^2 + 2\left(\frac{14}{15}\right)^2 + \left(\frac{29}{15}\right)^2 + 3\left(\frac{44}{15}\right)^2}{121 \cdot \frac{104}{225}} =$$

$$= \frac{16\,510}{12\,585} \neq 1{,}38,$$

поэтому статистический материал довольно хорошо согласуется с предположением постоянной вероятности и нет оснований считать, что в различных группах словаря отклонение m от $15p = \frac{121}{15}$ вызвано той или иной опре-

деленно меняющейся тенденцией составителей сокращать или удлинять биографии, а не случаем.

7) Рассматривается $(s+1)$ группа независимых опытов, по n в каждой, при чем в i-й $(i=0, 1, \ldots, s)$ группе вероятность события A равна $p_i = P - \frac{\rho}{s}\left(i - \frac{s}{2}\right)$; таким образом средняя вероятность для всех групп равна P, а групповые вероятности растут в арифметической прогрессии от $P - \frac{\rho}{2}$ до $P + \frac{\rho}{2}$; требуется вычислить теоретический коэффициент дисперсии E.

Отв. $E = 1 + \frac{n-1}{12PQ}\rho^2\frac{s+2}{s}$. Для вывода пользуемся формулой (148), замечая, что

$$\delta = \frac{\rho^2}{s^2}\cdot\frac{s\left(\frac{s}{2}+1\right)(s+1)}{6(s+1)} = \frac{\rho^2(s+2)}{12s}.$$

ЧАСТЬ ЧЕТВЕРТАЯ

ЗАКОН НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

ГЛАВА ПЕРВАЯ

ТЕОРЕМА ЛАПЛАСА

1. При выводе закона больших чисел мы пользовались неравенством [1]) Чебышева (или соответствующим его видоизменением), которое позволяло нам установить нижнюю границу для вероятности P неравенства вида

$$|x_1+x_2+\ldots+x_n-(a_1+a_2+\ldots+a_n)|<t\sqrt{B}. \quad (153)$$

Задача возможно точного вычисления вероятности P этого неравенства представляет большой теоретический и практический интерес. Если значения числа n невелики, то P зависит от n и от t, так что $P=P(n, t)$ существенным образом также зависит и от свойств рассматриваемых величин. Напротив, оказывается, что по мере возрастания n, когда закон больших чисел уравнивает случайные уклонения отдельных x_i от их математических ожиданий a_i, вероятность $P(n, t)$ в большинстве случаев стремится к определенному пределу $P(t)$, являющемуся функцией одной только величины t.

Рамки настоящего курса не позволяют нам исследовать в полном объеме рассматриваемую задачу; поэтому ограничимся пока ее решением для частного случая, соответствующего теореме Бернулли, когда неравенство (153) получает форму

$$|m-np|<t\sqrt{npq}; \quad (154)$$

мы в дальнейшем укажем лишь (без доказательства) соответствующие обобщения полученных результатов, которыми наука обязана главным образом трудам А. М. Ляпунова и А. А. Маркова.

Предельная теорема в той частной формулировке, которую мы ей теперь дадим, была впервые установлена Лапласом,

[1]) Стр. 148.

а потому и функцию $P(t)$, найденную им, называют **функцией Лапласа.**

Теорема Лапласа. **Если производится весьма большое число независимых опытов n, при каждом из которых вероятность наступления события A равна p, то вероятность $P(n, t_0, t_1)$, что число m появлений события A удовлетворит неравенству**

$$t_0\sqrt{npq} < m - np < t_1\sqrt{npq}, \tag{155}$$

имеет пределом

$$\frac{1}{\sqrt{2\pi}}\int_{t_0}^{t_1} e^{-\frac{t^2}{2}}\,dt, \tag{156}$$

когда n бесконечно возрастает.

Прежде чем приступить к доказательству, заметим, что из высказанной теоремы непосредственно вытекает, что вероятность неравенства (154), которое равнозначно

$$-t\sqrt{npq} < m - np < t\sqrt{npq}, \tag{154 bis}$$

имеет пределом

$$\frac{1}{\sqrt{2\pi}}\int_{-t}^{t} e^{-\frac{t^2}{2}}\,dt = \frac{2}{\sqrt{2\pi}}\int_{0}^{t} e^{-\frac{t^2}{2}}\,dt; \tag{156 bis}$$

поэтому функция Лапласа, которую мы обозначили выше через $P(t)$, равна

$$P(t) = \frac{2}{\sqrt{2\pi}}\int_{0}^{t} e^{-\frac{t^2}{2}}\,dt. \tag{157}$$

Переходим теперь к доказательству.

Вероятность $P(n, t_0, t_1)$ неравенства (155) вообще равна, согласно теореме сложения, сумме вероятностей, что m окажется равным какому-нибудь целому числу, удовлетворяющему условию (155), т.-е. $P(n, t_0, t_1) = \sum I_n^m$, где $I_n^m = C_n^m p^m q^{n-m}$, распространяя суммирование $\sum$ на все целые значения m, удовлетворяющие неравенству (155); все эти значения m можем представить в виде

$$m = np + t\sqrt{npq}, \tag{158}$$

при условии, что

$$t_0 < t < t_1.$$

Выражение $I_n^m = C_n^m p^m q^{n-m} = \frac{n!}{m!(n-m)!} p^m q^{n-m}$ может быть преобразовано при помощи формулы Стирлинга, на основании которой

$$x! = x^x e^{-x} \sqrt{2\pi x}\,(1+\alpha_x), \tag{159}$$

где α_x быстро стремится к 0 с возрастанием x.

Таким образом, применяя (159) к $n!$, $m!$ и $(n-m)!$, находим

$$I_n^m = \frac{n^n e^{-n} \sqrt{2\pi n}\,(1+\alpha_n)\, p^m q^{n-m}}{m^m e^{-m} \sqrt{2\pi m}\,(n-m)^{n-m} e^{-n+m} \sqrt{2\pi(n-m)}\,(1+\alpha_m)(1+\alpha_{n-m})},$$

и после очевидных сокращений

$$I_n^m = \frac{n^n p^m q^{n-m}}{m^m (n-m)^{n-m}} \cdot \sqrt{\frac{n}{2\pi m(n-m)}}\,(1+\alpha), \tag{160}$$

где $1+\alpha = \frac{1+\alpha_n}{(1+\alpha_m)(1+\alpha_{n-m})}$, так что α стремится к 0, когда m и $n-m$ (а следовательно и n) безгранично возрастают.

Остановим теперь наше внимание на первом множителе в полученном для I_n^m выражении и положим

$$\frac{1}{H} = \frac{n^n p^m q^{n-m}}{m^m (n-m)^{n-m}} = \left(\frac{np}{m}\right)^m \cdot \left(\frac{nq}{n-m}\right)^{n-m}.$$

Следовательно,

$$H = \left(\frac{m}{np}\right)^m \left(\frac{n-m}{nq}\right)^{n-m} = \left(1+t\sqrt{\frac{q}{np}}\right)^m \left(1-t\sqrt{\frac{p}{nq}}\right)^{n-m}, \tag{161}$$

благодаря равенству (158) и вытекающему из него

$$n-m = nq - t\sqrt{npq}. \tag{158 bis}$$

Отсюда

$$\log H = m \log\left(1+t\sqrt{\frac{q}{np}}\right) + (n-m)\log\left(1-t\sqrt{\frac{p}{nq}}\right) =$$

$$= (np + t\sqrt{npq}) \log\left(1+t\sqrt{\frac{q}{np}}\right) +$$

$$+ (nq - t\sqrt{npq}) \log\left(1-t\sqrt{\frac{p}{nq}}\right).$$

Но для любого x, заключенного между -1 и 1,

$$\log(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}-\cdots$$

Поэтому, придавая t определенное значение и полагая n достаточно большим, чтобы $\left|t\sqrt{\frac{q}{np}}\right|<\frac{1}{2}$ и $\left|t\sqrt{\frac{p}{nq}}\right|<\frac{1}{2}$, находим:

$$\log\left(1+t\sqrt{\frac{q}{np}}\right)=t\sqrt{\frac{q}{np}}-\frac{qt^2}{2np}+\rho,$$

$$\log\left(1-t\sqrt{\frac{p}{nq}}\right)=-t\sqrt{\frac{p}{nq}}-\frac{pt^2}{2nq}-\rho',$$

где

$$|\rho|\leqslant\frac{1}{3}\left|t\sqrt{\frac{q}{np}}\right|^3+\frac{1}{4}\left|t\sqrt{\frac{q}{np}}\right|^4+\cdots<\frac{A|t^3|}{n^{3/2}},\qquad(\rho')<\frac{A|t^3|}{n^{3/2}},$$

обозначая [1]) через A некоторое определенное положительное число не зависящее ни от n ни от t.

Таким образом

$$\log H=(np+t\sqrt{npq})\left(t\sqrt{\frac{q}{np}}-\frac{qt^2}{2np}+\rho\right)-$$

$$-(nq-t\sqrt{npq})\left(t\sqrt{\frac{p}{nq}}+\frac{pt^2}{2nq}+\rho'\right)=$$

$$=t\sqrt{npq}-\frac{qt^2}{2}+qt^2-\frac{qt^3}{2}\sqrt{\frac{q}{np}}+$$

$$+\rho\,[np+t\sqrt{npq}]-t\sqrt{npq}-\frac{pt^2}{2}+pt^2+$$

$$+\frac{t^3}{2}p\sqrt{\frac{p}{nq}}-\rho'\,[nq-t\sqrt{npq}]=\frac{t^2}{2}-\frac{qt^3}{2}\sqrt{\frac{q}{np}}+$$

$$+\frac{pt^3}{2}\sqrt{\frac{p}{nq}}+\rho\,[np+t\sqrt{npq}]-\rho'(nq-t\sqrt{npq})=\frac{t^2}{2}+\delta,$$

---

[1]) Так как $\left|t\sqrt{\frac{q}{np}}\right|<\frac{1}{2}$, то члены ряда ρ убывают быстрее, чем члены геометрической прогрессии с знаменателем 1/2, а потому значение ρ менее удвоенного его первого члена. Ввиду того, что и ρ' обладает тем же свойством, можем положить A равным наибольшему из двух чисел

$$\frac{2}{3}\left(\frac{q}{p}\right)^{3/2}\text{ и }\frac{2}{3}\left(\frac{p}{q}\right)^{3/2}.$$

при чем $|\delta| < \frac{h|t^3|}{n^{\frac{1}{2}}}$, где h — не зависящая от n и t постоянная величина [1]). Следовательно,

$$H = e^{\frac{t^2}{2}+\delta} = e^{\frac{t^2}{2}}(1+\gamma), \tag{162}$$

где $|\gamma| < 2|\delta| < \frac{2h|t^3|}{n^{1/2}}$, лишь бы $\frac{ht^3}{n^{1/2}} < \frac{1}{2}$, т.-е., при всяком данном t, γ стремится к 0, когда n неограниченно возрастает.

С другой стороны, имеем также

$$\frac{m(n-m)}{n} = n\left[p + t\sqrt{\frac{pq}{n}}\right]\left[q - t\sqrt{\frac{pq}{n}}\right] = npq(1+\beta), \tag{163}$$

где β также стремится к 0 при всяком данном t, когда n неограниченно возрастает.

Из (162) и (163) заключаем, что

$$I_n^m = C_n^m p^m q^{n-m} = \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi npq}}(1+\varepsilon), \tag{164}$$

где ε стремится равномерно к 0 с возрастанием n, для всех целых значений

$$m = np + t\sqrt{npq},$$

каково бы ни было t в определенном промежутке (t_0, t_1).

---

[1]) Замечая, что $\left(\frac{q}{2}\sqrt{\frac{q}{p}} - \frac{p}{2}\sqrt{\frac{p}{q}}\right) = \frac{1}{2}\sqrt{pq}\left(\frac{q}{p} - \frac{p}{q}\right) = \frac{q-p}{2\sqrt{pq}}$,

$$\left|\rho\left(np + t\sqrt{npq}\right)\right| < \frac{A|t^3|}{\sqrt{n}}p\left(1 + t\sqrt{\frac{q}{np}}\right) < \frac{3}{2}A\frac{|t^3|p}{\sqrt{n}}$$

и $\left|\rho'\left(nq - t\sqrt{npq}\right)\right| < \frac{3}{2}\frac{A|t^3|q}{\sqrt{n}}$), заключаем, что $h < \frac{|q-p|}{2\sqrt{pq}} + \frac{3}{2}A <$ $< \left(\frac{1}{2} + \frac{3}{2}\right)A$, т.-е. меньше наибольшего из чисел $\frac{4}{3}\left(\frac{q}{p}\right)^{3/2}$ и $\frac{4}{3}\left(\frac{p}{q}\right)^{3/2}$.

Пусть, например, $n = 10\,000$, $p = q = \frac{1}{2}$; в таком случае $m = 5\,000$ соответствует $t = 0$; тогда, отбрасывая в (164) множитель $(1 + \varepsilon)$, мы находим:

$$I_{10\,000}^{5\,000} \neq \frac{1}{50\sqrt{2\pi}} \neq 0{,}007\,978.$$

Таким образом, хотя $m = 5\,000$ является наиболее вероятным числом появлений события A, имеющего вероятность $\frac{1}{2}$, однако такое точное совпадение является маловероятным; точно так же, полагая последовательно $t = 1, 2, 3, 4$, находим:

$$I_{10\,000}^{4\,950} = I_{10\,000}^{5\,050} \neq \frac{e^{-\frac{1}{2}}}{50\sqrt{2\pi}} \neq 0{,}004\,894;$$

$$I_{10\,000}^{4\,900} = I_{10\,000}^{5\,100} \neq \frac{e^{-2}}{50\sqrt{2\pi}} \neq 0{,}001\,080;$$

$$I_{10\,000}^{4\,850} = I_{10\,000}^{5\,150} \neq \frac{e^{-4{,}5}}{50\sqrt{2\pi}} \neq 0{,}000\,100;$$

$$I_{10\,000}^{4\,800} = I_{10\,000}^{5\,200} \neq \frac{e^{-8}}{50\sqrt{2\pi}} \neq 0{,}000\,002.$$

Заметим, кроме того, что, если рассматриваем два последовательных целых числа m и $m + 1$, то приращение Δt, которое получает t при переходе от первого ко второму, определяется вычитанием равенства

$$m = np + t\sqrt{npq}$$

из равенства $$m + 1 = np + (t + \Delta t)\sqrt{npq},$$

откуда находим

$$1 = \Delta t\sqrt{npq},$$

т.-е.

$$\Delta t = \frac{1}{\sqrt{npq}}.$$

Следовательно, формуле (164) можно придать и такой вид:

$$I_n^m = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \Delta t (1+\varepsilon), \qquad (164\ \text{bis})$$

который допускает простую геометрическую интерпретацию. Строим кривую

$$Y = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}},$$

считая t абсциссами, а Y ординатами точек на плоскости. Отмечая на данной кривой точку M, имеющую некоторую абсциссу t, соответствующую, согласно формуле (158), определенному целому числу m, и затем — точку M_1 с абсциссой $t_1 = t + \Delta t$, соответствующей следующему целому числу $m+1$, мы видим, что

$$\Delta S = Y\Delta t = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \Delta t$$

представляет площадь прямоугольника, имеющего высотой ординату точки M на кривой, а основанием — промежуток $\Delta t = t_1 - t$. Таким образом вероятность I_n^m с точностью до множителя $(1+\varepsilon)$ измеряется площадью ΔS; иначе говоря,

$$I_n^m = \Delta S (1+\varepsilon).$$

Благодаря этому вероятность неравенства (155) равна

$$\sum I_n^m = \sum_{t_0}^{t_1} \Delta S (1+\varepsilon),$$

а следовательно, при бесконечном возрастании n

$$\text{пред.} \sum I_n^m = \text{пред.} \sum_{t_0}^{t_1} \Delta S (1+\varepsilon) = \text{пред.} \sum_{t_0}^{t_1} \Delta S,$$

где $\sum_{t_0}^{t_1}$ распространяется на смежные, указанные выше прямоугольники с равными основаниями Δt, заполняющими отрезок оси

абсцисс от [1]) $t_0+\delta$ до $t_1+\delta'$, где $\delta \leqslant \Delta t$ и $\delta' < \Delta t$. Но, согласно математическому определению площади, пред. $\sum_{t_0}^{t_1} \Delta S$ суммы рассматриваемых прямоугольников, когда число их увеличивается, а основания стремятся к нулю, представляет собою не что иное как площадь криволинейной фигуры, ограниченной снизу осью абсцисс, сверху кривой линией

$$Y=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$$

и с боков перпендикулярами к оси абсцисс в точках с абсциссами t_0 и t_1, которая аналитически выражается определенным интегралом

$$\text{пред.} \sum_{t_0}^{t_1} \Delta S=\frac{1}{\sqrt{2\pi}}\int_{t}^{t_1} e^{-\frac{t^2}{2}}dt.$$

Следовательно, предел, к которому стремится при неограниченном возрастании n вероятность неравенства (155), действительно равен

$$\frac{1}{\sqrt{2\pi}}\int_{t}^{t_1} e^{-\frac{t^2}{2}}dt=\Phi(t_1)-\Phi(t_0), \tag{165a}$$

где [2])

$$\Phi(t)=\frac{1}{\sqrt{2\pi}}\int_{0}^{t} e^{-\frac{t^2}{2}}dt,$$

что и требовалось доказать.

---

[1]) Имея в виду, что наименьшее целое число m, удовлетворяющее неравенству (155), заключено между

$$np+t_0\sqrt{npq} \quad \text{и} \quad np+t_0\sqrt{npq}+1=np+(t_0+\Delta t)\sqrt{npq},$$

и, аналогичным образом, наибольшее значение m, удовлетворяющее (155), заключено между

$$np+t_1\sqrt{npq}-1=np+(t_1-\Delta t)\sqrt{npq} \quad \text{и} \quad np+t_1\sqrt{npq}.$$

[2]) Функция $\Phi(t)$ является, очевидно, нечетной функцией, т.-е.

$$\Phi(-t)=-\Phi(t).$$

2. Теорема, или, как ее иногда называют, формула Лапласа чрезвычайно важна для практики, несмотря на то, что она является лишь предельной (для $n=\infty$) формулой, и, следовательно, для конечных значений n дает лишь приближенное значение искомой вероятности. Очевидно, для конечных значений n вероятность $F(n, t)$ (т.-е. интегральная функция распределения вероятностей величин t) неравенства

$$m-np<t\sqrt{npq}$$

является прерывной функцией t, потому что, когда t, возрастая, проходит через значения, которые соответствуют целым числам m, удовлетворяющим равенству

$$m=np+t\sqrt{npq},$$

вероятность $F(n, t)$ сразу увеличивается на $I_n^m=C_n^m p^m q^{n-m}$ и затем остается неизменной, пока t не получит приращения $\frac{1}{\sqrt{npq}}$, приводящего к следующему целому числу m. Таким образом интегральная функция распределения вероятностей $F(n, t)$ для величины t есть прерывная функция t; но если n достигает нескольких тысяч, то скачки ее $I_n^m=\frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi npq}}(1+\varepsilon)$ настолько малы, что $F(n, t)$ с вполне достаточной практически точностью можно отождествить с непрерывной функцией

$$F(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{t^2}{2}}\,dt=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} e^{-\frac{t^2}{2}}\,dt+$$

$$+\frac{1}{\sqrt{2\pi}}\int_{0}^{t} e^{-\frac{t^2}{2}}\,dt=\frac{1}{2}+\Phi(t). \qquad (166)$$

Кроме того, следует отметить, что, при данном n, $F(n, t)$ остается тождественно равной нулю, пока $np+t\sqrt{npq}<0$, т.-е. $t<-\sqrt{\frac{nq}{p}}$, и затем, при $t>\sqrt{\frac{np}{q}}$, $F(n, t)=1$, так как все возможные значения m удовлетворяют условию $0\leqslant m\leqslant n$. Предельная же функ-

ция $F(t)$ указанным свойством не обладает, так как $0 < F(t) < 1$ и только $F(-\infty) = 0$, а

$$F(\infty) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = 1.$$

Но это замечание не имеет практического значения, так как функция $\Phi(t)$ при сравнительно небольших (положительных) t очень близка к $\frac{1}{2}$, именно:

$t =$	4	5	6	10
$\Phi(t) \neq$	$\frac{1}{2} - \frac{3}{10^5}$	$\frac{1}{2} - \frac{2}{10^7}$	$\frac{1}{2} - \frac{1}{10^9}$	$\frac{1}{2} - \frac{1}{10^{22}}$

Поэтому значения $|t| > 6$, например, можно считать фактически невозможными.

Вообще, формула Лапласа дает достаточное для всех практических задач приближение [1]), когда npq достигает нескольких сотен.

---

[1]) Можно доказать, что существует значение α ($|\alpha| \leqslant 1$), для которого вероятность неравенства

$$\left| m - np - \frac{t^2}{6}(q-p) \right| < t\sqrt{npq} + \alpha$$

равна $2\left[\Phi(t) + \theta e^{-\sqrt[3]{2npq}}\right]$, где $|\theta| < 1$ (каковы бы ни были n и t, лишь бы $\frac{t^6}{16} \leqslant npq \geqslant 365$).

Пусть, например,

$$p = q = \frac{1}{2},\ n = 2\,500;\ \text{тогда}\ 2e^{-\sqrt[3]{1250}} < 0{,}000\,045,$$

так что $2\Phi(t)$ представляет с точностью до 4-го десятичного знака включительно вероятность неравенства

$$|m - np| < t\sqrt{npq} + \alpha,\ \text{если}\ t \leqslant \sqrt[6]{10\,000} \neq 4{,}65;$$

в частности, $2\Phi(3) \neq 0{,}9973$ (последний знак точен) есть вероятность, по крайней мере, одного из трех неравенств

$$1\,174 < m < 1\,326;\ 1\,175 < m < 1\,325 \text{ или } 1\,176 < m < 1\,324.$$

Применение формулы Лапласа при небольших значениях npq (не менее, однако, 20—30) также допустимо в большинстве случаев, где не требуется очень большой точности.

В дальнейшем мы будем пользоваться формулой Лапласа при любых значениях n как совершенно точной, но не следует забывать о сделанных выше оговорках.

Величину

$$t = \frac{m - np}{\sqrt{npq}} \tag{167}$$

мы будем называть нормированным отклонением числа m. Таким образом нормированное отклонение равно отклонению $m - np$ числа m (от его математического ожидания np), деленному на штандарт этого отклонения. Следовательно, М. О. $t = 0$ и М. О. $t^2 = 1$. Можно проверить эти два последних свойства, заметив, что

$$\text{М. О. } t = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-\frac{t^2}{2}} dt = 0,$$

$$\text{М. О. } t^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{t^2}{2}} dt = 1.$$

Допущение, что формула Лапласа применима для конечных значений n, означает, что плотность распределения вероятностей нормированного отклонения t не *зависит* от n и равна

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}, \tag{168}$$

и, в частности, что вероятность неравенства

$$\left| \frac{m - np}{\sqrt{npq}} \right| < t \tag{154 bis}$$

равна $2\Phi(t)$.

Остановивши наше внимание на таблице значений функции $\Phi(t)$, замечаем, что, когда t мало, то, как видно из таблицы (стр. 361), вероятность неравенства (154) или (154 bis) невелика и остается меньше $\frac{1}{2}$,

пока t не превышает значения $\mu = 0{,}67$ (или точнее [1]) 0,674), которое называется вероятным отклонением $|t|$, так как одинаково вероятно, что $|t| < \mu$, и то, что, напротив, $|t| > \mu$. Согласно закону больших чисел, мы должны ожидать, что, рассматривая очень большое число S серий опытов по n в каждой, приблизительно для $\frac{1}{2}$ всех серий $|t| = \left|\frac{m - np}{\sqrt{npq}}\right| < \mu$. По мере дальнейшего возрастания t, $2\Phi(t)$ растет, и в частности $2\Phi(1) = 0{,}6826 \neq \frac{2}{3}$. Таким образом можно запомнить, что вероятность неравенства

$$\frac{m - np}{\sqrt{npq}} < 1,$$

означающего, что уклонение $|m - np|$ остается меньше своего штандарта, равна $\frac{2}{3}$ (с точностью до 0,02).

После этого $2\Phi(t)$ все быстрее приближается к 1; в частности, уже при $t = 4$, $2\Phi(t) \neq 0{,}999\,936$, и соответствующее неравенство

$$\left|\frac{m - np}{\sqrt{npq}}\right| < 4 \tag{169}$$

обычно считают практически достоверным. Отсюда следует, например, что, если мы произведем 2 500 бросаний монеты, то можно быть практически уверенным, что число m появлений „орла“ будет удовлетворять неравенству

$$|m - 1\,250| < 4 \cdot 25, \tag{170}$$

т.-е.

$$1\,150 < m < 1\,350;$$

и, с другой стороны, столь же вероятно, что

$$|m - 1\,250| < 0{,}674 \cdot 25,$$

т.-е.

$$1\,233{,}2 < m < 1\,268{,}7,$$

как то, что m в указанных пределах не заключается

[1]) Полезно помнить приближенное значение $\mu \neq \frac{2}{3} = 0{,}666\ldots$, верное с точностью до 0,01.

С увеличением числа опытов n в 4 раза соответствующие пределы неравенства (154), которому мы (при том же t) должны приписать ту же вероятность, увеличатся в $2=\sqrt{4}$ раза, так что при 10 000 бросаниях монеты мы можем, например, быть практически уверены, что

$$|m-5\,000|<4\cdot 50, \tag{171}$$

а вероятное отклонение m будет равно $0{,}674\cdot 50=33{,}7$ вместо 16,85.

Мы можем коротко выразить это, говоря, что отклонение $m-np$ возрастает пропорционально $\sqrt{n}$. Отсюда следует, что $\frac{m-np}{n}=\frac{m}{n}-p$, т.-е. разность между статистической и теоретической вероятностью пропорциональна $\frac{\sqrt{n}}{n}=\frac{1}{\sqrt{n}}$ и изменяется таким образом (убывает) обратно пропорционально $\sqrt{n}$. Так, при 2 500 опытах неравенство (170) равнозначно

$$\left|\frac{m}{n}-\frac{1}{2}\right|<\frac{1}{25}, \tag{170 bis}$$

и при $n=10\,000$ неравенство (171) равнозначно

$$\left|\frac{m}{n}-\frac{1}{2}\right|<\frac{1}{50} \tag{171 bis}$$

(оба неравенства при 2500 и 10 000 бросаниях монеты, соответственно, практически достоверны).

Упражнения. 1) Определить, применяя формулу Лапласа, вероятность P того, что при 8 000 бросаниях игральной кости статистическая вероятность $\frac{m}{8\,000}$ появления 6 очков отклонится от $p=\frac{1}{6}$ меньше, чем на $\frac{1}{80}$.

Ответ. $P=0{,}9973$. Действительно, $\left|\frac{m}{n}-\frac{1}{6}\right|<\frac{1}{80}$ равнозначно $\left|m-\frac{n}{6}\right|<100=t\sqrt{8000\cdot\frac{1}{6}\cdot\frac{5}{6}}$, если $t=100:\frac{200}{6}=3$; но $2\Phi(3)=0{,}9973$.

2) Показать, что, если к величине x, имеющей штандарт σ, применима теорема Лапласа, то, в среднем, из 20 наблюдаемых значений x только одно будет уклоняться от своего математического ожидания a более, чем на $1{,}96\sigma$; из 500 значений — только одно уклонение превысит 3σ; из 10 000 значений — только одно превысит $3{,}9\sigma$, и потребуется, в среднем, 100 000 наблюдений, чтобы превысить 1 раз уклонение $4{,}4\sigma$.

Ответ. $1-2\Phi(1{,}96)\neq 0{,}05$ и т. д.

3) Пусть вероятность p события A равна $^1/_3$ и пусть число опытов $n = 45\,000$. Каково вероятное отклонение числа m появлений события A от 15 000?

Ответ. 67. Действительно, штандарт m равен $\sqrt{npq} = \sqrt{\frac{45\,000 \cdot 2}{3 \cdot 3}} = 100$; так как вероятное отклонение соответствует $t = 0{,}67$, то вероятное отклонение m равно $0{,}67 \cdot 100 = 67$.

4) Какова вероятность при бросании монеты 90 000 раз, что

$$43\,500 < m < 45\,000?$$

Ответ. $\frac{1}{2} - \frac{1}{10^{22}}$; действительно, штандарт $\sqrt{npq} = 150$; поэтому наше неравенство равнозначно $0 > m - np > -10 \cdot \sqrt{npq}$, вероятность которого равна $\Phi(0) - \Phi(-10) = \Phi(10) = \frac{1}{2} - \frac{1}{10^{22}}$.

5) Сколько раз надо бросить монету, чтобы вероятность того, что разность $\left| \frac{m}{n} - \frac{1}{2} \right| < \frac{1}{50}$, была не менее $2\Phi(4) \neq 0{,}999\,936$?

Ответ. $n \geq 10\,000$. Действительно, требуемое неравенство, равнозначное $\left| m - \frac{n}{2} \right| < \frac{n}{50}$, будет иметь вероятность $2\Phi(4)$, если $\frac{n}{50} = 4\sqrt{npq} = 2\sqrt{n}$, откуда $\sqrt{n} = 100$ и $n = 10\,000$.

6) Игрок выигрывает 7 рублей, если при бросании игральной кости появляется 6 очков, и уплачивает 1 рубль, когда появляется меньшее число очков. Требуется указать, с вероятностью равною $2\Phi(4) \neq 0{,}999\,936$, между какими пределами будет заключен его выигрыш R при повторении игры $n = 8\,000$ раз.

Ответ. $1\,600 < R < 3\,733$. Действительно, обозначая через m число выигранных партий, с требуемой вероятностью можно утверждать, что $\left| m - \frac{8\,000}{6} \right| < 4\sqrt{\frac{8\,000 \cdot 5}{6 \cdot 6}}$, т.-е. что $1\,200 < m < \frac{4\,400}{3}$. Но выигрыш $R = 7m - n + m = 8m - 8\,000$, откуда $m = \frac{1}{8} R + 1\,000$; следовательно, подставляя это значение m в написанное неравенство, находим:

$$1\,200 < \frac{1}{8} R + 1\,000 < \frac{4\,400}{3}, \quad \text{т.-е.} \quad 200 < \frac{1}{8} R < \frac{1\,400}{3}.$$

7) Сохраняя условия предыдущей задачи, вычислить вероятность P, что игрок окажется в убытке.

Ответ. $P = \frac{1}{2} - \Phi(10) \neq \frac{1}{10^{22}}$. Действительно, для этого надо, чтобы $m < 1\,000$, т.-е. $m - \frac{8\,000}{6} < -\frac{2\,000}{6} = t\sqrt{\frac{8\,000 \cdot 5}{6 \cdot 6}}$, откуда $t = -10$; поэтому $P = \Phi(-10) - \Phi(-\infty) = \frac{1}{2} - \Phi(10)$. Применяя неравенство Чебышева, мы

бы знали только, что $P < \frac{1}{10^2}$, а применяя неравенство (106), мы могли бы утверждать, что $P < e^{-25} < \frac{2}{10^{11}}$.

8) В различное время ботаники Mendel, Correns, Tchermak, Hurst, Bateson и Lokk скрещивали желтый (гибридный) горох и получили в общей сложности 25 647 желтых семян и 8 506 зеленых. Требуется выяснить, совместим ли этот результат с гипотезой Менделя, что вероятность появления зеленого гороха во всех опытах была равна $\frac{1}{4}$.

Ответ. Совместим, так как $\frac{8\,506}{34\,153} - \frac{1}{4} \neq -0{,}0012$, между тем как

$$= \sqrt{\frac{3}{16 \cdot 34\,153}} \neq 0{,}0024.$$

ГЛАВА ВТОРАЯ

РАСПРОСТРАНЕНИЕ ТЕОРЕМЫ ЛАПЛАСА НА СЛУЧАЙ УРНЫ С НЕВОЗВРАЩАЕМЫМИ ШАРАМИ

1. Из различных обобщений теоремы Лапласа заслуживает особого внимания распространение ее на ряд опытов, связь между которыми может быть схематически представлена следующим образом (схема урны с невозвращаемыми шарами).

В урне имеется весьма большое число N шаров, из которых pN белых и qN черных шаров $(p+q=1)$, отличающихся между собой только цветом, так что вероятность появления белого шара при извлечении одного шара из урны равна p. Из урны вынимается весьма большое число n шаров одновременно или последовательно, но так, что вынутый однажды шар обратно в урну не возвращается и, следовательно, вторично не может появиться. При этом предполагается, что $\frac{n}{N}<\lambda$, где λ — некоторое определенное число, меньшее чем 1.

Хотя в данном случае вероятность любому из вынутых шаров быть белым равна p, однако, если известно, что первый вынутый шар был, например, белым, то вероятность следующему шару быть белым уже будет немного меньше, чем p, а именно,

$$\frac{pN-1}{N-1}=\frac{p(N-1)-q}{N-1}=p-\frac{q}{N-1};$$

вообще, если при нескольких извлечениях пропорция вынутых белых шаров к общему числу произведенных извлечений более p, то пропорция оставшихся белых шаров к общему числу оставшихся в урне шаров будет менее p, и, следовательно, вероятность появления белого шара при следующем вынимании будет немного менее p (обратное имело бы место, если бы отношение вынутых белых шаров

к общему числу произведенных извлечений было менее p). Благодаря этому, в случае, когда вынутые шары в урну не возвращаются, последовательные извлечения не являются независимыми между собой, и без всяких вычислений очевидно, что чем $\frac{n}{N}$ будет больше, тем более эта зависимость будет тормозить рост уклонения числа m появившихся белых шаров от их математического ожидания np. Кроме того, нетрудно вычислить дисперсию m, т.-е.

$$B = \text{М. О. } [m - np]^2 = \text{М. О. } [(x_1 - p) + \ldots + (x_n - p)]^2 = \\ = npq + 2\sum \text{М. О. } (x_i - p)(x_k - p),$$

где x_i равно 1 или 0, смотря по тому, оказывается ли i-й шар белым или черным. Поэтому

$$\text{М. О. } (x_i - p)(x_k - p) = \text{М. О. } x_i x_k - p^2 = \\ = p\left(p - \frac{q}{N-1}\right) - p^2 = -\frac{pq}{N-1},$$

ибо М. О. $x_i x_k$ равно вероятности совмещения белого шара в i-м извлечении с белым шаром при k-м извлечении, и, как было замечено выше, после того, как известно, что i-й шар был белым, вероятность k-му шару оказаться белым равна $p - \frac{q}{N-1}$. Следовательно,

$$B = npq - \frac{n(n-1)pq}{N-1} = npq\left(1 - \frac{n-1}{N-1}\right) = npq\,\frac{N-n}{N-1}. \qquad (172)$$

Таким образом, как и следовало ожидать, коэффициент дисперсии в данном случае

$$\frac{B}{npq} = 1 - \frac{n-1}{N-1} = \frac{N-n}{N-1}$$

меньше 1, и дисперсия оказывается поднормальною, т.-е. меньшею, чем в ряде независимых опытов с тою же вероятностью p. Случай независимых опытов является частным или, вернее, предельным случаем рассматриваемой нами схемы невозвращаемых шаров, когда число N бесконечно велико (по сравнению с n).

2. Докажем теперь следующее обобщение предельной теоремы Лапласа.

Теорема. Вероятность неравенства

$$t_0\sqrt{npq\left(\frac{N-n}{N}\right)}<m-np<t_1\sqrt{npq\left(\frac{N-n}{N}\right)} \qquad (173)$$

имеет пределом

$$\Phi(t_1)-\Phi(t_0)=\frac{1}{\sqrt{2\pi}}\int_{t_0}^{t_1}e^{-\frac{t^2}{2}}dt, \qquad (156)$$

если числа n и N (имеющие вышеуказанный смысл) бесконечно возрастают, при чем $\frac{n}{N}<\lambda<1$.

Замечаем прежде всего, согласно (46), что вероятность I_n^m появления m белых шаров при n извлечениях определяется формулой

$$I_n^m=\frac{C_{pN}^m C_{qN}^{n-m}}{C_N^n}, \qquad (46\ \text{bis})$$

или, полагая

$$N-n=M \text{ и } m=pn+x,$$

$$I_n^m=\frac{n!\,M!}{N!}\cdot\frac{(pN)!\,(qN)!}{(pn+x)!\,(pM-x)!\,(qn-x)!\,(qM+x)!}, \qquad (46\ \text{ter})$$

так как

$$pN-m=p(N-n)-x=pM-x,$$
$$n-m=n(1-p)-x=qn-x$$

и

$$qN-n+m=qN-qn+x=qM+x.$$

Таким образом

$$I_n^{m+1}=I_n^m\cdot\frac{(pM-x)(qn-x)}{(pn+x+1)(qM+x+1)}, \qquad (174)$$

откуда

$$\frac{I_n^{m+1}-I_n^m}{I_n^m}=\frac{(pM-x)(qn-x)}{(pn+x+1)(qM+x+1)}-1=$$
$$=\frac{(pM-x)(qn-x)-(pn+x+1)(qM+x+1)}{(pn+x+1)(qM+x+1)}=$$
$$=\frac{-(N+2)x-(pn+qM+1)}{(pn+x+1)(qM+x+1)}. \qquad (175)$$

Прежде чем преобразовывать дальше полученное равенство, напишем соответствующую ему формулу для случая Бернулли (возвращаемых шаров).

Тогда

$$I_n^{m+1} = I_n^m \cdot \frac{n-m}{m+1} \cdot \frac{p}{q};$$

поэтому[1])

$$\frac{I_n^{m+1} - I_n^m}{I_n^m} = \frac{(n-m)p - (m+1)q}{(m+1)q} =$$

$$= \frac{np - m - q}{(m+1)q} = -\frac{x+q}{(pn+x+1)q}. \qquad (175\ \text{bis})$$

Полагая $x = z\sqrt{n}$, заметим, что и в том и в другом случае нас интересуют только конечные значения z, ибо мы знаем, что вероятность неравенства $|m - np| > z\sqrt{n}$ очень близка к 0, когда z велико, ввиду того, что штандарт m равен или менее $\sqrt{npq}$. Но придавая z какое-нибудь определенное конечное значение, мы видим, что отношение

$$\frac{I_n^{m+1}}{I_n^m} = \frac{(pM - z\sqrt{n})(qn - z\sqrt{n})}{(pn + z\sqrt{n} + 1)(qM + z\sqrt{n} + 1)} =$$

$$= \frac{\left(1 - \frac{z\sqrt{n}}{pM}\right)\left(1 - \frac{z}{q\sqrt{n}}\right)}{\left(1 + \frac{z}{p\sqrt{n}} + \frac{1}{pn}\right)\left(1 + \frac{z\sqrt{n}}{qM} + \frac{1}{qM}\right)} \qquad (176)$$

с возрастанием n стремится к 1, так как

$$\frac{\sqrt{n}}{M} = \frac{\sqrt{n}}{N-n} = \frac{1}{\sqrt{n}} \cdot \frac{1}{\frac{N}{n} - 1} < \frac{1}{\sqrt{n}} \cdot \frac{\lambda}{1-\lambda}.$$

---

[1]) Формулу (175 bis) можно было бы непосредственно получить из (175), деля числитель и знаменатель во 2-й части на N и безгранично увеличивая N, при чем все дроби вида $\frac{a}{N}$ стремятся к 0, если только a не имеет множителем N или M (очевидно, что пред. $\frac{M}{N} = \frac{N-n}{N} = 1$, так как в случае Бернулли пред. $\frac{n}{N} = 0$).

С другой стороны, применяя к логарифму известную формулу (конечных приращений)

$$f(b)-f(a)=f'(c)(b-a),$$

имеем вообще

$$\log b-\log a=\frac{1}{c}(b-a)=\frac{b-a}{a}\cdot\frac{a}{c},$$

где c — какое-то число, заключенное между a и b $\left(\text{ибо } (\log x)'=\frac{1}{x}\right)$.

Следовательно,

$$\frac{I_n^{m+1}-I_n^m}{I_n^m}=\left[\log I_n^{m+1}-\log I_n^m\right]\frac{C}{I_n^m},$$

где C заключено между I_n^{m+1} и I_n^m, так что $\frac{C}{I_n^m}$ заключается между 1 и $\frac{I_n^{m+1}}{I_n^m}$; поэтому, принимая во внимание, что $\frac{I_n^{m+1}}{I_n^m}$ с возрастанием n стремится к 1, можем написать, что

$$(\log I_n^{m+1}-\log I_n^m)=(1+\varepsilon)\frac{I_n^{m+1}-I_n^m}{I_n^m}, \qquad (177)$$

где $|\varepsilon|$ становится сколь угодно малым при достаточно больших значениях n.

Прилагая это замечание сначала к равенству (175 bis), получим новое доказательство теоремы Лапласа.

Действительно, (175 bis) преобразуется в

$$\log I_n^{m+1}-\log I_n^m=-\frac{x+q}{(pn+x+1)q}(1+\varepsilon); \qquad (178)$$

но, полагая $x=t\sqrt{npq}$ и замечая, что увеличение m на 1 соответствует также увеличению x на 1 (так как $m=pn+x$), а по-

тому t получает при этом приращение $\Delta t = \frac{1}{\sqrt{npq}}$, видим, что

$$\frac{x+q}{(pn+x+1)q} = \frac{t\sqrt{npq}+q}{(pn+t\sqrt{npq}+1)q} =$$

$$= \frac{1}{\sqrt{npq}} \cdot \frac{t+\frac{q}{\sqrt{npq}}}{1+t\sqrt{\frac{q}{pn}}+\frac{1}{np}} =$$

$$= \Delta t \frac{t+\frac{q}{\sqrt{npq}}}{1+t\sqrt{\frac{q}{pn}}+\frac{1}{np}} = \Delta t(t+\varepsilon'),$$

где $|\varepsilon'|$ стремится к 0 с возрастанием n.

Поэтому равенство (178) принимает вид:

$$\log I_n^{m+1} - \log I_n^m = -\Delta t(t+\alpha), \tag{179}$$

где α стремится (равномерно) к 0 при возрастании n, если t не превышает некоторого произвольно большого, но определенного предела.

Придавая m последовательные целые значения от $m_0 = np + x_0$, где $0 \leqslant x_0 < 1$, до $m = m_1 - 1 = np + x_1 - 1 = np + t_1\sqrt{npq} - 1 = np + \left(t_1 - \frac{1}{\sqrt{npq}}\right)\sqrt{npq}$ и складывая все соответствующие равенства (179), получим

$$\log I_n^{m_1} - \log I_n^{m_0} = -\sum_{t=\frac{x_0}{\sqrt{npq}}}^{t=t_1-\frac{1}{\sqrt{npq}}} (t\Delta t + \alpha\Delta t),$$

так как все промежуточные члены в первой части при сложении попарно сокращаются. Поэтому, при безграничном возрастании n, находим

$$\underset{n=\infty}{\text{пред.}} \left[\log I_n^{m_1} - \log I_n^{m_0}\right] = -\int_0^{t_1} t\,dt = -\frac{t_1^2}{2},$$

откуда следует, что

$$\underset{n=\infty}{\text{пред.}} \frac{I_n^{m_1}}{I_n^{m_0}} = e^{-\frac{t_1^2}{2}};$$

или, иначе говоря, вообще при $m = np + t\sqrt{npq}$

$$I_n^m = I_n^{m_0} e^{-\frac{t^2}{2}} (1 + \delta), \tag{180}$$

где δ стремится равномерно к 0 при неограниченном возрастании n, если t не превышает произвольно большого данного числа.

Что же касается неизвестного пока множителя $I_n^{m_0}$ (не зависящего от t), то он определяется из замечания, что, при t достаточно большом, сумма

$$\sum_{m=np-t\sqrt{npq}}^{m=np+t\sqrt{npq}} I_n^m,$$

согласно теореме Чебышева, становится сколь угодно близкой к 1; поэтому, если, взявши t достаточно большим, станем затем безгранично увеличивать n, то

$$\left| I_n^{m_0} \sqrt{npq} \sum_{-t}^{t} e^{-\frac{t^2}{2}} \Delta t - 1 \right| < \eta,$$

так что и

$$\left| \underset{n=\infty}{\text{пред.}} I_n^{m_0} \sqrt{npq} \int_{-t}^{+t} e^{-\frac{t^2}{2}} dt - 1 \right| < \eta,$$

как бы мало ни было данное число η.

Следовательно,

$$\text{пред. } I_n^{m_0} \sqrt{npq} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \text{пред. } I_n^{m_0} \sqrt{2\pi npq} = 1.$$

Отсюда

$$I_n^{m_0} = \frac{1+\gamma}{\sqrt{2\pi npq}}$$

вследствие (180),

$$I_n^m = \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi npq}} (1 + \gamma'), \tag{164 bis}$$

где γ и γ' стремятся к нулю с возрастанием n. Из (164 bis) теорема Лапласа выводится, как было указано в предыдущей главе.

3. Тот же способ рассуждений применим и в общем случае. Действительно, если, как мы предполагаем, $\frac{|x|}{\sqrt{n}} < L$, где L—произвольно большое, но определенное число, то при безграничном возрастании n

$$\text{пред.} \left(1 + \frac{x+1}{pn}\right)\left(1 + \frac{x+1}{qM}\right) = 1.$$

Поэтому при помощи (177) равенство (175) преобразуется в

$$\log I_n^{m+1} - \log I_n^m = -\frac{(N+2)x + (pn + qM + 1)}{pqnM}(1+\varepsilon'), \quad (181)$$

где ε' стремится к 0 с возрастанием n. Следовательно, полагая $x = t\sqrt{\frac{pqnM}{N}}$, так что t получает приращение $\Delta t = \sqrt{\frac{N}{pqnM}}$, когда m увеличивается на 1, находим

$$\log I_n^{m+1} - \log I_n^m = -\Delta t\left(t + \frac{2t}{N} + \frac{pn + qM + 1}{\sqrt{pqnMN}}\right)(1+\varepsilon') =$$

$$= -\Delta t(t + \alpha), \quad (181 \text{ bis})$$

где α равномерно стремится к 0 с возрастанием n.

Отсюда заключаем, воспроизводя буквально сделанное выше рассуждение, что, при $m = np + t\sqrt{\frac{pqnM}{N}}$,

$$I_n^m = I_n^{m_0} e^{-\frac{t^2}{2}}(1+\delta),$$

где δ стремится равномерно к 0 с возрастанием n $\left(\frac{n}{N} < \lambda < 1\right)$, если $|t|$ не превышает данного произвольно большого числа.

Для определения коэффициента пропорциональности $I_n^{m_0}$ замечаем, как и раньше, что, при произвольно малом η,

$$\left| I_n^{m_0}\sqrt{npq\frac{N-n}{N}} \sum_{-t}^{+t} e^{-\frac{t^2}{2}}\Delta t - 1 \right| < \eta,$$

если t достаточно велико; поэтому

$$\underset{n=\infty}{\text{пред.}}\ I_n^{m_0}\sqrt{npq\frac{N-n}{N}}\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}\,dt=1,$$

откуда

$$I_n^{m_0}=\frac{1+\gamma}{\sqrt{2\pi npq\frac{N-n}{N}}},$$

$$I_n^{m}=\frac{(1+\gamma')e^{-\frac{t^2}{2}}}{\sqrt{2\pi npq\frac{N-n}{N}}}=(1+\gamma')e^{-\frac{t^2}{2}}\Delta t, \tag{182}$$

где γ и γ' стремятся к 0 при безграничном увеличении n.

Следовательно,

$$\underset{n=\infty}{\text{пред.}}\sum_{m=np+x_0}^{m=np+x_1} I_n^{m}=\frac{1}{\sqrt{2\pi}}\int_{t_0}^{t_1} e^{-\frac{t^2}{2}}\,dt,$$

где $x_0=t_0\sqrt{npq\frac{N-n}{N}}$, $x_1=t_1\sqrt{npq\frac{N-n}{N}}$; иными словами, вероятность неравенства

$$t_0\sqrt{npq\frac{N-n}{N}}<m-np<t_1\sqrt{npq\frac{N-n}{N}} \tag{173}$$

имеет пределом

$$\frac{1}{\sqrt{2\pi}}\int_{t_0}^{t_1} e^{-\frac{t^2}{2}}\,dt,$$

если n безгранично возрастает, что и требовалось доказать.

Предположим, например, что $N=45\,000$, $p=\frac{1}{2}$, $n=40\,000$.

В таком случае

$$\sqrt{npq}=100,\ \text{а}\ \sqrt{npq\frac{N-n}{N}}=\frac{100}{3}.$$

Таким образом, если шары обратно в урну не возвращаются, то штандарт уклонения m в 3 раза меньше нормального. Полагая $t=3$, видим, что в этом случае вероятность того, что

$$19\,900<m<20\,100 \tag{183}$$

равна $2\Phi(3) \neq 0{,}9973$, т.-е. имеем больше 997 шансов против 3, что это неравенство осуществится; напротив, если бы шары обратно возвращались, вероятностью $2\Phi(3)$ обладало бы неравенство

$$19\,700 < m < 20\,300;$$

неравенство же (183) обладало бы сравнительно небольшой вероятностью $2\Phi(1) \neq \frac{2}{3}$.

Применяя способ дисперсии к подобным сериям опытов, мы могли бы констатировать поднормальную дисперсию. Однако столь значительное уменьшение дисперсии на практике редко встречается, обычно отношение $\frac{n}{N}$ не так близко к 1, как в нашем примере $\left(\frac{n}{N} = \frac{8}{9}\right)$; если бы мы положили, например, что первоначальное число шаров равно $N = 270\,400$, так что $\frac{n}{N} = \frac{40\,000}{270\,400} = \frac{25}{169} \neq 0{,}15$ (15%), то $\sqrt{\frac{N-n}{N}} = \sqrt{\frac{144}{169}} = \frac{12}{13}$; поэтому вероятность $2\Phi(3) \neq 0{,}9973$ следовало бы приписать неравенству

$$19\,723 < m < 20\,277,$$

так что статистически этот случай весьма трудно было бы отличить от случая Бернулли; ясно, что чем N (при данном n) будет больше, тем это отличие будет меньше.

Идея данного нами доказательства обобщенной теоремы Лапласа заимствована у Пирсона, который, однако, использовал ее для другой цели, а именно для установления новых типов кривых распределения вероятностей, обобщающих кривую Лапласа (допустивши при этом некоторые математические неправильности).

ГЛАВА ТРЕТЬЯ

ОБЩИЕ УСЛОВИЯ ПРИЛОЖИМОСТИ ЗАКОНА НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

1. Кривая

$$Y = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}, \tag{184}$$

к которой нас привело исследование предыдущих двух глав, называется основной (или нормированной) нормальной кривой распределения вероятностей. Величина t, для которой эта кривая служит кривой распределения, обладает, в частности, свойствами, что М. О. $t = 0$ и М. О. $t^2 = 1$, т.-е. штандарт величины t равен 1.

Если некоторая другая величина x связана с t равенством

$$x = a + t\sigma, \tag{185}$$

то

$$\text{М. О. } x = \text{М. О. } a + \text{М. О. } t\sigma = a,$$

а дисперсия, или квадрат штандарта, величины x

$$\beta = \text{М. О. } (x - a)^2 = \text{М. О. } t^2\sigma^2 = \sigma^2.$$

Таким образом коэффициент σ при t в равенстве (185) является штандартом величины x.

Из равенства (185) следует, что неравенство

$$t_0\sigma < x - a < t_1\sigma_1$$

равнозначно неравенству

$$t_0 < t < t_1,$$

а потому вероятность соблюдения этого неравенства равна

$$\Phi(t_1) - \Phi(t_0) = \frac{1}{\sqrt{2\pi}} \int_{t_0} e^{--} dt.$$

Примерами подобных величин x могут служить (приближенно) числа m или $\frac{m}{n}$, рассмотренные в предыдущих главах; так, например, для случая Бернулли

$$m = np + t\sqrt{npq},$$

$$\frac{m}{n} = p + t\sqrt{\frac{pq}{n}}.$$

Мы будем говорить, что величина x также подчиняется закону нормального распределения (или закону Гаусса) с центром a и штандартом σ.

Нетрудно получить кривую распределения (плотностей) вероятностей для величин x; для этого замечаем, что вероятность $F(x_1)$ неравенства

$$x < x_1 = a + t_1\sigma$$

равна

$$F(x_1) = \Phi\left(\frac{x_1 - a}{\sigma}\right) - \Phi\left(-\infty\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x_1 - a}{\sigma}} e^{-\frac{t^2}{2}}\,dt;$$

поэтому, согласно (79), функция

$$f(x) = F'(x) = \frac{1}{\sigma}\Phi'\left(\frac{x-a}{\sigma}\right) = \frac{e^{-\frac{(x-a)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

представляет плотность распределения вероятностей величины x.

Кроме того, можно непосредственно заметить, что вероятность $f(x_0)\,dx$ неравенства

$$x_0 < x < x_0 + dx,$$

т.-е.

$$\frac{x_0 - a}{\sigma} < t < \frac{x_0 - a + dx}{\sigma},$$

равна

$$\frac{1}{\sqrt{2\pi}}\int_{\frac{x_0 - a}{\sigma}}^{\frac{x_0 - a + dx}{\sigma}} e^{-\frac{t^2}{2}}\,dt = \frac{dx}{\sigma\sqrt{2\pi}}\,e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Итак,

$$Y = \frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{(x-a)^2}{2\sigma^2}} \tag{186}$$

есть нормальная кривая распределения (плотностей) вероятностей [1]) с центром a и штандартом σ.

Очевидно, что абсцисса a центра не влияет на форму кривой, определяя лишь значение x, соответствующее максимальной ординате нормальной кривой; таким образом все нормальные кривые, имеющие одинаковый штандарт σ, получаются посредством параллельного перемещения вдоль оси абсцисс кривой

$$Y=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}. \qquad (186\ \text{bis})$$

Напротив, изменение штандарта σ существенно влияет на форму нормальной кривой.

Возьмем нормированную (основную) нормальную кривую

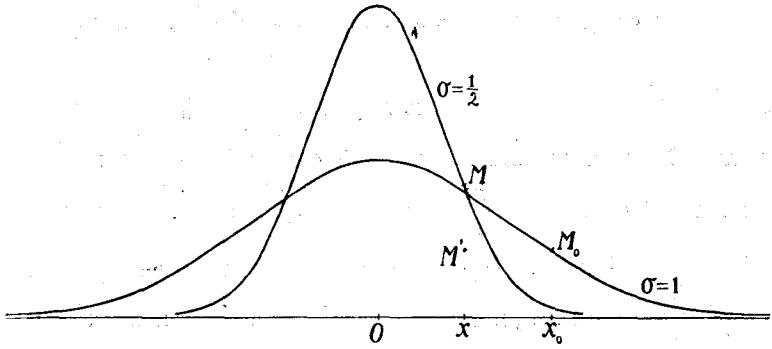
$$Y_0=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},$$

штандарт которой равен 1.

Беря на последней произвольную точку M_0 с координатами (x_0, Y_0), заменим ее сначала точкой M' с абсциссой $x=\sigma x_0$ и с той же самой ординатой; в таком случае, сделавши то же преобразование над всеми точками основной кривой, мы получим новую кривую

$$Y'=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}},$$

которая будет растянута по сравнению с основной кривой, если $\sigma>1$, и сжата, если $\sigma<1$. Однако полученная после этого преобразования кривая (черт. 6) не будет кривою распределения вероятностей,



Черт. 6.

[1]) Рекомендуем читателю перед чтением дальнейшего возобновить памяти содержание IV главы второй части.

так как вся площадь, ограниченная ею, не будет равна 1, а будет, очевидно, равна σ, т.-е. увеличена или уменьшена пропорционально произведенному растяжению или сжатию.

Для того, чтобы компенсировать произведенное изменение площади, нам остается заменить ординаты Y' всех точек M' через $Y = \frac{Y'}{\sigma}$, т.-е. при $\sigma > 1$ компенсировать продольное растяжение фигуры соответствующим поперечным сжатием, и наоборот в случае $\sigma < 1$. Таким образом равенство всей площади единице будет восстановлено, и мы получим кривую распределения вероятностей, точки которой M определяются уравнением

$$Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}, \qquad (186 \text{ bis})$$

соответствующим штандарту σ.

Из сказанного следует, что при увеличении штандарта σ кривая делается более сплюснутой к середине $\left(\text{максимальная ордината ее равна } \frac{1}{\sigma\sqrt{2\pi}}\right)$ и распределение вероятностей становится более равномерным, ибо, для больших значений x, Y будет тем больше, чем больше будет σ. Напротив, если бы σ было очень мало, то мы имели бы высокую и узкую кривую, которая даже для небольших значений x уже сливалась бы графически с осью абсцисс (так, при $\sigma = 0{,}1$, мы имели бы, для $x \geqslant \frac{1}{2}$, $Y \leqslant \leqslant \frac{10}{\sqrt{2\pi}} e^{-12{,}5} \neq 0{,}000\,015$, между тем как $Y_0 = \frac{10}{\sqrt{2\pi}} \neq 3{,}99$ при $x = 0$); в частности, этим свойством обладает распределение вероятностей отношения $\frac{m}{n}$ (числа появлений события с постоянной вероятностью к числу n независимых опытов), штандарт которого $\sigma = \sqrt{\frac{pq}{n}}$ весьма мал, когда n достаточно велико, и в соответствии с теоремой Бернулли сколько-нибудь значительные отклонения $\frac{m}{n}$ (от p) практически невозможны.

Рекомендуя читателю построить при помощи указанного выше приема несколько нормальных кривых, соответствующих различным значениям штандарта σ, укажем еще на геометрическое свойство величины σ. Все кривые распределения выпуклы кверху вблизи центра, но при $x=\pm\sigma$ они имеют точку перегиба, после которой выпуклость кривой направляется вниз. Действительно, дифференцируя дважды уравнение (186 bis), находим

$$Y''=\frac{e^{-\frac{x^2}{2\sigma^2}}}{\sigma^3\sqrt{2\pi}}\left(\frac{x^2}{\sigma^2}-1\right),$$

откуда видно, что $Y''<0$ при $|x|<\sigma$, $Y''=0$ при $|x|=\sigma$ и $Y''>0$ при $|x|>\sigma$.

Заканчивая эти общие элементарные замечания, напомним еще раз, что ордината Y сама по себе отнюдь не представляет вероятности (например, при $x=0$, возможно, что $Y>1$, если σ достаточно мало); значение вероятности имеет только соответствующая площадь или интеграл

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{x_0}^{x_1} e^{-\frac{x^2}{2\sigma^2}}dx,$$

который представляет вероятность, что $x_0<x<x_1$.

2. Установим теперь следующую общую лемму.

Пусть $f(x)$ есть функция распределения вероятностей величины x; пусть, далее, $z=\varphi(x)$ есть некоторая возрастающая [1]) функция x, а $x=\psi(z)$ — обратная функция; в таком случае функция распределения вероятностей для величины z равна $f[\psi(z)]\,\psi'(z)$.

В самом деле, вероятность, что $x_0<x<x_1$, равна $\int_{x_0}^{x_1} f(x)\,dx$; поэтому вероятность неравенства

$$z_0<z<z_1,$$

которое равнозначно неравенству

$$\psi(z_0)<\psi(z)=x<\psi(z_1),$$

[1]) Если z есть функция убывающая, то лемма остается в силе, с той лишь разницей, что перед $f[\psi(z)]\,\psi'(z)$ нужно поставить знак —.

равна

$$\int_{\psi(z_0)}^{\psi(z_1)} f(x)\,dx = \int_{z_0}^{z_1} f[\psi(z)]\,\psi'(z)\,dz; \qquad (187)$$

другими словами, $f[\psi(z)]\,\psi'(z)$ есть функция распределения (плотностей) вероятностей для $z=\varphi(x)$ [где $\varphi(x)$ есть функция обратная $\psi(z)$].

Применим доказанную лемму к случаю, когда распределение вероятностей величины x нормально, так что

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{(x-a)^2}{2\sigma^2}}, \qquad (186)$$

и положим, что $z=\varphi(x)$ есть линейная функция x, которую можем написать в виде

$$z=A+\lambda(x-a),$$

откуда

$$x=\psi(z)=a+\frac{1}{\lambda}(z-A) \quad \text{и} \quad \psi'(z)=\frac{1}{\lambda}.$$

Поэтому, согласно формуле (187), вероятностью неравенства

$$z_0<z<z_1$$

будет

$$\int_{z_0}^{z_1} f\left(a+\frac{z-A}{\lambda}\right)\frac{dz}{\lambda},$$

т.-е., принимая во внимание (186),

$$\frac{1}{\lambda\sigma\sqrt{2\pi}}\int_{z_0}^{z_1} e^{-\frac{(z-A)^2}{2\lambda^2\sigma^2}}\,dz.$$

Таким образом, *если распределение вероятностей x нормально* (т.-е. если распределение вероятностей x удовлетворяет закону Гаусса), *то распределение вероятностей всякой линейной функции $z=A+\lambda(x-a)$ удовлетворяет также закону Гаусса; при этом, центром распределения является* М. О. $z=A$, *а штандартом служит штандарт величины z*, равный $\sqrt{\text{М. О.}(z-A)^2}=|\lambda|\sigma$.

Предположим, например, что игрок повторяет n раз игру, при которой он может выиграть A рублей с вероятностью p и проиграть B рублей с вероятностью q, так что $\nu=Ap-Bq$ есть ма-

тематическое ожидание его выигрыша в каждой партии. Определим закон распределения вероятностей для его суммарного выигрыша nX. Очевидно, что $nX = Am - B(n - m) = (A + B)m - Bn = (A + B)(m - np) + Anp + Bnp - Bn = (A + B)(m - np) + n\nu$. Поэтому, принимая во внимание, что $m - np$ подчиняется закону Гаусса с штандартом $\sigma = \sqrt{npq}$, заключаем, что кривая распределения для nX также нормальна и имеет центром $n\nu$ и штандартом $(A + B)\sqrt{npq}$, т.-е. вероятность, что

$$z_0 < nX < z,$$

равна

$$\frac{1}{(A+B)\sqrt{2\pi npq}} \int_{z_0}^{z_1} e^{-\frac{(z-n\nu)^2}{2npq(A+B)^2}} dz,$$

или, иначе, вероятность, что

$$t_0 (A + B)\sqrt{npq} < nX - n\nu < t_1 (A + B)\sqrt{npq}$$

равна

$$\frac{1}{\sqrt{2\pi}} \int_{t_0}^{t_1} e^{-\frac{t^2}{2}} dt.$$

Напротив, если функция $\varphi(x)$ не линейна, то кривая распределения уже не будет нормальна. Пусть, например,

$$z = (x - a)^3,$$

так что $x = a + \sqrt[3]{z}$ и $\varphi'(z) = \frac{1}{3z^{2/3}}$; следовательно, вероятность неравенства

$$z_0 < z < z_1$$

равна

$$\frac{1}{3\sigma\sqrt{2\pi}} \int_{z_0}^{z_1} \frac{e^{-\frac{z^{\frac{2}{3}}}{2\sigma^2}}}{z^{2/3}} dz.$$

В данном примере производная $\varphi'(x) = 3(x - a)^2$ обращается в нуль при $x = a$; если же это обстоятельство не имеет места и $\varphi'(a) = \lambda \gtrless 0$, то в тех случаях, когда значение штандарта σ мало, так что значительные отклонения $x - a$ невероятны, можно заменять приближенно любую функцию

$$z = \varphi(x) = \varphi(a) + \varphi'(a)(x - a) + \frac{\varphi''(a)}{2}(x - a)^2 +$$

первыми ее двумя членами и полагать

$$z = \varphi(x) \neq \varphi(a) + \varphi'(a)(x - a).$$

В таком случае **приближенным законом распределения вероятностей для** z **будет снова нормальная кривая с центром** $A = \varphi(a)$ **и с штандартом** $\Sigma = \sigma|\varphi'(a)|$.

3. Перейдем теперь к рассмотрению функции $z = \varphi(x, y)$ двух переменных, ограничиваясь случаем, когда эти последние **независимы** между собой, при чем распределение вероятностей каждой из них подчиняется закону Гаусса.

Предположим сначала, что z есть линейная функция, которую можем представить в виде

$$z = A + \lambda(x - a) + \lambda_1(y - a_1),$$

где $a =$ М. О. x и $a_1 =$ М. О. y. Полагая, что σ есть штандарт x, а σ_1 есть штандарт y, видим, что М. О. $z = A$, а **штандарт** $\sigma(z)$ **величины** z **равен**

$$\sigma(z) = \sqrt{\text{М. О.}(z - A)^2} = \sqrt{\lambda^2\sigma^2 + \lambda_1^2\sigma_1^2},$$

так как М. О. $(x - a)(y - a_1) = 0$, в силу независимости x и y.

Покажем, что **распределение вероятностей величины** z **также подчиняется закону Гаусса**; в силу вышеизложенного, если это утверждение правильно, то центром кривой распределения будет A, а штандарт будет равен $\sqrt{\lambda^2\sigma^2 + \lambda_1^2\sigma_1^2}$. Поэтому, заметив, что $t = \lambda(x - a)$ и $t_1 = \lambda_1(y - a_1)$ подчиняются закону Гаусса, с соответственными штандартами $k = |\lambda|\sigma$ и $k_1 = |\lambda_1|\sigma_1$, достаточно будет проверить, что

$$z = t + t_1$$

удовлетворяет закону Гаусса, когда кривые распределения вероятностей каждой из независимых величин t и t_1 нормальны и имеют центром 0 (так как если z удовлетворяет закону Гаусса, то и $z + A$ ему удовлетворяет).

Итак, по предположению, вероятность, что t находится в бесконечно малом интервале $(t, t + dt)$, равна

$$\frac{e^{-\frac{t^2}{2k^2}}}{k\sqrt{2\pi}}\,dt,$$

и, независимо от этого, вероятность, что t_1 находится в промежутке (t_1, t_1+dt_1), равна

$$\frac{e^{-\frac{t_1^2}{2k_1^2}}}{k_1\sqrt{2\pi}}dt_1.$$

Поэтому вероятность совмещения этих двух фактов, т.-е. вероятность того, что точка с прямоугольными координатами t и t_1 находится внутри бесконечно малого прямоугольника с вершинами (t, t_1), (t, t_1+dt_1), $(t+dt, t_1)$, $(t+dt, t_1+dt_1)$ равна

$$\frac{e^{-\frac{t^2}{2k^2}-\frac{t_1^2}{2k_1^2}}}{2\pi k k_1}dt\,dt_1,$$

а потому вероятность, что точка с абсциссой t и ординатой t_1 окажется внутри данной площади S, выразится двойным интегралом, распространенным на эту площадь:

$$P_S=\frac{1}{2\pi k k_1}\iint\limits_S e^{-\frac{t^2}{2k^2}-\frac{t_1^2}{2k_1^2}}dt\,dt_1. \tag{188}$$

Нас интересует вероятность неравенства

$$z_0<z=t+t_1<z_1,$$

где z_1 и z_0 — данные числа. Геометрически говоря, это значит, что мы хотим определить вероятность того, что точка с координатами (t, t_1) окажется внутри полосы S, заключенной между двумя параллельными прямыми $t+t_1=z_0$ и $t+t_1=z_1$. Для вычисления интеграла P_S внутри этой полосы S делаем замену одной из переменных $t_1=z-t$, так что

$$P_S=\frac{1}{2\pi k k_1}\int\limits_{-\infty}^{\infty}\int\limits_{z_0}^{z_1} e^{-\frac{t^2}{2k^2}-\frac{(z-t)^2}{2k_1^2}}dt\,dz=$$

$$=\frac{1}{2\pi k k_1}\int\limits_{-\infty}^{\infty}\int\limits_{z_0}^{z_1} e^{-\frac{t^2(k^2+k_1^2)-2k^2tz+k^2z^2}{2k^2k_1^2}}dt\,dz=$$

$$=\frac{1}{2\pi k k_1}\int\limits_{-\infty}^{\infty}\int\limits_{z_0}^{z_1} e^{-\frac{k^2+k_1^2}{2k^2k_1^2}\left(t-\frac{k^2z}{k+k_1^2}\right)^2-\frac{z^2}{2(k^2+k_1^2)}}dt\,dz;$$

полагая затем $t - \frac{k^2 z}{k^2 + k_1^2} = v$, получаем наконец

$$P_S = \frac{1}{2\pi k k_1} \int_{-\infty}^{\infty} e^{-\frac{k^2 + k_1^2}{2k^2 k_1^2} v^2} dv \int_{z_0}^{z_1} e^{-\frac{z^2}{2(k^2 + k_1^2)}} dz =$$

$$= \frac{1}{\sqrt{2\pi \left(k^2 + k_1^2\right)}} \int_{z_0}^{z_1} e^{-\frac{z^2}{2(k^2 + k_1^2)}} dz.$$

Следовательно, вероятность неравенства

$$z_0 < z < z_1$$

равна

$$\frac{1}{\sqrt{2\pi \left(k^2 + k_1^2\right)}} \int_{z_0}^{z_1} e^{-\frac{z^2}{2(k^2 + k_1^2)}} dz,$$

так что распределение вероятностей для величины z нормально с штандартом $\sqrt{k^2 + k_1^2}$.

Из того, что сумма z двух величин t и t_1, подчиняющихся закону Гаусса, также удовлетворяет закону Гаусса, следует, что вообще, если

$$z = t_1 + t_2 + \cdots + t_n$$

есть сумма любого числа независимых величин, подчиняющихся закону Гаусса, то распределение вероятностей величины z также нормально.

Действительно, полагая, что наше утверждение справедливо для суммы $n - 1$ слагаемых $t_1 + t_2 + \ldots + t_{n-1}$, заключаем, что оно правильно и для n слагаемых, так как

$$z = (t_1 + t_2 + \ldots + t_{n-1}) + t_n$$

можно рассматривать как сумму двух независимых величин, подчиняющихся закону Гаусса.

Очевидно таким образом, что вообще

$$z = A + \lambda_1 (x_1 - a_1) + \lambda_2 (x_2 - a_2) + \ldots + \lambda (x_n - a_n)$$

удовлетворяет закону Гаусса с центром A и штандартом

$$\sigma(z) = \sqrt{\lambda_1^2 \sigma_1^2 + \ldots + \lambda_n^2 \sigma_n^2}, \tag{189}$$

если кривые распределения независимых величин $x_1, x_2, \ldots, x_n$ нормальны и имеют соответственно центры $a_1, a_2, \ldots, a_n$ и штандарты $\sigma_1, \sigma_2, \ldots, \sigma_n$.

Если z нелинейная функция n независимых величин $x_1, x_2, \ldots x_n$, то вообще распределение ее вероятностей ненормально. Однако, предполагая функцию $z = \varphi(x_1, x_2, \ldots, x_n)$ разлагающейся в строку Тэйлора, можно часто в первом приближении (как мы это сделали в случае функции одной переменной), пренебрегать степенями $(x_i - a_i)$ выше первой. В таком случае,

$$z = \varphi(a_1, a_2, \ldots, a_n) + \varphi'_{a_1}(x_1 - a_1) + \ldots + \varphi'_{a_n}(x_n - a_n);$$

поэтому можно принять в качестве приближенного закона распределения вероятностей величины z кривую Гаусса с центром $A = \varphi(a_1, a_2, \ldots, a_n)$ и с штандартом

$$\sigma(z) = \sqrt{(\varphi'_{a_1})^2 \sigma_1^2 + \ldots + (\varphi'_{a_n})^2 \sigma_n^2}. \qquad (189\ \text{bis})$$

Пусть, например,

$$z = \frac{x}{y},$$

где М. О. $x = a$, М. О. $(x - a)^2 = \sigma^2$, М. О. $y = a_1$, М. О. $(y - a_1)^2 = \sigma_1^2$, при чем a_1 настолько велико по сравнению с σ_1, что значения $y \leqslant 0$ являются фактически невозможными. В таком случае, в первом приближении (более точная формула указана в конце гл. II, пятой части) закон распределения вероятностей z считают нормальным, при центре и штандарте, соответственно равных

$$A = \frac{a}{a_1}, \quad \sigma(z) = \frac{1}{a_1}\sqrt{\sigma^2 + \sigma_1^2 \frac{a^2}{a_1^2}} \qquad (190)$$

согласно формуле (189 bis).

Полагая, в частности, $x = m$, $y = m_1$, где m и m_1 представляют числа появлений события B, имеющего вероятность p в двух независимых сериях опытов, по n опытов в каждой, мы получили бы

$$A = 1 \text{ и } \sigma\left(\frac{m}{m_1}\right) = \sqrt{\frac{2q}{np}}, \qquad (190\ \text{bis})$$

так как $\sigma^2 = \sigma_1^2 = npq$, $a = a_1 = np$.

Теоретически формулы (190) и (190 bis) конечно неправильны [1]), так как ясно, например, что значение $\frac{m}{m_1}=\infty$, соответствующее $m_1=0$, не лишено возможности, а поэтому М. О. $\frac{m}{m_1}=\infty$. Для теоретической точности и во избежание иногда грубых ошибок нужно помнить, что формула (190) в действительности соответствует не функции $z=\frac{x}{y}$, а функции $z_1=\frac{x}{a_1}-\frac{(y-a_1)a}{a_1^2}$, которые получаются, если в разложении z в строку Тэйлора по степеням $(x-a)$ и $(y-a_1)$ отбросить члены выше первой степени.

Из последних замечаний следует, что закон Гаусса, или нормального распределения вероятностей, не имеет универсального характера, и нужно признать, что нередко его прилагают без достаточных оснований главным образом ради его математической простоты. Тем более важно указать некоторые общие условия (аналогичные теореме Лапласа), при которых закон Гаусса применим.

4. *Теорема Ляпунова.* Пусть

$$z=x_1+x_2+\ldots+x_n,$$

где $x_1, x_2, \ldots, x_n$ — независимые величины, при чем М. О. $x_i=a_i$, М. О. $(x_i-a_i)^2=\beta_i$, М. О. $|x_i-a_i|^{2+\varepsilon}=\delta_i$, где ε — некоторое определенное положительное число; в таком случае вероятность неравенства

$$t_0\sqrt{B}<z-\sum a_i<t_1\sqrt{B}, \tag{153}$$

---

[1]) Заметим, в частности, что если $a=a_1$, $\sigma=\sigma_1$, то распределение вероятностей для z и $\frac{1}{z}$ будет, очевидно, тождественно; поэтому, за исключением случая, когда $a=a_1=0$, распределение z не может быть симметрично по отношению к $A=\frac{a}{a_1}=1$, которое будет медианой (см. следующую часть, гл. II) распределения, но не центром. Вообще в данном случае М. О. z лишено смысла, а тем более М. О. z^2. Читатель может проверить в виде упражнения, что при $a=a_1=0$ функция распределения $f(z)=\frac{1}{\pi}\cdot\frac{1}{1+z^2}$.

где $B = \sum \beta_i$, имеет пределом $\frac{1}{\sqrt{2\pi}} \int_{t_0}^{t_1} e^{-\frac{t^2}{2}} dt$ при неограниченном возрастании n, если

$$\frac{\delta_1 + \delta_2 + \ldots + \delta_n}{B^{1+\frac{\varepsilon}{2}}}$$

стремится к 0.

Последнее условие означает, по существу, что среди рассматриваемых величин x_i не должно быть таких, которые были бы значительно больше большинства остальных.

Предположим, например, что $M =$ максим. $|x_i - a_i|$; в таком случае, для применимости теоремы Ляпунова достаточно, чтобы [1] $\frac{M}{\sqrt{B}}$ стремилось к 0 с возрастанием n. Действительно, $\delta_i =$ М. О. $|x_i - a_i|^\varepsilon (x_i - a_i)^2 \leqslant \beta_i M^\varepsilon$, следовательно, $\sum \delta_i \leqslant BM^\varepsilon$; поэтому $\frac{\delta_1 + \delta_2 + \ldots + \delta_n}{B^{1+\frac{\varepsilon}{2}}} \leqslant \left(\frac{M}{\sqrt{B}}\right)^\varepsilon$ стремится к 0.

Напротив, если бы среди наших величин была одна x_1, которая принимала бы лишь два значения $x_1 = a_1 \pm \frac{1}{2}\sqrt{B}$ того же порядка, что и штандарт $\sqrt{B}$ всей суммы, то понятно, что вмешательство этой величины искажало бы результат совместного действия прочих величин: присутствие такого рода слагаемых не допускается теоремой Ляпунова, так как тогда оказалось бы, что

$$\delta_1 = \text{М. О. } |x_1 - a_1|^{2+\varepsilon} = \left(\frac{1}{2}\sqrt{B}\right)^{2+\varepsilon}$$

а потому $\frac{\delta_1}{B^{1+\frac{\varepsilon}{2}}} = \left(\frac{1}{2}\right)^{2+\varepsilon}$ не стремилось бы к нулю; при таких условиях закон Гаусса не может быть применим.

---

[1]) Это условие, очевидно, соблюдено, если существует определенное положительное число h, обладающее свойством, что все $\beta_i > hM^2$, так как тогда $B > nhM^2$, откуда $\frac{M}{\sqrt{B}} < \frac{1}{\sqrt{nh}}$ стремится к 0 с возрастанием n.

В частности, из теоремы Ляпунова вытекает следующее следствие.

Если имеем ряд независимых опытов, при которых вероятности появления события A равны соответственно $p_1, p_2, \ldots, p_n$, где $\sum_1^n p_i q_i$ бесконечно возрастает с увеличением n, то вероятность неравенства

$$t_0 \sqrt{\sum p_i q_i} < m - \sum p_i < t_1 \sqrt{\sum p_i q_i} \qquad (153 \text{ bis})$$

имеет пределом, при бесконечном n, $\Phi(t_1) - \Phi(t_0)$. Указанное условие, очевидно, соблюдено, если ни в одном опыте вероятность p_i не приближается [1]) безгранично ни к 0, ни к 1.

Действительно, в данном случае неравенство (153) обращается в (153 bis). Кроме того, полагая $\varepsilon = 1$, имеем

$$\delta_i = \text{М. О.} |x_i - a_i|^3 = \text{М. О.} |x_i - p_i|^3 =$$
$$= (1 - p_i)^3 p_i + p_i^3 q_i = p_i q_i (q_i^2 + p_i^2) < p_i q_i.$$

Поэтому

$$\frac{\delta_1 + \delta_2 + \ldots + \delta_n}{B^{3/2}} < \frac{\sum p_i q_i}{\left(\sum p_i q_i\right)^{3/2}} < \frac{1}{\left(\sum p_i q_i\right)^{1/2}}$$

(где $\sum p_i q_i$ бесконечно возрастает) стремится к 0 с возрастанием n.

Пусть, например, из двух урн с шарами производится по 15 000 извлечений по одному шару, при чем вероятность появления бе-

---

[1]) Предельная теорема может быть справедлива и в том случае, когда вероятность p_i стремится к 0 или к 1: если π_i есть наименьшее из двух чисел p_i и $q_i = 1 - p_i$, то условие, что $\sum_1^n p_i q_i$ безгранично возрастает, равнозначно условию, что $\sum_1^n \pi_i$ бесконечно возрастает, так

$$\frac{1}{2} \sum \pi_i < \sum p_i q_i < \sum \pi_i.$$

лого шара из первой урны все время равна $\frac{3}{4}$, а из второй — равна $\frac{1}{4}$. В таком случае математическое ожидание общего числа вынутых белых шаров m будет равно 15 000 так же, как если бы все 30 000 извлечений производились бы из одной урны с вероятностью $p=\frac{1}{2}$ появления белого шара. Но разница будет та, что в рассматриваемом случае дисперсия поднормальна, $B=15\,000\cdot\frac{3\cdot 1}{16}+$ $+15\,000\cdot\frac{3\cdot 1}{16}=5\,625$ вместо $30\,000\cdot\frac{1}{4}=7\,500$ в случае Бернулли. И в том и в другом случае уклонение m от 15 000 будет подчиняться закону Гаусса [1]), но при различных урнах штандарт равен $\sqrt{5\,625}=75$, а для одной штандарт будет $\sqrt{7\,500}\neq 86{,}6$.

5. В качестве второго следствия из теоремы Ляпунова отметим следующее предложение.

Если М. О. $x=a$, то средняя арифметическая $X=\frac{x_1+x_2+\ldots+x_n}{n}$ из n наблюденных значений $x_1, x_2, \ldots, x_n$ величины x подчиняется (для достаточно большого n) закону Гаусса с штандартом $\frac{\sigma}{\sqrt{n}}$, если штандарт x равен σ (хотя бы величина x сама закону Гаусса не подчинялась), лишь бы только существовало некоторое положительное число ε такое, что $\delta=$ М. О. $|x-a|^{2+\varepsilon}$ не бесконечно.

Действительно, в данном случае неравенство (153) получает форму

$$t_0\sqrt{n\sigma}<nX-na<t_1\sqrt{n\sigma},$$

равнозначную

$$t_0\sqrt{\frac{\sigma}{n}}<X-a<t_1\sqrt{\frac{\sigma}{n}}, \qquad (191)$$

---

[1]) Заметим, что для данного примера применимость закона Гаусса и значение штандарта вытекают из формулы (189), так как, по теореме Лапласа, число появлений белого шара из каждой урны в отдельности подчиняется (в пределе) закону Гаусса, а потому и сумма m этих двух чисел подчиняется нормальному закону распределения вероятностей.

а выражение

$$\frac{\delta_1+\delta_2+\ldots+\delta_n}{B^{1+\frac{\varepsilon}{2}}}$$

обращается в

$$\frac{n\delta}{(n\sigma)^{1+\frac{\varepsilon}{2}}}=\frac{\delta}{\sigma^{1+\frac{\varepsilon}{2}}}\cdot\frac{1}{n^{\frac{\varepsilon}{2}}}$$

и стремится к 0 с возрастанием n.

Ограничительное условие относительно δ для практики мало стеснительно, так как оно могло бы оказаться нарушенным только тогда, когда безгранично возрастающие значения $|x-a|$ были бы фактически возможны. Поэтому указанное следствие имеет чрезвычайно большое практическое значение и в дальнейшем будет нами неоднократно применяемо.

Вообще закон Гаусса более или менее точно соблюдается во всех случаях, когда исследуемую величину можно рассматривать как сумму большого числа независимых слагаемых. Таким образом, если при измерении какой-нибудь величины a то или иное уклонение наблюдаемого значения z от истинной величины a получается в результате воздействия многочисленных причин, из которых каждая является источником независимой малой ошибки x_i, при чем М. О. $x_i=0$, то суммарная ошибка, или уклонение $z-a$, подчиняется закону Гаусса, носящему поэтому также название закона случайных ошибок.

Это же обстоятельство является основанием того, что при стрельбе в цель уклонение вправо и влево от намеченной цели с значительной точностью подчиняется закону Гаусса так же, как и вертикальное уклонение, которое оказывается независимым от горизонтального уклонения, при чем штандарты $\sigma(x)=\sigma_1$ и $\sigma(y)=\sigma_2$ обычно различны. Применяя теорему умножения вероятностей и принимая намеченную цель за начало координат, мы находим таким образом, что вероятность точке попадания оказаться внутри прямоугольника, заключенного между вертикалями $(x, x+dx)$ и горизонталями $(y, y+dy)$, равна

$$\frac{dx\,dy}{2\pi\sigma_1\sigma_2}e^{-\frac{x^2}{2\sigma_1^2}-\frac{y^2}{2\sigma_2^2}}$$

а вероятность попасть внутрь площади S равна поэтому двойному интегралу, взятому по площади S:

$$\frac{1}{2\pi\sigma_1\sigma_2}\iint_S e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_1^2}+\frac{y^2}{\sigma_2^2}\right)}dx\,dy. \qquad (188\ \text{bis})$$

В частности, если S есть площадь эллипса

$$\frac{x^2}{\sigma_1^2}+\frac{y^2}{\sigma_2^2}=R^2,$$

то интеграл (188 bis) после замены переменных $x=\sigma_1\rho\cos\theta$, $y=\sigma_2\rho\sin\theta$ преобразуется в

$$I=\frac{1}{2\pi}\int_0^{2\pi}\int_0^R e^{-\frac{\rho^2}{2}}\rho\,d\rho\,d\theta=\int_0^R e^{-\frac{\rho^2}{2}}\rho\,d\rho=1-e^{-\frac{R^2}{2}}.$$

Вероятный эллипс попадания (внутрь которого при многократной стрельбе попадает около $\frac{1}{2}$ всех снарядов) соответствует значению $I=\frac{1}{2}$, т.-е. $e^{-\frac{R^2}{2}}=\frac{1}{2}$, откуда $R^2\neq 1{,}39$ (полуоси этого эллипса равны $R\sigma_1\neq 1{,}17\sigma_1$ и $R\sigma_2\neq 1{,}17\sigma_2$).

6. Особенно важно приложение закона Гаусса к биологии. К этому вопросу мы еще вернемся в дальнейшем. Но для этого необходимо заметить теперь же, что теорема Ляпунова может во многих случаях быть распространена и на зависимые величины, при условии, что связь между величинами $x_1, x_2, \ldots, x_k, \ldots, x_n$ постепенно ослабевает по мере удаления их друг от друга, т.-е. при возрастании $(i-k)$. В данном случае дело представляется несколько сложнее, чем при распространении закона больших чисел. Мы ограничимся лишь формулировкой одной теоремы, из которой будет виден характер ограничений несколько иной, чем при обобщении теоремы Чебышева. Для большей отчетливости предположим величины x_i ограниченными, т.-е. $|x_i|<M$, а их математические ожидания — равными нулю.

Теорема. Если $\lambda g<$ М. О. $[x_{i+1}+\ldots+x_{i+g}]^2<Lg$, где L и λ данные положительные числа (т.-е. дисперсия суммы любого числа слагаемых $x_{i+1},\ldots,x_{i+g}$ того же порядка g, как если бы слагаемые были независимы), то закон Гаусса при n

весьма большом применим к сумме $z=\sum_1^n x_i$, лишь бы М. О. x_k и М. О. $x_k x_l$ не претерпевали бы изменений больших, чем $\frac{1}{n}$, каковы бы ни были известные уже значения каких-нибудь из прочих величин, до тех пор, пока неизвестно ни одно из x_i, для которого $|k-i|<n^\rho$ или $|l-i|<n^\rho$, где $\rho<\frac{1}{2}$ — данное положительное число. (Что касается взаимоотношений между более близкими x_k и x_i, то они ничем неограничены.)

Существенную роль здесь играет показатель $\rho<\frac{1}{2}$, который ограничивает сферу интенсивного влияния каждой величины на все прочие. Важно заметить, что, если бы $\rho=\frac{1}{2}$, то закон Гаусса вообще не был бы применим. Чтобы лучше это уяснить, возьмем грубо упрощенный пример.

Положим, что игрок повторяет n безобидных партий, при которых вероятность выиграть или проиграть по 1 рублю равна $\frac{1}{2}$. Допустим, что первые $\sqrt{n}$ партий связаны так, что, выигравши первую партию, игрок выиграет и все указанные $\sqrt{n}$ партий, которые он, напротив, проигрывает в случае первой неудачи; остальные же партии пусть будут независимы. Таким образом в нашем примере есть партии, сферы влияния которых определяются показателем $\rho=\frac{1}{2}$, и этого уже достаточно, чтобы закон Гаусса оказался неприменим. Действительно, мы можем выигрыш z нашего игрока представить в виде суммы

$$z=x+y,$$

где x получает лишь 2 значения $\pm\sqrt{n}$ (в зависимости от исхода первой партии), вероятности которых, следовательно, равны $\frac{1}{2}$. Напротив, сумма y, соответствующая прочим $n-\sqrt{n}$ партиям, подчиняется закону Гаусса с штандартом $\sqrt{n-\sqrt{n}}$ (при центре 0). Поэтому, так как неравенство

$$z_0<z<z_1 \tag{192}$$

имеет место в двух случаях: $x=\sqrt{n}$ и $z_0-\sqrt{n}<y<z_1-\sqrt{n}$, либо $x=-\sqrt{n}$ и $z_0+\sqrt{n}<y<z_1+\sqrt{n}$, заключаем, что вероятность неравенства (192) равна

$$\frac{1}{2}\cdot\frac{1}{\sqrt{2\pi(n-\sqrt{n})}}\left[\int_{z_0-\sqrt{n}}^{z_1-\sqrt{n}} e^{-\frac{z^2}{2(n-\sqrt{n})}}dz+\right.$$

$$\left.+\int_{z_0+\sqrt{n}}^{z_1+\sqrt{n}} e^{-\frac{z^2}{2(n-\sqrt{n})}}dz\right]=$$

$$=\frac{1}{2\pi(n-\sqrt{n})}\int_{z_0}^{z_1}\left[e^{-\frac{(z-\sqrt{n})^2}{2(n-\sqrt{n})^2}}+e^{-\frac{(z+\sqrt{n})^2}{2(n-\sqrt{n})^2}}\right]dz,$$

т.-е. кривая распределения вероятностей для z распадается на 2 кривые Гаусса с разными центрами [1]).

Указанная выше теорема включает в себя, в частности, опыты, образующие простую цепь, к которым, как доказал А. А. Марков, применима предельная теорема теории вероятностей. Таким образом, сохраняя обозначения гл. III (часть 3-я, стр. 194), при неограниченном возрастании n, вероятность неравенства

$$t_0\sigma<m-np<t_1\sigma, \tag{193}$$

где

$$\sigma=\sqrt{npq\frac{1+\alpha-\beta}{1+\beta-\alpha}}$$

имеет пределом

$$\frac{1}{\sqrt{2\pi}}\int_{t_0}^{t_1} e^{-\frac{t^2}{2}}dt.$$

А. А. Марков применил свою теорему к исследованию вопроса о частоте гласных и согласных в русском языке, произведя соответствующие подсчеты сначала для „Евгения Онегина", а потом для сочинения С. Т. Аксакова „Детские годы Багрова-внука". Приведем некоторые данные, полученные А. А. Марковым при рассмотрении 100000 букв в последнем сочинении.

Прежде всего, общее число m гласных оказалось равно $M=44\,898$ при $N=100\,000$.

---

[1]) Разумеется, закон больших чисел в данном примере все-таки применим, и средний выигрыш при большом n, стремится к 0.

Далее, беря $s = 1\,000$ групп по $n = 100$ последовательных букв, А. А. Марков получил следующие результаты:

Число m гласных	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52
Число h групп	1	2	5	21	33	69	123	163	196	171	109	50	35	10	10	2
$m - 45$	−8	−7	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	7
$(m - 45)^2$	64	49	36	25	16	9	4	1	0	1	4	9	16	25	36	49
$h(m - 45)^2$	64	98	180	525	528	621	492	163	0	171	436	450	560	250	360	98

Отсюда $\frac{\sum h(m-45)^2}{1\,000} = 4{,}996$. Поэтому среднее квадратичное уклонение, т.-е. $\frac{\sum h(m-np)^2}{1\,000} = \frac{\sum h(m-44{,}9)^2}{1\,000} = 4{,}996 - 0{,}01 \neq 5$. С другой стороны, $npq = 100 \cdot 0{,}449 \cdot 0{,}551 \neq 24{,}74$. Следовательно, эмпирический коэффициент дисперсии

$$E = \frac{5}{24{,}74} \neq 0{,}202,$$

и дисперсия явно поднормальная.

Этот результат вполне подтверждает теорию Маркова, согласно которой последовательности букв не являются независимыми, но образуют цепь, которую, в первом приближении, можно считать простой, полагая, что вероятность данной буквы быть гласной вполне определяется, если известно, какого рода буква (гласная или согласная) ей непосредственно предшествует, независимо от того, каковы более отдаленные предшествующие ей буквы. Статистические вероятности α (появления гласной после гласной) и β (появления гласной после согласной) Марков получает, подсчитывая число A пар (гласная, гласная) и число B пар (согласная, гласная). Оказалось, что $A = 6\,588$ и $B = 38\,310$. В таком случае

$$\alpha \neq \frac{A}{M} \neq \frac{6\,588}{44\,898} \neq 0{,}147 \quad \text{и} \quad \beta = \frac{B}{N-M} = \frac{38\,310}{55\,102} \neq 0{,}695.$$

Поэтому коэффициент корреляции между буквами одинакового рода $R = \alpha - \beta \neq -0{,}548$, откуда $\frac{1+\alpha-\beta}{1-\alpha+\beta} = \frac{0{,}452}{1{,}548} \neq 0{,}29$.

Таким образом вышеуказанный эмпирический коэффициент дисперсии 0,202 несколько мал даже для делаемой Марковым гипотезы простой цепи; в действительности, как замечает Марков, по поводу того, что и для „Евгения Онегина“ эмпирический коэффициент дисперсии оказался 0,208 вместо теоретического 0,3, надо было бы принять в расчет и предшествующую букву, т.-е. цепь последовательных букв, в сущности, не простая [1]).

[1]) Если бы цепь была простая, то вероятность 3 буквам под-ряд быть гласными была бы равна $\frac{M}{N}\alpha^2 = 0{,}065\,88 \cdot 0{,}147 \neq 0{,}0097$; таким образом оказалось бы, что во всей совокупности 100 000 букв следовало бы ожидать встретить около 1 000 последовательностей из 3 гласных — несомненно, что такого рода неблагозвучное сочетание произойдет гораздо реже.

То обстоятельство, что штандарт числа m появлений гласной в каждой группе из 100 букв не может быть равен $\sqrt{npq} = \sqrt{24{,}74} \neq 4{,}97$ (который соответствовал бы независимости последовательных букв), легко обнаружить, если заметить, что вероятное отклонение было бы равно $\frac{2}{3}\sigma = 3{,}32$; поэтому приблизительно в 500 группах число гласных было бы заключено между $44{,}9 \pm 3{,}32$, т.-е. оказалось бы не менее 42 и не более 48 гласных; между тем, указанным свойством обладает 881 группа ($881 = 69 + 123 + 163 + 196 + 171 + 109 + 50$). Вероятность, что при $s = 1\,000$ независимых опытов событие, вероятность которого равна $\frac{1}{2}$, произойдет 881 раз, т.-е., что $|m_s - sp| > 380 = t\sqrt{250}$, меньше, на основании (113) и (114), чем $2e^{-\frac{t^2}{2}} = 2e^{-288{,}8}$, т.-е. практически равна 0.

Упражнения. 1) Определить абсциссу $x_0 > 0$ точки пересечения двух нормальных кривых $y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$ и $y = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_1^2}}$; показать, что x_0 заключено между σ и σ_1.

Отв. $x_0 = \sigma\sigma_1\sqrt{\frac{2\log\frac{\sigma}{\sigma_1}}{\sigma^2 - \sigma_1^2}}$. Полагая $\sigma > \sigma_1$, можно написать:

$$x_0 = \sigma\sqrt{\frac{2\log\rho}{\rho^2 - 1}} = \sigma_1\sqrt{\frac{2\log\rho}{1 - \frac{1}{\rho^2}}},$$ где $\rho = \frac{\sigma}{\sigma_1} > 1$, откуда видно, что $\sigma > x_0 > \sigma_1$, так как $\rho^2 - 1 > 2\log\rho > 1 - \frac{1}{\rho^2}$.

2) Определить условия, необходимые и достаточные для того, чтобы при n весьма большом сумма $\sum x_k$ независимых величин $x_1, x_2, \cdots, x_n$ подчинялась закону Гаусса, если x_k может получать с одинаковой вероятностью лишь два значения $\pm k^\lambda$, где λ — данное число.

Отв. Необходимо и достаточно, чтобы $\lambda \geqslant -\frac{1}{2}$. Действительно, если $\lambda < -\frac{1}{2}$, то $B = \text{М. О.}\left(\sum x_k\right)^2 = \sum_1^n k^{2\lambda}$ стремится к конечному пределу при неограниченном возрастании n; с другой стороны, $\text{М. О.}\left(\sum x_k\right)^4 = \text{М. О.}\sum x_k^4 + 6\,\text{М. О.}\sum x_i^2 x_k^2 = 3\,\text{М. О.}\left(\sum x_k^2\right)^2 - 2\,\text{М. О.}\sum x_k^4 = 3B^2 - 2\,\text{М. О.}\sum x_k^4$, а потому $\text{М. О.}\left(\sum x_k\right)^4$ не имеет пределом $3B^2$, как это имеет место при нормальном распределении вероятностей. Итак, условие $\lambda \geqslant -\frac{1}{2}$ во всяком случае необходимо, но достаточность его следует из теоремы Ляпунова. Действительно, $\text{М. О.}\left|x_k^{2+\varepsilon}\right| = \delta_k = k^{\lambda(2+\varepsilon)}$; поэтому $\sum_1^n \delta_k \backsim \frac{n^{\lambda(2+\varepsilon)+1}}{\lambda(2+\varepsilon)+1}$, при $\lambda > -\frac{1}{2}$, между тем как $B = \sum_1^n k^{2\lambda} \backsim \frac{n^{2\lambda+1}}{2\lambda+1}$.

Следовательно, $\dfrac{\sum\limits_1^n \delta_k}{B^{1+\frac{\varepsilon}{2}}} \backsim \dfrac{(2\lambda+1)^{1+\frac{\varepsilon}{2}}}{\lambda(2+\varepsilon)+1} \cdot \dfrac{1}{n^{\frac{\varepsilon}{2}}}$ стремится к 0 с возрастанием n.

Если же $\lambda = -\dfrac{1}{2}$, то теорема Ляпунова также применима, так как при этом B бесконечно возрастает, а $\sum\limits_1^n \delta_k$ остается ограниченной.

Полученный результат может быть использован для нового доказательства того, что вероятность неравенства

$$|X| = \left|\frac{x_1 + x_2 + \ldots + x_n}{n}\right| < \varepsilon,$$

где ε — данное произвольно малое число, стремится к достоверности, при $n \to \infty$ тогда и только тогда, когда $\lambda < \dfrac{1}{2}$; иными словами, для применимости закона больших чисел к величинам $x_1, x_2, \ldots, x_n$ необходимо и достаточно, чтобы $\lambda < \dfrac{1}{2}$. Так как достаточность этого условия (стр. 167) не вызывает сомнений, рассмотрим лишь предположение $\lambda \geqslant \dfrac{1}{2}$, при котором X подчиняется закону Гаусса с штандартом $\dfrac{1}{n}\sqrt{B} \backsim \sqrt{\dfrac{n^{2\lambda-1}}{2\lambda+1}}$, не стремящимся к 0 с возрастанием n; отсюда ясно, что если ε мало, то и вероятность $|X| < \varepsilon$ близка к 0, а не к 1. Например, при $\lambda = \dfrac{1}{2}$ вероятность неравенства

$$\left|\frac{\pm 1 \pm \sqrt{2} \pm \sqrt{3} \pm \cdots \pm \sqrt{n}}{n}\right| < \frac{1}{\sqrt{2}}$$

имеет пределом $\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-1}^{+1} e^{-\frac{t^2}{2}}\,dt \neq \dfrac{2}{3}$. (См. конец главы II, часть 3-я.)

3) Определить штандарт $\sigma\left(\dfrac{x}{y}\right)$ с точностью того же порядка, что формула (190), если величины y и x не независимы и коэффициент корреляции между ними равен R.

Отв. $\sigma\left(\dfrac{x}{y}\right) = \dfrac{1}{a_1}\sqrt{\sigma^2 - 2R\sigma\sigma_1\dfrac{a}{a_1} + \sigma_1^2\left(\dfrac{a}{a_1}\right)^2}$. Эта формула есть следствие из общей формулы $\sigma[\varphi(x, y)] = \sqrt{\varphi'^2_x\sigma^2 + 2R\varphi'_x\varphi'_y\sigma\sigma_1 + \varphi'^2_y\sigma_1^2}$, которая получается тем же способом, что и (189 bis).

ГЛАВА ЧЕТВЕРТАЯ

ВЫБОРОЧНЫЙ МЕТОД В СТАТИСТИКЕ

1. Если требуется определить приближенно, какое количество объектов данной вполне определенной совокупности, состоящей из весьма большого числа N объектов, обладает определенным признаком A, то для этого с успехом может быть применен так называемый выборочный метод. В основе этого метода лежит теорема Лапласа или обобщающая ее теорема, относящаяся к схеме урны с невозвращаемыми шарами.

Предположим, что каждому объекту совокупности соответствует определенная карточка и все N карточек помещены в урну и тщательно перемешаны, так что всякая карточка имеет одинаковые шансы быть вынутой. Мы можем произвести n извлечений карточек из нашей урны, либо возвращая в урну каждую вынутую карточку (при условии тщательного смешивания всех карточек перед каждым извлечением), либо не возвращая обратно вынутой карточки. Если pN $(p < 1)$ объектов обладают признаком A, который мы можем приписать и соответствующей карточке, то в обоих случаях вероятность появления карточки с признаком A равна p; при этом в первом случае мы имеем дело со схемой ряда n опытов типа Бернулли (повторение n независимых опытов, при которых наступление события A имеет неизменную вероятность p), между тем как второй случай приводит нас к схеме урны с невозвращаемыми шарами (глава II). Вынутые n карточек представляют собой выборку или частичную совокупность из данной (основной) совокупности.

Не останавливаясь на технике осуществления выборки [1]), которая должна быть произведена в условиях, объективно эквивалентных

---

[1]) Например, при рассмотрении только что указанных в конце предыдущей главы результатов подсчета букв А. А. Марковым, ни в коем случае нельзя смотреть на его группы из 100 последовательных букв, как на

вышеуказанным схемам, мы можем формулировать следующие математические положения, вытекающие из ранее доказанных теорем.

Пусть m есть число объектов, обладающих в произведенной выборке признаком A. Тогда, в первом случае (схема Бернулли или возвращаемых шаров), вероятность, что p удовлетворяет неравенству

$$t_0\sqrt{\frac{pq}{n}}<\frac{m}{n}-p<t_1\sqrt{\frac{pq}{n}}, \tag{194}$$

а во втором случае (схема невозвращаемых шаров), вероятность неравенства

$$t_0\sqrt{\frac{pq}{n}\left(1-\frac{n}{N}\right)}<\frac{m}{n}-p<t_1\sqrt{\frac{pq}{n}\left(1-\frac{n}{N}\right)}, \tag{195}$$

равна (с достаточной точностью для больших значений n)

$$\frac{1}{\sqrt{2\pi}}\int_{t_0}^{t_1}e^{-\frac{t^2}{2}}dt. \tag{196}$$

В действительности, вместо неравенств (194) и (195) пользуются, соответственно, следующими неравенствами:

$$t_0\sqrt{\frac{m(n-m)}{n^3}}<\frac{m}{n}-p<t_1\sqrt{\frac{m(n-m)}{n^3}} \tag{194 bis}$$

и

$$t_0\sqrt{\frac{m(n-m)}{n^3}\left(1-\frac{n}{N}\right)}<\frac{m}{n}-p<$$

$$<t_1\sqrt{\frac{m(n-m)}{n^3}\left(1-\frac{n}{N}\right)}, \tag{195 bis}$$

которые получаются из (194) и (195) путем замены в их левых и правых частях неизвестных величин p и q приближенными значениями $\frac{m}{n}$ и $\frac{n-m}{n}$.

выборки из совокупности 100 000 букв; необходимо было бы перемещать все 100 000 букв, — выборка, по определению, должна уничтожить все следы той реальной связи между объектами совокупности, которая обычно существует в действительности.

Полное обоснование этих формул требует непосредственного доказательства предельной теоремы, обратной теореме Лапласа, с помощью формулы Байесса [1]).

---

[1]) Читатель найдет это доказательство в курсе Маркова, делающего однако частное предположение, что a priori распределение вероятностей для p $(0 < p < 1)$ равномерно. В моем литографированном курсе (изд. Харьковского университета 1917 года) дано более общее доказательство, предполагающее лишь, что дифференциальная функция распределения вероятностей $f(p)$ a priori отлична от 0 для значений промежутка $\left(\frac{m}{n}+\varepsilon,\ \frac{m}{n}-\varepsilon\right)$ где ε — данная произвольно малая величина.

Не вдаваясь в детали вычислений, можем рассуждать следующим образом. На основании теоремы, обратной теореме Бернулли (глава IV, часть III) выбираем L достаточно большим, чтобы вероятность того, что p окажется вне промежутка $\frac{m}{n} \pm \frac{L}{n}$ была произвольно мала. В силу этого, полагая $|t_0| \leqslant L$ и $|t_1| \leqslant L$, заключаем, что вероятность того, что p удовлетворяет неравенству (194 bis) должна иметь тот же предел при неограниченном возрастании n, что и вероятность неравенства (194). Но, допуская, что априорное распределение вероятностей величины p на отрезке 01 выражается некоторой непрерывной функцией, мы можем считать это распределение равномерным внутри всякого промежутка, длина которого не более $\frac{2L}{\sqrt{n}}$, если n неограниченно возрастает. Поэтому, благодаря формуле Байесса (58 ter), вероятность гипотезы, что p находится в некотором промежутке $(p,\ p+\Delta p)$, после того как становится известным значение $\frac{m}{n}$, пропорциональна условной вероятности, получаемой этим значением $\frac{m}{n}$ при рассматриваемом значении p, т.-е.

$$\frac{e^{-\frac{(m-np)^2}{2npq}}}{\sqrt{2\pi npq}} = \frac{e^{-\frac{(m-np)^2}{2m(1-m/n)}}}{\sqrt{2\pi m(1-m/n)}} = \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}\,\Delta t,$$

(пренебрегая бесконечно малыми высших порядков).

Отсюда следует, что вероятность неравенства (134 bis), когда дано $\frac{m}{n}$, должна иметь пределом $\frac{\lambda}{\sqrt{2\pi}}\int_{t_0}^{t_1} e^{-\frac{t^2}{2}}\,dt$, где λ постоянный множитель, который должен быть равен 1, так как, ввиду сделанного выше замечания, рассматриваемая вероятность должна произвольно мало отличаться от единицы, если $t_0 = -L$, $t_1 = L$, когда L достаточно велико.

Заметим, что, если значения n не очень велики (примерно не более 200), то применение формулы Лапласа к неравенству (194 bis) приводит к дополнительной погрешности, которой нет в неравенстве (194). Полагая $\frac{m}{n} - p = y$ и умножая последнее неравенство на $\sqrt{\frac{m(n-m)}{n^2pq}}$, можем придать ему форму

$$t_0\sqrt{\frac{m(n-m)}{n^3}} < \frac{y}{\sqrt{1+\frac{(2m-n)ny}{m(n-m)}-\frac{n^2y^2}{m(n-m)}}} <$$

$$< t_1\sqrt{\frac{m(n-m)}{n^3}}, \qquad (194\ \text{ter})$$

так как

$$pq = \left(\frac{m}{n}-y\right)\left(\frac{n-m}{n}+y\right) = \frac{m(n-m)}{n^2} + \frac{2m-n}{n}y - y^2.$$

Пусть, например, $n = 100$, $m = 80$; тогда неравенство (194 bis) обращается в

$$\frac{t_0}{25} < y < \frac{t_1}{25},$$

в то время как неравенство (194 ter) имеет вид

$$\frac{t_0}{25} < \frac{y}{\sqrt{1+\frac{15}{4}y-\frac{25}{4}y^2}} < \frac{t_1}{25};$$

таким образом, например, вероятность $\Phi(2) = 0{,}4772$, вместо неравенства $0 < y < \frac{2}{25}$ (т.-е. $0{,}72 < p < 0{,}8$), будет точнее приписать неравенству

$$0 < \frac{y}{\sqrt{1+\frac{15}{4}y-\frac{25}{4}y^2}} < \frac{2}{25},$$

т.-е.

$$0 < y < \frac{0{,}3+\sqrt{4{,}25}}{26} \neq 0{,}091$$

(т.-е. $0{,}709 < p < 0{,}8$), и, с другой стороны, неравенству $0{,}8 < p < 0{,}868$, вместо $0{,}8 < p < 0{,}88$.

2. На основании (194 bis) и (195 bis), приравнивая неизвестную вероятность p полученному из выборки значению $p' = \frac{m}{n}$ и полагая $\sigma = \sqrt{\frac{m(n-m)}{n^3}} = \sqrt{\frac{p'q'}{n}}$ и $\sigma' = \sqrt{\frac{m(n-m)}{n^3}\left(1-\frac{n}{N}\right)} = \sqrt{\frac{p'q'}{n}\left(1-\frac{n}{N}\right)}$,

мы можем утверждать, что погрешность $p-p'$ подчиняется закону Гаусса с штандартом равным σ (при выборке первого рода) и с штандартом σ' (при выборке второго рода).

Наиболее обычна выборка второго рода (схема невозвращаемых шаров); при том же числе n выбранных объектов она дает немного бо́льшую точность, чем схема Бернулли, так как $\sigma'=\sigma\sqrt{1-\frac{n}{N}}$ т.-е. $\sigma'<\sigma$. Однако, в большинстве случаев N настолько велико, что полученное от этого преимущество невелико; например, если $\frac{n}{N}=0{,}1$, то $\sqrt{1-\frac{n}{N}}\neq 0{,}95$, и замена σ через $0{,}95\,\sigma$ значения не имеет; только тогда, когда процент выбранных объектов велик, если, например, $\frac{n}{N}=\frac{1}{2}$, то $\sigma'=\frac{\sigma}{\sqrt{2}}$ дает существенное уточнение. Вообще же говоря, точность определения p посредством выборки лишь в ничтожной степени зависит от отношения $\frac{n}{N}$ числа выбранных объектов к числу объектов совокупности, а главным образом характеризуется абсолютной величиной числа n, так как штандарт σ $\left(\text{при том же найденном значении } p'=\frac{m}{n}\right)$ обратно пропорционален $\sqrt{n}$. Например, если $N=1\,000\,000$, то выборка при $n=8\,000$ $\left(\frac{n}{N}=0{,}008\right)$ дает вдвое более точный результат, чем выборка $n=2\,000$ объектов из $N=20\,000$ $\left(\frac{n}{N}=0{,}1\right)$. [Понятно, впрочем, что штандарт погрешности в определении абсолютного числа pN объектов основной совокупности, обладающих признаком A, равен (в схеме невозвращаемых шаров) $N\sigma'=\sqrt{p'q'\frac{N}{n}(N-n)}$, и для его уменьшения необходимо было бы, чтобы $\frac{N}{n}$ было близко к 1.]

Как было указано раньше, благодаря тому, что $p - p'$ следует закону Гаусса, после определения $p' = \frac{m}{n}$ можно быть практически уверенным, что

$$p' - 4\sigma < p < p' + 4\sigma; \tag{197}$$

если хотим, например, определить p с точностью[1]) до 0,01, то достаточно, чтобы $4\sigma < 0{,}01$, т.-е. $\sqrt{\frac{p'q'}{n}} \leqslant \frac{1}{400}$, откуда $n \geqslant 160\,000\, p'q'$, и так как, во всяком случае, $p'(1 - p') \leqslant \frac{1}{4}$, то достаточно, чтобы $n \geqslant 40\,000$.

Таким образом частичная совокупность, состоящая из 40 000 объектов, обычно совершенно достаточно характеризует любую основную совокупность. Поэтому, за исключением случая, когда признак A редок (и ожидаемое значение p не превышает 0,01—0,02), статистик не должен стремиться еще увеличивать значение n, но взамен этого сконцентрировать свой технический аппарат на сравнительно небольшой группе обследуемых объектов, обратив все свое внимание на то, чтобы выборка соответствовала указанной выше теоретической схеме и выбранные случайно объекты, по возможности, без всякого исключения, подверглись точному обследованию. Если, вопреки всем усилиям, некоторое число α из входящих в частичную совокупность индивидов окажется неправильно обследованным, то погрешность значения p' увеличится на $\frac{\alpha}{n}$, и если бы $\frac{\alpha}{n}$ достигало 4—5%, то выборку пришлось бы признать неточной, как бы велико ни было значение n (а если $p' < 0{,}1$, то и 1% уклонившихся от обследования или плохо обследованных объектов достаточен, чтобы подорвать теоретическое значение выборки). Поэтому лучше подвергнуть выборочному обследованию $n = 20\,000$ объектов, приняв меры, гарантирующие не более

погрешность p равна $\frac{4\sigma}{p'} = 4\sqrt{\frac{q'}{p'n}}$; поэтому, если
ньше, чем $\frac{12}{\sqrt{n}}$.

1% сомнительных показаний, чем производить обследование 180 000 объектов в расчете на то, что будет не более 5% непроверенных показаний: если, например, $p' = \frac{1}{3}$, то в первом случае можно гарантировать, что $\left|p - \frac{1}{3}\right| < \frac{4}{300} + \frac{1}{100} \neq 0{,}023$, а во втором — только $\left|p - \frac{1}{3}\right| < \frac{4}{900} + \frac{1}{20} \neq 0{,}054$.

Возвращаясь к неравенству (197), заметим, что на практике можно без ущерба коэффициент 4 при σ заменить коэффициентом 3 (так как и тогда вероятность этого неравенства $2\Phi(3) = 0{,}9973$ очень близка к 1); кроме того, значение n может быть во много раз уменьшено, если допустимую погрешность в определении p несколько увеличить; так, например, если хотим, чтобы вероятность неравенства

$$|p - p'| < \frac{1}{20}$$

была равна $2\Phi(3)$, то достаточно взять $n = 900$, так как нужно, чтобы $3\sqrt{\frac{p'q'}{n}} \leqslant \frac{1}{20}$. Из свойств кривой Гаусса следует, кроме того, что при $n = 900$ вероятная погрешность $p - p'$ равна $\frac{2}{3}\sqrt{\frac{p'q'}{n}} < \frac{1}{90}$; таким образом приблизительно столь же часто будет оказываться, что $|p - p'| < \frac{1}{45}\sqrt{p'q'}$, как и противоположное неравенство.

Приведем следующий пример, заимствованный у Bowley[1]).

Из географического указателя, содержащего $N = 31\,210$ городов, была сделана выборка (по схеме урны с невозвращаемыми шарами), включающая $n = 500$ названий, при чем признаком $A_1, A_2, \ldots$ было, соответственно, нахождение города между 0° и 10° широты, 10° и 20° широты и т. д. Нижеследующая таблица дает результаты выборки и действительные значения p, соответствующие каждой широте, при чем для удобства письма даны значения $1\,000\,p$ и $1\,000\,p'$ вместо p и p'.

	Широта северная или южная								
	0° — 10°	10° — 20°	20° — 30°	30° — 40°	40° — 50°	50° — 60°	60° — 70°	70° — 80°	80° — 90°
Число городов в выборке при $n = 500$	22	56	104	103	93	112	9	1	0
$1\,000\,p'$	44	112	208	206	186	224	18	2	0
$1\,000\sqrt{\frac{p'q'}{n}}$	9	14	18	18	17	19	6	2	?
$1\,000\,p$ (Фактическое число городов на 1 000)	51	111	201	200	200	215	18	3,4	0,9

Принимая во внимание, что $\frac{n}{N} < \frac{1}{60}$, здесь не было надобности усложнять вычисление применением значения σ' вместо σ. Можно заметить, что разность $1\,000\,p - 1\,000\,p'$ только 2 раза превышает свое вероятное значение $\frac{2\,000\,\sigma}{3}$ (в первом и пятом столбцах).

3. Если для двух основных совокупностей s и s_1 путем выборки установлены соответственно приближенные значения p' и p_1' частот p и p_1, которыми обладает признак A в каждой из совокупностей s и s_1, то нетрудно дать приближенное значение разности $p - p_1$.

Принимая во внимание, что если из двух независимых выборок найдены величины p' и p_1', — p и p_1 подчиняются закону Гаусса с центрами p' и p_1' и со штандартами σ и σ_1, заключаем, что разность $p - p_1$ также подчиняется закону Гаусса с центром, равным $p' - p_1'$ и штандартом $\sqrt{\sigma^2 + \sigma_1^2}$.

Предположим, что посредством выборок, произведенных для совокупностей s и s_1 всех членов данного профессионального союза в 2 различных срока, установлены приближенные значения для процента женщин $100\,p' = 21{,}5$ в первый срок и $100\,p_1' = 20{,}07$ — во второй срок; в первом случае выборка содержала $n = 1\,000$, во втором — $n' = 1\,500$ индивидов. Полагая, что общее число членов союза более 100 000, можно пренебречь дробью $\frac{n}{N}$.

Таким образом в данном случае

$$100\,\sigma = \sqrt{\frac{21{,}5 \cdot 78{\cdot}5}{1000}} \neq 1{,}29,$$

$$100\,\sigma_1 = \sqrt{\frac{20{,}07 \cdot 79{,}93}{1500}} \neq 1{,}034;$$

поэтому штандарт разности $100\,p_1' - 100\,p'$ равен

$$\sqrt{1{,}29^2 + 1{,}034^2} \neq 1{,}31.$$

Следовательно, разность

$$100\,p' - 100\,p_1' = 1{,}43$$

едва превышает значение своего штандарта, и потому нет серьезных оснований считать, что процент женщин действительно уменьшился. Иначе обстояло бы дело, если бы выборки опирались на более многочисленные данные; те же значения p' и p_1' были бы показателями несомненного (практически) уменьшения числа женщин, если бы $n = n' = 10\,000$. Тогда $100\,\sigma \neq 0{,}41$, $100\,\sigma_1 = 0{,}40$, откуда $100\sqrt{\sigma^2 + \sigma_1^2} = 0{,}57$; поэтому $\frac{1{,}43}{0{,}57} \neq 2{,}5$, и вероятность, что столь значительное уклонение могло получиться случайно, равна только $1 - 2\Phi(2{,}5) \neq 0{,}012$.

4. Выборочный метод применяется также для определения среднего размера какой-нибудь величины (или экстенсивного признака), которая у различных индивидов данной основной совокупности имеет разные значения. Это может быть биологический признак, как, например, рост всех мужчин (не моложе 20 лет) данной страны, или экономический, как месячный заработок каждого члена данного профессионального союза, и т. п. Так же, как и в случае качественного или альтернативного признака A, который мы рассматривали до сих пор, необходимо точно ограничить ту совокупность индивидов, для которой определяется средняя величина признака, ни в какой мере не предполагая, однако, что исследуемая нами совокупность биологически или экономически однородна, и оставляя в стороне вопрос о методологических основаниях, которыми тот или иной исследователь может руководствоваться, объединяя всех индивидов данной совокупности. Для нас в настоящий момент важно лишь, чтобы

указанное объединение было технически возможно и чтобы в отношении рассматриваемого признака все объединенные в данную совокупность индивиды признавались нами равноценными, так что, независимо от ее практического значения, мы ставим себе только задачу, не производя индивидуального измерения всех без исключения рассматриваемых индивидов, найти приближенно величину

$$X = \frac{x_1 + x_2 + \ldots + x_N}{N},$$

где x_i $(i = 1, \ldots, N)$ есть фактическое количественное значение признака x каждого из N индивидов выделенной совокупности.

Для этой цели мы, подобно предыдущему, заменяем каждого индивида соответствующей карточкой и, осуществляя условия, при которых карточка каждого из N индивидов имеет одинаковые шансы быть вытянутой, извлекаем n карточек из всей совокупности, т.-е. делаем выборку n элементов. Имея в виду, что на практике $\frac{n}{N}$ обычно мало, можем для некоторого упрощения вычислений ограничиться сначала схемой Бернулли; в таком случае мы имеем n независимых величин $\xi_1, \xi_2, \ldots, \xi_n$, из которых каждая с вероятностью $\frac{1}{N}$ может получить любое из N значений $(x_1, x_2, \ldots, x_N)$. Поэтому

$$\text{М. О. } \xi_i = \frac{x_1 + x_2 + \ldots + x_N}{N} = X, \tag{198}$$

$$\sigma^2 = \text{М. О. } (\xi_i - X)^2 =$$

$$= \frac{(x_1 - X)^2 + (x_2 - X)^2 + \ldots + (x_N - X)^2}{N}. \tag{199}$$

Следовательно, применяя теорему Ляпунова, заключаем, что вероятность неравенства

$$t_0 \frac{\sigma}{\sqrt{n}} < \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n} - X < t_1 \frac{\sigma}{\sqrt{n}} \tag{200}$$

при неограниченном возрастании n имеет пределом

$$\frac{1}{\sqrt{2\pi}}\int_{t_0}^{t_1} e^{-\frac{t^2}{2}}\,dt. \tag{156}$$

Подобно тому как было замечено раньше, это утверждение остается в силе и тогда, когда мы считаем данным

$$X' = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n},$$

а неизвестной величиной является X.

Кроме того

$$\text{М. О. } (\xi_1 - X')^2 = \text{М. О. } \left[\frac{(n-1)\,\xi_1 - \xi_2 - \ldots - \xi_n}{n}\right]^2 =$$

$$= \frac{(n-1)^2 + (n-1)}{n^2}\,\sigma^2 = \frac{n-1}{n}\,\sigma^2;$$

поэтому, при возрастании n, величина

$$\frac{\sum\limits_1^n (\xi_i - X')^2}{(n-1)\,\sigma^2},$$

согласно закону больших чисел, безгранично приближается к 1, которая является ее математическим ожиданием. А потому, полагая

$$\sigma' = \sqrt{\frac{\sum (\xi_i - X')^2}{n-1}}, \tag{201}$$

можем утверждать, что и неравенство

$$t_0\frac{\sigma'}{\sqrt{n}} < X' - X < t_1\frac{\sigma'}{\sqrt{n}}, \tag{200 bis}$$

рассматриваемое вместо (200), имеет пределом

$$\frac{1}{\sqrt{2\pi}}\int_{t_0}^{t_1} e^{-\frac{t^2}{2}}\,dt,$$

когда n бесконечно возрастает, так как $\frac{\sigma'}{\sigma} \longrightarrow 1$.

Понятно, что, поскольку мы имеем дело с большим значением n, безразлично, берется ли знаменателем $n-1$ в выражении (201) или n; считая эту деталь практически несущественной, будем и в том и в другом случае называть σ'^2 **средним квадратичным уклонением величины** ξ_i **от их общей средней** X', замечая, что σ'^2 представляет собой достаточно точное для практических целей приближенное значение квадрата штандарта σ [т.-е. М. О. $(\xi_i - X)^2$]. Очевидно кроме того, что погрешность этого приближенного значения σ может быть определена аналогичным образом.

Действительно, согласно равенству (199), σ^2 представляет среднее квадратичное уклонение $\frac{\sum (x_i - X)^2}{N}$; поэтому та же выборка может служить для приближенного определения этой средней величины объектов всей совокупности, и мы имеем, по той же теореме Ляпунова, что вероятность неравенства

$$\frac{z_0 \sigma_2}{\sqrt{n}} < \frac{(\xi_1 - X)^2 + \ldots + (\xi_n - X)^2}{n} - \sigma^2 < z_1 \frac{\sigma_2}{\sqrt{n}} \tag{202}$$

имеет пределом (при возрастании n) $\frac{1}{\sqrt{2\pi}} \int_{z_0}^{z_1} e^{-\frac{t^2}{2}} dt$, где

$$\sigma_2^2 = \text{М. О. } [(\xi_i - X)^2 - \sigma^2]^2 = \text{М. О. } (\xi_i - X)^4 - 2\sigma^2 \text{ М. О. } (\xi_i - X)^2 + \sigma^4 =$$
$$= \text{М. О. } (\xi_i - X)^4 - \sigma^4. \tag{203}$$

Но, принимая во внимание, что

$$\sum_1^n (\xi_i - X)^2 = \sum (\xi_i - X')^2 + 2(X' - X) \sum_1^n (\xi_i - X') + n(X' - X)^2 =$$
$$= \sum (\xi_i - X')^2 + n(X' - X)^2$$

(так как $\sum_1^n \xi_i = nX'$, а потому $\sum_1^n (\xi_i - X') = 0$), видим, что

$$\frac{1}{n} \sum_1^n (\xi_i - X)^2 = \frac{1}{n} \sum_1^n (\xi_i - X')^2 + (X' - X)^2,$$

и, замечая в силу (200 bis), что, как бы мало ни было ε, вероятность неравенства

$$(X' - X)^2 > \frac{\varepsilon^2}{\sqrt{n}}$$

стремится к 0 с возрастанием [1]) n, заключаем, что вероятность неравенства

$$z_0 \frac{\sigma_2}{\sqrt{n}} < \frac{1}{n} \sum_1^n (\xi_i - X')^2 - \sigma^2 < z_1 \frac{\sigma_2}{\sqrt{n}} \qquad (202\ \text{bis})$$

имеет тот же предел $\frac{1}{\sqrt{2\pi}} \int_{z_0}^{z_1} e^{-\frac{t^2}{2}} dt$, что и вероятность неравенства (202).

Таким образом, после того как по выборке определены значения ξ_i и их средняя X', квадрат неизвестного штандарта (или дисперсия величин x_i) σ^2 подчиняется закону Гаусса с центром σ'^2 и штандартом σ_2. Поэтому, в частности, вероятная погрешность равенства

$$\sigma'^2 \neq \sigma^2$$

равна $\pm \frac{2}{3} \frac{\sigma_2}{\sqrt{n}}$ (т.-е. столь же вероятно, что $|\sigma'^2 - \sigma^2| < \frac{2}{3} \frac{\sigma_2}{\sqrt{n}}$, как и противоположное неравенство), и, имея в виду, что σ_2 есть независимая от n постоянная, заключаем, что погрешность указанного выше способа вычисления штандарта на основании выборки стремится к 0 с возрастанием n.

Замечая, кроме того, что неравенство (202 bis) равнозначно

$$\frac{z_0}{\sigma' + \sigma} \cdot \frac{\sigma^2}{\sqrt{n}} < \sigma' - \sigma < \frac{z_1}{\sigma' + \sigma} \cdot \frac{\sigma_2}{\sqrt{n}} \qquad (202\ \text{ter})$$

и для больших значений n пред. $\frac{\sigma'}{\sigma} = 1$, можем заменить последнее неравенство любым из неравенств

$$\frac{z_0 \sigma_2}{2\sigma \sqrt{n}} < \sigma' - \sigma < \frac{z_1 \sigma_2}{2\sigma \sqrt{n}} \quad \text{или} \quad \frac{z_0 \sigma_2}{2\sigma' \sqrt{n}} < \sigma' - \sigma < \frac{z_1 \sigma_2}{2\sigma' \sqrt{n}}, \qquad (204)$$

откуда видно, что $\sigma' - \sigma$ также подчиняется закону Гаусса со штандартом $\frac{\sigma_2}{2\sigma \sqrt{n}}$ $\left(\text{или } \frac{\sigma_2}{2\sigma' \sqrt{n}}\right)$. При этом, по тем же соображениям, и величина σ_2 может быть заменена весьма мало от нее отличающимся приближенным значением:

$$\sigma'_2 = \sqrt{\frac{\sum (\xi_i - X')^4}{n} - \sigma'^4}. \qquad (203\ \text{bis})$$

Прежде чем перейти к рассмотрению примеров, следует заметить, что если n не очень велико (не превышает нескольких сотен),

---

[1]) Так как это неравенство равнозначно $|X' - X| > \frac{\varepsilon}{\sqrt[4]{n}} = \frac{\varepsilon \sqrt[4]{n}}{\sqrt{n}}$.

то пользование вышеустановленными формулами дает лишь грубое приближение, и на результатах такой выборки нельзя строить серьезных выводов, а нужно видеть в них лишь некоторые качественные указания, требующие проверки на основании более обширных статистических материалов.

5. Предположим, что требуется определить средний урожай X ржи с десятины в данном году в некотором районе, охватывающем 1 000 000 подлежащих учету десятин. Для этого производится точное выборочное обследование [1]) (по схеме урны с невозвращаемыми шарами) 5 000 десятин. Найдено, что $X' = 78$ пуд.,

$$\sigma'^2 = \frac{\sum_1^{5\,000} (\xi_i - 78)^2}{5\,000} = 312{,}5.$$

В таком случае

$$\frac{\sigma'}{\sqrt{n}} = \sqrt{\frac{312{,}5}{5\,000}} = \frac{1}{4}.$$

Поэтому вероятность, что средний урожай с десятины X по всему району удовлетворяет неравенству

$$\frac{t_0}{4} < 78 - X < \frac{t_1}{4}$$

равна $\Phi(t_1) - \Phi(t_0)$. В частности, вероятность, что

$$77 < X < 79 \tag{205}$$

равна $2\Phi(4) = 0{,}999\,936$, т.-е. можно быть практически уверенным, что найденный по выборке результат дает действительный средний урожай всего района с точностью до 1 пуда.

Если бы, например, было найдено, что $\sum \frac{(\xi_i - 78)^2}{5\,000} < 2\,000\,000$, то $\sigma^2 < 1\,400$; поэтому штандарт погрешности при замене σ^2 через 312,5 менее, чем $\frac{1\,400}{\sqrt{5\,000}} < 20$. Следовательно, можно быть уверенным, что $\sigma^2 < 312{,}5 + 4 \cdot 20 = 392{,}5$, а потому $\frac{\sigma}{\sqrt{n}} < \sqrt{\frac{392{,}5}{5\,000}} \neq 0{,}28$, что не изменит по существу нашего вывода, так как и в худшем случае вероятность неравенства (205) равна $2\Phi(3{,}6) = 0{,}999\,682$.

---

[1]) Каждой из $N = 1\,000\,000$ десятин должна соответствовать отдельная карточка.

Основное требование выборочного метода заключается в том, чтобы всякий из индивидов основной совокупности имел одинаковые шансы быть отобранным. Ясно, что если мы рассматриваем две выборки из одной и той же совокупности, то соответствующие им средние X' и X'' не будут вполне тождественны; вообще, каковы бы ни были обе выборки (лишь бы они были независимы), мы знаем на основании предыдущего, что вероятность, равную $\frac{1}{\sqrt{2\pi}}\int_{t_0}^{t_1} e^{-\frac{t^2}{2}}\,dt$, будет иметь неравенство

$$t_0\sqrt{\frac{\sigma'^2}{n}+\frac{\sigma''^2}{m}}<X'-X''<t_1\sqrt{\frac{\sigma'^2}{n}+\frac{\sigma''^2}{m}}, \qquad (206)$$

где σ'^2 и σ''^2 — соответственно, средние квадратичные уклонения каждой выборки, а n и m — числа объектов, вошедших в них, при чем безразлично, произведены ли выборки из той же самой или различных основных совокупностей [1]). Однако результат будет иной, если выборки не независимы между собой. Допустим, что $n<m$ и первая выборка является частью второй; пусть $m-n=n_1$. Тогда мы можем рассматривать, как независимые, выборки с n элементами и с остальными $n_1=m-n$ элементами, предполагая, что число N объектов основной совокупности велико по сравнению с m. Поэтому, если X_1' есть средняя этой дополнительной совокупности (с n_1 элементами), то

$$nX'+n_1X_1'=mX''$$

(где X'' есть средняя суммарной выборки с $m=n+n_1$ элементами) и разность

$$mX''-mX'=(n-m)X'+n_1X_1'=n_1X_1'-n_1X',$$

являясь разностью между независимыми величинами, подчиняется закону Гаусса со штандартом

$$n_1\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n}}=\sigma\sqrt{\frac{n_1m}{n}}.$$

---

[1]) Таким образом, если найдено, что $|X'-X''|>4\sqrt{\frac{\sigma'^2}{n}+\frac{\sigma''^2}{m}}$, то можно быть уверенным, что выборки соответствуют различным совокупностям или, что техника выполнения по крайней мере одной из них была неудовлетворительна.

Следовательно, $X''-X'$ также подчиняется закону Гаусса со штандартом[1])

$$\sigma\sqrt{\frac{n_1}{nm}}=\sigma\sqrt{\frac{1}{n}-\frac{1}{m}}. \tag{207}$$

Полученной формулой можно, очевидно, пользоваться и в том случае, когда выбранная совокупность с m элементами рассматривается как основная и отношение $\frac{n}{m}$ не настолько мало, чтобы им можно было пренебречь.

Предположим, например, что хотят определить средний головной индекс (отношение ширины черепа к длине) X у 20 000 новобранцев, призываемых в данном районе. Чтобы не подвергать всех измерению, измеряют индекс только у каждого 5-го по порядку индивида, так что $m=20\,000$ и $n=4\,000$; пусть средний индекс у 4 000 новобранцев $X'=76{,}5$, а $\sigma'=2{,}8$; тогда, замечая, что

$$\sigma'\sqrt{\frac{1}{n}-\frac{1}{m}}=2{,}8\cdot\sqrt{\frac{1}{5\,000}}\neq 2{,}8\cdot 0{,}0141\neq 0{,}0395,$$

заключаем, что вероятность неравенства $\varepsilon_0<X-76{,}5<\varepsilon_1$ равна

$$\frac{1}{0{,}0395\sqrt{2\pi}}\int_{\varepsilon_0}^{\varepsilon_1} e^{-\frac{z^2}{2\cdot(0{,}0395)^2}}dz=\Phi\left(\frac{\varepsilon_1}{0{,}0395}\right)-\Phi\left(\frac{\varepsilon_0}{0{,}0395}\right);$$

таким образом среднее значение 76,5 индекса всех 20 000 новобранцев можно считать точным до 0,1, так как вероятность, что $|X-76{,}5|<0{,}1$ больше, чем $2\Phi(2{,}5)\neq 0{,}9876$.

Часто вместо выборки, основным условием которой является соблюдение равновозможности обследования любого индивида основной совокупности, отбор частичной совокупности произво-

---

[1]) Значение σ на основании суммарной выборки с m элементами следует заменить приближенной величиной $\sigma''=\sqrt{\sum_1^m\frac{(\xi_i-X'')^2}{m}}$; если же σ'' неизвестно, но определено только $\sigma'=\sqrt{\sum_1^n\frac{(\xi_i-X')^2}{n}}$, то можно пользоваться последним.

дится по какому-нибудь определенному, например территориальному, признаку. Ясно, что тогда нет никаких оснований a priori ожидать, что обследование данной частной совокупности дает соответствующую характеристику всей основной совокупности. Существенным вопросом является тогда, эквивалентен ли произведенный отбор случайному или нет. На этот вопрос нетрудно ответить, если для основной совокупности известно среднее значение X и штандарт σ. В таком случае, если n невелико по отношению к N, то, как мы видели, штандарт разности $X' - X$ должен быть равен $\frac{\sigma}{\sqrt{n}}$, и, принимая во внимание, что эта разность подчиняется закону Гаусса, можно считать результат отбора эквивалентным выборке, если $|X' - X| < \frac{\sigma}{\sqrt{n}}$; при $\frac{\sigma}{\sqrt{n}} < |X' - X| < \frac{2\sigma}{\sqrt{n}}$ нет еще оснований считать отбор симптоматическим, а не случайным; если $\frac{2\sigma}{\sqrt{n}} < |X' - X| < \frac{3\sigma}{\sqrt{n}}$, то симптоматичность (т.-е. зависимость средней от способа отбора) приобретает значительную вероятность, предполагая, что нет априорных данных для противоположного суждения; наконец, если $|X' - X| > \frac{3\sigma}{\sqrt{n}}$, то симптоматичность отклонения можно в большинстве случаев считать практически доказанной, имея, конечно, в виду, что если вывод должен иметь характер общего суждения, распространяемого на другие аналогичные совокупности, то он нуждается в дальнейшей проверке. Если число $N = m$ не очень велико по сравнению с n, то вместо $\frac{\sigma}{\sqrt{n}}$ нужно брать $\sigma\sqrt{\frac{1}{n} - \frac{1}{m}}$.

Упражнения. 1) При выборочном [1]) 10-процентном обследовании деревенских хозяйств Украины весной 1924 г. из 300 000 хозяйств Волынской губ. и из 600 000 хозяйств Полтавской губ. было обследовано, соответственно, 30 000 хозяйств в первой и 60 000 во второй; при этом хозяйств без рабочего скота в первой оказалось 11 250, а во второй 23 400. Требуется опре-

[1]) Все данные нами закруглены для удобства вычислений; поэтому получаемые здесь численные результаты для непосредственного применения к указанным конкретным материалам должны быть соответствующим образом уточнены.

делить приближенно процент хозяйств, лишенных скота в указанных губерниях, погрешность этого приближения, а также установить, следует ли считать, что процент хозяйств, необеспеченных скотом, действительно больше в Полтавской губ., чем в Волынской.

Ответ. $p'=37{,}5\%$; $p''=39\%$; штандарт первой $\sigma'=\sqrt{\frac{9p'(100-p')}{300\,000}}\neq 0{,}27$ и $\sigma''=\sqrt{\frac{9p''(100-p'')}{600\,000}}\neq 0{,}19$. Разность $p''-p'$ имеет штандартом $\sigma=\sqrt{\sigma'^2+\sigma''^2}\neq 0{,}32$. Допущение, что при этом штандарте разность $p''-p'$ случайно достигла значения 1,5, неправдоподобно, так как $\frac{1{,}5}{0{,}32}\neq 4{,}7$; поэтому практически несомненно, что процент необеспеченных скотом хозяйств в Полтавской губ. был действительно больше, чем в Волынской.

2) Для определения среднего урожая с одной десятины определенного района в 1 000 000 десятин, которые обрабатываются 500 000 хозяйствами различных размеров, предлагается произвести выборочное обследование 5 000 хозяйств. Какова должна быть теоретическая схема этой выборки и какова ее погрешность?

Ответ. Полагая, что $f_1, f_2, \ldots, f_{500\,000}$ представляют соответственно число десятин каждого из хозяйств, так что $f_1+f_2+\ldots+f_{500\,000}=N=1\,000\,000$, находим точное значение среднего урожая с десятины $a=\frac{\sum f_i x_i}{\sum f_i}$, где $f_i x_i$ — полный урожай i-го хозяйства, а средний урожай с десятины для i-го хозяйства равен x_i. В основу выборки можно положить одну из двух следующих схем. Первая схема: производим выборку $n=5\,000$ хозяйств из всех $M=500\,000$ хозяйств так, чтобы вероятность каждому хозяйству быть обследованным была пропорциональна его размерам (в десятинах). В таком случае, обозначая через ξ_k средний урожай с десятины обследованного k-го хозяйства, имеем М. О. $\xi_k=a$; поэтому $X'=\frac{\xi_1+\ldots+\xi_n}{n}$ будет приближенным значением a, подчиняющимся закону Гаусса со штандартом $\frac{\sigma}{\sqrt{n}}$, где

$$\sigma^2=\frac{\sum_1^M f_i(x_i-a)}{N}=\frac{1}{N}\sum f_i x_i^2-a^2.$$

Второй прием, практически более простой, но менее точный, состоит в том, что производится обычным способом выборка, в которой каждое хозяйство имеет равные шансы быть обследованным, и в качестве неизвестной величины принимаем $A=\frac{\sum f_i x_i}{M}=a\frac{N}{M}$, т.-е. средний уро-

жай хозяйства всей совокупности. В таком случае получаем приближенное значение

$$\frac{f_1\xi_1+\ldots+f_n\xi_n}{n}$$

для A, где $f_k\xi_k$ — весь урожай k-го хозяйства, откуда находим приближенное значение для a

$$X_1=\frac{M}{N}\cdot\frac{\sum\limits_1^n f_k\xi_k}{n},$$

которого штандарт $\frac{\sigma_1}{\sqrt{n}}$ определяется из равенства

$$\sigma_1^2=\frac{M^2}{N^2}\sum_1^M\frac{(f_ix_i-A)^2}{M}=\frac{M}{N^2}\sum_1^M f_i^2x_i^2-a^2.$$

Нужно заметить, однако, что штандарт σ_1 обыкновенно значительно больше σ (когда размеры хозяйств могут быть весьма различны); поэтому, если значения f_i были известны до выборки, первый прием представляет то преимущество, что при выборке с меньшим числом n хозяйств он дает не меньшую точность, чем второй прием при более обширной выборке.

Предположим, например, что имеются хозяйства в 1, 2, 3, 4, 5 десятин, числа которых равны соответственно: 250 000, 100 000, 75 000, 50 000, 25 000 (так что число $M=500\,000$, $N=1\,000\,000$). Пусть средние урожаи с десятины для каждого из этих 5 классов хозяйств равны соответственно: $a_1=72$, $a_2=78$, $a_3=80$, $a_4=75$, $a_5=82$ при штандартах $s_1=17$, $s_2=s_3=s_4=16$, $s_5=18$. Чтобы произвести выборку по первой схеме, нужно взять $N=1\,000\,000$ карточек так, чтобы карточек, соответствующих каждому хозяйству, было столько, сколько в нем десятин (понятно, техника случайного отбора при данном приеме усложняется, если число различных классов увеличить; вообще, для того, чтобы учесть возможность непрерывного изменения размеров, целесообразно было бы пользоваться геометрическими моделями, отожествляя, например, каждое хозяйство с отрезком ленты, пропорциональным величине хозяйства). В данном случае

$$a=\frac{\sum f_ix_i}{\sum f_i}=\frac{72\cdot250+78\cdot200+80\cdot225+75\cdot200+82\cdot125}{1000}=76{,}85,$$

$$\sigma^2=\frac{1}{N}\sum_1^M f_i(x_i-a)^2=\frac{(17^2+4{,}85^2)\cdot250}{1000}+\frac{(16^2+1{,}15^2)\cdot200}{1000}+$$

$$+\frac{(16^2+3{,}15^2)\cdot225}{1000}+\frac{(16^2+1{,}85^2)\cdot200}{1000}+\frac{(18^2+5{,}15^2)\cdot125}{1000}=$$

$$=78{,}130\,625+51{,}4645+59{,}832\,562+51{,}8845+43{,}815\,313=285{,}1275$$

поэтому, при $n=5\,000$, $\frac{\sigma}{\sqrt{n}}=\sqrt{\frac{285{,}1275}{5000}}\neq0{,}239$, и, следовательно,

можно быть уверенным, что выборка определит a с точностью до одного пуда $\left(\frac{4\sigma}{\sqrt{n}}\right)$.

Применяя второй прием, мы находим

$$\sigma_1^2 = \frac{1}{M}\sum_1^M \left(f_i \frac{M}{N} x_i - a\right)^2 = \left[\frac{1}{4}\cdot 17^2 + (36 - 76{,}85)^2\right]\cdot 0{,}5 +$$

$$+ [16^2 + 1{,}15^2]\cdot 0{,}2 + \left[\frac{9}{4}\cdot 16^2 + 43{,}15^2\right]\cdot 0{,}15 + [4\cdot 16^2 + 73{,}15^2]\cdot 0{,}1 +$$

$$+ [2025 + 128{,}15^2]\cdot 0{,}05 = 2847{,}5025,$$

откуда (при $n = 5\,000$)

$$\frac{\sigma_1}{\sqrt{n}} = \sqrt{\frac{2\,847{,}5025}{5\,000}} \neq 0{,}754$$

более чем втрое превышает погрешность первого приема.

Ввиду большей простоты второго приема и невозможности применения первого, если заранее неизвестны все f_i, т.-е. размеры всех хозяйств, можно, сохраняя схему равновозможности всех хозяйств, брать также в качестве приближенного значения для a величину

$$X_2 = \frac{\sum_1^n f_k \xi_k}{\sum_1^n f_k},$$

штандарт которой равен $\frac{\sigma_2}{\sqrt{n}}$ (для больших значений n), где σ_2 определяется формулой

$$\sigma_2^2 = \frac{M}{N^2}\sum_1^M f_i^2 (x_i - a)^2,$$

получаемой из выражения для $\sigma\left(\frac{x}{y}\right)$, данного в 3-м упражнении предыдущей главы. (Можно доказать, что, для очень больших значений n, X_2 также подчиняется закону Гаусса.) Обычно σ_2 меньше σ_1 и того же порядка, приблизительно, что и σ; поэтому, при применении на практике второй схемы, следует пользоваться значением X_2 предпочтительно перед значением X_1. В нашем примере

$$\sigma_2^2 = 0{,}5\cdot(78{,}130\,625) + (51{,}4645) + 1{,}5\cdot(59{,}832\,562) + 2\cdot(51{,}8845) +$$
$$+ 2{,}5\cdot(43{,}815\,313) \neq 392{,}636;$$

поэтому

$$\frac{\sigma_2}{\sqrt{n}} = \sqrt{\frac{392{,}636}{5\,000}} \neq 0{,}28;$$

таким образом, вторая схема выборки при применении последней формулы (X_2) дает почти ту же точность, что и первая схема (X).

3) В порт A в течение года прибыло 2 500 судов с товарами. Выборочным путем была обследована стоимость товаров на 500 судах; при этом оказалось 400 судов по 10 000 руб., 80 судов по 50 000 руб., 15 судов по 100 000 руб., 4 — по 250 000 руб. и 1 — на 500 000 руб. Требуется определить приближенно ценность годового импорта в порт A.

Ответ. 55 миллионов с вероятной ошибкой $\frac{2}{3}S = 2{,}4$ миллионов.

Средняя стоимость товара на 1 судне

$$a = \frac{10\,000 \cdot 400 + 50\,000 \cdot 80 + 100\,000 \cdot 15 + 250\,000 \cdot 4 + 500\,000}{500} = 22\,000;$$

среднее квадратичное уклонение

$$\sigma^2 = \frac{10\,000^2 \cdot 400 + 50\,000^2 \cdot 80 + 100\,000^2 \cdot 15 + 250\,000^2 \cdot 4 + 500\,000^2}{500} -$$

$$- 22\,000^2 = 1\,296 \cdot 10^6.$$

Поэтому по формуле (207) штандарт ошибки величины a равен $\sigma\sqrt{\frac{1}{500} - \frac{1}{2500}} = \frac{36\,000}{25} = 1\,440$, а штандарт S ошибки приближенного значения всего импорта 2 500 a равен $S = 2\,500 \cdot 1\,440 = 3{,}6$ миллионов.

В данном примере выборка дает не особенно точный результат, благодаря очень большому значению σ, соответствующему чрезвычайному разнообразию обследуемых элементов (от 10 000 до 500 000); в подобных случаях более целесообразно произвести группировку элементов совокупности на несколько классов и проводить выборку по классам.

ЧАСТЬ ПЯТАЯ

ОСНОВЫ ТЕОРИИ КРИВЫХ И ПОВЕРХНОСТЕЙ РАСПРЕДЕЛЕНИЯ

ГЛАВА ПЕРВАЯ

ПОСТРОЕНИЕ НОРМАЛЬНОЙ КРИВОЙ НА ОСНОВАНИИ СТАТИСТИЧЕСКИХ ДАННЫХ

1. В предыдущей части мы установили, что закон нормального распределения вероятностей (Гаусса) играет существенную роль в теории вероятностей и в ее приложениях, поскольку в действительности более или менее точно соблюдаются те весьма общие условия, которые нужны для приложимости предельных теорем Лапласа, Ляпунова и Маркова.

Если рассматривается весьма большое число n опытов (или объектов), в которых исследуемая величина x может получать различные значения с плотностью распределения вероятностей $f(x)$, то, на основании закона больших чисел, интеграл $\int_{\alpha}^{\beta} f(x)\,dx$, представляющий вероятность, что x находится в промежутке (α, β), будет очень мало отличаться от отношения $\frac{m(\alpha, \beta)}{n}$ числа $m(\alpha, \beta)$ объектов, для которых значение x заключено между α и β, к числу n всех объектов. Вполне точного совпадения обеих величин мы, однако, не в праве ожидать, особенно если интервал (α, β) очень мал и соответствующая ему вероятность ничтожна. Благодаря этому задача нахождения математической функции $f(x)$, которая бы соответствовала данному статистическому материалу, является в значительной мере неопределенной и тем в большей степени, чем меньше n. К указанной общей задаче мы вернемся в дальнейшем, выбирая по тем или иным теоретическим или техническим соображениям определенного вида функцию $f(x)$; но сначала мы возьмем в качестве теорети-

ческой функции распределения функцию Гаусса $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$ и, вычислив ее параметры a и σ на основании статистических данных, посмотрим, как выяснить, вмещаются ли уклонения теоретических чисел от наблюденных в рамки случайных уклонений, или же, напротив, несовпадение обоих рядов чисел настолько значительно, что гипотеза нормального распределения вероятностей должна быть отброшена.

Строго говоря, закон нормального распределения вероятностей не может соблюдаться, если величина x изменяется не непрерывно; тем не менее теорема Лапласа дает классический пример величины $\left(\frac{m-np}{\sqrt{npq}}\right)$, изменяющейся прерывно, к которой указанный закон применим с весьма большой точностью, лишь бы интервал (α, β) не был слишком мал. Но и тогда, когда x изменяется непрерывно, измерение ее при помощи данных приборов приводит к значениям, содержащим лишь некоторое определенное число точных десятичных знаков, а потому и в данном случае нет смысла рассматривать очень малые промежутки: если, например, величина x измеряется с точностью до 1 *мм*, то целые числа миллиметров играют фактически роль прерывных значений, которые она может получать. Поэтому мы не будем делать принципиального различия между случаями непрерывного и прерывного распределения вероятностей (см. главу V, часть II).

2. Вычисление центра a и штандарта σ вытекает из предшествующих теоретических выводов; приближенным значением a будет средняя арифметическая a' из наблюденных значений x_i, а приближенным значением дисперсии $\beta=\sigma^2$ является среднее квадратичное уклонение

$$\sigma'^2=\frac{\sum_1^n (x_i-a')^2}{n}.$$

Мы знаем, кроме того, что (при n достаточно большом) $\frac{\sigma'}{\sqrt{n}}$ будет приближенным штандартом погрешности $a'-a$, которая во всяком случае подчиняется закону Гаусса,

также как и разность $\sigma'^2 - \sigma^2$, имеющая штандартом $\frac{\sigma_2}{\sqrt{n'}}$, где

$$\sigma_2^2 = \text{M. O. } [(x_i - a)^2 - \sigma^2]^2 = \text{M. O. } (x_i - a)^4 - \sigma^4. \qquad (203)$$

В общем случае, как мы видели, σ_2^2 может быть, на основании статистического же материала, заменено приближенным значением

$$\sigma'^2_2 = \frac{\sum_1^n (x_i - a')^4}{n} - \sigma'^4. \qquad (203\text{ bis})$$

Однако, если мы предполагаем, что величина x подчиняется закону Гаусса, то

$$\beta_2 = \text{M. O. } (x_i - a)^4 = \text{M. O. } (x - a)^4 = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - a)^4 e^{-\frac{(x-a)^2}{2\sigma^2}} dx =$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^4 e^{-\frac{x^2}{2\sigma^2}} dx = \frac{\sigma x^3 e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi}} \Bigg|_{\infty}^{-\infty} +$$

$$+ \frac{3\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2\sigma^2}} dx = 3\sigma^4; \qquad (208)$$

поэтому

$$\sigma_2^2 = \beta_2 - \sigma^4 = 2\sigma^4 = 2\beta^2. \qquad (209)$$

Следовательно, в данном случае нет надобности вычислять (203 bis) и можно в качестве приближенного значения $\sigma_2 = \sigma^2\sqrt{2}$ взять $\sigma'^2\sqrt{2}$, а потому, согласно (204), приближенным штандартом погрешности $\sigma' - \sigma$ будет $\frac{\sigma'}{\sqrt{2n}}$.

Обычно, когда хотят выразить, что погрешность $A - A'$ равенства $A \neq A'$ подчиняется закону Гаусса с штандартом σ, пишут [1]

$$A = A' \pm \sigma.$$

[1]) Это значит, в частности, что вероятное значение разности $A - A'$ равно $\frac{2}{3}\sigma$, а значение большее чем 3σ практически невероятно (его вероятность менее чем 0,003).

Таким образом для определения центра a и штандарта σ кривой Гаусса имеем равенства

$$a = a' \pm \frac{\sigma}{\sqrt{n}}, \qquad \sigma = \sigma' \pm \frac{\sigma'}{\sqrt{2n}} \tag{210}$$

где

$$a' = \frac{\sum_1^n x_i}{n}, \qquad \sigma' = \sqrt{\frac{1}{n}\sum_1^n (x_i - a')^2}.$$

3. Рассмотрим следующий пример, заимствуя у Charlier численные данные, которые мы уже рассматривали на странице 210: m — число карт черной масти при 10 извлечениях из полной колоды карт (с возвращением вынутой карты, так что $p = \frac{1}{2}$ и схематические условия опыта тождественны 10-кратному бросанию монеты, где отмечают число m появлений „орла"). Произведено 1 000 таких опытов (т.-е. всего 10 000 извлечений). Найдено:

ТАБЛИЦА I.

$m =$	0	1	2	3	4	5	6	7	8	9	10	Сумма
Число опытов $Y' =$	3	10	43	116	221	247	202	115	34	9	0	1 000
$m - 5 = x =$	−5	−4	−3	−2	−1	0	1	2	3	4	5	0
$Y'x =$	−15	−40	−129	−232	−221	0	202	230	102	36	0	−67
$Y'x^2 =$	75	160	387	464	221	0	202	460	306	144	0	2 419

Для технического упрощения вычисления средней a' выбираем начало отсчета (провизорную среднюю) a_0 так, чтобы a_0 было круглым (в данном случае целым) числом, которое при беглом обозрении данных близко к искомой средней, и полагаем $x = m - a_0$; здесь естественно взять $a_0 = 5$.

В таком случае $a' = \frac{\sum Y'x}{n} = \frac{-67}{1\,000} = -0{,}067$, где числитель — 67 получился от сложения чисел 4-й строки, которые являются произведениями соответствующих чисел 2-й и 3-й строк таблицы.

С другой стороны,

$$n\sigma'^2 = \sum (x_i - a')^2 = \sum x_i^2 - 2\sum x_i a' + na'^2 = \sum x_i^2 - na'^2,$$

так как $\sum x_i = \sum Y'x = na'$; поэтому

$$\sigma'^2 = \frac{1}{n}\sum x_i^2 - a'^2 = \frac{1}{n}\sum Y'x^2 - a'^2 = 2{,}419 - 0{,}004\,489 =$$

$= 2{,}414\,511$, где $\sum Y'x^2 = 2\,419$ получена сложением чисел 5-й строки, которые представляют собой произведения соответствующих чисел 3-й и 4-й строк.

Таким образом

$$a = -0{,}067 \pm \frac{\sigma'}{\sqrt{1\,000}} = -0{,}067 \pm 0{,}0491,$$

$$\sigma' = 1{,}554 \pm 0{,}034\,75.$$

Полученные результаты достаточно хорошо согласуются с теоретической схемой опыта, где $a = 0$ и $\sigma = \sqrt{\frac{1}{4}\cdot 10} \neq 1{,}58$ (погрешность 0,067 центра лишь незначительно превышает штандарт 0,0491, а погрешность σ значительно меньше своего штандарта). Однако из этого еще не следует, что распределение вероятностей величины x действительно нормально, тем более, что и в теоретической схеме число извлечений 10 еще не так велико, чтобы можно было утверждать, что $x = m - 5$ подчиняется закону Гаусса.

На этом же примере мы разъясним, каким образом сделать проверку, допустима ли (в пределах данного опыта) гипотеза о нормальном распределении вероятностей. Если бы у нас не было теоретических оснований, чтобы приписать a и σ определенные значения, то мы построили бы кривую Гаусса, соответствующую только что вычисленным значениям a' и σ', но для лучшего выяснения различных допущений и погрешностей, которые при этом делаются, мы сравним сначала наше статистическое распределение с нормальной кривой

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}},$$

соответствующей теоретическим значениям $a=0$ и $\sigma=\sqrt{\frac{10}{4}}=1{,}58$, имея в виду, что рассматриваемые нами промежутки значений x не должны быть меньше 1.

Таким образом теоретическое число или, точнее, математическое ожидание числа опытов, в которых x заключено между α и β, равно

$$I_{\alpha}^{\beta}=\frac{1\,000}{\sigma\sqrt{2\pi}}\int_{\alpha}^{\beta}e^{-\frac{x^2}{2\sigma^2}}dx=\frac{1\,000}{\sqrt{2\pi}}\int_{\frac{\alpha}{\sigma}}^{\frac{\beta}{\sigma}}e^{-\frac{t^2}{2}}dt=$$

$$=1\,000\left[\Phi\left(\frac{\beta}{\sigma}\right)-\Phi\left(\frac{\alpha}{\sigma}\right)\right]$$

(так как общее число опытов $n=1\,000$), при чем мы ограничиваемся лишь значениями α и β вида $h+\frac{1}{2}$, где h — целое число, сравнивая теоретическое число $I_{h-\frac{1}{2}}^{h+\frac{1}{2}}=Y_h$ с наблюденным числом опытов Y'_h, в которых $x=h$.

Поэтому, пользуясь таблицей функции Лапласа $\Phi(t)$ и придавая t последовательные значения $\frac{-5{,}5}{\sigma}, \frac{-4{,}5}{\sigma}, \ldots, \frac{-0{,}5}{\sigma}, \frac{0{,}5}{\sigma}, \ldots \frac{5{,}5}{\sigma}$, где $\frac{1}{\sigma}=0{,}64$, составляем таблицу II.

2-я и 3-я строки служат для вычисления теоретических значений Y_h 4-й строки; 5-я строка содержит уклонения эмпирических значений Y'_h предшествующей таблицы от соответствующих теоретических значений Y_h. Попутно в 6-й строке даны теоретические числа, вытекающие из формулы бинома Ньютона $1\,000\left(\frac{1}{2}+\frac{1}{2}\right)^{10}$, которые совершенно точно соответствуют схеме произведенных опытов, при которой распределение вероятностей величины x прерывно; 7-я строка содержит уклонения эмпирических чисел Y'_h от последних теоретических чисел; беглого сравнения этой строки с 5-й

ТАБЛИЦА II.

$h =$		-5	-4	-3	-2	-1	0	1	2	3	4	5
$t =$	$-3{,}52$	$-2{,}88$	$-2{,}24$	$-1{,}60$	$-0{,}96$	$-0{,}32$	0,32	0,96	1,60	2,24	2,88	3,52
$1000\,\Phi(t) =$	$-499{,}8$	-498	$-487{,}5$	$-445{,}2$	$-331{,}5$	$-125{,}5$	125,5	331,5	445,2	487,5	498	499,8
$Y_h =$		1,8	9,5	42,3	113,7	206	251	206	113,7	42,3	9,5	1,8
$Y_h - Y'_h =$		1,2	$-0{,}5$	$-0{,}7$	$-2{,}3$	-15	4	4	$-1{,}3$	8,3	0,5	1,8
$\frac{1000}{1024}\,C_{10}^{h+5} =$		1	9,8	43,9	117,2	205,1	246,1	205,1	117,2	43,9	9,8	1
$\frac{1000}{1024}\,C_{10}^{h+5} - Y'_h =$		-2	$-0{,}2$	0,9	1,2	$-15{,}9$	$-0{,}9$	3,1	2,2	9,9	0,8	1
$\beta(h) =$		1,8	9,5	40,5	100,9	164	188	164	100,9	40,5	9,5	1,8
$\frac{(Y_h - Y'_h)^2}{\beta(h)} =$		0,8	0	0	0,1	1,4	0,1	0,1	0	1,7	0,1	1,8

достаточно, чтобы видеть, что приближение, достигаемое с помощью членов бинома, не лучше, чем то [1]), которое дает кривая Гаусса.

В 8-й строке помещена теоретическая (для кривой Гаусса) дисперсия

$$\beta(h) = \text{М. О.}(Y_h - Y'_h)^2 = 1\,000\frac{Y_h}{1\,000}\left(1 - \frac{Y_h}{1\,000}\right) = Y_h\left(1 - \frac{Y_h}{1\,000}\right);$$

последняя строка содержит отношение квадрата уклонения $(Y_h - Y'_h)$ к его математическому ожиданию; поэтому, применяя метод дисперсии, мы можем считать совпадение теоретических чисел с эмпирическими удовлетворительными, если средняя арифметическая D из величин $\frac{(Y_h - Y'_h)^2}{\beta(h)}$ не превышает значительно единицу [2]), например, не достигает 2. В данном примере совпадение очень хорошее, так как

$$D = \frac{6{,}1}{11} = 0{,}56.$$

Производя аналогичное вычисление для кривой Гаусса с эмпирическим центром $a' = -0{,}067$ и штандартом $\sigma' = 1{,}554$, мы получим теоретические значения

$$Y_h = \frac{1\,000}{\sigma'\sqrt{2\pi}}\int_{h-\frac{1}{2}}^{h+\frac{1}{2}} e^{-\frac{(x-a')^2}{2\sigma'^2}}\,dx = \frac{1\,000}{\sqrt{2\pi}}\int_{\frac{h-\frac{1}{2}+a'}{\sigma'}}^{\frac{h+\frac{1}{2}+a'}{\sigma'}} e^{-\frac{t^2}{2}}\,dt =$$

$$= 1\,000\left[\Phi\left(\frac{h+0{,}567}{\sigma'}\right) - \Phi\left(\frac{h-0{,}433}{\sigma'}\right)\right].$$

Поэтому, беря попрежнему $\frac{1}{\sigma'} = 0{,}64$, перемещаем лишь центр и составляем таблицу III.

---

[1]) Для того, чтобы обнаружить эмпирически те незначительные различия, которые проистекают от замены точных значений членов симметричной формулы десятой степени бинома интегралами Лапласа для соответствующих промежутков, нужно было бы взять n не менее 10 000.

[2]) См. гл. IV, часть III, и, в частности, стр. 216.

ТАБЛИЦА III.

h	t	$1\,000\,\Phi(t)$	Y_h	$Y_h - Y'_h$	$\frac{(Y_h - Y'_h)^2}{\beta(h)}$
	— 3,48	— 499,7			
— 5			2	— 1	0,5
	— 2,84	— 497,8			
— 4			11,6	1,6	0,2
	— 2,20	— 486,1			
— 3			45,5	2,5	0,2
	— 1,56	— 440,6			
— 2			119,4	— 3,4	0,1
	— 0,92	— 321,2			
— 1			210,9	— 10,1	0,6
	— 0,28	— 110,3			
0			250,9	3,7	0,1
	0,36	140,6			
1			200,7	— 1,3	0
	1	341,3			
2			108,2	— 6,8	0,5
	1,64	449,5			
3			39,2	5,2	0,7
	2,28	488,7			
4			9,5	0,5	0
	2,92	498,2			
5			1,6	1,6	1,6
	3,56	499,8			

Здесь

$$D = \frac{4,5}{11} \neq 0,41,$$

и построенная эмпирически кривая Гаусса еще лучше согласуется с статистическими данными, чем вычисленная раньше.

4. Применение способа дисперсии для выяснения степени пригодности той или иной теоретической кривой распределения вероятностей состоит, таким образом, в том вообще, что вычисляется средний коэффициент дисперсии, т.-е. средняя арифметическая

$$D = \frac{1}{s} \sum \frac{(Y_h - Y'_h)^2}{\beta(h)} = \frac{1}{s} \sum \frac{(Y_h - Y'_h)^2}{Y_h\left(1 - \frac{Y_h}{n}\right)}, \qquad (211)$$

где s — число интервалов, на которые разбит весь данный статистический материал, Y_h — теоретическое число, а Y'_h — наблюденное число объектов,

соответствующее интервалу $\left(h-\frac{\delta}{2},\ h+\frac{\delta}{2}\right)$, длиною δ.

Для упрощения вычислений коэффициент D можно заменить другим коэффициентом

$$H=\frac{1}{s-1}\sum\frac{(Y_h-Y'_h)^2}{Y_h}, \tag{212}$$

который назовем коэффициентом точности рассматриваемой теоретической кривой; коэффициент точности также имеет математическое ожидание равное 1 и при больших значениях s по закону больших чисел должен мало отличаться от 1.

Действительно,

$$\text{М. О. }\frac{(Y_h-Y'_h)^2}{Y_h}=1-\frac{Y_h}{n};$$

поэтому

$$\text{М. О. }\sum_1^s\frac{(Y_h-Y'_h)^2}{Y_h}=s-\frac{\sum Y_h}{n}=s-1,$$

а отсюда М. О. $H=1$.

Коэффициент точности H в данном примере для теоретической кривой Гаусса (при $a=-0{,}067$, $\sigma'=1{,}554$) равен

$$H=\frac{1}{10}\sum\frac{(Y_h-Y'_h)^2}{Y_h}=0{,}43;$$

таким образом значения D и H практически одинаково характеризуют достигнутое приближение.

Разумеется, указанное применение способа дисперсии нисколько не требует, чтобы все промежутки были равны, но понятно, что, вообще говоря, точность приближения теоретической кривой (т.-е. коэффициенты H или D) может, в согласии с сделанными ранее замечаниями, быть не вполне одинаковой в зависимости от того или иного подразделения материала на промежутки. Поэтому, указывая найденное значение коэффициента точности H (или D), желательно для теоретической полноты упоминать о том, каковы были рассматриваемые интервалы.

Для наиболее точного вычисления параметров теоретической кривой целесообразно деление всего материала на возможно мелкие промежутки [1]) так, чтобы каждая эмпирическая величина x была учтена со всей точностью, какую допускало ее фактическое измерение; напротив, производя сравнение теоретического распределения с эмпирическим, не следует брать слишком малых интервалов для того, чтобы теоретические значения Y_h во всяком случае были не меньше нескольких единиц. Когда данные эмпирические промежутки настолько малы, что соответствующие им значения Y'_h не велики, то для сравнения кривых можно рекомендовать объединение нескольких промежутков в интервалы, не равные между собой, но которым соответствуют теоретически равные значения Y_h, например, $Y_h = \frac{n}{10}$. Если кривая распределения дана уравнением

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma}},$$

то соответствующие точки деления на интервалы будут ($-\infty$, $-1{,}28\sigma$; $-0{,}842\sigma$; $-0{,}524\sigma$; $-0{,}253\sigma$; 0; $0{,}253\sigma$; $0{,}524\sigma$; $0{,}842\sigma$; $1{,}28\sigma$; ∞). Эти числа мы получаем из таблицы функции $\Phi(z)$ как значения z, соответствующие значениям $\Phi(z)$, равным $-0{,}4$; $-0{,}3$; $-0{,}2$; $-0{,}1$; 0; 0,1 и т. д.

В 25 провинциях Швеции в 1883 году даны числа родившихся двоен на 1 000 рождений. Эти числа в порядке возрастания были:

10,1	13,4	14,3	15,2	16,8
13,1	13,4	14,3	15,2	17,7
	13,5	14,4	15,4	19,8
	13,6	14,4	15,6	
	13,8	14,5	15,8	
	13,8	14,5		
	13,9	14,8		
		14,9		

[1]) Расширение промежутков диктуется лишь желанием упростить статистическую технику.

Применяя указанный выше прием для вычисления (не производя никакой группировки) центра и штандарта, находим: $a = 14{,}648$; $\sigma = 1{,}73$. Разделяя весь материал на $s = 5$ интервалов, соответствующих $Y_h = \frac{n}{s} = \frac{25}{5} = 5$, находим точки деления: $a - 0{,}842\sigma = 13{,}19$; $a - 0{,}253\sigma = 14{,}21$; $a + 0{,}253\sigma = 15{,}09$; $a + 0{,}842\sigma = 16{,}11$; в действительности, в указанных промежутках вместо 5 было, соответственно 2, 7, 8, 5, 3 провинции; таким образом

$$H = \frac{1}{4}\left[\frac{(2-5)^2}{5} + \frac{(7-5)^2}{5} + \frac{(8-5)^2}{5} + \frac{(3-5)^2}{5}\right] = 1{,}3,$$

так что является довольно вероятным, что те же 25 чисел могли бы быть получены, если бы было выбрано наудачу 25 значений величины x, подчиняющейся закону Гаусса с центром 14,648 и штандартом 1,73. (Разумеется, при столь малых значениях n невозможно сколько-нибудь точно выявить закон распределения величины x.)

Однако, если бы для получения равновероятных интервалов пришлось дробить первоначально данные элементарные промежутки, то обычно лучше этого дробления не производить, ограничиваясь лишь объединением соседних промежутков, для которых Y'_h незначительны, что и будет сделано нами в некоторых дальнейших примерах.

5. Следует сделать еще одно замечание по поводу погрешности, проистекающей оттого, что действительно наблюденные значения приурочиваются к определенной точке h, середине промежутка $h \pm \frac{\delta}{2}$, между тем как наша теоретическая функция распределения вероятностей $f(x)$ непрерывна.

Допуская, что эмпирическое распределение идеально совпадает с теоретическим, т.-е. что $\frac{1}{n} Y'_h = \frac{1}{n} Y_h = \int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} f(x)\,dx$, мы не получили бы, однако, вполне точного совпадения между величинами

$$a = \text{М. О. } x = \int_{-\infty}^{\infty} x f(x)\,dx \quad \text{и} \quad a = \frac{1}{n}\sum hY_h;$$

и вообще

$$a_k = \text{М. О. } x^k = \int_{-\infty}^{\infty} x^k f(x)\,dx \text{ не равно } a'_k = \sum \frac{h^k Y_h}{n}, \text{ так как}$$

$$\int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} x^k f(x)\,dx \neq h^k \int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} f(x)\,dx.$$

Величину $a_k = \text{М. О. } x^k$ называют теоретическим моментом k-й степени, а a'_k называют иногда грубым моментом.

Постараемся установить поправку, которую нужно внести в значение a'_k, т.-е. определим множитель λ_k, которым надо заменить h^k, чтобы получить точное равенство

$$\int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} x^k f(x)\,dx = \lambda_k \int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} f(x)\,dx. \tag{213}$$

Для этой цели замечаем, что

$$\int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} f(x)\,dx = F\left(h+\frac{\delta}{2}\right) - F\left(h-\frac{\delta}{2}\right) =$$

$$= \left[F(h) + \frac{\delta}{2} f(h) + \frac{\delta^2}{4}\frac{f'(h)}{2!} + \ldots\right] - \left[F(h) - \frac{\delta}{2} f(h) + \ldots\right] =$$

$$= \delta f(h) + \frac{\delta^3}{24} f''(h) + \cdots$$

и аналогичным образом

$$\int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} x^k f(x)\,dx = \delta\,(h^k f(h)) + \frac{\delta^3}{24}\left(h^k f(h)\right)'' + \cdots$$

Поэтому

$$\lambda_k = \frac{\int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} x^k f(x)\,dx}{\int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} f(x)\,dx} = \frac{h^k f(h) + \frac{\delta^2}{24}\left(h^k f(h)\right)'' + \cdots}{f(h) + \frac{\delta^2}{24} f''(h) + \cdots}, \tag{214}$$

и, пренебрегая членами не ниже 4-й степени относительно $\frac{\delta}{2}$, имеем

$$\lambda_k = h^k\left[1+\frac{\delta^2}{24f(h)}\left(f''(h)+\frac{2k}{h}f'(h)+\frac{k(k-1)}{h^2}f(h)\right)-\frac{\delta^2}{24}\frac{f''(h)}{f(h)}\right]=$$

$$=h^k\left[1+\frac{\delta^2}{24}\left(\frac{2kf'(h)}{hf(h)}+\frac{k(k-1)}{h^2}\right)\right],$$

т.-е.

$$\int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} x^k f(x)\,dx = h^k\int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} f(x)\,dx+$$

$$+\frac{\delta^2}{24}\left[k(k-1)h^{k-2}+2kh^{k-1}\frac{f'(h)}{f(h)}\right]\int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} f(x)\,dx. \qquad (215)$$

Предположим сначала, что теоретическая функция $f(x)$ есть функция Гаусса $f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$; тогда $\frac{f'(h)}{f(h)}=\frac{(a-h)}{\sigma^2}$; поэтому, применяя равенство (215) ко всем промежуткам и складывая их, получим

$$a_k = a'_k+\frac{\delta^2}{24}\left[k(k-1)a'_{k-2}+\frac{2k}{\sigma^2}\left(aa'_{k-1}-a'_k\right)\right]. \qquad (216)$$

Полагая $k=1$, находим ($a_0=a'_0=1$)

$$a = a'+\frac{\delta^2}{12\sigma^2}\left(a-a'\right),$$

откуда $a=a'$ (с точностью порядка $\frac{\delta^4}{16}$), так что погрешностью в определении центра a, проистекающей от замены интеграла суммой, можно пренебречь и, перенеся начало координат в центр, положить $a=a'=0$. В таком случае равенство (216) примет вид

$$a_k = a'_k+\frac{\delta^2}{24}\left[k(k-1)a'_{k-2}-\frac{2ka'_k}{\sigma^2}\right], \qquad (217)$$

и, в частности, для $k=2$ ($a_2=\sigma^2$, $a'_2=\sigma'^2$),

$$\sigma^2=\sigma'^2+\frac{\delta^2}{12}\left(1-\frac{2\sigma'^2}{\sigma^2}\right),$$

откуда, пренебрегая опять величинами порядка $\frac{\delta^4}{16}$, получим

$$\sigma^2=\sigma'^2+\frac{\delta^2}{12}\left(1-2\right)=\sigma'^2-\frac{\delta^2}{12}. \qquad (218)$$

Возвращаясь к общему равенству (213), где мы допускаем лишь, что теоретическая функция распределения $f(x)$ обладает конечными производными первых трех порядков, можно, пользуясь тем же приемом, который привел нас к равенству (215), заменить

$$\frac{h^{k-1}f'(h)}{f(h)}\int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} f(x)\,dx$$

через

$$\int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} x^{k-1}f'(x)\,dx = x^{k-1}f(x)\Big|_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} - (k-1)\int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} x^{k-2}f(x)\,dx, \qquad (219)$$

пренебрегая членами порядка $\frac{\delta^2}{4}$. Тогда, благодаря тому, что рассматриваемое выражение входит в (215) с множителем $\frac{k\delta^2}{12}$, порядок погрешности равенства (215) после этой замены останется попрежнему равным $\frac{\delta^4}{16}$, и мы получим

$$\int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} x^k f(x)\,dx = h^k\int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} f(x)\,dx +$$

$$+\frac{\delta^2}{24}\Bigg[k(k-1)h^{k-2}\int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} f(x)\,dx -$$

$$-2k(k-1)\int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} x^{k-2}f(x)\,dx + 2kx^{k-1}f(x)\Bigg|_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}}\Bigg]$$

и с тою же степенью абсолютной и относительной точности

$$\int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} x^k f(x)\,dx = h^k\int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} f(x)\,dx -$$

$$-\frac{k(k-1)\delta^2}{24}\left[h^{k-2}\int_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}} f(x)\,dx - \frac{2x^{k-1}}{k-1}f(x)\Bigg|_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}}\right]$$

Распространяя это равенство на все промежутки и складывая их, находим наконец

$$a_k = a'_k - \frac{k(k-1)}{24}a'_{k-2}\delta^2. \qquad (220)$$

так как члены вида $x^k f(x) \Big|_{h-\frac{\delta}{2}}^{h+\frac{\delta}{2}}$ взаимно сокращаются и, кроме того, раз мы допускаем, что М. О. x^k конечно, то $x^k f(x)$ стремится к 0 при $x = \pm\infty$.

Равенство (220) представляет собою поправочную формулу [1]) Шеппарда, соответствующую приближению с точностью до $\frac{\delta^4}{16}$. Следовательно, в частности,

$$a = a', \quad \sigma^2 = \sigma'^2 - \frac{\delta^2}{12}, \tag{218 bis}$$

какова бы ни была функция $f(x)$, удовлетворяющая вышеуказанным условиям.

Упражнения. 1) В биографическом словаре Poggendorf'a рассмотрено 225 пар страниц. Число биографий, начинающихся на каждой паре страниц, дано в таблице.

Число биографий x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
„ пар страниц	1	3	1	7	18	21	34	36	32	26	19	10	7	5	2	0	2	1

Вычислить центр a и штандарт σ распределения вероятностей; проверить, согласуются ли эти данные с допущением нормального распределения.

Ответ: $a = 7{,}44$; $\sigma = 2{,}76$. Имея в виду, что, при столь незначительном $n = 225$, дробление на 18 промежутков является чрезмерным, объединяем два соседние промежутка в 1 интервал и получаем эмпирические и теоретические ряды:

	Эмпирич. данные	Теоретич. данные	Теоретич. вероятн.	$\frac{x-a}{\sigma} = t$
$x < 1{,}5$	4	3,6	0,016	$t < -2{,}156$
$1{,}5 < x < 3{,}5$	8	13,5	0,060	$-2{,}156 < t < -1{,}43$
$3{,}5 < x < 5{,}5$	39	36,8	0,162	$-1{,}43 < t < -0{,}705$
$5{,}5 < x < 7{,}5$	70	60	0,267	$-0{,}705 < t < 0{,}02$
$7{,}5 < x < 9{,}5$	58	59,4	0,264	$0{,}02 < t < 0{,}746$
$9{,}5 < x < 11{,}5$	29	35,3	0,157	$0{,}746 < t < 1{,}47$
$11{,}5 < x < 13{,}5$	12	12,8	0,057	$1{,}47 < t < 2{,}195$
$13{,}5 < x$	5	3,2	0,014	$2{,}195 < t$

[1]) Формула (220) с тою же степенью точности эквивалентна, конечно, равенству (217), так как, для $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}}$, $a_k = (k-1)a_{k-2}\sigma^2$; поэтому, заменяя в скобках второй части (217) $\frac{a'_k}{\sigma^2}$ через $(k-1)\,a'_{k-2}$, получаем (220).

Коэффициент точности $H \neq 0{,}9$ удовлетворителен.

2) В связи с своими исследованиями по теории наследственности, английский ученый Гальтон измерил рост 928 родительских пар, определяя рост родительской пары по формуле $\frac{x+1{,}08\,y}{2}$, где x — рост отца, а y — рост матери. Измерения, произведенные в дюймах, дали

37	144	430	251	62	4	род. пары
63 — 65	65 — 67	67 — 69	69 — 71	71 — 73	73 — 75	рост

Определить соответствующую этому распределению нормальную кривую и выяснить степень ее пригодности.

Ответ. Средний рост $a = 68{,}364''$ при штандарте $\sigma = 1{,}853$. Поэтому теоретические числа равны (в порядке возрастания роста) 31; 184; 374; 269; 66; 5. Следовательно, коэффициент точности

$$H = \frac{1}{5}\left[\frac{36}{31} + \frac{1\,600}{184} + \frac{3\,136}{374} + \frac{324}{269} + \frac{16}{66} + \frac{1}{5}\right] \neq 4$$

слишком велик, чтобы можно было считать данное распределение роста нормальным.

ГЛАВА ВТОРАЯ

АНАЛИЗ СТАТИСТИЧЕСКИХ КРИВЫХ РАСПРЕДЕЛЕНИЯ, УКЛОНЯЮЩИХСЯ ОТ НОРМАЛЬНОГО ТИПА

1. Если нам дана весьма многочисленная совокупность, состоящая из n объектов, у которых измерен некоторый признак x, то прежде всего для характеристики этой совокупности относительно рассматриваемого признака следует вычислить среднюю величину $a' = \frac{1}{n}\sum x$ признака и его штандарт $\sigma' = \sqrt{\frac{1}{n}\sum (x-a)^2}$. Построив после этого нормальную кривую, имеющую указанные параметры a и σ', мы нередко убеждаемся, что соответствующий коэффициент точности H настолько велик, что отклонения от нормального распределения нельзя признать случайными.

В таком случае перед нами стоят две задачи: 1) указать возможно простую теоретическую функцию распределения вероятностей $f(x)$, уклонение которой от данного распределения, измеряемое, например, коэффициентом точности H, было бы достаточно мало, чтобы его можно было приписать случайности; 2) дать теоретическое обоснование выбора указанной функции $f(x)$, руководствуясь принципами теории вероятностей и элементарными физическими или биологическими (или экономическими) законами, на почве которых происходит рассматриваемое явление. Вторая задача не допускает единообразного математического решения, и, откладывая до конца главы некоторые принципиальные соображения по этому поводу, займемся первой задачей — интерполирования или приближенного изображения данной статистической кривой при помощи простых математических формул.

Как было разъяснено выше, достаточно вообще ограничиться предположением, что действительное (эмпирическое) распределение вероятностей непрерывно и плотность вероятностей выражается некоторой непрерывной положительной функцией $\varphi(x)$, которая стремится к 0 при $x = \pm\infty$.

Если бы на функцию $\varphi(x)$ не налагали никаких других ограничений, то и тогда возможно было бы представить ее с какой угодно точностью при помощи произведения $P(x)e^{-\frac{(x-a)^2}{2\sigma^2}}$, где a и σ — произвольные данные числа, а $P(x)$ — соответствующим образом подобранный многочлен, так что при всех значениях x

$$\left|\varphi(x) - P(x)e^{-\frac{(x-a)^2}{2\sigma^2}}\right| < \varepsilon, \tag{221}$$

как бы мало ни было положительное число ε

В частности, можем взять $a = \text{М. О. } x$, $\sigma^2 = \text{М. О. } (x-a)^2$ на основании данного статистического распределения, соответствующего действительной функции распределения $\varphi(x)$.

Если бы, взявши вместо многочлена $P(x)$ постоянную величину, можно было удовлетворить неравенству (221), то распределение было бы нормальным (с соответствующей погрешностью ε). Вообще, полагая

$$\varphi(x) = \frac{F(x)e^{-\frac{(x-a)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} \tag{222}$$

(после указанного выбора a и σ), назовем $F(x)$ пертурбационной функцией, так как она является множителем[1]), на который надо умножить нормальную плотность (соответствующую тому же центру и штандарту) для того, чтобы получить действительную плотность распределения вероятностей.

В силу вышесказанного, эту пертурбационную функцию $F(x)$ всегда можно заменить приближенным многочленом $P(x)$ так, чтоб

$$\left|\frac{P(x)}{\sigma\sqrt{2\pi}} - \frac{F(x)}{\sigma\sqrt{2\pi}}\right| e^{-\frac{(x-a)^2}{2\sigma^2}} < \varepsilon. \tag{221 bis}$$

---

[1]) Очевидно, что для тех значений x, для которых $F(x) > 1$, действительная кривая распределения имеет ординату бо́льшую, чем соответствующая кривая Гаусса, и, напротив, ее ордината меньше, когда $F(x) < 1$. Функция $F(x)$ характеризует таким образом дополнительные причины, которые отклоняют действительное распределение частоты признака от нормального.

Вводя вместо величины x нормированную величину $\frac{x-a}{\sigma}$, можем свести общий случай к тому, когда центр находится в начале координат, а штандарт равен 1. Поэтому будем полагать вообще, что

$$\varphi(x)=\frac{F(x)}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}. \qquad (222\text{ bis})$$

2. Как было разъяснено в предыдущей главе, осуществление неравенства (221) для всякого x не имеет практического значения и не может быть экспериментально проверено: для сравнения всегда берутся конечные, хотя и небольшие, промежутки (x_i, x_{i+1}), и если n есть общее число индивидов статистической совокупности, то для каждого промежутка рассматриваются разности

$$Y-Y'=n\int_{x_i}^{x_{i+1}} f(x)\,dx-n\int_{x_i}^{x_{i+1}}\varphi(x)\,dx.$$

Обозначая, кроме того, через s число промежутков, определяем коэффициент точности

$$H=\frac{1}{s-1}\sum\frac{(Y-Y')^2}{Y}=\frac{n}{s-1}\sum\frac{\left[\int_{x_i}^{x_{i+1}}[f(x)-\varphi(x)]\,dx\right]^2}{\int_{x_i}^{x_{i+1}} f(x)\,dx}. \qquad (212\text{ bis})$$

Предполагая премежутки (x_i, x_{i+1}) достаточно малыми, можно каждый член последней суммы заменить через

$$\frac{[f(x)-\varphi(x)]^2(x_{i+1}-x_i)^2}{f(x)(x_{i+1}-x_i)}=\frac{[f(x)-\varphi(x)]^2}{f(x)}(x_{i+1}-x_i).$$

Таким образом, если бы промежутки были выбраны так, что соответствующие им теоретические числа $Y=1$, т.-е. $n=s$, то

предел H, при $n=\infty$, был бы равен

$$\text{пред. } H=H'=\int_{-\infty}^{\infty}\frac{[f(x)-\varphi(x)]^2}{f(x)}\,dx; \tag{223}$$

в частности, если теоретической функцией является

$$f(x)=\frac{P(x)}{\sqrt{2\pi}}\,e^{-\frac{x^2}{2}}, \tag{224}$$

где $P(x)$ — многочлен, то последний интеграл можем записать в виде

$$H'=\sqrt{2\pi}\int_{-\infty}^{\infty}\left[\frac{P(x)e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}-\varphi(x)\right]^2\frac{e^{\frac{x^2}{2}}}{P(x)}\,dx=$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left[P(x)-F(x)\right]^2\frac{e^{-\frac{x^2}{2}}}{P(x)}\,dx; \tag{225}$$

естественно поэтому для возможно лучшего с точки зрения теории вероятностей приближения функции $\varphi(x)$ при помощи функции вида (224), в которой степень многочлена $P(x)$ не выше данного числа k, так подыскивать коэффициенты $P(x)$, чтобы обратить интеграл (225) в минимум.

Во всяком случае, если хотим, чтобы коэффициент точности H' был близок к 1, то необходимо, чтобы H' или соответствующее ему приближенное выражение было мало.

Обычно среднее значение $\frac{n}{s}$ числа Y индивидов, приходящихся на данный интервал, равно по крайней мере нескольким десяткам; однако, если, выбирая степень k многочлена $P(x)$ достаточно высокой, возможно H' сделать менее всякой наперед заданной величины, то, как видно из (212 bis), и коэффициент точности H' $\left(\text{при данном даже весьма большом } \frac{n}{s}\right)$ станет произвольно малым, каковы бы ни были промежутки. В действительности, отыскание минимума интеграла H', где $\varphi(x)$ [или $F(x)$] есть данная функция, представляет сложную математическую задачу, которую без существенного ущерба можно значительно упростить, отбросивши знаменатель

$P(x)$ в подынтегральной функции (225) и рассматривая таким образом вместо H' интеграл

$$I=\sqrt{2\pi}\int_{-\infty}^{\infty}\left[\frac{P(x)e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}-\varphi(x)\right]^2 e^{\frac{x^2}{2}}dx=$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left[P(x)-F(x)\right]^2 e^{-\frac{x^2}{2}}dx. \qquad (226)$$

Прежде чем перейти к решению задачи отыскания коэффициентов многочлена $P(x)$ степени не выше k, при которых интеграл I обращается в минимум, следует заметить, что на функцию $\varphi(x)$ необходимо наложить существенные, с теоретической точки зрения, ограничения для того, чтобы с увеличением степени k значение интеграла I (или H') могло быть сделано малым.

Действительно, если данная функция $\varphi(x)=\lambda(x)e^{-\frac{x^2}{4}}$, где $\lambda(x)>\lambda>0$ [т.-е. $\lambda(x)$ не стремится к 0 при $x=\pm\infty$], то интеграл

$$I=\sqrt{2\pi}\int_{-\infty}^{\infty}\left[\frac{P(x)e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}-\varphi(x)\right]^2 e^{\frac{x^2}{2}}dx=$$

$$=\sqrt{2\pi}\int_{-\infty}^{\infty}\left[\frac{P(x)e^{-\frac{x^2}{4}}}{\sqrt{2\pi}}-\lambda(x)\right]^2 dx \qquad (226\text{ bis})$$

будет бесконечно велик, как бы мы ни выбирали многочлен данной степени $P(x)$, так как при $x=\pm\infty$ подынтегральная функция не стремится к 0. Отсюда видно, что порядок убывания функции $\varphi(x)$ не может быть произвольным. Можно было бы доказать следующую теорему.

Для того, чтобы с увеличением степени k многочлена $P(x)$ интеграл I (а также H') мог быть сделан произвольно малым, достаточно, чтобы $\lambda(x)<\frac{A}{(1+x^2)^\alpha}$, где $\alpha>\frac{1}{4}$ и $A>0$ — некоторые произвольные, но определенные числа.

Не останавливаясь на доказательстве этой теоремы, которая вытекает из ранее *формулированного общего положения* (221), заметим, что хотя с узко практической точки зрения указанное

ограничение не имеет значения, поскольку вообще пределы изменяемости всякой наблюдаемой величины конечны, но оно показывает, что прием, к изложению которого мы теперь перейдем, может оказаться непригодным, если по тем или иным теоретическим соображениям, основанным на свойствах изучаемого явления, необходимо было бы признать, что функция распределения $\varphi(x)$ убывает не столь быстро, как этого требует высказанная теорема[1]).

3. Итак, ищем для функции (224) многочлен k-й степени $P(x)=A_0+A_1x+\ldots+A_kx^k$, по условию, чтобы интеграл

$$I=\sqrt{2\pi}\int_{-\infty}^{\infty}\left[(A_0+A_1x+\ldots+A_kx^k)\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}-\varphi(x)\right]^2 e^{\frac{x^2}{2}}\,dx \quad (226)$$

был минимум.

Так как I является непрерывной и дифференцируемой функцией $k+1$ переменных $A_0, A_1, \ldots, A_k$, то для минимума, который несомненно существует (ибо I не может быть отрицательным), необходимо, чтобы

$$\frac{\partial I}{\partial A_0}=\frac{\partial I}{\partial A_1}=\cdots=\frac{\partial I}{\partial A_k}=0,$$

т.-е.

$$\int_{-\infty}^{\infty}[f(x)-\varphi(x)]\,x^h\,dx=$$

$$=\int_{-\infty}^{\infty}\left[\frac{P(x)}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}-\varphi(x)\right]x^h\,dx=0 \quad (h=0,1,\ldots,k),$$

---

[1]) Заметим также, что если многочлен $P(x)$ данной степени k, обращающий в минимум интеграл I, для некоторых значений x приближается к 0 (или тем более, если он становится отрицательным), то H' может при этом стать очень большим. Практически это будет не существенно лишь в том случае, когда соответствующие значения x в несколько раз превышают штандарт σ, так что найденный закон распределения характеризует главное ядро совокупности, за исключением немногих ее элементов, для которых величина признака очень удалена от центра.

откуда, полагая

$$\int_{-\infty}^{\infty} x^h \varphi(x)\,dx = a_h,$$

выводим $k+1$ уравнений

$$\int_{-\infty}^{\infty} f(x) x^h\,dx = a_h, \tag{227}$$

которые носят название уравнений моментов и служат для определения $k+1$ коэффициентов A_h. В развернутой форме уравнения (227) имеют вид[1]):

$$\begin{aligned}
A_0 + A_1 M_1 + \ldots + A_k M_k &= a_0 = 1\\
A_0 M_1 + A_1 M_2 + \ldots + A_k M_{k+1} &= a_1 = 0\\
A_0 M_2 + A_1 M_3 + \ldots + A_k M_{k+2} &= a_2 = 1\\
A_0 M_3 + \ldots\ldots\ldots + A_k M_{k+3} &= a_3\\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
A_0 M_k + \ldots\ldots\ldots + A_k M_{2k} &= a_k,
\end{aligned} \tag{227 bis}$$

где

$$M_h = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^h e^{-\frac{x^2}{2}}\,dx.$$

Нетрудно видеть, что определитель коэффициентов при неизвестных A_h отличен[2]) от 0, т.-е. уравнения (227 bis) допускают только одну систему решений; следовательно, минимум интеграла I осуществляется именно при этих значениях A_h, удовлетворяющих уравнениям (227 bis).

---

[1]) Равенства $a_1 = 0$, $a_2 = 1$ вытекают из сделанного выше предположения, что центр распределения находится в начале координат, а штандарт равен 1.

[2]) Для этого достаточно заметить, что если $\varphi(x) = 0$, то, складывая все $k+1$ уравнений, помноживши $(h+1)$-е из них соответственно на A_h, получим $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} P^2(x)\, e^{-\frac{x^2}{2}}\,dx = 0$, откуда видно, что $P(x)$ тождественно равно 0, а потому уравнения (227 bis), которые в этом случае однородны, не имеют других решений, кроме $A_h = 0$; как известно, это означает, что определитель из коэффициентов при A_h не равен нулю.

Замечая, что вообще интегрирование по частям дает

$$\int_{\infty}^{\infty} x^{h+2} e^{-\frac{x^2}{2}} dx = -x^{h+1} e^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} + (h+1) \int_{-\infty}^{\infty} x^h e^{-\frac{x^2}{2}} dx =$$

$$= (h+1) \int_{-\infty}^{\infty} x^h e^{-\frac{x^2}{2}} dx,$$

находим, что, при h нечетном, $\int_{-\infty}^{\infty} x^h e^{-\frac{x^2}{2}} dx = 0$, а в случае $h = 2p$

$$M_{2p} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2p} e^{-\frac{x^2}{2}} dx = (2p-1)(2p-3) \ldots 3 \cdot 1. \qquad (228)$$

Положим сначала $k = 2$; тогда наши уравнения получают вид

$$A_0 + A_2 = 1, \quad A_1 = 0, \quad A_0 + 3A_2 = 1, \qquad (229)$$

откуда заключаем, что $A_1 = A_2 = 0$ и $A_0 = 1$. Иными словами, если многочлен $P(x)$ не должен быть выше 2-й степени, то $P(x) = 1$, и минимум интеграла I осуществляется функцией Гаусса $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

Беря $k = 3$, получим 4 уравнения

$$A_0 + A_2 = 1, \quad A_1 + 3A_3 = 0, \quad A_0 + 3A_2 = 1,$$
$$3A_1 + 3 \cdot 5A_3 = a_3, \qquad (230)$$

откуда

$$A_0 = 1, \quad A_2 = 0, \quad A_1 = -\frac{a_3}{2}, \quad A_3 = \frac{a_3}{6}.$$

Рассмотрим еще случай, когда степень многочлена $P(x)$ равна $k = 4$.

Тогда мы имеем 5 уравнений:

$$A_0 + A_2 + 3A_4 = 1, \quad A_0 + 3A_2 + 15a_4 = 1;$$
$$3A_0 + 15A_2 + 105A_4 = a_4, \qquad (231)$$
$$A_1 + 3A_3 = 0, \quad 3A_1 + 15A_3 = a_3.$$

Положим

$$\frac{a_3}{2}=S,\ \frac{a_4-3}{8}=E; \tag{232}$$

величину S будем называть асимметрией данного распределения, а величину E назовем эксцессом. Введя эти обозначения, находим сразу из последних двух уравнений (231)

$$A_1=-S,\ A_3=\frac{S}{3}; \tag{233}$$

для определения же коэффициентов с четными значками складываем первые 3 уравнения, предварительно умноживши первое на 3, а второе на — 6, откуда находим

$$24A_4=a_4-3,\ \text{или}\ A_4=\frac{E}{3}, \tag{234}$$

и, подставляя это значение в первые два уравнения, получаем

$$A_0=1+E,\quad A_2=-2E. \tag{234 bis}$$

Таким образом пертурбационный многочлен 4-й степени

$$P(x)=1+E-Sx-2Ex^2+\frac{S}{3}x^3+\frac{E}{3}x^4=$$
$$=1-S\left(x-\frac{x^3}{3}\right)+E\left(1-2x^2+\frac{x^4}{3}\right). \tag{235}$$

Можно было бы еще повысить степень многочлена $P(x)$, но на практике этого обычно не делают, так как погрешность вычисления последовательных величин a_h, т.-е. последовательных моментов, на основании статистических данных увеличивается с увеличением степени h благодаря тому, что случайное присутствие или отсутствие одного или нескольких индивидов, у которых размеры x сильно уклоняются от центра, оказывают тем большее влияние на приближенную величину a_h, чем больше h; никакие поправки Шеппарда, о которых мы говорили выше, не могут существенно улучшить положение вещей, если только статистическая совокупность не исключительно велика.

По той же причине трудно дать достаточно обоснованное значение штандарта погрешности статистического определения моментов a_k при $k>2$ и связанных с ними величин S и, в особенности, E.

Вообще, как и для всех статистических величин, погрешность убывает обратно пропорционально $\sqrt{n}$; если же n очень велико и эмпирическая кривая мало уклоняется от нормальной, то штандарты погрешностей S и E приближенно равны

$$\sigma(S) \neq \frac{1{,}225}{\sqrt{n}}, \quad \sigma(E) \neq \frac{0{,}6124}{\sqrt{n}}. \tag{236}$$

4. Если допустим, что действительное значение $F(x)$ точно совпадает с $P(x)$, то плотность распределения вероятностей выразится таким образом функцией

$$\varphi(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\left(1 + E - Sx - 2Ex^2 + \frac{S}{3}x^3 + \frac{E}{3}x^4\right). \tag{235 bis}$$

Кривая будет симметрична, если $S=0$, и будет совпадать с кривой Гаусса, когда, кроме того, эксцесс $E=0$.

Из формы функции $\varphi(x)$ видно, что если $E>0$, то центральная ордината кривой ($x=0$) равна $\frac{1+E}{\sqrt{2\pi}}$, т.-е. больше ординаты соответствующей нормальной кривой (отсюда происхождение термина „эксцесс"), и, наоборот, меньше последней, когда $E<0$. Нетрудно проверить [1]) (ограничиваясь случаем $S=0$), что функция $\varphi(x)$ не будет становиться отрицательной лишь при условии, что $0<E\leqslant\frac{1}{2}$; кроме того, если $E>\frac{1}{2}$, то $\varphi(x)$ делается отрицательной уже при $|x|=\sqrt{3}$ или даже раньше; поэтому указанная форма (235 bis) невозможна при $E>\frac{1}{2}$. Что касается отрицательных эксцессов, то вообще никогда не бывает $E<-\frac{1}{4}$, так как $a_4 \geqslant a_2^2 = 1$; но при $E<0$ $\varphi(x)$ всегда становится отрицательным для больших значений $|x|$; только если $|E|$ очень мал, то это не является практическим затруднением; во всяком случае, при $E<0$ практи-

[1]) Так как $1+E-2Ex^2+\frac{E}{3}x^4 = \frac{E}{3}\left[(x^2-3)^2+\frac{3}{E}-6\right]$.

чески, для применимости (235 bis), большею частью достаточно, чтобы $E| \leqslant \frac{1}{10}$, так как тогда $\varphi(x)$ становится отрицательной лишь при $|x| \geqslant 3$.

Если $S \gtrless 0$, то кривая асимметрична: при положительной асимметрии $S > 0$, ордината вблизи центра убывает слева направо; наоборот, при отрицательной асимметрии $S < 0$, кривая вблизи центра возрастает слева направо $\left(\text{так как } \varphi'(0) = -\frac{S}{\sqrt{2\pi}}\right)$. Если S и E достаточно малы, то нетрудно приближенно вычислить абсциссу μ (моду), соответствующую максимальной ординате кривой $\varphi(x)$. Действительно, мода μ удовлетворяет уравнению $\varphi'(x) = 0$, т.-е.

$$\left(-S - 4Ex + Sx^2 + \frac{4}{3}Ex^3\right) -$$

$$-x\left(1 + E - Sx - 2Ex^2 + \frac{S}{3}x^3 + \frac{E}{3}x^4\right) = 0, \qquad (237)$$

или, пренебрегая высшими степенями x и произведением Ex, получаем приближенно

$$\mu \neq -S. \qquad (238)$$

Таким образом во всяком нормированном (т.-е. имеющем штандарт $\sigma = 1$) распределении вероятностей, мало уклоняющемся от нормального, расстояние от центра до моды приближенно равно асимметрии S, при чем мода находится влево от центра при положительной асимметрии и вправо — при отрицательной.

Если признак не нормирован и М. О. $x = a$, М. О. $(x-a)^2 = \sigma^2$, то функция $\varphi(x)$, очевидно, имеет форму

$$\varphi(x) = \frac{e^{-\frac{(x-a)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}\left[1 + E - S\frac{x-a}{\sigma} - 2E\left(\frac{x-a}{\sigma}\right)^2 + \right.$$

$$\left. + \frac{S}{3}\left(\frac{x-a}{\sigma}\right)^3 + \frac{E}{3}\left(\frac{x-a}{\sigma}\right)^4\right], \qquad (239)$$

при чем нужно иметь в виду, что асимметрия и эксцесс

$$S=\frac{1}{2}\frac{\text{M. O. }(x-a)^3}{\sigma^3},\quad E=\frac{1}{8}\left[\frac{\text{M. O. }(x-a)^4}{\sigma^4}-3\right] \qquad (232\text{ bis})$$

являются отвлеченными числами, не зависящими от единиц измерения. Напротив, μ представляет собой расстояние или, точнее, разность между двумя значениями признака x и измеряется в тех же единицах, что x и его штандарт; поэтому в общем случае

$$\mu \neq -S\sigma. \qquad (238\text{ bis})$$

5. Переходя от дифференциальной функции распределения $\varphi(x)$ к интегральной (возвращаясь опять для простоты письма к случаю $a=0$, $\sigma=1$), находим, что вероятность неравенства

$$x_0 < x < x_1$$

равна интегралу

$$\frac{1}{\sqrt{2\pi}}\int_{x_0}^{x_1} e^{-\frac{x^2}{2}} P(x)\,dx = F_1(x_1) - F_1(x_0), \qquad (240)$$

где

$$F_1(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{x^2}{2}}\left(1+E-Sx-2Ex^2+\frac{S}{3}x^3+\frac{E}{3}x^4\right)dx=$$

$$=\frac{1}{2}+\Phi(x)-\frac{S}{3}\Phi'''(x)+\frac{E}{3}\Phi^{(IV)}(x)=$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{x^2}{2}}dx+\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\left[\frac{S}{3}(1-x^2)+\frac{E}{3}(3x-x^3)\right], \qquad (241)$$

так как из значений последовательных производных функции $\Phi(x)$

$$\Phi'(x)=\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}},\quad \Phi''(x)=\frac{-xe^{-\frac{x^2}{2}}}{\sqrt{2\pi}},\quad \Phi'''(x)=\frac{x^2-1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},$$

$$\Phi^{(IV)}(x)=\frac{3x-x^3}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},\quad \Phi^{(V)}(x)=\frac{3-6x^2+x^4}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

вытекает, что подъинтегральная функция

$$\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}P(x)=\Phi'(x)-\frac{S}{3}\Phi^{(IV)}(x)+\frac{E}{3}\Phi^{(V)}(x). \qquad (235\ \text{ter})$$

Из выражения $F_1(x)$ видим, что медиана распределения, т.-е. абсцисса $x=M$, для которой $F_1(x)=\frac{1}{2}$ (так что столь же вероятно, что величина x менее M, как то, что она больше M), не совпадает с центром 0, если $S \gtrless 0$. Для определения M, вследствие (241), имеем уравнение

$$\Phi(x)-\frac{S}{3}\Phi'''(x)+\frac{E}{3}\Phi^{(IV)}(x)=0,$$

или, пренебрегая высшими степенями x и произведением Ex, находим в первом приближении

$$x+\frac{S}{3}=0,$$

т.-е.

$$M \neq -\frac{S}{3}. \qquad (242)$$

Следовательно, для кривых, мало уклоняющихся от нормальной, медиана находится между центром и модой на расстоянии вдвое большем от моды, чем от центра.

В качестве примера рассмотрим измерения ширины 12 000 фасолей (Phaseolus vulgaris), произведенные проф. Johannsen в Копенгагене, численные результаты которых приведены в книге Charlier. Численные данные и результаты вычислений указаны в следующей таблице. Центр $a=8{,}562$ *мм* и $\sigma=0{,}62$ *мм* вычислены согласно вышепоказанному приему средней арифметической.

Ширина z в миллиметрах	Число объектов	Теорет. числа нормальн. распределения	Исправл. теоретич. числа
6,25 — 6,50	3 } 8	21,6	9,6
6,50 — 6,75	5		
6,75 — 7	24	49,2	28,8
7 — 7,25	103	133,2	98,4
7,25 — 7,5	259	319,2	288
7,5 — 7,75	624	618	615,6

Ширина z в миллиметрах	Число объектов	Теорет. числа нормальн. распределения	Исправл. теоретич. числа
7,75 — 8	1 187	1 035,6	1 101,4
8 — 8,25	1 650	1 525,2	1 651,2
8,25 — 8,5	1 883	1 820,4	1 936,8
8,5 — 8,75	1 930	1 892,4	1 921,2
8,75 — 9	1 638	1 681,2	1 602
9 — 9,25	1 130	1 302	1 168,8
9,25 — 9,5	737	816	720
9,5 — 9,75	427	448,8	416,4
9,75 — 10	221	214,8	236,4
10 — 10,25	110	82,8	117,8
10,25 — 10,5	57	27,6	54
10,5 — 10,75	24		
10,75 — 11	6 } 32	12	34,8
11 — 11,25	2		

Теоретические [1]) числа (3-й столбец), соответствующие нормальной кривой с указанными центром и штандартом, при внимательном сравнении со 2-м столбцом обнаруживают значительные расхождения. Если объединить в одну группу объекты, у которых признак уклоняется от средней более, чем на 3σ, особенно ярко бросается в глаза неудовлетворительность обеих крайних групп: в то время как нормальное распределение требовало бы, чтобы лишь 0,001 всех фасолей была шире 10,5 *мм*, таких оказалось 32 на 12 000; уклонение $32 - 12 = 20$ при штандарте $\sqrt{12 \cdot 0{,}999} \neq 3{,}4$ практически почти невозможно. Но не только на краях, где мы видим определенную тенденцию преувеличенного, по сравнению с нормальным, числа широких фасолей за счет узких, некоторые средние числа так же невероятны, как, например, число 1 130 вместо 1 302 для ширины 9 — 9,25 *мм*; здесь значение $\frac{(Y-Y')^2}{Y} = \frac{172^2}{1\,302} > 22$ так же неправдоподобно. Короче говоря, вычисляя коэффициент H для 17 групп, мы находим неприемлемое значение $H \neq 10$. Вычисляя после этого значение коэффициента асимметрии S и эксцесс E, мы находим

$$S = 0{,}141, \qquad E = 0{,}03.$$

[1]) Сначала при помощи функции Лапласа вычисляются вероятности, соответствующие каждому интервалу, а затем они умножаются на $n = 12\,000$; эти числа получаются непосредственно из соответствующих чисел 2-го столбца следующей таблицы.

Поэтому нормированная теоретическая функция

$$\frac{1}{2}+\Phi(x)-\frac{S}{3}\Phi'''(x)+\frac{E}{3}\Phi^{(IV)}(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}P(x)\,e^{-\frac{x^2}{2}}\,dx=F_1(x)$$

имеет пертурбационный многочлен

$$P(x)=1{,}03-0{,}141\,x-0{,}06\,x^2+0{,}047\,x^3+0{,}01\,x^4.$$

Для вычисления теоретических чисел, соответствующих исправленной функции $F_1(x)$, пользуемся [1]) таблицей.

$x=\frac{z-a}{\sigma}$	$\frac{1}{2}+\Phi(x)$	x^2-1	$\Phi'(x)$	$-\frac{S}{3}\Phi'''(x)$	$\frac{E}{3}\Phi^{(IV)}(x)$	$F_1(x)$	$\Delta F_1(x)$
—2,92	18	7,52	56	— 19	+ 9	8	8
—2,52	59	5,35	167	— 41	+ 14	32	24
—2,12	170	3,49	422	— 69	+ 13	114	82
—1,71	436	1,92	925	— 82	0	354	240
—1,31	951	0,71	1 691	— 56	—28	867	513
—0,91	1 814	—0,18	2 637	+ 22	—52	1 784	917
—0,50	3 085	—0,75	3 521	+ 124	—49	3 160	1 376
—0,10	4 602	—0,99	3 970	+ 184	—12	4 774	1 614
0,30	6 179	—0,91	3 814	+ 163	+33	6 375	1 601
0,70	7 580	—0,51	3 123	+ 75	+55	7 710	1 335
1,11	8 665	0,23	2 155	— 23	+42	8 684	974
1,51	9 345	1,28	1 276	— 76	+15	9 284	600
1,91	9 719	2,64	644	— 80	— 8	9 631	347
2,32	9 898	4,38	270	— 55	—15	9 828	197
2,72	9 967	6,39	99	— 29	—12	9 926	98
3,12	9 990	8,73	31	— 12	— 7	9 971	45
∞	10 000		0	0	0	29	29

Числа последнего столбца, умноженные на 1,2, дают теоретические значения последнего столбца предшествующей таблицы. Вычисляя коэффициент точности H для 17 промежутков, получаем

$$H=\frac{1}{16}\left[\frac{(1{,}6)^2}{9{,}6}+\frac{(4{,}8)^2}{28{,}8}+\frac{(4{,}6)^2}{98{,}4}+\frac{(49)^2}{288}+\frac{(8{,}4)^2}{615{,}6}+\frac{(85{,}6)^2}{1101{,}4}+\right.$$

[1]) Все числа таблицы умножены на 10 000 (кроме 1-го и 3-го столбцов).

$$+\frac{(1,2)^2}{1651,2}+\frac{(53,8)^2}{1936,8}+\frac{(8,8)^2}{1921,2}+\frac{(36)^2}{1602}+\frac{(38,8)^2}{1168,8}+\frac{(17)^2}{720}+$$
$$+\frac{(10,6)^2}{416,4}+\frac{(15,4)^2}{236,4}+\frac{(7,8)^2}{117,8}+\frac{(3)^2}{54}+\frac{(2,8)^2}{34,8}\Bigg]\neq 1,42,$$

значение довольно удовлетворительное.

6. Недостатком изложенного нами метода приближения статистических кривых является его непригодность для случаев, когда S и E более значительны, так как тогда необходимо повысить степень пертурбационного многочлена и для этого нужно прибегнуть к практически менее надежным моментам более высоких степеней. Однако если число n индивидов совокупности не очень велико, то распределение вероятностей является достаточно хорошо охарактеризованным своими первыми четырьмя моментами, т.-е. центром a, штандартом σ, асимметрией $S=\frac{1}{2}\frac{a_3}{\sigma^3}$ и эксцессом $E=\frac{1}{8}\left(\frac{a_4}{\sigma^4}-3\right)$, которые определяются, соответственно, из значений моментов 3-й и 4-й степени. Исходя из этого замечания, Пирсон ограничивается рассмотрением класса теоретических функций $y=f(x)$, удовлетворяющих дифференциальному уравнению

$$\frac{dy}{y\,dx}=\frac{a-x}{\beta+\gamma x+\delta x^2}, \tag{243}$$

зависящих только от четырех параметров a, β, γ, δ, которые вместе с коэффициентом пропорциональности определяются из 5 уравнений моментов ($k=4$); при этом существенным достоинством метода Пирсона является, что все фактически возможные значения S и E приводят к формально приемлемым функциям распределения, давая довольно хорошее приближение для статистических кривых, даже сильно уклоняющихся от кривой Гаусса.

Разумеется, кривые Пирсона [1]) могли бы быть заменены любым классом функций, имеющих те же моменты; таким образом

[1]) Уравнение Пирсона (243) является обобщением дифференциального уравнения, определяющего функцию Гаусса, для которой коэффициенты γ и δ равны 0. Однако лишь при малых значениях $\frac{\gamma}{\beta}$ и $\frac{\delta}{\beta}$ уравнение Пирсона может быть связано с соответствующей задачей теории вероятностей (см. формулу 175, стр. 239). Поэтому в настоящей книге, имеющей целью освещение принципиальных вопросов, я не имею возможности уделить больше места теории кривых Пирсона, представляющей в методологи-

вообще можно поставить себе задачу определить теоретическую функцию $f(x)$ данной формы, зависящую (кроме коэффициента пропорциональности) от k произвольных параметров, которые определяются из $k+1$ уравнений моментов

$$\int_{-\infty}^{\infty} f(x)\, x^h\, dx = a_h \quad (h=0, 1, \ldots, k). \tag{244}$$

Существенным теоретическим вопросом является при этом, возможно ли, располагая достаточным числом параметров, получить таким образом произвольно точное, с точки зрения теории вероятностей, изображение данной статистической функции распределения $\varphi(x)$; иными словами, возможно ли, удовлетворивши достаточно большому числу уравнений (244), сделать интеграл

$$H' = \int_{-\infty}^{\infty} \frac{[f(x)-\varphi(x)]^2}{f(x)}\, dx \tag{223}$$

произвольно малым.

Нетрудно видеть, что для этого во всяком случае необходимо, чтобы теоретическая функция $f(x)$ не обращалась в 0 в конечных промежутках, где $\varphi(x)$ отлична от 0, и, кроме того, H' всегда будет бесконечно возрастать, если функция $f(x)$ настолько быстро убывает по сравнению с $\varphi(x)$, что существует положительное число λ, для которого $\frac{\varphi^2(x)}{f(x)} > \lambda$ при достаточно больших x, так как

$$H' = \int_{-\infty}^{\infty} \left[f(x) - 2\varphi(x) + \frac{\varphi^2(x)}{f(x)} \right] dx = \int_{-\infty}^{\infty} \frac{\varphi^2(x)}{f(x)}\, dx - 1.$$

В этом случае интерполирование имело бы исключительно эмпирический характер.

ческом отношении не более как остроумный и прекрасно технически разработанный прием приближенного интерполирования, изложению которого должно быть уделено соответствующее место в специальном курсе теоретической статистики.

Допустим, напротив, что можно указать такое число A, что $\frac{\varphi(x)}{f(x)} < A$, и, кроме того, как теоретическая функция $f(x)$, так и действительная функция распределения $\varphi(x)$ убывает с возрастанием x не медленнее, чем в геометрической прогрессии, т.-е.

$$f(x) < Be^{-\alpha|x|}, \quad \varphi(x) < Be^{-\alpha|x|}, \tag{245}$$

где B и α — данные положительные числа; тогда, как бы мало ни было ε, можно выбрать k достаточно большим, чтобы осуществление уравнений

$$\int_{-\infty}^{\infty} f(x)\, x^h\, dx = a_h \quad (h = 0, 1, \ldots, k) \tag{244}$$

имело следствием, что

$$H' = \int_{-\infty}^{\infty} \frac{[f(x) - \varphi(x)]^2}{f(x)}\, dx < \varepsilon.$$

Действительно, из общих теорем о приближении функций вытекает что, каковы бы ни были данные положительные числа ρ и $\delta < \alpha$, можно всегда подыскать многочлен $P(x)$ достаточно высокой степени k, чтобы

$$\left| \frac{f(x) - \varphi(x)}{f(x)} - P(x) \right| < \rho e^{\delta|x|}. \tag{246}$$

Но если для всех значений $h \leqslant k$ удовлетворены уравнения моментов (244), то имеем также

$$\int_{-\infty}^{\infty} f(x)\, P(x)\, dx = \int_{-\infty}^{\infty} \varphi(x)\, P(x)\, dx,$$

т.-е.

$$\int_{-\infty}^{\infty} [f(x) - \varphi(x)]\, P(x)\, dx = 0.$$

Поэтому из (246) заключаем, что

$$\int_{-\infty}^{\infty} \frac{[f(x) - \varphi(x)]^2}{f(x)}\, dx < \int_{-\infty}^{\infty} |f(x) - \varphi(x)|\, \rho e^{\delta|x|}\, dx <$$

$$< 2B\rho \int_{-\infty}^{\infty} e^{(\delta - \alpha)|x|}\, dx = \frac{4B\rho}{\alpha - \delta}$$

стремится к 0 с возрастанием степени k многочлена $P(x)$. Из нашего рассуждения видно также, что при тех же условиях (245) функции $f(x)$ и $\varphi(x)$ окажутся тождественны для всех значений x, если все их моменты ($k=\infty$) будут равны.

Так как эмпирическую функцию $\varphi(x)$ можем всегда считать удовлетворяющей вышеуказанному условию (245), то, определяя теоретическую функцию распределения из достаточно большого числа уравнений моментов (244), всегда можно дать сколь угодно точное, с точки зрения теории вероятностей, приближение к любой статистической кривой, если только функция $f(x)$ не обращается в 0 в точках, где $\varphi(x)>0$ и удовлетворяет условию, что $f(x)<Be^{-\alpha|x|}$.

7. Способ моментов является наиболее распространенным для интерполирования статистических кривых. Нужно, однако, иметь в виду, что он налагает существенное ограничение на функции распределения, требуя, чтобы они обладали конечными моментами любой степени. Понятно, что a priori нельзя утверждать, что, сообразуясь с законами, лежащими в основе рассматриваемого явления, всегда будет целесообразно связывать теоретическую функцию таким ограничением. Кроме того и практически способ моментов может быть неудобен, если число промежутков, на которые разбит данный статистический материал, невелико; правда, в последнем случае функция распределения будет носить грубо интегральный характер.

Если весь материал разбит только на 3 класса, то всегда можно указать совершенно точно нормальное распределение, которое приводило бы как раз к наблюдаемой группировке.

Пусть $n_1+n_2+n_3=N$ будет общее число индивидов совокупности; из них у n_1 индивидов размер признака $x<x_1$, у n_1+n_2 индивидов $x<x_2$. В таком случае, при помощи таблицы функции

$\Phi(t)=\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{t^2}{2}}dt$, определяем t_1 и t_2 из уравнений

$$\frac{n_3}{N}=\frac{1}{\sqrt{2\pi}}\int_{t_1}^{\infty} e^{-\frac{t^2}{2}}dt, \quad \frac{n_2+n_3}{N}=\frac{1}{\sqrt{2\pi}}\int_{t_1}^{\infty} e^{-\frac{t^2}{2}}dt. \quad (247)$$

Тогда, полагая

$$\frac{x-a}{\sigma}=t, \tag{248}$$

из двух уравнений

$$x_1-a=\sigma t_1, \quad x_2-a=\sigma t_2$$

найдем штандарт

$$\sigma=\frac{x_2-x_1}{t_2-t_1} \tag{249}$$

и центр

$$a=x_1-\frac{t_1(x_2-x_1)}{t_2-t_1}=\frac{x_1t_2-x_2t_1}{t_2-t_1} \tag{250}$$

эквивалентного нормального распределения.

По данным [1]) о распределении браков в Ленинграде в первом полугодии 1920 года находим, что из 10 000 женихов [2]) 1 189 были не старше 22 лет (ранние браки), 7 020 были в возрасте от 23 до 37 лет включительно (средние браки) и 1 791 старше 37 лет (поздние браки).

Штандарт σ и центр a нормального распределения, соответствующего данной группировке, получаем из уравнений (247):

$$0{,}1791=\frac{1}{\sqrt{2\pi}}\int_{t_2}^{\infty}e^{-\frac{t^2}{2}}dt \quad \text{и} \quad 0{,}1189=\frac{1}{\sqrt{2\pi}}\int_{-t_1}^{\infty}e^{-\frac{t^2}{2}}dt,$$

откуда, пользуясь таблицей значений функции $\phi(t)$, находим:

$$t_1=-1{,}18, \qquad t_2=0{,}919;$$

поэтому из (249) и (250)

$$\sigma=\frac{15}{2{,}099}\neq 7{,}14, \quad a=\frac{23\cdot 0{,}919+38\cdot 1{,}18}{2{,}099}\neq 31{,}42.$$

Разумеется, при более дробном делении по группам мы убедились бы, что возраст лиц, вступающих в брак, не подчиняется закону Гаусса [3]) (и обнаружили бы положительную асимметрию распределения).

---

[1]) Материалы по статистике Ленинграда (3-й вып.) 1921 г.

[2]) В действительности зарегистрировано было 10 004 брака, но возраст 11 женихов неизвестен.

[3]) См. следующий §.

8. Но, каково бы ни было статистическое распределение, его можно охарактеризовать аналогичным образом, введя, на место постоянной дисперсии $\sigma^2 = \beta$, функцию дисперсии. Для этого поступаем следующим образом. Пусть h будет число интересующих нас или данных нам промежутков (x_i, x_{i+1}), которое, для определенности, положим равным четырем ($h = 4$), и пусть на каждый из них приходится, соответственно, n_1, n_2, n_3, n_4 индивидов ($n_1 + n_2 + n_3 + n_4 = N$), так что величине признака промежутка (x_i, x_{i+1}) соответствует n_i индивидов. Из уравнений

$$\frac{n_4}{N} = \frac{1}{\sqrt{2\pi}} \int_{t_3}^{\infty} e^{-\frac{t^2}{2}} dt, \quad \frac{n_4 + n_3}{N} = \frac{1}{\sqrt{2\pi}} \int_{t_2}^{\infty} e^{-\frac{t^2}{2}} dt \text{ и т. д.} \tag{247 bis}$$

определяем последовательно значения t_3, t_2, t_1; поэтому, если положим

$$t = \psi(x), \tag{251}$$

где $\psi(x)$ есть возрастающая (от $-\infty$ до $+\infty$) функция величины x, которую подчиним лишь условиям, чтобы

$$t_i = \psi(x_i), \quad (i = 1, 2, \ldots, h-1), \tag{252}$$

то вероятность каждого неравенства вида

$$x_l < x < x_k$$

будет точно выражена интегралом

$$\frac{1}{\sqrt{2\pi}} \int_{\psi(x_l)}^{\psi(x_k)} e^{-\frac{t^2}{2}} dt.$$

Очевидно, что если бы $h-1$ уравнениям (252) удовлетворяла приближенно линейная функция, то данное распределение (в пределах рассматриваемой группировки) было бы более или менее близко к нормальному, и мы имели бы приближенно

$$t_i = \frac{x_i - a}{\sigma}.$$

В общем случае мы можем тоже положить

$$t = \frac{x - a}{\sqrt{B(x)}}, \tag{251 bis}$$

но $\sqrt{B(x)}$ уже не будет постоянной величиной σ, а будет некоторой функцией переменной x, которая должна удовлетворять $h-1$ уравнениям

$$t_i = \frac{x_i - a}{\sqrt{B(x_i)}}. \tag{253}$$

Ясно, что при $x=a$ уравнение (251 bis) дает $t=0$; поэтому a является медианой распределения вероятностей. Что же касается функции $\sqrt{B(x)}$, то для каждого значения x_i она равна штандарту той нормальной кривой, имеющей центром a, для которой вероятность, что $x > x_i$ (а также $x < x_i$) та же самая, что и для данной кривой: иначе говоря, независимо от всех прочих элементов распределения, зная $B(x_i)$ для определенного x_i и медиану a, можно утверждать, что вероятность неравенства $x > x_i$ равна

$$\frac{1}{\sqrt{2\pi}} \int_{\frac{x_i - a}{\sqrt{B(x_i)}}}^{\infty} e^{-\frac{t^2}{2}} dt. \tag{254}$$

Функцию $B(x)$ назовем дисперсионной функцией или функцией дисперсии, а $\sqrt{B(x)}$ — штандартной функцией данного распределения.

Предположим, что для двух значений $x_1 = a + y$ и $x_2 = a - y$, где $y > 0$, $B(x_1) > B(x_2)$; это значит, что наше распределение асимметрично, при чем положительные уклонения, превышающие $|y|$, более вероятны, чем отрицательные, и, следовательно, наоборот, меньшие положительные уклонения менее вероятны, чем соответствующие отрицательные уклонения. В частности, если

$$B(x) = l + qx, \tag{255}$$

то это обстоятельство, характеризующее положительную или правую асимметрию, представится, когда $q > 0$ (при этом коэффициент S асимметрии положителен); если же $q < 0$, то асимметрия отрицательна. В первом случае вблизи медианы кривая опускается слева направо, а во втором — поднимается.

Предположим, напротив, что распределение симметрично (по отношению к медиане, которая тогда, очевидно, совпадает

с центром); в таком случае, если с возрастанием уклонения $|x-a|$ дисперсионная функция $B(x)$ убывает, то вблизи центра (при той же центральной ординате) распределение гуще нормального, и, напротив, если $B(x)$ возрастает, то распределение вблизи центра реже нормального: первый случай соответствует, вообще, отрицательному эксцессу, второй — положительному.

9. Если, по тем или иным теоретическим соображениям, дисперсионная функция $B(x)=B(y+a)$ представляется в виде

$$B(x)=l+q_1u_1(x)+\ldots+q_ku_k(x), \tag{256}$$

где $u_1(x),\ldots,u_k(x)$ данные функции, то для определения $k+2$ величин: $a, l, q_1,\ldots, q_k$ имеем нужное число уравнений, если $k=h-3$. Если $k<h-3$, то уравнениям (253) можно удовлетворить только приближенно; но мы ограничимся лишь рассмотрением случая, когда $k=h-3$; задача приводится тогда к решению квадратного уравнения относительно a.

Для определенности положим $k=1$ (т.-е. $h=4$), $u_1(x)=x$; уравнения (252 bis) по возведении в квадрат дают

$$t_1^2(l+qx_1)=(x_1-a)^2,\quad t_2^2(l+qx_2)=(x_2-a)^2,$$
$$t_3^2(l+qx_3)=(x_3-a)^2,$$

и, исключая из них l и q, получаем для определения a квадратное уравнение

$$\begin{vmatrix} t_1^2 & t_1^2x_1 & (x_1-a)^2 \\ t_2^2 & t_2^2x_2 & (x_2-a)^2 \\ t_3^2 & t_3^2x_3 & (x_3-a)^2 \end{vmatrix}=0,$$

или

$$\frac{(x_1-a)^2}{t_1^2}(x_3-x_2)+\frac{(x_2-a)^2}{t_2^2}(x_1-x_3)+$$
$$+\frac{(x_3-a)^2}{t_3^2}(x_2-x_1)=0, \tag{257}$$

которое, как нетрудно проверить, всегда имеет[1]) вещественные корни, когда $|t_2|<|t_1|$, $|t_2|<|t_3|$. После подстановки[2]) a в уравнения (253) (возведенные в квадрат) получаем для определения остальных параметров систему линейных уравнений, решение которой не представляет труда: в рассматриваемом частном случае из уравнений

$$t_1^2(l+qx_1)=(x_1-a)^2,\quad t_3^2(l+qx_3)=(x_3-a)^2,$$

которые записываем в виде

$$t_1^2(\beta+qy_1)=y_1^2,\quad t_3^2(\beta+qy_3)=y_3^2,$$

получаем

$$q=\frac{\dfrac{y_3^2}{t_3^2}-\dfrac{y_1^2}{t_1^2}}{y_3-y_1},\quad \beta=y_1y_3\frac{\dfrac{y_1}{t_1^2}-\dfrac{y_3}{t_3^2}}{y_3-y_1},\tag{258}$$

где

$$y_1=x_1-a,\quad y_3=x_3-a,\quad \beta=l+qa.$$

Введем в вышерассмотренном примере с браками еще один промежуток, а именно: разделим основной интервал от 23 до 37 лет включительно на два промежутка: на группу женихов моложе 30 лет и на группу не моложе этого возраста. Таким образом имеем 4 промежутка: $n_1=1\,189$ $(x<23)$, $n_2=4\,382$ $(23\leqslant x<30)$, $n_3=2\,638$ $(30\leqslant x<38)$, $n_4=1\,791$ $(x\geqslant 38)$. В таком случае

$$t_1=-1{,}18,\quad t_2=0{,}144,\quad t_3=0{,}919,$$

и уравнение (257) обращается в

$$\frac{8(23-a)^2}{1{,}39}-\frac{15(30-a)^2}{0{,}0207}+\frac{7(38-a)^2}{0{,}8445}=0,$$

---

[1]) Вообще условием необходимым и достаточным для вещественности корней (257) является неравенство

$$t_1^2(x_3-x_2)+t_2^2(x_1-x_3)+t_3^2(x_2-x_1)>0.$$

[2]) Из двух корней a_1 и a_2 уравнения (257), которые удовлетворяют неравенствам: $a_1<x_2$ и $a_2>x_2$, корень a_1 соответствует $t_2>0$, корень же a_2 соответствует $t_2<0$.

или, полагая $30 - a = z$,

$$710{,}7\,z^2 - 52{,}1\,z - 812{,}6 = 0.$$

Принимая во внимание, что $a < 30 = x_2$, так как $t_2 > 0$, заключаем, что искомым решением будет положительный корень $z \neq 1{,}1$, откуда медиана $a = 28{,}9$, и из формул (258) находим

$$q \neq 5, \qquad \beta \neq 48.$$

Таким образом дисперсионная функция

$$B(x) \neq 48 + 5\,(x - a),$$

и мы имеем действительно значительную положительную асимметрию, при чем полученная функция $t = \frac{x-a}{\sqrt{B(x)}}$, распределение которой подчиняется нормированному закону Гаусса, дает и для более дробного деления на промежутки довольно точное приближение рассматриваемой статистической кривой [1]).

9. До сих пор мы рассматривали задачу интреполирования статистической кривой как чисто математическую задачу возможно точной и полной характеристики данной совокупности при помощи той или иной функции или коэффициентов, определяющих эту функцию. Этого совершенно достаточно, если мы преследуем узко практическую цель арифметического описания данной конкретной статистической совокупности.

Но если мы хотим видеть в данной совокупности произвольного представителя (случайную выборку) из некоторой определенной основной весьма обширной или теоретически неограниченной совокупности, то статистические вероятности и средние величины приобретают характер приближенных значений соответствующих им реальных величин, и особую важность получает вторая указанная выше задача построения гипотетической схемы рассматриваемой

---

[1])

x	<20	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	≥40	Всего
n	188	245	277	478	586	614	674	627	657	615	605	500	372	439	317	268	318	224	199	229	157	1404	9993

Внутри ряда здесь наблюдаются некоторые неправильности (например, относительный максимум для 32 лет), анализ которых не входит в нашу задачу и которые, конечно, не могут быть обнаружены нашей интегральной теоретической кривой.

совокупности, основанной на свойствах рассматриваемого явления. Эта задача удовлетворительно разрешена во многих вопросах теоретической физики; для других же явлений, за исключением случаев, где применим закон Гаусса, отметим, в частности, схему (рассмотренную нами в главе I второй части, стр. 61), которая приводит к формуле распределения

$$f(x)=\frac{e^{-k}k^x}{x!}, \tag{47 bis}$$

а также функцию

$$f(x)=\lambda e^{-\lambda x},$$

где $x>0$, при которой интегральная функция распределения вероятностей равна $1-e^{-\lambda x}$, т.-е. вероятность, что $x>x_0$, равна $\varphi(x)=e^{-\lambda x}$. Последняя функция соответствует формуле (118, стр. 173) вероятности разорения игрока, обладающего капиталом x, и применима в тех случаях, когда $\varphi(x+y)=\varphi(x)\cdot\varphi(y)$, т.-е. тогда, когда для достижения признаком величины $x+y$ ему нужно сначала получить значение x и дальнейшее увеличение его на величину y не зависит от достигнутого уже значения x.

Укажем еще одну общую теоретическую схему [1]), которая приведет нас к формуле (254), и, в частности, к случаям, когда функция дисперсии $B(x)$ линейна или представляет многочлен 2-й степени. Для большей ясности рассмотрим конкретный пример.

Положим, что исследуется количество z некоторых определенных изделий, вырабатываемых в день рабочим данного производства; мы допускаем, что количество энергии (выраженное в каких угодно единицах), которое он может израсходовать в течение дня, равное $\sum$, подчиняется закону Гаусса, с центром A (М. О. $\sum=A$) и дисперсией $B=$ М. О. $(\sum-A)^2$; допустим, что в процессе производства энергия рабочего убывает так, что энергия его $\sum_z$, после того как он произвел определенное количество z изделий, также подчиняется закону Гаусса, но М. О. $\sum_z=A_z$, и значение дисперсии $B(z)=$ М. О. $\left[\sum_z-A_z\right]^2$ также некоторым образом изменилось, представляя ту или иную функцию $B(z)$ величины z; рабочий прекращает работу, когда им израсходована вся его рабочая энергия данного дня $\sum$, т.-е. когда $\sum_z=0$. Из указанных предпосылок нетрудно вывести закон распределения вероятностей количества z продукции, выработанной рабочим в день.

---

[1]) См. мою статью „Sur les courbes de distribution des probabilités". Mathem. Zeitschrift. 1925.

Действительно, пусть P_1 есть вероятность, что $\sum_z$ остается положительным, пока $z < z_1$. В таком случае, если $z_1 < z_2$ и P_2 есть вероятность, что $\sum_z$ продолжает оставаться положительными при $z < z_2$, то

$$P_1 = P + P_2, \tag{259}$$

где P есть вероятность, что $\sum_z$ перестает быть положительным, т.-е. обращается в нуль для значения z

$$z_1 < z < z_2$$

(так как убывающая величина $\sum_z$ может оставаться положительной при $z < z_1$, либо обратившись затем в нуль в промежутке $z_1 > z > z_2$, либо продолжая быть положительной и в этом промежутке).

Но P_1 есть вероятность, что при данном z_1 $\sum_{z_1} > 0$, или, что то же самое, вероятность неравенства

$$\sum_z - Az_1 > -Az_1 = t_1 \sqrt{B(z_1)},$$

где

$$t_1 = \frac{-Az_1}{\sqrt{B(z_1)}};$$

поэтому

$$P_1 = \frac{1}{\sqrt{2\pi}} \int_{t_1}^{\infty} e^{-\frac{t^2}{2}} dt$$

и, по тем же основаниям,

$$P_2 = \frac{1}{\sqrt{2\pi}} \int_{t_2}^{\infty} e^{-\frac{t^2}{2}} dt, \quad \text{где } t_2 = \frac{-Az_2}{\sqrt{B(z_2)}}.$$

Следовательно, вследствие (259),

$$P = \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-\frac{t^2}{2}} dt,$$

т.-е., иными словами, функция

$$t = \frac{-A_z}{\sqrt{B(z)}} \tag{251 ter}$$

подчиняется нормированному закону Гаусса.

Делая довольно общие допущения, можно притти к выводу, что A_z и $B(z)$ должны быть линейными функциями. Для упрощения сделаем следующие более ограничительные допущения, чем это необходимо. А именно, допустим, что математическое ожидание количества энергии φ_x, которую рабочий должен затратить, чтобы произвести количество x изделий, равно λx, где λ постоянная величина, независимо от того, какой энергией он обладает; таким образом М. О. $\varphi_z = \lambda z$, а потому

$$A_z = \text{М. О.} \sum_z = \text{М. О.} \left[\sum - \varphi_z\right] = A - \lambda z; \tag{260}$$

кроме того, обозначая через b дисперсию количества φ_1 энергии, затраченной на единицу продукции, т.-е. полагая $b = \text{М. О. } [\varphi_1 - \lambda]^2$, имеем, в силу допущенной выше независимости,

$$\text{М. О. } [\varphi_2 - 2\lambda]^2 = \text{М. О. } (\varphi_1 - \lambda)^2 + \text{М. О. } (\varphi_1 - \lambda)^2 = 2b$$

и вообще

$$\text{М. О. } [\varphi_z - \lambda z]^2 = bz.$$

Поэтому

$$B(z) = \text{М. О. } \left[\sum_z - A_z\right]^2 = \text{М. О. } \left[\sum - A - (\varphi_z - \lambda z)\right]^2 = B + bz. \quad (261)$$

Итак, при указанных предположениях формула (251 ter) получает вид

$$t = \frac{\lambda z - A}{\sqrt{B + bz}}. \quad (262)$$

В частности, медиана $a = \frac{A}{\lambda}$, и, обозначая через y уклонение $z - a = y$ от медианы, можем написать

$$t = \frac{y}{\sqrt{\beta + qy}}, \text{ где } q = \frac{b}{\lambda^2}, \ \beta = \frac{B + ab}{\lambda^2} = \frac{B}{\lambda^2} + \frac{Ab}{\lambda^3}. \quad (262 \text{ bis})$$

Следует заметить, что, при делаемых здесь допущениях, $q = \frac{b}{\lambda^2}$ должно быть мало $\left(\text{примерно не более } \frac{1}{10}\right)$ для того, чтобы предположенная нами независимость последовательных затрат энергии исключала практически возможность отрицательной энергии, когда z увеличивается на 1. Вообще, применяя выведенную формулу, нужно ожидать лучшей ее пригодности для больших интервалов, чем для очень малых, где иногда возможны противоположные колебания энергии.

Аналогичный метод рассуждений применим к различным биологическим и экономическим явлениям (z может быть, например, также числом плодов дерева данной породы или продолжительностью амортизационного срока машин данного образца и т. п.); при этом во многих случаях можно допустить, что функция дисперсии $B(z)$ есть многочлен 1-й или 2-й степени:

$$B(z) = B(y + a) = \beta + q_1 y + q_2 y^2. \quad (263)$$

Укажем одно общее свойство этих кривых, которое позволяет часто без особых вычислений проверить, пригодны ли они для приближенного изображения данной статистической кривой.

Назовем две величины y_0 и $-y'_0$ соответственными значениями при данном распределении, если вероятность, что $0 < y < y_0$ равна вероятности, что $0 > y > -y'_0$; эти значения соответствуют значениям t_0 и $-t_0$ нормальной кривой.

Поэтому имеем равенство

$$\frac{y_0}{\sqrt{\beta + q_1 y_0 + q_2 y_0^2}} = \frac{y'_0}{\sqrt{\beta - q_1 y'_0 + q_2 y'^2_0}},$$

из которого (после возведения в квадрат) нетрудно вывести соотношение

$$\frac{1}{y'_0} - \frac{1}{y_0} = \frac{q_1}{\beta}, \tag{264}$$

характеризующее природу асимметрии рассматриваемых кривых распределения. Напомню, что y представляет здесь уклонение от медианы.

Не останавливаясь на подробном исследовании рассматриваемых кривых распределения, отметим, что $q_2 < 0$ соответствует распределению вероятностей на ограниченном с обеих сторон отрезке; $q_2 = 0$ приводит к случаю, когда область изменения y ограничена лишь с одной стороны (если $q_1 \lesseqgtr 0$); наконец, если $q_2 > 0$, то кривая теоретически неограниченна [1]).

Наиболее простым является случай, когда $q_2 = 0$, т.-е. когда функция

$$B(z) = B(y + a) = \beta + q_1 y \tag{263 bis}$$

линейна; все неизвестные элементы a, β, q легко определяются тогда по способу моментов при помощи лишь первых трех моментов.

Полагая $a_1 =$ М. О. z, $a_2 =$ М. О. z^2, $a_3 =$ М. О. z^3, $\sigma^2 =$ М. О. $(z - a_1)^2 = a_2 - a_1^2$, $\gamma^3 =$ М. О. $(z - a_1)^3 = a_3 - 3a_1 a_2 + 2a_1^3$ и определяя эти величины обычным статистическим способом как соответствующие средние, находим посредством несложных вычислений

$$a_1 = a + \frac{1}{2} q_1 \tag{265}$$

---

[1]) Последний случай представляется, когда $z = \frac{u}{v}$, где u и v независимы или связаны нормальной корреляцией, при чем отрицательные значения z практически невозможны; при этом, полагая М. О. $u = A > 0$, М. О. $(u - A)^2 = B$, М. О. $v = A_1$, М. О. $(v - A_1)^2 = B_1$ и обозначая через k коэффициент корреляции между u, v, находим, что медиана равна $\frac{A}{A_1} = a$, и $\beta = a^2(\sigma^2 + \sigma_1^2 - 2k\sigma\sigma_1)$, $q_1 = 2a\sigma_1(\sigma_1 - k\sigma)$, $q_2 = \sigma_1^2$, где величины $\sigma = \frac{\sqrt{B}}{A}$ и $\sigma_1 = \frac{\sqrt{B_1}}{A_1}$ должны быть малы (менее 1/3).

Своеобразие случая $q_2 > 0$ состоит в том, что М. О. z теоретически не имеет смысла, и поэтому, строго говоря, метод моментов к нему неприменим; но когда $|q_2|$ мало, можно применить соответствующий приближенный способ вычислений, на котором мы здесь останавливаться не будем.

(таким образом q_1 есть удвоенное расстояние между центром и медианой, так что имеем приближенно $q_1 \neq \frac{2}{3} S\sigma$, где S есть асимметрия) и затем

$$\sigma^2 = \beta + \frac{5}{4} q_1^2, \quad \gamma^3 = 3q_1 \beta + \frac{11}{2} q_1^3; \tag{266}$$

исключая из этих двух уравнений β, находим для определения q_1 уравнение 3-й степени

$$7q_1^3 + 12\sigma^2 q_1 - 4\gamma^3 = 0. \tag{267}$$

После определения q_1 значения a и β получаются немедленно из уравнения (265) и первого из уравнений (266).

Применим указанный прием к примеру, статистические данные которого я заимствую из книги проф. Лахтина „Теория кривых распределения". В окрестностях Москвы было сорвано 1 158 ландышевых стеблей и подсчитано число z цветков на каждом из них; следующая таблица дает результаты подсчета (2-й столбец) и теоретические числа, соответствующие нормальной кривой (3-й столбец) и кривой с линейной дисперсионной функцией (4-й столбец).

Число цветков на стебле	Фактическое число стеблей		Теоретич. число нормальной кривой t		Теоретич. число по формуле (269)	
3	3	49		69	1,1	50
4	46				48,9	
5	270		236		274,6	
6	430		406		422	
7	280		318		278	
8	96	129		129	102,5	133,4
9	27				25,5	
10	5				4,6	
11	1				0,8	

Средняя $a_1 = 6{,}18$; $a_2 - a_1^2 = 1{,}174$; $\sigma = \sqrt{1{,}174} \neq 1{,}08$. Нормальная кривая приводит тогда к числам 3-го столбца, где мы объединили крайние интервалы, чтобы в каждом было теоретически не менее $\frac{1}{10}$ всех индивидов; тогда коэффициент точности

$$H = \frac{1}{4}\left[\frac{20^2}{69} + \frac{34^2}{236} + \frac{24^2}{406} + \frac{38^2}{318} + 0\right] \neq \frac{1}{4}\left[5{,}8 + 4{,}9 + 1{,}4 + 4{,}5\right] = 4{,}15.$$

обнаруживает **неудовлетворительность** этой кривой. Напротив, вторая кривая при любых интервалах дает для H малое значение и, в частности, при той же группировке

$$H = \frac{4}{1}\left[\frac{1}{50} + \frac{4{,}6^2}{276{,}6} + \frac{8^2}{422} + \frac{2^2}{278} + \frac{4{,}4^2}{133{,}4}\right] \neq 0{,}1. \tag{268}$$

Кривая, соответствующая последнему столбцу, определяется из соотношения

$$t = \frac{z - 6{,}092}{\sqrt{0{,}083 + 0{,}173\,z}} = \frac{y}{\sqrt{1{,}137 + 0{,}173\,y}}, \tag{269}$$

где величина t подчиняется нормированному закону Гаусса, так что вероятность того, например, что число цветков равно 8, т.-е., что $7{,}5 < z < 8{,}5$, получаем из таблицы функции $\Phi(t)$, беря разность $\Phi(t_1) - \Phi(t_0)$, где

$$t_1 = \frac{8{,}5 - 6{,}092}{\sqrt{0{,}083 + 0{,}173 \cdot 8{,}5}} = 1{,}932,$$

$$t_0 = \frac{7{,}5 - 6{,}092}{\sqrt{0{,}083 + 0{,}173 \cdot 7{,}5}} = 1{,}198.$$

Для получения численных значений коэффициентов формулы (269) сначала вычисляем $\gamma^3 = 0{,}6179$, а затем из уравнения

$$7q_1^3 + 14{,}088\,q_1 - 2{,}4716 = 0, \tag{267 bis}$$

находим (с точностью до 0,001) $q_1 = 0{,}173$, откуда медиана $a = a_1 - \frac{1}{2}q_1 = 6{,}092$ и, по (266), $\beta = 1{,}1366$.

Упражнение. Показать, что кривые Пирсона, соответствующие $\delta = 0$ в уравнении (243), для нормированного распределения ($a_1 = 0$, $\sigma = 1$) имеют вид $f(x) = y_0\, e^{-\frac{x}{\gamma}}(1 + \gamma x)^{\frac{1}{\gamma^2} - 1}$, где $\gamma = S = \frac{a_3}{2}$ есть асимметрия кривой, а y_0 — постоянная, определенная условием, чтобы $\int\limits_{-\frac{1}{\gamma}}^{\infty} f(x)\,dx = 1$.

Отв. Из уравнения (243) выводим

$$(\beta + \gamma x + \delta x^2)\,dy = y\,(a - x)\,dx; \tag{243 bis}$$

интегрируя обе части равенства (первую — по частям) между пределами, где y обращается в нуль, находим:

$$-\gamma - 2\delta a_1 = a - a_1,$$

откуда, при $a_1 = 0$, $-\gamma = a$. Умножая (243 bis) на x и x^2 и интегрируя, соответственно, между теми же пределами, получим (при $\delta = 0$, $\sigma = 1$)

$$\beta = 1 \text{ и } 2\gamma = a_3.$$

Следовательно, получаем дифференциальное уравнение

$$\frac{dy}{y} = -\frac{\gamma + x}{1 + \gamma x}\,dx,$$

интегрирование которого даст указанный результат.

ГЛАВА ТРЕТЬЯ.

ЭЛЕМЕНТЫ ТЕОРИИ КОРРЕЛЯЦИИ.

1. Если нам дана некоторая статистическая совокупность, состоящая из n объектов, которую мы исследуем одновременно относительно двух признаков x, y, то во всех случаях, когда рассматриваемая совокупность является произвольным представителем некоторой определенной, теоретически весьма обширной или неограниченной совокупности, мы можем говорить о математической вероятности того, что у данного объекта признак x, а также признак y получает определенное значение. При этом, если один или оба признака изменяются непрерывно, то для указанного признака следует, как мы видели, рассматривать вероятности неравенств вида $x_0 < x < x_1$, вместо точных равенств $x = x_0$. Каждый из признаков, рассматриваемый самостоятельно, имеет свой закон распределения вероятностей и, вообще, вероятность неравенства

$$x_0 < x < x_1 \tag{270}$$

равна $F(x_1) - F(x_0)$, где $F(x)$ есть интегральная функция распределения вероятностей величины x, и точно так же вероятность неравенства

$$y_0 < y < y_1 \tag{271}$$

выражается соответствующей разностью $F_1(y_1) - F_1(y_0)$.

Если оказывается, что, каковы бы ни были значения одной из величин, вероятности указанных неравенств для другой величины не меняют своего значения, то величины называются **независимыми**. В противном случае, величины x, y называются **зависимыми** или **коррелятивно связанными**.

Для полной характеристики коррелятивной зависимости между x и y необходимо и достаточно знать, с одной стороны, какую вероятность получает каждое неравенство (270), когда известно значение y, и, с другой стороны, какую вероятность приобретает (271), когда дано x.

Предположим, что, разбивая значения y на небольшое (два для определенности) число классов, мы констатируем, что все объекты, принадлежащие к каждому классу, подчиняются закону Гаусса по отношению к признаку x, при чем для одного класса центр и штандарт равны a и σ, а для другого — равны a_1 и σ_1. В таком случае, если p и q $(p+q=1)$ представляют вероятности, что рассматриваемые объекты нашей совокупности принадлежат, соответственно, к первому и второму классу, то функция распределения величины x будет иметь вид:

$$\frac{p}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}+\frac{q}{\sigma_1\sqrt{2\pi}}e^{-\frac{(x-a_1)^2}{2\sigma_1^2}}. \qquad (272)$$

Таким образом признак x для всей совокупности не будет подчиняться закону Гаусса. Это обстоятельство представится, например, если, исследуя размеры индивидов какой-нибудь чистой биологической расы, мы объединим в одну совокупность особи мужского и женского пола или же соединим вместе в определенной пропорции индивидов двух различных рас. Однако, если число рассматриваемых классов, внутри которых признак x подчиняется закону Гаусса, весьма велико, то не исключена возможность, что и для всей совокупности, охватывающей все указанные классы, закон Гаусса также будет соблюден. Этот именно случай имеет место, если величины x и y связаны так называемой нормальной корреляцией, или коррелятивной зависимостью.

Если непрерывные величины x и y обладают свойством, что в классе индивидов, для которых y находится в определенном весьма малом интервале dy, признак x подчиняется закону Гаусса, и наоборот, когда x находится в весьма малом промежутке dx, признак y также подчиняется закону Гаусса, то корреляция между x и y называется нормальной (при дополнительном теоретическом ограничении, что дифференциальные функции распределения дважды дифференцируемы и, кроме того, что при неограниченном возрастании y величина x не стремится к некоторому вполне определенному значению $x=x_0$).

A priori не очевидно, что такая зависимость между величинами x и y вообще возможна, ибо, как мы сейчас увидим, функция

$$F_1(x, y)\,dy,$$

выражающая вероятность, что y находится в промежутке $(y, y+dy)$, когда известно, что x имеет значение, заключенное в промежутке $(x, x+dx)$, связана определенным образом с функцией

$$\Phi_1(x, y)\,dx,$$

представляющей вероятность, что x находится в промежутке $(x, x+dx)$, когда дано, что y содержится в промежутке $(y, y+dy)$.

Действительно, если $f(x)\,dx$ представляет вероятность a priori (т.-е. для индивидов всей совокупности), что x находится в промежутке $(x, x+dx)$, и, с другой стороны, $\varphi(y)\,dy$ есть вероятность a priori, что y находится в промежутке $(y, y+dy)$, то, по теореме умножения вероятностей, одновременное нахождение x и y в указанных промежутках имеет вероятность

$$P(x, y)\,dx\,dy = f(x)\cdot F_1(x, y)\,dx\,dy = \varphi(y)\,\Phi_1(x, y)\,dx\,dy. \quad (273)$$

Поэтому должно быть соблюдено тождество

$$f(x)\cdot F_1(x, y) = \varphi(y)\cdot\Phi_1(x, y), \quad (274)$$

откуда, полагая

$$F(x, y) = \log F_1(x, y), \quad \Phi(x, y) = \log \Phi_1(x, y) \quad (275)$$

и логарифмируя (274), находим:

$$F(x, y) - \Phi(x, y) = \log\varphi(y) - \log f(x); \quad (276)$$

дифференцируя последнее равенство по x и по y, получаем:

$$\frac{\partial^2 F(x, y)}{\partial x\,\partial y} = \frac{\partial^2 \Phi(x, y)}{\partial x\,\partial y}. \quad (277)$$

В данном случае функции $F_1(x, y)$ и $\Phi_1(x, y)$ по условию должны иметь вид:

$$F_1(x, y) = \sqrt{\frac{k_1(x)}{2\pi}}\,e^{-\frac{[y-a_1(x)]^2}{2}k_1(x)},$$

$$\Phi_1(x, y) = \sqrt{\frac{k(y)}{2\pi}}\,e^{-\frac{[x-a(y)]^2}{2}k(y)}, \quad (278)$$

так как индивиды промежутка $(x, x+dx)$ должны подчиняться закону Гаусса относительно признака y, при чем параметры этого распределения могут быть функциями x[1]), и то же самое mutatis mutandis относится к признаку x.

Следовательно, применяя уравнение (277) и дифференцируя дважды $\log F_1(x, y)$ и $\log \Phi_1(x, y)$ (по x и по y), мы находим тождество

$$[(y-a_1(x))^2 k_1(x)]''_{xy} = [(x-a(y))^2 k(y)]''_{xy}, \tag{279}$$

которому должны удовлетворять неизвестные функции: $a_1(x)$, $k_1(x)$, $a(y)$, $k(y)$. Полагая,

$$a(y)\cdot k(y) = b(y), \quad a_1(x)\cdot k_1(x) = b_1(x),$$

можем придать тождеству (279) вид

$$yk_1'(x) - b_1'(x) = xk'(y) - b'(y). \tag{280}$$

Поэтому, дифференцируя еще по y и по x последнее тождество, находим:

$$k_1''(x) = k''(y) = 2c, \tag{281}$$

где c есть произвольная постоянная величина (так как функции двух различных переменных могут быть тождественны только, если они обращаются в постоянную величину). Отсюда заключаем, что

$$k_1(x) = cx^2 - 2d_1x + h_1, \qquad k(y) = cy^2 - 2dy + h, \tag{282}$$

где d, d_1, h, h_1 — новые произвольные постоянные коэффициенты. Подставляя (282) в (280), получаем:

$$2y(cx - d_1) - b_1'(x) = 2x(cy - d) - b'(y),$$

откуда

$$-2d_1y + b'(y) = -2dx + b_1'(x) = l, \tag{283}$$

где l — произвольная постоянная величина. Следовательно,

$$b'(y) = 2d_1y + l, \qquad b_1'(x) = 2dx + l;$$

---

[1]) Если бы эти параметры не зависели от x, то перед нами был бы случай двух независимых величин.

и наконец, интегрируя, находим:

$$b(y)=d_1y^2+ly+m_1, \quad b_1(x)=dx^2+lx+m, \tag{284}$$

где m и m_1 — новые постоянные, откуда

$$a_1(x)=\frac{b_1(x)}{k_1(x)}=\frac{dx^2+lx+m}{cx^2-2d_1x+h_1}, \quad a(y)=\frac{d_1y^2+ly+m_1}{cy^2-2dy+h}. \tag{285}$$

Таким образом, подставляя (282) и (285) в выражения (278), находим, что необходимая форма функций $F_1(x, y)$, $\Phi_1(x, y)$ такова:

$$F_1(x,y)=\sqrt{\frac{cx^2-2d_1x+h_1}{2\pi}}\,e^{-\frac{[y(cx^2-2d_1x+h_1)-(dx^2+lx+m)]^2}{2(cx^2-2d_1x+h_1)}},$$

$$\Phi_1(x,y)=\sqrt{\frac{cy^2-2dy+h}{2\pi}}\,e^{-\frac{[x(cy^2-2dy+h)-(d_1y^2+ly+m_1)]^2}{2(cy^2-2dy+h)}}. \tag{286}$$

Первая из формул (285) дает значение центра распределения y при данном x, а вторая — центр распределения x при данном y; штандарты, соответственно равные $\frac{1}{\sqrt{k_1(x)}}$ и $\frac{1}{\sqrt{k(y)}}$, определяются из формул (282). Поэтому, если $c \gtreqless 0$, то $c>0$, так как иначе штандарты становились бы мнимыми; поэтому при бесконечном возрастании x штандарт y стремился бы к 0 (и наоборот); это означает, что при $x \to \infty$ вероятность даже самого малого уклонения y от $a_1(\infty)=\frac{d}{c}$ становилась бы невозможной. Исключая этот случай, повидимому, не имеющий практического применения, мы полагаем $c=0$, и тогда для того, чтобы штандарт не становился мнимым, необходимо также, чтобы $d=d_1=0$ (при этом, конечно, $h>0$, $h_1>0$).

2. Таким образом формулы (286) приобретают окончательный вид:

$$F_1(x,y)=\sqrt{\frac{h_1}{2\pi}}\,e^{-\frac{h_1}{2}\left(y-\frac{lx+m}{h_1}\right)^2},$$

$$\Phi_1(x,y)=\sqrt{\frac{h}{2\pi}}\,e^{-\frac{h}{2}\left(x-\frac{ly+m_1}{h}\right)^2}, \tag{287}$$

и приводят нас к следующим выводам, которые удобнее будет формулировать, если мы будем представлять себе величины x и y как прямоугольные координаты точки.

1) При каждом значении x $(x, x+dx)$ центр a_1, соответствующий распределению вероятностей величины y, определяется из уравнения

$$Y=\frac{lx+m}{h_1}, \tag{288}$$

которое геометрически представляется в виде прямой линии. Эта прямая, каждая точка которой имеет ординатой М. О. $(y)_x = a_1(x) = \frac{lx+m}{h_1}$ при данном значении x, называется линией регрессии величины y по x. Угловой коэффициент

$$\frac{l}{h_1}=\rho_1 \tag{289}$$

называется коэффициентом регрессии y по x.

Аналогичным образом, прямая линия

$$X=\frac{ly+m_1}{h}, \tag{288 bis}$$

все точки которой, при данном y, имеют абсциссу, равную М. О. $(x)_y = a(y) = \frac{ly+m_1}{h_1}$, соответствующую центру нормального распределения, когда дано значение y $(y, y+dy)$, называется линией регрессии x по y; а коэффициент

$$\frac{l}{h}=\rho \tag{289 bis}$$

называется коэффициентом регрессии x по y.

2) Каково бы ни было x, штандарт нормального распределения величины y сохраняет постоянное значение $\frac{1}{\sqrt{h_1}}=\sigma_1^*$; точно так же, каково бы ни было

данное значение величины y, штандарт нормального распределения величины x сохраняет одно и то же значение $\frac{1}{\sqrt{h}}=\sigma^*$.

3) Выбирая начало координат таким образом, чтобы, при $x=0$, $a_1(x)=0$, т.-е. чтобы линия регрессии проходила через начало координат, можем, не нарушая общности, положить $m=0$, а выбирая за начало координат точку пересечения обеих линий регрессии, получим $m=m_1=0$ и приведем функции $F_1(x, y)$ и $\Phi_1(x, y)$ к виду

$$F_1(x, y)=\sqrt{\frac{h_1}{2\pi}}\, e^{-\frac{h_1}{2}\left(y-\frac{lx}{h_1}\right)^2},$$

$$\Phi_1(x, y)=\sqrt{\frac{h}{2\pi}}\, e^{-\frac{h}{2}\left(x-\frac{ly}{h}\right)^2}. \qquad (287\ \text{bis})$$

Из равенства (274) выводим далее

$$\frac{f(x)}{\varphi(y)}=\frac{\Phi_1(x, y)}{F_1(x, y)}=\sqrt{\frac{h}{h_1}}\, e^{\frac{h_1}{2}y^2+\frac{l^2x^2}{2h_1}-\frac{l^2y^2}{2h}-\frac{h}{2}x^2}=$$

$$=\sqrt{\frac{h}{h_1}}\,\frac{e^{x^2\left(\frac{l^2}{2h_1}-\frac{h}{2}\right)}}{e^{y^2\left(\frac{l^2}{2h}-\frac{h_1}{2}\right)}};$$

заключаем отсюда, что

$$\frac{f(x)}{\sqrt{h}\, e^{\frac{x^2}{2}\left(\frac{l^2}{h_1}-h\right)}}=\frac{\varphi(y)}{\sqrt{h_1}\, e^{\frac{y^2}{2}\left(\frac{l^2}{h}-h_1\right)}}=C, \qquad (290)$$

где постоянная C определяется условием, что

$$\int_{-\infty}^{\infty} f(x)\,dx=\int_{-\infty}^{\infty}\varphi(y)\,dy=C\sqrt{h}\int_{-\infty}^{\infty} e^{\frac{x^2}{2}\left(\frac{l^2}{h_1}-h\right)}dx=$$

$$=C\sqrt{h_1}\int_{-\infty}^{\infty} e^{\frac{y^2}{2}\left(\frac{l^2}{h}-h_1\right)}dy=1,$$

из которого видно, во-первых, что

$$\frac{1}{\sigma^2}=h-\frac{l^2}{h_1}>0 \quad \text{и} \quad \frac{1}{\sigma_1^2}=h_1-\frac{l^2}{h}>0, \qquad (291)$$

и, во-вторых, благодаря равенствам

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx = \sigma\sqrt{2\pi} \quad \text{и} \quad \int_{-\infty}^{\infty} e^{\frac{y^2}{2\sigma_1^2}} dy = \sigma_1\sqrt{2\pi}$$

находим:

$$C = \frac{1}{\sigma\sqrt{2\pi h}} = \frac{1}{\sigma_1\sqrt{2\pi h_1}} = \frac{\sqrt{1-\frac{l^2}{hh_1}}}{\sqrt{2\pi}}. \tag{292}$$

Таким образом в случае нормальной корреляции (как мы ее выше определили) величины x и y во всей совокупности, в целом, также подчиняются закону Гаусса, при чем центры их распределения равны 0, иначе говоря, точка с координатами равными М. О. x и М. О. y совпадает с началом координат, т.-е. с точкой пересечения обеих линий регрессии.

Штандарты же обоих (априорных) распределений соответственно равны σ и σ_1 и определяются из формул (291).

Что же касается функции $P(x, y)$, то она определится из равенства (273)

$$P(x, y) = f(x)\cdot F_1(x, y) \quad \text{или} \quad P(x, y) = \varphi(y)\cdot\Phi_1(x, y),$$

откуда при помощи (287 bis), (290), (291) и (292) находим, что

$$P(x, y) = \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}\cdot\sqrt{\frac{h_1}{2\pi}}\, e^{-\frac{h_1}{2}\left(y-\frac{lx}{h_1}\right)^2} =$$

$$= \frac{\sqrt{hh_1 - l^2}}{2\pi}\, e^{-\frac{1}{2}\left[\frac{x^2}{\sigma^2}+h_1\left(y-\frac{lx}{h_1}\right)^2\right]} =$$

$$= \frac{\sqrt{hh_1 - l^2}}{2\pi}\, e^{-\frac{1}{2}(hx^2+h_1y^2-2lxy)} \tag{293}$$

представляет при любых значениях $h > 0$, $h_1 > 0$ и l (при условии $hh_1 - l^2 > 0$) самую общую функцию распределения вероятностей двух величин x и y,

находящихся в нормальной коррелятивной зависимости, при выборе начала координат так, чтобы М. О. $x=$М. О. $y=0$. Поверхность $z=P(x, y)$ носит название поверхности нормального распределения вероятностей [1]).

Условие [вытекающее из (291)], что

$$hh_1 - l^2 > 0, \quad \text{или} \quad 1 > \frac{l^2}{hh_1} = \rho\rho_1, \tag{294}$$

выражает благодаря (289) и (289 bis), что произведение коэффициентов регрессии (которые, очевидно, всегда имеют одинаковый знак) должно быть меньше единицы. Геометрически это означает, что острый угол φ_1, образуемый линией регрессии y по x с осью абсцисс, всегда меньше острого угла φ, образуемого линией регрессии x по y с той же осью, так как

$$\operatorname{tg}\varphi_1 = \rho_1,$$

$$\operatorname{tg}\varphi = \frac{1}{\rho};$$

иными словами, линия регрессии y по x всегда проходит ближе к оси абсцисс, чем линия регрессии x по y.

3. Величина

$$R = \frac{l}{\sqrt{hh_1}} = \pm\sqrt{\rho\rho_1} \tag{295}$$

называется коэффициентом корреляции между величинами x и y, в согласии с ранее (часть III, глава III) данным общим определением этого коэффициента, в силу которого

$$R = \frac{\text{М. О. } xy}{\sqrt{\text{М. О. } x^2 \cdot \text{М. О. } y^2}} = \frac{\text{М. О. } xy}{\sigma\sigma_1}, \tag{124}$$

---

[1]) Эти поверхности характеризуются тем, что все их сечения плоскостями, параллельными оси OZ, представляют нормальные кривые, а все сечения, перпендикулярные к OZ, являются эллипсами.

(при условии, что М. О. $x=$ М. О. $y=0$). Действительно, вследствие (293)

$$\text{М. О. } xy=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\,P(x,y)\,dx\,dy=$$

$$=\frac{\sqrt{h_1}}{2\pi\sigma}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\,e^{-\frac{x^2}{2\sigma^2}}e^{-\frac{h_1}{2}\left(y-\frac{lx}{h_1}\right)^2}dx\,dy=$$

$$=\frac{\sqrt{h_1}}{2\pi\sigma}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\left(y+\frac{lx}{h_1}\right)e^{-\frac{x^2}{2\sigma^2}}\cdot e^{-\frac{h_1}{2}y^2}dx\,dy=$$

$$=\frac{\sqrt{h_1}}{2\pi\sigma}\int_{-\infty}^{\infty} xe^{-\frac{x^2}{2\sigma^2}}dx\int_{-\infty}^{\infty} ye^{-\frac{h_1}{2}y^2}dy+$$

$$+\frac{l}{2\pi\sigma\sqrt{h_1}}\int_{-\infty}^{\infty} x^2e^{-\frac{x^2}{2\sigma^2}}dx\int_{-\infty}^{\infty} e^{-\frac{h_1}{2}y^2}dy=\frac{l\sigma^2}{h_1}.$$

Следовательно,

$$\frac{\text{М. О. } xy}{\sigma\sigma_1}=\frac{l\sigma}{h_1\sigma_1}=\frac{l}{h_1}\sqrt{\frac{h_1-\frac{l^2}{h}}{h-\frac{l^2}{h_1}}}=\frac{l}{\sqrt{hh_1}}=R. \qquad (124\text{ bis})$$

Из (124 bis) и (289) вытекает также $\rho=\frac{l}{h}=R\frac{\sigma}{\sigma_1}$ и $\rho_1=\frac{l}{h_1}=R\frac{\sigma_1}{\sigma}$; поэтому

$$\rho=\frac{\text{М. О. } xy}{\text{М. О. } y^2},\quad \rho_1=\frac{\text{М. О. } xy}{\text{М. О. } x^2}. \qquad (296)$$

Введя в выражение функции $P(x, y)$ коэффициент корреляции R вместо параметра l и пользуясь формулой (124 bis), получим из (293):

$$P(x,y)=\sqrt{1-R^2}\frac{\sqrt{hh_1}}{2\pi}e^{-\frac{1}{2}(hx^2+h_1y^2-2R\sqrt{hh_1}\,xy)} \qquad (297)$$

или, обозначая через

$$\sigma^*=\frac{1}{\sqrt{h}},\quad \sigma_1{}^*=\frac{1}{\sqrt{h_1}}, \qquad (298)$$

соответственно, условный штандарт x при данном y и условный штандарт y при данном x, находим:

$$P(x, y) = \sqrt{1-R^2}\,\frac{1}{2\pi\sigma^*\sigma_1^*}\,e^{-\frac{1}{2}\left(\frac{x^2}{\sigma^{*2}}+\frac{y^2}{\sigma_1^{*2}}-\frac{2Rxy}{\sigma^*\sigma_1^*}\right)}. \qquad (299)$$

С другой стороны, из равенств (291) получаем:

$$\frac{1}{\sigma^2} = h\left(1-\frac{l^2}{hh_1}\right) = \frac{1}{\sigma^{*2}}(1-R^2),$$

$$\frac{1}{\sigma_1^2} = h_1\left(1-\frac{l^2}{hh_1}\right) = \frac{1}{\sigma_1^{*2}}(1-R^2),$$

и видим, что

$$\sigma^* = \sigma\sqrt{1-R^2}, \qquad \sigma_1^* = \sigma_1\sqrt{1-R^2}. \qquad (300)$$

Таким образом в случае нормальной корреляции штандарт σ величины x (а также y) для всей совокупности всегда в $\frac{1}{\sqrt{1-R^2}}$ раз больше, чем для определенной части ее, выделенной по признаку y. Условный штандарт σ^* тем меньше (по сравнению с σ), чем R больше; отбор из всей совокупности индивидов, обладающих данной величиной признака y, не только сдвигает центр распределения величины x на величину $\rho y = R\frac{\sigma}{\sigma_1}y$, но (независимо от смещения центра) уменьшает вероятные отклонения x от соответствующего центра тем в большей степени, чем $|R|$ ближе к 1.

Введя в формулу (299) вместо σ^* и σ_1^* значения σ и σ_1, можно также представить $P(x, y)$ в виде

$$P(x,y) = \frac{1}{2\pi\sigma\sigma_1\sqrt{1-R^2}}\,e^{-\frac{1}{2(1-R^2)}\left[\frac{x^2}{\sigma^2}+\frac{y^2}{\sigma^2}-\frac{2Rxy}{\sigma\sigma_1}\right]}. \qquad (299\ \text{bis})$$

Отсюда видно, что величины x и y будут независимы в том и только в том случае, когда коэффициент корре-

ляции $R=0$, так как тогда $P(x, y)$ представляется в виде произведения

$$P(x,y)=f(x)\cdot\varphi(y)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}\cdot\frac{1}{\sigma_1\sqrt{2\pi}}e^{-\frac{y^2}{2\sigma_1^2}}.$$

В общем же случае нормальной корреляции, согласно (293),

$$P(x,y)=f(x)\cdot\varphi(y-\rho_1 x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}\frac{1}{\sigma_1^*\sqrt{2\pi}}e^{-\frac{(y-\rho_1 x)^2}{2\sigma_1^{*2}}},$$

и независимыми оказываются величины x и $y-\rho_1 x$; аналогичным образом убеждаемся в независимости величин $x-\rho y$ и y.

Можно доказать, что и обратно, если связь между величинами x и y такова, что существуют 2 числа ρ и ρ_1, обладающие свойствами, что $y-\rho_1 x$ независимо от x и $x-\rho y$ независимо от y, то величины x и y находятся между собой в нормальной корреляции с коэффициентами регрессии ρ и ρ_1.

4. Важность нормальной корреляции, находящей обширные приложения (по крайней мере в первом приближении) особенно в биологических совокупностях, основывается на следующей общей теореме [1]):

Пусть

$$x=u_1+u_2+\ldots+u_n,\qquad y=v_1+v_2+\ldots+v_n$$

представляют суммы весьма большого числа n слагаемых, удовлетворяющих условиям теоремы Ляпунова, при чем [2]) *матем. ожид.* $u_i=0$, *матем. ожид.* $v_i=0$. В таком случае, полагая М. О. $u_i^2=\beta_i$, М. О. $v_i^2=\beta_i'$, М. О. $x^2=\sigma^2=\sum\beta_i$ и М. О. $y^2=\sigma_1^2=\sum\beta_i'$, М. О. $u_i v_i=\gamma_i$, веро-

---

[1]) Более общая форма теоремы дана в моей статье „Sur le théoreme limite du calcul des probabilités". Mathematische Annalen 1922.

[2]) Это условие вводится лишь для упрощения письма, так как при М. О. $u_i=\alpha_i$ и М. О. $v_i=\alpha'_i$ всегда возможно положить $u'_i=u_i-\alpha_i$ и $v'_i=v_i-\alpha'_i$, и теорема будет применима к $\sum u'_i$ и $\sum v'_i$.

ятность совместного осуществления неравенства

$$x_0 < x < x_1, \quad y_0 < y < y_1$$

имеет, при неограниченном возрастании n, пределом

$$\frac{1}{2\pi\sigma\sigma_1\sqrt{1-R^2}}\int_{x_0}^{x_1}\int_{y_0}^{y_1} e^{-\frac{1}{2(1-R^2)}\left[\frac{x^2}{\sigma^2}+\frac{y^2}{\sigma_1^2}-\frac{2Rxy}{\sigma\sigma_1}\right]}dx\,dy, \qquad (301)$$

где

$$R=\frac{\text{М. О. } xy}{\sigma\sigma_1}=\frac{\sum\gamma_i}{\sqrt{\sum\beta_i\sum\beta'_i}}. \qquad (302)$$

Если, например, рассматривается совокупность генотипически тождественных индивидов и исследуются размеры x и y двух (или более) различных их органов, то уклонения x и y от их средней, вызываемые многообразными внешними причинами, влияющими как на x, так и на y, слагаются, соответственно, из большого числа n слагаемых u_i и v_i, которые, вообще, не пропорциональны, но как-то связаны между собой функциональной или коррелятивной (не обязательно нормальной) зависимостью; характер последней связи совершенно не имеет значения, существенна лишь величина М. О. $u_iv_i=\gamma_i$; при этом, как видно из указанной теоремы, зависимость между x и y стремится с возрастанием n стать нормальной, и коэффициент нормальной корреляции R между x и y зависит не от каждого γ_i, а от их суммы $\sum\gamma_i$, так как

$$R=\frac{\sum\gamma_i}{\sqrt{\sum\beta_i\sum\beta'_i}}.$$

В некоторых случаях эта нормальная (или почти нормальная) корреляция очень велика, т.-е. R очень близко к 1, как в рассматриваемом ниже примере длины x бокового листка и длины y конечного листка Trifolium pratense[1]).

---

[1]) Таблица заимствована из книги Johannsen „Elemente der exakten Erblichkeitslehre".

Длина в миллиметрах x	Длина в миллиметрах y: 5	10	15	20	25	30	35	40	45	50	55	60	65	Всего
5	—	3	2	—	—	—	—	—	—	—	—	—	—	5
10	—	1	57	14	—	—	—	—	—	—	—	—	—	72
15	—	—	9	209	63	—	—	—	—	—	—	—	—	281
20	—	—	—	17	348	70	—	—	—	—	—	—	—	435
25	—	—	—	—	25	367	111	—	—	—	—	—	—	503
30	—	—	—	—	—	28	351	104	2	—	—	—	—	485
35	—	—	—	—	—	—	49	219	80	3	—	—	—	351
40	—	—	—	—	—	—	—	20	137	49	2	—	—	208
45	—	—	—	—	—	—	—	1	14	66	26	—	—	107
50	—	—	—	—	—	—	—	—	—	2	28	10	—	40
55	—	—	—	—	—	—	—	—	—	—	—	7	3	10
60	—	—	—	—	—	—	—	—	—	—	—	—	2	2
65	—	—	—	—	—	—	—	—	—	—	—	—	1	1
70														
Всего . . .	4	68	240	436	465	511	344	233	120	56	17	6	2 500	

Центр распределения величины x получается обычным методом средней $a = \frac{\sum x}{n} = 30{,}076$ *мм*, центр величины y равен $a' = \frac{\sum y}{2\,500} = 30{,}832$ *мм*. Аналогичным образом $\sigma = \frac{\sqrt{\sum (x-a)^2}}{n} = 9{,}284$, $\sigma_1 = 9{,}577$. Не останавливаясь здесь на техническом вычислении коэффициента корреляции, заметим, что М. О. $(x-a)(y-a') \neq \frac{\sum (x-a)(y-a')}{n}$ на основании общей теоремы о приближенном равенстве между математическим ожиданием величины и средней арифметической из наблюденных ее значений, а потому

$$R \neq \frac{\sum (x-a)(y-a')}{\sigma\sigma_1 n} \neq \frac{\sum (x-a)(y-a')}{\sqrt{\sum (x-a)^2 \sum (y-a')^2}} \neq 0{,}9627.$$

Подобным образом Пирсон вычислил ряд коэффициентов корреляции между величинами различных органов у человека, из которых приведем здесь некоторые:

		R		
Бедро и плечевая кость	от	0,84	до	0,87
Рост и плечо	„	0,77	„	0,81
Вес и рост (у детей)	„	0,62	„	0,64

5. Весьма важно также применение теории нормальной корреляции к исследованию вопроса о наследственной передаче величин признаков. Мы не можем, конечно, дать здесь сколько-нибудь полного обзора приложений теории вероятностей и, в частности, теории корреляции к этой области. Ограничимся лишь тем, что покажем (вопреки довольно распространенному среди биологов школы Менделя мнению), что так называемый закон наследственной регрессии Гальтона не находится в противоречии с экспериментально установленными элементарными законами наследственности Менделя[1]), а является математическим следствием из теории Менделя для случая сложных количественно измеряемых признаков, размеры которых зависят от значительного числа независимых элементов или ген.

Предположим (игнорируя, для упрощения, влияние внешней среды и полового отбора, которые соответствующим образом изменяют значение полученного дальше коэффициента корреляции), что индивиды данного населения (биотипа) по отношению к некоторому элементарному (в смысле Менделя) признаку могут быть 3 классов, а именно: чистой расы A, чистой расы B или гибридного класса C. Раньше[1]) было указано, каков закон Менделя наследственной передачи этого признака. Если таких элементарных признаков n, то общее число разновидностей индивидов данного биотипа равно 3^n. Предположим далее, что вероятность индивиду принадлежать по отношению к i-му признаку к классам A_i, B_i, C_i равна, соответственно, α_i, β_i, γ_i ($\alpha_i+\beta_i+\gamma_i=1$), и допустим, что распределение вероятностей α_i, β_i, γ_i (согласно установленному нами ранее результату) сохраняется неизменным из поколения в поколе-

[1]) См. стр. 67.

ние. Пусть, далее, рассматриваемый i-й признак заключается, между прочим, в том, что некоторая величина u_i в зависимости от того, к какому из 3 классов A_i, B_i или C_i принадлежит наш индивид, получает значения a_i, b_i, c_i. Тогда для индивидов рассматриваемой совокупности

$$\text{М. О. } u_i = \alpha_i a_i + \beta_i b_i + \gamma_i c_i,$$

при чем мы можем выбрать начало отсчета величины u_i таким образом, чтобы М. О. $u_i = 0$, т.-е.

$$\alpha_i a_i + \beta_i b_i + \gamma_i c_i = 0 \tag{303}$$

для любого поколения; обозначим

$$B_i = \text{М. О. } u_i^2 = \alpha_i a_i^2 + \beta_i b_i^2 + \gamma_i c_i^2.$$

Если какой-нибудь сложный признак [1]) индивида является суммой

$$X = u_1 + u_2 + \cdots + u_n,$$

то (при соответствующих ограничениях о сравнительных размерах и независимости величин u_i, согласно теореме Ляпунова) можем прежде всего утверждать, что при достаточно большом числе слагаемых n величина x будет подчиняться закону Гаусса с центром равным М. О. $x = 0$ и со штандартом $\sigma = \sqrt{\sum B_i}$.

Рассмотрим теперь, каков будет закон наследования указанного сложного признака x, и (при сделанных выше допущениях) докажем закон наследственной регрессии Гальтона, состоящий в том, что между величиной признака x у отца и сына существует нормальная корреляция.

Будем обозначать через x^0 величину признака x у отца, а через x' — величину того же признака у сына и, вообще, все слагаемые u_i у отца будем обозначать через u_i^0 и через u_i' — у сына. Таким образом: М. О. $u_i^0 =$ М. О. $u_i' =$ М. О. $x^0 =$ М. О. $x' = 0$ и $\sigma^0 = \sigma'$, так как, вообще, $B_i^0 = B_i'$. Вычислим $\Gamma_i =$ М. О. $u_i^0 \; u_i'$.

Для этого заметим, что: 1) если отец принадлежит, по отношению к u_i, к первой чистой расе (A_i), то

$$\alpha_i' = \alpha_i + \frac{\gamma_i}{2}, \; \beta_i' = 0, \; \gamma_i' = \beta_i + \frac{\gamma_i}{2},$$

[1]) Случай полимерии.

так как в случае скрещивания отца с самкой той же расы рождается индивид расы (A_i), в случае скрещивания с гибридом единственно возможен и равновероятен сын первой расы или гибрид, а в случае скрещивания с самкой противоположной расы всегда рождается гибрид;

2) если отец — второй чистой расы (B_i), то

$$\alpha^i = 0,\ \beta^i = \beta_i + \frac{\gamma_i}{2},\ \gamma^i = \alpha_i + \frac{\gamma_i}{2};$$

3) наконец, если отец — гибрид, то

$$\alpha^i = \frac{\alpha_i}{2} + \frac{\gamma_i}{4},\quad \beta^i = \frac{\beta_i}{2} + \frac{\gamma_i}{4},\quad \gamma^i = \frac{1}{2}.$$

Поэтому

$$\Gamma_i = \text{M. O. } u_i^0 u^i = a_i^2 \alpha_i \left(\alpha_i + \frac{\gamma_i}{2}\right) +$$

$$+ a_i c_i \alpha_i \left(\beta_i + \frac{\gamma_i}{2}\right) + b_i^2 \beta_i \left(\beta_i + \frac{\gamma_i}{2}\right) + b_i c_i \beta_i \left(\alpha_i + \frac{\gamma_i}{2}\right) +$$

$$+ a_i c_i \gamma_i \left(\frac{\alpha_i}{2} + \frac{\gamma_i}{4}\right) + b_i c_i \gamma_i \left(\frac{\beta_i}{2} + \frac{\gamma_i}{4}\right) + \frac{1}{2} c_i^2 \gamma_i =$$

$$= [a_i^2 \alpha_i^2 + b_i^2 \beta_i^2 + c_i^2 \gamma_i^2 + 2a_i b_i \alpha_i \beta_i + 2a_i c_i \alpha_i \gamma_i + 2b_i c_i \beta_i \gamma_i] +$$

$$+ \left[a_i^2 \frac{\alpha_i \gamma_i}{2} + b_i^2 \frac{\beta_i \gamma_i}{2} + c_i^2 \gamma_i \left(\frac{1}{2} - \gamma_i\right) - 2a_i b_i \alpha_i \beta_i - a_i c_i \alpha_i \gamma_i - b_i c_i \beta_i \gamma_i\right] +$$

$$+ (a_i c_i + b_i c_i)\left(\frac{\gamma_i^2}{4} + \alpha_i \beta_i\right);$$

и так как, по (303), $(a_i \alpha_i + b_i \beta_i + c_i \gamma_i)^2 = 0$, а вследствие (51 bis)

$$\frac{\gamma_i^2}{4} - \alpha_i \beta_i = 0,$$

то

$$\Gamma_i = \frac{a_i^2 \alpha_i \gamma_i + b_i^2 \beta_i \gamma_i + c_i^2 \gamma_i - a_i b_i \gamma_i^2}{2} - c_i \gamma_i (a_i \alpha_i + b_i \beta_i + c_i \gamma_i) +$$

$$+ \frac{a_i c_i + b_i c_i}{2} \gamma_i^2 = \frac{\gamma_i}{2} [a_i^2 \alpha_i + b_i^2 \beta_i + c_i^2 - a_i b_i \gamma_i + a_i c_i \gamma_i + b_i c_i \gamma_i] =$$

$$= \frac{\gamma_i}{2} [c_i^2 - a_i b_i \gamma_i - a_i b_i \beta_i - a_i b_i \alpha_i] = \frac{\gamma_i}{2} (c_i^2 - a_i b_i).$$

Таким образом, согласно указанной выше теореме, между признаками x_0 и x_0' отца и сына существует нормальная корреляция, при чем коэффициент R между ними равен

$$R=\frac{1}{2}\frac{\sum\gamma_i(c_i^2-a_ib_i)}{\sum(a_i^2\alpha_i+b_i^2\beta_i+c_i^2\gamma_i)}. \qquad (302\ \text{bis})$$

Нетрудно видеть, что корреляция всегда положительна, так как

$$c_2^i-a_ib_i\geqslant 0,$$

вследствие того, что иначе не могло бы быть соблюдено равенство

$$\alpha_ia_i+\beta_ib_i+\gamma_ic_i=\alpha_ia_i+\beta_ib_i+2c_i\sqrt{\alpha_i\beta_i}=0. \qquad (303\ \text{bis})$$

Кроме того, принимая во внимание, что

$$2\Gamma_i=\gamma_i(c_i^2-a_ib_i)=\gamma_i\left[\left(\frac{a_i\alpha_i+b_i\beta_i}{\gamma_i}\right)^2-a_ib_i\right]=$$

$$=\frac{1}{\gamma_i}\left[(a_i\alpha_i-b_i\beta_i)^2+4a_ib_i\alpha_i\beta_i-a_ib_i\gamma_i^2\right]=\frac{1}{\gamma_i}(a_i\alpha_i-b_i\beta_i)^2,$$

между тем как

$$B_i=a_i^2\alpha_i+b_i^2\beta_i+c_i^2\gamma_i=a_i^2\alpha_i+b_i^2\beta_i+\frac{(a_i\alpha_i+b_i\beta_i)^2}{\gamma_i}=$$

$$=a_i^2\alpha_i+b_i^2\beta_i+a_ib_i\gamma_i+\frac{(a_i\alpha_i-b_i\beta_i)^2}{\gamma_i},$$

получаем, что

$$\frac{2\Gamma_i}{B_i}=\frac{\gamma_i(c_i^2-a_ib_i)}{a_i^2\alpha_i+b_i^2\beta_i+c_i^2\gamma_i}=\frac{1}{1+\dfrac{\gamma_i(a_i^2\alpha_i+b_i^2\beta_i+a_ib_i\gamma_i)}{(a_i\alpha_i-b_i\beta_i)^2}}=$$

$$=\frac{1}{1+2\sqrt{\alpha_i\beta_i}\left(\dfrac{a_i\sqrt{\alpha_i}+b_i\sqrt{\beta_i}}{a_i\alpha_i-b_i\beta_i}\right)^2}\leqslant 1.$$

Отсюда следует, что максимальное возможное значение [1]) коэффициента R корреляции между отцом и сыном равно $\frac{1}{2}$;

[1]) Если $\gamma_i=0$, для некоторых значений i, то соответствующие слагаемые в числителе и знаменателе (302 bis) обращаются в нуль (так как тогда $\beta_i=a_i=0$ или $\alpha_i=b_i=0$) и поэтому не влияют на значение R. Этим обстоятельством объясняется то, что для генотипически чистых рас, где уклонения признака

достигается оно лишь при предположении, что $c_i = \frac{a_i + b_i}{2}$, так как для этого нужно, чтобы $a_i\sqrt{\alpha_i} + b_i\sqrt{\beta_i} = 0$, в то время как, по условию, $a_i\alpha_i + b_i\beta_i + 2c_i\sqrt{\alpha_i\beta_i} = 0$.

При сделанных нами допущениях коэффициент корреляции совпадает, очевидно, с коэффициентами регрессии.

Применяя те же соображения, можно было бы показать, что коэффициент корреляции между дедом и внуком равен $\frac{1}{2}R$ и, вообще, между каждым индивидом и его предком n-й степени существует нормальная корреляция с коэффициентом $\left(\frac{1}{2}\right)^n R$, где R дан тою же формулой (302 bis) (разумеется при условии устанановившегося распределения вероятностей для всех поколений).

6. Упражнения. 1) Показать, что при нормальной корреляции (полагая М. О. $x =$ М. О. $y = 0$) коэффициент корреляции $R = \cos q\pi$, где q есть вероятность, что x и y противоположны по знаку.

Отв. Для доказательства замечаем, что на основании формулы (299 bis)

$$q = \frac{1}{\pi\sigma\sigma_1\sqrt{1-R^2}}\int_0^{\infty}\int_{-\infty}^{0} e^{-\frac{x^2\sigma_1^2 + y^2\sigma^2 - 2R\,\sigma\sigma_1\,xy}{2\sigma^2\sigma_1^2(1-R^2)}}\,dx\,dy =$$

$$= \frac{1}{\pi\sigma\sigma_1\sqrt{1-R^2}}\int_{-\frac{\pi}{2}}^{0}\int_0^{\infty} e^{-\rho^2\left[\frac{\sigma_1^2\cos\theta + \sigma^2\sin^2\theta - 2R\,\sigma\sigma_1\cos\theta\sin\theta}{2\sigma^2\sigma_1^2(1-R^2)}\right]}\rho\,d\rho\,d\theta =$$

$$= \frac{\sigma\sigma_1\sqrt{1-R^2}}{\pi}\int_{-\frac{\pi}{2}}^{0}\frac{d\theta}{\sigma^2\sin^2\theta + \sigma_1^2\cos^2\theta - 2R\,\sigma\sigma_1\cos\theta\sin\theta} =$$

$$= \frac{\sigma\sigma_1\sqrt{1-R^2}}{\pi}\int_{-\infty}^{0}\frac{du}{\sigma^2u^2 - 2\sigma\sigma_1Ru + \sigma_1^2} =$$

$$= \frac{1}{\pi}\left[\frac{\pi}{2} - \operatorname{arc\,tg}\frac{R}{\sqrt{1-R^2}}\right] = \frac{1}{\pi}\arccos R.$$

от его средней вызываются только внешними причинами (иначе все индивиды были бы тождественны), коэффициент корреляции между предками и потомками оказывается равным нулю. Опыты Johannsen'a с чистыми расами, установившие упомянутый факт, послужили основанием для признания несостоятельности закона наследственной регрессии Гальтона; в действительности, из вышеприведенных вычислений вытекает, что закон Гальтона применим к генотипически сложным (не чистым) биотипам.

Замечая, что, при $p+q=1$, $\cos q\pi = \sin\frac{p-q}{2}$ получаем также $R = \sin\frac{p-q}{2}\pi \neq \frac{p-q}{2}\pi$; последнее приближенное значение дает R с точностью до 0,02, если $|R| < \frac{1}{2}$.

2) Величины y и x связаны функциональной зависимостью $y = x^m\sqrt{m}$ $\left(\text{откуда } x = y^{\frac{1}{m}} m^{\frac{1}{2m}}\right)$, при чем распределение вероятностей x равномерно на отрезке $(-1, +1)$. Вычислить коэффициент корреляции R между x и y и показать, что R стремится к 0, когда m неограниченно возрастает.

Отв. $R = \frac{\sqrt{6m+3}}{m+2}$. (Например: при $m=13$, $R=0{,}6$; при $m=181$, $R = \frac{11}{61} \neq 0{,}180$.) Действительно, М. О. $y =$ М. О. $x = 0$,

$$\sigma^2 = \text{М. О. } x^2 = \frac{1}{2}\int_{-1}^{+1} x^2\,dx = \frac{1}{3},$$

$$\sigma_1^2 = \text{М. О. } y^2 = \frac{m}{2}\int_{-1}^{+1} x^{2m}\,dx = \frac{m}{2m+1},$$

$$R\sigma\sigma_1 = \text{М. О. } xy = \frac{\sqrt{m}}{2}\int_{-1}^{+1} x^{m+1}\,dx = \frac{\sqrt{m}}{m+2},$$

откуда $R = \frac{\sqrt{m}}{m+2}\sqrt{3\cdot\frac{2m+1}{m}} = \frac{\sqrt{6m+3}}{m+2}$. Таким образом, даже при обратимой функциональной зависимости, коэффициент корреляции может быть сколь угодно близок к 0.

3) Показать, что если $y = x^2 - \frac{1}{3}$ и распределение вероятностей для x на отрезке $(-1, +1)$ равномерно, то коэффициент корреляции R между x и y равен 0.

Отв. М. О. $x =$ М. О. $y =$ М. О. $xy = 0$; М. О. $x^2 = \frac{1}{3}$, М. О. $y^2 = \frac{4}{45}$, поэтому $R = 0$.

4) Показать, что если величины x, y связаны нормальной корреляцией при коэффициенте корреляции R, то коэффициент корреляции R_1 между x и $\varphi(y)$, где $\varphi(y)$ какая-нибудь данная функция y, всегда менее R по абсолютному значению.

Отв. $$R_1 = \frac{\iint (x + Ry)\varphi(y) e^{-\frac{x^2}{2(1-R^2)} - \frac{y^2}{2}} dx\, dy}{\sqrt{2\pi(1-R^2)\int x^2 e^{-\frac{x^2}{2}} dx \int \varphi^2(y) e^{-\frac{y^2}{2}} dy}} =$$

$$= R \cdot \frac{\int y\, \varphi(y) e^{-\frac{y^2}{2}} dy}{\sqrt{\int y^2 e^{-\frac{y^2}{2}} dy \int \varphi^2(y) e^{-\frac{y^2}{2}} dy}}$$

и так как множитель при R представляет коэффициент корреляции между y и $\varphi(y)$, который по абсолютному значению меньше 1, то $|R_1| < |R|$.

5) Определить штандарт $\sigma(R)$ погрешности коэффициента корреляции R между x и y при применении приближенной формулы $R \neq -\cos\frac{m}{n}\pi$, где $\frac{m}{n}$ статистическая вероятность того, что x и y имеют одинаковый знак (корреляция предполагается нормальной).

Отв. $$\sigma(R) = \pi\sqrt{\frac{m(n-m)}{n^3}}\sin\frac{m}{n}\pi \neq \pi\sqrt{\frac{m(n-m)}{n^3}(1-R^2)}.$$

6) Вычислить штандарты погрешности коэффициента корреляции R и коэффициентов регрессии ρ и ρ_1 при статистическом вычислении их по приближенным формулам $R \neq \frac{\sum xy}{\sqrt{\sum x^2 \cdot \sum y^2}}$, $\rho \neq \frac{\sum xy}{\sum y^2}$, $\rho_1 \neq \frac{\sum xy}{\sum x^2}$ (корреляция предполагается нормальной) на основании n наблюдений пары (x, y)

Отв. $\sigma(R) \neq \frac{1-R^2}{\sqrt{n}}$, $\sigma(\rho) \neq \frac{\sigma}{\sigma_1}\sqrt{\frac{1-R^2}{n}}$; $\sigma(\rho_1) \neq \frac{\sigma_1}{\sigma}\sqrt{\frac{1-R^2}{n}}$. Указанные формулы получаем, применяя результат, найденный в 3-м упражнении III главы IV части. Вычислим, например, $\sigma(\rho_1)$. Сначала находим

$$\sigma^2\left(\sum xy\right) = \text{M. O.}\left[\sum xy - nR\sigma\sigma_1\right]^2 =$$

$$= n\left[\text{M. O. } x^2y^2 - R^2\sigma^2\sigma_1^2\right] = n\sigma^2\sigma_1^2 \cdot (1 + R^2),$$

так как

$$\text{M. O. } x^2y^2 = \frac{1}{2\pi\sigma\sigma_1\sqrt{1-R^2}}\iint x^2y^2 e^{-\frac{x^2\sigma_1^2 + y^2\sigma^2 - 2R\sigma\sigma_1 xy}{2\sigma^2\sigma_1^2(1-R^2)}} dx\, dy =$$

$$= \frac{1}{2\pi\sigma\sigma_1\sqrt{1-R^2}}\iint (x^2y^2 + \rho_1^2 x^4)\, e^{-\frac{x^2}{2\sigma^2} - \frac{y^2}{2\sigma_1^2(1-R^2)}} dx\, dy =$$

$$= \sigma^2\sigma_1^2(1-R^2) + 3\rho_1^2\sigma^4 = \sigma^2\sigma_1^2(1 + 2R^2).$$

Мы знаем (стр. 292), что $\sigma^2\left(\sum x^2\right)=2n\sigma^4$. Остается вычислить коэффициент r корреляции между $\sum xy$ и $\sum x^2$; находим, что

$$r=\frac{\text{М. О.}\left[\sum xy-nR\,\sigma\sigma_1\right]\left[\sum x^2-n\sigma^2\right]}{n\sigma^3\sigma_1\sqrt{2(1+R^2)}}=$$

$$=\frac{\text{М. О. }x^3y-R\,\sigma^3\sigma_1}{\sigma^3\sigma_1\sqrt{2(1+R^2)}}=R\sqrt{\frac{2}{1+R^2}},$$

так как

$$\text{М. О. }x^3y=\frac{1}{2\pi\sigma\sigma_1\sqrt{1-R^2}}\iint x^3(y+\rho_1x)\,e^{-\frac{x^2}{2\sigma^2}-\frac{y^2}{2\sigma_2^1(1-R^2)}}\,dx\,dy=$$

$$=3\rho_1\sigma^4=3R\,\sigma^3\sigma_1.$$

Следовательно,

$$\sigma\left[\frac{\sum xy}{\sum x^2}\right]=$$

$$=\frac{1}{n\sigma^2}\sqrt{n\sigma^2\sigma_1^2(1+R^2)-2R^2\sqrt{\frac{2}{1+R^2}}\sqrt{2n^2\sigma^4\sigma_1^4(1+R^2)}+2\,n\sigma^2\sigma_1^2\,R^2}=$$

$$=\frac{\sigma_1}{\sigma}\sqrt{\frac{1-R^2}{n}}$$

ТАБЛИЦА ЗНАЧЕНИЙ

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{1}{2}z^2} dz$$

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
0,00	0,0000	0,32	0,1255	0,64	0,2389	0,96	0,3315
0,01	0,0040	0,33	0,1293	0,65	0,2422	0,97	0,3340
0,02	0,0080	0,34	0,1331	0,66	0,2454	0,98	0,3365
0,03	0,0120	0,35	0,1368	0,67	0,2486	0,99	0,3389
0,04	0,0160	0,36	0,1406	0,68	0,2517	1,00	0,3413
0,05	0,0199	0,37	0,1443	0,69	0,2549	1,01	0,3438
0,06	0,0239	0,38	0,1480	0,70	0,2580	1,02	0,3461
0,07	0,0279	0,39	0,1517	0,71	0,2611	1,03	0,3485
0,08	0,0319	0,40	0,1554	0,72	0,2642	1,04	0,3508
0,09	0,0359	0,41	0,1591	0,73	0,2673	1,05	0,3531
0,10	0,0398	0,42	0,1628	0,74	0,2703	1,06	0,3554
0,11	0,0438	0,43	0,1664	0,75	0,2734	1,07	0,3577
0,12	0,0478	0,44	0,1700	0,76	0,2764	1,08	0,3599
0,13	0,0517	0,45	0,1736	0,77	0,2794	1,09	0,3621
0,14	0,0557	0,46	0,1772	0,78	0,2823	1,10	0,3643
0,15	0,0596	0,47	0,1808	0,79	0,2852	1,11	0,3665
0,16	0,0636	0,48	0,1844	0,80	0,2881	1,12	0,3686
0,17	0,0675	0,49	0,1879	0,81	0,2910	1,13	0,3708
0,18	0,0714	0,50	0,1915	0,82	0,2939	1,14	0,3729
0,19	0,0753	0,51	0,1950	0,83	0,2967	1,15	0,3749
0,20	0,0793	0,52	0,1985	0,84	0,2995	1,16	0,3770
0,21	0,0832	0,53	0,2019	0,85	0,3023	1,17	0,3790
0,22	0,0871	0,54	0,2054	0,86	0,3051	1,18	0,3810
0,23	0,0910	0,55	0,2088	0,87	0,3078	1,19	0,3830
0,24	0,0948	0,56	0,2123	0,88	0,3106	1,20	0,3849
0,25	0,0987	0,57	0,2157	0,89	0,3133	1,21	0,3869
0,26	0,1026	0,58	0,2190	0,90	0,3159	1,22	0,3888
0,27	0,1064	0,59	0,2224	0,91	0,3186	0,23	0,3907
0,28	0,1103	0,60	0,2257	0,92	0,3212	1,24	0,3925
0,29	0,1141	0,61	0,2291	0,93	0,3238	1,25	0,3944
0,30	0,1179	0,62	0,2324	0,94	0,3264	1,26	0,3962
0,31	0,1217	0,63	0,2357	0,95	0,3289	1,27	0,3980

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
1,28	0,3997	1,61	0,4463	1,94	0,4738	2,54	0,4945
1,29	0,4015	1,62	0,4474	1,95	0,4744	2,56	0,4948
1,30	0,4032	1,63	0,4484	1,96	0,4750	2,58	0,4951
1,31	0,4049	1,64	0,4495	1,97	0,4756	2,60	0,4953
1,32	0,4066	1,65	0,4505	1,98	0,4761	2,62	0,4956
1,33	0,4082	1,66	0,4515	1,99	0,4767	2,64	0,4959
1,34	0,4099	1,67	0,4525	2,00	0,4772	2,66	0,4961
1,35	0,4115	1,68	0,4535	2,02	0,4783	2,68	0,4963
1,36	0,4131	1,69	0,4545	2,04	0,4793	2,70	0,4965
1,37	0,4147	1,70	0,4554	2,06	0,4803	2,72	0,4967
1,38	0,4162	1,71	0,4564	2,08	0,4812	2,74	0,4969
1,39	0,4177	1,72	0,4573	2,10	0,4821	2,76	0,4971
1,40	0,4192	1,73	0,4582	2,12	0,4830	2,78	0,4973
1,41	0,4207	1,74	0,4591	2,14	0,4838	2,80	0,4974
1,42	0,4222	1,75	0,4599	2,16	0,4846	2,82	0,4976
1,43	0,4236	1,76	0,4608	2,18	0,4854	2,84	0,4977
1,44	0,4251	1,77	0,4616	2,20	0,4861	2,86	0,4979
1,45	0,4265	1,78	0,4625	2,22	0,4868	2,88	0,4980
1,46	0,4279	1,79	0,4633	2,24	0,4875	2,90	0,4981
1,47	0,4292	1,80	0,4641	2,26	0,4881	2,92	0,4982
1,48	0,4306	1,81	0,4649	2,28	0,4887	2,94	0,4984
1,49	0,4319	1,82	0,4656	2,30	0,4893	2,96	0,4985
1,50	0,4332	1,83	0,4664	2,32	0,4898	2,98	0,4986
1,51	0,4345	1,84	0,4671	2,34	0,4904	3,00	0,49865
1,52	0,4357	1,85	0,4678	2,36	0,4909	3,20	0,49931
1,53	0,4370	1,86	0,4686	2,38	0,4913	3,40	0,49966
1,54	0,4382	1,87	0,4693	2,40	0,4918	3,60	0,499841
1,55	9,4394	1,88	0,4699	2,42	0,4922	3,80	0,499928
1,56	0,4406	1,89	0,4706	2,44	0,4927	4,00	0,499968
1,57	0,4418	1,90	0,4713	2,46	0,4931	4,50	0,499997
1,58	0,4429	1,91	0,4719	2,48	0,4934	5,00	0,499997
1,59	0,4441	1,92	0,4726	2,50	0,4938		
1,60	0,4452	1,93	0,4732	2,52	0,4941		

ТАБЛИЦА ЗНАЧЕНИЙ

$$\Phi'(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$ (все значения умножены на 10 000)

	0	1	2	3	4	5	6	7	8	9
0,0	3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3725	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3	0044	0033	0024	0017	0012	0009	0006	0004	0003	0002
4	0001	0001	0001	0000	0000	0000	0000	0000	0000	0000