В. И. ГЛИВЕНКО

КУРС ТЕОРИИ ВЕРОЯТНОСТЕЙ

ГОНТИ · 1939

В.И. ГЛИВЕНКО · КУРС ТЕОРИИ ВЕРОЯТНОСТЕЙ

ОПЕЧАТКИ

Страница	Строка	Напечатано	Следует читать	По чьей вине
34	12 сн.	$C_n^{m_1}$	$C_n^{m_1}$	ред.
47	2 сн.	$\Rightarrow$	$\rightrightarrows$	»
48	1 св.	$-\frac{\sqrt{np}}{q}$	$-\frac{\sqrt{np}}{\sqrt{q}}$	тип.
49	5 сн.	$+\frac{p\sqrt{p}}{\sqrt{q}}$	$-\frac{p\sqrt{p}}{\sqrt{q}}$	ред.
136	7 сн.	$\lim\limits_{n\to\infty} F_n(x') \leqslant \lim\limits_{n\to\infty} F_n(x) \leqslant$	$\varliminf\limits_{n\to\infty} F_n(x') \leqslant \varliminf\limits_{n\to\infty} F_n(x) \leqslant$	»
142	7 сн.	неравенство	неравенства	»
152	13 св.	$\frac{\nu - A_n}{\sqrt{B_n}}$	$\frac{v - A_n}{\sqrt{B_n}}$	»
157	2 св.	$\leqslant \frac{1}{2}$	$\geqslant \frac{1}{2}$	»
184	11 сн.	$\sum\limits_{x_i < x}$	$\sum\limits_{m_i < x}$	»
192	15 сн.	$R_{AB} = R_{\bar{A}\bar{B}}R$	$R_{AB} = R_{\bar{A}\bar{B}}$	»
192	11 сн.	(123)	(108)	»
220	4 сн.	дисперсии DX	дисперсии Dx	тип.

Зак. 1725. В. И. Гливенко. Курс теории вероятностей.

В. И. ГЛИВЕНКО

Проф. Московского Педагогического института им. К. Либкнехта

КУРС
ТЕОРИИ ВЕРОЯТНОСТЕЙ

Допущено Всесоюзным комитетом по делам высшей школы при СНК СССР в качестве учебника для физико-математических факультетов государственных университетов

ГОСУДАРСТВЕННОЕ
ОБЪЕДИНЕННОЕ НАУЧНО-ТЕХНИЧЕСКОЕ ИЗДАТЕЛЬСТВО
РЕДАКЦИЯ ТЕХНИКО-ТЕОРЕТИЧЕСКОЙ ЛИТЕРАТУРЫ
Москва 1939 Ленинград

ОГЛАВЛЕНИЕ.

ГЛАВА I.

ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ.

§ 1. События.

1. Первое понятие, с которым мы встречаемся в теории вероятностей, это — понятие *события*. Мы называем событием всякое явление, о котором имеет смысл говорить, что оно происходит или что оно не происходит (в настоящем, в прошедшем или в будущем). В дальнейшем мы всюду будем обозначать события прописными буквами $A, B, C, \ldots$

Между событиями могут существовать некоторые соотношения, с которыми мы постоянно будем иметь дело и которые мы поэтому прежде всего опишем и изучим.

Может случиться, что из наступления некоторого события A с необходимостью вытекает наступление некоторого другого события B. Представим себе, например, что мы бросаем игральную кость, на каждой из шести граней которой выбито соответственно 1, 2, 3, 4, 5 и 6 очков. Обозначим через A событие, состоящее в том, что при бросании игральной кости на верхней ее грани выпадает одно очко, а через B — событие, состоящее в том, что выпадает нечетное число очков. Здесь при наступлении события A наступает и событие B.

Возьмем газ, заключенный в сосуде, и будем следить за движением какой-нибудь определенной молекулы этого газа. Выделим мысленно объем $\mathfrak{A}$ внутри сосуда и обозначим через A попадание молекулы в этот объем. Выделим затем объем $\mathfrak{B}$, в котором объем $\mathfrak{A}$ содержится как часть, и обозначим через B попадание молекулы в объем $\mathfrak{B}$. Здесь при наступлении события A наступает и событие B.

Возьмем совокупность предметов, часть которых обладает некоторым признаком X, и будем выбирать наугад один за другим два предмета из этой совокупности. Обозначим через A событие, состоящее в том, что оба выбранных предмета окажутся обладающими признаком X. Обозначим через B событие, состоящее в том, что по крайней мере один из выбранных предметов окажется обладающим признаком X. Здесь при наступлении события A наступает и событие B.

Когда из наступления события A с необходимостью вытекает наступление события B, мы будем говорить, что A *влечет за собой* B, и будем обозначать это символом

$$A \subset B.$$

Ясно, что

1° $A \subset A$;

2° *Если* $A \subset B$ *и* $B \subset C$, *то* $A \subset C$.

Действительно, соотношение 1° означает, что из наступления любого события A с необходимостью вытекает наступление его самого. Соотношение же 2° означает, что если событие A влечет за собой событие B, а событие B влечет за собой событие C, то тем самым событие A влечет за собой событие C.

Пусть теперь не только из наступления события A с необходимостью вытекает наступление события B, но и обратно, из наступления события B также с необходимостью вытекает наступление события A. Иначе говоря, пусть A происходит тогда и только тогда, когда происходит B.

В таком случае мы будем говорить, что события A и B *равносильны*, и будем обозначать это символом

$$A = B.$$

2. Условимся называть *произведением* двух событий A и B и обозначать символом

$$AB$$

событие, равносильное наступлению *обоих* событий A и B. Это может означать, например, просто одновременное наступление событий A и B. Но слово „одновременное" не всегда уместно для описания наступления обоих событий A и B. Представим себе, например, что мы бросаем игральную кость два раза подряд. Пусть при этом A есть выпадение одного очка при первом бросании и пусть B — выпадение тоже одного очка, но при втором бросании. Здесь наступлением обоих событий A и B будет выпадение одного очка при обоих бросаниях. Очевидно, неудобно было бы сказать, что одно очко выпадает одновременно сначала при первом бросании и потом при втором.

Возьмем газ, заключенный в сосуде, и будем следить за движением какой-нибудь определенной молекулы этого газа. Выделим мысленно два объема $\mathfrak{A}$ и $\mathfrak{B}$ внутри сосуда, частично перекрывающие друг друга, и обозначим через A и B соответственно попадание молекулы в объем $\mathfrak{A}$ и в объем $\mathfrak{B}$. Здесь наступлением обоих событий A и B будет попадание молекулы в общую часть объемов $\mathfrak{A}$ и $\mathfrak{B}$.

Возьмем совокупность предметов, часть которых обладает некоторым признаком X, и будем выбирать наугад один за другим два предмета из этой совокупности. Обозначим через A и B соответственно события, состоящие в том, что первый из выбранных предметов окажется обладающим признаком X, и в том, что второй из выбранных предметов окажется обладающим признаком X. Здесь наступлением обоих событий A и B будет событие, состоящее в том, что оба выбранных предмета окажутся обладающими признаком X.

Ясно, что

3° $AB \subset A$;
4° $AB \subset B$;
5° *Если* $C \subset A$ *и* $C \subset B$, *то* $C \subset AB$.

Действительно, соотношения 3° и 4° означают, что когда происходят оба события A и B, то происходит и каждое из них в отдельности. Соотношение же 5° означает, что если событие C влечет за собой каждое из событий A и B в отдельности, то оно влечет за собой оба события A и B. Соотношениями 3°, 4° и 5° произведение AB определяется вполне однозначно, так что если $A_1 = A_2$ и $B_1 = B_2$, то и $A_1B_1 = A_2B_2$. Действительно, если $A_1 = A_2$ и $B_1 = B_2$, то, в силу 3° и 4°, $A_1B_1 \subset A_2$ и $A_1B_1 \subset B_2$. а потому, в силу 5°, $A_1B_1 \subset A_2B_2$, и точно так же доказывается, что $A_2B_2 \subset A_1B_1$.

Очевидно, что $AA = A$, $AB = BA$ и $(AB)C = A(BC)$. В частности, событие $(AB)C$, или, что то же, $A(BC)$, равносильно наступлению всех трех событий A, B и C. Его можно обозначить просто через ABC. Тот же символ распространяется и на любое число событий.

Условимся затем называть *суммой* двух событий A и B и обозначать символом

$$A + B$$

событие, равносильное наступлению хотя бы одного из событий A или B. Так, обратимся снова к нашему примеру однократного бросания игральной кости и обозначим через A выпадение при этом бросании трех очков, а через B — выпадение шести очков. Здесь наступлением хотя бы одного из событий A или B будет выпадение числа очков, кратного трем.

Возьмем опять газ, заключенный в сосуде, и будем следить за движением какой-нибудь определенной молекулы этого газа. Выделим мысленно два объема $\mathfrak{A}$ и $\mathfrak{B}$ внутри сосуда и обозначим через $\mathfrak{C}$ объем, образованный соединением друг с другом объемов $\mathfrak{A}$ и $\mathfrak{B}$. Обозначим через A и B соответственно попадание молекулы в объем $\mathfrak{A}$ и в объем $\mathfrak{B}$. Здесь наступлением хотя бы одного из событий A или B будет попадание молекулы в объем $\mathfrak{C}$.

Возьмем опять совокупность предметов, часть которых обладает некоторым признаком X, и будем выбирать наугад один за другим два предмета из этой совокупности. Обозначим через A и B соответственно события, состоящие в том, что первый из выбранных предметов окажется обладающим признаком X, и в том, что второй из выбранных предметов окажется обладающим признаком X. Здесь наступлением хотя бы одного из событий A или B будет событие, состоящее в том, что по крайней мере один из выбранных предметов окажется обладающим признаком X.

Ясно, что

6° $A \subset A + B$;
7° $B \subset A + B$;
8° *Если* $A \subset C$ *и* $B \subset C$, *то* $A + B \subset C$.

Действительно, соотношения 6° и 7° означают, что когда происходит какое-либо из событий A или B, то тем самым происходит по крайней мере одно из них. Соотношение же 8° означает, что если из наступления каждого из событий A и B в отдельности вытекает событие C, то и из наступления по крайней мере одного из них вытекает событие C. Соотношениями 6°, 7° и 8° сумма $A+B$ определяется вполне однозначно, так что если $A_1=A_2$ и $B_1=B_2$, то и $A_1+B_1=A_2+B_2$. В этом можно убедиться таким же путем, каким мы убедились в том, что соотношения 3°, 4° и 5° определяют произведение AB.

Очевидно, что $A+A=A$, $A+B=B+A$ и $(A+B)+C=A+$ $+(B+C)$. В частности, событие $(A+B)+C$, или, что то же, $A+(B+C)$, равносильно наступлению хотя бы одного из трех событий A, B или C. Его можно обозначить просто через $A+B+C$. Тот же символ распространяется и на любое число событий. Для примера обратимся снова к однократному бросанию игральной кости и будем обозначать через A выпадение одного очка, через B — выпадение трех очков и через C — выпадение пяти очков; здесь событием $A+B+C$ будет выпадение нечетного числа очков.

Отметим еще одну важную связь между произведениями и суммами событий, а именно:

9° *Каковы бы ни были три события A, B и C,*

$$A(B+C)=AB+AC.$$

В самом деле, если произойдет событие A и хотя бы одно из событий B или C, то, очевидно, случится хотя бы одно из двух: либо произойдут события A и B, либо произойдут события A и C. Таким образом

$$A(B+C)\subset AB+AC.$$

Наоборот, если случится хотя бы одно из двух: либо произойдут события A и B, либо произойдут события A и C, то, очевидно, произойдет событие A и хотя бы одно из событий B или C. Таким образом и обратно

$$AB+AC\subset A(B+C).$$

То же справедливо и для любого числа событий, так что вообще

$$A(B+C+\ldots+S)=AB+AC+\ldots+AS.$$

3. Событие называется *достоверным*, если оно с необходимостью должно произойти. Так, когда мы ставим ту или иную задачу о бросании игральной кости, мы, естественно, считаем достоверным, что в результате этого бросания выпадет к а к о е-т о ч и с л о очков.

Событие называется *невозможным*, если достоверно его ненаступление. Так, при бросании игральной кости невозможно, чтобы в результате бросания не выпало н и к а к о г о ч и с л а очков.

Пусть даны: какое-нибудь событие A, достоверное событие U и невозможное событие V. Тогда, во-первых, вопрос о наступлении о б о и х событий A и U сводится к вопросу о наступлении одного только события A, так как U все равно должно произойти; во-вторых, вопрос о наступлении по крайней мере одного из событий A или V также сводится к вопросу о наступлении одного только события A, так как V все равно не может произойти:

10° *Каково бы ни было событие* A,

$$AU = A;$$

11° *Каково бы ни было событие* A,

$$A + V = A.$$

Нетрудно убедиться в том, что этими условиями 10° и 11° достоверное событие U и невозможное событие V определяются вполне однозначно, так что все достоверные события оказываются равносильными между собой и все невозможные — равносильными между собой. Действительно, пусть U_1 и U_2 — достоверные события; тогда, в силу 10°, $U_1 = U_1U_2 = U_2U_1 = U_2$; точно так же, пусть V_1 и V_2 — невозможные события; тогда, в силу 11°, $V_1 = V_1 + V_2 = V_2 + V_1 = V_2$.

Условимся теперь обозначать символом

$$\overline{A}$$

событие, равносильное ненаступлению события A. Так, если обратиться к приведенному выше примеру двукратного бросания игральной кости, где A есть выпадение одного очка при первом бросании, то $\overline{A}$ будет обозначать ненаступление этого, т. е. будет обозначать выпадение более чем одного очка при первом бросании. Событие $\overline{A}$ называется *противоположным* событию A.

Возьмем газ, заключенный в сосуде, и будем следить за движением какой-нибудь определенной молекулы этого газа. Выделим мысленно объем $\mathfrak{A}$ внутри сосуда и обозначим через $\overline{\mathfrak{A}}$ весь сосуд за исключением объема $\mathfrak{A}$. Обозначим через A попадание молекулы в объем $\mathfrak{A}$. Тогда событием, противоположным событию A, будет попадание молекулы в объем $\overline{\mathfrak{A}}$.

Возьмем совокупность предметов, часть которых обладает некоторым признаком X, и будем выбирать наугад один предмет из этой совокупности. Обозначаем через A событие, состоящее в том, что выбранный предмет окажется обладающим признаком X. Тогда событие, противоположное событию A, будет состоять в том, что выбранный предмет окажется не обладающим признаком X.

Рассмотрим теперь какое угодно событие A. Если мы возьмем событие, равносильное тому, что A либо наступит, либо не наступит, т. е. событие $A + \overline{A}$, то это последнее, очевидно, всегда будет достоверно; точно так же, если мы возьмем событие, равносильное тому, что A и наступит и не наступит, т. е. событие $A\overline{A}$, то это последнее, очевидно, всегда будет невозможно:

12° $A + \overline{A} = U$.

13° $A\overline{A} = V$.

Нетрудно убедиться в том, что этими соотношениями 12° и 13° противоположное событие $\overline{A}$ определяется вполне однозначно, так что если $A=B$, то и $\overline{A}=\overline{B}$. Действительно, если $A=B$, то на основании 9°, 10°, 11°, 12° и 13° мы имеем: $\overline{A}=\overline{A}(B+\overline{B})=\overline{A}B+\overline{A}\overline{B}=$ $=\overline{A}A+\overline{A}\overline{B}=\overline{A}\overline{B}$, и точно так же доказывается, что и $\overline{B}=\overline{A}\overline{B}$, так что $\overline{A}=\overline{B}$.

Отметим еще два важных свойства противоположных событий. Во-первых, $\overline{\overline{A}}=A$, т. е. ненаступление ненаступления события A есть то же, что его наступление. Во-вторых, $\overline{AB}=\overline{A}+\overline{B}$ и $\overline{A+B}=$ $=\overline{A}\overline{B}$, т. е., с одной стороны, мы имеем ненаступление обоих событий A и B либо тогда, когда не происходит A, либо тогда, когда не происходит B; и, с другой стороны, мы имеем ненаступление даже по крайней мере одного из событий A или B, когда не происходит ни A ни B. То же справедливо и для любого числа событий, так что вообще

$$\overline{AB\ldots S}=\overline{A}+\overline{B}+\ldots+\overline{S},$$

$$\overline{A+B+\ldots+S}=\overline{A}\overline{B}\ldots\overline{S}.$$

Наконец, $\overline{U}=V$ и $\overline{V}=U$: ненаступление достоверного события невозможно, ненаступление невозможного достоверно.

Все эти свойства очевидны. Впрочем, их нетрудно было бы и доказать, опираясь на соотношения 12° и 13°.

4. После всего сказанного мы можем ввести одно понятие, чрезвычайно важное для теории вероятностей, а именно — понятие подразделения события на частные случаи.

Пусть нам даны два события A и B. Составим их произведение AB и их сумму $A+B$. Мы будем различать два случая: случай, когда произведение AB будет возможным событием, и случай, когда оно будет невозможным событием.

Так, если взять пример двукратного бросания игральной кости и обозначить, как мы это делали выше, через A выпадение одного очка при первом бросании и через B—выпадение одного очка при втором бросании, то произведением AB будет выпадение одного очка при обоих бросаниях; это—возможное событие. Иначе говорят: здесь события A и B совместимы. Сумма $A+B$, по определению состоящая в наступлении хотя бы одного из событий A или B, здесь состоит в наступлении либо только события A, либо только события B, либо их обоих. Если же взять пример однократного бросания игральной кости и обозначить через B выпадение трех очков, а через A—выпадение шести очков, то произведением AB будет выпадение и трех и шести очков при одном и том же бросании одной и той же игральной кости, что, очевидно, невозможно. Иначе говоря, здесь события A и B *несовместимы*. Сумма $A+B$ здесь состоит в наступлении либо только события A, либо только события B, но не их обоих, поскольку это последнее здесь невозможно.

В этом последнем случае мы и будем говорить, что событие $A+B$ *подразделяется* на частные случаи A и B.

Возьмем еще раз газ, заключенный в сосуде, и будем следить за движением какой-нибудь определенной молекулы этого газа. Выделим мысленно два объема $\mathfrak{A}$ и $\mathfrak{B}$ внутри сосуда и обозначим через $\mathfrak{C}$ объем, образованный соединением друг с другом объемов $\mathfrak{A}$ и $\mathfrak{B}$. Обозначим через A и B соответственно попадание молекулы в объем $\mathfrak{A}$ и в объем $\mathfrak{B}$. Как мы видели, событием AB будет попадание молекулы в общую часть объемов $\mathfrak{A}$ и $\mathfrak{B}$, а событием $A+B$ попадание молекулы в объем $\mathfrak{C}$. Если объемы $\mathfrak{A}$ и $\mathfrak{B}$ частично перекрываются друг с другом, то событие AB возможно, т. е. события A и B совместимы. В этом случае событие $A+B$ состоит в попадании молекулы либо в объем $\mathfrak{A}$, но не в $\mathfrak{B}$, либо в объем $\mathfrak{B}$, но не в $\mathfrak{A}$, либо в общую часть объемов $\mathfrak{A}$ и $\mathfrak{B}$. Если же объемы $\mathfrak{A}$ и $\mathfrak{B}$ друг с другом не перекрываются, то событие AB, очевидно, невозможно, т. е. события A и B несовместимы. В этом случае событие $A+B$ состоит в попадании молекулы либо в объем $\mathfrak{A}$, но не в $\mathfrak{B}$, либо в объем $\mathfrak{B}$, но не в $\mathfrak{A}$. Здесь попадание молекулы в объем $\mathfrak{C}$ подразделяется на частные случаи: попадание ее в объем $\mathfrak{A}$ и попадание в объем $\mathfrak{B}$.

Вообще пусть нам дано несколько событий $A, B, \ldots, S$. Если каждое из них несовместимо с каждым другим, то мы будем говорить, что событие $A+B+\ldots+S$ подразделяется на частные случаи $A, B, \ldots, S$.

5. Особо важен тот случай группы событий $A, B, \ldots, S$, когда хотя бы одно из этих событий с необходимостью должно произойти. Иначе говоря—тот случай, когда сумма $A+B+\ldots+S$ есть достоверное событие. В этом случае мы говорим, что данная группа событий—*полная*. Возьмем, например, группу из шести событий, могущих произойти при однократном бросании игральной кости: выпадение одного очка, двух очков, трех очков, четырех очков, пяти очков и шести очков; это—полная группа событий.

В частности, полная группа несовместимых между собой событий—это группа событий $A, B, \ldots, S$, на которые подразделяется достоверное событие $A+B+\ldots+S$. Такова как раз группа в только что приведенном примере.

6. Ни в одной задаче нам, конечно, не придется иметь дело со всеми событиями, какие вообще возможны. В каждом отдельном случае мы будем рассматривать то или иное множество событий, рассмотрение которых необходимо для решения данной задачи.

При этом в зависимости от условий задачи одно и то же конкретное событие A может фигурировать в одной задаче как достоверное, в другой—нет; и одни и те же конкретные события A и B могут оказаться в одной задаче противоположными друг другу, в другой—нет. Например, относительно бросания двух игральных костей одной за другой можно поставить среди многих других следующие задачи: 1) задачу о сумме числа очков, которые выпадут на обеих костях—в предположении, что ни одна из костей еще не брошена; 2) задачу о той же сумме—в предположении, что одна из костей уже брошена и на ней выпало одно очко. В задаче 2) достоверным событием будет появление суммы, равной одному из чисел 2, 3, 4, 5, 6 или 7; и потому, например, появление суммы, равной одному из чисел 2, 3 или 4, будет событием, противоположным появлению суммы, равной одному из чисел 5, 6 или 7. В задаче же 1) появление суммы, равной одному из чисел 2, 3, 4, 5, 6 или 7, будет только возможным событием, но не достоверным, поскольку здесь будут возможны и бо́льшие значения суммы; и появление суммы, равной одному из чисел 2, 3 или 4, отнюдь не будет событием, противоположным появлению суммы, равной одному из чисел 5, 6 или 7.

Однако в каждой задаче нам во всяком случае будет необходимо иметь так называемое поле событий, т. е. такой запас событий, который обеспечивал бы возможность образования произведений, сумм и противоположных событий для всех имеющихся в этом запасе событий.

Точнее, *полем событий* называется множество S событий A, B, $C, \ldots$, удовлетворяющее следующим условиям:

1. Для каждых двух событий A и B, принадлежащих полю S, определено, влечет ли событие A за собой событие B, $A \subset B$ или нет.

2. Если полю S принадлежат события A и B, то ему принадлежит и их произведение AB.

3. Если полю S принадлежат события A и B, то ему принадлежит и их сумма $A+B$.

4. Полю S принадлежат достоверное и невозможное события.

5. Если полю S принадлежит событие A, то ему принадлежит и противоположное этому событию событие $\overline{A}$.

§ 2. Классическое определение вероятности.

7. Относительно двух или нескольких событий часто говорят, что они *равновероятны* между собой (или *равновозможны*, что то же самое). Так, например, при бросании идеально правильной игральной кости, имеющей точную форму куба, с центром тяжести в центре куба, будут равновероятными выпадения любого определенного числа очков, 1, 2, 3, 4, 5 или 6, поскольку здесь ни у одного числа очков объективно нет никакого преимущества перед другими.

Понятие равновероятности общепонятно и не нуждается в логическом определении, подобно, например, понятию прямой реального пространства. Мы говорим о равновероятности двух или нескольких событий, если имеем на то достаточные объективные основания, опирающиеся на данные практики, подобно тому, как мы говорим хотя бы о прямолинейности светового луча. Иное дело—более сложное понятие вероятности. Эта последняя есть прежде всего число. Каждое данное значение вероятности (например 0,1 или 0,5) есть числовая характеристика общего свойства всех событий, равновероятных некоторому данному событию,—характеристика, определяемая по известным правилам. Сейчас мы увидим — как именно.

8. Для определения вероятности $\mathbf{P}(A)$ [1]) события A прежде всего надо, чтобы была задана **полная** группа событий $E_1, E_2, \ldots, E_n$, **несовместимых и равновероятных** между собой.

Такая полная группа несовместимых и равновероятных между собой событий существует, например, при бросании игральной кости. Ее составляют выпадения определенного числа 1, 2, 3, 4, 5 или 6 очков.

Мы будем рассматривать события A, которые подразделяются на частные случаи $E'_1, E'_2, \ldots, E'_m$, входящие в состав полной группы несовместимых и равновероятных между собой событий $E_1, E_2, \ldots, E_n$; сюда мы причислим и невозможное событие, считая его событием, подразделяющимся на 0 частных случаев. Такого рода события A мы будем называть *допустимыми*.

Например, в случае бросания игральной кости одним из допустимых событий будет выпадение числа очков, кратного трем. Действительно, оно подразделяется на два частных случая: выпадения трех или шести очков. Одним из допустимых событий будет также выпадение нечетного числа очков. Оно подразделяется на три частных случая: выпадения одного, трех или пяти очков.

Множество всех допустимых событий есть поле событий. В самом деле, пусть A—допустимое событие, подразделяющееся на частные случаи $E'_1, E'_2, \ldots, E'_r$ и B—допустимое событие, подразделяющееся на частные случаи $E''_1, E''_2, \ldots, E''_s$. Тогда $A \subset B$ в том и только в том случае, если все события E'_i входят в число событий E''_j; событием AB будет сумма всех тех событий, которые являются одновременно событиями E'_i и E''_j; событием $A+B$—сумма всех событий E'_i и всех событий E''_j, событием $\overline{A}$—сумма всех тех событий E_k, которые не являются событиями E'_i. Мы видим, таким образом, что каждое из событий AB, $A+B$ и $\overline{A}$, так же как и события A и B, подразделяются на несколько частных случаев, входящих в состав группы событий E_k, и следовательно, каждое из них принадлежит множеству допустимых событий.

Для поля допустимых событий A и дается классическое определение их *вероятности* $\mathbf{P}(A)$. А именно, *когда событие A подразделяется на m частных случаев, входящих в состав полной группы из n несовместимых и равновероятных между собой событий, то полагают по определению*

$$\mathbf{P}(A) = \frac{m}{n}.$$

Например, вероятность выпадения числа очков, кратного трем, равна $\frac{2}{6} = \frac{1}{3}$. Вероятность выпадения нечетного числа очков равна $\frac{3}{6} = \frac{1}{2}$.

[1]) $\mathbf{P}$—Probabilitas (лат).

9. В теории вероятностей широко пользуются также следующей терминологией, к которой мы будем не раз прибегать в дальнейшем. Представляют себе, что для выяснения вопроса, произойдет или нет некоторое событие A (например, выпадение числа очков, кратного трем), надо сделать некоторое испытание, которое дало бы ответ на этот вопрос (в нашем примере надо бросать игральную кость). Полная группа несовместимых и равновероятных между собой событий, которые могут произойти при таком испытании, называется группой *возможных результатов испытания*. Те же из возможных результатов испытания, на которые подразделяется событие A, называются *результатами испытания, благоприятствующими* A.

Пользуясь этой терминологией, можно сказать, что *вероятность* $\boldsymbol{P}(A)$ *события* A *равна отношению числа результатов испытания, благоприятствующих событию* A, *к числу всех возможных результатов испытания.*

10. Из приведенного определения вероятности можно немедленно сделать ряд важных следствий.

Мы имеем шесть основных теорем о вероятностях. Мы приведем их доказательства, пользуясь терминологией, изложенной в п. 9.

Теорема I. — *Если событие* A *влечет за собой событие* B, *то вероятность* A *не превосходит вероятности* B:

$$\boldsymbol{P}(A) \leqslant \boldsymbol{P}(B).$$

Доказательство. — Пусть событиям A и B благоприятствуют соответственно m_1 и m_2 результатов испытания, так что

$$\boldsymbol{P}(A) = \frac{m_1}{n} \quad \text{и} \quad \boldsymbol{P}(B) = \frac{m_2}{n}.$$

Так как по условию A влечет за собой B, то, очевидно, каждый результат испытания, благоприятствующий A, благоприятствует также и B. Иначе говоря, все m_1 результатов испытания, благоприятствующих A, входят в число m_2 результатов испытания, благоприятствующих B. Отсюда $m_1 \leqslant m_2$, следовательно,

$$\frac{m_1}{n} \leqslant \frac{m_2}{n},$$

или, что то же, $\boldsymbol{P}(A) \leqslant \boldsymbol{P}(B)$.

Теорема II. — *Если события* A *и* B *равносильны, то их вероятности равны:*

$$\boldsymbol{P}(A) = \boldsymbol{P}(B).$$

Доказательство. — Эта теорема является следствием теоремы I. Действительно, если события A и B равносильны, то A влечет за собой B, а B влечет за собой A. Следовательно, в силу теоремы I

$$\boldsymbol{P}(A) \leqslant \boldsymbol{P}(B) \quad \text{и} \quad \boldsymbol{P}(B) \leqslant \boldsymbol{P}(A),$$

откуда и вытекает требуемое равенство $\boldsymbol{P}(A) = \boldsymbol{P}(B)$.

Теорема III (теорема сложения). — *Если событие* A *под-*

разделяется на частные случаи B и C, то вероятность A складывается из вероятностей B и C:

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

Доказательство. — По условию событие A равносильно событию $B+C$. Поэтому, в силу теоремы II,

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

Остается доказать, что

$$\boldsymbol{P}(B+C)=\boldsymbol{P}(B)+\boldsymbol{P}(C).$$

Пусть событиям B и C благоприятствуют соответственно m_1 и m_2 результатов испытания, так что

$$\boldsymbol{P}(B)=\frac{m_1}{n} \quad \text{и} \quad \boldsymbol{P}(C)=\frac{m_2}{n}.$$

Но по условию события B и C несовместимы. Поэтому результаты испытания, благоприятствующие одному из них, не могут благоприятствовать другому. Это обстоятельство дает возможность вычислить, сколько результатов испытания благоприятствуют событию $B+C$, т. е. наступлению хотя бы одного из событий B или C. Для этого, очевидно, достаточно взять m_1 результатов испытания, благоприятствующих событию B, и прибавить к ним m_2 отличных от них результатов испытания, благоприятствующих событию C. Итак, число результатов испытания, благоприятствующих событию $B+C$, есть m_1+m_2, так что

$$\boldsymbol{P}(B+C)=\frac{m_1+m_2}{n}.$$

Сопоставляя выражения для вероятностей $\boldsymbol{P}(B)$, $\boldsymbol{P}(C)$ и $\boldsymbol{P}(B+C)$, мы видим, что $\boldsymbol{P}(B+C)=\boldsymbol{P}(B)+\boldsymbol{P}(C)$.

Теорема, таким образом, доказана.

Обобщение теоремы III. — *Если событие A подразделяется на частные случаи $A_1, A_2, \ldots, A_s$, то вероятность A складывается из вероятностей $A_1, A_2, \ldots, A_s$:*

$$\boldsymbol{P}(A)=\boldsymbol{P}(A_1)+\boldsymbol{P}(A_2)+\ldots+\boldsymbol{P}(A_s).$$

Доказательство. — В силу теоремы III это верно для двух слагаемых. Докажем теперь, что если это верно для $s-1$, то верно и для s слагаемых. Итак, пусть $A'=A_1+A_2+\ldots+A_{s-1}$ и пусть

$$\boldsymbol{P}(A')=\boldsymbol{P}(A_1)+\boldsymbol{P}(A_2)+\ldots+\boldsymbol{P}(A_{s-1}).$$

Тогда, прежде всего $A=A'+A_s$. Затем события A' и A_s несовместимы, так как $A_sA'=A_s(A_1+A_2+\ldots+A_{s-1})=A_sA_1+A_sA_2+ +\ldots+A_sA_{s-1}$, по условию же события $A_sA_1, A_sA_2, \ldots, A_sA_{s-1}$ невозможны, следовательно, наступление хотя бы одного из них, очевидно, также невозможно. Таким образом, событие A подразделяется на частные случаи A' и A_s. Поэтому, в силу теоремы III,

$$\boldsymbol{P}(A)=\boldsymbol{P}(A')+\boldsymbol{P}(A_s).$$

Сопоставляя это с вышеприведенным выражением для $\boldsymbol{P}(A')$, мы и получаем требуемое равенство

$$\boldsymbol{P}(A)=\boldsymbol{P}(A_1)+\boldsymbol{P}(A_2)+\ldots+\boldsymbol{P}(A_s).$$

Теорема IV.—*Вероятность достоверного события равна единице.*

Доказательство.—Достоверному событию A благоприятствуют все n возможных результатов испытания. Поэтому

$$\boldsymbol{P}(A)=\frac{n}{n}=1.$$

Теорема V.—*Вероятность события, противоположного событию A, равна единице минус вероятность A:*

$$\boldsymbol{P}(\overline{A})=1-\boldsymbol{P}(A).$$

Доказательство.—Эта теорема является следствием теорем III и IV. Действительно, рассмотрим событие $A+\overline{A}$. Так как события A и $\overline{A}$ несовместимы, то событие $A+\overline{A}$ подразделяется на частные случаи A и $\overline{A}$. Следовательно, в силу теоремы III,

$$\boldsymbol{P}(A+\overline{A})=\boldsymbol{P}(A)+\boldsymbol{P}(\overline{A}).$$

С другой стороны, событие $A+\overline{A}$ достоверно. Следовательно, в силу теоремы IV,

$$\boldsymbol{P}(A+\overline{A})=1.$$

Сопоставляя написанные равенства, мы получаем $\boldsymbol{P}(A)+\boldsymbol{P}(\overline{A})=1$, или, что то же, $\boldsymbol{P}(\overline{A})=1-\boldsymbol{P}(A)$.

Теорема VI.—*Вероятность невозможного события равна нулю.*

Доказательство.—Эта теорема может рассматриваться как следствие теорем IV и V. Действительно, невозможность события A означает достоверность его ненаступления, т. е. достоверность противоположного ему события $\overline{A}$. Следовательно, в силу теоремы IV,

$$\boldsymbol{P}(\overline{A})=1.$$

С другой стороны, в силу теоремы V

$$\boldsymbol{P}(\overline{A})=1-\boldsymbol{P}(A).$$

Сопоставляя написанные равенства, мы получаем $1=1-\boldsymbol{P}(A)$, или, что то же, $\boldsymbol{P}(A)=0$.

Отметим еще, что *для любого события A*

$$0\leqslant\boldsymbol{P}(A)\leqslant 1.$$

В самом деле, пусть U — достоверное событие и V — невозможное событие; тогда

$$V\subset A+V=A=AU\subset U,$$

и следовательно,

$$0=\boldsymbol{P}(V)\leqslant\boldsymbol{P}(A)\leqslant\boldsymbol{P}(U)=1.$$

§ 3. Условные вероятности.

11. Пусть попрежнему задана полная группа событий $E_1, E_2, \ldots, E_n$, несовместимых и равновероятных между собой, и на основании ее установлены вероятности в поле допустимых событий A, B и т. д. Часто возникает вопрос, как должны измениться эти вероятности, если предположить, что станет достоверным наступление одного из допустимых событий, скажем A.

Будем, например, бросать две игральные кости одну за другой. Полная группа событий $E_1, E_2, \ldots, E_n$, несовместимых и равновероятных между собой, будет здесь образованным из 36 событий, состоящих в выпадении определенного числа 1, 2, 3, 4, 5 или 6 очков на первой кости и определенного числа 1, 2, 3, 4, 5 или 6 очков на второй кости, причем те и другие могут сочетаться друг с другом как угодно (см. табл. I).

Таблица I.

1,1	1,2*	1,3	1,4	1,5	1,6
2,1*	2,2	2,3	2,4	2,5	2,6
3,1	3,2	3,3	3,4	3,5	3,6
4,1	4,2	4,3	4,4	4,5	4,6
5,1	5,2	5,3	5,4	5,5	5,6
6,1	6,2	6,3	6,4	6,5	6,6

Первое число в каждой клетке этой таблицы есть число очков на первой кости, второе — на второй кости. Среди многих допустимых здесь событий отметим два. Во-первых, событие, которое мы обозначим через A, состоящее в том, что на первой кости выпадает одно очко. Этому благоприятствует вся первая строка нашей таблицы, т. е. шесть клеток, так что

$$P(A) = \frac{6}{36} = \frac{1}{6}.$$

Во-вторых, отметим событие, которое мы обозначим через B, состоящее в том, что на обеих костях выпадет число очков, в сумме равное трем. Этому благоприятствуют две отмеченные звездочками клетки нашей таблицы, так что

$$P(B) = \frac{2}{36} = \frac{1}{18}.$$

Поставим теперь вопрос следующим образом. Представим себе, что мы начали бросать кости, а именно — бросили первую из них, и получили на ней одно очко. Как изменится после этого вероятность того, что на обеих костях выпадет число очков, в сумме равное трем. Иначе говоря, пусть стало достоверно событие A. Как изменится после этого вероятность события B?

Вопросы такого рода постоянно возникают в теории вероятностей и в ее приложениях. Они требуют для своего разрешения нового понятия условной вероятности.

Предположение, что стало достоверным событие A, нарушает заданную полную группу несовместимых между собой событий $E_1, E_2, \ldots, E_n$ в том смысле, что благодаря этому предположению все события заданной группы уже не могут больше считаться равновероятными. А именно, предположение, что стало достоверным событие A, означает, что из всех событий $E_1, E_2, \ldots, E_n$ остаются возможными только те, на которые подразделяется A, остальные же становятся невозможными. Так, в нашем примере после выпадения на первой кости одного очка остается возможной только первая строка табл. I. При этом подразумевается, что те из событий $E_1, E_2, \ldots, E_n$, которые остаются возможными, остаются и равновероятными между собой. Именно так и обстоит дело в нашем примере.

Описанное нарушение заданной полной группы несовместимых и равновероятных между собой событий $E_1, E_2, \ldots, E_n$ сводится в сущности к замене ее другой полной группой несовместимых и равновероятных между собой событий, где роль достоверного события играет A и где фигурируют только те из событий $E_1, E_2, \ldots, E_n$, на которые подразделяется A. В нашем примере эта новая группа изобразится табл. II, состоящей из первой строки табл. I.

Таблица II.

1,1	1,2*	1,3	1,4	1,5	1,6

В этой новой группе в качестве частных случаев любого события B сохранятся, разумеется, уже не все те из событий $E_1, E_2, \ldots, E_n$, на которые подразделялось B, а только те из них, которые остались возможными. В нашем примере в качестве частных случаев события B сохраняются не обе клетки, отмеченные звездочками в табл. I, а только одна — та, которая осталась в табл. II. Иначе можно сказать, что в новой группе в качестве частных случаев любого события B сохраняются только те из событий $E_1, E_2, \ldots, E_n$, которые являлись частными случаями и события B и события A, т. е. те, которые являлись частными случаями события AB. В нашем примере сохраняется клетка, отмеченная звездочкой и находившаяся в первой строке табл. I; здесь это и есть событие AB, т. е. выпадение на первой кости одного очка и на обеих костях — числа очков, в сумме равного трем.

Отсюда понятно, каким образом следует ответить на вопрос, как должна измениться вероятность $\boldsymbol{P}(B)$ любого события B, если предположить, что станет достоверным событие A. Эта вероятность должна замениться *условной вероятностью* $\boldsymbol{P}(B/A)$ события B в предположении, что произошло событие A, которая определяется следующим образом: *когда событие A подразделяется на $m>0$ событий из полной группы несовместимых и равновероятных между собой событий, а событие AB подразделяется на l событий из той же группы, то*

$$\boldsymbol{P}(B/A)=\frac{l}{m}.$$

Для $m=0$ условная вероятность $\boldsymbol{P}(B/A)$ не определена.

В нашем примере событию A соответствует шесть клеток ($m=6$), а событию AB — одна ($l=1$). Поэтому в нашем примере

$$\boldsymbol{P}(B/A)=\frac{1}{6}.$$

Разумеется и здесь можно воспользоваться той терминологией, которой мы пользовались, начиная с п. 9. А именно можно сказать, что *условная вероятность* $\boldsymbol{P}(B/A)$ *события* B *в предположении, что произошло событие* A, *равна отношению числа результатов испытания, благоприятствующих событию* AB, *к числу результатов испытания, благоприятствующих событию* A.

12. На условные вероятности распространяются все предложения, установленные нами в п. 10 для безусловных вероятностей. В предположении, что произошло некоторое определенное событие A, всегда справедливы неравенства

$$0\leqslant\boldsymbol{P}(B/A)\leqslant 1$$

и всегда верны следующие теоремы:

Теорема I*. — *Если событие* B *влечет за собой событие* C, *то условная вероятность* B *не превосходит условной вероятности* C:

$$\boldsymbol{P}(B/A)\leqslant\boldsymbol{P}(C/A).$$

Теорема II*. — *Если события* B *и* C *равносильны, то их условные вероятности равны:*

$$\boldsymbol{P}(B/A)=\boldsymbol{P}(C/A).$$

Теорема III* (теорема сложения). — *Если событие* B *подразделяется на частные случаи* $B_1, B_2, \ldots, B_s$, *то условная вероятность* B *складывается из условных вероятностей* $B_1, B_2, \ldots, B_s$:

$$\boldsymbol{P}(B/A)=\boldsymbol{P}(B_1/A)+\boldsymbol{P}(B_2/A)+\ldots+\boldsymbol{P}(B_s/A).$$

Теорема IV*. — *Условная вероятность достоверного события равна единице.*

Теорема V*. — *Условная вероятность события, противоположного событию* B, *равна единице минус условная вероятность* B:

$$\boldsymbol{P}(\overline{B}/A)=1-\boldsymbol{P}(B/A).$$

Теорема VI*. — *Условная вероятность невозможного события равна нулю.*

Все эти предложения доказываются по существу так же, как и предложения, установленные в п. 10. Некоторые неизбежные здесь усложнения не представляют собой больших трудностей, и потому мы представляем доказательства самому читателю.

13. Условные вероятности могут быть прямо выражены через безусловные, на основании следующей теоремы.

Теорема VII (теорема умножения). — *Вероятность наступления обоих событий A и B равна произведению вероятности A на условную вероятность B в предположении, что произошло событие A:*

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

Доказательство. — Пусть событию A благоприятствует m результатов испытания, а событию AB благоприятствует l результатов испытания из общего числа n возможных результатов испытания. Тогда, в силу самого определения вероятностей $\boldsymbol{P}(AB)$ и $\boldsymbol{P}(A)$ и условной вероятности $\boldsymbol{P}(B/A)$,

$$\boldsymbol{P}(AB) = \frac{l}{n}, \quad \boldsymbol{P}(A) = \frac{m}{n}, \quad \boldsymbol{P}(B/A) = \frac{l}{m},$$

откуда и вытекает требуемое равенство $\boldsymbol{P}(AB) = \boldsymbol{P}(A)\,\boldsymbol{P}(B/A)$.

В силу этой теоремы, условная вероятность может быть выражена через безусловные следующим образом:

$$\boldsymbol{P}(B/A) = \frac{\boldsymbol{P}(AB)}{\boldsymbol{P}(A)}. \tag{1}$$

Обобщение теоремы VII. — *Вероятность наступления всех событий $A_1, A_2, \ldots, A_s$ может быть выражена формулой*

$$\boldsymbol{P}(A_1 A_2 \ldots A_s) = \boldsymbol{P}(A_1)\,\boldsymbol{P}(A_2/A_1)\,\boldsymbol{P}(A_3/A_1 A_2)\ldots\boldsymbol{P}(A_s/A_1 A_2 \ldots A_{s-1}).$$

Доказательство. — В силу теоремы VII это верно для двух множителей. Докажем теперь, что если это верно для $s-1$, то верно и для s множителей. Итак, пусть $A' = A_1 A_2 \ldots A_{s-1}$ и пусть

$$\boldsymbol{P}(A') = \boldsymbol{P}(A_1)\,\boldsymbol{P}(A_2/A_1)\,\boldsymbol{P}(A_3/A_1 A_2)\ldots\boldsymbol{P}(A_{s-1}/A_1 A_2 \ldots A_{s-2}).$$

Но $A = A' A_s$. Поэтому, в силу теоремы VII,

$$\boldsymbol{P}(A) = \boldsymbol{P}(A')\,\boldsymbol{P}(A_s/A').$$

Сопоставляя это с только что приведенными выражениями для A' и для $\boldsymbol{P}(A')$, мы и получаем требуемую формулу.

Из теоремы VII может быть выведена также другая теорема, которая дает возможность выражать условную вероятность $\boldsymbol{P}(A/B)$ через условную вероятность $\boldsymbol{P}(B/A)$.

Теорема VIII. — *Произведение вероятности A на условную вероятность B в предположении, что произошло A, равно произведению вероятности B на условную вероятность A в предположении, что произошло B:*

$$\boldsymbol{P}(A)\,\boldsymbol{P}(B/A) = \boldsymbol{P}(B)\,\boldsymbol{P}(A/B).$$

Доказательство. — $AB = BA$. Поэтому, в силу теорем VII и II,

$$\boldsymbol{P}(A)\,\boldsymbol{P}(B/A) = \boldsymbol{P}(AB) = \boldsymbol{P}(BA) = \boldsymbol{P}(B)\,\boldsymbol{P}(A/B).$$

В силу этой теоремы условная вероятность $P(A/B)$ может быть выражена через условную вероятность $P(B/A)$ следующим образом:

$$P(A/B)=\frac{P(A)\,P(B/A)}{P(B)}. \tag{2}$$

14. События A и B называются *независимыми*: 1) в случае если предположение, что произошло одно из них, не изменяет вероятности другого:

$$P(B/A)=P(B) \quad \text{и} \quad P(A/B)=P(A);$$

и 2) в случае, если вероятность по крайней мере одного из них равна нулю:

$$P(A)=0 \quad \text{или} \quad P(B)=0.$$

Заметим тут же, что в случае 1) из двух написанных равенств одно всегда является следствием другого, так что для установления независимости событий A и B всегда достаточно установить либо первое, либо второе из этих равенств. В самом деле, по теореме VIII

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

Таким образом если предположить, например, что имеет место равенство $P(B/A)=P(B)$, то мы получим

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

или, что то же, $P(A/B)=P(A)$; так же доказывается и обратное.

Для примера возьмем однократное бросание игральной кости и обозначим через A выпадение числа очков, не превосходящего двух, и через B — выпадение четного числа очков. Выпадению числа очков, не превосходящего двух (событие A), благоприятствуют два возможных результата испытания (выпадение 1 или 2 очков), выпадению же четного числа очков, не превосходящего двух (событие AB), благоприятствует один результат испытания (выпадение 2 очков); поэтому $P(B/A)=\frac{1}{2}$. С другой стороны, выпадению четного числа очков (событие B) благоприятствует половина всех возможных результатов испытания (выпадение 2, 4 или 6 очков); поэтому $P(B)=\frac{1}{2}$. Таким образом $P(B/A)=P(B)$. События A и B независимы.

Но если бы мы обозначили, например, через A выпадение числа очков, не превосходящего трех, а через B — попрежнему выпадение четного числа очков, то мы имели бы $P(B/A)=\frac{1}{3}$, тогда как $P(B)=\frac{1}{2}$. События A и B не были бы независимы.

Общее, несколько событий $A_1, A_2, \ldots, A_s$ называются *независимыми в их совокупности*, если, каково бы ни было событие A_q из их числа и каковы бы ни были другие события $A_i, A_j, \ldots, A_p$ из их же числа, событие A_q и произведение событий $A_iA_j\ldots A_p$ независимы между собой.

Заметим, что если события $A_1, A_2, \ldots, A_s$ независимы попарно, т. е. если каждые два из них независимы между собой, то отсюда еще отнюдь не следует, что они независимы в их совокупности. Пусть, например, нам дана совокупность предметов, из которых одна четверть обладает тремя признаками X, Y и Z, одна четверть — только признаком X, одна четверть — только признаком Y и одна четверть — только признаком Z. Будем выбирать наугад один предмет из этой совокупности. Обозначим через A, B и C соответственно события, состоящие в том, что выбранный предмет окажется обладающим признаком X, Y или Z. Нетрудно подсчитать, что $\boldsymbol{P}(A/B) = \boldsymbol{P}(A)$, $\boldsymbol{P}(B/C) = \boldsymbol{P}(B)$ и $\boldsymbol{P}(C/A) = \boldsymbol{P}(C)$ (все эти вероятности равны половине), так что события A, B и C попарно независимы. Однако, например, $\boldsymbol{P}(A/BC) \neq \boldsymbol{P}(A)$ (первая из этих вероятностей равна единице), так что события A, B и C не независимы в их совокупности.

15. Для независимых событий теорема умножения получает более простую и полную формулировку.

Теорема VII bis. — *Для того чтобы события A и B были независимы, необходимо и достаточно, чтобы вероятность наступления их обоих была равна произведению вероятностей каждого из них в отдельности:*

$$\boldsymbol{P}(AB) = \boldsymbol{P}(A)\,\boldsymbol{P}(B). \tag{3}$$

Доказательство. — Когда $\boldsymbol{P}(A) \neq 0$ и $\boldsymbol{P}(B) \neq 0$, то события A и B независимы, если

$$\boldsymbol{P}(B/A) = \boldsymbol{P}(B). \tag{4}$$

В силу же теоремы VII,

$$\boldsymbol{P}(AB) = \boldsymbol{P}(A)\,\boldsymbol{P}(B/A). \tag{5}$$

Подставляя (4) в (5), мы и получаем требуемое равенство (3). Если же события A и B независимы в силу того, что $\boldsymbol{P}(A) = 0$ или $\boldsymbol{P}(B) = 0$, то, поскольку $AB \subset A$ и $AB \subset B$, постольку и $\boldsymbol{P}(AB) = 0$, и мы опять получаем равенство (3).

Обратно, пусть имеет место это последнее. В случае, когда обе его части не равны 0, очевидно, также $\boldsymbol{P}(A) \neq 0$ и $\boldsymbol{P}(B) \neq 0$. Тогда, сравнивая (3) с (5), мы получаем (4), т. е. независимость событий A и B. В случае же, когда обе части равенства (3) равны 0, мы имеем $\boldsymbol{P}(A) = 0$ или $\boldsymbol{P}(B) = 0$, т. е. опять независимость событий A и B.

Обобщение теоремы VII bis. — *Для того чтобы события $A_1, A_2, \ldots, A_s$ были независимы в их совокупности, необходимо и достаточно, чтобы, какова бы ни была группа $A_i, A_j, \ldots, A_p$, составленная из этих событий, имело место равенство*

$$\boldsymbol{P}(A_iA_j \ldots A_p) = \boldsymbol{P}(A_i)\,\boldsymbol{P}(A_j) \ldots \boldsymbol{P}(A_p). \tag{6}$$

Доказательство. — В силу теоремы VII bis равенство (6) верно для двух независимых событий. Докажем теперь, что если оно верно для $s-1$, то верно и для s независимых событий. Итак, предположим, что для независимых событий $A_1, A_2, \ldots, A_{s-1}$ всегда верно, что, какова бы ни была группа $A_i, A_j, \ldots, A_p$, составленная из этих событий, для нее имеет место равенство (6). Пусть даны независимые события $A_1, A_2, \ldots, A_s$. Тогда, очевидно, независимы и события $A_1, A_2, \ldots, A_{s-1}$. Следовательно, в силу нашего предположения, для любой группы $A_i, A_j, \ldots, A_p$, составленной из этих последних, верно (6). Любую же группу, составленную из событий $A_1, A_2, \ldots, A_s$, можно получить, либо взяв просто одну из названных групп $A_i, A_j, \ldots, A_p$, либо взяв ту же группу и присоединив к ней событие A_s. В этом последнем случае в силу определения независимости событий $A_1, A_2, \ldots, A_s$ и теоремы VII bis

$$\boldsymbol{P}(A_iA_j \ldots A_pA_s) = \boldsymbol{P}(A_iA_j \ldots A_p)\,\boldsymbol{P}(A_s). \tag{7}$$

Подставляя (6) в (7), мы получаем требуемое равенство:

$$\boldsymbol{P}(A_iA_j \ldots A_pA_s) = \boldsymbol{P}(A_i)\,\boldsymbol{P}(A_j) \ldots \boldsymbol{P}(A_p)\,\boldsymbol{P}(A_s).$$

Обратно, пусть, какова бы ни была группа $A_i, A_j, \ldots, A_p$, составленная из событий $A_1, A_2, \ldots, A_s$, верно (6). Тогда для группы $A_i, A_j, \ldots, A_p, A_q$, составленной из событий $A_1, A_2, \ldots, A_s$, будет иметь место равенство

$$\boldsymbol{P}(A_iA_j \ldots A_pA_q) = \boldsymbol{P}(A_i)\,\boldsymbol{P}(A_j) \ldots \boldsymbol{P}(A_p)\,\boldsymbol{P}(A_q). \tag{8}$$

Сопоставляя (6) с (8), мы получаем

$$\boldsymbol{P}(A_iA_j \ldots A_pA_q) = \boldsymbol{P}(A_iA_j \ldots A_p)\,\boldsymbol{P}(A_q).$$

Отсюда, в силу теоремы VII bis и определения независимости событий $A_1, A_2, \ldots, A_s$, и вытекает эта независимость.

16. Необходимо сделать еще одно общее замечание о независимых событиях. Мы имеем для двух независимых событий равенство

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

Естественным обобщением этого равенства для многих независимых событий служат равенства вида

$$\boldsymbol{P}(A_s/A_1A_2 \ldots A_{s-1}) = \boldsymbol{P}(A_s). \tag{9}$$

Далее, мы имеем для двух независимых событий равенство

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

Прямым обобщением этого равенства для многих независимых событий являются равенства вида

$$\boldsymbol{P}(A_1A_2 \ldots A_s) = \boldsymbol{P}(A_1)\,\boldsymbol{P}(A_2) \ldots \boldsymbol{P}(A_s). \tag{10}$$

Таким образом для установления независимости событий, казалось бы, мы обязаны подсчитывать левые части равенств (9) или (10), что сводится к определению вероятностей целого ряда произведений нескольких событий.

Между тем, практически часто поступают как раз наоборот. Справедливость равенств (9), т. е. то обстоятельство, что наступление одних событий не изменяет вероятности других, часто бывает ясно без всяких подсчетов, просто по реальному смыслу поставленной задачи. Отсюда же заключают, в силу доказанных теорем, и о справедливости равенств (10). И тогда как раз эти последние равенства используют для вычисления вероятностей произведений нескольких событий.

Пусть, например, игральная кость бросается четыре раза подряд и мы ищем вероятность того, что при всех бросаниях окажется выпавшим по одному очку. Ясно, что выпадение одного очка при одних бросаниях не изменяет вероятности выпадения одного очка при других бросаниях; эта вероятность всегда остается равной $\frac{1}{6}$. Иначе говоря, ясно без всяких подсчетов, что выпадения одного очка при различных бросаниях представляют собой независимые события в смысле равенств (9). Поэтому здесь применимы и равенства (10). В силу же этих последних искомая вероятность выпадения одного очка при всех бросаниях будет равна

$$\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6}=\left(\frac{1}{6}\right)^4=\frac{1}{1296}.$$

Или пусть, например, игральная кость бросается четыре раза подряд и мы ищем вероятность того, что по крайней мере при одном бросании окажется выпавшим одно очко. Найдем сначала вероятность противоположного события — того, что при всех бросаниях окажется выпавшим более, чем одно очко. Ясно, что выпадение более чем одного очка при одних бросаниях не изменяет вероятности выпадения более чем одного очка при других бросаниях; эта вероятность всегда остается равной

$$1-\frac{1}{6}=\frac{5}{6}.$$

Иначе говоря, ясно без всяких подсчетов, что выпадения более чем одного очка при различных бросаниях представляют собой независимые события в смысле равенств (9). Поэтому здесь применимы и равенства (10). В силу же этих последних вероятность выпадения более чем одного очка при всех бросаниях будет равна

$$\left(1-\frac{1}{6}\right)\left(1-\frac{1}{6}\right)\left(1-\frac{1}{6}\right)\left(1-\frac{1}{6}\right)=\left(1-\frac{1}{6}\right)^4=\frac{625}{1296}.$$

Следовательно, искомая вероятность противоположного этому события — выпадения одного очка по крайней мере при одном бросании — равна

$$1-\left(1-\frac{1}{6}\right)^4=\frac{671}{1296}.$$

§ 4. Аксиоматика теории вероятностей.

17. Классическое определение вероятности неприменимо к событию A, если нет возможности указать полную группу, состоящую из *конечного числа* несовместимых и равновероятных между собой событий, по отношению к которой событие A было бы допустимым событием. Отсюда возникает ряд затруднений в применении теории вероятностей к вычислениям с непрерывными величинами — в физике, в технике, в биологии и т. д.

Мы приведем для примера две задачи, по существу ничем друг от друга не отличающиеся; разница между ними будет только в числовых данных. И, тем не менее, одна из этих задач будет решаться на основе классического определения вероятности, а в другой задаче это определение будет неприменимо.

В ряде наших примеров фигурировала игральная кость; испытание состояло в бросании кости, а результат его — в выпадении грани с тем или иным числом очков. Основная гипотеза о равновероятности, которую мы при этом делали, состояла в том, что мы считали равновероятными выпадения всех граней. В двух же задачах, которые мы сейчас приведем, мы будем брать вместо кости шарик с поверхностью, разделенной на несколько областей; испытание будет состоять в бросании шарика на плоскость, а результат его — в выпадении той или иной области, причем под выпадением области $\mathfrak{A}$ мы подразумеваем тот факт, что точка поверхности шарика, диаметрально противоположная той, которая прикасается к плоскости при падении на нее шарика, оказывается принадлежащей области $\mathfrak{A}$. Основная гипотеза о равновероятности, которую мы сделаем, будет состоять в том, что будут равновероятными выпадения областей, имеющих равные площади. Различие между двумя нашими задачами будет заключаться лишь в том, каким образом мы будем подразделять поверхность шарика на части.

В первой задаче, пользуясь географической терминологией, мы будем подразделять поверхность шарика экватором и двумя параллелями: 30-й северной и 30-й южной. Мы будем обозначать через $\mathfrak{A}_1$, $\mathfrak{B}_1$, $\mathfrak{C}_1$, $\mathfrak{D}_1$ области, рассматриваемые последовательно в направлении от северного полюса к южному.

Во второй задаче мы будем подразделять поверхность шарика тоже экватором и двумя параллелями, но на этот раз 60-й северной и 60-й южной. Мы будем обозначать через $\mathfrak{A}_2$, $\mathfrak{B}_2$, $\mathfrak{C}_2$, $\mathfrak{D}_2$ области, рассматриваемые попрежнему последовательно в направлении от северного полюса к южному.

Первая задача. — *Какова вероятность выпадения области* $\mathfrak{A}_1$? Задача эта решается без труда. Площадь каждой из областей $\mathfrak{A}_1$, $\mathfrak{B}_1$, $\mathfrak{C}_1$, $\mathfrak{D}_1$ равна πr^2, где r — радиус шарика. Поэтому, в силу нашего основного предположения о равновероятности, выпадения четырех областей $\mathfrak{A}_1$, $\mathfrak{B}_1$, $\mathfrak{C}_1$, $\mathfrak{D}_1$ образуют группу несовместимых и равновероятных между собой событий, очевидно, полную. Следовательно, вероятность каждого из них, в частности и выпадения области $\mathfrak{A}_1$, равна $\frac{1}{4}$.

Вторая задача. — *Какова вероятность выпадения области* $\mathfrak{A}_2$? Площадь области $\mathfrak{A}_2$, как и области $\mathfrak{D}_2$, равна $(2-\sqrt{3})\pi r^2$, т. е. составляет $\frac{(2-\sqrt{3})}{4}$ площади всей поверхности шарика. Площадь же области $\mathfrak{B}_2$, как и области $\mathfrak{C}_2$, равна $\sqrt{3}\pi r^2$, т. е. составляет $\frac{\sqrt{3}}{4}$ площади всей поверхности шарика. Так как $\frac{(2-\sqrt{3})}{4}$ и $\frac{\sqrt{3}}{4}$ — числа иррациональные, то, очевидно, поверхность шарика нельзя подразделить на конечное число областей равной площади так, чтобы хотя бы одна из наших областей $\mathfrak{A}_2$, $\mathfrak{B}_2$, $\mathfrak{C}_2$, $\mathfrak{D}_2$ составлялась в точности из некоторого числа этих областей. Таким образом здесь не существует полной группы из конечного числа событий, несовместимых и равновероятных между собой, на которые подразделялось бы событие, состоящее в выпадении хотя бы одной из наших областей, в частности и области $\mathfrak{A}_2$. Иначе говоря, здесь не существует полной группы несовместимых и равновероятных между собой событий, по отношению к которой выпадение области $\mathfrak{A}_2$ было бы допустимым событием. Классическое определение вероятности здесь неприменимо.

18. Необходимое расширение определения вероятности можно произвести при помощи *аксиоматического определения вероятности*.

Пусть дано поле S событий $A, B, C, \ldots$, удовлетворяющее следующим аксиомам:

I. — *Каждому событию A поля S поставлено в соответствие действительное число $P(A)$; это число называется вероятностью события A.*

II. — *Если событие A влечет за собой событие B, то*

$$P(A) \leqslant P(B).$$

III. — (Аксиома сложения). *Если событие A подразделяется на частные случаи B и C, то*

$$P(A) = P(B) + P(C).$$

IV. — *Если событие A достоверно, то*

$$P(A) = 1.$$

Поле S событий $A, B, C, \ldots$, удовлетворяющее аксиомам I IV, вместе с определенными для событий $A, B, C, \ldots$ вероятностями $P(A)$, $P(B)$, $P(C), \ldots$, называется *полем вероятностей* [1]).

Изложенная аксиоматика не является просто придуманным условным соглашением. Она (как и всякая аксиоматика) лишь отражает в форме, пригодной для дальнейших дедуктивных выводов, историческое развитие теории вероятностей. Действительно, существенная часть этой аксиоматики — свойства вероятностей, выраженные аксиомами II, III и IV (аксиома I служит для описания поля применения этих последних), это те свойства, которые в случае классического определения вероятности могут быть доказаны и были доказаны нами, в форме теорем I, III и IV. Отсюда же вытекает, что при данном аксиоматическом определении вероятности остаются в силе и все

[1]) Аксиому II можно заменить эквивалентной ей аксиомой: *Если события A и B равносильны, то* $P(A) = P(B)$.

доказанные нами шесть основных теорем о вероятностях, так как теоремы II, V и VI являются следствиями теорем I, III и IV. В прошлом исследования по теории вероятностей по мере ее развития все чаще и чаще опирались именно на перечисленные теоремы, которые применялись как исходные положения даже в тех случаях, когда, строго говоря, само классическое определение вероятности было уже неприменимо. Так обстояло дело в теории ошибок наблюдений и т. п., причем полученные таким путем выводы полностью оправдывались на практике. Аксиоматика лишь зафиксировала это исторически создавшееся положение в форме, сообщающей ему общее обоснование.

Однозначно ли определяется вероятность, если задать поле событий и потребовать, чтобы вероятности этих событий были определены так, чтобы они удовлетворяли аксиомам I—IV? Легко видеть, что такой однозначности нет. Возьмем хотя бы поле допустимых событий в задаче об однократном бросании игральной кости и определим вероятности этих событий иначе, чем мы это делали выше. А именно, для „основных" событий — выпадений определенного числа 1, 2, 3, 4, 5 или 6 очков — допустим, что вероятности выпадения 1, 2 и 3 очков равны $\frac{1}{4}$, а вероятности выпадения 4, 5 и 6 очков равны $\frac{1}{12}$; для каждого же из остальных допустимых событий A определим вероятность как сумму вероятностей тех „основных" событий, на которые подразделяется A. Легко видеть, что все аксиомы I—IV здесь будут выполнены. Между тем, числовые значения вероятностей здесь не те, что были выше.

Естественно поэтому возникает вопрос, не упущено ли в нашем аксиоматическом определении какое-нибудь общее свойство вероятностей, такое, что требование его соблюдения сделало бы наше определение однозначным. Только что приведенный пример показывает, что это не так. В самом деле, если бы вероятности выпадения того или иного числа очков на игральной кости определялись однозначно, то это значило бы, что те числовые значения вероятностей, которые мы только что допустили в нашем примере, вообще не были бы возможны. Между тем, они вполне возможны, например, тогда, когда игральная кость сделана из неоднородного материала и ее центр тяжести несколько смещен в сторону угла, в котором сходятся грани с одним, двумя и тремя очками. Таким образом однозначности определения вероятности нельзя достичь добавочным требованием о соблюдении еще какого-нибудь о б щ е г о свойства вероятностей без того, чтобы теория вероятностей не потеряла ряда своих возможных приложений. Однозначность определения вероятности можно получить только для каждой отдельной задачи, если прибавить к уже названным общим свойствам вероятностей какое-нибудь с п е ц и ф и ч е с к о е для данной задачи допущение. При классическом определении вероятностей таким допущением является предположение о равновероятности событий, принадлежащих той или иной группе. Как мы видим, в аналогичном положении находится и аксиоматическое определение вероятности: и здесь для однозначного определения вероятности придется делать какое-нибудь специфическое для данной задачи допущение. Фактически таким допущением нередко бывает опять-таки предположение о равновероятности тех или иных событий.

Когда задано поле вероятностей, то условные вероятности для событий, принадлежащих этому полю, при $P(A) > 0$, можно определить равенством

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

В случае $P(A) = 0$ условная вероятность $P(B/A)$ не определена.

При этом остаются в силе все теоремы I*—VI* п. 12, что легко доказать, подставляя приведенное выражение для условной вероятности в формулировки этих теорем и пользуясь теоремами I—VI п. 10. Остаются в силе также и теоремы VII и VIII п. 13, так как теорема VII просто равносильна приведенному выражению для условной вероятности, а теорема VIII является следствием теоремы VII. Наконец, остаются в силе определение независимости событий и все установленные в п. 14 и в п. 15 свойства независимых событий, так как все это опирается лишь на теоремы VII и VIII п. 13.

19. Чтобы получить пример применения аксиоматического определения вероятности, обратимся ко второй задаче п. 17, где было неприменимо классическое определение.

Эту задачу можно решить, включив фигурирующее в ней событие A_2, состоящее в выпадении области $\mathfrak{A}_2$, в некоторое бесконечное поле вероятностей. А именно, рассмотрим на поверхности шарика всевозможные области, ограниченные *квадрируемыми кривыми*, т. е. кривыми, площади которых равны нулю [1]), и будем рассматривать события, состоящие в выпадении таких областей. Эти события образуют поле событий. В самом деле, если A есть выпадение некоторой области $\mathfrak{A}$, а B — выпадение некоторой области $\mathfrak{B}$, то $A \subset B$ тогда и только тогда, когда область $\mathfrak{A}$ содержится как часть в области $\mathfrak{B}$, событием AB будет выпадение общей части $\mathfrak{A}\mathfrak{B}$ областей $\mathfrak{A}$ и $\mathfrak{B}$, событием $A + B$ — выпадение области $\mathfrak{A} + \mathfrak{B}$, образованной соединением областей $\mathfrak{A}$ и $\mathfrak{B}$, событием $\bar{A}$ — выпадение области $\bar{\mathfrak{A}}$, дополняющей область $\mathfrak{A}$ до полной поверхности шарика; если же области $\mathfrak{A}$ и $\mathfrak{B}$ ограничены квадрируемыми кривыми, то, очевидно, названные здесь области $\mathfrak{A}\mathfrak{B}$, $\mathfrak{A} + \mathfrak{B}$ и $\bar{\mathfrak{A}}$ также ограничены квадрируемыми кривыми. Теперь нам остается определить (пользуясь аксиомами I—IV и нашим предположением о равновероятности выпадения

---

[1]) Раньше думали, что никакой кривой нельзя приписать меру площади, отличную от нуля. Теперь, однако, известно, что существуют кривые, которым можно приписать такую меру. Об этом см., например, в готовящейся к печати книге акад. Н. Н. Лузина. *Начала теории функций действительного переменного*, гл. V. Вот эти-то кривые, площади которых не равны нулю, мы и исключаем из рассмотрения.

областей равной площади) вероятность каждого из событий A, состоящего в выпадении некоторой определенной области $\mathfrak{A}$, ограниченной квадрируемой кривой. Прежде всего представим себе, что вся поверхность шарика подразделена на области $\mathfrak{E}_1, \mathfrak{E}_2, \ldots \mathfrak{E}_n$, ограниченные квадрируемыми кривыми и имеющие равные площади. Сумма площадей областей $\mathfrak{E}_1, \mathfrak{E}_2, \ldots, \mathfrak{E}_n$ равна $4\pi r^2$. Поэтому, обозначая для краткости площадь области знаком абсолютной величины, мы имеем:

$$|\mathfrak{E}_1| = |\mathfrak{E}_2| = \ldots = |\mathfrak{E}_n| = \frac{4\pi r^2}{n}. \tag{11}$$

Обозначим затем через $E_1, E_2, \ldots, E_n$ соответственно выпадения областей $\mathfrak{E}_1, \mathfrak{E}_2, \ldots, \mathfrak{E}_n$. В силу аксиом VII и VIII, сумма вероятностей событий $E_1, E_2, \ldots, E_n$ должна быть равна 1. Поэтому в силу сделанного в данной задаче предположения о равновероятности выпадения областей равной площади мы имеем

$$\boldsymbol{P}(E_1) = \boldsymbol{P}(E_2) = \ldots = \boldsymbol{P}(E_n) = \frac{1}{n}. \tag{12}$$

Сравнивая равенства (11) и (12), мы получаем соотношение между площадями областей $\mathfrak{E}_1, \mathfrak{E}_2, \ldots, \mathfrak{E}_n$ и вероятностями соответствующих этим областям событий $E_1, E_2, \ldots, E_n$, а именно

$$\boldsymbol{P}(E_1) = \frac{|\mathfrak{E}_1|}{4\pi r^2}, \quad \boldsymbol{P}(E_2) = \frac{|\mathfrak{E}_2|}{4\pi r^2}, \quad \ldots, \quad \boldsymbol{P}(E_n) = \frac{|\mathfrak{E}_n|}{4\pi r^2}. \tag{13}$$

Сейчас мы увидим, что то же соотношение сохраняется и для любой данной области $\mathfrak{A}$, ограниченной квадрируемой кривой, и для соответствующего ей события A. В самом деле, очевидно, что вся поверхность шарика может быть подразделена на столь малые области $\mathfrak{E}_1, \mathfrak{E}_2, \ldots, \mathfrak{E}_n$, что если обозначить через $\mathfrak{E}_*$ соединение тех из областей $\mathfrak{E}_1, \mathfrak{E}_2, \ldots, \mathfrak{E}_n$, которые находятся внутри данной области $\mathfrak{A}$, а через $\mathfrak{E}^*$ соединение тех из областей $\mathfrak{E}_1, \mathfrak{E}_2, \ldots, \mathfrak{E}_n$, которые либо находятся внутри области $\mathfrak{A}$, либо хотя бы имеют с ней общие части, то разность между площадями областей $\mathfrak{E}^*$ и $\mathfrak{E}_*$ будет как угодно мала [1]). Обозначим через E_* и E^* соответственно выпадения областей $\mathfrak{E}_*$ и $\mathfrak{E}^*$. Так как площади областей $\mathfrak{E}_*$ и $\mathfrak{E}^*$ складываются из площадей областей $\mathfrak{E}_1, \mathfrak{E}_2, \ldots, \mathfrak{E}_n$, содержащихся в той и в другой, а вероятности событий E_* и E^* должны складываться, в силу аксиомы VII, из вероятностей соответствующих названным областям событий $E_1, E_2, \ldots, E_n$, то равенства (13) дают нам

$$\boldsymbol{P}(E_*) = \frac{|\mathfrak{E}_*|}{4\pi r^2}, \quad \boldsymbol{P}(E^*) = \frac{|\mathfrak{E}^*|}{4\pi r^2}. \tag{14}$$

Далее, область $\mathfrak{E}_*$ содержится в области $\mathfrak{A}$, а область $\mathfrak{A}$ — в области $\mathfrak{E}^*$. Поэтому

$$|\mathfrak{E}_*| \leqslant |\mathfrak{A}| \leqslant |\mathfrak{E}^*|. \tag{15}$$

---

[1]) Так как в противном случае площадь кривой, ограничивающей область $\mathfrak{A}$, не была бы равна нулю.

При этом событие E_* влечет за собой событие A, а событие A — событие E^*. Поэтому, в силу аксиомы VI, мы должны иметь

$$(16) \qquad P(E_*) \leqslant P(A) \leqslant P(E^*).$$

Сравнивая (14) и (16), мы получаем

$$(17) \qquad \frac{|\mathfrak{E}_*|}{4\pi r^2} \leqslant P(A) \leqslant \frac{|\mathfrak{E}^*|}{4\pi r^2}.$$

Поскольку же разность между $|\mathfrak{E}^*|$ и $|\mathfrak{E}_*|$ может быть как угодно мала, сравнение неравенств (15) и (17) дает нам

$$(18) \qquad P(A) = \frac{|\mathfrak{A}|}{4\pi r^2}.$$

Построенному таким образом полю вероятностей принадлежит, в частности, событие A_2, состоящее в выпадении области $\mathfrak{A}_2$, вероятность которого ищется во второй задаче п. 17. Эту вероятность непосредственно дает нам формула (18), а именно, так как

$$|\mathfrak{A}_2| = (2 - \sqrt{3})\,\pi r^2,$$

то

$$P(A_2) = \frac{2 - \sqrt{3}}{4}.$$

20. Аксиоматическое определение вероятности является обобщением классического определения. В тех случаях, когда применимо классическое определение, применимо также и аксиоматическое определение, и последнее дает то же самое значение вероятности, что и первое.

В самом деле, пусть дана полная группа событий $E_1, E_2, \ldots, E_n$, несовместимых и равновероятных между собой. Мы уже видели, что допустимые по отношению к этой группе события A образуют поле событий. Определим же, пользуясь аксиомами I—IV, вероятность любого такого события A. В силу аксиом III и IV сумма вероятностей событий $E_1, E_2, \ldots, E_n$ должна быть равна 1. Поэтому, в силу равновероятности событий $E_1, E_2, \ldots, E_n$,

$$P(E_1) = P(E_2) = \ldots = P(E_n) = \frac{1}{n}.$$

Обозначим теперь через $E'_1, E'_2, \ldots, E'_m$ те из событий $E_1, E_2, \ldots, E_n$, на которые подразделяется событие A. В силу аксиомы III, мы должны иметь

$$P(A) = P(E'_1) + P(E'_2) + \ldots + P(E'_m) = \frac{m}{n}.$$

Мы нашли таким образом для любого допустимого события A то же самое значение вероятности, которое дает классическое определение.

ГЛАВА II.

ПРИЕМЫ ВЫЧИСЛЕНИЯ ВЕРОЯТНОСТЕЙ.

§ 5. Комбинаторика.

21. Подсчет результатов испытания, благоприятствующих данному событию, часто бывает технически очень сложен. Здесь нередко приходится прибегать к помощи комбинаторики, т. е. учения о перестановках и сочетаниях.

Перестановкой (или размещением) из n предметов по m называется группа m предметов, выбранных из числа данных n предметов и расположенных в определенном порядке. Так, например, существует три перестановки из 3 предметов a, b, c по 1:

$$a \qquad b \qquad c;$$

существует шесть перестановок из 3 предметов по 2:

$$\begin{array}{ccc} a,\ b & b,\ a & c,\ a \\ a,\ c & b,\ c & c,\ b; \end{array}$$

существует шесть перестановок из 3 предметов по 3:

$$\begin{array}{ccc} a,\ b,\ c & b,\ a,\ c & c,\ a,\ b \\ a,\ c,\ b & b,\ c,\ a & c,\ b,\ a. \end{array}$$

Число всех возможных перестановок из n предметов по m обозначается через P_n^m и определяется следующим образом. Прежде всего ясно, что существует n перестановок из n предметов по 1:

$$P_n^1 = n.$$

Затем все возможные перестановки из n предметов по 2, очевидно, получаются, если брать любую перестановку из n предметов по 1 и присоединять к ней любой из оставшихся $n-1$ предметов, не вошедших в эту перестановку. Таким образом

$$P_n^2 = n(n-1).$$

Все возможные перестановки из n предметов по 3 получаются, если брать любую перестановку из n предметов по 2 и присоединять к ней любой из оставшихся $n-2$ предметов. Таким образом

$$P_n^3 = n(n-1)(n-2).$$

Продолжая дальше таким же образом, мы видим, что вообще

$$P_n^m = n(n-1)(n-2)\ldots(n-m+1). \tag{19}$$

Перестановка из n предметов по n называется просто перестановкой из n предметов. Число всех таких перестановок обозначается кратко через P_n и, как нетрудно видеть, равно

$$P_n = n(n-1)(n-2)\ldots 1. \tag{20}$$

Сочетанием из n предметов по m называется группа из m предметов, выбранных из числа данных n предметов (причем безразлично, в каком порядке они расположатся). Так, например, существует три сочетания из 3 предметов a, b, c по 1:

$$a \qquad b \qquad c;$$

существует три сочетания из 3 предметов по 2:

$$a,\ b \qquad b,\ c \qquad c,\ a,$$

существует одно сочетание из 3 предметов по 3:

$$a,\ b,\ c.$$

Число всех возможных сочетаний из n предметов по m обозначается через C_n^m и может быть определено из сравнения с числом перестановок. Каждое сочетание из n предметов по m, если располагать входящие в него предметы в том или ином определенном порядке, дает столько различных перестановок, сколько вообще существует перестановок из m предметов, т. е. P_m. Все же сочетания из n предметов по m дадут таким образом $C_n^m P_m$ перестановок. И все это будут, очевидно, различные перестановки из n предметов по m. Следовательно, $C_n^m P_m \leqslant P_n^m$. Наоборот, каждая перестановка из n предметов по m, очевидно, может быть получена таким образом из определенного сочетания из n предметов по m путем расположения входящих в него предметов в определенном порядке. Следовательно, $P_n^m \leqslant C_n^m P_m$. Итак:

$$C_n^m P_m = P_n^m,$$

откуда

$$C_n^m = \frac{P_n^m}{P_m},$$

или, что то же:

$$C_n^m = \frac{n(n-1)(n-2)\ldots(n-m+1)}{m(m-1)(m-2)\ldots 1}. \tag{21}$$

Изящные и легко запоминаемые выражения для числа P_n^m, P_n и C_n^m можно получить при помощи факториала:

$$n! = 1 \cdot 2 \cdot \ldots \cdot n.$$

Нетрудно вывести из формул (19), (20) и (21), что

(22) $$P_n^m = \frac{n!}{(n-m)!};$$

(23) $$P_n = n!;$$

(24) $$C_n^m = \frac{n!}{m!\,(n-m)!}.$$

По определению полагают еще $C_n^0 = 1$ и $0! = 1$. Этим достигается полная общность формул (22), (23) и (24).

22. В наиболее важных вопросах теории вероятностей и ее приложений комбинаторика применяется по следующей схеме.

Возьмем n предметов и подразделим их на два класса по каким-нибудь двум признакам X и Y, таким, что каждый из взятых нами предметов обладает одним и только одним из этих признаков. В результате такого подразделения мы получим: 1) определенное *распределение* n предметов по двум признакам, — если будет указано, какой именно предмет каким признаком обладает, и 2) определенный *состав* (или статистическое распределение) n предметов по двум признакам, — если будет указано только количество предметов, обладающих тем или другим признаком. Так, например, возможны восемь различных распределений 3 предметов a, b, c по любым двум признаком X и Y, а именно:

X	Y	X	Y
$a,\ b,\ c$	—	a	$b,\ c$
$a,\ b$	c	b	$a,\ c$
$a,\ c$	b	c	$a,\ b$
$b,\ c$	a	—	$a,\ b,\ c$

Но возможны только четыре различных состава (или статистических распределения) 3 предметов по двум признакам, а именно:

X	Y	X	Y
3	0	1	2
2	1	0	3

где цифры 0, 1, 2, 3 обозначают количество предметов, обладающих соответствующим признаком X или Y.

Очевидно, что различные распределения могут иметь одинаковый состав. Так, в только что приведенном примере три распределения имеют состав из 2 предметов с признаком X и из 1 предмета с признаком Y, равно как три же распределения имеют состав из 1 предмета с признаком X и из 2 предметов с признаком Y.

Спросим себя, сколько вообще распределений n предметов по двум признакам имеют данный состав из m предметов с признаком X и $n-m$ предметов с признаком Y? Ответ на этот вопрос нетруден. Для того чтобы распределение n предметов по двум признакам имело названный состав, очевидно, необходимо и достаточно, чтобы в этом распределении обладали признаком X ровно m каких-нибудь предметов из общего числа n предметов; иначе говоря, чтобы в это распределение входило с признаком X какое-нибудь сочетание из n предметов по m. Таким образом искомое число распределений равно числу всех возможных сочетаний из n предметов по m, т. е. равно

$$C_n^m,$$

или по формуле (24)

$$\frac{n!}{m!\,(n-m)!}.$$

Естественным обобщением только что изложенной схемы применения комбинаторики является следующая схема.

Возьмем n предметов и подразделим их на s классов ($s \geqslant 2$) по каким-нибудь s признакам $X_1, X_2, \ldots, X_s$, таким, что каждый из взятых нами предметов обладает одним и только одним из этих признаков. Будем говорить о распределении n предметов по s признакам и о составе n предметов по s признакам в таком же смысле, как выше в случае $s=2$.

Спросим себя, сколько распределений n предметов по s признакам имеют данный состав из m_1 предметов с признаком X_1, из m_2 предметов с признаком X_2 и т. д., наконец, из m_s предметов с признаком X_s? На этот вопрос можно ответить следующим образом. Для того чтобы распределение n предметов по s признакам имело названный состав, очевидно, необходимо и достаточно, чтобы в нем обладали признаком X_1 ровно m_1 каких-нибудь предметов из общего числа n предметов, затем — признаком X_2 ровно m_2 каких-нибудь из числа остальных $n-m_1$ предметов и т. д., наконец — признаком X_{s-1} ровно m_{s-1} каких-нибудь предметов из числа $n-m_1-\ldots-m_{s-2}$ предметов; иначе говоря, необходимо и достаточно, чтобы в это распределение входило с признаком X_1 какое-нибудь сочетание из n предметов по m_1, с признаком X_2 — какое-нибудь сочетание из остальных $n-m_1$ предметов по m_2 и т. д., наконец, с признаком X_{s-1} — какое-нибудь сочетание из $n-m_1-\ldots-m_{s-2}$ предметов по m_{s-1}. Таким образом искомое число распределений равно произведению

$$C_n^{m_1} C_{n-m_1}^{m_2} \ldots C_{n-m_1-\ldots-m_{s-2}}^{m_{s-1}}.$$

Это произведение по формуле (24) равно

$$\frac{n!}{m_1!\,(n-m_1)!}\;\frac{(n-m_1)!}{m_2!\,(n-m_1-m_2)!}\cdots\frac{(n-m_1-\ldots-m_{s-2})!}{m_{s-1}!\,(n-m_1-\ldots-m_{s-1})!}$$

или после очевидных сокращений

$$\frac{n!}{m_1!\,m_2!\ldots(n-m_1-\ldots-m_{s-1})!}.$$

А так как $n-m_1-\ldots-m_{s-1}=m_s$, то окончательно

$$\frac{n!}{m_1!\,m_2!\ldots m_s!}.$$

§ 6. Последовательные испытания.

23. Задача о независимых испытаниях. Будем производить последовательные испытания, в результате каждого из которых может наступить или не наступить некоторое событие A, и пусть при каждом отдельном испытании вероятность наступления события A одна и та же и не зависит от наступления или ненаступления этого события при других испытаниях; обозначим ее через p. Какова вероятность того, что событие A наступит ровно при m испытаниях (и, следовательно, не наступит ровно при $n-m$ испытаниях)?

Эта задача решается следующим образом. Представим себе все возможные комбинации из последовательных результатов наших испытаний. Так, например, в случае 3 испытаний возможны восемь таких комбинаций, а именно, если обозначать, как всегда, через $\overline{A}$ ненаступление события A:

$$\begin{array}{lllllll} A, & A, & A; & \qquad & A, & \overline{A}, & \overline{A}; \\ A, & A, & \overline{A}; & & \overline{A}, & A, & \overline{A}; \\ A, & \overline{A}, & A; & & \overline{A}, & \overline{A}, & A; \\ \overline{A}, & A, & A; & & \overline{A}, & \overline{A}, & \overline{A}. \end{array}$$

Выделим те комбинации, в которых событие A наступает ровно m раз (и, следовательно, не наступает ровно $n-m$ раз); назовем для краткости такие комбинации **допустимыми**. Определим вероятность появления каждой отдельной допустимой комбинации. Для этого заметим, что появление допустимой комбинации представляет собой произведение n событий, а именно: m наступлений события A при одних испытаниях и $n-m$ его ненаступлений при других испытаниях. Вероятность наступления события A при каждом отдельном испытании по условию равна p; следовательно, вероятность его ненаступления по теореме V равна $q=1-p$. По условию наступления или ненаступления события A при различных испытаниях представляют собой независимые события; следовательно, вероятность их произведения по теореме VII bis равна произведению их вероятностей. Таким образом вероятность появления каждой отдельной допустимой комбинации, представляющей собой произведение m наступлений и $n-m$ ненаступлений события A, равна $p^m q^{n-m}$.

Заметим теперь, что то, что нас интересует — наступление события A ровно при m испытаниях, — очевидно, равносильно появлению хотя бы одной из допустимых комбинаций. Поэтому, поскольку различные комбинации несовместимы, искомая вероятность наступления события A ровно при m испытаниях по теореме III равна сумме вероятностей появлений допустимых комбинаций. Следовательно, искомая вероятность равна $Kp^m q^{n-m}$, где K — число всех допустимых комбинаций.

Остается определить число K. Это, очевидно, есть число всех тех возможных распределений n последовательных испытаний по двум признакам (наступление и ненаступление A), которые имеют один и тот же состав из m наступлений и $n-m$ ненаступлений A. Следовательно, как мы видели в п. 22,

$$K=C_n^m=\frac{n!}{m!\,(n-m)!}\cdot$$

Итак, окончательно искомая вероятность наступления события A ровно при m испытаниях равна

$$C_n^m p^m q^{n-m} = \frac{n!}{m!\,(n-m)!}\, p^m q^{n-m} \qquad (q = 1 - p). \tag{25}$$

Это выражение известно в алгебре как общий член бинома Ньютона:

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

Поэтому это выражение называется *биномиальным законом* распределения вероятностей.

24. Если выписать коэфициенты C_n^m по схеме

$$\begin{array}{ccccccc}
 & & & C_0^0 & & & \\
 & & C_1^0 & & C_1^1 & & \\
 & C_2^0 & & C_2^1 & & C_2^2 & \\
C_3^0 & & C_3^1 & & C_3^2 & & C_3^3 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots,
\end{array}$$

то получится так называемый треугольник Паскаля:

$$\begin{array}{ccccccccccccccc}
 & & & & & & & 1 & & & & & & & \\
 & & & & & & 1 & & 1 & & & & & & \\
 & & & & & 1 & & 2 & & 1 & & & & & \\
 & & & & 1 & & 3 & & 3 & & 1 & & & & \\
 & & & 1 & & 4 & & 6 & & 4 & & 1 & & & \\
 & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & & \\
 & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 & \\
1 & & 7 & & 21 & & 35 & & 35 & & 21 & & 7 & & 1 \\
\ldots & & \ldots & & \ldots & & \ldots & & \ldots & & \ldots & & \ldots & & \ldots
\end{array}$$

Мы видим, что при каждом данном n эти коэфициенты с возрастанием m сначала возрастают до некоторого максимума, затем убывают. Можно ожидать, что и вероятности $C_n^m p^m q^{n-m}$ будут вести себя подобным же образом. Сейчас мы увидим, что это действительно так.

Чтобы точнее уяснить себе, как распределяются вероятности того, что событие A наступит ровно m раз при данных n, p и q, рассмотрим отношение этих вероятностей для двух последовательных значений, m и $m+1$. Оно равно

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

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

При переходе от значения m к значению $m+1$ вероятность возрастает или убывает, смотря по тому, что имеет бо́льшую величину в последнем выражении, числитель или знаменатель. Таким образом вероятность возрастает, когда

$$np-mp>mq+q,$$

или, что то же (так как $p+q=1$), когда

$$m<np-q;$$

и вероятность убывает, когда

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

или, что то же, когда

$$m>np-q.$$

Мы видим, что вероятность сначала возрастает (пока $m<np-q$), затем убывает.

Обозначим через m_0 наименьшее из значений m, для которого

$$m_0>np-q.$$

Тогда для значения m_0-1 остаются две возможности:

1° $$m_0-1<np-q.$$
2° $$m_0-1=np-q.$$

В случае 1° при переходе от значения m_0-1 к значению m_0 вероятность возрастает, а так как при переходе от значения m_0 к значению m_0+1 она уже убывает, то в этом случае m_0 — единственное значение m, для которого вероятность получает свое наибольшее значение, единственное „наивероятнейшее“ значение числа наступлений события A. Так как оно больше $np-q$, а оно же без единицы меньше $np-q$, то оно заключено внутри интервала $(np-q,\ np+p)$. Этим m_0 вполне определяется, так как оно — целое число, а названный интервал имеет длину $p+q=1$. В случае же 2°, когда $m_0-1=$ $=np-q$, при переходе от значения m_0-1 к значению m_0 вероятность не изменяется, так как при

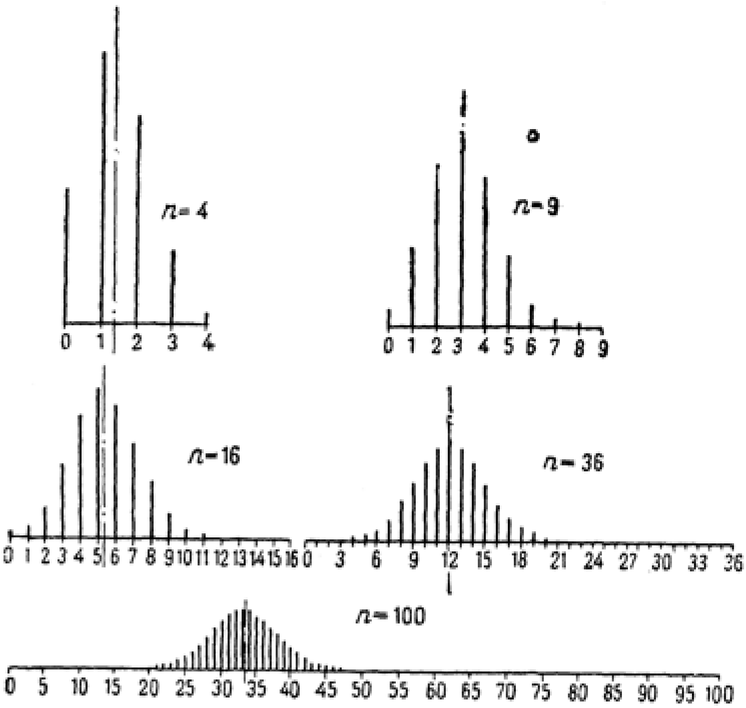
$$m=np-q$$

мы имеем

$$\frac{n-m}{m+1}\cdot\frac{p}{q}=\frac{n-np+q}{np-q+1}\cdot\frac{p}{q}=\frac{nq+q}{np+p}\cdot\frac{p}{q}=1.$$

Но при переходе от значения $m_0-2<np-q$ к значению $m_0-1=$ $=np-q$ вероятность возрастает, а так как при переходе от значения m_0 к значению m_0+1 она уже убывает, то в этом случае есть два и только два „наиболее вероятных“ числа наступлений события A. Это $m_0-1=np-q$ и $m_0=np+p$, т. е. концы интервала $(np-q,\ np+p)$.

На черт. 1 изображена[1]) диаграмма распределения вероятностей по биномиальному закону для нескольких различных значений n; здесь $p=\frac{1}{3}$.



Черт. 1.

25. Задача о статистической выборке. Пусть дано N предметов, каждый из которых обладает одним и только одним из двух признаков X и Y, причем признаком X обладают M предметов, а признаком Y обладают $L=N-M$ предметов. Будем выбирать последовательно n предметов из общего числа данных N предметов, не возвращая выбранных предметов обратно, и пусть после выбора каждых i предметов следующим за ними $(i+1)$-м предметом может оказаться с одинаковой вероятностью любой из еще невыбранных предметов. Какова вероятность того, что мы получим в выборке ровно m предметов с признаком X (и, следовательно, ровно $l=n-m$ предметов с признаком Y)?

Эта задача решается следующим образом. Обозначим данные N предметов через $a_1, a_2, \ldots, a_N$. При выборке из них n предметов мы

[1]) Черт. 1, так же как черт. 4, 5 и 7, заимствован из книги R. v. Mises, *Wahrscheinlichkeitsrechnung*, 1931.

получаем некоторое сочетание из N предметов по n:

$$(C) \qquad a_{k_1},\ a_{k_2},\ \ldots,\ a_{k_n}.$$

Так как по условию выборка производится наугад, то очевидно, что все такие сочетания (C) равновероятны между собой. Общее число возможных сочетаний (C) есть C_N^n. Далее, каждое из сочетаний (C), которое содержит m предметов с признаком X и l предметов с признаком Y, представляет собой соединение двух сочетаний: сочетания

$$(A) \qquad a_{i_1},\ a_{i_2},\ \ldots,\ a_{i_m},$$

содержащего исключительно предметы с признаком X, и сочетания

$$(B) \qquad a_{j_1},\ a_{j_2},\ \ldots,\ a_{j_l},$$

содержащего исключительно предметы с признаком Y. Общее число возможных сочетаний (A) есть C_M^m, а общее число возможных сочетаний (B) есть C_L^l; следовательно, общее число возможных попарных соединений сочетаний (A) с сочетаниями (B) есть $C_M^m C_L^l$. Итак, в силу самого определения вероятности, вероятность получения в выборке ровно m предметов с признаком X равна

$$(26) \qquad \frac{C_M^m C_L^l}{C_N^n} = \frac{M!}{m!(M-m)!}\ \frac{L!}{l!(L-l)!}\ \frac{n!(N-n)!}{N!}.$$

26. Распределение вероятностей по формуле (26) можно было бы подвергнуть такому же детальному изучению, какое мы провели выше для биномиального закона распределения вероятностей (25). Мы, однако, не будем здесь углубляться в эти детали, так как они не имеют большого практического значения. В самом деле, практически бывают интересны почти исключительно те случаи, когда количество n выбранных предметов очень невелико по сравнению с общим количеством N предметов в совокупности, т. е. когда дроби вида $\frac{K}{N}$, где $K \leqslant n$, очень малы, и при расчетах ими можно пренебрегать. Но выражение (26) можно написать в виде

$$\frac{n!}{m!\,l!}\ \frac{\frac{M}{N}\left(\frac{M}{N}-\frac{1}{N}\right)\cdots\left(\frac{M}{N}-\frac{m-1}{N}\right)\frac{L}{N}\left(\frac{L}{N}-\frac{1}{N}\right)\cdots\left(\frac{L}{N}-\frac{l-1}{N}\right)}{1\left(1-\frac{1}{N}\right)\cdots\left(1-\frac{n-1}{N}\right)},$$

откуда видно, что если пренебречь дробями вида $\frac{K}{N}$, где $K \leqslant N$, то (26) превращается в

$$(26\text{bis}) \qquad \frac{n!}{m!\,l!}\left(\frac{M}{N}\right)^m\left(\frac{L}{N}\right)^l.$$

С другой стороны, если бы мы, выбирая каждый предмет из совокупности, возвращали его обратно до того, как выбрать следующий, то вероятность выбрать предмет с признаком X каждый раз была бы равна $\frac{M}{N}$, а с признаком Y — равна $\frac{L}{N}$. Следовательно, мы находились бы в условиях независимых испытаний, с $p=\frac{M}{N}$ и $q=\frac{L}{N}$, и применяя формулу (25), мы имели бы для вероятности получить в выборке из n предметов ровно m предметов с признаком X как раз выражение (26 bis).

Таким образом практически мы почти всегда можем обойтись без формулы (26), заменяя ее более простой формулой (25), с $p=\frac{M}{N}$ и $q=\frac{L}{N}$. И в дальнейшем мы будем интересоваться преимущественно этой последней формулой (25).

§ 7. Формула полной вероятности.

27. Пусть дана группа несовместимых событий $A_1, A_2, \ldots, A_s$ и некоторое событие B, подразделяющееся на частные случаи BA_1, $BA_2, \ldots, BA_s$, и пусть даны вероятности $\boldsymbol{P}(A_1)$, $\boldsymbol{P}(A_2), \ldots, \boldsymbol{P}(A_s)$ и условные вероятности $\boldsymbol{P}(B/A_1)$, $\boldsymbol{P}(B/A_2), \ldots, \boldsymbol{P}(B/A_s)$. Тогда можно определить вероятность $\boldsymbol{P}(B)$ по так называемой *формуле полной вероятности*:

$$\boldsymbol{P}(B)=\boldsymbol{P}(A_1)\,\boldsymbol{P}(B/A_1)+\boldsymbol{P}(A_2)\,\boldsymbol{P}(B/A_2)+\ldots+\boldsymbol{P}(A_s)\,\boldsymbol{P}(B/A_s).$$

Действительно, в силу теоремы III

$$\boldsymbol{P}(B)=\boldsymbol{P}(BA_1)+\boldsymbol{P}(BA_2)+\ldots+\boldsymbol{P}(BA_s).$$

Но, в силу теоремы VII,

$$\begin{aligned}\boldsymbol{P}(BA_1)&=\boldsymbol{P}(A_1)\ \boldsymbol{P}(B/A_1),\\ \boldsymbol{P}(BA_2)&=\boldsymbol{P}(A_2)\ \boldsymbol{P}(B/A_2),\\ &\ldots\ldots\ldots\ldots\\ \boldsymbol{P}(BA_s)&=\boldsymbol{P}(A_s)\ \boldsymbol{P}(B/A_s).\end{aligned}$$

Подставляя это в предшествующее равенство, мы и получаем формулу полной вероятности.

28. Цепи Маркова. Будем производить ряд последовательных испытаний, в результате каждого из которых может наступить или не наступить некоторое событие A, и пусть в случае наступления события A при n-м по порядку испытании вероятность его наступления при $(n+1)$-м испытании равна a, а в случае его ненаступления при n-м испытании вероятность его наступления при $(n+1)$-м испытании равна b. Поставим задачу, зная вероятность p_1 его наступления при первом по порядку испытании, определить вероятность p_n его наступления при каждом n-м испытании.

Обозначим на минуту через A_n наступление события A при n-м испытании. Тогда, поскольку событие A_n и противоположное ему событие $\bar{A}_n$ при каждом данном значении n образуют полную группу и несовместимы между собой, событие A_{n+1} подразделяется на два частных случая $A_{n+1}A_n$ и $A_{n+1}\overline{A}_n$, и мы имеем по формуле полной вероятности

$$\boldsymbol{P}(A_{n+1}) = \boldsymbol{P}(A_n)\ \boldsymbol{P}(A_{n+1}/A_n) + \boldsymbol{P}(\overline{A}_n)\ \boldsymbol{P}(A_{n+1}/\overline{A}_n).$$

Мы же положили $\boldsymbol{P}(A_{n+1}/A_n) = a$, $\boldsymbol{P}(A_{n+1}/\overline{A}_n) = b$, $\boldsymbol{P}(A_n) = p_n$ и, следовательно, $\boldsymbol{P}(\overline{A}_n) = 1 - p_n$. Поэтому у нас

$$p_{n+1} = ap_n + b(1 - p_n). \tag{27}$$

Таким образом мы имеем для определения интересующей нас вероятности p_n уравнение в конечных разностях (27). Уравнение это можно решить следующим образом.

Заметим прежде всего, что оно имеет *стационарное*, т. е. не зависящее от n, решение $p_n = p$. Действительно, это последнее есть решение уравнения

$$v = ap + b(1 - p), \tag{28}$$

в которое превращается уравнение (27) при постоянном $p_n = p$. Уравнение же (28) удовлетворяется, если положить

$$p = \frac{b}{1 - a + b}.$$

Затем, зная стационарное решение p уравнения (27), можно получить и его общее решение. Для этого положим $p_n = p + p_n^*$. Подставляя это в (27), мы получим

$$p + p_{n+1}^* = ap + ap_n^* + b(1 - p) - bp_n^*,$$

или, принимая во внимание (28):

$$p_{n+1}^* = p_n^*(a - b).$$

Отсюда ясно, что

$$p_{n+1}^* = p_1^*(a - b)^n.$$

Переходя от p_n^* к p_n, мы получим

$$p_{n+1} - p = (p_1 - p)(a - b)^n,$$

или, что то же:

$$p_{n+1} = \frac{b}{1 - a + b} + \left(p_1 - \frac{b}{1 - a + b}\right)(a - b)^n.$$

Наши рассуждения неприменимы только в том случае, когда $a = 1$ и $b = 0$, так как тогда знаменатель в выражении для p обращается в нуль. В этом исключительном случае, как показывает само уравнение (27), $p_{n+1} = p_n$ и, следовательно,

$$p_{n+1} = p_1.$$

Если $a < 1$ и $b > 0$, то найденное нами выражение для общего решения p_{n+1} показывает, между прочим, что при $n \to \infty$ это решение имеет вполне определенный предел, не зависящий от p_1 а именно:

$$\lim_{n\to\infty} p_n = \frac{b}{1-a+b}.$$

Интересен частный случай, когда $a=b$; в этом случае p_n при всяком $n>1$ не зависит от p_1 и равно b, или, что то же, равно a.

Для примера рассмотрим следующий вопрос. Пусть по некоторой городской трассе размещены автоматические светофоры, отрегулированные применительно к некоторой средней скорости автомобилей, и притом с таким расчетом, чтобы автомобиль, задержавшийся у красного огня какого-нибудь светофора, наверняка попал к зеленому огню следующего светофора. Что же касается автомобиля, прошедшего под зеленым огнем какого-нибудь светофора, то мы предположим, схематически упрощая картину уличного движения, что этот автомобиль всегда имеет одну и ту же вероятность $a<1$ попасть к зеленому огню следующего светофора. Пусть p_n есть вероятность того, что автомобиль попадет к зеленому огню n-го из пройденных им светофоров и, следовательно, $1-p_n$ есть вероятность того, что он попадет к красному огню. Требуется, зная p_1, найти p_n. Здесь, очевидно,

$$p_{n+1} = ap_n + (1-p_n).$$

Это — формула (27), где $b=1$. Таким образом

$$p_{n+1} = \frac{1}{2-a} + \left(p_1 - \frac{1}{2-a}\right)(a-1)^n.$$

Мы видим, что искомая вероятность p_n асимптотически приближается к $\frac{1}{2-a}$, становясь попеременно то больше, то меньше этого значения.

В разобранной нами задаче вероятности p_n образуют простейший случай так называемой *цепи Маркова*.

Более общий случай мы получаем при решении следующей задачи. Пусть дана полная группа несовместимых между собой событий $A_1, A_2, \ldots, A_s$. Будем производить ряд последовательных испытаний, в результате каждого из которых наступает одно (и, следовательно, только одно) из событий нашей группы, и пусть при условии наступления события A_i при n-м по порядку испытании вероятность наступления события A_k при $(n+1)$-м испытании равна a_{ik}. Поставим задачу, зная все вероятности $p_1^{(1)}, p_2^{(1)}, \ldots, p_s^{(1)}$ наступления событий $A_1, A_2, \ldots, A_s$ при первом по порядку испытании, определить вероятности $p_1^{(n)}, p_2^{(n)}, \ldots, p_s^{(n)}$ наступления этих событий при любом n-м испытании.

Обозначим на минуту через A_i^n наступление события A_i при n-м испытании. Тогда, поскольку события A_i^n при каждом данном значении n образуют полную группу и несовместимы между собой, каждое из событий A_k^{n+1} можно подразделить на s частных случаев $A_k^{n+1} A_i^n$, где $i=1, 2, \ldots, s$, и мы имеем по формуле полной вероятности

$$P(A_1^{n+1}) = P(A_1^n)\, P(A_1^{n+1}/A_1^n) + \ldots + P(A_s^n)\, P(A_1^{n+1}/A_s^n),$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$P(A_s^{n+1}) = P(A_1^n)\, P(A_s^{n+1}/A_1^n) + \ldots + P(A_s^n)\, P(A_s^{n+1}/A_s^n).$$

Мы же положили $P(A_k^{n+1}/A_i^n) = a_{ik}$ и $P(A_i^n) = p_i^{(n)}$.

Поэтому у нас

$$p_1^{(n+1)} = a_{11}p_1^{(n)} + \ldots + a_{s1}p_s^{(n)},$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$p_s^{(n+1)} = a_{1s}p_1^{(n)} + \ldots + a_{s1}p_s^{(n)}.$$

Эти равенства дают возможность решить поставленную задачу, так как при помощи них можно, зная вероятности при первом испытании, найти вероятности при втором испытании, затем, зная вероятности при втором испытании, найти вероятности при третьем испытании и т. д.

Системы вероятностей, связанных такими равенствами, и называются *цепями Маркова* (подробное изложение цепей Маркова находится во II т. книги Fréchet).

§ 8. Формулы Бейеса.

29. Пусть дана группа несовместимых событий $A_1, A_2, \ldots, A_s$ и некоторое событие B, подразделяющееся на частные случаи BA_1, $BA_2, \ldots, BA_s$, и пусть даны вероятности $P(A_1), P(A_2), \ldots, P(A_s)$ и условные вероятности $P(B/A_1), P(B/A_2), \ldots, P(B/A_s)$. Тогда можно определить условные вероятности

$$P(A_1/B),\ P(A_2/B), \ldots,\ P(A_s/B)$$

по *формулам Бейеса*:

$$P(A_1/B) = \frac{P(A_1)\, P(B/A_1)}{P(A_1)\, P(B/A_1) + P(A_2)\, P(B/A_2) + \ldots + P(A_s)\, P(B/A_s)},$$
$$P(A_2/B) = \frac{P(A_2)\, P(B/A_2)}{P(A_1)\, P(B/A_1) + P(A_2)\, P(B/A_2) + \ldots + P(A_s)\, P(B/A_s)},$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$P(A_s/B) = \frac{P(A_s)\, P(B/A_s)}{P(A_1)\, P(B/A_1) + P(A_2)\, P(B/A_2) + \ldots + P(A_s)\, P(B/A_s)}.$$

Действительно, по формуле (2) п. 14

$$P(A_1/B) = \frac{P(A_1)\, P(B/A_1)}{P(B)},$$

а по формуле полной вероятности п. 27

$$P(B) = P(A_1)\, P(B/A_1) + P(A_2)\, P(B/A_2) + \ldots + P(A_s)\, P(B/A_s).$$

Подставляя второе из этих равенств в первое, мы получаем первую из формул Бейеса. Точно так же доказываются и остальные формулы Бейеса.

Интересен случай, когда события $A_1, A_2, \ldots, A_s$ равновероятны, т. е. когда $P(A_1) = P(A_2) = \ldots = P(A_s)$. Тогда, как нетрудно видеть, формулы Бейеса упрощаются и могут быть написаны так:

$$
\begin{aligned}
\boldsymbol{P}(A_1/B) &= Q\boldsymbol{P}(B/A_1),\\
\boldsymbol{P}(A_2/B) &= Q\boldsymbol{P}(B/A_2),\\
&\dots\dots\dots\\
\boldsymbol{P}(A_s/B) &= Q\boldsymbol{P}(B/A_s),
\end{aligned}
$$

где

$$
Q=\frac{1}{\boldsymbol{P}(B/A_1)+\boldsymbol{P}(B/A_2)+\ldots+\boldsymbol{P}(B/A_s)}.
$$

Таким образом в этом случае условные вероятности $\boldsymbol{P}(A_i/B)$ пропорциональны условным вероятностям $\boldsymbol{P}(B/A_i)$.

Часто называют безусловные вероятности $\boldsymbol{P}(A_1), \boldsymbol{P}(A_2), \ldots, \boldsymbol{P}(A_s)$ вероятностями a priori (*до* выяснения вопроса о том, произошло или нет событие B), а условные вероятности $\boldsymbol{P}(A_1/B), \boldsymbol{P}(A_2/B), \ldots, \boldsymbol{P}(A_s/B)$ — вероятностями a posteriori (*после* выяснения того, что произошло событие B).

30. Обращение задачи о статистической выборке. Пусть дана совокупность из N предметов, каждый из которых обладает одним и только одним из двух признаков X и Y, причем количество M предметов, обладающих признаком X, а следовательно, и количество $L=N-M$ предметов, обладающих признаком Y, неизвестно. Однако для каждого числа $M=0, 1, \ldots, N$ на основании предшествующего опыта известна вероятность p_M того, что данная совокупность содержит M предметов с признаком X. Будем выбирать наугад один за другим n предметов из данной совокупности, не возвращая выбранных предметов обратно, и пусть при этом среди выбранных предметов окажется m предметов, обладающих признаком X. Спрашивается, какова после этого нового испытания будет вероятность того, что данная совокупность содержит то или иное число M предметов с признаком X?

Обозначим через $A_0, A_1, \ldots, A_N$ события, состоящие в том, что в данной совокупности находится соответственно $0, 1, \ldots, N$ предметов, обладающих признаком X, и через B — событие, состоящее в том, что при выборе из этой совокупности наугад n предметов среди них оказывается m предметов с признаком X. Событие, которое нас интересует — то, что данная совокупность содержит M предметов с признаком X, есть A_M. Интересующая же нас вероятность есть условная вероятность этого события после того, как в данной совокупности среди выбранных наугад n предметов оказалось m предметов с признаком X, т. е. после того, как произошло событие B. Эту вероятность можно определить пользуясь формулами Бейеса, в силу которых она равна

$$
(29)\quad \boldsymbol{P}(A_M/B)=\frac{\boldsymbol{P}(A_M)\,\boldsymbol{P}(B/A_M)}{\boldsymbol{P}(A_0)\,\boldsymbol{P}(B/A_0)+\boldsymbol{P}(A_1)\,\boldsymbol{P}(B/A_1)+\ldots+\boldsymbol{P}(A_N)\,\boldsymbol{P}(B/A_N)}.
$$

Остается определить вероятности a priori $\boldsymbol{P}(A_M)$ и условные вероятности $\boldsymbol{P}(B/A_M)$, где $M=0, 1, \ldots, N$. Что касается вероятности a priori $P(A_M)$, то она, как мы упомянули выше, равна p_M. Что же касается условной вероятности $P(B/A_M)$, т. е. вероятности того, что если совокупность содержит всего N предметов и в том числе M предметов, обладающих признаком X, то при выборе из нее наугад n предметов появится m предметов с признаками X, — эта вероятность, как мы знаем из решенной нами в п. 25 задачи о статистической выборке, равна

$$\frac{C_M^m C_{N-M}^{n-m}}{C_N^n}.$$

Подставляя все эти данные в формулу (29), мы получаем следующее выражение для искомой вероятности:

$$\frac{p_M C_M^m C_{N-M}^{n-m}}{p_0 C_0^m C_N^{n-m} + p_1 C_1^m C_{N-1}^{n-m} + \ldots + p_N C_N^m C_0^{n-m}}. \tag{30}$$

При этом всегда должно быть $m \leqslant M$ и $n - m \leqslant N - M$; следовательно, условные вероятности $\boldsymbol{P}(B/A_M)$, или, что то же, их числители $C_M^m C_{N-M}^{n-m}$ в случае $M < m$, $M > m + N - n$ должно полагать равными нулю. Таким образом выражение (30), полученное нами для искомой вероятности, может быть окончательно написано в таком виде:

$$\frac{p_M C_M^m C_{N-M}^{n-m}}{\sum\limits_{K=m}^{m+N-n} p_K C_K^m C_{N-K}^{n-m}}.$$

ГЛАВА III.

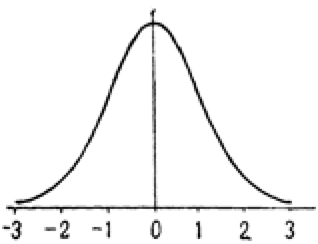
АСИМПТОТИЧЕСКИЕ ФОРМУЛЫ.

§ 9. Локальная теорема Лапласа.

31. Вычисление вероятностей по заданным формулам часто бывает очень сложным. Поэтому в теории вероятностей широко пользуются так называемыми *асимптотическими*, или *предельными, формулами*, которыми можно заменять истинные формулы, не вызывая этим заметных погрешностей в вычислениях.

Для определенного типа задач асимптотической формулой может служить так называемый *закон Гаусса* — интеграл функции

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$



Черт. 2.

График этой функции, называемый *нормальной кривой*, изображен на черт. 2.

Пусть μ — число наступлений события A при n независимых испытаниях, при каждом из которых вероятность этого события равна постоянной p, не равной нулю и не равной единице.

Будем обозначать через $P_n(m)$ вероятность того, что величина μ будет удовлетворять равенству $\mu = m$. На черт. 1 (стр. 38) даны графики величин $P_n(m)$ для нескольких значений n и для всех возможных значений $m = 0, 1, 2, \ldots, n$ величины μ.

Будем изменять эти графики по мере возрастания n при помощи трех преобразований: 1) перемещать графики в целом так, чтобы наибольшая ордината всегда приходилась на начало координат, для чего мы будем заменять абсциссы m новыми абсциссами $m' = m - l_n$, где числа l_n растут с возрастанием n; 2) уменьшать расстояние между соседними ординатами, для чего мы будем заменять абсциссы m' новыми абсциссами $m'' = \frac{m'}{k_n}$, где числа k_n тоже растут с возрастанием n (это изменение масштаба по оси абсцисс на нашем чертеже фактически осуществлено); 3) увеличивать сами ординаты, поддерживая их примерно на одном уровне, для чего мы будем заменять ординаты y ординатами $a_n y$, где числа a_n опять-таки растут с возрастанием n.

Иначе говоря, будем заменять на графиках вероятности $P_n(m)$ величинами

$$a_n P_n(m) = a_n P_n(l_n + m') = a_n P_n(l_n + m'' k_n).$$

По виду графиков на черт. 1 можно предположить, что если подбирать соответствующим образом числа a_n, k_n и l_n, то удастся достичь того, что измененные графики будут сходиться к нормальной кривой, изображенной на черт. 2. Иначе говоря, величины $a_n P_n(l_n + x_n k_n)$ будут стремиться к

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

если $n \to \infty$ и $x_n \to x$.

Сейчас мы докажем, что это так и будет [1]), если взять

$$a_n = k_n = \sqrt{npq} \text{ и } l_n = np.$$

32. Локальная теорема Лапласа. — *Пусть μ — число наступлений события при n независимых испытаниях, при каждом из которых вероятность этого события равна постоянной p, отличной от нуля и единицы, и пусть $P_n(m)$ есть вероятность того, что величина μ будет удовлетворять равенству*

$$\mu = m.$$

Тогда [2])

$$\text{(a)} \qquad \sqrt{npq} P_n(np + x_n \sqrt{npq}) : \frac{1}{\sqrt{2\pi}} e^{-\frac{x_n^2}{2}} \rightrightarrows 1 \qquad (q = 1 - p),$$

если $n \to \infty$, а x_n варьирует в каком угодно конечном интервале (a, b). В частности, отсюда следует, что

$$\text{(b)} \qquad \sqrt{npq}\, P_n(np + x_n \sqrt{npq}) \to \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

если $n \to \infty$ и $x_n \to x$.

Фигурирующие здесь числа x_n не могут быть какими угодно. При каждом данном n они должны быть такими, чтобы выражение

$$m = np + x_n \sqrt{npq}$$

было целым числом, лежащим между 0 и n, так что

$$x_n = \frac{m - np}{\sqrt{npq}} \qquad (m = 0, 1, \ldots, n).$$

Но число x может быть каким угодно. В самом деле, только что написанное равенство показывает, что при определенном индексе

---

[1]) Мы назовем относящиеся сюда теоремы, как это принято делать, теоремами Лапласа. На самом деле они были в существенных чертах много раньше Лапласа открыты Муавром.

[2]) Двойная стрелка $\rightrightarrows$ здесь и в дальнейшем обозначает равномерную сходимость.

и наименьшее из чисел x_n (соответствующее $m=0$) есть $-\frac{\sqrt{np}}{q}$, наибольшее (соответствующее $m=n$) есть $+\frac{\sqrt{nq}}{\sqrt{p}}$ и расстояние между двумя соседними числами x_n (соответствующими m и $m+1$) есть $\frac{1}{\sqrt{npq}}$. Таким образом при $n\to\infty$ наименьшее из чисел x_n стремится к $-\infty$, наибольшее к $+\infty$, и расстояние между двумя соседними к нулю. Отсюда ясно, что множество всех чисел x_n со всевозможными индексами $n=0, 1, 2, \ldots$ всюду плотно на прямой $-\infty<x<+\infty$. Следовательно, любое число x является пределом чисел x_n.

Известно несколько доказательств локальной теоремы Лапласа. Мы приведем прежде всего классическое доказательство, опирающееся на формулу Стирлинга.

Первое доказательство. — Введем краткое обозначение для величины, фигурирующей в формулировке теоремы: положим

$$\Pi_n(x_n)=\sqrt{npq}\,P_n(np+x_n\sqrt{npq}).$$

Так как выражение под знаком P_n равно m, а

$$P_n(m)=\frac{n!}{m!\,(n-m)!}p^mq^{n-m},$$

то

$$\Pi_n(x_n)=\sqrt{npq}\frac{n!}{m!\,(n-m)!}p^mq^{n-m}.$$

Применим теперь формулу Стирлинга [1]). Она гласит, что

$$n!=\sqrt{2\pi}\,n^n\sqrt{n}e^{-n}(1+\alpha_n),$$

где $\alpha_n\to 0$ при $n\to\infty$; следовательно,

$$\Pi_n(x_n)=\frac{\sqrt{npq}}{\sqrt{2\pi}}\frac{n^n\sqrt{n}}{m^m\sqrt{m}\,(n-m)^{n-m}\sqrt{n-m}}p^mq^{n-m}\frac{1+\alpha_n}{(1+\alpha_m)(1+\alpha_{n-m})}.$$

То же самое для дальнейшего удобнее представить в таком виде:

$$\Pi_n(x_n)=\frac{1}{\sqrt{2\pi}}R_n(x_n)\,S_n(x_n)\,T_n(x_n), \tag{31}$$

где

$$R_n(x_n)=\left(\frac{np}{m}\right)^m\left(\frac{nq}{n-m}\right)^{n-m},$$

$$S_n(x_n)=\sqrt{\frac{np}{m}}\sqrt{\frac{nq}{n-m}},$$

$$T_n(x_n)=\frac{1+\alpha_n}{(1+\alpha_m)(1+\alpha_{n-m})}.$$

---

[1]) Вывод формулы Стирлинга приведен в Добавлении II.

Мы изучим изменение при $n\to\infty$ и $a\leqslant x_n\leqslant b$ каждого из множителей $R_n(x_n)$, $S_n(x_n)$, $T_n(x_n)$ в отдельности.

Множитель $R_n(x_n)$.— Мы имеем

$$m=np+x_n\sqrt{npq}, \tag{32}$$

$$n-m=nq-x_n\sqrt{npq}. \tag{33}$$

Следовательно,

$$R_n(x_n)=\left(\frac{np}{np+x_n\sqrt{npq}}\right)^m\left(\frac{nq}{nq-x_n\sqrt{npq}}\right)^{n-m}=$$
$$=\left(1+x_n\frac{\sqrt{q}}{\sqrt{np}}\right)^{-m}\left(1-x_n\frac{\sqrt{p}}{\sqrt{nq}}\right)^{-(n-m)}.$$

Найдем логарифм этого последнего выражения, пользуясь известной формулой

$$\log(1+z)=z-\frac{z^2}{2}[1+\omega(z)],$$

где $\omega(z)\to 0$ при $z\to 0$. Мы получим

$$\log R_n(x_n)=-m\left[x_n\frac{\sqrt{q}}{\sqrt{np}}-x_n^2\frac{q}{2np}(1+\omega')\right]-$$
$$-(n-m)\left[-x_n\frac{\sqrt{p}}{\sqrt{nq}}-x_n^2\frac{p}{2nq}(1+\omega'')\right],$$

где $\omega'=\omega\left(x_n\frac{\sqrt{q}}{\sqrt{np}}\right)$ и $\omega''=\omega\left(x_n\frac{\sqrt{p}}{\sqrt{nq}}\right)$ и, следовательно, если $n\to\infty$ и $a\leqslant x_n\leqslant b$, то ω' и ω'' равномерно стремятся к нулю. Выразив здесь m и $n-m$ по формулам (32) и (33), мы получим

$$\log R_n(x_n)=-x_n^2q+x_n^2\frac{q}{2}(1+\omega')+x_n^3\frac{q\sqrt{q}}{2\sqrt{np}}(1+\omega')-$$
$$-x_n^2p+x_n^2\frac{p}{2}(1+\omega'')-x_n^3\frac{p\sqrt{p}}{2\sqrt{nq}}(1+\omega'').$$

Отсюда, принимая во внимание, что $p+q=1$, мы получим

$$\log R_n(x_n)=-\frac{x_n^2}{2}+\frac{x_n^2}{2}(q\omega'+p\omega'')+$$
$$+\frac{x_n^3}{2\sqrt{n}}\left[\frac{q\sqrt{q}}{\sqrt{p}}(1+\omega')+\frac{p\sqrt{p}}{\sqrt{q}}(1+\omega'')\right].$$

Таким образом если $n\to\infty$ и $a\leqslant x_n\leqslant b$, то

$$\log R_n(x_n)-\left(-\frac{x_n^2}{2}\right)\rightrightarrows 0,$$

а следовательно, и

$$R_n(x_n):e^{-\frac{x_n^2}{2}}\rightrightarrows 1. \tag{34}$$

Множитель $S_n(x_n)$. — Выразив m и $n-m$ по формулам (32) и (33), мы получим

$$S_n(x_n)=\sqrt{\frac{1}{1+x_n\frac{\sqrt{q}}{\sqrt{np}}}}\sqrt{\frac{1}{1-x_n\frac{\sqrt{p}}{\sqrt{nq}}}}.$$

Отсюда ясно, что если $n\to\infty$ и $a\leqslant x_n\leqslant b$, то

$$S_n(x_n)\rightrightarrows 1. \tag{35}$$

Множитель $T_n(x_n)$. — Формулы (32) и (33) показывают, что если $n\to\infty$ и $a\leqslant x_n\leqslant b$, то m и $n-m$ равномерно стремятся к бесконечности. Следовательно, не только $\alpha_n\to 0$, но и $\alpha_m\rightrightarrows 0$ и $\alpha_{n-m}\rightrightarrows 0$. Следовательно,

$$T_n(x_n)\rightrightarrows 1. \tag{36}$$

Сопоставляя (31), (34), (35) и (36), мы видим, что если $n\to\infty$ и $a\leqslant x_n\leqslant b$, то

$$\Pi_n(x_n):\frac{1}{\sqrt{2\pi}}e^{-\frac{x_n^2}{2}}\rightrightarrows 1,$$

что и доказывает основное утверждение (a) нашей теоремы.

Указанное в формулировке теоремы следствие (b) получается отсюда очевидным образом.

Чтобы получить наглядное представление о том, в какой мере можно пользоваться асимптотической формулой Лапласа при конечных n, т. е. заменять биномиальный закон нормальной кривой при вычислении вероятностей для конечных n, приведем пример. Для простоты рассмотрим случай $p=q=\frac{1}{2}$ и возьмем такие n, при которых возможно значение $x_n=1$; такими могут быть, например, $n=25, 100, 400, 1156$. Положим

$$P_n=[P_n(np+x_n\sqrt{npq})]_{\substack{p=q=\frac{1}{2}\\ x_n=1}};$$

это — вероятность того, что μ будет удовлетворять равенству

$$\mu=\frac{1}{2}n+\frac{1}{2}\sqrt{n}\qquad\left(\text{при } p=\frac{1}{2}\right).$$

Положим далее

$$Q_n=\left[\frac{1}{\sqrt{2\pi npq}}e^{-\frac{x_n^2}{2}}\right]_{\substack{p=q=\frac{1}{2}\\ x_n=1}}.$$

Согласно локальной теореме Лапласа отношение

$$\frac{P_n}{Q_n}=\frac{\Pi_n(1)}{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}}}$$

должно стремиться к единице, когда n стремится к бесконечности. Вычисление при вышеназванных значениях n дает:

n	P_n	Q_n	$P_n - Q_n$	$\frac{P_n}{Q_n}$
25	0,09742	0,09679	0,00063	1,0065
100	0,04847	0,04839	0,00008	1,0030
400	0,024207	0,024194	0,000013	1,0004
1156	0,014236	0,014234	0,000002	1,0001

33. Второе доказательство локальной теоремы Лапласа, на возможность которого мы хотим здесь указать, состоит в том, что классическим путем аналитических преобразований доказывается только фигурирующее в формулировке теоремы утверждение (b), а основное утверждение (a) выводится из него просто на основании общих теоретико-функциональных соображений.

При этом доказательство утверждения (b) можно провести также с помощью формулы Стирлинга и вообще тем же путем, каким в первом доказательстве было получено утверждение (a); но отдельные этапы можно упростить здесь в том смысле, чтобы нигде не прибегать к понятию и свойствам равномерной сходимости, а лишь к понятию и свойствам обыкновенной сходимости.

Как и в первом доказательстве, мы полагаем

$$\Pi_n(x_n) = \sqrt{npq}\, P_n(np + x_n\sqrt{npq}),$$
$$m = np + x_n\sqrt{npq},$$
$$n - m = nq - x_n\sqrt{npq},$$

и, применяя формулу Стирлинга, выводим отсюда, что

$$\Pi_n(x_n) = \frac{1}{\sqrt{2\pi}} R_n(x_n)\, S_n(x_n)\, T_n(x_n),$$

где

$$R_n(x_n) = \left(\frac{np}{m}\right)^m \left(\frac{nq}{n-m}\right)^{n-m},$$

$$S_n(x_n) = \sqrt{\frac{np}{m}}\sqrt{\frac{nq}{n-m}},$$

$$T_n(x_n) = \frac{1+\alpha_n}{(1+\alpha_m)(1+\alpha_{n-m})}$$

($\alpha_n \to 0$ при $n \to \infty$). Теми же преобразованиями, что и в первом доказательстве, мы получаем

$$\log R_n(x_n) = -\frac{x_n^2}{2} + \frac{x_n^2}{2}(q\omega' + p\omega'') + \frac{x_n^3}{2\sqrt{n}}\left[\frac{q\sqrt{q}}{\sqrt{p}}(1+\omega') + \frac{p\sqrt{p}}{\sqrt{q}}(1+\omega'')\right]$$

($\omega' \to 0$ и $\omega'' \to 0$ при $n \to \infty$). Отсюда ясно, что если $n \to \infty$ и $x_n \to x$, то

(34 bis) $$\log R_n(x_n) \to -\frac{x^2}{2}$$

и, следовательно,

$$R_n(x_n) \to e^{-\frac{x^2}{2}}.$$

Далее, пользуясь выражениями для m и $n-m$, мы имеем

$$S_n(x_n) = \sqrt{\frac{1}{1+x_n\frac{\sqrt{q}}{\sqrt{np}}}}\sqrt{\frac{1}{1-x_n\frac{\sqrt{p}}{\sqrt{nq}}}},$$

откуда ясно, что если $n \to \infty$ и $x_n \to x$, то

$$S_n(x_n) \to 1. \tag{35 bis}$$

Наконец, пользуясь теми же выражениями для m и $n-m$, мы видим, что если $n \to \infty$ и $x_n \to x$, то $m \to \infty$ и $n-m \to \infty$, следовательно, не только $\alpha_n \to 0$, но и $\alpha_m \to 0$ и $\alpha_{n-m} \to 0$; поэтому

$$T_n(x_n) \to 1. \tag{36 bis}$$

Сопоставляя же выражения для $\Pi_n(x_n)$ с (34 bis), (35 bis) и (36 bis), мы видим, что если $n \to \infty$ и $x_n \to x$, то

$$\sqrt{npq}\, P_n(np + x_n\sqrt{npq}) \to \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Это и есть утверждение (b).

Чтобы перейти теперь к основному утверждению (a), остается применить следующую общую лемму о равномерной сходимости:

Пусть дана последовательность функций

$$f_1(x),\ f_2(x),\ \ldots,\ f_n(x),\ \ldots \tag{S}$$

и непрерывная функция

$$g(x),$$

определенные на всей прямой $-\infty < x < +\infty$, *и пусть при* $n \to \infty$ и $x_n \to x$ *всегда* $f_n(x_n) \to g(x)$, *тогда отношение*

$$\frac{f_n(x_n)}{g(x_n)}$$

равномерно стремится к единице, если $n \to \infty$, *а* x_n *варьирует в каком угодно конечном интервале* (a, b), *в котором* $|g(x)| > c > 0$.

Достаточно положить

$$f_n(x) = \sqrt{npq}\, P_n(np + x\sqrt{npq}),\ g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Доказательство леммы. Если бы утверждение леммы было неверно, то нашлась бы положительная постоянная ε и бесконечная последовательность индексов $n_1, n_2, \ldots, n_k, \ldots$ и соответствующих им точек

$$x_{n_1}, x_{n_2}, \ldots, x_{n_k}, \ldots \tag{S}$$

некоторого интервала $a \leqslant x \leqslant b$, где $|g(x)| > c > 0$, для которых имело бы место неравенство

$$\left|\frac{f_{n_k}(x_{n_k})}{g(x_{n_k})} - 1\right| \geqslant \varepsilon,$$

а следовательно, и неравенство

$$|f_{n_k}(x_{n_k}) - g(x_{n_k})| > \varepsilon.$$

Последовательность (S) имела бы на интервале $a \leqslant x \leqslant b$ по крайней мере одну предельную точку x и, следовательно, нашлась бы подпоследовательность

$$x_{n_{k_1}}, x_{n_{k_2}}, \ldots, x_{n_{k_p}}, \ldots$$

последовательности (S), сходящаяся к этой точке: $x_{n_{k_p}} \to x$. Только что написанное неравенство имело бы место, в частности, и для этой подпоследовательности:

(37) $$|f_{n_{k_p}}(x_{n_{k_p}}) - g(x_{n_{k_p}})| > \varepsilon.$$

Но по условию

(38) $$f_{n_{k_p}}(x_{n_{k_p}}) \to g(x)$$

и, по условию же, $g(x)$ — непрерывная функция, так что

(39) $$g(x_{n_{k_p}}) \to g(x).$$

Однако

$$|f_{n_{k_p}}(x_{n_{k_p}}) - g(x_{n_{k_p}})| \leqslant |f_{n_{k_p}}(x_{n_{k_p}}) - g(x)| + |g(x) - g(x_{n_{k_p}})|.$$

Поэтому из (38) и (39) вытекает, что

$$|f_{n_{k_p}}(x_{n_{k_p}}) - g(x_{n_{k_p}})| \to 0.$$

Это противоречит неравенству (37). Полученное противоречие и доказывает лемму.

§ 10. Интегральная теорема Лапласа.

34. Мы назвали только что доказанную теорему локальной в отличие от другой, не менее важной теоремы Лапласа, которую мы назовем интегральной. К изложению этой последней мы сейчас и приступим.

Возьмем снова графики вероятностей $P_n(m)$ того, что величина μ будет удовлетворять равенству $\mu = m$, изображенные на черт. 1 (стр. 38). Подвергнем эти графики тем же преобразованиям, которые были описаны в начале п. 31, т. е. заменим величины $P_n(m)$ последовательно величинами $P_n(l_n + m')$, $P_n(l_n + m''k_n)$ и, наконец, $a_n P_n(l_n + m''k_n)$, где мы на этот раз сразу положим $a_n = k_n = \sqrt{npq}$ и $l_n = np$. Преобразованные графики мы будем обозначать через Γ_n.

Кроме того, мы еще пополним графики Γ_n так, чтобы получилась ступенчатая линия $y = \Pi_n(x)$. А именно, пусть

$$x_{n,m} = \frac{m - np}{\sqrt{npq}} \qquad (m = 0, 1, \ldots, n)$$

— абсциссы, фигурирующие на графике Γ_n, и пусть

$$y = \Pi_n(x_{n,m}) = \sqrt{npq}\, P_n(np + x_{n,m}\sqrt{npq})$$

— соответствующие этим абсциссам ординаты на графике Γ_n. Мы определим ординаты ступенчатой линии $y = \Pi_n(x)$, полагая в каждом интервале $x_{n,m} \leqslant x < x_{n,m+1}$, где $m = 0, 1, \ldots, n-1$:

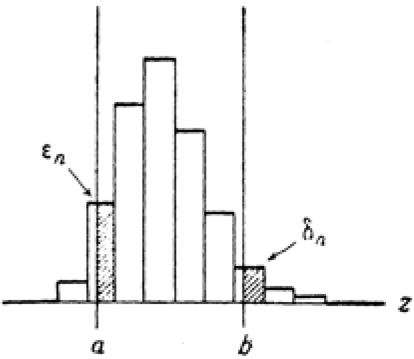
$$y = \Pi_n(x_{n,m}),$$

а для $x < x_{n,0}$ и $x \geqslant x_{n,n}$:

$$y = 0.$$

Для $n = 9$ и $p = \frac{1}{3}$ эта ступенчатая линия изображена на черт. 3.

Пусть $\Sigma_n(a, b)$ — площадь, заключенная между этой ступенчатой линией, осью абсцисс и двумя прямыми $x = a$ и $x = b$, так что



Черт. 3.

$$\Sigma_n(a, b) = \int_a^b \Pi_n(z)\, dz.$$

Площадь эта почти равна сумме площадей прямоугольников с основаниями $x_{n,m+1} - x_{n,m}$ и с высотами $\Pi_n(x_{n,m})$, где $a < x_{n,m} < b$; разница между ними не превосходит $\delta_n + \varepsilon_n$, где δ_n и ε_n — площади прямоугольников, заштрихованных на черт. 3. Точнее, если обозначить через $\underline{x}_n = x_{n,\underline{m}_n}$ точку $x_{n,m}$, ближайшую справа к a, и через $\overline{x}_n = x_{n,\overline{m}_n}$ — точку $x_{n,m}$, ближайшую справа к b, то

$$\Sigma_n(a, b) = \sum_{m=\underline{m}_n}^{\overline{m}_n - 1} \Pi_n(x_{n,m})(x_{n,m+1} - x_{n,m}) - \delta_n + \varepsilon_n,$$

где

$$\delta_n = \Pi_n(x_{n,\overline{m}_n - 1})(x_{m,\overline{m}_n} - b), \quad \varepsilon_n = \Pi_n(x_{n,\underline{m}_n - 1})(x_{n,\underline{m}_n} - a).$$

Но основание $x_{n,m+1} - x_{n,m}$ равно $\frac{1}{\sqrt{npq}}$, а высота $\Pi_n(x_{n,m})$ равна умноженной на $\sqrt{npq}$ вероятности $P_n(np + x_{n,m}\sqrt{npq})$ того, что величина μ будет удовлетворять равенству

$$\mu = np + x_{n,m}\sqrt{npq},$$

так что

$$\Pi_n(x_{n,m})(x_{n,m+1} - x_{n,m}) = P_n(np + x_{n,m}\sqrt{npq}).$$

В силу же теоремы сложения вероятностей сумма этих вероятностей для $a < x_{n,m} < b$, равна вероятности $P_n(a, b)$ того, что величина μ будет удовлетворять неравенствам

$$np + a\sqrt{npq} < \mu < np + b\sqrt{npq},$$

так что

$$\sum_{m=\underline{m}_n}^{\overline{m}_n - 1} \Pi_n(x_{n,m})(x_{n,m+1} - x_{n,m}) = P_n(a, b).$$

Таким образом площадь $\Sigma_n(a, b)$ почти равна названной вероятности $P_n(a, b)$:

(α) $$\Sigma_n(a, b) = P_n(a, b) - \delta_n + \varepsilon_n.$$

При этом, когда $n \to \infty$, то

(β) $$\delta_n \to 0 \quad \text{и} \quad \varepsilon_n \to 0.$$

Действительно, поскольку, в силу локальной теоремы Лапласа, графики Γ_n равномерно сходятся к нормальной кривой, постольку ординаты $\Pi_n(x_{n, \overline{m}_n - 1})$ и $\Pi_n(x_{n, \underline{m}_n - 1})$, очевидно, ограничены; разности же $x_{n, \overline{m}_n} - b$ и $x_{n, \underline{m}_n} - a$ не превосходят $\frac{1}{\sqrt{npq}}$, следовательно, стремятся к нулю.

Наконец, пусть $S(a, b)$ — площадь, заключенная между нормальной кривой и прямыми $x = a$ и $x = b$, так что

$$S(a, b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{z^2}{2}} dz.$$

Поскольку, в силу локальной теоремы Лапласа, графики Γ_n равномерно сходятся к нормальной кривой, постольку почти очевидно, что и построенные нами ступенчатые линии, пополняющие графики Γ_n равномерно сходятся к нормальной кривой. Отсюда, в свою очередь, следует, что

(γ) $$\Sigma_n(a, b) \to S(a, b).$$

Из соотношений (α), (β) и (γ) немедленно вытекает, что

$$P_n(a, b) \to S(a, b).$$

Это последнее утверждение и составляет сущность интегральной теоремы Лапласа, которую мы сейчас точно сформулируем и докажем.

35. Интегральная теорема Лапласа. — *Пусть μ — число наступлений события при n независимых испытаниях, при каждом из которых вероятность этого события равна постоянной p, отличной от нуля и единицы, и пусть $P_n(a, b)$ есть вероятность того, что величина μ будет удовлетворять неравенствам*

$$np + a\sqrt{npq} < \mu < np + b\sqrt{npq} \qquad (q = 1 - p).$$

Тогда

(a) $$P_n(a_n, b_n) - \frac{1}{\sqrt{2\pi}} \int_{a_n}^{b_n} e^{-\frac{z^2}{2}} dz \rightrightarrows 0,$$

если $n \to \infty$, *а* a_n *и* $b_n > a_n$ *варьируют как угодно между* $-\infty$ *и* $+\infty$. *В частности, для постоянных* a *и* $b > a$

(b) $$P_n(a, b) \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{z^2}{2}} dz,$$

если $n \to \infty$.

Первое доказательство. — Мы будем пользоваться обозначениями, введенными в п. 34. Только теперь мы будем обозначать через $\underline{x}_n = x_{n, \underline{m}_n}$ точку $x_{n,m}$, ближайшую справа к a_n, и через $\overline{x}_n = x_{n, \overline{m}_n}$ — точку $x_{n,m}$, ближайшую справа к b_n.

Доказательство мы проведем в два приема. Сначала докажем теорему в предположении, что a_n и b_n варьируют в каком-нибудь конечном интервале $(-C, C)$, и затем распространим результат на общий случай изменения a_n и b_n.

Первая часть доказательства. — Пусть

$$-C \leqslant a_n < b_n \leqslant C.$$

1° По теореме сложения вероятностей мы имеем

$$P_n(a_n, b_n) = \sum_{m=\underline{m}_n}^{\overline{m}_n - 1} \Pi_n(x_{n,m}) \frac{1}{\sqrt{npq}}.$$

А так как $\frac{1}{\sqrt{npq}} = x_{n,m+1} - x_{n,m}$ и при $\underline{m}_n \leqslant m \leqslant \overline{m}_n$ имеет место неравенство $|x_{n,m}| \leqslant C$, то по локальной теореме Лапласа мы имеем отсюда

$$P_n(a_n, b_n) = \sum_{m=\underline{m}_n}^{\overline{m}_n - 1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_{n,m}^2}{2}} (x_{n,m+1} - x_{n,m})(1 + \pi_{n,m}),$$

где

$$\pi_{n,m} \rightrightarrows 0.$$

Следовательно,

(40) $$P_n(a_n, b_n) = \frac{1}{\sqrt{2\pi}} \left[\sum_{m=\underline{m}_n}^{\overline{m}_n - 1} e^{-\frac{x_{n,m}^2}{2}} (x_{n,m+1} - x_{n,m}) \right] (1 + \pi_n),$$

где

$$\pi_n \rightrightarrows 0,$$

так как $|\pi_n| \leqslant \max |\pi_{n,m}| \; (\underline{m}_n \leqslant m \leqslant \overline{m}_n - 1)$.

2° Пользуясь обозначением

$$S(a, b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{z^2}{2}} dz,$$

мы имеем

$$S(a_n, b_n) = S(\underline{x}_n, \overline{x}_n) + S(a_n, \underline{x}_n) - S(b_n, \overline{x}_n).$$

А так как $\underline{x}_n - a_n \leqslant \frac{1}{\sqrt{npq}}$, $\overline{x}_n - b_n \leqslant \frac{1}{\sqrt{npq}}$ и $e^{-\frac{z^2}{2}} \leqslant 1$, то мы имеем отсюда

$$S(a_n, b_n) = S(\underline{x}_n, \overline{x}_n) + \rho_n, \tag{41}$$

где

$$\rho_n \rightrightarrows 0,$$

потому что $|\rho_n| \leqslant \frac{2}{\sqrt{2\pi npq}}$.

3° Мы имеем

$$S(\underline{x}_n, \overline{x}_n) = \frac{1}{\sqrt{2\pi}} \sum_{m=\underline{m}_n}^{\overline{m}_n - 1} \int_{x_{n,m}}^{x_{n,m+1}} e^{-\frac{z^2}{2}} dz.$$

Но

$$\int_{x_{n,m}}^{x_{n,m+1}} e^{-\frac{z^2}{2}} dz = e^{-\frac{x^2_{n,m}}{2}} (x_{n,m+1} - x_{n,m})(1 + \sigma_{n,m}),$$

где

$$1 + \sigma_{n,m} = e^{x^2_{n,m} - \xi^2_{n,m}},$$

где, в свою очередь,

$$x_{n,m} < \xi_{n,m} < x_{n,m+1},$$

так что при $|x_{n,m}| \leqslant C$ мы имеем $|x^2_{n,m} - \xi^2_{n,m}| < \frac{2C}{\sqrt{npq}}$ и, следовательно,

$$\sigma_{n,m} \rightrightarrows 0.$$

Поэтому

$$S(\underline{x}_n, \overline{x}_n) = \frac{1}{\sqrt{2\pi}} \left[\sum_{m=\underline{m}_n}^{\overline{m}_n - 1} e^{-\frac{x^2_{n,m}}{2}} (x_{n,m+1} - x_{n,m}) \right] (1 + \sigma_n), \tag{42}$$

где

$$\sigma_n \rightrightarrows 0,$$

так как $|\sigma_n| \leqslant \max |\sigma_{n,m}|$ $(\underline{m}_n \leqslant m \leqslant \overline{m}_n - 1)$.

4° Из (40) и (42) мы получаем

$$\frac{P_n(a_n, b_n)}{S(\underline{x}_n, \overline{x}_n)} = \frac{1 + \pi_n}{1 + \sigma_n}$$

и потому

$$\frac{P_n(a_n, b_n)}{S(\underline{x}_n, \overline{x}_n)} \rightrightarrows 1.$$

А так как $S(\underline{x}_n, \overline{x}_n) < 1$, то при этом и

$$P_n(a_n, b_n) - S(\underline{x}_n, \overline{x}_n) \rightrightarrows 0.$$

С другой стороны, по (41)

$$S(a_n, b_n) - S(\underline{x}_n, \overline{x}_n) \rightrightarrows 0.$$

Таким образом и

$$P_n(a_n, b_n) - S(a_n, b_n) \rightrightarrows 0,$$

что и требовалось доказать.

Вторая часть доказательства. — Распространим теперь этот результат на общий случай изменения a_n и $b_n > a_n$, когда они варьируют как угодно между $-\infty$ и $+\infty$.

Очевидно, что каков бы ни был конечный интервал $(-C, C)$ и как бы ни были расположены числа a_n и b_n относительно этого интервала, всегда

$$\begin{aligned}(43)\quad |S(a_n, b_n) - P_n(a_n, b_n)| \leqslant \\ \leqslant S(-\infty, -C) + P_n(-\infty, -C) + |S(\alpha_n, \beta_n) - P_n(\alpha_n, \beta_n)| + \\ + S(C, +\infty) + P_n(C, +\infty),\end{aligned}$$

где α_n и β_n — некоторые числа [1]), удовлетворяющие условию

$$-C \leqslant \alpha_n < \beta_n \leqslant C.$$

Возьмем произвольное положительное число ε и выберем какое-нибудь положительное число C, для которого имеет место неравенство

$$(43') \qquad S(-\infty, -C) + S(C, +\infty) = 1 - S(-C, C) < \frac{\varepsilon}{4}.$$

В силу первой части нашего доказательства найдется также такое число N, что при $n > N$ будет иметь место неравенство

$$(43'') \qquad |S(\alpha_n, \beta_n) - P_n(\alpha_n, \beta_n)| < \frac{\varepsilon}{4},$$

каковы бы ни были числа α_n и β_n, удовлетворяющие условию

$$-C \leqslant \alpha_n < \beta_n \leqslant C.$$

Полагая $\alpha_n = -C$ и $\beta_n = C$, мы получим отсюда также неравенство

$$\begin{aligned}(43''')\quad P_n(-\infty, -C) + P_n(C, +\infty) = 1 - P_n(-C, C) = \\ = [1 - S(-C, C)] + [S(-C, C) - P_n(-C, C)] < \frac{\varepsilon}{2}.\end{aligned}$$

Сопоставим теперь неравенство (43) с неравенствами (43′), (43″) и (43‴); мы получим для $n > N$ неравенство

$$|S(a_n, b_n) - P_n(a_n, b_n)| < \varepsilon,$$

что и требовалось доказать.

---

[1]) В зависимости от расположения чисел a_n и b_n относительно интервала $(-C, C)$, α_n может быть равно a_n или $-C$, а β_n может быть равно b_n или C.

Таким образом мы убедились в справедливости утверждения (a) нашей теоремы. Утверждение (b) вытекает отсюда непосредственно.

36. Второе доказательство интегральной теоремы Лапласа, которое мы хотим здесь привести, будет просто уточнением и пополнением тех соображений, которые мы привели в п. 34 в качестве подготовительных к пониманию сущности теоремы. В связи с этим оно будет отличаться от первого доказательства тем, что сначала будет доказываться утверждение (b), основное же утверждение (a) будет вытекать из него с помощью некоторых добавочных соображений.

Утверждение (b) гласит, что при $n \to \infty$

$$P_n(a, b) \to S(a, b),$$

где, как всюду выше,

$$S(a, b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{z^2}{2}} dz.$$

Это утверждение было уже почти доказано нами в п. 34; оно вытекало из сопоставления трех соотношений:

$$(\alpha) \qquad \Sigma_n(a, b) = P_n(a, b) - \delta_n + \varepsilon_n,$$

$$(\beta) \qquad \delta_n \to 0 \quad \text{и} \quad \varepsilon_n \to 0,$$

$$(\gamma) \qquad \Sigma_n(a, b) \to S(a, b),$$

из которых в п. 34 не было строго обосновано только последнее. Это обоснование и составит первую часть нашего доказательства.

Первая часть доказательства. — В силу локальной теоремы Лапласа, если $n \to \infty$ и x_n варьирует в каком угодно конечном интервале, то [1])

$$\Pi_n(x_n) - \frac{1}{\sqrt{2\pi}} e^{-\frac{x_n^2}{2}} \rightrightarrows 0.$$

Если, в частности, взять любую точку x некоторого конечного интервала (a, b) и обозначить через x_n ближайшую к ней слева точку $x_{n,m}$, то эти последние точки x_n будут находиться в конечном интервале $\left(a - \frac{1}{\sqrt{pq}}, b\right)$ и, следовательно, только что написанное соотношение будет для них справедливо. Затем, в силу непрерывности функции $e^{-\frac{x^2}{2}}$

---

[1]) В силу локальной теоремы Лапласа

$$\Pi_n(x_n) : \frac{1}{\sqrt{2\pi}} e^{-\frac{x_n^2}{2}} \rightrightarrows 1$$

и, сверх того, функция $e^{-\frac{x^2}{2}}$ ограничена; отсюда и вытекает написанное.

и неравенства $|x-x_n|<\frac{1}{\sqrt{npq}}$, в интервале (a, b) будет иметь место соотношение

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{x_n^2}{2}} \rightrightarrows \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

Наконец, по самому определению функции $\Pi_n(x)$,

$$\Pi_n(x)=\Pi_n(x_n).$$

Поэтому

$$\Pi_n(x)-\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}=\left[\Pi_n(x_n)-\frac{1}{\sqrt{2\pi}}e^{-\frac{x_n^2}{2}}\right]+\left[\frac{1}{\sqrt{2\pi}}e^{-\frac{x_n^2}{2}}-\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right],$$

и, следовательно,

$$\Pi_n(x) \rightrightarrows \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

Отсюда непосредственно вытекает справедливость соотношения (γ):

$$\sum_n(a, b)\to S(a, b).$$

Тем самым утверждение (b), вытекающее из (α), (β) и (γ), доказано.

Во второй части доказательства нам предстоит установить справедливость утверждения (a). Оно же, очевидно, равносильно утверждению, что при $n\to\infty$

$$P_n(a, b)\to S(a, b)$$

равномерно относительно a и $b>a$, каковы бы они ни были. Это последнее мы сейчас и покажем.

Вторая часть доказательства. — Здесь мы можем опираться на следующую общую лемму о равномерной сходимости:

Пусть дана последовательность неубывающих функций

$$\text{(S)}\qquad f_1(x),\ f_2(x),\ \ldots,\ f_n(x),\ \ldots$$

и непрерывная неубывающая функция

$$g(x),$$

определенные на всей прямой $-\infty<x<+\infty$, *пусть последовательность* (S) *сходится к функции* $g(x)$ *и пусть, кроме того, при каждом* n *имеет место равенство* [1])

$$f_n(+\infty)-f_n(-\infty)=g(+\infty)-g(-\infty)<+\infty;$$

[1]) Символы $f(+\infty)$ и $f(-\infty)$ здесь и в дальнейшем обозначают соответственно $\lim\limits_{x\to+\infty} f(x)$ и $\lim\limits_{x\to-\infty} f(x)$.

тогда последовательность (S) *сходится к функции* $g(x)$ *равномерно на всей прямой* $-\infty < x < +\infty$.

Положим для $x \geqslant 0$

$$f_n(x) = P_n(0, x), \qquad g(x) = S(0, x)$$

и для $x < 0$

$$f_n(x) = -P_n(x, 0), \quad g(x) = -S(x, 0).$$

Определенные таким образом функции $f_n(x)$ и $g(x)$ удовлетворяют всем условиям леммы. Действительно, $f_n(x) \to g(x)$ при всяком x в силу первой части нашего доказательства. Затем $f_n(+\infty) - f_n(-\infty) = 1$, так как эта разность равна $P_n(-\infty, +\infty)$, т. е. равна вероятности достоверного события, и $g(+\infty) - g(-\infty) = 1$, так как эта разность равна $S(-\infty, +\infty)$.

Таким образом, в силу леммы, для всякой положительной постоянной ε найдется такое число N, что при $n > N$ на всей прямой $-\infty < x < +\infty$

$$|f_n(x) - g(x)| < \frac{\varepsilon}{2},$$

и, следовательно, при любых a и $b > a$

$$|P_n(a, b) - S(a, b)| = |[f_n(b) - f_n(a)] - [g(b) - g(a)]| \leqslant$$
$$\leqslant |f_n(b) - g(b)| + |f_n(a) - g(a)| < \varepsilon.$$

Тем самым доказано и утверждение (a).

Доказательство леммы. — Возьмем произвольную положительную постоянную ε и подберем положительное число C таким образом, чтобы было

$$g(+\infty) - g(C) + g(-C) - g(-\infty) < \frac{\varepsilon}{2},$$

а следовательно, и

$$f_n(+\infty) - g(C) + g(-C) - f_n(-\infty) < \frac{\varepsilon}{2}.$$

Рассмотрим точки x интервала $(-C, C)$. Отметим на этом интервале конечное число $s+1$ точек x_k, включая $x_0 = -C$ и $x_s = C$, таким образом, чтобы при каждом k было

$$g(x_k) - g(x_{k-1}) < \frac{\varepsilon}{2}.$$

Найдется такое число N, что при $n > N$ во всех этих точках x_k будет

$$|f_n(x_k) - g(x_k)| < \frac{\varepsilon}{2}. \tag{43 bis}$$

Но каждая точка x интервала $(-C, C)$ находится между двумя точками x_{k-1} и x_k, и потому

$$f_n(x_{k-1}) - g(x_k) \leqslant f_n(x) - g(x) \leqslant f_n(x_k) - g(x_{k-1}),$$

откуда при $n > N$

$$-\varepsilon < [f_n(x_{k-1}) - g(x_{k-1})] + [g(x_{k-1}) - g(x_k)] \leqslant$$
$$\leqslant f_n(x) - g(x) \leqslant$$
$$\leqslant [f_n(x_k) - g(x_k)] + [g(x_k) - g(x_{k-1})] < \varepsilon.$$

Таким образом для каждой точки x интервала $(-C, C)$ при $n > N$ имеет место неравенство

$$|f_n(x) - g(x)| < \varepsilon.$$

Рассмотрим теперь точки x, лежащие правее интервала $(-C, C)$. Для них

$$f_n(C) - g(+\infty) \leqslant f_n(x) - g(x) \leqslant f_n(+\infty) - g(C),$$

следовательно, тем более:

$$[g(-\infty) - g(-C)] + [f_n(C) - g(+\infty)] \leqslant$$
$$\leqslant f_n(x) - g(x) \leqslant$$
$$\leqslant [f_n(+\infty) - g(C)] + [f_n(-C) - f_n(-\infty)],$$

и отсюда, пользуясь неравенством (43 bis), в применении к точкам $x_0 = -C$ и $x_s = C$, при $n > N$

$$-\varepsilon < [g(-\infty) - g(-C)] + [g(C) - g(+\infty)] - \frac{\varepsilon}{2} \leqslant$$
$$\leqslant f_n(x) - g(x) \leqslant$$
$$\leqslant [f_n(+\infty) - g(C)] + [g(-C) - f_n(-\infty)] + \frac{\varepsilon}{2} < \varepsilon.$$

Таким образом и для каждой точки x, лежащей правее интервала $(-C, C)$ при $n > N$ имеет место неравенство

$$|f_n(x) - g(x)| < \varepsilon.$$

Точно так же доказывается, что и для каждой точки x, лежащей левее интервала $(-C, C)$, при $n > N$ имеет место неравенство

$$|f_n(x) - g(x)| < \varepsilon.$$

§ 11. Применения интегральной теоремы Лапласа.

37. Для вычисления вероятностей по закону Гаусса требуются специальные таблицы, так как здесь мы имеем дело с интегралом, который не берется в конечном виде. Обычно пользуются таблицами, составленными для интеграла

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz.$$

Такая таблица приведена в конце книги [1]). Закон Гаусса приводится к интегралу $\Phi(x)$ по следующей очевидной формуле:

$$\frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{z^2}{2}} dz = \Phi(b) - \Phi(a).$$

Таблица составлена только для положительных значений x; для отрицательных значений вычисляют функцию $\Phi(x)$, пользуясь равенством

$$\Phi(-x) = -\Phi(x),$$

---

[1]) Таблица заимствована из книги акад. С. Н. Бернштейна, *Теория вероятностей*, 1934.

которое сразу получается из того, что подинтегральная функция — четная, так что ее график расположен симметрично относительно оси ординат. Наиболее просты вычисления, когда интеграл берется между пределами, различными по знаку, но одинаковыми по абсолютной величине; в этом случае мы имеем

$$\frac{1}{\sqrt{2\pi}}\int_{-a}^{+a} e^{-\frac{z^2}{2}}\,dz = 2\Phi(a).$$

Условимся писать $x_n \sim y_n$, если $(x_n - y_n) \to 0$. Тогда утверждение интегральной теоремы Лапласа может быть записано следующим образом:

$$(44)\qquad P\left(a_n < \frac{\mu - np}{\sqrt{npq}} < b_n\right) \sim \Phi(b_n) - \Phi(a_n).$$

Эта формула имеет смысл приближенного равенства, годного при больших n.

38. В теории и в практике чаще всего приходится иметь дело с линейными возрастающими функциями величины μ, т. е. с величинами

$$\nu = \frac{\mu - l_n}{k_n},$$

где k_n — положительные, l_n — произвольные постоянные. Таковы, например, сама величина μ, величина $\frac{\mu}{n}$ — относительная частота наступления события и т. п.

Пусть $a < b$. Заменим в (44) μ через

$$k_n\nu + l_n$$

и a_n, b_n соответственно через

$$\frac{k_n a + l_n - np}{\sqrt{npq}}, \quad \frac{k_n b + l_n - np}{\sqrt{npq}};$$

мы получим асимптотическое равенство для вероятности того, что величина ν будет находиться в интервале (a, b):

$$(45)\qquad P(a < \nu < b) \sim \Phi\left(\frac{k_n b + l_n - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_n a + l_n - np}{\sqrt{npq}}\right).$$

Например, если $\nu = \mu$, то $k_n = 1$, $l_n = 0$, и мы имеем

$$(46)\qquad P(a < \mu < b) \sim \Phi\left(\frac{b - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np}{\sqrt{npq}}\right).$$

Если $\nu = \frac{\mu}{n}$, то $k_n = n$, $l_n = 0$, и мы имеем

$$(47)\qquad P\left(a < \frac{\mu}{n} < b\right) \sim \Phi\left(\frac{b - p}{\sqrt{pq}}\sqrt{n}\right) - \Phi\left(\frac{a - p}{\sqrt{pq}}\sqrt{n}\right)$$

и т. п.

39. С тем, как применяются полученные нами асимптотические формулы (44) и сл. в теории и на практике, удобнее всего ознакомиться на каком-нибудь конкретном примере. Мы выберем для такого примера величину

$$\nu = \frac{\mu}{n} - p,$$

которая часто встречается в приложениях и будем изучать вероятность $P_n(\varepsilon)$ неравенства

$$\left|\frac{\mu}{n} - p\right| < \varepsilon,$$

где ε — положительная постоянная.

Поставим вопрос об изменении этой вероятности при $n \to \infty$. Так как в данном случае

$$\nu = \frac{\mu - l_n}{k_n},$$

где $k_n = n$ и $l_n = np$, а неравенство, о котором идет речь, равносильно неравенствам

$$a < \nu < b,$$

где $a = -\varepsilon$ и $b = \varepsilon$, то, в силу, (45), мы имеем

$$P_n(\varepsilon) \sim \Phi\left(\frac{\varepsilon}{\sqrt{pq}}\sqrt{n}\right) - \Phi\left(-\frac{\varepsilon}{\sqrt{pq}}\sqrt{n}\right),$$

или, что то же:

$$P_n(\varepsilon) \sim \frac{1}{\sqrt{2\pi}} \int\limits_{-\frac{\varepsilon}{\sqrt{pq}}\sqrt{n}}^{\frac{\varepsilon}{\sqrt{pq}}\sqrt{n}} e^{-\frac{z^2}{2}}\, dz.$$

При $n \to \infty$ пределы интегрирования стремятся соответственно к $-\infty$ и к $+\infty$, так что интеграл стремится к единице. Следовательно, и интересующая нас вероятность $P_n(\varepsilon)$ неравенства $\left|\frac{\mu}{n} - p\right| < \varepsilon$ при любой постоянной ε стремится к единице.

Это показывает, что с возрастанием числа испытаний относительная частота $\frac{\mu}{n}$ наступления события имеет тенденцию безгранично приближаться к вероятности p этого события в отдельном испытании. Для этого важного утверждения, называемого теоремой Бéрнулли, мы приведем ниже, в главе о законе больших чисел, более простое и прямое доказательство.

40. С типичными вычислительными задачами, которые решаются при помощи полученных нами асимптотических формул, удобно ознакомиться на том же примере вероятности $P_n(\varepsilon)$ неравенства $\left|\frac{\mu}{n} - p\right| < \varepsilon$.

Мы только что нашли асимптотическое равенство для этой вероятности:

$$P_n(\varepsilon) \sim \Phi\left(\frac{\varepsilon}{\sqrt{pq}}\sqrt{n}\right) - \Phi\left(-\frac{\varepsilon}{\sqrt{pq}}\sqrt{n}\right),$$

которое можно написать короче так:

$$P_n(\varepsilon) \sim 2\Phi\left(\frac{\varepsilon}{\sqrt{pq}}\sqrt{n}\right). \tag{48}$$

Приведем три типичные задачи, которые решаются при помощи этого равенства (48).

I. Пусть производится 600 испытаний, при каждом из которых вероятность наступления события A равна $\frac{1}{6}$. Спрашивается, какова вероятность того, что относительная частота наступления события A отклонится от $\frac{1}{6}$ меньше, чем на 0,025?

Здесь $n = 600$, $p = \frac{1}{6}$, $q = \frac{5}{6}$, $\varepsilon = 0{,}025$. Искомая вероятность есть вероятность неравенства

$$\left|\frac{\mu}{600} - \frac{1}{6}\right| < 0{,}025.$$

Применяя формулу (48) и взяв нужное значение функции Φ из таблицы, приведенной в конце книги, мы видим, что искомая вероятность равна

$$2\Phi\left(\frac{0{,}025\sqrt{600}}{\sqrt{\frac{1}{6}\cdot\frac{5}{6}}}\right) = 2\Phi(1{,}64) = 0{,}899.$$

II. Пусть производится 600 испытаний, при каждом из которых вероятность наступления события A равна $\frac{1}{6}$. Спрашивается, как велика та относительная частота наступления события A, вероятность отклонения которой от $\frac{1}{6}$ равна 0,999?

Здесь $n = 600$, $p = \frac{1}{6}$, $q = \frac{5}{6}$. Неизвестным является число ε, меньше которого должно быть названное отклонение и которое находится из условия, что вероятность неравенства

$$\left|\frac{\mu}{600} - \frac{1}{6}\right| < \varepsilon$$

должна быть равна 0,999. Применяя формулу (48), мы получим для определения ε уравнение

$$2\Phi\left(\frac{\varepsilon\sqrt{600}}{\sqrt{\frac{1}{6}\cdot\frac{5}{6}}}\right) = 0{,}999.$$

Отсюда, взяв нужное значение функции Φ из таблицы при помощи линейной интерполяции, мы получим

$$\frac{\varepsilon\sqrt{600}}{\sqrt{\frac{1}{6}\cdot\frac{5}{6}}} = 3{,}31.$$

Вычисление дает

$$\varepsilon = 0{,}05.$$

III. Пусть производится ряд испытаний, при каждом из которых вероятность наступления события A равна $\frac{1}{6}$. Спрашивается, сколько надо произвести испытаний, чтобы вероятность того, что относительная частота наступления события A отклонится от $\frac{1}{6}$ меньше чем на 0,025, была равна 0,999?

Здесь $p = \frac{1}{6}$, $q = \frac{5}{6}$, $\varepsilon = 0{,}025$. Число же n неизвестно и должно быть найдено из условия, что вероятность неравенства

$$\left| \frac{\mu}{n} - \frac{1}{6} \right| < 0{,}025$$

равна 0,999. Применяя формулу (48), мы получим для определения n уравнение

$$2\Phi\left(\frac{0{,}025\sqrt{n}}{\sqrt{\frac{1}{6}\cdot\frac{5}{6}}}\right) = 0{,}999.$$

Отсюда, найдя значение аргумента функции Φ, как в предыдущей задаче, мы получим

$$\frac{0{,}025\sqrt{n}}{\sqrt{\frac{1}{6}\cdot\frac{5}{6}}} = 3{,}31.$$

Вычисление дает

$$n = 2420.$$

41. Общая форма закона Гаусса. — Правые части асимптотических равенств, полученных нами для различных вероятностей, могут быть записаны еще в иной форме, неудобной для вычислений, но имеющей теоретическое значение. А именно, положим для краткости

$$S(a, b;\ \alpha, \beta) = \frac{1}{\sqrt{2\pi\beta}} \int_a^b e^{-\frac{(s-\alpha)^2}{2\beta}}\, ds,$$

где α — произвольный, β — положительный параметр. Закон Гаусса

$$\frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{z^2}{2}}\, dz,$$

с которым мы до сих пор имели дело, есть частный случай этого интеграла $S(a, b;\ \alpha, \beta)$, соответствующий значениям параметров $\alpha = 0$ и $\beta = 1$. Интеграл $S(a, b;\ \alpha, \beta)$ в общем случае тоже называется законом Гаусса с параметрами α и β [1]).

Подстановкой $s = \sqrt{\beta}z + \alpha$ легко получить формулу

$$S(a, b;\ \alpha, \beta) = \Phi\left(\frac{b-\alpha}{\sqrt{\beta}}\right) - \Phi\left(\frac{a-\alpha}{\sqrt{\beta}}\right).$$

[1]) Положительный параметр β часто обозначают через σ^2.

Обратно, подстановка $s = ks - l$ дает нам

$$\Phi(kb+l) - \Phi(ka+l) = S\left(a,\ b;\ -\frac{l}{k},\ \frac{1}{k^2}\right).$$

Поэтому мы можем написать выведенную выше формулу (45) для $\nu = \frac{\mu - l_n}{k_n}$, полагая $k = \frac{k_n}{\sqrt{npq}}$ и $l = \frac{l_n - np}{\sqrt{npq}}$, следующим образом:

$$\boldsymbol{P}(a < \nu < b) \sim S\left(a,\ b;\ \frac{np - l_n}{k_n},\ \frac{npq}{k_n^2}\right).$$

Например, если $\nu = \mu$, так что $k_n = 1$ и $l_n = 0$, то мы имеем

$$\boldsymbol{P}(a < \mu < b) \sim S(a,\ b;\ np,\ npq).$$

Если $\nu = \frac{\mu}{n}$, так что $k_n = n$ и $l_n = 0$, то мы имеем

$$\boldsymbol{P}\left(a < \frac{\mu}{n} < b\right) \sim S\left(a,\ b;\ p,\ \frac{pq}{n}\right),$$

и т. п.

Мы видим, что асимптотическим законом распределения для любой линейной возрастающей функции величины μ является закон Гаусса $S(a,\ b;\ \alpha,\ \beta)$ с одними и теми же пределами интегрирования a и b, но только с разными параметрами α и β.

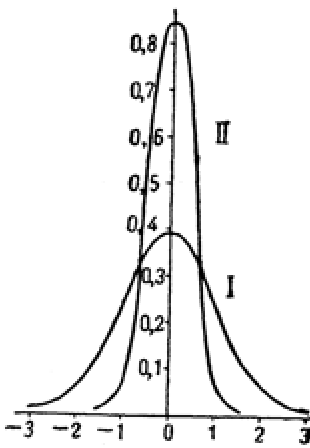
42. Различие между законами $S(a,\ b;\ \alpha,\ \beta)$ с разными параметрами α и β легко интерпретировать геометрически. Рассмотрим семейство кривых

$$w = \frac{1}{\sqrt{2\pi\beta}}\, e^{-\frac{(s-\alpha)^2}{2\beta}},$$

зависящих от параметров α и β. Тогда

$$S(a,\ b;\ \alpha,\ \beta) = \int_a^b w\, ds.$$

Следовательно, $S(a,\ b;\ \alpha,\ \beta)$ равно площади, ограниченной кривой названного семейства и прямыми $w = 0$, $s = a$ и $s = b$.



Черт. 4.

На черт. 3 была изображена кривая названного семейства, соответствующая значениям параметров $\alpha = 0$, $\beta = 1$. На черт. 4 наряду с той же кривой (I) изображена для сравнения другая кривая (II) того же семейства, соответствующая значениям параметров $\alpha = 0$ и $\beta = \frac{1}{4}$. Мы видим, что для меньших значений параметра β кривая поднимается выше при $s = 0$ и падает скорее при $s > 0$ и $s < 0$; при бо́льших значениях параметра β она поднимается и падает менее круто.

При $\alpha \neq 0$ кривые имеют ту же форму, но сдвинуты вправо (при $\alpha > 0$) или влево (при $\alpha < 0$).

§ 12. Теорема Пуассона.

43. Для определенного типа задач асимптотической формулой может служить так называемый *закон Пуассона*:

$$\frac{k^x}{x!}e^{-k},$$

где $k>0$ и $x=0, 1, 2, \ldots$

Представим себе, что мы производим некоторую серию независимых испытаний, состоящую из конечного числа испытаний, затем — новую серию, затем — еще новую и т. д. Мы будем иметь последовательность серий испытаний. Пусть при каждом испытании каждой серии может наступить или не наступить некоторое событие A и пусть вероятность наступления A при отдельном испытании остается постоянной внутри каждой серии, но может меняться от серии к серии. Для этого случая мы имеем следующую теорему.

Теорема Пуассона [1]). — *Пусть дана последовательность серий испытаний, состоящих последовательно из* $1, 2, \ldots, n, \ldots$ *испытаний, пусть* μ — *число наступлений события* A *в* n*-й серии испытаний и пусть вероятность* p *события* A *при каждом отдельном испытании* n*-й серии равна* $\frac{k}{n}$, *где* k — *постоянная, не зависящая от* n. *Тогда вероятность* $P_n(m)$ *того, что величина* μ *будет удовлетворять равенству*

$$\mu=m,$$

при $n\to\infty$ *и постоянном* m *стремится к*

$$\frac{k^m}{m!}e^{-k}.$$

Доказательство. — Как мы знаем,

$$P_n(m)=\frac{n!}{m!\,(n-m)!}p^m(1-p)^{n-m}.$$

В данном случае

$$P_n(m)=\frac{n!}{m!\,(n-m)!}\left(\frac{k}{n}\right)^m\left(1-\frac{k}{n}\right)^{n-m}=$$

$$=\frac{k^m}{m!}\left(1-\frac{k}{n}\right)^n\frac{n(n-1)\ldots(n-m+1)}{n^m\left(1-\frac{k}{n}\right)^m}=$$

$$=\frac{k^m}{m!}\left(1-\frac{k}{n}\right)^n\frac{1\cdot\left(1-\frac{1}{n}\right)\ldots\left(1-\frac{m-1}{n}\right)}{\left(1-\frac{k}{n}\right)^m}.$$

Но, как известно,

$$\lim_{n\to\infty}\left(1-\frac{k}{n}\right)^n=e^{-k}\quad\text{и}\quad\lim_{n\to\infty}\frac{1\cdot\left(1-\frac{1}{n}\right)\ldots\left(1-\frac{m-1}{n}\right)}{\left(1-\frac{k}{n}\right)^m}=1.$$

Отсюда и вытекает требуемое предельное соотношение.

---

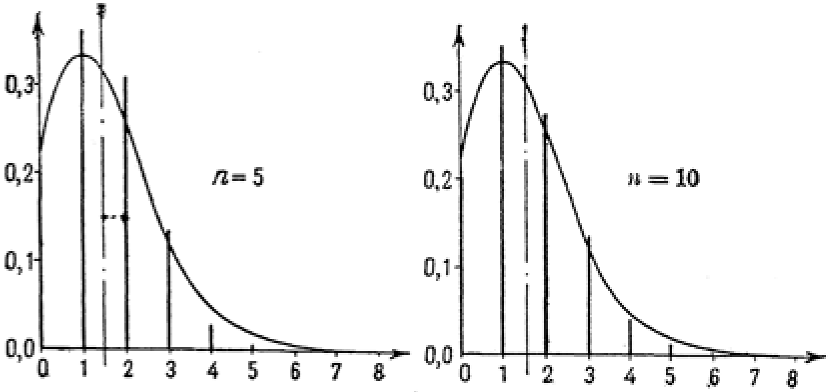
[1]) Во избежание недоразумений отметим, что имя теоремы Пуассона носит также один частный случай приводимой ниже теоремы Маркова.

Асимптотическая формула

$$P_n(m) \sim \frac{k^m}{m!}\, e^{-k},$$

естественно, будет пригодна, как таковая, при больших n, следовательно, при малых вероятностях $\frac{k}{n}$; иначе говоря, она будет пригодна для очень редких событий A.

Черт. 5 показывает, как ординаты $P_n(m)$ стремятся с возрастанием n к своим предельным значениям $\frac{k^m}{m!}\, e^{-k}$ в случае $k = 1{,}5$. При $n = 50$ эти ординаты на чертеже были бы уже неотличимы от своих предельных значений.



Черт. 5.

44. Интересно сопоставить на конкретном примере условия, при которых оказываются применимыми в роли асимптотических формул закон Гаусса и закон Пуассона.

Представим себе сосуд $\mathfrak{A}$, содержащий разреженный газ, и выделим мысленно в этом сосуде какую-нибудь определенную его часть, по терминологии статистической механики — „ячейку", $\mathfrak{B}$. Допустим, что молекулы газа, числом n, перемещаются в сосуде $\mathfrak{A}$ свободно и независимо друг от друга, так что для любой определенной молекулы вероятность найти ее в ячейке $\mathfrak{B}$ равна отношению p объемов $\mathfrak{B}$ и $\mathfrak{A}$, независимо от того, сколько других молекул уже находится в этой ячейке. Поставим прежде всего задачу определить вероятность $P_n(m)$ того, что в ячейке $\mathfrak{B}$ будет ровно m молекул. Эта задача легко приводится к задаче о независимых испытаниях. А именно, представим себе, что мы производим наблюдение за каждой из n молекул с целью установить, находится ли молекула в ячейке $\mathfrak{B}$. Эти наблюдения составят ряд из n испытаний, при каждом из которых может произойти или не произойти событие, состоящее в том, что данная молекула будет найдена в ячейке $\mathfrak{B}$. Вероятность этого события при каждом испытании будет равна p независимо от исхода остальных испытаний. Искомая вероятность $P_n(m)$ будет вероятностью того, что при наших n испытаниях молекулы будут найдены в ячейке $\mathfrak{B}$ ровно m раз. Следовательно:

$$P_n(m) = \frac{n!}{m!\,(n-m)!}\,p^m\,(1-p)^{n-m}.$$

Для того чтобы перейти к асимптотическим формулам, представим себе, что мы увеличиваем число n молекул рассматриваемого нами газа. Но мы будем производить это увеличение массы газа двумя различными способами: I при постоянном объеме и II при постоянном давлении. Это приведет нас к различным асимптотическим формулам. Рассмотрим оба случая отдельно.

I. Пусть масса газа увеличивается при постоянном объеме, т. е. новые молекулы вводятся в наш сосуд $\mathfrak{A}$, остающийся неизменным. В этом случае за исключением увеличения общего числа n молекул в газе вся картина распределения этих молекул остается также неизменной. В частности, для каждой отдельной молекулы сохраняется постоянная вероятность p найти ее в ячейке $\mathfrak{B}$. Отсюда ясно, что в этом случае асимптотической формулой для вероятности $P_n(m)$ будет закон Гаусса.

Примечание. — Реально, конечно, не может быть речи об увеличении общего числа n молекул до бесконечности. При уплотнении газа в сосуде наступит момент, когда уже нельзя будет не считаться с взаимодействием молекул и, следовательно, попадания молекул в данную ячейку уже нельзя будет считать событиями независимыми. При дальнейшем уплотнении мы вообще уже не будем иметь газа. Но до наступления названных степеней уплотнения общее число n молекул будет уже достаточно велико, чтобы асимптотической формулой можно было пользоваться.

II. Пусть теперь масса газа увеличивается при постоянном давлении — путем простого присоединения к данному сосуду $\mathfrak{A}$ новых и новых сосудов такого же объема и содержащих такой же газ. В этом случае будет изменяться не только общее число n молекул в газе, но и вероятность p найти отдельную молекулу в ячейке $\mathfrak{B}$. В самом деле, эта вероятность равна отношению объема $\mathfrak{B}$ к общему объему газа, а этот последний с присоединением новых и новых сосудов будет увеличиваться. Точнее, изменение вероятности p можно охарактеризовать следующим образом. Общий объем газа будет возрастать пропорционально общему числу n его молекул. Следовательно, вероятность p, которая обратно пропорциональна общему объему газа, будет также обратно пропорциональна n: мы будем иметь $p = \frac{k}{n}$, где k — постоянная. Отсюда ясно, что в этом случае, в обширном объеме разреженного газа, асимптотической формулой для вероятности $P_n(m)$ будет закон Пуассона.

Примечание. — Реально и здесь не может быть речи об увеличении общего числа n молекул до бесконечности. При увеличении объема газа наступит момент, когда уже нельзя будет считать попадания всех молекул в данную ячейку равновероятными. Для этого нужно, чтобы были равновероятны как малые, так и сколь угодно большие скорости молекул, чего, очевидно, нет. Но до наступления такой степени увеличения объема газа общее число n молекул и здесь будет, вообще говоря, уже достаточно велико, чтобы можно было пользоваться асимптотической формулой.

ГЛАВА IV.

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ.

§ 13. Расширенная аксиома сложения вероятностей.

45. Все события, вероятности которых мы до сих пор вычисляли, состояли в том, что те или иные величины принимали те или иные значения. Такого рода события являются пока почти единственным объектом, с которым имеет дело теория вероятностей, и поэтому заслуживают более подробного изучения.

Но это изучение наталкивается на целый ряд непреодолимых затруднений, если оставаться при той аксиоматике теории вероятностей, которой мы до сих пор пользовались. Затруднения возникают оттого, что в нашей аксиоматике предусматривалась возможность подразделения событий только на конечное число частных случаев; между тем, в дальнейшем мы будем встречаться с необходимостью подразделения событий на счетное множество частных случаев.

Мы будем говорить, что событие A подразделяется на счетное множество частных случаев $A_1, A_2, \ldots, A_n, \ldots$, если событие A равносильно наступлению хотя бы одного из событий $A_1, A_2, \ldots, A_n, \ldots$ и эти последние друг с другом попарно несовместимы. Для устранения упомянутых выше затруднений мы примем раз навсегда, что аксиома сложения нашей аксиоматики теории вероятностей может быть расширена, а именно — что справедлива следующая аксиома:

Расширенная аксиома сложения вероятностей. — *Если событие A подразделяется на конечное или счетное множество частных случаев $A_1, A_2, \ldots, A_n, \ldots$, то вероятность A складывается из вероятностей $A_1, A_2, \ldots, A_n, \ldots$:*

$$\boldsymbol{P}(A) = \boldsymbol{P}(A_1) + \boldsymbol{P}(A_2) + \ldots + \boldsymbol{P}(A_n) + \cdots$$

46. Поясним эту новую аксиому одним примером.

Пусть ξ — некоторая переменная величина, принимающая те или иные значения в зависимости от случая, и пусть $F(x)$ есть вероятность события, состоящего в том, что значение величины ξ будет меньше x:

$$F(x) = \boldsymbol{P}(\xi < x).$$

Функция $F(x)$, таким образом определенная, называется *функцией распределения* величины ξ.

Приращение функции распределения, $F(x_2) - F(x_1)$, где $x_2 > x_1$, всегда равно вероятности события, состоящего в том, что значение величины ξ будет находиться в интервале $x_1 \leqslant x < x_2$:

$$F(x_2) - F(x_1) = \boldsymbol{P}(x_1 \leqslant \xi < x_2).$$

Это вытекает из теоремы сложения вероятностей. В самом деле, в силу этой последней,

$$\boldsymbol{P}(\xi < x_2) = \boldsymbol{P}(\xi < x_1) + \boldsymbol{P}(x_1 \leqslant \xi < x_2).$$

Поэтому

$$F(x_2) - F(x_1) = \boldsymbol{P}(\xi < x_2) - \boldsymbol{P}(\xi < x_1) = \boldsymbol{P}(x_1 \leqslant \xi < x_2).$$

Отсюда вытекает, что $F(x)$ — *неубывающая функция*, так как ее приращения всегда неотрицательны.

Кроме того, как вероятность, $F(x)$ заключена между 0 и 1.

Без расширенной аксиомы сложения вероятностей больше ничего определенного о произвольной функции распределения сказать нельзя. Применяя же эту аксиому, можно сразу установить еще несколько общих свойств функций распределения.

Во-первых, функция распределения всегда *непрерывна слева*, т. е.

$$F(x - 0) = F(x).$$

В самом деле, возьмем любую возрастающую последовательность точек, сходящуюся к точке x:

$$x_0 < x_1 < \ldots < x_n < \ldots,$$
$$x_n \to x.$$

Тогда, в силу расширенной аксиомы сложения вероятностей,

$$\begin{aligned} F(x-0) &= \lim_{n \to \infty} F(x_n) = \\ &= F(x_0) + [F(x_1) - F(x_0)] + \ldots + [F(x_{n+1}) - F(x_n)] + \ldots = \\ &= \boldsymbol{P}(\xi < x_0) + \boldsymbol{P}(x_0 \leqslant \xi < x_1) + \ldots + \boldsymbol{P}(x_n \leqslant \xi < x_{n+1}) + \ldots = \\ &= \boldsymbol{P}(\xi < x) = F(x). \end{aligned}$$

Во-вторых, правое значение функции распределения равно вероятности события, состоящего в том, что значение величины ξ будет меньше или равно x:

$$F(x + 0) = \boldsymbol{P}(\xi \leqslant x).$$

В самом деле, возьмем любую убывающую последовательность точек, сходящуюся к точке x:

$$x_0 > x_1 > \ldots > x_n > \ldots,$$
$$x_n \to x.$$

Тогда, в силу расширенной аксиомы сложения вероятностей,

$$\begin{aligned} F(x+0) &= \lim_{n \to \infty} F(x_n) = \\ &= F(x_0) - [F(x_0) - F(x_1)] - \ldots - [F(x_n) - F(x_{n+1})] - \ldots = \\ &= P(\xi < x_0) - P(x_1 \leqslant \xi < x_0) - \ldots - P(x_{n+1} \leqslant \xi < x_n) - \ldots = \\ &= P(\xi < x_0) - P(x < \xi < x_0) = P(\xi \leqslant x). \end{aligned}$$

Наконец, функция распределения обладает тем свойством, что

$$F(-\infty)=0 \quad \text{и} \quad F(+\infty)=1.$$

В самом деле, с одной стороны, применяя расширенную аксиому сложения вероятностей и принимая во внимание, что неравенство $\xi<+\infty$ заведомо достоверно, мы имеем

$$\begin{aligned} F(+\infty) &= \lim_{n\to\infty} F(n) = \\ &= F(0)+[F(1)-F(0)]+\ldots+[F(n+1)-F(n)]+\ldots= \\ &= \boldsymbol{P}(\xi<0)+\boldsymbol{P}(0\leqslant\xi<1)+\ldots+\boldsymbol{P}(n\leqslant\xi<n+1)+\ldots= \\ &= \boldsymbol{P}(\xi<+\infty)=1. \end{aligned}$$

С другой стороны, применяя расширенную аксиому сложения вероятностей, мы имеем

$$\begin{aligned} F(-\infty) &= \lim_{n\to\infty} F(-n) = \\ &= F(0)-[F(0)-F(-1)]-\ldots-[F(-n)-F(-n-1)]-\ldots= \\ &= \boldsymbol{P}(\xi<0)-\boldsymbol{P}(-1\leqslant\xi<0)-\ldots-\boldsymbol{P}(-n-1\leqslant\xi<-n)-\ldots= \\ &= \boldsymbol{P}(\xi<0)-\boldsymbol{P}(\xi<0)=0. \end{aligned}$$

Этот пример применения расширенной аксиомы сложения вероятностей в дальнейшем будет неоднократно использован, так как в дальнейшем мы постоянно будем прибегать к понятию функции распределения.

§ 14. Законы распределения.

47. Мы будем рассматривать переменные величины, которые могут принимать те или иные значения в зависимости от случая.

Прежде всего мы подразделим эти переменные величины на два класса с точки зрения наиболее существенной для теории вероятностей. Мы будем различать:

1° Переменные, которые могут принимать конечное или счетное множество значений.

2° Переменные, которые могут принимать несчетное множество значений.

Мы встретили уже несколько примеров переменных, которые могут принимать *конечное* число значений:

1) Число очков, выпадающих на игральной кости. Эта переменная может принимать значения:

$$1,\ 2,\ 3,\ 4,\ 5,\ 6.$$

2) Число μ наступлений события A при n независимых испытаниях (см. п. 23). Эта переменная μ может принимать значения:

$$0,\ 1,\ 2,\ \ldots,\ n.$$

3) Число μ предметов, обладающих признаком X и вошедших в число n предметов, выбранных из некоторой совокупности (см. п. 25). Эта переменная μ может принимать значения:

$$0,\ 1,\ 2,\ \ldots,\ n.$$

Примером переменной, которая может принимать счетное множество значений, может служить любая величина, подчиняющаяся *закону Пуассона*, т. е. могущая принимать любое целое неотрицательное значение m с вероятностью

$$\frac{k^m}{m!} e^{-k}.$$

Примером переменной, которая может принимать несчетное множество значений, может служить любая величина, подчиняющаяся *закону Гаусса*, т. е. могущая принимать значения в любом интервале (a, b) с вероятностью

$$\frac{1}{\sqrt{2\pi\beta}} \int_a^b e^{-\frac{(s-\alpha)^2}{2\beta}} ds.$$

48. Пусть дана переменная ξ, которая может принимать конечное или счетное множество значений x, и определены вероятности $P(x)$ того, что она примет то или иное из возможных значений x. Иначе говоря, пусть для этой переменной можно составить таблицу ее возможных значений и их вероятностей (табл. III).

Таблица III.

Значения x	x_1	x_2	...	x_n	...
Вероятности $P(x)$	$P(x_1)$	$P(x_2)$	...	$P(x_n)$	...

Такую переменную ξ мы назовем *дискретной случайной величиной*. Функция $P(x)$ называется *законом распределения* этой величины.

Так, например, число очков, выпадающих на игральной кости, есть дискретная случайная величина с законом распределения

$$P(x) = \frac{1}{6} \qquad (x = 1,\ 2,\ 3,\ 4,\ 5,\ 6).$$

Число μ наступлений события A при n независимых испытаниях есть дискретная случайная величина со следующим законом распределения:

$$P(x) = C_n^x p^x (1-p)^{n-x} \qquad (x = 0,\ 1,\ 2,\ \ldots,\ n)$$

где p есть вероятность наступления события A при каждом отдельном испытании.

Число μ предметов, обладающих признаком X и вошедших в число n предметов, выбранных из некоторой совокупности, которая содержит N предметов и в том числе M предметов с признаком X, есть дискретная случайная величина с законом распределения

$$P(x) = \frac{C_M^x C_{N-M}^{n-x}}{C_N^n} \qquad (x = 0,\ 1,\ 2,\ \ldots,\ n).$$

Любая величина, подчиняющаяся закону Пуассона, есть дискретная случайная величина; закон Пуассона и есть ее закон распределения:

$$P(x)=\frac{k^x}{x!}e^{-k} \qquad (x=0,\ 1,\ 2,\ \ldots,\ \text{ad inf}).$$

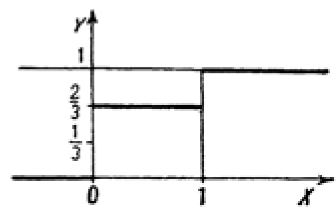
Для всякой дискретной случайной величины ξ имеет место равенство

$$\sum_x P(x)=1,$$

где суммирование распространяется на все возможные значения x величины ξ. Действительно, в силу теоремы III (или расширенной аксиомы сложения вероятностей) левая часть этого равенства представляет собой вероятность того, что величина ξ примет хотя бы одно из ее возможных значений; в силу же теоремы IV эта вероятность равна единице.

Вместо того чтобы задавать для дискретной случайной величины ξ ее закон распределения $P(x)$, мы можем задать ее функцию распределения $F(x)$, определенную в п. 46, т. е. вероятность того, что ξ примет значение, меньшее чем x:

$$F(x)=\boldsymbol{P}(\xi<x).$$



Черт. 6.

Действительно,

$$F(x)=\sum_{s<x}P(s), \tag{49}$$

где суммирование распространяется на все возможные значения s величины ξ, меньшие чем x; это немедленно вытекает из теоремы III (или из расширенной аксиомы сложения вероятностей). Обратно,

$$P(x)=F(x+0)-F(x), \tag{50}$$

так как $\boldsymbol{P}(\xi=x)=\boldsymbol{P}(\xi\leqslant x)-\boldsymbol{P}(\xi<x)$.

Сопоставляя (49) и (50), мы имеем

$$F(x)=\sum_{s<x}[F(s+0)-F(s)].$$

Отсюда видно, что функция распределения дискретной случайной величины разрывна, а именно — что она возрастает скачками при тех значениях x, которые являются возможными значениями данной величины ξ.

Возьмем, например, дискретную случайную величину μ_1 — число наступлений события A при одном испытании, при котором вероятность события A равна p. Величина μ_1 имеет два возможных значения, 0 и 1, причем $P(0)=1-p$ и $P(1)=p$. Поэтому в силу (49)

$$F(x)=\begin{cases} 0 & \text{при } x\leqslant 0,\\ 1-p & \text{при } 0<x\leqslant 1,\\ 1 & \text{при } x>1. \end{cases}$$

График этой функции изображен на черт. 6, где $p=\frac{1}{3}$.

Если два возможных значения величины разделены интервалом, в котором других возможных значений нет, то в этом интервале функция распределения, очевидно, постоянна. Таким образом, когда всех возможных значений, конечное число, например, s, то функция распределения есть ступенчатая функция, постоянная в $s+1$ интервалах. Когда же возможных значений счетное множество, то это последнее может быть и всюду плотным, так что интервалов постоянства у функции распределения может и не быть. Пусть, например, возможными значениями величины ξ являются те и только те числа x, которые имеют вид

$$\frac{m}{2^n},$$

где n — все возможные целые положительные числа, а m — нечетные числа $1, 3, \ldots, 2^n-1$, так что

$$x=\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}, \ldots,$$

и пусть

$$P\left(\frac{m}{2^n}\right)=\frac{1}{3^n},$$

так что только что указанным значениям x соответствуют вероятности

$$P(x)=\frac{1}{3}, \frac{1}{9}, \frac{1}{9}, \frac{1}{27}, \frac{1}{27}, \frac{1}{27}, \frac{1}{27}, \ldots$$

ξ — дискретная случайная величина, и ее функция распределения $F(x)$ имеет множество точек разрыва, всюду плотное на интервале $0<x<1$.

49. Вообще *случайной величиной* называется всякая переменная ξ, которая может принимать те или иные значения в зависимости от случая и для которой определена функция распределения $F(x)$ в смысле п. 46, т. е. вероятность того, что ξ примет значение, меньшее чем x:

$$F(x)=\boldsymbol{P}(\xi<x).$$

Мы знаем, что эта функция — неубывающая и непрерывная слева и что $F(-\infty)=0$ и $F(+\infty)=1$. Обратно, всякая функция $F(x)$ — неубывающая, непрерывная слева и такая, что $F(-\infty)=0$ и $F(+\infty)=1$, может рассматриваться как функция распределения некоторой случайной величины ξ. Действительно, имея такую функцию $F(x)$, достаточно определить вероятности равенством

$$\boldsymbol{P}(\xi<x)=F(x),$$

чтобы удовлетворить всем аксиомам теории вероятностей.

Не следует, однако, думать, что случайные величины можно отождествить с их функциями распределения. Может случиться, что заведомо неравные случайные величины будут иметь одну и ту же функцию распределения. Например, пусть ξ — случайная величина, которая может принимать два значения, 0 или 1, оба — с вероятностью $\frac{1}{2}$. Положим $\eta=1-\xi$. Ясно, что всегда

$$|\xi-\eta|=1.$$

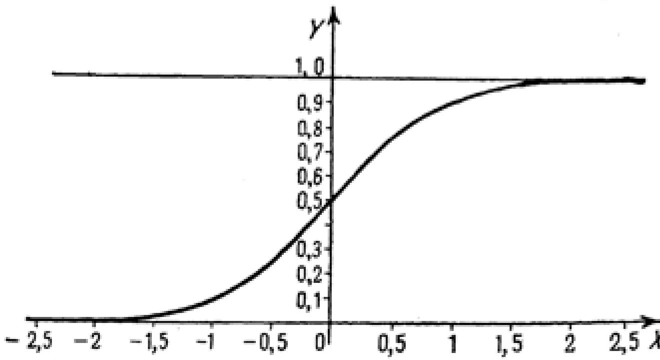
Однако функция распределения у величин ξ и η, очевидно, одна и та же:

$$F(x)=\begin{cases} 0 & \text{при } x \leqslant 0, \\ \frac{1}{2} & \text{при } 0 < x \leqslant 1, \\ 1 & \text{при } x > 1. \end{cases}$$

В качестве примера функции распределения недискретной случайной величины приведем функцию распределения величины, подчиняющейся закону Гаусса:

$$F(x)=\frac{1}{\sqrt{2\pi\beta}}\int_{-\infty}^{x} e^{-\frac{(s-\alpha)^2}{2\beta}}\,ds.$$

График функции распределения величины, подчиняющийся закону Гаусса с параметрами $\alpha=0$ и $\beta=1$, изображен на черт. 7.



Черт. 7.

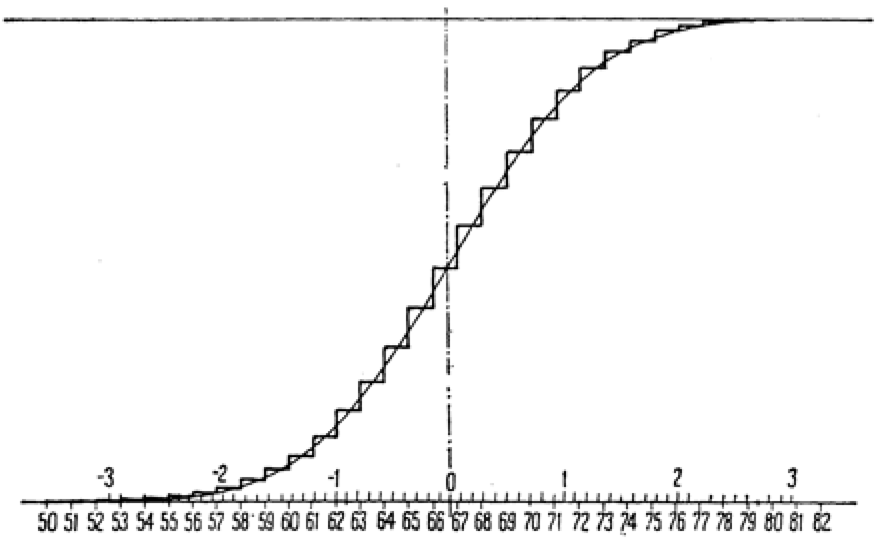
Иногда характеризуют поведение случайной величины не функцией распределений, а каким-нибудь другим выражением, лишь бы функция распределения могла быть получена из него по определенным правилам. Всякое такое выражение называется *законом распределения* случайной величины. Например, законом распределения случайной величины ξ всегда можно назвать функцию интервала $P(x_1, x_2)$ — вероятность того, что значение ξ будет находиться в интервале $x_1 \leqslant x < x_2$. Действительно, из $P(x_1, x_2)$ можно получить функцию распределения $F(x)$ и обратно по формулам

$$F(x)=P(-\infty, x),$$
$$P(x_1, x_2)=F(x_2)-F(x_1).$$

Мы фактически уже применяли название закона распределения в этом смысле. А именно, мы назвали законом распределения Гаусса интеграл

$$P(x_1, x_2)=\frac{1}{\sqrt{2\pi\beta}}\int_{x_1}^{x_2} e^{-\frac{(s-\alpha)^2}{2\beta}}\,ds.$$

Но характеристика поведения случайной величины при помощи функции распределения имеет общепризнанное преимущество одновременной общности и наглядности. Чтобы это стало ясно, достаточно взглянуть на черт. 8, где изображены два графика, иллюстрирующие интегральную теорему Лапласа: ступенчатый график функции распределения нормированного отклонения величины μ при $p=\frac{2}{3}$ и $n=100$ и непрерывный график функции распределения величины ξ, подчиняющейся закону Гаусса с параметрами $\alpha=0$ и $\beta=1$ [1]).



Черт. 8.

Примечание. — Мы условились называть и будем называть всюду в дальнейшем, за исключением гл. VII и VIII, **функцией распределения** вероятность того, что ξ примет значение, меньшее чем x. Но иногда называют функцией распределения вероятность того, что ξ примет значение, меньшее или равное x:

$$F(x)=\boldsymbol{P}(\xi\leqslant x).$$

Иногда называют функцией распределения также среднее арифметическое из обеих названных вероятностей:

$$F(x)=\frac{\boldsymbol{P}(\xi<x)+\boldsymbol{P}(\xi\leqslant x)}{2}.$$

То или иное из трех перечисленных определений функции распределения выбирается в зависимости от удобства, в связи с аналитическим аппаратом, которым пользуются в данном исследовании.

[1]) Чертеж 8 заимствован из кабинета статистики НИИМ при МГУ.

50. Пусть функция распределения $F(x)$ случайной величины ξ имеет непрерывную производную $p(x)$, так что

$$F(x)=\int_{-\infty}^{x} p(s)\,ds. \tag{51}$$

В таком случае ξ будет называться *непрерывной случайной величиной.*

Функция $p(x)$ называется *плотностью вероятности* этой величины в точке x. Нередко, впрочем, вместо того, чтобы прибегать к этому термину, пишут $p(x)\,dx$ и называют это последнее выражение „вероятностью того, что ξ примет значение в бесконечно малом промежутке между x и $x+dx$".

Величина, подчиняющаяся закону Гаусса, есть как раз непрерывная случайная величина. Ее плотность вероятности есть

$$p(x)=\frac{1}{\sqrt{2\pi\beta}}e^{-\frac{(x-\alpha)^2}{2\beta}}.$$

Что касается связи между законом распределения $P(x_1, x_2)$, определенным в п. 49, и плотностью вероятности $p(x)$, то в силу соотношения $P(x_1, x_2)=F(x_2)-F(x_1)$ эта связь описывается формулами

$$p(x)=\lim_{\Delta x\to 0}\frac{P(x, x+\Delta x)}{\Delta x},$$

$$P(x_1, x_2)=\int_{x_1}^{x_2} p(s)\,ds.$$

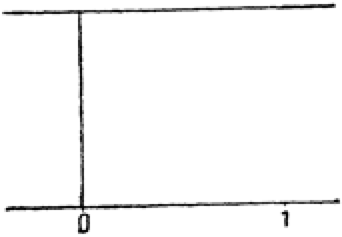
В частности, отметим, что всегда

$$\int_{-\infty}^{+\infty} p(s)\,ds=1.$$

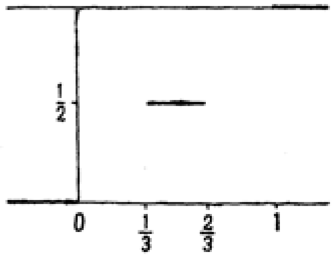
Существуют, разумеется, случайные величины, которые нельзя отнести ни к дискретным, ни к непрерывным. Не говоря уже о возможности случайных величин, которые в одних интервалах ведут себя как дискретные, а в других — как непрерывные, не будучи, следовательно, в целом ни теми ни другими, — существуют и такие случайные величины, которые не являются дискретными ни в каком интервале, но не являются и непрерывными. Представление о таких случайных величинах можно составить себе, рассмотрев хотя бы случайную величину ξ, которая ведет себя следующим образом. ξ может принимать только значения между нулем и единицей. Таким образом $P(0, 1)=1$. Следовательно, $F(x)=0$ при $x<0$ и $F(x)=1$ при $x>1$ (см. черт. 9а). Что касается интервала $(0, 1)$, то ξ может принимать значения только в первой его трети и в последней, то и другое — с одинаковой вероятностью. Таким образом $P\left(0, \frac{1}{3}\right)=\frac{1}{2}$ и $P\left(\frac{1}{3}, 1\right)=\frac{1}{2}$. Следовательно (см. черт. 9б).

$$F(x)=\frac{1}{2} \text{ при } \frac{1}{3}<x<\frac{2}{3}.$$

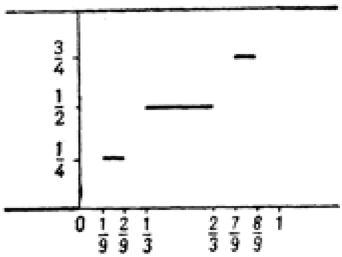
Далее, что касается интервалов $\left(0, \frac{1}{3}\right)$ и $\left(\frac{2}{3}, 1\right)$, то ξ опять-таки может принимать только значения в первой и в последней трети каждого из них, то и другое — с одинаковой вероятностью. Таким образом $P\left(0, \frac{1}{9}\right)=\frac{1}{4}$, $P\left(\frac{2}{9}, \frac{1}{3}\right)=\frac{1}{4}$, $P\left(\frac{2}{3}, \frac{7}{9}\right)=\frac{1}{4}$ и $P\left(\frac{8}{9}, 1\right)=\frac{1}{4}$.



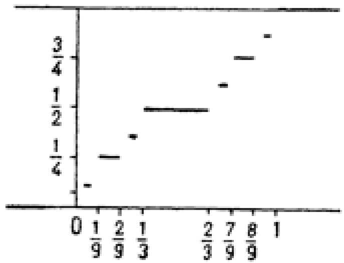
Черт. 9а.



Черт. 9б.



Черт. 9в.



Черт. 9г.

Следовательно (см. черт. 9в).

$$F(x)=\frac{1}{4} \text{ при } \frac{1}{9}<x<\frac{2}{9},$$

$$F(x)=\frac{3}{4} \text{ при } \frac{7}{9}<x<\frac{8}{9}.$$

Далее, в интервалах $\left(0, \frac{1}{9}\right)$, $\left(\frac{2}{9}, \frac{1}{3}\right)$, $\left(\frac{2}{3}, \frac{7}{9}\right)$ и $\left(\frac{8}{9}, 1\right)$ повторяется такая же картина распределения значений ξ (см. черт. 9г). И так далее. Представим себе, что мы повторили наше рассуждение бесконечное число раз. Тогда мы будем иметь счетное множество интервалов — назовем их интервалами γ, — на которых функция распределения $F(x)$ у нас будет определена. Нетрудно видеть, что интервалы γ являются смежными интервалами к некоторому совершенному нигде неплотному множеству $\mathfrak{G}$. Нам остается, следовательно, определить функцию распределения $F(x)$ на этом множестве $\mathfrak{G}$. Но на нем она должна получить вполне определенные значения в силу того, что в целом она должна быть всюду неубывающей. Величина ξ не является ни дискретной, ни непрерывной случайной величиной, ибо ее функция распределения $F(x)$ непрерывна, но, вообще говоря, не диференцируема.

51. Пусть теперь дана пара переменных величин ξ и η, каждая из которых может принимать конечное или счетное множество значений, и пусть определены вероятности $P(x; y)$ того, что ξ примет то или иное из возможных значений x и одновременно η примет то или иное из возможных значений y. Функцию $P(x; y)$ мы будем называть *двумерным законом распределения* пары величин ξ и η.

Каждая из величин ξ и η является в этом случае дискретной случайной величиной; их законы распределения, $P_\xi(x)$ и $P_\eta(y)$, определяются из двумерного закона распределения $P(x; y)$ по формулам

$$P_\xi(x)=\sum_y P(x; y), \quad P_\eta(y)=\sum_x P(x; y),$$

где суммирование распространяется в каждом случае на все возможные значения соответствующей величины. Формулы эти вытекают непосредственно из теоремы сложения вероятностей.

Напротив, законы распределения двух дискретных случайных величин не определяют двумерного закона распределения. Например, пусть ξ — число очков, выпадающее на одной, и η — число очков, выпадающее на другой, ничем не связанной с первой, игральной кости. Ясно, что $P_\xi(x)=\frac{1}{6}$ при любом x, что $P_\eta(y)=\frac{1}{6}$ при любом y и что $P(x; y)=\frac{1}{36}$ при любой паре x и y $(x=1, \ldots, 6;\ y=1, \ldots, 6)$. Напротив, пусть ξ — число очков, выпадающее на некоторой игральной кости, и η — число очков на грани той же кости, противоположной выпавшей грани. Ясно, что теперь $P_\xi(x)$ и $P_\eta(y)$ — те же, что и в предыдущем примере, но $P(x; y)=\frac{1}{6}$ для тех пар x и y, для которых $x+y=7$, и $P(x; y)=0$ для всех прочих пар x и y. Это — совсем не та функция $P(x; y)$, что в предыдущем примере.

Исключение составляет лишь случай независимых величин. Дискретные случайные величины ξ и η называются *независимыми*, если, каково бы ни было возможное значение x величины ξ и каково бы ни было возможное значение y величины η, событие A_x, состоящее в том, что ξ принимает значение x, и событие B_y, состоящее в том, что η принимает значение y, являются независимыми событиями. Из теоремы умножения вероятностей для независимых событий немедленно вытекает, что в этом случае

$$P(x; y)=P_\xi(x)\,P_\eta(y),$$

причем выполнение этого равенства для всех возможных пар значений x и y является не только необходимым, но и достаточным условием независимости величин ξ и η.

Пусть, далее, дана пара, вообще говоря, каких угодно переменных величин ξ и η, и пусть определены вероятности $F(x; y)$ того, что ξ примет значение, меньшее чем x, и одновременно η примет значение, меньшее чем y. Функция $F(x; y)$ называется *двумерной функцией распределения* пары величин ξ и η.

Каждая из величин ξ и η является здесь случайной величиной; их функции распределения, $F_\xi(x)$ и $F_\eta(y)$, определяются из двумерной функции распределения $F(x;\ y)$ по формулам

$$F_\xi(x) = \int_{-\infty}^{+\infty} d_y F(x;\ y) = F(x;\ +\infty),$$
$$F_\eta(y) = \int_{-\infty}^{+\infty} d_x F(x;\ y) = F(+\infty;\ y).$$

Напротив, функции распределения двух случайных величин не определяют двумерной функции распределения.

Исключение, как и в случае дискретных величин, составляют независимые величины. В общем случае две случайные величины ξ и η называются *независимыми*, если, какова бы ни была пара чисел x и y, — событие A^x, состоящее в том, что ξ примет значение, меньшее чем x, и событие B^y, состоящее в том, что η примет значение, меньшее чем y, являются независимыми событиями. Из теоремы умножения вероятностей для независимых событий немедленно вытекает, что в этом случае

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

Как и в одномерном случае, поведение пары случайных величин можно охарактеризовать и не двумерной функцией распределения, а каким-нибудь другим *двумерным законом распределения*, с нею связанным по определенным правилам.

Так, естественным аналогом закона распределения

$$P(x_1,\ x_2),$$

определенного нами в одномерном случае в п. 49, для пары случайных величин ξ и η является двумерный закон распределения $P(x_1,\ x_2;\ y_1,\ y_2)$ — вероятность того, что значение ξ будет находиться в интервале $x_1 \leqslant x < x_2$ и одновременно значение η — в интервале $y_1 \leqslant y < y_2$. Связь между $P(x_1,\ x_2;\ y_1,\ y_2)$ и двумерной функцией распределения $F(x;\ y)$ выражается формулами

$$F(x;\ y) = P(-\infty,\ x;\ -\infty,\ y),$$
$$P(x_1,\ x_2;\ y_1,\ y_2) = F(x_2;\ y_2) - F(x_2;\ y_1) - F(x_1;\ y_2) + F(x_1;\ y_1).$$

Пусть, наконец, существует непрерывная функция $p(x;\ y)$, такая, что двумерная функция распределения $F(x;\ y)$ пары величин ξ и η является ее интегралом:

$$F(x;\ y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p(s;\ t)\,dt\,ds. \tag{52}$$

Функция $p(x;\ y)$ называется *плотностью вероятности* в точке $(x;\ y)$. Или, иначе, пишут $p(x;\ y)\,dx\,dy$ и называют это последнее выражение „вероятностью того, что пара значений данных случайных величин окажется в бесконечно малом прямоугольнике $(x, x+dx; y, y+dy)$".

В этом случае величины ξ и η являются непрерывными случайными величинами, и их плотности вероятности $p_\xi(x)$ и $p_\eta(y)$ в точках x и y выражаются через плотность вероятности $p(x; y)$ в точке $(x; y)$ следующим образом:

$$p_\xi(x) = \int_{-\infty}^{+\infty} p(x; t)\,dt, \quad p_\eta(y) = \int_{-\infty}^{+\infty} p(s; y)\,ds.$$

Это легко выводится из приведенного выше выражения функций распределения $F_\xi(x)$ и $F_\eta(y)$ через двумерную функцию распределения $F(x; y)$ и из выражения этой последней интегралом (52).

Для независимых величин ξ и η в этом случае имеет место равенство

$$p(x; y) = p_\xi(x)\,p_\eta(y).$$

Оно без труда выводится из такого же равенства для функций распределения и из равенств (51) и (52).

Совершенно так же, как мы ввели двумерные законы распределения, могут быть введены, разумеется, и многомерные законы распределения любого числа измерений.

52. Понятие двумерного закона распределения находит естественную интерпретацию и естественное применение в задачах на так называемые геометрические вероятности — в задачах, в которых требуется определить вероятность, при тех или иных условиях, попадания „брошенной" наугад точки на ту или иную область плоскости. Вероятность попадания точки на прямоугольник $(x_1, x_2; y_1, y_2)$ выражается двумерным законом распределения $P(x_1, x_2; y_1, y_2)$, конкретная форма которого зависит от заданных условий „бросания" точки. Если таким образом задается не закон распределения, а плотность вероятности $p(x; y)$, то вероятность попадания на названный прямоугольник выражается интегралом

$$\int_{y_1}^{y_2}\int_{x_1}^{x_2} p(x; y)\,dx\,dy.$$

Подобным же образом, при заданной плотности вероятности $p(x; y)$ вероятность попадания точки на какую-нибудь область D плоскости должна быть равна интегралу

$$\iint_D p(x; y)\,dx\,dy.$$

В этом нетрудно убедиться, исходя из вышеприведенного интегрального выражения для вероятности попадания на прямоугольники и опираясь на аксиоматику теории вероятностей.

Аналогичную интерпретацию допускает, разумеется, и понятие трехмерного закона распределения.

§ 15. Интеграл Стильтьеса.

53. В теории случайных величин большое значение имеют так называемые числовые характеристики этих величин, т. е. постоянные числа, связанные со случайными величинами по тем или иным правилам. В определении же важнейших числовых характеристик основную роль играют так называемые интегралы Стильтьеса, с которыми мы поэтому и должны ознакомиться.

Пусть дан интервал (a, b) (в частности может быть $a = -\infty$ и $b = +\infty$) и пусть на нем определены: непрерывная функция $f(x)$ и какая-нибудь функция распределения случайной величины, $F(x)$. Подразделим интервал (a, b) на частичные интервалы вида

$$(x_i, \ x_i + \Delta x_i)$$

и составим сумму

$$\sum_i f(x_i)\,[F(x_i + \Delta x_i) - F(x_i)],$$

распространенную на все подразделение (в случаях $a = -\infty$ и $b = +\infty$ это будет бесконечный ряд). Если эта сумма имеет определенный предел при $\omega \to 0$, где ω есть верхняя грань колебаний функции $f(x)$ на частичных интервалах данного подразделения, — то этот предел называется *интегралом Стильтьеса функции* $f(x)$ с интегрирующей функцией $F(x)$ и обозначается через

$$\int_a^b f(x)\,dF(x).$$

Он наверное существует, когда функция $f(x)$ непрерывна и ограничена (даже в случаях $a = -\infty$ или $b = +\infty$); это доказывается так же, как и для обыкновенных интегралов.

Когда функция $f(x)$, будучи непрерывной, тем не менее не ограничена (случай, который может представиться при $a = -\infty$ или $b = +\infty$), этот интеграл может не существовать. При этом — и это важно отметить — интеграл функции $f(x)$ в том его определении, которое только что дано, существует тогда и только тогда, когда существует и интеграл ее абсолютной величины $|f(x)|$, с той же интегрирующей функцией:

$$\int_a^b |f(x)|\,dF(x);$$

это опять-таки доказывается так же, как для обыкновенных интегралов.

Если $F(x)$ есть функция распределения **дискретной** случайной величины, так что

$$F(x) = \sum_{s<x} F(s+0) - F(s),$$

то

$$\int_a^b f(x)\,dF(x) = \sum_{a<x<b} f(x)\,[F(x+0) - F(x)],$$

где x — возможные значения данной величины, причем интеграл в левой части равенства существует тогда и только тогда, когда ряд в правой части — абсолютно сходящийся.*

Если $F(x)$ есть функция распределения непрерывной случайной величины, так что она имеет непрерывную производную $p(x)$, плотность вероятности данной величины, то

$$\int_a^b f(x)\,dF(x) = \int_a^b f(x)\,p(x)\,dx,$$

причем интеграл в левой части равенства существует тогда и только тогда, когда интеграл в правой части — абсолютно сходящийся.

Нам придется также иметь дело с кратными интегралами Стильтьеса. Покажем, как они определяются, на примере двойных интегралов.

Пусть дан прямоугольник $(a_1,\ b_1;\ a_2,\ b_2)$ и пусть на нем определены: непрерывная функция $f(x)$ и какая-нибудь двумерная функция распределения пары случайных величин, $F(x;\ y)$. Подразделим прямоугольник $(a_1,\ b_1;\ a_2,\ b_2)$ на частичные прямоугольники вида

$$(x_i,\ x_i + \Delta x_i;\quad y_i,\ y_i + \Delta y_i)$$

и составим сумму

$$\sum_i f(x_i,\ y_i)\,[F(x_i + \Delta x_i;\ y_i + \Delta y_i) - F(x_i + \Delta x_i;\ y_i) - \\ - F(x_i;\ y_i + \Delta y_i) + F(x_i;\ y_i)],$$

распространенную на все подразделение. Если эта сумма имеет определенный предел при $\omega \to 0$, где ω есть верхняя грань колебаний функции $f(x,\ y)$ на частичных прямоугольниках данного подразделения, то этот предел называется двойным интегралом Стильтьеса и обозначается через

$$\int_{a_1}^{b_1}\int_{a_2}^{b_2} f(x,\ y)\,ddF(x;\ y).$$

Вопросы существования решаются для него так же, как для простого интеграла.

Если $f(x,\ y) = f_1(x)\,f_2(y)$, то

$$\int_{a_1}^{b_1}\int_{a_2}^{b_2} f(x,\ y)\,ddF(x;\ y) = \int_{a_1}^{b_1} f_1(x)\,d\int_{a_2}^{b_2} f_2(y)\,d_y F(x;\ y). \tag{53}$$

Если $F(x;\ y)=F_1(x)F_2(y)$, то

$$(54)\qquad \int_{a_1}^{b_1}\int_{a_2}^{b_2} f(x,\ y)\,ddF(x;\ y)=\int_{a_2}^{b_2}\left[\int_{a_1}^{b_1} f(x,\ y)\,dF_1(x)\right]dF_2(y).$$

54. Перечислим основные свойства интеграла Стильтьеса, которые понадобятся нам в дальнейшем. Доказательств мы, как и выше, не приводим, так как при желании их нетрудно восстановить, исходя из самого определения интеграла Стильтьеса.

1) Мы полагаем по определению

$$\int_b^a f(x)\,dF(x)=-\int_a^b f(x)\,dF(x).$$

2) Если $a<c_1<\ldots<c_n<b$, то

$$\int_a^b f(x)\,dF(x)=\int_a^{c_1} f(x)\,dF(x)+\ldots+\int_{c_n}^b f(x)\,dF(x).$$

3) При $f(x)=1$ мы имеем

$$\int_a^b dF(x)=F(b)-F(a).$$

4) При $f(x)=k_1f_1(x)+\ldots+k_nf_n(x)$, где $k_1,\ldots,\ k_n$ — произвольные постоянные, мы имеем

$$\int_a^b f(x)\,dF(x)=k_1\int_a^b f_1(x)\,dF(x)+\ldots+k_n\int_a^b f_n(x)\,dF(x).$$

5) Если $a<b$, то при $f(x)\geqslant 0$ мы имеем

$$\int_a^b f(x)\,dF(x)\geqslant 0.$$

6) При

$$F(x)=\int_a^x g(s)\,dG(s)$$

мы имеем

$$\int_a^b f(x)\,dF(x)=\int_a^b f(x)\,g(x)\,dG(x);$$

в частности, при

$$F(x)=\int_a^x g(s)\,ds$$

мы имеем

$$\int_a^b f(x)\,dF(x) = \int_a^b f(x)\,g(x)\,dx.$$

7) При $F(x) = K_1F_1(x) + \ldots + K_nF_n(x)$, где $K_1, \ldots, K_n$ — положительные постоянные, мы имеем

$$\int_a^b f(x)\,dF(x) = K_1\int_a^b f(x)\,dF_1(x) + \ldots + K_n\int_a^b f(x)\,dF_n(x).$$

8) Если $F(x)$ зависит от параметра y, так что

$$F(x) = F_y(x),$$

то

$$\int_A^B\left[\int_a^b f(x)\,dF_y(x)\right]dy = \int_a^b f(x)\,d\int_A^B F_y(x)\,dy.$$

Аналогичными свойствами обладают и кратные интегралы Стильтьеса.

§ 16. Среднее значение и дисперсия.

55. Главное место среди числовых характеристик случайной величины занимает *среднее значение*, или *математическое ожидание*, определение которого мы сейчас дадим. При этом мы остановимся на первом термине „среднее значение", хотя второй термин „математическое ожидание" больше распространен в теоретических исследованиях и в учебниках. Наш выбор объясняется тем, что первый термин все же вытесняет второй и становится общераспространенным в приложениях теории вероятностей к физике, к технике, к биологии и т. д. Обозначать мы будем среднее значение случайной величины ξ символом $\boldsymbol{E}\xi$. Здесь буква $\boldsymbol{E}$ — отнюдь не множитель, а лишь условный знак; в целом выражение $\boldsymbol{E}\xi$ обозначает некоторое постоянное число, определенным образом связанное с данной случайной величиной.

Если ξ — дискретная случайная величина с законом распределения $P(x)$, то это число $\boldsymbol{E}\xi$ определяется формулой

$$\boldsymbol{E}\xi = \sum_x x P(x),$$

где суммирование распространяется на все возможные значения величины ξ. При этом в случае, когда величина ξ может принимать счетное множество значений, среднее значение $\boldsymbol{E}\xi$ считается существующим лишь при том условии, что определяющая его сумма представляет собой не только сходящийся, но и абсолютно сходящийся ряд; иначе эта сумма зависела бы от порядка, в котором расположены члены ряда, и, следовательно, характеризовала бы не столько самую случайную величину, сколько пожелания вычислителя.

Если ξ — произвольная случайная величина с функцией распределения $F(x)$, то среднее значение $E\xi$ определяется формулой

$$E\xi = \int_{-\infty}^{+\infty} x\, dF(x).$$

В частности, если ξ — дискретная случайная величина, то эта общая формула дает выражение

$$E\xi = \sum_x x\,[F(x+0) - F(x)],$$

где суммирование распространяется на все возможные значения величины ξ; в силу равенства (50) п. 48 приведенная выше формула для среднего значения дискретной случайной величины совпадает с этим выражением.

Если ξ — непрерывная случайная величина, с плотностью вероятности $p(x)$, то

$$E\xi = \int_{-\infty}^{+\infty} xp(x)\, dx.$$

Поясним смысл среднего значения следующим примером. Пусть разыгрывается беспроигрышная лотерея. Допустим, что продается n билетов, а выигрыши распределяются следующим образом:

есть m_1 выигрышей по x_1 рублей каждый,
есть m_2 выигрышей по x_2 рублей каждый,
. . . .
есть m_s выигрышей по x_s рублей каждый,

причем

$$m_1 + m_2 + \ldots + m_s = n.$$

Поставим вопрос: какова справедливая цена одного билета? Это — некоторая средняя стоимость выигрыша. Точно она определяется так: справедливая цена a одного билета должна быть такова, чтобы сумма денег, вырученных за билеты, была равна сумме денег, затраченных на выигрыши. Таким образом должно быть

$$na = m_1x_1 + m_2x_2 + \ldots + m_sx_s,$$

откуда

$$a = x_1\frac{m_1}{n} + x_2\frac{m_2}{n} + \ldots + x_s\frac{m_s}{n}.$$

Теперь, выигрыш лица, купившего один билет, есть случайная величина ξ, которая может принять одно из значений $x_1, x_2, \ldots, x_s$ с некоторыми вероятностями $p_1, p_2, \ldots, p_s$. Среднее значение этой случайной величины ξ есть

$$E\xi = x_1p_1 + x_2p_2 + \ldots + x_sp_s.$$

При этом вероятности различных значений выигрыша, в силу самого определения вероятности, равны

$$p_1 = \frac{m_1}{n}, \quad p_2 = \frac{m_2}{n}, \quad \ldots, \quad p_s = \frac{m_s}{n}.$$

Таким образом

$$\boldsymbol{E}\xi = a.$$

Мы видим, что справедливая цена билета есть среднее значение выигрыша.

Понятие среднего значения возникло из задач о случайных величинах такого типа и затем было распространено и на всевозможные случайные величины.

Все случайные величины, с которыми мы до сих пор встречались, имеют среднее значение. Ниже мы вычислим наиболее важные из них.

Примером случайной величины, не имеющей определенного среднего значения, может служить величина, подчиняющаяся так называемому *закону Коши*:

$$F(x) = \frac{1}{2} + \frac{1}{\pi}\operatorname{arctg} x.$$

В самом деле, здесь

$$\int_{-\infty}^{+\infty} |x|\, dF(x) = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{|x|}{1+x^2}\, dx = \frac{2}{\pi}\int_{0}^{+\infty} \frac{x}{1+x^2}\, dx = +\infty.$$

В связи с определением среднего значения необходимо отметить одно обстоятельство, весьма важное для последующего. Если ξ есть какая-нибудь случайная величина, с функцией распределения

$$F(x) = \boldsymbol{P}(\xi < x),$$

и если на множестве всех возможных значений x этой величины задана непрерывная функция $f(x)$, то величина $\eta = f(\xi)$ есть также случайная величина, с функцией распределения

$$G(y) = \boldsymbol{P}[f(\xi) < y].$$

Среднее значение $\boldsymbol{E}\eta$, согласно общему определению среднего значения, равно

$$I = \int_{-\infty}^{+\infty} y\, dG(y).$$

На деле же обычно определяют среднее значение $\boldsymbol{E}\eta = \boldsymbol{E}f(\xi)$ как выражение

$$J = \int_{-\infty}^{+\infty} f(x)\, dF(x).$$

Для того чтобы это обычное определение не противоречило общему определению среднего значения, необходимо, чтобы всегда имело место равенство $I=J$, и прежде всего — чтобы интегралы I и J существовали или не существовали одновременно. Так оно и есть на самом деле, но доказательство этого требует применения довольно сложного аналитического аппарата, и было впервые дано только недавно (автором этой книги); мы не можем воспроизвести его здесь в общем виде.

Аналогично, если ξ и η есть какая-нибудь пара случайных величин, с двумерной функцией распределения $F(x; y)$, и если на множестве всех возможных пар значений $(x; y)$ этих величин задана непрерывная функция $f(x, y)$, то величина $\zeta=f(\xi, \eta)$ есть также случайная величина, со своей функцией распределения $G(z)$, и среднее значение $\boldsymbol{E}\zeta$ равно

$$\int_{-\infty}^{+\infty} z\,dG(z).$$

На деле же обычно определяют среднее значение $\boldsymbol{E}\zeta=\boldsymbol{E}f(\xi, \eta)$ как выражение

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x, y)\,dd F(x; y).$$

В этом двумерном случае дело обстоит точно так же, как и в одномерном. Точно так же обстоит дело и вообще в любом многомерном случае.

Приведем доказательство равенства интегралов I и J для того единственного случая, когда оно не представляет затруднений: для случая, когда $f(x)$ есть возрастающая функция. Здесь однозначно определена обратная ей функция $f^{-1}(y)$, также возрастающая. Будем выбирать подразделения прямой $-\infty<y<+\infty$ на частичные интервалы $(y_i, y_i+\Delta y_i)$ таким образом, чтобы верхняя грань ω длин Δy_i при переходе от подразделения к подразделению стремилась к нулю. При этом условии мы имеем

$$I=\int_{-\infty}^{+\infty} y\,dG(y)=\lim_{\omega\to 0}\sum_i y_i\,[G(y_i+\Delta y_i)-G(y_i)].$$

Положим $x_i=f^{-1}(y_i)$. Мы получим

$$G(y_i+\Delta y_i)-G(y_i)=\boldsymbol{P}[f(x_i)\leqslant\eta<f(x_i+\Delta x_i)]=$$
$$=\boldsymbol{P}(x_i\leqslant\xi<x_i+\Delta x_i)=F(x_i+\Delta x_i)-F(x_i).$$

Следовательно:

$$I=\lim_{\omega\to 0}\sum_i f(x_i)\,[F(x_i+\Delta x_i)-F(x_i)]=\int_{-\infty}^{+\infty} f(x)\,dF(x),$$

так что, действительно,

$$I=J.$$

Примечание. — Часто определяют среднее значение как выражение

$$L=\lim_{\substack{a\to-\infty\\ b\to+\infty}}\int_a^b x\,dF(x),$$

где предел берется при условии, что a стремится к $-\infty$ и b стремится к $+\infty$ независимо одно от другого. Это вполне равносильно вышеприведенному определению среднего значения, опирающемуся на наше прямое определение интеграла Стильтьеса в пределах от $-\infty$ до $+\infty$. Однако, когда речь идет о среднем значении $\boldsymbol{E}f(\xi)$, то выражение

$$M = \lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_a^b f(x)\, dF(x)$$

уже не равносильно, вообще говоря, интегралу J: оно может иметь смысл, в то время как названный интеграл его не имеет и, следовательно, среднее значение в нашем определении не существует. Тогда и только тогда, когда имеет смысл не одно лишь выражение M, но и выражение, получающееся из него заменой функции $f(x)$ ее абсолютной величиной $|f(x)|$, — имеет смысл и интеграл J; при этом $M = J$.

Например, пусть случайная величина ξ может принимать только неотрицательные значения, для которых ее функция распределения есть

$$F(x) = 1 - e^{-x},$$

и пусть для тех же значений определена непрерывная функция

$$f(x) = \varphi(1 - e^{-x}),$$

где $\varphi(s)$ есть функция, определенная на интервале $0 \leqslant s < 1$ следующим образом: когда

$$\frac{n}{n+1} \leqslant s < \frac{n+1}{n+2},$$

где $n = 0, 1, 2, \ldots$, то

$$\varphi(s) = (n+2)\pi \sin[(n+2)(n+1)\pi s - (n+1)n\pi].$$

Здесь

$$M = \lim_{b \to +\infty} \int_0^b f(x)\, dF(x) = \int_0^1 \varphi(s)\, ds = 2 \log 2,$$

между тем как

$$\lim_{b \to +\infty} \int_0^b |f(x)|\, dF(x) = +\infty,$$

и интеграл J здесь не существует.

56. Докажем несколько теорем о средних значениях, которые могут оказать помощь при вычислении средних значений и особенно существенны для дальнейшего развития теории случайных величин.

Теорема IX. — *Среднее значение постоянной есть сама эта постоянная:*

$$\boldsymbol{E}k = k.$$

Доказательство. — Постоянную k можно рассматривать как дискретную случайную величину, которая может принимать только одно значение k с вероятностью единица. Поэтому

$$\boldsymbol{E}k = k \cdot 1 = k.$$

Теорема X. — *Постоянный множитель можно выносить за знак среднего значения:*

$$\boldsymbol{E}(k\xi) = k\boldsymbol{E}\xi.$$

Доказательство для дискретных случайных величин:

$$\boldsymbol{E}(k\xi) = \sum_x kxP(x) = k\sum_x xP(x) = k\boldsymbol{E}\xi.$$

Доказательство для произвольных случайных величин:

$$\boldsymbol{E}(k\xi) = \int\limits_{-\infty}^{+\infty} kx\,dF(x) = k\int\limits_{-\infty}^{+\infty} x\,dF(x) = k\boldsymbol{E}\xi.$$

Теорема XI. — *Среднее значение суммы двух случайных величин складывается из средних значений этих величин:*

$$\boldsymbol{E}(\xi+\eta) = \boldsymbol{E}\xi + \boldsymbol{E}\eta.$$

Доказательство для дискретных случайных величин:

$$\begin{aligned}\boldsymbol{E}(\xi+\eta) &= \sum_x\sum_y (x+y)\,P(x;\,y) = \\ &= \sum_x x\Big[\sum_y P(x;\,y)\Big] + \sum_y y\Big[\sum_x P(x;\,y)\Big] = \\ &= \sum_x xP_\xi(x) + \sum_y yP_\eta(y) = \boldsymbol{E}\xi + \boldsymbol{E}\eta.\end{aligned}$$

Доказательство для произвольных случайных величин: по формуле (53) п. 53:

$$\begin{aligned}\boldsymbol{E}(\xi+\eta) &= \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} (x+y)\,ddF(x;\,y) = \\ &= \int\limits_{-\infty}^{+\infty} x\,d\int\limits_{-\infty}^{+\infty} d_yF(x;\,y) + \int\limits_{-\infty}^{+\infty} y\,d\int\limits_{-\infty}^{+\infty} d_xF(x;\,y) = \\ &= \int\limits_{-\infty}^{+\infty} x\,dF_\xi(x) + \int\limits_{-\infty}^{+\infty} y\,dF_\eta(y) = \boldsymbol{E}\xi + \boldsymbol{E}\eta.\end{aligned}$$

Обобщение теоремы XI. — *Среднее значение суммы любого конечного числа случайных величин складывается из средних значений этих величин:*

$$\boldsymbol{E}(\xi_1+\xi_2+\ldots+\xi_s) = \boldsymbol{E}\xi_1 + \boldsymbol{E}\xi_2 + \ldots + \boldsymbol{E}\xi_s.$$

Доказательство. — В силу теоремы XI это верно для двух слагаемых. Докажем, что если это верно для $s-1$ слагаемых, то верно и для s слагаемых. Итак, пусть $\xi' = \xi_1+\xi_2+\ldots+\xi_{s-1}$ и

$$\boldsymbol{E}\xi' = \boldsymbol{E}\xi_1 + \boldsymbol{E}\xi_2 + \ldots + \boldsymbol{E}\xi_{s-1}.$$

Тогда $\xi_1+\xi_2+\ldots+\xi_s = \xi'+\xi_s$. Поэтому в силу теоремы XI

$$\boldsymbol{E}(\xi_1+\xi_2+\ldots+\xi_s) = \boldsymbol{E}\xi' + \boldsymbol{E}\xi_s.$$

Подставляя сюда написанное выше выражение для $E\xi'$, мы и получаем требуемое равенство.

Теорема XII. — *Среднее значение произведения двух независимых случайных величин равно произведению их средних значений:*

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

Доказательство для дискретных случайных величин:

$$\begin{aligned} E(\xi\eta) &= \sum_x \sum_y xyP(x;\, y) = \\ &= \sum_x \sum_y xyP_\xi(x)\,P_\eta(y) = \sum_y y\,[\sum_x xP_\xi(x)]\,P_\eta(y) = \\ &= \sum_x xP_\xi(x) \cdot \sum_y yP_\eta(y) = E\xi \cdot E\eta. \end{aligned}$$

Доказательство для произвольных случайных величин: по формуле (54) п. 53

$$\begin{aligned} E(\xi\eta) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy\, ddF(x;\, y) = \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy\, dd\,[F_\xi(x)\,F_\eta(y)] = \int_{-\infty}^{+\infty} y\,[\int_{-\infty}^{+\infty} xdF_\xi(x)]\,dF_\eta(y) = \\ &= \int_{-\infty}^{+\infty} xdF_\xi(x) \cdot \int_{-\infty}^{+\infty} ydF_\eta(y) = E\xi \cdot E\eta. \end{aligned}$$

Приведем два простых примера применения доказанных теорем. Во-первых, покажем, что

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

Действительно, в силу теорем XI и X

$$E(\xi - \eta) = E[\xi + (-1)\,\eta] = E\xi + (-1)\,E\eta = E\xi - E\eta.$$

Во-вторых, назовем *отклонением* случайной величины ξ разность $\xi - E\xi$. Покажем, что среднее значение отклонения всегда равно нулю:

$$E(\xi - E\xi) = 0.$$

Действительно, так как $E\xi$ есть постоянная, то в силу только что доказанного равенства и теоремы IX

$$E(\xi - E\xi) = E\xi - E(E\xi) = E\xi - E\xi = 0.$$

57. Вычислим среднее значение величины ξ, подчиняющейся закону Пуассона:

$$E\xi = \sum_{x=0}^{\infty} x\left(\frac{k^x}{x!}\,e^{-k}\right) = e^{-k} \sum_{x=1}^{\infty} x\,\frac{k^x}{x!} = e^{-k}k \sum_{x=1}^{\infty} \frac{k^{x-1}}{(x-1)!} = e^{-k}ke^k = k.$$

Вычислим еще среднее значение величины ξ, подчиняющейся закону Гаусса:

$$\boldsymbol{E}\xi = \frac{1}{\sqrt{2\pi\beta}}\int\limits_{-\infty}^{+\infty} xe^{-\frac{(x-a)^2}{2\beta}}\,dx = \frac{\sqrt{2\beta}}{\sqrt{\pi}}\int\limits_{-\infty}^{+\infty} se^{-s^2} + \frac{a}{\sqrt{\pi}}\int\limits_{-\infty}^{+\infty} e^{-s^2}ds = a;$$

первый из двух интегралов, взятых по переменному s, исчезает вследствие того, что график подинтегральной функции расположен симметрично относительно начала координат.

58. Теорема XI *о сложении средних значений* находит изящное применение при вычислении среднего значения числа μ наступлений события A при n независимых испытаниях. Для этого мы представим случайную величину μ в виде суммы случайных величин с более простыми законами распределения. Пусть:

μ_1 есть число наступлений события A при первом испытании,
μ_2 есть число наступлений события A при втором испытании,
. . . .
μ_n есть число наступлений события A при n-м испытании.

Тогда

$$\mu = \mu_1 + \mu_2 + \ldots + \mu_n.$$

При каждом отдельном испытании событие A может только либо один раз наступить, либо ни одного раза не наступить; первое — с постоянной вероятностью p, второе — с постоянной вероятностью $1-p$. Таким образом при каждом i величина μ_i может принимать два значения: значение 1 с вероятностью p и значение 0 с вероятностью $1-p$.

Отсюда при каждом i

$$\boldsymbol{E}\mu_i = 1\cdot p + 0\cdot(1-p) = p. \tag{55}$$

Следовательно, в силу теоремы XI

$$\boldsymbol{E}\mu = \boldsymbol{E}\mu_1 + \boldsymbol{E}\mu_2 + \ldots + \boldsymbol{E}\mu_n = np.$$

Если воспользоваться еще теоремой X, то отсюда можно получить интересное следствие для случайной величины $\frac{\mu}{n}$ — относительной частоты наступления события A при n испытаниях. В силу теоремы X мы получим

$$\boldsymbol{E}\left(\frac{\mu}{n}\right) = \frac{1}{n}\boldsymbol{E}\mu = p. \tag{56}$$

59. Положим

$$\boldsymbol{D}\xi = \boldsymbol{E}(\xi - \boldsymbol{E}\xi)^2.$$

Эта постоянная $\boldsymbol{D}\xi$, среднее значение квадрата отклонения величины ξ, является одной из важных числовых характеристик случайной величины и называется *дисперсией* этой величины.

Роль дисперсии как числовой характеристики случайной величины может быть до известной степени выяснена на одном простом примере.

Возьмем случайную величину ξ, которая имеет плотность вероятности $p(x)$, равную положительной постоянной k в некотором интервале (a, b) и равную нулю вне этого интервала. Такая случайная величина называется величиной с *равномерным законом распределения*. Это — величина такого же типа, как непрерывные случайные величины; здесь только в двух точках $x = a$ и $x = b$ нарушается требование непрерывности функции $p(x)$, налагавшееся нами на непрерывные случайные величины, но для вопроса, который мы собираемся разобрать, это обстоятельство не играет никакой роли.

Определим прежде всего постоянную k. Так как

$$\int_{-\infty}^{+\infty} p(x)\,dx = 1,$$

что в данном случае означает

$$\int_a^b k\,dx = 1,$$

то

$$k = \frac{1}{b-a}.$$

Вычислим теперь среднее значение и дисперсию величины ξ. Мы получим последовательно

$$\boldsymbol{E}\xi = \int_{-\infty}^{+\infty} xp(x)\,dx = \frac{1}{b-a}\int_a^b x\,dx = \frac{a+b}{2},$$

$$\boldsymbol{D}\xi = \int_{-\infty}^{+\infty} \left[x - \int_{-\infty}^{+\infty} sp(s)\,ds\right]^2 dx = \frac{1}{b-a}\int_a^b \left(x - \frac{a+b}{2}\right)^2 dx = \frac{(b-a)^2}{12}.$$

Мы видим, что среднее значение есть середина интервала (a, b), в котором лежат возможные значения величины ξ. Дисперсия же зависит только от длины $b-a$ интервала (a, b) и является возрастающей функцией этой длины. Если мы будем увеличивать интервал (a, b) и тем самым уменьшать плотность вероятности $k = \frac{1}{b-a}$, т. е. если мы будем присоединять новые, более далекие от среднего, возможные значения величины ξ за счет уменьшения вероятности более близких к среднему значений, то дисперсия будет расти. Дисперсия играет роль, так сказать, меры случайного рассеяния значений данной величины вокруг среднего значения.

60. Для дисперсий справедливы следующие теоремы:

Теорема XIII. — *Дисперсия постоянной равна нулю:*

$$\boldsymbol{D}k = 0.$$

Доказательство. — В силу теорем X и IX

$$\boldsymbol{D}k = \boldsymbol{E}(k - \boldsymbol{E}k)^2 = \boldsymbol{E}(k-k)^2 = \boldsymbol{E}0 = 0.$$

Теорема XIV. — *Постоянный множитель можно выносить за знак дисперсии, возводя его в квадрат:*

$$D(k\xi) = k^2 D(\xi).$$

Доказательство. — В силу теоремы X

$$D(k\xi) = E[k\xi - E(k\xi)]^2 = E[k(\xi - E\xi)]^2 = k^2 E(\xi - E\xi)^2 = k^2 D\xi.$$

Теорема XV. — *Дисперсия суммы двух независимых случайных величин равна сумме дисперсий этих величин:*

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

Доказательство. — Когда величины ξ и η независимы, то независимы, очевидно, и их отклонения $\xi - E\xi$ и $\eta - E\eta$. Следовательно, по теореме XII

$$E[(\xi - E\xi)(\eta - E\eta)] = E(\xi - E\xi) \cdot E(\eta - E\eta) = 0.$$

Отсюда, на основании теорем XI и X

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

Первое обобщение теоремы XV. — *Если $\xi_1, \xi_2, \ldots, \xi_n$ — случайные величины, каждая из которых, ξ_k, независима от суммы предшествующих, $\xi_1 + \xi_2 + \ldots + \xi_{k-1}$, то дисперсия суммы всех этих величин складывается из дисперсий каждой из этих величин в отдельности:*

$$D(\xi_1 + \xi_2 + \ldots + \xi_n) = D\xi_1 + D\xi_2 + \ldots + D\xi_n.$$

Доказательство проводится точно так же, как доказательство данного выше обобщения теоремы XI: следует только заменить всюду средние значения дисперсиями.

Второе обобщение теоремы XV. — *Дисперсия суммы любого конечного числа попарно независимых случайных величин равна сумме дисперсий этих величин:*

$$D(\xi_1 + \xi_2 + \ldots + \xi_s) = D\xi_1 + D\xi_2 + \ldots + D\xi_s.$$

Доказательство точно такое же, как и для самой теоремы XV. Только вместо равенства $(\bar\xi + \bar\eta)^2 = \bar\xi^2 + \bar\eta^2 + 2\bar\xi\bar\eta$, где $\bar\xi = \xi - E\xi$ и $\bar\eta = \eta - E\eta$, здесь применяется более общее равенство:

$$\left(\sum_{i=1}^{s} \bar\xi_i\right)^2 = \sum_{i=1}^{s} \bar\xi_i^2 + \sum_{i \neq j} \bar\xi_i \bar\xi_j,$$

где $\bar\xi_i = \xi_i - E\xi_i$.

Теорема XVI. — *Дисперсия случайной величины всегда равна разности между средним значением квадрата величины и квадратом среднего значения самой величины:*

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

Доказательство. — На основании теорем XI, X, и IX, причем последние две применяются к постоянным $2E\xi$ и $(E\xi)^2$, мы имеем

$$D\xi = E(\xi - E\xi)^2 = E\xi^2 - 2E\xi \cdot E\xi + (E\xi)^2 = E\xi^2 - (E\xi)^2.$$

Дисперсия всегда неотрицательна:

$$D\xi \geqslant 0.$$

Действительно, $(\xi - E\xi)^2 \geqslant 0$, почему, очевидно, и $E(\xi - E\xi)^2 \geqslant 0$.

Отсюда и из теоремы XVI следует, что

$$E\xi^2 \geqslant (E\xi)^2.$$

Это неравенство, известно в анализе под именем неравенства Шварца.

Приведем еще два простых примера применения доказанных теорем о дисперсиях. Во-первых, покажем, что если случайные величины ξ и η независимы, то

$$D(\xi - \eta) = D\xi + D\eta.$$

Действительно, в силу теорем XV и XIV,

$$D(\xi - \eta) = D[\xi + (-1)\eta] = D\xi + (-1)^2 D\eta = D\xi + D\eta.$$

Во-вторых, назовем *нормированным отклонением* случайной величины ξ величину $\frac{\xi - E\xi}{\sqrt{D\xi}}$. Покажем, что дисперсия нормированного отклонения всегда равна единице:

$$D\left(\frac{\xi - E\xi}{\sqrt{D\xi}}\right) = 1.$$

Действительно, ξ и $E\xi$, рассматриваемые как случайные величины, независимы, так как вторая из них постоянная; поэтому в силу только что доказанного равенства и теорем XIII и XIV

$$D\left(\frac{\xi - E\xi}{\sqrt{D\xi}}\right) = \frac{D\xi + D(E\xi)}{D\xi} = \frac{D\xi}{D\xi} = 1.$$

Постоянные $E(\xi - E\xi)^\nu$, где ν — любое неотрицательное число, называются *центральными моментами*, постоянные $E|\xi - E\xi|^\nu$ — *абсолютными центральными моментами*, порядка ν, случайной величины ξ. Дисперсия есть центральный и одновременно, очевидно, абсолютный центральный момент второго порядка.

61. Вычислим дисперсии тех случайных величин, для которых мы вычисляли раньше средние значения. Для случайной величины ξ, подчиняющейся закону Пуассона, мы имели $\boldsymbol{E}\xi = k$, и мы получим

$$\boldsymbol{E}\xi^2 = \sum_{x=0}^{\infty} x^2 \frac{k^x}{x!} e^{-k} = e^{-k} k \sum_{x=1}^{\infty} x \frac{k^{x-1}}{(x-1)!} =$$

$$= e^{-k} k \sum_{x=1}^{\infty} [(x-1)+1] \frac{k^{x-1}}{(x-1)!} = e^{-k} k^2 \sum_{x=2}^{\infty} \frac{k^{x-2}}{(x-2)!} +$$

$$+ e^{-k} k \sum_{x=1}^{\infty} \frac{k^{x-1}}{(x-1)!} = e^{-k} k^2 e^k + e^{-k} k e^k = k^2 + k,$$

следовательно, по теореме XVI

$$\boldsymbol{D}\xi = \boldsymbol{E}\xi^2 - (\boldsymbol{E}\xi)^2 = k.$$

Для случайной величины ξ, подчиняющейся закону Гаусса, мы имели $\boldsymbol{E}\xi = a$, и мы получим

$$\boldsymbol{D}\xi = \frac{1}{\sqrt{2\pi\beta}} \int_{-\infty}^{+\infty} (x-a)^2 e^{-\frac{(x-a)^2}{2\beta}} dx = \frac{2\beta}{\sqrt{\pi}} \int_{-\infty}^{+\infty} s^2 e^{-s^2} ds;$$

введем вспомогательный параметр $\gamma > 0$ и вычислим интеграл

$$I(\gamma) = \frac{2\beta}{\sqrt{\pi}} \int_{-\infty}^{+\infty} s^2 e^{-\gamma s^2} ds =$$

$$= -\frac{2\beta}{\sqrt{\pi}} \frac{d}{d\gamma} \int_{-\infty}^{+\infty} e^{-\gamma s^2} ds = -\frac{2\beta}{\sqrt{\pi}} \frac{d}{d\gamma} \left(\frac{\sqrt{\pi}}{\sqrt{\gamma}} \right) = \frac{\beta}{\sqrt{\gamma^3}};$$

отсюда

$$\boldsymbol{D}\xi = I(1) = \beta.$$

Здесь случайное рассеяние значений данной величины вокруг среднего значения возрастает с возрастанием β; это наглядно видно на нашем черт. 4 (стр. 67), где показаны плотности вероятности величин, подчиняющихся закону Гаусса, со значениями $\beta = 1$ и $\beta = \frac{1}{4}$.

62. Перейдем к числу μ наступлений события A при n независимых испытаниях. Представим, как в п. 58, случайную величину μ в виде суммы величин

$$\mu = \mu_1 + \mu_2 + \ldots + \mu_n.$$

Что касается этих последних, то каждая из них, μ_i, может принимать только значения 1 и 0 с вероятностями p и $1-p$. Поэтому и квадраты этих величин, т. е. каждая из величин μ_i^2, могут принимать только значения 1 и 0, притом с теми же вероятностями p и $1-p$. Таким образом $\boldsymbol{E}\mu_i^2 = \boldsymbol{E}\mu_i$. Но мы имели $\boldsymbol{E}\mu_i = p$. Отсюда по теореме XVI

$$\boldsymbol{D}\mu_i = \boldsymbol{E}\mu_i^2 - (\boldsymbol{E}\mu_i)^2 = p - p^2 = p(1-p). \tag{57}$$

Наконец, случайные величины μ_i, очевидно, попарно независимы. Следовательно, в силу теоремы XV о сложении дисперсий

$$\boldsymbol{D}\mu = \boldsymbol{D}\mu_1 + \boldsymbol{D}\mu_2 + \ldots + \boldsymbol{D}\mu_n = np(1-p).$$

Что же касается относительной частоты $\frac{\mu}{n}$ наступления события A при n независимых испытаниях, то ее дисперсия в силу теоремы XIV равна

$$\boldsymbol{D}\left(\frac{\mu}{n}\right) = \frac{1}{n^2}\boldsymbol{D}\mu = \frac{p(1-p)}{n}. \tag{58}$$

Мы видим, что дисперсия числа наступлений события возрастает с увеличением числа испытаний, а дисперсия относительной частоты наступления события, наоборот, убывает с увеличением числа испытаний. Чем больше испытаний, тем больше и случайное рассеяние числа наступлений события вокруг его среднего значения np. И наоборот, чем больше испытаний, тем меньше случайное рассеяние относительной частоты наступления события вокруг его среднего значения p.

ГЛАВА V.

ЗАКОН БОЛЬШИХ ЧИСЕЛ.

§ 17. Вероятности и массовые явления.

63. Математическая теория, которую мы изложили в гл. I—IV, дает возможность лишь выводить те или иные значения вероятностей из других, заранее заданных значений вероятностей. Результатами, которые можно извлечь из этой теории, могут быть лишь утверждения вида „такое-то событие имеет такую-то вероятность".

Можем ли мы на основании этого предсказать что-либо относительно действительного наступления или ненаступления события? Вообще говоря, ничего. Так например, мы знаем, что при однократном бросании игральной кости вероятность выпадения одного очка значительно меньше, нежели вероятность выпадения более чем одного очка. И тем не менее, бросая один раз игральную кость, мы ничего не можем сказать заранее о том, выпадет или не выпадет как раз одно очко.

Но в известных случаях мы все же предсказываем, и даже больше, чем это буквально вытекает из данных математической теории. Мы руководствуемся при этом следующим принципом: *если при некоторых условиях вероятность события A очень мала, то можно быть практически уверенным в том, что при осуществлении этих условий событие A не произойдет, т. е. можно на практике поступать так, как если бы мы были уверены, что это событие не произойдет.*

Так, например, будем бросать игральную кость не один раз, а скажем, сто раз подряд. Событие, состоящее в том, что одно очко выпадет все сто раз, имеет очень малую вероятность; и приступая к бросанию игральной кости сто раз подряд, мы, конечно, будем практически уверены в том, что такое событие не произойдет.

Названный п р и н ц и п п р а к т и ч е с к о й у в е р е н н о с т и, и только он, дает возможность применять теорию вероятностей к действительности. Мы руководствуемся им во всех наших действиях и исследованиях, потому что он подтверждается всей нашей практикой. Разумеется, степень малости вероятности события A, при которой мы получаем практическую уверенность в невозможности события A, различна в различных конкретных условиях. Принимая ту или иную малую величину вероятности события за „практическую невозможность" этого события, всегда приходится делать некоторое допущение, гипотезу, и вопрос о том, верно или неверно сделанное допущение, решает согласие окончательных выводов с данными опыта.

Всем этим, однако, нисколько не умаляется роль математической теории, роль вычисленных с ее помощью различных, отнюдь не очень малых вероятностей. В этом можно убедиться уже на таком простом примере. Допустим, что событие, вероятность которого меньше ε, практически невозможно, и пусть на основании математической теории вычислено, что некоторое данное событие A имеет вероятность p, превосходящую $1-\varepsilon$:

$$p > 1 - \varepsilon.$$

Тогда противоположное ему событие $\overline{A}$ должно иметь вероятность $1-p$, меньшую, чем ε:

$$1 - p < \varepsilon.$$

Следовательно, событие $\overline{A}$ практически невозможно. Но ненаступление события $\overline{A}$ есть наступление события A; следовательно, событие A „практически достоверно“, т. е. можно на практике поступать так, как если бы мы были уверены в том, что оно произойдет.

Кстати сказать, в связи с этим на практике часто делают допущение не о степени малости вероятности, при которой событие оказывается практически невозможным, а о степени близости вероятности к единице, при которой событие оказывается практически достоверным. Как мы видим, одно к другому легко сводится.

Математическая теория позволяет делать и значительно более общие выводы. Так, например, в дальнейшем будет показано следующее. Пусть μ — число наступлений события A при n независимых испытаниях и пусть p — вероятность наступления этого события при отдельном испытании. Зададим как угодно малое положительное число ε и обозначим через B_n событие, состоящее в том, что случайная величина μ примет значение, удовлетворяющее неравенству

$$\left|\frac{\mu}{n} - p\right| < \varepsilon,$$

т. е. в том, что относительная частота наступления события A будет отличаться от p меньше, чем на ε. В дальнейшем будет показано (теорема Бернулли)[1], что

$$\lim_{n\to\infty} \boldsymbol{P}(B_n) = 1.$$

Следовательно, при достаточно большом n событие B_n непременно становится практически достоверным. Отсюда по самому смыслу события B_n мы вправе сделать такое заключение: увеличивая число испытаний, всегда можно достичь практической уверенности в том, что относительная частота наступления события A будет как угодно мало отличаться от p. Мы видим, таким образом, что когда математическая теория позволяет нам вычислить вероятность p какого-нибудь события A, мы можем с помощью принципа практической уве-

[1]) Впрочем, теорема Бернулли уже была доказана мимоходом в п. 39.

ренности предсказать кое-что и о действительном наступлении или ненаступлении этого события. Так, при бросании игральной кости вероятность выпадения одного очка равна $\frac{1}{6}$. Следовательно, при многократном бросании кости можно достичь практической уверенности в том, что относительная частота выпадения одного очка будет к к угодно мало отличаться от $\frac{1}{6}$. •

В настоящее время известно несколько обобщений теоремы Бернулли. Они имеют аналогичное значение для применения теории вероятностей к действительности. Теоремы такого типа носят общее имя *закона больших чисел.* Наиболее известные из них (теоремы Чебышева и Маркова) будут изучены в настоящей главе.

64. Предварительно, однако, вернемся еще раз к принципу практической уверенности. Возьмем наш пример бросания игральной кости сто раз подряд. Мы видели, что при этом практически невозможно выпадение такого ряда R чисел очков, который состоял бы из одних единиц; это имеет слишком малую вероятность $\left(\frac{1}{6}\right)^{100}$. Но в таком же положении находится и любой другой заранее определенный ряд чисел очков, как бы ни были расположены в нем эти числа; его выпадение также практически невозможно, так как имеет ту же вероятность $\left(\frac{1}{6}\right)^{100}$. Чего же мы ожидаем, когда отказываемся ожидать выпадения ряда R? Мы ожидаем, что выпадет хотя бы один из рядов, отличных от R. Иначе говоря, мы не ожидаем, что произойдет то или иное индивидуальное событие, а ожидаем, что произойдет событие, равносильное наступлению хотя бы одного из большого числа индивидуальных событий.

Приведем еще аналогичный пример. Возьмем (как в § 4) вместо игральной кости шарик и мысленно отметим на нем одну точку P. Выпадение (в смысле § 4) отмеченной точки уже при однократном бросании шарика имеет вероятность, равную нулю. Это еще не свидетельствует о том, что оно невозможно. Мы в своем месте доказали, что вероятность невозможного события равна нулю, но обратно, из того, что вероятность события равна нулю, еще не следует, что это событие невозможно. Однако ясно, что выпадение отмеченной точки в нашем смысле практически невозможно; можно держать какое угодно пари, что при бросании шарика наугад отмеченная точка не выпадет. Но точно в таком же положении находится и любая другая точка. Спросим себя опять, чего же мы ожидаем, когда отказываемся ожидать выпадения точки P? Подобно тому, как в предыдущем примере, мы ожидаем, что выпадет хотя бы одна из точек, отличных от P. Иначе говоря, мы опять не ожидаем, что произойдет то или иное индивидуальное событие, а ожидаем, что произойдет событие, равносильное наступлению хотя бы одного из большого (на этот раз даже бесконечного) числа индивидуальных событий.

Несложное рассуждение покажет нам, что приведенные примеры должны быть типичны для применения теории вероятностей. Если мы хотим иметь практическую уверенность в наших выводах, мы должны иметь дело с очень маловероятными событиями; или же мы будем иметь дело с событиями, вероятность которых очень близка к единице, но тогда в рассматриваемом нами поле событий автоматически появятся и противоположные им события, т. е. опять-таки очень мало вероятные события. Для того же, чтобы иметь при этом возможность сравнивать вероятности отдельных событий, мы, естественно, должны будем подразделять и все рассматриваемое нами поле вероятностей на очень мало вероятные события. Иначе говоря, чтобы иметь практическую уверенность в наших выводах, мы должны иметь дело с индивидуальными событиями очень малой вероятности; следовательно, мы сможем предсказывать, как правило, наступление не этих индивидуальных событий, а лишь событий, равносильных наступлению хотя бы одного из большого числа индивидуальных событий.

Мы приходим, таким образом, к выводу, что единственным объектом применения теории вероятностей к действительности могут быть не индивидуальные, а, так сказать, массовые события. В этом смысле теория вероятностей есть наука о массовых явлениях.

§ 18. Теоремы Бéрнулли, Чебышева и Маркова.

65. Теоремы Бéрнулли, Чебышева и Маркова проще всего выводятся из так называемого неравенства Чебышева. Его можно представить в различных формах. Мы приведем две формы этого неравенства. Первую мы докажем непосредственно, вторую выведем из первой, а из второй, в свою очередь, будем выводить названные теоремы.

Неравенство Чебышева (первая форма). — *Какова бы ни была случайная величина ξ, которая может принимать только неотрицательные значения и имеет конечное среднее значение*

$$\boldsymbol{P}(\xi \geqslant 1) \leqslant \boldsymbol{E}\xi.$$

Доказательство для дискретных случайных величин. — Пусть закон распределения величины ξ есть $P(x)$. Тогда

$$\boldsymbol{P}(\xi \geqslant 1) = \sum_{x \geqslant 1} P(x) \leqslant \sum_{x \geqslant 1} xP(x) \leqslant \sum_{x} xP(x) = \boldsymbol{E}\xi,$$

что можно пояснить шаг за шагом следующим образом:

1) В силу теоремы III (или расширенной аксиомы сложения вероятностей) вероятность того, что величина ξ примет значение, не меньшее чем 1, складывается из вероятностей того, что она примет то или иное из значений x, не меньших чем 1:

$$\boldsymbol{P}(\xi \geqslant 1) = \sum_{x \geqslant 1} P(x).$$

2) Только что написанная сумма распространяется лишь на те значения x, которые не меньше 1. Поэтому, если мы будем умножать

члены $P(x)$ этой суммы на соответствующие значения x, то эти члены могут разве лишь увеличиться. Следовательно, при этом и вся сумма может разве лишь увеличиться:

$$\sum_{x \geqslant 1} P(x) \leqslant \sum_{x \geqslant 1} xP(x).$$

3) По условию величина ξ может принимать только неотрицательные значения x, поэтому и выражения вида $xP(x)$, где x — любое возможное значение величины ξ, а $P(x)$ — вероятность этого значения, могут быть только неотрицательными. Следовательно, заменяя сумму выражений $xP(x)$, где x — значения величины, удовлетворяющие неравенству $x \geqslant 1$, суммой таких же выражений, где x — любые возможные значения величины ξ, мы можем разве лишь увеличить первую сумму:

$$\sum_{x \geqslant 1} xP(x) \leqslant \sum_{x} xP(x).$$

4) По определению среднего значения

$$\sum_{x} xP(x) = \boldsymbol{E}\xi.$$

Доказательство для произвольных случайных величин. — Пусть функция распределения величины ξ есть $F(x)$. Тогда

$$\boldsymbol{P}(\xi \geqslant 1) = \int\limits_{x \geqslant 1} dF(x) \leqslant \int\limits_{x \geqslant 1} x\,dF(x) \leqslant \int\limits_{-\infty}^{+\infty} x\,dF(x) = \boldsymbol{E}\xi,$$

что можно пояснить шаг за шагом так же, как выше.

Неравенство Чебышева (вторая форма). — *Какова бы ни была случайная величина ξ, которая имеет конечное среднее значение и конечную дисперсию, и какова бы ни была положительная постоянная ε:*

$$\boldsymbol{P}(|\xi - \boldsymbol{E}\xi| < \varepsilon) \geqslant 1 - \frac{\boldsymbol{D}\xi}{\varepsilon^2}.$$

Словесно это удобнее всего читать так: неравенство $|\xi - \boldsymbol{E}\xi| < \varepsilon$ имеет вероятность, не меньшую чем $1 - \frac{\boldsymbol{D}\xi}{\varepsilon^2}$.

Доказательство. — Рассмотрим сначала не само интересующее нас неравенство $|\xi - \boldsymbol{E}\xi| < \varepsilon$, а противоположное ему неравенство $|\xi - \boldsymbol{E}\xi| \geqslant \varepsilon$. Это последнее, очевидно, равносильно неравенству $\frac{(\xi - \boldsymbol{E}\xi)^2}{\varepsilon^2} \geqslant 1$. Имея в виду это обстоятельство, применяя первую форму неравенства Чебышева к неотрицательной случайной величине $\frac{(\xi - \boldsymbol{E}\xi)^2}{\varepsilon^2}$ и, наконец, опираясь на теоремы X и XVI, мы получаем

$$\boldsymbol{P}(|\xi - \boldsymbol{E}\xi| \geqslant \varepsilon) = \boldsymbol{P}\left[\frac{(\xi - \boldsymbol{E}\xi)^2}{\varepsilon^2} \geqslant 1\right] \leqslant \boldsymbol{E}\left[\frac{(\xi - \boldsymbol{E}\xi)^2}{\varepsilon^2}\right] = \frac{\boldsymbol{D}\xi}{\varepsilon^2}.$$

Отсюда, возвращаясь к интересующему нас неравенству, мы получаем

$$P(|\xi - E\xi| < \varepsilon) = 1 - P(|\xi - E\xi| \geqslant \varepsilon) \geqslant 1 - \frac{D\xi}{\varepsilon^2}.$$

66. Теорема Бéрнулли. — *Пусть μ — число наступлений события A при n независимых испытаниях и пусть p — постоянная вероятность наступления события A при отдельном испытании. Тогда, какова бы ни была положительная постоянная ε:*

$$\lim_{n \to \infty} P\left(\left|\frac{\mu}{n} - p\right| < \varepsilon\right) = 1.$$

Доказательство. — Пользуясь равенствами (56) п. 58 и (58) п. 62, в силу которых $E\left(\frac{\mu}{n}\right) = p$ и $D\left(\frac{\mu}{n}\right) = \frac{p(1-p)}{n}$, и применяя вторую форму неравенства Чебышева к случайной величине $\frac{\mu}{n}$, мы получаем

$$P\left(\left|\frac{\mu}{n} - p\right| < \varepsilon\right) = P\left(\left|\frac{\mu}{n} - E\left(\frac{\mu}{n}\right)\right| < \varepsilon\right) \geqslant 1 - \frac{D\left(\frac{\mu}{n}\right)}{\varepsilon^2} =$$

$$= 1 - \frac{p(1-p)}{\varepsilon^2 n}.$$

Поэтому, переходя к пределу при $n \to \infty$, мы получаем

$$\lim_{n \to \infty} P\left(\left|\frac{\mu}{n} - p\right| < \varepsilon\right) \geqslant 1,$$

а так как вероятность вообще не может быть больше единицы, то отсюда и вытекает требуемое равенство.

Значение этой теоремы выяснено в п. 63: увеличивая число испытаний, можно достичь практической уверенности в том, что относительная частота наступления события A будет как угодно мало отличаться от вероятности наступления события A при отдельном испытании.

67. Теорема Чебышева. — *Пусть $\xi_1, \ldots, \xi_n, \ldots$ — последовательность попарно независимых случайных величин, удовлетворяющих следующим условиям*

1° *Все величины $\xi_1, \ldots, \xi_n, \ldots$ имеют одно и то же среднее значение:*

$$E\xi_1 = \ldots = E\xi_n = \ldots = a.$$

2° *Дисперсии величин $\xi_1, \ldots, \xi_n, \ldots$ конечны и равномерно ограничены, т. е.*

$$D\xi_1 \leqslant C, \ \ldots, \ D\xi_n \leqslant C, \ \ldots,$$

где C — постоянная.

Тогда, какова бы ни была положительная постоянная ε:

$$\lim_{n \to \infty} P\left(\left|\frac{\xi_1 + \ldots + \xi_n}{n} - a\right| < \varepsilon\right) = 1.$$

Доказательство. — В силу теорем X и XI

$$E\left(\frac{\xi_1+\ldots+\xi_n}{n}\right)=\frac{E\xi_1+\ldots+E\xi_n}{n}=a.$$

Таким образом, применяя вторую форму неравенства Чебышева к случайной величине $\frac{\xi_1+\ldots+\xi_n}{n}$, мы получаем

$$P\left(\left|\frac{\xi_1+\ldots+\xi_n}{n}-a\right|<\varepsilon\right)=$$

$$=P\left[\left|\frac{\xi_1+\ldots+\xi_n}{n}-E\left(\frac{\xi_1+\ldots+\xi_n}{n}\right)\right|<\varepsilon\right]\geqslant 1-\frac{D\left(\frac{\xi_1+\ldots+\xi_n}{n}\right)}{\varepsilon^2}.$$

Отсюда, опираясь на теоремы XIV и XV, мы получаем

$$P\left(\left|\frac{\xi_1+\ldots+\xi_n}{n}-a\right|<\varepsilon\right)\geqslant 1-\frac{D\xi_1+\ldots+D\xi_n}{n^2\varepsilon^2}\geqslant 1-\frac{C}{n\varepsilon^2}.$$

Поэтому, переходя к пределу при $n\to\infty$, мы получаем

$$\lim_{n\to\infty}P\left(\left|\frac{\xi_1+\ldots+\xi_n}{n}-a\right|<\varepsilon\right)\geqslant 1,$$

откуда и вытекает требуемое равенство.

Значение этой теоремы можно пояснить на следующем примере. Пусть нам предстоит измерить некоторую неизвестную физическую постоянную a. Будем производить ряд независимых друг от друга измерений. Результат каждого из этих измерений будет случайной величиной. Пусть $\xi_1, \ldots, \xi_n, \ldots$ — эти случайные величины. Допустим, как это делает теория ошибок наблюдений, что средние значения этих величин одинаковы и равны как раз измеряемой постоянной a, т. е.

$$E\xi_1=\ldots=E\xi_n=\ldots=a.$$

В сущности, это — не что иное, как точная формулировка допущения, что наши измерения свободны от систематических погрешностей. Допустим, кроме того, что дисперсии этих величин равномерно ограничены, т. е.

$$D\xi_1<C, \ldots, D\xi_n<C, \ldots,$$

где C — постоянная. Это — не что иное, как точная формулировка допущения, что наши измерения производятся с гарантией какой-нибудь, хотя бы и небольшой, но не безгранично убывающей точности, так что хотя условия наблюдения и могут изменяться, но при этом случайные рассеяния результатов наблюдений вокруг измеряемой величины не могут безгранично возрастать. Сделанные два допущения дают возможность применить к величинам $\xi_1, \ldots, \xi_n, \ldots$ теорему Чебышева:

$$\lim_{n\to\infty}P\left(\left|\frac{\xi_1+\ldots+\xi_n}{n}-a\right|<\varepsilon\right)=1.$$

Это значит, что, увеличивая число наблюдений, можно достичь практической уверенности в том, что среднее арифметическое из результатов наблюдений будет как угодно мало отличаться от измеряемой постоянной.

Нетрудно убедиться в том, что теорема Бéрнулли есть частный случай теоремы Чебышева. Для этого положим, как в п. 58 и 62:

$$\mu = \mu_1 + \ldots + \mu_n.$$

Величины $\mu_1, \ldots, \mu_n$ попарно независимы, и согласно равенствам (55) п. 58 и (57) п. 62

1° $\boldsymbol{E}\mu_1 = \ldots = \boldsymbol{E}\mu_n = p$,
2° $\boldsymbol{D}\mu_1 = \ldots = \boldsymbol{D}\mu_n = p(1-p)$.

Таким образом здесь можно применить теорему Чебышева, и мы получим

$$\lim_{n \to \infty} \boldsymbol{P}\left(\left|\frac{\mu}{n} - p\right| < \varepsilon\right) = \lim_{n \to \infty} \boldsymbol{P}\left(\left|\frac{\mu_1 + \ldots + \mu_n}{n} - p\right| < \varepsilon\right) = 1,$$

т. е. получим теорему Бéрнулли.

68. Пользуясь неравенствами Чебышева и следуя тому же ходу мысли, что и при доказательстве теоремы Чебышева, легко получить и следующую, еще более общую формулировку закона больших чисел.

Теорема Маркова. — *Пусть $\xi_1, \ldots, \xi_n, \ldots$ — последовательность попарно независимых случайных величин, удовлетворяющих следующим условиям:*

1° *Каждая величина ξ_k имеет конечное среднее значение a_k.*
2° *Каждая величина ξ_k имеет конечную дисперсию b_k, и эти дисперсии таковы, что при $n \to \infty$*

$$\frac{b_1 + \ldots + b_n}{n^2} \to 0.$$

Тогда, какова бы ни была положительная постоянная ε,

$$\lim_{n \to \infty} \boldsymbol{P}\left(\left|\frac{\xi_1 + \ldots + \xi_n}{n} - \frac{a_1 + \ldots + a_n}{n}\right| < \varepsilon\right) = 1.$$

Доказательство. — В силу теорем X и XI

$$\boldsymbol{E}\left(\frac{\xi_1 + \ldots + \xi_n}{n}\right) = \frac{a_1 + \ldots + a_n}{n}.$$

Таким образом, применяя вторую форму неравенства Чебышева к случайной величине $\frac{\xi_1 + \ldots + \xi_n}{n}$, мы получаем

$$\boldsymbol{P}\left(\left|\frac{\xi_1 + \ldots + \xi_n}{n} - \frac{a_1 + \ldots + a_n}{n}\right| < \varepsilon\right) \geqslant 1 - \frac{\boldsymbol{D}\left(\frac{\xi_1 + \ldots + \xi_n}{n}\right)}{\varepsilon^2}.$$

Отсюда, опираясь на теоремы XIV и XV, мы получаем

$$\boldsymbol{P}\left(\left|\frac{\xi_1 + \ldots + \xi_n}{n} - \frac{a_1 + \ldots + a_n}{n}\right| < \varepsilon\right) \geqslant 1 - \frac{b_1 + \ldots + b_n}{n^2\varepsilon^2}.$$

Поэтому, принимая во внимание условие 2° и переходя к пределу при $n \to \infty$, мы получаем

$$\lim_{n\to\infty} \boldsymbol{P}\left(\left|\frac{\xi_1+\ldots+\xi_n}{n}-\frac{a_1+\ldots+a_n}{n}\right|<\varepsilon\right)\geqslant 1,$$

откуда и вытекает требуемое равенство.

Теорема Чебышева вытекает отсюда автоматически как частный случай. Действительно, когда $\boldsymbol{E}\xi_k=a$ и $\boldsymbol{D}\xi_k<C$, то, очевидно, выполняются оба условия 1° и 2° теоремы Маркова.

§ 19. Усиленный закон больших чисел.

69. Пусть μ — число наступлений события A при n независимых испытаниях и пусть p — постоянная вероятность события A при отдельном испытании. Тогда по теореме Бернулли, какова бы ни была положительная постоянная ε, имеет место равенство

$$\lim_{n\to\infty} \boldsymbol{P}\left(\left|\frac{\mu}{n}-p\right|<\varepsilon\right)=1.$$

Это значит, что, предпринимая неограниченную серию испытаний, мы всегда будем иметь практическую уверенность в том, что рано или поздно отношение $\frac{\mu}{n}$ будет отличаться от p меньше, чем на любое наперед заданное положительное число.

Это не значит, однако, что отношение $\frac{\mu}{n}$ стремится к p:

$$\frac{\mu}{n}\to p.$$

Такое утверждение было бы просто неверным. В самом деле, хотя и с очень малой вероятностью, но может случиться, например, что событие A будет происходить при всех испытаниях, и тогда предел отношения $\frac{\mu}{n}$ будет равен единице; так же как может случиться, что событие A не будет происходить ни при одном испытании, и тогда этот предел будет равен нулю. И вообще, каково бы ни было число θ, где $0\leqslant\theta\leqslant 1$, ничто не гарантирует нас от такого случайного стечения обстоятельств, при котором отношение $\frac{\mu}{n}$ будет стремиться к θ, равно как, наконец, в силу случайного стечения обстоятельств это отношение может оказаться не стремящимся ни к какому пределу. Таким образом, если мы представим себе неограниченную серию испытаний, то здесь может случиться, что отношение $\frac{\mu}{n}$ будет стремиться к p, но этого может и не случиться.

Естественно возникает вопрос: какова же вероятность того, что в неограниченной серии испытаний отношение $\frac{\mu}{n}$ будет стремиться к p? Оказывается, как мы вскоре увидим, что эта вероятность равна единице. Отметим тут же, что это нисколько не противоречит невозможности утверждать с достоверностью, что отношение $\frac{\mu}{n}$ стремится к p. Действительно, хотя достоверное событие имеет вероятность, равную единице, но, как мы знаем, обратное может быть и неверным.

70. Разрешим прежде всего некоторые общие вопросы. Пусть дана бесконечная последовательность случайных величин $\xi_1, \ldots, \xi_n \ldots$ и постоянная a и пусть

$$\text{(A)} \qquad \boldsymbol{P}(\xi_n \to a) = 1.$$

Что в точности означает выражение (A)? Чтобы ответить на этот вопрос, достаточно вспомнить определение предела. Ответ, очевидно, можно формулировать следующим образом: выражение (A) означает, что равна единице вероятность того, что для всякого целого положительного числа r найдется целое положительное n, для которого при всех целых положительных k будет иметь место неравенство

$$|\xi_{n+k} - a| < \frac{1}{r}.$$

В дальнейшем часто будет удобнее рассматривать вместо выражения (A) равносильное ему выражение

$$\text{(B)} \qquad \boldsymbol{P}(\xi_n \not\to a) = 0.$$

Выражение (B) означает, что равна нулю вероятность того, что найдется целое положительное число r, для которого при любом целом положительном n и хотя бы при одном целом положительном k будет иметь место неравенство

$$|\xi_{n+k} - a| \geqslant \frac{1}{r}.$$

71. Докажем две леммы, которые лягут в основу всего дальнейшего.

Лемма I. — *Если событие E представляет собой сумму конечного или счетного множества событий $E_1, E_2, E_3, \ldots$:*

$$E = E_1 + E_2 + E_3 + \ldots,$$

то вероятность E не превосходит суммы вероятностей $E_1, E_2, E_3 \ldots$:

$$\boldsymbol{P}(E) \leqslant \boldsymbol{P}(E_1) + \boldsymbol{P}(E_2) + \boldsymbol{P}(E_3) + \ldots$$

Доказательство. — Событие E можно представить в виде суммы несовместимых событий. А именно, очевидно, что

$$E = E_1 + \overline{E}_1 E_2 + \overline{E}_1 \overline{E}_2 E_3 + \ldots$$

Отсюда по расширенной аксиоме сложения вероятностей

$$\boldsymbol{P}(E) = \boldsymbol{P}(E_1) + \boldsymbol{P}(\overline{E}_1 E_2) + \boldsymbol{P}(\overline{E}_1 \overline{E}_2 E_3) + \ldots$$

Далее, ясно, что

$$\overline{E}_1 E_2 \subset E_2, \quad \overline{E}_1 \overline{E}_2 E_3 \subset E_3, \ldots$$

Поэтому
$$P(E) \leqslant P(E_1) + P(E_2) + P(E_3) + \ldots,$$
что и требовалось доказать.

Лемма II. — *Равенство* (A) *или, что то же,* (B), *имеет место, если ряд*
$$\sum_{n=1}^{\infty} P\left(|\xi_n - a| \geqslant \frac{1}{r}\right)$$
сходится при всяком целом положительном значении r.

Доказательство. — Предположим, что условие леммы выполнено, и докажем равенство (B).

Обозначим через E_n^r событие, состоящее в том, что величина ξ_n принимает значение, удовлетворяющее неравенству
$$\left| \xi_n - a \right| \geqslant \frac{1}{r}.$$
Фиксируем числа r и n. Рассмотрим событие
$$S_n^r = E_{n+1}^r + E_{n+2}^r + E_{n+3}^r + \ldots,$$
состоящее в том, что для данных r и n хотя бы при одном $k = 1$, 2, 3, ..., величина ξ_{n+k} примет значение, удовлетворяющее неравенству
$$\left| \xi_{n+k} - a \right| \geqslant \frac{1}{r}.$$
В силу леммы I
$$P(S_n^r) \leqslant P(E_{n+1}^r) + P(E_{n+2}^r) + P(E_{n+3}^r) + \ldots$$
Правая часть этого последнего неравенства есть
$$\sum_{k=1}^{\infty} P\left(\left| \xi_{n+k} - a \right| \geqslant \frac{1}{r}\right)$$
и потому по условию представляет собой остаток сходящегося ряда. Следовательно, она стремится к нулю при $n \to \infty$, а потому и
$$\lim_{n \to \infty} P(S_n^r) = 0.$$

Далее, сохраняя фиксированное значение r, рассмотрим событие
$$S^r = S_1^r S_2^r S_3^r \ldots,$$
состоящее в том, что при данном r, для любого $n = 1, 2, 3, \ldots$ и хотя бы при одном $k = 1, 2, 3, \ldots$ величина ξ_{n+k} будет принимать значения, удовлетворяющие неравенству

$$\left| \xi_{n+k} - a \right| \geqslant \frac{1}{r}.$$

Ясно, что

$$S^r \subset S^r_n,$$

каково бы ни было n. Поэтому

$$\boldsymbol{P}(S^r) \leqslant \boldsymbol{P}(S^r_n),$$

каково бы ни было n. Но мы только что видели, что $\lim\limits_{n \to \infty} \boldsymbol{P}(S^r_n) = 0$. Следовательно, и

$$\boldsymbol{P}(S^r) = 0.$$

Это равенство имеет место при всяком r.

Наконец, рассмотрим событие

$$S = S^1 + S^2 + S^3 + \ldots,$$

состоящее в том, что хотя бы при одном $r = 1, 2, 3, \ldots$, для любого $n = 1, 2, 3, \ldots$, и хотя бы при одном $k = 1, 2, 3, \ldots$ величина ξ_{n+k} будет принимать значение, удовлетворяющее неравенству

$$\left| \xi_{n+k} - a \right| \geqslant \frac{1}{r}.$$

Другими словами, как мы выяснили выше, событие S состоит в том, что

$$\xi_n \not\to a.$$

В силу леммы I

$$\boldsymbol{P}(S) \leqslant \boldsymbol{P}(S^1) + \boldsymbol{P}(S^2) + \boldsymbol{P}(S^3) + \ldots$$

Но мы только что видели, что $\boldsymbol{P}(S^r) = 0$ при всяком r. Следовательно, и

$$\boldsymbol{P}(S) = 0.$$

Другими словами:

$$\boldsymbol{P}(\xi_n \not\to a) = 0,$$

что и требовалось доказать.

72. Теорема Бореля. — *Пусть μ — число наступлений события A при n независимых испытаниях и пусть p — постоянная вероятность события A при отдельном испытании. Тогда*

$$\boldsymbol{P}\left(\frac{\mu}{n} \to p\right) = 1.$$

Доказательство[1]. — Соотношение

$$\frac{\mu}{n} \to p,$$

---

[1]) Эту теорему впервые доказал E. Borel (1909) для случая $p = \frac{1}{2}$. Приводимое здесь доказательство следует методу F. P. Cantelli (1917), у которого данная теорема фигурирует как частный случай значительно более общего предложения.

очевидно, равносильно соотношению

$$\left(\frac{\mu}{n}-p\right)^4 \to 0.$$

Мы вычислим вероятность этого последнего.

Для этого мы представим, как обычно, величину μ в виде суммы n независимых случайных величин:

$$\mu = \mu_1 + \mu_2 + \ldots + \mu_n,$$

где каждое из слагаемых μ_i — число появлений события A при одном i-м испытании — может принимать значение 1 с вероятностью p или значение 0 с вероятностью $1-p$. Пользуясь этим представлением величины μ, оценим среднее значение величины $\left(\frac{\mu}{n}-p\right)^4$:

$$\text{(59)}\quad \boldsymbol{E}\left(\frac{\mu}{n}-p\right)^4 = \frac{1}{n^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\boldsymbol{E}(\mu_i-p)(\mu_j-p)(\mu_k-p)(\mu_l-p).$$

Все произведения:

$$\text{(60)}\qquad (\mu_i-p)(\mu_j-p)(\mu_k-p)(\mu_l-p)$$

можно разбить на три группы.

1) Произведения (60), в которых все четыре индекса i, j, k, l между собой равны. Эти произведения имеют вид

$$(\mu_i-p)^4.$$

Общее число их равно числу возможных значений индекса i, т. е. равно n. Среднее значение каждого из них, очевидно, положительно, но не превосходит единицы, поскольку ни одно из их значений не превосходит единицы. Таким образом сумма средних значений произведений (60) этой первой группы **не превосходит** n.

2) Произведения (60), в которых содержится по две пары равных индексов, а именно: либо $i=j$ и $k=l$, либо $i=k$ и $l=j$, либо $i=l$ и $j=k$. Эти произведения имеют вид

$$(\mu_i-p)^2(\mu_k-p)^2, \quad (\mu_i-p)^2(\mu_l-p)^2, \quad (\mu_i-p)^2(\mu_j-p)^2.$$

Общее число их, за вычетом тех, которые уже отнесены нами к первой группе, равно $3(n^2-n)$. Среднее значение каждого из них положительно, но не превосходит единицы. Таким образом сумма средних значений произведений (60) этой второй группы **не превосходит** $3(n^2-n)$.

3) Все остальные произведения (60). Для этих произведений характерно то, что в них по крайней мере один из индексов i, j, k, l не равен ни одному из остальных. Поэтому среднее значение каждого из них равно нулю. В самом деле, если, например, индекс i не равен ни одному из индексов j, k, l, то в силу независимости множителей произведения (60) и в силу того, что $\boldsymbol{E}\mu_i = p$, мы имеем:

$$\boldsymbol{E}(\mu_i-p)(\mu_j-p)(\mu_k-p)(\mu_l-p) =$$
$$= \boldsymbol{E}(\mu_i-p)\cdot\boldsymbol{E}(\mu_j-p)(\mu_k-p)(\mu_l-p) = 0.$$

Таким образом и сумма средних значений произведений этой третьей группы равна нулю.

Сопоставляя полученные нами заключения о трех группах произведений (60), мы видим, что сумма средних значений всех произведений (60) положительна, но не превосходит $n+3(n^2-n)$. Но эта сумма — не что иное, как четырехкратная сумма в правой части равенства (59). Поэтому из равенства (59) мы получаем следующее неравенство:

$$E\left(\frac{\mu}{n}-p\right)^4 \leqslant \frac{n+3(n^2-n)}{n^4} < \frac{3}{n^2}.$$

Теперь, в силу первой формы неравенства Чебышева, при $r>0$

$$P\left[\left(\frac{\mu}{n}-p\right)^4 \geqslant \frac{1}{r}\right] = P\left[r\left(\frac{\mu}{n}-p\right)^4 \geqslant 1\right] \leqslant E\left[r\left(\frac{\mu}{n}-p\right)^4\right] = \\ = rE\left(\frac{\mu}{n}-p\right)^4,$$

и следовательно,

$$P\left[\left(\frac{\mu}{n}-p\right)^4 \geqslant \frac{1}{r}\right] < \frac{3r}{n^2}.$$

Отсюда ясно, что ряд

$$\sum_{n=1}^{\infty} P\left[\left(\frac{\mu}{n}-p\right)^4 \geqslant \frac{1}{r}\right],$$

сходится при всяком положительном значении r, и следовательно, в силу леммы II

$$P\left[\left(\frac{\mu}{n}-p\right)^4 \to 0\right] = 1.$$

Теорема доказана.

Подобно тому как теорема Бéрнулли является частным случаем теорем Чебышева и Маркова, теорема Бореля является частным случаем других теорем того же типа: Кантелли, А. Я. Хинчина и А. Н. Колмогорова. Теоремы этого типа носят общее имя *усиленного закона больших чисел*.

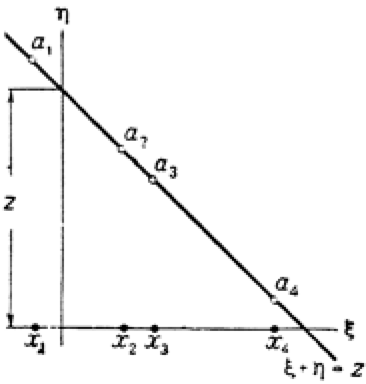
8 Зак. 1725. — В. И. Гливенко.

ГЛАВА VI.

ОДНОРОДНЫЕ СЛУЧАЙНЫЕ ПРОЦЕССЫ.

§ 20. Композиция законов распределения.

73. Всюду в дальнейшем нас будут интересовать, с той или с другой стороны, суммы независимых случайных величин. При этом нам придется не только составлять числовые характеристики сумм по данным числовым характеристикам слагаемых — чему нас учат теоремы XI и XV, — но и определять законы распределения сумм по данным законам распределения слагаемых.



Черт. 10.

Определение закона распределения суммы по законам распределения независимых слагаемых называется *композицией* законов распределения слагаемых. Оно осуществляется следующим образом.

Пусть сначала ξ и η — две д и с к р е т н ы е независимые случайные величины с законами распределения $P_\xi(x)$ и $P_\eta(y)$ и пусть ζ есть сумма этих величин; она, очевидно, также — дискретная случайная величина, имеющая некоторый закон распределения $P_\zeta(z)$. Для наглядности изобразим пару чисел (ξ, η) точкой плоскости с декартовыми координатами ξ и η. Возьмем какое-нибудь значение z величины ζ. Так как $\zeta=\xi+\eta$, то при взятом нами значении z величины ζ пара чисел (ξ, η) будет изображаться точками, лежащими на прямой $\xi+\eta=z$ (см. черт. 10). Это — точки $a=a_1, a_2, \ldots$ с координатами ξ и η, равными соответственно

$$x \quad \text{и} \quad z-x,$$

где $x=x_1, x_2, \ldots$ — возможные при $\zeta=z$ значения величины ξ. Обозначим множество всех таких точек a через $\mathfrak{A}$. Вероятность того, что точка, изображающая пару чисел (ξ, η), будет принадлежать множеству $\mathfrak{A}$, очевидно, складывается из вероятностей того, что эта точка окажется точкой a_1, точкой a_2 и т. д. Но вероятность того, что точка (ξ, η) будет принадлежать множеству $\mathfrak{A}$, есть не что иное, как

(61) $$P_\zeta(z).$$

Вероятности же того, что точка (ξ, η) окажется точкой a_1, a_2 и т. д., в силу независимости величин ξ и η соответственно равны

(62) $$P_\xi(x_1)\,P_\eta(z-x_1),\ P_\xi(x_2)\,P_\eta(z-x_2),\ \ldots$$

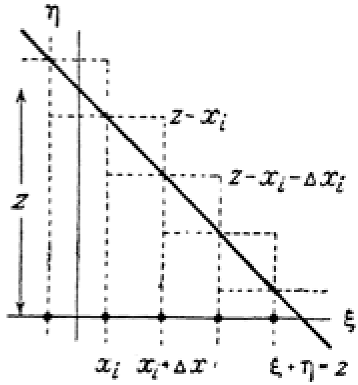
Таким образом выражение (61) складывается из выражений (62), т. е.

(63) $$P_\zeta(z)=\sum_x P_\xi(x)\,P_\eta(z-x),$$

где сумма распространена на все возможные при $\zeta=z$ значения x величины ξ. Аналогично можно было бы показать, что

(63 bis) $$P_\xi(z)=\sum_y P_\xi(z-y)\,P_\eta(y),$$

где сумма распространена на все возможные при $\zeta=z$ значения y величины η. Любая из формул (63) или (63 bis) и решает вопрос о композиции законов распределения в случае дискретных случайных величин.



Черт. 11.

74. Пусть теперь ξ и η — две произвольные независимые случайные величины с функциями распределения $F_\xi(x)$ и $F_\eta(y)$ и пусть ζ есть сумма этих величин и $F_\zeta(z)$ — ее функция распределения. Будем снова изображать пары чисел (ξ, η) точками плоскости с декартовыми координатами ξ и η. Возьмем какое-нибудь число z и будем рассматривать полупрямую $-\infty<\zeta<z$. Так как $\zeta=\xi+\eta$, то соответствующая этой полупрямой пара чисел (ξ, η) будет изображаться точками, лежащими в полуплоскости $\mathfrak{B}$, которая лежит ниже прямой $\xi+\eta=z$ (см. черт. 11). Заметив это, сделаем следующее построение. Возьмем ось ξ и подразделим ее на счетное множество конечных интервалов вида $x_i\leqslant\xi<x_i+\Delta x_i$. Тогда полуплоскость $\mathfrak{B}$ будет заключать в себе область $\mathfrak{B}'$ и будет сама заключена в области $\mathfrak{B}''$, которые определяются следующим образом.

Область $\mathfrak{B}'$ — это соединение прямоугольных полос b_i' со сторонами, параллельными осям координат и имеющими своими проекциями на этих осях интервалы

$$x_i\leqslant\xi<x_i+\Delta x_i \quad \text{и} \quad -\infty<\eta<z-x_i-\Delta x_i.$$

Область же $\mathfrak{B}''$ — это соединение прямоугольных полос b_i'' со сторонами, параллельными осям координат и имеющими своими проекциями на них интервалы

$$x_i\leqslant\xi<x_i+\Delta x_i \quad \text{и} \quad -\infty<\eta<z-x_i.$$

Вероятность $P(\mathfrak{B})$ того, что точка, изображающая пару чисел (ξ, η), попадет в полуплоскость $\mathfrak{B}$, очевидно, не меньше вероятности $P(\mathfrak{B}')$ того, что эта точка попадет в область $\mathfrak{B}'$ и не больше вероятности $P(\mathfrak{B}'')$ того, что она попадет в область $\mathfrak{B}''$:

$$P(\mathfrak{B}') \leqslant P(\mathfrak{B}) \leqslant P(\mathfrak{B}'').$$

Что же касается вероятностей $P(\mathfrak{B}')$ и $P(\mathfrak{B}'')$, то первая из них складывается из вероятностей $P(\mathfrak{b}_i')$ попадания в полосы $\mathfrak{b}_i'$, а вторая складывается из вероятностей $P(\mathfrak{b}_i'')$ попадания в полосы $\mathfrak{b}_i''$, так что

$$\sum_i P(\mathfrak{b}_i') \leqslant P(\mathfrak{B}) \leqslant \sum_i P(\mathfrak{b}_i'').$$

Но

$$P(\mathfrak{B}) = F_\zeta(z),$$

а в силу независимости величин ξ и η

$$P(\mathfrak{b}_i') = [F_\xi(x_i + \Delta x_i) - F_\xi(x_i)]\, F_\eta(z - x_i - \Delta x_i)$$

и

$$P(\mathfrak{b}_i'') = [F_\xi(x_i + \Delta x_i) - F_\xi(x_i)]\, F_\eta(z - x_i).$$

Следовательно:

$$(64)\quad \sum_i F_\eta(z - x_i - \Delta x_i)\,[F_\xi(x_i + \Delta x_i) - F_\xi(x_i)] \leqslant F_\zeta(z) \leqslant \\ \leqslant \sum_i F_\eta(z - x_i)\,[F_\xi(x_i + \Delta x_i) - F_\xi(x_i)].$$

При этом $F_\zeta(z)$ не зависит от Δx_i. Поэтому, если изменять подразделение оси ξ на интервалы так, чтобы максимальная из длин Δx_i этих интервалов стремилась к нулю, то неравенства (64) сохранятся и в пределе. Но суммы, стоящие в левой и в правой частях неравенств (64), будут иметь один и тот же предел [1]):

$$(65)\qquad F_\zeta(z) = \int_{-\infty}^{+\infty} F_\eta(z - x)\, dF_\xi(x).$$

Аналогично можно было бы показать, что

$$(65\text{ bis})\qquad F_\zeta(z) = \int_{-\infty}^{+\infty} F_\xi(z - y)\, dF_\eta(y).$$

Любая из формул (65) или (65 bis) и решает вопрос о композиции законов распределения для произвольных случайных величин.

---

[1]) Строго говоря, предел (65) имеет вторая из сумм (64), первая же сумма должна сходиться не к (65), а к выражению, получающемуся из (65) подстановкой $z - x - 0$ на место $z - x$, т. е. к интегралу функции $F_\eta(z - x - 0)$, а не функции $F_\eta(z - x)$. Но, в силу непрерывности слева функций распределения, эти последние всегда равны между собой.

75. Для непрерывных случайных величин ξ, η, ζ, с плотностями вероятности $p_\xi(x)$, $p_\eta(y)$, $p_\zeta(z)$ можно преобразовать формулы (65) и (65 bis), во-первых, заменив диференциалы $dF_\xi(x)$ и $dF_\eta(y)$ через $p_\xi(x)\,dx$ и $p_\eta(y)\,dy$ и, во-вторых, продиференцировав обе части равенств по z. Мы получим:

$$p_\zeta(z)=\int_{-\infty}^{+\infty} p_\eta(z-x)\,p_\xi(x)\,dx \tag{66}$$

и

$$p_\zeta(z)=\int_{-\infty}^{+\infty} p_\xi(z-y)\,p_\eta(y)\,dy. \tag{66 bis}$$

§ 21. Закон Пуассона.

76. Будем наблюдать событие A, повторяющееся в различные, случайно расположенные моменты времени, таким образом, что:

1° Для любых двух промежутков времени, не перекрывающихся между собой, наступление события A в одном из них не зависит от его наступления в другом.

2° Для любого промежутка времени вероятность наступления события A в этом промежутке зависит только от продолжительности этого промежутка.

3° Если обозначить через $\varphi(t)$ вероятность того, что в промежутке времени продожительности t событие A наступит по крайней мере один раз, то

$$\varphi(t)=at+\varepsilon t,$$

где a — положительное число, не зависящее от t, а ε стремится к нулю, когда t стремится к нулю.

4° Если обозначить через $\psi(t)$ вероятность того, что в промежутке времени продолжительности t событие A наступит по крайней мере два раза, то

$$\psi(t)=\delta t,$$

где δ стремится к нулю, когда t стремится к нулю.

Поставим задачу определить вероятность $p_m(t)$ того, что в промежутке времени продолжительности t событие A наступит ровно m раз.

Прежде чем решать эту задачу, покажем на двух примерах, какой реальный смысл могут иметь условия 1°—4° этой задачи.

Пример 1. — *Радиоактивность.* Как известно, радиоактивность состоит в том, что атомы химического элемента распадаются, теряя частицы своей материи, и в результате этого превращаются в атомы другого элемента. Каждый распад атома совершается мгновенно, наподобие взрыва. Будем наблюдать распады атомов в некоторой массе элемента. Фактически производившиеся наблюдения показывают, что распады атомов происходят совершенно случайно и независимо друг от друга, так что здесь ничто не противоречит допущению 1°. Допущение 2° несколько упрощает действительную картину явления, так как по мере уменьшения массы элемента вследствие радиоактивности сама радиоактивность постепенно ослабевает. Но для сравнительно небольших промежутков времени это допущение вполне законно. Допущение 3° здесь естественно потому, что если мы возьмем два промежутка времени, один, скажем, вдвое продолжительнее другого, то вероятность хотя бы одного распада атома в первом промежутке будет, очевидно, приблизительно вдвое больше, чем во втором. Мы говорим „приблизительно вдвое", так как мы не можем просто применить здесь теорему сложения вероятностей, поскольку распады атомов в различные промежутки времени отнюдь не являются несовместимыми событиями. Но если бы существовала т о ч н а я пропорциональность между продолжительностью промежутка времени и вероятностью хотя бы одного распада атома в этом промежутке, то мы имели бы равенство $\varphi(t) = at$, где a — положительное число, не зависящее от t. Если же, как мы допускаем, существует п р и б л и з и т е л ь н а я пропорциональность, то мы должны положить $\varphi(t) = (a + \varepsilon)t$, где $\varepsilon \to 0$ при $t \to 0$. Наконец, допущение 4° здесь также естественно, потому что при малом t вероятность $\psi(t)$ по крайней мере двух распадов атомов будет, очевидно, приблизительно квадратом вероятности по крайней мере одного распада, и во всяком случае будет бесконечно малой высшего порядка, чем вероятность одного распада, а эта последняя, как мы только что видели, одного порядка с t; таким образом отношение $\delta = \frac{\psi(t)}{t}$ будет стремиться к нулю вместе с t.

П р и м е р 2. — *Работа телефонной станции.* На телефонной станции перед каждой телефонисткой находится щит, на котором по временам появляется сигнал — вызов станции тем или другим абонентом. Возьмем один такой щит и будем наблюдать появление сигналов на нем. Ясно, что появление сигналов, как и распады атомов радиоактивного элемента, происходит случайно и независимо друг от друга. Допущения 1°, 3° и 4° обосновываются здесь точно такими же соображениями, как и в случае радиоактивности. В ином положении находится только допущение 2°. Вероятность вызова станции в некотором промежутке времени зависит не только от продолжительности этого промежутка, но и от его расположения на шкале времени, поскольку эта вероятность, очевидно, различна в разное время дня. Но если интересоваться только вызовами станции в определенное время дня, скажем в рабочие часы, то вероятность вызовов в том или другом промежутке в пределах этого времени дня уже может зависеть только от продолжительности данного промежутка. Здесь допущение 2° будет вполне законно.

Перейдем к решению поставленной выше общей задачи определения вероятности $p_m(t)$ того, что в промежутке времени продолжительности t событие A наступит ровно m раз. Мы решим эту задачу сначала для $m = 0$ и затем для любого $m > 0$.

Метод решения во всех случаях будет заключаться в том, что мы будем брать два смежных промежутка времени, а именно промежутки $(0, t)$ и $(t, t + \Delta t)$, а также образованный их соединением промежуток

$(0,\ t+\Delta t)$, и будем сравнивать между собой вероятности появлений события A в этих трех промежутках. Для случая $m=0$, т. е. для ненаступления события A, мы имеем, принимая во внимание независимость наступлений, а следовательно, и ненаступлений события A в промежутках времени $(0,\ t)$ и $(t,\ t+\Delta t)$:

$$p_0(t+\Delta t)=p_0(t)\,p_0(\Delta t).$$

Отсюда

$$p_0(t+\Delta t)-p_0(t)=p_0(t)\,[1-\varphi(\Delta t)]-p_0(t)=\\ =-\varphi(\Delta t)\,p_0(t)=-ap_0(t)\,\Delta t-\varepsilon p_0(t)\,\Delta t.$$

Следовательно, существует производная

$$\frac{dp_0(t)}{dt}=\lim_{\Delta t\to 0}\frac{p_0(t+\Delta t)-p_0(t)}{\Delta t}=\lim_{\Delta t\to 0}\,[-ap_0(t)-\varepsilon p_0(t)]=-ap_0(t).$$

Мы получили, таким образом, для определения вероятности $p_0(t)$ дифференциальное уравнение

$$\frac{dp_0(t)}{dt}=-ap_0(t).$$

Его решение есть $p_0(t)=Ce^{-at}$. Постоянную C можно исключить, заметив, что, очевидно, $p_0(0)=1$. Поэтому $C=1$, так что окончательно

$$p_0(t)=e^{-at}.$$

Пусть теперь $m>0$. Число m наступлений события A в промежутке времени $(0,\ t+\Delta t)$ складывается из числа n его наступлений в промежутке $(t,\ t+\Delta t)$ и из числа $m-n$ его наступлений в промежутке $(0,\ t)$. Поэтому здесь можно применить формулу композиции (63) п. 73, и мы получим, принимая во внимание, что здесь возможными значениями n являются $0,\ 1,\ \ldots,\ m$:

$$p_m(t+\Delta t)=\sum_{n=0}^{m}p_{m-n}(t)\,p_n(\Delta t).$$

Чтобы решить это уравнение, напишем его в таком виде:

$$p_m(t+\Delta t)=p_m(t)\,p_0(\Delta t)+p_{m-1}(t)\,p_1(\Delta t)+r,$$

где через r мы обозначаем сумму всех остальных $m-1$ слагаемых. Отсюда

$$p_m(t+\Delta t)-p_m(t)=\\ =p_m(t)\,[1-\varphi(\Delta t)]-p_m(t)+p_{m-1}(t)\,[\varphi(\Delta t)-\psi(\Delta t)]+r=\\ =-ap_m(t)\,\Delta t-\varepsilon p_m(t)\,\Delta t+ap_m(t)\,\Delta t+(\varepsilon-\delta)\,p_{m-1}(t)\,\Delta t+r.$$

Следовательно, существует производная

$$\frac{dp_m(t)}{dt}=\lim_{\Delta t\to 0}\frac{p_m(t+\Delta t)-p_m(t)}{\Delta t}=\\ =\lim_{\Delta t\to 0}\left[-ap_m(t)-\varepsilon p_m(t)+ap_{m-1}(t)+(\varepsilon-\delta)\,p_{m-1}(t)+\frac{r}{\Delta t}\right]=\\ =-ap_m(t)+ap_{m-1}(t),$$

ибо $\lim\limits_{\Delta t \to 0} \frac{r}{\Delta t} = 0$. Действительно,

$$r = p_{m-2}(t)\, p_2(\Delta t) + \ldots + p_0(t)\, p_m(\Delta t) \leqslant$$
$$\leqslant p_2(\Delta t) + \ldots + p_m(\Delta t) \leqslant \psi(\Delta t),$$

так что

$$\frac{r}{\Delta t} \leqslant \frac{\psi(\Delta t)}{\Delta t} = \delta.$$

Мы получили, таким образом, для определения вероятности $p_m(t)$ разностно-диференциальное уравнение

$$\frac{dp_m(t)}{dt} = -ap_m(t) + ap_{m-1}(t).$$

Оно проще всего решается следующим способом. Сделаем подстановку

$$p_m(t) = \overset{*}{p}_m(t)\, e^{-at}.$$

Тогда наше уравнение превратится в уравнение

$$\frac{d\overset{*}{p}_m(t)}{dt} = a\overset{*}{p}_{m-1}(t).$$

Отсюда

$$\overset{*}{p}_m(t) = a \int\limits_0^t \overset{*}{p}_{m-1}(u)\, du + C^*.$$

Однако при $m > 0$, очевидно, $\overset{*}{p}_m(0) = p_m(0) = 0$. Поэтому и $C^* = 0$. С другой же стороны, воспользовавшись уже полученным выражением для $p_0(t)$, мы получаем $\overset{*}{p}_0(t) = p_0(t)\, e^{at} = 1$. Поэтому мы имеем последовательно

$$\overset{*}{p}_1(t) = at,$$
$$\overset{*}{p}_2(t) = a^2\, \frac{t^2}{2},$$
$$\ldots\ldots\ldots$$

и вообще

$$\overset{*}{p}_m(t) = a^m\, \frac{t^m}{m!}.$$

Переходя от $\overset{*}{p}_m(t)$ к $p_m(t)$, мы получаем искомую вероятность

$$p_m(t) = \frac{(at)^m}{m!}\, e^{-at}.$$

Это — закон Пуассона:

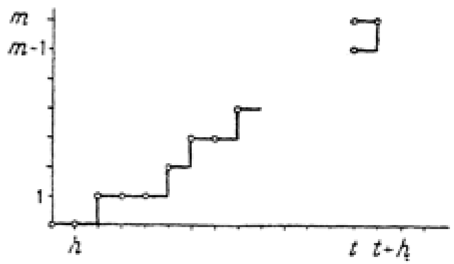
$$\frac{k^m}{m!}\, e^{-k}.$$

В данном случае $k = at$.

Если применить этот вывод хотя бы ко второму из примеров, который мы привели выше — к работе телефонной станции, то мы получим следующее: вероятность того, что за промежуток времени продолжительности t число вызовов станции будет m, равна только что полученной величине $p_m(t)$, т. е. определяется по закону Пуассона.

Если еще скомбинировать этот вывод с законом больших чисел, то мы будем иметь такой практический результат. Будем рассматривать не один промежуток времени данной продолжительности t, а много таких промежутков, скажем для определенности — большое число промежутков времени, начинающихся и оканчивающихся в одни и те же часы, но в разные дни. В силу теоремы Бéрнулли для большого числа дней мы будем иметь практическую уверенность в том, что относительное число дней, в которые в рассматриваемом промежутке времени будет наблюдаться данное число вызовов m, будет как угодно мало отличаться от найденной нами вероятности

$$\frac{(at)^m}{m!}e^{-at}.$$



Черт. 12а.

77. Появление закона Пуассона в качестве решения поставленной нами задачи и появление того же закона Пуассона в качестве асимптотической формулы в п° 43 — ни в коей мере не случайное совпадение. Можно очень наглядно показать связь между тем и другим, если обратиться к следующей схеме.

Возьмем прямую $-\infty < x < +\infty$ и представим себе, что по ней передвигается точка P, начинающая свое движение в 0 и следующим образом его продолжающая. В течение некоторого промежутка времени h она остается неподвижной; затем либо делает моментальный скачок на единицу вверх (см. черт. 12а), либо продолжает оставаться неподвижной до истечения времени $2h$; затем опять либо делает моментальный скачок на единицу вверх, либо продолжает оставаться неподвижной до истечения времени $3h$ и т. д. График движения точки P может оказаться тем или иным в зависимости от случая, но он всегда будет состоять из горизонтальных отрезков длины, кратной h, и из вертикальных отрезков длины 1, как это изображено на черт. 12а. Пусть теперь вероятность того, что точка P сделает скачок, в любой из моментов времени, кратный h, равна одному и тому же числу $p = ah$, и следовательно, вероятность того, что она останется неподвижной, равна $q = 1 - ah$, и то и другое — независимо от того, на какой высоте точка уже находится. Вычислим вероятность $f(m, n)$ того, что в момент $t = nh$ окончательное положение точки P будет на высоте m. Для этого установим связь между последовательными положениями точки в моменты $t = nh$ и $t + h = (n + 1)h$. Ясно, например, что в момент $t + h$ окончательное положение точки P может оказаться на высоте m только в двух случаях: либо 1) если окончательной ее высотой в момент t была высота m и в момент $t + h$ эта высота не изменилась, потому что в этот последний момент точка P осталась неподвижной; либо 2) если окончательной ее высотой в момент t была высота $m - 1$, но в момент $t + h$ точка P сделала скачок вверх. Вероятность неподвижного положения точки P в случае 1) по условию равна q, вероятность ее скачка в случае 2) равна p. Отсюда, пользуясь формулой полной вероятности, мы получаем

$$(\alpha) \qquad f(m, n + 1) = qf(m, n) + pf(m - 1, n).$$

Кроме того, нетрудно подсчитать, что всегда

$$(\beta) \qquad f(0, n) = q^n \quad \text{и} \quad f(m, m) = p^m \; (m > 0).$$

Наконец, по самому смыслу задачи достаточно определить функцию $f(m, n)$ только для пар целых неотрицательных значений m и n, удовлетворяющих неравенству $m \leqslant n$, ибо в течение времени $t = nh$ точка P может сделать не больше, чем n скачков. Можно проверить простым вычислением, что функция

$$f(m, n) = \frac{n!}{m!(n-m)!} p^m q^{n-m}$$

удовлетворяет уравнению (α) и начальным условиям (β) и определена в области $m \leqslant n$. Следовательно, она является одним из решений нашей задачи. Но нетрудно сообразить сверх того, что уравнение (α) при начальных условиях (β) может иметь в области $m \leqslant n$ только одно решение. Следовательно, указанное нами решение окончательное[1]).

Зафиксируем теперь момент t, но будем безгранично уменьшать промежуток времени h и одновременно безгранично увеличивать число n таких промежутков времени — так, чтобы оставалось в силе равенство $t = nh$. Соответственно будем разрешать точке P делать скачки по истечении каждого промежутка времени, кратного h; но пусть вероятность каждого скачка попрежнему равна $p = ah$, где a — то же самое, что было вначале, и следовательно, p уменьшается пропорционально h; это значит, что в среднем скачки останутся столь же редкими, как были вначале. Посмотрим, во что превратятся в пределе уравнения (α) начальные условия (β) и найденное нами решение при $h \to 0$, или, что то же, при $n \to \infty$. Для этого положим

$$p_m(t) = f(m, n),$$

где $t = nh$, и воспользуемся равенством $p = ah$. Тогда уравнение (α) примет вид

$$p_m(t+h) = (1-ah)\,p_m(t) + ah\,p_{m-1}(t),$$

или, что то же:

$$\frac{p_m(t+h) - p_m(t)}{h} = -ap_m(t) + ap_{m-1}(t).$$

При $h \to 0$ это дает

$$(\alpha^*) \qquad \frac{dp_m(t)}{dt} = -ap_m(t) + ap_{m-1}(t),$$

т. е. найденное нами выше разностно-диференциальное уравнение закона Пуассона. Начальные условия (β) примут вид

$$p_0(t) = (1-ah)^{\frac{t}{h}} \quad \text{и} \quad p_m(mh) = (ah)^m \ (m > 0).$$

При $h \to 0$ это дает

$$(\beta^*) \qquad p_0(t) = e^{-at} \quad \text{и} \quad p_m(0) = 0 \qquad (m > 0),$$

т. е. те же условия, при которых мы решали выше уравнение (α^*). Что же касается самого решения уравнения (α) при начальных условиях (β), то оно принимает вид

$$\frac{n!}{m!(n-m)!}\left(\frac{at}{n}\right)^m \left(1-\frac{at}{n}\right)^{n-m}.$$

По теореме Пуассона при $n \to \infty$ это дает закон Пуассона в виде

$$p_m(t) = \frac{(at)^m}{m!} e^{-at},$$

т. е. как раз решение уравнения (α^*) при условии (β^*).

Пусть теперь точка P может делать скачки на единицу вверх в **любой** момент времени, но пусть в среднем эти скачки будут столь же редкими,

---

[1]) Его, разумеется, можно было бы вывести и иначе, переведя описанную картину на язык независимых испытаний.

как мы полагали до сих пор. Мы окажемся в условиях 1°—4°, которые были приведены в начале п. 76; событием A будет здесь скачок точки P. Следовательно, здесь уравнение (α^*) будет уравнением для вероятности $p_m(t)$, то же относится и к начальным условиям (β^*). И поэтому закон Пуассона будет точным, а не только асимптотическим решением задачи о вероятности $p_m(t)$.

Но предположить, что точка P может делать скачки в любой момент времени, это значит предположить, что нижняя грань промежутка времени между двумя последовательными скачками, — которую мы обозначали выше через h, — здесь равна 0. Таким образом связь между предельным и точным законом Пуассона можно кратко охарактеризовать так: закон Пуассона одинаково получается как в пределе при $h \to 0$, так и при $h = 0$.

§ 22. Закон Гаусса.

78. Будем наблюдать величину, которая изменяется во времени, получая в различные промежутки времени случайные приращения ξ такие, что:

1° Для любых двух промежутков времени, не перекрывающихся между собой, приращение ξ в одном из них не зависит от приращения в другом,

2° Для любого промежутка времени вероятность того или иного приращения ξ в этом промежутке зависит только от продолжительности этого промежутка.

3° Законы распределения приращения ξ в различных промежутках времени мало отличаются один от другого, если сами промежутки времени мало различаются по продолжительности. Если обозначить через $F(x, t)$ функцию распределения приращения ξ в промежутке времени данной продолжительности $t > 0$, то существует плотность вероятности

$$p(x, t) = \frac{\partial F(x, t)}{\partial x},$$

в свою очередь имеющая частные производные по x до третьего порядка включительно, ограниченные при каждом данном $t > 0$.

4° В малых промежутках времени малые приращения ξ гораздо более вероятны, нежели большие приращения. Существуют определенные пределы:

$$(\mathrm{J}_1) \qquad \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{-\infty}^{+\infty} x p(x, \Delta t)\, dx = A,$$

$$(\mathrm{J}_2) \qquad \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{-\infty}^{+\infty} x^2 p(x, \Delta t)\, dx = B > 0,$$

$$(\mathrm{J}_3) \qquad \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{-\infty}^{+\infty} |x|^3 p(x, \Delta t)\, dx = 0.$$

Пояснение. — Качественной стороне этого условия 4° — тому, что в малых промежутках времени продолжительности Δt малые приращения ξ гораздо более вероятны, нежели большие приращения, —

соответствует то, что предел (J_2) больше нуля, тогда как предел (J_3) равен нулю. В самом деле, при малых x выражение $|x|^3$ меньше выражения x^2, при больших x — наоборот. Поэтому, если при малых Δt интеграл (J_3) оказывается меньше интеграла (J_2), то причиной этого может быть только то, что под знаком интеграла решающую роль играют малые значения x. А это может происходить только оттого, что значения диференциала $p(x, \Delta t)\,dx$ особенно велики при малых значениях x. Но диференциал $p(x, \Delta t)\,dx$ есть „вероятность того, что приращение ξ примет значение между x и $x + dx$". Таким образом действительно, при малых Δt вероятности малых приращений ξ значительно превосходят вероятности больших приращений.

Поставим задачу определить плотность вероятности $p(x, t)$.

Предварительно, однако, выясним, какой реальный смысл могут иметь условия 1°—4°. Оговоримся тут же, что в примерах, которые мы для этого приведем, непосредственно ясна будет только качественная сторона условий 3° и 4°. Уточнения количественного характера, которые мы ввели в формулировки этих условий, необходимы нам как исходные точки для дальнейшей математической дедукции. Но в наших примерах эти уточнения могут быть приняты только как рабочие гипотезы, которые оправдываются согласием окончательных выводов теории с данными опыта.

Пример 1. — *Диффузия газов.* Возьмем молекулу некоторого газа, свободно движущуюся среди других молекул того же или другого газа постоянной плотности и постоянной температуры. Отнесем все пространство к декартовым осям координат и будем следить, как изменяется с течением времени одна из координат данной молекулы, скажем координата x. При своем движении данная молекула будет подвергаться случайным столкновениям с другими молекулами. Вследствие этих столкновений координата x будет изменяться во времени, получая в различные промежутки времени случайные приращения ξ, причем все наши условия 1°—4° для ξ, очевидно, будут соблюдены (для условий 3° и 4° во всяком случае очевидна их качественная сторона).

Пример 2. — *Скорости молекул.* Возьмем снова молекулу некоторого газа, свободно движущуюся среди других молекул того же или другого газа постоянной плотности и постоянной температуры. Отнесем снова все пространство к декартовым осям координат и будем следить, как изменяется с течением времени компонента скорости данной молекулы по одной из координат; пусть эта компонента скорости будет x. Вследствие случайных столкновений данной молекулы с другими молекулами эта компонента скорости x будет изменяться во времени, получая в различные промежутки времени случайные приращения ξ. Очевидно, что для ξ все наши условия 1°—4° будут соблюдены (для условий 3° и 4° опять-таки очевидна их качественная сторона).

Перейдем к решению поставленной выше общей задачи определения плотности вероятности $p(x, t)$. Для этого возьмем два смежных промежутка времени, $(0, t)$ и $(t, t + \Delta t)$, а также образованный их соединением промежуток $(0, t+\Delta t)$, и сравним между собой вероятности приращений ξ в этих промежутках. Приращение в промежутке $(0, t+\Delta t)$ складывается из приращений в промежутках $(t, t+\Delta t)$ и $(0, t)$. Поэтому здесь можно применить формулу композиции (66) п. 74, и мы получим

$$p(x, t+\Delta t)=\int_{-\infty}^{+\infty} p(x-s, t)\, p(s, \Delta t)\, ds. \tag{67}$$

Мы имеем, таким образом, интегральное уравнение для определения $p(x, t)$.

Мы не будем непосредственно решать это уравнение, а выведем из него диференциальное уравнение, решение которого известно. С этой целью мы разложим первый из множителей под знаком интеграла (67), $p(x-s, t)$, по степеням $-s$ следующим образом:

$$p(x-s, t)=p(x, t)-\frac{\partial p}{\partial x}s+\frac{1}{2}\frac{\partial^2 p}{\partial x^2}s^2-\frac{1}{6}\left[\frac{\partial^3 p}{\partial x^3}\right]_\sigma s^3;$$

ненаписанными аргументами всегда являются x и t, а σ лежит между $x-s$ и x. Подставив это в (65), мы получим

$$p(x, t+\Delta t)=p(x, t)\int_{-\infty}^{+\infty} p(s, \Delta t)\, ds-\frac{\partial p}{\partial x}\int_{-\infty}^{+\infty} s p(s, \Delta t)\, ds+ \tag{68}$$
$$+\frac{1}{2}\frac{\partial^2 p}{\partial x_2}\int_{-\infty}^{+\infty} s^2 p(s, \Delta t)\, ds-\frac{1}{6}\int_{-\infty}^{+\infty}\left[\frac{\partial^3 p}{\partial x^3}\right]_\sigma s^3 p(s, \Delta t)\, ds.$$

Рассмотрим последовательно все четыре интеграла, фигурирующие в этом равенстве. Первый из этих интегралов, как мы знаем, равен единице. Что касается следующих двух интегралов, то следует обратиться к условию 4°, где мы имели пределы отношений этих интегралов к Δt при $\Delta t \to 0$. А именно, в силу (J_1) и (J_2) эти пределы соответственно равны A и B. Наконец, для последнего интеграла следует опять-таки обратиться к условию 4°. Действительно, этот интеграл по абсолютной величине не превосходит

$$T(t)\int_{-\infty}^{+\infty} |s|^3 p(s, \Delta t)\, ds,$$

где $T(t)$ есть верхняя грань функции $\left|\frac{\partial^3 p}{\partial x^3}\right|$ при данном t; следовательно, в силу (J_3) предел отношения этого интеграла к Δt при $\Delta t \to 0$ равен нулю. Установив все это, преобразуем равенство (66) следующим образом:

$$\frac{p(x, t+\Delta t)-p(x, t)}{\Delta t}=\frac{\partial p}{\partial x}\frac{1}{\Delta t}\int_{-\infty}^{+\infty} s p(s, \Delta t)\, ds+$$
$$+\frac{1}{2}\frac{\partial^2 p}{\partial x^2}\frac{1}{\Delta t}\int_{-\infty}^{+\infty} s^2 p(s, \Delta t)\, ds-\frac{1}{6}\frac{1}{\Delta t}\int_{-\infty}^{+\infty}\left[\frac{\partial^3 p}{\partial x^3}\right]_\sigma s^3 p(s, \Delta t)\, ds.$$

Переходя к пределу при $\Delta t \to 0$, мы видим, что существует производная

$$\frac{\partial p}{\partial t} = -A\frac{\partial p}{\partial x} + \frac{1}{2}B\frac{\partial^2 p}{\partial x^2}.$$

Мы имеем, таким образом, уравнение параболического типа для определения $p(x, t)$.

Это уравнение можно решить подстановкой:

$$p(x, t) = p^*(x^*, t^*),$$

где

$$x^* = x - At, \quad t^* = Bt.$$

Последняя переводит его в уравнение

$$\frac{\partial p^*}{\partial t^*} = \frac{1}{2}\frac{\partial^2 p^*}{\partial x^{*2}}.$$

Это — каноническое уравнение теплопроводности. Его решение, удовлетворяющее условию

$$\int_{-\infty}^{+\infty} p^*(x^*, t^*)\,dx^* = 1,$$

дается, как известно, формулой

$$p^*(x^*, t^*) = \frac{1}{\sqrt{2\pi t^*}}e^{-\frac{x^{*2}}{2t^*}}.$$

Возвращаясь к прежним переменным, мы получаем отсюда

$$p(x, t) = \frac{1}{\sqrt{2\pi Bt}}e^{-\frac{(x-At)^2}{2Bt}}.$$

Это — закон Гаусса:

$$\frac{1}{\sqrt{2\pi\beta}}e^{-\frac{(x-\alpha)^2}{2\beta}}.$$

В данном случае $\alpha = At$ и $\beta = Bt$.

Если применить этот вывод хотя бы ко второму из примеров, который мы привели выше — к скоростям молекул, то мы получим следующее: вероятность того, что за промежуток времени продолжительности t какая-нибудь компонента скорости одной определенной молекулы получит приращение, заключенное в бесконечно малом интервале $(x, x+dx)$ равна только что полученной величине $p(x, t)$, умноженной на dx, т. е. определяется по закону Гаусса.

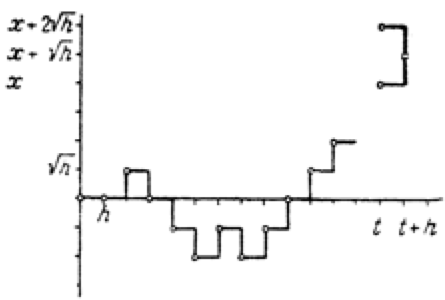
Если скомбинировать этот вывод с законом больших чисел, то мы получим такой результат. Будем рассматривать не одну молекулу, а много их, скажем для определенности — все молекулы в некоторой массе газа. В силу теоремы Бернулли для большого числа молекул мы будем иметь практическую уверенность в том, что *относительное число* молекул, для которых за промежуток времени данной продолжительности t компонента скорости получит приращение, заключенное в бесконечно малом интервале $(x, x+dx)$, будет как угодно мало отличаться от

$$\frac{1}{\sqrt{2\pi Bt}}e^{-\frac{(x-At)^2}{2Bt}}dx.$$

Это утверждение составляет содержание закона распределения скоростей Максвелла.

79. Появление закона Гаусса в качестве решения поставленной нами задачи и появление того же закона Гаусса в качестве асимптотической формулы в п. 32 — так же не случайное совпадение, как это было и с законом Пуассона. Это можно наглядно показать на следующей схеме.

Как и в случае закона Пуассона, возьмем прямую $-\infty < x < +\infty$ и представим себе, что по ней передвигается точка P, начинающая свое движение в 0, но на этот раз продолжающая его следующим образом. В течение некоторого промежутка времени h она останется неподвижной; затем делает моментальный скачок либо на расстояние $\sqrt{h}$ вверх, либо на такое же расстояние $\sqrt{h}$ вниз, после чего остается неподвижной до истечения времени $2h$; затем снова делает моментальный скачок либо на $\sqrt{h}$ вверх, либо на $\sqrt{h}$ вниз, после чего опять остается неподвижной до истечения времени $3h$ и т. д.



Черт. 12б.

График движения точки P может оказаться тем или иным в зависимости от случая, но он всегда будет состоять из горизонтальных отрезков длины h и из вертикальных отрезков длины $\sqrt{h}$, как это изображено на черт. 12б. Допустим для простоты, что в любой момент времени, кратный h, вероятности скачков на $\sqrt{h}$ вверх и на $\sqrt{h}$ вниз одинаковы и, следовательно, обе равны половине — независимо от того, на какой высоте точка P уже находится. Вычислим вероятность $f(r, n)$ того, что в момент $t = nh$ окончательное положение точки P будет на высоте $x = r\sqrt{h}$, положительной или отрицательной. Для этого установим связь между последовательными положениями точки в моменты $t = nh$ и $t + h = (n+1)h$. Ясно, например, что в момент $t+h$ окончательное положение точки может оказаться на высоте $x + \sqrt{h}$ только в двух случаях: либо 1) если окончательной ее высотой в момент t была высота x, а в момент $t+h$ точка P сделала скачок на $\sqrt{h}$ вверх, либо 2) если окончательной ее высотой в момент t была высота $x + 2\sqrt{h}$, а в момент $t+h$ точка P сделала скачок на $\sqrt{h}$ вниз. Вероятности скачков в обоих случаях по условию равны половине. Отсюда, пользуясь формулой полной вероятности, мы получаем

$$(\alpha) \qquad f(r+1, n+1) = \frac{1}{2}f(r, n) + \frac{1}{2}f(r+2, n).$$

Кроме того, нетрудно сообразить, что всегда

$$(\beta) \qquad f(-r, n) = f(r, n) \quad \text{и} \quad \sum_{r=-n}^{n} f(r, n) = 1.$$

При этом по самому смыслу задачи достаточно определить функцию $f(r, n)$ только для пар целых значений r и n, удовлетворяющих неравенствам $|r| \leqslant n$

и $n \geqslant 0$, ибо в течение времени $t = nh$ точка P может сделать не больше, чем n скачков вверх, и не больше, чем n скачков вниз. Можно проверить простым вычислением, что функция

$$f(r, n) = \frac{n!}{\left(\frac{n+r}{2}\right)!\left(\frac{n-r}{2}\right)!} \frac{1}{2^n}$$

удовлетворяет уравнению (α) и условиям (β) и определена в области $|r| \leqslant n$, $n \geqslant 0$. Следовательно, она является одним из решений нашей задачи. Но нетрудно убедиться, сверх того, в том, что уравнение (α) при условиях (β) может иметь в области $|r| \leqslant n$, $n \geqslant 0$ только одно решение. Следовательно, указанное нами решение — окончательное.

Зафиксируем теперь момент t, но будем безгранично уменьшать промежуток времени h и одновременно безгранично увеличивать число n таких промежутков времени — так, чтобы оставалось в силе равенство $t = nh$. Соответственно будем разрешать точке P делать скачки на $\sqrt{h}$ вверх или на $\sqrt{h}$ вниз по истечении каждого промежутка времени, кратного h; и пусть вероятность каждого скачка вверх и каждого скачка вниз попрежнему равна половине. Посмотрим, во что превратятся в пределе уравнение (α), условия (β) и найденное нами решение при $h \to 0$. Для этого прежде всего напишем уравнение (α) в виде

$$f(r+1, n+1) - f(r+1, n) = \frac{1}{2}[f(r+2, n) - 2f(r+1, n) + f(r, n)].$$

Положим затем

$$p(x, t) = \frac{1}{\sqrt{h}} f(r, n),$$

где $t = nh$ и $x = r\sqrt{h}$; тогда только что написанное уравнение примет вид

$$p(x+\sqrt{h}, t+h) - p(x+\sqrt{h}, t) =$$
$$= \frac{1}{2}[p(x+2\sqrt{h}, t) - 2p(x+\sqrt{h}, t) + p(x, t)].$$

Разделив обе части этого последнего уравнения на h и переходя к пределу при $h \to 0$, мы получим

$$(\alpha^*) \qquad \frac{\partial p(x, t)}{\partial t} = \frac{1}{2} \frac{\partial^2 p(x, t)}{\partial x^2},$$

т. е. частный случай найденного нами выше диференциального уравнения закона Гаусса. Условие (β) примет вид

$$p(-x, t) = p(x, t) \quad \text{и} \quad \sum_{x=-n\sqrt{h}}^{n\sqrt{h}} p(x, t)\sqrt{h} = 1.$$

Так как при $h \to 0$ мы имеем $n = \frac{t}{h} \to \infty$, следовательно, $n\sqrt{h} = \sqrt{nt} \to \infty$, то при $h \to 0$ эти условия дают

$$(\beta^*) \qquad p(-x, t) = p(x, t) \quad \text{и} \quad \int_{-\infty}^{+\infty} p(x, t)\, dx = 1.$$

Что же касается самого решения уравнения (α) при условиях (β), то его предел при $h \to 0$, или, что то же, при $n \to \infty$, можно найти так же, как был найден предел аналогичного выражения в локальной теореме Лапласа; мы получим для

$$\frac{1}{\sqrt{h}} f(r, n)$$

предел

$$p(x, t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}},$$

т. е. как раз решение уравнения (α^*) при условиях (β^*).

Пусть теперь точка P движется непрерывно и пусть расстояние, проходимое ею за малый промежуток времени Δt, имеет порядок $\sqrt{\Delta t}$, так что квадрат этого расстояния имеет порядок Δt, направление же движения вверх или вниз зависит от случая и в каждый момент времени то и другое одинаково вероятно. Мы окажемся в условиях 1°—4°, которые были приведены в начале п. 78; при этом, в частности, здесь будет $A = 0$ и $B = 1$. Следовательно, здесь уравнение (α^*) будет точным, а не только асимптотическим уравнением для $p(x, t)$; это же относится и к условиям (β^*). И поэтому закон Гаусса будет точным, а не только асимптотическим решением задачи.

Но предположить, что точка P движется непрерывно — это значит предположить, что нижняя грань промежутка времени между двумя последовательными изменениями положения этой точки, — которую мы обозначали выше через h, — здесь равна 0. Таким образом связь между предельным и точным законом Гаусса можно кратко охарактеризовать так: закон Гаусса одинаково получается как в пределе при $h \to 0$, так и при $h = 0$.

80. В заключение заметим, что в рассмотренных нами задачах, приводящих к законам Пуассона и Гаусса, мы имели простейшие типы *случайных процессов*. В обоих случаях мы имели величину, которая изменяется во времени, получая в различные промежутки времени случайные приращения. В задачах, приводящих к закону Пуассона, роль такого приращения играет число наступлений события A. В задачах, приводящих к закону Гаусса, — приращение ξ. В обоих случаях мы определяли вероятность $P(\mathfrak{E}, t)$ того, что названное приращение за промежуток времени продолжительности t примет значение, принадлежащее некоторому множеству $\mathfrak{E}$: в первом случае множество $\mathfrak{E}$ состояло из одного значения m; во втором случае множеством $\mathfrak{E}$ был интервал (x_1, x_2).

В рассмотренных нами случаях вероятность $P(\mathfrak{E}, t)$ зависела, действительно, только от $\mathfrak{E}$ и от t; зависимость этой вероятности от того, какое значение имеет величина в начале данного промежутка времени, и от того, как расположен сам этот промежуток на шкале времени, возможная в других случаях, у нас была исключена условиями 1° и 2°. Случайные процессы, удовлетворяющие этим условиям 1° и 2°, называются *однородными*.

Мы изучили наиболее интересные типы однородных случайных процессов. Общая форма однородных случайных процессов, определяемых „вероятностями перехода" $P(\mathfrak{E}, t)$, была найдена А. Н. Колмогоровым и П. Леви́, первый из которых является вообще основателем общей математической теории случайных процессов [1]).

---

[1]) Общее исследование основных типов случайных процессов можно найти в статьях А. Н. Колмогорова, А. Я. Хинчина и Феллера в *Успехах математических наук*, вып. V и в книге А. Я. Хинчина *Асимптотические законы теории вероятностей*, 1936.

ГЛАВА VII.

ХАРАКТЕРИСТИЧЕСКИЕ ФУНКЦИИ.

§ 23. Основные свойства характеристических функций.

81. Доказательство теоремы Лапласа, приведенное в гл. III, показывает, что установление асимптотических формул может требовать довольно сложных вычислений. Так обстоит дело даже в том простейшем случае, которым мы занимались — в случае последовательных испытаний. В более общих случаях трудности еще значительно возрастают. Поэтому важной задачей теории вероятностей является разработка методов, при помощи которых можно было бы подойти к вопросу об асимптотических формулах с какой-нибудь общей точки зрения и по возможности простым путем. Одним из наиболее общих и изящных является разработанный в последнее время метод характеристических функций, основы которого мы сейчас и изложим.

При этом для функций распределения $F(x)$ случайных величин ξ, — которыми нам постоянно придется пользоваться, — мы примем в этой главе не первое из трех возможных определений, указанных в п. 46, а последнее из них, т. е. будем полагать

$$F(x)=\frac{\boldsymbol{P}(\xi<x)+\boldsymbol{P}(\xi\leqslant x)}{2}.$$

Определенная таким образом функция распределения, очевидно, обладает тем свойством, что

$$F(x)=\frac{F(x-0)+F(x+0)}{2}.$$

Благодаря именно этому свойству этим определением удобно пользоваться в связи с аналитическим аппаратом, который нам придется применять (интегралы Фурье).

В определение тех числовых характеристик случайных величин, с которыми мы будем иметь дело, это не внесет никаких изменений. Мы будем иметь дело со средними значениями непрерывных функций $f(x)$, т. е. с интегралами вида

$$\int_{-\infty}^{+\infty} f(x)\,dF(x);$$

в определении же этих последних функция распределения $F(x)$ может одинаково фигурировать в обоих вышеназванных смыслах без того, чтобы это как-нибудь повлияло на величину и на самое существование интеграла [1]).

Характеристической функцией $\varphi(t)$ случайной величины ξ называется среднее значение функции $e^{it\xi}$, где t — действительная переменная:

$$\varphi(t) = \boldsymbol{E}e^{it\xi},$$

или, если обозначить через $F(x)$ функцию распределения величины ξ:

$$\varphi(t) = \int\limits_{-\infty}^{+\infty} e^{itx}\,dF(x).$$

Характеристическая функция определена для всякой случайной величины и при всяком t, так как $e^{itx} = \cos tx + i \sin tx$, а $\cos tx$ и $\sin tx$ при всяком t непрерывны и ограничены на всей прямой $-\infty < x < +\infty$.

Нетрудно видеть, что и сама характеристическая функция $\varphi(t)$ непрерывна, и $|\varphi(t)| \leqslant 1$ на всей прямой $-\infty < t < +\infty$. При этом всегда $\varphi(0) = 1$.

Обратно, если задать характеристическую функцию $\varphi(t)$, то будет вполне определена и функция распределения $F(x)$. А именно, имеет место следующая *формула обращения*:

$$F(x) - F(0) = \frac{1}{2\pi} \lim_{n\to\infty} \int\limits_{-n}^{n} \frac{1 - e^{-itx}}{it}\,\varphi(t)\,dt.$$

В самом деле, по известному свойству интеграла Дирихле

$$(\alpha)\qquad F(x) - F(0) = \frac{1}{\pi} \lim_{n\to\infty} \int\limits_{-\infty}^{+\infty} \frac{\sin nz}{z}\,[F(z+x) - F(z)]\,dz.$$

Но

$$\int\limits_{-\infty}^{+\infty} \frac{\sin nz}{z}\,[F(z+x) - F(z)]\,dz = \frac{1}{2} \int\limits_{-\infty}^{+\infty} \Big[\int\limits_{-n}^{n} e^{itz}\,dt\Big]\,[F(z+x) - F(z)]\,dz.$$

Меняя порядок интегрирования, мы получаем отсюда

$$(\beta)\qquad \int\limits_{-\infty}^{+\infty} \frac{\sin nz}{z}\,[F(z+x) - F(z)]\,dz = \frac{1}{2} \int\limits_{-n}^{n} \int\limits_{-\infty}^{+\infty} e^{itz}\,[F(z+x) - F(z)\,dz\;dt.$$

---

[1]) Строго говоря, мы будем иметь дело с интегралами от комплексных функций $f(x) = f_1(x) + if_2(x)$ действительного переменного x, где i, как и всюду на протяжении настоящей главы, обозначает мнимую единицу $\sqrt{-1}$. Такие интегралы определяют, полагая

$$\int\limits_a^b f(x)\,dF(x) = \int\limits_a^b f_1(x)\,dF(x) + i\int\limits_a^b f_2(x)\,dF(x).$$

Далее, по свойству 6) интеграла Стильтьеса, приведенному в п. 54:

$$\int\limits_{-\infty}^{+\infty} e^{itz}\left[F(z+x)-F(z)\right]dz=\int\limits_{-\infty}^{+\infty} e^{itz}\,d_z\int\limits_{z}^{z+x}F(u)\,du=$$

$$=\int\limits_{-\infty}^{+\infty} e^{itz}\,d_z\int\limits_{0}^{x}F(y+z)\,dy.$$

Отсюда, по свойству 8) интеграла Стильтьеса, приведенному в п. 54:

$$\int\limits_{-\infty}^{+\infty} e^{itz}\left[F(z+x)-F(z)\right]dz=\int\limits_{0}^{x}\Big[\int\limits_{-\infty}^{+\infty} e^{itz}\,dF(y+z)\Big]\,dy=$$

$$=\int\limits_{0}^{x}\Big[\int\limits_{-\infty}^{+\infty} e^{it(u-y)}\,dF(u)\Big]\,dy.$$

Отсюда окончательно

$$(\gamma)\quad \int\limits_{-\infty}^{+\infty} e^{itz}\left[F(z+x)-F(z)\right]dz=\int\limits_{0}^{x} e^{-ity}\,dy\int\limits_{-\infty}^{+\infty} e^{itu}\,dF(u)=\frac{1-e^{-itx}}{it}\,\varphi(t).$$

Сопоставляя (α), (β) и (γ), мы и получаем формулу обращения.

Законность сделанной в этом доказательстве ссылки на интеграл Дирихле и перемены порядка интегрирования обеспечивается тем, что функция $F(z+x)-F(z)$ при всяком x абсолютно интегрируема на всей прямой $-\infty<z<+\infty$. В самом деле, при любом конечном $c>0$

$$\int\limits_{-c}^{c}\left|F(z+x)-F(z)\right|dz=\Big|\int\limits_{-c}^{c}\left[F(z+x)-F(z)\right]dz\Big|=$$

$$=\Big|\int\limits_{c}^{c+x}F(z)\,dz-\int\limits_{-c}^{-c+x}F(z)\,dz\Big|=\left|\mu' x-\mu'' x\right|,$$

где μ' лежит между $F(c)$ и $F(c+x)$, а μ'' — между $F(-c)$ и $F(-c+x)$. Если же $c\to+\infty$, то $\mu'\to 1$ и $\mu''\to 0$, и потому

$$\int\limits_{-\infty}^{+\infty}\left|F(z+x)-F(z)\right|dz=|x|.$$

Таким образом, по данной характеристической функции $\varphi(t)$ определяется разность $F(x)-F(0)$. А так как $F(-\infty)=0$, то вполне определяется и сама функция распределения $F(x)=[F(x)-F(0)]-[F(-\infty)-F(0)]$.

82. Преимущества для вычислений, которые дает применение характеристических функций, основаны главным образом на следующих двух теоремах, XVII и XVIII. Так как в этих теоремах речь идет о нескольких случайных величинах ξ, η, ..., то мы будем обозначать характеристические функции этих величин, снабжая их соответствующими индексами: $\varphi_\xi(t)$, $\varphi_\eta(t)$, ...

Теорема XVII. — *Если $\eta = a\xi + b$, где a и b — постоянные, то* $\varphi_\eta(t) = \varphi_\xi(at)\,e^{ibt}$.

Доказательство.

$$\varphi_\eta(t) = \boldsymbol{E}e^{it\eta} = \boldsymbol{E}e^{it(a\xi+b)} = \boldsymbol{E}(e^{ibt}e^{iat\xi}) = e^{ibt}\boldsymbol{E}e^{iat\xi} = \varphi_\xi(at)\,e^{ibt}.$$

Теорема XVIII. — Если $\zeta = \xi + \eta$, где ξ *и* η — *независимые случайные величины, то*

$$\varphi_\zeta(t) = \varphi_\xi(t)\,\varphi_\eta(t).$$

Доказательство. — Так как вместе с ξ и η, очевидно, также $e^{it\xi}$ и $e^{it\eta}$ — независимые случайные величины, то

$$\varphi_\zeta(t) = \boldsymbol{E}e^{it\zeta} = \boldsymbol{E}e^{it(\xi+\eta)} = \boldsymbol{E}(e^{it\xi}e^{it\eta}) = \boldsymbol{E}e^{it\xi}\cdot\boldsymbol{E}e^{it\eta} = \varphi_\xi(t)\,\varphi_\eta(t).$$

Обобщение теоремы XVIII. — *Если* $\xi = \xi_1 + \xi_2 + \ldots + \xi_n$, *где каждое из слагаемых, ξ_k, есть случайная величина, независимая от суммы предшествующих слагаемых* $\xi_1 + \xi_2 + \ldots + \xi_{k-1}$, *то*

$$\varphi_\xi(t) = \varphi_{\xi_1}(t)\,\varphi_{\xi_2}(t)\,\ldots\,\varphi_{\xi_n}(t).$$

Доказательство проводится точно так же, как доказательство данного выше обобщения теоремы XI: следует только заменить всюду средние значения характеристическими функциями и суммы средних значений произведениями характеристических функций.

83. Пусть теперь случайная величина ξ такова, что существует ее *момент*

$$\boldsymbol{E}\xi^n = \int_{-\infty}^{+\infty} x^n dF(x)$$

порядка n. В соответствии с определением интеграла Стильтьеса этот момент существует только в том случае, если существует и *абсолютный момент*

$$\boldsymbol{E}|\xi|^n = \int_{-\infty}^{+\infty} |x|^n dF(x).$$

Тогда, очевидно, существуют и моменты всех порядков $k \leqslant n$, и характеристическую функцию

$$\varphi(t) = \int_{-\infty}^{+\infty} e^{itx}\,dF(x)$$

можно диференцировать $k \leqslant n$ раз, в результате чего мы получим

$$\varphi^{(k)}(t) = i^k \int_{-\infty}^{+\infty} x^k e^{itx}\,dF(x).$$

Полагая здесь $t=0$, мы получим

$$\varphi^{(k)}(0)=i^k \boldsymbol{E}\xi^k.$$

Отсюда формула для определения моментов при помощи характеристической функции:

$$\boldsymbol{E}\xi^k=\frac{1}{i^k}\varphi^{(k)}(0).$$

Среднее значение $\boldsymbol{E}\xi$ и дисперсия $\boldsymbol{D}\xi$ также просто выражаются при помощи л о г а р и ф м а характеристической функции. В самом деле, положим

$$\psi(t)=\log\varphi(t).$$

Тогда, принимая во внимание, что $\varphi(0)=1$, мы имеем

$$\psi'(0)=\varphi'(0)=i\boldsymbol{E}\xi,$$

$$\psi''(0)=\varphi''(0)-[\varphi'(0)]^2=-\boldsymbol{E}\xi^2+(\boldsymbol{E}\xi)^2=-\boldsymbol{D}\xi.$$

Отсюда

$$\boldsymbol{E}\xi=\frac{1}{i}\psi'(0),$$

$$\boldsymbol{D}\xi=-\psi''(0).$$

Производные $\psi^{(k)}(0)$ логарифма характеристической функции в точке 0, умноженные на i^k, называются *семиинвариантами* соответствующей случайной величины. При сложении независимых случайных величин их семиинварианты также складываются: в самом деле, пусть ξ и η — независимые случайные величины и пусть $\zeta=\xi+\eta$. Так как $\varphi_\zeta(t)=\varphi_\xi(t)\,\varphi_\eta(t)$, то $\psi_\zeta(t)=\psi_\xi(t)+\psi_\eta(t)$; отсюда путем k-кратного диференцирования в точке 0 и умножения на i^k и вытекает справедливость нашего утверждения.

Мы только что видели, что первыми двумя семиинвариантами являются среднее значение и дисперсия, т. е. моменты первого порядка и некоторая рациональная функция от моментов первого и второго порядков. Путем вычислений можно убедиться в том, что и каждый n-й семиинвариант есть рациональная функция от моментов порядка $\leqslant n$.

84. Применим сказанное о характеристических функциях к случайным величинам, подчиняющимся закону Гаусса. Рассмотрим сначала случай нормальной кривой распределения, где $\alpha=0$ и $\beta=1$. Здесь

$$F'(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}};$$

следовательно:

$$\varphi(t)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty}e^{itx-\frac{x^2}{2}}\,dx.$$

Подстановкой

$$x=s+it$$

мы получаем

$$\varphi(t)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty}e^{-\frac{t^2}{2}-\frac{s^2}{2}}\,ds=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\int\limits_{-\infty}^{+\infty}e^{-\frac{s^2}{2}}\,ds=e^{-\frac{t^2}{2}}.$$

Отсюда нетрудно перейти к общему случаю: для этого достаточно заметить, что если величина ξ подчиняется закону Гаусса с параметрами $\alpha=\boldsymbol{E}\xi$ и $\beta=\boldsymbol{D}\xi$, то ее нормированное отклонение

$$\overline{\xi}=\frac{\xi-\alpha}{\sqrt{\beta}}$$

тоже подчиняется закону Гаусса с параметрами $\overline{\alpha}=\boldsymbol{E}\overline{\xi}=0$ и $\overline{\beta}=$ $=\boldsymbol{D}\overline{\xi}=1$ и, следовательно, имеет характеристическую функцию

$$\varphi_{\overline{\xi}}(t)=e^{-\frac{t^2}{2}}.$$

Поэтому сама величина

$$\xi=\sqrt{\beta}\,\overline{\xi}+\alpha$$

имеет в силу теоремы XVII характеристическую функцию

$$\varphi_{\xi}(t)=\varphi_{\overline{\xi}}(\sqrt{\beta}\,t)\,e^{i\alpha t}=e^{-\frac{\beta t^2}{2}}e^{i\alpha t}=e^{i\alpha t-\frac{\beta t^2}{2}}.$$

Пусть теперь $\eta=a\xi+b$, где a и b — постоянные, и пусть ξ есть случайная величина, подчиняющаяся закону Гаусса с параметрами α и β. Тогда в силу теоремы XVII

$$\varphi_{\eta}(t)=\varphi_{\xi}(at)\,e^{ibt}=e^{i(a\alpha+b)t-\frac{a^2\beta t^2}{2}}.$$

Мы видим, что η, в свою очередь, подчиняется закону Гаусса, с параметрами $a\alpha+b$ и $a^2\beta$.

Пусть, наконец, $\zeta=\xi+\eta$, где ξ и η — независимые случайные величины, и пусть обе они подчиняются закону Гаусса, первая — с параметрами α_1 и β_1, вторая — с параметрами α_2 и β_2. Тогда в силу теоремы XVIII

$$\varphi_{\zeta}(t)=\varphi_{\xi}(t)\,\varphi_{\eta}(t)=e^{i(\alpha_1+\alpha_2)t-\frac{(\beta_1+\beta_2)t^2}{2}}.$$

Мы видим, что ζ, в свою очередь, подчиняется закону Гаусса, с параметрами $\alpha_1+\alpha_2$ и $\beta_1+\beta_2$.

Вычислим еще моменты случайной величины ξ, подчиняющейся закону Гаусса, причем мы здесь ограничимся случаем нормальной кривой распределения, где $\alpha=0$ и $\beta=1$. Мы имеем, пользуясь формулой Лейбница для производных произведения:

$$\varphi^{(n)}(t)=\left(e^{-\frac{t^2}{2}}\right)^{(n)}=\left(-te^{-\frac{t^2}{2}}\right)^{(n-1)}=$$
$$=-t\left(e^{-\frac{t^2}{2}}\right)^{(n-1)}-(n-1)\left(e^{-\frac{t^2}{2}}\right)^{(n-2)}=-t\varphi^{(n-1)}(t)-(n-1)\,\varphi^{(n-2)}(t).$$

Отсюда, полагая $t=0$:

$$\varphi^{(n)}(0)=-(n-1)\,\varphi^{(n-2)}(0).$$

Следовательно:

$$\boldsymbol{E}\xi^n=\frac{1}{i^n}\varphi^{(n)}(0)=-\frac{n-1}{i^n}\varphi^{(n-2)}(0)=\frac{n-1}{i^{n-2}}\varphi^{(n-2)}(0)=(n-1)\,\boldsymbol{E}\xi^{n-2}.$$

Отсюда, имея в виду, что $E\xi^0 = 1$, мы имеем

$$E\xi^2 = 1,$$
$$E\xi^4 = 1\cdot 3,$$
$$E\xi^6 = 1\cdot 3\cdot 5,$$
$$\ldots\ldots\ldots,$$

вообще для четных k

$$E\xi^k = 1\cdot 3\cdot 5\ldots(k-1).$$

Точно так же, имея в виду, что $E\xi = 0$, мы имеем для нечетных k

$$E\xi^k = 0.$$

§ 24. Предельные теоремы.

85. С точки зрения применения в теории асимптотических формул важнейшими теоремами теории характеристических функций являются две предельные теоремы, прямая и обратная. К их установлению мы сейчас и приступим; для этого, однако, удобно будет ввести некоторые вспомогательные понятия и предварительно доказать две теоремы чисто аналитического характера — первую и вторую теоремы Хелли (Helly).

Условимся говорить, что последовательность неубывающих функций

(69) $$F_1(x),\ F_2(x),\ \ldots,\ F_n(x),\ \ldots$$

сходится в основном к неубывающей функции

(70) $$G(x),$$

если она сходится к этой последней в каждой точке ее непрерывности.

Отметим сразу же, что для сходимости в основном достаточно, чтобы последовательность (69) сходилась к функции (70) на каком-нибудь всюду плотном множестве D. Действительно, пусть x — любая точка и пусть x' и x'' — какие-нибудь две точки множества D, такие, что $x' < x < x''$. Тогда и

$$F_n(x') \leqslant F_n(x) \leqslant F_n(x'').$$

Следовательно:

$$\varliminf_{n\to\infty} F_n(x') \leqslant \varliminf_{n\to\infty} F_n(x) \leqslant \varlimsup_{n\to\infty} F_n(x) \leqslant \varlimsup_{n\to\infty} F_n(x'').$$

А так как по предположению $\lim\limits_{n\to\infty} F_n(x') = G(x')$ и $\lim\limits_{n\to\infty} F_n(x'') = G(x'')$, то и

$$G(x') \leqslant \varliminf_{n\to\infty} F_n(x) \leqslant \varlimsup_{n\to\infty} F_n(x) \leqslant G(x'').$$

Но средние члены в этих неравенствах не зависят от x' и x'', и потому также

$$G(x-0) \leqslant \varliminf_{n\to\infty} F_n(x) \leqslant \varlimsup_{n\to\infty} F_n(x) \leqslant G(x+0).$$

Если теперь функция $G(x)$ в точке x непрерывна, то

$$G(x-0)=G(x)=G(x+0).$$

Поэтому в этом случае последние неравенства дают

$$\lim_{n\to\infty} F_n(x)=G(x),$$

что и требовалось установить.

Первая теорема Хелли. — *Всякая последовательность*

(S) $$F_1(x),\ F_2(x),\ \ldots,\ F_n(x),\ \ldots$$

функций распределения случайных величин содержит по крайней мере одну подпоследовательность

$$F_{n_1}(x),\ F_{n_2}(x),\ \ldots,\ F_{n_k}(x),\ \ldots,$$

которая сходится в основном к некоторой неубывающей функции $G(x)$.

Доказательство. — Пусть D — какое-нибудь счетное и всюду плотное множество точек $x_1',\ x_2',\ \ldots,\ x_n',\ \ldots$ Возьмем значения функций последовательности (S) в точке x_1':

$$F_1(x_1'),\ F_2(x_1'),\ \ldots,\ F_n(x_1'),\ \ldots$$

Так как множество этих значений ограничено (ибо все они лежат между нулем и единицей), то, как известно, это множество содержит по крайней мере одну последовательность, сходящуюся к некоторому предельному значению $G(x_1')$ (лежащему между нулем и единицей). Иначе говоря, последовательность (S) содержит по крайней мере одну подпоследовательность

(S_1) $$F_{11}(x),\ F_{12}(x),\ \ldots,\ F_{1k}(x),\ \ldots,$$

для которой $\lim\limits_{k\to\infty} F_{1k}(x_1')=G(x_1')$. Далее возьмем значения всех функций последовательности (S_1) в точке x_2':

$$F_{11}(x_2'),\ F_{12}(x_2'),\ \ldots,\ F_{1k}(x_2'),\ \ldots$$

Так как множество и этих значений ограничено, то и оно содержит по крайней мере одну последовательность, сходящуюся к некоторому предельному значению $G(x_2')$, т. е. последовательность (S_1) содержит по крайней мере одну подпоследовательность

(S_2) $$F_{21}(x),\ F_{22}(x),\ \ldots,\ F_{2k}(x)\ \ldots,$$

для которой $\lim\limits_{k\to\infty} F_{2k}(x_2')=G(x_2')$. Продолжая такое выделение подпоследовательностей из полученных ранее последовательностей, мы для каждой точки x'_n множества D получим последовательность

$$(\mathrm{S}_n) \qquad F_{n1}(x),\ F_{n2}(x),\ \ldots,\ F_{nk}(x),\ \ldots,$$

для которой $\lim\limits_{k\to\infty} F_{nk}(x'_n) = G(x'_n)$. Составим теперь диагональную последовательность

$$F_{11}(x),\ F_{22}(x),\ \ldots,\ F_{kk}(x),\ \ldots$$

Вся она в конечном счете выделена из последовательности (S_1); поэтому $\lim\limits_{k\to\infty} F_{kk}(x'_1) = \lim\limits_{k\to\infty} F_{1k}(x'_1) = G(x'_1)$. Далее, вся диагональная последовательность, за исключением лишь первого ее члена $F_{11}(x)$, выделена из последовательности (S_2); поэтому $\lim\limits_{k\to\infty} F_{kk}(x'_2) = \lim\limits_{k\to\infty} F_{2k}(x'_2) = G(x'_2)$. Вообще вся диагональная последовательность, за исключением ее первых $n-1$ членов $F_{11}(x)$, $F_{22}(x)$, …, $F_{n-1,\,n-1}(x)$, выделена из последовательности (S_n); поэтому $\lim\limits_{k\to\infty} F_{kk}(x'_n) = \lim\limits_{k\to\infty} F_{nk}(x'_n) = G(x'_n)$ при каждом n. Полученный результат можно формулировать так: последовательность (S) содержит по крайней мере одну подпоследовательность

$$(\mathrm{S}^*) \qquad F_{n_1}(x),\ F_{n_2}(x),\ \ldots,\ F_{n_k}(x),\ \ldots,$$

которая во всех точках x'_n множества D сходится к некоторой функции $G(x)$, определенной на этом множестве D. При этом, так как функции $F_{n_k}(x)$ — неубывающие и равномерно ограниченные (ибо значения их всех лежат между нулем и единицей), то, очевидно, и функция $G(x)$ будет неубывающей и ограниченной (ее значения также будут лежать между нулем и единицей).

Теперь ясно, что функцию $G(x)$, определенную на множестве D, можно продолжить так, что она будет определена на всей прямой $-\infty < x < +\infty$, оставаясь неубывающей и ограниченной: для этого достаточно положить в любой точке x, не принадлежащей множеству D, $G(x) = \lim\limits_{x'\to x} G(x')$, где x' принадлежит множеству D и находится слева от точки x. Последовательность (S^*) сходится к этой функции на всюду плотном множестве D; следовательно, она сходится к ней в основном, что и доказывает теорему.

Вторая теорема Хелли. — *Пусть $f(x)$ — непрерывная функция и пусть последовательность*

$$F_1(x),\ F_2(x),\ \ldots,\ F_n(x),\ \ldots$$

функций распределения случайных величин сходится в основном к функции $G(x)$ на некотором конечном интервале $a \leqslant x \leqslant b$, где

a и b — точки непрерывности функции $G(x)$; тогда

$$\lim_{n\to\infty}\int_a^b f(x)\,dF_n(x)=\int_a^b f(x)\,dG(x).$$

Доказательство. — Из непрерывности функции $f(x)$ вытекает, что, как бы мала ни была положительная постоянная ε, найдется подразделение интервала $a\leqslant x\leqslant b$ точками деления $a=x_0$, x_1, x_2, ..., $x_N=b$ на частичные интервалы длины Δx_0, Δx_1, ..., Δx_{N-1}, такое что в каждом интервале $x_k\leqslant x\leqslant x_k+\Delta x_k$ будет выполняться неравенство $|f(x)-f(x_k)|<\varepsilon$. Пользуясь этим обстоятельством, мы можем ввести вспомогательную функцию $f_\varepsilon(x)$, принимающую только конечное число значений, а именно функцию $f_\varepsilon(x)$, определяемую равенствами $f_\varepsilon(x)=f(x_k)$ при $x_k\leqslant x<x_k+\Delta x_k$, для которой во всем интервале $a\leqslant x\leqslant b$ будет выполняться неравенство

$$|f(x)-f_\varepsilon(x)|<\varepsilon. \tag{71}$$

При этом мы можем, очевидно, заранее выбрать точки деления x_0, x_1, x_2, ..., x_N так, чтобы все они были точками непрерывности функции $G(x)$. Тогда в силу сходимости в основном функций $F_1(x)$, $F_2(x)$, ..., $F_n(x)$, ... к функции $G(x)$ при достаточно больших n во всех этих точках деления будут выполняться неравенства

$$|G(x_k)-F_n(x_k)|<\frac{\varepsilon}{MN}, \tag{72}$$

где M — максимум модуля $|f(x)|$ в интервале $a\leqslant x\leqslant b$.

Очевидно, что

$$\left|\int_a^b f(x)\,dG(x)-\int_a^b f(x)\,dF_n(x)\right|\leqslant\left|\int_a^b f(x)\,dG(x)-\int_a^b f_\varepsilon(x)\,dG(x)\right|+$$
$$+\left|\int_a^b f_\varepsilon(x)\,dG(x)-\int_a^b f_\varepsilon(x)\,dF_n(x)\right|+\left|\int_a^b f_\varepsilon(x)\,dF_n(x)-\int_a^b f(x)\,dF_n(x)\right|.$$

Нетрудно видеть, что в этом неравенстве первое слагаемое правой части не превосходит $\varepsilon[G(b)-G(a)]$, а третье не превосходит $\varepsilon[F_n(b)-F_n(a)]$; это вытекает из неравенства (71). Что же касается второго слагаемого, то оно, очевидно, равно

$$\left|\sum_{k=0}^{N-1} f(x_k)\,[G(x_k+\Delta x_k)-G(x_k)]-\sum_{k=0}^{N-1} f(x_k)\,[F_n(x_k+\Delta x_k)-F_n(x_k)]\right|=$$
$$=\left|\sum_{k=0}^{N-1} f(x_k)\,[G(x_k+\Delta x_k)-F_n(x_k+\Delta x_k)]-\sum_{k=0}^{N-1} f(x_k)\,[G(x_k)-F_n(x_k)]\right|,$$

и следовательно, при достаточно больших n оно не превосходит 2ε; это вытекает из неравенств (72). В силу равномерной ограниченности функций распределения $F_n(x)$ сумма

$$\varepsilon[G(b)-G(a)]+\varepsilon[F_n(b)-F_n(a)]+2\varepsilon$$

может быть сделана как угодно малой вместе с ε. Поэтому при достаточно больших n будет как угодно малой и левая часть нашего неравенства, что и доказывает теорему.

Обобщение второй теоремы Хелли. — *Пусть функция $f(x)$ непрерывна и ограничена на всей прямой $-\infty<x<+\infty$, и пусть последовательность*

$$F_1(x),\ F_2(x),\ \ldots,\ F_n(x),\ \ldots$$

функций распределения случайных величин сходится в основном к функции распределения $G(x)$; тогда

$$\lim_{n\to\infty}\int_{-\infty}^{+\infty} f(x)\,dF_n(x)=\int_{-\infty}^{+\infty} f(x)\,dG(x).$$

Доказательство. — Возьмем любые числа $A<0$ и $B>0$ и положим

$$I_{\mathrm{I}}=\left|\int_{-\infty}^{A} f(x)\,dG(x)-\int_{-\infty}^{A} f(x)\,dF_n(x)\right|,$$

$$I_{\mathrm{II}}=\left|\int_{A}^{B} f(x)\,dG(x)-\int_{A}^{B} f(x)\,dF_n(x)\right|,$$

$$I_{\mathrm{III}}=\left|\int_{B}^{+\infty} f(x)\,dG(x)-\int_{B}^{+\infty} f(x)\,dF_n(x)\right|.$$

Тогда, очевидно,

$$\left|\int_{-\infty}^{+\infty} f(x)\,dG(x)-\int_{-\infty}^{+\infty} f(x)\,dF_n(x)\right|\leqslant I_{\mathrm{I}}+I_{\mathrm{II}}+I_{\mathrm{III}}.$$

Величины I_{I} и I_{III} могут быть сделаны как угодно малыми, если подобрать A и B достаточно большими по абсолютной величине и притом такими, чтобы точки A и B были точками непрерывности функции $G(x)$ и взять затем достаточно большое n. В самом деле, пусть M — верхняя грань модуля $|f(x)|$ на всей прямой $-\infty<x<$ $<+\infty$. Тогда

$$I_{\mathrm{I}}\leqslant M[G(A)+F_n(A)],$$

$$I_{\mathrm{III}}\leqslant M[1-G(B)+1-F_n(B)].$$

Но

$$\lim_{A\to-\infty} G(A)=0,\quad \lim_{B\to+\infty} G(B)=1,$$

и в силу сходимости в основном функции $F_n(x)$ к функции $G(x)$

$$\lim_{n\to\infty} F_n(A) = G(A), \quad \lim_{n\to\infty} F_n(B) = G(B).$$

Это и доказывает наше утверждение об I_{I} и I_{III}.

Величина же I_{II} может быть сделана как угодно малой, как бы ни были подобраны A и B, — лишь бы точки A и B были точками непрерывности функции $G(x)$ — если взять достаточно большое n. В самом деле, из непрерывности функции $f(x)$ на всей прямой $-\infty < x < +\infty$ вытекает, что она непрерывна в интервале $A \leqslant x \leqslant B$; поэтому здесь применима вторая теорема Хелли для ограниченного интервала.

Ясно, что все требования, которые мы предъявляем к подбору A, B и затем n, совместимы, и потому сумма $I_{\text{I}} + I_{\text{II}} + I_{\text{III}}$ может быть сделана как угодно малой. Тем самым обобщение второй теоремы Хелли полностью доказано.

86. Перейдем к предельным теоремам теории характеристических функций.

Прямая предельная теорема (Леви́). — *Если последовательность*

$$F_1(x),\ F_2(x),\ \ldots,\ F_n(x),\ \ldots$$

функций распределения случайных величин $\zeta_1, \zeta_2, \ldots, \zeta_n, \ldots$ *сходится в основном к функции распределения* $F(x)$ *случайной величины* ζ, *то и последовательность*

$$\varphi_1(t),\ \varphi_2(t),\ \ldots,\ \varphi_n(t),\ \ldots$$

характеристических функций величин $\zeta_1, \zeta_2, \ldots, \zeta_n$ *сходится к характеристической функции* $\varphi(t)$ *величины* ζ.

Доказательство. — Так как

$$\varphi_n(t) = \int_{-\infty}^{+\infty} e^{itx}\, dF_n(x), \quad \varphi(t) = \int_{-\infty}^{+\infty} e^{itx}\, dF(x)$$

и так как $e^{itx} = \cos tx + i \sin tx$, а $\cos tx$ и $\sin tx$ при каждом t непрерывны и ограничены на всей прямой $-\infty < x < +\infty$, то утверждение нашей теоремы есть непосредственное следствие второй теоремы Хелли [1]).

Обратная предельная теорема (Леви́). — *Если последовательность*

$$(\sigma) \qquad \varphi_1(t),\ \varphi_2(t),\ \ldots,\ \varphi_n(t),\ \ldots$$

характеристических функций случайных величин $\zeta_1, \zeta_2, \ldots, \zeta_n, \ldots$ *сходится к непрерывной функции* $\varphi(t)$, *то последовательность*

$$(\mathrm{S}) \qquad F_1(x),\ F_2(x),\ \ldots,\ F_n(x),\ \ldots$$

---

[1]) Можно доказать и более сильное утверждение, а именно — что функции $\varphi_n(t)$ сходятся к функции $\varphi(t)$ равномерно в каждом конечном интервале; мы не приводим его доказательства, так как нигде им не воспользуемся.

функций распределения величин $\zeta_1, \zeta_2, \ldots, \zeta_n, \ldots$ сходится в основном к функции распределения $F(x)$ некоторой случайной величины ζ (и в силу прямой предельной теоремы, $\varphi(t)$ является характеристической функцией этой величины ζ).

Доказательство. — На основании первой теоремы Хелли мы заключаем, что последовательность (S) непременно содержит подпоследовательность

$$(S') \qquad F_{n_1}(x),\ F_{n_2}(x), \ldots,\ F_{n_k}(x), \ldots,$$

которая сходится в основном к некоторой неубывающей функции $F(x)$. При этом понятно, что эту последнюю можно подобрать так, чтобы в каждой точке x выполнялось равенство

$$F(x) = \frac{F(x-0)+F(x+0)}{2}.$$

Вообще говоря, эта функция $F(x)$ могла бы не быть функцией распределения какой бы то ни было случайной величины, так как для этого еще требуется, чтобы было $F(-\infty)=0$ и $F(+\infty)=1$. А если положить, например,

$$F_n(x) = \begin{cases} 0 \text{ при } x < n, \\ 1 \text{ при } x > n, \end{cases}$$

то мы имели бы тождественно $F(x)=0$, следовательно, $F(-\infty)=0$, но не $F(+\infty)=1$. Однако в условиях нашей теоремы с необходимостью будет $F(-\infty)=0$ и $F(+\infty)=1$. В самом деле, если бы это было не так, то, принимая во внимание, что во всяком случае $F(-\infty)\geqslant 0$ и $F(+\infty)\leqslant 1$, мы имели бы

$$\delta = F(+\infty) - F(-\infty) < 1.$$

Возьмем же какое-нибудь положительное число ε, меньшее чем $1-\delta$. Так как по условию нашей теоремы последовательность характеристических функций (σ) сходится к функции $\varphi(t)$, то $\varphi(0)=1$. А так как, сверх того, функция $\varphi(t)$ непрерывна, то можно выбрать достаточно малое положительное число τ, такое, что будут иметь места неравенство

$$(\Delta) \qquad \frac{1}{2\tau}\left|\int_{-\tau}^{\tau}\varphi(t)\,dt\right| > 1-\frac{\varepsilon}{2} > \delta+\frac{\varepsilon}{2}.$$

Но в то же время можно выбрать $X > \frac{4}{\tau\varepsilon}$ и настолько большое K, чтобы при $k > K$ было

$$\delta_k = F_{n_k}(X) - F_{n_k}(-X) < \delta + \frac{\varepsilon}{4}.$$

А так как $\varphi_{n_k}(t)$ есть характеристическая функция величины ζ_{n_k}, то

$$\int_{-\tau}^{\tau}\varphi_{n_k}(t)\,dt = \int_{-\infty}^{+\infty}\left(\int_{-\tau}^{\tau}e^{itx}\,dt\right)dF_{n_k}(x).$$

Написанный в правой части этого равенства интеграл по переменному t можно оценить следующим образом. С одной стороны, так как $|e^{itx}|=1$, то всегда

$$\left|\int_{-\tau}^{\tau} e^{itx}\,dt\right| \leqslant 2\tau.$$

С другой стороны, так как

$$\int_{-\tau}^{\tau} e^{itx}\,dt = \frac{e^{i\tau x}-e^{-i\tau x}}{ix} = \frac{2}{x}\sin\tau x,$$

а $|\sin\tau x| \leqslant 1$, то при $|x|>X$

$$\left|\int_{-\tau}^{\tau} e^{itx}\,dt\right| < \frac{2}{X}.$$

Отсюда, применяя первую оценку при $|x|\leqslant X$ и вторую при $|x|>X$, мы имеем

$$\left|\int_{-\tau}^{\tau}\varphi_{n_k}(t)\,dt\right| \leqslant \left|\int_{|x|\leqslant X}\left(\int_{-\tau}^{\tau} e^{itx}\,dt\right)dF_{n_k}(x)\right| +$$

$$+\left|\int_{|x|>X}\left(\int_{-\tau}^{\tau} e^{itx}\,dt\right)dF_{n_k}(x)\right| < 2\tau\delta_k+\frac{2}{X},$$

и следовательно,

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau}\varphi_{n_k}(t)\,dt\right| < \delta_k+\frac{1}{\tau X} < \delta+\frac{\varepsilon}{2}.$$

Наконец, благодаря тому, что функции $\varphi_{n_k}(t)$ ограничены в их совокупности и сходятся к функции $\varphi(t)$, это неравенство сохранилось бы и в пределе, в виде

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau}\varphi(t)\,dt\right| \leqslant \delta+\frac{\varepsilon}{2},$$

а это противоречило бы неравенству (Δ).

Итак, функция $F(x)$, к которой сходится в основном последовательность (S'), есть функция распределения некоторой случайной величины; в силу прямой предельной теоремы это — величина, характеристическая функция которой есть $\varphi(t)$. Чтобы закончить доказательство теоремы, нам остается показать, что и вся последовательность (S) сходится в основном к функции $F(x)$. Но если бы это было не так, то нашлась бы точка c непрерывности функции $F(x)$ и некоторая подпоследовательность последовательности (S), которая сходится в этой точке c к числу, не равному $F(c)$. В этой же подпоследовательности, в свою очередь, нашлась бы подпоследовательность

$$(S^*) \qquad F^*_{n_1}(x),\ F^*_{n_2}(x),\ \ldots,\ F^*_{n_k}(x),\ \ldots,$$

которая сходится в основном к некоторой функции $F^*(x)$; для этой последней, в частности, мы имели бы

$$F^*(c) \neq F(c).$$

Но применяя к последовательности (S^*) те же рассуждения, которые мы применили выше к последовательности (S'), мы бы убедились в том, что ее предельная функция $F^*(x)$ является функцией распределения случайной величины, имеющей характеристическую функцию $\varphi(t)$; следовательно, мы имели бы $F^*(x) = F(x)$ и, в частности,

$$F^*(c) = F(c),$$

что противоречило бы только что написанному неравенству.

Следствия обратной предельной теоремы. — Условия теоремы, очевидно, выполнены, если известно, что последовательность (σ) сходится к некоторой функции $\varphi(t)$ *равномерно в каждом конечном интервале*.

Условие теоремы, очевидно, также выполнено, если известно, что последовательность (σ) сходится к некоторой функции $\varphi(t)$, которая является *характеристической функцией какой-нибудь случайной величины*.

87. Доказанные предельные теоремы применяются обычно в сочетании с некоторыми неравенствами, содержащими функцию e^{iy}. Наиболее употребительны два типа этих неравенств, справедливых для действительных y.

I тип:

$$|e^{iy} - 1| \leqslant |y|, \tag{73}$$

$$|e^{iy} - 1 - iy| \leqslant \frac{y^2}{2}, \tag{74}$$

$$\left| e^{iy} - 1 - iy + \frac{y^2}{2} \right| \leqslant \frac{|y|^3}{6}, \tag{75}$$

и т. д.

Первое из этих неравенств можно вывести из равенства $|e^{iy}| = 1$, а именно:

$$|e^{iy} - 1| = \left| \int_0^{iy} e^z\, dz \right| = \left| i \int_0^y e^{iu}\, du \right| \leqslant |y|;$$

второе — из первого:

$$|e^{iy} - 1 - iy| = \left| \int_0^{iy} (e^z - 1)\, dz \right| = \left| i \int_0^y (e^{iu} - 1)\, du \right| \leqslant \frac{y^2}{2};$$

третье — из второго:

$$\left| e^{iy} - 1 - iy + \frac{y^2}{2} \right| = \left| \int_0^{iy} (e^z - 1 - z)\,dz \right| =$$

$$= \left| i \int_0^y (e^{iu} - 1 - iu)\,du \right| \leqslant \frac{|y|^3}{6};$$

и т. д.

II тип:

$$|e^{iy} - 1| \leqslant 2, \tag{76}$$

$$|e^{iy} - 1 - iy| \leqslant 2|y|, \tag{77}$$

$$\left| e^{iy} - 1 - iy + \frac{y^2}{2} \right| \leqslant y^2, \tag{78}$$

и т. д.

Первое из этих неравенств непосредственно вытекает из равенства $|e^{iy}| = 1$, второе — из первого неравенства I типа, третье — из второго неравенства I типа; и т. д.

88. В качестве примера применения предельных теорем приведем вывод интересной связи между законом Гаусса и законом Пуассона.

Пусть ξ — случайная величина, подчиняющаяся закону Пуассона:

$$\frac{k^x}{x!} e^{-k},$$

где $k > 0$ и $x = 0, 1, 2, \ldots$ Рассмотрим нормированное отклонение $\bar{\xi} = \frac{\xi - k}{\sqrt{k}}$ величины ξ и покажем, что *при $k \to \infty$ закон распределения этого нормированного отклонения стремится к закону Гаусса:*

$$\frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{z^2}{2}}\,dz. \tag{G}$$

В самом деле, характеристическая функция величины ξ есть

$$\boldsymbol{E} e^{it\xi} = \sum_{x=0}^{\infty} e^{itx} \frac{k^x}{x!} e^{-k} = e^{-k} \sum_{x=0}^{\infty} \frac{(ke^{it})^x}{x!} = e^{-k} e^{ke^{it}} = e^{k(e^{it}-1)}.$$

Следовательно, в силу теоремы XVII характеристическая функция нормированного отклонения $\bar{\xi}$ есть

$$e^{k\left(e^{\frac{it}{\sqrt{k}}} - 1 - \frac{it}{\sqrt{k}}\right)}. \tag{79}$$

Характеристическая же функция величины, подчиняющейся закону Гаусса (G), есть

$$e^{-\frac{t^2}{2}}. \tag{80}$$

Таким образом в силу обратной предельной теоремы, для того чтобы закон распределения нормированного отклонения $\overline{\xi}$ стремился к закону Гаусса (G), достаточно, чтобы функция (79) сходилась к функции (80) или, что то же, чтобы логарифм функции (79) сходился к логарифму функции (80):

$$k\left(e^{\frac{it}{\sqrt{k}}}-1-\frac{it}{\sqrt{k}}\right)\to-\frac{t^2}{2}.$$

Чтобы убедиться в том, что это соотношение действительно имеет место, напишем его в таком виде:

$$k\left(e^{\frac{it}{\sqrt{k}}}-1-\frac{it}{\sqrt{k}}+\frac{t^2}{2k}\right)\to 0.$$

Мы знаем, что при действительном y всегда

$$\left|e^{ity}-1-iy+\frac{y^2}{2}\right|\leqslant\frac{|y|^3}{6}.$$

Поэтому, полагая $y=\frac{t}{\sqrt{k}}$, мы имеем

$$k\left|e^{\frac{it}{\sqrt{k}}}-1-\frac{it}{\sqrt{k}}+\frac{t^2}{2k}\right|\leqslant\frac{|t|^3}{6\sqrt{k^3}}.$$

Это и показывает, что соотношение, о котором идет речь, действительно имеет место при $k\to\infty$.

Доказанное предложение является частным случаем теоремы А. Н. Колмогорова, согласно которой *для всякой случайной величины* ξ, *изменяющейся со временем и процесс изменения которой есть однородный (случайный) процесс с конечной дисперсией в каждый момент времени* t, *закон распределения нормированного отклонения* $\overline{\xi}$ *при* $t\to\infty$ *стремится к закону Гаусса* (G). А мы видели, что закон Пуассона есть закон распределения как раз одной из таких случайных величин с дисперсией k, пропорциональной t.

ГЛАВА VIII.

ПРЕДЕЛЬНЫЕ ЗАКОНЫ РАСПРЕДЕЛЕНИЯ.

§ 25. Закон Гаусса как предельный закон распределения.

89. В гл. III мы доказали теорему Лапласа, согласно которой при $n \to \infty$

$$P\left(\frac{\mu - np}{\sqrt{npq}} < x\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz,$$

где μ есть число наступлений события при n независимых испытаниях, p — вероятность его наступления при каждом отдельном испытании и $q = 1 - p$. Этой теореме можно придать несколько иную форму, если заметить, — как мы это уже неоднократно делали, — что случайную величину μ можно рассматривать, как сумму n независимых случайных величин:

$$\mu = \mu_1 + \mu_2 + \ldots + \mu_n,$$

где μ_k есть число наступлений события при одном k-ом испытании, и что

$$np = E\mu = E\mu_1 + E\mu_2 + \ldots + E\mu_n,$$
$$npq = D\mu = D\mu_1 + D\mu_2 + \ldots + D\mu_n.$$

Что касается случайных величин μ_k, то каждая из них может принимать только два значения: 1 с постоянной вероятностью p и 0 с постоянной вероятностью q, — и очевидно, что теорема остается справедливой для суммы таких величин, независимо от связи с конкретной схемой независимых испытаний. Таким образом, теорему Лапласа можно формулировать следующим образом:

Если независимые случайные величины $\mu_1, \mu_2, \ldots, \mu_n, \ldots$ *принимают каждая только два значения* 1 *и* 0, *и то и другое с постоянной вероятностью, отличной от нуля и единицы, то при* $n \to \infty$

$$P\left(\frac{\sum_{k=1}^{n} \mu_k - \sum_{k=1}^{n} E\mu_k}{\sqrt{\sum_{k=1}^{n} D\mu_k}} < x\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz.$$

Естественно возникает вопрос, останется ли в силе эта теорема, если заменить величины μ_k величинами ξ_k, подчиненными менее стеснительным условиям. В решении этого вопроса заинтересована математическая статистика, а также ряд отделов физики, техники и биологии.

Теоретический интерес и практическое значение обобщения теоремы Лапласа на суммы различных независимых случайных величин были осознаны уже давно. Точную постановку задачи о законе Гаусса, как предельном законе распределения для сумм независимых величин, и указания путей к ее решению мы находим у Чебышева. Воспользовавшись этими указаниями, Марков решил задачу в довольно общих условиях, но в предположении о существовании у данных величин ξ_k моментов всех порядков. После этого, во многих отношениях интересным методом и в более общих условиях, ее решил Ляпунов.

Завершено было исследование и решение задачи о законе Гаусса, как предельном законе для сумм независимых величин, в наше время. Ниже мы познакомимся с наиболее простыми и важными современными результатами.

90. Возможность обобщения теоремы Лапласа на суммы различных независимых случайных величин и все практическое значение этого обобщения легко немедленно иллюстрировать на примере следующей теоремы [1]):

Теорема Леви́. — *Если независимые случайные величины ξ_1, $\xi_2, \ldots, \xi_n, \ldots$ ограничены в их совокупности и при $n \to \infty$ дисперсия суммы $\xi_1 + \xi_2 + \ldots + \xi_n$ стремится к бесконечности, то при $n \to \infty$*

$$\boldsymbol{P}\left(\frac{\sum\limits_{k=1}^{n} \xi_k - \sum\limits_{k=1}^{n} \boldsymbol{E}\xi_k}{\sqrt{\sum\limits_{k=1}^{n} \boldsymbol{D}\xi_k}} < x\right) \to \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{z^2}{2}}\, dz.$$

Независимость случайных величин $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ здесь следует понимать в том смысле, что каждая из этих величин, ξ_n, независима от суммы ей предшествующих, $\xi_1 + \xi_2 + \ldots + \xi_{n-1}$.

Доказательство. — 1° Введем вспомогательные случайные величины:

$$\xi_{nk} = \frac{\xi_k - \boldsymbol{E}\xi_k}{\sqrt{\sum\limits_{k=1}^{n} \boldsymbol{D}\xi_k}}.$$

Нетрудно подсчитать, что

$$\boldsymbol{E}\xi_{nk} = 0, \quad \sum_{k=1}^{n} \boldsymbol{E}\xi_{nk}^2 = 1.$$

---

1) Мы условно называем эту теорему теоремой Леви́, так как доказываем ее предложенным им методом. На самом деле эта теорема является частным случаем более общего предложения, доказанного много раньше Ляпуновым (см. ниже в п. 97 и 98).

По условию случайные величины ξ_k ограничены в их совокупности: $|\xi_k| < C$. Положим

$$C_n = \frac{2C}{\sqrt{\sum_{k=1}^{n} D\xi_k}}.$$

Нетрудно убедиться в том, что

$$|\xi_{nk}| < C_n, \quad C_n \to 0.$$

Нетрудно видеть, наконец, что доказать формулированную теорему — то же самое, что доказать следующее предложение: вероятность

$$\text{(P)} \qquad P\left(\sum_{k=1}^{n} \xi_{nk} < x\right)$$

при $n \to \infty$ стремится к

$$\text{(G)} \qquad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz.$$

2° Обозначим характеристическую функцию величины ξ_{nk} через $\varphi_{nk}(t)$. Тогда характеристическая функция суммы

$$\sum_{k=1}^{n} \xi_{nk},$$

имеющей функцию распределения (P), будет равна произведению

$$\prod_{k=1}^{n} \varphi_{nk}(t).$$

Что же касается характеристической функции величины, подчиняющейся закону Гаусса (G), то она, как мы знаем, равна

$$e^{-\frac{t^2}{2}}.$$

В силу теорем о характеристических функциях, доказать формулированную теорему или, что то же, только что приведенное равносильное ей предложение о функциях (P) и (G) — то же самое, что доказать следующее предложение: при $n \to \infty$

$$\text{(X)} \qquad \prod_{k=1}^{n} \varphi_{nk}(t) \to e^{-\frac{t^2}{2}}.$$

3° Это последнее предложение мы сейчас и докажем. При этом для простоты вычислений мы введем логарифмы характеристических функций $\psi_{nk}(t) = \log \varphi_{nk}(t)$, так что

$$\log \prod_{k=1}^{n} \varphi_{nk}(t) = \sum_{k=1}^{n} \psi_{nk}(t),$$

и установим, очевидно, равносильное соотношению (X) соотношение

$$\text{(Y)} \qquad \sum_{k=1}^{n} \psi_{nk}(t) \to -\frac{t^2}{2}.$$

С этой целью разложим функции $\psi_{nk}(t)$ по формуле Маклорена:

$$\psi_{nk}(t) = \psi_{nk}(0) + \psi'_{nk}(0)\,t + \psi''_{nk}(0)\,\frac{t^2}{2} + \psi'''_{nk}(\theta)\,\frac{t^3}{6},$$

где $0 < |\theta| < |t|$. Это разложение возможно, так как, в силу ограниченности величин ξ_{nk}, существуют их моменты всех порядков, следовательно, существуют и производные их характеристических функций всех порядков.

Сверх того, как мы знаем:

$$\psi_{nk}(0) = 0, \quad \psi'_{nk}(0) = i\boldsymbol{E}\xi_{nk} = 0, \quad \psi''_{nk}(0) = -\boldsymbol{E}\xi^2_{nk}.$$

Отсюда мы имеем

$$\sum_{k=1}^{n} \psi_{nk}(t) = -\frac{t^2}{2} + \frac{t^3}{6}\sum_{k=1}^{n} \psi'''_{nk}(\theta).$$

Следовательно:

$$\left|\sum_{k=1}^{n} \psi_{nk}(t) + \frac{t^2}{2}\right| \leqslant \frac{|t|^3}{6}\sum_{k=1}^{n} |\psi'''_{nk}(\theta)|.$$

Таким образом соотношение (Y), которое мы хотим установить, будет установлено, если мы покажем, что для всех t

$$\sum_{k=1}^{n} |\psi'''_{nk}(\theta)| \to 0.$$

4° Остается показать это последнее. Для этого мы вычислим производные $\psi'''_{nk}(t)$ и оценим их модули.

Так как $\psi_{nk}(t) = \log \varphi_{nk}(t)$, то

$$\psi'''_{nk}(t) = 2\left(\frac{\varphi'_{nk}(t)}{\varphi_{nk}(t)}\right)^3 - 3\,\frac{\varphi'_{nk}(t)\,\varphi''_{nk}(t)}{(\varphi_{nk}(t))^2} + \frac{\varphi'''_{nk}(t)}{\varphi_{nk}(t)}.$$

Поэтому

$$\text{(81)} \qquad |\psi'''_{nk}(t)| \leqslant 2\,\frac{|\varphi'_{nk}(t)|\,|\varphi'_{nk}(t)|^2}{|\varphi_{nk}(t)|^3} + 3\,\frac{|\varphi'_{nk}(t)|\,|\varphi''_{nk}(t)|}{|\varphi_{nk}(t)|^2} + \frac{|\varphi'''_{nk}(t)|}{|\varphi_{nk}(t)|}.$$

Что касается величин $|\varphi_{nk}(t)|$, $|\varphi'_{nk}(t)|$, $|\varphi'_{nk}(t)|^2$, $|\varphi''_{nk}(t)|$ и $|\varphi'''_{nk}(t)|$, то для них нетрудно получить ряд неравенств, которые приведут нас к цели. Пусть $F_{nk}(x)$ — функция распределения величины ξ_{nk}. Очевидно, что $F_{nk}(x) = 0$ при $x < -C_n$ и $F_{nk}(x) = 1$ при $x > C_n$. Далее, так как при всяком действительном y мы имеем

$$|e^{iy} - 1| \leqslant |y|,$$

то, полагая $y = tx$, мы имеем также

$$|\varphi_{nk}(t) - 1| = \left| \int_{-C_n}^{C_n} (e^{itx} - 1)\, dF_{nk}(x) \right| \leqslant C_n |t|.$$

Поэтому

$$|\varphi_{nk}(t)| \geqslant 1 - C_n |t|. \tag{82}$$

Затем, для всякого целого положительного p мы имеем для производных функции $\varphi(t)$ порядка p:

$$|\varphi_{nk}^{(p)}(t)| = \left| \int_{-C_n}^{C_n} x^p e^{itx} dF_{nk}(x) \right| \leqslant \int_{-C_n}^{C_n} |x|^p dF_{nk}(x).$$

Применяя это неравенство последовательно при $p = 1$, 1, 2 и 3, мы имеем [1])

$$|\varphi'_{nk}(t)| \leqslant 2C_n, \tag{83}$$

$$|\varphi'_{nk}(t)|^2 \leqslant \int_{-C_n}^{C_n} x^2 dF_{nk}(x), \tag{84}$$

$$|\varphi''_{nk}(t)| \leqslant \int_{-C_n}^{C_n} x^2 dF_{nk}(x), \tag{85}$$

$$|\varphi'''_{nk}(t)| \leqslant 2C_n \int_{-C_n}^{C_n} x^2 dF_{nk}(x). \tag{86}$$

Сопоставим теперь неравенства (81), (82), (83), (84), (85) и (86) и примем во внимание, что

$$\sum_{k=1}^{n} \int_{-C_n}^{C_n} x^2 dF_{nk}(x) = 1.$$

Мы получим

$$\sum_{k=1}^{n} |\psi'''_{nk}(t)| \leqslant 2C_n \left[\frac{2}{(1 - C_n|t|)^3} + \frac{3}{(1 - C_n|t|)^2} + \frac{1}{1 - C_n|t|} \right].$$

Так как $C_n \to 0$, то из этого последнего неравенства видно, что сумма

$$\sum_{k=1}^{n} |\psi'''_{nk}(\theta)|,$$

где $0 < |\theta| < |t|$, действительно, стремится к нулю для всех t.

[1]) При выводе неравенства (84) мы пользуемся неравенством Шварца: $(\mathbf{E}|\xi_{nk}|)^2 \leqslant \mathbf{E}\xi_{nk}^2$.

Тем самым наша теорема доказана.

91. Практическое значение только что доказанной теоремы легко будет иллюстрировать рядом примеров. Для этого, однако, целесообразно будет придать теореме несколько иную форму.

Дело в том, что утверждение теоремы может быть записано в виде

$$\boldsymbol{P}\left(\frac{\sum\limits_{k=1}^{n}\xi_k-\sum\limits_{k=1}^{n}\boldsymbol{E}\xi_k}{\sqrt{\sum\limits_{k=1}^{n}\boldsymbol{D}\xi_k}}<x\right)-\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x}e^{-\frac{z^2}{2}}dz\to 0,$$

причем *это соотношение имеет место равномерно на всей прямой* $-\infty<x<+\infty$; последнее непосредственно следует из леммы о равномерной сходимости приведенной в п. 36. Отсюда, в свою очередь, следует, что названное соотношение остается в силе и в том случае, если x зависит от n, какова бы ни была эта зависимость. Положим же $x=\frac{t-A_n}{\sqrt{B_n}}$ и сделаем попутно замену переменного под знаком интеграла, $z=\frac{v-A_n}{\sqrt{B_n}}$, где

$$A_n=\sum_{k=1}^{n}\boldsymbol{E}\xi_k,\quad B_n=\sum_{k=1}^{n}\boldsymbol{D}\xi_k.$$

Мы получим следующее соотношение для сумм случайных величин:

$$\boldsymbol{P}\left(\sum_{k=1}^{n}\xi_k<t\right)-\frac{1}{\sqrt{2\pi B_n}}\int\limits_{-\infty}^{t}e^{-\frac{(v-A_n)^2}{2B_n}}dv\to 0.$$

Отсюда видно, что, вообще говоря, всякая случайная величина, которая складывается из очень большого числа незначительных, независимых друг от друга случайных величин, должна иметь закон распределения, близкий к закону Гаусса с теми или другими параметрами α и β. Приведем два типичных примера.

Пример 1. — *Ошибки наблюдений.* Когда мы производим какое-нибудь наблюдение с целью измерить ту или иную физическую постоянную, то на наше наблюдение неизбежно влияет громадное количество факторов, которые мы не в состоянии учесть и которые порождают ошибки в измерении. Сюда относятся ошибки, вызываемые состоянием измерительного прибора, которое может нечувствительно меняться под влиянием различных атмосферных воздействий или незаметных механических причин. Сюда относятся личные ошибки наблюдателя, вызываемые особенностями его зрения или слуха и также могущие нечувствительно меняться в зависимости от физического или психического состояния наблюдателя, и т. д. Каждый из этих факторов сам по себе породил бы только ничтожную „элементарную ошибку". Но на измерении сказывается сразу вся сумма этих „элементарных ошибок". Иначе говоря, фактическая ошибка наблюдения будет случайной величиной, которая является суммой громадного числа незначительных, независимых друг от друга случайных величин. И хотя эти последние неизвестны и изменчивы по абсолютной величине и по знаку, их сумма в силу только что установленной теоремы должна иметь закон распределения, близкий к закону Гаусса.

Пример 2. — *Массовое производство.* В процессе массового производства, существующего во многих отраслях промышленности, выделываются большие количества одинаковых штук того или иного продукта. Обратим внимание на какую-нибудь числовую характеристику одного из таких продуктов. Поскольку продукт производится в соответствии с определенными техническими нормами, существует некоторая „нормальная" величина, которую должна была бы иметь интересующая нас числовая характеристика у каждой штуки данного продукта. На деле, однако, всегда наблюдаются некоторые отклонения от этой „нормальной" величины. При правильно поставленном производственном процессе такие отклонения могут вызываться только случайными причинами, каждая из которых производит лишь незаметный эффект. Но суммарным эффектом всех таких причин может быть уже заметное случайное отклонение от нормы. В силу только что установленной теоремы это отклонение должно иметь закон распределения, близкий к закону Гаусса.

Подобные примеры можно было бы умножать до бесконечности. Случайные отклонения от нормы, вызванные наложением друг на друга большого числа незначительных, независимых между собой случайных отклонений, встречаются повсюду в природе и в технике; благодаря этому закон Гаусса очень часто встречается в приложениях.

92. Переходя к дальнейшему обобщению теоремы Лапласа, мы начнем с того, что обобщим самую схему, к которой применяется теорема. А именно, мы будем рассматривать не последовательность случайных величин:

$$(s) \qquad \xi_1,\ \xi_2,\ \ldots,\xi_n,\ \ldots,$$

а систему, которая представляет собой последовательность серий случайных величин:

$$(S) \qquad \begin{array}{c} \xi_{11},\ \xi_{12},\ \ldots,\ \xi_{1k_1}, \\ \xi_{21},\ \xi_{22},\ \ldots,\ \xi_{2k_2}, \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ \xi_{n1},\ \xi_{n2},\ \ldots,\ \xi_{nk_n}, \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \end{array}$$

С подобной последовательностью серий случайных величин мы уже встречались в теореме Пуассона (п. 43).

Основной побудительной причиной для того, чтобы рассматривать системы (S) вместо последовательностей (s), исторически послужило следующее обстоятельство. Только что, в доказательстве теоремы Леви, мы фактически свели вопрос о последовательности величин ξ_n к вопросу о последовательности серий величин ξ_{nk}, и это внесло в доказательство теоремы значительное упрощение. Так же было бы целесообразно поступать и при дальнейшем обобщении теоремы Лапласа. Следовательно, проще прямо ставить вопрос о последовательностях серий.

При этом мы заведомо ничего не потеряем в общности результатов, а, напротив, только выиграем. В случае последовательностей (s) речь должна итти о суммах

$$(\sigma) \qquad \xi_1 + \xi_2 + \ldots + \xi_n,$$

а в случае систем (S) — о суммах

$$(\Sigma) \qquad \xi_{n1} + \xi_{n2} + \ldots + \xi_{nk_n}.$$

Но нетрудно видеть, что суммы (σ) представляют собой частный случай сумм (Σ). В самом деле, пусть дана последовательность (s). Составим следующую систему (S):

$$\begin{array}{l} \xi_1, \\ \xi_1, \ \xi_2, \\ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \\ \xi_1, \ \xi_2, \ \ldots, \ \xi_n, \\ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \end{array}$$

т. е. систему величин $\xi_{nk} = \xi_k$, где $k_n = n$. Ясно, что суммы (Σ) для этой системы — не что иное, как суммы (σ) для последовательности (s). Поэтому всякую теорему о суммах, доказанную для систем (S), можно будет применять, как к частному случаю, и к последовательностям (s).

Обратное же было бы неверно. Известно, например, что закон Пуассона, вообще говоря, не может быть пределом для сумм независимых случайных величин вида (σ). Но он может быть таковым для сумм вида (Σ), что мы видели в п. 43.

С другой стороны, и в приложениях суммы вида (Σ) часто лучше отражают излучаемое явление, нежели суммы вида (σ). Например, в теореме Леви́ величины ξ_n были ограничены в их совокупности, но в сумме могли безгранично возрастать, тогда как величины ξ_{nk} были равномерно бесконечно малы, а в сумме имели во всяком случае ограниченную дисперсию. В приложении же хотя бы к ошибкам наблюдений отдельные случайные слагаемые как раз бывают очень малы, сумма же их ограничена. Таким образом естественно рассматривать отдельные случайные слагаемые, из которых составляется ошибка наблюдения, именно как величины ξ_{nk}, а не как величины ξ_n, и понимать здесь утверждение теоремы в том смысле, что чем меньше слагаемые, тем ближе распределение суммы к гауссовскому.

93. В дальнейшем мы ограничимся системами (S) независимых случайных величин, которые мы назовем *элементарными системами*. Это — системы

$$\begin{array}{l} \xi_{11}, \ \xi_{12}, \ \ldots, \ \xi_{1k_1}, \\ \xi_{21}, \ \xi_{22}, \ \ldots, \ \xi_{2k_2}, \\ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \\ \xi_{n1}, \ \xi_{n2}, \ \ldots, \ \xi_{nk_n}, \\ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \end{array}$$

где все величины ξ_{nk} имеют средние значения, равные нулю, и конечные дисперсии b_{nk}, равномерно стремящиеся к нулю при $n \to \infty$; независимость этих величин всегда будет пониматься в том смысле, что каждая из них, ξ_{nk}, независима от суммы ей предшествующих с тем же индексом n, т. е. от $\xi_{n1} + \xi_{n2} + \ldots + \xi_{nk-1}$.

Для вопроса о предельном законе распределения для сумм $\xi_n = \xi_{n1} + \xi_{n2} + \ldots + \xi_{nk_n}$ единственным существенным ограничением, содержащимся в понятии элементарной системы, является требование о существовании конечных дисперсий b_{nk}.

Что касается требования, чтобы дисперсии b_{nk} равномерно стремились к нулю при $n \to \infty$, то это добавочное ограничение неизбежно по сути самого поставленного вопроса. В самом деле, нам интересно выяснить, как влияет на закон распределения величины ξ_n тот факт, что она является суммой большого числа независимых слагаемых ξ_{nk}. И мы, очевидно, отнюдь не приблизились бы к этому выяснению, а, напротив, отдалились бы от него, если бы допустили присутствие среди слагаемых ξ_{nk} таких величин, которые сами по себе оказывали бы заметное влияние на закон распределения суммы ξ_n, затемняя тем самым влияние совокупности остальных слагаемых. Но как раз так и обстояло бы дело, если бы величины ξ_{nk} обладали конечными дисперсиями b_{nk}, не стремящимися равномерно к нулю при $n \to \infty$. В этом случае в составе сумм ξ_n время от времени появлялось бы по крайней мере по одному слагаемому ξ_{nl_n}, дисперсия которого b_{nl_n} превышала бы некоторое определенное число $b > 0$. Если теперь представить себе, что мы удалили из состава суммы ξ_n такое слагаемое ξ_{nl_n}, то дисперсия суммы ξ_n^* оставшихся слагаемых в силу независимости слагаемых отличалась бы от дисперсии первоначальной суммы ξ_n по крайней мере на это число b. Следовательно, законы распределения сумм ξ_n^* отличались бы от законов распределения первоначальных сумм ξ_n так, что это отличие не исчезало бы с возрастанием n. Слагаемые ξ_{nl_n} оказывали бы исключительное влияние на закон распределения сумм ξ_n, затемняющее влияние совокупности остальных слагаемых.

Мы будем называть систему (S) независимых случайных величин *нормированной системой*, если сумма дисперсий этих величин при каждом данном n равна единице:

$$\sum_{k=1}^{k_n} b_{nk} = 1.$$

При наличии конечных дисперсий у всех величин данной системы эту последнюю, разумеется, всегда можно нормировать простым линейным преобразованием принадлежащих ей величин. При исследовании закона Гаусса предварительная нормировка систем рассматриваемых величин приводит к наиболее простым и ясным формулировкам теорем.

Для примера укажем, что система величин ξ_{nk} в доказательстве теоремы Леви была нормированной элементарной системой.

94. Лемма Г. М. Бавли. — *Пусть дана система (S) независимых случайных величин, пусть $\varphi_{nk}(t)$ — характеристическая функция величины ξ_{nk} и пусть при $n \to \infty$*

$$\sum_{k=1}^{k_n} |\varphi_{nk}(t) - 1|^2 \to 0.$$

Тогда для того, чтобы функция распределения суммы

$$\xi_n = \xi_{n1} + \xi_{n2} + \ldots + \xi_{nk_n}$$

сходилась в основном к функции распределения случайной величины, характеристическая функция которой есть $\varphi(t)$, необходимо и достаточно, чтобы имело место соотношение: при $n \to \infty$

$$\sum_{k=1}^{k_n} [\varphi_{nk}(t) - 1] \to \log \varphi(t).$$

Доказательство. — В силу теорем о характеристических функциях, функция распределения суммы ξ_n сходится в основном к функции распределения случайной величины, характеристическая функция которой есть $\varphi(t)$, тогда и только тогда, когда

$$\prod_{k=1}^{k_n} \varphi_{nk}(t) \to \varphi(t),$$

или, что то же, когда

$$\sum_{k=1}^{k_n} \log \varphi_{nk}(t) \to \log \varphi(t).$$

Поэтому для доказательства леммы достаточно убедиться в том, что в условиях леммы разность

$$R_n(t) = \sum_{k=1}^{k_n} \log \varphi_{nk}(t) - \sum_{k=1}^{k_n} [\varphi_{nk}(t) - 1]$$

стремится к нулю.

Положим

$$r_{nk}(t) = \log \varphi_{nk}(t) - [\varphi_{nk}(t) - 1],$$

так что

$$|R_n(t)| \leqslant \sum_{k=1}^{k_n} |r_{nk}(t)|.$$

Мы имеем

$$r_{nk}(t) = \log \{1 + [\varphi_{nk}(t) - 1]\} - [\varphi_{nk}(t) - 1] =$$

$$= -\frac{1}{2}[\varphi_{nk}(t) - 1]^2 + \frac{1}{3}[\varphi_{nk}(t) - 1]^3 - \ldots$$

и потому для всех n, для которых $\sum_{k=1}^{k_n} |\varphi_{nk}(t)-1|^2 \leqslant \frac{1}{4}$ и, следовательно, $|\varphi_{nk}(t)-1| \leqslant \frac{1}{2}$:

$$|r_{nk}(t)| \leqslant \frac{1}{2}|\varphi_{nk}(t)-1|^2 + \frac{1}{3}|\varphi_{nk}(t)-1|^3 + \ldots \leqslant$$
$$\leqslant \frac{1}{2}|\varphi_{nk}(t)-1|^2\{1+|\varphi_{nk}(t)-1|+\ldots\} \leqslant |\varphi_{nk}(t)-1|^2,$$

так что

$$|R_n(t)| \leqslant \sum_{k=1}^{k_n} |\varphi_{nk}(t)-1|^2.$$

Отсюда видно, что $R_n(t)$, действительно, стремится к нулю. Лемма доказана.

Для нас будет важно, что *условие леммы Бавли выполнено в нормированных элементарных системах*. В самом деле, пусть $F_{nk}(x)$ — функция распределения величины ξ_{nk}. В элементарной системе мы имеем

$$\int_{-\infty}^{+\infty} x\,dF_{nk}(x) = 0.$$

Поэтому

$$\varphi_{nk}(t)-1 = \int_{-\infty}^{+\infty} (e^{itx}-1)\,dF_{nk}(x) = \int_{-\infty}^{+\infty} (e^{itx}-1-itx)\,dF_{nk}(x).$$

Затем, при действительном y всегда

$$|e^{iy}-1-iy| \leqslant \frac{y^2}{2}.$$

Поэтому, полагая $y=tx$, мы имеем

$$|\varphi_{nk}(t)-1| \leqslant \int_{-\infty}^{+\infty} \frac{t^2x^2}{2}\,dF_{nk}(x) = \frac{t^2}{2}\int_{-\infty}^{+\infty} x^2 dF_{nk}(x).$$

Но

$$\int_{-\infty}^{+\infty} x^2 dF_{nk}(x) = b_{nk}.$$

Таким образом, если обозначить через $\overline{b}_n$ наибольшую из дисперсий $b_{n1}, b_{n2}, \ldots, b_{nk_n}$:

$$\sum_{k=1}^{k_n} |\varphi_{nk}(t)-1|^2 \leqslant \frac{t^4}{4}\sum_{k=1}^{k_n} b_{nk}^2 \leqslant \frac{t^4}{4}\overline{b}_n \sum_{k=1}^{k_n} b_{nk}.$$

В элементарной же системе $\overline{b}_n \to 0$, а в нормированной $\sum_{k=1}^{k_n} b_{nk} = 1$. Отсюда и вытекает условие леммы

$$\sum_{k=1}^{k_n} |\varphi_{nk}(t) - 1|^2 \to 0.$$

95. Пользуясь только что доказанной леммой, для нормированных элементарных систем будет нетрудно установить необходимое и достаточное условие для того, чтобы функция распределения суммы $\xi_{n1} + \xi_{n2} + \ldots + \xi_{nk_n}$ сходилась к закону Гаусса. Как мы увидим ниже, таким условием будет следующее:

Условие Линдеберга. — *Для любой положительной постоянной τ при $n \to \infty$*

$$\text{(L)} \qquad \sum_{k=1}^{k_n} \int\limits_{|x|>\tau} x^2 dF_{nk}(x) \to 0,$$

где $F_{nk}(x)$ — функция распределения величины ξ_{nk}.

Пояснение. — Смысл условия Линдеберга может быть лучше всего выяснен на одном частном случае. А именно, пусть все величины ξ_{nk} с одним и тем же индексом n имеют одну и ту же функцию распределения $F_{nk}(x) = F_{(n)}(x)$. Тогда в нормированной элементарной системе

$$(87) \qquad n \int\limits_{-\infty}^{+\infty} x^2\, dF_{(n)}(x) = 1,$$

а условие Линдеберга имеет вид

$$(88) \qquad n \int\limits_{|x|>\tau} x^2\, dF_{(n)}(x) \to 0.$$

Это значит, что при больших n малые значения величин ξ_{nk} гораздо более вероятны, нежели большие значения. В самом деле, если при больших n интеграл (88) оказывается значительно меньше интеграла (87), то причиной этого может быть только то, что под знаком интеграла сравнительно незначительную роль играют большие значения x, так что решающая роль принадлежит малым значениям x. А это может происходить только от того, что значения диференциала $dF_{(n)}(x)$ особенно велики при малых значениях x. Но диференциал $dF_{(n)}(x)$, очевидно, характеризует вероятность тех значений x, которые лежат в ближайших окрестностях точки x. Таким образом, действительно, при больших n вероятности малых значений величины ξ_{nk} значительно превосходят вероятности ее больших значений.

96. Теорема Линдеберга — Феллера. — *Пусть дана нормированная элементарная система независимых случайных величин.*

Тогда для того, чтобы функция распределения суммы $\xi_{n1}+\xi_{n2}+$ $+\ldots+\xi_{nk_n}$ *сходилась к закону Гаусса*

$$\text{(G)} \qquad \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-z^2}dz,$$

необходимо и достаточно, чтобы было выполнено условие Линдеберга (L).

Линдебергом была установлена достаточность условия (L), Феллером — его необходимость.

Мы докажем и то и другое методом, предложенным Б. В. Гнеденко и основанным на преобразовании интеграла Стильтьеса, примененном ранее по другому поводу А. Н. Колмогоровым.

1° Прежде всего заметим, что, как мы показали в п. 94, в нормированной элементарной системе выполнено условие леммы Бавли. Таким образом здесь функция распределения суммы $\xi_{n1}+\xi_{n2}+$ $+\ldots+\xi_{nk_n}$ сходится к закону Гаусса (G) тогда и только тогда, когда

$$\sum_{k=1}^{k_n}[\varphi_{nk}(t)-1]\to-\frac{t^2}{2},$$

где $\varphi_{nk}(t)$ — характеристическая функция величины ξ_{nk}.

Преобразуем это выражение, введя некоторые вспомогательные функции. А именно, положим

$$E_n(x)=\sum_{k=1}^{k_n}\int\limits_{-\infty}^{x}u^2\,dF_{nk}(u).$$

Очевидно, что $E_n(x)$ — неубывающая функция. Сверх того, $E_n(-\infty)=0$ и $E_n(+\infty)=1$. Так как в нормированной элементарной системе

$$\varphi_{nk}(t)-1=\int\limits_{-\infty}^{+\infty}(e^{itx}-1-itx)\,dF_{nk}(x),$$

то, как нетрудно видеть [1]):

$$\text{(89)} \qquad \sum_{k=1}^{k_n}[\varphi_{nk}(t)-1]=\int\limits_{-\infty}^{+\infty}(e^{itx}-1-itx)\frac{1}{x^2}\,dE_n(x).$$

С другой стороны, положим

$$E(x)=\begin{cases}0 \text{ при } x<0,\\ 1 \text{ при } x>0.\end{cases}$$

---

[1]) На основании свойств 6) и 7) интеграла Стильтьеса, приведенных в п. 54.

Простое вычисление показывает, что[1])

$$(90)\qquad -\frac{t^2}{2}=\int_{-\infty}^{+\infty}(e^{itx}-1-itx)\frac{1}{x^2}\,dE(x).$$

Таким образом функция распределения суммы $\xi_{n1}+\xi_{n2}+\ldots+\xi_{nk_n}$ сходится к закону Гаусса (G) тогда и только тогда, когда интеграл (89) стремится к интегралу (90) при любом t.

Заметим, что функция

$$(e^{itx}-1-itx)\frac{1}{x^2}$$

под знаком интегралов (89) и (90) непрерывна и, в силу неравенства $|e^{itx}-1-itx|\leqslant\frac{x^2}{2}$, ограничена на всей прямой $-\infty<x<+\infty$.

2° Заметим еще, что уравнение

$$(\Delta)\qquad \varphi(t)=\int_{-\infty}^{+\infty}(e^{itx}-1-itx)\frac{1}{x^2}\,dF(x)$$

однозначно определяет функцию распределения $F(x)$ по заданной функции $\varphi(t)$.

Чтобы в этом убедиться, достаточно дважды продифференцировать функцию $\varphi(t)$; мы получим

$$\varphi''(t)=-\int_{-\infty}^{+\infty}e^{itx}\,dF(x).$$

Но если задана функция $\varphi(t)$, то задана и ее вторая производная $\varphi''(t)$. Последняя же, как это видно из ее выражения, является, с точностью до знака, характеристической функцией для функции распределения $F(x)$; следовательно, как мы знаем из теории характеристических функций, и эта последняя вполне определена.

3° Достаточность условия (L). — Пусть выполнено условие (L). Так как

$$\sum_{k=1}^{k_n}\int_{|x|>\tau}x^2\,dF_{nk}(x)=1-E_n(\tau)+E_n(-\tau),$$

то это значит, что, каково бы ни было $\tau>0$,

$$E_n(\tau)-E_n(-\tau)\to 1.$$

Но $0\leqslant E_n(-\tau)\leqslant E_n(\tau)\leqslant 1$. Отсюда ясно, что, каково бы ни было $\tau>0$, $E_n(-\tau)\to 0$ и $E_n(\tau)\to 1$. Иначе говоря, $E_n(x)$ сходится

---

[1]) Так как интеграл с интегрирующей функцией $E(x)$, очевидно, равен значению подынтегральной функции в точке 0; это же последнее в данном случае, как видно из разложения e^{itx} в степенной ряд, равно $-\frac{t^2}{2}$.

в основном к $E(x)$. Таким образом интегралы (89) и (90) при любом t удовлетворяют условиям второй теоремы Хелли. Следовательно, интеграл (89), действительно, стремится к интегралу (90) при любом t.

4° Необходимость условия (L). — Пусть теперь интеграл (89) при любом t стремится к интегралу (90). В силу первой теоремы Хелли последовательность

$$(S) \qquad E_1(x),\ E_2(x),\ \ldots,\ E_n(x),\ \ldots$$

содержит подпоследовательность

$$(S') \qquad E_{n_1}(x),\ E_{n_2}(x),\ \ldots,\ E_{n_k}(x)\ \ldots,$$

которая сходится в основном к некоторой функции $E_\infty(x)$. В силу второй теоремы Хелли интегралы вида (89) для этой подпоследовательности (S') стремятся к интегралу вида (90), где на месте функции $E(x)$ стоит функция $E_\infty(x)$. Таким образом интеграл (90) и такой же интеграл вида (90), но где на месте $E(x)$ стоит $E_\infty(x)$, будучи пределом одних и тех же интегралов, равны между собой. В силу единственности решения $F(x)$ уравнения (Δ) отсюда вытекает, что $E_\infty(x) = E(x)$. Итак, подпоследовательность (S') сходится в основном к функции $E(x)$.

Но тогда и вся последовательность (S) сходится в основном к функции $E(x)$. В самом деле, если бы это было не так, то нашлась бы точка c непрерывности функции $E(x)$ и некоторая подпоследовательность последовательности (S), которая сходится в этой точке c к числу, не равному $E(c)$. В этой же подпоследовательности, в свою очередь, нашлась бы подпоследовательность

$$(S^*) \qquad E^*_{n_1}(x),\ E^*_{n_2}(x),\ \ldots,\ E^*_{n_k}(x),\ \ldots,$$

которая сходится в основном к некоторой функции $E^*_\infty(x)$; для этой последней, в частности, мы имели бы

$$E^*_\infty(c) \neq E(c).$$

Но применяя к последовательности (S*) те же рассуждения, которые мы только что применили к последовательности (S'), мы бы убедились в том, что ее предельная функция $E^*_\infty(x)$ совпадает с $E(x)$; в частности, мы имели бы

$$E^*_\infty(c) = E(c),$$

что противоречило бы вышенаписанному неравенству.

Итак, каково бы ни было $\tau > 0$, $E_n(-\tau) \to 0$ и $E_n(\tau) \to 1$. Следовательно, каково бы ни было $\tau > 0$:

$$1 - E_n(\tau) + E_n(-\tau) \to 0,$$

или, что то же:

$$\sum_{k=1}^{k_n} \int\limits_{|x|>\tau} x^2 dF_{nk}(x) \to 0,$$

т. е. выполнено условие (L).

Теорема доказана.

Примечание. — Для применений теоремы Линдеберга — Феллера, доказанной нами для нормированных элементарных систем величин ξ_{nk}, полезно знать, что одно из условий, входящих в определение элементарной системы — равномерное стремление к нулю дисперсий b_{nk} — автоматически выполняется, когда выполнено условие Линдеберга; таким образом, если известно, что выполнено последнее, то нет надобности проверять выполнение первого. Это немедленно вытекает из того, что при любом, сколь угодно малом $\tau > 0$

$$b_{nk} \leqslant \tau^2 \int\limits_{|x|\leqslant\tau} dF_{nk}(x) + \int\limits_{|x|>\tau} x^2 dF_{nk}(x) \leqslant \tau^2 + \int\limits_{|x|>\tau} x^2 dF_{nk}(x),$$

так что, если обозначить через $\overline{b}_n$ наибольшую из дисперсий $b_{n1}, b_{n2}, \ldots, b_{nk_n}$:

$$\overline{b}_n \leqslant \tau^2 + \sum_{k=1}^{k_n} \int\limits_{|x|>\tau} x^2 dF_{nk}(x).$$

97. Из теоремы Линдеберга—Феллера для нормированных элементарных систем легко извлечь различные, более или менее удобные для применений, достаточные условия для того, чтобы функция распределения суммы $\xi_{n1} + \xi_{n2} + \ldots + \xi_{nk_n}$ сходилась к закону Гаусса. Одним из таких условий является следующее:

Условие Ляпунова. — При $n \to \infty$ *для хотя бы одного постоянного* $\nu > 2$

$$\text{(Л)} \qquad \sum_{k=1}^{k_n} \int\limits_{-\infty}^{+\infty} |x|^\nu dF_{nk}(x) \to 0,$$

где $F_{nk}(x)$ — *функция распределения величины* ξ_{nk}.

Иначе говоря, верна следующая теорема:

Теорема Ляпунова. — *Пусть дана нормированная элементарная система независимых случайных величин.*

Тогда для того, чтобы функция распределения суммы $\xi_{n1} + \xi_{n2} + \ldots + \xi_{nk_n}$ *сходилась к закону Гаусса*

$$\text{(G)} \qquad \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{z^2}{2}} dz,$$

достаточно, чтобы было выполнено условие (Л).

Чтобы в этом убедиться, достаточно установить, что выполнение условия (Л) влечет за собой выполнение условия Линдеберга (L). Это же немедленно вытекает из того, что

$$\int\limits_{|x|>\tau} x^2 dF_{nk}(x) \leqslant \frac{1}{\tau^{\nu-2}} \int\limits_{|x|>\tau} |x|^\nu dF_{nk}(x) \leqslant \frac{1}{\tau^{\nu-2}} \int\limits_{-\infty}^{+\infty} |x|^\nu dF_{nk}(x),$$

так что и

$$\sum_{k=1}^{k_n} \int\limits_{|x|>\tau} x^2 dF_{nk}(x) \leqslant \frac{1}{\tau^{\nu-2}} \sum_{k=1}^{k_n} \int\limits_{-\infty}^{+\infty} |x|^\nu dF_{nk}(x).$$

Интересно сравнить условия теоремы Ляпунова, в которой закон Гаусса появляется как предельный закон распределения для сумм независимых случайных величин, с теми условиями, в которых мы вывели выше, в гл. VI, закон Гаусса как точный закон распределения для сумм независимых приращений непрерывно изменяющейся во времени случайной величины.

Для этого обратимся к тому частному случаю, когда все величины ξ_{nk} с одним и тем же индексом n имеют одну и ту же функцию распределения $F_{nk}(x) = F_{(n)}(x)$. В этом случае условия теоремы Ляпунова можно написать в следующем виде, выбрав $\nu = 3$:

$$n \int\limits_{-\infty}^{+\infty} x dF_{(n)}(x) = 0,$$

$$n \int\limits_{-\infty}^{+\infty} x^2 dF_{(n)}(x) = 1,$$

$$n \int\limits_{-\infty}^{+\infty} |x|^3 dF_{(n)}(x) \to 0.$$

Легко видеть аналогию между условиями и равенствами (J_1), (J_2), (J_3), фигурирующими в условии 4° п. 77.

Примечание. — Равномерное стремление к нулю дисперсий b_{nk} при выполнении условия Ляпунова, разумеется, автоматически имеет место, потому что оно имеет место уже при выполнении условия Линдеберга. Здесь это имеет тем большее значение, что речь идет только о достаточных условиях, так что равномерное стремление к нулю дисперсий b_{nk} может вообще не упоминаться в формулировке теоремы.

98. Из теоремы Линдеберга—Феллера легко извлечь заключение о законе Гаусса, как предельном законе распределения, для любых систем независимых случайных величин, обладающих конечными дисперсиями, и то же из теоремы Ляпунова — для величин, обладающих конечными абсолютными моментами какого-нибудь порядка $\nu > 2$. Эти

заключения получаются при помощи того линейного преобразования данных систем к нормированным, которое мы уже применяли при доказательстве теоремы Леви.

А именно, пусть ξ_{nk} — величины данной системы, a_{nk} — их средние значения, b_{nk} — их дисперсии, c_{nk} — их абсолютные моменты ν-го порядка, где ν — какое-нибудь определенное число > 2, и пусть

$$A_n = \sum_{k=1}^{k_n} a_{nk}, \quad B_n = \sum_{k=1}^{k_n} b_{nk}, \quad C_n = \sum_{k=1}^{k_n} c_{nk}.$$

Линейное преобразование, о котором идет речь, состоит в том, что вводят величины

$$\bar{\xi}_{nk} = \frac{\xi_{nk} - a_{nk}}{\sqrt{B_n}}.$$

Будем обозначать числовые характеристики этих величин $\bar{\xi}_{nk}$ так же, как такие же характеристики величин ξ_{nk}, но только снабжая их чертой сверху. Мы будем иметь

$$\bar{a}_{nk} = 0, \quad \bar{b}_{nk} = \frac{b_{nk}}{B_n}, \quad \bar{c}_{nk} = \frac{c_{nk}}{B_n^{\nu/2}},$$

и, следовательно:

$$\bar{A}_n = 0, \quad \bar{B}_n = 1, \quad \bar{C}_n = \frac{C_n}{B_n^{\nu/2}}.$$

Ясно, что функция распределения суммы $\bar{\xi}_{n1} + \bar{\xi}_{n2} + \ldots + \bar{\xi}_{nk_n}$ равна вероятности

$$\text{(P)} \qquad P\left(\frac{\sum\limits_{k=1}^{k_n} \xi_{nk} - A_n}{\sqrt{B_n}} < x\right).$$

Наконец, нетрудно проверить, что выполнение для величин $\bar{\xi}_{nk}$ условия Линдеберга (L) равносильно выполнению для величин ξ_{nk} условия: *при* $n \to \infty$ *для всякого* $\tau > 0$

$$\text{(L*)} \qquad \frac{1}{B_n} \sum_{k=1}^{k_n} \int (x - a_{nk})^2 \, dF_{nk}(x) \to 0,$$

где интегрирование распространено на область

$$|x - a_{nk}| > \tau \sqrt{B_n}.$$

Точно так же выполнение для величины $\bar{\xi}_{nk}$ условия Ляпунова (Л) равносильно выполнению для величин ξ_{nk} условия: *при* $n \to \infty$

$$\text{(Л*)} \qquad \frac{C_n}{B_n^{\nu/2}} \to 0.$$

Таким образом при наличии у величин ξ_{nk} конечных дисперсий, для того чтобы вероятность (P) стремилась к закону Гаусса (G), достаточно, чтобы было выполнено условие (L*). В случае, когда величины ξ_{nk} таковы, что отношения $\frac{b_{nk}}{B_n}$ равномерно стремятся к нулю это условие также и необходимо.

При наличии же у величин ξ_{nk} конечных абсолютных моментов порядка $\nu > 2$, для того чтобы вероятность (P) стремилась к закону Гаусса (G), достаточно, чтобы было выполнено условие (Л*).

99. Закон Пуассона. — Тем же методом, каким мы доказали теорему Линдеберга—Феллера, для нормированных элементарных систем можно легко установить также необходимое и достаточное условие для того, чтобы функция распределения суммы $\xi_{n1} + \xi_{n2} + \ldots + \xi_{nk_n}$ сходилась к закону Пуассона, соответственно нормированному. Мы имеем следующую теорему:

Теорема Б. В. Гнеденко. — *Пусть дана нормированная элементарная система независимых случайных величин. Тогда для того, чтобы функция распределения суммы* $\xi_{n1} + \xi_{n2} + \ldots + \xi_{nk_n}$ *сходилась к* $P(x+1)$, *где*

$$P(x) = \sum_{m=0}^{x} \frac{e^{-1}}{m!},$$

необходимо и достаточно, чтобы было выполнено условие: для любой положительной постоянной τ *при* $n \to \infty$

$$(\Gamma) \qquad \sum_{k=1}^{k_n} \int\limits_{|x-1|>\tau} x^2 dF_{nk}(x) \to 0,$$

где — $F_{nk}(x)$ — *функция распределения величины* ξ_{nk}.

Чтобы в этом убедиться, заметим, что величина ξ, имеющая функцию распределения $P(x)$ — закон Пуассона с дисперсией $k=1$, — имеет характеристическую функцию $e^{e^{it}-1}$. Следовательно, величина, имеющая функцию распределения $P(x+1)$, т. е. величина $\xi - 1$, имеет характеристическую функцию $e^{e^{it}-1}e^{-it}$, так что логарифм этой последней есть $e^{it} - 1 - it$.

Дальше доказательство проводится так же, как и для теоремы Линдеберга—Феллера. В частности, вводятся те же вспомогательные функции $E_n(x)$ и $E(x)$. В силу леммы Бавли дело сводится к доказательству соотношения

$$\sum_{k=1}^{k_n} [\varphi_{nk}(t) - 1] \to e^{it} - 1 - it,$$

где $\varphi_{nk}(t)$ — характеристическая функция величины ξ_{nk}. Но

$$\sum_{k=1}^{k_n} [\varphi_{nk}(t) - 1] = \int\limits_{-\infty}^{+\infty} (e^{itx} - 1 - itx) \frac{1}{x^2} dE_n(x)$$

и

$$e^{it}-1-it=\int\limits_{-\infty}^{+\infty}(e^{itx}-1-itx)\ \frac{1}{x^2}\,dE(x-1).$$

Таким образом дело сводится к доказательству того, что $E_n(x)$ сходится в основном к $E(x-1)$, т. е. того, что каково бы ни было $\tau>0$, $E_n(1-\tau)\to 0$ и $E_n(1+\tau)\to 1$. Это же последнее имеет место тогда, когда выполнено условие (Г), и только тогда, потому что

$$\sum_{k=1}^{k_n}\int\limits_{|x-1|>\tau}x^2dF_{nk}(x)=1-E_n(1+\tau)-E_n(1-\tau).$$

100. В заключение заметим, что в этом параграфе мы ограничились исследованием только наиболее интересных предельных законов распределения: закона Гаусса и закона Пуассона. При этом мы рассматривали только системы таких случайных величин ξ_{nk}, которые имеют конечные дисперсии b_{nk}.

Естественно, конечно, поставить вопрос и шире, и исследовать предельные законы распределения для сумм независимых случайных величин во всей их совокупности, не ограничиваясь упомянутыми двумя законами. Естественно также подвергнуть изучению всевозможные системы случайных величин, не делая ограничительного предположения о существовании конечных дисперсий.

Но решение вопроса, поставленного в такой общей форме, значительно сложнее. Оно было получено лишь в последнее время, в работах П. Леви́ и А. Я. Хинчина; большую помощь оказали здесь также работы А. Н. Колмогорова о связи предельных законов распределения со случайными процессами. Изложение названных работ не могло бы найти места в настоящем курсе [1]).

Мы еще остановимся только, в следующем параграфе, на одном интересном вопросе. А именно, на вопросе о том, каким образом к общей теории предельных законов распределения примыкает закон больших чисел.

§ 26. Закон больших чисел как предельный закон распределения.

101. Принято говорить, что последовательность случайных величин $\zeta_1, \zeta_2, \ldots, \zeta_n, \ldots$ *сходится по вероятности* к случайной величине ζ, если при любой положительной постоянной ε и при $n\to\infty$

$$P(|\zeta_n-\zeta|<\varepsilon)\to 1.$$

В частности, последовательность $\zeta_1, \zeta_2, \ldots, \zeta_n, \ldots$ сходится по

[1]) Исчерпывающее изложение относящихся сюда результатов можно найти в книге А. Я. Хинчина *Предельные законы для сумм независимых случайных величин*, 1938.

вероятности к нулю, если при любой положительной постоянной ε и при $n \to \infty$

$$P(|\zeta_n| < \varepsilon) \to 1.$$

В этих терминах закон больших чисел, в его общей форме теоремы Маркова, может быть формулирован следующим образом:

Пусть ξ_n — независимые случайные величины, имеющие средние значения a_n и дисперсии b_n, удовлетворяющие условию

$$\frac{\sum_{k=1}^{n} b_k}{n^2} \to 0.$$

Тогда последовательность величин

$$\frac{\sum_{k=1}^{n} \xi_k}{n} - \frac{\sum_{k=1}^{n} a_k}{n}$$

сходится по вероятности к нулю.

102. Формулировку закона больших чисел легко еще несколько обобщить — таким способом, который позволит нам в дальнейшем сразу увидеть связь между законом больших чисел и предельными законами распределения для сумм независимых случайных величин. А именно, будем рассматривать не последовательности отдельных случайных величин, а последовательности серий их. Кстати сказать, эти последние и появились впервые в теории вероятностей в явной форме, в работах А. Н. Колмогорова, именно в связи с исследованиями о законе больших чисел. Тогда будет верна следующая теорема:

Пусть дана система (S) *независимых случайных величин* ξ_{nk}, *имеющих среднее значение* 0 *и дисперсии* b_{nk}, *удовлетворяющие условию*

(M) $$\sum_{k=1}^{k_n} b_{nk} \to 0.$$

Тогда последовательность сумм

$$\sum_{k=1}^{k_n} \xi_{nk}$$

сходится по вероятности к нулю.

Это сразу вытекает из неравенства Чебышева, так как названные суммы имеют среднее значение 0 и дисперсии $\sum_{k=1}^{k_n} b_{nk}$; следовательно:

$$P\left(\left|\sum_{k=1}^{k_n} \xi_{nk}\right| < \varepsilon\right) \geqslant 1 - \frac{\sum_{k=1}^{k_n} b_{nk}}{\varepsilon^2}.$$

Теорема Маркова является частным случаем этой теоремы, потому что, если мы положим

$$\xi_{nk} = \frac{\xi_k - a_k}{n}$$

то мы будем иметь

$$\frac{\sum_{k=1}^{n} \xi_k}{n} - \frac{\sum_{k=1}^{n} a_k}{n} = \sum_{k=1}^{n} \xi_{nk}.$$

А если при этом величины ξ_k удовлетворяют условиям Маркова, то, как нетрудно подсчитать, величины ξ_{nk} будут иметь среднее значение 0 и дисперсии $\frac{b_k}{n^2}$, удовлетворяющие условию (М); следовательно, суммы

$$\sum_{k=1}^{n} \xi_{nk}$$

будут сходиться по вероятности к нулю.

103. Назовем *единичной функцией распределения* функцию $E(x)$, которую мы ввели при доказательстве теоремы Линдеберга—Феллера

$$E(x) = \begin{cases} 0 \text{ при } x < 0, \\ 1 \text{ при } x > 0. \end{cases}$$

Это, очевидно, функция распределения некоторой случайной величины, а именно — постоянной, равной нулю.

Лемма. — *Последовательность случайных величин $\zeta_1, \zeta_2, \ldots, \zeta_n, \ldots$ сходится по вероятности к нулю тогда и только тогда, когда последовательность функций распределения $F_1(x), F_2(x), \ldots, F_n(x), \ldots$ этих величин сходится в основном к единичной функции распределения.*

Доказательство. — Пусть D — множество точек $x \neq 0$, обладающих тем свойством, что как в самих этих точках x, так и в точках $-x$ все функции распределения $F_1(x), F_2(x), \ldots, F_n(x), \ldots$ непрерывны. Это множество, очевидно, всюду плотно. Пусть δ — любая положительная точка, принадлежащая множеству D. Тогда

$$P(|\zeta_n| < \delta) = F_n(\delta) - F_n(-\delta).$$

Теперь, соотношение

$$P(|\zeta_n| < \varepsilon) \to 1$$

имеет место для всех положительных точек ε, очевидно, тогда и только тогда, когда

$$P(|\zeta_n| < \delta) \to 1$$

для всех точек δ. Следовательно, тогда и только тогда, когда

$$F_n(\delta) - F_n(-\delta) \to 1,$$

Следовательно, в силу того, что $0 \leqslant F_n(-\delta) \leqslant F_n(\delta) \leqslant 1$, — тогда и только тогда, когда

$$F_n(-\delta) \to 0 \quad \text{и} \quad F_n(\delta) \to 1.$$

Иначе говоря, тогда и только тогда, когда функции $F_n(x)$ сходятся к функции $E(x)$ на всюду плотном множестве D.

В силу этой леммы теорему п. 102 можно формулировать так: при условии (M) функции распределения сумм

$$\sum_{k=1}^{k_n} \xi_{nk}$$

сходятся в основном к единичной функции распределения. Мы видим, таким образом, что закон больших чисел есть не что иное, как утверждение о наличии, при известных условиях, определенного предельного закона распределения для сумм независимых случайных величин: именно — закона распределения $E(x)$.

104. Это обстоятельство открывает возможность применения к исследованию закона больших чисел теории характеристических функций. Ниже мы увидим некоторые из таких применений.

Разумеется, надо знать заранее характеристическую функцию величины ζ, имеющей функцию распределения $E(x)$. Найти ее очень просто. Названная величина ζ есть постоянная, равная нулю. Поэтому для ее характеристической функции $\varphi(t)$ мы имеем

$$\varphi(t) = \boldsymbol{E}e^{it\zeta} = \boldsymbol{E}e^{0} = 1,$$
$$\log \varphi(t) = 0.$$

105. Естественно поставить вопрос о необходимых и достаточных условиях для применимости закона больших чисел. Условия для сходимости по вероятности последовательности случайных величин были указаны впервые С. Н. Бернштейном. Но естественно также поставить вопрос о таких необходимых и достаточных условиях для применимости закона больших чисел к суммам независимых случайных величин, которые выражались бы через числовые характеристики отдельных слагаемых. Такие условия были указаны в самом общем виде А. Н. Колмогоровым.

Мы не имеем возможности входить здесь в эти довольно сложные исследования. Ограничимся одним интересным частным случаем.

Теорема А. Я. Хинчина. — *Пусть $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ — независимые случайные величины, имеющие одну и ту же функцию распределения. Тогда необходимым и достаточным условием для того, чтобы сходилась по вероятности к нулю величина*

$$\frac{\sum\limits_{k=1}^{n} \xi_k}{n} - a,$$

где a — среднее значение величин ξ_k, является существование этого среднего значения.

Необходимость условия ясна из самой формулировки теоремы. Остается доказать его достаточность.

1° Положим

$$\bar{\xi}_k = \xi_k - a.$$

Величины $\bar{\xi}_k$ имеют одну и ту же функцию распределения, которую мы будем обозначать через $F(x)$, и среднее значение, равное нулю.

При этом

$$\frac{\sum\limits_{k=1}^{n} \xi_k}{n} - a = \sum_{k=1}^{n} \frac{\bar{\xi}_k}{n}.$$

Дело сводится к сходимости по вероятности к нулю последовательности сумм, стоящих в правой части этого равенства.

Введем еще характеристические функции $\varphi(t)$ и $\varphi_n(t)$ величин $\bar{\xi}_k$ и $\frac{\bar{\xi}_k}{n}$. Очевидно, что

$$\varphi_n(t) = \varphi\left(\frac{t}{n}\right).$$

При этом для характеристических функций $\varphi_n(t)$ слагаемых $\frac{\bar{\xi}_k}{n}$ будет выполнено условие леммы Бавли:

$$\sum_{k=1}^{n} |\varphi_n(t) - 1|^2 \to 0.$$

Это видно из того, что

$$\sum_{k=1}^{n} |\varphi_n(t) - 1|^2 = n\left|\varphi\left(\frac{t}{n}\right) - 1\right|^2 = n\left|\int_{-\infty}^{+\infty} \left(e^{\frac{itx}{n}} - 1\right) dF(x)\right|^2.$$

А так как при всяком действительном y справедливо неравенство $|e^{iy} - 1| \leqslant |y|$, то, полагая $y = \frac{tx}{n}$, мы имеем

$$\sum_{k=1}^{n} |\varphi_n(t) - 1|^2 \leqslant n\left(\frac{|t|}{n} \int_{-\infty}^{+\infty} |x|\, dF(x)\right)^2 = \frac{|t|^2}{n}\left(\int_{-\infty}^{+\infty} |x|\, dF(x)\right)^2.$$

Вспомним, что логарифм характеристической функции величины, имеющей единичную функцию распределения, равен нулю. Таким образом в силу леммы п. 103 и леммы Бавли дело сводится окончательно к установлению соотношения

$$\sum_{k=1}^{n} [\varphi_n(t) - 1] \to 0.$$

2° Чтобы убедиться в этом, заметим, что так как

$$\int_{-\infty}^{+\infty} x\,dF(x) = 0,$$

то

$$\sum_{k=1}^{n} [\varphi_n(t) - 1] = n\left[\varphi\left(\frac{t}{n}\right) - 1\right] = n \int_{-\infty}^{+\infty} \left(e^{\frac{itx}{n}} - 1 - \frac{itx}{n}\right) dF(x).$$

А так как при всяком действительном y справедливы неравенства $|e^{iy} - 1 - iy| \leqslant \frac{y^2}{2}$ и $|e^{iy} - 1 - iy| \leqslant 2|y|$, то, взяв произвольно малую положительную постоянную τ, полагая $y = \frac{tx}{n}$ и применяя первое неравенство при $|x| \leqslant \tau n$ и второе при $|x| > \tau n$, мы имеем

$$\left|\sum_{k=1}^{n} [\varphi_n(t) - 1]\right| \leqslant \frac{t^2}{2n} \int\limits_{|x| \leqslant \tau n} x^2 dF(x) + 2|t| \int\limits_{|x| > \tau n} |x|\,dF(x) \leqslant$$

$$\leqslant \frac{\tau t^2}{2} \int_{-\infty}^{+\infty} |x|\,dF(x) + 2|t| \int\limits_{|x| > \tau n} |x|\,dF(x).$$

Первый член в этой последней сумме может быть сделан сколь угодно малым вместе с τ, второй стремится к нулю вследствие существования среднего значения величины ξ_k. Таким образом требуемое соотношение доказано.

106. Обратимся теперь еще раз к нормированным элементарным системам, чтобы показать интересную связь между законом больших чисел и законом Гаусса, которая была открыта в последнее время. А именно, имеет место следующая теорема:

Теорема Д. А. Райкова. — *Пусть дана нормированная элементарная система независимых случайных величин. Тогда для того, чтобы функция распределения суммы*

$$\sum_{k=1}^{k_n} \xi_{nk}$$

при $n \to \infty$ *стремилась к закону Гаусса* (G), *необходимо и достаточно, чтобы последовательность сумм квадратов*

$$\sum_{k=1}^{k_n} \xi_{nk}^2$$

сходилась по вероятности к единице.

Для доказательства мы воспользуемся теоремой Линдеберга—Феллера, согласно которой функция распределения суммы

$$\sum_{k=1}^{k_n} \xi_{nk}$$

при $n \to \infty$ стремится к закону Гаусса (G) тогда и только тогда, когда выполнено условие: для любой положительной постоянной τ при $n \to \infty$

$$\text{(L)} \qquad \sum_{k=1}^{k_n} \int\limits_{|x|>\tau} x^2 dF_{nk}(x) \to 0,$$

где $F_{nk}(x)$ — функция распределения величины ξ_{nk}.

1° Введем характеристические функции $\widehat{\varphi}_{nk}(t)$ квадратов ξ_{nk}^2. Покажем, что для них будет выполнено условие леммы Бавли:

$$\sum_{k=1}^{k_n} |\widehat{\varphi}_{nk}(t) - 1|^2 \to 0.$$

Действительно, так как $\widehat{\varphi}_{nk}(t) = Ee^{it\xi_{nk}^2}$, то

$$\widehat{\varphi}_{nk}(t) - 1 = \int\limits_{-\infty}^{+\infty} (e^{itx^2} - 1)\, dF_{nk}(x).$$

А так как при всяком действительном y справедливо неравенство $|e^{iy} - 1| \leqslant |y|$, то, полагая $y = tx^2$, мы имеем

$$|\widehat{\varphi}_{nk}(t) - 1| \leqslant |t| \int\limits_{-\infty}^{+\infty} x^2 dF_{nk}(x) = |t|\, b_{nk},$$

где b_{nk} — дисперсия величины ξ_{nk}. Отсюда

$$\sum_{k=1}^{k_n} |\widehat{\varphi}_{nk}(t) - 1|^2 \leqslant t^2 \sum_{k=1}^{k_n} b_{nk}^2 \leqslant t^2\, \overline{b}_n \sum_{k=1}^{k_n} b_{nk},$$

где $\overline{b}_n$ — наибольшая из дисперсий $b_{n1}, b_{n2}, \ldots, b_{nk_n}$. Отсюда ясно, что условие леммы Бавли, действительно, выполнено, так как система величин ξ_{nk} — элементарная и нормированная.

2° Суммы квадратов

$$(\Sigma) \qquad \sum_{k=1}^{k_n} \xi_{nk}^2$$

сходятся по вероятности к единице, очевидно, тогда и только тогда, когда величины

$$(\Sigma_1) \qquad \sum_{k=1}^{k_n} \xi_{nk}^2 - 1$$

сходятся по вероятности к нулю. В силу леммы п. 103 это последнее равносильно тому, что функции распределения величин (Σ_1) сходятся в основном к единичной функции распределения $E(x)$, что имеет место, очевидно, тогда и только тогда, когда функции распределения сумм квадратов (Σ) сходятся в основном к $E(x-1)$. Но величина, имеющая функцию распределения $E(x-1)$, есть постоянная, равная единице. Поэтому ее характеристическая функция есть e^{it}. Таким образом в силу леммы Бавли суммы квадратов

$$\sum_{k=1}^{k_n} \xi_{nk}^2$$

сходятся по вероятности к единице тогда и только тогда, когда при $n \to \infty$

$$\text{(Z)} \qquad \sum_{k=1}^{k_n} [\widehat{\varphi}_{nk}(t) - 1] \to it.$$

Остается показать, следовательно, что выполнение условия (L) влечет за собой (Z) и обратно.

3° Пусть выполнено условие (L). В нормированной системе

$$\sum_{k=1}^{k_n} [\widehat{\varphi}_{nk}(t) - 1] - it = \sum_{k=1}^{k_n} \int_{-\infty}^{+\infty} (e^{itx^2} - 1 - itx^2)\, dF_{nk}(x).$$

А так как при всяком действительном y справедливы неравенства $|e^{iy} - 1 - iy| \leqslant \frac{y^2}{2}$ и $|e^{iy} - 1 - iy| \leqslant 2|y|$, то, взяв произвольно малую положительную постоянную τ, полагая $y = tx^2$ и применяя первое неравенство при $|x| \leqslant \tau$ и второе при $|x| > \tau$, мы имеем

$$\left| \sum_{k=1}^{k_n} [\widehat{\varphi}_{nk}(t) - 1] - it \right| \leqslant$$

$$\leqslant \frac{t^2}{2} \sum_{k=1}^{k_n} \int_{|x| \leqslant \tau} x^4 dF_{nk}(x) + 2|t| \sum_{k=1}^{k_n} \int_{|x| > \tau} x^2 dF_{nk}(x) \leqslant$$

$$\leqslant \frac{\tau^2 t^2}{2} \sum_{k=1}^{k_n} \int_{|x| \leqslant \tau} x^2 dF_{nk}(x) + 2|t| \sum_{k=1}^{k_n} \int_{|x| > \tau} x^2 dF_{nk}(x) \leqslant$$

$$\leqslant \frac{\tau^2 t^2}{2} + 2|t| \sum_{k=1}^{k_n} \int_{|x| > \tau} x^2 dF_{nk}(x).$$

В силу произвольной малости τ и условия (L) отсюда и вытекает (Z).

4° Пусть теперь имеет место соотношение (Z). Напишем снова

$$\sum_{k=1}^{k_n} [\widehat{\varphi}_{nk}(t)-1]-it=\sum_{k=1}^{k_n} \int_{-\infty}^{+\infty} (e^{itx^2}-1-itx^2)\,dF_{nk}(x).$$

Так как имеет место соотношение (Z), то обе части этого равенства стремятся к нулю; следовательно, стремится к нулю и мнимая часть правой части этого равенства, так что

$$\sum_{k=1}^{k_n} \int_{-\infty}^{+\infty} (tx^2-\sin tx^2)\,dF_{nk}(x)\to 0$$

при каждом t. Для положительных t отсюда вытекает, что при любом положительном τ тем более

$$\sum_{k=1}^{k_n} \int_{|x|>\tau} (tx^2-\sin tx^2)\,dF_{nk}(x)\to 0.$$

Но при достаточно большом t и при $|x|>\tau$ выполняются неравенства

$$tx^2-\sin tx^2>tx^2\left(1-\frac{1}{t\tau^2}\right)>0,$$

а потому и

$$t\left(1-\frac{1}{t\tau^2}\right)\sum_{k=1}^{k_n} \int_{|x|>\tau} x^2\,dF_{nk}(x)\to 0.$$

Это показывает, что выполняется и условие (L).

Сам Д. А. Райков нашел этот результат, исходя из условия, полученного А. А. Бобровым для применимости закона больших чисел к положительным случайным величинам, и доказал, что утверждение теоремы верно для любых систем случайных величин. Но доказательство в общем случае требует более сложного аналитического аппарата.

ГЛАВА IX.

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И СТАТИСТИКА.

§ 27. Эмпирические законы распределения.

107. В §§ 20 и 21 мы выяснили с помощью закона больших чисел, что, производя достаточно большое число независимых испытаний, при каждом из которых вероятность некоторого события A равна p, можно достичь практической уверенности в том, что в этом ряду испытаний относительная частота наступления события A будет как угодно мало отличаться от p.

На этом утверждении основана возможность эмпирической проверки выводов теории вероятностей. Пусть мы имеем дело с определенным родом предметов, которые могут обладать или не обладать некоторым признаком X. Будем выбирать наугад один за другим некоторое количество этих предметов и обозначать через A появление при каждой выборке предмета с признаком X и через p вероятность этого появления. В силу только что сказанного при достаточно большом количестве выбранных предметов мы будем иметь практическую уверенность в том, что относительная доля предметов с признаком X среди выбранных предметов будет как угодно мало отличаться от p. Если значение p вычислено заранее на основе теории, то относительная доля предметов с признаком X среди выбранных предметов даст нам представление о правильности этой теории. Если же p заранее неизвестно, то эта относительная доля даст нам представление о возможном значении p.

108. Чаще всего приходится иметь дело не с одним признаком, а со многими признаками $X_1, X_2, \ldots, X_s$. Пусть $p_1, p_2, \ldots, p_s$ — вероятности того, что предмет данного рода обладает названными признаками. Будем выбирать наугад n предметов и обозначим через $m_1, m_2, \ldots, m_s$ количества выбранных предметов, обладающих этими признаками, так что

$$m_1 + m_2 + \ldots + m_s = n.$$

При достаточно большом количестве выбранных предметов мы будем иметь практическую уверенность в том, что относительные доли $\frac{m_2}{n}$ $\frac{m_s}{n}$ будут все одновременно как угодно мало отличаться от вероятностей $p_1, p_2, \ldots, p_s$. Иначе говоря, в том, что наибольшая из абсолютных величин разностей

$$\frac{m_1}{n}-p_1, \quad \frac{m_2}{n}-p_2, \quad \ldots, \quad \frac{m_s}{n}-p_s$$

будет как угодно мала. В этом можно убедиться следующим рассуждением. Возьмем произвольные положительные постоянные ε и η. Наибольшая из абсолютных величин названных разностей будет $\geqslant \varepsilon$ тогда и только тогда, когда по крайней мере одна из этих разностей будет по абсолютной величине $\geqslant \varepsilon$. Вероятность этого не превосходит суммы вероятностей каждого в отдельности из неравенств [1])

$$\left|\frac{m_1}{n}-p_1\right| \geqslant \varepsilon, \quad \left|\frac{m_2}{n}-p_2\right| \geqslant \varepsilon, \quad \ldots, \quad \left|\frac{m_s}{n}-p_s\right| \geqslant \varepsilon.$$

По теореме же Бернулли вероятность каждого из этих неравенств при достаточно большом n не превосходит $\frac{\eta}{s}$; следовательно, сумма этих вероятностей при достаточно большом n не превосходит η. Итак, вероятность того, что наибольшая из абсолютных величин названных разностей будет $\geqslant \varepsilon$, при достаточно большом n меньше η; следовательно, вероятность того, что она будет $< \varepsilon$, при достаточно большом n больше $1-\eta$. Так как это верно при произвольно малом η, то тем самым наше утверждение доказано.

Мы видим, таким образом, что эмпирическая проверка выводов теории вероятностей, применимая к отдельному признаку, применима и ко многим признакам: Если теоретический закон распределения признаков, т. е. вероятности $p_1, p_2, \ldots, p_s$, установлен заранее, то *эмпирический закон распределения* тех же признаков, т. е. относительные доли $\frac{m_1}{n}, \frac{m_2}{n}, \ldots, \frac{m_s}{n}$, даст нам представление о правильности теории. Если же теоретический закон заранее неизвестен, то эмпирический закон даст нам представление о том, какой вид может иметь этот теоретический закон.

109. Распределение признаков поддается наиболее полному изучению тогда, когда признаки выражаются числами: $x_1, x_2, \ldots, x_s$. В этом случае мы имеем дело со случайной величиной, которая может принимать значения $x_1, x_2, \ldots, x_s$.

На практике, как правило, не учитывают отдельно каждого из возможных числовых значений признака, а объединяют их в группы. Вспомним, например, задачу о статистической выборке (§ 6) и ее обращение (§ 8). В них шла речь о совокупности, в которой содержится N предметов. Мы решали эти задачи, учитывая отдельно каждое возможное число $0, 1, 2, \ldots, N$ предметов, обладающих некоторым признаком X. На практике, однако, часто вычисляют только

---

[1]) На основании неравенства

$$\boldsymbol{P}(E_1+E_2+\ldots+E_s) \leqslant \boldsymbol{P}(E_1)+\boldsymbol{P}(E_2)+\ldots+\boldsymbol{P}(E_s),$$

доказанного в лемме I п. 71.

процент этих предметов, т. е. разбивают все возможные числа предметов на 101 группу, принимая условно:

$$\text{от } 0 \quad \text{до } \frac{1}{2}\frac{N}{100} \text{ предметов — за } 0\%,$$

$$\text{от } \frac{1}{2}\frac{N}{100} \text{ до } \frac{3}{2}\frac{N}{100} \text{ предметов — за } 1\%,$$

$$\text{от } \frac{3}{2}\frac{N}{100} \text{ до } \frac{5}{2}\frac{N}{100} \text{ предметов — за } 2\%,$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$\text{от } \frac{199}{2}\frac{N}{100} \text{ до } N \text{ предметов — за } 100\%.$$

Именно процент и будет здесь числовым значением изучаемого признака. К нему, разумеется, будут применимы те же рассуждения и формулы, что и к самому числу предметов, но при больших N, например, при $N = 1000$, количество нужных для решения задачи вычислений здесь будет значительно меньше.

Такое объединение значений признака в группе особенно естественно в том случае, когда возможное значение признака является непрерывной случайной величиной.

В этом случае группировка значений признака обычно состоит в том, что разбивают множество возможных значений на равные интервалы, точнее, задают определенный интервал (a, b), подразделяют его на $s-1$ частичных интервалов равной длины h и вместо непрерывной случайной величины ξ рассматривают дискретную случайную величину ξ^*, которая:

$$\text{при } -\infty < \xi < a + \frac{1}{2}h \text{ принимает значение } x_1 = a,$$

$$\text{при } a + \frac{1}{2}h \leqslant \xi < a + \frac{3}{2}h \text{ принимает значение } x_2 = a + h,$$

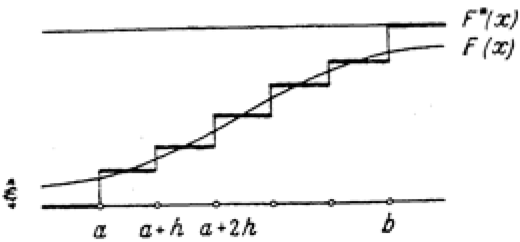
$$\text{при } a + \frac{3}{2}h \leqslant \xi < a + \frac{5}{2}h \text{ принимает значение } x_3 = a + 2h,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\text{при } b - \frac{1}{2}h \leqslant \xi < +\infty \text{ принимает значение } x_s = b.$$

При соответствующим образом подобранных a и b и при достаточно малой длине h частичных интервалов закон распределения величины ξ^* будет как угодно мало отличаться от закона распределения величины ξ. Действительно, пусть функции распределения той и другой величины будут $F^*(x)$ и $F(x)$. Возьмем произвольное положительное число ε, меньшее единицы. Выберем точки a и b так, чтобы удовлетворялись равенства $F(a) = \varepsilon$ и $F(b) = 1 - \varepsilon$ (что всегда возможно благодаря непрерывности функции $F(x)$). Интервал (a, b) между этими точками подразделим на $s-1$ равных между собой частичных интервалов настолько малой длины h, чтобы в каждом из них колебание функции $F(x)$ было меньше ε (что, опять-таки, всегда возможно благодаря непрерывности функции $F(x)$). Ступенчатая функция распределения $F^*(x)$ величины ξ^*, соответствующей такому подразделению, изображена на черт. 13. Ясно, что значение функции $F^*(x)$ отличается от значения функции $F(x)$ в любой точке x не больше, чем на колебание функции $F(x)$ в том частичном интервале, в котором находится точка x, т. е. меньше, чем на ε.

Так же и эмпирический закон распределения может быть изображен ступенчатой *эмпирической функцией распределения* $F_n(x)$, значения которой последовательно равны $0, \frac{m_1}{n}, \frac{m_1}{n}+\frac{m_2}{n}, \ldots, \frac{m_1}{n}+\frac{m_2}{n}+ + \ldots + \frac{m_s}{n}$. При соответствующим образом подобранных a и b, достаточно малой длине h частичных интервалов и достаточно большом n мы будем иметь практическую уверенность в том, что $F_n(x)$ будет как угодно мало отличаться от $F(x)$, притом равномерно на всей прямой $-\infty < x < +\infty$, т. е. в том, что будет как угодно мала величина



Черт. 13.

$$\sup_x |F_n(x)-F(x)|,$$

где $\sup_x$ обозначает верхнюю грань для всех значений x. В самом деле, пусть ε — сколь угодно малая положительная постоянная. Подберем a, b и h так, чтобы было

$$\sup_x |F^*(x)-F(x)| < \frac{\varepsilon}{2}.$$

Тогда

$$\sup_x |F_n(x)-F^*(x)| \leqslant \sup_x |F_n(x)-F^*(x)| + \frac{\varepsilon}{2}.$$

Далее, всякому x соответствует некоторое $i \leqslant s$, такое, что

$$F_n(x)-F^*(x) = \left(\frac{m_1}{n}-p_1\right)+\left(\frac{m_2}{n}-p_2\right)+\ldots+\left(\frac{m_i}{n}-p_i\right);$$

поэтому при всяком x

$$|F_n(x)-F^*(x)| \leqslant \left|\frac{m_1}{n}-p_1\right|+\left|\frac{m_2}{n}-p_2\right|+\ldots+\left|\frac{m_s}{n}-p_s\right|,$$

и следовательно,

$$\sup_x |F_n(x)-F^*(x)| \leqslant \max_i \left|\frac{m_i}{n}-p_i\right|,$$

где $\max\limits_i$ обозначает максимум для всех $i=1, 2, \ldots, s$. Таким образом мы будем иметь при достаточно большом n практическую уверенность в том, что будет выполнено неравенство

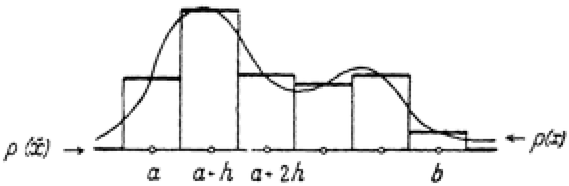
$$\sup_x |F_n(x) - F(x)| < \varepsilon,$$

если у нас будет при достаточно большом n практическая уверенность в выполнении неравенства

$$\max_i \left| \frac{m_i}{n} - p_i \right| < \frac{\varepsilon}{2}.$$

Это же последнее было доказано в п. 108.

110. Существует геометрическая интерпретация, позволяющая еще более наглядно сравнить законы распределения непрерывной случайной величины ξ и дискретной случайной величины ξ^*. Для величины ξ мы возьмем кривую $y=p(x)$, где $p(x)$ — плотность вероятности величины ξ. Для величины же ξ^* мы возьмем ступенчатую линию $y=p^*(x)$, где при $a-\frac{1}{2}h \leqslant x < b+\frac{1}{2}h$ функция $p^*(x)$ равна деленной на h вероятности ближайшего к x значения величины ξ^*, а при $x < a-\frac{1}{2}h$ и при $x > b+\frac{1}{2}h$ она равна нулю. Таким образом для величины ξ^*, как и для величины ξ, сами вероятности ее значений будут изображаться не ординатами, а площадями (черт. 14).



Черт. 14.

При соответствующим образом подобранных a и b и при достаточно малой длине h частичных интервалов функция $p^*(x)$ будет как угодно мало отличаться от функции $p(x)$. Это доказывается точно так же, как было доказано аналогичное предложение о функциях $F^*(x)$ и $F(x)$ в п. 109.

Так же и эмпирический закон распределения может быть изображен — и это очень часто делается в статистической литературе — ступенчатой линией, ординаты которой $p_n(x)$ равны деленным на h относительным долям $\frac{m_1}{n}, \frac{m_2}{n}, \ldots, \frac{m_s}{n}$; такая линия называется **гистограммой**. При соответствующим образом подобранных a и b, достаточно малой длине h частичных интервалов и достаточно большом n мы будем иметь практическую уверенность в том, что $p_n(x)$ будет как угодно мало отличаться от $p(x)$, притом равномерно на всей прямой $-\infty < x < +\infty$. Это доказывается так же, как было доказано аналогичное предложение о функциях $F_n(x)$ и $F(x)$ в п. 109. Разница состоит только в том, что вместо данной там оценки разности $F_n(x)-F^*(x)$ мы имеем здесь более простую оценку разности $p_n(x)-p^*(x)$, а именно: всякому x соответствует некоторое i, такое, что

$$p_n(x)-p^*(x)=\frac{1}{h}\left(\frac{m_i}{n}-p_i\right)$$

и, следовательно,

$$\sup_x |p_n(x)-p^*(x)|=\frac{1}{h}\max_i\left|\frac{m_i}{n}-p_i\right|.$$

Соотношения между $p_n(x)$ и $p(x)$ были изучены и более подробно. В частности, В. И. Гливенко было показано, что если, увеличивая до бесконечности число n независимых испытаний, одновременно увеличивать до бесконечности также и число s делений интервала (a, b), но настолько медленно, чтобы выражение $\left(\frac{s^2}{n}\right)^2$ было членом сходящегося ряда, то здесь применяется усиленный закон больших чисел, т. е. для интервала (a, b) мы имеем

$$P(\sup_x|p_n(x)-p(x)|\to 0)=1.$$

Н. В. Смирновым найден предельный закон распределения: если в интервале (a, b) нижняя грань функции $p(x)$ положительна и если выражение $\frac{m\log^3 m}{n}$ стремится к нулю, то в интервале (a, b)

$$P\left(\sup_x\frac{|p_n(x)-p^*(x)|}{\sqrt{p^*(x)}}<\frac{l+\frac{\lambda}{l}}{\sqrt{nh}}\right)\to e^{-2e^{-\lambda}},$$

где l есть корень уравнения

$$\frac{1}{\sqrt{2\pi}}\int_l^{+\infty}e^{-\frac{z^2}{2}}\,dz=\frac{1}{m}.$$

111. Приведем типичный образец применения развитых выше соображений об эмпирической проверке законов распределения. Пусть дана случайная величина ξ, относительно которой теоретически можно допустить, что она подчиняется закону распределения определенного вида (например, закону Гаусса), однако, с неизвестными нам средним значением $\boldsymbol{E}\xi$ и дисперсией $\boldsymbol{D}\xi$. Произведем группировку значений величины ξ, т. е. заменим ее дискретной случайной величиной ξ^*, и установим эмпирический закон распределения этой последней, т. е. определим относительно доли $\frac{m_1}{n}, \frac{m_2}{n}, \ldots, \frac{m_s}{n}$. Для эмпирического закона распределения можно ввести числовые характеристики, аналогичные тем, которые мы вводили для теоретических законов распределения. Это, во-первых, *эмпирическое среднее значение*:

$$Ex=\sum_{i=1}^{s}x_i\frac{m_i}{n}=\frac{1}{n}\sum_{i=1}^{s}x_im_i \tag{91}$$

и, во-вторых, *эмпирическая дисперсия*

$$Dx = \sum_{i=1}^{s} x_i^2 \frac{m_i}{n} - \left(\sum_{i=1}^{s} x_i \frac{m_i}{n}\right)^2 = \sum_{i=1}^{s}\left(x_i - \sum_{j=1}^{s} x_j \frac{m_j}{n}\right)^2 \frac{m_i}{n} =$$

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

Так как при достаточно большом количестве выбранных предметов мы будем иметь практическую уверенность в том, что относительные доли $\frac{m_1}{n}$, $\frac{m_2}{n}, \ldots, \frac{m_s}{n}$ будут все одновременно как угодно мало отличаться от соответствующих вероятностей, то, как нетрудно видеть, мы будем иметь практическую уверенность и в том, что названные эмпирические характеристики будут как угодно мало отличаться от соответствующих теоретических характеристик. Таким образом числа Ex и Dx дадут нам приближенное представление о $\boldsymbol{E}\xi^*$ и $\boldsymbol{D}\xi^*$, а следовательно, и о $\boldsymbol{E}\xi$ и $\boldsymbol{D}\xi$.

После этого остается подставить найденные приближенные значения Ex и Dx среднего значения $\boldsymbol{E}\xi$ и дисперсии $\boldsymbol{D}\xi$ в формулу предполагаемого закона распределения величины ξ и вычислить по этой формуле вероятности, соответствующие найденным нами относительным долям. Если те и другие будут незначительно отличаться друг от друга, то мы получим двойной результат: во-первых, убедимся в непротиворечивости сделанного допущения о виде закона распределения, которому подчиняется величина ξ и, во-вторых, пополним наше знание о нем тем, что будем знать, хотя бы приближенно, среднее значение и дисперсию этой величины.

Приведем один конкретный пример. Американский ученый Тоуэр исследовал колорадского жука Leptinotarsa decemlineata. Представители этого вида жуков отличаются развитием темных полос на более светлой поверхности надкрылий, причем эта окраска сильно варьирует у различных индивидов. Тоуэр различает несколько степеней развития темных полос, несколько „классов окраски". Исследование 100 взятых наугад жуков дало результат, показанный в табл. IV.

Таблица IV.

Класс окраски	1	2	3	4	5	6	7	8	9
Число жуков	1	2	5	20	42	18	9	2	1

Так как та или иная степень развития темных полос складывается из незначительных изменений окраски под влиянием громадного числа независимых друг от друга обстоятельств, то теоретически можно допустить, что она является непрерывной случайной величиной ξ, подчиняющейся закону Гаусса. Введенный Тоуэром „класс окраски" есть не что иное, как заменяющая эту непрерывную случайную величину дискретная случайная величина ξ^*. Таблица Тоуэра дает эмпирический закон распределения этой последней; в ней даны числа $m_1, m_2, \ldots, m_s$, от которых можно перейти к относительным долям $\frac{m_1}{n}, \frac{m_2}{n}, \ldots, \frac{m_s}{n}$ делением этих чисел на 100. Вычислим по формуле (91) эмпирическое среднее значение Ex, для чего перемножим между собой числа, стоящие в каждом столбце таблицы, сложим все полученные произведения и разделим полученную сумму на 100. Мы получим $Ex = 5{,}02$, или приближенно

$$Ex = 5.$$

Вычислим затем по формуле (92) эмпирическую дисперсию Dx, для чего вычтем из каждого числа, стоящего в первой строке таблицы, найденное среднее значение Ex, возведем каждую полученную разность в квадрат, помножим каждый из этих квадратов на соответствующее число из второй строки таблицы, сложим все полученные произведения и разделим полученную сумму на 100. Мы получим

$$Dx = 1{,}62.$$

Таким образом закон Гаусса, которому предположительно подчиняется интересующая нас степень развития темных полос, имеет параметры α и β, приближенно равные: $\alpha = Ex = 5$ и $\beta = Dx = 1{,}62$.

Черт. 15.

Остается вычислить, пользуясь формулой закона распределения Гаусса, вероятности, соответствующие найденным Тоуэром относительным долям, и сравнить те и другие. При этом следует иметь в виду, что в один и тот же „класс окраски" x зачисляются, очевидно, индивиды, степень окраски которых несколько варирует в обе стороны от какой-то стандартной величины, характеризующей этот класс; в него зачисляются индивиды со степенью окраски, грубо говоря, лежащей между $x - 0{,}5$ и $x + 0{,}5$ (черт. 15).

Поэтому за вероятность $P(x)$, соответствующую „классу окраски" x, естественно принять вероятность неравенств

$$x - 0{,}5 < \xi < x + 0{,}5.$$

Эта вероятность равна

$$\frac{1}{\sqrt{2\pi\beta}} \int\limits_{x-0{,}5}^{x+0{,}5} e^{-\frac{(s-\alpha)^2}{2\beta}}\, ds = \frac{1}{\sqrt{2\pi}} \int\limits_{\frac{x-0{,}5-\alpha}{\sqrt{\beta}}}^{\frac{x+0{,}5-\alpha}{\sqrt{\beta}}} e^{-\frac{z^2}{2}}\, dz =$$

$$= \Phi\left(\frac{x+0{,}5-\alpha}{\sqrt{\beta}}\right) - \Phi\left(\frac{x-0{,}5-\alpha}{\sqrt{\beta}}\right),$$

или в данном случае:

$$\Phi(0{,}8x - 3{,}6) - \Phi(0{,}8x - 4{,}4).$$

Взяв нужные значения функции Φ из таблицы, приложенной в конце книги, мы получим табл. V, где вероятности $P(x)$, вычисленные с точностью до 0,01, затем умножены на 100 для удобства сравнения с вышеприведенной таблицей Тоуэра.

Таблица V.

Класс окраски x	1	2	3	4	5	6	7	8	9
$100\,P(x)$	0	2	5	23	46	23	5	2	0

Сравнение показывает, что расхождение между относительными долями и соответствующими им вероятностями не превосходит 0,05. Принимая во внимание незначительное число и небольшую детализацию данных, этого достаточно, чтобы признать, что допущение о наличии здесь закона распределения Гаусса вполне оправдано.

112. Во всех предыдущих рассуждениях молчаливо предполагалось, что появление в выборке предмета с каким-нибудь признаком X имеет постоянную вероятность p, не зависящую от того, какие предметы уже были выбраны раньше. Это, разумеется, в точности верно лишь в том случае, когда каждый выбранный предмет возвращается в совокупность, из которой он выбран. В противном случае в совокупности, из которой выбираются предметы, будет постепенно изменяться как общее количество N предметов, так и количество M предметов с признаком X, что повлечет за собой и изменение вероятности p, которая равна отношению M к N. В § 6, решая задачу о статистической выборке, мы видели, каким образом будет изменяться названная вероятность.

Однако мы там же отметили, что когда количество выбранных предметов очень невелико по сравнению с общим количеством предметов в совокупности, это изменение вероятности незначительно и не оказывает влияния на вычисления. Так как на практике дело по большей части так и обстоит, то мы не станем подвергать здесь детальному исследованию тот случай, когда приходится считаться с изъятием из совокупности последовательно выбираемых из нее предметов.

§ 28. Новая концепция эмпирических законов распределения.

113. В предыдущем параграфе мы изложили классическую теорию эмпирических законов распределения. Характерной ее чертой являлось получение эмпирического закона распределения путем предварительного подразделения возможных значений изучаемой случайной величины на некоторое число s отдельных групп. После того как сделано такое подразделение и произведено n независимых испытаний, подсчитывают числа $m_1, m_2, \ldots, m_s$ появлений значений данной величины, принадлежащих соответственно первой, второй, ..., s-й группе. Эмпирическим законом распределения называется совокупность чисел

$$\frac{m_1}{n}, \quad \frac{m_2}{n}, \ldots, \quad \frac{m_s}{n}.$$

Под это определение целиком подходит и то понятие, с которого мы начали, до того, как зашла речь о группировке значений, — понятие эмпирического закона распределения дискретной случайной величины, определяемого по числам $m_1, m_2, \ldots, m_s$ появлений каждого из возможных значений $x_1, x_2, \ldots, x_s$ данной величины. Здесь просто любая группа значений содержит по одному значению.

В современной статистике, однако, мы встречаем и иное понятие эмпирического закона распределения. Это — понятие, которое мы находим у Мизеса. Здесь не производят никакого подразделения значений на отдельные группы, а поступают следующим образом. Пусть x — произвольное число. Произведем n независимых испытаний и подсчитаем число появлений при этом тех значений изучаемой случайной величины, которые меньше x; обозначим это число через $m(x)$. *Эмпирической функцией распределения* $S_n(x)$ данной величины будет называться функция

$$S_n(x) = \frac{m(x)}{n}.$$

Нетрудно видеть, что эта эмпирическая функция распределения всегда — неубывающая функция. При этом всегда $S_n(-\infty) = 0$ и $S_n(+\infty) = 1$.

Для дискретных случайных величин это определение не дает ничего существенно нового. Эмпирическая функция распределения дискретной случайной величины в смысле Мизеса может быть получена из ее ранее определенного закона распределения без группировки значений по формуле

$$S_n(x) = \sum_{x_i < x} \frac{m_i}{n}.$$

Наоборот, эмпирический закон распределения может быть получен из эмпирической функции распределения в смысле Мизеса следующим образом:

$$\frac{m_i}{n} = S_n(x_i + 0) - S_n(x_i - 0).$$

Но эмпирическая функция распределения в смысле Мизеса не сводима к эмпирическим законам распределения, в определении которых участвует объединение различных значений данной величины в группы.

114. Пусть $F(x)$ — теоретическая функция распределения изучаемой случайной величины. Как и для эмпирических функций распределения $F_n(x)$, определенных в предыдущем параграфе, так же точно и для только что определенных эмпирических функций распределения в смысле Мизеса, мы при достаточно большом числе испытаний n будем иметь практическую уверенность в том, что $S_n(x)$ будет как угодно мало отличаться от $F(x)$, притом равномерно на всей прямой $-\infty < x < +\infty$, т. е. в том, что будет как угодно мала величина

$$\sup_x |S_n(x) - F(x)|,$$

где $\sup\limits_x$ обозначает верхнюю грань для всех значений x. В самом деле, пусть ε — сколь угодно малая положительная постоянная. Возьмем целое положительное $s \geqslant \frac{2}{\varepsilon}$ и обозначим через x_i корень уравнения [1])

$$F(x_i) = \frac{i}{s}.$$

Тогда при всяком $i = 1, 2, \ldots, s$

$$F(x_i) - F(x_{i-1}) \leqslant \frac{\varepsilon}{2}.$$

Но очевидно, что при $x_{i-1} < x < x_i$

$$S_n(x_{i-1}) - F(x_i) \leqslant S_n(x) - F(x) \leqslant S_n(x_i) - F(x_{i-1}).$$

Следовательно:

$$\sup_x |S_n(x) - F(x)| \leqslant \max_i |S_n(x_i) - F(x_i)| + \frac{\varepsilon}{2},$$

где $\max\limits_i$ обозначает максимум для всех $i = 0, 1, \ldots, s$. Далее, пусть p_i — вероятность появления такого значения изучаемой случайной величины, которое меньше x_i, и пусть m_i — число появлений таких значений при n независимых испытаниях. Очевидно, что $p_i = F(x_i)$ и $\frac{m_i}{n} = S_n(x_i)$. Следовательно:

$$\max_i |S_n(x_i) - F(x_i)| = \max_i \left|\frac{m_i}{n} - p_i\right|.$$

Таким образом мы будем иметь при достаточно большом n практическую уверенность в том, что будет выполнено неравенство

$$\sup_x |S_n(x) - F(x)| < \varepsilon,$$

если у нас будет при достаточно большом n практическая уверенность в выполнении неравенства

$$\max_i \left|\frac{m_i}{n} - p_i\right| < \frac{\varepsilon}{2}.$$

Это же последнее было доказано в п. 108.

---

[1]) Доказательство проводится для простоты, как и всюду выше, для непрерывных функций $F(x)$, так что корень написанного уравнения всегда существует; и если корней больше одного, то берется наименьший из них.

Соотношения между $S_n(x)$ и $F(x)$ были изучены и более подробно. В частности, В. И. Гливенко было показано, что здесь применяется усиленный закон больших чисел, т. е. на всей прямой $-\infty < x < +\infty$ мы имеем

$$\boldsymbol{P}(\sup_x |S_n(x) - F(x)| \to 0) = 1.$$

А. Н. Колмогоровым найден предельный закон распределения: для непрерывных функций $F(x)$:

$$\boldsymbol{P}\left(\sup_x |S_n(x) - F(x)| < \frac{\lambda}{\sqrt{n}}\right) \to \sum_{-\infty}^{+\infty} (-1)^k e^{-2k^2\lambda^2}.$$

При определении эмпирической функции распределения в смысле Мизеса эмпирические числовые характеристики изучаемой случайной величины получаются следующим образом. Пусть $x_1, x_2, \ldots, x_s$ — различные значения этой величины, которые появились при n независимых испытаниях, и пусть $m_1, m_2, \ldots, m_s$ — числа появлений каждого из этих значений. Так как появлению каждого значения соответствует скачок функции $S_n(x)$ на $\frac{1}{n}$, то эмпирическое среднее значение и эмпирическая дисперсия выражаются здесь такими же формулами, как выше:

$$Ex = \frac{1}{n} \sum_{i=1}^{s} x_i m_i,$$

$$Dx = \frac{1}{n} \sum_{i=1}^{s} (x_i - Ex)^2 m_i.$$

Ясно, что и здесь при достаточно большом числе испытаний мы будем иметь практическую уверенность в том, что названные эмпирические числовые характеристики будут как угодно мало отличаться от соответствующих теоретических характеристик.

§ 29. Сравнение эмпирического среднего значения с теоретическим.

115. В п. 111 и 114 мы утверждали, опираясь на закон больших чисел, что эмпирическое среднее значение Ex и эмпирическая дисперсия Dx дают приближенное представление о среднем значении $E\xi$ и о дисперсии $D\xi$ случайной величины ξ. Практически часто важно знать, насколько надежно это приближение, особенно для среднего значения, которым мы и займемся теперь с этой точки зрения.

Эмпирическое среднее значение

$$Ex = \frac{1}{n} \sum_{i=1}^{s} x_i m_i; \tag{93}$$

есть, очевидно, случайная величина. Нам важно выяснить закон распределения этой случайной величины, хотя бы асимптотический; его можно найти следующим образом.

Чтобы образовать величину Ex мы начинаем с того, что выбираем наугад n раз по одному из значений величины ξ, затем подсчитываем число m_1 появлений значения x_1, число m_2 появлений значения x_2 и т. д. до числа m_s появлений значения x_s; наконец, подставляем все это в формулу (93). Но тот же процесс образования величины Ex из значений величины ξ можно описать и иначе. Обозначим через ξ_1 то из значений $x_1, x_2, \ldots, x_s$, которое попадается при выборке первым. Величина ξ_1, очевидно, есть случайная величина и притом такая же, как ξ, т. е. могущая принимать те же значения $x_1, x_2, \ldots, x_s$ и с теми же вероятностями, что и величина ξ. Обозначим затем через ξ_2 то из значений величины ξ, которое попадается при выборке вторым. Величина ξ_2 снова есть случайная величина, такая же, как ξ. Будем продолжать таким образом до ξ_n — опять-таки такой же случайной величины, как ξ. Нетрудно видеть, что

$$Ex = \frac{1}{n}\sum_{i=1}^{n}\xi_i = \sum_{i=1}^{n}\frac{\xi_i}{n}. \tag{94}$$

Положим теперь для краткости

$$a = \boldsymbol{E}\xi \quad \text{и} \quad b = \boldsymbol{D}\xi.$$

Тогда нормированное отклонение величины Ex будет равно

$$\frac{Ex-a}{\sqrt{\frac{b}{n}}} = \sum_{i=1}^{n}\frac{\frac{\xi_i}{n}-\frac{a}{n}}{\sqrt{\frac{b}{n}}}. \tag{95}$$

Нетрудно видеть, что среднее значение каждого слагаемого суммы (95) равно 0, дисперсия равна $\frac{1}{n}$ и сумма дисперсий равна 1. Кроме того, обозначим через $F(x)$ функцию распределения величины ξ и через $F_{ni}(x)$ — функцию распределения i-го слагаемого суммы (95). Нетрудно видеть, что

$$F_{ni}(x) = F(a + x\sqrt{nb}).$$

Поэтому при любом положительном τ

$$\sum_{i=1}^{n}\int\limits_{|x|>\tau} x^2\,dF_{ni}(x) = n\int\limits_{|x|>\tau} x^2\,dF(a+x\sqrt{nb}) = \frac{1}{b}\int\limits_{|s-a|>\tau\sqrt{nb}} (s-a)^2\,dF(s). \tag{96}$$

Отсюда ясно, что для сумм (95) выполнено условие Линдеберга, так как выражение (96) при $n\to\infty$, очевидно, стремится к нулю. Таким образом асимптотическим законом распределения для сумм (95) является закон Гаусса:

$$(97) \qquad P\left(\frac{Ex-a}{\sqrt{\frac{b}{n}}}<z\right)\sim\frac{2}{\sqrt{2\pi}}\int\limits_{-\infty}^{z}e^{-\frac{x^2}{2}}\,dx,$$

а из этого асимптотического равенства (97) вытекает, что

$$(98) \qquad P(Ex<z)\sim\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\frac{z-a}{\sqrt{\frac{b}{n}}}}e^{-\frac{x^2}{2}}\,dx.$$

Это и есть искомый асимптотический закон распределения величины Ex. Мы видим, что она подчиняется закону Гаусса со средним значением $\boldsymbol{E}\xi$ и с дисперсией $\frac{\boldsymbol{D}\xi}{n}$.

Теперь уже нетрудно оценить, насколько надежно приближенное представление, которое дает эмпирическое среднее значение Ex о среднем значении $\boldsymbol{E}\xi$ случайной величины ξ.

Из формулы (98) мы без труда получаем

$$(99) \qquad P(|Ex-\boldsymbol{E}\xi|<\varepsilon)\sim 2\Phi\left(\frac{\varepsilon}{\sqrt{\frac{\boldsymbol{D}\xi}{n}}}\right).$$

Когда дисперсия $\boldsymbol{D}\xi$, фигурирующая в правой части этой формулы (99), нам не известна, то в силу закона больших чисел мы не сделаем заметной погрешности, если при вычислениях будем заменять ее известной нам эмпирической дисперсией Dx, т. е. если будем пользоваться формулой

$$(100) \qquad P(|Ex-\boldsymbol{E}\xi|<\varepsilon)\sim 2\Phi\left(\frac{\varepsilon}{\sqrt{\frac{Dx}{n}}}\right).$$

Формулы (99) и (100) и дают нужную нам оценку.

116. Покажем, как обычно пользуются формулой (100) на практике. При этом в дальнейшем мы будем полагать для краткости

$$\sigma=\sqrt{\frac{Dx}{n}},$$

а в формуле (100) выразим ε через σ, полагая $\varepsilon=k\sigma$, и будем писать эту формулу в виде

$$(101) \qquad P(|Ex-\boldsymbol{E}\xi|>k\sigma)\sim 1-2\Phi(k).$$

Зададимся столь малой вероятностью, чтобы событие, имеющее эту или еще меньшую вероятность, можно было считать практически невозможным. Пусть такая вероятность есть, например, 0,0027.

Подсчитаем, при каком значении k левая часть формулы (101) будет как раз этой вероятностью 0,0027. Для этого достаточно приравнять числу 0,0027 правую часть $1-2\Phi(k)$ формулы (101) и найти соответствующие значения k по таблице для функции $\Phi(x)$, приложенной в конце книги. Мы получим $k=3$. Неравенство $|Ex-E\xi|>3\sigma$, вероятность которого равна $1-2\Phi(3)$, будет таким образом, практически невозможно.

Пусть теперь та или иная теория дает нам для среднего значения $E\xi$ величину a. Произведем n наблюдений и вычислим из них Ex и σ. Если окажется, что $|Ex-a|>3\sigma$, то это будет свидетельствовать о том, что теоретически определенная величина a среднего значения $E\xi$ невероятна и теорию следует пересмотреть. Если же окажется, что $|Ex-a|<3\sigma$, то наши наблюдения будут в согласии с теорией в том смысле, что они не будут ее опровергать или хотя бы ставить под сомнение.

Приведем один простой пример. Пусть на основании теории вычислено, что вероятность некоторого события A равна $\frac{3}{4}$. Произведем некоторое количество, скажем 580, испытаний, и пусть при этом событие A произойдет 428 раз и не произойдет 152 раза. Поставим в соответствие событию A случайную величину ξ — число наступлений этого события при одном испытании. Это число равно 1, если событие A наступает, и равно 0, если оно не наступает. Таким образом величина ξ может принимать два значения: значение 1 с вероятностью $\frac{3}{4}$ и значение 0 с вероятностью $\frac{1}{4}$, так что $a=\frac{3}{4}$. При описанной же нами совокупности 580 испытаний и их результатов величина ξ примет значение 1 в 428 случаях и значение 0 в 152 случаях. Отсюда, как нетрудно подсчитать, $Ex=0{,}74$ и $3\sigma=0{,}055$. Таким образом $|Ex-a|<3\sigma$, и мы можем утверждать, что вычисленная теоретически величина a, а следовательно, и вероятность события A, не противоречит результатам наших испытаний.

Эмпирическая числовая характеристика σ называется *средним квадратическим отклонением*. Практически ее удобно вычислять, не производя предварительного вычисления Dx, а прямо по формуле

$$\sigma=\frac{1}{n}\sqrt{\sum_{i=1}^{n}(x_i-Ex)\,m_i},$$

которая легко получается из определения σ и из формулы (92) для Dx.

За величину вероятности $1-2\Phi(k)$ чаще всего принимают такую величину, которой соответствует целое значение k: 1, 2, 3 или 4. Вот эти вероятности:

$$\begin{aligned}1-2\Phi(1)&=0{,}3174;\\1-2\Phi(2)&=0{,}0456;\\1-2\Phi(3)&=0{,}0027;\\1-2\Phi(4)&=0{,}000064.\end{aligned}$$

Но иногда поступают и наоборот: принимают за величину вероятности $1-2\Phi(k)$ число, имеющее простое выражение в десятичной системе, например, 0,1; 0,01; 0,001; 0,0001 и т. п. Этим числам соответствуют следующие значения k:

$$\begin{aligned}
&\text{если } 1-2\Phi(k)=0{,}1, &&\text{то } k=1{,}645;\\
&\text{если } 1-2\Phi(k)=0{,}01, &&\text{то } k=2{,}57;\\
&\text{если } 1-2\Phi(k)=0{,}001, &&\text{то } k=3{,}31;\\
&\text{если } 1-2\Phi(k)=0{,}0001, &&\text{то } k=3{,}91.
\end{aligned}$$

§ 30. Корреляция.

117. До сих пор мы интересовались преимущественно независимыми событиями и независимыми случайными величинами. Нельзя, однако, обойти молчанием один чрезвычайно важный для статистики вопрос о зависимых событиях и зависимых случайных величинах, а именно — вопрос о возможности измерения зависимости между событиями или величинами, т. е. о выражении степени этой зависимости числом.

Начнем с событий. Возьмем два любых события A и B. Мы знаем, что, для того чтобы эти события были независимы, необходимо и достаточно, чтобы выполнялось равенство

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

или, что то же:

$$\boldsymbol{P}(AB)-\boldsymbol{P}(A)\,\boldsymbol{P}(B)=0.$$

Таким образом наличие зависимости между событиями A и B характеризуется тем, что разность $\boldsymbol{P}(AB)-\boldsymbol{P}(A)\,\boldsymbol{P}(B)$ не равна нулю. Отсюда естественно считать меру зависимости между событиями пропорциональной этой разности. Эту меру называют *коэфициентом корреляции* [1]) между событиями A и B, обозначают через $\boldsymbol{R}_{AB}$ и определяют так:

$$\boldsymbol{R}_{AB}=\frac{\boldsymbol{P}(AB)-\boldsymbol{P}(A)\,\boldsymbol{P}(B)}{\sqrt{\boldsymbol{P}(A)\,[1-\boldsymbol{P}(A)]\,\boldsymbol{P}(B)\,[1-\boldsymbol{P}(B)]}}. \tag{102}$$

Представим себе, например, что мы бросаем игральную кость три раза подряд, и пусть событие A состоит в выпадении при втором бросании вдвое большего числа очков, нежели при первом бросании, а событие B — в выпадении при третьем бросании вдвое большего числа очков, нежели при первом. Подсчитывая число результатов испытания, благоприятствующих событиям A, B и AB из общего числа 216 возможных результатов испытания, мы получаем соответственно 18, 18 и 3. Следовательно:

$$\boldsymbol{P}(A)=\frac{1}{12},\quad \boldsymbol{P}(B)=\frac{1}{12},\quad \boldsymbol{P}(AB)=\frac{1}{72}.$$

Подставляя эти значения в формулу (102), получаем

$$\boldsymbol{R}_{AB}=\frac{1}{11}.$$

---

[1]) Correlatio (лат.) — связь.

Видоизменим теперь наш пример следующим образом. Будем снова бросать игральную кость три раза подряд, и пусть попрежнему событие A состоит в выпадении при втором бросании вдвое большего числа очков, нежели при первом бросании, но событие B — в выпадении при третьем бросании вдвое большего числа очков, нежели при втором, а не при первом бросании, как это было выше. Подсчитывая снова число результатов испытания, благоприятствующих событиям A, B и AB, мы получаем соответственно 18, 18 и 1. Следовательно, в этом случае

$$\boldsymbol{P}(A)=\frac{1}{12}, \quad \boldsymbol{P}(B)=\frac{1}{12}, \quad \boldsymbol{P}(AB)=\frac{1}{216}.$$

Отсюда по формуле (102) получаем

$$\boldsymbol{R}_{AB}=-\frac{1}{33}.$$

Коэфициент корреляции между двумя событиями положителен, если вероятность одного из этих событий увеличивается после наступления другого; напротив, коэфициент корреляции отрицателен, если вероятность одного из событий уменьшается после наступления другого. В самом деле, знак коэфициента корреляции $\boldsymbol{R}_{AB}$, очевидно, совпадает со знаком его числителя:

$$d=\boldsymbol{P}(AB)-\boldsymbol{P}(A)\,\boldsymbol{P}(B).$$

Но мы знаем, что $\boldsymbol{P}(AB)=\boldsymbol{P}(A)\,\boldsymbol{P}(B/A)$. Отсюда

$$d=\boldsymbol{P}(A)\,[\boldsymbol{P}(B/A)-\boldsymbol{P}(B)],$$

что и доказывает наше утверждение.

Так, если обратиться к только что приведенным примерам троекратного бросания игральной кости, то оказывается, что в случае, если при втором бросании выпало вдвое больше очков, нежели при первом, вероятность получить и при третьем бросании вдвое больше очков, нежели при первом, становится больше, чем была до начала испытания, так как здесь коэфициент корреляции положителен. Напротив, в том же случае вероятность получить при третьем бросании вдвое больше очков, нежели при втором, уменьшается, так как здесь коэфициент корреляции отрицателен.

118. Коэфициент корреляции (102) можно представить также в другой форме, если привлечь к рассмотрению события $\overline{A}$ и $\overline{B}$, противоположные событиям A и B, а именно, можно написать

$$\boldsymbol{R}_{AB}=\frac{\begin{vmatrix}\boldsymbol{P}(AB) & \boldsymbol{P}(\overline{A}B)\\ \boldsymbol{P}(A\overline{B}) & \boldsymbol{P}(\overline{A}\overline{B})\end{vmatrix}}{\sqrt{\boldsymbol{P}(A)\,\boldsymbol{P}(\overline{A})\,\boldsymbol{P}(B)\,\boldsymbol{P}(\overline{B})}}. \tag{103}$$

Легко видеть, как получается знаменатель выражения (103) из знаменателя выражения (102). Числитель же выражения (103) получается из числителя выражения (102) следующим образом. Прежде всего, ясно, что событие A подразделяется на частные случаи AB и $A\overline{B}$, а событие B — на частные случаи AB и $\overline{A}B$. Поэтому

$$\boldsymbol{P}(B)=\boldsymbol{P}(AB)+\boldsymbol{P}(A\overline{B}),$$

$$\boldsymbol{P}(B)=\boldsymbol{P}(AB)+\boldsymbol{P}(\overline{A}B).$$

Отсюда

$$\boldsymbol{P}(A)\,\boldsymbol{P}(B)=\boldsymbol{P}(AB)\,[\boldsymbol{P}(AB)+\boldsymbol{P}(A\overline{B})+\boldsymbol{P}(\overline{A}B)]+\\+\boldsymbol{P}(A\overline{B})\,\boldsymbol{P}(\overline{A}B).$$

С другой стороны, ясно, что если мы возьмем событие $\overline{A}\overline{B}$, то противоположным ему будет событие $AB+A\overline{B}+\overline{A}B$, и что это последнее подразделяется на частные случаи AB, $A\overline{B}$ и $\overline{A}B$. Поэтому

$$\boldsymbol{P}(\overline{A}\overline{B})=1-[\boldsymbol{P}(AB)+\boldsymbol{P}(A\overline{B})+\boldsymbol{P}(\overline{A}B)].$$

Отсюда и из предыдущей формулы

$$\boldsymbol{P}(A)\,\boldsymbol{P}(B)=\boldsymbol{P}(AB)\,[1-\boldsymbol{P}(\overline{A}\overline{B})]+\boldsymbol{P}(A\overline{B})\,\boldsymbol{P}(\overline{A}B).$$

Наконец, отсюда уже без труда вытекает то, что нам нужно:

$$\boldsymbol{P}(AB)-\boldsymbol{P}(A)\,\boldsymbol{P}(B)=\boldsymbol{P}(AB)\,\boldsymbol{P}(\overline{A}\overline{B})-\boldsymbol{P}(A\overline{B})\,\boldsymbol{P}(\overline{A}B).$$

Из формулы (123) непосредственно видно, что

$$\boldsymbol{R}_{AB}=\boldsymbol{R}_{\overline{A}\overline{B}}\boldsymbol{R}=-\boldsymbol{R}_{A\overline{B}}=-\boldsymbol{R}_{\overline{A}B}.$$

Формула (103) дает возможность убедиться в том, что всегда имеет место неравенство

(104) $$|\boldsymbol{R}_{AB}|\leqslant 1.$$

Действительно, в силу формулы (123)

(105) $$\boldsymbol{R}_{AB}^2=\frac{\begin{vmatrix}\boldsymbol{P}(AB) & \boldsymbol{P}(\overline{A}B)\\ \boldsymbol{P}(A\overline{B}) & \boldsymbol{P}(\overline{A}\overline{B})\end{vmatrix}}{\boldsymbol{P}(A)\,\boldsymbol{P}(\overline{A})}\cdot\frac{\begin{vmatrix}\boldsymbol{P}(BA) & \boldsymbol{P}(B\overline{A})\\ \boldsymbol{P}(\overline{B}A) & \boldsymbol{P}(\overline{B}\overline{A})\end{vmatrix}}{\boldsymbol{P}(B)\,\boldsymbol{P}(\overline{B})}.$$

Это последнее равенство можно написать еще так:

(106) $$\boldsymbol{R}_{AB}^2=\begin{vmatrix}\boldsymbol{P}(B/A) & \boldsymbol{P}(\overline{B}/A)\\ \boldsymbol{P}(\overline{B}/A) & \boldsymbol{P}(\overline{B}/\overline{A})\end{vmatrix}\cdot\begin{vmatrix}\boldsymbol{P}(A/B) & \boldsymbol{P}(\overline{A}/B)\\ \boldsymbol{P}(A/\overline{B}) & \boldsymbol{P}(\overline{A}/\overline{B})\end{vmatrix}.$$

Каждый же из этих последних детерминантов по модулю не превосходит единицы, так как все элементы этих детерминантов являются условными вероятностями и, следовательно, неотрицательны и не превосходят единицы.

Оба множителя в правой части равенства (105) называются *коэфициентами регрессии*: первый — коэфициентом регрессии события B относительно A (обозначается через ρ_B), второй — события A относительно B (обозначается через ρ_A). Из равенства (105) немедленно вытекает, что

(107) $$R_{AB} = \pm\sqrt{\rho_A \rho_B}.$$

119. Перейдем к случайным величинам. Возьмем две случайные величины ξ и η. Мы знаем, что в том случае, когда эти величины независимы, имеет место равенство.

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

Отсюда естественно считать меру зависимости между любыми двумя случайными величинами ξ и η пропорциональной разности $E(\xi\eta) - E\xi \cdot E\eta$. Эту меру называют *коэфициентом корреляции* между величинами ξ и η, обозначают через $R_{\xi,\eta}$ и определяют так:

(108) $$R_{\xi,\eta} = \frac{E(\xi\eta) - E\xi \cdot E\eta}{\sqrt{D\xi \cdot D\eta}}.$$

Возьмем, например, две независимые случайные величины ξ и η и вычислим коэфициент корреляции между ξ и $\xi - \eta$. Так как в этом случае

$$E[\xi(\xi-\eta)] = E\xi^2 - E\xi \cdot E\eta,$$

$$E\xi \cdot E(\xi-\eta) = (E\xi)^2 - E\xi \cdot E\eta,$$

$$D\xi \cdot D(\xi-\eta) = D\xi(D\xi + D\eta),$$

то по формуле (108)

$$R_{\xi,\xi-\eta} = \frac{\sqrt{D\xi}}{\sqrt{D\xi + D\eta}}.$$

Возьмем теперь те же величины ξ и η и вычислим коэфициент корреляции между ξ и $\eta - \xi$. Мы получим

$$R_{\xi,\eta-\xi} = -\frac{\sqrt{D\xi}}{\sqrt{D\xi + D\eta}}.$$

Коэфициент корреляций между случайными величинами обладает одним важным свойством, которое в случае событий не могло быть формулировано, а именно, представим себе, что мы подвергли величины ξ и η линейному преобразованию вида

$$\varphi = a\xi + c,$$

$$\psi = b\eta + d,$$

где a и b — положительные постоянные, c и d — произвольные постоянные. Оказывается, что такое преобразование не изменяет коэфициента корреляции, т. е. что

(109) $$R_{\varphi,\psi} = R_{\xi,\eta}.$$

В этом легко убедиться непосредственным вычислением. В самом деле,

$$R_{\varphi, \psi} = \frac{E(ab\xi\eta + ad\xi + bc\eta + cd) - E(a\xi + c)\, E(b\eta + d)}{\sqrt{D(a\xi + c)\, D(b\eta + d)}} =$$

$$= \frac{ab\,[E(\xi\eta) - E\xi \cdot E\eta]}{\sqrt{a^2 D\xi \cdot b^2 D\eta}} = R_{\xi, \eta}.$$

Таким образом при вычислении коэфициента корреляции между двумя случайными величинами мы можем выражать каждую из них в любых единицах измерения и пользоваться любым началом отсчета.

Что касается знака коэфициента корреляции, то можно установить следующее общее предложение:

Пусть ξ и η — две случайные величины и пусть $E(\eta/x)$ — *условное среднее значение* величины η, вычисленное в предположении, что величина ξ имеет значение x. Коэфициент корреляции между ξ и η *положителен*, если существует такое число l, что $E(\eta/x)$ больше $E\eta$, когда $x > l$, и меньше $E\eta$, когда $x \leqslant l$; наоборот, коэфициент корреляции *отрицателен*, если существует такое l, что $E(\eta/x)$ меньше $E\eta$, когда $x > l$, и больше $E\eta$, когда $x \leqslant l$.

Чтобы убедиться в этом, заметим прежде всего, что ничего не изменится, если мы заменим величину ξ величиной $\xi - l$. Действительно, как мы только что видели, коэфициенты корреляции $R_{\xi, \eta}$ и $R_{\xi - l, \eta}$ равны между собой и, следовательно, одновременно положительны или отрицательны. Но когда величина ξ превосходит или, наоборот, не превосходит числа l, то величина $\xi - l$ соответственно превосходит или, наоборот, не превосходит нуля. Таким образом мы можем, не нарушая общности дальнейших выводов, положить $l = 0$.

Знак коэфициента корреляции $R_{\xi, \eta}$ совпадает со знаком его числителя

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

Предположим для простоты, что ξ и η — дискретные случайные величины. Пусть $P_\xi(x)$ и $P_\eta(y)$ — соответственно законы распределения величин ξ и η и пусть $P(x; y)$ — вероятность того, что ξ примет значение x и одновременно η — значение y. По теореме умножения вероятностей $P(x; y) = P_\xi(x)\, P_\eta(y/x)$, где $P_\eta(y/x)$ есть *условный закон распределения* η, устанавливаемый в предположении, что ξ имеет значение x. Поэтому

$$E(\xi\eta) = \sum_x \sum_y xyP(x;\ y) = \sum_x xP_\xi(x)\, [\sum_y yP_\eta(y/x)] =$$

$$= \sum_x xP_\xi(x)\, E(\eta/x).$$

Затем

$$E\xi \cdot E\eta = \sum_x xP_\xi(x)\, E\eta.$$

Отсюда

(110) $$\delta = \sum_x xP_\xi(x)\, [E(\eta/x) - E\eta].$$

Пусть теперь всякий раз, как величина ξ принимает определенное значение

$$x > 0,$$

$E(\eta/x)$ становится больше $E\eta$, т. е.

$$E(\eta/x) - E\eta > 0,$$

и всякий раз, как величина ξ принимает значение

$$x < 0,$$

$E(\eta/x)$ становится меньше $E\eta$, т. е.

$$E(\eta/x) - E\eta < 0.$$

При этих предположениях все слагаемые в правой части равенства (110) будут положительны, и следовательно:

$$\delta > 0.$$

Обратные же предположения привели бы таким же путем к неравенству

$$\delta < 0.$$

Тем самым наше утверждение о знаке коэфициента корреляции полностью доказано.

Отсюда можно сделать следующий важный вывод: коэфициент корреляции между величинами ξ и η положителен, если $E(\eta/x)$ есть возрастающая функция от x; напротив, коэфициент корреляции между ξ и η отрицателен, если $E(\eta/x)$ есть убывающая функция от x.

В самом деле, пусть $E(\eta/x)$ — возрастающая функция от x. Тогда существует число l такое, что при $x > l$

$$E(\eta/x) > E\eta,$$

а при $x < l$

$$E(\eta/x) < E\eta.$$

Поэтому $R_{\xi, \eta} > 0$.

Аналогичным образом доказывается, что если $E(\eta/x)$ — убывающая функция от x, то $R_{\xi, \eta} < 0$.

Возьмем, например, приведенный выше случай, когда даны две независимые случайные величины ξ и η и ищется коэфициент корреляции между ξ и $\xi - \eta$, а также коэфициент корреляции между ξ и $\eta - \xi$. Так как

$$E(\xi - \eta/x) = E(x - \eta) = x - E\eta,$$

то $E(\xi - \eta/x)$ есть возрастающая функция от x. Следовательно, коэфициент корреляции $R_{\xi, \xi-\eta}$ должен быть положительным. Напротив, так как

$$E(\eta - \xi/x) = E(\eta - x) = E\eta - x,$$

то $E(\eta - \xi/x)$ есть убывающая функция от x, и коэфициент корреляции $R_{\xi, \eta-\xi}$ должен быть отрицательным. Приведенные выше вычисления этих коэфициентов корреляции находятся в полном согласии с нашими выводами.

120. Покажем, что введенный нами ранее коэфициент корреляции (102) между событиями можно рассматривать как частный случай коэфициента корреляции (108) между случайными величинами.

Условимся ставить в соответствие каждому событию C случайную величину $\chi(C)$ — число наступлений этого события при одном испытании. Это число равно 1, если событие C наступает, и равно 0, если оно не наступает. Таким образом величина $\chi(C)$ может принимать два значения: значение 1 с вероятностью $\boldsymbol{P}(C)$ и значение 0 с вероятностью $1-\boldsymbol{P}(C)$.

Положим теперь $\xi=\chi(A)$ и $\eta=\chi(B)$. Нетрудно видеть, что тогда $\xi\eta=\chi(AB)$. Затем, нетрудно подсчитать, что

$$\boldsymbol{E}\xi=\boldsymbol{P}(A),\quad \boldsymbol{E}\eta=\boldsymbol{P}(B),\quad \boldsymbol{E}(\xi\eta)=\boldsymbol{P}(AB),$$

$$\boldsymbol{D}\xi=\boldsymbol{P}(A)[1-\boldsymbol{P}(A)],\quad \boldsymbol{D}\eta=\boldsymbol{P}(B)[1-\boldsymbol{P}(B)].$$

Подставив эти данные в правую часть равенства (108), мы получим правую часть равенства (102).

121. Коэфициент корреляции (108) можно выразить в очень удобной для многих вычислений форме, если привлечь к рассмотрению отклонения $\bar{\xi}=\xi-\boldsymbol{E}\xi$ и $\bar{\eta}=\eta-\boldsymbol{E}\eta$ случайных величин ξ и η, а именно, можно написать

(111) $$\boldsymbol{R}_{\xi,\eta}=\frac{\boldsymbol{E}(\bar{\xi}\bar{\eta})}{\sqrt{\boldsymbol{E}\bar{\xi}^2\cdot\boldsymbol{E}\bar{\eta}^2}},$$

что нетрудно проверить вычислением.

Формула (111) дает возможность легко убедиться в том, что для коэфициентов корреляции случайных величин остается в силе неравенство

(112) $$|\boldsymbol{R}_{\xi,\eta}|\leqslant 1.$$

Действительно в силу формулы (111)

$$\boldsymbol{E}\left(\frac{\bar{\xi}}{\sqrt{\boldsymbol{E}\bar{\xi}^2}}\pm\frac{\bar{\eta}}{\sqrt{\boldsymbol{E}\bar{\eta}^2}}\right)^2=1\pm 2\boldsymbol{R}_{\xi,\eta}+1.$$

Но левая часть этого последнего равенства, очевидно, неотрицательна. Следовательно, неотрицательна и его правая часть. В силу же этого последнего обстоятельства

$$-1\leqslant\boldsymbol{R}_{\xi,\eta}\leqslant 1.$$

Из формулы (111) вытекает, что

(113) $$\boldsymbol{R}^2_{\xi,\eta}=\frac{\boldsymbol{E}(\bar{\xi}\bar{\eta})}{\boldsymbol{E}\bar{\xi}^2}\cdot\frac{\boldsymbol{E}(\bar{\eta}\bar{\xi})}{\boldsymbol{E}\bar{\eta}^2}.$$

Оба множителя в правой части равенства (113) называются *коэфи-*

циентами регрессии: первый — коэфициентом регрессии величины η относительно ξ (обозначается ρ_η), второй — величины ξ относительно η (обозначается ρ_ξ). Из равенства (113) немедленно вытекает, что, как и в случае событий:

(114) $$R_{\xi,\eta} = \pm\sqrt{\rho_\xi \rho_\eta}.$$

122. Коэфициенты регрессии могут иметь самостоятельное значение в описании связи между случайными величинами. Это особенно ясно в практически важном случае так называемой линейной корреляции, которую мы сейчас определим.

Величина $\boldsymbol{E}(\eta/x)$ является функцией от x, величина $\boldsymbol{E}(\xi/y)$ — функцией от y. Кривые, заданные уравнениями

$$y = \boldsymbol{E}(\eta/x), \quad x = \boldsymbol{E}(\xi/y),$$

называются *линиями регрессии*: первая — линия регрессии величины η относительно ξ, вторая — величина ξ относительно η.

Пусть теперь $\boldsymbol{E}(\eta/x)$ и $\boldsymbol{E}(\xi/y)$ — линейные фукции, т. е.

$$\boldsymbol{E}(\eta/x) = ax + c, \quad \boldsymbol{E}(\xi/y) = by + d,$$

где a, b, c, d — постоянные; иначе говоря, пусть обе линии регрессии — прямые. В таком случае мы будем говорить, что *корреляция между величинами ξ и η линейная*.

В этом случае, как мы сейчас увидим:

$$\rho_\eta = a, \quad \rho_\xi = b.$$

Иначе говоря, коэфициенты регрессии являются угловыми коэфициентами линий регрессии.

Чтобы убедиться в этом, заметим прежде всего, что в нашем случае в силу формулы полной вероятности

$$\boldsymbol{E}\eta = \sum_y y P_\eta(y) = \sum_x P_\xi(x) \left[\sum_y y P_\eta(y/x)\right] =$$
$$= \sum_x P_\xi(x)\, \boldsymbol{E}(\eta/x) = a \sum_x x P_\xi(x) + c \sum_x P_\xi(x) = a\boldsymbol{E}\xi + c.$$

Поэтому

$$\boldsymbol{E}(\bar{\eta}/x) = \boldsymbol{E}(\eta/x) - \boldsymbol{E}\eta = (ax + c) - (a\boldsymbol{E}\xi + c) = a(x - \boldsymbol{E}\xi).$$

Поэтому

$$\boldsymbol{E}(\bar{\xi}\bar{\eta}) = \sum_x \sum_y (x - \boldsymbol{E}\xi)(y - \boldsymbol{E}\eta) P(x;\, y) =$$
$$= \sum_x (x - \boldsymbol{E}\xi) P_\xi(x) \left[\sum_y (y - \boldsymbol{E}\eta) P_\eta(y/x)\right] =$$
$$= \sum_x (x - \boldsymbol{E}\xi) P_\xi(x)\, \boldsymbol{E}(\bar{\eta}/x) = a \sum_x (x - \boldsymbol{E}\xi)^2 P_\xi(x) = a\boldsymbol{E}\bar{\xi}^2.$$

Отсюда уже непосредственно вытекает, что $\rho_\eta = a$. Точно так же $\rho_\xi = b$.

Важность случая линейной корреляции можно иллюстрировать хотя бы следующим предложением, которое мы приведем без доказательства. Пусть случайные величины ξ и η подчиняются нормальному закону распределения, т. е. при каждом данном значении x величины ξ условный закон распределения величины η есть закон Гаусса (с параметрами α и β, зависящими от x), и обратно, при каждом данном значении y величины η условный закон распределения величины ξ есть закон Гаусса (с параметрами α и β, зависящими от y); тогда корреляция между величинами ξ и η необходимо линейная.

123. Коэфициент корреляции не характеризует полностью зависимости между случайными величинами. Чтобы убедиться в этом, достаточно привести такой пример. Пусть ξ — случайная величина с равномерным законом распределения в интервале $(-1, +1)$ (см. § 16 п. 59) и пусть

$$\eta = \xi^2 - \frac{1}{3}.$$

Здесь связь между ξ и η очень тесная: это — точная функциональная зависимость. И тем не менее, здесь, как и в случае независимых случайных величин:

$$\boldsymbol{R}_{\xi, \eta} = 0.$$

Действительно:

$$\boldsymbol{E}(\xi\eta) = \frac{1}{2}\int_{-1}^{+1} x\left(x^2 - \frac{1}{3}\right)dx = 0.$$

$$\boldsymbol{E}\xi = \frac{1}{2}\int_{-1}^{+1} x\,dx = 0, \quad \boldsymbol{E}\eta = \frac{1}{2}\int_{-1}^{+1}\left(x^2 - \frac{1}{3}\right)dx = 0.$$

Таким образом числитель в правой части формулы коэфициента корреляции (108) равен нулю.

Поэтому кроме коэфициента корреляции предлагались и другие меры зависимости между случайными величинами, более чувствительно реагирующие на всякую связь между последними.

Однако коэфициент корреляции остается наиболее распространенной характеристикой во всех приложениях теории вероятностей. Объясняется это тем, что в основном для этих приложений вопросе — в вопросе о применимости закона больших чисел и предельных законов распределения — часто оказывается имеющей значение не зависимость между случайными величинами, как таковая, а только та сторона этой зависимости, которая измеряется коэфициентом корреляции.

В основе этого лежит большею частью тот факт, что дисперсия суммы нескольких, вообще говоря зависимых, случайных величин может быть выражена через дисперсии слагаемых и коэфициенты корреляции между слагаемыми. А именно, мы имеем формулу

$$\boldsymbol{D}(\xi_1 + \ldots + \xi_n) = \sum_{i, k=1}^{n} \boldsymbol{R}_{\xi_i, \xi_k} \sqrt{\boldsymbol{D}\xi_i}\sqrt{\boldsymbol{D}\xi_k}. \tag{115}$$

Или иначе, имея в виду, что коэфициент корреляции между каждым слагаемым и им самим равен единице:

$$\boldsymbol{D}(\xi_1 + \ldots + \xi_n) = \boldsymbol{D}\xi_1 + \ldots + \boldsymbol{D}\xi_n + \sum_{i \neq k} \boldsymbol{R}_{\xi_i, \xi_k} \sqrt{\boldsymbol{D}\xi_i}\sqrt{\boldsymbol{D}\xi_k}.$$

Для двух слагаемых ξ и η эта формула доказывается следующим образом:

$$\begin{aligned} D(\xi+\eta) &= E(\xi+\eta)^2-(E\xi+E\eta)^2 = \\ &= E\xi^2+E\eta^2+2E(\xi\eta)-(E\xi)^2-(E\eta)^2-2E\xi\cdot E\eta = \\ &= D\xi+D\eta+2[E(\xi\eta)-E\xi\cdot E\eta] = \\ &= D\xi+D\eta+2R_{\xi,\eta}\sqrt{D\xi\cdot D\eta}. \end{aligned}$$

Доказательство для любого числа слагаемых $\xi_1, \ldots, \xi_n$ такое же, но только вместо равенства $(\xi+\eta)^2=\xi^2+\eta^2+2\xi\eta$, естественно, следует применить равенство

$$(\xi_1+\ldots+\xi_n)^2=\sum_{i=1}^{n}\xi_i^2+\sum_{i\neq k}\xi_i\xi_k.$$

Приведем примеры применения коэфициентов корреляции к предельным теоремам.

Условимся называть последовательность случайных величин *однородно коррелированной*, если коэфициент корреляции R_{ξ_i,ξ_k} между любыми величинами этой последовательности зависит только от расстояния этих величин в последовательности, т. е. от модуля разности их индексов, так что

$$R_{\xi_i,\xi_k}=R(|i-k|).$$

Функция $R(x)$, определенная для целых неотрицательных значений аргумента, называется *корреляционной функцией* последовательности.

Теорема С. Н. Бернштейна. — *Пусть $\xi_1, \ldots, \xi_n, \ldots$ однородно коррелированная последовательность случайных величин, удовлетворяющая следующим условиям:*

1° *Все величины $\xi_1, \ldots, \xi_n, \ldots$ имеют одно и то же среднее значение:*

$$E\xi_1=\ldots=E\xi_n=\ldots=a.$$

2° *Дисперсии величин $\xi_1, \ldots, \xi_n, \ldots$ равномерно ограничены, т. е.*

$$D\xi_1\leqslant C, \ \ldots, \ D\xi_n\leqslant C, \ \ldots,$$

где C — постоянная.

3° *Корреляционная функция стремится к нулю:*

$$\lim_{x\to\infty}R(x)=0.$$

Тогда, какова бы ни была положительная постоянная ε:

$$\lim_{n\to\infty}P\left(\left|\frac{\xi_1+\ldots+\xi_n}{n}-a\right|<\varepsilon\right)=1.$$

Доказательство. В силу второй формы неравенства Чебышева

$$(116)\qquad P\left(\left|\frac{\xi_1+\ldots+\xi_n}{n}-a\right|<\varepsilon\right)\geqslant 1-\frac{D\left(\frac{\xi_1+\ldots+\xi_n}{n}\right)}{\varepsilon^2}$$

и, далее, в силу формулы (115)

(117) $$D\left(\frac{\xi_1+\ldots+\xi_n}{n}\right)=\frac{1}{n^2}D(\xi_1+\ldots+\xi_n)=$$
$$=\frac{1}{n^2}\sum_{i,\,k=1}^{n}R(|i-k|)\sqrt{D\xi_i}\,\sqrt{D\xi_k}.$$

Возьмем теперь произвольно малое положительное число δ, и пусть m — целое положительное число, такое, что $|R(x)|<\delta$ при $x>m$. Разобьем последнюю сумму в равенствах (117) на две части: 1) на сумму тех слагаемых, где $|i-k|\leqslant m$, и 2) на сумму тех, где $|i-k|>m$. Слагаемых в первой из этих частей будет, очевидно, не больше $2nm$, так что вся эта часть не превзойдет $2nmC$; слагаемых же во второй части будет не больше n^2, так что вся она не превзойдет $n^2\delta$. Итак, все последнее выражение в равенствах (117) не превосходит

(118) $$\frac{1}{n^2}(2nmC+n^2\delta)=\frac{2mC}{n}+\delta.$$

Сопоставляя (116), (117) и (118), мы видим, что

$$P\left(\left|\frac{\xi_1+\ldots+\xi_n}{n}-a\right|<\varepsilon\right)\geqslant 1-\frac{2mC}{n\varepsilon^2}-\frac{\delta}{\varepsilon^2},$$

и следовательно:

$$\lim_{n\to\infty}P\left(\left|\frac{\xi_1+\ldots+\xi_n}{n}-a\right|<\varepsilon\right)\geqslant 1-\frac{\delta}{\varepsilon^2}.$$

Наконец, отсюда, принимая во внимание, что δ произвольно мало, мы и получаем непосредственно то, что требовалось доказать.

Самим С. Н. Бернштейном было доказано несколько более общее предложение. Только что установленная теорема воспроизводит лишь наиболее существенную его часть.

Другим примером роли коэфициентов корреляции может служить следующее предложение, которое мы приведем без доказательства. Пусть даны две последовательности случайных величин:

$$\xi_1',\ \xi_2',\ \ldots,\ \xi_n',\ \ldots,$$
$$\xi_1'',\ \xi_2'',\ \ldots,\ \xi_n'',\ \ldots,$$

причем каждая из этих последовательностей в отдельности состоит из независимых величин и удовлетворяет условию Линдеберга. Кроме того, пусть

$$\zeta_n'=\xi_1'+\xi_2'+\ldots+\xi_n',$$
$$\zeta_n''=\xi_1''+\xi_2''+\ldots+\xi_n'',$$

затем

$$A_n'=E\zeta_n',\ B_n'=D\zeta_n',$$
$$A_n''=E\zeta_n'',\ B_n''=D\zeta_n'',$$

наконец:

$$R_n=R_{\zeta_n',\,\zeta_n''}.$$

причем отношения $\frac{B_i'}{B_i''}$ заключены между двумя положительными числами, а $|R_n|$ не превосходит не зависящей от n постоянной, меньшей единицы. Тогда, если обозначить через $F_n(z'; z'')$ двумерную функцию распределения величин ζ_n' и ζ_n'', то

$$\lim_{n\to\infty}\left\{F_n(z'; z'') - \frac{1}{2\pi\sqrt{B_n' B_n''(1-R_n^2)}}\int_{-\infty}^{z'}\int_{-\infty}^{z''}\exp\frac{-1}{2(1-R_n^2)}\left[\frac{(x'-A_n')^2}{B_n'}+\right.\right.$$
$$\left.\left.+\frac{(x''-A_n'')^2}{B_n''} - 2R_n\frac{(x'-A_n')(x''-A_n'')}{\sqrt{B_n'}\sqrt{B_n''}}\right]dx'dx''\right\} = 0,$$

где мы пишем $\exp u$ вместо e^u.

124. Часто пользуются сокращенными обозначениями, полагая

$$\sigma_\xi^2 = \boldsymbol{E}\bar{\xi}^2, \quad \sigma_\eta^2 = \boldsymbol{E}\bar{\eta}^2, \quad \tau_{\xi,\eta} = \boldsymbol{E}(\bar{\xi}\,\bar{\eta}).$$

Если применить эти сокращенные обозначения, то формула (111) для коэфициента корреляции примет вид

$$\boldsymbol{R}_{\xi,\eta} = \frac{\tau_{\xi,\eta}}{\sigma_\xi\,\sigma_\eta},$$

а коэфициенты регрессии выразятся следующим образом:

$$\rho_\xi = \frac{\tau_{\xi,\eta}}{\sigma_\eta^2}, \quad \rho_\eta = \frac{\tau_{\xi,\eta}}{\sigma_\xi^2},$$

или, что, как нетрудно видеть, то же самое:

$$\rho_\xi = \boldsymbol{R}_{\xi,\eta}\frac{\sigma_\xi}{\sigma_\eta}, \quad \rho_\eta = \boldsymbol{R}_{\xi,\eta}\frac{\sigma_\eta}{\sigma_\xi}.$$

125. Посмотрим, как определяется корреляция в случае эмпирических законов распределения.

Пусть даны какие-нибудь случайные величины ξ и η. Заменим их, как мы это делали в п. 109, дискретными случайными величинами ξ^* и η^*, которые могут принимать значения $x_1, x_2, \ldots, x_s$ и $y_1, y_2, \ldots, y_t$. Пусть p_{11} — вероятность того, что ξ^* примет значение x_1 и одновременно η^* — значение y_1; затем p_{12} — вероятность того, что ξ^* примет значение x_1 и одновременно η^* — значение y_2; и т. д. Иначе говоря, пусть теоретический закон распределения пары величин ξ^* и η^* выражается табл. VI (см. стр. 202).

Установим теперь эмпирический закон распределения пары величин ξ^* и η^*, т. е. произведем n независимых испытаний, в результате каждого из которых ξ^* и η^* будут принимать определенные значения, и отметим число m_{11} испытаний, при которых ξ^* приняла значение x_1, а η^* — значение y_1, затем число m_{12} испытаний, при которых ξ^* при-

Таблица VI.

	x_1	x_2	...	x_s
y_1	p_{11}	p_{21}	...	p_{s1}
y_2	p_{12}	p_{22}	...	p_{s2}
...	...	...	...	...
y_t	p_{1t}	p_{2t}	...	p_{st}

няла значение x_1, а η^* — значение y_2, и т. д. Мы получим так называемую корреляционную таблицу (табл. VII).

Таблица VII.

	x_1	x_2	...	x_s
y_1	m_{11}	m_{21}	...	m_{s1}
y_2	m_{12}	m_{22}	...	m_{s2}
...	...	...	...	...
y_t	m_{1t}	m_{2t}	...	m_{st}

При достаточно большом числе n произведенных испытаний мы будем иметь практическую уверенность в том, что относительные доли $\frac{m_{ik}}{n}$ будут все одновременно как угодно мало отличаться от вероятностей p_{ik}.

Положим для краткости

$$u_i = \sum_{k=1}^{t} m_{ik},$$

$$v_k = \sum_{i=1}^{s} m_{ik}$$

и введем эмпирические характеристики:

$$Ex = \frac{1}{n}\sum_{i=1}^{s} x_i u_i, \qquad Ey = \frac{1}{n}\sum_{k=1}^{t} y_k v_k,$$

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{s}(x_i - Ex)^2 u_i, \quad \sigma_y^2 = \frac{1}{n}\sum_{k=1}^{t}(y_k - Ey)^2 v_k,$$

$$\tau_{x,y} = \frac{1}{n}\sum_{i=1}^{s}\sum_{k=1}^{t}(x_i - Ex)(y_k - Ey)\, m_{ik}.$$

Тогда *эмпирический коэфициент корреляции* будет

(119) $$R_{x,y} = \frac{\tau_{x,y}}{\sigma_x \sigma_y},$$

или, что то же:

(120) $$R_{x,y} = \frac{\sum\limits_{i=1}^{s}\sum\limits_{k=1}^{t}(x_i - Ex)(y_k - Ey)\, m_{ik}}{\sqrt{\sum\limits_{i=1}^{s}(x_i - Ex)^2 u_i \cdot \sum\limits_{k=1}^{t}(y_k - Ey)^2 v_k}}.$$

При достаточно большом числе испытаний мы будем иметь практическую уверенность в том, что этот эмпирический коэфициент корреляции будет как угодно мало отличаться от теоретического коэфициента корреляции R_{ξ^*, η^*}. Этот же последний даст нам приближенное представление об $R_{\xi, \eta}$.

Для фактического вычисления формула (119) не очень удобна. Статистикой разработан целый ряд приемов для облегчения вычислений по ней, но они только отчасти упрощают дело. Однако в случае, когда есть основание допустить, что налицо линейная корреляция, можно вычислить сначала эмпирические коэфициенты регрессии:

(121) $$\rho_x = \frac{\tau_{x,y}}{\sigma_y^2}, \quad \rho_y = \frac{\tau_{x,y}}{\sigma_x^2},$$

откуда эмпирический коэфициент корреляции определится по формуле

(122) $$R_{x,y} = \pm\sqrt{\rho_x \rho_y}.$$

Разумеется, вычисления по формулам (121) не менее сложны, чем по формуле (119). Но в случае линейной корреляции можно избежать вычислений по формулам (121), опираясь на доказанное выше предложение, что коэфициенты регрессии равны угловым коэфициентам линий регрессии. Как это делается, мы покажем сейчас на одном конкретном примере.

Пусть ξ^* — ширина черепа кролика, η^* — его длина, обе в миллиметрах. Приведем корреляционную таблицу для этих величин ξ^* и η^*, составленную по 44 наблюдениям (табл. VIII).

Таблица VIII.

y \ x	39	40	41	42	43	44
62	1					
65	3	1	1			
68		4	3	5		
71		1	6	4	2	
74			1	7	2	1
77					1	1

Выясним, есть ли основание допустить, что здесь налицо линейная корреляция. Для этого составим прежде всего эмпирические характеристики (для $i = 1, 2, \ldots, s$):

$$E(y/x_i) = \frac{1}{n_i} \sum_{k=1}^{t} y_k m_{ik}. \tag{123}$$

Они дадут нам приближенное представление об $E(\eta^*/x)$. Поэтому если точки с координатами $x = x_i$ и $y = E(y/x_i)$ расположатся примерно на одной прямой, то это даст нам основание допустить, что линия регрессии величины η^* относительно ξ^*, определяемая уравнением

$$y = E(\eta^*/x),$$

есть прямая, и именно та, около которой располагаются названные точки. В нашем примере мы имеем $s = 6$ и

$$x_1 = 39, \quad x_2 = 40, \quad x_3 = 41, \quad x_4 = 42, \quad x_5 = 43, \quad x_6 = 44,$$

а вычисление по формуле (123) дает

$$E(y/x_1) = 64, \quad E(y/x_2) = 68, \quad E(y/x_3) = 70;$$
$$E(y/x_4) = 71, \quad E(y/x_5) = 75, \quad E(y/x_6) = 76.$$

Полученные шесть точек с координатами x_i и $E(y/x_i)$ изображены кружками на черт. 16а. Они расположены примерно на прямой, изображенной на том же чертеже.

Составим еще эмпирические характеристики (для $k = 1, 2, \ldots, t$):

$$E(x/y_k) = \frac{1}{v_k}\sum_{i=1}^{s} x_i m_{ik}. \tag{124}$$

Они дадут нам приближенное представление об $E(\xi^*/y)$. Таким образом, если точки с координатами $y = y_k$ и $x = E(x/y_k)$ также расположатся примерно на одной прямой, то это даст нам основание допустить, что и линия регрессии величины ξ^* относительно η^*, определяемая уравнением

$$x = E(\xi^*/y),$$

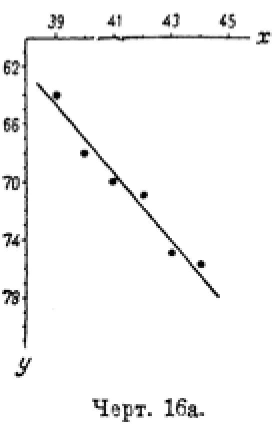
есть прямая. В нашем примере мы имеем $t = 6$ и

$$y_1 = 62,\quad y_2 = 65,\quad y_3 = 68,\quad y_4 = 71,\quad y_5 = 74,\quad y_6 = 77,$$

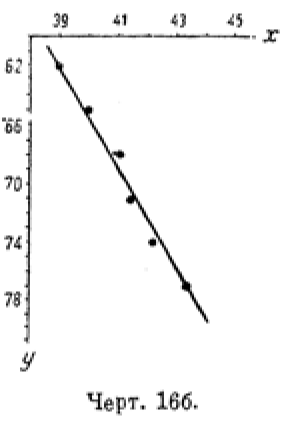
а вычисление по формуле (124) дает

$$E(x/y_1) = 39{,}0;\quad E(x/y_2) = 40{,}0;\quad E(x/y_3) = 41{,}1;$$
$$E(x/y_4) = 41{,}5;\quad E(x/y_5) = 42{,}3;\quad E(x/y_6) = 43{,}5.$$

Полученные шесть точек с координатами y_k и $E(x/y_k)$ изображены кружками на черт. 16б. Они расположены примерно на прямой, изображенной на том же чертеже.



Черт. 16а.



Черт. 16б.

Теперь у нас есть основание допустить, что корреляция в нашем примере линейная. При этом наклон прямой, изображенный на черт. 16а, к оси x, очевидно, дает приближенное значение половины [1]) эмпирического коэфициента регрессии ρ_y, а наклон прямой, изображенной на черт. 16б, к оси y — удвоенного эмпирического коэфициента регрессии ρ_x. Измеряя на чертежах тот и другой, мы получаем

$$2\rho_x = 0{,}6;\qquad \frac{1}{2}\rho_y = 1{,}2.$$

[1]) Так как масштаб по оси x вдвое крупнее масштаба по оси y.

Отсюда по формуле (122) эмпирический коэфициент корреляции

$$R_{x,\,y}=+\sqrt{\rho_x\rho_y}=+\sqrt{0{,}72}=0{,}85.$$

Положительный знак корня мы выбрали потому, что в данном случае $E(y/x_i)$ есть возрастающая функция от x_i и $E(x/y_k)$ — возрастающая функция от y_k.

Подсчет по формуле (120) дал бы в этом случае для эмпирического коэфициента корреляции то же самое значение.

Мы ввели целый ряд эмпирических характеристик: Ex, Ey, σ_x, σ_y, $\tau_{x,\,y}$, $R_{x,\,y}$, $E(x/y_k)$, $E(y/x_i)$, ρ_x, ρ_y. Все это — случайные величины и, как таковые, дают лишь приближенное представление о соответствующих теоретических числовых характеристиках величин ξ^* и η^*. Статистикой изучены законы распределения всех названных эмпирических характеристик, что позволяет оценить, в какой мере и с какой степенью уверенности можно считать их выражением теоретических числовых характеристик. Но эти вопросы уже выходят за рамки настоящего курса.

Точно так же и вопрос о поведении линий регрессии по заданным лежащим около них точкам может решаться не на-глаз, как мы это сделали в нашем примере, а более точно, с оценкой возможной погрешности, если пользоваться так называемым методом наименьших квадратов. Его изложение также не входит в задачу настоящего курса.

126. В п. 102 мы сказали, что если теоретический закон распределения признаков заранее неизвестен, то эмпирический закон даст нам представление о том, как этот теоретический закон может выглядеть. Только что, в п. 117, мы встретились с аналогичным положением: мы не имели заранее никаких оснований предвидеть ту или иную форму линии регрессии в примере черепов кроликов, а подсчет эмпирических характеристик дал нам основание высказать определенное допущение об этой форме.

По этому поводу необходимо сделать следующую оговорку. Неоднократно делались и делаются попытки придавать статистическим закономерностям некое самодовлеющее значение — так, как будто эти закономерности регулируют извне ход различных жизненных процессов, а не являются простым отражением в массовых явлениях законов, специфических для этих жизненных процессов. На этом основании делались и делаются попытки предсказывать течение различных процессов в будущем на основании одной только статистической картины их течения в прошлом, без учета реальных сил, вызывающих и изменяющих в действительности эти процессы. Такая точка зрения не имеет никаких практически оправданных оснований и нередко приносила прямой вред. История статистических исследований в конкретных науках показала, что только там, где исследуемый материал наряду со статистическим изучением подвергался и изучению по существу в данной конкретной науке, получались плодотворные результаты и действительное проникновение в природу вещей.

ДОБАВЛЕНИЕ I.

СОБЫТИЯ КАК ЭЛЕМЕНТЫ СТРУКТУРЫ И СОБЫТИЯ КАК МНОЖЕСТВА.

1. В гл. I мы определили основные соотношения между событиями ($A \subset B$, $A = B$) и основные операции над событиями (AB, $A + B$, $\bar{A}$), причем были установлены свойства операций, которыми эти последние определяются однозначно. Совокупность относящихся сюда предложений, которые мы отметили номерами от 1° до 13°, может рассматриваться как аксиоматическое описание понятия события.

Проанализируем эту аксиоматику. Мы увидим, что формальные свойства событий во многом совпадают с формальными свойствами различных других объектов изучения математических наук.

Пусть S есть множество каких угодно объектов $A, B, C, \ldots$, обладающее тем свойством, что для каждых двух элементов A и B этого множества определено, находятся ли они, или нет, в некотором соотношении $A \subset B$, которое обладает следующими свойствами:

1° $A \subset A$;

2° *Если* $A \subset B$ и $B \subset C$, *то* $A \subset C$.

Такое множество S называется *частично упорядоченным.*

В общем случае соотношение $A \subset B$ словесно читается так: A „предшествует" B, или B „следует" за A. Если одновременно $A \subset B$ и $B \subset A$, то пишут $A = B$.

Примеров частично упорядоченных множеств, с которыми мы имеем дело в математических науках, можно было бы привести сколько угодно. Мы не будем на них останавливаться.

Пусть теперь на частично упорядоченном множестве S для каждых двух его элементов A и B, — безразлично, находящихся в соотношении $A \subset B$ или нет, — определены две операции AB и $A + B$, дающие элементы того же множества S, которые обладают следующими свойствами:

3° $AB \subset A$;

4° $AB \subset B$;

5° *Если* $C \subset A$ и $C \subset B$, *то* $C \subset AB$;

6° $A \subset A + B$;

7° $B \subset A + B$;

8° *Если* $A \subset C$ и $B \subset C$, *то* $A + B \subset C$.

Такое частично упорядоченное множество S называется *структурой.*

Мы видим, что „произведение“ AB есть элемент структуры S, предшествующий одновременно элементам A и B, но следующий за всяким другим элементом, который предшествует A и B. Точно так же „сумма“ $A+B$ есть элемент структуры S, следующий одновременно за элементами A и B, но предшествующий всякому другому элементу, который следует за A и B.

Структуры, являющиеся объектом серьезного изучения, мы встречаем уже в геометрии и в алгебре. Так, пусть S_1 есть множество всех элементов проективной геометрии, т. е. пространство, плоскости, прямые, точки и пустое множество; условимся писать для этих элементов $A \subset B$, когда A лежит на B, и пусть AB есть пересечение A с B, а $A+B$ — наименьший элемент, на котором лежат одновременно A и B. Или, пусть S_2 есть множество всех подгрупп какой-нибудь группы G; условимся писать для этих подгрупп $A \subset B$, когда A есть подгруппа подгруппы B, и пусть AB есть пересечение A с B, а $A+B$ — наименьшая подгруппа группы G, содержащая одновременно A и B. Ясно, что при этих условиях как для S_1, так и для S_2, будут налицо все свойства структуры 1° — 8°.

Пусть еще структура S обладает тем свойством, что для любых трех ее элементов A, B и C имеет место соотношение

9° $A(B+C) = AB + AC$.

Такая структура S называется *дистрибутивной*.

Вместо 9° можно было бы потребовать, чтобы для любых трех элементов A, B и C имело место соотношение

$$9_1^\circ \qquad A + BC = (A+B)(A+C),$$

отличающееся от 9° тем, что здесь на месте произведений стоят суммы и обратно. Для структур оба эти требования совершенно равносильны.

Мы встречаемся с дистрибутивными структурами как с объектами изучения в теории чисел и в теории множеств. Так, пусть S_3 есть множество всех целых положительных чисел; условимся писать $A \subset B$, когда число B делится на число A, и обозначать через AB общий наибольший делитель чисел A и B, а через $A+B$ — их общее наименьшее кратное. Или, пусть S_4 есть какое-нибудь кольцо множеств, т. е. семейство множеств, обладающее тем свойством, что вместе с каждыми двумя множествами их общая часть и их соединение принадлежат семейству. Таковы, например, семейство всех многогранников, включая пустое множество, семейство всех замкнутых множеств на прямой, и т. д., и т. д. Будем, как обычно, писать $A \subset B$, когда множество A содержится в множестве B, и обозначать через AB общую часть множеств A и B и через $A+B$ их соединение. Нетрудно видеть, что при названных обозначениях S_3 и S_4 будут обладать всеми свойствами 1° — 9° дистрибутивных структур, причем, как мы видим, обозначения соотношений и операций в общей теории структур просто заимствованы из теории множеств.

Пусть, наконец, дистрибутивная структура S содержит „первый“ элемент V и „последний“ элемент U, и пусть каждому ее элементу A

соответствует ее же элемент $\bar{A}$, так что удовлетворяются следующие условия:

10° *Каков бы ни был элемент* A, $AU = A$;

11° *Каков бы ни был элемент* A, $A + V = A$;

12° $A + \bar{A} = U$;

13° $A\bar{A} = V$.

Такая дистрибутивная структура S называется *булевской алгеброй*, по имени Буля, который, как думали, впервые применил эту „алгебру" в исследованиях по математической логике (1847). Только недавно стало известно, что эта „алгебра" была открыта, в аналогичных исследованиях, еще раньше братьями Бернулли (1685).

В общем случае элемент $\bar{A}$ называется „дополнением" к элементу A.

Классическим примером булевской алгебры является множество предложений. В математической логике предложением называется всякое высказывание, о котором имеет смысл говорить, что оно истинно или что оно ложно. Условимся писать $A \subset B$, когда предложение B является следствием предложения A, и пусть AB есть предложение, истинность которого равносильна истинности обоих предложений A и B, $A + B$ — предложение, истинность которого равносильна истинности по крайней мере одного из предложений A или B, $\bar{A}$ — отрицание предложения A; пусть затем U есть (какое-нибудь) предложение, которое всегда истинно, V — (какое-нибудь) предложение, которое всегда ложно. Ясно, что здесь будут налицо все свойства 1°—13° булевской алгебры; в частности, 12° и 13° здесь — не что иное, как закон исключенного третьего и закон противоречия.

В качестве другого примера булевской алгебры можно привести любое тело подмножеств какого-нибудь множества M, т. е. семейство подмножеств, содержащее вместе с каждыми двумя множествами A и B их общую часть AB и их соединение $A + B$ и вместе с каждым множеством A его дополнение $\bar{A}$ до множества M, а также содержащее само множество M, а следовательно, и пустое множество N. Если положить $U = M$ и $V = N$, то здесь будут налицо все свойства 1°—13° булевской алгебры.

2. Как и аксиоматика понятия события, аксиоматика вероятности имеет свои аналоги в теории структур.

Булевская алгебра S называется *нормированной*, если она удовлетворяет следующим условиям:

I. Каждому элементу A из S поставлено в соответствие действительное число $|A|$; в общем случае это число называется *нормой* элемента A;

II. Если $A \subset B$, то $|A| \leqslant |B|$;

III. Если $A = B + C$ и $BC = V$, то $|A| = |B| + |C|$;

IV. $|U| = 1$.

Множество предложений является нормированной булевской алгеброй, притом с самой простой нормой, какая только возможна. Здесь каждому предложению A приписывают „логическое значение" $|A|$, равное или 1 или 0, в зависимости от того, истинно это предложение или ложно.

Нормированной же булевской алгеброй является тело всех измеримых подмножеств сегмента [0, 1]. Здесь каждое подмножество A имеет меру $|A|$.

3. Нормированная булевская алгебра предложений послужила образцом для построения первой аксиоматики вероятности. Эта аксиоматика была предложена С. Н. Бернштейном (1917).

Самые вероятности С. Н. Бернштейн предложил рассматривать как вероятности истинности предложений. При этом отпадает необходимость в формулировке специальной аксиоматики для понятия события, и мы можем пользоваться готовой аксиоматикой предложений. Такая постановка вопроса вполне законна, поскольку каждому событию A можно поставить в соответствие предложение „событие A происходит" (в настоящем, в прошедшем или в будущем); говорить о вероятности события A или о вероятности истинности названного предложения, очевидно, одно и то же.

Вероятность истинности предложения обладает теми же формальными свойствами I—IV, что и логическое значение предложения, но, разумеется, может принимать уже не два значения 1 и 0, а весь континуум значений между 0 и 1.

Впрочем, в последнее время в самой математической логике введено понятие такой многозначной нормы предложения, т. е. рассматриваются не только предложения просто истинные и просто ложные, но и предложения „более истинные" и „менее истинные". В связи с этим некоторые современные авторы, трактуя вероятность также как функцию предложения, предлагают рассматривать ее прямо как возможное логическое значение предложения.

4. Нормированная булевская алгебра измеримых подмножеств сегмента [0, 1] послужила образцом для построения той аксиоматики вероятности, которая наиболее известна в настоящее время. Эта аксиоматика была предложена А. Н. Колмогоровым (1929).

Ее возникновению предшествовало обобщение теории меры множества, которое выяснило, что эта теория может быть построена не только для точечных множеств, но и для абстрактных множеств, т. е. для множеств, элементы которых могут быть какой угодно природы. А. Н. Колмогоров предложил рассматривать вероятность как одну из возможных мер. При этом отпадает необходимость в формулировке специальной аксиоматики не только для понятия события, но, строго говоря, и для самого понятия вероятности: мы можем пользоваться готовым понятием множества и готовой аксиоматикой меры. Такая постановка вопроса законна постольку, поскольку все события, бывшие когда-либо предметом исследования в теории вероятностей, подразделяются на частные случаи, которые фигурируют в исследовании уже как неделимые; следовательно, события могут рассматриваться как множества некоторых элементов. Так, во всех случаях, где применимо классическое определение вероятности, в основу которого кладется полная группа несовместимых между собой событий

$E_1, E_2, \ldots, E_n$, допустимые события могут рассматриваться как конечные множества, составленные всевозможными способами из элементов $E_1, E_2, \ldots, E_n$. Во всех случаях, где речь идет о случайных величинах, подлежащими изучению событиями являются попадания значений случайной величины ξ на те или иные множества точек числовой оси, и эти события, естественно, могут рассматриваться как множества „элементарных событий", каковыми являются попадания значений ξ на отдельные точки числовой оси, и т. п.

Точно аксиоматика А. Н. Колмогорова может быть формулирована так: события всегда представляют собой какое-нибудь тело подмножеств некоторого множества M „элементарных событий", и вероятность есть функция множества, обладающего общими свойствами меры, т. е. свойствами I—III, причем само множество M имеет вероятность 1. В силу всего сказанного выше в аксиоматике А. Н. Колмогорова события и их вероятности обладают всеми свойствами 1°—13° и I—IV, где U есть множество M и V — пустое множество.

Аксиоматика А. Н. Колмогорова есть частный случай аксиоматики, приведенной в нашей книге, так как в аксиоматике А. Н. Колмогорова требуется, чтобы события были множествами, чего не предполагается у нас. Но для каждого поля вероятностей S, удовлетворяющего нашей аксиоматике, найдется изоморфное ему поле вероятностей, удовлетворяющее аксиоматике А. Н. Колмогорова. В самом деле, рассмотрим всевозможные нормы $|A|$ событий A поля S, удовлетворяющие условиям I—IV, но имеющие для каждого события A поля S только одно из двух значений 1 или 0, и обозначим через $\mathfrak{A}(A)$ множество всех тех норм $|A|$, которые имеют для A значение 1. Нетрудно видеть, что если взять множества $\mathfrak{A}(A)$, соответствующие всем событиям A поля S, и положить $P[\mathfrak{A}(A)] = P(A)$, то мы получим поле вероятностей, удовлетворяющее аксиоматике А. Н. Колмогорова.

ДОБАВЛЕНИЕ II.

ФОРМУЛА СТИРЛИНГА.

1. Речь идет о формуле

$$\text{(S)} \qquad n! = \sqrt{2\pi}\, n^n \sqrt{n}\, e^{-n}(1 + \alpha_n) \quad (\alpha_n \to 0 \text{ при } n \to \infty).$$

Ее можно вывести, следуя Э. Гурса, таким образом [1]).

1° Прежде всего нетрудно убедиться в том, что верна следующая формула Эйлера:

$$\text{(E)} \qquad n! = \int_0^{+\infty} x^n e^{-x}\, dx.$$

[1]) Э. Гурса, *Курс математического анализа*, т. I, ч. I, стр. 268—270 русск. изд. 1933 года.

Действительно, при $n=0$

$$\int_0^{+\infty} e^{-x}\,dx = -e^{-x}\Big|_0^{+\infty} = 1 = 0!$$

Если же формула (E) верна для какого-нибудь n, то для $n+1$

$$\int_0^{+\infty} x^{n+1}e^{-x}\,dx = -x^{n+1}e^{-x}\Big|_0^{+\infty} + (n+1)\int_0^{+\infty} x^n e^{-x}\,dx = (n+1)n! = (n+1)!$$

2° Покажем теперь, что в формуле (E) можно произвести замену переменного x на t по формуле

$$x^n e^{-x} = n^n e^{-n} e^{-t^2}, \tag{F}$$

причем полупрямой $0 < x < +\infty$ соответствует вся прямая $-\infty < < t < +\infty$.

Действительно, функция $x^n e^{-x}$ имеет максимум при $x = n$, и ее наибольшее значение равно $n^n e^{-n}$. При возрастании x от 0 до n эта функция возрастает от 0 до $n^n e^{-n}$, а при дальнейшем возрастании x от n до $+\infty$ она убывает до 0. Точно так же функция $n^n e^{-n} e^{-t^2}$ имеет максимум при $t=0$, и ее наибольшее значение равно $n^n e^{-n}$. При возрастании t от $-\infty$ до 0 эта функция также возрастает от 0 до $n^n e^{-n}$, и при дальнейшем возрастании t от 0 до $+\infty$ она убывает до 0.

3° Чтобы произвести замену переменного под знаком интеграла (E), надо еще выразить dx через dt.

С этой целью возьмем логарифмические диференциалы от обеих частей формулы (F); мы получим

$$dx = 2t\,\frac{x}{x-n}\,dt.$$

С другой стороны, из той же формулы (F) мы имеем

$$t^2 = x - n - n\log\left(\frac{x}{n}\right).$$

Разлагая логарифм по формуле Тейлора, мы получим

$$\log\left(\frac{x}{n}\right) = \log\left(1 + \frac{x-n}{n}\right) = \frac{x-n}{n} - \frac{1}{2\left(1+\theta\,\frac{x-n}{n}\right)^2}\left(\frac{x-n}{n}\right)^2,$$

причем θ заключается между нулем и единицей. Следовательно, мы имеем

$$t^2 = \frac{n}{2\left(1+\theta\,\frac{x-n}{n}\right)^2}\left(\frac{x-n}{n}\right)^2 = \frac{n(x-n)^2}{2\,[n+\theta(x-n)]^2}.$$

Отсюда мы найдем последовательно

$$\frac{1}{t} = \sqrt{\frac{2}{n}}\,\frac{n+\theta(x-n)}{x-n} = \sqrt{\frac{2}{n}}\,\frac{n}{x-n} + \sqrt{\frac{2}{n}}\,\theta,$$

$$\frac{n}{x-n} = \frac{1}{t}\sqrt{\frac{n}{2}} - \theta,$$

$$2t\,\frac{x}{x-n} = 2t\left(\frac{n}{x-n}+1\right) = 2\left[\sqrt{\frac{n}{2}} + (1-\theta)\,t\right].$$

Отсюда окончательно

(G) $$dx = 2\left[\sqrt{\frac{n}{2}} + (1-\theta)\,t\right]dt.$$

4° Остается сделать замену переменного под знаком интеграла (E) пользуясь формулами (F) и (G).

Мы получим

$$n! = 2n^n e^{-n}\sqrt{\frac{n}{2}}\int_{-\infty}^{+\infty} e^{-t^2}dt + 2n^n e^{-n}\int_{-\infty}^{+\infty} e^{-t^2}(1-\theta)\,t\,dt.$$

Первый интеграл, по формуле Пуассона, равен $\sqrt{\pi}$. Что касается второго интеграла, то его можно оценить следующим образом:

$$\left|\int_{-\infty}^{+\infty} e^{-t^2}(1-\theta)\,t\,dt\right| \leqslant 2\int_{0}^{+\infty} e^{-t^2}t\,dt,$$

интеграл же в правой части этого последнего неравенства равен $\frac{1}{2}$. Отсюда окончательно

$$n! = \sqrt{2\pi}\,n^n\sqrt{n}\,e^{-n}\left(1+\frac{\omega}{\sqrt{2\pi n}}\right),$$

причем ω заключается между -1 и 1; это и доказывает формулу (S).

2. Помимо того теоретического применения формулы Стирлинга, которое мы имеем в гл. III, ее применяют также в приближенных вычислениях.

Формула (S) дает приближенное равенство

$$n! \approx \sqrt{2\pi}\,n^n\sqrt{n}\,e^{-n},$$

или

$$\lg n! \approx \left(n+\frac{1}{2}\right)\lg n - n\lg e + \lg\sqrt{2\pi},$$

причем

$$\lg e = 0{,}43429\ldots,$$

$$\lg\sqrt{2\pi} = 0{,}39909\ldots$$

В применении, например, к вычислению числа сочетаний это дает

$$C_n^m \approx \frac{1}{\sqrt{2\pi}} \frac{n^n \sqrt{n}}{m^m \sqrt{m}\,(n-m)^{n-m} \sqrt{n-m}},$$

или

$$\lg C_n^m \approx \left(n+\frac{1}{2}\right)\lg n - \left(m+\frac{1}{2}\right)\lg m -$$
$$-\left(n-m+\frac{1}{2}\right)\lg(n-m) - \lg\sqrt{2\pi}.$$

В применении к закону Пуассона $P(m)=\frac{k^m}{m!}e^{-k}$ мы имеем

$$P(m) \approx \frac{k^m}{\sqrt{2\pi}\, m^m \sqrt{m}} e^{m-k},$$

или

$$\lg P(m) \approx m \lg k - \left(m+\frac{1}{2}\right)\lg m + (m-k)\lg e - \lg\sqrt{2\pi}$$

Таблица значений функции

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz.$$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,0000						
0,01	0,0040	0,41	0,1591	0,81	0,2910	1,21	0,3869
0,02	0,0080	0,42	0,1628	0,82	0,2939	1,22	0,3888
0,03	0,0120	0,43	0,1664	0,83	0,2967	1,23	0,3907
0,04	0,0160	0,44	0,1700	0,84	0,2995	1,24	0,3925
0,05	0,0199	0,45	0,1736	0,85	0,3023	1,25	0,3944
0,06	0,0239	0,46	0,1772	0,86	0,3051	1,26	0,3962
0,07	0,0279	0,47	0,1808	0,87	0,3078	1,27	0,3980
0,08	0,0319	0,48	0,1844	0,88	0,3106	1,28	0,3997
0,09	0,0359	0,49	0,1879	0,89	0,3133	1,29	0,4015
0,10	0,0398	0,50	0,1915	0,90	0,3159	1,30	0,4032
0,11	0,0438	0,51	0,1950	0,91	0,3186	1,31	0,4049
0,12	0,0478	0,52	0,1985	0,92	0,3212	1,32	0,4066
0,13	0,0517	0,53	0,2019	0,93	0,3238	1,33	0,4082
0,14	0,0557	0,54	0,2054	0,94	0,3264	1,34	0,4099
0,15	0,0596	0,55	0,2088	0,95	0,3289	1,35	0,4115
0,16	0,0636	0,56	0,2123	0,96	0,3315	1,36	0,4131
0,17	0,0675	0,57	0,2157	0,97	0,3340	1,37	0,4147
0,18	0,0714	0,58	0,2190	0,98	0,3365	1,38	0,4162
0,19	0,0753	0,59	0,2224	0,99	0,3389	1,39	0,4177
0,20	0,0793	0,60	0,2257	1,00	0,3413	1,40	0,4192
0,21	0,0832	0,61	0,2291	1,01	0,3438	1,41	0,4207
0,22	0,0871	0,62	0,2324	1,02	0,3461	1,42	0,4222
0,23	0,0910	0,63	0,2357	1,03	0,3485	1,43	0,4236
0,24	0,0948	0,64	0,2389	1,04	0,3508	1,44	0,4251
0,25	0,0987	0,65	0,2422	1,05	0,3531	1,45	0,4265
0,26	0,1026	0,66	0,2454	1,06	0,3554	1,46	0,4279
0,27	0,1064	0,67	0,2486	1,07	0,3577	1,47	0,4292
0,28	0,1103	0,68	0,2517	1,08	0,3599	1,48	0,4306
0,29	0,1141	0,69	0,2549	1,09	0,3621	1,49	0,4319
0,30	0,1179	0,70	0,2580	1,10	0,3643	1,50	0,4332
0,31	0,1217	0,71	0,2611	1,11	0,3665	1,51	0,4345
0,32	0,1255	0,72	0,2642	1,12	0,3686	1,52	0,4357
0,33	0,1293	0,73	0,2673	1,13	0,3708	1,53	0,4370
0,34	0,1331	0,74	0,2703	1,14	0,3729	1,54	0,4382
0,35	0,1368	0,75	0,2734	1,15	0,3749	1,55	0,4394
0,36	0,1406	0,76	0,2764	1,16	0,3770	1,56	0,4406
0,37	0,1443	0,77	0,2794	1,17	0,3790	1,57	0,4418
0,38	0,1480	0,78	0,2823	1,18	0,3810	1,58	0,4429
0,39	0,1517	0,79	0,2852	1,19	0,3830	1,59	0,4441
0,40	0,1554	0,80	0,2881	1,20	0,3849	1,60	0,4452

Продолжение

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
1,61	0,4463	1,86	0,4686	2,22	0,4868	2,72	0,4967
1,62	0,4474	1,87	0,4693	2,24	0,4875	2,74	0,4969
1,63	0,4484	1,88	0,4699	2,26	0,4881	2,76	0,4971
1,64	0,4495	1,89	0,4706	2,28	0,4887	2,78	0,4973
1,65	0,4405	1,90	0,4713	2,30	0,4893	2,80	0,4974
1,66	0,4515	1,91	0,4719	2,32	0,4898	2,82	0,4976
1,67	0,4525	1,92	0,4726	2,34	0,4904	2,84	0,4977
1,68	0,4535	1,93	0,4732	2,36	0,4909	2,86	0,4979
1,69	0,4545	1,94	0,4738	2,38	0,4913	2,88	0,4980
1,70	0,4554	1,95	0,4744	2,40	0,4918	2,90	0,4981
1,71	0,4564	1,96	0,4750	2,42	0,4922	2,92	0,4982
1,72	0,4573	1,97	0,4756	2,44	0,4927	2,94	0,4984
1,73	0,4582	1,98	0,4761	2,46	0,4931	2,96	0,4985
1,74	0,4591	1,99	0,4767	2,48	0,4934	2,98	0,4986
1,75	0,4599	2,00	0,4772	2,50	0,4938	3,00	0,49865
1,76	0,4608	2,02	0,4783	2,52	0,4941	3,20	0,49931
1,77	0,4616	2,04	0,4793	2,54	0,4945	3,40	0,49966
1,78	0,4625	2,06	0,4803	2,56	0,4948	3,60	0,499841
1,79	0,4633	2,08	0,4812	2,58	0,4951	3,80	0,499928
1,80	0,4641	2,10	0,4821	2,60	0,4953	4,00	0,499968
1,81	0,4649	2,12	0,4830	2,62	0,4956	4,50	0,499997
1,82	0,4656	2,14	0,4838	2,64	0,4959	5,00	0,49999997
1,83	0,4664	2,16	0,4846	2,66	0,4961		
1,84	0,4671	2,18	0,4854	2,68	0,4963		
1,85	0,4678	2,20	0,4861	2,70	0,4965		

УКАЗАТЕЛЬ ЛИТЕРАТУРЫ.

Общие руководства.

Марков А. А., *Исчисление вероятностей*, 4-е изд., ГИЗ, 1924.
Бернштейн С. Н., *Теория вероятностей*, 3-е изд., ГТТИ, 1935.
von Mises R., *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*, Wien, 1931.
Fréchet M., *Recherches théoriques modernes sur la Théorie des probabilités* v. I, II, Paris, 1937.

Монографии.

Колмогоров А. Н., *Основные понятия теории вероятностей*, ОНТИ, 1936.
Хинчин А. Я., *Основные законы теории вероятностей*, 2-е изд., ГТТИ, 1932.

К гл. VI.

Хинчин А. Я., *Асимптотические законы теории вероятностей*, ОНТИ, 1936.

К гл. VII—VIII.

Хинчин А. Я., *Предельные законы для сумм независимых случайных величин*, ОНТИ, 1938.
Lévy P., *Théorie de l'addition des variables aléatoires*, Paris, 1937.
Cramér H., *Random variables and probability distributions*, Cambridge, 1937.

К гл. IX.

Романовский В. И., *Математическая статистика*, ОНТИ, 1938.
Jordan Ch., *Statistique mathématique*, Paris, 1927.

К доп. I.

Glivenko V., *Théorie générale des structures*, Paris, 1938.

Статьи.

К гл. VI.

Колмогоров А. Н., *Об аналитических методах в теории вероятностей*, Успехи матем. наук, в. V (1938).
Хинчин А. Я., *Теория корреляции стационарных стохастических процессов*, там же.
Феллер В., *К теории стохастических процессов* (теоремы существования и единственности), там же.

К гл. VII.

Kolmogoroff A., *Sulla forma generale di un processo stocastico omogeneo*, Rend. Lincei (6), 15 (1932), стр. 805—808 и 866—869.

К гл. VIII.

Lindeberg J. W., *Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung*, Math. Zeitschr. 15 (1922), стр. 211—215.
Feller W., *Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung*, Math. Zeitschr. 40 (1935), стр. 521—559.
Bawly G. M., *Über eine Verallgemeinerung der Grenzwertsätze der Wahrscheinlichkeitsrechnung*, Мат. Сб., новая серия 1 (1936), стр. 917—929.
Гнеденко Б. В., *Предельные теоремы для сумм независимых случайных величин*, Изв. АН СССР, серия математическая, 1939, стр. 181—232.
Kolmogoroff A., *Über die Summen durch den Zufall bestimmter unabhängiger Grössen*, Math. Ann. 99 (1928), стр. 309—319 и 102 (1929), стр. 484—488.
Райков Д. А., *О связи между центральным предельным законом теории вероятностей и законом больших чисел*, Изв. АН СССР, серия математическая, 1938, стр. 323—336.

К гл. IX.

Glivenko V., *Sur la loi des grands nombres dans l'espace fonctionnel*, Confér intern. des sciences mathém., v. V, Paris, 1938.
Смирнов Н. В., *О распределении ω^2 — критерия Мизеса*, Мат. Сб., новая серия, 2 (1937), стр. 973—993.
Glivenko V., *Sulla determinazione empirica delle leggi di probabilita*. Giorn. Attuari, IV (1933), стр. 92—99.
Kolmogoroff A., *Sulla determinazione empirica di una legge di distribuzione* Guorn. Attuari, IV (1933), стр. 83—91.

К доп. I.

Бернштейн С. Н., *Опыт аксиоматического обоснования теории вероятностей*, Сообщ. Харьк. Мат. Общ., 15 (1917), стр. 209—274.
Колмогоров А. Н., *Общая теория меры и исчисление вероятностей*. Сборн. матем. разд. Комм. Ак., 1 (1929), стр. 8—21.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ.

Редактор *Б. В. Гнеденко* Техн. редактор *В. Ф. Зазульская.*

Индекс Т 26-5-2
Сдано в производство 7/VI 1939 г.
Подписано к печати 29/VII 1939 г.
Тираж 15 000 экз.
Печ. л. $13^3/_4$
Уч.-авт. л. 16,3
Тип. зн. в 1 бум. листе 102400

Прот. ТКК № 29
Формат $60\times92^1/_{16}$
Изд. № 213
Учетный № 4887
Уполном. Главлита № А-14988
Заказ № 1725
Бумага Вишерского бумкомбината

2-я тип. ОНТИ им. Евг. Соколовой. Ленинград, пр. Красных Командиров 29