МАТЕМАТИКА в МОНОГРАФИЯХ

А. Н. КОЛМОГОРОВ

ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ

ОНТИ • НКТП • СССР • 1936

Н И

МАТЕМАТИКА В МОНОГРАФИЯХ

ПОД РЕДАКЦИЕЙ

акад. И. М. ВИНОГРАДОВА, проф. А. Н. КОЛМОГОРОВА, акад. Н. Н. ЛУЗИНА, проф. Л. А. ЛЮСТЕРНИКА, проф. А. И. ПЛЕСНЕРА

СЕРИЯ ОБЗОРОВ

КНИГА II

А. Н. КОЛМОГОРОВ

ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ

ОБЪЕДИНЕННОЕ

НАУЧНО-ТЕХНИЧЕСКОЕ ИЗДАТЕЛЬСТВО НКТП СССР

А. Н. КОЛМОГОРОВ

ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ

ПЕРЕВОД С НЕМЕЦКОГО
Г. М. БАВЛИ

ГЛАВНАЯ РЕДАКЦИЯ ОБЩЕТЕХНИЧЕСКОЙ ЛИТЕРАТУРЫ И НОМОГРАФИИ
МОСКВА 1936 ЛЕНИНГРАД

Т-26-5-4(2)
ТКК 111

Редакция *Д. А. Райкова.* Оформление *В. Ф. Зазульской.*
Корректура *А. С. Стихно.* Наблюдал за выпуском *Тимофеев.*
Сдано в производство 22/VIII 1935 г. Подписано в печать 16/III 1936 г.
Кол. бум. л. 1¼. Тираж 3000. Формат 82 × 110. Кол. печ. зн. в 1 б. л. 98000.
Заказ № 2748. Уч. авт. л. 5. Общетехн. дисц. № 52. Уполн. Главл. № В-32224.

4-я тип. ОНТИ НКТП СССР «Кр. Печатник». Ленинград, Международный, 75а.

ПРЕДИСЛОВИЕ

Целью предлагаемой работы является аксиоматическое обоснование теории вероятностей. Ведущей мыслью автора было при этом естественное включение основ теории вероятностей, считавшихся еще недавно совершенно своеобразными, в ряд общих понятий современной математики. До возникновения лебеговой теории меры и интеграла эта задача была почти безнадежна. После исследований Лебега стала ясной аналогия между мерой множества и вероятностью события, а также между интегралом от функции и математическим ожиданием случайной величины. Эта аналогия допускает и дальнейшее продолжение: так, например, многие свойства независимых случайных величин вполне аналогичны соответствующим свойствам ортогональных функций. Для того чтобы, исходя из этой аналогии, обосновать теорию вероятностей, следовало еще освободить теорию меры и теорию интегрирования от геометрических элементов, которые еще имелись у Лебега. Это освобождение было осуществлено Фреше.

Попытки построения основ теории вероятностей, исходящие из этой общей точки зрения, уже имеются, и весь круг идей, излагаемых здесь, уже успел приобрести известную популярность в узком кругу специалистов; однако отсутствовало полное и свободное от излишних усложнений изложение всей системы (подготовляется, впрочем, к печати книга Фреше, см. указатель литературы, Fréchet [2]).

Я хотел бы еще указать здесь на те места в дальнейшем изложении, которые выходят за пределы упомянутого выше круга идей, уже достаточно знакомого в общих чертах специалистам. Эти места следующие: распределения вероятностей в бесконечномерных пространствах (глава третья, § 4), диференцирование и интегрирование математических ожиданий по параметру (глава четвертая, § 5) и особенно теория условных вероятностей и математических ожиданий (глава пятая). Следует при этом отметить, что все эти новые понятия и проблемы с необходимостью возникают при рассмотрении вполне конкретных физических задач [1]).

Шестая глава содержит обзор отдельных результатов А. Я. Хинчина и автора, касающихся условий применимости простого и усиленного закона больших чисел. В указателе литературы приведены некоторые новые работы, представляющие интерес с точки зрения вопросов обоснования теории вероятностей.

Приношу свою сердечную благодарность А. Я. Хинчину, внимательно прочитавшему всю рукопись и предложившему целый ряд улучшений.

Клязьма близ Москвы, 1 мая 1933 г.

А. Колмогоров

---

[1]) Ср., например, цитированную в сноске [1]) к стр. 55 работу М. А. Леонтовича и автора, а также *M. Leontowitsch*, Zur Statistik der kontinuierlichen Systeme und des zeitlichen Verlaufes der physikalischen Vorgänge, „Physik. Zeitschr. d. Sowjetunion“, т. 3, 1933, стр. 35—63.

ОГЛАВЛЕНИЕ

Стр.

I. Элементарная теория вероятностей

§ 1. Аксиомы 10
§ 2. Отношение к данным опыта 11
§ 3. Терминологические замечания 13
§ 4. Непосредственные следствия из аксиом, условные вероятности, теорема Байеса 14
§ 5. Независимость 16
§ 6. Условные вероятности как случайные величины, цепи Маркова 20

II. Бесконечные поля вероятностей

§ 1. Аксиома непрерывности 22
§ 2. Борелевские поля вероятностей 24
§ 3. Примеры бесконечных полей вероятностей 26

III. Случайные величины

§ 1. Вероятностные функции 30
§ 2. Определение случайных величин, функции распределения . 31
§ 3. Многомерные функции распределения 33
§ 4. Вероятности в бесконечномерных пространствах 36
§ 5. Эквивалентные случайные величины, разные виды сходимости 43

IV. Математические ожидания

§ 1. Абстрактные интегралы Лебега 46
§ 2. Абсолютные и условные математические ожидания 48
§ 3. Неравенство Чебышева 50
§ 4. Некоторые признаки сходимости 52
§ 5. Диференцирование и интегрирование математических ожиданий по параметру 53

V. Условные вероятности и математические ожидания

§ 1. Условные вероятности 56
§ 2. Объяснение одного парадокса Бореля 60
§ 3. Условные вероятности относительно случайной величины . 60
§ 4. Условные математические ожидания 62

Стр.

VI. Независимость. Закон больших чисел

§ 1. Независимость 66
§ 2. Независимые случайные величины 67
§ 3. Закон больших чисел 70
§ 4. Замечания к понятию математического ожидания 73
§ 5. Усиленный закон больших чисел, сходимость рядов 75

Дополнение. Одна замечательная теорема теории вероятностей 78
Указатель литературы 80

I. ЭЛЕМЕНТАРНАЯ ТЕОРИЯ ВЕРОЯТНОСТЕЙ

Мы называем элементарной теорией вероятностей ту часть теории вероятностей, в которой приходится иметь дело с вероятностями лишь конечного числа событий. Теоремы, которые здесь выводятся, естественно применяются также и к вопросам, связанным с бесконечным числом случайных событий, однако при изучении этих последних применяются также существенно новые принципы. Поэтому единственная аксиома математической теории вероятностей, относящаяся именно к случаю бесконечного числа случайных событий, вводится лишь в начале второй главы (аксиома VI).

Теория вероятностей как математическая дисциплина может и должна быть аксиоматизирована совершенно в том же смысле, как геометрия или алгебра. Это означает, что, после того как даны названия изучаемым объектам и их основным отношениям, а также аксиомы, которым эти отношения должны подчиняться, все дальнейшее изложение должно основываться исключительно лишь на этих аксиомах, не опираясь на обычное конкретное значение этих объектов и их отношений.

Соответственно этому в § 1 определяется понятие *поля вероятностей* как системы множеств, удовлетворяющей известным условиям. Что представляют собой элементы этих множеств, совершенно безразлично для чисто математического развития теории вероятностей (ср. введение основных геометрических понятий в „Основах геометрии" Гильберта, или определение групп, колец и тел в абстрактной алгебре).

Всякая аксиоматическая (абстрактная) теория допускает, как известно, бесконечное число конкретных интерпретаций. Таким образом и математическая теория вероятностей допускает наряду с теми интерпретациями, из которых она возникла, также много других. Так мы приходим к приложениям математической теории вероятностей к таким областям науки, которые не имеют отношения к понятиям случая и вероятности в собственном смысле этого слова.

Аксиоматизация основ теории вероятностей может быть проведена различными способами как в отношении выбора аксиом, так и выбора основных понятий и основных соотношений. Если преследовать цель возможной простоты как самой системы аксиом, так и построения из нее дальнейшей теории, то представляется наиболее целесообразным аксиоматизирование понятий случайного события и его вероятности. Существуют также другие системы обоснования теории вероятностей, а именно такие, в которых понятие вероятности не относится к числу основных понятий, а само выражается через другие понятия [1]). При этом стремятся, однако, к другой цели, а именно по возможности к наиболее тесному смыканию математической теории с эмпирическим возникновением понятия вероятности.

§ 1. Аксиомы [2])

Пусть E — множество элементов ξ, η, $\zeta, \ldots$, которые мы будем называть *элементарными событиями*, а $\mathfrak{F}$ — множество подмножеств из E; элементы множества $\mathfrak{F}$ будем называть *случайными событиями.*

I. *$\mathfrak{F}$ является телом множеств* [3]).

II. *$\mathfrak{F}$ содержит множество E.*

III. *Каждому множеству A из $\mathfrak{F}$ поставлено в соответствие неотрицательное действительное число $\mathsf{P}(A)$. Это число называется вероятностью события A.*

IV. $\mathsf{P}(E) = 1$.

V. *Если A и B не пересекаются, то*

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

Система множеств $\mathfrak{F}$ вместе с поставленными к ней в определенное соответствие числами $\mathsf{P}(A)$, удовлетворяющая аксиомам I—V, называется *полем вероятностей.*

Наша система аксиом I—V *непротиворечива.* Это показывает следующий пример: E состоит из единственного элемента ξ, $\mathfrak{F}$ — из E и нулевого множества 0, при этом положено $\mathsf{P}(E) = 1$, $\mathsf{P}(0) = 0$.

Наша система аксиом, однако, не является *полной*: в разных вопросах теории вероятностей рассматриваются различные поля вероятностей.

---

1) Ср., например, *R. v. Mises* [1] и [2] и *С. Н. Бернштейн* [1].

2) Читатель, желающий сразу же придать конкретный смысл нижеследующим аксиомам, должен сначала читать § 2.

3) *Ср. Hausdorff*, Mengenlehre, 1927, стр. 78. Система множеств называется *телом*, если сумма, пересечение и разность двух множеств системы опять принадлежат этой системе. Каждое не пустое тело множеств содержит нулевое множество 0. Мы обозначаем, следуя Гаусдорфу, пересечение A и B через AB, сумму множеств A и B в случае $AB = 0$ через $A + B$, в общем случае через $A \dotplus B$, и разность A и B через $A - B$. Дополнительное множество $E - A$ к множеству A обозначается через $\bar{A}$. Элементарные правила действий над множествами и их пересечениями, суммами и разностями мы в дальнейшем будем предполагать известными. Множества из $\mathfrak{F}$ будем в дальнейшем обозначать большими латинскими буквами.

Построение поля вероятностей. Простейшие поля вероятностей строятся следующим образом. Берутся произвольное конечное множество $E = \{\xi_1, \xi_2, \ldots, \xi_k\}$ и произвольное множество $\{P_1, P_2, \ldots, P_k\}$ неотрицательных чисел с суммою $P_1 + P_2 + \ldots + P_k = 1$. За $\mathfrak{F}$ принимается совокупность всех подмножеств из E, и полагается $\mathsf{P}\{\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_\lambda}\} = P_{i_1} + P_{i_2} + \ldots + P_{i_\lambda}$. В этом случае говорят, что $P_1, P_2, \ldots, P_k$ суть вероятности элементарных событий $\xi_1, \xi_2, \ldots, \xi_k$, или просто *элементарные вероятности*. Так получаются все возможные *конечные* поля вероятностей, в которых $\mathfrak{F}$ состоит из совокупности всех подмножеств из E (при этом поле вероятностей называют конечным, если множество E конечно). По поводу дальнейших примеров ср. вторую главу, § 3.

§ 2. Отношение к данным опыта [1])

Применение теории вероятностей к действительному миру опытов происходит по следующей схеме:

1. Предполагается данным некоторый комплекс $\mathfrak{S}$ условий, допускающий неограниченное число повторений.

2. Изучают определенный круг событий, которые могут наступить в результате осуществления условий $\mathfrak{S}$. В отдельных случаях реализации условий $\mathfrak{S}$ упомянутые события протекают, вообще говоря, различным образом. Пусть E — множество различных возможных вариантов $\xi_1, \xi_2, \ldots$ исхода указанных событий. Некоторые из этих вариантов могут при этом вообще не реализоваться. Мы включаем во множество E все варианты, которые мы a priori рассматриваем как возможные.

3. Если после реализации условий $\mathfrak{S}$ осуществившийся на практике вариант нашего события окажется принадлежащим к (определенному какими-либо условиями) множеству A, то мы говорим, что наступило событие A.

Пример. Комплекс условий $\mathfrak{S}$ заключается в том, что бросают два раза монету. Круг событий, о котором шла речь в п. 2, состоит в том, что при каждом бросании могут появиться решетка или герб. Отсюда

---

[1]) Читатель, который интересуется лишь чисто математическим развитием теории, может этот параграф не читать — дальнейшее изложение основывается на аксиомах § 1 и не использует рассуждений этого параграфа. В нем мы ограничимся лишь простым указанием на эмпирическое возникновение аксиом теории вероятностей и сознательно оставляем поэтому в стороне глубокие философские изыскания о понятии вероятности в мире опытов. В изложении необходимых предпосылок для приложимости теории вероятностей к миру действительных событий автор в значительной мере следует выводам Мизеса, в частности ср. R. v. Mises [1], стр. 21—27, параграф „Das Verhältnis der Theorie zur Erfahrungswelt“.

следует, что всего возможно четыре различных варианта (элементарных события), именно: решетка — решетка, решетка — герб, герб — решетка, герб — герб. В качестве события A рассматривается „повторение". Это событие состоит из соединения первого и четвертого элементарных событий. Таким образом можно каждое событие рассматривать как множество элементарных событий.

4. При известных условиях, в которые мы здесь не будем глубже вдаваться, можно предположить, что некоторым событиям A, которые могут наступить или же не наступить после осуществления условий $\mathfrak{S}$, поставлены в соответствие определенные действительные числа $\mathsf{P}(A)$, которые обладают следующими свойствами:

A. Можно практически быть уверенным, что, если комплекс условий $\mathfrak{S}$ будет повторен большое число n раз и если при этом через m обозначено число случаев, при которых событие A наступило, то отношение $\frac{m}{n}$ будет очень мало отличаться от $\mathsf{P}(A)$.

B. Если $\mathsf{P}(A)$ очень мало, то можно практически быть уверенным, что при однократной реализации условий $\mathfrak{S}$ событие A не будет иметь места.

Эмпирическая дедукция аксиом. Обычно можно предполагать, что система $\mathfrak{F}$ рассматриваемых событий $A, B, C, \ldots$, которым приписаны определенные вероятности, образует тело множеств, содержащее в качестве элемента множество E (аксиомы I и II, а также первая часть аксиомы III — существование вероятностей). Далее ясно, что $0 \leqslant \frac{m}{n} \leqslant 1$, так что вторая часть аксиомы III оказывается вполне естественной. Для события E всегда $m = n$, благодаря чему естественно положить $\mathsf{P}(E) = 1$ (аксиома IV). Если, наконец, A и B несовместны между собой (т. е. множества A и B не пересекаются), то $m = m_1 + m_2$, где m, m_1, m_2 обозначают соответственно число опытов, в которых происходят события $A + B$, A и B. Отсюда следует:

$$\frac{m}{n} = \frac{m_1}{n} + \frac{m_2}{n}.$$

Следовательно, является уместным положить $\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B)$ (аксиома V).

Примечание I. Из практической достоверности двух утверждений следует практическая достоверность утверждения об их одновременной правильности, хотя степень достоверности при этом несколько понижается. Если, однако, число утверждений очень велико, то из практической достоверности каждого отдельного из этих утверждений вообще нельзя вывести никаких заключений относительно одновременной правильности всех этих утверждений. Поэтому из принципа A никоим образом не следует, что при очень большом числе серий по n испытаний в *каждой* серии отношение $\frac{m}{n}$ будет мало отличаться от $\mathsf{P}(A)$.

Примечание II. Невозможному событию (пустому множеству) соответствует в силу наших аксиом вероятность $\mathsf{P}(0) = 0$ [1]), в то время как, наоборот, из $\mathsf{P}(A) = 0$ не следует еще невозможность события A; согласно принципу В из обращения вероятности в нуль следует только, что при однократной реализации условий $\mathfrak{S}$ событие A практически невозможно. Это, однако, никоим образом не означает, что при достаточно длинном ряде испытаний событие A также не наступит. С другой стороны, можно согласно принципу А лишь утверждать, что при $\mathsf{P}(A) = 0$ и очень большом n отношение $\frac{m}{n}$ будет очень мало $\left(\text{оно может, например, равняться } \frac{1}{n}\right)$.

§ 3. Терминологические замечания

Мы определили объекты нашего дальнейшего изучения — случайные события — как множества. Многие теоретико-множественные понятия обозначаются, однако, в теории вероятностей другими именами. Мы приведем здесь краткий указатель таких понятий.

В теории множеств	*Для случайных событий*
1. A и B не пересекаются, то-есть $AB = 0$.	1. События A и B несовместны.
2. $AB \ldots N = 0$.	2. События $A, B, \ldots, N$ несовместны.
3. $AB \ldots N = X$.	3. Событие X заключается в одновременной реализации всех событий $A, B, \ldots, N$.
4. $A \dot{+} B \dot{+} \ldots \dot{+} N = X$.	4. Событие X заключается в наступлении по крайней мере одного из событий $A, B, \ldots, N$.
5. Дополнительное множество $\overline{A}$.	5. Противоположное событие $\overline{A}$, состоящее в ненаступлении события A.
6. $A = 0$.	6. A невозможно.
7. $A = E$.	7. A должно необходимо наступить.
8. Система $\mathfrak{A}$ множеств $A_1, A_2, \ldots, A_n$ образует *разложение* множества E, если $A_1 + A_2 + \ldots + A_n = E$. (Это уже предполагает, что множества A_i попарно не пересекаются.)	8. Испытание $\mathfrak{A}$ заключается в том, что устанавливают, какое из событий $A_1, A_2, \ldots, A_n$ происходит. $A_1, A_2, \ldots, A_n$ называются при этом возможными исходами испытания $\mathfrak{A}$.
9. B является подмножеством A: $B \subset A$.	9. Из осуществления события B следует с необходимостью осуществление A.

[1]) Ср. § 4, формула (3).

§ 4. Непосредственные следствия из аксиом, условные вероятности, теорема Байеса

Из $A+\overline{A}=E$ и аксиом IV и V следует:

$$\mathsf{P}(A)+\mathsf{P}(\overline{A})=1, \tag{1}$$

$$\mathsf{P}(\overline{A})=1-\mathsf{P}(A). \tag{2}$$

Так как $\overline{E}=0$, то в частности получаем:

$$\mathsf{P}(0)=0. \tag{3}$$

Если $A, B, \ldots, N$ несовместны, то из аксиомы IV следует формула:

$$\mathsf{P}(A+B+\ldots+N)=\mathsf{P}(A)+\mathsf{P}(B)+\ldots+\mathsf{P}(N) \tag{4}$$

(*теорема сложения*).

Если $\mathsf{P}(A)>0$, то частное

$$\mathsf{P}_A(B)=\frac{\mathsf{P}(AB)}{\mathsf{P}(A)} \tag{5}$$

называют *условной вероятностью* события B при условии A.

Из (5) непосредственно следует:

$$\mathsf{P}(AB)=\mathsf{P}(A)\,\mathsf{P}_A(B). \tag{6}$$

Заключение по индукции дает затем общую формулу:

$$\mathsf{P}(A_1A_2\ldots A_n)=$$
$$=\mathsf{P}(A_1)\,\mathsf{P}_{A_1}(A_2)\,\mathsf{P}_{A_1A_2}(A_3)\ldots\mathsf{P}_{A_1A_2\ldots A_{n-1}}(A_n) \tag{7}$$

(*теорема умножения*).

Легко доказываются также следующие формулы:

$$\mathsf{P}_A(B)\geqslant 0, \tag{8}$$

$$\mathsf{P}_A(E)=1, \tag{9}$$

$$\mathsf{P}_A(B+C)=\mathsf{P}_A(B)+\mathsf{P}_A(C). \tag{10}$$

Сравнивая формулы (8) и (10) с аксиомами III—V, получаем, что система множеств $\mathfrak{F}$ вместе с функцией множеств $\mathsf{P}_A(B)$ (при закрепленном множестве A) образует поле вероятностей. Следовательно, *все доказанные для вероятностей* $\mathsf{P}(B)$ *общие теоремы справедливы для условных вероятностей* $\mathsf{P}_A(B)$ (при фиксированном событии A). Легко также заметить, что

$$\mathsf{P}_A(A)=1. \tag{11}$$

Из (6) и из аналогичной формулы

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

получаем важную формулу:

$$\mathsf{P}_B(A) = \frac{\mathsf{P}(A)\,\mathsf{P}_A(B)}{\mathsf{P}(B)}, \tag{12}$$

содержащую, собственно, теорему Байеса.

ТЕОРЕМА О ПОЛНОЙ ВЕРОЯТНОСТИ. Пусть $A_1 + A_2 + \ldots + A_n = E$ (это предполагает, что события $A_1, A_2, \ldots, A_n$ несовместны), а X произвольно. Тогда

$$\mathsf{P}(X) = \mathsf{P}(A_1)\,\mathsf{P}_{A_1}(X) + \mathsf{P}(A_2)\,\mathsf{P}_{A_2}(X) + \ldots + \mathsf{P}(A_n)\,\mathsf{P}_{A_n}(X). \tag{13}$$

Доказательство.

$$X = A_1X + A_2X + \ldots + A_nX;$$

согласно (4) получаем:

$$\mathsf{P}(X) = \mathsf{P}(A_1X) + \mathsf{P}(A_2X) + \ldots + \mathsf{P}(A_nX);$$

согласно (6) при этом имеет место

$$\mathsf{P}(A_iX) = \mathsf{P}(A_i)\,\mathsf{P}_{A_i}(X).$$

ТЕОРЕМА БАЙЕСА. Пусть $A_1 + A_2 + \ldots + A_n = E$, а X произвольно, тогда имеет место:

$$\mathsf{P}_X(A_i) = \frac{\mathsf{P}(A_i)\,\mathsf{P}_{A_i}(X)}{\mathsf{P}(A_1)\,\mathsf{P}_{A_1}(X) + \mathsf{P}(A_2)\,\mathsf{P}_{A_2}(X) + \ldots + \mathsf{P}(A_n)\,\mathsf{P}_{A_n}(X)} \tag{14}$$

$$(i = 1, 2, \ldots, n).$$

При этом $A_1, A_2, \ldots, A_n$ часто называют „гипотезами" и говорят, что формула (14) дает вероятность $\mathsf{P}_X(A_i)$ гипотезы A_i после наступления события X. [$\mathsf{P}(A_i)$ обозначает при этом априорную вероятность A_i.]

Доказательство. Согласно формуле (12) имеем:

$$\mathsf{P}_X(A_i) = \frac{\mathsf{P}(A_i)\,\mathsf{P}_{A_i}(X)}{\mathsf{P}(X)}.$$

Для получения формулы (14) остается только заменить вероятность $\mathsf{P}(X)$ ее выражением (13) по теореме о полной вероятности.

§ 5. Независимость

Понятие о взаимной *независимости* двух или нескольких опытов занимает в известном смысле центральное место в теории вероятностей. В самом деле, мы уже видели, что теорию вероятностей с математической точки зрения можно рассматривать как специальное применение общей теории аддитивных функций множеств. Естественно задать вопрос, каким же образом теория вероятностей развилась в большую, обладающую своими собственными методами самостоятельную науку.

Для ответа на этот вопрос следует указать на ту специализацию, которую получают в теории вероятностей общие проблемы, касающиеся аддитивных функций множеств.

То обстоятельство, что наша аддитивная функция множеств $\mathsf{P}(A)$ неотрицательна и удовлетворяет условию $\mathsf{P}(E)=1$, не обусловливает еще собой возникновения новых глубоких проблем. Случайные величины (ср. третью главу) с математической точки зрения представляют не что иное, как функции, измеримые по отношению к $\mathsf{P}(A)$, а их математические ожидания являются абстрактными интегралами Лебега. [Эта аналогия была впервые полностью разъяснена в работах Фреше [1].] Введение упомянутых понятий не может, следовательно, еще доставить никакого базиса для развития большой оригинальной теории.

Исторически независимость испытаний и случайных величин явилась тем математическим понятием, которое придало теории вероятностей своеобразный отпечаток. Классические работы Лапласа, Пуассона, Чебышева, Маркова, Ляпунова, Мизеса и Бернштейна действительно посвящены в основном изучению рядов независимых случайных величин. Если в новейших исследованиях (Марков, Бернштейн и др.) часто отказываются от предположения полной независимости, то оказываются принужденными для получения достаточно содержательных результатов ввести аналогичные ослабленные предположения (ср. в этой главе § 6, о цепях Маркова). Мы приходим, следовательно, к тому, чтобы в понятии независимости видеть по крайней мере первый зародыш своеобразной проблематики теории вероятностей — обстоятельство, которое в этой книге будет мало выделяться, так как в ней мы занимаемся, главным образом, только логической подготовкой к собственно теоретико-вероятностным исследованиям.

Соответственно этому одной из важнейших задач философии естественных наук, после разъяснения пресловутого вопроса о сущности самого понятия вероятности, являются выяснение и уточнение тех предпосылок, при которых можно какие-либо данные действительные явления рассматривать как независимые. Этот вопрос выходит, однако, за пределы нашей книги.

---

[1] Ср. *Fréchet* [1] и [2].

Перейдем к определению независимости.

Пусть даны n испытаний $\mathfrak{A}^{(1)}, \mathfrak{A}^{(2)}, \ldots, \mathfrak{A}^{(n)}$, т. е. n разложений

$$E = A_1^{(i)} + A_2^{(i)} + \ldots + A_{r_i}^{(i)} \quad (i = 1, 2, \ldots, n)$$

основного множества E. Тогда можно задать $r = r_1 r_2 \ldots r_n$ вероятностей

$$p_{q_1 q_2 \ldots q_n} = \mathsf{P}(A_{q_1}^{(1)} A_{q_2}^{(2)} \ldots A_{q_n}^{(n)}) \geqslant 0$$

вообще произвольно, при единственном условии

$$\sum_{q_1, q_2, \ldots, q_n} p_{q_1 q_2 \ldots q_n} = 1 \text{ }^{1)}. \qquad (1)$$

Определение I. n испытаний $\mathfrak{A}^{(1)}, \mathfrak{A}^{(2)}, \ldots, \mathfrak{A}^{(n)}$ называются *взаимно независимыми*, если для любых $q_1, q_2, \ldots, q_n$ имеет место равенство:

$$\mathsf{P}(A_{q_1}^{(1)} A_{q_2}^{(2)} \ldots A_{q_n}^{(n)}) = \mathsf{P}(A_{q_1}^{(1)}) \mathsf{P}(A_{q_2}^{(2)}) \ldots \mathsf{P}(A_{q_n}^{(n)}). \qquad (2)$$

Среди r уравнений (2) имеется только $r - r_1 - r_2 - \ldots - r_n + n - 1$ независимых [2]).

ТЕОРЕМА I. Если n испытаний $\mathfrak{A}^{(1)}, \mathfrak{A}^{(2)}, \ldots, \mathfrak{A}^{(n)}$ взаимно независимы, то из них любые m $(m < n)$ испытаний $\mathfrak{A}^{(i_1)}, \mathfrak{A}^{(i_2)}, \ldots, \mathfrak{A}^{(i_m)}$ также независимы [3]).

---

[1]) Поле вероятностей с произвольными вероятностями, удовлетворяющими только упомянутым условиям, можно построить следующим образом: множество E составляется из r элементов $\xi_{q_1 q_2 \ldots q_n}$, соответствующие элементарные вероятности пусть будут $p_{q_1 q_2 \ldots q_n}$, и $A_q^{(i)}$ определяется как множество всех $\xi_{q_1 q_2 \ldots q_n}$, для которых $q_i = q$.

[2]) В самом деле, в случае независимости можно выбрать произвольно только $r_1 + r_2 + \ldots + r_n$ вероятностей $p_q^{(i)} = \mathsf{P}(A_q^{(i)})$, притом так чтобы соблюдались n условий

$$\sum_q p_q^{(i)} = 1.$$

Следовательно, в общем случае имеем $r - 1$ степеней свободы, а в случае независимости только $r_1 + r_2 + \ldots + r_n - n$.

[3]) Для доказательства достаточно показать, что из взаимной независимости n разложений следует взаимная независимость первых $n - 1$ из них. Примем, что уравнения (2) удовлетворены. Тогда

$$\mathsf{P}(A_{q_1}^{(1)} A_{q_2}^{(2)} \ldots A_{q_{n-1}}^{(n-1)}) = \sum_{q_n} \mathsf{P}(A_{q_1}^{(1)} A_{q_2}^{(2)} \ldots A_{q_n}^{(n)}) =$$

$$= \mathsf{P}(A_{q_1}^{(1)}) \mathsf{P}(A_{q_2}^{(2)}) \ldots \mathsf{P}(A_{q_{n-1}}^{(n-1)}) \sum_{q_n} \mathsf{P}(A_{q_n}^{(n)}) =$$

$$= \mathsf{P}(A_{q_1}^{(1)}) \mathsf{P}(A_{q_2}^{(2)}) \ldots \mathsf{P}(A_{q_{n-1}}^{(n-1)}).$$

В случае независимости имеют, следовательно, место равенства:

$$\mathsf{P}(A_{q_1}^{(i_1)} A_{q_2}^{(i_2)} \ldots A_{q_m}^{(i_m)}) = \mathsf{P}(A_{q_1}^{(i_1)})\,\mathsf{P}(A_{q_2}^{(i_2)}) \ldots \mathsf{P}(A_{q_m}^{(i_m)}) \qquad (3)$$

(все i_k должны быть при этом различны).

Определение II. n событий $A_1, A_2, \ldots, A_n$ называются *взаимно независимыми*, если разложения (испытания)

$$E = A_k + \bar{A}_k \quad (k = 1, 2, \ldots, n)$$

являются независимыми.

В этом случае $r_1 = r_2 = \ldots = r_n = 2$; $r = 2^n$, следовательно, из 2^n уравнений (2) имеется независимых только $2^n - n - 1$. Для независимости событий $A_1, A_2, \ldots, A_n$ необходимы и достаточны следующие уравнения [1]):

$$\mathsf{P}(A_{i_1} A_{i_2} \ldots A_{i_m}) = \mathsf{P}(A_{i_1})\,\mathsf{P}(A_{i_2}) \ldots \mathsf{P}(A_{i_m}), \qquad (4)$$
$$m = 1, 2, \ldots, n, \; 1 \leqslant i_1 < i_2 < \ldots < i_m \leqslant n.$$

Все эти уравнения независимы между собой.

В случае $n = 2$ получаем из (4) только одно ($2^2 - 2 - 1 = 1$) условие для независимости двух событий A_1 и A_2:

$$\mathsf{P}(A_1 A_2) = \mathsf{P}(A_1)\,\mathsf{P}(A_2). \qquad (5)$$

Система уравнений (2) состоит в этом случае, кроме (5), еще из трех уравнений:

$$\mathsf{P}(A_1 \bar{A}_2) = \mathsf{P}(A_1)\,\mathsf{P}(\bar{A}_2),$$
$$\mathsf{P}(\bar{A}_1 A_2) = \mathsf{P}(\bar{A}_1)\,\mathsf{P}(A_2),$$
$$\mathsf{P}(\bar{A}_1 \bar{A}_2) = \mathsf{P}(\bar{A}_1)\,\mathsf{P}(\bar{A}_2),$$

которые, очевидно, следуют из (5) [2]).

Следует при этом еще заметить, что из попарной независимости событий $A_1, A_2, \ldots, A_n$, т. е. из соотношений:

$$\mathsf{P}(A_i A_j) = \mathsf{P}(A_i)\,\mathsf{P}(A_j) \quad (i \neq j),$$

---

[1]) Ср. *С. Н. Бернштейн* [1], стр. 47—57. Впрочем, читатель может это сам проверить без труда (заключение по индукции).

[2]) $$\mathsf{P}(A_1 \bar{A}_2) = \mathsf{P}(A_1) - \mathsf{P}(A_1 A_2) = \mathsf{P}(A_1) - \mathsf{P}(A_1)\,\mathsf{P}(A_2) =$$
$$= \mathsf{P}(A_1)\,\{1 - \mathsf{P}(A_2)\} = \mathsf{P}(A_1)\,\mathsf{P}(\bar{A}_2)$$

и т. д.

в случае $n > 2$ отнюдь не следует независимость этих событий [1]) [для этого необходимо наличие всех уравнений (4)].

При введении понятия независимости мы не пользовались условными вероятностями. Нашей целью было при этом возможно яснее изложить чисто математически сущность этого понятия. Его приложения опираются, однако, главным образом на свойства некоторых условных вероятностей. Если мы предположим, что все вероятности положительны, то из уравнений (3) следует, что

$$\mathsf{P}_{A_{q_1}^{(i_1)} A_{q_2}^{(i_2)} \ldots A_{q_{m-1}}^{(i_{m-1})}} (A_{q_m}^{(i_m)}) = \mathsf{P}(A_{q_m}^{(i_m)}) \text{ [2])}. \qquad (6)$$

Из наличия формул (6) следуют обратно, по теореме умножения [формула (7), § 4] формулы (2). Мы имеем, следовательно, следующую теорему.

ТЕОРЕМА II. В случае положительных вероятностей $\mathsf{P}(A_q^{(i)})$ необходимо и достаточно для независимости испытаний $\mathfrak{A}^{(1)}$, $\mathfrak{A}^{(2)}, \ldots, \mathfrak{A}^{(n)}$ следующее условие: условная вероятность исхода $A_q^{(i)}$ испытания $\mathfrak{A}^{(i)}$ при той гипотезе, что некоторые другие испытания $\mathfrak{A}^{(i_1)}$, $\mathfrak{A}^{(i_2)}, \ldots, \mathfrak{A}^{(i_k)}$ получили определенные исходы $A_{q_1}^{(i_1)}$, $A_{q_2}^{(i_2)}, \ldots, A_{q_k}^{(i_k)}$, равняется абсолютной вероятности $\mathsf{P}(A_q^{(i)})$.

На основе формул (4) доказывается аналогично следующая теорема.

ТЕОРЕМА III. Если все вероятности $\mathsf{P}(A_k)$ положительны, то для взаимной независимости событий $A_1, A_2, \ldots, A_n$ необходимо и достаточно наличие уравнений

$$\mathsf{P}_{A_{i_1} A_{i_2} \ldots A_{i_k}} (A_i) = \mathsf{P}(A_i) \qquad (7)$$

для любых попарно различных $i_1, i_2, \ldots, i_k, i$.

---

1) Это доказывается следующим простым примером (С. Н. Бернштейн): множество E состоит из четырех элементов $\xi_1, \xi_2, \xi_3, \xi_4$, соответствующие элементарные вероятности p_1, p_2, p_3, p_4 все полагаются равными $\frac{1}{4}$, и

$$A = \{\xi_1, \xi_2\}, \quad B = \{\xi_1, \xi_3\}, \quad C = \{\xi_1, \xi_4\}.$$

Легко тогда сосчитать, что

$$\mathsf{P}(A) = \mathsf{P}(B) = \mathsf{P}(C) = \frac{1}{2},$$

$$\mathsf{P}(AB) = \mathsf{P}(BC) = \mathsf{P}(AC) = \frac{1}{4} = \left(\frac{1}{2}\right)^2,$$

$$\mathsf{P}(ABC) = \frac{1}{4} \neq \left(\frac{1}{2}\right)^3.$$

2) Для доказательства следует вспомнить определение условной вероятности [формула (5), § 4] и заменить вероятности пересечений через произведения вероятностей по формуле (3).

В случае $n=2$ условия (7) сводятся к двум уравнениям:

$$\left.\begin{aligned} \mathsf{P}_{A_1}(A_2) &= \mathsf{P}(A_2), \\ \mathsf{P}_{A_2}(A_1) &= \mathsf{P}(A_1). \end{aligned}\right\} \quad (8)$$

Легко усмотреть, что уже одно лишь первое уравнение (8) представляет необходимое и достаточное условие независимости A_1 и A_2, если только $\mathsf{P}(A_1) > 0$.

§ 6. Условные вероятности как случайные величины, цепи Маркова

Пусть $\mathfrak{A}$ является разложением основного множества E:

$$E = A_1 + A_2 + \ldots + A_r,$$

а x — действительная функция элементарного события ξ, которая на каждом множестве A_q равняется соответствующей константе a_q. В этом случае говорят, что x — *случайная величина*, а сумму

$$\mathsf{E}(x) = \sum_q a_q\, \mathsf{P}(A_q)$$

называют *математическим ожиданием* величины x. Теория случайных величин и их математических ожиданий будет развита в третьей и четвертой главах, не ограничиваясь при этом случайными величинами, которые могут принимать только конечное число различных значений.

Случайную величину, которая на каждом множестве A_q принимает значение $\mathsf{P}_{A_q}(B)$, мы назовем *условной вероятностью события B после данного испытания* $\mathfrak{A}$ и обозначим через $\mathsf{P}_{\mathfrak{A}}(B)$. Два испытания тогда и только тогда независимы, если

$$\mathsf{P}_{\mathfrak{A}^{(1)}}(A_q^{(2)}) = \mathsf{P}(A_q^{(2)}) \quad (q = 1, 2, \ldots, r_2).$$

Если даны какие-либо разложения (испытания) $\mathfrak{A}^{(1)}, \mathfrak{A}^{(2)}, \mathfrak{A}^{(n)}$, то через

$$\mathfrak{A}^{(1)}\mathfrak{A}^{(2)}\ldots\mathfrak{A}^{(n)}$$

мы обозначим разложение множества E на произведения $A_{q_1}^{(1)}\ A_{q_2}^{(2)} \ldots A_{q_n}^{(n)}$. Испытания $\mathfrak{A}^{(1)}, \mathfrak{A}^{(2)}, \ldots, \mathfrak{A}^{(n)}$ тогда и только тогда взаимно независимы, если

$$\mathsf{P}_{\mathfrak{A}^{(1)}\mathfrak{A}^{(2)}\ldots\mathfrak{A}^{(k-1)}}(A_q^{(k)}) = \mathsf{P}(A_q^{(k)})$$

при любом выборе k и q [1]).

[1]) Необходимость этих условий следует из теоремы II, § 5, а что они также являются достаточными, можно заключить непосредственно из теоремы умножения [формула (7), § 4].

Определение. Последовательность испытаний

$$\mathfrak{A}^{(1)}, \mathfrak{A}^{(2)}, \ldots, \mathfrak{A}^{(n)}, \ldots$$

образует цепь Маркова, если при любых n и q

$$\mathsf{P}_{\mathfrak{A}^{(1)}\mathfrak{A}^{(2)}\ldots\mathfrak{A}^{(n-1)}}(A_q^{(n)}) = \mathsf{P}_{\mathfrak{A}^{(n-1)}}(A_q^{(n)}).$$

Цепи Маркова образуют естественное обобщение последовательностей взаимно независимых испытаний. Если положить

$$p_{q_n q_m}(m, n) = \mathsf{P}_{A_{q_m}^{(m)}}(A_{q_n}^{(n)}) \quad (m < n),$$

то основная формула теории цепей Маркова будет иметь вид:

$$p_{q_k q_n}(k, n) = \sum_{q_m} p_{q_k q_m}(k, m)\, p_{q_m q_n}(m, n) \quad (k < m < n). \tag{1}$$

Обозначив матрицу $\|p_{q_m q_n}(m, n)\|$ через $p(m, n)$, можно (1) записать в следующем виде:

$$p(k, n) = p(k, m)\, p(m, n) \quad (k < m < n)\ ^{1}). \tag{2}$$

---

[1]) По поводу дальнейшего развития теории цепей Маркова ср. *R. v. Mises* [1], § 16 и *B. Hostinsky*, Méthodes générales du calcul des probabilités. „Mém. Sci. math.“, вып. 52, Paris 1931.

II. БЕСКОНЕЧНЫЕ ПОЛЯ ВЕРОЯТНОСТЕЙ

§ 1. Аксиома непрерывности

Мы обозначаем, как обычно, через $\mathfrak{D}_m A_m$ пересечение множеств A_m (в конечном или бесконечном числе) и через $\mathfrak{S}_m A_m$ их сумму. Только в случае непересекающихся множеств A_m пишут $\sum_m A_m$ вместо $\mathfrak{S}_m A_m$. Следовательно,

$$\mathfrak{S}_m A_m = A_1 \dot{+} A_2 \dot{+} A_3 \dot{+} \ldots,$$

$$\sum_m A_m = A_1 + A_2 + A_3 + \ldots,$$

$$\mathfrak{D}_m A_m = A_1 A_2 A_3 \ldots$$

При всех дальнейших рассмотрениях мы предполагаем, что кроме аксиом I—V выполняется еще следующая аксиома.

VI. *Для убывающей последовательности*

$$A_1 \supset A_2 \supset A_3 \supset \ldots \supset A_n \supset \ldots \tag{1}$$

событий из $\mathfrak{F}$, *такой, что*

$$\mathfrak{D}_n A_n = 0, \tag{2}$$

имеет место равенство:

$$\lim_{n \to \infty} \mathsf{P}(A_n) = 0. \tag{3}$$

Во всем дальнейшем изложении мы называем *полем вероятностей* только такое поле вероятностей в смысле главы первой, которое, кроме того, удовлетворяет аксиоме VI. Поля вероятностей в смысле главы первой можно называть *полями вероятностей в расширенном смысле*.

Если система множеств $\mathfrak{F}$ конечна, аксиома VI следует из аксиом I—V. В самом деле, в этом случае существует только конечное число различных множеств в последовательности (1);

пусть A_k — наименьшее из них, тогда все множества A_{k+p} совпадают с A_k, и мы, следовательно, получаем:

$$A_k = A_{k+p} = \mathfrak{D}_n (A_n) = 0,$$

$$\lim \mathsf{P}(A_n) = \mathsf{P}(0) = 0.$$

Все примеры с конечными полями вероятностей из первой главы удовлетворяют, следовательно, также аксиоме VI. Система аксиом I—VI является, таким образом, *непротиворечивой* и *неполной*.

Напротив, для бесконечных полей аксиома непрерывности VI является независимой от аксиом I—V. Так как новая аксиома существенна лишь для бесконечных полей вероятностей, то является почти невозможным разъяснить ее эмпирическое значение, например, так, как это было вкратце проделано для аксиом I—V в § 2 главы первой. При описании какого-либо действительно наблюдаемого случайного процесса можно получать только конечные поля вероятностей. Бесконечные поля вероятностей появляются только как идеализированные схемы действительных случайных явлений. Мы произвольно ограничиваемся при этом такими схемами, которые удовлетворяют аксиоме VI. Это ограничение оказалось целесообразным в самых различных исследованиях.

РАСШИРЕННАЯ ТЕОРЕМА СЛОЖЕНИЯ. Если $A_1, A_2, \ldots, A_n, \ldots$ и A принадлежат к $\mathfrak{F}$, то из

$$A = \sum_n A_n \tag{4}$$

следует равенство:

$$\mathsf{P}(A) = \sum_n \mathsf{P}(A_n). \tag{5}$$

Доказательство. Положим

$$R_n = \sum_{m > n} A_m.$$

Тогда, очевидно,

$$\mathfrak{D}_n (R_n) = 0,$$

и, следовательно, по аксиоме VI

$$\lim_{n \to \infty} \mathsf{P}(R_n) = 0. \tag{6}$$

С другой стороны, по теореме сложения

$$\mathsf{P}(A) = \mathsf{P}(A_1) + \mathsf{P}(A_2) + \ldots + \mathsf{P}(A_n) + \mathsf{P}(R_n). \tag{7}$$

Из (6) и (7) следует непосредственно (5).

Итак, мы доказали, что *вероятность* $\mathsf{P}(A)$ *является на* $\mathfrak{F}$ *вполне аддитивной функцией множеств.* Обратно, аксиомы V и VI имеют место для всякой определенной на каком-либо теле множеств $\mathfrak{F}$ вполне аддитивной функции множеств [1]). Можно, следовательно, понятие поля вероятностей определить следующим образом. *Пусть* E — *произвольное множество,* $\mathfrak{F}$ — *тело подмножеств из* E, *содержащее* E, *а* $\mathsf{P}(A)$ — *неотрицательная определенная на* $\mathfrak{F}$ *вполне аддитивная функция множеств; тело множеств* $\mathfrak{F}$ *совместно с функцией множеств* $\mathsf{P}(A)$ *образуют тогда поле вероятностей.*

ТЕОРЕМА О ПОКРЫТИЯХ. Если $A_1, A_2, \ldots, A_n, \ldots$ принадлежат $\mathfrak{F}$ и

$$A \subset \mathfrak{S}_n A_n, \tag{8}$$

то

$$\mathsf{P}(A) \leqslant \sum_n \mathsf{P}(A_n). \tag{9}$$

Доказательство.

$$A = A \mathfrak{S}_n A_n = AA_1 + A(A_2 - A_2A_1) + A(A_3 - A_3A_2 - A_3A_1) + \ldots$$

$$\mathsf{P}(A) = \mathsf{P}(AA_1) + \mathsf{P}\{A(A_2 - A_2A_1)\} + \ldots \leqslant \mathsf{P}(A_1) + \mathsf{P}(A_2) + \ldots$$

§ 2. Борелевские поля вероятностей

Тело множеств $\mathfrak{F}$ называется *борелевским телом*, если все счетные суммы $\sum_n A_n$ множеств A_n из $\mathfrak{F}$ принадлежат $\mathfrak{F}$. Борелевские тела называют также вполне аддитивными системами множеств. Из формулы

$$\mathfrak{S}_n A_n = A_1 + (A_2 - A_2A_1) + (A_3 - A_3A_2 - A_3A_1) + \ldots \tag{1}$$

можно заключить, что борелевское тело содержит также все суммы $\mathfrak{S}_n A_n$, составленные из счетного числа принадлежащих ему множеств A_n. Из формулы

$$\mathfrak{D}_n A_n = E - \mathfrak{S}_n \bar{A}_n \tag{2}$$

следует то же для пересечений множеств.

Поле вероятностей является борелевским полем вероятностей, если соответствующее тело множеств является борелевским. Только при борелевских полях теория вероятностей получает полную свободу действия, не связанную с опасностью

[1]) См., например, *O. Nikodym*, Sur une généralisation des intégrales de M. J. Radon, „Fundam. Math.", т. 15, 1930, стр. 136.

притти к событиям, которые не имеют никакой вероятности. Мы теперь покажем, что можно ограничиться рассмотрением борелевских полей вероятностей. Это будет следовать из так называемой теоремы о продолжении, к которой мы сейчас перейдем.

Пусть дано поле вероятностей $(\mathfrak{F}, \mathsf{P})$. Как известно [1], существует наименьшее борелевское тело $B\mathfrak{F}$, содержащее $\mathfrak{F}$. Имеет место

ТЕОРЕМА О ПРОДОЛЖЕНИИ. *Определенную на $\mathfrak{F}$ неотрицательную вполне аддитивную функцию множеств $\mathsf{P}(A)$ можно всегда продолжить с сохранением обоих свойств (неотрицательности и полной аддитивности) на все множества $B\mathfrak{F}$, и притом единственным способом.*

Расширенное тело $B\mathfrak{F}$ образует вместе с продолженной функцией множеств $\mathsf{P}(A)$ некоторое поле вероятностей $(B\mathfrak{F}, \mathsf{P})$. Это поле вероятностей $(B\mathfrak{F}, \mathsf{P})$ назовем *борелевским расширением поля* $(\mathfrak{F}, \mathsf{P})$.

Доказательство этой теоремы, которая относится к теории аддитивных функций множеств и которая в основном должна быть известна в различных других трактовках, проводится следующим образом.

Пусть A — некоторое произвольное подмножество E; мы определяем $\mathsf{P}^*(A)$ как нижнюю границу сумм

$$\sum_n \mathsf{P}(A_n)$$

для всех покрытий

$$A \subset \mathfrak{S}_n A_n$$

множества A конечным или счетным числом множеств A_n из $\mathfrak{F}$. Легко доказать, что $\mathsf{P}^*(A)$ является внешней мерой в смысле Каратеодори [2]. Согласно теореме о покрытиях (§ 1), $\mathsf{P}^*(A)$ совпадает с $\mathsf{P}(A)$ для всех множеств из $\mathfrak{F}$. Далее доказывается, что все множества из $\mathfrak{F}$ измеримы в смысле Каратеодори. Так как все измеримые множества образуют борелевское тело, то, следовательно, все множества из $B\mathfrak{F}$ являются измеримыми. Функция множеств $\mathsf{P}^*(A)$ является, следовательно, вполне аддитивной на $B\mathfrak{F}$, и мы можем положить на $B\mathfrak{F}$

$$\mathsf{P}(A) = \mathsf{P}^*(A).$$

Этим доказано существование продолжения. Единственность продолжения следует сразу же из минимальных свойств тела $B\mathfrak{F}$.

---

[1] *Hausdorff*, Mengenlehre, 1927, стр. 79.

[2] *Carathéodory*, Vorlesungen über reelle Funktionen, 1918, стр. 237—258.

Замечание. Если множества (события) A из $\mathfrak{F}$ могут иметь смысл в качестве действительных и наблюдавшихся (хотя бы приближенно) событий, то отсюда еще не следует, что множества из расширенного тела $B\mathfrak{F}$ допускают такое же разумное истолкование в качестве действительно наблюдавшихся событий. Может случиться, что поле вероятностей $(\mathfrak{F}, \mathsf{P})$ можно рассматривать в качестве (хотя бы идеализированного) образа действительных случайных событий, в то время как расширенное поле вероятностей $(B\mathfrak{F}, \mathsf{P})$ остается чисто математическим построением.

Множества из $B\mathfrak{F}$ мы рассматриваем только как „идеальные события", которым ничего не соответствует во внешнем мире. Если, однако, рассуждение, которое использует вероятности таких идеальных событий, приводит к определению вероятностей действительного события из $\mathfrak{F}$, то это определение, очевидно, автоматически будет непротиворечивым и с эмпирической точки зрения.

§ 3. Примеры бесконечных полей вероятностей

I. Еще в первой главе, § 1, мы строили различные конечные поля вероятностей. Пусть теперь

$$E = \{\xi_1, \xi_2, \ldots, \xi_n, \ldots\}$$

— счетное множество, а $\mathfrak{F}$ совпадает с совокупностью всех подмножеств множества E.

Все возможные поля вероятностей с таким множеством $\mathfrak{F}$ получаются следующим образом: берется последовательность неотрицательных чисел p_n при условии

$$p_1 + p_2 + \ldots + p_n + \ldots = 1$$

и полагается для каждого множества A

$$\mathsf{P}(A) = \sum_n{}' p_n,$$

причем суммирование $\sum{}'$ распространяется на все те индексы n, для которых ξ_n принадлежит A. Эти поля вероятностей, очевидно, являются борелевскими.

II. Мы теперь предположим, что E представляет собой действительную числовую ось. Сначала пусть при этом $\mathfrak{F}$ образовано из всевозможных сумм полуоткрытых интервалов $[a, b) = \{a \leqslant \xi < b\}$ (при этом мы рассматриваем наряду с собственными интервалами с конечными a и b также и несобственные $[-\infty, a)$, $[a, +\infty)$ и $[-\infty, +\infty)$). Легко убедиться, что $\mathfrak{F}$ является телом. По теореме о продолжении можно, однако, каждое поле вероятностей из $\mathfrak{F}$ расширить в подобное поле из $B\mathfrak{F}$. Система множеств $B\mathfrak{F}$ в нашем случае является не чем иным, как системой всех борелевских множеств числовой прямой. Перейдем теперь к следующему случаю.

III. Предполагается, что E — действительная числовая ось, а $\mathfrak{F}$ состоит из вех борелевских точечных множеств этой прямой. Для построения поля вероятностей с данным телом $\mathfrak{F}$ достаточно определить любую неотрицательную вполне аддитивную функцию множеств $\mathsf{P}(A)$ из $\mathfrak{F}$, удовлетворяющую условию $\mathsf{P}(E) = 1$. Такая функция, как известно [1]), однозначно определяется своими значениями

$$\mathsf{P}[-\infty, x) = F(x) \tag{1}$$

для специальных интервалов $[-\infty, x)$. Функцию $F(x)$ называют *функцией распределения* ξ. Далее доказывается (третья глава, § 2), что $F(x)$ не убывает, непрерывна слева и имеет следующие предельные значения:

$$\begin{gathered}\lim_{x \to -\infty} F(x) = F(-\infty) = 0,\\ \lim_{x \to +\infty} F(x) = F(+\infty) = 1.\end{gathered} \tag{2}$$

Обратно, если дана функция $F(x)$, удовлетворяющая этим условиям, то она всегда определяет неотрицательную вполне аддитивную функцию множеств $\mathsf{P}(A)$, такую, что $\mathsf{P}(E) = 1$ [2]).

IV. Пусть теперь за основное множество E принимается n-мерное евклидово координатное пространство R^n, т. е. множество всех упорядоченных комплексов $\xi = \{x_1, x_2, \ldots, x_n\}$ из n действительных чисел. $\mathfrak{F}$ пусть состоит при этом из всех борелевских точечных множеств [3]) пространства R^n. На основании рассуждений, аналогичных приведенным в примере II, мы можем отказаться от рассмотрения более узких систем множеств, например системы всех n-мерных интервалов.

За вероятностную функцию $\mathsf{P}(A)$ здесь, как всегда, можно принять любую неотрицательную и вполне аддитивную функцию множеств, определенную на $\mathfrak{F}$ и удовлетворяющую условию $\mathsf{P}(E) = 1$. Такая функция множеств однозначно определяется, если даны ее значения:

$$\mathsf{P}(L_{a_1, a_2, \ldots, a_n}) = F(a_1, a_2, \ldots, a_n) \tag{3}$$

для специальных множеств $L_{a_1, a_2, \ldots; a_n}$, где $L_{a_1, a_2, \ldots; a_n}$ означает множество всех ξ, для которых $x_i < a_i$ $(i = 1, 2, \ldots, n)$.

---

[1]) Ср., например, *Лебег*, Интегрирование и отыскание примитивных функций, ГТТИ, 1934, стр. 127—132.

[2]) Ср. предыдущую сноску.

[3]) Определение борелевских множеств в R^n ср. *Hausdorff*, Mengenlehre, 1937, стр. 177—181.

За функцию $F(a_1, a_2, \ldots, a_n)$ можно при этом выбрать любую функцию, не убывающую по всем переменным и непрерывную слева, которая удовлетворяет следующим условиям:

$$\left.\begin{aligned} &\lim_{a_i \to -\infty} F(a_1, a_2, \ldots, a_n) = \\ &= F(a_1, \ldots, a_{i-1}, -\infty, \ldots, a_n) = 0 \quad (i = 1, 2, \ldots, n), \\ &\lim_{a_1 \to +\infty,\, a_2 \to +\infty,\, \ldots,\, a_n \to +\infty} F(a_1, a_2, \ldots, a_n) = F(+\infty, +\infty, \ldots, +\infty) = 1. \end{aligned}\right\} \quad (4)$$

$F(a_1, a_2, \ldots, a_n)$ называют функцией распределения величин $x_1, x_2, \ldots, x_n$.

Рассмотрение полей вероятностей вышеопределенного типа достаточно для всех классических проблем теории вероятностей[1]). В частности, вероятностная функция в R^n может быть определена так: берется любая определенная в R^n неотрицательная функция точки $f(x_1, x_2, \ldots, x_n)$, такая, что

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\ldots\int_{-\infty}^{+\infty} f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n = 1,$$

и полагается

$$\mathsf{P}(A) = \int\int_A \ldots \int f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n. \qquad (5)$$

$f(x_1, x_2, \ldots, x_n)$ является в этом случае *плотностью вероятности* в точке $(x_1, x_2, \ldots, x_n)$ (ср. глава третья, § 2).

Другой тип вероятностных функций в R^n получается следующим образом. Пусть $\{\xi_i\}$ — последовательность точек из R^n и $\{p_i\}$ — последовательность неотрицательных действительных чисел, такая, что $\sum p_i = 1$, тогда, так же как и в примере I, полагаем

$$\mathsf{P}(A) = \sum{}' p_i,$$

причем суммирование $\sum'$ распространяется на все те индексы, для которых ξ принадлежит A. Оба упомянутых здесь типа

[1]) Ср., например, *R. v. Mises* [1], стр. 13—19. Здесь требуется существование вероятностей для „всех практически возможных" множеств n-мерного пространства.

вероятностных функций в R^n не исчерпывают всех возможностей, хотя ими обычно довольствуются в приложениях теории вероятностей.

Можно, однако, себе представить кроме этой классической области также и другие, интересные для приложений задачи, в которых элементарные события определяются с помощью бесконечного числа координат. Соответствующие поля вероятностей мы исследуем ближе после введения некоторых необходимых для этого вспомогательных понятий (ср. глава третья, § 3).

III. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

§ 1. Вероятностные функции

Пусть дано отображение множества E в множество E', состоящее из каких-либо элементов, т. е. определенная на E однозначная функция $u(\xi)$, значения которой принадлежат множеству E'. Каждому подмножеству A' из E' мы ставим в соответствие в качестве его прообраза в E множество $u^{-1}(A')$ всех элементов из E, которые отображаются в один из элементов A'. Пусть далее $\mathfrak{F}^{(u)}$ — система всех подмножеств A' из E', прообразы которых принадлежат к телу множеств $\mathfrak{F}$. $\mathfrak{F}^{(u)}$ тогда также является телом; если при этом $\mathfrak{F}$ — борелевское тело, то то же имеет место и для $\mathfrak{F}^{(u)}$. Мы полагаем теперь

$$\mathsf{P}^{(u)}(A') = \mathsf{P}\{u^{-1}(A')\}. \tag{1}$$

Эта определенная на $\mathfrak{F}^{(u)}$ функция множеств $\mathsf{P}^{(u)}$ удовлетворяет относительно тела $\mathfrak{F}^{(u)}$ всем нашим аксиомам I—VI и, следовательно, является вероятностной функцией на $\mathfrak{F}^{(u)}$. Прежде чем перейти к доказательству всех только что указанных фактов, мы сформулируем уже теперь следующее определение.

Определение. Пусть дана однозначная функция $u(\xi)$ случайного события ξ. Тогда функция $\mathsf{P}^{(u)}(A')$, определенная формулой (1), называется *вероятностной функцией* от u.

Примечание 1. При исследовании поля вероятностей $(\mathfrak{F}, \mathsf{P})$ функцию $\mathsf{P}(A)$ называют просто вероятностной функцией, а $\mathsf{P}^{(u)}(A')$ — вероятностной функцией от u. В случае $u(\xi) = \xi$, $\mathsf{P}^{(u)}(A')$ совпадает с $\mathsf{P}(A)$.

Примечание 2. Событие $u^{-1}(A')$ состоит в том, что $u(\xi)$ принадлежит множеству A'. Следовательно, $\mathsf{P}^{(u)}(A')$ есть вероятность того, что $u(\xi) \subset A'$.

Нам осталось доказать вышеупомянутые свойства $\mathfrak{F}^{(u)}$ и $\mathsf{P}^{(u)}$. Они следуют, однако, из одного единственного факта, а именно следующего:

ЛЕММА. *Сумма, пересечение и разность каких-либо прообразных множеств $u^{-1}(A')$ являются прообразами соответствующих сумм, пересечений и разностей множества A'.*

Доказательство этой леммы предоставляется читателю.

Пусть теперь A' и B' — два множества из $\mathfrak{F}^{(u)}$, их прообразы A и B принадлежат тогда $\mathfrak{F}$; так как $\mathfrak{F}$ — тело, то множества AB, $A+B$ и $A-B$ также принадлежат $\mathfrak{F}$, но эти множества являются прообразами множеств $A'B'$, $A'+B'$ и $A'-B'$, следовательно, последние множества принадлежат к $\mathfrak{F}^{(u)}$. Итак, мы доказали, что $\mathfrak{F}^{(u)}$ — тело. Так же доказывается, что, если $\mathfrak{F}$ является борелевским телом, то то же справедливо и для $\mathfrak{F}^{(u)}$.

Далее ясно, что

$$\mathsf{P}^{(u)}(E') = \mathsf{P}\{u^{-1}(E')\} = \mathsf{P}(E) = 1.$$

Что $\mathsf{P}^{(u)}$ всегда неотрицательна, понятно само собой. Следовательно, остается доказать, что $\mathsf{P}^{(u)}$ вполне аддитивна (ср. конец § 1 главы второй).

Итак, пусть множества A'_n, а следовательно, и их прообразы $u^{-1}(A'_n)$ не пересекаются, отсюда следует, что

$$\mathsf{P}^{(u)}\left(\sum_n A'_n\right) = \mathsf{P}\left\{u^{-1}\left(\sum_n A'_n\right)\right\} = \mathsf{P}\left\{\sum_n u^{-1}(A'_n)\right\} =$$
$$= \sum_n \mathsf{P}\left\{u^{-1}(A'_n)\right\} = \sum \mathsf{P}^{(u)}(A'_n),$$

чем доказана полная аддитивность $\mathsf{P}^{(u)}$.

В заключение заметим еще следующее. Пусть $u_1(\xi)$ — функция, отображающая E в E_1, а $u_2(\xi')$ — другая функция, отображающая E' в E''. Тогда сложная функция $u_2 u_1(\xi)$ отображает множество E в E''. Мы рассмотрим теперь вероятностные функции $\mathsf{P}^{(u_1)}(A')$ и $\mathsf{P}^{(u)}(A'')$ для функций $u_1(\xi)$ и $u(\xi) = u_2 u_1(\xi)$. Эти две вероятностные функции связаны, как легко подсчитать, следующим соотношением:

$$\mathsf{P}^{(u)}(A'') = \mathsf{P}^{(u_1)}\{u_2^{-1}(A'')\}. \qquad (2)$$

§ 2. Определение случайных величин, функции распределения

Определение. Однозначную действительную функцию $x(\xi)$, определенную на основном множестве E, называют *случайной величиной*, если при каждом выборе действительного числа a множество $\{x < a\}$ всех ξ, для которых справедливо неравенство $x < a$, принадлежит к системе множеств $\mathfrak{F}$.

Эта функция $x(\xi)$ отображает основное множество E на множество R^1 всех действительных чисел. Наше определение случайной величины можно теперь сформулировать так. Действительная функция $x(\xi)$ является тогда и только тогда случайной величиной, если $\mathfrak{F}^{(x)}$ содержит каждый интервал вида $(-\infty; a)$.

Так как $\mathfrak{F}^{(x)}$ — тело, то оно содержит наряду с интервалами $(-\infty, a)$ также всевозможные конечные суммы полуоткрытых интервалов $[a; b)$. Если наше поле вероятностей борелевское, то $\mathfrak{F}$ и $\mathfrak{F}^{(x)}$ являются борелевскими телами, *следовательно, в этом случае* $\mathfrak{F}^{(x)}$ *содержит все борелевские множества* R^1. Вероятностную функцию случайной величины мы будем в дальнейшем обозначать через $\mathsf{P}^{(x)}(A')$. Она определена для всех множеств тела $\mathfrak{F}^{(x)}$. В частности в важнейшем случае борелевского поля вероятностей $\mathsf{P}^{(x)}$ определена для всех борелевских множеств R^1.

Определение. Функция

$$F^{(x)}(a) = \mathsf{P}^{(x)}(-\infty, a) = \mathsf{P}\{x < a\},$$

причем $-\infty$ и $+\infty$ допускаются в качестве значений a, называется *функцией распределения случайной величины* x.

Из определения непосредственно следует, что

$$F^{(x)}(-\infty) = 0, \quad F^{(x)}(+\infty) = 1. \tag{1}$$

Вероятность выполнения обоих неравенств $a \leqslant x < b$, очевидно, задается формулой:

$$\mathsf{P}\{x \subset [a, b)\} = F^{(x)}(b) - F^{(x)}(a). \tag{2}$$

Отсюда следует, что для $a < b$

$$F^{(x)}(a) \leqslant F^{(x)}(b),$$

это означает, что $F^{(x)}(a)$ — *неубывающая функция*.

Пусть далее $a_1 < a_2 < a_3 < \ldots < a_n < \ldots \to b$, тогда

$$\underset{n}{\mathfrak{D}}\{x \subset [a_n, b)\} = 0,$$

следовательно, согласно аксиоме непрерывности

$$F^{(x)}(b) - F^{(x)}(a_n) = \mathsf{P}\{x \subset [a_n, b)\}$$

стремится к нулю при $n \to \infty$. Отсюда видно, что $F^{(x)}(a)$ *непрерывна слева*.

Аналогично можно доказать формулы:

$$\lim_{a \to -\infty} F^{(x)}(a) = F^{(x)}(-\infty) = 0, \tag{3}$$

$$\lim_{a \to +\infty} F^{(x)}(a) = F^{(x)}(+\infty) = 1. \tag{4}$$

Если поле вероятностей $(\mathfrak{F}, \mathsf{P})$ — борелевское, то значения вероятностной функции $\mathsf{P}^{(x)}(A)$ для всех борелевских множеств A из R^1 однозначно определяются знанием функции распределения $F^{(x)}(a)$ (ср. глава вторая, § 3, III). Так как, главным образом, интересуются только этими значениями $\mathsf{P}^{(x)}(A)$, то функции распределения играют во всем дальнейшем изложении существенную роль.

Если функция распределения $F^{(x)}(a)$ диференцируема, то ее производную по a:

$$f^{(x)}(a) = \frac{d}{da} F^{(x)}(a)$$

называют *плотностью вероятности* x в точке a. Если при этом

$$F^{(x)}(a) = \int_{-\infty}^{a} f^{(x)}(a)\, da$$

для каждого a, то для каждого борелевского множества A вероятностную функцию можно выразить через $f^{(x)}(a)$ следующим образом:

$$\mathsf{P}^{(x)}(A) = \int_A f^{(x)}(a)\, da. \tag{5}$$

В этом случае говорят, что *распределение* x *непрерывно*. В общем случае пишут по аналогии:

$$\mathsf{P}^{(x)}(A) = \int_A dF^{(x)}(a). \tag{6}$$

Все введенные понятия допускают обобщение на случай условных вероятностей. Функция множеств

$$\mathsf{P}_B^{(x)}(A) = \mathsf{P}_B(x \subset A)$$

является условной вероятностной функцией от x при гипотезе B. Неубывающая функция

$$F_B^{(x)}(a) = \mathsf{P}_B(x < a)$$

есть соответствующая функция распределения, и, наконец [(в случае диференцируемости $F_B^{(x)}(a)$)],

$$f_B^{(x)}(a) = \frac{d}{da} F_B^{(x)}(a)$$

— условная плотность вероятности x в точке a при гипотезе B.

§ 3. Многомерные функции распределения

Пусть теперь даны n случайных величин $x_1, x_2, \ldots, x_n$. Точка $x = (x_1, x_2, \ldots, x_n)$ n-мерного пространства R^n является функцией элементарного события ξ. Следовательно, по общим правилам § 1, получаем тело множеств $\mathfrak{F}^{(x_1, x_2, \ldots, x_n)}$, состоящее из подмножеств пространства R^n, и определенную на $\mathfrak{F}^{(x_1, x_2, \ldots, x_n)}$ вероятностную функцию $\mathsf{P}^{(x_1, x_2, \ldots, x_n)}(A')$. Эту вероятностную функцию называют *n-мерной вероятностной функцией случайных величин* $x_1, x_2, \ldots, x_n$.

Тело $\mathfrak{F}^{(x_1, x_2, \ldots, x_n)}$ содержит, как это прямо следует из определения случайной величины, при каждом выборе i и a_i $(i = 1, 2, \ldots, n)$ множество всех точек R^n, для которых $x_i < a_i$. Следовательно, $\mathfrak{F}^{(x_1, x_2, \ldots, x_n)}$ содержит также и пересечение указанных множеств, т. е. множество $L_{a_1 a_2 \ldots a_n}$ всех точек R^n, для которых выполняются все неравенства $x_i < a_i$, $i = 1, 2, \ldots, n$ [1]). Если назвать n-мерным полуоткрытым интервалом

$$[a_1, a_2, \ldots, a_n; \quad b_1, b_2, \ldots, b_n)$$

множество всех точек R^n, для которых выполняются все неравенства $a_i \leqslant x_i < b_i$, то видно сразу, что каждый такой интервал также принадлежит телу $\mathfrak{F}^{(x_1, x_2, \ldots, x_n)}$, так как

$$[a_1, a_2, \ldots, a_n; \quad b_1, b_2, \ldots, b_n) =$$
$$= L_{b_1 b_2 \ldots b_n} - L_{a_1 b_2 \ldots b_n} - L_{b_1 a_2 b_3 \ldots b_n} - \ldots - L_{b_1 b_2 \ldots b_{n-1} a_n}.$$

Борелевское расширение системы всех n-мерных полуоткрытых интервалов состоит из всех борелевских множеств R^n. Отсюда следует, что *тело* $\mathfrak{F}^{(x_1, x_2, \ldots, x_n)}$ *в случае борелевского поля вероятностей содержит все борелевские множества пространства* R^n.

ТЕОРЕМА. *В случае борелевского поля вероятностей каждая борелевская функция* $x = f(x_1, x_2, \ldots, x_n)$ *конечного числа случайных величин* $x_1, x_2, \ldots, x_n$ *тоже является случайной величиной.*

Для доказательства достаточно заметить, что множество всех точек $(x_1, x_2, \ldots, x_n)$ из R^n, для которых $x = f(x_1, x_2, \ldots, x_n) < a$, является борелевским. В частности, все конечные суммы и произведения случайных величин тоже являются случайными величинами.

Определение. Функцию

$$F^{(x_1, x_2, \ldots, x_n)}(a_1, a_2, \ldots, a_n) = \mathsf{P}^{(x_1, x_2, \ldots, x_n)}(L_{a_1 a_2 \ldots a_n})$$

мы называем *n-мерной функцией распределения* случайных величин $x_1, x_2, \ldots, x_n$.

---

[1]) a_i могут принимать также бесконечные значения $\pm\infty$.

Доказывается, как в одномерном случае, что n-мерная функция распределения $F^{(x_1, x_2, \ldots, x_n)}(a_1, a_2, \ldots, a_n)$ не убывает по всем переменным и непрерывна слева. Аналогично равенствам (3) и (4), § 2, имеем здесь:

$$\lim_{a_i \to -\infty} F(a_1, a_2, \ldots, a_n) = F(a_1, \ldots, a_{i-1}, -\infty, a_{i+1}, \ldots, a_n) = 0, \quad (7)$$

$$\lim_{a_1 \to +\infty,\, a_2 \to +\infty, \ldots,\, a_n \to +\infty} F(a_1, a_2, \ldots, a_n) = F(+\infty, +\infty, \ldots, +\infty) = 1. \quad (8)$$

Функция распределения $F^{(x_1, x_2, \ldots, x_n)}$ непосредственно дает нам значения $\mathsf{P}^{(x_1, x_2, \ldots, x_n)}$ только для специальных множеств $L_{a_1 a_2 \ldots a_n}$. Если, однако, наше поле вероятностей — борелевское, то $\mathsf{P}^{(x_1, x_2, \ldots, x_n)}$ *однозначно определяется для борелевских множеств через функцию распределения* $F^{(x_1, x_2, \ldots, x_n)}$ [1]).

Если существует производная

$$f(a_1, a_2, \ldots, a_n) = \frac{\partial^n}{\partial a_1\, \partial a_2 \ldots \partial a_n} F^{(x_1, x_2, \ldots, x_n)}(a_1, a_2, \ldots, a_n),$$

то эту производную f называют *n-мерной плотностью вероятности* случайных величин $x_1, x_2, \ldots, x_n$ в точке $(a_1, a_2, \ldots, a_n)$. Если при этом для каждой точки $(a_1, a_2, \ldots, a_n)$

$$F^{(x_1, x_2, \ldots, x_n)}(a_1, a_2, \ldots, a_n) =$$
$$= \int_{-\infty}^{a_1}\int_{-\infty}^{a_2} \ldots \int_{-\infty}^{a_n} f(a_1, a_2, \ldots, a_n)\, da_1\, da_2 \ldots da_n,$$

то распределение $x_1, x_2, \ldots, x_n$ называют *непрерывным*. Для каждого борелевского множества $A \subset R^n$ имеет место равенство:

$$\mathsf{P}^{(x_1, x_2, \ldots, x_n)}(A) = \int\int_A \ldots \int f(a_1, a_2, \ldots, a_n)\, da_1\, da_2 \ldots da_n. \quad (9)$$

В заключение этого параграфа сделаем еще одно замечание о соотношениях между различными вероятностными функциями и функциями распределения. Пусть дана сперва подстановка

$$S = \begin{pmatrix} 1, & 2, & \ldots, & n \\ i_1, & i_2, & \ldots, & i_n \end{pmatrix},$$

и пусть r_S означает преобразование

$$x_k' = x_{i_k} \quad (k = 1, 2, \ldots, n)$$

1) Ср. глава вторая, § 3, IV.

пространства R^n в самого себя. Тогда ясно, что

$$P^{(x_{i_1}, x_{i_2}, \ldots, x_{i_n})}(A) = P^{(x_1, x_2, \ldots, x_n)}\{r_S^{-1}(A)\}. \tag{10}$$

Пусть теперь $x' = p_k(x)$ — „проекция" пространства R^n в пространство R^k $(k < n)$, при которой точка $(x_1, x_2, \ldots, x_n)$ отображается в точку $(x_1, x_2, \ldots, x_k)$. Тогда вследствие формулы (2) § 1

$$P^{(x_1, x_2, \ldots, x_k)}(A) = P^{(x_1, x_2, \ldots, x_n)}\{p_k^{-1}(A)\}. \tag{11}$$

Для соответствующих функций распределения следуют из (10) и (11) равенства:

$$F^{(x_{i_1}, x_{i_2}, \ldots, x_{i_n})}(a_{i_1}, a_{i_2}, \ldots, a_{i_n}) =$$
$$= F^{(x_1, x_2, \ldots, x_n)}(a_1, a_2, \ldots, a_n),$$
$$F^{(x_1, x_2, \ldots, x_k)}(a_1, a_2, \ldots, a_k) =$$
$$= F^{(x_1, x_2, \ldots, x_n)}(a_1, a_2, \ldots, a_k, +\infty, \ldots, +\infty).$$

§ 4. Вероятности в бесконечномерных пространствах

В § 3 главы второй мы видели, как строятся различные применяющиеся в теории вероятностей поля вероятностей. Можно, однако, представить себе интересные также и для приложений проблемы, в которых элементарные события определяются с помощью бесконечного числа координат. Итак, пусть выбрано множество M индексов μ любой мощности $\mathfrak{m}$. Совокупность всех систем

$$\xi = \{x_\mu\}$$

действительных чисел x_μ, причем μ пробегает все множество M, назовем пространством R^M (для определения элемента ξ пространства R^M следует каждому элементу μ множества M поставить в соответствие действительное число x_μ или, что то же самое, задать определенную на M однозначную действительную функцию x_μ элемента μ) [1]).

Если множество M состоит из n первых натуральных чисел $1, 2, \ldots, n$, то R^M есть обычное n-мерное пространство R^n. Если выбрать в качестве множества M множество всех действительных чисел R^1, то соответствующее пространство $R^M = R^{R^1}$ состоит из всех действительных функций

$$\xi(\mu) = x_\mu$$

действительного переменного μ.

[1]) Ср. *Hausdorff*, Mengenlehre, 1927, стр. 23.

Множество R^M при произвольном множестве M мы примем сейчас за основное множество E. Пусть $\xi = \{x_\mu\}$ — элемент E; мы обозначаем тогда через $p_{\mu_1 \mu_2 \dots \mu_n}(\xi)$ точку $(x_{\mu_1}, x_{\mu_2}, \dots, x_{\mu_n})$ n-мерного пространства R^n. Подмножество A из E мы назовем *цилиндрическим множеством*, если оно представимо в форме:

$$A = p^{-1}_{\mu_1 \mu_2 \dots \mu_n}(A'),$$

где A' есть подмножество R^n. Класс всех цилиндрических множеств совпадает, следовательно, с классом всех таких множеств, которые могут быть определены соотношениями вида:

$$f(x_{\mu_1}, x_{\mu_2}, \dots, x_{\mu_n}) = 0. \qquad (1)$$

Для того чтобы произвольное цилиндрическое множество $p^{-1}_{\mu_1 \mu_2 \dots \mu_n}(A')$ определить таким соотношением, достаточно принять за f функцию, которая на A' равна нулю, а вне A' равна единице.

Цилиндрическое множество является борелевским цилиндрическим множеством, если соответствующее множество A' — борелевское. Все борелевские цилиндрические множества пространства R^M образуют тело, которое в дальнейшем будет обозначаться через $\mathfrak{F}^M$ [1]).

Борелевское расширение тела $\mathfrak{F}^M$ мы обозначим, как всегда, через $B\mathfrak{F}^M$. Множества из системы $B\mathfrak{F}^M$ мы называем *борелевскими множествами пространства* R^M.

Далее будет дан метод для построения и оперирования с вероятностными функциями на $\mathfrak{F}^M$ и, следовательно, по теореме о продолжении также и на $B\mathfrak{F}^M$. В случае счетности множества M получаемые таким образом поля вероятностей достаточны для всех целей. Мы овладеваем, следовательно, всеми вопросами, касающимися счетной последовательности случайных величин.

---

[1]) Из вышеприведенных рассуждений следует, что борелевские цилиндрические множества — это те, которые могут быть определены борелевскими соотношениями (1). Пусть теперь A и B — два цилиндрических множества, определенных соотношениями:

$$f(x_{\mu_1}, x_{\mu_2}, \dots, x_{\mu_n}) = 0, \quad g(x_{\lambda_1}, x_{\lambda_2}, \dots, x_{\lambda_m}) = 0.$$

Тогда множества $A + B$, AB и $A - B$ можно определить соответственно следующими соотношениями:

$$f \cdot g = 0, \quad f^2 + g^2 = 0, \; f^2 + \omega(g) = 0,$$

где $\omega(x) = 0$ для $x \neq 0$ и $\omega(0) = 1$. Если f и g — борелевские функции, то таковыми же являются $f \cdot g$, $f^2 + g^2$ и $f^2 + \omega(g)$, следовательно, $A+B$, AB и $A - B$ являются борелевскими цилиндрическими множествами. Этим доказано, что система множеств $\mathfrak{F}^M$ является телом.

Если же M несчетно, то тогда многие простые и интересные подмножества R^M остаются вне системы $B\mathfrak{F}^M$. Например, множество всех элементов ξ, для которых x_μ при любом выборе индекса μ остается меньше какой-либо постоянной величины, не принадлежит к системе $B\mathfrak{F}^M$ в случае несчетности множества M.

Следовательно, всегда стоит добиваться, если это возможно, приведения всякой проблемы к такой форме, при которой пространство всех элементарных событий ξ имеет только счетное множество координат.

Пусть на $\mathfrak{F}^M$ определена вероятностная функция $\mathsf{P}(A)$. Каждую координату x_μ элементарного события ξ можно тогда рассматривать как случайную величину. Следовательно, всякая конечная группа $(x_{\mu_1}, x_{\mu_2}, \ldots, x_{\mu_n})$ этих координат имеет n-мерную вероятностную функцию $\mathsf{P}_{\mu_1 \mu_2 \ldots \mu_n}(A)$ и соответствующую функцию распределения $F_{\mu_1 \mu_2 \ldots \mu_n}(a_1, a_2, \ldots, a_n)$. Ясно, что для всякого борелевского цилиндрического множества

$$A = p^{-1}_{\mu_1 \mu_2 \ldots \mu_n}(A')$$

имеет место равенство:

$$\mathsf{P}(A) = \mathsf{P}_{\mu_1 \mu_2 \ldots \mu_n}(A'),$$

причем A' является борелевским множеством R^n. Таким образом вероятностная функция P однозначно определяется на теле $\mathfrak{F}^M$ всех цилиндрических множеств через значения всех конечномерных вероятностных функций $\mathsf{P}_{\mu_1 \mu_2 \ldots \mu_n}$ для всех борелевских множеств соответствующего пространства R^n. Однако для борелевских множеств значения вероятностной функции $\mathsf{P}_{\mu_1 \mu_2 \ldots \mu_n}$ однозначно определяются через соответствующие функции распределения. Следовательно, мы доказали следующую теорему.

Совокупность всех конечномерных функций распределения $F_{\mu_1 \mu_2 \ldots \mu_n}$ однозначно определяет вероятностную функцию $\mathsf{P}(A)$ для всех множеств из F^M. $\mathsf{P}(A)$ (по теореме о продолжении) определяется однозначно на $B\mathfrak{F}^M$ значениями функций распределения $F_{\mu_1 \mu_2 \ldots \mu_n}$.

Теперь можно поставить вопрос, при каких условиях a priori заданная система функций распределения $F_{\mu_1 \mu_2 \ldots \mu_n}$ определяет поле вероятностей на $\mathfrak{F}^M$ (и, следовательно, на $B\mathfrak{F}^M$).

Сперва отметим, что всякая функция распределения $F_{\mu_1 \mu_2 \ldots \mu_n}$ должна удовлетворять условиям, данным в главе второй, § 3, III, что, конечно, содержится в самом понятии функции распределения. Кроме того, вследствие формул (13) и (14) из § 2 имеют место еще следующие соотношения:

$$F_{\mu_{i_1}\mu_{i_2}\ldots\mu_{i_n}}(a_{i_1}, a_{i_2}, \ldots, a_{i_n}) = F_{\mu_1\mu_2\ldots\mu_n}(a_1, a_2, \ldots, a_n), \tag{2}$$

$$F_{\mu_1\mu_2\ldots\mu_k}(a_1, a_2, \ldots, a_k) = \\ = F_{\mu_1\mu_2\ldots\mu_n}(a_1, a_2, \ldots, a_k, +\infty, \ldots, +\infty), \tag{3}$$

причем $k < n$ и $\begin{pmatrix} 1, & 2, & \ldots, & n \\ i_1, & i_2, & \ldots, & i_n \end{pmatrix}$ — произвольная подстановка. Эти необходимые условия, однако, являются также и достаточными, как это явствует из следующей теоремы.

ОСНОВНАЯ ТЕОРЕМА. *Всякая система функций распределения* $F_{\mu_1\mu_2\ldots\mu_n}$, *удовлетворяющих условиям* (2) и (3), *определяет вероятностную функцию* $\mathsf{P}(A)$ *на* $\mathfrak{F}^M$, *которая удовлетворяет всем аксиомам* I—VI. *Эта вероятностная функция* $\mathsf{P}(A)$ *может быть продолжена* (*по теореме о продолжении*) *также и на* $B\mathfrak{F}^M$.

Доказательство. Итак, пусть даны функции распределения $F_{\mu_1\mu_2\ldots\mu_n}$, удовлетворяющие общим условиям главы второй, § 3, III, и условиям (2) и (3). Всякая функция распределения $F_{\mu_1\mu_2\ldots\mu_n}$ однозначно определяет соответствующую вероятностную функцию $\mathsf{P}_{\mu_1\mu_2\ldots\mu_n}$ для всех борелевских множеств из R^n (ср. § 3). В дальнейшем мы будем рассматривать только борелевские множества R^n и борелевские цилиндрические множества в E.

Для всякого цилиндрического множества

$$A = p^{-1}_{\mu_1\mu_2\ldots\mu_n}(A')$$

полагаем

$$\mathsf{P}(A) = \mathsf{P}_{\mu_1\mu_2\ldots\mu_n}(A'). \tag{4}$$

Так как одно и то же цилиндрическое множество A может быть определено через различные множества A', то нужно сперва доказать, что формула (4) дает всегда одинаковое значение для $\mathsf{P}(A)$.

Пусть $(x_{\mu_1}, x_{\mu_2}, \ldots, x_{\mu_n})$ — конечная система случайных величин x_μ. Исходя из вероятностной функции $\mathsf{P}_{\mu_1\mu_2\ldots\mu_n}$ этих случайных величин, мы можем согласно правилам § 2 определить вероятностную функцию $\mathsf{P}_{\mu_{i_1}\mu_{i_2}\ldots\mu_{i_k}}$ каждой подсистемы $(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \ldots, x_{\mu_{i_k}})$. Равенства (2) и (3) имеют своим следствием, что эта определенная по § 2 вероятностная функция совпадает с a priori заданной функцией $\mathsf{P}_{\mu_{i_1}\mu_{i_2}\ldots\mu_{i_k}}$. Мы теперь предположим, что цилиндрическое множество A опре деляется через

$$A = p^{-1}_{\mu_{i_1}\mu_{i_2}\ldots\mu_{i_k}}(A')$$

и одновременно через

$$A = p^{-1}_{\mu_{j_1}\mu_{j_2}\ldots\mu_{j_m}}(A''),$$

причем все случайные величины x_{μ_i} и x_{μ_j} принадлежат системе $(x_{\mu_1}, x_{\mu_2}, \ldots, x_{\mu_n})$, что, очевидно, не является существенным ограничением. Условия

$$(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \ldots, x_{\mu_{i_k}}) \subset A'$$

и

$$(x_{\mu_{j_1}}, x_{\mu_{j_2}}, \ldots, x_{\mu_{j_m}}) \subset A''$$

равнозначащи. Следовательно,

$$\mathsf{P}_{\mu_{i_1}\mu_{i_2}\ldots\mu_{i_k}}(A') = \mathsf{P}_{\mu_1\mu_2\ldots\mu_n}\{(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \ldots, x_{\mu_{i_k}}) \subset A'\} =$$

$$= \mathsf{P}_{\mu_1\mu_2\ldots\mu_n}\{(x_{\mu_{j_1}}, x_{\mu_{j_2}}, \ldots, x_{\mu_{j_m}}) \subset A''\} = \mathsf{P}_{\mu_{j_1}\mu_{j_2}\ldots\mu_{j_m}}(A''),$$

что доказывает наше утверждение относительно однозначности определения $\mathsf{P}(A)$.

Докажем теперь, что поле вероятностей $(\mathfrak{F}^M, \mathsf{P})$ удовлетворяет всем аксиомам I — VI. Аксиома I утверждает только, что $\mathfrak{F}^M$ должно быть телом; это обстоятельство было уже доказано выше. Далее, при любом μ

$$E = p^{-1}_{\mu}(R^1),$$

$$\mathsf{P}(E) = \mathsf{P}_{\mu}(R^1) = 1,$$

что доказывает применимость аксиом II и IV. Наконец, из определения $\mathsf{P}(A)$ непосредственно следует, что $\mathsf{P}(A)$ неотрицательна (аксиома III).

Несколько сложнее доказывается применимость аксиомы V. Для этой цели рассмотрим два цилиндрических множества:

$$A = p^{-1}_{\mu_{i_1}\mu_{i_2}\ldots\mu_{i_k}}(A')$$

и

$$B = p^{-1}_{\mu_{j_1}\mu_{j_2}\ldots\mu_{j_m}}(B').$$

При этом мы предположим, что все величины x_{μ_i} и x_{μ_j} принадлежат к объемлющей их конечной системе $(x_{\mu_1}, x_{\mu_2}, \ldots, x_{\mu_n})$.

Если множества A и B не пересекаются, то соотношения

$$(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \ldots, x_{\mu_{i_k}}) \subset A',$$

$$(x_{\mu_{j_1}}, x_{\mu_{j_2}}, \ldots, x_{\mu_{j_m}}) \subset B'$$

несовместны. Следовательно,

$$\mathsf{P}(A+B) = \mathsf{P}_{\mu_1 \mu_2 \ldots \mu_n} \{(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \ldots, x_{\mu_{i_k}}) \subset A'$$

$$\text{или } (x_{\mu_{j_1}}, x_{\mu_{j_2}}, \ldots, x_{\mu_{j_m}}) \subset B'\} =$$

$$= \mathsf{P}_{\mu_1 \mu_2 \ldots \mu_n} \{(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \ldots, x_{\mu_{i_k}}) \subset A'\} +$$

$$+ \mathsf{P}_{\mu_1 \mu_2 \ldots \mu_n} \{(x_{\mu_{j_1}}, x_{\mu_{j_2}}, \ldots, x_{\mu_{j_m}}) \subset B'\} = \mathsf{P}(A) + \mathsf{P}(B).$$

Этим наше доказательство закончено.

Остается аксиома VI. Пусть

$$A_1 \supset A_2 \supset \ldots \supset A_n \supset \ldots$$

— убывающая последовательность цилиндрических множеств, удовлетворяющая условию:

$$\lim \mathsf{P}(A_n) = L > 0.$$

Мы докажем, что пересечение всех множеств A_n не пусто. Можно без существенного ограничения постановки вопроса предположить, что в определение n первых цилиндрических множеств A_k входят только n первых координат x_{μ_k} последовательности

$$x_{\mu_1}, x_{\mu_2}, \ldots, x_{\mu_k}, \ldots,$$

т. е. что

$$A_n = p^{-1}_{\mu_1 \mu_2 \ldots \mu_n}(B_n).$$

Мы полагаем для сокращения

$$\mathsf{P}_{\mu_1 \mu_2 \ldots \mu_n}(B) = \mathsf{P}_n(B);$$

тогда, очевидно,

$$\mathsf{P}_n(B_n) = \mathsf{P}(A_n) \geqslant L > 0.$$

В каждом множестве B_n можно найти замкнутое ограниченное множество U_n такое, что

$$\mathsf{P}_n(B_n - U_n) \leqslant \frac{\varepsilon}{2^n}.$$

Из этого неравенства следует для множества

$$V_n = p^{-1}_{\mu_1 \mu_2 \ldots \mu_n}(U_n)$$

неравенство:

$$\mathsf{P}(A_n - V_n) \leqslant \frac{\varepsilon}{2^n}. \tag{5}$$

Пусть далее

$$W_n = V_1 V_2 \ldots V_n.$$

Из (5) следует:

$$\mathsf{P}(A_n - W_n) \leqslant \varepsilon;$$

так как

$$W_n \subset V_n \subset A_n,$$

то

$$\mathsf{P}(W_n) \geqslant \mathsf{P}(A_n) \quad \varepsilon \geqslant L - \varepsilon.$$

Если ε достаточно мало, то $\mathsf{P}(W_n) > 0$, и W_n не пусто. Мы выберем теперь в каждом множестве W_n точку $\xi^{(n)}$ с координатами $x_\mu^{(n)}$. Всякая точка $\xi^{(n+p)}$, $p = 0, 1, 2, \ldots,$ принадлежит множеству V_n, следовательно,

$$\left(x_{\mu_1}^{(n+p)}, x_{\mu_2}^{(n+p)}, \ldots, x_{\mu_n}^{(n+p)}\right) = p^{-1}_{\mu_1 \mu_2 \ldots \mu_n}\left(\xi^{(n+p)}\right) \subset U_n.$$

Так как множества U_n ограничены, можно (диагональный метод) из последовательности $\{\xi^{(n)}\}$ выбрать подпоследовательность

$$\xi^{(n_1)}, \ \xi^{(n_2)}, \ldots, \ \xi^{(n_i)}, \ldots,$$

для которой соответствующие координаты $x_{\mu_k}^{(n_i)}$ стремятся при любом k к определенному пределу x_k. Пусть, наконец, ξ — точка множества E с координатами

$$x_{\mu_k} = x_k,$$

$$x_\mu = 0 \quad (\mu \neq \mu_k, \ k = 1, 2, 3, \ldots).$$

Точка $(x_1, x_2, \ldots, x_k)$ принадлежит как предел последовательности $(x_1^{(n_i)}, x_2^{(n_i)}, \ldots, x_k^{(n_i)})$, $i = 1, 2, 3, \ldots,$ множеству U_k. Следовательно, ξ принадлежит к

$$A_k \subset V_k = p^{-1}_{\mu_1 \mu_2 \ldots \mu_k}(U_k)$$

при любом k, а следовательно, и к пересечению

$$A = \underset{k}{\mathfrak{D}} A_k.$$

§ 5. Эквивалентные случайные величины, разные виды сходимости

С этого параграфа мы рассматриваем исключительно борелевские поля вероятностей. Это не делает, как было уже разъяснено в § 2 главы второй, никакого существенного ограничения для наших исследований.

Две случайные величины x и y называются *эквивалентными*, если вероятность соотношения $x \neq y$ равна нулю. Ясно, что две эквивалентные случайные величины имеют одну и ту же вероятную функцию

$$\mathsf{P}^{(x)}(A) = \mathsf{P}^{(y)}(A).$$

Следовательно, функции распределения $F^{(x)}$ и $F^{(y)}$ также совпадают. Во многих вопросах теории вероятностей можно заменять какую-либо случайную величину любой эквивалентной ей величиной. Пусть теперь

$$x_1, x_2, \ldots, x_n, \ldots \tag{1}$$

—последовательность случайных величин. Рассмотрим множество A всех элементарных событий ξ, для которых последовательность (1) сходится. Если через $A_{np}^{(m)}$ обозначить множество всех ξ, для которых выполняются все неравенства

$$|x_{n+k} - x_n| < \frac{1}{m} \quad (k = 1, 2, \ldots, p),$$

то непосредственно получаем, что

$$A = \underset{m}{\mathfrak{D}}\,\underset{n}{\mathfrak{S}}\,\underset{p}{\mathfrak{D}}\,A_{np}^{(m)}. \tag{2}$$

Согласно § 3 множества $A_{np}^{(m)}$ принадлежат всегда телу множеств $\mathfrak{F}$. Соотношение (2) показывает нам, что множество A также принадлежит $\mathfrak{F}$. *Следовательно, всегда имеет вполне определенный смысл говорить о вероятности сходимости последовательности случайных величин.* Пусть теперь вероятность $\mathsf{P}(A)$ множества A равна единице. Тогда мы утверждаем, что последовательность (1) сходится с вероятностью единица к случайной величине x, при этом случайная величина x определяется однозначно с точностью до эквивалентных величин. Для определения этой случайной величины полагаем

$$x = \lim_{n \to \infty} x_n$$

на A и $x = 0$ вне A. Нам остается доказать, что x — случайная величина, т. е. что множество $A(a)$ элементов ξ, для которых $x < a$, принадлежит телу $\mathfrak{F}$. В самом деле,

$$A(a) = A\underset{n}{\mathfrak{S}}\underset{p}{\mathfrak{D}}\{x_{n+p} < a\}$$

в случае $a \leqslant 0$ и

$$A(a) = A\underset{n}{\mathfrak{S}}\underset{p}{\mathfrak{D}}\{x_{n+p} < a\} + \bar{A}$$

в противоположном случае, откуда непосредственно следует наше утверждение.

Если вероятность сходимости последовательности (1) к x равна единице, то мы говорим, что последовательность (1) *почти наверное* сходится к x. Однако для теории вероятностей, пожалуй, еще важнее другой вид сходимости.

Определение. Последовательность $x_1, x_2, \ldots, x_n, \ldots$ случайных величин *сходится по вероятности* (convèrge en probabilité) к случайной величине x, если при любом $\varepsilon > 0$ вероятность

$$\mathsf{P}\{|x_n - x| > \varepsilon\}$$

стремится к нулю при $n \to \infty$ [1]).

I. *Если последовательность* (1) *сходится по вероятности одновременно к* x *и к* x', *то* x *и* x' *эквивалентны.*

В самом деле,

$$\mathsf{P}\left\{|x' - x| > \frac{1}{m}\right\} \leqslant \mathsf{P}\left\{|x_n - x| > \frac{1}{2m}\right\} + \mathsf{P}\left\{|x_n - x'| > \frac{1}{2m}\right\};$$

так как последние вероятности при достаточно большом n сколько угодно малы, то отсюда следует, что

$$\mathsf{P}\left\{|x - x'| > \frac{1}{m}\right\} = 0,$$

а теперь уже непосредственно получаем, что

$$\mathsf{P}\{x \neq x'\} \leqslant \sum_m \mathsf{P}\left\{|x - x'| > \frac{1}{m}\right\} = 0.$$

II. *Если последовательность* (1) *почти наверное сходится к* x, *то она сходится к* x *также и по вероятности.* Пусть A — множество сходимости последовательности (1); тогда

$$1 = \mathsf{P}(A) \leqslant \lim_{n\to\infty} \mathsf{P}\{|x_{n+p} - x| < \varepsilon,\ p = 0, 1, 2, \ldots\} \leqslant$$
$$\leqslant \lim_{n\to\infty} \mathsf{P}\{|x_n - x| < \varepsilon\},$$

откуда следует сходимость по вероятности.

---

[1]) Это понятие восходит в основном еще к Бернулли, однако в полной общности было введено Е. Е. Слуцким (ср. Slutsky [1]).

III. *Для сходимости по вероятности последовательности* (1) *необходимо и достаточно следующее условие. Для любого* $\varepsilon > 0$ *существует такое* n, *что для каждого* $p > 0$ *имеет место неравенство*:

$$\mathsf{P}\{|x_{n+p} - x_n| > \varepsilon\} < \varepsilon.$$

Пусть теперь $F_1(a)$, $F_2(a), \ldots, F_n(a), \ldots, F(a)$ — функции распределения случайных величин $x_1, x_2, \ldots, x_n, \ldots, x$. Если последовательность x_n сходится по вероятности к x, то функция распределения $F(a)$ однозначно определяется знанием функций $F_n(a)$, так как имеет место

ТЕОРЕМА. *Если последовательность* $x_1, x_2, \ldots, x_n$ *сходится по вероятности к* x, *то последовательность соответствующих функций распределения* $F_n(a)$ *сходится к* $F(a)$, *функции распределения* x, *в каждой точке непрерывности* $F(a)$.

Что $F(a)$ действительно определяется через $F_n(a)$, следует из того, что $F(a)$ как напрерывная слева монотонная функция однозначно определяется ее значениями в точках непрерывности [1]). Для доказательства теоремы мы предположим, что a является точкой непрерывности F. Пусть $a' < a$, тогда в случае $x < a'$, $x_n \geqslant a$ необходимо $|x_n - x| > a - a'$. Следовательно,

$$\lim \mathsf{P}(x < a',\ x_n \geqslant a) = 0,$$

$$F(a') = \mathsf{P}(x < a') \leqslant \mathsf{P}(x_n < a) + \mathsf{P}(x < a',\ x_n \geqslant a) =$$
$$= F_n(a) + \mathsf{P}(x < a',\ x_n \geqslant a),$$

$$F(a') \leqslant \lim\inf F_n(a) + \lim \mathsf{P}(x < a',\ x_n \geqslant a),$$

$$F(a') \leqslant \lim\inf F_n(a). \tag{3}$$

Аналогично доказывается, что из $a'' > a$ следует соотношение:

$$F(a'') \geqslant \lim\sup F_n(a). \tag{4}$$

Так как $F(a')$ и $F(a'')$ при $a' \to a$ и $a'' \to a$ стремятся к $F(a)$, то из (3) и (4) следует, что

$$\lim F_n(a) = F(a),$$

чем наша теорема доказана.

---

[1]) В самом деле она имеет самое большее лишь счетное множество точек разрыва (ср. *Лебег*, Интегрирование и отыскание примитивных функций, 1934, стр. 70). Поэтому точки непрерывности лежат всюду плотно, и значение функции $F(a)$ в точке разрыва определяется как предел ее значений в лежащих слева точках непрерывности.

IV. МАТЕМАТИЧЕСКИЕ ОЖИДАНИЯ [1]

§ 1. Абстрактные интегралы Лебега

Пусть x — случайная величина и A — множество из $\mathfrak{F}$. Образуем для положительного λ сумму

$$S_\lambda = \sum_{k=-\infty}^{+\infty} k\lambda \mathsf{P}\{k\lambda \leqslant x < (k+1)\lambda, \ \xi \subset A\}. \tag{1}$$

Если этот ряд абсолютно сходится при любом λ, то при $\lambda \to 0$ S_λ стремится к определенному пределу, который по определению есть интеграл

$$\int_A x \mathsf{P}(dE). \tag{2}$$

В этой абстрактной форме понятие интеграла было введено Фреше [2]), в частности она необходима для теории вероятностей [читатель, впрочем, увидит в следующих параграфах, что обычное определение условного математического ожидания величины x при гипотезе A совпадает с точностью до постоянного множителя с определением интеграла (2)].

Мы дадим здесь краткий перечень важнейших свойств интегралов формы (2). Читатель найдет их доказательство в каждом учебнике по теории функций действительного переменного, хотя они большей частью проведены для случая, когда $\mathsf{P}(A)$ является лебеговой мерой множеств в R^n. Перенесение этих доказательств на общий случай не представляет никакой новой математической задачи; они остаются большей частью дословно теми же.

I. Если случайная величина x интегрируема на A, то она интегрируема на каждом подмножестве A' из A, принадлежащем $\mathfrak{F}$.

II. Если x интегрируема на A и A распадается на не более чем счетное число непересекающихся множеств A_n из $\mathfrak{F}$, то

$$\int_A x \mathsf{P}(dE) = \sum_n \int_{A_n} x \mathsf{P}(dE).$$

---

[1]) Как было упомянуто в § 5 главы третьей, мы рассматриваем в этой и во всех последующих главах только борелевские поля вероятностей.

[2]) *Fréchet*, Sur l'intégrale d'une fonctionnel étendue à un ensamble abstrait, „Bull. Soc. Math. France", т. 43, 1915, стр. 248.

III. Вместе с x интегрируем всегда и $|x|$, и при этом

$$\left|\int\limits_A x \mathsf{P}(dE)\right| \leqslant \int\limits_A |x| \mathsf{P}(dE).$$

IV. Если в каждом случае ξ выполняются неравенства $0 \leqslant y \leqslant x$, то вместе с x интегрируемо также и y [1]), и при этом

$$\int\limits_A y \mathsf{P}(dE) \leqslant \int\limits_A x \mathsf{P}(dE).$$

V. Если $m \leqslant x \leqslant M$, причем m и M — две постоянные, то

$$m \mathsf{P}(A) \leqslant \int\limits_A x \mathsf{P}(dE) \leqslant M\mathsf{P}(A).$$

VI. Если x и y интегрируемы, а K и L — две действительные постоянные, то $Kx + Ly$ также интегрируемо, и при этом

$$\int\limits_A (Kx + Ly) \mathsf{P}(dE) = K \int\limits_A x \mathsf{P}(dE) + L \int\limits_A y \mathsf{P}(dE).$$

VII. Если ряд

$$\sum_n \int\limits_A |x_n| \mathsf{P}(dE)$$

сходится, то ряд

$$\sum_n x_n = x$$

сходится в каждой точке множества A с точностью до некоторого множества B, такого, что $\mathsf{P}(B) = 0$. Если положить $x = 0$ всюду кроме $A - B$, то

$$\int\limits_A x \mathsf{P}(dE) = \sum_n \int\limits_A x_n \mathsf{P}(dE).$$

VIII. Если x и y эквивалентны ($\mathsf{P}\{x \neq y\} = 0$), то для каждого множества A из $\mathfrak{F}$

$$\int\limits_A x \mathsf{P}(dE) = \int\limits_A y \mathsf{P}(dE). \tag{3}$$

IX. Если (3) имеет место для каждого множества A из $\mathfrak{F}$, то x и y эквивалентны.

Из вышеупомянутого определения интеграла получается еще следующее свойство, которого нет в обычной теории Лебега.

[1]) При этом предполагается, что y — случайная величина, т. е. в терминологии общей теории интегрирования измерима по отношению к $\mathfrak{F}$.

X. Пусть $\mathsf{P}_1(A)$ и $\mathsf{P}_2(A)$ — две вероятностные функции, определенные на одном и том же теле $\mathfrak{F}$, $\mathsf{P}(A) = \mathsf{P}_1(A) + \mathsf{P}_2(A)$ и x интегрируема на A относительно $\mathsf{P}_1(A)$ и $\mathsf{P}_2(A)$. Тогда

$$\int\limits_A x\mathsf{P}(dE) = \int\limits_A x\mathsf{P}_1(dE) + \int\limits_A x\mathsf{P}_2(dE).$$

XI. Всякая ограниченная случайная величина интегрируема.

§ 2. Абсолютные и условные математические ожидания

Пусть x — случайная величина. Интеграл

$$\mathsf{E}(x) = \int\limits_E x\mathsf{P}(dE)$$

называют в теории вероятностей *математическим ожиданием* величины x.

Из свойств III, IV, V, VI, VII, VIII, IX следует, что

I. $|\mathsf{E}(x)| \leqslant \mathsf{E}(|x|)$.

II. $\mathsf{E}(y) \leqslant \mathsf{E}(x)$, если всегда $0 \leqslant y \leqslant x$.

III. $\inf x \leqslant \mathsf{E}(x) \leqslant \sup x$.

IV. $\mathsf{E}(Kx + Ly) = K\mathsf{E}(x) + L\mathsf{E}(y)$.

V. $\mathsf{E}\left(\sum\limits_n x_n\right) = \sum\limits_n \mathsf{E}(x_n)$, если ряд $\sum\limits_n \mathsf{E}(|x_n|)$ сходится.

VI. Если x и y эквивалентны, то

$$\mathsf{E}(x) = \mathsf{E}(y).$$

VII. Всякая ограниченная случайная величина имеет математическое ожидание.

По определению интеграла имеем:

$$\mathsf{E}(x) = \lim \sum_{k=-\infty}^{+\infty} k\lambda \mathsf{P}\{k\lambda \leqslant x < (k+1)\lambda\} =$$

$$= \lim \sum_{k=-\infty}^{+\infty} k\lambda \{F((k+1)\lambda) - F(k\lambda)\}.$$

Вторая строка есть не что иное, как обычное определение интеграла Стильтьеса.

$$\int\limits_{-\infty}^{+\infty} a\, dF^{(x)}(a) = \mathsf{E}(x). \tag{1}$$

Формула (1) может, следовательно, служить определением математического ожидания $\mathsf{E}(x)$.

Пусть теперь u — функция элементарного события ξ, а x — случайная величина, определенная как однозначная функция от u, $x = x(u)$. Тогда

$$\mathsf{P}\{k\lambda \leqslant x < (k+1)\lambda\} = \mathsf{P}^{(u)}\{k\lambda \leqslant x(u) < (k+1)\lambda\},$$

где $\mathsf{P}^{(u)}(A)$ — вероятная функция от u. Отсюда следует по определению интеграла, что

$$\int\limits_E x\mathsf{P}(dE) = \int\limits_{E^{(u)}} x\mathsf{P}^{(u)}(dE^{(u)}),$$

и, следовательно,

$$\mathsf{E}(x) = \int\limits_{E^{(u)}} x\mathsf{P}^{(u)}(dE^{(u)}), \tag{2}$$

причем $E^{(u)}$ означает множество всех возможных значений u.

В частности, когда u само является случайной величиной, то мы имеем:

$$\mathsf{E}(x) = \int\limits_E x\mathsf{P}(dE) = \int\limits_{R^1} x(u)\mathsf{P}^{(u)}(dR^1) =$$
$$= \int\limits_{-\infty}^{+\infty} x(a)\,dF^{(u)}(a). \tag{3}$$

Последний интеграл в формуле (3) является в случае непрерывности функции $x(u)$ обыкновенным интегралом Стильтьеса. Однако отметим при этом, что интеграл

$$\int\limits_{-\infty}^{+\infty} x(a)\,dF^{(u)}(a).$$

может существовать также и в случае отсутствия математического ожидания $\mathsf{E}(x)$. Для существования $\mathsf{E}(x)$ необходима и достаточна конечность интеграла

$$\int\limits_{-\infty}^{+\infty} |x(a)|\,dF^{(u)}(a)\ ^{1)}.$$

Если u — точка $(u_1, u_2, \ldots, u_n)$ пространства R^n, то вследствие (2)

$$\mathsf{E}(x) = \int\int\ldots\int\limits_{R^n} x(u_1, u_2, \ldots, u_n)\mathsf{P}^{(u_1, u_2, \ldots, u_n)}(dR^n). \tag{4}$$

---

[1]) Ср. *V. Glivenko*, Sur les valeurs probables de fonctions, „Rend. Accad. Lincei", т. 8, 1928, стр. 480—483.

Мы уже видели, что условная вероятность $\mathsf{P}_B(A)$ обладает всеми свойствами вероятностной функции. Соответствующий интеграл

$$\mathsf{E}_B(x) = \int_E x \mathsf{P}_B(dE) \tag{5}$$

мы называем *условным математическим ожиданием случайной величины x относительно события B*. Так как

$$\mathsf{P}_B(\overline{B}) = 0,$$

$$\int_{\overline{B}} x \mathsf{P}_B(dE) = 0,$$

то из (5) следует равенство:

$$\mathsf{E}_B(x) = \int_E x \mathsf{P}_B(dE) = \int_B x \mathsf{P}_B(dE) + \int_{\overline{B}} x \mathsf{P}_B(dE) = \int_B x \mathsf{P}_B(dE)$$

Вспомнив, что в случае $A \subset B$

$$\mathsf{P}_B(A) = \frac{\mathsf{P}(AB)}{\mathsf{P}(B)} = \frac{\mathsf{P}(A)}{\mathsf{P}(B)},$$

мы получаем:

$$\mathsf{E}_B(x) = \frac{1}{\mathsf{P}(B)} \int_B x \mathsf{P}(dE), \tag{6}$$

$$\int_B x \mathsf{P}(dE) = \mathsf{P}(B)\, \mathsf{E}_B(x). \tag{7}$$

Из (6) и из равенства

$$\int_{A+B} x \mathsf{P}(dE) = \int_A x \mathsf{P}(dE) + \int_B x \mathsf{P}(dE)$$

следует, наконец:

$$\mathsf{E}_{A+B}(x) = \frac{\mathsf{P}(A)\, \mathsf{E}_A(x) + \mathsf{P}(B)\, \mathsf{E}_B(x)}{\mathsf{P}(A+B)}. \tag{8}$$

В частности, имеет место формула:

$$\mathsf{E}(x) = \mathsf{P}(A)\, \mathsf{E}_A(x) + \mathsf{P}(\overline{A})\, \mathsf{E}_{\overline{A}}(x). \tag{9}$$

§ 3. Неравенство Чебышева

Пусть $f(x)$ — неотрицательная функция действительного аргумента x, которая при $x \geqslant a$ остается не меньше $b > 0$. Тогда для любой случайной величины x

$$\mathsf{P}(x \geqslant a) \leqslant \frac{\mathsf{E}\{f(x)\}}{b}, \tag{1}$$

если только существует математическое ожидание $\mathsf{E}\{f(x)\}$.

В самом деле,

$$\mathsf{E}\{f(x)\} = \int\limits_E f(x)\,\mathsf{P}(dE) \geqslant \int\limits_{\{x \geqslant a\}} f(x)\,\mathsf{P}(dE) \geqslant b\mathsf{P}(x \geqslant a),$$

откуда непосредственно следует (1). Например, для любого положительного c

$$\mathsf{P}(x \geqslant a) \leqslant \frac{\mathsf{E}(e^{cx})}{e^{ca}}. \tag{2}$$

Пусть теперь $f(x)$ — неотрицательная, четная и при положительном x неубывающая функция. Тогда для каждой случайной величины x и при любом выборе постоянной $a > 0$ имеет место неравенство:

$$\mathsf{P}(|x| \geqslant a) \leqslant \frac{\mathsf{E}\{f(x)\}}{f(a)}. \tag{3}$$

В частности,

$$\mathsf{P}(|x - \mathsf{E}(x)| \geqslant a) \leqslant \frac{\mathsf{E}\{f(x - \mathsf{E}(x))\}}{f(a)}. \tag{4}$$

Особо важным является случай $f(x) = x^2$. В этом случае из (3) и (4) получаем:

$$\mathsf{P}(|x| \geqslant a) \leqslant \frac{\mathsf{E}(x^2)}{a^2}, \tag{5}$$

$$\mathsf{P}(|x - \mathsf{E}(x)| \geqslant a) \leqslant \frac{\mathsf{E}\{x - \mathsf{E}(x)\}^2}{a^2} = \frac{\sigma^2(x)}{a^2}. \tag{6}$$

При этом

$$\sigma^2(x) = \mathsf{E}\{x - \mathsf{E}(x)\}^2$$

называют *дисперсией* величины x. Легко подсчитать, что

$$\sigma^2(x) = \mathsf{E}(x^2) - \{\mathsf{E}(x)\}^2.$$

Если $f(x)$ ограничена:

$$|f(x)| \leqslant K,$$

то $\mathsf{P}(|x| \geqslant a)$ можно оценить также снизу. В самом деле,

$$\mathsf{E}\{f(x)\} = \int\limits_E f(x)\,\mathsf{P}(dE) =$$

$$= \int\limits_{\{|x| < a\}} f(x)\,\mathsf{P}(dE) + \int\limits_{\{|x| \geqslant a\}} f(x)\,\mathsf{P}(dE) \leqslant$$

$$\leqslant f(a)\,\mathsf{P}(|x| < a) + K\mathsf{P}(|x| \geqslant a) \leqslant f(a) + K\mathsf{P}(|x| \geqslant a),$$

и, следовательно,

$$\mathsf{P}(|x| \geqslant a) \geqslant \frac{\mathsf{E}\{f(x)\} - f(a)}{K}. \tag{7}$$

Если вместо $f(x)$ случайная величина x сама ограничена:

$$|x| \leqslant M,$$

то $f(x) \leqslant f(M)$, и вместо (7) мы получаем формулу:

$$\mathsf{P}(|x| \geqslant a) \geqslant \frac{\mathsf{E}\{f(x)\} - f(a)}{f(M)}. \tag{8}$$

В случае $f(x) = x^2$ получаем из (8):

$$\mathsf{P}(|x| \geqslant a) \geqslant \frac{\mathsf{E}(x^2) - a^2}{M^2}. \tag{9}$$

§ 4. Некоторые признаки сходимости

Пусть

$$x_1,\ x_2, \ldots,\ x_n, \ldots \tag{1}$$

— последовательность случайных величин, и $f(x)$ — неотрицательная, четная и при положительном x монотонно возрастающая функция [1]. Тогда справедливы следующие теоремы.

I. Для сходимости последовательности (1) по вероятности достаточно следующее условие: для каждого $\varepsilon > 0$ существует такое n, что для каждого $p > 0$ справедливо неравенство:

$$\mathsf{E}\{f(x_{n+p} - x_n)\} < \varepsilon. \tag{2}$$

II. Для сходимости по вероятности последовательности (1) к случайной величине x достаточно условие:

$$\lim_{n \to +\infty} \mathsf{E}\{f(x_n - x)\} = 0. \tag{3}$$

III. Если $f(x)$ ограничена и непрерывна и $f(0) = 0$, то условия I и II являются также и необходимыми.

IV. Если $f(x)$ непрерывна, $f(0) = 0$ и все

$$x_1,\ x_2, \ldots,\ x_n, \ldots,\ x$$

ограничены в своей совокупности, то условия I и II являются также и необходимыми.

Из II и IV следует, в частности:

V. Для сходимости по вероятности последовательности (1) к x достаточным является условие

$$\lim \mathsf{E}(x_n - x)^2 = 0. \tag{4}$$

Если при этом $x_1, x_2, \ldots, x_n, \ldots, x$ ограничены в своей совокупности, то это условие является и необходимым.

Для доказательства I — IV ср. Slutsky [1] и Fréchet [1]. Впрочем, эти теоремы следуют почти непосредственно из формул (3) и (8) предшествующего параграфа.

---

[1] Следовательно, $f(x) > 0$, если $a \neq 0$.

§ 5. Диференцирование и интегрирование математических ожиданий по параметру

Пусть каждому элементарному событию ξ поставлена в соответствие определенная действительная функция $x(t)$ действительного переменного t. Мы говорим, что $x(t)$ есть *случайная функция*, если при каждом фиксированном t величина $x(t)$ является случайной величиной. Возникает тогда вопрос, при каких условиях знак математического ожидания является переместительным со знаками интегрирования и диференцирования. Две следующие теоремы могут, не исчерпывая всей проблемы, дать во многих простых случаях удовлетворительный ответ на этот вопрос.

ТЕОРЕМА I. Если математическое ожидание $\mathsf{E}[x(t)]$ конечно для любого t, $x(t)$ всегда для любого t диференцируема и производная $x'(t)$ от $x(t)$ по t всегда по абсолютному значению меньше некоторой определенной постоянной M, то

$$\frac{d}{dt}\mathsf{E}[x(t)] = \mathsf{E}[x'(t)].$$

ТЕОРЕМА II. Если $x(t)$ по абсолютному значению всегда остается меньше некоторой постоянной K и интегрируема по Риману, то

$$\int_a^b \mathsf{E}[x(t)]\,dt = \mathsf{E}\left[\int_a^b x(t)\,dt\right],$$

если только $\mathsf{E}[x(t)]$ интегрируемо в смысле Римана.

Доказательство теоремы I. Прежде всего заметим, что $x'(t)$ как предел случайных величин

$$\frac{x(t+h)-x(t)}{h}, \quad h = 1, \frac{1}{2}, \ldots, \frac{1}{n}, \ldots$$

является случайной величиной. Так как $x'(t)$ ограничена, то существует математическое ожидание $\mathsf{E}[x'(t)]$ (свойство VII математических ожиданий из § 2). Выберем теперь постоянное t и обозначим через A событие

$$\left|\frac{x(t+h)-x(t)}{h} - x'(t)\right| > \varepsilon.$$

Вероятность $\mathsf{P}(A)$ стремится к нулю при $h \to 0$ при любом $\varepsilon > 0$. Так как всегда

$$\left|\frac{x(t+h)-x(t)}{h}\right| \leqslant M \qquad |x'(t)| \leqslant M$$

и, кроме того, в случае A

$$\left| \frac{x(t+h)-x(t)}{h} - x'(t) \right| \leqslant \varepsilon,$$

то

$$\left| \frac{\mathsf{E}x(t+h)-\mathsf{E}x(t)}{h} - \mathsf{E}x'(t) \right| \leqslant \mathsf{E} \left| \frac{x(t+h)-x(t)}{h} - x'(t) \right| =$$

$$= \mathsf{P}(A)\,\mathsf{E}_A \left| \frac{x(t+h)-x(t)}{h} - x'(t) \right| +$$

$$+ \mathsf{P}(\bar{A})\,\mathsf{E}_{\bar{A}} \left| \frac{x(t+h)-x(t)}{h} - x'(t) \right| \leqslant 2M\mathsf{P}(A) + \varepsilon.$$

$\varepsilon > 0$ можно при этом выбрать произвольно, а $\mathsf{P}(A)$ сколь угодно мало при достаточно малом h. Следовательно,

$$\frac{d}{dt}\mathsf{E}x(t) = \lim_{n \to 0} \frac{\mathsf{E}x(t+h)-\mathsf{E}x(t)}{h} = \mathsf{E}x'(t),$$

что и требовалось доказать.

Доказательство теоремы II. Пусть

$$S_n = \frac{1}{n}\sum_{k=1}^{k=n} x(t+kh), \qquad h = \frac{b-a}{n}.$$

Так как S_n сходится к $J = \int\limits_a^b x(t)\,dt$, то можно для любого $\varepsilon > 0$ выбрать такое N, что из $n \geqslant N$ следует неравенство:

$$\mathsf{P}(A) = \mathsf{P}\{|S_n - J| > \varepsilon\} < \varepsilon.$$

Если положить

$$S_n^* = \frac{1}{n}\sum_{k=1}^{k=n} \mathsf{E}x(t+kh) = \mathsf{E}S_n,$$

то

$$|S_n^* - \mathsf{E}(J)| = |\mathsf{E}(S_n - J)| \leqslant \mathsf{E}|S_n - J| =$$

$$= \mathsf{P}(A)\,\mathsf{E}_A|S_n - J| + \mathsf{P}(\bar{A})\,\mathsf{E}_{\bar{A}}|S_n - J| \leqslant 2K\mathsf{P}(A) +$$

$$+ \varepsilon \leqslant (2K+1)\,\varepsilon.$$

Следовательно, S_n^* стремится к $\mathsf{E}(J)$, откуда следует равенство

$$\int\limits_a^b \mathsf{E}x(t)\,dt = \lim S_n^* = \mathsf{E}(J).$$

Теорема II может быть без всяких новых затруднений обобщена для двойных и тройных интегралов. Мы дадим применение этой теоремы для одной задачи геометрической теории вероятностей. Пусть G — квадрируемая область на плоскости, вид которой зависит от случая, т. е. пусть каждому элементарному событию ξ из поля вероятностей поставлена в соответствие определенная квадрируемая область G на плоскости. Через J обозначим площадь области G, а через $\mathsf{P}(x, y)$ — вероятность того, что точка (x, y) принадлежит области G. Тогда

$$\mathsf{E}(J) = \int\int \mathsf{P}(x, y)\, dx\, dy.$$

Для доказательства достаточно заметить, что

$$J = \int\int f(x, y)\, dx\, dy,$$

$$\mathsf{P}(x, y) = \mathsf{E}f(x, y),$$

где $f(x, y)$ — характеристическая функция области G [$f(x, y) = 1$ на G и $f(x, y) = 0$ вне G [1])].

---

[1]) Ср. *A. Kolmogoroff* und *M. Leontowitsch*, Zur Berechnung der mittleren Brownschen Fläche, „Physik. Zeitschr. d. Sowjetunion“, т. 4, 1933.

V. УСЛОВНЫЕ ВЕРОЯТНОСТИ И МАТЕМАТИЧЕСКИЕ ОЖИДАНИЯ

§ 1. Условные вероятности

В главе первой, § 6, мы определили условную вероятность $\mathsf{P}_{\mathfrak{A}}(B)$ события B относительно испытания $\mathfrak{A}$. При этом в главе первой было предположено, что $\mathfrak{A}$ допускает лишь конечное число различных возможных исходов. Можно, однако, определить $\mathsf{P}_{\mathfrak{A}}(B)$ также и для случая испытания $\mathfrak{A}$ с бесконечным множеством возможных исходов, т. е. для разложения множества E на бесконечное число непересекающихся подмножеств. В частности, такое разложение получается, если рассматривать произвольную функцию u от ξ и определять в качестве элементов разложения $\mathfrak{A}_u$ множества $u = \mathrm{const}$.

Условная вероятность $\mathsf{P}_{\mathfrak{A}_u}(B)$ будет обозначаться также через $\mathsf{P}_u(B)$. Любое разложение $\mathfrak{A}$ множества E можно определить как разложение $\mathfrak{A}_u$, которое „индуцируется“ функцией u от ξ, если каждому ξ поставлено в соответствие в качестве $u(\xi)$ то множество из разложения $\mathfrak{A}$, которое содержит ξ.

Две функции u и u_1 от ξ определяют тогда и только тогда одно и то же разложение $\mathfrak{A}_u = \mathfrak{A}_{u_1}$ множества E, если существует такое взаимнооднозначное соответствие $u_1 = f(u)$ между их значениями, при котором тождественно $u_1(\xi) = f u(\xi)$. Читатель может легко показать, что определяемые ниже случайные величины $\mathsf{P}_u(B)$ и $\mathsf{P}_{u_1}(B)$ в этом случае совпадают; следовательно, в основном они определяются самим разложением $\mathfrak{A}_u = \mathfrak{A}_{u_1}$. Для определения $\mathsf{P}_u(B)$ можно применить следующее равенство:

$$\mathsf{P}_{\{u \subset A\}}(B) = \mathsf{E}_{\{u \subset A\}} \mathsf{P}_u(B). \tag{1}$$

Легко показать, что в случае конечности множества $E^{(u)}$ возможных значений u равенство (1) выполняется при любом выборе множества A (причем $\mathsf{P}_u(B)$ определяется согласно § 6 главы первой). В общем случае (для которого $\mathsf{P}_u(B)$ еще не определена) мы докажем, что всегда существует одна и, с точностью до эквивалентных величин, только одна случайная величина $\mathsf{P}_u(B)$, определяемая как функция от u и удовлетворяющая при каждом A из $\mathfrak{F}^{(u)}$, для которого $\mathsf{P}^{(u)}(A) > 0$, уравнению (1). *Определенную таким образом с точностью до эквивалентности функцию* $\mathsf{P}_u(B)$ *от* u *мы называем условной вероятностью* B *относительно* u (или при данном u). Значение $\mathsf{P}_u(B)$ при $u = a$ мы обозначаем при этом через $\mathsf{P}_u(a; B)$.

Доказательство существования и единственности $\mathsf{P}_u(B)$. Умножив (1) на $\mathsf{P}\{u \subset A\} = \mathsf{P}^{(u)}(A)$, получаем слева:

$$\mathsf{P}\{u \subset A\}\,\mathsf{P}_{\{u \subset A\}}(B) = \mathsf{P}(B\{u \subset A\}) = \mathsf{P}(Bu^{-1}(A)),$$

а справа:

$$\mathsf{P}\{u \subset A\}\mathsf{E}_{\{u \subset A\}}\mathsf{P}_u(B) = \int\limits_{\{u \subset A\}} \mathsf{P}_u(B)\,\mathsf{P}(dE) = \int\limits_A \mathsf{P}_u(B)\,\mathsf{P}^{(u)}(dE^{(u)}$$

получаем, следовательно, формулу:

$$\mathsf{P}(Bu^{-1}(A)) = \int\limits_A \mathsf{P}_u(B)\,\mathsf{P}^{(u)}(dE^{(u)}). \tag{2}$$

Обратно из (2) следует формула (1).

В случае $\mathsf{P}^{(u)}(A) = 0$, при котором (1) не имеет смысла, равенство (2) тривиально. Требование (2), следовательно, эквивалентно (1). По свойству IX интеграла (глава четвертая, § 1) случайная величина x с точностью до эквивалентности однозначно определяется через значения интеграла

$$\int\limits_A x\mathsf{P}(dE)$$

для всех множеств из $\mathfrak{F}$. Так как $\mathsf{P}_u(B)$ — случайная величина, определенная на поле вероятностей $(\mathfrak{F}^{(u)}\,\mathsf{P}^{(u)})$, то отсюда следует, что формула (2) однозначно определяет эту величину $\mathsf{P}_u(B)$ с точностью до эквивалентности.

Нам остается доказать существование $\mathsf{P}_u(B)$. Для этой цели мы применим следующую теорему Никодима [1]):

Пусть $\mathfrak{F}$ — борелевское тело множеств, $\mathsf{P}(A)$ — определенная на $\mathfrak{F}$ неотрицательная вполне аддитивная функция множеств [в теоретико-вероятностной терминологии — случайная величина на $(\mathfrak{F}, \mathsf{P})$], а $Q(A)$ — вторая определенная на $\mathfrak{F}$ вполне аддитивная функция множеств, причем из $Q(A) \neq 0$ следует неравенство $\mathsf{P}(A) > 0$. Тогда существует измеримая по отношению

[1]) *O. Nikodym,* Sur une généralisation des intégrales de M. J. Radon. „Fund. Math.", т. 15, 1930, стр. 168 (théorème III).

к $\mathfrak{F}$ функция $f(\xi)$ (в теоретико-вероятностной терминологии — случайная величина), удовлетворяющая для каждого множества A из $\mathfrak{F}$ равенству:

$$Q(A) = \int\limits_A f(\xi)\,\mathsf{P}(dE).$$

Для применения этой теоремы к нашему случаю остается доказать: 1° что $Q(A) = \mathsf{P}(Bu^{-1}(A))$ вполне аддитивна на $\mathfrak{F}^{(u)}$, 2° что из $Q(A) \neq 0$ следует неравенство: $\mathsf{P}^{(u)}(A) > 0$.

Прежде всего 2° следует из

$$0 \leqslant \mathsf{P}(Bu^{-1}(A)) \leqslant \mathsf{P}(u^{-1}(A)) = \mathsf{P}^{(u)}(A).$$

Для доказательства 1° полагаем

$$A = \sum_n A_n.$$

Тогда

$$u^{-1}(A) = \sum_n u^{-1}(A_n)$$

и

$$Bu^{-1}(A) = \sum_n Bu^{-1}(A_n).$$

Так как P вполне аддитивна, то отсюда следует, что

$$\mathsf{P}(Bu^{-1}(A)) = \sum_n \mathsf{P}(Bu^{-1}(A_n)),$$

что и требовалось доказать.

Из равенства (1) следует, в частности (если положить $A = E^{(u)}$) важная формула:

$$\mathsf{P}(B) = \mathsf{E}(\mathsf{P}_u(B)). \tag{3}$$

Теперь мы перейдем к доказательству двух следующих фундаментальных свойств условных вероятностей:

ТЕОРЕМА I. Почти наверное можно утверждать, что

$$0 \leqslant \mathsf{P}_u(B) \leqslant 1. \tag{4}$$

ТЕОРЕМА II. Если B разложено на суммы множеств B_n так, что в каждой сумме не более счетного числа слагаемых

$$B = \sum_n B_n,$$

то почти наверное имеет место равенство:

$$\mathsf{P}_u(B) = \sum_n \mathsf{P}_u(B_n). \tag{5}$$

Эти два свойства $P_u(B)$ соответствуют двум характеристическим свойствам вероятностной функции $\mathsf{P}(B)$: всегда $0 \leqslant \mathsf{P}(B) \leqslant 1$

и $\mathsf{P}(B)$ вполне аддитивна. Они позволяют перенести на условные вероятности P_u многие дальнейшие существенные свойства абсолютных вероятностей $\mathsf{P}(B)$. Однако при этом не следует забывать, что $\mathsf{P}_u(B)$ при фиксированном множестве B является величиной, определенной лишь с точностью до эквивалентности.

Доказательство теоремы I. Если мы предположим — в противоположность доказываемому утверждению, — что на множестве M с $\mathsf{P}^{(u)}(M)>0$ выполняется неравенство $\mathsf{P}_u(B)\geqslant 1+\varepsilon$, $\varepsilon>0$, то по формуле (1)

$$\mathsf{P}_{\{u\subset M\}}(B)=\mathsf{E}_{\{u\subset M\}}\mathsf{P}_u(B)\geqslant 1+\varepsilon,$$

что, очевидно, невозможно. Так же доказывается, что почти наверное $\mathsf{P}_u(B)\geqslant 0$.

Доказательство теоремы II. Из сходимости ряда

$$\sum_n \mathsf{E}\left|\mathsf{P}_u(B_n)\right|=\sum_n \mathsf{E}(\mathsf{P}_u(B_n))=\sum_n \mathsf{P}(B_n)=\mathsf{P}(B)$$

следует по свойству V математических ожиданий (глава четвертая, § 2), что ряд

$$\sum_n \mathsf{P}_u(B_n)$$

сходится почти наверное. Так как ряд

$$\sum_n \mathsf{E}_{\{u\subset A\}}\left|\mathsf{P}_u(B_n)\right|=\sum_n \mathsf{E}_{\{u\subset A\}}(\mathsf{P}_u(B_n))=$$
$$=\sum_n \mathsf{P}_{\{u\subset A\}}(B_n)=\mathsf{P}_{\{u\subset A\}}(B)$$

сходится при каждом выборе множества A такого, что $\mathsf{P}^{(u)}(A)>0$, то из упомянутого свойства V математических ожиданий следует, что для каждого множества A указанного вида имеет место соотношение:

$$\mathsf{E}_{\{u\subset A\}}\left(\sum \mathsf{P}_u(B_n)\right)=\sum \mathsf{E}_{\{u\subset A\}}\left(\mathsf{P}_u(B_n)\right)=$$
$$=\mathsf{P}_{\{u\subset A\}}(B)=\mathsf{E}_{\{u\subset A\}}\left(\mathsf{P}_u(B_n)\right),$$

а отсюда непосредственно следует равенство (5).

В заключение этого параграфа укажем на два частных случая. В первом $u(\xi)=C$ — постоянная, тогда почти наверное $\mathsf{P}_C(A)=\mathsf{P}(A)$. Если положить, напротив, $u(\xi)=\xi$, то получим сразу, что $\mathsf{P}_\xi(A)$ почти наверное равна единице на A и нулю на $\bar{A}$. $\mathsf{P}_\xi(A)$ является, следовательно, *характеристической функцией* множества A.

§ 2. Объяснение одного парадокса Бореля

Пусть за основное множество E выбрано множество всех точек сферической поверхности. За $\mathfrak{F}$ мы примем совокупность всех борелевских множеств сферических поверхностей. Наконец, пусть $\mathsf{P}(A)$ пропорциональна мере множества A. Выберем теперь две диаметрально противоположные точки в качестве полюсов, тогда каждый меридианный круг однозначно определяется соответствующей географической долготой ψ $(0 \leqslant \psi < \pi)$. Так как ψ изменяется только от 0 до π, т. е. мы рассматриваем полные меридианные круги (а не полуокружности), то и широта θ должна изменяться от $-\pi$ до $+\pi$ $\left(\text{а не от } -\frac{\pi}{2} \text{ до } +\frac{\pi}{2}\right)$. Борелем поставлена следующая задача: определить „условное распределение вероятностей" для широты θ, $-\pi \leqslant \theta < +\pi$ при заданной долготе ψ. Легко подсчитать, что

$$\mathsf{P}_{\psi}(\theta_1 \leqslant \theta < \theta_2) = \frac{1}{4} \int_{\theta_1}^{\theta_2} |\cos \theta| \, d\theta.$$

Распределение вероятностей для θ при заданном ψ неравномерно.

Если предположить, что условное распределение вероятностей для θ „при гипотезе, что ξ лежит на меридианном круге", должно быть равномерным, то получается противоречие.

Это обстоятельство показывает, что понятие условной вероятности относительно изолированно заданной гипотезы, вероятность которой равна нулю, является недопустимым: только тогда мы получим на меридианном круге распределение вероятностей для θ, если будем рассматривать этот меридианный круг в качестве элемента разложения всей сферической поверхности на меридианные круги с заданными полюсами.

ловные вероятности относительно случайной величины

Если x — случайная величина и $\mathsf{P}_x(B)$ как функция x измерима в смысле Бореля, то $\mathsf{P}_x(B)$ можно определить также элементарным путем. В самом деле, формуле (2), § 1, можно придать следующий вид:

$$\mathsf{P}(B)\,\mathsf{P}_B^{(x)}(A) = \int_A \mathsf{P}_x(B)\,\mathsf{P}^{(x)}(dE). \tag{1}$$

В нашем случае из (1) получается непосредственно, что

$$\mathsf{P}(B)\,F_B^{(x)}(a) = \int_{-\infty}^{a} \mathsf{P}_x(a;\,B)\,dF^{(x)}(a). \tag{2}$$

Согласно одной теореме Лебега[1]) из (2) следует, что

$$\mathsf{P}_x(a;\, B) = \mathsf{P}(B) \lim_{h \to 0} \frac{F_B^{(x)}(a+h) - F_B^{(x)}(a)}{F^{(x)}(a+h) - F^{(x)}(a)} \tag{3}$$

с точностью до множества H точек a такого, что $\mathsf{P}^{(x)}(H) = 0$. $\mathsf{P}^{(x)}(a;\, B)$ была определена в § 1 лишь с точностью до множества G такого, что $\mathsf{P}^{(x)}(G) = 0$. Если рассматривать теперь формулу (3) как определение $\mathsf{P}_x(a;\, B)$, причем в случае несуществования предела в правой части (3) положить $\mathsf{P}_x(a;\, B) = 0$, то эта новая величина удовлетворяет всем требованиям § 1.

Если, кроме того, существуют плотности вероятностей $f^{(x)}(a)$ и $f_B^{(x)}(a)$ и если $f^{(x)}(a) > 0$, то формула (3) превращается в следующую:

$$\mathsf{P}_x(a;\, B) = \mathsf{P}(B) \frac{f_B^{(x)}(a)}{f^{(x)}(a)}. \tag{4}$$

Кроме того, из формулы (3) следует, что существование предела (3) и плотности вероятности $f^{(x)}(a)$ имеет своим следствием существование $f_B^{(x)}(a)$. При этом

$$\mathsf{P}(B) f_B^{(x)}(a) \leqslant f^{(x)}(a). \tag{5}$$

Если $\mathsf{P}(B) > 0$, то из (4) следует равенство:

$$f_B^{(x)}(a) = \frac{\mathsf{P}_x(a;\, B)\, f^{(x)}(a)}{\mathsf{P}(B)}. \tag{6}$$

В случае $f^{(x)}(a) = 0$ согласно (5) $f_B^{(x)}(a) = 0$, и, следовательно, (6) также верно. Если при этом распределение x непрерывно, то имеем:

$$\mathsf{P}(B) = \mathsf{E}\mathsf{P}_x(B) = \int_{-\infty}^{+\infty} \mathsf{P}_x(a;\, B)\, dF^{(x)}(a) = \int_{-\infty}^{+\infty} \mathsf{P}_x(a;\, B)\, f^{(x)}(a)\, da. \tag{7}$$

Из (6) и (7) получаем:

$$f_B^{(x)}(a) = \frac{\mathsf{P}_x(a;\, B)\, f^{(x)}(a)}{\int_{-\infty}^{+\infty} \mathsf{P}_x(a;\, B)\, f^{(x)}(a)\, da}. \tag{8}$$

---

[1]) *Лебег*, Интегрирование и отыскание примитивных функций, 1934, стр. 245—246.

Это равенство дает нам так называемую *теорему Байеса для непрерывных распределений.* Предположения, при которых эта теорема доказана, следующие: $\mathsf{P}_x(B)$ измерима в смысле Бореля и в точке a определена по формуле (3), распределение x непрерывно, и в точке a существует плотность вероятности $f^{(x)}(a)$

§ 4. Условные математические ожидания

Пусть u — произвольная функция от ξ, а y — случайная величина. Случайная величина $\mathsf{E}_u(y)$, представимая как функция u и удовлетворяющая для любого множества A из $\mathfrak{F}^{(u)}$ с $\mathsf{P}^{(u)}(A)>0$ условию

$$\mathsf{E}_{\{u\subset A\}}(y)=\mathsf{E}_{\{u\subset A\}}\mathsf{E}_u(y), \tag{1}$$

называется (в случае, если она существует) *условным математическим ожиданием величины y при известном значении u.*

Умножив (1) на $\mathsf{P}^{(u)}(A)$, получаем:

$$\int\limits_{\{u\subset A\}}\mathsf{P}y(dE)=\int\limits_{A}\mathsf{E}_u(y)\mathsf{P}^{(u)}(dE^{(u)}). \tag{2}$$

Обратно из (2) следует формула (1). В случае $\mathsf{P}^{(u)}(A)=0$, при котором (1) не имеет смысла, (2) — тривиально. Так же, как и в случае условных вероятностей (ср. § 1), доказывается, что $\mathsf{E}_u(y)$ однозначно — с точностью до эквивалентности — определяется через (2).

Значение $\mathsf{E}_u(y)$ при $u=a$ мы обозначаем через $\mathsf{E}_u(a;y)$. Заметим еще, что $\mathsf{E}_u(y)$, так же как и $\mathsf{P}_u(y)$, зависит только от разложения $\mathfrak{A}_u$ и может быть обозначено через $\mathsf{E}_{\mathfrak{A}_u}(y)$.

При определении $\mathsf{E}_u(y)$ уже предположено существование $\mathsf{E}(y)$ [если положить $A=E^{(u)}$, тогда $\mathsf{E}_{\{u\subset A\}}(y)=\mathsf{E}(y)$]. Мы теперь докажем, что *существования* $\mathsf{E}(y)$ *также и достаточно для существования* $\mathsf{E}_u(y)$. Для этого достаточно показать, что по теореме Никодима (ср. § 1) функция множеств

$$Q(A)=\int\limits_{\{u\subset A\}}y\mathsf{P}(dE)$$

вполне аддитивна на $\mathfrak{F}^{(u)}$ и абсолютно непрерывна относительно $\mathsf{P}^{(u)}(A)$. Первое обстоятельство доказывается дословно так же, как и в случае условных вероятностей (ср. § 1). Второе требование—абсолютной непрерывности заключается в том, что из $Q(A) \neq 0$ должно следовать неравенство $\mathsf{P}^{(u)}(A) > 0$. Если мы предположим, что $\mathsf{P}^{(x)}(A) = \mathsf{P}\{u \subset A\} = 0$, то ясно, что

$$Q(A) = \int\limits_{\{u \subset A\}} y\mathsf{P}(dE) = 0.$$

Таким образом наше второе требование также выполнено.

Если в равенстве (1) положить $A = E^{(u)}$, то получается формула:

$$\mathsf{E}(y) = \mathsf{E}\mathsf{E}_u(y). \tag{3}$$

Далее доказывается, что почти наверное

$$\mathsf{E}_u(ay + bz) = a\mathsf{E}_u(y) + b\mathsf{E}_u(z), \tag{4}$$

где a и b — две произвольные постоянные. (Проведение доказательств предоставляется читателю.)

Если u и v — две функции элементарного события ξ, то пара (u, v) может также рассматриваться как функция ξ. Здесь тогда имеет место следующее важное равенство:

$$\mathsf{E}_u\mathsf{E}_{(u, v)}(y) = \mathsf{E}_u(y). \tag{5}$$

В самом деле, $\mathsf{E}_u(y)$ определяется через соотношение:

$$\mathsf{E}_{\{u \subset A\}}(y) = \mathsf{E}_{\{u \subset A\}}\mathsf{E}_u(y).$$

Следовательно, нужно доказать, что $\mathsf{E}_u\mathsf{E}_{(u, v)}(y)$ удовлетворяет равенству:

$$\mathsf{E}_{\{u \subset A\}}(y) = \mathsf{E}_{\{u \subset A\}}\mathsf{E}^{\,u}\mathsf{E}_{(u, v)}(y). \tag{6}$$

Из определения $\mathsf{E}_{(u, v)}(y)$ следует:

$$\mathsf{E}_{\{u \subset A\}}(y) = \mathsf{E}_{\{u \subset A\}}\mathsf{E}_{(u, v)}(y). \tag{7}$$

Из определения $\mathsf{E}_u\mathsf{E}_{(u, v)}(y)$ следует далее, что

$$\mathsf{E}_{\{u \subset A\}}\mathsf{E}_{(u, v)}(y) = \mathsf{E}_{\{u \subset A\}}\mathsf{E}_u\mathsf{E}_{(u, v)}(y). \tag{8}$$

Равенство (6) является следствием равенств (7) и (8), чем наше утверждение доказано.

Если положить $y = \mathsf{P}_\xi(B)$ равным единице на B и нулю вне B, то

$$\mathsf{E}_u(y) = \mathsf{P}_u(B),$$

$$\mathsf{E}_{(u, v)}(y) = \mathsf{P}_{(u, v)}(B).$$

В этом случае из формулы (5) получается формула:

$$\mathsf{E}_u \mathsf{P}_{(u,v)}(B) = \mathsf{P}_u(B). \tag{9}$$

Условные математические ожидания можно определить также и непосредственно через соответствующие условные вероятности. Для этой цели рассматриваются следующие суммы:

$$S_\lambda(u) = \sum_{k=-\infty}^{+\infty} k\lambda \mathsf{P}_u\{k\lambda \leqslant y < (k+1)\lambda\} = \sum_k R_k. \tag{10}$$

Если $\mathsf{E}(y)$ существует, то ряд (10) сходится почти наверное. В самом деле, по формуле (3), § 1,

$$\mathsf{E}|R_k| = |k\lambda| \mathsf{P}\{k\lambda \leqslant y < (k+1)\lambda\},$$

а сходимость ряда

$$\sum_{k=-\infty}^{+\infty} |k\lambda| \mathsf{P}\{k\lambda \leqslant y < (k+1)\lambda\} = \sum_k \mathsf{E}|R_k|$$

является необходимым условием существования $\mathsf{E}(y)$ (ср. главу четвертую, § 1). Из этой сходимости следует, что ряд (10) сходится почти наверное (ср. главу четвертую, § 2, V). Далее доказываем, так же как и в теории интеграла Лебега, что из сходимости (10) при каком-либо λ следует сходимость при всяком λ и что в случае сходимости ряда (10) $S_\lambda(u)$ стремится к определенному пределу при $\lambda \to 0$ [1]). Можно тогда в порядке определения положить

$$\mathsf{E}_u(y) = \lim_{\lambda \to 0} S_\lambda(u). \tag{11}$$

Для того чтобы доказать, что определенное соотношением (11) условное ожидание $\mathsf{E}_u(y)$ удовлетворяет поставленным прежде требованиям, нужно только убедиться в том, что определенная согласно (11) величина $\mathsf{E}_u(y)$ удовлетворяет равенству (1). Это доказательство протекает следующим образом:

$$\mathsf{E}_{\{u \subset A\}} \mathsf{E}_u(y) = \lim_{\lambda \to 0} \mathsf{E}_{\{u \subset A\}} S_\lambda(u) =$$

$$= \lim_{\lambda \to 0} \sum_{k=-\infty}^{+\infty} k\lambda \mathbf{P}_{\{u \subset A\}}\{k\lambda \leqslant y < (k+1)\lambda\} = \mathsf{E}_{\{u \subset A\}}(y).$$

---

[1]) При этом мы рассматриваем только счетную последовательность значений λ: тогда все вероятности $\mathbf{P}\{k\lambda \leqslant y < (k+1)\lambda\}$ определены почти наверное для всех этих значений λ.

Перестановка знаков математического ожидания и предела допустима при этой выкладке, так как $S_\lambda(u)$ сходится равномерно к $\mathsf{E}_u(y)$ при $\lambda \Rightarrow 0$ (простое следствие свойства V математических ожиданий из § 2). Перестановка знаков математического ожидания и суммирования также оправдана, так как ряд

$$\sum_{k=-\infty}^{=+\infty} \mathsf{E}_{\{u \subset A\}}\{|k\lambda| \mathsf{P}_u [k\lambda \leqslant y < (k+1)\lambda]\} =$$

$$= \sum_{k=-\infty}^{k=+\infty} |k\lambda| \mathsf{P}_{\{u \subset A\}}[k\lambda \leqslant y < (k+1)\lambda]$$

сходящийся (непосредственное применение свойства V математических ожиданий).

Вместо (11) можно написать:

$$\mathsf{E}_u(y) = \int_E y \mathsf{P}_u(dE); \tag{12}$$

не следует, однако, при этом забывать, что (12) не является интегралом в смысле § 1 главы четвертой, так что (12)—только символическое обозначение.

Если x — случайная величина, то функцию от x и a:

$$F_x^{(y)}(a) = \mathsf{P}_x(y < a)$$

мы называем *условной функцией распределения y при известном x*. $F_x^{(y)}(a)$ определена почти наверное при всяком a. Если $a < b$, то почти наверное

$$F_x^{(y)}(a) \leqslant F_x^{(y)}(b);$$

из (11) и (10) следует, что почти наверное

$$\mathsf{E}_x(y) = \lim_{\lambda \to 0} \sum_{k=-\infty}^{k=+\infty} k\lambda\, [F_x^{(y)}((k+1)\lambda) - F_x^{(y)}(k\lambda)]. \tag{13}$$

Это обстоятельство можно символически выразить формулой:

$$\mathsf{E}_x(y) = \int_{-\infty}^{+\infty} a\, dF_x^{(y)}(a). \tag{14}$$

С помощью нового определения математического ожидания (10)—(11) легко доказать, что для действительной функции от u

$$\mathsf{E}_u[f(u)y] = f(u)\mathsf{E}_u(y).$$

VI. НЕЗАВИСИМОСТЬ. ЗАКОН БОЛЬШИХ ЧИСЕЛ

§ 1. Независимость

Определение 1. Две функции u и v от ξ взаимно независимы, если для любых двух множеств A из $\mathfrak{F}^{(u)}$ и B из $\mathfrak{F}^{(v)}$ справедливо следующее равенство:

$$\mathsf{P}(u \subset A, v \subset B) = \mathsf{P}(u \subset A)\,\mathsf{P}(v \subset B) = \mathsf{P}^{(u)}(A)\,\mathsf{P}^{(v)}(B). \quad (1)$$

Если множества $E^{(u)}$ и $E^{(v)}$ состоят лишь из конечного числа элементов

$$E^{(u)} = u_1 + u_2 + \ldots + u_n,$$
$$E^{(v)} = v_1 + v_2 + \ldots + v_m,$$

то наше определение независимости u и v совпадает с определением независимости разложений

$$E = \sum_k \{u = u_k\},$$
$$E = \sum_k \{v = v_k\}$$

согласно § 5 главы первой.

Для независимости u и v *необходимо* и *достаточно* следующее условие: при любом выборе множества A из $\mathfrak{F}^{(u)}$ почти наверное выполнено равенство:

$$\mathsf{P}_v(u \subset A) = \mathsf{P}(u \subset A). \quad (2)$$

В самом деле, в случае, когда $\mathsf{P}^{(v)}(B) = 0$, оба равенства (1) и (2) удовлетворяются, следовательно, достаточно доказать их эквивалентность в случае $\mathsf{P}^{(v)}(B) > 0$. В этом случае (1) эквивалентно соотношению

$$\mathsf{P}_{\{v \subset B\}}(u \subset A) = \mathsf{P}(u \subset A) \quad (3)$$

и, следовательно, соотношению

$$\mathsf{E}_{\{v \subset B\}}\mathsf{P}_v(u \subset A) = \mathsf{P}(u \subset A). \quad (4)$$

С другой стороны, видно, что из (2) следует равенство (4). Обратно, так как $\mathsf{P}_v(u \subset A)$ однозначно определяется из (4) с точностью до вероятности нуль, то из (4) почти наверное следует равенство (2).

Определение 2. Пусть M — множество функций $u_\mu(\xi)$ от ξ. Эти функции называются взаимно независимыми в своей совокупности, если выполняется следующее условие. Пусть M' и M'' — два непересекающихся подмножества из M, A' (соотв. A'') — множество из $\mathfrak{F}$, определяемое соотношением между u_μ из M' (соотв. M''); тогда выполняется равенство:

$$\mathsf{P}(A'A'') = \mathsf{P}(A')\,\mathsf{P}(A'').$$

Совокупность всех u_μ из M' (соотв. из M'') можно рассматривать как координаты некоторой функции u' (соотв. u''). Определение 2 требует лишь независимости u' и u'' в смысле определения 1 при каждом выборе непересекающихся множеств M' и M''.

Если $u_1, u_2, \ldots, u_n$ взаимно независимы, то всегда

$$\mathsf{P}\{u_1 \subset A_1,\ u_2\, A_2 \subset, \ldots, u_n \subset A_n\} = \\ = \mathsf{P}(u_1 \subset A_1)\,\mathsf{P}(u_2 \subset A_2)\ldots\mathsf{P}(u_n \subset A_n), \qquad (5)$$

если только множества A_k принадлежат соответствующим $\mathfrak{F}^{(u_k)}$ (доказательство посредством индукции). Этого равенства, однако, в общем случае никоим образом не достаточно для взаимной независимости $u_1, u_2, \ldots, u_n$.

Равенство (5) без затруднений обобщается на случай счетного произведения.

Из взаимной независимости u_{μ_k} в каждой конечной группе $(u\tau_\delta \ldots, u_{\mu_k})$, вообще говоря, еще не следует, что все u_μ взаимно независимы.

Наконец, легко заметить, что взаимная независимость функций u_μ, собственно говоря, является свойством соответствующих разложений $\mathfrak{A}_{u_\mu}$. Если далее u'_μ — однозначные функции соответствующих u_μ, то из взаимной независимости u_μ следует то же относительно u'_μ.

§ 2. Независимые случайные величины

Если $x_1, x_2, \ldots, x_n$ — взаимно независимые случайные величины, то из равенства (2) предыдущего параграфа следует, в частности, формула:

$$F^{(x_1, x_2, \ldots, x_n)}(a_1, a_2, \ldots, a_n) = F^{(x_1)}(a_1)\,F^{(x_2)}(a_2)\ldots F^{(x_n)}(a_n). \quad (1)$$

Если при этом тело множеств $\mathfrak{F}^{(x_1, x_2, \ldots, x_n)}$ *состоит только из борелевских множеств пространства* R^n, *то условие* (1) *является также достаточным для взаимной независимости величин* $x_1, x_2, \ldots, x_n$.

Доказательство. Пусть $x' = (x_{i_1}, x_{i_2}, \ldots, x_{i_k})$ и $x'' = (x_{j_1}, x_{j_2}, \ldots, x_{j_m})$ — две непересекающиеся подсистемы величин $x_1, x_2, \ldots, x_n$. На основании формулы (1) следует доказать, что для любых двух борелевских множеств A' и A'' из R^k (соотв. R^m) выполняется равенство:

$$\mathsf{P}(x' \subset A', \; x'' \subset A'') = \mathsf{P}(x' \subset A')\, \mathsf{P}(x'' \subset A''). \tag{2}$$

Это следует непосредственно из (1) для множеств вида:

$$A' = \{x_{i_1} < a_1, \; x_{i_2} < a_2, \ldots, x_{i_k} < a_k\},$$
$$A'' = \{x_{j_1} < b_1, \; x_{j_2} < b_2, \ldots, x_{j_m} < b_m\}.$$

Далее доказывается, что это свойство множеств A' и A'' сохраняется при обозначении сумм и разностей, откуда равенство (2) следует для всех борелевских множеств.

Пусть теперь $x = \{x_\mu\}$ — произвольная (вообще бесконечная) совокупность случайных величин. *Если тело множеств* $\mathfrak{F}^{(x)}$ *совпадает с телом* $B\mathfrak{F}^M$ *(*M *есть множество всех* μ*), то совокупность равенств:*

$$F_{\mu_1 \mu_2 \ldots \mu_n}(a_1, a_2, \ldots, a_n) = F_{\mu_1}(a_1)\, F_{\mu_2}(a_2) \ldots F_{\mu_n}(a_n) \tag{3}$$

необходима и достаточна для взаимной независимости величин x_μ.

Необходимость этого условия следует непосредственно из формулы (1). Мы теперь докажем, что оно также и достаточно. Пусть M' и M'' — два непересекающихся подмножества из множества M всех индексов μ, A' (соотв. A'') — множество из $B\mathfrak{F}^M$, определенное соотношением между x_μ с индексами μ из M' (соотв. из M''); следует доказать, что тогда выполнено равенство:

$$\mathsf{P}(A'A'') = \mathsf{P}(A')\, \mathsf{P}(A''). \tag{4}$$

Действительно, если A' и A'' — цилиндрические множества, то мы имеем дело с соотношениями между конечным числом величин x_μ; равенство (4) представляет в этом случае простое следствие предшествующих результатов [(формула (2)]. А так как соотношение (4) сохраняется для сумм и разностей множеств A' (соотв. A''), то (4) доказано также для всех множеств из $B\mathfrak{F}^M$.

Пусть теперь для каждого μ из множества M задана a priori функция распределения $F_\mu(a)$; *можно тогда построить такое поле вероятностей, что в нем заданные случайные величины x_μ будут взаимно независимы, причем x_μ будет иметь в качестве функции распределения a priori данную функцию $F_\mu(a)$*. Для того чтобы это доказать, достаточно принять R^M за основное множество E и $B\mathfrak{F}^M$ за тело $\mathfrak{F}$, а функции распределения $F_{\mu_1,\mu_2\ldots\mu_n}$ (ср. главу третью, § 4) определить через равенство (3).

Заметим еще, что, как было видно выше, из взаимной независимости всякой конечной группы величин x_μ [равенство (3)] следует взаимная независимость всех x_μ на $B\mathfrak{F}^M$. В объемлющих полях вероятностей это свойство может утеряться.

В заключение этого параграфа дадим еще некоторые признаки независимости для двух случайных величин.

Если две случайные величины x и y взаимно независимы и если $\mathsf{E}(x)$, и $\mathsf{E}(y)$, конечны, то почти наверное

$$\left.\begin{aligned} \mathsf{E}_x(y) &= \mathsf{E}(y), \\ \mathsf{E}_y(x) &= \mathsf{E}(x). \end{aligned}\right\} \tag{5}$$

Эти формулы представляют непосредственное следствие второго определения условных математических ожиданий [формулы (10) и (11) § 4 пятой главы]. Следовательно, в случае независимости обе величины

$$f^2 = \frac{\mathsf{E}[y - \mathsf{E}_x(y)]^2}{\sigma^2(y)}, \quad g^2 = \frac{\mathsf{E}[x - \mathsf{E}_y(x)]^2}{\sigma^2(x)}$$

равны нулю, [если только $\sigma^2(x) > 0$ и $\sigma^2(y) > 0$]. Число f^2 называется *корреляционным отношением* y по x, g^2 — то же для x по y (Пирсон).

Из (5) далее следует, что

$$\mathsf{E}(xy) = \mathsf{E}(x)\,\mathsf{E}(y). \tag{6}$$

Для доказательства применяется формула (15) из § 4 пятой главы

$$\mathsf{E}(xy) = \mathsf{E}\,\mathsf{E}_x(xy) = \mathsf{E}[x\,\mathsf{E}_x(y)] = \mathsf{E}[x\,\mathsf{E}(y)] = \mathsf{E}(x)\,\mathsf{E}(y).$$

Следовательно, в случае независимости

$$r = \frac{\mathsf{E}(xy) - \mathsf{E}(x)\,\mathsf{E}(y)}{\sigma(x)\,\sigma(y)}$$

также равно нулю; r, как известно, — *коэфициент корреляции* между x и y.

Если две случайные величины x и y удовлетворяют равенству (6), то они называются *некоррелированными*. Для суммы

$$S = x_1 + x_2 + \ldots + x_n$$

попарно некоррелированных величин $x_1, x_2, \ldots, x_n$ легко сосчитать, что

$$\sigma^2(S) = \sigma^2(x_1) + \sigma^2(x_2) + \ldots + \sigma^2(x_n). \qquad (7)$$

В частности, равенство (7) справедливо для независимых величин x_k.

§ 3. Закон больших чисел

Случайные величины s последовательности

$$s_1, s_2, \ldots, s_n, \ldots$$

называются *стабильными*, если существует такая числовая последовательность

$$d_1, d_2, \ldots, d_n, \ldots,$$

что для любого положительного ε

$$\mathsf{P}\{|s_n - d_n| \geqslant \varepsilon\}$$

стремится к нулю при $n \to \infty$. Если существуют все $\mathsf{E}(s_n)$ и если можно положить

$$d_n = \mathsf{E}(s_n),$$

то стабильность *нормальная*.

Если все s_n равномерно ограничены, то из

$$\mathsf{P}\{|s_n - d_n| \geqslant \varepsilon\} \Rightarrow 0 \qquad n \Rightarrow \infty \qquad (1)$$

следует соотношение:

$$|\mathsf{E}(s_n) - d_n| \Rightarrow 0 \qquad n \to \infty,$$

и, следовательно,

$$\mathsf{P}\{|s_n - \mathsf{E}(s_n)| \geqslant \varepsilon\} \Rightarrow 0 \qquad n \to \infty. \qquad (2)$$

Стабильность ограниченной стабильной последовательности необходимо нормальна. Пусть

$$\mathsf{E}[s_n - \mathsf{E}(s_n)]^2 = \sigma^2(s_n) = \sigma_n^2.$$

По неравенству Чебышева

$$\mathsf{P}\{|s_n - \mathsf{E}(s_n)| \geqslant \varepsilon\} \leqslant \frac{\sigma_n^2}{\varepsilon^2}.$$

Следовательно, *условие Маркова*

$$\sigma_n^2 \to 0 \tag{3}$$

достаточно для нормальной стабильности.

Если $s_n - \mathsf{E}(s_n)$ равномерно ограниченны:

$$|s_n - \mathsf{E}(s_n)| \leqslant M,$$

то по неравенству (9) из § 3 четвертой главы

$$\mathsf{P}\{|s_n - \mathsf{E}(s_n)| \geqslant \varepsilon\} \geqslant \frac{\sigma_n^2 - \varepsilon^2}{M^2}.$$

Следовательно, в этом случае *условие Маркова* (3) является также и необходимым для стабильности x_n.

Если

$$s_n = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

и величины x_n попарно некоррелированы, то имеем:

$$\sum\nolimits_n^2 \sigma_n^2 = \frac{1}{n^2}\{\sigma^2(x_1) + \sigma^2(x_2) + \ldots + \sigma^2(x_n)\}.$$

Следовательно, в этом случае для стабильности средних арифметических s_n достаточно условие:

$$n^2\sigma_n^2 = \sum\nolimits_n^2 = \sigma^2(x_1) + \sigma^2(x_2) + \ldots + \sigma^2(x_n) = o(n^2) \tag{4}$$

(*теорема Чебышева*). В частности, условие (4) выполнено, если все величины x_n равномерно ограниченны.

Можно обобщить эту теорему на случай слабо коррелированных величин x_n. Если предположить, что коэфициент корреляции r_{mn} [1]) между x_m и x_n удовлетворяет неравенству:

$$r_{mn} \leqslant c(|m-n|)$$

и что

$$C_n = \sum_{k=0}^{n-1} c(k),$$

то для нормальной стабильности средних арифметических достаточно условие [2]):

$$C_n \sum\nolimits_n^2 = o(n^2). \tag{5}$$

---

[1]) Ясно, что всегда $r_{nn} = 1$.

[2]) Ср. *A. Khintchine*, Sur la loi forte des grandes nombres, „C. R. de l'acad. sci. Paris", т. 186, 1928, стр. 285.

В случае независимых слагаемых x_n можно дать также необходимое и достаточное условие для стабильности средних арифметических s_n. Для каждого x_n существует константа m_n (медиана x_n), удовлетворяющая следующим условиям:

$$\mathsf{P}(x_n < m_n) \leqslant \frac{1}{2},$$
$$\mathsf{P}(x_n > m_n) \leqslant \frac{1}{2}.$$

Мы полагаем

$$x_{nk} = \begin{cases} x_k, & \text{если } |x_k - m_k| \leqslant n, \\ 0 & \text{в противном случае,} \end{cases}$$

$$s_n^* = \frac{x_{n1} + x_{n2} + \ldots + x_{nn}}{n}.$$

Тогда соотношения

$$\sum_{k=1}^{n} \mathsf{P}\{|x_k - m_k| > n\} = \sum_{k=1}^{n} \mathsf{P}(x_{nk} \neq x_k) \to 0 \qquad n \to \infty, \tag{6}$$

$$\sum\nolimits_n^{*2} = \sum_{k=1}^{k=n} \sigma^2(x_{nk}) = o(n^2) \tag{7}$$

необходимы и достаточны для стабильности величин s_n[1]).

Можно при этом принять постоянные d_n равными величинам $\mathsf{E}(s_n^*)$, так что в случае

$$\mathsf{E}(s_n^*) - \mathsf{E}(s_n) \to 0 \qquad n \to \infty$$

(и только в этом случае) стабильность *нормальная*.

Дальнейшее обобщение теоремы Чебышева получается, если предположить, что s_n каким-нибудь образом зависят от исходов каких-либо n испытаний

$$\mathfrak{A}_1, \mathfrak{A}_2, \ldots, \mathfrak{A}_n,$$

так что после каждого определенного исхода всех этих n испытаний s_n принимает определенное значение. Общая идея всех теорем, известных под названием *закона больших чисел*, состоит в том, что, если зависимость величины s_n от каждого отдельного испытания $\mathfrak{A}_k$ ($k = 1, 2, \ldots, n$) очень мала при большом n, то величины s_n стабильны. Если рассматривать

$$\beta_{nk}^2 = \mathsf{E}\left[\mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_k}(s_n) - \mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{k-1}}(s_n)\right]^2$$

---

[1]) Ср. *A. Kolmogorov*, Über die Summen durch den Zufall bestimmter unabhängiger Grössen, „Math. Ann.", т. 99, 1928, стр. 309—319, (исправления и замечания к этой работе, т. 102, 1929, стр. 484—488), теорема VIII и добавление к ней на стр. 318.

как разумную меру зависимости величины s_n от испытания $\mathfrak{A}_k$, то вышеупомянутая общая идея закона больших чисел может быть конкретизирована следующими рассуждениями [1]).

Пусть

$$z_{nk} = \mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_k}(s_n) - \mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{k-1}}(s_n).$$

Тогда

$$s_n - \mathsf{E}(s_n) = z_1 + z_2 + \ldots + z_n,$$

$$\mathsf{E}(z_{nk}) = \mathsf{E}\mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_k}(s_n) - \mathsf{E}\mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{k-1}}(s_n) =$$
$$= \mathsf{E}(s_n) - \mathsf{E}(s_n) = 0,$$

$$\sigma^2(z_{nk}) = \mathsf{E}(z_{nk}^2) = \beta_{nk}^2.$$

Легко далее подсчитать, что случайные величины z_{nk} $(k = 1, 2, \ldots, n)$ некоррелированы. В самом деле, пусть $i < k$, тогда [2])

$$\mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{k-1}}(z_{ni} z_{nk}) = z_{ni} \mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{k-1}}(z_{nk}) =$$
$$= z_{ni} \mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{k-1}}[\mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_k}(s_n) - \mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{k-1}}(s_n)] =$$
$$= z_{ni}[\mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{k-1}}(s_n) - \mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{k-1}}(s_n)] = 0,$$

и, следовательно,

$$\mathsf{E}(z_{ni} z_{nk}) = 0.$$

Итак, мы имеем:

$$\sigma^2(s_n) = \sigma^2(z_{n1}) + \sigma^2(z_{n2}) + \ldots + \sigma^2(z_{nn}) = \beta_{n1}^2 + \beta_{n2}^2 + \ldots + \beta_{nn}^2.$$

Следовательно, условие

$$\beta_{n1}^2 + \beta_{n2}^2 + \ldots + \beta_{nn}^2 \to 0 \qquad n \to \infty$$

достаточно для нормальной стабильности величин s_n.

§ 4. Замечания к понятию математического ожидания

Мы определили математическое ожидание величины x как

$$\mathsf{E}(x) = \int_E x \mathsf{P}(dE) = \int_{-\infty}^{+\infty} a\, dF^{(x)}(a).$$

---

[1]) Ср. *A. Kolmogoroff*, Sur la loi des grandes nombres, „Rend. Accad. Lincei", т. 9, 1929, стр. 470—474.

[2]) Применение формулы (15) § 4 пятой главы.

При этом интеграл в правой части понимается в смысле

$$\mathsf{E}(x) = \int_{-\infty}^{+\infty} a\,dF^{(x)}(a) = \lim \int_{b}^{c} a\,dF^{(x)}(a) \quad \begin{matrix} b \to -\infty, \\ c \to +\infty. \end{matrix} \qquad (1)$$

Напрашивается мысль о том, чтобы рассматривать выражение

$$\mathsf{E}^*(x) = \lim_{b \to \infty} \int_{-b}^{+b} a\,dF^{(x)}(a) \qquad (2)$$

в качестве *обобщенного* математического ожидания. При этом, конечно, теряются некоторые простые свойства математических ожиданий. Например, в этом случае формула

$$\mathsf{E}(x+y) = \mathsf{E}(x) + \mathsf{E}(y)$$

не всегда верна. Однако следует заметить, что при некоторых ограничивающих предположениях определение (2) является вполне естественным и пригодным.

Именно можно поставить вопрос следующим образом. Пусть

$$x_1,\ x_2,\ \ldots,\ x_n,\ \ldots$$

— последовательность взаимно независимых величин, имеющих ту же функцию распределения $F^{(x)}(a) = F^{(x_n)}(a)$ $(n = 1, 2, \ldots)$, что и x. Пусть далее

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

Спрашивается, существует ли постоянная $\mathsf{E}^*(x)$ такая, что при каждом $\varepsilon > 0$

$$\lim_{n \to +\infty} \mathsf{P}(|s_n - \mathsf{E}^*(x)| > \varepsilon) = 0. \qquad (3)$$

Ответ гласит: *если такая постоянная* $\mathsf{E}^*(x)$ *существует, то она выражается формулой* (2). Необходимое и достаточное условие для справедливости формулы (3) заключается в существовании предела (2) и соотношении

$$\mathsf{P}(|x| > n) = o\left(\frac{1}{n}\right). \qquad (4)$$

Для доказательства применяется теорема о том, что условие (4) необходимо и достаточно для стабильности средних арифметических s_n, причем в случае стабильности можно положить

$$d_n = \int_{-n}^{+n} a\,dF^{(x)}(a)\ ^{1)}.$$

1) Ср. *A. Kolmogoroff*, Bemerkungen zu meiner Arbeit „Über die Summen zufälliger Grössen", „Mathem. Ann.", т. 102, 1929, стр. 484—488, теорема XII.

Еслисуществует математическое ожидание в прежнем смысле [формула (1)], то условие (4) всегда выполнено [1]).

Так как в этом случае $\mathsf{E}(x) = \mathsf{E}^*(x)$, то условие (3) действительно определяет обобщение понятия математического ожидания. Для *обобщенного математического ожидания* сохраняются свойства I — VII (глава четвертая, § 2), однако в общем случае из существования $\mathsf{E}^*(x)$ не следует существование $\mathsf{E}^*|x|$.

Для того чтобы доказать, что новое понятие математического ожидания является действительно более общим, чем предыдущее достаточно следующего примера. Полагаем плотность вероят ности $f^{(x)}(a)$ равной

$$f^x(a) = \frac{C}{(|a|+2)^2 \ln(|a|+2)},$$

причем постоянная C определяется из условия:

$$\int_{-\infty}^{+\infty} f^{(x)}(a)\, da = 1.$$

Легко далее подсчитать, что в этом случае выполнено условие (4). Формула (2) дает значение

$$\mathsf{E}^*(x) = 0,$$

однако интеграл

$$\int_{-\infty}^{+\infty} |a|\, dF^{(x)}(a) = \int_{-\infty}^{+\infty} |a|\, f^{(x)}(a)\, da$$

расходится.

§ 5. Усиленный закон больших чисел, сходимость рядов

Случайные величины s_n последовательности

$$s_1, s_2, \ldots, s_n, \ldots$$

усиленно стабильны, если существует такая числовая последовательность чисел

$$d_1, d_2, \ldots, d_n, \ldots,$$

что случайные величины

$$s_n - d_n$$

---

1) Ср. *A. Kolmogoroff*, Bemerkungen zu meiner Arbeit „Über die Summen zufälliger Grössen, „Mathem. Ann.", т. 102, 1929, стр. 484—488, теорема XII.

почти наверное стремятся к нулю при $n \to \infty$. Из усиленной стабильности следует, очевидно, обычная стабильность. Если можно выбрать

$$d_n = \mathsf{E}(s_n),$$

то усиленная стабильность *нормальна.*

В случае Чебышева

$$s_n = \frac{x_1 + x_2 + \ldots + x_n}{n},$$

причем величины x_n взаимно независимы. Для усиленной нормальной стабильности средних арифметических s_n достаточна сходимость ряда

$$\sum_{n=1}^{\infty} \frac{\sigma^2(x_n)}{n} \text{ [1]}. \tag{1}$$

Это условие является наилучшим в том смысле, что для любого ряда постоянных b_n таких, что

$$\sum_{n=1}^{\infty} \frac{b_n}{n^2} = +\infty,$$

можно построить такой ряд взаимно независимых случайных величин x_n, что

$$\sigma^2(x_n) = b_n,$$

а соответствующие средние арифметические s_n не будут усиленно стабильны.

Если все x_n имеют одну и ту же функцию распределения $F^{(x)}(a)$, то существование математического ожидания

$$\mathsf{E}(x) = \int_{-\infty}^{+\infty} a\, dF^{(x)}(a)$$

необходимо и достаточно для усиленной стабильности s_n; стабильность в этом случае всегда нормальна [2].

Пусть теперь опять

$$x_1, x_2, \ldots, x_n, \ldots$$

— любые взаимно независимые случайные величины.

---

[1]) Ср. *A. Kolmogoroff,* Sur la loi forte des grandes nombres, „C. R. Acad. Sci. Paris“, т. 191, 1930, стр. 910—911.

[2]) Доказательство этого предложения еще не опубликовано.

Тогда вероятность сходимости ряда

$$\sum_{n=1}^{\infty} x_n \tag{2}$$

равна либо единице, либо нулю. В частности, эта вероятность равна единице, когда оба ряда

$$\sum_{n=1}^{\infty} \mathsf{E}(x_n) \quad \text{и} \quad \sum_{n=1}^{\infty} \sigma^2(x_n)$$

сходятся. Пусть далее положено

$$y_n = \begin{cases} x_n & \text{в случае } |x_n| \leqslant 1, \\ 0 & \text{в случае } |x_n| > 1. \end{cases}$$

Тогда для сходимости ряда (1) с вероятностью единица необходима и достаточна одновременная сходимость рядов[1]):

$$\sum_{n=1}^{\infty} \mathsf{P}\{|x_n| > 1\}, \quad \sum_{n=1}^{\infty} \mathsf{E}(y_n) \quad \text{и} \quad \sum_{n=1}^{\infty} \sigma^2(y_n).$$

---

[1]) Ср. *A. Khintchine* und *A. Kolmogoroff*, Über Konvergenz von Reihen, Математический сборник", т. 32, 1925, стр. 668—677.

ДОПОЛНЕНИЕ

ОДНА ЗАМЕЧАТЕЛЬНАЯ ТЕОРЕМА ТЕОРИИ ВЕРОЯТНОСТЕЙ

Наблюдались уже многие случаи, в которых известные предельные вероятности с необходимостью равны нулю или единице. Например, вероятность сходимости ряда независимых случайных величин может принимать только эти два значения [1]. Мы докажем теперь общую теорему, обнимающую много таких случаев.

ТЕОРЕМА. Пусть $x_1, x_2, \ldots, x_n, \ldots$ — какие-либо случайные величины, а $f(x_1, x_2, \ldots, x_n, \ldots)$ — бэровская функция [2]) переменных $(x_1, x_2, \ldots, x_n, \ldots)$ такая, что условная вероятность

$$\mathsf{P}_{x_1 x_2 \ldots x_n}\{f(x) = 0\}$$

соотношения

$$f(x_1, x_2, \ldots, x_n, \ldots) = 0$$

при известных n первых величинах $x_1, x_2, \ldots, x_n$ остается равной абсолютной вероятности

$$\mathsf{P}\{f(x) = 0\} \tag{1}$$

для каждого n. При этих условиях вероятность (1) равна нулю или единице.

В частности, предпосылки этой теоремы выполняются, если все величины x_n взаимно независимы и значение функции $f(x)$ остается неизменным при изменении лишь конечного числа величин x_n.

Доказательство теоремы. Обозначим через A событие

$$f(x) = 0.$$

---

[1]) Ср. главу шестую, § 5. То же имеет место и для вероятности

$$\mathsf{P}\{s_n - d_n \to 0\}$$

в усиленном законе больших чисел, по крайней мере, когда величины x_n взаимно независимы.

[2]) При этом под бэровской функцией понимается функция, которая может быть представлена, исходя из полиномов, через последовательные итерированные предельные переходы.

Наряду с этим событием мы рассмотрим тело $\mathfrak{R}$ всех событий, которые могут быть определены через какие-либо соотношения между конечным числом величин x_n. Если событие B принадлежит $\mathfrak{R}$, то по условиям теоремы

$$\mathsf{P}_B(A) = \mathsf{P}(A). \tag{2}$$

В случае $\mathsf{P}(A) = 0$ наша теорема уже доказана. Пусть теперь $\mathsf{P}(A) > 0$. Тогда из (2) следует формула:

$$\mathsf{P}_A(B) = \frac{\mathsf{P}_B(A)\,\mathsf{P}(B)}{\mathsf{P}(A)} = \mathsf{P}(B). \tag{3}$$

Итак, $\mathsf{P}(B)$ и $\mathsf{P}_A(B)$ — две вполне аддитивные функции множеств, совпадающие на $\mathfrak{R}$, следовательно, они должны оставаться равными друг другу на каждом множестве борелевского расширения $B\mathfrak{R}$ тела $\mathfrak{R}$. Поэтому в частности

$$\mathsf{P}(A) = \mathsf{P}_A(A) = 1,$$

чем наша теорема доказана.

Некоторые другие случаи, в которых можно утверждать об известных вероятностях, что они принимают только значения нуль и единица, были открыты P. Levy. Ср. по этому поводу *P. Lévy*. Sur un théorème de M. Khintchine („Bull. des Sci. math.“, т. 55, 1931, стр. 145—160, теорема II).

УКАЗАТЕЛЬ ЛИТЕРАТУРЫ

[1] *Бернштейн С. Н.*, Опыт аксиоматического обоснования теории вероятностей, «Зап. Харьк. мат. об-ва», 1917, стр. 209—274.

[2] — «Теория вероятностей», 2-е изд., Москва, ГТТИ, 1934.

[1] *Borel E.*, Les probabilités dénombrables et leurs applications arithmétiques, «Rend. Circ. mat. Palermo», т. 27, 1909, стр. 247—271.

[2] — Principes et formules classiques, fasc. 1 du tome I du Traité des probabilités par E. Borel et divers auteurs. Paris: Cauthier-Villars; 1925.

[3] — Applications à l'arithmétique et à la théorie des fonctions, fasc. 1 du tome II du Traité des probabilités par E. Borel et divers auteurs. Paris: Cauthier - Villars, 1926.

[1] *Cantelli F. P.*, Una teoria astratta del Calcolo delle probabilita, «Giorn. Ist. Ital. Attuari», т. 3, 1932, стр. 257—265.

[2] — Sulla legge dei grandi numeri, «Mem. Acad. Lincei», т. 11, 1916.

[3] — Sulla probabilità come limite della frequenza, «Rend. Accad. Lincei», т. 26, 1917, стр. 39—45.

[1] *Copeland H.*, The theory of probability from the point of view of admissible numbers, «Ann. Math. Statist.», т. 3, 1932, стр. 143—156.

[1] *Dörge K.*, Zu der von R. von Mises gegebenen Begründung der Wahrscheinlichkeitsrechnung, «Math. Z.», т. 32, 1930, стр. 232—258.

[1] *Fréchet M.*, Sur la convergence en probabilité, «Metron», т. 8, 1930, стр. 1—48.

[2] — Recherches théoriques modernes, fasc. 3 du tome I du Traité des probabilités par E. Borel et divers auteurs. Paris: Gauthier — Villars (готовится к печати).

[1] *Kolmogoroff A.*, Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung, «Math. Ann.», т. 104, 1931, стр. 415—458.

[2] *Колмогоров А.*, Общая теория меры и теория вероятностей. Сборник трудов секции точных наук Коммунистической академии, т. I, 1929, стр. 8—21.

[1] *Lévy P.*, Calcul des probabilités, Paris: Gauthier - Villars.

[1] *Lomnicki A.*, Nouveaux fondements du calcul des probabilités, «Fundam. Math.», т. 4, 1923, стр. 34—71.

[1] *Mises, R.* v., Wahrscheinlichkeitsrechnung, Leipzig u. Wien: Fr. Deuticke, 1931.

[2] — Grundlagen der Wahrscheinlichkeitsrechnung, «Math. Z.», т. 5, 1919, стр. 52—99.

[3] — Wahrscheinlichkeitsrechnung, Statistik und Wahrheit. Wien: Julius Springer, 1928.

[1] *Reichenbach H.*, Axiomatik der Wahrscheinlichkeitsrechnung, «Math. Z.», т. 34, 1932, стр. 568—619.

[1] *Slutsky E.*, Über stochastische Asymptoten und Grenzwerte, «Metron», т. 5, 1925, стр. 3—89.

[2] *Слуцкий Е.*, К вопросу о логических основаниях теории вероятности, «Вестник статистики». т. 12, 1922, стр. 13—21.

[1] *Steinhaus H.*, Les probabilités dénombrables et leur rapport à la théorie de la mesure, «Fundam. Math.», т. 4, 1923, стр. 286—310.

[1] *Tornier E.*, Wahrscheinlichkeitsrechnung und Zahlentheorie, «J. reine angew. Math.», т. 160, 1929, стр. 177—198.

[2] — Grundlagen der Wahrscheinlichkeitsrechnung, «Acta math.», т. 60, 1933, стр. 239—380.

Опечатка

Стр.	Строка	Напечатано	Следует читать
50	9 сверху	$+\int_{B} x\mathsf{P}_B\,(dE)$	$+\int_{\overline{B}} x\mathsf{P}_B\,(dE)$

Колмогоров

Опечатки

Стр.	Строка	Напечатано	Следует читать
54	1 сверху	A	$\overline{A}$
54	9 „	$\lim\limits_{n\to 0}$	$\lim\limits_{h\to 0}$

Колмогоров

Опечатки

Стр.	Строка	Напечатано	Следует читать
62	16 св.	$\int_{\{u \subset A\}} \mathsf{P}y\,(dE)$	$\int_{\{u \subset A\}} y\mathsf{P}\,(dE)$
63	3 „	$\mathsf{P}^{(x)}(A)$	$\mathsf{P}^{(u)}(A)$
63	11 снизу	$\mathsf{E}_{\{u \subset A\}}\,\mathsf{E}^u\mathsf{E}_{(u,v)}(y)$	$\mathsf{E}_{\{u \subset A\}}\,\mathsf{E}_u\mathsf{E}_{(u,v)}(y)$.

Колмогоров

Опечатки

Стр.	Строка	Напечатано	Следует читать
67	18 св.	$u_2A_2 \subset$	$u_2 \subset A_2$
67	12 сн.	$(u\tau_\delta, \ldots, u_{\mu_k})$	$(u_{\mu_1}, u_{\mu_2}, \ldots, u_{\mu_k})$