В. В. Горяйнов

Лекции по теории вероятностей

Содержание

§ 1. Пространство элементарных событий. Вероятность

Теория вероятностей представляет математические модели, которые отражают свойства случайных явлений. Как и в любой другой теории (например, классической механике, геометрии и др.), основные понятия теории вероятностей отражают некоторые свойства реального мира, но являются идеальными объектами. Для иллюстрации этого будем исходить из понятия случайного эксперимента, результат которого нельзя заранее предсказать, но сам эксперимент можно провести в одних и тех же условиях любое количество раз. Примеры: подбрасывание монеты, подбрасывание игральной кости. В результате эксперимента мы можем наблюдать те или иные события, которые будем обозначать $A, B, C, \ldots$. Если проводить эксперимент в одинаковых условиях большое количество n раз, то частота n_A/n (n_A — число появления A) появление события A будет стремиться к некоторой величине. Это позволяет предположить, что событие A обладает некоторой скрытой характеристикой, которую называют вероятностью и обозначают $\mathsf{P}(A)$.

1.1. Теоретико–множественная модель. Построение теоретико-множественной модели теории вероятностей связано с выделением пространства Ω *элементарных событий* ω (*исходов*). В результате каждого проведения случайного эксперимента может наблюдаться только одно элементарное событие ω. При этом любое событие A ассоциируется с некоторым подмножеством Ω, т. е. $A \subset \Omega$. Само Ω также рассматривается как событие. Его называют *достоверным* событием, поскольку оно происходит всегда. Наряду с этим пустое множество $\varnothing$ связывают с *невозможным* событием.

Таким образом, результат эксперимента описывается элементарным событием ω. Если $\omega \in A$, то говорят, что событие A произошло, если же $\omega \notin A$, то событие A в этом эксперименте не произошло. Теоретико множественные операции и отношения соответствуют логическим операциям над событиями. Если $A \subset B$, то это означает, что наступление события A влечет наступление события B. Событие $\overline{A} = \Omega \setminus A$ называется отрицанием события A. Объединению $A \cup B$ соответствует событие, которое происходит в случае, когда произошло A или B. Пересечению $A \cap B$ соответствует событие, которое происходит лишь в случае, когда происходят одновременно и A и B. Аналогичный смысл имеют объединение и пересечение любого числа событий. Разность $A \setminus B$ означает событие, когда A произошло, но B в этом эксперименте не произошло.

Заметим, что операции объединения и пересечения множеств обладают свойством дистрибутивности

$$A \bigcap \left(\bigcup_{i \in I} B_i \right) = \bigcup_{i \in I} (AB_i), \qquad A \bigcup \left(\bigcap_{i \in I} B_i \right) = \bigcap_{i \in I} \left(A \bigcup B_i \right).$$

Следующие соотношения известны как правила де Моргана:

$$\overline{\bigcup_{i \in I} A_i} = \bigcap_{i \in I} \overline{A_i}, \qquad \overline{\bigcap_{i \in I} A_i} = \bigcup_{i \in I} \overline{A_i}.$$

Если Ω конечно или счетно, то говорят о дискретной модели теории вероятностей (или случайного эксперимента). Пусть $\Omega = \{\omega_1, \omega_2, \ldots\}$ конечно или счетно и допустим, что каждому элементарному событию ω_k соответствует вероятность p_k. Тогда $p_k \geqslant 0$ и $\sum p_k = 1$. Кроме того, для любого события $A \subset \Omega$ вероятность должна определяться как сумма

$$\mathsf{P}(A) = \sum_{k:\, \omega_k \in A} p_k.$$

В дискретных моделях особенно выделяется случай *классического определения* вероятности. Допустим, что Ω состоит из конечного числа элементарных событий $\omega_1, \ldots, \omega_n$, которые в силу каких либо причин (например, симметрии) являются «равновозможными». Такая ситуация возникает при подбрасывании монеты или игральной кости. Тогда $p_k = 1/n$ для всех $k = 1, \ldots, n$, а подсчет вероятности события $A \subset \Omega$ сводится к вычислению числа $|A|$ элементарных событий, составляющих A. В частности, $|\Omega| = n$, а для $\mathsf{P}(A)$ в этих терминах можно записать формулу

$$\mathsf{P}(A) = \frac{|A|}{|\Omega|}.$$

Эта формула и составляет классическое определение вероятности.

1.2. Элементы комбинаторики. При использовании классического определения вероятности основным инструментом являются комбинаторные методы. Отметим некоторые из них.

Основное правило комбинаторики. Пусть имеется m групп элементов, причем i-тая группа состоит из k_i элементов. Выберем по одному элементу из каждой группы. Тогда общее число N способов, которыми можно произвести такой выбор, определяется равенством

$$N = k_1 \cdot k_2 \cdot \ldots \cdot k_m.$$

Доказательство легко проводится по индукции.

Выборки. Многие комбинаторные вычисления укладываются в следующую схему. Допустим, что мы имеем занумерованных шаров: $a_1, \ldots, a_n$. Осуществляется выборка объема k из этой совокупности шаров. Нас интересует количество способов, которыми можно осуществить эту выборку. При этом возникает четыре различные ситуации: выборка может быть упорядоченной или неупорядоченной и с возвращением или без возвращения.

♣ Количество упорядоченных выборок с возвращением (или размещений из n по k с повторениями) равно n^k. Это сразу же следует из основного правила комбинаторики.

♣ Количество упорядоченных выборок без возвращения (или размещений из n по k без повторений) равно

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

Эта формула также следует из основного правила комбинаторики.

♣ Количество неупорядоченных выборок без возвращения (или сочетаний из n по k без повторений) равно

$$C_n^k = \frac{n!}{k!(n-k)!} = \frac{A_n^k}{k!}.$$

♣ Количество неупорядоченных выборок с возвращением (или сочетаний из n по k с повторениями) равно C_{n+k-1}^k.

Для подсчета числа таких выборок воспользуемся следующей конструкцией. Каждую такую выборку однозначно мы можем представить упорядоченной цепочкой нулей и единиц следующим образом. Вначале напишем столько единиц, сколько в выборке присутствует a_1, после чего поставим нуль; затем напишем столько единиц, сколько в выборке присутствует a_2, после чего снова поставим нуль и так далее. Таким образом, каждой выборке будет соответствовать последовательность из k единиц и $n-1$ нулей. Чтобы зафиксировать выборку, достаточно указать места расположения единиц. Это эквивалентно осуществлению неупорядоченной выборке из $n+k-1$ элементов объема k, т. е. их количество равно C_{n+k-1}^k.

Полученные результаты можно представить в виде таблицы.

выборки	с возвращением	без возвращения
упорядоченные	n^k	$A_n^k = \dfrac{n!}{(n-k)!}$
неупорядоченные	$C_{n+k-1}^k = \dfrac{(n+k-1)!}{k!(n-1)!}$	$C_n^k = \dfrac{n!}{k!(n-k)!}$

Размещение частиц по ячейкам. В физических приложениях приведенная выше комбинаторная схема встречается в другой интерпретации. Размещается k частиц по n ячейкам. В статистической физике в качестве частиц могут быть электроны, протоны, а в качестве ячеек, например, энергетические уровни. Здесь также выделяется четыре случая: различимые или неразличимые частицы и размещение с ограничением (не более одной частицы в одной ячейке) или без ограничений. Размещению без ограничений различимых частиц соответствует статистика Максвелла— Больцмана, размещению без ограничений неразличимых частиц соответствует статистика Бозе— Эйнштейна, а размещению с ограничениями неразличимых частиц соответствует статистика Ферми— Дирака. Случай размещений с ограничениями различимых частиц в статистической физике не используется.

Размещение частиц по ячейкам легко сводится к выборкам. Будем считать, что ячейки занумерованы: $a_1, \ldots, a_n$. Тогда под размещением можно понимать назначение каждой частице номера ячейки. В результате приходим к следующему результату.

♠ Количество размещений без ограничений различимых частиц (статистика Максвелла— Больцмана) равно n^k.

♠ Количество размещений без ограничений неразличимых частиц (статистика Бозе— Эйнштейна) равно C_{n+k-1}^k.

♠ Количество размещений с ограничением неразличимых частиц (статистика Ферми— Дирака) равно C_n^k.

♠ Количество размещений с ограничениями различимых частиц равно A_n^k.

1.3. Геометрические вероятности. Рассмотрим теперь задачу, в которой пространство элементарных событий не является даже счетным. Допустим, что два лица X и Y условились встретиться между 14 и 15 часами в определенном месте. Пришедший первым ждет другого в течение 10 минут, а затем уходит. Какова вероятность их встречи, если приход каждого лица в любой момент времени из условленного промежутка равновероятен?

Выберем в качестве единицы измерения один час и пусть x — промежуток времени (в долях часа) от 14 часов до момента прихода X, а y — промежуток времени от 14 часов до момента прихода Y. Тогда пространством элементарных событий будет квадрат

$$\Omega = \{(x, y) \colon\ 0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 1\},$$

а событие A, которое отвечает встрече, представляет собой подмножество Ω, выделяемое условием: $|x - y| \leqslant 1/6$.

В этой задаче нельзя построить вероятностную модель, просто приписав вероятность каждому элементарному событию. С другой стороны, логично считать, что вероятность события A равняется отношению ее площади к площади всего Ω. Поскольку площадь Ω равна 1, а площадь A равна

$$1-\left(1-\frac{1}{6}\right)^2 = \frac{11}{36},$$

то $\mathsf{P}(A)=11/36$.

В общем случае, когда пространство элементарных событий Ω представляет собой область в $\mathbb{R}^n$, $n=1,2,\ldots$, и все элементарные события считаются равновероятными, вероятность события $A\subset\Omega$ будем вычислять по формуле

$$\mathsf{P}(A) = \frac{\mathrm{mes}(A)}{\mathrm{mes}(\Omega)},$$

где под $\mathrm{mes}(A)$ понимается мера Жордана множества A, хотя в дальнейшем мы выясним, что для более сложных множеств, чем области в $\mathbb{R}^n$, следует рассматривать меру Лебега.

§ 2. Алгебры событий и свойства вероятности

В случае, когда пространство элементарных событий Ω не является конечным или счетным, нет удовлетворительного способа определить вероятность на всех его подмножествах. Поэтому возникает необходимость выяснить свойства класса подмножеств Ω, которые будут отвечать событиям. Как мы видели, теоретико множественные операции соответствуют логическим операциям над событиями. Поэтому такой класс подмножеств Ω должен быть замкнут относительно теоретико множественных операций.

Определение 2.1. Класс $\mathscr{A}$ подмножеств Ω называется алгеброй, если $\Omega\in\mathscr{A}$ и он замкнут относительно операций объединения, пересечения и взятия дополнения.

Следующее утверждение оказывается полезным при проверке, является ли выделенный класс подмножеств Ω алгеброй.

Предложение 2.1. *Пусть $\mathscr{A}$ — класс подмножеств Ω, для которого выполняются следующие условия:*

(i) $\Omega\in\mathscr{A}$;
(ii) Если $A\in\mathscr{A}$, то $\overline{A}\in\mathscr{A}$;
(iii) Если $A,B\in\mathscr{A}$, то $A\cup B\in\mathscr{A}$.

Тогда $\mathscr{A}$ является алгеброй подмножеств Ω.

Доказательство. Нам нужно показать, что $\mathscr{A}$ замкнуто относительно всех теоретико множественных операций. Из условий (ii) и (iii) следует, что класс $\mathscr{A}$ замкнут относительно операции взятия дополнения и объединения. Остается показать, что класс $\mathscr{A}$ также замкнут относительно операции пересечения. Пусть A и B из $\mathscr{A}$. Тогда в силу условия (ii) дополнения $\overline{A}$, $\overline{B}$ принадлежат $\mathscr{A}$. Представление

$$AB = \overline{\overline{A}\cup\overline{B}}$$

и условие (iii) доказывают принадлежность пересечения AB классу $\mathscr{A}$. □

Приведем один важный пример получения алгебры событий (множеств).

Определение 2.2. Совокупность событий $\mathscr{D} = \{D_1, \dots, D_n\}$ будем называть *разбиением*, если $D_iD_j = \varnothing$ при $i \neq j$ и $\cup_{i=1}^n D_i = \Omega$. Множества D_i называются в этом случае атомами разбиения $\mathscr{D}$.

Пусть $\mathscr{D} = \{D_1, \dots, D_n\}$ — некоторое разбиение. Через $\mathscr{A}(\mathscr{D})$ обозначим класс событий, который включает $\varnothing$ и всевозможные объединения атомов разбиения. Очевидно, что $\mathscr{A}(\mathscr{D})$ представляет собой алгебру, которую мы будем называть *алгеброй, порожденной разбиением* $\mathscr{D}$.

Анализируя свойства частоты и примеры, приходим к выводу, что вероятность P есть функция, определенная на некоторой алгебре $\mathscr{A}$ событий, которая обладает следующими свойствами:

1°. $\mathsf{P}(A) \geqslant 0$ для всех A из $\mathscr{A}$ (неотрицательность);
2°. $\mathsf{P}(\Omega) = 1$ (нормированность);
3°. Если A и B — два события из $\mathscr{A}$ и $AB = \varnothing$, то $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B)$ (аддитивность).

Эти простые три свойства вероятности приводят к более сложным свойствам.

Следствие 2.1. *Пусть $\mathscr{A}$ — алгебра подмножеств Ω и $\mathsf{P}\colon \mathscr{A} \to \mathbb{R}$ — функция, удовлетворяющая свойствам $1° - 3°$. Тогда имеют место утверждения:*

1. *Для любого $A \in \mathscr{A}$ имеет место равенство $\mathsf{P}(\overline{A}) = 1 - \mathsf{P}(A)$ и $\mathsf{P}(\varnothing) = 0$;*
2. *Если $A, B \in \mathscr{A}$ и $A \subset B$, то $\mathsf{P}(B \setminus A) = \mathsf{P}(B) - \mathsf{P}(A)$;*
3. *Если $A, B \in \mathscr{A}$ и $A \subset B$ то $\mathsf{P}(A) \leqslant \mathsf{P}(B)$;*
4. *Для любого $A \in \mathscr{A}$ имеют место неравенства $0 \leqslant \mathsf{P}(A) \leqslant 1$;*
5. *Для любых $A, B \in \mathscr{A}$ имеет место равенство*

$$\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(AB). \tag{2.1}$$

Доказательство. Пусть $A \in \mathscr{A}$. Из равенства $\Omega = A \cup \overline{A}$ и в силу свойств $2°, 3°$ имеем $1 = \mathsf{P}(A) + \mathsf{P}(\overline{A})$, что и доказывает первое утверждение. В частности, $\mathsf{P}(\varnothing) = \mathsf{P}(\overline{\Omega}) = 0$.

Допустим теперь, что $A, B \in \mathscr{A}$ и $A \subset B$. Тогда $B = A \cup (B \setminus A)$ и снова в силу свойства 3° получаем $\mathsf{P}(B) = \mathsf{P}(A) + \mathsf{P}(B \setminus A)$. Отсюда следует второе утверждение.

Третье утверждение следует из равенства $\mathsf{P}(B) = \mathsf{P}(A) + \mathsf{P}(B \setminus A)$ и неотрицательности вероятности.

Непосредственно из третьего утверждения, которое составляет свойство монотонности вероятности, следует

$$0 = \mathsf{P}(\varnothing) \leqslant \mathsf{P}(A) \leqslant \mathsf{P}(\Omega) = 1.$$

Пусть теперь A, B — произвольные из алгебры $\mathscr{A}$. Заметим, что $A \cup B$ можно представить в виде объединения трех непересекающихся множеств

$$A \cup B = (A \setminus (AB)) \cup (B \setminus (AB)) \cup (AB).$$

Но тогда по свойству аддитивности вероятности с использованием второго утверждения получаем

$$\mathsf{P}(A \cup B) = \mathsf{P}(A) - \mathsf{P}(AB) + \mathsf{P}(B) - \mathsf{P}(AB) + \mathsf{P}(AB) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(AB),$$

и равенство (2.1) доказано. □

Равенство (2.1) допускает обобщение на случай произвольного числа событий.

Теорема 2.1. *[Сложения]. Пусть $A_1, \dots, A_n$ — произвольные события, т. е. принадлежат алгебре $\mathscr{A}$. Тогда имеет место равенство*

$$\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathsf{P}(A_i) - \sum_{1 \leqslant i_1 < i_2 \leqslant n} \mathsf{P}(A_{i_1} A_{i_2}) + \ldots + (-1)^{n-1} \mathsf{P}(A_1 \cdot \ldots \cdot A_n).$$

Доказательство. В случае $n = 2$ доказываемое утверждение совпадает с равенством (2.1). Допустим теперь, что наше утверждение верно для всех натуральных чисел вплоть до $n-1$ и покажем, что оно тогда будет верным и для n. Используя вначале равенство (2.1), а затем предположение индукции, получаем

$$\mathsf{P}\left(A_n \cup \left(\bigcup_{i=1}^{n-1} A_i\right)\right) = \mathsf{P}(A_n) + \mathsf{P}\left(\bigcup_{i=1}^{n-1} A_i\right) - \mathsf{P}\left(\bigcup_{i=1}^{n-1} A_n A_i\right)$$
$$= \mathsf{P}(A_n) + \left[\sum_{i=1}^{n-1} \mathsf{P}(A_i) - \sum_{1 \leqslant i_1 < i_2 \leqslant n-1} \mathsf{P}(A_{i_1} A_{i_2}) + \ldots + (-1)^{n-2} \mathsf{P}(A_1 \cdot \ldots \cdot A_{n-1})\right]$$
$$- \left[\sum_{i=1}^{n-1} \mathsf{P}(A_n A_i) - \sum_{1 \leqslant i_1 < i_2 \leqslant n-1} \mathsf{P}(A_n A_{i_1} A_{i_2}) + \ldots + (-1)^{n-2} \mathsf{P}(A_1 \cdot \ldots \cdot A_n)\right]$$
$$= \sum_{i=1}^{n} \mathsf{P}(A_i) - \sum_{1 \leqslant i_1 < i_2 \leqslant n} \mathsf{P}(A_{i_1} A_{i_2}) + \ldots + (-1)^{n-1} \mathsf{P}(A_1 \cdot \ldots \cdot A_n).$$

Теорема доказана. □

Следующий результат выражает свойство полуаддитивности вероятности для произвольного набора событий.

Предложение 2.2. *Пусть $A_1, \dots, A_n$ — произвольные события. Тогда*

$$\mathsf{P}\left(\bigcup_{k=1}^{n} A_k\right) \leqslant \sum_{k=1}^{n} \mathsf{P}(A_k).$$

Доказательство. Рассмотрим события $B_1 = A_1$, $B_2 = A_2 \overline{A_1}, \dots, B_n = A_n \overline{A_1} \cdot \ldots \cdot \overline{A_{n-1}}$. Очевидно, что $B_i B_j = \varnothing$ при $i \neq j$ и $\cup_{i=1}^{n} A_i = \cup_{i=1}^{n} B_i$. Кроме того, поскольку $B_i \subset A_i$, то в силу свойства монотонности вероятности $\mathsf{P}(B_i) \leqslant \mathsf{P}(A_i)$, $i = 1, \dots, n$. Используя эти неравенства и свойство аддитивности, получаем

$$\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \mathsf{P}\left(\bigcup_{i=1}^{n} B_i\right) = \sum_{i=1}^{n} \mathsf{P}(B_i) \leqslant \sum_{i=1}^{n} \mathsf{P}(A_i).$$

□

В дальнейшем под элементарной вероятностной моделью мы будем понимать тройку $(\Omega, \mathscr{A}, \mathsf{P})$, где Ω — некоторое множество (пространство элементарных событий), $\mathscr{A}$ — алгебра подмножеств Ω и P — функция (вероятность), определенная на $\mathscr{A}$ и удовлетворяющая условиям 1° – 3°. Коротко тройку $(\Omega, \mathscr{A}, \mathsf{P})$ будем называть *вероятностное пространство*. Вскоре мы несколько расширим понятия вероятностной модели и вероятностного пространства.

§ 3. Условная вероятность и независимость

Допустим, что мы не знаем точный результат (элементарное событие) случайного эксперимента, но знаем, что произошло событие A. Эта информация должна повлиять на переоценку вероятностей всех других событий. Принимая во внимание то, что вероятность события проявляется через частоту его появления, допустим, что при N испытаниях события A, B и AB наблюдались в N_A, N_B и N_{AB} числе случаев, соответственно. Тогда отношение N_{AB}/N_A будет выражать частоту появления события B при условии, что произошло событие A. Поскольку

$$\frac{N_{AB}}{N_A} = \frac{N_{AB}/N}{N_A/N},$$

то эта частота будет отражением величины $\mathsf{P}(AB)/\mathsf{P}(A)$.

Определение 3.1. Пусть $(\Omega, \mathscr{A}, \mathsf{P})$ — вероятностное пространство и событие A имеет положительную вероятность, $\mathsf{P}(A) \neq 0$. Тогда для каждого B из $\mathscr{A}$ под *условной вероятностью* этого события при условии A будем понимать величину

$$\mathsf{P}(B|A) = \frac{\mathsf{P}(AB)}{\mathsf{P}(A)}. \tag{3.1}$$

Условная вероятность $\mathsf{P}(B|A) = \mathsf{P}_A(B)$, как функция на $\mathscr{A}$, обладает всеми свойствами вероятности. Таким образом, информация о том, что произошло событие A, позволяет перейти к новому вероятностному пространству $(\Omega, \mathscr{A}, \mathsf{P}_A)$.

На практике чаще легче бывает вычислить условную вероятность $\mathsf{P}(B|A)$ и тогда, используя равенство (3.1), получить вероятность совместного наступления этих событий

$$\mathsf{P}(AB) = \mathsf{P}(A)\mathsf{P}(B|A). \tag{3.2}$$

Это равенство известно как теорема умножения и имеет следующее обобщение.

Теорема 3.1. *Пусть события $A_1, \ldots, A_n$ таковы, что* $\mathsf{P}(A_1 \cdot \ldots \cdot A_n) > 0$. *Тогда*

$$\mathsf{P}(A_1 \cdot \ldots \cdot A_n) = \mathsf{P}(A_1) \cdot \mathsf{P}(A_2|A_1) \cdot \ldots \cdot \mathsf{P}(A_n|A_1 \cdot \ldots \cdot A_{n-1}). \tag{3.3}$$

Доказательство. Заметим прежде всего, что все условные вероятности в правой части равенства (3.3) корректно определены, поскольку $\mathsf{P}(A_1 \cdot \ldots \cdot A_n) > 0$ и $A_1 \cdot \ldots \cdot A_n$ является подмножеством каждого условия. При $n = 2$ равенство (3.3) совпадает с равенством (3.2). Допустим теперь, что равенство (3.3) выполняется для всех номеров, вплоть до $n-1$. Обозначим $A = A_1, \dots, A_{n-1}$, $B = A_n$. Тогда, используя равенство(3.2) и предположение индукции, получаем

$$\begin{aligned} \mathsf{P}(A_1 \cdot \ldots \cdot A_n) &= \mathsf{P}(A_1 \cdot \ldots \cdot A_{n-1})\mathsf{P}(A_n|A_1 \cdot \ldots \cdot A_{n-1}) \\ &= \mathsf{P}(A_1) \cdot \mathsf{P}(A_2|A_1) \cdot \ldots \cdot \mathsf{P}(A_n|A_1 \cdot \ldots \cdot A_{n-1}). \end{aligned}$$

□

Следующий результат известен как «*формула полной вероятности*».

Теорема 3.2. *Пусть $H_1, \dots, H_n$ — события, удовлетворяющие следующим условиям $H_iH_j = \varnothing$ при $i \neq j$, $\mathsf{P}(H_i) > 0$, $i = 1, \dots, n$, и $\sum_{i=1}^n \mathsf{P}(H_i) = 1$. Тогда для любого события A имеет место равенство*

$$\mathsf{P}(A) = \sum_{i=1}^n \mathsf{P}(H_i)\mathsf{P}(A|H_i). \tag{3.4}$$

Доказательство. Введем в рассмотрение также событие $H_0 = \Omega \setminus \cup_{i=1}^n H_i$. Очевидно, что $\mathsf{P}(H_0) = 0$ и $\mathscr{D} = \{H_0, H_1, \dots, H_n\}$ является разбиением. Поскольку

$$A = A\Omega = AH_0 \cup AH_1 \cup \ldots \cup AH_n$$

представляет собой объединение попарно непересекающихся множеств (попарно несовместных событий) и $\mathsf{P}(AH_0) = 0$, то

$$\mathsf{P}(A) = \sum_{i=1}^n \mathsf{P}(AH_i).$$

Применяя теперь к каждому слагаемому в последней сумме теорему умножения $\mathsf{P}(AH_i) = \mathsf{P}(H_i)\mathsf{P}(A|H_i)$, приходим к формуле (3.4). □

Формула (3.4) является эффективным инструментом при вычислении вероятностей сложных событий. При этом события $H_1, \dots, H_n$ в этой формуле обычно называют *гипотезами*.

Теорема 3.3. *[Формула Байеса] Пусть события $H_1, \dots, H_n$ удовлетворяют условиям теоремы 3.2 и $A \in \mathscr{A}$, $\mathsf{P}(A) > 0$. Тогда для $k = 1, \dots, n$ имеют место равенства*

$$\mathsf{P}(H_k|A) = \frac{\mathsf{P}(H_k)\mathsf{P}(A|H_k)}{\sum\limits_{i=1}^n \mathsf{P}(H_i)\mathsf{P}(A|H_i)}.$$

Доказательство. Непосредственно из определения условной вероятности и теоремы умножения получаем

$$\mathsf{P}(H_k) = \frac{\mathsf{P}(H_kA)}{\mathsf{P}(A)} = \frac{\mathsf{P}(H_k)\mathsf{P}(A|H_k)}{\mathsf{P}(A)}.$$

Остается знаменатель записать по формуле полной вероятности (3.4) и мы приходим к требуемому утверждению. □

Вероятности гипотез $\mathsf{P}(H_k)$ называют обычно *априорными* вероятностями, а условные вероятности $\mathsf{P}(H_k|A)$ — *апостериорными.* Таким образом, формула Байеса позволяет «переоценить» априорную вероятность гипотезы при наличии информации, что произошло событие A.

Независимость. Понятие независимости относится к одному из основных в теории вероятностей. С информационной точки зрения под независимостью события B от события A естественно понимать то, что знание о наступлении события A не влияет на вероятность наступления события B. Это выражается равенством $\mathsf{P}(B|A) = \mathsf{P}(B)$. При таком определении мы должны предполагать, что $\mathsf{P}(A) > 0$. Если при этом и $\mathsf{P}(B) > 0$, то из независимости B от A следует

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

Это означает, что тогда и событие A не зависит от события B, т.е. свойство независимости двух событий является симметричным. Кроме того, независимость событий A и B в названном выше смысле сразу же следует из равенства

$$\mathsf{P}(AB) = \mathsf{P}(A)\mathsf{P}(B). \tag{3.5}$$

Определение 3.2. Для произвольных двух событий $A, B \in \mathscr{A}$ будем говорить, что они независимы, если выполняется равенство (3.5).

Заметим, что в отличие от «информационного» подхода к определению независимости мы здесь не требуем условия положительности вероятностей событий A и B.

Предложение 3.1. *Если события A и B независимы, то независимы также пары событий A и $\overline{B}$, $\overline{A}$ и B, $\overline{A}$ и $\overline{B}$.*

Доказательство. Утверждение достаточно доказать для пары A и $\overline{B}$. Используя равенство (3.5), получаем

$$\mathsf{P}(A\overline{B}) = \mathsf{P}(A \setminus AB) = \mathsf{P}(A) - \mathsf{P}(AB) = \mathsf{P}(A)(1 - \mathsf{P}(B)) = \mathsf{P}(A)\mathsf{P}(\overline{B}).$$

□

Определение 3.3. Будем говорить, что события $A_1, \dots, A_n$ независимы в совокупности, если для любых $k = 2, \dots, n$ и $1 \leqslant i_1 < i_2 < \dots < i_k \leqslant n$ выполняются соотношения

$$\mathsf{P}(A_{i_1} \cdot \ldots \cdot A_{i_k}) = \mathsf{P}(A_{i_1}) \cdot \ldots \cdot \mathsf{P}(A_{i_k}). \tag{3.6}$$

Заметим, что из попарной независимости событий не следует, вообще говоря, независимость в совокупности. Кроме того, при проверке независимости в совокупности не достаточно ограничиться проверкой лишь самых длинных цепочек в равенстве (3.6). Сказанное подтверждается следующими двумя примерами.

Пример 1. Пусть $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, где все элементарные события ω_i равновероятны. Определим $A = \{\omega_1, \omega_2\}$, $B = \{\omega_1, \omega_3\}$, $C = \{\omega_1, \omega_4\}$. Эти события попарно независимы, но не выполняется равенство

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

Пример 2. Подбрасывается две игральные кости. Рассмотрим три события:

$$A = \{\text{на первой кости выпала «единица», «двойка» или «пятерка»}\},$$

$$B = \{\text{на первой кости выпала «четверка», «пятерка» или «шестерка»}\},$$

$$C = \{\text{сумма выпавших очков на двух игральных костях равна девяти}\}.$$

В этом случае выполняется равенство

$$\mathsf{P}(ABC) = \mathsf{P}(A)\mathsf{P}(B)\mathsf{P}(C),$$

но нет даже попарной независимости.

Определение 3.4. Классы событий $\mathscr{F}_1, \ldots, \mathscr{F}_n$ называются независимыми, если для любого набора $A_1 \in \mathscr{F}_1, \ldots, A_n \in \mathscr{F}_n$ события $A_1, \ldots, A_n$ независимы в совокупности.

Заметим, что условие независимости алгебр $\mathscr{F}_1, \ldots, \mathscr{F}_n$ можно сформулировать просто: для любых $A_1 \in \mathscr{F}_1, \ldots, A_n \in \mathscr{F}_n$ должно выполняться равенство

$$\mathsf{P}(A_1 \cdot \ldots \cdot A_n) = \mathsf{P}(A_1) \cdot \ldots \cdot \mathsf{P}(A_n).$$

В действительности, здесь содержатся все цепочки из формулы (3.6), поскольку выбор в качестве одного или нескольких событий $A_1, \ldots, A_n$ достоверного события Ω приводит к более короткой цепочке.

Теорема 3.4. *Пусть $\mathscr{D}_1 = \{D_1^{(1)}, \ldots, D_{k_1}^{(1)}\}, \ldots, \mathscr{D}_n = \{D_1^{(n)}, \ldots, D_{k_n}^{(n)}\}$ — разбиения пространства Ω и $\mathscr{A}(\mathscr{D}_1), \ldots, \mathscr{A}(\mathscr{D}_n)$ — порожденные ими алгебры событий. Тогда $\mathscr{A}(\mathscr{D}_1), \ldots, \mathscr{A}(\mathscr{D}_n)$ независимы, если для любых $i_1, \ldots, i_n$, $1 \leqslant i_1 \leqslant k_1$, $\ldots$, $1 \leqslant i_n \leqslant k_n$, выполняются равенства*

$$\mathsf{P}(D_{i_1}^{(1)} \cdot \ldots \cdot D_{i_n}^{(n)}) = \mathsf{P}(D_{i_1}^{(1)}) \cdot \ldots \cdot \mathsf{P}(D_{i_n}^{(n)}). \tag{3.7}$$

Доказательство. Нам нужно доказать, что если $A_1 \in \mathscr{A}(\mathscr{D}_1), \ldots, A_n \in \mathscr{A}(\mathscr{D}_n)$, то выполняется равенство

$$\mathsf{P}(A_1 A_2 \ldots A_n) = \mathsf{P}(A_1)\mathsf{P}(A_2)\ldots\mathsf{P}(A_n).$$

Согласно условию теоремы, это равенство выполняется в случае, когда в качестве $A_1, \ldots, A_n$ выбраны атомы разбиений. Поскольку любое событие алгебры в нашем случае представляет собой объединение атомов соответствующего разбиения, а сами атомы одного разбиения взаимно не пересекаются, то для доказательства теоремы достаточно установить, что при выполнении равенств

$$\mathsf{P}(AH_1 \ldots H_m) = \mathsf{P}(A)\mathsf{P}(H_1)\ldots\mathsf{P}(H_m), \ \mathsf{P}(BH_1 \ldots H_m) = \mathsf{P}(B)\mathsf{P}(H_1)\ldots\mathsf{P}(H_m)$$

для некоторых $H_1, \ldots, H_m$ и двух непересекающихся A, B следует

$$\mathsf{P}((A \cup B)H_1 \ldots H_m) = \mathsf{P}(A \cup B)\mathsf{P}(H_1)\ldots\mathsf{P}(H_m).$$

Последнее очевидно, поскольку при этих условиях

$$\begin{aligned} \mathsf{P}((A \cup B)H_1 \ldots H_m) &= \mathsf{P}(AH_1 \ldots H_m) + \mathsf{P}(BH_1 \ldots H_m) \\ &= \mathsf{P}(A)\mathsf{P}(H_1)\ldots\mathsf{P}(H_m) + \mathsf{P}(B)\mathsf{P}(H_1)\ldots\mathsf{P}(H_m) \\ &= \mathsf{P}(A \cup B)\mathsf{P}(H_1)\ldots\mathsf{P}(H_m) \end{aligned}$$

и теорема доказана. □

§ 4. Независимые испытания. Схема Бернулли

Допустим, что проводится серия из n независимых экспериментов (испытаний), в каждом из которых с вероятностью p, $0 < p < 1$, может наступить и с вероятностью $q = 1 - p$ может не наступить некоторое событие A. Появление события A обычно называют «успехом», а не появление — «неудачей». Основной вопрос в схеме Бернулли состоит в вычислении вероятности события $B_n(k)$, которое заключается в том, что в проведенной серии испытаний произошло ровно k успехов, $k = 0, 1, \ldots, n$.

Для решения поставленной задачи построим вероятностное пространство $(\Omega, \mathscr{A}, \mathsf{P})$, соответствующее схеме Бернулли. Каждый результат серии экспериментов можно представить n-мерной точкой $\omega = (\delta_1, \ldots, \delta_n)$, где δ_i принимает значение 1 в случае успеха в i-том испытании и значение 0 в случае неуспеха. В силу независимости испытаний вероятность элементарного события $\omega = (\delta_1, \ldots, \delta_n)$ будет определяться равенством

$$\mathsf{P}(\{\omega\}) = p^{\sum_{i=1}^n \delta_i} q^{n - \sum_{i=1}^n \delta_i}. \tag{4.1}$$

Если $\omega \in B_n(k)$, то из (4.1) следует, что $\mathsf{P}(\{\omega\}) = p^k q^{n-k}$. Замечая теперь, что $B_n(k)$ состоит из C_n^k элементарных событий, приходим к формуле

$$\mathsf{P}(B_n(k)) = C_n^k p^k q^{n-k}. \tag{4.2}$$

События $B_n(k)$, $k = 0, 1, \ldots, n$, образуют разбиение. Это подтверждается равенством

$$\sum_{k=0}^n \mathsf{P}(B_n(k)) = \sum_{k=0}^n C_n^k p^k q^{n-k} = (p+q)^n = 1.$$

Формула (4.2) выражает распределение вероятностей числа успехов в n независимых испытаниях. Это распределение называют *биномиальным.*

При больших значениях n реализация формулы (4.2) сопряжена с трудоемкими вычислениями. Поэтому ее пытаются заменить приближенными формулами. Следующий результат относится к случаю, когда p мало, а n велико. В связи с малостью p этот результат иногда называют законом редких событий.

Теорема 4.1. *[Пуассона.] Если $n \to \infty$ и $p \to 0$ так, что $np \to \lambda$, $0 < \lambda < \infty$, то при всех $k = 0, 1, 2, \ldots$ выполняется соотношение*

$$\mathsf{P}(B_n(k)) = C_n^k p^k q^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda}.$$

Доказательство. Из преобразований

$$\begin{aligned}\frac{n!}{k!(n-k)!} p^k q^{n-k} &= \frac{1}{k!} \frac{(n-1) \cdot \ldots \cdot (n-k+1)}{n^{k-1}} n^k p^k (1-p)^{n-k} \\ &= \frac{(np)^k}{k!} \left(1 - \frac{1}{n}\right) \cdot \ldots \cdot \left(1 - \frac{k-1}{n}\right) (1-p)^{-k} (1-p)^{\frac{1}{p} np}\end{aligned}$$

и предельного соотношения

$$\lim_{p \to 0} (1-p)^{1/p} = e^{-1}$$

следует утверждение теоремы. □

Заметим, что в условиях теоремы Пуассона 4.1 имеет место оценка

$$\left|\mathsf{P}(B_n(k)) - \frac{\lambda^k}{k!}e^{-\lambda}\right| \leqslant \frac{\lambda^2}{n}, \qquad \lambda = np,$$

т. е. результат применим, когда np^2 мало.

Доказанная теорема относится к так называемым предельным теоремам в схеме Бернулли. Распределение вероятностей

$$p_k = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots,$$

к которому стремится в условиях теоремы биномиальное распределение, называют *пуассоновским распределением.* В случае, когда вероятность успеха фиксирована, а число экспериментов стремится к бесконечности, используют предельные теоремы Муавра— Лапласа.

Теорема 4.2. *[Локальная теорема Муавра-Лапласа.] Пусть в схеме Бернулли $0 < p < 1$ фиксировано, $\sigma = \sqrt{npq}$, $x = x(k) = (k - np)/\sigma$, $k = 0, 1, 2, \ldots$. Тогда для любого $M > 0$ равномерно по всем k таким, что $|x(k)| \leqslant M$, при $n \to \infty$ имеет место асимптотическое равенство*

$$\mathsf{P}(B_n(k)) = \frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2}(1 + o(1)).$$

Доказательство. В основе доказательства теоремы лежит формула Стирлинга

$$n! = \sqrt{2\pi n}\cdot n^n \cdot e^{-n}\cdot e^{\frac{\theta}{12n}}, \qquad 0 < \theta < 1.$$

Из этой формулы следует, что

$$\ln n! = \ln\sqrt{2\pi n} + n\ln n - n + O\left(\frac{1}{n}\right).$$

Пусть k таково, что $|x(k)| \leqslant M$. Тогда из равенства

$$k = np + x\sigma = np\left(1 + \frac{xq}{\sigma}\right)$$

следует, что $k \to \infty$ при $n \to \infty$ $(\sigma \to \infty)$. Аналогично, из равенства

$$n - k = nq - x\sigma = nq\left(1 - \frac{xp}{\sigma}\right)$$

видно, что $(n - k) \to \infty$ при $n \to \infty$. Следовательно, к $(n - k)!$ и $k!$ также можно применить формулу Стирлинга. Кроме того,

$$\frac{1}{n} = O\left(\frac{1}{\sigma^2}\right), \quad \frac{1}{k} = O\left(\frac{1}{\sigma^2}\right), \quad \frac{1}{n-k} = O\left(\frac{1}{\sigma^2}\right).$$

Таким образом,

$$\begin{aligned}\ln\mathsf{P}(B_n(k)) &= \ln n! - \ln k! - \ln(n-k)! + k\ln p + (n-k)\ln q \\ &= \frac{1}{2}\ln\frac{n}{2\pi k(n-k)} + n\ln n - k\ln k - (n-k)\ln(n-k) \\ &+ k\ln p + (n-k)\ln q + O\left(\frac{1}{\sigma^2}\right) \\ &= \frac{1}{2}\left(\ln\frac{1}{2\pi} + \ln\frac{n}{k(n-k)}\right) - k\ln\frac{k}{np} - (n-k)\ln\frac{n-k}{nq} + O\left(\frac{1}{\sigma^2}\right).\end{aligned}$$

Далее,

$$\ln \frac{n}{k(n-k)} = \ln \frac{1}{npq} - \ln\left(1+\frac{xq}{\sigma}\right) - \ln\left(1-\frac{xp}{\sigma}\right) = 2\ln\frac{1}{\sigma} + O\left(\frac{1}{\sigma}\right),$$

а также

$$\begin{aligned} k\ln\frac{k}{np} + (n-k)\ln\frac{n-k}{nq} &= (np+x\sigma)\ln\left(1+\frac{xq}{\sigma}\right) + (nq-x\sigma)\ln\left(1-\frac{xp}{\sigma}\right) \\ &= (np+x\sigma)\left(\frac{xq}{\sigma} - \frac{x^2q^2}{2\sigma^2} + o\left(\frac{1}{\sigma^2}\right)\right) \\ &+ (nq-x\sigma)\left(-\frac{xp}{\sigma} - \frac{x^2p^2}{2\sigma^2} + o\left(\frac{1}{\sigma^2}\right)\right) \\ &= x\sigma - \frac{x^2q}{2} + x^2q - x\sigma - \frac{x^2p}{2} + x^2p + O\left(\frac{1}{\sigma}\right) \\ &= \frac{x^2}{2} + O\left(\frac{1}{\sigma}\right). \end{aligned}$$

В результате получаем

$$\ln \mathsf{P}(B_n(k)) = \ln\frac{1}{\sigma\sqrt{2\pi}} - \frac{x^2}{2} + O\left(\frac{1}{\sigma}\right),$$

что и доказывает теорему. □

Теорема 4.3. *[Интегральная теорема Муавра-Лапласа.] Пусть в схеме Бернулли $0 < p < 1$ фиксировано и S_n — число успехов в серии из n независимых экспериментов. Тогда для $-\infty < a < b < +\infty$ при $n \to \infty$ имеет место предельное соотношение*

$$\lim_{n\to\infty} \mathsf{P}\left(a \leqslant \frac{S_n - np}{\sqrt{npq}} \leqslant b\right) = \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-\frac{x^2}{2}}\,dx.$$

Доказательство этой теоремы следует из центральной предельной теоремы, которая будет рассмотрена позже.

Предельные теоремы Пуассона и Муавра— Лапласа используют для приближенных вычислений вероятностей $\mathsf{P}(S_n = k)$ и $\mathsf{P}(k_1 \leqslant S_n \leqslant k_2)$ в схеме Бернулли, когда n велико. В теореме Пуассона появилось распределение Пуассона: $p_k = e^{-\lambda}\lambda^k/k!$, $k = 0, 1, 2, \ldots$. В теореме Муавра— Лапласа возникает нормальное распределение, которое описывается плотностью $\varphi(x) = e^{-x^2/2}/\sqrt{2\pi}$. Для распределения Пуассона и интеграла Лапласа

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{u^2}{2}}\,du$$

имеются таблицы.

Допустим, что нам нужно вычислить вероятность $\mathsf{P}(k_1 \leqslant S_n \leqslant k_2)$ в схеме Бернулли с n независимыми испытаниями и вероятностью p успеха в отдельном испытании. Если n велико, то можно использовать интегральную теорему

Муавра—Лапласа следующим образом:

$$\mathsf{P}(k_1 \leqslant S_n \leqslant k_2) = \mathsf{P}\left(\frac{k_1 - np}{\sqrt{npq}} \leqslant \frac{S_n - np}{\sqrt{npq}} \leqslant \frac{k_2 - np}{\sqrt{npq}}\right) \approx \Phi(x_2) - \Phi(x_1),$$

где

$$x_1 = \frac{k_1 - np}{\sqrt{npq}}, \quad x_2 = \frac{k_2 - np}{\sqrt{npq}}, \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{u^2}{2}} du.$$

Если считать, что $\Phi_0(x) = -\Phi_0(|x|)$ при отрицательных x, то $\Phi(x_2) - \Phi(x_1) = \Phi_0(x_2) - \Phi_0(x_1)$ для любых значений x_1, x_2. Это следует из четности функции $\varphi(x)$.

Полиномиальная схема является обобщением схемы Бернулли. Здесь результатом каждого испытания может быть один из r взаимоисключающих исходов $A_1, \ldots, A_r$ с вероятностями появления $p_1, \ldots, p_r$, соответственно, $p_1 + \ldots + p_r = 1$. Элементарное событие, соответствующее серии независимых n испытаний, можно представить в виде $\omega = (\delta_1, \ldots, \delta_n)$, где каждое δ_i может принимать одно из значений $1, \ldots, r$. Вероятность элементарного события в силу независимости испытаний будет определяться равенством

$$\mathsf{P}(\{\omega\}) = p_{\delta_1} \cdot \ldots \cdot p_{\delta_n}.$$

Подобно событию $B_n(k)$, которое рассматривалось в схеме Бернулли, в полиномиальной схеме вводится $B_n(k_1, \ldots, k_r)$ — событие, состоящее в том, что в серии из n экспериментов произошло k_1 исходов с номером 1, …, k_r исходов с номером r, $k_1 + \ldots + k_r = n$. Из общей формулы видно, что если $\omega \in B_n(k_1, \ldots, k_r)$, то

$$\mathsf{P}(\{\omega\}) = p_1^{k_1} \cdot \ldots \cdot p_r^{k_r}.$$

Используя основное правило комбинаторики, легко подсчитывается количество элементарных событий в $B_n(k_1, \ldots, k_r)$:

$$C_n^{k_1} \cdot C_{n-k_1}^{k_2} \cdot \ldots \cdot C_{k_r}^{k_r} = \frac{n!}{k_1! \cdot \ldots \cdot k_r!}.$$

Таким образом, мы приходим к основному результату полиномиальной схемы — *полиномиальному распределению*

$$\mathsf{P}(B_n(k_1, \ldots, k_r)) = \frac{n!}{k_1! \cdot \ldots \cdot k_r!} p_1^{k_1} \cdot \ldots \cdot p_r^{k_r}, \qquad k_1 + \ldots + k_r = n.$$

В случае $r = 2$ мы снова получаем схему Бернулли.

§ 5. Дискретные случайные величины

Часто с результатом случайного эксперимента связывают число. В связи с этим возникает важное понятие случайной величины.

Пусть $(\Omega, \mathscr{A}, \mathsf{P})$ — некоторое вероятное пространство. Под *случайной величиной*, определенной на этом вероятностном пространстве, будем понимать числовую функцию $\xi\colon\ \Omega \to \mathbb{R}$ такую, что

$$\{\omega\colon\ \xi(\omega) \leqslant x\} \in \mathscr{A}$$

для всех $x \in \mathbb{R}$. Другими словами, $\{\xi \leqslant x\}$ является событием и мы можем говорить о вероятности этого события. В этом параграфе мы будем изучать, в основном, дискретные случайные величины, которые принимают конечное или счетное множество значений. Если $\xi\colon\, \Omega \to \{x_1, x_2, \dots\}$, то условие того, что ξ — случайная величина, выражается просто: $\{\xi = x_k\} \in \mathscr{A}$, $k = 1, 2, \dots$.

Вначале рассмотрим случай, когда ξ принимает конечное число значений. Такие случайные величины называют *простыми*. Если ξ принимает только значения $x_1, \dots, x_n$, то события $D_k = \{\xi = x_k\}$ образуют разбиение $\mathscr{D}_\xi = \{D_1, \dots, D_n\}$. Через $\mathscr{A}_\xi$ будем обозначать алгебру, порожденную этим разбиением. Заметим также, что $\mathsf{P}(\xi = x_k) = \mathsf{P}(D_k)$.

Простейшей случайной величиной является индикатор события $A \in \mathscr{A}$, который определяется равенством

$$\mathbb{1}_A(\omega) \;=\; \begin{cases} 1, & \text{если} \quad \omega \in A, \\ 0, & \text{если} \quad \omega \notin A. \end{cases}$$

Отметим основные свойства индикаторов, которые следуют непосредственно из определения.

$$\mathbb{1}_\Omega(\omega) \,\equiv\, 1, \quad \mathbb{1}_\varnothing(\omega) \,\equiv\, 0, \quad \mathbb{1}_{\overline{A}}(\omega) \,=\, 1 - \mathbb{1}_A(\omega).$$

В дальнейшем мы будем опускать зависимость от элементарного события и писать просто $\mathbb{1}_A$. Имеют место также следующие соотношения

$$\mathbb{1}_{\cap A_i} \,=\, \prod \mathbb{1}_{A_i}, \quad \mathbb{1}_{\cup A_i} \,=\, 1 - \mathbb{1}_{\overline{\cup A_i}} \,=\, 1 - \mathbb{1}_{\cap \overline{A_i}} \,=\, 1 - \prod \left(1 - \mathbb{1}_{A_i}\right).$$

Пусть ξ — случайная величина, принимающая значения $x_1, \dots, x_n$ соответственно на $D_1, \dots, D_n$, т. е. $\mathscr{D}_\xi = \{D_1, \dots, D_n\}$. Тогда

$$\xi \,=\, \sum_{i=1}^{n} x_i \mathbb{1}_{D_i}. \tag{5.1}$$

Другими словами, простую случайную величину можно представить как линейную комбинацию индикаторов.

Допустим теперь, что проводится большое количество N экспериментов, соответствующих вероятностному пространству $(\Omega, \mathscr{A}, \mathsf{P})$. Тогда значение x_i будет приниматься случайной величиной ξ, определяемой равенством (5.1), приблизительно в Np_i, $p_i = \mathsf{P}(D_i)$, случайных экспериментах. Следовательно, среднее наблюдавшихся значений случайной величины ξ будет приблизительно равно величине

$$\frac{1}{N}\left(x_1 N p_1 + \ldots + x_n N p_n\right) \,=\, \sum_{i=1}^{n} x_i p_i.$$

Это делает интуитивно понятным следующее определение.

Определение 5.1. Математическим ожиданием случайной величины ξ, определяемой равенством (5.1), называется число

$$\mathsf{E}\xi \,=\, \sum_{i=1}^{n} x_i \mathsf{P}(D_i) \,=\, \sum_{i=1}^{n} x_i \mathsf{P}(\xi = x_i). \tag{5.2}$$

Замечание 5.1. Если случайная величина ξ, определяемая равенством (5.1), допускает также представление

$$\xi = \sum_{j=1}^{N} y_j \mathbb{1}_{H_j},$$

где $\{H_1, \ldots, H_N\}$ — разбиение, а $y_1, \ldots, y_N$ не обязательно все все различные, то

$$\mathsf{E}\xi = \sum_{j=1}^{N} y_j \mathsf{P}(H_j).$$

Доказательство. Действительно, множество индексов $\{1, \ldots, N\}$ можно разбить на подмножества $J_k = \{j\colon\ y_j = x_k\}$, $k = 1, \ldots, n$. Заметим, что при этом $D_k = \cup_{j \in J_k} H_j$. Следовательно,

$$\sum_{j=1}^{N} y_j \mathsf{P}(H_j) = \sum_{k=1}^{n} \sum_{j \in J_k} x_k \mathsf{P}(H_j) = \sum_{k=1}^{n} x_k \mathsf{P}(D_k) = \mathsf{E}\xi.$$

$\square$

Замечание 5.2. Если ξ определена равенством (5.1), а $\varphi\colon \mathbb{R} \to \mathbb{R}$ — некоторая функция, то математическое ожидание случайной величины $\eta = \varphi(\xi)$ можно вычислить по формуле

$$\mathsf{E}\eta = \sum_{i=1}^{n} \varphi(x_i) \mathsf{P}(D_i).$$

Свойства математического ожидания простых случайных величин.

1°. Для любого $A \in \mathscr{A}$ имеет место равенство $\mathsf{E}\mathbb{1}_A = \mathsf{P}(A)$;

2°. (Линейность.) Если ξ, η — случайные величины и $\alpha \in \mathbb{R}$, то $\mathsf{E}(\alpha\xi) = \alpha\mathsf{E}\xi$ и $\mathsf{E}(\xi + \eta) = \mathsf{E}\xi + \mathsf{E}\eta$;

3°. (Монотонность.) Если $\xi \geqslant 0$, то $\mathsf{E}\xi \geqslant 0$ и равенство $\mathsf{E}\xi = 0$ возможно лишь в случае $\mathsf{P}(\{\omega\colon\ \xi(\omega) \neq 0\}) = 0$, т. е. случайная величина ξ почти наверное равна нулю;

4°. Для любой случайной величины ξ имеет место неравенство $|\mathsf{E}\xi| \leqslant \mathsf{E}|\xi|$;

5°. (Неравенство Шварца.) Для любых случайных величин ξ и η имеет место неравенство $(\mathsf{E}\xi\eta)^2 \leqslant (\mathsf{E}\xi^2)(\mathsf{E}\eta^2)$.

Доказательство. Свойство 1° очевидно. Равенство $\mathsf{E}(\alpha\xi) = \alpha\mathsf{E}\xi$ следует из замечания 5.2. Допустим теперь, что $\xi = \sum_{i=1}^{n} x_i \mathbb{1}_{D_i}$, $\eta = \sum_{j=1}^{m} y_j \mathbb{1}_{H_j}$ — две случайные величины, а $\{D_1, \ldots, D_n\}$, $\{H_1, \ldots, H_m\}$ — соответствующие им разбиения. Тогда случайная величина $\xi + \eta$ может быть представлена в виде

$$\xi + \eta = \sum_{i,j} (x_i + y_j) \mathbb{1}_{D_i H_j}.$$

Но тогда, в силу замечания 5.1 имеем

$$\begin{aligned}
\mathsf{E}(\xi+\eta) &= \sum_{i,j}(x_i+y_j)\mathsf{P}(D_iH_j)\\
&= \sum_{i,j} x_i\mathsf{P}(D_iH_j) + \sum_{i,j} y_j\mathsf{P}(D_iH_j)\\
&= \sum_i x_i \sum_j \mathsf{P}(D_iH_j) + \sum_j y_j \sum_i \mathsf{P}(D_iH_j)\\
&= \sum_i \mathsf{P}(D_i) + \sum_j y_j\mathsf{P}(H_j)\\
&= \mathsf{E}\xi + \mathsf{E}\eta
\end{aligned}$$

и свойство 2° доказано.

Условие $\xi \geqslant 0$ означает, что в представлении (5.1) все x_i, $i = 1, \dots, n$, неотрицательны. Но тогда и $\mathsf{E}\xi$, как сумма неотрицательных слагаемых, также будет неотрицательной. Равенство $\mathsf{E}\xi = 0$ возможно лишь в случае, когда $\mathsf{P}(D_i) = 0$ для тех индексов i, которые соответствуют строго положительным значениям x_i. Это доказывает свойство 3°.

Свойство 4° является непосредственным следствием неравенства треугольника. Для доказательства 5° заметим вначале, что в случае равенства нулю правой части неравенства Шварца обращается в нуль и левая его часть. Если, например, $\mathsf{E}\xi^2 = 0$, то $\mathsf{P}(\xi = 0) = 1$ и $\mathsf{E}\xi\eta = 0$. Допустим теперь, что $\mathsf{E}\xi^2 > 0$ и $\mathsf{E}\eta^2 > 0$. Тогда корректно определены случайные величины

$$\hat{\xi} = \frac{|\xi|}{\sqrt{\mathsf{E}\xi^2}}, \qquad \hat{\eta} = \frac{|\eta|}{\sqrt{\mathsf{E}\eta^2}}.$$

Замечая, что $\mathsf{E}\hat{\xi}^2 = \mathsf{E}\hat{\eta}^2 = 1$, и используя очевидное неравенство $2\hat{\xi}\hat{\eta} \leqslant \hat{\xi}^2 + \hat{\eta}^2$, получаем

$$2\mathsf{E}\hat{\xi}\hat{\eta} \leqslant \mathsf{E}\hat{\xi}^2 + \mathsf{E}\hat{\eta}^2 = 2, \qquad \mathsf{E}\hat{\xi}\hat{\eta} \leqslant 1,$$

что эквивалентно неравенству Шварца. □

Другой важной числовой характеристикой случайной величины является дисперсия, которая характеризует степень «разброса» значений случайной величины.

Определение 5.2. Дисперсией случайной величины ξ называется число

$$\mathsf{D}\xi = \mathsf{E}(\xi - \mathsf{E}\xi)^2.$$

Величина $\sigma_\xi = \sqrt{\mathsf{D}\xi}$ называется стандартным (или среднеквадратическим) отклонением.

Свойства дисперсии.

1°. $\mathsf{D}\xi = \mathsf{E}\xi^2 - (\mathsf{E}\xi)^2$;

2°. Для любого вещественного числа c и случайной величины ξ имеют место равенства

$$\mathsf{D}(c\xi) = c^2\mathsf{D}\xi, \qquad \mathsf{D}(\xi + c) = \mathsf{D}\xi;$$

3°. Равенство $\mathsf{D}\xi = 0$ возможно лишь в случае $\mathsf{P}(\xi = \mathsf{E}\xi) = 1$, т. е. случайная величина ξ почти наверное равна постоянной.

Доказательство. Используя свойство линейности математического ожидания, получаем

$$\mathsf{D}\xi \;=\; \mathsf{E}(\xi - \mathsf{E}\xi)^2 \;=\; \mathsf{E}\xi^2 - 2\mathsf{E}\xi \cdot \mathsf{E}\xi + (\mathsf{E}\xi)^2 \;=\; \mathsf{E}\xi^2 - (\mathsf{E}\xi)^2.$$

Пусть теперь $c \in \mathbb{R}$ и ξ — случайная величина. Тогда

$$\mathsf{D}(c\xi) \;=\; \mathsf{E}(c\xi - c\mathsf{E}\xi)^2 \;=\; c^2\mathsf{E}(\xi - \mathsf{E}\xi)^2 \;=\; c^2\mathsf{D}\xi,$$

а также

$$\mathsf{D}(\xi + c) \;=\; \mathsf{E}(\xi + c - \mathsf{E}(\xi + c))^2 \;=\; \mathsf{E}(\xi - \mathsf{E}\xi)^2 \;=\; \mathsf{D}\xi.$$

Из свойства монотонности математического ожидания 3° следует, что $\mathsf{D}\xi = \mathsf{E}(\xi - \mathsf{E}\xi)^2 \geqslant 0$ и равенство $\mathsf{D}\xi = 0$ возможно лишь в случае $\xi - \mathsf{E}\xi = 0$ почти наверное. □

Независимость случайных величин.

Пусть ξ — случайная величина с представлением (5.1). Информация о том, какое из значений $x_1, \ldots, x_n$ приняла случайная величина ξ, эквивалентна тому, что мы знаем, какое из событий $D_1, \ldots, D_n$ разбиения $\mathscr{D}_\xi$ произошло. Но тогда мы можем сказать о любом событии из алгебры $\mathscr{A}_\xi$ произошло оно или нет. Другими словами, алгебра $\mathscr{A}_\xi$ отражает информацию, связанную со случайной величиной ξ. В связи с этим естественно ввести следующее определение.

Определение 5.3. Случайные величины ξ и η, определенные на вероятностном пространстве $(\Omega, \mathscr{A}, \mathsf{P})$, называются независимыми, если независимы алгебры $\mathscr{A}_\xi$ и $\mathscr{A}_\eta$.

Напомним, что условие независимости алгебр $\mathscr{A}_\xi$ и $\mathscr{A}_\eta$ эквивалентно тому, что для любых $A \in \mathscr{A}_\xi$ и $B \in \mathscr{A}_\eta$ выполняется равенство $\mathsf{P}(AB) = \mathsf{P}(A)\mathsf{P}(B)$. Кроме того, в силу теоремы 3.4 для того, чтобы ξ и η были независимы, достаточно выполнения условия $\mathsf{P}(DH) = \mathsf{P}(D)\mathsf{P}(H)$ для любых $D \in \mathscr{D}_\xi$ и $H \in \mathscr{D}_\eta$.

Теорема 5.1. *Если ξ и η — независимые случайные величины, то*

$$\mathsf{E}\xi\eta \;=\; \mathsf{E}\xi \cdot \mathsf{E}\eta, \qquad \mathsf{D}(\xi + \eta) \;=\; \mathsf{D}\xi + \mathsf{D}\eta.$$

Доказательство. Пусть

$$\xi \;=\; \sum_{i=1}^{n} x_i \mathbb{1}_{D_i}, \qquad \eta \;=\; \sum_{j=1}^{m} y_j \mathbb{1}_{H_j}.$$

Тогда, в силу независимости случайных величин, выполняются равенства $\mathsf{P}(D_i H_j) = \mathsf{P}(D_i)\mathsf{P}(H_j)$ и, следовательно,

$$\begin{aligned}\mathsf{E}(\xi\eta) \;&=\; \sum_{i,j} x_i y_j \mathsf{P}(D_i H_j) \;=\; \sum_{i,j} x_i y_j \mathsf{P}(D_i)\mathsf{P}(H_j) \\ &=\; \left(\sum_{i=1}^{n} x_i \mathsf{P}(D_i)\right)\left(\sum_{j=1}^{m} y_j \mathsf{P}(H_j)\right) \;=\; \mathsf{E}\xi\mathsf{E}\eta.\end{aligned}$$

Далее,

$$\mathsf{D}(\xi+\eta) \;=\; \mathsf{E}(\xi+\eta)^2-(\mathsf{E}(\xi+\eta))^2 \;=\; \mathsf{E}\xi^2+2\mathsf{E}(\xi\eta)+\mathsf{E}\eta^2-(\mathsf{E}\xi)^2-2\mathsf{E}\xi\mathsf{E}\eta-(\mathsf{E}\eta)^2.$$

Отсюда с использованием равенства $\mathsf{E}\xi\eta = \mathsf{E}\xi\mathsf{E}\eta$ получаем, что $\mathsf{D}(\xi+\eta) = \mathsf{D}\xi+\mathsf{D}\eta$ и теорема доказана. □

Замечание 5.3. Если ξ и η — две независимые случайные величины, а $\varphi\colon \mathbb{R}\to\mathbb{R}$ — некоторая функция, то $\varphi(\xi)$ и η также являются независимыми случайными величинами.

Действительно, это сразу же следует из включения $\mathscr{A}_{\varphi(\xi)}\subset\mathscr{A}_\xi$.

Ковариация и коэффициент корреляции. Эти числовые характеристики можно рассматривать как меру зависимости случайных величин.

Определение 5.4. Пусть ξ и η — две случайные величины, определенные на одном вероятностном пространстве. Под *ковариацией* этих случайных величин понимается число

$$\mathsf{cov}(\xi,\eta) \;=\; \mathsf{E}(\xi-\mathsf{E}\xi)(\eta-\mathsf{E}\eta).$$

Для вычисления ковариации нужно знать совместное распределение случайных величин. Пусть $\xi=\sum_{i=1}^n x_i\mathbb{1}_{D_i}$, $\eta=\sum_{j=1}^m y_j\mathbb{1}_{H_j}$, т. е. $\mathscr{D}_\xi=\{D_1,\dots,D_n\}$, $\mathscr{D}_\eta=\{H_1,\dots,H_m\}$. Рассмотрим случайный вектор (ξ,η), который принимает значение (x_i,y_j) на D_iH_j, $i=1,\dots,n$, $j=1,\dots,m$. Другими словами, случайному вектору (ξ,η) можно сопоставить разбиение

$$\mathscr{D}_{\xi,\eta} \;=\; \{D_iH_j\colon\; i=1,\dots,n, j=1,\dots,m\}.$$

Вероятности $p_{ij}=\mathsf{P}(D_iH_j)=\mathsf{P}(\xi=i,\eta=j)$ определяют *совместное распределение* случайных величин ξ и η. Совместное распределение двух простых случайных величин удобно представить таблицей.

$\xi\diagdown\eta$	y_1	y_2	$\dots$	y_m
x_1	p_{11}	p_{12}	$\dots$	p_{1m}
x_2	p_{11}	p_{12}	$\dots$	p_{1m}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
x_n	p_{n1}	p_{n2}	$\dots$	p_{nm}

Отметим некоторые особенности этой таблицы. Поскольку $\mathscr{D}_{\xi\eta}$ является разбиением, то $\sum_{i,j}p_{ij}=1$. Кроме того,

$$\sum_{j=1}^m p_{ij} \;=\; p_i^x \;=\; \mathsf{P}(\xi=x_i), \qquad \sum_{i=1}^n p_{ij} \;=\; p_j^y \;=\; \mathsf{P}(\eta=y_j).$$

Таким образом, совместное распределение случайных величин ξ и η позволяет получить их индивидуальные распределения. Условие независимости случайных величин ξ и η эквивалентна выполнению равенств $p_{ij}=p_i^xp_j^y$, $i=1,\dots,n$, $j=1,\dots,m$. Заметим также, что любая такая таблица с выполнением условий $p_{ij}\geqslant 0$, $\sum_{i,j}p_{ij}=1$ определяет совместное распределение двух случайных величин.

Наряду с ковариацией также рассматривают коэффициент корреляции. Его вводят в рассмотрение при условии, что $\mathsf{D}\xi > 0$ и $\mathsf{D}\eta > 0$. Заметим, что если дисперсия одной из случайных величин равна нулю, например $\mathsf{D}\xi = 0$, то $\xi = \mathsf{E}\xi$ почти наверное и в этом случае $\mathsf{cov}(\xi, \eta) = 0$. Допустим теперь, что $\mathsf{D}\xi > 0$ и $\mathsf{D}\eta > 0$. Тогда корректно определены случайные величины

$$\hat{\xi} = \frac{\xi - \mathsf{E}\xi}{\sqrt{\mathsf{D}\xi}}, \qquad \hat{\eta} = \frac{\eta - \mathsf{E}\eta}{\sqrt{\mathsf{D}\eta}},$$

которые удовлетворяют условиям

$$\mathsf{E}\hat{\xi} = \mathsf{E}\hat{\eta} = 0, \qquad \mathsf{D}\hat{\xi} = \mathsf{D}\hat{\eta} = 1.$$

Коэффициент корреляции $\varrho(\xi, \eta)$ определяется равенством

$$\varrho(\xi, \eta) = \mathsf{E}(\hat{\xi}\hat{\eta}) = \frac{\mathsf{cov}(\xi, \eta)}{\sqrt{\mathsf{D}\xi} \cdot \sqrt{\mathsf{D}\eta}}.$$

Свойства ковариации и коэффициента корреляции.

1°. $\mathsf{cov}(\xi, \eta) = \mathsf{E}\xi\eta - \mathsf{E}\xi \cdot \mathsf{E}\eta$;

2°. Если ξ и η независимы, то $\mathsf{cov}(\xi, \eta) = 0$;

3°. $|\varrho(\xi, \eta)| \leqslant 1$ и $|\varrho(\xi, \eta)| = 1$ в том и только том случае, если $\mathsf{P}(\eta = a\xi + b) = 1$ при некоторых $a, b \in \mathbb{R}$.

Доказательство. Свойства 1° и 2° очевидны. Остается доказать свойство 3°. В силу неотрицательности дисперсии имеем

$$0 \leqslant \mathsf{D}(\hat{\xi} \pm \hat{\eta}) = \mathsf{D}\hat{\xi} + \mathsf{D}\hat{\eta} \pm 2\varrho(\xi, \eta) = 2(1 \pm \varrho(\xi, \eta)).$$

Отсюда следует, что $-1 \leqslant \varrho(\xi, \eta) \leqslant 1$. Если же $\varrho(\xi, \eta) = 1$, то $\mathsf{D}(\hat{\xi} - \hat{\eta}) = 0$ и тогда $\hat{\xi} - \hat{\eta} \equiv \text{const}$ почти наверное. Это влечет линейную зависимость ξ и η. Аналогично, если $\varrho(\xi, \eta) = -1$, то $\mathsf{D}(\hat{\xi} + \hat{\eta}) = 0$ и тогда $\hat{\xi} + \hat{\eta} \equiv \text{const}$ почти наверное. □

Пример. Пусть ζ — случайная величина, которая с равными вероятностями $1/3$ принимает значения $0, \pi/2$ и π. Тогда случайные величины $\xi = \cos\zeta$ и $\eta = \sin\zeta$ функционально связаны, но $\mathsf{cov}(\xi, \eta) = 0$.

Определение 5.5. Случайные величины ξ и η называются некоррелированными, если $\mathsf{cov}(\xi, \eta) = 0$.

Для вычисления ковариации случайных величин достаточно знать таблицу их совместного распределения, поскольку

$$\mathsf{E}\xi = \sum_{i=1}^{n} x_i \sum_{j=1}^{m} p_{ij}, \qquad \mathsf{E}\eta = \sum_{j=1}^{m} y_j \sum_{i=1}^{n} p_{ij}, \qquad \mathsf{E}\xi\eta = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j p_{ij}.$$

Отметим также, что непосредственно из определений следуют равенства

$$\mathsf{D}(\xi + \eta) = \mathsf{D}\xi + \mathsf{D}\eta + 2\mathsf{cov}(\xi, \eta), \qquad \mathsf{cov}(\xi, \eta) = \sqrt{\mathsf{D}\xi} \cdot \sqrt{\mathsf{D}\eta} \cdot \varrho(\xi, \eta).$$

Ковариационная матрица.

Пусть $\xi_1, \ldots, \xi_n$ — случайные величины, определенные на одном вероятностном пространстве. Тогда степень их попарной зависимости отражает ковариационная матрица $\mathbb{V} = (v_{ij})$, $v_{ij} = \mathsf{cov}(\xi_i, \xi_j)$, $i, j = 1, \ldots, n$. В случае $i = j$ элемент v_{ij} является дисперсией $\mathsf{D}\xi$. Отметим некоторые особенности ковариационной матрицы. Это симметрическая матрица, на диагонали которой стоят неотрицательные числа. Кроме того, она неотрицательно определена, т. е.

$$\mathbf{x}^t \mathbb{V} \mathbf{x} = \sum_{i,j=1}^{n} v_{ij} x_i x_j \geqslant 0$$

для всех $\mathbf{x} = (x_1, \ldots, x_n)$ из $\mathbb{R}^n$. Будем $\mathbf{x} \in \mathbb{R}^n$ представлять как вектор-столбец. Также сформируем случайный вектор-столбец $\vec{\xi} = (\xi_1, \ldots, \xi_n)^t$ и пусть $\mathsf{E}\vec{\xi} = (\mathsf{E}\xi, \ldots, \mathsf{E}\xi_n)^t$ — его математическое ожидание. Тогда

$$\mathbb{V} = \mathsf{E}\{(\vec{\xi} - \mathsf{E}\vec{\xi})(\vec{\xi} - \mathsf{E}\vec{\xi})^t\},$$

а значение квадратичной формы на векторе $\mathbf{x} \in \mathbb{R}^n$ запишется в виде

$$\mathbf{x}^t \mathbb{V} \mathbf{x} = \mathsf{E}\{\mathbf{x}^t(\vec{\xi} - \mathsf{E}\vec{\xi})(\vec{\xi} - \mathsf{E}\vec{\xi})^t \mathbf{x}\} = \mathsf{E}(\mathbf{x}^t(\vec{\xi} - \mathsf{E}\vec{\xi}))^2 \geqslant 0.$$

Замечание 5.4. Симметричность и неотрицательная определенность являются характеристическими свойствами ковариационной матрицы.

Если $\mathsf{D}\xi_i = v_{ii} > 0$ при всех $i = 1, \ldots, n$, то определена также корреляционная матрица $\mathbb{K} = (\varrho_{ij})$, $\varrho_{ij} = v_{ij}/(\sigma_i \sigma_j)$, где $\sigma_i = \sqrt{\mathsf{D}\xi_i}$. Корреляционная матрица также является симметрической и неотрицательно определенной. Ее отличительной чертой является то, что на главной диагонали стоят единицы.

Задача линейного оценивания.

Пусть ξ и η — две случайные величины, из которых лишь ξ является наблюдаемой. Если ξ и η коррелированы, то естественно предположить, что знание значения ξ позволит получить некоторые выводы о значениях ненаблюдаемой случайной величины η. Всякую функцию $\varphi\colon \mathbb{R} \to \mathbb{R}$ в контексте этой задачи называют оценкой для η. Оценка φ^* называется оптимальной в смысле среднеквадратического отклонения в классе оценок Φ, если

$$\mathsf{E}(\eta - \varphi^*(\xi))^2 = \inf_{\varphi \in \Phi} \mathsf{E}(\eta - \varphi(\xi))^2.$$

Оказывается, что для отыскания оптимальной оценки в классе линейных функций $\varphi(x) = ax + b$ достаточно знания ковариации $\mathsf{cov}(\xi, \eta)$.

Рассмотрим функцию двух переменных

$$\psi(a, b) = \mathsf{E}(\eta - (a\xi + b))^2,$$

которая представляет собой неотрицательную квадратичную форму относительно переменных a и b. Поэтому она достигает минимума в некоторой точке (a^*, b^*), которая является решением системы уравнений

$$\frac{\partial \psi}{\partial a}(a, b) = 0, \qquad \frac{\partial \psi}{\partial b}(a, b) = 0.$$

Вычисляя частные производные функции ψ, получаем

$$\begin{cases} \mathsf{E}\xi\eta - a\mathsf{E}\xi^2 - b\mathsf{E}\xi \;=\; 0, \\ \quad \mathsf{E}\eta - a\mathsf{E}\xi - b \quad\;=\; 0. \end{cases}$$

Умножая второе равенство на $\mathsf{E}\xi$ и вычитая из первого, приходим к соотношению

$$\mathsf{cov}(\xi, \eta) - a\mathsf{D}\xi \;=\; 0,$$

откуда находим

$$a^* \;=\; \frac{\mathsf{cov}(\xi, \eta)}{\mathsf{D}\xi}$$

и далее

$$b^* \;=\; \mathsf{E}\eta - \frac{\mathsf{cov}(\xi, \eta)}{\mathsf{D}\xi}\mathsf{E}\xi.$$

Следовательно, оптимальной оценкой в классе линейных функций является

$$\eta^* \;=\; \varphi^*(\xi) \;=\; \mathsf{E}\eta + \frac{\mathsf{cov}(\xi, \eta)}{\mathsf{D}\xi}(\xi - \mathsf{E}\xi).$$

Это равенство называют также уравнением регрессии η на ξ.

Заметим также, что среднеквадратическая ошибка линейного оценивания вычисляется по формуле

$$\begin{aligned} \mathsf{E}(\eta - \eta^*)^2 \;&=\; \mathsf{E}\left((\eta - \mathsf{E}\eta) - \frac{\mathsf{cov}(\xi, \eta)}{\mathsf{D}\xi}(\xi - \mathsf{E}\xi)\right)^2 \\ &=\; \mathsf{D}\eta - 2\frac{(\mathsf{cov}(\xi, \eta))^2}{\mathsf{D}\xi} + \frac{(\mathsf{cov}(\xi, \eta))^2}{\mathsf{D}\xi} \\ &=\; \mathsf{D}\eta\left(1 - (\varrho(\xi, \eta))^2\right). \end{aligned}$$

Отсюда видно, что оценка тем точнее, чем ближе коэффициент корреляции $\varrho(\xi, \eta)$ по модулю к единице.

Целочисленные случайные величины и производящие функции.

С точки зрения распределения вероятностей легко ввести в рассмотрение случайные величины, принимающие счетное число значений: $\mathsf{P}(\xi = x_k) = p_k$, $p_k \geqslant 0$ и $\sum_{k=1}^{\infty} p_k = 1$. Однако в этом случае математическое ожидание и дисперсия не всегда определены.

Определение 5.6. Пусть случайная величина ξ принимает счетное число значений: $\mathsf{P}(\xi = x_k) = p_k$, $k = 1, 2, \ldots$. Будем говорить, что для ξ определено математическое ожидание, если сходится числовой ряд $\sum_{k=1}^{\infty} |x_k| p_k$. В этом случае определяется

$$\mathsf{E}\xi \;=\; \sum_{k=1}^{\infty} x_k p_k.$$

Дискретную случайную величину ξ, принимающую только целые неотрицательные значения, называют *целочисленной* случайной величиной. Ее распределение вероятностей $\mathsf{P}(\xi = k) = p_k$, $k = 0, 1, 2, \ldots$, удобно представлять *производящей функцией*

$$g_\xi(x) = \mathsf{E}x^\xi = \sum_{k=0}^{\infty} p_k x^k.$$

Заметим, что степенной ряд, определяющий производящую функцию, сходится при $|x| \leqslant 1$. При этом $g_\xi(1) = 1$ и

$$p_k = \frac{1}{k!} g_\xi^{(k)}(0), \qquad k = 0, 1, 2, \ldots.$$

Если $p_k \neq 0$ лишь для конечного числа индексов, то $g_\xi(x)$ представляет собой полином, а ξ принимает конечное число значений. В терминах производных от производящей функции легко вычисляются математическое ожидание и дисперсия случайной величины. Действительно,

$$\mathsf{E}\xi = \sum_{k=1}^{\infty} k p_k = g'_\xi(1),$$

и аналогично

$$\mathsf{D}\xi = \sum_{k=1}^{\infty} k^2 p_k - \left(\sum_{k=1}^{\infty} k p_k\right)^2 = g''_\xi(1) + g'_\xi(1) - \big(g'_\xi(1)\big)^2.$$

Наиболее часто встречающиеся в приложениях дискретные случайные величины являются целочисленными. Перечислим некоторые из них и найдем их производящие функции и числовые характеристики.

◇ Бернуллиевское распределение.

$$\mathsf{P}(\xi = 1) = p, \qquad \mathsf{P}(\xi = 0) = q, \quad q = 1 - p, \quad 0 < p < 1.$$

$$g_\xi(x) = px + q, \quad \mathsf{E}\xi = p, \quad \mathsf{D}\xi = pq.$$

◇ Биномиальное распределение.

$$\mathsf{P}(\xi = k) = C_n^k p^k q^{n-k}, \qquad 0 < p < 1, \qquad q = 1 - p, \qquad k = 0, 1, \ldots, n.$$

$$g_\xi(x) = (px + q)^n, \qquad \mathsf{E}\xi = np, \qquad \mathsf{D}\xi = npq.$$

◇ Пуассоновское распределение.

$$\mathsf{P}(\xi = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \qquad \lambda > 0, \qquad k = 0, 1, 2, \ldots.$$

$$g_\xi(x) = e^{\lambda(x-1)}, \qquad \mathsf{E}\xi = \lambda, \qquad \mathsf{D}\xi = \lambda.$$

◇ Геометрическое распределение.

$$\mathsf{P}(\xi = k) = pq^k, \qquad 0 < p < 1, \qquad q = 1 - p, \qquad k = 0, 1, \ldots.$$

$$g_\xi(x) = \frac{p}{1 - qx}, \qquad \mathsf{E}\xi = \frac{q}{p}, \qquad \mathsf{D}\xi = \frac{q}{p^2}.$$

§ 6. Пространство с мерой и общая модель вероятностного пространства

Напомним, что под алгеброй подмножеств Ω мы понимаем такой класс подмножеств Ω, который замкнут относительно конечного числа теоретико-множественных операций. При работе с бесконечными множествами часто приходится иметь дело с последовательностями подмножеств и производить с ними счетное число операций. Например, если $A_1, A_2, \ldots$ — последовательность множеств, то естественно возникают верхний и нижний пределы этой последовательности

$$A^* = \varlimsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k\geqslant n} A_k, \qquad A_* = \varliminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k\geqslant n} A_k.$$

Если $A_1, A_2, \ldots$ интерпретировать как события, то верхний предел A^* выражает то, что произошло бесконечно много событий последовательности, а нижний предел A_* соответствует тому, что произошли все события последовательности, за исключением, быть может, конечного числа. В случае $A_1 \subset A_2 \subset \ldots$ последовательность $\{A_n\}$ называется монотонно возрастающей и тогда $A^* = A_* = \cap_{n=1}^{\infty} A_n$. Аналогично, в случае $A_1 \supset A_2 \supset \ldots$ последовательность $\{A_n\}$ называется монотонно убывающей и $A^* = A_* = \cup_{n=1}^{\infty} A_n$. В общем случае последовательность $\{A_n\}$ называется сходящейся, если $A^* = A_*$.

Определение 6.1. Класс $\mathscr{A}$ подмножеств Ω называется σ-алгеброй, если он является алгеброй и замкнут относительно счетных объединений.

Замечание 6.1. Класс $\mathscr{A}$ подмножеств Ω является σ-алгеброй в том и только том случае, если выполнены следующие условия:

1° $\Omega \in \mathscr{A}$;
2° Если $A \in \mathscr{A}$, то $\bar{A} \in \mathscr{A}$;
3° Если $\{A_n\} \subset \mathscr{A}$, то $\cup_{n=1}^{\infty} A_n \in \mathscr{A}$.

Другими словами, σ-алгебра замкнута относительно счетного числа теоретико-множественных операций. Действительно, замкнутость относительно счетных объединений содержится в пункте 3°, а замкнутость относительно счетных пересечений выводится с использованием правил де Моргана.

Определение 6.2. Пусть $\mathscr{K}$ — некоторый класс подмножеств Ω. Тогда под $\sigma(\mathscr{K})$ будем понимать σ-алгебру подмножеств Ω, которая удовлетворяет следующим условиям:

(i) $\mathscr{K} \subset \sigma(\mathscr{K})$;
(ii) Если $\mathscr{F}$ — σ-алгебра подмножеств Ω и $\mathscr{K} \subset \mathscr{F}$, то $\sigma(\mathscr{K}) \subset \mathscr{F}$.

В силу условия (ii) σ-алгебру $\sigma(\mathscr{K})$ называют *минимальной σ-алгеброй, порожденной $\mathscr{K}$*. Поскольку 2^{Ω}, множество всех подмножеств Ω, является σ-алгеброй и пересечение любой совокупности σ-алгебр также является σ-алгеброй, то для любого семейства $\mathscr{K}$ подмножеств Ω существует и единственна $\sigma(\mathscr{K})$.

Если $\Omega = \mathbb{R}$, а $\mathscr{K}$ — совокупность открытых множеств в $\mathbb{R}$, то в качестве $\sigma(\mathscr{K})$ мы получаем *борелевскую* σ-алгебру подмножеств $\mathbb{R}$. Ее обычно обозначают через $\mathscr{B}(\mathbb{R})$ или $\mathscr{B}$. Множества из $\mathscr{B}$ называют *борелевскими* множествами.

Замечание 6.2. Если $\mathscr{T}$ — класс полуинтервалов вида $(-\infty, x]$, $x \in \mathbb{R}$, то $\sigma(\mathscr{T}) = \mathscr{B}$.

Действительно, $\sigma(\mathscr{T})$ должна содержать полуинтервалы вида $(a, b]$, поскольку $(a, b] = (-\infty, b] \setminus (-\infty, a]$. Но тогда замкнутость $\sigma(\mathscr{T})$ относительно счетных объединений влечет принадлежность и интервалов $(a, b) = \cup_{n=1}^{\infty}(a, b - \varepsilon/n]$, $0 < \varepsilon < (b - a)/2$. С другой стороны, любое открытое множество в $\mathbb{R}$ можно представить в виде объединения не более, чем счетного числа интервалов. Это влечет равенство $\sigma(\mathscr{T}) = \mathscr{B}$.

Определение 6.3. Пусть Ω — некоторое множество и $\mathscr{A}$ — некоторая σ-алгебра его подмножеств. Пара $(\Omega, \mathscr{A})$ называется измеримым пространством.

Введем еще несколько определений, связанных с понятием вероятности.

Определение 6.4. Пусть $(\Omega, \mathscr{A})$ — измеримое пространство. Неотрицательная функция $\mu\colon \mathscr{A} \to \mathbb{R}_+$ называется *аддитивной*, если $\mu(\varnothing) = 0$ и для всех $A, B \in \mathscr{A}$ таких, что $AB = \varnothing$, выполняется равенство $\mu(A \cup B) = \mu(A) + \mu(B)$. Функция $\mu\colon \mathscr{A} \to \mathbb{R}_+$ называется *счетно-аддитивной*, если $\mu(\varnothing) = 0$ и для любой последовательности $\{A_N\} \subset \mathscr{A}$ такой, что $A_i A_j = \varnothing$ при $i \neq j$, выполняется равенство

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

Определение 6.5. Пусть $(\Omega, \mathscr{A})$ — измеримое пространство. Функция $\mu\colon \mathscr{A} \to \mathbb{R}_+$ называется мерой на $(\Omega, \mathscr{A})$, если она счетно-аддитивна. Тройка $(\Omega, \mathscr{A}, \mu)$ в этом случае называется пространством с мерой. Мера μ на $(\Omega, \mathscr{A})$ называется *вероятностной* мерой, если $\mu(\Omega) = 1$.

В дальнейшем под *вероятностным пространством* будем понимать тройку $(\Omega, \mathscr{A}, \mathsf{P})$, где $(\Omega, \mathscr{A})$ — измеримое пространство, а P — вероятностная мера на $(\Omega, \mathscr{A})$. Это определение вероятностного пространства составляет *аксиоматику Колмогорова.*

Определение 6.6. Пусть $(\Omega, \mathscr{A})$ — измеримое пространство. Функция $\mu\colon \mathscr{A} \to \mathbb{R}_+$ называется непрерывной, если $\mu(\varnothing) = 0$ и для всякой монотонной исчезающей последовательности $H_n \searrow \varnothing$ в $\mathscr{A}$ выполняется условие

$$\mu(H_n) \to 0 \qquad \text{при} \qquad n \to \infty.$$

Теорема 6.1. *Пусть $(\Omega, \mathscr{A})$ — измеримое пространство и $\mu\colon \mathscr{A} \to \mathbb{R}_+$ — аддитивная функция. Для того, чтобы μ была счетно-аддитивной, необходимо и достаточно, чтобы она была непрерывной.*

Доказательство. Допустим, что μ является счетно-аддитивной и $\{H_n\} \subset \mathscr{A}$ — исчезающая последовательность. Определим $A_1 = H_1\bar{H}_2, A_2 = H_2\bar{H}_3, \ldots$.

Поскольку $H_n \searrow \varnothing$, то $A_1, A_2, \ldots$ образуют последовательность попарно непересекающихся множеств в $\mathscr{A}$. При этом для $n = 1, 2, \ldots$ выполняется равенство $H_n = \cup_{k=n}^{\infty} A_k$. В силу счетной аддитивности получаем

$$\mu(H_n) = \sum_{k=n}^{\infty} \mu(A_k)$$

и $\mu(H_n) \to 0$ при $n \to \infty$ как остаток сходящегося ряда.

Обратно, допустим теперь, что μ удовлетворяет условию непрерывности и $A_1, A_2, \ldots$ — последовательность попарно непересекающихся множеств из $\mathscr{A}$. Обозначим $A = \cup_{n=1}^{\infty} A_n$, $B_n = \cup_{k=1}^{n} A_k$. Тогда $B_n \nearrow A$ и в силу аддитивности μ выполняются равенства

$$\mu(B_n) = \sum_{k=1}^{n} \mu(A_k), \qquad \mu(A \setminus B_n) = \mu(A) - \sum_{k=1}^{n} \mu(A_k).$$

Замечая также, что $H_n = A \setminus B_n$, $n = 1, 2, \ldots$, образуют исчезающую последовательность, с использованием непрерывности μ получаем

$$\mu(A) = \mu(H_n) + \sum_{k=1}^{n} \mu(A_k) \to \sum_{k=1}^{\infty} \mu(A_k)$$

при $n \to \infty$. □

Следствие 6.1. *Пусть $(\Omega, \mathscr{A}, \mu)$ — пространство с мерой и $\{A_n\} \subset \mathscr{A}$ — монотонная последовательность с пределом $A = \lim_{n\to\infty} A_n$. Тогда*

$$\lim_{n\to\infty} \mu(A_n) = \mu(A).$$

Доказательство. Докажем это для монотонно возрастающей последовательности. Допустим, что $A_1 \subset A_2 \subset \ldots$. В этом случае множества $H_n = A \setminus A_n$, $n = 1, 2, \ldots$, образуют исчезающую последовательность и, следовательно, получаем

$$\mu(H_n) = \mu(A) - \mu(A_n) \to 0$$

при $n \to \infty$, что и доказывает наше утверждение. □

Уточним теперь понятие случайной величины, определенной на вероятностном пространстве $(\Omega, \mathscr{A}, \mathsf{P})$. Естественно исходить из того, что случайная величина — это функция от элементарного события, т. е. $\xi\colon \Omega \to \mathbb{R}$. Однако такое понимание случайной величины требует уточнения. Чтобы можно было говорить о вероятности того, что случайная величина ξ примет значение из интервала (a, b), необходимо, чтобы

$$\xi^{-1}((a,b)) = \{\omega \in \Omega\colon\ \xi(\omega) \in (a,b)\}$$

было событием, т. е. принадлежало σ-алгебре $\mathscr{A}$. Поскольку минимальным классом подмножеств $\mathbb{R}$, замкнутым относительно счетного числа теоретико-множественных операций и содержащим интервалы, является борелевская σ-алгебра $\mathscr{B}(\mathbb{R})$, то мы естественным образом приходим к следующему определению.

Определение 6.7. Пусть $(\Omega, \mathscr{A}, \mathsf{P})$ — вероятностное пространство. Функция $\xi\colon \Omega \to \mathbb{R}$ называется случайной величиной, определенной на этом вероятностном пространстве, если для всякого борелевского множества $B \in \mathscr{B}(\mathbb{R})$ выполняется условие

$$\{\omega \in \Omega\colon\ \xi(\omega) \in B\} \in \mathscr{A},$$

т. е. $\xi^{-1}(B) \in \mathscr{A}$.

Поскольку операция взятия прообраза обладает свойством

$$\xi^{-1}\left(\bigcup_{\alpha} E_\alpha\right) = \bigcup_{\alpha} \xi^{-1}(E_\alpha), \quad \xi^{-1}\left(\bigcap_{\alpha} E_\alpha\right) = \bigcap_{\alpha} \xi^{-1}(E_\alpha)$$

для любого семейства множеств E_α, то

$$\sigma(\xi) := \{A = \xi^{-1}(B)\colon\ B \in \mathscr{B}(\mathbb{R})\}$$

является σ-алгеброй подмножеств Ω и $\sigma(\xi) \subset \mathscr{A}$.

Теорема 6.2. *Пусть* $(\Omega, \mathscr{A}, \mathsf{P})$ *— вероятностное пространство и* $\xi\colon \Omega \to \mathbb{R}$ *такая, что для всех* $x \in \mathbb{R}$ *выполняется условие*

$$\{\omega \in \Omega\colon\ \xi(\omega) \leqslant x\} \in \mathscr{A}.$$

Тогда ξ *является случайной величиной.*

Доказательство. Нам нужно показать, что $\xi^{-1}(B) \in \mathscr{A}$ для всех $B \in \mathscr{B}$. Обозначим

$$\mathscr{F} = \{B \subset \mathbb{R}\colon\ \xi^{-1}(B) \in \mathscr{A}\}$$

и покажем, что $\mathscr{F}$ является σ-алгеброй подмножеств $\mathbb{R}$. Действительно, $\mathbb{R} \in \mathscr{F}$ поскольку $\xi^{-1}(\mathbb{R}) = \Omega \in \mathscr{A}$. Если $B \in \mathscr{F}$, то

$$\xi^{-1}(\mathbb{R} \setminus B) = \Omega \setminus \xi^{-1}(B) \in \mathscr{A}$$

и $\mathbb{R} \setminus B$ также принадлежит $\mathscr{F}$. Наконец, если $B_n \in \mathscr{F}$, $n = 1, 2, \ldots$, то $\xi^{-1}(\cup_{n=1}^{\infty} B_n) = \cup_{n=1}^{\infty} \xi^{-1}(B_n)$ принадлежит $\mathscr{A}$, т. е. $\cup_{n=1}^{\infty} B_n$ принадлежит $\mathscr{F}$. Таким образом выполнены условия, из которых следует, что $\mathscr{F}$ является σ-алгеброй. Далее, по условию теоремы совокупность $\mathscr{T}$ полуинтервалов $(-\infty, x]$, $x \in \mathbb{R}$, содержится в $\mathscr{F}$. Но тогда и $\sigma(\mathscr{T}) = \mathscr{B} \subset \mathscr{F}$. □

Замечание 6.3. Заключение теоремы будет верным, если в ее предположениях полуинтервалы $(-\infty, x]$ заменить полуинтервалами любого из трех видов: $(-\infty, x)$, (x, ∞), $[x, \infty)$, поскольку минимальная σ-алгебра, порожденная семейством полуинтервалов любого из этих трех видов также совпадает с борелевской σ-алгеброй.

Определение 6.8. Случайные величины $\xi_1, \ldots, \xi_n$, определенные на одном и том же вероятностном пространстве $(\Omega, \mathscr{A}, \mathsf{P})$, называются независимыми, если независимы σ-алгебры $\sigma(\xi_1), \ldots, \sigma(\xi_n)$, т. е. для любых борелевских множеств $B_1, \ldots, B_n$ выполняется условие

$$\mathsf{P}(\{\omega \in \Omega\colon\ \xi_1(\omega) \in B_1, \ldots, \xi_n(\omega) \in B_n\}) = \prod_{k=1}^{n} \mathsf{P}(\{\omega \in \Omega\colon\ \xi_k(\omega) \in B_k\}).$$

Заметим теперь, что случайную величину ξ, определенную на вероятностном пространстве $(\Omega, \mathscr{A}, \mathsf{P})$, можно рассматривать как измеримое отображение измеримых пространств $(\Omega, \mathscr{A})$ и $(\mathbb{R}, \mathscr{B})$. При этом вероятностную меру P, определенную на $\mathscr{A}$, можно перенести на борелевскую σ-алгебру $\mathscr{B}(\mathbb{R})$ посредством равенства

$$\mathsf{P}_\xi(B) := \mathsf{P}(\xi^{-1}(B)) = \mathsf{P}(\{\omega \in \Omega\colon\ \xi(\omega) \in B\}), \qquad B \in \mathscr{B}.$$

Определенная таким образом P_ξ является вероятностной мерой на измеримом пространстве $(\mathbb{R}, \mathscr{B})$. Действительно,

$$\mathsf{P}_\xi(\mathbb{R}) = \mathsf{P}(\{\omega \in \Omega\colon\ \xi(\omega) \in \mathbb{R}\}) = \mathsf{P}(\Omega) = 1,$$

а если $B_1, B_2, \ldots$ — попарно непересекающиеся борелевские множества, то

$$\mathsf{P}_\xi\left(\bigcup_{n=1}^{\infty} B_n\right) = \mathsf{P}\left(\xi^{-1}\left(\bigcup_{n=1}^{\infty} B_n\right)\right) = \mathsf{P}\left(\bigcup_{n=1}^{\infty} \xi^{-1}(B_n)\right)$$
$$= \sum_{n=1}^{\infty} \mathsf{P}(\xi^{-1}(B_n)) = \sum_{n=1}^{\infty} \mathsf{P}_\xi(B_n).$$

Здесь мы воспользовались свойством взятия прообраза и тем, что множества $\xi^{-1}(B_1), \xi^{-1}(B_2), \ldots$ попарно не пересекаются.

Таким образом, каждая случайная величина ξ порождает вероятностное пространство $(\mathbb{R}, \mathscr{B}, \mathsf{P}_\xi)$. При этом мера P_ξ называется *распределением вероятностей* случайной величины ξ.

Определение 6.9. Функцию

$$F_\xi(x) = \mathsf{P}_\xi((-\infty, x]) = \mathsf{P}(\{\omega \in \Omega\colon\ \xi(\omega) \leqslant x\}),$$

определенную на $\mathbb{R}$, называют *функцией распределения* случайной величины ξ.

Посредством функции распределения F_ξ сразу же определяется мера полуинтервалов

$$\mathsf{P}_\xi((a, b]) = F_\xi(b) - F_\xi(a).$$

С полуинтервалов мера P_ξ однозначно продолжается до меры на измеримом пространстве $(\mathbb{R}, \mathscr{B})$. Отметим характеристические свойства функции распределения:

1°. $F_\xi(x)$ не убывает на $\mathbb{R}$;

2°. $F_\xi(x)$ непрерывна справа в каждой точке $x \in \mathbb{R}$;

3°. $F_\xi(-\infty) = \lim\limits_{x \to -\infty} F_\xi(x) = 0, \qquad F_\xi(\infty) = \lim\limits_{x \to \infty} F_\xi(x) = 1.$

Доказательство. Монотонность функции F_ξ следует из свойства монотонности меры P_ξ. Действительно, если $x' < x''$, то $(-\infty, x'] \subset (-\infty, x'']$ и $\mathsf{P}_\xi((-\infty, x']) \leqslant \mathsf{P}_\xi((-\infty, x''])$, т. е. $F_\xi(x') \leqslant F_\xi(x'')$. Допустим теперь, что $x_n \searrow x$ при $n \to \infty$. Тогда $(-\infty, x_n] \searrow (-\infty, x]$ и по свойству непрерывности меры P_ξ имеем

$$\lim_{n \to \infty} F_\xi(x_n) = \lim_{n \to \infty} \mathsf{P}_\xi((-\infty, x_n]) = \mathsf{P}_\xi((-\infty, x]) = F_\xi(x).$$

Тем самым доказана непрерывность справа функции F_ξ. Свойство 3° также следует из непрерывности меры P_ξ поскольку $(-\infty, x_n] \searrow \varnothing$ при $x_n \searrow -\infty$ и $(-\infty, x_n] \nearrow \mathbb{R}$ при $x_n \nearrow \infty$. □

Заметим, что любая функция F, определенная на $\mathbb{R}$ и удовлетворяющая условиям 1° — 3°, может рассматриваться как функция распределения некоторой случайной величины.

В случае дискретной случайной величины ξ ее функция распределения является ступенчатой. Среди непрерывных функций распределения выделяются *абсолютно непрерывные*, которые определяются плотностью

$$F_\xi(x) \;=\; \int\limits_{-\infty}^{x} f_\xi(u)\,du.$$

В качестве плотности f_ξ может выступать любая неотрицательная функция f на $\mathbb{R}$, которая интегрируема на всей числовой оси и

$$\int\limits_{-\infty}^{\infty} f(x)\,dx \;=\; 1.$$

В терминах плотности f_ξ просто выражаются следующие вероятности

$$\mathsf{P}(\{\omega\colon\ a < \xi(\omega) < b\}) \;=\; \mathsf{P}(\{\omega\colon\ a \leqslant \xi(\omega) \leqslant b\}) \;=\; \int\limits_{a}^{b} f_\xi(x)\,dx.$$

Кроме того, в точках непрерывности плотности функция F_ξ дифференцируема и выполняется равенство

$$F_\xi'(x) \;=\; f_\xi(x).$$

Среди наиболее часто встречающихся в приложениях абсолютно непрерывных распределений отметим следующие:

♢ Нормальное (или гауссовское) с плотностью

$$f_\xi(x) \;=\; \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/2\sigma^2}, \qquad a \in \mathbb{R}, \quad \sigma > 0;$$

♢ Равномерное на $[a, b]$

$$f_\xi(x) \;=\; \frac{1}{b-a}\mathbb{1}_{[a,b]}(x);$$

♢ Показательное

$$f_\xi(x) \;=\; \lambda e^{-\lambda x}\mathbb{1}_{[0,\infty)}(x), \qquad \lambda > 0;$$

♢ Коши

$$f_\xi(x) \;=\; \frac{1}{\pi}\frac{1}{x^2+1}.$$

Совместное распределение случайных величин также можно описать функцией распределения. Пусть ξ и η — две случайные величины, определенные на одном вероятностном пространстве $(\Omega, \mathscr{A}, \mathsf{P})$. Тогда под их *совместной функцией распределения* понимают

$$F_{\xi,\eta}(x,y) \;=\; \mathsf{P}(\{\omega \in \Omega \colon\; \xi(\omega) \leqslant x,\; \eta(\omega) \leqslant y\}),$$

$(x,y) \in \mathbb{R}^2$. Отметим свойства $F_{\xi,\eta}$ как функции двух переменных.

1*. $F_{\xi,\eta}(x,y)$ не убывает по каждой переменной;

2*. $F_{\xi,\eta}(x,y)$ непрерывна справа по каждой переменной и

$$F_{\xi,\eta}(-\infty,y) \;=\; F_{\xi,\eta}(x,-\infty) \;=\; 0, \qquad F_{\xi,\eta}(+\infty,+\infty) \;=\; 1;$$

3*. Для всех $x, y \in \mathbb{R}$, $\Delta x \geqslant 0$, $\Delta y \geqslant 0$ выполняется неравенство

$$F_{\xi,\eta}(x+\Delta x, y+\Delta y) - F_{\xi,\eta}(x, y+\Delta y) - F_{\xi,\eta}(x+\Delta x, y) + F_{\xi,\eta}(x,y) \;\geqslant\; 0.$$

Свойства $1^* - 3^*$ являются характеристическими для совместной функции распределения. Кроме того, совместная функция распределения содержит в себе индивидуальные (одномерные) функции распределения:

$$F_\xi(x) \;=\; \lim_{y\to\infty} F_{\xi,\eta}(x,y), \qquad F_\eta(y) \;=\; \lim_{x\to\infty} F_{\xi,\eta}(x,y).$$

Важный класс непрерывных совместных функций распределения составляют *абсолютно непрерывные*, которые определяются *плотностью*

$$F_{\xi,\eta}(x,y) \;=\; \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f_{\xi,\eta}(u,v)\,du dv.$$

В терминах плотности $f_{\xi,\eta}$ удобно вычислять вероятность попадания случайной точки (ξ,η) в область $D \subset \mathbb{R}^2$:

$$\mathsf{P}((\xi,\eta) \in D) \;=\; \iint\limits_{D} f_{\xi,\eta}(x,y)\,dxdy.$$

В качестве плотности совместного распределения двух случайных величин можно рассматривать любую неотрицательную функцию $f(x,y)$, для которой

$$\iint\limits_{\mathbb{R}^2} f(x,y)\,dxdy \;=\; 1.$$

В точках непрерывности плотности $f_{\xi,\eta}(x,y)$ выполняется равенство

$$\frac{\partial^2}{\partial x \partial y} F_{\xi,\eta}(x,y) \;=\; f_{\xi,\eta}(x,y).$$

Если совместная функция распределения $F_{\xi,\eta}$ абсолютно непрерывна, то также абсолютно непрерывными будут одномерные функции распределения F_ξ и F_η. При этом

$$f_\xi(x) \;=\; \int\limits_{-\infty}^{\infty} f_{\xi,\eta}(x,y)\,dy, \qquad f_\eta(y) \;=\; \int\limits_{-\infty}^{\infty} f_{\xi,\eta}(x,y)\,dx.$$

Замечание 6.4. Случайные величины ξ и η независимы в том и только том случае, если для всех $x, y \in \mathbb{R}$ выполняется равенство

$$F_{\xi,\eta}(x,y) \;=\; F_\xi(x)F_\eta(y).$$

В абсолютно непрерывном случае это равенство эквивалентно следующему

$$f_{\xi,\eta}(x,y) \;=\; f_\xi(x)f_\eta(y).$$

Аналогично двумерному случаю рассматривается совместное распределение n случайных величин.

§ 7. Математическое ожидание

Ранее мы определили математическое ожидание для дискретных случайных величин. В общем случае математическое ожидание (если оно существует) вводится путем аппроксимации случайной величины последовательностями простых случайных величин и предельным переходом.

Теорема 7.1. *[Аппроксимационная] Пусть ξ — неотрицательная случайная величина, определенная на вероятностном пространстве $(\Omega, \mathscr{A}, \mathsf{P})$. Тогда найдется последовательность $\{\xi_n\}$ простых неотрицательных случайных величин такая, что $\xi_n(\omega) \nearrow \xi(\omega)$ при $n \to \infty$ для всех $\omega \in \Omega$.*

Доказательство. Для каждого натурального n введем разбиение $\mathscr{D}_n = \left\{\Delta_n, D_1^{(n)}, \ldots, D_{n2^n}^{(n)}\right\}$, где

$$\Delta_n \;=\; \{\omega\colon\ \xi(\omega) \geqslant n\}, \quad D_k^{(n)} \;=\; \left\{\omega\colon\ \frac{k-1}{2^n} \leqslant \xi(\omega) < \frac{k}{2^n}\right\}, \quad k = 1, 2, \ldots, n2^n.$$

Далее определим простые случайные величины

$$\xi_n \;=\; n \cdot \mathbb{1}_{\Delta_n} + \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \cdot \mathbb{1}_{D_k^{(n)}}$$

и покажем, что $\xi_n(\omega) \nearrow \xi(\omega)$ при $n \to \infty$ для всех $\omega \in \Omega$. Вначале докажем монотонность этой последовательности простых случайных величин. Пусть $\omega \in \Omega$ фиксировано. Если $\xi(\omega) \geqslant n$, то $\xi_n(\omega) = n$, а $\xi_{n+1} \geqslant n$, как видно из определения последовательности случайных величин ξ_n. В случае $\xi(\omega) < n$ найдется $k \leqslant n2^n$ такое, что $\omega \in D_k^{(n)}$. Но тогда $\xi_n(\omega) = (k-1)2^{-n}$, а

$$D_k^{(n)} \;=\; D_{2k-1}^{(n+1)} \cup D_{2k}^{(n+1)}.$$

Если $\omega \in D_{2k-1}^{(n+1)}$, то

$$\xi_{n+1}(\omega) \;=\; (2k-2)\cdot 2^{-(n+1)} \;=\; (k-1)\cdot 2^{-n} \;=\; \xi_n(\omega),$$

а если $\omega \in D_{2k}^{(n+1)}$, то

$$\xi_{n+1}(\omega) \;=\; (2k-1)\cdot 2^{-(n+1)} \;=\; (k-1)\cdot 2^{-n} + 2^{-(n+1)} \;=\; \xi_n(\omega) + 2^{-(n+1)}.$$

Монотонность последовательности $\xi_n(\omega)$ таким образом доказана.

Заметим теперь, что для фиксированного $\omega \in \Omega$ найдется такое натуральное N, что $\xi(\omega) < N$. Из построения последовательности следует, что при $n \geqslant N$ будут выполняться неравенства

$$\xi_n(\omega) \leqslant \xi(\omega) < \xi_n(\omega) + 2^{-n}.$$

Отсюда следует, что $\xi_n(\omega) \nearrow \xi(\omega)$ при $n \to \infty$. □

Для доказательства корректности вводимого ниже определения математического ожидания нам потребуется также следующий результат.

Теорема 7.2. *Пусть $\eta, \xi_1, \xi_2, \ldots$ — простые неотрицательные случайные величины и $\xi_n(\omega) \nearrow \xi(\omega) \geqslant \eta(\omega)$ при $n \to \infty$ для всех $\omega \in \Omega$. Тогда*

$$\mathsf{E}\eta \leqslant \lim_{n\to\infty} \mathsf{E}\xi_n.$$

Доказательство. Фиксируем произвольно $\varepsilon > 0$ и определим

$$A_n = \{\omega\colon\ \xi_n(\omega) \geqslant \eta(\omega) - \varepsilon\}, \qquad n = 1, 2, \ldots.$$

Поскольку

$$\xi_n = \xi_n \mathbb{1}_{A_n} + \xi_n \mathbb{1}_{\overline{A}_n} \geqslant (\eta - \varepsilon)\mathbb{1}_{A_n} = \eta - \eta\mathbb{1}_{\overline{A}_n} - \varepsilon\mathbb{1}_{A_n} \geqslant \eta - M\mathbb{1}_{\overline{A}_n} - \varepsilon,$$

где $M = \max\{\eta(\omega)\colon\ \omega \in \Omega\}$, то

$$\mathsf{E}\xi_n \geqslant \mathsf{E}\eta - \varepsilon - M\mathsf{P}(\overline{A}_n).$$

Из монотонности последовательности $\{\xi_n\}$ и условия $\xi \geqslant \eta$ следует, что $A_n \nearrow \Omega$ при $n \to \infty$. Но тогда $\overline{A}_n \searrow \varnothing$ и в силу непрерывности вероятностной меры $\mathsf{P}(\overline{A}_n) \to 0$ при $n \to \infty$. Поскольку числовая последовательность $\{\mathsf{E}\xi_n\}$ не убывает, то она либо стремится к ∞, либо сходится к конечному пределу. Следовательно,

$$\lim_{n\to\infty} \mathsf{E}\xi_n \geqslant \mathsf{E}\eta - \varepsilon.$$

Поскольку $\varepsilon > 0$ выбиралось произвольно, то

$$\lim_{n\to\infty} \mathsf{E}\xi_n \geqslant \mathsf{E}\eta$$

и теорема доказана. □

Следствие 7.1. *Пусть $\{\xi_n\}$, $\{\eta_n\}$ — две последовательности простых неотрицательных случайных величин и $\xi_n \nearrow \xi$, $\eta_n \nearrow \xi$. Тогда*

$$\lim_{n\to\infty} \mathsf{E}\xi_n = \lim_{n\to\infty} \mathsf{E}\eta_n.$$

Доказательство. Из теоремы следует, что для каждого $k = 1, 2, \ldots$ выполняется неравенство

$$\lim_{n\to\infty} \mathsf{E}\xi_n \geqslant \mathsf{E}\eta_k.$$

Но тогда

$$\lim_{n\to\infty} \mathsf{E}\xi_n \geqslant \lim_{k\to\infty} \mathsf{E}\eta_k.$$

Меняя ролями эти последовательности, получим обратное неравенство, что и влечет требуемое равенство. □

Определение 7.1. Пусть ξ — неотрицательная случайная величина и $\xi_n \nearrow \xi$, где $\{\xi_n\}$ — последовательность простых неотрицательных случайных величин. Тогда

$$\mathsf{E}\xi \;:=\; \lim_{n\to\infty} \mathsf{E}\xi_n,$$

если этот предел конечен.

Заметим, что существование последовательности $\{\xi_n\}$ следует из теоремы 7.1, а то, что определение $\mathsf{E}\xi$ не зависит от выбора аппроксимирующей последовательности, следует из теоремы 7.2 и следствия 7.1.

Для произвольной случайной величины ξ рассматриваются две неотрицательные случайные величины

$$\xi^+ \;=\; \max\{\xi, 0\}, \qquad \xi^- \;=\; \max\{-\xi, 0\}.$$

Очевидно, что $\xi = \xi^+ - \xi^-$. В случае, когда $\mathsf{E}\xi^+$ и $\mathsf{E}\xi^-$ конечны, будем говорить, что ξ имеет конечное математическое ожидание

$$\mathsf{E}\xi \;:=\; \mathsf{E}\xi^+ - \mathsf{E}\xi^-.$$

Из определения следует, что в случае, когда случайная величина ξ имеет конечное математическое ожидание, то и $|\xi|$ также имеет конечное математическое ожидание, поскольку $|\xi| = \xi^+ + \xi^-$.

Введенное таким образом математическое ожидание сохраняет основные свойства, которые были установлены для простых случайных величин:

1° Линейность: $\mathsf{E}(a\xi + b\eta) = a\mathsf{E}\xi + b\mathsf{E}\eta$;

2° Монотонность: если $\xi \geqslant 0$, то $\mathsf{E}\xi \geqslant 0$ и равенство $\mathsf{E}\xi = 0$ влечет $\xi \overset{\text{п.н.}}{=} 0$, т. е. $\mathsf{P}(\{\omega\colon\ \xi(\omega) = 0\}) = 1$;

3° Если ξ имеет конечное математическое ожидание, то $|\mathsf{E}\xi| \leqslant \mathsf{E}|\xi|$, а в случае конечных математических ожиданий $\mathsf{E}\xi^2$ и $\mathsf{E}\eta^2$ выполняется неравенство Шварца $(\mathsf{E}\xi\eta)^2 \leqslant (\mathsf{E}\xi^2)(\mathsf{E}\eta^2)$.

Кроме того, если $\xi_1, \ldots, \xi_n$ — независимые случайные величины с конечными математическими ожиданиями, то $\xi = \prod_{k=1}^n \xi_k$ также имеет конечное математическое ожидание и выполняется равенство

$$\mathsf{E}\xi \;=\; \prod_{k=1}^{n} \mathsf{E}\xi_k.$$

Сформулируем также без доказательства основные теоремы о предельном переходе под знаком математического ожидания.

Теорема 7.3. *[О монотонной сходимости] Пусть $\xi_1, \xi_2, \ldots$ — неотрицательные случайные величины и $\xi_n \nearrow \xi$. Тогда $\mathsf{E}\xi_n \nearrow \mathsf{E}\xi$ при $n \to \infty$.*

Здесь для удобства формулировки теоремы допускается случай $\mathsf{E}\xi = +\infty$. Тогда и $\mathsf{E}\xi_n \to \infty$ при $n \to \infty$. Это допущение предполагается и в последующих формулировках теорем.

Теорема 7.4. *[Лемма Фату] Пусть $\xi_1, \xi_2, \ldots$ — неотрицательные случайные величины. Тогда*

$$\mathsf{E} \varliminf_{n\to\infty} \xi_n \;\leqslant\; \varliminf_{n\to\infty} \mathsf{E}\xi_n.$$

Теорема 7.5. *[Лебега о мажорируемой сходимости] Пусть $\eta, \xi_1, \xi_2, \ldots$ — случайные величины, для которых выполняются следующие условия: $|\xi_n| \leqslant \eta$ для всех n, $\xi_n \to \xi$ при $n \to \infty$ и $\mathsf{E}\eta < \infty$. Тогда*

$$\lim_{n\to\infty} \mathsf{E}\xi_n \,=\, \mathsf{E}\xi, \qquad \lim_{n\to\infty} \mathsf{E}|\xi_n - \xi| \,=\, 0.$$

В действительности, определение математического ожидания, приведенное выше, является интегралом Лебега от функции $\xi\colon\ \Omega \to \mathbb{R}$ по мере P и

$$\mathsf{E}\xi \,=\, \int\limits_{\Omega} \xi(\omega) d\mathsf{P}(\omega) \,=\, \int\limits_{\Omega} \xi(\omega) \mathsf{P}(d\omega).$$

Как ранее отмечалось, с каждой случайной величиной ξ ассоциируется новое вероятностное пространство $(\mathbb{R}, \mathscr{B}, \mathsf{P}_\xi)$, где $\mathscr{B}$ — борелевская σ-алгебра на $\mathbb{R}$, а P_ξ — распределение вероятностей (борелевская мера) случайной величины ξ. Все вычисления, связанные со случайной величиной ξ, мы можем перенести на $(\mathbb{R}, \mathscr{B}, \mathsf{P}_\xi)$. В частности, если ξ имеет конечное математическое ожидание $\mathsf{E}\xi$, то

$$\mathsf{E}\xi \,=\, \int\limits_{\mathbb{R}} x d\mathsf{P}_\xi(x).$$

Кроме того, если $\varphi\colon\ \mathbb{R} \to \mathbb{R}$ — борелевская функция (для любого борелевского множества B прообраз $\varphi^{-1}(B)$ также является борелевским множеством), то $\eta = \varphi(\xi)$ также будет случайной величиной. При этом, если η имеет конечное математическое ожидание $\mathsf{E}\eta = \mathsf{E}\varphi(\xi)$, то

$$\mathsf{E}\varphi(\xi) \,=\, \int\limits_{\mathbb{R}} \varphi(x) d\mathsf{P}_\xi(x). \tag{7.1}$$

В случае $\varphi(x) = x$ мы снова приходим к формуле для вычисления $\mathsf{E}\xi$. Формула (7.1) является основной при вычислениях, связанных со случайной величиной ξ. В частности, для вычисления дисперсии $\mathsf{D}\xi = \mathsf{E}\xi^2 - (\mathsf{E}\xi)^2$ требуется существование интеграла

$$\mathsf{E}\xi^2 \,=\, \int\limits_{\mathbb{R}} x^2 d\mathsf{P}_\xi(x).$$

В случае дискретной случайной величины $\xi = \sum_{k=1}^n x_k \mathbb{1}_{D_k}$ формула (7.1) принимает вид

$$\mathsf{E}\varphi(\xi) \,=\, \sum_{k=1}^n \varphi(x_k)\mathsf{P}_\xi(\{x_k\}) \,=\, \sum_{k=1}^n \varphi(x_k)\mathsf{P}(\{\omega\colon\ \xi(\omega) = x_k\}).$$

Если же распределение вероятностей случайной величины ξ описывается плотностью f_ξ, то формула (7.1) записывается в виде

$$\mathsf{E}\varphi(\xi) \,=\, \int\limits_{\mathbb{R}} \varphi(x) f_\xi(x) dx \,=\, \int\limits_{-\infty}^{\infty} \varphi(x) f_\xi(x) dx. \tag{7.2}$$

Интуитивно этот интеграл можно трактовать как бесконечную «сумму» произведений значений $\varphi(x)$ случайной величины $\varphi(\xi)$ на «вероятности» $f_\xi(x)dx$, с которыми эти значения принимаются. Формально интеграл в равенстве (7.2) понимается в смысле Лебега. Однако в случае, когда он существует, функция $|\varphi(x)|f_\xi(x)$ должна быть интегрируемой по Риману в несобственном смысле на $(-\infty, \infty)$ и несобственный интеграл Римана от $\varphi(x)f_\xi(x)$ будет совпадать с интегралом Лебега.

Допустим теперь, что ξ и η — две случайные величины, определенные на одном вероятностном пространстве $(\Omega, \mathscr{A}, \mathsf{P})$. Тогда мы можем рассмотреть измеримое отображение $(\xi, \eta)\colon (\Omega, \mathscr{A}) \to (\mathbb{R}^2, \mathscr{B}(\mathbb{R}^2))$, где $\mathscr{B}(\mathbb{R}^2)$ — минимальная σ-алгебра подмножеств в $\mathbb{R}^2$, порожденная открытыми множествами. Как и в случае одной случайной величины, мы можем определить на $\mathscr{B}(\mathbb{R}^2)$ меру посредством равенства

$$\mathsf{P}_{\xi,\eta}(B) = \mathsf{P}(\{\omega\colon (\xi(\omega), \eta(\omega)) \in B\}), \qquad B \in \mathscr{B}(\mathbb{R}^2).$$

Мера $\mathsf{P}_{\xi,\eta}$ называется *совместным распределением* случайных величин ξ и η, или распределением вероятностей случайного вектора (ξ, η). Посредством совместной функции распределения $F_{\xi,\eta}$ легко определяется мера прямоугольников

$$\mathsf{P}_{\xi,\eta}((a_1, a_2] \times (b_1, b_2]) = F_{\xi,\eta}(a_2, b_2) - F_{\xi,\eta}(a_1, b_2) - F_{\xi,\eta}(a_2, b_1) + F_{\xi,\eta}(a_1, b_1).$$

В случае абсолютно непрерывного совместного распределения, когда существует плотность $f_{\xi,\eta}$, вероятность попадания случайной точки (ξ, η) в область $D \subset \mathbb{R}^2$ определяется двойным интегралом

$$\mathsf{P}(\{\omega\colon (\xi(\omega), \eta(\omega)) \in D\}) = \iint\limits_D f_{\xi,\eta}(x, y)dxdy.$$

Если $\varphi(x, y)$ — борелевская функция и $\zeta = \varphi(\xi, \eta)$ имеет конечное математическое ожидание, то

$$\mathsf{E}\varphi(\xi, \eta) = \iint\limits_{\mathbb{R}^2} \varphi(x, y) f_{\xi,\eta}(x, y)dxdy.$$

В частности,

$$\mathsf{E}\xi\eta = \iint\limits_{\mathbb{R}^2} xy f_{\xi,\eta}(x, y)dxdy.$$

Это дает способ вычисления ковариации $\mathsf{cov}(\xi, \eta)$ случайных величин ξ и η, которые имеют абсолютно непрерывное совместное распределение.

§ 8. Неравенство Чебышева. Закон больших чисел

Неравенство Чебышева позволяет оценить вероятность отклонения случайной величины от ее математического ожидания.

Теорема 8.1. *Пусть случайная величина ξ имеет конечную дисперсию (конечный второй момент). Тогда для любого $\varepsilon > 0$ выполняются следующие неравенства*

$$\mathsf{P}(\{\omega\colon\ |\xi(\omega)| \geqslant \varepsilon\}) \leqslant \frac{\mathsf{E}|\xi|}{\varepsilon}, \tag{8.1}$$

$$\mathsf{P}(\{\omega\colon\ |\xi(\omega) - \mathsf{E}\xi| \geqslant \varepsilon\}) \leqslant \frac{\mathsf{D}\xi}{\varepsilon^2}. \tag{8.2}$$

Доказательство. Из очевидного неравенства

$$|\xi| = |\xi|\mathbb{1}_{\{|\xi|\geqslant\varepsilon\}} + |\xi|\mathbb{1}_{\{|\xi|<\varepsilon\}} \geqslant \varepsilon\mathbb{1}_{\{|\xi|\geqslant\varepsilon\}}$$

и свойства монотонности математического ожидания получаем

$$\mathsf{E}|\xi| \geqslant \varepsilon\mathsf{P}(\{|\xi| \geqslant \varepsilon\}),$$

откуда следует неравенство (8.1). Доказательство неравенства (8.2) основывается на доказанном неравенстве (8.1). Действительно,

$$\mathsf{P}(\{|\xi - \mathsf{E}\xi| \geqslant \varepsilon\}) = \mathsf{P}(\{(\xi - \mathsf{E}\xi)^2 \geqslant \varepsilon^2\}) \leqslant \frac{\mathsf{D}\xi}{\varepsilon^2}$$

и теорема доказана. □

Неравенство (8.1) иногда называют неравенством Маркова. Для его выполнения достаточно конечности первого момента, т. е. $\mathsf{E}|\xi| < \infty$. Конечность второго момента, т. е. $\mathsf{E}\xi^2 < \infty$, влечет и конечность первого момента. Это следует из неравенства Шварца

$$(\mathsf{E}\xi\eta)^2 \leqslant (\mathsf{E}\xi^2)(\mathsf{E}\eta^2),$$

которое легко переносится с простых случайных величин на произвольные. Неравенство (8.2) известно как неравенство Чебышева.

Приведем одно из приложений неравенства Чебышева, известное как «правило трех сигм». Пусть ξ — случайная величина с математическим ожиданием $\mathsf{E}\xi = m$ и дисперсией $\mathsf{D}\xi = \sigma^2$. Тогда применение неравенства Чебышева (8.2) с $\varepsilon = 3\sigma$ дает следующую оценку вероятности отклонения случайной величины от ее математического ожидания

$$\mathsf{P}(\{|\xi - m| \geqslant 3\sigma\}) \leqslant \frac{1}{9}.$$

Это общее неравенство существенно улучшается в случае конкретных распределений. В частности, если случайная величина ξ имеет нормальное распределение, то вероятность отклонения ξ от m на 3σ очень близка к нулю.

Теорема 8.2. *[Закон больших чисел Чебышева] Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин, для которых $\mathsf{D}\xi_n \leqslant C$, $n = 1, 2, \ldots$, при некотором $C > 0$. Тогда для любого $\varepsilon > 0$ выполняется следующее предельное соотношение*

$$\lim_{n\to\infty} \mathsf{P}\left(\left|\frac{S_n}{n} - \frac{\mathsf{E}S_n}{n}\right| \geqslant \varepsilon\right) = 0,$$

где $S_n = \xi_1 + \ldots + \xi_n$.

Доказательство. Применяя неравенство Чебышева (8.2), получаем

$$\mathsf{P}\left(\left|\frac{S_n}{n} - \frac{\mathsf{E}S_n}{n}\right| \geqslant \varepsilon\right) \leqslant \frac{\mathsf{D}S_n}{\varepsilon^2 n^2}.$$

Поскольку случайные величины $\xi_1, \ldots, \xi_n$ независимы, то

$$\mathsf{D}S_n = \mathsf{D}\xi_1 + \ldots + \mathsf{D}\xi_n \leqslant Cn.$$

Следовательно,

$$\mathsf{P}\left(\left|\frac{S_n}{n} - \frac{\mathsf{E}S_n}{n}\right| \geqslant \varepsilon\right) \leqslant \frac{Cn}{\varepsilon^2 n^2} = \frac{C}{\varepsilon^2 n} \to 0$$

при $n \to \infty$. Теорема доказана. □

Следствие 8.1. *Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин с одинаковыми математическими ожиданиями $\mathsf{E}\xi_n = a$ и одинаковыми дисперсиями $\mathsf{D}\xi_n = \sigma^2$, $n = 1, 2, \ldots$. Тогда для любого $\varepsilon > 0$ выполняется соотношение*

$$\lim_{n\to\infty} \mathsf{P}\left(\left|\frac{S_n}{n} - a\right| \geqslant \varepsilon\right) = 0.$$

Предельные соотношения в теореме 8.2 и следствии 8.1 принято называть *законами больших чисел*. Содержание закона больших чисел состоит в том, что среднее арифметическое суммы независимых случайных величин становится близким к постоянной с вероятностью, сколь угодно близкой к единице, при числе слагаемых, стремящимся к бесконечности. Один из первых законов больших чисел был получен Бернулли.

Теорема 8.3. *[Бернулли] Пусть S_n — число успехов в серии из n независимых испытаний с вероятностью p, $0 < p < 1$, в отдельном испытании. Тогда для любого $\varepsilon > 0$ выполняется соотношение*

$$\lim_{n\to\infty} \mathsf{P}\left(\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right) = 0.$$

Доказательство. Пусть ξ_k — индикатор успеха в k-том испытании, т. е. $\xi_k = 1$, если в k-том испытании произошел успех, и $\xi_k = 0$ в противном случае. Тогда $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин. При этом

$$\mathsf{P}(\xi_k = 1) = p, \quad \mathsf{P}(\xi_k = 0) = 1 - p, \quad \mathsf{E}\xi_k = p, \quad \mathsf{D}\xi_k = p(1-p).$$

Применяя к этой последовательности следствие 8.1, получаем требуемое утверждение. □

Замечая, что в условиях теоремы 8.3 S_n — количество успехов в n независимых испытаниях, а S_n/n — частота появления успеха, приходим к теоретическому обоснованию того, что вероятность события проявляется через частоту его появления.

§ 9. Характеристические функции. Центральная предельная теорема

Аналитический аппарат производящих функций является эффективным инструментом, но имеет ограниченную сферу применения в рамках целочисленных случайных величин. В общем случае аналогичную роль играют характеристические функции.

Для определения характеристической функции нам нужно перейти к комплексным случайным величинам. Пусть $\xi = \xi_1 + i\xi_2$, где ξ_1 и ξ_2 — вещественные случайные величины с конечными математическими ожиданиями. Тогда определим $\mathsf{E}\xi = \mathsf{E}\xi_1 + i\mathsf{E}\xi_2$. Основные свойства математического ожидания (линейность и мультипликативное свойство) переносятся и на комплексные случайные величины. Используя тригонометрическую форму записи комплексных чисел

$$z \;=\; |z|(\cos\varphi + i\sin\varphi) \;=\; |z|e^{i\varphi},$$

где φ — аргумент комплексного числа, устанавливается неравенство

$$|\mathsf{E}\xi| \;=\; |\mathsf{E}\xi_1 + i\mathsf{E}\xi_2| \;\leqslant\; \mathsf{E}\left(\sqrt{\xi_1^2+\xi_2^2}\right) \;=\; \mathsf{E}|\xi|.$$

Действительно, пусть $\mathsf{E}\xi = |\mathsf{E}\xi|e^{i\theta}$. Тогда

$$|\mathsf{E}\xi| \;=\; e^{-i\theta}\mathsf{E}\xi \;=\; \mathsf{E}(e^{-i\theta}\xi) \;=\; \mathsf{Re}\{\mathsf{E}(e^{-i\theta}\xi)\} \;=\; \mathsf{E}\mathsf{Re}\{e^{-i\theta}\xi\} \;\leqslant\; \mathsf{E}|\xi|.$$

Под *характеристической функцией* вещественной случайной величины понимается функция

$$h_\xi(t) \;=\; \mathsf{E}e^{it\xi}$$

от переменной $t \in \mathbb{R}$. Поскольку $|e^{it\xi}| = 1$, то математическое ожидание в определении h_ξ существует, т. е. характеристическая функция (как и функция распределения) определена для каждой случайной величины. Характеристическая функция вполне определяется распределением вероятностей случайной величины. Если ξ имеет дискретное распределение $\mathsf{P}(\xi = x_k) = p_k, k = 1, \ldots, n$, то

$$h_\xi(t) \;=\; \sum_{k=1}^{n} p_k e^{ix_k t} \;=\; \sum_{k=1}^{n} p_k(\cos x_k t + i\sin x_k t).$$

В случае абсолютно непрерывного распределения, т. е. когда распределение вероятностей случайной величины ξ описывается плотностью $f_\xi(x)$, характеристическая функция определяется по формуле

$$h_\xi(t) \;=\; \int\limits_{-\infty}^{\infty} e^{itx} f_\xi(x)dx.$$

Свойства характеристических функций.

1°. $|h_\xi(t)| \leqslant t$ и $h_\xi(0) = 1$;

2°. $h_\xi(-t) = \overline{h_\xi(t)}$;

3°. $h_{a\xi+b}(t) = e^{itb}h_\xi(at), \quad a, b \in \mathbb{R}$;

4°. Если $\xi_1, \ldots, \xi_n$ — независимые случайные величины и $S_n = \xi_1 + \ldots + \xi_n$, то $h_{S_n}(t) = \prod_{k=1}^n h_{\xi_k}(t)$;

5°. Функция $h_\xi(t)$ равномерно непрерывна на $\mathbb{R}$ и если $\mathsf{E}|\xi|^n < \infty$, то $h_\xi(t)$ имеет производные до n-го порядка включительно и

$$\mathsf{E}\xi^n = \frac{1}{i^n} h_\xi^{(n)}(0).$$

Свойства $1^\circ - 4^\circ$ следуют непосредственно из определения характеристической функции и свойств математического ожидания. Доказательство последнего свойства опирается на теорему Лебега о предельном переходе под знаком интеграла.

Теорема 9.1. *[Бохнера] Для того, чтобы непрерывная на $\mathbb{R}$ функция $h(t)$ с $h(0) = 1$ была характеристической функцией некоторого распределения, необходимо и достаточно, чтобы она была положительно определенной, т. е. для любых $t_1, \ldots, t_n$ из $\mathbb{R}$ и $z_1, \ldots, z_n$ из $\mathbb{C}$, $n = 1, 2, \ldots$, выполнялось условие*

$$\sum_{k,l=1}^n h(t_k - t_l) z_k \overline{z}_l \geqslant 0.$$

Доказательство. Необходимость условия следует из простых преобразований.

$$\begin{aligned}
\sum_{k,l} h_\xi(t_k - t_l) z_k \overline{z}_l &= \mathsf{E}\left(\sum_{k,l} z_k \overline{z}_l e^{i(t_k - t_l)\xi}\right) \\
&= \mathsf{E}\left(\sum_k z_k e^{it_k\xi}\right)\left(\overline{\sum_l z_l e^{it_l\xi}}\right) \\
&= \mathsf{E}\left|\sum_k z_k e^{it_k\xi}\right|^2 \geqslant 0.
\end{aligned}$$

Доказательство достаточности условия теоремы выходит за рамки данного курса. □

Приведем теперь характеристические функции некоторых часто встречающихся в приложениях распределений.

♣ Вырожденное распределение $\mathsf{P}(\xi = a) = 1$.

Непосредственно из определения характеристической функции получаем ее вид

$$h_\xi(t) = \mathsf{E}e^{it\xi} = e^{iat}.$$

♣ Равномерное на отрезке распределение с плотностью

$$f_\xi(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x).$$

Ее характеристическая функция также получается непосредственно из определения. В случае $t \neq 0$ имеем

$$h_\xi(t) = \frac{1}{b-a} \int_a^b e^{itx} dx = \left.\frac{e^{itx}}{it(b-a)}\right|_{x=a}^{x=b} = \frac{e^{itb} - e^{ita}}{it(b-a)}.$$

При $t = 0$ имеем $h_\xi(0) = 1$.

♣ Экспоненциальное распределение зависит от параметра $\lambda > 0$ и определяется плотностью

$$f_\xi(x)\lambda e^{-\lambda x}\mathbb{1}_{[0,\infty)}(x).$$

Здесь снова воспользуемся определением характеристической функции

$$h_\xi(t) = \lambda \int\limits_0^\infty e^{itx} e^{-\lambda x} dx = \left. \frac{\lambda e^{(it-\lambda)x}}{it-\lambda} \right|_{x=0}^{x=\infty} = \frac{\lambda}{\lambda - it}.$$

♣ Нормальное распределение, определяемое плотностью

$$f_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/2\sigma^2}$$

с двумя параметрами $a \in \mathbb{R}$ и $\sigma > 0$, имеет важное значение как в теории, так и в приложениях.

Для вычисления характеристической функции нормального распределения вначале рассмотрим случай, когда $a = 0$ и $\sigma = 1$. В этом случае говорят, что случайная величина ξ имеет стандартное нормальное распределение. При этом

$$\begin{aligned} h_\xi(t) &= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} e^{itx} e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \left[\int\limits_{-\infty}^{\infty} \cos tx \cdot e^{-x^2/2} dx + i \int\limits_{-\infty}^{\infty} \sin tx \cdot e^{-x^2/2} dx \right] \\ &= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} \cos tx \cdot e^{-x^2/2} dx, \end{aligned}$$

поскольку интеграл в мнимой части берется по симметричному промежутку от нечетной функции. Замечая, что выполнены условия дифференцирования несобственного интеграла, зависящего от параметра, получаем

$$\begin{aligned} h'_\xi(t) &= -\frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} x \sin tx \cdot e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} \sin tx \cdot de^{-x^2/2} \\ &= -\frac{t}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} \cos tx \cdot e^{-x^2/2} dx = -t h_\xi(t). \end{aligned}$$

Таким образом, характеристическая функция стандартного нормального распределения является решением дифференциального уравнения

$$h'_\xi(t) = -t h_\xi(t).$$

Это вместе с условием $h_\xi(0) = 1$ приводит к ее виду

$$h_\xi(t) \,=\, e^{-t^2/2}.$$

Пусть теперь $\sigma > 0$ и $a \in \mathbb{R}$. Случайная величина $\eta = \sigma\xi + a$ будет иметь, в силу свойства 3°, характеристическую функцию

$$h_\eta(t) \,=\, e^{iat}e^{-\sigma^2 t^2/2}.$$

С другой стороны,

$$F_\eta(x) \,=\, \mathsf{P}(\sigma\xi + a \leqslant x) \,=\, \mathsf{P}(\xi \leqslant \frac{x-a}{\sigma}) \,=\, \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{(x-a)/\sigma} e^{-u^2/2} du,$$

т. е.

$$f_\eta(x) \,=\, \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

и η имеет нормальное распределение с параметрами a, σ^2.
Зная вид характеристической функции, легко вычислить математическое ожидание и дисперсию случайной величины с использованием свойства 5°. В частности, если ξ имеет стандартное нормальное распределение, то из равенств

$$h'_\xi(t) \,=\, -th_\xi(t), \qquad h''_\xi(t) \,=\, -h_\xi(t) - th'_\xi(t)$$

следует, что $h'_\xi(0) = 0$, $h''_\xi(0) = -1$. Отсюда получаем

$$\mathsf{E}\xi \,=\, 0, \qquad \mathsf{D}\xi \,=\, \mathsf{E}\xi^2 \,=\, 1.$$

Для произвольного нормального распределения, когда $\eta = \sigma\xi + a$, получаем

$$\mathsf{E}\eta \,=\, a, \qquad \mathsf{D}\eta \,=\, \sigma^2.$$

Таким образом, параметры нормального распределения имеют вероятностный смысл.

♣ Целочисленная случайная величина с производящей функцией $g_\xi(x)$ имеет характеристическую функцию $h_\xi(t) = g_\xi(e^{it})$. Это следует непосредственно из определений характеристической и производящей функций.

Важным моментом в применении характеристических функций является то, что они вполне определяют распределение случайной величины. Приведем два результата, известные как *формулы обращения.* Напомним вначале, что распределение случайной величины ξ представляет собой вероятностную меру P_ξ, определенную на σ-алгебре борелевских множеств, т. е. на измеримом пространстве $(\mathbb{R}, \mathscr{B})$. Эта мера связана с функцией распределения F_ξ равенствами

$$F_\xi(x) \,=\, \mathsf{P}_\xi((-\infty, x]), \qquad \mathsf{P}_\xi((a, b]) \,=\, F_\xi(b) - F_\xi(a).$$

Теорема 9.2. *Для любых $a < b$ выполняется соотношение*

$$\lim_{R\to\infty} \frac{1}{2\pi} \int\limits_{-R}^{R} \frac{e^{-iat} - e^{-ibt}}{it} h_\xi(t)dt = \frac{1}{2}\left[\mathsf{P}_\xi([a,b)) + \mathsf{P}_\xi((a,b])\right]$$
$$= \frac{1}{2}\left[F_\xi(b-0) - F_\xi(a-0) + F_\xi(b) - F_\xi(a)\right].$$

Теорема 9.3. *Пусть характеристическая функция h_ξ абсолютно интегрируема, т. е.*

$$\int\limits_{-\infty}^{\infty} |h_\xi(t)|dt < \infty.$$

Тогда ξ имеет абсолютно непрерывное распределение, а ее плотность определяется по формуле

$$f_\xi(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-ixt} h_\xi(t)dt.$$

Из формулы обращения следует, что между множеством функций распределения F и множеством характеристических функций h имеет место взаимно-однозначное соответствие. Это соответствие является даже непрерывным относительно соответствующих сходимостей.

Определение 9.1. Будем говорить, что последовательность функций распределения F_n, $n = 1, 2, \ldots$, *слабо сходится* к функции распределения F и писать $F_n \Rightarrow F$, если $F_n(x) \to F(x)$ при $n \to \infty$ в каждой точке непрерывности функции F.

Теорема 9.4. *Пусть $F_1, F_2, \ldots$ — функции распределения, а $h_1, h_2, \ldots$ — соответствующие им характеристические функции. Тогда имеют место следующие утверждения:*

(i) Если $F_n \Rightarrow F$ и h — характеристическая функция, соответствующая функции распределения F, то $h_n(t) \to h(t)$ при $n \to \infty$ для всех $t \in \mathbb{R}$;

(ii) Если $h_n(t) \to h(t)$ при $n \to \infty$ для всех $t \in \mathbb{R}$ и функция h непрерывна в точке $t = 0$, то h является характеристической функцией некоторой функции распределения F и $F_n \Rightarrow F$.

Важную роль в приложениях играют результаты, утверждающие нормальность распределения сумм независимых случайных величин. Обычно их объединяют в одну группу с названием *центральной предельной теоремы.* Мы рассмотрим один из вариантов центральной предельной теоремы.

Теорема 9.5. *Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин с конечными $\mathsf{E}\xi = a$ и $\mathsf{D}\xi = \sigma^2$. Тогда для $S_n = \xi_1 + \ldots + \xi_n$ выполняется следующее предельное соотношение*

$$\lim_{n\to\infty} \mathsf{P}\left(\frac{S_n - na}{\sigma\sqrt{n}} \leqslant x\right) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-u^2/2}du$$

для всех $x \in \mathbb{R}$

Доказательство. Обозначим $\zeta_n = (S_n - na)/\sigma\sqrt{n}$, $n = 1, 2, \ldots$. Тогда утверждение теоремы эквивалентно тому, что $F_{\zeta_n} \Rightarrow \Phi$, где Φ — функция Лапласа, т. е. функция распределения случайной величины η, имеющей стандартное нормальное распределение. В силу теоремы сходимости это эквивалентно условию $h_{\zeta_n}(t) \to h_\eta(t)$ при $n \to \infty$ для всех $t \in \mathbb{R}$.

Рассмотрим случайные величины $\hat{\xi}_n = (\xi_n - a)/\sigma$, $n = 1, 2, \ldots$. Очевидно, что $\mathsf{E}\hat{\xi}_n = 0$ и $\mathsf{D}\hat{\xi}_n = \mathsf{E}\hat{\xi}_n^2 = 1$. При этом

$$\zeta_n = \frac{\hat{\xi}_1}{\sqrt{n}} + \ldots + \frac{\hat{\xi}_n}{\sqrt{n}}.$$

Пусть

$$h(t) = h_{\hat{\xi}_n}(t) = \mathsf{E}e^{it\hat{\xi}_n}.$$

Поскольку $h(0) = 1$, $h'(0) = i\mathsf{E}\hat{\xi}_n = 0$, $h''(0) = -\mathsf{E}\hat{\xi}_n^2 = -1$, то

$$h(t) = 1 - \frac{1}{2}t^2 + o(t^2).$$

Но тогда

$$h_{\zeta_n}(t) = \left(h_{\hat{\xi}_n/\sqrt{n}}(t)\right)^n = \left(h\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{1}{2}\frac{t^2}{n} + o\left(\frac{t^2}{n}\right)\right)^n.$$

При фиксированном $t \in \mathbb{R}$ имеем $t^2/n \to 0$ при $n \to \infty$ и

$$\lim_{n\to\infty} h_{\zeta_n}(t) = \lim_{n\to\infty} e^{n\ln(1 - \frac{t^2}{n} + o(\frac{t^2}{n}))} = e^{-t^2/2}.$$

Замечая, что $h_\eta(t) = e^{-t^2/2}$, приходим к утверждению теоремы. □

Следствие 9.1 (теорема Муавра—Лапласа). *Пусть S_n — число успехов в схеме Бернулли из n независимых испытаний с вероятностью p, $0 < p < 1$, успеха в отдельном испытании. Тогда для $-\infty < a < b < \infty$ выполняется предельное соотношение*

$$\lim_{n\to\infty} \mathsf{P}\left(a \leqslant \frac{S_n - np}{\sqrt{npq}} \leqslant b\right) = \Phi(b) - \Phi(a),$$

где $q = 1 - p$.

Доказательство. Заметим, что S_n можно представить как сумму независимых случайных величин ξ_k, которые являются индикаторами успеха в k-ом испытании: $\mathsf{P}(\xi_k = 1) = p$, $\mathsf{P}(\xi_k = 0) = q$, $k = 1, 2, \ldots$. Тогда $\mathsf{E}\xi_k = p$, $\mathsf{D}\xi_k = pq$. Применение центральной предельной теоремы дает требуемый результат. □

§ 10. Виды сходимости последовательностей случайных величин

В этом параграфе рассматриваются основные виды сходимости последовательностей случайных величин и устанавливается взаимосвязь между ними.

Определение 10.1. Будем говорить, что последовательность случайных величин $\xi_1, \xi_2, \ldots$ сходится *по вероятности* к случайной величине ξ, и писать $\xi_n \xrightarrow{\mathsf{P}} \xi$, если для всякого $\varepsilon > 0$ выполняется предельное соотношение

$$\lim_{n\to\infty} \mathsf{P}(|\xi_n - \xi| \geqslant \varepsilon) \,=\, 0.$$

Именно этот вид сходимости имеет место в законе больших чисел Чебышёва. Заметим, что из сходимости по вероятности последовательности случайных величин мы ничего не можем сказать о том, как ведет себя последовательность в отдельных экспериментах. В этом отношении более содержательную информацию дает следующий вид сходимости.

Определение 10.2. Будем говорить, что последовательность случайных величин $\xi_1, \xi_2, \ldots$ сходится к случайной величине ξ с вероятностью 1 или *почти наверное*, и писать $\xi_n \xrightarrow{\text{п.н.}} \xi$, если

$$\mathsf{P}(\{\omega\colon \lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)\}) \,=\, 1.$$

Следующий вид сходимости требует от случайных величин существования средних порядка p.

Определение 10.3. Будем говорить, что последовательность случайных величин $\xi_1, \xi_2, \ldots$ сходится к случайной величине ξ *в среднем порядка* p, $p > 0$, и писать $\xi_n \xrightarrow{L_p} \xi$, если $\mathsf{E}|\xi_n - \xi|^p \to 0$ при $n \to \infty$.

При $p = 2$ эта сходимость называется также *в среднем квадратичном.*

Определение 10.4. Будем говорить, что последовательность случайных величин $\xi_1, \xi_2, \ldots$ сходится к случайной величине ξ *по распределению*, и писать $\xi_n \xrightarrow{d} \xi$, если $F_{\xi_n} \Rightarrow F_\xi$.

Теорема 10.1. *Пусть $\xi_1, \xi_2, \ldots$ — последовательность случайных величин. Тогда имеют место следующие утверждения:*

(i) Если $\xi_n \xrightarrow{\text{п.н.}} \xi$, то $\xi_n \xrightarrow{\mathsf{P}} \xi$;

(ii) Если $\xi_n \xrightarrow{L_p} \xi$, то $\xi_n \xrightarrow{\mathsf{P}} \xi$;

(iii) Если $\xi_n \xrightarrow{\mathsf{P}} \xi$, то $\xi_n \xrightarrow{d} \xi$.

Доказательство. (i) Для $\varepsilon > 0$ введем в рассмотрение последовательность событий

$$A_n^\varepsilon \,=\, \{\omega\colon\, |\xi_n(\omega) - \xi(\omega)| \geqslant \varepsilon\},$$

$n = 1, 2, \ldots$. Заметим, что условие $\xi_n \xrightarrow{\mathsf{P}} \xi$ эквивалентно тому, что для любого $\varepsilon > 0$ должно выполняться предельное соотношение $\mathsf{P}(A_n^\varepsilon) \to 0$ при $n \to \infty$.

Также для каждого $\varepsilon > 0$ и $n = 1, 2, \ldots$ рассмотрим

$$B_n^\varepsilon \,=\, \left\{\omega\colon\, \sup_{k\geqslant n} |\xi_n(\omega) - \xi(\omega)| \geqslant \varepsilon\right\}.$$

Очевидно, что $A_n^\varepsilon \subset B_n^\varepsilon$ и $\{B_n^\varepsilon\}_{n=1}^\infty$ является монотонно убывающей последовательностью. Пусть $B^\varepsilon = \bigcap_{n=1}^\infty B_n^\varepsilon$ — предел этой последовательности. В силу свойства непрерывности вероятностной меры $\mathsf{P}(B_n^\varepsilon) \to \mathsf{P}(B^\varepsilon)$ при $n \to \infty$. С другой стороны, если $\omega \in B^\varepsilon$, то $\xi_n(\omega) \not\to \xi(\omega)$. Это означает, что

$$B^\varepsilon \subset \{\omega\colon\ \xi_n(\omega) \not\to \xi(\omega)\}.$$

Однако, из условия $\xi_n \overset{\text{п.н.}}{\to} \xi$ следует

$$\mathsf{P}(\{\omega\colon\ \xi_n(\omega) \not\to \xi(\omega)\}) = 0.$$

Но тогда

$$\lim_{n\to\infty} \mathsf{P}(B_n^\varepsilon) = \mathsf{P}(B^\varepsilon) = 0.$$

Из неравенств

$$0 \leqslant \mathsf{P}(A_n^\varepsilon) \leqslant \mathsf{P}(B_n^\varepsilon)$$

следует, что и $\mathsf{P}(A_n^\varepsilon) \to 0$ при $n \to \infty$, т. е. $\xi_n \overset{\mathsf{P}}{\to} \xi$.

(ii) Допустим теперь,что $\xi_n \overset{L_p}{\to} \xi$, т. е. $\mathsf{E}|\xi_n - \xi|^p \to 0$ при $n \to \infty$. Фиксируем произвольно $\varepsilon > 0$. Используя неравенство (8.1) Маркова, получаем

$$\mathsf{P}(|\xi_n - \xi| \geqslant \varepsilon) = \mathsf{P}(|\xi_n - \xi|^p \geqslant \varepsilon^p) \leqslant \frac{\mathsf{E}|\xi_n - \xi|^p}{\varepsilon^p} \to 0$$

при $n \to \infty$, т. е. $\xi_n \overset{\mathsf{P}}{\to} \xi$.

(iii) Пусть $\xi_n \overset{\mathsf{P}}{\to} \xi$, т. е. $\mathsf{P}(A_n^\varepsilon) \to 0$ при $n \to \infty$ для всех $\varepsilon > 0$. Обозначим $F_n = F_{\xi_n}$ и $F = F_\xi$. Доказательство $F_n \Rightarrow F$ будет следовать из того, что для всех $\varepsilon > 0$ выполняются неравенства

$$F(x-\varepsilon) \leqslant \varliminf_{n\to\infty} F_n(x) \leqslant \varlimsup_{n\to\infty} F_n(x) \leqslant F(x+\varepsilon), \qquad x \in \mathbb{R}. \tag{10.1}$$

Эти неравенства, в свою очередь, сразу же следуют из включений

$$\{\xi \leqslant x - \varepsilon\} \subset \{\xi_n \leqslant x\} \cup A_n^\varepsilon, \qquad \{\xi_n \leqslant x\} \subset \{\xi \leqslant x + \varepsilon\} \cup A_n^\varepsilon$$

и того, что $\mathsf{P}(A_n^\varepsilon) \to 0$ при $n \to \infty$. В точках непрерывности функции F предельный переход в неравенствах (10.1) при $\varepsilon \to 0$ дает

$$F(x) \leqslant \varliminf_{n\to\infty} F_n(x) \leqslant \varlimsup_{n\to\infty} F_n(x) \leqslant F(x),$$

что означает $F_n(x) \to F(x)$ при $n \to \infty$ и теорема доказана. □

Следующие два примера показывают, что из сходимости последовательности случайных величин почти наверное не следует сходимость в среднем, а из сходимости в среднем не следует сходимость почти наверное.

Пример 1. Пусть $\Omega = [0,1]$, $\mathscr{A} = \mathscr{B}([0,1])$, P — мера Лебега. Рассмотрим последовательность

$$\xi_n(\omega) = n^{1/p} \cdot \mathbb{1}_{[0,1/n]}(\omega),$$

$n = 1, 2, \ldots$. Тогда $\xi_n \overset{\text{п.н.}}{\to} \xi$, $\xi(\omega) \equiv 0$, но $\xi_n \not\to \xi$ в L_p, $p > 0$.

Пример 2. Пусть $\Omega = \mathbb{T}$ — единичная окружность, $\mathscr{A} = \mathscr{B}(\mathbb{T})$, $\mathsf{P} = \mu/2\pi$ — нормированная мера Лебега. Для $n = 1, 2, \ldots$ рассмотрим

$$E_n = \left\{ \omega = e^{i\theta} \colon \sum_{k=1}^{n-1} \frac{1}{k} \leqslant \theta < \sum_{k=1}^{n} \frac{1}{k} \right\} \subset \mathbb{T}.$$

Заметим, что $\mathsf{P}(E_n) = 1/(2\pi n) \to 0$ при $n \to \infty$. Следовательно, для любого $p > 0$ последовательность $\xi_n = \mathbb{1}_{E_n}$ сходится в среднем порядка p и $\xi(\omega) \equiv 0$.

С другой стороны, поскольку $\sum_{k=1}^{\infty} 1/k = \infty$, то каждое элементарное событие ω принадлежит бесконечному числу событий из последовательности $\{E_n\}$, т. е. $\overline{\lim}_{n\to\infty} E_n = \Omega$. Это означает, что $\xi_n(\omega) \not\to \xi(\omega)$ ни при каком $\omega \in \Omega$, т, е. последовательность ξ_n не сходится почти наверное.

Закон больших чисел Чебышёва касался последовательности независимых случайных величин с равномерно ограниченными дисперсиями. Метод характеристических функций позволяет снять условие конечности вторых моментов для одинаково распределенных случайных величин.

Теорема 10.2 (Хинчина). *Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин с конечным $\mathsf{E}\xi = a$. Тогда для $S_n = \xi_1 + \ldots + \xi_n$ имеет место предельное соотношение*

$$\frac{S_n}{n} \xrightarrow{\mathsf{P}} a \qquad \text{при} \qquad n \to \infty.$$

Доказательство. Поскольку все ξ_k, $k = 1, 2, \ldots$, одинаково распределены, то они имеют одну и ту же характеристическую функцию

$$h(t) = h_{\xi_k}(t) = \mathsf{E} e^{it\xi_k}.$$

В силу того, что ξ_k имеет конечное математическое ожидание, характеристическая функция h дифференцируема и

$$h(t) = 1 + iat + o(t) \qquad \text{при} \qquad t \to 0.$$

По условию $\xi_1, \ldots, \xi_n$ независимы. Поэтому

$$h_{S_n}(t) = (h(t))^n$$

и

$$h_{\frac{S_n}{n}}(t) = \left(h\left(\frac{t}{n}\right)\right)^n = \left(1 + ia\frac{t}{n} + o\left(\frac{t}{n}\right)\right)^n.$$

Используя асимптотическую формулу $\ln(1+z) = z + o(|z|)$, при $z \to 0$, получаем

$$\lim_{n\to\infty} h_{\frac{S_n}{n}}(t) = \lim_{n\to\infty} e^{n \ln\left(1 + \frac{iat}{n} + o\left(\frac{t}{n}\right)\right)} = e^{iat}.$$

Предельная функция e^{iat} непрерывна и по теореме сходимости она является характеристической функцией некоторой случайной величины η, причем

$S_n/n \xrightarrow{d} \eta$. Заметим также, что $h_\eta(t) = e^{iat}$ является характеристической функцией вырожденного распределения, сосредоточенного в точке a, т. е.

$$F_\eta(x) = \mathbb{1}_{[a,\infty)}(x).$$

Остается показать, что $S_n/n \xrightarrow{\mathsf{P}} \eta$. Для $\varepsilon > 0$ имеем

$$\begin{aligned}\mathsf{P}\left(\left|\frac{S_n}{n} - a\right| \geqslant \varepsilon\right) &= \mathsf{P}\left(\frac{S_n}{n} \leqslant a - \varepsilon\right) + \mathsf{P}\left(\frac{S_n}{n} \geqslant a + \varepsilon\right) \\ &\leqslant F_n(a-\varepsilon) + \left(1 - F_n\left(a + \frac{\varepsilon}{2}\right)\right),\end{aligned}$$

где F_n — функция распределения случайной величины S_n/n. Поскольку точки $a-\varepsilon$ и $a+\varepsilon/2$ являются точками непрерывности функции F_η, то

$$F_n(a-\varepsilon) \to 0, \qquad F_n(a+\varepsilon/2) \to 1$$

при $n \to \infty$. Отсюда следует, что

$$\lim_{n\to\infty} \mathsf{P}\left(\left|\frac{S_n}{n} - a\right| \geqslant \varepsilon\right) = 0$$

и $S_n/n \xrightarrow{\mathsf{P}} a$. Теорема доказана. $\square$