В. В. Горяйнов

Лекции по теории вероятностей

Содержание

§ 1. Случайный эксперимент и аксиоматика Колмогорова

Понятия случайность и вероятность естественно возникают в повседневной жизни. Мы с ними сталкиваемся, когда играем в азартные игры, участвуем в розыгрыше лотереи или покупаем акции. При этом нам приходится оценивать вероятность наступления тех или иных событий.

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. Мы 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) события A будет стремиться к некоторой величине. Это позволяет предположить, что событие A обладает некоторой скрытой характеристикой, которую называют вероятностью и обозначают $\mathsf{P}(A)$.

Чтобы построить математическую модель конкретного случайного эксперимента, нужно пройти три этапа: определить пространство элементарных событий Ω, выделить в нем класс $\mathscr{F}$ событий (подмножеств Ω), определить на выделенном классе подмножеств вероятность $\mathsf{P}(A)$, $A \in \mathscr{F}$.

1.1. Пространство элементарных событий. В результате проведения случайного эксперимента может наблюдаться только одно элементарное событие (исход). Совокупность всех возможных элементарных событий образует множество Ω, которое называют *пространством элементарных событий.* Элементы этого множества (*элементарные события*) обозначают ω. Обычно нас интересует не то, какое именно элементарное событие $\omega \in \Omega$ произойдет, а появление некоторого подмножества $A \subset \Omega$, которое мы будем называть *событием.* Если исход ω принадлежит A, т. е. $\omega \in A$, то говорят, что событие A произошло, если же $\omega \notin A$, то событие A в этом случайном 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, т. е. из A следует 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 $A \cap B = AB$ соответствует событие, которое происходит лишь в случае, когда происходят одновременно и A и B. Аналогичный смысл имеют объединение и пересечение любого числа событий (подмножеств Ω). Разность $A \setminus B = A\overline{B}$ означает событие, когда A произошло, но B в этом эксперименте не произошло. Само пространство элементарных событий Ω также рассматривается как событие. Его называют *достоверным* событием, поскольку оно происходит всегда. Наряду с этим пустое множество $\varnothing$ связывают с *невозможным* событием, которое никогда не происходит.

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:

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

Прежде, чем мы приступим к обсуждению общей ситуации, рассмотрим более простой частный случай, когда вопрос класса $\mathscr{F}$ и вероятности $\mathsf{P}(A)$, $A \in \mathscr{F}$, решается сравнительно просто.

1.2. Дискретная модель. Если Ω конечно или счетно, то говорят о дискретной модели теории вероятностей (или случайного эксперимента). Рассмотрим этот случай, т. е. допустим, что $\Omega = \{\omega_1, \omega_2, \ldots\}$ конечно или счетно. В этом случае в качестве $\mathscr{F}$ можно рассмотреть совокупность всех подмножеств Ω. Множество всех подмножеств Ω часто обозначают символом 2^{Ω}.

Допустим, что каждому элементарному событию ω_k соответствует вероятность p_k. Очевидно, что числа p_k, $k = 1, 2, \ldots$, должны быть неотрицательными и их сумма $\sum p_k$ должна равняться 1. Такой набор чисел p_k называют *распределением вероятностей.* Вероятность любого события $A \subset \Omega$ можно теперь определить равенством

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

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) можно считать равновероятными. 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

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

которая и составляет классическое определение вероятности.

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 n 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) можно получить, поделив A_n^k на $k!$, число способов упорядочить k элементов. В результате получаем

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

♣ 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	$A_n^k = \frac{n!}{(n-k)!}$
неупорядоченные	$C_{n+k-1}^k = \frac{(n+k-1)!}{k!(n-1)!}$	$C_n^k = \frac{n!}{k!(n-k)!}$

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 Лебега множества A, хотя в приложениях можно считать, что это мера Жордана.

1.4. Вероятностное пространство. Вернемся теперь к общему случаю выделения класса событий $\mathscr{F}$, т. е. подмножеств Ω, и определения вероятностной меры P. Как показывает пример задачи о встрече, мы не можем в общем случае определить $\mathscr{F}$ и P, как это было сделано в дискретной модели. Кроме того, теоретически доказано, что множество всех подмножеств 2^Ω слишком велико, чтобы его можно было взять в качестве $\mathscr{F}$. Это, например, показывает результат Витали (1905 г.) для случайного эксперимента с бесконечным подбрасыванием монеты. Упомянем также результат Банаха и Куратовского (1929 г.), согласно которому не существует вероятностной меры P, определенной на всех подмножествах отрезка $[0,1]$ и удовлетворяющей условию $\mathsf{P}(\{x\}) = 0$ для каждого $x \in [0,1]$.

Пусть теперь Ω — некоторое непустое множество, которое мы будем рассматривать как пространство элементарных событий. Класс $\mathscr{F}$ подмножеств Ω, которые мы будем называть событиями, должен быть замкнут относительно теоретико–множественных операций, поскольку они соответствуют логическим операциям над событиями.

Определение 1.1. Класс $\mathscr{A}$ подмножеств Ω называется *алгеброй* подмножеств Ω, если 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 $A \cup B \in \mathscr{A}$.

Условия (i)–(iii) достаточны для того, чтобы класс $\mathscr{A}$ был замкнут относительно всех теоретико–множественных операций. Из условий (ii) и (iii) следует, что он замкнут относительно объединения и операции взятия дополнения. Покажем, что он замкнут и 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 $\mathscr{A}$.

В случае бесконечного пространства элементарных событий Ω приходится рассматривать последовательности $(A_n\colon n \in \mathbb{N})$ подмножеств Ω. Если $A_1, A_2, \ldots$ рассматриваются как события, то естественно возникает событие A^*, которое состоит в том, что произошло бесконечно много событий последовательности $(A_n\colon n \in \mathbb{N})$, т. е. $\omega \in A^*$ в том и только том случае, если найдется подпоследовательность $(A_{n_k}\colon k \in \mathbb{N})$ такая, что $\omega \in A_{n_k}$ для всех $k = 1, 2, \ldots$. Это событие A^* называют верхним пределом последовательности $(A_n\colon n \in \mathbb{N})$ и обозначают $\overline{\lim\limits_n} A_n$. Кроме того, рассматривают также событие A_*, которое состоит в том, что произошли все события последовательности $(A_n\colon n \in \mathbb{N})$ за исключением, быть может, конечного числа, т. е. $\omega \in A_*$ в том и только том случае, если $\omega \in \cap_{k \geqslant N} A_k$ при некотором N. Событие A_* называют нижним пределом последовательности $(A_n\colon n \in \mathbb{N})$ и обозначают $\underline{\lim\limits_n} A_n$. Заметим, что

$$A^* = \overline{\lim_n} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k \geqslant n} A_k, \qquad A_* = \underline{\lim_n} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k \geqslant n} A_k.$$

В случае $A^* = A_*$ говорят, что последовательность $(A_n\colon n \in \mathbb{N})$ сходится и через $\lim\limits_n A_n$ обозначают ее верхний и нижний пределы, которые в данной ситуации совпадают.

Таким образом, к условию замкнутости $\mathscr{F}$ относительно теоретико–множественных операций нужно добавить замкнутость относительно счетного числа операций.

Определение 1.2. Класс $\mathscr{F}$ подмножеств Ω называется σ-алгеброй, если он является алгеброй и для любой последовательности $(A_n\colon n \in \mathbb{N}) \subset \mathscr{F}$ объединение $\cup_{n=1}^{\infty} A_n$ принадлежит $\mathscr{F}$.

Другими словами, σ-алгебра подмножеств Ω замкнута относительно счетного числа теоретико–множественных операций. Действительно, замкнутость относительно счетных объединений содержится в определении, а замкнутость относительно счетных пересечений выводится, как и в случае доказательства замкнутости относительно пересечений алгебры, с использованием правил де Моргана.

Определение 1.3. Пусть Ω — некоторое непустое множество и $\mathscr{F}$ — σ-алгебра его подмножеств. Пара $(\Omega, \mathscr{F})$ называется *измеримым пространством.*

Для завершения аксиоматического построения теории вероятностей остается дополнить измеримое пространство вероятностной мерой. При формировании аксиом вероятности естественно проанализировать свойства частоты появления события $A \in \mathscr{F}$. При этом возникает вопрос минимальности постулируемых свойств.

Определение 1.4. Пусть $(\Omega, \mathscr{F})$ — измеримое пространство. Неотрицательная функция $\mathsf{P}\colon \mathscr{F} \to \mathbb{R}_+$ называется вероятностной мерой (вероятностью), если выполняются следующие условия:

(i) $\mathsf{P}(\Omega) = 1$ (нормированность);

(ii) если $A_1, A_2, \ldots$ — принадлежат $\mathscr{F}$ и $A_i A_j = \varnothing$ при $i \neq j$, то $\mathsf{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathsf{P}(A_n)$ (счетная аддитивность).

Тройка $(\Omega, \mathscr{F}, \mathsf{P})$ называется *вероятностным пространством.*

Понятие вероятностного пространства и является математической моделью случайного эксперимента. Это так называемая аксиоматика Колмогорова теории вероятностей. Из постулированных свойств вероятности выводится ряд других важных свойств.

Предложение 1.1. *Пусть P — вероятностная мера, определенная на измеримом пространстве $(\Omega, \mathscr{F})$. Тогда имеют место утверждения:*

1. *$\mathsf{P}(\varnothing) = 0$ и для $A, B \in \mathscr{F}$, $AB = \varnothing$, имеет место равенство $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B)$ (конечная аддитивность);*
2. *Если $A, B \in \mathscr{F}$ и $A \subset B$, то $\mathsf{P}(B \setminus A) = \mathsf{P}(B) - \mathsf{P}(A)$ и $\mathsf{P}(A) \leqslant \mathsf{P}(B)$ (монотонность);*
3. *Для произвольных $A_1, \ldots, A_n$ из $\mathscr{F}$ выполняется неравенство*

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

(полуаддитивность).

Доказательство. Рассмотрим последовательность событий $(A_n\colon n \in \mathbb{N})$, где все $A_n = \varnothing$. По свойству счетной аддитивности $\mathsf{P}(\varnothing) = \sum_{n=1}^{\infty} \mathsf{P}(\varnothing)$, откуда следует, что $\mathsf{P}(\varnothing) = 0$. В случае $A, B \in \mathscr{F}$ и $AB = \varnothing$ применение свойства счетной аддитивности к последовательности $A_1 = A$, $A_2 = B$, $A_k = \varnothing$, $k = 3, 4, \ldots$, приводит к свойству конечной аддитивности вероятности.

Если $A, B \in \mathscr{A}$ и $A \subset B$, то $B = A \cup (B \setminus A)$ и в силу свойства конечной аддитивности получаем $\mathsf{P}(B) = \mathsf{P}(A) + \mathsf{P}(B \setminus A)$, откуда следует, что $\mathsf{P}(B \setminus A) = \mathsf{P}(B) - \mathsf{P}(A)$ и $\mathsf{P}(A) \leqslant \mathsf{P}(B)$.

Для доказательства третьего утверждения рассмотрим события $B_1 = A_1$, $B_2 = A_2\overline{A_1}, \ldots, B_n = A_n\overline{A_1} \cdot \ldots \cdot \overline{A_{n-1}}$. Очевидно, что $B_iB_j = \varnothing$ 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 $\mathsf{P}(B_i) \leqslant \mathsf{P}(A_i)$, $i = 1, \ldots, n$. Используя эти неравенства и свойство аддитивности, получаем

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

□

Теорема 1.1 (Сложения). *Пусть $A_1, \ldots, A_n$ — произвольные события, т. е. принадлежат $\mathscr{F}$. 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. Рассмотрим вначале случай $n = 2$. Из представления

$$A_1 \cup A_2 = (A_1 \setminus (A_1A_2)) \cup (A_2 \setminus (A_1A_2)) \cup (A_1A_2)$$

и свойств вероятности следует

$$\begin{aligned} \mathsf{P}(A_1 \cup A_2) &= \mathsf{P}(A_1) - \mathsf{P}(A_1A_2) + \mathsf{P}(A_2) - \mathsf{P}(A_1A_2) + \mathsf{P}(A_1A_2) \\ &= \mathsf{P}(A_1) + \mathsf{P}(A_2) - \mathsf{P}(A_1A_2). \end{aligned}$$

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 доказанное равенство, а затем предположение индукции, получаем

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

Теорема доказана. □

Отметим еще одно свойство вероятности, которое называется непрерывностью меры. Допустим, что $(A_n\colon n \in \mathbb{N})$ — монотонно возрастающая последовательность множеств, т. е. $A_1 \subseteq A_2 \subseteq \ldots$. Тогда для всех $n = 1, 2, \ldots$ выполняются равенства

$$\bigcap_{k\geqslant n} A_k = A_n, \qquad \bigcup_{k\geqslant n} = \bigcup_{n=1}^{\infty} A_n$$

и, следовательно,

$$\overline{\lim_n} A_n = \underline{\lim_n} A_n = \lim_n A_n = \bigcup_{n=1}^{\infty} A_n.$$

Если $(A_n\colon n \in \mathbb{N})$ — монотонно убывающая последовательность, т. е. $A_1 \supseteq A_2 \supseteq \ldots$, то

$$\bigcap_{k\geqslant n} A_k = \bigcap_{n=1}^{\infty} A_n, \qquad \bigcup_{k\geqslant n} A_k = A_n$$

при всех $n = 1, 2, \ldots$ и

$$\overline{\lim_n} A_n = \underline{\lim_n} A_n = \lim_n A_n = \bigcap_{n=1}^{\infty} A_n.$$

Теорема 1.2. *Пусть последовательность* $(A_n\colon n \in \mathbb{N}) \subset \mathscr{F}$ *является монотонной. Тогда*

$$\mathsf{P}(\lim_n A_n) = \lim_{n\to\infty} \mathsf{P}(A_n).$$

Доказательство. Допустим вначале, что последовательность $(A_n\colon n \in \mathbb{N})$ является монотонно возрастающей. Рассмотрим новую последовательность событий $B_1 = A_1$, $B_n = A_n\overline{A_{n-1}}$, $n = 2, 3, \ldots$. Заметим, что $B_iB_j = \varnothing$ при $i \neq j$ и для каждого $n = 1, 2, \ldots$ выполняются равенства

$$\bigcup_{k=1}^{n} B_k = \bigcup_{k=1}^{n} A_k = A_n.$$

Кроме того,

$$\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n = \lim_n A_n.$$

Поэтому в силу счетной аддитивности вероятности имеем

$$\mathsf{P}(\lim_n A_n) = \sum_{n=1}^{\infty} \mathsf{P}(B_n) = \lim_{n\to\infty} \sum_{k=1}^{n} \mathsf{P}(B_k) = \lim_{n\to\infty} \mathsf{P}(A_n).$$

Если $(A_n\colon n \in \mathbb{N})$ является монотонно убывающей последовательностью, то $(\overline{A_n}\colon n \in \mathbb{N})$ монотонно возрастает и по доказанному

$$\mathsf{P}\left(\bigcup_{n=1}^{\infty} \overline{A_n}\right) = \lim_{n\to\infty} \mathsf{P}(\overline{A_n}) = 1 - \lim_{n\to\infty} \mathsf{P}(A_n).$$

Но тогда, используя правила де Моргана, получаем

$$\lim_{n\to\infty} \mathsf{P}(A_n) = 1 - \mathsf{P}\left(\bigcup_{n=1}^{\infty} \overline{A_n}\right) = \mathsf{P}\left(\overline{\bigcup_{n=1}^{\infty} \overline{A_n}}\right) = \mathsf{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \mathsf{P}(\lim_n A_n).$$

Теорема доказана. □

Заметим также, что конечная аддитивность меры и свойство непрерывности влекут ее счетную аддитивность.

§ 2. Условные вероятности и независимость событий

Понятие условной вероятности имеет важное значение. Во-первых, часто требуется переоценить исходные вероятности при наличии некоторой информации. Во-вторых, условные вероятности используются даже при отсутствии дополнительной информации для более простого вычисления исходных вероятностей.

2.1. Условные вероятности. Наряду с основной вероятностной мерой P часто возникает дополнительная мера, которая обусловлена наличием частичной информации об исходе случайного эксперимента. Допустим, что A, B — два события и A произошло. При наличии информации о наступлении события A вероятность события B следует изменить, поскольку в этом случае событие B может произойти, если только произойдет AB. Поэтому исправленная вероятность должна быть пропорциональна величине $\mathsf{P}(AB)$.

Определение 2.1. Пусть $(\Omega, \mathscr{F}, \mathsf{P})$ — вероятностное пространство и событие A имеет положительную вероятность, $\mathsf{P}(A) \neq 0$. Тогда для каждого B из $\mathscr{F}$ под *условной вероятностью* этого события при условии A будем понимать величину

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

Заметим, что коэффициент пропорциональности обеспечивает естественное равенство $\mathsf{P}(A|A) = 1$.

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

На практике чаще легче бывает вычислить условную вероятность $\mathsf{P}(B|A)$ и с ее помощью получить вероятность совместного наступления этих событий, т. е. воспользоваться равенством

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

Это равенство известно как теорема умножения и имеет следующее обобщение.

Теорема 2.1. *Пусть события* $A_1, \ldots, A_n$ *таковы, что* $\mathsf{P}(A_1 \cdot \ldots \cdot A_n) > 0$. *Тогда*

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

Доказательство. Заметим прежде всего, что все условные вероятности в доказываемом равенстве корректно определены. Это обеспечивает условие $\mathsf{P}(A_1 \cdot \ldots \cdot A_n) > 0$ и то, что $A_1 \cdot \ldots \cdot A_n$ является подмножеством каждого условия. При $n = 2$ доказываемое равенство следует, как отмечалось выше, непосредственно из определения условных вероятностей. Допустим теперь, что это равенство выполняется для всех номеров, вплоть до $n-1$. Тогда, применяя его для двух событий $A = A_1, \ldots, A_{n-1}$ и $B = A_n$ с последующим использованием предположения 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 2.2. *Пусть $H_1, \ldots, H_n$ — события, удовлетворяющие следующим условиям $H_i H_j = \varnothing$ при $i \neq j$, $\mathsf{P}(H_i) > 0$, $i = 1, \ldots, n$, и $\sum_{i=1}^n \mathsf{P}(H_i) = 1$. Тогда для любого события A имеет место равенство*

$$\mathsf{P}(A) = \sum_{i=1}^{n} \mathsf{P}(H_i)\mathsf{P}(A|H_i). \tag{2.1}$$

Доказательство. Введем в рассмотрение также событие $H_0 = \Omega \setminus \cup_{i=1}^n H_i$. Очевидно, что $\mathsf{P}(H_0) = 0$ и $\mathscr{D} = \{H_0, H_1, \ldots, H_n\}$ 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 (2.1). □

Формула (2.1) является эффективным инструментом при вычислении вероятностей сложных событий. При этом события $H_1, \ldots, H_n$ в этой формуле обычно называют *гипотезами*. Иногда вместо условия $\sum_{i=1}^n \mathsf{P}(H_i) = 1$ требуют, чтобы $\cup_{i=1}^n H_i = \Omega$, и совокупность $H_1, \ldots, H_n$ называют *полной группой событий*.

Теорема 2.3 (Формула Байеса). *Пусть события $H_1, \ldots, H_n$ удовлетворяют условиям теоремы 2.2 и $A \in \mathscr{F}$, $\mathsf{P}(A) > 0$. Тогда для $k = 1, \ldots, n$ 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

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

Остается знаменатель записать по формуле полной вероятности (2.1) 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.

Наивная интерпретация условных вероятностей может привести к ошибочным выводам о причинно–следственных связях. При работе с условными вероятностями очень важно понимать то, в каком виде мы получили дополнительную информацию для переоценки вероятностей. Рассмотрим два примера.

Пример 1. Рассмотрим семью с двумя детьми. Известно, что один из них мальчик. Считается, что рождение мальчика и девочки равновероятно. Какова вероятность того, что второй ребёнок тоже мальчик?

Решение. В данном случае $\Omega = \{(\text{м,м}),\ (\text{м,д}),\ (\text{д,м}),\ (\text{д,д})\}$ известная информация $A = \{(\text{м,м}),\ (\text{м,д}),\ (\text{д,м})\}$. Нас интересует вероятность события $B = \{(\text{м,м})\}$ при условии A. Поскольку $|\Omega| = 4$, $|A| = 3$, $|AB| = |B| = 1$, то $\mathsf{P}(B|A) = 1/3$.

Пример 2. Известно, что у коллеги по работе двое детей. Мы встречаем его на прогулке с мальчиком, который представлен его сыном. Какова вероятность, что второй ребёнок коллеги тоже мальчик?

Решение. Изначально мы имеем три гипотезы: H_1 — в семье коллеги два сына, H_2 — сын и дочь, H_3 — две дочери. Событие A — коллега вышел на прогулку с сыном. В этом случае нас интересует условная вероятность $\mathsf{P}(H_1|A)$. Замечая, что $\mathsf{P}(H_1) = \mathsf{P}(H_3) = 1/4$, $\mathsf{P}(H_2) = 1/2$, $\mathsf{P}(A|H_1) = 1$, $\mathsf{P}(A|H_2) = 1/2$, $\mathsf{P}(A|H_3) = 0$, по формуле Байеса получаем

$$\mathsf{P}(H_1|A) = \frac{1/4\cdot 1}{1/4\cdot 1 + 1/2\cdot 1/2 + 0} = \frac{1}{2}.$$

2.2. 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

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

которое само является следствием независимости. Однако, требование выполнения равенства $\mathsf{P}(AB) = \mathsf{P}(A)\mathsf{P}(B)$ не предполагает условий ни $\mathsf{P}(A) > 0$, ни $\mathsf{P}(B) > 0$.

Определение 2.2. Для произвольных двух событий $A, B \in \mathscr{F}$ будем говорить, что они независимы, если выполняется равенство

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

Условие независимости событий $A_1, \ldots, A_n$ при $n > 2$ выглядит сложнее. С точки зрения предыдущих рассуждений можно сказать, что информация о наступлении некоторых событий из этой совокупности не влияет на переоценку вероятностей остальных событий.

Определение 2.3. Будем говорить, что события $A_1, \ldots, A_n$ независимы (независимы в совокупности), если

$$\mathsf{P}\left(\bigcap_{i\in I} A_i\right) = \prod_{i\in I} \mathsf{P}(A_i) \tag{2.2}$$

для любого подмножества $I \subseteq \{1, 2, \ldots, n\}$.

Заметим, что из попарной независимости событий не следует, вообще говоря, независимость в совокупности, а условие (2.2) связано с проверкой выполнения $\sum_{k=2}^{n} C_n^k = 2^n - n - 1$ равенств. Кроме того, при проверке независимости в совокупности не достаточно ограничиться проверкой лишь самых длинных цепочек в равенстве (2.2). 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

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

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.

Следующий результат показывает, что замена одного или нескольких из событий A_k на отрицание $\overline{A_k}$ не нарушает независимости.

Предложение 2.1. *Пусть события $A_1, \ldots, A_n$ независимы (в совокупности). Тогда независимы также события $\overline{A_1}, A_2, \ldots, A_n$.*

Доказательство. Пусть $I \subseteq \{2, 3, \ldots, n\}$ и $H = \cap_{i\in I} A_i$. Тогда, используя свойства вероятности и условие независимости событий $A_1, \ldots, A_n$, получаем

$$\begin{aligned}\mathsf{P}(\overline{A_1}H) &= \mathsf{P}(H \setminus HA_1) = \mathsf{P}(H) - \mathsf{P}(HA_1)\\ &= \mathsf{P}(H)(1 - \mathsf{P}(A_1)) = \mathsf{P}(H)\mathsf{P}(\overline{A_1}), = \mathsf{P}(\overline{A_1})\prod_{i\in I}\mathsf{P}(A_i),\end{aligned}$$

что означает выполнение условий (2.2) для событий $\overline{A_1}, A_2, \ldots, A_n$. □

Определение 2.4. Будем говорить, что бесконечное семейство событий независимо, если любое его конечное подсемейство независимо.

Следующий результат известен как *лемма Бореля–Кантелли*, которая утверждает, что для последовательности событий в зависимости от условий верхний предел имеет вероятность либо 0, либо 1.

Теорема 2.4. *Пусть $(A_n\colon n \in \mathbb{N})$ — последовательность событий. Тогда имеют место следующие утверждения:*

(i) Если $\sum_{n=1}^{\infty}\mathsf{P}(A_n) < \infty$, то $\mathsf{P}(\overline{\lim_n} A_n) = 0$;

(ii) Если $\sum_{n=1}^{\infty}\mathsf{P}(A_n) = +\infty$ и последовательность $(A_n\colon n \in \mathbb{N})$ независима, то $\mathsf{P}(\overline{\lim_n} A_n) = 1$.

Доказательство. (i) Пусть $B_n = \cup_{k\geqslant n} A_k$, $n = 1, 2, \ldots$. Тогда $(B_n\colon n \in \mathbb{N})$ является монотонно убывающей последовательностью событий и

$$\bigcap_{n=1}^{\infty} B_n = \lim B_n = \overline{\lim_n} A_n.$$

Используя свойство непрерывности вероятностной меры и свойство полуаддитивности, получаем

$$\mathsf{P}(\overline{\lim_n} A_n) = \lim_{n\to\infty}\mathsf{P}(B_n) \leqslant \lim_{n\to\infty}\sum_{k\geqslant n}\mathsf{P}(A_k) = 0.$$

(ii) Заметим вначале, что

$$\overline{\overline{\lim_n} A_n} = \underline{\lim_n}\, \overline{A_n} = \bigcup_{n=1}^{\infty} C_n,$$

где $C_n = \cap_{k\geqslant n}\overline{A_k}$. Из условия независимости последовательности $(A_n\colon n \in \mathbb{N})$ следует, в силу предложения 2.1, независимость последовательности $(\overline{A_n}\colon n \in \mathbb{N})$. Замечая, что при всех $m = 1, 2, \ldots$ имеет место включение $C_n \subseteq \cap_{k=n}^{n+m}\overline{A_k}$, в силу независимости $(\overline{A_n}\colon n \in \mathbb{N})$, получаем

$$\mathsf{P}(C_n) \leqslant \mathsf{P}\left(\bigcap_{k=n}^{n+m}\overline{A_k}\right) = \prod_{k=n}^{n+m}\mathsf{P}(\overline{A_k}) = \prod_{k=n}^{n+m}(1 - \mathsf{P}(A_k)).$$

Используя теперь неравенство $1 - x \leqslant e^{-x}$, $x > 0$, выводим

$$\mathsf{P}(C_n) \leqslant \exp\left\{-\sum_{k=n}^{n+m}\mathsf{P}(A_k)\right\}$$

для всех $m = 1, 2, \ldots$. Однако, правая часть последнего неравенства стремится к нулю при $m \to \infty$. Это означает, что $\mathsf{P}(C_n) = 0$ для всех n, и

$$\mathsf{P}\left(\underline{\lim}_n \overline{A_n}\right) \;=\; \lim_{n\to\infty} \mathsf{P}(C_n) \;=\; 0.$$

Отсюда получаем

$$\mathsf{P}\left(\overline{\lim}_n A_n\right) \;=\; 1 - \mathsf{P}\left(\underline{\lim}_n \overline{A_n}\right) \;=\; 1.$$

□

2.3. Независимые испытания. *Схема Бернулли* представляет собой простую математическую модель независимых испытаний, которая широко используется в приложениях. 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 $(\Omega, \mathscr{F}, \mathsf{P})$, 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

$$\mathsf{P}(\{\omega\}) \;=\; p^{\sum_{i=1}^n \delta_i} q^{n - \sum_{i=1}^n \delta_i}. \tag{2.3}$$

Если $\omega \in B_n(k)$, то из (2.3) 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

$$\mathsf{P}(B_n(k)) \;=\; C_n^k p^k q^{n-k}. \tag{2.4}$$

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

При больших значениях n реализация формулы (2.4) 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 2.5 (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

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

и предельного соотношения

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

следует утверждение теоремы. □

Заметим, что в условиях теоремы Пуассона 2.5 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 2.6 (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 2.7 (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.

§ 3. Дискретные случайные величины

Часто с результатом случайного эксперимента связывают число. В связи с этим возникает важное понятие случайной величины. В этом параграфе мы рассмотрим так называемый дискретный случай, когда случайная величина принимает конечное или счётное число значений. Позже некоторые определения этого параграфа будут сформулированы несколько иначе при снятии ограничения конечности множества множества значений случайной величины, но они будут эквивалентными приведенным здесь для дискретных случайных величин.

Определение 3.1. Под *дискретной случайной величиной*, определённой на вероятностном пространстве $(\Omega,\mathscr{F},\mathsf{P})$, будем понимать отображение $\xi\colon\ \Omega\to\mathbb{R}$, удовлетворяющее условиям: $\xi(\Omega)$ — конечное или счётное множество в $\mathbb{R}$ и для каждого $x\in\mathbb{R}$ множество $\{\omega\colon\ \xi(\omega)=x\}$ принадлежит $\mathscr{F}$, т. е. является событием. Если $\xi(\Omega)$ — конечное множество, то ξ называется *простой случайной величиной.*

Заметим, что если $x\notin\xi(\Omega)$, то множество $\{\omega\colon\ \xi(\omega)=x\}$ пусто и условие его принадлежности σ-алгебре $\mathscr{F}$ очевидно выполняется. Это условие содержательно, когда $x\in\xi(\Omega)$. Поскольку $\xi(\Omega)$ конечно или счётно, то его можно представить в виде $\xi(\Omega)=\{x_i\}_{i\in I}$, где в качестве индексного множества I выбирается $\{1,2,\ldots,n\}$ в случае простой случайной величины или множество натуральных чисел $\mathbb{N}$, если $\xi(\Omega)$ счётно. Каждому x_i соответствует подмножество $D_i=\{\omega\colon\ \xi(\omega)=x_i\}$ пространства элементарных событий Ω. Поскольку $D_iD_j=\varnothing$ при $i\neq j$ и $\cup_{i\in I}D_i=\Omega$, то $\mathscr{D}_\xi:=\{D_i\}_{i\in I}$ образует разбиение пространства элементарных событий Ω, соответствующее дискретной случайной величине ξ. Вероятность $\mathsf{P}(D_i)$ атома разбиения $\mathscr{D}_\xi$ означает вероятность того, что случайная величина ξ примет значение x_i в результате проведения случайного эксперимента. Числа

$$\mathsf{P}_\xi(x_i) \;:=\; \mathsf{P}(D_i) \;=\; \mathsf{P}(\xi=x_i), \qquad i\in I,$$

образуют *распределение вероятностей* случайной величины ξ.

Определение 3.2. Дискретные случайные величины ξ и η, определённые на вероятностном пространстве $(\Omega,\mathscr{F},\mathsf{P})$, будем называть *независимыми*, если для любых $x,y\in\mathbb{R}$ выполняется равенство

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

Другими словами, события $\{\xi = x\}$ и $\{\eta = y\}$ должны быть независимыми. Очевидно, что это условие содержательно, когда $x \in \xi(\Omega)$ и $y \in \eta(\Omega)$. В противном случае обе части равенства обращаются в нуль.

Простые случайные величины играют особую роль как в построении теории, так и в приложениях. Поэтому мы начнем изучение случайных величин с них. Простейшей случайной величиной является индикатор события $A \in \mathscr{F}$, 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. Для достоверного и невозможного событий имеем $\mathbb{1}_\Omega(\omega) \equiv 1$ и $\mathbb{1}_\varnothing(\omega) \equiv 0$. Кроме того, $\mathbb{1}_{\overline{A}}(\omega) = 1 - \mathbb{1}_A(\omega)$. В дальнейшем мы будем опускать зависимость от элементарного события и писать просто $\mathbb{1}_A$. Если $\{A_i\}_{i=1}^n$ — некоторая совокупность событий, то

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

Пусть теперь $\xi\colon \Omega \to \mathbb{R}$ — простая случайная величина и $\xi(\Omega) = \{x_1, \dots, x_n\} \subset \mathbb{R}$. Если $\mathscr{D}_\xi = \{D_1, \dots, D_n\}$ — разбиение, соответствующее ξ, т.е. $D_i = \{\xi = x_i\}$, $i = 1, \dots, n$, то ξ можно представить в виде линейной комбинации индикаторов атомов разбиения

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

Допустим теперь, что проводится большое количество N экспериментов, соответствующих вероятностному пространству $(\Omega, \mathscr{F}, \mathsf{P})$. Тогда значение x_i случайная величина ξ будет принимать приблизительно в $N\mathsf{P}_\xi(x_i)$ случаях, а среднее наблюдавшихся значений случайной величины ξ будет приблизительно равно

$$\frac{1}{N} \sum_{i=1}^n x_i N \mathsf{P}_\xi(x_i) = \sum_{i=1}^n x_i \mathsf{P}_\xi(x_i).$$

Эти рассуждения делают интуитивно понятным следующее определение.

Определение 3.3. Пусть ξ — простая случайная величина, определённая на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$, и $\xi(\Omega) = \{x_1, \dots, x_n\}$. *Математическим ожиданием* (или *средним*) этой случайной величины называется число

$$\mathsf{E}\xi := \sum_{i=1}^n x_i \mathsf{P}_\xi(x_i) = \sum_{i=1}^n x_i \mathsf{P}(\xi = x_i).$$

Прежде чем приступить к рассмотрению свойств математического ожидания, заметим, что представление случайной величины ξ в виде линейной комбинации индикаторов не является единственным. Например, если один или несколько атомов разбиения $\mathscr{D}_\xi$ представить как объединение непересекающихся множеств из $\mathscr{F}$, то мы получим новое разбиение, на атомах которого ξ принимает постоянные значения. Следовательно, ξ можно представить в виде линейной комбинации индикаторов нового разбиения. При этом на некоторых атомах нового разбиения случайная величина ξ будет принимать одинаковые значения. Следующий результат показывает, что математическое ожидание простой случайной величины можно вычислять, используя любое её представление в виде линейной комбинации индикаторов разбиения.

Предложение 3.1. *Пусть*

$$\mathscr{D}' = \{D'_1, \dots, D'_n\}, \qquad \mathscr{D}'' = \{D''_1, \dots, D''_m\}$$

— разбиения пространства элементарных событий Ω такие, что

$$\xi = \sum_{i=1}^{n} x_i \mathbb{1}_{D'_i} = \sum_{j=1}^{m} y_j \mathbb{1}_{D''_j}.$$

Тогда

$$\sum_{i=1}^{n} x_i \mathsf{P}(D'_i) = \sum_{j=1}^{m} y_j \mathsf{P}(D''_j) = \mathsf{E}\xi.$$

Доказательство. Поскольку $\mathscr{D}'$ и $\mathscr{D}''$ являются разбиениями пространства элементарных событий Ω, то выполняются равенства

$$\mathsf{P}(D'_i) = \sum_{j=1}^{m} \mathsf{P}(D'_i D''_j), \qquad \mathsf{P}(D''_j) = \sum_{i=1}^{n} \mathsf{P}(D'_i D''_j).$$

Следовательно,

$$\sum_{i=1}^{n} x_i \mathsf{P}(D'_i) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i \mathsf{P}(D'_i D''_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} y_j \mathsf{P}(D'_i D''_j) = \sum_{j=1}^{m} y_j \mathsf{P}(D''_j).$$

поскольку $x_i = y_j$, если $D'_i D''_j \neq \varnothing$, а в противном случае $\mathsf{P}(D'_i D''_j) = 0$. □

Заметим также, что если ξ — простая случайная величина, определённая на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$, и $\varphi\colon \mathbb{R} \to \mathbb{R}$ — некоторая функция, то $\eta = \varphi \circ \xi = \varphi(\xi)$ также будет простой случайной величиной на $(\Omega, \mathscr{F}, \mathsf{P})$. При этом, если

$$\xi = \sum_{i=1}^{n} x_i \mathbb{1}_{D_i}, \qquad \mathscr{D}_\xi = \{D_1, \dots, D_n\},$$

то

$$\eta = \sum_{i=1}^{n} \varphi(x_i) \mathbb{1}_{D_i}$$

и, следовательно,

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

Свойства математического ожидания простых случайных величин.

1°. Для любого $A \in \mathscr{F}$ 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 о функции от случайной величины. 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 предложения 3.1 имеем

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

и свойство 2° доказано.

Условие $\xi \geqslant 0$ означает, что в $\xi(\Omega) = \{x_1, \ldots, x_n\}$ все x_i, $i = 1, \ldots, n$, неотрицательны. Но тогда и $\mathsf{E}\xi$, как сумма неотрицательных слагаемых, также будет неотрицательной. Равенство $\mathsf{E}\xi = 0$ возможно лишь в случае, когда $\mathsf{P}_\xi(x_i) = 0$ 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 3.4. *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. □

Отметим еще одно свойство математического ожидания и дисперсии, связанное с независимыми случайными величинами.

Теорема 3.1. *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. □

Ковариация и коэффициент корреляции.

Эти числовые характеристики можно рассматривать как меру зависимости случайных величин.

Определение 3.5. 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. Подбрасывают две игральные кости. Пусть ξ_1 — число очков, выпавших на первой игральной кости, а ξ_2 — на второй. Определим другие две случайные величины: $\eta_1 = \xi_1 + \xi_2$ и $\eta_2 = \xi_1 - \xi_2$. Тогда

$$\mathsf{cov}(\eta_1, \eta_2) = \mathsf{E}(\xi_1 + \xi_2)(\xi_1 - \xi_2) - \mathsf{E}(\xi_1 + \xi_2)\mathsf{E}(\xi_1 - \xi_2) = \mathsf{D}\xi_1 - \mathsf{D}\xi_2 = 0.$$

Однако, η_1 и η_2 одновременно принимают четные и нечетные значения, т. е. они не являются независимыми.

Определение 3.6. 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 $\mathsf{D}\xi_i$. Отметим некоторые особенности ковариационной матрицы. Это симметрическая матрица, на диагонали которой стоят неотрицательные числа. Кроме того, она неотрицательно определена, т. е.

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

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

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

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

$$\mathbb{x}^t \mathbb{V} \mathbb{x} = \mathsf{E}\{\mathbb{x}^t(\mathbb{X} - \mathsf{E}\mathbb{X})(\mathbb{X} - \mathsf{E}\mathbb{X})^t \mathbb{x}\} = \mathsf{E}(\mathbb{x}^t(\mathbb{X} - \mathsf{E}\mathbb{X}))^2 \geqslant 0. \qquad (3.1)$$

Замечание 3.1. 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.

Целочисленные случайные величины и производящие функции.

Если дискретная случайная величина ξ принимает счётное число значений $x_1, x_2, \ldots$ с вероятностями $p_1, p_2, \ldots$, соответственно, т. е. $\mathsf{P}_\xi(x_k) p_k$, $k = 1, 2, \ldots$, то математическое ожидание $\mathsf{E}\xi$ может не существовать даже в случае сходимости ряда $\sum_{n=1}^{\infty} x_k p_k$. Действительно, если этот ряд сходится условно, то по теореме Римана путем перестановки его членов (перенумерацией значений случайной величины) можно получить любое значение в качестве суммы этого ряда. Поэтому при определении математического ожидания дискретной случайной величины со счётным числом значений требуют абсолютную сходимость соответствующего числового ряда.

Определение 3.7. 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

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

Непосредственно из определения производящей функции следует также, что если $\xi_1, \ldots, \xi_n$ — независимые целочисленные случайные величины и $S_n = \xi_1 + \ldots + \xi_n$, то

$$g_{S_n}(x) = \prod_{k=1}^{n} g_{\xi_k}(x).$$

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

$\diamond$ Бернуллиевское распределение.

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

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

$\diamond$ Биномиальное распределение.

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

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

$\diamond$ Пуассоновское распределение.

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

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

$\diamond$ Геометрическое распределение.

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

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

§ 4. Структура сигма–алгебр, продолжение меры, независимость классов событий

На примере простых случайных величин мы видели, что все вычисления, связанные со случайной величиной ξ, переносятся на числовую ось $\mathbb{R}$. Это достигается введением распределения P_ξ, которое представляет собой вероятностную меру на $\mathbb{R}$, сосредоточенную в точках множества $\xi(\Omega)$. При этом $\mathsf{P}_\xi(x)$ для $x \in \xi(\Omega)$ определяется как вероятность

$$\mathsf{P}_\xi(x) \;=\; \mathsf{P}(\{\omega\colon\; \xi(\omega) = x\})$$

и после определения P_ξ нет необходимости обращаться к исходному вероятностному пространству $(\Omega, \mathscr{F}, \mathsf{P})$.

В случае, когда $\xi(\Omega)$ не является счетным (или конечным) множеством, естественно ожидать, что P_ξ будет вероятностной мерой, определённой на некоторой σ-алгебре подмножеств числовой прямой $\mathbb{R}$. Как определить эту σ-алгебру? Вначале уместно было бы выделить наиболее простой класс множеств, на котором точно должна быть определена мера P_ξ. В нашем случае к такому классу множеств следует отнести интервалы $(a, b) \subset \mathbb{R}$, поскольку

$$\mathsf{P}_\xi((a, b)) \;=\; \mathsf{P}(\{\omega\colon\; a < \xi(\omega) < b\})$$

выражает вероятность того факта, что случайная величина ξ примет значение из этого интервала. Учитывая то, что любое открытое множество в $\mathbb{R}$ можно представить в виде объединения не более, чем счётного, множества непересекающихся интервалов, эта σ-алгебра должна содержать и все открытые множества. Прежде чем приступить к определению этой σ-алгебры, рассмотрим некоторые общие результаты о структуре σ-алгебр, порождённых тем или иным классом подмножеств Ω.

Определение 4.1. Класс $\mathscr{T}$ подмножеств Ω называется π-системой, если из условия $A, B \in \mathscr{T}$ следует, что $AB \in \mathscr{T}$.

Другими словами, π-система является замкнутой относительно операции пересечения. В отличие от σ-алгебры проверка выполнения условий π-системы значительно проще. Введем в рассмотрение еще один класс множеств.

Определение 4.2. Класс $\mathscr{G}$ подмножеств Ω называется δ-системой, если выполняются следующие условия:

(i) $\Omega \in \mathscr{G}$;

(ii) Если $A, B \in \mathscr{G}$ и $A \subseteq B$, то $B \setminus A \in \mathscr{G}$;

(iii) Если последовательность множеств $(A_n\colon\; n \in \mathbb{N}) \subset \mathscr{G}$ монотонно возрастает, то её предел $\lim\limits_n A_n = \cup_{n=1}^\infty A_n$ также принадлежит $\mathscr{G}$.

Очевидно, что любая σ-алгебра подмножеств Ω одновременно является π-системой и δ-системой.

Теорема 4.1. *Класс $\mathscr{G}$ подмножеств Ω является σ-алгеброй в том и только том случае, если $\mathscr{G}$ одновременно является π-системой и δ-системой.*

Доказательство. Как следует из замечания выше, нам нужно доказать, что если $\mathscr{G}$ замкнуто относительно операции пересечения и выполнены условия (i)—(iii) из определения δ-системы, то $\mathscr{G}$ является σ-алгеброй. Для этого нам достаточно проверить выполнение условий из определения σ-алгебры. Из условия (i) в определении δ-системы следует, что $\Omega \in \mathscr{G}$. Это совместно с условием (ii) влечёт замкнутость $\mathscr{G}$ относительно операции взятия дополнения. Действительно, если $A \in \mathscr{G}$, то $\Omega \setminus A = \overline{A}$ также принадлежит $\mathscr{G}$. Замкнутость $\mathscr{G}$ относительно конечного числа объединений следует из равенства

$$A \cup B = \overline{\overline{A} \cap \overline{B}}.$$

Пусть теперь последовательность $(A_n\colon n \in \mathbb{N})$ содержится в $\mathscr{G}$. Тогда множества $B_n = \cup_{k=1}^n A_k$ принадлежат $\mathscr{G}$ и образуют монотонно возрастающую последовательность. В силу условия (iii) в определении δ-системы её предел

$$\lim_n B_n = \bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$$

принадлежит $\mathscr{G}$. Таким образом, $\mathscr{G}$ замкнута относительно счётных объединений и теорема доказана. □

Определение 4.3. 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 $\mathscr{G}$ — σ-алгебра подмножеств Ω и $\mathscr{K} \subset \mathscr{G}$, то $\sigma(\mathscr{K}) \subset \mathscr{G}$.

В связи с условием (ii) $\sigma(\mathscr{K})$ называют *минимальной σ-алгеброй, порождённой классом $\mathscr{K}$*. Аналогично определяется $\delta(\mathscr{K})$ — *минимальная δ-система, порождённая классом $\mathscr{K}$*.

Замечание 4.1. Для любого класса $\mathscr{K}$ подмножеств Ω существует единственная минимальная σ-алгебра $\sigma(\mathscr{K})$ (минимальная δ-система $\delta(\mathscr{K})$), порождённая классом $\mathscr{K}$. Это следует из того, что 2^{Ω} (множество всех подмножеств Ω) является σ-алгеброй (δ-системой) и пересечение любого семейства σ-алгебр (δ-систем) также является σ-алгеброй (δ-системой).

Одним из наиболее важных результатов, касающихся структуры σ-алгебр, является следующая теорема, известная в литературе как лемма Дынкина.

Теорема 4.2. *Пусть класс $\mathscr{T}$ подмножеств Ω является π-системой. Тогда $\delta(\mathscr{T}) = \sigma(\mathscr{T})$.*

Доказательство. Включение $\delta(\mathscr{T}) \subseteq \sigma(\mathscr{T})$ очевидно, поскольку σ-алгебра одновременно является и δ-системой. Для доказательства включения $\sigma(\mathscr{T}) \subseteq \delta(\mathscr{T})$ достаточно показать, что $\delta(\mathscr{T})$ является π-системой, и утверждение будет следовать из теоремы 4.1.

Введём в рассмотрение класс множеств

$$\mathscr{E} = \{E \in \delta(\mathscr{T})\colon\ ED \in \delta(\mathscr{T}) \text{ для всех } D \in \mathscr{T}\}.$$

Из определения этого класса видно, что $\mathscr{T} \subseteq \mathscr{E}$. Покажем, что $\mathscr{E}$ является δ-системой. Для этого достаточно проверить выполнение условий (i)—(iii) из определения δ-системы. Условие (i) выполняется, поскольку $\Omega D = D$ для любого $D \in \mathscr{T}$ и $\mathscr{T} \subseteq \delta(\mathscr{T})$. Пусть теперь $A, B \in \mathscr{E}$ и $A \subseteq B$. Тогда для произвольного $D \in \mathscr{T}$ будем иметь включения $AD \in \delta(\mathscr{T})$, $BD \in \delta(\mathscr{T})$ и, следовательно,

$$(B \setminus A)D = B\overline{A}D = (BD) \setminus (AD)$$

также принадлежит $\delta(\mathscr{T})$ и $B \setminus A \in \mathscr{E}$. Условие (ii) из определения δ-системы для $\mathscr{E}$ выполнено.

Допустим, наконец, что последовательность $(A_n\colon\ n \in \mathbb{N}) \subset \mathscr{E}$ является монотонно возрастающей и $\lim\limits_n A_n = A$. Тогда для любого $D \in \mathscr{T}$ множества $A_n D$ будут принадлежать $\delta(\mathscr{T})$, а поскольку последовательность $(A_n D\colon\ n \in \mathbb{N})$ также монотонно возрастает и $\lim\limits_n (A_n D) = AD$, то $AD \in \delta(\mathscr{T})$. Таким образом, условие (iii) также выполняется для $\mathscr{E}$ и, следовательно, $\mathscr{E}$ является δ-системой. Из условия минимальности следует, что $\mathscr{E} = \delta(\mathscr{T})$.

Определим теперь класс множеств

$$\mathscr{G} = \{G \in \delta(\mathscr{T})\colon\ GD \in \delta(\mathscr{T})\,\text{для всех}\,D \in \delta(\mathscr{T})\}$$

и покажем, что $\mathscr{G} = \delta(\mathscr{T})$. Условие $\mathscr{T} \subseteq \mathscr{G}$ следует из определения $\mathscr{E}$ и равенства $\mathscr{E} = \delta(\mathscr{T})$. Проверка выполнения условий (i)—(iii) для $\mathscr{G}$ проводится аналогично тому, как это было сделано для класса $\mathscr{E}$. Таким образом, $\mathscr{G}$ является δ-системой, что влечёт равенство $\mathscr{G} = \delta(\mathscr{T})$.

Заметим теперь, что равенство $\mathscr{G} = \delta(\mathscr{T})$ означает замкнутость $\delta(\mathscr{T})$ относительно операции пересечения множеств. Следовательно, $\delta(\mathscr{T})$ является π-системой и теорема доказана. □

Определение 4.4. Пусть $\Omega = \mathbb{R}$ (или $\mathbb{R}^n$) и $\mathscr{K}$ — совокупность открытых множеств. Минимальная σ-алгебра $\sigma(\mathscr{K})$ называется *борелевской* σ-алгеброй и обозначается $\mathscr{B}(\mathbb{R}) = \mathscr{B}$ (соответственно, $\mathscr{B}(\mathbb{R}^n) = \mathscr{B}^n$). Множества B из $\mathscr{B}$ называются *борелевскими* множествами, а измеримое пространство $(\mathbb{R}, \mathscr{B})$ называется борелевским пространством.

Борелевская σ-алгебра является более широким классом множеств, чем может показаться на первый взгляд. В частности, открытые и замкнутые множества, а также всевозможные их счётные объединения и пересечения являются борелевскими множествами. Однако этим далеко не исчерпывается σ-алгебра $\mathscr{B}$ и её конструктивное описание достаточно сложное. С другой стороны, мы приходим к борелевской σ-алгебре $\mathscr{B}$, если значительно сузим порождающий класс множеств.

Предложение 4.1. *Пусть $\mathscr{T}$ — класс полуинтервалов вида $(-\infty, x]$ (или $(-\infty, x), [x, \infty), (x, \infty)$), $x \in \mathbb{R}$. Тогда минимальная σ-алгебра $\sigma(\mathscr{T})$ совпадает с борелевской σ-алгеброй $\mathscr{B}$.*

Доказательство. Нам достаточно показать, что открытые множества содержатся в $\sigma(\mathscr{T})$. Проведём рассуждения, когда $\mathscr{T}$ состоит из полуинтервалов вида $(-\infty, x]$, $x \in \mathbb{R}$. Случаи других видов полуинтервалов рассматриваются аналогично.

Заметим вначале, что $\sigma(\mathscr{T})$ содержит полуинтервалы вида $(a,b]$, где $a<b$. Действительно, это следует из замкнутости $\sigma(\mathscr{T})$ относительно теоретико-множественных операций и равенства $(a,b] = (-\infty,b] \setminus (-\infty,a]$. Далее, поскольку

$$\bigcup_{n=1}^{\infty} \left(a, b - \frac{\varepsilon}{n}\right] = (a,b), \qquad \varepsilon = \frac{b-a}{3},$$

то $\sigma(\mathscr{T})$ содержит все открытые интервалы (a,b). С другой стороны, каждое открытое множество $G \subset \mathbb{R}$ можно представить в виде объединения (не более счётного числа) непересекающихся открытых интервалов. Следовательно, открытые множества содержатся в $\sigma(\mathscr{T})$ и утверждение доказано. $\square$

Из результатов этого параграфа следует, что естественным измеримым пространством, на котором должна быть определена мера P_ξ в общем случае, является борелевское пространство $(\mathbb{R},\mathscr{B})$. Однако, для определения P_ξ через вероятностную меру P вероятностного пространства $(\Omega,\mathscr{F},\mathsf{P})$, на котором определена случайная величина ξ, потребуется также, чтобы прообразы $\xi^{-1}(B)$, $B \in \mathscr{B}$, принадлежали σ-алгебре $\mathscr{F}$.

Определение 4.5. Пусть $(\Omega,\mathscr{F},\mathsf{P})$ — вероятностное пространство. Отображение $\xi\colon \Omega \to \mathbb{R}$ называется случайной величиной, определённой на этом вероятностном пространстве, если для любого борелевского множества $B \in \mathscr{B}$ его прообраз

$$\xi^{-1}(B) = \{\omega\colon \xi(\omega) \in B\}$$

принадлежит σ-алгебре $\mathscr{F}$.

Как следует из определения, чтобы установить, что некоторая функция $\xi\colon \Omega \to \mathbb{R}$ является случайной величиной, нужно проверить условие $\xi^{-1}(B) \in \mathscr{F}$ для всех борелевских множеств B. В действительности, проверку этого условия можно существенно сократить.

Отметим вначале некоторые свойства операции взятия прообраза. Пусть Ω и E — произвольные множества и $f\colon \Omega \to E$ — некоторое отображение. Тогда имеют место следующие утверждения.

(a) Если $\{G_i\}_{i\in I}$ — семейство подмножеств E, то

$$f^{-1}\left(\bigcup_{i\in I} G_i\right) = \bigcup_{i\in I} f^{-1}(G_i), \qquad f^{-1}\left(\bigcap_{i\in I} G_i\right) = \bigcap_{i\in I} f^{-1}(G_i).$$

(b) Если $G \subset E$, то $f^{-1}(E\setminus G) = \Omega \setminus f^{-1}(G)$.

Предложение 4.2. *Пусть $(\Omega,\mathscr{F},\mathsf{P})$ — вероятностное пространство и $\xi\colon \Omega \to \mathbb{R}$ — отображение, удовлетворяющее условию $\xi^{-1}(D) \in \mathscr{F}$ для всех $D \in \mathscr{T}$, где $\mathscr{T}$ — класс подмножеств $\mathbb{R}$ такой, что $\sigma(\mathscr{T}) = \mathscr{B}$. Тогда ξ является случайной величиной.*

Доказательство. Нам нужно показать, что $\xi^{-1}(B) \in \mathscr{F}$ для всех борелевских множеств B. Рассмотрим класс подмножеств $\mathbb{R}$, выделяемый условием

$$\mathscr{G} = \{B \subseteq \mathbb{R}\colon \xi^{-1}(B) \in \mathscr{F}\}.$$

Покажем, что $\mathscr{G}$ является σ-алгеброй. Действительно, $\mathbb{R} \in \mathscr{G}$, поскольку $\xi^{-1}(\mathbb{R}) = \Omega$. Если $G \in \mathscr{G}$, то в силу (b) $\xi^{-1}(\mathbb{R} \setminus G) = \Omega \setminus \xi^{-1}(G)$ принадлежит $\mathscr{F}$, а $\mathbb{R} \setminus G$ принадлежит $\mathscr{G}$. Наконец, если последовательность $\{G_n\}$ содержится в $\mathscr{G}$, т. е. $\xi^{-1}(G_n) \in \mathscr{F}$, $n = 1, 2, \ldots$, то в силу (a)

$$\xi^{-1}\left(\bigcup_{n=1}^{\infty} G_n\right) = \bigcup_{n=1}^{\infty} \xi^{-1}(G_n),$$

откуда следует замкнутость $\mathscr{G}$ относительно счётных объединений. Таким образом, $\mathscr{G}$ является σ-алгеброй.

Замечая, что по условию $\mathscr{T} \subset \mathscr{G}$ и $\sigma(\mathscr{T}) = \mathscr{B}$, приходим к включению $\mathscr{B} \subseteq \mathscr{G}$, что и доказывает наше утверждение. □

Как следует из предложения 4.1, в качестве $\mathscr{T}$ можно взять класс полуинтервалов вида $(-\infty, x]$ (или $(-\infty, x)$, $[x, \infty)$, (x, ∞)), $x \in \mathbb{R}$. В частности, $\xi\colon \Omega \to \mathbb{R}$ будет случайной величиной, если для всех $x \in \mathbb{R}$ выполняется условие

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

Определение 4.6. Пусть ξ — случайная величина, определённая на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$. Под *распределением* (или *распределением вероятностей*) случайной величины ξ будем понимать вероятностную меру P_ξ, определённую на измеримом пространстве $(\mathbb{R}, \mathscr{B})$ равенством

$$\mathsf{P}_\xi(B) = \mathsf{P}(\xi^{-1}(B)) = \mathsf{P}(\{\omega\colon \xi(\omega) \in B\})$$

для всех $B \in \mathscr{B}$.

То, что для P_ξ выполняются все условия из определения вероятностной меры, следует из свойств (a), (b) операции взятия прообраза и того, что P является вероятностной мерой. Если ξ — дискретная случайная величина и $\xi(\Omega) = (x_i\colon i \in I)$, то для любого $B \in \mathscr{B}$ имеет место равенство

$$\mathsf{P}_\xi(B) = \sum_{i\colon x_i \in B} \mathsf{P}_\xi(x_i).$$

Если $(\Omega, \mathscr{F})$ — измеримое пространство, а $\mu\colon \mathscr{F} \to \mathbb{R}_+$ — счётно-аддитивная функция с $\mu(\Omega) < \infty$ (т. е. не обязательно нормированная), то будем говорить, что μ — конечная мера, определённая на измеримом пространстве $(\Omega, \mathscr{F})$. Поскольку μ отличается от вероятностной меры лишь положительным множителем, то она обладает всеми основными свойствами: конечная аддитивность, монотонность, непрерывность. Как отмечалось в начале параграфа, обычно мера определяется на некотором простом классе множеств, а затем продолжается на σ-алгебру. В частности, борелевская мера μ (т. е. определённая на $(\mathbb{R}, \mathscr{B})$) продолжается с класса $\mathscr{T}$ полуинтервалов вида $(a, b]$, $a < b$, по следующей схеме. Вначале с учётом свойства аддитивности μ продолжается на алгебру $\mathscr{A}$ подмножеств $\mathbb{R}$, которые представимы в виде объединения конечного числа непересекающихся полуинтервалов из $\mathscr{T}$. После этого на всех подмножествах $\mathbb{R}$ определяется внешняя мера μ^* и по схеме Лебега или Каратеодори выделяется σ-алгебра измеримых множеств, на которых и определяется μ, как продолжение с алгебры $\mathscr{A}$. Следующая теорема показывает, что мера единственным образом продолжается с порождающей π-системы.

Теорема 4.3. *Пусть* $(\Omega, \mathscr{F})$ — *измеримое пространство и* $\mathscr{T}$ — *некоторая* π*-система подмножеств* Ω*, для которой* $\sigma(\mathscr{T}) = \mathscr{F}$*. Допустим также, что на* $(\Omega, \mathscr{F})$ *определены две меры* μ *и* ν*, удовлетворяющие условиям* $\mu(\Omega) = \nu(\Omega)$ *и* $\mu(T) = \nu(T)$ *для всех* $T \in \mathscr{T}$*. Тогда* μ *и* ν *совпадают на* $\mathscr{F}$*, т. е.* $\mu(A) = \nu(A)$ *для всех* $A \in \mathscr{F}$*.*

Доказательство. Введём в рассмотрение класс множеств

$$\mathscr{G} = \{G \in \mathscr{F} \colon \mu(G) = \nu(G)\}$$

и покажем, что $\mathscr{G}$ является δ-системой. Действительно, поскольку $\mu(\Omega) = \nu(\Omega)$, то $\Omega \in \mathscr{G}$. Далее, если $A, B \in \mathscr{G}$ и $A \subseteq B$, то

$$\mu(B \setminus A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B \setminus A)$$

и $B \setminus A$ также принадлежит $\mathscr{G}$. Наконец, если $(G_n \colon n \in \mathbb{N})$ — монотонно возрастающая последовательность множеств из $\mathscr{G}$ и $G = \lim_n G_n$, то в силу непрерывности меры получаем

$$\mu(G) = \lim_{n\to\infty} \mu(G_n) = \lim_{n\to\infty} \nu(G_n) = \nu(G),$$

откуда следует, что $G \in \mathscr{G}$. Таким образом, $\mathscr{G}$ является δ-системой. Но тогда $\delta(\mathscr{T}) \subseteq \mathscr{G}$, поскольку $\mathscr{T} \subset \mathscr{G}$ по условию теоремы. В силу леммы Дынкина 4.2 $\delta(\mathscr{T}) = \sigma(\mathscr{T}) = \mathscr{F}$ и теорема доказана. □

Определение 4.7. Пусть $(\Omega, \mathscr{F}, \mathsf{P})$ — вероятностное пространство и $\mathscr{K}_i \subset \mathscr{F}$, $i = 1, \ldots, n$. Классы $\mathscr{K}_1, \ldots, \mathscr{K}_n$ называются независимыми, если для всякого выбора $A_1 \in \mathscr{K}_1, \ldots, A_n \in \mathscr{K}_n$ получаем независимые (в совокупности) события. Бесконечное семейство классов событий называется независимым, если независимо любое его конечное подсемейство.

Теорема 4.4. *Пусть классы* $\mathscr{T}_1, \ldots, \mathscr{T}_n$ *подмножеств* Ω *являются* π*-системами и независимы. Тогда независимы также порождённые ими* σ*-алгебры* $\sigma(\mathscr{T}_1), \ldots, \sigma(\mathscr{T}_n)$*.*

Доказательство. Докажем вначале, что $\sigma(\mathscr{T}_1), \mathscr{T}_2, \ldots, \mathscr{T}_n$ являются независимыми. Пусть $I \subseteq \{2, \ldots, n\}$ и для каждого $i \in I$ выберем в $\mathscr{T}_i$ произвольно T_i. Определим на $\sigma(\mathscr{T}_1)$ две меры

$$\mu(A) = \mathsf{P}\left(A \bigcap_{i\in I} T_i\right), \qquad \nu(A) = \mathsf{P}(A) \prod_{i\in I} \mathsf{P}(T_i).$$

По условию теоремы эти меры совпадают на $\mathscr{T}_1$ и $\mu(\Omega) = \nu(\Omega)$. По теореме 4.3 эти меры совпадают на $\sigma(\mathscr{T}_1)$, т. е. $\mu(A) = \nu(A)$ для всех $A \in \sigma(\mathscr{T}_1)$. Это эквивалентно тому, что

$$\mathsf{P}\left(A \bigcap_{i\in I} T_i\right) = \mathsf{P}(A) \prod_{i\in I} \mathsf{P}(T_i)$$

для любых $A \in \sigma(\mathscr{T}_1)$ и $I \subseteq \{2, \ldots, n\}$, т. е. $\sigma(\mathscr{T}_1), \mathscr{T}_2, \ldots, \mathscr{T}_n$ являются независимыми.

Далее аналогичными рассуждениями приходим к независимости классов $\sigma(\mathscr{T}_1), \sigma(\mathscr{T}_2), \mathscr{T}_3 \ldots, \mathscr{T}_n$. Продолжая процесс замены π-систем на их минимальные σ-алгебры, приходим к утверждению теоремы. □

Пусть ξ — случайная величина, определённая на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$. Информация о том, какое значение приняла случайная величина ξ, позволяет определить произошло или нет каждое из событий $\xi^{-1}(B) \in \mathscr{F}$, $B \in \mathscr{B}$. Таким образом, с каждой случайной величиной ξ связана σ-алгебра

$$\sigma(\xi) = \{\xi^{-1}(B) \colon B \in \mathscr{B}\}.$$

Из предыдущего видно, что эта σ-алгебра порождается π-системой

$$\pi(\xi) = \{\xi^{-1}((-\infty, x]) \colon x \in \mathbb{R}\},$$

т. е. $\sigma(\pi(\xi)) = \sigma(\xi)$.

Определение 4.8. Случайные величины $\xi_1, \ldots, \xi_n$, определённые на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$, называются независимыми, если независимы $\sigma(\xi_1), \ldots, \sigma(\xi_n)$.

Заметим, что из теоремы 4.4 следует, что случайные величины независимы, если независимы их π-системы $\pi(\xi_1), \ldots, \pi(\xi_n)$.

Если $\xi_1, \ldots, \xi_n$ — случайные величины и $\sigma(\xi_1), \ldots, \sigma(\xi_n)$ — их σ-алгебры, то объединение $\cup_{k=1}^n \sigma(\xi_k)$ вовсе не обязано быть σ-алгеброй. Через $\sigma(\xi_1, \ldots, \xi_n)$ будем обозначать минимальную σ-алгебру, порождённую этим объединением.

Пусть $\xi_1, \xi_2, \ldots$ — последовательность случайных величин. Определим следующие σ-алгебры, связанные с этой последовательностью,

$$\mathscr{A}_n = \sigma(\xi_1, \ldots, \xi_n), \qquad \mathscr{X}_n = \sigma(\xi_{n+1}, \xi_{n+2}, \ldots), \qquad \mathscr{X} = \bigcap_{n=1}^{\infty} \mathscr{X}_n.$$

При этом $\mathscr{X}$ называется *хвостовой* σ-алгеброй для последовательности $(\xi_n \colon n \in \mathbb{N})$. Эта σ-алгебра содержит ряд важных событий. Например, сходимость самой последовательности или ряда $\sum_{n=1}^{\infty} \xi_n$ относятся к событиям хвостовой σ-алгебры.

Теорема 4.5 (Закон «0 или 1» Колмогорова). *Пусть $(\xi_n \colon n \in \mathbb{N})$ — последовательность независимых случайных величин и $\mathscr{X}$ — её хвостовая σ-алгебра. Тогда для любого $A \in \mathscr{X}$ вероятность $\mathsf{P}(A)$ равна нулю, либо единице.*

Доказательство. Покажем, что $\mathscr{X}$ не зависит от себя, т. е. для любых $A, B \in \mathscr{X}$ выполняется равенство $\mathsf{P}(AB) = \mathsf{P}(A)\mathsf{P}(B)$. Но тогда для любого $A \in \mathscr{X}$ будет выполняться равенство $\mathsf{P}(A) = (\mathsf{P}(A))^2$, откуда будет следовать утверждение теоремы.

Докажем вначале, что $\mathscr{A}_n$ и $\mathscr{X}_n$ (в обозначениях перед формулировкой теоремы) являются независимыми. Для этого рассмотрим две π-системы: $\mathscr{T}_n$ — совокупность множеств вида

$$T = \bigcap_{i \in I} \{\xi_i \leqslant x_i\}, \qquad x_i \in \mathbb{R}, \qquad I \subseteq \{1, \ldots, n\},$$

и $\mathscr{S}_n$ — совокупность множеств вида

$$S = \bigcap_{i \in I} \{\xi_{n+i} \leqslant x_i\}, \qquad x_i \in \mathbb{R},$$

где I — любое конечное подмножество в $\mathbb{N}$. Из условия теоремы следует, что $\mathscr{T}_n$ и $\mathscr{S}_n$ независимы. Кроме того, $\sigma(\mathscr{T}_n) = \mathscr{A}_n$, $\sigma(\mathscr{S}_n) = \mathscr{X}_n$. Следовательно, по теореме 4.4 σ-алгебры $\mathscr{A}_n$ и $\mathscr{X}_n$ независимы.

Далее, из включения $\mathscr{X} \subseteq \mathscr{X}_n$ следует, что $\mathscr{X}$ и $\mathscr{A}_n$ независимы при всех $n = 1, 2, \ldots$.

Заметим теперь, что $\mathscr{A}_1 \subseteq \mathscr{A}_2 \subseteq \ldots$. Поэтому $\mathscr{A} = \cup_{n=1}^{\infty} \mathscr{A}_n$ является алгеброй (не обязательно σ-алгеброй). Но тогда из независимости $\mathscr{X}$ и $\mathscr{A}_n$, $n = 1, 2, \ldots$, следует независимость $\mathscr{X}$ и $\mathscr{A}$. Однако, $\sigma(\mathscr{A}) = \mathscr{X}_1$ и снова по теореме 4.4 приходим к независимости $\mathscr{X}$ и $\mathscr{X}_1$. С другой стороны, $\mathscr{X} \subseteq \mathscr{X}_1$, откуда следует, что $\mathscr{X}$ независима от себя и теорема доказана. □

§ 5. Случайные величины (общий случай)

5.1. Распределение вероятностей. Пусть ξ — случайная величина, определённая на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$. На практике вовсе не важно, как устроено вероятностное пространство $(\Omega, \mathscr{F}, \mathsf{P})$ при изучении числовых характеристик случайной величины ξ, поскольку все вычисления переносятся на вероятностное пространство $(\mathbb{R}, \mathscr{B}, \mathsf{P}_\xi)$. Напомним, что P_ξ является борелевской мерой, $\mathsf{P}_\xi(B) = \mathsf{P}(\xi^{-1}(B))$, $B \in \mathscr{B}$, и называется распределением вероятностей случайной величины ξ. Как следует из результатов предыдущего параграфа, P_ξ вполне определяется своими значениями на полуинтервалах вида $(-\infty, x]$, $x \in \mathbb{R}$.

Определение 5.1. Пусть ξ — случайная величина, определённая на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$. Функция

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

называется *функцией распределения* случайной величины ξ.

Отметим характеристические свойства функции распределения.

1° $F_\xi(x)$ является неубывающей функцией на $\mathbb{R}$;

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

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

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 $1^\circ - 3^\circ$, может рассматриваться как функция распределения некоторой случайной величины ξ. Мера полуинтервалов $(a, b]$, $a < b$, определяется посредством равенства

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

Далее она единственным способом продолжается на σ-алгебру $\mathscr{B}$, поскольку, как было показано в предыдущем параграфе, борелевские меры, совпадающие на π-системе $\mathscr{T}$ полуинтервалов $(-\infty, x]$, $x \in \mathbb{R}$, совпадают и на $\mathscr{B}$.

В случае простой случайной величины ξ функция распределения F_ξ представляет собой « ступенчатую» функцию с множеством точек разрыва $\xi(\Omega)$. Среди непрерывных функций распределения выделяются так называемые абсолютно непрерывные.

Определение 5.2. Будем говорить, что случайная величина ξ имеет *абсолютно непрерывное* распределение, если существует неотрицательная функция f_ξ, которая интегрируема на $\mathbb{R}$ и для всех $x \in \mathbb{R}$ выполняется равенство

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

При этом f_ξ называется *плотностью* распределения случайной величины ξ.

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.

◊ Равномерное на отрезке $[a, b]$

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

◊ Показательное (или экспоненциальное)

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

◇ Распределение Коши
$$f_\xi(x) \,=\, \frac{1}{\pi}\frac{1}{x^2+1}.$$

◇ Нормальное (или гауссовское) с параметрами $a \in \mathbb{R}, \quad \sigma > 0$
$$f_\xi(x) \,=\, \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/2\sigma^2}.$$

В этом случае пишут $\xi \in \mathscr{N}(a, \sigma^2)$. Если $a = 0$ и $\sigma = 0$, то говорят, что ξ имеет стандартное нормальное распределение.

Все свойства плотности для приведенных примеров распределений проверяются непосредственными вычислениями. При проверке равенства интеграла единице в случае нормального распределения используется известный несобственный интеграл
$$\int\limits_{-\infty}^{\infty} e^{-x^2} dx \,=\, \sqrt{\pi}.$$

В случае, когда ξ является простой случайной величиной и $\varphi\colon\ \mathbb{R} \to \mathbb{R}$ — произвольная функция, мы можем рассмотреть композицию $\varphi \circ \xi$, которая также будет простой случайной величиной. В общем случае возникают ограничения на функцию φ.

Определение 5.3. Функция $\varphi\colon\ \mathbb{R} \to \mathbb{R}$ называется борелевской, если для любого борелевского множества $B \in \mathscr{B}$ его прообраз $\varphi^{-1}(B)$ также является борелевским множеством.

Непосредственная проверка показывает, что если ξ — случайная величина, определённая на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$, а $\varphi\colon\ \mathbb{R} \to \mathbb{R}$ — борелевская функция, то композиция $\varphi \circ \xi$ также является случайной величиной, определённой на том же вероятностном пространстве.

Класс борелевских функций достаточно широк. В частности, всякая непрерывная функция является борелевской. Действительно, в этом случае прообразом $\varphi^{-1}(G)$ любого открытого множества $G \subset \mathbb{R}$ является открытое множество. Поскольку $\mathscr{B}$ является минимальной σ-алгеброй, порождённой классом открытых множеств, то утверждение следует из рассуждений, аналогичных проведенным при доказательстве предложения 4.2 из предыдущего параграфа. Выбирая соответствующим образом функцию φ, получаем, что если ξ — случайная величина, то также случайными величинами являются $|\xi|, \xi^2$ и т. д.

5.2. Математическое ожидание. Рассмотрим теперь вопрос выделения класса случайных величин, для которых можно определить математическое ожидание. Вначале мы это сделаем для неотрицательных случайных величин. При этом важную роль будут играть простые случайные величины, для которых математическое ожидание уже определено.

Теорема 5.1. *Пусть ξ — неотрицательная случайная величина, определённая на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$. Тогда найдётся последовательность $(\xi_n\colon\ n \in \mathbb{N})$ простых случайных величин такая, что $0 \leqslant \xi_n \leqslant \xi_{n+1}$,*

$\sigma(\xi_n) \subseteq \sigma(\xi)$, $n = 1, 2, \ldots$, *и*

$$\lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)$$

для всех $\omega \in \Omega$.

Доказательство. Для каждого $n \in \mathbb{N}$ введём в рассмотрение разбиение $\mathscr{D}_n = \{\Delta_n, D_1^{(n)}, \ldots, D_{n2^n}^{(n)}\}$, где

$$\Delta_n = \xi^{-1}([n, \infty)), \qquad D_k^{(n)} = \xi^{-1}\left(\left[(k-1)2^{-n}, k2^{-n}\right)\right), \qquad k = 1, \ldots, n2^n,$$

и определим простую случайную величину

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

Из определения ξ_n видно, что $\sigma(\xi_n) \subseteq \sigma(\xi)$ и $0 \leqslant \xi_n \leqslant \xi$. Остаётся показать, что при всех $\omega \in \Omega$ числовая последовательность $(\xi_n(\omega) \colon n \in \mathbb{N})$ является неубывающей и $\xi_n(\omega) \to \xi(\omega)$ при $n \to \infty$.

Если $\omega \in \Delta_n$, то $\xi_n(\omega) = n$, а $\xi_{n+1}(\omega)$ определяется в зависимости от того $\omega \in \Delta_{n+1}$ (тогда $\xi_{n+1}(\omega) = n+1$) или $\omega \in D_k^{(n+1)}$ при $k > n2^{n+1}$ (тогда $\xi_{n+1}(\omega) \geqslant n$), т. е. $\xi_n(\omega) \leqslant \xi_{n+1}(\omega)$. Если $\omega \in D_k^{(n)}$ при $k \leqslant n2^n$, то поскольку при переходе к разбиению $\mathscr{D}_{n+1}$ атом $D_k^{(n)}$ делится на атомы $D_{2k-1}^{(n+1)}$ и $D_{2k}^{(n+1)}$, то $\xi_{n+1}(\omega)$ равняется либо $(k-1)2^{-n} = \xi_n(\omega)$, либо $(2k-1)2^{-(n+1)} > \xi_n(\omega)$.

Далее, фиксируем произвольно $\omega \in \Omega$, и пусть N такое, что $\xi(\omega) < N$. Тогда при $n \geqslant N$ выполняются неравенства

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

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

Заметим теперь, что в силу свойств математического ожидания для простых случайных величин числовая последовательность $(\mathsf{E}\xi_n \colon n \in \mathbb{N})$ математических ожиданий аппроксимирующей последовательности из теоремы 5.1 является неубывающей. Поэтому она либо стремится к бесконечности, либо имеет конечный предел. Оказывается, что предел математических ожиданий (конечный или бесконечный) не зависит от выбора аппроксимирующей монотонной последовательности простых неотрицательных случайных величин.

Теорема 5.2. *Пусть* $\eta, \xi_1, \xi_2, \ldots$ — *простые неотрицательные случайные величины и последовательность* $(\xi_n \colon n \in \mathbb{N})$ *является неубывающей. Если при этом*

$$\lim_{n\to\infty} \xi_n(\omega) = \xi(\omega) \geqslant \eta(\omega)$$

для всех $\omega \in \Omega$, *то*

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

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

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

Из условий теоремы следует, что $(A_n\colon\ n \in \mathbb{N})$ — неубывающая последовательность событий и $\lim\limits_n A_n = \Omega$. Поэтому $\mathsf{P}(A_n) \to 1$ при $n \to \infty$. Далее, пусть

$$M \;=\; \max\{\eta(\omega)\colon\ \omega \in \Omega\}.$$

Тогда

$$\xi_n \;\geqslant\; (\eta - \varepsilon)\mathbb{1}_{A_n} \;\geqslant\; \eta - \varepsilon - M\mathbb{1}_{\overline{A_n}},$$

откуда следует неравенство

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

Поскольку $\mathsf{P}(\overline{A_n}) \to 0$ при $n \to \infty$, то $\lim\limits_{n\to\infty} \mathsf{E}\xi_n \geqslant \mathsf{E}\eta - \varepsilon$. Учитывая теперь произвольность в выборе ε, получаем $\lim\limits_{n\to\infty} \mathsf{E}\xi_n \geqslant \mathsf{E}\eta$ и теорема доказана. □

Следствие 5.1. *Пусть $(\xi_n\colon\ n \in \mathbb{N})$ и $(\eta_n\colon\ n \in \mathbb{N})$ — две неубывающие последовательности простых неотрицательных случайных величин и*

$$\lim_{n\to\infty} \xi_n(\omega) \;=\; \lim_{n\to\infty} \eta_n(\omega) \;=\; \xi(\omega)$$

для всех $\omega \in \Omega$. Тогда

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

либо оба предела равны бесконечности.

Доказательство. Из теоремы 5.2 следует, что для каждого $k = 1, 2, \ldots$ выполняется неравенство

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

Осуществляя в этом неравенстве предельный переход при $k \to \infty$, получаем

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

Меняя ролями эти последовательности, получим обратное неравенство, что и доказывает наше утверждение. □

Определение 5.4. Пусть ξ — неотрицательная случайная величина и $(\xi_n\colon\ n \in \mathbb{N})$ — неубывающая последовательность простых неотрицательных случайных величин такая, что $\lim\limits_{n\to\infty} \xi_n = \xi$. Тогда под математическим ожиданием случайной величины ξ будем понимать

$$\mathsf{E}\xi \;:=\; \lim_{n\to\infty} \mathsf{E}\xi_n. \tag{5.1}$$

Корректность этого определения следует из теоремы 5.1 и следствия 5.1 Предел в (5.1), конечный или бесконечный, существует в силу монотонности последовательности $(\mathsf{E}\xi_n\colon\ n \in \mathbb{N})$. Часто допускают значение бесконечности в (5.1) из соображений удобства, но мы будем считать, что $\mathsf{E}\xi$ для неотрицательной случайной величины существует, если этот предел конечный.

Для произвольной случайной величины ξ рассматриваются две неотрицательные случайные величины

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

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

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

Введённое таким образом математическое ожидание (в случае его существования) обладает теми же свойствами, которые были установлены для простых случайных величин. Покажем, например, как устанавливается свойство мультипликативности математического ожидания.

Предложение 5.1. *Пусть ξ и η — независимые случайные величины с конечными математическими ожиданиями. Тогда их произведение $\xi\eta$ также имеет конечное математическое ожидание и выполняется равенство*

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

Доказательство. Допустим вначале, что ξ и η — независимые неотрицательные случайные величины с конечными математическими ожиданиями $\mathsf{E}\xi$ и $\mathsf{E}\eta$. В силу теоремы 5.1 найдутся такие неубывающие последовательности $(\xi_n\colon\ n \in \mathbb{N})$ и $(\eta_n\colon\ n \in \mathbb{N})$ простых неотрицательных случайных величин, что

$$\lim_{n\to\infty} \xi_n \,=\, \xi, \qquad \lim_{n\to\infty} \eta_n \,=\, \eta.$$

При этом $\sigma(\xi_n) \subseteq \sigma(\xi)$, $\sigma(\eta_n) \subseteq \sigma(\eta)$ для всех $n = 1, 2, \ldots$. Из независимости ξ и η следует независимость ξ_n и η_n и потому $\mathsf{E}\xi_n\eta_n = \mathsf{E}\xi_n\mathsf{E}\eta_n$ по свойству математического ожидания для простых случайных величин. Кроме того, $(\xi_n\eta_n\colon n \in \mathbb{N})$ является неубывающей последовательностью простых неотрицательных случайных величин, для которой $\lim\limits_{n\to\infty} \xi_n\eta_n = \xi\eta$. Но тогда по определению математического ожидания для неотрицательных случайных величин получаем

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

В случае произвольных независимых случайных величин с конечными математическими ожиданиями заметим, что $\sigma(\xi^\pm) \subseteq \sigma(\xi)$ и $\sigma(\eta^\pm) \subseteq \sigma(\eta)$. Следовательно, $\xi^\pm$ и $\eta^\pm$ являются неотрицательными независимыми случайными величинами и по доказанному с учётом свойства линейности математического ожидания получаем

$$\begin{aligned}
\mathsf{E}\xi\eta &= \mathsf{E}(\xi^+ - \xi^-)(\eta^+ - \eta^-)\\
&= \mathsf{E}\xi^+\mathsf{E}\eta^+ - \mathsf{E}\xi^+\mathsf{E}\eta^- - \mathsf{E}\xi^-\mathsf{E}\eta^+ + \mathsf{E}\xi^-\mathsf{E}\eta^-\\
&= (\mathsf{E}\xi^+ - \mathsf{E}\xi^-)(\mathsf{E}\eta^+ - \mathsf{E}\eta^-)\\
&= \mathsf{E}\xi\mathsf{E}\eta.
\end{aligned}$$

□

Замечание 5.1. Свойство мультипликативности математического ожидания имеет место и для n случайных величин. Если $\xi_1, \ldots, \xi_n$ — независимые случайные величины с конечными математическими ожиданиями, то их произведение также имеет конечное математическое ожидание и

$$\mathsf{E}(\xi_1 \cdot \ldots \cdot \xi_n) \,=\, \mathsf{E}\xi_1 \cdot \ldots \cdot \mathsf{E}\xi_n.$$

При определении математического ожидания $\mathsf{E}\xi$ была реализована конструкция интеграла Лебега от функции $\xi\colon \Omega \to \mathbb{R}$, измеримой относительно σ-алгебры $\mathscr{F}$, по мере P. Таким образом,

$$\mathsf{E}\xi = \int_{\Omega} \xi(\omega)\mathsf{P}(d\omega) = \int_{\Omega} \xi d\mathsf{P}.$$

Как и в случае дискретных случайных величин все вычисления, связанные со случайной величиной ξ, можно перенести на вероятностное пространство $(\mathbb{R}, \mathscr{B}, \mathsf{P}_\xi)$.

Теорема 5.3. *Пусть ξ — случайная величина, определённая на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$, с распределением вероятностей P_ξ и $\varphi\colon \mathbb{R} \to \mathbb{R}$ — борелевская функция. Тогда из существования одного из интегралов*

$$\int_{\Omega} \varphi(\xi) d\mathsf{P}, \qquad \int_{\mathbb{R}} \varphi d\mathsf{P}_\xi$$

следует существование другого и их равенство, т. е.

$$\int_{\mathbb{R}} \varphi d\mathsf{P}_\xi = \int_{\mathbb{R}} \varphi(x)\mathsf{P}_\xi(dx) = \mathsf{E}\varphi(\xi). \tag{5.2}$$

Доказательство. Допустим вначале, что $\varphi(x) = \mathbb{1}_B(x)$, где $B \in \mathscr{B}$. Тогда $\varphi \circ \xi = \mathbb{1}_B \circ \xi = \mathbb{1}_{\xi^{-1}(B)}$ и

$$\mathsf{E}\varphi(\xi) = \mathsf{P}(\xi^{-1}(B)) = \mathsf{P}_\xi(B) = \int_{\mathbb{R}} \varphi d\mathsf{P}_\xi,$$

т. е. равенство (5.2) в этом случае доказано. Если φ — простая борелевская функция $\varphi = \sum_{k=1}^{n} x_k \mathbb{1}_{B_k}$, $B_k \in \mathscr{B}$ и $B_i B_j = \varnothing$ при $i \neq j$, то в силу свойства линейности интеграла снова приходим к равенству (5.2). Для неотрицательной борелевской функции φ равенство (5.2) устанавливается посредством предельного перехода для аппроксимирующей последовательности $(\varphi_n\colon n \in \mathbb{N})$ простых неотрицательных борелевских функций. После этого используется представление $\varphi = \varphi^+ - \varphi^-$. □

Замечание 5.2. В случае абсолютно непрерывного распределения равенство (5.2) принимает вид

$$\mathsf{E}\varphi(\xi) = \int_{\mathbb{R}} \varphi(x) f_\xi(x) dx, \tag{5.3}$$

где f_ξ — плотность распределения случайной величины ξ, а интеграл понимается, вообще говоря, в смысле Лебега, хотя в большей части приложений он совпадает с интегралом Римана.

Интеграл Лебега имеет много преимуществ по сравнению с интегралом Римана. Одним из таких преимуществ являются теоремы о предельном переходе под знаком интеграла. Ниже мы сформулируем две наиболее часто используемые такие теоремы. Вначале сделаем некоторые замечания о терминологии. О свойстве, которое выполняется на событии, имеющем вероятность 1, говорят, что оно выполняется *почти наверное* (сокращенно п. н.). В частности, применительно к сходимости случайных величин введём следующее определение.

Определение 5.5. Будем говорить, что последовательность случайных величин $\xi_1, \xi_2, \ldots$ сходится к случайной величине ξ с вероятностью 1 или *почти наверное*, и писать $\xi_n \overset{\text{п.н.}}{\to} \xi$, если

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

В дальнейшем мы не будем различать случайные величины, которые совпадают почти наверное, и не будем это каждый раз оговаривать.

Теорема 5.4 (О монотонной сходимости). *Пусть* $(\xi_n\colon n \in \mathbb{N})$ — *неубывающая последовательность неотрицательных случайных величин и* $\lim\limits_{n\to\infty} \xi_n = \xi$. *Тогда*

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

либо обе части равенства равны бесконечности.

Для неотрицательной случайной величины ξ математическое ожидание либо существует, либо его можно считать бесконечным. Поэтому условие его существования можно коротко представить неравенством $\mathsf{E}\xi < \infty$. В общем случае условие существования математического ожидания будем записывать в виде $\mathsf{E}|\xi| < \infty$, поскольку $|\xi| = \xi^+ + \xi^-$.

Теорема 5.5 (Лебега о мажорируемой сходимости). *Пусть случайные величины* $\eta, \xi, \xi_1, \xi_2, \ldots$ *удовлетворяют условиям:* $|\xi_n| \leqslant \eta$, $n = 1, 2, \ldots$, $\xi_n \overset{\text{п.н.}}{\to} \xi$ *и* $\mathsf{E}\eta < \infty$. *Тогда*

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

5.3. Характеристические функции. Наряду с функцией распределения F_ξ используются и другие способы описания распределения вероятностей P_ξ случайной величины ξ. Некоторые из них применимы не ко всем распределениям. Например, производящая функция моментов

$$M_\xi(t) = \mathsf{E}e^{t\xi},$$

которая определяется для тех $t \in \mathbb{R}$, для которых в правой части равенства определено математическое ожидание. В случае, когда ξ имеет стандартное нормальное распределение,

$$M_\xi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx = \frac{e^{t^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-t)^2/2} dx = e^{t^2/2}.$$

Заметим также, что здесь в качестве переменной можно было вместо вещественного t взять комплексное z и получить голоморфную во всей комплексной плоскости функцию $M_\xi(z) = e^{z^2/2}$.

В отличие от производящей функции моментов характеристическая функция определена для любого распределения.

Определение 5.6. Пусть ξ — случайная величина. Тогда

$$h_\xi(t) = \mathsf{E}e^{it\xi} = \mathsf{E}(\cos t\xi + i\sin t\xi), \qquad t \in \mathbb{R},$$

называется *характеристической функцией* случайной величины ξ.

В зависимости от характера распределения P_ξ производится вычисление $h_\xi(t)$. Если ξ — простая случайная величина и $\mathsf{P}_\xi(x_k) = p_k$, $k = 1, \ldots, n$, $\sum_{k=1}^n p_k = 1$, то

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

В случае абсолютно непрерывного распределения, когда существует плотность f_ξ, характеристическая функция определяется по формуле

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

Свойства характеристических функций.

1°. $|h_\xi(t)| \leqslant 1$ и $h_\xi(0) = 1$;
2°. $h_\xi(-t) = \overline{h_\xi(t)}$ (здесь черта сверху означает комплексное сопряжение);
3°. Для $a, b \in \mathbb{R}$ и случайной величины ξ выполняется равенство $h_{a\xi+b}(t) = e^{itb}h_\xi(at)$;
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 $\mathbb{R}$;
6°. Если $\mathsf{E}|\xi|^n < \infty$, то $h_\xi(t)$ имеет производные до n-го порядка включительно и

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

7°. Если существует чётная производная $h_\xi^{(2k)}(0)$, то $\mathsf{E}\xi^{2k} < \infty$.

Утверждения $1° - 4°$ являются непосредственными следствиями свойств математического ожидания. Остальные свойства доказываются с использованием теоремы Лебега о мажорируемой сходимости.

Приведём 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 $[a, b]$ распределение. 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 с параметром $\lambda > 0$. Здесь снова воспользуемся определением характеристической функции

$$h_\xi(t) = \lambda\int_0^\infty e^{itx}e^{-\lambda x}dx = \left.\frac{\lambda e^{(it-\lambda)x}}{it-\lambda}\right|_{x=0}^{x=\infty} = \frac{\lambda}{\lambda - it}.$$

♣ Распределение Коши имеет характеристическую функцию

$$h_\xi(t) = e^{-|t|}.$$

Вычисления, связанные с получением этого равенства, проводятся с использованием теории вычетов.

♣ Нормальное распределение с параметрами $a \in \mathbb{R}$ и $\sigma > 0$ имеет характеристическую функцию

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

Действительно, в случае стандартного нормального распределения, когда $a = 0$ и $\sigma = 1$, характеристическая функция получается из производящей функции моментов

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

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

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

С другой стороны,

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

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.

♣ 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.

Между характеристическими функциями и функциями распределения имеет место взаимно-однозначное соответствие. В частности, для любых $a < b$ выполняется соотношение

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

$$= \frac{1}{2}\left[F_\xi(b-0) - F_\xi(a-0) + F_\xi(b) - F_\xi(a)\right].$$

Отметим также, что если характеристическая функция h_ξ абсолютно интегрируема, т. е.

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

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

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

5.4. Неравенства. В основе многих неравенств лежит идея выпуклости. Напомним, что функция $\varphi\colon (a,b) \to \mathbb{R}$ (допускается $a = -\infty, b = \infty$) называется выпуклой, если для всех $x', x'' \in (a,b)$ и $\lambda \in [0,1]$ выполняется неравенство

$$\varphi(\lambda x' + (1-\lambda)x'') \leqslant \lambda\varphi(x') + (1-\lambda)\varphi(x'').$$

Лемма 5.1. *Пусть $\varphi\colon (a,b) \to \mathbb{R}$ — выпуклая функция. Тогда для каждого $x_0 \in (a,b)$ найдётся функция $l(x) = \alpha x + \beta$, $\alpha, \beta \in \mathbb{R}$, такая, что $l(x) \leqslant \varphi(x)$ для всех $x \in (a,b)$ и $l(x_0) = \varphi(x_0)$.*

Доказательство. Пусть $a < x' < x_0 < x'' < b$. Тогда в силу выпуклости функции φ будет выполняться неравенство

$$\varphi(x_0) = \varphi\left(\frac{x''-x_0}{x''-x'}x' + \frac{x_0-x'}{x''-x'}x''\right) \leqslant \frac{x''-x_0}{x''-x'}\varphi(x') + \frac{x_0-x'}{x''-x'}\varphi(x'').$$

Отсюда получаем

$$\frac{x''-x_0}{x''-x'}(\varphi(x_0)-\varphi(x')) \leqslant \frac{x_0-x'}{x''-x'}(\varphi(x'')-\varphi(x_0)),$$

и далее

$$\frac{\varphi(x_0)-\varphi(x')}{x_0-x'} \leqslant \frac{\varphi(x'')-\varphi(x_0)}{x''-x_0}.$$

Поскольку x', x'' выбирались произвольно по разные стороны от x_0, то найдётся α между инфимумом правой части и супремумом левой полученного неравенства. Но тогда для этого α и всех $x \in (a,b)$ будет выполняться неравенство

$$\varphi(x) - \varphi(x_0) \geqslant \alpha(x - x_0).$$

□

Теорема 5.6 (Неравенство Иенсена). *Пусть* $\varphi\colon (a,b)\to\mathbb{R}$ — *выпуклая функция и* ξ — *случайная величина такие, что* $\mathsf{P}(\xi\in(a,b))=1$, $\mathsf{E}|\xi|<\infty$ *и* $\mathsf{E}|\varphi(\xi)|<\infty$. *Тогда выполняется неравенство*

$$\varphi(\mathsf{E}\xi)\ \leqslant\ \mathsf{E}\varphi(\xi).$$

Доказательство. Поскольку $\mathsf{P}(\xi\in(a,b))=1$, то $\mathsf{E}\xi\in(a,b)$. В силу леммы 5.1 найдется функция $l(x)=\alpha x+\beta$ такая, что $l(x)\leqslant\varphi(x)$ для всех $x\in(a,b)$ и $l(\mathsf{E}\xi)=\varphi(\mathsf{E}\xi)$. Но тогда

$$\mathsf{E}\varphi(\xi)\ \geqslant\ \mathsf{E}l(\xi)\ =\ \alpha\mathsf{E}\xi+\beta\ =\ l(\mathsf{E}\xi)\ =\ \varphi(\mathsf{E}\xi),$$

и теорема доказана. □

Выбирая соответствующим образом выпуклую функцию, можно получать различные неравенства. Например, выбор функции $\varphi(x)=|x|$ приводит к неравенству $|\mathsf{E}\xi|\leqslant\mathsf{E}|\xi|$, а в случае $\varphi(x)=x^2$ получаем $(\mathsf{E}\xi)^2\leqslant\mathsf{E}\xi^2$.

Будем говорить, что случайная величина ξ имеет конечный r-тый момент ($r\geqslant 1$), если $\mathsf{E}|\xi|^r<\infty$. Совокупность случайных величин ξ, определённых на вероятностном пространстве $(\Omega,\mathscr{F},\mathsf{P})$ и имеющих конечный r-тый момент, будем обозначать $L_r(\Omega,\mathscr{F},\mathsf{P})$ или, коротко, L_r. Напомним, что мы отождествляем случайные величины, которые совпадают почти наверное. Поэтому под элементом в L_r понимается класс случайных величин, совпадающих почти наверное. Заметим также, что L_r образует линейное пространство. Действительно, если $\mathsf{E}|\xi|^r<\infty$, то для $\alpha\in\mathbb{R}$ имеем $\mathsf{E}|\alpha\xi|^r=|\alpha|^r\mathsf{E}|\xi|^r<\infty$. Замкнутость L_r относительно суммы случайных величин следует из неравенства

$$|\xi+\eta|^r\ \leqslant\ 2^r(|\xi|^r+|\eta|^r).$$

В линейном пространстве L_r вводится норма

$$||\xi||_{L_r}\ :=\ (\mathsf{E}|\xi|^r)^{1/r}.$$

Особо выделяется случай $r=2$, поскольку в L_2 норма определяется скалярным произведением

$$\langle\xi,\eta\rangle\ :=\ \mathsf{E}\xi\eta$$

и L_2 является гильбертовым пространством.

Следующий результат показывает, в частности, что существование r-го момента, $r>1$ влечёт существование моментов q-го порядка для всех $q\in[1,r]$.

Теорема 5.7 (Неравенство Ляпунова). *Пусть* $1\leqslant q<r<\infty$ *и* $\xi\in L_r$. *Тогда* $\xi\in L_q$ *и выполняется неравенство*

$$||\xi||_{L_q}\ \leqslant\ ||\xi||_{L_r}.$$

Доказательство. Пусть $\xi_n=(\min\{|\xi|,n\})^q$. Тогда $(\xi_n\colon n\in\mathbb{N})$ — неубывающая последовательность неотрицательных случайных величин и $\xi_n\to|\xi|^q$, $\xi_n^{r/q}\to|\xi|^r$ при $n\to\infty$. Поскольку $\varphi(x)=x^{r/q}$ является выпуклой функцией на $(0,\infty)$, то из неравенства Иенсена следует $(\mathsf{E}\xi_n)^{r/q}\leqslant\mathsf{E}\xi_n^{r/q}$. Используя теперь теорему 5.4 о монотонной сходимости, получаем

$$\mathsf{E}|\xi|^r\ =\ \lim_{n\to\infty}\mathsf{E}\xi_n^{r/q}\ \geqslant\ \Big(\lim_{n\to\infty}\mathsf{E}\xi_n\Big)^{r/q}\ =\ (\mathsf{E}|\xi|^q)^{r/q},$$

откуда следует утверждение теоремы. □

Из неравенства Ляпунова следует, что если $\mathsf{E}\xi^2 < \infty$, то $\mathsf{E}|\xi| < \infty$ и для случайной величины ξ определена дисперсия $\mathsf{D}\xi = \mathsf{E}\xi^2 - (\mathsf{E}\xi)^2$. Следующие два важных неравенства позволяют оценить вероятность отклонения случайной величины от её математического ожидания через первые два момента.

Теорема 5.8. *Пусть ξ — случайная величина, определённая на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$ и $\varepsilon > 0$. Тогда*

(i) если $\mathsf{E}|\xi| < \infty$, то

$$\mathsf{P}(|\xi| \geqslant \varepsilon) \leqslant \frac{\mathsf{E}|\xi|}{\varepsilon}; \tag{5.4}$$

(ii) если $\mathsf{E}\xi^2 < \infty$, то

$$\mathsf{P}(|\xi - \mathsf{E}\xi| \geqslant \varepsilon) \leqslant \frac{\mathsf{D}\xi}{\varepsilon^2}. \tag{5.5}$$

Доказательство. Из очевидного неравенства

$$|\xi| \geqslant \varepsilon \mathbb{1}_{\{|\xi| \geqslant \varepsilon\}}$$

и свойства монотонности математического ожидания получаем

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

что эквивалентно неравенству (5.4).

Далее, используя неравенство (5.4), получаем

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

□

Неравенство (5.4) называют *неравенством Маркова*, а неравенство (5.5) известно как *неравенство Чебышёва*. Из (5.5) следует широко известное «*правило трёх сигм*». Пусть ξ — случайная величина с $\mathsf{E}\xi = m$ и $\mathsf{D}\xi = \sigma^2$. Тогда

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

Это общее для всех распределений неравенство существенно улучшается для конкретных распределений.

5.5. Совместное распределение и независимость.

Определение 5.7. Пусть $(\Omega, \mathscr{F}, \mathsf{P})$ — вероятностное пространство. Отображение $\xi = (\xi_1, \ldots, \xi_n)\colon \Omega \to \mathbb{R}^n$ называется случайным вектором, если $\xi^{-1}(B) \in \mathscr{F}$ для всех $B \in \mathscr{B}^n$.

В пространстве $\mathbb{R}^n$ борелевская σ-алгебра $\mathscr{B}^n$ порождается π-системой $\mathscr{T}^n$ n-мерных полуинтервалов

$$Q_{(-\infty,x]} = (-\infty, x_1] \times \ldots \times (-\infty, x_n],$$

где $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, т. е. $\sigma(\mathscr{T}^n) = \mathscr{B}^n$.

Предложение 5.2. *Для того чтобы отображение $\xi = (\xi_1 \ldots, \xi_n)\colon \Omega \to \mathbb{R}^n$ было случайным вектором, необходимо и достаточно, чтобы каждая координатная функция $\xi_i\colon \Omega \to \mathbb{R}$, $i = 1, \ldots, n$, являлась случайной величиной.*

Доказательство. Пусть $\xi = (\xi_1 \dots, \xi_n)\colon \Omega \to \mathbb{R}^n$ — случайный вектор. Покажем, что его координатные функции являются случайными величинами. Для удобства записи сделаем это для ξ_1. Выберем произвольно $B \in \mathscr{B}$ и заметим, что

$$\xi_1^{-1}(B) = \{\omega\colon \xi_1(\omega) \in B\} = \{\omega\colon \xi(\omega) \in B \times \mathbb{R}^{n-1}\} \in \mathscr{F}.$$

Таким образом, ξ_1 — случайная величина.

Обратно, пусть $\xi_1, \dots, \xi_n$ — случайные величины и $\xi = (\xi_1, \dots, \xi_n)$. Фиксируем произвольно $x = (x_1, \dots, x_n)$ в $\mathbb{R}^n$. Для каждого $i = 1, \dots, n$ множества $\xi_i^{-1}((-\infty, x_i])$ принадлежат $\mathscr{F}$. Но тогда и

$$\xi^{-1}(Q_{(-\infty,x]}) = \bigcap_{i=1}^{n} \xi^{-1}((-\infty, x_i])$$

также принадлежит $\mathscr{F}$. Это означает, что $\xi^{-1}(\mathscr{T}^n) \subset \mathscr{F}$, а поскольку $\sigma(\mathscr{T}^n) = \mathscr{B}^n$, то рассуждения, аналогичные проведенным при доказательстве предложения 4.2, приводят к включению $\xi^{-1}(\mathscr{B}^n) \subset \mathscr{F}$. □

Предложение 5.3. *Пусть $f\colon \mathbb{R}^n \to \mathbb{R}$ — борелевская функция, а $\xi_1, \dots, \xi_n$ — случайные величины, определённые на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$. Тогда $\eta = f(\xi_1, \dots, \xi_n)$ является случайной величиной.*

Доказательство. Из предложения 5.2 следует, что $\xi = (\xi_1, \dots, \xi_n)$ — случайный вектор. При этом $\eta = f \circ \xi$. Для борелевского множества $B \in \mathscr{B}$ получаем

$$\eta^{-1}(B) = \{\omega\colon f \circ \xi(\omega) \in B\} = \{\omega\colon \xi(\omega) \in f^{-1}(B)\} = \xi^{-1}(f^{-1}(B)) \in \mathscr{F},$$

поскольку $f^{-1}(B) \in \mathscr{B}^n$. □

Как и в одномерном случае, для случайного вектора $\xi = (\xi_1, \dots, \xi_n)$ определяется вероятностная мера $\mathsf{P}_\xi(B) = \mathsf{P}(\xi^{-1}(B))$ на измеримом пространстве $(\mathbb{R}^n, \mathscr{B}^n)$. Эта мера называется *совместным распределением* случайных величин $\xi_1, \dots, \xi_n$. Она также однозначно характеризуется n-мерной функцией распределения

$$F_{\xi_1,\dots,\xi_n}(x_1, \dots, x_n) = \mathsf{P}(\xi_1 \leqslant x_1, \dots, \xi_n \leqslant x_n), \qquad x = (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Здесь также выделяются абсолютно непрерывные распределения, которые определяются посредством плотности

$$F_{\xi_1,\dots,\xi_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\xi_1,\dots,\xi_n}(u_1, \dots, u_n) du_1 \cdot \ldots \cdot du_n.$$

Плотность совместного распределения случайных величин $\xi_1, \dots, \xi_n$ является неотрицательной функцией и

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\xi_1,\dots,\xi_n}(x_1, \dots, x_n) dx_1 \cdot \ldots \cdot dx_n = 1.$$

Предложение 5.4. *Случайные величины $\xi_1, \ldots, \xi_n$, определённые на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$, независимы в том и только том случае, если для всех $x = (x_1, \ldots, x_n)$ из $\mathbb{R}^n$ выполняется равенство*

$$F_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{\xi_i}(x_i). \tag{5.6}$$

Кроме того, если случайные величины $\xi_1, \ldots, \xi_n$ независимы и имеют плотности $f_{\xi_1}, \ldots, f_{\xi_n}$, то случайный вектор $\xi = (\xi_1, \ldots, \xi_n)$ также имеет плотность и

$$f_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{\xi_i}(x_i). \tag{5.7}$$

Доказательство. Если случайные величины $\xi_1, \ldots, \xi_n$ независимы и $x = (x_1, \ldots, x_n)$ из $\mathbb{R}^n$, то

$$\begin{aligned} F_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) &= \mathsf{P}(\xi_1 \leqslant x_1, \ldots, \xi_n \leqslant x_n) \\ &= \prod_{i=1}^{n} \mathsf{P}(\xi_i \leqslant x_i) = \prod_{i=1}^{n} F_{\xi_i}(x_i). \end{aligned}$$

Пусть теперь для всех $x = (x_1, \ldots, x_n)$ из $\mathbb{R}^n$ выполняется равенство (5.6). Выберем произвольно множество индексов $I \subseteq \{1, \ldots, n\}$. Для всех $j \notin I$ в равенстве (5.6) осуществим предельный переход при $x_j \to \infty$. В результате получим

$$\mathsf{P}\left(\bigcap_{i \in I} \{\xi_i \leqslant x_i\}\right) = \prod_{i \in I} \mathsf{P}(\{\xi_i \leqslant x_i\}).$$

Это означает, что события $\{\xi_1 \leqslant x_1\}, \ldots, \{\xi_n \leqslant x_n\}$ независимы. Отсюда следует независимость π-систем $\pi(\xi_1), \ldots, \pi(\xi_n)$, а в силу теоремы 4.4 получаем независимость случайных величин $\xi_1, \ldots, \xi_n$.

Допустим, что $\xi_1, \ldots, \xi_n$ независимы и имеют плотности $f_{\xi_1}, \ldots, f_{\xi_n}$. Тогда

$$\begin{aligned} F_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) &= \prod_{i=1}^{n} F_{\xi_i}(x_i) \\ &= \prod_{i=1}^{n} \int_{-\infty}^{x_i} f_{\xi_i}(u_i) du_i = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} \left(\prod_{i=1}^{n} f_{\xi_i}(u_i)\right) du_1 \cdot \ldots \cdot du_n. \end{aligned}$$

Это означает, что случайный вектор $\xi = (\xi_1, \ldots, \xi_n)$ имеет абсолютно непрерывное распределение, а его плотность определяется равенством (5.7). □

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

$$\mathsf{P}(\{\omega\colon\ (\xi(\omega), \eta(\omega)) \in D\}) = \iint_D f_{\xi,\eta}(x, y) dx dy.$$

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.

§ 6. Законы больших чисел и центральная предельная теорема

Значительная часть классической теории вероятностей и её применений касается предельных теорем, т. е. асимптотического поведения последовательностей случайных величин. При этом выделяются различные виды сходимости. Один из них уже встречался в теоремах о предельном переходе под знаком интеграла Лебега (см. определение 5.5). В этом параграфе мы введём еще три вида сходимости и установим связь между ними. Начнем с характеристики сходимости почти наверное.

Лемма 6.1. *Пусть $\xi, \xi_1, \xi_2, \ldots$ — случайные величины, определённые на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$. Для $\varepsilon > 0$ и $n \in \mathbb{N}$ определим*

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

Тогда $\xi_n \overset{\text{п.н.}}{\to} \xi$ в том и только том случае, если для всех $\varepsilon > 0$ выполняется условие $\mathsf{P}(\overline{\lim\limits_n} A_n^\varepsilon) = 0$.

Доказательство. Условие $\xi_n \overset{\text{п.н.}}{\to} \xi$ можно сформулировать в виде $\mathsf{P}(A) = 0$, где $A = \{\omega\colon\ \xi_n(\omega) \not\to \xi(\omega)\}$. С другой стороны, $\omega \in A$ в том и только том случае, если $\omega \in A^\varepsilon = \overline{\lim\limits_n} A_n^\varepsilon$ при некотором $\varepsilon > 0$. Заметим также, что $A^{\varepsilon_2} \subseteq A^{\varepsilon_1}$, если $\varepsilon_1 < \varepsilon_2$. Следовательно, последовательность $(A^{1/k}\colon\ k \in \mathbb{N})$ является неубывающей и $\lim\limits_k A^{1/k} = A$. В силу свойства непрерывности вероятностной меры

$$\mathsf{P}(A) = \lim_{k\to\infty} \mathsf{P}(A^{1/k}).$$

Но тогда $\mathsf{P}(A) = 0$ в том и только том случае, когда $\mathsf{P}(A^{1/k}) = 0$ при всех $k \in \mathbb{N}$. Последнее эквивалентно тому, что $\mathsf{P}(A^\varepsilon) = 0$ при всех $\varepsilon > 0$. □

Определение 6.1. Будем говорить, что последовательность случайных величин $\xi_1, \xi_2, \ldots$ сходится *по вероятности* к случайной величине ξ, и писать $\xi_n \overset{\mathsf{P}}{\to} \xi$, если для всякого $\varepsilon > 0$ выполняется предельное соотношение

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

Замечание 6.1. Поскольку для любой последовательности множеств выполняется неравенство

$$\mathsf{P}(\overline{\lim_n} A_n) \geqslant \overline{\lim_{n\to\infty}} \mathsf{P}(A_n),$$

то из сходимости последовательности случайных величин почти наверное следует её сходимость по вероятности.

Действительно, сходимость почти наверное эквивалентна условию $\mathsf{P}(\overline{\lim_n} A_n^\varepsilon) = 0$ для всех $\varepsilon > 0$. Условие сходимости по вероятности можно записать в виде $\mathsf{P}(A_n^\varepsilon) \to 0$ при $n \to \infty$ для всех $\varepsilon > 0$.

Теорема 6.1. *Последовательность случайных величин $\xi_1, \xi_2, \ldots$ сходится по вероятности к случайной величине ξ в том и только том случае, если всякая её подпоследовательность содержит сходящуюся почти наверное подпоследовательность.*

Доказательство. Допустим, что $\xi_n \xrightarrow{\mathsf{P}} \xi$, и покажем, что можно выделить подпоследовательность $(\xi_{n_k}\colon\ k \in \mathbb{N})$, которая сходится к ξ почти наверное. Для $k = 1$ выберем n_1 так, чтобы выполнялось условие $\mathsf{P}(|\xi_{n_1} - \xi| \geqslant 1) < 1/2$. Далее выберем n_k так, чтобы $n_k > n_{k-1}$ и $\mathsf{P}(|\xi_{n_k} - \xi| \geqslant 1/k) < 1/2^k$, $k = 2, 3, \ldots$. Пусть теперь $\varepsilon > 0$. Выберем N так, чтобы $1/N < \varepsilon$. Но тогда

$$\begin{aligned}
\sum_{k=1}^\infty \mathsf{P}(A_{n_k}^\varepsilon) &= \sum_{k=1}^\infty \mathsf{P}(|\xi_{n_k} - \xi| \geqslant \varepsilon) \\
&\leqslant \sum_{k=1}^N \mathsf{P}(|\xi_{n_k} - \xi| \geqslant \varepsilon) + \sum_{k=N+1}^\infty \mathsf{P}(|\xi_{n_k} - \xi| \geqslant 1/k) \\
&\leqslant \sum_{k=1}^N \mathsf{P}(A_{n_k}^\varepsilon) + \sum_{k=N+1}^\infty \frac{1}{2^k} < \infty,
\end{aligned}$$

и по лемме Бореля-Кантелли (теорема 2.4) $\mathsf{P}(\overline{\lim_k} A_{n_k}^\varepsilon) = 0$. Следовательно, $\xi_{n_k} \xrightarrow{\text{п.н.}} \xi$ при $k \to \infty$.

Допустим теперь, что из всякой подпоследовательности последовательности $(\xi_n\colon\ n \in \mathbb{N})$ можно выделить подпоследовательность, сходящуюся почти наверное к ξ. Фиксируем произвольно $\varepsilon > 0$ и пусть $\alpha_n = \mathsf{P}(A_n^\varepsilon)$. По предположению из любой подпоследовательности $(\alpha_{n_k}\colon\ k \in \mathbb{N})$ можно выделить подпоследовательность, которая стремится к нулю. Это влечёт стремление к нулю и самой последовательности, т. е. $\mathsf{P}(A_n^\varepsilon) \to 0$ при $n \to \infty$. □

Лемма 6.2. *Пусть $\xi_n \xrightarrow{\mathsf{P}} \xi$ и $f\colon \mathbb{R} \to \mathbb{R}$ — непрерывная функция. Тогда $f(\xi_n) \xrightarrow{\mathsf{P}} f(\xi)$.*

Доказательство. Заметим, что если $\xi_{n_k} \xrightarrow{\text{п.н.}} \xi$, то в силу непрерывности f имеем также $f(\xi_{n_k}) \xrightarrow{\text{п.н.}} f(\xi)$. Но тогда утверждение леммы следует из теоремы 6.1. □

Теорема 6.2 (Закон больших чисел Чебышёва). *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*

$$\frac{1}{n}(S_n - \mathsf{E}S_n) \xrightarrow{\mathsf{P}} 0,$$

где $S_n = \xi_1 + \ldots + \xi_n$.

Доказательство. Пусть $\varepsilon > 0$. Используя неравенство Чебышёва (5.5), 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 $\mathsf{D}S_n = \mathsf{D}\xi_1 + \ldots + \mathsf{D}\xi_n \leqslant nC$, и

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

при $n \to \infty$. □

Один из первых законов больших чисел был получен Бернулли.

Теорема 6.3. *Пусть* S_n *— число успехов в серии из* n *независимых испытаний с вероятностью* $p, 0 < p < 1$, *успеха в отдельном испытании. Тогда* $\frac{S_n}{n} \xrightarrow{\mathsf{P}} p$.

Доказательство. Пусть ξ_k — индикатор успеха в k-том испытании. Тогда $\xi_1, \xi_2, \ldots$ — независимые одинаково распределённые случайные величины с $\mathsf{E}\xi_k = p$ и $\mathsf{D}\xi_k = p(1-p)$. Поэтому выполнены условия теоремы 6.2, из которой следует доказываемое утверждение. □

Заметим, что доказанная теорема даёт теоретическое обоснование того, что вероятность события проявляется через частоту его появления.

Теоремы Чебышёва и Бернулли относят к *слабым* законам больших чисел, поскольку в них утверждается сходимость по вероятности средних к математическому ожиданию. Теоремы, в которых утверждается сходимость почти наверное, называют *усиленными* законами больших чисел. Приведём один такой результат.

Теорема 6.4. *Пусть* $\xi_1, \xi_2, \ldots$ *— независимые одинаково распределённые случайные величины с* $\mathsf{E}\xi_n = a$ *и* $\mathsf{D}\xi_n = \sigma^2$. *Тогда*

$$\frac{S_n}{n} \overset{\text{п.н.}}{\to} a,$$

где $S_n = \xi_1 + \ldots + \xi_n$.

Доказательство. Заметим вначале, что доказательство достаточно провести в предположении $\xi_n \geqslant 0$. В противном случае можно перейти к последовательностям $(\xi_n^+ \colon n \in \mathbb{N})$ и $(\xi_n^- \colon n \in \mathbb{N})$. Поэтому предположим, что $\xi_n \geqslant 0$ и для $\varepsilon > 0$ определим

$$A_n^\varepsilon = \left\{\omega \colon \left|\frac{S_n(\omega)}{n} - a\right| \geqslant \varepsilon\right\},$$

$n = 1, 2, \ldots$. Из неравенства Чебышёва (5.5) следует, что $\mathsf{P}(A_n^\varepsilon) \leqslant \sigma^2/(\varepsilon^2 n)$. Рассмотрим подпоследовательность событий $(A_{k^2}^\varepsilon\colon\ k \in \mathbb{N})$ и заметим, что

$$\sum_{k=1}^{\infty} \mathsf{P}(A_{k^2}^\varepsilon) \leqslant \sum_{k=1}^{\infty} \frac{\sigma^2}{\varepsilon^2 k^2} < \infty.$$

В силу леммы Бореля-Кантелли 2.4 $\mathsf{P}(\overline{\lim_k}\, A_{k^2}^\varepsilon) = 0$. Поскольку это выполняется для любого $\varepsilon > 0$, то из леммы 6.1 следует

$$\frac{S_{k^2}}{k^2} \xrightarrow{\text{п.н.}} a.$$

Пусть

$$E = \left\{\omega\colon\ \lim_{k\to\infty} \frac{S_{k^2}(\omega)}{k^2} = a\right\}.$$

По доказанному $\mathsf{P}(E) = 1$. Допустим теперь, что $\omega \in E$ и $k^2 \leqslant n < (k+1)^2$. Тогда в силу неотрицательности случайных величин будут выполняться неравенства

$$\frac{S_{k^2}(\omega)}{(k+1)^2} \leqslant \frac{S_n(\omega)}{n} \leqslant \frac{S_{(k+1)^2}(\omega)}{k^2}$$

или, что эквивалентно,

$$\frac{k^2}{(k+1)^2}\frac{S_{k^2}(\omega)}{k^2} \leqslant \frac{S_n(\omega)}{n} \leqslant \frac{(k+1)^2}{k^2}\frac{S_{(k+1)^2}(\omega)}{(k+1)^2}.$$

Поскольку при $n \to \infty$ имеем также $k \to \infty$ и

$$\frac{k^2}{(k+1)^2} \to 1, \qquad \frac{S_{k^2}(\omega)}{k^2} \to a,$$

то $S_n(\omega)/n \to a$. Таким образом, $S_n/n \xrightarrow{\text{п.н.}} a$, поскольку $\mathsf{P}(E) = 1$. □

В действительности условие конечности второго момента (существование дисперсии) требовать необязательно. Однако тогда доказательство теоремы несколько усложнится.

Определение 6.2. Пусть случайные величины $\xi, \xi_1, \xi_2, \ldots$ имеют конечные r-тые моменты. Будем говорить, что последовательность $(\xi_n\colon\ n \in \mathbb{N})$ сходится к ξ *в среднем порядка* r и писать $\xi_n \xrightarrow{L_r} \xi$, если $\mathsf{E}|\xi_n - \xi|^r \to 0$ при $n \to \infty$.

Наиболее важные частные случаи: $r = 1$ — сходимость в среднем и $r = 2$ — сходимость в среднем квадратическом.

Лемма 6.3. *Пусть $\xi_n \xrightarrow{L_r} \xi$ при некотором r. Тогда $\xi_n \xrightarrow{\mathsf{P}} \xi$.*

Доказательство. Пусть $\varepsilon > 0$. Используя неравенство Маркова (5.4), получаем

$$\mathsf{P}(|\xi_n - \xi| \geqslant \varepsilon) = \mathsf{P}(|\xi_n - \xi|^r \geqslant \varepsilon^r) \leqslant \frac{1}{\varepsilon^r}\mathsf{E}|\xi_n - \xi|^r \to 0$$

при $n \to \infty$. □

Замечание 6.2. В условиях теоремы 6.4 $\frac{S_n}{n} \xrightarrow{L_2} a$.

Доказательство. Действительно, в силу независимости случайных величин $\mathsf{E}S_n = n\sigma^2$ и

$$\mathsf{E}\left(\frac{S_n}{n} - a\right)^2 = \mathsf{D}\frac{S_n}{n} = \frac{\sigma^2}{n} \to 0$$

при $n \to \infty$. □

Как отмечалось ранее, между множеством функций распределения F и множеством характеристических функций h имеет место взаимно-однозначное соответствие. Это соответствие оказывается и непрерывным относительно соответствующих сходимостей.

Определение 6.3. Будем говорить, что последовательность функций распределения $(F_n\colon n \in \mathbb{N})$ *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 6.5. *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 $F_n \Rightarrow F$.

В приложениях важную роль играют результаты, утверждающие нормальность распределения больших сумм независимых случайных величин. Обычно их объединяют в одну группу с названием *центральна предельная теорема.* Рассмотрим один из вариантов центральной предельной теоремы.

Теорема 6.6. *Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин с конечными $\mathsf{E}\xi_n = a$ и $\mathsf{D}\xi_n = \sigma^2$. 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 6.5 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
$$h(t) \,=\, h_{\hat{\xi}_n}(t) \,=\, \mathsf{E}e^{it\hat{\xi}_n}.$$

Из свойств характеристических функций следует, что h является дифференцируемой до второго порядка функцией. Поскольку $h(0) = 1$, $h'(0) = i\mathsf{E}\hat{\xi}_n = 0$, $h''(0) = -\mathsf{E}\hat{\xi}_n^2 = -1$, то по формуле Тейлора
$$h(t) \,=\, 1 - \frac{1}{2}t^2 + o(t^2)$$

при $t \to 0$. 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
$$\lim_{n\to\infty} h_{\zeta_n}(t) \,=\, \lim_{n\to\infty} e^{n\ln(1 - \frac{t^2}{2n} + o(\frac{t^2}{n}))} \,=\, e^{-t^2/2}.$$

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

Следствие 6.1 (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. □

Определение 6.4. 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 $F_{\xi_n} \Rightarrow F_\xi$.

Значение сходимости по распределению обусловлено уже тем, что вид предельного распределения определяет все важнейшие характеристики предельной случайной величины. Ранее было установлено, что из сходимости почти наверное, а также из сходимости в среднем, следует сходимость по вероятности.

Теорема 6.7. *Если $\xi_n \xrightarrow{\mathsf{P}} \xi$, то $\xi_n \xrightarrow{d} \xi$.*

Доказательство. Фиксируем произвольно $\varepsilon > 0$ и определим
$$A_n^\varepsilon \,=\, \{\omega\colon\ |\xi_n(\omega) - \xi(\omega)| \geqslant \varepsilon\}.$$

Из условия $\xi_n \xrightarrow{\mathsf{P}} \xi$ следует, что $\mathsf{P}(A_n^\varepsilon) \to 0$ при $n \to \infty$. Пусть $F_n = F_{\xi_n}$ и $F = F_\xi$. Заметим, что для $x \in \mathbb{R}$ выполняются неравенства

$$F(x-\varepsilon) \leqslant \varliminf_{n\to\infty} F_n(x) \leqslant \varlimsup_{n\to\infty} F_n(x) \leqslant F(x+\varepsilon), \qquad x \in \mathbb{R}. \tag{6.1}$$

Действительно, это следует из включений

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

и условия $\mathsf{P}(A_n^\varepsilon) \to 0$ при $n \to \infty$. Если x — точка непрерывности функции F, то предельный переход в (6.1) при $\varepsilon \to 0$ приводит к неравенствам

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

откуда следует, что $F_n(x) \to F(x)$ при $n \to \infty$. Это означает слабую сходимость $F_{\xi_n} \Rightarrow F_\xi$. Теорема доказана. □

Метод характеристических функций позволяет получить слабый закон больших чисел для независимых одинаково распределённых случайных величин без условия существования вторых моментов.

Теорема 6.8 (Закон больших чисел Хинчина). *Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин с $\mathsf{E}\xi_n = a$. 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. □

Завершим этот параграф одной теоремой Бернштейна о равномерном приближении непрерывной на отрезке функции полиномами, которая хорошо иллюстрирует связи между теорией вероятностей и анализом.

Пусть f — вещественная функция, непрерывная на отрезке $[0,1]$. В силу теоремы Кантора она равномерно непрерывна на $[0,1]$ и ее модуль непрерывности

$$\omega(\delta; f) = \sup\{|f(x') - f(x'')| \colon |x'-x''| \leqslant \delta, \quad x', x'' \in [0,1]\}$$

стремится к нулю при $\delta \searrow 0$. Кроме того,

$$\sup\{|f(x)| \colon x \in [0,1]\} = M < \infty.$$

Для $n = 1, 2, \ldots$ определим полиномы Бернштейна

$$B_n(x; f) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) C_n^k x^k (1-x)^{n-k}.$$

Их структура связана с законом больших чисел для схемы Бернулли. Действительно, пусть S_n — количество успехов в схеме Бернулли с вероятностью $x \in [0,1]$ успеха в отдельном испытании. Тогда

$$B_n(x; f) = \mathsf{E} f\left(\frac{S_n}{n}\right).$$

Теорема 6.9 (Бернштейна). *Для любой непрерывной на* $[0,1]$ *функции* f *выполняется предельное соотношение*

$$\max_{x\in[0,1]} |B_n(x;f)-f(x)| \to 0$$

при $n\to\infty$.

Доказательство. Пусть $\delta>0$ и $n\in\mathbb{N}$. Тогда

$$\begin{aligned} |B_n(x;f)-f(x)| &= \left|\sum_{k=0}^{n} C_n^k x^k(1-x)^{n-k}\left(f\left(\frac{k}{n}\right)-f(x)\right)\right| \\ &\leqslant \omega(\delta;f)\sum_{k:\,|\frac{k}{n}-x|\leqslant\delta} C_n^k x^k(1-x)^{n-k} \\ &+ 2M\sum_{k:\,|\frac{k}{n}-x|>\delta} C_n^k x^k(1-x)^{n-k}. \end{aligned}$$

Поскольку $\mathsf{D}S_n = nx(1-x)\leqslant \frac{1}{4}n$, то в силу неравенства Чебышева получаем

$$\sum_{k:\,|\frac{k}{n}-x|>\delta} C_n^k x^k(1-x)^{n-k} = \mathsf{P}\left(\left|\frac{S_n}{n}-x\right|>\delta\right) \leqslant \frac{\mathsf{D}\frac{S_n}{n}}{\delta^2} \leqslant \frac{1}{4\delta^2 n}.$$

Следовательно,

$$\max_{x\in[0,1]} |B_n(x;f)-f(x)| \leqslant \omega(\delta;f)+\frac{M}{2\delta^2 n}.$$

Пусть теперь $\varepsilon>0$. Выберем $\delta>0$ так, чтобы $\omega(\delta;f)<\varepsilon/2$. Затем выберем N так, чтобы выполнялось неравенство

$$\frac{M}{2\delta^2 N} < \frac{\varepsilon}{2}.$$

Но тогда при $n\geqslant N$ и всех $x\in[0,1]$ будет выполняться неравенство

$$|B_n(x;f)-f(x)| \leqslant \frac{\varepsilon}{2}+\frac{\varepsilon}{2} = \varepsilon$$

и теорема доказана. □

§ 7. Задача оценивания и условные математические ожидания

В ряде случаев возникает ситуация, когда по наблюдением за одной или несколькими случайными величинами нужно делать предположение о ненаблюдаемой случайной величине. Другими словами, по информации, которую нам даёт наблюдение за одними случайными величинами, необходимо давать прогноз на другие случайные величины. При этом в качестве меры отклонения прогноза от истинного значения часто рассматривается математическое ожидание квадрата разности прогноза и ненаблюдаемой случайной величины.

Поэтому мы начнем со случая, когда все рассматриваемые случайные величины имеют конечный второй момент. В этом случае полезными оказываются методы гильбертова пространства.

Пусть $(\Omega, \mathscr{F}, \mathsf{P})$ — вероятностное пространство. Через $L_2(\Omega, \mathscr{F}, \mathsf{P})$ (или просто L_2) будем обозначать линейное пространство случайных величин ξ, определённых на $(\Omega, \mathscr{F}, \mathsf{P})$ и имеющих конечный второй момент, т. е. $\mathsf{E}\xi^2 < \infty$. Напомним, что мы отождествляем случайные величины, совпадающие почти наверное, в один элемент пространства L_2. В этом пространстве вводится скалярное произведение $\langle \xi, \eta \rangle = \mathsf{E}\xi\eta$ и норма $||\xi|| = \sqrt{\langle \xi, \xi \rangle} = \sqrt{\mathsf{E}\xi^2}$. Все аксиомы скалярного произведения (следовательно, и нормы) выполняются. Кроме того, пространство L_2 является полным, т. е. всякая фундаментальная последовательность в нём сходится. Если $\xi_1, \ldots, \xi_n$ — произвольная система элементов в L_2, то через $\mathscr{L}(\xi_1, \ldots, \xi_n)$ будем обозначать линейную оболочку этой системы, т. е. совокупность всех линейных комбинаций $\sum_{k=1}^n \alpha_k \xi_k$, $\alpha_k \in \mathbb{R}$. Две случайные величины ξ и η из L_2 называются ортогональными и пишут $\xi \perp \eta$, если $\langle \xi, \eta \rangle = 0$. Важным результатом в геометрии гильбертовых пространств является так называемая теорема о проекции.

Теорема 7.1. *Пусть $\mathscr{M}$ — замкнутое подпространство в L_2 и $\xi \in L_2$. Тогда существует единственный элемент $\widehat{\xi} \in \mathscr{M}$ такой, что*

$$||\xi - \widehat{\xi}|| = \inf\{||\xi - \eta||\colon \eta \in \mathscr{M}\}.$$

При этом $(\xi - \widehat{\xi}) \perp \mathscr{M}$.

Элемент $\widehat{\xi}$ называют ортогональной проекцией элемента ξ на подпространство $\mathscr{M}$. Заметим, что ортогональной проекцией ξ на подпространство постоянных случайных величин является постоянная $\mathsf{E}\xi$.

Рассмотрим теперь задачу о наилучшей аппроксимации случайной величины $\eta \in L_2$ линейными комбинациями случайных величин $\xi_1, \ldots, \xi_n$ из L_2. Более точно, нас интересует ортогональная проекция $\widehat{\eta}$ случайной величины η на подпространство $\mathscr{M} = \mathscr{L}(1, \xi_1, \ldots, \xi_n)$. Эта проекция должна иметь вид

$$\widehat{\eta} = \mu + \sum_{k=1}^n \alpha_k \xi_k, \qquad \alpha_k \in \mathbb{R}, \quad k = 1, \ldots, n,$$

и удовлетворять соотношениям

$$\langle \eta - \widehat{\eta}, 1 \rangle = 0, \qquad \langle \eta - \widehat{\eta}, \xi_k \rangle = 0, \qquad k = 1, \ldots, n.$$

Первое равенство означает, что $\mathsf{E}\eta = \mathsf{E}\widehat{\eta}$ и, следовательно,

$$\mu = \mathsf{E}\eta - \sum_{k=1}^n \alpha_k \mathsf{E}\xi_k,$$

а саму проекцию можно записать в виде

$$\widehat{\eta} = \mathsf{E}\eta + \sum_{k=1}^n \alpha_k (\xi_k - \mathsf{E}\xi_k).$$

Далее, используя свойства скалярного произведения и условие $\langle \eta - \widehat{\eta}, 1\rangle = 0$, получаем

$$\begin{aligned}\langle \eta - \widehat{\eta}, \xi_k\rangle &= \langle \eta - \widehat{\eta}, \xi_k - \mathsf{E}\xi_k\rangle \\ &= \langle \eta - \mathsf{E}\eta - \sum_{k=1}^{n} \alpha_i(\xi_i - \mathsf{E}\xi_i), \xi_k - \mathsf{E}\xi_k\rangle \\ &= \mathsf{cov}(\eta, \xi_k) - \sum_{i=1}^{n} \alpha_i \mathsf{cov}(\xi_i, \xi_k).\end{aligned}$$

Таким образом, равенство $\langle \eta - \widehat{\eta}, \xi_k\rangle = 0$ можно записать в виде

$$\mathsf{cov}(\eta, \xi_k) = \sum_{i=1}^{n} \alpha_i \mathsf{cov}(\xi_i, \xi_k).$$

Полученные соотношения удобно записать в матричном виде. Пусть $\mathbb{V} = (v_{i,j})$ — ковариационная матрица случайного вектора $(\xi_1, \ldots, \xi_n)$. Через $\mathbb{a}$ обозначим вектор с координатами неизвестных коэффициентов $\alpha_1, \ldots, \alpha_n$, а через $\mathbb{c}$ вектор с координатами $c_i = \mathsf{cov}(\eta, \xi_i)$, $i = 1, \ldots, n$. Тогда полученные соотношения принимают вид $\mathbb{V}\mathbb{a} = \mathbb{c}$.

Если случайные величины $\xi_1, \ldots, \xi_n$ образуют линейно независимую систему, то матрица $\mathbb{V}$ строго положительно определена и $\det \mathbb{V} \neq 0$. Действительно, если для некоторого ненулевого вектора $(x_1, \ldots, x_n)$ из $\mathbb{R}^n$ выполняется равенство

$$\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j v_{ij} = 0,$$

то (см. (3.1))

$$\mathsf{E}\left(\sum_{i=1}^{n} x_i(\xi_i - \mathsf{E}\xi_i)\right)^2 = 0,$$

откуда следует, что $\sum_{i=1}^{n} x_i(\xi_i - \mathsf{E}\xi_i) = 0$ почти наверное. Это означает, что одну из случайных величин $\xi_1, \ldots, \xi_n$ можно представить как линейную комбинацию остальных. Поэтому в задаче линейного оценивания можно считать, что $\det \mathbb{V} \neq 0$. Но тогда матрица $\mathbb{V}$ обратима и вектор $\mathbb{a}$ с компонентами неизвестных коэффициентов однозначно определяется $\mathbb{a} = \mathbb{V}^{-1}\mathbb{c}$.

Особый интерес представляет случай $n = 1$. Пусть ξ — наблюдаемая случайная величина и $\mathsf{D}\xi \neq 0$. Тогда наилучшая линейная оценка для случайной величины η будет определяться как $\widehat{\eta} = \alpha\xi + \mu$, где

$$\mu = \mathsf{E}\eta - \alpha\mathsf{E}\xi, \qquad \alpha = \frac{\mathsf{cov}(\xi, \eta)}{\mathsf{D}\xi}.$$

Таким образом,

$$\widehat{\eta} = \mathsf{E}\eta + \frac{\mathsf{cov}(\xi, \eta)}{\mathsf{D}\xi}(\xi - \mathsf{E}\xi).$$

Это так называемое *уравнение линейной регрессии* η на ξ. Величина $\mathsf{E}(\eta - \widehat{\eta})^2$ называется среднеквадратической ошибкой оценивания. В последнем случае

$n=1$

$$\mathsf{E}(\eta-\widehat{\eta})^2 \;=\; \mathsf{E}\left(\eta-\mathsf{E}\eta-\frac{\mathsf{cov}(\xi,\eta)}{\mathsf{D}\xi}(\xi-\mathsf{E}\xi)\right)^2 \;=\; \mathsf{D}\eta(1-\varrho^2(\xi,\eta)).$$

Решение задачи нелинейного оценивания, т. е. поиск φ в классе всех борелевских функций $\varphi\colon\ \mathbb{R}\to\mathbb{R}$ (или в многомерном случае $\varphi\colon\ \mathbb{R}^n\to\mathbb{R}$), которая минимизировала бы $\mathsf{E}(\eta-\varphi(\xi))^2$ (соответственно, $\mathsf{E}(\eta-\varphi(\xi_1,\dots,\xi_n))^2$), связано с понятием условного математического ожидания. Если $\varphi\colon\ \mathbb{R}\to\mathbb{R}$ — борелевская функция, то случайная величина $\varphi(\xi)$ называется *оценкой* η по ξ. Условие того, что случайная величина $\widehat{\eta}$ является функцией от ξ можно сформулировать следующим образом.

Теорема 7.2. *Пусть ξ и η — две случайные величины, определённые на вероятностном пространстве $(\Omega,\mathscr{F},\mathsf{P})$, и η является $\sigma(\xi)$-измеримой, т. е. $\sigma(\eta)\subseteq\sigma(\xi)$. Тогда найдётся такая борелевская функция $\varphi\colon\ \mathbb{R}\to\mathbb{R}$, что $\eta=\varphi(\xi)$.*

Доказательство. Вначале заметим, что случайную величину η можно считать ограниченной. В противном случае можно перейти к $\operatorname{arctg}\eta$.

Обозначим через Φ_ξ совокупность случайных величин, которые допускают представление в виде $\varphi(\xi)$, где $\varphi\colon\ \mathbb{R}\to\mathbb{R}$ — борелевская функция. Нам нужно показать, что $\eta\in\Phi_\xi$.

Пусть $A\in\sigma(\xi)$, что означает $A=\xi^{-1}(B)$ для некоторого борелевского множества B. Тогда $\mathbb{1}_A\in\Phi_\xi$, поскольку $\mathbb{1}_A=\mathbb{1}_B(\xi)$, а $\mathbb{1}_B\colon\ \mathbb{R}\to\mathbb{R}$ — борелевская функция. Поскольку Φ_ξ является линейным пространством (замкнуто относительно линейных операций), то все простые $\sigma(\xi)$-измеримые случайные величины принадлежат Φ_ξ. Допустим теперь, что $\eta\geqslant 0$. В силу аппроксимационной теоремы 5.1 найдётся неубывающая последовательность простых неотрицательных $\sigma(\xi)$-измеримых случайных величин $(\eta_n\colon\ n\in\mathbb{N})$, сходящаяся к η. Для каждой η_n найдётся борелевская функция φ_n такая, что $\eta_n=\varphi_n(\xi)$. Но тогда

$$\varphi(x) \;=\; \varlimsup_{n\to\infty}\varphi_n(x)$$

также будет борелевской функцией. При этом для всех $\omega\in\Omega$ будем иметь

$$\varphi(\xi(\omega)) \;=\; \lim_{n\to\infty}\varphi_n(\xi(\omega)) \;=\; \lim_{n\to\infty}\eta_n(\omega) \;=\; \eta(\omega),$$

т. е. $\eta\in\Phi_\xi$. В общем случае доказательство завершается с помощью представления $\eta=\eta^+-\eta^-$ и применением доказанного к η^+ и η^-. □

Замечание 7.1. Аналогичный результат имеет место и для n случайных величин $\xi_1,\dots,\xi_n$. Если η является $\sigma(\xi_1,\dots,\xi_n)$-измеримой, то найдётся борелевская функция $\varphi\colon\ \mathbb{R}^n\to\mathbb{R}$ такая, что $\eta=\varphi(\xi_1,\dots,\xi_n)$.

Итак, вопрос о наилучшей оценке случайной величины η сводится к отысканию случайной величины $\widehat{\eta}$, которая минимизирует $\mathsf{E}(\eta-\zeta)^2$ в классе всех случайных величин $\zeta\in L_2$, которые являются $\sigma(\xi)$-измеримыми.

Пусть теперь $\mathscr{G}$ — произвольная σ-алгебра подмножеств Ω и $\mathscr{G}\subset\mathscr{F}$. Тогда $L_2(\Omega,\mathscr{G},\mathsf{P})$ будет замкнутым подпространством гильбертова пространства

$L_2(\Omega, \mathscr{F}, \mathsf{P})$. Если η — случайная величина из $L_2(\Omega, \mathscr{F}, \mathsf{P})$, то по теореме о проекции 7.1 существует единственный элемент $\widehat{\eta}$ в $L_2(\Omega, \mathscr{G}, \mathsf{P})$, который является наилучшим приближением случайной величины η, т. е. $||\eta - \widehat{\eta}|| \leqslant ||\eta - \zeta||$ для всех $\zeta \in L_2(\Omega, \mathscr{G}, \mathsf{P})$. Случайная величина $\widehat{\eta}$ называется условным математическим ожиданием η относительно σ-алгебры $\mathscr{G}$, а если $\mathscr{G} = \sigma(\xi)$, то $\widehat{\eta}$ — условное математическое ожидание η относительно случайной величины ξ и $\widehat{\eta} = \varphi(\xi)$, где φ — некоторая борелевская функция. Условное математическое ожидание даёт более детальную характеристику зависимости между случайными величинами, чем ковариация. Кроме того, условное математическое ожидание можно ввести и в случае, когда ξ и η не имеют конечных вторых моментов.

Отметим вначале характерные свойства случайной величины $\widehat{\eta}$, полученной как ортогональная проекция η на подпространство $L_2(\Omega, \mathscr{G}, \mathsf{P})$. Кроме $\mathscr{G}$-измеримости $\widehat{\eta}$ обладает следующим свойством. Для любого $G \in \mathscr{G}$ выполняется равенство

$$\mathsf{E}\eta\mathbb{1}_G = \mathsf{E}\widehat{\eta}\mathbb{1}_G,$$

которое является следствием ортогональности $\eta - \widehat{\eta}$ подпространству $L_2(\Omega, \mathscr{G}, \mathsf{P})$. Однако для требования выполнения этих условий достаточно существования лишь первого момента.

Теорема 7.3. *Пусть η — случайная величина с конечным математическим ожиданием, определённая на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$, и $\mathscr{G}$ — некоторая σ-алгебра, содержащаяся в $\mathscr{F}$. Тогда существует единственная (с точностью до равенства п. н.) случайная величина $\widehat{\eta}$, удовлетворяющая условиям:*

(i) $\mathsf{E}|\widehat{\eta}| < \infty$;

(ii) $\widehat{\eta}$ *является $\mathscr{G}$-измеримой;*

(iii) для любого $G \in \mathscr{G}$ выполняется равенство $\mathsf{E}\eta\mathbb{1}_G = \mathsf{E}\widehat{\eta}\mathbb{1}_G$.

Доказательство. (Единственность) Пусть $\widehat{\eta}_1$ и $\widehat{\eta}_2$ удовлетворяют условиям (i)—(iii). Нам нужно доказать, что $\mathsf{P}(\widehat{\eta}_1 \neq \widehat{\eta}_2) = 0$. Допустим противное. Тогда хотя бы одно из событий $\{\widehat{\eta}_1 - \widehat{\eta}_2 > 0\}$, $\{\widehat{\eta}_2 - \widehat{\eta}_1 > 0\}$ имеет положительную вероятность. Пусть для определённости $\mathsf{P}(\{\widehat{\eta}_1 - \widehat{\eta}_2 > 0\}) > 0$, и рассмотрим монотонно возрастающую последовательность событий $A_n = \{\widehat{\eta}_1 - \widehat{\eta}_2 > 1/n\}$, $n = 1, 2, \ldots$. В силу свойства непрерывности вероятностной меры

$$\lim_{n\to\infty} \mathsf{P}(A_n) = \mathsf{P}(\widehat{\eta}_1 - \widehat{\eta}_2 > 0) > 0.$$

Но тогда найдётся номер n_0 такой, что $\mathsf{P}(A_{n_0}) > 0$, и

$$\mathsf{E}(\widehat{\eta}_1 - \widehat{\eta}_2)\mathbb{1}_{A_{n_0}} \geqslant \frac{1}{n_0}\mathsf{P}(A_{n_0}) > 0.$$

Однако $A_{n_0} \in \mathscr{G}$, поскольку $\widehat{\eta}_1$ и $\widehat{\eta}_2$ являются измеримыми, и полученное неравенство противоречит условию (iii).

(Существование) Допустим вначале, что $\eta \geqslant 0$. Тогда в силу аппроксимационной теоремы 5.1 найдётся неубывающая последовательность простых неотрицательных случайных величин $(\eta_n\colon n \in \mathbb{N})$, которая сходится к η. Поскольку при каждом n случайная величина η_n принадлежит $L_2(\Omega, \mathscr{F}, \mathsf{P})$, то для неё найдётся измеримая случайная величина $\widehat{\eta}_n$ такая, что для любого $G \in \mathscr{G}$ выполняется равенство

$$\mathsf{E}\eta_n \mathbb{1}_G = \mathsf{E}\widehat{\eta}_n \mathbb{1}_G.$$

Из этого равенства и неотрицательности η_n следует, что и $\widehat{\eta}_n \geqslant 0$ п. н. Заметим также, что поскольку $\eta_{n+1} - \eta_n \geqslant 0$, то и $\widehat{\eta}_{n+1} - \widehat{\eta}_n \geqslant 0$ п. н. для всех $n = 1, 2, \ldots$. В силу монотонности последовательности $(\widehat{\eta}_n\colon n \in \mathbb{N})$ определена почти наверное неотрицательная измеримая случайная величина

$$\widehat{\eta} = \varlimsup_{n\to\infty} \widehat{\eta}_n.$$

Для любого $G \in \mathscr{G}$ по теореме 5.4 о монотонной сходимости получаем

$$\mathsf{E}\eta \mathbb{1}_G = \lim_{n\to\infty} \mathsf{E}\eta_n \mathbb{1}_G = \lim_{n\to\infty} \widehat{\eta}_n \mathbb{1}_G = \mathsf{E}\widehat{\eta} \mathbb{1}_G.$$

Таким образом, $\widehat{\eta}$ удовлетворяет условиям (i)—(iii).

В общем случае доказательство завершается представлением η в виде разности двух неотрицательных случайных величин $\eta = \eta^+ - \eta^-$. □

Определение 7.1. Пусть η — случайная величина, определённая на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$, с конечным математическим ожиданием и пусть $\mathscr{G}$ — σ-алгебра, содержащаяся в $\mathscr{G}$. *Условным математическим ожиданием η относительно $\mathscr{G}$* называется $\mathscr{G}$-измеримая случайная величина $\zeta := \mathsf{E}(\eta|\mathscr{G})$ такая, что $\mathsf{E}\zeta \mathbb{1}_G = \mathsf{E}\eta \mathbb{1}_G$ для всех $G \in \mathscr{G}$.

Корректность этого определения обеспечивает теорема 7.3.

Если σ-алгебра $\mathscr{G}$ порождена случайной величиной ξ, $\mathscr{G} = \sigma(\xi)$, то $\mathsf{E}(\eta|\sigma(\xi))$ называют условным математическим ожиданием η относительно ξ и обозначают $\mathsf{E}(\eta|\xi)$. В этом случае, как следует из теоремы 7.2, $\mathsf{E}(\eta|\xi) = \varphi(\xi)$, где $\varphi\colon \mathbb{R} \to \mathbb{R}$ — борелевская функция. При этом совершенно не важно какие значения φ принимает вне $\xi(\Omega)$. Поэтому обычно полагают $\varphi(x) = 0$ при $x \notin \xi(\Omega)$.

Допустим теперь, что ξ — простая случайная величина, и пусть

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

где $\mathscr{D}_\xi = \{D_1, \ldots, D_n\}$ — разбиение. В этом случае $\xi(\Omega) = \{x_1, \ldots, x_n\}$, а функция φ определяется из представления

$$\zeta = \mathsf{E}(\eta|\xi) = \sum_{i=1}^{n} \varphi(x_i) \mathbb{1}_{D_i}.$$

Поскольку $D_i \in \sigma(\xi)$, $i = 1, \ldots, n$, то из определения условного математического ожидания следует

$$\mathsf{E}\eta \mathbb{1}_{D_i} = \mathsf{E}\zeta \mathbb{1}_{D_i} = \mathsf{E}\left(\mathbb{1}_{D_i} \sum_{k=1}^{n} \varphi(x_k) \mathbb{1}_{D_k}\right) = \varphi(x_i) \mathsf{P}(D_i).$$

В силу того, что мы не различаем случайные величины, совпадающие почти наверное, то можно считать, что все атомы разбиения $\mathscr{D}_\xi$ имеют положительную вероятность. Тогда из равенства выше находим, что

$$\varphi(x_i) \;=\; \frac{1}{\mathsf{P}(D_i)}\mathsf{E}\eta\mathbb{1}_{D_i}.$$

В связи с этим используется обозначение $\varphi(x_i) = \mathsf{E}(\eta|\xi = x_i)$, т. е. $\varphi(x)$ интерпретируется как математическое ожидание случайной величины η при условии $\xi = x$.

Допустим теперь, что совместное распределение ξ и η описывается плотностью $f_{\xi,\eta}(x,y)$. В этом случае событие $\{\xi = x\}$ имеет нулевую вероятность для любого $x \in \mathbb{R}$. Чтобы понять как должна выглядеть функция φ в этом случае, заметим, что определяющим для неё является равенство

$$\mathsf{E}\eta\mathbb{1}_A \;=\; \mathsf{E}\varphi(\xi)\mathbb{1}_A, \tag{7.1}$$

которое должно выполняться для всех $A \in \sigma(\xi)$. Поскольку $\sigma(\xi) = \xi^{-1}(\mathscr{B})$, то $\mathbb{1}_A = \mathbb{1}_B(\xi)$, $B \in \mathscr{B}$, и равенство (7.1) можно переписать в виде

$$\mathsf{E}\eta\mathbb{1}_B(\xi) \;=\; \mathsf{E}\varphi(\xi)\mathbb{1}_B(\xi). \tag{7.2}$$

Пусть $f_\xi(x)$ и $f_\eta(y)$ — плотности распределения вероятностей случайных величин ξ и η. Определим плотность условного распределения вероятностей

$$f_{\eta|\xi}(y|x) \;=\; \begin{cases} \dfrac{f_{\xi,\eta}(x,y)}{f_\xi(x)}, & \text{если} \quad f_\xi(x) \neq 0, \\ 0, & \text{если} \quad f_\xi(x) = 0. \end{cases}$$

Тогда

$$\varphi(x) \;=\; \mathsf{E}(\eta|\xi = x) \;=\; \int\limits_{\mathbb{R}} y f_{\eta|\xi}(y|x)dy. \tag{7.3}$$

Действительно, если $B \in \mathscr{B}$, то с использованием теоремы Фуббини получаем

$$\begin{aligned} \mathsf{E}\eta\mathbb{1}_B(\xi) \;&=\; \iint\limits_{\mathbb{R}\times\mathbb{R}} y\mathbb{1}_B(x) f_{\xi,\eta}(x,y)dxdy \\ &=\; \int\limits_{\mathbb{R}} \mathbb{1}_B(x) f_\xi(x) \left(\int\limits_{\mathbb{R}} y f_{\eta|\xi}(y|x)dy\right) dx \\ &=\; \int\limits_{\mathbb{R}} \mathbb{1}_B(x) f_\xi(x)\varphi(x)dx \\ &=\; \mathsf{E}\varphi(\xi)\mathbb{1}_B(\xi), \end{aligned}$$

т. е. для функции φ из (7.3) и любого $B \in \mathscr{B}$ выполняется равенство (7.2). Переход к $f_{\eta|\xi}(y|x)$ в интеграле обоснован, поскольку из равенства

$$f_\xi(x) \;=\; \int\limits_{\mathbb{R}} f_{\xi,\eta}(x,y)dy$$

видно, что если $f_\xi(x)=0$, то $f_{\xi,\eta}(x,y)=0$ для почти всех y.

Свойства условного математического ожидания. Поскольку условные математические ожидания являются случайными величинами, то равенства с ними следует понимать как равенства почти наверное. Как и обычное математическое ожидание, условное математическое ожидание обладает свойством линейности, которое проверяется непосредственно. Свойство монотонности ($\mathsf{E}(\eta|\mathscr{G})\geqslant 0$, если $\eta\geqslant 0$) было установлено при доказательстве теоремы 7.3.

Теорема 7.4 (Неравенство Иенсена). *Пусть η — случайная величина, а $\varphi\colon \mathbb{R}\to\mathbb{R}$ — выпуклая функция, такие, что $\mathsf{E}|\eta|<\infty$ и $\mathsf{E}|\varphi(\eta)|<\infty$. Тогда*

$$\varphi(\mathsf{E}(\eta|\mathscr{G}))\;\leqslant\;\mathsf{E}(\varphi(\eta)|\mathscr{G}).$$

Доказательство. Заметим, что выпуклую функцию можно представить в виде $\varphi(x)=\sup l(x)$, где супремум берется по всем функциям l вида $l(x)=ax+b$ и удовлетворяющих условию $l(x)\leqslant\varphi(x)$. Для любой такой функции l имеем

$$\mathsf{E}(\varphi(\eta)|\mathscr{G})\;\geqslant\;\mathsf{E}(l(\eta)|\mathscr{G})\;=\;l(\mathsf{E}(\eta|\mathscr{G})),$$

и, переходя в правой части этого неравенства к супремуму, приходим к неравенству Иенсена. □

Теорема 7.5. *Пусть ξ,η — случайные величины с конечными математическими ожиданиями, а $\mathscr{G},\mathscr{H}$ — σ-алгебры, содержащиеся в $\mathscr{F}$. Тогда*

(i) если η и $\mathscr{G}$ независимы, то $\mathsf{E}(\eta|\mathscr{G})=\mathsf{E}\eta$;

(ii) если $\mathscr{G}\subseteq\mathscr{H}$, то $\mathsf{E}(\mathsf{E}(\eta|\mathscr{H})|\mathscr{G})=\mathsf{E}(\eta|\mathscr{G})$;

(iii) если ξ $\mathscr{G}$-измерима и $\mathsf{E}|\xi\eta|<\infty$, то $\mathsf{E}(\xi\eta|\mathscr{G})=\xi\mathsf{E}(\eta|\mathscr{G})$.

Доказательство. (i) Нужно проверить, что $\zeta\equiv\mathsf{E}\eta$ удовлетворяет всем свойствам условного математического ожидания $\mathsf{E}(\eta|\mathscr{G})$. Условия $\mathsf{E}|\eta|<\infty$ и $\mathscr{G}$-измеримости очевидны. Пусть $G\in\mathscr{G}$. Тогда в силу свойства мультипликативности математического ожидания

$$\mathsf{E}\eta\mathbb{1}_G\;=\;\mathsf{E}\eta\cdot\mathsf{E}\mathbb{1}_G\;=\;\mathsf{E}\zeta\mathbb{1}_G.$$

(ii) Пусть $\zeta_1=\mathsf{E}(\eta|\mathscr{H})$, $\zeta_2=\mathsf{E}(\eta|\mathscr{G})$. Нужно показать, что $\mathsf{E}(\zeta_1|\mathscr{G})=\zeta_2$. Выберем произвольно $G\in\mathscr{G}$. Тогда G принадлежит также и σ-алгебре $\mathscr{H}$. Следовательно,

$$\mathsf{E}\zeta_1\mathbb{1}_G\;=\;\mathsf{E}\eta\mathbb{1}_G\;=\;\mathsf{E}\zeta_2\mathbb{1}_G.$$

(iii) Пусть $\zeta=\mathsf{E}(\eta|\mathscr{G})$. Нужно показать, что $\mathsf{E}(\xi\eta|\mathscr{G})=\xi\zeta$. Поскольку ξ и η можно представить в виде разности неотрицательных случайных величин, то в силу линейности условного математического ожидания можно считать, что $\xi\geqslant 0$, $\eta\geqslant 0$. При этом остаётся проверить, что для любого $G\in\mathscr{G}$ выполняется равенство

$$\mathsf{E}(\xi\zeta\mathbb{1}_G)\;=\;\mathsf{E}(\xi\eta\mathbb{1}_G). \tag{7.4}$$

Если $\xi=\mathbb{1}_A$, то в силу $\mathscr{G}$-измеримости $A\in\mathscr{G}$. Но тогда $AG\in\mathscr{G}$ и

$$\mathsf{E}(\xi\zeta\mathbb{1}_G)\;=\;\mathsf{E}(\mathbb{1}_A\zeta\mathbb{1}_G)\;=\;\mathsf{E}(\zeta\mathbb{1}_{AG})\;=\;\mathsf{E}(\eta\mathbb{1}_{AG})\;=\;\mathsf{E}(\xi\eta\mathbb{1}_G).$$

В силу линейности математического ожидания равенство (7.4) будет выполняться и для простых $\mathscr{G}$-измеримых случайных величин ξ. Из аппроксимационной теоремы 5.1 следует, что существует неубывающая последовательность простых неотрицательных случайных величин $(\xi_n\colon\ n \in \mathbb{N})$, которая сходится к ξ, и каждая ξ_n является $\mathscr{G}$-измеримой. Поэтому для каждого $n \in \mathbb{N}$ будет выполняться равенство

$$\mathsf{E}(\xi_n \zeta \mathbb{1}_G) \ = \ \mathsf{E}(\xi_n \eta \mathbb{1}_G). \tag{7.5}$$

В силу свойства монотонности условного математического ожидания $\zeta \geqslant 0$. Следовательно, $(\xi_n \zeta \mathbb{1}_G\colon\ n \in \mathbb{N})$ — неубывающая последовательность неотрицательных случайных величин, сходящаяся к $\xi\zeta\mathbb{1}_G$, а $(\xi_n \eta \mathbb{1}_G\colon\ n \in \mathbb{N})$ — такая же последовательность, сходящаяся к $\xi\eta\mathbb{1}_G$. Поэтому, осуществляя предельный переход при $n \to \infty$ в равенстве (7.5) в силу теоремы 5.4 о монотонной сходимости получаем равенство (7.4). □

§ 8. Цепи Маркова

Пусть некоторая система меняет свое состояние случайным образом в моменты времени $t = 1, 2, \dots, T$. Множество возможных состояний системы является конечным множеством $E = \{e_1, \dots, e_r\}$ и называется *фазовым пространством*. Среди реальных систем важный класс образуют такие, у которых вероятности перехода из одного состояния в другое в данный момент времени не зависит от того, как вела себя система в предыдущие моменты времени. Такие системы называют *марковскими* или *цепями Маркова*. Эволюция изучаемой системы описывается траекторией $\omega = (\omega_0, \omega_1, \dots, \omega_T)$, где $\omega_t = i$, если система в момент времени t находилась в состоянии e_i. Поэтому под элементарным событием будем понимать $\omega = (\omega_0, \omega_1, \dots, \omega_T)$, т. е. последовательность номеров состояний, в которых находилась система в моменты времени $0, 1, \dots, T$. Можно считать, что цепь Маркова является обобщением схемы независимых испытаний, в которой условие независимости заменяется на некоторые другие естественные предположения. Рассмотрим эти предположения и введем некоторые понятия, связанные с цепями Маркова.

Для вычисления вероятности того, что траектория $\omega = (\omega_0, \omega_1, \dots, \omega_T)$ будет реализована как $i_0, i_1, \dots, i_T$, можно воспользоваться теоремой умножения

$$\begin{aligned}&\mathsf{P}(\omega_0 = i_0, \dots, \omega_T = i_T) \ = \\ &\mathsf{P}(\omega_0 = i_0)\mathsf{P}(\omega_1 = i_1 |\, \omega_0 = i_0) \cdot \ldots \cdot \mathsf{P}(\omega_T = i_T |\, \omega_0 = i_0, \dots, \omega_{T-1} = i_{T-1}).\end{aligned}$$

Здесь мы считаем, что условная вероятность относительно события с нулевой вероятностью равна нулю.

Условие марковости (независимость от прошлого) выражается в том, что для любых $s < t$ выполняются равенства

$$\mathsf{P}(\omega_t = j |\, \omega_0 = i_0, \dots, \omega_{s-1} = i_{s-1}, \omega_s = i) \ = \ \mathsf{P}(\omega_t = j |\, \omega_s = i),$$

$i, j = 1, \dots, r$. Другими словами, вероятность перехода системы, находившейся в состоянии i в момент времени s, в состояние j в момент времени t не зависит от того, как она себя вела до момента времени s.

Кроме того, мы будем рассматривать *однородные* марковские цепи, для которых условные вероятности перехода из состояния i в состояние j за время t не зависят от того, в какой момент времени s она находилась в состоянии i, т. е.

$$\mathsf{P}(\omega_{s+t} = j |\, \omega_s = i) \;=\; \mathsf{P}(\omega_t = j |\, \omega_0 = i) \;=\; p_{ij}(t).$$

Вероятности $p_{ij}(t)$ обладают следующими свойствами:

1°. $p_{ij}(t) \geqslant 0$;

2°. $\sum\limits_{j=1}^{r} p_{ij}(t) \;=\; 1$;

3°. $p_{ij}(0) \;=\; \delta_{ij}$.

Матрицу, составленную из вероятностей $p_{ij}(t)$, будем обозначать $\Pi(t)$. Свойства 1° и 2° означают, что $\Pi(t)$ является *стохастической матрицей.* Вероятности перехода системы из состояния i в состояние j за единичное время $p_{ij}(1)$ называются *переходными вероятностями* и обозначаются p_{ij}. Матрица $\Pi(1) = \Pi$ называется *матрицей переходных вероятностей.*

Однородная марковская цепь вполне определяется вектором-строкой $\overrightarrow{p}(0) = (p_1(0), \ldots, p_r(0))$ начальных вероятностей ($p_i(0)$ — вероятность того, что в начальный момент времени система находилась в состоянии i) и матрицей переходных вероятностей Π. Действительно, вероятность $p_j(t)$ того, что система будет находиться в состоянии j в момент времени t определяется равенством

$$p_j(t) \;=\; \mathsf{P}(\omega_t = j) \;=\; \sum_{i=1}^{r} p_i(0) p_{ij}(t).$$

Если $\overrightarrow{p}(t)$ — вектор-строка вероятностей состояний в момент времени t, то

$$\overrightarrow{p}(t) \;=\; \overrightarrow{p}(0)\Pi(t).$$

С другой стороны, матрица $\Pi(t)$ выражается через Π. Это следует из уравнений Колмогорова — Чепмена

$$p_{ij}(s+t) \;=\; \sum_{k=1}^{r} p_{ik}(s) p_{kj}(t), \tag{8.1}$$

которые выводятся непосредственно из свойств марковости и однородности

$$\begin{aligned}
p_{ij}(s+t) \;&=\; \mathsf{P}(\omega_{s+t} = j |\, \omega_0 = i) \\
&=\; \frac{\mathsf{P}(\omega_{s+t} = j, \omega_0 = i)}{\mathsf{P}(\omega_0 = i)} \\
&=\; \sum_{k=1}^{r} \frac{\mathsf{P}(\omega_{s+t} = j, \omega_0 = i, \omega_s = k)}{\mathsf{P}(\omega_0 = i)} \\
&=\; \sum_{k=1}^{r} \frac{\mathsf{P}(\omega_s = k, \omega_0 = i)\mathsf{P}(\omega_{s+t} = j |\, \omega_0 = i, \omega_s = k)}{\mathsf{P}(\omega_0 = i)} \\
&=\; \sum_{k=1} p_{ik}(s) p_{kj}(t).
\end{aligned}$$

Уравнения (8.1) можно также записать в матричном виде

$$\Pi(s+t) = \Pi(s)\Pi(t).$$

Поскольку $\Pi(1) = \Pi$, то $\Pi(t) = \Pi^t$. Следовательно,

$$\overrightarrow{p}(t) = \overrightarrow{p}(0)\Pi^t.$$

Важное значение в теории однородных марковских цепей имеет следующий результат.

Теорема 8.1. *[О предельных вероятностях.] Пусть при некотором $t_0 > 0$ все элементы матрицы $\Pi^{t_0} = (p_{ij}(t_0))$ являются строго положительными. Тогда для каждого $j = 1, \ldots, r$ существует предел*

$$p_j = \lim_{t\to\infty} p_{ij}(t).$$

При этом $p_j > 0$, $j = 1, \ldots, r$, не зависят от i и являются единственным решением системы

$$\sum_{k=1}^{r} x_k p_{kj} = x_j, \qquad \sum_{k=1}^{r} x_k = 1. \tag{8.2}$$

Доказательство. Введем в рассмотрение

$$M_j(t) = \max_{1\leqslant i\leqslant r} p_{ij}(t), \qquad m_j(t) = \min_{1\leqslant i\leqslant r} p_{ij}(t).$$

В силу уравнений (8.1)

$$p_{ij}(1+t) = \sum_{k=1}^{r} p_{ik} p_{kj}(t).$$

Отсюда и непосредственно из определения $m_j(t)$ и $M_j(t)$ получаем неравенства

$$m_j(t) \leqslant p_{ij}(t+1) \leqslant M_j(t).$$

Из этих неравенств, в свою очередь, следуют неравенства

$$m_j(t) \leqslant m_j(t+1) \leqslant M_j(t+1) \leqslant M_j(t).$$

Это означает, что $m_j(t)$ не убывает по t, а $M_j(t)$ не возрастает по t.

Обозначим

$$\varepsilon = \min_{1\leqslant i,j\leqslant r} p_{ij}(t_0).$$

По условию теоремы $0 < \varepsilon < 1$. Для $t > 0$, используя уравнения (8.1), получаем

$$\begin{aligned} p_{ij}(t_0+t) &= \sum_{k=1}^{r} p_{ik}(t_0) p_{kj}(t) \\ &= \sum_{k=1}^{r} [p_{ik}(t_0) - \varepsilon p_{jk}(t)] p_{kj}(t) + \varepsilon \sum_{k=1}^{r} p_{jk}(t) p_{kj}(t) \\ &= \sum_{k=1}^{r} [p_{ik}(t_0) - \varepsilon p_{jk}(t)] p_{kj}(t) + \varepsilon p_{jj}(2t). \end{aligned}$$

Замечая, что выражения в квадратных скобках неотрицательны, получаем неравенство

$$p_{ij}(t_0+t) \leqslant M_j(t)\sum_{k=1}^{r}[p_{ik}(t_0) - \varepsilon p_{jk}(t)] + \varepsilon p_{jj}(2t) = (1-\varepsilon)M_j(t) + \varepsilon p_{jj}(2t),$$

из которого следует, что

$$M_j(t_0+t) \leqslant (1-\varepsilon)M_j(t) + \varepsilon p_{jj}(2t).$$

Аналогично получаются неравенства

$$p_{ij}(t_0+t) \geqslant (1-\varepsilon)m_j(t) + \varepsilon p_{jj}(2t)$$

и

$$m_j(t_0+t) \geqslant (1-\varepsilon)m_j(t) + \varepsilon p_{jj}(2t).$$

Сравнивая полученные выше неравенства, приходим к следующему

$$M_j(t_0+t) - m_j(t_0+t) \leqslant (1-\varepsilon)(M_j(t) - m_j(t)).$$

Применяя последовательно k раз последнее неравенство, получаем

$$M_j(kt_0+t) - m_j(kt_0+t) \leqslant (1-\varepsilon)^k(M_j(t) - m_j(t)).$$

Отсюда сразу же следует, что

$$\lim_{k\to\infty}[M_j(kt_0+t) - m_j(kt_0+t)] = 0,$$

а поскольку последовательность $\{M_j(t) - m_j(t)\}_{t=1}^{\infty}$ монотонно убывает, то

$$\lim_{t\to\infty}[M_j(t) - m_j(t)] = 0.$$

Таким образом, доказано существование пределов

$$p_j = \lim_{t\to\infty} M_j(t) = \lim_{t\to\infty} m_j(t) = \lim_{t\to\infty} p_{ij}(t),$$

$j = 1, \ldots, r$. Заметим также, что поскольку $m_j(t)$ не убывают по t, то

$$p_j \geqslant m_j(t_0) \geqslant \varepsilon > 0,$$

т. е. все p_j строго положительны. При этом

$$\sum_{j=1}^{r} p_j = \lim_{t\to\infty}\sum_{j=1}^{r} p_{ij}(t) = 1.$$

Осуществляя в равенстве

$$p_{ij}(t+1) = \sum_{k=1} p_{ik}(t)p_{kj}$$

предельный переход при $t \to \infty$, получаем

$$p_j = \sum_{k=1}^{r} p_k p_{kj},$$

т. е. $p_1, \dots, p_r$ — решение системы (8.2).

Допустим теперь, что $q_1, \dots, q_r$ также является решением этой системы, т. е.

$$q_j = \sum_{k=1}^{r} q_k p_{kj}, \qquad \sum_{k=1}^{r} q_k = 1.$$

Заметим, что тогда $q_1, \dots, q_r$ будет решением и системы

$$q_j = \sum_{k=1}^{r} q_k p_{kj}(t) \tag{8.3}$$

при $t = 1, 2, \dots$. Действительно,

$$\sum_{k=1}^{r} q_k p_{kj}(t) = \sum_{k=1}^{r} q_k \sum_{l=1}^{r} p_{kl} p_{lj}(t-1) = \sum_{l=1}^{r} q_l p_{lj}(t-1) = \dots = \sum_{k=1}^{r} q_k p_{kj} = q_j.$$

Осуществляя в равенстве (8.3) предельный переход при $t \to \infty$, получаем

$$q_j = \sum_{k=1}^{r} q_k p_j = p_j$$

и единственность решения системы (8.2) доказана. □

Доказанная теорема относится к так называемым эргодическим теоремам. Ее смысл состоит в том, что за длительный промежуток времени система как бы забывает, из какого начального состояния она стартовала.

Условие положительности элементов матрицы Π^{t_0} означает, что с положительной вероятностью из любого состояния в любое состояние система может перейти за время t_0.

§ 9. Ветвящиеся процессы

Немного истории. История возникновения теории ветвящихся процессов связывается с публикацией Гальтона 1873 года. Постановке задачи предшествовали следующие замечания: „Исчезновение фамилий лиц, которые занимали видное положение в истории, это факт, неоднократно отмечавшийся в прошлом; по этому поводу строились разные догадки... Слишком многочисленны были примеры фамилий, которые, будучи распространенными, становились редкими или совсем исчезали“. Чтобы разобраться в природе исчезновения фамилий, Гальтон поставил задачу создания математической модели, описывающей этот процесс. Такую модель построил Ватсон в 1874 году. Математическая модель Гальтона и Ватсона была вскоре забыта в связи с элементарной ошибкой Ватсона, которая приводила к выводу, что каждая фамилия должна выродиться даже в том случае, если средняя численность популяции растет от поколения к поколению. Более подробно об этом можно почитать в монографии Т. Харриса „Теория ветвящихся случайных процессов“ . Интересно, что в монографии 1934 года Н. Н. Семенов использовал модель Гальтона—Ватсона как элемент своей теории химических (неядерных) цепных реакций. За эти исследования Н. Н. Семенов в 1956 году получил Нобелевскую премию. Сам термин „ветвящийся процесс“ впервые появился в работе А. Н. Колмогорова и Н. А. Дмитриева 1947 года.

Описание модели. Первоначально задача формулировалась следующим образом. Пусть $p_0, p_1, p_2, \ldots$ — вероятности того, что отец имеет соответственно $0, 1, 2, \ldots$ сыновей, каждый из которых с теми же вероятностями может иметь своих сыновей и т. д. Какова вероятность, что мужская линия выродится к n-ому поколению?

В общем случае можно рассматривать однотипные частицы, под которыми могут пониматься люди, животные, бактерии или нейтроны в цепных реакциях. Вероятности p_k, $k = 0, 1, \ldots$, будут связываться с событиями, что одна частица в следующем поколении превращается в k частиц. Будем считать, что частицы размножаются независимо друг от друга и для каждой из них вероятности p_k, $k = 0, 1, \ldots$, одни и те же.

Обозначим через $\xi_0, \xi_1, \xi_2, \ldots$ число частиц в нулевом, первом, втором,... поколениях соответственно. Другими словами, ξ_n — объем популяции в n-ом поколении. Будем считать (без особого ограничения общности), что $\xi_0 = 1$, т. е. начальное поколение состоит из одной частицы. Объем последующих поколений $\xi_1, \xi_2, \ldots$ будет случайным. Заметим, что при сделанных предположениях $\{p_0, p_1, \ldots\}$ будет распределением вероятностей случайной величины ξ_1, т. е. $\mathsf{P}(\xi_1 = k) = p_k$, $k = 0, 1, \ldots$. Поскольку частицы размножаются независимо друг от друга и число потомков каждой из них имеет одно и то же распределение вероятностей, то распределение случайной величины ξ_n должно определяться из начальных условий.

Важной характеристикой ветвящегося процесса является среднее число непосредственных потомков одной частицы $\mathsf{E}\xi_1 = \sum_{k=0}^{\infty} k p_k = m$. В зависимости от значения этой величины ветвящийся процесс называется:

- докритическим, если $m < 1$;
- критическим, если $m = 1$;
- надкритическим, если $m > 1$ (или $m = \infty$).

Через σ^2 будем также обозначать дисперсию $\mathsf{D}\xi_1$. Пусть A — событие, которое состоит в том, что процесс $\xi_0 = 1, \xi_1, \xi_2, \ldots$ в конце концов выродится. Его вероятность $q = \mathsf{P}(A)$ будем называть *вероятностью вырождения* процесса. Ближайшей нашей целью будет найти зависимость q от начальных данных.

Производящая функция процесса. Поскольку $\xi_1, \xi_2, \ldots$ являются целочисленными случайными величинами, то их распределения вероятностей естественно представлять производящими функциями. В частности, обозначим

$$f(x) = g_{\xi_1}(x) = \sum_{k=0}^{\infty} p_k x^k$$

и будем называть f производящей функцией процесса. Для отыскания производящих функций g_{ξ_n}, $n = 2, 3, \ldots$, нам потребуется следующий результат.

Теорема 9.1. *Пусть η_1, η_2, ... — независимые одинаково распределенные целочисленные случайные величины с производящей функцией $g_\eta(x)$ и ν — независимая от них целочисленная случайная величина с производящей функцией $g_\nu(x)$. Тогда производящая функция целочисленной случайной величины $S_\nu = \eta_1 + \ldots + \eta_\nu$ $(S_0 = 0)$ определяется равенством*

$$g_{S_\nu}(x) \;=\; g_\nu \circ g_\eta(x).$$

Доказательство. Непосредственно из определения производящей функции с использованием свойств математического ожидания получаем

$$\begin{aligned} g_{S_\nu}(x) \;&=\; \mathsf{E} x^{S_\nu} \;=\; \sum_{k=0}^{\infty} \mathsf{E}\left(x^{S_\nu} \mathbb{1}_{\{\nu=k\}}\right) \\ &=\; \sum_{k=0}^{\infty} \mathsf{E}\left(x^{\eta_1+\ldots+\eta_k} \mathbb{1}_{\{\nu=k\}}\right) \\ &=\; \sum_{k=0}^{\infty} \left(g_\eta(x)\right)^k \mathsf{P}(\nu = k) \;=\; g_\nu\left(g_\eta(x)\right). \end{aligned}$$

□

Теорема 9.2. *Пусть $\xi_0 = 1$, ξ_1, ξ_2, ... — ветвящийся процесс Гальтона—Ватсона с производящей функцией $g_{\xi_1}(x) = f(x)$. Тогда производящая функция $g_{\xi_n}(x)$ представляет собой n-ую итерацию f^n функции f, т. е.*

$$g_{\xi_n}(x) \;=\; f^n(x) \;=\; f \circ \ldots \circ f(x).$$

Доказательство. Представим n-ое поколение частиц в виде $\xi_n = \eta_1 + \ldots + \eta_{\xi_{n-1}}$, где η_1 — число потомков от первой частицы, η_2 — от второй, ..., $\eta_{\xi_{n-1}}$ — от последней из $(n-1)$-го поколения. Тогда по предыдущей теореме

$$g_{\xi_n}(x) \;=\; g_{\xi_{n-1}} \circ g_\eta(x) \;=\; g_{\xi_{n-1}} \circ f(x).$$

По индукции получаем требуемый результат. □

Вероятность вырождения. Пусть A_n — событие, которое заключается в том, что n-ое поколение отсутствует, т. е. $A_n = \{\xi_n = 0\}$. Тогда $A_n \nearrow A = \cup_{n=1}^{\infty} A_n$. Поэтому в силу непрерывности вероятностной меры

$$q \;=\; \mathsf{P}(A) \;=\; \lim_{n\to\infty} \mathsf{P}(A_n).$$

Теорема 9.3. *Вероятность q вырождения процесса Гальтона—Ватсона с производящей функцией $f(x)$ определяется равенством*

$$q \;=\; \lim_{b\to\infty} f^n(0)$$

и является наименьшим неотрицательным корнем уравнения

$$f(x) \;=\; x.$$

Доказательство. Поскольку $\mathsf{P}(\xi_n = 0) = f^n(0)$, то $f^n(0) \nearrow q$. Осуществляя предельный переход при $n \to \infty$ в равенстве

$$f(f^n(0)) = f^{n+1}(0),$$

получаем $f(q) = q$, т. е. q — корень уравнения $f(x) = x$. Остается показать, что q — наименьший неотрицательный корень уравнения $f(x) = x$. Допустим противное, т. е. $f(\alpha) = \alpha$ и $0 \leqslant \alpha < q$. Тогда в силу монотонности функции f будем иметь

$$f(0) \leqslant f(\alpha) = \alpha < q = f(q).$$

Снова, используя монотонность функции f, получаем

$$f^2(0) \leqslant \alpha < q.$$

Повторение этого процесса приводит к неравенствам

$$f^n(0) \leqslant \alpha < q, \qquad n = 2, 3, \ldots.$$

Но тогда

$$\lim_{n\to\infty} f^n(0) \leqslant \alpha,$$

что противоречит определению q. □

Теорема 9.4. *Если процесс Гальтона—Ватсона критический или докритический, то $q = 1$. В случае надкритического процесса $0 \leqslant q < 1$.*

Доказательство. Допустим, что $m = f'(1) \leqslant 1$. Если процесс не вырожденный, то $f'(x) < 1$ при $0 \leqslant x < 1$. Фиксируем произвольно $x_0 \in [0, 1)$. По теореме о среднем найдется такое $x^* \in (x_0, 1)$, что

$$1 - f(x_0) = f'(x^*)(1 - x_0).$$

Но тогда

$$1 - f(x_0) < 1 - x_0$$

и $f(x_0) > x_0$. Т. о. в интервале $[0, 1)$ корней уравнение $f(x) = x$ не имеет.

Если $m = f'(1) > 1$, то найдется такое $x_0 \in (0, 1)$, что $f'(x_0) > 1$. Используя теорему о среднем, получаем

$$1 - f(x_0) > 1 - x_0,$$

т. е. $f(x_0) - x_0 < 0$. Если $f(0) = 0$, то $q = 0$. В противном случае $f(0) > 0$ и функция $\varphi(x) = f(x) - x$ принимает на концах промежутка $[0, x_0]$ значения разных знаков. Следовательно, в этом промежутке найдется решение уравнения $f(x) = x$. □

Пример геометрического распределения. Рассмотрим случай, когда число потомков от одной частицы в следующем поколении имеет геометрическое распределение. Эта модель ветвящегося процесса зависит от одного параметра и допускает явный вид итераций производящей функции.

Допустим, что $\mathsf{P}(\xi_1 = k) = p(1-p)^k$, $k = 0, 1, 2, \ldots$. Тогда производящая функция процесса $f(x) = \mathsf{E}x^{\xi_1}$ будет дробно-линейной функцией

$$f(x) = \sum_{k=0}^{\infty} p(1-p)^k x^k = \frac{p}{1-(1-p)x}.$$

Здесь p, $0 < p < 1$, является параметром, определяющим процесс Гальтона—Ватсона. При этом

$$m = \mathsf{E}\xi_1 = f'(1) = \frac{1-p}{p}.$$

Следовательно, в случае $0 < p < 1/2$ процесс будет надкритическим ($m > 1$), $p = 1/2$ соответствует критическому процессу, а при $1/2 < p < 1$ мы получаем докритический процесс. В надкритическом случае вероятность вырождения процесса (корень уравнения $f(x) = x$) выражается равенством

$$q = \frac{p}{1-p} = \frac{1}{m}.$$

Для получения явного вида итераций производящей функции f представим её в виде

$$f(x) = \frac{1}{m+1-mx} = 1 - \frac{m(1-x)}{1+m(1-x)}.$$

Тогда производящая функция числа частиц в n-том поколении будет определяться по формуле

$$f^n(x) = 1 - \frac{m^n(1-x)}{1+(1-x)\sum\limits_{k=1}^{n} m^k},$$

$n = 1, 2, \ldots$. Проверку этого равенства легко провести с использованием метода математической индукции.

§ 10. Броуновское движение

В 1785 г. Ян Ингенхауз (Jan Ingenhousz) обнаружил хаотическое движение частицы в жидкости. При наблюдении в микроскопе взвеси цветочной пыльцы в 1827 г. Р. Браун (Robert Brown) переоткрыл это явление. В результате многочисленных наблюдений он пришел к ряду выводов. Движение частицы не ослабевает со временем и не зависит от химических свойств среды, его интенсивность увеличивается с ростом температуры среды и с уменьшением её вязкости и размеров частицы.

Количественная теория броуновского движения была дана Эйнштейном и Смолуховским 1905 и 1906 гг. на основе молекулярно-кинетической теории. Выводы этой теории были подтверждены измерениями Ж. Перрена и Т. Сведберга в 1906 г. Наконец, в 1923 г. Н. Винер построил строгую математическую модель броуновского движения. Интересно, что в 1900 г. к модели броуновского движения пришёл Л. Башелье при описании динамики цен акций. В настоящее время трудно переоценить роль броуновского движения в стохастическом анализе и приложениях.

Простейший способ представления броуновского движения состоит в описании « инфинитезимального случайного блуждания». Именно так оно возникает в различных приложениях. Рассмотрим частицу, которая движется влево или вправо с равной вероятностью на расстояние Δx. Она это делает через каждый промежуток времени Δt. Другими словами, положение частицы меняется лишь в дискретные моменты времени $k\Delta t$, $k = 1, 2, \ldots$. Находясь в точке x, частица независимо от предшествующего поведения переходит равновероятно в одну из соседних точек $x \pm \Delta x$.

Считая начальное положение частицы $x = 0$ при $t = 0$, через $X(t)$ обозначим её положение в момент времени t. Поскольку решение двигаться влево или вправо является случайным, мы будем моделировать его с использованием последовательности независимых одинаково распределённых случайных величин $\xi_1, \xi_2, \ldots$ с распределением $\mathsf{P}(\xi_k = 1) = \mathsf{P}(\xi_k = -1) = 1/2$. Тогда

$$X(t) \;=\; \Delta x \sum_{k=1}^{\infty} \xi_k \mathbb{1}_{[0,\infty)}(t - k\Delta t), \qquad t \geqslant 0.$$

Это равенство можно переписать в другом виде. Для $n = 1, 2, \ldots$ обозначим $S_0 = 0$, $S_n = \xi_1 + \ldots + \xi_n$. Пусть также $n(t)$ обозначает целую часть числа $t/\Delta t$, т. е.

$$n(t) \;=\; \left[\frac{t}{\Delta t}\right], \qquad t \;\geqslant\; 0.$$

В этих обозначениях $X(t) = \Delta x S_{n(t)}$. Отсюда сразу же следует, что $\mathsf{E}X(t) = 0$ и $\mathsf{D}X(t) = (\Delta x)^2 n(t)$.

Устремляя к нулю Δx и Δt соответствующим образом, мы можем надеяться получить случайное блуждание, непрерывное во времени и пространстве. Если $t > 0$ фиксировано, то $n(t) \to \infty$ при $\Delta t \to 0$. При этом $n(t)\Delta t \to t$. Таким образом, если Δx и Δt будут стремиться к нулю с одинаковой скоростью, то $\mathsf{D}X(t)$ будет стремиться к нулю и мы не получим содержательной модели. Поэтому будем считать, что $(\Delta x)^2/\Delta t = \sigma^2$ постоянно при $\Delta t \to 0$, $\Delta x \to 0$. В физических приложениях постоянная σ^2 называется *коэффициентом диффузии.*

Заметим теперь, что в силу центральной предельной теоремы 6.6

$$\frac{S_{n(t)}}{\sqrt{n(t)}} \;\xrightarrow{d}\; \eta_t$$

при $n(t) \to \infty$ $(\Delta t \to 0)$, где η_t имеет стандартное нормальное распределение. Следовательно, с учётом предельного соотношения $\Delta x\sqrt{n(t)} \to \sigma\sqrt{t}$, получаем

$$X(t) \;=\; \Delta x\sqrt{n(t)}\,\frac{S_{n(t)}}{\sqrt{n(t)}} \;\xrightarrow{d}\; \sigma\sqrt{t}\eta_t$$

при $\Delta t \to 0$ $((\Delta x)^2/\Delta t = \sigma^2)$. Эти рассуждения естественно приводят к следующему определению.

Определение 10.1. Под броуновским движением с коэффициентом диффузии σ^2 будем понимать семейство случайных величин $B = (B(t)\colon\, t \geqslant 0)$, удовлетворяющее условиям

(i) $B(0) = 0$;

(ii) для любых $0 \leqslant t_1 \leqslant t_2 \leqslant \ldots \leqslant t_n$ приращения $B(t_n) - B(t_{n-1}), \ldots, B(t_2) - B(t_1)$ являются независимыми случайными величинами;

(iii) для всех $t \geqslant 0$ и $h > 0$ приращение $B(t+h) - B(t)$ имеет нормальное распределение с нулевым средним и дисперсией $\sigma^2 h$;

(iv) почти наверное функция $t \mapsto B(t)$ является непрерывной.

В случае $\sigma = 1$ процесс броуновского движения будем называть стандартным винеровским процессом и обозначать $W = (W(t)\colon\ t \geqslant 0)$.

Отметим некоторые особенности приведенного определения. Мы определили броуновское движение как случайный процесс $B = (B(t)\colon\ t \geqslant 0)$, который представляет собой семейство случайных величин $\omega \mapsto B(t, \omega)$, определённых на одном вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$. С другой стороны, на семейство B можно посмотреть как на отображение $B\colon\ \mathbb{R}_+ \times \Omega \to \mathbb{R}$. Если фиксировать элементарное событие $\omega \in \Omega$, то мы получим функцию $t \mapsto B(t, \omega)$, которая называется *траекторией*, *реализацией* или *выборочной функцией* процесса B. Условие (iv) означает, что траектории броуновского движения являются непрерывными функциями.

Под *конечномерными распределениями* случайного процесса $B = (B(t)\colon\ t \geqslant 0)$ мы понимаем распределения всех конечномерных случайных векторов

$$(B(t_1), \ldots, B(t_n))$$

для всех $0 \leqslant t_1 < t_2 < \ldots < t_n$, $n = 2, 3, \ldots$. Первые три пункта в определении броуновского движения позволяют найти все конечномерные распределения. Поскольку броуновское движение имеет нормальные конечномерные распределения, то броуновское движение относят к так называемым *гауссовским* процессам.

Вопрос существования случайного процесса с теми или иными свойствами является важным. Имеется много различных способов построения броуновского движения. Согласно принципу инвариантности Донскера—Прохорова броуновское движение можно получить, выбирая масштабированные копии случайного блуждания и переходя к пределу по распределению. Этот результат называется принципом инвариантности, поскольку все случайные блуждания, приращения которых имеют нулевое среднее и конечную дисперсию, по существу, приводят к одному и тому же пределу — броуновскому движению.

Рассмотрим теперь некоторые свойства стандартного винеровского процесса $W = (W(t)\colon\ t \geqslant 0)$. Вначале отметим, что $W(t)$ как гауссовская случайная величина имеет конечные моменты. Действительно, если $n = 2k + 1$ нечётно, то

$$\mathsf{E}(W(t))^n = \frac{1}{\sqrt{2\pi t}} \int\limits_{-\infty}^{\infty} x^{2k+1} e^{-x^2/2t} dx = 0,$$

как интеграл от нечётной функции по симметричному промежутку. Для чётных $n = 2k$ имеем

$$\begin{aligned}
\mathsf{E}(W(t))^n &= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} x^{2k} e^{-x^2/2t} dx \\
&= \sqrt{\frac{2}{\pi t}} \int_{0}^{\infty} x^{2k} e^{-x^2/2t} dx \\
&= \frac{2^k t^k}{\sqrt{\pi}} \int_{0}^{\infty} y^{k-1/2} e^{-y} dy \\
&= \frac{2^k t^k \Gamma(k+1/2)}{\sqrt{\pi}},
\end{aligned}$$

где Γ — гамма-функция. Далее, если $0 < s < t$, то используя независимость приращений, получаем

$$\mathsf{cov}(W(t), W(s)) = \mathsf{E}W(t)W(s) = \mathsf{E}W(s)[W(t) - W(s) + W(s)] = \mathsf{E}(W(s))^2 = s.$$

Свойства траекторий винеровского процесса. Пусть $W = (W(t)\colon t \geqslant 0)$ — стандартный винеровский процесс. Для $T > 0$ и $n \in \mathbb{N}$ введём в рассмотрение следующие случайные величины

$$V_n(T) = \sum_{i=1}^{n} |W(t_i) - W(t_{i-1})|, \qquad V_n^2(T) = \sum_{i=1}^{n} |W(t_i) - W(t_{i-1})|^2,$$

где $t_i = iT/n$, $i = 0, 1, \ldots, n$.

Лемма 10.1. *$V_n^2(T) \xrightarrow{\mathsf{P}} T$ при $n \to \infty$.*

Доказательство. Для удобства введём следующие обозначения

$$\delta = \frac{T}{n}, \qquad \Delta W_i = W(t_i) - W(t_{i-1})$$

и заметим, что $\mathsf{E}(\Delta W_i)^2 = \delta$, $\mathsf{E}(\Delta W_i)^4 = 3\delta^2$, $i = 1, \ldots, n$. Отсюда с использованием свойства независимости приращений получаем

$$\mathsf{E}V_n^2(T) = \mathsf{E}\left(\sum_{i=1}^{n} (\Delta W_i)^2\right) = n\delta = T.$$

Используя этот результат, получаем выражение для дисперсии

$$\begin{aligned}
\mathsf{D}V_n^2(T) &= \mathsf{E}\left(\sum_{i=1}^{n} (\Delta W_i)^2\right)^2 - T^2 \\
&= \sum_{i=1}^{n} \mathsf{E}(\Delta W_i)^4 + \sum_{i \neq j} \mathsf{E}(\Delta W_i)^2 (\Delta W_j)^2 - T^2 \\
&= 3n\delta^2 + (n^2 - n)\delta^2 - T^2 = 2n\delta^2 = 2T\delta.
\end{aligned}$$

Пусть теперь $\varepsilon > 0$ фиксировано. Поскольку $\delta \to 0$ при $n \to \infty$, то, используя неравенство Чебышёва (5.5), получаем

$$\mathsf{P}(|V_n^2(T) - T| \geqslant \varepsilon) \;\leqslant\; \frac{1}{\varepsilon^2}\mathsf{D}V_n^2(T) \;=\; \frac{2T\delta}{\varepsilon^2} \;\to\; 0$$

при $n \to \infty$. □

Напомним определение вариации функции на интервале. Пусть $\varphi\colon\ [0,T] \to \mathbb{R}$ — детерминированная функция и $\mathscr{P} = \{0 = t_0 < t_1 < \ldots < t_n = T\}$ — разбиение отрезка $[0,T]$. Обозначим

$$V(\varphi, [0,T], \mathscr{P}) \;=\; \sum_{i=1}^{n} |\varphi(t_i) - \varphi(t_{i-1})|.$$

Величину

$$V(\varphi, [0,T]) \;=\; \sup V(\varphi, [0,T], \mathscr{P}),$$

где супремум берётся по всем разбиениям $\mathscr{P}$, называют вариацией (или полной вариацией) функции φ на промежутке $[0,T]$.

Лемма 10.2. *Почти все траектории винеровского процесса имеют бесконечную вариацию, т. е. для любого $T > 0$ почти наверное $V(W,[0,T]) = \infty$.*

Доказательство. Сохраняя принятые при доказательстве леммы 10.1 обозначения, заметим, что в силу свойства непрерывности траекторий винеровского процесса

$$\max_{1\leqslant i\leqslant n} |\Delta W_i| \;\to\; 0$$

при $n \to \infty$. Здесь, как и выше,

$$\Delta W_i \;=\; W(t_i) - W(t_{i-1}), \qquad t_i \;=\; \frac{iT}{n}, \quad i \;=\; 0, 1, \ldots, n.$$

Из леммы 10.1 и теоремы 6.1 следует, что найдется подпоследовательность $(n_k\colon\ k \in \mathbb{N})$ такая, что

$$V_{n_k}^2(T) \;\overset{\text{п.н.}}{\to}\; T$$

при $k \to \infty$. С другой стороны,

$$V_{n_k}^2(T) \;=\; \sum_{i=1}^{n_k} (\Delta W_i)^2 \;\leqslant\; \max_{1\leqslant i\leqslant n} |\Delta W_i| \sum_{j=1}^{n_k} |\Delta W_j| \;=\; V_{n_k}(T) \max_{1\leqslant i\leqslant n} |\Delta W_i|.$$

Отсюда следует, что $V_{n_k}(T) \overset{\text{п.н.}}{\to} \infty$ при $k \to \infty$. Это означает, что $V(W,[0,T]) = \infty$ почти наверное. □

Этот результат имеет важные следствия в теории стохастического интеграла.

Моменты остановки и марковское свойство. С винеровским процессом связывают также семейство σ-алгебр $\mathscr{F}_t = \sigma(\{W(s)\colon\ 0 \leqslant s \leqslant t\})$. Каждую σ-алгебру $\mathscr{F}_t$ можно интерпретировать как информацию, связанную с наблюдением винеровского процесса до момента времени t. Семейство $(\mathscr{F}_t\colon\ t \geqslant 0)$ называют *фильтрацией* и оно обладает свойством $\mathscr{F}_s \subseteq \mathscr{F}_t$ при $s \leqslant t$.

Заметим теперь, что для любого $a > 0$ процесс

$$W_a = (W_a(t) = W(t+a) - W(a)\colon\ t \geqslant 0)$$

также является винеровским процессом, что легко устанавливается проверкой свойств (i)—(iv) из определения 10.1. Кроме того, процесс W_a не зависит от σ-алгебры $\mathscr{F}_a$. Это так называемое *марковское свойство* винеровского процесса. Наряду с этим винеровский процесс обладает «строгим марковским свойством». Его формулировка связана с важным понятием момента остановки.

Определение 10.2. Неотрицательная случайная величина τ, определённая на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$, называется моментом остановки относительно фильтрации $(\mathscr{F}_t\colon\ t \geqslant 0)$, если для всех $t \geqslant 0$ выполняется условие $\{\omega\colon\ \tau(\omega) \leqslant t\} \in \mathscr{F}_t$.

Условие $\{\tau \leqslant t\} \in \mathscr{F}_t$ можно трактовать как то, что событие $\{\tau \leqslant t\}$ зависит только от истории винеровского процесса до момента времени t. Оказывается, что если τ — момент остановки относительно фильтрации $(\mathscr{F}_t\colon\ t \geqslant 0)$, то

$$W_\tau = (W_\tau(t) = W(t+\tau) - W(\tau)\colon\ t \geqslant 0)$$

также будет винеровским процессом. Этот результат, который мы не будем доказывать, составляет так называемое *строго марковское свойство* винеровского процесса.

Одним из важных моментов остановки, связанных с винеровским процессом, является

$$\tau_a = \inf\{t\colon\ W(t) = a\}, \qquad a > 0,$$

— момент первого достижения уровня a.

Лемма 10.3. *Пусть* $a > 0$, $t > 0$. *Тогда* $\mathsf{P}(\tau_a < t) = 2\mathsf{P}(W(t) > a)$.

Доказательство. Допустим, что выполнено условие $W(t) > a$. Тогда в силу непрерывности траектории винеровского процесса будет также выполняться условие $\tau_a < t$. Это означает, что

$$\{W(t) > a\} \subseteq \{\tau_a < t\}, \qquad \{W(t) > a, \tau_a < t\} = \{W(t) > a\}.$$

Далее, в силу симметрии и строго марковского свойства винеровского процесса имеет место равенство

$$\mathsf{P}(W(t) > a | \tau_a < t) = \mathsf{P}(W(t) < a | \tau_a < t) = \frac{1}{2},$$

поскольку при выходе из точки a за время от τ_a до t броуновская частица равновероятно окажется левее или правее точки a. Но тогда

$$\mathsf{P}(W(t) > a) = \mathsf{P}(W(t) > a, \tau_a < t) = \mathsf{P}(\tau_a < t)\mathsf{P}(W(t) > a | \tau_a < t) = \frac{1}{2}\mathsf{P}(\tau_a < t)$$

и лемма доказана. □

Полученный результат позволяет найти распределение случайной величины τ_a. Действительно, для $t > 0$ имеем

$$F_{\tau_a}(t) = \mathsf{P}(\tau_a \leqslant t) = 2\mathsf{P}(W(t) \geqslant a) = \frac{2}{\sqrt{2\pi t}} \int_a^\infty e^{-x^2/2t} dx = \sqrt{\frac{2}{\pi}} \int_{a/\sqrt{t}}^\infty e^{-u^2/2} du.$$

Дифференцируя полученное выражение, находим плотность распределения

$$f_{\tau_a}(t) = \frac{a}{\sqrt{2\pi}} t^{-3/2} e^{-a^2/2t} \mathbb{1}_{(0,\infty)}(t).$$

Максимум винеровского процесса на промежутке $[0, t]$. Найдём теперь распределение вероятностей максимального смещения вправо броуновской частицы за фиксированное время, т. е. распределение случайной величины

$$M_t := \max\{W(s) \colon 0 \leqslant s \leqslant t\}.$$

Из определения видно, что $M_t \geqslant 0$ и является неубывающей функцией по t. Заметим также, что для $x > 0$ имеет место равенство

$$\{\omega \colon M_t(\omega) \geqslant x\} = \{\omega \colon \tau_x(\omega) \leqslant t\}.$$

Следовательно,

$$\mathsf{P}(M_t \geqslant x) = \mathsf{P}(\tau_x \leqslant t) = \sqrt{\frac{2}{\pi}} \int_{x/\sqrt{t}}^\infty e^{-u^2/2} du,$$

откуда получаем плотность распределения

$$f_{M_t}(x) = \frac{d}{dx}(1 - \mathsf{P}(M_t \geqslant x)) = \sqrt{\frac{2}{\pi t}} e^{-x^2/2t} \mathbb{1}_{(0,\infty)}(x).$$

Это так называемый удвоенный нормальный закон распределения вероятностей. При этом для всех $x > 0$ выполняется равенство (Башелье)

$$\mathsf{P}(M_t \geqslant x) = 2\mathsf{P}(W(t) \geqslant x).$$

Аналогично определяется распределение вероятностей случайной величины

$$M_t^- := \min\{W(s) \colon 0 \leqslant s \leqslant t\}.$$

Из соображений симметрии получаем плотность распределения этой случайной величины

$$f_{M_t^-}(x) = \sqrt{\frac{2}{\pi t}} e^{-x^2/2t} \mathbb{1}_{(-\infty,0)}(x).$$

Полученные результаты приводят к следующему свойству траекторий винеровского процесса. Для любого $t > 0$ имеют место равенства

$$\mathsf{P}(M_t^- < 0) = \mathsf{P}(M_t > 0) = 2\mathsf{P}(W(t) > 0) = 1.$$

Это означает, что, выходя из начала координат, броуновская частица за сколь угодно малое время с вероятностью 1 побывает как правее точки $x = 0$, так и левее этой точки.

§ 11. Случайные выборки и оценки параметров распределения

Задача математической статистики — получить определённые выводы из экспериментальных данных. В некотором смысле эти задачи являются обратными по отношению к теории вероятностей. Коротко их можно охарактеризовать следующим образом. Даны некоторые наблюдения (статистические данные), по которым мы хотим восстановить вероятностную модель, соответствующую полученным данным. Одна из первых задач математической статистики — это задача оценивания параметров распределений. Чаще всего исходные данные представляют собой результаты наблюдений случайных величин. В связи с этим возникает следующая терминология.

Под (случайной) *выборкой* объёма n, отвечающей функции распределения F, будем понимать набор (или вектор) из n независимых случайных величин $\mathbb{X} = (X_1, \ldots, X_n)$, каждая из которых имеет распределение F, т. е. $F_{X_i}(x) = F(x)$, $i = 1, \ldots, n$. Реализацию случайной выборки $\mathbb{X} = (X_1, \ldots, X_n)$ будем обозначать $\mathbb{x} = (x_1, \ldots, x_n)$. Если $\varphi\colon \mathbb{R}^n \to \mathbb{R}$ — борелевская функция, то случайная величина $\varphi(X_1, \ldots, X_n)$ называется *статистикой* выборки $\mathbb{X}$. Одной из основных задач математической статистики является построение по случайным выборкам $\mathbb{X}$ статистик, которые могли бы служить оценками параметров распределений.

Каждой реализации $\mathbb{x}$ выборки $\mathbb{X}$ можно поставить в соответствие упорядоченную последовательность значений $x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(n)}$, где $x_{(1)}$ — минимальное значение из $x_1, \ldots, x_n$, а $x_{(n)}$ — максимальное. Упорядочивание каждой реализации $\mathbb{x} = (x_1, \ldots, x_n) \mapsto (x_{(1)}, \ldots, x_{(n)})$ приводит к преобразованию $(X_1, \ldots, X_n) \mapsto (X_{(1)}, \ldots, X_{(n)})$. Случайный вектор $(X_{(1)}, \ldots, X_{(n)})$ называется *вариационным рядом* выборки $\mathbb{X}$, а случайные величины $X_{(i)}$, $i = 1, \ldots, n$, называются порядковыми статистиками выборки $\mathbb{X}$. В частности, $X_{(1)} = \min\{X_1, \ldots, X_n\}$ и $X_{(n)} = \max\{X_1, \ldots, X_n\}$. Заметим, что порядковые статистики имеют каждая свою функцию распределения

$$F_{X_{(i)}}(x) = \mathsf{P}(X_{(i)} \leqslant x) = \sum_{k=i}^{n} C_n^k (F(x))^k (1 - F(x))^{n-k}.$$

Здесь выражен тот факт, что в n независимых испытаниях случайная величина X с функцией распределения F приняла не менее i раз значение, не превышающее x.

Эмпирическая функция распределения $\mathbb{F}_n$ является оценкой для функции распределения F случайной выборки $\mathbb{X} = (X_1, \ldots, X_n)$. Она определяется равенством

$$\mathbb{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty, x]}(X_i), \qquad -\infty < x < \infty.$$

Заметим, что при каждом $x \in \mathbb{R}$ эмпирическая функция распределения является случайной величиной. Для конкретной реализации $\mathbb{x} = (x_1, \ldots, x_n)$ мы получаем обычную ступенчатую функцию распределения

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty, x]}(x_i), \qquad -\infty < x < \infty.$$

Значение $F_n(x)$ выражает долю тех значений $x_1, \dots, x_n$, которые не превосходят x. Для визуализации реализации выборки часто используют *гистограмму*. В некотором смысле она является дискретным аналогом плотности распределения.

Предложение 11.1. *Пусть $\mathbb{X} = (X_1, \dots, X_n)$ — выборка, отвечающая функции распределения F. Тогда для любого $x \in \mathbb{R}$*

$$\mathbb{F}_n(x) \overset{\text{п.н.}}{\to} F(x),$$

где $\mathbb{F}_n$ — эмпирическая функция распределения выборки $\mathbb{X}$.

Доказательство. Пусть $x \in \mathbb{R}$ фиксировано. Обозначим $\xi_i = \mathbb{1}_{\{X_i \leqslant x\}}$. Случайные величины $\xi_1, \xi_2, \dots$ являются независимыми и имеют распределение

$$\mathsf{P}(\xi_i = 1) = F(x), \qquad \mathsf{P}(\xi_i = 0) = 1 - F(x).$$

Следовательно,

$$\mathsf{E}\xi_i = F(x), \qquad \mathsf{D}\xi_i = F(x)(1 - F(x)),$$

т. е. выполнены условия теоремы 6.4. Но тогда

$$\mathbb{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \xi_i \overset{\text{п.н.}}{\to} \mathsf{E}\xi_1 = F(x).$$

□

Выборочные моменты. Эмпирическая функция распределения является некоторой сводной характеристикой выборки. Наряду с этим с каждой выборкой связывают также выборочные моменты, характеризующие выборку в целом. Под *выборочным моментом* порядка r, $r \in \mathbb{N}$, выборки $\mathbb{X} = (X_1, \dots, X_n)$ понимается

$$X^{(r)} = \frac{1}{n}\sum_{i=1}^{n} (X_i)^r.$$

Выборочный момент первого порядка называют *выборочным средним* и обозначают $\overline{X}$. В качестве *выборочной дисперсии* рассматривают

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \overline{X})^2 \qquad \text{или} \qquad \overline{S}^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^2.$$

Для вывода соотношений, связывающих разные статистики, полезен следующий результат.

Лемма 11.1. *Пусть $x_1, \dots, x_n$ и a — произвольные вещественные числа и $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Тогда имеют место следующие равенства*

(i) $\sum_{i=1}^{n} (x_i - a)^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 + n(\overline{x} - a)^2$;

(ii) $\sum_{i=1}^{n} (x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2$.

Доказательство. Поскольку $\sum_{i=1}^{n} (x_i - \overline{x}) = 0$, то

$$\sum_{i=1}^{n} (x_i - a)^2 = \sum_{i=1}^{n} (x_i - \overline{x} + \overline{x} - a)^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 + n(\overline{x} - a)^2,$$

что доказывает (i). Равенство (ii) следует из (i) при $a = 0$. □

Предложение 11.2. *Пусть выборка $\mathbb{X} = (X_1, \dots, X_n)$ отвечает распределению с конечным математическим ожиданием a и конечной дисперсией σ^2. Тогда*

$$\mathsf{E}\overline{X} = a, \quad \mathsf{D}\overline{X} = \frac{\sigma^2}{n}, \quad \mathsf{E}S^2 = \sigma^2, \quad \mathsf{E}\overline{S}^2 = \frac{n-1}{n}\sigma^2.$$

Доказательство. Равенство $\mathsf{E}\overline{X} = a$ следует из свойства линейности математического ожидания и условия $\mathsf{E}X_i = a$. Из независимости случайных величин в выборке и условия $\mathsf{D}X_i = \sigma^2$ получаем

$$\mathsf{D}\overline{X} = \mathsf{D}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathsf{D}X_i = \frac{1}{n}\sigma^2.$$

Для вычисления $\mathsf{E}S^2$ воспользуемся равенством (ii) из леммы 11.1

$$\begin{aligned}\mathsf{E}S^2 &= \frac{1}{n-1}\mathsf{E}\left(\sum_{i=1}^{n}(X_i - \overline{X})^2\right) \\ &= \frac{1}{n-1}\left(\sum_{i=1}^{n}\mathsf{E}X_i^2 - n\mathsf{E}\overline{X}^2\right) \\ &= \frac{1}{n-1}\left(n\sigma^2 + na^2 - \sigma^2 - na^2\right) = \sigma^2.\end{aligned}$$

Замечая также, что $\overline{S}^2 = \frac{n-1}{n}S^2$, получаем последнее равенство. □

Часто в приложениях бывает известно, что выборка $\mathbb{X}$ относится к некоторому классу распределений, который определяется одним или несколькими параметрами (например, пуассоновское, показательное, нормальное и т. д.). При этом априори некоторые параметры неизвестны. Возникает задача оценивания неизвестного параметра θ. Под *оценкой* неизвестного параметра θ понимается функция от случайных величин выборки, т. е. статистика. Однако не все оценки в таком понимании имеют практическую ценность. Поэтому вводятся понятия, которые выделяют практически значимые оценки. Одним из таких понятий является несмещённость.

Статистика $\varphi_n(\mathbb{X})$ называется *несмещённой* оценкой параметра θ, если $\mathsf{E}\varphi_n(\mathbb{X}) = \theta$. Из предложения 11.2 следует, что выборочное среднее $\overline{X}$ является несмещённой оценкой математического ожидания, а выборочная дисперсия S^2 — несмещённой оценкой дисперсии распределения, которому соответствует выборка. При этом $\overline{S}^2$ не является несмещённой оценкой дисперсии.

Другое понятие связано с асимптотическим поведением оценки при $n \to \infty$. Статистика $\varphi_n(\mathbb{X})$ называется *состоятельной* оценкой параметра θ, если $\varphi_n(\mathbb{X}) \xrightarrow{\mathsf{P}} \theta$ при $n \to \infty$. В случае, когда $\varphi_n(\mathbb{X}) \xrightarrow{\text{п.н.}} \theta$, говорят о *сильной* состоятельности оценки.

Замечание 11.1. Из усиленного закона больших чисел следует, что $\overline{X}$ является сильно состоятельной оценкой математического ожидания, если выборка отвечает распределению с конечным вторым моментом.

ПРЕДЛОЖЕНИЕ 11.3. *Пусть $\mathbb{X}$ — выборка, отвечающая распределению с конечным четвёртым моментом. Тогда S^2 и $\overline{S}^2$ являются сильно состоятельными оценками дисперсии σ^2 этого распределения.*

ДОКАЗАТЕЛЬСТВО. Поскольку $S^2 = \frac{n}{n-1}\overline{S}^2$, то достаточно доказать состоятельность любой из выборочных дисперсий. Докажем это для оценки $\overline{S}^2$. Из равенства (ii) леммы 11.1 и определения $\overline{S}^2$ следует

$$\overline{S}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2.$$

Как отмечалось в замечании выше, $\overline{X} \overset{\text{п.н.}}{\to} a$ при $n \to \infty$. Но тогда $\overline{X}^2 \overset{\text{п.н.}}{\to} a^2$. Обозначим $X_i^2 = \eta_i$. По условию $\eta_1, \eta_2, \ldots$ являются независимыми одинаково распределёнными случайными величинами с конечными вторыми моментами. Но тогда по усиленному закону больших чисел

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \frac{1}{n}\sum_{i=1}^{n} \eta_i \overset{\text{п.н.}}{\to} \mathsf{E}\eta_1 = \sigma^2 + a^2.$$

Это вместе с условием $\overline{X}^2 \overset{\text{п.н.}}{\to} a^2$ доказывает наше утверждение. □

Методы получения оценок. В некоторых случаях можно догадаться о виде оценки. Например, в качестве оценки математического ожидания распределения естественно взять выборочное среднее. Однако в более сложных ситуациях требуются специальные методы. Одним из первых методов получения оценок неизвестных параметров является *метод моментов*, который восходит к Пирсону и достаточно прост. Суть метода моментов состоит в том, чтобы связать неизвестные параметры распределения системой уравнений с моментами неизвестной случайной величины. Разрешив эту систему относительно неизвестных параметров и заменив моменты случайной величины выборочными моментами, получаем оценки неизвестных параметров.

Пример. Пусть ξ — дискретная случайная величина, равномерно распределённая на $\{1, 2, \ldots, \theta\}$, т. е. $\mathsf{P}(\xi = k) = 1/\theta$, $k = 1, \ldots, \theta$. Найти оценки параметра θ и дисперсии $\mathsf{D}\xi$.

Для удобства будем считать $\theta_1 = \theta$ и $\theta_2 = \mathsf{D}\xi$ неизвестными параметрами, для которых нужно найти оценки. Обозначим $m_1 = \mathsf{E}\xi$ и $m_2 = \mathsf{E}\xi^2$. Тогда

$$m_1 = \frac{1+\theta_1}{2}, \qquad m_2 = \theta_2 + m_1^2,$$

откуда следуют равенства

$$\theta_1 = 2m_1 - 1, \qquad \theta_2 = m_2 - m_1^2.$$

Таким образом,

$$\widehat{\theta}_1(\mathbb{X}) = 2\overline{X} - 1, \qquad \widehat{\theta}_2(\mathbb{X}) = X^{(2)} - \overline{X}^2.$$

Для применения этого метода требуется существование достаточного количества моментов. Примером распределения, для которого этот метод не работает, является распределение Коши.

Метод *наибольшего правдоподобия* является одним из наиболее широко используемых методов построения оценок. Допустим $X_1, \dots, X_n$ — выборка из распределения с плотностью $f(x;\theta)$ Здесь θ — скалярный или векторный параметр, который нужно определить. Для $\mathbb{x} = (x_1, \dots, x_n)$ — реализации выборки строится так называемая *функция правдоподобия*

$$f(\mathbb{x};\theta) = \prod_{i=1}^{n} f(x_i;\theta).$$

В случае, если распределение дискретное, в качестве $f(x;\theta)$ рассматривается $\mathsf{P}_\theta(\xi = x)$ — вероятность значения x при вероятностном распределении с параметром θ.

Пусть $\widehat{\theta}(\mathbb{x})$ — значение параметра, на котором $f(\mathbb{x};\theta)$ достигает максимума как функция θ при фиксированном $\mathbb{x}$. Оценкой наибольшего правдоподобия параметра θ тогда полагают $\widehat{\theta}(\mathbb{X})$.

Доверительные интервалы. Другой подход в определении неизвестного параметра распределения связан с понятием доверительного интервала. По выборке $\mathbb{X} = (X_1, \dots, X_n)$ строится интервал, содержащий неизвестный параметр. Пусть $\underline{\theta}(\mathbb{X})$ и $\overline{\theta}(\mathbb{X})$ — две функции от случайной выборки $\mathbb{X}$ такие, что $\underline{\theta}(\mathbb{X}) \leqslant \overline{\theta}(\mathbb{X})$.

Интервал $(\underline{\theta}(\mathbb{X}), \overline{\theta}(\mathbb{X}))$ называется *доверительным интервалом* для неизвестного параметра θ, отвечающим *доверительной вероятности* α (или уровню доверия α), если

$$\mathsf{P}\left(\underline{\theta}(\mathbb{X}) < \theta < \overline{\theta}(\mathbb{X})\right) \geqslant \alpha.$$

Методы построения доверительных интервалов основываются на центральной предельной теореме.

Доверительные интервалы для неизвестного $a = \mathsf{E}X$ при известной дисперсии $\sigma^2 = \mathsf{D}X$.

В силу свойства асимптотической нормальности выборочного среднего $\overline{X}$ имеем

$$\mathsf{P}\left(\frac{\sqrt{n}}{\sigma}\left(\overline{X} - a\right) < x\right) \approx \Phi(x),$$

где $\Phi(x)$ — функция распределения стандартного нормального закона. Отсюда получаем

$$\mathsf{P}\left(-l < \frac{\sqrt{n}}{\sigma}\left(\overline{X} - a\right) < l\right) \approx \Phi(l) - \Phi(-l),$$

или

$$\mathsf{P}\left(\overline{X} - \frac{l\sigma}{\sqrt{n}} < a < \overline{X} + \frac{l\sigma}{\sqrt{n}}\right) \approx \Phi(l) - \Phi(-l).$$

Поскольку $\Phi(-l) = 1 - \Phi(l)$, то $\Phi(l) - \Phi(-l) = 2\Phi(l) - 1$, и условие $\Phi(l) - \Phi(-l) \geqslant \alpha$ будет выполнено в случае $l \geqslant x_\gamma$, где x_γ — решение уравнения $\Phi(x) = \gamma$, $\gamma = (1+\alpha)/2$. Таким образом, доверительный интервал с уровнем доверия α можно получить, определив

$$\underline{\theta}(\mathbb{X}) = \overline{X} - \frac{\sigma x_\gamma}{\sqrt{n}}, \qquad \overline{\theta}(\mathbb{X}) = \overline{X} + \frac{\sigma x_\gamma}{\sqrt{n}}.$$

Доверительный интервал для $a = \mathsf{E}X$ при неизвестной дисперсии естественно выбрать в виде

$$\underline{\theta}(\mathbb{X}) = \overline{X} - \frac{x_\gamma\sqrt{S^2}}{\sqrt{n}}, \qquad \overline{\theta}(\mathbb{X}) = \overline{X} + \frac{x_\gamma\sqrt{S^2}}{\sqrt{n}},$$

где

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

§ 12. Мартингалы

Как отмечалось выше, под случайным процессом мы понимаем семейство случайных величин, определённых на одном вероятностном пространстве. Характер случайного процесса и методы его изучения во многом определяются тем, как задана зависимость между случайными величинами. Важный класс случайных процессов, который мы рассмотрим в этом параграфе, выделяется тем, что зависимость между прошлым и будущим определяется посредством условного математического ожидания.

Будем считать время дискретным и наряду с вероятностным пространством $(\Omega, \mathscr{F}, \mathsf{P})$ рассматривать последовательность σ-алгебр $(\mathscr{F}_n\colon n \geqslant 0)$ такую, что $\mathscr{F}_m \subseteq \mathscr{F}_n \subseteq \mathscr{F}$, $0 \leqslant m \leqslant n$. В этом случае говорят, что на $(\Omega, \mathscr{F}, \mathsf{P})$ определена фильтрация $(\mathscr{F}_n)$, а вероятностное пространство с фильтрацией обозначают $(\Omega, \mathscr{F}, \mathsf{P}, (\mathscr{F}_n))$. Интуитивно $\mathscr{F}_n$ можно интерпретировать как информацию, которая становится доступной в момент времени n, т. е. в момент времени n становится известно о каждом событии из $\mathscr{F}_n$ произошло оно или нет. С течением времени эта информация расширяется.

Определение 12.1. Случайный процесс $X = (X_n\colon n \geqslant 0)$ будем называть согласованным с фильтрацией $(\mathscr{F}_n)$, если X_n является $\mathscr{F}_n$-измеримой для всех $n = 0, 1, 2, \ldots$. Если же для всех $n \geqslant 0$ случайные величины X_n являются $\mathscr{F}_{n-1}$-измеримыми (считаем, что $\mathscr{F}_{-1} = \mathscr{F}_0$), то процесс X называется предсказуемым.

Для иллюстрации вводимых определений рассмотрим следующий пример.

Пример. Представим себе игру в казино, где игрок перед каждой игрой может делать ставки. Допустим, что $\xi_1, \xi_2, \ldots$ — независимые одинаково распределённые случайные величины, где ξ_n — сумма выигрыша в n-той игре на единичной ставке. Тогда $S_0 = 0$, $S_n = \xi_1 + \ldots + \xi_n$, $n = 1, 2, \ldots$, будет выражать чистый выигрыш (который может принимать и отрицательные значения) игрока после n-той игры, если он делает каждый раз единичную ставку. С другой стороны, игрок может придерживаться некоторой стратегии, делая на n-тую игру ставку H_n, $n = 1, 2, \ldots$. Тогда его чистый выигрыш после n-той игры составит $X_n = H_1\xi_1 + \ldots + H_n\xi_n$. В этом примере естественно возникает фильтрация $(\mathscr{F}_n)$, где $\mathscr{F}_0 = \{\varnothing, \Omega\}$, $\mathscr{F}_n = \sigma(\xi_1, \ldots, \xi_n)$, $n = 1, 2, \ldots$. В терминологии определения 12.1 $S = (S_n\colon n \geqslant 0)$, и $X = (X_n\colon n \geqslant 0)$ будут процессами, согласованными с фильтрацией $\mathscr{F}_n$, а $H = (H_n\colon n \geqslant 0)$ — предсказуемый процесс, поскольку выбор H_n формируется до n-той игры на основании информации о результатах предыдущих игр. Здесь полагаем $H_0 = 0$.

Определение 12.2. Пусть $(\Omega, \mathscr{F}, \mathsf{P}, (\mathscr{F}_n))$ — пространство с фильтрацией. Процесс $X = (X_n\colon n \geqslant 0)$, определённый на этом пространстве называется *мартингалом*, если выполняются следующие условия.

(i) X является согласованным с фильтрацией $(\mathscr{F}_n)$;
(ii) $\mathsf{E}|X_n| < \infty$ для всех $n \geqslant 0$;
(iii) $\mathsf{E}(X_n|\mathscr{F}_{n-1}) = X_{n-1}$ п. н. для всех $n \geqslant 1$.

Если в условии (iii) равенство заменить на неравенство $\mathsf{E}(X_n|\mathscr{F}_{n-1}) \geqslant X_{n-1}$ ($\mathsf{E}(X_n|\mathscr{F}_{n-1}) \leqslant X_{n-1}$), то процесс X называется *субмартингалом* (*супермартингалом*).

Условие (iii) можно также записать в виде $\mathsf{E}(\Delta X_n|\mathscr{F}_{n-1}) = 0$ п. н. (или $\mathsf{E}(\Delta X_n|\mathscr{F}_{n-1}) \geqslant 0$ в случае субмартингала), где $\Delta X_n = X_n - X_{n-1}$.

Заметим, что в приведенном выше примере в случае $\mathsf{E}\xi_n = 0$ процесс $S = (S_n\colon n \geqslant 0)$ является мартингалом, а игра называется честной. Действительно, из свойств условного математического ожидания (теорема 7.5) получаем

$$\mathsf{E}(\Delta S_n|\mathscr{F}_{n-1}) = \mathsf{E}(\xi_n|\mathscr{F}_{n-1}) = 0.$$

В случае $\mathsf{E}\xi_n \geqslant 0$ ($\mathsf{E}\xi_n \leqslant 0$) процесс S является субмартингалом (супермартингалом), а игра считается предпочтительной (непредпочтительной) для игрока.

Из определения 12.2 также видно, что если $X = (X_n\colon n \geqslant 0)$ является субмартингалом, то $-X = (-X_n\colon n \geqslant 0)$ будет супермартингалом. Согласованный процесс X является мартингалом в том и только том случае, если он одновременно является субмартингалом и супермартингалом. Это часто используется при доказательстве утверждений. Например, покажем, что условие (iii) выполняется и для несоседних индексов. Пусть X — субмартингал и $m < n$. Тогда поскольку $\mathscr{F}_m \subseteq \mathscr{F}_n$, то, используя свойства условных математических ожиданий (теорема 7.5), получаем

$$\mathsf{E}(X_n|\mathscr{F}_m) = \mathsf{E}(\mathsf{E}(X_n|\mathscr{F}_{n-1})|\mathscr{F}_m) \geqslant \mathsf{E}(X_{n-1}|\mathscr{F}_m).$$

Продолжая эти рассуждения, приходим к неравенству

$$\mathsf{E}(X_n|\mathscr{F}_m) \geqslant X_m,$$

т. е. свойство субмартингальности выполняется и для всех индексов $m < n$. Пусть теперь X — супермартингал. Применяя доказанное свойство для субмартингала $-X$, получаем $\mathsf{E}(-X_n|\mathscr{F}_m) \geqslant -X_m$, откуда следует $\mathsf{E}(X_n|\mathscr{F}_m) \leqslant X_m$. Для мартингала равенство $\mathsf{E}(X_n|\mathscr{F}_m) = X_m$ при $m < n$ следует теперь из того, что он одновременно является субмартингалом и супермартингалом.

Предложение 12.1. *Пусть $X = (X_n\colon n \geqslant 0)$ — мартингал, а $\varphi\colon \mathbb{R} \to \mathbb{R}$ — выпуклая функция. Допустим также, что $\mathsf{E}|\varphi(X_n)| < \infty$ для всех $n \geqslant 0$. Тогда $\varphi \circ X = (\varphi(X_n)\colon n \geqslant 0)$ является субмартингалом. Утверждение остаётся в силе, если X — субмартингал, а функция φ выпукла и не убывает.*

Доказательство. Пусть X — мартингал и $m < n$. Используя неравенство Иенсена 7.4, получаем

$$\varphi(X_m) = \varphi(\mathsf{E}(X_n|\mathscr{F}_m)) \leqslant \mathsf{E}(\varphi(X_n)|\mathscr{F}_m),$$

что означает субмартингальность процесса $\varphi \circ X$.

В случае, когда X — субмартингал, а φ является ещё и неубывающей, имеем

$$\varphi(X_n) \leqslant \varphi(\mathsf{E}(X_n|\mathscr{F}_m)) \leqslant \mathsf{E}(\varphi(X_n)|\mathscr{F}_m).$$

□

Следствие 12.1. *Если $X = (X_n\colon n \geqslant 0)$ — мартингал, то $|X| = (|X_n|\colon n \geqslant 0)$ — субмартингал. В случае, когда X является квадратично интегрируемым мартингалом, т. е. $\mathsf{E}X_n^2 < \infty$ для всех $n \geqslant 0$, то $X^2 = (X_n^2\colon n \geqslant 0)$ — субмартингал.*

Утверждения следуют из доказанного предложения и выпуклости функций $|x|$ и x^2.

Пусть теперь H — предсказуемый процесс, а X — согласованный процесс. Определим процесс $Y = (Y_n\colon n \geqslant 0)$, положив

$$Y_0 = 0, \qquad Y_n = \sum_{k=1}^{n} H_k \Delta X_k, \qquad n = 1, 2, \ldots.$$

Процесс Y называют преобразованием X посредством H и обозначают $H \bullet X = ((H \bullet X)_n\colon n \geqslant 0)$, т. е. $Y_n = (H \bullet X)_n$. Если X — мартингал, то $H \bullet X$ называют *мартингальным преобразованием.*

Теорема 12.1. *Пусть $H = (H_n\colon n \geqslant 0)$ — предсказуемый ограниченный процесс, т. е. $|H_n| \leqslant M$ при некотором фиксированном M и всех n. Тогда имеют место следующие утверждения.*

(i) Если X — мартингал, то $H \bullet X$ также является мартингалом.

(ii) Если X — субмартингал (супермартингал), а предсказуемый процесс H дополнительно является неотрицательным ($H_n \geqslant 0$ для всех n), то $H \bullet X$ — субмартингал (супермартингал).

Доказательство. Ограниченность процесса H влечёт условие $\mathsf{E}|Y_n| < \infty$ для всех n. Остаётся проверить условие мартингальности для процесса $Y = H \bullet X$. Если X — мартингал, то

$$\mathsf{E}(\Delta Y_n|\mathscr{F}_{n-1}) = \mathsf{E}(H_n \Delta X_n|\mathscr{F}_{n-1}) = H_n \mathsf{E}(\Delta X_n|\mathscr{F}_{n-1}) = 0,$$

что доказывает (i).

В случае, когда X — субмартингал, используя неотрицательность процесса H, получаем

$$\mathsf{E}(\Delta Y_n|\mathscr{F}_{n-1}) = H_n \mathsf{E}(\Delta X_n|\mathscr{F}_{n-1}) \geqslant 0,$$

т. е. $H \bullet X$ — субмартингал. Если X — супермартингал, то последнее неравенство меняется на противоположное и (ii) доказано. □

Полученный результат иногда называют теоремой об устойчивости, поскольку $H \bullet X$ сохраняет свойства процесса X. В случае нашего примера чистый выигрыш игрока, если он придерживается стратегии H, представляет собой преобразование $H \bullet S$. Из теоремы об устойчивости следует, что любая стратегия игрока не меняет качества игры.

Моменты остановки и остановленный процесс. Понятие момента остановки является одним из фундаментальных в теории мартингалов. Этот термин пришёл из теории игр и тесно связан с понятием стратегии, например, с моментом выхода из игры.

Определение 12.3. Случайная величина $\tau\colon\ \Omega \to \mathbb{Z}_+ \cup \{\infty\}$ называется *моментом остановки* относительно фильтрации $(\mathscr{F}_n)$, если $\mathsf{P}(\tau < \infty) = 1$ и для всех $n \geqslant 0$ выполняется условие

$$\{\tau \ = \ n\} \in \mathscr{F}_n. \tag{12.1}$$

Предложение 12.2. *В определении момента остановки условие* (12.1) *можно заменить на эквивалентное*

$$\{\tau \ \leqslant \ n\} \in \mathscr{F}_n. \tag{12.2}$$

для всех $n \geqslant 0$, Кроме того, для момента остановки τ и любого $n \in \mathbb{N}$ выполняются следующие соотношения

$$\{\tau > n\} \in \mathscr{F}_n, \qquad \{\tau < n\} \in \mathscr{F}_{n-1}, \qquad \{\tau \geqslant n\} \in \mathscr{F}_{n-1}.$$

Доказательство. Пусть τ — момент остановки, т.е. для каждого $n \geqslant 0$ выполняется условие (12.1). Тогда из равенства

$$\{\tau \leqslant n\} \ = \ \bigcup_{k=0}^{n} \{\tau = k\}$$

и определения фильтрации следует выполнение условия (12.2).

Обратно, если для всех n выполняется условие (12.2), то из равенства

$$\{\tau = n\} = \{\tau \leqslant n\} \setminus \{\tau \leqslant n-1\}$$

следует выполнение условия (12.1).

Остальная часть утверждения следует из равенств

$$\{\tau > n\} \ = \ \overline{\{\tau \leqslant n\}}, \qquad \{\tau < n\} \ = \ \{\tau \leqslant n-1\}, \qquad \{\tau \geqslant n\} \ = \ \overline{\{\tau < n\}}.$$

□

Теорема 12.2. *Пусть τ и ν — моменты остановки. Тогда $\tau \wedge \nu = \min\{\tau, \nu\}$, $\tau \vee \nu = \max\{\tau, \nu\}$ и $\tau + \nu$ также будут моментами остановки.*

Доказательство. Утверждение теоремы следует из соотношений

$$\begin{aligned}
\{\tau \wedge \nu \leqslant n\} \ &= \ \{\tau \leqslant n\} \cup \{\nu \leqslant n\}, \\
\{\tau \vee \nu \leqslant n\} \ &= \ \{\tau \leqslant n\} \cap \{\nu \leqslant n\}, \\
\{\tau + \nu = n\} \ &= \ \bigcup_{k=0}^{n} \{\tau = k\}\{\nu = n-k\}.
\end{aligned}$$

□

Введём теперь понятие остановленного процесса. Пусть X — согласованный процесс, а τ — момент остановки. Под остановленным процессом будем понимать $X^\tau = (X_n^\tau\colon\ n \geqslant 0)$, где

$$X_n^\tau \;=\; X_{\tau\wedge n} \;=\; \sum_{k=0}^{n-1} X_k \mathbb{1}_{\{\tau=k\}} + X_n \mathbb{1}_{\{\tau\geqslant n\}}.$$

Теорема 12.3. *Пусть X — субмартингал, а τ — момент остановки. Тогда X^τ также будет субмартингалом.*

Доказательство. Заметим вначале, что

$$X_n^\tau \;=\; \sum_{k=0}^{n-1} X_k(\mathbb{1}_{\{\tau\geqslant k\}} - \mathbb{1}_{\{\tau\geqslant k+1\}}) + X_n \mathbb{1}_{\{\tau\geqslant n\}} \;=\; X_0 + \sum_{k=1}^{n} \Delta X_k \mathbb{1}_{\{\tau\geqslant k\}},$$

а $H = (H_n = \mathbb{1}_{\{\tau\geqslant n\}}\colon\ n \geqslant 1)$ — предсказуемый ограниченный неотрицательный процесс (см. предложение 12.2). Поэтому $X^\tau = X_0 + H \bullet X$ и утверждение следует из теоремы 12.1. □

Следствие 12.2. *Если X — супермартингал (мартингал), то X^τ также будет супермартингалом (мартингалом).*

Следующий результат показывает, что свойство мартингальности сохраняется и на ограниченных моментах остановки.

Теорема 12.4. *Пусть X — субмартингал, а τ и ν — моменты остановки, удовлетворяющие условию $\nu \leqslant \tau \leqslant N$, где $N \in \mathbb{N}$. Тогда $\mathsf{E}X_\nu \leqslant \mathsf{E}X_\tau$.*

Доказательство. Покажем, что $Y = X^\tau - X^\nu$ является субмартингалом. Действительно, используя для X_n^τ и X_n^ν представление из доказательства предыдущей теоремы, получаем $Y_0 = 0$ и

$$Y_n \;=\; X_n^\tau - X_n^\nu \;=\; \sum_{k=1}^{n} \left(\mathbb{1}_{\{\tau\geqslant k\}} - \mathbb{1}_{\{\nu\geqslant k\}}\right) \Delta X_k,$$

$n = 1, 2, \ldots$. Из условия теоремы следует, что при всех n имеет место включение $\{\nu \geqslant n\} \subseteq \{\tau \geqslant n\}$. Следовательно, $C_n = \mathbb{1}_{\{\tau\geqslant n\}} - \mathbb{1}_{\{\nu\geqslant n\}}$ является неотрицательной $\mathscr{F}_{n-1}$-измеримой случайной величиной. Замечая, что $Y = C \bullet X$, где $C = (C_n\colon\ n \geqslant 1)$ — предсказуемый ограниченный неотрицательный процесс, приходим к утверждению, что Y — субмартингал. Но тогда

$$0 \;=\; \mathsf{E}Y_0 \;\leqslant\; \mathsf{E}Y_N \;=\; \mathsf{E}X_\tau - \mathsf{E}X_\nu$$

и теорема доказана. □

Следствие 12.3. *Если в условиях теоремы X — супермартингал (мартингал), то $\mathsf{E}X_\nu \geqslant \mathsf{E}X_\tau$ ($\mathsf{E}X_\nu = \mathsf{E}X_\tau$).*

Неравенства. Приведём некоторые неравенства, связанные с мартингалами.

Теорема 12.5. *Пусть $X = (X_n\colon\ n \geqslant 0)$ — субмартингал. Тогда для любых $a > 0$ и $n \in \mathbb{N}$ выполняется неравенство*

$$\mathsf{P}\left(\max_{0\leqslant k\leqslant n} X_k \geqslant a\right) \leqslant \frac{1}{a}\mathsf{E}X_n^+. \tag{12.3}$$

Доказательство. Пусть $a > 0$ и $n \in \mathbb{N}$ фиксированы. Обозначим

$$A = \left\{\omega\colon\ \max_{0\leqslant k\leqslant n} X_k(\omega) \geqslant a\right\}$$

и введём в рассмотрение $\tau = \inf\{k\colon\ X_k \geqslant a\}$ $(\inf\varnothing = \infty)$ и $\nu = \tau \wedge n$. Легко видеть, что ν является моментом остановки. Поэтому по теореме 12.4 получаем

$$\mathsf{E}X_n \geqslant \mathsf{E}X_\nu = \mathsf{E}X_\nu \mathbb{1}_A + \mathsf{E}X_\nu \mathbb{1}_{\overline{A}} \geqslant a\mathsf{P}(A) + \mathsf{E}X_n \mathbb{1}_{\overline{A}},$$

откуда следует неравенство

$$a\mathsf{P}(A) \leqslant \mathsf{E}X_n - \mathsf{E}X_n\mathbb{1}_{\overline{A}} = \mathsf{E}X_n\mathbb{1}_A.$$

Замечая теперь, что $X_n\mathbb{1}_A \leqslant X_n^+$, приходим к неравенству (12.3). □

Теорема 12.6. *Пусть X — квадратично интегрируемый мартингал. Тогда для любых $a > 0$ и $n \in \mathbb{N}$ выполняется неравенство*

$$\mathsf{P}\left(\max_{0\leqslant k\leqslant n} |X_k| \geqslant a\right) \leqslant \frac{1}{a^2}\mathsf{E}X_n^2. \tag{12.4}$$

Доказательство. Поскольку $X^2 = (X_n^2\colon\ n \geqslant 0)$ является субмартингалом, то, применяя неравенство (12.3), получаем

$$\mathsf{P}\left(\max_{0\leqslant k\leqslant n} |X_k| \geqslant a\right) = \mathsf{P}\left(\max_{0\leqslant k\leqslant n} X_k^2 \geqslant a^2\right) \leqslant \frac{1}{a^2}\mathsf{E}X_n^2.$$

□

Теорема 12.7 (Неравенство Колмогорова). *Пусть $\xi_1, \ldots, \xi_n$ — независимые случайные величины с нулевыми средними и конечными дисперсиями $\mathsf{D}\xi_k = \sigma_k^2$, $k = 1, \ldots, n$. Тогда для любого $a > 0$ выполняется неравенство*

$$\mathsf{P}\left(\max_{0\leqslant k\leqslant n} |\xi_1 + \ldots + \xi_k| \geqslant a\right) \leqslant \frac{1}{a^2}\sum_{k=1}^n \sigma_k^2. \tag{12.5}$$

Доказательство. Пусть $\xi_1, \xi_2 \ldots$ — независимые случайные величины с нулевыми средними и конечными вторыми моментами. Определим фильтрацию $(\mathscr{F}_n)$, $\mathscr{F}_0 = \{\varnothing, \Omega\}$, $\mathscr{F}_n = \sigma(\xi_1, \ldots, \xi_n)$, $n \geqslant 1$, и рассмотрим процесс $S = (S_n\colon\ n \geqslant 0)$, положив $S_0 = 0$ и $S_n = \xi_1 + \ldots + \xi_n$, $n \geqslant 1$. Процесс S является мартингалом относительно фильтрации $(\mathscr{F}_n)$ и

$$\mathsf{E}S_n^2 = \mathsf{D}S_n = \sum_{k=1}^n \mathsf{D}\xi_k,$$

т. е. S — квадратично интегрируемый мартингал. Применяя к нему неравенство (12.4), получаем

$$\mathsf{P}\left(\max_{0\leqslant k\leqslant n}|S_k| \geqslant a\right) \leqslant \frac{1}{a^2}\mathsf{E}S_n^2 = \frac{1}{a^2}\sum_{k=1}^{n}\mathsf{D}\xi_k.$$

□

Неравенство Колмогорова содержит в качестве частного случая неравенство Чебышёва и используется для доказательства усиленного закона больших чисел Колмогорова.

§ 13. Таблица нормального распределения

Важность нормального (или гауссовского) распределения следует из центральной предельной теоремы. Его популярность так велика, что на одной из самых распространенных купюр в десять немецких марок (до ввода евро) был изображен портрет Гаусса и приведен график плотности нормального распределения.

Рис. 1. 10 DM Serie4 Vorderseite

Напомним, что нормальное распределение определяется двумя параметрами $a \in \mathbb{R}$ и $\sigma > 0$ посредством плотности

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Случай $a = 0$ и $\sigma = 1$ относится к стандартному нормальному нормальному распределению. Его функция распределения

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-u^2/2} du$$

связана с функцией распределения $F(x)$, отвечающего параметрам a и σ равенством

$$F(x) = \Phi\left(\frac{x-a}{\sigma}\right).$$

Ниже приводится таблица значений функции

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{0}^{x} e^{-u^2/2} du, \qquad x > 0.$$

Таблица нормального распределения

В таблице приводятся значения функции $\Phi_0(x) = \dfrac{1}{\sqrt{2\pi}} \int_0^x e^{-y^2/2}\, dy$ для аргументов $x \in [0; 3{,}99]$. Значение функции $\Phi_0(x)$ находится на пересечении строки, дающей целую часть и десятые доли аргумента x, и столбца, дающего сотые доли аргумента x.

Если $\eta \sim \mathcal{N}(0, 1)$ — случайная величина со стандартным нормальным распределением, то $\mathbb{P}(a < \eta < b) = \Phi_0(b) - \Phi_0(a)$. Функция $\Phi_0(x)$ отличается от функции распределения η на константу $\frac{1}{2}$, т.е. $\mathbb{P}(\eta < x) = \frac{1}{2} + \Phi_0(x)$.

Функция $\Phi_0(x)$ является нечетной, т.е. $\Phi_0(-x) = -\Phi_0(x)$. В практических расчетах значения $\Phi_0(x)$ при $x \geq 4$ можно считать равными $\frac{1}{2}$.

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	x
0,0	0,00000	0,00399	0,00798	0,01197	0,01595	0,01994	0,02392	0,02790	0,03188	0,03586	0,0
0,1	0,03983	0,04380	0,04776	0,05172	0,05567	0,05962	0,06356	0,06749	0,07142	0,07535	0,1
0,2	0,07926	0,08317	0,08706	0,09095	0,09483	0,09871	0,10257	0,10642	0,11026	0,11409	0,2
0,3	0,11791	0,12172	0,12552	0,12930	0,13307	0,13683	0,14058	0,14431	0,14803	0,15173	0,3
0,4	0,15542	0,15910	0,16276	0,16640	0,17003	0,17364	0,17724	0,18082	0,18439	0,18793	0,4
0,5	0,19146	0,19497	0,19847	0,20194	0,20540	0,20884	0,21226	0,21566	0,21904	0,22240	0,5
0,6	0,22575	0,22907	0,23237	0,23565	0,23891	0,24215	0,24537	0,24857	0,25175	0,25490	0,6
0,7	0,25804	0,26115	0,26424	0,26730	0,27035	0,27337	0,27637	0,27935	0,28230	0,28524	0,7
0,8	0,28814	0,29103	0,29389	0,29673	0,29955	0,30234	0,30511	0,30785	0,31057	0,31327	0,8
0,9	0,31594	0,31859	0,32121	0,32381	0,32639	0,32894	0,33147	0,33398	0,33646	0,33891	0,9
1,0	0,34134	0,34375	0,34614	0,34849	0,35083	0,35314	0,35543	0,35769	0,35993	0,36214	1,0
1,1	0,36433	0,36650	0,36864	0,37076	0,37286	0,37493	0,37698	0,37900	0,38100	0,38298	1,1
1,2	0,38493	0,38686	0,38877	0,39065	0,39251	0,39435	0,39617	0,39796	0,39973	0,40147	1,2
1,3	0,40320	0,40490	0,40658	0,40824	0,40988	0,41149	0,41308	0,41466	0,41621	0,41774	1,3
1,4	0,41924	0,42073	0,42220	0,42364	0,42507	0,42647	0,42785	0,42922	0,43056	0,43189	1,4
1,5	0,43319	0,43448	0,43574	0,43699	0,43822	0,43943	0,44062	0,44179	0,44295	0,44408	1,5
1,6	0,44520	0,44630	0,44738	0,44845	0,44950	0,45053	0,45154	0,45254	0,45352	0,45449	1,6
1,7	0,45543	0,45637	0,45728	0,45818	0,45907	0,45994	0,46080	0,46164	0,46246	0,46327	1,7
1,8	0,46407	0,46485	0,46562	0,46638	0,46712	0,46784	0,46856	0,46926	0,46995	0,47062	1,8
1,9	0,47128	0,47193	0,47257	0,47320	0,47381	0,47441	0,47500	0,47558	0,47615	0,47670	1,9
2,0	0,47725	0,47778	0,47831	0,47882	0,47932	0,47982	0,48030	0,48077	0,48124	0,48169	2,0
2,1	0,48214	0,48257	0,48300	0,48341	0,48382	0,48422	0,48461	0,48500	0,48537	0,48574	2,1
2,2	0,48610	0,48645	0,48679	0,48713	0,48745	0,48778	0,48809	0,48840	0,48870	0,48899	2,2
2,3	0,48928	0,48956	0,48983	0,49010	0,49036	0,49061	0,49086	0,49111	0,49134	0,49158	2,3
2,4	0,49180	0,49202	0,49224	0,49245	0,49266	0,49286	0,49305	0,49324	0,49343	0,49361	2,4
2,5	0,49379	0,49396	0,49413	0,49430	0,49446	0,49461	0,49477	0,49492	0,49506	0,49520	2,5
2,6	0,49534	0,49547	0,49560	0,49573	0,49585	0,49598	0,49609	0,49621	0,49632	0,49643	2,6
2,7	0,49653	0,49664	0,49674	0,49683	0,49693	0,49702	0,49711	0,49720	0,49728	0,49736	2,7
2,8	0,49744	0,49752	0,49760	0,49767	0,49774	0,49781	0,49788	0,49795	0,49801	0,49807	2,8
2,9	0,49813	0,49819	0,49825	0,49831	0,49836	0,49841	0,49846	0,49851	0,49856	0,49861	2,9
3,0	0,49865	0,49869	0,49874	0,49878	0,49882	0,49886	0,49889	0,49893	0,49896	0,49900	3,0
3,1	0,49903	0,49906	0,49910	0,49913	0,49916	0,49918	0,49921	0,49924	0,49926	0,49929	3,1
3,2	0,49931	0,49934	0,49936	0,49938	0,49940	0,49942	0,49944	0,49946	0,49948	0,49950	3,2
3,3	0,49952	0,49953	0,49955	0,49957	0,49958	0,49960	0,49961	0,49962	0,49964	0,49965	3,3
3,4	0,49966	0,49968	0,49969	0,49970	0,49971	0,49972	0,49973	0,49974	0,49975	0,49976	3,4
3,5	0,49977	0,49978	0,49978	0,49979	0,49980	0,49981	0,49981	0,49982	0,49983	0,49983	3,5
3,6	0,49984	0,49985	0,49985	0,49986	0,49986	0,49987	0,49987	0,49988	0,49988	0,49989	3,6
3,7	0,49989	0,49990	0,49990	0,49990	0,49991	0,49991	0,49992	0,49992	0,49992	0,49992	3,7
3,8	0,49993	0,49993	0,49993	0,49994	0,49994	0,49994	0,49994	0,49995	0,49995	0,49995	3,8
3,9	0,49995	0,49995	0,49996	0,49996	0,49996	0,49996	0,49996	0,49996	0,49997	0,49997	3,9
x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	x