Конспект лекций по теории вероятностей

Оглавление

§ 1. Введение

В окружающем мире существуют явления, которые даже при практически полном повторении начальных условий могут принимать разные формы (или вообще не происходить). При бросании игрального кубика может выпасть 1, 2, 3, 4, 5 или 6 очков; при розыгрышах тиражей лотереи из лототрона выкатываются разные наборы выигрышных номеров, и т.п.

Явления, о которых заранее нельзя сказать, произойдут они или нет, а если произойдут — то в какой форме, мы называем *случайными*.

Несмотря на видимое отсутствие причинных связей, многие случайные события подчиняются некоторым закономерностям. Одна из них называется *законом устойчивости частот*.

Пусть один и тот же эксперимент с несколькими возможными исходами (например, бросание монеты или игрального кубика) многократно проводится практически в одних и тех же условиях; это называется *серией экспериментов*. Обозначим возможные исходы эксперимента числами 1, 2, $\ldots$, M (в примерах с бросанием монеты или кубика соответственно $M = 2$ или $M = 6$), и пусть ξ_1, ξ_2, $\ldots$ — исходы первого, второго, $\ldots$ экспериментов. Для натурального числа T найдем числа появлений (*частоты*, или, точнее, *абсолютные частоты*) появлений каждого исхода эксперимента в T испытаниях:

$$\nu_k(T) = \sum_{t=1}^{T} \mathbb{I}\left\{\xi_t = k\right\}, \quad k \in \{1, \ldots, M\};$$

здесь и далее символом $\mathbb{I}\{A\}$ обозначается *индикатор* события (или условия) A: $\mathbb{I}\{A\} = 1$, если событие A происходит (условие A выполняется), и $\mathbb{I}\{A\} = 0$ в противном случае.

Закон устойчивости частот проявляется в том, что с ростом числа испытаний T последовательность значений вектора *относительных частот*

$$\nu(T) = \left(\frac{\nu_1(T)}{T}, \frac{\nu_2(T)}{T}, \ldots, \frac{\nu_M(T)}{T}\right), \quad \frac{\nu_1(T)}{T} + \frac{\nu_2(T)}{T} + \ldots + \frac{\nu_M(T)}{T} = 1,$$

ведет себя как сходящаяся к какому-то «предельному» вектору

$$p = (p_1, \ldots, p_M), \quad p_1, \ldots, p_M \geqslant 0, \quad p_1 + \ldots + p_M = 1.$$

Величины $p_1, \ldots, p_M$ называют *вероятностями исходов* 1, 2, ..., M. Например, при бросании правильного кубика вероятность каждого из 6 исходов равна $\frac{1}{6}$.

Величины $p_k = \lim_{T\to\infty} \frac{1}{T} \sum\limits_{t=1}^{T} \mathbb{I}\{\xi_t = k\}$ с точки зрения математического анализа являются пределами по Чезаро последовательностей $\mathbb{I}\{\xi_1 = k\}, \mathbb{I}\{\xi_2 = k\}, \ldots$, состоящих из 0 и 1. Легко привести примеры числовых (даже двоичных) последовательностей, для которых предел по Чезаро не существует, т. е. закон устойчивости частот для них не выполняется.

Закон устойчивости частот — свойство, которое допускает практическую проверку (если при этом сохраняются условия проводимых экспериментов). Разумная математическая модель случайных явлений должна обладать свойствами, соответствующими этому эмпирическому закону.

Математическое изучение свойств случайных явлений началось в XVII веке в связи с практическими потребностями организации лотерей, страхового дела и т. п., и привело к возникновению и развитию специального раздела математики — теории вероятностей.

Математика отличается от других естественных наук тем, что она не столько изучает окружающий нас мир, сколько строит и исследует формальные аксиоматические модели, с разной степенью точности отражающие те или иные свойства реального мира. Например, наряду с геометрией евклидовых пространств рассматриваются геометрии Лобачевского, Римана и т.п. Если математическая модель достаточно хорошо отражает свойства реального явления, то свойствам, обнаруженным при исследовании математической модели, могут соответствовать свойства реального явления, в том числе и те, на которые ранее не обращали внимание.

Современная аксиоматика теории вероятностей была создана А.Н.Колмогоровым в начале 30-х годов XX века. Вероятностная аксиоматика (в отличие от привычных аксиом геометрии, математического анализа, механики) приводит к моделям, которые на первый взгляд имеют мало общего с нашими интуитивными представлениями о случайности. В колмогоровских аксиомах теории вероятностей по сути дела заложены следующие неформальные принципы:

а) все события полностью определяются «начальными условиями»,

б) неопределенность порождается недостатком информации о «начальных условиях».

Эти принципы соответствуют определению случайности в терминах материалистической философии, согласно которой все явления в окружающем мире подчиняются объективным законам природы: *случайность — это форма проявления необходимости, возникающая под воздействием большого числа посторонних (для необходимости) причин.*

Например, массовое пренебрежение правилами пожарной безопасности создает условия для возникновения пожаров, однако где именно произойдет очередной пожар, зависит от конкретного стечения обстоятельств.

Заметим, что в религиозных учениях, предполагающих, что все происходит по воле Бога, тоже нет места для случайности.

§ 2. Основные понятия теории вероятностей

Основными понятиями теории вероятностей являются *вероятностное пространство*, *событие* и *случайная величина*. Понятия «случайность» в математической теории вероятностей нет; оно может возникать только при описании свойств практических ситуаций, находящих отражение в свойствах абстрактных математических объектов.

Исходным объектом вероятностной модели является *пространство элементарных событий*. Его принято обозначать прописной греческой буквой Ω, произвольную точку этого пространства называют *элементарным событием* и обозначают строчной буквой ω: $\omega \in \Omega$. События — это некоторые совокупности элементарных событий.

При построении вероятностной модели простого случайного явления пространство элементарных событий Ω можно представлять себе как множество всех возможных реализаций ω этого явления. Как только реализация ω выбрана, все, что к этому явлению относится, становится однозначно определенным.

Для модели однократного бросания игрального кубика пространство элементарных событий можно выбрать состоящим из 6 элементов: $\Omega = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$; каждый элемент $\omega \in \Omega$ соответствует числу очков, выпадающему при бросании кубика. Для модели двукратного бросания кубика пространством элементарных событий может быть прямое произведение $\{1,\ 2,\ 3,\ 4,\ 5,\ 6\} \times \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$ или множество пар $\{(i,j)\colon i,j \in \{1,\dots,6\}\}$.

Пространство элементарных событий для одной и той же задачи можно выбирать разными способами. Например, в качестве пространства элементарных событий в модели бросания игрального кубика можно выбрать $\Omega = [0,1)$ и считать, что выпадает k очков, если $\omega \in \left[\frac{k-1}{6}, \frac{k}{6}\right)$. Этот же полуинтервал можно использовать как пространство элементарных событий для модели многократного бросания кубика, если представить числа $\omega \in [0,1)$ в виде дробей в 6-ичной системе счисления: $\omega = \sum_{t\geqslant 1} \frac{x_t(\omega)}{6^t}$, где $x_t(\omega) \in \{0,1,\dots,5\}$ при всех $t \geqslant 1$, и считать, что при t-м бросании кубика выпадает k очков, если $x_t(\omega) = k-1$. Тогда, например, событию «при втором бросании кубика выпало 1 очко» соответствует множество $A_2(1) = \bigcup_{m=0}^{5} \left[\frac{m}{6}, \frac{m}{6} + \frac{1}{36}\right)$. Удачный выбор пространства элементарных событий может упростить решение задачи.

Следующий шаг в построении вероятностного пространства состоит в выделении из множества 2^Ω всех подмножеств Ω некоторой совокупности $\mathcal{F}$ подмножеств, которые далее будут называться *событиями*. Каждому событию сопоставляется число из отрезка $[0,\ 1]$, которое называют *вероятностью события*. Множество событий должно быть замкнуто относительно стандартных теоретико-множественных операций: объединения, пересечения и дополнения, а вероятности событий удовлетворять некоторым естественным условиям.

Когда говорят «происходит событие $A \in \mathcal{F}$», то имеют в виду, что рассматривается множество $A \subset \Omega$ или какое-нибудь его подмножество. Объединение $A \cup B$ событий A и B иногда называют их *суммой*, пересечение $A \cap B$ — *произведением* и обозначают AB, дополнение $A^c = \overline{A} = \Omega \setminus A$ — *противоположным событием.* Если A и B — события, то говорят, что на множестве $A \cap B$ события A и B *происходят одновременно*. Если $A \cap B = \varnothing$, то события A и B называют *несовместными*. Если $A \subseteq B$, то говорят, что событие A *влечет* событие B.

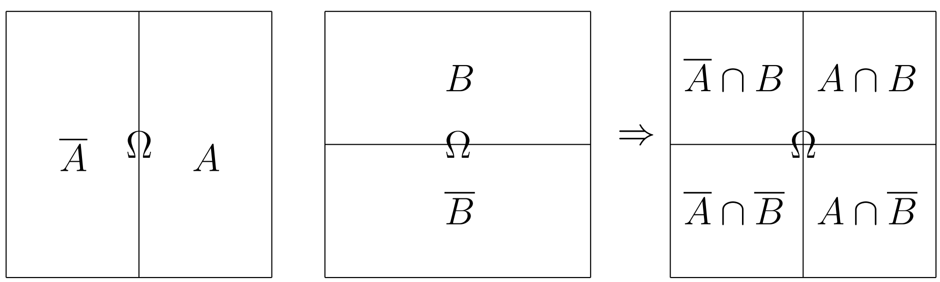
Операции $\cup$ и $\cap$ коммутативны, ассоциативны и транзитивны: для любых подмножеств $A, B, C \subseteq \Omega$

$$AB = BA, \quad A \cup B = B \cup A,$$
$$(AB)C = A(BC), \quad (A \cup B) \cup C = A \cup (B \cup C),$$
$$A(B \cup C) = AB \cup AC, \quad A \cup (BC) = (A \cup B)(A \cup C).$$

При решении вероятностных задач часто используются соотношения

$$\text{если } A \subset B, \text{ то } \overline{B} \subset \overline{A},$$
$$\overline{AB} = \overline{A} \cup \overline{B}, \quad \overline{A \cup B} = \overline{A} \cap \overline{B}, \quad A \cup B = A \cup (B \setminus A), \quad A \backslash B = A \cap \overline{B},$$

которые можно проверить по рисунку



а также их обобщения на произвольные множества событий

$$\overline{\bigcap_{k \geqslant 1} A_k} = \bigcup_{k \geqslant 1} \overline{A_k}, \quad \overline{\bigcup_{k \geqslant 1} A_k} = \bigcap_{k \geqslant 1} \overline{A_k}.$$

Определение 1. Совокупность $\mathcal{A}$ подмножеств Ω называется *алгеброй множеств*, если выполнены следующие условия:

1) $\varnothing \in \mathcal{A}, \;\; \Omega \in \mathcal{A}$;

2) если $A \in \mathcal{A}$, то и $\Omega \setminus A = A^c = \overline{A} \in \mathcal{A}$;

3) если $A, B \in \mathcal{A}$, то $A \cup B \in \mathcal{A}$ и $A \cap B \in \mathcal{A}$.

В условиях 1) и 3) вторые части являются следствиями первых и условия 2).

Примеры алгебр множеств:
а) совокупность 2^Ω всех подмножеств произвольного множества Ω,
б) совокупность всех конечных объединений замкнутых справа полуинтервалов $(a, b])$, содержащихся в $(0, 1]$.

Определение 2. Алгебра $\mathcal{F}$ подмножеств множества Ω называется *σ-алгеброй*, если для любой конечной или счетной совокупности множеств $A_1, A_2, \ldots \in \mathcal{F}$

$$\bigcup_{n \geqslant 1} A_n \in \mathcal{F} \quad \text{и} \quad \bigcap_{n \geqslant 1} A_n \in \mathcal{F}.$$

Множество всех подмножеств Ω, очевидно, является и алгеброй, и σ-алгеброй. Его выбирают в качестве множества событий $\mathcal{F}$, если пространство элементарных событий Ω конечно или счетно.

Пусть $\mathcal{F}_k, k \in K$, — конечная или бесконечная совокупность алгебр (или σ-алгебр) подмножеств множества Ω. *Пересечением* алгебр (соответственно, σ-алгебр) $\mathcal{F}_k, k \in K$, называется совокупность всех таких подмножеств $A \subseteq \Omega$, что $A \in \mathcal{F}_k$ при любом $k \in K$.

Задача: проверить, что пересечение алгебр (σ-алгебр) является алгеброй (σ-алгеброй) и что обычное теоретико-множественное объединение алгебр (σ-алгебр) может не быть алгеброй (σ-алгеброй).

Определение. Пусть $\mathcal{M}$ — совокупность подмножеств множества Ω. Пересечение всех алгебр подмножеств множества Ω, которые содержат $\mathcal{M}$, называется *наименьшей алгеброй*, порожденной $\mathcal{M}$. Пересечение всех σ-алгебр подмножеств множества Ω, которые содержат $\mathcal{M}$, называется *наименьшей σ-алгеброй*, порожденной $\mathcal{M}$.

Указанные в определении пересечения не пусты, так как, во-первых, алгебры (σ-алгебры), содержащие любую совокупность подмножеств $\mathcal{M}$, существуют (например, множество 2^Ω всех подмножеств Ω является и алгеброй, и σ-алгеброй и содержит любую совокупность $\mathcal{M}$), а во-вторых, все подмножества, содержащиеся в $\mathcal{M}$, входят в эти пересечения.

Если Ω — действительная прямая $\mathbb{R}$ (или евклидово пространство $\mathbb{R}^d$), а $\mathcal{M}$ — совокупность интервалов (или открытых множеств), то наименьшая σ-алгебра, порожденная $\mathcal{M}$, называется *борелевской σ-алгеброй*.

Определение. Пусть $\mathcal{F}$ — алгебра или σ-алгебра подмножеств Ω. Функция $\mu\colon \mathcal{F} \to \mathbb{R}$ называется мерой, если она *аддитивна*, т. е. $\mu(A \cup B) = \mu(A) + \mu(B)$ для любых таких $A, B \in \mathcal{F}$, что $A \cap B = \varnothing$.

Мера μ называется *счётно-аддитивной*, если $\mu\left(\cup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k)$ для любой последовательности попарно не пересекающихся множеств $A_1, A_2, \ldots \in \mathcal{F}$.

Примеры мер на $\mathbb{R}$:
а) мера Лебега $\mathrm{mes}(A)$,
б) $\mu(A)$ — количество целых чисел, принадлежащих A,
в) $\mathrm{mes}(A) - \mu(A)$ (принимает как положительные, так и отрицательные значения).

Определение. Мера μ на σ-алгебре $\mathcal{F}$ подмножеств Ω называется *вероятностной*, если все ее значения лежат в отрезке $[0,\ 1]$ и $\mu(\Omega) = 1$.

Обычно меры определяют их значениями на совокупности множеств, порождающих алгебру или σ-алгебру; с этой совокупности множеств мера распространяется на остальные множества (σ) алгебры с помощью свойства аддитивности. Теорема Каратеодори указывает условия, при которых можно продолжить счетно-аддитивную меру с алгебры на σ-алгебру.

Теорема Каратеодори. *Если на алгебре $\mathcal{A}$ подмножеств пространства Ω определена вероятностная мера μ, обладающая свойством счетной аддитивности, то эту меру можно продолжить на наименьшую σ-алгебру $\mathcal{F} = \sigma(\mathcal{A})$, порожденную алгеброй $\mathcal{A}$, с сохранением свойства счетной аддитивности.*

Доказательство теоремы Каратеодори можно найти в учебнике А.А.Боровкова [1].

Примером вероятностной не счетно аддитивной меры на алгебре является мера μ на множестве $\mathbb{Q}_1 = \mathbb{Q} \cap [0,\ 1]$ рациональных чисел отрезка $[0,\ 1]$, определенная равенствами $\mu(\mathbb{Q}_1 \cap [a,\ b]) = b - a,\ 0 \leqslant a \leqslant b \leqslant 1$. Для любого рационального числа $r \in \mathbb{Q}_1$ по определению $\mu(r) = 0$, и так как множество рациональных чисел счетно, то $\sum_{r \in \mathbb{Q}_1} \mu(r) = 0 \neq \mu(\mathbb{Q}_1) = 1$.

Определение 3 (Основная часть аксиоматики Колмогорова). *Вероятностным пространством* называется тройка $(\Omega,\ \mathcal{F},\ \mathbf{P})$, где Ω — *пространство элементарных событий*, $\mathcal{F}$ — σ-алгебра подмножеств Ω, называемых *событиями*, $\mathbf{P}\colon \mathcal{F} \to \mathbb{R}$ — функция, называемая *вероятностью* и удовлетворяющая следующим условиям:

1^o. Неотрицательность: $\mathbf{P}(A) \geq 0$ при любом $A \in \mathcal{F}$.

2^o. Нормированность: $\mathbf{P}(\Omega) = 1$.

3^o. Счетная аддитивность: Если события не более чем счетной совокупности A_1, A_2, ... попарно несовместны, то

$$\mathbf{P}\left(\bigcup_{n\geqslant 1} A_n\right) = \sum_{n\geqslant 1} \mathbf{P}(A_n).$$

Еще одним важным понятием теории вероятностей является понятие случайной величины. Термин «величина» не означает, что значениями случайной величины обязательно являются числа; например, случайная подстановка или случайная точка в единичном квадрате тоже являются случайными величинами. Дадим пока упрощенное определение случайной величины.

Определение. Если $(\Omega, \mathcal{F}, \mathbf{P})$ — вероятностное пространство, то *случайной величиной* $\xi = \xi(\omega)$ называется функция, определенная на пространстве элементарных событий Ω.

Модель трехкратного бросания монеты можно построить на пространстве элементарных событий $\Omega = \{0, 1, 2, 3, 4, 5, 6, 7\}$ и определить результаты (0 или 1) трех бросаний тремя функциями $f_1, f_2, f_3 \colon \Omega \to \{0, 1\}$, полагая для $\omega \in \Omega$

$$f_1(\omega) = \omega \bmod 2, \quad f_2(\omega) = [\omega/2] \bmod 2, \quad f_3(\omega) = [\omega/4] \bmod 2.$$

Функции f_1, f_2, f_3 в этой модели являются случайными величинами, значения которых определяются выбором элемента $\omega \in \Omega$. С точки зрения философского определения случайности функции f_1, f_2, f_3 — это закономерности (необходимость), а выбор элемента $\omega \in \Omega$ и тем самым значений случайных величин определяются посторонними для этих закономерностей причинами.

Пример: Конечные вероятностные пространства.

Пусть в тройке $(\Omega, \mathcal{F}, \mathbf{P})$ пространство элементарных событий $\Omega = \{\omega\}$ — конечное множество, σ-алгебра (в данном случае — алгебра) $\mathcal{F}$ — совокупность всех подмножеств Ω. Число элементов любого конечного множества B будем обозначать записью $|B|$. Вероятность $\mathbf{P}$ можно задать, определив для каждого $\omega \in \Omega$ неотрицательное число $\mathbf{P}(\omega)$ так, чтобы

$$\sum_{\omega\in\Omega} \mathbf{P}(\omega) = 1.$$

Для произвольного события $A \subseteq \Omega$ положим

$$\mathbf{P}(A) = \sum_{\omega \in A} \mathbf{P}(\omega).$$

Выполнение условий 1^o–3^o в этом случае очевидно.

Частным случаем является так называемое *классическое определение вероятности*, когда $\mathbf{P}(\omega) = \frac{1}{|\Omega|}$ для всех $\omega \in \Omega$. Тогда

$$\mathbf{P}(A) = \frac{|A|}{|\Omega|} \text{ для любого } A \subseteq \Omega.$$

Теория вероятностей — это раздел математики, связанный с методами построения совокупностей событий и исследования их вероятностей.

Основные свойства вероятности.

1) **Вероятности противоположных событий:** $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$, в частности, $\mathbf{P}\{\varnothing\} = 0$ и $\mathbf{P}(A) \leqslant 1$ для любого $A \in \mathcal{F}$.

   Первое равенство следует из условия 3^o, так как $A \cup A^c = \Omega$, $A \cap A^c = \varnothing$. При $A = \Omega$ получаем второе равенство. Наконец, $\mathbf{P}(A) \leqslant 1$, так как должно выполняться условие $\mathbf{P}(A^c) = 1 - \mathbf{P}(A) \geqslant 0$.

2) **Монотонность по включению:** Если $A \subseteq B$, то $\mathbf{P}\{A\} \leqslant \mathbf{P}\{B\}$ и $\mathbf{P}\{B \setminus A\} = \mathbf{P}\{B\} - \mathbf{P}\{A\}$.

   Действительно, если $A \subseteq B$, то $B = A \cup (B \setminus A)$ и $A \cap (B \setminus A) = \varnothing$. Применим условие 3^o к последовательности A, $B \setminus A$:

   $$\mathbf{P}\{B\} = \mathbf{P}\{A\} + \mathbf{P}\{B \setminus A\}, \tag{1}$$

   что эквивалентно доказываемому равенству. Неравенство $\mathbf{P}\{A\} \leqslant \mathbf{P}\{B\}$ следует из соотношения (1) и условия 1°.

3) **Конечная (и счетная) полуаддитивность:** Для любых событий $A_1, A_2, \ldots \subset \Omega$ и для любого $n \leqslant \infty$

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

Положим $B_1 = A_1$, $B_k = A_k \backslash (A_1 \cup \ldots \cup A_{k-1})$, $k = 2, 3, \ldots$. Докажем, что при любом $m \leqslant \infty$

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

Из включений $B_k \subset A_k, k = 1, 2, \ldots$, следует, что $\cup_{k=1}^{m} B_k \subset \cup_{k=1}^{m} A_k$. С другой стороны, для каждого элемента $\omega \in \cup_{k=1}^{m} A_k$ определено конечное значение $k(\omega) = \min\{k \geqslant 1 \colon \omega \in A_k\} \leqslant m$, и тогда $\omega \in B_{k(\omega)} \subseteq \cup_{k=1}^{m} B_k$, поэтому $\cup_{k=1}^{m} A_k \subseteq \cup_{k=1}^{m} B_k$.

События $B_1, \ldots, B_n$ попарно не пересекаются: если $1 \leqslant m < n < \infty$, то

$$\{\omega \in B_n\} \Leftrightarrow \{\omega \in A_n, \omega \notin A_j \, \forall j = 1, ..., n-1\} \Rightarrow \{\omega \notin A_m\} \Rightarrow \{\omega \notin B_m\}.$$

Используем теперь условие 3^o и свойство 2):

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

4) **Вероятность объединения двух событий:** Для любых событий A и B

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B). \tag{2}$$

Так как $A \cup B = (A \setminus AB) \cup (B \setminus AB) \cup AB$ и события $A \backslash AB$, $B \backslash AB$ и AB попарно не пересекаются, то по условию 3^o

$$\mathbf{P}(A \cup B) = \mathbf{P}(A \setminus AB) + \mathbf{P}(B \setminus AB) + \mathbf{P}(AB). \tag{3}$$

Так как $AB \subseteq A$ и $AB \subseteq B$, то по свойству монотонности $\mathbf{P}(A \setminus AB) = \mathbf{P}(A) - \mathbf{P}(AB)$ и $\mathbf{P}(B \setminus AB) = \mathbf{P}(B) - \mathbf{P}(AB)$. Подставляя эти равенства в (3) и приводя подобные члены, получаем (2).

5) **Формула включения-исключения.** Для любых событий

$A_1,\ A_2,\ \ldots,\ A_n$

$$\mathbf{P}\left\{\bigcup_{k=1}^{n} A_k\right\} =$$

$$= \sum_{k=1}^{n} \mathbf{P}\left\{A_k\right\} - \sum_{1\leqslant i_1<i_2\leqslant n} \mathbf{P}\{A_{i_1}A_{i_2}\} + \sum_{1\leqslant i_1<i_2<i_3\leqslant n} \mathbf{P}\{A_{i_1}A_{i_2}A_{i_3}\} - \ldots =$$

$$= \sum_{m=1}^{n} (-1)^{m-1} \sum_{1\leqslant i_1<\ldots<i_m\leqslant n} \mathbf{P}\left\{A_{i_1}A_{i_2}\ldots A_{i_m}\right\}. \tag{4}$$

Для доказательства воспользуемся принципом математической индукции. При $n = 2$ утверждение совпадает с 4). Допустим, что при некотором $n \geqslant 2$ утверждение справедливо, и покажем, что тогда оно справедливо и при $n + 1$.

Пусть $A_1, \ldots, A_{n+1}$ — произвольные события. Мы должны показать, что

$$\mathbf{P}\left\{\bigcup_{k=1}^{n+1} A_k\right\} = \sum_{m=1}^{n+1} (-1)^{m-1} \sum_{1\leqslant i_1<\ldots<i_m\leqslant n+1} \mathbf{P}\left\{A_{i_1}A_{i_2}\ldots A_{i_m}\right\}. \tag{5}$$

Если $B = A_1 \cup \ldots \cup A_n$, то $\cup_{k=1}^{n+1} A_k = B \cup A_{n+1}$ и в силу 4) и предположения индукции

$$\mathbf{P}\left\{\bigcup_{k=1}^{n+1} A_k\right\} = \mathbf{P}\{B\} + \mathbf{P}\{A_{n+1}\} - \mathbf{P}\{B \cap A_{n+1}\}, \tag{6}$$

где $\mathbf{P}\{B\}$ есть правая часть (4). Правая часть (5) отличается от правой части (4) только слагаемыми, содержащими A_{n+1}, и нам осталось доказать, что

$$\mathbf{P}\{A_{n+1}\} - \mathbf{P}\{B \cap A_{n+1}\} = \sum_{m=1}^{n+1} (-1)^{m-1} \sum_{1\leqslant i_1<\ldots<i_m=n+1} \mathbf{P}\left\{A_{i_1}A_{i_2}\ldots A_{i_m}\right\}.$$

Слагаемое в правой части, соответствующее $m = 1$, есть $\mathbf{P}\{A_{n+1}\}$. Используя предположение индукции, покажем, что остальные слагаемые соответствуют $(-\mathbf{P}\{B\cap A_{n+1}\})$:

$$-\mathbf{P}\{B\cap A_{n+1}\}=-\mathbf{P}\Big\{\Big(\bigcup_{k=1}^{n}A_k\Big)\cap A_{n+1}\Big\}=-\mathbf{P}\Big\{\bigcup_{k=1}^{n}(A_k\cap A_{n+1})\Big\}=$$

$$=-\sum_{m=1}^{n}(-1)^{m-1}\sum_{1\leqslant i_1<\ldots<i_m\leqslant n}\mathbf{P}\left\{\bigcap_{j=1}^{m}(A_{i_j}\cap A_{n+1})\right\}=$$

$$=\sum_{m=1}^{n}(-1)^{m}\sum_{1\leqslant i_1<\ldots<i_m<i_{m+1}=n+1}\mathbf{P}\left\{A_{i_1}\ldots A_{i_m}A_{i_{m+1}}\right\}=$$

$$=\sum_{m'=2}^{n+1}(-1)^{m'-1}\sum_{1\leqslant i_1<\ldots<i_{m'-1}<i_{m'}=n+1}\mathbf{P}\left\{A_{i_1}\ldots A_{i_{m'-1}}A_{i_{m'}}\right\}.$$

6) **Теорема о непрерывности вероятностной меры**. 1. *Для любой неубывающей последовательности событий* $A_1\subseteq A_2\subseteq\ldots$

$$\mathbf{P}\left\{\bigcup\nolimits_{n=1}^{\infty}A_n\right\}=\lim_{n\to\infty}\mathbf{P}\{A_n\}.$$

2. *Для любой невозрастающей последовательности событий* $B_1\supseteq B_2\supseteq\ldots$

$$\mathbf{P}\left\{\bigcap\nolimits_{n=1}^{\infty}B_n\right\}=\lim_{n\to\infty}\mathbf{P}\{B_n\}.$$

В частности, если $B_1\supseteq B_2\supseteq\ldots$ *и* $\cap_{n=1}^{\infty}B_n=\varnothing$, *то*

$$\lim_{n\to\infty}\mathbf{P}\{B_n\}=0.$$

Доказательство. 1. Аналогично доказательству свойства 3) по неубывающей последовательности событий $\{A_n\}$ построим последовательность несовместных событий $B_1=A_1$, $B_2=A_2\setminus A_1$, $B_3=A_3\setminus A_2$ и т.д. Так же, как в доказательстве свойства 3), можно показать, что $\cup_{n=1}^{m}A_n=\sqcup_{n=1}^{m}B_n$, $m\leqslant\infty$, кроме того, $\mathbf{P}\{B_n\}=\mathbf{P}\{A_n\}-\mathbf{P}\{A_{n-1}\}$ при $n\geqslant 2$ по свойству 2). Тогда по условию счетной аддитивности

$$\mathbf{P}\left\{\cup_{n=1}^{\infty}A_n\right\}=\mathbf{P}\left\{\sqcup_{n=1}^{\infty}B_n\right\}=\sum_{n=1}^{\infty}\mathbf{P}\{B_n\}=$$

$$=\lim_{N\to\infty}\sum_{n=1}^{N}\mathbf{P}\{B_n\}=\lim_{N\to\infty}\sum_{n=1}^{N}(\mathbf{P}\{A_n\}-\mathbf{P}\{A_{n-1}\})=\lim_{N\to\infty}\mathbf{P}\{A_N\},$$

что и требовалось доказать.

2. Докажем, что из п.1 следует п.2. Если $B_1 \supseteq B_2 \supseteq \ldots$ и $A_n = \overline{B_n} = \Omega \setminus B_n$, $n \geqslant 1$, то $A_1 \subseteq A_2 \subseteq \ldots$ и по доказанному в п.1 и по свойству 1)

$$\mathbf{P}\left\{\cup_{n=1}^{\infty} A_n\right\} = \lim_{N\to\infty} \mathbf{P}\{A_N\} = \lim_{N\to\infty} (1 - \mathbf{P}\{B_N\}) = 1 - \lim_{N\to\infty} \mathbf{P}\{B_N\}.$$

С другой стороны, из тождества $\overline{\cup_{n\geqslant 1} C_n} = \cap_{n\geqslant 1} \overline{C_n}$ следует, что

$$\mathbf{P}\left\{\cup_{n=1}^{\infty} A_n\right\} = \mathbf{P}\left\{\overline{\overline{\left(\cup_{n=1}^{\infty} A_n\right)}}\right\} =$$
$$= \mathbf{P}\left\{\overline{\cap_{n=1}^{\infty} \overline{A_n}}\right\} = \mathbf{P}\left\{\overline{\cap_{n=1}^{\infty} B_n}\right\} = 1 - \mathbf{P}\{\cap_{n=1}^{\infty} B_n\}.$$

Из этих двух выражений для $\mathbf{P}\left\{\cup_{n=1}^{\infty} A_n\right\}$ вытекает первое утверждение п.2. Второе утверждение следует из первого и из свойства 1). Теорема доказана.

§ 3. Простейшие вероятностные модели

Пусть в тройке $(\Omega, \mathcal{F}, \mathbf{P})$ пространство элементарных событий $\Omega = \{\omega\}$ — конечное множество, состоящее из $|\Omega|$ элементов, σ-алгебра (в данном случае — алгебра) $\mathcal{F}$ — совокупность всех подмножеств Ω.

При *классическом определении вероятности* $\mathbf{P}(\omega) = \frac{1}{|\Omega|}$ для всех $\omega \in \Omega$ и

$$\mathbf{P}(A) = \frac{|A|}{|\Omega|} \text{ для любого } A \subseteq \Omega.$$

Пример 1: выборка с возвращением. Пусть в урне содержится N шаров, занумерованных числами 1, 2, ..., N, причем шары с номерами от 1 до M белые, а остальные $N - M$ шаров – черные. Из урны последовательно извлекают t шаров, после извлечения каждого шара отмечают его номер и возвращают шар обратно в урну. Возможные исходы t извлечений шаров (элементарные события) можно отождествить с конечными последовательностями $(k_1, k_2, \ldots, k_t)$ чисел из множества $\{1, \ldots, N\}$. Поэтому в качестве пространства элементарных событий в данном случае можно выбрать множество $\Omega = \{(k_1, \ldots, k_t) :\ k_1, \ldots, k_t \in \{1, \ldots, N\}\} = \{1, \ldots, N\}^t$. Количество наборов $(k_1, \ldots, k_t)$, образующих такое пространство элементарных событий Ω, равно N^t. При классическом определении вероятности каждому набору $(k_1, \ldots, k_t) \in \Omega$ сопоставляется вероятность $\frac{1}{N^t}$.

Найдем вероятность события

$$A_s = \{\text{при выборе с возвращением } t \text{ шаров появилось } s \text{ белых шаров}\}.$$

Количество наборов чисел $(k_1,\, k_2,\, \ldots,\, k_t) \in \{1,\, \ldots,\, N\}^t$, содержащих ровно s чисел из множества $\{1,\, \ldots,\, M\}$ и $t - s$ чисел из множества $\{M + 1,\, \ldots,\, N\}$, равно $C_t^s M^s (N - M)^{t-s}$, поэтому

$$\mathbf{P}\,(A_s) = C_t^s \left(\frac{M}{N}\right)^s \left(1 - \frac{M}{N}\right)^{t-s}.$$

Важный частный случай t-кратной выборки с возвращением при $N = 2$, $M = 1$ часто интерпретируют как t-кратное бросание *идеальной симметричной монеты*; при этом исходы бросаний называют «гербом» и «решеткой». Другая интерпретация этого же случая состоит в том, что белый шар отождествляется с 0, а черный — с 1; тогда выборка соответствует *равновероятной случайной двоичной последовательности* и событие $B_s = \{$равновероятная случайная двоичная последовательность длины t содержит ровно s единиц$\}$ имеет вероятность

$$\mathbf{P}\{B_s\} = \tfrac{1}{2^t}\, C_t^s.$$

Стоит обратить внимание на то, что все 2^t реализаций равновероятной двоичной последовательности длины t имеют одну и ту же вероятность $\frac{1}{2^t}$. Поэтому с точки зрения теории вероятностей все реализации двоичных равновероятных последовательностей равноправны: как регулярные, например, $(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)$ так и «хаотичные», например, $(1,0,0,1,0,1,1,1,1,1,1,1,1,0,0,1,1,1,1,0)$ (k-й член второй последовательности равен $[k!\,\pi] \pmod 2$). В равновероятной случайной двоичной последовательности знаки 0 и 1 появляются с одной и той же вероятностью.

Пример 2: выборка без возвращения. Пусть из той же урны опять извлекают t шаров, но теперь извлеченные шары в урну не возвращаются (при этом должно быть $t \leqslant N$ и $s \leqslant M$). Теперь в качестве элементарных событий следует рассматривать конечные последовательности попарно различных чисел $(k_1,\, k_2,\, \ldots,\, k_t) \in \{1, \ldots, N\}^t$, а в качестве пространства элементарных событий можно выбрать $\Omega = \{(k_1, \ldots, k_t) : k_1, \ldots, k_t \in \{1, \ldots, N\}, k_i \neq k_j, 1 \leqslant i < j \leqslant t\}$. Число элементов этого пространства элементарных событий Ω вычисляется по другой формуле:

$$|\Omega| = N\,(N-1)\ldots(N-t+1) \stackrel{\mathrm{def}}{=} N^{[t]}.$$

Вероятность каждого элементарного события в схеме выбора t шаров без возвращения равна $\frac{1}{N^{[t]}}$. Число элементарных событий, составляющих аналогичное примеру 1 событие

$$A_s^* = \{\text{при выборе } t \text{ шаров без возвращения появилось } s \text{ белых шаров}\}.$$

равно

$$C_t^s M^{[s]}(N-M)^{[t-s]}.$$

Поэтому

$$\mathbf{P}\,(A_s^*) = C_t^s\,\frac{M^{[s]}(N-M)^{[t-s]}}{N^{[t]}}\,.$$

Например, при $N = 100$, $M = 30$, $t = 25$, $s = 7$ в схеме выбора с возвращением получаем $\mathbf{P}\,(A_7) \approx 0,1712$, а в схеме выбора без возвращения вероятность такого же события больше: $\mathbf{P}\,(A_7^*) \approx 0,1947$. Соотношение между вероятностями противоположных к A_7 и A_7^* событий обратное: $\mathbf{P}\left(\overline{A_7}\right) \approx 0,8288$, $\mathbf{P}\left(\overline{A_7^*}\right) \approx 0,8053$.

Выборку без возвращения объема 7 можно использовать как вероятностную модель совокупности номеров 7 шаров, выкатывающихся из лототрона при розыгрыше одного тиража лотереи, а выборку с возвращением — как вероятностную модель совокупности номеров первых шаров в 7 разных тиражах лотереи.

Пример 3: случайные равновероятные подстановки. В качестве пространства элементарных событий Ω при построении модели случайных подстановок можно рассматривать множество S_n всех $n!$ подстановок на множестве $\{1, 2, \ldots, n\}$:

$$\sigma = \begin{pmatrix} 1 & 2 & \ldots & n \\ \sigma_1 & \sigma_2 & \ldots & \sigma_n \end{pmatrix}.$$

Нижняя строка подстановки является *перестановкой* элементов множества $\{1, 2, \ldots, n\}$. В классической вероятностной схеме каждой подстановке сопоставляется вероятность, равная $\frac{1}{n!}$. Такие случайные подстановки (как и соответствующие им перестановки) называются *равновероятными.* Случайную равновероятную перестановку можно интерпретировать как последовательность номеров шаров, извлекаемых по схеме выбора без возвращения из урны с шарами, занумерованными числами от 1 до n, до исчерпания урны.

Пусть $\sigma \in S_n$ — случайная равновероятная подстановка. Найдем вероятность события

$$D = \{\sigma \in S_n \text{ не имеет неподвижных точек}\} = \{\sigma_m \neq m \ \forall\, m \in \{1, \ldots, n\}\}.$$

Так как $D = \cap_{m=1}^{n} \{\sigma_m \neq m\}$, то $\overline{D} = \cup_{m=1}^{n} \{\sigma_m = m\}$, и $\mathbf{P}\{D\} = 1 - \mathbf{P}\{\overline{D}\}$. Вероятность объединения событий можно вычислить с помощью формулы включения-исключения:

$$\mathbf{P}\{\overline{D}\} = \mathbf{P}\left\{\bigcup_{m=1}^{n} \{\sigma_m = m\}\right\} = \sum_{m=1}^{n} (-1)^{m-1} \sum_{1 \leqslant i_1 < \ldots < i_m \leqslant n} \mathbf{P}\{\sigma_{i_j} = i_j \ (j = 1, \ldots, m)\}.$$

Число слагаемых во внутренней сумме равно C_n^m, а каждое слагаемое равно отношению числа подстановок, которые оставляют на месте заданные элементы $i_1, \ldots, i_m$ (число таких подстановок равно $(n-m)!$), к общему числу подстановок $n!$. Так как $C_n^m \cdot \frac{(n-m)!}{n!} = \frac{1}{m!}$, то

$$\mathbf{P}\{\overline{D}\} = \sum_{m=1}^{n} (-1)^{m-1} \frac{1}{m!}$$

и

$$\mathbf{P}\{\sigma \in S_n \text{ не имеет неподвижных точек}\} = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \ldots + (-1)^n \frac{1}{n!}$$

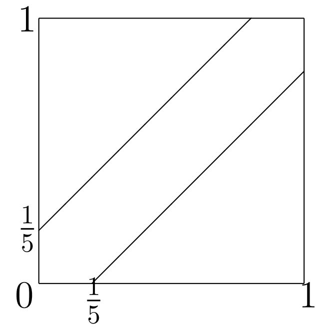
быстро стремится к $\frac{1}{e}$ при $n \to \infty$.

Пример 4. Геометрические вероятности.

Геометрическими вероятностями называют вероятности, связанные со свойствами случайных геометрических объектов. Пространством элементарных событий Ω в задачах о геометрических вероятностях обычно является область евклидова пространства R^n, $n \geqslant 1$, множеством событий — борелевская σ-алгебра подмножеств R^n с мерой Лебега; мы будем обозначать меру Лебега множества A символом $|A|$. В задачах о геометрических вероятностях, если не оговаривается противное, вероятность события $A \in \mathcal{F}$ полагается равной отношению n-мерных объемов множества A и области Ω:

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

Пример 4: случайная точка в квадрате. Допустим, что точка $A = (\xi, \eta)$ выбирается случайно в единичном квадрате $[0, 1] \times [0, 1]$. Чему равна вероятность того, что $|\xi - \eta| < \frac{1}{5}$?



В качестве пространства элементарных событий в этом случае естественно выбрать единичный квадрат: $\Omega = [0, 1] \times [0, 1]$, в качестве σ-алгебры $\mathcal{F}$ — совокупность его борелевских подмножеств, в качестве вероятностной меры $\mathbf{P}$ — меру Лебега на $\Omega = [0, 1] \times [0, 1]$. Событие $B = \{|\xi - \eta| < \frac{1}{5}\} = \bigcup\limits_{t \in (-1/5, +1/5)} \{\xi = \eta + t\}$ соответствует полосе S, ограниченной отрезками, которые соединяют точку $(0, \frac{1}{5})$ с точкой $(\frac{4}{5}, 1)$ и точку $(\frac{1}{5}, 0)$ с точкой $(1, \frac{4}{5})$. Дополнение к этой области до квадрата $[0, 1]^2$ состоит из двух равнобедренных прямоугольных треугольников с длинами катетов, равными $\frac{4}{5}$. Сумма их площадей равна $\left(\frac{4}{5}\right)^2 = \frac{16}{25}$, значит, площадь полосы S равна $|S| = 1 - \frac{16}{25} = \frac{9}{25}$. Так как площадь квадрата равна 1, то $\mathbf{P}\{B\} = \frac{|S|}{|\Omega|} = \frac{9}{25}$.

Этот пример можно использовать как вероятностную модель следующей ситуации. Два разведчика договорились встретиться возле газетного киоска в 10-минутном промежутке между 12.00 и 12.10; чтобы не обращать на себя внимание, каждый ждет другого около киоска ровно 2 минуты. Пара (t_1, t_2) моментов прихода разведчиков имеет равномерное распределение в множестве $[12.0, 12.1] \times [12.0, 12.1]$. Встреча происходит, если $|t_1 - t_2| \leqslant 0.02$.

§ 4. Независимость событий

Термин «независимость» отражает эмпирически наблюдаемое свойство *причинно независимых* событий, т.е. событий, порождаемых не связанными друг с другом причинами. Примером пары причинно независимых событий является четность номера вагона метро, в котором студент вчера возвращался домой, и четность числа изюминок в булочке, купленной им сегодня в буфете.

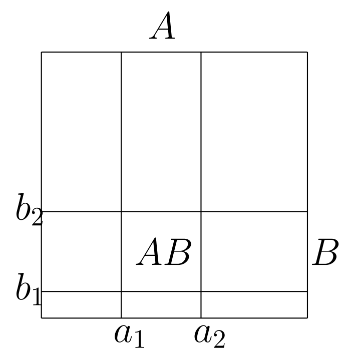
Эксперименты показывают, что если A и B — два причинно независимых события и N_{AB}, N_A и N_B — числа появлений событий $A \cap B$, A и B соответственно в N разных опытах, проводившихся в одинаковых условиях, то при достаточно больших N выполняется приближенное равенство

$$\frac{N_{AB}}{N} \approx \frac{N_A}{N} \cdot \frac{N_B}{N}.$$

В соответствии с законом устойчивости частот величины N_{AB}/N, N_A/N и N_B/N при увеличении N стабилизируются около предельных значений, которые и считают вероятностями событий AB, A и B. Этот эффект отражается в формальном определении теоретико-вероятностной независимости.

Определение. Пусть $(\Omega, \mathcal{F}, \mathbf{P})$ – вероятностное пространство. События $A, B \in \mathcal{F}$ называются *независимыми*, если

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



Пример. Пусть пространство элементарных событий $\Omega = \{(x, y)\colon x, y \in [0, 1]\}$ — единичный квадрат, $\mathcal{F}$ — σ-алгебра измеримых по Лебегу подмножеств Ω, $\mathbf{P}$ — мера Лебега на Ω (обобщающая геометрическое понятие площади). Если

$$A = \{(x,y) \in \Omega\colon a_1 \leqslant x \leqslant a_2\},\ B = \{(x,y) \in \Omega\colon b_1 \leqslant y \leqslant b_2\},$$

то $AB = \{(x,y) \in \Omega\colon a_1 \leqslant x \leqslant a_2,\ b_1 \leqslant y \leqslant b_2\}$, и

$$\mathbf{P}\{A\} = \frac{a_2 - a_1}{|\Omega|} = a_2 - a_1,\ \mathbf{P}\{B\} = \frac{b_2 - b_1}{|\Omega|} = b_2 - b_1,$$

$$\mathbf{P}\{AB\} = \frac{(a_2 - a_1)(b_2 - b_1)}{|\Omega|} = (a_2 - a_1)(b_2 - b_1) = \mathbf{P}\{A\}\mathbf{P}\{B\},$$

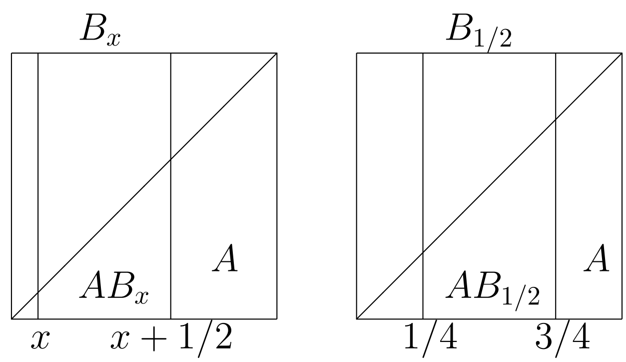
т. е. события A и B независимы.

Математическое определение независимости событий не связано с причинной независимостью: оно просто выделяет класс пар событий, для которых вероятность их одновременного появления равна произведению вероятностей отдельных событий. События могут явно порождаться одной «причиной» и быть независимыми. Например, указанные выше события A и B с точки зрения теории вероятностей независимы, хотя порождаются общим для них выбором случайной точки квадрата.

Приведем еще один пример, показывающий, что в теории вероятностей независимость — это формальное свойство. На том же единичном квадрате рассмотрим множества

$$A = \{(x, y)\colon 0 \leqslant y \leqslant x \leqslant 1\},$$
$$B_t = \{(x, y)\colon t \leqslant x \leqslant x + \frac{1}{2}, 0 \leqslant y \leqslant 1\}, \quad 0 \leqslant t \leqslant \frac{1}{2}.$$

Тогда $\mathbf{P}\{A\} = \mathbf{P}\{B_t\} = \frac{1}{2}$, $\mathbf{P}\{A \cap B_t\} = \frac{1}{8} + \frac{t}{2}$ растет линейно по t от $\frac{1}{8}$ при $t = 0$ до $\frac{3}{8}$ при $t = \frac{1}{2}$, и $\mathbf{P}\{AB_t\} = \mathbf{P}\{A\}\mathbf{P}\{B_t\} = \frac{1}{4} \Leftrightarrow t = \frac{1}{4}$.



Простейшие свойства независимых событий.

а) Если $A \in \mathcal{F}$ – событие, то события A и $\varnothing$, а также A и Ω независимы:

$$\mathbf{P}\left(A\varnothing\right) = 0 = \mathbf{P}\left(A\right) \cdot 0 = \mathbf{P}\left(A\right)\mathbf{P}\left(\varnothing\right),$$
$$\mathbf{P}\left(A\Omega\right) = \mathbf{P}\left(A\right) \cdot 1 = \mathbf{P}\left(A\right)\mathbf{P}\left(\Omega\right).$$

б) Если независимы события A и B, то независимы также события A и $\overline{B}$, $\overline{A}$ и B, $\overline{A}$ и $\overline{B}$:

$$\mathbf{P}\left(A\overline{B}\right) = \mathbf{P}\left(A \setminus AB\right) = \mathbf{P}\left(A\right) - \mathbf{P}\left(AB\right) =$$
$$= \mathbf{P}\left(A\right)\left(1 - \mathbf{P}\left(B\right)\right) = \mathbf{P}\left(A\right)\mathbf{P}\left(\overline{B}\right),$$

аналогично, $\mathbf{P}\left(\overline{A}B\right) = \mathbf{P}\left(\overline{A}\right)\mathbf{P}(B)$; наконец,

$$\mathbf{P}\left(\overline{A}\overline{B}\right) = \mathbf{P}\left(\overline{A} \setminus \overline{A}B\right) = \mathbf{P}\left(\overline{A}\right) - \mathbf{P}\left(\overline{A}B\right) =$$
$$= \mathbf{P}\left(\overline{A}\right)\left(1 - \mathbf{P}\left(B\right)\right) = \mathbf{P}\left(\overline{A}\right)\mathbf{P}\left(\overline{B}\right).$$

Понятие независимости можно разными способами обобщать на совокупности, состоящие более чем из двух событий.

Определение. Пусть на одном и том же вероятностном пространстве определены события $A_1,\ A_2,\ \ldots$ Говорят, что $A_1,\ A_2, \ldots$ – совокупность *попарно независимых событий*, если

$$\mathbf{P}\left(A_i A_j\right) = \mathbf{P}\left(A_i\right)\mathbf{P}\left(A_j\right) \text{ для любых } i \neq j.$$

События $A_1,\ A_2, \ldots$ называют *независимыми в совокупности* (или *совместно независимыми*), если для любого натурального m и для любых $i_1,\ \ldots, i_m,\ 1 \leqslant i_1 < \ldots < i_m$ выполняется равенство

$$\mathbf{P}\{A_{i_1} A_{i_2} \ldots A_{i_m}\} = \mathbf{P}\{A_{i_1}\}\mathbf{P}\{A_{i_2}\} \ldots \mathbf{P}\{A_{i_m}\}.$$

Из определения следует, что если события $A_1,\ \ldots,\ A_n$ независимы в совокупности и числа $i_1,\ \ldots, i_r,\ j_1,\ \ldots, j_s \in \{1, \ldots, n\}$ попарно различны, то события $A_{i_1} \ldots A_{i_r}$ и $A_{j_1} \ldots A_{j_s}$ независимы.

Обычно термин «независимость» понимают как независимость в совокупности. Если его понимают иначе, то это оговаривается особо.

§ 5. Независимость разбиений, алгебр и σ-алгебр

Определение. Пусть $\gamma = \{\gamma_1,\ \gamma_2,\ \ldots\}$, $\gamma_k \subseteq \Omega$, $k = 1, 2, \ldots,$ — совокупность подмножеств Ω. Наименьшая алгебра (σ-алгебра) множеств $\mathcal{F}(\gamma)$, содержащая γ, называется *алгеброй (σ-алгеброй), порожденной* γ.

Определение. Пусть $(\Omega,\ \mathcal{F},\ \mathbf{P})$ — вероятностное пространство. Система событий $A_1,\ A_2,\ \ldots \in \mathcal{F}$ называется *разбиением* пространства Ω, если эти события попарно несовместны, а их объединение совпадает с Ω:

$$A_i \cap A_j = \varnothing \ \ (i \neq j), \quad \bigcup_{n \geqslant 1} A_n = \Omega.$$

Разбиения называют *полными системами несовместных событий.*

Замечание. Если $\gamma = \{\gamma_1,\ \gamma_2,\ \ldots,\ \gamma_n\}$ — конечное разбиение Ω, то алгебра $\mathcal{F}(\gamma)$, порожденная разбиением γ, является конечной и состоит из пустого множества $\varnothing$ и всех множеств вида

$$\gamma_{i_1} \cup \gamma_{i_2} \cup \ldots \cup \gamma_{i_k}, \ \ k = 1, \ldots, n, \ \ 1 \leqslant i_1 < \cdots < i_k \leqslant n.$$

Из этого замечания следует, что *если в разбиении* $\gamma = \{\gamma_1,\ \gamma_2,\ \ldots,\ \gamma_n\}$ *все n множеств не пустые, то $\mathcal{F}(\gamma)$ состоит из 2^n элементов.* Если разбиение счетное, то порожденная им алгебра состоит из всех конечных объединений множеств, входящих в разбиение, и дополнений к таким объединениям; значит, множество ее элементов тоже счетное. Элементам σ-алгебры, порожденной счетным разбиением, можно взаимно однозначно сопоставить бесконечные последовательности из нулей и единиц (единицы соответствуют номерам элементов разбиения, входящим в объединение), поэтому такая σ-алгебра имеет мощность континуума. Таким образом, счетные алгебры событий существуют, а счетные σ-алгебры не существуют.

Определение. Пусть $(\Omega,\, \mathcal{F},\, \mathbf{P})$ — вероятностное пространство. Разбиения

$$\{A_{11}, A_{12}, \ldots\},\ \{A_{21}, A_{22}, \ldots\}, \ldots, \{A_{m1}, A_{m2}, \ldots\}$$

пространства Ω на множества $A_{ij} \in \mathcal{F}$ называются *независимыми*, если

$$\mathbf{P}\left(A_{1i_1} \cap A_{2i_1} \cap \ldots \cap A_{mi_m}\right) = \mathbf{P}\left(A_{1i_1}\right) \mathbf{P}\left(A_{2i_2}\right) \ldots \mathbf{P}\left(A_{mi_m}\right)$$

для любых допустимых значений $i_1, \ldots, i_m \geqslant 1$. Алгебры или σ-алгебры событий $\mathcal{F}_1,\, \ldots,\, \mathcal{F}_m \subseteq \mathcal{F}$ называются *независимыми*, если

$$\mathbf{P}\left(A_1 \cap \ldots \cap A_m\right) = \mathbf{P}\left(A_1\right) \ldots \mathbf{P}\left(A_m\right)$$

для любых событий $A_k \in \mathcal{F}_k,\ \ k = 1, \ldots, m.$

Теорема. *Если* $(\Omega,\, \mathcal{F},\, \mathbf{P})$ *— вероятностное пространство и σ-алгебры $\mathcal{F}_1,\, \ldots,\, \mathcal{F}_m \subseteq \mathcal{F}$ порождаются конечными или счетными разбиениями Ω, то σ-алгебры $\mathcal{F}_1,\, \ldots,\, \mathcal{F}_m$ независимы тогда и только тогда, когда независимы порождающие их разбиения.*

Доказательство. Если σ-алгебры $\mathcal{F}_1,\, \ldots,\, \mathcal{F}_m$ независимы, то порождающие их разбиения тоже независимы, поскольку элементы разбиений принадлежат алгебрам.

Пусть теперь независимы разбиения

$$\{A_{11}, A_{12}, \ldots\},\ \{A_{21}, A_{22}, \ldots\}, \ldots, \{A_{m1}, A_{m2}, \ldots\}$$

порождающие σ-алгебры $\mathcal{F}_1, \ldots, \mathcal{F}_m$, и $B_1 \in \mathcal{F}_1, \ldots, B_m \in \mathcal{F}_m$. Покажем, что $\mathbf{P}\left(B_1 B_2 \ldots B_m\right) = \mathbf{P}\left(B_1\right) \mathbf{P}\left(B_2\right) \ldots \mathbf{P}\left(B_m\right)$. Так как σ-алгебра $\mathcal{F}_k$ порождается разбиением $\{A_{k1}, A_{k2}, \ldots\}$, то каждому событию $B_k \in \mathcal{F}_k$ соответствует такое множество натуральных чисел $\{i_k(1), i_k(2), \ldots\}$, что

$$B_k = \bigsqcup_{v \geqslant 1} A_{k i_k(v)}, \quad k = 1, \ldots, m.$$

Тогда

$$B_1 B_2 \ldots B_m = \bigcap_{k=1}^{m} \left(\bigsqcup_{v \geqslant 1} A_{k i_k(v)} \right) = \bigsqcup_{v_1 \geqslant 1} \ldots \bigsqcup_{v_m \geqslant 1} A_{1 i_1(v_1)} \ldots A_{m i_m(v_m)},$$

поскольку пересечения элементов разбиений в правой части попарно не пересекаются: если $(v_1, \ldots, v_m) \neq (u_1, \ldots, u_m)$, то существует $k \in \{1, \ldots, m\}$, при котором $v_k \neq u_k$; тогда $A_{k i_k(v_k)} \bigcap A_{k i_k(u_k)} = \varnothing$, $A_{1 i_1(v_1)} \ldots A_{m i_m(v_m)} \subset A_{k i_k(v_k)}$, $A_{1 i_1(u_1)} \ldots A_{m i_m(u_m)} \subset A_{k i_k(u_k)}$ и

$$(A_{1 i_1(v_1)} \ldots A_{m i_m(v_m)}) \cap (A_{1 i_1(u_1)} \ldots A_{m i_m(u_m)}) \subseteq \left(A_{k i_k(v_k)} \cap A_{k i_k(u_k)} \right) = \varnothing.$$

Так как вероятность объединения несовместных событий равна сумме вероятностей этих событий, а разбиения $\mathcal{F}_1, \ldots, \mathcal{F}_m$ независимы, то

$$\begin{aligned} \mathbf{P}\{B_1 B_2 \ldots B_m\} &= \sum_{v_1 \geqslant 1} \ldots \sum_{v_m \geqslant 1} \mathbf{P}\{A_{1 i_1(v_1)} \ldots A_{m i_m(v_m)}\} = \\ &= \sum_{v_1 \geqslant 1} \ldots \sum_{v_m \geqslant 1} \mathbf{P}\{A_{1 i_1(v_1)}\} \ldots \mathbf{P}\{A_{m i_m(v_m)}\} = \\ &= \prod_{k=1}^{m} \left(\sum_{v \geqslant 1} \mathbf{P}\{A_{k i_k(v)}\} \right) = \prod_{k=1}^{m} \mathbf{P}\{B_k\}. \end{aligned}$$

Теорема доказана.

§ 6. Независимые испытания. Произведения вероятностных пространств

Допустим, что нужно построить вероятностное пространство, соответствующее совокупности n независимых испытаний (не обязательно имеющих одинаковые вероятности исходов: например, испытания с нечетными номерами — это бросания монеты, а с четными номерами — бросания кубика). Пусть k-му ($k = 1, \ldots, n$) испытанию соответствует вероятностное пространство $(\Omega_k,\, \mathcal{F}_k,\, \mathbf{P}_k)$ (например, $\Omega_{2k-1} = \{0, 1\}, \Omega_{2k} = \{1, 2, 3, 4, 5, 6\}$, $k = 1, 2, \ldots$). Стандартный способ состоит в использовании *произведения вероятностных пространств*, т.е. такого вероятностного пространства $(\Omega,\, \mathcal{F}, \mathbf{P})$, в котором

$$\Omega = \Omega_1 \times \ldots \times \Omega_n = \{\omega = (\omega_1, \ldots, \omega_n) : \; \omega_k \in \Omega_k, \quad k = 1, \ldots, n\},$$

σ-алгебра $\mathcal{F} = \mathcal{F}_1 \times \ldots \times \mathcal{F}_n$ — минимальная σ-алгебра, содержащая все «цилиндрические»[1] события вида $A'_k = \Omega_1 \times \ldots \times \Omega_{k-1} \times A_k \times \Omega_{k+1} \times \ldots \times \Omega_n$, точнее:

$$A'_k = \{\omega = (\omega_1, \ldots, \omega_n) \in \Omega\colon \omega_k \in A_k \in \mathcal{F}_k,\ \omega_j \in \Omega_j,\ j \in \{1, \ldots, n\} \backslash \{k\}\},\qquad k = 1, \ldots, n, \tag{7}$$

а вероятностная мера $\mathbf{P} = \mathbf{P}_1 \times \ldots \times \mathbf{P}_n$ на $\mathcal{F} = \mathcal{F}_1 \times \ldots \times \mathcal{F}_n$ для множеств вида $A_1 \times \ldots \times A_n = \bigcap_{k=1}^n A'_k$ определяется равенствами

$$\mathbf{P}\left\{\bigcap_{k=1}^n A'_k\right\} = \mathbf{P}\{A_1 \times \ldots \times A_n\} \stackrel{\text{def}}{=} \prod_{k=1}^n \mathbf{P}_k\{A_k\},\ A_k \in \mathcal{F}_k,\ k = 1, \ldots, n; \tag{8}$$

по аддитивности эта мера продолжается на алгебру событий, порожденную событиями A'_k, а затем на порожденную ими же σ-алгебру (обоснованием такого построения является теорема Каратеодори).

Таким образом, для любого $k = 1, 2, \ldots, n$ каждому событию $A_k \in \mathcal{F}_k$ сопоставлено «цилиндрическое» событие в пространстве $(\Omega,\, \mathcal{F},\, \mathbf{P})$

$$A'_k = \{\omega = (\omega_1, \ldots,\, \omega_n) : \omega_k \in A_k\} = \\ = \Omega_1 \times \ldots \times \Omega_{k-1} \times A_k \times \Omega_{k+1} \times \ldots \times \Omega_n.$$

Тогда, пользуясь (8), находим: $\mathbf{P}\,(A'_k) = \mathbf{P}\,(A_k)$, и σ-алгебре $\mathcal{F}_k$ будет взаимно однозначно соответствовать σ-алгебра

$$\mathcal{F}'_k = \{A'_k = \Omega_1 \times \ldots \times \Omega_{k-1} \times A_k \times \Omega_{k+1} \times \ldots \times \Omega_n :\ A_k \in \mathcal{F}_k\}$$

подмножеств пространства Ω. В силу определения (8) любые события $A'_1, \ldots, A'_n$ независимы:

$$\mathbf{P}\{A'_1 \cap \ldots \cap A'_n\} = \mathbf{P}\{A'_1\} \ldots \mathbf{P}\{A'_n\} = \mathbf{P}\{A'_1\} \ldots \mathbf{P}\{A'_n\},$$

т.е. описанная конструкция действительно позволяет построить на одном вероятностном пространстве *независимые* события $A'_1,\, \ldots,\, A'_n$, имеющие соответственно такие же вероятности, что и события $A_1,\, \ldots,\, A_n$.

Примеры. а) *Схемой Бернулли* называют последовательность независимых испытаний с двумя исходами, которые обычно обозначают 1

[1] Термин подразумевает аналогию с цилиндром в трехмерном пространстве $\{(x, y, z)\colon x^2 + y^2 \leqslant 1, z \in \mathbb{R}\} = \{(x, y)\colon x^2 + y^2 \leqslant 1\} \times \mathbb{R}$

и 0 (или «успех» и «неудача», или «герб» и «решетка»). Испытанию с номером k соответствует вероятностное пространство $(\Omega_k,\, \mathcal{F}_k,\, \mathbf{P}_k)$:

$$\Omega_k = \{0,\, 1\}, \quad \mathcal{F}_k = \{\varnothing,\, \{0\},\, \{1\},\, \Omega_k\},$$
$$\mathbf{P}_k\{1\} = p = p^1 q^0, \quad \mathbf{P}_k\{0\} = 1 - p = q = p^0 q^1.$$

Произведение $(\Omega, \mathcal{F}, \mathbf{P})$ вероятностных пространств $(\Omega_1,\, \mathcal{F}_1,\, \mathbf{P}_1), \ldots, (\Omega_n,\, \mathcal{F}_n,\, \mathbf{P}_n)$ соответствует совокупностям исходов n независимых испытаний:

$$\Omega = \Omega_1 \times \ldots \times \Omega_n = \{\omega = (\omega_1, \ldots, \omega_n) : \ \omega_k \in \Omega_k, \ \ k = 1, \ldots, n\}\,,$$

$\mathcal{F}$ — множество всех подмножеств Ω, и для $\omega = (\omega_1,\, \ldots,\, \omega_n) \in \Omega$

$$\mathbf{P}\{\omega\} = \prod_{k=1}^{n} p^{\omega_k} q^{1-\omega_k} = p^{w_1+\ldots+w_n} q^{n-(w_1+\ldots+w_n)}. \tag{9}$$

Вероятность $\mathbf{P}\{\omega\}$ зависит только от суммы $\nu = \nu(\omega) = \omega_1 + \ldots + \omega_n$, равной числу единиц (числу успехов) в n испытаниях и не зависит от того, в каких именно испытаниях эти успехи появились. Поэтому легко найти вероятность того, что среди n испытаний ровно k закончатся успехом:

$$\mathbf{P}\{\nu = k\} = \sum_{\omega\in\Omega:\, \nu(\omega)=k} \mathbf{P}(\omega) = C_n^k p^k q^{n-k}.$$

Более общей моделью независимых испытаний с двумя исходами является *неоднородная схема Бернулли*, когда вероятность успеха в k-м испытании зависит от k: $\mathbf{P}_k(\{1\}) = p_k$, $\mathbf{P}_k(\{0\}) = q_k = 1 - p_k$. В этом случае формула (9) заменяется соотношением

$$\mathbf{P}\,(\omega) = \prod_{j=1}^{n} p_j^{w_j} q_j^{1-w_j} \quad \text{и} \quad \mathbf{P}\{\nu = k\} = \sum_{(\omega_1,\ldots,\omega_n):\ \sum_{1\leqslant j\leqslant n}\omega_j = k} \ \prod_{j=1}^{n} p_j^{\omega_j} q_j^{1-\omega_j}.$$

б) *Полиномиальная схема* отличается от схемы Бернулли тем, что число исходов в каждом испытании равно натуральному числу $N \geq 2$. Испытанию с номером k соответствует вероятностное пространство $(\Omega_k,\, \mathcal{F}_k,\, \mathbf{P}_k)$, где $\Omega_k = \{1,\, \ldots,\, N\}$, $\mathcal{F}_k$ состоит из всех 2^N подмножеств Ω_k,

$$\mathbf{P}_k\{j\} = p_j \geqslant 0, \quad j \in \Omega_k = \{1, \ldots, N\}, \quad \sum_{j=1}^{N} p_j = 1.$$

Произведение n таких пространств соответствует n независимым испытаниям с N исходами:

$$\Omega = \Omega_1 \times \ldots \times \Omega_n =$$
$$= \{\omega = (\omega_1, \ldots, \omega_n) : \ \omega_k \in \Omega = \{1, \ldots, N\}, \ \ k = 1, \ldots, n\},$$

$\mathcal{F}$ – множество всех подмножеств Ω. Чтобы задать вероятностную меру на произведении пространств в виде, аналогичном схеме Бернулли, используем символ Кронекера $\delta(i, j)$, т.е. $\delta(i, j) = 1$ при $i = j$ и $\delta(i, j) = 0$ при $i \neq j$. Если обозначить через

$$\nu_j(\omega) = \delta(j, \omega_1) + \delta(j, \omega_2) + \ \ldots \ + \delta(j, \omega_n)$$

число появлений исхода j в n испытаниях, то для любого $\omega = (\omega_1, \ \ldots, \ \omega_n) \in \Omega$

$$\mathbf{P}\{\omega\} = \prod_{k=1}^{n} \mathbf{P}\{\omega_k\} = \prod_{k=1}^{n} p_{\omega_k} = \prod_{j=1}^{N} p_j^{\nu_j(\omega)}.$$

Как и в случае схемы Бернулли, $\mathbf{P}\{\omega\}$ зависит только от чисел $\nu_j(\omega)$ появлений исходов в n испытаниях и не зависит от порядка их появления. Так как число наборов из n символов, принимающих значения $1, 2, \ \ldots, \ N$ и содержащих m_j раз символ j, $j = 1, \ \ldots, \ N$, равно полиномиальному коэффициенту $\frac{n!}{m_1! \ldots m_N!}$, то

$$\mathbf{P}\{\text{ исход } j \text{ в } n \text{ испытаниях появился } m_j \text{ раз}, \ j = 1, \ldots, N\} =$$
$$= \mathbf{P}\{\nu_j(\omega) = m_j, \ j = 1, \ldots, N\} =$$
$$= \sum_{\omega:\ \nu_j(\omega) = m_j,\, j = 1, \ldots, N} \mathbf{P}\{\omega\} = \frac{n!}{m_1! \ldots m_N!}\, p_1^{m_1} \ldots p_N^{m_N}.$$

Выше описан способ построения произведения конечного числа конечных вероятностных пространств, соответствующего проведению конечного числа независимых испытаний. Однако можно представить себе, что последовательность испытаний продолжается бесконечно или что количество испытаний заранее нельзя ограничить (например, если испытания Бернулли проводятся до первого успеха). Естественно возникает вопрос о том, как строить вероятностное пространство, соответствующее бесконечной последовательности независимых испытаний.

Опишем схему такого построения. Пусть, как и раньше, испытанию с номером $k = 1, 2, \ldots$ соответствует вероятностное пространство $(\Omega_k,\, \mathcal{F}_k,\, \mathbf{P}_k)$. В качестве пространства элементарных событий, соответствующего *бесконечной* последовательности независимых испытаний, можно выбрать прямое произведение бесконечной последовательности множеств Ω_k:

$$\Omega = \Omega_1 \times \Omega_2 \times \ldots = \{(\omega_1,\, \omega_2, \ldots) : \omega_k \in \Omega_k,\ k = 1, 2, \ldots\}.$$

Для каждого натурального n строится произведение вероятностных пространств $(\Omega^{(n)}, \mathcal{F}^{(n)}, \mathbf{P}^{(n)}) = (\Omega_1 \times \ldots \times \Omega_n,\, \mathcal{F}_1 \times \ldots \times \mathcal{F}_n,\, \mathbf{P}_1 \times \ldots \times \mathbf{P}_n)$. Такому произведению сопоставляется σ-алгебра подмножеств пространства Ω вида $B = B_n \times \Omega_{n+1} \times \Omega_{n+2} \times \ldots,\ B_n \in \mathcal{F}^{(n)}$, и по определению полагается $\mathbf{P}\{B\} = \mathbf{P}^{(n)}\{B_n\}$. После этого σ-алгебра $\mathcal{F}$ подмножеств Ω определяется как минимальная σ-алгебра, содержащая все σ-алгебры $\mathcal{F}^{(1)}, \mathcal{F}^{(2)}, \ldots$ Меры $\mathbf{P}^{(n)}$ продолжаются на σ-алгебру $\mathcal{F}$ с помощью теоремы Каратеодори.

§ 7. Условные вероятности относительно событий

Пусть $(\Omega,\, \mathcal{F},\, \mathbf{P})$ — вероятностное пространство и $B \in \mathcal{F}$ — событие, $\mathbf{P}\{B\} > 0$. Вероятностная мера $\mathbf{P}$ — это функция, определенная на σ-алгебре $\mathcal{F}$. Можно рассмотреть ограничение вероятностного пространства на множество B: $(\Omega_B, \mathcal{F}_B, \mathbf{P}^*_B)$, где $\Omega_B = B$,

$$\mathcal{F}_B = \{C \in \mathcal{F} \colon C \subseteq B\} = \{C \cap B \colon C \in \mathcal{F}\}$$

и $\mathbf{P}^*_B\{C\} = \mathbf{P}\{C\}$ для всех множеств $C \subset B$, $C \in \mathcal{F}$. Полученная мера $\mathbf{P}^*_B$ — неотрицательная, но при $\mathbf{P}\{B\} < 1$ не вероятностная, так как $\mathbf{P}^*_B\{\Omega_B\} = \mathbf{P}\{B\}$. От нее можно перейти к вероятностной мере $\mathbf{P}_B$, используя нормировку всех значений $\mathbf{P}^*_B$ числом $\mathbf{P}\{B\}$.

Определение. *Условными вероятностями событий* $A \in \mathcal{F}$ *при условии события* B, имеющего положительную вероятность: $\mathbf{P}(B) > 0$, называются величины

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

Условные вероятности, как и обычные (безусловные) счетно-аддитивны: если события $A_1, A_2, \ldots$ попарно несовместны, то

$$\mathbf{P}\left\{\bigcup_{k\geqslant 1} A_k \middle| B\right\} = \frac{\mathbf{P}\left\{\left(\bigcup_{k\geqslant 1} A_k\right) \cap B\right\}}{\mathbf{P}\{B\}} = \frac{\mathbf{P}\left\{\bigcup_{k\geqslant 1}(A_k \cap B)\right\}}{\mathbf{P}\{B\}} =$$
$$= \frac{\sum_{k\geqslant 1} \mathbf{P}\{A_k \cap B\}}{\mathbf{P}\{B\}} = \sum_{k\geqslant 1} \frac{\mathbf{P}\{A_k \cap B\}}{\mathbf{P}\{B\}} = \sum_{k\geqslant 1} \mathbf{P}\{A_k \,|\, B\}.$$

Для независимых событий A, B условные вероятности совпадают с безусловными: если $\mathbf{P}\{A\} > 0$, $\mathbf{P}\{B\} > 0$ и $\mathbf{P}\{AB\} = \mathbf{P}\{A\}\mathbf{P}\{B\}$, то

$$\mathbf{P}\{A \,|\, B\} = \mathbf{P}\{A \,|\, \overline{B}\} = \mathbf{P}\{A\},\ \mathbf{P}\{\overline{A} \,|\, B\} = \mathbf{P}\{\overline{A} \,|\, \overline{B}\} = \mathbf{P}\{\overline{A}\}. \tag{10}$$

Из определения следует формула

$$\mathbf{P}\,(AB) = \mathbf{P}\,(B)\,\mathbf{P}\,(A\,|B\,)\,, \tag{11}$$

которая является частным случаем *теоремы умножения.*

Теорема умножения. *Для любых таких событий* A_1, A_2, …, A_n, *что* $\mathbf{P}\,(A_1A_2\ldots A_{n-1}) > 0$, *справедливо равенство*

$$\mathbf{P}\{A_1A_2\ldots A_n\} = \tag{12}$$
$$= \mathbf{P}\{A_1\}\mathbf{P}\{A_2 \,|\, A_1\}\mathbf{P}\{A_3 \,|\, A_1A_2\}\ldots\mathbf{P}\{A_n \,|\, A_1A_2\ldots A_{n-1}\}.$$

Доказательство. Так как $A_1\ldots A_{n-1} \subseteq A_1\ldots A_k$ при любом $k \leqslant n-1$, то из условия $\mathbf{P}\{A_1A_2\ldots A_{n-1}\} > 0$ следует, что $\mathbf{P}\{A_1A_2\ldots A_k\} > 0$ при любом $k = 1, 2, \ldots, n-1$, т.е. что все условные вероятности в правой части (12) определены.

При $n = 2$ формула (12) совпадает с (11). Далее воспользуемся методом математической индукции. Допустим, что при некотором $n \geq 2$ формула (12) верна для любых n событий, и, используя (11), докажем, что тогда формула (12) верна и для любых $n+1$ событий:

$$\mathbf{P}\{A_1A_2\ldots A_{n+1}\} = \mathbf{P}\{(A_1A_2\ldots A_n)A_{n+1}\} =$$
$$= \mathbf{P}\{A_1A_2\ldots A_n\}\mathbf{P}\{A_{n+1} \,|\, A_1A_2\ldots A_n\} =$$
$$= \mathbf{P}\{A_1\}\mathbf{P}\{A_2 \,|\, A_1\}\ldots\mathbf{P}\{A_n \,|\, A_1A_2\ldots A_{n-1}\}\mathbf{P}\{A_{n+1} \,|\, A_1A_2\ldots A_n\}.$$

Теорема доказана.

Следствие. *Если события* $A_1, \ldots, A_n$ *независимы в совокупности, числа* $i_1, \ldots, i_r, j_1, \ldots, j_s \in \{1, \ldots, n\}$ *попарно различны и* $\mathbf{P}\left(A_{i_1} \ldots A_{i_r}\right) > 0$, *то*

$$\mathbf{P}\{A_{j_1}A_{j_2}\ldots A_{j_s} \,|\, A_{i_1}A_{i_2}\ldots A_{i_r}\} = \mathbf{P}\{A_{j_1}A_{j_2}\ldots A_{j_s}\} = \\ = \mathbf{P}\{A_{j_1}\}\mathbf{P}\{A_{j_2}\}\cdots\mathbf{P}\{A_{j_s}\}.$$

Доказательство следует из (10) и (11).

Теорема (Формула полной вероятности). *Если* $A_1, A_2, \ldots$ — *не более чем счетное разбиение* Ω *и* $\mathbf{P}\{A_k\} > 0$ *при всех* $k = 1, 2, \ldots$, *то для любого события* B *справедливо равенство*

$$\mathbf{P}\{B\} = \sum_{k\geqslant 1}\mathbf{P}\{A_kB\} = \sum_{k\geqslant 1}\mathbf{P}\{A_k\}\mathbf{P}\{B \,|\, A_k\}. \tag{13}$$

Доказательство. Так как $A_1, A_2, \ldots$ — разбиение Ω, то

$$B = B\bigcap\Omega = B\bigcap\left(\bigcup_{k\geqslant 1}A_k\right) = \bigcup_{k\geqslant 1}(BA_k) \text{ и } BA_k\bigcap BA_j = \varnothing \text{ при } k \neq j.$$

Отсюда, пользуясь аддитивностью вероятности и теоремой умножения, находим:

$$\mathbf{P}\{B\} = \sum_{k=1}^{n}\mathbf{P}\{BA_k\} = \sum_{k=1}^{n}\mathbf{P}\{A_k\}\mathbf{P}\{B \,|\, A_k\}.$$

Теорема доказана.

Теорема (Формула Байеса). *Если* $A_1, A_2, \ldots$ — *полная система несовместных событий и событие* B *таково, что* $\mathbf{P}(B) > 0$, *то*

$$\mathbf{P}\{A_k \,|\, B\} = \frac{\mathbf{P}\{A_k\}\mathbf{P}\{B \,|\, A_k\}}{\sum\limits_{j\geqslant 1}\mathbf{P}\{A_j\}\mathbf{P}\{B \,|\, A_j\}}.$$

Доказательство. Достаточно воспользоваться определением условной вероятности, теоремой умножения и формулой полной вероятности:

$$\mathbf{P}\{A_k \,|\, B\} = \frac{\mathbf{P}\{A_kB\}}{\mathbf{P}\{B\}} = \frac{\mathbf{P}\{A_k\}\mathbf{P}\{B \,|\, A_k\}}{\sum\limits_{j\geqslant 1}\mathbf{P}\{A_j\}\mathbf{P}\{B \,|\, A_j\}}.$$

Теорема доказана.

Формулу полной вероятности удобно использовать при решении вероятностных задач, если можно разбить пространство элементарных событий Ω_B на события $A_1, A_2, \ldots \in \mathcal{F}$ так, что условные вероятности на ограничениях вероятностного пространства $(\Omega, \mathcal{F}, \mathbf{P})$ на эти события: $(A_1, \mathcal{F}_{A_1}, \mathbf{P}_{A_1}), (A_2, \mathcal{F}_{A_2}, \mathbf{P}_{A_2}), \ldots$ легко находятся.

§ 8. Случайные величины

Одним из центральных понятий теории вероятностей является понятие случайной величины.

Определение. Если $(\Omega, \mathcal{F}, \mathbf{P})$ — конечное или счетное вероятностное пространство, а $\mathcal{F}$ — совокупность всех подмножеств Ω, то *случайной величиной* $\xi = \xi(\omega)$ называется любая функция, определенная на пространстве элементарных событий Ω.

Для произвольного вероятностного пространства $(\Omega, \mathcal{F}, \mathbf{P})$ *случайной величиной*, $\xi = \xi(\omega)$ принимающей значения в множестве S с σ-алгеброй подмножеств Σ, называется любая $(\mathcal{F}, \Sigma)$-измеримая функция $\xi\colon \Omega \to S$, т. е. такая функция, что для любого подмножества $A \subseteq S$, $A \in \Sigma$, выполняется условие

$$\xi^{-1}(A) = \{\omega \in \Omega\colon \xi(\omega) \in A\} \in \mathcal{F}.$$

Условие $(\mathcal{F}, \Sigma)$-измеримости случайной величины ξ как функции, отображающей Ω в S, позволяет сводить вычисление вероятностей событий, заданных условиями на значения случайной величины, к вычислению вероятностей прообразов этих событий — измеримых подмножеств пространства элементарных событий Ω, для которых определена вероятностная мера.

Это определение еще раз показывает, что математическая модель случайных явлений — детерминистическая: значение любой случайной величины однозначно определяется точкой пространства элементарных событий.

Термин «величина» не означает, что значениями случайной величины обязательно являются числа. Примерами случайных величин, принимающих значения из конечного множества, являются число успехов в n испытаниях Бернулли, число появлений заданного исхода в n испытаниях по полиномиальной схеме. В примере на геометрические вероятности случайными величинами являются координаты случайной точки в единичном квадрате или разность этих координат. Примерами нечисловых случайных величин являются случайная подстановка, многочлен со случанми коэффициентами, поворот трехмерного пространства на случайный угол вокруг случайно выбранной оси.

Если $\xi_1(\omega),\ \ldots,\ \xi_n(\omega)$ — случайные величины, заданные на одном и том же конечном или счетном вероятностном пространстве $(\Omega,\ \mathcal{F},\ \mathbf{P})$, а $g\,(X_1,\ \ldots,\ X_n)$ — функция, заданная на прямом произведении множеств значений этих случайных величин, то $\eta(\omega) = g\,(\xi_1(\omega),\ \ldots,\ \xi_n(\omega))$ — тоже случайная величина, определенная на $(\Omega,\ \mathcal{F},\ \mathbf{P})$. Если $(\Omega,\ \mathcal{F},\ \mathbf{P})$ – произвольное вероятностное пространство, то для того, чтобы суперпозиция $g\,(\xi_1,\ \ldots,\ \xi_n)$ была случайной величиной, нужно налагать на функцию $g\colon \Omega \to S$ те или иные условия, обеспечивающие измеримость суперпозиции $g\,(\xi_1,\ \ldots,\ \xi_n)$.

При решении многих вероятностных задач удобно использовать простые случайные величины: *индикаторы событий*. Для вероятностного пространства $(\Omega, \mathcal{F}, \mathbf{P})$ индикатор события $A \in \mathcal{F}$ определяется соотношениями

$$I_A = I\,(A) = I_A\,(\omega) = \begin{cases} 1 \text{ при } \omega \in A, \\ 0 \text{ при } \omega \notin A. \end{cases}$$

Свойства индикаторов.

1) $I\,(\varnothing) \equiv 0,\ I\,(\Omega) \equiv 1,\ I(A^c) = 1 - I(A)$.

2) Для любых событий $A_1,\ \ldots,\ A_n$

$$I\,(A_1 \cap \ldots \cap A_n) = I\,(A_1) \ldots I\,(A_n)\,.$$

3) Если события $A_1,\ \ldots,\ A_n$ попарно несовместны, то

$$I\,(A_1 \cup \ldots \cup A_n) = \sum_{k=1}^{n} I\,(A_i).$$

4) Если $A \cap B \neq \varnothing$, то $I\,(A \cup B) \neq I(A) + I(B)$.

5) В общем случае

$$I\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{s=1}^{n} (-1)^{s-1} \sum_{1 \leqslant k_1 < \ldots < k_s < n} I(A_{k_1} \ldots A_{k_s}). \tag{14}$$

Формулу (14) можно вывести с помощью уже перечисленных свойств и равенства

$$\overline{\bigcup_{k=1}^{n} A_k} = \bigcap_{k=1}^{n} \overline{A_k}.$$

Действительно:

$$I\left(\bigcup_{k=1}^{n} A_k\right) = 1 - I\left(\overline{\bigcup_{k=1}^{n} A_k}\right) = 1 - I\left(\bigcap_{k=1}^{n} \overline{A_k}\right) =$$

$$= 1 - \prod_{k=1}^{n} I\left(\overline{A_k}\right) = 1 - \prod_{k=1}^{n} (1 - I(A_k)) =$$

$$= \sum_{k=1}^{n} I(A_k) - \sum_{1 \leqslant k_1 < k_2 \leqslant n} I(A_{k_1}) I(A_{k_2}) +$$

$$+ \sum_{1 \leqslant k_1 < k_2 < k_3 \leqslant n} .I(A_{k_1}) I(A_{k_2}) I(A_{k_3}) - \ldots + (-1)^{n-1} I(A_1) \ldots I(A_n).$$

Заменяя произведения индикаторов индикаторами пересечений событий, получаем:

$$I\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{s=1}^{n} (-1)^{s-1} \sum_{1 \leqslant k_1 < \ldots < k_s < n} I(A_{k_1} \ldots A_{k_s}).$$

§ 9. Закон распределения случайной величины

Пусть на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbf{P})$ задана случайная величина ξ, и множество ее значений $S = \{x_1, x_2, \ldots\}$ конечно или счетно. Тогда события $A_j = \{\omega \colon \xi(\omega) = x_j\}$, $j = 1, 2, \ldots$, образуют конечное или счетное разбиение Ω. Это разбиение порождает алгебру событий (подмножеств Ω)

$$A(\xi) = \{A = \{\omega \in \Omega \colon \xi(\omega) \in B\} : \ B \subseteq S\} \in \mathcal{F},$$

поскольку для любого $B \subseteq S$

$$\{\omega \in \Omega \colon \xi(\omega) \in B\} = \left\{\omega \in \Omega \colon \xi(\omega) \in \bigcup_{j \colon x_j \in B} \{x_j\}\right\} =$$

$$= \bigcup_{j \colon x_j \in B} \{\omega \in \Omega \colon \xi(\omega) = x_j\} = \bigcup_{j \colon x_j \in B} A_j.$$

Алгебру $A(\xi)$ называют *алгеброй событий, порожденной случайной величиной* ξ.

Так как события A_j не пересекаются, то

$$\xi = \xi(\omega) = \sum_{k=1}^{n} I_{A_k}(\omega) x_k.$$

Такое представление случайной величины будет использоваться в дальнейшем.

В общем случае σ-алгебра событий, порожденная случайной величиной ξ, которая определена на вероятностном пространстве $(\Omega,\ \mathcal{F},\ \mathbf{P})$ и принимает значения в измеримом пространстве (S, Σ), определяется равенством

$$A(\xi) = \{A = \{\omega \in \Omega \colon \xi(\omega) \in B\},\ \ B \in \Sigma\} \in \mathcal{F}.$$

Случайная величина ξ со значениями в множестве S кроме алгебры подмножеств $A(\xi)$ порождает также меру на пространстве (S, Σ) по формуле

$$\mathbf{P}_\xi\{B\} = \mathbf{P}\{\omega \colon \xi(\omega) \in B\}, \quad B \in \Sigma.$$

Эта мера называется *законом распределения* (или просто *распределением*) случайной величины ξ. Как правило, для вероятностей событий, определяемых значениями случайных величин, используется сокращенная запись:

$$\mathbf{P}\{\omega \colon \xi(\omega) \in B\} = \mathbf{P}\{\xi \in B\}.$$

Закон распределения является мерой на пространстве значений случайной величины и не связан жестко с каким-либо вероятностным пространством. Случайные величины с одним и тем же законом распределения могут быть определены на разных вероятностных пространствах, и на одном вероятностном пространстве можно определить разные случайные величины с одним и тем же законом распределения. Например, вероятностную модель бросания идеального кубика можно построить на вероятностном пространстве с пространством элементарных событий $\Omega_1 = \{1, 2, 3, 4, 5, 6\}$ или с пространством элементарных событий $\Omega_2 = [0,\ 1]$. С другой стороны, на вероятностном пространстве $([0, 1], \mathcal{B}, \mathbf{P})$, где $\mathcal{B}$ — борелевская σ-алгебра, а $\mathbf{P}$ — мера Лебега, можно определить разные случайные величины, принимающие значения 0 и 1 с равными вероятностями:

$$\xi_1(\omega) = \begin{cases} 0, & 0 \leqslant \omega \leqslant \frac{1}{2}, \\ 1, & \frac{1}{2} < \omega \leqslant 1, \end{cases} \quad \xi_2(\omega) = \begin{cases} 1, & 0 \leqslant \omega \leqslant \frac{1}{2}, \\ 0, & \frac{1}{2} < \omega \leqslant 1. \end{cases}$$

В теории вероятностей, как правило, изучаются только свойства распределений случайных величин, поэтому при решении задач можно выбирать вероятностное пространство наиболее удобным для решения задачи образом.

Если случайная величина ξ принимает лишь конечное или счетное множество значений $X = \{x_1, x_2, \ldots\}$, то закон ее распределения полностью определяется значениями $x_1, x_2, \ldots$ и соответствующими им вероятностями

$$p_j = \mathbf{P}\{\xi = x_j\} \geqslant 0, \quad j = 1, 2, \ldots, \quad \sum\nolimits_{j\geqslant 1} p_j = 1. \tag{15}$$

Иногда для наглядности закон распределения задают таблицей:

x_1	x_2	$\ldots$	x_n	$\ldots$
p_1	p_2	$\ldots$	p_n	$\ldots$

По заданному формулой (15) или таблицей закону распределения ξ легко вычисляются вероятности любых событий, связанных со случайной величиной ξ:

$$\mathbf{P}_\xi\{B\} = \mathbf{P}\{\xi(\omega) \in B\} = \sum_{j:\, x_j \in B} p_j.$$

Если случайная величина ξ принимает действительные значения, то закон ее распределения удобно описывать функцией распределения

$$F_\gamma(x) = \mathbf{P}\{\xi \leqslant x\}, \quad -\infty < x < \infty.$$

Из определения функции распределения следует, что при любых $a < b$

$$\mathbf{P}\{a < \xi \leqslant b\} = \mathbf{P}\{\{\xi \leqslant b\} \backslash \{\xi \leqslant a\}\} = \mathbf{P}\{\xi \leqslant b\} - \mathbf{P}\{\xi \leqslant a\} = F_\xi(b) - F_\xi(a).$$

Теорема. *Функция распределения $F(x) = \mathbf{P}\{\xi \leqslant x\}$ обладает следующими свойствами:*

1. *$F(x)$ не убывает,*

2. $\lim\limits_{x\to-\infty} F(x) = 0, \quad \lim\limits_{x\to\infty} F(x) = 1,$

3. $F(x)$ *непрерывна справа и имеет пределы слева*: $F(x) = \lim_{y\downarrow x} F(y)$, $\mathbf{P}\{\xi < x\} = F(x-0) = \lim\limits_{y\uparrow x} F(y)$.

Доказательство. 1. Если $x < y$, то $\{\xi \leqslant y\} \supseteq \{\xi \leqslant x\}$ и

$$F(y) - F(x) = \mathbf{P}\{\xi \leqslant y\} - \mathbf{P}\{\xi \leqslant x\} = \mathbf{P}\{x < \xi \leqslant y\} \geq 0.$$

2. Рассмотрим неубывающую последовательность событий $C_0 = \varnothing, C_n = \{\omega\colon -n < \xi \leqslant n\}$, $n = 1, 2, \ldots$. События $C_n \backslash C_{n-1}$, $n = 1, 2, \ldots$, попарно несовместны и образуют счетное разбиение: $\Omega = \bigsqcup_{n\geqslant 1}(C_n \backslash C_{n-1})$. Поэтому

$$1 = \mathbf{P}\{\Omega\} = \sum_{n\geqslant 1} \mathbf{P}\{C_n \backslash C_{n-1}\} = \lim_{N\to\infty} \sum_{n=1}^{N} \mathbf{P}\{C_n \backslash C_{n-1}\} =$$
$$= \mathbf{P}\left\{C_N\right\} = \lim_{N\to\infty}(F(N) - F(-N)).$$

Из монотонности $F(x)$ и того, что $F(x) \in [0, 1]$ при всех x, следует существование пределов $\lim_{N\to\infty} F(N) \leqslant 1$ и $\lim_{N\to\infty} F(-N) \geqslant 0$, а из последнего равенства — что

$$\lim_{x\to\infty} F(-\infty) = 0, \quad \lim_{x\to\infty} F(\infty) = 1.$$

3. Рассмотрим последовательность не пересекающихся множеств

$$B_n = \left(\omega\colon x + \tfrac{1}{n+1} < \xi\left(\omega\right) \leqslant x + \tfrac{1}{n}\right), \quad n = 1, 2, \ldots$$

Тогда $\{\omega\colon x < \xi \leqslant x + 1\} = \bigsqcup_{n\geqslant 1} B_n$ и

$$F\left(x+1\right) - F\left(x\right) = \mathbf{P}\left\{x < \xi \leqslant x + 1\right\} = \mathbf{P}\left\{\bigsqcup_{n\geqslant 1} B_n\right\} =$$
$$= \sum_{n\geqslant 1} \mathbf{P}\left\{B_n\right\} = \sum_{n\geqslant 1}\left(F\left(x+\tfrac{1}{n}\right) - F\left(x + \tfrac{1}{n+1}\right)\right) =$$
$$= \lim_{N\to\infty} \sum_{n=1}^{N}\left(F\left(x+\tfrac{1}{n}\right) - F\left(x + \tfrac{1}{n+1}\right)\right) =$$
$$= \lim_{N\to\infty}\left(F(x+1) - F\left(x + \tfrac{1}{N+1}\right)\right) = F(x+1) - \lim_{N\to\infty} F\left(x + \tfrac{1}{N+1}\right).$$

Следовательно,

$$F\left(x\right) = \lim_{N\to\infty} F\left(x + \tfrac{1}{N+1}\right) = \lim_{y\downarrow x} F\left(y\right) = F\left(x+0\right).$$

Аналогично, рассматривая не пересекающиеся множества

$$A_1 = \{\omega\colon \xi \leqslant x-1\}\,, \quad A_n = \left\{x - \tfrac{1}{n-1} < \xi_n \leqslant x - \tfrac{1}{n}\right\}, \quad n \geqslant 2,$$

для которых $\{\omega\colon \xi < x\} = \bigsqcup_{n\geqslant 1} A_n$, находим

$$\begin{gathered}
\mathbf{P}\{\xi < x\} = \sum_{n\geqslant 1} \mathbf{P}\{A_n\} = \\
= F(x-1) + \sum_{n\geqslant 2} \left(F\left(x - \tfrac{1}{n}\right) - F\left(x - \tfrac{1}{n-1}\right)\right) = \\
= \lim_{N\to\infty} \left(F(x-1) + \sum_{n=2}^{N} \left(F\left(x - \tfrac{1}{n}\right) - F\left(x - \tfrac{1}{n-1}\right)\right)\right) = \\
= \lim_{N\to\infty} F\left(x - \tfrac{1}{N}\right) = \lim_{y\uparrow x} F(y) = F(x-0)\,.
\end{gathered}$$

В каждо точке x функция распределения $F(x)$ имеет пределы слева и справа. Их разность имеет простой смысл.

Следствие. $F_\xi(x+0) - F_\xi(x-0) = \mathbf{P}\{\xi \leqslant x\} - \mathbf{P}\{\xi < x\} = \mathbf{P}\{\xi = x\}$ *при любом* x.

Таким образом, функция распределения случайной величины ξ, принимающей действительные значения, не убывает и имеет разрывы в точках, где $\mathbf{P}\{\xi = x\} > 0$.

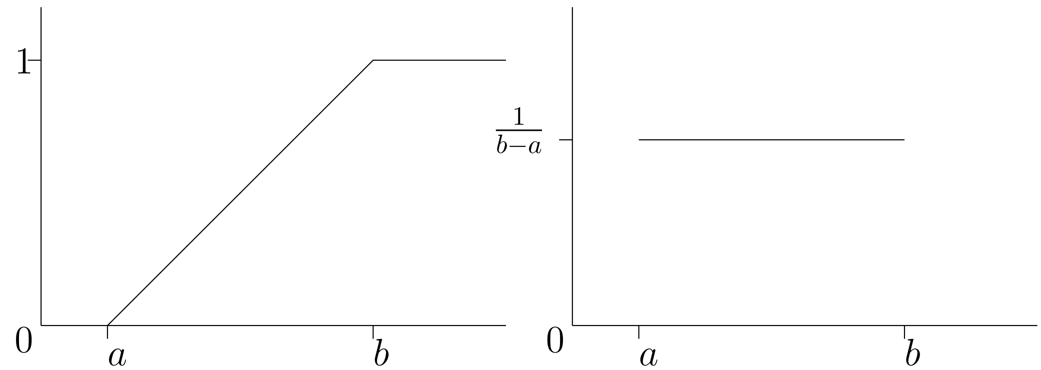
Функция распределения однозначно определяет значения $\mathbf{P}\{x < \xi \leqslant y\} = F_\xi(y) - F_\xi(x)$ меры на $\mathbb{R} = (-\infty, \infty)$, порожденной случайной величиной $\xi\colon \Omega \to \mathbb{R}$. Счетная аддитивность вероятностной меры, порожденной случайной величиной ξ, позволяет определить $F_\xi\{B\}$ для любых подмножеств $B \subset \mathbb{R}$ из борелевской σ-алгебры, порожденной такими полуинтервалами. При этом продолжении меры используется теорема Каратеодори.

Если существует такая функция $f_\xi(x)$, что $F_\xi(x) = \int_{-\infty}^{x} f_\xi(u)du$, то она называется *плотностью распределения* ξ, а само распределение — абсолютно непрерывным.

Пример. Говорят, что случайная величина ξ *имеет равномерное распределение* на отрезке $[a,b] \subset (-\infty, \infty)$, если

$$\mathbf{P}\{\xi \in [c,d]\} = \frac{d-c}{b-a} \text{ для любого отрезка } [c,d] \subset [a,b].$$

Функция распределения $F_\xi(x) = \mathbf{P}\{\xi \leqslant x\} = \mathbf{P}\{a \leqslant \xi \leqslant d\}$ такой случайной величины равна 0 при $x < a$, равна 1 при $x > b$ и равна $\frac{x-a}{b-a}$ при $a \leqslant x \leqslant b$. Плотность распределения этой случайной величины равна $\frac{1}{b-a}$ при $a \leqslant x \leqslant b$ и равна 0 вне отрезка $[a,b]$.



Функция распределения и плотность распределения равномерного распределения на $[a, b]$

Перечислим несколько распределений, часто возникающих в вероятностных задачах.

1. *Вырожденное распределение* — это распределение случайной величины, которая с вероятностью 1 принимает какое-то одно значение (т. е. как функция на пространстве элементарных событий Ω она равна константе на множестве меры 1).

2. Пусть $\eta_1, \eta_2, \ldots$ — индикаторы успехов в последовательности независимых испытаний Бернулли с вероятностью успеха p:

$$\mathbf{P}\{\eta_k = 1\} = p, \quad \mathbf{P}\{\eta_k = 0\} = 1 - p,\ k = 1, 2, \ldots$$

Биномиальное распределение с параметрами $(n,\, p)$ соответствует распределению числа $\mu_n = \eta_1 + \ldots + \eta_n$ успехов в n независимых испытаниях по схеме Бернулли, если в каждом испытании вероятность успеха равна p:

$$\mathbf{P}\{\mu_n = m\} = C_n^m p^m (1-p)^{n-m}, \quad m = 0,\, 1,\, \ldots,\, n.$$

3. *Геометрическое распределение* с параметром p определяется равенствами

$$\mathbf{P}\{\nu = m\} = p(1-p)^{m-1}, \quad m = 1, 2, \ldots$$

Одна из интерпретаций случайной величины с геометрическим распределением — порядковый номер первого закончившегося успехом испытания в схеме Бернулли с вероятностью успеха p: $\nu = \min\{n \geqslant 1 :\ \eta_n = 1\}$, где $\eta_1, \eta_2, \ldots$ — те же, что в предыдущем примере. Легко проверить, что

$$\mathbf{P}\{\nu > m\} = \sum_{k=m+1}^{\infty} p(1-p)^{k-1} = p\sum_{k=m}^{\infty}(1-p)^k = p\,\frac{(1-p)^m}{1-(1-p)} = (1-p)^m,$$

поэтому

$$\mathbf{P}\{\nu \leqslant m\} = 1 - (1-p)^m, \quad m = 0, 1, \ldots$$

4. *Равновероятное распределение* на конечном множестве $\{1, \ldots, N\}$ определяется равенствами

$$\mathbf{P}\{\xi = m\} = \frac{1}{N}, \quad m = 1, 2, \ldots, N.$$

5. *Гипергеометрическое распределение* с параметрами n, M и N является распределением числа ξ белых шаров среди n шаров, извлеченных по схеме равновероятного выбора без возвращения урны, содержавшей вначале M белых и $N - M$ черных шаров:

$$\mathbf{P}\{\xi = m\} = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}, \quad \max\{0, n - N + M\} \leqslant m \leqslant \min\{n, M\}.$$

6. *Распределение Пуассона* с параметром λ:

$$\mathbf{P}\{\xi = m\} = \left(\frac{\lambda^m}{m!}\right) e^{-\lambda}, \quad m = 0, 1, 2, \ldots$$

Очевидно,

$$\sum_{m \geqslant 0} \mathbf{P}\{\xi = m\} = \sum_{m \geqslant 0} \frac{\lambda^m}{m!} e^{-\lambda} = e^{\lambda} e^{-\lambda} = 1.$$

§ 10. Совместные распределения

Рассмотрим сначала простой пример. Пусть на одном вероятностном пространстве $(\Omega, \mathcal{F}, \mathbf{P})$ определены случайные величины $\xi = \xi(\omega)$ и $\eta = \eta(\omega)$. Пусть $X = \{x_1, \ldots, x_m\}$ и $Y = \{y_1, \ldots, y_n\}$ — множества всех возможных значений случайных величин ξ и η соответственно:

$$\mathbf{P}\{\omega \colon \xi = x_r\} = p_r \geqslant 0, \quad r = 1, \ldots, m, \ p_1 + \ldots + p_m = 1,$$
$$\mathbf{P}\{\omega \colon \eta = y_s\} = q_s \geqslant 0, \quad s = 1, \ldots, n, \ q_1 + \ldots + q_n = 1.$$

Множеством возможных значений вектора $(\xi(\omega), \eta(\omega))$ в этом примере является, очевидно, множество $X \times Y = \{(x_r, y_s), r \in \{1, \ldots, m\}, s \in \{1, \ldots, n\}\}$, и *совместное распределение* случайных величин ξ и η определяется вероятностями

$$v_{rs} = \mathbf{P}\{\omega \colon \xi(\omega) = x_r, \eta(\omega) = y_s\}, \quad r \in \{1, \ldots, m\}, \quad s \in \{1, \ldots, n\},$$

причем все v_{rs} неотрицательны и их сумма равна 1. Связь между совместным распределением $\{v_{rs}\}$, распределением $\{p_r\}$ случайной величины ξ и распределением $\{q_s\}$ случайной величины η описывается соотношениями

$$p_r = \mathbf{P}\{\xi = x_r\} = \mathbf{P}\left\{\bigcup_{s=1}^{n}\{\xi = x_r, \eta = y_s\}\right\} = \mathbf{P}\{(\xi,\eta)\in\{x_r\}\times Y\} = \sum_{s=1}^{n} v_{rs},$$
$$q_s = \mathbf{P}\{\eta = y_s\} = \mathbf{P}\left\{\bigcup_{r=1}^{m}\{\xi = x_r, \eta = y_s\}\right\} = \mathbf{P}\{(\xi,\eta)\in X\times\{y_s\}\} = \sum_{r=1}^{m} v_{rs},$$

которые должны выполняться для любых $r = 1,\ldots,m,\ s = 1,\ldots,n.$

Распределения случайных величин ξ и η называются *маргинальными* распределениями их совместного распределения (один из смыслов английского слова «margin» — поле на листе тетради или в таблице, так что маргинальные распределения — это проекции совместного распределения на множества значений компонент).

Рассмотрим теперь общий случай. Пусть на вероятностном пространстве $(\Omega,\ \mathcal{F},\ \mathbf{P})$ задано несколько случайных величин

$$\xi_1 = \xi_1(\omega),\ldots,\xi_k = \xi_k(\omega)$$

с измеримыми пространствами значений $(S_1,\Sigma_1),\ \ldots\ ,(S_k,\ \Sigma_k)$ (возможно, разными: например, $\xi_1(\omega)$ может быть случайной матрицей, а $\xi_2(\omega)$ быть рангом матрицы $xi_1(\omega)$). По этому набору случайных величин (функций, отображающих одно и то же пространство элементарных событий Ω соответственно в $S_1,\ \ldots,\ S_k$) можно построить новую случайную величину: случайный «вектор»

$$\Xi = \Xi(\omega) = (\xi_1,\ldots,\xi_k), \quad \omega\in\Omega,$$

принимающий значения в пространстве $S = S_1\times\ldots\times S_k$. В пространстве S рассматривается σ-алгебра подмножеств, порожденная прямыми произведениями элементов σ-алгебр $\Sigma_1,\ \ldots,\ \Sigma_k$. Распределение вектора Ξ (т.е. мера, порожденная этим «вектором» в пространстве S) называется *совместным распределением* случайных величин $\xi_1,\ \ldots,\ \xi_k$.

Разумеется, возможен и обратный переход: если на каком-то вероятностном пространстве $(\Omega,\ \mathcal{F},\ \mathbf{P})$ задан случайный вектор

$$\Xi = \Xi(\omega) = (\xi_1,\ldots,\xi_k), \quad \omega\in\Omega,$$

со значениями в пространстве $S = S_1 \times \ldots \times S_k$, то по нему однозначно определяются образующие его случайные компоненты

$$\xi_1 = \xi_1(\omega), \ldots, \xi_k = \xi_k(\omega),$$

а по распределению вектора Ξ в пространстве $S = S_1 \times \ldots \times S_k$ — распределения его компонент:

$$\mathbf{P}\{\xi_j \in B_j\} = \mathbf{P}\{\Xi \in S_1 \times \ldots \times S_{j-1} \times B_j \times S_{j+1} \times \ldots \times S_k\}$$

для любого $j = 1, \ldots, k$ и для любого $B_j \in \Sigma_j$. Эти («одномерные») распределения компонент случайного вектора являются его *маргинальными* распределениями.

§ 11. Математическое ожидание случайной величины с дискретным распределением

Пусть $(\Omega, \mathcal{F}, \mathbf{P})$ — вероятностное пространство, и пространство элементарных событий Ω конечное или счетное. Тогда мера $\mathbf{P}$ определяется своими значениями $\mathbf{P}(\omega)$ на элементах ω пространства элементарных событий Ω. Пусть, далее, $\xi = \xi(\omega)$ — случайная величина, определенная на $(\Omega, \mathcal{F}, \mathbf{P})$ и принимающая действительные значения.

Определение. *Математическим ожиданием* (или *средним значением*) случайной величины ξ, определенной на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbf{P})$ с конечным или счетным пространством элементарных событий Ω, называется сумма ряда

$$\mathbf{M}\xi = \sum_{\omega \in \Omega} \xi(\omega)\mathbf{P}(\omega),$$

если он сходится абсолютно: $\sum\limits_{\omega \in \Omega} |\xi(\omega)|\mathbf{P}(\omega) < \infty$; если же ряд не сходится абсолютно, то говорят, что случайная величина ξ *не имеет математического ожидания*.

Математическое ожидание обозначают также символом $\mathbf{E}$: $\mathbf{E}\,\xi$.

Математическое ожидание является естественным аналогом обычного среднего значения набора величин. Например, пусть в городе живет N человек и k-й человек получает зарплату $m(k)$ рублей. Тогда средняя зарплата M жителя города вычисляется по формуле

$$M = \frac{m(1) + \ldots + m(N)}{N} = \sum_{k=1}^{N} m(k)\frac{1}{N}.$$

В терминах теории вероятностей этой ситуации можно сопоставить пространство элементарных событий $\Omega = \{1, \ldots, N\}$ и ввести на нем вероятностную меру $\mathbf{P}\{k\} = \frac{1}{N}$ для всех $k = 1, \ldots, N$. Тогда $m(k)$ можно рассматривать как случайную величину, определенную на Ω, а дроби $\frac{1}{N}$ в сумме — как вероятности элементарных событий.

Определение математического ожидания для случайных величин, определенных на произвольном (не обязательно конечном или счетном) вероятностном пространстве и принимающих любые действительные значения, будет дано позже.

Рассмотрим основные свойства математического ожидания.

1) Если I_A – индикатор события A, то

$$\mathbf{M}I_A = \mathbf{P}\{A\}.$$

Действительно:

$$\mathbf{M}I_A = \sum_{\omega\in\Omega} I_A(\omega)\mathbf{P}(\omega) = \sum_{\omega\in A} \mathbf{P}(\omega) = \mathbf{P}\{A\}.$$

2) Конечная аддитивность: если каждая из случайных величин $\xi_1, \ldots, \xi_n$ имеет конечное математическое ожидание, то

$$\mathbf{M}(\xi_1 + \ldots + \xi_n) = \mathbf{M}\xi_1 + \ldots + \mathbf{M}\xi_n.$$

Доказательство. Рассмотрим сначала сумму двух случайных величин. Ввиду абсолютной сходимости рядов $\sum\limits_{\omega\in\Omega} \xi_1(\omega)\mathbf{P}(\omega)$ и $\sum\limits_{\omega\in\Omega} \xi_2(\omega)\mathbf{P}(\omega)$ их члены можно переставлять, не изменяя значения суммы, и поэтому

$$\mathbf{M}\xi_1 + \mathbf{M}\xi_2 = \sum_{\omega\in\Omega} \xi_1(\omega)\mathbf{P}(\omega) + \sum_{\omega\in\Omega} \xi_2(\omega)\mathbf{P}(\omega) =$$
$$= \sum_{\omega\in\Omega} (\xi_1(\omega) + \xi_2(\omega))\mathbf{P}(\omega) = \sum_{\omega\in\Omega} (\xi_1 + \xi_2)(\omega)\mathbf{P}(\omega) = \mathbf{M}(\xi_1 + \xi_2);$$

в последнем переходе использовано определение математического ожидания.

Общий случай следует из этого частного индукцией по числу слагаемых:

$$\mathbf{M}(\xi_1 + \ldots + \xi_n) = \mathbf{M}((\xi_1 + \ldots + \xi_{n-1}) + \xi_n) =$$
$$= (\mathbf{M}\xi_1 + \ldots + \mathbf{M}\xi_{n-1}) + \mathbf{M}\xi_n.$$

Условие конечности математических ожиданий слагаемых существенно. Например, если $\gamma = \gamma(\omega)$ — случайная величина, не имеющая математического ожидания, а $\xi(\omega) = -\gamma(\omega)$, то $\xi(\omega) + \gamma(\omega) = 0$ при любом $\omega \in \Omega$, и поэтому $\mathbf{M}(\xi + \gamma) = 0$, однако $\mathbf{M}(\xi + \gamma) \neq \mathbf{M}\xi + \mathbf{M}\gamma$, так как правая часть не определена.

Замечание. В § 2 методом математической индукции была доказана формула для вероятности объединения событий $A_1, \ldots, A_n$ (формула включения-исключения):

$$\mathbf{P}\left\{\bigcup_{k=1}^{n} A_k\right\} = \sum_{m=1}^{n} (-1)^{m-1} \sum_{1 \leqslant i_1 < \ldots < i_m \leqslant n} \mathbf{P}\left\{A_{i_1} A_{i_2} \ldots A_{i_m}\right\}.$$

В § 10 с помощью простых свойств индикаторов была доказана формула для индикатора объединения событий:

$$I\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{s=1}^{n} (-1)^{s-1} \sum_{1 \leqslant k_1 < \ldots < k_s < n} I(A_{k_1} \ldots A_{k_s}).$$

Эти формулы тесно связаны: первое равенство получается из второго вычислением математического ожидания от левой и правой частей.

3) Счетная аддитивность: если случайные величины $\xi_1, \xi_2, \ldots$ удовлетворяют условию

$$\sum_{n=1}^{\infty} \mathbf{M}|\xi_n| < \infty, \tag{16}$$

то

$$\mathbf{M}\left\{\sum_{n=1}^{\infty} \xi_n\right\} = \sum_{n=1}^{\infty} \mathbf{M}\xi_n. \tag{17}$$

Действительно, правую часть (17) можно представить в виде двойного ряда:

$$\sum_{n=1}^{\infty} \mathbf{M}\xi_n = \sum_{n=1}^{\infty} \sum_{\omega \in \Omega} \xi_n(\omega) \mathbf{P}(\omega);$$

этот ряд по условию (16) сходится абсолютно:

$$\sum_{n=1}^{\infty} \sum_{\omega \in \Omega} |\xi_n(\omega) \mathbf{P}(\omega)| = \sum_{n=1}^{\infty} \mathbf{M}|\xi_n| < \infty,$$

поэтому можно изменить порядок суммирования:

$$\sum_{n=1}^{\infty}\mathbf{M}\xi_n=\sum_{n=1}^{\infty}\sum_{\omega\in\Omega}\xi_n(\omega)\mathbf{P}(\omega)=\sum_{\omega\in\Omega}\left\{\sum_{n=1}^{\infty}\xi_n(\omega)\right\}\mathbf{P}(\omega)=\mathbf{M}\left\{\sum_{n=1}^{\infty}\xi_n\right\},$$

что и требовалось доказать.

Замечание. Если условие (16) не выполняется, то (17) может не выполняться. Например, если $\Omega=\{0,\,1\}$, $\mathbf{P}\{0\}=\mathbf{P}\{1\}=\frac{1}{2}$ и

$$\xi_1(0)=\xi_2(0)=\ldots=1,\quad \xi_1(1)=\xi_2(1)=\ldots=-1,$$

то $\mathbf{M}\xi_k=0$, $\mathbf{M}|\xi_k|=1$, $k=1,2,\ldots$, так что (16) не выполняется. В этом случае правая часть (17) равна 0, а в левой части

$$\sum\nolimits_{k=1}^{\infty}\xi_k(0)=+\infty,\quad \sum\nolimits_{k=1}^{\infty}\xi_k(1)=-\infty,$$

т. е. левая часть (17) не определена.

4) Линейность: если $b,c\in R$ — константы, а ξ — случайная величина, то

$$\mathbf{M}(c\xi+b)=c\mathbf{M}\xi+b.$$

Действительно,

$$\mathbf{M}(c\xi+b)=\sum_{\omega\in\Omega}(c\xi(\omega)+b)\mathbf{P}(\omega)=$$
$$=c\sum_{\omega\in\Omega}\xi(\omega)\mathbf{P}(\omega)+b\sum_{\omega\in\Omega}\mathbf{P}(\omega)=c\mathbf{M}\xi+b.$$

5) Монотонность: если $\mathbf{M}\xi$ и $\mathbf{M}\eta$ существуют и $\xi\geqslant\eta$, то $\mathbf{M}\xi\geqslant\mathbf{M}\eta$.

Доказательство:

$$\mathbf{M}\xi-\mathbf{M}\eta=\sum_{\omega\in\Omega}\xi(\omega)\mathbf{P}(\omega)-\sum_{\omega\in\Omega}\eta(\omega)\mathbf{P}(\omega)=\sum_{\omega\in\Omega}(\xi(\omega)-\eta(\omega))\mathbf{P}(\omega)\geqslant 0,$$

так как $\xi(\omega)\geq\eta(\omega)$ при всех $\omega\in\Omega$.

6) Если $\xi\geqslant 0$ и $\mathbf{M}\xi=0$, то $\mathbf{P}\{\xi=0\}=1$.

Доказательство. По условию $0=\mathbf{M}\xi=\sum\limits_{\omega\in\Omega}\xi(\omega)\mathbf{P}(\omega)$. Так как все слагаемые в этой сумме неотрицательны, то она может быть равна

0 только если $\xi(\omega)\mathbf{P}(\omega) = 0$ при всех $\omega \in \Omega$. Значит, если $\xi(\omega) > 0$, то $\mathbf{P}(\omega) = 0$, и поэтому

$$\mathbf{P}\{\xi > 0\} = \sum_{\omega\in\Omega\colon \xi(\omega)>0} \mathbf{P}(\omega) = 0,$$

а $\mathbf{P}\{\xi = 0\} = 1 - \mathbf{P}\{\xi > 0\} = 1$.

7) Если $\mathbf{M}\xi$ существует, то $|\mathbf{M}\xi| \leqslant \mathbf{M}|\xi|$.

Действительно,

$$|\mathbf{M}\xi| = \left|\sum\nolimits_{\omega\in\Omega} \xi(\omega)\mathbf{P}(\omega)\right| \leqslant \sum\nolimits_{\omega\in\Omega} |\xi(\omega)|\mathbf{P}(\omega) = \mathbf{M}|\xi|.$$

8) Пусть случайная величина ξ принимает значения x_1, x_2, … и

$$\mathbf{P}\{\xi = x_k\} = p_k,\ k \geqslant 1, \quad \sum_{k\geqslant 1} p_k = 1, \quad \sum_{k\geqslant 1} |x_k|\, p_k < \infty.$$

Тогда

$$\mathbf{M}\xi = \sum_{k\geqslant 1} x_k\, \mathbf{P}\{\xi = x_k\} = \sum_{k\geqslant 1} x_k\, p_k.$$

Доказательство. Положим $A_k = \{\omega \in \Omega : \xi(\omega) = x_k\}$, $k = 1, 2, \ldots$, и представим случайную величину ξ в виде суммы индикаторов с коэффициентами:

$$\xi = \sum_{k\geqslant 1} x_k\, I\{\xi = x_k\} = \sum_{k\geqslant 1} x_k\, I\{A_k\}.$$

Теперь вычислим математическое ожидание от обеих частей этого равенства, пользуясь счетной аддитивностью $\big(\mathbf{M}|x_k\, I\{A_k\}| = |x_k|\, p_k,\ k \geqslant 1$, и по условию $\sum_{k\geqslant 1} \mathbf{M}|x_k\, I\{A_k\}| = \sum_{k\geqslant 1} |x_k|\, p_k < \infty\big)$ и линейностью математического ожидания:

$$\mathbf{M}\xi = \mathbf{M}\sum_{k\geqslant 1} x_k\, I\{\xi = x_k\} = \sum_{k\geqslant 1} \mathbf{M}x_k\, I\{\xi = x_k\} =$$

$$= \sum_{k\geqslant 1} x_k\, \mathbf{M}I\{\xi = x_k\} = \sum_{k\geqslant 1} x_k\, \mathbf{P}\{\xi = x_k\}.$$

Свойство 8) позволяет частично обобщить определение математического ожидания на случайные величины с не более счетным множеством значений, определенные на произвольном (не обязательно не более чем счетном) пространстве элементарных событий.

Определение. *Математическим ожиданием* (или *средним значением*) случайной величины ξ, определенной на вероятностном пространстве $(\Omega,\, \mathcal{F},\, \mathbf{P})$ и принимающей значения из не более чем счетного множества $\{x_1, x_2, \ldots\} \subset (-\infty, \infty)$, называется величина

$$\mathbf{M}\xi = \sum_{k\geqslant 1} x_k \, \mathbf{P}\{\xi = x_k\},$$

если $\sum_{k\geqslant 1} |x_k| \, \mathbf{P}\{\xi = x_k\} < \infty$, т. е. ряд сходится абсолютно; в противном случае говорят, что случайная величина ξ не имеет математического ожидания.

Из свойства 8) следует, что для вероятностных пространств с не более чем счетным пространством элементарных событий новое определение совпадает с исходным.

Свойства 1) – 7) справедливы и для нового определения математического ожидания, но их доказательства нужно немного изменить, так как в общем случае пространство элементарных событий Ω может не быть счетным и суммировать по всем $\omega \in \Omega$ нельзя.

В качестве примера докажем равенство

$$\mathbf{M}(\xi + \eta) = \mathbf{M}\xi + \mathbf{M}\eta.$$

Пусть дискретные случайные величины ξ и η заданы на одном вероятностном пространстве и $S_1, S_2 \subset \mathbb{R}$ — такие счетные множества, что $\mathbf{P}\{\xi \in S_1\} = \mathbf{P}\{\eta \in S_2\} = 1$. Для любых $u \in S_1, v \in S_2$ положим $p_{u,v} = \mathbf{P}\{\xi = u, \eta = v\}$, тогда

$$\mathbf{P}\{\xi = u\} = \sum_{v\in S_2} p_{u,v},\ \mathbf{P}\{\eta = v\} = \sum_{u\in S_1} p_{u,v},$$

$$\mathbf{P}\{\xi + \eta = z\} = \sum_{u\in S_1 \colon z-u\in S_2} p_{u,z-u},$$

множество $S = \{z = u + v \colon u \in S_1, v \in S_2\}$ значений z, для которых $\mathbf{P}\{\xi + \eta = z\} > 0$, не более чем счетно. Если $\mathbf{M}\xi$ и $\mathbf{M}\eta$ существуют, то ряды

$$\sum_{u\in S_1} u\mathbf{P}\{\xi = u\} = \mathbf{M}\xi, \quad \sum_{v\in S_2} v\mathbf{P}\{\eta = v\} = \mathbf{M}\eta$$

сходятся абсолютно, поэтому их слагаемые можно объединять в группы и переставлять:

$$\mathbf{M}\xi + \mathbf{M}\eta = \sum_{u\in S_1} u\mathbf{P}\{\xi = u\} + \sum_{v\in S_2} v\mathbf{P}\{\eta = v\} =$$
$$= \sum_{u\in S_1} u \sum_{v\in S_2} p_{u,v} + \sum_{v\in S_2} v \sum_{u\in S_1} p_{u,v} =$$
$$= \sum_{z\in S}\left(\sum_{u\in S_1\colon z-u\in S_2} u p_{u,z-u} + \sum_{u\in S_1\colon z-u\in S_2} (z-u) p_{u,z-u}\right) =$$
$$= \sum_{z\in S} z \sum_{u\in S_1\colon z-u\in S_2} p_{u,z-u} = \sum_{z\in S} z\mathbf{P}\{\xi+\eta = z\} = \mathbf{M}(\xi+\eta).$$

9) **Формула полного математического ожидания.** Математическое ожидание случайной величины $\xi = \xi(\omega)$ со значениями $x_1, x_2, \ldots$, определенной на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbf{P})$, может вычисляться не только относительно вероятностной меры $\mathbf{P}$, но и относительно других вероятностных мер. Например, если $B \in \mathcal{F}$ — событие и $\mathbf{P}\{B\} > 0$, то определены условные вероятности $\mathbf{P}\{\xi = x_i \,|\, B\} = \frac{\mathbf{P}\{\{\xi=x_i\}\cap B\}}{\mathbf{P}\{B\}}$ для $i = 1, 2, \ldots$, и математическое ожидание ξ относительно этой меры называется *условным математическим ожиданием при условии события* B:

$$\mathbf{M}\{\xi \,|\, B\} = \sum_{i\geqslant 1} x_i \mathbf{P}\{\xi = x_i \,|\, B\}.$$

Теорема (Формула полного математического ожидания). *Если* A_1, A_2, ... — *не более чем счетное разбиение* Ω *и* $\mathbf{P}\{A_k\} > 0$ *при всех* k, *то для любой случайной величины* ξ, *определенной на* $(\Omega, \mathcal{F}, \mathbf{P})$ *с не более чем счетным множеством значений и имеющей конечное математическое ожидание, справедливо равенство*

$$\mathbf{M}\xi = \sum_{k\geqslant 1} \mathbf{M}\xi\mathbb{I}\{A_k\} = \sum_{k\geqslant 1} \mathbf{P}\{A_k\}\mathbf{M}\{\xi \,|\, A_k\}. \qquad (18)$$

Доказательство. Пусть случайная величина ξ принимает значения $x_1, x_2, \ldots$ и $C_i = \{\omega\colon \xi(\omega) = x_i\}, i = 1, 2, \ldots$ Так как $A_1, A_2, \ldots$ — разбиение Ω, то в силу аддитивности математического ожидания

$$\xi = \sum_{k\geqslant 1} \xi\mathbb{I}_{A_k}, \quad \mathbf{M}\xi = \sum_{k\geqslant 1} \mathbf{M}\xi\mathbb{I}_{A_k}.$$

Далее, каждое слагаемое $\mathbf{M}\xi\mathbb{I}_{A_k}$ можно представить в виде

$$\mathbf{M}\xi\mathbb{I}_{A_k} = \mathbf{M}\sum_{i\geqslant 1} x_i\mathbb{I}_{C_i}\mathbb{I}_{A_k} = \sum_{i\geqslant 1} x_i\mathbf{P}\{C_i\cap A_k\} =$$

$$\sum_{i\geqslant 1} x_i\mathbf{P}\{A_k\}\mathbf{P}\{C_i\,|\,A_k\} = \mathbf{P}\{A_k\}\sum_{i\geqslant 1} x_i\mathbf{P}\{C_i\,|\,A_k\} = \mathbf{P}\{A_k\}\mathbf{M}\{\xi\,|\,A_k\}.$$

Из полученных равенств следует (18). Теорема доказана.

Замечание. Условие существования математического ожидания случайной величины ξ существенно. Например, если $\Omega = \mathbb{Z} = \{0,\pm1,\pm2,\ldots\}$,

$$\mathbf{P}\{k\} = \begin{cases} 0, & k=0,\\ \frac{1}{2^{|k|+1}}, & k\neq 0, \end{cases} \quad \xi(k) = \begin{cases} 0, & k=0,\\ 2^k, & k>0,\\ -2^k, & k<0, \end{cases}$$

то $\mathbf{M}\xi$ не существует, но для разбиения $A_0 = \{0\}, A_k = \{k,-k\}$, $k = 1,2,\ldots$ имеем:

$$\mathbf{P}\{k\,|\,A_k\} = \mathbf{P}\{-k\,|\,A_k\} = \frac{1}{2}, \quad \mathbf{M}\{\xi\,|\,A_k\} = 0,\ k=1,2,\ldots,$$

так что равенство (18) не имеет места.

Этот пример показывает, что при вычислении математического ожидания случайной величины по формуле полной вероятности нужно проверять, что для этой случайной величины математическое ожидание определено.

§ 12. Степенные моменты случайных величин

Пусть ξ — случайная величина, принимающая значения из множества $\{x_1,x_2,\ldots\}$, а $g(x)$ — числовая функция. Тогда $\eta = g(\xi)$ — тоже случайная величина,

$$\eta = \sum_{k\geqslant 1} g\,(x_k) I\,\{\xi = x_k\}\,.$$

Если $\sum_{k\geqslant 1}|g(x_k)|\mathbf{P}\{\xi = x_k\} < \infty$, то по свойству счетной аддитивности

$$\mathbf{M}g(\xi) = \mathbf{M}\eta = \sum_{k\geqslant 1}^{n} g\,(x_k)\mathbf{P}\,\{\xi = x_k\}\,.$$

Будем предполагать, не оговаривая это специально, что все рассматриваемые здесь математические ожидания существуют.

Полагая $g(x) = x^n$, $g(x) = |x|^n$, $g(x) = (x - \mathbf{M}\xi)^n$, мы получаем величины, каждая из которых имеет свое название:

$$\mathbf{M}\xi^n = \sum_{k\geqslant 1} x_k^n \mathbf{P}\left\{\xi = x_k\right\}$$

— *момент* n-го порядка,
для случайной величины ξ, принимающей только целочисленные неотрицательные значения $0, 1, \ldots,$

$$\mathbf{M}\xi^{[n]} = \mathbf{M}\xi(\xi-1)\ldots(\xi-n+1) = \sum_{k\geqslant 0} k(k-1)\ldots(k-n+1)\mathbf{P}\left\{\xi = k\right\}$$

— факториальный момент n-го порядка,

$$\mathbf{M}|\xi|^n = \sum_{k\geqslant 1}^{n} |x_k|^n \mathbf{P}\left\{\xi = x_k\right\}$$

— *абсолютный* момент n-го порядка,

$$\mathbf{M}(\xi - \mathbf{M}\xi)^n = \sum_{k\geqslant 1}^{n} (x_k - \mathbf{M}\xi)^n \mathbf{P}\left\{\xi = x_k\right\}$$

— *центральный* момент n-го порядка.

Случайная величина $\xi^* = \xi - \mathbf{M}\xi$ называется *центрированной*; ее математическое ожидание равно 0.

Центральный момент второго порядка всегда неотрицателен и имеет собственное название — *дисперсия* — и два обозначения: $\mathbf{D}\xi$ и $\sigma^2(\xi)$, поскольку является важной характеристикой распределения случайной величины:

$$\mathbf{D}\xi = \sigma^2\left(\xi\right) = \mathbf{M}\left(\xi - \mathbf{M}\,\xi\right)^2.$$

Так же, как математическое ожидание является средним значением случайной величины, дисперсия является средним значением *квадрата отклонения* случайной величины от ее математического ожидания. Величина $\sigma(\xi) = \sqrt{\mathbf{D}\xi}$ называется *среднеквадратическим отклонением* случайной величины ξ и характеризует степень разброса значений случайной величины ξ около $\mathbf{M}\xi$.

Свойства дисперсии:

1. Выражение через степенные моменты:

$$\mathbf{D}\xi = \mathbf{M}\xi^2 - (\mathbf{M}\,\xi)^2.$$

Доказательство:

$$\begin{gathered}\mathbf{D}\xi = \mathbf{M}(\xi - \mathbf{M}\xi)^2 = \mathbf{M}(\xi^2 - 2\xi\mathbf{M}\xi + (\mathbf{M}\xi)^2) = \\ = \mathbf{M}\xi^2 - 2\mathbf{M}\xi\mathbf{M}\xi + (\mathbf{M}\xi)^2 = \mathbf{M}\xi^2 - (\mathbf{M}\xi)^2.\end{gathered}$$

2. $\mathbf{D}\xi \geq 0$; если $\mathbf{D}\xi = 0$, то существует такая константа C, что

$$\mathbf{P}\left\{\xi = C\right\} = 1.$$

Первое неравенство очевидно, так как $\mathbf{D}\xi = \mathbf{M}(\xi - \mathbf{M}\xi)^2$ является математическим ожиданием неотрицательной случайной величины $(\xi - \mathbf{M}\xi)^2$, а второе утверждение следует из свойства 4) математического ожидания: если $\mathbf{M}(\xi - \mathbf{M}\xi)^2 = 0$, то, поскольку

$$\mathbf{P}\left\{(\xi - \mathbf{M}\xi)^2 \geqslant 0\right\} = 1,$$

постольку $\mathbf{P}\left\{(\xi - \mathbf{M}\xi)^2 = 0\right\} = 1$, т.е. $\mathbf{P}\{\xi = \mathbf{M}\xi\} = 1$.

3. При любом действительном C

$$\mathbf{D}(\xi + C) = \mathbf{D}\xi, \quad \mathbf{D}(C\xi) = C^2\mathbf{D}\xi.$$

Действительно,

$$\begin{gathered}\mathbf{D}(\xi + C) = \mathbf{M}\left(\xi + C - \mathbf{M}(\xi + C)\right)^2 = \mathbf{M}(\xi - \mathbf{M}\xi)^2 = \mathbf{D}\xi, \\ \mathbf{D}(C\xi) = \mathbf{M}\left(C\xi - \mathbf{M}(C\xi)\right)^2 = \mathbf{M}(C\xi - C\mathbf{M}\xi)^2 = \\ = \mathbf{M}C^2(\xi - \mathbf{M}\xi)^2 = C^2\mathbf{M}(\xi - \mathbf{M}\xi)^2 = C^2\mathbf{D}\xi.\end{gathered}$$

4. Для любой случайной величины ξ с $\mathbf{D}\xi < \infty$ и любого $c \in R$

$$\mathbf{M}(\xi - c)^2 \geqslant \mathbf{D}\xi \quad \text{т. е.} \quad \mathbf{D}\xi = \min_{c\in\mathbb{R}} \mathbf{M}(\xi - c)^2.$$

Действительно,

$$\begin{gathered}\mathbf{M}(\xi - c)^2 = \mathbf{M}\left(\xi - \mathbf{M}\xi + \mathbf{M}\xi - c\right)^2 = \\ = \mathbf{M}(\xi - \mathbf{M}\xi)^2 + 2\mathbf{M}(\xi - \mathbf{M}\xi)(\mathbf{M}\xi - c) + \mathbf{M}(\mathbf{M}\xi - c)^2 = \\ = \mathbf{D}\xi + (\mathbf{M}\xi - c)^2 \geqslant \mathbf{D}\xi,\end{gathered}$$

так как $2\mathbf{M}(\xi - \mathbf{M}\xi)(\mathbf{M}\xi - c) = 2(\mathbf{M}\xi - c)(\mathbf{M}\xi - \mathbf{M}\xi) = 0$.

§ 13. Неравенства для моментов и распределений

А. Неравенства Маркова и Чебышёва.

Неравенство Маркова. *Для любой случайной величины* ξ *и любого* $x > 0$

$$\mathbf{P}\left\{|\xi| \geqslant x\right\} \leqslant \frac{\mathbf{M}|\xi|}{x}.$$

Доказательство. Рассмотрим разбиение Ω на события $\{|\xi| \geq x\}$ и $\{|\xi| < x\}$. Тогда

$$\begin{gathered}\mathbf{M}|\xi| = \mathbf{M}|\xi|(I\{|\xi| \geqslant x\} + I\{|\xi| < x\}) = \\ = \mathbf{M}|\xi| I\{|\xi| \geqslant x\} + \mathbf{M}|\xi| I\{|\xi| < x\} \geqslant \\ \geqslant \mathbf{M} x I\{|\xi| \geqslant x\} + \mathbf{M} 0 \cdot I\{|\xi| < x\} = x\mathbf{P}\{|\xi| \geqslant x\}.\end{gathered}$$

Разумеется, неравенство Маркова содержательно только при $x > \mathbf{M}|\xi|$.

Неравенства Чебышёва (1874). *Для любой случайной величины* ξ *и для любого* $x > 0$

$$\mathbf{P}\left\{|\xi - \mathbf{M}\xi| \geqslant x\right\} \leqslant \frac{\mathbf{D}\xi}{x^2}, \quad \mathbf{P}\left\{|\xi - \mathbf{M}\xi| < x\right\} \geqslant 1 - \frac{\mathbf{D}\xi}{x^2}.$$

Доказательство. Первое неравенство — следствие неравенства Маркова, примененного к квадрату центрированной случайной величины: $\eta = (\xi - \mathbf{M}\xi)^2$. Второе неравенство эквивалентно первому, так как

$$\mathbf{P}\{|\xi - \mathbf{M}\xi| < x\} = 1 - \mathbf{P}\{|\xi - \mathbf{M}\xi| \geqslant x\}.$$

Содержательный смысл неравенства Чебышёва в том, что основная масса распределения случайной величины ξ с конечной дисперсией $\mathbf{D}\xi$ сосредоточена в интервале, длина которого имеет тот же порядок, что $\sqrt{\mathbf{D}\xi}$. В частности, если $\mathbf{D}\xi$ мало, то почти все распределение ξ сосредоточено около $\mathbf{M}\xi$.

Б. Неравенство Коши–Буняковского.

Неравенство Коши–Буняковского. *Для любых случайных величин* ξ *и* η

$$|\mathbf{M}\xi\eta| \leqslant \mathbf{M}|\xi\eta| \leqslant \sqrt{\mathbf{M}\xi^2\,\mathbf{M}\eta^2}.$$

Доказательство. Первое неравенство — свойство математического ожидания. Рассмотрим квадратичную форму

$$Q(x,y) = \mathbf{M}(x|\xi| + y|\eta|)^2 = x^2\mathbf{M}\xi^2 + 2xy\mathbf{M}|\xi\eta| + y^2\mathbf{M}\eta^2.$$

Так как $(x|\xi|+y|\eta|)^2 \geqslant 0$, то квадратичная форма $Q(x,y)$ неотрицательно определена. Следовательно, неотрицательно определена ее матрица

$$\begin{pmatrix} \mathbf{M}\xi^2 & \mathbf{M}|\xi\eta| \\ \mathbf{M}|\xi\eta| & \mathbf{M}\eta^2 \end{pmatrix},$$

а тогда неотрицателен определитель этой матрицы:

$$\det \begin{vmatrix} \mathbf{M}\xi^2 & \mathbf{M}|\xi\eta| \\ \mathbf{M}|\xi\eta| & \mathbf{M}\eta^2 \end{vmatrix} = \mathbf{M}\xi^2\mathbf{M}\eta^2 - (\mathbf{M}|\xi\eta|)^2 \geqslant 0.$$

Отсюда непосредственно следует неравенство Коши–Буняковского.

В. Неравенства для математического ожидания функции от случайной величины.

Лемма. *Пусть ξ – случайная величина, принимающая значения из множества $A \subseteq R$ и имеющая конечное математическое ожидание, $g(x): \ A \to R$ — функция и существует такое $c \in R$, что*

$$g\,(x) \geqslant g(\mathbf{M}\xi) + (x - \mathbf{M}\xi)\,c \quad \textit{для всех } x \in A. \tag{19}$$

Тогда

$$\mathbf{M}g(\xi) \geqslant g(\mathbf{M}\xi).$$

Доказательство леммы следует из свойств монотонности и линейности математического ожидания. Заменим в (19) x на ξ и вычислим математическое ожидание от обеих частей:

$$\mathbf{M}g(\xi) \geqslant \mathbf{M}\left(g\left(\mathbf{M}\xi\right) + \left(\xi - \mathbf{M}\xi\right)c\right) = g\left(\mathbf{M}\xi\right) + \mathbf{M}\left(\xi - \mathbf{M}\xi\right)c = g\left(\mathbf{M}\xi\right),$$

что и требовалось доказать.

Простым следствием этой леммы является часто используемое неравенство Иенсена для математических ожиданий выпуклых функций от случайных величин, т.е. для таких функций $g(x)$, что для любого x существует такое число $c(x)$, что $g(y) \geq g(x) + c(x)(y-x)$ для всех y. Если функция $g(x)$ дважды дифференцируема и $g''(x) \geq 0$ при всех x, то $g(x)$ выпукла.

Неравенство Иенсена. *Если ξ – случайная величина с конечным математическим ожиданием и $g(x)$ – выпуклая функция, то*

$$\mathbf{M}g(\xi) \geqslant g(\mathbf{M}\xi).$$

Неравенство Ляпунова. *Если ξ – случайная величина и $0 < \alpha < \beta$, то*
$$(\mathbf{M}|\xi|^{\alpha})^{1/\alpha} \leqslant \left(\mathbf{M}|\xi|^{\beta}\right)^{1/\beta}.$$

Доказательство. Так как по условию $0 < \alpha < \beta$, то $\frac{\beta}{\alpha} > 1$ и $g(x) = x^{\frac{\beta}{\alpha}}$ — выпуклая функция на $[0, \infty)$. Применяя неравенство Иенсена к случайной величине $\eta = |\xi|^{\alpha}$, находим:

$$\mathbf{M}g\left(|\xi|^{\alpha}\right) \geqslant g\left(\mathbf{M}|\xi|^{\alpha}\right).$$

Но левая часть этого неравенства есть $\mathbf{M}\left(|\xi|^{\alpha}\right)^{\beta/\alpha} = \mathbf{M}\,|\xi|^{\beta}$, а левая равна $\left(\mathbf{M}\,|\xi|^{\alpha}\right)^{\beta/\alpha}$. Следовательно,

$$(\mathbf{M}|\xi|^{\alpha})^{\beta/\alpha} \leqslant \mathbf{M}\,|\xi|^{\beta}.$$

Возводя обе части последнего неравенства в степень $1/\beta$, получаем неравенство Ляпунова.

§ 14. Независимость случайных величин

Определение. Случайные величины $\xi_1, \ldots, \xi_n$, заданные на одном и том же вероятностном пространстве и принимающие значения в измеримых пространствах $(S_1, \Sigma_1), \ldots, (S_n, \Sigma_n)$ соответственно, называются *независимыми в совокупности*, если независимы порожденные ими σ-алгебры $\mathcal{F}_{\xi_1}, \ldots, \mathcal{F}_{\xi_n}$ или, что то же самое, если для любых событий $B_1 \in \Sigma_1, \ldots, B_n \in \Sigma_n$

$$\mathbf{P}\left\{\xi_1 \in B_1, \ldots, \xi_n \in B_n\right\} = \mathbf{P}\left\{\xi_1 \in B_1\right\} \ldots \mathbf{P}\left\{\xi_n \in B_n\right\}.$$

Если множества значений $S_1, S_2, \ldots, S_n$ случайных величин $\xi_1, \ldots, \xi_n$ конечны или счетны, то независимость этих случайных величин эквивалентна выполнению при любых $x_1 \in S_1, \ldots, x_n \in S_n$ условия

$$\mathbf{P}\left\{\xi_1 = x_1, \ldots, \xi_n = x_n\right\} = \mathbf{P}\left\{\xi_1 = x_1\right\} \ldots \mathbf{P}\left\{\xi_n = x_n\right\}.$$

Случайные величины $\xi_1, \ldots, \xi_n$ называются *попарно независимыми*, если при любых $1 \leqslant j < k \leqslant n$ и любых $B_j \in \Sigma_j$, $B_k \in \Sigma_k$

$$\mathbf{P}\{\xi_j \in B_j,\, \xi_k \in B_k\} = \mathbf{P}\{\xi_j \in B_j\}\mathbf{P}\{\xi_k \in B_k\}.$$

В § 6 для задания независимых событий строились произведения вероятностных пространств. В некоторых случаях можно обойтись без такой конструкции.

Пример. Рассмотрим вероятностное пространство $(\Omega,\ \mathcal{F},\ \mathbf{P})$, где $\Omega = [0,1)$, $\mathcal{F}$ — σ-алгебра борелевских подмножеств полуинтервала $[0,1)$, $\mathbf{P}$ — мера Лебега. Любое число $\omega \in [0,1)$ можно записать в двоичной системе счисления:

$$\omega = (0,\ \xi_1(\omega)\xi_2(\omega)\xi_3(\omega)\ldots)_2 = \sum\nolimits_{k\geqslant 1} 2^{-k}\xi_k(\omega), \tag{20}$$

где $\xi_1(\omega)$, $\xi_2(\omega)$, ... принимают значения 0 или 1 и

$$\xi_k\,(\omega) \equiv \left[2^k\omega\right]\ (\operatorname{mod} 2)\,, \quad k = 1, 2, \ldots \tag{21}$$

Для двоично-рациональных чисел вида $\frac{m}{2^n}$, где m и n — целые положительные, а $m = a_1 2^{n-1} + a_2 2^{n-2} + \ldots + a_{n-1}2^1 + 1$ нечетное, возможны два варианта представления (20):

$$\frac{m}{2^n} = \sum_{k=1}^{n-1}\frac{a_k}{2^k} + \frac{1}{2^n} + \sum_{k=n+1}^{\infty}\frac{0}{2^k} = \sum_{k=1}^{n-1}\frac{a_k}{2^k} + \frac{0}{2^n} + \sum_{k=n+1}^{\infty}\frac{1}{2^k};$$

из них обычно рассматривается только первое, что соответствует формуле (21).

Таким образом, на вероятностном пространстве $(\Omega,\ \mathcal{F},\ \mathbf{P})$ с $\Omega = [0,1)$ можно построить бесконечную последовательность $\xi_1,\ \xi_2,\ \ldots$ случайных величин, принимающих значения 0 и 1 и определенных формулой (21).

Покажем, что $\mathbf{P}\{\xi_k = 0\} = \mathbf{P}\{\xi_k = 0\} = 1$ при всех $k = 1, 2, \ldots$ и что $\xi_1, \xi_2, \ldots$ независимы в совокупности.

При любых $n \geqslant 1$ и $a_1, \ldots, a_n \in \{0,\ 1\}$

$$\begin{gathered}\{\omega:\ \xi_1(\omega) = a_1, \ldots, \xi_n(\omega) = a_n\} = \\ = \bigsqcup_{b_{n+1}, b_{n+2}, \ldots \in \{0,1\}} \left\{\sum\nolimits_{k=1}^{n} 2^{-k}a_k + \sum\nolimits_{k=n+1}^{\infty} 2^{-k}b_k\right\} = \\ = \left\{\omega:\ \sum\nolimits_{k=1}^{n} 2^{-k}a_k \leqslant \omega < \sum\nolimits_{k=1}^{n} 2^{-k}a_k + 2^{-n}\right\} = \\ = \left[\sum\nolimits_{k=1}^{n} 2^{-k}a_k \leqslant \omega < \sum\nolimits_{k=1}^{n} 2^{-k}a_k + 2^{-n}\right),\end{gathered}$$

поэтому $\mathbf{P}\{\xi_1 = a_1, \ldots, \xi_n = a_n\} = 2^{-n}$ при любых $a_1, \ldots, a_n \in \{0,\ 1\}$.

Тогда для любого $n \geqslant 1$ и любого $a \in \{0,1\}$ по формуле полной вероятности по значениям $\xi_1, \dots, \xi_{n-1}$

$$\mathbf{P}\{\xi_n = a\} = \mathbf{P}\left\{ \bigsqcup_{b_1,\dots,b_{n-1}\in\{0,1\}} \{\xi_1 = b_1, \dots, \xi_{n-1} = b_{n-1}, \xi_n = a\} \right\} =$$
$$= \sum_{b_1,\dots,b_{n-1}\in\{0,1\}} \mathbf{P}\{\xi_1 = b_1, \dots, \xi_{n-1} = b_{n-1}, \xi_n = a\} = 2^{n-1}\frac{1}{2^n} = \frac{1}{2},$$

т. е. все случайные величины ξ_n принимают значения 0 и 1 с вероятностями $\frac{1}{2}$. Кроме того, эти случайные величины независимы в совокупности. Например, при $1 \leqslant m < n$ и $a_m, a_n \in \{0,1\}$

$$\mathbf{P}\{\xi_m = a_m, \xi_n = a_n\} =$$
$$= \mathbf{P}\left\{ \bigsqcup_{b_m\in\{0,1\}\colon\, m\in\{1,\dots,n\}\setminus\{m,n\}} \{xi_1 = b_1, \dots, \xi_{m-1} = b_{m-1}, \xi_m = a_m,\right.$$
$$\left.\xi_{m+1} = b_{m+1}, \dots, \xi_{n-1} = a_{n-1}, \xi_n = a_n\}\right\} =$$
$$= \sum_{b_m\in\{0,1\}\colon\, m\in\{1,\dots,n\}\setminus\{m,n\}} \mathbf{P}\{\xi_1 = b_1, \dots, \xi_{m-1} = b_{m-1}, \xi_m = a_m,$$
$$\xi_{m+1}=b_{m+1}, \dots, \xi_{n-1}=b_{n-1}, \xi_n = a_n\} = \frac{2^{n-2}}{2^n} = \frac{1}{4} = \mathbf{P}\{\xi_m=a_m\}\mathbf{P}\{\xi_n=a_n\}.$$

В общем случае при любых $1 \leqslant i_1 < i_2 < \ldots < i_k$ и $a_1, \dots, a_k \in \{0,\, 1\}$ по формуле полной вероятности

$$\mathbf{P}\{\xi_{i_j} = a_j\,(j = 1, \dots, k)\} =$$
$$= \sum_{\substack{b_m\in\{0,\,1\}\colon \\ m\in\{1,\dots,i_k\}\setminus\{i_1,\dots,i_k\}}} \mathbf{P}\left\{ \xi_v = \begin{cases} a_j, & v = i_j,\ j = 1, \dots, k, \\ b_m, & v = m \in \{1, \dots, i_k\} \setminus \{i_1, \dots, i_k\} \end{cases} \right\} =$$
$$= \sum_{\substack{b_m\in\{0,\,1\}\colon \\ m\in\{1,\dots,i_k\}\setminus\{i_1,\dots,i_k\}}} \frac{1}{2^{i_k}} = \frac{2^{i_k-k}}{2^{i_k}} = \frac{1}{2^k},$$

т. е. случайные величины $\xi_1, \dots, \xi_n$ независимы при любом n и $\mathbf{P}\{\xi_k = 1\} = \mathbf{P}\{\xi_k = 0\} = \frac{1}{2}$ при любом $k = 1, 2, \dots$

Теорема. *Пусть* $\xi_1, \ldots, \xi_n$ *— независимые случайные величины, определенные на вероятностном пространстве* $(\Omega, \mathcal{F}, \mathbf{P})$ *и принимающие значения соответственно в пространствах* (S_k, Σ_k), $k = 1, \ldots, n$, *а* (Q_k, Θ_k) *— некоторые измеримые пространства. Если для всех* $k = 1, \ldots, n$ *функции* $g_k \colon S_k \to Q_k$ *являются* (Σ_k, Θ_k)-*измеримыми, то*

$$\eta_1(\omega) = g_1(\xi_1(\omega)), \ldots, \eta_n(\omega) = g_n(\xi_n(\omega))$$

– независимые случайные величины.

Доказательство. а) Докажем, что $\eta_1(\omega), \ldots, \eta_n(\omega)$ – случайные величины, т. е. $(\mathcal{F}, \Theta_k)$-измеримые функции, отображающие Ω в $Q_1, \ldots, Q_n$. Действительно, для любого $k \in \{1, \ldots, n\}$ и любого множества $B \in \Theta_k$ имеем:

$$\{\omega : \ \eta_k(\omega) \in B\} = \left\{\omega : \ \xi_k(\omega) \in g_k^{-1}(B)\right\}, \tag{22}$$

где $g_k^{-1}(B)$ — полный прообраз множества B при отображении g_k. В силу (Σ_k, Θ_k)-измеримости функции g_k этот полный прообраз содержится в σ-алгебре Σ_k, и в силу $(\mathcal{F}, \Sigma_k)$-измеримости ξ_k множества (22) $\mathcal{F}$-измеримы при любых $k \in \{1, \ldots, n\}$ и $B \in \Theta_k$. Следовательно, функции $\eta_k \colon \Omega \to Q_k$ измеримы, т. е. являются случайными величинами.

б) Докажем независимость $\eta_1, \ldots, \eta_n$. Пусть $B_1 \in \Theta_1, \ldots, B_n \in \Theta_n$ – измеримые множества (события). Тогда множества

$$A_1 = \{x \in S_1 : \ g_1(x) \in B_1\}, \ldots, A_n = \{x \in S_n : \ g_n(x) \in B_n\}$$

принадлежат σ-алгебрам $\Sigma_1, \ldots, \Sigma_n$ соответственно, и из независимости $\xi_1, \ldots, \xi_n$ следует, что

$$\mathbf{P}\{\eta_1 = g_1(\xi_1) \in B_1, \ldots, \eta_n = g_n(\xi_n) \in B_n\} = \mathbf{P}\{\xi_1 \in A_1, \ldots, \xi_n \in A_n\} = \\ = \mathbf{P}\{\xi_1 \in A_1\} \ldots \mathbf{P}\{\xi_n \in A_n\} = \mathbf{P}\{\eta_1 \in B_1\} \ldots \mathbf{P}\{\eta_n \in B_n\}.$$

Теорема доказана.

§ 15. Мультипликативное свойство математических ожиданий для независимых случайных величин

Теорема. *Если случайные величины* $\xi_1, \ldots, \xi_n$ *определены на одном вероятностном пространстве* $(\Omega, \mathcal{F}, \mathbf{P})$, *принимают значения из конечного или счетного множества действительных чисел, имеют конечные математические ожидания и независимы, то*

$$\mathbf{M}(\xi_1 \ldots \xi_n) = \mathbf{M}\xi_1 \cdot \ldots \cdot \mathbf{M}\xi_n.$$

Доказательство проведем индукцией по числу сомножителей. Основным случаем является база индукции: $n = 2$. Пусть ξ и η – независимые случайные величины; будем обозначать через x_1, x_2, … все возможные значения ξ, через y_1, y_2, … – все возможные значения η. Случайные величины ξ и η порождают разбиения

$$\{A_j = \{\omega \in \Omega : \ \xi(\omega) = x_j\}, \ j \geqslant 1\},$$
$$\{B_k = \{\omega \in \Omega : \ \eta(\omega) = y_k\}, \ k \geqslant 1\}$$

пространства элементарных событий Ω; тогда

$$\xi = \sum_{j\geqslant 1} x_j I(A_j), \quad \mathbf{M}\xi = \sum_{j\geqslant 1} x_j \mathbf{P}\{A_j\},$$
$$\eta = \sum_{k\geqslant 1} y_k I(B_k), \quad \mathbf{M}\eta = \sum_{k\geqslant 1} y_k \mathbf{P}\{B_k\}$$

и

$$\xi\eta = \left(\sum_{j\geqslant 1} x_j I(A_j)\right) \cdot \left(\sum_{k\geqslant 1} y_k I\{B_k\}\right) =$$
$$= \sum_{j\geqslant 1}\sum_{k\geqslant 1} x_j y_k I(A_j) I(B_k) = \sum_{j\geqslant 1}\sum_{k\geqslant 1} x_j y_k I(A_j B_k).$$

Используем почленное умножение абсолютно сходящихся рядов для $\mathbf{M}\xi$ и $\mathbf{M}\eta$, независимостью ξ и η и определением математического ожидания:

$$\mathbf{M}\xi\mathbf{M}\eta = \left(\sum\nolimits_{j\geqslant 1} x_j \mathbf{P}\{A_j\}\right) \left(\sum\nolimits_{k\geqslant 1} y_k \mathbf{P}\{B_k\}\right) =$$
$$= \sum_{j\geqslant 1}\sum_{k\geqslant 1} x_j y_k \mathbf{P}\{A_j\}\mathbf{P}\{B_k\} = \sum_{j\geqslant 1}\sum_{k\geqslant 1} x_j y_k \mathbf{P}\{A_j B_k\} = \mathbf{M}\xi\eta.$$

Если утверждение теоремы верно для $n-1$, то по уже доказанному

$$\mathbf{M}(\xi_1, \ldots, \xi_n) = \mathbf{M}((\xi_1 \ldots \xi_{n-1})\xi_n) = \mathbf{M}(\xi_1 \ldots \xi_{n-1})\mathbf{M}\xi_n = \mathbf{M}\xi_1 \cdot \ldots \cdot \mathbf{M}\xi_n.$$

(Из независимости $\xi_1, \ldots, \xi_n$ следует, что набор случайных величин $(\xi_1, \ldots, \xi_{n-1})$ и случайная величина ξ_n независимы, а произведение $\xi_1 \ldots, \xi_{n-1}$ и ξ_n являются функциями от независимых случайных величин $(\xi_1, \ldots, \xi_{n-1})$ и ξ_n и поэтому независимы.) Теорема доказана.

§ 16. Ковариации. Дисперсия суммы случайных величин

Определение. Пусть ξ и η — случайные величины с действительными значениями и конечными математическими ожиданиями.
Ковариацией случайных величин ξ и η называется величина

$$\operatorname{cov}(\xi, \eta) = \mathbf{M}(\xi - \mathbf{M}\xi)(\eta - \mathbf{M}\eta).$$

Очевидно, $\operatorname{cov}(\xi,\,\eta)=\operatorname{cov}(\eta,\,\xi)$. Кроме того, для ковариации справедлива формула, аналогичная формуле $\mathbf{D}\xi=\mathbf{M}\xi^2-(\mathbf{M}\xi)^2$ для дисперсии:

$$\operatorname{cov}\,(\xi,\,\eta)=\mathbf{M}\xi\eta-\mathbf{M}\xi\,\mathbf{M}\eta.$$

Отметим еще несколько свойств ковариации.

1. Если случайная величина ξ имеет математическое ожидание, то $\operatorname{cov}(\xi,\,\xi)=\mathbf{D}\xi$.

Утверждение следует из определений ковариации и дисперсии:

$$\operatorname{cov}\,(\xi,\,\xi)=\mathbf{M}\,(\xi-\mathbf{M}\xi)\,(\xi-\mathbf{M}\xi)=\mathbf{M}(\xi-\mathbf{M}\xi)^2=\mathbf{D}\xi.$$

2. Для любых случайных величин с конечными математическими ожиданиями

$$|\operatorname{cov}\,(\xi,\,\nu)|\leqslant\sqrt{\mathbf{D}\xi\mathbf{D}\eta}.$$

Доказательство. Нужно доказать, что

$$|\mathbf{M}\,(\xi-\mathbf{M}\xi)\,(\eta-\mathbf{M}\eta)|\leqslant\sqrt{\mathbf{D}\xi\mathbf{D}\eta}=\sqrt{\mathbf{M}\,(\xi-\mathbf{M}\xi)^2\,\mathbf{M}\,(\eta-\mathbf{M}\eta)^2}.$$

Введем центрированные случайные величины $\xi^*=\xi-\mathbf{M}\xi$, $\eta^*=\eta-\mathbf{M}\eta$. Тогда, используя неравенство $|\mathbf{M}\zeta|\leqslant\mathbf{M}\,|\zeta|$ и неравенство Коши–Буняковского, получаем:

$$|\operatorname{cov}(\xi,\,\eta)|=|\mathbf{M}\xi^*\eta^*|\leqslant\sqrt{\mathbf{M}|\xi^*|^2\mathbf{M}|\eta^*|^2}=\sqrt{\mathbf{D}\xi\mathbf{D}\eta}.$$

Определение. Отношение

$$\frac{\operatorname{cov}(\xi,\,\eta)}{\sqrt{\mathbf{D}\xi\mathbf{D}\eta}}=\operatorname{corr}(\xi,\,\eta)$$

называют *корреляцией* случайных величин ξ и η.

Из свойства 2 следует, что корреляция (если она определена, т.е. если $\mathbf{D}\xi$ и $\mathbf{D}\eta$ конечны) по модулю не превосходит 1 и что она имеет такой же знак, как ковариация.

3. *Если случайные величины ξ и η независимы и имеют конечные математические ожидания, то* $\operatorname{cov}(\xi,\,\eta)=0$. *Существуют зависимые случайные величины, ковариация которых равна 0.*

Доказательство. Первое утверждение следует из мультипликативности математического ожидания для произведения независимых случайных величин:

$$\operatorname{cov}(\xi,\, \eta) = \mathbf{M}\left\{(\xi - \mathbf{M}\xi)(\eta - \mathbf{M}\eta)\right\} =$$
$$= \mathbf{M}(\xi - \mathbf{M}\xi)\mathbf{M}(\eta - \mathbf{M}\eta) = (\mathbf{M}\xi - \mathbf{M}\xi)(\mathbf{M}\eta - \mathbf{M}\eta) = 0.$$

Для доказательства второго утверждения достаточно привести пример двух случайных величин, которые не являются независимыми, но имеют нулевую ковариацию. Пусть распределение пары случайных величин $(\xi,\, \eta)$ состоит из 4 атомов веса $1/4$, расположенных в концах единичных векторов координатных осей:

$$\mathbf{P}\{(\xi,\, \eta) = (0,\, 1)\} = \mathbf{P}\{(\xi,\, \eta) = (0,\, 1)\} =$$
$$= \mathbf{P}\{(\xi,\, \eta) = (-1,\, 0)\} = \mathbf{P}\{(\xi,\, \eta) = (0,\, -1)\} =$$
$$= \mathbf{P}\{(\xi,\, \eta) = (-1,\, 0)\} = \mathbf{P}\{(\xi,\, \eta) = (0,\, -1)\} = \frac{1}{4}\,.$$

Тогда

$$\mathbf{P}\{\xi = 1\} = \mathbf{P}\{\eta = 1\} = \mathbf{P}\{\xi = -1\} = \mathbf{P}\{\eta = -1\} = \frac{1}{4}\,,$$
$$\mathbf{P}\{\xi = 0\} = \mathbf{P}\{\eta = 0\} = \frac{1}{2}\,,$$

и

$$0 = \mathbf{P}\{\xi = 0,\, \eta = 0\} \neq \mathbf{P}\{\xi = 0\}\mathbf{P}\{\eta = 0\} = \frac{1}{4}\,,$$

так что случайные величины ξ и η не являются независимыми. Далее, $\mathbf{M}\xi = \mathbf{M}\eta = 0$ и

$$\operatorname{cov}(\xi,\, \eta) = \mathbf{M}\xi\eta = 0,$$

так как $\mathbf{P}\{\xi\eta = 0\} = 1$. Значит, пара зависимых случайных величин ξ, η является искомым примером.

Таким образом, если случайные величины ξ и η независимы и имеют конечные математические ожидания, то их ковариация и корреляция равны 0. С другой стороны, если случайные величины ξ и η линейно связаны: $\eta = a\xi + b$, то $\mathbf{M}\eta = a\mathbf{M}\xi + b$, $\mathbf{D}\eta = a^2\mathbf{D}\xi$ и

$$\operatorname{cov}(\xi,\, \eta) = \mathbf{M}(\xi - \mathbf{M}\xi)(\eta - \mathbf{M}\eta) =$$
$$= \mathbf{M}(\xi - \mathbf{M}\xi)(a\xi + b - (a\mathbf{M}\xi + b)) = a\mathbf{M}(\xi - \mathbf{M}\xi)^2 = a\mathbf{D}\xi,$$
$$\operatorname{corr}(\xi,\, \eta) = \frac{\operatorname{cov}(\xi,\, \eta)}{\sqrt{\mathbf{D}\xi\mathbf{D}\eta}} = \frac{a\mathbf{D}\xi}{\sqrt{\mathbf{D}\xi\, a^2\mathbf{D}\xi}} = \operatorname{sign}(a),$$

т. е. ковариация и корреляция имеют тот же знак, что коэффициент a, и при этом корреляция линейно связанных случайных величин с невырожденными распределениями равна либо 1, либо -1. Значит, ковариация и корреляция в какой-то мере характеризуют степень зависимости между случайными величинами.

Случайные величины, ковариация которых равна 0, называют некоррелированными.

Свойства независимости и некоррелированности одно из другого не следуют. Независимые случайные величины не являются некоррелированными, если они не имеют математического ожидания.

4. Дисперсия суммы случайных величин.

Теорема. *Для любых случайных величин $\xi_1,\, \ldots,\, \xi_n$ с конечными дисперсиями*

$$\mathbf{D}(\xi_1 + \ldots + \xi_n) = \sum_{k=1}^{n} \mathbf{D}\xi_k + 2 \sum_{1 \leqslant k < m \leqslant n} \operatorname{cov}(\xi_k,\, \xi_m).$$

Если случайные величины $\xi_1,\, \ldots,\, \xi_n$ попарно независимы (или хотя бы попарно некоррелированы), то

$$\mathbf{D}(\xi_1 + \ldots + \xi_n) = \mathbf{D}\xi_1 + \ldots + \mathbf{D}\xi_n.$$

Доказательство. Используя определение дисперсии и аддитивность математического ожидания, находим:

$$\mathbf{D}(\xi_1 + \ldots + \xi_n) = \mathbf{M}((\xi_1 + \ldots + \xi_n) - \mathbf{M}(\xi_1 + \ldots + \xi_n))^2 =$$
$$= \mathbf{M}\left(\sum\nolimits_{k=1}^{n} \xi_k - \sum\nolimits_{k=1}^{n} \mathbf{M}\xi_k\right)^2 = \mathbf{M}\left(\sum\nolimits_{k=1}^{n} (\xi_k - \mathbf{M}\xi_k)\right)^2 =$$
$$= \mathbf{M}\left\{\sum_{k=1}^{n} (\xi_k - \mathbf{M}\xi_k)^2 + 2 \sum_{1 \leqslant k < m \leqslant n} (\xi_k - \mathbf{M}\xi_k)(\xi_m - \mathbf{M}\xi_m)\right\} =$$
$$= \sum_{k=1}^{n} \mathbf{D}\xi_k + 2 \sum_{1 \leqslant k < m \leqslant n} \operatorname{cov}(\xi_k, \xi_m).$$

Тем самым доказано первое утверждение.

Для доказательства второго утверждения заметим, что если $\xi_1, \ldots, \xi_n$ попарно независимы, то все слагаемые второй суммы равны 0 как ковариации независимых случайных величин.

5. Ковариации сумм случайных величин.

Теорема. *Для любых случайных величин* $\xi_1, \ldots, \xi_n$ *и* $\eta_1, \ldots, \eta_m$ *с конечными математическими ожиданиями и конечными попарными ковариациями*

$$\mathrm{cov}(\xi_1 + \ldots + \xi_n,\, \eta_1 + \ldots + \eta_m) = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathrm{cov}(\xi_i, \eta_j).$$

Если пары случайных величин $(\xi_1, \eta_1), \ldots, (\xi_n, \eta_n)$ *попарно независимы и у всех случайных величин существуют математические ожидания, то*

$$\mathrm{cov}(\xi_1 + \ldots + \xi_n,\, \eta_1 + \ldots + \eta_n) = \mathrm{cov}(\xi_1, \eta_1) + \ldots + \mathrm{cov}(\xi_n, \eta_n).$$

Доказательство. Используя определение ковариации и аддитивность математического ожидания, находим:

$$\begin{gathered} \mathrm{cov}(\xi_1 + \ldots + \xi_n,\, \eta_1 + \ldots + \eta_m) = \\ = \mathbf{M}(\xi_1 + \ldots + \xi_n - \mathbf{M}(\xi_1 + \ldots + \xi_n))(\eta_1 + \ldots + \eta_m - \mathbf{M}(\eta_1 + \ldots + \eta_m)) = \\ = \mathbf{M}\left(\sum\nolimits_{i=1}^{n}(\xi_i - \mathbf{M}\xi_i)\right)\left(\sum\nolimits_{j=1}^{m}(\eta_j - \mathbf{M}\eta_j)\right) = \qquad (23) \\ = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathbf{M}(\xi_i - \mathbf{M}\xi_i)(\eta_j - \mathbf{M}\eta_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathrm{cov}(\xi_i, \eta_j). \end{gathered}$$

Если пары $(\xi_1, \eta_1), \ldots, (\xi_n, \eta_n)$ попарно независимы, то случайные величины ξ_i и η_j независимы при любых $i \neq j$ (как функции $f(x, y) = x$ и $g(x, y) = y$ от независимых пар (ξ_i, η_i) и (ξ_j, η_j)) и поэтому $\mathrm{cov}(\xi_i, \eta_j) = 0$ при $i \neq j$. Из этого замечания и формулы (23) следует второе утверждение. Теорема доказана.

Следствие (билинейность ковариации). *Для любых случайных величин* $\xi_1, \ldots, \xi_n$ *и* $\eta_1, \ldots, \eta_m$ *с конечными математическими ожиданиями и конечными попарными ковариациями при любых числах* $a_1, \ldots, a_n, b_1, \ldots, b_m \in \mathbb{R}$

$$\mathrm{cov}(a_1\xi_1 + \ldots + a_n\xi_n,\, b_1\eta_1 + \ldots + b_m\eta_m) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \mathrm{cov}(\xi_i, \eta_j).$$

Утверждение следует из (23) и того, что $\mathrm{cov}(a\xi, b\eta) = ab\,\mathrm{cov}(\xi, \eta)$.

6. **Определение.** Если $\bar{\xi} = (\xi_1, \ldots, \xi_d)$ — набор случайных величин, определенных на одном вероятностном пространстве (или, что то же самое, случайный вектор), то матрица

$$B = B(\bar{\xi}) = \|\mathrm{cov}(\xi_i, \xi_j)\| = \begin{vmatrix} \mathbf{D}\xi_1 & \mathrm{cov}(\xi_1, \xi_2) & \ldots & \mathrm{cov}(\xi_1, \xi_d) \\ \mathrm{cov}(\xi_2, \xi_1) & \mathbf{D}\xi_2 & \ldots & \mathrm{cov}(\xi_2, \xi_d) \\ \ldots & \ldots & \ldots & \ldots \\ \mathrm{cov}(\xi_d, \xi_1) & \mathrm{cov}(\xi_d, \xi_2) & \ldots & \mathbf{D}\xi_d \end{vmatrix}$$

называется его *матрицей ковариаций* или *ковариационной матрицей.*

Утверждение. *Матрица ковариаций* $B(\bar{\xi}) = \|\mathrm{cov}\,(\xi_i, \xi_j)\,\|$ *случайного вектора* $\bar{\xi} = (\xi_1, \ldots, \xi_d)$ *симметрична и неотрицательно определена. Более того, матрица ковариаций* $B(\bar{\xi})$ *вырождена* ($\det B = 0$) *тогда и только тогда, когда существуют такие вектор* $x \in \mathbb{R}^d\ \{\bar{0}\}$ *и число* $c \in \mathbb{R}$, *что* $\mathbf{P}\{\sum\limits_{i=1}^{d} x_i\xi_i = c\} = 1$.

Доказательство. Симметричность матрицы ковариаций следует из того, что $\mathrm{cov}(\xi, \eta) = \mathrm{cov}(\eta, \xi)$. Неотрицательная определенность матрицы $B = \|\mathrm{cov}(\xi_i, \xi_j)\|$ означает, что соответствующая ей квадратичная форма неотрицательна:

$$Q_B\,(\bar{x}) = \bar{x}B\bar{x}^T \geqslant 0$$

для любого вектора $\bar{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$ той же размерности, что и порядок матрицы B. Покажем, что матрица $Q_B\,(\bar{x})$ неотрицательно определена:

$$Q_B\,(\bar{x}) = (x_1, \ldots, x_d)\|\mathrm{cov}(\xi_1, \xi_j)\|(x_1, \ldots, x_d)^T =$$
$$= \sum_{i,j=1}^{d} x_i \mathrm{cov}(\xi_i, \xi_j) x_j = \sum_{i,j=1}^{d} x_i x_j \mathbf{M}(\xi_i - \mathbf{M}\xi_i)(\xi_j - \mathbf{M}\xi_j) =$$
$$= \mathbf{M}\sum_{i,j=1}^{d} x_i x_j (\xi_i - \mathbf{M}\xi_i)(\xi_j - \mathbf{M}\xi_j) = \mathbf{M}\left(\sum\nolimits_{i=1}^{d} x_i(\xi_i - \mathbf{M}\xi_i)\right)^2 \geqslant 0$$

при любом векторе $\bar{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$.

Наконец, вырожденность матрицы B означает, что соответствующая ей квадратичная форма Q_B обращается в нуль хотя бы на одном ненулевом векторе $\bar{x}$. Но по свойствам математического ожидания равенство в неравенстве

$$Q_B(\bar{x}) = \mathbf{M}\left(\sum\nolimits_{i=1}^{d} x_i(\xi_i - \mathbf{M}\xi_i)\right)^2 \geqslant 0$$

возможно только в случае, когда

$$\mathbf{P}\left\{\sum\nolimits_{i=1}^{d} x_i(\xi_i - \mathbf{M}\xi_i) = 0\right\} = 1,$$

т. е.

$$\mathbf{P}\left\{\sum\nolimits_{i=1}^{d} x_i\xi_i = \sum\nolimits_{i=1}^{d} x_i\mathbf{M}\xi_i = c\right\} = 1.$$

Утверждение полностью доказано.

7. В п.4 и п.5 было доказано, что дисперсии и ковариации сумм независимых случайных величин равны соответственно суммам дисперсий и ковариаций слагаемых. Аналогичное утверждение верно для ковариационных матриц.

Теорема. *Если случайные d-мерные векторы $\bar{\xi}^{(1)} = (\xi_1^{(1)}, \ldots, \xi_d^{(1)}), \ldots, \bar{\xi}^{(n)} = (\xi_1^{(n)}, \ldots, \xi_d^{(n)})$ независимы, $B(\bar{\xi}^{(1)}), \ldots, B(\bar{\xi}^{(n)})$ — их ковариационные матрицы и $\bar{\Xi} = \bar{\xi}^{(1)} + \ldots + \bar{\xi}^{(n)} = (\Xi_1, \ldots, \Xi_d)$, то*

$$B(\bar{\Xi}) = B(\bar{\xi}^{(1)}) + \ldots + B(\bar{\xi}^{(n)}).$$

Доказательство. Из независимости векторов $\bar{\xi}^{(1)}, \ldots, \bar{\xi}^{(n)}$ по теореме из § 14 следует, что при любых $k, m \in \{1, \ldots, d\}$ их компоненты $\xi_k^{(i)}$ и $\xi_m^{(j)}$ независимы, если $i \neq j$. Для доказательства теоремы рассмотрим произвольный элемент $\mathrm{cov}(\Xi_k, \Xi_m)$ ковариационной матрицы вектора

$$\bar{\Xi} = (\Xi_1, \ldots, \Xi_d) = \left(\sum\nolimits_{i=1}^{n} \xi_1^{(i)}, \ldots, \sum\nolimits_{i=1}^{n} \xi_d^{(i)}\right).$$

Этот элемент имеет вид

$$\mathrm{cov}(\Xi_k, \Xi_m) = \mathrm{cov}(\xi_k^{(1)} + \ldots + \xi_k^{(n)}, \xi_m^{(1)} + \ldots + \xi_m^{(n)}),$$

причем пары $(\xi_k^{(1)}, \xi_m^{(i)}), \ldots, (\xi_k^{(n)}, \xi_m^{(n)})$ независимы. По п.5

$$\mathrm{cov}(\Xi_k, \Xi_m) = \mathrm{cov}(\xi_k^{(1)}, \xi_m^{(1)}) + \ldots + \mathrm{cov}(\xi_k^{(n)}, \xi_m^{(n)}).$$

Таким образом, каждый элемент ковариационной матрицы вектора Ξ равен сумме соответствующих элементов ковариационных матриц векторов $\bar{\xi}^{(1)}, \ldots, \bar{\xi}^{(n)}$. Тем самым теорема доказана.

8. При умножении случайной величины на константу b ее дисперсия умножается на b^2. Обобщение этого свойства на ковариационные матрицы произведений вектора и константы или вектора и матрицы выглядит следующим образом.

Утверждение 2. *Если $\bar{\gamma} = (\gamma_1, \ldots, \gamma_d) \in \mathbb{R}^d$ — случайный вектор, $B(\bar{\gamma})$ — его матрица ковариаций, $b \in \mathbb{R}$, $\bar{a} \in \mathbb{R}^d$, $C = \|c_{ij}\|$ — $(d \times d)$-матрица, то*

$$B(b\bar{\gamma} + \bar{a}) = b^2 B \quad \textit{и} \quad B(C\bar{\gamma} + \bar{a}) = CB(\bar{\gamma})C^T.$$

Доказательство. Первое равенство следует из того, что при любых i, j

$$\mathrm{cov}(b\gamma_i + a_i, b\gamma_j + a_j) = \mathbf{M}(b\gamma_i - \mathbf{M}b\gamma_i)(b\gamma_j - \mathbf{M}b\gamma_j) = b^2\mathrm{cov}(\gamma_i, \gamma_j).$$

Для доказательства второго равенства заметим, что $(C\bar{\gamma})_i = \sum_{j=1}^d c_{ij}\gamma_j$, поэтому в силу билинейности оператора ковариации

$$\mathrm{cov}((C\bar{\gamma} + \bar{a})_r, (C\bar{\gamma} + \bar{a})_s) =$$

$$= \mathbf{M}\left(\sum_{j=1}^d c_{rj}\gamma_j - \mathbf{M}\sum_{j=1}^d c_{rj}\gamma_j\right)\left(\sum_{m=1}^d c_{sm}\gamma_m - \mathbf{M}\sum_{m=1}^d c_{sm}\gamma_m\right) =$$

$$= \sum_{j,m=1}^d \mathbf{M}(c_{rj}\gamma_j - \mathbf{M}c_{rj}\gamma_j)(c_{sm}\gamma_m - \mathbf{M}c_{sm}\gamma_m) =$$

$$= \sum_{j,m=1}^d c_{rj}\mathrm{cov}(\gamma_j, \gamma_m)c_{sm} = (CB(\bar{\gamma})C^T)_{rs},$$

что и требовалось доказать.

§ 17. Закон больших чисел

Теорема (Закон больших чисел). *Пусть случайные величины $\xi_1, \xi_2, \ldots$ попарно независимы или попарно некоррелированы и имеют одинаковые математические ожидания и дисперсии: $\mathbf{M}\xi_t = a$, $\mathbf{D}\xi_t = \sigma^2 < \infty$. Тогда для каждого $\varepsilon > 0$*

$$\mathbf{P}\left\{\left|\frac{1}{n}(\xi_1 + \ldots + \xi_n) - a\right| < \varepsilon\right\} \to 1, \quad n \to \infty.$$

Замечание. Для случайных величин ξ_t, множество значений которых конечно, условие $\mathbf{D}\xi_t < \infty$ выполняется автоматически.

Закон больших чисел в указанной форме является следствием более общего утверждения.

Теорема. *Пусть случайные величины* $\xi_1, \xi_2, \ldots$ *попарно независимы или попарно некоррелированы и* $\sup_t \mathbf{D}\xi_t \leqslant \sigma^2 < \infty$. *Тогда для любого* $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbf{P}\left\{\left|\frac{1}{n}(\xi_1 + \ldots + \xi_n) - \frac{1}{n}(\mathbf{M}\xi_1 + \ldots + \mathbf{M}\xi_n)\right| < \varepsilon\right\} = 1.$$

Доказательство. Положим $\zeta_n = \xi_1 + \ldots + \xi_n$. Тогда

$$\mathbf{M}\zeta_n = \mathbf{M}\xi_1 + \ldots + \mathbf{M}\xi_n, \quad \mathbf{D}\zeta_n = \mathbf{D}\xi_1 + \ldots + \mathbf{D}\xi_n \leqslant n\sigma^2.$$

Следовательно,

$$]\mathbf{M}\left(\tfrac{1}{n}\zeta_n\right) = \tfrac{1}{n}\left(\mathbf{M}\xi_1 + \ldots + \mathbf{M}\xi_n\right),$$
$$\mathbf{D}\left(\tfrac{1}{n}\zeta_n\right) \leqslant \tfrac{1}{n^2}\left(\mathbf{D}\xi_1 + \ldots + \mathbf{D}\xi_n\right) \leqslant \left(\tfrac{1}{n^2}\right) n\sigma^2 = \tfrac{\sigma^2}{n}.$$

Применим к случайной величине ζ_n/n неравенство Чебышева:

$$\mathbf{P}\left\{\left|\tfrac{1}{n}\zeta_n - \mathbf{M}\left(\tfrac{1}{n}\zeta_n\right)\right| \geqslant \varepsilon\right\} \leqslant \frac{\mathbf{D}\left(\tfrac{1}{n}\zeta_n\right)}{\varepsilon^2} \leqslant \tfrac{\sigma^2}{n\varepsilon^2}.$$

Правая часть этого неравенства стремится к 0 при $n \to \infty$. Следовательно, при $n \to \infty$

$$\mathbf{P}\left\{\left|\tfrac{1}{n}\zeta_n - \mathbf{M}\left(\tfrac{1}{n}\zeta_n\right)\right| < \varepsilon\right\} = 1 - \mathbf{P}\left\{\left|\tfrac{1}{n}\zeta_n - \mathbf{M}\left(\tfrac{1}{n}\zeta_n\right)\right| \geqslant \varepsilon\right\} \to 1,$$

что завершает доказательство.

Следствие (Теорема Бернулли). *Пусть проводятся испытания по схеме Бернулли с вероятностью успеха* $p \in (0, 1)$ *и* μ_n *– число успехов в первых* n *испытаниях. Тогда при любом* $\varepsilon > 0$

$$\mathbf{P}\left\{\left|\tfrac{1}{n}\mu_n - p\right| < \varepsilon\right\} \to 1, \quad n \to \infty.$$

Доказательство. Свяжем с k-м испытанием Бернулли индикатор $\mathbb{I}_k$, который равен 1, если испытание оканчивается успехом, и равен 0 в противном случае. Тогда $\mu_n = \mathbb{I}_1 + \ldots + \mathbb{I}_n$, случайные величины $\mathbb{I}_1$, $\mathbb{I}_2$, … независимы в совокупности,

$$\mathbf{P}\{\mathbb{I}_k = 1\} = p_k, \quad \mathbf{P}\{\mathbb{I}_k = 0\} = 1 - p_k, \quad k = 1, 2, \ldots,$$

значит, $\mathbf{M}\mathbb{I}_k = p_k$, $\mathbf{D}\mathbb{I}_k = \mathbf{M}\mathbb{I}_k^2 - (\mathbf{M}\mathbb{I}_k)^2 = p - p^2 = p(1-p) \leqslant \frac{1}{4}$. Поэтому утверждение следствия — частный случай закона больших чисел.

§ 18. Предельные теоремы Пуассона и Муавра–Лапласа

Пусть проведена серия из n независимых испытаний по схеме Бернулли с вероятностью успеха p в каждом испытании. Число μ_n успехов в этой серии испытаний имеет биномиальное распределение с параметрами $(n,\, p)$:

$$\mathbf{P}\{\mu_n = m\} = C_n^m p^m (1-p)^{n-m}, \quad m = 0, 1, \ldots, n.$$

Рассмотрим теперь последовательность серий испытаний: пусть в n-й серии проводится n независимых испытаний Бернулли с вероятностью успеха $p = p(n)$. Следующая теорема указывает условия, при которых распределения μ_n при $n \to \infty$ сходятся к просто описываемому распределению.

Теорема Пуассона. *Если $n \to \infty$ и $p = p(n) \to 0$ так, что $np \to \lambda \in (0,\, \infty)$, то для любого фиксированного $m = 0, 1, 2, \ldots$*

$$\mathbf{P}\{\mu_n = m\} \to \frac{\lambda^m}{m!}\, e^{-\lambda}.$$

Доказательство. Если m фиксировано, то

$$\mathbf{P}\{\mu_n = m\} = C_n^m p^m (1-p)^{n-m} =$$
$$= \left(\frac{(np)^m}{m!}\right) \left(\frac{n}{n(1-p)}\, \frac{n-1}{n(1-p)}\, \frac{n-2}{n(1-p)} \ldots \frac{n-m+1}{n(1-p)}\right) (1-p)^n.$$

Первый сомножитель в правой части при наших условиях стремится к $\frac{\lambda^m}{m!}$, второй – к 1, а логарифм последнего равен $n\ln(1-p) = n(-p + O(p^2)) = -np(1+O(p))$, поэтому последний сомножитель стремится к $e^{-\lambda}$.

Появившиеся в этой теореме распределения на множестве целых неотрицательных чисел

$$\mathbf{P}\,\{\mu = m\} = p_m = \left(\frac{\lambda^m}{m!}\right) e^{-\lambda}, \quad m = 0, 1, 2, \ldots,$$

называются *распределениями Пуассона.* Каждому неотрицательному значению параметра λ соответствует свое распределение Пуассона. Это распределение используется как приближение для распределения числа успехов в длинной серии независимых испытаний Бернулли, когда математическое ожидание λ числа успехов ограничено.

Утверждение. *Если случайная величина μ имеет распределение Пуассона с параметром λ, то $\mathbf{M}\mu = \mathbf{D}\mu = \lambda$.*

Доказательство. Действительно:

$$\mathbf{M}\mu = \sum_{m\geqslant 0} mp_m = \sum_{m\geqslant 0} m\left(\frac{\lambda^m}{m!}\right)e^{-\lambda} = \lambda \sum_{m\geqslant 0}\left(\frac{\lambda^m}{m!}\right)e^{-\lambda} = \lambda.$$

Аналогично,

$$\mathbf{M}\mu^{[2]} = \sum_{m\geqslant 0} m^{[2]}p_m = \sum_{m\geqslant 0} m(m-1)\left(\frac{\lambda^m}{m!}\right)e^{-\lambda} = \lambda^2 \sum_{m\geqslant 0}\left(\frac{\lambda^m}{m!}\right)e^{-\lambda} = \lambda^2$$

и поэтому

$$\mathbf{D}\mu = \mathbf{M}\mu^2 - (\mathbf{M}\mu)^2 = \mathbf{M}\mu(\mu-1) + \mathbf{M}\mu - (\mathbf{M}\mu)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Пусть $\mathbb{I}_1$, $\mathbb{I}_2$, ... — последовательность независимых одинаково распределенных индикаторов,

$$\mathbf{P}\{\mathbb{I}_k = 1\} = p, \quad \mathbf{P}\{\mathbb{I}_k = 0\} = 1 - p = q, \quad k = 1, 2, \ldots$$

Случайная величина $\mu_n = \mathbb{I}_1 + \ldots + \mathbb{I}_n$ имеет биномиальное распределение с параметрами (n, p):

$$\mathbf{P}\{\mu_n = k\} = b_{n,p}(k) = C_n^k p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n,$$

и $\mathbf{M}\mu_n = n\mathbf{M}\mathbb{I}_1 = np$, $\mathbf{D}\mu_n = n\mathbf{D}\mathbb{I}_1 = npq$.

Локальная предельная теорема Муавра–Лапласа (1738–1810 гг.). *Пусть μ_n имеет биномиальное распределение с параметрами (n, p). Тогда если $p = \mathrm{const}$, $p \in (0, 1)$, $q = 1 - p$ и $n, k \to \infty$ так, что $k - np = o(n^{2/3})$, то*

$$\mathbf{P}\{\mu_n = k\} = \frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{(k-np)^2}{2npq}\right\}(1 + o(1)), \tag{24}$$

а при любых соотношениях между n и $k \in \{1, \ldots, n-1\}$

$$\mathbf{P}\{\mu_n = k\} = \frac{1}{\sqrt{2\pi n \frac{k}{n} \frac{n-k}{n}}} \exp\left\{-nH\left(\frac{k}{n}, p\right) + \varepsilon(n, k)\right\}, \tag{25}$$

где $H(x, p) = x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{1-p}$ и $|\varepsilon(n, k)| \leqslant \frac{n}{9k(n-k)}$.

Доказательство.

Лемма. *Функция* $H(x,p)=x\ln\frac{x}{p}+(1-x)\ln\frac{1-x}{1-p}$, $x,p\in(0,\,1)$, *обладает следующими свойствами:*

$$e^{H(x,p)}=\left(\frac{x}{p}\right)^x\left(\frac{1-x}{1-p}\right)^{1-x},\quad H(p,p)=0,\quad H(x,p)\geqslant 0,$$

$$H(x,p)=\frac{(x-p)^2}{2p(1-p)}+O((x-p)^3),\ x\to p. \tag{26}$$

Доказательство леммы. Первые две формулы проверяются непосредственно. Далее, дифференцированием находим, что

$$\frac{\partial}{\partial x}H(x,p)=\ln\frac{x}{p}+\frac{x}{x}-\ln\frac{1-x}{1-p}-\frac{1-x}{1-x}=\ln\frac{x(1-p)}{(1-x)p}\,,$$

$$\frac{\partial^2}{\partial x^2}H(x,p)=\frac{1}{x}+\frac{1}{1-x}=\frac{1}{x(1-x)}\geqslant 4,\quad 0<x<1.$$

Так как при фиксированном $p\in(0,\,1)$ вторая производная $H(x,p)$ по x строго положительна, то $H(x,p)$ строго выпукла вниз, отсюда и из равенств $H(p,p)=\frac{\partial}{\partial x}H(x,p)\big|_{x=p}=0$ следует, что $H(x,p)\geqslant 0$ при всех $x,p\in(0,\,1)$ и что $H(x,p)=0$ только при $x=p$. Далее, применяя формулу Тейлора к функции $H(x,p)$ в окрестности точки p, получаем (26):

$$H(x,p)=\frac{1}{2}\,(x-p)^2H''_{xx}(p,p)+O((x-p)^3)=\frac{(x-p)^2}{2p(1-p)}+O((x-p)^3).$$

Лемма доказана.

Сначала докажем (25), используя лемму и уточненную формулу Стирлинга (см., например, [**?**])

$$n!=\left(\frac{n}{e}\right)^n\sqrt{2\pi n}e^{\theta(n)},\quad \text{где}\quad \frac{1}{12n+1}<\theta(n)<\frac{1}{12n},$$

а из нее выведем (24).

Так как $\mathbf{P}\{\mu_n=k\}=C_n^kp^k(1-p)^{n-k}=\frac{n!p^k(1-p)^{n-k}}{k!(n-k)!}$, то

$$C_n^kp^k(1-p)^{n-k}=\frac{\left(\frac{n}{e}\right)^n\sqrt{2\pi n}e^{\theta(n)}p^k(1-p)^{n-k}}{\left(\frac{k}{e}\right)^k\sqrt{2\pi k}\left(\frac{n-k}{e}\right)^{n-k}\sqrt{2\pi(n-k)}e^{\theta(k)+\theta(n-k)}}=$$

$$=\frac{e^{\varepsilon(n,k)}}{\sqrt{2\pi n\,\frac{k}{n}\,\frac{n-k}{n}}}\left(\frac{p}{k/n}\right)^k\left(\frac{1-p}{(n-k)/n}\right)^{n-k}= \tag{27}$$

$$=\frac{1}{\sqrt{2\pi n\,\frac{k}{n}\,\frac{n-k}{n}}}\,\exp\left\{-nH\left(\tfrac{k}{n}\,,p\right)+\varepsilon(n,k)\right\},$$

где $\varepsilon(n,k) = \theta(n) - \theta(k) - \theta(n-k)$ и $|\varepsilon(n,k)| \leqslant \frac{1}{12}\left(\frac{1}{n} + \frac{1}{k} + \frac{1}{n-k}\right) < \frac{1}{12}\left(\frac{n}{4k(n-k)} + \frac{n}{k(n-k)}\right) < \frac{n}{9k(n-k)}$, так как $(n-2k)^2 \geqslant 0$, т. е. $n^2 \geqslant 4k(n-k)$ при $0 < k < n$ и $\frac{1}{n} < \frac{n}{4k(n-k)}$. Тем самым (25) доказано.

Чтобы доказать равенство (24), достаточно заметить, что если $k - np = o(n^{2/3})$, то $\frac{k}{n} \to p$, $\frac{n-k}{n} \to 1-p = q$, $\varepsilon(n,k) = O(\frac{1}{n})$, и использовать разложение (26):

$$nH(\tfrac{k}{n}, p) = \frac{n}{2}\,\frac{(\frac{k}{n}-p)^2}{p(1-p)} + nO\left(\left(\frac{k}{n}-p\right)^3\right) = \frac{(k-np)^2}{2npq} + O\left(\frac{(k-np)^3}{n^2}\right),$$

в котором при $k - np = o(n^{2/3})$ остаточный член есть $o(1)$. Теорема доказана.

Замечание 2. Функция $\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$, $-\infty < x < \infty$, возникшая в формулировке локальной предельной теоремы Муавра–Лапласа, является плотностью *стандартного нормального распределения.*

Для того чтобы неотрицательная функция была плотностью распределения случайной величины, необходимо и достаточно, чтобы интеграл от нее был равен 1. Докажем, что $\varphi(x)$ — плотность распределения, вычислив квадрат интеграла $\int_{-\infty}^{\infty}\varphi(x)dx$ с помощью перехода к полярным координатам $(r,\alpha) \in [0,\infty) \times [0,2\pi]$, связанным с декартовыми координатами формулами $x = r\cos\alpha$, $y = \sin\alpha$. Якобиан этого преобразования (от полярных к декартовым координатам) равен $\begin{vmatrix} \cos\alpha & \sin\alpha \\ -r\sin\alpha & r\cos\alpha \end{vmatrix} = r$, поэтому на него нужно не делить, а умножать:

$$\left(\int_{-\infty}^{\infty}\varphi(x)\,dx\right)^2 = \left(\int_{-\infty}^{\infty}\varphi(x)\,dx\right)\left(\int_{-\infty}^{\infty}\varphi(y)\,dy\right) =$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}\,dydx = \int_0^{\infty}\int_0^{2\pi}\frac{1}{2\pi}re^{-\frac{r^2}{2}}\,d\alpha dr =$$
$$= \int_0^{\infty} re^{-\frac{r^2}{2}}dr = \int_0^{\infty} e^{-u}du = 1.$$

Функция распределения стандартного нормального закона обозначается $\Phi(x)$:

$$\Phi(x) = \int_{-\infty}^{x}\varphi(u)du = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{u^2}{2}}du.$$

Интегральная предельная теорема Муавра–Лапласа. *Если μ_n имеет биномиальное распределение с параметрами (n,p), $0 < p < 1$,*

$p = \mathrm{const}$, $p \in (0,1)$, $q = 1-p$, *то*

$$\lim_{n\to\infty} \mathbf{P}\left\{\frac{\mu_n - np}{\sqrt{npq}} \leqslant x\right\} = \Phi\left(x\right) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-u^2/2}\, du, \quad -\infty < x < \infty.$$

Доказательство. Покажем, что для любого фиксированного $x \in \mathbb{R}$ для любого $\varepsilon > 0$ при всех достаточно больших n справедливо неравенство

$$\left|\mathbf{P}\left\{\frac{\mu_n - np}{\sqrt{npq}} \leqslant x\right\} - \Phi(x)\right| < \varepsilon.$$

При любом отрицательном $z < x$

$$\left|\mathbf{P}\left\{\frac{\mu_n - np}{\sqrt{npq}} \leqslant x\right\} - \Phi(x)\right| = \tag{28}$$

$$= \left|\left(\mathbf{P}\left\{\frac{\mu_n - np}{\sqrt{npq}} \leqslant z\right\} + \mathbf{P}\left\{z < \frac{\xi_n - np}{\sqrt{npq}} \leqslant x\right\}\right) - (\Phi(z) + (\Phi(x) - \Phi(z)))\right| \leqslant$$

$$\leqslant \mathbf{P}\left\{\frac{\mu_n - np}{\sqrt{npq}} \leqslant z\right\} + \Phi(z) + \left|\mathbf{P}\left\{z < \frac{\xi_n - np}{\sqrt{npq}} \leqslant x\right\} - (\Phi(x) - \Phi(z))\right|.$$

Покажем, что при $n \to \infty$ можно выбрать $z = z(n)$ так, чтобы все слагаемые в правой части (28) стремились к 0.

Так как $\mathbf{M}\mu_n = np$ и $\mathbf{D}\mu_n = npq$, то по неравенству Чебышёва при отрицательном $z < -\frac{2}{\sqrt{\varepsilon}}$

$$\mathbf{P}\left\{\frac{\mu_n - np}{\sqrt{npq}} \leqslant z\right\} = \mathbf{P}\left\{\mu_n - np \leqslant z\sqrt{npq}\right\} \leqslant$$

$$\leqslant \mathbf{P}\left\{|\mu_n - np| > |z|\sqrt{npq}\right\} \leqslant \frac{\mathbf{D}\mu_n}{(z\sqrt{npq})^2} = \frac{1}{z^2} < \frac{\varepsilon}{4}\,.$$

Далее, $\Phi(z) = \int_{-\infty}^{z} \varphi(u)du \to 0$ при $z \to -\infty$, поэтому можно выбрать $z < -\frac{2}{\sqrt{\varepsilon}}$ так, что $\Phi(z) < \varepsilon/4$.

Наконец, по локальной теореме Муавра–Лапласа при фиксированных x и z справедливо равенство, в котором $o(1)$ стремится к 0 равномерно по области суммирования, в которой $m - np = O(\sqrt{n})$:

$$\mathbf{P}\left\{z < \frac{\mu_n - np}{\sqrt{npq}} \leqslant x\right\} =$$

$$= \sum_{m:\ z < \frac{m-np}{\sqrt{npq}} \leqslant x} \mathbf{P}\{\mu_n = m\} = \sum_{m:\ z < \frac{m-np}{\sqrt{npq}} \leqslant x} \frac{1 + o(1)}{\sqrt{2\pi npq}} \exp\left(-\frac{1}{2}\left(\frac{m - np}{\sqrt{npq}}\right)^2\right).$$

Используя обозначения $x_{n,m} = \frac{m-np}{\sqrt{npq}}$, последнее равенство можно представить в виде

$$\mathbf{P}\left\{z < \frac{\mu_n - np}{\sqrt{npq}} \leqslant x\right\} = \frac{1+o(1))}{\sqrt{2\pi}} \sum_{m:\ z<x_{n,m}\leqslant x} (x_{n,m+1} - x_{n,m}) \exp\left\{-\frac{x_{n,m}^2}{2}\right\}.$$

Но правая часть этого равенства представляет собой интегральную сумму для интеграла Римана

$$\int_z^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du = \Phi(x) - \Phi(z),$$

значит,

$$\lim_{n\to\infty} \left|\mathbf{P}\left\{z < \frac{\mu_n - np}{\sqrt{npq}} \leqslant x\right\} - (\Phi(x) - \Phi(z))\right| = 0.$$

Из (28) и полученных оценок следует, что для любых $x \in \mathbb{R}$ и $\varepsilon > 0$ при достаточно больших n

$$\left|\mathbf{P}\left\{\frac{\mu_n - np}{\sqrt{npq}} \leqslant x\right\} - \Phi(x)\right| \leqslant \varepsilon.$$

Тем самым интегральная теорема Муавра–Лапласа доказана.

§ 19. Производящие функции и их свойства

Удобным аналитическим аппаратом для исследования свойств распределений целочисленных неотрицательных случайных величин и доказательства предельных теорем для них являются производящие функции.

Определение. *Производящая функция* $f_\xi(s)$ распределения случайной величины ξ, принимающей целые неотрицательные значения, определяется степенным рядом

$$f_\nu(s) = \mathbf{M}s^\nu = \sum_{r=0}^{\infty} \mathbf{P}\{\nu = r\} s^r.$$

Производящая функция как степенной ряд с неотрицательными коэффициентами, сумма которых равна 1, определена при всех s, $|s| \leqslant 1$, и является аналитической функцией в круге радиуса 1 (а возможно, и в более широкой области).

Отметим несколько простых свойств производящих функций.

1) $f_\nu(1) = 1, f_\nu(0) = \mathbf{P}\{\nu = 0\}$.

2) $f_\nu(s)$ и все ее производные монотонно не убывают на отрезке [0,1] (как сходящиеся степенные ряды с неотрицательными коэффициентами).

3) Производящая функция $f(s)$ случайной величины ξ и распределение ξ однозначно определяют друг друга, так как разложение в степенной ряд единственно.

4) $\frac{d}{ds^k} f_\nu(s)\Big|_{s=1-} = \mathbf{M}\nu^{[k]} = \mathbf{M}\nu(\nu-1)\ldots(\nu-k+1),\ k = 1, 2, \ldots$

Доказательство свойства 3) проводится непосредственно:

$$\frac{d^k}{ds^k} f_\nu(s)\Big|_{s=1-} = \frac{d^k}{ds^k} \sum_{r=0}^{\infty} \mathbf{P}\{\nu = r\} s^r \Big|_{s=1-} =$$

$$= \sum_{r=0}^{\infty} r^{[k]} \mathbf{P}\{\nu = r\} s^{r-k} \Big|_{s=1-} = \sum_{r=0}^{\infty} r^{[k]} \mathbf{P}\{\nu = r\} = \mathbf{M}\nu^{[k]}.$$

Величины $\mathbf{M}\nu^{[k]} = \mathbf{M}\nu(\nu-1)\ldots(\nu-k+1)$ называются *факториальными моментами* целочисленной случайной величины ν.

5) Если $\xi_1, \ldots, \xi_n$ — независимые неотрицательные целочисленные случайные величины, то по мультипликативному свойству математического ожидания

$$\mathbf{M}s^{\xi_1+\ldots+\xi_n} = \mathbf{M}s^{\xi_1} \ldots s^{\xi_n} = \prod_{k=1}^{n} \mathbf{M}s^{\xi_k}.$$

Для многих стандартных дискретных распределений производящие функции имеют простой аналитический вид.

Если ξ — индикатор: $\mathbf{P}\{\xi = 1\} = p$, $\mathbf{P}\{\xi = 0\} = 1 - p$, то

$$\mathbf{M}s^\xi = 1 - p + ps.$$

Если β имеет биномиальное распределение с параметрами (n, p):

$$\mathbf{P}\{\beta = k\} = C_n^k p^k q^{n-k}, \quad k = 0, 1, \ldots, n, \quad p, q > 0, p + q = 1,$$

то β имеет такое же распределение, как сумма $\xi_1 + \ldots + \xi_n$, где $\xi_1, \ldots, \xi_n$ независимых одинаково распределенных индикаторов: $\mathbf{P}\{\xi_k = 1\} = p$, $\mathbf{P}\{\xi_k = 0\} = 1 - p$, $k = 1, \ldots, n$, и в силу свойства 5)

$$\mathbf{M}s^\beta = \sum_{k=0}^{n} \mathbf{P}\{\beta = k\} s^k = \sum_{k=0}^{n} C_n^k p^k q^{n-k} s^k = (1 - p + ps)^n = (\mathbf{M}s^{\xi_1})^n.$$

По свойству 4)

$$\mathbf{M}\beta^{[k]} = \frac{d^k}{ds^k}\,(1-p+ps)^n\Big|_{s=1} = n^{[k]}p^k, \quad k=1,\ldots,n.$$

Если κ имеет распределение Пуассона с параметром λ:

$$\mathbf{P}\{\kappa=k\} = \frac{\lambda^k}{k!}\,e^{-\lambda}, \quad k=0,1,\ldots,$$

то

$$\mathbf{M}s^{\kappa} = \sum_{k=0}^{\infty}\mathbf{P}\{\kappa=k\}s^k = \sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\,e^{-\lambda}s^k = e^{-\lambda(s-1)}.$$

По свойству 4) в этом случае

$$\mathbf{M}\kappa^{[k]} = \frac{d^k}{ds^k}\,e^{-\lambda(s-1)}\Big|_{s=1} = \lambda^k, \quad k=1,2,\ldots$$

Если γ имеет геометрическое распределение с параметром p:

$$\mathbf{P}\{\gamma=k\} = pq^k, \quad k=1,2,\ldots, \quad p,q>0, p+q=1,$$

то

$$\mathbf{M}s^{\gamma} = \sum_{k=1}^{\infty}\mathbf{P}\{\gamma=k\}s^k = \sum_{k=0}^{\infty}pq^{k-1}s^k = \frac{ps}{1-qs}.$$

Так как $\frac{d}{ds}\frac{ps}{1-qs} = \frac{p}{1-qs} + \frac{pqs}{(1-qs)^2} = \frac{p}{(1-qs)^2}$, то по свойству 4) в этом случае при $k \geqslant 1$

$$\mathbf{M}\gamma^{[k]} = \frac{d^k}{ds^k}\,\frac{ps}{1-qs}\Big|_{s=1} = \frac{d^{k-1}}{ds^{k-1}}\,\frac{p}{(1-qs)^2}\Big|_{s=1} = \frac{pq^k}{(1-qs)^{k+1}}\Big|_{s=1} = \left(\frac{q}{p}\right)^k k!$$

В частности,

$$\mathbf{M}\gamma = \frac{1}{p}, \quad \mathbf{D}\gamma = \mathbf{M}\gamma^{[2]} + \mathbf{M}\gamma - (\mathbf{M}\gamma)^2 = 2\left(\frac{q}{p}\right)^2 + \frac{q}{p} - \left(\frac{q}{p}\right)^2 = \frac{q}{p^2}\,.$$

Пусть проводятся независимые однородные испытания Бернулли с вероятностью успеха p и неудачи $q=1-p$. Случайная величина ν_r имеет *отрицательное биномиальное распределение* с параметрами p и r, если она равна порядковому номеру испытания, при котором происходит r-й успех, т. е.

$$\mathbf{P}\{\nu_r=k\} = C_{k-1}^{r-1}p^kq^{n-k}, \quad k=r,r+1,\ldots$$

Чтобы найти производящую функцию отрицательного биномиального распределения, заметим, что промежутки времени между моментами появления соседних успехов независимы и имеют такое же распределение, как ν_1. Поэтому ν_r имеет такое же распределение, как сумма r независимых случайных величин, имеющих такое же распределение, как ν_1:

$$\mathbf{P}\{\nu_1 = k\} = p^k q^{n-k}, \quad k = 1, 2, \ldots$$

Значит, распределение ν_1 получается из геометрического распределения с параметром p и имеет производящую функцию $\mathbf{M}s^{\nu_1} = \frac{ps}{1-qs}$. По свойству 5) производящих функций тогда $\mathbf{M}s^{\nu_r}$ есть r-я степень производящей функции распределения ν_1:

$$\mathbf{M}s^{\nu_r} = \frac{p^r s^r}{(1-qs)^r},$$

т. е. *отрицательной степенью бинома.*

В различных вероятностных задачах возникают суммы зависимых случайных величин, в частности, суммы зависимых индикаторов. Зависимость слагаемых, как правило, усложняет исследование распределений сумм случайных величин. В частности, для вычисления производящих функций сумм зависимых случайных величин приведенное выше свойство 5) неприменимо. Поэтому используются другие подходы: вычисление моментов, оценивание моментов, доказательство предельных теорем. Удобный способ вычисления факториальных моментов сумм зависимых индикаторов дает следующая лемма.

Лемма о факториальных моментах. *Если $\xi = \chi_1 + \ldots + \chi_n$ — сумма случайных индикаторов, то при любом натуральном k*

$$\mathbf{M}\xi^{[k]} = k! \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} \mathbf{P}\{\chi_{i_1} = \chi_{i_2} = \ldots = \chi_{i_k} = 1\}. \tag{29}$$

Доказательство. Очевидно,

$$\mathbf{M}\xi^{[k]} = k!\, \mathbf{M}\frac{\xi^{[k]}}{k!} = k!\mathbf{M}C_\xi^k. \tag{30}$$

Преобразуем сумму в правой части (29), используя свойства аддитивности и формулу полного математического ожидания по разбиению, порожденному случайной величиной $\xi = \chi_1 + \ldots + \chi_n$:

$$\sum_{1\leqslant i_1<\ldots<i_k\leqslant n} \mathbf{P}\{\chi_{i_1}=\ldots=\chi_{i_k}=1\} = \sum_{1\leqslant i_1<\ldots<i_k\leqslant n} \mathbf{M} I\{\chi_{i_1}=\ldots=\chi_{i_k}=1\} =$$

$$= \mathbf{M} \sum_{1\leqslant i_1<\ldots<i_k\leqslant n} \mathbb{I}\{\chi_{i_1} = \ldots = \chi_{i_k} = 1\} = \tag{31}$$

$$= \sum_{m=0}^{n} \mathbf{P}\{\xi=m\} \mathbf{M} \left\{ \sum_{1\leqslant i_1<\ldots<i_k\leqslant n} \mathbb{I}\{\chi_{i_1}=\ldots=\chi_{i_k}=1\} \,\middle|\, \xi=m \right\}.$$

При любом таком $\omega \in \Omega$, что $\xi(\omega) = m$, среди индикаторов $\chi_1(\omega), \ldots, \chi_n(\omega)$ имеется ровно m индикаторов, равных 1, и $n-m$ индикаторов, равных 0, поэтому в сумме по k индексам в условном математическом ожидании имеется ровно C_m^k ненулевых (единичных) слагаемых, значит, она равна C_m^k и

$$\sum_{1\leqslant i_1<\ldots<i_k\leqslant n} \mathbf{P}\{\chi_{i_1}=\ldots=\chi_{i_k}=1\} = \sum_{m=0}^{n} \mathbf{P}\{\xi = m\} C_m^k = \mathbf{M} C_\xi^k.$$

Отсюда, из (31) и (30) следует утверждение леммы.

Применим эту лемму к следующей схеме: пусть по N ячейкам случайно размещаются T частиц, так что каждая частица независимо от остальных попадает в любую из N ячеек с вероятностью $\frac{1}{N}$. Обозначим через $\mu_0(T,N)$ случайную величину, равную числу пустых ячеек (т. е. таких, в которые не попало ни одной частицы).

Следствие 1. *При любом натуральном* k

$$\mathbf{M}\mu_0^{[k]}(T,N) = k!\, C_N^k \left(1 - \frac{k}{N}\right)^T. \tag{32}$$

Доказательство. Пусть $\xi_1, \xi_2, \ldots, \xi_T$ — номера ячеек, в которые попали 1-я, 2-я,..., T-я частицы; эти случайные величины независимы и имеют равномерное распределение на множестве $\{1, \ldots, N\}$. Обозначим через $\eta_j = \sum_{t=1}^{T} I\{\xi_t = j\}$ — число частиц, попавших в j-ю ячейку. Тогда

$$\mu_0(T,N) = I\{\eta_1(T) = 0\} + \ldots + I\{\eta_N(T) = 0\},$$

и согласно лемме о факториальных моментах

$$\mathbf{M}\mu_0^{[k]}(T,N) = k! \sum_{1\leqslant i_1<\ldots<i_k\leqslant N} \mathbf{P}\{\eta_{i_j}(T) = 1,\ j = 1, \ldots, k\}.$$

Остается заметить, что при любых попарно различных $i_1, \ldots, i_k \in \{1, \ldots, N\}$

$$\mathbf{P}\{I\{\eta_{i_j}(T) = 1\},\, j = 1, \ldots, k\} = \mathbf{P}\{\eta_{i_j}(T) = 0, j = 1, \ldots, k\} =$$
$$= \mathbf{P}\{\xi_t \notin \{i_1, \ldots, i_k\},\, t = 1, \ldots, T\} = \left(1 - \frac{k}{N}\right)^T.$$

Лемма о связи вероятности нулевого значения и факториальных моментов. *Для любой целочисленной неотрицательной случайной величины ξ и любого такого целого $r \geqslant 1$, что $\mathbf{M}\xi^r < \infty$, справедливо равенство*

$$\mathbf{P}\{\xi = 0\} = 1 + \sum_{k=1}^{r-1} \frac{(-1)^k}{k!}\mathbf{M}\xi^{[k]} + \frac{(-1)^r}{r!}\,\theta\,\mathbf{M}\xi^{[r]} \quad \text{где } 0 \leqslant \theta \leqslant 1. \qquad (33)$$

Доказательство. Разложим производящую функцию $f(s) = \mathbf{M}s^\xi$ по формуле Тейлора в точке $s = 1$:

$$f(s) = f(1) + \sum_{k=1}^{r-1} \frac{1}{k!}\, f^{(k)}(1)(s-1)^k + \frac{1}{r!}\, f^{(r)}(u_s)(s-1)^r, \qquad (34)$$

где $u_s \in [s, 1]$. Полагая в этом равенстве $s = 0$ и замечая, что

$$f(0) = \mathbf{M}s^\xi|_{s=0} = \mathbf{P}\{\xi = 0\}, \quad 0 \leqslant f^{(k)}(u) \leqslant f^{(k)}(1) = \mathbf{M}\xi^{[k]},\ u \in [0, 1],$$

так как все производные производящей функции монотонно возрастают на отрезке [0, 1], получаем равенство (33). Лемма доказана.

Следствие 2 (Формула включения-исключения и неравенства Бонферрони). *Если $A_1, \ldots, A_N$ — события, определенные на одном и том же вероятностном пространстве, и $\xi = \sum_{k=1}^{N} \mathbb{I}\{A_k\}$ — число одновременно происходящих событий, а $S_k = \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant N} \mathbf{P}\{A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}\} = \frac{1}{k!}\mathbf{M}\xi^{[k]}$, то*

$$\mathbf{P}\{A_1 \cup \ldots \cup A_N\} = \mathbf{P}\{\xi > 0\} = 1 - \mathbf{P}\{\xi = 0\} = \sum_{k=1}^{N} (-1)^{k+1} S_k. \qquad (35)$$

Для любой целочисленной неотрицательной случайной величины ξ неравенства

$$\sum_{k=1}^{2r} \frac{(-1)^{k+1}}{k!}\,\mathbf{M}\xi^{[k]} \leqslant \mathbf{P}\{\xi > 0\} \leqslant \sum_{k=1}^{2r+1} \frac{(-1)^{k+1}}{k!}\,\mathbf{M}\xi^{[k]}, \tag{36}$$

если входящие в эти неравенства факториальные моменты конечны.

Доказательство. При условиях Следствия производящая функция $\mathbf{M}s^{\xi}$ является многочленом степени не выше N. Поэтому формула включения-исключения (35) следует из (33) при $r = N$ и того, что в силу леммы $\mathbf{M}\xi^{[k]} = k!\,S_k$. Неравенства Бонферрони (36) следуют из формулы (33) и из того, что в силу неотрицательности всех производных производящей функции на отрезке [0,1] знаки остаточных членов зависят только от четности r. Следствие доказано.

§ 20. Производящие функции и предельные теоремы

Производящие функции часто используются при доказательстве предельных теорем. Основу для этого дает следующее утверждение.

Теорема непрерывности для производящих функций. *Пусть $f_n(s) = \sum\limits_{k\geqslant 0} p_{n,k}s^k = \mathbf{M}s^{\nu_n}$ $(n = 1, 2, \ldots)$ и $f(s) = \sum\limits_{k\geqslant 0} p_k s^k = \mathbf{M}s^{\nu}$ – производящие функции неотрицательных случайных величин. Условие*

$$p_{n,k} \to p_k \quad \textit{при} \quad n \to \infty \quad \textit{при всех} \quad k \geqslant 0 \tag{37}$$

эквивалентно условию

$$f_n(s) \to f(s) \quad \textit{при} \quad n \to \infty \quad \textit{при всех} \quad s \in [0, 1). \tag{38}$$

Доказательство. Пусть выполнено (37). Тогда для любого $s \in [0, 1)$ и для любого натурального N имеем

$$|f_n(s) - f(s)| \leqslant \sum_{k\geqslant 0} |p_{n,k} - p_k|s^k \leqslant$$

$$\leqslant \sum_{k=0}^{N} |p_{n,k} - p_k| + \sum_{k\geqslant N} s^k \leqslant \sum_{k=0}^{N} |p_{n,k} - p_k| + s^N/(1 - s).$$

Пусть $\varepsilon > 0$ – произвольное малое число. Выберем N так, чтобы выполнялось неравенство $s^N/(1 - s) < \varepsilon/2$, а затем выберем $n_0 < \infty$

так, что $\sum\limits_{0\leqslant k<N}|p_{n,k}-p_k|<\varepsilon/2$ при всех $n>n_0$. Тогда при всех $n>n_0$ выполняется неравенство $|f_n(s)-f(s)|\leqslant\varepsilon$, что и означает выполнение (38).

Пусть теперь выполнено (38). Допустим, что (37) не выполнено хотя бы при одном значении k. Пусть для определенности (37) не выполняется при $k=0$ (в других случаях рассуждения аналогичны).

Каждая из последовательностей $\{p_{n,k},\ n=1,2,\ldots\}$ ограничена, поэтому из них можно выбирать сходящиеся подпоследовательности. Выберем возрастающую последовательность $\{n_0(t),\ t=1,2,\ldots\}$ так, чтобы последовательность $\{p_{n_0(t),0}\}$ сходилась к какому-то неотрицательному числу $q_0\neq p_0$ (из предположения о том, что $p_{n,0}$ не сходится к p_0, следует, что у ограниченной последовательности $p_{n,0}$ существует предельная точка, отличная от p_0). Затем из $\{n_0(t)\}_{t\geqslant1}$ выберем подпоследовательность $\{n_1(t)\}_{t\geqslant1}$ так, чтобы последовательность $\{p_{n_1(t),1}\}$ сходилась к некоторому неотрицательному числу q_1 (при этом, конечно, $p_{n_1(t),0}\to q_0$). Продолжая эту процедуру, получим такую бесконечную последовательность вложенных одна в другую последовательностей $\{n_0(t)\}_{t\geqslant1}\supseteq\{n_1(t)\}_{t\geqslant1}\supseteq\{n_2(t)\}_{t\geqslant1}\ldots$, что $p_{n_k(t),m}\to q_m$ при $t\to\infty$ для любых $k\geqslant0$ и $m\in\{0,1,\ldots,k\}$. Тогда в «диагональной» последовательности $\{n_t(t),\ t=1,2,\ldots\}$ каждый элемент $n_t(t)$ принадлежит всем последовательностям $\{n_k(t)\}, k\leqslant t$. Поэтому выполняется условие

$$p_{n_t(t),k}\to q_k\quad\text{при}\quad t\to\infty\quad\text{для всех}\quad k\geqslant0.$$

По доказанному тогда

$$f_{n_t(t)}(s)\to\sum_{k=0}^{\infty}q_ks^k\quad\text{при}\quad t\to\infty\quad\text{при всех}\quad s\in[0,1),$$

причем $q_0\neq p_0$, и поэтому $\sum_{k\geqslant0}q_ks^k\neq\sum_{k\geqslant0}p_ks^k=f(s)$. Но из условия (38) следует, что $f_{n_t(t)}(s)\to f(s)$ при $t\to\infty$. Полученное противоречие доказывает, что $q_0=p_0$.

В общем случае, предположив, что (37) не выполняется, можно найти значение $K=\min\{k:\ p_{n,k}\not\to p_k\,(n\to\infty)\}$ и повторить проведенные выше рассуждения, начиная с подпоследовательности $p_{n_K(t),K}$, выбрав ее сходящейся к числу $q_k\neq p_k$. Тогда

$$f_{n_t(t)}(s)\to f_q(s)=\sum_{k=0}^{\infty}q_ks^k\quad\text{при}\quad t\to\infty\quad\text{при всех}\quad s\in[0,1),$$

где $q_j = p_j$ при $j < K$ и $q_K \neq p_K$, и поэтому $f_q(s) \neq f(s)$, т. е. $f_{n_t(t)}(s) \not\to f_q(s)$, $t \to \infty$, что противоречит условию (38). Теорема доказана.

Таким образом, если $\nu_1, \nu_2, \ldots$ – последовательность неотрицательных целочисленных случайных величин с производящими функциями $f_n(s) = \mathbf{M}s^{\nu_n}$, то сходимость $f_n(s) \to f(s)$, $0 \leqslant s < 1$, означает, что распределения ν_n сходятся к распределению целочисленной случайной величины с производящей функцией $f(s)$.

В качестве первого примера применения этого свойства докажем предельную теорему Пуассона для сумм независимых *неодинаково распределенных* индикаторов.

Теорема Пуассона. *Если* $\xi_n = \mathbb{I}_1^{(n)} + \ldots + \mathbb{I}_n^{(n)}$ *– последовательность целочисленных случайных величин, представленных в виде сумм независимых индикаторов* $\mathbb{I}_j^{(n)}$, $1 \leqslant j \leqslant n, n = 1, 2, \ldots,$

$$\mathbf{P}\{\mathbb{I}_j^{(n)} = 1\} = p_{nj}, \quad \mathbf{P}\{\mathbb{I}_j^{(n)} = 0\} = 1 - p_{nj},$$

и выполняются условия

$$\max_{1 \leqslant j \leqslant n} p_{nj} \to 0 \ (n \to \infty), \quad \sum_{j=1}^{n} p_{nj} \to \lambda \in (0, \infty),$$

то

$$\mathbf{P}\{\xi_n = k\} \to \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \ldots$$

Доказательство. Так как по условию слагаемые, образующие ξ_n, являются индикаторами и независимы, то

$$\mathbf{M}s_n^{\xi} = \prod_{j=1}^{n} \mathbf{M}s^{\mathbb{I}_j(n)} = \prod_{j=1}^{n} (1 - p_{nj} + p_{nj}s) = \prod_{j=1}^{n} (1 - p_{nj}(1 - s)).$$

Из оценки

$$\sum_{j=1}^{n} (p_{nj})^2 < (\max_{1 \leqslant j \leqslant n} p_{nj}) \sum_{j=1}^{n} p_{nj}$$

следует, что если выполняются условия теоремы, то $\sum_{j=1}^{n} (p_{nj})^2 \to 0$ при $n \to \infty$. Поэтому при любом $s \in [0, 1]$

$$\ln \mathbf{M}s^{\xi_n} = \sum_{j=1}^{n} \ln(1 - p_{nj}(1 - s)) = -\sum_{j=1}^{n} p_{nj}(1 - s) + \sum_{j=1}^{n} (1 - s)^2 O((p_{nj})^2) =$$

$$= -(1 - s)\lambda + o(1) + O\left((1 - s)^2 \max_{1 \leqslant j \leqslant n} p_{nj}\right) \to (s - 1)\lambda, \quad n \to \infty.$$

Следовательно, $\mathbf{M}s^{\xi_n} \to e^{\lambda(s-1)} = \sum\limits_{k\geqslant 0} \frac{\lambda^k}{k!}\, s^k e^{-\lambda}$, и (согласно теореме непрерывности для производящих функций) распределения ξ_n сходятся к распределению Пуассона с параметром λ.

Эквивалентность сходимости производящих функций и сходимости распределений целочисленных случайных величин позволяет сформулировать удобные для проверки условия на моменты случайных величин, достаточные для сходимости последовательностей распределений.

Теорема (метод моментов). *Пусть $\nu_1, \nu_2, \ldots$ – последовательность неотрицательных целочисленных случайных величин, $m_k^{(n)} = \mathbf{M}\nu_n^{[k]}$, $n, k = 1, 2, \ldots$ Если существуют пределы*

$$m_k = \lim_{n\to\infty} m_k^{(n)}, \quad k = 1, 2, \ldots, \tag{39}$$

и

$$m_k = O(k!) \quad \textit{при} \quad k \to \infty, \tag{40}$$

то существует такая случайная величина ν, что

$$\lim_{n\to\infty} \mathbf{P}\{\nu_n = j\} = \mathbf{P}\{\nu = j\}, \quad j = 0, 1, \ldots,$$

и $\mathbf{M}\nu^{[k]} = m_k$, $k = 1, 2, \ldots$

Доказательство. Из условия $m_k = O(k!)$, $k \to \infty$, следует, что ряд $f(s) = 1 + \sum_{k=1}^{\infty} \frac{m_k}{k!}\,(s-1)^k$ сходится при любом $s \in (0, 1]$; при этом $f^{(k)}(1) = m_k$, $k = 1, 2, \ldots$ Покажем, что

$$f_n(s) = \mathbf{M}s^{\nu_n} \to f(s),\ n \to \infty, \quad \text{при любом } s \in [0, 1].$$

Если $s = 1$, то $f_n(1) = f(1) = 1$ при всех n. Фиксируем произвольные $s \in (0, 1]$ и $\varepsilon > 0$. Выберем $N = N(\varepsilon) < \infty$ так, чтобы выполнялось неравенство $\sum_{k\geqslant N} \frac{m_k}{k!}\,(1-s)^k < \frac{\varepsilon}{3}$. Тогда

$$f(s) = 1 + \sum_{k=1}^{N-1} \frac{m_k}{k!}\,(s-1)^k + \theta\,\frac{\varepsilon}{3}, \quad 0 < \theta \leqslant 1. \tag{41}$$

Выпишем формулу Тейлора для $f_n(s)$ с остаточным членом в форме Лагранжа:

$$f_n(s) = 1 + \sum_{k=1}^{N-1} \frac{m_k^{(n)}}{k!}\,(s-1)^k + \theta_n \frac{m_N^{(n)}}{N!}\,(s-1)^N, \tag{42}$$

где $0 < \theta_n \leqslant 1$, так как производные $f_n(s)$ монотонно не убывают на отрезке $[0,\, 1]$ и $f_n^{(k)}(1) = m_k^{(n)}$. При всех достаточно больших n

$$\sum_{k=1}^{N-1} \frac{|m_k^{(n)} - m_k|}{k!}\,(s-1)^k < \frac{\varepsilon}{3} \quad \text{и} \quad \frac{m_N^{(n)}}{N!}\,(s-1)^N < \frac{\varepsilon}{3}.$$

Из этих оценок (41), (42) и произвольности выбора $s \in [0,1)$ и $\varepsilon > 0$ следует, что $f_n(s) \to f(s)$, $n \to \infty$ при любом $s \in (0,\, 1]$. Сходимость при $s = 0$ — следствие непрерывности и монотонности $f_n(s)$ на $[0,\, 1]$.

Остается заметить, что непрерывным пределом последовательности производящих функций (как степенных рядов с неотрицательными коэффициентами) может быть только производящая функция. Последнее утверждение теоремы следует из построения производящей функции $f(s)$. Теорема доказана.

Метод моментов часто используется при доказательстве сходимости к распределению Пуассона, поскольку факториальные моменты случайной величины ν, имеющей распределение Пуассона с параметром λ, имеют особенно простой вид:

$$\mathbf{M}\nu^{[k]} = \frac{d^k}{ds^k}\mathbf{M}s^\nu\Big|_{s=1} = \frac{d^k}{ds^k}e^{\lambda(s-1)}\Big|_{s=1} = \lambda^k e^{\lambda(s-1)}\Big|_{s=1} = \lambda^k.$$

Следствие 3. *Если $\nu_n, n = 1, 2, \ldots,$ — такая последовательность целочисленных неотрицательных случайных величин, что $\mathbf{M}\nu_n^{[k]} \to \lambda^k$ при $n \to \infty$ для любого $k = 1, 2, \ldots,$ то распределения случайных величин ν_n сходятся к распределению Пуассона с параметром λ:*

$$\lim_{n\to\infty} \mathbf{P}\{\nu_n = j\} = \frac{\lambda^j}{j!}\,e^{-\lambda}, \quad j = 0, 1, \ldots$$

Доказательство. При указанных в следствии условиях

$$\mathbf{M}s^\nu = \sum_{k\geqslant 0} \frac{\lambda^k}{k!}\,(s-1)^k = e^{-\lambda(s-1)},$$

т.е. предельная случайная величина ν имеет распределение Пуассона с параметром λ.

В качестве примера применения метода моментов докажем предельную теорему для числа $\mu_0(T, N)$ пустых ячеек в равновероятной схеме независимого размещения T частиц по N ячейкам.

Теорема 2. *Если $T, N \to \infty$ так, что $T = N\ln N - N\ln\lambda + o(N)$, то*

$$\lim_{n\to\infty} \mathbf{P}\{\mu_0(T,N) = k\} = \frac{\lambda^k}{k!}\,e^{-\lambda}, \quad k = 0, 1, \ldots$$

Доказательство. Согласно (32)

$$\mathbf{M}\mu_0^{[m]}(T,N) = N^{[m]}\left(1 - \frac{m}{N}\right)^T = N^{[m]}\exp\left\{T\ln\left(1 - \frac{m}{N}\right)\right\}.$$

При указанных в теореме условиях для каждого фиксированного m

$$\ln N^{[m]} = \ln\left(N^m \prod_{j=0}^{m}\left(1 - \frac{j}{N}\right)\right) = m\ln N + O\left(\frac{1}{N}\right),$$

$$\ln\exp\left(T\ln\left(1 - \frac{m}{N}\right)\right) = -\left(\frac{m}{N} + O\left(\frac{1}{N^2}\right)\right)(N\ln N - N\ln\lambda + o(N)) =$$

$$= -\left(1 + O\left(\frac{1}{N}\right)\right)m(\ln N - \ln\lambda + o(1)),$$

откуда следует, что

$$\ln\mathbf{M}\mu_0^{[m]}(T,N) = m\ln\lambda + O(N^{-1}\ln N) = m\ln\lambda + o(1),$$

т.е.

$$\lim_{N,T\to\infty}\mathbf{M}\mu_0^{[m]}(T,N) = \lambda^m,\ m = 1, 2, \ldots$$

Остается заметить, что $\mathbf{M}\mu_0^{[m]}(T,N) < N^m e^{-mT/N}$, т.е. $\sup\limits_{N,T\to\infty}\mathbf{M}\mu_0^{[m]}(T,N) = O(\lambda^m)$ при $m\to\infty$. Отсюда и из следствия из теоремы 1 вытекает утверждение теоремы 2.

§ 21. Основные виды распределений. Смеси распределений

Случайная величина $\xi = \xi(\omega)$, $\omega \in \Omega$, — это измеримая функция, отображающая пространство элементарных событий $(\Omega, \mathcal{F})$ в пространство $(B, \mathcal{B})$ ее значений. Эта функция переводит меру $\mathbf{P}$ — составную часть вероятностного пространства $(\Omega, \mathcal{F}, \mathbf{P})$ — в меру $\mathbf{P}_\xi$ на подмножествах из σ-алгебры $\mathcal{B}$ по формуле $\mathbf{P}_\xi(C) = \mathbf{P}\{\xi \in C\}$, $C \in \mathcal{B}$.

Разным случайным величинам, принимающим значения в одном и том же пространстве $(B, \mathcal{B})$, могут соответствовать разные меры на множествах из $\mathcal{B}$. Поэтому естественно считать, что на пространстве $(B, \mathcal{B})$

задана некоторая «каноническая» неотрицательная мера μ, с которой можно сравнивать другие меры (скажем, порожденные случайными величинами). Если $(B, \mathcal{B})$ — действительная прямая (или евклидово пространство) с борелевской σ-алгеброй подмножеств, то обычно в качестве μ выбирают меру Лебега; если $(B, \mathcal{B})$ — поверхность сферы, то в качестве μ выбирают меру, соответствующую площади на сфере; если $(B, \mathcal{B})$ — конечное или счетное множество, то в качестве $\mu(C)$ можно выбирать число элементов $|C|$ множества $C \subset B$.

Пример. Пусть $\Omega = [0,\, 1]$ — единичный отрезок с борелевской σ-алгеброй борелевских подмножеств и мерой Лебега. Случайные величины $\xi_1(\omega) = 2 + \omega^2$ и $\xi_2(\omega) = 2 + (1-\omega)^2$ разные как функции на Ω, но одинаково распределены и порождают одну и ту же меру на отрезке $[2,\, 3]$, а случайная величина $\xi_3(\omega) = 2 + \omega^3$ порождает другую меру на том же отрезке.

В множестве распределений случайных величин со значениями в множестве действительных чисел с определенной на нем мерой Лебега μ (в общем случае — со значениями в измеримом пространстве $(B, \mathcal{B})$ с неотрицательной мерой на $\mathcal{B}$, которую мы тоже будем обозначать символом μ) можно выделить три существенно разных класса: дискретные, абсолютно непрерывные относительно меры μ и сингулярные распределения относительно меры μ.

Определение. Распределение случайной величины ξ со значениями в множестве B называется *дискретным*, если существует такое не более чем счетное множество $A = \{a_1,\, a_2, \ldots\} \in B$, что $\mathbf{P}\{\xi \in A\} = 1$. Тогда

$$\mathbf{P}\left\{\xi = a_k\right\} = p_k \geqslant 0, \quad \sum_{k \geqslant 1} p_k = 1.$$

Случайная величина ξ порождает меру, сосредоточенную на счетном множестве A, и элементы a_k этого множества, для которых $p_k > 0$, называют *атомами* распределения ξ, а значения p_k – *весами* этих атомов.

Если B — действительная прямая, то функция распределения $F(x) = \mathbf{P}\{\xi \leqslant x\}$ чисто разрывна в том смысле, что сумма величин ее скачков равна 1.

Если, например, $A = \mathbb{Z}$, то функция распределения $F(x)$ — ступенчатая с разрывами в целых точках. Если множество A атомов распределения случайной величины ξ всюду плотно, то ее функция распределения не имеет интервалов постоянства. Примером такой случайной величины является отношение $(-1)^{\nu_2}\frac{\nu_1}{\nu_2}$, где ν_1 и ν_2 — независимые случайные величины, имеющие геометрическое распределение с параметром $p \in (0,1)$ ($\mathbf{P}\{\nu_1 = k\} = \mathbf{P}\{\nu_2 = k\} = (1-p)p^{k-1}, k = 1, 2, \ldots$): такая случайная величина с положительной вероятностью принимает любое ненудевое рациональное значение, а это множество рациональных чисел всюду плотно на положительной полуоси.

Определение. Распределение случайной величины ξ с действительными значениями называется *абсолютно непрерывным относительно меры Лебега* μ, если существует такая измеримая функция $p(x)$, что

$$F\,(x) = \mathbf{P}\,\{\xi \leqslant x\} = \int\limits_{-\infty}^{x} p\,(x)\mu(dx) \quad \text{при всех} \quad x \in \mathbb{R}.$$

Если случайная величина ξ принимает значения в измеримом пространстве $(B, \mathcal{B})$ с заданной на нем неотрицательной мерой μ, то ее распределение называется *абсолютно непрерывным относительно меры* μ, если существует такая $\mathcal{B}$-измеримая функция $p(x)$, $x \in B$, что для любого множества $A \in \mathcal{B}$

$$\mathbf{P}\{\xi \in A\} = \int\limits_{A} p\,(x)\mu(dx).$$

Функция $p(x)$ называется *плотностью распределения* случайной величины ξ относительно меры μ.

Замечание. В этих определениях в общем случае имеются в виду интегралы не по Риману, а по Лебегу, так как плотность может быть неограниченной функцией.

Если случайная величина ξ принимает действительные значения и ее распределение имеет плотность $p(x)$ относительно меры Лебега, то $\mathbf{P}\{\xi = x\} = 0$ для любого действительного x, $F(x) = \mathbf{P}\{\xi \leqslant x\} = \int_{-\infty}^{x} p(u)du$ и $p(x) = F'(x)$, если $p(x)$ непрерывна в точке x.

Примеры:

— стандартное нормальное распределение имеет функцию распределения

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\,du, \quad -\infty < x < \infty,$$

и плотность распределения

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\,e^{-x^2/2},$$

— равномерное распределение на отрезке $[a, b]$ имеет функцию распределения

$$F(x) = \begin{cases} 0 & \text{при } x < a, \\ \frac{x-a}{b-a} & \text{при } a \leqslant x \leqslant b, \\ 1 & \text{при } x > b, \end{cases}$$

и плотность распределения

$$p(x) = \begin{cases} \frac{1}{b-a} & \text{при } a \leqslant x \leqslant b \\ 0 & \text{в остальных случаях,} \end{cases}$$

— показательное распределение с параметром α имеет функцию распределения

$$F(x) = \begin{cases} 0 & \text{при } x < a, \\ 1 - e^{-\alpha x} & \text{при } x \geqslant 0 \end{cases}$$

и плотность

$$p(x) = \begin{cases} 0 & \text{при } x < a, \\ \alpha e^{-\alpha x} & \text{при } x \geqslant 0. \end{cases}$$

Определение. Распределение случайной величины ξ, принимающей значения в измеримом пространстве $(B, \mathcal{B})$ с неотрицательной мерой μ называется *сингулярным относительно меры* μ, если $\mathbf{P}\{\xi = x\} = 0$ при любом $x \in B$ и для каждого $\varepsilon > 0$ существует такое множество $A \in \mathcal{B}$ с $\mu(A) \leqslant \varepsilon$, что

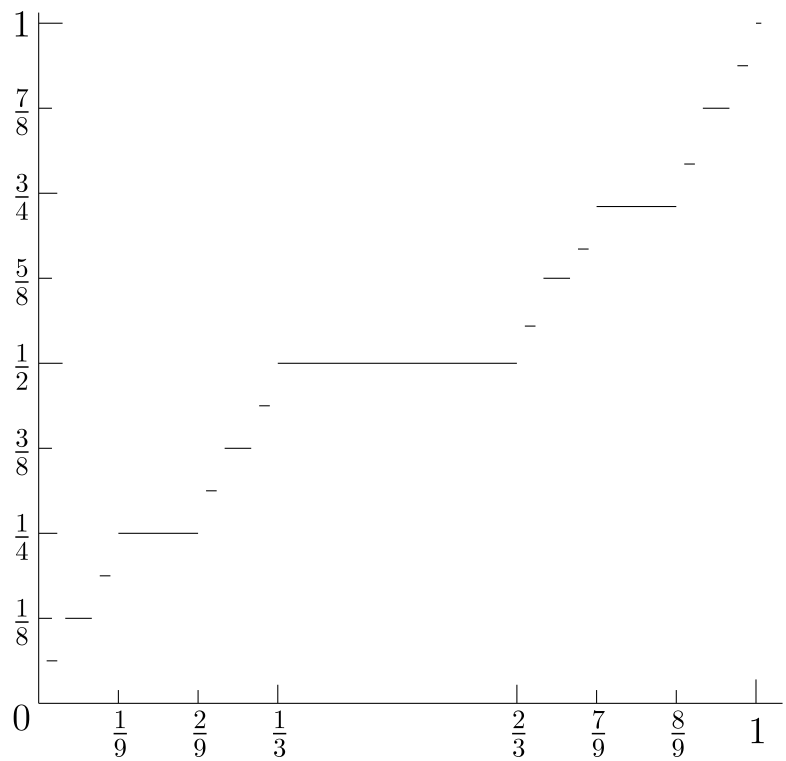
$$\mathbf{P}\left\{\xi \in A\right\} > 1 - \varepsilon.$$

Например, если случайная величина ξ имеет равномерное распределение на отрезке $[0,\ 1]$, то распределение двумерного случайного вектора (ξ, ξ), сосредоточенное на диагонали квадрата $[0,\ 1] \times [0,\ 1]$, сингулярно относительно меры Лебега на плоскости.

Если распределение случайной величины ξ, принимающей действительные значения, сингулярно относительно меры Лебега, то ее функция распределения $F(x) = \mathbf{P}\{\xi \leqslant x\}$ непрерывна, но для любого $\varepsilon > 0$ существует такой набор интервалов $\{(a_k, b_k)\}_{k=1}^{\infty}$ суммарной длины $\sum_{k=1}^{\infty}(b_k - a_k) < \varepsilon$, что

$$\sum_{k=1}^{\infty}(F(b_k) - F(a_k)) > 1 - \varepsilon.$$

Классический пример сингулярного распределения — распределение на канторовом совершенном множестве[2] меры 0 на отрезке $[0, 1]$. Его функция распределения — это «канторова лестница», т. е. непрерывная функция, которая на отрезке $[0,1]$ растет от 0 до 1 и почти всюду имеет нулевую производную. На графике изображено несколько самых длинных участков постоянства непрерывной «канторовой лестницы».



Смесью конечного или счетного множества вероятностных распределений $P_1, P_2, \ldots$ на измеримом пространстве $(B, \mathcal{B})$ называется их выпуклая линейная комбинация, т. е. вероятностное распределение $P^{(\alpha)}$ на $(B, \mathcal{B})$, определенное равенством

$$P^{(\alpha)}(C) = \sum_{j \geqslant 1} \alpha_j P_j(C) \quad \forall C \in \mathcal{B},$$

где $\alpha = (\alpha_1, \alpha_2, \ldots)$ — совокупность неотрицательных чисел и $\alpha_1 + \ldots = 1$.

[2]Канторово совершенное множество получается из отрезка $[0,1]$ в результате последовательного удаления из него средней трети, т. е. интервала $(\frac{1}{3}, \frac{2}{3})$, затем удаления средней трети из каждого из двух оставшихся отрезков, затем удаления средней трети из каждого из четырех оставшихся отрезков, и т. д. до бесконечности. Нетрудно проверить, что сумма длин удаленных интервалов равна 1, поэтому мера Лебега оставшегося множества равна 0.

Например, если P_j — такая мера на $\{0,1,\ldots\}$, что $P_j(j)=1, P_j(m)=0\,\forall m\neq j,\ j\in\{0,1,\ldots\}$, а $\alpha_j=\frac{\lambda^j}{j!}\,e^{-\lambda}, j=0,1,\ldots$, то смесь $\sum_{j\geqslant 0}\alpha_j P_j$ — распределение Пуассона с параметром λ.

Теорема Лебега. *Любое распределение вероятностей $\mathbf{P}$ на измеримом пространстве $(\Omega,\mathcal{F})$ с мерой μ на $\mathcal{F}$ можно представить в виде смеси дискретного, абсолютно непрерывного и сингулярного (относительно меры μ) распределений, т. е. для распределения $\mathbf{P}$ существуют такие дискретное $D_{\mathbf{P}}$, абсолютно непрерывное $C_{\mathbf{P}}$ и сингулярное $S_{\mathbf{P}}$ распределения и неотрицательные константы $\alpha_1,\alpha_2,\alpha_3$, сумма которых равна 1, что для любого множества $A\in\mathcal{F}$*

$$\mathbf{P}\{A\}=\alpha_1 D_{\mathbf{P}}(A)+\alpha_2 C_{\mathbf{P}}(A)+\alpha_3 S_{\mathbf{P}}(A).$$

Доказательство этой теоремы можно найти в учебнике А.А.Боровкова [1].

§ 22. Преобразования случайных величин и их распределений

Опишем правило преобразования распределения действительной случайной величины γ с абсолютно непрерывным распределением при переходе к случайной величине $\kappa=g(\gamma)$, где g — монотонная дифференцируемая функция.

Предложение 1. *Пусть случайная величина γ имеет абсолютно непрерывную функцию распределения $F_\gamma(x)=\boldsymbol{P}\{\gamma\leqslant x\}, x\in\mathbb{R}$, и плотность $p_\gamma(x)=F'_\gamma(x)$. Пусть, далее, $g(x)$ – строго монотонная непрерывно дифференцируемая функция и $\kappa=g(\gamma)$. Тогда распределение случайной величины κ абсолютно непрерывно и имеет плотность $p_\kappa(x)$; плотности $p_\gamma(x)$ и $p_\kappa(x)$ связаны соотношениями*

$$p_\kappa\left(g\left(x\right)\right)=\frac{p_\gamma\left(x\right)}{\left|g'\left(x\right)\right|},\quad p_\kappa\left(x\right)=\left.\frac{p_\gamma(y)}{\left|g'\left(y\right)\right|}\right|_{y=g^{-1}(x)},$$

где $g^{-1}(x)$ – функция, обратная к $g(x)$.

Замечание. В точках, где производная функции g равна 0, плотность распределения κ доопределяется по непрерывности.

Пример. Пусть γ имеет равномерное распределение на отрезке $[0,\,1]$ и плотность $p_\gamma(x)=\begin{cases}1,\ x\in[0,\,1],\\ 0,\ x\notin[0,\,1],\end{cases}$ а $g(x)=2x$. Тогда $g'(x)=$

2, $g(\gamma)$ имеет равномерное распределение на $[0,2]$ и плотность $p_{g(\gamma)}(x) =$ $\begin{cases} 1/2, \ x \in [0,\, 2], \\ 0, \ x \notin [0,\, 2], \end{cases}$ т. е. $p_{g(\gamma)}(x) = \frac{p_\gamma(x/2)}{2} = \frac{p_\gamma(g^{-1}(x))}{g'(g^{-1}(x))}$.

Доказательство. Для определенности рассмотрим случай, когда $g(x)$ строго возрастает. Тогда

$$F_\kappa(x) = \mathbf{P}\left\{\kappa \leqslant x\right\} = \mathbf{P}\left\{g\left(\gamma\right) \leqslant x\right\} = \mathbf{P}\left\{\gamma \leqslant g^{-1}\left(x\right)\right\} = F_\gamma\left(g^{-1}\left(x\right)\right).$$

Если функция распределения дифференцируема, то плотность распределения — это ее производная. Поэтому плотность распределения случайной величины κ можно вычислить по правилу дифференцирования сложной функции:

$$\begin{gathered} p_\kappa\left(x\right) = \tfrac{d}{dx}F_\kappa\left(x\right) = \tfrac{d}{dx}F_\gamma\left(g^{-1}\left(x\right)\right) = \\ = \tfrac{d}{dy}F_\gamma(y)\Big|_{y=g^{-1}(x)} \cdot \tfrac{d}{dx}g^{-1}\left(x\right) = \tfrac{p_\gamma\left(g^{-1}(x)\right)}{g'(y)|_{y=g^{-1}(x)}} \end{gathered}$$

формулу $\frac{d}{dx}g^{-1}\left(x\right) = \frac{1}{g'(y)|_{y=g^{-1}(x)}}$ тоже легко получить по правилу дифференцирования сложной функции:

$$x = g\left(g^{-1}(x)\right),$$

значит,

$$1 = \frac{d}{dx}g\left(g^{-1}\left(x\right)\right) = g'\left(y\right)\big|_{y=g^{-1}(x)} \cdot \frac{d}{dx}g^{-1}\left(x\right).$$

Тем самым доказана вторая формула; первая получается из нее заменой переменных.

Случай, когда $g(x)$ убывает, рассматривается точно так же, отличие состоит только в замене функции распределения ее дополнением:

$$F_\kappa\left(x\right) = \mathbf{P}\left\{\kappa \leqslant x\right\} = \mathbf{P}\left\{g\left(\gamma\right) \leqslant x\right\} = \mathbf{P}\left\{\gamma \geqslant g^{-1}\left(x\right)\right\} = 1 - F_\gamma\left(g^{-1}\left(x\right)\right).$$

Предложение 1 доказано.

Замечание. Если функция $g(x)$ дифференцируема, но не монотонна, то плотность распределения случайной величины $\kappa = g(\gamma)$ можно вычислять по формуле

$$p_\kappa\left(g\left(x\right)\right) = \sum_{y:\ g(y)=g(x)} \frac{p_\gamma(y)}{|g'(y)|},$$

т.е. проводить суммирование по всем прообразам значений κ.

Практически в случаях, когда функция g не монотонна, бывает удобно просто повторять рассуждения, проведенные при доказательстве предложения 1. Например, пусть $g(x) = x^2$ и $\kappa = g(\gamma) = \gamma^2$. Тогда

$$F_\kappa(x) = \mathbf{P}\{\kappa \leqslant x\} = \mathbf{P}\left\{\gamma^2 \leqslant x\right\} = \mathbf{P}\left\{|\gamma| \leqslant \sqrt{x}\right\} = F_\gamma\left(\sqrt{x}\right) - F_\gamma\left(-\sqrt{x}\right)$$

и поэтому по правилу дифференцирования сложной функции

$$p_\kappa(x) = \tfrac{d}{dx}F_\kappa(x) = \tfrac{d}{dx}\left(F_\gamma(\sqrt{x}) - F_\gamma(-\sqrt{x})\right) = \\ = \left(\tfrac{1}{2\sqrt{x}}\right)\left(p_\gamma(\sqrt{x}) + p_\gamma(-\sqrt{x})\right).$$

Важный пример. Если ξ имеет непрерывную функцию распределения $F(x) = \mathbf{P}\{\xi \leqslant x\}$, то случайная величина $\gamma = F(\xi)$ имеет равномерное распределение на отрезке $[0,1]$: при любом $z \in [0,\ 1]$

$$\mathbf{P}\{\gamma \leqslant z\} = \mathbf{P}\{F(\xi) \leqslant z\} = \mathbf{P}\{\xi \leqslant F^{-1}(z)\} = F(F^{-1}(z)) = z.$$

Обратно, если случайная величина γ имеет равномерное распределение на отрезке $[0,1]$, то для того, чтобы получить случайную величину ξ, имеющую заданную непрерывную функцию распределения $F(x)$, достаточно положить $\xi = F^{-1}(\gamma)$, тогда

$$\mathbf{P}\{F^{-1}(\gamma) \leqslant z\} = \mathbf{P}\{F(F^{-1}(\gamma)) \leqslant F(z)\} = \mathbf{P}\{\gamma \leqslant F(z)\} = F(z).$$

Это свойство используется при статистическом моделировании для преобразования значений, порождаемых генератором случайных чисел и рассматриваемых как реализации случайных величин с равномерным распределением на отрезке $[0,1]$, в значения с распределением с заданной функцией распределения $F(x)$.

§ 23. Распределения векторнозначных (многомерных) случайных величин

Пусть на вероятностном пространстве $(\Omega,\ \mathcal{F},\ \mathbf{P})$ задано несколько случайных величин $\xi_1(\omega),\ \ldots,\ \xi_n(\omega):\ \ \Omega \to R$; они определяют новую случайную величину: случайный вектор

$$\xi(\omega) = (\xi_1(\omega), \ldots, \xi_n(\omega))\colon \Omega \to R^n.$$

Случайный вектор ξ порождает на совокупности измеримых по Лебегу подмножеств R^n вероятностную меру (распределение)

$$P(B) = \mathbf{P}\{\xi \in B\}, \quad B \subset \mathbb{R}^n.$$

Распределение случайного вектора $\xi = (\xi_1, \dots, \xi_n)$ называют также *совместным распределением* случайных величин $\xi_1, \dots, \xi_n$, а распределения отдельных величин $\xi_1, \dots, \xi_n$ называются *маргинальными* распределениями этого совместного распределения. Фактически маргинальные распределения являются проекциями совместного распределения в R^n на координатные оси.

Так же, как и для одномерных случайных величин, распределение случайного вектора может быть дискретным (сосредоточенным на счетном множестве точек), абсолютно непрерывным или сингулярным.

Функцией распределения случайного вектора $\xi = (\xi_1, \dots, \xi_n)$ называется

$$F(x_1, \dots, x_n) = \mathbf{P}\{\xi_1 \leqslant x_1, \dots, \xi_n \leqslant x_n\}, \quad x_1, \dots, x_n \in \mathbb{R}.$$

Отметим несколько простых свойств функций распределения случайных векторов.

1) $F_\xi(x_1, \dots, x_n)$ не убывает по каждому аргументу.

Это следует из того, что при любых $x_1, \dots, x_n \in \mathbb{R}$, $i \in \{1, \dots, n\}$ и $\Delta > 0$

$$\{\omega \in \Omega \colon \xi_j(\omega) \leqslant x_j (j = 1, \dots, n)\} \subseteq$$
$$\subseteq \{\omega \in \Omega \colon \xi_i(\omega) \leqslant x_i + \Delta, \xi_j(\omega) \leqslant x_j (j \in \{1, \dots, n\} \setminus \{i\})\}.$$

2) $F_\xi(x_1, \dots, x_n) \to 0$, если $\min\{x_1, \dots, x_n\} \to -\infty$,
$F_\xi(x_1, \dots, x_n) \to 1$, если $\min\{x_1, \dots, x_n\} \to \infty$,
$F_\xi(x_1, \dots, x_n) \to \mathbf{P}\{\xi_i \leqslant x_i\} = F_{\xi_i}(x_i)$, если $x_i = \mathrm{const}$, а $\min_{j \neq i} x_j \to \infty$.

Для доказательства первого соотношения заметим, что если $F_{\xi_j}(x)$ — функция распределения ξ_j, $j = 1, \dots, n$, то $F^*(x) = \max_{1 \leqslant j \leqslant n} F_{\xi_j}(x) \to 0$ при $x \to -\infty$, так как $F_{\xi_j}(x) \to 0$ при $x \to -\infty$ для каждого $j = 1, \dots, n$. Далее, при любом $j \in \{1, \dots, n\}$

$$F_\xi(x_1, \dots, x_n) = \mathbf{P}\{\xi_1 \leqslant x_1, \dots, \xi_n \leqslant x_n\} \leqslant \mathbf{P}\{\xi_j \leqslant x_j\} = F_{\xi_j}(x_j),$$

поэтому

$$F_\xi(x_1, \dots, x_n) \leqslant \min_{1 \leqslant j \leqslant n} F_{\xi_j}(x_j) \leqslant \min_{1 \leqslant j \leqslant n} F^*(x_j) = F^*(\min_{1 \leqslant j \leqslant n} x_j) \to 0$$

при $\min\{x_1, \dots, x_n\} \to -\infty$.

Второе соотношение следует из того, что при $\min\{x_1, \dots, x_n\} \to \infty$

$$\{\omega \in \Omega \colon \xi_1(\omega) \leqslant x_1, \dots, \xi_n(\omega) \leqslant x_n\} \uparrow \Omega.$$

Аналогично доказывается третье соотношение: если $x_i = \text{const}$, а $\min_{j\neq i} x_j \to \infty$, то

$$\{\omega \in \Omega\colon \xi_1(\omega) \leqslant x_1, \ldots, \xi_n(\omega) \leqslant x_n\} \uparrow \{\omega \in \Omega\colon \xi_i \leqslant x_i\}.$$

3) Определим действие операторов $\Delta_h^{(i)}$ на функции $G\,(x_1, \ldots x_n)$ соотношениями

$$\Delta_h^{(i)} G(x_1, \ldots, x_n) = G\,(x_1, \ldots, x_{i-1}, x_i + h, x_{i+1}, \ldots, x_n) - G\,(x_1, \ldots, x_n)\,.$$

Тогда для любых положительных $h_1, \ldots, h_n$

$$\Delta_{h_1}^{(1)} \Delta_{h_2}^{(2)} \ldots \Delta_{h_n}^{(n)} F(x_1, \ldots, x_n) \geqslant 0.$$

При $n = 1$ это неравенство верно:

$$\Delta_{h_1}^{(1)} F(x_1) = F(x_1 + h_1) - F(x_1) = \mathbf{P}\{x_1 < \xi \leqslant x_1 + h_1\} \geqslant 0.$$

В общем случае заметим, что

$$\begin{gathered} \Delta_{h_n}^{(n)} F\,(x_1, \ldots, x_n) = \\ = \mathbf{P}\,\{\xi_1 \leqslant x_1, \ldots, \xi_{n-1} \leqslant x_{n_1},\ \ \xi_n \leqslant x_n + h_n\} - \mathbf{P}\,\{\xi_1 \leqslant x_1, \ldots, \xi_n \leqslant x_n\} = \\ = \mathbf{P}\,\{\xi_1 \leqslant x_1, \ldots, \xi_{n-1} \leqslant x_{n-1}, x_n < \xi_n \leqslant x_n + h_n\}\,. \end{gathered}$$

Аналогично получаем равенство

$$\begin{gathered} \Delta_{h_{n-1}}^{(n-1)} \Delta_{h_n}^{(n)} F\,(x_1, \ldots, x_n) = \\ = \mathbf{P}\,\{\xi_1 \leqslant x_1, \ldots, x_{n-1} < \xi_{n-1} \leqslant x_n + h_{n-1},\ \ x_n < \xi_n \leqslant x_n + h_n\}\,, \end{gathered}$$

и по индукции доказываем, что

$$\begin{gathered} \Delta_{h_1}^{(1)} \Delta_{h_2}^{(2)} \ldots \Delta_{h_n}^{(n)} F\,(x_1, \ldots, x_n) = \\ = \mathbf{P}\,\{x_1 < \xi_1 \leqslant x_1 + h_1, x_2 < \xi_2 \leqslant x_2 + h_2, \ldots,\ \ x_n < \xi_n \leqslant x_n + h_n\} \geqslant 0. \end{gathered}$$

Таким образом, в многомерном случае функция распределения должна удовлетворять ряду дополнительных (по сравнению с одномерным случаем) условий.

Определение. Распределение случайного вектора $\overline{\xi} = (\xi_1, \ldots, \xi_n)$ со значениями в $\mathbb{R}^n$ называется *абсолютно непрерывным* (относительно меры Лебега), если существует такая функция $p(\overline{x})$, $\overline{x} \in \mathbb{R}^n$, что

$$\mathbf{P}\{\overline{\xi} \in B\} = \int\limits_B p(\overline{x})\, d\overline{x} \text{ для любой измеримой области } B \subset \mathbb{R}^n.$$

Функция $p(\overline{x}) = p(x_1, \ldots, x_n)$ называется *плотностью* распределения случайного вектора $\overline{\xi}$ относительно меры Лебега в $\mathbb{R}^n$. Плотность $p(\overline{x})$ неотрицательна и удовлетворяет условию $\int\limits_{\mathbb{R}^n} p(\overline{x})\, d\overline{x} = 1$.

Из определения плотности следует формула для функции распределения:

$$F(x_1, \ldots, x_n) = \mathbf{P}\{\xi_1 \leqslant x_1, \ldots, \xi_n \leqslant x_n\} = \int\limits_{-\infty}^{x_1} \ldots \int\limits_{-\infty}^{x_n} p\left(\overline{x}\right)\, dx_n \ldots dx_1.$$

Если производная $\frac{\partial^n}{\partial x_1 \ldots \partial x_n} F(\overline{x})$ в точке $\overline{x}$ существует и непрерывна, то

$$\frac{\partial^n}{\partial x_1 \ldots \partial x_n} F(\overline{x}) = p(\overline{x}). \tag{43}$$

Замечание. Из существования и непрерывности плотности равенство (43), вообще говоря, не следует. Например, если ξ имеет равномерное распределение в области

$$S = \left\{(x, y)\colon\ -1 \leqslant x \leqslant 1,\ y^2 \leqslant \frac{1}{|x|}\right\}$$

(плотность $p_\xi(x, y)$ распределения ξ равна $1/8$ в области S и равна 0 вне S), то функция распределения ξ в области $|x|\,,\ |y| \leqslant 1$ имеет вид

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p_\xi(u, v)\, du\, dv = \left(\frac{1}{8}\right) \left(2 + \frac{2x}{\sqrt{|x|}} + y\,(x+1)\right).$$

Плотность распределения в точке $(0, 0)$ равна $1/8$ и непрерывна, но выполнение равенства (43) при $x = y = 0$ зависит от способа вычисления $\frac{\partial^2}{\partial x \partial y} F(x, y)$: формула

$$\frac{\partial^2}{\partial x \partial y} F(x, y) = \frac{\partial}{\partial y} \left(\frac{\partial}{\partial x} F(x, y)\right)$$

при $x = 0$ неприменима, так как при $x = 0$ частная производная по x равна ∞ при любом y; с другой стороны, формула

$$\frac{\partial^2}{\partial x \partial y} F(x, y) = \frac{\partial}{\partial x} \left(\frac{\partial}{\partial y} F(x, y)\right)$$

при $x = y = 0$ дает значение плотности $\frac{1}{8}$. Пример легко модифицировать так, что оба способа вычисления смешанной производной окажутся неприменимыми.

Этот пример показывает также, что возможны случаи, когда совместное распределение двух случайных величин имеет ограниченную плотность, а плотность маргинального распределения не ограничена.

В предыдущем параграфе было показано, что если случайная величина ξ имеет абсолютно непрерывное распределение на $\mathbb{R}$ с функцией распределения $F_\xi(x) = \mathbf{P}\{\xi \leqslant x\}$ и плотностью $f_\xi(x) = F'_\xi(x)$, а $g(x)$ — монотонная дифференцируемая функция, то случайная величина $g(\xi)$ имеет плотность распределения

$$f_{g(\xi)}(x) = \frac{f_\xi(g^{-1}(x))}{g'(g^{-1}(x))}\,.$$

Получим аналогичную формулу для преобразования плотности распределения в многомерном случае.

Утверждение 1. *Если $\overline{\gamma}$ — случайный вектор со значениями в $\mathbb{R}^d$ и плотностью распределения $p_{\overline{\gamma}}(\overline{x}), \overline{x} \in \mathbb{R}^d$, а $\overline{g}(\overline{x}) = (g_1(\overline{x}), \ldots, g_d(\overline{x}))$ — непрерывно дифференцируемое обратимое отображение $\mathbb{R}^d \to \mathbb{R}^d$, то случайный вектор $\overline{\kappa} = \overline{g}(\overline{\gamma})$ имеет плотность распределения*

$$p_{\overline{\kappa}}(\overline{x}) = \frac{p_{\overline{\gamma}}(g^{-1}(\overline{x}))}{|\det G(\overline{g}^{-1}(\overline{x}))|}, \tag{44}$$

где $G(\overline{z}) = \left\| \frac{\partial g_i}{\partial x_j}\Big|_{\overline{x}=\overline{z}} \right\|_{i,j=1}^{d}$.

Доказательство. По определению плотность распределения вектора $\overline{\kappa} = \overline{g}(\overline{\gamma})$ — это такая функция $p_{\overline{\kappa}}(\overline{x})$, что

$$\mathbf{P}\{\overline{\kappa} \in B\} = \int\limits_B p_{\overline{\kappa}}(\overline{y})d\overline{y} \quad \text{для любой измеримой области } B \subset \mathbb{R}^d.$$

Но в силу взаимной однозначности отображения $\overline{g}$

$$\mathbf{P}\{\overline{\kappa} \in B\} = \mathbf{P}\{\overline{\gamma} \in \overline{g}^{-1}(B)\} = \int\limits_{\overline{g}^{-1}(B)} p_{\overline{\gamma}}(\overline{x})d\overline{x}.$$

Сделаем в последнем интеграле замену переменных $\overline{y} = \overline{g}(\overline{x})$. Из курса математического анализа известно, что при этом (по аналогии с одномерным случаем) $d\overline{y} = |\det G(\overline{x})|d\overline{x}$, где

$$\det G(\overline{z}) = \left\| \frac{\partial g_i}{\partial x_j}\Big|_{\overline{x}=\overline{z}} \right\|_{i,j=1}^{d}$$

— якобиан отображения $\overline{g}$. Поэтому для любой измеримой области $B \in \mathbf{R}^d$

$$\mathbf{P}\{\overline{\kappa} \in B\} = \int\limits_{\overline{g}^{-1}(B)} p_{\overline{\gamma}}(\overline{x})d\overline{x} = \int\limits_{B} \frac{p_{\overline{\gamma}}\left(\overline{g}^{-1}\left(\overline{y}\right)\right)}{|\det G(\overline{x})|}\,d\overline{y} = \int\limits_{B} \frac{p_{\overline{\gamma}}\left(\overline{g}^{-1}\left(\overline{y}\right)\right)}{|\det G(\overline{g}^{-1}(\overline{y}))|}\,d\overline{y},$$

т. е. действительно $p_{\overline{\kappa}}(\overline{y}) = \frac{p_{\overline{\gamma}}(\overline{g}^{-1}(\overline{y}))}{|\det G(\overline{g}^{-1}(\overline{y}))|}$. Тем самым утверждение доказано.

§ 24. Независимые случайные величины и распределение их суммы

Определение. Случайные величины $\gamma_1, \ldots, \gamma_n$ *независимы в совокупности*, если

$$\mathbf{P}\{\gamma_1 \in B_1, \ldots, \gamma_n \in B_n\} = \mathbf{P}\{\gamma_1 \in B_1\} \ldots \mathbf{P}\{\gamma_n \in B_n\}$$

для любых измеримых множеств $B_1, \ldots, B_n$. Случайные величины $\gamma_1, \ldots, \gamma_n$ *попарно независимы*, если

$$\mathbf{P}\{\gamma_i \in B_i, \gamma_j \in B_j\} = \mathbf{P}\{\gamma_i \in B_i\}\mathbf{P}\{\gamma_j \in B_j\}$$

для любых $1 \leqslant i < j \leqslant n$ и любых измеримых множеств $B_1, \ldots, B_n$.

Обычно независимость случайных величин понимается как независимость в совокупности.

Поэтому функция распределения случайного вектора $\gamma = (\gamma_1, \ldots, \gamma_n)$ с независимыми в совокупности компонентами является произведением функций распределения отдельных компонент:

$$F_\gamma(x_1, \ldots, x_n) = \mathbf{P}\{\gamma_1 \leqslant x_1, \ldots, \gamma_n \leqslant x_n\} =$$
$$= \mathbf{P}\{\gamma_1 \leqslant x_1\} \ldots \mathbf{P}\{\gamma_n \leqslant x_n\} = F_{\gamma_1}(x_1) \ldots F_{\gamma_n}(x_n),$$

а если распределения случайных величин $\gamma_1, \ldots, \gamma_n$ абсолютно непрерывны, то и распределение вектора $\gamma = (\gamma_1, \ldots, \gamma_n)$ с независимыми компонентами имеет плотность, равную произведению плотностей:

$$p_\gamma\left(x_1, \ldots, x_n\right) = \frac{\partial^n}{\partial x_1 \ldots x_n} \prod_{k=1}^{n} F_{\gamma_k}\left(x_k\right) = \prod_{k=1}^{n} \frac{\partial}{\partial x_k} F_{\gamma_k}\left(x_k\right) = \prod_{k=1}^{n} p_{\gamma_k}\left(x_k\right).$$

Если случайные величины ξ и η независимы и принимают только целые значения:

$$\mathbf{P}\left\{\xi = k\right\} = p_k, \quad \mathbf{P}\left\{\eta = k\right\} = r_k, \quad k = 0,\, \pm 1,\, \ldots,$$

то по формуле полной вероятности

$$\begin{gathered}\mathbf{P}\left\{\xi + \eta = n\right\} = \sum_k \mathbf{P}\left\{\xi = k,\ \ \eta = n - k\right\} = \\ = \sum_k \mathbf{P}\left\{\xi = k\right\}\mathbf{P}\left\{\eta = n - k\right\} = \sum_k p_k r_{n-k}.\end{gathered}$$

Если случайные величины γ и κ имеют плотность совместного распределения $p(x, y)$, то функция распределения суммы $\gamma + \kappa$ вычисляется по формуле

$$F_{\gamma+\kappa}\left(u\right) = \mathbf{P}\left\{\gamma + \kappa \leqslant u\right\} = \int\limits_{x+y \leqslant u} p\left(x, y\right) dxdy.$$

Если случайные величины γ и κ независимы и имеют абсолютно непрерывные распределения с плотностями $p_\gamma(x)$ и $p_\kappa(y)$, то эту формулу можно упростить:

$$\begin{gathered}F_{\gamma+\kappa}\left(u\right) = \int\limits_{x+y \leqslant u} p\left(x, y\right) dxdy = \int\limits_{x+y \leqslant u} p_\gamma\left(x\right) p_\kappa\left(y\right) dxdy = \\ = \int\limits_R p_\gamma\left(x\right) dx \int\limits_{-\infty}^{u-x} p_\kappa\left(y\right) dy = \int\limits_R p_\gamma\left(x\right) F_\kappa\left(u - x\right) dx.\end{gathered}$$

Дифференцируя обе части этого равенства по u, находим формулу для плотности распределения суммы $\gamma + \kappa$:

$$\begin{gathered}p_{\gamma+\kappa}\left(u\right) = \tfrac{\partial}{\partial u} F_{\gamma+\kappa}\left(u\right) = \tfrac{\partial}{\partial u} \int\limits_R p_\gamma\left(x\right) F_\kappa\left(u - x\right) dx = \\ = \int\limits_R p_\gamma\left(x\right) \tfrac{\partial}{\partial u} F_\kappa\left(u - x\right) dx = \int\limits_R p_\gamma\left(x\right) p_\kappa\left(u - x\right) dx.\end{gathered}$$

Таким образом, плотность распределения суммы независимых случайных величин является сверткой плотностей распределения слагаемых.

§ 25. Математическое ожидание случайной величины с произвольным распределением

Случайные величины, принимающие лишь конечное число разных значений, будем называть *простыми.*

Для простой (в этом смысле) случайной величины $\xi = \xi(\omega)$, принимающей значения из конечного множества $\{x_1, \dots, x_n\}$ с вероятностями $p_1, \dots, p_n$, математическое ожидание было определено равенством

$$\mathbf{M}\xi = \sum_{k=1}^{n} x_k\, p_k = \sum_{k=1}^{n} x_k\, \mathbf{P}\{\xi = x_k\}.$$

При распространении понятия математического ожидания на произвольные случайные величины, принимающие действительные значения, оно определяется сначала для неотрицательных случайных величин, а потом — для произвольных по следующей схеме.

Определение. Если $\xi = \xi(\omega)$ — неотрицательная случайная величина и $\xi_1(\omega), \xi_2(\omega), \dots$ — неубывающая последовательность простых случайных величин, сходящаяся к $\xi(\omega)$:

$$\xi_n(\omega) \uparrow \xi(\omega) \text{ при } n \uparrow \infty \text{ для каждого } \omega \in \Omega,$$

то

$$\mathbf{M}\xi = \lim_{n\to\infty} \mathbf{M}\xi_n.$$

Если $\xi = \xi(\omega)$ — действительная случайная величина и $\xi = \xi_+ - \xi_-$, где $\xi_+ = \max(\xi,\, 0)$, $\xi_- = \max(-\xi,\, 0)$ — неотрицательные случайные величины, то при $\mathbf{M}\xi_+ < \infty$ и $\mathbf{M}\xi_- < \infty$ полагают $\mathbf{M}\xi = \mathbf{M}\xi_+ - \mathbf{M}\xi_-$; если же $\max\{\mathbf{M}\xi_+, \mathbf{M}\xi_-\} = \infty$, то считается, что случайная величина ξ не имеет математического ожидания.

Построить монотонно не убывающую последовательность случайных величин, сходящуюся к неотрицательной случайной величине ξ, можно по формулам

$$\xi_n = \min\left\{n,\, [2^n\xi]\,/2^n\right\}, \quad n = 1, 2, \dots,$$

где $[x]$ обозначает целую часть числа x. Монотонное неубывание этой последовательности следует из того, что $[2^{n+1}x] \geqslant 2[2^n x]$ при любом $x \geqslant 0$, а сходимость $\xi_n \to \xi$ при $n \to \infty$ — из того, что $0 \leqslant \xi - \xi_n < 1/2^n$.

Корректность определения 1 математического ожидания обосновывается следующим утверждением.

Теорема. *Если* $\xi = \xi(\omega)$ — *неотрицательная случайная величина, а* $\xi_n = \xi_n(\omega)$ *и* $\eta_n = \eta_n(\omega)$ — *неубывающие последовательности простых случайных величин, сходящиеся к* $\xi(\omega)$ *при каждом* $\omega \in \Omega$, *то*

$$\lim_{n\to\infty} \mathbf{M}\xi_n = \lim_{n\to\infty} \mathbf{M}\eta_n.$$

Доказательство.

Лемма. *Пусть* $\eta, \xi_1, \xi_2, \ldots$ — *простые неотрицательные случайные величины,* $\xi_n \uparrow \xi \geqslant \eta, n \uparrow \infty$. *Тогда* $\lim\limits_n \mathbf{M}\xi_n \geqslant \mathbf{M}\eta$.

Доказательство леммы. Фиксируем произвольно $\varepsilon > 0$. События $A_n = \{\omega : \xi_n(\omega) \geqslant \eta(\omega) - \varepsilon\}$ монотонно не убывают по n и в силу условия леммы $\overline{A_n} \downarrow \varnothing$, значит (по теореме о непрерывности вероятностной меры), $\mathbf{P}\left\{\overline{A_n}\right\} \to 0$, $\mathbf{P}\left\{A_n\right\} \to 1$, $n \to \infty$. Далее, так как $\mathbb{I}_{A_n} = 1 - \mathbb{I}_{\overline{A_n}}$, то

$$\xi_n \geqslant \xi_n \mathbb{I}_{A_n} \geqslant (\eta - \varepsilon)\mathbb{I}_{A_n} = \eta - \eta\mathbb{I}_{\overline{A_n}} - \varepsilon\mathbb{I}_{A_n}.$$

По условию, η — простая случайная величина, поэтому существует такое $z < \infty$, что $\mathbf{P}\{\eta < z\} = 1$. Тогда по свойствам математических ожиданий простых случайных величин

$$\mathbf{M}\xi_n \geqslant \mathbf{M}\left(\eta - \eta\mathbb{I}_{\overline{A_n}} - \varepsilon\mathbb{I}_{A_n}\right) \geqslant$$
$$\geqslant \mathbf{M}\eta - z\mathbf{P}\{\overline{A_n}\} - \varepsilon\mathbf{P}\{A_n\} \geqslant \mathbf{M}\eta - z\mathbf{P}\{\overline{A_n}\} - \varepsilon.$$

Переходя к пределу по $n \to \infty$, получаем:

$$\lim_{n\to\infty} \mathbf{M}\xi_n \geqslant \mathbf{M}\eta - \varepsilon \text{ при любом } \varepsilon > 0,$$

что эквивалентно утверждению леммы.

Перейдем к доказательству теоремы.

Для каждого $m < \infty$ имеем по условию $\xi_n \uparrow \xi \geqslant \eta_m$, $n \uparrow \infty$, и в силу леммы

$$\lim_n \mathbf{M}\xi_n \geqslant \mathbf{M}\eta_m \text{ при любом } m = 1, 2, \ldots;$$

переходя к пределу по $m \to \infty$, получаем:

$$\lim_{n\to\infty} \mathbf{M}\xi_n \geqslant \lim_{m\to\infty} \mathbf{M}\eta_m.$$

В силу симметрии $\lim\limits_{m\to\infty} \mathbf{M}\eta_m \geqslant \lim\limits_{n\to\infty} \mathbf{M}\xi_n$. Следовательно,

$$\lim_n \mathbf{M}\xi_n = \lim_n \mathbf{M}\eta_n,$$

что и требовалось доказать.

Расширение определения математического ожидания позволяет использовать, например, понятия моментов, дисперсий и т.п. для случайных величин с произвольными распределениями на действительной прямой. При переходах, описанных в пп. а) и б), сохраняются все установленные ранее свойства математических ожиданий (в частности, монотонность и линейность) и соответствующие этим свойствам формулы. Доказательства этого проводятся по одной и той же схеме. В качестве примера покажем, что для независимых случайных величин определенное с помощью предельного перехода математическое ожидание обладает свойством мультипликативности.

Теорема. *Если ξ и η — независимые случайные величины с конечными математическими ожиданиями, то*

$$\mathbf{M}\xi\eta = \mathbf{M}\xi\mathbf{M}\eta. \tag{45}$$

Доказательство. а) Для простых независимых случайных величин равенство (45) уже доказано.

б) Пусть ξ и η — независимые неотрицательные случайные величины. Введем монотонно не убывающие последовательности простых случайных величин

$$\xi_n = \min\{n,\, [2^n\xi]/2^n\} \uparrow \xi, \quad \eta_n = \min\{n,\, [2^n\eta]/2^n\} \uparrow \eta, \quad n \uparrow \infty.$$

При каждом n простые случайные величины ξ_n и η_n независимы как функции от независимых случайных величин ξ и η. Последовательность $\xi_n\eta_n$ состоит из простых случайных величин, монотонно не убывает и сходится к случайной величине $\xi\eta$. Из соотношений

$$\mathbf{M}\xi_n \to \mathbf{M}\xi, \quad \mathbf{M}\eta_n \to \mathbf{M}\eta, \quad \mathbf{M}\xi_n\eta_n \to \mathbf{M}\xi\eta, \quad n \to \infty,$$

и из того, что $\mathbf{M}\xi_n\eta_n = .\mathbf{M}\xi_n\mathbf{M}\eta_n$, следует, что

$$\mathbf{M}\xi\eta = \lim_{n\to\infty}\mathbf{M}\xi_n\eta_n = \lim_{n\to\infty}\mathbf{M}\xi_n\mathbf{M}\eta_n = \lim_{n\to\infty}\mathbf{M}\xi_n \lim_{n\to\infty}\mathbf{M}\eta_n = \mathbf{M}\xi\mathbf{M}\eta.$$

в) Пусть ξ и η — независимые случайные величины, принимающие как положительные, так и отрицательные значения. Тогда

$$\xi = \xi^+ - \xi^-, \quad \eta = \eta^+ - \eta^-, \quad x^+ = \max\{0,\, x\}, \quad x^- = \max\{0,\, -x\}.$$

Случайные векторы (ξ^+, ξ^-), (η^+, η^-) независимы как функции от независимых случайных величин ξ и η и имеют неотрицательные компоненты. Поэтому

$$\begin{gathered}\mathbf{M}\xi\eta = \mathbf{M}(\xi^+ - \xi^-)(\eta^+ - \eta^-) = \\ = \mathbf{M}\xi^+\eta^+ - \mathbf{M}\xi^+\eta^- - \mathbf{M}\xi^-\eta^+ + \mathbf{M}\xi^-\eta^- = \\ = \mathbf{M}\xi^+\mathbf{M}\eta^+ - \mathbf{M}\xi^+\mathbf{M}\eta^- - \mathbf{M}\xi^-\mathbf{M}\eta^+ + \mathbf{M}\xi^-\mathbf{M}\eta^- = \\ = (\mathbf{M}\xi^+ - \mathbf{M}\xi^-)(\mathbf{M}\eta^+ - \mathbf{M}\eta^-) = \mathbf{M}\xi\mathbf{M}\eta.\end{gathered}$$

При доказательстве различных теорем теории вероятностей часто используются два утверждения, фактически относящиеся к математическому анализу: теоремы о монотонной и мажорируемой сходимости.

Теорема о монотонной сходимости. *Если неубывающая последовательность $\{\xi_n\}$ неотрицательных случайных величин такова, что $\xi_n \uparrow \xi$ при $n \uparrow \infty$, то*

$$\lim_{n\to\infty} \mathbf{M}\xi_n = \mathbf{M}\xi.$$

Эта теорема отличается от аналогичного утверждения при обосновании корректности определения математического ожидания тем, что случайные величины ξ_n не предполагаются простыми.

Доказательство. В силу свойства монотонности математического ожидания из условия $0 \leqslant \xi_n \leqslant \xi$ следует, что $0 \leqslant \mathbf{M}\xi_n \leqslant \mathbf{M}\xi$ и

$$\lim_{n\to\infty} \mathbf{M}\xi_n \leqslant \mathbf{M}\xi.$$

Докажем, что верно и обратное неравенство. Для каждого $n = 1, 2, \ldots$ введем монотонную не убывающую последовательность $\{\xi_{n\,1}, \xi_{n\,2}, \ldots\}$ простых случайных величин так, что $0 \leqslant \xi_{n\,k} \uparrow \xi_n$ при $k \uparrow \infty$. Случайные величины η_k:

$$\eta_k = \max\{\xi_{1\,k}, \xi_{2\,k}, \ldots, \xi_{k\,k}\} \leqslant \max\{\xi_1, \xi_2, \ldots, \xi_k\} = \xi_k$$

тоже являются простыми, причем последовательность η_k не убывает:

$$0 \leqslant \eta_k = \max\{\xi_{1\,k}, \xi_{2\,k}, \ldots, \xi_{k\,k}\} \leqslant \max\{\xi_{1\,k+1}, \xi_{2\,k+1}, \ldots, \xi_{k+1\,k+1}\} = \eta_{k+1}.$$

Значит, для всех $\omega \in \Omega$ существует $\lim\limits_{k\to\infty} \eta_k(\omega) = \eta(\omega)$. Так как $\eta_k \leqslant \xi_k$ при всех k и η_k — простые случайные величины, то

$$\lim_{k\to\infty} \mathbf{M}\eta_k = \mathbf{M}\eta \leqslant \lim_{k\to\infty} \mathbf{M}\xi_k \leqslant \mathbf{M}\xi.$$

С другой стороны, $\xi_{n\,k} \leqslant \eta_k \leqslant \eta$; переходя к пределу по $k \to \infty$, находим: $\xi_n \leqslant \eta$ при всех n, значит, $\xi = \lim_{k\to\infty} \xi_k \leqslant \eta$ и $\mathbf{M}\xi \leqslant \mathbf{M}\eta$. Таким образом,

$$\mathbf{M}\xi \leqslant \mathbf{M}\eta \leqslant \lim_{k\to\infty} \mathbf{M}\xi_k \leqslant \mathbf{M}\xi,$$

и теорема доказана.

Следствие 1. *Если случайные величины ξ_n неотрицательны, то*

$$\mathbf{M} \sum_{n \geqslant 1} \xi_n = \sum_{n \geqslant 1} \mathbf{M}\xi_n.$$

Для доказательства достаточно рассмотреть монотонно не убывающую последовательность случайных величин $\eta_k = \sum_{n=1}^{k} \xi_n$ и применить теорему о монотонной сходимости.

Следствие 2. *Если $\mathbf{M}\eta$ конечно и последовательность событий $A_n \downarrow \varnothing$ при $n \uparrow \infty$, то*

$$\lim_{n\to\infty} \mathbf{M}\eta I\{A_n\} = 0.$$

Доказательство. По условию $\mathbf{M}\,|\eta| < \infty$. Представим $|\eta|$ в виде $|\eta| = \eta_n + \eta_n'$, где $\eta_n = |\eta| I\left\{A_n\right\}$, $\eta_n' = |\eta| I\left\{\overline{A_n}\right\}$. Тогда

$$\mathbf{M}\,|\eta| = \mathbf{M}\eta_n + \mathbf{M}\eta_n'.$$

Так как $\overline{A_n} \uparrow \Omega$, то $0 \leqslant \eta_n' \uparrow |\eta|$ и по теореме о монотонной сходимости

$$\mathbf{M}\eta_n' \uparrow \mathbf{M}\,|\eta|\,, \text{ значит, } \mathbf{M}\eta_n \to 0 \text{ при } n \to \infty.$$

Но тогда $|\mathbf{M}\eta I\left\{A_n\right\}| \leqslant \mathbf{M}\,|\eta I\left\{A_n\right\}| = \mathbf{M}\eta_n \to 0$.

Теорема Лебега о мажорируемой сходимости. *Если*

$$\lim_{n\to\infty} \xi_n(\omega) = \xi(\omega) \ \ \text{и} \ \ |\xi_n(\omega)| \leqslant \eta(\omega) \ \ \text{при всех} \ \ \omega \in \Omega,$$

причем $\mathbf{M}\eta < \infty$, то

$$\lim_{n\to\infty} \mathbf{M}\xi_n = \mathbf{M}\xi.$$

Доказательство. Выберем произвольно $\varepsilon > 0$ и рассмотрим последовательность событий

$$A_n = \left\{\omega \in \Omega\colon\ \sup_{m>n} |\xi_m\,(\omega) - \xi(\omega)| < \varepsilon\right\}, \quad n = 1, 2, \ldots$$

Из условия теоремы следует, что $\overline{A_n} \downarrow \varnothing$ при $n \uparrow \infty$ для любого $\varepsilon > 0$. Тогда, как в следствии 2,

$$\xi_n = \xi_n I\left\{A_n\right\} + \xi_n I\left\{\overline{A_n}\right\}$$

и

$$\begin{gathered} \xi I\{A_n\} - \varepsilon \leqslant \xi_n I\{A_n\} \leqslant \xi I\{A_n\} + \varepsilon, \\ -\eta I\left\{\overline{A_n}\right\} \leqslant \xi_n I\left\{\overline{A_n}\right\} \leqslant \eta I\left\{\overline{A_n}\right\}. \end{gathered}$$

Складывая почленно эти неравенства с учетом равенства $\xi I\{A_n\} = \xi - \xi I\left\{\overline{A_n}\right\}$ и того, что $|\xi I_{\overline{A_n}} \leqslant \eta I_{\overline{A_n}}$, получаем

$$\xi - \varepsilon - 2\eta I\left\{\overline{A_n}\right\} \leqslant \xi_n \leqslant \xi + \varepsilon + 2\eta I\left\{\overline{A_n}\right\};$$

далее вычислим математическое ожидание от всех частей неравенства:

$$\mathbf{M}\xi - \varepsilon - 2\mathbf{M}\eta I\left\{\overline{A_n}\right\} \leqslant \mathbf{M}\xi_n \leqslant \mathbf{M}\xi + \varepsilon + 2\mathbf{M}\eta I\left\{\overline{A_n}\right\}.$$

Переходя здесь к пределу по $n \to \infty$ и учитывая следствие 2, получим:

$$\mathbf{M}\xi - \varepsilon \leqslant \liminf_{n\to\infty} \mathbf{M}\xi_n \leqslant \limsup_{n\to\infty} \mathbf{M}\xi_n \leqslant \mathbf{M}\xi + \varepsilon.$$

Так как $\varepsilon > 0$ можно выбирать произвольно малым, то $\lim_{n\to\infty} \mathbf{M}\xi_n = \mathbf{M}\xi$; теорема доказана.

§ 26. Формулы для вычисления математического ожидания

Формулы для математических ожиданий случайных величин с дискретными распределениями выведены раньше. Для неотрицательной случайной величины ξ с действительными значениями и произвольным распределением математическое ожидание определяется по общей схеме с помощью неубывающей последовательности простых случайных величин, сходящейся к ξ при $n \to \infty$, например, с помощью последовательности $\xi_n(\omega) = \min\left\{n, \frac{1}{2^n}\left[2^n \xi(\omega)\right]\right\}$:

$$\mathbf{M}\xi = \lim_{n\to\infty} \mathbf{M}\xi_n = \lim_{n\to\infty} \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbf{P}\left\{\omega : \frac{k-1}{2^n} \leqslant \xi(\omega) < \frac{k}{2^n}\right\}.$$

Этот предел является интегралом Лебега по пространству элементарных событий Ω с мерой $\mathbf{P}$ и обозначается $\int\limits_{\Omega} \xi(\omega) d\mathbf{P}\{\omega\}$.

Неотрицательная случайная величина определяет меру $P_\xi(B) = \mathbf{P}\{\xi \in B\}$ на полуоси $[0, \infty)$ и функцию распределения $F_\xi(x)$. В этих терминах формулу для $\mathbf{M}\xi$ можно представить в виде

$$\mathbf{M}\xi = \lim_{n\to\infty} \mathbf{M}\xi_n = \lim_{n\to\infty} \sum_{k=1}^{n2^n} \tfrac{k-1}{2} \mathbf{P}\left\{\left(\tfrac{k-1}{2^n} < \xi(\omega) \leqslant \tfrac{k}{2^n}\right)\right\} =$$
$$= \lim_{n\to\infty} \sum_{k=1}^{n2^n} \tfrac{k-1}{2} \left(F_\xi\left(\tfrac{k}{2^n}\right) - F_\xi\left(\tfrac{k-1}{2^n}\right)\right) = \int_0^\infty x dF(x),$$

т. е. в виде интеграла Лебега–Стилтьеса. Случайную величину ξ, принимающую значения обоих знаков, можно представить в виде разности двух неотрицательных случайных величин $\xi^+ = \max\{\xi, 0\}$ и $\xi^- = \max\{-\xi, 0\}$, применяя к каждой из них последнюю формулу и объединяя результаты, получим следующее выражение:

$$\mathbf{M}\xi = \int_R x dF(x).$$

Из него следует формула для вычисления математического ожидания случайной величины, имеющей абсолютно непрерывное распределение.

Утверждение 1. *Если распределение случайной величины* ξ *имеет плотность* $p_\xi(x)$ *и* $\int_{\mathbb{R}} |x| p_\xi(x) dx < \infty$, *то*

$$\mathbf{M}\xi = \int_{\mathbb{R}} x p_\xi(x) dx.$$

Доказательство. При условиях утверждения

$$F_\xi(x) = \mathbf{P}\{\xi \leqslant x\} = F_\xi(0) + \int_0^x p_\xi(u) du,$$

и поэтому по правилам интегрального исчисления

$$\mathbf{M}\xi = \int_{\mathbb{R}} x dF_\xi(x) = \int_{\mathbb{R}} x d\left(F_\xi(0) + \int_0^x p_\xi(u) du\right) = \int_{\mathbb{R}} x p_\xi(x) dx.$$

Немного сложнее (фактически с повторением процедуры определения математического ожидания) проводится доказательство формулы для математического ожидания функции от случайной величины.

Утверждение 2. *Если распределение случайной величины ξ имеет плотность $p_\xi(x)$, функция $g(x)$ кусочно-непрерывна и $\int\limits_{\mathbb{R}} |g(x)|p_\xi(x)dx < \infty$, то*

$$\mathbf{M}g(\xi) = \int\limits_{\mathbb{R}} g(x)p_\xi(x)dx. \tag{46}$$

Доказательство. Пусть случайная величина ξ определена на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbf{P})$. Как и при определении математического ожидания, достаточно рассмотреть случай, когда $g(x) \geqslant 0$ при всех x.

Прежде всего заметим, что формула (46) верна, если $g(x)$ — простая ступенчатая функция с не более чем счетным множеством интервалов постоянства. Действительно, в этом случае существует такое разбиение действительной оси на не более чем счетное множество интервалов $\Delta_k = (a_k, b_k), k = 1, 2, \ldots,$ что $g(x) = g_k$ для всех $x \in \Delta_k$. Множество концов этих интервалов относительно непрерывного распределения имеет нулевую меру и поэтому не влияет на значения математического ожидания и интеграла. Отсюда, из равенства $g(\xi) = \sum_{k\geqslant 1} g_k I\{\xi \in \Delta_k\}$ и из свойств счетной аддитивности математического ожидания и интеграла следует, что

$$\mathbf{M}g(\xi) = \sum_{k\geqslant 1} \mathbf{M}g_k\mathbb{I}\{\xi \in \Delta_k\} =$$

$$= \sum_{k\geqslant 1} g_k\mathbf{P}\{\xi \in \Delta_k\} = \sum_{k\geqslant 1} g_k \int\limits_{\Delta_k} p_\xi(x)dx = \int\limits_{\mathbb{R}} g(x)p_\xi(x)dx.$$

Если функция $g(x)$ кусочно-непрерывна, то последовательность $g_n(x) = \frac{1}{2^n}[2^n g(x)]$, $n = 1, 2, \ldots$, состоит из ступенчатых функций и сходится к $g(x)$, монотонно не убывая. Поэтому случайные величины $\xi_n = g_n(\xi)$, монотонно не убывая, сходятся к ξ. По теореме о монотонной сходимости тогда

$$\lim_{n\to\infty} \mathbf{M}g_n(\xi) = \mathbf{M}g(\xi),$$

но, как было показано выше, $\mathbf{M}g_n(\xi) = \int_a^b g_n(x)p_\xi(x)dx$, а эти интегралы при $n \to \infty$ сходятся к $\int_a^b g(x)p_\xi(x)dx$ по построению функций $g_n(x)$. Утверждение доказано.

§ 27. Математические ожидания комплекснозначных случайных величин

Распространим понятие математического ожидания на комплекснозначные случайные величины.

Пусть $\zeta = \xi + i\eta$, где $\xi = \mathrm{Re}\zeta$ и $\eta = \mathrm{Im}\zeta$ – случайные величины с действительными значениями, причем существуют и конечны $\mathbf{M}\xi$ и $\mathbf{M}\eta$. Тогда по определению полагают

$$\mathbf{M}\zeta = \mathbf{M}\xi + i\mathbf{M}\eta.$$

Таким образом, $\mathrm{Re}\mathbf{M}\zeta = \mathbf{M}\mathrm{Re}\zeta, \quad \mathrm{Im}\mathbf{M}\zeta = \mathbf{M}\mathrm{Im}\zeta.$

Основные свойства математического ожидания (аддитивность, мультипликативность для произведения независимых случайных величин) переносятся на математические ожидания комплекснозначных случайных величин.

Лемма 1. *Если существует математическое ожидание комплекснозначной случайной величины ζ, то*

$$|\mathbf{M}\zeta| \leqslant \mathbf{M}|\zeta|.$$

Доказательство. Рассмотрим функцию $m\,(x+iy) = |x+iy| = \sqrt{x^2+y^2}$, $x, y \in \mathbb{R}$, как функцию от двух действительных переменных x и y. Для любых точек $(a,b), (x,y) \in \mathbb{R}^2$ $(a,b) \neq (0,0)$,

$$z = z\,(x,y) \stackrel{\mathrm{def}}{=} \frac{xa+yb}{(a^2+b^2)^{1/2}} \leqslant m\,(x+iy) = \sqrt{x^2+y^2},$$

так как по неравенству Коши–Буняковского

$$((x,y),(a,b)) = ax + by \leqslant \sqrt{(x^2+y^2)(a^2+b^2)}.$$

Пусть теперь ζ – комплекснозначная случайная величина и $\mathbf{M}\zeta = a + ib$. Тогда

$$\mathbf{M}|\zeta| = \mathbf{M}m(\zeta) \geqslant \mathbf{M}z(\mathrm{Re}\,\zeta,\, \mathrm{Im}\,\zeta) = \tfrac{\mathbf{M}(a\mathrm{Re}\,\zeta + b\mathrm{Im}\,\zeta)}{\sqrt{a^2+b^2}} =$$
$$= \sqrt{a^2+b^2} = |a+ib| = |\mathbf{M}\zeta|.$$

§ 28. Характеристические функции

Определение. *Характеристической функцией* случайной величины ξ, принимающей действительные значения, называется функция

$$f(t) = \mathbf{M} \exp\{it\xi\}, \quad -\infty < t < \infty.$$

Перечислим несколько простых свойств характеристических функций.

1. $|f(t)| \leqslant 1$ при всех $t \in \mathbb{R}$, $f(0) = 1$.

Первое неравенство — простое следствие леммы, так как $|\exp\{it\xi\}| = 1$ и $|f(t)| = |\mathbf{M}\exp\{it\xi\}| \leqslant \mathbf{M}|\exp\{it\xi\}| = 1$. Второе равенство очевидно.

2. Если $\eta = a\xi + b$, где a, b – константы, то

$$f_\eta(t) = e^{itb} f_\xi(at).$$

Действительно,

$$f_\eta(t) = \mathbf{M}\exp\{it\eta\} = \mathbf{M}\exp\{it(a\xi + b)\} = e^{itb}\mathbf{M}\exp\{ita\xi\} = e^{itb} f_\xi(at).$$

3. $f_\gamma(-t) = \overline{f_\gamma(t)}$, где черта — знак комплексного сопряжения. Имеем: $f_\gamma(-t) = \mathbf{M}\exp\{i(-t)\gamma\} = \mathbf{M}\exp\{-it\gamma\} = \mathbf{M}\overline{\exp\{it\gamma\}} = \overline{\mathbf{M}\exp\{it\gamma\}} = \overline{f_\gamma(t)}$.

Следствие. *Если распределение случайной величины γ симметрично, т.е. γ и $-\gamma$ одинаково распределены, то*

$$f_\gamma(t) = \mathbf{M}\exp\{it\gamma\} = \mathbf{M}\exp\{-it\gamma\} = f_\gamma(-t) = \overline{f_\gamma(t)},$$

т. е. характеристическая функция случайной величины с симметричным распределением принимает только действительные значения.

4. Если случайные величины $\gamma_1, \ldots, \gamma_n$ независимы, то

$$f_{\gamma_1+\ldots+\gamma_n}(t) = \prod_{k=1}^{n} f_{\gamma_k}(t).$$

Так как $\gamma_1, \ldots, \gamma_n$ независимы, то случайные величины $\exp\{it\gamma_1\}, \ldots, \exp\{it\gamma_n\}$ тоже независимы, и поэтому

$$f_{\gamma_1+\ldots+\gamma_n}(t) = \mathbf{M}\exp\{it(\gamma_1 + \ldots + \gamma_n)\} = \mathbf{M}\prod_{k=1}^{n}\exp\{it\gamma_k\} =$$
$$= \prod_{k=1}^{n}\mathbf{M}\exp\{it\gamma_k\} = \prod_{k=1}^{n} f_\gamma(t).$$

Примеры. а) Пусть ξ имеет симметричное двухточечное распределение: $\mathbf{P}\{\xi = a\} = \mathbf{P}\{\xi = -a\} = \frac{1}{2}$. Тогда

$$\mathbf{M}\exp\{it\xi\} = \mathbf{P}\{\xi = a\}e^{ita} + \mathbf{P}\{\xi = -a\}e^{-ita} = \frac{\left(e^{ita} + e^{-ita}\right)}{2} = \cos(at).$$

б) Пусть ξ имеет равномерное распределение на $[-a, a]$:

$$p_\xi(x) = \frac{1}{2a} \text{ при } x \in [-a, a] \quad \text{и} \quad p_\xi(x) = 0 \text{ при } |x| > a.$$

Тогда, так как интеграл от ограниченной нечетной функции $\sin(tx)$ по симметричному относительно 0 интервалу $[-a, a]$ равен 0, то

$$\mathbf{M}e^{it\xi} = \int\limits_{-a}^{a} e^{itx}\frac{1}{2a}dx = \frac{1}{2a}\int\limits_{-a}^{a}(\cos(tx) + i\sin(tx))dx =$$
$$= \frac{1}{2at}\sin(tx)\Big|_{-a}^{a} = \frac{\sin(at)}{at}.$$

в) Пусть ξ имеет стандартное нормальное распределение с плотностью $\varphi(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$. Так как плотность $\varphi(x)$ – четная функция, то

$$f(t) = \mathbf{M}\exp\{it\xi\} = \int\limits_{\mathbb{R}} e^{itx}\varphi(x)dx =$$
$$= \int\limits_{\mathbb{R}}(\cos(tx) + i\sin(tx))\varphi(x)dx = \int\limits_{\mathbb{R}}\varphi(x)\cos(tx)dx.$$

Составим уравнение для производной $f(t)$, пользуясь тем, что $-e^{-x^2/2}x\,dx = d\left(e^{-x^2/2}\right)$::

$$f'(t) = \frac{d}{dt}\int\limits_{\mathbb{R}}\frac{e^{-x^2/2}}{\sqrt{2\pi}}\cos(tx)dx = -\frac{1}{\sqrt{2\pi}}\int\limits_{\mathbb{R}}e^{-x^2/2}x\sin(tx)dx =$$
$$= \frac{1}{\sqrt{2\pi}}\int\limits_{\mathbb{R}}\sin(tx)d\left(e^{-x^2/2}\right) =$$
$$= \frac{1}{\sqrt{2\pi}}\left(\sin(tx)e^{-x^2/2}\Big|_{-\infty}^{\infty} - t\int\limits_{\mathbb{R}}\cos(tx)e^{-x^2/2}dx\right) = -tf(t).$$

Дифференциальное уравнение $f'(t) = -tf(t)$ с начальным условием $f(0) = 1$ решается методом разделения переменных:

$$\frac{f'(t)}{f(t)} = (\ln f(t))' = -t \Rightarrow \ln f(t) = \frac{-t^2}{2},$$

т.е. $f(t) = \exp\left(-\frac{t^2}{2}\right), \;\; t \in \mathbb{R}$.

Если случайная величина ξ имеет стандартное нормальное распределение с математическим ожиданием 0 и дисперсией 1, то случайная величина $\eta = \sigma\xi + a$ имеет нормальное распределение с параметрами (a, σ^2) (математическим ожиданием a и дисперсией σ^2). По свойству 2) получаем:

$$\mathbf{M} \exp(it\eta) = e^{ita} f(\sigma t) = \exp\left(ita - \frac{\sigma^2 t^2}{2}\right).$$

Случайная величина η имеет функцию распределения $\mathbf{P}\{\eta \leqslant x\} = \Phi\left(\frac{x-a}{\sigma}\right)$ и плотность распределения $\varphi_{a,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}$.

Как и функция распределения, характеристическая функция существует для *любой* случайной величины с действительными значениями. Более того, характеристические функции обладают рядом аналитических свойств.

Лемма 2. *Для любых $t \in R$ и целых неотрицательных n*

$$\left| e^{it} - \sum_{k=0}^{n-1} \frac{(it)^k}{k!} \right| \leqslant \frac{|t|^n}{n!}.$$

Доказательство проведем индукцией по n. Если $n = 0$, то сумма равна 0 и $|e^{it}| = 1 = \frac{|t|^0}{0!}$, при $n = 1$ имеем

$$\left|e^{it} - 1\right| = \left|\int_0^t \frac{1}{i} e^{iu} du\right| \leqslant \int_0^{|t|} \left|\frac{1}{i} e^{iu}\right| du \, |t| \quad \text{при всех} \;\; t \in \mathbb{R}.$$

Допустим, что для некоторого целого неотрицательного n утверждение леммы справедливо, и сделаем индуктивный переход. Заметим, что

$$\frac{d}{dt}\left(e^{it} - \sum_{k=0}^{n} \frac{(it)^k}{k!}\right) = i\, e^{it} - \sum_{k=0}^{n} \frac{ik(it)^{k-1}}{k!} = i\left(e^{it} - \sum_{k=0}^{n-1} \frac{(it)^k}{k!}\right)$$

и что при $t = 0$ разность в круглых скобках в правой части равна 0. Поэтому

$$\left| e^{it} - \sum_{k=0}^{n} \frac{(it)^k}{k!} \right| = \left| i \int\limits_0^t \left(e^{iu} - \sum_{k=0}^{n-1} \frac{(iu)^k}{k!} \right) du \right|.$$

Оценивая подынтегральное выражение с помощью индуктивного предположения, находим

$$\left|e^{it} - \sum_{k=0}^{n} \frac{(it)^k}{k!}\right| = \left|\int_0^t \left(e^{iu} - \sum_{k=0}^{n-1} \frac{(iu)^k}{k!}\right) du\right| \leqslant \int_0^{|t|} \frac{|u|^n}{n!} du = \frac{|t|^{n+1}}{(n+1)!}.$$

Лемма доказана.

5. **Лемма 3.** *Пусть случайная величина ξ имеет функцию распределения $F(x) = \mathbf{P}\{\xi \leqslant x\}$, $x \in \mathbb{R}$, и $\mathbf{M}|\xi|^n < \infty$ при некотором натуральном n. Положим $\mathbf{M}\xi^k = m_k, k = 0, 1, \ldots, n$. Тогда при $k = 1, \ldots, n$ существуют производные $f^{(k)}(t)$ и*

$$f^{(k)}(0) = i^k \mathbf{M}\xi^k = i^k m_k, \quad |f^{(k)}(t)| \leqslant |f^{(k)}(0)| \ \forall t \in (-\infty, \infty), \quad k = 1, \ldots, n;$$

более того, имеют место разложения

$$f(t) = \sum_{k=0}^{n} m_k \frac{(it)^k}{k!} + R_n(t), \quad R_n(t) = o(t^n), \quad t \to 0, \tag{47}$$

$$f(t) = \sum_{k=0}^{n-1} m_k \frac{(it)^k}{k!} + r_n(t), \quad |r_n(t)| \leqslant \mathbf{M}|\xi|^n \cdot \frac{|t|^n}{n!}. \tag{48}$$

Доказательство. Формальным дифференцированием по t равенства

$$f(t) = \mathbf{M} \exp\{it\xi\} = \int_R e^{itx} dF(x)$$

получаем:

$$f^{(k)}(t) = \int_R (ix)^k e^{itx} dF(x); \tag{49}$$

законность дифференцирования следует из того, что полученный интеграл сходится абсолютно при любом t:

$$\int_{\mathbb{R}} |(ix)^k e^{itx}| dF(x) = \int_{\mathbb{R}} |x|^k dF(x) < 1 + \mathbf{M}|\xi|^n \infty.$$

Полагая $t = 0$ в (49), доказываем первое утверждение:

$$f^{(k)}(0) = \int_{\mathbb{R}} (ix)^k dF(x) = i^k m_k, \quad k = 1, \ldots, n.$$

Очевидно, $|f^{(k)}(t)| \leqslant |f^{(k)}(0)|$ при всех t.

Так как n-кратная дифференцируемость $f(t)$ доказана, для доказательства равенства (47) достаточно использовать формулу Тейлора с остаточным членом в форме Пеано. Оценка остаточного члена в (48) следует из разложения

$$\mathbf{M}\exp\{it\xi\} = \mathbf{M}\left[\sum_{k=0}^{n-1}\frac{(it\xi)^k}{k!} + \rho_n(t\xi)\right] = \sum_{k=0}^{n-1}\frac{m_k(it)^k}{k!} + \mathbf{M}\rho_n(t\xi),$$

где $\rho_n(x) = e^{ix} - \sum_{k=0}^{n-1}\frac{(ix)^k}{k!}$ и по лемме $|\rho_n(t\xi)| \leqslant \frac{|t\xi|^n}{n!}$. Так как для комплекснозначной случайной величины ζ справедлива оценка $|\mathbf{M}\zeta| \leqslant \mathbf{M}|\zeta|$, то

$$|\mathbf{M}\rho_n(t\xi)| = |r_n(t)| \leqslant \mathbf{M}|\xi|^n \cdot \frac{|t|^n}{n!}.$$

Лемма доказана.

Важными частными случаями соотношения (47) являются

$$\begin{gathered} f(t) = f(0) + tf'(0) + o(t) = 1 + it\mathbf{M}\xi + o(t), \\ f(t) = f(0) + tf'(0) + t^2\tfrac{f''(0)}{2!} + o(t^2) = \\ = 1 + it\mathbf{M}\xi - \tfrac{t^2\mathbf{M}\xi^2}{2} + o(t^2), \quad t \to 0; \end{gathered}$$

если $\mathbf{M}\xi = 0$, то $\mathbf{M}\xi^2 = \mathbf{D}\xi = \sigma^2$ и (47), (48) принимают вид

$$\begin{gathered} f(t) = 1 - \tfrac{t^2\sigma^2}{2} + o(t^2) \quad (t \to 0); \\ f(t) = 1 - \tfrac{t^2\sigma^2}{2} + \Theta t^3\mathbf{M}|\xi|^3, \quad |\Theta| \leqslant \tfrac{1}{6}\,. \end{gathered}$$

Следующие два свойства лежат в основе почти всех применений аппарата характеристических функций в теории вероятностей.

6. **Теорема единственности.** *Различным распределениям вероятностей соответствуют различные характеристические функции.*

Доказательство этой теоремы основывается на формуле обращения для характеристических функций, которая выглядит немного сложнее формулы обращения для преобразования Фурье, так как функции распределения могут иметь разрывы (доказательства можно найти в учебниках Б.А.Севастьянова, А.А.Боровкова, а также во втором томе книги В.Феллера).

Формула обращения. *Пусть случайная величина ξ имеет функцию распределения $F_\xi(x) = \mathbf{P}\{\xi \leqslant x\}$ и характеристическую функцию*

$f_\xi(t) = \mathbf{M}e^{it\xi}$. *Если* $a+h$ *и* $a-h$ *— точки непрерывности* $F_\xi(x)$, *то*

$$F_\xi(a+h) - F_\xi(a-h) = \lim_{\sigma\to 0} \frac{1}{\pi} \int_{-\infty}^{\infty} e^{-iat} f_\xi(t) \, \frac{\sin th}{t} \, e^{-\frac{\sigma^2 t^2}{2}} dt.$$

Пример. Пусть $\xi_1, \ldots, \xi_n$ — независимые случайные величины, имеющие нормальные распределения: ξ_k имеет нормальное распределение с параметрами (a_k, σ_k^2), $k = 1, \ldots, n$; им соответствуют характеристические функции

$$f_k(t) = \mathbf{M}e^{it\xi_k} = \exp\left\{iat - \frac{\sigma_k^2 t^2}{2}\right\}, \; k = 1, \ldots, n.$$

Тогда сумма $\zeta = \xi_1 + \ldots + \xi_n$ имеет характеристическую функцию

$$f(t) = \mathbf{M}e^{it(\xi_1+\ldots+\xi_n)} =$$
$$= \prod_{k=1}^{n} \mathbf{M}e^{it\xi_k} = \exp\left\{i(a_1 + \ldots + a_n)t - \frac{1}{2}(\sigma_1^2 + \ldots + \sigma_n^2)t^2\right\},$$

которая совпадает с характеристической функцией нормального распределения с параметрами $(a_1 + \ldots + a_n, \sigma_1^2 + \ldots + \sigma_n^2)$. Таким образом, (согласно теореме единственности) *сумма независимых случайных величин, имеющих нормальные распределения, имеет нормальное распределение, параметры которого являются суммами параметров слагаемых.*

Случайные величины — это функции на пространстве элементарных событий. В математическом анализе рассматриваются разные виды сходимости функций: поточечная, равномерная, почти всюду, в том или ином функциональном пространстве и т.д. В теории вероятностей тоже рассматриваются разные виды сходимости случайных величин. Пока рассмотрим только сходимость распределений (слабую сходимость).

Слабая сходимость. **Определение.** Последовательность $\xi_1, \xi_2, \ldots$ случайных величин с действительными значениями слабо сходится к случайной величине ξ, если функции распределения $F_n(x) = \mathbf{P}\{\xi_n \leqslant x\}$ сходятся к функции распределения $F(x) = \mathbf{P}\{\xi \leqslant x\}$ в каждой точке непрерывности $F(x)$.

В общем случае последовательность случайных величин $\{\xi_n\}$ со значениями в измеримом пространстве $(B, \mathcal{B})$ с метрикой ρ называют *слабо* *сходящейся* к случайной величине ξ, если для любой непрерывной ограниченной функции $g\colon B \to \mathbb{R}$

$$\lim_{n\to\infty} \mathbf{M} g(\xi_n) = \mathbf{M} g(\xi).$$

Для действительных случайных величин эти определения эквивалентны.

Пример. При слабой сходимости функции распределения допредельных случайных величин могут не сходиться к предельной функции в точках ее разрыва. Пусть случайные величины ξ_n — двухточечные:

$$\mathbf{P}\{\xi_n = -\tfrac{1}{n}\} = p_n, \quad \mathbf{P}\{\xi_n = \tfrac{1}{n}\} = 1 - p_n, \quad n = 1, 2, \ldots,$$

а $\mathbf{P}\{\xi = 0\} = 1$. Так как $\mathbf{P}\{\xi_n \in \{-\frac{1}{n}, \frac{1}{n}\}\} = 1$, то естественно считать, что при $n \to \infty$ распределения случайных величин ξ_n сходятся к распределению, сосредоточенному в 0 и имеющему функцию распределения $F(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geqslant 0. \end{cases}$ Для любых значений p_n при каждом $x \neq 0$ последовательность

$$F_n(x) = \mathbf{P}\{\xi_n \leqslant x\} = \begin{cases} 0, & x < -\frac{1}{n}, \\ p_n, & -\frac{1}{n} \leqslant x < \frac{1}{n}, \\ 1, & x \geqslant \frac{1}{n}, \end{cases}$$

при $n \to \infty$ сходится к соответствующему значению $F(x)$ (к 0 при $x < 0$ и к 1 при $x > 0$), но последовательность $F_n(0) = p_n$ может (в зависимости от выбора значений p_n) сходиться к любому числу из отрезка $[0, 1]$, а может не иметь предела.

7. **Теорема непрерывности.** а) *Если последовательность F_n распределений случайных величин γ_n слабо сходится при $n \to \infty$ к распределению F случайной величины γ, то последовательность соответствующих характеристических функций $f_n(t) = \mathbf{M}e^{it\gamma_n}$ сходится к характеристической функции $f(t)$ распределения F при каждом t, и эта сходимость равномерна на любом конечном интервале значений t.*

б) *Обратно, если последовательность характеристических функций $f_n(t)$ при каждом t сходится к функции $f(t)$, непрерывной в нуле, то последовательность соответствующих распределений F_n слабо сходится к распределению F и $f(t)$ есть характеристическая функция распределения F.*

Доказательство теоремы непрерывности, как и доказательство теоремы единственности, можно найти в учебниках Б.А.Севастьянова, А.А.Боровкова и во втором томе книги В.Феллера.

Приведем пример, показывающий, что при нарушении условия непрерывности в 0 в п.б) слабой сходимости может не быть. Пусть ξ_n имеет равномерное распределение на отрезке $[-n,\, n]$: $\mathbf{P}\{\xi_n \leqslant x\} = \frac{1}{2} + \frac{x}{2n}$, $|x| \leqslant n$, $n = 1, 2, \ldots$; тогда (см. приведенный в начале параграфа пример б))

$$f_n(t) = \mathbf{M}e^{it\xi_n} = \frac{\sin(nt)}{nt},$$

и $f_n(0) = 1$ при всех $n = 1, 2, \ldots$, но $f_n(t) \to 0$ при $n \to \infty$ для любого $t \neq 0$. Пределом функций распределения $F_n(t) = \mathbf{P}\{\xi_n \leqslant x\} = \frac{1}{2} + \frac{x}{2n}$ при $n \to \infty$ является функция, тождественно равная $\frac{1}{2}$, не являющаяся функцией распределения собственной случайной величины (т. е. такой, которая с вероятностью 1 принимает конечные значения).

Следствие. *Непрерывная функция, являющаяся поточечным пределом последовательности характеристических функций, сама является характеристической.*

Теорема непрерывности позволяет сводить доказательство слабой сходимости случайных величин или их функций распределения к доказательству поточечной сходимости характеристических функций. Обычно такое доказательство оказывается проще доказательства, использующего прямое вычисление распределений.

§ 29. Центральная предельная теорема

Ранее теорема Муавра–Лапласа о сходимости распределения центрированного и нормированного числа успехов в n независимых испытаниях Бернулли к нормальному распределению была доказана с помощью довольно громоздких вычислений, а результат относился к узкому классу случаев. С помощью характеристических функций легко доказываются значительно более общие утверждения.

Теорема. *Если случайные величины $\xi_1, \xi_2, \ldots$ независимы, одинаково распределены, $\mathbf{M}\xi_1 = a$, $\mathbf{D}\xi_1 = \sigma^2 \in (0,\, \infty)$, то*

$$\lim_{n\to\infty} \mathbf{P}\left\{\frac{\xi_1 + \ldots + \xi_n - na}{\sigma\sqrt{n}} \leqslant x\right\} = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-u^2/2}\, du.$$

Доказательство. Согласно теореме о непрерывности соответствия между функциями распределения и характеристическими функциями достаточно показать, что характеристические функции $f_n(t) = \mathbf{M}e^{it\zeta_n}$ случайных величин

$$\zeta_n = \frac{\xi_1 + \ldots + \xi_n - na}{\sigma\sqrt{n}}$$

при $n \to \infty$ для каждого $t \in \mathbb{R}$ сходятся к характеристической функции стандартного нормального распределения, равной $\exp\left\{-t^2/2\right\}$.

Очевидно,

$$\zeta_n = \frac{\xi_1 + \ldots + \xi_1 - na}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\frac{\xi_k - a}{\sigma};$$

слагаемые в сумме в правой части независимы и имеют такое же распределение, как ζ_1. Поэтому если $\mathbf{M}e^{it(\xi_1-a)/\sigma} = f_1(t)$, то

$$\mathbf{M}\exp\left\{it\sum_{k=1}^{n}\frac{\xi_k - a}{\sigma}\right\} = \left(\mathbf{M}\exp\left\{it\frac{\xi_1 - a}{\sigma}\right\}\right)^n = f_1^n(t).$$

Так как $\mathbf{M}\,\frac{\xi_1-a}{\sigma} = 0$, $\mathbf{D}\,\frac{\xi_1-a}{\sigma} = \frac{1}{\sigma^2}\,\mathbf{D}\,\xi_1 = 1$, то по свойству 6 характеристических функций

$$f_1(t) = \mathbf{M}\exp\left\{it\frac{\xi_1 - a}{\sigma}\right\} = 1 - \frac{t^2}{2} + o(t^2), \quad t \to 0.$$

Следовательно, при любом фиксированном $t \in \mathbb{R}$ и $n \to \infty$

$$\mathbf{M}\exp\{it\zeta_n\} = \mathbf{M}\exp\left\{\frac{it}{\sqrt{n}}\sum_{k=1}^{n}\frac{\xi_k - a}{\sigma}\right\} = f_1^n\left(\frac{t}{\sqrt{n}}\right) =$$
$$= \left(1 - \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 + o\left(\frac{t^2}{n}\right)\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{2n}\right)\right)^n \to e^{-t^2/2},$$

так как $\ln\left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{2n}\right)\right)^n = n\left(-\frac{t^2}{2n} + o\left(\frac{t^2}{2n}\right)\right) \to -\frac{t^2}{2}$. Отсюда и из теорем непрерывности и единственности следует утверждение теоремы.

Метод характеристических функций позволяет доказывать предельные теоремы и в других ситуациях. В качестве примера рассмотрим теорему Ляпунова: в ней рассматривается не последовательность нарастающих сумм независимых одинаково распределенных случайных величин, а схема серий, в которой при разных числах слагаемых распределения слагаемых могут быть разными.

Теорема Ляпунова. *Пусть при каждом* $n = 1, 2, \ldots$ *случайные величины* $\xi_{n,1}, \ldots, \xi_{n,n}$ *независимы,*

$$\mathbf{M}\xi_{n,k} = a_{n,k}, \quad \mathbf{D}\xi_{n,k} = \sigma^2_{n,k}, \quad \mathbf{M}|\xi_{n,k} - a_{n,k}|^3 = c^3_{n,k}, \quad n, k = 1, 2, \ldots,$$

и $S_n = \xi_{n,1} + \ldots + \xi_{n,n}$. *Если*

$$A_n = \sum_{k=1}^{n} a_{n,k}, \quad B_n^2 = \sum_{k=1}^{n} \sigma^2_{n,k}, \quad C_n^3 = \sum_{k=1}^{n} c^3_{n,k}$$

и распределения слагаемых изменяются так, что $\frac{C_n}{B_n} \to 0$, $n \to \infty$, *то*

$$\lim_{n\to\infty} \mathbf{P}\left\{\frac{S_n - A_n}{B_n} \leqslant x\right\} = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-u^2/2}\, du.$$

Доказательство. Достаточно доказать, что характеристические функции случайных величин $\frac{1}{B_n}(S_n - A_n)$ в каждой точке t сходятся к характеристической функции стандартного нормального распределения, равной $\exp\{-t^2/2\}$.

Аналогично случаю, когда слагаемые одинаково распределены, представим характеристическую функцию в виде произведения характеристических функций центрированных и нормированных слагаемых и используем разложения характеристических функций с явной оценкой остаточного члена: при любом $t \in \mathbb{R}$

$$\mathbf{M}\exp\left\{it\frac{S_n - A_n}{B_n}\right\} = \mathbf{M}\exp\left\{it\sum_{k=1}^{n}\frac{\xi_{n,k} - a_{n,k}}{B_n}\right\} =$$

$$= \prod_{k=1}^{n} \mathbf{M}\exp\left\{it\frac{\xi_{n,k} - a_{n,k}}{B_n}\right\} = \prod_{k=1}^{n}\left(1 - \frac{t^2}{2B_n^2}\,\sigma^2_{n,k} + \theta_{n,k}\frac{t^3}{B_n^3}\,c^3_{n,k}\right),$$

где $|\theta_{n,k}| \leqslant \frac{1}{6}, \quad k = 1, \ldots, n$ (отметим, что остаточный член θ, вообще говоря, — комплексное число). Согласно неравенству Ляпунова

$$\sigma_{n,k} = \sqrt{\mathbf{M}(\xi_{n,k} - a_{n,k})^2} \leqslant \left(\mathbf{M}|\xi_{n,k} - a_{n,k}|^3\right)^{1/3} = c_{n,k}, \quad k = 1, \ldots, n.$$

По условию $\frac{C_n}{B_n} \to 0$ при $n \to \infty$, значит,

$$\max_{1\leqslant k\leqslant n}\frac{c^3_{n,k}}{B_n^3} \leqslant \left(\frac{C_n}{B_n}\right)^3 \to 0, \quad \text{т. е.} \quad \max_{1\leqslant k\leqslant n}\frac{c_{n,k}}{B_n} \to 0.$$

Следовательно,
$$\max_{1\leqslant k\leqslant n}\frac{\sigma_{n,k}}{B_n}\leqslant \max_{1\leqslant k\leqslant n}\frac{c_{n,k}}{B_n}\to 0.$$
Таким образом, для каждого фиксированного $t\in\mathbb{R}$
$$\max_{1\leqslant k\leqslant n}\left(\frac{t^2}{2B_n^2}\,\sigma_{n,k}^2+\theta_{n,k}\frac{t^3}{B_n^3}\,c_{n,k}^3\right)\to 0 \quad \text{при } n\to\infty. \tag{50}$$
Применяя соотношение $\ln(1+x)=x(1+o(1))$ при $x\to 0$, находим, что при $n\to\infty$ и $k=1,\dots,n$
$$\ln\left(1-\frac{t^2}{2B_n^2}\,\sigma_{n,k}^2+\theta_{n,k}\frac{t^3}{B_n^3}\,c_{n,k}^3\right)=\left(-\frac{t^2}{2B_n^2}\,\sigma_{n,k}^2+\theta_{n,k}\frac{t^3}{B_n^3}\,c_{n,k}^3\right)(1+o(1)),$$
где в силу (50) остаточные члены стремятся к 0 при $n\to\infty$ равномерно по $k\in\{1,\dots,n\}$. Отсюда следует, что для любого $t\in\mathbb{R}$ при некоторых $\theta=\theta(n,t), |\theta|\in[-\frac{1}{6},\frac{1}{6}]$,
$$\ln\mathbf{M}\exp\left\{it\frac{\xi_{n,1}+\ldots+\xi_{n,n}-A_n}{B_n}\right\}=\sum_{k=1}^{n}\ln\left(1-\frac{t^2}{2B_n^2}\,\sigma_{n,k}^2+\theta_{n,k}\frac{t^3}{B_n^3}\,c_{n,k}^3\right)=$$
$$=(1+o(1))\sum_{k=1}^{n}\left(-\frac{t^2}{2B_n^2}\,\sigma_{n,k}^2+\theta_{n,k}\frac{t^3}{B_n^3}\,c_{n,k}^3\right)=$$
$$=(1+o(1))\left(-t^2\frac{B_n^2}{2B_n^2}+\theta t^3\frac{C_n^3}{B_n^3}\right)=-\frac{t^2}{2}\,(1+o(1)).$$
Значит, $\mathbf{M}\exp\left\{it\frac{S_n-A_n}{B_n}\right\}\to e^{-t^2/2}$, $n\to\infty$. Теорема доказана.

В качестве примера применения теоремы Ляпунова докажем следующее утверждение.

Утверждение. *Пусть $\{\xi_{n1},\dots,\xi_{nn}\}_{n=1}^{\infty}$ – последовательность серий независимых случайных величин,*
$$\mathbf{P}\{\xi_{nk}=1\}=p_{nk},\ \mathbf{P}\{\xi_{nk}=0\}=1-p_{nk},\quad k,n\geq 1,$$
и $\zeta_n=\xi_{n1}+\ldots+\xi_{nn}$. Если $\mathbf{D}\zeta_n\to\infty$ при $n\to\infty$, то
$$\mathbf{P}\left\{\frac{\zeta_n-\mathbf{E}\zeta_n}{\sqrt{\mathbf{D}\zeta_n}}\leqslant x\right\}\to\Phi(x),\quad x\in(-\infty,\infty).$$

Доказательство. Покажем, что для сумм ζ_n выполняются условия теоремы Ляпунова об асимптотической нормальности, т.е. что отношение суммы третьих центральных абсолютных моментов слагаемых

$\left(\sum_{k=1}^{n} \mathbf{E}|\xi_{nk} - p_{nk}|^3\right)$ к кубу среднеквадратичного отклонения $\left((\mathbf{D}\zeta_n)^{3/2}\right)$ стремится к 0 при $n \to \infty$. Действительно,

$$\mathbf{E}\xi_{nk} = p_{nk}, \quad \mathbf{D}\xi_{nk} = p_{nk}(1 - p_{nk}),$$
$$\mathbf{E}|\xi_{nk} - \mathbf{E}\xi_{nk}|^3 = p_{nk}(1 - p_{nk})(1 - 2p_{nk} + 2p_{nk}^2) < \mathbf{D}\xi_{nk}.$$

Таким образом,

$$\sum_{k=1}^{n} \mathbf{E}|\xi_{nk} - \mathbf{E}\xi_{nk}|^3 < \sum_{k=1}^{n} \mathbf{D}\xi_{nk} = \mathbf{D}\zeta_n = o((\mathbf{D}\zeta_n)^{3/2}),$$

если $\mathbf{D}\zeta_n \to \infty$. Утверждение доказано.

Нормальное распределение наблюдается на практике, например, при обработке статистических данных. В частности, распределение ошибок измерений часто оказывается близким к нормальному. Одной из причин этого явления может быть то, что ошибки измерений возникают при взаимодействии многих слабых возмущений, которые в первом приближении можно считать независимыми.

§ 30. Характеристические функции распределений случайных векторов

Характеристические функции распределений случайных векторов вводятся аналогично скалярному случаю с заменой произведения аргумента характеристической функции и случайной величины скалярным произведением векторного аргумента характеристической функции и случайного вектора.

Определение. *Характеристической функцией* распределения случайного вектора $\overline{\gamma} = (\gamma_1, \ldots, \gamma_d)$, принимающего значения в R^d, называется

$$f(\overline{t}) = \mathbf{M}e^{i(\overline{t}, \overline{\gamma})} = \mathbf{M}\exp\{i(t_1\gamma_1 + \ldots + t_d\gamma_d)\}, \quad \overline{t} = (t_1, \ldots, t_d) \in \mathbb{R}^d.$$

Свойства характеристических функций случайных векторов аналогичны свойствам характеристических функций одномерных случайных величин, и доказываются они так же, с заменой произведений чисел скалярными произведениями векторов.

1) $|f(\overline{t})| \leqslant 1$ при всех $\overline{t} \in R^d$, $f(\overline{0}) = 1$.

Первая оценка следует из того, что по лемме, доказанной в § 27, $|\mathbf{M}e^{i(\bar{t},\overline{\gamma})}| \leqslant \mathbf{M}|e^{i(\bar{t},\overline{\gamma})}| = 1$, второе равенство очевидно.

Если $d = 1$ и $f(t_0) = \mathbf{M}e^{it_0\gamma} = e^{i\lambda}$ при некоторых $t_0 \neq 0$, $\lambda \in \mathbb{R}$, то $\mathbf{M}e^{i(t_0\gamma-\lambda)} = 1$, т. е.

$$\operatorname{Re}\mathbf{M}e^{i(t_0\gamma-\lambda)} = \mathbf{M}\operatorname{Re}e^{i(t_0\gamma-\lambda)} = \mathbf{M}\cos(t_0\gamma-\lambda) = 1,$$

но $\cos x \leqslant 1$ при всех $x \in \mathbb{R}$ и $\cos x = 1$ только при $x \in \{2\pi k,\, k \in \mathbb{Z}\}$, поэтому

$$\mathbf{P}\{t_0\gamma \in \{\lambda + 2\pi k,\, k \in \mathbb{Z}\}\} = 1,$$

т. е. распределение γ сосредоточено на арифметической прогрессии $\{\frac{\lambda}{t_0} + \frac{2\pi}{t_0}k,\, k \in \mathbb{Z}\}$.

Аналогично, если $d > 1$ и $f(\overline{t_0}) = \mathbf{M}e^{i(\overline{t_0},\overline{\gamma})} = e^{i\lambda}$ при некотором $\overline{t_0} \neq \overline{0}$, то

$$\mathbf{P}\{(\overline{t_0},\overline{\gamma}) \in \{\lambda + 2\pi k,\, k \in \mathbb{Z}\}\} = 1,$$

т. е. распределение случайного вектора $\overline{\gamma}$ сосредоточено на семействе параллельных гиперплоскостей вида $\{\overline{x} \in \mathbb{R}^d \colon (\overline{t_0},\overline{x}) = \lambda + 2\pi k\}, k \in \mathbb{Z}$.

2) Если $\overline{\gamma}$ — случайный вектор (столбец) и $\overline{\kappa} = B\overline{\gamma} + \overline{a}$, где $B = \|b_{ij}\|_{i,j=1}^d$, $\overline{a} \in R^d$, то

$$f_{\overline{\kappa}}(\bar{t}) = \mathbf{M}e^{i(\bar{t},\overline{\kappa})} = e^{i(\bar{t},\overline{a})}f_{\overline{\gamma}}(B^\top\bar{t}),$$

где $^\top$ — знак транспонирования.

Действительно, так как

$$(\bar{t}, B\overline{\gamma}^\top) = \sum_{i=1}^d t_i \sum_{j=1}^d b_{ij}\gamma_j = \sum_{j=1}^d \left(\sum_{i=1}^d b_{ij}t_i\right)\gamma_j = (B^\top\bar{t},\overline{\gamma}),$$

то

$$f_{\overline{\kappa}}(\bar{t}) = \mathbf{M}e^{i(\bar{t},\overline{\kappa})} = \mathbf{M}e^{i(\bar{t},B\overline{\gamma}+\overline{a})} =$$
$$= e^{i(\bar{t},\overline{a})}\mathbf{M}e^{i(\bar{t},B\overline{\gamma})} = e^{i(\bar{t},\overline{a})}\mathbf{M}e^{i(B^\top\bar{t},\overline{\gamma})} = e^{i(\bar{t},\overline{a})}f_{\overline{\gamma}}(B^\top\bar{t}).$$

3) При любом $\bar{t} \in \mathbb{R}^d$ значения $f(-\bar{t}) = \mathbf{M}e^{i(-\bar{t},\overline{\gamma})} = \mathbf{M}e^{-i(\bar{t},\overline{\gamma})}$ и $f(\bar{t}) = \mathbf{M}e^{i(\bar{t},\overline{\gamma})}$ комплексно сопряжены.

4) Если случайные векторы $\overline{\gamma}^{(1)}, \ldots, \overline{\gamma}^{(n)}$ независимы, то

$$f_{\overline{\gamma}^{(1)}+\ldots+\overline{\gamma}^{(n)}}(\bar{t}) = \mathbf{M}e^{i(\bar{t},\overline{\gamma}^{(1)}+\ldots+\overline{\gamma}^{(n)})} = \mathbf{M}\prod_{k=1}^n e^{i(\bar{t},\overline{\gamma}^{(k)})} = \prod_{k=1}^n f_{\overline{\gamma}^{(k)}}(\bar{t}).$$

Замечание. Из свойств 3) и 4) следует, что если $f(t) = \mathbf{M}e^{i(\bar{t},\overline{\gamma})}$, то $|f(t)|^2$ — характеристическая функция случайного вектора $\overline{\gamma}_1 - \overline{\gamma}_2$, где $\overline{\gamma}_1$ и $\overline{\gamma}_2$ независимы и имеют такое же распределение, как $\overline{\gamma}$. Распределение разности $\overline{\gamma}_1 - \overline{\gamma}_2$ симметрично: $-(\overline{\gamma}_1 - \overline{\gamma}_2) = \overline{\gamma}_2 - \overline{\gamma}_1$, и характеристическая функция любого симметричного распределения принимает только действительные неотрицательные значения.

5) **Теорема единственности.** *Различным распределениям вероятностей соответствуют различные характеристические функции.*

6) **Теорема непрерывности.** а) *Если последовательность F_n распределений случайных векторов $\overline{\gamma}_n$ со значениями в $\mathbb{R}^d$ слабо сходится при $n \to \infty$ к распределению F случайного вектора $\overline{\gamma}$, то последовательность соответствующих характеристических функций $f_n(\bar{t}) = \mathbf{M}e^{i(\bar{t},\overline{\gamma_n})}$ сходится к характеристической функции $f(\bar{t})$ распределения F при каждом $\bar{t}$.*

б) *Обратно, если последовательность характеристических функций $f_n(\bar{t})$ при каждом $\bar{t}$ сходится к функции $f(\bar{t})$, непрерывной в нуле, то последовательность соответствующих распределений F_n слабо сходится к распределению F и $f(\bar{t})$ есть характеристическая функция распределения F.*

Как и в скалярном случае, теоремы единственности и непрерывности в этом курсе приводятся без доказательств.

Следствие. *Непрерывная функция, являющаяся поточечным пределом последовательности характеристических функций, сама является характеристической.*

7) **Лемма 1.** *Если случайный вектор $\overline{\gamma} = (\gamma_1, \ldots, \gamma_d)$ имеет конечные смешанные моменты*

$$m_{k_1,\ldots,k_d} = \mathbf{M}\gamma_1^{k_1} \ldots \gamma_d^{k_d}, \quad k_1, \ldots, k_d \in \{0, 1, \ldots\}, \quad \sum_{j=1}^{d} k_j \le r,$$

и характеристическую функцию $f(\bar{t}) = \mathbf{M}e^{i(\bar{t},\overline{\gamma})}, \bar{t} = (t_1, \ldots, t_d) \in \mathbb{R}^d$, то существуют частные производные характеристической функции вплоть до r-го порядка и

$$\frac{\partial^{k_1+\ldots+k_d}}{\partial t_1^{k_1} \ldots \partial t_d^{k_d}} f(t_1, \ldots, t_d) = i^k \mathbf{M}e^{i(\bar{t},\overline{\gamma})}\gamma_1^{k_1} \ldots \gamma_d^{k_d}, \quad k = \sum_{j=1}^{d} k_j \le r,$$

$$m_{k_1,\ldots,k_d} = i^{-k}\left.\frac{\partial^k f(\bar t)}{\partial t_1^{k_1}\ldots\partial t_d^{k_d}}\right|_{\bar t=\bar 0}, \quad k=\sum_{j=1}^{d} k_j \leqslant r, \tag{51}$$

и

$$f(\bar t) = 1 + \sum_{k=1}^{r} i^k \sum_{k_1+\ldots+k_d=r} m_{k_1,\ldots,k_d}\frac{t_1^{k_1}\ldots t_d^{k_d}}{k_1!\ldots k_d!} + R_r(\bar t), \tag{52}$$

где $R_r(\bar t) = o(|\bar t|^r)$ *при* $|\bar t| = |t_1| + \ldots + |t_d| \to 0$.

Доказательство леммы. Доказательство проводится индукцией по порядку производных. Формула для частной производной первого порядка, например, по t_1 следует из равенства

$$\frac{\partial}{\partial t_1} f(t_1,t_2,\ldots,t_d) = \lim_{h\to 0}\frac{f(t_1+h,t_2,\ldots,t_d)-f(t_1,t_2,\ldots,t_d)}{h} =$$
$$= \lim_{h\to 0}\frac{\mathbf{M}e^{i(\bar t+h\overline{e_1},\overline{\gamma})}-\mathbf{M}e^{i(\bar t,\overline{\gamma})}}{h} = \mathbf{M}e^{i(\bar t,\overline{\gamma})}\frac{e^{ih\gamma_1}-1}{h}.$$

Согласно лемме 2 из § 28 дробь под знаком математического ожидания не превосходит $|\gamma_1|$ по абсолютной величине и стремится к $i\gamma_1$ при $h\to 0$, поэтому под знаком математического ожидания можно перейти к пределу:

$$\frac{\partial}{\partial t_1} f(t_1,\ldots,t_d) = i\,\mathbf{M}e^{i(\bar t,\overline{\gamma})}\gamma_1.$$

Получив такие формулы для всех частных производных первого порядка, аналогично можно доказывать формулы для производных второго порядка, и по индукции — для частных производных всех таких порядков, при которых конечны соответствующие смешанные моменты. Лемма доказана.

Формулы для моментов получаются из формул леммы 1 подстановкой $\bar t = \bar 0$. После того, как доказаны формулы для частных производных, формула (52) становится формулой Тейлора для функции d переменных.

Следствие. *Пусть случайный вектор* $\overline{\gamma} = (\gamma_1,\ldots,\gamma_d)$ *имеет конечные математическое ожидание* $\overline{a} = (a_1,\ldots,a_d) = (\mathbf{M}\gamma_1,\ldots,\mathbf{M}\gamma_d)$ *и матрицу ковариаций* $C(\overline{\gamma})$. *Тогда при* $|\bar t| = |t_1|+\ldots+|t_d|\to 0$

$$f(\bar t) = e^{i(\bar t,\overline{a})}\left(1-\frac{1}{2}\,\bar t C(\overline{\gamma})\bar t^{\top} + o(|\bar t|^2)\right).$$

Доказательство. Представим вектор $\overline{\gamma}$ в виде $\overline{\gamma} = \overline{a} + (\overline{\gamma}-\overline{a})$. По свойству 2)

$$\mathbf{M}e^{i(\bar t,\overline{\gamma})} = e^{i(\bar t,\overline{a})}\mathbf{M}e^{i(\bar t,\overline{\gamma}-\overline{a})}.$$

Далее, $\mathbf{M}(\overline{\gamma}-\overline{a})=(m_1,\ldots,m_d)=0$ и

$$C(\overline{\gamma})=\|c_{ij}\|=\|\mathrm{Cov}(\gamma_i,\gamma_j)\|=\|\mathbf{M}(\gamma_i-a_i)(\gamma_j-a_j)\|,$$

т.е. ковариации компонент вектора $\overline{\gamma}$ совпадают со смешанными моментами второго порядка вектора $\overline{\gamma}-\overline{a}$; поэтому по свойству 7)

$$\mathbf{M}e^{i(\overline{t},\overline{\gamma}-\overline{a})}=1-\sum_{k_1+\ldots+k_d=2} m_{k_1,\ldots,k_d}\frac{t_1^{k_1}\ldots t_d^{k_d}}{k_1!\ldots k_d!}+o(|\overline{t}|^2)=$$

$$=1-\frac{1}{2}\sum_{i,j=1}^{d} c_{ij}t_it_j+o(|\overline{t}|^2),\quad |\overline{t}|\to 0.$$

(Слагаемому, соответствующему набору $(k_1,\ldots,k_d)$, в котором отлично от нуля только $k_i=2$, соответствует одно слагаемое с коэффициентом c_{ii}, а набору с двумя ненулевыми элементами $k_i=k_j=1, i<j,$ — слагаемые с коэффициентами c_{ij} и c_{ji}.) Тем самым следствие доказано.

Пример: многомерные нормальные распределения. Случайная величина ζ имеет стандартное нормальное распределение с параметрами $(0,1)$ (математическим ожиданием 0 и дисперсией 1), если ее плотность распределения равна $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}, -\infty<x<\infty$; функция стандартного нормального распределения есть

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-u^2/2}du,$$

а его характеристическая функция

$$\psi(t)=\mathbf{M}e^{it\zeta}=e^{-t^2/2},\quad -\infty<t<\infty.$$

Среди вероятностных распределений в $\mathbb{R}^d$ аналогом стандартного нормального распределения является распределение случайного вектора $\overline{\zeta}=(\zeta_1,\ldots,\zeta_d)$, компоненты которого — независимые случайные величины, имеющие стандартное нормальное распределение. Такой случайный вектор имеет нулевой вектор математических ожиданий и единичную матрицу ковариаций:

$$\mathbf{M}\overline{\zeta}=\overline{0},\quad \mathrm{Cov}(\overline{\zeta})=\|\mathrm{cov}(\zeta_i,\zeta_j)\|=\begin{Vmatrix}1&0&0&\ldots&0\\0&1&0&\ldots&0\\ \ldots&\ldots&\ldots&\ldots&\ldots\\0&0&0&\ldots&1\end{Vmatrix}.$$

Плотность и характеристическая функция распределения вектора $\overline{\zeta} = (\zeta_1, \ldots, \zeta_d)$ имеют вид

$$\varphi(x_1, \ldots, x_d) = \prod_{k=1}^{d} \varphi(x_k) = \frac{1}{(2\pi)^{d/2}}\, e^{-\frac{x_1^2 + \ldots + x_d^2}{2}},$$

$$\psi(t_1, \ldots, t_d) = \mathbf{M} e^{i(\overline{t}, \overline{\zeta})} = \mathbf{M} e^{i(t_1\zeta_1 + \ldots + t_d\zeta_d)} = \mathbf{M} \prod_{k=1}^{d} e^{it_k\zeta_k} = e^{-\frac{t_1^2 + \ldots + t_d^2}{2}}.$$

Плотность $\varphi(\overline{x})$ распределения $\overline{\zeta}$ зависит только от расстояния точки $\overline{x} = (x_1, \ldots, x_d)$ от начала координат, т.е. инвариантна относительно ортогональных преобразований $\mathbb{R}^d$, оставляющих точку $\overline{0}$ на месте. Такие распределения называются *сферически симметричными*. Если случайный вектор имеет сферически симметричное распределение, то ортогональное преобразование с неподвижной точкой $\overline{0}$ переводит этот случайный вектор в другой случайный вектор, имеющий то же самое распределение.

Согласно определению случайный вектор $\overline{\zeta}$ с $\mathbf{M}\overline{\zeta} = \overline{0}$, $\mathrm{Cov}(\overline{\zeta}) = I$ можно представить в виде $\overline{\zeta} = \zeta_1\overline{e_1} + \ldots + \zeta_d\overline{e_d}$, где $\zeta_1, \ldots, \zeta_d$ — независимые случайные величины, имеющие нормальное распределение с параметрами $(0, 1)$, а $\overline{e_1}, \ldots, \overline{e_d}$ — единичные векторы координатных осей. Из сферической симметричности распределения $\overline{\zeta}$ следует, что для любого другого ортонормированного базиса $\overline{e_1'}, \ldots, \overline{e_d'}$ в соответствующем разложении $\overline{\zeta} = \zeta_1'\overline{e_1'} + \ldots + \zeta_d'\overline{e_d'}$ случайные величины $\zeta_1', \ldots, \zeta_d'$ тоже независимы и имеют то же самое нормальное распределение с параметрами $(0, 1)$.

Распределение случайного вектора $c\overline{\zeta}$ с математическим ожиданием $\overline{0}$ и матрицей ковариаций $c^2 I$ тоже является нормальным и сферически симметричным. Такое распределение имеет плотность

$$\varphi_{\overline{0}, c^2 I}(x_1, \ldots, x_d) = \frac{1}{(2\pi)^{d/2} c^d}\, e^{-\frac{x_1^2 + \ldots + x_d^2}{2c^2}}$$

и характеристическую функцию

$$\psi_{\overline{0}, c^2 I}(t_1, \ldots, t_d) = e^{-\frac{(t_1^2 + \ldots + t_d^2)c^2}{2}}.$$

Определение. Распределение случайного d-мерного вектора $\overline{\kappa}$ называется *многомерным нормальным*, если оно совпадает с распределением линейного преобразования случайного вектора $\overline{\zeta}_{\overline{0}, I}$, имеющего нулевое математическое ожидание и единичную матрицу ковариаций: $\overline{\kappa} = B\overline{\zeta}_{\overline{0},I} + \overline{a}$, где $\overline{a} \in \mathbb{R}^d$ — неслучайный вектор и B — $(d \times d)$-матрица.

Замечание 1. Так как множество линейных преобразований образует полугруппу относительно операции суперпозиции, то *линейное преобразование любого случайного вектора, имеющего многомерное нормальное распределение, имеет многомерное нормальное распределение.*

Найдем $\mathbf{M}\overline{\kappa}$ и ковариационную матрицу $\mathrm{Cov}(\overline{\kappa})$. Переход от вектора $\overline{\zeta}_{\overline{0},I}$ к вектору $\overline{\kappa}$ осуществляется преобразованием $g(\overline{x}) = B\overline{x} + \overline{a}$. Значит,

$$\mathbf{M}\overline{\kappa} = \mathbf{M}g(\overline{\zeta}_{\overline{0},I}) = \mathbf{M}(B\overline{\zeta}_{\overline{0},I} + \overline{a}) = \overline{a}.$$

В силу утверждения 2 из § 16

$$\mathrm{Cov}(\overline{\kappa}) = B\mathrm{Cov}(\overline{\zeta}_{\overline{0},I})B^\top = BB^\top.$$

В частности, отсюда следует, что $\det C(\overline{\kappa}) = (\det B)^2 \geqslant 0$.

Согласно утверждению 1 из § 22 при невырожденной $d \times d$-матрице B плотность распределения вектора $\overline{\kappa}$ имеет вид

$$\varphi_{\overline{\kappa}}(\overline{x}) = \frac{\varphi_{\overline{\zeta}_{\overline{0},I}}(B^{-1}(\overline{x} - \overline{a}))}{|\det B|} =$$
$$= \frac{1}{(2\pi)^{d/2}|\det B|}\exp\left\{-\frac{1}{2}\left(B^{-1}(\overline{x} - \overline{a}), B^{-1}(\overline{x} - \overline{a})\right)\right\} =$$
$$= \frac{1}{\sqrt{(2\pi)^d \det C(\overline{\zeta})}}\exp\left\{-\frac{1}{2}(\overline{x} - \overline{a})C^{-1}(\overline{\kappa})(\overline{x} - \overline{a})^\top\right\}. \qquad (53)$$

По теореме из линейной алгебры для любой неотрицательно определенной симметричной матрицы C существует такая матрица B, что $C = BB^\top$. Поэтому для любой симметричной неотрицательно определенной $d \times d$-матрицы C и любого вектора $\overline{a} \in \mathbb{R}^d$ существует d-мерное нормальное распределение с математическим ожиданием $\overline{a}$ и ковариационной матрицей C.

Так как плотность $\varphi_{\overline{\kappa}}(\overline{x})$ выражается через $\overline{a} = \mathbf{M}\zeta$ и $\mathrm{Cov}(\overline{\zeta})$, то *многомерное нормальное распределение однозначно определяется вектором математических ожиданий и матрицей ковариаций.*

В отличие от плотности сферически симметричного нормального распределения, у которой поверхностями уровня являлись сферы с центром в начале координат, для плотности многомерного нормального распределения (53) поверхностями уровня являются эллипсоиды.

Покажем, что если случайный вектор имеет многомерное нормальное распределение и какие-то две его координаты некоррелированы, то они независимы (для произвольных распределений случайных векторов это не так!). Достаточно рассмотреть случай двумерного нормального распределения: совместное распределение любых двух координат вектора с многомерным нормальным распределением является нормальным, так как оно получается из исходного многомерного распределения линейным преобразованием (проекцией на двумерное подпространство).

Если вектор (ξ, η) имеет невырожденное двумерное нормальное распределение с вектором средних (a, b) и матрицей ковариаций $\Sigma = \left(\begin{smallmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{smallmatrix}\right)$, то $\sigma_{12} = \sigma_{21}$,

$$\Sigma^{-1} = \frac{1}{\det\Sigma} \begin{pmatrix} \sigma_{22} & -\sigma_{21} \\ -\sigma_{12} & \sigma_{11} \end{pmatrix}$$

и плотность его распределения равна

$$\varphi(x, y) = \frac{1}{2\pi\sqrt{\det\Sigma}} \exp\left\{-\frac{1}{2\det\Sigma}\left(\sigma_{22}x^2 - 2\sigma_{12}xy + \sigma_{11}y^2\right)\right\}.$$

Случайные величины ξ, η с абсолютно непрерывными распределениями независимы тогда и только тогда, когда плотность их совместного распределения распадается в произведение одномерных плотностей, а в рассматриваемом случае это происходит тогда и только тогда, когда $\sigma_{12} = \operatorname{cov}(\xi, \eta) = 0$.

Пользуясь свойством 2) характеристических функций, найдем характеристическую функцию многомерного нормального распределения с плотностью (53): так как $\overline{\kappa} = B\overline{\zeta}_{\overline{0},I}^{\top} + \overline{a}$, то

$$\psi_{\overline{\kappa}}(\overline{t}) = \mathbf{M}e^{i(\overline{t},\overline{\kappa})} = e^{i(\overline{t},\overline{a})}\psi_{\overline{\zeta}_{\overline{0},I}}(B^{\top}\overline{t}) = e^{i(\overline{t},\overline{a})}e^{-|B^{\top}\overline{t}|^2/2} = \tag{54}$$
$$= e^{i(\overline{t},\overline{a})}e^{-(B^{\top}\overline{t},B^{\top}\overline{t})/2} = e^{i(\overline{t},\overline{a})}e^{-(BB^{\top}\overline{t},\overline{t})/2} = e^{i(\overline{t},\overline{a})}e^{-t\mathrm{Cov}(\overline{\kappa})\overline{t}^{\top}/2}, \quad \overline{t} \in \mathbb{R}^d.$$

Равенство (54) можно использовать как определение многомерного нормального распределения с математическим ожиданием $\overline{a}$ и ковариационной матрицей $C(\overline{\kappa})$. Нужно иметь в виду, что плотность (53) существует только для многомерных нормальных распределений с невырожденными матрицами ковариаций.

Теперь докажем центральную предельную теорему для сумм независимых одинаково распределенных случайных векторов.

Центральная предельная теорема. *Пусть* $\overline{\xi^{(n)}} = (\xi_1^{(n)}, \ldots, \xi_d^{(n)})$, $n = 1, 2, \ldots,$ *— независимые одинаково распределенные случайные векторы с* $\mathbf{M}\overline{\xi^{(1)}} = \overline{a} \in \mathbb{R}^d$ *и матрицей ковариаций* $C(\overline{\xi})$*; пусть* $\overline{\zeta_n} = \overline{\xi^{(1)}} + \ldots + \overline{\xi^{(n)}}$*. Последовательность случайных векторов*

$$\overline{\zeta_n'} = \frac{\overline{\zeta_n} - n\overline{a}}{\sqrt{n}}$$

слабо сходится к случайному вектору $\overline{\zeta}$*, который имеет многомерное нормальное распределение с* $\mathbf{M}\overline{\zeta} = \overline{0}$, $C(\overline{\zeta}) = C(\overline{\xi})$.

Доказательство. Введем центрированные случайные векторы

$$\overline{\xi^{(k)\,\prime}} = \overline{\xi^{(k)}} - \overline{a}, \quad k = 1, 2, \ldots,$$

тогда $\overline{\zeta_n'} = \frac{1}{\sqrt{n}}\,(\overline{\xi^{(1)\,\prime}} + \ldots + \overline{\xi^{(n)\,\prime}})$. Так как $\mathbf{M}\overline{\xi^{(k)\,\prime}} = \overline{0}$ и ковариационная матрица $\overline{\xi^{(k)\,\prime}}$ равна $C(\overline{\xi})$, то по следствию из свойства 7 при $|\overline{t}| \to 0$ характеристическая функция $f_{\overline{\xi^{(1)\,\prime}}}(\overline{t}) = \mathbf{M}e^{i(\overline{t}, \overline{\xi^{(1)\,\prime}})}$ центрированных слагаемых удовлетворяет соотношению

$$f_{\overline{\xi^{(1)\,\prime}}}(\overline{t}) = 1 - \frac{1}{2}\,\overline{t}C(\overline{\xi})\overline{t}^{\top} + o(|\overline{t}|^2).$$

Следовательно, при любом $\overline{t} \in \mathbb{R}^d$ и $n \to \infty$

$$\mathbf{M}e^{i(\overline{t}, \overline{\zeta_n'})} = \mathbf{M}\exp\left\{i\left(\overline{t}, \frac{1}{\sqrt{n}}\,(\overline{\xi^{(1)\,\prime}} + \ldots + \overline{\xi^{(n)\,\prime}})\right)\right\} =$$
$$= \mathbf{M}\exp\left\{i\left(\frac{\overline{t}}{\sqrt{n}}, \overline{\xi^{(1)\,\prime}} + \ldots + \overline{\xi^{(n)\,\prime}}\right)\right\} = \left(f_{\overline{\xi^{(1)\,\prime}}}\left(\frac{\overline{t}}{\sqrt{n}}\right)\right)^n =$$
$$= \left(1 - \frac{1}{2n}\,\overline{t}C(\overline{\xi})\overline{t}^{\top} + o\left(\frac{|\overline{t}|^2}{n}\right)\right)^n = (1 + o(1))\exp\left\{-\frac{1}{2}\,\overline{t}C(\overline{\xi})\overline{t}^{\top}\right\}.$$

Таким образом, характеристические функции ζ_n' сходятся к характеристической функции многомерного нормального распределения с нулевым математическим ожиданием и ковариационной матрицей $C(\overline{\xi})$. Отсюда и из теоремы об однозначном и непрерывном соответствии между множествами распределений и характеристических функций следует утверждение теоремы.

§ 31. Виды сходимости последовательностей случайных величин

Разным видам сходимости функций в математическом анализе соответствуют разные виды сходимости последовательностей случайных величин. Ранее уже вводилось понятие слабой сходимости.

Определение. Последовательность $\xi_1, \xi_2, \ldots$ случайных величин с действительными значениями слабо сходится к случайной величине ξ, если функции распределения $F_n(x) = \mathbf{P}\{\xi_n \leqslant x\}$ сходятся к функции распределения $F(x) = \mathbf{P}\{\xi \leqslant x\}$ в каждой точке непрерывности $F(x)$.

В общем случае последовательность случайных величин $\{\xi_n\}$ со значениями в измеримом пространстве $(B, \mathcal{B})$ с метрикой ρ называют *слабо сходящейся* к случайной величине ξ, если для любой непрерывной ограниченной функции $g: B \to \mathbb{R}$

$$\lim_{n\to\infty} \mathbf{M}g(\xi_n) = \mathbf{M}g(\xi).$$

б) *Сходимость почти наверное* (или *сходимость с вероятностью* 1). Последовательность $\xi_1, \xi_2, \ldots$ сходится *почти наверное* (п.н.) к случайной величине ξ (это обозначается так: $\xi_n \overset{\text{п.н.}}{\to} \xi$), если

$$\mathbf{P}\{\omega : \lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)\} = 1. \tag{55}$$

Утверждение 1. *Последовательность случайных величин $\xi_1, \xi_2, \ldots$ сходится почти наверное к случайной величине ξ тогда и только тогда, когда при любом $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbf{P}\{\omega : \sup_{m\geqslant n} |\xi_m(\omega) - \xi(\omega)| > \varepsilon\} = 0.$$

Доказательство. Положим $A(n,\varepsilon) = \left\{\omega : \sup\limits_{m\geqslant n} |\xi_m(\omega) - \xi(\omega)| > \varepsilon\right\}$. Тогда последовательность событий $A(n,\varepsilon)$, $n = 1, 2, \ldots$, монотонно не возрастает: $A(1,\varepsilon) \supseteq A(2,\varepsilon) \supseteq \ldots$, значит,

$$\bigcap_{n\geqslant 1} A(n,\varepsilon) = \{\omega : \forall n < \infty\, \exists m > n,\ |\xi_m(\omega) - \xi(\omega)| > \varepsilon\} =$$

$$= \left\{\omega \colon \overline{\lim}_{n\to\infty} |\xi_n(\omega) - \xi(\omega)| \geqslant \varepsilon\right\}$$

и по свойству непрерывности вероятностной меры

$$\lim_{n\to\infty} \mathbf{P}\left\{\sup_{m\geqslant n}|\xi_m(\omega)-\xi(\omega)|>\varepsilon\right\} = \lim_{n\to\infty}\mathbf{P}\{A(n,\varepsilon)\} = \mathbf{P}\left\{\bigcap_{n\geqslant 1} A(n,\varepsilon)\right\} =$$
$$= \mathbf{P}\left\{\omega\colon \overline{\lim}_{n\to\infty}|\xi_n(\omega)-\xi(\omega)| \geqslant \varepsilon\right\}.$$

Правая часть равна 0 при любом $\varepsilon > 0$ тогда и только тогда, когда ξ_n сходятся к ξ почти наверное:

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

Поэтому указанное в Утверждении условие эквивалентно сходимости почти наверное. Тем самым утверждение доказано.

в) *Сходимость по вероятности.* Последовательность случайных величин ξ_n *сходится по вероятности* к случайной величине ξ (это обозначается так: $\xi_n \xrightarrow{P} \xi$), если для любого $\varepsilon > 0$

$$\lim_{n\to\infty}\mathbf{P}\{|\xi_n-\xi|>\varepsilon\} = 0.$$

Утверждение 2. *Если последовательность случайных величин ξ_n сходится к случайной величине ξ с вероятностью* 1, *то она сходится и по вероятности.*

Доказательство. Действительно, пусть $\xi_n \overset{\text{п. н.}}{\to} \xi$, $n\to\infty$. Тогда, как было показано в Утверждении 1,

$$\lim_{n\to\infty}\mathbf{P}\left\{\sup_{m\geqslant n}|\xi_n-\xi|>\varepsilon\right\} = 0.$$

Так как $\sup\limits_{m\geqslant n}|\xi_m-\xi| \geqslant |\xi_n-\xi|$, то

$$\lim_{n\to\infty}\mathbf{P}\{|\xi_n-\xi|>\varepsilon\} = 0,$$

что и требовалось доказать.

Приведем простой пример последовательности случайных величин, сходящейся к 0 по вероятности, но не с вероятностью 1 (почти наверное): пусть $\Omega = [0,1]$, $\mathcal{F}$ — борелевская σ-алгебра, $\mathbf{P}$ — мера Лебега на [0, 1] и $N_k = 1+2+\ldots+(k-1) = C_k^2, k = 1,2,\ldots$; положим

$$\xi_{N_k+m}(\omega) = \begin{cases} 1, & \frac{m-1}{k} \leqslant \omega < \frac{m}{k}, \\ 0 & \text{в остальных случаях,}\end{cases} \qquad 1\leqslant m\leqslant k,\ k = 0,1,\ldots$$

Тогда, как нетрудно проверить,

$$\mathbf{P}\{\xi_{N_k+m} = 1\} \leqslant \frac{1}{k} \to 0 \quad \text{при } n \to \infty$$

и

$$\liminf_{n\to\infty} \xi_n(\omega) = 0, \quad \limsup_{n\to\infty} \xi_n(\omega) = 1 \quad \text{при всех } \omega \in [0,1).$$

Утверждение 3. *Если* $\xi_n \xrightarrow{P} \xi$, $n \to \infty$, *то* $\xi_n \xrightarrow{w} \xi$, $n \to \infty$.

Доказательство. Пусть $\varepsilon > 0$ — произвольное положительное число и $A_n(\varepsilon) = \{\omega : |\xi_n - \xi| > \varepsilon\}$. Тогда $\mathbf{P}\{A_n(\varepsilon)\} \to 0$ и

$$\xi - \varepsilon \leqslant \xi_n \leqslant \xi + \varepsilon \quad \text{при } \omega \notin A_n(\varepsilon).$$

Значит,

$$\{\omega : \xi_n \leqslant x\} \subset \{\omega : \xi \leqslant x + \varepsilon\} \cup A_n(\varepsilon),$$
$$\{\omega : \xi \leqslant x - \varepsilon\} \subset \{\xi_n \leqslant x\} \cup A_n(\varepsilon),$$

т.е.

$$\mathbf{P}\{\omega : \xi_n \leqslant x\} \leqslant \mathbf{P}\{\omega : \xi \leqslant x + \varepsilon\} + \mathbf{P}\{A_n(\varepsilon)\},$$
$$\mathbf{P}\{\omega : \xi \leqslant x - \varepsilon\} \leqslant \mathbf{P}\{\xi_n \leqslant x\} + \mathbf{P}\{A_n(\varepsilon)\}.$$

Поэтому

$$\mathbf{P}\{\xi \leqslant x - \varepsilon\} - \mathbf{P}\{A_n(\varepsilon)\} \leqslant \mathbf{P}\{\xi_n \leqslant x\} \leqslant \mathbf{P}\{\xi \leqslant x + \varepsilon\} + \mathbf{P}\{A_n(\varepsilon)\}.$$

Переходя к пределу по $n \to \infty$, получаем, что при любом $\varepsilon > 0$

$$\mathbf{P}\{\xi \leqslant x - \varepsilon\} \leqslant \liminf_{n\to\infty} \mathbf{P}\{\xi_n \leqslant x\} \leqslant$$
$$\leqslant \limsup_{n\to\infty} \mathbf{P}\{\xi_n \leqslant x\} \leqslant \mathbf{P}\{\xi \leqslant x + \varepsilon\}.$$

Из этих неравенств следует, что если x — точка непрерывности функции распределения $\mathbf{P}\{\xi \leqslant x\}$, то

$$\lim_{n\to\infty} \mathbf{P}\{\xi_n \leqslant x\} = \mathbf{P}\{\xi \leqslant x\}.$$

Вз слабой сходимости следует сходимость по вероятности, если предельная случайная величина с вероятностью 1 равна какой-нибудь константе.

Утверждение 4. *Если $\xi_n \xrightarrow{w} \xi$ и распределение ξ вырождено, т.е. $\mathbf{P}\{\xi = c\} = 1$ при некотором c, то $\xi_n \xrightarrow{P} c$.*

Доказательство. Условие $\xi_n \xrightarrow{w} c$ означает, что при любом $\varepsilon > 0$

$$\mathbf{P}\{\xi_n \leqslant c + \varepsilon\} \to 1 \quad \text{и} \quad \mathbf{P}\{\xi_n \leqslant c - \varepsilon\} \to 0 \quad \text{при } n \to \infty.$$

Из этих соотношений следует, что $\mathbf{P}\{|\xi_n - c| > \varepsilon\} \to 0$, т.е. $\xi_n \xrightarrow{P} c$.

г) *Сходимость в среднем.* Последовательность случайных величин $\{\xi_n\}_{n=1}^{\infty}$ сходится к случайной величине ξ *в среднем порядка* $r > 0$ (это обозначается так: $\xi_n \xrightarrow{r} \xi$), если

$$\mathbf{M}|\xi_n - \xi|^r \to 0, \quad n \to \infty.$$

В случае, когда $r = 2$, сходимость называется *сходимостью в среднем квадратическом*, или *среднеквадратической сходимостью.*

При $r > 0$ из неравенства Маркова

$$\mathbf{P}\{|\xi_n - \xi| > \varepsilon\} = \mathbf{P}\{|\xi_n - \xi|^r > \varepsilon^r\} \leqslant \frac{\mathbf{M}|\xi_n - \xi|^r}{\varepsilon^r}$$

следует, что сходимость в среднем влечет сходимость по вероятности.

Таким образом, рассмотренные виды сходимости связаны между собой следующим образом:

$$\xi_n \xrightarrow{\text{п.н.}} \xi \Rightarrow \xi_n \xrightarrow{P} \xi \Rightarrow \xi_n \xrightarrow{w} \xi,$$
$$\xi_n \xrightarrow{r} \xi \Rightarrow \xi_n \xrightarrow{P} \xi \Rightarrow \xi_n \xrightarrow{w} \xi.$$

§ 32. Обычный и усиленный законы больших чисел

Ранее с помощью неравенства Чебышева был доказан закон больших чисел, который во введенных обозначениях можно записать так:

если случайные величины $\xi_1, \xi_2, \ldots$ независимы, $S_n = \xi_1 + \ldots + \xi_n$ и $\sup\limits_k \mathbf{D}\xi_k < c < \infty$, то

$$\frac{1}{n}(S_n - \mathbf{M}S_n) \xrightarrow{P} 0 \quad \textit{при } n \to \infty.$$

Это утверждение можно интерпретировать как асимптотическую вырожденность распределений $\frac{1}{n} S_n$.

Метод характеристических функций позволяет доказать справедливость закона больших чисел без условия конечности дисперсий.

Теорема Хинчина. *Если $\xi_1, \xi_2, \ldots$ независимы, одинаково распределены, $\mathbf{M}\xi_1 = a \in (-\infty, \infty)$ и $S_n = \xi_1 + \ldots + \xi_n$, то*

$$\frac{S_n}{n} \xrightarrow{P} a \quad \text{\textit{при} } n \to \infty.$$

Доказательство. Представим утверждение теоремы в форме $\frac{S_n - na}{n} \xrightarrow{P} 0$. Пусть $f(t) = \mathbf{M}\exp\{it(\xi_1 - a)\}$ — характеристическая функция распределения каждого из центрированных слагаемых; тогда

$$f(t) = 1 + it\mathbf{M}(\xi_1 - a) + o(t) = 1 + o(t),\ t \to \infty.$$

Поэтому при любом $t \in \mathbb{R}$

$$\mathbf{M}\exp\left\{it\frac{S_n - na}{n}\right\} = \left(\mathbf{M}\exp\left\{it\frac{\xi_1 - a}{n}\right\}\right)^n = \\ = f^n\left(\frac{t}{n}\right) = \left(1 + o\left(\frac{t}{n}\right)\right)^n \to 1.$$

Функция, тождественно равная 1, является характеристической функцией распределения, сосредоточенного в 0. Значит, распределение $\frac{1}{n}(S_n - na) = \frac{1}{n}S_n - a$ сходится к распределению, сосредоточенному в точке 0, т.е. $\frac{1}{n}S_n - a \xrightarrow{w} 0$. Согласно утверждению 4 из § 31 из сходимости по распределению к константе следует сходимость по вероятности к той же константе, что и требовалось доказать.

Соотношение $\frac{1}{n}S_n \xrightarrow{P} a$ при $n \to \infty$, соответствующее слабой сходимости, называют *слабым законом больших чисел*, чтобы отличать его от *усиленного закона больших чисел*, согласно которому

$$\frac{1}{n}S_n \xrightarrow{\text{п. н.}} a \quad \text{при } n \to \infty.$$

Доказательство усиленного закона больших чисел использует неравенство Колмогорова.

Неравенство Колмогорова. *Пусть $\xi_1, \ldots, \xi_n$ — независимые случайные величины, которые имеют конечные математические ожидания и дисперсии, и $S_n = \xi_1 + \ldots + \xi_n$. Тогда*

$$\mathbf{P}\left\{\max_{1 \leqslant k \leqslant n} |S_k - \mathbf{M}S_k| \geqslant x\right\} \leqslant \frac{\mathbf{D}S_n}{x^2}.$$

Замечание. Если оценивать левую часть как вероятность объединения событий и использовать для каждого события неравенство Чебышёва, то неравенство окажется значительно более грубым: в числителе вместо дисперсии последней суммы будет стоять сумма дисперсий всех сумм от S_1 до S_n.

Доказательство. Так как $S_k - \mathbf{M}S_k = \sum_{m=1}^{k}(\xi_m - \mathbf{M}\xi_m)$ и

$$\mathbf{M}(\xi_m - \mathbf{M}\xi_m) = 0, \quad \mathbf{D}(\xi_m - \mathbf{M}\xi_m) = \mathbf{D}\xi_m,$$

то без ограничения общности можно считать, что $\mathbf{M}\xi_1 = \ldots = \mathbf{M}\xi_n = 0$, и доказывать, что при этом условии

$$\mathbf{P}\left\{\max_{1\leqslant k\leqslant n} |S_k| \geqslant x\right\} \leqslant \frac{\mathbf{D}S_n}{x^2}\,.$$

Введем случайную величину

$$\tau = \begin{cases} k, & \text{если } \max\{S_1^2, \ldots, S_{k-1}^2\} < x^2,\ S_k^2 \geqslant x^2, \\ n+1, & \text{если } S_1^2, \ldots, S_n^2 < x^2. \end{cases}$$

Так как события $\{\tau = k\}$, $k = 1, \ldots, n+1$, попарно несовместны, то

$$S_n^2 \geqslant S_n^2 \sum_{k=1}^{n} \mathbb{I}\{\tau = k\},$$

и в силу монотонности и аддитивности математического ожидания

$$\begin{gathered} \mathbf{D}S_n = \mathbf{M}S_n^2 \geqslant \sum_{k=1}^{n} \mathbf{M}S_n^2\mathbb{I}\{\tau = k\} = \\ = \sum_{k=1}^{n} \mathbf{M}(S_k + S_n - S_k)^2\mathbb{I}\{\tau = k\} \geqslant \\ \geqslant \sum_{k=1}^{n} \mathbf{M}S_k^2\mathbb{I}\{\tau = k\} + 2\sum_{k=1}^{n} \mathbf{M}S_k(S_n - S_k)\mathbb{I}\{\tau = k\} \end{gathered} \tag{56}$$

(отброшенное слагаемое $\sum_{k=1}^{n} \mathbf{M}(S_n - S_k)^2\mathbb{I}\{\tau = k\}$ неотрицательно).

Сделаем два замечания. Во-первых, если $\tau = k$, то $S_k^2 \geqslant x^2$, значит,

$$\mathbf{P}\{S_k^2\mathbb{I}\{\tau = k\} \geqslant x^2\} = \mathbf{P}\{\tau = k\},$$

т. е. $\mathbf{M}S_k^2\mathbb{I}\{\tau = k\} \geqslant x^2\mathbf{P}\{\tau = k\}$, $k = 1, \ldots, n$. Во-вторых, при любом фиксированном k индикатор $\mathbb{I}\{\tau = k\}$ и $S_k = \xi_1 + \ldots + \xi_k$ есть функции от $\xi_1, \ldots, \xi_k$, и поэтому разность $S_n - S_k = \xi_{k+1} + \ldots + \xi_n$ не зависит от значений $\mathbb{I}\{\tau = k\}$ и S_k. Значит, при любом $k = 1, \ldots, n$

$$\mathbf{M}S_k(S_n - S_k)\mathbb{I}\{\tau = k\} = \mathbf{M}S_k\mathbb{I}\{\tau = k\} \cdot \mathbf{M}(S_n - S_k) =$$
$$= \mathbf{M}S_k\mathbb{I}\{\tau = k\} \cdot \mathbf{M}(\xi_{k+1} + \ldots + \xi_n) = 0.$$

Объединяя эти замечания с оценкой (56), получаем:

$$\mathbf{D}S_n \geqslant \sum_{k=1}^{n} \mathbf{M}S_k^2\mathbb{I}\{\tau = k\} \geqslant \sum_{k=1}^{n} x^2\mathbf{P}\{\tau = k\} =$$
$$= x^2\mathbf{P}\{\tau \leqslant n\} = x^2\mathbf{P}\left\{\max_{1\leqslant k\leqslant n} |S_k| \geqslant x\right\}.$$

Деля обе части неравенства на x^2, завершаем доказательство неравенства Колмогорова.

Усиленный закон больших чисел. *Пусть случайные величины* $\xi_1, \xi_2, \ldots$ *независимы,* $\mathbf{M}\xi_n \equiv 0$, $\mathbf{D}\xi_n = \sigma_n^2$, $S_n = \xi_1 + \ldots + \xi_n$ *и* $\sum\limits_{n\geqslant 1} \frac{\sigma_n^2}{n^2} < \infty$. *Тогда*

$$\frac{1}{n}\, S_n \overset{\text{п. н.}}{\to} 0.$$

Доказательство. Согласно утверждению 1 из §31 для доказательства того, что последовательность $\frac{1}{n}\, S_n$ сходится с вероятностью 1, достаточно показать, что при любом $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbf{P}\left\{\omega : \sup_{k\geqslant n} \left|\tfrac{1}{k}\, S_k\right| > \varepsilon\right\} = 0. \qquad (57)$$

Введем события

$$A_m = \left\{\omega : \max_{2^{m-1}\leqslant k<2^m} \left|\tfrac{1}{k}\, S_k\right| > \varepsilon\right\}, \quad m = 1, 2, \ldots$$

Тогда (57) равносильно тому, что

$$\mathbf{P}\left\{\bigcup_{m\geqslant n} A_m\right\} \to 0, \quad n \to \infty. \qquad (58)$$

Вероятность объединения событий не больше суммы вероятностей этих событий:

$$\mathbf{P}\left\{\bigcup_{m\geqslant n} A_m\right\} \leqslant \sum_{m\geqslant n} \mathbf{P}\{A_m\},$$

поэтому (58) будет доказано, если показать, что ряд $\sum_{m\geqslant 1}\mathbf{P}\{A_m\}$ сходится. Оценим члены этого ряда с помощью неравенства Колмогорова:

$$\mathbf{P}\{A_m\} = \mathbf{P}\left\{\max_{2^{m-1}\leqslant k<2^m}\left|\tfrac{1}{k}\,S_k\right| > \varepsilon\right\} \leqslant \mathbf{P}\left\{\max_{2^{m-1}\leqslant k<2^m}|S_k| > 2^{m-1}\varepsilon\right\} \leqslant$$
$$\leqslant \mathbf{P}\left\{\max_{1\leqslant k<2^m}|S_k| > 2^{m-1}\varepsilon\right\} \leqslant \frac{\mathbf{D}S_{2^m}}{(\varepsilon 2^{m-1})^2} = \frac{1}{\varepsilon^2 2^{2(m-1)}}\sum_{k=1}^{2^m}\sigma_k^2.$$

Из полученных оценок и условия теоремы следует, что

$$\sum_{m\geqslant 1}\mathbf{P}\{A_m\} \leqslant \frac{1}{\varepsilon^2}\sum_{m\geqslant 1}\sum_{k=1}^{2^m}\frac{\sigma_k^2}{2^{2(m-1)}} = \frac{1}{\varepsilon^2}\sum_{k\geqslant 1}\sigma_k^2\sum_{m\colon 2^m\geqslant k}\frac{1}{2^{2(m-1)}} <$$
$$< \frac{1}{\varepsilon^2}\sum_{k\geqslant 1}\frac{4}{1-\frac{1}{4}}\frac{\sigma_k^2}{k^2} < \frac{16}{3\varepsilon^2}\sum_{k\geqslant 1}\frac{\sigma_k^2}{k^2} < \infty.$$

Теорема доказана.

Следствие. *Для числа ν_n успехов в n испытаниях по схеме Бернулли с вероятностью успеха p справедливы как слабый закон больших чисел*

$$\frac{\nu_n}{n} \xrightarrow{P} p, \quad n\to\infty,$$

так и усиленный закон больших чисел

$$\frac{\nu_n}{n} \xrightarrow{\text{п. н.}} p, \quad n\to\infty.$$

Проведенные рассуждения можно дополнить и доказать, что усиленный закон больших чисел для *одинаково распределенных* случайных величин верен при тех же условиях, при которых верен слабый закон больших чисел.

Усиленный закон больших чисел Колмогорова. *Пусть случайные величины $\xi_1, \xi_2, \ldots$ независимы и одинаково распределены. Для справедливости усиленного закона больших чисел $\frac{1}{n}\,S_n \xrightarrow{\text{п.н.}} a$ необходимо и достаточно, чтобы существовало конечное математическое ожидание $\mathbf{M}\xi_k = a$.*

§ 33. Лемма Бореля–Кантелли

Лемма Бореля–Кантелли. *Пусть на вероятностном пространстве* $(\Omega, \mathcal{F}, \mathbf{P})$ *задана последовательность событий* $\{A_1, A_2, \ldots\}$ *и*

$$A^* = \left\{\omega \in \Omega : \ \sum\nolimits_{n \geq 1} \mathbb{I}(A_n) = \infty\right\},$$

где $\mathbb{I}(A_n)$ *– индикатор события* A_n*.*

а) *Если* $\sum\limits_{n\geq 1} \mathbf{P}\{A_n\} < \infty$, *то* $\mathbf{P}\{A^*\} = 0$.

б) *Если события* $A_1, A_2, \ldots$ *попарно независимы и* $\sum\limits_{n\geq 1} \mathbf{P}\{A_n\} = \infty$, *то* $\mathbf{P}\{A^*\} = 1$.

Доказательство. Пусть $\nu_T = \sum_{1\leqslant n\leqslant T} \mathbb{I}(A_n)$ — число одновременно происходящих событий в совокупности $A_1, \ldots, A_T$. Последовательность ν_T монотонно не убывает, поэтому при каждом $\omega \in \Omega$ существует $\lim_{T\to\infty} \nu_T \stackrel{\text{def}}{=} \nu = \sum_{n\geq 1} \mathbb{I}\{A_n\} \leqslant \infty$. Тогда

$$\mathbf{M}\nu_T = \sum_{k=1}^{T} \mathbf{P}\{A_k\}, \quad \mathbf{M}\nu = \sum_{k\geq 1} \mathbf{P}\{A_k\} \quad \text{и} \quad A^* = \{\nu = \infty\}.$$

а) Если ряд $\sum\limits_{n\geq 1} \mathbf{P}\{A_n\}$ сходится, то $\mathbf{M}\nu < \infty$ и поэтому $\mathbf{P}\{\nu < \infty\} = 1$, т. е. $\mathbf{P}\{A^*\} = 0$.

б) Покажем, что $\mathbf{D}\nu_T \leq \mathbf{M}\nu_T$. По общей формуле для дисперсии суммы случайных величин

$$\mathbf{D}\nu_T = \mathbf{D}\sum_{n=1}^{T} \mathbb{I}(A_n) = \sum_{n=1}^{T} \mathbf{D}\mathbb{I}(A_n) + \sum_{\substack{k,n=1\\k\neq n}}^{T} \operatorname{cov}(\mathbb{I}(A_k), \mathbb{I}(A_n)).$$

В силу попарной независимости событий A_n их индикаторы тоже попарно независимы, значит, $\operatorname{cov}(\mathbb{I}(A_k), \mathbb{I}(A_n)) = 0$ при $k \neq n$. Кроме того, $\mathbf{D}\mathbb{I}(A) = \mathbf{P}\{A\}(1 - \mathbf{P}\{A\}) \leqslant \mathbf{P}\{A\}$. Следовательно,

$$\mathbf{D}\nu_T = \sum_{k=1}^{T} \mathbf{D}\mathbb{I}(A_k) = \sum_{k=1}^{T} \mathbf{P}\{A_k\}(1 - \mathbf{P}\{A_k\}) \leqslant \mathbf{M}\nu_T.$$

Используя неравенство Чебышева и эту оценку, находим, что при любом $a < \mathbf{M}\nu_T$

$$\mathbf{P}\{\nu_T < a\} \leqslant \mathbf{P}\left\{\{\nu_T < a\} \bigcup \{\nu_T > \mathbf{M}\nu_T + (\mathbf{M}\nu_T - a)\}\right\} =$$
$$= \mathbf{P}\{|\nu_T - \mathbf{M}\nu_T| > \mathbf{M}\nu_T - a\} \leqslant \frac{\mathbf{D}\nu_T}{(\mathbf{M}\nu_T - a)^2} \leqslant \frac{\mathbf{M}\nu_T}{(\mathbf{M}\nu_T - a)^2}.$$

В случае б) $\nu_T \uparrow \nu$ и $\mathbf{M}\nu_T \uparrow \infty$ при $T \uparrow \infty$, значит, при любом $a < \infty$

$$\mathbf{P}\{\nu < a\} \leqslant \lim_{T\to\infty} \mathbf{P}\{\nu_T < a\} = 0 \quad \text{и} \quad \mathbf{P}\{\nu \geqslant a\} = 1 - \mathbf{P}\{\nu < a\} = 1.$$

Наконец,

$$\mathbf{P}\{\nu = \infty\} = \mathbf{P}\left\{\bigcap_{a=1}^{\infty}\{\nu \geqslant a\}\right\} = 1$$

как вероятность пересечения счетной совокупности событий, вероятность каждого из которых равна 1. Лемма доказана.

Список литературы

[1] Боровков А.А. Теория вероятностей. — М., Книжный дом "ЛИБРОКОМ 2009. — 656 с.

[2] Севастьянов Б.А. Курс теории вероятностей и математической статистики. — М., Наука, 1982. — 256 с.

[3] Феллер В. Введение в теорию вероятностей и ее приложения, тт.1,2. — М., Мир, 1984.

[4] Ширяев А.Н. Вероятность–1. — М., МЦНМО, 2007. – 552 с.