МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ «МЭИ»

А.Н. АРХАНГЕЛЬСКИЙ, А.А. БОБОДЖАНОВ,

С.Ф. КУДИН

ОСНОВЫ ТЕОРИЯ ВЕРОЯТНОСТЕЙ

МОСКВА Издательство МЭИ 2022

УДК 51
ББК 22.1

Рецензенты: канд. физ.-мат. наук, доц. Анисимова Т.В.,
докт. физ.-мат., зав. каф. Качалов В.И.

А.Н. Архангельский

Б 72 Основы теория вероятностей: учебное пособие / А.Н. Архангельский, А.А. Бободжанов, С. Ф. Кудин М.: Издательство МЭИ 2022. 139 с.

ISBN 978-5-7046-1863-8

В книге излагается курс теории вероятностей для технических вузов. Авторы стремились не перегружать учебник доказательствами теорем, оставляя лишь некоторые важные для понимания материала. Учебник рассчитан на изучение материала на втором курсе.

Авторы надеются , что книга будет полезна студентам технических вузов.

УДК 51
ББК 22.1

Содержание

Глава 1.Классическая вероятность

В этом параграфе приводятся основные формулы комбинаторики, которые в дальнейшем изложении будут использоваться для нахождения числа различных исходов в таких экспериментах, как подбрасывание игральных костей, различных урновых схемах и в других подобных примерах.

§1. Основные элементы комбинаторики

1. В основе всех получаемых ниже формул лежит так называемый ***основной комбинаторный принцип***. Пусть имеется несколько множеств элементов:

$$\{a_1, a_2, ..., a_t\}, \{b_1, b_2, ..., b_s\}, \{c_1, c_2, ..., c_k\}, \ldots .$$

Вопрос: сколькими способами можно составить множества $\{a, b, c, ...\}$ взяв из каждого исходного множества по одному элементу? Ответ на этот вопрос дают следующие рассуждения.

Элемент a из первого множество можно выбрать t способами, элемент b — из второго множества s способами, элемент c — из третьего множества k способами и т. д. Пару элементов ab можно составить ts способами. Это следует из табл. 1.1 в которой перечислены все способы такого выбора.

Таблица 1.1

b \ a	a_1	a_2	...	a_t
b_1	$a_1\ b_1$	$a_2\ b_1$	...	$a_t\ b_1$
b_2	$a_1\ b_2$	$a_2\ b_2$	...	$a_t\ b_2$
...			...	
b_s	$a_1\ b_s$	$a_2\ b_s$	...	$a_t\ b_s$

Способы выбора трех элементов abc перечислены в табл. 1.2.

Таблица 1.2

c \ ab	$a_1\ b_1$	$a_2\ b_1$	$a_3\ b_1$	...	$a_t\ b_s$
c_1	$a_1\ b_1\ c_1$	$a_2\ b_2\ c_1$	$a_3\ b_1\ c_1$	...	$a_t\ b_s\ c_1$
c_2	$a_1\ b_1\ c_2$	$a_2\ b_2\ c_2$	$a_3\ b_1\ c_2$	...	$a_t\ b_s\ c_2$
...				...	
c_k	$a_1\ b_1\ c_k$	$a_2\ b_2\ c_k$	$a_3\ b_1\ c_k$	...	$a_t\ b_s\ c_k$

В этой таблице k строк и st столбцов, поэтому искомые число способов выбора трех элементов abc равно $t \cdot s \cdot k.$ Продолжая рассуждать подобным образом, получим следующее утверждение.

Основной комбинаторный принцип. Если некоторый первый выбор можно сделать t способами, для каждого первого выбора некоторый второй можно сделать $\mathbf{s}$ способами, для каждой пары первых двух – третий выбор можно сделать k способами и т.д., то число способов для таких выборов равно $t \cdot s \cdot k \ldots.$

Сформулируем его в виде теоремы.

Теорема 1 (основной комбинаторный принцип). *Пусть имеется любое целое число m групп элементов, причём в i-ой группе* $(1 \leq i \leq m)$ *содержится* n_i *элементов. Будем выбирать из каждой группы по одному элементу. Пусть результат этого выбора обозначается в виде набора элементов* $d_1, d_2, \ldots, d_m$*, где* d_i *- выбранный элемент из i-ой группы. Тогда общее число различных таких наборов равно* $N = n_1 \cdot n_2 \cdot \ldots \cdot n_m$.

2. Рассмотрим кратко основные понятия комбинаторики.

i. Предположим, что имеется группа из n занумерованных от 1 до n элементов. Разместим все эти элементы в виде строки $(a_1, a_2, \ldots, a_n)$, где a_i - порядковый номер элемента из группы, размещённый на i-ой позиции. Будем называть этот объект ***размещением***. Возникает естественный вопрос: каково общее количество различных размещений? Для ответа на этот вопрос, считаем, что имеется n пустых занумерованных мест, куда будут размещаться элементы из заданной группы. Заполнение пустых мест производится по возрастанию номеров мест. Ясно, что на первое место можно разместить любой элемент из n, имеющихся в группе. После заполнения первого места, в группе остаётся $n-1$ элементов и именно таким количеством способов можно заполнить второе место. Продолжая эти рассуждения, замечаем, что k-е место заполняется $n-k+1$ способом. Теперь воспользовавшись теоремой 1, получаем, что общее количество различных размещений

$$N = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1 = n!$$

.

ii. Предположим, что также, как и выше, имеется группа из n занумерованных от 1 до n элементов и пусть имеется $k,\ k \leq n$ от 1 до k мест, куда будут размещаться элементы из заданной группы. Объект $(b_1, b_2, \ldots, b_k)$, где b_i - порядковый номер элемента из заданной группы, размещённый на i-ой позиции будем называть *выборкой объёма* k *из* n *элементов* или короче ***выборкой из n по k***.

Будем говорить, что выборка из n по k *упорядоченная*, если учитываются порядок, в котором эти элементы стоять в выборке. В противном случае будем говорить, что выборка *неупорядоченная.*

Найдём сначала количество упорядоченных выборок из n по k. Действуем точно так же, как выше в *i*. Считаем общее количество выборок $(b_1, b_2, \ldots, b_k)$. Ясно, что элемент b_1 можно выбрать n способами - именно такое количество элементов во всей группе. После выбора b_1 в группе остаётся $n-1$ элементов и именно столько способов выбрать элемент b_2. И так далее, продолжаем подсчитывать количество способов выбора всех последующих элементов. Перед выбором последнего элемента b_k в группе остаётся $n-k+1$ элементов и, поэтому, этот последний элемент выборки можем выбрать $n-k+1$ способами. Остаётся применить теорему 1, чтобы получить общее количество упорядоченных выборок из n по k

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

Это число называется ***числом размещений из n по k***. Заметим, что $A_n^n = n!$, поскольку $0! = 1$. Это есть, как отмечено выше, общее число размещений или кратко общее число перестановок n различимых элементов.

iii. Пусть вся схема построения будет такой же, как и выше в пункте *ii*. Будем искать теперь количество неупорядоченных выборок из n по k. Из определения ясно, что любая перестановка фиксированного набора элементов даёт одну и туже неупорядоченную выборку. Пусть C_n^k обозначает количество всех неупорядоченных выборок из n элементов по k. Из пункта *i* ясно, что для каждой выборки данного состава (состоящей из k элементов) можно получить $k!$ выборок, отличающихся друг от друга только порядком следования элементов. То есть, число выборок, различающихся ещё и порядком, в $k!$ раз больше, чем число неупорядоченных выборок, различающихся только составом. Поэтому, $A_n^k = k! \cdot C_n^k$ или

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

Это число обычно называют ***числом сочетаний из n элементов по k***. Из определения, в частности, следует, что $C_n^k = C_n^{n-k}$, $C_n^0 = 1$, где $0! = 1$.

iv. Все рассмотренные выше выборки были ***бесповторными***, т.е. каждый элемент из заданной группы в конкретной выборке мог встречаться не более одного раза. Рассмотрим теперь случай, когда группа содержит n различных (различимых) элементов и будем создавать объект, который схематично обозначим как $(c_1, c_2, \ldots, c_k)$, где все $c_i, i = 1, \ldots, k$ выбираются из заданной группы, причём компоненты c_i могут повторяться. Такой объект будем называть ***повторной выборкой объёма k***. Найдём общее количество таких повторных выборок. Ясно, что c_1 можно выбрать n способами из исходной группы, и c_2 тоже можно выбрать n способами и т.д., т.е. каждый компонент c_i можно выбрать n способами. Воспользовавшись теперь теоремой 1, заключаем, что общее количество повторных выборок равно n^k.

v. В заключение этого параграфа рассмотрим один пример. Пусть имеется конечное множество, содержащее n различных элементов. Подсчитаем, сколько всего неупорядоченных подмножеств у данного множества? Расположим все элементы заданного множества в некотором порядке, т.е. присвоим каждому элементу исходного множества по одному порядковому номеру от *1* до n. Ясно, что любое неупорядоченное подмножество можно получить из исходного множества, например, путём вычёркивания ненужных элементов. А можно ещё поступить таким образом: выписать для каждого конкретного подмножества бинарный вектор (это вектор, состоящий только из нулей и единиц) длины n $(\boldsymbol{d_1, d_2, \ldots, d_n})$, где $\boldsymbol{d_i = 1}$, если i-й элемент содержится в этом подмножестве и $\boldsymbol{d_i = 0}$, если не содержится для всех $\boldsymbol{i = 1, 2, \ldots, n}$. Таким образом, задача нахождения числа всех неупорядоченных подмножеств у данного множества сводится к нахождению числа всех бинарных векторов длины n. Осталось воспользоваться результатом предыдущего пункта *iv* и понять, что искомое число равно $\boldsymbol{2^n}$.

§2. Основные понятия теории вероятностей

В теории вероятностей изучаются закономерности, проявляющиеся в случайных экспериментах. Случайным обычно называют такой эксперимент, результат которого объективно нельзя предсказать заранее — основное свойство, которое отличает случайность от детерминированности (не случайности). Не все случайные эксперименты могут быть изучены методами теории вероятностей, а лишь те, которые обладают так называемым свойством *статистической устойчивости*: если некоторое событие может произойти или нет в результате некоторого эксперимента, то отношение числа экспериментов, в которых данное событие произошло к общему числу экспериментов, должно стремится к некоторому фиксированному числу с ростом общего количества экспериментов. Это число показывает «степень возможности» появления рассматриваемого события. В дальнейшем изложении всегда будем предполагать, что все рассматриваемые эксперименты удовлетворяют свойству статистической устойчивости.

1. Пространство элементарных событий. Операции над событиями

Определение 1. Пространство элементарных событий $\boldsymbol{\Omega}$ (омега большая)— это произвольное множество, которое в случайном эксприменте рассматривается как совокупность взаимо исключающих исходов данного эксперимента. Элементы этого множества называют элементарными событиями и обозначают $\boldsymbol{\omega}$ (омега малая) с индексами или без них.

Определение 2. Событиями будем называть некоторое подмножество $\boldsymbol{\Omega}$. Всегда будем события обозначать заглавными латинскими буквами. Говорят, что в результате эксперимента произошло событие $\boldsymbol{A \subset \Omega}$, если в эксперименте произошло одно из элементарных событий, входящих в $\boldsymbol{A}$.

Пример 1. Подбрасывается одна игральная кость один раз. Пространством элементарных событий будет множество $\boldsymbol{\Omega = \{1, 2, 3, 4, 5, 6\}}$, элементарные события соответствуют числу выпавших очков. Примеры событий: $\boldsymbol{A = \{2\}}$ — выпала двойка;

$\boldsymbol{A = \{1,3,5\}}$ — выпало нечетное число очков; $\boldsymbol{A = \{2,4,6\}}$ — выпало четное число очков.

Пример 2. Два раза подбрасывается одна игральная кость. Пространством элементарных событий здесь будет совокупность из 36 упорядоченных пар чисел $\boldsymbol{(i,j)}$, в которой i и j есть число очков, выпавших при первом и втором подбрасывании соответственно.

Примеры событий: $\boldsymbol{A = \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)\}}$ — при первом подбрасывании выпадает единица; $\boldsymbol{A = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}}$ — при первом и втором подбрасываниях выпадает одинаковые очки.

Пример 3. Подбрасывается монета. Будем считать, что результатом одного подбрасывания могут быть два исхода: о (орёл) и р (решка) (любые другие исходы, как то: монета встала на ребро, пропала и т.д. исключаются). Монета подбрасывается до тех пор, пока не выпадет орёл. Пространство элементарных событий состоит из бесконечного, числа элементарных событий: $\boldsymbol{\Omega} = \boldsymbol{\{}$*о, ро, рро, рррo, рррро, ррррро, …* $\boldsymbol{\}}$, где р и о обозначают выпадение решки и орла при одном подбрасывании, соответственно.

Определение 3.

i. Достоверным называется событие, которое обязательно происходит в результате эксперимента, т.е. единственное событие, включающее все элементарные события — событие $\boldsymbol{\Omega}$.

ii. Невозможным называется событие, которое не может произойти в результате эксперимента, т.е. событие, не содержащее ни одного элементарного исхода в данном эксприменте («пустое множество» $\boldsymbol{\varnothing}$). Ясно, что всегда $\boldsymbol{\varnothing \subset \Omega}$.

Определение 4. Пусть A и B — два выбранных события.

i. Объединением $\boldsymbol{C = A \cup B}$ событий A и B называется событие C, которое происходит тогда и только тогда, когда произошло либо A, либо B, либо оба события A и B одновременно. Обычно говорят, что событие C есть событие A **или** событие B.

ii. Пересечением $\boldsymbol{D = A \cap B}$ событий A и B называется событие D, которое происходит тогда и только тогда, когда произошли оба события A и B одновременно. Обычно говорят, что событие D есть событие A **и** событие B.

iii. Разностью $\boldsymbol{E = A \backslash B}$ событий A и B называется событие E, которое происходит тогда и только тогда, когда произошло событие A, но не произошло событие B. Обычно говорят, что событие E есть событие A **и не** событие B.

iv. Противоположным (или дополнительным) к событию A называется событие

$$\boldsymbol{F = \bar{A} = \Omega \backslash A,}$$

которое происходит тогда и только тогда, когда событие A в результате эксперимента не произошло. Обычно говорят, что событие F есть **не** событие A.

v. Симметрической разностью $\boldsymbol{E = A \Delta B}$ событий A и B называется событие E, которое происходит тогда и только тогда, когда происходит событие A и не происходит событие B или когда происходит событие B и не происходит событие A. Ясно, что

$\boldsymbol{E = (A \backslash B) \cup (B \backslash A)}$. Например, в случае, когда два игрока играют в шахматы и события A=**{***выиграл первый игрок***}**, B=**{***выиграл второй игрок***}**, то событие E означает отсутствия ничьей.

Легко проверить, что, например,

$\boldsymbol{A \cup B = B \cup A}$, $\boldsymbol{A \cap B = B \cap A}$, $\boldsymbol{A \cap (B \cup C) = (A \cap B) \cup (A \cap C)}$, $\boldsymbol{\bar{\bar{A}} = A}$,
$\boldsymbol{A \backslash B = A \cap \bar{B}}$, $\boldsymbol{A \Delta B = (A \cap \bar{B}) \cup (\bar{A} \cap B)}$=$\boldsymbol{(A \cup \bar{B}) \backslash (A \cap B)}$**.**

Определение 5.

i. События A и B называются *совместными*, если события A и B могут произойти одновременно. Например, при однократном подбрасывании игральной кости событие A=*{выпала шестёрка}* и событие B=*{выпало чётное число очков}* являются совместными.

ii. События A и B называются *несовместными*, если $\boldsymbol{A \cap B = \varnothing}$. Например, при однократном подбрасывании игральной кости событие A=*{выпала шестёрка}* и событие B=*{выпало нечётное число очков}* являются несовместными.

iii. События $\boldsymbol{A_1, \ldots, A_n}$ называются *попарно несовместными*, если для любых $\boldsymbol{i \neq j}$, $\boldsymbol{1 \leq i, j \leq n}$, события $\boldsymbol{A_i}$ и $\boldsymbol{A_j}$ несовместны.

iv. Говорят, что событие А влечёт событие В, и записывают это в виде $\boldsymbol{A \subset B}$, если всегда, как только происходит событие A, происходит и событие B. Это означает, что любое элементарное событие, входящее в A, одновременно входит и в событие B.

2. Вероятность на дискретном пространстве элементарных событий

Будем рассматривать дискретные пространства элементарных событий, которые состоят из конечного или бесконечного числа элементарных событий: $\boldsymbol{\Omega = \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\}}$.

Определение 6. Поставим каждому элементарному исходу $\boldsymbol{\omega_i \in \Omega}$ в соответствие число $\boldsymbol{p(\omega_i) \in [0, 1]}$ так, чтобы $\boldsymbol{\sum\limits_{\omega_i \in \Omega} p(\omega_i) = 1}$. Назовем число $\boldsymbol{p(\omega_i)}$ *вероятностью* элементарного события $\boldsymbol{\omega_i}$. Вероятностью события $\boldsymbol{A \subseteq \Omega}$ называется число, равное сумме вероятностей элементарных событий, входящих в событие A, т.е. $\boldsymbol{P(A) = \sum\limits_{\omega_i \in A} p(\omega_i)}$.

Очевидно, в случае дискретного пространства элементарных событий, справедливы следующие свойства вероятности:

i. $\boldsymbol{0 \leq P(A) \leq 1}$**;**
ii. $\boldsymbol{P(\Omega) = 1}$**;**
iii. $\boldsymbol{P(\varnothing) = 0}$**;**
iv. $\boldsymbol{P(\bar{A}) = 1 - P(A)}$**;**
v. $\boldsymbol{P(A \cup B) = P(A) + P(B)}$, если события A и B несовместны;
vi. $\boldsymbol{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$ для любых событий A и B;
vii. если $\boldsymbol{A \subseteq B}$, то $\boldsymbol{P(A) \leq P(B)}$.

3. Классическое определение вероятностей

Пусть пространство элементарных событий состоит из конечного числа N элементарных событий: $\boldsymbol{\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}}$ и пусть все элементарные события будут равновозможными, т.е $P(\omega_1) = P(\omega_2) = \dots = P(\omega_N) = 1/N$. Такого типа пространство элементарных событий будет, например, в случае, если описывать эксперимент с подбрасыванием правильной игральной кости. Под правильностью игральной кости здесь понимается то, что кость сделана из однородного материала постоянной плотности с центром тяжести, находящимся точно в центре кости. В этом случае, при произвольном подбрасывании (так, чтобы кость, например, закрутилась в воздухе, а не просто положить кость выбранной гранью на горизонтальную поверхность) будем иметь шесть различных элементарных событий: $\omega_i = \{$выпадение грани, на которой i очков$\}$, $P(\omega_i) = 1/6$, $i = 1, 2, \dots, 6$.

Если событие $A = \{\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_k}\}$ состоит из k элементарных событий, то вероятность этого события A равняется отношению k/N:

$$P(A) = P(\omega_{i_1}) + \ldots + P(\omega_{i_k}) = k \cdot \frac{1}{N} = \frac{\mu(A)}{\mu(\Omega)},$$

где функция $\mu(A)$ обозначает *считающую меру* события A, т.е. число элементарных событий события A.

Определение 7. Скажем, что эксперимент удовлетворяет *классической вероятностной схеме*, если пространство элементарных событий состоит из конечного числа $\mu(\Omega) = N$ равновероятных элементарных событий. В этом случае вероятности любого события A вычисляется по формуле $P(A) = \frac{\mu(A)}{\mu(\Omega)}$, называемой *классическим определением вероятности*, т.е. вероятность события A равна отношению числа благоприятствующих элементарных событий событию A к общему числу элементарных событий.

§3. Урновые схемы

Пусть имеется некоторый ящик, который в классических задачах теории вероятностей принято называть урной.Пусть урна содержит n занумерованных шаров. Выбирают произвольно из этой урны k шаров. Сколькими способами можно выбрать k шаров из урны, первоначально содержащей n неразличимых на ощупь шаров. Для однозначного ответа необходимо несколько уточнений. Рассмотрим следующие возможные схемы выбора.

i. Выбор без возвращения: выбранные шары снова в урну не возвращаются, и в полученной выборке не могут встречаться одни и те же номера (выборка без повторений). Как показано выше в §1, число таких *упорядоченных* выборок равно A_n^k, а *неупорядоченных* выборок равно C_n^k.

ii. Выбор с возвращением: каждый извлечённый из урны шар возвращается в ту же урну. Поэтому, каждый из k извлекаемых шаров выбирается из первоначально содержащегося набора из n шаров в урне. В полученной выборке, состоящей из k номеров шаров, могут встречаться одни и те же номера (выборка с повторениями). И как показано выше в §1, число таких упорядоченных выборок равно n^k.

Найдём теперь число *неупорядоченных* выборок в той же урновой схеме с возвращениями. Задача состоит в подсчёте числа различных исходов $k_1, k_2, \ldots, k_n$, где k_i обозначает число извлечений шара с номером i. Ясно, что $k_1 + k_2 + \cdots + k_n = k$ и если $k_i = 0$ для некоторого i, то это означает, что шар с номером i из урны не извлекался. Проще всего это количество исходов подсчитать, введя следующую вспомогательную схему. Представим, что имеется прямоугольная коробка с $n-1$ внутренними передвижными перегородками, которые создают n занумерованных коробочек, расположенных в ряд. Дальше, предположим, что в нашем распоряжении есть k одинаковых шаров, которые раскладываются по этим коробочкам. Если в коробочку с номером i попадают k_i шаров, то, фактически для рассматриваемой выше схеме извлечений шаров из урны это означает, что шар с номером i появлялся k_i раз. Ясно, что если $k_i = 0$, то i-я коробочка пуста. Верно также и обратно, что если шар с номером i извлечён k_i раз, то в коробочке с номером i окажется ровно k_i шаров. Для подсчёта числа размещений шаров по ящичкам можно считать, что $n-1$ перегородок переставляются между k шарами (число неупорядоченных выборок *из* $n+k-1$ - общее количество перегородок и шаров *по* $n-1$ - число перегородок), разложенными в прямоугольном ящике в ряд, или, что тоже самое шары перекладывают между перегородками (число неупорядоченных выборок *из* $n+k-1$ *по* k) Из §1 *iii* получаем, что это число есть $C_{n+k-1}^{n-1} = C_{n+k-1}^{k}$.

Пример 4. Рассмотрим случай упорядоченных выборок в схеме выбора N занумерованных шаров из урны с возвращением. Предположим, что все выборки равновозможны. Дополнительно предположим, что все шары с номерами $1, 2, \ldots, M$ $M \leq N$ окрашены в белый цвет, а остальные шары - в чёрный цвет. Требуется *найти вероятность того, что в выборке объёма* n *окажется ровно* $m,\ 0 \leq m \leq n$ *белых шаров.*

Решение. Пусть событие A=\{*в выборке объёма n окажется ровно m белых шаров*\}. Для решения задачи воспользуемся определением классической вероятности. Ясно, что число всех элементарных событий здесь равно N^n, а число благоприятствующих элементарных событий можно посчитать следующим образом. Для начала предположим, что извлекаются только белые шары (шары, с номерами от *1*-го до M-го) на шагах от первого до m-го (число различных возможностей равно M^m), а с шага $(m+1)$-го до последнего n-го извлекаем только щары с номерами от $(M+1)$-го до N-го - чёрные шары (число таких возможностей равно $(N-M)^{n-m}$. Поскольку такой порядок не единственный, то число других возможных порядков извлечения белых и чёрных шаров равно C_n^m - числу неупорядоченного размещения m белых шаров среди чёрных. Применив теперь основной комбинаторный принцип, получим, что число благоприятствующих элементарных событий для события A равно $C_n^m \cdot M^m (N-M)^{n-m}$. Поэтому искомая вероятность события будет равна

$$P(A) = \frac{C_n^m M^m (N-M)^{n-m}}{N^n} = C_N^m \cdot p^k \cdot (1-p)^{n-k},$$

где $p = M/N$ - доля белых шаров в урне (или вероятность извлечь белый шар при одном извлечении).

iii. Гипергеометрическое распределение

Пример 5. Рассмотрим случай упорядоченных выборок в схеме выбора N занумерованных шаров из урны без возвращения и предположим, что все выборки равновозможны. Как и в *Примере 4*, предположим, что все шары с номерами $1, 2, \ldots, M$ $M \leq N$ окрашены в белый цвет, а остальные шары - в чёрный цвет. Требуется *найти вероятность того, что в выборке объёма n окажется ровно m, $0 \leq m \leq n$ белых шаров.*

Решение.

Введём событие B={*в выборке объёма n окажется ровно m белых шаров*}, вероятность которого надо найти. Результатом эксперимента является набор из n шаров. При этом можно учитывать или не учитывать порядок следования шаров.

a. Выбор с учётом порядка. Ясно, что число всех элементарных событий равно $\mu(\Omega) = A_N^n$, где, как и выше, $\mu(\Omega)$ обозначает считающую меру, обозначающую количество элементарных событий в Ω. При подсчёте числа благоприятных элементарных событий нужно учесть как число способов выбрать нужное число шаров, так и число способов расположить эти шары среди n. Нужно посчитать число способов выбрать m мест среди n (равное C_n^m), затем число способов разместить на этих m местах из M белых шаров (равное A_M^m), а затем число способов разместить на оставшихся $n - m$ местах чёрные шары из общего количества $N - M$ чёрных шаров (равное A_{N-M}^{n-m}). Перемножая (по основному комбинаторному принципу) эти числа, получим

$$\mu(B) = C_n^m \cdot A_M^m \cdot A_{N-M}^{n-m}, \qquad P(B) = \frac{C_n^m \cdot A_M^m \cdot A_{N-M}^{n-m}}{A_N^n} =$$

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

b. Выбор без учёта порядка. Общее число элементарных событий есть число подмножеств, содержащих n элементов множества, состоящего из N элементов, то есть $\mu(\Omega) = C_N^n$ (§1 п.2 *iii*). Событию B благоприятствует появление любого набора, содержащего m белых шаров и $n - m$ чёрных шаров. Число благоприятных элементарных событий равно (по основному комбинаторному принципу) произведению числа способов выбрать m белых шаров из M и числа способов выбрать $(n - m)$ чёрных шаров из $N - M$: $\mu(B) = C_M^m \cdot C_{N-M}^{n-m}$. Вероятность события B равна $P(B) = \frac{C_M^m \cdot C_{N-M}^{n-m}}{C_N^n}$.

Таким образом, видим, что для данного примера 5 не важно, какую выборку рассматривать - упорядоченную или не упорядоченную.

Определение 8. Набор вероятностей

$$P(A_{m,n,M,N}) = \frac{C_M^m \cdot C_{N-M}^{n-m}}{C_N^n},$$

$$m = 0, 1, \ldots, \min\{n, M\}, \quad n - m \leq N - M$$

называется ***гипергеометрическим распределением***.

Гипергеометрическую схему можно обобщить следующим образом. Пусть имеется урна, в которой находятся шары, окрашенные в k различных цветов, причём M_i - шаров i-го цвета и пусть $\sum\limits_{i=1}^{k} M_i = M$. Пусть событие A означает, что в выборке без возвращения объёма n, шаров i-го цвета будет n_i, $\quad i = 1, \ldots, k, \quad \sum\limits_{i=1}^{k} n_i = n$. Тогда вероятность этого события равна

$$P(A) = \frac{C_{M_1}^{n_1} \cdot C_{M_2}^{n_2} \cdot \ldots \cdot C_{M_k}^{n_k}}{C_M^n} = \frac{\prod\limits_{i=1}^{n} C_{M_i}^{n_i}}{C_M^n}.$$

§4. Геометрические вероятности

Пусть задана некоторая область $\Omega \in R^l$ ($l = 1, 2, 3$ - на прямой, на плоскости или в трёхмерного пространства). Пусть «мера» Ω (длина, площадь, объём, соответственно) конечна. Пусть случайный эксперимент состоит в том, что наудачу бросается (выбираются) в эту область точка. Здесь *наудачу* означает, что вероятность попадания точки в любое подмножество $A \in \Omega$ не зависит от формы или расположения A внутри Ω, а зависит лишь от «меры» области A (если A *измеримо*, т.е. имеет конечную «меру»).

Определение 9. Будем говорить, что определена геометрическая вероятность, если пространство элементарных событий Ω задаётся некоторой геометрической измеримой областью из R^l и для любого события A, которое представляется измеримым подмножеством из Ω, вероятность этого события определяется следующим образом:

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

где $\mu(\cdot)$ обозначает меру соответствующего множества.

Под *мерой* здесь будем понимать длину, площадь, объём для множеств из соответствующих пространств $R^l, \quad l = 1, 2, 3.$ Этими случаями ограничимся. *Измеримым* множеством, как отмечено выше, будем называть множество с конечной мерой.

Пример 6. На отрезке длины l наудачу выбирают две точки. Какова вероятность того, что из трёх отрезков, на которые делится исходный отрезок выбранными точками, можно составить треугольник?

Решение. Обозначим длину первого отрезка x, длину второго – y. Ясно, что третий отрезок имеет длину $l - x - y$. Длина первого отрезка может быть $0 \le x \le l$, длина второго – $0 \le y \le l$ и $0 \le x + y \le l$. На плоскости изобразим область, которая задаёт пространство элементарных событий Ω. Для этого в выбранной прямоугольной системе координат отметим множество точек, удовлетворяющих указанным неравенствам. Ясно, что эта область представляет из себя все точки, лежащие внутри прямоугольного равнобедренного треугольника, лежащего в первом квадранте. Для того, чтобы из трёх отрезков можно было составить треугольник, нужно, чтобы сумма длин любых двух отрезков была бы не меньше длины третьего отрезка. Поэтому, имеем три неравенства:

$$x + y \ge l - x - y, \qquad x + l - x - y \ge y, \qquad y + l - x - y \ge x,$$

или

$$x + y \ge l/2, \qquad y \le l/2, \qquad x \le l/2.$$

Ясно, что прямые $y = l/2 - x, \quad y = l/2, \quad x = l/2$ разбивают треугольник Ω на четыре одинаковых прямоугольных равнобедренных треугольника, из которых центральный является областью события $A =${из трёх отрезков можно составить треугольник}, т.е координаты любой внутренней точки этого треугольника задают три отрезка, как описано выше, из которых можно построить треугольник. Поэтому, по определению геометрической вероятности получаем,

$$P(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{\frac{1}{2} \cdot (\frac{l}{2})^2}{\frac{1}{2} \cdot l^2} = \frac{1}{4}.$$

Пример 7 (задача Бюффона). На плоскость, разграфлённую параллельными прямыми, отстоящими друг от друга на расстоянии 2a наудачу бросается игла длиною $2l < 2a$. Какова вероятность того, что игла пересечёт одну из проведенных прямых?

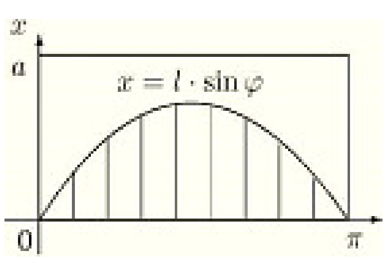

Рис. 1.1

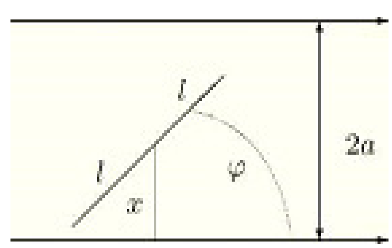

Рис. 1.2

Решение. Что означает *наудачу брошена игла*? Выберем прямоугольную систему координат. Возможные положения иглы на плоскости полностью определяются положением середины иглы и углом поворота иглы относительно какой-либо оси. Причём две эти переменные (положение центра и угол поворота) меняются независимо друг от друга. Обозначим через $x \in [0, a]$ (см., рис 1.1 и рис 1.2) расстояние от середины иглы до ближайшей прямой, а через $\varphi \in [0, \pi]$ – угол между прямой и иглой. Множество возможных положений иглы целиком определяется выбором наудачу точки из прямоугольника $\Omega = [0, \pi] \times [0, a]$ (обозначает множество упорядоченных пар). Игла пересекает ближайшую прямую, если координаты выбранной наудачу точки удовлетворяют неравенству: $x \leq l \cdot \sin(\varphi)$. Площадь области $A \subset \Omega$, точки которой удовлетворяют такому неравенству, равна

$$\mu(A) = \int_0^{\pi} l \cdot \sin(\varphi) d\varphi = -l\cos(\varphi)\Big|_0^{\pi} = 2l.$$

Поскольку $\mu(\Omega) = a\pi$, то, по определению геометрической вероятности, получаем

$$P(A) = \frac{2l}{a\pi}.$$

Пример 8 (Парадокс Бертрана). В круге единичного радиуса наудачу выбирается хорда. Какова вероятность того, что её длина будет больше, чем длина стороны вписанного в круг правильного треугольника?

Решение. Отметим, что формулировка задачи не корректна с математической точки зрения. «Фраза наудачу выбирается хорда в круге» (см., рис 1.3) может быть по-разному понято с точки зрения геометрического определения вероятности, что и будет показано ниже.

Есть по крайней мере три способа «выбрать наудачу хорду в круге».

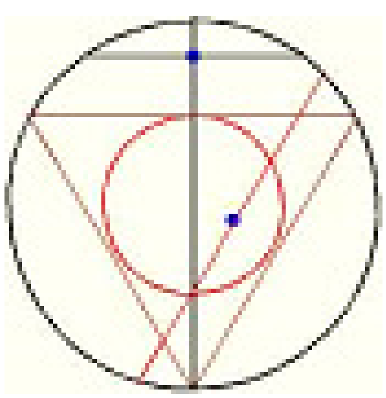

Рис. 1.3

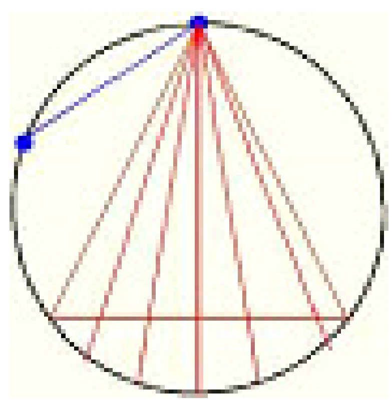

Рис. 1.4

i. Зафиксируем одну точку (конец хорды) на окружности и выберем наудачу на окружности другую точку (второй конец хорды). Ясно, что $\Omega = [0, 2\pi]$, а благоприятными являются положения второй точки на отрезке $[\frac{2}{3}\pi, \frac{4}{3}\pi]$. Вероятность получить «длинную» хорду равна $\frac{1}{3}$.

ii. Существует ровно одна хорда, для которой данная точка в круге является её серединой . Поэтому (см., рис 1.4) можно выбирать наудачу хорду, бросая наудачу точку в круг – середину хорды. Здесь $\mathbf{\Omega}$ — круг радиуса 1, $\boldsymbol{\mu(\Omega) = \pi}$, а благоприятными являются положения середины хорды внутри вписанного в треугольник круга радиусом $\frac{1}{2}$. Вероятность получить «длинную» хорду равна отношению площадей кругов, то есть $\frac{1}{4}$.

iii. Можно ограничиться рассмотрением только хорд, перпендикулярных какому-либо диаметру (остальные могут быть получены поворотом). То есть, эксперимент может состоять в выборе середины хорды наудачу на диаметре круга — отрезке длиной 2. Благоприятными будут положения середины хорды на отрезке длиной 1. Искомая вероятность для такого эксперимента равна $\frac{1}{2}$ (см., рис 1.4).

Глава 2. Аксиоматика теории вероятностей Колмогорова

§1. σ-алгебра событий

Пусть $\mathbf{\Omega}$ — пространство элементарных событий некоторого случайного эксперимента, то есть множество произвольной природы. Определим множества таких подмножеств $\mathbf{\Omega}$, которые будут называться событиями, и затем зададим вероятность как функцию от события, т.е. функцию, определённую только на множестве событий. То есть событием будет называться не любые подмножество $\mathbf{\Omega}$, а лишь подмножество из некоторой совокупности подмножеств $\boldsymbol{\mathcal{A}}$. При этом необходимо, чтобы это множество $\boldsymbol{\mathcal{A}}$ подмножеств $\mathbf{\Omega}$ было «замкнуто» относительно введённых выше в §2 Главы 1 операций над событиями, то есть чтобы объединение, пересечение, дополнение, разность и симметрическая разность событий из $\boldsymbol{\mathcal{A}}$ снова давало событие из той же совокупности $\boldsymbol{\mathcal{A}}$.

Определение 10. Множество $\boldsymbol{\mathcal{A}}$, состоящее из подмножеств множества $\mathbf{\Omega}$ называется **σ**-алгеброй событий (**σ**-алгеброй подмножеств $\mathbf{\Omega}$), если выполняются следующие условия:

(a) $\mathbf{\Omega} \in \boldsymbol{\mathcal{A}}$ (**σ**-алгебра событий содержит достоверное событие);

(b) если $\boldsymbol{A} \in \boldsymbol{\mathcal{A}}$, то $\bar{\boldsymbol{A}} \in \boldsymbol{\mathcal{A}}$ (вместе с любым событием **σ**-алгебра содержит ему противоположное событие);

(c) если $\boldsymbol{A_1, A_2, \ldots} \in \boldsymbol{\mathcal{A}}$, то $\bigcup\limits_{i=1}^{\infty} \boldsymbol{A_i} \in \boldsymbol{\mathcal{A}}$ (вместе с любым конечным или счётным набором событий **σ**-алгебра содержит и их объединение).

Условия *(a)*, *(b)* и *(c)* называются *аксиомами* **σ**-*алгебры.*

Проверим, что этого набора аксиом достаточно для замкнутости совокупности $\boldsymbol{\mathcal{A}}$ относительно других операций над событиями.

Свойства σ-алгебры.

i. $\varnothing \in \boldsymbol{\mathcal{A}}$ (**σ**-алгебра событий содержит невозможное событие).

Действительно, в силу аксиомы *(a)* $\mathbf{\Omega} \in \boldsymbol{\mathcal{A}}$ и в силу аксиомы *(b)* имеем $\varnothing = \bar{\mathbf{\Omega}} \in \boldsymbol{\mathcal{A}}$.

ii. При выполнении *(a)* и *(b)* свойство *(c)* эквивалентно свойству

(d) если $A_1, A_2, \dots \in \mathcal{A}$, то $\bigcap\limits_{i=1}^{\infty} A_i \in \mathcal{A}$ (вместе с любым конечным или счётным набором событий σ-алгебра событий содержит и их пересечение).

Покажем, что при выполнении *(a)* и *(b)* из *(c)* следует *(d)*. Если $A_1, A_2, \dots \in \mathcal{A}$, то при всех $i = 1, 2, \dots$ из свойства *(b)* следует $\overline{A_i} \in \mathcal{A}$. Тогда из *(c)* следует, что $\bigcup\limits_{i=1}^{\infty} \overline{A_i} \in \mathcal{A}$, и, по *(b)*, дополнение к этому множеству также принадлежит $\mathcal{A}$, т.е. $\overline{\bigcup\limits_{i=1}^{\infty} \overline{A_i}} \in \mathcal{A}$. Но, в силу закона де'Моргана получаем $\overline{\bigcup\limits_{i=1}^{\infty} \overline{A_i}} = \bigcap\limits_{i=1}^{\infty} A_i \in \mathcal{A}$, что и требовалось. Рассуждения в обратную сторону проводятся аналогично.

iii. Если $A, B \in \mathcal{A}$, то $A \backslash B \in \mathcal{A}$.

Действительно, $A \backslash B = A \cap \bar{B} \in \mathcal{A}$, так как $A \in \mathcal{A},\ \bar{B} \in \mathcal{A}$, а из *(d)* пересечение событий тоже принадлежит $\mathcal{A}$.

iv. Если $A, B \in \mathcal{A}$, то $A \triangle B \in \mathcal{A}$.

Действительно, $A \backslash B \in \mathcal{A}$ и $B \backslash A \in \mathcal{A}$ и в силу *(c)* объединение двух событий есть событие, т.е. $A \triangle B = (A \backslash B) \bigcup (B \backslash A) \in \mathcal{A}$.

Пример 9. Пусть $\Omega = \{1, 2, 3, 4, 5, 6\}$ — пространство элементарных событий при бросании игральной кости. Несложно показать, что следующие совокупности подмножеств Ω будут являться σ-алгебрами:

i. $\mathcal{A} = \{\Omega, \varnothing\} = \{\{1, 2, 3, 4, 5, 6\}, \varnothing\}$ — тривиальная σ-алгебра.

ii. $\mathcal{A} = \{\Omega, \varnothing, \{1\}, \{2, 3, 4, 5, 6\}\} = \{\{1, 2, 3, 4, 5, 6\}, \varnothing, \{1\}, \{2, 3, 4, 5, 6\}\}$.

iii. $\mathcal{A} = \{\Omega, \varnothing, A, \bar{A}\} = \{\{1, 2, 3, 4, 5, 6\}, \varnothing, A, \bar{A}\}$, где A — произвольное подмножество Ω (в *ii* $A = \{1\}$).

iv. $\mathcal{A}$ — множество всех подмножеств Ω. Ясно, что в $\mathcal{A}$ всего $2^6 = 64$ события.

В этом параграфе был определён класс $\mathcal{A}$ подмножеств пространства элементарных событий Ω, который называется σ-алгеброй событий, причём применение счётного числа операций (объединение, пересечение, дополнение, симметрическая разность) к множествам из $\mathcal{A}$ снова даёт множество из $\mathcal{A}$, т.е не выводит за рамки этого класса. Множества $A \in \mathcal{A}$ называются *событиями.*

В следующем параграфе будет определено понятие *вероятности* как функции, заданной на множестве событий, т.е. на σ-алгебре событий $\mathcal{A}$.

§2. Вероятность

Начнём этот параграф с общего определения меры.

Определение 11. Пусть Ω — некоторое множество и $\mathcal{A}$ — σ-алгебра его подмножеств. Функция $\mu : \mathcal{A} \to R$ называется *мерой* на измеримом пространстве $(\Omega, \mathcal{A})$, если эта функция удовлетворяет условиям:

(j) Для любого множества $\boldsymbol{A \in \mathcal{A}}$ его мера неотрицательна: $\boldsymbol{\mu(A) \geq 0}$.

(jj) Для любого не более, чем счётного набора попарно непересекающихся множеств

$$\boldsymbol{A_1, A_2, A_3, \ldots \in \mathcal{A}: \; A_i \cap A_j = \varnothing \quad \forall i \neq j,}$$

мера их объединения равна сумме их мер:

$$\boldsymbol{\mu\left(\bigcup_{i=1}^{N} A_i\right) = \sum_{i=1}^{N} \mu(A_i),}$$

где $\boldsymbol{N}$— натуральное число или символ $\boldsymbol{+\infty}$. Это свойство называется аддитивностью (для конечных $\boldsymbol{N}$) или $\boldsymbol{\sigma}$-*аддитивностью* (для бесконечного $\boldsymbol{N}$). То есть, мера это неотрицательная, счётно-аддитивная функция множеств.

Определение 12. Пусть $\boldsymbol{\Omega}$ есть пространство элементарных событий и $\boldsymbol{\mathcal{A}}$ — $\boldsymbol{\sigma}$-алгебра событий. *Вероятностью* или *вероятностной мерой* на измеримом пространстве $\boldsymbol{(\Omega, \mathcal{A})}$ называется мера $\boldsymbol{P : \mathcal{A} \to [0, 1]}$, обладающая свойствами:

(i) Для любого события $\boldsymbol{A \in \mathcal{A}}$ выполняется неравенство $\boldsymbol{P(A) \geq 0}$;

(ii) Для любого не более, чем счётного набора попарно несовместных событий $\boldsymbol{A_1, A_2, \ldots \in \mathcal{A}}$ имеет место равенство

$$\boldsymbol{P\left(\bigcup_{i=1}^{N} A_i\right) = \sum_{i=1}^{N} P(A_i),}$$

где $\boldsymbol{N}$— натуральное число или символ $\boldsymbol{+\infty}$.

(iii) Вероятность достоверного события равна единице: $\boldsymbol{P(\Omega) = 1}$.

(i), *(ii)* и *(iii)* называют *аксиомами вероятности.*

Определение 13. Тройка объектов $\boldsymbol{(\Omega, \mathcal{A}, P)}$, в которой $\boldsymbol{\Omega}$ – пространство элементарных событий, $\boldsymbol{\mathcal{A}}$ – $\boldsymbol{\sigma}$-алгебра его подмножеств и P – вероятностная мера, определённая на $\boldsymbol{\mathcal{A}}$, называется *вероятностным пространством.*

Рассмотрим свойства вероятности.

j). $\boldsymbol{P(\varnothing) = 0}$

Покажем, что это так. События $\boldsymbol{A_i = \varnothing, \; i = 1, 2, \ldots}$ попарно несовместны, т.е. $\boldsymbol{A_i \cap A_j = \varnothing \; \forall i \neq j}$ и $\boldsymbol{\bigcup\limits_{i=1}^{\infty} A_i = \varnothing}$, поэтому из $\boldsymbol{(ii)}$ следует, что $\boldsymbol{P(\varnothing) = \sum\limits_{i=1}^{\infty} P(\varnothing)}$, что возможно только в случае, если $\boldsymbol{P(\varnothing) = 0}$.

jj). Для любого конечного набора попарно несовместных событий $\boldsymbol{A_1, \ldots, A_n \in \mathcal{A}}$ имеет место равенство $\boldsymbol{P\left(\bigcup\limits_{i=1}^{n} A_i\right) = \sum\limits_{i=1}^{n} P(A_i)}$.

Покажем это. Пусть $\boldsymbol{B_i = A_i, \; i = 1, \ldots, n}$ и $\boldsymbol{B_i = \varnothing, \; i = n+1, n+2, \ldots}$. Ясно, что $\boldsymbol{\bigcup\limits_{i=1}^{n} A_i = \bigcup\limits_{i=1}^{\infty} B_i}$ и бесконечная совокупность событий $\boldsymbol{B_i}$ тоже будет набором попарно несовместных событий. Поэтому, из (ii) и предыдущего свойства $j)$ получаем

$$P(\bigcup_{i=1}^{n} A_i) = P(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} P(B_i) = \sum_{i=1}^{n} P(A_i).$$

jjj). $P(\bar{A}) = 1 - P(A)$.

Действительно, поскольку $A \cup \bar{A} = \Omega$, а события A и $\bar{A}$ несовместны, то по аксиоме (iii) и предыдущему свойству *jj)* получаем, $1 = P(\Omega) = P(A) + P(\bar{A})$.

jv). Если $A \subseteq B$, то $P(B \backslash A) = P(B) - P(A)$.

Действительно, поскольку $B = A \cup (B \backslash A)$ и события A и $B \backslash A$ несовместны, то по аксиоме *(ii)* и свойству *jj)* получаем, $P(B) = P(A) + P(B \backslash A)$.

v). Если $A \subseteq B$, то $P(A) \leq P(B)$.

Действительно, по предыдущему свойству *jv)* имеем $P(A) = P(B) - P(B \backslash A)$, но из *(i)* имеем $P(B \backslash A) \geq 0$, откуда из последнего равенства вытекает рассматриваемое свойство.

vj). $0 \leq P(A) \leq 1$.

Действительно, $P(A) \geq 0$ по *(i)*, и поскольку $A \subseteq \Omega$, то по свойству *v)* $P(A) \leq P(\Omega) = 1$.

vjj). $P(A \cup B) = P(A) + P(B) - P(A \bigcap B)$.

Действительно, $A \cap B \subseteq B$, поэтому $P(B \backslash (A \cap B)) = P(B) - P(A \cap B)$. Но события A и $B \backslash (A \cap B)$ несовместны, поэтому

$$P(A \cup B) = P(A \cup B \backslash (A \cap B)) = P(A) + P(B \backslash (A \cap B)) =$$

$$= P(A) + P(B) - P(A \cap B).$$

vjjj). $P(A \triangle B) = P(A) + P(B) - 2P(A \bigcap B)$.

Ясно, что $A \triangle B = (A \backslash B) \cup (B \backslash A) = (A \backslash (A \cap B) \cup (B \backslash (A \cap B))$, $A \cap B \subseteq A$ и $A \cap B \subseteq B$, поэтому $P(A \backslash (A \cap B)) = P(A) - P(A \cap B)$ и $P(B \backslash (A \cap B)) = P(B) - P(A \cap B)$. Но события $A \backslash (A \cap B)$ и $B \backslash (A \cap B)$ несовместны, поэтому

$$P(A \triangle B) = P(A \backslash (A \cap B)) + P(B \backslash (A \cap B)) =$$

$$= P(A) + P(B) - 2P(A \cap B).$$

jx). $P(A \triangle B) \leq P(A \cup B)$.

Свойство сразу следует из свойств *vjj)* и *vjjj).*

x). $P(A \cup B) \leq P(A) + P(B)$.

Свойство сразу следует из свойства *vjj)* и аксиомы (i).

xj). $P(A \triangle B) \leq P(A) + P(B)$.

Свойство сразу следует из свойства *vjij)* и аксиомы (i).

xjj). $P(A_1 \cup \dots \cup A_n) \le \sum\limits_{i=1}^{n} P(A_i)$.

Легко показать методом математической индукции.

xjjj). $P\left(\bigcup\limits_{i=1}^{n} A_i\right) = \sum\limits_{i=1}^{n} P(A_i) - \sum\limits_{1\le i_1<i_2\le n} P(A_{i_1}\cap A_{i_2}) +$
$+ \sum\limits_{1\le i_1<i_2<i_3\le n} P(A_{i_1}\cap A_{i_2}\cap A_{i_3}) + \dots + (-1)^{n-1} P(A_1\cap A_2\cap\dots\cap A_n)$.

Доказательство этого свойства проводится методом математической индукции. При $n = 2$ — свойство *vjj).* Пусть доказываемое свойство *xjjj)* верно при $n = k-1$. Докажем, что тогда оно верно и при n = k.

$$P\left(\bigcup_{i=1}^{k} A_i\right) = P\left(\bigcup_{i=1}^{k-1} A_i \cup A_k\right) =$$

$$= P\left(\bigcup_{i=1}^{k-1} A_i\right) + P(A_k) - P\left(A_k \cap \bigcup_{i=1}^{k-1} A_i\right).$$

По предположению индукции, первое слагаемое в правой части предыдущего равенства равно

$P\left(\bigcup\limits_{i=1}^{k-1} A_i\right) = \sum\limits_{i=1}^{k-1} P(A_i) - \sum\limits_{1\le i_1<i_2\le k-1} P(A_{i_1}\cap A_{i_2}) +$
$+ \sum\limits_{1\le i_1<i_2<i_3\le k-1} P(A_{i_1}\cap A_{i_2}\cap A_{i_3}) + \dots + (-1)^{k-2} P(A_1\cap A_2\cap\dots\cap A_{k-1})$.

Вычитаемое в правой части первого равенства равно

$P\left(A_k \cap \bigcup\limits_{i=1}^{k-1} A_i\right) = P\left(\bigcup\limits_{i=1}^{k-1} (A_i\cap A_k)\right) = \sum\limits_{i=1}^{k-1} (P(A_i\cap A_k) - \sum\limits_{1\le i_1<i_2\le k-1} P(A_{i_1}\cap$
$A_{i_2}\cap A_k) +$
$+ \sum\limits_{1\le i_1<i_2<i_3\le k-1} P(A_{i_1}\cap A_{i_2}\cap A_{i_3}\cap A_k) + \dots + (-1)^{k-2} P(A_1\cap A_2\cap\dots\cap A_{k-1}\cap A_k)$.

Подставляя последние два равенства в первое, получаем требуемое утверждение.

xjv). $P(A_1 \triangle A_2 \triangle \dots \triangle A_n) = \sum\limits_{i=1}^{n} P(A_i) - 2\sum\limits_{1\le i_1<i_2\le n} P(A_{i_1}\cap A_{i_2}) +$
$+ 4\sum\limits_{1\le i_1<i_2<i_3\le n} P(A_{i_1}\cap A_{i_2}\cap A_{i_3}) + \dots + (-2)^{n-1} P(A_1\cap A_2\cap\dots\cap A_n)$.

Доказательство этого свойства проводится методом математической индукции аналогично свойству *xjjj).*

Пример 10. Пусть имеются n занумерованных ящиков и n занумерованных шаров. Шары наудачу размещают по одному в каждый незанятый ящик. Найти вероятность того, что хотя бы один шар попадёт в ящик с тем же номером, что и шар. Найти предел этой вероятности при $n \to \infty$.

Решение. Пусть событие $A_i, i = 1, \dots, n$ означает, что i-й шар попал в ящик с тем же номером. Тогда событие $A = \{$хотя бы один шар попадёт в ящик с тем же номером, что и шар$\} = A_1 \cup \dots \cup A_n$. События $A_1, \dots, A_n$ совместны, поэтому используем формулу свойства *xjjj*. Легко проверяется, что

$P(A_i) = \frac{1}{n} \quad i = 1, \dots, n;$

$P(A_i \cap A_j) = \frac{(n-2)!}{n!} \quad 1 \le i < j \le n;$

$P(A_i \cap A_j \cap A_k) = \frac{(n-3)!}{n!} \quad 1 \le i < j < k \le n;$

$\dots$

$P\left(\bigcap_{i=1}^{n} A_i\right) = \frac{1}{n!}.$

Количество слагаемых в суммах правой части формулы свойства *xjjj*: первая сумма - n слагаемых, вторая сумма - C_n^2 слагаемых, третья сумма - C_n^3 слагаемых и т.д.

Подставляя всё найденное в эту формулу, получаем:

$P(A) = n\cdot\frac{1}{n} - C_n^2\cdot\frac{(n-2)!}{n!} + C_n^3\cdot\frac{(n-3)!}{n!} + \dots + (-1)^{n-1}\cdot\frac{1}{n!} = 1 - \frac{1}{2!} + \frac{1}{3!} + \dots + (-1)^{n-1}\frac{1}{n!}$.
Воспользовавшись разложением e^{-1} по формуле Тейлора, получаем, $P(A) \to 1 - e^{-1} = 0,632\dots$ *при* $n \to \infty$.

§3. Некоторые элементы функционального анализа

Этот параграф при первом чтении можно опустить и обращаться к материалу параграфа только в случае, если есть желание понять излагаемый материал глубже.

1. Борелевская σ-алгебра на прямой

Начнём этот раздел параграфа с простого примера.

Пример 11. Пусть $\Omega = R^1$ — вещественная прямая. Рассмотрим некоторые совокупности подмножеств на прямой, которые,вообще говоря, не являются σ-алгеброй, и посмотрим, как эту совокупность можно пополнить до σ-алгебры.

i. Совокупность подмножеств $\mathcal{F} = \{\Omega, \varnothing, [0,1], \{0\}\} = \{R^1, \varnothing, [0,1], \{0\}\}$ не является σ-алгеброй, поскольку, например, $\overline{[0,1]} = R^1 \backslash [0,1] = (-\infty, 0) \cup (1, +\infty) \notin \mathcal{F}$. Набор множеств, целиком содержащий совокупность $\mathcal{F}$ и являющийся σ-алгеброй, получится, если включить в него всевозможные объединения, пересечения и дополнения множеств из $\mathcal{F}$:

$\mathcal{A} = \{R^1, \varnothing, [0,1], \{0\}, (-\infty, 0) \cup (1, +\infty), (0,1], (-\infty, 0] \cup (1, +\infty), (-\infty, 0) \cup (0, +\infty)\}$.

Кстати, попутно заметим, что если какая-то σ-алгебра состоит из конечного числа множеств, то число этих множеств *обязано* быть степенью двойки, что несложно показать.

Определение 14. Минимальной σ-алгеброй, содержащей совокупность множеств $\mathcal{F}$, называется пересечение всех σ-алгебр, содержащих $\mathcal{F}$. Такую минимальную σ-алгебру будем обозначать $\sigma(\mathcal{F})$.

Ясно, что минимальная σ-алгебра, содержащая некоторую фиксированную совокупность, единственна. Поэтому, под σ-алгеброй, порождённой некоторой совокупностью всегда будем понимать именно минимальную σ-алгебру, содержащую эту совокупность.

Таким образом, понятие минимальной $\boldsymbol{\sigma}$-алгебры и $\boldsymbol{\sigma}$-алгебры, порождённой совокупностью множеств, есть синонимы.

Ясно, что $\boldsymbol{\sigma}$-алгебра $\boldsymbol{\mathcal{A}}$, построенная в примере 11 есть $\boldsymbol{\sigma(\mathcal{F})}$.

Пусть $\boldsymbol{\mathcal{F}}$ – совокупность подмножеств вещественной прямой $\boldsymbol{R^1}$, состоящая из всевозможных открытых интервалов: $\boldsymbol{\mathcal{F} = \{(a,b)\ :\ -\infty < a < b < +\infty\}}$. Легко проверяется, что совокупность $\boldsymbol{\mathcal{F}}$ не является $\boldsymbol{\sigma}$-алгеброй.

Определение 15. Минимальная $\boldsymbol{\sigma}$-алгебра $\boldsymbol{\sigma(\mathcal{F})}$, содержащая совокупность $\boldsymbol{\mathcal{F}}$ всех интервалов на вещественной прямой, называется борелевской $\boldsymbol{\sigma}$-алгеброй в $\boldsymbol{R^1}$ и обозначается $\boldsymbol{\mathfrak{B}}$.

Определение 16. Любое множество $\boldsymbol{B \in \mathfrak{B}}$ называется *борелевским* множеством.

Ясно, что вся прямая $\boldsymbol{R^1}$ есть борелевское множество. Или, например, замкнутый отрезок $\boldsymbol{[a,b],\ \ -\infty < a < b < +\infty}$ является борелевским множеством. Действительно, $\boldsymbol{[a,b] = \bigcap\limits_{n=1}^{\infty}(a-\frac{1}{n}, b+\frac{1}{n})}$. Аналогично показывается, что множества $\boldsymbol{[a,b)}$ и $\boldsymbol{(a,b]}$ являются борелевскими. Ясно также, что любое одноточечное множество $\boldsymbol{\{c\} = (b,c]\backslash(b,c)}$ также борелевское.

Определение 17. Минимальная $\boldsymbol{\sigma}$-алгебра $\boldsymbol{\sigma(\mathcal{F})}$, где совокупность $\boldsymbol{\mathcal{F}}$ всех многомерных открытых параллелепипедов $\boldsymbol{(a_1,b_1)\bigotimes(a_2,b_2)\bigotimes\cdots\bigotimes(a_k,b_k) \subset R^k}$, называется *борелевской* $\boldsymbol{\sigma}$*-алгеброй на* $\boldsymbol{R^k}$ и обозначается $\boldsymbol{\mathfrak{B}(R^k)}$.

2. Мера Лебега

Когда выше вводилось понятие *геометрической вероятности*, было использовано понятие меры измеримого множества $\boldsymbol{A \subset R^k}$, имея ввиду *длину* на прямой, *площадь* на плоскости и *объём* в трёхмерном пространстве. Под понятием *измеримого* множества далее всегда будем понимать *борелевское* множество. Возникает вопрос, будут ли понятия длина, площадь и объём настоящими мерами? Рассмотрим здесь для простоты только одномерный случай.

Одномерным измеримым вещественным пространством будем называть пару $\boldsymbol{(R^1, \mathfrak{B})}$, состоящую из вещественной прямой и борелевской $\boldsymbol{\sigma}$-алгеброй. Борелевская $\boldsymbol{\sigma}$-алгебра, по определению, является наименьшей $\boldsymbol{\sigma}$-алгеброй, содержащей любые интервалы. Для любого интервала $\boldsymbol{(a,b) \subset R}$ будем называть число $\boldsymbol{b-a}$ длиной интервала $\boldsymbol{(a,b)}$.

Справедлива следующая теорема.

Теорема 2. Существует единственная мера, т.е. неотрицательная и $\boldsymbol{\sigma}$-аддитивная функция $\boldsymbol{\lambda}$ на измеримом вещественном пространстве $\boldsymbol{(R^1, \mathfrak{B})}$, значение которой на любом интервале равно его длине: $\boldsymbol{\lambda((a,b)) = b-a}$. Эта мера называется *мерой Лебега.*

Эту теорему оставляем без доказательства. Утверждение теоремы 2 является следствием теоремы Каратеодори, см. А.Н.Колмогоров, С.В.Фомин, Функциональный анализ. В этой же книге можно найти обобщение этого утверждения на многомерный случай.

Таким образом, множество представляет собой совокупность борелевских множеств, и вероятность определяется только для на них.

В заключение этого параграфа приведём один пример неизмеримого множества, для которого понятие меры не существует.

Пример 12 (множество Витали). Рассмотрим следующее отношение эквивалентности $\sim$ на отрезке $[\mathbf{0},\ \mathbf{1}]$: $\boldsymbol{x} \sim \boldsymbol{y}$, если разность $\boldsymbol{x} - \boldsymbol{y}$ рациональна. Как обычно, это отношение эквивалентности разбивает отрезок $[\mathbf{0},\ \mathbf{1}]$ на классы эквивалентности, каждый из которых имеет счётную мощность, но их количество имеет мощность континуума. Далее, из каждого класса эквивалентности выберем по представителю — одной точке (здесь используется *аксиомой выбора* теории множеств: *для данного произвольного множества попарно непересекающихся непустых множеств существует по крайней мере одно множество, которое содержит точно один элемент, общий с каждым из непустых множеств*). Тогда полученное множество $\boldsymbol{E}$ представителей будет *неизмеримым.*

Действительно, если сдвинуть $\boldsymbol{E}$ счётное число раз на все рациональные числа из отрезка $[\mathbf{-1}, \mathbf{1}]$, то объединение будет содержать весь отрезок $[\mathbf{0}, \mathbf{1}]$, но при этом оно будет содержаться в отрезке $[\mathbf{-1}, \mathbf{2}]$. При этом «сдвинутые копии» множества $\boldsymbol{E}$ не будут пересекаться друг с другом, что непосредственно следует из построения отношения эквивалентности и множества $\boldsymbol{E}$.

Предположим, что множество $\boldsymbol{E}$ измеримо по Лебегу (иммет конечную меру Лебега). Тогда возможны два варианта.

i. Мера множества $\boldsymbol{E}$ равна нулю. Тогда мера отрезка $[\mathbf{0}, \mathbf{1}]$, как счётного объединения множеств меры нуль, тоже будет равна нулю.

ii. Мера $\boldsymbol{E}$ больше нуля. Тогда аналогично заключаем, что мера отрезка $[\mathbf{-1}, \mathbf{2}]$, в силу счётной аддитивности меры Лебега, будет бесконечна.

В обоих случаях получается противоречие. Таким образом, множество Витали не измеримо по Лебегу.

Глава 3. Условная вероятность. Независимость

§1. Условная вероятность

Начнём этот параграф с простого примера.

Пример 13. Игральная кость подбрасывается один раз. Известно, что выпало более трёх очков. Какова при этом вероятности того, что выпало четное число очков?

В данном примере пространство элементарных событий состоит из трёх равновозможных элементарных событий: $\boldsymbol{\Omega} = \{\mathbf{4}, \mathbf{5}, \mathbf{6}\}$, а событие $\boldsymbol{A} = \{$*выпало четное число* *очков*$\}$ состоит из двух элементарных событий: $\boldsymbol{A = \{4, 6\}}$. Поэтому по классической схеме, получаем, $\boldsymbol{P(A) = 2/3}$.

Посмотрим на этот пример с другой точки зрения. Пространство элементарных событий при однократном подбрасывании игральной кости состоит из шести элементарных события: $\boldsymbol{\Omega = \{1, 2, 3, 4, 5, 6\}}$. Фраза *«известно, что выпало более трёх очков»* означают, что в эксперименте произошло событие $\boldsymbol{B = \{4, 5, 6\}}$. Фраза *«какова при этом вероятность того, что выпало четное число очков?»* означают, что нас интересует, в какой доле случаев при осуществлении события $\boldsymbol{B}$ происходит и событие $\boldsymbol{A}$. Вероятность события $\boldsymbol{A}$, вычисленную в предположении, что событие $\boldsymbol{B}$ произошло, обозначается как $\boldsymbol{P(A|B)}$. Нужно вычислить отношение вероятности одновременного появление событий $\boldsymbol{A}$ и $\boldsymbol{B}$ к вероятности появление события $\boldsymbol{B}$.

$$\boldsymbol{P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{2/6}{3/6} = \frac{2}{3}.}$$

Определение 18. Условной вероятностью события A, при условии, что произошло событие B, называется величина

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

Будем считать, что условная вероятность определена только в случае, когда $\boldsymbol{P(B) > 0}$.

Теорема 3 (Теорема умножения). Если $\boldsymbol{P(B) > 0}$, то $\boldsymbol{P(A \cap B) = P(B)P(A|B)}$ и если $\boldsymbol{P(A) > 0}$, то $\boldsymbol{P(A \cap B) = P(A)P(B|A)}$.

Действительно, первая часть утверждения теоремы 3 непосредственно следует из определения 18. Вторая часть получается, если в определении 18 поменять местами события $\boldsymbol{A}$ и $\boldsymbol{B}$.

Теорема 4 (Обобщённая теорема умножения).

$$\boldsymbol{P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)P(A_2|A_1)P(A_3|A_1 \bigcap A_2)\ldots P\left(A_n|\bigcap_{i=1}^{n-1} A_i\right),}$$

если соответствующие условные вероятности определены.

Проверим эту формулу хотя бы для случая $\boldsymbol{n = 3}$.

$$\boldsymbol{P\left(A_1 \bigcap A_2 \bigcap A_3\right) = P(A_1) \cdot \frac{(P(A_1 \bigcap A_2)}{P(A_1)} \cdot \frac{(P(A_1 \bigcap A_2 \bigcap A_3)}{P(A_1 \bigcap A_2)} =}$$

$$\boldsymbol{= P(A_1)P(A_2|A_1)P(A_3|A_1 \bigcap A_2).}$$

§2. Независимость

Определение 19. События A и B называются независимыми, если $P(A \bigcap B) = P(A)P(B).$

Из этого определения, в частности, следует, что если $P(B) > 0,$ то $P(A|B) = P(A)$ или если $P(A) > 0,$ то $P(B|A) = P(A).$

Лемма. Если события A и B независимы, то независимы и события A и $\bar{B}$, $\bar{A}$ и B, $\bar{A}$ и $\bar{B}$.

Доказательство. Так как $A = A \cap B \bigcup A \cap \bar{B}$, и события $A \cap B$ и $A \cap \bar{B}$ несовместны, то $P(A) = P(A \cap B) + P(A \cap \bar{B})$. Поэтому, $P(A \cap \bar{B}) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(\bar{B})$.

Остальные утверждения доказываются аналогично.

Определение 20. События $A_1, A_2, \dots, A_n$ называются *независимыми в совокупности*, если для любого набора индексов $1 \le i_1 < \dots < i_k \le n$

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j}).$$

Очевидно, что если события $A_1, \dots, A_n$ независимы в совокупности, то они и попарно независимы, что, следует из последнего определения. Как показывает следующий классический пример, обратное неверно.

Пример 14.(С. Н. Бернштейн). Пусть имеется правильный тетраэдр, три грани которого окрашены, соответственно, в красный, синий и зелёный цвета, а на четвертой грани присутствуют все три цвета одновременно. При однократном подбрасывании этого тетраэдра события A, B и C означают, что выпала грань, содержащая красный, синий или зелёный цвета соответственно. Вероятность каждого из этих событий равна 1/2, так как каждый из цветов присутствуют на двух гранях из четырёх. Вероятность появления любых двух выбранных цветов одновременно равна 1/4, так как только одна грань содержит эти два выбранные цвета. А поскольку $1/4 = 1/2 \cdot 1/2$, то все эти события являются попарно независимыми. Но вероятность появления всех трёх событий тоже равна $1/4$, а не $1/8$. Таким образом, события A, B и C не являются независимыми в совокупности.

§3. Формула полной вероятности

Определение 21. Совокупность событий $H_1, H_2, \dots$ таких, что они попарно несовместны, т.е. $H_i \cap H_j = \varnothing \ \forall\, i \ne j,$ $P(H_i) \ge 0 \ \forall\, i = 1, 2 \dots$ и $\bigcup\limits_{i=1}^{N} H_i = \Omega$, называется *полной группой событий* или разбиением пространства элементарных событий $\Omega,$ где $N-$ либо натуральное число, либо символ $\infty.$

События $H_1, H_2, \ldots$, образующие полную группу событий, часто называют *гипотезами.* При подходящем выборе гипотез могут быть сравнительно просто вычислены их вероятности $P(H_i)$, а также для произвольного события A и условные вероятности $P(A|H_i)$ - вероятность появления события A при выполнении гипотезы H_i.

Теорема 5 (Формула полной вероятности). Пусть $H_1, H_2, \ldots$ — произвольная полная группа событий. Тогда вероятность любого события A может быть найдена по формуле:

$$P(A) = \sum_{i=1}^{N} P(A \big| H_i) P(H_i).$$

Доказательство. Поскольку $\{H_i\}$ - полная группа, то воспользовавшись свойством ассоциативности операций «или» и «и», получаем, что

$$A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^{N} H_i \right) = \bigcup_{i=1}^{N} (A \cap H_i).$$

События $A \cap H_1, A \cap H_2, \ldots$ попарно несовместны, поскольку $A \cap H_i \subset H_i$. Поэтому, используя σ-аддитивность вероятностной меры и теорему умножения, получаем,

$$P(A) = \sum_{i=1}^{N} P(A \cap H_i) = \sum_{i=1}^{N} P(A \big| H_i) P(H_i).$$

Примечание. Конечно, $P(A \cap H_i) = P(A|H_i)P(H_i)$ справедлива лишь при $P(H_i) > 0,$ но если $P(H_i) = 0,$ то это означает, что событие $P(H_i)$ либо невозможно в условиях данного эксперимента, либо "мера"его равна нулю. Но тогда и событие $A \cap H_i$ либо невозможно, либо его мера, тем более, равна 0. Поэтому, если принять, $P(A|H_i)$ равным любому числу (например, 0,5), то равенство $P(A \cap H_i) = (A|H_i)P(H_i)$ сохранится $(0 = 0)$ и, тем самым, сохранит свою справедливость и формула полной вероятности.

Пример 15. В группе из 30 студентов 5 отличников, 10 хорошистов и 15 троечников. Отличник на экзамене получает оценку «отлично» с вероятностью 4/5 и оценку «хорошо» с вероятностью 1/5; хорошист – «отлично» и «хорошо» с одинаковыми вероятностями и троечник на экзамене может получить оценки «хорошо», «удовлетворительно» и «неудовлетворительно» с одинаковыми вероятностями. Найти вероятность того, что первый произвольно вызванный студент получит оценку «хорошо»?

Решение. Пусть событие $A = \{$*вызванный студент получил оценку «хорошо»*$\}$ и пусть гипотеза $H_1 = \{$*вызван отличник*$\}$, $H_2 = \{$*вызван хорошист*$\}$ и $H_3 = \{$*вызван троечник*$\}$. Из условия, очевидно, следует, что $P(H_1) = 1/6$, $P(H_2) = 1/3$ и $P(H_3) = 1/2$. Кроме этого, из условия имеем: $P(A \big| H_1) = 1/5$, $P(A \big| H_2) = 1/2$

и $P(A|H_3) = 1/3$. Из формулы полной вероятности, получаем

$$P(A) = \frac{1}{5} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2} = \frac{11}{30}.$$

§4. Формула Байеса

Теорема 6 (Формула Байеса). Пусть $H_1, H_2, \ldots$ — полная группа гипотез и A — некоторое событие положительной вероятности. Тогда условная вероятность того, что событие A произошло при гипотезе H_k, может быть вычислена по формуле:

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

Доказательство. По определению условной вероятности,

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

Последнее равенство следует из теоремы умножения и формулы полной вероятности.

Отметим, что теорему 6 иногда называют *теоремой гипотез*.

Пример 16. Пусть условие такое же, как и в примере 15. Дополнительно известно, что вызванный на экзамене студент получил оценку «хорошо». Требуется найти вероятность того, что был вызван троечник.

Решение. В этом примере требуется найти условную вероятность $P\left(H_3|A\right)$, где все события определены в примере 15. По формуле Байеса, получаем:

$$P\left(H_3|A\right) = \frac{P(A|H_3)P(H_3)}{P(A)} = \frac{\frac{1}{3} \cdot \frac{1}{2}}{\frac{11}{30}} = \frac{5}{11}.$$

Глава 4. Схема Бернулли

§1. Распределение числа успехов в n испытаниях

Определение 22. Схемой Бернулли называется последовательность независимых испытаний, в каждом из которых возможны лишь два события – «успех» и «неудача», при этом в одном испытании «успех» происходит с вероятностью $p \in [0,1]$, а «неудача» – с вероятностью $q = 1 - p$.

Такая схема возникает, например, в экспериментах с подбрасыванием монеты, при этом отмечаются только события «орёл» и «решка», а любые другие события, которые могут произойти при бросаниях, будем считать невозможными. Пусть при каждом таком подбрасывании событие, скажем, «орёл» будет «успех», а противоположное событие «решка» – «неудача».

Теорема 7 (Формула Бернулли).

Пусть ν_n обозначает число успехов в n испытаниях схемы Бернулли. Тогда для любого $k = 0, 1, \ldots, n$

$$P(\nu_n = k) = C_n^k p^k (1-p)^{n-k} = C_n^k p^k q^{n-k}.$$

Доказательство. Событие $A = \{\nu_n = k\}$ означает, что в n испытаниях схемы Бернулли произошло ровно k успехов. Для простоты обозначений, пусть 1 обозначает «успех», а 0 – «неудача». Тогда результат n испытаний представляется бинарным вектором. Одно из возможных благоприятствующих событию A элементарных событий есть бинарный вектор $(1, 1, \ldots 1, 0, 0, \ldots, 0)$, у которого первых k местах стоять единицы, а следом идут $n - k$ нулей. Поскольку испытания независимы, вероятность такого элементарного события (первые k испытаний завершились успехом, остальные – неудачей) равна $p^k(1-p)^{n-k}$. Все другие элементарные события A отличаются от рассмотренного лишь расположением k единиц на n местах. Ясно, что имеется ровно C_n^k способов расположить k единиц на n местах. Поэтому, событие A содержит ровно C_n^k элементарных событий, вероятность каждого из которых равна $p^k(1-p)^{n-k}$.

Определение 23. Набор вероятностей $\{C_n^k p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n\}$ называется *биномиальным распределением вероятностей* и часто обозначается $B_{n,p}$ или $B(n,p)$.

§2. Наивероятнейшее число успехов

По формуле Бернулли, событие $\{$*произошло* 0 *успехов в* n *испытаниях*$\}$ имеет вероятность q^n, *1 успех* — вероятность npq^{n-1} и т.д. Какое же число успехов наиболее вероятно? Иначе говоря, при каком k достигается максимум вероятности $P(\nu_n = k)$?

Чтобы ответить на этот вопрос, сравним отношение $P(\nu_n = k)/P(\nu_n = k-1)$ с единицей. $\frac{P(\nu_n=k)}{P(\nu_n=k-1)} = \frac{n!}{k!(n-k!)}\cdot\frac{(k-1)!(n-k+1)!}{n!}\cdot\frac{p^k q^{n-k}}{p^{k-1}q^{n-k+1}} = \frac{(n-k+1)p}{kq} = 1+\frac{(n-k+1)p}{kq}-1 = 1 + \frac{np+p-k}{kq}$. Ясно, что

(i) $P(\nu_n = k) > P(\nu_n = k-1)$ при $np + p - k > 0$, т.е. при $k < np + p$;

(ii) $P(\nu_n = k) < P(\nu_n = k-1)$ при $np + p - k < 0$, т.е. при $k > np + p$;

(iii) $P(\nu_n = k) = P(\nu_n = k-1)$ при $np + p - k = 0$, что возможно только в случае, если $np + p$ есть целое число.

Рассмотрим два случая: $np + p$ – целое и $np + p$ – нецелое. В первом случае пусть $k_0 = np + p$. Из полученных выше неравенств сразу следует, что

$$\ldots < P(\nu_n = k_0 - 2) \overset{(i)}{<} P(\nu_n = k_0 - 1) \overset{(iii)}{=} P(\nu_n = k_0) \overset{(ii)}{>}$$

$$P(\nu_n = k_0 + 1) > \ldots$$

Во втором случае пусть $k_0 = [np + p]$ – целая часть числа $np + p$, то есть наибольшее целое число, не превосходящее $np + p$. Из двух неравенств (i) и (ii) следует, что

$$\ldots < P(\nu_n = k_0 - 2) \overset{(i)}{<} P(\nu_n = k_0 - 1) \overset{(i)}{<} P(\nu_n = k_0) \overset{(ii)}{>}$$

$$P(\nu_n = k_0 + 1) > \ldots$$

Действительно, неравенство $P(\nu_n = k_0) > P(\nu_n = k_0 + 1)$, например, следует из (ii), применённого для $k = k_0 + 1 > np + p$. Поэтому, в зависимости от того, будет ли число $np + p$ целым или нет, имеется либо два равновероятных «наиболее вероятных» числа успехов $k_0 = np + p$ и $k_0 - 1 = np + p - 1$, либо одно «наиболее вероятное» число успехов $k_0 = [np + p]$. Сформулируем это утверждение в виде теоремы.

Теорема 8. В n испытаниях схемы Бернулли с вероятностью успеха p наивероятнейшим числом успехов является

$i)$ единственное число $k_0 = [np + p]$, если число $np + p$ не целое;

$ii)$ два числа $k_0 = np + p$ и $k_0 - 1 = np + p - 1$, если число $np + p$ целое.

Пример 17. В схеме Бернулли, если $p = q = 1/2$, то при чётном числе испытаний n число $np + p = n/2 + 1/2$ не целое, так что наивероятнейшее число успехов есть единственное число успехов $[n/2+1/2] = n/2$. При нечётном числе испытаний n число $np + p = n/2 + 1/2$ целое, так что наивероятнейшими числами успехов являются два числа $n/2 + 1/2$ и $n/2 - 1/2$.

§3. Геометрическое распределение вероятностей

Пусть производится последовательность независимых испытаний Бернулли с вероятностью успеха p в каждом испытании. Испытания проводятся до появления первого успеха. Введем величину $\boldsymbol{\tau}$, принимающую значения из множества целых чисел $\{1, 2, 3, \dots\}$, равную номеру первого успешного испытания, т.е. номер испытания, при котором первый раз произошёл «успех». Справедливо следующее утверждение:

Теорема 9. Вероятность того, что первый успех произойдет в испытании с номером k, равна $P(\tau = k) = p \cdot q^{k-1}$.

Доказательство. Действительно, первый успех произойдет при k-м испытании, если при первых $k - 1$ испытаниях появлялись неудачи: $P(\tau = k) = P(0, 0, \dots, 0, 1) = p \cdot q^{k-1}$, где ноль, как и выше означает «неудачу», а единица - «успех».

Определение 24. Набор чисел $\{pq^{k-1}, k = 1, 2, 3, \dots\}$ называется *геометрическим распределением вероятностей* и обозначается G_p или $G(p)$.

Геометрическое распределение вероятностей обладает интересным свойством, которое можно назвать свойством «нестарения». Пусть величина τ обозначает, например, время, измеряемое целым числом часов безотказной работы некоторого устройства. Предположим, что для величины τ вероятность принять любое значение k в точности равна pq^{k-1}. Справедливо следующее утверждение.

Теорема 10. Пусть $P(\tau = k) = pq^{k-1} \forall k = 1, 2, \dots$. Тогда для произвольных n и $k \geq 0$

$$P(\tau > n + k | \tau > n) = P(\tau > k).$$

Это означает, что если известно, что устройство уже проработало без отказа n часов, то вероятность ему работать еще не менее k часов точно такая же, как вероятность проработать не менее k часов для нового устройства.

Доказательство. По определению условной вероятности имеем,

$$P\left(\tau > n + k \middle| \tau > n\right) = \frac{P(\{\tau > n + k\} \cap \{\tau > n\})}{P(\tau > n)} =$$

$$= \frac{P(\tau > n + k)}{P(\tau > n)}.$$

Последнее равенство следует из того, что $\{\tau > n+k\} \subset \{\tau > n\}$, так что одновременное выполнение этих событий есть событие $\{\tau > n + k\}$. Для произвольного целого $m > 0$, воспользовавшись формулой бесконечной суммы членов геометрической прогрессии, найдём,

$$P(\tau > m) = \sum_{i=m+1}^{\infty} P(\tau = i) = \sum_{i=m+1}^{\infty} pq^{i-1} = \frac{pq^m}{1 - q} = q^m.$$

Поэтому, окончательно получаем

$$P\left(\tau > n + k \middle| \tau > n\right) = \frac{P(\tau > n + k)}{P(\tau > n)} = \frac{q^{n+k}}{q^n} = q^k = P(\tau > k).$$

§4. Приближение гипергеометрического распределения

Пусть урна содержит N шаров, из которых K шаров – белые, а остальные $N-K$ шаров – чёрные. Из урны наудачу без возвращения извлекают n шаров. Вероятность $P_{N,K}(n,k)$ того, что будет выбрано ровно k белых и $n-k$ чёрных шаров, находится по формуле гипергеометрического распределения вероятностей:

$$P_{N,K}(n,k) = \frac{C_K^k \cdot C_{N-K}^{n-k}}{C_N^n}.$$

Если число шаров в урне велико, то извлечение небольшого количества шаров почти не меняет пропорцию белых и чёрных шаров в урне, так что вероятности $P_{N,K}(n,k)$ практически не отличаются от вероятностей в схеме выбора *с возвращением.* Положим $A_{n,k} = \{$*извлечь ровно* k *белых шаров при выборе* n *шаров с возвращением*$\}$. Тогда

$$P(A_{n,k}) = C_n^k \left(\frac{K}{N}\right)^k \left(1 - \frac{K}{N}\right)^{n-k}.$$

Теорема 11. Если $N \to \infty$ и $K \to \infty$ так, что $K/N \to p \in (0,1)$, то $\forall\ n,\ 0 \le k \le n$

$$P_{N,K}(n,k) \to C_n^k\, p^k (1-p)^{n-k}.$$

Доказательство. Из курса математического анализа вспомним, что две бесконечные числовые последовательности $\{a_n\}$ и $\{b_n\}$ называют эквивалентными и обозначают $a_n \underset{n\to\infty}{\sim} b_n$, если $\frac{a_n}{b_n} \underset{n\to\infty}{\longrightarrow} 1$. Покажем, что $C_K^k \underset{K\to\infty}{\sim} \frac{K^k}{k!}$. Действительно,

$$\frac{C_K^k k!}{K^k} = \frac{K!k!}{k!(K-k)!K^k} = \frac{A_K^k}{K^k} = \frac{K(K-1)\ldots(K-k+1)}{K^k} \underset{K\to\infty}{\longrightarrow} 1.$$

Из условия теоремы следует, что $K \sim pN$, поэтому $N - K \underset{N\to\infty}{\longrightarrow} \infty$, следовательно

$$C_{N-K}^{n-k} \underset{N\to\infty}{\sim} \frac{(N-K)^{n-k}}{(n-k)!}.$$

Из полученных эквивалентностей и условия теоремы, получаем

$$P_{N,K}(n,k) = \frac{C_K^k \cdot C_{N-K}^{n-k}}{C_N^n} \underset{N\to\infty}{\sim} \frac{K^k}{k!} \cdot \frac{(N-K)^{n-k}}{(n-k)!} \cdot \frac{n!}{N^n} =$$

$$= C_n^k \cdot \frac{K^k}{N^k} \cdot \frac{(N-K)^{n-k}}{N^{n-k}} = C_n^k \cdot \left(\frac{K}{N}\right)^k \cdot \left(1 - \frac{K}{N}\right)^{n-k} \underset{N\to\infty}{\longrightarrow}$$

$$C_n^k\, p^k (1-p)^{n-k}.$$

Отсюда, исходя из свойств пределов эквивалентных последовательностей (если $a_n \underset{n\to\infty}{\sim} b_n$ и существует предел $\lim\limits_{n\to\infty} b_n$, то $\lim\limits_{n\to\infty} a_n = \lim\limits_{n\to\infty} b_n$), получаем утверждение теоремы.

Пример 18. Среди 300 изделий 15 бракованные. Для проверки наугад выбрали 5 изделий. Какова вероятность, что среди них нет бракованных? Сравнить точное значение вероятности с приближённым, найденным по формуле Бернулли.

Решение. Пусть событие $A = \{$*среди пяти извлечённых изделий нет бракованных*$\}$. Выбрать 5 изделий из 300 можно C_{300}^5 способами. Событию A удовлетворяют те способы выбора, при которых пять изделий выбираются из 285 небракованных изделий. Число таких способов равно C_{285}^5. Приближённое значение вероятности события A с точностью до десятитысячной равно $P(A) = C_{285}^5/C_{300}^5 \approx 0.7724$.

Можно считать, что партия изделий велика, поэтому выбор одного за другим небракованных изделий сильно не меняет пропорции в этой партии. Поэтому вероятность выбора бракованного изделия на малом количестве шагов примерно равна $p = 15/300 = 0.05$. Поэтому по формуле Бернулли, получаем $P(A) \approx C_5^0 \cdot 0.05^0 \cdot 0.95^5 \approx 0.7738.$

§5. *Обобщённая схема Бернулли*

Пусть при одном испытании возможны m различных исходов. Обозначим эти исходы цифрами $1, 2, \ldots, m$ по их порядковому номеру. Пусть i-ый исход в одном испытании появляется с вероятностью $p_i,\ 1 \le i \le m,\ \sum\limits_{i=1}^{m} p_i = 1$. Обозначим через $P(n_1, \ldots, n_m)$ вероятность того, что в $n = n_1 + \cdots + n_m$ независимых испытаниях 1-ый исход появился n_1 раз, 2-ой исход – n_2 раз, …, m-ый исход – n_m раз. Такую серию из n независимых испытаний назовём *обобщённой схемой Бернулли.*

Теорема 12. Для любого целого положительного n и любых целых $n_1 \ge 0, \ldots,\ n_m \ge 0$ таких, что $n_1 + \cdots + n_m = n$ справедлива формула:

$$P(n_1, \ldots, n_m) = \frac{n!}{n_1! n_2! \ldots n_m!} \cdot p_1^{n_1} \cdot p_2^{n_2} \cdot \ldots \cdot p_m^{n_m}$$

Доказательство. Рассмотрим одно элементарное событие, благоприятствующее появлению первого исхода n_1 раз, второго исхода n_2 раз, …, m-го исхода n_m раз:

$$(1, 1, \ldots, 1, 2, \ldots, 2, \ldots, m, \ldots, m).$$

Это результат n экспериментов, когда все исходы появились в описанном заранее заданном порядке. Вероятность такого результата n независимых обобщённых испытаний Бернулли равна, очевидно, $p_1^{n_1} \cdot _2^{n_2} \cdot \ldots \cdot p_m^{n_m}$. Все остальные благоприятные исходы отличаются лишь расположением чисел $1, 2, \ldots, m$ на n местах. Число таких исходов, очевидно, равно числу способов расставить на n местах n_1 единиц, n_2 двоек, …, n_m чисел m, то есть,

$$C_n^{n_1} \cdot C_{n-n_1}^{n_2} \cdot \dots \cdot C_{n-n_1-\dots-n_{m-1}}^{n_m} = \frac{n!}{n_1! n_2! \dots n_m!}.$$

Пример 19. Правильную игральную кость подбросили 10 раз. Какова вероятность того, что два раза выпадет «шестёрка» и три раза – «пятёрка».

То, что кость правильная означает, что любое число очков (от единицы до шести) выпадает с вероятностью $1/6$. Обозначим событие $A = \{$из десяти подбрасываний два раза выпадет «шестёрка» и три раза – «пятёрка»$\}$ Ясно, что в пяти подбрасываниях в серии из десяти не появятся «пятёрка» и «шестёрка». Используя формулу теоремы 12, получаем:

$$P(A) = \frac{10!}{2!3!5!} \cdot \frac{1}{6^2} \cdot \frac{1}{6^3} \cdot \left(\frac{4}{6}\right)^5 = 280 \cdot \frac{4^5}{6^{10}} \approx 0.00474.$$

§6. Теорема Пуассона для схемы Бернулли

В этом параграфе рассматривается утверждение о приближённом вычислении вероятности какого-либо числа успехов в большом числе испытаний схемы Бернулли с маленькой вероятностью успеха (вероятность «редкого» события). Под словами «большое число» применительно к числу испытаний фактически понимают, что $n \to \infty$. Если при этом $p = p_n \not\to 0$, то, ясно, что вероятность получить любое конечное число успехов при растущем числе испытаний стремится к нулю. Чтобы этого не было, нужно чтобы вероятность успеха $p = p_n \to 0$ одновременно с ростом числа испытаний. Но от испытания к испытанию в схеме Бернулли вероятность успеха меняться не может. Поэтому, чтобы обойти эту проблему, будем рассматривать «схему серий»:

одно испытание с вероятностью успеха p_1;

два испытания с вероятностью успеха p_2;

…

n испытаний с вероятностью успеха p_n.

То есть, вероятность успеха меняется не внутри одной серии испытаний, а от серии к серии, когда меняется общее число испытаний. Обозначим ν_n число успехов в n−й серии испытаний.

Теорема 13 (Теорема Пуассона). Пусть $n \to \infty$ и $p_n \to 0$ так, чтобы $np_n \underset{n\to\infty}{\longrightarrow} \lambda$. Тогда $\forall\ k \geq 0$ вероятность k успехов в n испытаниях схемы Бернулли с вероятностью успеха p_n стремится к величине $\frac{\lambda^k}{k!} e^{-\lambda}$:

$$P(\nu_n = k) = C_n^k\, p_n^k\, (1 - p_n)^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda} \ \textit{ при } n \to \infty,$$

$p_n \to 0$ *так, что* $np_n \to \lambda > 0$.

Доказательство. Положим $\lambda_n = n\, p_n \underset{n\to\infty}{\longrightarrow} \lambda$. В доказательстве теоремы 11 показано, что $C_n^k \underset{n\to\infty}{\sim} \frac{n^k}{k!}$ при фиксированном значении k. Тогда имеем

$$C_n^k\, p_n^k\, (1-p_n)^{n-k} = C_n^k\, \frac{\lambda_n^k}{n^k} \left(1 - \frac{\lambda_n}{n}\right)^{n-k} \sim$$

$$\sim \frac{\cancel{n}^k}{k!}\, \frac{\lambda_n^k}{\cancel{n}^k} \left(1-\frac{\lambda_n}{n}\right)^n \left(1-\frac{\lambda_n}{n}\right)^{-k} \underset{n\to\infty}{\longrightarrow} \frac{\lambda^k}{k!}\, e^{-\lambda},$$

поскольку $\lambda_n^k \underset{n\to\infty}{\longrightarrow} \lambda^k$, $\left(1-\frac{\lambda_n}{n}\right)^n \underset{n\to\infty}{\longrightarrow} e^{-\lambda}$ и $\left(1-\frac{\lambda_n}{n}\right)^{-k} \underset{n\to\infty}{\longrightarrow} 1$, т.к.

$$\ln\left(1-\frac{\lambda_n}{n}\right)^n = n\ln\left(1-\frac{\lambda_n}{n}\right) \underset{n\to\infty}{\sim} n\left(-\frac{\lambda_n}{n}\right) \underset{n\to\infty}{\longrightarrow} -\lambda$$

и $\frac{\lambda_n}{n} \underset{n\to\infty}{\longrightarrow} 0$, поэтому $\quad \left(1-\frac{\lambda_n}{n}\right)^{-k} \underset{n\to\infty}{\longrightarrow} 1.$

Определение 25. Пусть $\lambda > 0$ – некоторая положительная постоянная. Набор чисел

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

называется *распределением Пуассона с параметром* λ.

Пример 20. Вероятность того, что изделие при транспортировке с завода повредится равно 0.0005. С завода было отправлено 4000 изделий. Какова вероятность того, что в пути повредится больше двух изделий?

Решение. Транспортировку каждого изделия можно рассматривать как независимый эксперимент, число которых ($n = 4000$) велико. Вероятность же появления события (повредить изделие) в каждом эксперименте мала ($p = 0.0005$) мала. Это даёт основание воспользоваться формулой Пуассона. Ясно, что $\lambda = np = 4000 \cdot 0.0005 = 2$. Необходимо найти вероятность

$$P(\nu_{4000} > 2) = 1 - P(\nu_{4000} = 0) - P(\nu_{4000} = 1) - P(\nu_{4000} = 2) =$$

$$= 1 - \frac{2^0}{0!}\, e^{-2} - \frac{2^1}{1!}\, e^{-2} - \frac{2^2}{2!}\, e^{-2} == 1 - 5e^{-2} \approx 0.31.$$

Теорема 14 (оценка погрешности в теореме Пуассона). Пусть $A \subseteq \{0, 1, 2, \ldots, n\}$, ν_n — число успехов в n испытаниях схемы Бернулли с вероятностью успеха p, $\lambda = np$. Тогда

$$\left|P(\nu_n \in A) - \sum_{k\in A} \frac{\lambda^k}{k!}\, e^{-\lambda}\right| = |\sum_{k\in A} C_n^k\, p^k (1-p)^{n-k} -$$

$$-\sum_{k\in A} \frac{\lambda^k}{k!}\, e^{-\lambda}| \le np^2 = \frac{\lambda^2}{n}.$$

Эту теорему оставляем пока без доказательства.

Глава 5. Случайные величины и их распределения

§1. Случайные величины

Пусть имеется случайный эксперимент и задано вероятностное пространство $(\Omega,\ \mathcal{A},\ P)$.

Определение 26. Функция $\xi : \Omega \to R^1$ называется *случайной величиной*, если $\forall\ x \in R^1$ множество $\{\xi < x\} = \{\omega \in \Omega : \xi(\omega) < x\} \in \mathcal{A}$ является событием, то есть принадлежит σ-алгебре событий $\mathcal{A}$.

Определение 27. Функция $\xi : \Omega \to R^1$ называется $\mathcal{A}$-*измеримой*, если $\{\omega \in \Omega : \xi(\omega) < x\} \in \mathcal{A}$ для любого $x \in R^1$.

Формально имеем, что случайная величина есть $\mathcal{A}$-измеримая функция, ставящая в соответствие каждому элементарному событию $\omega \in \Omega$ число $\xi(\omega) \in R^1$.

Пример 21. Один раз подбрасывается игральная кость. Пусть $\Omega = \{1, 2, 3, 4, 5, 6\}$ и пусть заданы две функции из Ω в R^1: $\xi(\omega) = \omega$ и $\zeta(\omega) = \omega^2$.

i. Если $\mathcal{A}$ есть множество всех подмножеств Ω (содержит $2^6 = 64$ событий), то ξ и ζ являются случайными величинами, поскольку любое множество элементарных событий принадлежит $\mathcal{A}$, в том числе и $\{\omega \in \Omega : \xi(\omega) < x\}$ и $\{\omega \in \Omega : \zeta(\omega) < x\}$. Можно записать соответствие между значениями случайных величин ξ и ζ и вероятностями $P(\xi = 1) = P(\xi = 2) = \cdots = P(\xi = 6) = P(\zeta = 1) = P(\zeta = 4) = \cdots = P(\zeta = 36) = \frac{1}{6}$, а можно представить эти значения в виде *«таблицы распределения вероятностей»* или, коротко, *«таблицы распределения»:*

ξ	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

ζ	1	4	9	16	25	36
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

.

ii. Пусть σ-алгебра событий $\mathcal{A}$ состоит из четырёх множеств: $\mathcal{A} = \{\Omega, \varnothing, \{1, 3, 5\}, \{2, 4, 6\}\ \}$, т.е. событием для этой σ-алгебры является, кроме достоверного и невозможного событий, ещё выпадение чётного числа очков, а ещё и нечётного числа очков. Убедимся, что при такой σ-алгебре ни ξ, ни ζ не являются случайными величинами, так как эти функции не $\mathcal{A}$-измеримы. Действительно, возьмём, например число $x = 3.2$. Заметим, что $\{\omega \in \Omega : \ \xi(\omega) < 3.2\} = \{1, 2, 3\} \notin \mathcal{A}$ и $\{\omega \in \Omega : \zeta(\omega) < 3.2\} = \{1\} \notin \mathcal{A}$.

iii. Пусть σ-алгебра событий $\mathcal{A}$ есть *тривиальная* σ-алгебра: $\mathcal{A} = \{\Omega, \varnothing\}$. Легко проверить, что измеримы относительно такой σ-алгебры только функции, принимающие единственное значение, т.е. постоянные.

§2. Дискретные случайные величины

Определение 28. Случайная величина ξ называется *дискретно распределённой* или имеет *дискретное распределение*, если существует конечный или счётный набор чисел $\{a_1, a_2, \dots\}$ такой, что:

i) $p_j = P(\xi = a_j) > 0 \ \ \forall \ \ j$; *ii)* $\sum\limits_{j=1}^{\infty} p_j = 1$. Таким образом, случайная величина ξ имеет дискретное распределение, если она принимает не более чем счётное множество значений.

Определение 29. Распределением дискретной случайной величины ξ называется соответствие $a_j \longleftrightarrow p_j$, которое часто представляют в виде таблицы.

§3. Примеры дискретных распределений

В этом параграфе перечислены некоторые основные дискретные случайные величины, часто используемые в примерах и во многих прикладных задачах. Естественно, класс дискретных случайных величин широк и перечислить всевозможные группы распределений не представляется возможным.

Определение 30. Случайная величина, принимающая единственное значение называются *вырожденной*.

Если вырожденная случайная величина ξ принимает значение a, то говорят, что с.в. вырождена в a, и пишут $\xi \sim I_a$. Ясно, что $P(\xi = a) = 1$ и таблица распределения имеет вид

ξ	a
P	1

.

Определение 31. Случайная величина ξ *имеет* распределение Бернулли *с параметром* p, и пишут $\xi \sim Bi(1, p)$, если ξ принимает значения 1 и 0 с вероятностями p и

$\boldsymbol{1-p}$, соответственно. Случайная величина ξ с таким распределением равна числу успехов в одном испытании схемы Бернулли с вероятностью успеха $\boldsymbol{p}$. Таблица распределения ξ имеет вид

ξ	0	1
$\boldsymbol{P}$	1-$\boldsymbol{p}$	$\boldsymbol{p}$

.

Определение 32. Случайная величина ξ имеет *биномиальное распределение* с параметрами $\boldsymbol{n}$ и $\boldsymbol{p}$, где $\boldsymbol{0 \le p \le 1}$, и пишут $\xi \sim \boldsymbol{Bi(n,p)}$, если ξ принимает значения $\boldsymbol{0,1,\ldots,n}$ с вероятностями $\boldsymbol{P(\xi=k)=C_n^k\,p^k\,(1-p)^{n-k}}$. Случайная величина ξ с таким распределением имеет смысл числа успехов в $\boldsymbol{n}$ независимых испытаниях Бернулли с вероятностью успеха $\boldsymbol{p}$. Таблица распределения ξ имеет вид

ξ	0	1	...	$\boldsymbol{k}$	...	$\boldsymbol{n}$
$\boldsymbol{P}$	$\boldsymbol{(1-p)^n}$	$\boldsymbol{np(1-p)^{n-1}}$	...	$\boldsymbol{C_n^k\,p^k\,(1-p)^{n-k}}$	...	$\boldsymbol{p^n}$

.

Определение 33. Случайная величина τ имеет *геометрическое распределение* с параметром $\boldsymbol{p}$, где $\boldsymbol{0 \le p \le 1}$, и пишут $\tau \sim \boldsymbol{G(p)}$, если τ принимает значения 1,2,3,... с вероятностями $\boldsymbol{P(\tau=k)=p(1-p)^{k-1}}$. Случайная величина τ с таким распределением имеет смысл номера первого успешного испытания в схеме испытаний Бернулли с вероятностью успеха $\boldsymbol{p}$. Таблица распределения τ имеет вид

τ	1	2	...	$\boldsymbol{k}$	...
$\boldsymbol{P}$	$\boldsymbol{p}$	$\boldsymbol{p(1-p)}$	...	$\boldsymbol{p\,(1-p)^{k-1}}$	...

.

Определение 34. Случайная величина ξ имеет *распределение Пуассона* с параметром $\boldsymbol{\lambda}$, где $\boldsymbol{\lambda > 0}$, и пишут $\xi \sim \boldsymbol{\Pi(\lambda)}$, если ξ принимает значения 0,1,2,... с вероятностями $\boldsymbol{P(\xi=k)=\frac{\lambda^k}{k!}\,e^{-\lambda}}$. Таблица распределения ξ имеет вид

ξ	0	1	...	$\boldsymbol{k}$	...
$\boldsymbol{P}$	$\boldsymbol{e^{-\lambda}}$	$\boldsymbol{\lambda e^{-\lambda}}$	...	$\boldsymbol{\frac{\lambda^k}{k!}e^{-\lambda}}$	...

.

Определение 35. Случайная величина ξ имеет *гипергеометрическое распределение* с параметрами $\boldsymbol{n,\ N-K}$, где $\boldsymbol{K \le N}$, $\boldsymbol{n \le N}$, если ξ принимает целые значения от $\boldsymbol{\max\{0, N-K-n\}}$ до $\boldsymbol{\min\{n,K\}}$ с вероятностями $\boldsymbol{P(\xi=k)=\frac{C_K^k\,C_{N-K}^{n-k}}{C_N^n}}$. Случайная величина ξ с таким распределением имеет смысл числа белых шаров среди $\boldsymbol{n}$ шаров, выбранных наудачу и без возвращения из урны, содержащей $\boldsymbol{K}$ белых шаров и $\boldsymbol{N-K}$ не белых.

В практических задачах встречаются не только дискретные случайные величины, но и так называемые абсолютно-непрерывные и сингулярные случайные величины, а также их смеси. Но об этом будем говорить позже.

Глава 6. Функция распределения

§1. Определение и свойства функции распределения

Разумно описать распределение случайной величины, задав для любого множества вероятность принять значения из этого множества. И это действительно полная характеризация распределения, но с ней невозможно работать, поскольку слишком много разных множеств на прямой. А можно ли ограничиться заданием вероятностей попадания в какую-нибудь «меньшую» совокупность множеств на прямой? Известно, что можно ограничиться только вероятностями попадания в интервалы $(-\infty, x)$ для всех $x \in R^1$.

Определение 36. Функцией распределения случайной величины ξ называется функция $F_\xi(x):\ R^1 \to [0,1]$, при каждом $x \in R^1$ равная

$$F_\xi(x) = P(\xi < x) = P\{\omega \in \Omega:\ \xi(\omega) < x\}.$$

Пример 22. Случайная величина $\xi \sim I_c$ имеет вырожденное распределение в точке c. Тогда,

$$F_\xi(x) = P(\xi < x) = P(c < x) = \begin{cases} 0, & x \le c; \\ 1, & x > c. \end{cases}$$

Пример 23. Случайная величина $\xi \sim Bi(1,p)$ имеет распределение Бернулли с вероятностью успеха p. Тогда,

$$F_\xi(x) = P(\xi < x) = \begin{cases} 0, & x \le 0; \\ 1-p, & 0 < x \le 1; \\ 1, & x > 1. \end{cases}$$

Пример 24. Случайная величина ξ имеет равномерное распределение на отрезке $[a,b]$ и пишем $\xi \sim R(a,b)$, если ξ – координата точки, брошенной наудачу на отрезок [a, b] числовой прямой. Это распределение задаётся функцией распределения:

$$F_\xi(x) = P(\xi < x) = \begin{cases} 0, & x < a; \\ \frac{x-a}{b-a}, & a \le x \le b; \\ 1, & x > b. \end{cases}$$

Теорема 15 (свойства функции распределения).

Функция распределения $F_\xi(x)$ обладает следующими общими свойствами:

(i) Функция распределения $F_\xi(x)$ не убывает: если $x_1 < x_2$, то $F(x_1) \le F(x_2)$.

(ii) Существуют пределы $\lim\limits_{x\to-\infty} F_\xi(x) = 0$ и $\lim\limits_{x\to+\infty} F_\xi(x) = 1$.

(iii) Функция распределения $F_\xi(x)$ непрерывна слева: $F_\xi(x_0-0) = \lim\limits_{x\to x_0-0} F_\xi(x) = F_\xi(x_0)$.

Доказательство свойства (i).

Если $x_1 < x_2$, то $\{\omega \in \Omega : \xi(\omega) < x_1\} \subseteq \{\omega \in \Omega : \xi(\omega) < x_2\}$. Поэтому $F_\xi(x_1) = P(\xi < x_1) \leq P(\xi < x_2) = F_\xi(x_2)$.

Доказательство свойства (ii).

Из курса математического анализа, хорошо известно, что если существует конечный предел $\lim\limits_{x\to a} f(x)$, то для произвольной числовой последовательности x_n, такой, что $x_n \to a$, $x_n \neq a$ справедливо равенство $\lim\limits_{x\to a} f(x) = \lim\limits_{n\to\infty} f(x_n)$.

Поэтому, существование пределов в свойстве (ii) вытекает из монотонности и ограниченности функции $F_\xi(x)$. Для доказательства равенства $\lim\limits_{x\to-\infty} F_\xi(x) = 0$ достаточно показать, что $\lim\limits_{n\to\infty} F_\xi(-n) = 0$. Представим событие $\{\xi < -k\}$ для любого целого положительного k как счётное объединение несовместных событий:

$$\{\xi < -k\} = \bigcup_{i=k}^{\infty} \{-i-1 \leq \xi < -i\}.$$

В силу σ-аддитивности и ограниченности вероятностной меры, имеем

$$P(\xi < -k) = \sum_{i=k}^{\infty} P(-i-1 \leq \xi < -i) \leq 1,$$

и по свойству сходящегося ряда имеем то, что «хвост» сходится к нулю. Поэтому, получаем

$$\lim_{n\to\infty} P(\xi < -n) = \lim_{n\to\infty} \sum_{i=n}^{\infty} P(-i-1 \leq \xi < -i) = 0.$$

Для доказательства второй части свойства (ii) $\lim\limits_{x\to+\infty} F_\xi(x) = 1$ достаточно доказать, что $\lim\limits_{n\to\infty} F_\xi(n) = 1$ или, что аналогично, $\lim\limits_{n\to\infty} P(\xi \geq n) = 0$. Представим событие $\{\xi \geq k\}$ для любого целого положительного k как счётное объединение несовместных событий:

$$\{\xi \geq k\} = \bigcup_{i=k}^{\infty} \{i \leq \xi < i+1\}.$$

В силу σ-аддитивности и ограниченности вероятностной меры, имеем

$$P(\xi \geq k) = \sum_{i=k}^{\infty} P(i \leq \xi < i+1) \leq 1,$$

и снова, по свойству сходящегося ряда имеем то, что «хвост» сходится к нулю. Поэтому, получаем

$$\lim_{n\to\infty} P(\xi \geq n) = \lim_{n\to\infty} \sum_{i=n}^{\infty} P(i \leq \xi < i+1) = 0.$$

Свойство (ii) полностью доказано.

Доказательство свойства (iii).

Также, как и в (ii) достаточно доказать, что $\lim\limits_{n\to\infty} F_\xi\left(x_0 - \frac{1}{n}\right) = F_\xi(x_0)$ или, что то же самое,

$$\lim_{n\to\infty}\left[F_\xi(x_0) - F_\xi\left(x_0 - \frac{1}{n}\right)\right] = \lim_{n\to\infty} P\left(x_0 - \frac{1}{n} \le \xi < x_0\right) = 0$$

Представим событие $\{\xi < x_0\}$ как счётное объединение несовместных событий

$$\{\xi < x_0\} = \{\xi < x_0 - 1\} \cup \bigcup_{i=1}^{\infty}\left\{x_0 - \frac{1}{i} \le \xi < x_0 - \frac{1}{i+1}\right\}.$$

В силу σ-аддитивности вероятностной меры, получаем

$$P(\xi < x_0) = P(\xi < x_0 - 1) + \sum_{i=1}^{\infty} P\left(x_0 - \frac{1}{i} \le \xi < x_0 - \frac{1}{i+1}\right) \le 1,$$

откуда заключаем, что $\lim\limits_{n\to\infty}\left[F_\xi(x_0) - F_\xi\left(x_0 - \frac{1}{n}\right)\right] = \lim\limits_{n\to\infty} P\left(x_0 - \frac{1}{n} \le \xi < x_0\right) = \lim\limits_{n\to\infty}\sum\limits_{i=n}^{\infty} P\left(x_0 - \frac{1}{i} \le \xi < x_0 - \frac{1}{i+1}\right) = 0.$ Отсюда следует, что $F_\xi(x_0) = F_\xi(x_0 - 0)$, что и означает непрерывность слева и, тем самым, свойство (iii) доказано.

Теорема 16. Если функция $F : R^1 \to [0,1]$ удовлетворяет свойствам $(i) - (iii)$, то F есть функция распределения некоторой случайной величины ξ, то есть найдется вероятностное пространство $(\Omega, \mathcal{A}, P)$ и случайная величина ξ на этом пространстве, что $F(x) \equiv F_\xi(x)$.

Эту теорему оставляем без доказательства.

Рассмотрим ещё два свойства функции распределения.

(iv) В любой точке x_0 разность $F_\xi(x_0 + 0) - F_\xi(x_0)$ равна вероятности $P(\xi = x_0)$:

$$F_\xi(x_0 + 0) - F_\xi(x_0) = \lim_{x\to x_0+0} F_\xi(x) - F_\xi(x_0) = P(\xi = x_0), \text{ или,}$$

$$F_\xi(x_0 + 0) = \lim_{x\to x_0+0} F_\xi(x) = F_\xi(x_0) + P(\xi = x_0) = P(\xi \le x_0).$$

Доказательство этого свойства проводится аналогично доказательствам свойств (ii) и (iii).

Отметим, что разность $F_\xi(x_0 + 0) - F_\xi(x_0)$ между пределом при стремлении к x_0 справа и значением в точке x_0 равна величине скачка функции распределения, и равна нулю, если функция распределения непрерывна в точке x_0, т.е. $P(\xi = x_0) = 0$.

(v) Для любой случайной величины ξ имеет место равенство $P(a \le \xi < b) = F_\xi(b) - F_\xi(a)$. Если же функция распределения $F_\xi(x)$ непрерывна (для любого x, или только в точках a и b), то

$$P(a \le \xi < b) = P(a < \xi < b) = P(a \le \xi \le b) =$$

$$= P(a < \xi \le b) = F_\xi(b) - F_\xi(a).$$

Доказательство. Доказать нужно только равенство $P(a \le \xi < b) = F_\xi(b) - F_\xi(a)$, поскольку все остальные равенства следуют из него с учётом свойства iv. Этим равенством мы уже несколько раз пользовались, когда доказывали свойства (ii) и (iii).

Замечаем, что $\{\xi < a\} \cup \{a \le \xi < b\} = \{\xi < b\}$, поскольку события в левой части последнего равенства являются несовместными. Поэтому

$$P(\xi < a) + P(a \le \xi < b) = P(\xi < b),$$

или $F_\xi(a) + P(a \le \xi < b) = F_\xi(b)$, что и требовалось доказать.

(vi) Случайная величина ξ имеет дискретное распределение тогда и только тогда, когда функция распределения F является неубывающей ступенчатой функцией. При этом, возможные значения ξ — точки a_i скачков F_ξ, и $p_i = P(\xi = a_i) = F_\xi(a_i + 0) - F_\xi(a_i)$ — величины скачков. Это свойство непосредственно следует из свойств (iv) и (iv)

Из математического анализа известен тот факт, что функция, монотонная на области определения может иметь на ней не более, чем счётное множество точек разрыва, и все они первого рода («скачков»).

Действительно, скачков величиной более 1/2 функция распределения может иметь не более одного, величиной более 1/3 не более двух, ..., величиной более $1/n$ не более $n-1$, ... Поэтому, максимальное множество «скачков» даёт счётное объединение конечных множеств, а это множество не может быть более чем счётным.

Глава 7. Абсолютно непрерывные распределения

§1. Определение и свойства абсолютно непрерывных распределений

Определение 37. Случайная величина ξ имеет *абсолютно непрерывное распределение*, если существует неотрицательная функция $f_\xi(x)$ такая, что для любого $x \in R^1$ функция распределения $F_\xi(x)$ представима в виде $F_\xi(x) = \int\limits_{-\infty}^{x} f_\xi(t)dt$. Функция $f_\xi(x)$ называется *плотностью распределения* случайной величины ξ.

Теорема 17. Плотность распределения обладает свойствами:

$(i)\quad f_\xi(x) \geq 0\ \forall x \in R^1$; $\qquad (ii)\quad \int\limits_{-\infty}^{+\infty} f_\xi(t)dt = 1$.

Доказательство. (i) выполнено по определению плотности. Докажем (ii). $\int\limits_{-\infty}^{+\infty} f_\xi(t)dt = \lim\limits_{x\to+\infty} \int\limits_{-\infty}^{x} f_\xi(t)dt = \lim\limits_{x\to+\infty} F_\xi(x) = 1$ по свойству (ii) функции распределения.

Эти два свойства полностью характеризуют класс плотностей:

Лемма. Если функция f обладает свойствами (i) и (ii), то существует вероятностное пространство и случайная величина ξ на нём, для которой f является плотностью распределения.

Доказательство. Пусть Ω есть область, заключенная между осью абсцисс и графиком функции f. Площадь области Ω равна 1 по свойству (ii). Пусть случайная величина ξ есть (см., рис 7.1) абсцисса точки, наудачу брошенной в эту область. Тогда для любого $x \in R^1$

$$F_\xi(x) = P(\xi < x) = P(\text{точка попала в область } D_x) =$$

$$= \frac{\text{площадь } D_x}{\text{площадь } \Omega} = \int\limits_{-\infty}^{x} f(t)dt,$$

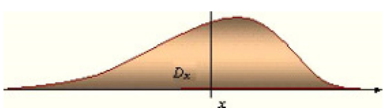

Рис. 7.1

т.е. f является плотностью распределения с.в. ξ.

Продолжим рассматривать свойства плотностей.

(iii) Если случайная величина ξ имеет абсолютно непрерывное распределение, то её функция распределения всюду непрерывна.

Доказательство. Этот факт следует из представления $F_\xi(x) = \int\limits_{-\infty}^{x} f_\xi(t)dt$ и непрерывности интеграла как функции верхнего предела если f хотя бы интегрируема.

Следствие 1. Если случайная величина ξ имеет абсолютно непрерывное распределение, то $P(\xi = x) = 0\ \ \forall\, x \in R^1$.

(iv) Если случайная величина ξ имеет абсолютно непрерывное распределение, то её функция распределения дифференцируема почти всюду и $f_\xi(x) = F'_\xi(x) = \frac{d}{dx}F_\xi(x)$ для почти всех x.

Замечание. Термин для «почти всех» означает «для всех, кроме (*возможно*) x из некоторого множества нулевой меры (длины)». Заметим, что стоящую под интегралом функцию можно изменить в одной точке (или на множестве нулевой длины), и интеграл от этого не изменится.

(v) Если случайная величина ξ имеет абсолютно непрерывное распределение, то

$$P(a < \xi < b) = P(a \le \xi < b) = P(a < \xi \le b) =$$

$$= P(a \le \xi \le b) = \int_a^b f(t)dt.$$

Доказательство. Действительно,

$$P(a \le \xi < b) = F_\xi(b) - F_\xi(a) = \int_{-\infty}^{b} f(t)dt - \int_{-\infty}^{a} f(t)dt = \int_a^b f(t)dt.$$

Остальные равенства вытекают из следствия 1.

§2. Примеры абсолютно непрерывных распределений

(j) **Равномерное распределение.** Это распределение уже встречалось выше. С.в. ξ имеет *равномерное распределение* на отрезке $[a, b]$, и обозначают это $\xi \sim R(a, b)$, если

$$F_\xi(x) = P(\xi < x) = \begin{cases} 0, & x < a; \\ \frac{x-a}{b-a}, & a \le x \le b; \\ 1, & x > b; \end{cases}$$

$$f_\xi(x) = \begin{cases} 0, & x < a \\ \frac{1}{b-a}, & a \le x \le b \\ 0, & x > b \end{cases} = \frac{1_{[a,b]}(x)}{b-a}.$$

Здесь $1_A(x)$ обозначает индикатор множества A, т.е.

$$1_A(x) = \begin{cases} 1, & x \in A; \\ 0, & x \notin A. \end{cases}$$

Заметим, что в точках a и b функция распределения не дифференцируема, и плотность в этих точках можно задать произвольно.

(jj) **Показательное распределение.** С.в. ξ имеет *показательное* распределение с параметром $\alpha,\ \alpha > 0$ и обозначают это $\xi \sim E(\alpha)$, если

$$F_\xi(x) = P(\xi < x) = \begin{cases} 0, & x < 0; \\ 1 - e^{-\alpha x}, & x \geq 0, \end{cases}$$

$$f_\xi(x) = \begin{cases} 0, & x < 0 \\ \alpha e^{-\alpha x}, & x \geq 0 \end{cases} = \alpha e^{-\alpha x} 1_{[0,+\infty)}(x).$$

Показательное распределение является единственным абсолютно непрерывным распределением, для которого выполнено свойство «нестарения» и в этом смысле оно является непрерывным аналогом дискретного геометрического распределения.

Теорема 18 (свойство «нестарения»). Пусть $\xi \sim E(a)$. Тогда $\forall\ x,\ y > 0$

$$P(\xi > x + y \mid \xi > x) = P(\xi > y).$$

Доказательство. $P(\xi > x + y \mid \xi > x) = \frac{P(\{\xi > x+y\} \cap P\{\xi > x\})}{P(\xi > x)} = \frac{P(\xi > x+y)}{P(\xi > x)} = \frac{e^{-\alpha(x+y)}}{e^{-\alpha x}} = = e^{-\alpha y} = 1 - F_\xi(y) = P(\xi > y).$

(jjj) **Нормальное распределение.** С.в. ξ имеет нормальное распределение с параметрами a и σ^2, где $a \in R^1$ и $\sigma^2 > 0$, и обозначают $\xi \sim N(a, \sigma^2)$, если ξ имеет плотность распределения:

$$f_\xi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right) \quad \forall\ x \in R^1.$$

Убедимся, что для $f_\xi(x)$ является плотностью распределения. Поскольку $f_\xi(x) > 0$ для всех $x \in R^1$, то свойство (i) для плотности выполнено. Проверим выполнение свойства (ii). Известно, что

$$\int\limits_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{2\pi} \quad - \quad \text{интеграл Пуассона.}$$

Поэтому,

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int\limits_{-\infty}^{+\infty} exp\left(-\frac{(x-a)^2}{2\sigma^2}\right) dx = \left[u = \frac{x-a}{\sigma},\ dx = \sigma du\right] =$$

$$= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} e^{-u^2/2} du = 1.$$

Для полноты картины, напомним, как вычисляется интеграл Пуассона.

$$\left(\int\limits_{-\infty}^{+\infty} e^{-x^2/2}dx\right)^2 = \int\limits_{-\infty}^{+\infty} e^{-x^2/2}dx \cdot \int\limits_{-\infty}^{+\infty} e^{-y^2/2}dy =$$

$$\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} \exp\left(-\frac{x^2+y^2}{2}\right) dx\, dy =$$

$=$(*полярная замена координат* $x = \rho\cos(\varphi)$, $y = \rho\sin(\varphi)$, $dx\, dy = \rho d\rho d\varphi$) $=$

$$= \int\limits_{0}^{2\pi}\int\limits_{0}^{+\infty} \rho\exp\left(-\frac{\rho^2}{2}\right) d\varphi\, d\rho = \int\limits_{0}^{2\pi} d\varphi \int\limits_{0}^{+\infty} \exp\left(-\frac{\rho^2}{2}\right) d\left(\frac{\rho^2}{2}\right) = 2\pi.$$

(jv) Гамма распределение. С.в. ξ имеет гамма распределение с параметрами α и λ, где $\alpha > 0$ и $\lambda > 0$, и обозначают $\xi \sim \Gamma(\alpha, \lambda)$, если ξ имеет плотность распределения:

$$f_\xi(x) = \frac{x^{\lambda-1}\, e^{-\frac{x}{\alpha}}}{\alpha^\lambda\, \Gamma(\lambda)}\, 1_{[0,+\infty)}(x),$$

где $\Gamma(\lambda) = \int\limits_0^{+\infty} x^{\lambda-1}e^{-\lambda}dx$ - гамма функция Эйлера. Легко показать, что для целых $k > 0$ справедливо равенство $\Gamma(k+1) = k\ \Gamma(k) = k!$ и $0! = \Gamma(1) = 1.$ Кстати, равенством $\gamma! = \Gamma(\gamma+1)$ задают определение факториала любого положительного числа γ, не обязательно целого.

Ясно, что гамма распределение $\Gamma(\alpha, 1)$ совпадает с показательным распределением $E(1/\alpha)$.

(v) Распределение Коши. Пусть распределение случайной величины ξ задаётся плотностью $f_\xi(x)$, имеющей вид:

$$f_\xi(x) = \frac{1}{\pi\gamma\left[1+\left(\frac{x-x_0}{\gamma}\right)^2\right]} = \frac{1}{\pi}\left[\frac{\gamma}{(x-x_0)^2+\gamma^2}\right],$$

где

$x_0 \in R^1$ — параметр сдвига; $\gamma > 0$ — параметр масштаба. Тогда говорят, что ξ имеет распределение Коши и пишут $\xi \sim C(x_0, \gamma)$. Если $x_0 = 0$ и $\gamma = 1$, то такое распределение называется стандартным распределением Коши.

Функция распределения Коши имеет вид:
$F_\xi(x) = \frac{1}{\pi}\operatorname{arctg}\left(\frac{x-x_0}{\gamma}\right) + \frac{1}{2}$.

§3. Свойства нормального распределения

Нормальное распределение (иногда называется *гауссовским* по имени Карла Гаусса) играет исключительно важную роль в теории вероятностей, поэтому будут подробно рассмотрены все свойства этого распределения.

Нормальное распределение задаётся при помощи плотности распределения. Связано это с тем, что нельзя выписать первообразную от функции $e^{-x^2/2}$ в виде элементарной функции, а можно, например, в виде интеграла. Поэтому, функция распределения нормального закона записывается в виде:

$$F_\xi(x) = \Phi_{a,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int\limits_{-\infty}^{x} \exp\left(-\frac{(u-a)^2}{2\sigma^2}\right) du.$$

Часто будет использоваться обозначение $\Phi_{a,\sigma}(x)$ для функции распределения нормального распределения с параметрами a и σ^2.

Определение 38. Нормальное распределение $N(0,1)$ (при $a=0$ и $\sigma^2=1$ называется *стандартным нормальным распределением.*

Плотность случайной величины $\xi \sim N(0,1)$ имеет вид $f_\xi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ при любом $x \in R^1$, а функция распределения $\Phi_{0,1}(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-u^2/2} du$ табулирована – таблицы её значений приводятся в более подробных руководствах по теории вероятностей.

Теорема 19. Для любого $x \in R^1$ справедливо равенство $\Phi_{a,\sigma}(x) = \Phi_{0,1}\left(\frac{x-a}{\sigma}\right)$.

Доказательство.

$$\Phi_{a,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int\limits_{-\infty}^{x} \exp\left(-\frac{(u-a)^2}{2\sigma^2}\right) du = \left[\begin{array}{c} v=\frac{u-a}{\sigma},\ du=\sigma dv \\ \hline u=x \longrightarrow v=\frac{x-a}{\sigma} \end{array}\right] =$$

$$= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\frac{x-a}{\sigma}} \exp\left(-\frac{v^2}{2}\right) dv = \Phi_{0,1}\left(\frac{x-a}{\sigma}\right).$$

Следствие 1. Если с.в. $\xi \sim N(a,\sigma^2)$, то с.в. $\zeta = \frac{\xi-a}{\sigma} \sim N(0,1)$.

Доказательство. $F_\zeta(x) = P(\zeta < x) = P\left(\frac{\xi-a}{\sigma} < x\right) = P(\xi < \sigma x + a) =$
$= \Phi_{a,\sigma}(\sigma x + a) = \Phi_{0,1}(\frac{\sigma x + a - a}{\sigma}) = \Phi_{0,1}(x).$

Следствие 2. Если с.в. $\xi \sim N(a,\sigma^2)$, то

$$P(x_1 < \xi < x_2) = \Phi_{a,\sigma}(x_2) - \Phi_{a,\sigma}(x_1) =$$

$$= \Phi_{0,1}\left(\frac{x_2 - a}{\sigma}\right) - \Phi_{0,1}\left(\frac{x_1 - a}{\sigma}\right).$$

Из приведённых утверждений заключаем,что нахождение любых вероятностей для нормально распределённой случайной величины сводится к вычислению функции распределения $\Phi_{0,1}(x)$.

Рассмотрим несколько свойств стандартного нормального распределения.

(i) $\Phi_{0,1}(-x) = 1 - \Phi_{0,1}(x), \quad \forall x \geq 0$. Равенство «хвостов» ф.р. стандартного нормального распределения. Очевидным образом следует из чётности плотности.

(ii) $\Phi_{0,1}(0) = 1/2$. Очевидно, если в равенстве предыдущего свойства положить $x = 0$, то получаем это равенство.

(iii) Если $\xi \sim N(0,1)$, то $P(|\xi| < x) = 1 - 2\Phi_{0,1}(-x) = 2\Phi_{0,1}(x) - 1.$

Действительно, $P(|\xi| < x) = P(-x < \xi < x) = \Phi_{0,1}(x) - \Phi_{0,1}(-x) =$
$=$ *(по свойству i)* $= 1 - 2\Phi_{0,1}(-x) = 2\Phi_{0,1}(x) - 1$.

(iv) ***(правило трёх сигм)*** Если $\xi \sim N(a, \sigma^2)$, то $P(|\xi - a| > 3\sigma) \approx 0.0027$.

Доказательство. $P(|\xi - a| > 3\sigma) = 1 - P(|\frac{\xi - a}{\sigma}| < 3) = 1 - (1 - 2\Phi_{0,1}(-3)) = 2\Phi_{0,1}(-3) \approx 2 \cdot 0.00135 = 0.0027.$

Смысл этого свойства заключается в том, что «почти всё» распределение с.в. ξ сосредоточено на отрезке $[a - 3\sigma, a + 3\sigma]$. Иначе говоря, вероятность принять с.в. ξ значения вне этого промежутка крайне мала.

Глава 8. Случайные векторы и их распределения

§1. Свойства функции совместного распределения

Определение 39. Если случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ заданы на одном вероятностном пространстве, то вектор $\overrightarrow{\xi} = (\xi_1, \xi_2, \ldots, \xi_n)$ называется *случайным вектором.*

Определение 40. Функция
$F_{\overrightarrow{\xi}}(\overrightarrow{x}) = P\left\{\omega \in \Omega : \xi_1(\omega) < x_1,\ \xi_2(\omega) < x_2, \ldots,\ \xi_n(\omega) < x_n\right\} =$
$= P(\xi_1 < x_1, \xi_2 < x_2, \ldots, \xi_n < x_n)$ называется функцией распределения случайного вектора $\overrightarrow{\xi}$ или функцией совместного распределения случайных величин $\xi_1, \xi_2, \ldots, \xi_n$.

Для простоты обозначений все дальнейшие рассуждения и формулировки приводятся в случае $n = 2$ для случайного вектора $\overrightarrow{\xi} = (\xi_1, \xi_2)$.

Свойства функции совместного распределения:

(i) $\quad 0 \leq F_{\xi_1,\xi_2}(x, y) \leq 1$;

(ii) Ф.р. $F_{\xi_1,\xi_2}(x,y)$ не убывает по каждой координате вектора (x,y);

(iii) $\lim\limits_{x\to-\infty} F_{\xi_1,\xi_2}(x,y)=0,\ \lim\limits_{y\to-\infty} F_{\xi_1,\xi_2}(x,y)=0$;

(iv) $\lim\limits_{(x,y)\to+\infty} F_{\xi_1,\xi_2}(x,y)=1$;

(v) $F_{\xi_1,\xi_2}(+\infty,y)=\lim\limits_{x\to+\infty} F_{\xi_1,\xi_2}(x,y)=F_{\xi_2}(y)$,

$F_{\xi_1,\xi_2}(x,+\infty)=\lim\limits_{y\to+\infty} F_{\xi_1,\xi_2}(x,y)=F_{\xi_1}(x)$ –

функции $F_{\xi_1}(x)$ и $F_{\xi_2}(y)$ называются маргинальными распределениями (это по существу функции распределения одномерных с.в. ξ_1 и ξ_2 соответственно);

(vi) Ф.р. $F_{\xi_1,\xi_2}(x,y)$ по каждой координате вектора (x,y) непрерывна слева.

Доказательство этих свойств абсолютно аналогично одномерному случаю.

§2. Типы многомерных распределений

Будем рассматривать только два случая, когда совместное распределение координат случайного вектора $\overrightarrow{\xi}$ либо дискретно, либо абсолютно непрерывно. Многообразие различных возможных вариантов случайных векторов эти случаи, конечно, не покрывают, но являются двумя основными классами распределений.

Определение 41. Случайные величины $\xi,\ \zeta$, заданные на одном вероятностном пространстве, имеют *совместное дискретное распределение*, если существует конечный или счётный набор пар чисел $\{a_i,b_j\}$ такой, что

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} P(\xi=a_i,\ \zeta=b_j)=1.$$

В случае, когда хотя бы одна из с.в. принимает конечное множество значений, соответствующий символ ∞ заменяется натуральным числом. Это надо иметь в виду и в дальнейшем, когда речь идёт о конечном или счётном множестве значений.

Определение 42. *Таблицей совместного распределения* случайных величин ξ и ζ называют таблицу, на пересечении i-й строки и j-го столбца которой (или наоборот, столбца - строки) стоит число $P(\xi=a_i,\ \zeta=b_j)$.

Таблицы распределений каждой из случайных величин ξ и ζ в отдельности, т.е. таблицы частных, или *маргинальных* распределений восстанавливаются по таблице совместного распределения при помощи очевидных формул:

$$P(\xi=a_i)=\sum_{j=1}^{\infty} P(\xi=a_i,\ \zeta=b_j) \quad \text{и}$$

$$P(\zeta=b_j)=\sum_{i=1}^{\infty} P(\xi=a_i,\ \zeta=b_j).$$

Определение 43. С.в. $\xi,\ \zeta$, заданные на одном вероятностном пространстве, имеют *совместное абсолютно непрерывное распределение*, если существует функция $f_{\xi,\zeta}(x,y) \geq 0$ такая, что для любой точки $(x,y) \in R^2$

$$F_{\xi,\zeta}(x,y) = P(\xi < x,\ \zeta < y) = \int\limits_{-\infty}^{x} \left(\int\limits_{-\infty}^{y} f_{\xi,\zeta}(u,v)dv \right) du.$$

Если такая функция $f_{\xi,\zeta}(x,y) \geq 0$ существует, то она называется *плотностью совместного распределения* случайных величин $\xi,\ \zeta$.

Плотность совместного распределения обладает свойствами, аналогичными свойствам плотности распределения одной случайной величины:

$$(i)\ f_{\xi,\zeta}(x,y) \geq 0\ \forall\ (x,y) \in R^2; \quad (ii) \int\limits_{-\infty}^{+\infty} \left(\int\limits_{-\infty}^{+\infty} f_{\xi,\zeta}(u,v)dv \right) du = 1.$$

Фактически, любая функция, обладающая этими свойствами, является совместной плотностью некоторого распределения. Если совместное распределение абсолютно непрерывно, то по совместной функции распределения его совместная плотность находится как смешанная частная производная:

$(iii)\ f_{\xi,\zeta}(x,y) = \frac{\partial^2}{\partial x\, \partial y} F_{\xi,\zeta}(x,y)$ для точек непрерывности плотности.

Теорема 20. Если случайные величины $\xi,\ \zeta$ имеют совместное абсолютно непрерывное распределение с плотностью $f_{\xi,\zeta}(x,y)$, то ξ и ζ также имеют абсолютно непрерывные распределения с плотностями:

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

Доказательство. Из свойства (ii), имеем:

$$F_{\xi}(x) = \lim_{y \to +\infty} F_{\xi,\zeta}(x,y) = \int\limits_{-\infty}^{x} \left(\int\limits_{-\infty}^{+\infty} f_{\xi,\zeta}(u,v)dv \right) du = \int\limits_{-\infty}^{x} f_{\xi}(u)du,$$

и аналогично:

$$F_{\zeta}(y) = \lim_{x \to +\infty} F_{\xi,\zeta}(x,y) = \int\limits_{-\infty}^{+\infty} \left(\int\limits_{-\infty}^{y} f_{\xi,\zeta}(u,v)dv \right) du =$$

$$\int\limits_{-\infty}^{y} \left(\int\limits_{-\infty}^{-\infty} f_{\xi,\zeta}(u,v)du \right) dv = \int\limits_{-\infty}^{y} f_{\zeta}(v)dv,$$

что и доказывает утверждение теоремы.

§3. Независимость случайных величин

Определение 44. Случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ называются *независимыми*, если для любых борелевских множеств $B_1, B_2, \ldots, B_n$ имеет место равенство:

$$P(\xi_1 \in B_1,\ \xi_2 \in B_2, \ldots, \xi_n \in B_n) =$$

$$= P(\xi_1 \in B_1) \cdot P(\xi_2 \in B_2) \cdot \ldots \cdot P(\xi_n \in B_n).$$

Это же определение можно сформулировать в терминах функций распределения.

Определение 45. Случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ называются *независимыми*, если для любых $x_1, x_2, \ldots, x_n$ имеет место равенство:

$$F_{\xi_1,\xi_2,\ldots,\xi_n}(x_1, x_2, \ldots, x_n) = F_{\xi_1}(x_1) \cdot F_{\xi_2}(x_2) \cdot \ldots \cdot F_{\xi_n}(x_n).$$

Для случайных величин с дискретным распределением эквивалентное определение независимости будет следующее

Определение 46. Дискретные случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ называются *независимыми*, если для любого набора чисел $a_1, a_2, \ldots, a_n$ имеет место равенство:

$$P(\xi_1 = a_1,\ \xi_2 = a_2, \ldots, \xi_n = a_n) =$$

$$= P(\xi_1 = a_1) \cdot P(\xi_2 = a_2) \cdot \ldots \cdot P(\xi_n = a_n).$$

Для случайных величин с совместным абсолютно непрерывным распределением определение независимости можно сформулировать таким образом:

Определение 47. Случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ с совместным абсолютно непрерывным распределением независимы, если плотность совместного распределения равна произведению маргинальных плотностей этих случайных величин, то есть для любых $x_1, x_2, \ldots, x_n$, имеет место равенство: $f_{\xi_1,\xi_2,\ldots,\xi_n}(x_1, x_2, \ldots, x_n) = f_{\xi_1}(x_1) \cdot f_{\xi_2}(x_2) \cdot \ldots \cdot f_{\xi_n}(x_n)$.

Для того, чтобы последнее определение имело право на существование, докажем эквивалентность определений 45 и 47. По теореме 20, если совместное распределение с.в. $\xi_1, \xi_2, \ldots, \xi_n$ абсолютно непрерывно, то и в отдельности каждая с.в. $\xi_1, \xi_2, \ldots, \xi_n$ также имеют абсолютно непрерывное распределение. Пусть случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ независимы в смысле определения 45, то есть для любых $x_1, x_2, \ldots, x_n$

$$\int\limits_{-\infty}^{x_1}\int\limits_{-\infty}^{x_2}\cdots\int\limits_{-\infty}^{x_n} f_{\xi_1,\xi_2,\ldots,\xi_n}(u_1, u_2, \ldots, u_n) du_1 du_2 \ldots du_n =$$

$$= F_{\xi_1,\xi_2,\dots,\xi_n}(x_1, x_2, \dots, x_n) =$$

$$= \left[\text{по определению 45}\right] = F_{\xi_1}(x_1) \cdot F_{\xi_2}(x_2) \cdot \ldots \cdot F_{\xi_n}(x_n) =$$

$$= \int\limits_{-\infty}^{x_1} f_{\xi_1}(u_1)du_1 \int\limits_{-\infty}^{x_2} f_{\xi_2}(u_2)du_2 \cdots \int\limits_{-\infty}^{x_n} f_{\xi_n}(u_n)du_n =$$

$$= \int\limits_{-\infty}^{x_1} \cdots \int\limits_{-\infty}^{x_n} f_{\xi_1}(u_1)\dots f_{\xi_n}(u_n)du_1\dots du_n.$$

Равенство двух крайних интегралов в левой и правой частях последней цепочки равенств при всех значениях $x_1, x_2, \dots, x_n$ влечёт равенство подынтегральных выражений, то есть независимость в смысле определения 47. Доказать в обратную сторону можно используя те же равенства, но в другом порядке.

Глава 9. Функции от случайных величин

§1. Преобразование одной случайной величины

В этом параграфе будем рассматривать преобразования случайных величин с абсолютно непрерывными распределениями. Пусть с. в. ξ имеет функцию распределения $F_\xi(x)$ и плотность распределения $f_\xi(x)$. Построим с помощью функции $g : R^1 \to R^1$ случайную величину $\zeta = g(\xi)$. Требуется найти функцию распределения и, если существует, плотность распределения с.в. ζ.

Заметим,что плотность распределения с.в. $\zeta = g(\xi)$ существует далеко не при любых функциях g. Так, если функция $g(x)$ кусочно-постоянна, то с. в. ζ имеет дискретное распределение, и плотность её распределения не существует вовсе, даже, если $g(x)$ – борелевская функция.

Плотность распределения $g(\xi)$ заведомо существует, если, например, функция $g(x)$ строго монотонна.

Теорема 21. Пусть абсолютно непрерывная с.в. ξ имеет плотность распределения $f_\xi(x)$. Тогда с.в. $\zeta = a\xi + b,\ a \neq 0$ имеет плотность распределения $f_\zeta(x) = \frac{1}{|a|}\, f_\xi\left(\frac{x-b}{a}\right)$.

Доказательство. *i).* Пусть $a > 0$. Тогда, очевидно, имеем:

$$F_\zeta(x) = P(\zeta < x) = P(a\xi + b < x) =$$

$$= P\left(\xi < \frac{x-b}{a}\right) = F_\xi\left(\frac{x-b}{a}\right) = \int\limits_{-\infty}^{\frac{x-b}{a}} f_\xi(u)du =$$

$$= \left[\text{замена: } v = au + b,\ du = \frac{1}{a}dv\right] = \int\limits_{-\infty}^{x} \frac{1}{a}\, f_\xi\left(\frac{v-b}{a}\right)\, dv,$$

т.е. с.в. $\zeta = a\xi + b$ имеет плотность $f_\zeta(x) = \frac{1}{a}\, f_\xi\left(\frac{x-b}{a}\right)$.

ii). Пусть теперь $a < 0$. Тогда получаем:

$$F_\zeta(x) = P(\zeta < x) = P(a\xi + b < x) = P\left(\xi > \frac{x-b}{a}\right) =$$

$$= \int\limits_{\frac{x-b}{a}}^{+\infty} f_\xi(u)du = \left[\text{замена: } v = au + b,\ du = \frac{1}{a}dv\right] =$$

$$= \int\limits_{x}^{-\infty} \frac{1}{a}\, f_\xi\left(\frac{v-b}{a}\right)\, dv = -\int\limits_{-\infty}^{x} \frac{1}{a}\, f_\xi\left(\frac{v-b}{a}\right)\, dv,$$

т.е. с.в. $\zeta = a\xi + b$ имеет плотность $f_\zeta(x) = -\frac{1}{a}\, f_\xi\left(\frac{x-b}{a}\right)$. Таким образом, объединяя случаи *i)* и *ii)* окончательно получаем, что в случае $a \neq 0$ с.в. $\zeta = a\xi + b$ имеет плотность $f_\zeta(x) = \frac{1}{|a|}\, f_\xi\left(\frac{x-b}{a}\right)$, что и требовалось доказать.

Последнюю теорему можно обобщить на более широкий класс функций. Именно, справедлива следующая

Теорема 22. Пусть абсолютно непрерывная с.в. ξ имеет плотность распределения $f_\xi(x)$. Тогда для строго монотонной функции $g(x)$ с.в. $\zeta = g(\xi)$, имеет плотность распределения

$$f_\zeta(x) = \frac{1}{\left|g'\left(g^{-1}(x)\right)\right|}\, f_\xi\left(g^{-1}(x)\right),$$

где g^{-1} обозначает обратную функцию к g.

Доказательство почти дословно повторяет доказательство предыдущей теоремы. Рассматриваются отдельно случаи строго монотонно возрастающих и строго монотонно убывающих функций.

Приведём несколько следствий теоремы 21.

Следствие 4. Если с.в. $\xi \sim N(0,1)$, то с.в. $\zeta = a + \sigma\xi \sim N(a, \sigma^2)$.

Раньше уже возникало это свойство. Непосредственно из теоремы 21 имеем,
$f_\zeta(x) = \frac{1}{\sigma} f_\xi\left(\frac{x-a}{\sigma}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$.

Следствие 5. Если с.в. $\zeta \sim N(a, \sigma^2)$, то с.в. $\xi = \frac{\zeta - a}{\sigma} \sim N(0,1)$.

Следствие 6. Если с.в. $\xi \sim E(\alpha)$, то с.в. $\zeta = \alpha\xi \sim E(1)$.

§2. Функции от двух случайных величин

Пусть ξ, ζ — случайные величины, заданные на одном и том же вероятностном пространстве с совместной плотностью распределения $f_{\xi,\zeta}(x, y)$. Предположим, что задана функция $g : R^2 \to R^1$. Требуется найти функцию распределения, а если существует, то и плотность распределения случайной величины $\eta = g(\xi, \zeta)$.

Теорема 23. Пусть $x \in R^1$ и область $D_x = \{(u, v) \in R^2 : g(u, v) < x\} \subseteq R^2$. Тогда с.в. $\eta = g(\xi, \zeta)$ имеет функцию распределения

$$F_\eta(x) = P\left((\xi, \zeta) \in D_x\right) = \iint\limits_{D_x} f_{\xi,\zeta}(u, v)dudv$$

Доказательство этой теоремы вытекает из определения абсолютно непрерывной случайной величины и того факта, что вероятность попасть случайному вектору в заданную область можно найти как объём трёхмерной области, расположенной под поверхностью, которая задаётся функцией совместной плотности распределения над этой областью.

Применение этой теоремы для суммы независимых случайных величин запишем в виде отдельной теоремы.

Теорема 23 (формула свёртки). Если с. в. ξ и ζ независимы и имеют абсолютно непрерывное распределение с плотностями $f_\xi(u)$ и $f_\zeta(v)$, то плотность распределения суммы $\xi + \zeta$ равна «свёртке» плотностей $f_\xi(u)$ и $f_\zeta(v)$:

$$f_{\xi+\zeta}(x) = \int\limits_{-\infty}^{+\infty} f_\xi(u)\ f_\zeta(x-u)\ du = \int\limits_{-\infty}^{+\infty} f_\xi(x-v)\ f_\zeta(v)\ dv.$$

Доказательство. Для получения формулы свёртки из теоремы 22 для независимых с.в. ξ и ζ, получаем

$$F_{\xi+\zeta}(x) = P\left(\xi + \zeta < x\right) = \iint\limits_{\substack{(u,v)\in R^2 :\\ u+v<x}} f_\xi(u)f_\zeta(v)dudv =$$

$$= \int\limits_{-\infty}^{+\infty} f_\xi(u) \int\limits_{-\infty}^{x-u} f_\zeta\ (v)dv\ du =$$

$$= \left[\frac{\text{замена во внутреннем интеграле}}{v=w-u,\ dv=dw} \right] = \int\limits_{-\infty}^{+\infty} f_\xi(u) \int\limits_{-\infty}^{x} f_\zeta\ (w-u)dw\ du =$$

$= \int\limits_{-\infty}^{x} \left(\int\limits_{-\infty}^{+\infty} f_\xi(u)\ f_\zeta\ (w-u)\ du \right) dw = \int\limits_{-\infty}^{x} f_{\xi+\zeta}(w)dw,$ откуда следует утверждение теоремы. Второе равенство утверждения теоремы, очевидно будет иметь место, если в доказательстве формально поменять местами плотности для ξ и ζ, поскольку сумма $\xi+\zeta$ является коммутативной операцией.

Теорема 24. Пусть с. в. ξ и ζ независимы и имеют абсолютно непрерывное распределение с плотностями $f_\xi(u)$ и $f_\zeta(v)$ соответственно. Тогда

$$f_{\xi/\zeta}(x) = \int\limits_{-\infty}^{+\infty} f_\xi(xv) f_\zeta(v)|v|dv.$$

Доказательство. Найдём сначала функцию распределения для случайной величины ξ/ζ. Из теоремы 22, получаем

$$F_{\xi/\zeta}(x) = P\left(\frac{\xi}{\zeta} < x\right) = \iint\limits_{\substack{(u,v)\in R^2\,: \\ \frac{u}{v}<x}} f_\xi(u) f_\zeta(v) du dv =$$

$$= \iint\limits_{v<0,\ u>xv} f_\xi(u) f_\zeta(v) du dv + \iint\limits_{v>0,\ u<xv} f_\xi(u) f_\zeta(v) du dv =$$

$$= \int\limits_{-\infty}^{0} f_\zeta(v) \left(\int\limits_{xv}^{+\infty} f_\xi(u) du \right) dv + \int\limits_{0}^{+\infty} f_\zeta(v) \left(\int\limits_{-\infty}^{xv} f_\xi(u) du \right) dv$$

Из курса математического анализа известно, что

$$\left(\int\limits_{\varphi_1(x)}^{\varphi_2(x)} g(u)\ du \right)'_x = g(\varphi_2(x))\varphi_2'(x) - g(\varphi_1(x))\varphi_1'(x),$$

откуда, в частности, следует

$$\left(\int\limits_{-\infty}^{\varphi(x)} g(u)\ du \right)'_x = g(\varphi(x))\ \varphi'(x) \quad \text{и}$$

$$\left(\int\limits_{\varphi(x)}^{+\infty} g(u)\,du\right)'_x = -g(\varphi(x))\,\varphi'(x).$$

Поэтому, пользуясь приведёнными формулами и из свойства (vi) для плотностей, получаем:

$$f_{\xi/\zeta}(x) = \int\limits_{-\infty}^{0} f_\zeta(v)\left(\int\limits_{xv}^{+\infty} f_\xi(u)du\right)'_x dv +$$

$$+\int\limits_{0}^{+\infty} f_\zeta(v)\left(\int\limits_{-\infty}^{xv} f_\xi(u)du\right)'_x dv = -\int\limits_{-\infty}^{0} f_\zeta(v)f_\xi(xv)v\,dv +$$

$$+\int\limits_{0}^{+\infty} f_\zeta(v)f_\xi(xv)v\,dv = \int\limits_{-\infty}^{+\infty} f_\xi(xv)f_\zeta(v)|v|\,dv.$$

Совершенно аналогично доказывается следующая

Теорема 25. Если с. в. ξ и ζ независимы и имеют абсолютно непрерывное распределение с плотностями $f_\xi(u)$ и $f_\zeta(v)$ соответственно, то

$$f_{\xi\cdot\zeta}(x) = \int\limits_{-\infty}^{+\infty} f_\xi(x/v)\; f_\zeta(v)\;\frac{dv}{|v|}.$$

Пример 25. Пусть независимые случайные величины ξ и ζ имеют стандартное нормальное распределение. Покажем, что их сумма имеет тоже нормальное распределение, но с параметрами 0 и 2.

По формуле свёртки, имеем

$$f_{\xi+\zeta}(x) = \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty} e^{-\frac{u^2}{2}}\; e^{-\frac{(x-u)^2}{2}}\,du =$$

$$= \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty} \exp\left\{-\left(u^2 - xu + \frac{x^2}{2}\right)\right\}\,du$$

Выделим полный квадрат в круглых скобках в последнем выражении:
$u^2 - xu + \frac{x^2}{2} = \left(u - \frac{x}{2}\right)^2 + \frac{x^2}{4}$. Тогда получаем

$$f_{\xi+\zeta}(x) = \frac{1}{2\pi}e^{-\frac{x^2}{4}}\int\limits_{-\infty}^{+\infty} e^{-\left(u-\frac{x}{2}\right)^2}\,du =$$

$$= \left|v = u - x/2,\ dv = du\right| = \frac{1}{2\pi}e^{-\frac{x^2}{4}}\int\limits_{-\infty}^{+\infty} e^{-v^2}\,dv = \frac{1}{2\sqrt{\pi}}\,\frac{1}{\sqrt{\pi}}\,e^{-\frac{x^2}{4}}\int\limits_{-\infty}^{+\infty} e^{-v^2}\,dv =$$

$= \frac{1}{2\sqrt{\pi}}e^{-\frac{x^2}{4}}$, поскольку $\frac{1}{\sqrt{\pi}}\int\limits_{-\infty}^{+\infty} e^{-v^2}\,dv = 1$ (интеграл от плотности нормального распределения с параметрами 0,$1/\sqrt{2}$). Таким образом, окончательно получено нормальное распределение с параметрами 0 и 2.

Пример 26. Пусть независимые случайные величины ξ и ζ имеют стандартное нормальное распределение. Покажем, что их частное имеет стандартное распределение Коши, т.е. $\xi/\zeta \sim \mathrm{C}(0,1)$.

Воспользуемся формулой теоремы 24:

$$f_{\xi/\zeta}(x) = \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty} e^{-\frac{x^2v^2}{2}}\ e^{-\frac{v^2}{2}}|v|dv = \frac{1}{\pi}\int\limits_{0}^{+\infty} e^{-\frac{v^2(1+x^2)}{2}}\ v\,dv =$$

$$= \frac{1}{\pi(1+x^2)}\int\limits_{0}^{+\infty} e^{-\frac{v^2(1+x^2)}{2}}\ d\left(\frac{v^2(1+x^2)}{2}\right) = \left|w = \frac{v^2(1+x^2)}{2} \geq 0\right| =$$

$$= \frac{1}{\pi(1+x^2)}\int\limits_{0}^{+\infty} e^{-w}\,dw = \frac{1}{\pi(1+x^2)}.$$

§3. Классы распределений, воспроизводимых по параметру

Определение 48. Если свёртка двух независимых случайных величин из одного и того же класса распределений принадлежит тому же классу, а параметры складываются, то такой класс распределений называется *воспроизводимым по параметру.*

j. Проверим, что класс распределений Пуассона воспроизводим по параметру. Пусть независимые случайные величины $\xi_1 \sim \Pi(\lambda_1)$ и $\xi_2 \sim \Pi(\lambda_2)$. Тогда

$$P(\xi_1 + \xi_2 = n) = (\text{ формула полной вероятности }) =$$

$$= \sum_{i=0}^{\infty} P(\xi_1 = i)P(\xi_1 + \xi_2 = n \mid \xi_1 = i) =$$

$$= (\text{ из независимости } \xi_1 \text{ и } \xi_2\ ) = \sum_{i=0}^{n} P(\xi_1 = i)P(\xi_2 = n - i) =$$

$$= \sum_{i=0}^{n}\ e^{-\lambda_1}\frac{\lambda_1^i}{i!}\ e^{-\lambda_2}\frac{\lambda_2^{n-i}}{(n-i)!} = \frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{i=0}^{n}\ \frac{n!}{i!(n-i)!}\lambda_1^i\ \lambda_2^{n-i}$$

$$= (\text{ бином Ньютона }) = e^{-(\lambda_1+\lambda_2)}\,\frac{(\lambda_1+\lambda_2)^n}{n!},$$

т.е. $\xi_1 + \xi_2 \sim \Pi(\lambda_1 + \lambda_2)$, что и требовалось показать.

jj. Покажем теперь, что класс биномиальных распределений с фиксированной вероятностью успеха воспроизводим по первому параметру. Итак, пусть независимые случайные величины $\xi \sim Bi(k,\ p)$ и $\zeta \sim Bi(n,\ p)$. Надо показать, что $\xi+\zeta \sim Bi(k+n,\ p)$. Пусть $m = 0, 1, 2, \ldots, k+n$, тогда по формуле полной вероятности получаем:
$P(\xi+\zeta = m) = \sum\limits_{i=0}^{k} P(\xi = i)P(\xi+\zeta = m \mid \xi = i)$ Из независимости случайных величин, следует, что $P(\xi+\zeta = m \mid \xi = i) = P(\zeta = m-i)$. Для того, чтобы произведение вероятностей $P(\xi = i)\cdot P(\zeta = m-i) \neq 0$ требуется, чтобы неравенства $0 \le i \le k$ и $0 \le m-i \le n$ выполнялись одновременно. Второе из этих неравенств, очевидно, эквивалентно неравенству $m-n \le i \le m$. Поэтому, чтобы указанные неравенства выполнялись одновременно, требуется, чтобы $\max(0,\ m-n) \le i \le \min(k,\ m)$. Поэтому, продолжая начатые выше равенства, получаем

$$P(\xi+\zeta = m) = \sum_{i=\max(0,\ m-n)}^{\min(k,m)} P(\xi = i)P(\zeta = m-i) =$$

$$= \sum_{i=\max(0,m-n)}^{\min(k,m)} C_k^i p^i q^{k-i} C_n^{m-i} p^{m-i} q^{n-m+i} =$$

$$= C_{k+n}^m p^m q^{k+n-m} \sum_{i=\max(0,m-n)}^{\min(k,m)} \frac{C_k^i\ C_n^{m-i}}{C_{k+n}^m} = C_{k+n}^m p^m q^{k+n-m},$$

где в последнем равенстве было использовано свойство гипергеометрического распределения, т.е. $\sum\limits_{i=\max(0,m-n)}^{\min(k,m)} \frac{C_k^i\ C_n^{m-i}}{C_{k+n}^m} = 1$. Таким образом, показано что $\xi+\zeta \sim Bi(k+n,\ p)$, т.е. то, что и требовалось.

На самом деле, последнее свойство для биномиальных распределений можно показать проще, исходя из определения биномиального закона. Вспомним, что биномиальная случайная величина равна числу успехов в заданной серии независимых испытаний Бернулли с вероятностью успеха p при каждом испытании. Поэтому, если ξ есть число успехов в первой серии из k испытаний, а ζ - число успехов во второй независимой серии из n испытаний, то сумма с.в. $\xi+\zeta$ есть общее число успехов в объединённой серии из $k+n$ независимых испытаний Бернулли, т.е. $\xi+\zeta \sim Bi(k+n,\ p)$.

jjj. Пусть независимые случайные величины $\xi_1 \sim N(a_1,\ \sigma_1^2)$ и $\xi_2 \sim N(a_2,\ \sigma_2^2)$. Тогда $\xi_1+\xi_2 \sim N(a_1+a_2,\ \sigma_1^2+\sigma_2^2)$, т.е. класс нормальных распределений является воспроизводимым по обоим параметрам. Этот факт будет доказан позже.

jv. Наконец, покажем, что класс гамма распределений при фиксированном первом параметре является воспроизводимым по второму параметру, т.е. если случайные величины $\xi_i \sim \Gamma(\alpha, \lambda_i), \;\; i = 1, 2$ независимы, то $\xi_1 + \xi_2 \sim \Gamma(\alpha, \lambda_1 + \lambda_2)$.

Предварительно, потребуются некоторые сведения из математического анализа.

Бета-функцией ($\mathbf{B}$-функцией, бета-функцией Эйлера или интегралом Эйлера I рода) называется следующая специальная функция от двух переменных:

$$\mathrm{B}(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}\, dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)},$$

определённая при $x > 0, \; y > 0$.

Исходя из свойств интеграла, несложно понять, что

$$\int_{-\infty}^{+\infty} g(u)\, \mathbf{1}_{[a,+\infty)}(u)\, \mathbf{1}_{[0,+\infty)}(b-u)\, du = \int_a^b g(u)\, du.$$

Итак, воспользовавшись формулой свёртки и упомянутыми равенствами, получаем,

$$f_{\xi_1+\xi_2}(x) = \int_{-\infty}^{+\infty} f_{\xi_1}(u) f_{\xi_2}(x-u) du =$$

$$= \int_{-\infty}^{+\infty} \frac{u^{\lambda_1-1}e^{-\frac{u}{\alpha}}}{\alpha^{\lambda_1}\Gamma(\lambda_1)}\, \mathbf{1}_{[0,+\infty)}(u)\, \frac{(x-u)^{\lambda_2-1}e^{-\frac{x-u}{\alpha}}}{\alpha^{\lambda_2}\Gamma(\lambda_2)}\, \mathbf{1}_{[0,+\infty)}(x-u)\, du =$$

$$= \frac{e^{-\frac{x}{\alpha}}}{\alpha^{\lambda_1+\lambda_2}\Gamma(\lambda_1)\Gamma(\lambda_2)} \int_0^x u^{\lambda_1-1}(x-u)^{\lambda_2-1} du =$$

$$= \left[\text{замена: } v = u/x, \; du = x dv \right] =$$

$$= \frac{x^{\lambda_1+\lambda_2-1}\, e^{-\frac{x}{\alpha}}}{\alpha^{\lambda_1+\lambda_2}\Gamma(\lambda_1)\Gamma(\lambda_2)} \int_0^1 v^{\lambda_1-1}(1-v)^{\lambda_2-1} dv =$$

$$= \frac{x^{\lambda_1+\lambda_2-1}\, e^{-\frac{x}{\alpha}}}{\alpha^{\lambda_1+\lambda_2}\Gamma(\lambda_1)\Gamma(\lambda_2)}\, \mathrm{B}(x, y) = \frac{x^{\lambda_1+\lambda_2-1}\, e^{-\frac{x}{\alpha}}}{\alpha^{\lambda_1+\lambda_2}\Gamma(\lambda_1+\lambda_2)},$$

т.е. $\xi_1 + \xi_2 \sim \Gamma(\alpha, \lambda_1 + \lambda_2)$, что и требовалось показать.

Глава 10. Числовые характеристики случайных величин

§1.Математическое ожидание случайной величины

Для дальнейшего рассмотрения, вообще говоря, потребуется понятие интеграла Лебега–Стилтьеса. Подробно на этом понятии не станем останавливаться. Отметим только, что интеграл Лебега–Стилтьеса - обычный интеграл Лебега относительно меры, известной как мера Лебега-Стилтьеса, которая может быть связана с любой функцией распределения на вещественной прямой. В свою очередь, интеграл Лебега — это обобщение интеграла Римана на более широкий класс функций. Все функции, определённые на конечном отрезке числовой прямой и интегрируемые по Риману, являются также интегрируемыми по Лебегу, причём в этом случае оба интеграла равны. Однако существует большой класс функций, определённых на отрезке и интегрируемых по Лебегу, но не интегрируемых по Риману. Такие функции здесь рассматриваться не будут и, поэтому все интегралы, которые будут встречаться понимаем исключительно как обычные интегралы Римана.

Определение 49. Математическим ожиданием $\boldsymbol{E\xi}$ случайной величины $\boldsymbol{\xi}$ с заданной функцией распределения $\boldsymbol{F_\xi(x)}$ называют интеграл Лебега-Стилтьеса

$$\boldsymbol{E\xi = \int\limits_{-\infty}^{+\infty} x\, dF_\xi(x)}\,,$$

который в случае *(j) дискретно распределённой величины* $\boldsymbol{\xi}$: $\boldsymbol{P(\xi = a_i) = p_i, i = 1, 2, \dots}$, превращается в равенство:

$$\boldsymbol{E\xi = \sum_{i=1}^{\infty} a_i\, P(\xi = a_i) = \sum_{i=1}^{\infty} a_i\, p_i,} \quad \text{если этот ряд } \textit{абсолютно} \text{ сходится,}$$

а в случае *(jj) абсолютно непрерывного распределения* превращается в равенство:

$$\boldsymbol{E\xi = \int\limits_{-\infty}^{+\infty} x\, f_\xi(x)\, dx,} \quad \text{если этот интеграл } \textit{абсолютно} \text{ сходится,}$$

где $\boldsymbol{f_\xi(x)}$ плотность с.в. $\boldsymbol{\xi}$.

Заметим, что, в случае дискретного распределения, если $\boldsymbol{\sum\limits_{i=1}^{\infty} |a_i|\, p_i = \infty}$, то говорят, что математическое ожидание не существует. Аналогично, в случае абсолютно непрерывного распределения, если $\boldsymbol{\int\limits_{-\infty}^{+\infty} |x|\, f_\xi(x)\, dx = \infty}$, то говорят, что математическое ожидание не существует.

Иногда математическое ожидание называют средним значением или первым моментом.

Пример 27. Пусть случайная величина ξ равна числу очков, выпадающих при одном подбрасывании игральной кости. Тогда $E\xi = \sum\limits_{i=1}^{6} i\,\frac{1}{6} = 3.5$, т.е. в среднем при одном подбрасывании кости выпадает 3.5 очка.

Пример 28. Пусть случайная величина $\xi \sim R(a,b),\ \ a < b$ имеет равномерное распределение на отрезке $[a,b]$. Плотность распределения с.в. ξ есть функция $f_\xi(x) = \frac{1_{[a,b]}(x)}{b-a}$. Тогда

$\mathsf{E}\,\xi = \int\limits_a^b x \cdot \frac{1}{b-a}\,dx = \frac{x^2}{2(b-a)}\Big|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$ есть середина отрезка.

§2. Свойства математического ожидания

Во всех свойствах предполагается, что рассматриваемые математические ожидания существуют. Никогда нельзя забывать, что математическое ожидание случайной величины есть число.

Если выбирать какую-то функцию g, чтобы для произвольной случайной величины $\xi(\omega)$ на некотором вероятностном пространстве суперпозиция $g(\xi(\omega))$ тоже была бы случайной величиной на том же вероятностном пространстве, то эта функция должна быть *борелевской.*

Определение 50. Функция $g : R^1 \to R^1$ называется *борелевской*, если прообраз любого борелевского множества является борелевским. Формально это записывается так: $\forall\, B \in \mathfrak{B} \quad g^{-1}(B) \in \mathfrak{B}$ где $\mathfrak{B}$ - борелевская σ-алгебра.

Не будем подробно углубляться в эту тему. Отметим только, что класс борелевских функций широк. Далее все функции для суперпозиций считаем *борелевскими.*

(i). Для произвольной борелевской функции $g : R^1 \to R^1$

$$Eg(\xi) = \int\limits_{-\infty}^{+\infty} g(x)\,dF_\xi(x)\,, \qquad \text{для произвольных с.в. и}$$

$$Eg(\xi) = \sum_{i=1}^{\infty} g(a_i)\,P(\xi = a_i), \qquad \text{если с.в. дискретная и}$$

$$Eg(\xi) = \int\limits_{-\infty}^{+\infty} g(x)\,f_\xi(x)\,dx, \qquad \text{если с.в. абсолютно непрерывная.}$$

Доказательство. Пусть ξ -дискретная с.в. и пусть $g(\xi)$ принимает значения

$c_1,\ c_2, \ldots$ с вероятностями $P(g(\xi) = c_j) = \sum\limits_{k:\ g(a_k)=c_j} P(\xi = a_k)$. Тогда

$$Eg(\xi) = \sum_j c_j\, P(g(\xi) = c_j) = \sum_j c_j \sum_{k:\ g(a_k)=c_j} P(\xi = a_k) =$$

$$= \sum_j \sum_{k:\ g(a_k)=c_j} g(a_k)\, P(\xi = a_k) = \sum_k g(a_k)\, P(\xi = a_k).$$

Остальные случаи оставляем без доказательства.

Следствие 7. Математическое ожидание ξ существует тогда и только тогда, когда $\mathsf{E}\,|\xi| < \infty$.

Доказательство. Условием существование математического ожидания является абсолютная сходимость ряда или интеграла. Это в точности есть условие $\mathsf{E}\, g(\xi) < \infty$ при $g(x) = |x|$.

(ii). Математическое ожидание вырожденной случайной величины равно самой случайной величине.
Или то же самое другими словами:
Математическое ожидание постоянной равно этой постоянной: $E\,c = c$.

Это есть простое следствие свойства *(i)*.

(iii). Константу можно вынести за знак м.о.: $Ec\xi = c\,E\xi$.
Это свойство сразу следует из *(i)* для функции $g(x) = cx$.

(iv). Математическое ожидание суммы с.в. ξ и ζ равно сумме их математических ожиданий: $E(\xi + \zeta) = E\xi + E\zeta$.

Доказательство проведём для с.в. с дискретными распределениями: пусть x_i и y_j — значения с.в. ξ и ζ, соответственно. Для борелевской функции $g : R^2 \to R^1$ можно доказать свойство, аналогичное *(i)*. Воспользуемся этим свойством для функции $g(x,\, y) = x + y$:

$$E(\xi + \zeta) = \sum_i \sum_j (x_i + y_j) P(\xi = x_i, \zeta = y_j) =$$

$$= \sum_i\ x_i\ \sum_j P(\xi = x_i, \zeta = y_j) + \sum_j\ y_j\ \sum_i P(\xi = x_i, \zeta = y_j) =$$

$$= \sum_i\ x_i\ P(\xi = x_i) + \sum_j\ y_j\ P(\zeta = y_j) = E\xi + E\zeta.$$

Определение 51. Если для события $A \in \mathcal{A}$ на некотором вероятностном пространстве $(\Omega, \mathcal{A}, P)$ выполняется равенство $P(A) = 1$, то говорят, что событие A произошло *почти наверное* и кратко пишут A п.н.

(v). j) Если $\xi \geq 0$ п.н., то $E\xi \geq 0$; *jj)* Если $\xi \geq 0$ п.н., и при этом $E\xi = 0$, то $\xi = 0$ п.н.

Из этого свойства, в частности, следует:

j) Если $\xi \geq \zeta$ п.н., то $E\xi \geq E\zeta$; *jj)* Если $\xi \geq \zeta$ п.н., и при этом $E\xi = E\zeta$, то $\xi = \eta$ п.н.

(vi). Математическое ожидание произведения *независимых* случайных величин равно произведению их математических ожиданий, т.е. если ξ и ζ независимы, то $\mathsf{E}\xi\,\zeta = \mathsf{E}\xi\,\mathsf{E}\zeta$.

Покажем справедливость этого свойства для дискретных с.в.

$$\mathsf{E}\xi\zeta = \sum_i \sum_j x_i\, y_j \mathsf{P}(\xi = x_i,\, \zeta = y_j) =$$

$$= \sum_i x_i\, \mathsf{P}(\xi = x_i) \sum_j y_j\, \mathsf{P}(\zeta = y_j) = \mathsf{E}\xi\,\mathsf{E}\zeta,$$

поскольку, для независимых с.в. ξ и ζ, имеем: $\mathsf{P}(\xi = x_i,\, \zeta = y_j) = \mathsf{P}(\xi = x_i)\,\mathsf{P}(\zeta = y_j)$.

Обратное утверждение, вообще говоря, неверно, т.е. из равенства $\mathsf{E}\xi\,\zeta = \mathsf{E}\xi\ \mathsf{E}\zeta$ не следует независимость с.в. ξ и ζ.

Пример 29. Пусть с.в.$\varphi \sim R[0, 2\pi]$. Несложно проверить, что с.в. $\xi = \sin(\varphi)$ и $\zeta = \cos(\varphi)$ являются зависимыми. Например:

$$\mathsf{P}\Big(\xi > \frac{1}{\sqrt{2}},\, \eta > \frac{1}{\sqrt{2}}\Big) = 0 \neq \mathsf{P}\Big(\xi > \frac{1}{\sqrt{2}}\Big)\,\mathsf{P}\Big(\eta > \frac{1}{\sqrt{2}}\Big) > 0.$$

Исходя из свойства *(i)*, найдём:

$$\mathsf{E}\xi = \frac{1}{2\pi}\int\limits_0^{2\pi} \sin(u)du = 0, \quad \mathsf{E}\zeta = \frac{1}{2\pi}\int\limits_0^{2\pi} \cos(u)du = 0, \quad \mathsf{E}\xi\zeta =$$

$$= \frac{1}{2\pi}\int\limits_0^{2\pi} \sin(u)\cos(u)du = 0,$$

т.е. $\mathsf{E}\xi\,\zeta = \mathsf{E}\xi\ E\zeta$, хотя ξ и ζ зависимые с.в.

Но если дополнительно потребовать, чтобы с.в. ξ и ζ имели бы нормальные распределения, то из равенства $\mathsf{E}\xi\,\zeta = \mathsf{E}\xi\ \mathsf{E}\zeta$ следует независимость с.в. ξ и ζ.

§3. Моменты случайных величин

Начнём этот параграф с важного определения.

Определение 52. Если $\mathsf{E}|\xi|^k < \infty$, то число

$\mathsf{E}\xi^k$ называется *моментом порядка* k (или короче, k-м моментом) случайной величины ξ;

$\mathsf{E}|\xi|^k$ называется *абсолютным моментом порядка* k (абсолютным k-м моментом) случайной величины ξ;

$\mathsf{E}(\xi - \mathsf{E}\xi)^k$ называется *центральным моментом порядка* k (центральным k-м моментом) случайной величины ξ;

$\mathsf{E}|\xi - \mathsf{E}\xi|^k$ называется *абсолютным центральным моментом порядка* k (абсолютным центральным k-м моментом) случайной величины ξ.

Число $\mathsf{D}\xi = \mathsf{E}(\xi - \mathsf{E}\xi)^2$ (центральный момент второго порядка) называется *дисперсией* случайной величины ξ.

Дисперсия задаёт среднее квадратичное отклонение от математического ожидания, т.е. задаёт некоторую величину отклонения от среднего значения.

Определение 53. Величина $\sigma = \sqrt{\mathsf{D}\xi}$ называется *средним квадратическим отклонением* с.в. ξ.

Чтобы прояснить связь моментов разных порядков, рассмотрим несколько неравенств.

Теорема 26. Если существует момент порядка $t > 0$ с.в. ξ, то существуют и её моменты порядка s при $0 < s < t$.

Доказательство. Для любого числа x верно неравенство $|x|^s \leq \max\{\,|x|^t,\ 1\} \leq |x|^t + 1$. Действительно, $|x|^s \leq |x|^t$ при $|x| > 1$, и $|x|^s \leq 1$ при $|x| \leq 1$. Из этого неравенства следует, что $|\xi(\omega)|^s \leq |\xi(\omega)|^t + 1$ для всех $\omega \in \Omega$. Но исходя из свойства (v) математического ожидания, получаем: $\mathsf{E}\,|\xi|^s \leq \mathsf{E}\,|\xi|^t + 1$. Момент порядка t существует по условию теоремы, т.е. $\mathsf{E}\,|\xi|^t < \infty$. Поэтому и $\mathsf{E}\,|\xi|^s < \infty$.

Теорема 27 (неравенство Йенсена). Пусть вещественнозначная функция g выпукла вниз. Тогда для любой случайной величины ξ с конечным математическим ожиданием верно неравенство: $\mathsf{E}\,g(\xi) \geq g(\mathsf{E}\,\xi)$. Для выпуклых вверх функций знак неравенства меняется на противоположный.

Замечание. Напомним, что дифференцируемую на интервале (a, b) функцию называют выпуклой вниз (вверх), если её график на этом интервале всюду лежит выше (ниже) любой касательной прямой, проведённой к этому графику.

Доказательство. Понадобится следующее свойство: если функция g выпукла, то для всякого x_0 найдется число $c(x_0)$ такое, что при всех x $g(x) \geq g(x_0) + c(x_0)(x - x_0)$. Это свойство очевидно и означает, что график выпуклой функции лежит полностью выше любой из касательных к этому графику. Возьмём в этом свойстве $x_0 = \mathsf{E}\,\xi,\ x = \xi$. Тогда $g(\xi) \geq g(\mathsf{E}\,\xi) + c(\mathsf{E}\,\xi)(\xi - \mathsf{E}\,\xi)$. Вычислим математическое ожидание обеих частей последнего неравенства. Так как $\mathsf{E}\,(\xi - \mathsf{E}\,\xi) = 0$, и неравенство между математическими ожиданиями сохраняется, то $\mathsf{E}\,g(\xi) \geq g(\mathsf{E}\,\xi)$.

Следующее неравенство связывает моменты разных порядков.

Теорема 28 (неравенство Ляпунова). Если $\mathsf{E}\,|\xi|^t < \infty$, то для любого $0 < s < t$

$$\sqrt[s]{\mathsf{E}\,|\xi|^s} \le \sqrt[t]{\mathsf{E}\,|\xi|^t}.$$

Доказательство. Поскольку $0 < s < t$, то $g(x) = |x|^{t/s}$ - выпуклая функция. По неравенству Йенсена для $\eta = |\xi|^s$,

$$(\mathsf{E}\,|\xi|^s)^{t/s} = (\mathsf{E}\,\eta)^{t/s} = g(\mathsf{E}\,\eta) \le \mathsf{E}\,g(\eta) =$$

$$= \mathsf{E}\,|\eta|^{t/s} = \mathsf{E}\,|\xi|^{s\cdot t/s} = \mathsf{E}\,|\xi|^t.$$

Осталось извлечь из обеих частей корень степени t.

Из неравенства Йенсена вытекают, например, неравенства:

$$\mathsf{E}\,e^{\xi} \ge e^{\mathsf{E}\,\xi}, \qquad \mathsf{E}\,\xi^2 \ge (\mathsf{E}\,\xi)^2, \qquad \mathsf{E}\,|\xi| \ge |\mathsf{E}\,\xi|,$$

$$\mathsf{E}\,\ln\xi \le \ln(\mathsf{E}\,\xi), \qquad \mathsf{E}\,\frac{1}{\xi} \ge \frac{1}{\mathsf{E}\,\xi}, \qquad \mathsf{E}\,\sqrt{\xi} \le \sqrt{\mathsf{E}\,\xi}.$$

Последние три неравенства верны для положительных ξ.

§4. Свойства дисперсии

Свойства дисперсии вытекают из соответствующих свойств математического ожидания. Из существования второго момента следует существование математического ожидания случайной величины и конечность дисперсии. Во всех свойствах дисперсии предполагается существование вторых моментов случайных величин.

(i) Дисперсия может вычисляться по формуле: $\mathsf{D}\,\xi = \mathsf{E}\,\xi^2 - (\mathsf{E}\,\xi)^2$.

Доказательство. Положим для удобства выкладок $a = \mathsf{E}\,\xi$. Тогда $\mathsf{D}\,\xi = \mathsf{E}\,(\xi - a)^2 = \mathsf{E}\,(\xi^2 - 2a\xi + a^2) = \mathsf{E}\,\xi^2 - 2a\mathsf{E}\,\xi + a^2 = \mathsf{E}\,\xi^2 - a^2$.

(ii) При умножении случайной величины на константу c дисперсия умножается на c^2: $\mathsf{D}\,(c\xi) = c^2\,\mathsf{D}\,\xi$.

(iii) Дисперсия всегда неотрицательна: $\mathsf{D}\,\xi \ge 0$. Дисперсия обращается в нуль лишь для вырожденного распределения: если $\mathsf{D}\,\xi = 0$, то $\xi = c$ п.н. и наоборот, где c - некоторая константа.

Доказательство. Дисперсия есть математическое ожидание почти наверное (далее п.н.) неотрицательной случайной величины $(\xi - \mathsf{E}\,\xi)^2$, и неотрицательность дисперсии следует из свойства (v) математического ожидания. Далее, по свойству (vi) математического ожидания из равенства дисперсии нулю вытекает $(\xi - \mathsf{E}\,\xi)^2 = 0$ п.н., т.е. $\xi = \mathsf{E}\,\xi$ п.н. И наоборот, если $\xi = c$ п.н., то $\mathsf{D}\,\xi = \mathsf{E}\,(c - \mathsf{E}\,c)^2 = 0$.

(iv) Дисперсия не зависит от сдвига случайной величины на константу: $\mathsf{D}\,(\xi + c) = \mathsf{D}\,\xi$.

(v) Если ξ и ζ независимы, то $\mathsf{D}\,(\xi + \zeta) = \mathsf{D}\,\xi + \mathsf{D}\,\zeta$.

Доказательство. Действительно,

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

$$-(\mathsf{E}\,\xi)^2 - (\mathsf{E}\,\zeta)^2 - 2\mathsf{E}\,\xi\mathsf{E}\,\zeta = \mathsf{D}\,\xi + \mathsf{D}\,\zeta,$$

так как математическое ожидание произведения независимых случайных величин равно произведению их математических ожиданий.

Следствие 8. Если ξ и ζ независимы, то

$$\mathsf{D}\,(\xi - \zeta) = \mathsf{D}\,(\xi + \zeta) = \mathsf{D}\,\xi + \mathsf{D}\,\zeta.$$

Доказательство. Из свойств (ii) и (v) дисперсии, получаем

$$\mathsf{D}\,(\xi - \zeta) = \mathsf{D}\,(\xi + (-\zeta)) = \mathsf{D}\,\xi + \mathsf{D}\,(-\zeta) = \mathsf{D}\,\xi + (-1)^2\mathsf{D}\,\zeta = \mathsf{D}\,\xi + \mathsf{D}\,\zeta.$$

Следствие 9. Для произвольных случайных величин ξ и ζ с конечными вторыми моментами имеет место равенство

$$\mathsf{D}\,(\xi + \zeta) = \mathsf{D}\,\xi + \mathsf{D}\,\zeta + 2\mathsf{cov}(\xi, \zeta),$$

где $\mathsf{cov}(\xi, \zeta) = \mathsf{E}\,(\xi\zeta) - \mathsf{E}\,\xi\,\mathsf{E}\,\zeta$.

(vi) $\mathsf{D}\,\xi = \mathsf{E}\,(\xi - \mathsf{E}\,\xi)^2 = \min\limits_a \mathsf{E}\,(\xi - a)^2$.

Доказательство. Сравним величину $E(\xi - a)^2$ с дисперсией:
$\mathsf{E}\,(\xi - a)^2 = \mathsf{E}\,\big((\xi - \mathsf{E}\,\xi) + (\mathsf{E}\,\xi - a)\big)^2 = \mathsf{D}\,\xi + \big(\mathsf{E}\,\xi - a\big)^2 + 2(\mathsf{E}\,\xi - \mathsf{E}\,\xi)\,\big(\mathsf{E}\,\xi - a\big) =$
$= \mathsf{D}\,\xi + \big(\mathsf{E}\,\xi - a\big)^2 \geq \mathsf{D}\,\xi,$
и последнее неравенство превращается в равенство лишь при $a = \mathsf{E}\,\xi$.

§5. *Примеры вычисления математических ожиданий и дисперсий*

Начнём этот параграф с простейших примеров.

Пример 30 (вырожденное распределение). Для с.в. $\xi \sim \mathsf{I}_c$ математическое ожидание и дисперсию этого распределения уже известны из свойства *(ii)* для математических ожиданий и из свойства *(iii)* для дисперсий: $\mathsf{E}\,c = c, \;\; \mathsf{D}\,c = 0$.

Пример 31 (распределение Бернулли). Пусть $\xi \sim \mathsf{Bi}(1, p)$. Вычислим два момента и дисперсию: $\mathsf{E}\,\xi = 1 \cdot p \,+\, 0 \cdot q = p$; $\mathsf{E}\,\xi^2 = 1^2 \cdot p \,+\, 0^2 \cdot q = p$; $\mathsf{D}\,\xi = \mathsf{E}\,\xi^2 - (\mathsf{E}\,\xi)^2 = p - p^2 = pq$.

Пример 32 (биномиальное распределение). Пусть $\xi \sim \mathsf{Bi}(n, p)$. Используем свойство воспроизводимости биномиального распределения по первому параметру. Пусть независимые с.в. $\xi_1, \xi_2, \ldots, \xi_n$, имеют одинаковое распределение Бернулли $\mathrm{B}i(1, p)$. Тогда их сумма $S_n = \xi_1 + \xi_2 + \ldots + \xi_n$ имеет распределение $\mathrm{B}i(n, p)$, и по свойству (iv) математического ожидания получаем $\mathsf{E}\, S_n = \sum\limits_{i=1}^{n} \mathsf{E}\,\xi_i = n\mathsf{E}\,\xi_1 = np$. А поскольку с.в. ξ_i независимы, и дисперсия каждой с.в. равна pq, то $\mathsf{D}\, S_n = \sum\limits_{i=1}^{n} \mathsf{D}\,\xi_i = n\mathsf{D}\,\xi_1 = npq$.

Таким образом, $\mathsf{E}\,\xi = np$, $\mathsf{D}\,\xi = npq$.

Пример 33 (геометрическое распределение). Пусть $\xi \sim \mathsf{G}(p)$. Вычислим математическое ожидание ξ:

$$\mathsf{E}\,\xi = \sum_{k=1}^{\infty} k\, p\, q^{k-1} = p \sum_{k=1}^{\infty} k\, q^{k-1} = p \sum_{k=1}^{\infty} \frac{dq^k}{dq} = p\, \frac{d}{dq}\left(\sum_{k=1}^{\infty} q^k\right) =$$

$$= p\, \frac{d}{dq}\left(\frac{q}{1-q}\right) =$$

$= p\, \frac{1}{(1-q)^2} = \frac{1}{p}$. Вычислим так называемый *второй факториальный момент* ξ:

$$\mathsf{E}\,\xi(\xi-1) = \sum_{k=2}^{\infty} k(k-1)\, p\, q^{k-2} = p\, q \sum_{k=0}^{\infty} \frac{d^2 q^k}{dq^2} =$$

$$= p\, q\, \frac{d^2}{dq^2}\left(\sum_{k=0}^{\infty} q^k\right) = p\, q\, \frac{d^2}{dq^2}\left(\frac{1}{1-q}\right) =$$

$= p\, q\, \frac{2}{(1-q)^3} = \frac{2q}{p^2}$. Найдем теперь дисперсию через второй факториальный момент:
$\mathsf{D}\,\xi = \mathsf{E}\,\xi(\xi-1) + \mathsf{E}\,\xi - (\mathsf{E}\,\xi)^2 = \dfrac{2q}{p^2} + \dfrac{1}{p} - \dfrac{1}{p^2} = \dfrac{2q-1+p}{p^2} = \dfrac{q}{p^2}$.

Пример 34 (распределение Пуассона). Пусть $\xi \sim \Pi(\lambda)$. Вычислим математическое ожидание ξ:

$$\mathsf{E}\,\xi = \sum_{k=0}^{\infty} k\, \frac{\lambda^k}{k!}\, e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{\infty} k\, \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} =$$

$$= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} =$$

$= \lambda e^{-\lambda} e^{\lambda} = \lambda$. Моменты более высоких порядков легко находятся через факториальные моменты $\mathsf{E}\,\xi^{[m]} = \mathsf{E}\,\xi(\xi-1)\ldots(\xi-m+1)$ порядка m. Так, второй факториальный момент ξ равен

$$\mathsf{E}\,\xi(\xi-1)=\sum_{k=0}^{\infty}k(k-1)\,\frac{\lambda^k}{k!}\,e^{-\lambda}=\lambda^2e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}=\lambda^2e^{-\lambda}e^{\lambda}=\lambda^2.$$

Поэтому $\mathsf{E}\,\xi^2=\mathsf{E}\,\xi(\xi-1)+\mathsf{E}\,\xi=\lambda^2+\lambda$ и $\mathsf{D}\,\xi=\mathsf{E}\,\xi^2-(\mathsf{E}\,\xi)^2=\lambda$.

Пример 35 (равномерное распределение). Пусть $\xi\sim\mathrm{R}(a,b)$. Математическое ожидание $\mathsf{E}\,\xi=\frac{a+b}{2}$ найдено выше в примере 28. Вычислим второй момент:

$$\mathsf{E}\,\xi^2=\int\limits_{-\infty}^{\infty}x^2f_\xi(x)\,dx=\int\limits_{a}^{b}x^2\,\frac{1}{b-a}\,dx=\frac{b^3-a^3}{3(b-a)}=\frac{a^2+ab+b^2}{3}.$$

Дисперсия равна $\mathsf{D}\,\xi=\mathsf{E}\,\xi^2-(\mathsf{E}\,\xi)^2=(b-a)^2/\,12$.

Пример 36 (стандартное нормальное распределение). Пусть $\xi\sim\mathrm{N}(0,\,1)$.

Математическое ожидание ξ равно нулю:

$$\mathsf{E}\,\xi=\int\limits_{-\infty}^{\infty}x f_\xi(x)\,dx=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}x\,e^{-x^2/2}\,dx=0,$$

так как под знаком сходящегося интеграла стоит нечетная функция. Далее,

$$\mathsf{E}\,\xi^2=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}x^2\,e^{-x^2/2}\,dx=\frac{2}{\sqrt{2\pi}}\int\limits_{0}^{\infty}x^2\,e^{-x^2/2}\,dx=$$

$$=-\frac{2}{\sqrt{2\pi}}\int\limits_{0}^{\infty}x\,de^{-x^2/2}=-\frac{2x}{\sqrt{2\pi}}\,e^{-x^2/2}\Big|_0^{\infty}+$$

$$+2\int\limits_{0}^{\infty}\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}\,dx=0+\int\limits_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}\,dx=1.$$

Поэтому $\mathsf{D}\,\xi=\mathsf{E}\,\xi^2-(\mathsf{E}\,\xi)^2=1-0=1$.

Пример 37. (нормальное распределение $\mathrm{N}(a,\,\sigma^2)$). Ранее было показано, что если $\xi\sim\mathrm{N}(a,\,\sigma^2)$, то $\zeta=\frac{\xi-a}{\sigma}\sim\mathrm{N}(0,\,1)$. Математическое ожидание $\mathsf{E}\,\zeta=0$ и дисперсия $\mathsf{D}\,\zeta=1$ стандартного нормального распределения вычислены выше. Тогда

$$\mathsf{E}\,\xi=\mathsf{E}\,(\sigma\zeta+a)=\sigma\mathsf{E}\,\zeta+a=a;\quad \mathsf{D}\,\xi=\mathsf{D}\,(\sigma\zeta+a)=\sigma^2\mathsf{D}\,\zeta=\sigma^2.$$

Таким образом, параметры a и σ^2 нормального распределения являются его математическим ожиданием и дисперсией.

Пример 38 (показательное распределение). Пусть $\xi \sim \mathbf{E}(\alpha)$. Найдем для произвольного целого положительного k момент порядка k:

$$\mathsf{E}\,\xi^k = \int_{-\infty}^{\infty} x^k f_\xi(x)\,dx = \int_0^{\infty} x^k\,\alpha\,e^{-\alpha x}\,dx =$$

$$= \frac{1}{\alpha^k}\int_0^{\infty} (\alpha x)^k\,e^{-\alpha x}\,d(\alpha x) = \frac{k!}{\alpha^k}.$$

В последнем равенстве использовалась гамма-функцией Эйлера:
$\Gamma(k+1) = \int\limits_0^{\infty} u^k\,e^{-u}\,du = k!$. Из формулы для момента порядка k находим:
$\mathsf{E}\,\xi = \frac{1}{\alpha}$, $\quad \mathsf{E}\,\xi^2 = \frac{2}{\alpha^2}$, $\quad \mathsf{D}\,\xi = \mathsf{E}\,\xi^2 - (\mathsf{E}\,\xi)^2 = \frac{1}{\alpha^2}$.

Пример 39 (стандартное распределение Коши). Пусть $\xi \sim \mathbf{C}(0,\,1)$. Математическое ожидание распределения Коши не существует, так как расходится интеграл

$$\mathsf{E}\,|\xi| = \int_{-\infty}^{\infty} |x|\,\frac{1}{\pi(1+x^2)}\,dx = \int_0^{\infty} \frac{1}{\pi(1+x^2)}\,dx^2 =$$

$$= \lim_{x\to+\infty}\frac{1}{\pi}\ln(1+x^2) = +\infty.$$

Расходится он потому, что подынтегральная функция ведет себя на бесконечности как $1/x$. Поэтому не существуют ни дисперсия, ни моменты более высоких порядков этого распределения. То же самое можно сказать про распределение Коши $\mathbf{C}(x_0,\gamma)$.

Глава 11. Числовые характеристики зависимых случайных величин

§1. Ковариация случайных величин

Из свойства (v) дисперсий известно, что для независимых случайных величин с конечными вторыми моментами дисперсия их суммы равна сумме их дисперсий. В общем случае дисперсия суммы равна (см. следствие 9):

$$\mathsf{D}\,(\xi+\zeta) = \mathsf{D}\,\xi + \mathsf{D}\,\zeta + 2\big(\mathsf{E}\,(\xi\zeta) - \mathsf{E}\,\xi\,\mathsf{E}\,\zeta\big). \tag{1}$$

Величина $\mathsf{E}\,(\xi\zeta) - \mathsf{E}\,\xi\,\mathsf{E}\,\zeta$ равняется нулю, если случайные величины ξ и ζ независимы. С другой стороны, из равенства её нулю вовсе не следует независимость, как показывает пример 29. Эту величину используют как характеристику *наличия зависимости* между двумя случайными величинами.

Определение 54. Ковариацией $\mathbf{cov}(\xi,\,\zeta)$ случайных величин ξ и ζ называется число $\mathbf{cov}(\xi,\,\zeta) = \mathsf{E}\left[(\xi - \mathsf{E}\,\xi)(\zeta - \mathsf{E}\,\zeta)\right]$.

Свойства ковариаций. Справедливы равенства:

$$j)\;\; \mathbf{cov}(\xi,\,\zeta) = \mathsf{E}\,(\xi\zeta) - \mathsf{E}\,\xi\,\mathsf{E}\,\zeta; \quad jj)\;\; \mathbf{cov}(\xi,\,\xi) = \mathsf{D}\,\xi;$$
$$jjj)\;\; \mathbf{cov}(c\cdot\xi,\,\zeta) = c\cdot\mathbf{cov}(\xi,\,\zeta); \quad jv)\;\; \mathbf{cov}(\xi,\,\zeta) = \mathbf{cov}(\zeta,\,\xi).$$

Справедливо обобщённое свойство дисперсии.

Свойство (vii) дисперсии. Дисперсия суммы нескольких случайных величин вычисляется по любой из следующих формул:

$$\mathsf{D}\,(\xi_1 + \ldots + \xi_n) = \sum_{i=1}^{n} \mathsf{D}\,\xi_i + \sum_{i\neq j} \mathbf{cov}(\xi_i, \xi_j) =$$
$$= \sum_{i=1}^{n} \mathsf{D}\,\xi_i + 2\sum_{i<j} \mathbf{cov}(\xi_i, \xi_j) = \sum_{i,j} \mathbf{cov}(\xi_i, \xi_j).$$

Эти равенства доказываются непосредственным возведением суммы в квадрат.

Итак, что имеем? Если ковариация $\mathbf{cov}(\xi,\,\zeta) \neq 0$, то с.в. ξ и η зависимы. Чтобы судить о зависимости согласно любому из определений независимости, требуется знать совместное распределение пары ξ и ζ. Но найти совместное распределение достаточно часто сложнее, чем просто посчитать математическое ожидание произведения ξ и ζ. Если $\mathsf{E}\,\xi\,\zeta \neq \mathsf{E}\,\xi\,\mathsf{E}\,\zeta$, то устанавливается зависимость между ξ и ζ не находя их совместного распределения.

Пример 40. Этот пример показывает, что с помощью ковариации можно судить о зависимости даже в том случае, когда для нахождения совместного распределения недостаточно данных. Пусть с.в. ξ и ζ - независимы и $\mathsf{D}\,\xi > 0$. Покажем, что с.в. ξ и $\xi + \zeta$ зависимы:

$$\mathsf{E}\left[\xi(\xi + \zeta)\right] = \mathsf{E}\,\xi^2 + \mathsf{E}\,\xi\,\mathsf{E}\,\zeta, \qquad \mathsf{E}\,\xi\,\mathsf{E}\,(\xi + \zeta) = (\mathsf{E}\,\xi)^2 + \mathsf{E}\,\xi\,\mathsf{E}\,\zeta.$$

Вычитая одно из другого, получим $\mathbf{cov}(\xi, \xi + \zeta) = \mathsf{D}\,\xi > 0$. Следовательно, с.в. ξ и $\xi + \zeta$ зависимы.

Пример 41. Найдём математическое ожидание и дисперсию гипергеометрического распределения. Раньше это сделать было проблематично, так как требовалось вычислять суммы

$$\mathsf{E}\,\xi = \sum_k k\,\frac{C_K^k C_{N-K}^{n-k}}{C_N^n}, \qquad \mathsf{E}\,\xi^2 = \sum_k k^2\,\frac{C_K^k C_{N-K}^{n-k}}{C_N^n},$$

где суммирование ведётся по целым k таким, что $0 \le k \le K$ и $0 \le n-k \le N-K$.

Итак, пусть имеется урна, содержащая K белых шаров и $N-K$ чёрных. Пусть из этой урны без возвращения извлекают по одному n шаров. Свяжем случайную величину ξ, равную числу белых шаров среди n выбранных, с результатами отдельных извлечений шаров.

Обозначим ξ_i, где $i = 1, \ldots, n$, *индикатор* того, что i-й по счёту вынутый шар оказался белым: $\xi_i = 1$, если при i-м извлечении извлекли белый шар, иначе $\xi_i = 0$. Тогда $\xi = \xi_1 + \ldots + \xi_n$ – число появившихся белых шаров, и математическое ожидание считается достаточно просто:

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

Убедимся, что случайные величины $\xi_1, \ldots, \xi_n$ имеют одно и то же распределение Бернулли $\mathrm{B}i(1,p)$, где $p = K\,/\,N$.

Пронумеруем шары: белые - номерами от одного до K, чёрные - номерами от $K+1$ до N. Элементарным событием опыта является набор из n номеров шаров в схеме *упорядоченного* выбора n элементов из N без возвращения. Общее число элементарных событий равно $|\Omega| = C_N^n$.

Вычислим вероятность события $A_i = \{\xi_i = 1\}$. Событие A_i включает в себя элементарные события, в которых на i-м месте стоит любой из номеров белых шаров, а остальные $n-1$ мест занимают любые из оставшихся $N-1$ номеров. По основному комбинаторному принципу (теорема 1) число элементарных событий события A_i есть произведение K и A_{N-1}^{n-1}, где K – число способов поставить на i-е место один из номеров белых шаров, A_{N-1}^{n-1} - число способов после этого разместить на оставшихся $n-1$ местах остальные $N-1$ номеров шаров. Но тогда

$$p = \mathsf{P}(\xi_i = 1) = \mathsf{P}(A_i) = \frac{|A_i|}{|\,\Omega\,|} = \frac{K C_{N-1}^{n-1}}{C_N^n} = \frac{K}{N}$$

что совершенно очевидно: вероятность m-му шару быть белым, если ничего не известно про цвет первых $m-1$ шаров, точно такая же, как вероятность первому шару быть белым и равна отношению числа белых шаров к числу всех.

Вернёмся к математическому ожиданию:

$$\mathsf{E}\,\xi = \mathsf{E}\,\xi_1 + \ldots + \mathsf{E}\,\xi_n = n\mathsf{E}\,\xi_1 = np = \frac{nK}{N}\,.$$

Вычислим дисперсию ξ. До сих пор не говорили о совместном распределении $\xi_1, \ldots, \xi_n$: для вычисления математического ожидания их суммы было достаточно знания маргинальных распределений этих величин. Но дисперсия суммы уже не всегда равна сумме дисперсий. Зависимость величин $\xi_1, \ldots, \xi_n$ очевидна: если случилось событие $A_1 = \{\xi_1 = 1\}$, то вероятность второму шару быть белым уже не равна отношению K / N:

$$\mathsf{P}\big(\{\xi_2 = 1\} \mid \{\xi_1 = 1\}\big) = \frac{K-1}{N-1} \neq \frac{K}{N} = \mathsf{P}(\xi_2 = 1).$$

Поэтому при вычислении дисперсии будем пользоваться свойством (vii) дисперсии. Вычислим ковариацию величин ξ_i и ξ_j, $i \neq j$. Для этого сначала найдём $\mathsf{E}\,(\xi_i\xi_j)$. Произведение $\xi_i\xi_j$ снова имеет распределение Бернулли: $\xi_i\xi_j = 1$, если при i-м и j-м извлечениях появились белые шары. Вероятность этого события равна

$$\mathsf{P}(\xi_i\xi_j = 1) = \mathsf{P}(A_i \cap A_j) = \frac{|A_i \cap A_j|}{|\,\Omega\,|} = \frac{K(K-1)C_{N-2}^{n-2}}{C_N^n} =$$

$$= \frac{K(K-1)}{N(N-1)}.$$

Тогда

$$\operatorname{cov}(\xi_i,\, \xi_j) = \mathsf{E}\,(\xi_i\xi_j) - \mathsf{E}\,\xi_i \mathsf{E}\,\xi_j =$$
$$= \frac{K(K-1)}{N(N-1)} - \frac{K}{N}\,\frac{K}{N} = -\frac{K(N-K)}{N^2(N-1)}.$$

Подставляя одинаковые дисперсии $\mathsf{D}\,\xi_i = p(1-p)$ и эти, не зависящие от i и j ковариации в формулу дисперсии суммы, получаем

$$\mathsf{D}\,\xi = \mathsf{D}\,(\xi_1 + \ldots + \xi_n) = \sum_{i=1}^{n} \mathsf{D}\,\xi_i + \sum_{i \neq j} \operatorname{cov}(\xi_i, \xi_j) =$$

$$= np(1-p) + n(n-1)\operatorname{cov}(\xi_1,\, \xi_2) =$$
$$= n\,\frac{K}{N}\Big(1 - \frac{K}{N}\Big) - n(n-1)\frac{K(N-K)}{N^2(N-1)} = n\,\frac{K}{N}\Big(1 - \frac{K}{N}\Big)\Big(1 - \frac{n-1}{N-1}\Big).$$

Заметим, что если извлекать шары с возвращением, то испытания станут независимыми испытаниями в схеме Бернулли; ставшие независимыми с.в. ξ_i в сумме дадут число белых шаров, имеющее биномиальное распределение с параметрами n и $p = \frac{K}{N}$ и точно такое же математическое ожидание $np = \frac{nK}{N}$, как и у числа белых шаров при выборе без возвращения. Дисперсия же у числа белых шаров при выборе без возвращения меньше, чем при выборе с возвращением - за счёт отрицательной коррелированности слагаемых ξ_i и ξ_j при $i \neq j$.

По значению величины $\operatorname{cov}(\xi, \zeta)$ нельзя судить о степени зависимости случайных величин, что не очень удобно для практических задач. Нужно каким-то образом нормировать ковариацию, получив из неё такую величину, абсолютное значение которой:

j. не менялось бы при умножении случайных величин на число;

jj. свидетельствовало бы о *величине зависимости* случайных величин.

Замечание. Говоря о величине зависимости между случайными величинами, имеется в виду следующее. Самая сильная зависимость - функциональная, а из функциональных - линейная зависимость, когда $\xi = a\zeta + b$ п.н. Бывают гораздо более слабые зависимости. Так, если по последовательности независимых с.в. $\xi_1, \xi_2, \ldots$ построить с.в. $\xi = \xi_1 + \ldots + \xi_{24} + \xi_{25}$ и с.в. $\zeta = \xi_{25} + \xi_{26} + \ldots + \xi_{90}$, то эти с.в. зависимы, но очень «слабо»: через единственное общее слагаемое ξ_{25}. Например, сильно ли зависимы число гербов в первых 25 подбрасываниях монеты и число гербов в испытаниях с 25-го по 90-е?

§2. Коэффициент корреляции

Определение 55. Коэффициентом корреляции $\rho(\xi, \zeta)$ с.в. ξ и ζ, дисперсии которых существуют и отличны от нуля, называется число

$$\rho(\xi, \zeta) = \frac{\operatorname{cov}(\xi, \zeta)}{\sqrt{\mathsf{D}\,\xi}\sqrt{\mathsf{D}\,\zeta}} = \frac{\mathsf{E}\left[(\xi - \mathsf{E}\,\xi)(\zeta - \mathsf{E}\,\zeta)\right]}{\sqrt{\mathsf{E}\,(\xi - \mathsf{E}\,\xi)^2}\sqrt{\mathsf{E}\,(\zeta - \mathsf{E}\,\zeta)^2}}\,.$$

Пример 42 (продолжение примера 40). Пусть независимые и одинаково распределённые с.в. ξ и ζ, и пусть $\mathsf{D}\,\xi = \mathsf{D}\,\zeta > 0$. Найдем коэффициент корреляции с.в. ξ и $\xi + \zeta$:

$$\rho(\xi, \xi + \zeta) = \frac{\operatorname{cov}(\xi, \xi + \zeta)}{\sqrt{\mathsf{D}\,\xi}\sqrt{\mathsf{D}\,(\xi + \zeta)}} = \frac{\mathsf{D}\,\xi}{\sqrt{\mathsf{D}\,\xi}\sqrt{\mathsf{D}\,\xi + \mathsf{D}\,\zeta}} =$$

$$= \frac{\mathsf{D}\,\xi}{\sqrt{\mathsf{D}\,\xi}\sqrt{2\mathsf{D}\,\xi}} = \frac{1}{\sqrt{2}}\,.$$

Коэффициент корреляции величин ξ и $\xi + \zeta$ равен косинусу угла 45°, образованного «векторами» ξ и $\xi + \zeta$, когда ξ и ζ «ортогональны» и их «длина» одинакова.

Теорема 29 (свойства коэффициента корреляции).

Коэффициент корреляции обладает следующими свойствами:

(i) если ξ и ζ независимы, то $\rho(\xi, \zeta) = 0$;

(ii) $|\rho(\xi, \zeta)| \leq 1$;

(iii) $|\rho(\xi, \zeta)| = 1$ тогда и только тогда, когда ξ и ζ п.н. линейно связаны, т.е. существуют числа $a \neq 0$ и b такие, что $\mathsf{P}(\zeta = a\xi + b) = 1$.

(iv) Для любых случайных величин ξ и ζ с конечной и ненулевой дисперсией при любых постоянных $a \neq 0$ и b имеет место равенство

$$\rho(a\xi + b,\, \zeta) = \operatorname{sgn}(a) \cdot \rho(\xi,\, \zeta), \quad \text{где } \operatorname{sgn}(a) = \frac{a}{|a|} \ - \ \text{знак } a.$$

Доказательство. Свойство (i) уже было доказано в свойствах для математических ожиданий. Пример 29 – один из многих возможных примеров того, что свойство (i) в обратную сторону неверно.

Докажем свойство (ii). Рассмотрим преобразование с.в. $\widehat{\xi} = \frac{\xi - \mathsf{E}\,\xi}{\sqrt{\mathsf{D}\,\xi}}$. С.в. $\widehat{\xi}$ имеет нулевое математическое ожидание и единичную дисперсию:

$$\mathsf{E}\,\widehat{\xi} = \mathsf{E}\,\frac{\xi - \mathsf{E}\,\xi}{\sqrt{\mathsf{D}\,\xi}} = \frac{\mathsf{E}\,\xi - \mathsf{E}\,\xi}{\sqrt{\mathsf{D}\,\xi}} = 0;$$

$$\mathsf{E}\,\widehat{\xi}^{\,2} = \mathsf{D}\,\widehat{\xi} = \mathsf{D}\,\frac{\xi - \mathsf{E}\,\xi}{\sqrt{\mathsf{D}\,\xi}} = \frac{\mathsf{D}\,(\xi - \mathsf{E}\,\xi)}{\mathsf{D}\,\xi} = 1$$

Коэффициент корреляции теперь запишется так: $\rho(\xi,\,\zeta) = \mathsf{E}\left(\widehat{\xi}\cdot\widehat{\zeta}\right)$. Далее, неравенство $(x \pm y)^2 \geq 0$ равносильно двустороннему неравенству $-\frac{1}{2}(x^2+y^2) \leq xy \leq \frac{1}{2}(x^2+y^2)$. Подставляя в него $\widehat{\xi}$ вместо x и $\widehat{\zeta}$ вместо y и взяв математические ожидания всех частей неравенства, получим свойство (ii):

$$-1 = -\frac{1}{2}\mathsf{E}\left(\widehat{\xi}^{\,2} + \widehat{\zeta}^{\,2}\right) \leq \rho(\xi,\,\zeta) = \mathsf{E}\left(\widehat{\xi}\cdot\widehat{\zeta}\right) \leq \frac{1}{2}\mathsf{E}\left(\widehat{\xi}^{\,2} + \widehat{\zeta}^{\,2}\right) = 1. \qquad (2)$$

Докажем свойство (iii). В одну сторону утверждение проверяется непосредственно: если $\zeta = a\xi + b$, то

$$\rho(\xi,\,a\xi + b) = \frac{\mathsf{E}\,(\xi(a\xi+b)) - \mathsf{E}\,\xi\cdot\mathsf{E}\,(a\xi+b)}{\sqrt{\mathsf{D}\,\xi}\sqrt{\mathsf{D}\,(a\xi+b)}} =$$

$$= \frac{a\mathsf{D}\,\xi}{\sqrt{\mathsf{D}\,\xi}\sqrt{a^2\mathsf{D}\,\xi}} = \begin{cases} 1, & a > 0, \\ -1, & a < 0. \end{cases}$$

Докажем вторую часть свойства (iii): если $|\rho(\xi,\,\zeta)| = 1$, то существуют числа $a \neq 0$ и b такие, что $\mathsf{P}(\zeta = a\xi + b) = 1$.

Рассмотрим сначала случай $\rho(\xi,\,\zeta) = \rho\left(\widehat{\xi},\,\widehat{\zeta}\right) = 1$. Тогда второе неравенство в формуле (2) превращается в равенство:

$$\mathsf{E}\left(\widehat{\xi}\cdot\widehat{\zeta}\right) = \frac{1}{2}\mathsf{E}\left(\widehat{\xi}^{\,2} + \widehat{\zeta}^{\,2}\right), \quad \text{т.е.} \quad \mathsf{E}\left(\widehat{\xi} - \widehat{\zeta}\right)^2 = 0.$$

Если математическое ожидание неотрицательной случайной величины $\left(\widehat{\xi} - \widehat{\zeta}\right)^2$ равно нулю, то $\left(\widehat{\xi} - \widehat{\zeta}\right)^2 = 0$, п.н. Поэтому с единичной вероятностью

$$\frac{\zeta - \mathsf{E}\,\zeta}{\sqrt{\mathsf{D}\,\zeta}} = \frac{\xi - \mathsf{E}\,\xi}{\sqrt{\mathsf{D}\,\xi}}, \qquad \zeta = \frac{\sqrt{\mathsf{D}\,\zeta}}{\sqrt{\mathsf{D}\,\xi}}\,\xi + \mathsf{E}\,\zeta - \frac{\sqrt{\mathsf{D}\,\zeta}}{\sqrt{\mathsf{D}\,\xi}}\,\mathsf{E}\,\xi = a\xi + b.$$

В случае $\rho(\xi,\,\zeta) = -1$ нужно рассмотреть первое неравенство в формуле (2) и повторить рассуждения.

Докажем свойство (iv). Используя свойства дисперсии, получаем:

$$\rho(a\xi + b,\, \zeta) = \frac{\mathrm{cov}(a\xi + b,\, \zeta)}{\sqrt{\mathsf{D}\,(a\xi + b)}\,\sqrt{\mathsf{D}\,\zeta}} =$$

$$= \frac{a\,\mathrm{cov}(\xi,\, \zeta)}{\sqrt{a^2 \mathsf{D}\,\xi}\,\sqrt{\mathsf{D}\,\zeta}} = \frac{a}{|a|} \cdot \rho(\xi,\, \zeta).$$

Тем самым теорема 29 доказана полностью.

Определение 56. С.в. ξ и ζ называются *отрицательно коррелированными*, если $\rho(\xi,\, \zeta) < 0$; *положительно коррелированными*, если $\rho(\xi,\, \zeta) > 0$; некоррелированными, если $\rho(\xi,\, \zeta) = 0$.

Смысл знака $\rho(\xi,\, \zeta)$ виден в случае $\rho(\xi,\, \zeta) = \pm 1$. Тогда знак ρ равен знаку a в равенстве $\zeta = a\xi + b$ п.н. Так, $\rho(\xi,\, \zeta) = 1$ означает, что чем больше ξ, тем больше и ζ. Напротив, $\rho(\xi,\, \zeta) = -1$ означает, что чем больше ξ, тем меньше ζ. Похожим образом можно трактовать знак коэффициента корреляции и в случае, когда $|\rho(\xi,\, \zeta)| < 1$, помня при этом, что зависимость между ξ и ζ теперь уже не линейная и, возможно, даже не функциональная.

Так, величины ξ и $\xi + \zeta$ в примерах 40 и 41 положительно коррелированы, но их зависимость не функциональная.

Пример 42. Если случайная точка $(\xi,\ \zeta)$ имеет равномерное распределение в треугольнике D с вершинами $(2, 0)$, $(0, 0)$ и $(0, 1)$, то коэффициент корреляции $\rho(\xi,\, \eta)$ отрицателен.

Действительно,

$$f_\xi(x) = (1 - \frac{x}{2})\ 1_{[0,2]}(x), \quad f_\zeta(y) = 2(1 - y)\ 1_{[0,1]}(y);$$

$$\mathsf{E}\,\xi = \int\limits_0^2 x\,(1 - \frac{x}{2})dx = \frac{2}{3}, \quad \mathsf{E}\,\zeta = 2\int\limits_0^1 y\,(1 - y)dy = \frac{1}{3}.$$

По определению многомерного равномерного распределения в области D,

$$\mathsf{E}\,(\xi\,\zeta) = \iint\limits_D xy \cdot 1\ dx\,dy = \int\limits_0^2 \int\limits_0^{1-x/2} x\,y\ dy\,dx = \frac{1}{6}.$$

Поэтому, $\mathrm{cov}(\xi, \zeta) = -\frac{1}{18} \ \Rightarrow\ \rho < 0$.

Пример 43. Найдем коэффициент корреляции между числом выпадений единицы и числом выпадений шестерки при n подбрасываниях правильной игральной кости.

Обозначим для $i \in \{1,\, \ldots,\, 6\}$ через ξ_i случайную величину, равную числу выпадений грани с i очками при n подбрасываниях кости. Найдём $\mathrm{cov}(\xi_1,\, \xi_6)$. Каждая из с.в.

$\xi_i \sim Bi(n, \frac{1}{6})$, поэтому $\mathsf{E}\,\xi_i = \frac{n}{6}$, $\mathsf{D}\,\xi_i = \frac{5n}{36}$.

Далее заметим, что $\xi_1 + \ldots + \xi_6 = n$. Из-за симметрии кости математические ожидания $\mathsf{E}\,\xi_1\xi_2, \mathsf{E}\,\xi_1\xi_3, \ldots, \mathsf{E}\,\xi_1\xi_6$ одинаковы, но отличаются от $\mathsf{E}\,\xi_1\xi_1 = \mathsf{E}\,\xi_1^2 = \mathsf{D}\,\xi_1 + (\mathsf{E}\,\xi_1)^2 = \frac{5n}{36} + \frac{n^2}{36}$. Найдём $\mathsf{E}\,\xi_1(\xi_1 + \cdots + \xi_6)$. С одной стороны, это число равно $\mathsf{E}\,\xi_1(\xi_1 + \ldots + \xi_6) = \mathsf{E}\,\xi_1 \cdot n = \frac{n^2}{6}$. С другой стороны, $\mathsf{E}\,\xi_1(\xi_1 + \ldots + \xi_6) = \mathsf{E}\,\xi_1^2 + 5\mathsf{E}\,\xi_1\xi_6 = \frac{5n}{36} + \frac{n^2}{36} + 5\mathsf{E}\,\xi_1\xi_6$. Отсюда $5\mathsf{E}\,\xi_1\xi_6 = \frac{n^2}{6} - \frac{5n}{36} - \frac{n^2}{36}, ..\mathsf{E}\,\xi_1\xi_6 = \frac{n^2-n}{36}$. Следовательно, искомый коэффициент корреляции равен $\rho(\xi_1, \xi_6) = \frac{\mathsf{E}\,\xi_1\xi_6 - \mathsf{E}\,\xi_1\mathsf{E}\,\xi_6}{\sqrt{\mathsf{D}\,\xi_1\mathsf{D}\,\xi_6}} = \frac{(n^2-n)/36-n^2/36}{5n/36} = -\frac{1}{5}$.

Заметим, что этот коэффициент корреляции не зависит от n.

Глава 12. Сходимость последовательностей случайных величин

§1. Сходимость почти наверное *и сходимость* по вероятности

Вспомним, что случайная величина есть измеримая функция из некоторого непустого множества Ω в множество действительных чисел. Последовательность с.в. $\xi_1, \xi_2, \ldots$ есть последовательность функций, определённых на одном и том же множестве Ω. Существуют разные виды сходимости последовательности функций. Давать определение любой сходимости будем, опираясь на сходимость числовых последовательностей, как на уже известное понятие математического анализа.

В частности, при каждом $\omega \in \Omega$ имеем свою числовую последовательность $\xi_1(\omega)$, $\xi_2(\omega)$, $\xi_3(\omega)$, $\ldots$. Поэтому можно говорить о сходимости последовательности значений функций в данной точке ω, а также во всех остальных точках $\omega \in \Omega$. В теории вероятностей можно не обращать внимание на «неприятности», происходящие с нулевой вероятностью. Поэтому вместо сходимости "всюду"принято рассматривать сходимость «почти всюду» или «почти наверное».

Пусть $(\Omega, \mathcal{A}, \mathsf{P})$ – некоторое заданное вероятностное пространство.

Определение 57. Последовательность с.в. $\xi_1, \xi_2, \ldots$ сходится почти наверное к случайной величине ξ при $n \to \infty$, и пишут: $\xi_n \to \xi$ *п.н.*, если

$$\mathsf{P}\left\{\omega \in \Omega : \lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)\right\} = 1\,.$$

Иначе говоря, если $\xi_n(\omega) \to \xi(\omega)$ при $n \to \infty$ для всех $\omega \in \Omega$, кроме, возможно, $\omega \in A$, где A – событие, такое, что $\mathsf{P}(A) = 0$.

Отметим сразу: определение сходимости *почти наверное* требует знания того, как устроены отображения $\omega \to \xi_n(\omega)$. В задачах же теории вероятностей, как правило, известны не сами случайные величины, а лишь их распределения.

Возможно ли, обладая только информацией о распределениях, говорить о какой-либо сходимости последовательности случайных величин $\{\xi_n\}$ к случайной величине ξ?

Можно, например, потребовать, чтобы вероятность тех элементарных событий ω, для которых $\xi_n(\omega)$ не попадает в ε-*окрестность* числа $\xi(\omega)$, уменьшалась до нуля с ростом n. Такая сходимость в функциональном анализе называется сходимостью *по мере*, а в теории вероятностей – сходимостью *по вероятности*.

Определение 58. Последовательность случайных величин $\{\xi_n\}$ называется *сходящейся по вероятности* к случайной величине ξ при $n \to \infty$, и пишут $\xi_n \xrightarrow{\mathrm{p}} \xi$, если

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} \mathsf{P}\left(|\xi_n - \xi| \geq \varepsilon\right) = 0$$

$$\left(\text{ или } \quad \lim_{n\to\infty} \mathsf{P}\left(|\xi_n - \xi| < \varepsilon\right) = 1 \right).$$

Пример 44. Рассмотрим последовательность $\xi_1, \xi_2, \ldots$, в которой все величины имеют разные распределения: величина ξ_n принимает значения 0 и n^7 с вероятностями $\mathsf{P}\bigl(\xi_n = n^7\bigr) = 1/n = 1 - \mathsf{P}(\xi_n = 0)$. Покажем, что эта последовательность сходится по вероятности к нулю.

Выберем произвольное $\varepsilon > 0$. Для всех n начиная с некоторого n_0 такого, что $n_0^7 > \varepsilon$, верно равенство $\mathsf{P}(\xi_n \geq \varepsilon) = \mathsf{P}(\xi_n = n^7) = 1/n$. Поэтому

$$\mathsf{P}\bigl(|\xi_n - 0| \geq \varepsilon\bigr) = \mathsf{P}\bigl(\xi_n \geq \varepsilon\bigr) = \mathsf{P}\bigl(\xi_n = n^7\bigr) = \frac{1}{n} \to 0 \text{ при } n \to \infty.$$

Таким образом, с.в. ξ_n с ростом n могут принимать все большие и большие значения, но со все меньшей и меньшей вероятностью. Например, последовательность $\{\xi_n\}$ можно задать на вероятностном пространстве $(\Omega, \mathcal{A}, \mathsf{P}) = ([0, 1], \mathfrak{B}([0, 1]), \mu)$ так: положим $\xi_n(\omega) = 0$ для $\omega \in [0, 1 - 1/n]$ и $\xi_n(\omega) = n^7$ для $\omega \in (1 - 1/n, 1]$.

Отметим, что сходимость по вероятности имеет место совершенно независимо от того, как именно заданы случайные величины на Ω, поскольку определяется лишь их распределениями.

Замечание. Другая ситуация со сходимостью *почти наверное.* Если, например, задать с.в. как указано выше, то п.н. сходимость будет иметь место. Действительно, $\forall\ \omega \in [0, 1)\ \exists\ n_0 : \omega \in [0, 1 - 1/n_0]$, и поэтому для всех $n \geq n_0$ все $\xi_n(\omega)$ равны нулю.

Можно попробовать задать случайные величины ξ_n на отрезке $[0, 1]$ как-нибудь иначе, чтобы не было сходимости почти наверное. Для этого нужно заставить отрезок длины $1/n$, на котором $\xi_n(\omega) = n^7$, передвигаться по всему отрезку $[0, 1]$, чтобы любая точка $\omega \in [0, 1]$ попадала внутрь этого отрезка бесконечное число раз, и, тем самым, для любого ω существовала подпоследовательность $\xi_{n_k}(\omega) \to \infty$.

Сходимость по вероятности не обязательно сопровождается сходимостью математических ожиданий или моментов других порядков: из $\xi_n \xrightarrow{\mathrm{p}} \xi$ не следует, что $\mathsf{E}\,\xi_n \to \mathsf{E}\,\xi$. Действительно, в примере 44 имеет место сходимость $\xi_n \xrightarrow{\mathrm{p}} \xi = 0$, но $\mathsf{E}\,\xi_n = n^6 \not\to \mathsf{E}\,\xi = 0$. При этом, вообще говоря, последовательность $\mathsf{E}\,\xi_n$ неограниченно возрастает.

А если вместо значения n^7 взять n (с той же вероятностью $1/n$), то получим $\mathsf{E}\,\xi_n = 1 \not\to \mathsf{E}\,\xi = 0$. Но теперь хотя бы предел у последовательности математических ожиданий конечен.

Если же ξ_n принимает значения 0 и $\sqrt{n}$ с вероятностями из примера 44, то $\mathsf{E}\,\xi_n = 1/\sqrt{n} \to \mathsf{E}\,\xi = 0$, но уже вторые моменты сходиться ко второму моменту ξ не будут:

$\mathsf{E}\,\xi_n^2 = 1 \not\to \mathsf{E}\,\xi^2 = 0$.

Однако сходимость математических ожиданий и других моментов сходящихся последовательностей бывает чрезвычайно важна в различных задачах математической статистики. Существуют условия, при выполнении которых сходимость по вероятности $\xi_n \xrightarrow{\mathrm{p}} \xi$ влечёт сходимость математических ожиданий $\mathsf{E}\,\xi_n \to \mathsf{E}\,\xi$.

Теорема 30. Пусть $\xi_n \xrightarrow{\mathrm{p}} \xi$ при $n \to \infty$. Тогда для сходимости $\mathsf{E}\,\xi_n \to \mathsf{E}\,\xi$ достаточно выполнения любого из следующих условий:

j) все члены последовательности ограничены одной и той же постоянной: $|\xi_n| \leq C$;

jj) все члены последовательности ограничены одной и той же случайной величиной с конечным первым моментом: $|\xi_n| \leq \zeta,\ \mathsf{E}\zeta < \infty$;

jjj) существует число $\alpha > 1$ такое, что $\mathsf{E}\,|\xi_n|^\alpha < C \ \ \forall\, n$.

Эту теорему оставляем без доказательства.

Самым слабым в этом списке является условие *jjj)*, а *j)* наиболее ограничительным. Ни одно из этих условий не является необходимым для сходимости математических ожиданий.

Сформулируем свойства сходимости по вероятности в виде отдельных теорем.

Теорема 31. Пусть функция $g:\ R^1 \to R^1$.

(i) Если $\xi_n \xrightarrow{\mathrm{p}} \xi$ и функция $g(x)$ непрерывна, то $g(\xi_n) \xrightarrow{\mathrm{p}} g(\xi)$.

(ii) Если $\xi_n \xrightarrow{\mathrm{p}} c$ и $g(x)$ непрерывна в точке c, то $g(\xi_n) \xrightarrow{\mathrm{p}} g(c)$.

Доказательство. Простое доказательство первого утверждения можно предложить в двух случаях, которыми и ограничимся: если $\xi = c$ п.н. (тогда достаточно, чтобы g была непрерывна в точке c) или если функция g равномерно непрерывна.

В обоих случаях $\forall\ \varepsilon > 0\ \exists\ \delta > 0$, такое. что $\forall\ \omega \in \Omega$, удовлетворяющего условию $|\xi_n(\omega) - \xi(\omega)| < \delta$, выполняется неравенство $|g(\xi_n(\omega)) - g(\xi(\omega))| < \varepsilon$.

Другими словами, событие $\{|\xi_n - \xi| < \delta\}$ влечёт за собой событие $\{|g(\xi_n) - g(\xi)| < \varepsilon\}$. Следовательно, вероятность первого события не больше вероятности второго события. Но, какое бы ни было $\delta > 0$, вероятность первого события стремится к единице по определению сходимости по вероятности:

$$1 \longleftarrow \mathsf{P}\big(|\xi_n - \xi| < \delta\big) \leq \mathsf{P}\big(|g(\xi_n) - g(\xi)| < \varepsilon\big) \leq 1.$$

Тогда вероятность второго события также стремится к единице.

Теорема 32. Пусть функция $g:\ R^2 \to R^1$.

(iii) Если $\xi_n \xrightarrow{\mathrm{p}} \xi$, $\zeta_n \xrightarrow{\mathrm{p}} \zeta$ при $n \to \infty$ и функция $g(x,\, y)$ всюду непрерывна, то $g(\xi_n,\, \zeta_n) \xrightarrow{\mathrm{p}} g(\xi,\, \zeta)$.

(iv) Если $\xi_n \xrightarrow{\mathrm{p}} c_1$, $\zeta_n \xrightarrow{\mathrm{p}} c_2$ при $n \to \infty$ и функция $g(x, y)$ непрерывна в точке $(c_1,\, c_2)$, то $g(\xi_n,\, \zeta_n) \xrightarrow{\mathrm{p}} g(c_1,\, c_2)$.

Доказательство. Докажем снова только второе свойство. Воспользуемся определением непрерывности функции двух переменных: $\forall\, \varepsilon > 0\, \exists\, \delta > 0$, что $\forall \omega \in \Omega$, принадлежащего одновременно двум событиям

$$A_n = \{|\xi_n(\omega) - c_1| < \delta\}, \qquad B_n = \{|\zeta_n(\omega) - c_2| < \delta\},$$

выполняется неравенство $|g(\xi_n(\omega), \zeta_n(\omega)) - g(c_1, c_2)| < \varepsilon$. Тогда событие $A_n \cap B_n$ влечёт событие $C = \{|g(\xi_n, \zeta_n) - g(c_1, c_2)| < \varepsilon\}$, поэтому вероятность первого события не больше вероятности второго события. Но вероятность пересечения двух событий, вероятности которых стремятся к единице, также стремится к единице:

$$\mathsf{P}(A_n \cap B_n) = 1 - \mathsf{P}(\overline{A_n} \cup \overline{B_n}) \geq 1 - \mathsf{P}(\overline{A_n}) - \mathsf{P}(\overline{B_n}) \to 1.$$

Поэтому $\mathsf{P}(C) \geq \mathsf{P}(A_n \cap B_n) \to 1$ при $n \to \infty$.

Из свойств (iii) и (iv) вытекают обычные свойства пределов, хорошо знакомые из курса математического анализа. Например, функции $g(x, y) = x + y$ и $g(x, y) = xy$ непрерывны в R^2, поэтому предел суммы или, соответственно, произведения сходящихся по вероятности последовательностей равен сумме или, соответственно, произведению пределов:

Если $\xi_n \xrightarrow{p} \xi$ и $\zeta_n \xrightarrow{p} \zeta$, то
(v) $\xi_n + \zeta_n \xrightarrow{p} \xi + \zeta$ и
(vi) $\xi_n \cdot \zeta_n \xrightarrow{p} \xi \cdot \zeta$.

Покажем, что п.н. сходимость сильнее сходимости по вероятности.

Теорема 33. Если $\xi_n \to \xi$ п.н., то отсюда сразу же следует сходимость $\xi_n \xrightarrow{p} \xi$.

Доказательство. Ограничимся для простоты доказательства лишь случаем, когда $\xi_n(\omega) \to \xi(\omega)\ \ \forall\, \omega \in \Omega$.

Зафиксируем $\omega \in \Omega$. По определению предела, $\xi_n(\omega) \to \xi(\omega)$ при $n \to \infty$, если $\forall\, \varepsilon > 0\ \exists\, N = N(\omega, \varepsilon) \geq 0$ такое, что $\forall\, n > N$ выполняется неравенство $|\xi_n(\omega) - \xi(\omega)| < \varepsilon$. Событие $A = \{n > N(\omega, \varepsilon)\}$ влечет событие $B = \{|\xi_n(\omega) - \xi(\omega)| < \varepsilon\}$. Тогда

$$1 \geq \mathsf{P}(B) \geq \mathsf{P}(A) = \mathsf{P}\big(N(\omega, \varepsilon) < n\big) =$$

$$= F_{N(\varepsilon,\omega)}(n) \to 1 \text{ при } n \to \infty$$

по свойству (ii) функций распределения. Таким образом, получаем, что $\mathsf{P}(B) \to 1$, т.е. $\xi_n \xrightarrow{p} \xi$.

§2. Неравенства Чебышёва

Чтобы доказывать сходимость последовательности с.в. по вероятности, требуется уметь вычислять $\mathsf{P}\,(|\xi_n - \xi| \geq \varepsilon)$ при больших n. Но для этого нужно знать распределение ξ_n, что не всегда возможно.

Полезно иметь неравенства, позволяющие оценивать вероятность $\mathsf{P}\left(|\xi_n - \xi| \geq \varepsilon\right)$ сверху. Тогда для доказательства сходимости по вероятности было бы достаточно устремить к нулю эту оценку.

Все неравенства в этом параграфе принято относить к одному классу, называемому неравенством Чебышёва.

Теорема 34 (неравенство Маркова). Если $\mathsf{E}\,|\xi| < \infty$, то $\forall\ x > 0$

$$\mathsf{P}\big(|\xi| \geq x\big) \leq \frac{\mathsf{E}\,|\xi|}{x}.$$

Доказательство. По определению индикатора, с.в. $\mathbf{1}_A(\omega),\ \omega \in \Omega$ имеет распределение Бернулли с параметром $p = \mathsf{P}(\omega \in \Omega:\ \mathbf{1}_A(\omega) = 1) = \mathsf{P}(A)$ и её математическое ожидание равно вероятности успеха $p = \mathsf{P}(A)$. Ясно, что индикаторы прямого и противоположного событий связаны равенством $\mathbf{1}_A(\omega) + \mathbf{1}_{\overline{A}}(\omega) = 1$. Поэтому

$$|\xi| = |\xi| \cdot \mathbf{1}_{(|\xi|<x)} + |\xi| \cdot \mathbf{1}_{(|\xi|\geq x)} \geq |\xi| \cdot \mathbf{1}_{(|\xi|\geq x)} \geq x \cdot \mathbf{1}_{(|\xi|\geq x)}.$$

Тогда

$$\mathsf{E}\,|\xi| \geq \mathsf{E}\left(x \cdot \mathbf{1}_{(|\xi|\geq x)}\right) = x \cdot \mathsf{P}\big(|\xi| \geq x\big).$$

Осталось разделить обе части этого неравенства на положительное число x.

Следствие 10 (обобщенное неравенство Чебышёва). Пусть функция g не убывает и положительна на R^1. Если $\mathsf{E}\,g(\xi) < \infty$, то $\forall x \in R^1$

$$\mathsf{P}\big(\xi \geq x\big) \leq \frac{\mathsf{E}\,g(\xi)}{g(x)}.$$

Доказательство. Заметим, что $\mathsf{P}\big(\xi \geq x\big) \leq \mathsf{P}\big(g(\xi) \geq g(x)\big)$, поскольку функция g не убывает. Оценим последнюю вероятность по неравенству Маркова, которое можно применять в силу положительности g:

$$\mathsf{P}\big(g(\xi) \geq g(x)\big) \leq \frac{\mathsf{E}\,g(\xi)}{g(x)}.$$

Следствие 11 (неравенство Чебышёва). Если $\mathsf{D}\,\xi$ существует, то для любого $x > 0$

$$\mathsf{P}\big(|\xi - \mathsf{E}\,\xi| \geq x\big) \leq \frac{\mathsf{D}\,\xi}{x^2}.$$

Доказательство. Для $x > 0$ неравенство $|\xi - \mathsf{E}\,\xi| \geq x$ равносильно неравенству $(\xi - \mathsf{E}\,\xi)^2 \geq x^2$, поэтому

$$\mathsf{P}\big(|\xi - \mathsf{E}\,\xi| \geq x\big) = \mathsf{P}\big((\xi - \mathsf{E}\,\xi)^2 \geq x^2\big) \leq \frac{\mathsf{E}\left(\xi - \mathsf{E}\,\xi\right)^2}{x^2} = \frac{\mathsf{D}\,\xi}{x^2}.$$

Неравенство Чебышёва позволяет, помимо всего прочего, получать абсолютные оценки для вероятности того, что стандартизованная случайная величина(т.е. с.в. $\frac{\xi - \mathsf{E}\,\xi}{\sqrt{\mathsf{D}\,\xi}}$) превзойдет некоторое значение: для любого $x > 0$

$$\mathsf{P}\left(\left|\frac{\xi - \mathsf{E}\,\xi}{\sqrt{\mathsf{D}\,\xi}}\right| \geq x\right) = \mathsf{P}\left(|\xi - \mathsf{E}\,\xi| \geq x\sqrt{\mathsf{D}\,\xi}\right) \leq \frac{\mathsf{D}\,\xi}{x^2 \mathsf{D}\,\xi} = \frac{1}{x^2}.$$

Например, при x=10 эта вероятность не превышает **0.01**.

§3. Законы больших чисел

Определение 59. Последовательность с.в. $\xi_1, \xi_2, \ldots$ с конечными первыми моментами удовлетворяет закону больших чисел (ЗБЧ), если

$$\frac{\xi_1 + \ldots + \xi_n}{n} - \frac{\mathsf{E}\,\xi_1 + \ldots + \mathsf{E}\,\xi_n}{n} \xrightarrow{\text{p}} 0 \text{ при } n \to \infty. \tag{3}$$

Законами больших чисел обычно называют утверждения, где говорится об условиях, при которых последовательность с.в. удовлетворяет закону больших чисел.

Посмотрим сначала, когда выполнен ЗБЧ для последовательности независимых и одинаково распределённых случайных величин.

Теорема 35 (ЗБЧ в форме Чебышёва). Для любой последовательности $\xi_1, \xi_2, \ldots$ попарно независимых и одинаково распределённых с.в. с конечным вторым моментом $\mathsf{E}\,\xi_1^2 < \infty$ имеет место сходимость

$$\frac{\xi_1 + \ldots + \xi_n}{n} \xrightarrow{\text{p}} \mathsf{E}\,\xi_1. \tag{4}$$

Заметим, что если с.в. одинаково распределены, то их математические ожидания одинаковы (и равны, например, $\mathsf{E}\,\xi_1$), поэтому свойство (3) можно записать в виде (4).

ЗБЧ утверждает, что среднее арифметическое большого числа с.в. *стабилизируется* с ростом этого числа. Как бы сильно каждая случайная величина ни отклонялась от своего среднего значения, при суммировании эти отклонения компенсируются, так что среднее арифметическое с.в. приближается к постоянной величине (по вероятности).

Дальше станет понятным, что требование конечности второго момента (или дисперсии) связано исключительно со способом доказательства, и что утверждение останется верным, если требовать существования только первого момента.

Доказательство. Обозначим $S_n = \xi_1 + \ldots + \xi_n$ сумму первых n случайных величин. Из свойства линейности математического ожидания получим

$$\mathsf{E}\left(\frac{S_n}{n}\right) = \frac{\mathsf{E}\,\xi_1 + \ldots + \mathsf{E}\,\xi_n}{n} = \frac{n\,\mathsf{E}\,\xi_1}{n} = \mathsf{E}\,\xi_1.$$

Пусть $\varepsilon > 0$. Воспользуемся неравенством Чебышёва (*следствие 11*):

$$\mathsf{P}\left(\left|\frac{S_n}{n} - \mathsf{E}\left(\frac{S_n}{n}\right)\right| \geq \varepsilon\right) \leq \frac{\mathsf{D}\left(\dfrac{S_n}{n}\right)}{\varepsilon^2} = \frac{\mathsf{D}S_n}{n^2\varepsilon^2} =$$

$$= \frac{\mathsf{D}\xi_1 + \ldots + \mathsf{D}\xi_n}{n^2\varepsilon^2} = \frac{n\,\mathsf{D}\xi_1}{n^2\varepsilon^2} = \frac{\mathsf{D}\xi_1}{n\varepsilon^2} \to 0$$

при $n \to \infty$, так как $\mathsf{D}\,\xi_1 < \infty$. Дисперсия суммы превратилась в сумму дисперсий в силу попарной независимости слагаемых, из-за которой все ковариации $\mathbf{cov}(\xi_i,\, \xi_j)$ в обобщённом свойстве дисперсии (vii) обратились в ноль при $i \neq j$.

Замечание. Была доказана не только сходимость по вероятности, но и получена оценка для вероятности среднего арифметического любого числа попарно независимых и одинаково распределённых величин отличаться от $\mathsf{E}\,\xi_1$ более чем на заданное ε:

$$\mathsf{P}\left(\left|\frac{\xi_1 + \ldots + \xi_n}{n} - \mathsf{E}\,\xi_1\right| \geq \varepsilon\right) \leq \frac{\mathsf{D}\,\xi_1}{n\varepsilon^2}. \tag{5}$$

Попарную независимость слагаемых в ЗБЧ Чебышёва можно заменить их попарной некоррелированностью, ничего не меняя в доказательстве. ЗБЧ может выполняться и для последовательности зависимых и разнораспределённых слагаемых. Из неравенства Чебышёва сразу вытекает следующее достаточное условие выполнения ЗБЧ для последовательности произвольных случайных величин.

Теорема 36 (ЗБЧ в форме Маркова). Последовательность случайных величин $\xi_1, \xi_2, \ldots$ с конечными вторыми моментами удовлетворяет ЗБЧ, если $\mathsf{D}\,S_n = o(n^2)$, т.е. если $\frac{\mathsf{D}\,S_n}{n^2} \to 0$ при $n \to \infty$.

Теорема Маркова утверждает, что ЗБЧ выполнен, если дисперсия суммы n слагаемых растет не слишком быстро с ростом n.

Сильная зависимость слагаемых приводит обычно к невыполнению ЗБЧ. Если, например, $\mathsf{D}\,\xi_1 \neq 0$ и $\xi_n \equiv \xi_1$, то $S_n = n\xi_1$, и свойство (4) не выполнено. В этом случае не выполнено и достаточное условие для ЗБЧ: $\mathsf{D}\,S_n = \mathsf{D}\,(n\xi_1) = cn^2$. Для одинаково распределённых слагаемых дисперсия суммы ещё быстрее расти уже не может.

Теорема 37 (ЗБЧ в форме Хинчина). Для любой последовательности $\xi_1,\, \xi_2,\, \ldots$ независимых в совокупности и одинаково распределённых с.в. с конечным первым моментом $\mathsf{E}\,|\xi_1| < \infty$ имеет место сходимость по вероятности:

$$\frac{\xi_1 + \ldots + \xi_n}{n} \xrightarrow{\text{p}} \mathsf{E}\,\xi_1.$$

Это утверждение будет доказано позже. Отметим сейчас только то, что в отличие от теоремы Чебышева здесь не требуется существования вторых моментов, но в теореме Хинчина с.в.одинаково распределены.

Итак, чтобы последовательность независимых и одинаково распределённых случайных величин удовлетворяла ЗБЧ, достаточно существования первого момента слагаемых. Более того, в условиях теоремы 37 имеет место и сходимость п.н. последовательности $(\xi_1 + \ldots + \xi_n)/n$ к $\mathsf{E}\,\xi_1$. Это утверждение называется усиленным законом больших чисел (УЗБЧ) Колмогорова. Доказывать УЗБЧ не будем.

Получим в качестве следствия из ЗБЧ Чебышёва ЗБЧ Бернулли. В отличие от ЗБЧ Чебышёва, описывающего предельное поведение среднего арифметического случайных величин с произвольными распределениями, ЗБЧ Бернулли - утверждение только для схемы Бернулли.

Теорема 38 (ЗБЧ в форме Бернулли). Пусть событие A (*успех*) может произойти в любом из n независимых испытаний с одной и той же вероятностью p, и пусть $\nu_n(A)$ - число появлений события A в n испытаниях. Тогда $\frac{\nu_n(A)}{n} \xrightarrow{\mathrm{p}} p$. При этом $\forall\, \varepsilon > 0$

$$\mathsf{P}\left(\left|\frac{\nu_n(A)}{n} - p\right| \geq \varepsilon\right) \leq \frac{p(1-p)}{n\varepsilon^2}.$$

Доказательство. Заметим, что $\nu_n(A)$ равна сумме независимых, одинаково распределённых с.в., имеющих распределение Бернулли с параметром $p = \mathsf{P}(A)$ (индикаторов того, что в соответствующем испытании произошло A): $\nu_n(A) = \xi_1 + \ldots + \xi_n$, где

$$\xi_i = \begin{cases} 1, & \text{если } A \text{ произошло в } i\text{-м испытании;} \\ 0, & \text{если } A \text{ не произошло в } i\text{-м испытании;} \end{cases}$$

и $\mathsf{E}\,\xi_1 = \mathsf{P}(A) = p,\ \mathsf{D}\,\xi_1 = p(1-p)$. Осталось воспользоваться ЗБЧ в форме Чебышёва и неравенством (5).

Пример 45. Правильная монета подбрасывается 10^4 раз. Оценим вероятность того, что частота выпадения герба отличается от $1/2$ на 0.01 или более.

Пусть $\xi_1, \ldots, \xi_n$ - независимые с.в., каждая из которых имеет распределение Бернулли с параметром $p = 1/2$ и равна единице, если при соответствующем подбрасывании выпал герб, и нулю иначе. Нужно оценить $\mathsf{P}\left(\left|\frac{\nu_n}{n} - \frac{1}{2}\right| \geq 0{,}01\right)$, где $n = 10^4$, а $\nu_n = \sum\limits_{i=1}^{n} \xi_i$ - число выпадений герба. Поскольку $\mathsf{D}\,\xi_1 = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$, искомая оценка сверху выглядит так:

$$\mathsf{P}\left(\left|\frac{\nu_n}{n} - \frac{1}{2}\right| \geq 0{,}01\right) \leq \frac{\mathsf{D}\,\xi_1}{n \cdot 0{,}01^2} = \frac{1}{4 \cdot 10^4 \cdot 10^{-4}} = \frac{1}{4}.$$

Итак, неравенство Чебышёва позволяет заключить, что в среднем не более чем в четверти случаев при 10000 подбрасываниях монеты частота выпадения герба будет отличаться от $1/2$ на одну сотую или больше. Мы увидим ниже, насколько это грубая оценка.

§4. Слабая сходимость

Пусть задана последовательность случайных величин $\xi_1, \xi_2, \ldots$ с заданными функциями распределения $F_{\xi_i}(x)$ и пусть ξ - некоторая с.в., имеющая функцию распределения $F_\xi(x)$.

Определение 60. Последовательность с.в. $\xi_1, \xi_2, \ldots$ сходится *слабо* или сходится *по распределению* к случайной величине ξ и пишут: $\xi_n \Rightarrow \xi$, если для любого x такого, что функция распределения F_ξ непрерывна в точке x, имеет место сходимость $F_{\xi_n}(x) \to F_\xi(x)$ при $n \to \infty$ во всех точках непрерывности предельной функции распределения.

Замечание. Сходимость $\xi_n \Rightarrow \xi$ есть сходимость распределений, а не случайных величин: если *предельную* величину ξ заменить на другую величину ζ с тем же распределением, ничего не изменится: в том же смысле $\xi_n \Rightarrow \zeta$.

Теорема 39 (Свойство *(i)* слабой сходимости). Если $\xi_n \Rightarrow \xi$, и функция распределения F_ξ непрерывна в точках a и b, то $\mathsf{P}(\xi_n \in (a, b)) \to \mathsf{P}(\xi \in (a, b))$. Если во всех точках a и b непрерывности функции распределения F_ξ имеет место сходимость $\mathsf{P}(\xi_n \in (a, b)) \to \mathsf{P}(\xi \in (a, b))$, то $\xi_n \Rightarrow \xi$.

Доказательство очевидным образом следует из определения и свойств функций распределения.

В последней теореме вместо открытого интервала (a, b) можно взять полуоткрытые интервалы или отрезок.

Теорема 40 (Свойства слабой сходимости *(ii)*, *(iii)*).
(ii) Если $\xi_n \xrightarrow{\mathrm{p}} \xi$, то $\xi_n \Rightarrow \xi$.
(iii) Если $\xi_n \Rightarrow c$, то $\xi_n \xrightarrow{\mathrm{p}} c$.

Доказательство. Свойство *(ii)* будет доказано позже.

Докажем *(iii)*. Пусть $\xi_n \Rightarrow c$, т.е.

$$F_{\xi_n}(x) \to F_c(x) = \begin{cases} 0, & x \le c; \\ 1, & x > c \end{cases}$$

при любом x, являющимся точкой непрерывности предельной функции $F_c(x)$, т.е. при всех $x \ne c$. Возьмём произвольное $\varepsilon > 0$ и докажем, что $\mathsf{P}(|\xi_n - c| < \varepsilon) \to 1$.

$$\mathsf{P}(-\varepsilon < \xi_n - c < \varepsilon) = \mathsf{P}(c - \varepsilon < \xi_n < c + \varepsilon) \ge$$

$$\ge \mathsf{P}(c - \varepsilon/2 \le \xi_n < c + \varepsilon) =$$

$$= F_{\xi_n}(c+\varepsilon) - F_{\xi_n}(c-\varepsilon/2) \to F_c(c+\varepsilon) - F_c(c-\varepsilon/2) = 1 - 0 = 1,$$

поскольку в точках $c+\varepsilon$ и $c-\varepsilon/2$ функция F_c непрерывна, и, следовательно, имеет место сходимость последовательностей $F_{\xi_n}(c+\varepsilon)$ к $F_c(c+\varepsilon) = 1$ и $F_{\xi_n}(c-\varepsilon/2)$ к $F_c(c-\varepsilon/2) = 0$. Осталось заметить, что $\mathsf{P}(|\xi_n - c| < \varepsilon)$ не может быть больше 1, так что по свойству предела зажатой последовательности $\mathsf{P}(|\xi_n - c| < \varepsilon) \to 1$.

Замечание. Свойство *предел суммы равен сумме пределов* для слабой сходимости бессмысленно: сходимости $\xi_n \Rightarrow \xi$, $\zeta_n \Rightarrow \zeta$ означают, что известны предельные одномерные распределения этих последовательностей. Но предельное распределение их суммы может быть различным в зависимости от совместного распределения ξ_n и ζ_n. Другое дело, когда одно из предельных распределений вырождено. Тогда предельная функция распределения суммы или произведения определена однозначно.

Теорема 41 (Свойства слабой сходимости *(iv)*,(v)).
(iv) Если $\xi_n \xrightarrow{p} c$ и $\zeta_n \Rightarrow \zeta$, то $\xi_n \cdot \zeta_n \Rightarrow c\zeta$.
(v) Если $\xi_n \xrightarrow{p} c$ и $\zeta_n \Rightarrow \zeta$, то $\xi_n + \zeta_n \Rightarrow c + \zeta$.

Доказательство. Заметим вначале, что если $\zeta_n \Rightarrow \zeta$, то $c\zeta_n \Rightarrow c\zeta$ и $c + \zeta_n \Rightarrow c + \zeta$, что просто доказать. Поэтому, достаточно доказать *(iv)* при $c = 1$, а *(v)* - при $c = 0$.

Рассмотрим *(v)*, оставив *(iv)* для самостоятельного доказательства. Пусть $\xi_n \xrightarrow{p} 0$ и $\zeta_n \Rightarrow \zeta$. Докажем, что тогда $\xi_n + \zeta_n \Rightarrow \zeta$.

Пусть x_0 - точка непрерывности функции распределения $F_\zeta(x)$. Требуется доказать, что имеет место сходимость $F_{\xi_n+\zeta_n}(x_0)$ к $F_\zeta(x_0)$. Зафиксируем $\varepsilon > 0$ такое, что $F_\zeta(x)$ непрерывна в точках $x_0 \pm \varepsilon$.

События $H_1 = \{|\xi_n| \geq \varepsilon\}$ и $H_2 = \{|\xi_n| < \varepsilon\}$ образуют полную группу, поэтому

$$F_{\xi_n+\zeta_n}(x_0) = \mathsf{P}(\xi_n + \zeta_n < x_0,\ H_1) + \mathsf{P}(\xi_n + \zeta_n < x_0,\ H_2) = P_1 + P_2.$$

Оценим $P_1 + P_2$ сверху и снизу. Для P_1 имеем

$$0 \leq P_1 = \mathsf{P}(\xi_n + \zeta_n < x_0,\ H_1) \leq \mathsf{P}(H_1) = \mathsf{P}(|\xi_n| \geq \varepsilon),$$

и последняя вероятность может быть выбором n сделана сколь угодно малой. Для P_2, с одной стороны,

$$P_2 = \mathsf{P}(\xi_n + \zeta_n < x_0,\ -\varepsilon < \xi_n < \varepsilon) \leq$$

$$\leq \mathsf{P}(-\varepsilon + \zeta_n < x_0) = F_{\zeta_n}(x_0 + \varepsilon).$$

Выше воспользовались тем, что если $-\varepsilon < \xi_n$ и $\xi_n + \zeta_n < x_0$, то тем более $-\varepsilon + \zeta_n < x_0$. С другой стороны,

$$P_2 = \mathsf{P}(\xi_n + \zeta_n < x_0,\ -\varepsilon < \xi_n < \varepsilon) \geq$$

$$\geq \mathsf{P}(\varepsilon + \zeta_n < x_0,\ -\varepsilon < \xi_n < \varepsilon) \geq$$

$$\geq \mathsf{P}(\varepsilon + \zeta_n < x_0) - \mathsf{P}(|\xi_n| \geq \varepsilon) = F_{\zeta_n}(x_0 - \varepsilon) - \mathsf{P}(|\xi_n| \geq \varepsilon).$$

Здесь первое неравенство объясняется включением

$$\{\varepsilon + \zeta_n < x_0,\ -\varepsilon < \xi_n < \varepsilon\} \subseteq \{\xi_n + \zeta_n < x_0,\ -\varepsilon < \xi_n < \varepsilon\},$$

которое получилось заменой в событии $\{\varepsilon + \zeta_n < x_0\}$ числа ε на меньшую величину $\xi_n,\ \xi_n < \varepsilon$. Второе неравенство следует из свойств:

$$\mathsf{P}(A\overline{B}) \leq \mathsf{P}(\overline{B}),$$

$$\text{поэтому}\ \ \mathsf{P}(AB) = \mathsf{P}(A) - \mathsf{P}(A\overline{B}) \geq \mathsf{P}(A) - \mathsf{P}(\overline{B}).$$

Получена оценка снизу и сверху для $P_1 + P_2$, т.е. для $F_{\xi_n+\zeta_n}(x_0)$:

$$F_{\zeta_n}(x_0 - \varepsilon) - \mathsf{P}(|\xi_n| \geq \varepsilon) \leq F_{\xi_n+\zeta_n}(x_0) \leq \mathsf{P}(|\xi_n| \geq \varepsilon) + F_{\zeta_n}(x_0 + \varepsilon).$$

Устремляя n к бесконечности и вспоминая, что $x_0 \pm \varepsilon$ - точки непрерывности функции распределения F_ζ, получаем

$$F_\eta(x_0 - \varepsilon) \leq \underline{\lim}\, F_{\xi_n+\eta_n}(x_0) \leq \overline{\lim}\, F_{\xi_n+\eta_n}(x_0) \leq F_\eta(x_0 + \varepsilon). \qquad (6)$$

У любой функции распределения не более чем счётное множество точек разрыва. Поэтому можно выбрать такую уменьшающуюся до нуля последовательность ε, что в точках $x_0 \pm \varepsilon$ функция распределения F_ζ будет непрерывной и, следовательно, останутся верны неравенства (6). Переходя к пределу по такой последовательности $\varepsilon \to 0$ и помня, что x_0 - точка непрерывности функции F_ζ, получаем, что нижний и верхний пределы $F_{\xi_n+\zeta_n}(x_0)$ при $n \to \infty$ совпадают и равны $F_\zeta(x_0)$.

В качестве простого следствия из только что доказанного свойства *(v)* покажем, что сходимость $\xi_n \xrightarrow{\mathrm{p}} \xi$ по вероятности влечёт слабую сходимость $\xi_n \Rightarrow \xi$, а это есть свойство *(ii)*, доказательство которого выше было отложено.

Представим ξ_n в виде суммы $\xi_n = (\xi_n - \xi) + \xi$. Здесь последовательность $\xi_n - \xi$ по вероятности стремится к нулю, а «последовательность» ξ слабо сходится к ξ. Поэтому их сумма слабо сходится к ξ, что и требовалось доказать.

Асимптотический анализ распределений сумм и слабую сходимость к предельному распределению удобно изучать с помощью центральной предельной теоремы, к которой переходим.

Глава 13. Центральная предельная теорема

§1. ЦПТ в форме Ляпунова

Следующую теорему будем рассматривать для простоты понимания в частном случае для последовательностей независимых и одинаково распределённых случайных величин. Обозначим $S_n = \xi_1 + \ldots + \xi_n$ сумму первых n с.в. в последовательности.

Теорема 42 (ЦПТ в форме Ляпунова). Пусть $\xi_1,\ \xi_2,\ \ldots$ - независимые и одинаково распределённые невырожденные с.в. с конечной дисперсией: $0 < \mathsf{D}\,\xi_1 < \infty$. Тогда имеет место следующая слабая сходимость

$$\frac{S_n - n\,\mathsf{E}\,\xi_1}{\sqrt{n\,\mathsf{D}\,\xi_1}} \Rightarrow\ \zeta\ \sim\ \mathrm{N}(0,1)$$

последовательности центрированных и нормированных сумм с.в. к стандартному нормальному распределению.

Используя определение и свойства слабой сходимости и замечая, что функция распределения $\Phi_{a,\sigma^2}(x)$ любого нормального закона непрерывна всюду на R^1, утверждение ЦПТ Ляпунова можно сформулировать любым из следующих способов.

Следствие 12. Пусть $\xi_1, \xi_2, \ldots$ - независимые и одинаково распределённые с.в. с конечной и ненулевой дисперсией. Тогда выполнены утверждения: $\forall\ x,\ y \in R^1 :\ x < y$ при $n \to \infty$ имеет место сходимость

$$\mathsf{P}\left(x \leq \frac{S_n - n\,\mathsf{E}\,\xi_1}{\sqrt{n\,\mathsf{D}\,\xi_1}} \leq y\right) \to \Phi_{0,1}(y) - \Phi_{0,1}(x) = \int\limits_x^y \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\,du;$$

если ζ - произвольная случайная величина со стандартным нормальным распределением, то

$$\sqrt{n}\ \left(\frac{S_n}{n} - \mathsf{E}\,\xi_1\right) = \frac{S_n - n\,\mathsf{E}\,\xi_1}{\sqrt{n}} \Rightarrow \sqrt{\mathsf{D}\,\xi_1}\cdot\zeta \sim \mathrm{N}(0, \mathsf{D}\,\xi_1).$$

Докажем центральную предельную теорему и закон больших чисел в форме Хинчина в следующем параграфе. Для этого потребуется понятие и свойства характеристических функций.

Получим в качестве следствия из ЦПТ Ляпунова предельную теорему Муавра–Лапласа. Как и ЗБЧ Бернулли, это утверждение годится только для схемы Бернулли.

Теорема 43 (предельная теорема Муавра–Лапласа). Пусть событие A может произойти в любом из n независимых испытаний с одной и той же вероятностью p и пусть $\nu_n(A)$ - число появлений события A в n испытаниях. Тогда

$$\frac{\nu_n(A) - np}{\sqrt{np\,(1-p)}}\ \Rightarrow\ \zeta\ \sim\ \mathrm{N}(0,1)\ \text{ при }\ n \to \infty,$$

т.е. для любых вещественных $x < y$ имеет место сходимость

$$\mathsf{P}\left(x \le \frac{\nu_n(A) - np}{\sqrt{np\,(1-p)}} \le y\right) \to$$

$$\to \Phi_{0,1}(y) - \Phi_{0,1}(x) = \int\limits_x^y \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\,du\,.$$

Доказательство. С.в. $\nu_n(A)$ равна сумме независимых, одинаково распределённых с.в., имеющих распределение Бернулли с параметром, равным вероятности успеха p : $\nu_n(A) = \xi_1 + \ldots + \xi_n$, где $\mathsf{E}\,\xi_1 = p,\ \ \mathsf{D}\,\xi_1 = p(1-p)$. Осталось воспользоваться теоремой 42.

В завершении этого параграфа продемонстрируем применение ЦПТ на простом примере.

Пример 46. Вернёмся к примеру 45. Требуется найти

$$\mathsf{P}\left(\left|\frac{\nu_n}{n} - \frac{1}{2}\right| \ge 0.01\right),$$

где $n = 10\,000$, ν_n - число выпадений герба. Вычислим вероятность дополнительного события. Домножим обе части неравенства под знаком вероятности на $\sqrt{n} = 100$ и поделим на $\sqrt{p\,(1-p)} = 1/2$.

$$\mathsf{P}\left(\left|\frac{\nu_n}{n} - \frac{1}{2}\right| < 0{,}01\right) =$$

$$= \mathsf{P}\left(\frac{\sqrt{n}}{\sqrt{p\,(1-p)}}\left|\frac{\nu_n}{n} - p\right| < 0{,}01\frac{\sqrt{n}}{\sqrt{p\,(1-p)}}\right) =$$

$$= \mathsf{P}\left(\frac{\sqrt{n}}{\sqrt{p\,(1-p)}}\left|\frac{\nu_n}{n} - p\right| < 2\right) =$$

$$= \mathsf{P}\left(-2 < \frac{\nu_n - np}{\sqrt{np\,(1-p)}} < 2\right) \approx$$

$$\approx \Phi_{0,1}(2) - \Phi_{0,1}(-2) = 1 - 2\Phi_{0,1}(-2) = 1 - 2\cdot 0{,}0228 = 1 - 0{,}0456.$$

Искомая вероятность примерно равна 0.0456:

$$\mathsf{P}\left(\left|\frac{\nu_n}{n} - \frac{1}{2}\right| \ge 0{,}01\right) = 1 - \mathsf{P}\left(\left|\frac{\nu_n}{n} - \frac{1}{2}\right| < 0{,}01\right) \approx 0{,}0456.$$

Обычно центральной предельной теоремой пользуются для приближённого вычисления вероятностей, связанных с суммами большого числа независимых и одинаково распределённых величин. При этом распределение центрированной и нормированной суммы заменяют на стандартное нормальное распределение. Насколько велика такая погрешность приближения?

В примере 46 найдена вероятность приближённо. Следующая теорема позволяет оценить погрешность приближения в ЦПТ.

Теорема 44 (неравенство Берри - Эссеена). В условиях ЦПТ $\forall\, x \in R^1$ и для любого распределения ξ_1 с конечным третьим моментом

$$\left|\mathsf{P}\left(\frac{S_n - n\mathsf{E}\,\xi_1}{\sqrt{n\,\mathsf{D}\,\xi_1}} < x\right) - \Phi_{0,1}(x)\right| \le C \cdot \frac{\mathsf{E}\,|\xi_1 - \mathsf{E}\,\xi_1|^3}{\sqrt{n}\big(\sqrt{\mathsf{D}\,\xi_1}\,\big)^3}.$$

В качестве константы C можно взять, например, число 0.4.

В связи с большой сложностью доказательство этой теоремы опустим.

Продолжение примера 46. Легко проверить, что для с.в. ξ_1 с распределением Бернулли

$$\mathsf{E}\,|\xi_1 - \mathsf{E}\,\xi_1|^3 = |0-p|^3\,\mathsf{P}(\xi_1 = 0) +$$
$$+ |1-p\,|^3\,\mathsf{P}(\xi_1 = 1) = p\,q(p^2+q^2).$$

Поэтому разница между левой и правой частями приближённого равенства в примере 46 при $n = 10^4$ и $p = q = \frac{1}{2}$ не превышает величины

$$C \cdot \frac{p\,q(p^2+q^2)}{\sqrt{npq}\big(\sqrt{pq}\big)^3} = C \cdot \frac{p^2+q^2}{\sqrt{n}\,\sqrt{pq}} \le 0{,}4 \cdot \frac{1}{100} = 0{,}004,$$

т.е. искомая вероятность $\mathsf{P}\left(\left|\frac{\nu_n}{n} - \frac{1}{2}\right| > 0{,}01\right)$ не больше, чем $0.0456 + 0.004$. Интересно сравнить этот результат с грубой оценкой $\frac{1}{4}$, полученной с помощью ЗБЧ в примере 45.

§2. Характеристические функции

В этом параграфе рассматривается функция, которая используется в аналитических методах теории вероятностей, применяемых в доказательствах утверждений, связанных с суммами случайных величин.

Всюду в этом параграфе i обозначает мнимую единицу, t - действительную переменную. Из теории функций комплексного переменного известна формула Эйлера $e^{it} = \cos t + i\sin t$. Из свойства линейности математического ожидания следует $\mathsf{E}\,(\xi + i\zeta) = \mathsf{E}\,\xi + i\,\mathsf{E}\,\zeta$ - математическое ожидание комплекснозначной с.в. $\xi + i\zeta$, если математические ожидания её действительной ξ и мнимой ζ частей существуют.

Напомним, что модулем комплексного числа $z = x + iy$ называется неотрицательное число $|z| = \sqrt{x^2+y^2}$, так что $\left|e^{it}\right| = 1$.

Определение 61. Функция $f_\xi(t) = \mathsf{E}\, e^{it\xi}$ действительной переменной t называется *характеристической функцией* случайной величины ξ.

Пример 47. Пусть с.в. ξ имеет распределение Бернулли с параметром p. Её характеристическая функция равна

$$f_\xi(t) = \mathsf{E}\, e^{it\xi} = e^{it\cdot 0}\ \mathsf{P}(\xi = 0) + e^{it\cdot 1}\ \mathsf{P}(\xi = 1) = 1 - p + pe^{it}.$$

Пример 48. Пусть с.в. ξ имеет биномиальное распределение с параметрами n и p. Её характеристическая функция равна

$$f_\xi(t) = \mathsf{E}\, e^{it\xi} = \sum_{k=0}^{n} e^{it\cdot k}\ \mathsf{P}(\xi = k) = \sum_{k=0}^{n} e^{it\cdot k}\, C_n^k\, p^k\, (1-p)^{n-k} =$$

$$= \sum_{k=0}^{n} C_n^k \left(pe^{it}\right)^k (1-p)^{n-k} = \left(1 - p + pe^{it}\right)^n,$$

где в последнем равенстве используется формула бинома Ньютона.

Пример 49. Пусть с.в. ξ имеет распределение Пуассона с параметром λ. Её характеристическая функция равна

$$f_\xi(t) = \mathsf{E}\, e^{it\xi} = \sum_{k=0}^{\infty} e^{it\cdot k}\ \mathsf{P}(\xi = k) = \sum_{k=0}^{\infty} e^{it\cdot k}\, \frac{\lambda^k}{k!}\, e^{-\lambda} =$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\left(\lambda e^{it}\right)^k}{k!} = e^{-\lambda} e^{\lambda e^{it}} = \exp\{\lambda \left(e^{it} - 1\right)\}.$$

Пример 50. Пусть с.в. ξ имеет гамма-распределение с параметрами α и λ. Её характеристическая функция равна

$$f_\xi(t) = \mathsf{E}\, e^{it\xi} = \int\limits_0^\infty e^{it\cdot x}\, f_\xi(x)\, dx = \int\limits_0^\infty e^{itx}\, \frac{\alpha^\lambda}{\Gamma(\lambda)}\, x^{\lambda-1}\, e^{-\alpha x}\, dx =$$

$$= \frac{\alpha^\lambda}{\Gamma(\lambda)} \int\limits_0^\infty x^{\lambda-1}\, e^{-x(\alpha - it)}\, dx = \left(\frac{\alpha}{\alpha - it}\right)^\lambda = \left(1 - \frac{it}{\alpha}\right)^{-\lambda}.$$

Интеграл вычисляется с помощью гамма-функции: замена $y = x(\alpha - it)$ даёт

$$\int\limits_0^\infty x^{\lambda-1}\, e^{-x(\alpha-it)}\, dx = \frac{1}{(\alpha - it)^\lambda} \int\limits_0^\infty y^{\lambda-1}\, e^{-y}\, dy = \frac{\Gamma(\lambda)}{(\alpha - it)^\lambda}.$$

В качестве следствия, очевидно, получим, что для с.в. ξ с показательным распределением

$\mathbf{E}(\alpha) = \Gamma(\alpha, 1)$ характеристическая функция равна $f_\xi(t) = \frac{\alpha}{\alpha - it}$.

Пример 51. Пусть с.в. ξ имеет стандартное нормальное распределение. Её характеристическая функция равна

$$f_\xi(t) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} e^{-t^2/2}\, e^{-(x-it)^2/2}\, dx =$$

$$= e^{-t^2/2} \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} e^{-(x-it)^2/2}\, d(x-it) \;=\; e^{-t^2/2}.$$

При интегрировании выделялся полный квадрат в показателе экспоненты и использовалось то, что интеграл по всей прямой R^1 от функции $\exp\left(-u^2/2\right)/\sqrt{2\pi}$ равен единице.

§3. Свойства характеристических функций

(i). Характеристическая функция всегда существует и иммет место неравенство

$$|f_\xi(t)| = \left|\mathsf{E}\, e^{it\xi}\right| \leq 1.$$

Заметим, что $\mathsf{E}\,\xi$ существует не всегда.

Доказательство. Воспользуемся свойством $\mathsf{D}\,\zeta \geq 0$, равносильным неравенству $\left(\mathsf{E}\,\zeta\right)^2 \leq \mathsf{E}\,\zeta^2$:

$$|f_\xi(t)|^2 = \left|\mathsf{E}\,\cos(t\xi) + i\mathsf{E}\,\sin(t\xi)\right|^2 = \left(\mathsf{E}\,\cos(t\xi)\right)^2 + \left(\mathsf{E}\,\sin(t\xi)\right)^2 \leq$$

$$\leq \mathsf{E}\,\cos^2(t\xi) + \mathsf{E}\,\sin^2(t\xi) = \mathsf{E}\left(\cos^2(t\xi) + \sin^2(t\xi)\right) = \mathsf{E}\,1 = 1.$$

(ii). По характеристической функции однозначно восстанавливается распределение (функция распределения, плотность или таблица распределения). Другими словами, если две случайные величины имеют одинаковые характеристические функции, то и распределения этих величин совпадают.

Формулы, с помощью которых по характеристической функции восстанавливается распределение, в теории вероятностей называют формулами обращения. Если модуль характеристической функции интегрируем на всей прямой, то у с.в. с этой характеристической функцией имеется плотность распределения и она находится по формуле

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

Отметим, что в этой формуле для характеристической функции и плотности используется одна и та же буква f для обозначения по сути двух разных функций. Здесь используем правило, которым будем пользоваться и далее: если аргумент функции обозначается буквой t, то эта функция обозначает характеристическую функцию, если для аргумента функции используется любая другая буква - то это плотность. Например, характеристическая функция для абсолютно непрерывного распределения с плотностью $f_\xi(x)$ равна

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

(iii). Характеристическая функция случайной величины $a + b\xi$ связана с характеристической функцией случайной величины ξ равенством

$$f_{a+b\xi}(t) = \mathsf{E}\, e^{it(a+b\xi)} = e^{ita}\mathsf{E}\, e^{i(tb)\xi} = e^{ita} f_\xi(tb).$$

Пример 52. Найдём характеристическую функцию случайной величины $\xi \sim N(a, \sigma^2)$. Воспользуемся примером 51, в котором найдена характеристическая функция стандартного нормального распределения. Ясно, что характеристическая функция $\zeta = (\xi - a)/\sigma$ равна $f_\zeta(t) = e^{-t^2/2}$. Тогда характеристическая функция с.в. $\xi = a + \sigma\zeta$ равна по свойству *(iii)*

$$f_\xi(t) = f_{a+\sigma\zeta}(t) = e^{ita} f_\zeta(t\sigma) = \exp\left(ita - \frac{t^2\sigma^2}{2}\right).$$

(iv). Характеристическая функция суммы независимых с.в. равна произведению характеристических функций слагаемых: если с.в. ξ и ζ независимы, то, по свойству *(vii)* математических ожиданий,

$$f_{\xi+\zeta}(t) = \mathsf{E}\, e^{it(\xi+\zeta)} = \mathsf{E}\, e^{it\xi}\, \mathsf{E}\, e^{it\zeta} = f_\xi(t)\, f_\zeta(t).$$

Замечание. Для того, чтобы характеристическая функция суммы n случайных величин записывалась в виде произведения их характеристических функций, попарной независимости слагаемых не хватает. То же самое можно сказать про свойство *(vii)* математических ожиданий. Если же с.в. независимы в совокупности, то их совместное распределение записывается в виде произведения распределений, и тогда математическое ожидание произведения можно записать в виде произведения математических ожиданий.

Пример 53. Воспользуемся свойством *(iv)* для доказательства воспроизводимости нормального распределения по двум параметрам. Пусть с.в. $\xi \sim \mathrm{N}(a_1, \sigma_1^2)$ и

$\zeta \sim \mathrm{N}(a_2,\, \sigma_2^2)$ независимы. Характеристическая функция суммы $\xi + \zeta$ равна

$$f_{\xi+\zeta}(t) = f_\xi(t)\, f_\zeta(t) = \exp\left(ita_1 - \frac{t^2\sigma_1^2}{2}\right) \cdot \exp\left(ita_2 - \frac{t^2\sigma_2^2}{2}\right) =$$

$$= \exp\left(it(a_1 + a_2) - \frac{t^2(\sigma_1^2 + \sigma_2^2)}{2}\right).$$

Видим, что характеристическая функция суммы есть характеристическая функция нормального распределения с параметрами $a_1 + a_2$ и $\sigma_1^2 + \sigma_2^2$. Следовательно, $\xi + \zeta \sim \mathrm{N}(a_1 + a_2,\, \sigma_1^2 + \sigma_2^2)$ по свойству *(ii)*.

Пример 54. Выведим свойство воспроизводимости распределения Пуассона по параметру. Для независимых с.в. с распределениями Пуассона $\Pi(\lambda)$ и $\Pi(\mu)$ характеристическая функция суммы

$$f_{\xi+\eta}(t) = f_\xi(t)\, f_\eta(t) = \exp\left\{\lambda\left(e^{it} - 1\right)\right\} \exp\left\{\mu\left(e^{it} - 1\right)\right\} =$$

$$= \exp\left\{(\lambda + \mu)\left(e^{it} - 1\right)\right\}$$

равна характеристической функции распределения Пуассона с параметром $\Pi_{\lambda+\mu}$.

Пример 55. Выведим свойство воспроизводимости биномиального распределения по первому параметру. Для независимых с.в. с биномиальными распределениями $\mathrm{B}(n, p)$ и $\mathrm{B}(m, p)$ характеристическая функция суммы

$$f_{\xi+\eta}(t) = f_\xi(t)\, f_\eta(t) = \left(1 - p + pe^{it}\right)^n \left(1 - p + pe^{it}\right)^m =$$

$$= \left(1 - p + pe^{it}\right)^{n+m}$$

равна характеристической функции биномиального распределения с параметрами $n+m$ и p.

Пример 56. Выведим свойство воспроизводимости гамма-распределения по второму параметру. Для независимых с.в. с гамма-распределениями $\Gamma(\alpha,\, \lambda_1)$ и $\Gamma(\alpha,\, \lambda_2)$ характеристическая функция суммы

$$f_{\xi+\eta}(t) = \left(1 - \frac{it}{\alpha}\right)^{-\lambda_1} \left(1 - \frac{it}{\alpha}\right)^{-\lambda_2} = \left(1 - \frac{it}{\alpha}\right)^{-(\lambda_1+\lambda_2)}$$

равна характеристической функции гамма-распределения $\Gamma(\alpha,\, \lambda_1 + \lambda_2)$.

(v). Пусть существует момент целого положительного порядка k с.в. ξ, т.е. $\mathsf{E}\,|\xi|^k < \infty$. Тогда характеристическая функция $f_\xi(t)$ непрерывно дифференцируема k раз и её

k-я производная в нуле связана с моментом порядка k равенством

$$f_\xi^{(k)}(0) = \left(\frac{d^k}{d\,t^k}\,\mathsf{E}\,e^{it\xi}\right)\Big|_{t=0} = \left(\mathsf{E}\,i^k\,\xi^k\,e^{it\xi}\right)\Big|_{t=0} = i^k\,\mathsf{E}\,\xi^k.$$

Не будем доказывать существование и непрерывность k-й производной, а также законность переноса производной под знак математического ожидания. Примем этот факт без доказательства.

Используя последнее свойство, легко проверить, что для случайной величины ξ со стандартным нормальным распределением момент чётного порядка $2k$ равен

$$\mathsf{E}\,\xi^{2k} = (2k-1)!! \;=\; (2k-1)\cdot(2k-3)\cdot\ldots\cdot 3\cdot 1,$$

а все моменты нечётных порядков того же стандартного нормального распределения существуют и равны нулю.

(vi). Пусть существует момент целого положительного порядка k с.в. ξ, т.е. $\mathsf{E}\,|\xi|^k < \infty$. Тогда характеристическая функция $f_\xi(t)$ в окрестности точки $t = 0$ разлагается по формуле Тейлора

$$f_\xi(t) = f_\xi(0) + \sum_{j=1}^{k}\frac{t^j}{j!}\,f_\xi^{(j)}(0) + o(|t^k|) = 1 + \sum_{j=1}^{k}\frac{i^j t^j}{j!}\,\mathsf{E}\,\xi^j + o(|t^k|) =$$

$$= 1 + it\,\mathsf{E}\,\xi - \frac{t^2}{2}\,\mathsf{E}\,\xi^2 + \ldots + \frac{i^k t^k}{k!}\,\mathsf{E}\,\xi^k + o(|t^k|).$$

Теорема 43 (теорема Леви о непрерывном соответствии). С.в. ξ_n слабо сходятся к случайной величине ξ тогда и только тогда, когда для любого t характеристические функции $f_{\xi_n}(t)$ сходятся к характеристической функции $f_\xi(t)$.

Эту теорему оставляем без доказательства.

Эта теорема устанавливает непрерывное соответствие между классами $\left\{F_\xi \;\Rightarrow\; \cdot\,\right\}$ функций распределения со слабой сходимостью и $\left\{f_\xi \;\to\; \cdot\,\right\}$ характеристических функций со сходимостью в каждой точке. «Непрерывность» этого соответствия состоит в том, что пределу в одном классе относительно заданной в этом классе сходимости соответствует предел в другом классе относительно сходимости, заданной в этом другом классе.

Завершим этот параграф доказательством двух теорем, сформулированных выше.

Доказательство ЗБЧ Хинчина Пусть $\xi_1, \xi_2, \ldots$ - последовательность независимых в совокупности и одинаково распределённых с.в. с конечным первым моментом $\mathsf{E}\,|\xi_1| < \infty$. Обозначим через a математическое ожидание $\mathsf{E}\,\xi_1$. Требуется доказать, что

$$\frac{S_n}{n} = \frac{\xi_1 + \ldots + \xi_n}{n} \xrightarrow{\text{p}} a.$$

По свойству, рассмотренному выше, сходимость по вероятности к константе эквивалентна слабой сходимости. Так как a - константа, то достаточно доказать слабую сходимость $\frac{S_n}{n}$ к a. По теореме о непрерывном соответствии, эта сходимость имеет место тогда и только тогда, когда $\forall\, t \in R^1$ сходятся характеристические функции

$$f_{S_n/n}(t) \to f_a(t) = \mathsf{E}\, e^{ita} = e^{ita}.$$

Найдем характеристическую функцию с.в. $\frac{S_n}{n}$. Пользуясь свойствами (iii) и (iv), получаем

$$f_{S_n/n}(t) = f_{S_n}\Big(\frac{t}{n}\Big) = \Big(f_{\xi_1}\Big(\frac{t}{n}\Big)\Big)^n.$$

Поскольку первый момент ξ_1 существует, то воспользовавшись свойством (vi) разложим $f_{\xi_1}(t)$ по формуле Тейлора в окрестности нуля:

$$f_{\xi_1}(t) = 1 + it\mathsf{E}\,\xi_1 + o(|t|) = 1 + ita + o(|t|).$$

В точке t/n соответственно

$$f_{\xi_1}\Big(\frac{t}{n}\Big) = 1 + \frac{ita}{n} + o\Big(\Big|\frac{t}{n}\Big|\Big), f_{S_n/n}(t) = \Big(f_{\xi_1}\Big(\frac{t}{n}\Big)\Big)^n =$$

$$= \Big(1 + \frac{ita}{n} + o\Big(\Big|\frac{t}{n}\Big|\Big)\Big)^n.$$

При $n \to \infty$, пользуясь «замечательным пределом» $\big(1 + \frac{x}{n}\big)^n \to e^x$, получаем

$$f_{S_n/n}(t) = \Big(1 + \frac{ita}{n} + o\Big(\Big|\frac{t}{n}\Big|\Big)\Big)^n \to e^{ita},$$

что и требовалось доказать.

Доказательство центральной предельной теоремы Пусть $\xi_1, \xi_2, \ldots$ - последовательность независимых в совокупности и одинаково распределённых с.в. с конечной и ненулевой дисперсией. Пусть a – математическое ожидание $\mathsf{E}\,\xi_1$ и σ^2 - дисперсия $\mathsf{D}\,\xi_1$. Требуется доказать, что

$$\frac{S_n - na}{\sigma\sqrt{n}} = \frac{\xi_1 + \ldots + \xi_n - na}{\sigma\sqrt{n}} \Rightarrow \zeta \ \sim\ \mathrm{N}(0,1).$$

Положим с.в. $\zeta_i = (\xi_i - a)/\sigma$ - независимые случайные величины с нулевыми математическими ожиданиями и единичными дисперсиями. Пусть $Z_n = \sum\limits_{i-1}^{n} \zeta_i$:

$$Z_n = \zeta_1 + \ldots + \zeta_n = \frac{S_n - na}{\sigma}.$$

Требуется доказать, что последовательность с.в. $\frac{Z_n}{\sqrt{n}}$ слабо сходится к стандартному нормальному распределению. Характеристическая функция с.в. $\frac{Z_n}{\sqrt{n}}$ равна

$$f_{Z_n/\sqrt{n}}(t) = f_{Z_n}\Big(\frac{t}{\sqrt{n}}\Big) = \Big(f_{\zeta_1}\Big(\frac{t}{\sqrt{n}}\Big)\Big)^n. \tag{7}$$

Характеристическую функцию случайной величины ζ_1 можно разложить по формуле Тейлора, в коэффициентах которого использовать известные моменты $\mathsf{E}\,\zeta_1 = 0, \mathsf{E}\,\zeta_1^2 = \mathsf{D}\,\zeta_1 = 1$:

$$f_{\zeta_1}(t) = 1 + it\mathsf{E}\,\zeta_1 - \frac{t^2}{2}\mathsf{E}\,\zeta_1^2 + o(t^2) = 1 - \frac{t^2}{2} + o(t^2).$$

Подставим это разложение, взятое в точке $t/\sqrt{n}$, в равенство (7) и устремим n к бесконечности. Ещё раз воспользуемся замечательным пределом:

$$\Big(1 - \frac{t^2}{2n} + o\Big(\frac{t^2}{n}\Big)\Big)^n \to e^{-t^2/2} \text{ при } n \to \infty.$$

В пределе получается характеристическую функцию стандартного нормального распределения. По теореме о непрерывном соответствии можно сделать вывод о слабой сходимости

$$\frac{Z_n}{\sqrt{n}} = \frac{S_n - na}{\sigma\sqrt{n}} \Rightarrow \mathrm{N}_{0,1}.$$

Литература

1. Ширяев А.Н. Вероятность. В 2-х кн. М., 2007, 552 с., 416 с.

2. Прохоров Ю.В., Пономаренко Л.С. Лекции по теории вероятностей и математической статистике. М., 2019, 219 с.

3. Ивченко Г.И., Медведев Ю.И. и др. Теория вероятностей и математическая статистика в задачах. М., 2003, 328 с.