Простейшие случайные величины

Высшая школа цифровой культуры
Университет ИТМО
dc@itmo.ru

Содержание

1 Случайные величины

1.1 Понятие случайной величины

Во многих рассматриваемых задачах природа пространства элементарных исходов не важна, а скорее важны некоторые численные характеристики, связанные с этими элементарными исходами. Например, количество выпавших на кубике очков, или количество успехов при подбрасывании монеты n раз. В частности по этой причине в теории вероятностей вводят понятие случайной величины. Одна из целей данного пункта – перенести соответствующие понятия, введенные для событий, на случайные величины, и обобщить их.

Определение 1.1.1 *Пусть Ω – конечное пространство элементарных исходов. Тогда произвольная вещественнозначная функция $\xi : \Omega \to \mathbb{R}$ называется случайной величиной.*

Заметим, что при рассмотрении конечного пространства элементарных исходов, множество значений случайной величины всегда конечно.

Пример 1.1.1 *Пусть случайный эксперимент заключается в двукратном бросании правильного тетраэдра, в вершинах которого написаны числа $1, 2, 3, 4$. Пространство элементарных исходов $\Omega = \{(i, j) : i, j \in \{1, 2, 3, 4\}\}$ состоит из 16 равновозможных исходов $\omega = (i, j)$ означающих, что при первом бросании выпало число i, а при втором число j. Рассмотрим две случайных величины, обозначающих сумму и произведение выпавших очков: $\xi(i, j) = i + j$, $\eta(i, j) = i \cdot j$.*

Так как случайная величина – это функция на пространстве элементарных исходов, а элементарные исходы возникают с некоторой вероятностью, то и случайная величина принимает свои значения с некоторой вероятностью. Тем самым появляется так называемое распределение случайной величины.

Определение 1.1.2 *Распределением случайной величины ξ называется набор вероятностей, с которыми она принимает свои значения $\mathsf{P}(\xi = a)$.*

Что такое $\mathsf{P}(\xi = a)$? Это в точности вероятность события $\{\omega \in \Omega : \xi(\omega) = a\}$.

Для случайных величин, имеющих конечное число значений $a_1, a_2, ..., a_n$ (а пока что других у нас быть и не может), распределение случайной величины часто записывают в виде таблицы (ее еще называют таблицей или рядом распределения):

ξ	a_1	a_2	$\dots$	a_n
P	$\mathsf{P}(\xi = a_1)$	$\mathsf{P}(\xi = a_2)$	$\dots$	$\mathsf{P}(\xi = a_n)$

Отметим очевидное замечание, что если множество значений случайной величины ξ – это множество, состоящее из элементов $a_1, a_2, ..., a_n$, то

$$\mathsf{P}(\xi = a_1) + \mathsf{P}(\xi = a_2) + ... + \mathsf{P}(\xi = a_n) = 1,$$

а значит набор написанных выше вероятностей задает вероятностное пространство.

Пример 1.1.2 *Вернемся к примеру с двукратным бросанием правильного тетраэдра. Случайная величина ξ, равная сумме выпавших очков, принимает значения $2, 3, 4, 5, 6, 7, 8$, а случайная величина η, равная произведению выпавших очков, принимает значения $1, 2, 3, 4, 6, 8, 9, 12, 16$. Напишем их ряды распределения. Подробно обсудим, как написать ряд распределения для величины ξ, для величины η сделайте это сами.*

Всего у нас 16 равновозможных элементарных исходов. Значению 2 величины ξ благоприятствует всего один элементарный исход $\omega = (1, 1)$, поэтому $\mathsf{P}(\xi = 2) = \frac{1}{16}$. Значению 3 благоприятствуют элементарные исходы $\omega = (1, 2)$ и $\omega = (2, 1)$, поэтому $\mathsf{P}(\xi = 3) = \frac{2}{16}$. Рассуждая аналогичным образом дальше, получим

ξ	2	3	4	5	6	7	8
P	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

Аналогичные рассуждения для случайной величины η дают

η	1	2	3	4	6	8	9	12	16
P	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

1.2 Совместное распределение случайных величин. Независимость.

Введем естественным образом возникающее понятие совместного распределения. Ограничимся двумя случайными величинами (совместное распределение большего числа случайных величин вводится аналогичным образом). Пусть случайные величины ξ, η заданы на одном вероятностном пространстве.

Определение 1.2.1 *Совместным распределением случайных величин ξ и η называется набор вероятностей $\mathsf{P}(\xi = a, \eta = b)$, где числа a пробегают всевозможные значения $a_1, ..., a_n$ случайной величины ξ, а числа b – всевозможные значения $b_1, ..., b_k$ случайной величины η, причем*

$$\sum_{i=1}^{n} \sum_{j=1}^{k} \mathsf{P}(\xi = a_i, \eta = b_j) = 1.$$

Пусть случайная величина ξ принимает значения $a_1, ..., a_n$, а случайная величина η принимает значения $b_1, ..., b_k$. Совместное распределение случайных величин часто записывают в виде таблицы

$\xi \setminus \eta$	b_1	b_2	...	b_k
a_1	$\mathsf{P}(\xi = a_1, \eta = b_1)$	$\mathsf{P}(\xi = a_1, \eta = b_2)$	...	$\mathsf{P}(\xi = a_1, \eta = b_k)$
a_2	$\mathsf{P}(\xi = a_2, \eta = b_1)$	$\mathsf{P}(\xi = a_2, \eta = b_2)$	...	$\mathsf{P}(\xi = a_2, \eta = b_k)$
...	...	...	...	...
a_n	$\mathsf{P}(\xi = a_n, \eta = b_1)$	$\mathsf{P}(\xi = a_n, \eta = b_2)$	...	$\mathsf{P}(\xi = a_n, \eta = b_k)$

Пример 1.2.1 *Вернемся к рассмотрению примера с бросанием тетраэдра и запишем совместное распределение двух случайных величин ξ и η.*

$\xi \setminus \eta$	1	2	3	4	6	8	9	12	16
2	$\frac{1}{16}$	0	0	0	0	0	0	0	0
3	0	$\frac{2}{16}$	0	0	0	0	0	0	0
4	0	0	$\frac{2}{16}$	$\frac{1}{16}$	0	0	0	0	0
5	0	0	0	$\frac{2}{16}$	$\frac{2}{16}$	0	0	0	0
6	0	0	0	0	0	$\frac{2}{16}$	$\frac{1}{16}$	0	0
7	0	0	0	0	0	0	0	$\frac{2}{16}$	0
8	0	0	0	0	0	0	0	0	$\frac{1}{16}$

Поясним, как получается эта таблица. Например, как найти вероятность $\mathsf{P}(\xi = 5, \eta = 6)$*? Мы знаем, что всего имеется* 16 *равновозможных элементарных исходов. Какие исходы благоприятствуют событию* $\{\omega : \xi(\omega) = 5, \eta(\omega) = 6\}$*? Таких исходов всего 2:* $\omega_1 = (2, 3)$, $\omega_2 = (3, 2)$. *Значит,* $\mathsf{P}(\xi = 5, \eta = 6) = \frac{2}{16}$. *Аналогично вычисляются другие вероятности.*

Зная совместное распределение, мы можем восстановить так называемые маргинальные (они же – просто обычные одномерные) распределения случайных величин ξ и η по правилам

$$\mathsf{P}(\xi = a_i) = \sum_{j=1}^{k} \mathsf{P}(\xi = a_i, \eta = b_j),\ i \in \{1, ..., n\},$$

$$\mathsf{P}(\eta = b_j) = \sum_{i=1}^{n} \mathsf{P}(\xi = a_i, \eta = b_j),\ j \in \{1, ..., k\}.$$

Пример 1.2.2 *Восстановим, например, распределение случайной величины ξ. Для этого просуммируем значения по строкам в таблице совместного распределения:*

$\xi \setminus \eta$	1	2	3	4	6	8	9	12	16
2	$\frac{1}{16}$	0	0	0	0	0	0	0	0
3	0	$\frac{2}{16}$	0	0	0	0	0	0	0
4	0	0	$\frac{2}{16}$	$\frac{1}{16}$	0	0	0	0	0
5	0	0	0	$\frac{2}{16}$	$\frac{2}{16}$	0	0	0	0
6	0	0	0	0	0	$\frac{2}{16}$	$\frac{1}{16}$	0	0
7	0	0	0	0	0	0	0	$\frac{2}{16}$	0
8	0	0	0	0	0	0	0	0	$\frac{1}{16}$

и получим

ξ	2	3	4	5	6	7	8
P	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

Подобно понятию независимости событий, можно ввести понятие независимости случайных величин.

Определение 1.2.2 *Случайные величины ξ со значениями из множества $A = \{a_1, ..., a_n\}$ и η со значениями из множества $B = \{b_1, ..., b_k\}$ называются независимыми, если*

$$\mathsf{P}(\xi = a, \eta = b) = \mathsf{P}(\xi = a)\mathsf{P}(\eta = b)$$

при всех $a \in A$, $b \in B$.

Иными словами (или на языке событий) это означает, что события $\{\omega \in \Omega : \xi(w) = a\}$ и $\{\omega \in \Omega : \eta(\omega) = b\}$ независимы при всех $a \in A$, $b \in B$. Аналогично независимости в совокупности событий можно ввести независимость в совокупности случайных величин. Обдумайте и сформулируйте это определение сами.

Пример 1.2.3 *Рассматриваемые нами случайные величины ξ и η независимыми не являются. Например,*

$$\frac{2}{16} = \mathsf{P}(\xi = 5, \eta = 6) \neq \mathsf{P}(\xi = 5)\mathsf{P}(\eta = 6) = \frac{4}{16} \cdot \frac{2}{16}.$$

Отметим еще одну важную роль совместного распределения. Зная совместное распределение случайных величин, мы можем написать распределение различных функций от этих случайных величин, в частности распределение суммы, разности или произведения. Знание маргинальных распределений не позволяет этого сделать, как показывает следующий пример.

Пример 1.2.4 *Пусть задано совместное распределение случайных величин ξ и η следующей таблицей ($r \in [0, 0.5]$)*

$\xi \setminus \eta$	0	1
0	r	$\frac{1}{2} - r$
1	$\frac{1}{2} - r$	r

.

Маргинальные распределения у случайных величин ξ и η одинаковы и не зависят от r:

ξ	0	1
P	$\frac{1}{2}$	$\frac{1}{2}$

.

η	0	1
P	$\frac{1}{2}$	$\frac{1}{2}$

.

Найдем распределение случайной величины $\eta + \xi$. Ясно, что сумма может принимать значения $0, 1, 2$, причем

$$\mathsf{P}(\eta + \xi = 0) = \mathsf{P}(\eta = 0, \xi = 0) = r,$$

$$\mathsf{P}(\eta + \xi = 1) = \mathsf{P}((\eta = 1, \xi = 0) \cup (\eta = 0, \xi = 1)) =$$
$$\mathsf{P}(\eta = 1, \xi = 0) + \mathsf{P}(\eta = 0, \xi = 1) = 1 - 2r$$

и

$$\mathsf{P}(\eta + \xi = 2) = \mathsf{P}(\eta = 1, \xi = 1) = r.$$

Тем самым,

$\xi + \eta$	0	1	2
P	r	$1 - 2r$	r

.

Видно, что распределение зависит от r при неизменных маргинальных распределениях.

2 Некоторые числовые характеристики случайных величин

Со случайными величинами связывают много числовых характеристик. В этом пункте отметим некоторые самые популярные и часто используемые в дальнейшем.

2.1 Медиана

Начнем с характеристики, называемой медианой.

Определение 2.1.1 *Число $a \in \mathbb{R}$ называется медианой случайной величины ξ, если*

$$\mathsf{P}(\xi \le a) \ge \frac{1}{2} \text{ и } \mathsf{P}(\xi \ge a) \ge \frac{1}{2}.$$

По сути своей медиана – это такое число, что случайная величина как минимум с вероятностью $\frac{1}{2}$ не больше и не меньше нее. Ясно, что медиана всегда существует, но не всегда является единственной. Если медиана не единственна, часто в качестве медианы берут полусумму наименьшей и наибольшей возможных медиан.

Пример 2.1.1 *Рассмотрим распределение ранее изученной величины ξ, равной сумме очков при двукратном подбрасывании правильного тетраэдра:*

ξ	2	3	4	5	6	7	8
P	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

.

Ясно, что в данном случае число 5 является единственной медианой случайной величины ξ, так как

$$\mathsf{P}(\xi \le 5) = \mathsf{P}(\xi = 2) + \mathsf{P}(\xi = 3) + \mathsf{P}(\xi = 4) + \mathsf{P}(\xi = 5) = \frac{10}{16} > \frac{1}{2}$$

и

$$\mathsf{P}(\xi \ge 5) = \mathsf{P}(\xi = 5) + \mathsf{P}(\xi = 6) + \mathsf{P}(\xi = 7) + \mathsf{P}(\xi = 8) = \frac{10}{16} > \frac{1}{2},$$

но исключив из рассмотрения событие $\xi = 5$, сумма вероятностей сразу становится меньше, чем 0.5.

Теперь вспомним распределение ранее изученной случайной величины η, равной произведению очков при двукратном подбрасывании правильного тетраэдра.

η	1	2	3	4	6	8	9	12	16
P	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

Прямой проверкой убеждаемся, что любое число $a \in [4, 6]$ является медианой случайной величины η, так как

$$\mathsf{P}(\eta \le 4) = \frac{1}{2}, \quad \mathsf{P}(\eta \ge 4) = \frac{11}{16} \ge \frac{1}{2}$$

и

$$\mathsf{P}(\eta \le 6) = \frac{10}{16} \ge \frac{1}{2}, \quad \mathsf{P}(\eta \ge 6) = \frac{1}{2}.$$

В качестве медианы можно взять и полусумму наибольшего и наименьшего значений:

$$\frac{4+6}{2} = 5.$$

2.2 Математическое ожидание

Математическое ожидание – одна из наиболее популярных характеристик случайных величин. Оно показывает что-то вроде среднего значения, так называемое среднее вероятностное значение. Почему так сложно? Давайте рассмотрим простой пример. Пусть случайная величина ξ задана таблицей распределения

ξ	0	100
P	$\frac{99}{100}$	$\frac{1}{100}$

Среднее (среднее арифметическое) ее значение равно $\frac{0+100}{2} = 50$. Алгебраически все честно, а по смыслу? Значение 100 встречается в среднем один раз из ста. Конечно, встречается, но резонно ли считать его настолько же полноправным, как и значение 0? Классически вводят следующее определение.

Определение 2.2.1 *Математическим ожиданием случайной величины ξ называется число*

$$\mathsf{E}\xi = \sum_{\omega\in\Omega} \xi(\omega)\mathsf{P}(\omega) = \sum_{x\in\xi(\Omega)} x\mathsf{P}(\xi = x).$$

В нашем случае математическое ожидание ξ будет равно

$$\mathsf{E}\xi = 0 \cdot \frac{99}{100} + 100 \cdot \frac{1}{100} = 1,$$

что куда ближе к нашим ожиданиям. Кстати, математическое ожидание станет средним арифметическим, если все n значений случайной величины равновозможны, то есть их вероятности равны $\frac{1}{n}$.

Как уже было сказано, по сути дела математическое ожидание есть не что иное, как среднее вероятностное значение случайной величины ξ. Слово значение употреблено несколько неточно, так как математическое ожидание может вовсе не совпадать ни с одним из значений рассматриваемой случайной величины. Второе же равенство дает удобный практический способ вычисления математического ожидания и получается из первого группировкой тех элементарных исходов, при которых случайная величина принимает конкретное значение x.

Можно дать и механическую интерпретацию. Пусть на оси в точках с координатами x_i сосредоточены массы, равные p_i, сумма которых равна 1. Тогда координата так называемого центра масс рассматриваемой системы определяется из соотношения

$$x_c = x_1p_1 + x_2p_2 + ... + x_np_n,$$

что и является аналогом выражения для математического ожидания.

Пример 2.2.1 *Снова вернемся к ранее рассмотренным величинам ξ и η и вычислим их математические ожидания. Вспомним ряд распределения случайной величины ξ*

ξ	2	3	4	5	6	7	8
P	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

и найдем ее математическое ожидание. Оно равно

$$\mathsf{E}\xi = 2\cdot\frac{1}{16}+3\cdot\frac{2}{16}+4\cdot\frac{3}{16}+5\cdot\frac{4}{16}+6\cdot\frac{3}{16}+7\cdot\frac{2}{16}+8\cdot\frac{1}{16}=\frac{80}{16}=5.$$

Аналогично, так как η имеет распределение

η	1	2	3	4	6	8	9	12	16
P	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

то ее математическое ожидание равно

$$\mathsf{E}\eta = 1\cdot\frac{1}{16}+2\cdot\frac{2}{16}+...+16\cdot\frac{1}{16}=\frac{100}{16}=6.25.$$

Замечание 2.2.1 *Заметим, что в случае случайной величины ξ, ее математическое ожидание совпало с ее значением. В случае случайной величины η ситуация противоположная.*

2.3 Небольшое сравнение математического ожидания и медианы

Рассмотрим такой простой пример. Предположим, что имеется некоторая фирма, в которой работает 100 человек, один из которых начальник. Заработная плата начальника равна 101000 долларов в месяц, а заработная плата каждого работника равна 1000 долларов в месяц. Пусть случайная величина ξ – зарплата работника, тогда ее распределение может быть задано, как

ξ	1000	101000
P	$\frac{99}{100}$	$\frac{1}{100}$

Легко понять, что математическое ожидание случайной величины ξ равно

$$\mathsf{E}\xi = 1000\cdot\frac{99}{100}+101000\cdot\frac{1}{100}=2000,$$

то есть средняя зарплата равна 2000 долларов в месяц. В то же время, медиана равна 1000 и медианная зарплата равна 1000. Ясно, что в этом примере гораздо более «честной» является медиана, нежели математическое ожидание, из-за такого сильного выброса в зароботной плате начальника.

2.4 Свойства математического ожидания

Отметим свойства математического ожидания, которыми мы неоднократно будем пользоваться в дальнейшем.

Теорема 2.4.1 (Свойства математического ожидания)
Математическое ожидание обладает следующими свойствами.

1. *Пусть* $\xi \geq 0$, *тогда* $\mathsf{E}\xi \geq 0$. *Иными словами, если случайная величина неотрицательна, то и ее среднее вероятностное неотрицательно.*
2. $\mathsf{E}(a\xi + b\eta) = a\mathsf{E}\xi + b\mathsf{E}\eta$. *Иными словами, математическое ожидание линейно.*
3. *Пусть* $\xi \geq \eta$ *(это значит, что* $\forall \omega \in \Omega$ *выполняется* $\xi(\omega) \geq \eta(\omega)$*), тогда* $\mathsf{E}\xi \geq \mathsf{E}\eta$. *Иными словами, математическое ожидание монотонно.*
4. $|\mathsf{E}\xi| \leq \mathsf{E}|\xi|$. *Иными словами, модуль математического ожидания не превосходит математического ожидания модуля.*
5. $(\mathsf{E}|\xi\eta|)^2 \leq \mathsf{E}\xi^2\mathsf{E}\eta^2$, *то есть для математического ожидания справедливо неравенство Коши-Буняковского.*
6. *Если* ξ, η *независимы, то* $\mathsf{E}(\xi\eta) = \mathsf{E}\xi\mathsf{E}\eta$. *Иными словами, в случае независимых случайных величин, математическое ожидание произведения равно произведению математических ожиданий.*
7. *Если* $\xi = \mathsf{const} \in \mathbb{R}$, *то* $\mathsf{E}(\mathsf{const}) = \mathsf{const}$.

Доказательство. 1. Свойство легко следует из определения, так как если $\xi \geq 0$, то все слагаемые суммы

$$\mathsf{E}\xi = \sum_{x \in \xi(\Omega)} x\mathsf{P}(\xi = x)$$

неотрицательны, а значит и вся сумма неотрицательна.
2.

$$\mathsf{E}(a\xi + b\eta) = \sum_{\omega \in \Omega}(a\xi(\omega) + b\eta(\omega))\mathsf{P}(\omega) = a\sum_{\omega \in \Omega}\xi(\omega)\mathsf{P}(\omega) + b\sum_{\omega \in \Omega}\eta(\omega)\mathsf{P}(\omega) =$$

$$a\mathsf{E}\xi + b\mathsf{E}\eta.$$

3. Это свойство легко следует из первых двух. Так как случайная величина $\xi - \eta \geq 0$, значит, согласно свойству 1, $\mathsf{E}(\xi - \eta) \geq 0$. Тогда по свойству 2 $\mathsf{E}\xi - \mathsf{E}\eta \geq 0$ и $\mathsf{E}\xi \geq \mathsf{E}\eta$.

4. Так как $-|\xi| \le \xi \le |\xi|$, то по свойству 2 и 3 получим $-\mathsf{E}|\xi| \le \mathsf{E}\xi \le \mathsf{E}|\xi|$.
5. Для доказательства данного свойства рассмотрим

$$0 \le \mathsf{E}(\xi + t\eta)^2 = \mathsf{E}\xi^2 + 2t\mathsf{E}(\xi\eta) + t^2\mathsf{E}\eta^2.$$

Так как это квадратный трехчлен с неотрицательным коэффициентом при старшей степени, то его дискриминант неположителен, откуда

$$\left(\mathsf{E}(\xi\eta)\right)^2 \le \mathsf{E}\xi^2\mathsf{E}\eta^2.$$

Подставляя вместо ξ ее модуль, аналогично вместо η ее модуль, получаем требуемое.
6. Так как для независимых случайных величин

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

то

$$\mathsf{E}(\xi\eta) = \sum_{x\in\xi(\Omega), y\in\eta(\Omega)} xy\mathsf{P}(\xi = x, \eta = y) = \sum_{x\in\xi(\Omega), y\in\eta(\Omega)} xy\mathsf{P}(\xi = x)\mathsf{P}(\eta = y) =$$

$$\sum_{x\in\xi(\Omega)} x\mathsf{P}(\xi = x) \sum_{y\in\eta(\Omega)} y\mathsf{P}(\eta = y) = \mathsf{E}\xi\mathsf{E}\eta.$$

7. Это следует из того, что случайная величина $\xi = \mathsf{const}$ имеет вырожденное распределение

ξ	const
P	1

и определения математического ожидания. □

Как мы видели раньше, для того, чтобы написать распределение случайной величины $a\xi + b\eta$, требуется знание совместного распределения. Однако для вычисления математического ожидания этой случайной величины знание совместного распределения вовсе необязательно, достаточно знания маргинальных распределений.

Пример 2.4.1 *Снова обратимся к примеру с двукратным бросанием тетраэдра и вычислим математическое ожидание случайной величины $3\xi + 4\eta$. Так как $\mathsf{E}\xi = 5$, $\mathsf{E}\eta = 6.25$, то*

$$\mathsf{E}(3\xi + 4\eta) = 15 + 25 = 40.$$

Пример 2.4.2 *Покажем, что условие*

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

не является достаточным для независимости случайных величин. Для этого предположим, что пространство элементарных исходов Ω *имеет вид* $\Omega = \{-\frac{\pi}{2}, 0, \frac{\pi}{2}\}$, *и на нем заданы две заведомо зависимые случайные величины* $\xi = \sin\omega$ *и* $\eta = \cos\omega$. *Запишем законы распределения случайных величин:*

ξ	-1	0	1
P	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

η	0	1
P	$\frac{2}{3}$	$\frac{1}{3}$

Ясно, что $\mathsf{E}\xi = 0$, $\mathsf{E}\eta = \frac{1}{3}$, *значит* $\mathsf{E}\xi\mathsf{E}\eta = 0$. *Кроме того, на рассматриваемом пространстве* Ω *всегда* $\xi\eta = 0$, *значит и* $\mathsf{E}(\xi\eta) = 0$. *Тем самым условие*

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

выполнено, но величины зависимы.

Обратимся снова к примеру с тетраэдрами.

Пример 2.4.3 *Вычислим* $\mathsf{E}(\xi\eta)$. *Для этого, пользуясь совместным распределением, напишем распределение* $\xi\eta$.

$\xi \setminus \eta$	1	2	3	4	6	8	9	12	16
2	$\frac{1}{16}$	0	0	0	0	0	0	0	0
3	0	$\frac{2}{16}$	0	0	0	0	0	0	0
4	0	0	$\frac{2}{16}$	$\frac{1}{16}$	0	0	0	0	0
5	0	0	0	$\frac{2}{16}$	$\frac{2}{16}$	0	0	0	0
6	0	0	0	0	0	$\frac{2}{16}$	$\frac{1}{16}$	0	0
7	0	0	0	0	0	0	0	$\frac{2}{16}$	0
8	0	0	0	0	0	0	0	0	$\frac{1}{16}$

Понятно, что

$\xi\eta$	2	6	12	16	20	30	48	54	84	128
P	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

,

а значит

$$\mathsf{E}(\xi\eta) = \frac{1}{16}(2+12+24+16+40+60+72+54+168+128) = 36 \neq 6.25\cdot 5 = 31.25,$$

что еще раз указывает на зависимость случайных величин ξ *и* η. *В дальнейшем нам потребуется значение*

$$\mathsf{E}(\xi\eta) - \mathsf{E}\xi\mathsf{E}\eta = 36 - 31.25 = 4.75.$$

2.5 Дисперсия, ковариация и корреляция

Еще одной важной характеристикой случайной величины является дисперсия. Она характеризует средний квадрат отклонения значений случайной величины от среднего, а корень из дисперсии – так называемый разброс. Вернемся к примеру случайной величины ξ, распределение которой задано следующей таблицей:

ξ	0	100
P	$\frac{99}{100}$	$\frac{1}{100}$

Так как $\mathsf{E}\xi = 1$, то разброс $\max|\xi - \mathsf{E}\xi|$, исходя из логики, равен аж 99, а квадрат его 99^2. Но опять же, насколько резонно его таким считать, ведь значение 100 является чрезвычайно «редким». Вводят следующее определение.

Определение 2.5.1 *Дисперсией случайной величины ξ называется число*

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

В нашем случае величина $(\xi - \mathsf{E}\xi)^2$ имеет распределение

$(\xi - \mathsf{E}\xi)^2$	1	99^2
P	$\frac{99}{100}$	$\frac{1}{100}$

а значит

$$\mathsf{D}\xi = \frac{99}{100} + \frac{99^2}{100} = 99,$$

что намного меньше, чем 99^2. Сама же оценка разброса будет $\sqrt{99} \approx 9.94$.

Еще раз отметим, что, как видно из определения, дисперсия показывает математическое ожидание квадрата отклонения значений случайной величины от ее математического ожидания. В некотором смысле дисперсия характеризует разброс значений случайной величины от ее математического ожидания. Правда дисперсия измеряется не в тех же единицах, что и сама случайная величина, поэтому часто вводят в рассмотрение следующую величину.

Определение 2.5.2 *Средним квадратическим отклонением или стандартным отклонением случайной величины ξ называется число*

$$\sigma_\xi = \sqrt{\mathsf{D}\xi}.$$

Из свойств дисперсии будет видно, что последнее определение корректно, то есть что под корнем не может возникнуть отрицательных чисел.

Теорема 2.5.1 (Свойства дисперсии) *Дисперсия обладает следующими свойствами:*

1. *Она неотрицательна, то есть* $\mathsf{D}\xi \geq 0$.

2. *Она может быть вычислена по формуле* $\mathsf{D}\xi = \mathsf{E}\xi^2 - (\mathsf{E}\xi)^2$.

3. *Она инвариантна относительно сдвига, то есть* $\mathsf{D}(\xi + a) = \mathsf{D}\xi$.

4. *Константа из-под знака дисперсии выносится с квадратом, то есть* $\mathsf{D}(c\xi) = c^2\mathsf{D}\xi$.

5. *Если* ξ *и* η *– независимые случайные величины, то дисперсия суммы равна сумме дисперсий, то есть* $\mathsf{D}(\xi + \eta) = \mathsf{D}\xi + \mathsf{D}\eta$.

Доказательство. Остановимся подробно только на доказательствах 2 свойства, которое позволяет упростить вычисление дисперсии, а также важного 5 свойства.
2. Раскроем скобки и воспользуемся свойствами математического ожидания

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

5. Используя 2 свойство,

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

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

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

Если случайные величины независимы, то $\mathsf{E}(\xi\eta) = \mathsf{E}\xi\mathsf{E}\eta$ и последнее слагаемое равно нулю, что и доказывает утверждение. □

Пример 2.5.1 *Вычислим дисперсию все тех же величин* ξ *и* η. *Вспомним их ряды распределения и напишем распределения их квадратов*

ξ	2	3	4	5	6	7	8
P	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

η	1	2	3	4	6	8	9	12	16
P	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

ξ^2	4	9	16	25	36	49	64
P	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

η^2	1	4	9	16	36	64	81	144	256
P	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

Тогда

$$\mathsf{E}\xi^2 = \frac{1}{16}(4 + 18 + 48 + 100 + 108 + 98 + 64) = 27.5$$

и

$$\mathsf{E}\eta^2 = \frac{1}{16}(1 + 8 + 18 + 48 + 72 + 128 + 81 + 288 + 256) = 56.25.$$

Тогда

$$\mathsf{D}\xi = \mathsf{E}\xi^2 - (\mathsf{E}\xi)^2 = 27.5 - 5^2 = 2.5,\ \sigma_\xi \approx 1.58.$$

$$\mathsf{D}\eta = 56.25 - (6.25)^2 = 17.1875,\ \sigma_\eta \approx 4.15.$$

Еще раз обратимся к выражению для дисперсии суммы двух случайных величин. Как было получено в доказательстве, она равна

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

Определение 2.5.3 *Величина* $\mathsf{E}(\xi\eta) - \mathsf{E}\xi\mathsf{E}\eta$ *называется ковариацией случайных величин* ξ *и* η *и обозначается* $\mathsf{cov}(\xi, \eta)$*.*

Пример 2.5.2 *Как уже было замечено, в примере с тетраэдрами*

$$\mathsf{cov}(\xi, \eta) = \mathsf{E}(\xi\eta) - \mathsf{E}\xi\mathsf{E}\eta = 36 - 31.25 = 4.75.$$

Замечание 2.5.1 *Прямые выкладки показывают, что ковариация может быть определена и следующим способом:*

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

Отметим свойства ковариации, моментально следующие из ее определения и свойств математического ожидания.

Теорема 2.5.2 *Ковариация обладает следующими свойствами.*

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

2. $\mathsf{cov}(a\xi, b\eta) = ab\,\mathsf{cov}(\xi, \eta)$;

3. $\mathsf{cov}(\xi + c, \eta + d) = \mathsf{cov}(\xi, \eta)$;

4. $\mathsf{cov}(\xi, \xi) = \mathsf{D}\xi$.

Обсудим некоторые достоинства ковариации (в прочем, основанные на уже изученном). Если ковариация отлична от нуля, то величины заведомо зависимы. Если же ковариация равна нулю, то о зависимости величин сделать никакого вывода нельзя. В итоге, ковариация характеризует связь между случайными величинами, является некоторым «индикатором» их зависимости.

Ковариация не является безразмерной величиной, измеряется в квадратах тех же единицах, что и исходные величины. В этом случае увеличение значений случайной величины в 1000 раз приведет к увеличению ковариации также в 1000 раз, однако «сила» зависимости не изменится. Эту проблему решает так называемый коэффициент корреляции.

Определение 2.5.4 *Коэффициентом корреляции двух величин ξ и η с отличными от нуля дисперсиями, называется величина*

$$\rho(\xi,\eta)=\frac{\mathsf{cov}(\xi,\eta)}{\sqrt{\mathsf{D}\xi}\sqrt{\mathsf{D}\eta}}=\frac{\mathsf{cov}(\xi,\eta)}{\sigma_\xi\sigma_\eta}$$

Пример 2.5.3 *В нашем примере*

$$\rho(\xi,\eta)\approx\frac{4.75}{1.58\cdot 4.15}\approx 0.72.$$

Легко видеть, что коэффициент корреляции безразмерен. Отметим и другие важные свойства коэффициента корреляции.

Лемма 2.5.1 *Коэффициент корреляции обладает следующими свойствами:*

1. *его абсолютное значение не превосходит единицы, то есть* $|\rho(\xi,\eta)|\leq 1$*;*
2. *Если ξ,η – независимы, то $\rho(\xi,\eta)=0$;*
3. $|\rho(\xi,\eta)|=1$ *тогда и только тогда, когда $\xi=a\eta+b$, причем $a\cdot\rho(\xi,\eta)>0$.*

Доказательство. 1. Доказательство немедленно следует из свойства математического ожидания

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

ведь

$$|\rho(\xi,\eta)|=\frac{|\mathsf{cov}(\xi,\eta)|}{\sigma_\xi\sigma_\eta}\leq\frac{\mathsf{E}(|\xi-\mathsf{E}\xi||\eta-\mathsf{E}\eta|)}{\sigma_\xi\sigma_\eta}\leq$$

$$\frac{\sqrt{\mathsf{E}(|\xi - \mathsf{E}\xi|)^2}\sqrt{\mathsf{E}(|\eta - \mathsf{E}\eta|)^2}}{\sigma_\xi \sigma_\eta} = 1$$

2. Верно в силу того, что ковариация независимых величин равна нулю
3. В одну сторону это свойство доказывается элементарно, сделайте это сами. В другую сторону доказательство технично, его можно найти в дополнительных материалах □

Из свойства 3 видно, что коэффициент корреляции показывает степень линейной зависимости случайных величин, но может совершенно «не чувствовать» какой-то другой функциональной зависимости. У нас уже был такой пример с синусом и косинусом.

2.6 Еще один пример на вычисление всего-всего

Давайте еще на одном примере вычислим все те параметры, которые мы только что изучили. Пусть эксперимент заключается в двукратном бросании игрального кубика, случайная величина ξ – количество выпавших единиц, а случайная величина η – количество выпавших шестерок. Пространство элементарных исходов в случае, когда все исходы равновозможны, записывается следующим образом

$$\Omega = \left\{ \begin{matrix} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{matrix} \right\}.$$

Составим таблицу совместного распределения пары случайных величин (ξ, η):

$\xi \setminus \eta$	0	1	2
0	$\frac{16}{36}$	$\frac{8}{36}$	$\frac{1}{36}$
1	$\frac{8}{36}$	$\frac{2}{36}$	0
2	$\frac{1}{36}$	0	0

Что означает, например, что $\xi = 0, \eta = 0$. Это значит, что не выпало ни одной шестерки и ни одной единицы. Таких исходов 16, всего исходов 36, а значит

$$\mathsf{P}(\xi = 0, \eta = 0) = \frac{16}{36}.$$

Аналогично вычисляются остальные вероятности. Вычислим маргинальные распределения случайных величин ξ и η. Ясно, что они будут одинаковыми, ведь нет разницы что рассматривать: шестерку или единицу (или вообще какую-то другую цифру):

ξ, η	0	1	2
P	$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$

Ясно, что случайные величины ξ и η не являются независимыми в «бытовом» смысле, ведь количество выпавших шестерок зависит от того: выпала единица или нет, и наоборот. Это подтверждается и «формально», например, так как

$$\frac{16}{36} = \mathsf{P}(\xi = 0, \eta = 0) \neq \mathsf{P}(\xi = 0) \cdot \mathsf{P}(\eta = 0) = \left(\frac{25}{36}\right)^2.$$

Вычислим теперь моменты случайных величин ξ и η. Так как они имеют одинаковые распределения, то у них совпадают и моменты. Итак,

$$\mathsf{E}\xi = \mathsf{E}\eta = 0 \cdot \frac{25}{36} + 1 \cdot \frac{10}{36} + 2 \cdot \frac{1}{36} = \frac{12}{36} = \frac{1}{3}.$$

Для вычисления дисперсии найдем распределение величин ξ^2 и η^2. Ясно, что она задается таблицей

ξ^2, η^2	0	1	4
P	$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$

.

Тогда

$$\mathsf{E}\xi^2 = \mathsf{E}\eta^2 = \frac{14}{36} = \frac{7}{18}.$$

Теперь можно найти дисперсии:

$$\mathsf{D}\xi = \mathsf{D}\eta = \frac{7}{18} - \frac{1}{9} = \frac{5}{18}.$$

Вычислим корреляцию. Для этого найдем распределение случайной величины $\xi \cdot \eta$:

$\xi \cdot \eta$	0	1
P	$\frac{34}{36}$	$\frac{2}{36}$

.

Математическое ожидание этой случайной величины равно

$$\mathsf{E}\,(\xi \cdot \eta) = \frac{2}{36} = \frac{1}{18}.$$

Тогда коэффициент корреляции случайных величин ξ и η равен

$$\rho(\xi, \eta) = \frac{\mathsf{cov}(\xi, \eta)}{\sqrt{\mathsf{D}\xi}\sqrt{\mathsf{D}\eta}} = \frac{\frac{1}{18} - \frac{1}{9}}{\frac{5}{18}} = -\frac{1}{5}.$$

2.7 Параметры схемы Бернулли

Покажем, как свойства математического ожидания и дисперсии могут помочь при вычислении характеристик для схемы Бернулли. Пусть случайная величина $\xi = S_n$ показывает количество успехов в серии из n испытаний схемы Бернулли с вероятностью успеха p в каждом испытании. Напишем ее распределение

ξ	0	1	...	n
P	$C_n^0 p^0 (1-p)^n$	$C_n^1 p(1-p)^{n-1}$	...	$C_n^n p^n (1-p)^0$

Если пытаться вычислять математическое ожидание, что называется, «в лоб», то придется суммировать такое выражение

$$\mathsf{E}\xi = \sum_{i=0}^{n} i C_n^i p^i (1-p)^{n-i},$$

что является довольно сложной задачей. Поступим иначе. Представим нашу случайную величину ξ, как сумму n независимых случайных величин

$$\xi = \xi_1 + \xi_2 + ... + \xi_n,$$

где случайная величина ξ_i принимает значение 0, если произошла неудача в i-ом испытании (с вероятностью $1-p$) и значение 1, если произошел успех (с вероятностью p). Все величины ξ_i одинаково распределены и имеют ряд распределения

ξ_i	0	1
P	$1-p$	p

,

причем $\mathsf{E}\xi_i = p$, $\mathsf{D}\xi_i = p - p^2 = p(1-p) = pq$. Тем самым,

$$\mathsf{E}\xi = \mathsf{E}\xi_1 + ... + E\xi_n = np,$$

$$\mathsf{D}\xi = \mathsf{D}\xi_1 + ... + \mathsf{D}\xi_n = npq.$$

3 Закон больших чисел

3.1 Неравенства Маркова и Чебышева в частном случае

Часто в задачах достаточно лишь оценить вероятность некоторого события. Кроме того, иногда вероятность не может быть вычислена точно. Приведем несколько удобных неравенств.

Теорема 3.1.1 (Неравенство Маркова) *Если $\xi \geq 0$ и $t > 0$, то*

$$\mathsf{P}(\xi \geq t) \leq \frac{\mathsf{E}\xi}{t}.$$

Доказательство. Для понятия случайной величины, введенной нами ранее, доказательство достаточно просто. Пусть случайная величина ξ имеет распределение, задаваемое таблицей

ξ	a_1	a_2	...	a_n
P	p_1	p_2	...	p_n

Тогда

$$\mathsf{E}\xi = \sum_{i=1}^{n} a_i \cdot p_i = \sum_{i:\ a_i < t} a_i \cdot p_i + \sum_{i:\ a_i \geq t} a_i \cdot p_i \geq \sum_{i:\ a_i \geq t} a_i \cdot p_i \geq \sum_{i:\ a_i \geq t} t \cdot p_i =$$

$$= t \cdot \sum_{i:\ x_i \geq t} p_i = t \cdot \mathsf{P}\left(\xi \geq t\right).$$

Таким образом,

$$\mathsf{P}\left(\xi \geq t\right) \leq \frac{\mathsf{E}\xi}{t}.$$

□

Неравенство Маркова может давать, например, такие грубые оценки.

Пример 3.1.1 *Пусть студенты всегда опаздывают на целое количество минут ξ и их опоздание в среднем равно 2 минутам. Какова вероятность, что студент опоздает больше, чем преподаватель, который опаздывает на 10 минут и не пускает в аудиторию никого после себя?*

Можно воспользоваться неравенством Маркова «в лоб»:

$$\mathsf{P}(\xi \geq 10) \leq \frac{\mathsf{E}\xi}{10} = \frac{2}{10} = \frac{1}{5} = 0.2.$$

Из неравенства Маркова моментально следует интересующее нас неравенство Чебышёва.

Теорема 3.1.2 (Неравенство Чебышёва)

$$\mathsf{P}((\xi - \mathsf{E}\xi)^2 \geq t) \leq \frac{\mathsf{D}\xi}{t},\ t > 0$$

$$\mathsf{P}(|\xi - \mathsf{E}\xi| \geq t) \leq \frac{\mathsf{D}\xi}{t^2},\ t > 0$$

Доказательство. Положим $\eta = (\xi - \mathsf{E}\xi)^2 \geq 0$. Для нее справедливо неравенство Маркова, то есть

$$\mathsf{P}(\eta \geq t) \leq \frac{\mathsf{E}\eta}{t}.$$

Так как $\mathsf{E}\eta = \mathsf{E}(\xi - \mathsf{E}\xi)^2 = \mathsf{D}\xi$, то первое неравенство доказано.
Для доказательства второго неравенства достаточно заметить, что

$$\mathsf{P}(|\xi - \mathsf{E}\xi| \geq t) = \mathsf{P}((\xi - \mathsf{E}\xi)^2 \geq t^2)$$

$\square$

3.2 Индикаторы множеств

Данный пункт является скорее техническим, нежели идейным. Для дальнейших рассуждений нам необходимо ввести понятие индикатора события $A \subset \Omega$.

Определение 3.2.1 *Индикатором события A называется функция $\mathsf{I}(A)$, которая равна 1, если событие A произошло, и 0 иначе.*

Отметим очевидные свойства индикатора, которые элементарно проверяются и (или) иллюстрируются геометрически

Лемма 3.2.1 *Справедливы следующие свойства*

1. $\mathsf{I}(A \cap B) = \mathsf{I}(A)\mathsf{I}(B)$.
2. $\mathsf{I}(A \cup B) = \mathsf{I}(A) + \mathsf{I}(B) - \mathsf{I}(A)\mathsf{I}(B)$.
3. $\mathsf{E}\ \mathsf{I}(A) = \mathsf{P}(A)$.

3.3 Неравенства Маркова и Чебышева

Часто в задачах достаточно лишь оценить вероятность некоторого события. Кроме того, иногда вероятность не может быть вычислена точно. Приведем несколько удобных неравенств.

Теорема 3.3.1 (Неравенство Маркова) *Если $\xi \geq 0$ и $t > 0$, то*

$$\mathsf{P}(\xi \geq t) \leq \frac{\mathsf{E}\xi}{t}.$$

Доказательство. Рассмотрим событие $A = \{\omega \in \Omega : \ \xi(w) \geq t\}$, тогда $1 = \mathsf{I}(A) + \mathsf{I}(\overline{A})$. Кроме того, так как $\xi \geq 0$, то

$$\xi = \xi \cdot \mathsf{I}(A) + \xi \cdot \mathsf{I}(\overline{A}) \geq \xi \cdot \mathsf{I}(A) \geq t \cdot \mathsf{I}(A).$$

Пользуясь свойствами математического ожидания, получим

$$\mathsf{E}\xi \geq t\mathsf{E}\ \mathsf{I}(A) = t\mathsf{P}(A) \Rightarrow \mathsf{P}(A) \leq \frac{\mathsf{E}\xi}{t}.$$

$\square$

Неравенство Маркова может давать, например, такие грубые оценки.

Пример 3.3.1 *Пусть студенты всегда опаздывают на целое количество минут ξ и их опоздание в среднем равно 2 минутам. Какова вероятность, что студент опоздает больше, чем преподаватель, который опаздывает на 10 минут и не пускает в аудиторию никого после себя?*

Можно воспользоваться неравенством Маркова «в лоб»:

$$\mathsf{P}(\xi \geq 10) \leq \frac{\mathsf{E}\xi}{10} = \frac{2}{10} = \frac{1}{5} = 0.2.$$

Из неравенства Маркова моментально следует интересующее нас неравенство Чебышёва.

Теорема 3.3.2 (Неравенство Чебышёва)

$$\mathsf{P}((\xi - \mathsf{E}\xi)^2 \geq t) \leq \frac{\mathsf{D}\xi}{t},\ t > 0$$

$$\mathsf{P}(|\xi - \mathsf{E}\xi| \geq t) \leq \frac{\mathsf{D}\xi}{t^2},\ t > 0$$

Доказательство. Положим $\eta = (\xi - \mathsf{E}\xi)^2 \geq 0$. Для нее справедливо неравенство Маркова, то есть

$$\mathsf{P}(\eta \geq t) \leq \frac{\mathsf{E}\eta}{t}.$$

Так как $\mathsf{E}\eta = \mathsf{E}(\xi - \mathsf{E}\xi)^2 = \mathsf{D}\xi$, то первое неравенство доказано.
Для доказательства второго неравенства достаточно заметить, что

$$\mathsf{P}(|\xi - \mathsf{E}\xi| \geq t) = \mathsf{P}((\xi - \mathsf{E}\xi)^2 \geq t^2)$$

$\square$

Из неравенства Чебышёва очень легко получаются следующие важные следствия

Лемма 3.3.1 *Пусть $a > 0$. Случайная величина ξ лежит в интервале $(\mathsf{E}\xi - a\sqrt{\mathsf{D}\xi}, \mathsf{E}\xi + a\sqrt{\mathsf{D}\xi})$ с вероятностью не меньшей, чем*

$$\frac{a^2 - 1}{a^2}.$$

Доказательство. Так как

$$\mathsf{P}(\xi \notin (\mathsf{E}\xi - a\sqrt{\mathsf{D}\xi}, \mathsf{E}\xi + a\sqrt{\mathsf{D}\xi})) = \mathsf{P}(|\xi - \mathsf{E}\xi| \geq a\sqrt{\mathsf{D}\xi}) \leq \frac{\mathsf{D}\xi}{a^2\mathsf{D}\xi} = \frac{1}{a^2},$$

то

$$\mathsf{P}(\xi \in (\mathsf{E}\xi - a\sqrt{\mathsf{D}\xi}, \mathsf{E}\xi + a\sqrt{\mathsf{D}\xi})) \geq 1 - \frac{1}{a^2} = \frac{a^2-1}{a^2}.$$

□

Интервал, который мы получили, тесно связан с так называемыми доверительными интервалами, широко изучаемыми в дальнейшем в статистике. Часто формулируют еще такое следствие неравенства Чебышёва, часто называемое правилом «3 сигма».

Лемма 3.3.2 *Случайная величина ξ лежит в интервале* $(\mathsf{E}\xi - 3\sqrt{\mathsf{D}\xi}, \mathsf{E}\xi + 3\sqrt{\mathsf{D}\xi})$ *с вероятностью не меньшей, чем* $\frac{8}{9}$.

3.4 Закон больших чисел

Для объяснения значения происходящего в этом пункте, приведем следующий часто использующийся пример. Пусть в некоторых одинаковых условиях производится измерение какой-то случайной величины a без так называемой систематической ошибки. Случайные же ошибки, которые совершаются в отдельно взятых измерениях, независимы и, так как условия неизменны, одинаково распределены. Резонно ожидать, что при большом числе измерений среднее арифметическое измерений будет сколь угодно близко к истинному значению a. Что же, об этом и говорит закон больших чисел.

3.4.1 Закон больших чисел для схемы Бернулли

В этом пункте мы обсудим закон больших чисел только на примере схемы Бернулли. Пусть S_n – количество успехов в n испытаниях схемы Бернулли. Мы можем записать, что

$$S_n = \xi_1 + \xi_2 + .. + \xi_n,$$

где ξ_k принимает значение 0, если при k-ом испытании произошла неудача и ξ_k принимает значение 1, если при k-ом испытании произошел успех. Как уже было вычислено ранее, $\mathsf{E}S_n = np$, $\mathsf{D}S_n = npq$. Но тогда

$$\mathsf{E}\left(\frac{S_n}{n}\right) = p, \ \mathsf{D}\left(\frac{S_n}{n}\right) = \frac{npq}{n^2} = \frac{pq}{n}.$$

Согласно неравенству Чебышёва и тому, что

$$pq = p(1-p) = \frac{1}{4} - \left(p - \frac{1}{2}\right)^2 \leq \frac{1}{4},$$

получим

$$\mathsf{P}\left(\left|\frac{S_n}{n}-p\right|\geq\varepsilon\right)\leq\frac{\mathsf{D}\left(\frac{S_n}{n}\right)}{\varepsilon^2}=\frac{pq}{\varepsilon^2 n}\leq\frac{1}{4n\varepsilon^2}\xrightarrow[n\to+\infty]{}0.$$

Тем самым, вероятность того, что среднее значение количества успехов при n подбрасываниях отклонилось от вероятности успеха при одном подбрасывании боле, чем на ε, стремится к нулю, при $n\to+\infty$. Кроме того, мы получаем явную оценку сверху для фиксированного ε:

$$\mathsf{P}\left(\left|\frac{S_n}{n}-p\right|\geq\varepsilon\right)\leq\frac{p(1-p)}{n\varepsilon^2}.$$

3.4.2 Общая формулировка закона больших чисел

В случае конечного пространства элементарных исходов общая формулировка закона больших чисел очень проста и лаконична, а доказательство ничем не отличается от того, что приведено для схемы Бернулли.

Теорема 3.4.1 (Закон больших чисел Чебышёва) *Пусть имеется последовательность $\xi_1,\xi_2,...,\xi_n,...$ попарно независимых и одинаково распределенных случайных величин. Тогда*

$$\forall\varepsilon>0\ \ \mathsf{P}\left(\left|\frac{\xi_1+\xi_2+...+\xi_n}{n}-\mathsf{E}\xi_1\right|\geq\varepsilon\right)\xrightarrow[n\to+\infty]{}0.$$

Итак, закон больших чисел утверждает, что в случае попарно независимых и одинаково распределенных случайных величин, вероятность того, что их среднее арифметическое отклоняется от математического ожидания, стремится к нулю с ростом n.

Последнее условие часто записывают в более общей форме (в нашем случае, так как случайные величины одинаково распределены, то их математические ожидания одинаковы)

$$\forall\varepsilon>0\ \ \mathsf{P}\left(\left|\frac{\xi_1+\xi_2+...+\xi_n}{n}-\frac{\mathsf{E}\xi_1+\mathsf{E}\xi_2+...+\mathsf{E}\xi_n}{n}\right|\geq\varepsilon\right)\xrightarrow[n\to+\infty]{}0.$$

4 Предельные теоремы для схемы Бернулли

Еще один важнейший результат теории вероятностей – так называемая центральная предельная теорема. Идея этого пункта – немного пролить на нее свет. Центральная предельная теорема находится, можно сказать, в непосредственной близости к теоремам типа закона больших чисел. В законе больших чисел, как вы видели, устанавливается, что при некоторых общих условиях среднее арифметическое случайных величин становится близко к постоянной величине. Интересно задаться вопросом о характеристиках уклонения сумм от этой постоянной и о вероятности, с которой эти отклонения наблюдаются. На эти вопросы и дает ответы центральная предельная теорема. Пока что мы не можем показать всей ее мощи, поэтому ограничимся схемой Бернулли (интегральной теоремой Муавра-Лапласа). Впрочем, попутно обсудим и теорему Пуассона.

4.1 Теорема Пуассона

Схема Бернулли обладает достаточно существенным недостатком: вычисления вероятностей $C_n^k p^k (1-p)^{n-k}$ довольно трудоемки. Особенно, когда приходится вычислять суммы со слагаемыми такого вида.

Пример 4.1.1 *Пусть требуется найти вероятность события A, заключающегося в том, что произошло не менее 5 успехов в серии из 100 испытаний в схеме Бернулли, где вероятность успеха в каждом испытании постоянна и равна 0.03. Ясно, что такая вероятность может быть вычислена любым из двух возможных способов:*

$$\mathsf{P}(A) = \sum_{k=5}^{100} \mathsf{P}(B(100,k)) = \sum_{k=5}^{100} C_{100}^k \cdot 0.03^k \cdot 0.97^{100-k} =$$

$$= 1 - \sum_{k=0}^{4} \mathsf{P}(B(100,k)) = 1 - \sum_{k=0}^{4} C_{100}^k \cdot 0.03^k \cdot 0.97^{100-k}.$$

Но как эти суммы вычислить?

Сформулируем и докажем первую предельную теорему, называемую теоремой Пуассона.

Теорема 4.1.1 (Теорема Пуассона) *Пусть имеется последовательность схем Бернулли, причем $n \to \infty$, $p_n \cdot n \sim \lambda > 0$, где n – количество испытаний в схеме Бернулли, p_n – вероятность успеха в схеме Бернулли с n испытаниями. Тогда*

$$\lim_{n\to\infty} \mathsf{P}(B(n,k)) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in \mathbb{N} \cup \{0\}.$$

Доказательство. Так как $\lim\limits_{n\to\infty} n \cdot p_n = \lambda$, то $n \cdot p_n = \lambda + o(1)$, а значит

$$p_n = \frac{\lambda}{n} + o\left(\frac{1}{n}\right).$$

Тогда

$$\mathsf{P}(B(n_k)) = C_n^k p_n^k (1-p_n)^{n-k} = \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^k \left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^{n-k} =$$

$$\frac{\lambda^k}{k!} \cdot \frac{n!}{(n-k)!}\left(\frac{1}{n} + o\left(\frac{1}{n}\right)\right)^k \cdot \left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^{n-k}.$$

Так как

$$\frac{n!}{(n-k)!}\left(\frac{1}{n} + o\left(\frac{1}{n}\right)\right)^k = \frac{n!}{n^k(n-k)!}\left(1 + o\left(\frac{1}{n^2}\right)\right)^k \sim 1$$

и

$$\left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^{n-k} \sim e^{-\lambda},$$

получаем требуемое. □

Пример 4.1.2 *Вернемся к нашему примеру. Используя теорему Пуассона, получим* $n \cdot p = 100 \cdot 0.03 = 3 = \lambda$. *Тогда искомая вероятность*

$$\mathsf{P}(A) \approx 1 - \sum_{k=0}^{4} \frac{3^k}{k!} e^{-3} \approx 0.185.$$

Естественно возникает вопрос: как оценить погрешность? Ведь предыдущий пример рассматривается вовсе не в пределе. На это дает ответ следующее интересное утверждение, часто называемое уточненной теоремой Пуассона.

Теорема 4.1.2 (Уточненная теорема Пуассона) *Пусть* $A \subset \mathbb{N} \cup \{0\}$, $np = \lambda$. *Тогда*

$$\left| \sum_{k \in A} \mathsf{P}(B(n,k)) - \sum_{k \in A} \frac{\lambda^k}{k!} e^{-\lambda} \right| \le \min(p, np^2).$$

Пример 4.1.3 *В нашем примере* $p = 0.03$, $np^2 = 0.09$, *а значит* $\min(p, np^2) = 0.03$ *и искомая вероятность события* A *находится в интервале*

$$\mathsf{P}(A) \in (0.182, 0.188),$$

что является достаточно хорошим результатом (ошибка в третьем знаке после запятой).

4.2 Локальная теорема Муавра-Лапласа

Начнем с так называемой локальной теоремы Муавра-Лапласа.

Теорема 4.2.1 (Локальная теорема Муавра-Лапласа) *Пусть S_n – количество успехов в серии из n испытаний схемы Бернулли с вероятностью успеха p в каждом испытании,*

$$x = \frac{k - np}{\sqrt{npq}} = \frac{k - ES_n}{\sqrt{DS_n}},$$

и $k \in \mathbb{N} \cup \{0\}$ меняется так, что существует число M, что $|x| \leq M$ сразу при всех $n \in \mathbb{N}$. Тогда

$$\mathsf{P}(S_n = k) \sim \frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad n \to \infty.$$

Доказательство. Так как $k = np + x\sqrt{npq}$ и $x \geq -M$, то при $n \to \infty$ и $k \to \infty$. Кроме того, так как

$$n - k = n - np - x\sqrt{npq} = nq - x\sqrt{npq},$$

то, рассуждая аналогично, при $n \to \infty$ и $(n-k) \to \infty$. Для дальнейшего напомним факт, называемый формулой Стирлинга:

$$n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n}, \quad n \to \infty.$$

Итак, применяя три раза формулу Стирлинга, получим

$$\mathsf{P}(S_n = k) = C_n^k p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \sim$$

$$\frac{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}}{\left(\frac{k}{e}\right)^k \sqrt{2\pi k} \left(\frac{n-k}{e}\right)^{n-k} \sqrt{2\pi(n-k)}} p^k (1-p)^{n-k}.$$

Из того, что $k = np + x\sqrt{npq}$ и $n - k = nq - x\sqrt{npq}$ получим, что

$$\frac{k}{n} = p + O\left(\frac{1}{\sqrt{n}}\right), \quad \frac{n-k}{n} = q + O\left(\frac{1}{\sqrt{n}}\right),$$

а тогда

$$\frac{\sqrt{2\pi n}}{\sqrt{2\pi k}\sqrt{2\pi(n-k)}} = \frac{1}{\sqrt{2\pi n}} \frac{1}{\sqrt{\left(p + O\left(\frac{1}{\sqrt{n}}\right)\right)\left(q + O\left(\frac{1}{\sqrt{n}}\right)\right)}} \sim \frac{1}{\sqrt{2\pi}\sqrt{npq}}.$$

Далее,

$$\frac{\left(\frac{n}{e}\right)^n}{\left(\frac{k}{e}\right)^k\left(\frac{n-k}{e}\right)^{n-k}}p^k(1-p)^{n-k} = \frac{p^k(1-p)^{n-k}}{\left(\frac{k}{n}\right)^k\left(\frac{n-k}{n}\right)^{n-k}} = \left(\frac{pn}{k}\right)^k\left(\frac{(1-p)n}{n-k}\right)^{n-k}.$$

Логарифмируя последнее равенство, получим

$$-k\ln\frac{k}{np}-(n-k)\ln\frac{n-k}{nq}.$$

Далее,

$$\ln\frac{k}{np} = \ln\frac{np+x\sqrt{npq}}{np} = \ln\left(1+x\sqrt{\frac{q}{np}}\right) = x\sqrt{\frac{q}{np}}-\frac{1}{2}x^2\frac{q}{np}+O\left(\frac{1}{n^{3/2}}\right).$$

Аналогично,

$$\ln\frac{n-k}{nq} = \ln\frac{nq-x\sqrt{npq}}{nq} = \ln\left(1-x\sqrt{\frac{p}{nq}}\right) = -x\sqrt{\frac{p}{nq}}-\frac{1}{2}x^2\frac{p}{nq}+O\left(\frac{1}{n^{3/2}}\right).$$

Тогда, после несложных тождественных преобразований получим, что

$$-k\ln\frac{k}{np}-(n-k)\ln\frac{n-k}{nq} = -\frac{x^2}{2}+o(1).$$

Объединяя все написанное, приходим к требуемому. □

Пример 4.2.1 *В нашем примере* $np=3$, $npq=2.91$, $k=0,1,2,3,4$. *Тогда*

$$\mathsf{P}(A)\approx 1-\frac{1}{\sqrt{2\pi}\sqrt{npq}}\sum_{k=0}^{4}e^{-x^2/2}\approx 0.205.$$

Интересно, что иной способ подсчета

$$\mathsf{P}(A)\approx \frac{1}{\sqrt{2\pi}\sqrt{npq}}\sum_{k=5}^{100}e^{-x^2/2}\approx 0.186$$

дает более точный результат (сравнить с теоремой Пуассона).

4.3 Интегральная теорема Муавра-Лапласа

Интегральная теорема Муавра-Лапласа, как уже отмечалось, является частным случаем центральной предельной теоремы, которая будет подробно изучена в дальнейшем.

Теорема 4.3.1 *Пусть S_n – количество успехов в серии из n испытаний схемы Бернулли с вероятностью успеха p в каждом испытании. Тогда*

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

причем сходимость равномерна по $a < b$.

Эта теорема может быть доказана с использованием локальной теоремы Муавра-Лапласа, но мы не будем этого делать, так как потом она получится, как частный случай уже не раз упоминавшейся центральной предельной теоремы.

Замечание 4.3.1 *Иначе теорема может быть записана следующим образом:*

$$\lim_{n\to\infty} \mathsf{P}\left(a \leq \frac{S_n - \mathsf{E}S_n}{\sqrt{\mathsf{D}S_n}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-x^2/2} dx,$$

Иными словами, при $a = -b < 0$ она устанавливает вероятность (при $n \to \infty$) попадания величины S_n в интервал

$$(\mathsf{E}S_n - b\sqrt{\mathsf{D}S_n}, \mathsf{E}S_n + b\sqrt{\mathsf{D}S_n}).$$

Замечание 4.3.2 *Обозначим*

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^x e^{-x^2/2} dx.$$

В задачах часто требуется найти $\mathsf{P}(A \leq S_n \leq B)$. Тогда при больших значениях n можно записать

$$\mathsf{P}(A \leq S_n \leq B) \approx \left(\Phi\left(\frac{B - np}{\sqrt{npq}}\right) - \Phi\left(\frac{A - np}{\sqrt{npq}}\right)\right).$$

Пример 4.3.1 *В нашем случае $A = 5$, $B = 100$. Тогда остается вычислить*

$$\left(\Phi\left(\frac{100 - 3}{\sqrt{2.91}}\right) - \Phi\left(\frac{5 - 3}{\sqrt{2.91}}\right)\right) \approx 0.119$$

Естественно, рассматривать приближение без оценки погрешности, достаточно безыдейная затея. Сформулируем важную теорему, показывающую, насколько хорошо работает интегральная теорема Муавра-Лапласа.

Теорема 4.3.2 (Уточнение интегральной теоремы Муавра-Лапласа) *Справедлива оценка*

$$\left| \mathsf{P}\left(\frac{S_n - np}{\sqrt{npq}} \in [a,b]\right) - \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-x^2/2} dx \right| < \frac{1}{pq\sqrt{n}}.$$

в последнем примере

$$\frac{1}{pq\sqrt{n}} = \frac{1}{0.03 \cdot 0.97 \cdot 10} < \frac{1}{0.29} < 3.5,$$

то есть оценка погрешности чрезвычайно большая, достоверность результата гарантировать нельзя. Однако, из полученных оценок можно понять, когда и какую теорему применять: Пуассона или интегральную теорему Муавра-Лапласа. Можно сформулировать следующее достаточно популярное «правило»: если n велико, а np «сравнимо с единицей», применяем теорему Пуассона; если же при этом и величина np велика, – теорему Муавра–Лапласа.