

Дисперсия

Ранее мы уже познакомились с математическим ожиданием случайной величины. А на этом уроке мы пройдем *дисперсию*, которая показывает "разброс" случайной величины относительно её математического ожидания.

Обычно, если вы хотите как-то описать случайную величину всего двумя числами, то вы назовёте её математическое ожидание и дисперсию.

Дополнительный материал. На последнем шаге этого урока есть ссылка на сайт с визуализациями, не пропустите ;)

Дисперсия

Определим дисперсию случайной величины.

Определение. Дисперсия случайной величины X это число $E[(X - E[X])^2]$, обозначаемое $Var(X)$ (от английского "variance").

Давайте поймём формулу $E[(X - E[X])^2]$. Проследите за тем, где у нас числа, а где случайные величины:

1. Сначала мы вычисляем математическое ожидание X , то есть получаем число $E[X]$.
2. Затем мы смотрим на разницу между X и числом $E[X]$, где число $E[X]$ мы воспринимаем как постоянную случайную величину. В результате получаем $X - E[X]$. Заметим, что $X - E[X]$ это случайная величина.
3. Потом возводим $X - E[X]$ в квадрат, и получаем случайную величину $(X - E[X])^2$.
4. У случайной величины $(X - E[X])^2$ берём математическое ожидание и получаем число $E[(X - E[X])^2]$.

Заметим, что случайные величины X , $(X - E[X])$ и $(X - E[X])^2$ определены на одном и том же вероятностном пространстве.

Другими словами, дисперсия $Var(X)$ это средний квадрат отклонения X от своего среднего значения $E[X]$. Почему мы используем именно средний квадрат отклонения, мы обсудим позже в этом уроке. Давайте посмотрим на два примера вычисления дисперсии.

Пример 1. Вы каждый день едете из пункта A в пункт B . С вероятностью 0.7 дорога свободна, и вы доезжаете за 10 минут. С вероятностью 0.3 вы попадаете в пробку, и доезжаете за 40 минут. Обозначим ваше время в пути за X . Это случайная величина, определённая на вероятностном пространстве с двумя исходами $\Omega = \{\text{свободно, пробка}\}$. Тогда

$$E[X] = 0.7 \cdot 10 + 0.3 \cdot 40 = 7 + 12 = 19.$$

Случайная величина $(X - E[X])^2$ определена на том же вероятностном пространстве с двумя исходами $\Omega = \{\text{свободно, пробка}\}$. По формуле для математического ожидания получаем:

$$\begin{aligned} Var(X) &= E[(X - E[X])^2] = P(\text{свободно}) \cdot (X(\text{свободно}) - E[X])^2 + P(\text{пробка}) \cdot (X(\text{пробка}) - E[X])^2 = \\ &= 0.7 \cdot (10 - 19)^2 + 0.3 \cdot (40 - 19)^2 = 0.7 \cdot 81 + 0.3 \cdot 441 = 189. \end{aligned}$$

Пример 2. Мы бросаем честную монетку. Пусть $X = 1$, если на честной монетке выпал орёл, и $X = 0$, если выпала решка. Тогда $E[X] = 0.5$. Случайная величина $(X - E[X])^2$ определена на вероятностном пространстве с двумя исходами $\Omega = \{\text{орёл, решка}\}$, вероятность каждого из исходов равна 0.5. По формуле для математического ожидания получаем:

$$\begin{aligned} Var(X) &= E[(X - E[X])^2] = P(\text{орёл}) \cdot (X(\text{орёл}) - E[X])^2 + P(\text{решка}) \cdot (X(\text{решка}) - E[X])^2 = \\ &= 0.5 \cdot (1 - 0.5)^2 + 0.5 \cdot (0 - 0.5)^2 = 0.5 \cdot 0.25 + 0.5 \cdot 0.25 = 0.25 \end{aligned}$$

Неформально. Чем дисперсия полезна для анализа данных? Если дисперсия маленькая, то X часто принимает значения, близкие к $E[X]$. Если дисперсия большая, то X часто принимает значения далёкие от $E[X]$. В частности, дисперсия помогает понять, насколько хорошо число $E[X]$ приближает случайную величину X .

Комментарий 1. В обозначении $E[X]$ обычно используются квадратные скобки, а в обозначении $Var(X)$ используются круглые скобки. Мы не знаем, почему так.

Комментарий 2. В русскоязычных источниках можно встретить обозначение дисперсии $D[X]$. Мы это обозначение использовать не будем.

Задача с проверкой. Дисперсия 1

Задача. Мы бросаем честный кубик. Пусть Y это число, выпавшее на кубике. Найдите $Var(Y)$.

Комментарий. В этой задаче может потребоваться калькулятор.

План решения.

1. Сначала находим математическое ожидание Y , то есть число $E[Y]$.
2. Посмотрим на случайную величину $Y - E[Y]$. Какие значения она принимает и с какими вероятностями?
3. Потом возводим $Y - E[Y]$ в квадрат, и получаем случайную величину $(Y - E[Y])^2$. Какие значения эта величина принимает и с какими вероятностями? Например, если величина $Y - E[Y]$ на каком-то исходе принимала значение (-2) , то величина $(Y - E[Y])^2$ на этом исходе примет значение $(-2)^2 = 4$.
4. И наконец находим число $Var(Y) = E[(Y - E[Y])^2]$.

Проверка. Введите ответ, округлив до третьего знака после запятой

Введите численный ответ

Другая формула для дисперсии

Вычислять дисперсию иногда удобно по другой формуле, которую мы сейчас выведем.

Важно. На протяжении всего вывода формулы внимательно следите за тем, где мы ставим возведение в степень – до скобки или после, внутри знака математического ожидания или вне. Например, $E[X^2]$ это математическое ожидание случайной величины X^2 , а $E[X]^2$ это квадрат числа $E[X]$.

1. Начинаем с $Var(X) = E[(X - E[X])^2]$.
2. Раскроем скобки внутри квадрата: $E[(X - E[X])^2] = E[X^2 - 2X \cdot E[X] + E[X]^2]$.
3. Воспользуемся [линейностью](#) математического ожидания. Математическое ожидание от суммы нескольких слагаемых равно сумме математических ожиданий этих слагаемых: $E[X^2 - 2X \cdot E[X] + E[X]^2] = E[X^2] - 2E[X \cdot E[X]] + E[E[X]^2]$.
4. Посмотрим на $E[X \cdot E[X]]$ – это математическое ожидание случайной величины X , умноженной на константу $E[X]$. Как мы знаем, константу можно выносить за знак математического ожидания. То есть $E[X \cdot E[X]] = E[X] \cdot E[X] = E[X]^2$.
5. Посмотрим на третье слагаемое – $E[E[X]^2]$ – это математическое ожидание квадрата постоянной случайной величины $E[X]$. Так как для любой постоянной случайной величины c выполнено $E[c^2] = c^2$, имеем $E[E[X]^2] = E[X]^2$.
6. Значит, $E[X^2] - 2E[X \cdot E[X]] + E[E[X]^2] = E[X^2] - 2E[X]^2 + E[X]^2 = E[X^2] - E[X]^2$.

Итого,

$$Var(X) = E[X^2] - E[X]^2.$$

Пример. На [шаге](#), где мы определяли дисперсию, мы нашли дисперсию числа орлов на монетке. Она оказалась равна 0.25. Давайте найдём эту дисперсию по новой формуле. Ясно, что в нашем случае $X^2 = X$, поэтому

$$Var(X) = E[X^2] - E[X]^2 = E[X] - E[X]^2 = 0.5 - 0.5^2 = 0.25.$$

На следующих шагах мы потренируемся вычислять дисперсию и докажем её свойства.

Комментарий. Для вычисления математического ожидания X достаточно знать только распределение X . Нам не обязательно знать вероятностное пространство, на котором определена случайная величина X – достаточно знать значения, которые принимает X , и вероятности, соответствующие этим значениям. Аналогично, чтобы вычислить дисперсию X , достаточно знать распределение X .

Мы бросаем честный кубик. Пусть Y это число, выпавшее на кубике. Найдите $Var(Y)$, используя формулу $Var(Y) = E[Y^2] - E[Y]^2$.

Введите ответ, округлив до третьего знака после запятой

Комментарий. Да, в эту задачу можно просто скопировать ответ, который вы получили два шага назад. Но мы надеемся, что вы получите ответ новым способом)

Введите численный ответ

Случайная величина Z имеет пять равновероятных исходов: 2, 3, 4, 5, 6. Найдите $Var(Z)$.

Введите численный ответ

Задача. Пусть $X = 1$ с вероятностью p и $X = 0$ с вероятностью $1 - p$. Такое распределение случайной величины называется *распределением Бернулли* с вероятностью успеха p . Найдите $Var(X)$.

Комментарий. Это можно сделать, воспользовавшись любой из двух эквивалентных формул:

- $Var(X) = E[(X - E[X])^2]$
- $Var(X) = E[X^2] - E[X]^2$

Формулы. Напомним инструкцию о том, как вводить формулы:

1. Знаки $+$, $-$, $*$, $/$ это сложение, вычитание, умножение и деление.
2. **Важно:** вставляйте знак умножения между числом и переменной. Чтобы написать «десять икс» нужно ввести «10*x». Выражение «10x» автоматическая проверка не поймёт.
3. Знак $^$ это возведение в степень. То есть x^4 будет интерпретировано как x^4 .

Введите математическую формулу

Напишите ваш ответ здесь...

Задача. Дисперсия 2

1. Докажите, что для любой случайной величины X выполнено $Var[X] \geq 0$.
 2. Докажите, что $Var[X] = 0$ если и только если X это постоянная случайная величина.
-

Задача с проверкой. Дисперсия 3

Задача.

1. Пусть X и Y это независимые случайные величины. Докажите, что $Var(X + Y) = Var(X) + Var(Y)$.
2. Приведите пример случайных величины X и Y , таких что $Var(X + Y) \neq Var(X) + Var(Y)$.

Задача для проверки.

Заполните пропуски

1. Мы бросаем два кубика. Пусть X и Y это числа, выпавшие на первом и втором броске соответственно. Тогда $Var(X + Y) =$

(ответ округлите до третьего знака после запятой). Будьте внимательны – если вы используете ответы

из предыдущих задач, то используйте сами ответы, а не их округлённые значения.

2. В классе 26 стульев. Пусть случайная величина X – число учеников, пришедших на урок. Пусть случайная величина Y – число мест, оставшихся свободными. Тогда $Var(X + Y) =$

Дисперсия суммы совместно независимых величин

На предыдущем шаге мы доказали, что если X и Y это независимые случайные величины, то $Var(X + Y) = Var(X) + Var(Y)$.

Верно более общее утверждение, которое мы доказывать не будем (его доказательство аналогично).

Утверждение. Пусть X_1, \dots, X_n это совместно независимые случайные величины. Тогда

$$Var(X_1 + \dots + X_n) = Var(X_1) + \dots + Var(X_n)$$

Честный кубик бросают 10 раз подряд. Найдите дисперсию суммы выпавших значений.

Ответ округлите до 3 знаков после запятой.

Введите численный ответ

1. Первый стрелок попадает в мишень с вероятностью 0.4, второй с вероятностью 0.7. Стрелки одновременно стреляют в одну и ту же мишень, их попадания независимы. Найдите дисперсию количества стрел, попавших в мишень.
2. Первый стрелок попадает в мишень с вероятностью 0.3, второй с вероятностью 0.5, третий с вероятностью 0.9. Стрелки одновременно стреляют в одну и ту же мишень, их попадания независимы. Найдите дисперсию количества стрел, попавших в мишень.

Заполните пропуски

1. Дисперсия в первом случае равна .
 2. Дисперсия во втором случае равна .
-

Задача с проверкой. Дисперсия 4

Задача. Пусть c это произвольное действительное число. Выразите дисперсию случайной величины (cX) через дисперсию случайной величины X .

Задача для проверки. Подбрасывается честная монетка. Если выпал орёл – вы получаете 100 рублей. Если выпадает решка – ничего не получаете. Найдите дисперсию количества рублей, которое вы получите после одного броска.

(Ответ получится странный. Подумайте, в каких единицах измерения этот ответ? В рублях? Мы ответим на этот вопрос на шестом [шаге](#) следующего урока)

Введите численный ответ

Вычисления с дисперсиями

Пример 1

Пусть X и Y – независимые случайные величины. При этом $Var(X) = 5$ и $Var(Y) = 7$. Найдём $Var(2X + 3Y)$.

Решение.

- Легко доказать, что если X и Y – независимые случайные величины, то и $(2X)$ и $(3Y)$ – тоже независимые случайные величины.
- Для независимых случайных величин дисперсия суммы равна сумме дисперсий, то есть

$$Var(2X + 3Y) = Var(2X) + Var(3Y).$$

- По задаче с предыдущего шага,

$$Var(2X) = 2^2 \cdot Var(X) = 4Var(X).$$

Аналогично

$$Var(3Y) = 3^2 \cdot Var(Y) = 9Var(Y).$$

- Поэтому:

$$Var(2X + 3Y) = Var(2X) + Var(3Y) = 4Var(X) + 9Var(Y) = 4 \cdot 5 + 9 \cdot 7 = 20 + 63 = 83$$

Пример 2

Более общий случай. Пусть X и Y – независимые случайные величины, а n и m – любые действительные числа. Тогда

$$Var(nX + mY) = Var(nX) + Var(mY) = n^2Var(X) + m^2Var(Y)$$

Доказательство. Аналогично решению Примера 1.

Пример 3

Ещё более общий случай. Мы уже [писали](#) (без доказательства), что дисперсия суммы совместно независимых величин это сумма дисперсий этих величин.

Поэтому если X_1, X_2, \dots, X_n это совместно независимые случайные величины, и $a_1, \dots, a_n \in \mathbb{R}$ это любые числа, то

$$Var(a_1X_1 + a_2X_2 + \dots + a_nX_n) = Var(a_1X_1) + Var(a_2X_2) + \dots + Var(a_nX_n) = a_1^2Var(X_1) + a_2^2Var(X_2) + \dots + a_n^2Var(X_n)$$

Доказательство. Аналогично решению Примера 1.

Заполните пропуски

1. Даны две независимые случайные величины X и Y , причём $Var(X) = 3$, $Var(Y) = 8$. Чему равно $Var(10X + 5Y)$?

2. Пусть мы три раза подряд кидаем честный кубик, после чего получаем $5X_1 + 20X_2 - 10X_3$ рублей, где X_1, X_2, X_3 – результаты первого, второго и третьего броска соответственно. Чему равно матожидание количества полученных денег?

А чему равна дисперсия? (Будьте внимательны – если вы используете ответы из предыдущих задач, то используйте сами ответы, а не их округлённые значения)

Выберите все подходящие ответы из списка

Для любых случайных величин X и Y верно, что $Var(X + Y) = Var(X) + Var(Y)$.

$Var(k \cdot X) = k \cdot Var(X)$

Дисперсия всякой случайной величины не меньше нуля.

Если дисперсия случайной величины равна нулю, то эта величина постоянна.

Что мы прошли на этом уроке

На этом уроке мы

- познакомились с дисперсией, вот две формулы для неё:

$$Var(X) = E[(X - E[X])^2] \text{ и } Var(X) = E[X^2] - E[X]^2.$$

- доказали несколько свойств дисперсии
- посчитали дисперсию на примерах

Что нас ждёт на следующем уроке

На следующем уроке мы

- узнаем, что такое *биномиальное распределение*
- познакомимся со *стандартным отклонением*

Дополнительный материал – визуализации

Поиграйте с первыми шестью визуализациями на этом сайте: [Seeing Theory: a visual introduction to probability and statistics](#). Это займёт около 20 минут. Вы увидите визуализации про: вероятность, математическое ожидание, дисперсию, теорию множеств, перестановки и условную вероятность.

Неформально, можно выделить два способа думать про вероятности:

- математический – вводить формальные определения вероятностного пространства, случайной величины и т.д.
- экспериментальный – проводить многократные испытания и смотреть на вероятность через результаты испытаний.

В нашем курсе мы начинаем с первого способа. В конце курса, когда мы будем говорить о статистике, мы обсудим и второй способ. Авторы сайта же сразу подходят к понятиям из теории вероятностей с позиции многократных испытаний. Поэтому сайт нам особенно полезен – там объясняют теорию вероятности не с той стороны, с которой мы её видели в первых двух неделях курса.

Комментарий. Всего этот сайт состоит из 18 визуализаций, они все хорошие. Но некоторые из них относятся к понятиям, которые мы пока не прошли. Можете посмотреть соответствующие визуализации сейчас, или позже, когда мы дойдём до этих понятий в курсе.

