

Классическое определение вероятности

Ключевым объектом изучения в теории вероятностей является *вероятность*. Это числовая характеристика *события*. Сначала люди заметили, что про какие-то события можно говорить, что они точно произойдут, про какие-то говорить, что они точно не произойдут. Потом они заметили, что некоторые события могут как произойти, так и не произойти, но есть смысл ожидать реализацию одного события больше, чем другого. Например, если вы играете в покер и у вас на руках пара тузов, то у вас есть больше надежды на победу, чем если у вас семёрка и тройка.

Чтобы сравнивать какое-то свойство событий, разумно каждому событию поставить в соответствие число. Рассмотрим очень простую модель. Существует N исходов, мы делаем некоторый эксперимент, и в результате этого эксперимента может произойти один из этих исходов, причём **равновероятно**. Для нас «благоприятными» исходами являются какие-то m из этих N . Тогда *вероятностью* того события, что реализуется один из этих m исходов, называют число

$$P = \frac{m}{N} \quad (1)$$

Формула (1) называется *классическим* определением вероятности. Оно работает ровно в тех случаях, когда у нас есть конечное множество равновероятных исходов.

Мы подбрасываем монетку, и у нас есть два исхода - выпадет орёл или выпадет решка. Или же мы подбрасываем игральный кубик, и у нас есть шесть исходов - выпадет 1, 2, 3, 4, 5 или 6.

Посчитать вероятность в такой модели — это комбинаторная задача. Нужно найти количество благоприятных исходов, и разделить их на общее число исходов.

Пример 1. Бросаем два игральных кубика. Найти вероятность того, что сумма выпавших чисел равна 10.

Решение. Всего исходов 36, благоприятных 3: $\{(4; 6), (5; 5), (6; 4)\}$.

Значит, вероятность равна $\frac{3}{36} = \frac{1}{12} \approx 8.3\%$. ■

Пример 2. Найти вероятность того, что перестановка длины n содержит только один цикл. У нас есть $n!$ всевозможных перестановок, считаем их всех равновероятными.

Теперь посчитаем, сколько есть «хороших» перестановок. Если перестановка — это простой цикл длины n , то выйдя из элемента 1 и идя из элемента i в элемент p_i , мы должны вернуться в 1 через n шагов и не раньше.

Давайте строить такой цикл. Выходя из 1, у нас есть $n - 1$ вариант куда пойти. Пусть мы пошли в p_1 . Тогда из p_1 мы можем пойти куда угодно, кроме 1 и $p_1 - n - 2$ варианта. Рассуждая аналогично, придём к ситуации когда остаётся один не посещённый элемент — идём в него, затем в 1. Итого у нас получилось $(n - 1) \cdot (n - 2) \cdot \dots \cdot 2 \cdot 1 \cdot 1 = (n - 1)!$ перестановок.

Значит, искомая вероятность равна $\frac{(n-1)!}{n!} = \frac{1}{n}$. ■

1. (56) Пусть мы имеем **последовательность** из m случайных чисел, каждое из которых равновероятно принимает значение от 1 до n .

1. Посчитайте вероятность того, что хотя бы два числа совпадают.
2. Оцените Θ -асимптотику величины m , при которой эта вероятность равна 50%.
3. Оцените, какой должна быть величина m , чтобы эта вероятность была равна p .

Нужно получить правильный константный множитель.

2. (26) Подбрасываем «честную» монету 10 раз. Подсчитайте вероятности следующих событий:

1. число выпавших «орлов» равно числу «решек»;
2. выпало больше «орлов», чем «решек»;

3. при $i = 1, \dots, 5$ одинаковы результаты i -го и $11 - i$ -го бросков;
4. «орел» выпал не менее четырех раз подряд.

Современное (аксиоматическое) определение вероятности

Классическое определение вероятности работает только тогда, когда есть конечное пространство равновероятных исходов. Когда множество исходов счётно, она уже не работает. Кроме того, основываясь на таком определении сложно строить полноценную теорию. Яркой иллюстрацией этого служит [парадокс Бертрона](#), показывающий, что вероятность может существенно зависеть от метода выбора случайной величины.

Приведём аксиоматику Колмогорова, из которой выводится современная теория вероятности.

Опр. σ -алгебра над множеством X - это такое семейство его подмножеств $\Omega \subseteq 2^X$, что:

- $X \in \Omega$;
- если $\omega \in \Omega$, то $X \setminus \omega \in \Omega$;
- пересечение любого счётного подсемейства Ω принадлежит Ω .

Опр. Вероятностное пространство - это тройка (X, Ω, \mathbb{P}) , где:

- X - множество элементарных исходов;
- Ω - σ -алгебра над X (её элементы называются *событиями*);
- \mathbb{P} - вероятностная мера или вероятность - такая функция $\mathbb{P} : \Omega \rightarrow [0, 1]$, что:

$$- \mathbb{P}(X) = 1;$$

$$- \text{Если } \omega_1, \omega_2, \dots \in \Omega \text{ — попарно не пересекаются, то } \mathbb{P}\left(\bigcup_{i=1}^{\infty} \omega_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(\omega_i).$$

В определении сказано, что σ -алгебра замкнута относительно счётного пересечения. Из этого также следует, что σ -алгебра замкнута относительно счётного объединения. Действительно, если $\omega_1, \omega_2, \dots \in \Omega$, то $\bigcup_{i=1}^{\infty} \omega_i = X \setminus \left(\bigcap_{i=1}^{\infty} (X \setminus \omega_i)\right) \in \Omega$. Более того, если бы заменили в определении «пересечение» на «объединение», это было бы эквивалентное определение.

Очевидно, что относительно конечного пересечения и объединения она тоже замкнута. Также σ -алгебра замкнута относительно разности множеств: если $A, B \in \Omega$, то $A \setminus B = A \cap (X \setminus B) \in \Omega$.

Рассмотрим несколько примеров вероятностных пространств.

Пример 3. Бросается два игральных кубика. На каждом кубике равновероятно выпадает натуральное число от 1 до 6.

Здесь множество элементарных исходов - это множество пар (i, j) , где i и j пробегает значения от 1 до 6, то есть это конечное множество из 36 элементов. σ -алгеброй здесь будет множество всех подмножеств X . Вероятностная мера имеет вид $\mathbb{P}(\omega) = \frac{|\omega|}{36}$, где $|\omega|$ - количество элементов в подмножестве ω .

Рассмотрим примеры событий:

- «Выпало две шестёрки».

$$\omega = \{(6; 6)\}$$

$$\mathbb{P}(\omega) = \frac{1}{36}$$

- «Выпала тройка и пятёрка».

$$\omega = \{(3; 5), (5; 3)\}$$

$$\mathbb{P}(\omega) = \frac{2}{36} = \frac{1}{18}$$

- «Сумма выпавших чисел равна 10».

$$\omega = \{(4; 6), (5; 5), (6; 4)\}$$

$$\mathbb{P}(\omega) = \frac{3}{36} = \frac{1}{12}$$

- «Выпала хотя бы одна шестёрка».

$$\omega = \{(1; 6), (2; 6), (3; 6), (4; 6), (5; 6), (6; 6), (6; 1), (6; 2), (6; 3), (6; 4), (6; 5)\}$$

$$\mathbb{P}(\omega) = \frac{11}{36}$$

- «Сумма выпавших чисел меньше 13».

$$\omega = X$$

$$\mathbb{P}(\omega) = \frac{36}{36} = 1$$

- «Сумма выпавших чисел больше 13».

$$\omega = \emptyset$$

$$\mathbb{P}(\omega) = \frac{0}{36} = 0$$

Из этого примера видно, что рассмотренное выше «классическое» определение вероятности является частным случаем аксиоматики Колмогорова, когда $X = \{1, 2, \dots, N\}$, $\Omega = 2^X$, $\mathbb{P}(\omega) = \frac{|\omega|}{N}$. Здесь обозначение 2^X означает «множество всех подмножеств множества X ». ■

Пример 4. Пусть $X = \mathbb{R}$, Ω — минимальная (по включению) σ -алгебра, содержащая любой интервал (a, b) (это называется *борелевская σ -алгебра*).

Вероятностная мера задаётся следующим образом. Рассмотрим монотонно неубывающую функцию $F : \mathbb{R} \rightarrow [0, 1]$. Тогда $\forall t \in \mathbb{R} : \mathbb{P}\left((-\infty, t]\right) = F(t)$.

Можно показать, что любой элемент борелевской σ -алгебры (в дальнейшем будем её обозначать буквой \mathcal{B}) представляет собой объединение непересекающихся отрезков, полуинтервалов, интервалов, точек и лучей на числовой прямой. Также можно показать, что вероятностную меру любого такого множества можно вычислить, зная F . Значит, зная только функцию F , мы можем вычислить вероятностную меру любого подмножества \mathcal{B} . ■

Условная вероятность и формула Байеса ([2], §1.2)

Опр. Событие A невозможно, если $\mathbb{P}(A) = 0$.

Опр. Событие A достоверно, если $\mathbb{P}(A) = 1$.

Опр. События A и B несовместны, если $\mathbb{P}(A \cap B) = 0$.

Опр. События A и B независимы, если $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$;

Опр. События A_1, \dots, A_n попарно независимы, если любые два из них независимы.

Опр. События A_1, \dots, A_n *независимы в совокупности*, если

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

Очевидно, что из независимости в совокупности следует попарная независимость. А вот обратное не всегда верно. Приведём простейший контрпример.

Пример 5. Пусть $X = \{x_1, x_2, x_3, x_4\}$, $\Omega = 2^X$, $\mathbb{P}(x_1) = \mathbb{P}(x_2) = \mathbb{P}(x_3) = \mathbb{P}(x_4) = \frac{1}{4}$.

Рассмотрим три события: $A = \{x_1, x_2\}$, $B = \{x_1, x_3\}$, $C = \{x_2, x_3\}$.

Тогда

$$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2},$$

$$\mathbb{P}(A \cap B) = \mathbb{P}(x_1) = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(B),$$

$$\mathbb{P}(A \cap C) = \mathbb{P}(x_2) = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(C),$$

$$\mathbb{P}(B \cap C) = \mathbb{P}(x_3) = \frac{1}{4} = \mathbb{P}(B)\mathbb{P}(C).$$

Получается, что события A, B, C попарно независимы.

Но $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\emptyset) = 0$, хотя $\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \frac{1}{8}$, значит они не являются независимыми в совокупности. ■

Опр. Условная вероятность события A при условии события B (такого, что $\mathbb{P}(B) > 0$) — это величина

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

Пример 6 ([5], задача 1.130). Найти условную вероятность того, что при бросании трёх игральных костей хотя бы на одной выпадет 6 очков, при условии, что на всех костях выпали грани с чётным числом очков.

Решение. Событие B — «на всех костях выпали грани с чётным числом очков». Ему соответствует $3^3 = 27$ комбинаций (на каждом кубике может выпасть 2, 4 или 6).

Событие A — «хотя бы на одной грани выпало 6 очков».

Удобнее рассмотреть событие $\neg A = X \setminus A$ — «ни на одной грани не выпало 6 очков». Тогда $(\neg A) \cap B$ — «на всех гранях выпало чётное число очков, но не равное 6», т.е. на всех гранях выпало 2 или 4, таких комбинаций $2^3 = 8$.

По определению,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B) - \mathbb{P}((\neg A) \cap B)}{\mathbb{P}(B)} = \frac{27 - 8}{27} = \frac{19}{27}.$$

Обратите внимание, что когда множество элементарных исходов $|X|$ конечно, а сами исходы равновероятны, мера какого-то события — это количество исходов (комбинаций), соответствующих этому событию, делённая на $|X|$ ($|X| = 6^3$ в этом примере). При делении эта константа сокращается, поэтому здесь мы использовали количество исходов в качестве меры события. ■

Из определения условной вероятности следует, что для любых событий A и B

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B). \quad (2)$$

Обратите внимание, что здесь мы уже не требуем, чтобы $\mathbb{P}(B) > 0$, потому что если $\mathbb{P}(B) = 0$, то т.к. $A \cap B \subseteq B$, $\mathbb{P}(A \cap B) \leq \mathbb{P}(B)$, а значит $\mathbb{P}(A \cap B) = 0$, и равенство всё равно верно.

Пусть имеется набор попарно непересекающихся событий B_1, \dots, B_n , такой что $\bigcup_{i=1}^n B_i = X$.

Знак \sqcup вместо обычного знака объединения означает, что множества попарно не пересекаются, т.е. $B_i \cap B_j = \emptyset$ при $i \neq j$. Если мы используем этот знак, можно дополнительно не оговаривать, что они попарно не пересекаются.

Тогда для любого события A можно записать:

$$A = \bigcup_{i=1}^n (A \cap B_i).$$

Заметим, что все объединяемые множества попарно не пересекаются. Действительно, если два множества не пересекались, а мы их пересекли с третьим множеством, в них не может добавиться новых точек, и они не могут пересекаться. Значит, по определению вероятностной меры, можно записать:

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A \cap B_i).$$

Применяя формулу (2), имеем:

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i). \quad (3)$$

Формула (3) называется *формулой полной вероятности*.

Ещё раз запишем формулу (2): $\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B)$. Поменяв местами A и B , получим $\mathbb{P}(B \cap A) = \mathbb{P}(B|A) \cdot \mathbb{P}(A)$. Имеем:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B) = \mathbb{P}(B|A) \cdot \mathbb{P}(A).$$

Выразим отсюда $\mathbb{P}(B|A)$:

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B) \cdot \mathbb{P}(B)}{\mathbb{P}(A)} \quad (4)$$

Выражение (4) называется *формулой Байеса*. Это очень важная теорема в теории вероятности, давайте посмотрим как её можно использовать.

Пример 7 ([5], задача 1.117). Два стрелка стреляют по мишени. Один из них попадает с вероятностью 50%, а второй — с вероятностью 80%. Перед выстрелом они бросают правильную монету для определения очерёдности. Посторонний наблюдатель знает условия стрельбы, но не знает, кто в данный момент стреляет. Вот он видит, что стрелок попал в цель. Какова вероятность того, что стрелял первый стрелок?

Решение. Введём обозначения для событий: S_1 — стрелял первый стрелок, S_2 — стрелял второй стрелок, A — в цель попали. Тогда по условию задачи

$$\mathbb{P}(S_1) = \mathbb{P}(S_2) = 0.5;$$

$$\mathbb{P}(A|S_1) = 0.5;$$

$$\mathbb{P}(A|S_2) = 0.8.$$

Нам требуется вычислить $\mathbb{P}(S_1|A)$. Запишем формулу Байеса (4):

$$\mathbb{P}(S_1|A) = \frac{\mathbb{P}(A|S_1) \cdot \mathbb{P}(S_1)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|S_1) \cdot \mathbb{P}(S_1)}{\mathbb{P}(A|S_1) \cdot \mathbb{P}(S_1) + \mathbb{P}(A|S_2) \cdot \mathbb{P}(S_2)} = \frac{0.5 \cdot 0.5}{0.5 \cdot 0.5 + 0.8 \cdot 0.5} = \frac{5}{13} \approx 38.5\%.$$

Обратите внимание, что знаменатель мы расписали, воспользовавшись формулой полной вероятности. ■

3. (16) Бросают два кубика. Пусть событие A — «на первом кубике выпало 6», событие B — «сумма выпавших очков равна 7». Найдите $\mathbb{P}(A|B)$ и $\mathbb{P}(B|A)$.
4. (16) Независимы ли события: при броске кубика «выпало чётное число очков» и «выпало число очков, кратное трём»?

Случайная величина

Определение

Опр. *Случайная величина* — это измеримая числовая функция на вероятностном пространстве, т.е. такая функция $\xi : X \rightarrow \mathbb{R}$, что для любого элемента борелевской σ -алгебры $b \in \mathcal{B}$ выполнено $\xi^{-1}(b) \in \Omega$.

Разберёмся, что же такое измеримость и зачем она нужна. Рассмотрим элемент борелевской σ -алгебры $b \in \mathcal{B}$. То что $\xi^{-1}(b) \in \Omega$ значит, что мы можем посчитать вероятностную меру множества $\xi^{-1}(b) \equiv \{\omega \in \Omega | \xi(\omega) \in b\}$, а это то же самое, что вероятность события $\xi \in b$.

Теперь вспомним, что элементы борелевской σ -алгебры — это всевозможные точки, отрезки, полуинтервалы, интервалы, лучи и их счётные объединение и пересечения. То есть мы хотим, чтобы для любого такого подмножества числовой прямой мы могли посчитать вероятность того, что случайная величина ξ принадлежит ему — это довольно разумное желание.

Пример 8 (Распределение Бернулли). Пусть $X = \{A, B\}$, $\Omega = 2^X$, $\mathbb{P}(A) = p$, $\mathbb{P}(B) = 1 - p$. Рассмотрим такую функцию $\xi : X \rightarrow \{0, 1\}$, что $\xi(A) = 1$, $\xi(B) = 0$.

Мы получили случайную величину, которая принимает значение 1 с вероятностью p и значение 0 с вероятностью $1 - p$. ■

Функция распределения

Опр. *Функция распределения* случайной величины ξ — это такая функция $F_\xi : \mathbb{R} \rightarrow [0, 1]$, что

$$F_\xi(x) = \mathbb{P}(\xi \leq x).$$

Заметим, что совсем формально нужно записать $F_\xi(x) = \mathbb{P}(\{w | \xi(w) \leq x\})$. Мы помним, что в аксиоматике событие — это некоторое подмножество множества элементарных исходов, но на практике событием у нас всегда будет некоторое логическое выражение, содержащее случайную величину. Когда мы говорим о вероятности события, мы говорим о вероятностной мере такого множества $\omega \in \Omega$, что если его элементы подставлять в это логическое выражение как аргумент случайной величины, оно будет истинным, а если подставлять элементы из $X \setminus \omega$, то оно будет ложным.

Функция распределения всегда определена, потому что лучи вида $(-\infty, x]$ принадлежат \mathcal{B} . Более того, счётное число раз пересекая и объединяя такие лучи, можно получить любой элемент $b \in \mathcal{B}$, а значит, имея только функцию распределения, можно определить вероятность того, что случайная величина принадлежит b . А так как большего мы не требуем, то функция распределения полностью определяет случайную величину.

Например, мы не всегда сможем узнать, какова вероятность того, что случайная величина принимает иррациональное значение, но мы этого и не требовали в определении случайной величины.

Вообще говоря, когда мы имеем дело со случайной величиной, можно считать, что мы имеем дело с вероятностным пространством $\{\mathbb{R}, \mathcal{B}, \mathbb{P}\}$, описанным в примере 4, причём функция F — не что иное, как функция распределения.

Пример 9. Функция распределения случайной величины Бернулли имеет вид

$$F_{Be(p)}(x) = \begin{cases} 0, & x < 0 \\ 1 - p, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases} \quad (5)$$

Пример 10. Рассмотрим равномерное распределение на отрезке $[a, b]$. Это значит, что случайная величина принимает значения на отрезке $[a, b]$, и вероятность того, что она принадлежит некоторому подотрезку этого отрезка пропорциональна его длине. То есть вероятность принадлежать отрезку $[a, x]$ при $x \in [a, b]$ должна быть пропорциональна $x - a$.

Функция распределения будет иметь вид:

$$F_{U([a,b])}(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases} \quad (6)$$

Так как случайная величина однозначно задаётся функцией распределения, говорят о «распределениях», и это то же самое, что случайная величина. Некоторые распределения имеют общепринятые обозначения, и для того, чтобы указать, что случайная величина имеет такое распределение, используют тильду. Например, запись $h \sim Be(p)$ читается «случайная величина h имеет распределение Бернулли с параметром p » и значит, что функция распределения случайной величины h имеет вид (5), а запись $z \sim U[a, b]$ читается «случайная величина z равномерно распределена на отрезке $[a, b]$ » и означает, что функция распределения случайной величины z имеет вид (6).

Математическое ожидание и дисперсия

Опр. Если случайная величина принимает не более чем счётное количество значений, то такая случайная величина называется *дискретной*.

Распределение Бернулли — пример дискретного распределения, так как случайная величина Бернулли принимает всего два значения.

В этом задании и в дальнейшем в курсе АМВ мы будем иметь дело только с дискретными случайными величинами. Для дискретной величины множество элементарных исходов можно отождествить с множеством её значений: $X = \text{dom} \xi \equiv \{\xi(x) | x \in X\}$, считать что $\Omega = 2^X$, а чтобы задать вероятностную меру достаточно для каждого $x_i \in \text{dom} \xi$ указать $\mathbb{P}(\xi = x_i)$.

Если мы рассматриваем несколько (не более чем счётное число) случайных величин, в качестве X можно взять декартово произведение множеств значений этих величин.

Ниже мы формально определим математическое ожидание и докажем его свойства для дискретных случайных величин, однако те же свойства имеют место для любых случайных величин, только сумму нужно заменить на интеграл.

Опр. Математическое ожидание дискретной случайной величины ξ — это число

$$\mathbb{E} \xi = \sum_{x \in \text{dom} \xi} x \cdot \mathbb{P}(\xi = x) \quad (7)$$

Смысл математического ожидания в следующем: если мы проведём много одинаковых независимых испытаний, и посчитаем среднее арифметическое значений случайной величины, то оно будет стремиться к математическому ожиданию этой случайной величины при количестве испытаний, стремящемся к бесконечности.

Можно дать эквивалентное определение через сумму по элементам множества элементарных исходов (его будет удобнее использовать при доказательствах):

$$\mathbb{E}\xi = \sum_{x \in X} \xi(x) \mathbb{P}(x) \quad (8)$$

Докажем линейность математического ожидания:

$$\begin{aligned} \mathbb{E}(\xi + \eta) &= \sum_{x \in X} (\xi(x) + \eta(x)) \mathbb{P}(x) = \sum_{x \in X} \xi(x) \mathbb{P}(x) + \sum_{x \in X} \eta(x) \mathbb{P}(x) = \mathbb{E}\xi + \mathbb{E}\eta \\ \mathbb{E}(a\xi) &= \sum_{x \in X} (a\xi(x)) = a \sum_{x \in X} \xi(x) = a \cdot \mathbb{E}\xi \end{aligned}$$

Полезно уметь вычислять математическое ожидание функции от случайной величины:

$$\mathbb{E}f(\xi) = \sum_{x \in X} f(\xi(x)) \mathbb{P}(x) = \sum_{x \in \text{dom}\xi} f(x) \cdot \mathbb{P}(\xi = x)$$

Опр. *Дисперсия* случайной величины — это средний квадрат отклонения случайной величины от её математического ожидания:

$$\mathbb{D}\xi = \mathbb{E}(\xi - \mathbb{E}\xi)^2 \quad (9)$$

Дисперсия показывает разброс случайной величины. В частности, $\mathbb{D}\xi = 0$ тогда и только тогда, когда ξ всегда принимает одно и то же значение.

Дисперсию можно выразить через математическое ожидание случайной величины и её квадрата:

$$\mathbb{D}\xi = \mathbb{E}(\xi - \mathbb{E}\xi)^2 = \mathbb{E}(\xi^2 - 2\xi\mathbb{E}\xi + (\mathbb{E}\xi)^2) = \mathbb{E}(\xi^2) - 2\mathbb{E}\xi \cdot \mathbb{E}\xi + (\mathbb{E}\xi)^2 = \mathbb{E}(\xi^2) - (\mathbb{E}\xi)^2 \quad (10)$$

Посмотрим, как изменится дисперсия, если случайную величину умножить на число:

$$\mathbb{D}(a\xi) = \mathbb{E}(a\xi - \mathbb{E}(a\xi))^2 = \mathbb{E}(a(\xi - \mathbb{E}\xi))^2 = a^2 \mathbb{E}(\xi - \mathbb{E}\xi)^2 = a^2 \mathbb{D}\xi \quad (11)$$

Пусть ξ_1, ξ_2 — некоторые случайные величины. Посчитаем дисперсию их суммы:

$$\begin{aligned} \mathbb{D}(\xi_1 + \xi_2) &= \mathbb{E}((\xi_1 + \xi_2) - \mathbb{E}(\xi_1 + \xi_2))^2 = \mathbb{E}((\xi_1 - \mathbb{E}\xi_1) + (\xi_2 - \mathbb{E}\xi_2))^2 = \\ &= \mathbb{E}(\xi_1 - \mathbb{E}\xi_1)^2 + 2\mathbb{E}(\xi_1 - \mathbb{E}\xi_1)(\xi_2 - \mathbb{E}\xi_2) + \mathbb{E}(\xi_2 - \mathbb{E}\xi_2)^2 = \\ &= \mathbb{D}\xi_1 + \mathbb{D}\xi_2 + 2\text{cov}(\xi_1, \xi_2) \end{aligned} \quad (12)$$

Здесь мы использовали обозначение

$$\text{cov}(\xi_1, \xi_2) = \mathbb{E}(\xi_1 - \mathbb{E}\xi_1)(\xi_2 - \mathbb{E}\xi_2) = \mathbb{E}(\xi_1\xi_2) - \mathbb{E}\xi_1\mathbb{E}\xi_2. \quad (13)$$

Величина (13) называется *ковариацией* случайных величин ξ_1 и ξ_2 .

Пример 11. Найти математическое ожидание и дисперсию случайной величины $\xi \sim \text{Be}(p)$.

$$\mathbb{E}\xi = 0 \cdot \mathbb{P}(\xi = 0) + 1 \cdot \mathbb{P}(\xi = 1) = 0 \cdot (1 - p) + 1 \cdot p = p.$$

$$\mathbb{D}\xi = \mathbb{E}(\xi - p)^2 = (0 - p)^2 \cdot (1 - p) + (1 - p)^2 \cdot p = p^2(1 - p) + p(1 - p)^2 = p(1 - p)(p + 1 - p) = p(1 - p).$$

Можно было воспользоваться формулой (10) и тем, что $\mathbb{E}(\xi^2) = \mathbb{E}(\xi) = p$:

$$\mathbb{D}\xi = \mathbb{E}(\xi^2) - (\mathbb{E}\xi)^2 = p - p^2 = p(1 - p).$$

Пример 12. Бросают n кубиков. Найти математическое ожидание суммы выпавших очков.

Решение. Имеем N независимых случайных величин X_1, X_2, \dots, X_N — сколько очков выпало на каждом кубике. Математическое ожидание каждой из них равно $\frac{1+2+3+4+5+6}{6} = \frac{7}{2}$. Тогда по линейности математического ожидания,

$$\mathbb{E}\left(\sum_{i=1}^N X_i\right) = \sum_{i=1}^N \mathbb{E}X_i = N \cdot \mathbb{X}_1 = \frac{7N}{2}$$

Опр. Условное математическое ожидание случайной величины ξ при условии события A вычисляется по формуле:

$$\mathbb{E}(\xi|A) = \sum_{x \in X} \xi(x) \mathbb{P}(x|A) \quad (14)$$

Пусть имеется набор попарно непересекающихся событий B_1, \dots, B_n , такой что $\mathbb{P}\left(\bigsqcup_{i=1}^n B_i\right) = 1$. Используя (3) можно вывести формулу полной вероятности для математического ожидания:

$$\mathbb{E}\xi = \sum_{x \in X} \xi(x) \mathbb{P}(x) = \sum_{x \in X} \xi(x) \sum_{i=1}^n \mathbb{P}(x|B_i) \mathbb{P}(B_i) = \sum_{i=1}^n \mathbb{P}(B_i) \sum_{x \in X} \xi(x) \mathbb{P}(x|B_i) = \sum_{i=1}^n \mathbb{P}(B_i) \mathbb{E}(\xi|B_i). \quad (15)$$

Пример 13. Монетку бросают до тех пор, пока не выпадет решка два раза подряд. Найти математическое ожидание числа бросков.

Решение. Пусть Z — количество бросков. Будем обозначать $p_1 = 0$, если при первом броске выпал орёл, и $p_1 = 1$ если решка.

Применим формулу полной вероятности для математического ожидания (15):

$$\mathbb{E}Z = \frac{1}{2} \mathbb{E}(Z|p_1 = 0) + \frac{1}{2} \mathbb{E}(Z|p_1 = 1).$$

Если при первом броске выпал орёл, то это ни на что не влияет — мы просто потеряли один ход и начинаем игру заново, т.е.

$$\mathbb{E}(Z|P_1 = 0) = 1 + \mathbb{E}Z$$

Если же при первом броске выпала решка, то с вероятностью $\frac{1}{2}$ на следующей броске тоже выпадет решка, и мы будем иметь $Z = 2$. В противном случае выпадет орёл, и мы теряем два хода и начинаем игру снова:

$$\mathbb{E}(Z|P_1 = 1) = \frac{1}{2} \cdot 2 + \frac{1}{2} \cdot (2 + \mathbb{E}Z).$$

Обозначая $x = \mathbb{E}Z$, $a = (Z|P_1 = 0)$, $b = (Z|P_1 = 1)$, имеем систему:

$$\begin{cases} x = \frac{a+b}{2}, \\ a = 1 + x, \\ b = 2 + \frac{x}{2}. \end{cases}$$

Из неё получаем $x = \frac{3}{2} + \frac{3}{4}x$, откуда $x = 6$.

Итак, $\mathbb{E}Z = 6$.

Независимость случайных величин

Опр. Две случайные величины ξ и η называются *независимыми*, если для любых чисел c и d события $\xi = c$ и $\eta = d$ независимы.

Из определения следует свойство:

$$\begin{aligned} \mathbb{E}(\xi \cdot \eta) &= \sum_{(x,y) \in \text{dom} \xi \times \text{dom} \eta} \xi(x)\eta(y) \cdot \mathbb{P}(\xi = x \wedge \eta = y) = \\ &= \sum_{x \in \text{dom} \xi} \sum_{y \in \text{dom} \eta} \xi(x)\eta(y)\mathbb{P}(\xi = x)\mathbb{P}(\eta = y) = \sum_{x \in \text{dom} \xi} \xi(x)\mathbb{P}(\xi = x) \cdot \sum_{y \in \text{dom} \eta} \eta(y)\mathbb{P}(\eta = y) = \mathbb{E}\xi \cdot \mathbb{E}\eta \end{aligned} \quad (16)$$

Подставляя (16) в определение ковариации (13) получаем, что ковариация независимых случайных величин равна нулю. Применив этот факт к выражению (12) получаем важное утверждение: если случайные величины ξ_1 и ξ_2 независимы, то их дисперсия аддитивна:

$$\mathbb{D}(\xi_1 + \xi_2) = \mathbb{D}\xi_1 + \mathbb{D}\xi_2. \quad (17)$$

Аналогично можно перенести на случайные величины понятия попарной независимости и независимости в совокупности.

Опр. Случайные величины $\xi_1, \xi_2, \dots, \xi_n$ попарно независимы, если любые две из них независимы.

Опр. Случайные величины $\xi_1, \xi_2, \dots, \xi_n$ независимы в совокупности, если для любых чисел x_1, x_2, \dots, x_n события $\{\xi_i = x_i\}_{i=1}^n$ независимы в совокупности.

5. (16) При двух бросках игральной кости выпало X_1 и X_2 .

Вычислите $\mathbb{E}[\max\{X_1, X_2\}] + \mathbb{E}[\min\{X_1, X_2\}]$

6. (46)

1. Найдите математическое ожидание числа бросаний кости до первого выпадения двух шестерок.
2. Симметричную монетку бросают неограниченное число раз. Какая из последовательностей встретится раньше с большей вероятностью: РОР или РРО?

7. (46)

1. Из бара на улицу выходит турист. В одном конце улицы находится гостиница, в другом — полицейский участок. Вдоль улицы горят фонари и турист идёт от одного фонаря к другому случайным образом: с вероятностью p в сторону гостиницы, с вероятностью $1-p$ в сторону участка. Найдите вероятность того, что в конце концов он дойдёт до гостиницы, не заходя в участок.

Считайте, что фонари пронумерованы от 0 (гостиница) до n (участок), и бар находится у фонаря с номером m . Вычислите значение вероятности для $n = 10, m = 5, p = \frac{1}{3}$.

2. При тех же условиях найдите среднее расстояние (в фонарях), которое пройдёт турист.

8. (16) Покажите, что из попарной независимости случайных величин не следует их независимость в совокупности. Приведите явный контрпример.

9. (16) Найти математическое ожидание числа простых циклов длины r в случайном графе на n вершинах. Считаем, что в графе рёбра между каждой парой вершин независимо генерируются с вероятностью p . (Такая модель случайного графа называется моделью *Эрдёша-Реньи*.)

10. (26) Найдите математическое ожидание и дисперсию числа простых циклов длины r в случайной перестановке n элементов в предположении, что все возможные перестановки равновероятны.

Статистические свойства случайных величин

Закон больших чисел

Пусть имеется набор независимых, одинаково распределённых случайных величин ξ_1, ξ_2, \dots с конечной дисперсией.

Тогда можно показать, что $\forall \varepsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \frac{\sum_{i=1}^n \xi_i}{n} - \mathbb{E}\xi_1 \right| < \varepsilon \right) = 1. \quad (18)$$

Утверждение (18) называется *законом больших чисел*. Его смысл в том, что если мы возьмём много независимых реализаций одной и той же случайной величины, и посчитаем их среднее, то это среднее будет стремиться к математическому ожиданию этой случайной величины. Например, если мы будем много раз подбрасывать «честную» монетку, то количество выпавших «орлов», делённое на число бросков, будет стремиться к $\frac{1}{2}$.

Неравенства Маркова и Чебышева

Итак, мы знаем, что в среднем случайная величина принимает значения, близкие к своему математическому ожиданию. Однако иногда мы хотим знать, какова вероятность **уклонения** случайной величины от своего математического ожидания. Рассмотрим несколько неравенств, отвечающие на этот вопрос.

Пусть ξ — **неотрицательная** случайная величина, $a > 0$ — некоторое число. Тогда

$$\mathbb{E}\xi = \sum_{x \in X} \xi(x) \mathbb{P}(x) \geq \sum_{x: \xi(x) \geq a} \xi(x) \mathbb{P}(x) \geq a \cdot \sum_{x: \xi(x) \geq a} \mathbb{P}(x) = a \cdot \mathbb{P}(x \geq a).$$

Отсюда получаем *неравенство Маркова*:

$$\mathbb{P}(\xi \geq a) \leq \frac{\mathbb{E}\xi}{a} \quad (19)$$

Теперь применим неравенство Маркова к случайной величине $(\xi - \mathbb{E}\xi)^2$ (которая всегда неотрицательна), и $a = \varepsilon^2$, где $\varepsilon > 0$:

$$\begin{aligned} \mathbb{P}((\xi - \mathbb{E}\xi)^2 \geq \varepsilon^2) &\leq \frac{\mathbb{E}(\xi - \mathbb{E}\xi)^2}{\varepsilon^2}. \\ \mathbb{P}(|\xi - \mathbb{E}\xi| \geq \varepsilon) &\leq \frac{\mathbb{D}\xi}{\varepsilon^2}. \end{aligned} \quad (20)$$

Неравенство (20) называется *неравенством Чебышева*. Его смысл в том, что вероятность того, что случайная величина примет значение вне некоторой окрестности математического ожидания быстро убывает с ростом размера этой окрестности (скорость убывания — хотя бы квадратична).

Эквивалентная формулировка неравенства Чебышева:

$$\mathbb{P}(|\xi - \mathbb{E}\xi| < \varepsilon) \geq 1 - \frac{\mathbb{D}\xi}{\varepsilon^2}. \quad (21)$$

Вероятностные алгоритмы

Рассмотрим, зачем нужна теория вероятности при изучении алгоритмов.

Во-первых, в реальной жизни в работе компьютерных программ может иметь место элемент неопределённости.

Пример 14. Мы хотим получить веб-страницу с сильно нагруженного сервера. Мы посылаем запрос. Ровно через секунду с вероятностью $\frac{1}{2}$ мы получаем веб-страницу, и с вероятностью $\frac{1}{2}$ получаем ошибку 408. Во втором случае мы автоматически посылаем запрос снова. Найти математическое ожидание времени, через которое мы получим страницу.

Решение. Обозначим время получения страницы t . Тогда $\mathbb{P}(t = 1) = \frac{1}{2}$, $\mathbb{P}(t = 2) = \frac{1}{4}$, и т.д. В общем случае, $\mathbb{P}(t = k) = \frac{1}{2^k}$ для $k = 1, 2, \dots$

По определению математического ожидания,

$$\mathbb{E}t = \sum_{k=1}^{\infty} k\mathbb{P}(t = k) = \sum_{k=1}^{\infty} \frac{k}{2^k} = 2$$

Мы использовали тот факт, что если $a \in (0, 1)$, то

$$\sum_{k=0}^{\infty} ka^k = \frac{a}{(1-a)^2}. \quad (22)$$

Это можно доказать следующим образом. Пусть $f(a) = \frac{1}{1-a}$. Тогда f можно разложить в ряд Тейлора: $f(a) = \sum_{k=0}^{\infty} a^k$.

Продифференцируем этот ряд по a : $f'(a) = \sum_{k=1}^{\infty} k \cdot a^{k-1} = \frac{1}{a} \sum_{k=1}^{\infty} k \cdot a^k$.

С другой стороны, $f'(a) = \frac{1}{(1-a)^2}$, так что $\frac{1}{a} \sum_{k=1}^{\infty} ka^k = \frac{1}{(1-a)^2}$. ■

Иногда элемент неопределённости мы можем вносить искусственно с помощью *генератора случайных чисел*. Пока будем считать, что это некоторая внешняя по отношению к алгоритму процедура, которая выдаёт реализации некоторой случайной величины с известным распределением.

Пример 15 (Вычисление площади методом Монте-Карло). Имеется непрерывная функция $f : [0, 1] \rightarrow [0, 1]$. Требуется вычислить $S = \int_0^1 f(x)dx$.

Давайте много раз равномерно случайно «бросать» точку и проверять, попала ли она под нашу кривую. Тогда отношение количества таких «попавших» точек к количеству всех точек будет оценкой площади.

Более формально, рассмотрим набор независимых одинаково распределённых случайных величин

$$x_1, \dots, x_N, y_1, \dots, y_n \sim U[0, 1]$$

и случайную величину

$$S' = \frac{1}{N} \sum_{i=1}^N [y_i \leq f(x_i)] \quad (23)$$

Квадратными скобками обозначается индикатор логического выражения. Он равен 1 если выражение истинно, и 0, если оно ложно.

Заметим, что $\mathbb{P}(y_i \leq f(x_i)) = S$, так что каждое слагаемое - Бернуллиевская случайная величина с параметром S . Значит, согласно закону больших чисел (18), увеличивая N мы можем сделать S' сколь угодно близкой к S с вероятностью, сколь угодно близкой к единице.

На практике мы сгенерируем $2N$ вещественных чисел с помощью генератора случайных чисел (равномерно на $[0, 1]$), и подставим их в формулу (23). ■

Иногда мы можем не использовать генератор случайных чисел, но использовать процедуру, обладающую вероятностными свойствами.

Пример 16. Пусть имеется вероятностный тест на простоту — некоторый алгоритм $q : \{2, 3, \dots\} \rightarrow 0, 1$, такой, что:

- $\mathbb{P}(q(x) = 1 | x \text{ — простое}) = 1$;
- $\mathbb{P}(q(x) = 1 | x \text{ — составное}) \leq \varepsilon < 1$.

Другими словами, q всегда точно распознаёт простые числа, но составные числа иногда он может по ошибке называть простыми, но вероятность такой ошибки всегда не больше ε .

Построим следующий алгоритм проверки числа x на простоту.

```

1 def isPrime(x):
2     for i in range(n):
3         if q(x)==0:
4             return False
5     return True

```

Если x — простое, то алгоритм гарантированно вернёт True. А вот если x — составное, то алгоритм даст правильный ответ, если тест q не ошибся **хотя бы один раз**.

Вероятность того, что алгоритм q ошибается не больше ε . Значит, вероятность того, что он ошибается **каждый раз**, не больше ε^N .

Даже если $\varepsilon = 0.9$, если взять хотя бы $N = 500$, вероятность ошибки итогового алгоритма будет не больше $1.3 \cdot 10^{-23}$, чем можно пренебречь при любом практическом применении. ■

Итак, будем называть вероятностным алгоритм, последовательность операций которого зависит от случайных событий. В более узком смысле вероятностный алгоритм — алгоритм использующий генератор случайных чисел.

Существует много самых разных задач, решать которые с помощью вероятностных алгоритмов намного проще (как по сложности реализации, так и по асимптотике времени работы), чем с помощью детерминированных.

Более того, для многих важных задач не существует детерминированных алгоритмов, которые выдают правильный ответ за разумное время, но есть детерминированные алгоритмы, которые быстро дают приближённый, но пригодный для практических целей ответ.

Несколько таких алгоритмов вы изучите в курсе АМВ, но ещё больше в дальнейшем, в самых различных областях (особенно в машинном обучении, где почти все алгоритмы вероятностные).

Время в среднем и в худшем случае

На лекции 1 вы узнали, как оценивать время работы алгоритмов. Когда мы имеем дело с вероятностными алгоритмами и хотим применять их на практике, мы тоже должны доказать, что алгоритм будет выполняться за разумное время. Однако бывает так, что мы можем себе позволить, чтобы иногда, очень редко, алгоритм работал долго. Поэтому при оценке времени работы рассматривают оценку «в среднем» и «в худшем случае». Формализуем эти понятия.

Опр. Алгоритм имеет время работы *в худшем случае* T , если время его работы на любом входе не больше T .

Опр. Алгоритм имеет время работы *в среднем* T , если мат. ожидание времени его работы равно T .

Например, в примере 14 время работы в худшем случае равно бесконечности.

Пример 17. Имеется массив из $2N \gg 1$ битов, в котором число нулей и единиц одинаково. Необходимо найти какой-нибудь элемент массива, равный единице. Пусть алгоритм последовательно проверяет биты от первого до последнего и останавливается, когда находит единицу. В худшем случае все единицы будут сконцентрированы в правой части, и потребуются сделать $N + 1$ проверок, т.е. сложность алгоритма в худшем случае будет $\mathcal{O}(N)$. Однако в среднем можно считать появление единицы на i -м шаге бернуллиевской случайной величиной. Тогда математическое ожидание номера первой единицы будет равно 2, а сложность алгоритма в среднем $\mathcal{O}(1)$.

Со временем работы в худшем случае вы уже знакомы. Время же работы в среднем обычно оказывается бесполезным на практике, так как предполагает, что вам известно распределение входных данных как случайной величины, что обычно не так. Но эту проблему иногда можно решить, искусственно добавив элемент случайности. Так, в примере выше будет достаточно идти по элементам не последовательно, а в некотором случайном порядке. Это позволит нам симитировать равномерное распределение на входных данных для исходного алгоритма. Обобщить данный подход можно понятием *рандомизированной* сложности.

Опр. Пусть помимо непосредственно входа *input* алгоритм каким-либо образом получил набор r случайных бит. Тогда *рандомизированной* сложностью алгоритма называют худшее по всем входам мат. ожидание времени работы как случайной величины определённой на пространстве наборов r .

Чтобы придать этим определениям более явный смысл, скажем, что $T(r, \text{input})$ – время работы алгоритма на входе *input* при наборе случайных бит r . В частности, пустой набор случайных бит будем обозначать \emptyset . Тогда худшее, среднее и рандомизированное время работы как функция от размера входа n запишутся следующим образом:

$$\begin{aligned} f_{\text{worst}}(n) &= \max_{|\text{input}|=n} T(\emptyset, \text{input}), \\ f_{\text{avg}}(n) &= \mathbb{E}_{|\text{input}|=n} T(\emptyset, \text{input}), \\ f_{\text{rand}}(n) &= \max_{|\text{input}|=n} \mathbb{E}_r T(r, \text{input}) \end{aligned}$$

Рандомизированное время работы алгоритма говорит о нём куда больше, чем время работы в среднем, потому что хотя рандомизированный алгоритм и может работать долго на некоторых конкретных наборах входов и случайных бит, не существует какого-то входного набора, который гарантировал бы плохое время работы. Иначе говоря, если наш алгоритм работает медленно на некотором входе, мы можем запустить его заново и ожидать, что вероятность повторения такой ситуации будет достаточно мала.

Здесь в свою очередь также будет играть важную роль дисперсия, так как она позволит оценить вероятность того, что нам попадётся достаточно плохой набор бит r , чтобы отдалить нас от мат. ожидания на некоторую фиксированную величину ε . При известной дисперсии для этого можно использовать, например, неравенство Чебышева.

11. (26) Предложите вероятностный алгоритм, который по заданному числу n генерирует случайную перестановку из чисел $1, \dots, n$ (с равномерным распределением) и работает за линейное время $\mathcal{O}(n)$.

12. (26) Предложите алгоритм, имеющий рандомизированное время работы $\mathcal{O}(n)$, который для данного массива из n элементов будет проверять есть ли в нём элемент, который встречается не меньше $\frac{n}{2}$ раз.

1*. (16) Предложите вероятностный алгоритм, который по заданному графу G генерирует случайное остовное дерево графа (с равномерным распределением) и работает за время $\mathcal{O}(E + V)$, где E – число рёбер графа, V – число вершин графа. То есть если у данного графа N различных остовных деревьев, ваш алгоритм должен выдавать каждое из них с вероятностью $\frac{1}{N}$.

Список литературы

- [1] Каноническое задание по курсу АМВ, 2018 год, [pdf](#).
- [2] *М.Я. Кельберт, Ю.М. Сухов*. Основные понятия теории вероятностей и математической статистики. М.:МЦНМО, 2007. - 456с.
- [3] *М.Б. Лагутин*. Наглядная математическая статистика. М.:Бином, 2015. - 472с.
- [4] *А.А. Натан, О.Г. Горбачев, С.А. Гуз*. Теория вероятностей. М.:МЗ Пресс, 2007. - 253 с.
- [5] *Прохоров А.В., Ушаков В.Г., Ушаков Н.Г.* Задачи по теории вероятностей. М.:Наука, 1986. - 328с.
- [6] *Hoeffding, Wassily*. Probability inequalities for sums of bounded random variables. Journal of the American statistical association 58.301 (1963): 13-30.