

Содержание

Кратко про теорию множеств и отображения	1
Множества	1
Отображения множеств	3
Мощность множества	5
Вероятностное пространство	6
Мотивировка	6
Математический подход	6
Как задать вероятность P ?	9
Классическое распределение	9
Четыре задачи о шарах	11
Еще задачи	14
Условная вероятность	17
Независимость событий	24
Схема Бернулли	25
Задачи	26
Случайные величины	28
Мотивация	28
Начало	28
Определение	29
Базовые распределения	35
Равномерное распределение, $X \sim U\{0, \dots, N\}$	35
Распределение Бернулли, $X \sim Bern(p)$	35
Биномиальное распределение, $X \sim Binom(n, p)$	36
Геометрическое распределение, $X \sim Geom(p)$	37
Распределение Пуассона, $X \sim Poisson(\lambda)$	37
Совместные распределения	38
Математическое ожидание и дисперсия	42
Коэффициент корреляции Пирсона	46
Задачи	46
Пределы	52
Закон больших чисел	52
Центральная предельная теорема	53
Задачи	53
Сюжеты	53
Счастливые билеты	53

Кратко про теорию множеств и отображения

Множества

Опр. Множество элементов (set) — неопределяемое понятие, описываемое синонимами (группа, класс, набор). Нет определения, но есть понимание.

Обозначается, как правило, заглавными буквами: A, B, \dots . Элементы множества обозначаются обычно маленькими буквами: a, b, \dots .

Приведем небольшой словарь обозначений:

- $a \in A$ — "элемент a принадлежит множеству A " (от буквы ϵ в слове "elementum")
- $A \ni a$ — "множество A содержит a "
- $a \notin A$ — " a не принадлежит A "
- $A \subset B$ или $A \subseteq B$ — "множество A является подмножеством множества B "
- $B \supset A$ или $B \supseteq A$ — " B содержит A "

Если множество не содержит элементов, то оно называется пустым и обозначается \emptyset .

Множества A и B называются равными, если $A \subset B$ и $B \subset A$. Обозначается $A = B$. По-простому говоря, множества равны, если они состоят из одних и тех же элементов.

Множества можно задавать явно (перечислением) и неявно (заданием свойств, которому оно должно удовлетворять). В случае явного задания используется перечисление элементов внутри фигурных скобок: $\{a, b, c, d, \dots\}$. Тип скобок важен, потому что круглые скобки используются для обозначения упорядоченного массива элементов: (a, b, c, d, \dots) .

Если элементы множества A проиндексированы элементами другого множества I , то пишут $A = \{a_i\}_{i \in I}$. Например, если $A = \{a_1, a_2, a_3, \dots, a_{10}\}$, а $I = \{1, 2, 3, \dots, 10\}$, то пишут $A = \{a_i\}_{i \in I}$ или же так $A = \{a_i\}_{i=1}^{10}$.

Часто множество задается неявно описанием свойств тех элементов, из которых оно состоит. Записывается это так: $C = \{c : \text{перечисление свойств}\}$, а читается "множество C состоит из элементов c , таких, что каждый элемент c удовлетворяет свойствам...".

Упражнение. Какие из формул верны:

$$1) \emptyset \in \{\emptyset, \{\emptyset\}\}; \quad 2) \{\emptyset\} \in \{\{\emptyset\}\}; \quad 3) \emptyset \in \{\{\emptyset\}\}?$$

Над множествами можно определить такие часто используемые операции:

- $A \cup B = \{c : c \in A \text{ или } c \in B\}$ — объединение
- $A \cap B = \{c : c \in A \text{ и } c \in B\}$ — пересечение
- $A \setminus B = \{c : c \in A \text{ и } c \notin B\}$ — вычитание

Упражнение. Доказать, что

$$A \cap A = A, \quad A \cup A = A, \quad A \cup \emptyset = A, \quad A \cap \emptyset = \emptyset.$$

▷ Докажем, например, первое утверждение, чтобы почувствовать, что здесь требуется сделать.

Итак, требуется показать, что $A \cap A = A$. По определению

$$A \cap A = A \Leftrightarrow A \cap A \subset A \text{ и } A \subset A \cap A.$$

То есть нужно доказать вложенность в обе стороны.

Итак, $A \cap A \subset A \Leftrightarrow \forall a \in A \cap A$ верно, что $a \in A$. Но утверждение $a \in A \cap A$ по определению означает, что $a \in A$ и $a \in A$, что эквивалентно тому, что просто $a \in A$, что и требовалось.

Теперь посмотрим вложенность в обратную сторону. Утверждение $A \subset A \cap A \Leftrightarrow \forall a \in A$ верно, что $a \in A \cap A$. Это верно по определению пересечения. Значит, верно и требуемое утверждение.

Обе вложенности верны, значит, все доказано. \triangleleft

Упражнение. Доказать, что операции пересечения и объединения коммутативны:

$$A \cap B = B \cap A, \quad A \cup B = B \cup A.$$

Упражнение. Доказать, что операции пересечения и объединения ассоциативны:

$$(A \cap B) \cap C = A \cap (B \cap C), \quad (A \cup B) \cup C = A \cup (B \cup C).$$

Упражнение. Доказать, что операции пересечения и объединения дистрибутивны:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C), \quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

Упражнение. Доказать, что

- $A \cap B = \overline{\overline{A} \cup \overline{B}}, \quad A \cup B = \overline{\overline{A} \cap \overline{B}};$
- $A \setminus (\bigcup_{\alpha} A_{\alpha}) = \bigcap_{\alpha} (A \setminus A_{\alpha}), \quad A \setminus (\bigcap_{\alpha} A_{\alpha}) = \bigcup_{\alpha} (A \setminus A_{\alpha}).$

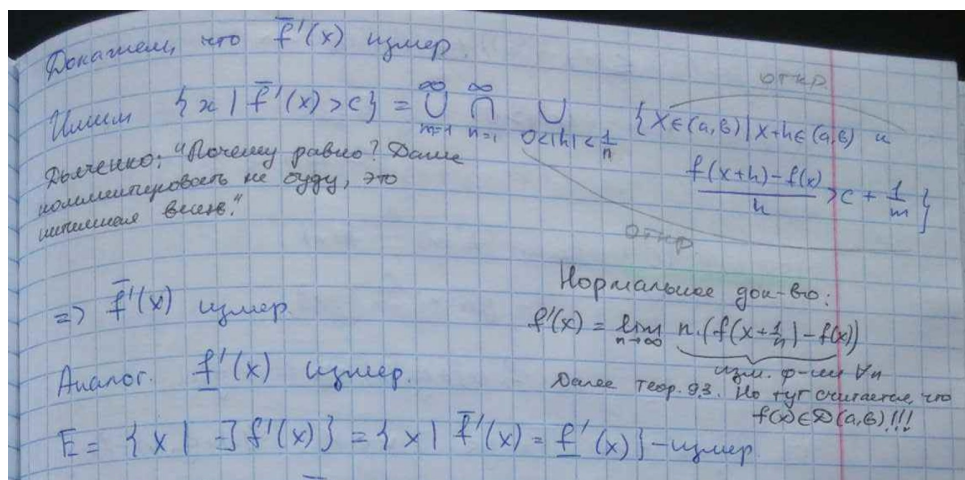


Рис. 1: Пример страшного равенства множеств из лекций по действительному анализу М.И. Дьяченко

Отображения множеств

Опр. Отображением F множества X в множество Y (mapping) называется сопоставление каждому элементу $x \in X$ некоторого элемента $y \in Y$. Тогда пишут, что $y = F(x)$, а $F : X \rightarrow Y$ (" F действует из X в Y ").

Если хотят полностью описать отображение, то используется обозначение

$$F : \begin{cases} X \rightarrow Y & (\text{откуда и куда действует}) \\ x \mapsto y = F(x) & (\text{правило действия}) \end{cases}$$

Понятие отображения обобщает понятие числовой функции, которая является отображением числовой прямой в саму себя ($X = Y = \mathbb{R}$). Например, функция $f(x) = x + 1$ есть отображение

$$f : \begin{cases} \mathbb{R} \rightarrow \mathbb{R} \\ x \mapsto x + 1 = y \end{cases}$$

Заметим, что в данном случае существует обратное отображение

$$f^{-1} : \begin{cases} \mathbb{R} \rightarrow \mathbb{R} \\ y \mapsto y - 1 = x \end{cases}$$

Но если бы мы взяли функцию $f(x) = 0$, отображающую все числа в ноль, то для нее уже не существует обратной, потому что все числа "склеились" в одно.

Опр. Отображение $F : X \rightarrow Y$ называется инъективным, если никакие два различных элемента $x_1, x_2 \in X$ не склеиваются в один элемент $y \in Y$, т.е. $F(x_1) \neq F(x_2)$.

Опр. Отображение $F : X \rightarrow Y$ называется сюръективным, если каждому элементу $y \in Y$ соответствует хотя бы один элемент $x \in X$.

Опр. Отображение $F : X \rightarrow Y$ называется взаимно однозначным (или биективным, или обратимым), если оно одновременно и инъективно, и сюръективно.

* * *

Упражнение. Построить все возможные отображения $F : X \rightarrow Y$, где

- 1) $X = \{x\}, Y = \{y\},$ 2) $X = \{x_1, x_2\}, Y = \{y\},$
- 3) $X = \{x\}, Y = \{y_1, y_2\},$ 4) $X = \{x_1, x_2\}, Y = \{y_1, y_2\}$

и указать, являются ли эти отображения инъективными, сюръективными и биективными.

Вопрос. Пусть X — множество всех людей, Y — все даты рождения (без года), F — сопоставление человеку даты его рождения. Является ли это отображение инъективным, сюръективным и биективным?

* * *

Признаться честно, мы дали "наивное" определение отображения. И для текущих целей этого достаточно, поэтому следующие несколько абзацев можно пропускать. Здесь мы дадим еще несколько формальных определений, чтобы дать определение отображению на теоретико-множественном языке.

Опр. Множество вида $\{a, b\}$ называется парой.

Опр. Множество вида $\{a, \{a, b\}\}$ называется упорядоченной парой и обозначается (a, b) , элемент a называется первым, а элемент b — вторым.

Опр. Пусть X, Y — произвольные множества. Множество всех упорядоченных пар $\{(x, y) : x \in X, y \in Y\}$ называется прямым произведением X и Y и обозначается $X \times Y$.

Опр. Пусть $F \subset X \times Y$. Если $\forall (x', y'), (x'', y'') \in F$ выполняется условие

$$x' = x'' \Rightarrow y' = y'',$$

(т.е. одному x не может соответствовать сразу два y), то F называется отображением из X в Y .

Множество всех x , таких, что $\exists y \in Y : (x, y) \in F$ называется областью определения F , обозначается D_F .

Множество всех y , таких, что $\exists x \in X : (x, y) \in F$ называется образом F , обозначается $F(X)$ или $\text{Im}(F)$ (от слова image).

Упр. Сформулировать определение инъективности, сюръективности и биективности на языке теории множеств.

Мощность множества

В интуитивном понимании мощность множества (cardinality) это количество элементов в нем. Но это понимание годится только если мы имеем дело с множествами из конечного числа элементов. А это очень сильное ограничение, учитывая, что хотя бы числовые множества (начиная с натуральных чисел) уже не конечны. Поэтому приходится копать глубже.

Идея здесь такова. Конечные множества мы и так хорошо понимаем с точки зрения их "размера", это нужно сохранить. В иных случаях посчитать количество элементов мы не можем, но зато можем их сравнивать, пусть и не имея конкретного числа, характеризующего количество элементов в нем.

* * *

Займемся формализацией приведенной идеи.

Опр. Пусть X, Y — множества. Если существует взаимно однозначное отображение $F : X \rightarrow Y$, то говорят, что множества X и Y равномощны (имеют одинаковую мощность). Обозначается $|X| = |Y|$.

Опр. Если X равномощно множеству $\{1, 2, \dots, n\}$, то X называется n -элементным, а n называется количеством элементов в X . Обозначается $|X| = n$ (иногда $\text{card}(X)$).

Опр. Количество элементов в \emptyset называется ноль и обозначается 0.

Опр. Пустое множество и n -элементное называются конечными множествами (пишут $|X| < \infty$). Если множество не является конечным, то оно называется бесконечным.

Опр. Если X равномощно \mathbb{N} , то X называется счетным множеством.

* * *

Хорошо, есть конечные множества, есть бесконечные. В бесконечных мы выделили счетные. А есть ли бесконечные, но не счетные? Можно показать, что \mathbb{N} , \mathbb{Z} , \mathbb{Q} — счетные множества, но \mathbb{R} — несчетно. Причем \mathbb{R} равномощно отрезку $[0, 1]$. Множества, равномощные $[0, 1]$, называются множествами мощности континуум.

А вот есть ли среди бесконечных множества, имеющие не счетную и не континуумную мощность — хороший вопрос. Если интересна эта тема, стоит почитать про континуум-гипотезу.

* * *

Упражнение. Является ли множество $\{\{\emptyset\}\}$ одноэлементным?

Упражнение. Сколько элементов содержат следующие множества?

- | | | |
|---|---|-------------------------------------|
| 1) $\{1, 2, 1\}$; | 2) $\{1, 2, \{1, 2\}\}$; | 3) $\{\{2\}\}$; |
| 4) $\{\{1\}, 1\}$; | 5) $\{1, \emptyset\}$; | 6) $\{\{\emptyset\}, \emptyset\}$; |
| 7) $\{\{\emptyset\}, \{\emptyset\}\}$; | 8) $\{x, 3x - 1\}$, где $x \in \mathbb{R}$. | |

Упражнение. Найти мощность множества всех подмножеств данного множества X , если $|X| = N$.

Упражнение. Найти мощность множества всех отображений $X \rightarrow Y$, если $|X| = N$, $|Y| = M$.

* * *

Подробнее про теорию множеств можно почитать в книге [17], более продвинуто — в книге Верещагина и Шеня [15] или еще подробнее в книге Куратовского и Мостовского [10].

Вероятностное пространство

Мотивировка

Представим себя в роли исследователя какого-то явления, вообще говоря, неважно какой природы. Например, в роли физика-экспериментатора. Задача заключается в том, чтобы понять, каким будет результат эксперимента. В некоторых случаях у нас есть достаточно инструментов для того, чтобы точно указать ответ. То есть результат можно предсказать достоверно. Конечно, всегда может упасть луна с неба, и наш расчет не поможет, но такие погрешности нам скорее всего простят.

Однако не всегда есть достаточно информации, умений, возможностей, чтобы дать достоверный ответ. Тогда можно либо развести руками, либо засучить рукава и искать способ получить достоверный ответ, либо попробовать что-то придумать в имеющихся условиях и получить ответ, хоть и не достоверный, но тем не менее полезный.

Тогда разумно возникает идея разделения достоверности на две составляющие: детерминированную (определенную) и случайную (неопределенную). Детерминировано то, что мы знаем о явлении, а случайно то, чего не знаем. Отсюда видно, что разделение это субъективно: у инсайдера гораздо больше информации, чем у обычного трейдера.

А теперь уже и становится ясно, что под вероятностью следует понимать меру того неполного знания, которое имеется у нас об исследуемом явлении.

Итак, концепция должна быть ясна, но теперь нужно придать введенным терминам больше точности, чтобы мочь с ними работать. Для этого нужно перевести внимание со случайности на само исследуемое явление, то есть на его свойства. Мы будем предполагать, что нам известны возможные состояния этого явления, иными словами, результаты эксперимента. А различные состояния принимаются с различной степенью достоверности.

Например, мы хотим узнать, какой завтра будет погода. У нас может быть метеостанция, а может не быть, но мы все равно можем с уверенностью сказать, что завтра не будет +50.

Таким образом, неопределенность распределяется между возможными состояниями. Если состояния равнозначны (подбрасывание монетки, подкидывание кости), то по ним распределяется неопределенность в равной степени. Но с прогнозом погоды это уже не так.

Математический подход

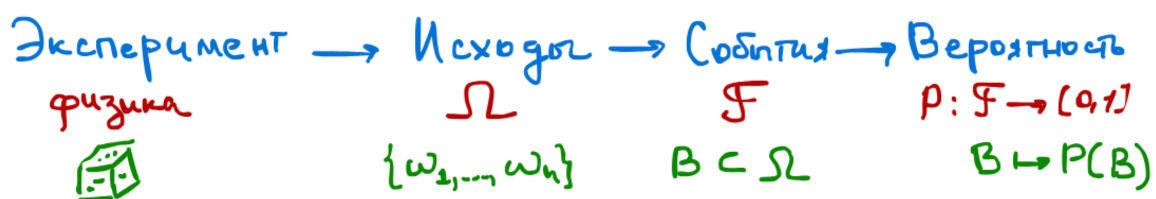


Рис. 2: Схема построения вероятностного пространства

Здесь мы переведем на язык теории множеств то, о чем было сказано выше. Изложение будет упрощенное, но все равно достаточно нагруженное новой терминологией. Этого хватит для того, чтобы решать задачи с конечным числом исходов и понять основные методы работы с вероятностью.

Опр. Пространством элементарных исходов (sample space) называется множество, содержащее все возможные взаимоисключающие результаты данного случайного эксперимента, обычно обозначается Ω . Элементы Ω называются элементарными исходами (outcomes) и обозначаются через ω .

Опр. Любое подмножество $A \subset \Omega$ называется событием (event). Говорят, что произошло событие A , если эксперимент завершился одним из элементарных исходов $\omega \in A$.

Опр. Пусть \mathcal{M} — система подмножеств Ω . Функция $m : \mathcal{M} \rightarrow [0, +\infty)$ называется мерой (measure), если $\forall A, B \in \mathcal{M} : A \cap B = \emptyset$ и $A \cup B \in \mathcal{M}$ верно, что

$$m(A \cup B) = m(A) + m(B).$$

Если $m(\Omega) = 1$, то мера m называется вероятностной.

Опр. Тройка (Ω, \mathcal{F}, P) называется вероятностным пространством (probability space), если Ω — пространство элементарных исходов, \mathcal{F} — множество событий (event space), P — вероятностная мера на \mathcal{F} (или просто — вероятность).

При этом система \mathcal{F} должна обладать свойством замкнутости относительно операций объединения и пересечения (объединение/пересечение двух множеств из \mathcal{F} есть множество из \mathcal{F}), ведь если мы умеем вычислять вероятность двух событий, то хотим уметь вычислять и вероятность их одновременного выполнения или что хотя бы одно из них выполнилось. Такие системы множеств называются алгебрами.

Задание вероятности P на алгебре \mathcal{F} называется заданием распределения вероятности.

* * *

Итак, для построения теории вероятностей, мы использовали теорию множеств (чтобы задать исходы Ω и события \mathcal{F} , а также чтобы использовать понятие отбраковки для задания вероятности P), плюс к этому потребовали от P выполнения трех свойств:

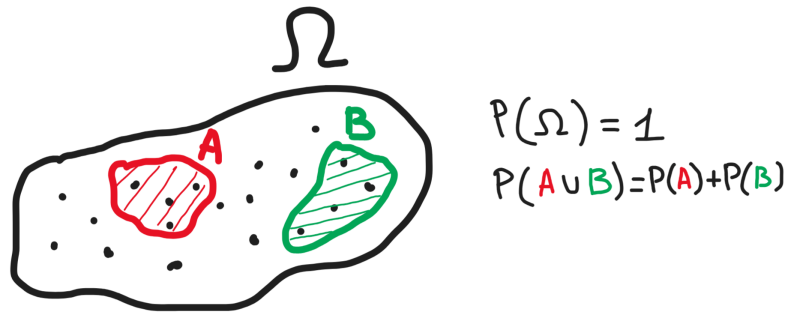
1. $\forall B \in \mathcal{F} \quad P(B) \geq 0$
2. $P(A \cup B) = P(A) + P(B)$ для непересекающихся событий A и B
3. $P(\Omega) = 1$

Эти свойства являются аксиомами теории вероятностей. Они были введены А.Н. Колмогоровым в 1933 году в статье, которая в последующем была опубликована на русском языке в [6].

Здесь может быть полезно немного остановиться и подумать, почему именно такие условия наложены на вероятность.

Первое условие объясняется тем, что мы хотим, чтобы наименьшее, что мы можем отмерить есть ничего, то есть ноль. Чуть ниже мы покажем, что мера пустого множества всегда будет нулевой в рамках данного определения. Отмерить множество, которое будет меньше пустого, нельзя.

Второе условие тоже понятно: оно показывает, что мы хотим уметь вычислять вероятности "сложных" событий, полученных из простых теоретико-множественными



операциями, причем вполне естественным образом: мера объединения непересекающихся множеств равна сумме их мер.

Третье условие утверждает, что событие, состоящее в том, что произойдет хоть что-то, является достоверным. Достоверности приписывается число 1, хотя, конечно, это не принципиально с математической точки зрения. Важно тут то, что эта аксиома специфична именно для теории вероятностей, потому что она по сути реализует концепцию "достоверности".

Вообще хотелось бы заметить, что на этой паре страниц произошла магия. Мы совершенно простыми — как философски, так и математически — рассуждениями сумели перевести в мир чисел то, что, если вдуматься, бесконечно от них далеко. И дальше из этих трех аксиом вытекает огромная чисто математическая теория.

Обратим еще внимание, что в аксиомах нет никакого намека на связь вероятности с частотностью. Эту связь надо доказывать, и делается это нетривиально. Теорема об этом называется Законом больших чисел.

* * *

Несколько слов об обозначениях.

Когда рассматривается вероятностное пространство с конечным числом исходов, алгебра событий \mathcal{F} обычно определяется как множество всех подмножеств Ω . Обозначается это так: $\mathcal{F} = 2^\Omega$.

Также, для краткости, часто опускают фигурные скобки под знаком $P(\cdot)$, например, вместо $P(\{\omega\})$ пишут $P(\omega)$. Строго говоря, это неграмотно, но подобных общепринятых сокращений будет в дальнейшем немало.

Обозначение $A \sqcup B$ означает, что объединяются непересекающиеся множества, т.е. $A \cap B = \emptyset$.

* * *

Теорема (свойства вероятности).

- (1) $P(\emptyset) = 0$
- (2) $P(\bar{A}) = 1 - P(A)$
- (3) $A \subset B \Rightarrow P(A) \leq P(B)$, $P(B \setminus A) = P(B) - P(A)$
- (4) $P(A \cup B) = P(A) + P(B) - P(AB)$

- (1) $P(A \sqcup \emptyset) = P(A) = P(A) + P(\emptyset) \Rightarrow P(\emptyset) = 0$.
 (2) $P(\Omega) = P(A) + P(\bar{A}) = 1 \Rightarrow P(\bar{A}) = 1 - P(A)$.
 (3) $P(B) = P(A) + P(B \setminus A) \Rightarrow P(A) \leq P(B)$, $P(B \setminus A) = P(B) - P(A)$.

$$(4) P(A \cup B) = P(A) + P(B \setminus A), \text{ а } P(B \setminus A) = P(B) - P(AB). \quad \blacksquare$$

Пример. Подбрасывание монетки. Определим вероятностное пространство:

$$\begin{aligned}\Omega &= \{\text{орел, решка}\}, \\ \mathcal{F} &= \{\emptyset, \{\text{орел}\}, \{\text{решка}\}, \{\text{орел, решка}\}\}, \\ P(\{\text{орел}\}) &= P(\{\text{решка}\}) = 1/2.\end{aligned}$$

Вопрос. Почему именно $1/2$? Можно ли задать P как-нибудь иначе?

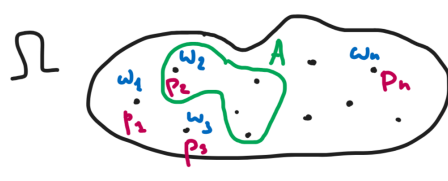
Как задать вероятность P ?

На одном и том же пространстве исходов, вероятность P можно задать различными способами, главное только, чтобы выполнялись аксиомы. Практически, то, как задано распределение, влияет на точность вероятностной модели относительно реального эксперимента. Но пока что мы посмотрим на этот вопрос с чисто теоретической точки зрения.

Допустим, мы как-то хотим задать распределение на Ω , где $|\Omega| = n$. Тогда, формально, надо определить P на всех 2^n подмножествах Ω . Но на самом деле эту задачу можно значительно упростить.

Пусть $\Omega = \{\omega_1, \dots, \omega_n\}$, $\mathcal{F} = 2^\Omega$. Тогда вероятность достаточно задать на каждом из элементарных исходов $P(\omega_i) = p_i$, а не на всем \mathcal{F} . Действительно, любое событие $A \in \mathcal{F}$ можно представить как объединение одноэлементных множеств, состоящих из элементарных исходов $\omega_{i_1}, \dots, \omega_{i_k}$ и тогда, по определению меры,

$$P(A) = \sum_{j=1}^k P(\omega_{i_j}) = \sum_{j=1}^k p_{i_j}.$$



$$\begin{aligned}A &= \{\omega_2, \omega_5, \omega_8, \omega_{17}, \omega_{20}\} \\ P(A) &= P(\omega_2) + P(\omega_5) + P(\omega_8) + P(\omega_{17}) + P(\omega_{20}) = \\ &= p_2 + p_5 + p_8 + p_{17} + p_{20}\end{aligned}$$

Классическое распределение

Опр. Если $|\Omega| = N$ и все $p_i = \frac{1}{N}$, то такое распределение называется классическим.

Классическое распределение соответствует эксперименту, в котором все элементарные исходы равновероятны: монетка считается абсолютно симметричной, кубик обладает идеальной формой, шары в урне неразличимы на ощупь и идеально перемешаны, карты в колоде равномерно перетасованы и т.д.

Понятно, что такие случаи, во-первых, в жизни в чистом виде не встречаются вообще, во-вторых, даже если простить монетке ее неидеальность, все равно классическое распределение покрывает лишь малую часть возможных экспериментов. Но слона нужно есть по кусочкам, поэтому мы и порешаем некоторые задачи на классическую вероятность. Не зря же она классическая.

Заметим только, что классическая вероятность приводит к школьной формуле вычисления вероятности события A :

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{число благоприятных исходов}}{\text{число всех исходов}}$$

* * *

Рассмотрим несколько поучительных примеров, показывающих тонкости работы с вероятностными пространствами с равновозможными исходами.

Пример. Шестигранный кубик имеет две красные стороны, три зеленых, одну белую. Эксперимент такой: бросаем кубик, смотрим, какой выпал цвет.

□ Построим вероятностное пространство:

$$\Omega = \{\text{Красный, Зеленый, Белый}\}, \quad \mathcal{F} = 2^\Omega.$$

Можно ли задать вероятность, как

$$P(\{\text{Красный}\}) = P(\{\text{Зеленый}\}) = P(\{\text{Белый}\}) = 1/3?$$

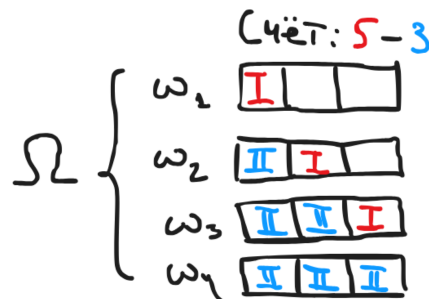
Математически — да, это никак не противоречит аксиомам. Однако, в этом случае выпадения разных граней кубика не будут равновозможными. Поэтому, если мы исходим из предположения о равновозможности выпадения граней кубика, то разумно распределить вероятности так:

$$P(\{\text{Красный}\}) = 2/6, \quad P(\{\text{Зеленый}\}) = 3/6, \quad P(\{\text{Белый}\}) = 1/6.$$

■

Пример. Два игрока подбрасывают монетку. Тот, кто первый победит 6 раз, тот побеждает. Игроков прервали на счете 5 : 3. Нужно разделить выигрыш. Как это сделать?

□ Выпишем возможные продолжения игры:



То есть в трех случаях из четырех побеждает первый игрок. Тогда, если рассматривать Ω как множество продолжений игры, то можно сделать вывод, что первый победит с вероятностью $3/4$.

Но в этом случае мы почему-то использовали то, что исходы равновероятны. С математической точки зрения, вероятность, определенная таким образом, вполне допустима. Однако, такое распределение не является разумным, так как, физически, равновероятными правильнее считать тройки подбрасываний монеты.

Тогда, первому исходу соответствует сразу четыре тройки, второму исходу — две, третьему — одна, четвертому — одна. И тогда,

$$P(\omega_1) = 4/8, \quad P(\omega_2) = 2/8, \quad P(\omega_3) = 1/8, \quad P(\omega_4) = 1/8.$$

А вероятность выигрыша первого игрока равна

$$P(\{\omega_1, \omega_2, \omega_3\}) = 7/8.$$

■

Пример (Задача Банаха о спичках). Есть два коробка спичек по n штук. Какова вероятность, что первый раз вынув пустой коробок, во втором кармане будет k спичек?

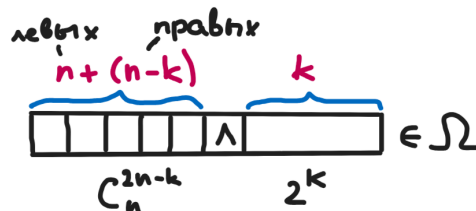
□ Итак, сначала нужно задать вероятностное пространство, соответствующее условию.

В качестве равновозможных исходов определим последовательности попыток вытащить спичку из правого или левого кармана. Максимальная длина такой последовательности равна $2n + 1$, поэтому

$$\Omega = \{(a_1, \dots, a_{2n+1}) : a_i \in \{\Pi, Л\}\}$$

Тогда $|\Omega| = 2^{2n+1}$, а вероятность любого исхода — $1/2^{2n+1}$. Теперь надо посчитать, сколько исходов будут удовлетворять заданному событию.

Предположим сначала, что пустым оказался коробок из левого кармана, а в правом при этом осталось k . Это означает, что к этому моменту были вынуты n левых спичек и $n - k$ правых.



То есть уже достали $n + (n - k) = 2n - k$ спичек, из которых n левых, затем попытались достать спичку из левого кармана, а потом осталось k раз достать оставшиеся спички из правого. Однако, эти k раз не равновозможны, им соответствует 2^k исходов из Ω .

Сколько же комбинаций (a_1, \dots, a_{2n-1}) , среди которых n штук Л и $n - k$ штук П? Расставить n штук Л по $2n - k$ местам это все равно что достать из $2n - k$ предметов n предметов с номерами этих мест. То есть C_{2n-k}^n . Подробнее см. ниже задачи о шарах.

Таким образом, количество исходов, соответствующих событию, что коробок в левом кармане оказался пуст, а в правом при этом осталось k спичек, равно

$$C_{2n-k}^n \cdot 2^k.$$

Для симметричной ситуации, когда правый коробок оказался пустым, а в левом осталось k спичек, ответ тот же. Поэтому всего исходов, удовлетворяющих заданному событию,

$$2 \cdot C_{2n-k}^n \cdot 2^k = 2^{k+1} C_{2n-k}^n.$$

Таким образом, ответ будет такой:

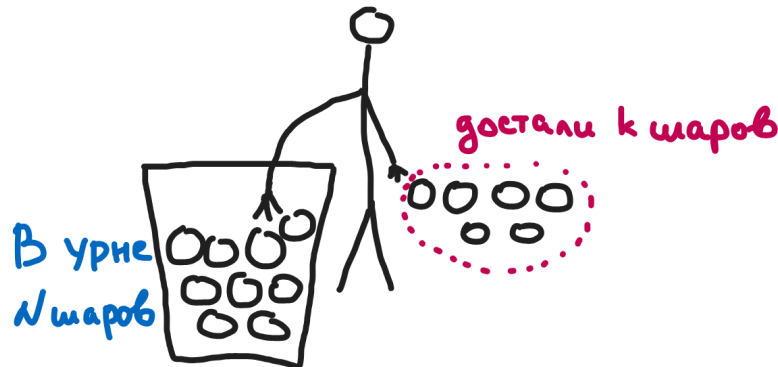
$$\frac{2^{k+1} C_{2n-k}^n}{2^{2n+1}} = \frac{C_{2n-k}^n}{2^{2n-k}}.$$

■

Четыре задачи о шарах

Знание основ комбинаторики необходимо как для решения самых простых задач на нахождение вероятностей, так и просто для понимания построения математической теории вероятностей. Но, конечно, комбинаторика не привязана именно к вероятности, она существует вполне самостоятельно и пригождается во многих задачах.

Итак, перейдем к модельной задаче, которая в себе содержит вопросы, на которые нам надо получить ответы. Представим, что имеется урна с шарами в размере N штук, и человеку требуется достать из урны k шаров. Вопрос в том, какова вероятность каждого варианта.



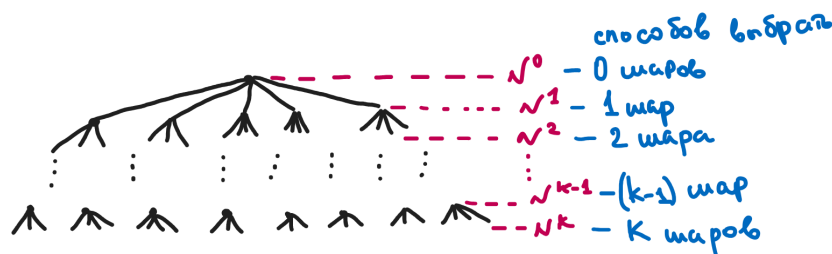
Но в формулировке задачи есть нюанс. Мы еще не уточнили, как именно проводится эксперимент, различаются ли как-либо шары в урне, и возвращает ли человек шар после того, как его достал. Отсюда возникает четыре различных случая, которые мы и рассмотрим далее.

Задача 1. Упорядоченный выбор с возвращением.

□ Строим вероятностное пространство.

$$\Omega_1 = \{(i_1, \dots, i_k) : 1 \leq i_s \leq N\}, \quad \mathcal{F}_1 = 2^{\Omega_1}.$$

Мощность $|\Omega_1| = N \cdot \dots \cdot N = N^k$.



Число благоприятных исходов (когда вытащили набор (i_1, \dots, i_k)) — один. Поэтому

$$P_1(((i_1, \dots, i_k))) = \frac{1}{N^k}.$$

Вероятностное пространство $(\Omega_1, \mathcal{F}_1, P_1)$ задано. ■

Задача 2. Упорядоченный выбор без возвращения.

□ Действуем аналогично.

$$\Omega_2 = \{(i_1, \dots, i_k) : 1 \leq i_s \leq N, i_j \neq i_l\}, \quad \mathcal{F}_2 = 2^{\Omega_2}.$$

Мощность $|\Omega_2| = N \cdot (N-1) \cdot (N-2) \cdot \dots \cdot (N-k+1) = \frac{N!}{(N-k)!} = A_n^k$.

Число благоприятных исходов (когда вытащили набор (i_1, \dots, i_k)) — опять один. Поэтому

$$P_2(((i_1, \dots, i_k))) = \frac{1}{A_n^k}.$$

Вероятностное пространство $(\Omega_2, \mathcal{F}_2, P_2)$ задано. ■

Задача 3. Неупорядоченный выбор без возвращения.

□ Действуем аналогично.

$$\Omega_3 = \{\{i_1, \dots, i_k\} : 1 \leq i_s \leq N, i_j \neq i_l\}, \quad \mathcal{F}_3 = 2^{\Omega_3}$$

Мощность $|\Omega_3| = \frac{A_n^k}{k!} = \frac{N!}{k!(N-k)!} = C_N^k$.

Число благоприятных исходов (когда вытащили набор (i_1, \dots, i_k)) — опять один. Поэтому

$$P_3(((i_1, \dots, i_k))) = \frac{1}{C_n^k}$$

Вероятностное пространство $(\Omega_3, \mathcal{F}_3, P_3)$ задано. ■

Следствие. Так как сумма вероятностей элементарных исходов должна равняться единице, то мы получаем формулу

$$\sum_{k=0}^N \frac{1}{C_n^k} = \sum_{k=0}^N \frac{1}{A_n^k} = 1.$$

Замечание. Неупорядоченный выбор без возвращения k шаров из n шаров это то же самое, что расставить k шаров по n ячейкам с номерами, которые заранее подписаны на вытасканных шарах. Поэтому количество способов расставить k одинаковых предметов по n местам — тоже C_n^k .

Задача 4. Неупорядоченный выбор с возвращением.

□ Здесь пойдем несколько другим путем. Будем представлять исходы в виде вектора чисел, характеризующих, сколько шаров какого типа (сколько первых шаров, сколько вторых и т.д.) в данном наборе шаров.

$$\Omega_4 = \{(i_1, \dots, i_N) : 0 \leq i_j \leq k, i_1 + \dots + i_N = k\}, \quad \mathcal{F} = 2^{\Omega_4}.$$

Кстати, линейные уравнения с целыми коэффициентами с поиском решений в целых числах называются диофантовыми. Таким является и наше получившееся уравнение

$$x_1 + \dots + x_N = k.$$

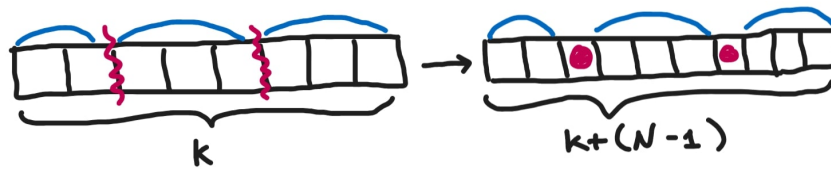
Здесь мы видим, что по сути задача сводится к тому, чтобы посчитать количество разбиений числа k в сумму N чисел от 0 до k .

Тогда можно представить, как в ряд выложены k шаров, а между ними расположены $N-1$ перегородка, каждая из которых отделяет i_j белых шаров. Но представим еще, что эти перегородки — тоже шары, допустим черные.

Таким образом, каждому нашему исходу можно однозначно сопоставить такой набор из белых и черных шаров, выложенных в ряд. То есть способов получить нужное разбиение ровно столько же, сколько есть способов расставить $N-1$ черный шар между k белыми.

Сколько же способов это сделать? Столько же, сколько способов достать $N-1$ шар из урны с $N+k-1$ одинаковыми шарами. Действительно, представим, что у нас выложено $N+k-1$ белых шаров в ряд, и все они занумерованы. Теперь сваливаем все шары в урну, перемешиваем, достаем $N-1$ шар и красим в черный. Затем достаем все остальные белые шары и выкладываем все шары в порядке подписанных заранее номеров. Получаем некоторое разбиение.

Итак, любое разбиение вида $k = i_1 + \dots + i_N$ означает, что из урны с $N+k-1$ шарами достали и перекрасили в черный шары с номерами $i_1+1, i_2+1, \dots, i_N+1$. То есть число разбиений равно числу способов достать $N-1$ шар из урны с $N+k-1$ шарами, а это число, по задаче 3, равно C_{N+k-1}^{N-1} .



Значит, мощность $|\Omega_4| = C_{N-k+1}^{N-1}$.

Поэтому

$$P((i_1, \dots, i_k)) = \frac{1}{C_{N-k+1}^{N-1}}.$$

Вероятностное пространство $(\Omega_4, \mathcal{F}_4, P_4)$ задано. ■

Упражнение*. Как еще можно построить вероятностное пространство в задаче 4?

Задача 1. Сколько подмножеств у n -элементного множества? (Указание: использовать первую задачу о шарах).

Задача 2. а) Сколько m -элементных подмножеств у n -элементного множества? б) Доказать, что

$$\sum_{k=0}^n C_n^k = 2^n.$$

Задача 3. а) Сколько встреч будет проведено в турнире 20 команд по круговой системе (в 1 круг)?

б) Сколькими способами можно собрать команду программистов из 5 человек, когда имеется 30 программистов?

Задача 4. Сколькими способами можно разделить десять задач в Jira по пяти программистам? (Указание: применить идею из четвертой задачи о шарах).

Задача 5. Доказать, что

$$C_n^k = C_n^{n-k}$$

используя и не используя прямой подсчет. Отсюда вытекает симметричность треугольника Паскаля.

Задача 6. Доказать рекуррентную формулу (треугольник Паскаля)

$$C_n^k = C_{n-1}^{k-1} + C_{n-1}^k$$

используя и не используя прямой подсчет.

Еще задачи

Задача 1.1 из [9]. Из ящика, содержащего три билета с номерами 1, 2, 3, вынимают по одному все билеты. Предполагается, что все последовательности номеров билетов имеют одинаковые вероятности. Найти вероятность того, что хотя бы у одного билета порядковый номер совпадает с собственным.

□ Определим вероятностное пространство:

$$\Omega = \{(a, b, c) : a, b, c \in \{1, 2, 3\} - \text{различны}\}$$

$$\mathcal{F} = 2^\Omega$$

$$P((a, b, c)) = \frac{1}{|\Omega|} = \frac{1}{6}$$

Выпишем все варианты последовательностей:

$$\begin{aligned}(1, 2, 3) \\ (1, 3, 2) \\ (2, 1, 3) \\ (2, 3, 1) \\ (3, 1, 2) \\ (3, 2, 1)\end{aligned}$$

Видим, что подходящих вариантов четыре, значит, ответ $4/6 = 2/3$. ■

Задача 1.2 из [9]. Колода из 36 карт хорошо перемешана (т.е. все возможные расположения карт равновероятны). Найти вероятности событий

$$\begin{aligned}A &= \{ \text{четыре туза расположены рядом} \} \\ B &= \{ \text{места расположения тузов образуют} \\ &\quad \text{арифметическую прогрессию с шагом 7} \}\end{aligned}$$

□ Определим вероятностное пространство:

$$\begin{aligned}\Omega &= \{(i_1, \dots, i_{36}) : i_j \in \{1, \dots, 36\} - \text{различны}\} \\ \mathcal{F} &= 2^\Omega \\ P((i_1, \dots, i_{36})) &= \frac{1}{|\Omega|} = \frac{1}{36!}\end{aligned}$$

Тогда

$$A = \{(i_1, \dots, i_{36}) : \exists j \leq 33 : i_j, i_{j+1}, i_{j+2}, i_{j+3} - \text{номера тузов}\}$$

Таким образом, имеем 33 позиции для четверки тузов и еще $4! = 24$ способов их между собой переставить и $32!$ способа переставить оставшиеся карты. Значит, $|A| = 33 \cdot 4! \cdot 32!$.

$$P(A) = \frac{|A|}{|\Omega|} = \frac{33 \cdot 4! \cdot 32!}{36!} = \frac{33 \cdot 24}{33 \cdot 34 \cdot 35 \cdot 36} = \frac{24}{34 \cdot 35 \cdot 36} \approx 0.00056.$$

Теперь зададим событие B :

$$B = \{(i_1, \dots, i_{36}) : \exists j \leq 15 : i_j, i_{j+7}, i_{j+14}, i_{j+21} - \text{номера тузов}\}$$

Значит, можно поставить первый туз на первые 15 мест, переставлять их между собой и переставлять оставшиеся карты, т.е.

$$|B| = 15 \cdot 4! \cdot 32!$$

Поэтому

$$P(B) = \frac{|B|}{|\Omega|} = \frac{15 \cdot 4! \cdot 32!}{36!} = \frac{15 \cdot 24}{33 \cdot 34 \cdot 35 \cdot 36} \approx 0.0002546. \quad \blacksquare$$

Задача 1 из листка 1 лекций Шапошникова.. N человек принесли подарки друг для друга. Затем этим подарки сложили в мешок и наугад каждый вынул из мешка себе подарок. Какова вероятность того, что конкретный человек вынул подарок, который он принес? Какова вероятность того, что никто не вытащил подарок, который сам принес?

□ Вероятностное пространство состоит из пространства исходов

$$\Omega = \{(a_1, \dots, a_N) : 1 \leq a_i \leq N, a_i \neq a_j\}$$

с классической вероятностью. Наборы упорядоченные, так как важно кто какой подарок получил.

Событие $A = \{\text{конкретный человек вынул свой подарок}\}$ имеет мощность

$$|A| = 1 \cdot (N-1) \cdot (N-2) \cdots 2 = (N-1)!,$$

таким образом,

$$P(A) = \frac{|A|}{|\Omega|} = \frac{(N-1)!}{N!} = \frac{1}{N}.$$

Рассмотрим теперь событие $B = \{\text{никто не вытащил свой подарок}\}$. Представим его в виде пересечения событий $B_i = \{i\text{-ый человек вытащил не свой подарок}\}$, тогда

$$B = \bigcap_{i=1}^N B_i.$$

Мощность пересечения считать сложно, а вот мощность объединения можно вычислить с помощью формулы включений-исключений. Поэтому попробуем найти мощность

$$\overline{B} = \overline{\bigcap_{i=1}^N B_i} = \bigcup_{i=1}^N \overline{B_i}.$$

Тогда по формуле включений-исключений имеем

$$|\overline{B}| = \sum_{i=1}^N |\overline{B_i}| - \sum_{i < j} |\overline{B_i} \cap \overline{B_j}| + \sum_{i < j < k} |\overline{B_i} \cap \overline{B_j} \cap \overline{B_k}| + \cdots + (-1)^{N+1} |\overline{B_1} \cap \cdots \cap \overline{B_N}|.$$

Итак, мощность $\overline{B_i}$ мы уже считали (для события A), т.е. $|\overline{B_i}| = (N-1)!$.

Пересечение событий $\overline{B_i}$ означает, что несколько человек взяли свои подарки. Аналогично рассуждая, получаем, что

$$\forall i < j \quad |\overline{B_i} \cap \overline{B_j}| = (N-2)!$$

и так далее для всех пересечений. Значит,

$$\begin{aligned} |\overline{B}| &= \sum_{i=1}^N (N-1)! - \sum_{i < j} (N-2)! + \sum_{i < j < k} (N-3)! + \cdots + (-1)^{N+1} = \\ &= N(N-1)! - C_N^2 (N-2)! + C_N^3 (N-3)! + \cdots + (-1)^{N+1} = \\ &= N! - \frac{N!}{2!} + \frac{N!}{3!} - \cdots + (-1)^{N+1} \end{aligned}$$

Значит,

$$\begin{aligned} P(\overline{B}) &= \frac{|\overline{B}|}{|\Omega|} = \frac{N! - \frac{N!}{2!} + \frac{N!}{3!} - \cdots + (-1)^{N+1}}{N!} = \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + \frac{(-1)^{N+1}}{N!}. \end{aligned}$$

Отсюда

$$P(B) = 1 - P(\overline{B}) = \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^N}{N!}.$$

Но это еще не все. Если знать, что

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^k}{k!} + \cdots \approx 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^N}{N!}, \quad N \rightarrow \infty,$$

то, можно заметить, что

$$P(B) \approx e^{-1} \approx 0.368.$$

■

Задача 1 (Задача 2 из листка 1 Шапошникова). Сто мудрецов по одному заходят в комнату, в которой стоят сто закрытых коробок. В каждой коробке лежит табличка с именем одного из мудрецов. Все имена различны. Мудрец открывает 50 коробок одну за другой в произвольном порядке. Если в одной из открытых им коробок есть его имя, то он выживает, а если нет, то погибает. После каждого мудреца все коробки закрывают и оставшиеся мудрецы не знают о судьбе ушедших в комнату. Изначально мудрецы находятся вместе и могут продумать план действий. Придумайте план, который гарантирует выживание всех мудрецов с вероятностью не менее $1/4$.

□

■

Задача 1.39. Найти вероятность того, что при случайной расстановке двух ладей на шахматной доске они не будут угрожать друг другу.

Задача 1.40*. Найти вероятность P_k того, что при случайной расстановке $2 \leq k \leq 8$ ладей на шахматной доске они не будут угрожать друг другу. При каких k вероятность меньше $1/2$? Менее $1/1000$?

Условная вероятность

Мы говорили о том, что вероятность субъективна, то есть зависит от того, какую информацию о физике эксперимента мы имеем. Поэтому важно уметь пересчитывать вероятность события после того, как были получены новые сведения.

Когда игрок в покер читает по лицу оппонента, пришла ли тому хорошая карта или плохая, он получает дополнительную информацию, которая влияет на принятие решения о дальнейших действиях.

Пересчитанную вероятность с учетом новых знаний называют условной, то есть вероятность события при условии, что уже известно что-то еще.

* * *

Пример. Кидается кубик. Какова вероятность, что выпало число меньше или равно трем, если известно, что выпало четное?

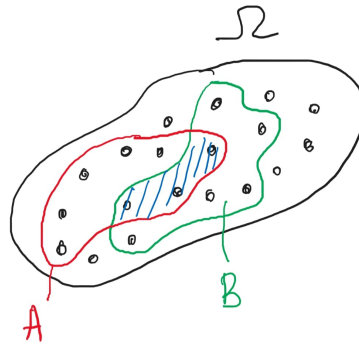
□ Отбрасываем все нечетные исходы, остаются только 2, 4, 6. Среди четных подходит только двойка. Значит вероятность равна $1/3$.

■

Вопрос. Перед тем, как идти дальше, стоит попробовать сформулировать общее определение условной вероятности в случае равновероятных исходов.

Опр. Пусть $P(B) > 0$. Тогда вероятностью события A при условии события B (conditional probability) называется число

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



Почему именно такая формула? Идея в следующем. Изначально было известно, что вероятность как-то распределена по исходам Ω , но когда стало известно B , то теперь про распределение стало известно точно, что попасть в \bar{B} вероятность равна нулю. То есть в некотором смысле пространство исходов сузилось, и теперь уже вероятность события A ищется в мире, где исходы извне B невозможны.

Поэтому если вначале вероятностью A мы называли, говоря интуитивным языком, долю знания о A относительно Ω , то теперь, зная, что B произошло, вероятность A это доля знания о нем относительно B . Об этом и говорит определение.

Утв. Отображение

$$P_B(\cdot) \equiv P(\cdot | B) : \begin{cases} 2^\Omega \rightarrow [0, 1] \\ A \mapsto P(A | B) \end{cases}$$

является вероятностной мерой на Ω . Причем $P_B(B) = P(B | B) = 1$.

Упр. Доказать предыдущее утверждение (проверить аксиомы).

Пример. В урне 15 белых и 13 черных шаров. Вынимается два шара (упорядоченно без возвращения). Какова вероятность, что оба — белые?

□

$$\begin{aligned} P(\{\text{оба белые}\}) &= P(\{\text{первый белый}\} \cap \{\text{второй белый}\}) = \\ &= P(\{\text{второй белый}\} | \{\text{первый белый}\})P(\{\text{первый белый}\}) = \\ &= \frac{14}{27} \cdot \frac{15}{28}. \end{aligned}$$

■

Пример. Какова вероятность, что второй шар — белый?

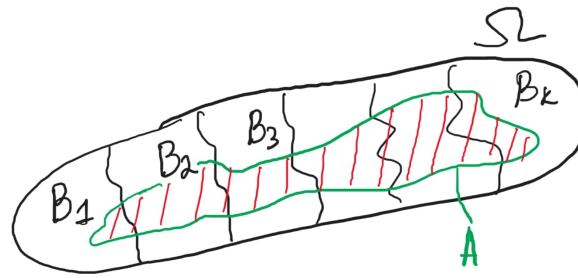
□

$$\begin{aligned} P(\text{2-й белый}) &= P(\text{2-й белый, 1-й белый}) + P(\text{2-й белый, 1-й черный}) = \\ &= P(\text{2-й белый} | \text{1-й белый})P(\text{1-й белый}) + \\ &+ P(\text{2-й белый} | \text{1-й черный})P(\text{1-й черный}) = \\ &= \frac{14}{27} \cdot \frac{15}{28} + \frac{15}{27} \cdot \frac{13}{28} = \frac{15}{28}. \end{aligned}$$

■

Теорема (Формула полной вероятности / Law of Total Probability). Пусть

$$\Omega = \bigsqcup_{k=1}^n B_k, \quad P(B_k) > 0.$$



Тогда

$$\forall A \in 2^\Omega \quad P(A) = \sum_{k=1}^n P(A \mid B_k)P(B_k).$$

□ Разобьем A на кусочки

$$A = \bigsqcup_{k=1}^n A \cap B_k.$$

Тогда, по аксиоме аддитивности,

$$P(A) = \sum_{k=1}^n P(A \cap B_k).$$

Но по определению условной вероятности

$$P(A \cap B_k) = P(A \mid B_k)P(B_k).$$

Пример. Монета лежит орлом вверх, а что с другой стороны — не знаем (либо тоже орел, либо решка). Бросили монету три раза. Найти вероятность того, что выпадет три раза орел. ■

□ Обозначим события

$$B_1 = \{\text{монета с двумя орлами}\}$$

$$B_2 = \{\text{монета с орлом и решкой}\}$$

$$A = \{\text{выпало три орла}\}$$

Тогда по формуле полной вероятности получаем

$$\begin{aligned} P(A) &= P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) = \\ &= 1 \cdot \frac{1}{2} + \frac{1}{8} \cdot \frac{1}{2} = \frac{9}{16}. \end{aligned}$$

Пример. Какая вероятность того, что монета с двумя орлами, зная, что выпало три орла? ■

□

$$\begin{aligned} P(B_1 \mid A) &= \frac{P(B_1 A)}{P(A)} = \frac{P(A \mid B_1)P(B_1)}{P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2)} = \\ &= \frac{1 \cdot \frac{1}{2}}{1 \cdot \frac{1}{2} + \frac{1}{8} \cdot \frac{1}{2}} = \frac{8}{9}. \end{aligned}$$

Упр. Что будет, если рассмотреть произвольное число бросков n и устремить его к бесконечности?

Теорема (Формула Байеса, Thomas Bayes, 1763).

$$\Omega = \bigsqcup_{i=1}^n B_i \Rightarrow P(B_j | A) = \frac{P(A | B_j)P(B_j)}{\sum_{i=1}^n P(A | B_i)P(B_i)}.$$

□ Очевидно (см. пример выше): комбинация определения и формулы полной вероятности. ■

Смысл формулы в том, чтобы связать условные вероятности $P(A | B)$ и $P(B | A)$. Это удобно, когда условие сложное, а событие простое. Дальше в примерах посмотрим, насколько это мощный инструмент.

Пример (Парадокс Байеса). Человек подозревает, что он болен. Имеется тест, который может ошибаться. Обозначим

$$\begin{aligned} B_1 &= \{\text{болен}\} \\ B_2 &= \{\text{здоров}\} \end{aligned}$$

Дано, что $P(B_1) = 0.01$. Также пусть $A = \{\text{тест сказал, что болен}\}$. Известно, что $P(A | B_1) = 0.99$, $P(A | B_2) = 0.02$. Найти $P(B_1 | A)$.

□ Применим формулу Байеса:

$$P(B_1 | A) = \frac{P(A | B_1)P(B_1)}{P(A | B_1)P(B_1) + P(A | B_2)P(B_2)} = \frac{0.99 \cdot 0.01}{0.99 \cdot 0.01 + 0.02 \cdot 0.99} = \frac{1}{3}.$$

Таким образом, если тест сказал, что человек болен, лишь в **трети** случаев это окажется правдой. Разберемся, в чем тут дело.

Посмотрим, за счет чего вероятность так мала. Вероятности $P(B_1)$ и $P(B_2)$ относятся к природе заболевания, а не к тесту. Вероятность $P(A | B_1)$ показывает точность теста, но она и так довольно высока: даже если бы она была равна единице, результат был бы не сильно лучше. А вот $P(A | B_2)$, играет ключевую роль: при уменьшении этой вероятности, результат приближается к единице.

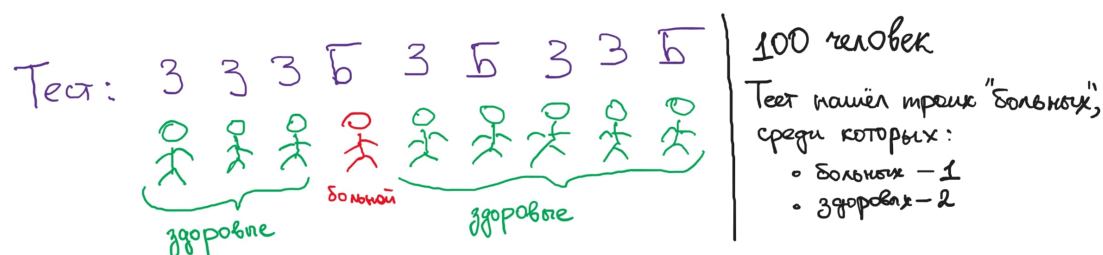
Отсюда можно сделать вывод, что точность теста в случае, если человек здоров, слишком мала для такой распространенности заболевания. Значит ли это, что тест плох? Что было бы лучше взять тест, который просто всегда говорит "здоров" и тем самым угадывает с вероятностью 0.99? Не совсем.

Пусть $C = \{\text{тест показал, что здоров}\}$. Посчитаем теперь вероятность

$$P(B_1 | C) = \frac{P(C | B_1)P(B_1)}{P(C | B_1)P(B_1) + P(C | B_2)P(B_2)} = \frac{0.01 \cdot 0.01}{0.01 \cdot 0.01 + 0.98 \cdot 0.99} \approx 10^{-4}.$$

Значит, в случае если тест показал, что человек здоров, то в это можно верить, а если показал, что болен, то нужно продолжить исследование — скорее всего человек здоров. Пример такого теста — маммография. ■

Пример (Парадокс Монти-Холла). Перед игроком три двери. За одной из них автомобиль, за двумя другими — козы. Он выбирает первую дверь, после чего ведущий открывает одну из других дверей, за которой коза. Ведущий предлагает игроку изменить свой выбор. Имеет ли смысл соглашаться?



□ Интуитивно кажется, что вероятность выигрыша — $1/2$, так как дверей осталось две, за одной точно автомобиль, а за другой — коза. Однако, это не так.

Есть несколько способов решать эту задачу. Мы рассмотрим два. Первый опирается больше на логику и рассуждения, второй использует формулу Байеса.

Решение 1 (логическое).

Итак, изначально вероятность выигрыша равняется $1/3$, так как автомобиль с равной возможностью находится за одной из трех дверей. Значит вероятность проигрыша, то есть того, что автомобиль находится за второй или третьей дверью, равна $2/3$.

Далее ведущий открывает третью дверь, за которой находится коза. Значит, вероятность того, что автомобиль за второй дверью теперь равна $2/3$, ведь за третьей дверью коза.

Поэтому игрок повысит свои шансы на выигрыш в два раза, если поменяет выбор.

Чтобы лучше прочувствовать, что произошло, можно рассмотреть несколько иную постановку задачи. Предположим, что дверей теперь 100. За одной дверью — авто. Игрок выбирает первую дверь. Тогда ведущий открывает 98 дверей, за которыми козы. Тот же вопрос, стоит ли игроку изменить выбор на ту дверь, которую ведущий оставил закрытой?

Становится понятно, что ведущий своим действием дает подсказку (или, скажем точнее, дополнительную информацию) относительно тех дверей, которые не выбраны игроком.

Решение 2 (формула Байеса).

Если приведенное выше рассуждение не кажется убедительным, то это решение должно все расставить на свои места.

Итак, чтобы применить формулу Байеса, нужно определить, какие события рассматриваются. Нам надо сравнить вероятности двух событий:

- автомобиль находится за **первой** дверью при условии, что ведущий открыл третью дверь *после того, как игрок выбрал первую дверь*
- автомобиль находится за **второй** дверью при условии, что ведущий открыл третью дверь *после того, как игрок выбрал первую дверь*

Найдем вероятность второго события. Из формулировки события сразу напрашиваются обозначения:

$A_i = \{\text{автомобиль находится за } i\text{-ой дверью}\}$

$B_i = \{\text{ведущий открыл } i\text{-ую дверь после того, как игрок выбрал первую}\}$

Тогда требуется найти $P(A_2 | B_3)$. Так как события A_i образуют полную систему событий, т.е. $\Omega = A_1 \sqcup A_2 \sqcup A_3$, можно применить формулу Байеса

$$\begin{aligned}
 P(A_2 | B_3) &= \frac{P(B_3 | A_2)P(A_2)}{P(B_3 | A_1)P(A_1) + P(B_3 | A_2)P(A_2) + P(B_3 | A_3)P(A_3)} = \\
 &= \frac{1 \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} = \frac{2}{3}.
 \end{aligned}$$

Следовательно,

$$P(A_1 | B_3) = P(B_3 \setminus A_2 | B_3) = P(B_3 | B_3) - P(A_2 | B_3) = 1 - \frac{2}{3} = \frac{1}{3}.$$

Из формулы Байеса видно, в чем различия между двумя выборами. А именно, видно, что изначальный выбор игрока влияет на выбор ведущего: если игрок выбрал дверь, за которой находится автомобиль, то ведущему все равно какую из дверей открыть, а если выбрана дверь, за которой коза, то ведущий может открыть только одну дверь.

Поэтому наша интуиция, которая говорит о том, что открытие двери ведущим никак не добавляет информации о том, за какой дверью автомобиль, не верна.

А что если мы знаем, что ведущий ленивый, стоит рядом с третьей дверью и всегда открывает ее, если за ней нет приза? Тогда

$$P(B_3 | A_1) = P(B_3 | A_2) = 1, \quad P(B_3 | A_3) = 0.$$

Значит,

$$P(A_2 | B_3) = \frac{1 \cdot \frac{1}{3}}{1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} = \frac{1}{2}.$$

А вот

$$\begin{aligned} P(A_3 | B_2) &= \frac{P(B_2 | A_3)P(A_3)}{P(B_2 | A_1)P(A_1) + P(B_2 | A_2)P(A_2) + P(B_2 | A_3)P(A_3)} = \\ &= \frac{1 \cdot \frac{1}{3}}{0 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}} = 1. \end{aligned}$$

Это понятно, ведь если ведущий пошел открывать вторую дверь, то это гарантированно означает, что за третьей — приз.

P.S. Решение просмотрено в посте в ЖЖ. ■

Пример (Парадокс мальчика и девочки). У мистера Брауна два ребенка, хотя бы один из них мальчик. Какова вероятность, что оба ребенка — мальчики? *[Вероятность рождения мальчика/девочки равна 1/2.]*

□ Здесь $\Omega = \{(c_1, c_2) : c_i \in \{b, g\}\}$ (c — child, b — boy, g — girl).

По формуле Байеса

$$\begin{aligned} P(c_1 = c_2 = b | c_1 = b \text{ или } c_2 = b) &= \\ &= \frac{P(c_1 = b \text{ или } c_2 = b | c_1 = c_2 = b)P(c_1 = c_2 = b)}{P(c_1 = b \text{ или } c_2 = b | c_1 = c_2 = g)P(c_1 = c_2 = g) + \dots} = \\ &= \frac{1 \cdot \frac{1}{4}}{0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4}} = \frac{1/4}{3/4} = \frac{1}{3}. \end{aligned}$$

Пример. У мистера Грина два ребенка. Один из них — мальчик, родившийся в среду. Какова вероятность того, что оба ребенка — мальчики? *[При решении этой задачи вероятность рождения детей в каждый день недели следует считать одинаковой, вероятности рождения мальчика и девочки считать одинаковыми, рождением близнецов пренебречь и т.п.]*

□ Будем рассматривать неделю из k дней. При $k = 1$ мы получаем задачу про мистера Брауна, при $k = 7$ — про мистера Грина, при $k = 365$ — задачу "По крайней мере один из детей мистера Блю — мальчик, рожденный 1 апреля".

Тогда имеем классическое вероятностное пространство с пространством исходов, состоящим из пар пар:

$$\Omega = \{((c_1, k_1), (c_2, k_2)) : c_i \in \{b, g\}, k_i \in \{1, \dots, k\}\}.$$

Заметим, что $|\Omega| = (2k)^2$.

Решение 1 (классическое).

Найдем вероятность того, что в семье есть девочка, зная, что есть и мальчик, родившийся в среду. Тогда, беря вероятность дополнения к этому событию, получим требуемую вероятность.

Итак, пусть мальчик, родившийся в среду — старший ребенок. Тогда для второго ребенка имеется $2k$ вариантов (возможны оба пола и k дней на каждого), из них k — девочки. Аналогично для случая, когда мальчик, родившийся в среду — младший ребенок.

Тогда вероятность, что в семье есть девочка равна

$$P(\text{есть девочка} \mid \text{есть мальчик, родившийся в среду}) = \frac{k + k}{2k + 2k - 1} = \frac{2k}{4k - 1},$$

здесь в знаменателе вычли единицу, так как в обоих рассмотренных случаях есть варианты, когда оба ребенка — мальчики, поэтому один из двух надо вычесть, чтобы не засчитать дважды.

Значит,

$$P(\text{оба мальчики} \mid \text{есть мальчик, родившийся в среду}) = 1 - \frac{2k}{4k - 1} = \frac{2k - 1}{4k - 1}.$$

При $k = 1$ это $1/3$, при $k = 7$ это $13/27$. Заметим, что

$$\lim_{k \rightarrow \infty} \frac{2k - 1}{4k - 1} = \frac{1}{2}.$$

Решение 2 (формула Байеса).

Обозначим для удобства:

$$A = \{\text{один из двух детей - мальчик}\}$$

$$B_i = \{\text{в семье } i \text{ мальчиков}\}, i = 0, 1, 2.$$

Нам надо найти $P(B_2 \mid A)$. Заметим, что

$$\Omega = B_0 \sqcup B_1 \sqcup B_2,$$

поэтому можно применять формулу Байеса:

$$\begin{aligned} P(B_2 \mid A) &= \frac{P(A \mid B_2)P(B_2)}{P(A \mid B_0)P(B_0) + P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2)} = \\ &= \frac{\left(1 - \frac{(k-1)^2}{k^2}\right) \cdot \frac{1}{4}}{0 \cdot \frac{1}{4} + 2 \cdot \frac{1}{k} \cdot \frac{1}{4} + \left(1 - \frac{(k-1)^2}{k^2}\right) \cdot \frac{1}{4}} = \frac{2k - 1}{4k - 1}. \end{aligned}$$

■

P.S. Два примера выше и решения взяты из этого ЖЖ поста. Эти же задачи привел Шапошников на лекциях в НМУ.

Замечание. TODO

Независимость событий

Опр. События A и B называются независимыми, если

$$P(A | B) = P(A) \quad \text{или} \quad P(AB) = P(A)P(B).$$

То есть знание о B не влияет на знание об A .

Заметим, что определение через условную вероятность по сути объясняет физический смысл удобного с алгебраической точки зрения определения $P(AB) = P(A)P(B)$, которое само по себе было бы трудно интерпретируемо.

Контрольный вопрос. Верно ли, что если $P(A | B) = P(A)$, то и $P(B | A) = P(B)$?

Н.В. Независимость — неинтуитивная вещь!

Пример. Пусть $\Omega = \{1, \dots, 100\}$. Берем одно число a . Рассмотрим два события

$$A = \{2|a\}, \quad B = \{5|a\}.$$

Тогда $A \cap B = \{10|a\}$, значит, $P(A \cap B) = \frac{1}{10}$. При этом

$$P(A) = \frac{1}{2}, \quad P(B) = \frac{1}{5}.$$

Таким образом, A и B — независимы.

Но теперь возьмем $\Omega = \{1, \dots, 101\}$ и те же события. Тогда

$$P(A \cap B) = \frac{10}{101}, \quad P(A) = \frac{50}{101}, \quad P(B) = \frac{20}{101}.$$

Отсюда видим, что теперь A и B зависимы.

Н.В. Часто независимость путают с несовместностью. Наоборот, если события несовместны ($A \cap B = \emptyset$), то события очень даже зависимы ($P(A | B) = 0$).

Пример. Имеется колода из 52 карт. Обозначим события

$$A = \{\text{взять даму}\}, \quad B = \{\text{взять пику}\}.$$

Являются ли эти события независимыми?

□

$$\Omega = \{1, \dots, 52\}$$

$$A = \{1, 14, 27, 40\} \text{ — номера дам}$$

$$B = \{1, \dots, 13\} \text{ — номера пиковых карт}$$

Значит,

$$P(A) = \frac{4}{52} = \frac{1}{13}, \quad P(B) = \frac{13}{52} = \frac{1}{4}, \quad P(AB) = \frac{1}{52}.$$

Таким образом,

$$P(AB) = P(A)P(B),$$

значит, A и B независимы. ■

Пример. Пусть теперь в колоду добавили два джокера (джокер может считаться любой картой). Тот же вопрос: зависимы ли A и B ?

□ Теперь

$$P(A) = \frac{6}{54}, \quad P(B) = \frac{15}{54}, \quad P(AB) = \frac{3}{54}.$$

Получаем, что

$$P(A)P(B) = \frac{90}{54^2} \neq \frac{3}{54} = P(AB),$$

т.е. события зависимы. ■

Упр. События A и B независимы. Являются ли независимыми а) A и \bar{B} , б) \bar{A} и \bar{B} ?

Схема Бернулли

Одним из основных применений теории вероятности является эксперимент с повторением независимых испытаний.

Предположим, что подкидывается несимметричная монетка. С вероятностью p выпадает решка, с вероятностью $q = 1 - p$ — орел. Рассмотрим два подряд подбрасывания монетки. Для каждого испытания можно задать идентичные вероятностные пространства:

$$\begin{aligned}\Omega_1 &= \{O, P\} & P_1(O) &= p & P_1(P) &= q \\ \Omega_2 &= \{O, P\} & P_2(O) &= p & P_2(P) &= q\end{aligned}$$

Для рассмотрения последовательности из двух испытаний, определим вероятностное пространство из пар:

$$\Omega = \Omega_1 \times \Omega_2 = \{(i, j) : i, j \in \{O, P\}\}$$

Так как испытания должны быть независимыми, необходимо, чтобы

$$P((i, j)) = P_1(i)P_2(j),$$

поэтому

$$\begin{aligned}P((O, O)) &= q^2 \\ P((O, P)) &= qp \\ P((P, O)) &= pq \\ P((P, P)) &= p^2\end{aligned}$$

* * *

Теперь рассмотрим более общий случай, так называемую схему Бернулли. Это схема эксперимента с повторяющимися независимыми n испытаниями, у каждого из которых возможно два исхода (success/fail): с вероятностями p и $q = 1 - p$. Итак, определим пространство исходов:

$$\Omega = \{(i_1, \dots, i_n) : i_j \in \{S, F\}\} \quad (S - \text{success}, F - \text{fail}).$$

Распределение, как и в случае с двумя испытаниями, естественно задать следующим образом:

$$P((i_1, \dots, i_n)) = p^{\text{число успехов}} (1 - p)^{\text{число неудач}}$$

Тогда

$$P(\text{ровно } k \text{ успехов}) = C_n^k p^k (1 - p)^{n-k},$$

так как любой исход, в котором ровно k успехов имеет вероятность $p^k (1 - p)^{n-k}$, а таких исходов столько же, сколько способов расставить k элементов по n ячейкам, то есть C_n^k .

Упр. Доказать, что заданное распределение действительно является распределением. То есть надо проверить, что $P(\Omega) = 1$.

Упр. Доказать, что события $A = \{i_k = F\}$ и $B = \{i_m = S\}$, $k \neq m$, независимы.

Замечание. Схема Бернулли легко обобщается на случай, когда испытания имеют не два, а некоторое конечное число исходов.

Задачи

Задача 1 (Задача 5 из листка 1 Шапошникова). Колоду из 52 карт раздают на четверых игроков. Один из игроков объявляет, что у него есть туз. Какова вероятность, что у него есть еще хотя бы один туз? Какова вероятность, что у него есть еще хотя бы один туз, если он объявил, что у него туз пик?

□

■

Задача 2 (Задача 14 из листка 1 Шапошникова). На посадку в самолет стоят $N \geq 2$ пассажиров, среди которых сумасшедшая старушка. Старушка расталкивает всех пассажиров и садится в самолет на произвольное место. Затем пассажиры, когда заходят в самолет, садятся на свое место, если оно свободно, и на произвольное свободное место в противном случае. Какова вероятность того, что последний пассажир сядет на свободное место?

□ TODO: добавить решение.

Решение см. в листке Шкляева [5].

■

Задача 3 (Задача 17 из листка 1 Шапошникова). Алиса и Боб бросают монету n раз. Какова вероятность, что у Алисы выпало больше орлов, чем у Боба? Каков будет ответ, если Алиса $n + 1$ раз бросила монету?

□ Определим пространство исходов:

$$\Omega = \{((a_1, \dots, a_n), (b_1, \dots, b_n)) : a_i, b_i \in \{0, 1\}\}.$$

Единица означает орел, а ноль — решка. Т.к. монетка правильная, Распределение классическое:

$$P((a_1, \dots, a_n), (b_1, \dots, b_n)) = \frac{1}{2^{2n}}.$$

Нам требуется вычислить вероятность события

$$\left\{ \sum_{i=1}^n a_i > \sum_{i=1}^n b_i \right\}.$$

Итак, по формуле полной вероятности,

$$P\left(\sum_{i=1}^n a_i > \sum_{i=1}^n b_i\right) = \sum_{k=0}^n P\left(\sum_{i=1}^n a_i > k \mid \sum_{i=1}^n b_i = k\right) P\left(\sum_{i=1}^n b_i = k\right)$$

Здесь, на основе рассуждений о схеме Бернулли, нам понятно, что

$$P\left(\sum_{i=1}^n b_i = k\right) = C_n^k \frac{1}{2^k} \cdot \frac{1}{2^{n-k}} = \frac{C_n^k}{2^n}.$$

Теперь, так как Алиса и Боб бросают монетку независимо друг от друга (это можно проверить формально, если неочевидно), имеем

$$P\left(\sum_{i=1}^n a_i > k \mid \sum_{i=1}^n b_i = k\right) = P\left(\sum_{i=1}^n a_i > k\right).$$

Вычислим эту вероятность. Воспользуемся аддитивностью вероятности (аксиомой):

$$P\left(\sum_{i=1}^n a_i > k\right) = \sum_{m=k+1}^n P\left(\sum_{i=1}^n a_i = m\right) = \sum_{m=k+1}^n \frac{C_n^m}{2^n}.$$

Собираем все, получаем, что

$$P\left(\sum_{i=1}^n a_i > \sum_{i=1}^n b_i\right) = \sum_{k=0}^n \sum_{m=k+1}^n \frac{C_n^m C_n^k}{2^n 2^n} = \frac{1}{4^n} \sum_{k=0}^n \sum_{m=k+1}^n C_n^m C_n^k.$$

Эту двойную сумму можно вычислить. Сначала заметим, что

$$\sum_{k=0}^n \sum_{m=k+1}^n C_n^m C_n^k = \sum_{0 \leq k < m \leq n} C_n^m C_n^k.$$

Но если суммировать по всем парам k, m , то такую сумму можно представить в виде

$$4^n = \sum_{0 \leq k, m \leq n} C_n^m C_n^k = 2 \sum_{0 \leq k < m \leq n} C_n^m C_n^k + \sum_{k=0}^n (C_n^k)^2.$$

Отсюда,

$$\sum_{k=0}^n \sum_{m=k+1}^n C_n^m C_n^k = \frac{4^n - \sum_{k=0}^n (C_n^k)^2}{2}.$$

Но это еще не все, хочется вычислить сумму этих квадратов. Для этого используем равенство

$$(1+x)^n (1+x)^n = (1+x)^{2n}.$$

По биному Ньютона имеем

$$(1+x)^n = \sum_{k=0}^n C_n^k x^k.$$

Тогда

$$\begin{aligned} (1+x)^n (1+x)^n &= \sum_{k=0}^n C_n^k x^k \sum_{m=0}^n C_n^m x^m = \sum_{k=0}^n \sum_{m=0}^n C_n^k C_n^m x^{k+m} = \\ &= \sum_{0 \leq k, m \leq n} C_n^k C_n^m x^{k+m} = \sum_{s=0}^{2n} \left(\sum_{\substack{0 \leq k, m \leq n \\ k+m=s}} C_n^k C_n^m \right) x^s \end{aligned}$$

Посмотрим, какой коэффициент стоит при x^n , т.е. когда $s = n$:

$$\sum_{\substack{0 \leq k, m \leq n \\ k+m=n}} C_n^k C_n^m = \sum_{\substack{0 \leq k \leq n \\ m=n-k}} C_n^k C_n^m = \sum_{0 \leq k \leq n} C_n^k C_n^{n-k} = \sum_{k=0}^n (C_n^k)^2.$$

С другой стороны,

$$(1+x)^{2n} = \sum_{k=0}^{2n} C_{2n}^k x^k,$$

и тут коэффициент при x^n равен C_{2n}^n .

Таким образом,

$$\sum_{k=0}^n (C_n^k)^2 = C_{2n}^n.$$

И тогда ответ получается следующий:

$$P\left(\sum_{i=1}^n a_i > \sum_{i=1}^n b_i\right) = \frac{1}{2} \left(1 - \frac{C_{2n}^n}{4^n}\right).$$

■

Случайные величины

Мотивация

Теория вероятностей началась с изучения азартных игр, где результат описывался количественно. Например, выпало k шестерок за n бросаний кубика. Затем получило развитие страхование жизни и проч. Поэтому практически любое событие, вероятность которого интересно было изучать, описывалось в количественном, числовом выражении. О таких количествах мы говорим, что это случайная величина. Слово "величина" несет в себе именно числовой оттенок, то есть ее можно измерить.

Мы же построили теорию, в которой события могут быть любыми, главное, чтобы распределение было задано в соответствии с аксиомами. Это построение позволило заложить базу для того, чтобы количественно оценивать вероятность, то есть, имеющуюся информацию о событии.

Отсюда логично сделать следующий шаг и попробовать придать событиям количественный смысл. Во-первых, нам в прикладном смысле практически всегда нужны именно события, описываемые численно, во-вторых, арифметическая структура пространства исходов позволит применить к теории вероятностей не только теорию множеств и теорию меры, но и анализ функций (дифференцирование, интегрирование).

Начало

Итак, пусть у нас есть вероятностное пространство (Ω, \mathcal{F}, P) . Пространство исходов

$$\Omega = \{\omega_1, \dots, \omega_n\}.$$

Исходы — абстрактны, они могут быть чем угодно, иметь любую природу: звезды, люди, песчинки, точки пространства, нейроны мозга, мысли и т.д. Но если, как в данном случае, их n штук, то их легко можно занумеровать, что, впрочем, уже и сделано в записи выше.

И тогда ничего не мешает "закодировать", дать числовые имена каждому исходу: исход ω_1 назовем числом 1 и т.д. Получится множество исходов

$$\Omega' = \{1, \dots, n\}.$$

При этом распределение на этом пространстве зададим так

$$P'(\{i\}) = P(\{\omega_i\}).$$

Что произошло? Вместо элементов какого-то множества неизвестной природы, теперь наши исходы — элементы арифметического пространства. Мы их можем складывать, вычитать, умножать, делить. Пока непонятно чем это хорошо, но ясно, что этот момент меняет суть дела. Это то, что было приобретено.

При этом за бенефиты было заплачено тем, что была потеряна природа эксперимента. Так, вместо условных орла и решки, мы будем рассматривать часто исходы $\{0, 1\}$. Но вероятностная структура была сохранена, и потому, с позиции исследования вероятностей, было все сохранено.

Так в чем смысл вообще этой деятельности? Что дает эта кодировка в практическом смысле?

Пример. Монетку подбросили n раз. Какова вероятность того, что орел выпал k раз?

□ Понятно, что ответ $\frac{C_n^k}{2^n}$. Но обозначим теперь исходы 0 – решка, 1 – орел, а результат i -го подбрасывания – a_i . Тогда нужно найти

$$P\left(\sum_{i=1}^n a_i = k\right),$$

что в сравнении с

$$P(|\{i \in \{1, \dots, n\} : a_i = \{\text{орел}\}\}| = k)$$

очевидно является более красивой и удобной записью. ■

При этом способ кодировки играет ключевую роль, как и выбор вероятностного пространства. Можно задать множество различных пространств, какие-то из них будут одинаково хорошо описывать эксперимент, а какие-то будут просто невалидными.

Так и здесь. Если поменять местами и задать 0 – решка, 1 – орел, хуже кодировка не станет, просто поменяется формулировка интересующего события. Но если задать решка – 0, орел – 100, то такой подход будет неудобным в предыдущей задаче, но станет удачным, если за орла дают сто рублей, и вопрос ставится о выигрыше.

Определение

Опр. Случайной величиной X (random variable) называют функцию на элементарных исходах

$$X : \Omega \rightarrow \mathbb{R}.$$

Пример. Два броска симметричной монеты.

$$\Omega = \{(i_1, i_2) : i_1, i_2 \in \{0, 1\}\}, \quad P((i_1, i_2)) = \frac{1}{4}.$$

Тогда

$$\begin{aligned} X_1((i_1, i_2)) &= i_1 && \text{результат первого броска} \\ X_2((i_1, i_2)) &= i_2 && \text{результат второго броска} \\ X_3((i_1, i_2)) &= i_1 + i_2 && \text{число орлов} \end{aligned}$$

Задание случайной величины это назначение каждому исходу некоторого числа. Причем разным исходам можно задать одно и то же число. Тогда эти исходы как бы склеются. Например, $X_3((0, 1)) = X_3((1, 0)) = 1$.

Случайная величина это датчик, который показывает "температуру" в разных точках пространства исходов. Такая склейка исходов суммирует их вероятности. Так,

$$P(\omega : X_3(\omega) = 1) = P((0, 1)) + P((1, 0)) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Если продолжать аналогию с температурой, то точки пространства с одинаковой температурой стали неразличимы, а их суммарная площадь стала больше. И тогда, имея такой датчик, мы можем задавать вопрос: с какой вероятностью муха окажется в точке с температурой 10 градусов?

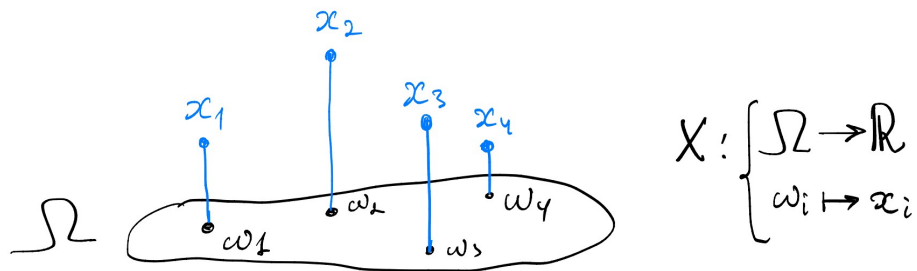


Рис. 3: Задание случайной величины, как функцию от исходов

Заметим, что случайная величина задает новое вероятностное пространство.

Обозначим буквой \mathcal{X} множество всех значений случайной величины X . То есть

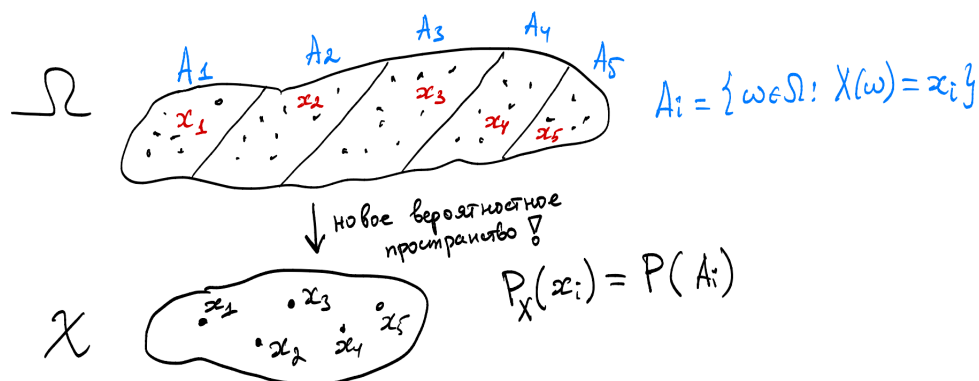
$$\mathcal{X} = \{x_1, \dots, x_n\}.$$

Тогда, измеряя температуру, мы проводим эксперимент с исходами \mathcal{X} , а распределение вероятности P_X будет следующим

$$P_X(x_i) = P(\omega \in \Omega : X(\omega) = x_i)$$

Замечание. Часто запись $P(\omega \in \Omega : X(\omega) = x_i)$ сокращают до $P(X = x_i)$. Тогда

$$P_X(x_i) = P(X = x_i).$$



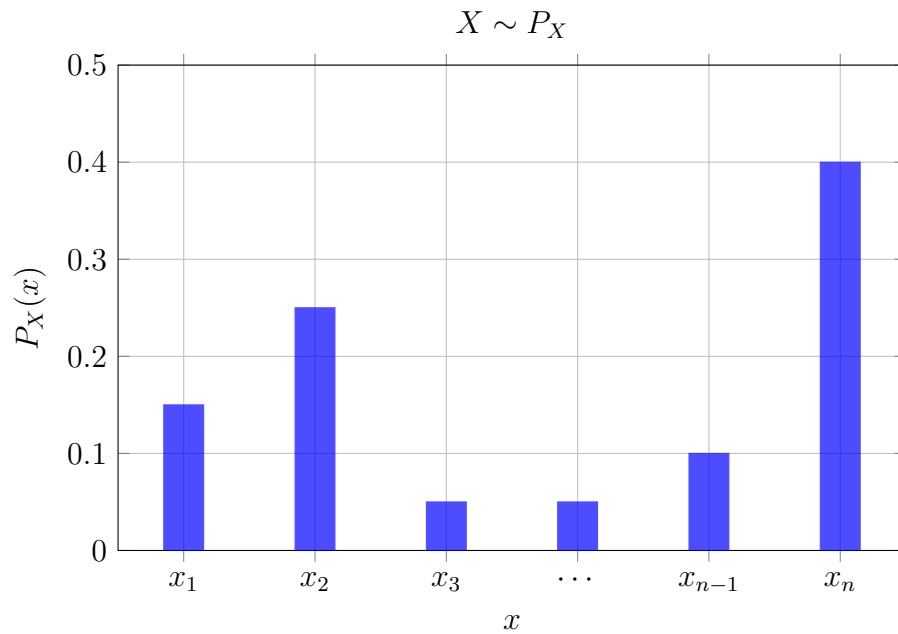
Опр. Пространство $(\mathcal{X}, 2^{\mathcal{X}}, P_X)$ называется вероятностным пространством, заданным случайной величиной X .

Опр. Распределение P_X называется распределением случайной величины X .

Обозначение. Если случайная величина X имеет распределение P , то пишут $X \sim P$.

Распределение случайной величины часто задают в виде таблицы или графика:

$x \in \mathcal{X}$	x_1	x_2	\dots	x_{n-1}	x_n
$P_X(x)$	p_1	p_2	\dots	p_{n-1}	p_n



* * *

Н.В. На практике нас интересует распределение именно случайной величины, а распределение вероятности на исходном вероятностном пространстве – нет. Поэтому при исследовании случайных величин даже не обязательно знать Ω , достаточно знать только, что оно есть.

Упр. Обратимся к примеру с бросанием двух монет и случайным величинам X_1, X_2, X_3 . Описать для каждой из них вероятностное пространство $(\mathcal{X}_i, 2^{\mathcal{X}_i}, P_{X_i})$, которое они задают.

* * *

Остановимся на понятии распределения случайной величины.

Распределение – главная и определяющая характеристика случайной величины, которая говорит о том, с какой вероятностью мы будем наблюдать каждое возможное значение. Попробуем почувствовать это чуть глубже.

Снова представим себя физиком-экспериментатором. Нам надо исследовать некую среду, хоть бы например и помещение, где мы находимся. Тогда мы возьмем любую точку пространства в этом помещении и будем разглядывать ее, пытаясь исследовать ее свойства, а значит и свойства всех остальных точек помещения.

Загвоздка в том, что у точки огромное количество свойств: температура, влажность, скорость ветра, давление воздуха, расстояние до Амстердама, количество проходивших через нее за все время людей и т.д.

Назовем совокупность всех этих свойств состоянием. Это состояние есть у каждой точки, и оно случайно. Напомню, что под случайностью надо понимать субъективную нехватку информации для достоверного знания о чем-либо. А про состояние точки мы что-то знаем, может быть ее положение, а многое не знаем.

Так вот, состояние точки оказывается случайным исходом в нашем уже привычном понимании вероятностного пространства, то бишь вероятностной модели эксперимента:

$$\Omega = \{\text{все возможные состояния}\}$$

$$\mathcal{F} = 2^\Omega$$

P – какое-то.

Распределение P указывает на то, какие состояния более, а какие менее вероятны.

Но изучать состояния вообще – неконкретная задача. Непонятно, что именно изучать. Поэтому разумно наблюдать за специфическими свойствами состояния, которые уже можно подвергать анализу.

Естественно выделяются те свойства, которые можно описать количественно, то есть численно. А значит, нужен датчик, реальный или воображаемый (для теоретических исследований), который бы умел вычислять конкретную величину, характеризующую некоторое свойство общего состояния, например температуру.

В этой аналогии с датчиком легко провести все параллели. Итак, само состояние измеряемого объекта это исход $\omega \in \Omega$, значение датчика это X , механизм работы датчика это отображение

$$X : \Omega \rightarrow \mathbb{R}.$$

Причем, механизм всегда работает одинаково, но за счет того, что сам исход случаен, то и значение датчика случайно.

Но случайность бывает разная. Рост первого попавшегося прохожего – случайная величина, но это не значит, что про эту случайность мы ничего не знаем. Например, мы знаем, что скорее всего рост этого человека будет около 160-180 сантиметров.

Поэтому распределение вероятности на обычном вероятностном пространстве или на пространстве, порожденном случайной величиной, это именно описание самой случайности, того, какова эта случайность.

* * *

Задача (Задача 3.1 из [9]). Распределение случайной величины X определяется формулами

$$P(X = i) = \frac{1}{5}, \quad i = -2, -1, 0, 1, 2.$$

Найти распределение величин $Y_1 = -X$, $Y_2 = |X|$.

□ Найдем распределение Y_1 , записав его в виде таблички:

k	-2	-1	0	1	2
$P(Y_1 = k)$	1/5	1/5	1/5	1/5	1/5

Из интересного заметим, что Y имеет то же самое распределение, что и X , но это другая случайная величина, они не равны как функции!

Аналогично сделаем с величиной Y_2 :

k	0	1	2
$P(Y_2 = k)$	1/5	2/5	2/5

■

Задача (Задача 3.2 из [9]). Распределение случайной величины X определяется формулами

$$P(X = k) = \frac{C}{k(k+1)}, \quad k = 1, 2, \dots$$

Найти: а) постоянную C ; б) $P(X \leq 3)$; в) $P(n_1 \leq X \leq n_2)$.

□ Чтобы найти C , нужно воспользоваться определением P :

$$\sum_{k=1}^{\infty} P(X = k) = 1.$$

То есть должно выполняться равенство

$$\sum_{k=1}^{\infty} \frac{C}{k(k+1)} = 1.$$

Для вычисления этого ряда воспользуемся тем, что

$$\frac{1}{k(k+1)} = \frac{1}{k} - \frac{1}{k+1}.$$

Тогда

$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)} = \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \dots = 1.$$

А значит

$$\sum_{k=1}^{\infty} \frac{C}{k(k+1)} = C = 1,$$

то есть $C = 1$.

Теперь вычислим $P(X \leq 3)$:

$$\begin{aligned} P(X \leq 3) &= P(X = 1) + P(X = 2) + P(X = 3) = \\ &= \frac{1}{2} + \frac{1}{6} + \frac{1}{12} = \frac{3}{4}. \end{aligned}$$

И осталось посчитать $P(n_1 \leq X \leq n_2)$:

$$\begin{aligned} P(n_1 \leq X \leq n_2) &= \sum_{k=n_1}^{n_2} P(X = k) = \sum_{k=n_1}^{n_2} \left(\frac{1}{k} - \frac{1}{k+1} \right) = \\ &= \left(\frac{1}{n_1} - \frac{1}{n_1+1} \right) + \left(\frac{1}{n_1+1} - \frac{1}{n_1+2} \right) + \dots + \left(\frac{1}{n_2} - \frac{1}{n_2+1} \right) = \\ &= \frac{1}{n_1} - \frac{1}{n_2+1} = \frac{n_2 - n_1 + 1}{n_1(n_2+1)}. \end{aligned}$$

■

Пример. X – случайная величина, задающая число бросков монеты до последнего выпадения орла среди 10 бросков монеты с вероятностью орла p (если орлов не было, то величина равна 0). Построить \mathcal{X} , найти распределение X , найти $P(X < 3)$.

□ X может принимать значения

$$\mathcal{X} = \{0, 1, \dots, 10\}.$$

Случайная величина задана на вероятностном пространстве. Определим для удобства в дальнейшем множество исходов:

$$\Omega = \{(a_1, \dots, a_{10}) : a_i \in \{O, P\}\}.$$

Теперь найдем распределение X . Для этого нужно вычислить вероятности $P(X = i)$, где $i \in \mathcal{X}$. Итак,

$$P(X = 0) = (1 - p)^{10},$$

так как это означает, что выпали только решки.

Далее, событие $X = 1$ означает, что первым выпал орел и дальше были только решки, т.е.

$$P(X = 1) = p(1 - p)^9.$$

Событие $X = 2$ означает, что вторым выпал орел, дальше только решки. А на первом месте могло быть что угодно. Распишем это по формуле полной вероятности

$$\begin{aligned} P(X = 2) &= P(X = 2 \mid a_1 = O)P(a_1 = O) + P(X = 2 \mid a_1 = P)P(a_1 = P) = \\ &= p(1 - p)^8 \cdot p + p(1 - p)^8 \cdot (1 - p) = p(1 - p)^8. \end{aligned}$$

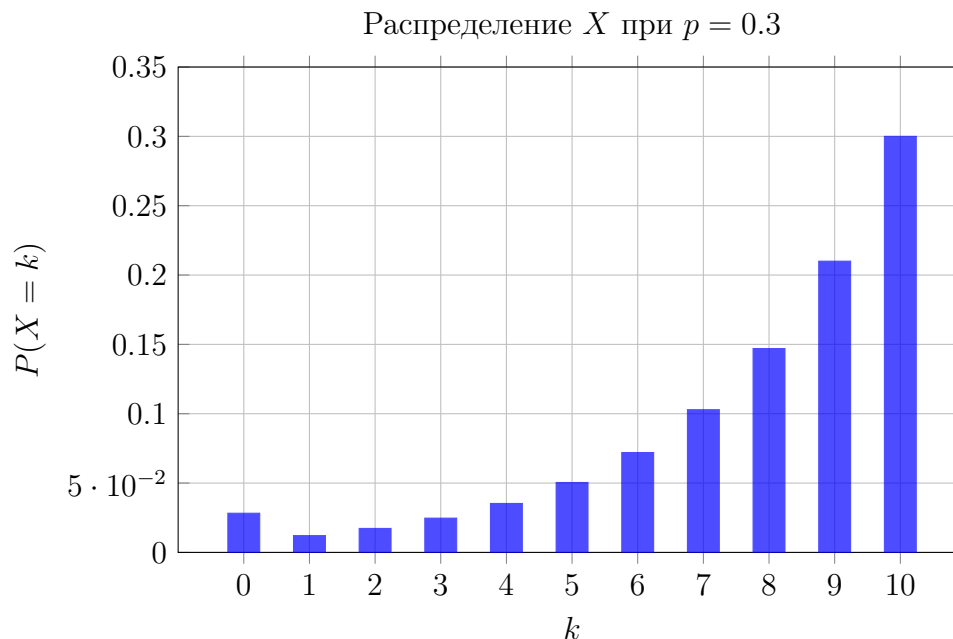
Аналогично, для $X = 3$ можно разбить на случаи по всем возможным исходам первых двух бросков. Получим

$$P(X = 3) = p(1 - p)^7.$$

И так далее

$$P(X = i) = p(1 - p)^{10-i}, \quad i = 1, \dots, 10.$$

Таким образом, распределение X найдено.



Вероятность события $X < 3$, т.е. того, что последний орел выпал не позже второго броска, считается просто:

$$\begin{aligned} P(X < 3) &= P(X = 0) + P(X = 1) + P(X = 2) = \\ &= (1 - p)^{10} + p(1 - p)^9 + p(1 - p)^8 = \\ &= (1 - p)^8((1 - p)^2 + p(1 - p) + p) = \\ &= (1 - p)^8. \end{aligned}$$

■

Базовые распределения

Распределение случайной величины описывает математическую модель эксперимента, как и в случае с вероятностным пространством. Но, как уже было сказано, случайная величина указывает не на любые исходы, а на численные показатели исхода.

В этом разделе мы рассмотрим базовые распределения, которые чаще всего встречаются в моделировании.

Равномерное распределение, $X \sim U\{1, \dots, N\}$

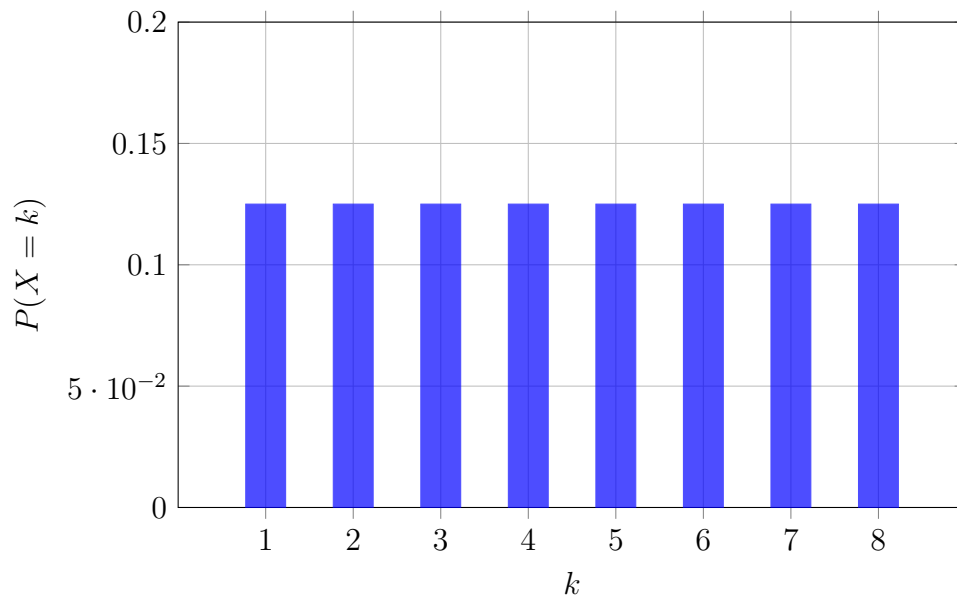
Равномерное распределение соответствует эксперименту, где все значения случайной величины равновероятны. Тогда все возможные значения это

$$\mathcal{X} = \{1, \dots, N\},$$

с распределением

$$P_X(k) = \frac{1}{N}, \quad 1 \leq k \leq N.$$

$$X \sim U\{1, \dots, N\}, \quad N = 8$$



Например, такая величина описывает количество очков, которое выпадет на шестигранном кубике, или номер шара, который вынут из урны.

Распределение Бернулли, $X \sim \text{Bern}(p)$

Это распределение моделирует подбрасывание несимметричной монетки, или, более абстрактно, любого эксперимента, где есть только два исхода: успех и неудача, причем вероятность успеха равна p .

Таким образом, множество значений случайной величины есть

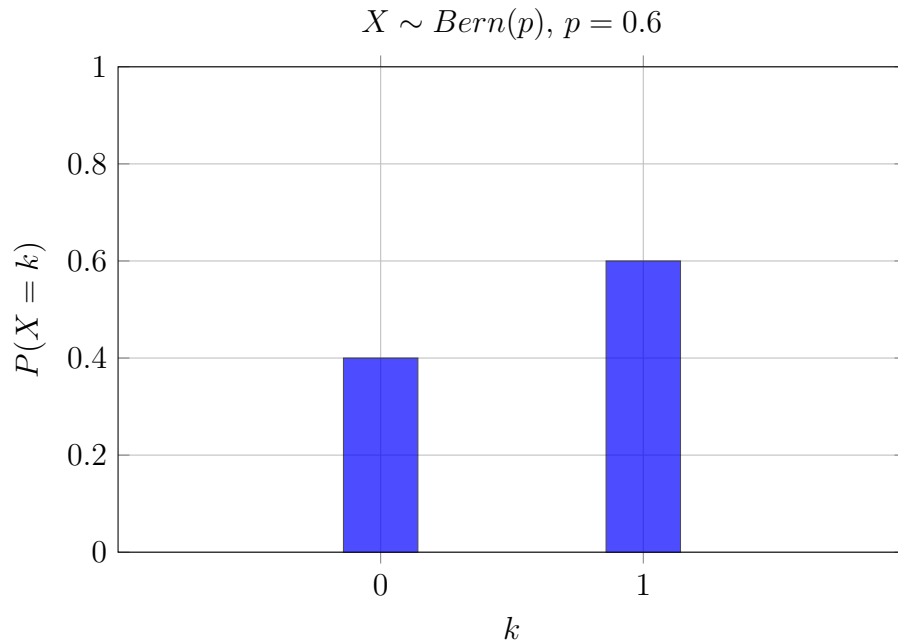
$$\mathcal{X} = \{0, 1\},$$

а распределение таково

$$P_X(1) = p$$

$$P_X(0) = 1 - p.$$

Распределение названо в честь Якоба Бернулли, одного из основоположников теории вероятностей. Он изучал в основном именно такие эксперименты с двумя исходами и доказал для этого случая вариант закона больших чисел.



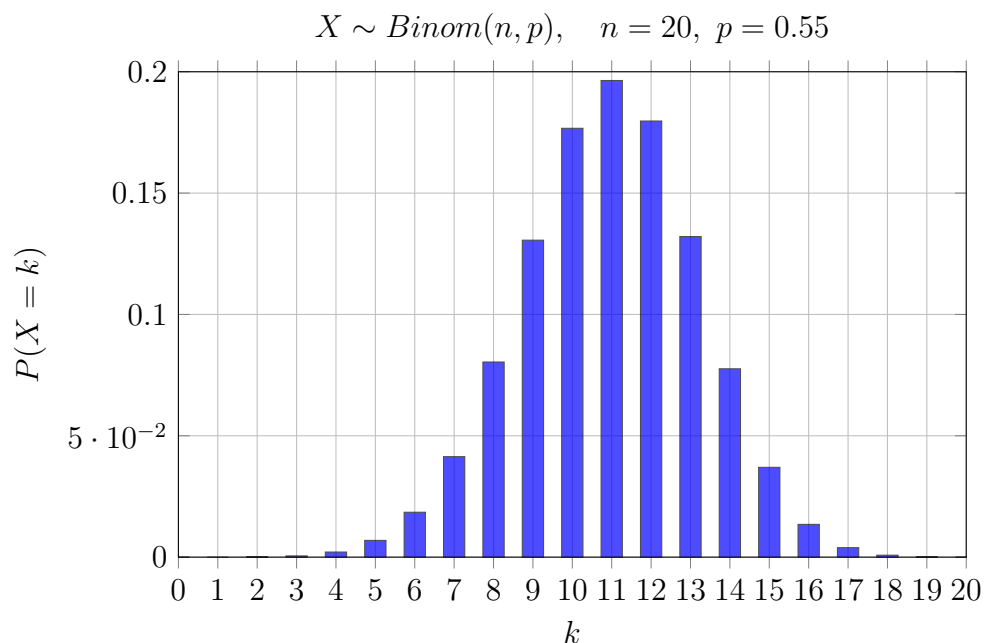
Биномиальное распределение, $X \sim \text{Binom}(n, p)$

Биномиальное распределение моделирует количество успехов в n независимых испытаниях. Таким образом, множество значений случайной величины X есть

$$\mathcal{X} = \{0, 1, \dots, n\},$$

а распределение X будет следующим:

$$P_X(k) = C_n^k p^k (1-p)^{n-k}.$$



Геометрическое распределение, $X \sim \text{Geom}(p)$

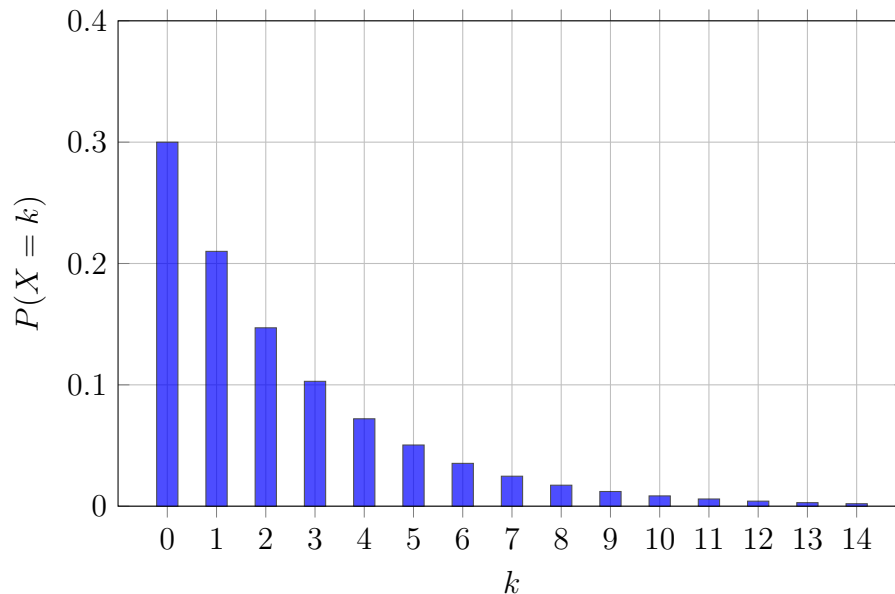
Случайная величина $X \sim \text{Geom}(p)$ моделирует число неудач до первого успеха, когда вероятность успеха равна p . Таким образом, множество значений X есть

$$\mathcal{X} = \{0, 1, 2, \dots\} = \mathbb{N} \cup \{0\},$$

а распределение будет иметь вид

$$P_X(k) = p(1-p)^k.$$

$$X \sim \text{Geom}(p), p = 0.3$$



Геометрическим это распределение называется потому, что сумма

$$\sum_{k=0}^{\infty} P_X(k) = p \sum_{k=0}^{\infty} (1-p)^k = \frac{p}{1-(1-p)} = 1$$

является геометрической прогрессией.

Распределение Пуассона, $X \sim \text{Poisson}(\lambda)$

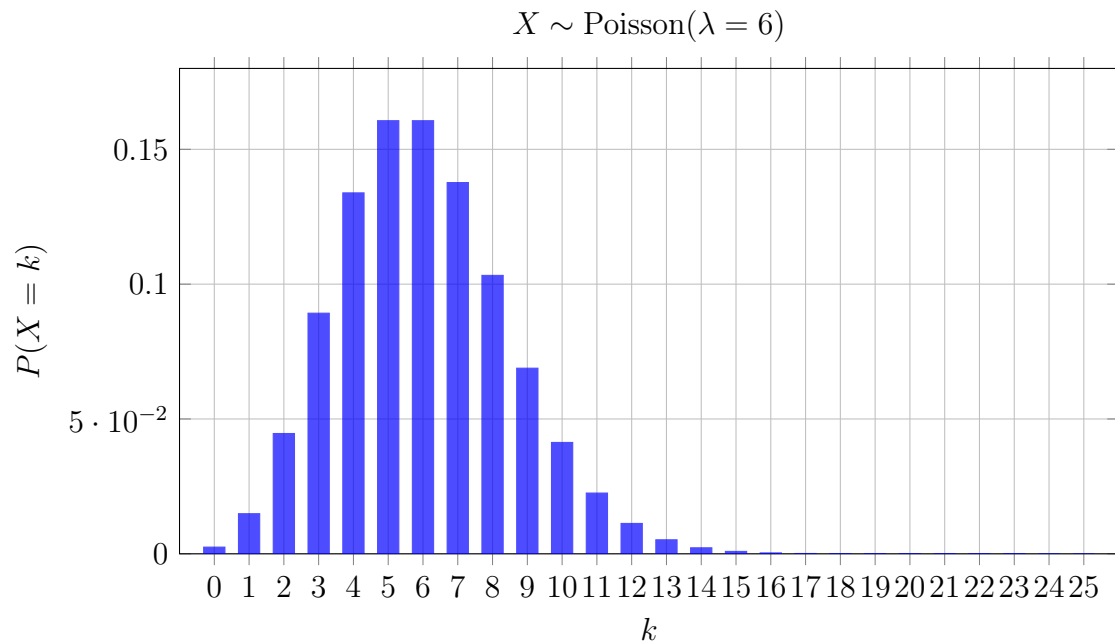
Пуассоновское распределение моделирует количество событий, произошедших за некоторое время, причем о событии известно среднее количество λ его появлений за этот промежуток времени.

Итак, множество значений величины $X \sim \text{Poisson}(\lambda)$ будет

$$\mathcal{X} = \{0, 1, 2, \dots\},$$

а распределение задается следующим образом

$$P_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \lambda > 0.$$



Примеров, которые описывались бы пуассоновской случайной величиной, много. Нужно, чтобы события были маловероятными, но при этом их было достаточно много. Например, число автомобильных аварий в Москве за день, число забиваемых голов и т.д.

Полезный источник: учебник Яндекса по ML.

Совместные распределения

Если на одном вероятностном пространстве (Ω, \mathcal{F}, P) задано несколько случайных величин X_1, \dots, X_n , то говорят, что на этом вероятностном пространстве задан случайный вектор (X_1, \dots, X_n) .

Такой случайный вектор задает новое вероятностное пространство

$$\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_n, \quad \mathcal{F}_X = 2^{\mathcal{X}}$$

с вероятностной мерой P_X , заданной формулой

$$P_X((x_{1,i_1}, \dots, x_{n,i_n})) = P(X_1 = x_{1,i_1}, \dots, X_n = x_{n,i_n}).$$

Эта мера называется совместным распределением случайных величин X_1, \dots, X_n .

Задача (Задача 3.16 из [9]). Совместное распределение $p_{ij} = P(X_1 = i, X_2 = j)$ случайных величин X_1, X_2 задано таблицей:

$X_1 \backslash X_2$	-1	0	1
-1	1/8	1/12	7/24
1	5/24	1/6	1/8

Найти а) одномерные распределения $p_i = P(X_1 = i)$ и $p_j = P(X_2 = j)$; б) совместное распределение $q_{ij} = P(Y_1 = i, Y_2 = j)$ случайных величин $Y_1 = X_1 + X_2$, $Y_2 = X_1 X_2$, в) одномерные распределения $q_i = P(Y_1 = i)$, $q_j = P(Y_2 = j)$.

□ Пункт а). Найдем $P(X_1 = -1)$. Для этого воспользуемся разбиением этого события (почти формула полной вероятности):

$$\begin{aligned} P(X_1 = -1) &= P(X_1 = -1, X_2 = -1) + P(X_1 = -1, X_2 = 0) + P(X_1 = -1, X_2 = 1) = \\ &= \frac{1}{8} + \frac{1}{12} + \frac{7}{24} = \frac{1}{2}. \end{aligned}$$

Отсюда становится понятно, что маргинальные распределения легко найти из таблички совместного распределения путем простого сложения вероятностей по строчкам или столбцам.

Теперь легко получаем:

$$\begin{aligned} P(X_1 = 1) &= \frac{5}{24} + \frac{1}{6} + \frac{7}{24} = \frac{1}{2} \\ P(X_2 = -1) &= \frac{1}{8} + \frac{5}{24} = \frac{1}{3} \\ P(X_2 = 0) &= \frac{1}{12} + \frac{1}{6} = \frac{1}{4} \\ P(X_2 = 1) &= \frac{7}{24} + \frac{1}{8} = \frac{5}{12}. \end{aligned}$$

Пункт б). Теперь найдем совместное распределение величин Y_1 и Y_2 . Для этого по сути нужно заполнить табличку

$Y_1 \backslash Y_2$	-1	0	1
-2			
-1			
0			
1			
2			

Начнем.

$$\begin{aligned} P(Y_1 = -2, Y_2 = -1) &= P(X_1 + X_2 = -2, X_1 X_2 = -1) = \\ &= P(X_1 = -1, X_2 = -1, X_1 X_2 = -1) = 0. \end{aligned}$$

Далее, уже очевидно, что

$$\begin{aligned} P(Y_1 = -2, Y_2 = 0) &= 0, \\ P(Y_1 = -2, Y_2 = 1) &= P(X_1 = -1, X_2 = -1) = \frac{1}{8}. \end{aligned}$$

Таким образом, первую строчку заполнили.

Далее, так как $Y_1 = -1 \Leftrightarrow X_1 = -1, X_2 = 0$, то получаем вторую строчку таблицы

$$\begin{aligned} P(Y_1 = -1, Y_2 = -1) &= 0, \\ P(Y_1 = -1, Y_2 = 0) &= P(X_1 = -1, X_2 = 0) = \frac{1}{12}, \\ P(Y_1 = -1, Y_2 = 1) &= 0. \end{aligned}$$

Для третьей строки таблицы заметим, что $Y_1 = 0 \Leftrightarrow X_1 = -1, X_2 = 1$ или $X_1 = 1, X_2 = -1$. Тогда

$$\begin{aligned} P(Y_1 = 0, Y_2 = -1) &= P(X_1 = -1, X_2 = 1) + P(X_1 = 1, X_2 = -1) = \frac{7}{24} + \frac{5}{24} = \frac{1}{2}, \\ P(Y_1 = 0, Y_2 = 0) &= 0, \\ P(Y_1 = 0, Y_2 = 1) &= 0. \end{aligned}$$

Для четвертой строчки: $Y_1 = 1 \Leftrightarrow X_1 = 1, X_2 = 0$, поэтому

$$P(Y_1 = 1, Y_2 = -1) = 0,$$

$$P(Y_1 = 1, Y_2 = 0) = P(X_1 = 1, X_2 = 0) = \frac{1}{6},$$

$$P(Y_1 = 1, Y_2 = 1) = 0.$$

И наконец для пятой строчки $Y_1 = 2 \Leftrightarrow X_1 = 1, X_2 = 1$, значит,

$$P(Y_1 = 2, Y_2 = -1) = 0,$$

$$P(Y_1 = 2, Y_2 = 0) = 0,$$

$$P(Y_1 = 2, Y_2 = 1) = P(X_1 = 1, X_2 = 1) = \frac{1}{8}.$$

Итого, получаем совместное распределение Y_1 и Y_2 :

$Y_1 \backslash Y_2$	-1	0	1
-2	0	0	1/8
-1	0	1/12	0
0	1/2	0	0
1	0	1/6	0
2	0	0	1/8

Пункт в) Маргинальные распределения тогда находятся легко:

i	-2	-1	0	1	2
$P(Y_1 = i)$	1/8	1/12	1/2	1/6	1/8

j	-1	0	1
$P(Y_2 = j)$	1/2	1/4	1/4

■

Задача. Привести пример, когда нельзя однозначно восстановить совместное распределение по маргинальным.

□ Пусть X_1, X_2 – случайные величины с распределениями $Bern(1/2)$:

$$P(X_1 = 0) = P(X_1 = 1) = 1/2,$$

$$P(X_2 = 0) = P(X_2 = 1) = 1/2.$$

Тогда нужно подобрать числа в таблицу совместного распределения, чтобы все суммы сходились. Это легко сделать (попробуйте сами):

$X_1 \backslash X_2$	0	1
0	1/6	1/3
1	1/3	1/6

$X_1 \backslash X_2$	0	1
0	1/2	0
1	0	1/2

■

Задача (Задача 3.49 из [9]). В схеме Бернулли с вероятностью успеха p обозначим ν_k ($k = 1, 2, \dots$) номер испытания, при котором происходит k -й успех, и положим

$$\tau_1 = \nu_1, \quad \tau_k = \nu_k - \nu_{k-1} \quad (k = 2, 3, \dots).$$

Найти совместное распределение величин τ_1, τ_k . Являются ли эти величины независимыми?

□ Введем случайные величины $X_i \sim \text{Bern}(p)$ – результат i испытания. Тогда

$$\begin{aligned} P(\tau_1 = l_1, \tau_k = l_2) &= P(X_1 = 0, \dots, X_{l_1-1} = 0, X_{l_1} = 1, \\ &\quad X_{\nu_{k-1}} = 0, \dots, X_{\nu_{k-1}+l_2-1} = 0, X_{\nu_k} = 1) = \\ &= (1-p)^{l_1-1} p \cdots (1-p)^{l_2-1} p = (1-p)^{l_1+l_2-2} p^2. \end{aligned}$$

При этом

$$P(\tau_1 = l_1, \tau_k = l_2) = P(\tau_1 = l_1)P(\tau_k = l_2),$$

значит, величины τ_1 и τ_k независимы. ■

Задача (Задача 3.51 из [9]). Игральную кость бросают до того момента, когда впервые выпадает меньше пяти очков. Обозначим через θ число очков, выпавших при последнем бросании игральной кости, и через ν – число бросаний кости. Найти совместное распределение θ и ν . Являются ли случайные величины θ и ν зависимыми?

□ Определимся с тем, какие значения принимают случайные величины θ и ν . Ясно, что

$$\theta \in \{1, 2, 3, 4\}, \quad \nu \in \{1, 2, 3, \dots\}.$$

Нам нужно найти вероятности $P(\theta = i, \nu = j)$. Событие $\{\theta = i, \nu = j\}$ означает, что $j-1$ раз выпадали числа ≥ 5 , а на j -ый раз выпало i . Так как последовательные бросания кости считаются независимыми, то получаем, что вероятность этого события равна

$$P(\theta = i, \nu = j) = \left(\frac{2}{6}\right)^{j-1} \cdot \frac{1}{6} = \frac{1}{6} \left(\frac{1}{3}\right)^{j-1}.$$

В задачнике предлагается выразить то же событие в терминах новой случайной величины ξ_i – результат i -го броска кубика. Тогда получим

$$\begin{aligned} P(\theta = i, \nu = j) &= P(\xi_1, \dots, \xi_{j-1} \leq 4, \xi_j = i) = \\ &= \left(\frac{2}{6}\right)^{j-1} \cdot \frac{1}{6} = \frac{1}{6} \left(\frac{1}{3}\right)^{j-1}. \end{aligned}$$

Найдем маргинальные распределения.

$$P(\theta = i) = \sum_{j=1}^{\infty} \frac{1}{6} \left(\frac{1}{3}\right)^{j-1} = \frac{1}{6} \cdot \frac{1}{1 - \frac{1}{3}} = \frac{1}{4}.$$

И для ν :

$$P(\nu = j) = \sum_{i=1}^4 P(\theta = i, \nu = j) = 4 \cdot \frac{1}{6} \left(\frac{1}{3}\right)^{j-1} = \frac{2}{3} \left(\frac{1}{3}\right)^{j-1}.$$

Отсюда мы видим, что выполняется равенство

$$P(\theta = i, \nu = j) = P(\theta = i)P(\nu = j),$$

значит, величины θ и ν независимы. ■

Математическое ожидание и дисперсия

Представим, случайная величина X с распределением

x	x_1	x_2	\cdots	x_{n-1}	x_n
$P(X = x)$	p_1	p_2	\cdots	p_{n-1}	p_n

замеряется N раз. Будем считать, что у нас N случайных величин X_1, \dots, X_N . Каково будет среднее арифметическое всех результатов?

Разумно предположить следующую логику. Величины p_i интуитивно должны означать долю испытаний, когда X приняло значение x_i . То есть за N повторений мы будем ожидать $N \cdot p_i$ раз, когда выпадет x_i . Тогда среднее арифметическое будет таким

$$\frac{X_1 + X_2 + \cdots + X_N}{N} \approx \frac{x_1 \cdot Np_1 + x_2 \cdot Np_2 + \cdots + x_n \cdot Np_n}{N} = x_1p_1 + x_2p_2 + \cdots + x_np_n.$$

Опр. Число

$$EX = x_1p_1 + x_2p_2 + \cdots + x_np_n$$

называется математическим ожиданием случайной величины X (expectation).

Опр. Число

$$\text{Var}X = E(X - EX)^2$$

называется дисперсией случайной величины X (variance).

Пример. Распределение Бернулли: $X \sim \text{Bern}(p)$ — бросание монетки. X принимает два значения: 1 с вероятностью p (успех) и 0 с вероятностью $1 - p$ (неуспех).

□

$$EX = 0 \cdot (1 - p) + 1 \cdot p = p$$

■

Пример. Биномиальное распределение: $X \sim \text{Binom}(n, p)$ — число орлов при бросании монетки n раз.

□

$$\begin{aligned} EX &= \sum_{k=0}^n k \cdot P(X = k) = \sum_{k=0}^n k \cdot C_n^k p^k (1 - p)^{n-k} \\ &= \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} p^k (1 - p)^{n-k} = \\ &= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1 - p)^{(n-1)-(k-1)} = np \end{aligned}$$

Но можно и проще, если заметить, что $X = X_1 + \dots + X_n$, где $X_k \sim \text{Bern}(p)$:

$$EX = EX_1 + \dots + EX_n = np.$$

■

Теорема. Пусть X, Y — дискретные случайные величины. Тогда

1. $E(cX) = cEX$, где $c \in \mathbb{R}$
2. $E(X + Y) = EX + EY$

$$3. X \geq 0 \Rightarrow EX \geq 0$$

$$4. Y \leq X \Rightarrow EY \leq EX$$

□ 1. Так как случайная величина X принимает конечное число значений, то ее можно представить в виде линейной комбинации индикаторов:

$$X = \sum_{i=1}^n a_i I_{A_i}.$$

Тогда

$$\begin{aligned} cX &= \sum_{i=1}^n (ca_i) I_{A_i} \\ E(cX) &= \sum_{i=1}^n ca_i P(A_i) = c \sum_{i=1}^n a_i P(A_i) = cEX. \end{aligned}$$

2. Аналогично, запишем случайные величины как суммы индикаторов:

$$X = \sum_{i=1}^n a_i I_{A_i}, \quad Y = \sum_{i=1}^m b_i I_{B_i}.$$

Тогда сумма $X + Y$ принимает значения $a_i + b_j$ на множествах $A_i \cap B_j$, т.е.

$$X + Y = \sum_{i=1}^n \sum_{j=1}^m (a_i + b_j) I_{A_i \cap B_j}.$$

Тогда математическое ожидание

$$\begin{aligned} E(X + Y) &= \sum_{i=1}^n \sum_{j=1}^m (a_i + b_j) P(A_i \cap B_j) = \\ &= \sum_{i=1}^n \sum_{j=1}^m a_i P(A_i \cap B_j) + \sum_{i=1}^n \sum_{j=1}^m b_j P(A_i \cap B_j) = \\ &= \sum_{i=1}^n a_i \sum_{j=1}^m P(A_i \cap B_j) + \sum_{j=1}^m b_j \sum_{i=1}^n P(A_i \cap B_j) = \\ &= \sum_{i=1}^n a_i P(A_i) + \sum_{j=1}^m b_j P(B_j) = EX + EY. \end{aligned}$$

3. $X \geq 0 \Rightarrow X(\omega) \geq 0 \quad \forall \omega \in \Omega$. Поэтому $\forall i \quad a_i \geq 0$ и

$$EX = \sum_{i=1}^n a_i I_{A_i} \geq 0.$$

4. $Y \leq X \Rightarrow X - Y \geq 0 \Rightarrow E(X - Y) = E(X + (-Y)) = EX + E(-Y) = EX - EY \geq 0$. ■

Теорема. Пусть X, Y – независимые случайные величины. Тогда

$$E(XY) = EXEY.$$

□ Представим случайные величины в виде комбинации индикаторов:

$$X = \sum_{i=0}^n a_i I_{A_i}, \quad Y = \sum_{j=0}^m b_j I_{B_j}.$$

Тогда

$$\begin{aligned} E(XY) &= E\left(\sum_{i=0}^n a_i I_{A_i} \sum_{j=0}^m b_j I_{B_j}\right) = E\left(\sum_{i=0}^n \sum_{j=0}^m a_i b_j I_{A_i} I_{B_j}\right) = \\ &= E\left(\sum_{i=0}^n \sum_{j=0}^m a_i b_j I_{A_i \cap B_j}\right) = \sum_{i=0}^n \sum_{j=0}^m a_i b_j P(A_i \cap B_j) = \\ &= \sum_{i=0}^n \sum_{j=0}^m a_i b_j P(A_i) P(B_j) = \sum_{i=0}^n a_i P(A_i) \sum_{j=0}^m b_j P(B_j) = \\ &= EXEY. \end{aligned}$$

■

Пример. Геометрическое распределение: $X \sim \text{Geom}(p)$ — число бросков до первого орла (успеха).

□ Сначала отметим, что $P(X = k) = p(1 - p)^k$. Тогда

$$EX = \sum_{k=0}^{\infty} kp(1 - p)^k.$$

Чтобы такое вычислить, нужно использовать трюк. Рассмотрим функцию

$$\varphi_X(s) = \sum_{k=0}^{\infty} s^k p(1 - p)^k$$

(эта функция называется производящей).

Тогда $\varphi'_X(1) = EX$.

Считаем:

$$\begin{aligned} \varphi_X(s) &= \sum_{k=0}^{\infty} s^k p(1 - p)^k = p \sum_{k=0}^{\infty} (s(1 - p))^k = p \frac{1}{1 - s(1 - p)} \\ \varphi'_X(s) &= \frac{p(1 - p)}{(1 - s(1 - p))^2} \\ \varphi'_X(1) &= \frac{1 - p}{p} \end{aligned}$$

■

Пример. Пусть есть N ячеек, в которые бросают n шаров. Шары могут попадать в одни и те же ячейки. X — число занятых ячеек.



□ Представим $X = X_1 + \dots + X_N$, где X_k — занята ли k -я ячейка. Тогда

$$EX = \sum EX_k = NEX_1$$

так как все равно какую ячейку рассматриваем.

Далее,

$$EX_1 = 0 \cdot P(X_1 = 0) + 1 \cdot P(X_1 = 1) = P(X_1 = 1) = 1 - P(X_1 = 0) = 1 - \left(\frac{N-1}{N}\right)^n$$

Таким образом,

$$EX = N \left(1 - \left(\frac{N-1}{N}\right)^n\right)$$

■

Задача (Рулетка). В комнате n человек, у каждого револьвер, по звонку каждый стреляет в случайного человека (кроме себя). Пусть X_n — число выживших. Найти EX_n .

□ Введем Y_i :

$$Y_i = \begin{cases} 1, & \text{если } i\text{-ый выжил,} \\ 0, & \text{если } i\text{-ый не выжил} \end{cases}$$

Тогда $X_n = \sum_{k=1}^n Y_k$ и $EX_n = \sum_{k=1}^n EY_k$.

$$EY_k = P(k\text{-ый выжил})$$

Чтобы посчитать вероятность, нужно задать вероятностное пространство. Зададим

$$\Omega = \{(x_1, \dots, x_n) : x_i \neq i\}$$

Тогда

$$\begin{aligned} EY_k &= P(k\text{-ый выжил}) = \frac{(n-2) \cdots (n-2)(n-1)(n-2) \cdots (n-2)}{(n-1)^n} = \\ &= \frac{(n-2)^{n-1}(n-1)}{(n-1)^n} = \left(\frac{n-2}{n-1}\right)^{n-1} \end{aligned}$$

Значит,

$$EX_n = n \cdot \left(\frac{n-2}{n-1}\right)^{n-1}$$

Отсюда можно заметить, что доля выживших

$$\frac{EX_n}{n} = \left(\frac{n-2}{n-1}\right)^{n-1} = \left(1 - \frac{1}{n-1}\right)^{n-1} \rightarrow \frac{1}{e}.$$

■

Вернемся к дисперсии.

Теорема. Пусть X — случайная величина. Тогда

$$\text{Var} X = EX^2 - (EX)^2.$$

□

$$\text{Var} X = E(X - EX)^2 = E(X^2 - 2XEX + (EX)^2) = EX^2 - (EX)^2.$$

■

Теорема. Пусть X — случайная величина. Тогда

$$1. \operatorname{Var}(cX) = c^2 \operatorname{Var} X$$

$$2. \operatorname{Var}(X_1 + X_2) = \operatorname{Var} X_1 + \operatorname{Var} X_2 + 2\operatorname{Cov}(X_1, X_2), \text{ где}$$

$$\operatorname{Cov}(X_1, X_2) = E(X_1 - EX_1)(X_2 - EX_2)$$

$$3. \operatorname{Cov}(X_1, X_2) = EX_1X_2 - EX_1EX_2$$

$$4. \operatorname{Cov}(\lambda X_1, X_2) = \lambda \operatorname{Cov}(X_1, X_2)$$

$$\operatorname{Cov}(X_1, X_2) = \operatorname{Cov}(X_2, X_1)$$

$$\operatorname{Cov}(X_1, X_2 + X_3) = \operatorname{Cov}(X_1, X_2) + \operatorname{Cov}(X_1, X_3)$$

$$5. \operatorname{Cov}(X, X) = \operatorname{Var} X \geq 0$$

$$6. \text{ Если } X, Y \text{ — независимы, то } \operatorname{Cov}(X, Y) = 0$$

□ Докажем (6). Так как X, Y независимы, то и $X - EX$ и $Y - EY$ независимы. Поэтому

$$\operatorname{Cov}(X, Y) = E(X - EX)(Y - EY) = (EX - EX)(EY - EY) = 0.$$

■

Коэффициент корреляции Пирсона

Опр. Коэффициент корреляции Пирсона

$$\rho = \frac{\operatorname{Cov}(X_1, X_2)}{\sqrt{\operatorname{Var} X_1} \cdot \sqrt{\operatorname{Var} X_2}}$$

Можно заметить, что $\rho \in [-1, 1]$.

TODO: геометрия корреляции и основные теоремы.

Задачи

Задача (Туфли). У Ярославы 10 пар туфель. Собака утащила 12 случайных туфель. Сколько в среднем целых пар осталось у Ярославы?

□ Обозначим

$$X_i = \begin{cases} 1, & \text{i-я пара не тронута} \\ 0, & \text{в i-й паре не хватает хотя бы одной туфли} \end{cases}$$

Тогда $X = \sum X_i$, значит, по линейности математического ожидания

$$EX = \sum EX_i = nEX_1.$$

Осталось посчитать EX_1 .

$$\begin{aligned} EX_1 &= P(\text{обе туфли в 1-й паре нетронуты}) = \\ &= \frac{C_{2n-2}^k}{C_{2n}^k}. \end{aligned}$$

В случае $n = 10$, $k = 12$ получаем

$$EX_1 = \frac{C_{18}^{12}}{C_{20}^{12}} = \frac{14}{95}.$$

Тогда

$$EX = nEX_1 = 10 \cdot \frac{14}{95} = \frac{28}{19} \approx 1.47.$$

■

Задача (Круглый стол). 10 пар супругов сели за круглый стол. Определим X – число сидящих напротив пар. Найти а) EX , б) $\text{Var}X$.

□ Найдем сначала математическое ожидание EX .

Для этого снова представим X как сумму бернуллиевских случайных величин. Определим

$$X_i = \begin{cases} 1, & \text{и-я пара сидит напротив друг друга} \\ 0, & \text{иначе} \end{cases}$$

Тогда $X = \sum_{i=1}^n X_i$, т.е.

$$EX = \sum_{i=1}^n EX_i = nEX_1.$$

Здесь мы воспользовались тем, что X_i имеют одинаковое распределение, а значит и математические ожидания.

Вычислим EX_1 :

$$EX_1 = P(\text{1-я пара сидит напротив}) = \frac{2n}{A_n^2} = \frac{2 \cdot 10}{A_{20}^2} = \frac{1}{19}.$$

Значит,

$$EX = nEX_1 = 10 \cdot \frac{1}{19} = \frac{10}{19}.$$

Теперь вычислим дисперсию $\text{Var}X$.

$$\text{Var}X = EX^2 - (EX)^2 = EX^2 - \left(\frac{10}{19}\right)^2.$$

Таким образом, нужно вычислить EX^2 .

$$\begin{aligned} EX^2 &= E\left(\sum_{i=1}^n X_i^2 + 2 \sum_{i < j} X_i X_j\right) = \\ &= E\left(\sum_{i=1}^n X_i^2\right) + 2 \sum_{i < j} E(X_i X_j) = \\ &= \sum_{i=1}^n EX_i^2 + 2 \cdot C_n^2 \cdot E(X_1 X_2). \end{aligned}$$

Заметим, что т.к. $EX_i = EX_i^2$, поэтому

$$\sum_{i=1}^n EX_i^2 = \sum_{i=1}^n EX_i = EX = \frac{10}{19}.$$

Произведение X_1X_2 будет снова бернуллиевской случайной величиной:

$$X_1X_2 = \begin{cases} 1, & \text{1 и 2 пары сидят напротив} \\ 0, & \text{иначе} \end{cases}$$

Таким образом,

$$E(X_1X_2) = P(1 \text{ и } 2 \text{ пары сидят напротив}) = \frac{2n \cdot (2n - 2)}{A_n^4} = \frac{20 \cdot 18}{A_{20}^4} = \frac{1}{19 \cdot 17}.$$

Получили, что

$$EX^2 = \frac{10}{19} + 2 \cdot \frac{10 \cdot 9}{2} \frac{1}{19 \cdot 17} = \frac{10}{19} + \frac{90}{19 \cdot 17} = \frac{260}{19 \cdot 17}.$$

А дисперсия

$$\text{Var}X = \frac{260}{19 \cdot 17} - \frac{100}{19 \cdot 19} = \frac{10}{19} \left(\frac{26}{17} - \frac{10}{19} \right) \approx 0.528.$$

■

Задача (Бегемотики). Внутри шоколадного яйца находится одна из 10 фигурок бегемотиков, причем все встречаются одинаково часто. Сколько а) в среднем их нужно купить, чтобы собрать всех животных, б) какова дисперсия этой величины?

□ Будем опять же решать в общем виде и затем подставим $n = 10$.

Пусть X – количество яиц, которое нужно купить для того, чтобы собрать всех бегемотиков.

Для вычисления EX используем линейность математического ожидания, для чего представим X как сумму X_i , где X_i – число яиц до выпадения i -го по очереди нового бегемотика после выпадения $(i - 1)$ -го. Тогда

$$X = \sum_{i=1}^n X_i.$$

Заметим, что величины X_i имеют разное распределение, но они независимы (кажется, мы это доказывали в одном из упражнений выше). Это нам пригодится при вычислении $\text{Var}X$.

Займемся вычислением EX :

$$EX = \sum_{i=1}^n EX_i.$$

Далее,

$$\begin{aligned} EX_i &= 1 \cdot P(X_i = 1) + 2 \cdot P(X_i = 2) + \dots + k \cdot P(X_i = k) + \dots = \\ &= 1 \cdot \frac{n - i + 1}{n} + 2 \cdot \frac{i - 1}{n} \cdot \frac{n - i + 1}{n} + \dots + \\ &\quad + k \cdot \left(\frac{i - 1}{n} \right)^{k-1} \cdot \frac{n - i + 1}{n} + \dots = \\ &= \frac{n - i + 1}{n} \left(1 + 2 \cdot \frac{i - 1}{n} + \dots + k \cdot \left(\frac{i - 1}{n} \right)^{k-1} + \dots \right) \end{aligned}$$

Для вычисления суммы в скобке заметим, что она равна значению функции

$$p(x) = 1 + 2x + 3x^2 + \dots + kx^{k-1} + \dots$$

в точке $x = \frac{i-1}{n}$. При этом заметим, что

$$p(x) = q'(x),$$

где

$$q(x) = x + x^2 + x^3 + \dots + x^k + \dots = \frac{x}{1-x},$$

по формуле суммы геометрической прогрессии. Отсюда

$$p(x) = q'(x) = \frac{1}{(1-x)^2}.$$

Подставляя $x = \frac{i-1}{n}$, получаем

$$EX_i = \frac{n-i+1}{n} \cdot \frac{1}{\left(1 - \frac{i-1}{n}\right)^2} = \frac{n}{n-i+1}.$$

Тогда

$$EX = \sum_{i=1}^n EX_i = \sum_{i=1}^n \frac{n}{n-i+1} = n \sum_{i=1}^n \frac{1}{n-i+1} = n \sum_{i=1}^n \frac{1}{i}.$$

При $n = 10$ получаем $EX \approx 29.29$.

Если вспомнить асимптотическую формулу суммы гармонического ряда

$$\sum_{i=1}^n \frac{1}{i} = \ln n + \gamma + o(1), \quad n \rightarrow \infty$$

(здесь $\gamma \approx 0.5772$ – постоянная Эйлера–Маскерони), то можно получить асимптотическую формулу и для EX при больших n :

$$EX = n(\ln n + \gamma) + o(n), \quad n \rightarrow \infty.$$

Найдем теперь дисперсию величины X . Начнем со стандартной формулы

$$\text{Var} X = EX^2 - (EX)^2.$$

Необходимо вычислить EX^2 . Для этого снова пользуемся стандартным представлением

$$EX^2 = \sum_{i=1}^n EX_i^2 + 2 \sum_{i < j} E(X_i X_j)$$

Напомним, что величины X_i, X_j независимы, а значит

$$EX^2 = \sum_{i=1}^n EX_i^2 + 2 \sum_{i < j} EX_i EX_j,$$

и второе слагаемое мы знаем. Остается вычислить EX_i^2 .

$$EX_i^2 = \frac{n-i+1}{n} \left(1^2 + 2^2 \frac{i-1}{n} + \dots + k^2 \left(\frac{i-1}{n} \right)^{k-1} + \dots \right)$$

Сумма в скобке равна значению функции

$$p(x) = 1^2 + 2^2 x + 3^2 x^2 + \dots + k^2 x^{k-1} + \dots$$

в точке $x = \frac{i-1}{n}$.

Рассмотрим ее первообразную

$$\begin{aligned} q(x) &= \int p(x)dx = x + 2x^2 + 3x^3 + \dots kx^k + \dots = \\ &= x(1 + 2x + 3x^2 + \dots + kx^{k-1} + \dots) = \\ &= x \cdot \frac{1}{(1-x)^2} = \frac{x}{(1-x)^2} \end{aligned}$$

Здесь мы воспользовались ранее посчитанной суммой.

Теперь можем найти $p(x)$:

$$p(x) = q'(x) = \frac{1+x}{(1-x)^3}$$

Подставляя $x = \frac{i-1}{n}$, получаем

$$EX_i^2 = \frac{n-i+1}{n} \cdot \frac{1 + \frac{i-1}{n}}{(1 - \frac{i-1}{n})^3} = n \cdot \frac{n+i-1}{(n-i+1)^2}$$

Подсчет дает ответ

$$EX^2 = \sum_{i=1}^{10} EX_i^2 + 2 \sum_{i < j} EX_i EX_j \approx 983.57.$$

Значит,

$$\text{Var}X = EX^2 - (EX)^2 \approx 983.57 - 857.9 = 125.67.$$

Красивый метод решения.

Величины X_i имеют геометрическое распределение $X_i \sim \text{Geom}(p_i)$, где $p_i = \frac{n-i+1}{n}$. Это было показано выше. Зная свойства геометрического распределения, можно быстро вычислить EX_i и $\text{Var}X_i$:

$$EX_i = \frac{1}{p_i}, \quad \text{Var}X_i = \frac{1-p_i}{p_i^2}.$$

Так как X_i независимы, то

$$\begin{aligned} \text{Var}X &= \sum_{i=1}^n \text{Var}X_i = \sum_{i=1}^n \frac{1-p_i}{p_i^2} = \sum_{i=1}^n \frac{n(i-1)}{(n-i+1)^2} = \\ &= |j = n - k + 1| = \sum_{j=1}^n \frac{n(n-j)}{j^2} = \\ &= n^2 \sum_{j=1}^n \frac{1}{j^2} - n \sum_{j=1}^n \frac{1}{j} = \\ &= n^2 H_n^{(2)} - n H_n, \end{aligned}$$

где

$$\begin{aligned} H_n &= \sum_{k=1}^n \frac{1}{k} = \ln n + \gamma + o(1) \\ H_n^{(2)} &= \sum_{k=1}^n \frac{1}{k^2} = \frac{\pi^2}{6} - \frac{1}{n} + O\left(\frac{1}{n^2}\right). \end{aligned}$$

Откуда получаем асимптотическую формулу для дисперсии

$$\mathrm{Var} X = \frac{n^2 \pi^2}{6} - n \ln n + O(n).$$

■

Асимптотические формулы хороши тем, что показывают поведение "последовательности на бесконечности", давая не точную формулу значения n -го члена последовательности, а приближенную формулу с главной частью и остатком. Остаток всегда можно уточнить, но не всегда это технически реально осуществить.

Пределы

На данный момент построенная теория дает скорее базовые элементы вероятностного языка, можно сказать, морфемы, из которых можно строить осмысленные слова. Осмысленные с точки зрения исходной идеи о прикладной, физической природе понятия вероятности как мере знания о том, что некоторое событие произойдет.

Задумаемся над тем, что нам доступно на текущем этапе, чему можно было научиться, пройдя с начала до этого момента? На этот вопрос ответить нетрудно: достаточно перечислить пройденные темы и сделать правильный вывод. Мы подробно рассмотрели понятие вероятности, то бишь научились считать вероятность событий. Случайные величины можно рассматривать как мощный язык опять же для нахождения вероятностей событий, описываемых удобным способом.

Да, характеристики случайных величин, такие как математическое ожидание, дисперсия, ковариация и корреляция, несут новую смысловую нагрузку, не чисто вероятностную. Но тогда какую?

Так как все упомянутые понятия являются производными от математического ожидания, то и говорить стоит именно о нем. Его значение нам кажется очевидным: EX это как бы "среднее" значение случайной величины X . Но если вспомнить мотивацию определения $EX = \sum x_i p_i$, можно заметить, что была использована идея о том, что p_i – не просто абстрактная вероятность значения x_i . Было сделано предположение, что при "многократном повторении эксперимента", доля тех исходов, при которых $X = x_i$, будет равна p_i . Но с какой стати это должно выполняться и что вообще значит "многократное повторение эксперимента"?

Напомним, что теория вероятностей строится на принципе меры, меры имеющейся информации об эксперименте. Вероятность субъективна. Субъективность кроется в построении модели, в нашем случае, моделью выступает вероятностное пространство. В рамках выбранной модели уже можно вычислять вероятности, определять случайные величины и находить их математические ожидания. То есть все это также субъективно. Это было важно подчеркнуть.

Идея предельных переходов состоит в том, чтобы рассматривать не один эксперимент, а бесконечное количество копий одного и того же эксперимента, то бишь для нас – вероятностного пространства. И тогда можно говорить о свойствах всех этих копий в совокупности.

На романтическом языке можно было бы сказать, что мы хотим изучать не одну конкретную вселенную внутри самой себя, а свойства вселенной как объекта в окружении бесконечного количества ее копий, в которых могут волею случая происходить разные исходы одного и того же эксперимента.

Закон больших чисел говорит о том, что если случайным образом брать все большее количество копий модели и смотреть в них значение эксперимента, то среднее арифметическое полученных результатов будет в некотором вероятностном смысле приближаться к математическому ожиданию: мера знания о том, что отклонение среднего арифметического от мат. ожидания будет велико, будет стремиться к нулю.

В следующих параграфах разговор пойдет о предельных теоремах, а задачи будут о пределах последовательностей случайных величин.

Закон больших чисел

Законом больших чисел (The Law of Large Numbers) или, сокращенно, ЗБЧ (LLN), принято называть класс теорем о сходимости средних арифметических случайных величин. Мы рассмотрим и докажем только одну такую теорему, которая входит в

стандартный курс теории вероятностей.

Итак, дадим формулировку.

Теорема (ЗБЧ). Пусть $\{X_i\}_{i=1}^{\infty}$ – последовательность независимых одинаково распределенных случайных величин с конечным математическим ожиданием. Обозначим через

$$S_n = \sum_{i=1}^n X_i.$$

Тогда

$$\frac{S_n}{n} \xrightarrow{P} EX_1.$$

Прежде чем переходить к доказательству (оно будет состоять из нескольких шагов), детально разберемся с формулировкой.

1. Последовательность случайных величин $\{X_i\}_{i=1}^{\infty}$ определяется как любая последовательность из курса математического анализа: последовательность это биекция из \mathbb{N} в пространство, в нашем случае, это пространство случайных величин.
2. Независимость и одинаковая распределенность в практическом смысле означает то, что эксперимент проводится полностью одинаково. Формальный же смысл прямо следует из определений. Важно понимать, что если величины имеют одинаковое распределение, то они будут иметь одинаковое математическое ожидание.
3. Случайные величины S_n называются частичными суммами последовательности $\{X_i\}_{i=1}^{\infty}$.
4. Самое важное. Стрелка \xrightarrow{P} означает сходимость "по вероятности". Мы подробно обсудим это понятие ниже, а также см. задачи в конце раздела.

* * *

Центральная предельная теорема

Задачи

Сюжеты

Счастливые билеты

Счастливым билетом считается билет, у которого сумма первых трех цифр совпадает с суммой последних трех цифр. Номера билетов бывают от 000000 до 999999. Задача состоит в том, чтобы найти вероятность того, что человеку, севшему в трамвай, попадется счастливый билет.

При решении этой задачи выяснилось, что ей посвящено несколько статей в журнале "Квант" и еще несколько интересных статей с другими подходами к решению.

Возникла идея собрать все это вместе в единый сюжет, объединяющий в себе сразу несколько любопытнейших тем. Здесь будут и признаки делимости, и рекуррентные соотношения, и производящие функции, и даже интегралы. А в более сложных обобщениях этой задачи применяется даже Центральная предельная теорема.

При этом допускается решение и более менее в лоб. Впрочем, кто-то с этим утверждением поспорит. Об этом решении мы тоже поговорим, но не сразу. Будем идти по нарастающей в смысле сложности техники.

Итак, начнем мы с того, что обратимся к статье [3].

Понятно, что самое простое, что можно придумать, это перебор: просто перебрать весь миллион вариантов, проверить каждый, подсчитать и дать ответ. Имея компьютер, это сделать совсем просто. Но задача возникла раньше, когда трамваи и автобусы уже были, а компьютеров в каждом доме не было.

Как же можно себе упростить жизнь?

Соображения делимости

Утв. Если суммы цифр двух чисел a и b совпадают, то равны их остатки от деления на 9, т.е. $a \equiv b \pmod{9}$.

▷ Рассмотрим трехзначные числа для краткости записи.

$$a = 100k_1 + 10l_1 + m_1$$

$$b = 100k_2 + 10l_2 + m_2$$

Но тогда сразу получаем, что

$$a = k_1 + l_1 + m_1 \pmod{9}$$

$$b = k_2 + l_2 + m_2 \pmod{9}$$

То есть при взятии остатка от деления на 9 числа дают просто сумму цифр. Значит, если равна сумма цифр, то равны и остатки. ◁

Тогда посчитаем, сколько билетиков будут удовлетворять свойству равенства первого и второго трехзначных чисел по модулю 9. Такие билетики назовем красивыми.

Сколько трехзначных чисел дают остаток 0 при делении на 9, т.е. попросту делятся на 9? Таких ненулевых чисел $999/9 = 111$. И еще 000. Тогда билетиков, у которых первая половина делится на 9 и вторая половина делится на 9, будет $112^2 = 12544$.

Трехзначных чисел, дающих остаток 1, будет опять 111 (все те же, что давали остаток на ноль, только с прибавкой единицы, и без $999 + 1 = 1000$). Значит, билетиков, у которых обе части дают остаток 1 при делении на 9, будет $111^2 = 12321$. И аналогично со всеми остатками до 8.

Таким образом, красивых билетиков будет

$$12321 \cdot 8 + 12544 = 111112.$$

Итак, мы получили, что счастливых билетиков не больше, чем 111112. Но можно эту оценку улучшить.

Утв. Пусть n — натуральное число. Сложим все цифры, стоящие на нечетных разрядах, потом сложим цифры, стоящие в четных разрядах. Если разность полученных сумм делится на 11, то и число n делится на 11. И наоборот, если число n делится на 11, то оно обладает этим свойством.

▷ Представим произвольное натуральное число в виде разложения

$$n = 10^k a_k + 10^{k-1} a_{k-1} + 10^{k-2} a_{k-2} + \cdots + 10a_1 + a_0.$$

Тогда возьмем это выражение по модулю 11:

$$n \equiv (-1)^k a_k + (-1)^{k-1} a_{k-1} + (-1)^{k-2} a_{k-2} + \cdots + (-1)a_1 + a_0 \pmod{11}.$$

Отсюда и следует утверждение. ◁

Как же это нам может помочь? Дело в том, что если билет считался бы счастливым, если сумма цифр на четных местах равнялась сумме цифр на нечетных местах, то ответ в задаче бы не изменился. Таких счастливых билетов столько же, сколько и исходных. Москвичи называли такие билеты счастливыми по-ленинградски, а ленинградцы — счастливыми по-московски.

А если такие суммы совпадают, то их разность равна нулю, а значит, номер билета делится на 11. Таким образом, таких билетов меньше чем количество чисел от 0 до 999999, делящихся на 11. А таких чисел $999999/11 = 90909$. Поэтому и интересующих нас билетов не больше 90909. Вот так и улучшили оценку.

Мораль такова: задача не решена, но очень эффективно получена оценка. Неожиданными, но элементарными методами мы отбросили больше 900 тысяч вариантов. Никакой комбинаторики и прочих хитростей.

* * *

Теперь посмотрим, какими способами можно все-таки посчитать точное число счастливых билетов.

Рассмотрим практически лобовой вариант, использующий немного комбинаторики.

Итак, счастливый билет имеет равную сумму цифр в первой и второй половине шестизначного номера. То есть два трехзначных числа должны иметь одинаковую сумму цифр.

Сумма трех цифр может принимать значения от 0 до 27. Поэтому можно посчитать количество $N(k)$ всех трехзначных цифр с суммой цифр k . Тогда число билетов, у которых первая и вторая суммы будут равны k , будет $N(k)^2$. А значит, всего счастливых билетов будет

$$\sum_{k=0}^{27} N(k)^2.$$

Как же посчитать? Заметим сразу, что считать для всех k не нужно, достаточно для половины из них. Действительно, если сумма цифр трехзначного числа $\overline{a_1 a_2 a_3}$ равна k , то число $\overline{a'_1 a'_2 a'_3}$, где $a'_i = 9 - a_i$, будет иметь сумму цифр, равную $27 - k$. Таким образом, есть взаимно однозначное соответствие, то есть

$$N(k) = N(27 - k).$$

Хорошо, надо сосчитать $N(k)$ для $k \in \{0, 1, \dots, 13\}$. Чтобы сосчитать количество трехзначных чисел с суммой цифр k , нужно найти количество решений уравнения

$$a_1 + a_2 + a_3 = k,$$

где $a_1, a_2, a_3 \in \{0, \dots, 9\}$ — неизвестные.

Можно, конечно, как предлагается в [13], пройти перебором уже по трехзначным числам, что гораздо меньше (тысяча вариантов против миллиона). Но можно подумать о том, что эта задача очень напоминает четвертую задачу о шарах. Там мы выяснили, что количество решений уравнения

$$x_1 + \dots + x_N = k,$$

где $x_i \in \{0, \dots, k\}$ — неизвестные, равно C_{N+k-1}^{k-1} . Но проблема в том, что в нашем случае ограничение на слагаемые всегда $x_i \leq 9$. А значит,

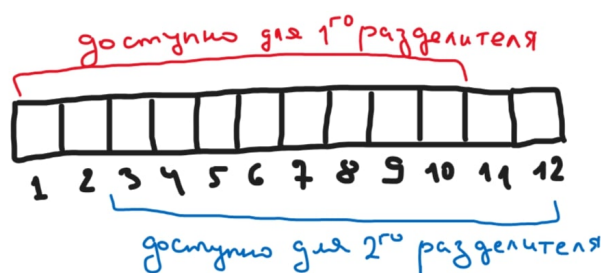
$$k \leq 9 \Rightarrow N(k) = C_{k+2}^2,$$

но для $k = 10, 11, 12, 13$ эта формула не подходит, придется посчитать отдельно.

Рассмотрим частный случай $k = 10$. Будем суммировать все варианты в зависимости от позиции первого разделителя. Если первый разделитель стоит на позиции 1, то второй разделитель должен стоять в любой доступной позиции, кроме тех, которые будут отделять среднюю часть больше 9. То есть второй разделитель не может стоять на позиции 12. Значит, ему доступно 9 позиций: от 3 до 11.

Если первый разделитель стоит в позиции 2, то второй разделитель может стоять на любой из 10 позиций от 3 до 12.

Если первый разделитель стоит в позициях 3-10, то второй может либо тоже стоять в позиции 3-10, но тогда они не должны совпадать, то есть вариантов всего $-\frac{8 \cdot 7}{2}$, либо второй стоит в позиции 11-12 и тогда вариантов $8 \cdot 2$.



Итого,

$$N(10) = 1 \cdot 9 + 1 \cdot 10 + \frac{8 \cdot 7}{2} + 8 \cdot 2 = 63.$$

Аналогично вычисляются $N(11), N(12), N(13)$.

И тогда можно выписать таблицу:

Результаты вычислений														
k	0	1	2	3	4	5	6	7	8	9	10	11	12	13
N(k)	1	3	6	10	15	21	28	36	45	55	63	69	73	75

Остается просто просуммировать

$$\sum_{k=0}^{27} N(k)^2 = 55252,$$

а тогда вероятность получить счастливый билет равна

$$\frac{55252}{1000000} = 0.055252,$$

то есть примерно каждый 18-ый билет будет счастливым.

Задачу можно считать решенной, все же выглядит это решение достаточно кустарно. Есть более красивые методы, рассмотрим сначала идею из [13].

Список литературы

- [1] Knuth D. *The Art of Computer Programming, Volume 2, Seminumerical Algorithms, 3rd edition*. 1997.
- [2] Левин А. *Что такое комбинаторика*. МЦНМО, 1995. URL: <https://kvant.mccme.ru/pdf/1999/05/kv0599levin.pdf>.
- [3] Савин А. и Финк Л. *Разговор в трамвае*. 1975. URL: https://kvant.mccme.ru/1975/07/razgovor_v_tramvae.htm (дата обр. 18.06.2025).
- [4] Булинский А.В. *Лекции по теории вероятностей на мехмате МГУ*. 2018.
- [5] Шкляев А.В. *Материалы к семинарам по теории вероятностей на мехмате МГУ*. 2024.
- [6] Колмогоров А.Н. *Основные понятия теории вероятностей*. Наука, 1974. URL: <https://www.klex.ru/q8k>.
- [7] Колмогоров А.Н., Журбенко И.Г. и Прохоров А.В. *Введение в теорию вероятностей*. Библиотечка "Квант", выпуск 141, 2024.
- [8] Подольский В.Е. *Лекции по математическому анализу на мехмате МГУ*. 2016.
- [9] Зубков, Севастьянов и Чистяков. *Сборник задач по теории вероятностей*. Издательство "Наука", 1989. URL: https://pdf.lib.vntu.edu.ua/books/2015/Zybko_1989_320.pdf.
- [10] Куратовский К. и Мостовский А. *Теория множеств*. Мир, 1970. URL: <https://ikfia.ysn.ru/wp-content/uploads/2018/01/KuratovskijMostovskij1970ru.pdf>.
- [11] Ландо С. К. *Счастливые билеты*. Матем. просв., выпуск 2, 127–132, 1998. URL: <https://www.mathnet.ru/links/dc27b0ef2b87d4a693fa509dd7bf4693/mp27.pdf>.
- [12] Журнал Квант. *Интегралом - по счастливым билетам*. 1978. URL: https://kvant.mccme.ru/1978/11/integralom_-_po_schastlivym_bi.htm (дата обр. 18.06.2025).
- [13] Финк Л. *Еще раз о счастливых билетах*. 1976. URL: https://kvant.mccme.ru/1976/12/eshche_raz_o_schastlivyh_bilet.htm (дата обр. 18.06.2025).
- [14] Дьяченко М.И. *Лекции по действительному анализу на мехмате МГУ*. 2018.
- [15] Н.К.Верецагин и А.Шень. *Начала теории множеств*. МЦНМО, 2012. URL: <https://old.mccme.ru/free-books/shen/shen-logic-part1-2.pdf>.
- [16] Н.К.Верецагин и А.Шень. *Языки и исчисления*. МЦНМО, 2012. URL: <https://old.mccme.ru/free-books/shen/shen-logic-part2-2.pdf>.
- [17] Виро О.Я. и др. *Элементарная топология*. МЦНМО. URL: <http://www.pdmi.ras.ru/~olegviro/topoman/rus-book.pdf>.
- [18] Чукова Ю. *Распределение Пуассона*. 1988. URL: https://kvant.mccme.ru/1988/08/raspredelenie_puassona.htm (дата обр. 03.09.2025).
- [19] Бернулли Я. *О законе больших чисел*. "Наука", 1986. URL: <https://www.klex.ru/29dd>.