Лекция 1. Вероятностные пространства и дискретные случайные величины

17 февраля 2022 г.

1 Вероятностное пространство

Вероятностное пространство — это тройка $(\Omega, \Sigma, \Pr)$

1. Ω — множество элементарных исходов
2. Σ — σ-алгебра событий
3. $\Pr$ — вероятностная мера

Рассмотрим каждый элемент тройки отдельно. В теорвере мы обычно рассматриваем какой-то эксперимент, у которого возможны разные исходы. Множество всех возможных исходов и есть Ω.

- Эксперимент не может закончиться сразу двумя исходами одновременно (исходы — *взаимоисключающие*).
- Эксперимент не может закончиться исходом не из Ω (Ω — *полное*).
- Множество Ω не должно быть черезчур подробным (Ω — *неизбыточное*).

Вот некоторые примеры Ω.

- Подбрасывание монеты:
  - $\Omega = \{$орел, решка$\}$
- Хоккейный матч (КХЛ, NHL):
  - $\Omega = \{$В, ВО, ВБ, ПБ, ПО, П$\}$
  - Если нас интересуют только очки одной команды, то $\Omega = \{0, 1, 2\}$
- Время ожидания автобуса на остановке:

– $\Omega = [0, +\infty)$

Событие — любое подмножество Ω (то есть, множество исходов). На Ω должна быть задана σ-алгебра событий Σ, то есть множество подмножеств, такое, что

1. $\emptyset \in \Sigma$,
2. $A \in \Sigma \Rightarrow \bar{A} \in \Sigma$,
3. $\{A_1, \ldots, A_n, \ldots\} \in \Sigma \Rightarrow \bigcap\limits_{i \in \mathbb{N}} A_i \in \Sigma$ и $\bigcup\limits_{i \in \mathbb{N}} A_i \in \Sigma$,

Примеры событий:

- Домашняя команда выиграла {В, ВО, ВБ}
- Монета упала орлом {орел}
- Автобус приехал в течение 5 минут $[0, 5]$ (если считаем время в минутах)

Вероятностная мера Pr — это функция, заданная на Σ, со следующими свойствами

1. *Неотрицательность:* $\Pr(A) \geq 0$ для любого события A
2. *Нормализация:* $\Pr(\Omega) = 1$
3. *Счетная аддитивнотсъ*: $A_1, \ldots, A_n, \ldots$ — последовательность попарно непересекающихся событий, тогда

$$\Pr\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} \Pr(A_i).$$

NB: В литературе встречаются обозначения $P, p, \mathbb{P}$, их можно использовать

Свойства Pr, следующие из аксиом:

- $\Pr(A) \leq 1$
- $\Pr(\emptyset) = 0$
- $\Pr(A) + \Pr(\bar{A}) = 1$
- $A \subset B \Rightarrow \Pr(A) \leq \Pr(B)$
- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
- $\Pr(A \cup B) \leq \Pr(A) + \Pr(B)$ — *Union bound*

В рамках данного курса мы будем работать с вероятностным пространством, а именно будем описывать множество элементарных исходов, определять события, определять вероятностную меру и потом делать какие-то выводы о построенной модели. Все, что мы будем делать — это на основе аксиом выводить разные теоремы вроде "вероятность события A равна p".

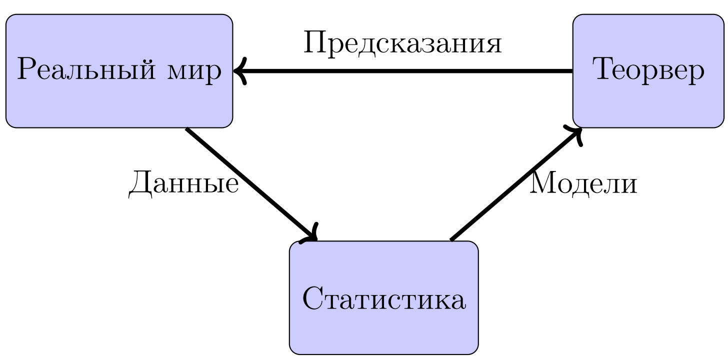
Однако несмотря на такую абстрактность, теорвер имеет прямое отношение к реальному миру, как и, например, интегральное исчисление, которое позволяет нам описывать очень много разных вещей в физике и предсказывать поведение объектов в тех или иных условиях. Теорвер описывает так называемые события и предсказывает их исход. В данном случае *вероятность* события можно интерпретировать двумя различными способами.

- Вероятность как *частота* события при многократном повторении эксперимента (*Пример:* если много раз кидать честную монетку, то примерно половина результатов будет "орел").

- Вероятность как *степень нашей веры в* событие (*Пример:* с вероятностью 0.1 Ливерпуль выиграет в этом году Лигу Чемпионов).

И хотя в нашем мире очень мало источников настоящей случайности (на самом деле они есть только на квантовом уровне), многие процессы и события реального мира могут быть описаны как случайные в силу того, что они зависят от слишком большого числа факторов, которые мы не способны учесть. И теория вероятностей позволяет нам предсказывать такие события с определенной степенью уверенности. Этим пользуются много в каких областях, например, для предсказания результатов выборов, для инвестиций, для азартных игр и во многих других областях.

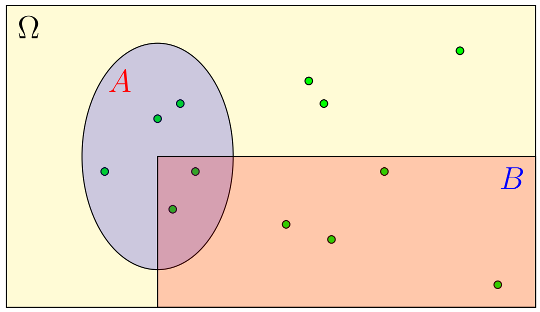
В целом взаимодействие теорвера с реальным миром можно описать следующей схемой:



Реальный мир дает некоторые данные, которые собирает статистика и на основе этих данных строится какая-то вероятностная модель (то есть наша тройка $(\Omega, \Sigma, \Pr)$). Теорвер работает с этой моделью и на ее основе делает какие-то предсказания о будущем в реальном мире. Затем реальный мир дает новые данные, которые могут совпадать или расходиться с предсказаниями, эти данные обрабатываются статистикой и так далее.

2 Условная вероятность

Рассмотрим понятие условной вероятности (то есть вероятности события при условии, что произошло другое событие) на следующем примере.



Пусть все исходы равновероятны (вероятность каждого $\frac{1}{12}$). Вероятность события A есть $\Pr(A) = 5 \cdot \frac{1}{12} = \frac{5}{12}$ Нам стало известно, что произошло событие B. Какой стала вероятность события A? В событие B входят 6 равновероятных исходов, из которых только 2 входят в событие A. То есть теперь вероятность события A есть $2 \cdot \frac{1}{6} = \frac{1}{3}$.

При появлении новой информации мы хотим изменить свою модель, чтобы она соответствовала новым условиям. Пусть мы знаем, что произошло событие B. Тогда мы хотим:

- $A : A \cap B = \emptyset \Rightarrow \Pr(A \mid B) = 0$
- $A \subset B \Rightarrow \Pr(A \mid B) = \frac{\Pr(A)}{\Pr(B)}$ (сохраняем нормализацию)

Любое событие можно представить как $A = (A \cap B) \cup (A \cap \bar{B})$, тогда по аддитивности вероятности

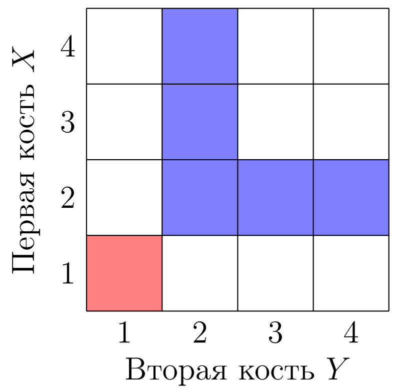
$$\Pr(A \mid B) = \Pr(A \cap B \mid B) + \Pr(A \cap \bar{B} \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} + 0.$$

Таким образом, мы пришли к определению вероятности события A при условии события B.

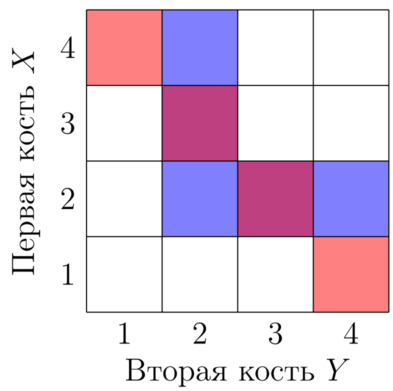
$$\boxed{\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}}$$

NB: $\Pr(\cdot \mid B)$ — это новая вероятностная мера, заданная на той же самой σ-алгебре, то есть для нее выполняются все аксиомы меры и следствия из них

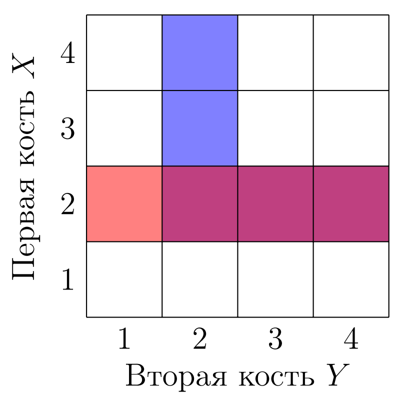
Пример условных вероятностей: бросок двух тетраэдральных костей. Пусть событие B есть $\min\{X, Y\} = 2$ (на рисунках обозначено синим).



$\Pr(\max\{X, Y\} = 1 \mid B) = 0$



$\Pr(X + Y = 5 \mid B) = \frac{2}{5}$



$\Pr(X = 2 \mid B) = \frac{3}{5}$

Часто, когда мы работаем с условными вероятностями, удобно описать модель с помощью дерева, в котором каждая вершина соответствует какой-то цепочке событий. Рассмотрим пример. В HoMM3 у вашего героя экспертный навык удачи. Ваш титан атакует дендройда.

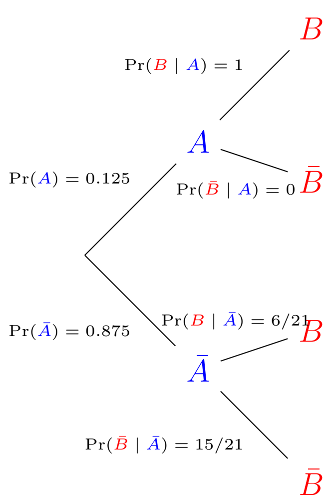
- Урон титана 40-60 (выбирается равновероятно)

- Здоровье дендройда 55
- Пусть навыки атаки и защиты одинаковы
- Удача срабатывает с вероятностью 0.125 и удваивает урон

Событие A: сработала удача, событие B: дендройд пал. Давайте найдем условную вероятность $\Pr[A \mid B]$ по формуле

$$\Pr[B \mid A] = \frac{\Pr[A \cap B]}{\Pr[B]}.$$

Для начала нарисуем такое дерево событий:



Чтобы найти $\Pr[A \cap B]$, мы можем переписать формулу условной вероятности:

$$\Pr[A \cap B] = \Pr[A] \Pr[B \mid A] = 0.125 \cdot 1 = 0.125,$$

так как все вероятности справа нам известны. Теперь найдем вероятность $\Pr[B]$ с помощью такого трюка. Заметим, что события $A \cap B$ и $\bar{A} \cap B$ не пересекаются, а их объединение равно B. Поэтому по свойствам меры можем написать:

$$\begin{aligned}\Pr[B] &= \Pr[A \cap B] + \Pr[\bar{A} \cap B] = \Pr[A] \Pr[B \mid A] + \Pr[\bar{A}] \Pr[B \mid \bar{A}] \\ &= 0.125 \cdot 1 + 0.875 \cdot \frac{6}{21} = 0.375.\end{aligned}$$

Наконец, подставляем это все в формулу $\Pr[A \mid B]$, и получаем

$$\Pr[A \mid B] = \frac{0.125}{0.375} = \frac{1}{3}.$$

2.1 Формула умножения вероятностей

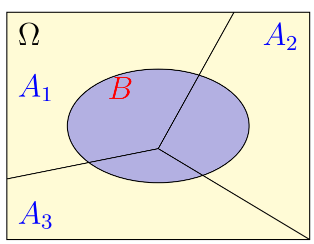
Мы только что пользовались трюком, где показали, что $\Pr[A \cap B] = \Pr[A] \Pr[B \mid A]$. Его можно обобщить:

$$\Pr\left(\bigcap_{i=1}^{n} A_i\right) = \Pr(A_1) \prod_{i=2}^{n} \Pr(A_i \mid A_1 \cap \dots \cap A_{i-1})$$

Доказать легко по индукции. База уже есть, докажем переход:

$$\begin{aligned}
\Pr\left(\bigcap_{i=1}^{n} A_i\right) &= \Pr\left(A_n \cap \left(\bigcap_{i=1}^{n-1} A_i\right)\right) \\
&= \Pr\left(\bigcap_{i=1}^{n-1} A_i\right) \Pr\left(A_n \mid \left(\bigcap_{i=1}^{n-1} A_i\right)\right) \\
&= \Pr(A_1) \prod_{i=2}^{n} \Pr(A_i \mid A_1 \cap \dots \cap A_{i-1})
\end{aligned}$$

2.2 Формула полной вероятности



Пусть у нас есть разбиение Ω на события $A_1, A_2, A_3, \dots$. И есть событие B. И пусть мы знаем все вероятности $\Pr[A_i]$ и все условные вероятности $\Pr[B \mid A_i]$. Тогда верно, что

$$\Pr(B) = \sum_i \Pr(A_i) \Pr(B \mid A_i)$$

Докажем это. Заметим, что $B = \bigcup_i (B \cap A_i)$ — объединение непересекающихся множеств. При этом $\Pr(B \cap A_i) = \Pr(A_i) \Pr(B \mid A_i)$ — из правила умножения. Поэтому,

$$\Pr[B] = \sum_i \Pr[B \cap A_i] = \sum_i \Pr[A_i] \Pr[B \mid A_i].$$

NB: Это верно как для конечного, так и для счетного разбиения Ω

2.3 Формула Байеса

Она работает в тех же условиях, что и формула полной вероятности, то есть у нас есть разбиение Ω на события $A_1, A_2, A_3, \dots$. И есть событие B. И пусть мы знаем все вероятности $\Pr[A_i]$ и все условные вероятности $\Pr[B \mid A_i]$. Тогда для всех i верно, что

X
$$\Pr(A_i \mid B) = \frac{\Pr(A_i)\Pr(B \mid A_i)}{\sum_j \Pr(A_j)\Pr(B \mid A_j)}$$

Доказательство:

$$Pr[A_i \mid B] = \frac{\Pr[A_i \cap B]}{\Pr[B]} = \frac{\Pr[A_i]\Pr[B \mid A_i]}{\sum_j \Pr[A_j]\Pr[B \mid A_j]}.$$

3 Независимость событий

Рассмотрим эксперимент, в котором мы два раза бросаем нечестную монетку.

- $\Pr(P) = p \neq 0.5$
- $\Pr(O) = (1-p)$

Мы получаем один из четырех исходов $\{PP, PO, OP, OO\}$, причем их вероятности

- $\Pr(PP) = \Pr(P*)\Pr(*P \mid P*) = p \cdot p = p^2$
- $\Pr(PO) = \Pr(OP) = p(1-p)$
- $\Pr(OO) = (1-p)^2$

Теперь ответим на несколько вопросов. Какова вероятность, что второй бросок упадет решкой? Ответ:

$$\Pr(*P) = p$$

Какова вероятность, что второй бросок упадет решкой, если первый упал орлом? Ответ:

$$\Pr(*P \mid O*) = p,$$

то есть условие *не влияет* на вероятность. Но какова вероятность, что второй бросок упадет решкой, если есть ровно одна решка? Ответ:

$$\Pr(*P \mid PO \cup OP) = \frac{\Pr(OP)}{\Pr(PO \cup OP)} = \frac{p(1-p)}{2p(1-p)} = \frac{1}{2},$$

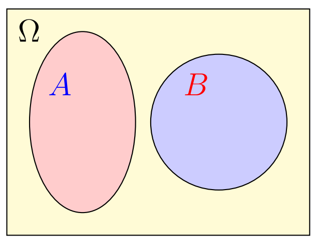
условие *влияет* на вероятность. Мы приходим к следующему интуитивному определению независимости событий. События A и B независимы, если $\Pr(A \mid B) = \Pr(A)$. Другими словами, B не несет никакой информации о событии A. Но это определение не работает, если $\Pr[B] = 0$. Поэтому мы называем события независимыми, если

$$\Pr(A \cap B) = \Pr(A)\Pr(B)$$

Чем оно лучше:

- Симметрично относительно событий A и B
- Из него следует и $\Pr(A \mid B) = \Pr(A)$, и $\Pr(B \mid A) = \Pr(B)$ (если условные вероятности определены)
- Корректно и при $\Pr(A) = 0$, и при $\Pr(B) = 0$

Рассмотрим типичную ошибку в понимании независимости. Являются ли независимыми непересекающиеся события?



Ответ нет. Если произошло событие A, то вероятность B равна нулю.

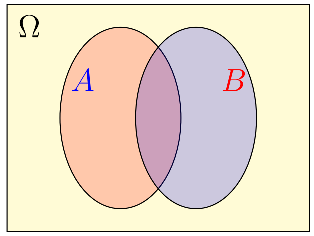
Если A и B независимы, то независимы и A и $\bar{B}$. Интуитивное доказательство: если “событие B произошло” не дает никакой инфы про A, то и “событие B не произошло” не должно ее давать. Формально:

$$\begin{aligned} P(A) &= \Pr(A \cap B) + \Pr(A \cap \bar{B}) \\ &= \Pr(A)\Pr(B) + \Pr(A \cap \bar{B}) \\ \Rightarrow \Pr(A \cap \bar{B}) &= P(A) - \Pr(A)\Pr(B) = \Pr(A)(1 - \Pr(B)) \\ &= \Pr(A)\Pr(\bar{B}) \end{aligned}$$

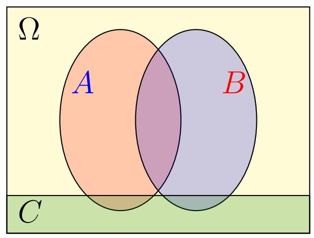
Также события могут быть зависимыми или независимыми при каком-то условии, так как условная вероятностная мера – тоже вероятностная мера. События A и B независимы при условии C, если

$$\Pr(A \cap B \mid C) = \Pr(A \mid C)\Pr(B \mid C)$$

Независимость и условная независимость не особо связаны друг с другом. Например, пусть A и B — независимы:



Однако если ввести условие C, как на картинке ниже, эти события не будут независимыми (что легко видно из того, что они не пересекаются).



Также нам иногда хочется ввести понятие независимости множества событий. Формально, множество событий должно быть независимо, если информация о том, что произошло какое-то подмножество событий не влияет на вероятность происхождения событий не из этого подмножества:

$$\Pr(A_1 \cap \bar{A}_4) = \Pr(A_1 \cap \bar{A}_4 \mid A_2 \cap (A_3 \cup \bar{A}_5))$$

разные индексы слева и справа от “|”

Поэтому вводится такое вот формальное определение. События $\{A_1, \ldots, A_n\}$ *независимы* (по совокупности), если для любого набора индексов $I \subset [1..n]$ верно

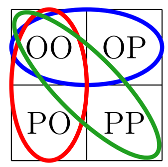
$$\boxed{\Pr\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \Pr(A_i)}$$

для набора из трех событий достаточно показать, что

- $\Pr(A_1 \cap A_2) = \Pr(A_1)\Pr(A_2)$
- $\Pr(A_1 \cap A_3) = \Pr(A_1)\Pr(A_3)$
- $\Pr(A_2 \cap A_3) = \Pr(A_2)\Pr(A_3)$
- $\Pr(A_1 \cap A_2 \cap A_3) = \Pr(A_1)\Pr(A_2)\Pr(A_3)$

Причем первые три пункта называются *попарной* независимостью. Причем попарная независимость событий слабее, чем независимость по совокупности (то есть из независимости по совокупности следует попарная, а обратное в общем случае неверно). Для иллюстрации расмотрим пример с подбрасыванием двух монет и следующие события:

- A — первая монета орлом
- B — вторая монета орлом
- C — обе монеты одинаковы



Вероятности событий:

- $\Pr(A) = \frac{1}{2}$
- $\Pr(B) = \frac{1}{2}$
- $\Pr(C) = \frac{1}{2}$

Вероятности комбинаций:

- $\Pr(A \cap B) = \frac{1}{4} = \Pr(A)\Pr(B)$
- $\Pr(A \cap C) = \frac{1}{4} = \Pr(A)\Pr(C)$
- $\Pr(B \cap C) = \frac{1}{4} = \Pr(B)\Pr(C)$
- $\Pr(A \cap B \cap C) = \frac{1}{4} \neq \Pr(A)\Pr(B)\Pr(C)$

То есть в данном случае события A, B и C попарно независимы, но не независимы по совокупности.

4 Случайные величины

Случайная величина (с.в.) — *функционал*, заданный на Ω.

$$X : \Omega \to \mathbb{R}$$

Единственное условие: этот функционал должен быть *измерим* по используемой мере Pr.

Договоримся об обозначениях:

- Большая буква (X, Y, Z, etc) — случайная величина (как *отображение*).
- Маленькая буква (x, y, z, etc) — значение случайной величины (*число*).

Примеры с.в.:

- Бросок кости: X — число на верхней грани
  - $\{1 \to 1,\, 2 \to 2,\, 3 \to 3,\, 4 \to 4,\, 5 \to 5,\, 6 \to 6\}$
- Хоккейный матч: X — очки домашней команды
  - $\{\text{В} \to 2,\, \text{ВО} \to 2,\, \text{ВБ} \to 2,\, \text{ПБ} \to 1,\, \text{ПО} \to 1,\, \text{П} \to 0\}$

На счетной Ω:

- Бросаем монету до первого орла: X — число бросков

На несчетной Ω:

- Бросаем дротик в мишень для дартса: X — очки согласно правилам игры
- Крутим волчок в ЧГК: X — угол, на котором он остановится
  - Y — население города, из которого пришел вопрос
  - Z — вероятность того, что на выбранный вопрос знатоки ответят

Дискретными называются те с.в., которые имеют не более, чем счетное число значений. Для них можно ввести понятие *функции вероятностей*. Можно рассматривать $(X = x)$ как событие, так как оно задает какое-то множество исходов из Ω. Тогда у этого события должна быть вероятность.

$$\boxed{p_X(x) = \Pr(X = x) = \Pr(\{\omega \in \Omega \mid X(\omega) = x\})}$$

p_X называется *функцией вероятности* с.в. X. Ее свойства:

- $p_X(x) \geq 0$
- $\sum_x p_X(x) = 1$

5 Известные распределения

Распределением мы будем называть любую вероятностную меру на $\mathbb{R}$. Мы говорим, что дискретная с.в. X следует распределнию $\mathcal{D}$ (при этом пишем $X \sim \mathcal{D}$), если вероятностная мера Pr, соответствующая $\mathcal{D}$ такова, что для любого $x \in \mathbb{R}$ верно, что

$$p_X(x) = \Pr[x].$$

То есть по сути для дискретных с.в. распределение просто определяет функцию вероятностей. Примеры известных распределений:

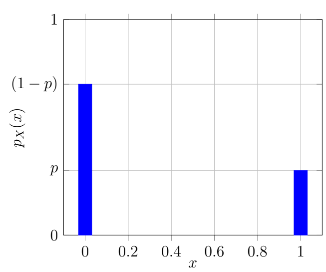
- Бернулли
- Равномерное
- Биномиальное
- Геометрическое
- Гипер-геометрическое
- Степенное

Распределение Бернулли с параметром p определяет случайную величину, которая равна единице в случае успеха какого-то случайного эксперимента и равна нулю иначе, если вероятность успеха эксперимента равна p. Формально,

$$X \sim \mathrm{Bern}(p) \Leftrightarrow \begin{cases} p_X(0) = (1-p), \\ p_X(1) = p, \\ p_X(x) = 0 \text{ для всех остальных } x. \end{cases}$$

Иногда успехом мы можем назвать то, что в результате какого-то эксперимента произошло событие A. Тогда такая с.в. будет называться *индикатором события* A.

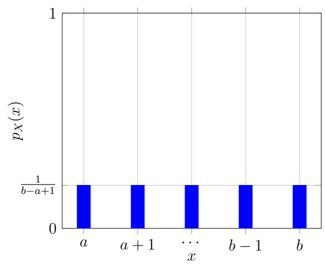
Для описания дискретных распределений часто используют такие графики, где на всех возможных значениях с.в. нарисованы столбики высотой, равной вероятности этого значени. Для с.в. Бернулли этот график выглядит так:



Равномерное дискретное распределение это распределение на целочисленном интервале $[a..b]$, где у каждого значения одинаковая функция вероятности.

$$X \sim U(a,b) = \begin{cases} p_X(x) = \frac{1}{b-a+1}, & \text{если } x \in [a..b], \\ p_X(x) = 0, & \text{иначе.} \end{cases}$$

Заметим, что a, b — целочисленные параметры ($a \leq b$), $\Omega = [a..b]$ и для любого исхода ω верно, что $X(\omega) = \omega$. График функции вероятностей выглядит так:



Биномиальное распределение с параметрами n и p есть случайная величина, равная числу успехов в n независимых опытах по схеме Бернулли с параметром p. Другими словами,

$$X \sim \mathrm{Bin}(n, p) \Leftrightarrow X = \sum_{i=1}^{n} X_i,$$

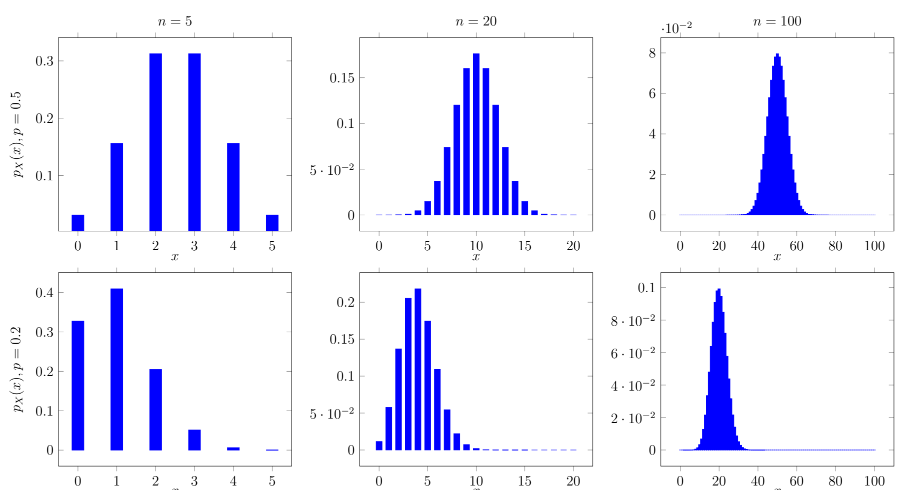
где все $X_i \sim \mathrm{Bern}(p)$. Посчитаем функцию вероятностей в каком-то x (понятно, что x может быть только целым числом из $[0..n]$). Вероятность конкретного исхода, в котором ровно x успехов:

$$\Pr(\omega) = p^x(1-p)^{n-x}.$$

Всего таких (непересекающихся) исходов $\binom{n}{x}$. Поэтому,

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Вот примеры графиков функции вероятностей биномиалки в зависимости от ее параметров n и p:



NB: везде разные масштабы оси OY

Геометрическое распределение – номер первого успешного опыта, когда мы последовательно проводим бесконечно много независимых опытов по схеме Бернулли с вероятностью успеха p.

$$X \sim Geom(p) \Leftrightarrow X = \min\{i \in \mathbb{N} \mid X_i = 1\},$$

где $X_i \sim \mathrm{Bern}(p)$. Посчитаем функцию вероятностей в каком-то $x \in \mathbb{N}$. Вероятность, что первые $x-1$ неуспешны:

$$\Pr(X_1 = X_2 = \dots X_{x-1} = 0) = (1-p)^{x-1}$$

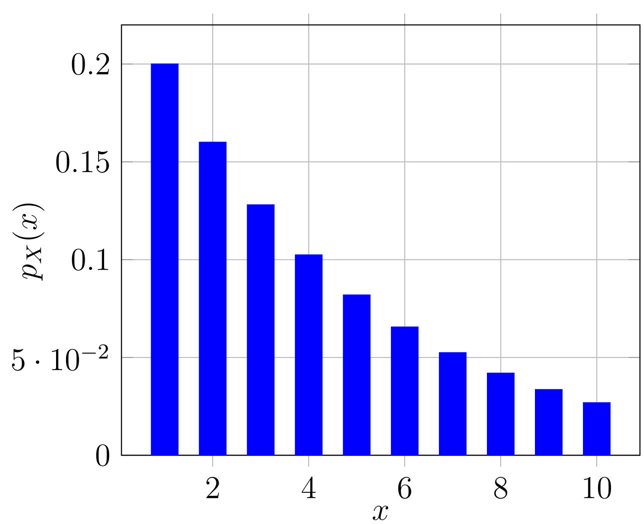
Вероятность, что x-ый исход успешен:

$$\Pr(X_x = 1) = p$$

Перемножая вероятности этих независимых событий, получаем:

$$p_X(x) = p(1-p)^{x-1}$$

И по традиции, вот график функции вероятностей (неполный, так как число возможных исходов бесконечно):



Есть исход, при котором $X = +\infty$, но его вероятность равна нулю.

$$\Pr(X = \infty) \le \Pr(X_1 = \dots = X_n = 0) = (1-p)^n$$

для любого n. Если допустить, что $\Pr(X = \infty) = q > 0$, то возьмем $n \ge \log_{1-p} q$, получим, что вероятность подсобытия больше, чем вероятность события (*Противоречие* со свойствами вероятностной меры).

6 Матожидание

Если много раз повторить один и тот же эксперимент, то каков будет средний результат? Например, если мы бросаем кость d6, то каждая грань выпадет примерно $n/6$ раз, где n — число бросков (напомним, что именно так мы и воспринимаем вероятность в реальном мире).

Тогда среднее значение будет равно

$$\frac{1}{n}\left(\frac{n}{6}\cdot 1+\frac{n}{6}\cdot 2+\frac{n}{6}\cdot 3+\frac{n}{6}\cdot 4+\frac{n}{6}\cdot 5+\frac{n}{6}\cdot 6\right)=3.5$$

Поэтому мы вводим понятие матожидания дискретной случайной величины как среднего (ожидаемого) ее значения, которое считается так:

$$\boxed{E(X)=\sum_x x\cdot p_X(x)}$$

NB: Матожидание определено только когда этот ряд *абсолютно сходится*
NB-2: Матожидание — *функционал* на множестве с.в.
По определению легко посчитать матожидание некоторых с.в.:
$X \sim \mathrm{Bern}(p)$

$$E(X)=0\cdot(1-p)+1\cdot p=p$$

$X \sim U(a,b)$

$$E(X)=\sum_{i=a}^{b}\frac{i}{b-a+1}=\frac{a+b}{2}$$

Элементарные свойства матожидания:

- *Неотрицательность:* $X \geq 0 \Rightarrow E(X) \geq 0$
- *Ограниченность:* $X \in [a,b] \Rightarrow E(X) \in [a,b]$
- *Матожидание константы:* $E(c)=c$

Заметим, что константой мы называем случайную величину X, для которой есть какая-то c такая, что $p_X(c)=1$. Но в наших выкладках мы просто пишем c.

Функции от с.в. и их матожидание. Пусть есть с.в. X и функция $g:\mathbb{R}\to\mathbb{R}$. Тогда $Y=g(X)$ — тоже с.в. Как посчитать $E[Y]$? Можно по определению: $E(Y)=\sum_y yp_Y(y)$, но для этого надо сначала определить множество занчений Y. Но можно проще:

$$\boxed{E(Y) = \sum_x g(x)p_X(x)}$$

Доказательство:

$$\begin{aligned}\sum_x g(x)p_X(x) &= \sum_y \sum_{x:g(x)=y} g(x)p_X(x) \\ &= \sum_y y \sum_{x:g(x)=y} p_X(x) = \sum_y y p_Y(y) = E(Y)\end{aligned}$$

Эта формула позволяет нам вывести важное свойство матожидания — *линейность* (хотя пока что только в усеченном формате).

$$\boxed{E(aX + b) = aE(X) + b}$$

Благодаря этому мы часто можем быстро считать матожидания с.в., являющихся линейными преобразованиями известных нам распределений. Например, X принимает четные значения от 0 до 100 с равной вероятностью. То есть $X = 2Y$, где $Y \sim U(0, 50)$. Поэтому $E[X] = 2E[Y] = 50$.

NB: Для всех линейных функций g выполнено $E(g(X)) = g(E(X))$, но в общем случае это неверно

Мы еще вернемся к линейности матожидания и дополним это свойство чуть позже.

7 Дисперсия

Иногда хотим оценить то, насколько с.в. *отклоняется* от своего среднего значения μ. Первая мысль: посчитать $E(X - \mu)$, но

$$E(X - \mu) = E(X) - \mu = 0.$$

Поэтому величина отклонения с.в. от своего матожидания (иногда говорят, ее концентрация или разброс) оценивется через так называемую *дисперсию* (или в некоторых источниках — вариацию):

$$\boxed{\mathrm{Var}(X) = E((X - \mu)^2)}$$

Неокторые замечания по поводу этого определения:

- Дисперсия считается как матожидание *функции от с.в.*
- Всегда *неотрицательна*, равна нулю только у *детерминированной* с.в.

- Почему не посчитать $|X - \mu|$? Просто так удобнее. Чем — узнаем позже
- Часто полезно знать *среднеквадратичное отклонение*

$$\boxed{\sigma(X) = \sqrt{\operatorname{Var}(X)}}$$

- В дальнейшем часто будем обозначать $\mu \coloneqq E(X)$, если понятно, про какую с.в. речь

У дисперсии нет линейности, как у матожидания (что неудивительно, ведь это матожидание квадратичной функции от с.в.), но есть такое полезное свойство:

$$\boxed{\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X)}$$

- Умножая с.в. на a — увеличиваем ее вариацию в a^2 раз
- Прибавляя к с.в. b — не меняем вариацию

Доказательство:

$$\begin{aligned}
\operatorname{Var}(aX + b) &= E\left((aX + b - E(aX + b))^2\right) \\
&= E\left((aX + b - aE(X) - b)^2\right) \\
&= E\left((aX - aE(X))^2\right) \\
&= E\left(a^2(X - E(X))^2\right) \\
&= a^2 E\left((X - E(X))^2\right) \\
&= a^2 \operatorname{Var}(X)
\end{aligned}$$

Дисперсию часто удобно считать по вот такой полезной формуле:

$$\boxed{\operatorname{Var}(X) = E(X^2) - (E(X))^2}$$

Ее доказательство:

$$\begin{aligned}
\operatorname{Var}(X) &= E\left((X - \mu)^2\right) \\
&= E\left(X^2 - 2X\mu + \mu^2\right) \\
&= E(X^2) - 2\mu^2 + \mu^2 \\
&= E(X^2) - (E(X))^2
\end{aligned}$$