

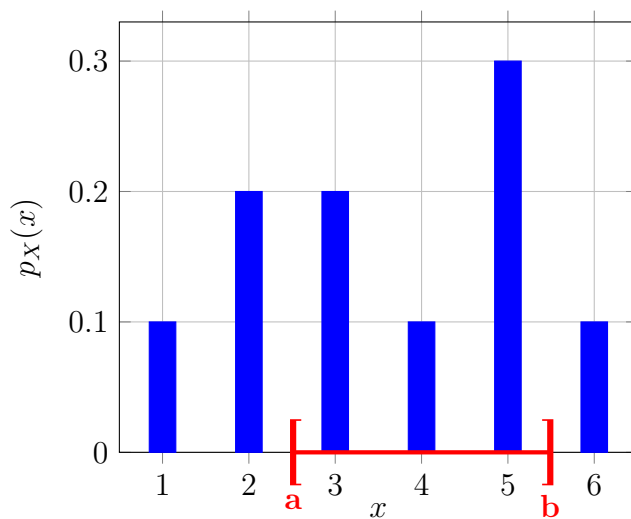
Лекция 3. Непрерывные с.в., часть 1

1 апреля 2022 г.

1 Непрерывные случайные величины

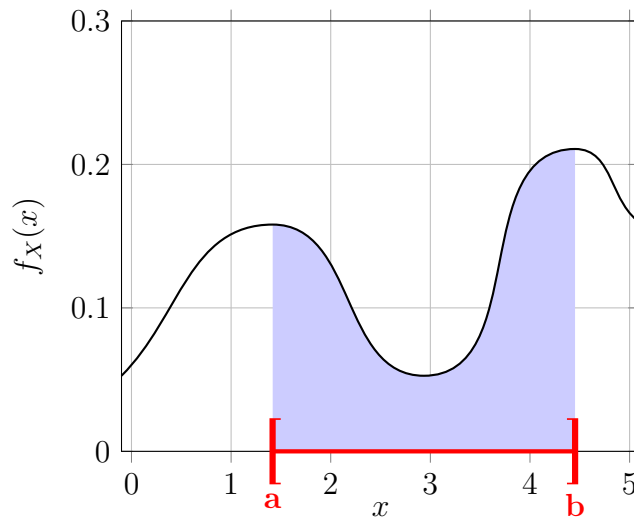
Определение непрерывных с.в.

В случае с дискретными величинами:



$$\Pr[a \leq X \leq b] = \sum_{x:a \leq x \leq b} p_X(x)$$

Но если X может принимать любые вещественные значения из этого интервала? Тогда нам нужна функция, которая показывает, сколько вероятностной массы лежит на каждом элементарном отрезке.



$$\Pr[a \leq X \leq b] = \int_a^b f_X(x) dx \quad (1)$$

Определение: Случайная величина называется непрерывной, если для нее существует такая функция $f_X(x)$, что для любых $a, b \in \mathbb{R}$ (где $a \leq b$) верно (??).

$f_X(x)$ — плотность вероятности с.в. X :

$$\Pr(a \leq X \leq a + \varepsilon) \approx f_X(a) \cdot \varepsilon$$

Плотность вероятности — аналог функции вероятностей для непрерывных с.в.:

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

То же самое

- $f_X(x) \geq 0$
- $\int_{-\infty}^{+\infty} p_X(x) = 1$

NB:

$$\Pr(X = a) = \Pr(a \leq X \leq a) = \int_a^a f_X(x) dx = 0$$

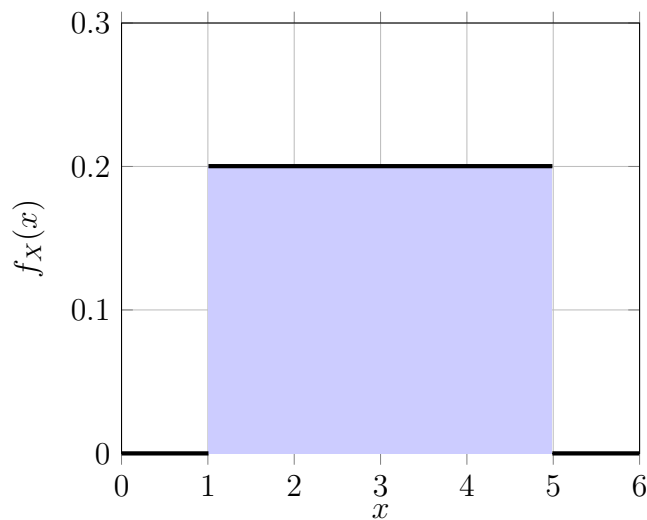
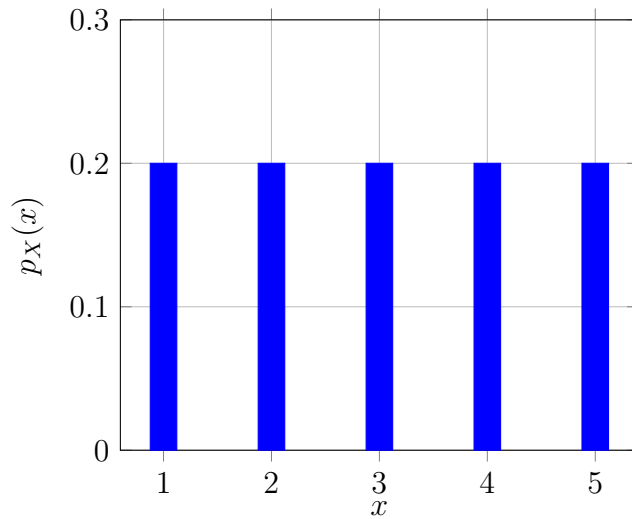
Поэтому:

$$\Pr(a \leq X \leq b) = \Pr(X = a) + \Pr(X = b) + \Pr(a < X < b) = \Pr(a < X < b)$$

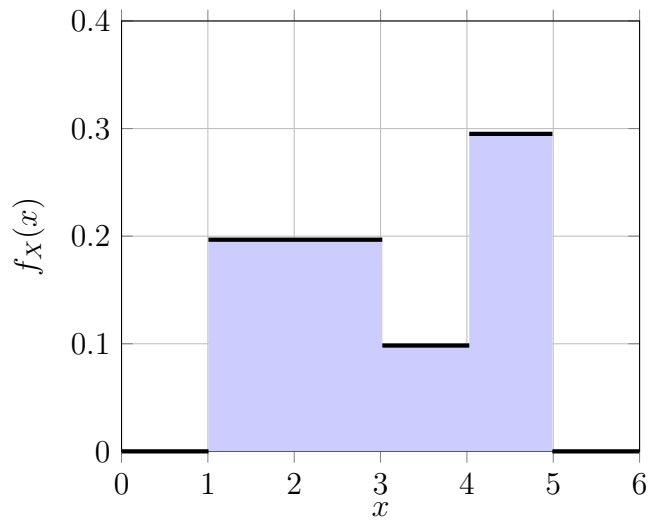
NB: Мы ушли от понятия событий, но у нас по-прежнему есть какая-то Ω , на которой и задана с.в. X . Просто нам сейчас проще быть чисто в терминах с.в.

NB: Переопределим дискретные с.в. как с.в., для которых есть функция вероятностей, то есть число возможных значений которых счетно.

NB: Вы уже могли догадаться, что с.в. могут быть и смешанные, но про это позже.
Пример: равномерное распределение



Обобщение: частично равномерное распределение



2 Матожидание

Для дискретных величин:

$$E(X) = \sum_x x p_X(x)$$

Для непрерывных: заменяем сумму на интеграл, а функцию вероятностей на плотность вероятности

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx$$

Важно: интеграл должен сходиться абсолютно

Интерпретация: центр масс вероятностной массы

Свойства матожидания

- $X \geq 0 \Rightarrow E[X] \geq 0$
- $X \in [a, b] \Rightarrow E[X] \in [a, b]$
- Матожидание функции от с.в.:

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$$

- Пример

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x) dx$$

- Линейность: $E(aX + b) = aE(X) + b$

3 Дисперсия

Как и для дискретных:

$$\text{Var}(X) = E((X - \mu)^2) = \int_{-\infty}^{+\infty} (x - \mu)^2 f_X(x) dx$$

среднеквадратичное отклонение:

$$\sigma_X = \sqrt{\text{Var}(X)}$$

Свойства — те же:

- $\text{Var}(aX + b) = a^2 \text{Var}(X)$
- $\text{Var}(X) = E(X^2) - (E(X))^2$

4 Моменты стандартных распределений

NB: i -й момент распределения — $\int_{-\infty}^{+\infty} x^i f_X(x) dx$ (для дискретных с.в — сумма)

Равномерное распределение

$X \sim U(a, b)$

Матожидание:

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}.$$

Дисперсия:

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{+\infty} x^2 f_X(x) dx = \int_a^b x^2 \frac{1}{b-a} dx \\ &= \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3} \end{aligned}$$

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{4b^2 + 4ab + 4a^2 - 3b^2 - 6ab - 3a^2}{12} = \frac{(b-a)^2}{12}.$$

Экспоненциальное распределение

Говорим, что X следует экспоненциальному распределению $\text{Exp}(\lambda)$ с параметром λ , если

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

NB: Это аналог геометрического распределения с параметром $p = \lambda$.

Матожидание:

$$\begin{aligned} E[X] &= \int_0^{+\infty} x \lambda e^{-\lambda x} dx = - \int_0^{+\infty} x de^{-\lambda x} = -xe^{-\lambda x} \Big|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx \\ &= 0 - \frac{1}{\lambda} \Big|_0^{+\infty} = \frac{1}{\lambda}. \end{aligned}$$

Дисперсия (два раза интегрируем по частям):

$$\begin{aligned} E[X^2] &= \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx = \frac{2}{\lambda}. \\ \text{Var}(X) &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}. \end{aligned}$$

Довольно хорошо сконцентрирована, так как вероятность хвоста экспоненциально падает:

$$\Pr(X \geq a) = \int_a^{+\infty} \lambda e^{-\lambda x} dx = e^{-\lambda a}.$$

5 Функция распределения

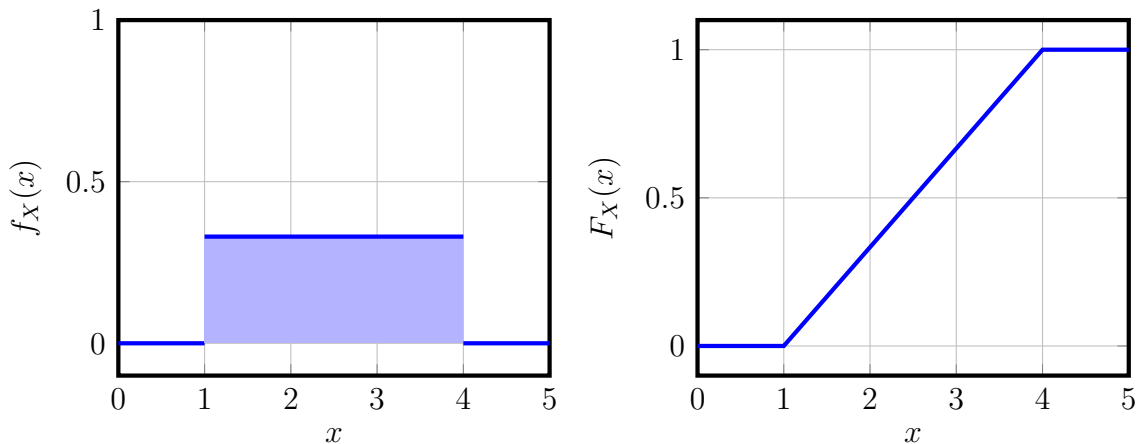
$F_X(x) = \Pr(X \leq x)$ — функция распределения с.в. X (как дискретной, так и непрерывной).

Как считать:

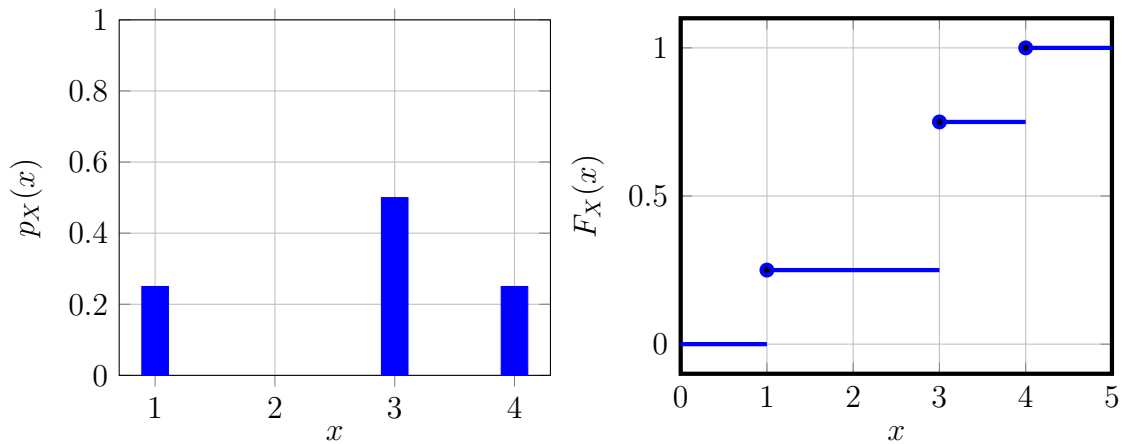
$$F_X(x) = \int_{-\infty}^x f_X(t) dt$$

Легко заметить: $F'_X(x) = f_X(x)$

Пример: равномерное распределение.



Функцию распределения можно считать и для дискретной случайной величины:



Свойства функции распределения:

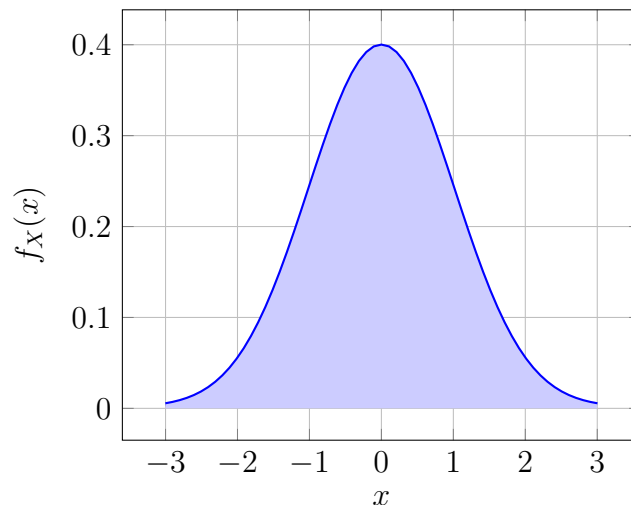
- Неубывающая
- $\lim_{x \rightarrow +\infty} F_X(x) = 1$
- $\lim_{x \rightarrow -\infty} F_X(x) = 0$

6 Нормальное распределение (распределение Гаусса)

Очень важная штука:

- Важна в центральной предельной теореме
- Часто на практике неизвестные распределения приближаются нормальным

Стандартное нормальное: $X \sim N(0, 1) \leftrightarrow f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$



Откуда берется коэффициент нормализации $\frac{1}{\sqrt{2\pi}}$? Из интеграла Гаусса

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$$

Свойства $N(0, 1)$:

$$E[X] = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0,$$

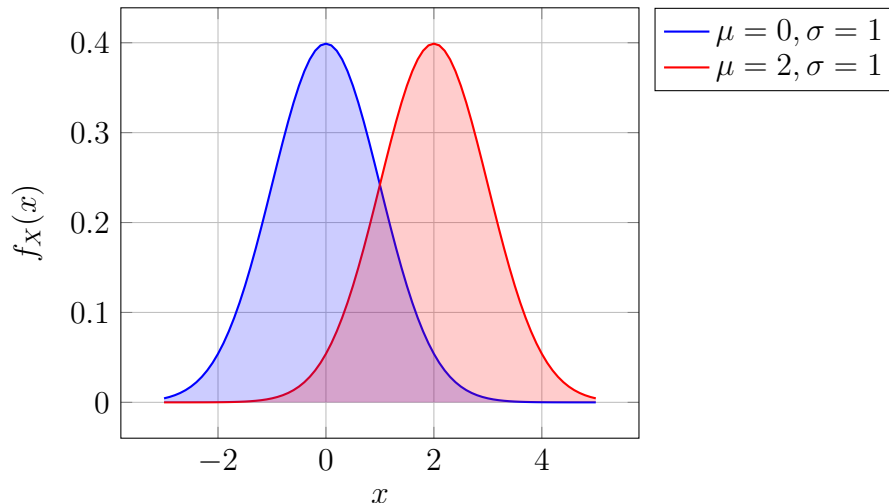
так как это интеграл нечетной функции, которая в бесконечности очень маленькая.

$$\begin{aligned} \text{Var}(X) = E[X^2] &= \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} d\frac{x^2}{2} \\ &= -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x de^{-x^2/2} \\ &= -\frac{1}{\sqrt{2\pi}} x e^{-x^2/2} \Big|_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} dx \\ &= 0 + \frac{\sqrt{2\pi}}{\sqrt{2\pi}} = 1. \end{aligned}$$

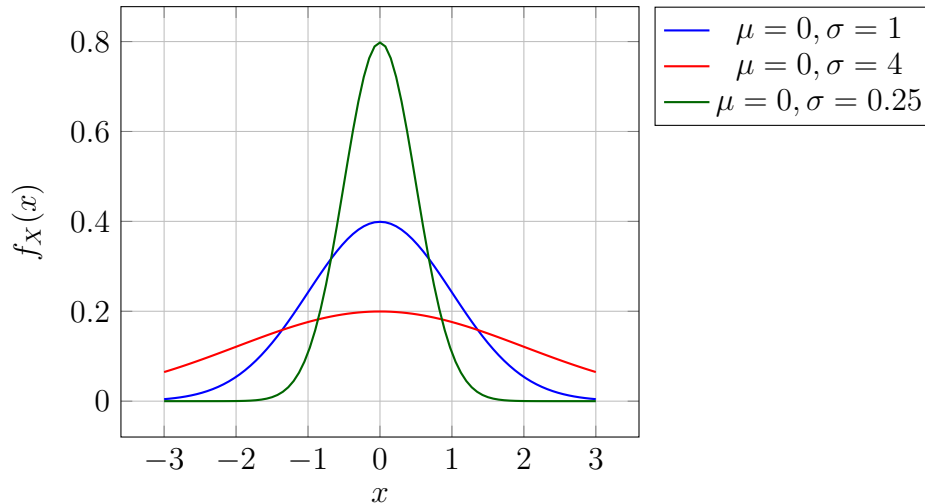
В записи $N(0, 1)$ нулик как раз обозначает матожидание, а единица — дисперсию

Обобщенное нормальное распределение: $X \sim N(\mu, \sigma^2) \Leftrightarrow f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

- μ — новое матожидание, насколько мы сдвигаем ось симметрии X .
- σ — новая дисперсия, насколько мы растягиваем распределение от оси симметрии.



Чем меньше σ (среднеквадратичное отклонение), тем больше распределение сжато вокруг оси симметрии



Полезное свойство нормальной случайной величины: если $X \sim N(\mu, \sigma^2)$ и при этом $Y = aX + b$, то Y тоже имеет нормальное распределение. Это мы докажем позже, но пока давайте посмотрим, какому именно распределению.

- $E(Y)$ должно равняться $E(aX + b) = aE(X) + b = a\mu + b$
- $\text{Var}(Y)$ должна равняться $\text{Var}(aX + b) = a^2 \text{Var}(X) = a^2\sigma^2$

Значит, $Y \sim N(a\mu + b, a^2\sigma^2)$.

Функция распределения $X \sim N(0, 1)$ обозначается буквой Φ :

$$\Phi(x) = F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Так как интеграл неберущийся, ее значения считаются с помощью таблиц.

Как работать с такими таблицами. Пусть мы хотим посчитать вероятность того, что $X \sim N(0, 1)$ не больше, чем 2.39. Находим строчку 2.3, столбец 0.09, смотрим на число в их пересечении, добавляем к нему 0.5. Если хотим посчитать $\Pr(X \leq -2.39)$, то вычитаем это число из 0.5.

Как пользоваться таблицей, если $X \sim N(\mu, \sigma^2)$? И пусть мы хотим найти вероятность $\Pr(X \in [a, b])$. Для этого Давайте рассмотрим $Y = \frac{X-\mu}{\sigma}$. Заметим, что $Y \sim N(0, 1)$. Теперь запишем интересное нас событие следующим образом

Таблица 1: Таблица для вычисления функции распределения стандартного нормального распределения $N(0, 1)$. В i -й строке и j -ом столбце число равно $\int_0^{0.1i+0.01j} f_X(x)dx$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

$$\begin{array}{ll}
a \leq X & \leq b \\
a - \mu \leq X - \mu & \leq b - \mu \\
\frac{a - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} & \leq \frac{b - \mu}{\sigma}
\end{array}$$

То есть искомая вероятность равна $\Pr(Y \in [\frac{a-\mu}{\sigma}, \frac{b-\mu}{\sigma}])$, а это уже вычисляется по таблице.

7 Условная плотность вероятности

Напомним трактовку плотность вероятности. Это то, сколько вероятностной массы приходится на маленький интервал значений с.в.:

$$f_X(x)\delta \approx \Pr(X \in [x, x + \delta])$$

Но вероятностная мера может меняться при условии, что произошло событие A . В этом случае определяем

$$f_{X|A}(x)\delta \approx \Pr(X \in [x, x + \delta] \mid A)$$

Точное определение: плотность вероятности с.в. X при условии A есть такая функция $f_{X|A}(x)$, что для любого измеримого множества B верно, что

$$\Pr(X \in B \mid A) = \int_B f_{X|A}(x)dx.$$

NB: у нас опять просто поменялась вероятностная мера. То есть у условной плотности вероятности будут все те же свойства:

- $f_{X|A}(x) \geq 0$
- $\int_{-\infty}^{+\infty} f_{X|A}(x)dx = 1$

Вычисляем аналогично условной функции вероятности. Пусть $x \in A$, тогда

$$\begin{aligned}
f_{X|A}(x)\delta \approx \Pr(X \in [x, x + \delta] \mid A) &= \frac{\Pr(X \in [x, x + \delta] \cap X \in A)}{\Pr(A)} \\
&= \frac{\Pr(X \in [x, x + \delta])}{\Pr(A)} \approx \frac{f_X(x)\delta}{\Pr(A)}
\end{aligned}$$

Поэтому строго говоря, она вычисляется так:

$$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{\Pr(A)}, & \text{если } x \in A, \\ 0, & \text{иначе.} \end{cases}$$

То есть масштабируем плотность вероятности по событию A .

Иногда придется иметь дело с условной функцией распределения:

$$F_{X|A}(x) = \Pr(X \leq x | A)$$

8 Условное матожидание

Условное матожидание определяем аналогично:

$$E(X | A) = \int_{-\infty}^{+\infty} x f_{X|A}(x) dx$$

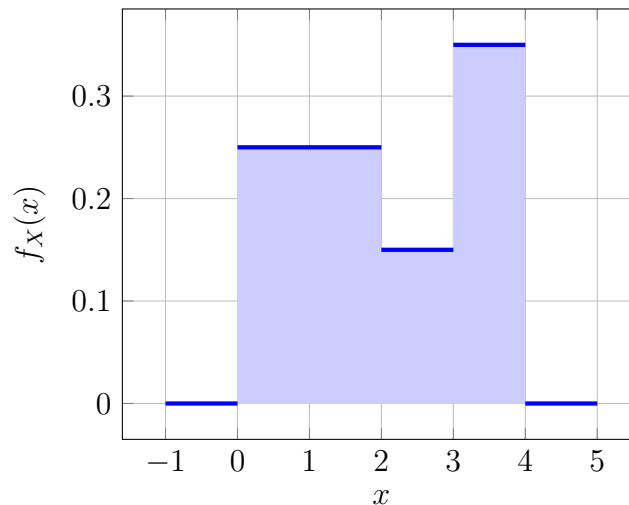
Работает то же самое правило и для функций от с.в. (у нас просто новая плотность вероятности при условии A):

$$E(g(X) | A) = \int_{-\infty}^{+\infty} g(x) f_{X|A}(x) dx$$

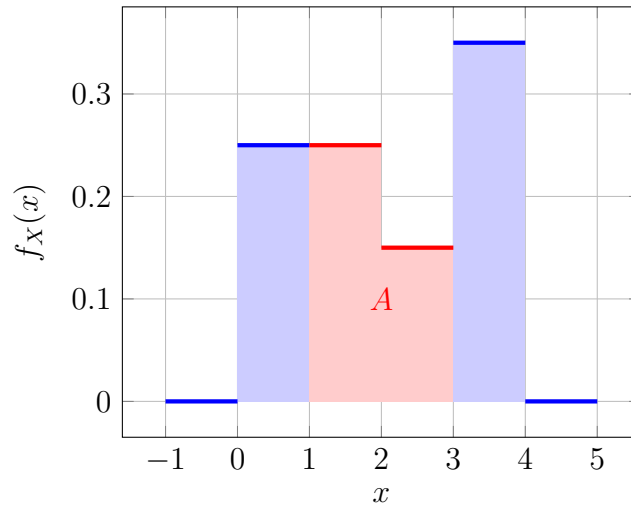
9 Пример условных с.в.

Рассмотрим частично равномерное распределение:

$$f_X(x) = \begin{cases} 0.25, & x \in [0, 2) \\ 0.15, & x \in [2, 3) \\ 0.35, & x \in [3, 4] \\ 0, & \text{иначе.} \end{cases}$$

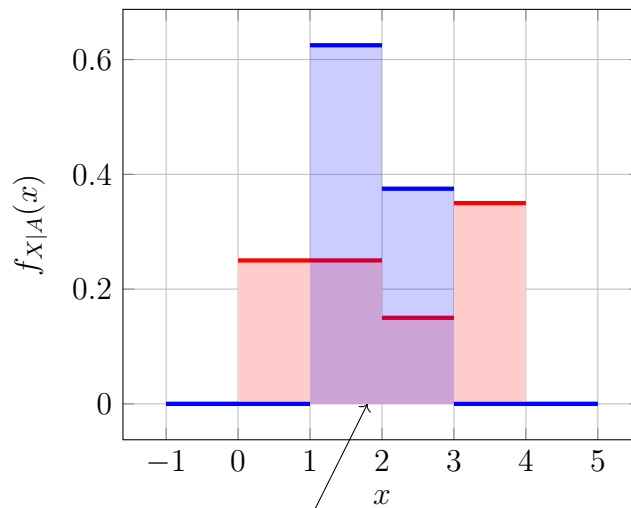


Пусть $A = [1, 3]$. Тогда $\Pr(A) = 0.25 \cdot 1 + 0.15 \cdot 1 = 0.4$



Значит, новая плотность вероятности выглядит так:

$$f_{X|A}(x) = \begin{cases} \frac{0.25}{0.4} = 0.625, & x \in [1, 2) \\ \frac{0.15}{0.4} = 0.375, & x \in [2, 3) \\ 0, & \text{иначе.} \end{cases}$$



Вот тут центр масс

$$E(X | A) = \int_1^2 t \cdot 0.625 dt + \int_2^3 t \cdot 0.375 dt = 1.875$$

10 Беспамятство экспоненциального распределения

Как уже говорилось, экспоненциальное распределение очень похоже на геометрическое. В том числе вот почему. Пусть продолжительность жизни лампочки T следует $\text{Exp}(\lambda)$. Следует ли поменять лампочку после того, как она проработала время t ? Посмотрим, сколько она еще проживет, то есть распределение $T - t$ при условии $T \geq t$.

$$\begin{aligned} F_{(T-t)|T \geq t}(x) &= \Pr(T - t \geq x \mid T > t) = \frac{\Pr(T - t \geq x \cap T > t)}{\Pr(T > t)} \\ &= \frac{\Pr(T \geq x + t)}{\Pr(T > t)} = \frac{e^{-\lambda(x+t)}}{e^{-\lambda t}} = e^{-\lambda x}, \end{aligned}$$

то есть распределение точно то же, как если мы заменим лампочку на новую.

11 Полные вероятность и матожидание

Напомним: пусть есть разбиение Ω на $\{A_i\}$ (не более, чем счетное), тогда

$$\begin{aligned} \Pr(B) &= \Pr(A_1) \Pr(B \mid A_1) + \dots + \Pr(A_n) \Pr(B \mid A_n) + \dots \\ p_X(x) &= \Pr(A_1) p_{X|A_1}(x) + \dots + \Pr(A_n) p_{X|A_n}(x) + \dots \end{aligned}$$

Ничего не меняется и в непрерывном случае. Сначала функция распределения:

$$\begin{aligned} F_X(x) &= \Pr(X \leq x) = \Pr(A_1) \Pr(X \leq x \mid A_1) + \dots + \Pr(A_n) \Pr(X \leq x \mid A_n) + \dots \\ &= \Pr(A_1) F_{X|A_1}(x) + \dots + \Pr(A_n) F_{X|A_n}(x) + \dots \end{aligned}$$

Дифференцируем, получаем:

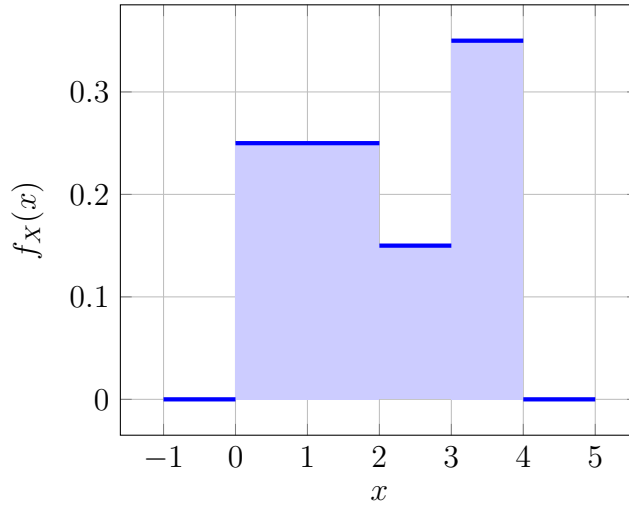
$$f_X(x) = \Pr(A_1) f_{X|A_1}(x) + \dots + \Pr(A_n) f_{X|A_n}(x) + \dots$$

Умножаем на x и интегрируем по всему \mathbb{R} :

$$E(X) = \Pr(A_1) E(X \mid A_1) + \dots + \Pr(A_n) E(X \mid A_n) + \dots$$

Пример (который уже был):

$$f_X(x) = \begin{cases} 0.25, & x \in [0, 2) \\ 0.15, & x \in [2, 3) \\ 0.35, & x \in [3, 4] \\ 0, & \text{иначе.} \end{cases}$$



Посчитаем матожидание

$$\begin{aligned} E(X) &= \Pr(X \in [0, 2))E(X \mid X \in [0, 2)) \\ &\quad + \Pr(X \in [2, 3))E(X \mid X \in [2, 3)) \\ &\quad + \Pr(X \in [3, 4))E(X \mid X \in [3, 4)) \end{aligned}$$

Заметим, что на каждом отрезке матожидание – это положение центра масс, то есть середина отрезка. Поэтому

$$E(X) = 0.5 \cdot 1 + 0.15 \cdot 2.5 + 0.35 \cdot 3.5 = 2.1$$

12 Смешанные распределения

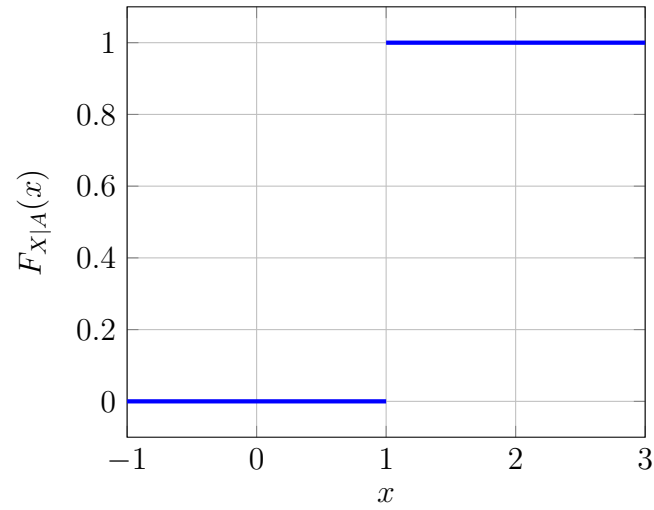
Иногда с.в. могут быть ни дискретными, ни непрерывными, например. Пусть у нас есть следующий эксперимент. Сначала бросаем честную монетку, потом, если выпал орел, то выбираем случайное число из отрезка $[0, 2]$ (равномерно). Случайная величина X при этом равна 1 в случае решки и равна выбранному числу в случае орла.

У данной с.в. нет функции вероятностей, как нет и плотности вероятности. Функция распределения все-таки есть. Как ее посчитать? По формуле полной вероятности, которая работает и для функции распределения. Пусть A — событие "выпала решка"

$$F_X(x) = F_{X|A}(x) \Pr(A) + F_{X|\bar{A}}(x) \Pr(\bar{A})$$

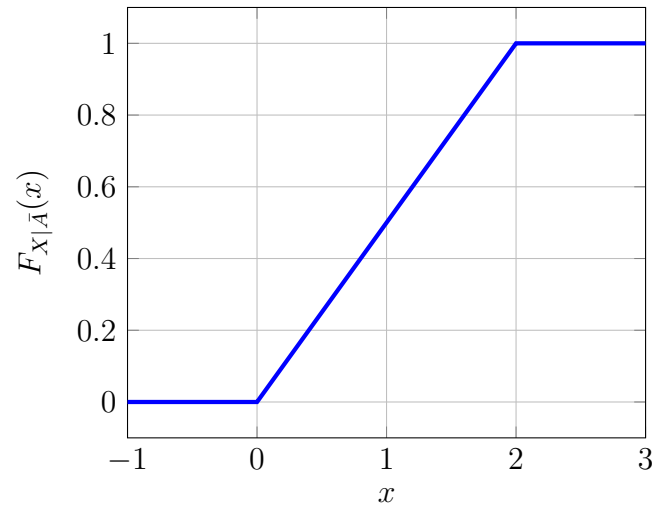
Заметим, что если A , то $X = 1$ с вероятностью 1. То есть,

$$F_{X|A}(x) = \begin{cases} 0, & x < 1 \\ 1, & x \geq 1 \end{cases}$$



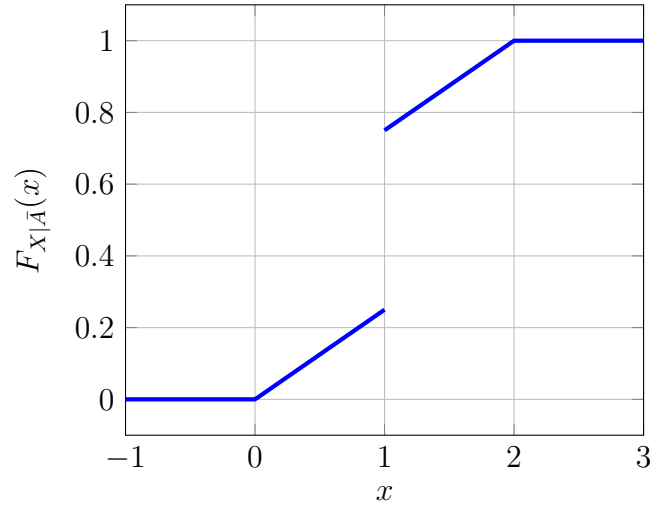
А если \bar{A} , то X следует равномерному распределению на отрезке $[0, 2]$.

$$F_{X|\bar{A}}(x) = \begin{cases} 0, & x < 0 \\ x/2, & x \in [0, 2] \\ 1, & x > 2 \end{cases}$$



Итоговая функция распределения выглядит так:

$$F_X(x) = \frac{1}{2}F_{X|A}(x) + \frac{1}{2}F_{X|\bar{A}}(x) = \begin{cases} 0, & x < 0 \\ x/4, & x \in [0, 1) \\ x/4 + 1/2, & x \in [1, 2] \\ 1, & x > 2 \end{cases}$$



13 Векторы из непрерывных с.в.

Пусть у нас есть 2 непрерывных с.в. X и Y . Тогда можно говорить о совместной плотности вероятности:

$$f_{X,Y}(x, y)\delta^2 \approx \Pr(X \in [x, x + \delta] \cap Y \in [y, y + \delta])$$

Более строго: если у нас есть такая функция $f_{X,Y}(x, y)$, такая что для любого измеримого множества $A \subset \mathbb{R}^2$ верно

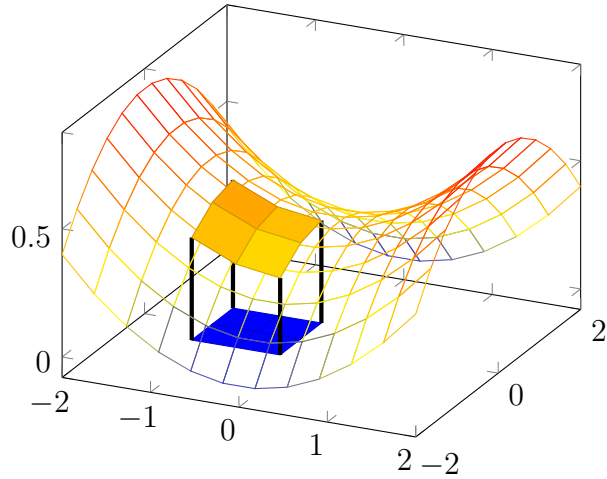
$$\Pr((X, Y) \in A) = \iint_A f_{X,Y}(x, y) dx dy,$$

то с.в. X и Y называются совместно непрерывными, а $f_{X,Y}(x, y)$ называется их совместной плотностью вероятности

Свойства, аналогичные совместной функции вероятности, только суммы заменены интегралами:

- $f_{X,Y}(x, y) \geq 0$
- $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx dy = 1$

Как визуализировать:



Чтобы посчитать вероятность, что X попадает в какое-то событие, надо просто посчитать объем подграфика на этом событии

NB: Одномерные события имеют вероятность ноль. Например, событие $Y = X$.

14 Маргинальные распределения

Покажем, что

$$\begin{cases} f_X(x) &= \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy \\ f_Y(y) &= \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx \end{cases}$$

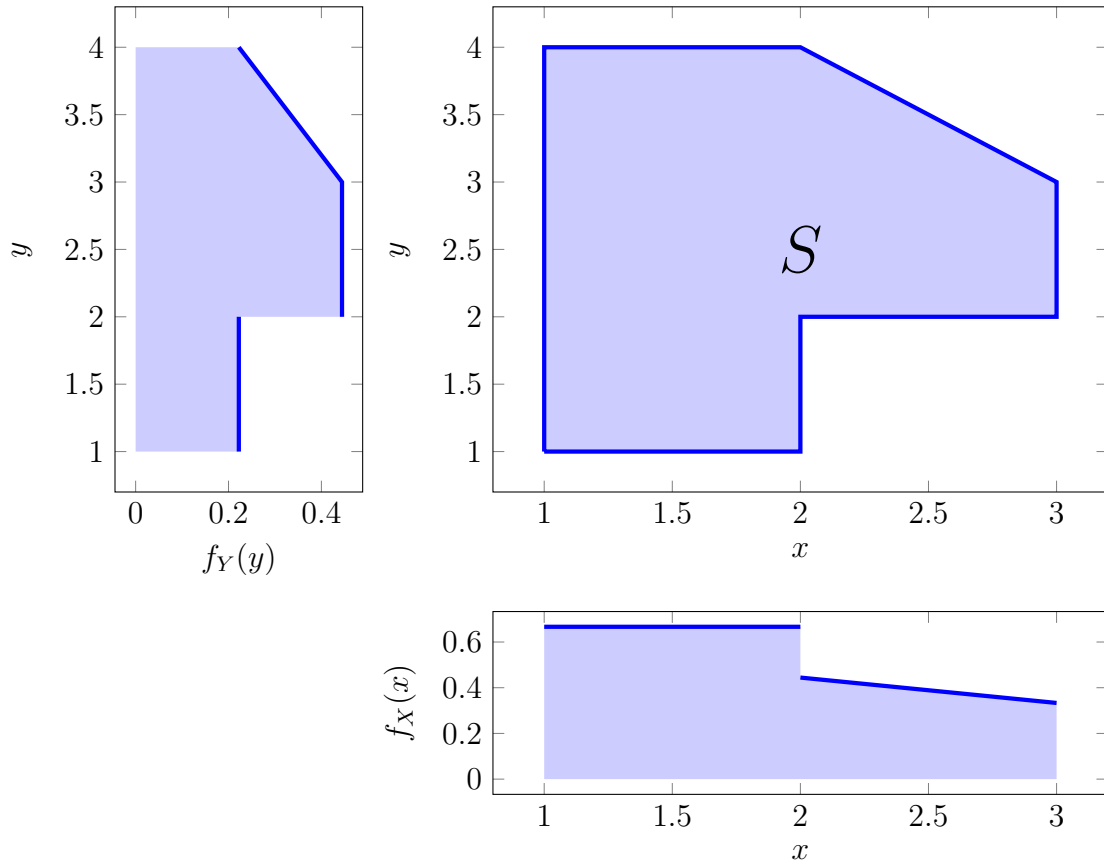
Для этого рассмотрим функцию распределения X :

$$F_X(x) = \Pr(X \leq x) = \int_{-\infty}^x \left(\int_{-\infty}^{+\infty} f_{X,Y}(t,y)dy \right) dt$$

$$f_X(x) = F'_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy$$

Пример: равномерное распределение на множестве S площадью 4.5

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{|S|}, & \text{если } (x,y) \in S, \\ 0, & \text{иначе.} \end{cases}$$



15 Более двух с.в. и функции от многих с.в.

Совместное распределение может быть задано на более, чем одной с.в., тогда есть плотность вероятности $f_{X,Y,Z}(x,y,z)$. Все то же самое, что было со многими дискретными с.в., только вместо сумм интегралы:

- $f_{X,Y,Z}(x,y,z) \geq 0$
- $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y,Z}(x,y,z) dx dy dz = 1$

И на многих с.в. мы также можем задавать функции, которые будут по сути новыми с.в.: $Z = g(X,Y)$. Их матожидание считается так:

$$E(g(X,Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y) f_{X,Y}(x,y) dx dy$$

При этом имеет место быть линейность матожидания:

$$E\left(\sum_i a_i X_i\right) = \sum_i a_i E(X_i)$$

16 Совместная функция распределения

Для нескольких с.в. определим функцию распределения:

$$F_{X,Y}(x, y) = \Pr(X \leq x \cap Y \leq y) = \int_{-\infty}^x \left(\int_{-\infty}^y f_{X,Y}(s, t) dt \right) ds$$

Также можно заметить:

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y}$$

17 С.в., условные на других с.в.

Было в дискретном случае:

- $p_{X,Y}(x, y)$ — совместная функция вероятности
- $p_{X|A}(x) = \frac{p_{X,Y}(x, y) \cdot [x \in A]}{\Pr(A)}$ — условная функция вероятности (где $[\cdot]$ — скобка Айверсона)
- $p_{X|Y}(x | y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}$ — функция вероятности X , условная на Y (определена только для таких y , что $p_Y(y) > 0$)

То же самое есть для непрерывного случая:

- $f_{X,Y}(x, y)$ — совместная плотность вероятности
- $f_{X|A}(x) = \frac{f_{X,Y}(x, y) \cdot [x \in A]}{\Pr(A)}$ — условная плотность вероятности
- $f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$ — плотность вероятности X , условная на Y (определена только для таких y , что $p_Y(y) > 0$)

Чуть ближе к формальному определению:

$$\begin{aligned} \Pr(X \in [x, x + \delta] | Y \in [y, y + \varepsilon]) &= \frac{\Pr(X \in [x, x + \delta] \cap Y \in [y, y + \varepsilon])}{\Pr(Y \in [y, y + \varepsilon])} \\ &\approx \frac{f_{X,Y}(x, y) \delta \varepsilon}{f_Y(y) \varepsilon} = \frac{f_{X,Y}(x, y)}{f_Y(y)} \delta \end{aligned}$$

И совсем формальное. Если существует такая функция $f_{X|Y}(x | y)$, что для всех y и для всех A верно

$$\Pr(X \in A | Y = y) = \int_A f_{X|Y}(x | y) dx,$$

то эта функция называется условной плотностью вероятности X при условии Y .

Опять при условиях у нас просто появляется новая вероятностная мера. Для всех $y : f_Y(y) > 0$ мы имеем

- $f_{X|Y}(x | y) \geq 0$
- $\int_{-\infty}^{+\infty} f_{X|Y}(x | y) dx = \frac{\int_{-\infty}^{+\infty} f_{X,Y}(x,y) dx}{f_Y(y)} = 1$

Из последнего видно, что стоит воспринимать условную плотность как срез совместной плотности по какому-то значению с.в.. Заметьте, что при этом вероятность этого среза равна нулю (одномерное множество), поэтому только терминами условности на события тут не обойтись.

Работает правило умножения:

$$\begin{aligned} f_{X,Y}(x, y) &= f_X(x) f_{Y|X}(y | x) \\ &= f_Y(y) f_{X|Y}(x | y) \end{aligned}$$

А значит, работают полные вероятность и матожидание. Полная вероятность (следует из правила умножения):

$$f_X(x) = \int_{-\infty}^{+\infty} f_Y(y) f_{X|Y}(x, y) dy$$

Также определено условное матожидание

$$E(X | Y = y) = \int_{-\infty}^{+\infty} x f_{X|Y}(x, y) dx$$

и работает теорема о полном матожидании (TODO: доказать)

$$E(X) = \int_{-\infty}^{+\infty} f_Y(y) E(X | Y = y) dy$$

Также можно посчитать условное матожидание функции с.в.

$$E(g(X) | Y = y) = \int_{-\infty}^{+\infty} g(x) f_{X|Y}(x, y) dx$$

18 Независимость с.в.

Было у дискретных:

$$p_{X,Y}(x, y) = p_X(x) p_Y(y)$$

Логично определить непрерывные с.в. X и Y независимыми, если

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

Это автоматически подразумевает, что для всех y верно, что $f_{X|Y}(x | y) = f_X(x)$. Свойств независимых с.в. те же, что и у дискретных

- $E(XY) = E(X)E(Y)$
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$
- $g(X)$ и $h(Y)$ тоже будут независимы

Пример про зависимые с.в.

Есть палка длиной ℓ . Ломаем ее в случайном месте $X \in [0, \ell]$, остаток тоже ломаем в случайном месте $Y \in [0, X]$. Найти:

- $f_{X,Y}$
- $f_Y(y)$
- $E[Y]$

Закончили на определении плотности вероятности с.в. X при известном значении с.в. Y .

$$\Pr(X \in A | Y = y) = \int_A f_{X|Y}(x | y)dx,$$

И ее свойствах:

- $f_{X|Y}(x | y) \geq 0$
- $\int_{-\infty}^{+\infty} f_{X|Y}(x | y)dx = \frac{\int_{-\infty}^{+\infty} f_{X,Y}(x|y)dx}{f_Y(y)} = 1$

Еще раз подчеркнем, что стоит воспринимать условную плотность как срез совместной плотности по какому-то значению с.в.. Заметьте, что при этом вероятность этого среза равна нулю (одномерное множество), поэтому только терминами условности на события тут не обойтись.

Работает правило умножения:

$$\begin{aligned} f_{X,Y}(x, y) &= f_X(x)f_{Y|X}(y | x) \\ &= f_Y(y)f_{X|Y}(x | y) \end{aligned}$$

А значит, работают полные вероятность и матожидание. Полная вероятность (следует из правила умножения и определения маргинальной плотности вероятности):

$$f_X(x) = \int_{-\infty}^{+\infty} f_Y(y) f_{X|Y}(x | y) dy$$

Также определено условное матожидание

$$E(X | Y = y) = \int_{-\infty}^{+\infty} x f_{X|Y}(x | y) dx$$

и работает теорема о полном матожидании:

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f_X(x) dx = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f_Y(y) f_{X|Y}(x | y) dy dx \\ &= \int_{-\infty}^{+\infty} f_Y(y) \left(\int_{-\infty}^{+\infty} x f_{X|Y}(x | y) dx \right) dy \\ &= \int_{-\infty}^{+\infty} f_Y(y) E(X | Y = y) dy \end{aligned}$$

Также можно посчитать условное матожидание функции с.в.

$$E(g(X) | Y = y) = \int_{-\infty}^{+\infty} g(x) f_{X|Y}(x, y) dx$$

19 Независимость с.в.

Было у дискретных:

$$p_{X,Y}(x, y) = p_X(x) p_Y(y)$$

Логично определить непрерывные с.в. X и Y независимыми, если

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

Это автоматически подразумевает, что для всех y верно, что $f_{X|Y}(x | y) = f_X(x)$. Свойства независимых с.в. те же, что и у дискретных

- $E(XY) = E(X)E(Y)$
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$
- $g(X)$ и $h(Y)$ тоже будут независимы

Пример независимых с.в.

Пусть есть два трамвая, которые ходят с интервалом 11 и 17 минут, независимо друг от друга. Вы приходите на остановку в случайный момент времени. Сколько ожидаемо вы будете ожидать трамвай, если вам подходит любой из двух?

Определим с.в. Пусть T_1 — время, через которое придет первый трамвай, а T_2 — второй. Заметим, что $T_1 \sim U(0, 11)$, а $T_2 \sim U(0, 17)$, то есть

$$F_{T_1}(t) = \begin{cases} 0, & t < 0 \\ \frac{t}{11}, & t \in [0, 11] \\ 1, & t > 11, \end{cases}$$
$$F_{T_2}(t) = \begin{cases} 0, & t < 0 \\ \frac{t}{17}, & t \in [0, 17] \\ 1, & t > 17, \end{cases}$$

А время, которое придется провести на остановке есть $Y = \min\{T_1, T_2\}$, функция от двух независимых с.в.

Определим функцию распределения Y .

$$\begin{aligned} F_Y(y) &= \Pr(Y \leq y) = \Pr(T_1 \leq y \cup T_2 \leq y) \\ &= 1 - \Pr(T_1 > y \cap T_2 > y) \\ &= 1 - \Pr(T_1 > y) \Pr(T_2 > y) = 1 - (1 - F_{T_1}(y))(1 - F_{T_2}(y)) \end{aligned}$$

Воспользуемся полезной формулой с прошлой практики для неотрицательных с.в. (которую доказали не все)

$$\begin{aligned} E(Y) &= \int_0^{+\infty} (1 - F_Y(y)) dy = \int_0^{+\infty} (1 - F_{T_1}(y))(1 - F_{T_2}(y)) dy \\ &= \int_0^{11} \left(1 - \frac{y}{11}\right) \left(1 - \frac{y}{17}\right) dy \approx 4.31 \end{aligned}$$

Пример про зависимые с.в.

Есть палка длиной ℓ . Ломаем ее в случайном месте $X \in [0, \ell]$, остаток тоже ломаем в случайном месте $Y \in [0, X]$. Какова длина остатка Y ? Заметим, что величины зависимы: если $X \geq \ell/2$, то есть шанс, что $Y \geq \ell/2$, а иначе нет.

Совместная функция распределения:

$$f_{X,Y}(x, y) = f_X(x) f_{Y|X}(y | x) = \frac{1}{\ell x}$$

Теперь можем найти плотность вероятности Y и его матожидание. Для этого интегрируем совместную ПВ по всем возможным X , то есть от y до ℓ

$$f_Y(y) = \int_y^\ell f_{X,Y}(x, y) dx = \int_0^\ell \frac{1}{\ell x} dx = \frac{1}{\ell} \ln \frac{\ell}{y}.$$

Два способа посчитать матожидание. Первый:

$$E(Y) = \int_0^\ell y f_Y(y) dy = \int_0^\ell \frac{y}{\ell} \ln \frac{\ell}{y} dy = \left(\frac{\ln(\ell)}{\ell} \cdot \frac{y^2}{2} - \frac{y^2 \ln(y)}{2\ell} + \frac{y^2}{4\ell} \right) \Big|_0^\ell = \frac{\ell}{4}$$

Второй:

$$E(Y) = \int_0^\ell f_X(x) E(Y | X = x) dx = \int_0^\ell \frac{1}{\ell} \cdot \frac{x}{2} dx = \frac{x^2}{4\ell} \Big|_0^\ell = \frac{\ell}{4}$$

Интуитивная логика: первый раз палка ломается в среднем посередине, потом снова в среднем посередине, то есть средняя длина должна быть $\ell/4$. Она работает не всегда, а только для Y , матожидание которых линейно относительно X . Пусть $E(Y | X = x) = g(x)$ тогда:

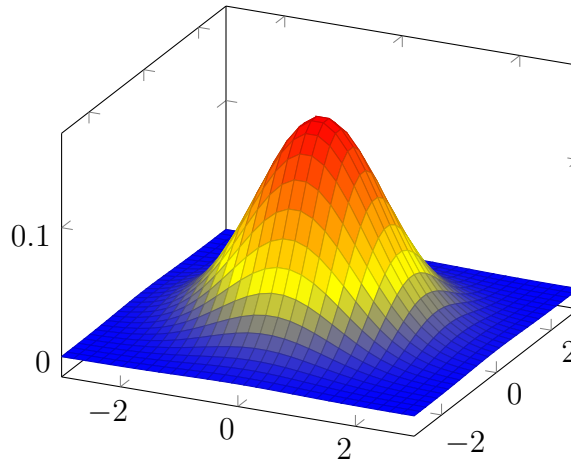
$$E(Y) = \int_{-\infty}^{+\infty} f_X(x) E(Y | X = x) dx = \int_{-\infty}^{+\infty} f_X(x) g(x) dx = E(g(X)) \neq g(E(X)).$$

Независимые нормальные распределения

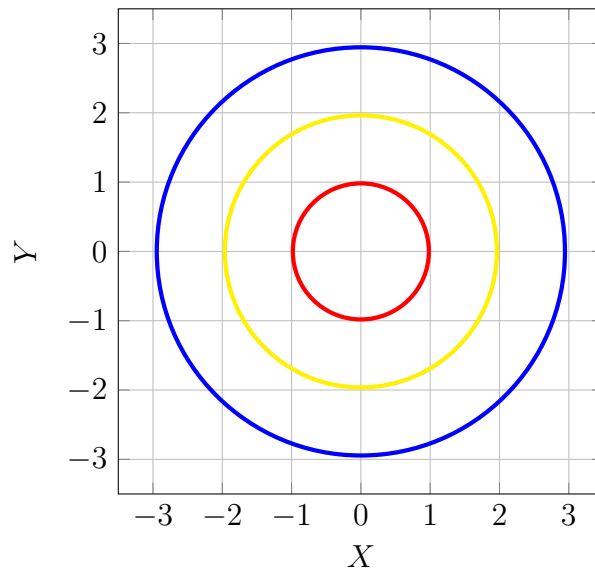
Пусть есть $X \sim N(0, 1)$ и $Y \sim N(0, 1)$, и они независимы. Тогда

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \frac{1}{2\pi} \exp \left(-\frac{x^2 + y^2}{2} \right).$$

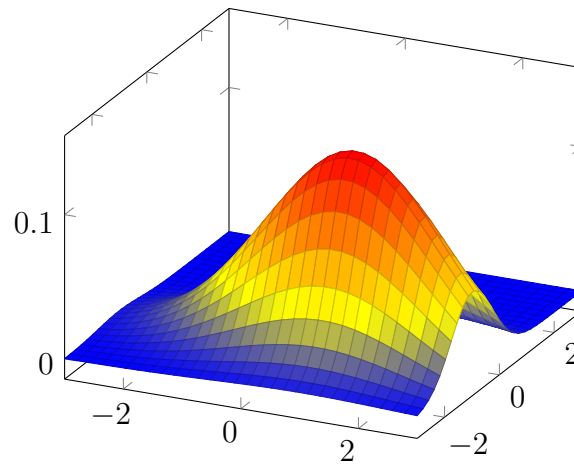
То есть их совместная плотность вероятности пропорциональна $e^{-r^2/2}$, где r — расстояние до точки $(0, 0)$.



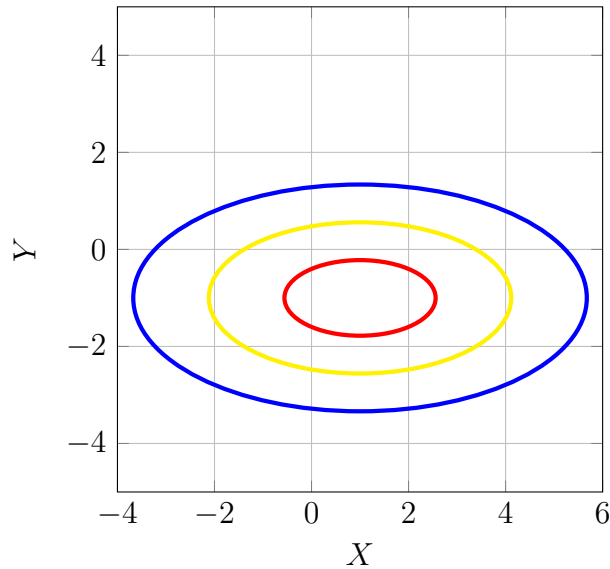
Эквиплотные линии:



Если у нас нестандартные нормальные распределения, то бугорок может сплющиваться или растягиваться вдоль оси, и его центр будет сдвигаться



Эквиплотные линии превратятся в эллипсы.



Заметим, что оси эллипса должны быть направлены вдоль осей, иначе с.в. будут зависимые

20 Формула Байеса для случайных величин

Напомним смысл формулы Байеса. С помощью нее мы выражаем вероятность события, которое не можем пронаблюдать (A_i в формуле) через априорные оценки вероятностей другого события(ий) (B в формуле), которое мы можем наблюдать, после его наблюдения.

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

Априорная оценка вероятности A_i Априорная оценка условной вероятности наблюдаемого события B

Апостериорная оценка вероятности A_i Полная вероятность B

Легко вывести из правила умножения и формулы полной вероятности для двух дискретных и двух непрерывных случайных величин. Начнем с дискретных

$$\begin{aligned}
p_{X,Y}(x, y) &= p_Y(y)p_{X|Y}(x | y) \\
&= p_X(x)p_{Y|X}(y | x) \\
\Rightarrow p_{X|Y}(x | y) &= \frac{p_X(x)p_{Y|X}(y | x)}{p_Y(y)} \\
&= \frac{p_X(x)p_{Y|X}(y | x)}{\sum_x p_X(x)p_{Y|X}(y | x)}
\end{aligned}$$

$$p_{X|Y}(x | y) = \frac{p_X(x)p_{Y|X}(y | x)}{\sum_x p_X(x)p_{Y|X}(y | x)}$$

То же самое для непрерывных с.в., но говорим про плотности вероятности, а не про функцию вероятности, и суммы заменяем интегралами

$$\begin{aligned}
f_{X,Y}(x, y) &= f_Y(y)f_{X|Y}(x | y) \\
&= f_X(x)f_{Y|X}(y | x) \\
\Rightarrow f_{X|Y}(x | y) &= \frac{f_X(x)f_{Y|X}(y | x)}{f_Y(y)} \\
&= \frac{f_X(x)f_{Y|X}(y | x)}{\int_{\mathbb{R}} f_X(x)f_{Y|X}(y | x)dx}
\end{aligned}$$

$$f_{X|Y}(x | y) = \frac{f_X(x)f_{Y|X}(y | x)}{\int_{\mathbb{R}} f_X(x)f_{Y|X}(y | x)dx}$$

Но иногда могут быть случаи, когда есть две с.в.: одна дискретная, другая непрерывная. Можем сделать примерно следующее. Пусть X — дискретная, а Y — непрерывная

$$\begin{aligned}
&\Pr(X = x \cap y \leq Y \leq y + \delta) \\
&= \Pr(X = x) \Pr(y \leq Y \leq y + \delta | X = x) \approx p_X(x)f_{Y|X}(y | x)\delta \\
&= \Pr(y \leq Y \leq y + \delta) \Pr(X = x | y \leq Y \leq y + \delta) \approx f_Y(y)\delta p_{X|Y}(x | y)
\end{aligned}$$

Откуда следует, что

$$p_X(x)f_{Y|X}(y | x) = f_Y(y)p_{X|Y}(x | y)$$

Чтобы доказать более строго, надо просто δ устремить к нулю, тогда вместо “ \approx ” будет “ $=$ ”. Получаем две формулы.

$$p_{X|Y}(x | y) = \frac{p_X(x)f_{Y|X}(y | x)}{f_Y(y)}$$

$$f_{Y|X}(y | x) = \frac{f_Y(y)p_{X|Y}(x | y)}{p_X(x)}$$

И теперь в левую формулу можно подставить формулу полной вероятности, которая работает для $f_Y(y)$:

$$f_Y(y) = \sum_{x'} p_X(x') f_{Y|X}(y | x').$$

С правой чуть посложнее, но пока поверьте наслову, что

$$p_X(x) = \int_{\mathbb{R}} f_Y(y') p_{X|Y}(x | y') dy.$$

Пример: наблюдаем непрерывную с.в., оцениваем дискретную

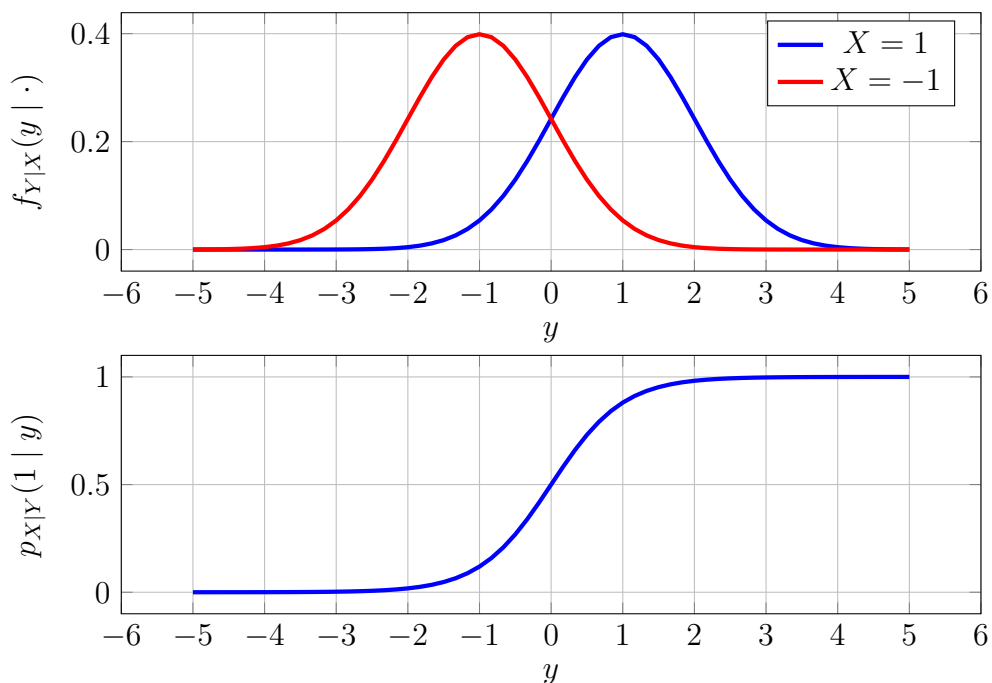
$$p_{X|Y}(x | y) = \frac{p_X(x)f_{Y|X}(y | x)}{f_Y(y)}$$

Ситуация: посылаем дискретный сигнал $X \in [-1, 1]$, но к нему добавляется шум $Z \sim N(0, 1)$. В итоге мы можем замерить только $Y = X + Z$. Давайте определим вероятности каждого варианта посланного сигнала, если изначально отправка каждого равновероятна.

- $p_X(-1) = p_X(1) = \frac{1}{2}$
- $Y \sim N(0, 1) + X$, то есть если $X = 1$, то $Y | X = 1 \sim N(1, 1)$. Аналогично $Y | X = -1 \sim N(-1, 1)$
- Из предыдущего понимаем, что
 - $f_{Y|X}(y, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-1)^2}{2}}$
 - $f_{Y|X}(y, -1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y+1)^2}{2}}$
- $f_Y(y) = \frac{1}{2} f_{Y|X}(y, 1) + \frac{1}{2} f_{Y|X}(y, -1)$

$$\begin{aligned} p_{X|Y}(1 | y) &= \frac{\frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-1)^2}{2}}}{\frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-1)^2}{2}} + \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(y+1)^2}{2}}} \\ &= \frac{1}{1 + e^{-\frac{(y+1)^2 - (y-1)^2}{2}}} = \frac{1}{1 + e^{-2y}} \end{aligned}$$

Иллюстрация: распределение наблюдаемого значения при разных сигналах и вероятность, что посланный сигнал равен единице, в зависимости от наблюдаемого значения



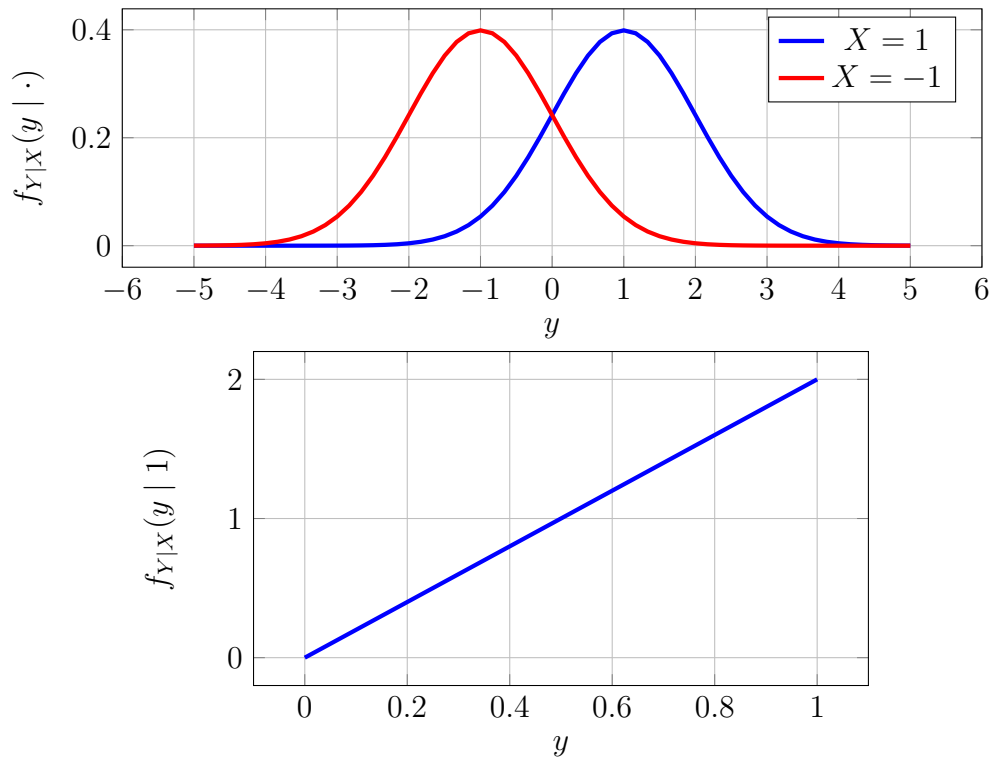
Пример: наблюдаем дискретную с.в., оцениваем непрерывную

$$f_{Y|X}(y | x) = \frac{f_Y(y)p_{X|Y}(x | y)}{p_X(x)}$$

Эксперимент: берем нечестную монету, бросаем ее и исходя из результата хотим оценить степень ее нечестности. Наблюдаемая с.в. $X \in \{0, 1\}$ и неизвестная с.в. $Y = \Pr(X = 1) \sim U(0, 1)$.

- $f_Y(y) = 1$ на отрезке $[0, 1]$
- $p_{X|Y}(1 | y) = y$
- $p_{X|Y}(0 | y) = 1 - y$
- $p_X(1) = \int_0^1 f_Y(y')p_{X|Y}(1 | y')dy' = \int_0^1 y'dy' = \frac{y^2}{2}\Big|_0^1 = \frac{1}{2}$

$$f_{Y|X}(y | 1) = \frac{1 \cdot y}{1/2} = 2y$$



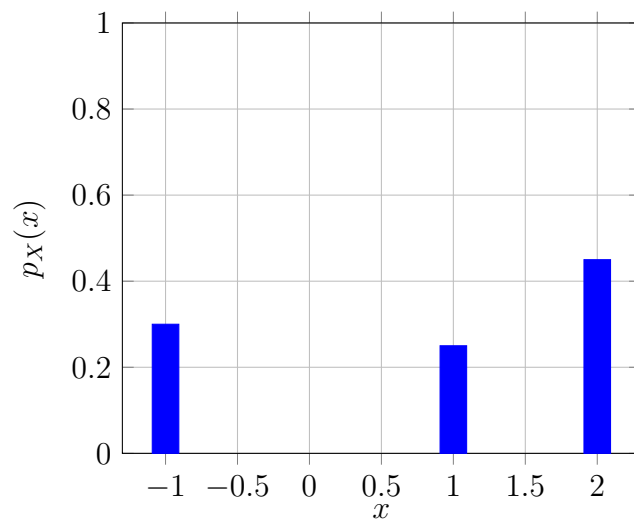
21 Линейные функции от с.в.

Дискретные с.в.

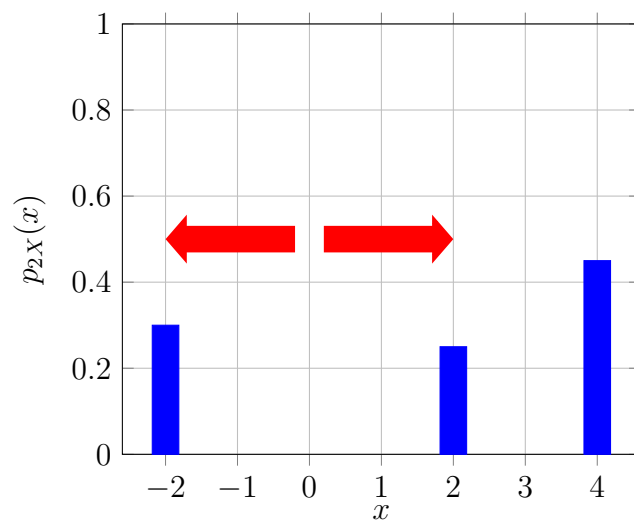
Если $Y = g(X)$ и мы знаем функцию вероятности $p_X(x)$, то не составит труда посчитать

$$p_Y(y) = \sum_{x:g(x)=y} p_X(x)$$

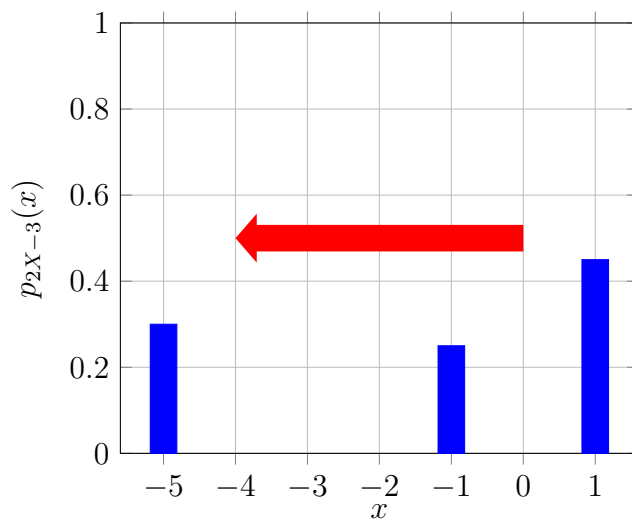
рассмотрим простой случай: $g(x) = ax + b$ — линейная функция. Что происходит с функцией распределения?



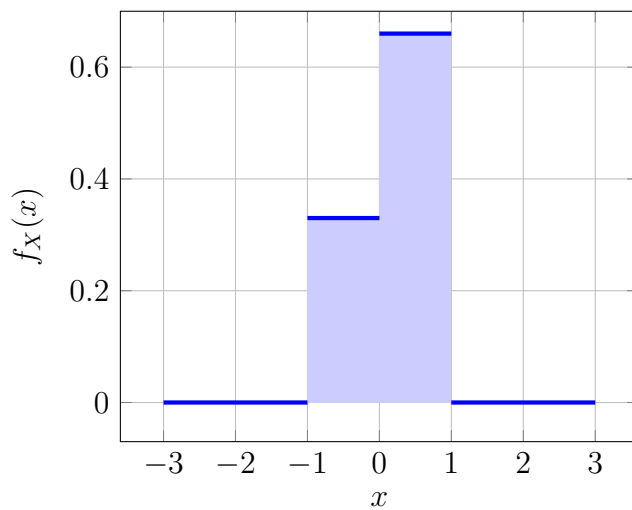
Рассмотрим сначала, например, $g(x) = 2x$. Легко понять, что функция вероятностей сохранила свою форму, просто столбики отъехали от оси OY в два раза.



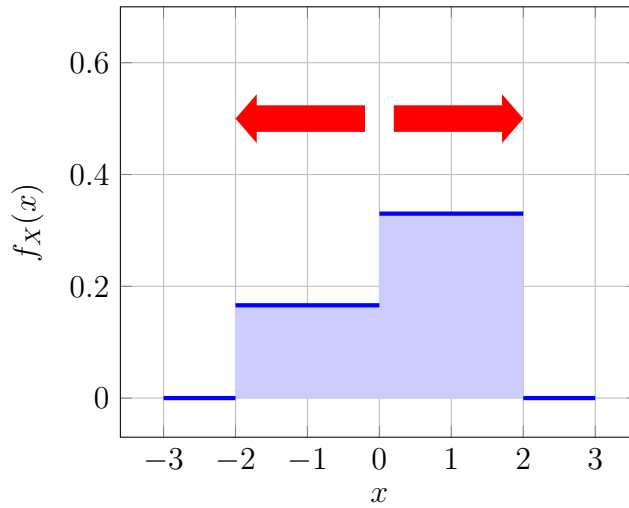
Если мы еще прибавим константу $b = -3$, то просто сдвинем все столбики влево на 3.



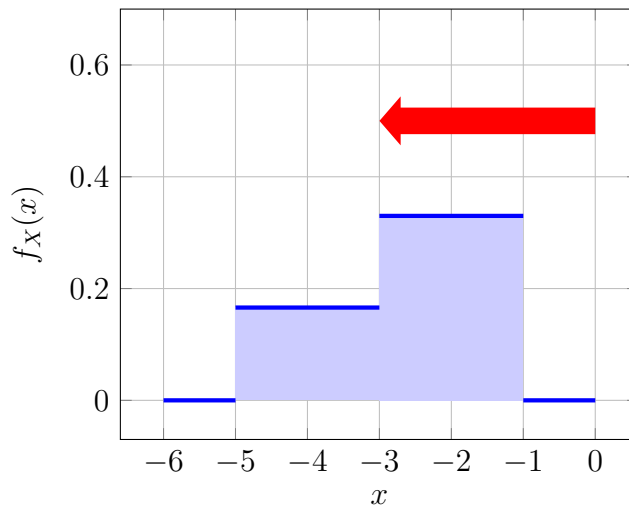
Очень похожая история с непрерывными. Рассмотрим пример.



Для того, чтобы получить функцию плотности вероятности для с.в. $Y = 2x$, надо растянуть ее от оси OY в два раза. Заметьте, что при растягивании сама плотность упадет также в два раза.



Ну и сдвиг при добавлении константы аналогичный. Рассмотрим, например, $g(x) = 2x - 3$



Как это в общем случае делать с.в.? Рассмотрим $Y = aX + b$ (где $a \neq 0$, иначе это скучный случай, когда $Y = b$ с вероятностью 1). Пусть сначала X дискретная, и нам известна ее функция вероятностей. Тогда

$$p_Y(y) = \Pr(Y = y) = \Pr(aX + b = y) = \Pr\left(X = \frac{y - b}{a}\right) = p_X\left(\frac{y - b}{a}\right)$$

С непрерывными такое не работает, так как вероятность, что непрерывная с.в. равна конкретному числу, есть ноль. Но мы можем работать с функциями распределения! Допустим мы знаем $f_X(x)$ и $F_X(x)$. Рассмотрим случай $a > 0$

$$F_Y(y) = \Pr(Y \leq y) = \Pr(aX + b \leq y) = \Pr\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right),$$

$$f_Y(y) = F'_Y(y) = f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a}.$$

И отдельно $a < 0$

$$F_Y(y) = \Pr(Y \leq y) = \Pr(aX + b \leq y) = \Pr\left(X \geq \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right),$$

$$f_Y(y) = F'_Y(y) = -f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a}.$$

Объединяя два случая, получаем:

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

Докажем теперь, что линейное преобразование нормального распределения оставляет его нормальным. Пусть $X \sim N(\mu, \sigma^2)$ и $Y = aX + b$. Значит,

$$\begin{aligned} f_Y(y) &= \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) \\ &= \frac{1}{|a|\sigma\sqrt{2\pi}} \exp\left(-\frac{\left(\frac{y-b}{a} - \mu\right)^2}{2\sigma^2}\right) \\ &= \frac{1}{(|a|\sigma)\sqrt{2\pi}} \exp\left(-\frac{(y - (b + a\mu))^2}{2(a\sigma)^2}\right), \end{aligned}$$

что есть функция плотности вероятности для $N(a\mu + b, (a\sigma)^2)$.