Лекция 6. Формула Байеса, функции от с.в.

21 марта 2022 г.

1 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 веротностей другого события(ий) (B в формуле), которое мы можем наблюдать, после его наблюдения.

Априорная оценка вероятности A_i

Априорная оценка условной вероятности наблюдаемого события B

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

Апостериорная оценка вероятности A_i

Полная вероятность B

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

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

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

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

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

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

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

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

Откуда следует, что

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

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

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

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

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

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

С правой чуть посложнее, но пока поверьте наслово, что

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

Пример: наблюдаем непрерывную с.в., оцениваем дискретную

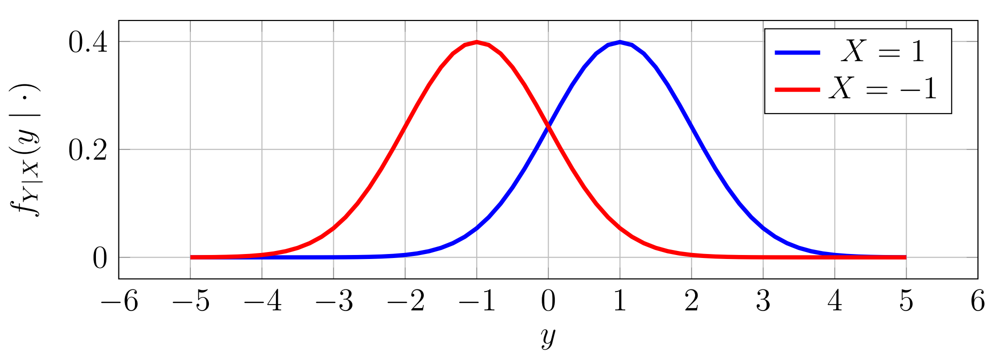
$$p_{X|Y}(x \mid y) = \frac{p_X(x) f_{Y|X}(y \mid x)}{f_Y(y)}$$

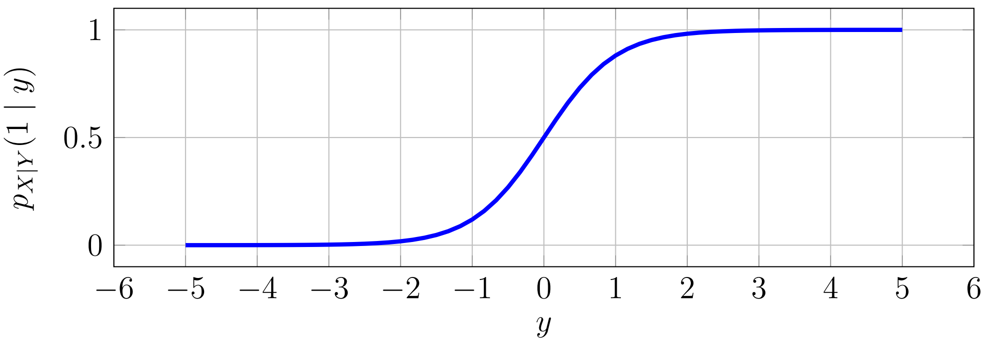
Ситуация: посылаем дискретный сигнал $X \in [-1, 1]$, но к нему добавляется шум $Z \sim N(0,1)$. В итоге мы можем замерить только $Y = X + Z$. Давайте определим вероятности каждого варианта посланного сигнала, если изначально отправка каждого равновероятна.

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

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

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





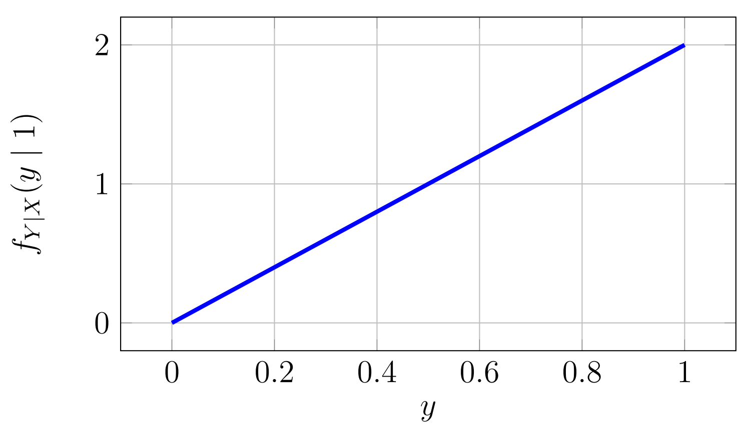
Пример: наблюдаем дискретную с.в., оцениваем непрерывную

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

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

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



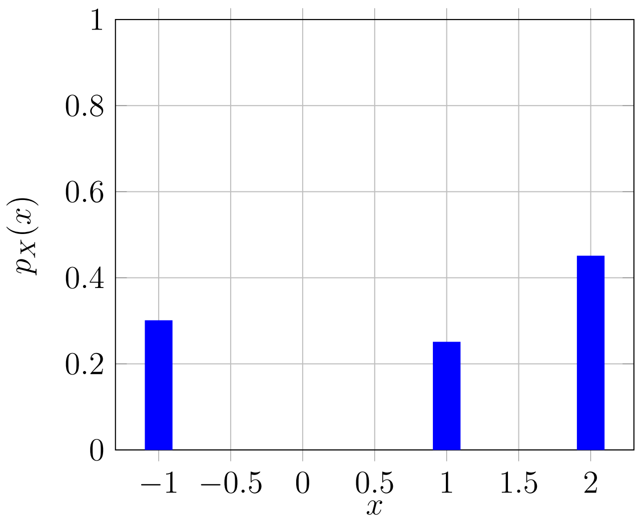
2 Линейные функции от с.в.

Дискретные с.в.

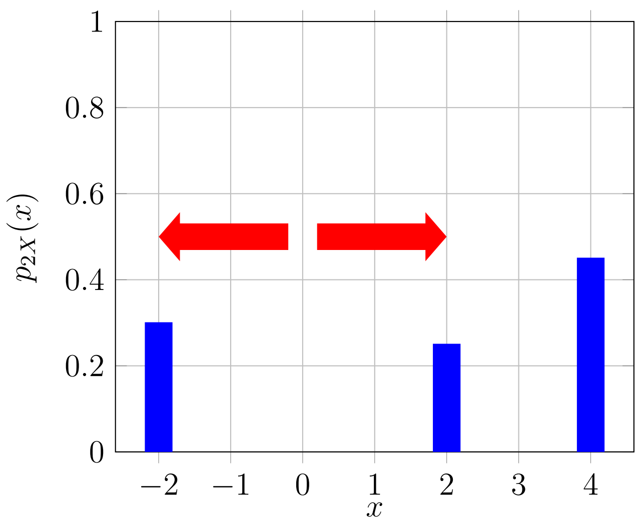
Если $Y = g(X)$ и мы знаем функцию вероятности $p_X(x)$, то не соствит труда посчитать

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

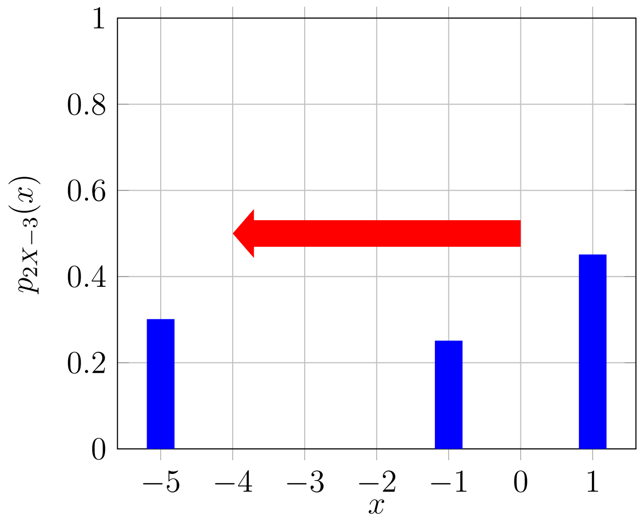
рассмотрим простой случай: $g(x) = ax + b$ — линейная функция. Что происходит с функцией распределения?



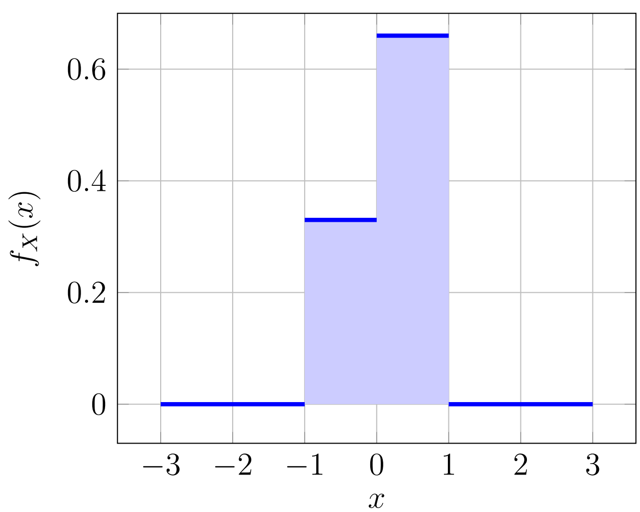
Расмотрим сначала, например, $g(x) = 2x$. Легко понять, что функция вероятностей сохранила свою форму, просто столбики отъехали от оси OY в два раза.



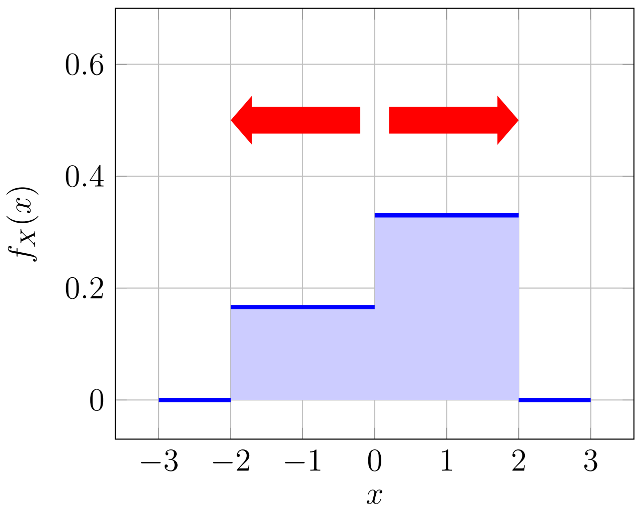
Если мы еще прибавим константу $b = -3$, то просто сдвинем все столбики влево на 3.



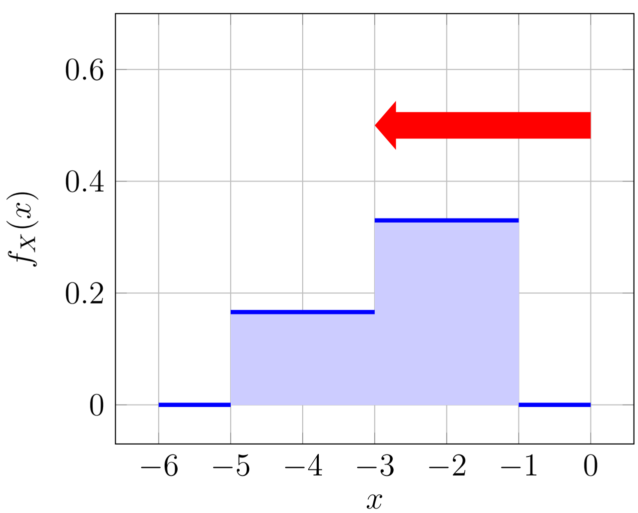
Очень похожая история с **непрерывными**. Рассмотрим пример.



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 $a > 0$

$$F_Y(y) = \Pr(Y \le y) = \Pr(aX + b \le y) = \Pr\left(X \le \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right),$$
$$f_Y(y) = F'_Y(y) = f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a}.$$

И отдельно $a < 0$

$$F_Y(y) = \Pr(Y \le y) = \Pr(aX + b \le y) = \Pr\left(X \ge \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right),$$
$$f_Y(y) = F'_Y(y) = -f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a}.$$

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

3 Нелинейные преобразования

Алгоритм поиска нового распределения $Y = g(X)$:

1. Найти функцию распределения $F_Y(y) = \Pr(Y \le y) = \Pr(g(X) \le y)$
2. Продифференцировать: $f_Y(y) = F'_Y(y)$.

Рассмотрим случай, когда $g(x)$ монотонно возрастает и дифференцируема.

$$F_Y(y) = \Pr(Y \le y) = \Pr(g(X) \le y) = \Pr(X \le g^{-1}(y)) = F_X(g^{-1}(y))$$
$$f_Y(y) = F_Y'(y) = F_X'(g^{-1}(y)) = f_X(g^{-1}(y))(g^{-1}(y))' = \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))}$$

Аналогично можно рассмотреть случай, когда $g(x)$ монотонно убывает и дифференцируема

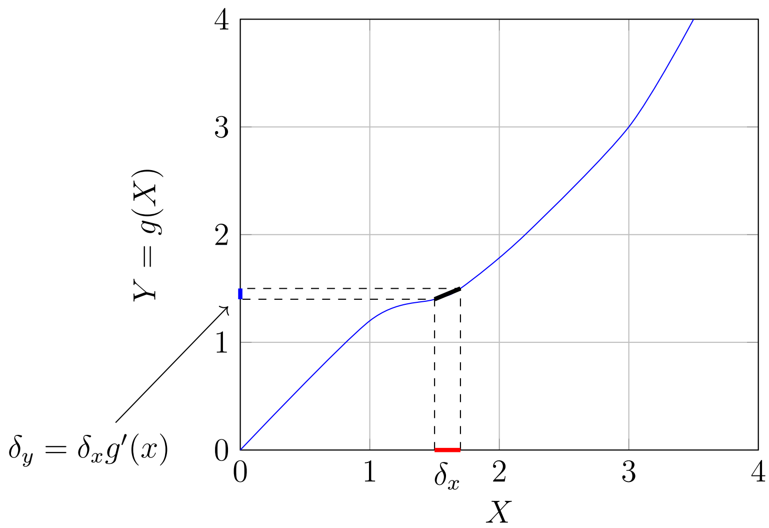
$$F_Y(y) = \Pr(Y \le y) = \Pr(g(X) \le y) = \Pr(X \ge g^{-1}(y)) = 1 - F_X(g^{-1}(y))$$
$$f_Y(y) = F_Y'(y) = -F_X'(g^{-1}(y)) = -f_X(g^{-1}(y))(g^{-1}(y))' = -\frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))} = \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|}$$

То есть в общем случае (когда $g(x)$ строго монотонна):

$$f_Y(y) = \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|}$$

Интуитивное объяснение (которое использовалось на практике). Пусть $y = g(x)$

$$f_Y(y)\delta_y \approx \Pr(y \le Y \le y + \delta_y) = \Pr(x \le X \le x + \delta_y) = f_X(x)\delta_x = f_X(x)\frac{\delta_y}{g'(x)}$$



Для немонотонных $g(x)$ алгоритм остается прежний, но нет гарантий, что все пройдет легко. рассмотрим $g(x) = x^2$ и $Y = g(X)$ (при известной плотности X).

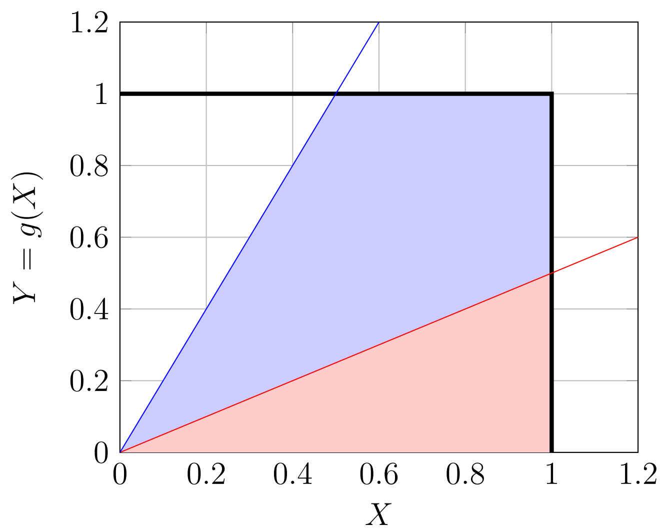
$$
\begin{aligned}
F_Y(y) &= \Pr(Y \le y) = \Pr(X^2 \le y) = 1 - \Pr(X^2 > y) \\
&= 1 - \Pr(X > \sqrt{y}) - \Pr(X < -\sqrt{y}) \\
&= \Pr(X \le \sqrt{y}) = \Pr(X < -\sqrt{y}) = F_X(\sqrt{y}) - F_X(-sqrty) \\
f_Y(y) &= F_Y'(y) = F_X'(\sqrt{y}) - F_X'(-\sqrt{y}) = f_X(\sqrt{y})\frac{1}{2\sqrt{y}} - f_X(-\sqrt{y})\left(-\frac{1}{2\sqrt{y}}\right) \\
&= \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}}
\end{aligned}
$$

4 Функции нескольких с.в.

Если есть $Z = g(X, Y)$, то алгоритм нахождения ее функции распределения или плотности такой же. Рассмотрим на примере. X, Y – независимые, обе следуют $U(0,1)$. Рассмотрим с.в. $Z = \frac{Y}{X}$.

Найдем функцию распределения $F_Z(z)$

$$
F_Z(z) = \Pr(\frac{Y}{X} \le z).
$$



Если z меньше единицы, то это просто интеграл по красному треугольнику, то есть его площадь (так как совместная плотность $f_{X,Y}(x,y) = 1$ на всем квадрате).

$$
F_Z(z) = \frac{z}{2}, \text{ если } z \in [0,1].
$$

Если $z > 1$, то это площадь всего квадрата минус площадь незакрашенного треугольника

$$F_Z(z) = 1 - \frac{1}{2z}.$$

Объединяя все вместе, получаем:

$$F_Z(z) = \begin{cases} 0, & \text{если } z < 0, \\ \frac{z}{2}, & \text{если } z \in [0, 1), \\ 1 - \frac{1}{2z}, & \text{если } z \geq 1. \end{cases}$$

4.1 Сумма независимых с.в.

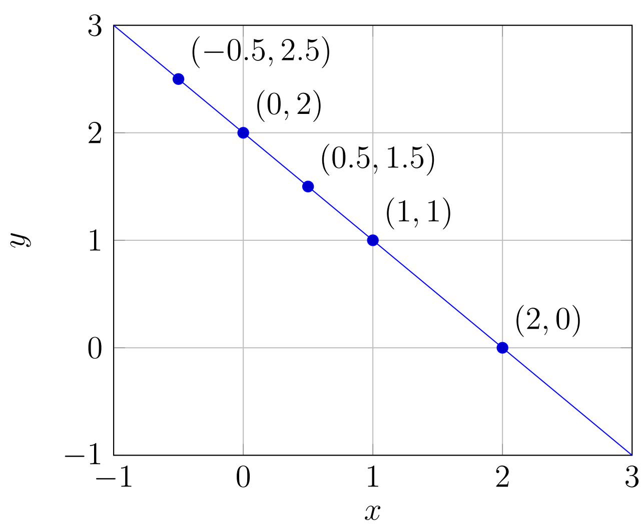
В данном подразделе мы имеем две независимых с.в. X и Y, знаем их распределение и хотим узнать распределение с.в. $Z = X + Y$.

Дискретный случай

Чтобы посчитать вероятность того, что $Z = z$, необходимо, чтобы одновременно выполнялось:

- $X = x$ (где x — какое-то число из множества значений X)
- $Y = z - x$.

То есть мы берем все пары (x, y) с прямой $x + y = z$ и суммируем вероятности этих пар.



Заметим, что из-за независимости X и Y события $X = x$ и $Y = z - x$ независимы, поэтому

$$\Pr(X = x \cap Y = z - x) = p_X(x)p_Y(z - x).$$

Таким образом, формула выглядит так:

$$p_Z(z) = \sum_x p_X(x) p_Y(z-x)$$

Непрерывный случай

Аналогичная формула для непрерывных с.в.:

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx$$

Чтобы ее доказать, обусловимся сначала на событии $X = x$. Тогда

$$f_{Z|X}(z \mid x) = f_{Y+x|X}(z \mid x) = f_{Y+x}(z) = f_Y(z-x).$$

Далее, по формуле полной вероятности имеем

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_{Z|X}(z \mid x) dx = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx/$$

Интересное следствие

Сумма независимых нормальных распределений есть нормальное распределение. Пусть $X \sim N(\mu_X, \sigma_X^2)$ и $Y \sim N(\mu_Y, \sigma_Y^2)$ — независимы. Посчитаем $f_Z(z)$, где $Z = X + Y$.

$$\begin{aligned}
f_Z(z) &= \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sigma_X\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right) \cdot \frac{1}{\sigma_Y\sqrt{2\pi}} \exp\left(-\frac{(z-x-\mu_Y)^2}{2\sigma_Y^2}\right) dx \\
&= \frac{1}{\sqrt{2\pi(\sigma_X^2+\sigma_Y^2)}} \exp\left(-\frac{(z-\mu_X-\mu_Y)^2}{2(\sigma_X^2+\sigma_Y^2)}\right)
\end{aligned}$$

Тут было много несложной, но громоздкой математики, которую мы опустили. Таким образом, $Z \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$. Единственное удивительное здесь – то, что мы по-прежнему имеем нормальное распределение. Его параметры при этом очевидны из линейности матожиданий и из свойств дисперсии независимых с.в.

По индукции легко доказать, что для любого конечного множества независимых нормальных с.в. их сумма будет также нормальной с.в.