Теория вероятностей и математическая статистика

Непрерывные случайные величины.

Глеб Карпов

ФКН ВШЭ

Мотивация для нового класса случайных величин

Пример 1: Физические измерения

- Рост человека
- Температура воздуха
- Скорость автомобиля

Пример 2: Время ожидания

- Время "жизни" детали
- Время между звонками в call-центр

Пример 3: Экономические показатели

- Доходы компании
- Цены на акции

Пример 4: Технические параметры

- Напряжение в сети
- Давление в шинах
- Концентрация вещества

Почему дискретные СВ не подходят?

- **Бесконечное количество значений**: Время может быть 1.234567 секунды
- **Непрерывность**: Между любыми двумя значениями есть промежуточные
- **Точность измерений**: Современные приборы дают очень точные результаты

Вероятность в точке теряет смысл

- Для непрерывных величин вероятность **точечного** события P(X = a) всегда равна нулю
- Отныне нас будует интересовать только вероятности попадания в **интервалы**:

$$P(a < X < b),\ P(X > c),\ P(X < d)$$

Физическое ~~Лирическое~~ вступление

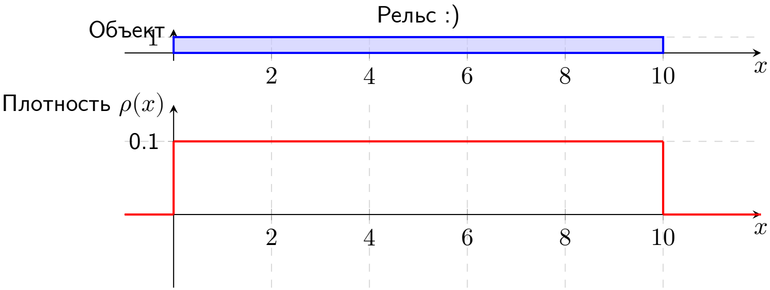
- Вспомним физику 8-го класса и формулу

$$m = \rho \cdot V, \quad [kg] = \left[\frac{kg}{m^3}\right][m^3]$$

- Можем перейти к линейной или погонной плотности $\rho_l = \left[\frac{kg}{m}\right]$, которая покажет нам массу единицы длины некого объекта (метр проволоки, метр провода, метр плитки шоколада). Тогда масса объекта длины L будет:

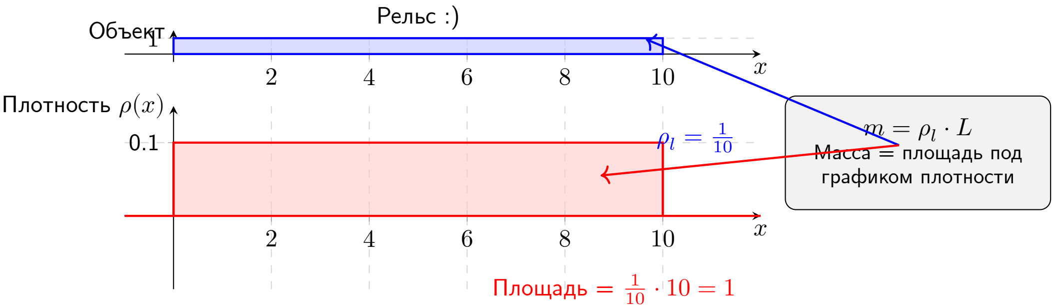
$$m = \rho_l \cdot L$$

Связь линейной плотности и массы



$$\rho(x) = \begin{cases} 0, & x < 0 \\ 0.1, & 0 \le x \le 10 \\ 0, & x > 10 \end{cases}$$

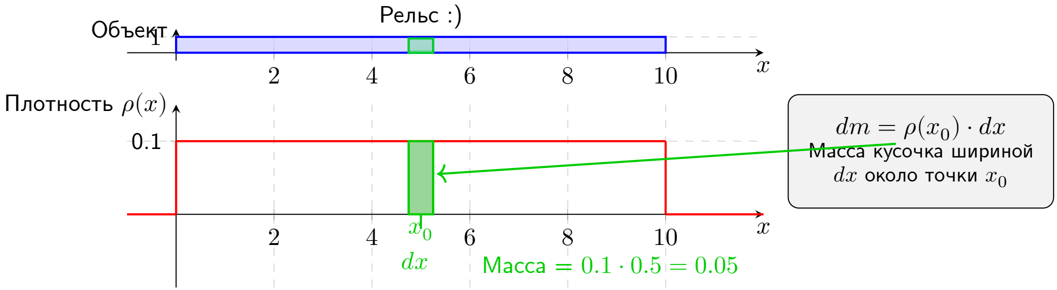
Связь линейной плотности и массы



Связь линейной плотности и массы

Значение плотности в точке

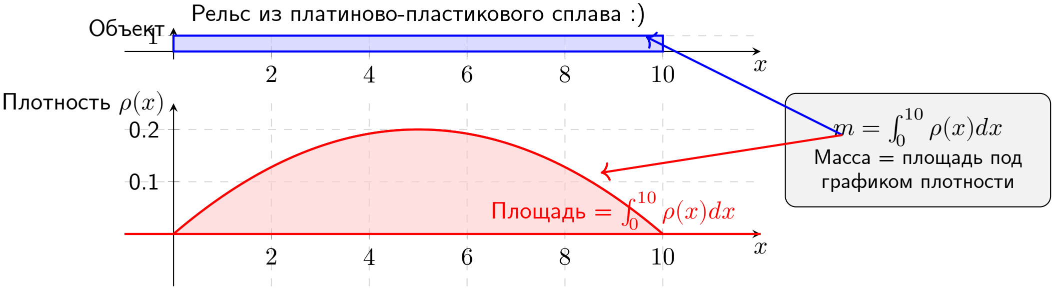
- Само по себе значение функции плотности $\rho(x_0)$ не обозначает массу в точке x_0. Смысл несет именно произведение плотности на длину, вспомним, чтобы сократились размерности $[kg] = \left[\frac{kg}{m}\right][m]$



Связь линейной плотности и массы

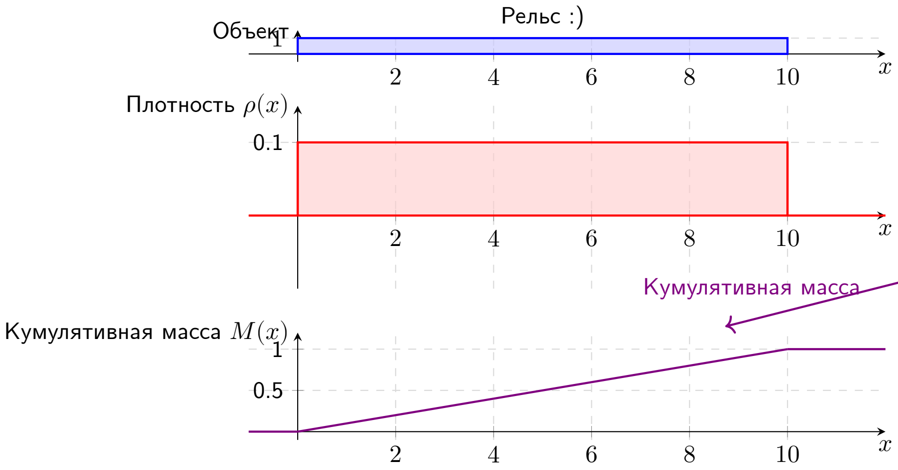
Переменная плотность

- Не всегда линейная плотность является постоянной! В нашей аналогии из физики мы можем себе представить рельс, у которого единица длины плавно становится тяжелее от краев к центру.



Кумулятивная масса

- Теперь мы хотим ввести функцию, которая покажет нам, сколько массы “скопилось” к координате x.



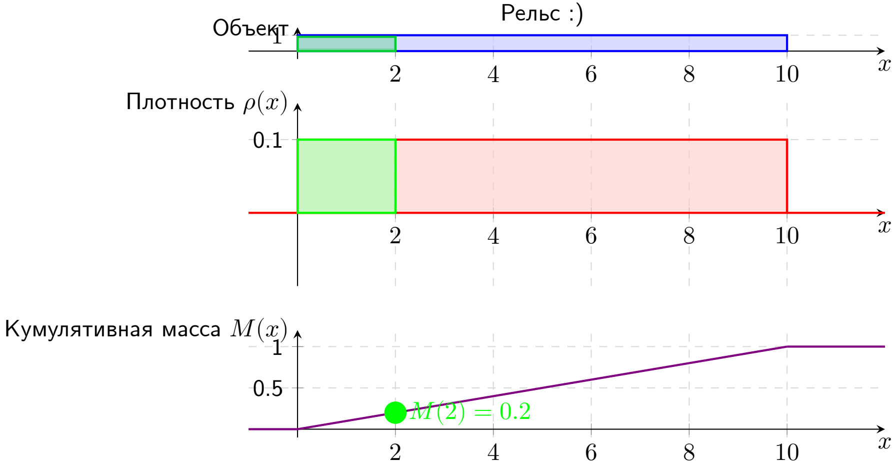
$$M(x) = \int_0^x \rho(t)dt$$

Кумулятивная масса

$$M(x) = \begin{cases} 0, & x < 0 \\ 0.1x, & 0 \le x \le 10 \\ 1, & x > 10 \end{cases}$$

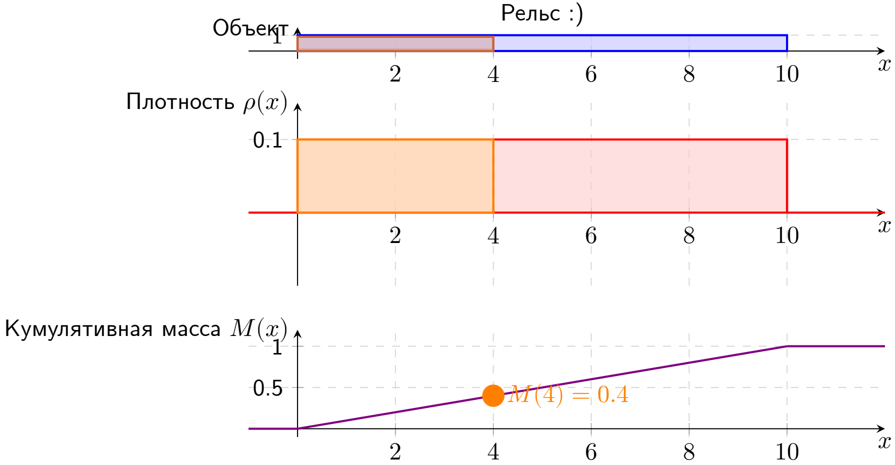
Кумулятивная масса

Пример 1



Кумулятивная масса

Пример 2



Функция плотности вероятности

- **Definition**

  Мы называем X непрерывной случайной величиной, если существует неотрицательная функция $f_X(x)$, определенная для $\forall x \in \mathbb{R}$, такая что любая вероятность вида $P(a \leq X \leq b)$ может быть найдена по формуле:

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx$$

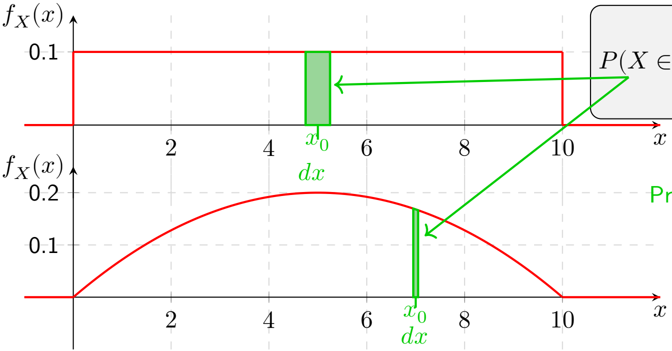
- Нормировка вероятности. Поскольку $P(\Omega_X)$ должно быть равно 1, здесь мы имеем:

$$1 = P(\Omega_X) = P\{X \in (-\infty, +\infty)\} = \int_{-\infty}^{+\infty} f_X(x)dx$$

Функция плотности вероятности

Значение плотности вероятности в точке

- По аналогии с физикой: само по себе значение функции плотности $f_X(x_0)$ не обозначает вероятность в точке x_0. Смысл несет именно произведение плотности на длину интервала.



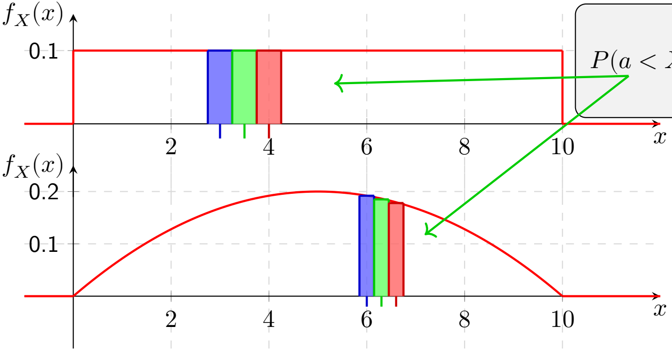
$$P(X \in (x_0 - \tfrac{dx}{2}, x_0 + \tfrac{dx}{2})) = f_X(x_0) \cdot dx$$

Probability = area = $f_X(x_0) \cdot dx$

Функция плотности вероятности

Вероятность попадания в интервал

- Последовательно складывая такие вероятности попадания в окрестность, получаем площадь под графиком функции плотности от точки a до точки b, которая и интерпретируется нами как $P(a < X < b)$.



$$P(a < X < b) = \sum f_X(x_i)dx = \int_a^b f_X(x)dx$$

Функция распределения (кумулятивная / интегральная)

- **Definition**

  Кумулятивная функция распределения случайной величины X - это неубывающая функция $F_X(x)$, определенная для $\forall x \in \mathbb{R}$, такая что:

$$F_X(x) = P\{X \in (-\infty, x]\} = P\{X \leq x\}$$

- Для непрерывной случайной величины мы можем переписать это как:

$$F_X(x) = P\{X \in (-\infty, x]\} = \int\limits_{-\infty}^{x} f_X(t)dt$$

- Основные свойства:

$$F_X(-\infty) = 0, \qquad F_X(\infty) = 1$$

Функция распределения как способ избежать интегрирования

- Предположим, нас интересует $P(a < X < b)$. Рассмотрим интервал $\mathcal{D} = (-\infty, b)$. Его можно разложить на объединение двух непересекающихся множеств: $\mathcal{D} = (-\infty,\ a] \cup (a,\ b)$.

- Согласно принципу аддитивности вероятности:

$$P\{X \in \mathcal{D}\} = P\{X \in (-\infty,\ a]\} + P\{X \in (a,\ b)\}$$

- В интегральной формулировке:

$$\int\limits_{-\infty}^{b} f_X(x)dx = \int\limits_{-\infty}^{a} f_X(x)dx + \int\limits_{a}^{b} f_X(x)dx$$

- Наконец:

$$\boxed{\int\limits_{a}^{b} f_X(x)dx = F_X(b) - F_X(a)}$$

Функция плотности вероятности является производной от функции распределения:

$$f_X(x) = F'_X(x) = \frac{d}{dx}F_X(x)$$

Пример: анализ функции плотности

- Пусть функция плотности случайной величины X задана в виде:

$$f_X(x) = \begin{cases} 0, & x < 0 \\ cx^2, & 0 \le x \le 3 \\ 0, & x > 3 \end{cases}$$

Найдите нормировочную константу, постройте функцию распределения, посчитайте вероятности $P(-5 < X < 2)$, $P(X > 1)$.

- Проверяем выполнение условия нормировки:

$$\int\limits_{-\infty}^{+\infty} f_X(x)dx = 1$$

$$\int\limits_{-\infty}^{0} {\color{red} f_X(x)}dx + \int\limits_{0}^{3} {\color{purple} f_X(x)}dx + \int\limits_{3}^{+\infty} {\color{blue} f_X(x)}dx = \int\limits_{-\infty}^{0} {\color{red} 0}\,dx + \int\limits_{0}^{3} {\color{purple} cx^2}dx + \int\limits_{3}^{+\infty} {\color{blue} 0}\,dx = 1$$

$$c\frac{x^3}{3}\Big|_0^3 = 9c = 1 \longrightarrow c = \frac{1}{9}$$

Пример: анализ функции плотности

- Используем формальное определение функции распределения: $F_X(x) = P(X \le x) = \int\limits_{-\infty}^{x} f_X(t)dt$
- Рассматриваем так же три области. Первая область $\forall x < 0$. Мы знаем, что функция плотности в этой области равна нулю.

$$\int\limits_{-\infty}^{x} f_X(t)dt = \int\limits_{-\infty}^{x} 0\,dt = 0$$

- Вторая область $\forall x \in [0,\, 3]$. Знаем, что на этом интервале у функции плотности определенный вид, плюс не забудем, что мы еще ранее нашли нормировочную константу.

$$\int\limits_{-\infty}^{x} f_X(t)dt = \int\limits_{-\infty}^{0} f_X(t)dt + \int\limits_{0}^{x} f_X(t)dt = \int\limits_{-\infty}^{0} 0\,dt + \int\limits_{0}^{x} \frac{1t^2}{9}dt = 0 + \frac{t^3}{27}\Big|_0^x = \frac{x^3}{27}$$

- Третья область $\forall x > 3$:

$$\int\limits_{-\infty}^{x} f_X(t)dt = \int\limits_{-\infty}^{0} f_X(t)dt + \int\limits_{0}^{3} f_X(t)dt + \int\limits_{3}^{x} f_X(t)dt = \int\limits_{-\infty}^{0} 0\,dt + \int\limits_{0}^{3} \frac{1t^2}{9}dt + \int\limits_{3}^{x} 0\,dt = 1$$

Пример: анализ функции плотности

- Финально, корректная запись функции распределения:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{x^3}{27}, & 0 \le x \le 3 \\ 1, & x > 3 \end{cases}$$

- Посчитаем вероятности через основное определение и через функцию распределения и сравним результаты.

$$P(-5 < X < 2) = \int\limits_{-5}^{2} f_X(x)dx = \int\limits_{-5}^{0} f_X(x)dx + \int\limits_{0}^{2} f_X(x)dx = \int\limits_{-5}^{0} 0\,dx + \int\limits_{0}^{2} \frac{1x^2}{9}dx = \frac{x^3}{27}\Big|_0^2 = \frac{8}{27}$$

$$P(-5 < X < 2) = F_X(2) - F_X(-5) = \frac{8}{27} - 0$$

$$P(X > 1) = \int\limits_{1}^{+\infty} f_X(x)dx = \int\limits_{1}^{3} f_X(x)dx + \int\limits_{3}^{+\infty} f_X(x)dx = \int\limits_{1}^{3} \frac{1x^2}{9}dx + \int\limits_{3}^{+\infty} 0\,dx = \frac{x^3}{27}\Big|_1^3 = \frac{26}{27}$$

$$P(X > 1) = F_X(+\infty) - F_X(1) = 1 - \frac{1}{27}$$