Математика для Data Science. Теория вероятностей. Шпаргалка

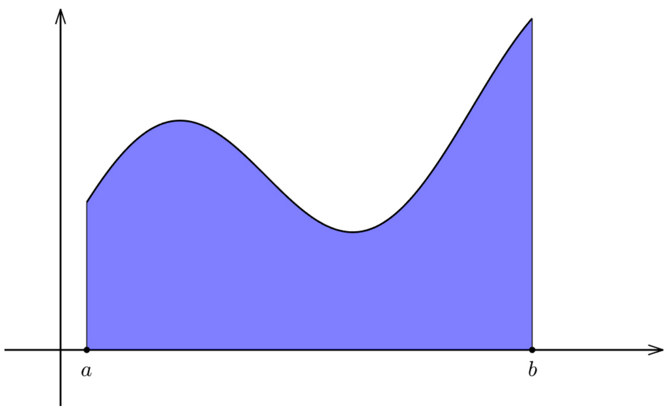
Содержание

Четвёртая неделя. Непрерывный случай

Определённый интеграл

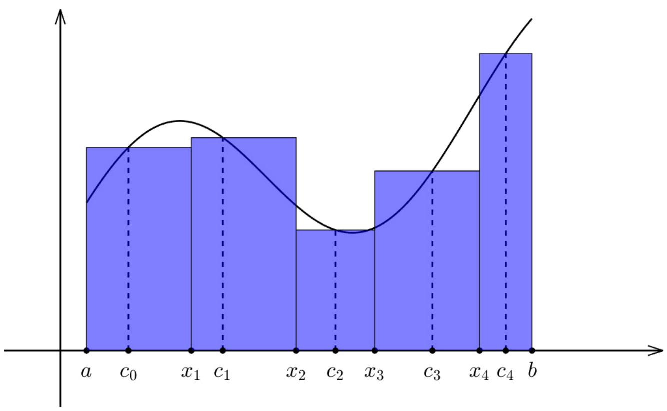
Пусть дана неотрицательная функция f со значениями в $\mathbb{R}$, определённая на отрезке $[a, b]$.

Давайте строго определим площадь фигуры, заключённой между графиком функции f, осью OX и вертикальными линиями $x = a, x = b$:



Выберем на отрезке $[a, b]$ несколько точек: $a = x_0 < x_1 < x_2 < \dots < x_k = b$. Будем называть это *разбиением* отрезка $[a, b]$.

Для каждого i на отрезке $[x_i, x_{i+1}]$ выберем произвольную точку c_i.



Интегральной суммой для такого разбиения и выбора точек c_i называется сумма $\sum\limits_{i=0}^{i=k-1} f(c_i)(x_{i+1} - x_i)$.

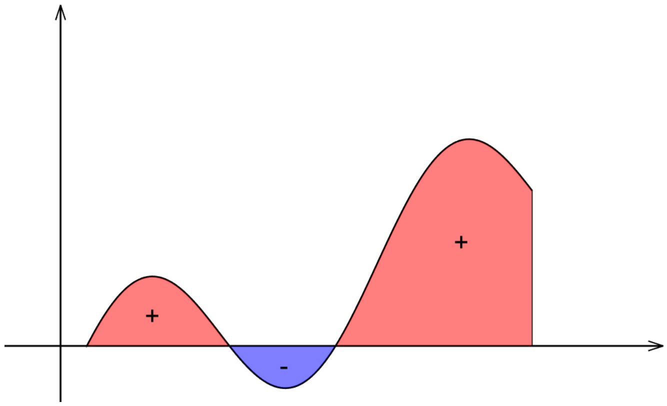
Рангом разбиения называется длина самого длинного из отрезков $[x_i, x_{i+1}]$, то есть $\max\limits_{0 \le i < k}(x_{i+1} - x_i)$.

Пусть для любой последовательности разбиений и выборов точек, такой что соответствующая последовательность рангов стремится к нулю, соответствующая последовательность интегральных сумм тоже имеет предел равный l.

Тогда l называется *определённым интегралом* функции f на отрезке $[a, b]$ и обозначается так: $\int\limits_a^b f(x)\,dx$.

Если определённый интеграл $\int\limits_a^b f(x)\,dx$ существует, то говорят, что f *интегрируема по Риману* на отрезке $[a, b]$.

Если же функция f не является неотрицательной, то определение интеграла никак не отличается. Но определённый интеграл функции f на отрезке $[a, b]$ будет равен площади фигуры, заключённой между графиком функции f, осью OX и вертикальными линиями $x = a, x = b$, где площадь фигур, расположенных ниже оси OX, считается со знаком минус:



Теорема. Для любой непрерывной функции f и любых $[a, b]$ определённый интеграл $\int\limits_a^b f(x)\,dx$ существует.

Следствие. Пусть у нас есть непрерывная функция f, определённая на отрезке $[a, b]$, и какая-то последовательность разбиений (и выборов c_i) с рангом, стремящемся к нулю. Тогда предел интегральных сумм этой последовательности существует и равен $\int\limits_a^b f(x)\,dx$.

Неопределённый интеграл

Неопределённым интегралом функции $f(x)$ называется любая функция $F(x)$, такая что производная $F(x)$ равна $f(x)$. То есть $F'(x) = f(x)$.

Эта функция $F(x)$ ещё называется *первообразной* функции $f(x)$.

Теорема (формула Ньютона-Лейбница). Пусть f – непрерывная функция, определённая на отрезке $[a, b]$. Пусть F – какая-то первообразная функции f. Тогда выполнено:

$$\int\limits_a^b f(x)\,dx = F(b) - F(a).$$

Обозначение. Часто выражение $F(b) - F(a)$ обозначают так: $F(x)\Big|_a^b$.

Линейность интеграла:

1. $\int\limits_a^b cf(x)\,dx = c\int\limits_a^b f(x)\,dx$ для любого $c \in \mathbb{R}$.
2. $\int\limits_a^b f(x)\,dx + \int\limits_a^b g(x)\,dx = \int\limits_a^b f(x) + g(x)\,dx$.

Несобственные итегралы первого рода

Если f это непрерывная функция на луче $[a, +\infty)$, то $\int\limits_a^{+\infty} f(x)\,dx := \lim\limits_{b\to+\infty} \int\limits_a^b f(x)\,dx$.

Запись $\lim\limits_{b\to+\infty}$ в правой части означает, что мы берём пределы для всевозможных последовательностей $b_0, b_1, b_2 \ldots$, стремящихся к $+\infty$, и все эти пределы совпадают. Если пределы не совпадают или их нет, то $\int\limits_a^{+\infty} f(x)\,dx$ не определён. В этом случае мы говорим, что интеграл расходится.

Если f это непрерывная функция на луче $(-\infty, b]$, то $\int\limits_{-\infty}^b f(x)\,dx := \lim\limits_{a\to-\infty} \int\limits_a^b f(x)\,dx$.

Если f это непрерывная функция на прямой $(-\infty, +\infty)$, то $\int\limits_{-\infty}^{+\infty} f(x)\,dx := \int\limits_{-\infty}^{c} f(x)\,dx + \int\limits_{-c}^{+\infty} f(x)\,dx$, где c это любое число.

Функция распределения

Для случайной величины ξ функция $F_\xi(x) := P(\xi \leqslant x)$ называется *функцией распределения.*
Функция F_ξ называется *функцией распределения*, если выполнены такие свойства:

- F_ξ не убывает, то есть для любых $a_1 < a_2$ выполнено $F_\xi(a_1) \leq F_\xi(a_2)$
- $\lim\limits_{a \to +\infty} F_\xi(a) = 1$
- $\lim\limits_{a \to -\infty} F_\xi(a) = 0$

Случайная величина называется *непрерывной*, если её функция распределения непрерывна в каждой точке.

Плотность вероятности

Гистограмма строится так:

1. Промежуток значений, которое может принимать измеряемая величина, разбивается на несколько интервалов — по-английски их называют *bins*, по-русски — карманы / корзины. Чаще всего эти интервалы берут одинаковыми.
2. Отложим полученные интервалы на горизонтальной оси. Над каждым карманом изобразим прямоугольник с высотой равной количеству участников, чей рост попал в данный карман.

Случайная величина ξ с функцией распределения F_ξ называется *абсолютно непрерывной*, если существует функция $p_\xi : \mathbb{R} \to \mathbb{R}_{\geqslant 0}$ такая, что для всех $x \in \mathbb{R}$ выполнено $F_\xi(x) = \int\limits_{-\infty}^{x} p_\xi(x)dx$. В этом случае функция p_ξ называется *плотностью вероятности.*

Кроме того верны формулы

- $F'_\xi(a) = p_\xi(a)$
- $P(\xi \in [a, b]) = \int\limits_{a}^{b} p_\xi(x)dx.$

Математическое ожидание

Пусть ξ — абсолютно непрерывная случайная величина с плотностью $p(\xi)$. Если интеграл $\int\limits_{-\infty}^{+\infty} xp_\xi(x)dx$ сходится, то его значение называют *математическим ожиданием* ξ. В противном случае говорят, что математическое ожидание не определено.

В случае абсолютно непрерывной случайной величины свойство линейности математического ожидания сохраняется, то есть соотношение $\forall a, b \in \mathbb{R} : \; E[a\xi + b\eta] \; = aE[\xi] + bE[\eta]$ выполнено, с оговоркой, что математическое ожидание случайных величин ξ и η должно быть определено.

Дисперсия

Пусть ξ — абсолютно непрерывная случайная величина с плотностью p_ξ. Если определено математическое ожидание ξ и интеграл $\int\limits_{-\infty}^{+\infty} x^2 \, p_\xi(x)dx$ сходится, то *дисперсия* $Var(\xi)$ определена и равняется $\int\limits_{-\infty}^{+\infty} x^2 \, p_\xi(x)dx - (E\xi)^2$.