Математика для Data Science. Теория вероятностей. Шпаргалка

Содержание

Первая неделя. Вероятностное пространство

Вероятностное пространство

Вероятностное пространство это тройка (Ω, F, P), где

1. Ω это любое множество. Оно называется *множеством элементарных исходов*, а его элементы – *элементарными исходами*. Мы пока будем заниматься только случаем, когда Ω это конечное множество. Оставшаяся часть определения написана только для этого случая.

2. Все подмножества Ω называются *событиями*. Множество событий обозначается F и называется *алгеброй событий*.

   Обозначение. Для обозначения событий мы будем использовать заглавные буквы из начала латинского алфавита: A, B, C, и так далее.

3. *Вероятность* P это функция из F в $[0, 1]$. Другими словами, P каждому событию сопоставляет число от 0 до 1. Это число называется *вероятностью* соответствующего события. Вероятность каждого события должна быть равна сумме вероятностей элементарных исходов, из которых состоит это событие. Сумма вероятностей всех элементарных исходов должна быть равна 1.

Обозначение. Если A это конечное множество, то мы обозначаем число элементов A через $|A|$.
Если число исходов в Ω равно k, то число событий равно 2^k, то есть $|\Omega| = k, k \in \mathbb{N} \implies |F| = 2^k$.
Пусть дано событие A. Тогда событием $\bar{A}$ называется событие, состоящие из всех элементарных исходов, которые не входят в A. Можно произносить $\bar{A}$ как "не A."Верно следующее равенство: $P(\bar{A}) = 1 - P(A)$.

Равновероятные исходы

Если вероятности всех элементарных исходов равны, то есть $P(a) = P(b)$ для всех $a, b \in \Omega$, то в этом случае мы говорим, что все элементарные исходы *равновероятны*.

Если все исходы равновероятны, то для любого события A выполнено $P(A) = \dfrac{|A|}{|\Omega|}$. В частности, вероятность каждого элементарного исхода равна $\frac{1}{k}$.

Условная вероятность и независимые события

Так как события являются множествами (состоящими из элементарных исходов), к ним можно применять все стандартные операции над множествами. В частности, события можно пересекать (символ $\cap$) и объединять (символ $\cup$).

Если $P(B) \neq 0$, то *условной вероятностью события A при условии события B* называется дробь $\dfrac{P(A \cap B)}{P(B)}$.
Обозначение:

$$P(A|B) := \frac{P(A \cap B)}{P(B)}.$$

Формула полной вероятности для двух событий:

$$P(B) = P(B \cap A) + P(B \cap \bar{A}) \text{ или, равносильно,}$$
$$P(B) = P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A})$$

Независимые события

Пусть $P(B) \neq 0$. Событие A называется *независимым от события B*, если и только если $P(A) = P(A|B)$.
Пусть $P(A) \neq 0$ и $P(B) \neq 0$. Тогда верна следующая цепочка эквивалентных утверждений:

$$\text{События } A \text{ и } B \textit{ независимы} \Leftrightarrow P(A) = P(A|B) \Leftrightarrow P(B) = P(B|A) \Leftrightarrow P(A \cap B) = P(A) \cdot P(B)$$

Два события, которые не являются независимыми, называют *зависимыми*.

Совместная независимость

События $A_1, \ldots, A_n$ называются *совместно независимыми*, если для любого k и любого набора индексов $1 \le i_1 < i_2 < \cdots < i_k \le n$ выполнено

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \cdots \cdot P(A_{i_k}).$$

События называются *попарно независимыми*, если любые два из них независимы.

Из совместной независимости событий следует их попарная независимость. Но в общем случае из попарной независимости событий не следует совместная независимость.

Вторая неделя. Комбинаторика и случайные величины

Теорема Байеса

Теорема Байеса. Для любых событий A и B, таких что $P(B) \neq 0$ выполнено

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}.$$

Формула полной вероятности.

Даны события $A_1, A_2, \ldots, A_n$, такие что каждый элементарный исход из Ω лежит ровно в одном из этих событий. Другими словами,

- эти события не пересекаются друг с другом, то есть $A_i \cap A_j = \emptyset$ для любых $i \neq j$
- и их объединение равно Ω, то есть $\bigcup\limits_{i=1}^{n} A_i = \Omega$.

Тогда выполнено:

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) \text{ или, эквивалентно,}$$

$$P(B) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i).$$

Правило суммы

События $A, B \in F$ называются *несовместными*, если $A \cap B = \emptyset$.

Правило суммы. Если события A и B несовместны, то $P(A \cup B) = P(A) + P(B)$.

Утверждение. Пусть $A, B \in F$ — события. Тогда

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Комбинаторика

Перестановка чисел от 1 до n — это некоторая упорядоченная последовательность чисел от 1 до n, где каждое число встречается ровно один раз.

Ещё одно определение перестановки — функция из множества $\{1, 2, \ldots, n\}$ в множество $\{1, 2, \ldots, n\}$, такая, что значение функции для двух различных чисел не может совпадать.

$n! := 1 \cdot 2 \cdot 3 \cdot 4 \cdot \cdots \cdot n$. Читается "$n$ факториал". При этом $0! := 1$.

Различных перестановок чисел от 1 до n всего $n!$

Биномиальные коэффициенты

Число $\frac{n!}{k!(n-k)!}$ обозначается $\binom{n}{k}$ и читается "n по k". Определено для целых неотрицательных n и k таких, что $n \geqslant k$. Все числа такого вида называются *биномиальными коэффициентами*. В русскоязычной литературе также используется обозначение C_n^k (цэ из n по k).

$\binom{n}{k} = \frac{n!}{k!(n-k)!}$ равно числу подмножеств размера k у n-элементного множества S.

Поскольку $0! := 1$, то $\binom{n}{k} = 1$.

Бином Ньютона. Для любых $a, b \in \mathbb{R}$ и $n \in \mathbb{N}$ выполнено $(a+b)^n = \sum\limits_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$.

Свойства биномиальных коэффициентов

1. Для любых n и k таких, что $k \leqslant n$ выполнено соотношение $\binom{n}{k} = \binom{n}{n-k}$
2. Для всех n выполнено

$$\sum_{k=0}^{n} \binom{n}{k} = \binom{n}{n} + \binom{n}{n-1} + \binom{n}{n-2} + \cdots + \binom{n}{k} + \cdots + \binom{n}{n-1} + \binom{n}{0} = 2^n$$

3. Для всех n и k таких, что $k \leqslant n$ выполнено $\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}$

Биномиальные коэффициенты можно визуализировать с помощью *треугольника Паскаля*, который получается так

- Первая строка состоит из одной 1
- Каждая следующая получается из предыдущей строки сложением её с самой собой со сдвигом 1

0							1							
1						1		1						
2					1		2		1					
3				1		3		3		1				
4			1		4		6		4		1			
5		1		5		10		10		5		1		
6	1		6		15		20		15		6		1	

Случайная величина и математическое ожидание

Как и раньше, рассматриваем случай, когда пространство исходов Ω конечно, и алгебра событий F состоит из всевозможных подмножеств Ω.

Случайная величина это функция из пространства исходов Ω в $\mathbb{R}$.

Пусть дано вероятностное пространство Ω, состоящее из n элементарных исходов, и случайная величина X. Обозначим через P_i вероятность i-ого исхода, и через x_i значение случайной величины X на i-ом исходе. Тогда *математическим ожиданием* случайной величины X называется число $E[X] := \sum\limits_{i=1}^{n} x_i P_i$.

Более коротко определение математического ожидания можно записать так: $E[X] := \sum\limits_{\omega \in \Omega} X(\omega)P(\omega)$.

Операции со случайными величинами

Случайные величины, определённые на одном вероятностном пространстве, можно складывать друг с другом, умножать друг на друга, как и любые функции. Например, сумма случайных величин X и Y обозначается $X+Y$ и определяется так: для любого исхода $\omega \in \Omega$ мы говорим $(X+Y)(\omega) := X(\omega) + Y(\omega)$.

Пусть дано число $c \in \mathbb{R}$ и пространство исходов Ω. Будем считать, что случайная величина, которая любому элементарному исходу сопоставляет число c, обозначается так же, как и число c. Тогда мы можем определить случайную величину, равную многочлену от случайных величин.

В дальнейшем, когда мы будем складывать или перемножать случайные величины, мы всегда будем предполагать, что они определены на одном и том же вероятностном пространстве.

Свойства математического ожидания

Пусть X и Y — случайные величины. Тогда

1. если $c \in \mathbb{R}$, то $E[cX] = c \cdot E[X]$;
2. $E[X+Y] = E[X] + E[Y]$;
3. $E[X \cdot Y]$ <u>не всегда равно</u> $E[X] \cdot E[Y]$

Третья неделя. Распределения, дисперсия и счётное пространство исходов

Распределения случайных величин

Функция $p_X : \mathbb{R} \to [0, 1]$, заданная условием $p_X(a) := P(X = a)$ для любого $a \in R$, называется *функцией вероятности случайной величины* X.

Мы говорим, что функция вероятности случайной величины X задаёт *распределение* X.

Две случайные величины с совпадающими функциями вероятности называются *одинаково распределёнными.*

Независимые случайные величины

Случайные величины X и Y *независимы*, если для любых $a \in \mathbb{R}, b \in \mathbb{R}$ события $X = a$ и $Y = b$ независимы. То есть:

$$\forall a, b \in \mathbb{R} : P(X = a \cap Y = b) = P(X = a) \cdot P(X = b).$$

Если случайные величины X и Y независимы, то $E[X \cdot Y] = E[X]E[Y]$.

Если случайные величины совместно независимы и одинаково распределены, то слово "совместно" часто опускают, и говорят просто "независимые и одинаково распределённые". В английском это звучит как "independent and identically distributed", что сокращают до i.i.d.

Пусть даны два вероятностных пространства с пространствами исходов Ω_1 и Ω_2, и функциями вероятности P_1 и P_2 соответственно. Тогда их *произведение* это вероятностное пространство $\Omega_1 \times \Omega_2$, определённое так:

- исходы в этом вероятностном пространстве это всевозможные пары (ω_1, ω_2) с $\omega_1 \in \Omega_1, \omega_2 \in \Omega_2$,
- вероятности соответствующих исходов равны $P(\omega_1, \omega_2) := P_1(\omega_1) \cdot P_2(\omega_2)$

Аналогично можно построить произведение любого числа вероятностных пространств.

Дисперсия

Дисперсия случайной величины X это число $E[(X - E[X])^2]$, обозначаемое $Var[X]$.
Эквивалентная формула дисперсии такая: $Var[X] = E[X^2] - E[X]^2$.

Свойства дисперсии

1. Для любой случайной величины X выполнено $Var[X] \geq 0$
2. $Var[X] = 0$ если и только если X это постоянная случайная величина
3. Если X и Y это <u>независимые</u> случайные величины, то $Var[X + Y] = Var[X] + Var[Y]$.
4. Если $c \in \mathbb{R}$ и X — случайная величина, то $Var[cX] = c^2Var[X]$.

Биномиальное распределение и стандартное отклонение

Стандартным отклонением случайной величины X называется $\sqrt{Var[X]}$.

Пусть $X = 1$ с вероятностью p и $X = 0$ с вероятностью $1 - p$. Такое распределение случайной величины называется *распределением Бернулли* с вероятностью успеха p.

Пусть $X_1, \ldots, X_n$ независимы и имеют распределение Бернулли с вероятностью успеха p. Тогда распределение случайной величины $S := X_1 + \cdots + X_n$ называется *биномиальным распределением с* n *степенями свободы.*

Биномиальное распределение обозначается $Bin(n, p)$. Фразу "S имеет биномиальное распределение с n степенями свободы" записывают так: $S \sim Bin(n, p)$.

Ряды

Рядом называется выражение вида $a_1+a_2+a_3+\ldots$, где $\{a_n\}$ это последовательность вещественных чисел. Также используется запись $\sum\limits_{n=1}^{\infty} a_n := a_1 + a_2 + a_3 + \ldots$.
Числа a_n называются *членами* ряда.
Частичными суммами ряда называются такие выражения:

- $S_1 := a_1$
- $S_2 := a_1 + a_2$
- $S_3 := a_1 + a_2 + a_3$

  ...
- $S_n = a_1 + a_2 + a_3 + \cdots + a_n$

  ...

Суммой ряда называется предел частичных сумм, то есть $\lim\limits_{n\to\infty} S_n$.
Ряд называется *сходящимся*, если предел $\lim\limits_{n\to\infty} S_n$ существует, и *расходящимся* в противном случае.
Например, ряд $\sum\limits_{n=1}^{\infty} \frac{1}{n}$ расходится. Этот ряд называют *гармоническим.*

Если ряд состоит из неотрицательных чисел и при этом все частичные суммы меньше некоторого числа B, то ряд сходится.

Свойства сходящихся рядов

1. **Необходимое условие сходимости ряда.** Пусть $\sum\limits_{n=1}^{\infty} a_n$ сходится. Тогда $\lim\limits_{n\to\infty} a_n = 0$.
2. Пусть $\sum\limits_{n=1}^{\infty} a_n$ и $\sum\limits_{n=1}^{\infty} b_n$ — два сходящихся ряда. Тогда
   - ряд $\sum\limits_{n=1}^{\infty} (a_n + b_n)$ сходится
   - ряд $\sum\limits_{n=1}^{\infty} (a_n - b_n)$ сходится
   - ряд $\sum\limits_{n=1}^{\infty} (ca_n)$ сходится для любого $c \in \mathbb{R}$
3. Если $0 \le a_n \le b_n$ для всех n, то говорят, что ряд $\sum\limits_{n=1}^{\infty} b_n$ *мажорирует* ряд $\sum\limits_{n=1}^{\infty} a_n$. Если, кроме того
   - $\sum\limits_{n=1}^{\infty} b_n$ сходится, то $\sum\limits_{n=1}^{\infty} a_n$ сходится
   - $\sum\limits_{n=1}^{\infty} a_n$ расходится, то $\sum\limits_{n=1}^{\infty} b_n$ расходится

Абсолютно сходящиеся ряды

Ряд $\sum\limits_{n=1}^{\infty} a_n$ назвается *абсолютно сходящимся*, если сходится ряд $\sum\limits_{n=1}^{\infty} |a_n|$.
Теорема. Любой абсолютно сходящийся ряд является сходящимся.
Ряд, который сходится, но не сходится абсолютно, называют *условно сходящимся.*

Теорема. Если ряд абсолютно сходится к сумме S, то любой ряд, полученный из него перестановкой слагаемых, тоже абсолютно сходится к той же сумме S.

Теорема Римана. Если ряд сходится условно, то его слагаемые можно переставить так, чтобы полученный ряд сходился к любому заранее заданному числу $c \in \mathbb{R}$.

Счётное пространство исходов

Множество A называется *счётным*, если существует функция $f : \mathbb{N} \to A$, такая что для любого $a \in A$ найдётся ровно одно $n \in N$, что $f(n) = a$.

Примеры счётных множеств: $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$

Примеры несчётных множеств: $\mathbb{R}, [0,1], \mathbb{R}^n, \mathbb{C}$

Когда говорят "множество несчётно", имеют в виду, что множество не является счётным или конечным.

Пусть есть счётное пространство исходов $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$. Как и в случае с конечным пространством исходов, будем называть событием любое подмножество Ω.

Пусть P_n — вероятность события, состоящего ровно из одного исхода ω_n. Потребуем, чтобы $P_n \geq 0$ и ряд $\sum\limits_{n=1}^{\infty} P_n$ сходился к 1. Вероятность каждого события как сумму вероятностей исходов, из которых это событие состоит. Эта сумма может быть бесконечной, то есть суммой ряда. Как и в случае конечного числа исходов, *случайная величина* X это функция из пространства исходов в $\mathbb{R}$, то есть $X : \Omega \to \mathbb{R}$.

Пусть дано вероятностное пространство Ω, состоящее из счётного количества элементарных исходов, и случайная величина X. Обозначим через P_i вероятность i-ого исхода, и через x_i значение случайной величины X на i-ом исходе. Если ряд $\sum\limits_{i=1}^{\infty} x_i P_i$ абсолютно сходится, его сумма называется *математическим ожиданием* случайной величины X и обозначается $E[X]$. Если ряд не сходится абсолютно, то математическое ожидание не определено.

Пусть дано вероятностное пространство Ω, состоящее из счётного количества элементарных исходов, и случайная величина X. Предположим, что $E[X]$ определено. Тогда *дисперсией* называется математическое ожидание случайной величины $(X - E[X])^2$, если это математическое ожидание определено. Если $E[X]$ или $E[(X - E[X])^2]$ не определено, то $Var(X)$ не определена.

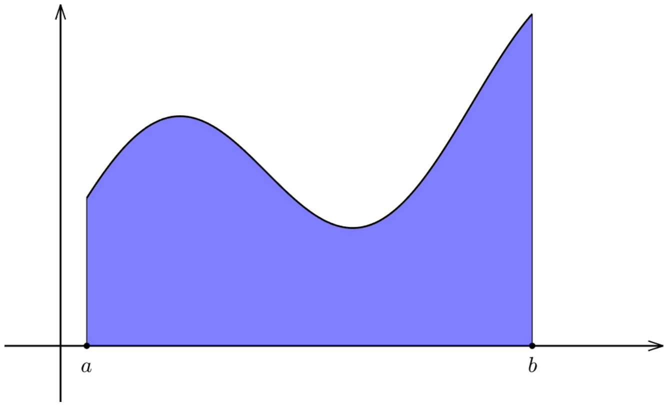
Вероятностные пространства с конечным или счётным количеством исходов называются *дискретными*. Аналогично, случайная величина, которая принимает конечное или счётное количество разных значений, называется *дискретной*.

Четвёртая неделя. Непрерывный случай

Определённый интеграл

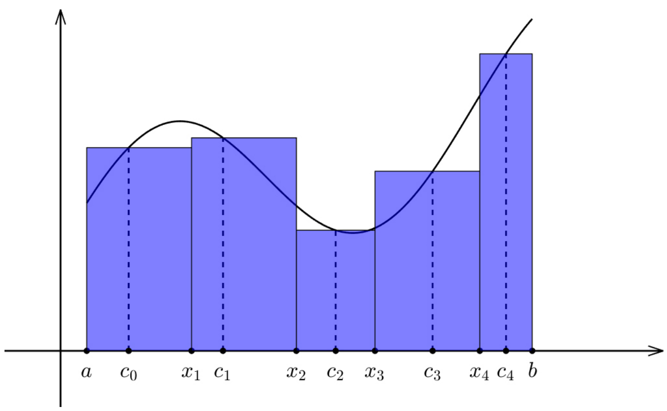
Пусть дана неотрицательная функция f со значениями в $\mathbb{R}$, определённая на отрезке $[a, b]$.

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: $a = x_0 < x_1 < x_2 < \cdots < x_k = b$. 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 $\sum\limits_{i=0}^{i=k-1} f(c_i)(x_{i+1} - x_i)$.

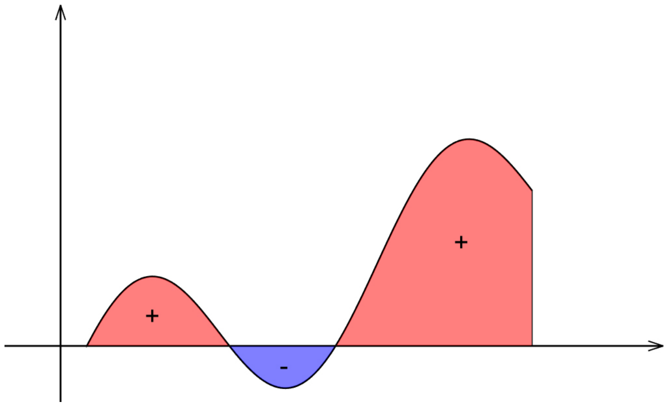
Рангом разбиения называется длина самого длинного из отрезков $[x_i, x_{i+1}]$, то есть $\max\limits_{0 \le i < k}(x_{i+1} - x_i)$.

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 $\forall a, b \in \mathbb{R} : \quad E[a\xi + b\eta] \quad = aE[\xi] + bE[\eta]$ 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 $\int\limits_{-\infty}^{+\infty} x^2\ p_\xi(x)dx - (E\xi)^2$.

Пятая неделя. Статистика

Арифметика случайных величин и нормальное распределение

Следующие формулы выполнены как для дискретных, так и для непрерывных случайных величин:

Математическое ожидание суммы. Математическое ожидание суммы двух случайных величин это сумма их математических ожиданий. То есть

$$E[X+Y] = E[X] + E[Y].$$

Дисперсия суммы. Если две случайных величины независимы, то дисперсия их суммы это сумма дисперсий. То есть

$$Var(X+Y) = Var(X) + Var(Y).$$

Сложение с константой. Для любой случайной величины X и любого числа $c \in \mathbb{R}$ выполнено

$$E[X+c] = E[X] + c$$

и

$$Var(X+c) = Var(X).$$

Умножение на константу. Для любой случайной величины X и любого числа $c \in \mathbb{R}$ выполнено

$$E[cX] = cE[X]$$

и

$$Var(cX) = c^2 Var(X).$$

Нормальное распределение

Нормальное распределение с математическим ожиданием 0 и дисперсией 1 обозначается $N(0,1)$. Оно имеет такую функцию плотности распределения:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{1}{2}\ x^2\right)}$$

Нормальное распределение с математическим ожиданием μ и дисперсией σ^2 обозначается $N(\mu, \sigma^2)$. У него такая функция плотности распределения:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Случайная величина, имеющая распределение $N(\mu, \sigma^2)$ для каких-то μ и σ, называется *нормально распределённой.*

Правило двух и трех сигм. Пусть случайная величина ξ имеет нормальное распределение со средним μ и дисперсией σ^2. Тогда:

- $P(\mu - \sigma < \xi < \mu + \sigma\ ) \ = 0.682\cdots \approx 0.68$
- $P(\mu - 2\sigma < \xi < \mu + 2\sigma\ ) = 0.954\cdots \approx 0.95$. Другими словами, вероятность получить результат, отклоняющийся от μ хотя бы на 2σ, меньше 0.05.
- $P(\mu - 3\sigma < \xi < \mu + 3\sigma\ ) = 0.997\cdots \approx 0.99$. Другими словами, вероятность получить результат, отклоняющийся от μ хотя бы на 3σ, меньше 0.01.

Утверждение. Сумма нескольких совместно независимых нормально распределённых случайных величин – это тоже нормально распределенная случайная величина.

Статистический тест

Реализация случайной величины – это конкретное число, которым стала эта случайная величина после измерения.

Шаблон статистических тестов

1. **Выборка.** Выборка это реализация набора случайных величин $x_1, \ldots, x_n$, то есть это n чисел. Обычно предполагают, что случайные величины $x_1, \ldots, x_n$ совместно независимы и имеют одинаковое распределение.

2. **Гипотезы и предположения.** Выбор гипотез H_0 и H_1, то есть сформулировать свой вопрос на языке теории вероятностей.

3. **Статистика.** Нам нужно как-то объединить величины $x_1, \ldots, x_n$, в одну случайную величину $T(x_1, \ldots, x_n)$. Эту величину называют *статистикой*. При условии что H_0 выполнена, нужно найти распределение случайной величины $T(x_1, \ldots, x_n)$.

4. **Уровень значимости.** *Уровень значимости* это число α отвечающее за вероятность *ошибки первого рода*. То есть за вероятность отвергнуть H_0 в случае, когда H_0 выполнена. Обычно берут $\alpha = 0.05$.

5. **Критическое множество.** Случайная величина $T(x_1, \ldots, x_n)$ принимает значения в $\mathbb{R}$. Нужно выделить подмножество $C_\alpha \subset \mathbb{R}$, по которому мы будем решать, принимать или отвергать H_0. Вероятность попадания T в множество C_α должна быть равна α. Обычно в качестве C_α берут множества вида:

   - $[a, +\infty)$ – если отклонение статистики T вверх свидетельствует в пользу H_1.
   - $(-\infty, b]$ – если отклонение статистики T вниз свидетельствует в пользу H_1.
   - $(-\infty, b] \cup [a, +\infty)$ – если отклонение статистики T от какого-то значения в любую сторону свидетельствует в пользу H_1.

6. **Статистический критерий.** Если реализация T не попала в множество C_α, то мы принимаем H_0. Если T попала в множество C_α, то мы отвергаем H_0 и принимаем H_1.

Ошибка первого рода (принять H_1 при верной H_0). Вероятность допустить ошибку первого рода всегда равна уровню значимости α. Это следует из нашего построения множества C_α.

Ошибка второго рода (принять H_0 при верной H_1). Найдём распределение T при условии, что выполнена H_1. Вероятность того, что так распределённая T не попала в критическое множество, это и есть вероятность ошибки второго рода. Другими словами $\beta = P(T \notin C_\alpha | H_1)$. Действительно, условие $T \notin C_\alpha$ как раз соответствует тому, что мы приняли H_0 и отвергли H_1.

Число $(1 - \beta)$ называют *мощностью* статистического критерия.

ЗБЧ и ЦПТ

Теорема [ЗБЧ]. Пусть $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ — бесконечная последовательность независимых одинаково распределённых случайных величин, имеющих конечное мат.ожидание μ. Обозначим среднее арифметическое первых n случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ так:

$$\bar{\xi}_n = \frac{1}{n} \sum_{i=1}^{n} \xi_i.$$

Тогда

$$\bar{\xi}_n \xrightarrow[n \to +\infty]{\text{по вероятности}} \mu.$$

То есть $\forall \varepsilon > 0$ выполнено

$$\lim_{n \to +\infty} P(|\bar{\xi}_n - \mu| > \varepsilon) = 0$$

Теорема [ЦПТ]. Пусть $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ — бесконечная последовательность независимых одинаково распределённых случайных величин, имеющих конечное мат.ожидание μ и дисперсию σ^2.

Тогда

$$\frac{\bar{\xi}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow[n\to+\infty]{\text{по распределению}} N(0,1)$$

где $N(0,1)$ — нормальное распределение со средним 0 и дисперсией 1.

Обозначение. (неформально) Стрелка

$$\eta_n \xrightarrow[n\to+\infty]{\text{по распределению}} F$$

означает, что при n стремящемся к плюс бесконечности распределение случайной величины η_n близко к распределению F.

Неформально ЦПТ можно сформулировать так:

$$\bar{\xi}_n \quad \xrightarrow[n\to+\infty]{\text{в неформальном смысле}} \quad N\left(\mu, \frac{\sigma^2}{n}\right).$$