Математические основы искусственного интеллекта
Непрерывные случайные величины. Функция и плотность распределения. Числовые характеристики случайных величин

Солодушкин Святослав Игоревич

Кафедра вычислительной математики и компьютерных наук,
УрФУ имени первого Президента России Б.Н. Ельцина

Октябрь 2021

Непрерывное распределение

Определение

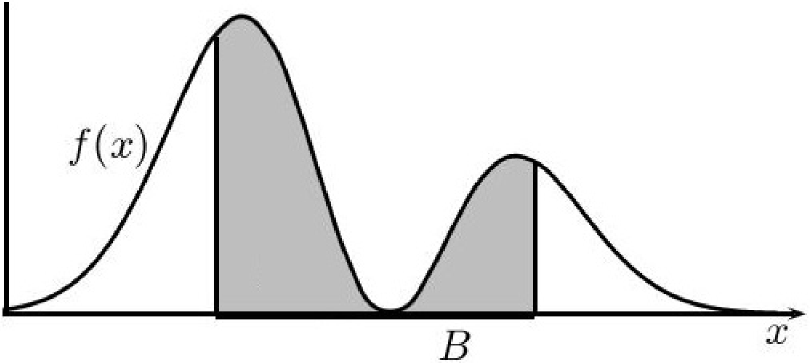
Случайная величина ξ имеет абсолютно непрерывное распределение, если существует неотрицательная функция $f_\xi(x)$ такая, что для любого борелевского множества B имеет место равенство:

$$\mathrm{P}(\xi \in B) = \int\limits_B f_\xi(x)\,dx.$$

Функцию $f_\xi(x)$ называют плотностью распределения величины ξ.

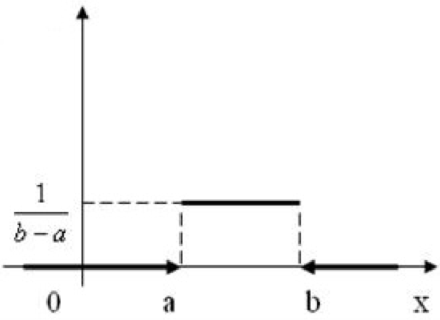
Замечание. Здесь интеграл понимается в смысле Лебега, а не Римана.

График непрерывного распределения



Равномерное распределение на отрезке [2, 4]

$$f_\xi(x) = \begin{cases} 0, & \text{если } x < 2, \\ 1/2, & \text{если } x \in [2,\, 4], \\ 0, & \text{если } x > 4. \end{cases}$$



Непрерывное распределение. Пример

Распределение может быть неравномерным

$$f_\xi(x) = \begin{cases} 0, & \text{если } x < 0, \\ 1/2 \sin x, & \text{если } x \in [0,\ \pi], \\ 0, & \text{если } x > \pi. \end{cases}$$

Вопрос. Можно ли взять вместо коэффициента 1/2 другой? Например, 2 или 1/3.

Теорема

Плотность распределения обладает свойствами:

(f1) $f_\xi(x) \geq 0$ для любого x;

(f2) $\displaystyle\int\limits_{-\infty}^{\infty} f_\xi(t)\,dt = 1.$

Вопрос. Каков вероятностный смысл свойства (f2)?
Ответ. Веротяность того, что случайная величина ξ примет хоть куда-нибудь значение, равна 1.

Как упростить описание распределения случайной величины?

Описание распределения набором вероятностей $\mathsf{P}(\xi \in B)$ весьма не удобно, так как существует очень много борелевских множеств. Мы описали дискретные распределения таблицей распределения, абсолютно непрерывные — плотностью распределения. Попробуем поискать какой-нибудь универсальный способ описать любое возможное распределение.
Можно поставить вопрос иначе: распределение есть набор вероятностей попадания в любые борелевские множества на прямой. Нельзя ли обойтись знанием вероятностей попадания в какой-нибудь меньший набор множеств на прямой? Борелевская σ-алгебра $\mathfrak{B}(\mathbb{R})$ порождается лучами $(-\infty, x)$, поэтому можно ограничиться только вероятностями попадания в такие лучи для всех $x \in \mathbb{R}$. А уже с их помощью можно будет определить и вероятность попасть в любое борелевское множество.

Функция распределения

Определение

Функцией распределения случайной величины ξ называется функция $F_\xi : \mathbb{R} \to [0, 1]$, при каждом $x \in \mathbb{R}$ равная вероятности случайной величине ξ принимать значения, меньшие x:

$$F_\xi(x) = \mathrm{P}(\xi < x) = \mathrm{P}\{\omega : \xi(\omega) < x\}.$$

Пусь ξ — случайная величина равная числу очков, выпавших на кубике. Построим функцию распределения $F_\xi(x)$.

$$F_\xi(x) = \begin{cases} 0, & \text{если } x \le 1, \\ 1/6, & \text{если } 1 < x \le 2, \\ 2/6, & \text{если } 2 < x \le 3, \\ 3/6, & \text{если } 3 < x \le 4, \\ 4/6, & \text{если } 4 < x \le 5, \\ 5/6, & \text{если } 5 < x \le 6, \\ 1, & \text{если } 6 \le x. \end{cases} \tag{1}$$

Вопрос. Можно ли в диапазонах знаки неравенств расставить так $1 \le x < 2$?

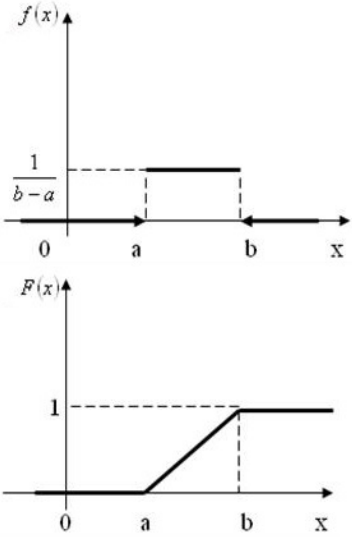
Функция распределения абсолбютно непрерывной случайной величины

Пусть случайная величина ξ имеет абсолютно непрерывное распределение с плотностью $f_\xi(t)$. Тогда функция распределения в любой точке x может быть найдена по плотности распределения так:

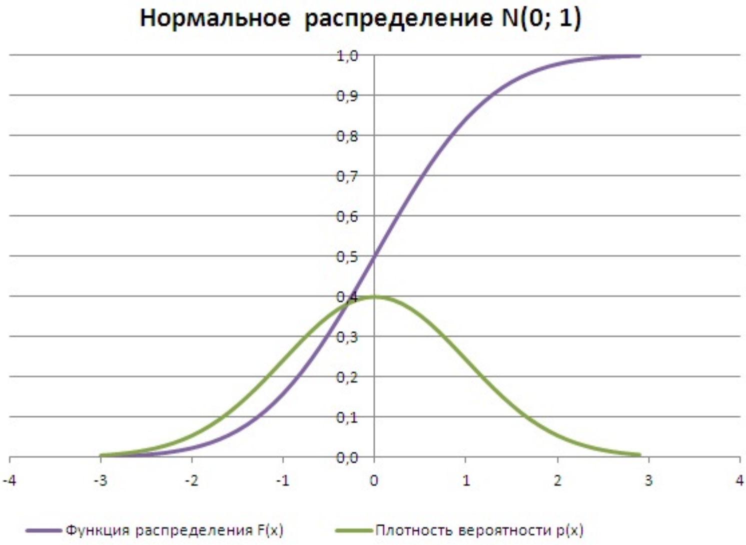
$$F_\xi(x) = \mathrm{P}(\xi < x) = \mathrm{P}(\xi \in (-\infty, x)) = \int\limits_{-\infty}^{x} f_\xi(t)\,dt. \qquad (2)$$

Поскольку функция распределения однозначно определяет распределение случайной величины, можно считать возможность представить функцию распределения интегралом от неотрицательной функции определением абсолютно непрерывного распределения.

Связь плотности и функции распределения



Связь плотности и функции распределения



Свойства функции распределения

Функция распределения обладает свойствами:

(F1) она не убывает: если $x_1 < x_2$, то $F_\xi(x_1) \leq F_\xi(x_2)$;

(F2) существуют пределы $\lim\limits_{x \to -\infty} F_\xi(x) = 0$ и $\lim\limits_{x \to +\infty} F_\xi(x) = 1$;

(F3) она в любой точке непрерывна слева

$$F_\xi(x_0 - 0) = \lim_{x \to x_0 - 0} F_\xi(x) = F_\xi(x_0).$$

И еще важное свойство: $\mathrm{P}(a \leq \xi < b) = F_\xi(b) - F_\xi(a)$.

Если случайная величина ξ имеет абсолютно непрерывное распределение, то для любых $a < b$ имеют место равенства:

$$\mathrm{P}(a < \xi < b) = \mathrm{P}(a \leq \xi < b) = \mathrm{P}(a < \xi \leq b) = \mathrm{P}(a \leq \xi \leq b) =$$

$$= \int\limits_a^b f_\xi(t)dt.$$

Определение

Математическим ожиданием $\mathsf{E}\,\xi$ случайной величины ξ с дискретным распределением называется число

$$\mathsf{E}\,\xi = \sum_k a_k p_k = \sum_k a_k\,\mathsf{P}(\xi = a_k),$$

если данный ряд абсолютно сходится, т. е. если $\sum |a_i| p_i < \infty$. Иначе говорят, что математическое ожидание не существует.

Пусть случайная величина ξ равна числу очков, выпадающих при одном подбрасывании кубика. Тогда

$$\mathsf{E}\,\xi = \sum_{k=1}^{6} k \cdot \frac{1}{6} = \frac{1}{6}(1+2+3+4+5+6) = 3.5.$$

В среднем при одном подбрасывании кубика выпадает 3.5 очка.

Определение

Математическим ожиданием $\mathsf{E}\,\xi$ случайной величины ξ с абсолютно непрерывным распределением с плотностью распределения $f_\xi(x)$ называется число

$$\mathsf{E}\,\xi = \int\limits_{-\infty}^{\infty} x\, f_\xi(x)\, dx,$$

если этот интеграл абсолютно сходится, т. е. если $\int\limits_{-\infty}^{\infty} |x|\, f_\xi(x)\, dx < \infty.$

В противном случае говорят, что математическое ожидание не существует.

Пьяница забыл каким ключом открывется дверь. В кармане у него связка из $n = 9$ ключей (только один подходит к двери). Пьяница наудачу вынимает ключ и пробует открыть им дверь. Если ключ не подходит...

а) он кладет ключ в другой карман.
б) в сердцах кидает ключ обратно в тот же карман (чтобы не потерять его) и начинает свои попытки заново.

Сколько попыток в среднем совершит пьяница?

Расчет мат. ожидания, a

Пусть ξ — случайная величина, равная количеству попыток (включая удачную). Составим распределение случайной величины ξ и вычислим математическое ожидание.

ξ	1	2	...	n
P	$\frac{1}{n}$	$\frac{1}{n}$	...	$\frac{1}{n}$

$$\mathrm{E}\,\xi = \sum_{k=1}^{n} k\frac{1}{n} = \frac{1}{n}\sum_{k=1}^{n} k = \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2}.$$

Расчет мат. ожидания, б

Обозначим вероятность успеха в одном испытании $p = \frac{1}{n}$ и неудачи — $q = 1 - p$.

ξ	1	2	...	k	...
P	p	$p(1-p)$	...	$p(1-p)^{k-1}$	...

.

Очевидно, что ξ может принимать сколь угодно большие значения.

$$\mathrm{E}\,\xi = \sum_{k=1}^{\infty} k\,p\,q^{k-1} = p\sum_{k=1}^{\infty} k\,q^{k-1} = p\sum_{k=1}^{\infty}\frac{dq^k}{dq} =$$

$$= p\,\frac{d}{dq}\left(\sum_{k=1}^{\infty} q^k\right) = p\,\frac{d}{dq}\left(\frac{q}{1-q}\right) = p\,\frac{1}{(1-q)^2} = \frac{1}{p}.$$

Пусть случайная величина ξ — координата точки, брошенной наудачу на отрезок $[a, b]$. Тогда

$$\mathrm{E}\,\xi = \int\limits_a^b x \cdot \frac{1}{b-a}\,dx = \frac{x^2}{2(b-a)}\Bigg|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$$

Центр тяжести равномерного распределения есть середина отрезка.

Свойство 1. Математическое ожидание постоянной величины равно самой постоянной:

$$\mathsf{E}\, C = C.$$

Свойство 2. Постоянный множитель можно выносить за знак математического ожидания :

$$\mathsf{E}\,(C\xi) = C\mathsf{E}\,(\xi).$$

Свойство 3. Математическое ожидание суммы любых случайных величин равно сумме их математических ожиданий, если только эти математические ожидания существуют:

$$\mathsf{E}\,(\xi + \eta) = \mathsf{E}\,\xi + \mathsf{E}\,\eta.$$

Мат. ожидание от произведения найдем, когда изучим системы случайных величин, их совместное распределение и понятие независимости.

Математическое ожидание суммы

ξ	x_1	x_2	$\dots$	x_n
P	p_1	p_2	$\dots$	p_n

η	y_1	y_2	$\dots$	y_m
P	q_1	q_2	$\dots$	p_m

$$\mathrm{E}\,(\xi+\eta)=$$

$$=(x_1+y_1)p_{11}+(x_1+y_2)p_{12}+\ldots+(x_1+y_m)p_{1m}+$$

$$+(x_2+y1)p_{21}+(x_2+y_2)p_{22}+\ldots+(x_2+y_m)p_{2m}+\ldots$$

$$+(x_n+y1)p_{n1}+(x_n+y_2)p_{n2}+\ldots+(x_n+y_m)p_{nm}=$$

$$=x_1(p_{11}+p_{12}+\ldots+p_{1m})+\ldots+x_n(p_{n1}+p_{n2}+\ldots+p_{nm})+$$

$$+y_1(p_{11}+p_{21}+\ldots+p_{n1})+\ldots+y_m(p_{1m}+p_{2m}+\ldots+p_{nm})=$$

$$=x_1p_1+\ldots x_np_n+y_1q_1+\ldots y_mq_m=\mathrm{E}\,\xi+\mathrm{E}\,\eta.$$

1. У стрелка есть 4 патрона. Вероятность попасть в мишень равна p. Стрелок стреляет до первого попадания, но, очевидно, не более четырех раз. Пусть ξ — случайная величина, равная числу выстрелов. Составить закон распределения и найти мат. ожидание случайной величины ξ.
2. Случайная величина ξ задана плотностью распределения

$$f_\xi(x) = \begin{cases} 0, & \text{если } x < 0, \\ a \sin x, & \text{если } 0 \le x \le \pi, \\ 0, & \text{если } \pi < x. \end{cases} \tag{3}$$

Найти параметр a, функцию распределения, вероятность попадания в интервал $[0, \pi/4]$, математическое ожидание.

Задачи посложнее

1. В группе детского сада n человек разного роста. Они встали в круг. Ребенок скажет, что он высокий, если он выше двух своих соседей. Сколько в среднем детсадовцев назовут себя высокими?
2. В комнате n ящиков, в каждом лежит по одному подарку. По очереди в комнату заходит m детей, каждый из которых случайным образом выбирает ящик и забирает оттуда подарок, если таковой там еще есть. Сколько в среднем детей уйдут без подарка?
3. При приеме в старшую группу детского сада проводится следующий экзамен. Перед ребенком случайной стороной раскладываются в ряд 35 карточек, на одной стороне карточек написана буква «м», на другой «а». Ребенок должен найти 4 подряд идущие карточки, на которых написано слово «мама». Сколько в среднем перед ним таких слов?

Самостоятельная работа

1. Есть 40 лотерейных билетов, среди которых 10 выигрышных, причем 2 из них выигрывают по 1000 рублей, а остальные – по 100 рублей. Вася получает один билет случайно. Пусть случайная величина ξ — размера выигрыша Васи. Составить закон распределения случайной величины ξ, найти мат. ожидание выигрыша.
2. Мистер X играет в европейскую рулетку по следующей системе. Подбрасывает монетку: если выпал орел — ставит 100 рублей на «красное», решка — 100 рублей на черное. Составить закон распределения его выигрыша. Вычислить математическое ожидание выигрыша и округлить его до копеек. Сколько в среднем выигрывает игрок с каждой поставленной сотни?

   Справка: европейская рулетка содержит 18 красных, 18 черных и 1 зеленый сектор («зеро»). В случае угадывания цвета игроку выплачивается удвоенная ставка, в противном случае она уходит в доход казино.

Самостоятельная работа

3. Во время эпиденмии все должны носить маски. По заявлению производителя системы наблюдения вероятность того, что умная камера распознает нарушителя масочного режима равна 0.8. В тестовом режиме перед камерой пройшли 6 людей без маски. Составить закон распределения числа распознанных нарушителей. Вычислить математическое ожидание.