Непрерывные случайные величины

октябрь–декабрь 2023 г.

Курс лекций «Теория вероятностей»

Оглавление

Непрерывные случайные величины: нестрогое определение

Определение
Случайная величина ξ называется непрерывной, если функция распределения $F_\xi(x)$ непрерывна на всей числовой прямой и дифференцируема всюду за исключением, возможно, конечного числа точек.

Функция $f_\xi(x) = F'_\xi(x)$ называется плотностью распределения случайной величины ξ или просто плотностью.

Непрерывные случайные величины: нулевая вероятность точки

Из определения следует, что непрерывная случайная величина каждое своё значение принимает с нулевой вероятностью — парадоксальный, на первый взгляд, результат.

Поскольку $P(\xi \in [a, b]) = F_\xi(b) - F_\xi(a-) = F_\xi(b) - F_\xi(a)$, то по формуле Ньютона – Лейбница получаем

$$P(\xi \in [a, b]) = \int_a^b f_\xi(x)dx = \int_{[a,b]} f_\xi(x)dx. \qquad (1)$$

Иными словами, плотность позволяет получать вероятность попадания случайной величины в отрезок (интервал, полуинтервал, объединение непересекающихся отрезков и т. п.) как интеграл от нее по соответствующему множеству, что геометрически представляет площадь соответствующей криволинейной трапеции.

Непрерывные случайные величины: строгое определение

В общей теории вероятности это свойство берется в качестве определения:

Определение
Случайная величина ξ называется непрерывной, если существует такая (измеримая) функция $f(x) = f_\xi(x)$, что

$$P(\xi \in B) = \int_B f(x)dx \tag{2}$$

для любого борелевского множества B. Функция $f(x) = f_\xi(x)$ называется плотностью распределения случайной величины ξ или просто плотностью.

Отступление: интеграл

Интеграл в (2) понимается, как интеграл Лебега, который является обобщением известного интеграла Римана. Можно показать, что если плотность $f(x)$ непрерывна и $B = [a, b]$, то интеграл Лебега $\int_B f(x)dx$ совпадает с обычным интегралом Римана $\int_a^b f(x)dx$.

Неоднозначность плотности

Замечание
Функция плотности в общем определении определяется неоднозначно: если изменить значения функции $f(x)$ в конечном (или счетном) числе точек, то равенство (1) сохранится.

Почему плотность?

Пусть $\Delta = [x_0 - \varepsilon, x_0 + \varepsilon]$, ε — маленькое число. Тогда

$$P(\xi \in \Delta) = \int_{x_0-\varepsilon}^{x_0+\varepsilon} f_\xi(x)dx \approx f_\xi(x_0) \cdot 2\varepsilon = f_\xi(x_0)|\Delta|.$$

Связь плотности и функции распределения

Если функция распределения $F_\xi(x)$ случайной величины ξ непрерывна на всей числовой прямой и непрерывно дифференцируема всюду за исключением конечного числа точек, то в силу теоремы Ньютона – Лейбница

$$P(\xi \in [a, b]) = F_\xi(b) - F_\xi(a) = \int_a^b F'_\xi(x)dx.$$

Значит, функция $f_\xi(x) = F'_\xi(x)$ является плотностью распределения случайной величины ξ. В любом случае

$$F_\xi(b) - F_\xi(a) = \int_a^b f_\xi(x)dx, \tag{3}$$

в частности,

$$F_\xi(x) = \int_{-\infty}^x f_\xi(t)dt,$$

т. е. функция распределения однозначно восстанавливается по функции плотности.

Свойства плотности

1. $f(x) \geqslant 0$;
2. $\displaystyle\int_{-\infty}^{+\infty} f(x)dx = 1$.

Первое свойство более точно следует понимать так: $f(x) \geqslant 0$ всюду за исключением множества, имеющего нулевую меру Лебега. Второе свойство получается, если в (3) устремить $a \to -\infty$, $b \to +\infty$ и воспользоваться свойствами функции распределения.

Теорема существования

Так же, как и для функции распределения, справедлива

Теорема
Если некоторая функция $f(x)$, $x \in \mathbb{R}$, удовлетворяет условиям 1) и 2), то существует случайная величина ξ, для которой эта функция является плотностью распределения.

Примеры непрерывных случайных величин: равномерная

Равномерная на отрезке $[a, b]$ величина:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & x \in [a, b], \\ 0, & x \notin [a, b]. \end{cases}$$

Нетрудно проверить, что функция распределения есть

$$F(x) = \begin{cases} 0, & x < a, \\ \dfrac{x-a}{b-a}, & a \leqslant x \leqslant b, \\ 1, & x > b. \end{cases}$$

В каком-то смысле равномерное распределение есть непрерывный аналог равновероятности для конечного вероятностного пространства.

Примеры непрерывных случайных величин: показательная

Показательная (экспоненциальная) случайная величина с параметром $\lambda > 0$:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geqslant 0, \\ 0, & x < 0. \end{cases}$$

Интегрированием получаем: $F(x) = 1 - e^{-\lambda x}$, $x \geqslant 0$. Функция «хвостов» $G(x) = 1 - F(x) = P(\xi > x) = e^{-\lambda x}$, $x \geqslant 0$.

Показательное распределение используется при моделировании «времени жизни».

Моделирование «времени жизни»

Пусть τ — случайная величина, интерпретируемая как «время жизни» некоторого объекта (время жизни индивидуума, время непрерывной работы прибора, время нахождения в состоянии безработицы и т. п.), имеющая плотность распределения $f_\tau(t)$, $t \geqslant 0$. Функция

$$\kappa(t) = \frac{f_\tau(t)}{1 - F_\tau(t)}$$

называется коэффициентом смертности. Содержательный смысл: величина $\kappa(t) \cdot \Delta$ есть условная вероятность того, что объект «умрет» в промежуток времени $[t, t + \Delta]$, если он «дожил» до момента t, или (что то же самое) доля выбывших за промежуток времени $t, t + \Delta$ среди тех, кто «дожил» до момента t:

$$P(\tau \in [t, t + \Delta] \mid \tau > t) = \frac{P(\tau \in [t, t + \Delta])}{P(\tau > t)} \approx \frac{f_\tau(t)}{1 - F_\tau(t)} \Delta.$$

Моделирование «времени жизни»: показательное распределение

Поэтому величина $\kappa(t)$ характеризует интенсивность «выбывания» в момент t, что и оправдывает терминологию. Легко проверить, что для показательного распределения

$$\kappa(t) = \frac{\lambda e^{-\lambda x}}{1 - (1 - e^{-\lambda x})} = \lambda,$$

коэффициент смертности постоянен. По коэффициенту смертности нетрудно восстановить функцию распределения. Действительно, обозначим $G_\tau(t) = 1 - F_\tau(t)$ — так называемая функция «хвостов» распределения. Очевидно, в силу положительности случайной величины τ, что $G(0) = 1$. По определению

$$\kappa(t) = -\frac{G'_\tau(t)}{G_\tau(t)} = -(\ln G_\tau(t))' \implies G_\tau(t) = \exp\left(-\int_0^t \kappa(s)ds\right).$$

Отсутствие последействия

Еще одно свойство показательной случайной величины τ:

$$P(\tau > s + t \mid \tau > s) = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(\tau > t).$$

Это свойство называют отсутствием последействия. Образно говоря, объект, живущий показательное время, «не стареет».

Примеры непрерывных случайных величин: нормальная

Нормальная (гауссовская) случайная величина с параметрами $m \in \mathbb{R}$, $\sigma^2 > 0$ (обозначение $\xi \sim N(m, \sigma^2)$):

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

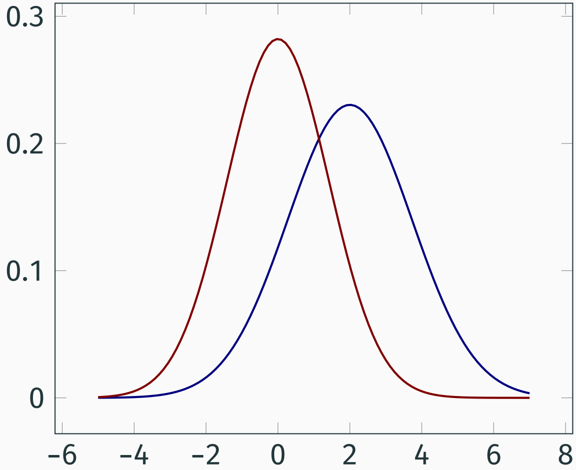
Можно показать, что плотность определена корректно, т. е. что $\int_{-\infty}^{+\infty} f(x)dx = 1$. Доказательство опирается на равенство $\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$ (интеграл Пуассона).

Действительно,

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx = \left| y = \frac{x-m}{\sqrt{2\sigma^2}}, dy = \frac{dx}{\sqrt{2\sigma^2}} \right| = \int_{-\infty}^{+\infty} \sqrt{2\sigma^2} e^{-y^2} dy = \sqrt{2\pi\sigma^2}.$$

Графики нормальных плотностей

Ниже приведены графики нормальных плотностей с $m = 2$, $\sigma^2 = 3$ и с $m = 0$, $\sigma^2 = 2$:



Стандартная нормальная случайная величина

Случайная величина $Z = N(0, 1)$ называется стандартной нормальной случайной величиной. Ее плотность распределения:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Ее функция распределения $\Phi(x) = \int_{-\infty}^{x} \varphi(t)dt$ называется функцией Лапласа.

Гаусс

Иога́нн Карл Фри́дрих Га́усс (нем. Johann Carl Friedrich Gauß; 30 апреля 1777, Брауншвейг — 23 февраля 1855, Гёттинген) — немецкий математик, механик, физик, астроном и геодезист. Считается одним из величайших математиков всех времён, «королём математиков». Лауреат медали Копли (1838), иностранный член Шведской (1821) и Российской (1824) Академий наук, английского Королевского общества.

Отступление: независимость случайных величин (в общем)

Определение
Случайные величины ξ и η называются независимыми, если для любых борелевских множеств A и B вероятность

$$P(\xi \in A, \eta \in B) = P(\xi \in A) \cdot P(\eta \in B).$$

(Подробнее про независимость непрерывных величин позже.)

Примеры непрерывных случайных величин: хи-квадрат

Случайная величина Q подчиняется распределению $\chi^2(n)$ (хи-квадрат с n степенями свободы), если она является суммой квадратов n независимых стандартных нормальных величин. Плотность:

$$f(x) = \begin{cases} \frac{(1/2)^{m/2}}{\Gamma(m/2)} x^{m/2-1} e^{-x/2}, & x \geqslant 0, \\ 0, & x < 0. \end{cases}$$

Это распределение полезно в статистике при построении интервальных оценок дисперсии нормального распределения.

Примеры непрерывных случайных величин: распределение Стьюдента

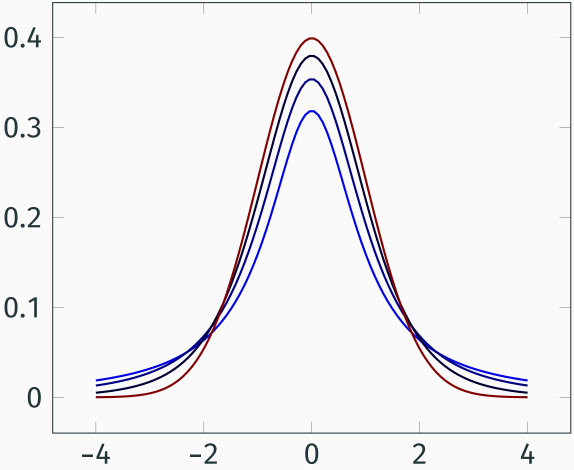
Случайная величина t подчиняется распределению Стьюдента $t(n)$ (с n степенями свободы), если она является отношением $\frac{z}{\sqrt{Q/n}}$, где z — стандартная нормальная величина, Q — величина с распределением $\chi^2(n)$, причем z и Q независимые. Плотность:

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

Это распределение полезно в статистике при построении интервальных оценок среднего для нормального распределения. Названо по псевдониму Уильяма Госсета.

Графики плотностей распределения Стьюдента

При $n \to +\infty$ распределение Стьюдента (приведено для $n = 1, 2, 5$) стремится к стандартному нормальному (красное).



Литература

[1] Sh. Ross. *A First Course in Probability*. 9th ed. London: Pearson, 2014.

[2] Б. В. Гнеденко. *Курс теории вероятностей*. М.: Издательство УРСС, 2005.

[3] С. В. Головань, П. К. Катышев, А. А. Пересецкий. *Сборник задач по курсу теории вероятностей*. М.: Издательский дом «Дело» РАНХиГС, 2021.

Для заметок