Теория вероятностей и статистика в Машинном Обучении

Непрерывные случайные величины

1. Определение непрерывной случайной величины
2. Меры связи случайных величин
3. Их числовые характеристики

Способы задания абсолютно непрерывных случайных величин

Случайная величина X называется *абсолютно непрерывной*, если ее функция распределения может быть представлена в виде

$$F_X(x) = \int_{-\infty}^{x} f_X(z)dz\,.$$

Плотность распределения

При этом функция $f_X(x)$ называется *плотностью распределения вероятностей* (или, короче, *плотностью распределения*) случайной величины X. График плотности распределения случайной величины X называется *кривой распределения вероятностей* (или, короче, *кривой распределения*) случайной величины X.

$$F'(x) = f(x).$$

Свойства плотности распределения

Неотрицательность плотности распределения. *Во всех точках $x \in \mathbb{R}$, в которых плотность распределения $f(x)$ абсолютно непрерывной случайной величины непрерывна,*

$$f(x) \geqslant 0.$$

Нормированность плотности распределения.

$$\int_{-\infty}^{+\infty} f(z)dz = 1.$$

Задача:

Годовой доход случайно выбранного налогоплательщика описывается случайной величиной X с плотностью распределения

$$f(x) = \begin{cases} 0, & x < 1, \\ \dfrac{c}{x^{3,5}}, & x > 1. \end{cases}$$

Найти значение параметра c и функцию распределения годового дохода.

$$1 = \int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{1} 0dx + \int_{1}^{+\infty} \frac{c}{x^{3,5}}dx = -\frac{c}{2,5} \cdot \frac{1}{x^{2,5}}\Bigg|_1^{+\infty} = \frac{c}{2,5} \Rightarrow C=2,5$$

$$F(x) = \int_{-\infty}^{x} f(z)dz = \int_{-\infty}^{x} 0dz = 0 \text{ при } x \leqslant 1,$$

$$F(x) = \int_{-\infty}^{x} f(z)dz = \int_{-\infty}^{1} f(z)dz + \int_{1}^{x} f(z)dz = \int_{1}^{x} \frac{2,5}{z^{3,5}}dt = -\frac{1}{t^{2,5}}\Bigg|_1^{x} = 1 - \frac{1}{x^{2,5}} \text{ при } x > 1.$$

Математическое ожидание, дисперсия, среднеквадратическое отклонение абсолютно *непрерывной* случайной величины

Математическое ожидание: $\mathbf{E}(X) = \int\limits_{-\infty}^{+\infty} x f(x) dx$.

Для *дискретной* случайной величины:

$$\mathbf{E}(X) = \sum_i x_i p_i = \sum_i x_i \mathbf{P}\{X = x_i\}$$

Свойства мат.ожидания:

Формула для математического ожидания константы. *Для любой неслучайной величины* $c \in \mathbb{R}$

$$\mathbf{E}(c) = c.$$

Линейность математического ожидания. *Для любой случайной величины* X *и любой неслучайной величины* $c \in \mathbb{R}$

$$\mathbf{E}(cX) = c\mathbf{E}(X).$$

Аддитивность математического ожидания. *Для любых случайных величин* X *и* Y

$$\mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y).$$

Формула для математического ожидания произведения независимых случайных величин. *Для любых независимых случайных величин* X *и* Y

$$\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y).$$

Дисперсия абсолютно непрерывных случайных величин

$$\mathbf{Var}(X) = \int_{-\infty}^{+\infty} [x - \mathbf{E}(X)]^2 f(x)dx$$

Свойства дисперсии:

Формула для дисперсии константы. *Для любой неслучайной величины* $c \in \mathbb{R}$

$$\mathbf{Var}(c) = 0;$$

Квадратичная нелинейность дисперсии. *Для любой случайной величины* X *и любой неслучайной величины* $c \in \mathbb{R}$

$$\mathbf{Var}(cX) = c^2\mathbf{Var}(X).$$

Аддитивность дисперсии независимых случайных величин

Для любых независимых случайных величин X *и* Y

$$\mathbf{Var}(X + Y) = \mathbf{Var}(X) + \mathbf{Var}(Y).$$

Для *дискретной* случайной величины:

$$\mathbf{Var}(X) = \mathbf{E}\{[X - \mathbf{E}(X)]^2\}$$

Равномерный закон распределения

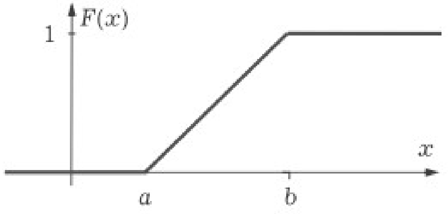
Случайные величины, которые могут принимать значения т о л ь к о в определенных границах некоторого отрезка [a; b], при этом все значения внутри этого отрезка одинаково возможны.

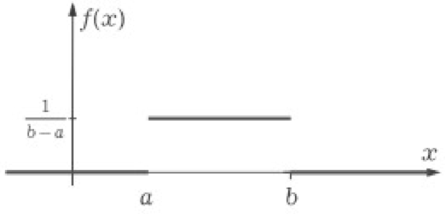
$$F(x)=\begin{cases} 0, & x<a, \\ \dfrac{x-a}{b-a}, & x\in[a;b], \\ 1, & x>b \end{cases}$$

называется распределенной *по равномерному закону* на отрезке $[a;\, b]$.

Плотность распределения:

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





$$\mathbf{E}(X)=\frac{a+b}{2},$$

$$\mathbf{Var}(X)=\frac{(b-a)^2}{12}.$$

Нормальный закон распределения (закон Гаусса)

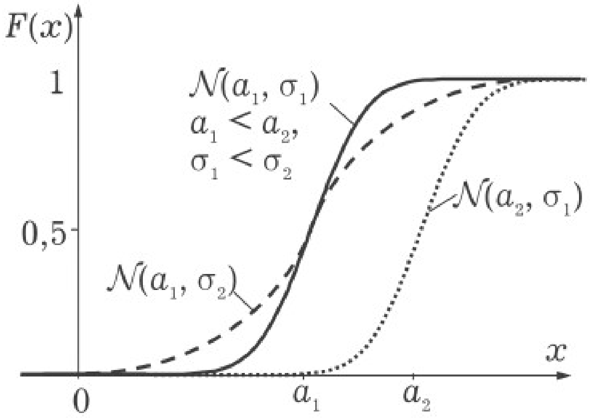
Нормальный закон является **предельным**, к которому приближаются другие законы распределения при часто встречающихся одинаковых условиях.

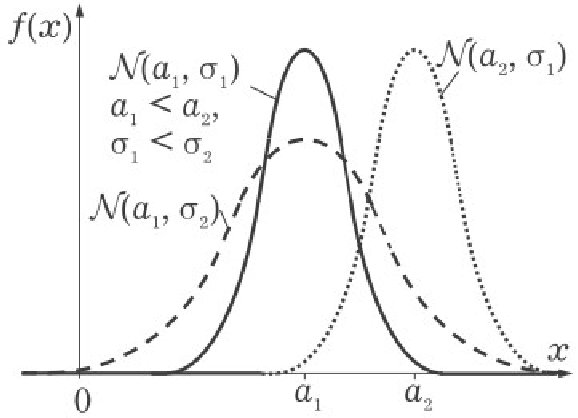
Случайная величина X называется распределенной *по нормальному закону* с параметрами a и σ, если ее плотность распределения имеет вид

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

$$\mathbf{E}(X)=a$$

$$\mathbf{Var}(X)=\sigma^2.$$





$$F(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{(t-a)^2}{2\sigma^2}}\,dt.$$

Стандартный нормальный закон

Нормальный закон распределения с параметрами $a = 0$ и $\sigma = 1$ называется *стандартным нормальным законом*. Плотность распределения и функция распределения случайной величины, распределенной по стандартному нормальному закону, равны соответственно

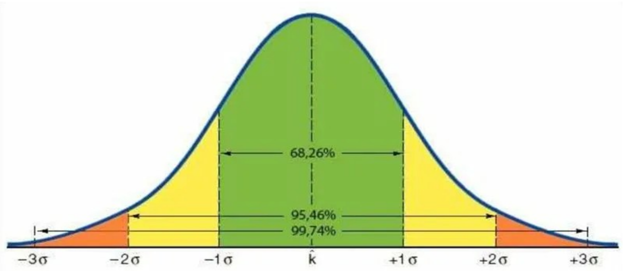
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

и

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt.$$

Правило трех сигм. *Для любой случайной величины* $X = \mathcal{N}(a; \sigma)$ *вероятность*

$$\mathbf{P}\{|X - \mathbf{M}X| < 3\sigma_X\} = 0{,}9973.$$



*Правило **3** сигм: на 10000 ед.товара 27 брака, при статистическом контроле качества сложной электроники в настоящее время используется правило **6** сигм*

МОМЕНТЫ И КРИТИЧЕСКИЕ ГРАНИЦЫ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Начальным моментом k-го порядка случайной величины X называется математическое ожидание k-й степени случайной величины X:

$$\nu_k(X) = \mathbf{E}(X^k), \quad k = 0, 1, 2, \ldots$$

Центральным моментом k-го порядка случайной величины X называется математическое ожидание k-й степени отклонения случайной величины X от своего математического ожидания:

$$\mu_k(X) = \mathbf{E}\{[X - \mathbf{E}(X)]^k\}, \quad k = 0, 1, 2, \ldots$$

Начальный момент первого порядка случайной величины X — это просто ее математическое ожидание: $\nu_1(X) = \mathbf{E}(X)$, а центральный момент второго порядка — это ее дисперсия: $\mu_2(X) = \mathbf{Var}(X)$.

Коэффициент асимметрии $\mathbf{A}_X$ случайной величины X характеризует скошенность кривой распределения этой случайной величины относительно ее математического ожидания и вычисляется по формуле

$$\mathbf{A}_X = \frac{\mu_3(X)}{\sigma_X^3}.$$

Эксцесс $\mathbf{E}_X$ случайной величины X характеризует тяжесть хвостов распределения этой случайной величины по сравнению с нормальным законом распределения и вычисляется по формуле

$$\mathbf{E}_X = \frac{\mu_4(X)}{\sigma_X^4} - 3.$$

Квантили и процентные точки случайной величины

Квантилью (или *левосторонней критической границей*) *уровня* α случайной величины X называется такое число x_α, что

$$F_X(x_\alpha) = \alpha,$$

(т. е. $\mathbf{P}\{X \leqslant x_\alpha\} = \alpha$), а *100α%-ной точкой* (или *правосторонней критической границей уровня* α) случайной величины X называется такое число ω_α, что

$$F_X(\omega_\alpha) = 1 - \alpha$$

(т. е. $\mathbf{P}\{X > \omega_\alpha\} = \alpha$).

Левосторонняя и правосторонняя критические границы одного и того же уровня связаны между собой очевидным соотношением

$$x_\alpha = \omega_{1-\alpha}.$$

Двусторонними критическими границами уровня α случайной величины X называются такие числа $\underline{x}_\alpha$, $\overline{x}_\alpha$, что одновременно

$$F_X(\underline{x}_\alpha) = \frac{\alpha}{2}, \qquad F_X(\overline{x}_\alpha) = 1 - \frac{\alpha}{2},$$

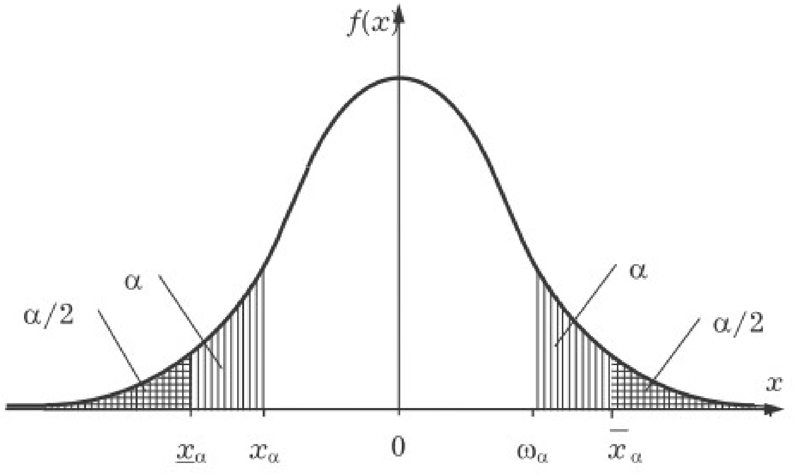
т. е. $\mathbf{P}\{X \leqslant \underline{x}_\alpha\} = \mathbf{P}\{X > \overline{x}_\alpha\} = \dfrac{\alpha}{2}$.

Односторонние и двусторонние критические границы случайной величины X связаны следующими соотношениями:

$$\underline{x}_\alpha = x_{\alpha/2} = \omega_{1-\alpha/2}; \qquad \overline{x}_\alpha = x_{1-\alpha/2} = \omega_{\alpha/2}.$$

Если плотность распределения абсолютно непрерывной случайной величины X с и м м е т р и ч н а относительно оси ординат и $\omega_{\alpha/2} = x_{1-\alpha/2}$ — 50α%-ная точка случайной величины X (т. е. $\mathbf{P}\{X > \omega_{\alpha/2}\} = \alpha/2$), то, очевидно, квантиль уровня $\alpha/2$ этой случайной величины будет равна $x_{\alpha/2} = -\omega_{\alpha/2}$, причем $\mathbf{P}\{X \leqslant -\omega_{\alpha/2}\} = \mathbf{P}\{X > \omega_{\alpha/2}\} = \alpha/2$ или

$$\mathbf{P}\{|X| > \omega_{\alpha/2}\} = \alpha.$$



Критические границы случайной величины $X = \mathcal{N}(0; 1)$

Ковариация и коэффициент корреляции

Для измерения «степени зависимости» случайных величин вводится *ковариация* случайных величин X и Y:

$$\operatorname{cov}(X, Y) = \mathbf{E}\{[X - \mathbf{E}(X)][Y - \mathbf{E}(Y)]\}.$$

Последняя формула легко преобразуется к виду

$$\operatorname{cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y).$$