

Нормальное распределение

Лемма

»

Функция $pdf(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ при любых μ и $\sigma^2 > 0$ задаёт корректную плотность

Определение

» Нормальное распределение со средним μ и дисперсией σ^2

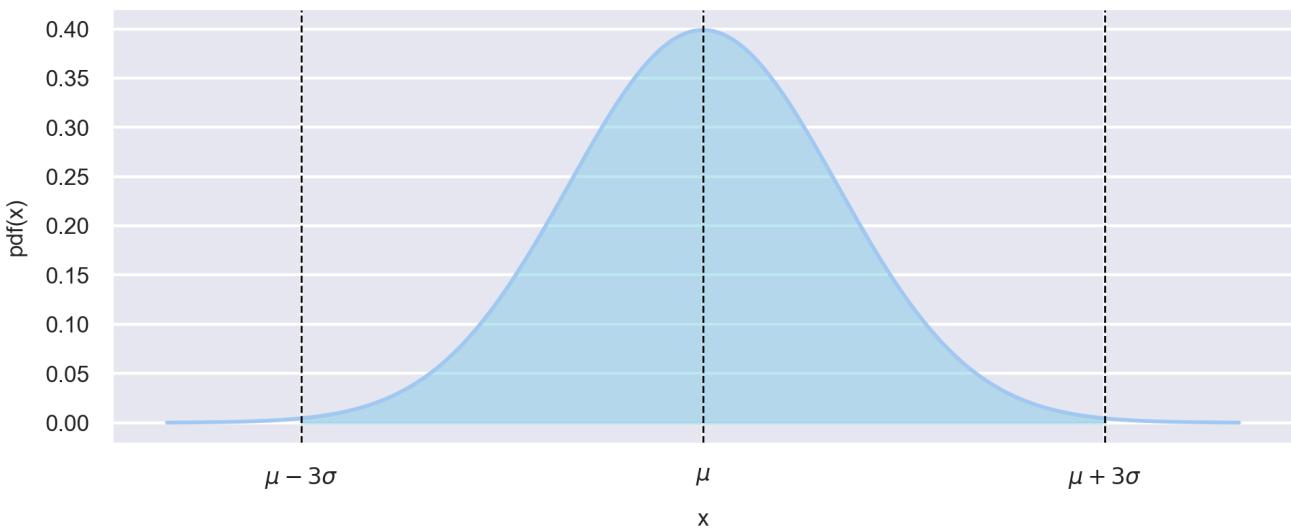
— это непрерывное распределение с плотностью

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

Обозначается $\mathcal{N}(\mu, \sigma^2)$

Носитель - \mathbb{R}

Нормальное распределение



В Python:

```
sps.norm(mu,sigma)
```

где σ - стандартное отклонение

Лемма

»» **Линейная комбинация двух независимых нормальных величин**

Если X, Y — независимые и нормально распределённые случайные величины, $a, b \in \mathbb{R}$ — числовые коэффициенты, причём хотя бы один коэффициент ненулевой, то:

1. $aX + bY$ распределена нормально;
2. $\mathbb{E}(aX + bY) = a \cdot \mathbb{E}X + b \cdot \mathbb{E}Y$;
3. $\mathbb{V}(aX + bY) = a^2 \cdot \mathbb{V}X + b^2 \cdot \mathbb{V}Y$.

Математически можно записать:

$$aX + bY \sim \mathcal{N}(a\mathbb{E}X + b\mathbb{E}Y, a^2\mathbb{V}X + b^2\mathbb{V}Y)$$

Лемма

»» **Сумма и среднее арифметическое выборки из нормального распределения**

Если X_1, X_2, \dots, X_n — случайная выборка из нормального распределения $\mathcal{N}(\mathbb{E}X, \mathbb{V}X)$, то

$$\sum_{i=1}^n X_i \sim \mathcal{N}(n\mathbb{E}X, n\mathbb{V}X)$$

$$\bar{X} \sim \mathcal{N}\left(\mathbb{E}X, \frac{\mathbb{V}X}{n}\right)$$

Лемма

»» **Симметрия нормального распределения**

$$P(X \leq \mu - h) = P(X \geq \mu + h)$$

$$pdf(\mu + h) = pdf(\mu - h), \quad \forall h$$

$$CDF(\mu + h) + CDF(\mu - h) = 1, \quad \forall h$$

Правило 3 сигм

Лемма

Правило трех сигм

Пусть $X \sim \mathcal{N}(\mu, \sigma^2)$

Тогда

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.9973$$

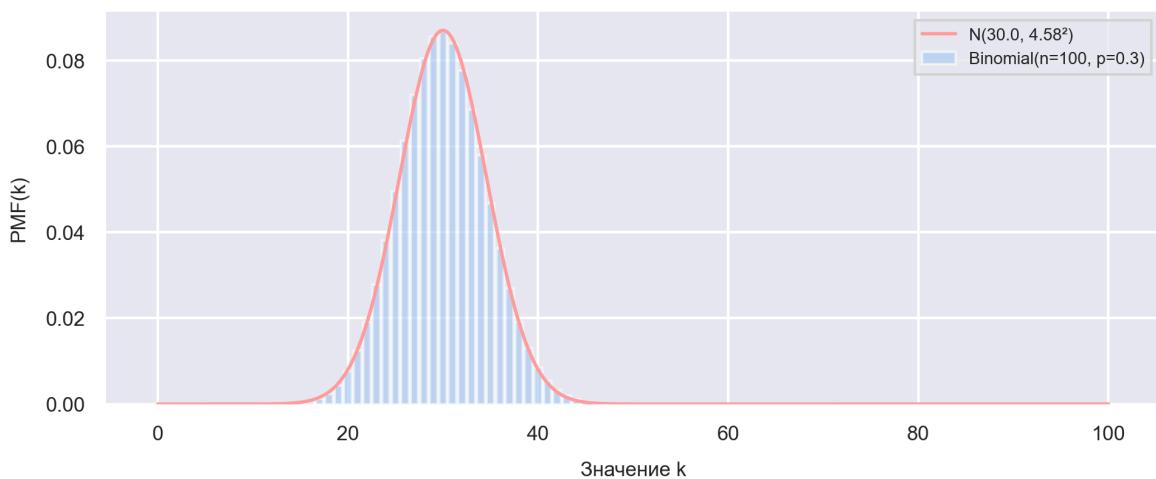
Лемма

Правило 68-95-99.7

- В пределах 1σ - около 68% значений
- В пределах 2σ - около 95% значений
- В пределах 3σ - около 99.7% значений

Приближение колоколообразного распределения нормальным

Биномиальное распределение: большое число испытаний



Матожидание - вершина графика

Выделяем промежуток, который содержит почти 100% вероятности

Считаем его длину и вычисляем стандартное отклонение

Важно

Не все колоколообразные распределения можно аппроксимировать нормальным:

- Асимметричные (скошенные) — нет
- С тяжёлыми хвостами — осторожно
- Дискретные (например, биномиальное при малых n или p далеко от 0.5) — могут плохо приближаться

Стандартное нормальное распределение

Определение

Стандартное нормальное распределение

$$\mathcal{N}(0, 1)$$

Обозначения:

- $Z \sim \mathcal{N}(0, 1)$
- $\Phi = CDF_Z$
- $\phi = pdf_Z$
- $Z_\alpha = PPF_Z(a)$

Лемма

Приведение произвольного нормального распределения к стандартному

Если $X \sim \mathcal{N}(\mathbb{E}X, \mathbb{V}X)$, то

$$\frac{X - \mathbb{E}X}{\sqrt{\mathbb{V}X}} \sim \mathcal{N}(0, 1)$$

Это преобразование называется **стандартизацией**, а результат — **стандартной нормальной величиной Z** .

Лемма

Связь произвольного и стандартного нормального распределения

$$CDF_{\mathcal{N}(\mu, \sigma^2)}(a) = \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$PPF_{\mathcal{N}(\mu, \sigma^2)}(a) = \mu + \sigma \cdot z_\alpha$$

Z-score

Определение

Z-score числа a относительно $\mathcal{N}(\mu, \sigma^2)$

$$\frac{a - \mu}{\sigma}$$

Интерпретация

Показывает, на сколько стандартных отклонений значение a больше матожидания

Пример

Рост человека: $X=180\text{см}$, $\mu=170$, $\sigma=5$

$z = \frac{180-170}{5} = 2 \rightarrow$ рост на 2 стандартных отклонения выше среднего.

Распределение хи-квадрат

Распределение Стьюдента