# Метод максимального правдоподобия. Функции распределения. Концентрация.

Леонид Иосипой

Курс «Вероятностные модели и статистика»
Центр непрерывного образования, ВШЭ

13 апреля 2021

- Повторение
- Функции распределения
- Неравенства концентрации I
- Распределение суммы случайных величин

# Повторение

1. Метод максимального правдоподобия.

   Допустим, что у нас есть реализация выборки из некоторого распределения, известного с точностью до одного или нескольких параметров.

   **Метод максимального правдоподобия:** чтобы оценить неизвестные параметры модели, нам необходимо найти максимум функции правдоподобия (то есть найти частные производные по всем параметрам и приравнять их к нулю).

# Повторение

1. Метод максимального правдоподобия.

Пусть дана реализация выборки $x_1, \dots, x_n$ из некоторого распределения с неизвестным (многомерным) параметром

$$\theta \in \Theta \subset \mathbb{R}^d.$$

Введем величину:

$$p(u, \theta) = \begin{cases} \mathbb{P}_\theta(X = u) & \text{в дискретном случае,} \\ f_\theta(u) & \text{в непрерывном случае.} \end{cases}$$

Функцией правдоподобия называется величина:

$$L(\theta) = p(x_1, \theta) \cdot \ldots \cdot p(x_n, \theta).$$

# Повторение

1. Метод максимального правдоподобия.

   В общем случае $L(\theta)$ характеризует вероятность получить реализацию $x_1, \dots, x_n$ выборки при заданном $\theta$.

   Представляется разумным в качестве оценки параметра $\theta$ взять наиболее правдоподобное значение, которое получается при максимизации функции $L(\theta)$.

# Метод максимального правдоподобия

## Задача

Пусть $x_1, \ldots, x_n$ – реализация из распределения Бернулли $\mathbf{B}_\theta$ с неизвестным параметром успеха $\theta \in [0, 1]$. Оценить $\theta$ с помощью метода максимального правдоподобия.

# Метод максимального правдоподобия

**Решение.** Найдем сначала функцию правдоподобия:

$$p(u,\theta) = \mathbb{P}_\theta(X = u) = \begin{cases} 1-\theta & \text{для } u = 0, \\ \theta & \text{для } u = 1, \end{cases} = (1-\theta)^{1-u}\theta^u,$$

$$L(\theta) = p(x_1,\theta)\cdot\ldots\cdot p(x_n,\theta) = (1-\theta)^{n-\sum_{i=1}^n x_i}\cdot\theta^{\sum_{i=1}^n x_i}.$$

Найдем логарифм функции правдоподобия:

$$\ln L(\theta) = \left(n - \sum_{i=1}^n x_i\right)\ln(1-\theta) + \sum_{i=1}^n x_i \ln\theta.$$

# Метод максимального правдоподобия

Дифференцируя ее по $\theta$, получаем

$$(\ln L(\theta))' = -\frac{n - \sum_{i=1}^{n} x_i}{1-\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta}.$$

Приравняем производную к нулю:

$$\frac{n - \sum_{i=1}^{n} x_i}{1-\theta} = \frac{\sum_{i=1}^{n} x_i}{\theta}.$$

Откуда

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

# Функции распределения

Функцией распределения случайной величины $X$ называется функция $F_X : \mathbb{R} \to [0, 1]$ задаваемая следующей формулой

$$F_X(u) = \mathbb{P}(X \leq u).$$

Значение $F$ в точке $u$ аккумулирует вероятности всех возможных значений $X$ вплоть до $u$ (включительно).

Мы иногда будем опускать нижний индекс $X$, когда понятно, о какой случайной величине идет речь.

# Функции распределения

В дискретном случае:

$$F_X(u) := \mathbb{P}(X \leq u) = \sum_{i:\ a_i \leq u} p_i,$$

В непрерывном случае:

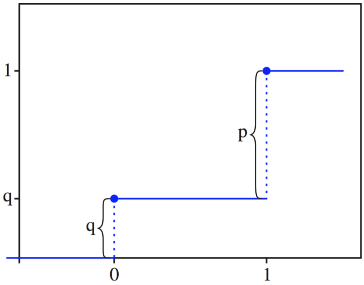
$$F_X(u) := \mathbb{P}(X \leq u) = \int_{-\infty}^{u} f(y) dy.$$

# Функции распределения

## Задача

Нарисовать график $F_X(u)$ для $X \sim \mathbf{B}_p$, $p \in [0, 1]$.

$$F_X(u) = \mathbb{P}(X \le u).$$



# Функции распределения

## Задача

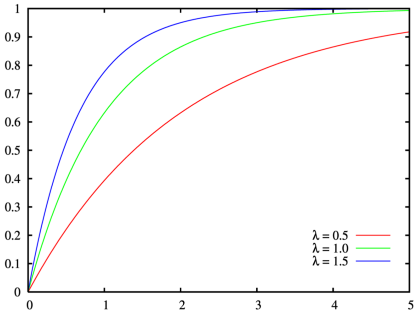
Нарисовать график $F_X(u)$ для $X \sim \text{Exp}(\lambda)$, $\lambda > 0$.

$$F_X(u) = \mathbb{P}(X \leq u).$$

# Функции распределения

**Решение.** Заметим, что для всех $u \leq 0$, $F_X(u) = 0$. А для $u > 0$ имеем

$$F_X(u) = \mathbb{P}(X \leq u) = \int_0^u \lambda e^{-\lambda y} dy = \left(-e^{-\lambda y}\right)\Big|_0^u = 1 - e^{-\lambda u}.$$



# Функции распределения

**Важное свойство функции распределения:**

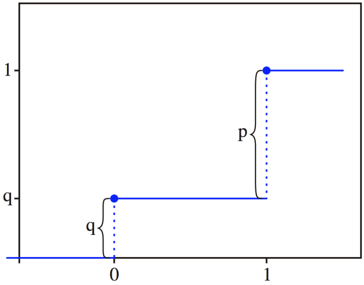
Любую случайную величину $X$ можно задать через функцию распределения. То есть по функции распределения можно восстановить распределение случайной величины $X$:

— в дискретном случае можно восстановить $a_k$ и $p_k$;

— в непрерывном случае можно восстановить $f(u)$.

# Функции распределения

## Дискретный случай



# Функции распределения

## Непрерывный случай

$$\boxed{F'_X(u) = f(u)}$$

Действительно,

$$\begin{aligned}
\lim_{\Delta u \to 0} \frac{F(u+\Delta u) - F(u)}{\Delta u} &= \lim_{\Delta u \to 0} \frac{\int_u^{u+\Delta u} f(y)dy}{\Delta u} \\
&\approx \lim_{\Delta u \to 0} \frac{f(u) \cdot \Delta u}{\Delta u} \\
&= f(u).
\end{aligned}$$

# Функции распределения

Этот факт позволит нам находить плотность непрерывной случайной величины после некоторого преобразования.

Как найти плотность $Y = g(X)$, зная плотность $X$?

1. Выразить функцию распределения $F_Y(u)$ через функцию распределения $F_X(u)$;
2. Продифференцировать $F_Y(u)$.

# Функции распределения

## Задача

Найти плотность $Y = 5 + X$, где $X \sim \mathrm{Exp}(\lambda)$ и $\lambda > 0$.

# Функции распределения

**Решение.** Воспользуемся нашим алгоритмом.

1. Найдем функцию распределения $F_Y(u)$:

   $$F_Y(u) = \mathbb{P}(Y \leq u) = \mathbb{P}(5+X \leq u) = \mathbb{P}(X \leq u-5) = F_X(u-5).$$

   Используя результаты предыдущей задачи:

   $$F_X(u-5) = \begin{cases} 1 - e^{-\lambda(u-5)}, & u \geq 5, \\ 0, & u < 5. \end{cases}$$

2. Продифференцируем:

   $$f_Y(u) = F'_Y(u) = \begin{cases} \lambda e^{-\lambda(u-5)}, & u \geq 5, \\ 0, & u < 5. \end{cases}$$

# Функции распределения

**Решение.** Воспользуемся нашим алгоритмом.

1. Найдем функцию распределения $F_Y(u)$:

   $F_Y(u) = \mathbb{P}(Y \leq u) = \mathbb{P}(5+X \leq u) = \mathbb{P}(X \leq u-5) = F_X(u-5);$

   Используя результаты предыдущей задачи:

$$F_X(u-5) = \begin{cases} 1 - e^{-\lambda(u-5)}, & u \geq 5, \\ 0, & u < 5. \end{cases}$$

2. Продифференцируем:

$$f_Y(u) = F_Y'(u) = \begin{cases} \lambda e^{-\lambda(u-5)}, & u \geq 5, \\ 0, & u < 5. \end{cases}$$

# Функции распределения

Явно вычислять $F_X(u)$ не обязательно:

$$\begin{aligned} F'_Y(u) &= \Big(F_X(u-5)\Big)' \\ &= F'_X(u-5)\cdot(u-5)' \\ &= f_X(u-5)\cdot 1 \\ &= \begin{cases} \lambda e^{-\lambda(u-5)}, & u \geq 5, \\ 0, & u < 5. \end{cases} \end{aligned}$$

# Функции распределения

## Задача

Найти плотность $Y \sim \mathcal{N}(a, \sigma^2)$ для некоторых $a \in \mathbb{R}$ и $\sigma > 0$.

# Функции распределения

**Решение.** По определению $Y = a + \sigma X$, где $X \sim \mathcal{N}(0, 1)$.

Воспользуемся нашим алгоритмом.

1. Найдем функцию распределения $F_Y(u)$:

   $$F_Y(u) = \mathbb{P}(Y \leq u) = \mathbb{P}(a+\sigma X \leq u) = \mathbb{P}\left(X \leq \frac{u-a}{\sigma}\right) = F_X\left(\frac{u-a}{\sigma}\right);$$

   Явно вычислить $F_X$ не получится!

2. Продифференцируем:

   $$f_Y(u) = F'_Y(u) = \left(F_X\left(\frac{u-a}{\sigma}\right)\right)' = f_X\left(\frac{u-a}{\sigma}\right) \cdot \frac{1}{\sigma} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(u-a)^2}{2\sigma^2}}.$$

# Неравенства концентрации

Неравенства концентрации — это оценка вероятности

$$\mathbb{P}(X \in A)$$

с помощью числовых характеристик $X$.

# Неравенства концентрации

## Теорема (Неравенство Маркова)

*Пусть $X$ — произвольная случайная величина. Тогда для любого $t > 0$*

$$\mathbb{P}(|X| > t) \leq \frac{\mathbb{E}|X|}{t}.$$

# Неравенства концентрации

**Доказательство.** Для произвольного события $A$, введем случайную величину $I(A)$, которую будем называть индикатором $A$:

$$I(A) = \begin{cases} 1, & A \text{ произошло}, \\ 0, & A \text{ не произошло, то есть произошло } \overline{A}. \end{cases}$$

Заметим, что $I(A)$ имеет распределение Бернулли и

$$\mathbb{E}\big[I(A)\big] = 0 \cdot \mathbb{P}(\overline{A}) + 1 \cdot \mathbb{P}(A) = \mathbb{P}(A).$$

# Неравенства концентрации

**Доказательство.** Пусть $A = \{|X| > t\}$, $\overline{A} = \{|X| \le t\}$.

Тогда

$$|X| = |X| \cdot \Big(I(A) + I(\overline{A})\Big) \ge |X| \cdot I(A) \ge t \cdot I(A).$$

Наконец, по свойствам математического ожидания:

$$\mathbb{E}|X| \ge \mathbb{E}(t \cdot I(A)) = t \cdot \mathbb{P}(A) = t \cdot \mathbb{P}\big(|X| > t\big).$$

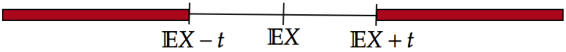
Осталось разделить обе части неравенства на $t > 0$.

# Неравенства концентрации

## Теорема (Неравенство Чебышёва)

*Пусть $X$ — произвольная случайная величина и $|\mathbb{E}X| < \infty$. Тогда для любого $t > 0$*

$$\mathbb{P}\big(|X - \mathbb{E}X| > t\big) \le \frac{\operatorname{Var} X}{t^2}.$$



# Неравенства концентрации

**Доказательство.** Действительно,

$$\begin{aligned}\mathbb{P}\big(|X - \mathbb{E}X| > t\big) &= \mathbb{P}\Big((X - \mathbb{E}X)^2 > t^2\Big) \quad \text{(равносильное неравенство)}\\ &\le \frac{\mathbb{E}(X - \mathbb{E}X)^2}{t^2} \quad \text{(неравенство Маркова)}\\ &= \frac{\operatorname{Var} X}{t^2} \quad \text{(определение дисперсии)}.\end{aligned}$$

# Неравенства концентрации

Заметим, что в доказательстве мы могли возвести обе части неравенства не в квадрат, а в другую степень $k \in \mathbb{N}$:

$$\mathbb{P}(|X - \mathbb{E}X| > t) = \mathbb{P}\left(|X - \mathbb{E}X|^k > t^k\right) \quad \text{(равносильное неравенство)}$$
$$\leq \frac{\mathbb{E}|X - \mathbb{E}X|^k}{t^k} \quad \text{(неравенство Маркова)}.$$

# Неравенства концентрации

Все эти характеристики могут встречаться в концентрации $X$:

- $\mathbb{E}X^k$ — $k$-ый момент $X$;
- $\mathbb{E}|X|^k$ — абсолютный $k$-ый момент $X$;
- $\mathbb{E}(X - \mathbb{E}X)^k$ — центральный $k$-ый момент $X$;
- $\mathbb{E}|X - \mathbb{E}X|^k$ — абсолютный центральный $k$-ый момент $X$.

# Неравенства концентрации

## Задача

Небольшая фабрика игрушек производит случайное число $X$ плюшевых жирафов в месяц. Мы знаем, что среднее число жирафов в месяц — 500.

1. Что можно сказать о вероятности произвести более 1000 жирафов в месяц?
2. Дисперсия случайной величины $X$ равна 100. Что можно сказать о вероятности произвести от 400 до 600 жирафов в месяц?

$$\boxed{\mathbb{P}(|X| > t) \leq \frac{\mathbb{E}|X|}{t}, \qquad \mathbb{P}(|X - \mathbb{E}X| > t) \leq \frac{\operatorname{Var} X}{t^2}.}$$

# Неравенства концентрации

**Решение:**

1. По неравенству Маркова:
$$\mathbb{P}(X > 1000) \le \frac{\mathbb{E}X}{1000} = \frac{500}{1000} = \frac{1}{2}.$$
Что это значит?

2. По неравенству Чебышёва:
$$\begin{aligned}\mathbb{P}(400 \le X \le 600) &= \mathbb{P}(|X - \mathbb{E}X| \le 100) \\ &= 1 - \mathbb{P}(|X - \mathbb{E}X| > 100) \\ &\ge 1 - \frac{\operatorname{Var} X}{100^2} \\ &= \frac{99}{100}.\end{aligned}$$
Что это значит?

# Распределение суммы случайных величин

Пусть даны две независимые случайные величины $X$ и $Y$:

| $X$ | 0 | 1 | ... | $n$ |
|---|---|---|---|---|
| $\mathbb{P}$ | $p_0$ | $p_1$ | ... | $p_n$ |

| $Y$ | 0 | 1 | ... | $n$ |
|---|---|---|---|---|
| $\mathbb{P}$ | $q_0$ | $q_1$ | ... | $q_n$ |

Каким будет распределение у суммы $X + Y$?

То есть чему будут равны вероятности $\mathbb{P}(X + Y = k)$, где $k$ — любое целое число от 0 до $2n$ (сумма не превышает $2n$)?

# Распределение суммы случайных величин

Пусть даны две независимые случайные величины $X$ и $Y$:

| $X$ | 0 | 1 | ... | $n$ |
|---|---|---|---|---|
| $\mathbb{P}$ | $p_0$ | $p_1$ | ... | $p_n$ |

| $Y$ | 0 | 1 | ... | $n$ |
|---|---|---|---|---|
| $\mathbb{P}$ | $q_0$ | $q_1$ | ... | $q_n$ |

$$\begin{aligned}
\mathbb{P}(X+Y=k) &= \sum_{j=0}^{k} \mathbb{P}(X=j, Y=k-j) \\
&= \sum_{j=0}^{k} \mathbb{P}(X=j)\cdot\mathbb{P}(Y=k-j) \\
&= \sum_{j=0}^{k} p_j \cdot q_{k-j}.
\end{aligned}$$

# Распределение суммы случайных величин

В общем виде «красивой» формулы нет, но идея такая:

чтобы найти вероятности $\mathbb{P}(X + Y = c)$ необходимо понять, при каких $a_i$ и $b_j$ получается $a_i + b_j = c$, и просуммировать соответствующие произведения вероятностей $p_i \cdot q_j$.

В случае, когда $X$ и $Y$ принимают значения из $\mathbb{Z}$, формула выглядит так: для $k \in \mathbb{Z}$

$$\mathbb{P}(X + Y = k) = \sum_{j=-\infty}^{+\infty} \mathbb{P}(X = j) \cdot \mathbb{P}(Y = k - j).$$

Эта формула называется свёрткой (дискретной).

# Распределение суммы случайных величин

Как будет выглядеть формула в непрерывном случае?

Пусть $X$ имеет плотность $f_X(u)$, а $Y$ — плотность $f_Y(u)$.

$$\mathbb{P}(X+Y=k) = \sum_{j=-\infty}^{+\infty} \mathbb{P}(X=j)\cdot\mathbb{P}(Y=k-j).$$

$$\downarrow$$

$$f_{X+Y}(u) = \int_{-\infty}^{+\infty} f_X(z) f_Y(u-z)dz.$$

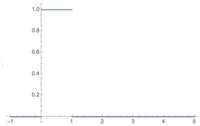
Эта формула называется свёрткой (непрерывной).

# Распределение суммы случайных величин

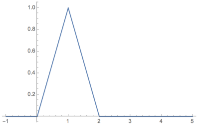
Поиск распределения суммы двух независимых случайных величин — сложная задача!

# Распределение суммы случайных величин

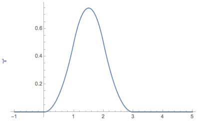
**Пример.** Пусть $X_1, X_2, X_3, X_4$ — независимые случайные величины, равномерно распределенные на $[0, 1]$. Какое распределение будет у $X_1$, $X_1 + X_2$, $X_1 + X_2 + X_3$ и $X_1 + X_2 + X_3 + X_4$?



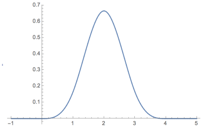
$X_1$



$X_1 + X_2$



$X_1 + X_2 + X_3$



$X_1 + X_2 + X_3 + X_4$

Что-то напоминает?

# Преобразования случайных величин

Базовые операции для работы со случайными величинами:

1. Центрирование: вычитание из $X$ ее мат. ожидания $\mathbb{E}X$,
$$\mathbb{E}[X - \mathbb{E}X] = \mathbb{E}[X] - \mathbb{E}[X] = 0;$$
2. Нормирование: деление $X$ на $\sqrt{\mathrm{Var}(X)}$,
$$\mathrm{Var}\left(\frac{X}{\sqrt{\mathrm{Var}(X)}}\right) = \left(\frac{1}{\sqrt{\mathrm{Var}(X)}}\right)^2 \cdot \mathrm{Var}(X) = 1;$$
3. Стандартизация: центрирование + нормирование,
$$\frac{X - \mathbb{E}X}{\sqrt{\mathrm{Var}(X)}}.$$

Спасибо за внимание!