# Прикладная статистика в машинном обучении
## Семинар 1

И. К. Козлов
(*Мехмат МГУ*)

2022

# Прикладная статистика в машинном обучении

Курс сдаётся в

**https://anytask.org**

!!! Нужно зарегистрироваться !!!

## Семинар 1

Постарайтесь не заснуть на семинаре и отвечать на вопросы!

Рис.: Не спи - замёрзнешь!

# MSE, проверка несмещённости и состоятельности

# Семинар 1

**Q1:** Доказать, что
$$\overline{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$$
является **несмещённой** оценкой $\mathbb{E}X$.

# Семинар 1

**Q1:** Доказать, что

$$\overline{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$$

является **несмещённой** оценкой $\mathbb{E}X$.

**A:** Следует из линейности матожидания:

$$\mathbb{E}\sum_i X_i = \sum_i \mathbb{E}X_i$$

(для любых $X_1, \dots, X_n$, не обязательно независимых).

**Q2**: Доказать, что
$$\overline{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$$
является **состоятельной** оценкой $\mathbb{E}X$.

**A**: Если
$$\text{Bias} \to 0, \qquad \mathbb{V}\overline{X} \to 0,$$
то оценка состоятельная.

Оценка несмещённая, поэтому
$$\text{Bias} = \mathbb{E}\bar{X} - \mathbb{E}X = 0.$$

# Семинар 1

- Осталось оценить дисперсию.
- Неравенство Чебышёва
$$\mathbb{P}(|\hat{\theta} - \theta| \geq \varepsilon) \leq \frac{\mathbb{V}(\hat{\theta})}{\varepsilon^2}.$$
- Дисперсия суммы *независимых* с.в. = сумма диспресий
$$\mathbb{V}\overline{X} = \frac{1}{n^2}\mathbb{V}\sum_i X_i = \frac{1}{n^2}\sum_i \mathbb{V}X_i = \frac{\sigma^2}{n}.$$
- Оценка состоятельна, т.к. $\text{bias} = 0, \mathbb{V}\overline{X} \to 0$.

# Семинар 1

**Q3**: Найти **MSE** для

$$\overline{X}_n = \frac{1}{n}(X_1 + \dots + X_n).$$

**A**: Если

$$MSE = \text{Bias}^2 + \mathbb{V}\overline{X} = 0 + \frac{\sigma^2}{n} = \frac{\sigma^2}{n}.$$

# Известные факты из теорвера

# Семинар 1

**Q:** Является выборочная дисперсия

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

**несмещённой** оценкой дисперсии?

# Семинар 1

**Q**: Является выборочная дисперсия

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

**несмещённой** оценкой дисперсии?

**A**: Нет, у несмещённой дисперсии другой коэффициент

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 .$$

# Семинар 1

Полезные неравенства:

- Неравенство Чебышёва. Если $\mathbb{V}X = \sigma^2 < \infty$ и $\mathbb{E}X = \mu$, то

$$\mathbb{P}\left(|X - \mu| \geqslant a\right) \leqslant \frac{\sigma^2}{a^2},$$

- Неравенство Маркова. Если $X \geq 0$ и $a > 0$, то

$$\mathrm{P}(X \geq a) \leq \frac{\mathrm{E}(X)}{a}.$$

## Метод моментов

# Метод моментов

Метод моментов.

- Моменты — это функции от параметров
$$\mu_j = \mathbb{E}(X^j) = g_j(\theta_1, \ldots, \theta_k).$$
- *Приравниваем моменты и выборочные моменты*
$$\widehat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} x_i^j$$
и решаем как уравнение на параметры $\theta_1, \ldots, \theta_k$.

# Семинар 1

Оценка методом параметров $\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_k$ — решение системы уравнений

$$
\begin{aligned}
\widehat{\mu}_1 &= g_1(\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_k),\\
\widehat{\mu}_2 &= g_2(\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_k),\\
&\vdots\\
\widehat{\mu}_k &= g_k(\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_k).
\end{aligned}
$$

# Семинар 1

**Пример.** Найдём параметры нормального распределения $\mathcal{N}(\mu, \sigma^2)$ с помощью метода моментов.

# Семинар 1

**Пример.** Найдём параметры нормального распределения $\mathcal{N}(\mu, \sigma^2)$ с помощью метода моментов.

**A:**

- Моменты: $\mathbb{E}X = \mu$. Чему равно $\mathbb{E}X^2$?.

**Пример.** Найдём параметры нормального распределения $\mathcal{N}(\mu, \sigma^2)$ с помощью метода моментов.

**A**:

- Моменты: $\mathbb{E}X = \mu$. Чему равно $\mathbb{E}X^2$?.

$$\mathbb{V}X = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}(X^2) - (\mathbb{E}X)^2.$$

- Итак, в данном случае

$$\mu_1 = \mu, \qquad \mu_2 = \sigma^2 + \mu^2.$$

# Семинар 1

Получаем систему уравнений

$$\mu = \frac{1}{n}\sum_i X_i$$
$$\mu^2 + \sigma^2 = \frac{1}{n}\sum_i X_i^2$$

**Ответ:**

$$\mu = \bar{X}$$
$$\sigma^2 = \frac{1}{n}\sum_i \left(X_i - \bar{X}\right)^2$$

## Семинар 1

*При необходимых нужных условий*

- Оценка методом момента $\hat{\theta}_n$ существует.
- Состоятельна $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$.
- Асимптотически нормальна.

# Оценка максимального правдоподобия

# Семинар 1

**Q**: $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$. Найти оценку максимального правдоподобия.

**A**: Плотность распределения

$$f_p(x) = p^x (1 - p)^{1-x}.$$

## Семинар 1

**Q:** $X_1, \dots, X_n \sim \text{Bernoulli}(p)$. Найти оценку максимального правдоподобия.

**A:** Плотность распределения

$$f_p(x) = p^x (1-p)^{1-x}.$$

1. Вычисляем логарифмическую функцию правдоподобия.

$$\ell_n = \sum_i \ln f_p(x_i) = S \ln p + (n - S) \ln(1-p), \qquad S = \sum x_i.$$

# Семинар 1

**Q**: $X_1, \ldots, X_n \sim \mathrm{Bernoulli}(p)$. Найти оценку максимального правдоподобия.

**A**: Плотность распределения

$$f_p(x) = p^x (1-p)^{1-x}.$$

1. Вычисляем логарифмическую функцию правдоподобия.

$$\ell_n = \sum_i \ln f_p(x_i) = S \ln p + (n-S)\ln(1-p), \qquad S = \sum x_i.$$

2. Находим экстремум

$$\frac{\partial \ell_n}{\partial p} = 0 \qquad \Rightarrow \qquad \frac{S}{p} - \frac{(n-S)}{(1-p)} = 0 \qquad \Rightarrow \qquad \hat{p} = \frac{S}{n}.$$

**Q**: Что ещё нужно сделать?

# Семинар 1

- По-хорошему, нужно проверить, что *максимум не достигается на границе*.

  Если в исходе все элементы равны, то формула верна. Иначе при $p = 0, 1$ правдоподобие равно 0, и это не максимум.

## Семинар 1

- По-хорошему, нужно проверить, что *максимум не достигается на границе*.

  Если в исходе все элементы равны, то формула верна. Иначе при $p = 0, 1$ правдоподобие равно 0, и это не максимум.

- Проверить, что *экстремум* — *максимум*. Варианты:

# Семинар 1

- По-хорошему, нужно проверить, что *максимум не достигается на границе*.

  Если в исходе все элементы равны, то формула верна. Иначе при $p = 0, 1$ правдоподобие равно 0, и это не максимум.

- Проверить, что *экстремум — максимум*. Варианты:
  - Матрица вторых производных (гессиан) отрицательно определена.
  - Т-ма Вейрештрассе. Непрерывная функция на компакте достигает своего минимума и максимума.

## Семинар 1

**Q**: $X_1, \ldots, X_n \sim \mathrm{U}(0, \theta)$. Найти ML оценку.

Плотность:

$$f(x, \theta) = \begin{cases} \frac{1}{\theta}, & 0 \le x \le \theta \\ 0, & \text{иначе} \end{cases}$$

# Семинар 1

**Q**: $X_1, \ldots, X_n \sim \mathrm{U}(0, \theta)$. Найти ML оценку.

Плотность:

$$f(x, \theta) = \begin{cases} \frac{1}{\theta}, & 0 \leq x \leq \theta \\ 0, & \text{иначе} \end{cases}$$

**A**: $X_n = \max(X_1, \ldots, X_n)$, потому что

$$\mathcal{L}_n = \begin{cases} \frac{1}{\theta^n}, & \theta \geq X_{(n)} \\ 0, & \text{иначе} \end{cases}$$

# Семинар 1

**Q**: $X_1, \ldots, X_n \sim \mathrm{U}(0, \theta)$. Найти ML оценку.

Плотность:

$$f(x, \theta) = \begin{cases} \frac{1}{\theta}, & 0 < x < \theta \\ 0, & \text{иначе} \end{cases}$$

# Семинар 1

**Q**: $X_1, \ldots, X_n \sim \mathrm{U}(0, \theta)$. Найти ML оценку.

Плотность:

$$f(x, \theta) = \begin{cases} \frac{1}{\theta}, & 0 < x < \theta \\ 0, & \text{иначе} \end{cases}$$

**A**: Её нет.

# Семинар 1

**Q**: Будет ли оценка $X_n = \max(X_1, \ldots, X_n)$ асимптотически нормальна?

## Семинар 1

**Q**: Будет ли оценка $X_n = \max(X_1, \ldots, X_n)$ асимптотически нормальна?

**A**: Нет, $\hat{\theta}$ всегда меньше $\theta$.

Задание на дом — *что пошло не так*?

# Экспоненциальное семейство распределений

# Семинар 1

Семейство распределений относится к экспоненциальному классу, если его плотность может быть представлена в следующем виде:

$$f(x|\theta) = \frac{h(x)}{g(\theta)} \exp\left(\theta^T u(x)\right),$$

где

- $h(x) \geq 0$,
- $\theta^T u(x)$ означает $\theta_1 u_1(x) + \cdots + \theta_k u_k(x)$,
- $u_j(x)$ — произвольные функции.

# Семинар 1

$$f(x|\theta) = \frac{h(x)}{g(\theta)} \exp\left(\theta^T u(x)\right)$$

*Контрольный вопрос*. Чему равно $g(\theta)$?

# Семинар 1

$$f(x|\theta) = \frac{h(x)}{g(\theta)} \exp\left(\theta^T u(x)\right)$$

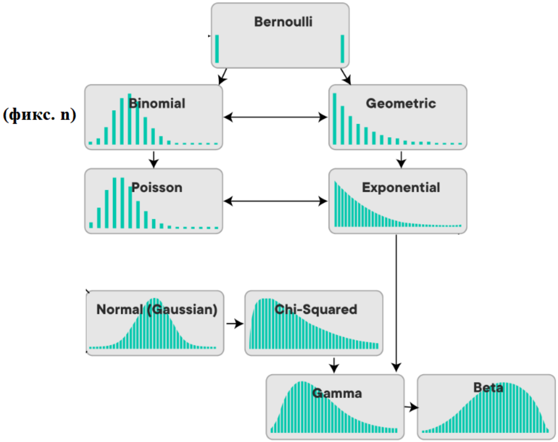
*Контрольный вопрос*. Чему равно $g(\theta)$?

*Ответ*. Чтобы получилась плотность распределения,

$$g(\theta) = \int h(x) \exp\left(\theta^T u(x)\right) dx.$$

# Семинар 1

Многие распределения лежат в экспоненциальном классе.




**Q**: Какого известного распределения нет?

## Семинар 1

НЕ лежат в экспоненциальном классе:

- Равномерные распределения $U[a, b]$.
- Смеси распределений = выпуклые комбинации плотностей

$$f(x) = \sum \alpha_i f_i(x), \qquad \alpha_j \geq 0, \qquad \sum_j \alpha_j = 1.$$

- Распределение Стьюдента.

# Семинар 1

**Q:** Показать, что биномиальное распределение $\mathrm{Bin}(n, p)$ при *фиксированном $n$* лежит в экспоненциальном семействе.

Плотность:

$$f(x, p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x \in \{0, 1, 2, \dots, n\}..$$

# Семинар 1

Хотим выражение

$$f(x|\theta) = \frac{h(x)}{g(\theta)} \exp\left(\theta^T u(x)\right),$$

В произведении нужно выделить 3 множителя:

- $h(x)$ — зависит только от $x$.
- $g(\theta)$ — зависит только от $\theta$.
- $\exp\left(\theta^T u(x)\right)$ — содержит всю зависимость между $x$ и параметрами.

# Семинар 1

Смотрим на

$$f(x,p) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} \exp\left[x \log p + (n-x)\log(1-p)\right].$$

# Семинар 1

Смотрим на

$$f(x,p) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} \exp\left[x \log p + (n-x)\log(1-p)\right].$$

Группируем слагаемые

$$f(x) = \binom{n}{x} \exp\left(x \log\left(\frac{p}{1-p}\right) + n\log(1-p)\right) =$$

$$= \frac{\binom{n}{x}}{\exp(-n\log(1-p))} \exp\left(x \log\left(\frac{p}{1-p}\right)\right)$$

# Семинар 1

$$f(x) = \frac{\binom{n}{x}}{\exp(-n\log(1-p))} \exp\left(x \log\left(\frac{p}{1-p}\right)\right) = \frac{h(x)}{g(\theta)} \exp\left(\theta^T u(x)\right),$$

**Ответ:**

- $h(x) = \binom{n}{x}$.

  Формально следует домножить на носитель $I(x \in \{0, 1, 2, \ldots, n\})$.

- $\theta = \log \frac{p}{1-p} \qquad \Rightarrow \qquad p = \frac{e^\theta}{1+e^\theta}$

- $u(x) = x$.

- $g(\theta) = \exp(n \log \frac{1}{1-p}) = \exp(n\log(1 + e^\theta))$.