Лекция 6

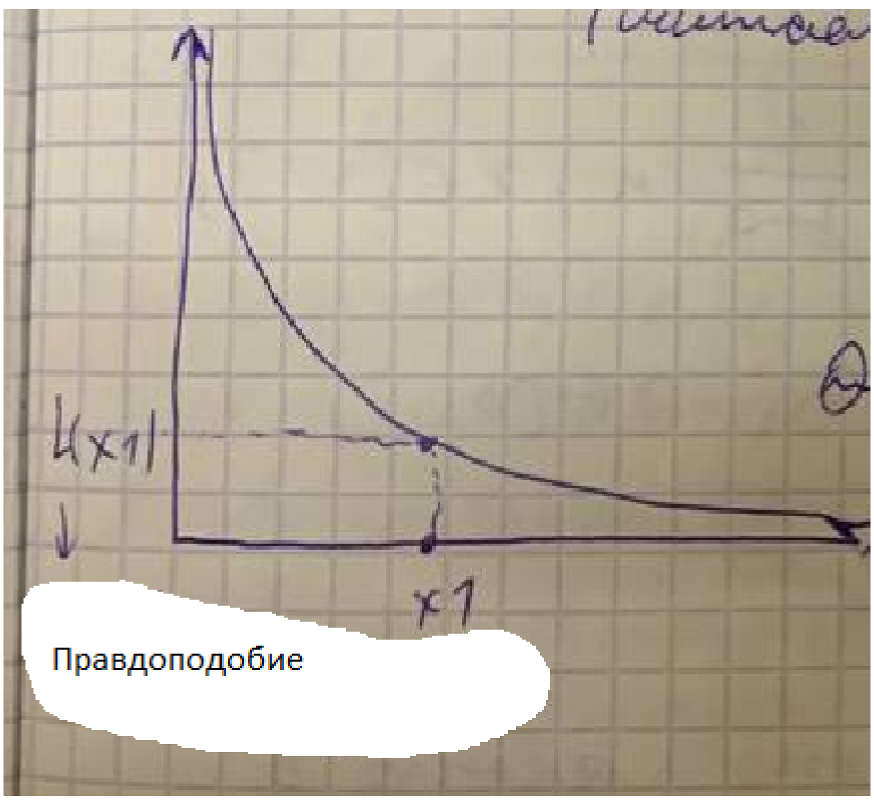
Есть генеральная совокупность, хотим оценивать её по выборке, но эти оценки зачастую оказываются смещёнными.

Метод максимального правдоподобия

Есть выборка и есть задачи:

1. Найти распределение. Предположим экспоненциальное
2. Подобрать параметр, чтобы распределение максимально точно подходила под выборку. То есть подобрать такое λ, чтобы для каждого значения выборки его вероятность была бы максимальной.

Решение:



$L(\lambda|x_0) = p(x_0, \lambda)$, где p - плотность вероятности.

$L(x_1, ..., x_n, \lambda) = L(\lambda|x_1, ..., x_n) = \prod_{i=1}^{n} p(x_i, \lambda) \to \max$

Экспоненциальное

$\prod_{i=1}^{n} p(x_i, \lambda) = \lambda^n \cdot e^{-\lambda \cdot (x_1 + .. + x_n)}$

$\ln = n \cdot \log \lambda - \lambda \cdot (x_1 + ... + x_n)$

$\frac{\partial \ln}{\partial \lambda} = \frac{n}{\lambda} - (x_1 + ... + x_n) = 0$ (так как ищем экстремум по λ)

$\lambda = \frac{1}{\bar{X}}$

Бернулли

$$\xi = \begin{cases} 0 \\ 1 \end{cases}$$

$$p(\xi) = \begin{cases} \theta, \ \xi = 1 \\ 1 - \theta, \ \xi = 0 \end{cases}$$

$$\ln(\xi) = \begin{cases} \ln \theta \\ \ln(1 - \theta) \end{cases}$$

$$L(\theta) = \theta^m \cdot (1 - \theta)$$

$$\ln(\theta) = m \cdot \log \theta + (n - m) \cdot \log(1 - \theta)$$

$$\ln'(\theta) = \frac{m}{\theta} - \frac{n - m}{1 - \theta} = 0 \Rightarrow \theta = \frac{m}{n}$$

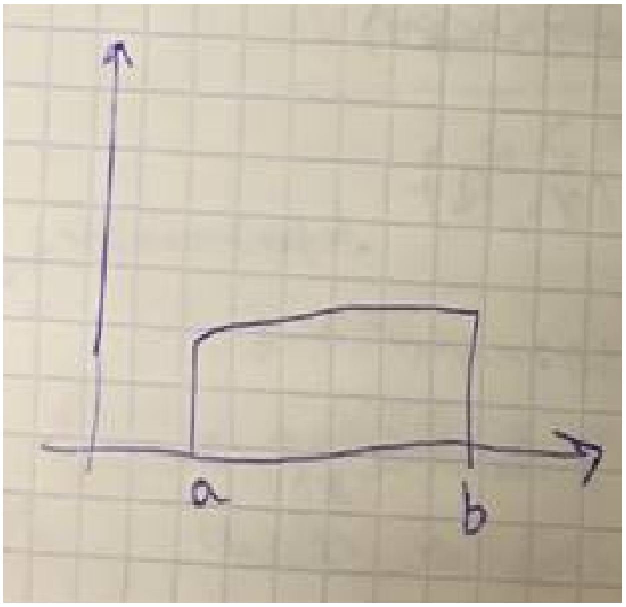
Нормальное распределение

$$L(\mu, \sigma | x) = \frac{1}{\sqrt{2 \cdot \pi} \cdot \sigma} \cdot e^{\frac{-(x - \mu)^2}{2 \cdot \sigma^2}}$$

$$L(\mu, \sigma | x_1, ..., x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2 \cdot \pi} \cdot \sigma} \cdot e^{\frac{-(x_i - \mu)^2}{2 \cdot \sigma^2}}$$

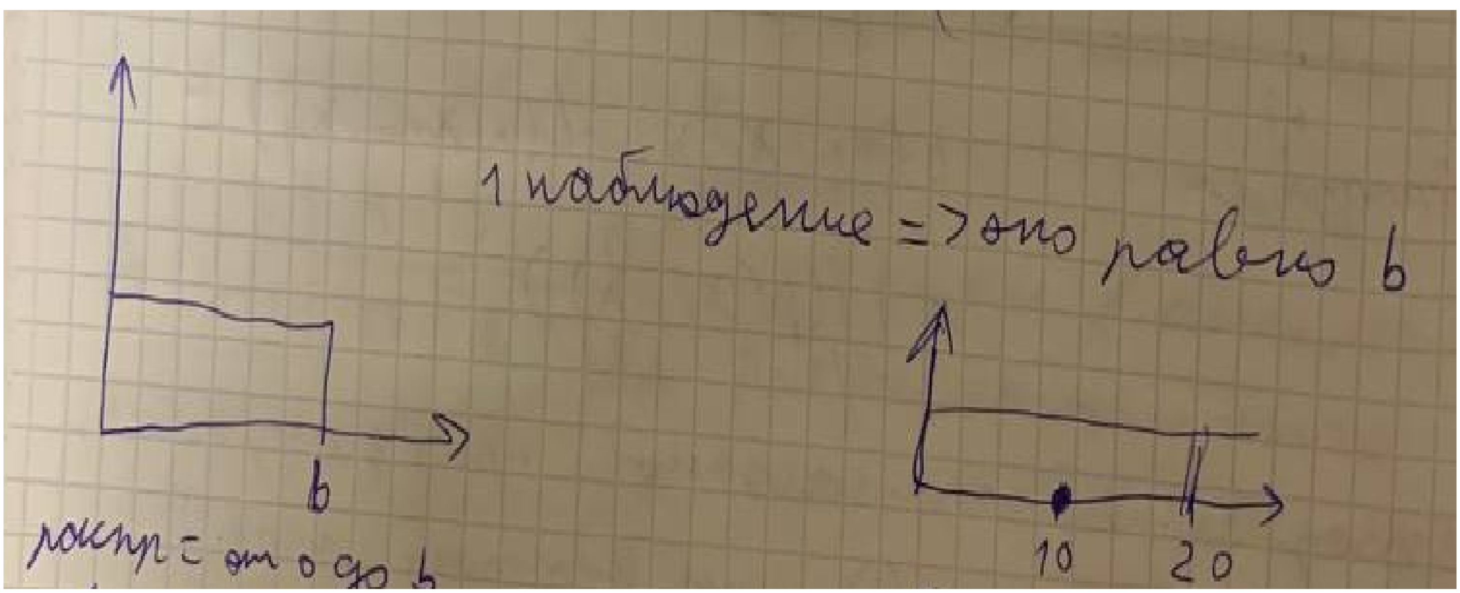
Максимальное значение достигается при $\mu = \bar{X}, \ \sigma = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}}$

Равномерное распределение



$$L(x_1, ..., x_n) = \begin{cases} \frac{1}{(b-a)}, \ x \in [a, b] \\ 0, \ x \notin [a, b] \end{cases}$$

При 1 эксперименте метод правдоподобности даст распределение от 0 до b, а метод моментов от 0 до $2 \cdot b$



Доверительные интервалы

Доверительный интервал для μ

Предположим, что μ - неизвестно, σ - известно

$Pr(\bar{X} - \frac{\sigma}{\sqrt{n}} \cdot Z_{\frac{\epsilon}{2}} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} \cdot Z_{1-\frac{\epsilon}{2}}) = 1 - \epsilon$ - доверительная вероятность (вероятность попасть в доверительный интервал).

Алгоритм:

1. Выбор ϵ.
2. Подсчёт среднего.
3. Подбор σ для соответствия заданной вероятности.

$1 - \epsilon$ - уровень доверия, вероятность того что μ попадёт в нужный интервал.

Если в гипотезах мы смотрели, чтобы значение принадлежало инервалу, то сейчас мы смотрим границы этого интревала.

Если σ^2 неизвестна, то $P_{\text{дов}} = Pr(\bar{X} - \frac{S}{\sqrt{n}} \cdot t^{n-1}_{\frac{\epsilon}{2}} < \mu < \bar{X} + \frac{S}{\sqrt{n}} \cdot t^{n-1}_{1-\frac{\epsilon}{2}})$

Доверительный интервал для σ^2

u_α - кватиль X^2 для $n-1$ степени свободы.

v_α - квантиль X^2 для n степеней свободы.

Если μ известно
$P_{\text{доверительный для дисперсии}}\{\frac{n \cdot s^2}{V_{1-\frac{\epsilon}{2}}} < \sigma^2 < \frac{n \cdot s^2}{V_{\frac{\epsilon}{2}}}\}$