MAXIMUM LIKELIHOOD ESTIMATORS

Idea: If we observe data sampled from a probability distribution depending from parameters θ and we want an estimator of θ, a good choiche is to consider $\hat{\theta}$ such that the observed data are "very likely".

Setting: Consider a population described by a random variable X with a law depending on some parameter $\theta \in \Theta$.

Example: $X \sim B_e(p) \Rightarrow \theta = p, \ \Theta = (0,1)$

Example: $X \sim B(n,p) \Rightarrow \theta = (n,p)$,
$$\Theta = (\mathbb{N} \setminus \{0\}) \times (0,1).$$

Let us assume now that the law that describes X is discrete.

Let $p(x;\theta)$ be the probability mass function:

$$p(x;\theta) = \mathbb{P}(X=x).$$

It depends on θ.

Assume that $(X_1, \dots, X_n)$ is a random sample drawn from the population. This means that $X_1, \dots, X_n$ are independent and identically distributed to X.

Assume that $(x_1, \dots, x_n) \in Range(X_1, \dots, X_n)$ are observations of the random sample (the data).
The probability that these data are indeed observed is:

$$\mathbb{P}(\{X_1 = x_1\} \cap \dots \cap \{X_n = x_n\}) =$$

$$= \mathbb{P}(\{X_1 = x_1\}) \cdot \dots \cdot \mathbb{P}(\{X_n = x_n\}) =$$

$$= p(x_1;\theta) \cdot \dots \cdot p(x_n;\theta) =: L(x_1, \dots, x_n;\theta)$$

This function is called <u>likelihood</u>.

We are interested in finding

$$\hat{\theta}(x_1, \dots, x_n) \in \underset{\theta \in \Theta}{\operatorname{argmax}}\, L(x_1, \dots, x_n ; \theta)$$

<u>Def</u>: The estimator (a random variable)

$$\hat{\theta}(X_1, \dots, X_n)$$

is called <u>maximum likelihood estimator</u>

<u>Example</u>: $X \sim Be(p)$

$$\mathbb{P}(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \end{cases}$$

The likelihood is

$$L(x_1, \dots, x_n ; \theta) = \prod_{i=1}^{n} \mathbb{P}(X = x_i) =$$

$$= \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i}$$

To maximize it, we take $-\log$:

$$-\log L(x_1, \dots, x_n; \theta) =$$

$$= -\sum_{i=1}^{n} \log\left(p^{x_i}(1-p)^{1-x_i}\right) =$$

$$= -\sum_{i=1}^{n}\left(x_i \log p + (1-x_i)\log(1-p)\right)$$

Let us impose:

$$0 = \frac{\partial L}{\partial p}(x_1, \dots, x_n; \theta) = -\sum_{i=1}^{n}\left(\frac{x_i}{p} - \frac{1-x_i}{1-p}\right) =$$

$$= -n\frac{\overline{x}}{p} + n\frac{(1-\overline{x})}{1-p}$$

$$\Rightarrow \frac{\overline{x}}{p} = \frac{1-\overline{x}}{1-p}$$

$$\Rightarrow (1-p)\overline{x} = p(1-\overline{x})$$

$$\Rightarrow \overline{x} - \cancel{\overline{x}p} = p - \cancel{\overline{x}p}$$

$$\Rightarrow p = \overline{x}$$

Hence $\hat{p}(X_1, \dots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a MLE.