Statistics
MDI220
1. Point Estimation

Thomas Bonald
Institut Polytechnique de Paris

2023 – 2024

The task of point estimation is to determine the parameter θ of a statistical model based on the observation x. We denote by $\hat{\theta}(x)$ this estimation.

1 Maximum likelihood

Assume that the statistical model is dominated. The probability density function p_θ gives the likelihood of each observation x. It is then natural to select the parameter θ that maximizes the likelihood of the observation. If several values of the parameter θ reach the maximum, one of them is selected arbitrarily.

> The Maximum Likelihood Estimator (MLE) is $\hat{\theta}(x) = \arg\max_{\theta\in\Theta} p_\theta(x)$.

Example. *For the Bernoulli model $\mathcal{P} = \{P_\theta \sim \mathcal{B}(\theta), \theta \in [0,1]\}$, we have $p_\theta(x) = \theta^x(1-\theta)^{1-x}$:*

- *If $x = 1$ then $p_\theta(x) = \theta$ and $\hat{\theta}(x) = 1$.*
- *If $x = 0$ then $p_\theta(x) = 1-\theta$ and $\hat{\theta}(x) = 0$.*

Thus $\hat{\theta}(x) = x$.

When n independent observations are available, say $x = (x_1, \ldots, x_n)$, the probability density function is a product and it is generally more convenient to use the log-likelihood.

> For n observations, the MLE is $\hat{\theta}(x) = \arg\max_{\theta\in\Theta} \log p_\theta(x) = \arg\max_{\theta\in\Theta} \sum_{i=1}^n \log p_\theta(x_i)$.

Example. *For the Bernoulli model with n observations, the log-likelihood is:*

$$\log p_\theta(x) = \sum_{i=1}^n (x_i \log\theta + (1-x_i)\log(1-\theta)) = S\log\theta + (n-S)\log(1-\theta),$$

with $S = \sum_{i=1}^n x_i$. The maximum is reached for:

$$\hat{\theta}(x) = \frac{1}{n}\sum_{i=1}^n x_i.$$

2 Method of moments

Another approach relies on the Strong Law of Large Numbers (SLLN). For simplicity, we assume that $\theta \in \mathbb{R}$. Let $x_1, \ldots, x_n$ be n observations, supposed to be i.i.d. samples of some random variable $X \sim P_\theta$. By the SLLN, the following approximation holds true for large values of n:

$$\mathrm{E}_\theta(X) \approx \frac{1}{n}\sum_{i=1}^n x_i.$$

The method of moments consists in finding the unique value of θ (if any) for which this approximation is exact.

The estimation of a real parameter θ by the method of moments is $\hat{\theta}(x) = f^{-1}(\frac{1}{n}\sum_{i=1}^n x_i)$, where the function $f : \theta \mapsto \mathrm{E}_\theta(X)$ is supposed to be injective.

Example. *For the Bernoulli model with n observations, we have $X \sim \mathcal{B}(\theta)$ so that $\mathrm{E}_\theta(X) = \theta$ and*

$$\hat{\theta}(x) = \frac{1}{n}\sum_{i=1}^n x_i.$$

3 Bias

The bias of an estimator $\hat{\theta}$ is the difference between its expected value (average over the observations) and the true parameter θ.

The bias of an estimator $\hat{\theta}$ is $b(\theta, \hat{\theta}) = \mathrm{E}_\theta(\hat{\theta}(X)) - \theta$. It depends on the parameter θ.

An estimator $\hat{\theta}$ is said to be *unbiased* if $b(\theta, \hat{\theta}) = 0$ for all $\theta \in \Theta$.

Example. *The estimator $\hat{\theta}(x) = \frac{1}{n}\sum_{i=1}^n x_i$ for the Bernoulli model is unbiased since*

$$\forall \theta \in [0, 1], \quad \mathrm{E}_\theta(\hat{\theta}(X)) = \frac{1}{n}\sum_{i=1}^n \mathrm{E}_\theta(X_i) = \theta.$$