I Maximum Likelihood Methods

Suppose that $X_1, \dots, X_n$ are i.i.d. random variables with common pdf $f(x, \theta)$, $\theta \in \Theta \subseteq \mathbb{R}^p$, $p \geq 1$.

1. Maximum Likelihood Estimation

Definition 1

The function $L : \Theta \to \mathbb{R}$ of the form

$$L(\theta) = L(\theta, \underline{x}) = \prod_{i=1}^{n} f(x_i, \theta), \quad \theta \in \Theta,$$

where $\underline{x} = (x_1, \dots, x_n)$ is called the likelihood function.

The function $\ell : \Theta \to \mathbb{R}$ of the form

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i, \theta), \quad \theta \in \Theta$$

is called the log-likelihood.

Example 1

Let $X_1, \dots, X_n$ denote a random sample from the distribution with pmf

$$p(x) = \begin{cases} \theta^x (1-\theta)^{1-x}, & x = 0, 1, \\ 0, & \text{elsewhere}, \end{cases}$$

where $0 < \theta < 1$. We have

$$P(\underline{X} = \underline{x}) = P((X_1, \dots, X_n) = (x_1, \dots, x_n)) = \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n - \sum_{i=1}^{n} x_i}.$$

Thus

$$L(\theta) = \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n - \sum_{i=1}^{n} x_i}, \quad \theta \in (0,1).$$

Problem: What value of θ maximize the probability $L(\theta)$ of obtaining ~~this particular~~ the observed sample $x_1, .., x_n$? Would it be a good estimate of θ?

We have

$$l(\theta) = \log L(\theta) = \left(\sum_{i=1}^{n} x_i\right) \log \theta + \left(n - \sum_{i=1}^{n} x_i\right) \log (1-\theta),$$

$$\frac{d l(\theta)}{d\theta} = \frac{\sum x_i}{\theta} - \frac{n - \sum x_i}{1-\theta} = 0,$$

$$(1-\theta)\sum_{i=1}^{n} x_i - \theta\left(n - \sum x_i\right) = \sum x_i - n\theta = 0 \Rightarrow \hat{\theta} = \frac{1}{n}\sum x_i$$

$$\frac{d^2 l(\theta)}{d\theta^2} = -\frac{\sum x_i}{\theta^2} - \frac{n - \sum x_i}{(1-\theta)^2} < 0$$

The statistic

$$\hat{\theta} = \bar{X}$$

is called the maximum likelihood estimator of θ.

Let θ_0 denote the true value of θ.

Assumptions (Regularity Conditions)

(R0): The pdfs are distinct, i.e., $\theta \neq \theta' \Rightarrow f(x_i; \theta) \neq f(x_i; \theta')$

(R1): The pdfs have common support for all θ.

(R2) $\theta_0 \in \operatorname{int} \Theta$.

Theorem 1

Under assumptions (R0) - (R1)

$$\lim_{n\to\infty} P_{\theta_0}\left[L(\theta_0, \underline{X}) > L(\theta, \underline{X})\right] = 1 \quad \text{for all } \theta \neq \theta_0.$$

Remark 1

Asymptotically, the likelihood function is maximized at the true value ϑ_0.

Definition 2

We say that $\hat{\theta} = \hat{\theta}(\underline{X})$ is a maximum likelihood estimator ~~of~~ (mle) of θ if

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\max}\, L(\theta, \underline{X}).$$

In other words

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta, \underline{X})$$

~~Remark 2~~

~~The mle can not exist or~~

Example 2

$X_1, \dots, X_n$, i.i.d $\quad X_i \sim f(x, \theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \mathbb{1}_{(0,+\infty)}(x)$

$$L(\theta) = \frac{1}{\theta^n} e^{-\frac{\sum_i x_i}{\theta}}$$

$$\ell(\theta) = -n \log \theta - \frac{1}{\theta} \sum_i x_i$$

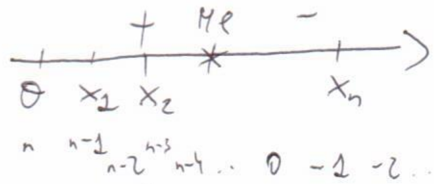
$$\frac{d\ell(\theta)}{d\theta} = -\frac{n}{\theta} + \frac{\sum_i x_i}{\theta^2} = 0 \iff n\theta = \sum_i x_i \iff \hat{\theta} = \bar{X}$$

$$\frac{d^2\ell(\theta)}{d\theta^2} = \frac{n}{\theta^2} - \frac{2\sum_i x_i}{\theta^3} = \frac{1}{\theta^3}\left(n\theta - 2\sum_i x_i\right)\Big|_{\theta = \bar{x}} = \frac{1}{\bar{x}^3}\left(n\bar{x} - 2n\bar{x}\right) < 0$$

Example 3

$X_1, \dots, X_n$ i.i.d $\quad X_i \sim f(x, \theta) = \frac{1}{2} e^{-|x-\theta|}$, $x, \theta \in \mathbb{R}$

$$L(\theta) = \left(\tfrac{1}{2}\right)^n e^{-\sum_i |x_i - \theta|}$$



$$\ell(\theta) = -n \log 2 - \sum_i |x_i - \theta|$$

$$\ell'(\theta) = \sum_i \operatorname{sgn}(x_i - \theta) = 0 \implies \hat{\theta} = \operatorname{med}\{X_1, \dots, X_n\}$$