I Maximum Likelihood Methods

Suppose that $X_1, \ldots, X_n$ 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 $\underline{x} = (x_1, \ldots, x_n)$ 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 $X_1, \ldots, X_n$ 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

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

Thus

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

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

We have

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

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

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

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

The statistic

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

is called the maximum likelihood estimator of θ.

Let θ_0 denote the true value of θ.

<u>Assumptions (Regularity Conditions)</u>

(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) $\theta_0 \in \operatorname{int} \Theta_0$.

<u>Theorem 1</u>

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

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

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

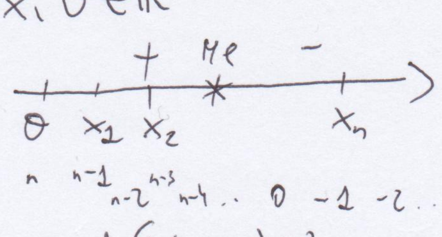
$$\frac{d\ell(\theta)}{d\theta} = -\frac{n}{\theta} + \frac{\sum x_i}{\theta^2} = 0 \iff n\theta = \sum x_i \iff \hat{\theta} = \bar{X}$$

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

Example 3

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

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



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

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

Example 4

$X_1, \dots, X_n$ i.i.d. $X_i \sim \mathcal{U}(0,\theta)$, $f(x,\theta) = \frac{1}{\theta} \mathbb{1}_{(0,\theta)}(x)$

$$L(\theta) = \left(\frac{1}{\theta}\right)^n \prod_{i=1}^{n} \mathbb{1}_{(0,\theta)}(x_i) = \left(\frac{1}{\theta}\right)^n \mathbb{1}_{[0,+\infty)}(x_{(1)})\, \mathbb{1}_{(-\infty,\theta]}(x_{(n)})$$

$\forall_i \quad 0 \leq x_i \leq \theta$

$0 \leq x_{(1)}$ & $x_{(n)} \leq \theta$

$$\hat{\theta} = X_{(n)}$$

Theorem 2.

Let $X_1, \dots, X_n$ be i.i.d. with the pdf $f(x,\theta)$, $\theta \in \Theta$. For a specified function $g: \Theta \to \mathbb{R}$, let $\eta = g(\theta)$ be a parameter of ~~interest~~ interest. Suppose $\hat{\theta}$ is the mle of θ. Then $g(\hat{\theta})$ is the mle of $\eta = g(\theta)$.

2. Rao-Cramér Lower Bound and Efficiency

Let X be a random variable with pdf $f(x,\theta)$, $\theta \in \Theta$, where Θ is a open set.

Assumptions (Additional Regularity Conditions)

(R3) The pdf $f(x,\theta)$ is twice differentiable as a function of θ.

(R4) The integral $\int_{\mathbb{R}} f(x,\theta)\,dx$ can be differentiated twice under the ~~sig~~ integral sign as a function of θ.

We have

$$1 = \int_{-\infty}^{+\infty} f(x,\theta)\,dx \quad \Big/ \frac{\partial}{\partial \theta}$$

$$0 = \int_{-\infty}^{+\infty} \frac{\partial f(x,\theta)}{\partial \theta}\,dx.$$

Equivalently,

$$0 = \int_{-\infty}^{+\infty} \frac{\frac{\partial f(x,\theta)}{\partial \theta}}{f(x,\theta)} f(x,\theta)\,dx = \int_{-\infty}^{+\infty} \frac{\partial \log f(x,\theta)}{\partial \theta} f(x,\theta)\,dx \quad (*)$$

Thus

$$\mathbb{E}\left[\frac{\partial \log f(X,\theta)}{\partial \theta}\right] = 0.$$

Furthermore, differentiate one more $(*)$, we obtain

$$0 = \int_{-\infty}^{+\infty} \frac{\partial^2 \log f(x,\theta)}{\partial \theta^2} f(x,\theta)\,dx + \int_{-\infty}^{+\infty} \frac{\partial \log f(x,\theta)}{\partial \theta} \underbrace{\frac{\partial \log f(x,\theta)}{\partial \theta} f(x,\theta)}_{\frac{\partial f(x,\theta)}{\partial \theta}}\,dx,$$

Therefore

$$-\int_{-\infty}^{+\infty} \frac{\partial^2 \log f(x,\theta)}{\partial \theta^2} f(x,\theta)\,dx = \mathbb{E}\left[\frac{\partial \log f(X,\theta)}{\partial \theta}\right]^2.$$

Definition 1

The number

$$I(\theta) = \mathbb{E}\left[\left(\frac{\partial \log f(X,\theta)}{\partial \theta}\right)^2\right]$$

is called Fisher information.

Corollary 1

Under the assumptions (R0) - (R4)

$$I(\theta) = -\int_{-\infty}^{\infty} \frac{\partial^2 \log f(x,\theta)}{\partial \theta^2} f(x,\theta)\,dx = Var\left[\frac{\partial \log f(X,\theta)}{\partial \theta}\right].$$

Example 1

$X \sim b(1,\theta)$, $f(x,\theta) = \theta^x (1-\theta)^{1-x}$

$$\log f(x,\theta) = x \log \theta + (1-x)\log(1-\theta)$$

$$\frac{\partial \log f(x,\theta)}{\partial \theta} = \frac{x}{\theta} - \frac{1-x}{1-\theta}$$

$$\frac{\partial^2 \log f(x,\theta)}{\partial \theta^2} = -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2}$$

Therefore,