Maximum Likelihood Asymptotics, Information, Cramer-Rao Lower Bound

A. Stepanov
D. Simpson
S. Culpepper

Definition: Maximum Likelihood Estimator (MLE)

p.m.f. or p.d.f. $f(x;\theta)$, $\theta \in \Omega$. Ω – parameter space.

Likelihood function for a sample of i.i.d. $X_1, \dots, X_n$,

$$L(\theta;\mathbf{x}) = L(\theta; x_1, \dots, x_n) = f(x_1, \dots, x_n; \theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

where $\mathbf{x} = (x_1, \dots, x_n)'$ is a vector of sample observations. The log-likelihood is,

$$\ell(\theta;\mathbf{x}) = \ln[L(\theta; x_1, \dots, x_n)] = \sum_{i=1}^{n} \ln[f(x_i;\theta)]$$

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) The true unknown point θ_0 is an interior point in Ω.

(R3) $f(x;\theta)$ is a twice differentiable function of θ.

(R4) $\int f(x;\theta)dx$ can be twice differentiable under the integral as a function of θ.

(R5) $\left|\frac{\partial^3}{\partial\theta^3}\ln[f(x;\theta)]\right| < M(x),\ \ E[M(x)] < \infty$

Theorem 6.1.3. Assume that $X_1, \dots, X_n$ satisfy regularity conditions (R0) to (R2), where θ_0 is the true parameter, and further that $f(x{:}\,\theta)$ is differentiable with respect to $\theta \in \Omega$, then,

$$\frac{\partial L(\theta;\mathbf{x})}{\partial\theta} = 0, \frac{\partial \ell(\theta;\mathbf{x})}{\partial\theta} = 0$$

has a solution $\hat{\theta}_n$, such that $\hat{\theta}_n \xrightarrow{P} \theta_0$.

Definition: The score function is $\frac{\partial}{\partial\theta}\ln[f(x;\theta)]$.

Note that,

$$E\left\{\frac{\partial}{\partial\theta}\ln[f(x;\theta)]\right\} = 0$$

Proof:

$$1 = \int f(x;\theta)dx$$

Taking a derivative of both sizes with respect to θ yields,

$$\Rightarrow 0 = \frac{\partial}{\partial\theta}\int f(x;\theta)dx = \int \frac{\partial}{\partial\theta}f(x;\theta)dx = \int \frac{\frac{\partial}{\partial\theta}f(x;\theta)}{f(x;\theta)}f(x;\theta)dx$$
$$= \int\left[\frac{\partial}{\partial\theta}\ln[f(x;\theta)]\right]f(x;\theta)dx = E\left\{\frac{\partial}{\partial\theta}\ln[f(x;\theta)]\right\} \quad \blacksquare$$

Definition: Fisher information, $I(\theta)$, is the variance of the score function,

$$I(\theta) = Var\left\{\frac{\partial}{\partial\theta}\ln[f(x;\theta)]\right\} = E\left\{\left(\frac{\partial}{\partial\theta}\ln[f(x;\theta)]\right)^2\right\} = -E\left\{\frac{\partial^2}{\partial\theta^2}\ln[f(x;\theta)]\right\}$$

Proof.

$$Var\left\{\frac{\partial}{\partial\theta}\ln[f(x;\theta)]\right\} = E\left\{\left(\frac{\partial}{\partial\theta}\ln[f(x;\theta)]\right)^2\right\} - \left(E\left\{\frac{\partial}{\partial\theta}\ln[f(x;\theta)]\right\}\right)^2$$

We showed above that,

$$\int\left[\frac{\partial}{\partial\theta}\ln[f(x;\theta)]\right]f(x;\theta)dx = 0$$

Let's take another derivative of both sides with respect to θ,

$$\frac{\partial}{\partial\theta}\int\left[\frac{\partial}{\partial\theta}\ln[f(x;\theta)]\right]f(x;\theta)dx = 0$$

$$\Rightarrow \int \frac{\partial}{\partial\theta}\left[\frac{\partial}{\partial\theta}\ln[f(x;\theta)]\right] f(x;\theta)dx = 0$$

Using the product rule yields,

$$\int \left[\frac{\partial^2}{\partial\theta^2}\ln[f(x;\theta)]\right] f(x;\theta)dx + \int \left[\frac{\partial}{\partial\theta}\ln[f(x;\theta)]\right]\frac{\partial}{\partial\theta}f(x;\theta) = 0$$

$$E\left\{\frac{\partial^2}{\partial\theta^2}\ln[f(x;\theta)]\right\} + \int \left[\frac{\partial}{\partial\theta}\ln[f(x;\theta)]\right]\frac{\frac{\partial}{\partial\theta}f(x;\theta)}{f(x;\theta)}f(x;\theta) = 0$$

$$\Rightarrow E\left\{\frac{\partial^2}{\partial\theta^2}\ln[f(x;\theta)]\right\} + E\left\{\left(\frac{\partial}{\partial\theta}\ln[f(x;\theta)]\right)^2\right\} = 0 \quad \blacksquare$$

Rao-Cramer Lower Bound: Consider $X_1, \dots, X_n$ iid $f(x;\theta)$ and a statistics $Y = u(X_1, \dots, X_n)$ such that $E(Y) = k(\theta)$, then,

$$Var(Y) \geq \frac{\left(k'(\theta)\right)^2}{nI(\theta)}.$$

Note that if Y is unbiased then $k(\theta) = \theta$, $k'(\theta) = 1$, and

$$Var(Y) \geq \frac{1}{nI(\theta)}.$$

Y is an efficient estimator of θ if and only if the variance of Y attains the Rao-Cramer lower bound.

Theorem 6.2.2. Assume $X_1, \dots, X_n$ are iid with pdf $f(x;\theta)$ for $\theta_0 \in \Omega$ and (R0) to (R5) are satisfied. Suppose Fisher information satisfies $0 < I(\theta_0) < \infty$. Then any consistent sequence of solutions of the mle satisfies,

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{D} N\left(0, \frac{1}{I(\theta_0)}\right)$$

$$\hat{\theta} \pm z_{\alpha/2}\sqrt{\frac{1}{nI\left(\hat{\theta}\right)}}$$

has an approximate $100(1-\alpha)\%$ confidence level for large n.

Intuition:

A second-order Taylor expansion of $\ell'\left(\hat{\theta}_n; \mathbf{x}\right)$ about θ_0 is,

$$\ell'\left(\hat{\theta}_n; \mathbf{x}\right) = \ell'(\theta_0; \mathbf{x}) + \left(\hat{\theta}_n - \theta_0\right)\ell''(\theta_0; \mathbf{x}) + \frac{1}{2}\left(\hat{\theta}_n - \theta_0\right)^2 \ell'''(\theta_n^*; \mathbf{x})$$

for θ_n^* between $\hat{\theta}_n$ and θ_0.

Note:

$$\ell'\left(\hat{\theta}_n; \mathbf{x}\right) = 0$$

$$\frac{1}{\sqrt{n}}\ell'(\theta_0; \mathbf{x}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln[f(x_i;\theta_0)] \xrightarrow{D} N\left(0, I(\theta_0)\right)$$

$$-\frac{1}{n}\ell''(\theta_0; \mathbf{x}) = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\ln[f(x_i;\theta_0)] \xrightarrow{P} I(\theta_0)$$

$$-\left(\hat{\theta}_n - \theta_0\right)\frac{1}{n}\ell'''(\theta_n^*; \mathbf{x}) \xrightarrow{P} 0$$

Example 1. Consider the $N(\mu, \sigma^2)$ distribution with σ^2 known.

a) Find $I(\mu)$.

$$f(x;\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$$

$$\ln[f(x;\mu)] = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}$$

$$\frac{\partial}{\partial\mu}\ln[f(x;\mu)] = \frac{x-\mu}{\sigma^2}, \qquad \frac{\partial^2}{\partial\mu^2}\ln[f(x;\mu)] = -\frac{1}{\sigma^2}$$

We have two options for computing $I(\mu)$:

$$I(\mu) = Var\left\{\frac{\partial}{\partial\mu}\ln[f(x;\mu)]\right\} = Var\left(\frac{x-\mu}{\sigma^2}\right) = \frac{1}{\sigma^2}$$

$$I(\mu) = -E\left\{\frac{\partial^2}{\partial\mu^2}\ln[f(x;\mu)]\right\} = \frac{1}{\sigma^2}$$

b) Is $\hat{\mu}$ an efficient estimator of μ?

The mle of μ is $\hat{\mu} = \bar{X}_n$.

$$Var(\hat{\mu}) = Var(\bar{X}_n) = \frac{\sigma^2}{n} = \frac{1}{nI(\mu)} \Rightarrow \hat{\mu} \text{ is an efficient estimator of } \mu.$$

c) What is the large sample distribution of $\hat{\mu}$?

$$\hat{\mu} = \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Example 2. Consider the $N(\mu, \sigma^2)$ distribution with μ known.

a) Find $I(\sigma)$.

$$f(x;\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$$

$$\ln[f(x;\sigma)] = -\frac{1}{2}\ln(2\pi) - \ln(\sigma) - \frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}$$

$$\frac{\partial}{\partial\sigma}\ln[f(x;\sigma)] = -\frac{1}{\sigma} + \frac{(x-\mu)^2}{\sigma^3}, \qquad \frac{\partial^2}{\partial\sigma^2}\ln[f(x;\sigma)] = \frac{1}{\sigma^2} - 3\frac{(x-\mu)^2}{\sigma^4}$$

$$I(\sigma) = -E\left\{\frac{\partial^2}{\partial\sigma^2}\ln[f(x;\sigma)]\right\} = -E\left\{\frac{1}{\sigma^2} - 3\frac{(x-\mu)^2}{\sigma^4}\right\} = \frac{2}{\sigma^2}$$

b) Find $I(\sigma^2)$.

$$f(x;\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$$

$$\ln[f(x;\sigma^2)] = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}$$

$$\frac{\partial\ln[f(x;\sigma^2)]}{\partial\sigma^2} = -\frac{1}{2\sigma^2} + \frac{1}{2}\frac{(x-\mu)^2}{(\sigma^2)^2}, \qquad \frac{\partial^2\ln[f(x;\sigma^2)]}{(\partial\sigma^2)^2} = \frac{1}{2(\sigma^2)^2} - \frac{(x-\mu)^2}{(\sigma^2)^3}$$

$$I(\sigma^2) = -E\left\{\frac{\partial^2}{(\partial\sigma^2)^2}\ln[f(x;\sigma^2)]\right\} = -E\left\{\frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6}\right\} = \frac{1}{2\sigma^4}$$

c) What is the asymptotic distribution of $S_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2}$?

The mle is $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$ and we know that $\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{D} N(0,2\sigma^4)$.

$$S_n = \sqrt{\frac{n}{n-1}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$$

$$g(x) = \sqrt{\frac{n}{n-1}}\sqrt{x}, \qquad g'(x) = \frac{1}{2}\sqrt{\frac{n}{n-1}}\frac{1}{\sqrt{x}}$$

$$\sqrt{n}(S_n - \sigma) \xrightarrow{D} N\left(0, \frac{1}{4}\frac{n}{n-1}\frac{1}{\sigma^2}2\sigma^4\right) \Rightarrow N\left(0, \frac{n}{n-1}\frac{\sigma^2}{2}\right)$$

Example 3. Let $X \sim Exponential(\theta)$. Find $I(\theta)$.

$$f(x;\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}, x > 0, \theta > 0$$

$$\ln[f(x;\theta)] = -\ln(\theta) - \frac{x}{\theta}$$

$$\frac{\partial}{\partial\theta}\ln[f(x;\theta)] = -\frac{1}{\theta} + \frac{x}{\theta^2}, \qquad \frac{\partial^2}{\partial\theta^2}\ln[f(x;\theta)] = \frac{1}{\theta^2} - 2\frac{x}{\theta^3}$$

We have two options for computing $I(\theta)$:

$$I(\theta) = Var\left\{\frac{\partial}{\partial\theta}\ln[f(x;\theta)]\right\} = Var\left[-\frac{1}{\theta} + \frac{x}{\theta^2}\right] = \frac{1}{\theta^2}$$

$$I(\mu) = -E\left\{\frac{\partial^2}{\partial\theta^2}\ln[f(x;\theta)]\right\} = \frac{1}{\theta^2}$$

Example 4. Let $X \sim Binomial(1, \theta)$, $0 < \theta < 1$.

a) Find $I(\theta)$.

$$f(x;\theta) = \theta^x (1-\theta)^{1-x}, \qquad 0 < \theta < 1, \qquad x = 0,1$$

$$\ln[f(x;\theta)] = x\ln(\theta) + (1-x)\ln(1-\theta)$$

$$\frac{\partial}{\partial\theta}\ln[f(x;\theta)] = \frac{x}{\theta} - \frac{1-x}{1-\theta} = \frac{x-\theta}{\theta(1-\theta)}, \qquad \frac{\partial^2}{\partial\theta^2}\ln[f(x;\theta)] = -\frac{x}{\theta^2} + \frac{1-x}{(1-\theta)^2}$$

We have two options for computing $I(\theta)$:

$$\begin{aligned} I(\theta) &= Var\left\{\frac{\partial}{\partial\theta}\ln[f(x;\theta)]\right\} \\ &= Var\left(\frac{x-\theta}{\theta(1-\theta)}\right) \\ &= \frac{1}{\theta(1-\theta)} \end{aligned} \qquad \begin{aligned} I(\theta) &= -E\left\{\frac{\partial^2}{\partial\theta^2}\ln[f(x;\theta)]\right\} \\ &= -E\left[-\frac{x}{\theta^2} + \frac{1-x}{(1-\theta)^2}\right] \\ &= \frac{1}{\theta(1-\theta)} \end{aligned}$$

b) Is $\hat{\theta}$ an efficient estimator θ?

The mle of theta is $\hat{\theta} = \bar{X}_n$, so $Var(\hat{\theta}) = \frac{\theta(1-\theta)}{n} = \frac{1}{nI(\theta)} \Rightarrow \hat{\theta}$ is an efficient estimator of θ.

c) What is the large sample distribution of $\hat{\theta}$?

$$\hat{\theta} = \bar{X}_n \sim N\left(\theta, \frac{\theta(1-\theta)}{n}\right)$$

Example 5. Let $X_1, \ldots, X_n$ be a random sample of size n from the distribution with probability density function,

$$f(x;\theta) = \begin{cases} \frac{1}{\theta} x^{\frac{1-\theta}{\theta}}, & 0 \le x \le 1 \\ 0, & otherwise \end{cases}$$

$$E[-\ln(X)] = \theta, \qquad Var[-\ln(X)] = \theta^2$$

a) Find $I(\theta)$.

$$\ln[f(x;\theta)] = -\ln(\theta) - \left(\frac{1}{\theta} - 1\right)\ln(x)$$

$$\frac{\partial}{\partial\theta}\ln[f(x;\theta)] = -\frac{1}{\theta} + \frac{\ln(x)}{\theta^2}, \qquad \frac{\partial^2}{\partial\theta^2}\ln[f(x;\theta)] = \frac{1}{\theta^2} - 2\frac{\ln(x)}{\theta^3}$$

$$\begin{aligned} I(\theta) &= Var\left\{\frac{\partial}{\partial\theta}\ln[f(x;\theta)]\right\} \\ &= Var\left(-\frac{1}{\theta} + \frac{\ln(x)}{\theta^2}\right) \\ &= \frac{1}{\theta^2} \end{aligned}$$

$$\begin{aligned} I(\theta) &= -E\left\{\frac{\partial^2}{\partial\theta^2}\ln[f(x;\theta)]\right\} \\ &= -E\left[\frac{1}{\theta^2} - 2\frac{\ln(x)}{\theta^3}\right] = \frac{1}{\theta^2} \end{aligned}$$

Example 6. Let $\lambda > 0$ and let $X_1, \dots, X_n$ be a random sample from the distribution with the probability density function, $f(x;\lambda) = 2\lambda^2 x^3 e^{-\lambda x^2},\ x > 0$.

Recall,

$$E(X^k) = \lambda^{-\frac{k}{2}}\Gamma\left(\frac{k}{2}+2\right), \qquad k > -4$$

$$Y = \sum_{i=1}^{n} X_i^2 \sim Gamma\left(\alpha = 2n, \theta = \frac{1}{\lambda}\right),$$

$$E\left(\frac{1}{Y}\right) = \frac{\lambda}{2n-1}, \qquad Var\left(\frac{1}{Y}\right) = \frac{\lambda^2}{(2n-1)^2(2n-2)}$$

Consider the estimator,

$$\hat{\hat{\lambda}} = \frac{2n-1}{\sum_{i=1}^{n} x_i^2}$$

Is $\hat{\hat{\lambda}}$ an efficient estimator of λ? If not, find its efficiency.

$$Var\left(\hat{\hat{\lambda}}\right) = Var\left(\frac{2n-1}{Y}\right) = \frac{\lambda^2}{2n-2}$$

$$\ln[f(x;\lambda)] = \ln 2 + 2\ln(\lambda) + 3\ln x - \lambda x^2$$

$$\frac{\partial}{\partial\lambda}\ln[f(x;\lambda)] = \frac{2}{\lambda} - x^2, \qquad \frac{\partial^2}{\partial\lambda^2}\ln[f(x;\lambda)] = -\frac{2}{\lambda^2}$$

$$\begin{aligned} I(\lambda) &= Var\left\{\frac{\partial}{\partial\lambda}\ln[f(x;\lambda)]\right\} \\ &= Var\left(\frac{2}{\lambda} - x^2\right) \\ &= E(x^4) - [E(x^2)]^2 \\ &= 3\lambda^{-2} - [\lambda^{-1}]^2 = \frac{2}{\lambda^2} \end{aligned}$$

$$\begin{aligned} I(\lambda) &= -E\left\{\frac{\partial^2}{\partial\lambda^2}\ln[f(x;\lambda)]\right\} \\ &= -E\left[-\frac{2}{\lambda^2}\right] = \frac{2}{\lambda^2} \end{aligned}$$

Rao-Cramer lower bound is $\frac{\lambda^2}{2n}$ so $\hat{\hat{\lambda}}$ is not efficient. It's efficiency is,

$$\frac{2n-2}{2n} = \frac{n-1}{n} \to 1\ as\ n \to \infty$$