Stat 4620: Day 16

(3/21)

Sec. 6.1 (con't.): Maximum Likelihood Estimation

Theorem 5. *Let $\mathbf{x} = (x_1, \ldots, x_n)$ be a random sample from a density $f(x;\theta)$, $\theta \in \Omega$. Let $\eta = g(\theta)$. Suppose that $\hat{\theta}$ is the mle of θ, then $g(\hat{\theta})$ is the mle of $\eta = g(\theta)$.*

Proof. Case 1: Suppose that $g(\cdot)$ is a one-to-one function. Then

$$L(\theta; \mathbf{x}) = \prod f(x_i;\theta) = \prod f(x_i; g^{-1}(\eta))$$

But the maximum occurs when $g^{-1}(\hat{\eta}) = \hat{\theta}$. This implies $\hat{\eta} = g(\hat{\theta})$. For example, suppose that $x_1, \ldots, x_n$ is a random sample from $N(\mu, 1)$ and $\overline{x} = 5.2$, say. Then

$$L(\mu; \mathbf{x}) = (2\pi)^{-n/2} e^{-(1/2)\sum(x_i - \mu)^2}$$

is maximum when $\hat{\mu} = 5.2$. If the parameter of interest is $\eta = e^{\mu}$, then

$$L(\eta; \mathbf{x}) = (2\pi)^{-n/2} e^{-(1/2)\sum(x_i - \ln\eta)^2}$$

is maximum when $\ln\hat{\eta} = 5.2$. This implies $\hat{\eta} = e^{5.2}$.

Case 2: (Heuristic) Suppose that $g(\cdot)$ is not one-to-one. For example, suppose the parameter of interest in our example above is $\eta = \cos(\mu)$. Note that μ is not a function of η, since $1/2 = \cos(\pi/3) = \cos(\pi/3 + 2\pi) = \cos(\pi/3 + 4\pi)$ implies that

$$\cos^{-1}(1/2) = ?$$

However, if we go ahead and write the likelihood in terms of η

$$L(\eta; \mathbf{x}) = (2\pi)^{-n/2} e^{-(1/2)\sum(x_i - \cos^{-1}(\eta)^2}$$

then this is maximized when $\hat{\eta}$ is such that $\cos^{-1}(\hat{\eta}) = 5.2$, which implies that

$$\hat{\eta} = \cos(5.2) = .4685 \text{ uniquely!}$$

□

Theorem 6. *Suppose that $\mathbf{x} = (x_1, \ldots, x_n)$ satisfy assumptions (R0) and (R1), plus*

(R2) the true value θ_0 is an interior point of Ω

then the likelihood equation

$$\frac{\partial}{\partial\theta} L(\theta) = 0$$

or equivalently

$$\frac{\partial}{\partial\theta} l(\theta) = 0$$

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

Proof. Let the constant $a > 0$ be such that $(\theta_0 - a, \theta_0 + a) \subset \Omega$. Let

$$S_n = \{\mathbf{x} : l(\theta_0; \mathbf{x}) > l(\theta_0 - a; \mathbf{x})\} \cap \{\mathbf{x} : l(\theta_0; \mathbf{x}) > l(\theta_0 + a; \mathbf{x})\}$$

Then $P(S_n) \to 1$ by Theorem 4 of Day 15 notes. So with probability approaching 1, there exists $\hat{\theta}_n$ such that

$$l'(\hat{\theta}_n) = 0 \text{ and } |\hat{\theta}_n - \theta_0| < a$$

Thus $P[|\hat{\theta}_n - \theta_0| < a] \to 1$, or $\hat{\theta}_n \xrightarrow{P} \theta_0$. □

Corollary (Consistency). *Suppose that $\mathbf{x} = (x_1, \dots, x_n)$ satisfy assumptions (R0)-(R2), and the likelihood equation has a unique solution $\hat{\theta}_n$. Then $\hat{\theta}_n \xrightarrow{P} \theta_0$.*

Sec. 6.2: Rao-Cramer Lower Bound and Efficiency

Example (Poisson) Let $\mathbf{x} = (x_1, x_2, x_3) = (5, 0, 1)$ be observations from a Poisson distribution with mean θ. Then the likelihood and log-likelihood are

$$L(\theta; \mathbf{x}) = \prod p(x_i; \theta) = \left(\frac{e^{-\theta}\theta^5}{5!}\right)\left(\frac{e^{-\theta}\theta^0}{0!}\right)\left(\frac{e^{-\theta}\theta^1}{1!}\right) = \frac{e^{-3\theta}\,\theta^6}{5!\,0!\,1!}$$

$$l(\theta) = \ln L(\theta) = -3\theta + 6\ln\theta - \ln(5!\,0!\,1!)$$

Comment:

- The shape of $L(\theta)$ and $\ln L(\theta)$ change with the data
- We are trying to find the value of θ that maximizes $L(\theta)$, so *steeper is better*.

In general, if $x_1, \dots, x_n$ is a random sample from Poisson(θ),

$$\ln L(\theta) = \ln \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!} = \ln \frac{e^{-n\theta}\theta^{\sum x_i}}{\prod x_i!} = -n\theta + \sum x_i \ln\theta - \sum \ln x_i!$$

$$\frac{\partial}{\partial\theta}\ln L(\theta) = -n + \frac{\sum x_i}{\theta} \equiv 0 \Rightarrow \hat{\theta} = \frac{\sum x_i}{n} = \overline{x}$$

Fisher information

First, we derive some relationships.

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

Taking derivative of both sides with respect to θ

$$0 = \int \frac{\partial}{\partial\theta} f(x;\theta)dx = \int \left[\frac{\frac{\partial}{\partial\theta} f(x;\theta)}{f(x;\theta)}\right] f(x;\theta)dx = \int \left[\frac{\partial}{\partial\theta}\ln f(x;\theta)\right] f(x;\theta)dx \quad (1)$$

Which gives Equation (6.2.2) in textbook

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

Example (Poisson con't.) Since $f(X;\theta) = \frac{e^{-\theta}\theta^X}{X!}$, then

$$E\left[\frac{\partial}{\partial\theta}\ln f(X;\theta)\right] = E\left[\frac{\partial}{\partial\theta}(-\theta + X\ln\theta - \ln X!)\right] = E\left[-1+\frac{X}{\theta}\right] = -1+\frac{\theta}{\theta} = 0$$

Taking derivative of both sides of equation (1),

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

which gives the following equivalent expressions for the *Fisher information* of X

$$-E\left[\frac{\partial}{\partial\theta^2}\ln f(X;\theta)\right] = E\left[\frac{\partial}{\partial\theta}\ln f(X;\theta)\right]^2 = V\left[\frac{\partial}{\partial\theta}\ln f(X;\theta)\right] \equiv \mathbf{I}(\theta)$$

Example (Poisson con't.) $f(x;\theta) = \frac{e^{-\theta}\theta^x}{x!}$

$$\ln f(x;\theta) = -\theta + x\ln\theta - \ln x!$$

$$\frac{\partial}{\partial\theta}\ln f(x;\theta) = -1 + \frac{x}{\theta}$$

$$\frac{\partial}{\partial\theta^2}\ln f(x;\theta) = -\frac{x}{\theta^2}$$

Two ways to calculate the Fisher information of X:

1. $\mathbf{I}(\theta) = -E\left[\frac{\partial}{\partial\theta^2}\ln f(X;\theta)\right] = -E\left[\frac{X}{\theta^2}\right] = \frac{\theta}{\theta^2} = \frac{1}{\theta}$
2. $\mathbf{I}(\theta) = V\left[\frac{\partial}{\partial\theta}\ln f(X;\theta)\right] = V\left[-1+\frac{X}{\theta}\right] = \left(\frac{1}{\theta}\right)^2 V(X) = \left(\frac{1}{\theta}\right)^2\theta = \frac{1}{\theta}$

Fisher information of a random sample

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be independent random variables with density $f(x;\theta)$. The Fisher information of the random sample $\mathbf{X}$ is defined as

$$\mathbf{I}_n(\theta) = V\left[\frac{\partial}{\partial\theta}\ln L(\theta;\mathbf{X})\right]$$

Since the variables are independent,

$$\mathbf{I}_n(\theta) = V\left[\frac{\partial}{\partial\theta}\sum_{i=1}^n \ln f(\theta;X_i)\right] = V\left[\sum_{i=1}^n \frac{\partial}{\partial\theta}\ln f(\theta;X_i)\right] = \sum_{i=1}^n V\left[\frac{\partial}{\partial\theta}\ln f(\theta;X_i)\right] = n\mathbf{I}(\theta)$$

Rao-Cramer Lower Bound

Theorem 7. *Let $x_1, \ldots, x_n$ be a random sample from $f(x;\theta), \theta \in \Omega$. Assume that (R0)-(R4) hold. Let $Y = U(x_1, \ldots, x_n)$ be a statistic with mean $k(\theta)$. Then*

$$V(Y) \geq \frac{[k'(\theta)]^2}{n\mathbf{I}(\theta)}$$

Corollary. *If $E(Y) = \theta$, then*

$$V(Y) \geq \frac{1}{n\mathbf{I}(\theta)}$$

Implication: Allows optimality results.

Example (Poisson con't.) $f(x;\theta) = \frac{e^{-\theta}\theta^x}{x!}$. Let $U(x_1, \ldots, x_n) = \overline{x}$. Note that $E(U) = \theta$. RCLB says that if Y is any unbiased estimator of θ, then

$$V(Y) \geq \frac{1}{n\mathbf{I}(\theta)} = \frac{1}{n(1/\theta)} = \frac{\theta}{n}$$

But $V(U) = V(\overline{x}) = \theta/n$, so $\overline{x}$ is *minimum variance unbiased.*

Proof.

$$k(\theta) = E\left[U(X_1, \ldots, X_n)\right] = \int \cdots \int U(x_1, \ldots, x_n) f(x_1;\theta) \cdots f(x_n;\theta) dx_1 \cdots dx_n$$

$$\begin{aligned}
k'(\theta) &= \int \cdots \int U(x_1, \ldots, x_n) \left[\sum_{i=1}^{n} \frac{f'(x_i;\theta)}{f(x_i;\theta)}\right] f(x_1;\theta) \cdots f(x_n;\theta) dx_1 \cdots dx_n \\
&= E\left[U(X_1, \ldots, X_n) \sum_{i=1}^{n} \frac{f'(X_i;\theta)}{f(X_i;\theta)}\right] \\
&= E\left[U(X_1, \ldots, X_n) \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f(X_i;\theta)\right] \\
&= \rho\left(U(X_1, \ldots, X_n), \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f(X_i;\theta)\right) \sqrt{V\left(U(X_1, \ldots, X_n)\right)} \sqrt{V\left(\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f(X_i;\theta)\right)}
\end{aligned}$$

$$\left[k'(\theta)\right]^2 \leq V\left(U(X_1, \ldots, X_n)\right) V\left(\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f(X_i;\theta)\right) = V\left(U(X_1, \ldots, X_n)\right) n\mathbf{I}(\theta)$$

which implies

$$V\left(U(X_1, \ldots, X_n)\right) \geq \frac{[k'(\theta)]^2}{n\mathbf{I}(\theta)}$$

□

Comments:

1. $\frac{\partial}{\partial\theta} \ln f(X;\theta)$ is called the *score function* of X
2. $E(\text{score function}) = 0$ and $V(\text{score function}) = \mathbf{I}(\theta)$
3. $\hat{\theta}_{mle}$ is solution to $\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f(X;\hat{\theta}) = 0$