Recorded Lecture for Friday April 16 — Lecture 37

Large Sample Properties of ML

$Y_1\ Y_2\ \ldots\ Y_m\ \ldots$ iid seq. $Y_k \in \Gamma$ and want to study props of ML est as $m \to \infty$.

$\hat{\theta}_{ML}(y, m)$

y: complete seq. of observations

m: number of measurements.

Questions of Interest

1) Is the ML estimator consistent: $\hat{\theta}_{ML} \to \theta_*$ as $m \to \infty$ (θ_*: true param.)

2) $Var_\theta\{\hat{\theta}_{ML}\}$ and its relationship to CRLB as $m \to \infty$.

3) Asymptotic dist. of $\hat{\theta}_{ML}$ as $m \to \infty$.

Thm: A ML estimator is a consistent estimator.

Sketch of Proof $Y^m = [Y_1^T, Y_2^T \dots Y_m^T]^T$ and say $\hat{\theta}_{ML}$ is the ML est. of a true param. θ_* ie

$$\hat{\theta}_{ML} = \underset{\theta}{\operatorname{argmax}} \log f_\theta(y^m)$$

the solution

$$\nabla_\theta \log f_\theta(y^m)\Big|_{\theta = \hat{\theta}_{ML}} = 0 .$$

As in the proof of CRLB

$$0 = \nabla_\theta 1 = \nabla_\theta \int f_\theta(y^m)\, dy^m = \int \nabla_\theta f_\theta(y^m)\, dy^m$$

$$= \int \left(\nabla_\theta \log f_\theta(y^m)\right) f_\theta(y^m)\, dy^m = E_\theta\left\{\nabla_\theta \log f_\theta(Y^m)\right\}$$

Expand $\nabla_\theta \log f_\theta(y^m)\Big|_{\theta = \hat{\theta}_{ML}}$ in a Taylor Series about the true param θ_*.

Notation: $\Psi(y^m, \theta) = \nabla_\theta \log f_\theta(y^m)$ gradient

$\Psi'(y^m, \theta) = \nabla_\theta^2 \log f_\theta(y^m)$ Hessian

Taylor Series

$$\Psi(y^m, \hat{\theta}_{ML}) = \Psi(y^m, \theta_*) + \Psi'(y^m, \tilde{\theta})(\hat{\theta}_{ML} - \theta_*). \quad (1)$$

for some $\tilde{\theta} = \lambda \theta_* + (1-\lambda)\hat{\theta}_{ML}$ $\quad 0 \le \lambda \le 1.$

We assume that $\hat{\theta}_{ML}$ satisfies likelihood eqn: That is

$$\Psi(y^m, \hat{\theta}_{ML}) = 0$$

Solve (1)

$$\Psi(y^m, \theta_*) = -\Psi'(y^m, \tilde{\theta})(\hat{\theta}_{ML} - \theta_*)$$

From indep. of the samples $Y_1 \dots Y_m$

$$f_\theta(y^m) = \prod_{i=1}^{m} f^*_\theta(y_i)$$

$$\Psi(y^m, \theta) = \sum_{i=1}^{m} \nabla_\theta \log f^*_\theta(y_i)$$

$$= \sum_{i=1}^{m} \varphi(y_i, \theta)$$

Similarly $\quad \Psi'(y^m, \theta) = \sum_{i=1}^{m} \varphi'(y_i, \theta)$

Then we use the strong law

$$\frac{1}{m} \sum_{i=1}^{m} \varphi(Y_i, \theta) \longrightarrow E_\theta\{\varphi(Y_1, \theta)\} = 0$$

$$\frac{1}{m} \sum_{i=1}^{m} \varphi'(Y_i, \theta) \longrightarrow E_\theta\{\varphi'(Y_1, \theta)\}$$

as $m \to \infty$ (with prob. one under P_θ).

Now go back to (1) and take lim. $m \to \infty$.

Left side of (1) $\to 0$ with prob one as $m \to \infty$.

Therefore

$$\lim_{m\to\infty} \Psi'(Y^m, \tilde{\theta})(\hat{\theta}_{ML} - \theta_*) = 0 \quad \text{wp1 under } P_{\theta_*}$$

Then assuming that all these Hessians are pos. def

$$\hat{\theta}_{ML} \longrightarrow \theta_* \quad \text{with prob. } 1.$$

Re: Asymptotically Unbiased

Oftentimes people prove $\hat{\theta}_{ML} \to \theta_*$ in P_{θ_*} prob.

If so would we be able to claim:

$$\lim_{n\to\infty} E_{\theta_*}\{\hat{\theta}_n(Y)\} = E_{\theta_*}\left\{\lim_{n\to\infty}\hat{\theta}_n(Y)\right\} = \theta_*$$

Typically such interchanges are not allowed without further assumptions.

One such assumption: Assume or show the existence of a r.v. X s.t. $E_{\theta_*}\{|X|\} < \infty$

and

$$|\hat{\theta}(Y)| \leq X \quad \forall n \quad \text{wp1 under } P_{\theta_*}$$

Then can use <u>dominated</u> <u>convergence</u> thm to prove unbiased.

Asymptotically Efficient → Argue plausibility via looking at the asymptotic dist of

$$\hat{\theta}_n - \theta_*$$

Asymptotic Normality Thm

$\{Y_k\}_{k=1}^{\infty}$ an iid seq., each with marg. pdf $f_{\theta_*}(\cdot)$.

Let $\{\hat{\theta}_n\}_{n=1}^{\infty}$ be a consistent seq. of roots of the likelihood eqn.

Also assume ⟶ see Poor text.

Then

$$P_{\theta_*}\left(\sqrt{n\, i_{\theta_*}}\left[\hat{\theta}_n - \theta_*\right] \leq x\right) \longrightarrow \Phi(x)$$

↳ cdf of standard gaussian.

$$\Longrightarrow \therefore \sqrt{n}\left[\hat{\theta}_n - \theta_*\right] \to N\left(0, 1/i_{\theta_*}\right).$$

in distribution.