Consistent Estimators

Def.: Let $\{\hat{\Theta}_n\}$ be a sequence of estimators, where $\hat{\Theta}_n$ is based on a sample of size n.

The sequence is consistent for θ if $\hat{\Theta}_n$ converges in probability to θ as $n \to \infty$:

for any $\varepsilon > 0$, $\lim_{n\to\infty} P(|\hat{\Theta}_n - \theta| > \varepsilon) = 0$

(or $\lim_{n\to\infty} P(|\hat{\Theta}_n - \theta| < \varepsilon) = 1$)

Intuition: If n is large enough, ~~p~~ probability 1 that $\hat{\Theta}_n$ is very close to θ.

Example: The law of large numbers says that $\bar{X}$ is a consistent estimator of μ.

Theorem: The method of moments estimator is consistent.

Theorem: The maximum likelihood estimator is consistent.

Sketch of proof:

- $\hat{\Theta}_n^{MLE}$ maximizes $\frac{1}{n} \ell(\theta | X_1, \ldots, X_n)$ (definition)
- For any θ, by LLN
  $\frac{1}{n} \ell(\theta | X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} \ell(\theta | X_i) \to E[\ell(\theta | X_i)]$
- The true parameter value θ_0 maximizes $E[\ell(\theta | X_i)]$

↳ as $n \to \infty$, $\frac{1}{n} \ell(\theta | X_1, \ldots, X_n)$ becomes more and more like $E[\ell(\theta | X_i)]$

↳ as this happens, the maximizer of $\frac{1}{n} \ell(\theta | X_1, \ldots, X_n)$ becomes closer and closer to $E[\ell(\theta | X_i)]$

↳ $\hat{\Theta}^{MLE} \to \theta_0$.

Non-proof that θ_0 maximizes $E[\ell(\theta|X_i)]$:

As one step in proof that $\hat{\Theta} \sim \text{Normal}(0, \frac{1}{nI(\theta_0)})$ for large n, we showed that $E[\frac{d}{d\theta}\ell(\theta|X_i)|_{\theta=\theta_0}] = 0$

$$0 = E\left[\frac{d}{d\theta}\ell(\theta|X_i)|_{\theta=\theta_0}\right]$$

$$= \int \frac{d}{d\theta}\ell(\theta|x_i)|_{\theta=\theta_0} \cdot f_{X_i}(x_i|\theta_0)\, dx_i$$

$$= \frac{d}{d\theta}\int \ell(\theta|x_i) \cdot f_{X_i}(x_i|\theta_0)\, dx_i \Big|_{\theta=\theta_0}$$

$$= \frac{d}{d\theta} E\left[\ell(\theta|X_i)\right]\Big|_{\theta=\theta_0}$$

↳ slope of $E[\ell(\theta|X_i)]$ is 0 at $\theta=\theta_0$.

↳ more work to show a maximum.

Cramér-Rao Lower Bound (CRLB)

Let $X_1, \ldots, X_n$ be iid random variables with pdf $f_X(x|\theta)$, and let $T = g(X_1, \ldots, X_n)$ be an unbiased estimator of θ.

Then, if $f_X(x|\theta)$ satisfies "regularity conditions",

$$Var(T) \geq \frac{1}{n I_i(\theta)}$$

Proof in book. Uses familiar facts:

- $Var\left[\frac{d}{d\theta} \ell(\theta | X_i)\right] = I_i(\theta)$, so
  $Var\left[\sum_{i=1}^{n} \frac{d}{d\theta} \ell(\theta|X_i)\right] = n\, I_i(\theta)$
- Exchange differentiation & integration if $f_{X_i}(x_i|\theta)$ smooth.

<u>Example:</u> Suppose $X_1, \ldots, X_n \sim \text{Poisson}(\lambda)$.

$f_{X_i}(x_i|\lambda) = e^{-\lambda} \cdot \frac{\lambda^{x_i}}{x_i!}$, MLE is $\hat{\lambda} = \bar{X}$, &

$E(X_i) = Var(X_i) = \lambda$

- Find Fisher information from one observation:
  - $\frac{d}{d\lambda} \log\left[f_{X_i}(x_i|\lambda)\right] = \frac{d}{d\lambda}\left[-\lambda + x_i \cdot \log(\lambda) - \log(x_i!)\right]$
    $= -1 + \frac{x_i}{\lambda}$

    $\frac{d^2}{d\lambda^2} \log\left[f_{X_i}(x_i|\lambda)\right] = \frac{-x_i}{\lambda^2}$

  $I_i(\lambda) = -E\left[\frac{d^2}{d\lambda^2}\log\left\{f_{X_i}(X_i|\lambda)\right\}\Big|_{\lambda=\lambda}\right] = -E\left[\frac{-X_i}{\lambda^2}\right] = \frac{1}{\lambda^2} E(X_i) = \frac{1}{\lambda^2} \cdot \lambda = \frac{1}{\lambda}$

- $E[\hat{\lambda}] = E\left[\frac{1}{n}\sum X_i\right] = \frac{1}{n} \cdot n\lambda = \lambda$
- $Var(\hat{\lambda}) = Var\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2} \cdot n \cdot Var(X_i) = \frac{1}{n} \cdot \lambda = \frac{\lambda}{n} = \frac{1}{n \cdot \frac{1}{\lambda}} = \frac{1}{n \cdot I_i(\lambda)}$
- By CRLB, no <u>unbiased</u> estimator has lower variance.