Math 408 - Mathematical Statistics

Lecture 28. Efficiency and the Cramer-Rao Lower Bound

April 10, 2013

Agenda

- Mean Squared Error
- Cramer-Rao Inequality
- Example: Poisson Distribution
- Summary

Measure of Efficiency: Mean Squared Error

In most estimation problems, there are many possible estimates $\hat{\theta}$ of θ.
For example, the MoM estimate $\hat{\theta}_{\mathrm{MoM}}$ or the MLE estimate $\hat{\theta}_{\mathrm{MLE}}$.

Question: How would we choose which estimate to use?

Qualitatively, it is reasonable to choose that estimate whose distribution is most highly concentrated about the true parameter value θ_0. To make this idea work, we need to define a quantitative measure of such concentration.

Definition

The **mean squared error** of $\hat{\theta}$ as an estimate of θ_0 is

$$\boxed{\mathrm{MSE}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta_0)^2]}$$

- The mean squared error can be also written as follows:

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{V}[\hat{\theta}] + \underbrace{(\mathbb{E}(\hat{\theta}) - \theta_0)^2}_{\text{squared bias}}$$

- If $\hat{\theta}$ is unbiased, then $\mathrm{MSE}(\hat{\theta}) = \mathbb{V}[\hat{\theta}]$.

Cramer-Rao Inequality

- Given two unbiased estimates, $\hat{\theta}$ and $\tilde{\theta}$, the **efficiency** of $\hat{\theta}$ relative to $\tilde{\theta}$ is defined to be

$$\mathrm{eff}(\hat{\theta}, \tilde{\theta}) = \frac{\mathbb{V}[\tilde{\theta}]}{\mathbb{V}[\hat{\theta}]}$$

- $\hat{\theta}$ is more efficient than $\tilde{\theta}$ $\Leftrightarrow$ $\mathrm{eff}(\hat{\theta}, \tilde{\theta}) > 1$
- In general, the mean squared error is a measure of efficiency of an estimate:

the smaller $\mathrm{MSE}(\hat{\theta})$, the better the estimate $\hat{\theta}$

Cramer-Rao Inequality

Let $X_1, \ldots, X_n$ be i.i.d. from $\pi(x|\theta)$. Let $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ be any unbiased estimate of a parameter θ whose true values is θ_0. Then, under smoothness assumptions on $\pi(x|\theta)$,

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{V}[\hat{\theta}] \geq \frac{1}{nI(\theta_0)}$$

Cramer-Rao Inequality

Cramer-Rao: $$\boxed{\mathrm{MSE}(\hat{\theta}) = \mathbb{V}[\hat{\theta}] \geq \frac{1}{nI(\theta_0)}}$$

Important Remarks:

- $\hat{\theta}$ can't have arbitrary small MSE
- The Cramer-Rao inequality gives a lower bound on the variance of any unbiased estimate.

Definition

An unbiased estimate whose variance achieves this lower bound is said to be **efficient**.

Recall that MLE is asymptotically Normal: $\hat{\theta}_{\mathrm{MLE}} \to \mathcal{N}\left(\theta_0, \frac{1}{nI(\theta_0)}\right)$

- Therefore, MLE is asymptotically efficient
- However, for a finite sample size n, MLE may not be efficient
- MLEs are not the only asymptotically efficient estimates.

Example: Poisson Distribution

Recall that the Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events k occurring in a fixed interval of time if these events occur with a known average rate λ and independently of the time since the last event.

$$\mathbb{P}(X = k|\lambda) = \frac{\lambda^k}{k!}e^{-\lambda} \qquad \mathbb{E}[X] = \lambda \qquad \mathbb{V}[X] = \lambda$$

Example

Let $X_1, \ldots, X_n \sim \mathrm{Pois}(\lambda)$.

- Find the MLE of λ
- Show that $\hat{\lambda}_{\mathrm{MLE}}$ is efficient.

- The theorem does not exclude the possibility that there is a biased estimator of λ that has a smaller MSE than $\hat{\lambda}_{\mathrm{MLE}}$

Summary

- Mean squared error is a measure of efficiency of an estimate

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta_0)^2]$$

- If $\hat{\theta}$ is unbiased, then

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{V}[\hat{\theta}]$$

- Cramer-Rao Inequality:

$$\boxed{\mathrm{MSE}(\hat{\theta}) = \mathbb{V}[\hat{\theta}] \geq \frac{1}{nI(\theta_0)}}$$

- An unbiased estimate whose variance achieves this lower bound is said to be efficient
- Any MLE is asymptotically efficient (as $n \to \infty$)
- Example: if $X_1, \ldots, X_n \sim \mathrm{Poisson}(\lambda)$, then $\hat{\lambda}_{\mathrm{MLE}}$ is efficient