MATH 503: Mathematical Statistics

Lecture 3: Estimator Variation, and Properties of Point Estimators I

Reading: Sections 6.1-6.2, 7.3

Kimberly F. Sellers

Department of Mathematics & Statistics

Today's Topics

- Information
- Cramér-Rao Lower Bound
- Efficiency
- Sufficiency and its properties

Cramér-Rao Lower Bound

- Gives lower bound on variance of unbiased estimator
- Let X be r.v. with pdf $f(x;\theta)$, $\theta \in \Omega$
- Regularity conditions:
  - Pdfs are distinct, i.e. $\theta \neq \theta' \Rightarrow f(x_i;\theta) \neq f(x_i;\theta')$
  - Pdfs have common support for all θ
  - The point θ_0 is an interior point in Ω
  - Pdf is twice differentiable as a function of θ
  - Integral $\int f(x;\theta)\mathrm{d}x$ can be differentiated twice under the integral sign as a function of θ

Fisher Information: Background

Consider random variable X with pdf $f(x;\theta)$. By definition, $\int_{-\infty}^{\infty} f(x;\theta)\mathrm{d}x = 1$.

Taking derivative wrt $\theta \Rightarrow \int_{-\infty}^{\infty} \frac{\partial f(x;\theta)}{\partial\theta}\mathrm{d}x = 0$.

Multiply w/in integral $\Rightarrow \underbrace{\int_{-\infty}^{\infty} \frac{\partial f(x;\theta)/\partial\theta}{f(x;\theta)} f(x;\theta)\mathrm{d}x}_{*} = 0$.

$$* = \int_{-\infty}^{\infty} \frac{\partial \log f(x;\theta)}{\partial\theta} f(x;\theta)\mathrm{d}x = E\left(\underbrace{\frac{\partial \log f(x;\theta)}{\partial\theta}}_{\text{score function}}\right) = 0.$$

Fisher Information: Bkgd (cont.)

Differentiating again wrt θ (product rule),

$$\therefore 0 = \underbrace{\int_{-\infty}^{\infty} \frac{\partial^2 \log f(x;\theta)}{\partial \theta^2} f(x;\theta)\mathrm{d}x}_{E\left(\frac{\partial^2 \log f(x;\theta)}{\partial \theta^2}\right)}$$

$$+ \underbrace{\int_{-\infty}^{\infty} \frac{\partial \log f(x;\theta)}{\partial \theta} \frac{\partial \log f(x;\theta)}{\partial \theta} f(x;\theta)\mathrm{d}x}_{E\left[\left(\frac{\partial \log f(x;\theta)}{\partial \theta}\right)^2\right] = I(\theta)}$$

Fisher Information

$$I(\theta) = \mathrm{Var}\left(\frac{\partial \log f(X;\theta)}{\partial\theta}\right)$$

$$= E\left[\left(\frac{\partial \log f(X;\theta)}{\partial\theta} - \underbrace{E\left(\frac{\partial \log f(X;\theta)}{\partial\theta}\right)}_{0}\right)^2\right]$$

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

- The amount of information that observable r.v. *X* carries about unobservable θ
  - The greater these derivatives (on average), the more information we get about θ

Example 1

Let X~Bernoulli(θ). Find the associated information.

Fisher Information: Bkgd (cont.)

- Can consider information for random sample $X_1, \ldots, X_n$
- $\frac{\partial \log L(\theta;\mathbf{X})}{\partial\theta} = \sum_{i=1}^{n} \frac{\partial \log f(X_i;\theta)}{\partial\theta}$ where $\mathrm{Var}\left(\frac{\partial \log f(X_i;\theta)}{\partial\theta}\right) = I(\theta)$ for all i, and X_is are iid
$\Rightarrow \mathrm{Var}\left(\frac{\partial \log L(\theta;\mathbf{X})}{\partial\theta}\right) = nI(\theta)$
- **Result:** information for a sample is n times information for one rv

Cramér-Rao Lower Bound

- Let $X_1, X_2, \ldots, X_n$ be iid with common pdf $f(x;\theta)$ for $\theta \in \Omega$. Assume that the regularity conditions hold. Let $Y = u(X_1, X_2, \ldots, X_n)$ be a statistic with mean $E(Y) = k(\theta)$. Then $\mathrm{Var}(Y) \geq \frac{[k'(\theta)]^2}{nI(\theta)}$.

- Corollary: if Y is unbiased for θ, then $\mathrm{k}(\theta) = \theta$, and the result becomes $\mathrm{Var}(Y) \geq \frac{1}{nI(\theta)}$

Efficiency

- Let Y be an unbiased estimator of a parameter θ in the case of point estimation. The statistic Y is an <u>efficient estimator</u> of θ iff. the variance of Y attains the Cramer-Rao lower bound.
- The ratio of the Cramer-Rao lower bound to the actual variance of any unbiased estimator of θ is called the <u>efficiency</u> of that estimator.

Example 1 (cont.)

Let $X_1, \ldots, X_n$ iid ~ Bernoulli(θ) distribution. Show that the MLE of θ attains the Cramer-Rao lower bound.

Example 2

Let $X_1, \ldots, X_n$ iid ~Poisson(θ), $\theta > 0$.

1. Find the MLE of θ.
2. Show that the MLE is an efficient estimator of θ.

Theorem

- **Additional regularity condition:** The pdf $f(x;\theta)$ is three times differentiable as a function of θ. Further, for all $\theta \in \Omega$, there exists a constant c and function $M(x)$ s.t. $\left|\frac{\partial^3}{\partial\theta^3}\log f(x;\theta)\right| \leq M(x)$, with $E_{\theta_0}[M(x)] < \infty$, for all $\theta_0 - c < \theta < \theta_0 + c$ and all x in the support of X.

Theorem (cont.)

- Assume $X_1, \ldots, X_n$ are iid with pdf $f(x;\theta_0)$ for $\theta_0 \in \Omega$ such that all regularity conditions hold. Suppose further that Fisher information satisfies $0 < I(\theta_0) < \infty$. Then any consistent sequence of solutions of the MLE equations satisfies $\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, \frac{1}{I(\theta_0)}\right)$
- **Corollary:** Under the above assumptions, suppose $g(x)$ is a continuous function of x which is differentiable at θ_0 s.t. $g'(\theta_0) \neq 0$. Then $\sqrt{n}\left(g(\hat{\theta}) - g(\theta_0)\right) \xrightarrow{d} N\left(0, \frac{g'(\theta_0)^2}{I(\theta_0)}\right)$

Asymptotic Efficiency

- Let $X_1, \ldots, X_n$ be iid with pdf $f(x;\theta)$. Suppose $\hat{\theta}_{1n} = \hat{\theta}_{1n}(X_1, \ldots, X_n)$ is an estimator of θ_0 st.

$$\sqrt{n}(\hat{\theta}_{1n} - \theta_0) \xrightarrow{d} N(0, \sigma^2_{\hat{\theta}_{1n}})$$

Then

a) The asymptotic efficiency of $\hat{\theta}_{1n}$ is defined to be

$$e(\hat{\theta}_{1n}) = \frac{1/I(\theta_0)}{\sigma^2_{\hat{\theta}_{1n}}}$$

b) The estimator is asymptotically efficient if the ratio equals 1.

Asymptotic Efficiency (cont.)

- Suppose $\hat{\theta}_{1n} = \hat{\theta}_{1n}(X_1, \ldots, X_n)$ is an estimator of θ_0 such that

$$\sqrt{n}(\hat{\theta}_{1n} - \theta_0) \xrightarrow{d} N(0, \sigma^2_{\hat{\theta}_{1n}})$$

Then

c) Let $\hat{\theta}_{2n}$ be another estimator st.

$$\sqrt{n}(\hat{\theta}_{2n} - \theta_0) \xrightarrow{d} N(0, \sigma^2_{\hat{\theta}_{2n}})$$

Then the <u>asymptotic relative efficiency</u> (ARE) of $\hat{\theta}_{1n}$ to $\hat{\theta}_{2n}$ is

$$e(\hat{\theta}_{1n}, \hat{\theta}_{2n}) = \frac{\sigma^2_{\hat{\theta}_{2n}}}{\sigma^2_{\hat{\theta}_{1n}}}$$

Recap

- Let $Y_n = u(X_1, \ldots, X_n)$ be point estimator based on $X_1, \ldots, X_n$
- Y_n is <u>consistent</u> if $Y_n \xrightarrow{p} \theta$ (i.e. Y_n is close to θ for large sample sizes)
- Y_n is <u>unbiased</u> if $E(Y_n) = \theta$

Maximum Likelihood Estimators

- Under suitable conditions, MLEs are consistent
- MLEs not necessarily unbiased, but generally asymptotically unbiased

Minimum Variance Unbiased Estimators (MVUEs)

- For a given positive integer n, $Y = u(X_1, \ldots, X_n)$ is a <u>minimum variance unbiased estimator</u> (MVUE) of the parameter θ
  - if Y is unbiased, and
  - if the variance of Y is less than or equal to the variance of every other unbiased estimator of θ.

Sufficiency

Let $X_1, \ldots, X_n$ denote a random sample of size n from a distribution that has pdf/pmf $f(x;\theta), \theta \in \Omega$. Let $Y_1 = u_1(X_1, \ldots, X_n)$ be a statistic whose pdf/pmf is $f_{Y_1}(y_1;\theta)$. Then Y_1 is a sufficient statistic for θ iff.

$$\frac{f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta)}{f_{Y_1}[u_1(x_1, \ldots, x_n);\theta]} = H(x_1, \ldots, x_n),$$

where $H(x_1, \ldots, x_n)$ does not depend on $\theta \in \Omega$.

Example 1

Let $X_1, \ldots, X_n \sim$ Bernoulli(p) iid. Show that $Y_1(\boldsymbol{X}) = \sum X_i$ is sufficient.

Example 2

Let $X_1, \dots, X_n$ be iid with pdf $f(x;\theta) = e^{-(x-\theta)} I_{(\theta,\infty)}(x)$. Show that $Y_1(\boldsymbol{X}) = X_{(1)}$ is sufficient.

Neyman-Fisher Factorization Thm

Let $X_1, \ldots, X_n$ denote a random sample from a distribution that has pdf/pmf $f(x;\theta), \theta \in \Omega$. The statistic $Y_1 = u_1(X_1, \ldots, X_n)$ is a sufficient statistic for θ iff. we can find two nonnegative functions, k_1 and k_2, such that

$$f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta) = k_1[u_1(x_1, \ldots, x_n);\theta] \cdot k_2(x_1, \ldots, x_n)$$

where $k_2(x_1, \ldots, x_n)$ does not depend on θ.

Example 3

Let $X_1, \ldots, X_n \sim$ Normal(θ, σ^2) iid, where $\sigma^2 > 0$ known. Show that $\bar{X}$ is sufficient for θ.

Example 4

Let $X_1, \ldots, X_n$ be iid with pdf $f(x;\theta) = \theta x^{\theta-1}$, $0 < x < 1$, $\theta > 0$. Show that $\prod X_i$ is sufficient for θ.

Example 5

Let $X_1, \ldots, X_n \sim$Uniform$(0, \theta)$ iid. Find a sufficient estimator for θ.

Notes re. Sufficient Statistics

- Not unique!
- Can lead to a “best” point estimator
- Given two statistics, Y_1 sufficient for θ and Y_2 unbiased, function $\varphi(Y_1)$ is an unbiased estimator of θ having a smaller variance than that of Y_2

Rao-Blackwell Theorem

Let $X_1, \ldots, X_n$, n a fixed positive integer, denote a random sample from a distribution that has pdf/pmf $f(x;\theta), \theta \in \Omega$. Let $Y_1 = u_1(X_1, \ldots, X_n)$ be a sufficient statistic for θ, and let $Y_2 = u_2(X_1, \ldots, X_n)$, not a function of Y_1 alone, be an unbiased estimator of θ. Then $E(Y_2 \mid y_1) = \varphi(y_1)$ defines a statistic $\varphi(Y_1)$. This statistic $\varphi(y_1)$ is a function of the sufficient statistic for θ; it is an unbiased estimator of θ; and its variance is less than that of Y_2.

Notes re. Rao-Blackwell Thm.

- If we know a sufficient statistic for the parameter exists, the MVUE will be a function of the sufficient statistic.
- This does not mean that we first need to find an unbiased statistic!
- Focus on functions of sufficient statistics

Theorem

- Let $X_1, \ldots, X_n$ denote a random sample from a distribution that has pdf/pmf $f(x;\theta), \theta \in \Omega$. If a sufficient statistic $Y_1 = u_1(X_1, \ldots, X_n)$ for θ exists, and if a MLE $\hat{\theta}$ of θ also exists uniquely, then $\hat{\theta}$ is a function of $Y_1 = u_1(X_1, \ldots, X_n)$.

- **The point:** MLEs are functions of sufficient statistics.

Proof

Let $f_{Y_1}(y_1;\theta)$ be the pdf/pmf of Y_1. Then by the definition of sufficiency, the likelihood function

$$\begin{aligned} L(\theta; x_1, \ldots, x_n) &= f(x_1;\theta) f(x_2;\theta) \cdots f(x_n;\theta) \\ &= f_{Y_1}[u_1(x_1, \ldots, x_n);\theta] H(x_1, \ldots, x_n), \end{aligned}$$

where $H(x_1, \ldots, x_n)$ does not depend on θ. Thus L and f_{Y_1} are maximized simultaneously. Since there is only one value of θ that maximizes these functions, that value of θ must be a function of $u_1(x_1, \ldots, x_n)$. Thus, the MLE $\hat{\theta}$ is a function of sufficient statistic $Y_1 = u_1(x_1, \ldots, x_n)$.

Example

Let $X_1, \dots, X_n$ denote a random sample from a distribution that has pdf $f(x;\theta) = \theta e^{-\theta x}, 0 < x < \infty$.

1. Find a sufficient statistic for θ.
2. Find the MLE of θ.
3. Determine a MVUE of θ.