Lecture 9: UMVUEs and the Cramér-Rao Lower Bound

MATH 667-01
Statistical Inference
University of Louisville

October 5, 2017

- We discuss uniform minimum variance unbiased estimators as discussed in Section 7.3 of Casella and Berger (2002)[1].
- We review correlation from Section 4.5.
- We discuss and prove the Cramér-Rao Inequality and some corollaries. The regularity conditions in these notes are from Section 7.3 of Casella and Berger (1990)[2].
- We present several examples to illustrate the results.

[1] Casella, G. and Berger, R. (2002). *Statistical Inference, Second Edition*. Duxbury Press, Belmont, CA.

[2] Casella, G. and Berger, R. (1990). *Statistical Inference*. Duxbury Press, Belmont, CA.

- In this lecture, we evaluate an estimator W of a parameter θ based on the squared error loss function.
- If we consider only unbiased estimators, then $\mathsf{E}_\theta[(W-\theta)^2] = \mathsf{Var}_\theta[W]$.
- *Definition L9.1* (Def 7.3.7 on p.334): An estimator W^* is a *best unbiased estimator* of $\tau(\theta)$ if it satisfies $\mathsf{E}_\theta[W^*] = \tau(\theta)$ for all θ and, for any other unbiased estimator W with $\mathsf{E}_\theta[W] = \tau(\theta)$, we have $\mathsf{Var}_\theta[W^*] \leq \mathsf{Var}_\theta[W]$ for all θ.
- W^* is also called a *uniform minimum variance unbiased estimator* (UMVUE) of $\tau(\theta)$.

- *Example L9.1*:
- *Answer to Example L9.1*:

Review: Correlation

- $\mathsf{E}[X] = \mu_X$, $\mathsf{E}[Y] = \mu_Y$, $\mathsf{Var}[X] = \sigma_X^2$, $\mathsf{Var}[Y] = \sigma_Y^2$
- Assume $0 < \sigma_X^2 < \infty$ and $0 < \sigma_Y^2 < \infty$
- *Definition L9.2* (Def 4.5.2 on p.169): The *correlation of X and Y* is the number defined by

$$\rho_{XY} = \frac{\mathsf{Cov}[X, Y]}{\sigma_X \sigma_Y}.$$

  The value ρ_{XY} is also called the *correlation coefficient*.
- *Theorem L9.1* (Thm 4.5.7 on p.172): For any random variables X and Y,
  (a) $-1 \leq \rho_{XY} \leq 1$.
  (b) $|\rho_{XY}| = 1$ if and only if there exists numbers $a \neq 0$ and b such that $P(Y = aX + b) = 1$. If $\rho_{XY} = 1$ then $a > 0$, and if $\rho_{XY} = -1$ then $a < 0$.

Cramér-Rao Lower Bound

- *Theorem L9.2* (p.335): Let $X_1, \ldots, X_n$ be a sample with pdf $f(\boldsymbol{x}|\theta)$, and let $W(\boldsymbol{X}) = W(X_1, \ldots, X_n)$ be any estimator where $\mathsf{E}_\theta[W(\boldsymbol{X})]$ is a differentiable function of θ. Suppose the joint pdf $f(\boldsymbol{x}|\theta)$ satisfies

$$\frac{d}{d\theta} \int \cdots \int h(\boldsymbol{x}) f(\boldsymbol{x}|\theta)\, d\boldsymbol{x} = \int \cdots \int h(\boldsymbol{x}) \frac{\partial}{\partial\theta} f(\boldsymbol{x}|\theta)\, d\boldsymbol{x},$$

for any function $h(\boldsymbol{x})$ with $\mathsf{E}_\theta[\,|h(\boldsymbol{X})|\,] < \infty$. Then

$$\mathsf{Var}_\theta[W(\boldsymbol{X})] \geq \frac{\left\{\frac{d}{d\theta}\mathsf{E}_\theta[W(\boldsymbol{X})]\right\}^2}{\mathsf{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(\boldsymbol{X}|\theta)\right)^2\right]}.$$

- The inequality is referred to as the Cramér-Rao inequality.
- If $W(\boldsymbol{X})$ is an unbiased estimator of $\tau(\theta)$, then the numerator becomes

$$\left(\frac{d}{d\theta}\mathsf{E}_\theta[W(\boldsymbol{X})]\right)^2 = \left(\tau'(\theta)\right)^2.$$

- *Proof of Theorem L9.2*: Since *Theorem L9.1*(a) implies

$$\left\{\mathsf{Cov}\left[W(\boldsymbol{X}), \frac{\partial}{\partial\theta}\ln f(\boldsymbol{X}|\theta)\right]\right\}^2 \leq \mathsf{Var}[W(\boldsymbol{X})]\mathsf{Var}\left[\frac{\partial}{\partial\theta}\ln f(\boldsymbol{X}|\theta)\right],$$

  it follows that

$$\mathsf{Var}[W(\boldsymbol{X})] \geq \frac{\left\{\mathsf{Cov}\left[W(\boldsymbol{X}), \frac{\partial}{\partial\theta}\ln f(\boldsymbol{X}|\theta)\right]\right\}^2}{\mathsf{Var}\left[\frac{\partial}{\partial\theta}\ln f(\boldsymbol{X}|\theta)\right]}.$$

- *Proof of Theorem L9.2 continued*: Note that

$$
\begin{aligned}
\mathsf{E}\left[\frac{\partial}{\partial\theta}\ln f(\boldsymbol{X}|\theta)\right] &= \int_{\mathcal{X}} \frac{\partial}{\partial\theta}\ln f(\boldsymbol{x}|\theta) f(\boldsymbol{x}|\theta)\ d\boldsymbol{x} \\
&= \int_{\mathcal{X}} \frac{\frac{\partial}{\partial\theta} f(\boldsymbol{X}|\theta)}{f(\boldsymbol{X}|\theta)} f(\boldsymbol{x}|\theta)\ d\boldsymbol{x} \\
&= \int_{\mathcal{X}} \frac{\partial}{\partial\theta} f(\boldsymbol{x}|\theta)\ d\boldsymbol{x} \\
&= \frac{\partial}{\partial\theta}\int_{\mathcal{X}} f(\boldsymbol{x}|\theta)\ d\boldsymbol{x} = \frac{\partial}{\partial\theta} 1 = 0.
\end{aligned}
$$

- *Proof of Theorem L9.2 continued*: Then we have

$$\begin{aligned}\mathsf{Cov}\left[W(\boldsymbol{X}), \frac{\partial}{\partial\theta}\ln f(\boldsymbol{X}|\theta)\right] &= \mathsf{E}\left[W(\boldsymbol{X})\frac{\partial}{\partial\theta}\ln f(\boldsymbol{X}|\theta)\right] \\ &= \mathsf{E}\left[W(\boldsymbol{X})\frac{\frac{\partial}{\partial\theta}f(\boldsymbol{X}|\theta)}{f(\boldsymbol{X}|\theta)}\right] \\ &= \int_{\mathcal{X}}\frac{d}{d\theta}[W(\boldsymbol{x})f(\boldsymbol{x}|\theta)]\ d\boldsymbol{x} \\ &= \frac{d}{d\theta}\mathsf{E}[W(\boldsymbol{X})]\end{aligned}$$

and

$$\mathsf{Var}\left[\frac{\partial}{\partial\theta}\ln f(\boldsymbol{X}|\theta)\right] = \mathsf{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(\boldsymbol{X}|\theta)\right)^2\right].$$

- *Theorem L9.3* (p.337): Let $X_1, \ldots, X_n$ be iid with pdf $f(x|\theta)$, and let $W(\boldsymbol{X}) = W(X_1, \ldots, X_n)$ be any estimator where $\mathsf{E}_\theta[W(\boldsymbol{X})]$ is a differentiable function of θ. If the joint pdf $f(\boldsymbol{x}|\theta) = \prod f(x_i|\theta)$ satisfies

$$\frac{d}{d\theta} \int \cdots \int h(\boldsymbol{x}) f(\boldsymbol{x}|\theta) \; d\boldsymbol{x} = \int \cdots \int h(\boldsymbol{x}) \frac{\partial}{\partial\theta} f(\boldsymbol{x}|\theta) \; d\boldsymbol{x},$$

for any function $h(\boldsymbol{x})$ with $\mathsf{E}_\theta[\ |h(\boldsymbol{X})|\ ] < \infty$, then

$$\mathsf{Var}_\theta[W(\boldsymbol{X})] \geq \frac{\left(\frac{d}{d\theta}\mathsf{E}_\theta[W(\boldsymbol{X})]\right)^2}{n\mathsf{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(X|\theta)\right)^2\right]}.$$

- *Proof of Theorem L9.3 continued*: If we also assume that $X_1, \ldots, X_n$ is iid, then we have

$$
\begin{aligned}
\mathsf{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(\boldsymbol{X}|\theta)\right)^2\right] &= \mathsf{E}\left[\left(\frac{\partial}{\partial\theta}\ln\prod_{i=1}^n f(X_i|\theta)\right)^2\right] \\
&= \mathsf{E}\left[\left(\frac{\partial}{\partial\theta}\sum_{i=1}^n \ln f(X_i|\theta)\right)^2\right] \\
&= \sum_{i=1}^n\sum_{j=1}^n \mathsf{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(X_i|\theta)\right)\left(\frac{\partial}{\partial\theta}\ln f(X_j|\theta)\right)\right] \\
&= \sum_{i=1}^n \mathsf{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(X_i|\theta)\right)^2\right] + \\
&\quad \sum_{i=1}^n \sum_{\substack{j=1 \\ j\neq i}}^n \mathsf{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(X_i|\theta)\right)\left(\frac{\partial}{\partial\theta}\ln f(X_j|\theta)\right)\right]
\end{aligned}
$$

Cramér-Rao Lower Bound (iid case)

- *Proof of Theorem L9.3 continued*:

$$
\begin{aligned}
&= \sum_{i=1}^{n} \mathsf{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(X_i|\theta)\right)^2\right] + \\
&\quad \sum_{i=1}^{n}\sum_{j\neq i} \mathsf{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(X_i|\theta)\right)\left(\frac{\partial}{\partial\theta}\ln f(X_j|\theta)\right)\right] \\
&= \sum_{i=1}^{n} \mathsf{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(X_i|\theta)\right)^2\right] + \\
&\quad \sum_{i=1}^{n}\sum_{j\neq i} \mathsf{E}\left[\frac{\partial}{\partial\theta}\ln f(X_i|\theta)\right]\mathsf{E}\left[\frac{\partial}{\partial\theta}\ln f(X_j|\theta)\right] \\
&= \sum_{i=1}^{n} \mathsf{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(X_i|\theta)\right)^2\right] \\
&= n\mathsf{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(X|\theta)\right)^2\right].
\end{aligned}
$$

- The quantity $\mathsf{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(X|\theta)\right)^2\right]$ is called the *information number*, or *Fisher information* of the sample.
- *Theorem L9.4* (Lem 7.3.11 on p.338): If $f(x|\theta)$ satisfies

$$\frac{d}{d\theta}\mathsf{E}_\theta\left[\frac{\partial}{\partial\theta}\ln f(X|\theta)\right] = \int \frac{\partial}{\partial\theta}\left[\left(\frac{\partial}{\partial\theta}\ln f(x|\theta)\right) f(x|\theta)\right]\, dx,$$

  then

$$\mathsf{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(X|\theta)\right)^2\right] = -\mathsf{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f(X|\theta)\right].$$

- The condition on $f(x|\theta)$, and consequently the result, is true for an exponential family.

- *Proof of Theorem L9.4*:

Cramér-Rao Lower Bound (iid case)

- *Example L9.3*: Let $X_1, \ldots, X_n$ be iid Poisson(λ). Find the Cramér-Rao lower bound on the variance of unbiased estimators of λ. Also, find the MLE and show that it is the UMVUE of λ.
- *Answer to Example L9.3*: The Cramér-Rao lower bound is $\dfrac{\lambda}{n}$.

  The MLE of λ is $\hat{\lambda} = \bar{X}$ and Var$[\bar{X}] = \dfrac{\lambda}{n}$.

Cramér-Rao Lower Bound (iid case)

- *Example L9.4*:
- *Answer to Example L9.4*:

- *Theorem L9.5* (Cor 7.3.15 on p.341): Let $X_1, \ldots, X_n$ be iid $f(x|\theta)$, where $f(x|\theta)$ satisfies the conditions of *Theorem L9.3*. Let $L(\theta|\boldsymbol{x}) = \prod_{i=1}^{n} f(x_i|\theta)$ denote the likelihood function. If $W(\boldsymbol{X}) = W(X_1, \ldots, X_n)$ is any unbiased estimator of $\tau(\theta)$, then $W(\boldsymbol{X})$ attains the Cramér-Rao Lower Bound if and only if
$$a(\theta)[W(\boldsymbol{x}) - \tau(\theta)] = \frac{\partial}{\partial\theta} \ln L(\theta|\boldsymbol{x})$$
for some function $a(\theta)$.
- *Proof of Theorem L9.5*:

Attainment

- *Example L9.4*:
- *Answer to Example L9.4*:

Attainment

- *Example L9.5*:
- *Answer to Example L9.5*: