Minimum Variance Estimators (LM Ch. 5.5)

February 19, 2010

1 Cramér-Rao Lower Bound

Given two unbiased estimators for θ, $\hat{\theta}_1$ and $\hat{\theta}_2$, we know that the one with the smaller variance is the *better* estimator. However, how do we know that there isn't yet another estimator $\hat{\theta}_3$ that has an even smaller variance?

It turns out there is a theoretical limit below which the variance of any unbiased estimator for θ cannot fall. This is called the Cramér Rao lower bound (CRLB).

1.1 Theorem 5.5.1

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from the continuous pdf $f_Y(y;\theta)$, where $f_Y(y;\theta)$ has continuous first-order and second-order partial derivatives at all but a finite set of points. Suppose that the set of y's for which $f_Y(y;\theta) \neq 0$ does not depend θ. Let $\hat{\theta} = h(Y_1, Y_2, \ldots, Y_n)$ be any unbiased estimator for θ. Then

$$\text{Var}(\hat{\theta}) \geq \left\{ n\text{E}\left[\left(\frac{\partial \ln f_Y(Y;\theta)}{\partial \theta} \right)^2 \right] \right\}^{-1} = \left\{ -n\text{E}\left[\left(\frac{\partial^2 \ln f_Y(Y;\theta)}{\partial \theta^2} \right) \right] \right\}^{-1}$$

(A similar statement holds if the n observations come from a discrete pdf, $p_X(k;\theta)$).

If the variance of an unbiased estimator $\hat{\theta}$ is equal to the CRLB, we know that the estimator is *optimal* in the sense that there is no unbiased $\hat{\theta}$ that can estimate θ with greater precision.

1.2 Example

Let $X_1, X_2, \ldots, X_n$ be n Bernoulli random variables, and $X = X_1 + X_2 + \cdots + X_n$. Then, $X \sim Bin(n, p)$. Show that $\hat{p} = \bar{X}$ achieves the Cramér-Rao lower bound.

Clearly, $\hat{p}$ is unbiased and,

$$p_{X_i}(k;p) = p^k (1-p)^{1-k} \quad k = 0, 1; \quad 0 < p < 1$$

$$\ln p_{X_i}(X_i;p) = X\ln p + (1-X_i)\ln(1-p)$$
$$\frac{\partial \ln p_{X_i}(X_i;p)}{\partial p} = \frac{X_i}{p} - \frac{1-X_i}{1-p}$$
$$\frac{\partial^2 \ln p_{X_i}(X_i;p)}{\partial p^2} = -\frac{X_i}{p^2} - \frac{1-X_i}{(1-p)^2}$$

Taking the expected value of the second derivative,

$$\begin{aligned}
\mathrm{E}\left[\frac{\partial^2 \ln p_{X_i}(X_i;p)}{\partial p^2}\right] &= -\frac{\mathrm{E}[X_i]}{p^2} - \frac{1-\mathrm{E}[X_i]}{(1-p)^2} \\
&= -\frac{p}{p^2} - \frac{1-p}{(1-p)^2} \\
&= -\frac{1}{p} - \frac{1}{1-p} \\
&= -\frac{1-p+p}{p(1-p)} = -\frac{1}{p(1-p)}
\end{aligned}$$

The CRLB is then,

$$\left\{-n\mathrm{E}\left[\frac{\partial^2 \ln p_X(X;p)}{\partial p^2}\right]\right\}^{-1} = \left\{-n\left(-\frac{1}{p(1-p)}\right)\right\}^{-1} = \frac{p(1-p)}{n}$$

The variance of $\hat{p} = \bar{X} = \frac{p(1-p)}{n}$, hence $\hat{p}$ achieves the CRLB.

What would be different if we had started with the Binomial distribution, $(p_X(k;p) = \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k}, k = 1, 2, \ldots)$ instead of the Bernoulli?

2 Minimum Variance Unbiased Estimators

Unbiased estimators whose variance are equal to the CRLB are called Minimum Variance Unbiased Estimators (MVUE).

2.1 Definition: Best Estimators

Let Θ denote the set of all estimators $\hat{\theta} = h(Y_1, Y_2, \ldots, Y_n)$ that are unbiased for the parameter θ in the continuous pdf $f_Y(y;\theta)$. We say that $\hat{\theta}^*$ is a *best* (or *minimum-variance*) estimator if $\hat{\theta}^* \in \Theta$ and

$$\mathrm{Var}(\hat{\theta}^*) \leq \mathrm{Var}(\hat{\theta}) \quad \text{for all} \quad \hat{\theta} \in \Theta$$

(Similar terminology applies to discrete distributions).

2.2 Definition: Efficient Estimators

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n drawn from the continuous pdf $f_Y(y; \theta)$. Let $\hat{\theta} = h(Y_1, Y_2, \ldots, Y_n)$ be an unbiased estimator for θ.

(a) The unbiased estimator $\hat{\theta}$ is said to be *efficient* if the variance of $\hat{\theta}$ equals the Cramér-Rao lower bound associated with $f_Y(y; \theta)$.

(b) The *efficiency* of an unbiased estimator $\hat{\theta}$ is the ratio of the Cramér-Rao lower bound for $f_Y(y; \theta)$ to the variance of $\hat{\theta}$.

Note that *best* and *efficient* are not synonymous. There are situations where no unbiased estimators achieve the Cramér-Rao lower bound. None of these will be *efficient* but one (or more) could still be termed *best*.

2.3 Example

Refer to the previous example. Suppose that $n = 10$. Derive the efficiency of the estimator $\tilde{p} = (X_1 + X_2)/2$.

First note that $\tilde{p}$ is unbiased as $\mathrm{E}[(X_1 + X_2)/2] = (p + p)/2 = p$.

$$\mathrm{Var}[(X_1 + X_2)/2] = 2p(1-p)/4 = p(1-p)/2$$

$$\text{Efficiency} = \frac{p(1-p)}{n} \times \frac{2}{p(1-p)} = \frac{2}{n} = \frac{1}{5}$$

3 MLEs and the Cramér-Rao Lower Bound

The maximum likelihood estimate is probably the most used estimation technique. As we have seen, maximum likelihoods estimates will always be functions of the sufficient statistics. According to the Rao-Blackwell Theorem (Lecture Notes 16.3 or LM page 405), unbiased estimators that are functions of the sufficient statistics will have less variance than unbiased estimators that are *not* functions of the sufficient statistics. In addition, MLEs are advantageous due to the following asymptotic properties:

1. **The maximum likelihood estimate is at least asymptotically unbiased.** It may be unbiased for any number of observations (for example if the MLE is the sample mean).

2. **The maximum likelihood estimate is consistent.** For larger and larger samples, its variance tends to 0 and its expectation tends to the true value of the parameter θ.

3. **The maximum likelihood estimate is asymptotically efficient.** As $n \to \infty$, the ratio of the variance of a MLE to the Cramér-Rao lower bound tends to 1. As a MLE is asymptotically unbiased, it is then also asymptotically efficient.

4. **The maximum likelihood estimate is asymptotically normally distributed.** As $n \to \infty$, the distribution of the MLE converges to a normal distribution. Even for moderately large samples, the distribution of MLE is approximately normal.

4 Practice Problems

1. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from a gamma pdf $f_Y(y;\theta) = \frac{1}{(r-1)!\theta^r} y^{r-1} e^{-y/\theta}, y > 0$. Show that $\hat{\theta}$, the MLE of θ, is an efficient estimator.

   First lets find the MLE.

$$\begin{aligned}
L(\theta) &= \prod \frac{1}{(r-1)!\theta^r} y_i^{r-1} e^{-y_i/\theta} \\
&= \left(\frac{1}{(r-1)!}\right)^n \frac{1}{\theta^{rn}} e^{-\frac{1}{\theta}\sum y_i} \prod y_i^{r-1} \\
\ln L(\theta) &= n \ln \frac{1}{(r-1)!} - rn \ln \theta - \frac{1}{\theta} \sum y_i + (r-1) \sum y_i \\
\frac{d \ln L(\theta)}{d\theta} &= -\frac{rn}{\theta} + \frac{1}{\theta^2} \sum y_i = 0 \\
\hat{\theta} &= \frac{\sum y_i}{rn} = \frac{\bar{Y}}{r}
\end{aligned}$$

   Next, we need to show it is unbiased, and calculate its variance. Since it is a gamma distribution, $E[Y] = r\theta$ and $\text{Var}[Y] = r\theta^2$.

$$\begin{aligned}
\text{E}[\hat{\theta}] &= \text{E}\left[\frac{\bar{Y}}{r}\right] \\
&= \frac{r\theta}{r} = \theta \\
\text{Var}[\hat{\theta}] &= \text{Var}\left[\frac{\bar{Y}}{r}\right] \\
&= \frac{r\theta^2}{nr^2} = \frac{\theta^2}{nr}
\end{aligned}$$

Now we need to calculate the CRLB.

$$
\begin{aligned}
\frac{d \ln f}{d\theta} &= -\frac{r}{\theta} + \frac{Y}{\theta^2} \\
\frac{d^2 \ln f}{d\theta^2} &= \frac{r}{\theta^2} - \frac{2Y}{\theta^3} \\
\mathrm{E}\left[\frac{d^2 \ln f}{d\theta^2}\right] &= \frac{r}{\theta^2} - \frac{2r\theta}{\theta^3} \\
&= \frac{r - 2r}{\theta^2} = -\frac{r}{\theta^2} \\
\text{CRLB} &= \frac{\theta^2}{nr}
\end{aligned}
$$

Thus the MLE achieves the CRLB and is an efficient estimator.

2. Let $Y_1, Y_2, \ldots, Y_n$ be independent exponentially distributed random variables with mean θ (or rate $1/\theta$).

   (a) Find the probability distribution function for $Y_{min} = min(Y_1, Y_2, \ldots, Y_n)$. (*Hint: Think of how we found the probability distribution for the max from a Uniform distribution and apply the same idea here.*)

   $$
   \begin{aligned}
   P(Y_{min} \leq y) &= 1 - P(Y_{min} \geq y) = 1 - P(Y_1 \geq y \cap Y_2 \geq y \ldots Y_n \geq y) \\
   &= 1 - P(Y_1 \geq y) \times P(Y_2 \geq y) \times \cdots \times P(Y_n \geq y) \quad \text{since independent} \\
   &= 1 - e^{-\frac{ny}{\theta}} \\
   f_{Y_{min}}(y) &= \frac{n}{\theta} e^{-\frac{ny}{\theta}}
   \end{aligned}
   $$

   Thus $Y_{min} \sim Exp(n/\theta)$.

   (b) Show that nY_{min} is an unbiased estimator of θ.

   $\mathrm{E}[Y_{min}] = \frac{\theta}{n}$, therefore $\mathrm{E}[nY_{min} = \frac{n\theta}{n} = \theta$ which is unbiased.

   (c) Find the variance of this estimator and calculate its efficiency relative to the Cramér-Rao lower bound.

   $$\mathrm{Var}[nY_{min}] = n^2 \mathrm{Var}[Y_{min}] = n^2 \left(\frac{\theta}{n}\right)^2 = \theta^2$$

$$
\begin{aligned}
f &= \frac{1}{\theta}e^{-\frac{Y}{\theta}} \\
\ln f &= -\ln\theta - \frac{Y}{\theta} \\
\frac{d\ln f}{d\theta} &= -\frac{1}{\theta} + \frac{Y}{\theta^2} \\
\frac{d^2\ln f}{d\theta^2} &= \frac{1}{\theta^2} - \frac{2Y}{\theta^3} \\
\mathrm{E}\left[\frac{d^2\ln f}{d\theta^2}\right] &= \frac{1}{\theta^2} - \frac{2\theta}{\theta^3} \\
&= \frac{1-2}{\theta^2} = -\frac{1}{\theta^2} \\
\text{CRLB} &= \frac{\theta^2}{n}
\end{aligned}
$$

$$
\text{Efficiency} = \frac{\theta^2}{n} \times \frac{1}{\theta^2} = \frac{1}{n}
$$

3. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from the uniform pdf $f_Y(y;\theta) = 1/\theta, 0 \le y \le \theta$.

   (a) Find the maximum likelihood estimator of θ. From last homework we know that $\hat{\theta} = Y_{max}$, since $L(\theta) = (1/\theta)^n$ is decreasing in θ, and the smallest value of θ that satisfies the constraint $0 \le y \le \theta$ is $\hat{\theta} = Y_{max}$.

   (b) Find an unbiased estimator of θ based on the maximum likelihood estimator.

$$
\begin{aligned}
P(Y_{max} \le y) &= P(\text{Every} Y_i \le y) = \left(\frac{y}{\theta}\right)^n \\
f_{Y_{max}} &= \frac{ny^{n-1}}{\theta^n} \\
\mathrm{E}[Y_{max}] &= \int_0^\theta \frac{ny^n}{\theta^n} = \frac{n\theta}{n+1} \\
\mathrm{E}\left[\frac{n+1}{n}Y_{max}\right] &= \theta
\end{aligned}
$$

   Thus, $\frac{n+1}{n}Y_{max}$ is an unbiased etimator of θ.

   (c) Compare the variance of this estimator to the Cramér-Rao lower bound.

$$\begin{aligned}
f &= 1/\theta \\
\ln f &= -\ln\theta \\
\frac{d\ln f}{d\theta} &= -\frac{1}{\theta} \\
\left(\frac{d\ln f}{d\theta}\right)^2 &= \frac{1}{\theta^2} \\
\text{CRLB} &= \frac{\theta^2}{n}
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}[\mathrm{Y}_{max}^2] &= \int_0^\theta \frac{ny^{n+1}}{\theta^n} = \frac{n\theta^2}{n+2} \\
\mathrm{E}\left[\left(\frac{n+1}{n}Y_{max}\right)^2\right] &= \left(\frac{n+1}{n}\right)^2 \times \frac{n\theta^2}{n+2} = \frac{(n+1)^2\theta^2}{n(n+2)} \\
\mathrm{Var}\left[\frac{n+1}{n}Y_{max}\right] &= \frac{(n+1)^2\theta^2}{n(n+2)} - \theta^2 \\
&= \frac{\theta^2}{n(n+2)}[n^2+2n+1-(n^2+2n)] \\
&= \frac{\theta^2}{n(n+2)}
\end{aligned}$$

This is smaller than the CRLB!. This occurs because Theorem 5.5.1 is not necessarily valid if the range of the pdf depends on the parameter as is the case with this problem.