Probability and Statistics

Fall 2000

LECT. 14: MINIMUM VARIANCE UNBIASED ESTIMATORS, CB 7.3.2, 7.3.3

A different characterization of the Cramér-Rao bound is in terms of the second derivative of the log of the density function:

Result 1 (CRAMÈR–RAO BOUND)
Let X be a random variable with pdf/pmf $f_X(x;\theta)$, and let W be an unbiased estimator for θ. Then

$$V(W) \geq -\frac{1}{E\left[\frac{\partial^2 \ln f}{\partial \theta^2}(x;\theta)\right]}.$$

This result relies on the information matrix equality:

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

To see why this holds recall that

$$1 = \int_x f_X(x;\theta)dx,$$

implying,

$$0 = \frac{\partial}{\partial \theta}\int_x f_X(x;\theta)dx.$$

and thus, assuming we can change the order of differentiation and integration, we get

$$0 = \int_x \frac{\partial f_X}{\partial \theta}(x;\theta)dx$$

$$= \int_x \frac{\partial \ln f_X}{\partial \theta}(x;\theta) \cdot f_X(x;\theta)dx.$$

Now differentiate again to get

$$0 = \int_x \frac{\partial^2 \ln f_X}{\partial \theta^2}(x;\theta) \cdot f_X(x;\theta)dx + \int_x \frac{\partial \ln f_X}{\partial \theta}(x;\theta) \cdot \frac{\partial f_X}{\partial \theta}(x;\theta)dx$$

$$= \int_x \frac{\partial^2 \ln f_X}{\partial \theta^2}(x;\theta) \cdot f_X(x;\theta)dx + \int_x \frac{\partial \ln f_X}{\partial \theta}(x;\theta) \cdot \frac{\partial \ln f_X}{\partial \theta}(x;\theta) f_X(x;\theta)dx$$

$$= \int_x \frac{\partial^2 \ln f_X}{\partial \theta^2}(x;\theta) \cdot f_X(x;\theta)dx + \int_x \left(\frac{\partial \ln f_X}{\partial \theta}(x;\theta)\right)^2 f_X(x;\theta)dx$$

$$= E\Big[\frac{\partial^2 \ln f}{\partial \theta^2}(x;\theta)\Big] + E\Big[\frac{\partial \ln f}{\partial \theta}(x;\theta)^2\Big].$$

Now back to the interpretation of the Cramér-Rao bound. The Cramér–Rao bound gives a lower bound for the variance of unbiased estimators. In some sense this is helpful only if we can find an unbiased estimator with variance equal to this bound. If that is the case we know this is the minimum variance unbiased estimator. If not, there are two possibilities. Either we missed the minimum variance unbiased estimator, or we have an minimum variance unbiased estimator with variance larger than the bound. In many cases a minimum variance unbiased estimator does not even exist. To demonstrate some of these possibilities, consider the following examples. We have already seen an example where the bound does not apply.

Example 1

Suppose X has a binomial distribution with parameters 1 and $\sqrt{\theta}$. Any estimator for θ can be written as

$$W = W(X) = W(0) + (W(1) - W(0) \cdot X = \alpha + \beta \cdot X.$$

Its expectation is for any α and β equal to

$$\alpha + \beta \cdot \sqrt{\theta}.$$

There are no α and β that make this equal to θ, and so there is no unbiased estimator for θ, let alone one that achieves the Cramer–Rao bound. □

Example 2

X_1 and X_2 are independent binomial random variable with parameters 1 and $\sqrt{\theta}$:

$$f_{X_1,X_2}(x_1, x_2|\theta) = (\sqrt{\theta})^{x_1+x_2} \cdot (1-\sqrt{\theta})^{2-x_1-x_2}.$$

What is the Cramer–Rao bound? The log of the density is

$$\ln f_{X_1,X_2}(x_1, x_2|\theta) = \frac{1}{2}(x_1 + x_2) \cdot \ln\theta + (2 - x_1 - x_2) \cdot \ln(1-\sqrt{\theta}).$$

The derivative, or the score function, is

$$\frac{\partial \ln f_{X_1,X_2}}{\partial\theta}(x_1, x_2|\theta) = \frac{1}{2\theta}(x_1 + x_2) \cdot \ln\theta - \frac{1}{2}(2 - x_1 - x_2) \cdot \frac{\theta^{-1/2}}{1-\sqrt{\theta}}$$

$$= \frac{x_1 + x_2 - 2\sqrt{\theta}}{2\theta(1-\sqrt{\theta})}.$$

The score function clearly has expectation zero. Its variance is the the inverse of the CR bound:

$$1/CR = E\left[\left(\frac{x_1 + x_2 - 2\sqrt{\theta}}{2\theta(1-\sqrt{\theta})}\right)^2\right] = \frac{1}{2\theta\sqrt{\theta}(1-\sqrt{\theta})},$$

and the CR bound is

$$CR = 2\theta\sqrt{\theta}(1-\sqrt{\theta}).$$

Now consider estimators for θ. Any estimator can be written as

$$W = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + a_3 \cdot X_1 \cdot X_2,$$

with expectation

$$E[W] = a_0 + a_1 \cdot \sqrt{\theta} + a_2 \cdot \sqrt{\theta} + a_3 \cdot \theta.$$

Unbiased estimators must have $a_0 = 0$, $a_3 = 1$ and $a_2 = -a_1$. Using Rao–Blackwell to make the estimator a function of the sufficient statistic $X_1 + X_2$ implies that we must have

$$W = X_1 \cdot X_2.$$

This estimator has mean θ and variance $\theta(1-\theta)$, which is strictly higher than the Cramer–Rao bound. Nevertheless, it is the minimum variance unbiased estimator. □

Now let us investigate when we have an unbiased estimator with variance equal to the Cramer–Rao bound. In that case we must have, in the notation of the proof of the CR bound, that the correlation of the score U and the estimator W is equal to one in absolute value. Hence it must be the case that the score is a linear function of W, with coefficients possibly depending on θ:

$$\frac{\partial \ln f}{\partial \theta}(X;\theta) = a(\theta) \cdot W(X) + b(\theta).$$

Because W is unbiased, or $E[W] = \theta$, it must be that $b(\theta) = -a(\theta) \cdot \theta$, or we must be able to write the score function as

$$\frac{\partial \ln f}{\partial \theta}(X;\theta) = a(\theta) \cdot (W(X) - \theta).$$

It turns out that this is both sufficient and necessary for the existence of an unbiased estimator with variance equal to the Cramer–Rao bound.

Result 1

An unbiased estimator with variance equal to the Cramer–Rao bound exists if and only if the score function can be written as

$$\frac{\partial \ln f}{\partial \theta}(X;\theta) = a(\theta) \cdot (W(X) - \theta),$$

for some function $W(X)$. The minimum variance unbiased estimator is then equal to the maximum likelihood estimator $W(X) = \hat{\theta}_{mle}$. □

Proof

We have already proven that the existence of an MVUE with variance equal to the CR bound implies the above characterization of the score function. Now let us consider the only if part of the result.

Suppose we can write the score as

$$\frac{\partial \ln f}{\partial \theta}(X;\theta) = a(\theta) \cdot (W(X) - \theta).$$

Because the score function has expectation zero, $W(X)$ is an unbiased estimator. Its variance is equal to $a(\theta)^2$ times the variance of the score function, which itself is equal to the inverse of the CR bound. At the same time, by the information matrix equality the expected second derivative of the log of the density is also equal to minus the expectation of the square of the first derivative of the log of the density. The second derivative is equal to

$$E\left[\frac{\partial^2 \ln f}{\partial \theta^2}(X;\theta)\right] = E[a'(\theta) \cdot (W(X) - \theta) - a(\theta)] = -a(\theta).$$

Hence

$$1/CR = a(\theta) = a(\theta)^2 \cdot V(W(X)),$$

implying that

$$V(W(X)) = 1/a(\theta) = CR$$

Finally, by setting the derivative of the log of the density equal to zero, combined with a negative second derivative, we have maximized the log of the density, or the log likelihood and so under these conditions the minimum variance unbiased estimator $W(X)$ is equal to the maximum likelihood estimator.