

Lecture Note 14: Cramér-Rao Bound

Another important result that helps us in the search for a minimum variance unbiased estimator is the Cramér–Rao bound:

Result 1 (CRAMÉR–RAO BOUND)

Let X be a random variable with pdf/pmf $f_X(x; \theta)$, and that we can interchange integration and differentiation in the following sense:

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

Let W be an unbiased estimator for θ with finite variance. Then

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

Proof: Recall that the square of the covariance of two random variables S and U is less than or equal to the product of the variances (that is the same as saying that the correlation coefficient is less than or equal to one in absolute value:

$$[Cov(S, U)]^2 \leq V(S) \cdot V(U).$$

Now let us take $S = W$ and $U = \frac{\partial \ln f}{\partial \theta}(X; \theta)$. First consider U , known as the score function, and its expectation. Because for all θ ,

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

we have,

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

Since we can change the order of differentiation and integration by assumption, we get

$$\begin{aligned} 0 &= \int_x \frac{\partial f_X}{\partial \theta}(x; \theta) dx \\ &= \int_x \frac{\partial \ln f_X}{\partial \theta}(x; \theta) \cdot f_X(x; \theta) dx \\ &= E \left[\frac{\partial \ln f_X}{\partial \theta}(x; \theta) \right] = E[U] = 0. \end{aligned}$$

Therefore the covariance of U and S is the expectation of the product of S and U :

$$\begin{aligned} E[SU] &= \int_x W \frac{\partial \ln f_X}{\partial \theta}(x; \theta) f_X(x; \theta) dx = \int_x W \frac{\partial f_X}{\partial \theta}(x; \theta) dx \\ &= \frac{\partial}{\partial \theta} \int_x W f(x; \theta) dx = \frac{\partial}{\partial \theta} \theta = 1. \end{aligned}$$

So

$$1 \leq V(W) \cdot V(U),$$

implying

$$V(W) \geq 1/V(U) = 1/E[U^2].$$

□

Of course finding a lower bound for the variance is not so hard. Zero is a lower bound that applies with no conditions attached. The interest in the Cramér–Rao bound stems largely from the fact that in many cases the bound can actually be reached; there are often estimators with variance equal to the bound.

Example

Suppose X has an exponential distribution with mean μ . Consider the estimator $\hat{\mu} = X$. This estimator is unbiased with variance μ^2 . To calculate the Cramér–Rao bound, consider the log of the density:

$$-\ln(\mu) - x/\mu.$$

The derivative of the log of the density—the score function—is

$$-1/\mu + x/\mu^2 = (x - \mu)/\mu^2.$$

Clearly this has expectation zero (this is a good sign that the regularity conditions are satisfied. If it did not hold either your algebra is wrong or one of the regularity conditions is not satisfied. See some of the examples below.) The variance is

$$E(X - \mu)^2/\mu^4 = 1/\mu^2,$$

and the Cramér–Rao bound is μ^2 . This is the variance of the unbiased estimator $\hat{\mu}$ suggested, so that estimator is the minimum variance unbiased estimator. □

A corollary of the Cramér–Rao bound is the following result for N iid random variables.

Result 2 *Let X_1, \dots, X_N be iid random variable with common pdf/pmf $f_X(x; \theta)$, and let W be an unbiased estimator for θ . Then*

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

Example

Suppose X_1, \dots, X_N are independent with normal distributions with mean μ and known variance σ^2 . The obvious estimator for the mean is the sample average \bar{x} with variance σ^2/N . Consider the log of the density function:

$$\ln f_X(x; \mu) = -\frac{1}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x - \mu)^2.$$

The score function is

$$\frac{\partial}{\partial \mu} \ln f_X(x; \mu) = \frac{1}{\sigma^2}(x - \mu).$$

Again the score clearly has expectation zero, and the variance is $\sigma^2/\sigma^4 = 1/\sigma^2$, and therefore the Cramér–Rao bound for the single observation case is σ^2 , and the Cramér–Rao bound for the N observation case is σ^2/N . \square

Example

Finally let us consider an example where the Cramér–Rao bound does not apply. Recall that in the proof we have to be able to reverse the order of integration and differentiation. That does not work if the argument of the function enters in the bounds of the integral. Suppose X has a uniform distribution on the interval from zero to θ , $X \sim U[0, \theta]$. The log of the density function is

$$\ln f_X(x; \theta) = -\ln \theta.$$

The derivative is

$$\frac{\partial}{\partial \theta} \ln f_X(x; \theta) = -1/\theta.$$

Note that this clearly does not have expectation zero, which is a property we used in the proof of the Cramér–Rao bound. Nevertheless, let us ignore this and proceed with the calculation. The expectation of the square is $1/\theta^2$, and the Cramér–Rao bound is equal to θ^2 . Now consider the estimator $2X$. It clearly is unbiased. Its variance is $\theta^2 \cdot 4/12 = \theta^2/3$, lower than the Cramér–Rao bound, which does not apply in this case. \square

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

Result 3 (CRAMÉR–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

$$\begin{aligned} 0 &= \int_x \frac{\partial f_X}{\partial \theta}(x; \theta) dx \\ &= \int_x \frac{\partial \ln f_X}{\partial \theta}(x; \theta) \cdot f_X(x; \theta) dx. \end{aligned}$$

Now differentiate again to get

$$\begin{aligned}
0 &= \int_x \frac{\partial^2 \ln f_X}{\partial \theta^2}(x; \theta) \cdot f_X(x; \theta) dx + \int_x \frac{\partial \ln f_X}{\partial \theta}(x; \theta) \cdot \frac{\partial f_X}{\partial \theta}(x; \theta) dx \\
&= \int_x \frac{\partial^2 \ln f_X}{\partial \theta^2}(x; \theta) \cdot f_X(x; \theta) dx + \int_x \frac{\partial \ln f_X}{\partial \theta}(x; \theta) \cdot \frac{\partial \ln f_X}{\partial \theta}(x; \theta) f_X(x; \theta) dx \\
&= \int_x \frac{\partial^2 \ln f_X}{\partial \theta^2}(x; \theta) \cdot f_X(x; \theta) dx + \int_x \left(\frac{\partial \ln f_X}{\partial \theta}(x; \theta) \right)^2 f_X(x; \theta) dx \\
&= E \left[\frac{\partial^2 \ln f}{\partial \theta^2}(x; \theta) \right] + E \left[\frac{\partial \ln f}{\partial \theta}(x; \theta)^2 \right].
\end{aligned}$$

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

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 Cramér-Rao bound. \square

Example

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

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

From the form of the density we can tell that $X_1 + X_2$ is a sufficient statistic. What is the Cramér-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

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

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 $a_2 = -a_1$, or

$$W = a \cdot (X_1 - X_2) + X_1 \cdot X_2.$$

Next, use Rao–Blackwell to make the estimator a function of the sufficient statistic $X_1 + X_2$.

$$E[X_1 - X_2 | X_1 + X_2] = 0,$$

(check this for $X_1 + X_2 = 0$, $X_1 + X_2 = 1$, and for $X_1 + X_2 = 2$), and

$$E[X_1 \cdot X_2 | X_1 + X_2] = 1\{X_1 + X_2 = 2\} = X_1 \cdot X_2,$$

implying 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 Cramér–Rao bound. Nevertheless, it is the minimum variance unbiased estimator. \square

Now let us investigate when we have an unbiased estimator with variance equal to the Cramér–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 $b(\theta) = -a(\theta) \cdot \theta$, and hence 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 Cramér–Rao bound.

Result 4 *An unbiased estimator with variance equal to the Cramér–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

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

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:

$$V(W(X)) = \frac{1}{a(\theta)^2} \cdot V(\mathcal{S}(X; \theta)) = \frac{1}{a(\theta)^2} \cdot CR^{-1}.$$

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

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

Hence

$$CR = \frac{1}{a(\theta)},$$

implying that

$$V(W(X)) = \frac{1}{a(\theta)^2} \cdot CR^{-1} = \frac{1}{a(\theta)}.$$

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. \square