Probability and Statistics

Fall 2000

Lect. 13: Rao–Blackwell Theorem, Cramér–Rao Bound, CB 6.1.4, CB 7.3.2–7.3.3

We are still in the context of a random variable X with a probability density (mass) function $f_X(x;\theta^*)$, for some unknown θ^*, and with the function $f_X(\cdot)$ known.

Suppose we have a sufficient statistic $T(X)$. How can we use this? The Rao-Blackwell Theorem deals with this question.

Result 1 *(*Rao–Blackwell Theorem*)*
Let $W = W(X)$ be any unbiased estimator for θ, and let $T = T(X)$ be a sufficient statistic for θ. Then

$$\tilde{W} = E[W|T],$$

is an unbiased estimator for θ with a variance less than or equal to that of W.

Proof:
First consider the expectation of $\tilde{W}$. The law of iterated expectations says in general that

$$E[Y] = E\left[E[Y|X]\right],$$

because

$$E[Y] = \int\int y f_{XY}(x,y)dydx = \int\int y f_{Y|X}(y|x)f_X(x)dydx$$

$$= \int E[Y|X]f_X(x)dx = E[E[Y|X]].$$

Therefore we have

$$\theta = E[W] = E[E[W|T]] = E[\tilde{W}],$$

showing that $\tilde{W}$ is unbiased. In addition,

$$V(W) = V(E[W|T]) + E[V(W|T)] = V(\tilde{W}) + E[V(W|T)] \geq V(\tilde{W}).$$

We use here the fact that

$$Y = E[Y|X] + Y - E[Y|X],$$

so

$$V(Y) = V(E[Y|X]) + V(Y - E[Y|X]) + 2 \cdot COV(E[Y|X], Y - E[Y|X])$$

$$= V(E[Y|X]) + E[V(Y|X)],$$

where we use the fact that

$$V(Y - E[Y|X]) = E[(Y - E[Y|X])^2] = E[E[(Y - E[Y|X])^2]] = E[V(Y|X)].$$

So one way of using sufficient statistics is to first look for any unbiased estimator and then take its conditional expectation given the sufficient statistic. This will never make you worse off, and can actually improve things. In practice this is not an easy thing to do. The conditional distribution of the unbiased estimator given the sufficient statistic is often difficult to calculate, and calculating its expectation is often even more difficult. The main value of the result is in formalizing the notion that we can do as well with estimators that are functions of the sufficient statistic as with the general set of all estimators.

The reason we can only use this trick with sufficient statistics is that in general $E[W|S]$ depends on the unknown θ, unless S is a sufficient statistic. Of course this still does not narrow things down very much. Another important result that helps us in the search for a minimum variance unbiased estimator is the Cramèr–Rao bound:

Result 2 (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

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

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:

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

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,

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

Assuming we can change the order of differentiation and integration, we get

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

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

$$= E[\frac{\partial \ln f_X}{\partial \theta}(x;\theta)] = E[U] = 0.$$

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

$$E[SU] = \int_x W\frac{\partial \ln f_X}{\partial \theta}(x;\theta)f_X(x;\theta)dx = \int_x W\frac{\partial f_X}{\partial \theta}(x;\theta)dx$$

$$= \frac{\partial}{\partial\theta}\int_x W f(x;\theta)dx = \frac{\partial}{\partial\theta}\theta = 1.$$

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

Result 3 Let $X_1, \ldots, X_N$ 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 $X_1, \ldots, X_N$ 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 Cramèr–Rao bound for the single observation case is σ^2, and the Cramèr–Rao bound for the N observation case is σ^2/N. $\square$

Example

Finally let us consider an example where the Cramèr–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 Cramèr–Rao bound. Nevertheless, let us ignore this and proceed with the calculation. The expectation of the square is $1/\theta^2$, and the Cramèr–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 Cramèr–Rao bound, which does not apply in this case. $\Box$