Handout 19: Cramér-Rao Lower Bound

Today we derive the most theoretical result of the semester: the **Cramér-Rao** lower bound called the on the variance of an unbiased estimator. Since the variance of an unbiased estimator gives the expected error of the estimator, this bound provides a best-case scenario on how well we can estimate a parameter. We will see on today's worksheet that many of the MLE estimators we have so far derived achieve this lower bound.

Consider a random variable X with a probability density function $f(\theta; x)$, which has a one univariate parameter θ. We can define a random variable V, called the **score function**, as the derivative of the logarithm of the density of X. [1] We see that this has a nice form by applying the chain rule:

$$\begin{aligned} V &= \frac{\partial}{\partial\theta}\left[\log(f(\theta; X))\right] \\ &= \frac{1}{f(\theta; X)} \cdot \frac{\partial}{\partial\theta}\left[f(\theta; X)\right]. \end{aligned}$$

The score, under some fairly general regularity conditions, will have an expected value of zero:

$$\begin{aligned} \mathbb{E}V &= \int f(\theta; x) \cdot \frac{1}{f(\theta; x)} \cdot \frac{\partial}{\partial\theta}\left[f(\theta; x)\right] dx \\ &= \int \frac{\partial}{\partial\theta} f(\theta; x) dx = \frac{\partial}{\partial\theta} \int f(\theta; x) dx = \frac{\partial}{\partial\theta}[1] = 0. \end{aligned}$$

The variance of the score is called the **Fisher information** and serves as a measurment of how much information about θ is provided by the data X. This is often written as $\mathcal{I}(\theta)$. Note that the variance is just the expected value of V^2.

Now, let $T = t(X)$ be a point estimator for the parameter θ with expectation $\mathbb{E}T = \psi(\theta)$. In other words, if T is unbiased, we would have $\psi(\theta) = \theta$ for all values of θ. If we look at the covariance of T and V, note that this is equal to just $\mathbb{E}[VT]$ since the expected value of V is zero.[2] This has, by construction, a nice form:

$$\begin{aligned} Cov(V, T) &= \int \left[t(x) \times \frac{1}{f(\theta; x)} \times \frac{\partial}{\partial\theta}\left[f(\theta; x)\right]\right] dx \\ &= \frac{\partial}{\partial\theta}\left[\int t(x) f(\theta, x) dx\right] = \frac{\partial}{\partial\theta}\mathbb{E}T = \psi'(\theta). \end{aligned}$$

Next, we need to use the **Cauchy-Schwartz Inequality**, which for probability spaces says that covariance of two random variables is always less in absolute value than the square-root of the product of their variances.[3] Applying this to T and V shows that:

[1] The score function is a bit unusual because we are evaluating the density of the random variable at the value of the random variable. In other words, we have $f(\theta; X)$ rather than the usual $f(\theta; x)$.

[2] Recall that the covariance in general would be $\mathbb{E}[(V - \mathbb{E}V)(T - \mathbb{E}T)]$.

[3] The more general form says that the squared inner product $|\langle u, v\rangle|^2$. is less than $\langle u, u\rangle \cdot \langle v, v\rangle$. Applying this to the integration with density f yields the probabilistic version.

$$\begin{aligned} Var(T) \cdot Var(V) &\geq |Cov(V,T)|^2 \\ Var(T) \cdot \mathcal{I}(\theta) &\geq |\psi'(\theta)|^2 \\ Var(T) &\geq \frac{|\psi'(\theta)|^2}{\mathcal{I}(\theta)}. \end{aligned}$$

Which implies, in the unbiased case, that the variance of T can never be less than the inverse of the Fisher information. So, this provides a bound on the best that we can hope to do in terms of estimating the parameter θ from the data X.

The **efficency** of an unbiased estimator, written $e(\widehat{\theta})$, provides a measurement of how far away the variance of the estimator is away from the Cramèr-Rao bound. Namely, we have:

$$e(\widehat{\theta}) = \frac{\mathcal{I}(\theta)^{-1}}{Var(\widehat{\theta})}.$$

We say that an estimator is **efficent** if it has an effiency of 1. Another way to state the Cramér-Rao bound is to simply say that the efficency is never greater than 1.

Under some regularity conditions—in particular, that the logarithm of the density function f is twice-differentiable—the Fisher information can be written in a somewhat simplified form:

$$\mathcal{I}(\theta) = \mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(\theta;x)\right].$$

Typically, squaring the log density requires having a number of cross terms, whereas the second derivative removes a number of terms, simplifying the calculation. This is the version that we will use on the worksheet.

It is possible to extend the result above to the case where X and θ are vectors. The extension for a vector X, which includes the important case of a random sample of size n, is fairly trivial. We just replace all of the single integrals above with n-dimensional integrals over $\mathbb{R}^n$. Generalizing to a vector value for θ is a bit more work, requiring some vector calculus that goes beyond the prerequisites for this course. The general idea, however, is very similar.