

5. Examples, examples, and counterexamples

GULP Winter 2015

February 11, 2015

1 Unbiased estimator that hits the C-R Bound

So far we know that if we have an unbiased estimator that has a variance equivalent to the C-R bound, then the unbiased estimator is the best among all the unbiased estimators. Here we show some examples, in both discrete and continuous case.

Binomial. Suppose we have one sample observation X coming from binomial distribution (n, p) and we need to estimate p . One natural unbiased estimator is X/n , with variance $p(1-p)/n$. Then we seek for the Fisher information of the observation:

$$\begin{aligned} I(p) &= -\mathbb{E}\left(\frac{\partial^2}{\partial p^2} \ln f(X, p)\right) \\ &= -\mathbb{E}\left(\frac{\partial^2}{\partial p^2} \ln\left(\binom{n}{k} p^X (1-p)^{n-X}\right)\right) \\ &= -\mathbb{E}\left(-\frac{X}{p^2} - \frac{n-X}{(1-p)^2}\right) \\ &= \frac{n}{p(1-p)} \end{aligned}$$

Since the C-R bound is just the inverse of the Fisher information, we know that the estimator is actually hitting the lower bound.

Normal Variance. Suppose we have n independent identically distributed samples from $N(\mu, \sigma^2)$ with known μ . We propose the following estimator:

$$S^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n}$$

Since for single observation we have $\mathbb{E}(X_i - \mu)^2 = \sigma^2$, S^2 is an unbiased estimator. Using the tool of χ^2 distribution we can justify that $\text{Var}(S^2) = 2\sigma^4/n$. Then we calculate the Fisher

information for one observation:

$$\begin{aligned}
I(\sigma^2) &= -\mathbb{E}\left(\left(\frac{\partial^2}{\partial(\sigma^2)^2} \ln f(X, \mu, \sigma^2)\right)\right) \\
&= -\mathbb{E}\left(\left(\frac{\partial^2}{\partial p^2} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}\right)\right)\right) \\
&= -\mathbb{E}\left(\frac{1}{2(\sigma^2)^2} - \frac{(X-\mu)^2}{(\sigma^2)^3}\right) \\
&= \frac{1}{2\sigma^4}
\end{aligned}$$

n samples contains $nI(\sigma^2)$ information therefore the C-R bound is $1/nI(\sigma^2) = 2\sigma^4/n$. Therefore S^2 is also the best unbiased estimator.

You may justify that the sample mean of the Poisson samples is the best unbiased estimator to estimate the λ , and sample mean is also best for normal samples.

2 Unbiased estimator that does not hit the C-R Bound, and other

One simple example shows that the C-R bound can not be reached every time. Suppose we want to estimate σ^2 with μ unknown, under normal distribution settings. It can be shown (at the next time) that the best unbiased estimator is

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}.$$

The variance for this estimator is $2\sigma^4/(n-1)$ which is strictly larger than the C-R bound shown before.

Also, sometimes the method will break down because the density function is badly behaved. And sometimes it is hard to tell that they are badly behaved. We have a typical example when this method breaks down.

Uniform distribution. Suppose we have samples from $U(0, \theta)$, and we are interested with θ . Some calculation will give you the Fisher information for one observation is $1/\theta^2$, then the C-R bound for n observations should be θ^2/n . However, we can construct an unbiased estimator which has a smaller variance.

Consider the estimator $(n+1)X_{(n)}/n$, where $X_{(n)}$ is the largest observation among n samples. It can be shown that this estimator is unbiased, with variance $\theta^2/(n^2+n)$. This is much less than the C-R bound.

The reason that the method fail is that we can not interchange the partial differentiation and the integration as we have done before. The critical point is that the support of the uniform distribution is depending on the parameter, which is not a good behavior.

There are a bunch of the assumptions that the density function has to satisfy, in order to make this C-R bound calculation valid. In general it can be roughly described as 'smooth enough against the parameter' and 'support independent of the parameter'.