Lecture 17. Relative Efficiency, Cramér-Rao Bound

11/19/2011

When a lot is known about the distribution of interest, there often are many ways to estimate the same parameter. One example is given at the end of Lecture 16. In that example, the distribution of interest is that of X and is known (assumed) to be $Uniform[0, \theta]$ with an unknown parameter θ. At least three estimators can be motivated by the sample analogue principle. Which one is better?

In **finite samples,** compare the MSE.

In the uniform example, the first estimator $\hat{\theta}_{1,n} = 2\bar{X}_n$. $E_\theta(2\bar{X}_n) = 2(\theta/2) = \theta$. $Var_\theta(2\bar{X}_n) = 4 \times (\theta^2/12)/n = \theta^2/(3n)$. Thus,

$$MSE_\theta(\hat{\theta}_{1,n}, \theta) = Bias^2(\hat{\theta}_{1,n}, \theta) + Var_\theta(\hat{\theta}_{1,n}) = \theta^2/(3n).$$

The third estimator $\hat{\theta}_{3,n} = \max_{i=1,\dots,n}\{X_i\}$ has cdf:

$$\begin{aligned} F_{\hat{\theta}_{3,n}}(x) &= \Pr{}_\theta(\max_{i=1,\dots,n}\{X_i\} \leq x) \\ &= \Pr{}_\theta(X_i \leq x \text{ for all } i) \\ &= \times_{i=1}^n \Pr{}_\theta(X_i \leq x) \quad \text{by i.i.d.} \\ &= (x/\theta)^n \cdot 1(x \in [0, \theta]) + 1(x > \theta). \quad \text{by Uniform.} \end{aligned} \tag{1}$$

Thus

$$
\begin{aligned}
E_\theta[\hat{\theta}_{3,n}] &= \int_0^\theta x d(x/\theta)^n \\
&= \int_0^\theta n x^n/\theta^n dx \\
&= \int_0^\theta (n/(\theta^n(n+1)) dx^{n+1} \\
&= (n/(\theta^n(n+1)) x^{n+1}|_0^\theta \\
&= n\theta/(n+1). \qquad (2)
\end{aligned}
$$

($\hat{\theta}_{3,n}$ is not unbiased.) And

$$
\begin{aligned}
E_\theta[\hat{\theta}_{3,n}^2] &= \int_0^\theta x^2 d(x/\theta)^n \\
&= \int_0^\theta n x^{n+1}/\theta^n dx \\
&= \int_0^\theta (n/(\theta^n(n+2)) dx^{n+2} \\
&= (n/(\theta^n(n+2)) x^{n+2}|_0^\theta \\
&= n\theta^2/(n+2). \qquad (3)
\end{aligned}
$$

Thus, $Var_\theta(\hat{\theta}_{3,n}^2) = E_\theta[\hat{\theta}_{3,n}^2] - (E_\theta[\hat{\theta}_{3,n}])^2 = n\theta^2/(n+2) - n^2\theta^2/(n+1)^2$. The MSE is

$$
\begin{aligned}
MSE_\theta(\hat{\theta}_{3,n}, \theta) &= Bias^2(\hat{\theta}_{3,n}, \theta) + Var_\theta(\hat{\theta}_{3,n}) \\
&= (\theta^2/(n+1)^2) + n\theta^2/(n+2) - n^2\theta^2/(n+1)^2 \\
&= \frac{(1-n^2)\theta^2}{(n+1)^2} + \frac{n}{n+2}\theta^2 \\
&= \frac{2\theta^2}{(n+1)(n+2)}. \qquad (4)
\end{aligned}
$$

Compare $MSE_\theta(\hat{\theta}_{3,n}, \theta)$ and $MSE_\theta(\hat{\theta}_{1,n}, \theta)$ and we can say that $\hat{\theta}_{3,n}$ is better than $\hat{\theta}_{1,n}$ for all $n > 2$ and all $\theta > 0$ and the difference widens as n grows. This example tells us that a little bias may go a long way reducing the variance!

MSE are usually hard to compute when the estimator is not a sample mean of i.i.d. random variables. For example, the cdf of $\hat{\theta}_{2,n}$ in the uniform distribution can be worked out, but it is too complicated to make calculating MSE a sensible thing to do. A much easier task is to restrict ourselves to consistent estimators and compare their asymptotic variances.

This only requires us to (1) show consistency and (2) derive the asymptotic distribution. The first task is aided by the law of large numbers and the Slutsky theorem and the second task is aided by the central limit theorem, the continuous mapping theorem and the delta method.

Use the uniform distribution example again.

We have learned the consistency and asymptotic distribution of a sample mean in previous lectures. Using those, we immediately obtain:

$$\hat{\theta}_{1,n} \to_p \theta \text{ and } \sqrt{n}(\hat{\theta}_{1,n} - \theta) \to_d N(0, \theta^2/3). \tag{5}$$

Deriving the asymptotic distribution of the second estimator can use some magic. The magic is the following Lemma:

Lemma 1. *Suppose* $Y_1, Y_2, ...Y_n) \sim i.i.d.Uniform[0,1]$*. Then the sample median has the same distribution as*

$$\frac{\sum_{i=1}^{\lceil n/2 \rceil} Z_i}{\sum_{i=1}^{n+1} Z_i},$$

where $(Z_1, ..., Z_{n+1}) \sim i.i.d.Exp(1)$*. (Ferguson 1996).*

The estimator $\hat{\theta}_{2,n}$ is the sample median of a uniform distribution.

...

Now we know how to judge which estimator is better, we can ask the question: what's the best an estimator can get?

Cramer-Rao Bound

In the case when we know the distribution of a random variable/vector of interest up to a finite dimensional parameter, the Cramer-Rao bound defines the lower bound of the variance of any unbiased estimator.