Lecture 18: Performance Analysis of Estimators and Cramer-Rao Lower Bound

15 Mar 2016

Continuation of last lecture

In the last lecture, we define conjugate family for a class of pdf's or pmf's. Now, we look at one example for this.

Example 0.1. (**Normal Bayes Estimators**)

Let $X_1, X_2, \ldots, X_n$ be i.i.d $\mathcal{N}(\theta, 1)$ and suppose that prior distribution on θ is $\mathcal{N}(\mu, \tau^2)$, (here we assume that μ and τ^2 are both known). The posterior distribution of θ is also normal with mean and variance given by,

$$\mathbb{E}(\theta|x) = \frac{\tau^2}{\frac{1}{n} + \tau^2}\bar{x} + \frac{\frac{1}{n}}{\frac{1}{n} + \tau^2}\mu, \text{ and } Var(\theta|x) = \frac{\frac{1}{n}\tau^2}{\frac{1}{n} + \tau^2}. \tag{1}$$

Note that posterior distribution belongs to the same class of distribution as that of prior distribution. Hence normal family is its own conjugate family for fixed variance case.

Remark 1. Posterior mean interpolates between the prior mean and the sample mean.

1 Performance Analysis of Estimators

So far, we have studied different types of estimators. in this section we will study the performance of these estimators with respect to some performance criteria. Although there are several such criteria, here we will focus our attention to the mean squared error (MSE) criterion.

1.1 Mean Squared Error Criterion

MSE measures the average squared difference between the estimator W and the parameter θ and is a reasonable measure of performance for a point estimator. Moreover, it is quite tractable analytically.

Definition 1.1. The MSE of a point estimator W of a parameter θ is a function of θ denoted by $MSE(W;\theta)$ and defined by

$$MSE(W;\theta) = \mathbb{E}_\theta[(W-\theta)^2]$$

Specifying the dependence of W on X explicitly, it turns out that

$$MSE(W;\theta) = \mathbb{E}_\theta[(W(X)-\theta)^2]$$

Applying the definition of expectation, the above equation can be written as

$$MSE(W;\theta) = \int_{\mathbb{R}^n} (W(x)-\theta)^2 f(x|\theta)dx \tag{2}$$

1.1.1 Bias - Variance Decomposition of MSE

Definition 1.2. Variance of a point estimator W, of a parameter θ is the average value of squared difference between the estimator W and its expected value $\mathbb{E}_\theta[W]$; that is

$$Var_\theta[W] = \mathbb{E}_\theta[(W-\mathbb{E}_\theta(W))^2]$$

Definition 1.3. Bias of a point estimator W of a parameter θ is the difference between the expected value of W and θ; that is

$$Bias_\theta[W] = \mathbb{E}_\theta[W] - \theta$$

Remark 2. Variance of an estimator represents its variability or fluctuation.

Remark 3. Bias represents average accuracy of the estimator.

Remark 4. For *unbiased* estimators, bias is identically equal to zero and satisfies $\mathbb{E}_\theta[W] = \theta$ for all θ in Θ.

By the definition of MSE,

$$
\begin{aligned}
MSE(W;\theta) &= \mathbb{E}_\theta[(W-\theta)^2], \\
&= \mathbb{E}_\theta[(W-\mathbb{E}_\theta(W)+\mathbb{E}_\theta(W)-\theta)^2], \\
&= \mathbb{E}_\theta[(W-\mathbb{E}_\theta(W))^2] + (\mathbb{E}_\theta(W)-\theta)^2, \\
&= Var_\theta[W] + (Bias_\theta[W])^2. \tag{3}
\end{aligned}
$$

Where, in the second equality, we have added and subtracted $\mathbb{E}_\theta[W]$. Third equality follows from the fact that the cross term is equal to zero, and last equality follows from the definition of variance and bias of an estimator.

Remark 5. For unbiased estimator $MSE[W; \theta] = Var_\theta[W]$ that is MSE is equal to the variance of the unbiased estimator.

Now, we look at some examples.

Example 1.4. (Normal MSE)
Let $X_1, X_2, \ldots, X_n$ be i.i.d samples from $\mathcal{N}(\mu, \sigma^2)$. The statistics $\bar{X}$ and S^2 are both unbiased estimators since

$$\mathbb{E}[\bar{X}] = \mu \ , \ \mathbb{E}[S^2] = \sigma^2 \ \forall \ \mu \text{ and } \sigma^2.$$

Now MSE's of these estimators are given by ,

$$MSE(\bar{X}; (\mu, \sigma^2)) = Var_{\mu,\sigma^2}(\bar{X}) = \mathbb{E}_{\mu,\sigma^2}[(\bar{X} - \mu)^2] = \frac{\sigma^2}{n}, \tag{4}$$

$$MSE(S^2; (\mu, \sigma^2)) = Var_{\mu,\sigma^2}(S^2) = \mathbb{E}_{\mu,\sigma^2}[(S^2 - \sigma^2)^2] = \frac{2\sigma^4}{n-1}. \tag{5}$$

An alternative estimator for σ^2 is the maximum likelihood estimator

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

$$\mathbb{E}[\hat{\sigma}^2] = \mathbb{E}\left[\frac{n-1}{n}S^2\right] = \frac{n-1}{n}\sigma^2,$$

$$Var[\hat{\sigma}^2] = Var\left[\frac{n-1}{n}S^2\right] = \frac{2(n-1)}{n^2}\sigma^4,$$

$$Bias[\hat{\sigma}^2] = (\mathbb{E}[\hat{\sigma}^2] - \sigma^2)^2 = \left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2 = \frac{\sigma^4}{n^2}. \tag{6}$$

Therefore,

$$MSE(\hat{\sigma}^2; (\mu, \sigma^2)) = \frac{2(n-1)}{n^2}\sigma^4 + \frac{\sigma^4}{n^2} = \frac{(2n-1)}{n^2}\sigma^4 < \frac{2\sigma^4}{n-1}, \tag{7}$$

which shows that $\hat{\sigma}^2$ has smaller MSE than S^2.

Remark 6. Unbiased estimator are natural but not necessarily MSE optimal.

2   Best Unbiased Estimators(BUE)

In the previous section we saw that there is no one "best MSE" estimator. The reason is that the class of all estimators is too large a class. One way of finding a "best estimator" is to limit the class of estimators. So here we will consider the class of unbiased estimators only.

Definition 2.1. An estimator W^* is a best unbiased estimator of $g(\theta)$ if it satisfies

1. $\mathbb{E}_\theta[W^*] = g(\theta)$ for all θ in Θ and,

2. For any other estimator W with ,

$$\mathbb{E}_\theta[W] = g(\theta) \quad \forall \theta \in \Theta, \tag{8}$$

we have $Var_\theta(W^*) \leq Var_\theta(W), \quad \forall \theta \in \Theta$.

W^* is also called a uniform minimum variance unbiased estimator(UMVUE) of $g(\theta)$.

Theorem 2.2. *(**Cramer-Rao Lower Bound**) Let $X_1, X_2, \ldots, X_n$ be a sample with pdf $f(x|\theta)$ and let $W(X) = W(X_1, X_2, \ldots, X_n)$ be any estimator satisfying*

1. *$\frac{d}{d\theta}\mathbb{E}_\theta[W(X)] = \int_{\mathbb{R}^n} \frac{\partial}{\partial\theta}[W(x)f(x|\theta)]dx$, and*

2. *$Var_\theta(W(X)) < \infty$,*

then

$$Var_\theta(W(X)) \geq \frac{\left(\frac{d}{d\theta}\mathbb{E}_\theta[W(X)]\right)^2}{\mathbb{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right)}. \tag{9}$$

Proof. The proof of this theorem follows from a clever application of Cauchy-Schwartz inequality. The basic Cauchy-Schwartz inequality is given by,

$$\sum_{i=1}^n a_i b_i \leq \sqrt{\sum_{i=1}^n a_i^2}\sqrt{\sum_{i=1}^n b_i^2}.$$

For two random variables A and B, this inequality can be written as,

$$\mathbb{E}(AB) \leq \sqrt{\mathbb{E}(A^2)}\sqrt{\mathbb{E}(B^2)}. \tag{10}$$

Substituting, $A \triangleq U - \mathbb{E}(U)$, and $B \triangleq V - \mathbb{E}(V)$ for some random variables U and V, it turns out that

$$COV(U, V) \leq \sqrt{Var(U)}\sqrt{Var(V)}$$
$$[COV(U, V)]^2 \leq Var(U)\,Var(V)$$
$$Var(U) \geq \frac{[COV(U, V)]^2}{Var(V)}. \tag{11}$$

Now, choose $U = W(X)$ and $V = \frac{\partial}{\partial\theta}\log f(X|\theta)$, we get

$$\frac{d}{d\theta}\mathbb{E}_\theta[W(X)] = \int_{\mathbb{R}^n} W(x)\Big[\frac{\partial}{\partial\theta}f(x|\theta)\Big]dx,$$

$$= \mathbb{E}_\theta\Big[W(X)\frac{\frac{\partial}{\partial\theta}f(X|\theta)}{f(X|\theta)}\Big],$$

$$= \mathbb{E}_\theta\Big[W(X)\frac{\partial}{\partial\theta}\log f(X|\theta)\Big]. \tag{12}$$

First equality follows from the hypothesis of CRLB, second equality is obtained by multiplying $\frac{f(X|\theta)}{f(X|\theta)}$ in the above equality and then using the definition of expectation, and last equality follows from the property of logarithm. Put $W(X) = 1$ in the above equation, we get,

$$\mathbb{E}_\theta\Big[\frac{\partial}{\partial\theta}\log f(X|\theta)\Big] = \frac{d}{d\theta}\mathbb{E}_\theta[1] = 0. \tag{13}$$

Therefore, $\mathbb{E}_\theta\Big[W(X)\frac{\partial}{\partial\theta}\log f(X|\theta)\Big]$ represents covariance between $W(X)$ and $\frac{\partial}{\partial\theta}\log f(X|\theta)$, i.e.,

$$COV_\theta\Big[W(X),\frac{\partial}{\partial\theta}\log f(X|\theta)\Big] = \mathbb{E}_\theta\Big[W(X)\frac{\partial}{\partial\theta}\log f(X|\theta)\Big] = \frac{d}{d\theta}\mathbb{E}_\theta[W(X)]. \tag{14}$$

Now,

$$Var_\theta\Big[\frac{\partial}{\partial\theta}\log f(X|\theta)\Big] = \mathbb{E}_\theta\Big(\frac{\partial}{\partial\theta}\log f(X|\theta)\Big)^2. \tag{15}$$

Using the Cauchy-Schwartz inequality together with the above two equalities, we obtain

$$Var_\theta(W(X)) \geq \frac{\Big(\frac{d}{d\theta}\mathbb{E}_\theta[W(X)]\Big)^2}{\mathbb{E}_\theta\Big(\big(\frac{\partial}{\partial\theta}\log f(X|\theta)\big)^2\Big)} \tag{16}$$

This completes the proof. □