MA 3140: Statistical Inference

Dr. Sameen Naqvi
Department of Mathematics, IIT Hyderabad
Email id: sameen@math.iith.ac.in

Example 4: Uniform Distribution

Let $X_1, \ldots, X_n$ be a random sample from a Uniform distribution on $[0, \theta]$. Find Unbiased and Consistent Estimators.

Solution: The pdf is

$$f_{X_i}(x) = \begin{cases} \frac{1}{\theta}, & 0 \leq x_i \leq \theta \\ 0, & \text{otherwise.} \end{cases}$$

Here, $E(X_i) = \frac{\theta}{2}$.

$$E(\overline{X}) = \frac{\theta}{2} \Longrightarrow E(2\overline{X}) = \theta.$$

Thus, $T_1 = 2\overline{X}$ is an unbiased and consistent estimator for θ.

Example 4 cont'd

- Now, consider $X_{(n)} = \max\{X_1, \ldots, X_n\}$.
  We know that the pdf of $X_{(n)}$ is:

$$\because f_{X_{(n)}}(x) = \begin{cases} \frac{nx^{n-1}}{\theta^n}, & 0 \leq x \leq \theta \\ 0, & \text{otherwise.} \end{cases}$$

$$E(X_{(n)}) = \int_0^\theta x \ f_{X_{(n)}}(x) dx = \int_0^\theta x \frac{nx^{n-1}}{\theta^n} dx = \frac{n}{n+1}\theta$$

$$\implies E\left[\frac{n+1}{n} X_{(n)}\right] = \theta$$

Thus, $T_2 = \frac{n+1}{n} \ X_{(n)}$ is an UE of θ.

Example 4 cont'd

Now, we check for consistency.

$$P(|X_{(n)} - \theta| > \epsilon) = P(\theta - X_{(n)} > \epsilon) = P(X_{(n)} < \theta - \epsilon)$$
$$= \left(\frac{\theta - \epsilon}{\theta}\right)^n \longrightarrow 0 \text{ as } n \longrightarrow \infty$$

Thus, $X_{(n)}$ is consistent for θ.
$\Longrightarrow T_2 = \frac{n+1}{n} X_{(n)}$ is also consistent for θ.

Hence, T_2 is also unbiased and consistent for estimating θ.

Methods of evaluating estimators

Methods of evaluating estimators

- Mean Squared Error
- Best Unbiased Estimators
- Sufficiency and Unbiasedness

Mean Squared Error (MSE)

- $E[T] - g(\theta)$: Expected Error or Bias
  $E|T - g(\theta)|$: Mean Absolute Error
  $E[T - g(\theta)]^2$: Mean Squared Error

- Interpretation of MSE

$$
\begin{aligned}
&E[T - g(\theta)]^2 \\
&= E[T - E(T) + E(T) - g(\theta)]^2 \\
&= E[T - E(T)]^2 + E[E(T) - g(\theta)]^2 + 2E[T - E(T)][E(T) - g(\theta \\
&= V(T) + Bias^2(T) + 0
\end{aligned}
$$

- Thus, $MSE(T) = E[T - g(\theta)]^2 = V(T) + Bias^2(T)$.

MSE cont'd

- Estimator T_1 is said to be better (more efficient) than estimator T_2 if

$$MSE(T_1) \leq MSE(T_2), \; \forall \; \theta \in \Theta$$

- If T is unbiased for $g(\theta)$, then

$$MSE(T) = Var(T)$$

Example 1: Normal Distribution

Let $X_1, \ldots, X_n$ be a random sample from a Normal distribution with mean μ and variance σ^2.

We know that

$$E(\overline{X}) = \mu, \qquad E(S^2) = \sigma^2, \ \forall \ \mu \text{ and } \sigma^2.$$

The MSEs of these estimators are:

$$E(\overline{X} - \mu)^2 = Var(\overline{X}) = \frac{\sigma^2}{n}$$

$$E(S^2 - \sigma^2)^2 = Var(S^2) = \frac{2}{n-1}\sigma^4$$

Proof for $Var(S^2)$:

We know that $Y = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$.

This implies $Var(Y) = 2(n-1)$.

$$\begin{aligned} \frac{(n-1)^2}{\sigma^4} Var(S^2) &= 2(n-1) \\ Var(S^2) &= \frac{2(n-1)}{(n-1)^2}\sigma^4 \\ &= \frac{2}{n-1}\sigma^4 \end{aligned}$$

Remark

Be aware that controlling bias does not guarantee that MSE is controlled.

In some situations, a trade-off occurs between variance and bias in such a way that a small increase in bias can be traded for a larger decrease in variance, resulting in an improvement of MSE.

Example 1 cont'd

Consider an alternative estimator for σ^2, i.e., its MLE

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

Now,

$$E\hat{\sigma}^2 = E\left(\frac{n-1}{n}S^2\right) = \frac{n-1}{n}\sigma^2,$$

which implies $\hat{\sigma}^2$ is a biased estimator of σ^2. The variance of $\hat{\sigma}^2$ is

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

Example 1 cont'd

Hence, its MSE is

$$E(\hat{\sigma}^2 - \sigma^2)^2 = \frac{2(n-1)\sigma^4}{n^2} + \left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2 = \left(\frac{2n-1}{n^2}\right)\sigma^4.$$

Now,

$$E(\hat{\sigma}^2 - \sigma^2)^2 = \left(\frac{2n-1}{n^2}\right)\sigma^4 < \left(\frac{2}{n-1}\right)\sigma^4 = E(S^2 - \sigma^2)^2$$

showing that

$$\text{MSE}(\hat{\sigma}^2) < \text{MSE}(S^2).$$

Thus, by trading off variance for bias, the MSE is improved.

Remark

However, we should not abandon S^2 as an estimator of σ^2.

> On an average, $\hat{\sigma}^2$ will be closer to σ^2 than S^2 if MSE is used as a measure.
>
> Note that since $\hat{\sigma}^2$ is biased, and will, on the average, underestimate σ^2, it is not a good choice for estimating σ^2.

MSE just provides more information about the estimators in the hope that a good estimator is chosen.

Since MSE is a function of the parameter, we cannot find one "best" estimator.

Best Unbiased Estimators

Since we cannot find one "best MSE" estimator, it is recommended to tackle the problem of finding a "best" estimator by limiting ourselves to the class of UEs.

If W_1 and W_2 are 2 UEs of a parameter θ, i.e., $EW_1 = EW_2 = \theta$, then their MSEs are equal to their variances.

Thus, we should choose the estimator which has smaller variance.

Goal: Find an UE with uniformly smallest variance, i.e., a best UE.

Best Unbiased Estimators contd.

An estimator W is said to be a *best unbiased estimator* of $g(\theta)$ if W is unbiased and, for any other unbiased estimator W^* of $g(\theta)$,

$$Var(W) \leq Var(W^*), \ \forall \ \theta \in \Theta.$$

W is also referred to a *Uniformly minimum variance unbiased estimator* (UMVUE) of $g(\theta)$.

Remark

- **Finding a UMVUE is not an easy task!**

  Example: Let $X_1, \ldots, X_n$ be a random sample from Poisson (λ).

  Since $E(X_i) = Var(X_i) = \lambda$, we know that

  $$E\overline{X} = \lambda, \text{ and } ES^2 = \lambda, \ \forall \ \lambda,$$

  implying that both $\overline{X}$ and S^2 are UEs of λ.

  In order to determine the better estimator, we need to compare the variances of $\overline{X}$ and S^2.

  Although we know $Var\overline{X} = \frac{\lambda}{n}$, finding $\mathrm{Var}S^2$ is quite a lengthy calculation.

Remark contd.

In fact, finding $\text{Var}S^2$ may be for nothing because you may end up getting that $Var\overline{X} \leq \text{Var}S^2$.

Now, consider

$$W(\overline{X}, S^2) = a\overline{X} + (1-a)S^2.$$

Here, for every constant a, $E[W(\overline{X}, S^2)] = \lambda$, implying that we have infinitely many UEs of λ.

So, even if we establish that $\overline{X}$ is better than S^2, will it better than every $W(\overline{X}, S^2)$?

Remark

A more comprehensive approach to find a best UE is:

- Specify a lower bound, say $B(\theta)$, on the variance of any UE of $g(\theta)$.
- Find an UE W satisfying $Var(W) = B(\theta)$.

This bound is referred to as Cramer-Rao Lower Bound (CRLB).

Cramer-Frechet-Rao (CFR) Inequality

Let $\Theta \in \mathbb{R}$ be an open interval and suppose that the family $\{f_\theta : \theta \in \Theta\}$ satisfies the following regularity conditions:

(i) It has common support set S, i.e., $S = \{\boldsymbol{x} : f_\theta(\boldsymbol{x}) > 0\}$ does not depend on θ.

(ii) For $\boldsymbol{x}$ and $\theta \in \Theta$, the derivative $\frac{\partial}{\partial\theta} \log f(\boldsymbol{x}, \theta)$ exists and is finite.

(iii) For any statistic h with $E[|h(\boldsymbol{X})|] < \infty$, the operations of integration (summation) and differentiation with respect to θ can be interchanged in $E[h(\boldsymbol{X})]$, i.e.,

$$\frac{\partial}{\partial\theta} \int h(\boldsymbol{x}) f_\theta(\boldsymbol{x}) d\boldsymbol{x} = \int h(\boldsymbol{x}) \frac{\partial}{\partial\theta} f_\theta(\boldsymbol{x}) d\boldsymbol{x}$$

whenever the RHS is finite.

CFR Inequality cont'd

Let $T(\boldsymbol{X})$ be such that $Var[T(\boldsymbol{X})] < \infty$ and define $\psi(\theta) = E[T(\boldsymbol{X})]$.

If

$$I(\theta) = E\left[\frac{\partial}{\partial\theta} \log f_\theta(\boldsymbol{X})\right]^2$$

satisfies $0 < I(\theta) < \infty$, then

$$Var[T(\boldsymbol{X})] \geq \frac{[\psi'(\theta)]^2}{I(\theta)}.$$

Note: The proof is based on Cauchy-Schwartz Inequality, i.e., for any two r.v.s X and Y,

$$Var(X)\ Var(Y) \geq [Cov(X, Y)]^2.$$

Claim 1:

$$E\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right] = 0$$

Proof:

$$\begin{aligned}
E\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right] &= \int \left[\frac{\partial}{\partial\theta}\log f_\theta(x)\right] f_\theta(x)dx \\
&= \int \frac{1}{f_\theta(x)}\left[\frac{\partial}{\partial\theta} f_\theta(x)\right] f_\theta(x)dx \\
&= \int \frac{\partial}{\partial\theta} f_\theta(x)dx \\
&= \frac{\partial}{\partial\theta}\int f_\theta(x)dx \\
&= \frac{\partial}{\partial\theta} 1 = 0
\end{aligned}$$

Thanks for your patience!