LECTURE - (2)

Agenda:

(1) Statistical Estimators and confidence intervals
(2) Bias and Mean Square Error

STATISTICAL ESTIMATORS

One of the main purposes of statistics is to use the sample/data to estimate some unknown characteristic of the population/phenomenon under consideration.

> An unknown characteristic of interest in the random experiment/phenomenon is known as a PARAMETER.

The only available information we have to estimate the unknown parameter is the sample/data.

> Given a parameter, say θ, a STATISTICAL ESTIMATOR for θ is a function of the sample/data that provides a guess for the true value of θ.

Note that we are in the context of a random experiment, and the observed sample/data is essentially a collection of the observed values of appropriate random variables. Hence, any statistical estimator is essentially a random variable.

Example: Suppose we have a coin which falls heads with probability p (p is unknown).

Random Expt: Toss the coin 1000 times independently.

Data: $X_1, X_2, \ldots, X_{1000}$, where $X_i = \begin{cases} 1 & \text{if } i^{\text{th}} \text{ toss is heads,} \\ 0 & \text{otherwise.} \end{cases}$

Note that $X_1, X_2, \ldots, X_{1000}$ are independent and have a common Bernoulli (p) distribution.

Parameter of interest: p

Most intuitive estimator: $\frac{1}{1000} \sum_{i=1}^{1000} X_i$.

Some people however are more comfortable with an interval estimator for the parameter of interest, i.e.,

they want to construct an interval (based on the data), so that the interval contains the true parameter with a very high probability. Suppose the parameter of interest is θ.
Suppose the interval that we have constructed is $(\hat{\theta}_L, \hat{\theta}_U)$.

Both are functions of the data and hence are random variables.

We typically would like to say that
$P(\hat{\theta}_L \le \theta \le \hat{\theta}_U)$ is large.

$\boxed{\text{Such interval estimators are called CONFIDENCE INTERVALS.}}$

We will come back to confidence intervals. But let us turn to single-valued estimators or point estimators, as they are commonly called.

BIAS AND MEAN SQUARED ERROR

One can very easily imagine that there may be several competing estimators for the parameter of interest.

In the coin tossing example, here are some possible competitoors for the average estimator

$$\hat{p}_{avg} = \frac{1}{1000} \sum_{i=1}^{1000} X_i .$$

$$\hat{p}_1 = \frac{1}{500} \sum_{i=1}^{500} X_i .$$

$$\hat{p}_2 = \frac{1}{500} \sum_{i=501}^{1000} X_i .$$

The question that arises is how do we compare different estimators? In this context, let us define some standard notions in statistics.

The estimator of a parameter θ is generally denoted by $\hat{\theta}$ (sometimes with an appropriate subscript if there are other estimators of θ to consider). RECALL THAT AN ESTIMATOR IS A FUNCTION OF THE DATA, AND HENCE IS A RANDOM VARIABLE.

Let $\hat{\theta}$ be a point estimator for a parameter θ. Then $\hat{\theta}$ is an UNBIASED ESTIMATOR if $E(\hat{\theta}) = \theta$. If $E(\hat{\theta}) \neq \theta$, then $\hat{\theta}$ is said to be BIASED.

The BIAS of a point estimator $\hat{\theta}$ is given by $B(\hat{\theta}) = E(\hat{\theta}) - \theta$.

Note that unbiased is a desired quality in an estimator. ~~[illegible]~~

However, a more important quantity in evaluating the performance of an estimator is the mean square error of an estimator.

The <u>mean square error</u> of a point estimator $\hat{\theta}$ for a parameter θ is defined as

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

The lower the mean squared error, the better the quality of the estimator.

RESULT: $\text{MSE}(\hat{\theta}) = V(\hat{\theta}) + [B(\hat{\theta})]^2$

Proof:

$$\text{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

$$= E[(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2]$$

$$= E[(\hat{\theta} - E(\hat{\theta}))^2 + (E(\hat{\theta}) - \theta)^2 - 2(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta)]$$

$$= E[(\hat{\theta} - E(\hat{\theta}))^2] + (E(\hat{\theta}) - \theta)^2 - 2(E(\hat{\theta}) - \theta)\, E[\hat{\theta} - E(\hat{\theta})]$$

$$= V(\hat{\theta}) + [B(\hat{\theta})]^2 - 0$$

$$= V(\hat{\theta}) + [B(\hat{\theta})]^2.$$

Hence, if an estimator $\hat{\theta}$ is unbiased, then

$$\boxed{\text{MSE}(\hat{\theta}) = V(\hat{\theta}) \text{ as } B(\hat{\theta}) = 0.}$$

For the coin toss example,

$\hat{p}_{avg} = \frac{1}{1000} \sum_{i=1}^{1000} X_i$, where $X_1, X_2, \ldots, X_{1000}$ are independent and have Bernoulli(p) distribution in common.

Hence, $E[\hat{p}_{avg}] = \frac{1}{1000}\sum_{i=1}^{1000} E[X_i] = p$, which means that $\hat{p}_{avg}$ is an unbiased estimator of p.

Also, $MSE(\hat{p}_{avg}) = V(\hat{p}_{avg})$

$$= V\left(\frac{1}{1000}\sum_{i=1}^{1000} X_i\right)$$

$$= \frac{1}{(1000)^2} V\left(\sum_{i=1}^{1000} X_i\right)$$

$$= \frac{1}{(1000)^2}\sum_{i=1}^{1000} V(X_i)$$

$$= \frac{p(1-p)}{1000}.$$

On the other hand, $E[\hat{p}_1] = E[\hat{p}_2] = p$, which means that $\hat{p}_1$ and $\hat{p}_2$ are both unbiased estimators of p, but

$V(\hat{p}_1) = V(\hat{p}_2) = \frac{p(1-p)}{500}$ (Why?), which means that $\hat{p}_{avg}$ is a better estimator of p as compared to $\hat{p}_1$ and $\hat{p}_2$.