# LECTURE - 4

**Agenda:**

Unbiased estimation for some standard problems.

**PROBLEM 1** Suppose $Y_1, Y_2, ---, Y_n$ is a random sample from a population with mean $\mu$ and variance $\sigma^2$ (both unknown).

(a) Find unbiased estimators of $\mu$ and $\sigma^2$.

Again, note that "$Y_1, Y_2, ---, Y_n$ is a random sample from a population" means that $Y_1, Y_2, ---, Y_n$ are independent random variables, and have a common distribution (unspecified here). We know that the mean of this common distribution is $\mu$ and the variance of this common distribution is $\sigma^2$, i.e.,

$$E[Y_i] = \mu \quad \text{and} \quad V(Y_i) = \sigma^2,$$

for every $i = 1, 2, ---, n$. The most obvious estimator for $\mu$ is

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} Y_i = \overline{Y}.$$

Note that

$$E[\hat{\mu}] = E\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[Y_i] = \mu.$$

Hence $\hat{\mu}$ is an unbiased estimator of $\mu$.

**DEFINITION:** The <u>standard error</u> of an estimator $\hat{\theta}$ of a parameter $\theta$ is defined as

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

The standard error of $\hat{\mu}$ is given by

$$SE(\hat{\mu}) = \sqrt{MSE(\hat{\mu})}$$

$$= \sqrt{V(\hat{\mu})} \quad \text{(because } \hat{\mu} \text{ is unbiased)}$$

$$= \sqrt{V\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)}$$

$$= \sqrt{\frac{1}{n^2} V\left(\sum_{i=1}^{n} Y_i\right)}$$

$$= \sqrt{\frac{1}{n^2}\sum_{i=1}^{n} V(Y_i)}$$

$$= \frac{\sigma}{\sqrt{n}}\,.$$

Now let us find an unbiased estimator for $\sigma^2$. Note that $\sigma^2$ is the population variance, i.e.,

$\sigma^2 =$ Average squared deviation from the mean $\mu$ in the population

The only information we have is the sample data $Y_1, Y_2, \ldots, Y_n$. Hence, the most obvious strategy is the use sample quantities in place of population quantities for our procedure. Hence

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 =$$ Average squared deviation from the sample mean $\bar{Y}$ in the sample.

Let us calculate $E[\hat{\sigma}^2]$.

$$
\begin{aligned}
E[\hat{\sigma}^2] &= E\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right] \\
&= E\left[\frac{1}{n}\sum_{i=1}^{n}\left(Y_i^2 - 2Y_i\bar{Y} + \bar{Y}^2\right)\right] \\
&= E\left[\frac{1}{n}\sum_{i=1}^{n}Y_i^2 - 2\bar{Y}^2 + \bar{Y}^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}E[Y_i^2] - E[\bar{Y}^2] \\
&= (\mu^2 + \sigma^2) - E[\bar{Y}^2] \quad (\text{Why??})
\end{aligned}
$$

Let us concentrate on evaluating $E[\bar{Y}^2]$ and then substitute it back into the expression above.

$$E[\bar{Y}^2] = E\left[\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^2\right]$$

$$= \frac{1}{n^2} E\left[\left(\sum_{i=1}^{n} Y_i\right)^2\right]$$

$$= \frac{1}{n^2} E\left[\sum_{i=1}^{n} Y_i^2 + \sum\sum_{1\le i\ne j\le n} Y_i Y_j\right]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} E[Y_i^2] + \frac{1}{n^2}\sum\sum_{1\le i\ne j\le n} E[Y_i Y_j]$$

$$= \frac{(\mu^2+\sigma^2)}{n} + \frac{1}{n^2}\sum\sum_{1\le i\ne j\le n} E[Y_i]\,E[Y_j]$$

(By independence)

$$= \frac{(\mu^2+\sigma^2)}{n} + \frac{1}{n^2}\sum\sum_{1\le i\ne j\le n} \mu^2$$

$$= \frac{(\mu^2+\sigma^2)}{n} + \frac{n(n-1)}{n^2}\mu^2$$

$$= \mu^2 + \frac{\sigma^2}{n}.$$

Hence, $E[\hat{\sigma}^2] = (\mu^2+\sigma^2) - \left(\mu^2 + \frac{\sigma^2}{n}\right) = \frac{(n-1)}{n}\sigma^2$.

This means $\hat{\sigma}^2$ is not unbiased for $\sigma^2$. However, a slight modification does the job.

Let $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2 = \frac{n}{n-1}\hat{\sigma}^2$. Then,

$$E[s^2] = \frac{n}{n-1}E[\hat{\sigma}^2] = \frac{n}{n-1}\times\frac{n-1}{n}\sigma^2 = \sigma^2.$$