POINT ESTIMATORS — FUNDAMENTAL FORMULAE

THE SAMPLE ARITHMETIC AVERAGE (arithmetic mean):

$$\bar{X} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} X_i$$

its expected value is equal to:

$$E\{\bar{X}\} = \frac{1}{n}[E\{X_1\} + E\{X_2\} + \ldots + E\{X_n\}] = \frac{1}{n} \cdot nE\{X\} \equiv \mu$$

and its variance equals

$$\begin{aligned}
\sigma^2(\bar{X}) &= E\{[\bar{X} - E\{\bar{X}\}]^2\} = E\left\{\left[\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu\right]^2\right\} \\
&= \frac{1}{n^2} E\{[(X_1 - \mu) + (X_2 - \mu) + \ldots (X_n - \mu)]^2\} = \ldots \\
& \quad E\{(X_i - \mu)^2\} = \sigma^2(X) \quad i = 1, 2 \ldots \\
& \quad E\{(X_i - \mu)(X_k - \mu)\} = 0 \quad i \neq k \\
\ldots &= \frac{1}{n^2} n\sigma^2(X) = \frac{1}{n}\sigma^2(X)
\end{aligned}$$

Conclusion: the sample mean $\bar{X}$ is a consistent estimator of $E(X)$ (the bigger the sample size the more effective is the estimator).

mean square deviation about the mean

$$S^2 \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\begin{aligned}
E\{S^2\} &= \frac{1}{n}\sum_{i=1}^{n} E\{(X_i - \bar{X})^2\} = \frac{1}{n}\sum_{i=1}^{n} E\{[(X_i - \mu) + (\mu - \bar{X})]^2\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[E\{(X_i - \mu)^2\} + 2E\{(\mu - \bar{X})(X_i - \mu)\} + E\{(\mu - \bar{X})^2\}\right]
\end{aligned}$$

$$\ldots \frac{2}{n}\sum_{i=1}^{n} E\{(X_i - \mu)(\bar{X} - \mu)\} = \frac{1}{n}E\{(\mu - \bar{X})(X_1 - \mu + \ldots + X_n - \mu)\} =$$

$$= 2E\left\{(\mu - \bar{X})\left(\frac{X_1 + \ldots X_n}{n} - \mu\right)\right\} = -2E\left\{(\mu - \bar{X})^2\right\}\ldots$$

$$\begin{aligned}
&= \frac{1}{n}n\sigma^2(X) - E\left\{(\mu - \bar{X})^2\right\} \\
&= \sigma^2(X) - \frac{1}{n}\sigma^2(X) = \frac{n-1}{n}\sigma^2(X)
\end{aligned}$$

THE SAMPLE MEAN SQUARE DEVIATION (ABOUT THE MEAN), cntd.

The bias B is equal to:

$$1 - \frac{n-1}{n} \to 0 \quad \text{for} \quad n \to \infty$$

For this reason we introduce the **unbiased variance estimator**:

$$S^{*2} \stackrel{\text{def}}{=} \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

THE SAMPLE CORRELATION COEFFICIENT – for a 2-D RV

recall:

$$CORR(X,Y) = \frac{E\{(X-\mu_X)(Y-\mu_Y)\}}{\sigma_X \sigma_Y}$$

it can be shown that the appropriate estimator of the correlation of the pair (X,Y) is

$$r = \frac{S_{xy}}{\sqrt{S_{xx}}\,\sqrt{S_{yy}}} \qquad \text{where:}$$

$$S_{xx} = \sum_{i=1}^{n}(X_i - \bar{X})^2 \qquad S_{yy} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

$$\text{and} \;\; S_{xy} = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

.