9 Estimation Theory and Applications:

(Reading Exercises: Montgomery and Runger – Chapter 7 and 8)
(Reading Exercises: Yates and Goodman – Chapter 12)

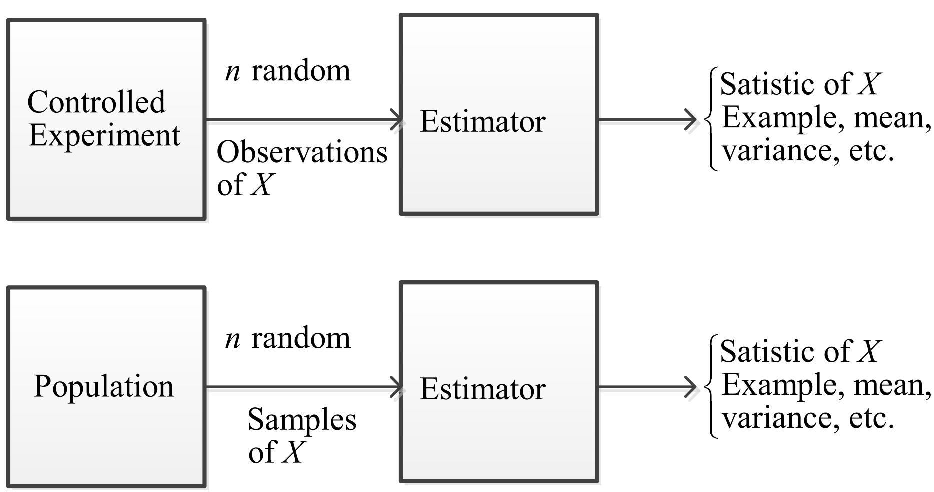
When we acquire samples taken from a population, we may need to estimate certain statistics or parameters of interest that help us make inferences about the population.

Learning Outcomes:

You will

- Know how to find a maximum likelihood estimator of a population parameter.
- Know how to check to see if an estimator is unbiased and/or minimum variance for a particular parameter.
- Understand the difference between point estimation and interval estimation
- Be able to establish a confidence interval for a sample mean
- Understand the steps involved in each of the proofs in the section.
- Be able to apply the methods learned in the section to new problems.

Definition: An estimator is a procedure that provides an estimate of a desired parameter from a set of random data samples drawn from a controlled experiment or a population.



For example, we can observe data samples X from n geostationary satellites and use them to calculate estimates of our location.

9.1 Desirable Properties of Good Estimators:

Bias:

Definition: Let $\hat{\theta}$ denote an estimator of a desired parameter, θ. The bias, denoted as B, of the estimator is defined as the difference between the expected value of the estimate and the true value of value of the parameter,

$$B = E\left[\hat{\theta}\right] - \theta$$

- If $B = 0$ then the estimator is said to be unbiased.
- If $B \neq 0$ the estimator is said to be biased.
- The desirable estimator is the unbiased estimator.

Variance:

Definition: The variance of an estimator of parameter θ, is defined as

$$Var\left(\hat{\theta}\right) = E\left[\left(\hat{\theta} - E\left[\hat{\theta}\right]\right)^2\right]$$

- It is desirable that the variance of the estimator, which is a measure of the spread of values of the estimate around the true parameter, be as small as possible – minimum variance estimator.

Mean Square Estimation Error:

Definition: The Mean Square Error (MSE) of an estimator of parameter θ, is defined as the expected value of the square of the difference between its estimate and its true value,

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

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

$$= E\left[\left(\hat{\theta} - E\left[\hat{\theta}\right]\right)^2\right] - 2BE\left[\hat{\theta} - E\left[\hat{\theta}\right]\right] + B^2 = Var\left(\hat{\theta}\right) + B^2$$

- The smaller the MSE, the more accurate the estimate is, in terms of precision and bias.

Definition: A consistent estimator is one whose variance and bias go to zero as the number of samples approach infinity.

Conclusion:

- The desired estimator is a minimum variance, unbiased estimator (MVUE).

Sample Mean and Sample Variance:

Taking the expectation of the sample mean gives

$$\mu_{\bar{X}} = E\left[\bar{X}\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[X_i\right] = \mu_X$$

Therefore, the sample mean is unbiased.

Consider the two sample variances defined earlier. For the first definition, let us first simplify as follows:

$$\begin{aligned}
S_X^2 &\triangleq \frac{1}{n}\sum_{j=1}^{n}\left(x_j - \bar{X}\right)^2 = \frac{1}{n}\sum_{j}\left(x_j - \mu_X + \mu_X - \bar{X}\right)^2 \\
&= \frac{1}{n}\sum_{j=1}^{n}\left(x_j - \mu_X\right)^2 - \frac{2}{n}\sum_{j=1}^{n}\left(x_j - \mu_X\right)(\bar{X} - \mu_X) + \frac{1}{n}\sum_{i=1}^{n}(\bar{X} - \mu_X)^2 \\
&= \frac{1}{n}\sum_{j=1}^{n}\left(x_j - \mu_X\right)^2 - \frac{2}{n}\sum_{j=1}^{n}\left(x_j - \mu_X\right)\left(\frac{1}{n}\sum_{i}\left(x_i - \mu_X\right)\right) + \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n}\sum_{i}\left(x_i - \mu_X\right)\right)^2 \\
&= \frac{1}{n}\sum_{j=1}^{n}\left(x_j - \mu_X\right)^2 - \frac{2}{n^2}\sum_{j=1}^{n}\left(x_j - \mu_X\right)\left(\sum_{i}\left(x_i - \mu_X\right)\right) + \frac{1}{n^2}\left(\sum_{i}\left(x_i - \mu_X\right)\right)^2
\end{aligned}$$

Taking the expectation, and noting that the samples are independent, gives

$$\begin{aligned}
\sigma_X^2 &= \frac{1}{n}\sum_{j=1}^{n} E\left[\left(x_j - \mu_X\right)^2\right] - \frac{2}{n^2}\sum_{j=1}^{n} E\left[\left(x_j - \mu_X\right)\right]^2 + \frac{1}{n^2}\sum_{i}\left(x_i - \mu_X\right)^2 \\
&= \sigma_X^2 - 2\frac{\sigma_X^2}{n} + \frac{\sigma_X^2}{n} = \sigma_X^2 - \frac{\sigma_X^2}{n} = \frac{n-1}{n}\sigma_X^2 \neq \sigma_X^2
\end{aligned}$$

The sample variance in this case is biased.

Consider now, the second definition. We may write the expression as

$$S_X^2 \triangleq \frac{1}{n-1}\sum_{j=1}^{n}\left(x_j - \overline{X}\right)^2 = \frac{n}{n-1}\left\{\frac{1}{n}\sum_{j=1}^{n}\left(x_j - \overline{X}\right)^2\right\}$$

The expression in the winged brackets is the same as the first definition. Therefore, taking the expected value gives

$$E\left[S_X^2\right] = \frac{n}{n-1}\left\{\frac{1}{n}\sum_{j=1}^{n}E\left[\left(x_j - \overline{X}\right)^2\right]\right\} = \frac{n}{n-1}\left(\frac{n-1}{n}\sigma_X^2\right) = \sigma_X^2$$

The expected value of the sample mean in this case is unbiased.

9.2 Methods for Obtaining Estimators:

There are two types of parameters to be estimated – deterministic and random parameters. Based on the parameter of interest, there are two basic estimators:

- Minimum mean square estimators (MMSE): used when the parameter to be estimated is random.
- Maximum likelihood estimators (MLE): used when the parameter is deterministic.

For example, suppose we want to estimate a parameter θ, from a random sample $X = \theta + v$, where v is Gaussian noise.

- If θ is a deterministic (constant) parameter, then use the MLE.

- If θ is a random variable, then use the MMSE.

In this course, the focus is on deterministic parameter estimation.

9.2.1 Maximum Likelihood Estimator:

Consider a set of collected IID random samples $X = \{x_1, x_2, \ldots, x_n\}$, that contain a desired deterministic parameter θ. The joint conditional pdf $f\left(x_1, x_2, \ldots, x_n \mid \theta\right)$, given θ, is known to as the likelihood function. The MLE estimate is the parameter that maximizes the likelihood function. Since the random variables are IID, the joint conditional PDF (the likelihood function) is the product of the individual PDFs,

$$f\left(x_1, x_2, \ldots, x_n \mid \theta\right) = \prod_{i=1}^{n} f\left(x_i \mid \theta\right)$$

A parameter that maximizes the logarithm of a function will also maximize the function itself.
Taking the natural log of the likelihood function gives

$$L\left(\mathbf{x} \mid \theta\right) = \ln f\left(x_1, x_2, \ldots, x_n \mid \theta\right) = \sum_{i=1}^{n} L\left(x_i \mid \theta\right)$$

The parameter, θ, that maximizes the log likelihood function must satisfy the following condition:

$$\frac{\partial}{\partial \theta} L\left(\mathbf{x} \mid \theta\right) = \frac{\partial}{\partial \theta} \ln f\left(x_1, x_2, \ldots, x_n \mid \theta\right) = 0$$

<u>Estimation of the Mean of X with known variance:</u>

Consider a set of n IID samples, $X = \{x_1,\ x_2,\ \ldots,\ x_n\}$, which are known to be Gaussian. The variance, σ_X^2, of each sample is assumed known but the mean, μ_X, is unknown. The likelihood function is

$$f\left(\mathbf{x} \mid \mu_X\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(-\frac{\left(x_i - \mu_X\right)^2}{2\sigma_X^2}\right) = \left(2\pi\sigma_X^2\right)^{-\frac{n}{2}} \exp\left(-\sum_{i=1}^{n} \frac{\left(x_i - \mu_X\right)^2}{2\sigma_X^2}\right)$$

Taking the natural log gives

$$L\left(\mathbf{x} \mid \mu_X\right) = -\frac{n}{2} \ln\left(2\pi\sigma_X^2\right) - \sum_{i=1}^{n} \frac{\left(x_i - \mu_X\right)^2}{2\sigma_X^2}$$

Taking the derivative with respect to μ_X and equating to zero gives

$$\frac{\partial L\left(\mathbf{x} \mid \mu_X\right)}{\partial \mu_X} = 2\sum_{i=1}^{n} \frac{x_i - \mu_X}{2\sigma_X^2} = 0 \Rightarrow \sum_{i=1}^{n} x_i - n\mu_X = 0$$

Therefore, the maximum likelihood estimate of the mean is the sample mean,

$$\hat{\mu}_X = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

The ML estimate of the mean is unbiased since $E\left[\bar{X}\right]=\mu_X$. It is also minimum variance for large sample size, since $\sigma_{\bar{X}}^2=\sigma_X^2/n\to 0$ with increasing n.

Estimation of the variance of X with known Mean:

Consider a set of n IID samples, $X=\{x_1,\ x_2,\ \ldots,\ x_n\}$, which are known to be Gaussian. The mean, μ_X, of each sample is assumed known but the variance, σ_X^2, is unknown. Find the MLE for the variance σ_X^2, of n IID samples, which are known to be Gaussian. The mean of each sample μ_X, is assumed known. The likelihood function is

$$\ell\left(\mathbf{x}\,|\,\sigma_X^2\right)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma_X^2}}\exp\left(-\frac{\left(x_i-\mu_X\right)^2}{2\sigma_X^2}\right)=\left(2\pi\sigma_X^2\right)^{-\frac{n}{2}}\exp\left(-\sum_{i=1}^{n}\frac{\left(x_i-\mu_X\right)^2}{2\sigma_X^2}\right)$$

Taking the derivative with respect to σ_X^2 and equating it to zero gives

$$\frac{\partial L\left(\mathbf{x}\,|\,\sigma_X^2\right)}{\partial\sigma_X^2}=\sum_{i=1}^{n}\frac{\left(x_i-\mu_X\right)^2}{2\left(\sigma_X^2\right)^2}-\frac{n}{2}\frac{1}{2\sigma_X^2}2\pi=0$$

Solving the above equation, the estimate of the variance is the sample variance

$$\hat{\sigma}_X^2=S_X^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\mu_X\right)^2$$

The ML estimate of the variance is biased for small sample size since $E\left[S_X^2\right]=\sigma_X^2-\sigma_X^2/n$. However, for large sample size, it is unbiased since $E\left[S_X^2\right]=\sigma_X^2-\sigma_X^2/n\to\sigma_X^2$ as $n\to\infty$.

Estimation of the Mean and Variance of X:

Consider a set of n IID samples, $X=\{x_1,x_2,\ldots,x_n\}$, which are known to be Gaussian. The variance, σ_X^2, and the mean, μ_X, are unknown. Equating to zero, the partial derivatives of the log likelihood function with respect to the mean and variance, respectively, and then solving the two equations simultaneously for the estimates gives

$$\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \hat{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \hat{\mu}_X\right)^2$$

9.3 Interval Estimation and Confidence Interval:

Instead of seeking a single value that we designate to be the estimate of the parameter of interest, we may want to specify an interval that is highly likely to contain the true value of the parameter. For example, we can assume that the desired parameter lies inside an interval $l(\mathbf{x}) \le \theta \le u(\mathbf{x})$. The objective then is to find the interval limits, $l(\mathbf{x})$ and $u(\mathbf{x})$, such that the interval contains θ with some degree of confidence or acceptable probability, say $1-\alpha$. That is,

$$P\left[l(\mathbf{x}) \le \theta \le u(\mathbf{x})\right] = 1-\alpha$$

<u>Definition:</u> Let α denote the probability that the parameter θ is not in the interval $\left[l(\mathbf{x}),\ u(\mathbf{x})\right]$. Then,

- $1-\alpha$ is called the degree of confidence, which may be also measured in percentage
- The width of $\left[l(\mathbf{x}),\ u(\mathbf{x})\right]$ is called the confidence interval and is a measure of the accuracy of the estimator. A narrow interval corresponds to a higher accuracy of the estimated parameter.

9.3.1 Interval Estimation of the Mean of X with Known Variance:

Consider a set of IID random samples, $X = \{x_1,\ x_2,\ \ldots,\ x_n\}$ obtained from a Gaussian population where the mean is unknown but the true variance σ_X^2, is known. Denote the true mean as μ_X. Let us use the sample mean $\bar{X}$, as an estimate of the true mean. We know that

$$E\left[\bar{X}\right] = \mu_X \text{ and } Var\left(\bar{X}\right) = \frac{\sigma_X^2}{n}$$

Since we have to use the Q-function table, we define the normalized sample estimate

$$Z = \frac{\bar{X} - \mu_X}{\sigma_X/\sqrt{n}}$$

The normalized sample estimate Z, is a standard Gaussian random variable (in accordance with the central limit theorem). The distribution being symmetric around zero, we can write the estimation interval for Z as $z_c \le Z \le z_c$. Given α, we can determine z_c by evaluating the probability that the normalized sample mean is inside this interval. That is,

$$P\left[-z_c \le Z \le z_c\right] = \int_{-z_c}^{z_c} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = 1 - 2Q\left(z_c\right) = 1 - \alpha \Rightarrow Q\left(z_c\right) = \alpha/2$$

We use the Q-function table to find $z_c = Q^{-1}\left(\alpha/2\right)$. For example, for $\alpha = 0.05$, $Q\left(z_c\right) = 0.025$ and the table gives $z_c \simeq 1.96$.

Confidence interval for the true mean:

In order to find the confidence interval for the true mean, we make the following substitution:

$$-z_c \le Z = \frac{\bar{X} - \mu_X}{\sigma_X/\sqrt{n}} \le z_c$$

Upon simplification, we obtain the confidence interval for the true mean as

$$\boxed{\bar{X} - \frac{\sigma_X}{\sqrt{n}} z_c \le \mu_X \le \bar{X} + \frac{\sigma_X}{\sqrt{n}} z_c}$$

This interval is valid even if the set of random samples are not Gaussian as long as the sample size n is large enough to invoke the central limit theorem.

Example: (a) Find the 99% confidence interval for the batch of data in the table below. (b) How will the 99% confidence limits change if the number of samples is increased to 100? The true variance is 4.

X_i	7.31	10.8	11.27	11.91	5.51	8.0	9.03	14.42	10.24	10.91

Solution: The sample mean is $\bar{X} = 9.94$ for 10 samples. For a confidence level of 99%, $\alpha = 0.01$ and $Q\left(z_c\right) = 0.5\alpha = 0.005$. From the Q-function table, $z_c \simeq 2.6$. The confidence interval for the true mean is

$$9.94 - \frac{\sqrt{4}}{\sqrt{10}} \times 2.6 \le \mu_X \le 9.94 + \frac{\sqrt{4}}{\sqrt{10}} \times 2.6 \Rightarrow 8.3 \le \mu_X \le 11.58$$

If we increase the number of samples to 100, we obtain

$$9.94-\frac{\sqrt{4}}{\sqrt{100}}\times 2.6\le \mu_X \le 9.94+\frac{\sqrt{4}}{\sqrt{100}}\times 2.6 \Rightarrow 9.12\le \mu_X \le 9.42$$

We observe that the confidence interval is tighter, and the accuracy is, therefore, increased with sample size.