Point Estimation (Chapter 6)

1. General Concepts of Point Estimation

   (a) Point Estimates

   - Objective: Statistical inference is often concerned with drawing conclusions about one or more population parameters based on data drawn from each population under study. Such conclusions are based on computed values of various sample quantities.
   - In this course we will consider three aspects of statistical inference: Point Estimation (Cha.6), Interval Estimation (Cha.7), and Hypothesis Testing (Cha.8).
   - **Point Estimate:** A point estimate of parameter θ is a single number, a sensible value for θ, obtained by computing the value of a suitable sample statistic.
   - This suitable sample statistic is called the **point estimator** of θ and is denoted by $\hat{\theta}$.

   (b) Unbiased Estimators

   - Point estimator, $\hat{\theta}$, is said to be an unbiased estimator of θ if

     $$E(\hat{\theta}) = \theta$$

     for every possible value of θ.
   - If $\hat{\theta}$ is not an unbiased estimator, then

     $$\left(Bias\ of\ \hat{\theta}\right) = E(\hat{\theta}) - \theta$$

   - When choosing among different estimators of θ, select the one that is unbiased. This is the Principle of Unbiased Estimation.

   (c) Estimators for Population Mean, Variance, and Proportion

   - If $X_1, X_2, ..., X_n$ is a random sample from a distribution with mean, μ, and variance, σ^2, then the sample mean, $\bar{x}$, is an unbiased estimator of μ, where

     $$\bar{x} = \frac{\sum x_i}{n}$$

   - If the distribution is continuous and symmetric, then $\tilde{x}$, the sample median, and any trimmed mean (mean computed with both the smallest and largest fraction, say 10%, of data values discarded) are also unbiased estimators of μ.
   - Variance: Also, the sample variance, S^2, is an unbiased estimator of σ^2, where

     $$S^2 = \hat{\sigma}^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$

   - Proportion: When X is a binomial rv with parameters n and p, the sample proportion, $\hat{p}$, is an unbiased estimator of p.

     $$\hat{p} = \frac{X}{n}$$

(d) Minimum Variance Estimators

- From among all estimators of θ that are unbiased, choose the one with minimum variance. The resulting $\hat{\theta}$ is called the **minimum variance unbiased estimator (MVUE)** of θ.
- If $X_1, X_2, ..., X_n$ is a random sample from a normal distribution with parameters, μ and σ^2, then the sample mean, $\bar{x} = \hat{\mu}$, is the MVUE for μ.

(e) Standard Error

- The **Standard Error** of estimator $\hat{\theta}$ is its standard deviation.

$$\sigma_{\hat{\theta}} = \sqrt{Var(\theta)}$$

- If the standard error involves unknown parameters whose values can be estimated, substitution into $\sigma_{\hat{\theta}}$ yields the estimated standard error of the estimator. This is denoted $\hat{\sigma}_{\hat{\theta}}$ or $S_{\hat{\theta}}$.
- **Reporting a Point Estimate:** Best to report both the value itself,$\hat{\theta}$, and some measure of the precision of the estimate.
- The standard error, $\sigma_{\hat{\theta}}$, is a common measure of the precision of the estimate.

(f) Point Estimation Methods

- Methods of point estimation provide, among other things, a way to derive the unbiased estimators for population (distribution) parameters presented above.
- Two common methods of point estimation are the **method of moments** and **maximum liklihood estimation**.
- These two methods are discussed in more detail below.

2. Method of Maximum Liklihood

(a) Maximum Liklihood Estimator (MLE)

- Suppose that X is an rv with pdf (or pmf), $f(x, \theta)$, where θ is a single unknown parameter. Let $X_1, X_2, ..., X_n$ be observed values of a random sample of size n.
- The **Liklihood Function** of the sample is then

$$L(\theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdot ... \cdot f(x_n, \theta)$$

- As $X_1, X_2, ..., X_n$ are iid, the sample liklihood is also it's joint probability distribution. That is,

$$L(\theta) = P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) = f(x_1, \theta) \cdot f(x_2, \theta) \cdot ... \cdot f(x_n, \theta)$$

- The maximum liklihood estimate of θ, mleθ, is the value of θ that maximizes the liklihood function, $L(\theta)$. This value of θ also maximizes the probability of occurance of the observed sample.

(b) Estimating Functions of Parameters

- For any function, $h(\theta)$, the mle of $h(\theta)$ is just $h(mle\theta)$. This property also holds when h is a function of many θ's.

(c) Large Sample Behavior of MLE

- Optimal properties: When the sample size is large, the maximum liklihood estimator (mle) of any parameter, θ, is approximately unbiased, that is

$$E(\hat{\theta}) \approx \theta$$

- The maximum liklihood estimator (mle) of θ, has a variance that is nearly as small as can be achieved by any estimator.

$$mle(\hat{\theta}) \approx MVUE\theta$$

(d) More than Two Random Variables

- Let $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_n$ be the MLE's of parameters $\theta_1, \theta_2, ..., \theta_n$, then the MLE of any function: $h(\theta_1, \theta_2, ..., \theta_n)$ of these parameters is:

$$h(\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_n)$$

- This is called the Invariance Principle

(e) Example: Consider a random sample $x_1, x_2, ..., x_n$ from a normal distribution $N(\mu, \sigma)$. Find the maximum liklihood estimators for μ and σ^2.

- Liklihood function for the normal distribution is

$$L(x_1, x_2, ..., x_n, \mu, \sigma^2) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2\right]$$

- Taking natural logs gives us

$$lnL(x_1, x_2, ..., x_n, \mu, \sigma^2) = \frac{n}{2}ln2\pi - \frac{n}{2}ln\sigma^2 - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2$$

$$\frac{\partial lnL}{\partial \mu} = \sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)$$

- The minimum of $L(x_1, x_2, ..., x_n, \mu, \sigma^2)$ occurs when the minimum of $lnL(x_1, x_2, ..., x_n, \mu,$ occurs, which takes place when its dirivative is set equal to zero:

$$\sum_{i=1}^{n} x_i - n\mu = 0$$

- Solving for μ gives the maximum liklihood estimate of μ, here denoted as $\hat{\mu}$.

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

- So $\bar{x}$, the sample mean, is shown to be the MLE of normal population mean, μ.
- The MLE estimator of normal population variance, σ^2, can be determined in a similar fashion with the following result

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}$$

- Note that the MLE estimate of σ^2 shown above is not the unbiased estimator, S^2 discussed earlier, since

$$S^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$

- Maximum likelihood estimators are not necessarily unbiased, but do enjoy The other good properties mentioned earlier.

3. Method of Moments

(a) Population (Distribution) Moments

- Let $X_1, X_2, ..., X_n$ be a random sample from pdf (or pmf) $f(x)$. For $k = 1, 2, ...$ the kth population moment, or kth moment of distribution $f(x)$ is $E(X^k)$.

(b) Sample Moments

- The kth sample moment is,

$$\left(\frac{1}{n}\right) \sum_{i=1}^{n} X_i^k$$

(c) Moment Estimators

- Moment Estimators, $\theta_1, ..., \theta_m$ are obtained by equating the first m sample moments to the corresponding first m population moments and solving for $\theta_1, ..., \theta_m$.

(d) Example: Let $X \sim N(\mu, \sigma^2)$, where μ and σ are unknown. Derive estimators for μ and σ using the method of moments.

- Recall that the first two moments for a normal distribution are:

$$E(X) = \mu; \qquad E(X^2) = \sigma^2 + \mu^2$$

- The first two sample moments are

$$m1 = \frac{1}{n} \sum_{i=1}^{n} x_i; \qquad m2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2$$

- Equating sample and distribution moments we get

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\sigma^2 + \mu^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2$$

- Solving these two equations simultaneously gives

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$$

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

- These results are identical to those determined earlier by the maximum likelihood method.