Point Estimation

Population: In Statistics, population is an aggregate of objects, animate or inanimate, under study. The population may be finite or infinite.

Sample: A part or a finite subset of population is called a sample and the number of units in the sample is called the sample size.

Parameter: The statistical constants of the population such as population mean (μ), population variance (σ^2) are referred as parameters.

Statistic: The statistical measures computed from the sample observations alone for example, sample mean ($\bar{x}$), sample variance (s^2) are known as statistics.

Here in the theory of point estimation, we consider that the population under study is described by a probability density function (pdf) or probability mass function (pmf), say, $f(x|\theta)$. The knowledge of parameter θ yields the knowledge of entire population but the problem of statistical parametric inference is that θ is unknown. In order to estimate this θ, we resort to take a random sample form the population and infer about the unknown parameter θ. It may also happen that instead of θ, our interest is to find an estimator for a function of θ, say, $g(\theta)$.

Estimator: Any function of the random sample which is used to estimate the unknown value of the given parametric function say $g(\theta)$ is called an estimator. If $\underline{X} = X_1, \ldots, X_n$ is a random sample from a population with the probability distribution P_θ, a function $d(\underline{X})$ used for estimating $g(\theta)$ is known as an estimator. Let $\underline{x} = x_1, \cdots, x_n$ be a realization of $\underline{X}$. Then, $d(\underline{x})$ is called an estimate.
For example, in estimating the average height of male students in a class, we may use the sample mean $\bar{X}$ as an estimator. Now, if a random sample of size 20 has a sample mean 170cm, then 170cm is an estimate of the average height of male students of that class.

Parameter Space: The set of all possible values of a parameter(s) is called parameter space. It is denoted by Θ.

Desirable Criteria for Estimators

Unbiasedness: Let $X_1, \cdots, X_n$ be a random sample from a population with probability distribution $P_\theta, \theta \in \Theta$. An estimator $T(\underline{X}), \underline{X} = X_1, \cdots, X_n$ is said to be unbiased for estimating $g(\theta)$, if

$$E_\theta(T(\underline{X})) = g(\theta), \forall \theta \in \Theta. \tag{1}$$

If for some $\theta \in \Theta$, we have

$$E_\theta(T(\underline{X})) = g(\theta) + b(\theta),$$

then, $b(\theta)$ is called bias of T. If $b(\theta) > 0, \forall \theta$, then T is said to overestimate $g(\theta)$. On the other hand if $b(\theta) < 0, \forall \theta$, then T is an underestimator of $g(\theta)$.

Example 1: Let $X_1, \cdots, X_n$ be a random sample from binomial distribution with parameters n and p, where, n is known and $0 \leq p \leq 1$. Find unbiased estimators for a) p, the population proportion, b) p^2 c) Variance of X.

Solution: a) Given that X follows $binomial(n, p)$, n is known and p, the population proportion is unknown. Let $T(\underline{X}) = \frac{X}{n}$, the sample proportion. Now,

$$E(T(\underline{X})) = E\left(\frac{X}{n}\right) = \frac{np}{n} = p.$$

Thus, the sample proportion is an unbiased estimator of population proportion.
b) We can compute that

$$E(X(X-1)) = n(n-1)p^2 \tag{2}$$

Hence, $\frac{X(X-1)}{n(n-1)}$ is an unbiased estimator for p^2.
c)Since, $\text{Var}(X) = np(1-p) = n(p - p^2)$.
Therefore, $T(\underline{X}) = n\left(\frac{X}{n} - \frac{X(X-1)}{n(n-1)}\right) = \frac{X(n-X)}{n-1}$ is an unbiased estimator of Variance of X.

Remarks:

1. The unbiased estimator need not be unique. For example, let $X_1, \cdots, X_n$ be a random sample form Poisson distribution with parameter $\lambda, \lambda > 0$. Then, $T_1(\underline{X}) = \bar{X}$, $T_2(\underline{X} = X_i)$, $T_3(\underline{X}) = \frac{X_1+2X_2}{3}$ are some unbiased estimators for λ.

2. If $E(X)$ exists, then the sample mean is an unbiased estimator of the population mean.

3. Let $E(X^2)$ exists, i.e. $Variance(X) = \sigma^2$ exists. Then, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is unbiased for σ^2.(Prove!)

4. Unbiased estimators may not always exist.For example, X follows binomial distribution with parameters n and p. Then, there exists no unbiased estimator for p^{n+1}.(Prove!)

5. Unbiased estimators may not be reasonable always. They may be absurd. For example $T(\underline{X}) = (-2)^X$ is an absurd unbiased estimator for $e^{-3\lambda}$, where, X follows Poisson distribution with parameter λ. (Why?)

Consistency: An estimator $T_n = T(X_1, \cdots, X_n)$ is said to be consistent for estimating $g(\theta)$ if for each $\epsilon > 0$, $P(|T_n - g(\theta)| > \epsilon) \to 0$ as $n \to \infty$, $\forall \theta \in \Theta$.
Example 1: Let $X_1, \cdots, X_n$ be a random sample from a population with mean μ and variance σ^2. Then,

$$P(|\bar{X} - \mu| > \epsilon) \leq \frac{Variance(\bar{X})}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0 \ \ as \ \ n \to \infty.$$

Hence, $\bar{X}$ is consistent for μ.

Example 2: Let $X_1, \cdots, X_n$ be a sequence of i.i.d. random variables with mean μ, then by weak law of large numbers (WLLN), $\bar{X}$ is consistent for μ.

Note: If population mean exists, sample mean is consistent for the population mean.

Theorem 1: If T_n is consistent for θ and h is a continuous function, then $h(T_n)$ is consistent for $h(\theta)$.

Theorem 2: If $E(T_n) = \theta_n \to \theta$, $V(T_n) = \sigma_n^2 \to 0$ as $n \to \infty$, then T_n is consistent for θ.

Theorem 3: If T_n is consistent for θ and $a_n \to 1$, $b_n \to 0$, then $a_n T_n + b_n$ is consistent for θ.

Note: The consistent estimator may not be unique. For example, if T_n is consistent for θ, then, $\frac{n}{n+1}T_n$, $\frac{n+2}{n+4}T_n$ are all consistent for θ.

Methods of Finding Estimators

There are various methods of finding estimators for the parameters, some of which are listed below.

- Method of Moments
- Method of Least Squares
- Method of Minimum Chi square
- Method of Maximum Likelihood Estimation

We will discuss the method of moments and method of maximum likelihood estimation in detail.

Method of Moments: Let $X_1, \cdots, X_n$ be a random sample from a population with probability distribution $P_{\underline{\theta}}; \theta \in \Theta; \underline{\theta} = (\theta_1, \cdots, \theta_k)$.
Consider first k non central moments,

$$\mu_1^{'} = E(X_1) = g_1(\underline{\theta})$$

$$\mu_2^{'} = E(X_1^2) = g_2(\underline{\theta})$$

$$\vdots$$

$$\mu_k^{'} = E(X_1^k) = g_k(\underline{\theta}).$$

Assume that the above system of equations have solution as

$$\theta_1 = h_1(\mu_1^{'}, \cdots, \mu_k^{'})$$

$$\theta_2 = h_2(\mu_1^{'}, \cdots, \mu_k^{'})$$

$$\vdots$$

$$\theta_k = h_k(\mu_1^{'}, \cdots, \mu_k^{'}).$$

Now, define the first k non central sample moments as

$$\alpha_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\alpha_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

$$\vdots$$

$$\alpha_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k.$$

In the method of moments, we estimate k^{th} population moment by k^{th} sample moment, i.e.,

$$\hat{\mu}_{j'} = \alpha_j \ ; \quad j = 1, \cdots, k.$$

Thus, the method of moments estimators of $\theta_1, \cdots, \theta_k$ are defined as

$$\hat{\theta}_1 = h_1(\alpha_1, \cdots, \alpha_k)$$

$$\hat{\theta}_2 = h_2(\alpha_1, \cdots, \alpha_k)$$

$$\vdots$$

$$\hat{\theta}_k = h_k(\alpha_1, \cdots, \alpha_k).$$

Example 1: Let $X_1, \cdots, X_n$ follow $N(\mu, \sigma^2)$; μ and σ^2 are unknown. Find the method of moments estimators μ and σ^2 .

Solution: We know, for normal distribution, $\mu_1^{'} = \mu$ and ${\mu_2}' = \mu^2 + \sigma^2$. Therefore, we have

$$\mu = \mu_1^{'}$$

and

$$\sigma^2 = \mu_2^{'} - \mu_1^{'\,2}.$$

Now, equating the population moments to sample moments, we get

$$\hat{\mu}_{MME} = \bar{X},$$

and

$$\hat{\sigma}^2_{MME} = \alpha_2 - \alpha_1^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2.$$

Exercise 1: Let $X_1, \cdots, X_k$ follow binomial distribution with parameters n and p. Then, find the moment estimators of (i) p, when n in known and (ii) n and p both, when n is unknown.
Exercise 2: Let $X_1, \cdots, X_n$ be a random sample from Poisson distribution with parameter λ, Find the moment estimator of λ.

Remarks:

1. The method moment estimators need not be unbiased always.
2. If the functions $g_i's$ are continuous and one-one then the functions $h_i's$ are also continuous and then the method of moment estimators will be consistent.

Method of Maximum Likelihood Estimation: Let $X_1, \cdots, X_n$ be a random sample from the probability distribution $f(x_i, \underline{\theta})$. Let $\underline{x} = (x_1, \cdots, x_n)$ is a realization of the random sample, then, the likelihood function is given by

$$L(\underline{\theta}, \underline{x}) = \prod_{i=1}^{n} f(x_i, \underline{\theta}).$$

The value of θ, say $\hat{\theta}(\underline{x})$ such that

$$L(\hat{\underline{\theta}}, \underline{x}) \geq L(\underline{\theta}, \underline{x}) \quad \forall\ \underline{\theta} \in \Theta,$$

is called the maximum likelihood estimator (MLE) of $\underline{\theta}$.

In practice, we may often consider maximization of log likelihood, i.e., $\log L(\underline{\theta}, \underline{x}) = l(\underline{\theta}, \underline{x})$ as log is an increasing function of $\underline{\theta}$. A useful approach often is applicable is to find solutions of the likelihood equations ($\frac{\partial l}{\partial \theta_1} = 0, \cdots, \frac{\partial l}{\partial \theta_k} = 0$).

Example: Let $X_1, \cdots, X_n$ follow Poisson distribution with parameter λ; $\lambda > 0$. Find the MLE for λ.
Solution: Let $\underline{x} = (x_1, \cdots, x_n)$ be a realization of a random sample. Then the likelihood function is given by

$$L(\lambda, \underline{x}) = \prod_{i=1}^{n} f(x_i, \lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}.$$

Therefore, the log likelihood function is given by

$$\log L(\lambda, \underline{x}) = l(\lambda) = -n\lambda + \sum_{i=1}^{n} x_i \log \lambda - \log \left(\prod_{i=1}^{n} x_i! \right).$$

The likelihood equation is

$$\frac{\partial l}{\partial \lambda} = -n + \frac{1}{\lambda} \sum_{i=1}^{n} x_i = 0.$$

Now, $\frac{\sum_{i=1}^{n} x_i - n\lambda}{\lambda} > 0$ if $\lambda < \bar{x}$ and $\frac{\sum_{i=1}^{n} x_i - n\lambda}{\lambda} < 0$ if $\lambda > \bar{x}$
Hence, the MLE for λ is $\hat{\lambda} = \bar{x}$.