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 specific characteristics of the population such as population mean (μ), population variance (σ^2) are referred as parameters.
Statistic: It is a function of sample observations, 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, $f(x|\underline{\theta})$. The knowledge of parameter(s) $\underline{\theta}$ yields the knowledge of entire population but the problem of statistical parametric inference is that $\underline{\theta}$ is unknown. In order to estimate this $\underline{\theta}$, we resort to take a random sample from the population and infer about the unknown parameter(s) $\underline{\theta}$. It may also happen that instead of $\underline{\theta}$, our interest is to find an estimator for a function of $\underline{\theta}$, say, $g(\underline{\theta})$.

Definition 1. ***Estimator:*** *Any function of the random sample which is used to estimate the unknown value of the given parametric function $g(\underline{\theta})$ is called an estimator. If $\underline{X} = X_1, \ldots, X_n$ is a random sample from a population with common distribution function $F_{\underline{\theta}}$, a function $t(\underline{X})$ used for estimating $g(\underline{\theta})$ is known as an estimator. Let $\underline{x} = x_1, \cdots, x_n$ be a realization of $\underline{X}$. Then, $t(\underline{x})$ 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

Given the sample, one may have multiple estimators to estimate the parametric function. For example, to estimate the population average, one may use sample mean/sample median/sample mode. So, in order to choose among the estimators, we should have certain desirable criteria which the estimator to be used should meet. Two such criteria unbiasedness and consistency are discussed as follows.

Definition 2. ***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

$$E_\theta(t(\underline{X})) = g(\theta) + b_n(\theta),$$

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

Definition 3. *An estimator $t(\underline{X})$ is said to be asymptotically unbiased estimator of θ if*

$$\lim_{n\to\infty} b_n(\theta) = 0, \quad \forall\theta \in \Theta. \tag{2}$$

Definition 4. *The quantity $E_\theta(t(\underline{X})-\theta)^2$ is called the mean square error (MSE) of $t(\underline{X})$ about θ.*

$$MSE(t(\underline{X})) = Var(t(\underline{X})) + (b_n(\theta))^2 .$$

If t is unbiased for θ, $MSE(t)$ reduces to $Var(t)$.

Example 5. *Let $X_1, \cdots, X_n$ be a random sample from binomial distribution with parameters n and p, where, n is known and $0 \le p \le 1$. 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

$$E(X(X-1)) = n(n-1)p^2 \tag{3}$$

Hence, $\frac{X(X-1)}{n(n-1)}$ is an unbiased estimator for p^2.
c)Since, 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.

Example 6. *Let $X_1, \cdots, X_n$ be a random sample from the population*

$$f(x,\theta) = \begin{cases} e^{-(x-\theta)}, & x > \theta \\ 0, & otherwise \end{cases}$$

.

Is $\bar{X}$ unbiased for θ?

Solution: Note that

$$E(X) = \int_0^\infty x e^{-(x-\theta)} = \theta + 1,$$

so that $E(\bar{X}) = E(X) = \theta + 1$. Thus, $\bar{X}$ is a biased estimator for θ. However, $E(\bar{X}-1) = \theta$.

Remark 7.

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. $Var(X) = \sigma^2$ 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?)*

It is intuitively clear that for $t_n(\underline{X})(= t(\underline{X}))$ to be a good estimator the difference $t_n - \theta$ should be as small as possible. However, t_n is a random variable and has its own sampling distribution whose range may be infinitely large. Therefore, it would be sufficient if the sampling distribution of t_n becomes more and more concentrated around θ as the sample size n increases. This means that for each fixed $\theta \in \Theta$, the probability

$$P_\theta[|T_n - \theta| \leq \epsilon]$$

for any given $\epsilon(> 0)$ should be an increasing function of n. This idea leads to the concept of consistency as a criterion of a good estimator.

Definition 8. *Consistency: A statistic t or rather a sequence $\{t_n\}$ is said to be consistent for θ if t_n converges in probability to θ ($t_n \to \theta$) as $n \to \infty$ for each fixed $\theta \in \Theta$. Thus, t_n is said to be consistent if for every fixed $\theta \in \Theta$ and every pair of positive quantities ϵ and η, however, small, it is possible to find an n_0, depending on ϵ and η, such that*

$$P_\theta[|t_n - \theta| < \epsilon] > 1 - \eta,$$

whenever $n \geq n_0(\epsilon, \eta)$.

If such statistics are used, the accuracy of the estimate increases with the increase in the value of n. It is to be noted that consistency is a large sample property as it is concerned with the behavior of an estimator as the sample size becomes infinitely large.

Example 9. *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 10. *Let $X_1, \cdots, X_n$ be a sequence of independently and identically distributed (iid) random variables with mean μ, then by weak law of large numbers (WLLN), $\bar{X}$ is consistent for μ.*

Example 11. *Let $\{X_n\}$ be a sequence of iid random variables with pdf*

$$f(x, \theta) = \begin{cases} e^{-(x-\theta)}, & x > \theta \\ 0, & otherwise \end{cases}$$

Show that $X_{(1)} = \min X_i$ is a consistent estimator of θ.

Solution: The pdf of $X_{(1)}$ is

$$g(x_{(1)}) = ne^{-(x_{(1)}-\theta)} \left(\int_{x_{(1)}}^{\infty} e^{-(x-\theta)dx} \right)^{n-1} = ne^{-n(x_{(1)})-\theta},$$

for $x_{(1)} > \theta$ and $g(x_{(1)}) = 0$ otherwise.
Now, $P(|X_{(1)} - \theta| < \epsilon) = P(\theta < X_{(1)} < \theta + \epsilon) = 1 - e^{-n\epsilon} \to 1$ as $n \to \infty$.
Hence, $X_{(1)}$ is consistent for θ.

Remark 12. *1. If population mean exists, sample mean is consistent for the population mean.*

2. 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.

Theorem 13. *Let $\{t_n\}$ be a sequence of estimates such that for every $\theta \in \Theta$, the expectation and variance of t_n exist and $E(t_n) = \theta_n \to \theta$ and $V(t_n) \to 0$ as $n \to \infty$. Then, t_n is consistent for θ.*

Proof. We have, by Chebyshev's inequality

$$P[|t_n - \theta| > \epsilon] < \frac{E(t_n - \theta)^2}{\epsilon^2} = \frac{V(t_n) + (E(t_n) - \theta)^2}{\epsilon^2} \to 0 \quad as \quad n \to \infty.$$

□

Theorem 14. *If t is consistent for θ and h is a continuous function of θ. Then, $h(t)$ is consistent for $h(\theta)$.*

Exercise 15. *1. Let $X_1, \ldots, X_n$ be random sample from uniform distribution*

$$f(x) = \begin{cases} \frac{1}{\theta}, & 0 < x < \theta \\ 0, & otherwise \end{cases}.$$

2. Let $X_1, \ldots, X_n$ be a random sample from the uniform disitribution

$$f(x) = \begin{cases} \frac{1}{\theta}, & 0 \leq x \leq \theta \\ 0, & otherwise \end{cases}.$$

Examine the consistencies of the estimators $T_1 = \max X_i$, $T_2 = (n+1) \min X_i$, $T_3 = \min X_i + \max X_i$, $T_4 = 2\bar{X}$ for estimating θ.