1.1 INTRODUCTION.

In this chapter we shall give the basic concepts which are needed for the studies involved in this dissertation. The topics in this chapter are described often briefly. This chapter is completely based on the book by Randles, R.H. and Wolfe, D.A. (1979), and for the completeness we have reproduced some of the proofs of the results as such.

Order statistics being one of the important subject matter in non-parametric study, has been discussed in Sec.1.2. Inverse of a distribution function and the probability integral transformation have been given in Sec.1.3. Sec.1.4 describes different methods of construction of Distribution-free statistics. This includes Counting statistics, Ranking statistics and the combination of these two. Another important class of statistics namely, U-Statistics have been discussed in Sec.1.5, which are used by many authors to get Distribution-free methods. It's distributional results both for One-sample case and Two-sample case have been discussed in length. The K-sample case has been described in Sec.1.6. Sec.1.7. describes the Power functions and their properties. The last section studies the Pitmann-Asymptotic-Relative Efficiency, which we will recall every
now and then for efficiency comparisons.

1.2. ORDER STATISTICS.

The fundamental principle of the development of non-parametric statistics both for hypothesis tests and for estimators is based on the ordered values of the sample. These ordered sample observations are referred to as order statistics. In this section we develop some of the basic properties of order statistics that are used throughout this dissertation.

Let the continuous random variables $x_1, x_2, x_3, ..., x_n$ denote a random sample from a population with Cumulative Density Function $F(x)$ and the density $f(x)$. Let $x_{(i)}$, $i = 1, 2, 3, ..., n$ be the $i^{th}$ smallest of these sample observations. Then $x_1 \leq x_2 \leq ... \leq x_n$ is known as the order statistics for the random sample $x'$. Unlike the $x'$s themselves, the order statistics are neither mutually independent nor identically distributed.

**Theorem 1.1.** Let $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$ be the order statistics for a random sample of continuous random variables from a distribution with c.d.f. $F(x)$ and density $f(x)$. The joint density for the order statistics is then

$$g(x_{(1)}, x_{(2)}, ..., x_{(n)}) = n! \prod_{i=1}^{n} f(x_{(i)}) , \quad -\infty \leq x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)} \leq \infty$$

$$= 0, \quad \text{elsewhere.} \quad (1.2.1)$$
Proof:

The theorem can be proved for a particular value of \( n \), say \( n = 2 \).

Define \( A \) as \( \{ (x_1, x_2, \ldots, x_n) / -\infty < x_i < \infty, x_i \neq x_j \text{ for } i \neq j \} \)
and \( B \) as \( \{ (x_{(1)}, x_{(2)}, \ldots, x_{(n)}) / -\infty < x_{(1)} < x_{(2)} < \cdots < x_{(n)} < \infty \} \).

The transformation maps \( A \) onto \( B \), but not in a one-to-one fashion, since each of the \( n! \) permutations of the observed values yields the same values for the order statistics.

When \( n = 2 \), let \( (x_1, x_2) = (2, 3, 3, 1) \) and \( (x_2, x_1) = (3, 2, 3, 1) \) both yield \( (x_{(1)}, x_{(2)}) = (2, 3, 3, 1) \). If we partition the set \( A \) into \( n! \) subsets each one corresponding to a particular ordering of the sample observations, then the order statistics transformation maps each of these partitioned sets onto \( B \) in a one-to-one fashion.

Let \( A \) be partitioned as
\[
A_1 = \{ (x_1, x_2) / -\infty < x_1 < x_2 < \infty \}
A_2 = \{ (x_1, x_2) / -\infty < x_2 < x_1 < \infty \}
\]

On \( A_1 \) set \( x_1 = x_{(1)} \) and \( x_2 = x_{(2)} \) then the Jacobian \( |J_1| = 1 \)
and on \( A_2 \) set \( x_1 = x_{(2)} \) and \( x_2 = x_{(1)} \) yielding \( |J_2| = 1 \)

Then in general for \( n \), it follows similarly that the Jacobian of each of the one-to-one transformation from one of the \( n! \) partition of \( A \) onto \( B \) has an absolute value equal to 1.

Then the joint density function of the order statistics is the sum of the contributions from each of the partitions.

i.e. \[ g(x_{(1)}, x_{(2)}) = f(x_{(1)})f(x_{(2)})|J_1| + f(x_{(2)})f(x_{(1)})|J_2| \]
\[ = 2f(x_{(1)})f(x_{(2)}) , -\infty < x_{(1)} < x_{(2)} < \infty \]

For general \( n \), the joint density is given by
\[ g(x_{(1)}, x_{(2)}, \ldots, x_{(n)}) = n! f(x_{(1)}) f(x_{(2)}) \ldots f(x_{(n)}). \]

\[ = n! \prod_{i=1}^{n} f(x_{(i)}). \]

**Remark:** The joint density of the order statistic corresponding to an underlying uniform distribution in the interval [0,1] is given by

\[ g(x_{(1)}, x_{(2)}, \ldots, x_{(n)}) = n!, \quad \alpha < x_{(1)} < x_{(2)} < \ldots < x_{(n)} < 1. \]

\[ = 0, \text{ elsewhere.} \hspace{1cm} (1.2.2) \]

This distribution plays an important role in nonparametric statistics. This is primarily due to a result referred to as the probability integral transformation.

### 1.3. Probability Integral Transformation.

For a random variable \( X \) with c.d.f. \( F(x) \), we define the inverse distribution function \( F^{-1}(.) \) by

\[ F^{-1}(y) = \inf \{ x \mid F(x) \geq y \}, \quad 0 < y < 1. \hspace{1cm} (1.3.1) \]

If \( F(x) \) is strictly increasing between 0 and 1, then there is only one \( x \) such that \( F(x) = y \).

Suppose there is some \( x \) such that \( F(x) = y \). Since \( F(.) \) is continuous from the right, \( F(F^{-1}(y)) = y \).

In particular, if \( F(.) \) is continuous then \( F(F^{-1}(y)) = y \), \( \forall y \) satisfying \( 0 < y < 1 \).

However, if \( F(.) \) is the c.d.f. of discrete distribution, then for a given \( y \) there may be no \( x \) for which \( F(x) = y \). In such cases
$F^{-1}(y)$ is the smallest $x$ yielding an $F(x)$ value larger than $y$, and hence we have the relationship,

$$y \leq F(F^{-1}(y)) \quad \text{for } 0 < y < 1.$$ 

**THEOREM 1.2.** Let $x$ be a continuous random variable with c.d.f. $F(x)$. Then the random variable $y = F(x)$ has a uniform distribution on $[0,1]$.

**Proof:** Since $F(x)$ is continuous, $F(F^{-1}(y)) = y$ for $0 < y < 1$.

Using the monotonicity property of $F(x)$ we see that:

$$<x \leq F^{-1}(y) \quad \Rightarrow \quad \{x \leq F(x) \leq F(F^{-1}(y)) = y\}$$

Also:

$$\{F(x) \leq y\} = \{x \leq F^{-1}(y)\} \cup \{x > F^{-1}(y)\} \quad \text{and} \quad F(x) = y$$

The continuous distribution of $X$ is $P(F(x) = y) = 0$ \hspace{1cm} (1.3.2)

Thus $P(F(x) \leq y) = P(x \leq F^{-1}(y)) + P(x > F^{-1}(y))$ and $F(x) = y$

Therefore,

$$P(F(x) \leq y) = P(x \leq F^{-1}(y))$$

Since $P(x > F^{-1}(y))$ and $F(x) = y = 0$ due to (1.3.2)

Let $H(y)$ be the distribution function of $y$.

Since $y$ takes values from $[0,1]$ only,

$$H(y) = 0 \quad , \text{for } y < 0$$

$$= 1 \quad , \text{for } y \geq 1. \hspace{1cm} (1.3.3)$$

Also

$$H(y) = P(Y \leq y)$$

$$= P(F(x) \leq y)$$

$$= P(x \leq F^{-1}(y))$$

$$= F(F^{-1}(y))$$

$$= y \, , \quad 0 < y < 1. \hspace{1cm} (1.3.4)$$
Using (1.3.3) and (1.3.4) and the nondecreasing nature of HCyD we see that HCyD is the distribution of a uniform on [0,1]. □

1.4. DISTRIBUTION-FREE STATISTICS.

Let $X_1, X_2, X_3, \ldots, X_n$ be random variables with a joint distribution denoted by $D$, where $D$ is a member of some collection $\mathcal{D}$ of possible joint distributions. We use $T(X_1, X_2, X_3, \ldots, X_n)$ to denote some statistic based on these random variables.

**Definition. 1.1.** The statistic $T$ as mentioned above is said to be distribution-free over $\mathcal{D}$ if the distribution of $T$ is the same for every joint distribution of $\mathcal{D}$.

**Example:** Let $\mathcal{D}_1$ denote the collection of joint distributions of $n$ independently and identically distributed variables with known mean $\mu_0$ and unknown variance $\sigma_0^2 < \sigma^2 < \infty$. Let $\mathcal{D}_2$ be a second collection of joint distributions of $n$ i.i.d. normals with unknown mean $\mu$, $-\infty < \mu < \infty$ and known variance $\sigma_0^2$.

Define

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

and $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$ be the sample mean and variance respectively.

Then consider the two statistics,

$$U_1 = n^{1/2} \frac{(\overline{X} - \mu_0)}{\sigma} \text{ and } U_2 = (n-1) \frac{S^2}{\sigma_0^2}.$$
Thus $U_1$ is the well known $t$-statistic for testing the null hypothesis $H_0: \mu = \mu_0$ and $U_2$ is used to test $H_1: \sigma^2 = \sigma_0^2$.

The statistic $U_1$ follows a $t$-distribution with $(n-1)$ degrees of freedom when $H_0$ is true and $U_2$ follows a chi-square distribution with $(n-1)$ degrees of freedom when $H_1$ is true.

Thus $U_1$ is distribution-free over $\mathcal{D}_1$ and $U_2$ is distribution-free over $\mathcal{D}_2$. That is each of these statistics has a distribution-free property over a parametric class consisting of i.i.d. normal variates.

1.4.1. COUNTING STATISTICS.

Let $x_1, x_2, x_3, \ldots, x_n$ be independent random variables such that $x_i$ has a continuous distribution with c.d.f. $F_i(x)$, $i = 1, 2, 3, \ldots, n$. Assume that $F_i(\theta) = \theta_i^p$, $0 < \theta_i < 1$, for $i = 1, 2, 3, \ldots, n$, for some unknown $\theta$.

Let $\theta_0$ be a known real number and define
\[ \psi_i = \psi(x - \theta_0), \quad i = 1, 2, 3, \ldots, n \ldots \quad (1.4.1) \]

Where \[ \psi(t) = 1, \quad \text{for } t > 0 \]
\[ = 0, \quad \text{for } t < 0. \]

THEOREM 1.3. Let $S(\psi_1, \psi_2, \psi_3, \ldots, \psi_n)$ be any statistic based on $\psi_1, \psi_2, \psi_3, \ldots, \psi_n$ only. Then if $\theta = \theta_0$, the statistics $\psi_1, \psi_2, \ldots, \psi_n$ are independent identically distributed Bernoulli random variables with parameter $1 - \theta_0$ and $S(\psi_1, \psi_2, \ldots, \psi_n)$ is distribution-free over the nonparametric class $\mathcal{D}_g$ consisting of all joint distribution of independent continuous random variables each with \( p_0 \) th quantile.
equal to \( \theta_0 \).

Proof. Since \( \psi_i \) is a function of \( x_i \) only, \( x_1, x_2, ..., x_n \) are independent, then \( \psi_1, \psi_2, ..., \psi_n \) are mutually independent variables. Moreover \( \psi_i \) has a Bernoulli distribution with parameter \( 1 - F_i(\theta_0) \), for \( i = 1, 2, ..., n \). Hence for the class \( \mathcal{D}_3 \), the \( \psi_i \) are independent and identically distributed Bernoulli variables with common parameter \( 1 - F_i(\theta_0) = 1 - p_0 \). By definition then any statistic \( \Sigma \psi_1, \psi_2, ..., \psi_n \) based on \( \psi_1, \psi_2, ..., \psi_n \) only, will be distributed as that of a function of \( n \) i.i.d. Bernoulli variables with parameter \( 1 - p_0 \), provided that the joint distribution belongs to \( \mathcal{D}_3 \). This is another way of saying that \( \Sigma \psi_1, \psi_2, ..., \psi_n \) is distribution-free over \( \mathcal{D}_3 \).

Sign test statistic is an example for these types of statistics.

i.e. if we set \( B = B(\psi_1, \psi_2, ..., \psi_n) = \sum_{i=1}^{n} \psi_i \) \hspace{1cm} (1.4.2.)

Then \( B \) is distributed as a binomial variable with parameters \( n \) and \( (1 - p_0) \), provided the joint distribution is \( \mathcal{D}_3 \). In particular, when \( p_0 = 0.5 \) and \( H_0 \): each \( x_i \) has median \( \theta_0 \), then \( B = \Sigma \psi_i \) is known as the sign test statistic. This is much useful for detecting the location alternatives for which each \( x_i \) has a median than is greater than \( \theta_0 \).

Thus Sign test is distribution-free under \( H_0 \) with a null distribution of Binomial \( (n, 1/2) \).
1.4.2. RANKING STATISTICS.

Let \( x_1, x_2, ..., x_n \) be a random sample from a continuous distribution with distribution function \( F(x) \), and let \( x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)} \) be the corresponding order statistics. Then

**Definition. 1.2.** The sample observation \( x_i \) is said to have Rank \( R_i \) among \( x_1, x_2, ..., x_n \) if \( x_i = x_{(R_i)} \), provided the \( R_i \)th order statistic is uniquely defined.

The following theorem gives an important result which enables to construct distribution-free rank statistics.

**Theorem 1.4.** Let \( x_1, x_2, ..., x_n \) be a random sample from a continuous distribution and let \( R = (R_1, R_2, ..., R_n) \) be the vector of ranks, where \( R_i \) is the rank of \( x_i \) among \( x_1, x_2, ..., x_n \). If \( R = (r_1, r_2, ..., r_n) \) is a permutation of the integers \( 1, 2, 3, ..., n \) then \( R \) is uniformly distributed over \( \{R\} \).

**Proof.** We note that there are \( n! \) elements in \( R \). Then it is sufficient to show that \( R^* \) assumes each of the permutations of \( (1, 2, 3, ..., n) \) with probability \( 1/n! \).

Let \( r = (r_1, r_2, ..., r_n) \) be an arbitrary element of \( \{R\} \).

Then \( P(R^* = r) = P\left( (z_1, z_2, ..., z_n) = (z_{(r_1)}, z_{(r_2)}, ..., z_{(r_n)}) \right) \)

\[ = P(z_{d_1}, z_{d_2}, ..., z_{d_n}) \] where \( d_i \) is the position of the
number in the permutation \( r \) for \( i = 1, 2, 3, \ldots, n \). Then
\[
(z_1, z_2, \ldots, z_n) \text{ and } (z'_1, z'_2, \ldots, z'_n)
\]
are equal in distribution, which implies that
\[
\Pr(z_1 < z_2 < \ldots < z_n) = \Pr(z'_1 < z'_2 < \ldots < z'_n).
\]

Thus we have \( \Pr(R^* = r) = \Pr(z_1 < z_2 < \ldots < z_n) = \Pr(R^* = r_0) \), where \( r_0 = (1, 2, \ldots, n) \). Since there are \( n! \) elements in \( R \) and that \( r \) is arbitrary, the result follows.

\[\Box\]

1.4.3. STATISTICS UTILIZING COUNTING AND RANKING.

We have already seen the two techniques for constructing distribution-free statistics, those of Counting and Ranking. Here we consider a general setting in which such a combination is useful.

Let \( x_1, x_2, \ldots, x_n \) be a random sample from a continuous distribution that is symmetric about zero.

Letting \( x = x - \theta_0 \) is symmetrically distributed about 0, define,
\[
y_i = \psi(x_i), i = 1, 2, 3, \ldots, n \tag{1.4.3}
\]
where \( \psi(t) \) is as per (1.4.1). Then we have the following lemma.

Lemma 1.1. Let \( x \) be a continuous random variable with a distribution that is symmetric about 0. Then the random variables \( |x| \) and \( \psi = \psi(x) \) are stochastically independent.

Proof. Let \( F(x) \) be the c.d.f. of \( x \).

Then case (i): When \( t \geq 0 \).
PC(\psi = 1, |x| \leq t) = PC(x > 0, |x| \leq t)
= PC(x \leq t)
= F(t) - F(0), \text{Since } x \text{ is continuous.}
= 1/2 \{ F(t) - F(-t) \}, \text{Since the distribution of } x \text{ is symmetric about 0.}
= 1/2 \{ PC(t \leq x \leq t) \}
= 1/2. PC(|x| \leq t)
= PC(\psi = 1) \cdot PC(|x| \leq t).

Since 0 is the median of the x distribution.

case (ii) : when \( t < 0 \)
PC(\psi = 0, |x| \leq t) = PC(|x| \leq t) - PC(\psi = 1, |x| \leq t)
= PC(|x| \leq t) \cdot [1 - PC(\psi = 1)], \text{Since}
PC(\psi = 1, |x| \leq t) \text{ is independent.}
= PC(|x| \leq t) \cdot PC(\psi = 0).

Hence \( \psi \text{ and } |x| \) are stochastically independent. 

\begin{definition}

For any random variables \( x_1, x_2, \ldots, x_n \), the absolute rank of \( x_i \), denoted by \( R_i^+ \), is the rank of \( |x_i| \) among \( |x_1|, |x_2|, \ldots, |x_n| \). The signed rank of \( x_i \) is then \( \psi_i R_i^+ \), where \( \psi_i \) is defined in (1.4.3).

i.e. The signed rank of positive observation is simply its absolute rank but the signed rank of a negative observation is zero.

Many distribution-free statistics are functions of \( \psi = (\psi_1, \psi_2, \ldots, \psi_n) \) and the absolute rank vector \( R^+ = (R_1^+, R_2^+, \ldots, R_n^+) \). The following theorem establishes an
important joint distributional structure of $\gamma$ and $R^*$.

THEOREM 1.5. Let $x_1, x_2, \ldots, x_n$ denote a random sample from a continuous distribution that is symmetric about zero. Let $R^*$ denote the vector of absolute ranks of the $x$'s. If $\gamma_i$ is defined by (1.4.3) for $i = 1, 2, \ldots, n$ then the $n + 1$ random variables $\gamma_1, \gamma_2, \ldots, \gamma_n, R^*$ are mutually independent. Moreover each $\gamma_i$ is a Bernoulli random variable with $p = 1/2$, and $R^*$ is uniformly distributed over $\mathcal{R}$, the set of all permutations of the integers $(1, 2, 3, \ldots, n)$.

Corollary: Let $S(\gamma, R^*)$ be a statistic that depends on the observations $x_1, x_2, \ldots, x_n$ only through $\gamma_1, \gamma_2, \ldots, \gamma_n$ and $R^*$. Then the statistic $S(\gamma, R^*)$ is distribution-free over $\mathbb{D}_4$, the collection of joint distributions of $n$ i.i.d. continuous random variables, each symmetrically distributed about zero. This is because $\gamma$ and $R^*$ have the same joint distribution for every joint distribution in $\mathbb{D}_4$.

1.5. U-STATISTICS.

In this section we study a class of unbiased estimators of characteristics of a population, which are generally known as $U$-Statistics and some of their properties.

1.5.1. ONE-SAMPLE U-STATISTICS.

Let $x_1, x_2, \ldots, x_n$ be $n$ independent observations from a distribution function $F$ (a vector valued variables). Let $F$ be a
class of cumulative distribution functions or $\mathbb{R}$, and $\Theta(\mathbb{F})$ be a real valued functional of the c.d.f. $F$ whose domain is $\mathbb{F}$ and whose Range is a subset of $\mathbb{R}$. $\Theta(\mathbb{F})$ may also be a vector valued having $s$ ($s \geq 1$) components.

Definition 1.4. A parameter $\Theta(\mathbb{F})$ is said to be estimable of degree $r$ for the family of distributions $\mathbb{F}$ if $r$ is the smallest sample size for which there exists a function $\phi(x_1, x_2, \ldots, x_n)$ such that

$$E_F[\phi(x_1, x_2, \ldots, x_n)] = \Theta(\mathbb{F}), \text{ for all } F \in \mathbb{F}.$$  \hfill (1.5.1)

The function $\phi(x_1, x_2, \ldots, x_n)$ defined in (1.5.1) is called the Kernel of the parameter $\Theta(\mathbb{F})$. Without loss of generality, we can assume that a Kernel is symmetric in its arguments. That is

$$\phi(x_1, x_2, \ldots, x_n) = \phi(x_{\alpha_1}, x_{\alpha_2}, \ldots, x_{\alpha_r}), \text{ for every permutation } (\alpha_1, \alpha_2, \ldots, \alpha_r) \text{ of } 1, 2, \ldots, r.$$  

Otherwise the symmetric Kernel is given by,

$$\phi(x_1, x_2, \ldots, x_n) = (\frac{1}{n!}) \sum \phi(x_{\alpha_1}, x_{\alpha_2}, \ldots, x_{\alpha_r}).$$  \hfill (1.5.2)

Where the summation is over all permutations of the integers $(1, 2, \ldots, n)$

Now, suppose we have a random sample $x_1, x_2, \ldots, x_n$, $n \geq r$, from a distribution with c.d.f. $F \in \mathbb{F}$ then the U-statistic have been defined as follows.

Definition 1.5. A U-statistic for the estimable parameter $\Theta(\mathbb{F})$ of degree $r$ is created with the symmetric Kernel $\phi(.)$ by forming
U = \binom{c(n)}{c} \sum_{\alpha \in A} \phi(x_{\alpha_1}, x_{\alpha_2}, \ldots, x_{\alpha_r}) \tag{1.5.30}

Where \( A = \{ \alpha/\alpha \text{ is one of the } c(n,r) \text{ unordered subsets of } r \} \), integers chosen without replacement from the set \((1,2,\ldots,n)\).

Note that a U-statistic is an unbiased estimator of \( \theta(F) \) for every \( F \in \mathcal{F} \) and is symmetric in its arguments.

Consider,

\[
\phi_c(x_1, x_2, \ldots, x_c) = E_F [\phi(x_{1}, x_{2}, \ldots, x_{c+1}, x_{c+1}, \ldots, x_r)] \tag{1.5.4}
\]

\[
\psi_c(x_1, x_2, \ldots, x_c) = \phi_c(x_1, x_2, \ldots, x_c) - \theta(F) \tag{1.5.5}
\]

and

Let \( \zeta_c(F) \) denote the covariance of these observations then,

\[
\zeta_c(F) = E_F [\psi_c(x_1, x_2, \ldots, x_c)^2] \tag{1.5.6}
\]

Where \( c = 1,2,\ldots,m \), \( F \in \mathcal{F} \), \( \phi_0 = \theta(F) \) and \( \zeta_0(F) = 0 \). Furthermore since \( \phi_c \) is symmetric in its arguments and the variables \( x_1, x_2, \ldots, x_n \) are i.i.d, it follows that

\[
\zeta_c(F) = \text{Cov} \{ \phi(x_{\alpha_1}, x_{\alpha_2}, \ldots, x_{\alpha_r}), \phi(x_{\alpha'_1}, x_{\alpha'_2}, \ldots, x_{\alpha'_r}) \} \tag{1.5.7}
\]

Where \( \alpha = (x_{\alpha_1}, x_{\alpha_2}, \ldots, x_{\alpha_r}) \) and \( \alpha' = (x_{\alpha'_1}, x_{\alpha'_2}, \ldots, x_{\alpha'_r}) \) are subsets of the integers \((1,2,\ldots,n)\) having exactly \( c \) integers in common. If \( \alpha \) and \( \alpha' \) have no integers in common, then \( \phi(x_{\alpha_1}, x_{\alpha_2}, \ldots, x_{\alpha_r}) \) and \( \phi(x_{\alpha'_1}, x_{\alpha'_2}, \ldots, x_{\alpha'_r}) \) are independent.

Hence \( \zeta_0(F) = 0 \), and \( 0 \leq \zeta_c(F) \leq \zeta_1(F) \), \( 1 \leq c \leq r \).

Now, the variance of the U-statistics is given by

\[
\text{Var}(U) = E \{ \binom{c(n)}{c} \sum_{\alpha \in A} (\phi(x_{\alpha_1}, x_{\alpha_2}, \ldots, x_{\alpha_r}) - \theta(F))^2 \}
\]
\[ = c(n, r)^{-2} \sum_{\alpha \in A} \sum_{\alpha' \in A} E \left[ (\phi(x_{\alpha_1}, x_{\alpha_2}, \ldots, x_{\alpha_r}) - \varepsilon(F)) \right] \\
\quad \left( (\phi(x_{\alpha_1}, x_{\alpha_2}, \ldots, x_{\alpha_r}) - \varepsilon(F)) \right) \]

\[ = c(n, r)^{-2} \sum_{\alpha \in A} \sum_{\alpha' \in A} \text{Cov}(\phi(x_{\alpha_1}, x_{\alpha_2}, \ldots, x_{\alpha_r}), \phi(x_{\alpha'_1}, x_{\alpha'_2}, \ldots, x_{\alpha'_r})) \]

(1.5.9)

All terms in (1.5.9) for which \( \alpha \) and \( \alpha' \) have exactly \( c \) integers in common have the same covariance, \( \zeta_c \). The number of such terms are \( \binom{n}{r} \binom{r}{c} \binom{n-r}{r-c} \).

Thus it follows that

\[ \text{Var}(U) = \binom{n}{r}^{-2} \sum_{c=0}^{r-1} \binom{r}{c} \binom{n-r}{r-c} \zeta_c(F) \]

\[ = \binom{n}{r}^{-1} \sum_{c=1}^{r} \binom{r}{c} \binom{n-r}{r-c} \zeta_c(F), \quad (1.5.10) \]

Since \( \zeta_0 = 0 \).

Result. As \( n \to \infty \), \( \lim_{n \to \infty} \text{Var}(U) = 0 \) and \( \lim_{n \to \infty} n \cdot \text{Var}(U) = r^2 \zeta_1(F) \). The asymptotic distribution of U-statistics is given in the following theorem.

**THEOREM 1.6.** Let \( \{U(x_1, x_2, \ldots, x_n)\} \) be the U-statistics for a symmetric kernel \( \phi(x_1, x_2, \ldots, x_r) \). If \( E[\phi^2(x_1, x_2, \ldots, x_r)] < \infty \) then,

\[ \lim_{n \to \infty} n \cdot \text{Var} \{U(x_1, x_2, \ldots, x_n)\} = r^2 \zeta_1(F). \]

**Proof.** Since \( E[\phi^2(x_1, x_2, \ldots, x_n)] < \infty \), \( \zeta_r = \text{Var}(\phi(x_1, x_2, \ldots, x_r)) \) exists. Thus using (1.5.8) and (1.5.10) we see that \( \text{Var}(U(x_1, x_2, \ldots, x_r)) \) exists.
Define a constant \( K_c = \frac{(r+c)^2}{(c!)(r-c)!^2} \), \( c = 1,2,\ldots,r \).

We have from (1.5.10)

\[
\text{Var}(U) = c(n,r)^{-4} \sum_{c=1}^{r} c(r,c) c(n-r,r-c) \xi_c(F).
\]

Consider the general term \( \binom{r}{c} \binom{n-r}{r-c} / \binom{n}{r} \xi_c(F) \).

Multiply by \( n \), then,

\[
\frac{n \binom{r}{c} \binom{n-r}{r-c}}{\binom{n}{r}} \xi_c(F) = K_c n \frac{(n-r)(n-r-1)\ldots(n-2r+c-1)}{n(n-1)(n-2)\ldots(n-r-1)} \xi_c.
\]

There are \( r-c+1 \) factors in the numerator involving \( n \), and \( r \) such factors in the denominator.

Thus if \( c = 1 \) and \( n \to \infty \) the term becomes \( K_1 \xi_1(F) \).

For \( c > 1 \) the term becomes zero and \( K \) becomes \( r^2 \).

Thus \( \lim_{n \to \infty} n \text{Var}(u) = r^2 \xi_1(F) \).

**THEOREM 1.7.** One sample U-statistics theorem (Hoeffding, 1948).

Let \( x_1, x_2, \ldots, x_n \) denote a random sample from some population.

Let \( \theta(F) \) be an estimable parameter of degree \( r \) with symmetric kernel \( \phi(x_1, x_2, \ldots, x_n) \). If \( E \{ \phi(x_1, x_2, \ldots, x_r) \} = \theta \) and if \( U(x_1, x_2, \ldots, x_n) \) is as per (1.5.3) then

\[
n^{-r/2} \{ U(x_1, x_2, \ldots, x_n) - \theta(F) \},
\]

has a limiting distribution with mean 0 and variance \( r^2 \xi_1 \).

For the Proof see Randles, R.H. and Wolfe, D.A. (1979, P.P.82).

1.5.2. TWO SAMPLE U-STATISTICS.

The technique used for constructing unbiased estimators based on a single sample extends directly to many other settings.
Let \( x_1, x_2, \ldots, x_m \) and \( y_1, y_2, \ldots, y_n \) be independent random samples from distributions with c.d.f.'s \( F(x) \) and \( G(y) \) respectively.

A parameter \( e \in \mathcal{F}, \mathcal{G} \) is said to be estimable of degree \((r,s)\) if \( r \) and \( s \) are the smallest sample sizes for which there exists an estimator \( \hat{e} \) of \( e \) that is unbiased for every \( (\mathcal{F},\mathcal{G}) \in \mathcal{F} \).

i.e. there exists a function \( \phi(x_1, x_2, \ldots, x_r, y_1, y_2, \ldots, y_s) \) such that
\[
E_{(\mathcal{F},\mathcal{G})}(\phi(x_1, x_2, \ldots, x_r, y_1, y_2, \ldots, y_s)) = e_{(\mathcal{F},\mathcal{G})},
\]
for every \( (\mathcal{F},\mathcal{G}) \in \mathcal{F} \), which is the two-sample Kernel and is assumed to be symmetric in its \( x \) components and \( y \) components separately. Then the U-statistics based on these two-sample Kernel is defined as follows.

**Definition 1.6.** An estimable parameter \( e \) of degree \((r,s)\) and with symmetric Kernel \( \phi \), for \( m \geq r \) and \( n \geq s \) has the form,

\[
U(x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n)
\]

\[
= \binom{m}{r}^{-1} \binom{n}{s}^{-1} \sum_{a \in A} \sum_{\beta \in B} \phi(x_{a_1}, x_{a_2}, \ldots, x_{a_r}; y_{\beta_1}, y_{\beta_2}, \ldots, y_{\beta_s}),
\]

Where \( A \) is the collection of all subsets of \( r \) integers chosen without replacement from the integers \( \{1, 2, \ldots, m\} \) and \( B \) is the collection of \( s \) integers chosen without replacement from the integers \( \{1, 2, \ldots, n\} \).

Let \( \xi_{c,d} \) denote the covariance between the two Kernel random variable with exactly \( c \) \( x \)'s and \( d \) \( y \)'s in common such that \( c \leq c \).
\[ \begin{align*}
\zeta_{c,d} &= \text{COV} \left\{ \phi(x_1, x_2, \ldots, x_c, x_{c+1}, \ldots, x_{2r-c}; y_1, y_2, \ldots, y_d, y_{d+1}, \ldots, y_e) \right\} \\
&= E \left\{ \phi(x_1, x_2, \ldots, x_{c-r}, x_{c-r+1}, \ldots, x_{2r-c}; y_1, y_2, \ldots, y_d, y_{d+1}, \ldots, y_{2e-d}) \right\} - \theta^2 \\
&= \mathcal{O} \left( \frac{1}{n} \right) \quad \text{as } n \to \infty
\end{align*}\]

(1.5.11)

Stating \( \zeta_{0,0} = 0 \),

\[ \begin{align*}
\text{Var} \left\{ \phi(x_1, x_2, \ldots, x_r; y_1, y_2, \ldots, y_n) \right\} \\
&= \left( \begin{array}{cc}
m & n \\
r & s \end{array} \right)^{-2} E \left\{ \sum_{\alpha \beta} \phi(x_1, x_2, \ldots, x_r; y_1, y_2, \ldots, y_n) \left( \phi(x_1, x_2, \ldots, x_r; y_1, y_2, \ldots, y_n) - \theta \right) \right\} \\
&= \left( \begin{array}{cc}
m & n \\
r & s \end{array} \right)^{-2} \sum_{\alpha \beta} \phi(x_1, x_2, \ldots, x_r; y_1, y_2, \ldots, y_n) \left( \phi(x_1, x_2, \ldots, x_r; y_1, y_2, \ldots, y_n) - \theta \right) \\
&= \left( \begin{array}{cc}
m^2 & n^2 \\
r & s \end{array} \right) \sum_{c=0}^{r} \sum_{d=0}^{s} \left( \begin{array}{cc}
m & n \\
r & s \end{array} \right) \left( \begin{array}{cc}
r & r \end{array} \right) \left( \begin{array}{cc}
r & r-c \end{array} \right) \left( \begin{array}{cc}
s & s \end{array} \right) \left( \begin{array}{cc}
s & s-d \end{array} \right) \xi_{c,d} \\
&= \left( \begin{array}{cc}
m & n \\
r & s \end{array} \right)^{-1} \sum_{c=0}^{r} \sum_{d=0}^{s} \left( \begin{array}{cc}
r & r \end{array} \right) \left( \begin{array}{cc}
r & r-c \end{array} \right) \left( \begin{array}{cc}
s & s \end{array} \right) \left( \begin{array}{cc}
s & s-d \end{array} \right) \xi_{c,d} \\
&= \mathcal{O} \left( \frac{1}{n} \right) \quad \text{as } n \to \infty
\end{align*}\]

(1.5.12)

Now let \( N = m + n \), and index the sample sizes by \( N \), \( x_i \), sequences by \( m \), and \( y_i \), sequences by \( n \).

Then \( \lim_{N \to \infty} \frac{c_m}{N} = \lambda \) and \( \lim_{N \to \infty} \frac{c_n}{N} = 1 - \lambda \), \( 0 < \lambda < 1 \).

Then the asymptotic behaviour of the two-sample U-statistic is established in the following theorem.
THEOREM 1.8. If $E \{ \phi^2(x_1, x_2, \ldots, x_n; y_1, y_2, \ldots, y_m) \} < \infty$ then

$$\lim_{N \to \infty} N \cdot \text{Var}(U_{ij} \mid x_1, x_2, \ldots, x_i; y_1, y_2, \ldots, y_j) = \frac{r^2 \xi_{1,0} + s^2 \xi_{0,1}}{\lambda (1-\lambda)}.$$ 


The asymptotic normality is due to Lehman (1951), of Hoeffding (1948) U-statistic theorem to two sample statistics.

THEOREM 1.9. Let $x_1, x_2, \ldots, x_m$ and $y_1, y_2, \ldots, y_n$ denote independent random variables from populations with c.d.f.'s $F(x)$ and $G(y)$ respectively. Let $\phi_r(s)$ denote a symmetric Kernel for an estimable parameter $\theta$ of degree $(r, s)$.

If $E \{ \phi^2(x_1, x_2, \ldots, x; y_1, y_2, \ldots, y) \} < \infty$ then,

$$N^{1/2} (U_{ij} \mid x_1, x_2, \ldots, x_i; y_1, y_2, \ldots, y_n) - \theta$$

has a limiting normal distribution with mean 0 and variance

$$\left[ \frac{r^2 \xi_{1,0}}{\lambda} + s^2 \xi_{0,1} \right],$$

provided the variance is positive, where $0 < \lambda = \lim_{N \to \infty} m/n$, and $\xi_{1,0}$ and $\xi_{0,1}$ are as per (1.5.11).

1.6. K-SAMPLE U-STATISTICS.

In this section we will be discussing the generalization of one sample and two sample U-statistics.

Let $x_{ij}, \ j = 1, 2, \ldots, n_i, \ i = 1, 2, \ldots, k$ be k independent random samples with the $i$th sample having c.d.f. $F_i(x)$. 

A parameter $\theta$ is said to be estimable of degree $(r_1, r_2, \ldots, r_k)$ for distributions $(F_1, F_2, \ldots, F_k)$ in some family $F$, if $(r_1, r_2, \ldots, r_k)$ are the smallest sample sizes for which there exists an estimator of $\theta$ which is unbiased for every
Let \( \phi(x_{i1}, x_{i2}, \ldots, x_{ir_i}; x_{i1}, x_{i2}, \ldots, x_{ir_i}; x_{k1}, x_{k2}, \ldots, x_{kr_k}) \) be a parametric function that is separately symmetric in the arguments \( x_{i1}, x_{i2}, \ldots, x_{ir_i}, i = 1, 2, \ldots, k \) if

\[
E_{(F_1, F_2, \ldots, F_k)} [\phi(x_{i1}, x_{i2}, \ldots, x_{ir_i}; \ldots; x_{k1}, x_{k2}, \ldots, x_{kr_k})] = 0
\]

for all \( (F_1, F_2, \ldots, F_k) \in \mathcal{F} \). Then \( \phi(\cdot) \) is called a \( k \)-sample symmetric Kernel for \( \epsilon \).

**Definition 1.7.** For any estimable parameter \( \epsilon \) of degree \( (r_1, r_2, \ldots, r_k) \) and symmetric Kernel \( \phi(\cdot) \), the \( k \)-sample U-statistic has for \( n_i \geq r_i \), the form

\[
U(x_{i1}, x_{i2}, \ldots, x_{ir_i}; \ldots; x_{k1}, x_{k2}, \ldots, x_{kr_k})
\]

\[
= \binom{n_i}{r_i} \sum_{\alpha \in \mathcal{A}_i} \sum_{\beta \in \mathcal{A}_i} \phi(x_{i1}, x_{i2}, \ldots, x_{ir_i}; \ldots; x_{k1}, x_{k2}, \ldots, x_{kr_k})
\]

Where \( \alpha_i = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{ir_i}) \) and \( \mathcal{A}_i \) is the collection of all subsets of \( r_i \) integers chosen without replacement from the integers \( \{1, 2, \ldots, n_i\} \), for \( i = 1, 2, \ldots, k \).

To obtain the asymptotic normality of a \( k \)-sample U-statistics we proceed as in the two sample setting.

Let \( i \) be a fixed integer satisfying \( 1 \leq i \leq k \),

Define, \( H_{1i} = \phi(x_{i1}, x_{i2}, \ldots, x_{ir_i}; \ldots; x_{k1}, x_{k2}, \ldots, x_{kr_k}) \)

and \( H_{2i} = \phi(x_{i1}, x_{i2}, \ldots, x_{ir_i}; \ldots; x_{k1}, x_{k2}, \ldots, x_{kr_k}) \)

(1.8.2)
When \( j \neq i \) the two sets \( \{a_j, a_{j+1}, \ldots, a_{j_i}\} \) and \( \{\beta_j, \beta_{j+1}, \ldots, \beta_{j_k}\} \) have no integers in common and when \( j = i \) there will be exactly one integer in common.

Set \( \xi_{0,0,0,0} = \text{Cov}[H_{i_1}, H_{i_2}] \)

\[
= E[H_{i_1} H_{i_2}] - \delta^2. \tag{1.6.3}
\]

Where subscript \( i \) of \( \xi_{0,0,0,0} \) is in the \( i^{th} \) position.

Taking \( N = n_1 n_2 \ldots n_k \), we have the following \( k \)-sample U-statistic theorem due to Lehmann (1963a).

**THEOREM 1.10.** Let \( U(x_{i_1}, x_{i_2}, \ldots, x_{i_1}; \ldots; x_{k_1}, x_{k_2}, \ldots, x_{k_n}) \) be a \( k \)-sample U-statistic for the parameter \( \theta \) of degree \( (r_1, r_2, \ldots, r_k) \). If \( \lim_{N \to \infty} n^i_i / N = \lambda_i, 0 < \lambda_i < 1 \), for \( i = 1, 2, \ldots, k \) and if

\[
E[\phi^2(x_{i_1}, x_{i_2}, \ldots, x_{1 r_1}; \ldots; x_{k_1}, x_{k_2}, \ldots, x_{k r_k})] < \infty,
\]

then

\[
N^{1/2} [U - \theta] \text{ has a limiting normal distribution with mean 0 and variance } \sigma^2 = \sum_{i=1}^{k} r_i \xi_{0,0,0,0,0,0} \text{, provided } \sigma^2 > 0.
\]

**1.7. POWER FUNCTIONS AND THEIR PROPERTIES.**

**Definition 1.8.** Consider a model which is indexed by parameter(s) \( \xi \). The power function of a test of hypothesis relevant to this model is given by

\[
P(\xi) = P_{\xi} (\text{the null hypothesis is rejected}).
\]
Definition 1.9. The size of a test with power function $P(\xi)$ is $\sup_{\xi \in \omega} P(\xi)$. The test is said to be of level $\alpha$ if its size is less than or equal to $\alpha$.

An important point is to be noticed here is, since our hypothesis are often indexed by a parameter say, $\theta$, with an underlying distribution say $F(\cdot)$, the parameters for such a problem are $\xi = (\theta, F)$ and the corresponding null hypothesis is composite.

The unbiasedness of such a test is given by,

Definition 1.10. The $\alpha$-level test with power function $P(\xi)$ is said to be unbiased at level $\alpha$ if $P(\xi) \geq \alpha$ for all $\xi \in \Omega - \omega$.

Let $x_1, x_2, \ldots, x_n$ be i.i.d $F(x-\theta)$

\begin{equation}
\text{Where location parameter } \theta \text{ is the median of underlying continuous distribution.}
\end{equation}

We wish to test

\begin{equation}
H_0: \theta = \theta_0 \text{ versus } H_1: \theta > \theta_0
\end{equation}

THEOREM 1.11. We reject $H_0$ for large(small) values of a test statistic $S(x_1, x_2, \ldots, x_n)$ that holds

\begin{equation}
S(x_1^k, x_2^k, \ldots, x_n^k) \gtrless \leq S(x_1, x_2, \ldots, x_n)
\end{equation}

for every $K \geq 0$ and $(x_1, x_2, \ldots, x_n)$. Then the test has a monotonic power function in $\theta$ for the one-sample location problem;

\[ P_\alpha(\theta, F) \leq P_\alpha(\theta', F) \quad \text{for } \theta \leq \theta' \]

and any continuous distribution with c.d.f. $F(\cdot)$. 

27
Proof. Given $x_1, x_2, \ldots, x_n$ are i.i.d. $F(x-\theta)$ so $x_1 + (\theta' - \theta), x_2 + (\theta' - \theta), \ldots, x_n + (\theta' - \theta)$ are i.i.d $F(x-\theta')$.

Consider the case where we reject for large values of $S\xi_\circ$ and let $c$ be the corresponding critical value of the test.

Then for any $\theta \leq \theta'$,

$$P_{\xi,\circ}(\theta, F) = P \{ S(x_1, x_2, \ldots, x) \geq c / \text{each } x_i < F(x-\theta) \}$$

$$\leq P \{ S(x_1 + (\theta' - \theta), \ldots, x_n + (\theta' - \theta)) \geq c / \text{each } x_i + F(x-\theta) \}$$

$$= P \{ S(x_1^*, x_2^*, \ldots, x_n^*) \geq c / \text{each } x_i^* + F(x-\theta) \}$$

$$= P_{\xi,\circ}(\theta', F).$$

Similarly we can reject for small values of $S\xi_\circ$. $\square$

Consider a two sample problem with two independent samples

$$x_1, x_2, \ldots, x_m \text{ are i.i.d } F(x) \text{ and }$$

$$y_1, y_2, \ldots, y_n \text{ are i.i.d } F(x-\Delta) \quad (1.7.4)$$

Then we wish to test

$$H_0: \Delta = 0 \text{ versus } H_1: \Delta > 0. \quad (1.7.5)$$

THEOREM 1.12. Suppose that we reject $H_0$ in favour of $H_1$ for large(small) values of the test statistic $S(x_1, x_2, \ldots, x_m; y_1, y_2, \ldots, y_n)$ satisfying $S(x_1, x_2, \ldots, x_m; y_1^k, \ldots, y_n^k) \geq (\leq)

S(x_1, x_2, \ldots, x_m; y_1, y_2, \ldots, y_n)$, \quad (1.7.6)

for every $k \geq 0$ and every $(x_1, x_2, \ldots, x_m; y_1, y_2, \ldots, y_n)$. Then the power function of the test based on $S\xi_\circ$ is monotone in $\Delta$.

That is $P_{\xi,\circ}(\Delta, F) \leq P_{\xi,\circ}(\Delta', F)$ for all $\Delta \leq \Delta'$.

Another important property of power function is given by its asymptotic behaviour.
Let $\{S_n\}$ denote a sequence of test statistics based on a sample of size $N$, which yields a power function $P_{S_n}(\xi)$, for each $N$.

**Definition 1.11.** If $\{S_n\}$ denotes a sequence of $\alpha$-level tests of $H_0: \xi \in \omega$ versus $H_1: \xi \in \Omega - \omega$, then the sequence is said to be consistent against the class $\Omega - \omega$ if

$$\lim_{N \to \infty} P_{S_n}(\xi) = 1,$$

for every $\xi \in \Omega - \omega$.

In reference to this follows the theorem by Lehmann (1951).

**Theorem 1.13.** Let $\{S_n\}$ denote a sequence of test statistics for an $\alpha$-level test of $H_0: \xi \in \omega$ versus $H_1: \xi \in \Omega - \omega$, such that the test based on $S_n$ rejects $H_0$ if $S_n \geq c_n$. Suppose there exists a function $k(\xi)$ such that $S_n$ converges in probability to $k(\xi)$ for every $\xi \in \Omega$. If in addition,

$$k(\xi) = k_0, \text{ for all } \xi \in \omega,$$

$$k(\xi) > k_0, \text{ for all } \xi \in \Omega - \omega,$$

and

$$\lim_{N \to \infty} c_n \leq k_0,$$

then $\{S_n\}$ is a consistent sequence of tests for all alternatives in $H_1: \xi \in \Omega - \omega$.

### 1.8. Asymptotic Relative Efficiency of Tests

Consider two consistent tests $S$ and $T$ for testing $H_0: \theta = \theta_0$ against $H_1: \theta > \theta_0$. The question now arises is which test one has to prefer. Since the powers of the both tests tends to one at each fixed alternative point, the asymptotic power comparison does not give any idea about which test one has to prefer.
To overcome this difficulty one may think of considering a sequence \( \{e_i\} \) of alternatives in \( H_i \) and sequences of sample sizes \( \{n_i\} \) and \( \{m_i\} \) such that \( e_i \to e_0 \) as \( i \to \infty \) and limiting powers at \( e_i \) as \( i \to \infty \), of the tests \( S_{n_i} \) and \( T_{m_i} \) are same and bounded away from zero and one. Then the limiting inverse ratio of sample sizes \( \lim_{i \to \infty} m_i/n_i \) required by the two competing tests to have same limiting power may be considered as measure of efficiency of \( S \) with respect to the test \( T \). This concept of A.R.E was first introduced by Pitman (1948) and is given in the following definition.

Definition: 1.12. Let \( \{S_{n_i}\} \) and \( \{T_{m_i}\} \) be two sequences of test statistics for testing \( H_{0_i}: e = e_0 \), where \( e_0 \) is some specified parameter value, against a class of alternatives \( H_i \) of size \( \alpha \) and let \( \{e_i\} \) be a sequence of alternatives to \( H_{0_i} \) such that \( \lim_{i \to \infty} e_i = e_0 \). Let \( \beta_{S_{n_i}}(e_i) \) and \( \beta_{T_{m_i}}(e_i) \) be the powers of the tests based on \( S_{n_i} \) and \( T_{m_i} \) respectively. Let \( \{n_i\} \) and \( \{m_i\} \) be increasing sequences of positive integers such that

\[
\alpha < \lim_{i \to \infty} \beta_{S_{n_i}}(e_i) = \lim_{i \to \infty} \beta_{T_{m_i}}(e_i) < 1. \tag{1.8.1}
\]

Then the asymptotic relative efficiency (ARE) of \( \{S_{n_i}\} \) relative to \( \{T_{m_i}\} \) is

\[
\text{A.R.E}(S, T) = \lim_{i \to \infty} m_i/n_i, \tag{1.8.2}
\]

provided that this limit is the same for all such sequences \( \{n_i\} \) and \( \{m_i\} \) and independent of the sequence \( \{e_i\} \).
Remark: Note that based on the computation of A.R.E., we are able to decide which test is more efficient than other, only for fixed size \( \alpha \) and for fixed asymptotic power \( \beta \). Thus a test which is more efficient than its competitor at some fixed size \( \alpha \) common asymptotic power \( \beta \) may fail to be so at some other values of \( \alpha \) and \( \beta \). Hence it is quite natural to obtain the conditions under which A.R.E is free from \( \alpha \) and \( \beta \). This led to the following theorem by Noether (1955). Also the theorem ease the computation of A.R.E.

**THEOREM 1.14.** Let \( \{ S_n \} \) and \( \{ T_m \} \) be two sequence of size \( \alpha \) tests which are consistent for testing \( H_0: \theta = \theta_0 \) against \( H_1: \theta > \theta_0 \). Let \( \{ \theta_i \} \) be a sequence of alternatives in \( H_1 \) such that \( \lim_{i \to \infty} \theta_i = \theta_0 \). Let \( \{ n_i \} \) and \( \{ m_i \} \) be two sequence of natural numbers as given in Def. 1.12. Let \( \{ \mu_{S_n}(\theta) \} \), \( \{ \mu_{T_m}(\theta) \} \), \( \{ \sigma_{S_n}^2(\theta) \} \) and \( \{ \sigma_{T_m}^2(\theta) \} \) be sequences associated with \( \{ S_n \} \) and \( \{ T_m \} \) such that the following assumptions hold.

A1. \( \lim_{i \to \infty} P \left[ \{ S_n - \mu_{S_n}(\theta) \} / \sigma_{S_n}(\theta) \leq x \right] = \lim_{i \to \infty} P \left[ \{ T_m - \mu_{T_m}(\theta) \} / \sigma_{T_m}(\theta) \leq x \right] = H(\alpha), \)

A2. \( \lim_{i \to \infty} P \left[ \{ S_n - \mu_{S_n}(\theta) \} / \sigma_{S_n}(\theta) \leq x \right] = \lim_{i \to \infty} P \left[ \{ T_m - \mu_{T_m}(\theta) \} / \sigma_{T_m}(\theta) \leq x \right] = H(\alpha), \)

A3. \( \lim_{i \to \infty} \sigma_{S_n}(\theta) / \sigma_{S_n}(\theta_0) = \lim_{i \to \infty} \sigma_{T_m}(\theta) / \sigma_{T_m}(\theta_0) = 1, \)

A4. \( d/d\theta \left[ \mu_{S_n}(\theta) \right] = \mu'_{S_n}(\theta) \), \( d/d\theta \left[ \mu_{T_m}(\theta) \right] = \mu'_{T_m}(\theta), \)

are assumed to exist and be continous in some closed interval about \( \theta = \theta_0 \) with \( \mu'_{S_n}(\theta_0) \) and \( \mu'_{T_m}(\theta_0) \) both non-zero.
A5. \[ \lim_{n \to \infty} \frac{\mu'_{S}(e)}{\mu'_{S}(e_0)} = \lim_{m_i \to \infty} \frac{\mu'_{T}(e)}{\mu'_{T}(e_0)} = 1, \]

A6. \[ \lim_{n \to \infty} \frac{\mu'_{S}(e)}{\sigma^2_{S}(e_0)} = K_s \quad \text{and} \quad \lim_{m_i \to \infty} \frac{\mu'_{T}(e)}{\sigma^2_{T}(e_0)} = K_T; \]

Where \( K_s \) and \( K_T \) are positive constants for all choices of sequences \( \{n_i\} \) and \( \{m_i\} \) such that 1.12 holds. Then

\[ \text{A.R.E}(S,T) = \frac{K_s^2}{K_T^2}. \] (1.8.3)

**Definition.** 1.13. The quantities \( K_s^2 \) and \( K_T^2 \) defined in A6 are called efficacies of test \( S \) and \( T \) respectively.

**Remark.** Note that the definition of Asymptotic Relative Efficiency is based on the parametric set up. In the non parametric problems, in order to define the concept of A.R.E for two competing tests we will need to consider a sequence of distributions under alternative hypothesis which converges to a distribution in the null hypothesis as sample size tends to infinity. Also, we should be able to find power functions of the tests for this sequence of distributions under the alternative. This seems to be an impossible task in general and so as opposed to considering much larger class of distributions under alternative, we, in general, introduce a parametric sub-family of this and restrict our considerations of A.R.E as this parametric sub-family.