This dissertation has been microfilmed exactly as received

WILLKE, Thomas Aloys. A CLASS OF MULTIVARIATE RANK STATISTICS.

The Ohio State University, Ph.D., 1960
Mathematics

University Microfilms, Inc., Ann Arbor, Michigan
A CLASS OF MULTIVARIATE RANK STATISTICS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

THOMAS ALOYS WILLKE, A. B., M. Sc.

* * * * * * *

The Ohio State University
1960

Approved by:

[Signature]
Adviser
Department of Mathematics
ACKNOWLEDGMENT

The author wishes to express his grateful appreciation to Professor D. Ransom Whitney for his helpful criticism and encouragement in the preparation of this dissertation.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>I. A CLASS OF SAMPLE STATISTICS</td>
<td>3</td>
</tr>
<tr>
<td>Notation</td>
<td></td>
</tr>
<tr>
<td>The Class, $\mathcal{N}$, of Sample Statistics</td>
<td></td>
</tr>
<tr>
<td>Examples of Statistics in $\mathcal{N}$</td>
<td></td>
</tr>
<tr>
<td>A Class, $\mathcal{N}^2$, of Two Sample Statistics</td>
<td></td>
</tr>
<tr>
<td>II. EQUIVALENCE CLASSES OF RANDOM VARIABLES IN $\mathbb{E}$</td>
<td>14</td>
</tr>
<tr>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td>Equivalence Classes</td>
<td></td>
</tr>
<tr>
<td>Properties of Equivalence Classes</td>
<td></td>
</tr>
<tr>
<td>Invariance of Statistics in $\mathcal{N}$</td>
<td></td>
</tr>
<tr>
<td>Cumulative Expected Values</td>
<td></td>
</tr>
<tr>
<td>III. A CONVERGENCE THEOREM</td>
<td>28</td>
</tr>
<tr>
<td>IV. APPLICATIONS OF THE CONVERGENCE THEOREM</td>
<td>35</td>
</tr>
<tr>
<td>Application to a Subset of $\mathcal{N}$</td>
<td></td>
</tr>
<tr>
<td>Consistency</td>
<td></td>
</tr>
<tr>
<td>Application to Two Sample Statistics</td>
<td></td>
</tr>
<tr>
<td>V. REMARKS AND EXAMPLES ON LIMITING DISTRIBUTIONS</td>
<td>57</td>
</tr>
<tr>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td>Normalization of $S_n$</td>
<td></td>
</tr>
<tr>
<td>Comparison of $\xi_n$ and $\eta_n$</td>
<td></td>
</tr>
<tr>
<td>List of References</td>
<td>71</td>
</tr>
<tr>
<td>Autobiography</td>
<td>72</td>
</tr>
</tbody>
</table>
INTRODUCTION

In the consideration of multivariate random variables, questions arise which are totally absent for univariate random variables. These are, of course, the questions concerning the relationships and interdependence of the components on one another, that is, concerning what could be called the essential multivariate structure of the random variable. The one dimensional problems still exist for each of the components; however, the questions on the multivariate structure compose an entirely different study. In fact, the latter type of question can often be considered separately of the marginal distributions of each component. For example, in a bivariate normal distribution the parameter which contains all the bivariate information is the correlation coefficient, \( \rho \), and the values of the means and variances do not influence the dependency relation between the variables fixed by \( \rho \). For another example, all continuous independent bivariate random variables are similar in the sense that, no matter what the marginal distributions of the components are, the nature of the relationship of the two components to each other is determined by the fact of independence.

In Chapter I a set of multivariate statistics is introduced. Its members have the property that their distributions are not dependent on the nature of the marginal distributions of the continuous random variable being sampled, but only on the multivariate structure of the random variable. To make clear these ideas, in Chapter II equivalence classes of \( r \)-dimensional continuous random variables are
defined by placing all the random variables with the same, at least in some sense, dependency relations between the components in one equivalence class. The members of the set of statistics previously introduced are then seen to have their probability distributions invariant for samples drawn from random variables within the same equivalence class. In Chapters III and IV the consistency of these statistics in estimation and testing hypotheses is investigated by means of a special convergence theorem. Chapter V contains some comments on the limiting distributions of these statistics.

In a parallel development a set of two sample statistics is also introduced, and conditions for consistency are given by means of the same convergence theorem.
CHAPTER I

A CLASS OF SAMPLE STATISTICS

Notation

The following notation will be used throughout unless otherwise stated. Small letters, x, y, z, will be points in $\mathbb{E}_I^r$, the r-dimensional Euclidean space. Thus

\[
x = (x_1, x_2, \ldots, x_r)
\]

\[
y = (y_1, y_2, \ldots, y_r)
\]

etc.

We let $X = (X_1, X_2, \ldots, X_r)$ be a random variable defined over $\mathbb{E}_I^r$ with a distribution function

\[
F(x) = P\{X_{\alpha} \leq x_{\alpha}, \beta = 1, \ldots, r\}
\]

where $P$ is the probability measure for $X$. The marginal distribution functions for $X$ will be given by $F$ with the appropriate subscripts, thus

\[
F_{\alpha}(x_{\alpha}) = P\{X_{\alpha} \leq x_{\alpha}\}
\]

\[
F_{\alpha\beta}(x_{\alpha}, x_{\beta}) = P\{X_{\alpha} \leq x_{\alpha}, X_{\beta} \leq x_{\beta}\}
\]

for $\alpha < \beta$, etc., on any number of components up to $r$.

There are $r$ marginal distribution functions on one component, \(\binom{r}{2}\) marginal distribution functions on two components, and, in general, \(\binom{r}{i}\) on $i$ components. Hence, there are $2^r - 1$ distribution functions associated with $X$. The $2^r - 1$ dimensional vector of
functions,

\[ \Phi = (F_1, F_2, \ldots, F_r, F_{12}, F_{13}, \ldots, F_{r-1,r}, F_{123}, \ldots, F) \]

will be called the distribution function vector for \( X \). Thus, for \( x \in \mathbb{R}^r \)

\[ \Phi(x] = (F_1(x_1), F_2(x_2), \ldots, F_r(x_r), F_{12}(x_1x_2), \ldots, F(x)) \]

is the distribution function vector evaluated at \( x \) and is an ordered set of \( 2^r - 1 \) numbers between 0 and 1.

A sample of size \( n \) from \( X \) is a set of \( n \) points \( x^{(k)} \in \mathbb{R}^r, \ k = 1, \ldots, n \) where

\[ x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \ldots, x_r^{(k)}). \]

The sample distribution function for this sample is

\[ F^{(n)}(x) = \frac{1}{n} \left\{ \text{number of } x^{(k)} \text{ such that } x_\lambda^{(k)} \leq x_\lambda \right\} \]

\( \lambda = 1, \ldots, r \} . \)

This may be written as

\[ F^{(n)}(x) = \frac{1}{n} \sum_{k=1}^{n} \prod_{\lambda=1}^{r} c(x_\lambda - x_\lambda^{(k)}) \]

when

\[ c(a) = \begin{cases} 0 & a < 0 \\ 1 & a \geq 0 \end{cases} \]

The marginal sample distribution functions on one component are

\[ F_{\lambda}^{(n)}(x_\lambda) = \frac{1}{n} \left\{ \text{number of } x^{(k)} \text{ such that } x_\lambda^{(k)} \leq x_\lambda \right\} \]
for $k = 1, 2, \ldots, r$; The sample marginal distribution functions for two components are

$$F_{\alpha \beta}^{(n)}(x, x) = \frac{1}{n} \left\{ \text{number of } x^{(k)} \exists x^{(k)} < x, x^{(k)} < x \right\}$$

(5)

and in a similar way the sample marginal distribution functions for any number of components up to $r$ will be written.

The sample distribution function vector for the sample is the $2^r - 1$ dimensional vector of functions

(6)

$$\mathbf{F}^{(n)} = \left( F_1^{(n)}, F_2^{(n)}, \ldots, F_r^{(n)}, F_{12}^{(n)}, \ldots, F^{(n)} \right)$$

and

(7)

$$\mathbf{F}^{(n)}[x] = \left( F_1^{(n)}(x), F_2^{(n)}(x), \ldots, F_r^{(n)}(x), F_{12}^{(n)}(x, x), \ldots, F^{(n)}(x) \right).$$

The Class, $\mathcal{I}$, of Sample Statistics

For a sample of size $n$ from a random variable $X$ a class of sample statistics may be defined as follows:

Definition: The class $\mathcal{I}$ consists of all sample statistics, $S_n$, which can be written in the form

(8)

$$S_n = \int_{E^r} I_n(x) dF^{(n)}(x)$$
where there is a real function, \( Q(t_1, t_2, \ldots, t_{2^r-1}) \), defined for \( 0 \leq t_A \leq 1, \ A = 1, 2, \ldots, 2^r - 1 \), such that

\[
I_n(x) = Q(x^{(n)}[x]),
\]

and the above Lesbesgue-Stieltjes integral exists.

Since \( I_n(x) \) may be discontinuous at the sample points, \( \{x^{(k)}\} \), the Riemann-Stieltjes integral will not exist, and consequently, the Lesbesgue-Stieltjes integral must be used. Notice also, that any \( S_n \in \mathcal{G} \) can be written as a sum

\[
S_n = \int_{E^r} I_n \, dF(n) = \frac{1}{n} \sum_{k=1}^{n} I_n(x^{(k)})
\]

since the measure on \( E^r \) defined by \( F(n)(x) \) is discrete with a measure of \( \frac{1}{n} \) at each of the sample points, \( x^{(k)} \).

\( \mathcal{G} \) is closed under linear transformations, that is, if \( S_n \in \mathcal{G} \), then for \( S_n \in \mathcal{G} \) and constants, \( a, b \), the statistic \( aS_n + b \in \mathcal{G} \). This follows because if

\[
S_n = \int_{E^r} I_n \, dF(n)
\]

then

\[
aS_n + b = a \int_{E^r} I_n \, dF(n) + b
\]

\[
= \int_{E^r} (aI_n + b) \, dF(n),
\]

and \( aS_n + b \) qualifies as a member of \( \mathcal{G} \).
Examples of Statistics in \( \mathcal{D} \)

Two examples of particular members of \( \mathcal{D} \) are given; these examples will be referred to later in the paper.

Example 1: Let \( r = 2 \) with \( F(x) \) a continuous distribution function.

If

\[
s(a) = \begin{cases} 
-1, & a < 0 \\
1, & a > 0,
\end{cases}
\]

then for a sample, \( x^{(1)}, x^{(2)}, \ldots, x^{(n)} \), the quantity

\[
(11) \quad \tau = \frac{1}{n(n-1)} \sum_{k=1}^{n} \sum_{i \neq j} s(x^{(k)}_1 - x^{(j)}_1)s(x^{(k)}_2 - x^{(j)}_2)
\]

is called the difference sign covariance, or Kendall's \( \tau \) [1].

This statistic, \( \tau \), is a member of \( \mathcal{D} \), in fact, there exists \( \tau \in \mathcal{D} \) such that

\[
\tau = \frac{\ln n}{n-1} \tau_s - \frac{n+3}{n-1}
\]

for \( n > 1 \). We find \( \tau \in \mathcal{D} \) in the following way. Since

\[
c(a) = \begin{cases} 
0, & a < 0 \\
1, & a > 0
\end{cases}
\]

then

\[
s(a) = 2c(a) - 1
\]

Thus, from equation (11)

\[
(12) \quad \tau = \frac{1}{n(n-1)} \sum_{k=1}^{n} \sum_{i \neq j} \left[ \left(2c(x^{(k)}_1 - x^{(j)}_1) - 1\right) \left(2c(x^{(k)}_2 - x^{(j)}_2) - 1\right) \right]
\]
\[
\frac{1}{n(n-1)} \sum_{k=1}^{n} \sum_{i=1, i \neq k}^{n} [4c(x^{(k)}_1 - x^{(j)}_1)c(x^{(k)}_2 - x^{(j)}_2)
- 2c(x^{(k)}_1 - x^{(j)}_1) - 2c(x^{(k)}_2 - x^{(j)}_2) - 1]
\]

but, for \( \Delta = 1, 2 \), using equation (4)

\[
\sum_{k=1}^{n} \sum_{i=1, i \neq k}^{n} c(x^{(k)}_1 - x^{(j)}_1) c(x^{(k)}_2 - x^{(j)}_2) = \sum_{k=1}^{n} \sum_{i=1}^{n} c(x^{(k)}_1 - x^{(j)}_1) - \sum_{k=1}^{n} c(x^{(k)}_1 - x^{(k)}_1)
\]

\[
= \sum_{k=1}^{n} n F^{(n)}(x^{(k)}_1) - n
\]

\[
= \frac{n(n-1)}{2} - n
\]

Moreover, using equation (5)

\[
\sum_{k=1}^{n} \sum_{i=1, i \neq k}^{n} c(x^{(k)}_1 - x^{(j)}_1) c(x^{(k)}_2 - x^{(j)}_2)
- \sum_{k=1}^{n} \sum_{i=1}^{n} c(x^{(k)}_1 - x^{(j)}_1) c(x^{(k)}_2 - x^{(j)}_2)
- \sum_{k=1}^{n} c(x^{(k)}_1 - x^{(k)}_1) c(x^{(k)}_2 - x^{(k)}_2)
\]

\[
= \sum_{k=1}^{n} nF^{(n)}(x^{(k)}_1) - n
\]

Hence, in equation (12)
\[ \tau = \frac{1}{n(n-1)} \left[ \ln \sum_{k=1}^{n} F^{(n)}(x^{(k)}) - \ln n - 2n(n-1) \cdot n(n-1) \right] \]

\[ = \frac{\ln}{n-1} \left[ \sum_{k=1}^{n} F^{(n)}(x^{(k)}) \right] - \frac{n+3}{n-1} \]

\[ = \frac{\ln}{n-1} \int_{E^2} F^{(n)}(x) dF^{(n)}(x) - \frac{n+3}{n-1} \]

Thus, by letting

(13) \[ \tau_s = \frac{1}{n-1} \int_{E^2} F^{(n)}(x) dF^{(n)} \]

we have \( \tau_s \in \mathcal{I} \) and

(14) \[ \tau = \frac{\ln}{n-1} \tau_s - \frac{n+3}{n-1} \]

Since \( \tau \) is a linear function of \( \tau_s \), it follows that \( \tau \in \mathcal{I} \).

**Example 2:**

For a sample from a random variable in 2 dimensions with a continuous distribution function, \( F(x) \), Spearman's rank correlation coefficient [1] is given by

\[ \rho = \frac{\sum_{k=1}^{n} (r^{(k)}_1 - \bar{r}_1)(r^{(k)}_2 - \bar{r}_2)}{\sqrt{\left[ \sum_{k=1}^{n} (r^{(k)}_1 - \bar{r}_1)^2 \right] \left[ \sum_{k=1}^{n} (r^{(k)}_2 - \bar{r}_2)^2 \right]}} \]

where \( r^{(k)}_d \) is the rank of \( x^{(k)}_d \) and \( \bar{r}_1 = \bar{r}_2 = \frac{\sum_{k=1}^{n} r^{(k)}_d}{n} = \frac{n+1}{2} \)
It is also commonly written as

\[(15) \quad \rho = 1 - \frac{6 \sum_{k=1}^{n} \alpha_k^2}{n(n^2 - 1)} \]

where \( \alpha_k = r_1(k) - r_2(k) \).

We now show that there exists a \( \rho_s \in \mathcal{B} \) such that

\[(16) \quad \rho = \frac{12n^2}{n^2 + 1} \rho_s - \frac{3n + 1}{n - 1} \]

Begin by noting that \( r_\alpha(k) = nF(n)(x_\alpha(k)) \) for \( \alpha = 1, 2 \). Then

\[
\sum_{k=1}^{n} \alpha_k^2 = \sum_{k=1}^{n} \left[ (r_1(k))^2 - 2r_1(k)r_2(k) + (r_2(k))^2 \right]
\]

\[
= 2 \sum_{k=1}^{n} k^2 - 2 \sum_{k=1}^{n} r_1(k)r_2(k)
\]

\[
= n(n - 1)(2n + 1) - 2n^2 \left[ \sum_{k=1}^{n} F_1(n)(x_1(k))F_2(n)(x_2(k)) \right]
\]

Letting

\[
\rho_s = \int_{\mathcal{B}} F_1(n)F_2(n)dF(n)
\]

we have \( \rho_s \in \mathcal{B} \) and

\[
\sum_{k=1}^{n} \alpha_k^2 = \frac{n(n + 1)(2n + 1)}{3} - 2n^3 \rho_s
\]
A Class, $\mathcal{S}^2$, of Two Sample Statistics

A class of statistics similar to $\mathcal{S}$ but for samples arising from two random variables can be defined. Let $x^{(1)}, x^{(2)}, \ldots, x^{(n)}$ be a sample of size $n$ taken from $X$ and $y^{(1)}, y^{(2)}, \ldots, y^{(m)}$ be a sample of size $m$ from a random variable $Y$. Let $G(y), y \in \mathbb{R}$ be the distribution function of $Y$. For the marginal distribution functions and the sample marginal distribution functions of $Y$ the same super and subscript notation will be used with "$G$" as is used with "$F$" for $X$. Moreover, $\mathcal{F}[x]$ and $\mathcal{F}^{(m)}[x]$ will be the distribution function vector and the sample distribution function vector, respectively, for $Y$.

Definition: The set $\mathcal{S}^2$ consists of all two sample statistics, $S_{mn}$, which can be written in the form

\begin{equation}
S_{mn} = \int_{\mathbb{R}^r} I_m(x) d\mathcal{F}^{(n)}(x)
\end{equation}

where there exists a real function, $Q(t_1, t_2, \ldots, t_{2r-1})$, defined for $0 \leq t_\alpha \leq 1, \alpha = 1, 2, \ldots, 2r - 1$, such that

\[ I_m(x) = Q[\mathcal{F}^{(m)}[x]] \]

and the above Lesbesgue–Stieltjes integral exists.
Notice that any $S_{mn} \in \mathcal{J}^2$ can be written as a sum

\[(18) \quad S_{mn} = \int_{\mathbb{R}} I_m dF(n) = \frac{1}{n} \sum_{k=1}^{n} I_m (x(k)) \]

\[= \frac{1}{n} \sum_{k=1}^{n} Q(\mathcal{H}(m)[x(k)]) \]

Also, any linear function of a member of $\mathcal{J}^2$ is again a member of $\mathcal{J}^2$, for if $a, b$ are constants and $S_{mn} \in \mathcal{J}^2$, then

\[aS_{mn} + b = a \int_{\mathbb{R}} I_m dF(n) + b \]

\[= \int_{\mathbb{R}} (aI_m + b) dF(n) \]

**Example 3:** Let $X$ and $Y$ be one dimensional continuous random variables. Then the statistic

\[U_s = \int_{\mathbb{R}} G^{(m)}(x) dF(n)(x) \]

is in $\mathcal{J}^2$. This statistic is essentially the Mann-Whitney $U$ statistic \(^{[2]}\) which can be written

\[U = \sum_{j=1}^{n} \left\{ \text{number of } y^{(k)} \text{ such that } y^{(k)} \leq x^{(j)} \right\} \]

or

\[U = \sum_{j=1}^{n} \sum_{k=1}^{n} c(x^{(j)} - y^{(k)}) \]
where, remember,

\[
    c(a) = \begin{cases} 
    0, & a < 0 \\
    1, & a \geq 0.
    \end{cases}
\]

However, by equation (4)

\[
    \sum_{k=1}^{m} c(x(j) - y(k)) = mG(m)x(j),
\]

hence

\[
    U = m \sum_{j=1}^{m} G(m)x(j)
\]

\[
    = nm \sum_{j=1}^{m} G(m) d_F(n)
\]

\[
    = nmU_g
\]

Thus, \( U \) is a linear function of \( U_g \in \mathbb{L}^2 \), and hence \( U \in \mathbb{L}^2 \).
CHAPTER II

EQUIVALENCE CLASSES OF RANDOM VARIABLES IN $E^r$

Introduction

In order to describe how statistics in $\mathcal{D}$ are useful in investigating the multivariate structure of a random variable in such a way as to be free of the nature of the marginal distributions we define equivalence classes of continuous random variables. These equivalence classes are defined so that members of the same class may differ in their marginal distributions in one dimension, but not, at least in some respects, in the dependence of any set of components on any other set. Then it is shown that the distribution of any statistic in $\mathcal{D}$ is the same for all random variables within an equivalence class. Cumulative expected values are introduced as a useful concept in describing the use of statistics of $\mathcal{D}$ in estimation and testing hypotheses on differences between equivalence classes.

Equivalence Classes

Denote by $C^r$ the $r$-dimensional unit cube in $E^r$, that is,

$$C^r = \{x \mid x \in E^r, \ 0 \leq x_r \leq 1, \ q = 1, 2, \ldots, r\}.$$

An equivalence relation is defined as follows:

Definition: Let $X$ and $X^*$ be continuous random variables over $E^r$ with distribution functions $F$ and $F^*$, and distribution function vectors $\mathcal{F}$ and $\mathcal{F}^*$ respectively. If, when $u = (u_1, u_2, \ldots, u_r)$,
\[ H(u) = P\{F_u(X_u) \leq u, \ u = 1,2,\ldots,r \} \]
\[ H^*(u) = P\{F_u^{*}(X_u) \leq u, \ u = 1,2,\ldots,r \} \]

then we say \( X \sim X^* \) (\( X \) is equivalent to \( X^* \)) whenever \( H(u) = H^*(u) \) for all \( u \in C^r \).

The relation just defined is clearly an equivalence relation for it follows immediately from the definition that

i) if \( X \sim X^* \), then \( X^* \sim X \),

ii) \( X \sim X \),

iii) if \( X \sim X^* \), \( X^* \sim X^{**} \), then \( X \sim X^{**} \).

Equivalence classes of continuous random variable in \( E^r \) are thus defined. Notice that every continuous random variable is equivalent to a random variable on \( C^r \) whose marginals on one dimension are uniform 0 to 1, for, if \( X \) is continuous, then \( X \sim U \), where

\[ U = (F_1(X_1), F_2(X_2), \ldots, F_r(X_r)) \]

and \( U \) is clearly a random variable on \( C^r \) with uniform marginals. It is, in fact, in terms of this member random variable on \( C^r \) that the equivalence class is defined, for the definition says that \( X \sim X^* \) whenever the distribution of \( U^* \) is the same as the distribution of \( U^* = (F_1^{*}(X_1^*), F_2^{*}(X_2^*), \ldots, F_r^{*}(X_r^*)) \).
Properties of Equivalence Classes

In order to make clear the significance of classifying random variables into equivalence classes and to show the connection of these equivalence classes to statistics in \( \mathcal{D} \), we now list a few simple properties associated with the equivalence classes. In the following presentation we always have \( X \) and \( X^* \) as random variables in \( \mathbb{R}^r \) with distribution functions \( F(x) \) and \( F^*(x) \) and distribution function vectors \( \mathbf{F} \) and \( \mathbf{F}^* \), respectively. Moreover, let

\[
U = (F_1(X_1), F_2(X_2), \ldots, F_r(X_r))
\]

have a distribution function \( h(u) \) and a distribution function vector \( \mathbf{X}[u] \), and

\[
U^* = (F_1^*(X_1^*), F_2^*(X_2^*), \ldots, F_r^*(X_r^*))
\]

have a distribution function \( h^*(u) \) and a distribution function vector \( \mathbf{X}^*[u] \).

If all of the random variables under consideration had strictly increasing marginal distribution functions on each component, then the following presentation would be simpler. But, in order to include all continuous random variables we must take into account the fact that some of the marginal distribution functions on one dimension may have intervals in which they are constant. Treating these intervals in terms of distribution functions does require some extra attention to details, but nothing essential is added for such intervals all
have a probability of 0. The following lemma dispenses with some of
the difficulty.

**Lemma 2.1:** For any $x \in E^r$, $F(x) = P\{F_d(x_d) \leq F_d(x_d), \ \forall = 1, \cdots, r\}$. This may be restated as: For any $x \in E^r$, and $u = (F_1(x_1), \cdots, F_r(x_r))$, it is true that $F(x) = H(u)$, and, consequently, $\mathcal{G}[x] = \mathcal{M}[u]$.

**Proof:** First $x_\perp \leq x_\perp$ implies $F_d(x_\perp) \leq F_d(x_\perp)$, thus

$$F(x) \leq P\{F_d(x_\perp) \leq F_d(x_\perp), \ \forall = 1, \cdots, r\}$$

Moreover,

$$P\{F_d(x_\perp) \leq F_d(x_\perp), \ \forall = 1, \cdots, r\}$$

$$= P\{F_d(x_\perp) \leq F_d(x_\perp), X_\perp \leq x_\perp, \ \forall = 1, \cdots, r\}$$

$$+ P\{F_d(x_\perp) \leq F_d(x_\perp), \ \forall = 1, \cdots, r, \ \text{and for some } \beta, X_\beta > x_\perp\}.$$ 

Now, $P\{F_d(x_\perp) \leq F_d(x_\perp), X_\perp \leq x_\perp, \ \forall = 1, \cdots, r\} \leq F(x)$, and also, if both $F_\beta(x_\perp) \leq F_\beta(x_\perp)$ and $X_\beta > x_\perp$, then $F_\beta(x_\perp) = F_\beta(x_\perp)$ for $F_\beta$ is monotone increasing. Thus

$$P\{F_d(x_\perp) \leq F_d(x_\perp), \ \forall = 1, \cdots, r, \ \text{and for some } \beta, X_\beta > x_\perp\}$$

$$\leq P\{F_\beta(x_\perp) = F_\beta(x_\perp)\} = 0$$

for $F_\beta(x_\perp)$ is uniformly distributed 0 to 1. Hence,

$$P\{F_d(x_\perp) \leq F_d(x_\perp), \ \forall = 1, \cdots, r\} \leq F(x).$$
Property 1: If $X \sim X^*$, then $\mathcal{H}(u) = \mathcal{H}^*(u)$ for all $u \in C^r$.

Proof: If $X \sim X^*$, then by the definition $H(u) = H^*(u)$, and, consequently, all the marginals on any number of dimensions are the same. Therefore, $\mathcal{H}(u) = \mathcal{H}^*(u)$.

Definition: $x$ and $x^*$ are co-image points for $X$ and $X^*$ if $F_\alpha(x_\alpha) = F_\alpha^*(x_\alpha^*)$, $\alpha = 1, \ldots, r$.

Property 2: If $x$ and $x^*$ are co-image points for $X$ and $X^*$, and $X \sim X^*$, then $\mathcal{F}(x) = \mathcal{F}^*(x^*)$.

Proof: Let $u_\alpha = F_\alpha(x_\alpha)$, and $u_\alpha^* = F_\alpha^*(x_\alpha^*)$, then by lemma 2.1

$$\mathcal{F}(x) = \mathcal{H}(u)$$

$$\mathcal{F}^*(x^*) = \mathcal{H}^*(u^*)$$

But $x$ and $x^*$ are co-image points, so $u = u^*$. Using this and the fact that $X \sim X^*$ we have

$$\mathcal{H}^*[u^*] = \mathcal{H}^*[u] = \mathcal{H}[u]$$

Hence, $\mathcal{F}(x) = \mathcal{F}^*(x^*)$

Property 2 implies that within an equivalence class the distribution function is determined completely by the set of one dimensional marginal distribution functions.
**Definition:** Sets \( e \subseteq E \) and \( e^* \subseteq E^* \) are co-image sets for \( X \) and \( X^* \) if, when

\[
    c = \{ u \mid u \in c^r, u_\lambda = F_\lambda(x_\lambda), \lambda = 1, \ldots, r, \text{ and } x \in e \} \\
    c^* = \{ u \mid u \in c^r, u_\lambda = F_\lambda^*(x_\lambda^*), \lambda = 1, \ldots, r, \text{ and } x^* \in e^* \}.
\]

It is true that \( c = c^* \).

**Property 3:** If \( e \) and \( e^* \) are co-image sets for \( X \) and \( X^* \), and \( X \sim X^* \), then \( P\{X \in e\} = P\{X^* \in e^*\} \).

**Proof:** Take \( c \) and \( c^* \) as given in the preceding definition and

\[
    u(x) = (F_1(x_1), F_2(x_2), \ldots, F_r(x_r)) \\
    u^*(x^*) = (F_1^*(x^*_1), F_2^*(x^*_2), \ldots, F_r^*(x^*_r))
\]

Then, using lemma 2.1,

\[
P\{X \in e\} = \int_{e^*} dF(x) = \int_{e^*} dH(u(x)) = \int_{e^*} dH(u) = P\{u(X) \in c\} .
\]

But \( e \) and \( e^* \) are co-image sets, so \( c = c^* \), and since \( X \sim X^* \), \( u(X) \) and \( u^*(X^*) \) are identically distributed, and thus

\[
P\{u(X) \in c\} = P\{u(X) \in c^*\} = P\{u^*(X^*) \in c^*\}.
\]

Moreover, as above, \( P\{u^*(X^*) \in c^*\} = P\{X^* \in e^*\} \), hence, it follows

\[
P\{X \in e\} = P\{X^* \in e^*\}.
\]
Property 4: If $e_1$ and $e_1^*$ are co-image sets for $X$ and $X^*$, and $e_2$, $e_2^*$ are co-image sets for $X$ and $X^*$, with $P\{X \in e_2\} > 0$, and if $X \sim X^*$, then

$$P\{X \in e_1 | X \in e_2\} = P\{X^* \in e_1^* | X^* \in e_2^*\}.$$

Proof: Since $e_1$, $e_1^*$, and $e_2$, $e_2^*$, are both pairs of co-image sets, then $e_1 \cap e_2$ and $e_1^* \cap e_2^*$ are also co-image sets. Using property 3

$$P\{X \in e_1 | X \in e_2\} = \frac{P\{X \in e_1 \cap e_2\}}{P\{X \in e_2\}}$$

and

$$= \frac{P\{X^* \in e_1^* \cap e_2^*\}}{P\{X^* \in e_2^*\}}$$

$$= P\{X^* \in e_1^* | X^* \in e_2^*\}.$$

The relationship of random variables within an equivalence class to each other can be viewed in terms of transformations. In the definition of the equivalence relation a random variable, $X$, is transformed into $U$, $X \rightarrow U$, by means of monotone transformations on each of the components separately. If another random variable, $X^*$, by similar transformation also goes to $U$, $X^* \rightarrow U$, then it was said that $X \sim X^*$. The inverse transformation, $U \rightarrow X^*$, has the same nature, and thus $X \rightarrow U \rightarrow X^*$, and $X^*$ is obtained from $X$ by monotone transformations on each of the components separately.

However, since the marginal distribution functions on one dimension of $X$ and of $X^*$ may have intervals of constant value, the
inverse transformation \( U \rightarrow X^* \) may be one to many, and, hence, \( X \rightarrow X^* \) would be many to many. If it were not for this awkwardness the above properties of equivalence classes could be more easily given in terms of the transformation \( X \rightarrow X^* \).

Properties 3 and 4 are significant for they provide motivation for dealing with the equivalence classes. Property 3 states that if \( X \sim X^* \), then the transformation \( X \rightarrow X^* \) is probability preserving. Property 4 shows that it is also conditional probability preserving as long as the conditioning set has positive probability. Thus, within an equivalence class the marginal distribution functions on one dimension determine not only the distribution function and all the marginal distribution functions on any number of dimensions, but also all the conditional distribution functions for any number of dimensions when the conditioning is by a set of positive probability.

It is in this sense that we say the multivariate structure of a random variable does not differ within an equivalence class.

Many multivariate statistics, especially rank order statistics, are unaffected by differences within an equivalence class, and in dealing with them equivalence classes, not individual random variables, are used. For \( r \geq 2 \) the set of all \( r \)-dimensional random variables with mutually independent components comprise an equivalence class. A test for independence would ideally discriminate between this class and all other classes.

The question of equivalence classes does not arise in one dimension because all continuous one dimensional random variables

\[\text{21}\]
belong to the same class. It is precisely for this reason, for instance, that the Kolmogorov-Smirnov statistic,

\[ D_n = \sup_x |F_n(x) - F(x)|, \]

has a distribution that does not depend on \( F \) as long as \( F \) is continuous. The value of \( D_n \) and its distribution are clearly invariant within an equivalence class. However, for \( r > 1 \) there are many classes and the natural extension of the Kolmogorov-Smirnov statistic,

\[ D(r) = \sup_{x \in \mathbb{R}_r} |F_n(x) - F(x)| \]

is not distributed independently of \( F \). It is, though, independent of \( F \) within an equivalence class.

It should be noted in this connection that not all measures and parameters of multivariate information are invariant within equivalence classes. For instance, for a bivariate random variable the correlation coefficient is not the same for all members of a class.

**Invariance of Statistics in \( \mathcal{S} \)**

It is now shown that any statistic of the class \( \mathcal{S} \) has the property that its distribution is the same for samples from any member of a given equivalence class.

The same notation as above is used throughout this section with \( \mathcal{F}(n) \) and \( \mathcal{F}^*(n) \) as the sample distribution function vectors for samples of size \( n \) from \( X \) and \( X^* \), respectively, and \( \mathcal{W}(n) \) and \( \mathcal{W}^*(n) \) as the sample distribution function vectors of \( U \) and \( U^* \), respectively.
Note that when \( u_\lambda = F_\lambda(x_\lambda), \lambda = 1, \ldots, r, \) and \( u_\lambda^* = F_\lambda^*(x_\lambda^*), \lambda = 1, \ldots, r, \) then
\[
\gamma^{(n)}(u) = \gamma^{(n)}(x) \quad \text{and} \quad \gamma^{*(n)}(u^*) = \gamma^{*(n)}(x^*).
\]

Suppose that \( S_n \) is a statistic in \( \mathcal{G} \) so that
\[
S_n = \iint_{E^r} Q(\gamma^{(n)}(x)) d\mathcal{F}(n)(x)
\]
and
\[
S_n^* = \iint_{E^r} Q(\gamma^{*(n)}(x)) d\mathcal{F}^*(n)(x)
\]
with
\[
\varphi(s) = P\{S_n \leq s\}
\]
\[
\varphi^*(s) = P\{S_n^* \leq s\}.
\]

**Theorem 2.1:** If \( X \sim X^* \), then for all \( s \in E^1 \), \( \varphi(s) = \varphi^*(s) \).

**Proof:** Consider the statistic
\[
T_n = \iint_{E^r} Q(\gamma^{(n)}(u)) d\mathcal{H}(n)(u)
\]
where \( u = u(x) \) is given by \( u_\lambda = F_\lambda(x_\lambda), \lambda = 1, \ldots, r. \) Then
\[
T_n = \iint_{E^r} Q(\gamma^{(n)}(u(x))) d\mathcal{H}(n)(u(x))
\]
\[
= \iint_{E^r} Q(\gamma^{*(n)}(x)) d\mathcal{F}(n)(x)
\]
\[
= S_n
\]
In the same way

\[ T_n^* = \int_{C^r} Q(\mathcal{U}^*(n)[u]) d\mathcal{H}^*(n) = S_n^*. \]

But since \( X \sim X^* \), then \( U \) and \( U^* \) are identically distributed and \( T_n \) and \( T_n^* \) are just the same statistic for samples of the same size from identically distributed random variables. Therefore, \( T_n \) and \( T_n^* \) have the same distribution and

\[ \phi(s) = P(S_n \leq s) = P(T_n \leq s) = P(T_n^* \leq s) = P(S_n^* \leq s) = \phi^*(s). \]

**Cumulative Expected Values**

Just as random variables are characterized by expected values, so also equivalence classes are characterized by a certain type of expected values which are invariant within a class. These expected values, called cumulative expected values, are now defined and shown to be the same for all members of an equivalence class.

**Definition:** Let \( Q(t_1, t_2, \ldots, t_{2^r - 1}) \) be a real function defined on \( C^r \). Let \( I(x) = Q(\mathcal{F}[x]) \), then \( E(I(X)) \), if it exists, is a cumulative expected value for \( X \).

**Theorem 2.2:** If \( E(Q(\mathcal{F}[X])) \) exists and \( X \sim X^* \), then

\[ E(Q(\mathcal{F}[X])) = E(Q(\mathcal{F}^*[X^*])). \]

**Proof:** Let \( u = u(x) \) and \( u^*(x^*) \) be given as above. Then \( U = u(X) \) and \( U^* = u^*(X^*) \), and from lemma 2.1 and property 1
\[
E(Q(F'[X'])) = E(Q(\Psi[u(X)])) \\
= E(Q(\Psi[u])) \\
= E(Q(\Psi*[u*])) \\
= E(Q(\Psi*[u*(X^*)])) \\
= E(Q(F^*[X^*])).
\]

A few items to be noted in connection with cumulative expected values which will be useful later are listed here. For all of these \(X\) is continuous in \(E^r\).

1) For \(r = 1\), \(E(F(X)) = \frac{1}{2}\)

2) For \(r \geq 2\), \(0 \leq E(F(X)) \leq \frac{1}{2}\), and there exist random variables, \(X'\) and \(X''\), such that \(E(F'(X')) = 0\) and \(E(F''(X'')) = \frac{1}{2}\).

Proof: Since \(F(x) \geq 0\), \(E(F(X)) \geq 0\). Also
\[
F(x) \leq F'_1(x_1) \text{ any } x_1,
\]

thus
\[
E(F(X)) \leq E(F'_1(X'_1)) = \frac{1}{2}.
\]

For \(r = 2\), a random variable \(X'\) such that \(E(F'(X')) = 0\) is given by \(X' = (X'_1, X'_2)\) where \(F'_1(x'_1) = x'_1, 0 \leq x'_1 \leq 1\) and \(X'_1 = 1 - X'_2\).

This random variable has all its probability distributed on the diagonal of the unit square, that is, on the straight line from the point \((0,1)\) to the point \((1,0)\). However, at every point on that diagonal \(F'(x) = 0\), and hence,
\[
E(F'(X')) = 0.
\]
For \( r = 2 \) a random variable \( X^r \) such that \( E(F^r(X^r)) = \frac{1}{2} \) is given by \( X^r = (X_1^r, X_2^r) \) where \( F_1(x_1) = x_1, 0 \leq x_1 \leq 1, \) and \( X_1^r = X_2^r. \) This one has all its probability on the other diagonal. On that diagonal \( F^r(x) = F_1^r(x_1), \) hence

\[
E(F^r(X^r)) = \int_0^1 \int_0^1 dF^r = \int_0^1 dF^r = \frac{1}{2}
\]

Similar examples can be constructed for \( r > 2 \) to get these extreme values for \( E(F). \)

3) For \( r > 1 \) if \( X_1, X_2, \ldots, X_r \) are mutually independent then

\[
E(F(X)) = \frac{1}{2r}
\]

since

\[
E(F(X)) = E\left(\prod_{j=1}^r F_j(X_j)\right) = \prod_{j=1}^r E(F_j(X_j)) = \frac{1}{2r}
\]

4) For \( r = 2, \ \frac{1}{6} \leq E(F_1(X_1)F_2(X_2)) \leq \frac{1}{3}. \)

Proof: The random variable \( (F_1(X_1), F_2(X_2)) \) is just a random variable on \( \mathbb{R}^2 \) with uniform marginals. If \( \rho_{F_1F_2} \) is the correlation coefficient of \( F_1(X_1) \) and \( F_2(X_2) \) then

\[
\rho_{F_1F_2} = \frac{E(F_1F_2) - E(F_1)E(F_2)}{\sigma_{F_1} \sigma_{F_2}},
\]

but \( E(F_1) = E(F_2) = \frac{1}{2}, \) and \( \sigma_{F_1} = \sigma_{F_2} = \sqrt{\frac{1}{12}} \) thus

\[
\rho_{F_1F_2} = \frac{E(F_1F_2) - \frac{1}{2}}{\frac{1}{12}}.
\]
But \(-1 \leq \rho_{F_1F_2} \leq 1\), hence it follows that \(\frac{1}{5} \leq E(F_1F_2) \leq \frac{1}{3}\). The quantity \(\rho_{F_1F_2}\) is sometimes called the grade correlation coefficient for \(X_1\) and \(X_2\). [1]
A CONVERGENCE THEOREM

Let the set of all continuous random variables defined over $\mathbb{F}^r$ be designated by $\mathbb{X}^r$. Suppose $T_n(X)$ is a sample statistic for samples of size $n$ from $X \in \mathbb{X}^r$, and $T(X)$ is a parameter associated with $X$. Then for some $\omega \in \mathbb{X}^r$ we say

$$\text{plim}_{n \to \infty} T_n(X) = T(X)$$

uniformly for $X \in \omega$ if, for any $\epsilon, \gamma > 0$ there exists an $N$ such that $n > N$ implies

$$P\left\{ |T_n(X) - T(X)| > \epsilon \right\} < \gamma$$

for all $X \in \omega$.

Let $y^{(1)}, y^{(2)}, \ldots, y^{(m)}$, be a sample from a random variable, $Y$, in $\mathbb{F}^r$.

**Definition:** A function, $K(y), y \in \mathbb{F}^r$, is a Borel simple sample function for $Y$ if the $m$ sample values for $Y$ are parameters in the definition of $K(y)$, and if there exist Borel sets, $e_1, e_2, \ldots, e_p$, a finite partition of $\mathbb{F}^r$, and constants, $h_1, h_2, \ldots, h_p$, such that

$$K(y) = \frac{1}{h_i} \text{ for } y \in e_i.$$ 

The value of a Borel simple sample function at any point is a function of the sample and, consequently, a random variable.

A theorem is now proved which can be used to show connections between cumulative expected values of a random variable and statistics from $\mathbb{F}$ for samples from that random variable. The theorem will also
be useful in connection with statistics from $\mathcal{L}^2$.

**Convergence Theorem:**

H-1: \( \{F^{(r)}(x)\} \) is a sequence of sample distribution functions for samples of size \( n \) drawn from a random variable, \( X \in \mathcal{X}^r \), with a distribution function, \( F(x) \).

H-2: \( \{I_m(y)\} \) is a sequence of Borel simple sample functions for samples of size \( m \) from a random variable, \( Y \in \omega \), such that

1) there exists a constant, \( B > 0 \), such that for all \( m \), all \( y \in \mathcal{E}^r \), and all \( Y \in \omega \), \( |I_m(y)| < B \) with probability 1,

ii) there exists a function associated with \( Y \), \( I_Y(y) \), such that

\[
\lim_{n \to \infty} \sup_{y \in \mathcal{E}^r} |I_m(y) - I(y)| = 0
\]

uniformly for \( Y \in \omega \).

Then

\[
\lim_{n \to \infty} \int_{\mathcal{E}^r} I_m(x)dF^n(x) = \int_{\mathcal{E}^r} I_Y(x)dF(x)
\]

uniformly for \( X \in \mathcal{X}^r \), \( Y \in \omega \).

**Proof:** We must show that for any \( \varepsilon, \gamma > 0 \) there exists an \( N \) such that for all \( X \in \mathcal{X}^r \), \( Y \in \omega \), \( m, n > N \) implies

\[
P \left\{ \left| \int_{\mathcal{E}^r} I_m dF^n - \int_{\mathcal{E}^r} I_Y dF \right| > \varepsilon \right\} < \gamma
\]
Now (all integrals in the proof are over \( E^r \) unless otherwise stated),

letting

\[
\begin{align*}
\Delta &= \mid \int_{I} dF(n) - \int \text{Id}F \mid \\
\Delta_1 &= \mid \int_{I} dF(n) - \int \text{Id}F \mid \\
\Delta_2 &= \mid \int \text{Id}F - \int \text{Id}F \mid
\end{align*}
\]

since \( \Delta \leq \Delta_1 + \Delta_2 \), and the event \( \{ \Delta_1 \cdot \Delta_2 > \epsilon \} \) implies at least one of the events \( \{ \Delta_1 > \frac{\epsilon}{2} \} \) or \( \{ \Delta_2 > \frac{\epsilon}{2} \} \), we have

\[
P \{ \Delta > \epsilon \} \leq P \{ \Delta_1 \cdot \Delta_2 > \epsilon \} \leq P \left\{ \Delta_1 > \frac{\epsilon}{2} \text{ or } \Delta_2 > \frac{\epsilon}{2} \right\} \leq P \left\{ \Delta_1 > \frac{\epsilon}{2} \right\} + P \left\{ \Delta_2 > \frac{\epsilon}{2} \right\}
\]

Hence, it is sufficient to show that there exists an \( N \) such that \( m, n > N \) implies for all \( X \in \mathcal{X}^r, Y \in \omega \),

\[
P \left\{ \Delta_i > \frac{\epsilon}{2} \right\} < \frac{1}{2} \quad i = 1, 2
\]

We work first with \( \Delta_1 \). Since the \( I_m \) are Borel simple sample functions, for each \( m \) there exists a partition, \( \{ e_{mi} \}, i = 1, 2, \ldots, k_m \) of \( E^r \),

and a set of constants, \( \{ \gamma_{mi} \}, i = 1, 2, \ldots, k_m \) such that

\[
I_m(x) = \gamma_{mi}
\]

for \( x \in e_{mi} \).
Then
\[ A_1 = \left| \int_{\mathbb{E}} dF(n) - \int_{\mathbb{E}} dF \right| \]
\[ \leq \sum_{i} \left| \int_{\mathbb{E}_i} dF(n) - \int_{\mathbb{E}_i} dF \right| \]
\[ = \sum_i \left| \int_{\mathbb{E}_i} dF(n) - \int_{\mathbb{E}_i} dF \right| + \sum_* \left| \int_{\mathbb{E}_*} dF(n) - \int_{\mathbb{E}_*} dF \right| \]

where \( \sum_1 \) is the sum over all values of \( i \) such that
\[ \int_{\mathbb{E}_i} dF(n) - \int_{\mathbb{E}_i} dF \geq 0 \]

and \( \sum_* \) is the sum over all values of \( i \) such that
\[ \int_{\mathbb{E}_i} dF(n) - \int_{\mathbb{E}_i} dF < 0 \]

Moreover, since \( \left| \int_{\mathbb{E}_i} dF(n) - \int_{\mathbb{E}_i} dF \right| \) is bounded with probability 1,
\[ A_1 < B \left\{ \sum_1 \left| \int_{\mathbb{E}_i} dF(n) - \int_{\mathbb{E}_i} dF \right| + \sum_* \left| \int_{\mathbb{E}_*} dF(n) - \int_{\mathbb{E}_*} dF \right| \right\} \]

and by the choice of summands in \( \sum_1 \) and \( \sum_* \)
\[ A_1 < B \left\{ \left| \int_{\mathbb{U}_1} dF(n) - \int_{\mathbb{U}_1} dF \right| + \left| \int_{\mathbb{U}_*} dF(n) - \int_{\mathbb{U}_*} dF \right| \right\} \]

where \( \mathbb{U}_1 \) is the union over all values of \( i \) used in the sum \( \sum_1 \) and \( \mathbb{U}_* \) is the union over all values of \( i \) used in the sum \( \sum_* \).
Thus, letting

\[ A_{11} = \int_{U'_{e_{mi}}}^{U''_{e_{mi}}} dF(n) - \int_{U'_{e_{mi}}}^{U''_{e_{mi}}} dF | \]

\[ A_{12} = \int_{U'_{e_{mi}}}^{U''_{e_{mi}}} dF(n) - \int_{U'_{e_{mi}}}^{U''_{e_{mi}}} dF | \]

by reasoning similar to that above, in order to show \( P\{A_1 > \frac{\epsilon}{2}\} < \frac{\eta}{4} \) it is sufficient to show that

\[ P\left\{ B_{\lambda_j} > \frac{\epsilon}{4} \right\} < \frac{\eta}{4} \quad j = 1, 2 \]

If either \( \int_{U'_{e_{mi}}}^{U''_{e_{mi}}} dF = 0 \) or \( \int_{U'_{e_{mi}}}^{U''_{e_{mi}}} dF = 0 \) then this condition is satisfied for both \( j = 1 \) and \( j = 2 \). If neither of these probabilities are 0 then consider the following.

Since \( \{F(n)\} \) is a sequence of sample distributions for \( X \), then for any Borel set, \( e \in E^F \),

\[ \int_{n}^{dF(n)} = \frac{1}{n} \left\{ \text{number of sample points in } e \right\} \]

\[ = \frac{1}{n} S_n \quad (say) \]

and

\[ \int_{e}^{dF} = P\{X \in e\} \]

\[ = p \quad (say) \]
Then $S_n$ is a binomial random variable with mean $np$ and variance $np(1 - p)$. Thus, if $0 < p < 1$,

$$
P \left\{ \left| \frac{S_n}{n} - p \right| > \frac{\varepsilon}{4B} \right\}
$$

$$
= P \left\{ \left| \frac{S_n - np}{\sqrt{np(1 - p)}} \right| > \frac{\varepsilon}{4B} \sqrt{\frac{n}{p(1 - p)}} \right\}
$$

$$
\leq P \left\{ \left| \frac{S_n - np}{\sqrt{np(1 - p)}} \right| > \frac{2\varepsilon}{4B} \right\}
$$

since $p(1 - p) \leq \frac{1}{4}$. Then, by Tchebycheff's inequality, this is

$$
\leq \frac{4B^2}{\varepsilon^2 n}.
$$

Thus, choose $N_1$ such that $n > N_1$ implies that $\frac{2B}{\varepsilon^2 n} < \frac{1}{4}$ and then for any Borel set, $\varepsilon$, any $X \in \mathcal{X}^\Gamma$,

$$
P \left\{ \left| \frac{dF(n)}{dF} - 1 \right| > \frac{\varepsilon}{4B} \right\} < \frac{1}{4}
$$

Therefore, for $n > N_1$, any $m$, by taking the Borel set, $\varepsilon$, to be first $U^{e_{mi}}$ and then $U^{e_{mi}}$ we have $P \left\{ A_{1j} > \frac{\varepsilon}{4B} \right\} < \frac{1}{4}$, for $j = 1, 2$.

Hence, for $n > N_1$, any $m$, any $X \in \mathcal{X}^\Gamma$,

$$
P \left\{ A_1 > \frac{\varepsilon}{2} \right\} < \frac{1}{2}.$$
Next, we see

\[ A_2 = | \int \left( I_m dF - \int IdF \right) | \]
\[ = | \int (I_m - I) dF | \]
\[ \leq \int |I_m - I| dF \]
\[ \leq \sup_{x \in \mathbb{R}} |I_m(x) - I(x)|. \]

But by H-2, for any \( \varepsilon, \gamma > 0 \) there exists an \( N_2 \) such that \( m > N_2 \) implies for any \( Y \in \omega \),

\[ P \left\{ \sup_{x \in \mathbb{R}} |I_m(x) - I(x)| > \frac{\varepsilon}{2} \right\} < \frac{\gamma}{2} \]

Thus \( m > N_2 \) implies

\[ P \left\{ A_2 > \frac{\varepsilon}{2} \right\} < \frac{\gamma}{2} \]

Therefore, we choose \( N \) as the larger of \( N_1 \) and \( N_2 \) and the theorem is proved.
Chapter IV

Applications of the Convergence Theorem

Application to a Subset of \( \mathcal{I} \)

A sample statistic, \( S_n \), for a sample of size \( n \) from a random variable, \( X \), was defined to be a member of \( \mathcal{I} \) if it could be written in the form

\[
S_n = \int_{\mathbb{R}^r} I_n(x) dF(n)(x)
\]

and

\[
I_n(x) = Q(C_{x}^{(n)}[x]).
\]

Now a subset \( \mathcal{I}_P \subset \mathcal{I} \) is defined.

Definition: The subset \( \mathcal{I}_P \subset \mathcal{I} \) consists of all \( S_n \in \mathcal{I} \) such that the integrand function, \( I_n(x) \), is of the form

\[
I_n(x) = Q_P(C_{x}^{(n)}[x])
\]

where \( Q_P(t_1, t_2, \ldots, t_{2^r - 1}) \) is a polynomial in \( t_1, t_2, \ldots, t_{2^r - 1} \).

Hence this subset, \( \mathcal{I}_P \), is just the set of all statistics in \( \mathcal{I} \) whose integrand functions are polynomials in the sample distribution functions.

Definition: A sequence of statistics, \( \{S_n\}, S_n \in \mathcal{I} \) for samples from \( X \) is called a related sequence if the integrand function for each \( n \)
is based on the same function, \( Q \). That is, there exists \( Q \) such that

\[
I_n(x) = Q(\mathcal{F}(n)[x]) \quad n = 1, 2, 3, \ldots.
\]

The convergence theorem can be used to evaluate the probability limit of related sequences in \( \mathcal{L}_p \). In order to show that the hypotheses of the theorem are satisfied a sort of weak multivariate Glivenko-Cantelli theorem is needed. The required result follows directly from a theorem by J. Kiefer and J. Wolfowitz.\(^{[5]}\)

**Theorem.** (Kiefer and Wolfowitz) Let \( X \) be a random variable in \( \mathcal{X}^r \) with distribution function \( F(x) \), and sample distribution function for a sample of size \( n, F^{(n)}(x) \). Then if

\[
D_n = \sup_{x \in \mathcal{P}} |F^{(n)}(x) - F(x)|
\]

and

\[
K_n(z; X) = P\{D_n \leq \frac{z}{\sqrt{n}}\}
\]

then for each \( r \) there exist positive constants \( c_0 \) and \( c \) such that, for all \( n \), all \( X \), all positive \( z \)

\[
1 - K_n(z; X) \leq c_0 e^{-cz^2}.
\]

**Corollary 1.1.** (to Kiefer-Wolfowitz Theorem) Let \( X \) be a random variable over \( \mathcal{X}^r \) with distribution function \( F(x) \) and sample distribution function \( F^{(n)}(x) \), then, uniformly for all \( X \in \mathcal{X}^r \)

\[
\limsup_{n \to \infty} \sup_{x \in \mathcal{X}^r} |F^{(n)}(x) - F(x)| = 0.
\]
Proof of corollary: Let $D_n$ and $K_n(x)$ be as defined above. Then for $\epsilon > 0$, by the previous theorem there exist positive constants, $c_0$ and $c$, such that, for all $x \in X^r$, 

$$P\left\{\sup_x |F^{(n)}(x) - F(x)| > \epsilon \right\} = P\{D_n > \epsilon\}$$

$$= 1 - K_n(\epsilon \sqrt{n}; x)$$

$$< c_0 e^{-cn^2}$$

$$\frac{n}{n} \to 0$$

and the corollary is proved.

Now a corollary concerning related sequences in $\mathcal{F}_p$ can be given.

Corollary 1.2: Let $X$ be a random variable with distribution function vector $\overline{F}[x]$ and sample distribution function vector $\overline{F}^{(n)}[x]$. Let $\{S_n\}$ be a related sequence of statistics in $\mathcal{F}_p$ with 

$$S_n = \left\{ \int_{\mathcal{F}} dF^{(n)} \right\}$$

where 

$$I_n(x) = Q_p(\overline{F}^{(n)}[x]).$$

If 

$$I_x(x) = Q_p(\overline{F}[x])$$

then 

$$\lim_{n \to \infty} S_n = E(I_x(x)), \quad \text{uniformly for } x \in X^r.$$ 

Proof: If it can be shown that the sequence $\{I_n\}$ satisfies the
conditions of $H-2$ in the hypothesis of the convergence theorem, then it follows that

$$\lim_{n \to \infty} S_n = \lim_{n \to \infty} \int_{\mathbb{F}^r} I_n dF(n) = \int_{\mathbb{F}^r} I_{\mathbb{F}} dF = E(I_{\mathbb{F}}), \quad \text{uniformly for } \mathbb{F} \in \mathbb{F}^r.$$

To fulfill $H-2$ the sequence $\{I_n\}$ must be a sequence of Borel simple sample functions such that

1) there exists a number, $B$, such that $|I_n(x)| < B$ for all $n$ and all $x$,

2) $\limsup_{n \to \infty} \|I_n(x) - I_{\mathbb{F}}(x)\| = 0$, uniformly for $\mathbb{F} \in \mathbb{F}^r$.

First of all, the $I_n$ are Borel simple sample functions since they are polynomials in the sample distribution functions which are themselves Borel simple sample functions, and finite sums and products of Borel simple sample functions are again Borel simple sample functions.

In order to show that (1) and (2) are satisfied by $\{I_n\}$ a slight change of notation is helpful. In this proof let the $A$th component of the distribution function vector $\vec{F}$ be written $F_A$. Thus, $F_1$ is still written $F_1$, $F_2$ is still $F_2$, $F_r$ is $F_r$, but $F_{12}$ is now written $F_r + \mathbb{1}$ and so on until $F$ becomes $F_2^{r-1}$. The same subscript notation is used for the components of $\vec{F}(n)$.

Since $I_{\mathbb{F}}(x)$ is a polynomial in the $F_A$ it can be written

$$I_{\mathbb{F}}(x) = \sum_p a_{p_1p_2\cdots p_r} F_1^{p_1} F_2^{p_2} \cdots F_r^{p_r} - I_{\mathbb{F}} \mathbb{1}.$$
where \( \sum^* \) is the sum over all sets, \( p_1, p_2, \ldots, p_2^r - 1 \), of non-negative integers which are less than some number, \( P \), say, the highest power appearing in \( I_X \) and \( a^1, a^2, \ldots, a^r \) are constants. Then for all \( n \) and all \( x \)

\[
|I_n(x)| = \left| \sum^* a_{p_1 p_2 \ldots p_2^r - 1} \left( \prod_{i=1}^{2^r-1} p_i^{x_i} \right) \right| \\
\leq \sum^* |a_{p_1 p_2 \ldots p_2^r - 1}|.
\]

Since \( \sum^* \) is a finite sum and the \( a_{p_1 p_2 \ldots p_2^r - 1} \) are finite constants condition (i) is satisfied.

For (ii) we must show that

\[
\limsup_n |I_n(x) - I_X(x)| = 0 \quad \text{uniformly for } x \in X^r.
\]

This is easily done by partitioning the difference \( |I_n - I_X| \) and using corollary 4.1 on each part. Let

\[
\Delta = \max_{a_{p_1 p_2 \ldots p_2^r - 1}} |a_{p_1 p_2 \ldots p_2^r - 1}|
\]

then

\[
\sup_x |I_n(x) - I_X(x)| = \sup_x \left| \sum^* a_{p_1 \ldots p_2^r - 1} \left( \prod_{i=1}^{2^r-1} p_i^{x_i} - \prod_{i=1}^{2^r-1} F(n)_i^{x_i} \right) \right| \\
< \Delta \sum^* \sup_x \left| \prod_{i=1}^{2^r-1} p_i^{x_i} - \prod_{i=1}^{2^r-1} F(n)_i^{x_i} \right|.
\]

Hence it is sufficient to show that for every set, \( (p_1, \ldots, p_2^r - 1) \),

\[
\limsup_{n x} \left| \prod_{i=1}^{2^r-1} p_i^{x_i} - \prod_{i=1}^{2^r-1} F(n)_i^{x_i} \right| = 0 \quad \text{uniformly for } x \in X^r.
\]
For this take $F_0 = F_2^{(n)} = P_0 = P_2 = 1$, and then
\[
\left| \prod_{i=1}^{2^n} F_i - \prod_{i=1}^{2^n} F_i^{(n)} \right| = \left| \sum_{i=1}^{2^n} \left( \prod_{j=1}^{i-1} F_j - F_j^{(n)} \right) \left( \prod_{j=1}^{i} F_j - F_j^{(n)} \right) \right|
\]
\[
\leq \sum_{i=1}^{2^n} \left| F_i - F_i^{(n)} \right|
\]
\[
\leq \sum_{i=1}^{2^n} p_i \left| F_i - F_i^{(n)} \right|
\]
since for any $i$, $|F_i| \leq 1$, and $|F_i^{(n)}| \leq 1$. Thus to show that equation (21) holds it is now sufficient to show that
\[
(22) \quad \limsup_{n \to \infty} \prod_{i=1}^{2^n} \left| F_i - F_i^{(n)} \right| = 0
\]
uniformly in $X$, for $i = 1, 2, \ldots, 2^n - 1$. However, for each $i$, $F_i$ is a distribution function and $F_i^{(n)}$ is the corresponding distribution function, and therefore, by corollary 4.1 we have that (22) holds. Hence (ii) is satisfied.

Since both (i) and (ii) are fulfilled corollary 4.2 is proved.

**Consistency**

Form corollary 4.2 some immediate consequences follow concerning the consistency of related sequences of statistics in $\mathcal{S}_p$ when used as estimators or in testing hypotheses.

**Corollary 4.3:** If $\{S_n\}$ is a related sequence in $\mathcal{S}_p$ and $I_X$ is given by (19), then $S_n$ is a consistent estimate of $E(I_X(X))$. 
Proof: \( S_n \) is a consistent estimate of \( E(I_X(X)) \) if
\[
\lim_{n \to \infty} S_n = E(I_X(X)),
\]
but this is true from corollary 4.2.

In order to more clearly discuss the consistency of related sequences in \( \mathcal{Q}_p \) when used in testing hypotheses a short presentation of these concepts is given.

Let \( \omega \) be some subset of \( \mathcal{X}^r \), then a possible hypothesis to be tested is that \( X \in \omega \) against the alternative that \( X \in \mathcal{X}^r - \omega \). This will be written
\[
\begin{align*}
H_0 & : X \in \omega \\
H_1 & : X \in \mathcal{X}^r - \omega
\end{align*}
\]
If a test statistic, \( T_n(X) \), is used to test this hypothesis with a critical region, \( e_n \), so that the test procedure is
- if \( T_n(X) \notin e_n \) accept the hypothesis
- if \( T_n(X) \in e_n \) accept the alternative,
then we identify the test by writing \( (T_n, e_n) \). The power function for the test \( (T_n, e_n) \) is
\[
P_X(e_n) = P \{ T_n(X) \in e_n | X \},
\]
and the test is of the exact size \( \lambda > 0 \) if
\[
P_X(e_n) \leq \lambda \quad \text{for all } X \in \omega,
\]
and if there exists at least one \( X \in \omega \) such that \( P_X(e_n) = \lambda \).
This may be restated by saying that $(T_n, e_n)$ is of exact size $\lambda$ if

$$\sup_{X \in \omega} P_X(e_n) = \lambda.$$

**Definition:** A sequence of tests, $\{(T_n, e_n)\}$, is consistent for $\mathcal{F} \subset \mathcal{X}$ if

$$\lim_{n \to \infty} P_X(e_n) = 1 \quad \text{for all } X \in \mathcal{F}.$$

Now suppose $T(X)$ is a real parameter defined over $\mathcal{X}$, and $\omega$ in the hypothesis of (23) is just the set of all $X \in \mathcal{X}$ such that $T(X) = \Theta_0$. Then this hypothesis with a one-sided alternative could be written

$$H_0: \quad T(X) = \Theta_0,$$

$$H_1: \quad T(X) > \Theta_0.$$

Here, $\mathcal{F}$, the alternative set, consists of all $X$ such that $T(X) > \Theta_0$. For this formulation, a simple condition for consistency is given by the following lemma, ([4], p 267, and [6]).

**Lemma 4.1:** If $\lim_{n \to \infty} T_n(X) = T(X)$ uniformly for $X \in \omega$, then the sequence of tests $\{(T_n, e_n)\}$, where the critical region, $e_n$, is given by

$$T_n > \Theta_0 + c_n, \quad c_n > 0,$$

is consistent for $\mathcal{F}$.

**Proof:** First we show that

$$\lim_{n \to \infty} \sup_{m \geq n} c_m < 0.$$
Suppose that \( \lim \sup c_m = c > 0 \). Then there exist arbitrarily large \( n \) such that \( c_n > \frac{c}{2} \). Since the tests are of the exact size \( \omega \), for any \( n \) there exists an \( X \in \omega \) such that

\[ P_X(e_n) = P\{T_n(X) - \theta_0 > c_n | X\} \]

Thus there exist arbitrarily large \( n \) such that

(26) \( 0 < \frac{1}{n} \leq P\{T_n(X) - \theta_0 > \frac{c}{2} | X\} \) for some \( X \in \omega \).

But by the hypothesis

\[ \liminf_{n \to \infty} T_n(X) = \Theta_0 \]

uniformly for \( X \in \omega \). This implies that

\[ \sup_{X \in \omega} P\{T_n(X) - \theta_0 > \frac{c}{2} | X\} \to 0 \]

and this is in contradiction to (26), hence

(27) \( \lim \sup c_m \leq 0 \).

Now, for \( X \in \mathcal{F} \), and \( 0 < \xi < T(X) - \theta_0 \), we have by the hypothesis

\[ \liminf_{n \to \infty} T_n(X) = T(X) > \xi + \theta_0 \]

However, by (27), there exists an \( N \) such that \( n > N \) implies \( c_n < \xi \), and hence
\[
\lim_{n \to \infty} P_n(e_n) = \lim_{n \to \infty} P\{T_n(X) - \Theta_0 > c_n |X\}
\]
\[
= \lim_{n \to \infty} P\{T_n(X) - \Theta_0 > \varepsilon |X\}
\]
\[
= \lim_{n \to \infty} P\{T_n(X) - \Theta_0 + \varepsilon |X\}
\]
\[
= 1
\]

Therefore, the sequence \(\{(T_n, e_n)\}\) is consistent for \(X \in \mathcal{Y}\).

This lemma can be extended to cover two-sided tests against the alternative, \(T(X) \neq \Theta_0\). By replacing the critical region, \(e_n\), in (25) with \(e_1\) given by

\[
|T_n - \Theta_0| > c_n
\]

and with a very similar argument, it can be shown that the sequence of tests, \(\{(T_n, e_n)\}\), is consistent for the set of all \(X\) such that \(T(X) \neq \Theta_0\).

This lemma can be used together with corollary 1.2 to show consistency for related sequences in \(\mathcal{Q}_p\).

**Corollary 1.4:** Suppose \(\{S_n\}\) is a related sequence of statistics in \(\mathcal{Q}_p\), and a test of

\[
H: E(I(X)) = \Theta_0
\]

\[
A: E(I(X)) > \Theta_0
\]

of exact size \(\alpha > 0\) is given by \((S_n, e_n)\), where the critical region, \(e_n\), is given by \(S_n > \Theta_0 + c_n\). Then the sequence of tests \(\{(S_n, e_n)\}\)
is consistent for the alternative.

Similarly, for the two-sided alternative we have,

**Corollary 4.5:** Suppose \( S_n \) is a sequence of related statistics in \( \mathcal{X}_p \) and a test of

\[
H_0 : E(I_X(x)) = \Theta_0 \\
A_1 : E(I_X(x)) \neq \Theta_0
\]

of exact size \( \alpha > 0 \) is given by \( (S_n, \epsilon_n^1) \), where the critical region, \( \epsilon_n^1 \), is given by \( |S_n - \Theta_0| > \epsilon_n^1 \). Then the sequence of tests \( \{(S_n, \epsilon_n)\} \) is consistent for the alternative.

**Example 4:** In Chapter I, Example 1, it was seen that Kendall's \( \tau \) is given by

\[
\tau(n) = \frac{\ln n}{n-1} \tau_s(n) - \frac{n+1}{n-1}
\]

where

\[
\tau_s(n) = \int_{F^2} F(n) dF(n)
\]

Since \( \tau_s(n) \in \mathcal{X}_p \) and \( \{\tau(n)\} \) is a related sequence, by corollary 4.2

\[
\lim_{n \to \infty} \tau_s(n) = E(F(X))
\]

or

\[
\lim_{n \to \infty} \tau(n) = 4E(F(X)) - 1.
\]
The statistic $\tau_s^{(n)}$ (or $\tau^{(n)}$) could be used in a test of

\[ H_0: E(F(X)) = \theta_0, \]
\[ H_1: E(F(X)) > \theta_0. \]

with $e_n$, the critical region of exact size $\alpha > 0$, given by

\[ \tau_s^{(n)} > \theta_0 + c_n. \]

The sequence of tests, $\{(\tau_s^{(n)}, e_n)\}$, is consistent for the above alternative. Similarly, the corresponding sequence of two-sided tests $\{(\tau_s^{(n)}, e_n')\}$ as given in corollary 4.5 is consistent against the two-sided alternative, $E(F(X)) \neq \theta_0$.

The value of the parameter $E(F(X))$, since it is a cumulative expected value, is the same for all members of any particular equivalence class. However, there may be many equivalence classes for which this parameter has the same value. Hence the hypothesis of (29) is generally composed of many different equivalence classes.

Although the distribution of $\tau_s^{(n)}$ is the same for all members of a particular equivalence class it will vary from class to class. Thus, with a composite hypothesis as in (29) the distribution of the sample statistic will not be the same for all members of the hypothesis set.

It was for this sort of situation that the uniformity of convergence was required in lemma 4.1 and corollary 4.2.

Let $\omega_i$ be the equivalence class of $X \in \chi^2$ composed of all the random variables with independent components. The common usage of $\tau^{(n)}$ is to test

\[ H_0: X \in \omega_i, \]
\[ H_1: X \notin \chi^2 - \omega_i. \]
Now let \( \omega^* \) be the set of all \( X \in \mathcal{X}^2 \) such that \( E(F(X)) = \frac{1}{4} \), then by the remarks in the last section of Chapter II we have \( \omega_1 \subset \omega^* \).

From above, the sequence of tests of

\[
\begin{align*}
H_0: & \quad X \in \omega^* \\
H_1: & \quad X \notin \mathcal{X}^2 - \omega^*
\end{align*}
\]

with critical regions \( |\tau_s^{(n)} - \frac{1}{4}| > c_n \), is consistent for any alternative \( \mathcal{Y} \subset \mathcal{X}^2 - \omega^* \). The same sequence of tests when used, as it usually is, for (30) will consequently be consistent for any \( \mathcal{Y} \subset \mathcal{X}^2 - \omega^* \). However, it will not necessarily be consistent for \( \mathcal{Y} \) containing some \( X \in \omega^* - \omega_1 \). This set, \( \omega^* - \omega_1 \) is the set of all \( X \in \mathcal{X}^2 \) such that \( E(F(X)) = \frac{1}{4} \) and the components are not independent.

A simple example of \( X \in \omega^* - \omega_1 \) is the following. Let \( X \) be defined over the unit square by means of the density function

\[
f(x_1, x_2) = \begin{cases} 
  a, & \text{for } 0 \leq x_1 \leq \frac{1}{2}, 0 \leq x_2 \leq \frac{1}{2} \\
  b, & \text{for } 0 \leq x_1 \leq \frac{1}{2}, \frac{1}{2} < x_2 \leq 1 \\
  c, & \text{for } \frac{1}{2} < x_1 \leq 1, 0 \leq x_2 \leq \frac{1}{2} \\
  d, & \text{for } \frac{1}{2} < x_1 \leq 1, \frac{1}{2} < x_2 \leq 1.
\end{cases}
\]

If \( b = \frac{1}{2}, d = 1, a = \frac{5 + \sqrt{15}}{4}, c = \frac{5 - \sqrt{15}}{4} \), then \( X_1 \) and \( X_2 \) are not independent, yet \( E(F(X)) = \frac{1}{4} \).

**Example 2**: A natural extension of Kendall's \( \tau \) to any dimension \( r \geq 2 \) is given by

\[
\tau_s^{(n)} = \int_{F} f(n) dF(n)
\]
for samples of size \( n \) from \( X \in \mathcal{X}^r \). This statistic, with a critical region of \( |\tau_s^{(n)} - \theta_0| > c_n \), gives a sequence of tests of

\[
H_0: \quad E(F(X)) = \theta_0
\]

that is consistent against any alternative such that \( E(F) \neq \theta_0 \).

As before let \( \omega_I \) be the equivalence class of \( X \in \mathcal{X}^r \) with mutually independent components. Let \( \omega^* \) be the set of \( X \in \mathcal{X}^r \) such that

\[
E(F(X)) = \frac{1}{2^r}.
\]

Then \( \omega^* \subset \omega_I \). From above, the sequence of two-sided tests of

\[
H_0: \quad X \in \omega^*
\]

\[
A: \quad X \in \mathcal{X}^r - \omega^*
\]

given by a critical region of \( |\tau_s^{(n)} - \frac{1}{2^r}| > c_n \) is consistent for any alternative \( \exists \subset \mathcal{X}^r - \omega^* \). Hence \( \tau_s^{(n)} \) could be used as a test of

\[
H_0: \quad X \in \omega_I
\]

\[
A: \quad X \in \mathcal{X}^r - \omega_I
\]

just as it is for \( r = 2 \). The sequence of tests is consistent for any alternative \( \exists \in \mathcal{X}^r - \omega^* \), but not necessarily for any \( \exists \) containing some \( X \in \omega^* - \omega_I \). Thus the test for independence using Kendall's \( \tau \) can be extended to any dimension, but, for any dimension, there are classes of non-independent random variables for which it may not be consistent.

**Example 6:** In Chapter I, Example 2, Spearman's rank correlation
coefficient was seen to be

$$\rho^{(n)} = \frac{12n^2}{n^2 - 1} \rho_s^{(n)} - \frac{2n + 1}{n - 1}$$

where

$$\rho_s^{(n)} = \int_{F_1}^{F_2}(n) dF(n)$$

Since $\rho_s^{(n)} \in \mathcal{L}_p$ and \{\rho^{(n)}\} is a related sequence, by corollary 4.2

$$\lim_{n \to \infty} \rho_s^{(n)} = E(F_1F_2)$$

or

$$\lim_{n \to \infty} \rho^{(n)} = 12E(F_1F_2) - 2$$

Remarks on $\rho_s^{(n)}$ can be made which are similar to those made for $\tau_s^{(n)}$. $\rho_s^{(n)}$ could be used in a test of

$$H_1: E(F_1F_2) = \theta_0$$

$$A_1: E(F_1F_2) > \theta_0$$

(35)

with a critical region of the form $\rho_s^{(n)} > \theta_0 + c_n$, of exact size \(\Delta > 0\). This sequence of tests is consistent for the alternative above. Similarly, the corresponding sequence of two-sided tests of this hypothesis is consistent against the alternative, $E(F_1F_2) \neq \theta_0$.

Spearman's $\rho$, just as Kendall's $\tau$, is commonly used to test

$$H_1: X \in \omega_1$$

$$A_1: X \notin \omega_1.$$  

(36)

Since $E(F_1F_2) = \frac{1}{4}$ for $X \in \omega_1$, the critical region is of the form

$$|\rho_s^{(n)} - \frac{1}{4}| > c_n.$$  

Now let $\omega^{**}$ be the set of all $X \in \mathcal{X}^2$ such that
\[ E(F_1F_2) = \frac{1}{4}, \text{ then } \omega \subset \omega^{**}. \] With this critical region, \( \rho_s^{(n)} \) gives a sequence of tests of

\[ (37) \begin{align*}
H_0: & \quad X \in \omega^{**} \\
A: & \quad X \in \chi^2 - \omega^{**}
\end{align*} \]

which is consistent for any \( \beta \subset \chi^2 - \omega^{**} \). This test when used for (36) will thus be consistent for any \( \beta \subset \chi^2 - \omega^{**} \), but may not be consistent for any \( \beta \) containing some \( X \in \omega^{**} - \omega_I \).

Here again, \( \omega^{**} - \omega_I \) is not an empty set. An example of a member is given by the random variable of the last example with the density function as in (32) but with \( b = \frac{1}{2}, d = 1, a = \frac{b}{2} + \sqrt{\frac{b}{4} - b}, \) and \( c = \frac{32 - \sqrt{1489}}{32} \). This yields \( E(F_1F_2) = \frac{1}{4} \), yet \( X_1 \) and \( X_2 \) are not independent.

**Example 7:** Spearman's \( \rho \) can also be extended in a natural way to dimensions \( r > 2 \) by

\[ \rho_s^{(n)} = \int_{E^r} F_1^{(n)}F_2^{(n)} \cdots F_r^{(n)} dF(n). \]

Then

\[ \text{plim } \rho_s^{(n)} = E(\sum_{d=1}^{r} F_d^{(r)}). \]

This statistic, with the usual one-sided (or two-sided) critical region provides a sequence of tests of

\[ H_0: \quad E(\sum_{d=1}^{r} F_d^{(r)}) = \theta_0 \]

which is consistent for alternatives such that \( E(\sum_{d=1}^{r} F_d^{(r)}) > \theta_0 \) (or \( E(\sum_{d=1}^{r} F_d^{(r)}) \neq \theta_0 \)). It could be used as a test for mutual independence.
of the components,

$$H_0: X \in \omega_1,$$

with a critical region of the form $|P_X(n) - \frac{1}{2^n}| > c_n$, since for $X \in \omega_1$, $E(\prod F_i) = \frac{1}{2^n}$. This sequence of tests is consistent for any alternative for which $E(\prod F_i) \neq \frac{1}{2^n}$, but need not be for alternatives containing some $X$ such that $E(\prod F_i) = \frac{1}{2^n}$.

**Example 8:** Let

$$S_n = \int_{E^2} (F(n) - F_1(n)F_2(n))^2 dF(n).$$

Since $S_n \in \mathcal{G}_p$, by corollary 4.2

$$\lim_{n \to \infty} S_n = E[(F - F_1F_2)^2].$$

A test of the hypothesis of independence in $\chi^2$ could be based on $S_n$ with a critical region of the form $S_n > c_n$. Since $E[(F - F_1F_2)^2] = 0$ if and only if $X \in \omega_1$, this sequence of tests is consistent for any alternative $\gamma \in \chi^2 - \omega_1$. Thus, there are no $X$ with dependent components for which this test is not consistent.

**Application to Two Sample Statistics**

A two sample statistic for a sample of size $n$ from $X$ and a sample of size $m$ from $Y$ was defined to be in $\mathcal{G}_2$ if it could be written as

$$S_{mn} = \int_{E^n} I_{m} dF(n).$$
where

\[ I_m(x) = Q(\mathbf{A}(m)[x]) \]

and \( \mathbf{A}(m) \) is the sample distribution function vector for the sample from \( Y \). A subset \( \mathcal{I}_p \subset \mathcal{I} \), analogous to \( \mathcal{I}_p \subset \mathcal{I} \), can also be defined.

**Definition:** The subset \( \mathcal{I}_p \subset \mathcal{I} \) consists of all \( S_{mn} \in \mathcal{I}^2 \) for which the integrand function, \( I_m \), is of the form

\[ I_m(x) = Q_p(\mathbf{A}(m)[x]) \]

where \( Q_p(t_1, t_2, \ldots, t_{2r-1}) \) is a polynomial in \( t_1, t_2, \ldots, t_{2r-1} \).

**Definition:** A double sequence, \( \{S_{mr}\} \), \( S_{mn} \in \mathcal{I}^2 \), is called a related sequence if the integrand function for each \( n, m \) is based on the same function, \( Q \). That is, there is a \( Q \) such that

\[ I_m(x) = Q(\mathbf{A}(m)[x]) \quad m = 1, 2, \ldots. \]

The convergence theorem can now be used to evaluate the probability limit of related sequences of statistics, in \( \mathcal{I}_p^2 \).

**Corollary 4.6:** Let \( \{S_{mn}\} \) be a related sequence of statistics in \( \mathcal{I}_p^2 \) such that

\[ S_{mn} = \int_{\mathcal{F}^r} I_m dP(n) \]

where

\[ I_m(x) = Q_p(\mathbf{A}(m)[x]). \]
If
\[ I_Y(x) = Q_p(\mathcal{A}[x]) \]
then
\[ \text{plim}_{m \to \infty} S_{mn} = E(I_Y(x)), \quad \text{uniformly for } x \in \mathcal{X}^F, \ y \in \mathcal{X}^F. \]

Proof: The proof is entirely analogous to the proof of corollary 4.2.

The \( I_m(x) \) are Borel simple sample functions which are uniformly bounded in \( m \) and \( x \). Moreover,
\[ \text{plim}_{m \to \infty} \sup_{x} |I_m(x) - I_y(x)| = 0, \quad \text{uniformly for } Y. \]

Then, by the convergence theorem
\[ \text{plim}_{m \to \infty} S_{mn} = \text{plim}_{m \to \infty} \int I_m dF(n) \]
\[ = \int I_m dF \]
\[ = E(I_Y(x)), \quad \text{uniformly for } x \in \mathcal{X}^F, \ y \in \mathcal{X}^F. \]

From this corollary follows a consequence regarding the consistency of related sequences in \( \mathcal{S}_p^2 \) when used in testing hypotheses. A lemma, similar to lemma 4.1 but for two-sample statistics, gives the required conditions for consistency.
Suppose $T_{mn}$ is a two-sample statistic for a sample of size $n$ from $X \in \mathcal{X}$ and a sample of size $m$ from $Y \in \mathcal{Y}$, and suppose that $T(X,Y)$ is a real valued function defined for $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$. The hypothesis to be tested is

$$H_0: T(X,Y) = \theta_0.$$ 

**Lemma 4.2:** If $\lim_{m,n} T_{mn} = T(X,Y)$ uniformly for all $(X,Y)$ such that $T(X,Y) = \theta_0$, then the sequence of tests, $\{(T_{mn}, a_{mn})\}$, of exact size $\alpha > 0$, where the critical region, $e_{mn}$, is of the form

$$T_{mn} > \theta_0 + c_{mn},$$

is consistent for any set $(X,Y)$ such that $T(X,Y) > \theta_0$.

This lemma can be extended to cover two-sided alternatives with the two-sided critical regions.

**Example 2:** In Chapter I, Example 3, the Mann-Whitney $U$ statistic was found to be

$$U(m,n) = \sum_{i=1}^{m} U_s(m,n)$$

where

$$U_s(m,n) = \int_{E^1} G(m)dF(n).$$

Since $U_s(m,n) \leq \frac{2}{m}$ and $\{U_s(m,n)\}$ is a related sequence, then

$$\lim_{m,n} U_s(m,n) = \int_{E^1} GdF = E(G(X))$$
uniformly for X and Y. Then the U statistic, with a critical region
as in lemma 4.2, provides a consistent sequence of tests of
\[ H_0: E(G(X)) = \theta_0 \]
which is consistent against any alternative such that \( E(G(X)) > \theta_0 \).
In particular, it gives a sequence of tests of
\[ (37) \quad H_0: E(G(X)) = \frac{1}{2} \]
which is consistent for all alternatives such that \( E(G(X)) > \frac{1}{2} \).
U is usually used to test
\[ (38) \quad H_0: F(x) = G(x). \]
Now if \( F = G \), then \( E(G(X)) = \frac{1}{2} \), thus the hypothesis set of (38) is a
subset of the hypothesis set of (37). Therefore, when U is used for a
test of (38) it gives a sequence of tests that is consistent against
any alternative such that \( E(G(X)) > \frac{1}{2} \).
Analogous comments to those above can be made for two-sided
alternatives with two-sided critical regions.

Example 10: A natural extension of \( u^{(m,n)}_s \) to two or more dimensions
is given by
\[ u^{(m,n)}_s(r) = \int_{E^r} G(m) dF(n) \]
where \( X \in \mathcal{X}^r \) and \( Y \in \mathcal{X}^r \). Under the hypothesis that \( F = G \)
\[ \lim_{m,n \to \infty} u^{(m,n)}_s(r) = \int_{E^r} GdF = \int_{E^r} FdF = E(F(X)). \]
In one dimension the U test is distribution free since under the hypo-
thesis the distribution of U does not depend on F as long as F is
continuous. However, when \( r > 1 \), \( E(F(X)) \) depends on \( F \) and, consequently, the distribution of \( U_{g}^{(m,n)}(r) \) could not be independent of \( F \). Thus, the \( U \) statistic, like the Kolmogoroff-Smirnov statistic, is one that suffers the loss of its distribution-free nature as it is extended, at least in this way, to two or more dimensions.
CHAPTER V

REMARKS AND EXAMPLES ON LIMITING DISTRIBUTIONS

Introduction

In Chapter IV it was found that a statistic, $S_n \in \mathcal{L}_p$, converges in probability to a constant. When these statistics are properly normalized they may have a non-trivial limiting distribution. In this chapter the rate at which $S_n$ converges to a constant is investigated in order to determine the appropriate normalization.

The limiting distribution of $S_n$ properly normalized is not in general known, but some important special cases, namely, Kendall's $\tau$ and Spearman's $\rho$ are known to have a normal distribution in the limit by the theory of U statistics ([4] p. 232). However, since the members of $\mathcal{L}_p$ are not always U statistics, this theory does not encompass all of $\mathcal{L}_p$. In the last part of this chapter there is some discussion on the convergence of moments of statistics in $\mathcal{L}_p$.

Normalization of $S_n$

Let $S_n \in \mathcal{L}_p$ be given by

$$S_n = \frac{1}{n} \sum_{k=1}^{n} \varphi_p(x^{(n)}(k))$$

and

$$\tilde{S}_n = \sqrt{n}S_n$$

Then the following can be shown.
Theorem 5.1: Let \( \{a_n\}, n = 1, 2, \cdots \) be any increasing, unbounded, sequence of positive numbers. Then there exists a sequence, \( \{b_n\} \), such that

\[
\lim_{n \to \infty} \frac{S_n - b_n}{a_n} = 0
\]

Proof: Let

\[
T_n = \frac{1}{n} \sum_{k=1}^{n} q_p(\sigma F[k(k)])
\]

(39)

\[
\gamma_n = \sqrt{n} T_n
\]

First it is shown that

(40) \( \lim_{n \to \infty} \frac{S_n - \eta_n}{a_n} = 0 \).

To do this, the difference, \( |S_n - \eta_n| \), must be decomposed exactly as a similar difference was decomposed in Corollary 4.2. The notation of that corollary is used here, and it was seen there that

\[
|S_n - \eta_n| \leq \frac{1}{\sqrt{n}} \sum_{k=1}^{n} q_p(\sigma F[k(k)]) - q_p(\sigma F[k(k)])
\]

\[
\leq \sqrt{n} \Delta \left( \sum_{n} \sup_{x} \left( \prod_{j=1}^{r} F_j(n) p_j - \prod_{j=1}^{r} F_j p_j \right) \right)
\]

Thus, in order to show (40) it is sufficient to show that for any \( \varepsilon > 0 \), and any set \( \{p_1, p_2, \cdots, p_{2r}, -1\} \),

(41) \( P \left\{ \sqrt{n} \sup_{x} \left( \prod_{j=1}^{r} F_j(n) p_j - \prod_{j=1}^{r} F_j p_j \right) > \varepsilon a_n \right\} \to 0 \),

but, as in (21), the probability above is less than or equal to

\[
P \left( \sqrt{n} \sum_{j=1}^{r} p_j \sup_{x} |F_j(n) - F_j| > \varepsilon a_n \right).
\]
Hence, it is now sufficient to show that for every $i$

\[(42) \quad P\left\{ \sup_x |F_i(n) - F_i| > \varepsilon_n \right\} \to 0,\]

but by the Wolfowitz-Kiefer Theorem

\[P\left\{ \sup_x |F_i(n) - F_i| > \varepsilon_n \right\} < c e^{-c_0 a_n} \to 0 \]

since $a_n \to 0$. Therefore, (42) is true and thus (10).

Next consider $T_n$ as given in (39). $T_n$ is the mean of $n$ independent identically distributed random variables, and hence, letting

\[\mu = E(\xi) \]
\[\sigma^2 = \sigma^2(\xi) \]

where $\sigma^2(\xi)$ is the notation for the variance of $\xi$, we have

\[
\begin{align*}
E(T_n) &= \mu \\
\sigma^2(T_n) &= \frac{\sigma^2}{n} \\
E(\eta_n) &= \sqrt{n}\mu \\
\sigma^2(\eta_n) &= \sigma^2.
\end{align*}
\]

Then by Tchebycheff, for any $\varepsilon > 0$,

\[P \left\{ \left| \frac{T_n - \sqrt{n}\mu}{a_n} \right| > \varepsilon \right\} = P \left\{ \left| \eta_n - \sqrt{n}\mu \right| > \varepsilon a_n \right\} \]
\[\leq \frac{\sigma^2}{\varepsilon^2 a_n^2} \]
\[\to 0 \]
Hence,

\[(\text{II}) \quad \lim_{n \to \infty} \frac{\eta_n - \sqrt{n} \mu}{\sigma_n} = 0\]

Now

\[\frac{\xi_n - \sqrt{n} \mu}{\sigma_n} = \frac{\xi_n - \eta_n}{\sigma_n} \cdot \frac{\eta_n - \sqrt{n} \mu}{\sigma_n}\]

and by (I.4) and (IV) both terms on the right have a probability limit of 0, therefore,

\[\lim_{n \to \infty} \frac{\xi_n - \sqrt{n} \mu}{\sigma_n} = 0\]

and theorem is proved.

Hence we see that for any related sequence, \(\{S_n\}\), in \(\mathcal{F}_p\) the sequence \(\{c_n S_n\}\) will converge in probability to a constant whenever

\[c_n n^{-\frac{1}{2}} \to 0.\]

This result is sharp for there do exist related sequences in \(\mathcal{F}_p\) such that \(\sqrt{n} S_n\) has a non-trivial limiting distribution for samples from some random variables. However, this is not always the case, for there also exist some \(S_n\) such that for samples from some random variables \(\sqrt{n} S_n\) still converges in probability to a constant.

Both of these cases are illustrated by Kendall's \(\tau\).

**Example 11:** In Example 1 Kendall's \(\tau\) was seen to be given by

\[\tau^{(n)} = \frac{\mu n \cdot \tau^{(n-1)}}{n - 1} - \frac{n \cdot 3}{n - 1}\]

where

\[\tau^{(n)} = \frac{1}{n} \sum_{k=1}^{n} f(k)(X(k)).\]
From the theory of U statistics it is known that \( \frac{1}{n^2}(\tau(n) - E(\tau(n))) \) has a limiting normal distribution. Moreover, the mean and the variance of \( \tau(n) \) are expressible in terms of cumulative moments. Let

\[
\sigma^{-12} = 2\left\{ E[2F + 2F^2 + F_1F_2 - 2FF_1 - 2FF_2] - 2[E(F)]^2 - \frac{1}{6} \right\},
\]

then

\[
E(\tau(n)) = 4E(F) - 1
\]

\[
\sigma^{-2}(\tau(n)) = \frac{16}{n(n-1)} \left[ n\sigma^{-12} - 2\sigma^{-12} + E(F) - 2[E(F)]^2 \right].
\]

Thus,

\[
E(\tau(n)) \xrightarrow{\mathcal{N}} 4E(F) - 1
\]

\[
\sigma^{-2}(\sqrt{n}\tau(n)) \xrightarrow{\mathcal{N}} 16\sigma^{-12}
\]

and

\[
E(\tau_s(n)) \xrightarrow{\mathcal{N}} E(F)
\]

\[
\sigma^{-2}(\sqrt{n}\tau_s(n)) \xrightarrow{\mathcal{N}} \sigma^{-12}.
\]

Consequently, \( \frac{1}{n^2}(\tau(n) - 4E(F) + 1) \) is asymptotically normal with mean 0 and variance \( 16\sigma^{-12} \), we write it

\[
n^2(\tau(n) - 4E(F) + 1) \xrightarrow{\mathcal{N}} N(0, 16\sigma^{-12}).
\]

Then

\[
n^2(\tau_s(n) - E(F)) \xrightarrow{\mathcal{N}} N(0, \sigma^{-12}).
\]

Case i) If the sample is drawn from a random variable, \( X \), with independent components, \( X_1 \) and \( X_2 \), then, from the remarks at the
end of Chapter II,

\[ E(F) = \frac{1}{4} \]

also

\[ E(F^2) = \frac{1}{9}, \quad E(F_1 F_2) = \frac{1}{4}, \quad E(F_1 F_2) = E(F_1^2) = E(F_2^2) = \frac{1}{6} \]

and hence

\[ \sigma^{'2} = \frac{1}{36} \]

so

\[ \frac{1}{n^2} (\tau_{s}^{(n)} - \frac{1}{4}) \to N(0, \frac{1}{36}) \]

and \( \tau_{s}^{(n)} \) has a non-trivial limiting distribution.

However, there are degenerate cases where \( \tau_{s}^{(n)} \) does have a unitary limiting distribution.

Case ii) Let the sample be drawn from a random variable, \( X \), with

\[ F_1(x_1) = x_1, \quad 0 \leq x_1 \leq 1 \]
\[ F_2(x_2) = x_2, \quad 0 \leq x_2 \leq 1 \]

with \( X_1 = X_2 \). This random variable has all of its probability along the diagonal of the unit square from \((0,0)\) to \((1,1)\). For this random variable \( F = F_1 = F_2 \) along that diagonal, and from the remarks in Chapter II

\[ E(F) = \frac{1}{2} \]

Since

\[ E(F^2) = E(F_1 F_2) = E(F_1^2) = E(F_2^2) = E(F_1^2) = \frac{1}{3} \]

it follows that

\[ \sigma^{'2} = 0, \]

and \( \tau_{s}^{(n)} \) has a unitary limiting distribution at \( \frac{1}{2} \),
\[ (46) \quad n^2 \left( \tau_s^{(n)} - \frac{1}{2} \right) \to N(0,0). \]

This is necessary because the terms of the sum \( \sum_{k=1}^{n} F(n)(x(k)) \) must be just the numbers \( \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n} \), in some permutation, and

\[ \tau_s^{(n)} = \frac{n + 1}{2n} \]

with probability 1.

Case iii) The other extreme is given by the random variable with all of its probability of the other diagonal of the unit square. Take

\[ F_1(x_1) = x_1, \quad 0 \leq x_1 \leq 1 \]
\[ F_2(x_2) = x_2, \quad 0 \leq x_2 \leq 1 \]

and \( X_1 = 1 - X_2 \). Here, for all possible outcomes, \( F(X) = 0 \), hence

\[ E(F) = E(FF_1) = E(FF_2) = E(F^2) = 0 \]
\[ E(F_1 F_2) = E(F_1 - F_1^2) = \frac{1}{2} - \frac{1}{3} = \frac{1}{6} \]

and

\[ \sigma^{-2} = 0 \]

and again

\[ n^2 \left( \tau_s^{(n)} \right) \to N(0,0). \]

In fact, \( n^2 \tau_s^{(n)} = 0 \) with probability 1 for every \( n \).

Comparison of \( \xi_n \) and \( \eta_n \)

In the investigation of the limiting distribution of \( \xi_n \), as in the last section, one of the obvious questions that arises is when, if
ever, do $\xi_n$ and $\eta_n$ have the same limiting distribution. Of course, $\eta_n$, by the central limit theorem, has a normal limiting distribution, hence anytime $\xi_n$ and $\eta_n$ are the same in the limit then the limiting distribution of $\xi_n$ is known. It is not known when the limiting distributions are the same, however, some insight into the relationship of $\xi_n$ and $\eta_n$ is given by looking at their moments. A lemma that is useful in this respect is proved first. In this section the notation can be shortened somewhat. $X$ is always some fixed random variable in $X^r$ with a distribution function vector, $\mathcal{F}$, and a sample distribution function vector, $\mathcal{F}(n)$. Let

$$q_{k}^{(n)} = q_{p}(\mathcal{F}(n) [X(k)])$$

$$q_{k} = q_{p}(\mathcal{F} [X(k)])$$

$$q = q_{p}(\mathcal{F} [X])$$

Thus,

$$S_n = \frac{1}{n} \sum_{k=1}^{n} q_{k}^{(n)}$$

$$T_n = \frac{1}{n} \sum_{k=1}^{n} q_{k}$$

**Lemma 5.1:** For any fixed $k$ and $p$, a positive integer,

$$\lim_{n} E[(q_{k}^{(n)})^p] = E(q^p).$$

**Proof:** Let

$$S_n^* = \frac{1}{n} \sum_{k=1}^{n} [q_{k}^{(n)}]^p$$

$$H_n(u) = P[S_n^* \leq u].$$
Since $S^* \in \mathcal{S}_p$, we have by the convergence theorem

$$\lim_{n \to \infty} S^*_n = E(Q^p).$$

Hence,

$$\lim_{n \to \infty} H_n(u) = H(u) = \begin{cases} 0, & u < E(Q^p) \\ 1, & u \geq E(Q^p). \end{cases}$$

Moreover, there is a number, $B$, such that $|S^*_n| < B$ for all $n$ with probability 1. Then, using the Helly-Bray lemma,

$$E(S^*_n) = \int_{-B}^{B} u dH_n \to \int_{-B}^{B} u dH = E(Q^p).$$

However,

$$E(S^*_n) = E\left[ \frac{1}{n} \sum_{k=1}^{n} (q_k^{(n)})^p \right] = E\left[ (q_k^{(n)})^p \right]$$

for any fixed $k$, hence,

$$E\left[ (q_k^{(n)})^p \right] \to E(Q^p)$$

and the lemma is proved.

First of all, from the lemma

(45) \hspace{1cm} \lim_{n \to \infty} E(S_n) = E(Q).

Moreover,

$$\sigma^2(S_n) = \frac{1}{n^2} \sigma^2\left( \sum_{k=1}^{n} q_k^{(n)} \right)$$

$$= \frac{1}{n^2} \left[ \sum_{k=1}^{n} \sigma^2(q_k^{(n)}) + \sum_{k=1}^{n} \sum_{j=1}^{n} \sigma^2(q_k^{(n)} q_j^{(n)}) \right]$$

where $\sigma(ZW)$ is the notation for the covariance of $Z$ and $W$. Since for
each $k, q_k^{(n)}$ is identically distributed

$$\sigma^2(S_n) = \frac{1}{n} \sigma^2(q_k^{(n)}) + \frac{n-1}{n} \sigma(q_k^{(n)} q_j^{(n)})$$

for any fixed $k$ and $j \neq k$. But, from the lemma

$$E[q^{(n)^2}] \Rightarrow E(q^2)$$

$$E[q^{(n)}] \Rightarrow E(q)$$

hence

$$\sigma^2(q_k^{(n)}) \Rightarrow \sigma^2(q).$$

Therefore, if the limits exist,

$$(l_6) \quad \lim \sigma^2(\gamma_n) = \lim n \sigma^2(S_n)$$

$$= \sigma^2(q) + \lim (n-1) \sigma(q_k^{(n)} q_j^{(n)})$$

Now, the moments of $\gamma_n$ were given by (l_3) to be

$$E(\gamma_n) = \sqrt{\sigma(q)}$$

$$\sigma^{2}(\gamma_n) = \sigma^2(q).$$

for any $n$. If $\gamma_n$ and $\gamma_n$ are to have their variances converging to the same limit, which amounts to

$$\lim_n \sigma^2(\gamma_n) = \sigma^2(q),$$

then from (l_6) it is necessary and sufficient that

$$(l_7) \quad \lim_n n \sigma(q_k^{(n)} q_j^{(n)}) = 0.$$
Thus, if the dependence between the terms of the sum composing \( \xi_n \) vanish fast enough as \( n \) gets large, then \( \xi_n \) and \( \eta_n \) have the same limits for their variances. The condition in (47) gives the rate at which this must occur.

If, indeed,

\[
\lim_{n \to \infty} E|\left( \xi_n - E(\xi_n) \right|^2 \cdot g | 
\]

does exist and is finite for some \( g > 0 \), then (47) becomes a necessary condition for \( \xi_n \) and \( \eta_n \) to have the same limiting distribution. For if (48) holds then \( \lim \sigma^2(\xi_n) \) must be the variance of the limiting distribution. ([3] p. 184)

Several simple examples will illustrate for some cases that (47) does not hold and \( \xi_n \) and \( \eta_n \) do not have the same limiting distribution.

**Example 12:** In Example 11, \( S_n = \tau_s^{(n)} \), and

\[
\xi_n = \sqrt{n} S_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} F^{(n)}(X(k)) \\
\eta_n = \sqrt{n} T_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} F(X(k)) \\
E(\eta_n) = \sqrt{n} E(F) \\
\sigma^2(\eta_n) = \sigma^2(F).
\]

From above,

\[
\lim_{n \to \infty} E(S_n) = E(F)
\]
and, also
\[ \lim_{n \to \infty} \sigma^2(\xi_n) = \sigma^2(F) \]
if and only if
\[ \lim_{n \to \infty} n \sigma(F^n(x(k)) F(x(j))) = 0 \]
for any fixed \( k \neq j \).

Now in case (ii) of that example, with all the probability on the diagonal from \((0,0)\) to \((1,1)\), the covariance in (49) can easily be computed.
\[ E(F^n(x(k))) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{n} = \frac{n \cdot 1}{2n} \]
and
\[ E(F^n(x(k)) F^n(x(j))) = \frac{1}{n(n-1)} \sum_{i \neq j}^{n} \frac{1}{n^2} \]
\[ = \frac{1}{n^3(n-1)} \left[ \sum_{i=1}^{n} \frac{1}{i} \sum_{j=1}^{n} \frac{1}{j} - \sum_{i=1}^{n} \frac{1}{i^2} \right] \]
\[ = \frac{3n^2 \cdot 5n^2}{12n^2} \cdot \frac{2}{12n^2} . \]

Therefore,
\[ n \sigma^2(F^n(x(k)) F^n(x(j))) \]
\[ = n \left( \frac{3n^2 \cdot 5n^2}{12n^2} - \frac{n^2 \cdot 2n \cdot 1}{4n^2} \right) \]
\[ = \frac{n \cdot 1}{12n} \]
\[ \rightarrow \frac{1}{12} \]
and clearly (47) does not hold. Therefore, the limits of the variances of $\xi_n$ and of $\eta_n$ are not the same. Actually, in this case, $F(X^{(k)})$ is uniformly distributed on the interval $[0,1]$, and

$$\left(\eta_n - \frac{\sqrt{n}}{2}\right) \to N(0, \frac{1}{12}),$$

whereas, we saw in Example 11, (46), that

$$\left(\xi_n - \frac{\sqrt{n}}{2}\right) \to N(0,0)$$

so that $\xi_n$ and $\eta_n$ do not have the same limiting distribution.

However, in case (iii) where the probability was on the other diagonal, both

$$F(X) = 0$$

and $F(n)(X(k)) = 0$

with probability 1, hence

$$\sigma^2(F(n)(X(k)) - F(n)(X(j))) = 0$$

and condition (49) is fulfilled. Moreover,

$$\xi_n \to N(0,0)$$

$$\eta_n \to N(0,0)$$

so, in a trivial fashion, they do have the same limiting distribution.

Case (i) provides an important, non-degenerate, example of when $\xi_n$ and $\eta_n$ do not have the same limiting distribution. In (45) it was seen that

$$\left(\xi_n - \frac{\sqrt{n}}{4}\right) \to N(0, \frac{1}{36}),$$
but

\[
\sigma^2(\eta_n) = \sigma^2(f) \\
= E(F^2) - [E(F)]^2 \\
= \frac{1}{9} - \frac{1}{6} \\
= \frac{7}{114}
\]

and then, by the central limit theorem,

\[
(\eta_n - \frac{\sqrt{n}}{4}) \rightarrow N(0, \frac{7}{114}).
\]


I, Thomas Aloys Willke, was born near Rome City, Indiana, on April 22, 1932. In 1935 I moved to Cincinnati, Ohio, and received my secondary education in the parochial schools there. I received a Bachelor of Arts degree with a major in classical languages from Xavier University, Cincinnati, in 1954, and a Master of Science degree with a major in mathematics from Ohio State University, Columbus, Ohio, in 1956. While in residence at Ohio State University, I served as graduate assistant during the years 1954-55, 1957-58, and instructor 1959-60. For two years beginning in 1955 I served as assistant to Professor D. Ransom Whitney in the Statistics Laboratory of the Department of Mathematics.