THE ASYMPTOTIC DISTRIBUTION OF SOME NON-LINEAR FUNCTIONS OF THE TWO-SAMPLE RANK VECTOR

BY

SIEGFRIED SCHACH

TECHNICAL REPORT NO. 6
FEBRUARY 1, 1968

SUPPORTED BY PUBLIC HEALTH SERVICE
GRANT USPHS-GM-14554-02

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
THE ASYMPTOTIC DISTRIBUTION OF SOME NON-LINEAR
FUNCTIONS OF THE TWO-SAMPLE RANK VECTOR

by

SIEGFRIED SCHACH

Technical Report No. 6
February 1, 1968

Prepared under U. S. Public Health
Service Grant USPHS-CM-14554-02

Reproduction in whole or in part is permitted
for any Purpose of the United States Government

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Preface

The present paper is concerned with a problem that arises in various areas of applied statistics, in particular in the field of biological time series.

The problem consists in testing whether two probability distributions on a circle are identical, when one is not willing to assume that the distributions are members of a given parametric class.

It is well-known that a large number of physiologic functions fluctuate in a 24 hour cycle (e.g., temperature, heart rate, excretion of certain steroids in human beings). The time of the day at which any of these functions attains the maximum depends on the day and on the particular individual. In other words: It can be considered as a random variable which takes its values on the circumference of a circle.

In many cases observations on these maxima have been taken from two different groups: sick and healthy persons, male and female persons, etc. The question then arises as to whether these two samples have the same underlying distribution or whether the distributions are different.

It turns out that for certain types of alternatives tests for this problem should be based on statistics of the form (1.1) of section 1. In order to make the tests available for people studying periodic phenomena, the distribution of the test statistic has to be known, at least for large samples. In the present paper this large sample distribution is obtained for a large class of statistics.
THE ASYMPTOTIC DISTRIBUTION OF SOME NON-LINEAR
FUNCTIONS OF THE TWO-SAMPLE RANK VECTOR

by

Siegfried Schach

1. Introduction.

Let \( X_1, \ldots, X_m \) and \( Y_1, \ldots, Y_n \) be independent samples from
the same continuous distribution. Let \( R_{N_i} \) (\( i = 1, 2, \ldots, m \)) be the
rank of \( X_i \) in the combined sample \( (N = m + n) \), and define

\[
Z_i = Z_i^{(N)} = \begin{cases} 
1 & \text{if the } i \text{th element of the combined} \\
& \text{ordered sample is a } X \text{ observation} \\
0 & \text{otherwise} .
\end{cases}
\]

Let \( A^{(N)} = (a_{i,j}^{(N)}) \) be a sequence of symmetric matrices. We find
conditions under which a statistic of the form

\[
S_N = \sum_{i=1}^{m} \sum_{j=1}^{m} a_{i,j}^{(N)} \sum_{i=1}^{N} \sum_{j=1}^{N} a_{i,j}^{(N)} Z_i Z_j
\]

converges in distribution (after suitable standardization), as \( N \to \infty \).
Examples of statistics of the form (1.1) are given in section 5.

2. Step Function Representation. Degenerate and Nondegenerate
Limit Functions.

Let \( h_N(\cdot, \cdot) \) be a function on the unit square which is constant on
sub-squares of the form \( \frac{i-1}{N} \leq x \leq \frac{i}{N} \), \( \frac{j-1}{N} \leq y \leq \frac{j}{N} \) and which is symmetric, i.e., \( h_N(x,y) = h_N(y,x) \). If we set \( h_N(\frac{i}{N}, \frac{j}{N}) = e_{i,j}^{(N)} \) then obviously every statistic of the form (1.1) can be written as

\[
S_N = \sum_{i=1}^{m} \sum_{j=1}^{m} h_N(\frac{R_{N_i}}{N}, \frac{R_{N_j}}{N}) = \sum_{i=1}^{N} \sum_{j=1}^{N} h_N(\frac{i}{N}, \frac{j}{N})Z_iZ_j
\]

and vice versa. Since conditions on convergence can be stated more easily in terms of sequences \( \{h_N\} \) we will use representation (2.1) in the sequel.

Assume that there exists a function \( h(\cdot, \cdot) \) such that \( h_N \to h \) in \( L_2(\mathbb{I}) \), i.e., in the sense of convergence in the Hilbert space of square integrable functions on the unit square. By Fubini's theorem the function

\[
g(x) = \int_0^1 h(x,y)dy
\]

exists a.e. For convenience we assume that

\[
\int_0^1 \int_0^1 h_N(x,y)dxdy = 0
\]

for each \( N \) (this can always be achieved by a suitable standardization of \( S_N \)). Then \( \int_0^1 \int_0^1 h(x,y)dxdy = 0 \).

1) Such a function is a step function. Whenever we talk about step functions (of two variables) we mean functions which are constant on squares of the form \( \frac{i-1}{N} \leq x \leq \frac{i}{N} \), \( \frac{j-1}{N} \leq y \leq \frac{j}{N} \). We use a corresponding terminology for functions of one variable.
Definition: We say \( h(\cdot, \cdot) \) is degenerate if \( g(x) = 0 \) a.e.

3. **Asymptotic Distribution of \( S_N \) if \( h \) is Nondegenerate.**

Lemma (3.1): If \( h_N(\cdot, \cdot) \to h(\cdot, \cdot) \) in \( L_2(\mathbb{R}) \), then

\[
\int_0^1 h_N(\cdot, y) \, dy \to \int_0^1 h(\cdot, y) \, dy
\]

in \( L_2 \) (i.e., in the space of square integrable functions on \([0,1]\)).

Proof: Define \( g_N(x) = \int_0^1 h_N(x, y) \, dy \). Then, using Schwarz's inequality,

\[
\int_0^1 (g_N(x) - g(x))^2 \, dx = \int_0^1 \left[ \int_0^1 h_N(x, y) \, dy - \int_0^1 h(x, y) \, dy \right]^2 \, dx
\]

\[
= \int_0^1 \left[ \int_0^1 (h_N(x, y) - h(x, y)) \, dy \right]^2 \, dx \leq \int_0^1 \int_0^1 (h_N(x, y) - h(x, y))^2 \, dx \, dy \to 0.
\]

If we set

\[
(3.2) \quad h'(x, y) = h(x, y) - g(x) - g(y)
\]

then \( h'(x, y) \) is symmetric and degenerate, since

\[
(3.3) \quad \int_0^1 h'(x, y) \, dy = \int_0^1 h(x, y) \, dy - g(x) - \int_0^1 g(y) \, dy = 0.
\]

Similarly we define

\[
(3.4) \quad g_N(x) = \int_0^1 h_N(x, y) \, dy,
\]

then \( g_N(\cdot) \) is a step function and
\[(3.5) \quad h_N'(x,y) = h_N(x,y) - g_N(x) - g_N(y)\]

is a symmetric degenerate step function in \(L_2(\mathbb{D})\). Now set

\[(3.6) \quad T_N = \sum_{i=1}^{N} \sum_{j=1}^{N} h_N'(\frac{i}{N}, \frac{j}{N})Z_i Z_j,\]

then

\[(3.7) \quad S_N = T_N + \sum_{i=1}^{N} \sum_{j=1}^{N} g_N(\frac{i}{N})Z_i Z_j + \sum_{i=1}^{N} \sum_{j=1}^{N} g_N(\frac{j}{N})Z_i Z_j \]

\[= T_N + 2m \sum_{i=1}^{N} g_N(\frac{i}{N})Z_i \]

\[= T_N + 2m U_N\]

where

\[(3.8) \quad U_N = \sum_{i=1}^{N} g_N(\frac{i}{N})Z_i = \sum_{i=1}^{N} \frac{R_i}{N}.\]

We always assume that \(\frac{m}{N} \to \lambda, \ 0 < \lambda < 1\).

**Theorem (3.2):** If \(h_N(\cdot, \cdot) \to h(\cdot, \cdot)\) then \(\frac{U_N}{\sqrt{N}}\) is asymptotically normal with

\[(3.9) \quad \mu_N = \frac{1}{\sqrt{N}} \mathbb{E} U_N = 0, \quad \sigma^2 = \lambda(1-\lambda) \int_0^1 g(x)^2 dx.\]

**Proof:** This follows immediately from Hájek's theorem V.1.6.a (Hájek and Sidák (1967), page 163) provided that

\[(3.10) \quad g_N(\cdot) \to g(\cdot) \text{ in } L_2, \text{ as } N \to \infty.\]

But by lemma (3.1) this condition is satisfied.
Lemma (3.3): If \( h_N \overset{L^2}{\rightarrow} h \), then \( h'_N \overset{L^2}{\rightarrow} h' \).

Proof: By lemma (3.1) \( g_N \overset{L^2}{\rightarrow} g \). Since
\[
h'_N(x,y) - h'(x,y) = h_N(x,y) - h(x,y) + g_N(x) - g(x) + g_N(y) - g(y)
\]
the desired result follows.

Lemma (3.4): If \( h_N \overset{L^2}{\rightarrow} h \), \( \int_0^1 h_N(x,x)^2 dx \) is bounded and
\[
\sum_{i=1}^{N} h_N^j \left( \frac{i}{N}, \frac{j}{N} \right) = 0 \text{ for each } j \text{ (i.e. } \int_0^1 h_N(x,y) dx = 0) \]
then for \( S_N \) defined by (2.1) we get
\[
\frac{1}{N} E S_N = \frac{m}{N} (1 - \frac{m-1}{N-1}) \int_0^1 h_N(x,x) dx = \lambda (1- \lambda) \int_0^1 h(x,x) dx + o_N(1),
\]
(3.12) \[ \frac{1}{N^2} \text{ Var } S_N = 2\lambda^2 (1- \lambda)^2 \|h\|_{L^2(\square)}^2 + o_N(1). \]

Proof: Obviously,
\[
E Z_i Z_j = \begin{cases} \frac{m}{N} & \text{if } i = j \\ \frac{m}{N} \frac{m-1}{N-1} & \text{if } i \neq j. \end{cases}
\]
Hence
\[
E S_N = \sum_{i=1}^{N} \sum_{j=1}^{N} h_N^j \left( \frac{i}{N}, \frac{j}{N} \right) E Z_i Z_j
\]
\[
= \frac{m}{N} \frac{m-1}{N-1} \sum_{i=1}^{N} \sum_{j=1}^{N} h_N \left( \frac{i}{N}, \frac{j}{N} \right) + \frac{m}{N} \frac{m-1}{N-1} \sum_{i=1}^{N} h_N \left( \frac{i}{N}, \frac{1}{N} \right)
\]
\[
= N \cdot \frac{m}{N}(1 - \frac{m-1}{N-1}) \int_0^1 h_N(x,x) dx.
\]
Since \( \int_0^1 h_N(x,x)dx \) is bounded \( (\int_0^1 h_N(x,x)dx)^2 \leq \int_0^1 h_N(x,x)^2dx \) and since \( \frac{m}{N} \to \lambda \) (3.11) follows.

A straightforward, but tedious, calculation (using (3.10) repeatedly) shows that

\[
(3.15) \quad E S_N^2 = 2N^2 \lambda^2 (1-\lambda)^2 \|h_N(\cdot,\cdot)\|_{L_2}^2 + N^2 \lambda^2 (1-\lambda)^2 (\int_0^1 h_N(x,x)dx)^2 + o(N).
\]

Hence

\[
\frac{1}{N^2} \text{Var } S_N = 2\lambda^2 (1-\lambda)^2 \|h\|_{L_2}^2 + o_N(1).
\]

Theorem (3.5): If \( h_N \overset{L_2}{\to} h \), \( h \) nondegenerate, \( \int h_N(x,x)^2dx \) bounded, then for \( S_N \) defined by (2.1) we have

\[
(3.16) \quad \lim_{N \to \infty} \text{P}[N^{-3/2} S_N \leq x\sigma] = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt
\]

with

\[
(3.17) \quad \sigma^2 = 4\lambda^3 (1-\lambda) \|g\|_{L_2}^2,
\]

where \( g(\cdot) \) is defined by (2.2).

Proof: By (3.7) we have

\[
S_N = T_N + 2mU_N.
\]

\( T_N \), defined by (3.6), is based on the function \( h'_N \), defined by (3.5).

By lemma (3.3) \( h'_N \overset{L_2}{\to} h' \), and thus \( h'_N \) satisfies the assumptions of lemma (3.4). Hence by (3.11) and (3.12),
\[(3.18) \quad N^{-3/2} E N^2_{-N} \to 0 \quad \text{as} \quad N \to \infty ,\]

which implies that \(N^{-3/2} T_N \to 0\) in probability. From theorem (3.4) we obtain asymptotic normality of \(N^{-3/2} m U_N / (\lambda^2 (1-\lambda) \|g\|_L^2 )\). Thus \(N^{-3/2} S_N\) is asymptotically \(N(0, \sigma^2)\), where \(\sigma^2\) is defined by (3.17).

**Remark (3.6):** Under the conditions of theorem (3.6) the asymptotic distribution of \(S_N\) depends on \(h_N\) only through the "marginals" \(g_N^i\); in fact it depends only on \(\|g\|^2 = \lim_{N \to \infty} \|g_N^i\|^2\).

4. **Asymptotic Distribution of \(S_N\) if \(h\) is Degenerate.**

In this section we obtain the asymptotic distribution of \(S_N\) if \(h\) is degenerate, provided that certain conditions on \(h\) and \(\omega\) the approximating sequence \(h_N\) are satisfied.

Let \(\{h_N(\cdot, \cdot)\}\) be a sequence of symmetric step function, such that
\[
\int_0^1 \int_0^1 h_N(x,y)dx dy = 0 \quad \text{for all} \quad N, \quad \text{and let} \quad h_N \to h \quad \text{in} \quad L_2(\Omega) \quad \text{norm.}
\]
Then by the definition of degeneracy
\[
(4.1) \quad \int_0^1 h(x,y)dy = 0 \quad a.e.
\]
Again define
\[
(4.2) \quad g_N(x) = \int_0^1 h_N(x,y)dy .
\]
By lemma (3.1) \(g_N \overset{L_2}{\to} 0\).

In order to make the limiting distribution independent of the particular approximating sequence we need somewhat stronger convergence conditions.
Lemma (4.1): Let the following conditions be satisfied: As \( N \to \infty \)

(4.3) \[ h_N \to 0 \text{ in } L^2 \]

(4.4) \[ N\|g_N\|^2 = N \int_0^1 g_N(x)^2 \, dx \to 0 \]

(4.5) \[ \int_0^1 h_N(x,x) \, dx \to 0 \]

(4.6) \[ |\int_0^1 h_N(x,x) \, dx| \leq K \text{ for some constant } K. \]

Then for \( S_N \) defined by (2.1) we have

(4.7) \[ N^{-1} S_N \to 0 \text{ in mean square}. \]

Proof: Define

(4.8) \[ h'_N(x,y) = h_N(x,y) - g_N(x) - g_N(y) \]

and let \( T_N \) be the statistic based on \( h'_N \), i.e.

(4.9) \[ T_N = \sum_{i=1}^N \sum_{j=1}^N h'_N(y_{i,N}, y_{i,j,N}) \, Z_i \, Z_j. \]

By direct calculation

(4.10) \[ \frac{1}{N^2} \mathbb{E}(S_N - T_N)^2 = \frac{4}{N^2} \frac{(m)}{N} \frac{N-m}{N-1} N \int_0^1 g_N(x)^2 \, dx \to 0, \]

since \( N \int_0^1 g_N(x)^2 \, dx \to 0. \)
By lemma (3.3) \( h_N^! L_2 \to h = 0 \). Also

\[
(4.11) \quad \left( \int_0^1 h_N^!(x,x)^2 dx \right)^{1/2} \leq \left( \int_0^1 h_N(x,x)^2 dx \right)^{1/2} + 2 \left( \int_0^1 \varepsilon(x)^2 dx \right)^{1/2} \leq K+1
\]

for \( N \) large enough. Since \( h_N^! \) satisfies (3.10) we may apply lemma (3.4) and obtain

\[
(4.12) \quad \frac{1}{N} E_T^N = \lambda(1-\lambda) \int_0^1 h_N^!(x,x) dx + o_N(1) = \lambda(1-\lambda) \int_0^1 h_N(x,x) dx + o_N(1) \to 0 \quad \text{by assumption (4.5)}, \quad \text{and}
\]

\[
(4.13) \quad \frac{1}{N^2} ET_N^2 = 2\lambda^2(1-\lambda)^2 \|a\| + o_N(1) \to 0.
\]

Thus \( \frac{1}{N} T_N \to 0 \) in mean square, and by (4.10) we obtain finally

\[
\frac{1}{N} S_N \to 0 \quad \text{in m.s.}
\]

**Remark (4.2):** If we take approximating sequences satisfying (3.10), then (4.4) is satisfied.

**Corollary (4.3):** Let \( \{h_N^!\}, \{h_N^\prime\} \) be sequences of step functions converging to \( h \) (in \( L_2 \)). Let \( S_N^! \) and \( S_N^\prime \) be the statistics defined by (2.1), based on \( h_N^! \) and \( h_N^\prime \) respectively.

If \( h_N = h_N^! - h_N^\prime \) converges to 0 fast enough to satisfy (4.4), (4.5), (4.6), then \( N^{-1} S_N^! \) and \( N^{-1} S_N^\prime \) have the same limiting distribution, provided that either one of them converges in law.

**Proof:** By lemma (4.1) \( N^{-1}(S_N^! - S_N^\prime) \to 0 \) in m.s., and hence the desired result follows.
Let \( h(\cdot, \cdot) \) be a symmetric element of \( L_2(\square) \). Then it is well-known that the relation \( H: f \to g \) defined by

\[
(4.14) \quad g(x) = \int_0^1 h(x,y) f(y) dy
\]

is a compact operator which maps \( L_2 \) into \( L_2 \). The spectrum of \( H \) consists of a countable number of real eigenvalues \( \lambda_k \), such that

\[
\sum_{k=1}^{\infty} \lambda_k^2 < \infty.
\]

Let \( \{\varphi_k\} \) be a corresponding sequence of eigenfunctions. Then by the spectral theorem

\[
(4.15) \quad h(x,y) = \sum_{k=1}^{\infty} \lambda_k \varphi_k(x)\varphi_k(y), \quad \int_0^1 \varphi_k(x)\varphi_k(x) dx = \delta_{kk}.
\]

**Definition (4.4):** We say \( h(\cdot, \cdot) \) is a trace-class function if

\[
(4.16) \quad \sum_{k=1}^{\infty} |\lambda_k| < \infty.
\]

**Lemma (4.5):** If \( h \) is degenerate, then \( \int_0^1 \varphi_k(x) dx = 0 \) for all \( k \).

**Proof:** Let \( \int_0^1 \varphi_k(x) dx = b_k \). In the expansion \( (4.15) \) we may assume w.l.o.g. \( \lambda_k \neq 0 \) for all \( k \). Then we have a.e.

\[
(4.17) \quad 0 = \int_0^1 h(x,y) dy = \int_0^1 \sum_{k=1}^{\infty} \lambda_k \varphi_k(x)\varphi_k(y) dy = \sum_{k=1}^{\infty} \lambda_k b_k \varphi_k(x).
\]

Since \( \{\varphi_k(\cdot)\} \) is orthonormal we must have \( \lambda_k b_k = 0 \) and hence \( b_k = 0 \) for all \( k \).

If \( h \) is any element of \( L_2(\square) \), then the projection \( h_N^P \) of \( h \) onto the space of step functions of order \( N \) is characterized by
\[ (4.18) \quad h_N^P(\frac{i}{N}, \frac{j}{N}) = N^2 \int_{\frac{i-1}{N}}^{\frac{i}{N}} \int_{\frac{j-1}{N}}^{\frac{j}{N}} h(x,y) \, dx \, dy, \quad 1 \leq i, j \leq N. \]

Using the continuity of the projection operator it is easy to see that from the expansion (4.15) we get

\[ (4.19) \quad h_N^P(x,y) = \sum_{k=1}^{\infty} \lambda_k \varphi_k^N(x) \varphi_k^N(y) \]

where \( \varphi_k^N(\cdot) \) is the projection of \( \varphi_k(\cdot) \) onto the space of step function of one variable, i.e.,

\[ (4.20) \quad \varphi_k^N(\frac{i}{N}) = N \int_{\frac{i-1}{N}}^{\frac{i}{N}} \varphi_k(x) \, dx, \quad 1 \leq i \leq N. \]

**Lemma (4.6):** \( h_N^P \to h \) in \( L_2(\mathbb{R}) \) as \( N \to \infty \). \( \varphi_k^N \to \varphi_k \) for each \( k \), as \( N \to \infty \).

**Proof:** We prove the first assertion. The proof of the second one is identical. Let \( \varepsilon > 0 \) be given. Since the continuous functions are dense in \( L_2(\mathbb{R}) \) there exists a continuous function \( g \in L_2(\mathbb{R}) \) such that \( \| h - g \| \leq \frac{\varepsilon}{2} \). Let \( \{ g_N^P \} \) be the corresponding sequence of projections. By uniform continuity of \( g \) it follows that there exists \( N_\varepsilon \) such that

\[ \sup_{0 \leq x, y \leq 1} | g_N^P(x,y) - g(x,y) | \leq \frac{\varepsilon}{2} \quad \text{for} \quad N \geq N_\varepsilon. \]

Hence \( \| g_N^P - g \| \leq \frac{\varepsilon}{2} \). The functions \( h_N^P \), and \( \| g_N^P - h_N^P \| \leq \| g_N^P - g \| + \| g - h \| \leq \varepsilon \). But since \( h_N^P \) is the projection we must have \( \| h_N^P - h \| \leq \| g_N^P - h \| \leq \varepsilon \) for \( N \geq N_\varepsilon \), q.e.d.

Set

\[ (4.21) \quad \nu_k = \sqrt{|\lambda_k|}, \quad k = 1,2,\ldots \]
\[ (4.22) \quad \eta_k^N = \frac{1}{\sqrt{N}} \nu_k \sum_{i=1}^{N} \varphi_k \left( \frac{i}{N} \right) Z_i, \quad k = 1, 2, \ldots \]

**Theorem (4.7):** Let \( \{ \varphi_k : k=1, 2, \ldots, K \} \) be an orthonormal set in \( L_2 \).

Define the step functions \( \varphi_k^N \) by (4.20). Then as \( N \to \infty \) the joint distribution of \( (\eta_1^N, \eta_2^N, \ldots, \eta_K^N) \) is asymptotically normal with mean vector \( \mathbf{0} \) and covariance matrix \( (\sigma_{k\ell}) \), where \( \sigma_{k\ell} = \lambda(1-\lambda) \delta_{k\ell} \nu_2^2 \).

**Proof:** It suffices to show that any linear combination \( \xi_N = \sum_{k=1}^{K} t_k \eta_k^N \) is asymptotically normal with \( \mathbb{E} \xi_N = 0, \) \( \text{Var}\xi_N \to \lambda(1-\lambda) \sum_{k=1}^{K} t_k^2 \nu_k^2 \).

\[ (4.23) \quad \xi_N = \frac{1}{\sqrt{N}} \sum_{k=1}^{K} t_k \nu_k \sum_{i=1}^{N} \varphi_k \left( \frac{i}{N} \right) Z_i = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{k=1}^{K} t_k \nu_k \varphi_k \left( \frac{i}{N} \right) Z_i. \]

Obviously \( \mathbb{E}\xi_N = 0 \). By lemma (4.6) we get \( \sum_{k=1}^{K} t_k \nu_k \varphi_k (\cdot) \to \sum_{k=1}^{K} t_k \nu_k \varphi_k (\cdot) \) in \( L_2 \). Hence Hájek's theorem mentioned on page 4 implies asymptotic normality of \( \xi_N \) with \( \mu = 0 \) and

\[ \sigma^2 = \lambda(1-\lambda) \int_{0}^{1} \left( \sum_{k=1}^{K} t_k \nu_k \varphi_k (x) \right)^2 dx. \]

Since \( \{ \varphi_k \} \) is orthonormal we get

\[ \sigma^2 = \lambda(1-\lambda) \sum_{k=1}^{K} t_k^2 \nu_k^2, \quad \text{q.e.d.} \]

**Corollary (4.8):** If \( h(\cdot, \cdot) \) is a degenerate function with a finite expansion

\[ (4.24) \quad h(x, y) = \sum_{k=1}^{K} \lambda_k \varphi_k (x) \varphi_k (y), \]

then the asymptotic distribution of

\[ (4.25) \quad N^{-1} S_N = N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} h \left( \frac{i}{N}, \frac{j}{N} \right) Z_i Z_j \]
exists and is equal to the distribution of $\sum_{k=1}^{K} \lambda_k \chi_k^2(k)$, where $\chi_k^2(k)$, $k=1,2,\ldots,K$ are independent $\chi^2$ random variables with 1 d.f.

**Proof:** Because of (4.19) we have

$$N^{-1} \mathbf{S}_N = N^{-1} \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_k \phi_k^N(i) \phi_k^N(j) Z_i Z_j$$

$$= \sum_{k=1}^{N} \text{sign}(\lambda_k)(\eta_k^N)^2,$$

and hence the result follows immediately from theorem (4.7).

**Theorem (4.9):** Let $h$ be a symmetric trace-class function. If

$$h_N(x,y) = \sum_{k=1}^{N} \lambda_k \phi_k^N(x) \phi_k^N(y),$$

then $h_N \to h$ in $L_2(\mathbb{R})$.

**Proof:** Let $\epsilon > 0$ be given. Then there exists a $K$ such that

$$\sum_{k=K+1}^{\infty} |\lambda_k| \leq \frac{\epsilon}{12}.$$  For each $N$ we have

$$\|\phi_k(x)\phi_k(y) - \phi_k^N(x)\phi_k^N(y)\| = \|\phi_k(x)\phi_k(y) - \phi_k(x)\phi_k^N(y) + \phi_k(x)\phi_k^N(y) - \phi_k^N(x)\phi_k^N(y)\|$$

$$\leq \|\phi_k\|\|\phi_k - \phi_k^N\| + \|\phi_k^N\|\|\phi_k - \phi_k^N\|.$$

Since $\phi_k^N$ is a projection of $\phi$ we have $\|\phi_k^N\| \leq \|\phi\| = 1$. Hence

$$\|\phi_k(x)\phi_k(y) - \phi_k^N(x)\phi_k^N(y)\| \leq \frac{\epsilon}{12}$$

for all $N$. By lemma (4.6) we have

$$\|\phi_k(x)\phi_k(y) - \phi_k^N(x)\phi_k^N(y)\| = 0$$

for each $k$, as $N \to \infty$. Let $N_\epsilon$ be such that

$$\sum_{k=1}^{K} |\lambda_k|\|\phi_k(x)\phi_k(y) - \phi_k^N(x)\phi_k^N(y)\| \leq \frac{\epsilon}{3}$$

for $N \geq N_\epsilon$. Then
\[ \|h(x,y) - h_N(x,y)\| = \left\| \sum_{k=1}^{\infty} \lambda_k \phi_k(x) \varphi_k(y) - \sum_{k=1}^{N} \lambda_k \phi_k(x) \varphi_k(y) \right\| \]

\[ \leq \left\| \sum_{k=N+1}^{\infty} \lambda_k \phi_k(x) \varphi_k(y) \right\| + \left\| \sum_{k=K+1}^{N} \lambda_k (\phi_k(x) \varphi_k(y) - \phi_k(x) \varphi_k(y)) \right\| \]

\[ + \left\| \sum_{k=1}^{K} \lambda_k (\phi_k(x) \varphi_k(y) - \phi_k(x) \varphi_k(y)) \right\| \leq \sum_{k=N+1}^{\infty} |\lambda_k| + 4 \sum_{k=K+1}^{N} |\lambda_k| + \frac{\epsilon}{3} \]

\[ \leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} \text{ for } N \geq \max(N_\epsilon, K), \text{ q.e.d.} \]

We use the theory of measures on separable Hilbert spaces in order to derive the limiting distribution of \( N^{-1}S_N \), where \( S_N \) is based on a sequence \( h_N \) defined by (4.27). Let \( H \) be the Hilbert space of real sequences with finite sums of squares. Then we define a sequence \( \{M_N\} \) of mappings from the space of the \( Z \)'s to \( H \) by

\[ (4.28) \quad M_N: (Z_1, \ldots, Z_n) \rightarrow (\eta_N^n, \ldots, \eta_N^n, 0, 0, \ldots) \]

where \( \eta_N \) is defined by (4.22). These mappings are measurable and hence they induce a sequence \( \{\mu_N\} \) of probability measures on \( H \).

Recall that the set of probability measures on \( H \) can be metrized in such a way that convergence in this metric is equivalent to weak convergence. With this metric the space of all probability measures on \( H \) is a complete separable metric space (Prokhorov (1956)).

**Theorem (4.10):** The sequence \( \{\mu_N\} \) of probability measures is sequentially compact.

**Proof:** According to Prokhorov's theorem 1.13 (Prokhorov (1956)) it suffices to show that
(i) \[ \sup_{N} \int_{H} \|x\|^2 \, d\mu_{N} < \infty \]

(ii) \[ \lim_{L \to \infty} \sup_{N} \int_{H} \sum_{\ell=L}^{\infty} x_{\ell}^2 \, d\mu_{N} = 0, \]

where \( x = (x_1, x_2, \ldots) \) is a generic element of \( H \). Since \( x_\ell = 0 \) a.e. \( (\mu_N) \) for \( \ell > N \), we have

\[ (4.29) \quad \int_{H} \sum_{\ell=L}^{\infty} x_{\ell}^2 \, d\mu_{N} = \int_{H} \sum_{\ell=L}^{N} x_{\ell}^2 \, d\mu_{N}. \]

For \( k \leq N \) we have

\[ (4.30) \quad \int_{H} x_{k}^2 \, d\mu_{N} = E(\eta_{k}^2) = \frac{1}{N} \sum_{i=1}^{N} \varphi_{k}^2 \sum_{j=1}^{N} \varphi_{k}^N(z_{i}^N) \varphi_{k}^N(z_{j}^N) \]

\[ = \varphi_{k}^2 m \left( \frac{m-1}{N-1} \right) \frac{1}{N} \sum_{i=1}^{N} \varphi_{k}^N(z_{i}^N) \leq \varphi_{k}^2 \| \varphi_{k}^N \|^2 \leq |\lambda_k|. \]

Combining (4.29) and (4.30) we obtain

\[ \int_{H} \sum_{\ell=L}^{\infty} x_{\ell}^2 \leq \sum_{\ell=L}^{\infty} |\lambda_{\ell}|, \]

which shows that (i) and (ii) are satisfied. This completes our proof.

Let \( (\cdot, \cdot) \) be the inner product on \( H \). For every \( f \in H \) and every probability measure \( \mu \) we define the characteristic functional

\[ (4.31) \quad \xi(f, \mu) = \int_{H} e^{i(f, x)} \, d\mu(x). \]

\( \xi(\cdot, \mu) \) is continuous on \( H \), and it determines \( \mu \) uniquely.

**Theorem (4.11):** \( \mu_{N} \) converges to the Gaussian measure with mean 0 and S-operator of the form \( [S]_{k, \ell} = \lambda(1-\lambda) \delta_{k, \ell} |\lambda_k|. \)
Proof: If \( \mu_N \to \mu \) in the sense of weak convergence, then obviously

\[ 1 = \mu_N(H) \to \mu(H), \quad \text{so that any limit has to be a probability measure.} \]

Let \( f^{(L)} = (f_1, f_2, \ldots, f_L, 0, 0, \ldots) \), then by theorem (4.7)

\begin{equation}
(4.32) \quad x(f^{(L)}, \mu_N) \to \exp\left(-\frac{1}{2} \lambda (1-\lambda) \sum_{k=1}^{L} |\lambda_k| f_k^2 \right), \quad \text{as } N \to \infty.
\end{equation}

If \( \mu \) is any limit measure of a suitably chosen subsequence, then by

the definition of weak convergence we must have

\begin{equation}
(4.33) \quad x(f^{(L)}, \mu) = \exp\left(-\frac{1}{2} \lambda (1-\lambda) \sum_{k=1}^{L} |\lambda_k| f_k^2 \right).
\end{equation}

Since the set of \( f^{(L)} \) elements is dense in \( H \), and since the left hand

side of (4.33) is continuous in the first argument, we must have

\begin{equation}
(4.34) \quad x(f, \mu) = \exp\left(-\frac{1}{2} \lambda (1-\lambda) \sum_{k=1}^{\infty} |\lambda_k| f_k^2 \right).
\end{equation}

The right hand side of this last equation characterizes \( \mu \) uniquely.

Hence any convergent subsequence of \( \{\mu_N\} \) converges to the same limit,

and since \( \{\mu_N\} \) is sequentially compact, we have \( \mu_N \to \mu \). (4.31) shows

that \( \mu \) is a Gaussian measure, which has mean 0 and S-operator of the

desired form.

Theorem (4.12): If \( h \) is a symmetric trace-class function and if \( S_N \)

is based on \( h_N \) defined by (4.27), then \( N^{-1} S_N \) converges in distribution

to the distribution with characteristic function

\begin{equation}
(4.35) \quad \varphi(t) = \prod_{k=1}^{\infty} \left(1 - 2i\lambda(1-\lambda)\lambda_k t\right)^{-1/2}.
\end{equation}
Proof: Because of (4.19) we have

\[ N^{-1} S_N = N^{-1} \sum_{k=1}^{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_{i,j} N^{i,j} \phi_k^{N}(\frac{i}{N}) \phi_k^{N}(\frac{j}{N}) Z_i Z_j \]

\[ = \sum_{k=1}^{N} \text{sign}(\lambda_k)(\eta_k^N)^2. \]

Let \( \varphi_N(\cdot) \) be the characteristic function of the distribution of \( N^{-1} S_N \). Then

\[ \varphi_N(t) = E_{\mu_N} \exp(it \sum_{k=1}^{\infty} \delta_{k} x_k^2), \text{ where } \delta_k = \text{sign} \lambda_k. \]

By the definition of weak convergence

\[ \varphi_N(t) = E_{\mu_N} \exp(it \sum_{k=1}^{\infty} \delta_{k} x_k^2) \Rightarrow E_{\mu} \exp(it \sum_{k=1}^{\infty} \delta_{k} x_k^2). \]

But this limit is continuous in \( t \), and hence \( N^{-1} S_N \) converges in distribution. Furthermore

\[ \varphi(t) = E_{\mu} \exp(it \sum_{k=1}^{\infty} \delta_{k} x_k^2). \]

We now evaluate \( \varphi(t) \). From theorem (4.11) it follows that for each finite \( K \)

\[ E_{\mu_N} \exp(it \sum_{k=1}^{K} \delta_{k} x_k^2) \rightarrow \prod_{k=1}^{K} (1-2i\lambda_k(1-\lambda_k \lambda_k) t)^{-1/2} = E_{\mu} \exp(it \sum_{k=1}^{K} \delta_{k} x_k^2). \]

By the dominated convergence theorem we can pass to the limit in \( K \) and get

\[ \varphi(t) = E_{\mu} \exp(it \sum_{k=1}^{\infty} \delta_{k} x_k^2) = \prod_{k=1}^{\infty} (1-2i\lambda_k(1-\lambda_k \lambda_k) t)^{-1/2}. \]

This finishes our proof.
So far we have studied the asymptotic behavior of $N^{-1}S_N$ only for the particular approximating sequence $\{h_N\}$ defined by (4.27). Since this is our standard sequence we will denote it by $\{h_N^s\}$, i.e.,

$$(4.41) \quad h_N^s(x,y) = \sum_{k=1}^{N} \lambda_k \varphi_k(x) \varphi_k(y).$$

By theorem (4.9) we have $h_N^s \to h$ in $L_2(\square)$, as $N \to \infty$. Since the diagonal of the unit square has Lebesgue measure 0, we can not, in general, expect that $h_N^s(x,x) \to h(x,x)$ in any sense.

We now find conditions under which the statistic $N^{-1}S_N^s$, based on an arbitrary sequence $\{h_N\}$, converges to the same law as the particular one based on $\{h_N^s\}$.

Theorem (4.13): Let $h(\cdot, \cdot)$ be a symmetric trace-class function. Let $h_N \to h$ in $L_2(\square)$ such that

$$(4.42) \quad h_N(x,x) - h_N^s(x,x) \to 0 \text{ in } L_2[0,1], \text{ as } N \to \infty$$

$$(4.43) \quad N\int_0^1 g_N(x)^2 \, dx \to 0 \text{ as } N \to \infty.$$

Then, if $S_N$ and $S_N^s$ are the statistics corresponding to $h_N$ and $h_N^s$, respectively, we have

$$(4.44) \quad N^{-1}(S_N - S_N^s) \to 0 \text{ in mean square}.$$

Proof: We apply lemma 4.1 to $h_N' = h_N - h_N^s$. By theorem (4.9) $h_N^s \to h$ in $L_2(\square)$, and hence $h_N' \to 0$ in $L_2(\square)$. Since
\[ (4.45) \quad \int_0^1 h^s_N(x,y)\,dy = \sum_{k=1}^{N} \lambda_k \varphi_k(x) \int_0^1 \varphi_k(y)\,dy = 0 \]

for all \( x \) and all \( N \), we have

\[ (4.46) \quad g^*_N(x) = \int_0^1 (h_N(x,y) - h^s_N(x,y))\,dy = g_N(x), \]

and hence

\[ (4.47) \quad N\|g^*_N\|^2 \to 0 \quad \text{by assumption (4.43)}. \]

Finally (4.5) and (4.6) are satisfies by (4.42).

**Remark (4.14):** Condition (4.42) seems to be satisfies in all cases of practical interest (provided that \( h(x,x) \) is defined in some natural way, e.g. by continuity), because we usually have

\[ h_N(x,x) \to h(x,x) \quad \text{and} \quad h^s_N(x,x) \to h(x,x) \quad \text{in} \ L_2[0,1]. \]

5. **Examples.**

Wheeler and Watson (1964) proposed a test for equality of two circular distributions which is based on a statistic of the form (1.1). In their case \( h_N(x,y) \) of (2.1) is equal to \( \cos(x-y) \). Matthes and Truax (1965) obtained a statistic of the form (1.1) when deriving a locally most powerful invariant test for two-sided shift alternatives. It has been shown (Schach (1967)) that in detecting rotation alternatives for circular distributions a locally most powerful invariant test under a suitable group of transformations is based on a statistic of the form (2.1), where \( h_N(x,y) \) is a function of the difference \( x-y \) only.
REFERENCES


