June 1967

NONPARAMETRIC TESTS OF LOCATION FOR CIRCULAR DISTRIBUTIONS[1]

Siegfried Schach

Technical Report No. 95

University of Minnesota

Minneapolis, Minnesota

[1]Research supported by the National Science Foundation under Grant No. GP-3816 and GP-6859, and U. S. Navy Grant NONR(G)-00003-66.

ABSTRACT

In this thesis classes of nonparametric tests for circular distributions are investigated. In the one-sample problem the null hypothesis states that a distribution is symmetric around the horizontal axis. An interesting class of alternatives consists of shifts of such distributions by a certain angle $\varphi \neq 0$. In the two-sample case the null hypothesis claims that two samples have originated from the same underlying distribution. The alternatives considered consist of pairs of underlying distributions such that one of them is obtained from the other by a shift of the probability mass by an angle $\varphi \neq 0$.

The general outline of the reasoning is about the same for the two cases: We first define (Sections 2 and 8) a suitable group of transformations of the sample space, which is large enough to make any invariant test nonparametric, in the sense that any invariant test statistic has the same distribution for all elements of the null hypothesis. We then derive, for parametric classes of distributions, the efficacy of the best parametric test, in order to have a standard of comparison for nonparametric tests (Sections 3 and 9). Next we find a locally most powerful invariant test against shift alternatives (Sections 4 and 10). In both cases this suggests a general class of nonparametric test statistics.

In the next step we find the large-sample distribution of the test statistics under the null hypothesis as well as under alternatives (Sections 5, 12, and 13). In the one-sample situation the differences between the linear and the circular case are only minor, and hence the arguments follow very closely the classical line of reasoning. The two-sample case poses a new type of problem, since the locally most powerful test

statistic has a quadratic character. It has been found useful to use the theory of measures on Hilbert spaces for determining the asymptotic distribution of the test statistic.

Finally, the asymptotic relative efficiency of these test sequences (as compared to best parametric tests) is derived (Sections 6 and 14). Bahadur's concept of efficiency seems to be appropriate and is used exclusively. The results obtained can be expressed in compact form in terms of inner products of vectors in certain Hilbert spaces. These expressions immediately yield a solution to the variational problem of maximizing the efficiency. Under certain regularity conditions an asymptotically efficient nonparametric test exists in both cases. We then illustrate the general theory by a few examples of test statistics, in particular by a two-sample test proposed by Wheeler and Watson (Section 15).

TABLE OF CONTENTS

I. PRELIMINARIES

1. Circular Distributions vs. Distributions on the Real Line.

For convenience we define the circle as the set of complex numbers $C = \{z: |z| = 1\}$. The relation $x \to s(x) = e^{ix}$, $0 \le x < 2\pi$, defines a one-to-one correspondence between C and $J = [0,2\pi)$. J is a measurable space if the measurable sets are defined to be the Borel sets on this interval. By means of the correspondence $s(\cdot)$ we define the measurable sets of C to be the images (by $s(\cdot)$) of the Borel sets on J. Any measure on one of the two measurable spaces induces a measure on the other. There is then a measure-theoretic isomorphism between the two spaces and hence for measure-theoretic consideration we are free to choose the copy which is most convenient for the problem in question. A circular distribution is then simply a distribution on C or, equivalently, on J.

With respect to other structural relations the situation is not so pleasant. C has a group structure (as a subgroup of the multiplicative group of the complex number system), but the group structure on J induced by $s^{-1}(\cdot)$ does not coincide with any of the basic algebraic structures on J (as a subset of the real line R).

C and J are topological spaces (with the topologies induced by the topological spaces of which they are subsets). With respect to these topologies $s(\cdot)$ is a continuous mapping, but it is not a homeomorphism, i.e., $s^{-1}(\cdot)$ is not continuous. In fact a homeomorphism (with respect to these topologies) cannot exist, since C is compact whereas J is not (using the fact that the continuous image of a compact set is compact).

One difference between C and J is of considerable importance in the investigation of nonparametric tests: J is an ordered set, C has no natural order relation. The one-to-one mapping $s(\cdot)$ induces an order relation on C, but this amounts to choosing a rather arbitrary cut-off point and a direction. This fact, for example, makes it impossible to define the rank of an observation in a satisfactory way, unless the distribution defines a "natural" cut-off point, as, e.g., the symmetric distributions considered in Chapter II.

Statistical analysis of circular distributions: Much work has been done in the analysis of circular distributions. There is, on one hand, a wide area of applications for these distributions (see, e.g., E. Batschelet (1965)); on the other hand probabilists have been attracted by the particular features of the circle group and they obtained a number of analogues of classical probability theorems for the circle group and for compact groups in general (see, e.g., U. Grenander (1963)). We are not concerned with either of these two areas of investigation--we restrict our interest to statistical problems which arise in connection with circular distributions.

Parametric classes of circular distributions: Among the parametric classes of circular distributions the von Mises distribution certainly is of extreme importance. It is sometimes called a circular normal distribution. Its density is

$$f_{k,\theta}(x) = \frac{1}{2\pi I_0(k)} \exp(k \cos(x-\theta)), \qquad 0 \le x < 2\pi; \quad 0 \le \theta < 2\pi, \quad k \ge 0. \tag{1.1}$$

θ is the mode of the distribution, and it is usually called the location parameter. k is a measure of concentration. In the particular case $k = 0$ we get the uniform distribution. $I_0(k)$ is a suitable normalizing constant.

In the Appendix (formula A.3) it is shown to be the Bessel function of purely imaginary argument of order zero. In the Appendix we also derive a few additional results about this distribution, which are useful in making comparisons between parametric and nonparametric tests for the location parameter.

Besides the class of von Mises distributions there is another type of unimodal circular distribution, but it is of minor importance: the sine-wave distribution. Its density function is simply

$$f_\theta(x) = \frac{1}{2\pi} + \frac{1}{2\pi} \cos (x-\theta), \qquad 0 \le x < 2\pi; \quad 0 \le \theta < 2\pi. \tag{1.2}$$

Finally it should be mentioned that any distribution on the real line determines a circular distribution if its probability is wrapped around a circle. Thus we get a "wrapped normal distribution", a "wrapped Cauchy distribution" and so on. For more details see Gumbel, Greenwood, and Durand (1953).

Moments of circular distributions: Various efforts have been made to define moments for circular distributions in a useful way. Some important information is contained in the notion of a mean vector. This is the vector from the center of the circle to the center of gravity of the distribution. Its direction certainly is a descriptive parameter of the distribution. In the particular case of a unimodal, symmetric (with respect to the mode) density the direction of the mode coincides with the direction of the mean vector (see, for example, the v. Mises distribution). The length of the mean vector, i.e., the distance between midpoint of the circle and the center of gravity of the distribution, is a descriptive measure of the concentration of the mass. It varies between 0 and 1. From the Appendix it follows easily that for the v. Mises distribution this measure of concentration is $\frac{I_1(k)}{I_0(k)}$, where $I_1(k)$ is the Bessel function of purely imaginary argument of order one. Further comments on these and other descriptive measures may be found in E. Batschelet (1965).

<u>Trigonometric moments</u>: A useful tool for analyzing circular distributions is the sequence of trigonometric moments defined by

$$(1.3) \qquad \mu_m = \int_0^{2\pi} e^{imx}\, dF(x), \qquad m = 0, \pm 1, \pm 2, \ldots$$

where F is the c.d.f. of the probability distribution (on J). Its real counterparts are the sequences $\mu_m^{(c)}$ and $\mu_m^{(s)}$, $m = 0, 1, 2, \ldots$, where

$$(1.4) \qquad \mu_m^{(c)} = \int_0^{2\pi} \cos(mx)\, dF(x), \qquad \mu_m^{(s)} = \int_0^{2\pi} \sin(mx)\, dF(x).$$

It can be seen that the mean vector defined above is the vector $(\mu_1^{(c)}, \mu_1^{(s)})$. The sequence $\{\mu_m, \; m = 0, \pm 1, \pm 2, \ldots\}$ is the sequence of Fourier coefficients of F, and it is a well-known fact that it in turn determines F, so that the trigonometric moments of a circular distribution characterize this distribution.

II. ONE-SAMPLE CASE

2. Invariance Considerations in the One-Sample Case.

Definition of a location parameter: Let $f(\cdot)$ be a positive, periodic function on the real line R, with period 2π and with the property

$$\int_0^{2\pi} f(x)\,dx = 1.$$

Let $I(x)$ be the indicator function of $[0,2\pi)$, i.e.,

$$I(x) = \begin{cases} 1, & \text{if } 0 \le x < 2\pi, \\ 0, & \text{otherwise.} \end{cases}$$

The family of densities

$$(2.1) \qquad \Theta' = \{f(x-\theta)\,I(x):\ 0 \le \theta < 2\pi\}$$

on $[0,2\pi)$ is a one-parameter family of densities and θ is called location parameter.

In this definition θ is defined quite arbitrarily. In many instances it seems natural to restrict this arbitrariness somewhat. Thus if $f(\cdot)$ is symmetric with respect to some point φ $(0 \le \varphi < 2\pi)$, then we redefine $f(\cdot)$ by shifting it by the amount φ, so that it becomes symmetric with respect to 0, and hence $f(x-\theta)$ is axially symmetric (as a density on C) with respect to the axis $\overrightarrow{Oe^{i\theta}}$ on the complex plane. If $f(\cdot)$ is symmetric with respect to 0 and periodic with period 2π, then it is also symmetric with respect to π, since

$$(2.2) \qquad f(\pi+x) = f(\pi+x-2\pi) = f(x-\pi) = f(\pi-x), \qquad \text{any } x.$$

This shows that even in the symmetric case θ is not defined uniquely. However, in most cases of interest we have $f(0) \neq f(\pi)$, and in these cases we will, w.l.o.g., assume that $f(0) > f(\pi)$. Thus, for example, in the case of the v. Mises distribution defined by (1.1) θ satisfies all our assumptions about a location parameter, provided $k > 0$.

Unless $k = 0$ these conventions define θ uniquely for the v. Mises distribution. This will be the case in all situations, where the density is unimodal and symmetric. There are, however, cases, e.g., the class of densities $\{\overline{C}(k) \exp(\cos(2x-\theta)): 0 \le \theta < 2\pi\}$, where a given density does not identify θ uniquely. In these cases statistical procedures can never detect the "correct" θ. The extremely degenerate case of a uniform distribution is another case in point. In order to exclude these difficulties we assume, for the one-sample case, (i.e., throughout this chapter) that $f(\cdot)$ is symmetric with respect to 0, that $f(0) \ge f(x)$ for $x \in (0,2\pi)$, and that $f(\pi) < f(0)$. Under these assumptions the location of $f(\cdot)$ is defined uniquely (up to multiples of 2π), and hence the location parameter θ is a uniquely defined parameter for the class of densities (2.1).

<u>Hypothesis of symmetry</u>: Let $f(\cdot)$ have the properties described above. If we are to test $H: \theta = 0$ vs. $K: \theta \ne 0$, then this is equivalent to testing H: the distribution is symmetric with respect to 0 vs. $K = H^c$. This equivalence of symmetry and $\theta = 0$ will allow us to apply invariance considerations to the problem of testing a simple hypothesis about the location parameter θ.

<u>Transformation group</u>: Let F (on $[0,2\pi]$) be the c.d.f. of a circular distribution, F continuous, strictly increasing. We now consider the more general problem of testing

$$(2.3)\qquad \begin{cases} H: & F(x) + F(2\pi-x) = 1 \quad \text{for all } 0 \le x \le 2\pi, \\ & \text{vs.} \\ K: & F(x) + F(2\pi-x) \not\equiv 1 \end{cases}$$

In other words, H is the class of continuous circular distributions symmetric with respect to the horizontal axis and K is the class of all continuous alternatives. In defining a suitable group of transformations we make use of the one-to-one correspondence $s(\cdot)$ between $[0,2\pi)$ and C defined in section 1 $(s(x) = e^{ix})$. By this isomorphism any group $T = \{t\}$ of transformations on C has a corresponding group $T' = \{t'\}$ of transformations on $[0,2\pi)$ and vice versa.

We define T' and generate T by means of this isomorphism. Let $T' = \{t'(\cdot)\}$, where $t'(\cdot)$ is any continuous, strictly increasing mapping of $[0,2\pi]$ onto $[0,2\pi]$ with the property

$$t'(2\pi-x) + t'(x) = 2\pi, \qquad \text{all } x \in [0,2\pi]. \tag{2.4}$$

In particular we have $t'(0) = 0$, $t'(\pi) = \pi$, $t'(2\pi) = 2\pi$, $t'((0,\pi)) \to (0,\pi)$ and $t'((\pi,2\pi)) \to (\pi,2\pi)$. I.e., the induced transformation t maps the upper half-circle onto the upper half-circle and the lower half-circle onto the lower half-circle. By (2.4) the "deformation" obtained by t is the same for upper and lower half-circle.

<u>Lemma 2.1</u>: The class T' of strictly increasing, continuous transformations of $[0,2\pi]$ onto $[0,2\pi]$ satisfying (2.4) is a group.

<u>Proof</u>: The existence of a unit element and of a strictly increasing continuous inverse function $(t')^{-1}$ is obvious. We only have to show that it satisfies (2.4). Assume the contrary. Let x be such that, say,

$$(t')^{-1}(2\pi-x) > 2\pi - (t')^{-1}(x).$$

Then, applying the strictly increasing t' on both sides, we obtain

$$2\pi-x > t'(2\pi-(t')^{-1}(x)) = 2\pi-t'\circ(t')^{-1}(x) = 2\pi-x,$$

a contradiction.

<u>Definition</u>: The transformation group T (on C) is the group of transformations corresponding to T' by means of the isomorphism $s(\cdot)$; i.e., $t(x) = s\circ t'\circ s^{-1}(x)$ for all $x \in C$.

Induced group: If X is a circular random variable[1] with a continuous c.d.f. $F(x)$, $0 \le x < 2\pi$, $t \in T$, then $t(X)$ is a circular random variable with c.d.f. $F \circ (t')^{-1}$, where $t \hat{=} t'$ under the isomorphism $s(\cdot)$. Define

$$(2.5) \qquad \Theta = \{F: F \text{ is a continuous, strictly increasing c.d.f. on } [0,2\pi)\} = H \cup K.$$

Then for every pair of elements $t \in T$, $F \in \Theta$ the c.d.f. $F \circ (t')^{-1}$ is in Θ. It is easy to see that the class of mappings

$$(2.6) \qquad \bar{t}: \ F \to F \circ (t')^{-1}$$

is itself a transformation group $\bar{T}$, acting on members of Θ, and that the relation $t \to \bar{t}$ is a homomorphism. $\bar{T}$ is called induced group.

Invariance of the hypothesis: If a testing problem exhibits certain symmetry relations, one would like to restrict the class of test statistics to those which match this type of symmetry, since otherwise a test would depend on a specific labeling of the sample points, which might be quite arbitrary. Such a restriction is usually possible, provided that the problem itself is invariant, i.e., provided that $\bar{T}$ maps H into H and K into K. We show that this requirement is satisfied in our case.

Lemma 2.2: For every $\bar{t} \in \bar{T}$ we have $\bar{t}H = H$ and $\bar{t}K = K$.

Proof: Since $\bar{t}$ has an inverse on Θ, we must have $\bar{t}\Theta = \Theta$. Let $t'(x) = y$, then $(t')^{-1}(2\pi - y) = 2\pi - (t')^{-1}(y) = 2\pi - x$, by (2.4). Since $t': [0,2\pi] \xrightarrow{\text{onto}} [0,2\pi]$ we get the relations

$$F \in H \Leftrightarrow F(x) + F(2\pi - x) \equiv 1 \Leftrightarrow F \circ (t')^{-1}(y) + F \circ (2\pi - (t')^{-1}(y)) \equiv 1$$
$$\Leftrightarrow F \circ (t')^{-1}(y) + F \circ (t')^{-1}(2\pi - y) \equiv 1 \Leftrightarrow F \circ (t')^{-1} \in H \Leftrightarrow \bar{t}F \in H.$$

Hence $\bar{t}H = H$, and consequently $\bar{t}K = \bar{t}(\Theta - H) = \bar{t}\Theta - \bar{t}H = \Theta - H = K$,

q.e.d.

---

[1] By definition a circular random variable is a random variable taking its values on C, or, equivalently, on J.

<u>Maximal invariant</u>: Let a sample outcome $x_1, x_2, \ldots, x_n$ be given on C. Construct a set of points $y_1, y_2, \ldots, y_n$, where

$$y_i = \begin{cases} x_i, & \text{if } x_i \text{ is on the upper half-circle,} \\ \overline{x_i}, & \text{(conjugate complex) if } x_i \text{ is on the lower half-circle.} \end{cases}$$

Let $r_1, r_2, \ldots, r_m$ be the ranks of the x_i on the upper half-circle among the y_i[1], where we use the positive direction (counter-clockwise) and 0 as the cut-off point to define the order. Then we get the following

<u>Lemma 2.3:</u> The ranks of the observations on the upper half-circle among the y_i $(i = 1, 2, \ldots, n)$ is a maximal invariant under T.

<u>Proof</u>: (a) Since any $t' \in T'$ is monotone increasing and skew-symmetric around (π, π), the corresponding t does not change the relative position of the observations, i.e., $r_1, \ldots, r_m$ is invariant under T.

(b) If two samples have the same rank sequence $r_1, r_2, \ldots, r_m$ then the corresponding y_i's may differ only with respect to the spacing of the observations. It is then obvious that there exists a (continuous, strictly increasing) polygon $t'(x)$ on $[0, \pi]$, with $t'(0) = 0$, $t'(\pi) = \pi$, which maps the first y-sequence into the second one. Extending $t'(\cdot)$ to the interval $[0, 2\pi]$ by skew-symmetry around (π, π) yields an element of T'; and since m and $r_1, \ldots, r_m$ are the same for the two samples, the corresponding element $t(\cdot) \in T$ maps the first sample into the second. Thus two samples with the same rank sequence $r_1, \ldots, r_m$ lie on the same orbit, which proves maximality of the invariant.

<u>Distribution of an invariant statistic</u>: E. L. Lehmann (1959, page 220, Thm. 3) shows that the distribution of an invariant statistic depends on the underlying distribution F only through the orbit (defined by the induced group $\overline{T}$) of which F is a member. We now show that all the elements of H belong to the same orbit.

---

[1] Sufficiency of the "order statistic" allows us to neglect the order in which the x_i's were obtained.

<u>Theorem 2.1</u>: Any test of H vs. K depending only on the ranks of the observations on the upper half-circle (among the "combined" sample, consisting of the y_i's) is nonparametric, i.e., it has the same distribution for all elements of H.

<u>Proof</u>: According to Lehmann's theorem referred to above it suffices to show that if $F \in H$ and $G \in H$, there exists a $\bar{t} \in \bar{T}$ such that $\bar{t}F = G$, or, equivalently, $F \circ (t')^{-1} = G$ for some element $t' \in T'$. Now set $t' = G^{-1} \circ F$. t' is a continuous, strictly increasing mapping and $t'([0,2\pi]) = [0,2\pi]$. Since $G(2\pi - y) = 1 - G(y)$ for any $y \in [0,2\pi]$, we have, upon applying G^{-1}, $2\pi - y = G^{-1}(1-G(y))$. For $y = G^{-1} \circ F(x)$ we obtain

$$G^{-1} \circ F(2\pi - x) = G^{-1}(1-F(x)) = G^{-1}(1-G(y)) = 2\pi - G^{-1} \circ F(x),$$

i.e., t' satisfies (2.4). Since $(t')^{-1} = F^{-1} \circ G$, we have $\bar{t}F = F \circ (t')^{-1}$ $= F \circ F^{-1} \circ G = G$, and this completes our proof.

3. Asymptotic Relative Efficiency of Tests for Location Parameters.

Concept of relative efficiency: Nonparametric tests of a hypothesis have the advantage that the probability α for a Type I error can be controlled even if only a minimum of information about the underlying distribution is available. But in general one has to pay for this advantage by a larger probability β for a Type II error, i.e., the power function for alternatives is lower. In many cases the powers of a parametric test and of a suitable nonparametric test can be made equal to a certain β for a specific $\theta \in K$ by taking a larger number of observations for the nonparametric test. The ratio of the two sample sizes $\frac{n'}{n}$, where n' corresponds to the best parametric test, is then called the (exact) efficiency of the nonparametric test. In general, if $\{T_n\}$, $\{T'_{n'}\}$ are two sequences of test statistics, this efficiency $e(T_n, T'_{n'})$ depends on α, β, θ, and, of course, on the underlying distribution. Various asymptotic efficiency measures (ARE) have been proposed, which are independent of α, β, and θ. The efficiency concept used most frequently has been suggested by Pitman (1948). However, the Pitman-efficiency is defined only under quite restrictive conditions. In particular T_n has to be asymptotically normally distributed under H as well as under K. A more general and for our purposes more suitable measure of efficiency, which is applicable for many non-normal cases, has been proposed by Bahadur (1960).

Bahadur-efficiency: Bahadur's concept of efficiency has been extended to an even more general class of test statistics by Gleser (1964). We state here Gleser's assumptions and show how the efficiency is computed.

Let T_n, T'_n be two test statistics for testing the hypothesis $\theta \in H$.

Assumption 1: There exist continuous cumulative distribution functions $F^{(i)}(x)$, $i = 1, 2$ such that for each $\theta \in H$

$$(3.1) \qquad \lim_{n \to \infty} P_\theta\{T_n^{(i)} \le x\} = F^{(i)}(x), \quad \text{all } x.$$

Assumption 2: There exist $t_i > 0$ and $a_i > 0$ such that

$$(3.2) \qquad -2 \log (1 - F^{(i)}(x)) = a_i x^{t_i}(1 + o(1)), \qquad x \to \infty, \quad i = 1, 2.$$

Assumption 3: There exist continuous, strictly increasing functions $b^{(i)}(x)$ mapping $(0,\infty)$ onto $(0,\infty)$, and functions $c_i(\theta)$, $0 < c_i(\theta) < \infty$, defined on K, such that

$$(3.3) \qquad \text{p lim } T_n^{(i)}/b^{(i)}(n) = c_i(\theta), \qquad \text{all } \theta \in K; \quad i = 1,2.$$

Assumption 4:

$$(3.4) \qquad (b^{(1)}(x))^{t_1} = (b^{(2)}(x))^{t_2}.$$

Bahadur has shown that under these conditions tests with critical regions $T_n^{(i)} \geq C_{n,\alpha}^{(i)}$ are optimal in a certain sense and that an (approximate) measure of asymptotic relative efficiency at $\theta \in K$ is given by

$$(3.5) \qquad e(T_n^{(1)}, T_n^{(2)}/\theta) = a_1 c_1^{t_1}(\theta)/(a_2 c_2^{t_2}(\theta)).$$

If we test a simple hypothesis about θ, e.g., $H: \theta = \theta_0$ vs. $K: \theta \neq \theta_0$, then we are primarily interested in evaluating

$$(3.6) \qquad e(T_n^{(1)}, T_n^{(2)}) = \lim_{\theta \to \theta_0} e(T_n^{(1)}, T_n^{(2)}/\theta),$$

since under these conditions most tests are consistent for all alternatives of interest and hence the discriminatory power for alternatives "close to the hypothesis" is the property that is of basic importance for the comparison.

$$(3.6a) \qquad s_i(\theta) = a_i c_i^{t_i}(\theta)$$

is sometimes called asymptotic slope of the test based on $\{T_n^{(i)}\}$, since for a suitably normalized sequence $K_n^{(i)}$ it is the rate at which the probability escapes to infinity, if $\theta \neq \theta_0$ is the true parameter. For more details see R. R. Bahadur (1960).

If the test statistic T_n is asymptotically normally distributed, there is another measure of the effectiveness of this sequence of tests, which is closely related to the asymptotic slope of the test sequence. This measure is called efficacy.

Definition: For testing H: $\theta = \theta_0$ vs. K: $\theta > \theta_0$ let T_n be a test statistic which is asymptotically normally distributed for all θ. Let $E_\theta T_n \to \mu_\theta$ and $\sqrt{n}\,\sigma_\theta(T_n) \to \sigma_\theta$ as $n \to \infty$. Assume that $\sigma_\theta \to \sigma_{\theta_0}$ as $\theta \to \theta_0$ and that $\left.\frac{\partial\mu_\theta}{\partial\theta}\right|_{\theta=\theta_0}$ exists and is different from 0. Then the quantity

$$\text{(3.7)} \qquad \text{eff}\ (T_n) = \left(\left.\frac{\partial\mu_\theta}{\partial\theta}\right|_{\theta=\theta_0}\right)^2 / \sigma_{\theta_0}{}^2$$

is called the efficacy of the sequence $\{T_n\}$ of test statistics. Under quite general conditions the Pitman ARE is equal to the ratio of the efficacies of the two test sequences.

If the assumptions for applying Pitman's concept of the ARE of two test sequences are satisfied, then Bahadur's assumptions are also satisfied and the two values coincide. It follows easily from the definitions that in this case we have

$$\text{eff}\ (T_n) = \frac{1}{2}\frac{d^2}{d\theta^2}\, s(\theta)\bigg|_{\theta=0},$$

where $s(\theta)$ is the asymptotic slope. Using this equation as a definition, we will extend the notion of efficacy of a test sequence $\{T_n\}$ to all the cases where Assumptions 1-4 are satisfied.

Some remarks on the consistency of a test sequence: A sequence $\{T_n\}$ of test statistics for testing H vs. K is called consistent for a particular $\theta_1 \in K$ if

$$\lim_{n \to \infty} P_{\theta_1} (T_n \in \text{Critical Region}) = 1$$

whatever the size of the test might be.

Lemma 3.1: A sequence of tests of the form $T_n \geq C_n$, where T_n satisfies (3.1) for $\theta \in H$ and (3.3) for $\theta \in K$ is consistent for any θ such that $c(\theta) > 0$, provided that $b(n) \to \infty$.

Proof: Because of (3.1) the critical points C_n converge to some value C. Hence $C_n/b(n) \to 0$. (3.3) is equivalent to $T_n/b(n) - c(\theta) \to 0$ in probability. Thus we get

$$P_\theta\{T_n > C_n\} = P_\theta\{\frac{T_n}{b(n)} > \frac{C_n}{b(n)}\} = P\{\frac{T_n}{b(n)} - c(\theta) > \frac{C_n}{b(n)} - c(\theta)\} \to 1$$

as $n \to \infty$.

Asymptotic slope of "best" parametric tests: In order to evaluate the merits of a certain nonparametric test of a hypothesis, one has to have a standard with which it is to be compared. Such a standard can usually be obtained if it is known that the underlying distribution is a member of a certain parametric class. If efficiency is the criterion used to compare different tests, then one would derive the asymptotic slope of the (parametric) test with highest slope and this value would then be used as standard of comparison. We shall derive this standard for a family of distributions with a location parameter θ. Since we are not primarily interested in parametric tests, we will be somewhat informal in deriving our results. In particular we will assume that the densities are sufficiently regular, so that the required derivatives exist, that Fishers "information" is finite and positive, and that the interchanges of limit processes are justified. For the v. Mises distribution, which is of basic interest to us, it is easily seen that these assumptions are satisfied.

Let $f(\cdot)$ be a positive, periodic function on R, with period 2π, such that

$$\int_0^{2\pi} f(x)\,dx = 1.$$

Define $I(x)$ to be the indicator function of $[0,2\pi)$. Then

$$\{f(x-\theta)\,I(x):\ 0 \le \theta < 2\pi\} \tag{3.8}$$

is a one-parameter family of densities of circular distributions. θ is a location parameter.

Let X be a circular random variable. If we are to test

H: X has density $f(x)\,I(x)$

vs.

K: X has density $f(x-\theta)\,I(x)$, some specific $\theta > 0$,

then according to the Neyman-Pearson lemma the most powerful test of H vs. K rejects H if

$$T_{n,\theta}' = \sum_{i=1}^{n} \log \frac{f(x_i-\theta)}{f(x_i)} > C_\theta \tag{3.9}$$

for some C_θ depending on α.

An equivalent test would be

$$T_{n,\theta} = \frac{1}{\theta} \sum_{i=1}^{n} \log \frac{f(x_i-\theta)}{f(x_i)} > C_\theta. \tag{3.10}$$

(C is to be understood as a generic constant.) By the Central Limit Theorem $T_{n,\theta}$ is asymptotically normally distributed.

Upper bound for the efficacy: $T_{n,\theta}$ is not a test statistic for testing H: $\theta = 0$ vs. K: $\theta > 0$, since it depends on a specific θ, but it is easy to see that the "efficacy" based on tests $T_{n,\theta} > C_\theta$ is an upper bound for the efficacy of any proper test statistic T_n, since each of these tests is tailor-made for its specific θ. We now compute this upper bound.

$$\frac{E_\theta T_{n,\theta} - E_0 T_{n,\theta}}{\theta} = \frac{n}{\theta^2}\left[\int_0^{2\pi} (\log \frac{f(x-\theta)}{f(x)}) f(x-\theta)dx - \int_0^{2\pi} (\log \frac{f(x-\theta)}{f(x)}) f(x)dx \right]$$

$$= n \int_0^{2\pi} \frac{\log f(x-\theta) - \log f(x)}{-\theta} \frac{f(x-\theta) - f(x)}{-\theta} dx$$

$$\underset{\theta\to 0}{\longrightarrow} n \int_0^{2\pi} \frac{\partial \log f(x+\theta)}{\partial\theta}\bigg|_{\theta=0} \frac{\partial f(x+\theta)}{\partial\theta}\bigg|_{\theta=0} dx$$

$$= n \int_0^{2\pi} \frac{f'(x)}{f(x)} f'(x)dx = n \text{ Inf } (f),$$

where $\text{Inf } (f) = \int_0^{2\pi} (\frac{f'(x)}{f(x)})^2 f(x)dx = E_0(\frac{f'(x)}{f(x)})^2.$

In the same way it can be shown that $\text{Var}_\theta\ T_{n,\theta} = \frac{n}{\theta^2} \text{Var}_\theta(\log \frac{f(x-\theta)}{f(x)}) \underset{\theta\to 0}{\longrightarrow} n \text{ Inf } (f).$

Hence we obtain, as an upper bound for the efficacy of any sequence of test statistics,

$$(3.11) \qquad \text{eff}_u(T_n) = \left(\frac{\partial E_\theta T_{n,\theta}}{\partial\theta}\bigg|_{\theta=0}\right)^2 \Big/ n \text{ Var}_{\theta=0}(T_{n,\theta}) = \text{Inf } (f).$$

<u>Example of an efficient parametric test</u>: We will now show that there actually exists a test statistic T_n whose efficacy achieves this upper bound. The heuristic approach we take consists in letting θ go to 0 in (3.10), hoping that the resulting test statistic, T_n say, is asymptotically as good as $T_{n,\theta}$. Since

$$\lim_{\theta\to 0} \frac{1}{\theta} (\log f(x-\theta) - \log f(x)) = - \frac{\partial \log f(x+\theta)}{\partial\theta}\bigg|_{\theta=0} = - \frac{f'(x)}{f(x)},$$

we get

$$(3.12) \qquad T_n = - \sum_{i=1}^{n} \frac{f'(x_i)}{f(x_i)} .$$

T_n is asymptotically normally distributed, hence we can compute its efficacy. We get

$$\frac{E_\theta T_n - E_0 T_n}{\theta} = -n \int_0^{2\pi} \frac{f'(x)}{f(x)} \frac{f(x-\theta) - f(x)}{\theta}\, dx \xrightarrow[\theta\to 0]{} n \text{ Inf } (f).$$

$$\text{Var}_0\, T_n = n \text{ Var}_0 \left(\frac{f'(x)}{f(x)}\right) = n\, E_0 \left(\frac{f'(x)}{f(x)}\right)^2 = n \text{ Inf } (f).$$

Thus for the particular test statistic T_n we get

(3.13) $$\text{eff } (T_n) = \text{Inf } (f).$$

Since this value coincides with the upper bound (3.11), it is the maximum achievable and will whence be used as the standard for comparison.

Likelihood ratio test for the v. Mises distribution: In this subsection we derive the likelihood ratio test for the v. Mises distribution and show that this is another efficient test.

First we note that in the case of the v. Mises distribution the test statistic T_n defined in (3.12) has the form

(3.14) $$T_n = k \sum_{i=1}^{n} \sin x_i > C_{k,\alpha}$$

for some suitably chosen constant $C_{k,\alpha}$.[1] This makes sense intuitively, since under H $\sin X$ has a symmetric distribution with $E \sin X = 0$ whereas under $K: \theta > 0$ we get $E_\theta \sin X > 0$ for small $\theta > 0$ (see Appendix (A.6)) and this shows, incidentally, that T_n is consistent against these local alternatives.

The likelihood ratio test has critical region

(3.16) $$\frac{\exp\left(k \sum_{i=1}^{n} \cos x_i\right)}{\exp\left(k \sum_{i=1}^{n} \cos (x_i - \hat\theta)\right)} < c$$

where $\hat\theta$ is the MLE.

[1] Since we are interested in location parameters only we will always assume that k is known.

Equivalently

$$T_n' = \sum_{i=1}^{n} \cos(x_i - \hat{\theta}) - \sum_{i=1}^{n} \cos x_i > C.$$

It is easy to see that the MLE $\hat{\theta}$ satisfies $\tan \hat{\theta} = \dfrac{\Sigma \sin x_i}{\Sigma \cos x_i}$

(i.e., the location θ is estimated by the direction of the resultant vector $(\sum_{i=1}^{n} \cos x_i, \sum_{i=1}^{n} \sin x_i)$, or "sample mean vector" $(\frac{1}{n} \sum_{i=1}^{n} \cos x_i, \frac{1}{n} \sum_{i=1}^{n} \sin x_i))$. Hence we get

(3.17)
$$\begin{aligned} T_n' &= \Sigma \cos x_i \cos \hat{\theta} + \Sigma \sin x_i \sin \hat{\theta} - \Sigma \cos x_i \\ &= \cos \hat{\theta} (\Sigma \cos x_i + \tan \hat{\theta}\, \Sigma \sin x_i) - \Sigma \cos x_i \\ &= \frac{1}{\sqrt{1 + \left(\dfrac{\Sigma \sin x_i}{\Sigma \cos x_i}\right)^2}} \left(\Sigma \cos x_i + \frac{(\Sigma \sin x_i)^2}{\Sigma \cos x_i}\right) - \Sigma \cos x_i \\ &= \sqrt{(\Sigma \cos x_i)^2 + (\Sigma \sin x_i)^2} - \Sigma \cos x_i \\ &= R - L, \end{aligned}$$

where R is the length of the resultant of the sample points, and P is the length of the projection of this resultant onto the axis determined by θ_0 of the hypothesis (in our case $\theta_0 = 0$, horizontal axis).

<u>Asymptotic slope of the LR test for the v. Mises distribution</u>: It is well-known from likelihood ratio theory that under H $2kT_n'$ has asymptotically a χ^2-distribution with one d.f. Hence Assumption 1 above is satisfied with $F(x) = F_{\chi^2}(2kx)$, where $F_{\chi^2}(\cdot)$ is the c.d.f. of a χ^2-variable with one d.f. It has been shown be R. R. Bahadur (1960) that F_{χ^2} satisfies Assumption 2 above with $a = 1$, $t = 1$. Hence $F(x)$ satisfies (3.2) with $a = 2k$, $t = 1$. To check Assumption 3 we take $b(x) = x$. Then from (3.17)

(3.18) $$T_n'/n = \left(\left[\frac{\sum_{i=1}^n \cos x_i}{n}\right]^2 + \left[\frac{\sum_{i=1}^n \sin x_i}{n}\right]^2\right)^{1/2} - \frac{\sum_{i=1}^n \cos x_i}{n}$$

$$\rightarrow \sqrt{(E_\theta \cos X)^2 + (E_\theta \sin X)^2} - E_\theta \cos X \quad \text{a.s.}$$

by the strong law of large numbers. According to Appendix (A.6), (A.7) $E_\theta \cos X = \cos\theta\, E_0 \cos X$, $E_\theta \sin X = \sin\theta\, E_0 \cos X$. Hence

(3.19) $$T_n'/n \rightarrow (1 - \cos\theta)\, E_0 \cos X$$

$$= (1 - \cos\theta) \frac{I_1(k)}{I_0(k)} \qquad \text{(by Appendix (A.5))}$$

$$= (1 - \cos\theta) \frac{1}{k} \operatorname{Inf}(f) \qquad \text{(by Appendix (A.10))}$$

$$= c(\theta), \text{ say.}$$

For the asymptotic slope $s(\theta) = ac(\theta)^t$ we get

(3.20) $$s(\theta) = 2k\,(1 - \cos\theta) \frac{1}{k} \operatorname{Inf}(f) \doteq \theta^2 \operatorname{Inf}(f)$$

for θ close to 0.

From this result we get the efficacy

(3.21) $$\operatorname{eff}(T_n') = \frac{1}{2} \frac{d^2}{d\theta^2} s(\theta)\Big|_{\theta=0} = \operatorname{Inf}(f).$$

Hence the likelihood ratio test is an efficient test in the case of the v. Mises distribution.

4. Locally Most Powerful Invariant Test for Shift Alternatives.

In Section 2 we derived the result that tests are invariant under T if and only if they are "rank tests", i.e., if they depend only on the ranks of the observations originally on the upper half-circle, among the combination of points obtained by reflecting the lower half-circle on the horizontal axis. We will denote these ranks by $R_1, R_2, \ldots, R_m$ and a specific realization by $r_1, \ldots, r_m$. Here m is a random variable, $0 \le m \le n$, where n is the sample-size. We will now, under certain conditions, derive an invariant test for a location parameter θ which is a locally most powerful (LMP) invariant test. That is, if θ is the location parameter, and $\theta = 0$ is the hypothesis to be tested, then we will obtain a neighborhood $U = [0, \epsilon]$, $\epsilon > 0$ and a test statistic T_n such that no invariant test has higher power for any $\theta \in U$ than the test $T_n \le C_{n,\alpha}$.

If a family of densities with a location parameter θ is given, we denote by $E_\theta X$ the expectation of any variable X, if the location parameter has the value θ. For the derivation of a LMP invariant test it is convenient to use a result, essentially due to W. Hoeffding, in the form presented by Lehmann (1959), page 254, prob. 22:

Let $Z_1, Z_2, \ldots, Z_n$ be independently distributed with densities $f_1, \ldots, f_n$ and let the rank of Z_i be denoted by T_i. If f is any probability density which is positive whenever at least one of the f_i is positive, then

$$(4.1) \qquad P\{T_1 = t_1, \ldots, T_n = t_n\} = \frac{1}{n!} E\left[\frac{f_1(V^{(t_1)})}{f(V^{(t_1)})} \cdots \frac{f_n(V^{(t_n)})}{f(V^{(t_n)})}\right],$$

where $V^{(1)} < \ldots < V^{(n)}$ is an ordered sample from a distribution with density f.

We now use this result to derive a theorem which is an analogue to a well-known theorem about locally most powerful tests for shift alternatives of distributions on the real line. But since shifts on the real line do not exactly correspond to shifts on the circle, our result is not a special case of that theorem, and hence it needs a separate proof.

<u>Theorem 4.1</u>: Let Θ be a family of densities, as defined in (2.1); if $f(\cdot)$ is symmetric with respect to 0, with a continuous derivative $f'(\cdot)$, and if $V^{(i)}$ is the i^{th} order statistic of a distribution with density $2f(\cdot)$ on $[0,\pi)$, then the test

$$(4.2) \qquad T_n = \sum_{i=1}^{m} E\left[\frac{f'(V^{(r_i)})}{f(V^{(r_i)})}\right] < C$$

is a LMP invariant test for the hypothesis $\theta = 0$ vs. $\theta > 0$.

<u>Proof</u>: We derive a LMP invariant test for every fixed m. Since m can take on only a finite number of values, it follows that the intersection of the corresponding neighborhoods of 0 is non-empty, and on this intersection all the conditional powers are maximized, hence the unconditional power is maximized too.

Now let $m = m_0$, fixed. Under the hypothesis $\theta = 0$ every possible assignment of the m_0 positions within the n locations is equally likely, since $f(\cdot)$ is symmetric w.r. to 0. Hence the probability is $1/\binom{n}{m_0}$ for every combination. Thus we maximize the power itself by maximizing the derivative of the power function, i.e., by taking into the critical region those rank combinations for which

$$(4.3) \qquad \frac{\partial}{\partial\theta} P_\theta \{R_1 = r_1, \ldots, R_m = r_m\}_{\theta=0}$$

is largest. (The number of points is determined by the size α of the test).

We now evaluate (4.3). Let $\int_0^\pi f(x-\theta)\, dx = D_\theta^{-1}$, then it follows immediately from Lehmann's result stated above and from the periodicity of $f(\cdot)$ that

$$(4.4)\qquad P_\theta(R_1 = r_1,\dots,R_m = r_m \mid m = m_0) =$$

$$\frac{D_\theta^{m_0}(1-D_\theta)^{n-m_0}}{2^n\binom{n}{m_0}} E\left[\frac{f(V^{(r_1)}-\theta)}{f(V^{(r_1)})} \cdots \frac{f(V^{(r_{m_0})}-\theta)}{f(V^{(r_{m_0})})} \frac{f(V^{(s_1)}+\theta)}{f(V^{(s_1)})} \cdots \frac{f(V^{(s_{n-m_0})}+\theta)}{f(V^{(s_{n-m_0})})}\right],$$

where $\{s_1, s_2,\dots,s_{n-m_0}\} = \{1, 2,\dots,n\} - \{r_1, r_2,\dots,r_{m_0}\}$.

(The factor $n!$ has been replaced by $\binom{n}{m_0}$, since each rank combination is obtained from $m_0!(n-m_0)!$ different rank permutations.) Taking the derivative under the expectation sign, which is justified by the Lebesgue dominated convergence theorem, we get

$$(4.5)\qquad \left.\frac{\partial P_\theta(R_1 = r_1,\dots,R_m = r_m \mid m = m_0)}{\partial\theta}\right|_{\theta=0} =$$

$$D + \left(-\sum_{i=1}^{m_0} E\left[\frac{f'(V^{(r_i)})}{f(V^{(r_i)})}\right] + \sum_{i=1}^{n-m_0} E\left[\frac{f'(V^{(s_i)})}{f(V^{(s_i)})}\right]\right)\binom{n}{m_0}^{-1} =$$

$$D' - D'' \sum_{i=1}^{m_0} E\left[\frac{f'(V^{(r_i)})}{f(V^{(r_i)})}\right]$$

for some constants D, D', $D'' > 0$. In the last step we used the fact that $\sum_{i=1}^{n} E\left[\frac{f'(V^{(i)})}{f(V^{(i)})}\right]$ is a constant. (4.5) shows that tests of the form (4.2) are UMP invariant at their corresponding level α, and this completes our proof.

The test statistic T_n which yields locally most powerful tests is a "linear rank order statistic" in the sense that it can be written in the form

$$(4.6) \qquad T_n = \sum_{i=1}^{n} a_{ni} z_i ,$$

where a_{ni} are double-sequences of constants and

$$(4.7) \qquad z_i = \begin{cases} 1, & \text{if the } i^{th} \text{ value of the "combined" (on the upper half-circle) sample corresponds to a value which was not reflected (i.e., it was on the upper half-circle originally)} \\ 0, & \text{otherwise.} \end{cases}$$

We now derive a few asymptotic properties for a class of linear rank order statistics of this form.

5. Some Asymptotic Properties of a Class of Linear Rank Order Statistics

In this section we will show that a large class of linear rank order statistics is asymptotically normally distributed. We will then derive the ARE of these tests and, in particular, we will show that, under quite general conditions, there exists an efficient test based on a linear rank order statistic. This implies the existence of an efficient nonparametric test for testing the location parameter of a symmetric circular distribution.

Some results obtained by Z. Govindarajulu: H. Chernoff and I. R. Savage (1958) derived asymptotic normality for a large class of two-sample rank order statistics. Z. Govindarajulu (1960) used their results to obtain similar theorems for the symmetric one-sample case. These results are useful for our problem. Before we state them we introduce some notation: Let $H_n(x)$ $0 \le x < \pi$ be the empirical c.d.f. of the "combined" sample, i.e., the "sample" obtained by reflecting the observations on the lower half-circle on the horizontal axis. F_m is defined to be the empirical distribution function of the observations on the upper half-circle (m is a random variable). Let $J_n(\cdot)$ be an arbitrary function on $[0,1]$, then it is easy to see that

$$(5.1) \qquad T_n' = m \int_0^{\pi} J_n(H_n(x))\, dF_m(x)$$

is the same linear rank order statistic as T_n defined by (4.6), provided that

$$(5.2) \qquad a_{ni} = J_n\left(\frac{i}{n}\right) \qquad i = 1, 2, \ldots, n.$$

The representation (5.1) is somewhat more handy in deriving asymptotic results.

It is convenient in this section to use $F(x)$, $0 \le x \le \pi$ as the (theoretical) c.d.f. of the (conditional) distribution on the upper half-circle. Correspondingly we use $G(x)$ and $G_{m'}(x)$ as the theoretical and empirical c.d.f. of the observations on the lower half-circle, where $m' = n - m$.

If we define $\lambda_n = \frac{m}{n}$, a random variable, then we get

(5.3) $$H_n(x) = \lambda_n F_m(x) + (1 - \lambda_n) G_{m'}(x) \qquad 0 \le x \le \Pi.$$

We define correspondingly

(5.4) $$H(x) = \lambda_n F(x) + (1 - \lambda_n) G(x) = G(x) + \lambda_n \Delta(x),$$

where

(5.5) $$\Delta(x) = F(x) - G(x).$$

For the statement of the basic theorem we also need the quantities

(5.6) $$p_n = E(\lambda_n), \qquad q_n = 1 - p_n.$$

(5.7) $$H^*(x) = G(x) + p_n \Delta(x) = p_n F(x) + (1 - p_n) G(x).$$

(5.8) $$L_i = \int_0^{\Pi} [\Delta(x)]^i J^{(i)} (H^*(x))\, dF(x) \qquad i = 0, 1.$$

(5.9) $$U = 2 \int\int_{0<x<y<\Pi} G(x)\, [1 - G(y)]\, J' (H^*(x))\, J' (H^*(y))\, dF(x)\, dF(y).$$

(5.10) $$V = 2 \int\int_{0<x<y<\Pi} F(x)\, [1 - F(x)]\, J' (H^*(x))\, J' (H^*(y))\, dG(x)\, dG(y).$$

(5.11) $$\alpha_n = n p_n L_0.$$

(5.12) $$\beta_n^2 = n p_n q_n\, [p_n U + q_n V + (L_0 + p_n L_1)^2].$$

From Theorems 3.1 and 3.3 and Lemma 5A.1 of Govindarajulu (1960) we obtain immediately

<u>Theorem 5.1</u>: If

(i) $E(\lambda_n) = p_n \to p_0$ such that $0 < p_0 < 1$;

(ii) $F(\cdot)$ and $G(\cdot)$ are absolutely continuous;

(iii) $J(\cdot)$ is a function on $[0,1]$ with a continuous second derivative, $J(0) = 0$, $J(\cdot)$ is not constant;

(iv) $p_0 U + q_0 V \neq 0$;

then for

(5.13) $$T_n = m \int_0^{\pi} J(H_n(x))\, dF_m(x)$$

we get

(5.14) $$\lim_{n \to \infty} P\{ \frac{T_n - \alpha_n}{\beta_n} \le w \} = \Phi(w) \qquad -\infty < w < \infty,$$

where $\Phi(\cdot)$ is the c.d.f. of the standardized normal distribution.

Remark: Govindarajulu's results allow for a more general class of approximating functions $J_n(\cdot)$, but we do not need his results in this generality.

Asymptotic normality of T_n in the null case:

Corollary 1: Let a circular distribution have density $f(x)$, $0 \le x \le 2\pi$, where $f(\cdot) > 0$ is continuous, $f(x + 2\pi k) = f(x)$, $k = \pm 1, \pm 2, \ldots$ and $f(x) = f(-x)$. Let J satisfy (iii) of Theorem 5.1. Then T_n defined by (5.13) is asymptotically normally distributed with normalizing constants

(5.15) $$\alpha_n = \frac{n}{2} \int_0^1 J(u)\, du$$

(5.16) $$\beta_n^2 = \frac{n}{4} \int_0^1 J^2(u)\, du.$$

Proof: We show that the conditions of Theorem 5.1 are satisfied:

(i) $E(\lambda_n) = \frac{1}{2}$ for all n.

(ii) F and G have densities $2f(x)$, $0 \le x \le \pi$.

(iii) is satisfied by assumption.

(iv) First we note that $H^* = F = G$. Hence $U = V$.

(5.17) $$U = 2 \int\int_{0<x<y<\pi} F(x)\,(1 - F(y))\, J'(F(x))\, J'(F(y))\, dF(x)\, dF(y)$$

$$= 2 \int\int_{0<u<v<1} u(1 - v)\, J'(u)\, J'(v)\, du\, dv$$

$$= \int_0^1 J^2(u)\, du - [\int_0^1 J(u)\, du]^2,$$

where the last step follows from an argument used by Chernoff and

Savage (1958), page 978. Hence $U > 0$, since $J(\cdot)$ is not a constant, and (iv) is satisfied. We now evaluate the normalizing constants:

$$(5.18) \qquad \alpha_n = \frac{n}{2}\int_0^{\pi} J(F(x))\, dF(x) = \frac{n}{2}\int_0^1 J(u)\, du$$

$$(5.19) \qquad \beta_n^2 = n \cdot \frac{1}{2} \cdot \frac{1}{2}\left[\frac{1}{2}U + \frac{1}{2}V + (L_0 + \frac{1}{2}L_1)^2\right].$$

Now $L_0 = \int_0^{\pi} J(F(x))\, dF(x) = \int_0^1 J(u)\, du$, $L_1 = 0$, since $\Delta(x) \equiv 0$.

$$(5.20) \qquad \beta_n^2 = \frac{n}{4}\left[\int_0^1 J^2(u)\, du - \left(\int_0^1 J(u)\, du\right)^2 + \left(\int_0^1 J(u)\, du\right)^2\right]$$

$$= \frac{n}{4}\int_0^1 J^2(u)\, du.$$

Asymptotic normality of T_n under alternatives $\theta \neq 0$: If we consider the class of densities $\{f(x-\theta)I(x),\ 0 \le x \le 2\pi\ ;\ 0 \le \theta < 2\pi\}$, then from Theorem 5.1 one would expect that T_n is asymptotically normally distributed for just about all the alternatives $\theta \neq 0$. However, for some $\theta \neq 0$ Assumption (iv) might be violated. Since it is difficult to single out these cases and since we are primarily interested in small deviations from the hypothesis (for efficiency-considerations) we will prove a "local" result only.

The quantities p_n, q_n, L_0, L_1, U, V now depend on θ. We indicate this dependence by a superscript (e.g., $p_n^{(\theta)}$) or a subscript (e.g., U_θ). It is easy to see that for a fixed θ p_n and q_n do not depend on n; hence we omit it.

Corollary 2: Let the assumptions on $f(\cdot)$ of Corollary 1 be satisfied. Then there exists a neighborhood U_ϵ of 0 such that T_n has asymptotically a normal distribution for all $\theta \in U_\epsilon$, where the normalizing constants are defined by (5.11) and (5.12).

Proof: Obviously Assumptions (i) - (iii) of Theorem 5.1 are satisfied. (iv) is satisfied for $\theta = 0$. It is easy to see that $p_0^{(\theta)}$, $q_0^{(\theta)}$, U_θ, V_θ are continuous functions of θ, hence $p_0^{(\theta)}U_\theta + q_0^{(\theta)}V_\theta > 0$ for some neighborhood U_ϵ of 0, and this completes our proof.

Bahadur efficiency of linear rank order tests: Before we derive the ARE of test statistics of the form (5.13) we first compute the approximate slope $ac^t(\theta)$ (see Section 3 for terminology) for a fixed alternative θ.

Lemma 5.1: Let the assumptions on f of Corollary 1 be satisfied, and let θ be such that $p_0^{(\theta)}U_\theta + q_0^{(\theta)}V_\theta \neq 0$. Then the function $c(\theta)$, defined in (3.3), is given by

$$(5.21) \qquad c(\theta) = 2\frac{p^{(\theta)}L_0^{(\theta)} - p^{(0)}L_0^{(0)}}{||J(\cdot)||},$$

where $||J(\cdot)||$ is the length of the function $J(\cdot)$ as an element of the Hilbert space of square integrable functions on $[0,1]$, i.e., $||J(\cdot)|| = (\int_0^1 J^2(u)du)^{\frac{1}{2}}$.

Proof: In this case the proper normalizing sequence $n^{-\delta}$, which makes $n^{-\delta}\frac{T_n - \alpha_n}{\beta_n}$ converge to a finite value $c(\theta)$, is obtained by setting $\delta = \frac{1}{2}$, because then we get

$$(5.22) \qquad \frac{T_n - \alpha_n}{n\,\beta_n} \simeq \frac{T_n - \alpha_n^{(\theta)}}{\frac{n}{2}\,||J||} + \frac{\alpha_n^{(\theta)} - \alpha_n}{\frac{n}{2}\,||J||}$$

$$\simeq \frac{1}{n}\frac{\beta_n^{(\theta)}}{\frac{1}{2}n\,||J||}\frac{T_n - \alpha_n^{(\theta)}}{\beta_n^{(\theta)}} + \frac{\alpha_n^{\theta} - \alpha_n}{\frac{n}{2}\,||J||}$$

$$\overset{pr.}{\rightarrow} 2\frac{p^{(\theta)}L_0^{(\theta)} - p^{(0)}L_0^{(0)}}{||J||},$$

since $\beta_n^{(\theta)} = O(n^{\frac{1}{2}})$ from (5.12), and $\frac{T_n - \alpha_n^{(\theta)}}{\beta_n^{(\theta)}} \rightarrow N(0,1)$, hence the first summand converges to 0 in probability; the second summand is independent of n, and the result follows on account of (5.11).

Lemma 5.2: Under the assumptions of Lemma 5.1 the asymptotic slope for a fixed alternative $\theta \neq 0$ is given by

$$(5.23) \qquad s(\theta) = 4\left(\frac{p^{(\theta)}L_0^{(\theta)} - p^{(0)}L_0^{(0)}}{||J||}\right)^2 .$$

Proof: The asymptotic slope is defined as $s(\theta) = ac^t(\theta)$, where the constants a and t are determined by (3.1) and (3.2). Here, however, T_n has to be replaced by $(T_n - \alpha_n^{(0)})/\beta_n^{(0)} = T_n'$, say. Then the c.d.f. F of (3.1) is equal to Φ, the c.d.f. of the standardized normal distribution. Bahadur (1960) has shown that for Φ Assumption (3.2) is satisfied with $a = 1$, $t = 2$. This result, combined with (5.21), yields (5.23).

In order to compute the efficiency of a test based on a statistic of the form (5.13), we have to analyze $s(\theta)$ in a neighborhood of 0. More specifically we need the efficacy of the sequence $\{T_n\}$ defined by

$$(5.24) \qquad \text{eff } (T_n) = \frac{1}{2} \frac{\partial^2}{\partial \theta^2} s(\theta)\Big|_{\theta=0} .$$

Theorem 5.2: If, in addition to the assumptions of Corollary 1, f has a continuous derivative, the efficacy of the sequence of test statistics $\{T_n\}$ is given by

$$(5.25) \qquad \text{eff } (T_n) = \frac{4[\int_0^{\pi} J(R(x))\, f'(x)\, dx]^2}{\int_0^1 J^2(x)\, dx} ,$$

where $R(x) = 2\int_0^x f(u)\, du$, $\quad 0 \leq x \leq \pi$.

Proof: Define $M_\theta(x) = \int_0^x f(y - \theta)\, dy = \int_{-\theta}^{x-\theta} f(u)\, du = M(x-\theta) - M(-\theta)$, where $M(\cdot) \equiv M_0(\cdot)$. Now $F(\cdot)$ and $G(\cdot)$, the conditional c.d.f. of the upper and lower half-circle, respectively, depend on θ; we indicate this by writing $F_\theta(\cdot)$, $G_\theta(\cdot)$. Then obviously

$$(5.26) \qquad F_\theta(x) = \frac{\int_0^x f(y-\theta)\,dy}{M_\theta(\pi)} = \frac{M(x-\theta) - M(-\theta)}{M_\theta(\pi)},$$

and

$$(5.27) \qquad G_\theta(x) = \frac{\int_x^0 f(y-\theta)\,dy}{1 - M_\theta(\pi)} = \frac{\int_{-x-\theta}^{-\theta} f(u)\,du}{1 - M_\theta(\pi)} = \frac{M(-\theta) - M(-x-\theta)}{1 - M_\theta(\pi)}.$$

Furthermore

$$(5.28) \qquad p^{(\theta)} = \int_0^\pi f(y-\theta)\,dy = M_\theta(\pi).$$

Hence

$$(5.29) \qquad p^{(\theta)}L_0^{(\theta)} = M_\theta(\pi) \int_0^\pi J(H_\theta^*(x))\,dF_\theta(x) \qquad \text{(by (5.8))}$$

$$= M_\theta(\pi) \int_0^\pi J\Big(M_\theta(\pi)\,\frac{M(x-\theta) - M(-\theta)}{M_\theta(\pi)} +$$

$$+ (1 - M_\theta(\pi))\,\frac{M(-\theta) - M(-x-\theta)}{1 - M_\theta(\pi)}\Big)\,\frac{f(x-\theta)}{M_\theta(\pi)}\,dx$$

$$= \int_0^\pi J(M(x-\theta) - M(-x-\theta))\,f(x-\theta)\,dx$$

$$= \int_{-\theta}^{\pi-\theta} J(M(u) - M(-u-2\theta))\,f(u)\,du.$$

By using the result just derived we obtain

$$(5.30) \qquad \frac{\partial}{\partial\theta}\, p^{(\theta)}L_0^{(\theta)}\Big|_{\theta=0} =$$

$$= -J(M(\pi) - M(-\pi))\,f(\pi) + J(M(0) - M(0))\,f(0) + 2\int_0^\pi J'(M(u) - M(-u))f(u)^2 du$$

$$= -J(1)\,f(\pi) + J(0)\,f(0) + 2\int_0^\pi J'(2M(u))\,f(u)^2\,du,$$

since $f(\cdot)$ is symmetric, and therefore $M(\cdot)$ is skew-symmetric.

If we express everything in terms of the conditional c.d.f. $R(\cdot) = 2M(\cdot)$ and conditional density $r(\cdot) = 2f(\cdot)$ for the upper (or lower) half-circle under $H: \theta = 0$, we get

$$(5.31) \qquad J(0)\, f(0) - J(1)\, f(\pi) = -J(R(x))\, f(x)\Big|_0^{\pi} =$$

$$-\left[\int_0^{\pi} J'(R(x))\, r(x)\, f(x)\, dx + \int_0^{\pi} J(R(x))\, f'(x)\, dx\right].$$

Hence we obtain from (5.30)

$$(5.32) \qquad \frac{\partial}{\partial\theta} p^{(\theta)}{}_{L_0}{}^{(\theta)}\Big|_{\theta=0} = -\frac{1}{2}\int_0^{\pi} J'(R(x)) r^2(x)\, dx - \int_0^{\pi} J(R(x))\, f'(x)\, dx$$

$$+ \frac{1}{2}\int_0^{\pi} J(R(x))\, r^2(x)\, dx = -\int_0^{\pi} J(R(x))\, f'(x)\, dx.$$

Since $s(0) = 0$, (5.23) and (5.24), combined with (5.32) yield the desired result (5.25).

<u>Lemma 5.3</u>: If the efficacy given by expression (5.25) is positive, then there exists a neighborhood U of 0 such that the tests based on $\{T_n\}$ are consistent for all $\theta \in U$.

<u>Proof</u>: Since the efficacy is equal to $\frac{1}{2}\frac{d^2}{d\theta^2} s(\theta)_{\theta=0}$, where $s(\theta)$ is the asymptotic slope, and since $s(0) = 0$, $\frac{d}{d\theta} s(\theta)_{\theta=0} = 0$ (this follows easily from (3.6a), here $a = 1$, $t = 2$) we must have $s(\theta) > 0$ in some neighborhood U of 0. In this neighborhood we then have $c(\theta) \neq 0$, since $c(\theta) = g\, s(\theta)$ for some constant g. The result follows from Lemma 3.1.

The space of square-integrable functions $h(\cdot)$ on $[0,1]$ is a Hilbert space If we denote the inner product of two functions h_1, h_2 of this space by (h_1, h_2), and the length of a function h by $||h||$, i.e.,

$$(h_1, h_2) = \int_0^1 h_1(x)\, h_2(x)\, dx, \qquad ||h|| = \left(\int_0^1 h(x)^2\, dx\right)^{\frac{1}{2}},$$

then we obtain directly the following

Theorem 5.3: If under H a circular distribution has a symmetric positive density f with a continuous derivative, then the Bahadur efficiency against shift alternatives of a test sequence based on

$$T_n = \sum_{i=1}^{n} J\left(\frac{i}{n}\right) z_i$$

is

$$(5.33) \qquad e(T_n, \text{ best test}/f) = \frac{\left(\int_0^1 J(v) \frac{r'}{r} \circ R^{-1}(v)dv\right)^2}{\int_0^1 J(v)^2 dv \int_0^1 \left(\frac{r'}{r} \circ R^{-1}(v)\right)^2 dv}$$

$$= \frac{\left(J(\cdot), \frac{r'}{r} \circ R^{-1}(\cdot)\right)^2}{||J(\cdot)||^2 \, ||\frac{r'}{r} \circ R^{-1}(\cdot)||^2} ,$$

where r,R are the density and c.d.f. of the distribution on the upper half-circle, provided that $J(\cdot)$ has two continuous derivatives, $J(0) = 0$ and $J(\cdot)$ is not a constant.

Proof: Upon replacing $R(x)$ by v we obtain

$$(5.34) \qquad 4\left(\int_0^{\pi} J(R(x)) \, f'(x) \, dx\right)^2 = \left(\int_0^1 J(v) \frac{r'}{r} \circ R^{-1}(v) \, dv\right)^2.$$

Also from (3.11) and (3.13) we know that the efficacy of the best (parametric) test for detecting shift alternatives is given by

$$(5.35) \qquad \text{eff} = \text{Inf }(f) = \int_0^{2\pi} \left(\frac{f'(x)}{f(x)}\right)^2 f(x) \, dx = \int_0^{\pi} \left(\frac{r'(x)}{r(x)}\right)^2 r(x) \, dx$$

$$= \int_0^1 \left(\frac{r'}{r} \circ R^{-1}(x)\right)^2 dx.$$

Since the efficiency is equal to the ratio of the efficacies of the two corresponding test sequences, we obtain the desired result by combining (5.25), (5.34) and (5.35).

6. Existence of an Efficient Nonparametric Test.

Interpreting the efficiency of T_n as the square of the inner product of two normalized functions gives us an immediate solution to the variational problem of finding a function $J(\cdot)$ which maximizes this efficiency.

Theorem 6.1: If for a symmetric non-uniform circular distribution $r(\cdot) = 2f(\cdot)$ is positive and $\frac{r'}{r} \circ R^{-1}(v)$ has two continuous derivatives, then the sequence of test statistics $\{T_n\}$ based on $J(v) = \frac{r'}{r} \circ R^{-1}(v)$ maximizes the expression (5.33). The maximum efficiency is equal to 1.

Proof: By the Schwarz inequality (and by the definition of efficiency) (5.33) can never exceed 1. For our particular choice it is equal to 1, and it is easy to see that this choice is legitimate: $r'(0) = 0$ by symmetry of r; $r'(\cdot)$ is not constant, since the distribution is not uniform; finally, the required derivatives exist.

Remark: It follows from Lemma 5.3 that the sequence of test statistics $\{T_n\}$ based on $J(v) = \frac{r'}{r} \circ R^{-1}(v)$ is consistent in a neighborhood of 0.

7. Efficiencies of a Few Standard Tests.

In this section we derive the Bahadur-efficiency of the Wilcoxon test and of the sign test for the case of a v. Mises distribution. These efficiencies will depend on the concentration parameter k of the distribution, and hence we will get efficiency-curves.

<u>Wilcoxon test</u>: This test corresponds to the function $J(u) = u$. The hypothesis is rejected if the sum of the ranks of the reflected (at the horizontal axis) values is too large. For the computation of the numerator of the efficacy we first evaluate (5.30). $J(0)\, f(0) - J(1)\, f(\pi) +$

$$2\int_0^{\pi} J'(2M(u))\, f(u)^2 du = 2\int_0^{\pi} C(k)^2 e^{2k \cos x} dx - e^{-k} C(k) =$$

$$\frac{C(k)^2}{C(2k)} - e^{-k} C(k) = \frac{1}{2\pi I_0(k)} \left(\frac{I_0(2k)}{I_0(k)} - e^{-k}\right), \qquad \text{by Appendix (A.3).}$$

Furthermore

$$\int_0^1 J^2(u)du = \int_0^1 u^2 du = \frac{1}{3}.$$

By (5.25) and (5.31) we obtain the efficacy of the Wilcoxon test

(7.1) eff(Wilcoxon; v. Mises distribution) =

$$\frac{3}{\pi^2 I_0(k)^2} \left(\frac{I_0(2k)}{I_0(k)} - e^{-k}\right)^2.$$

In Sec. 3 it has been shown that the efficacy of the "best" test for the location parameter θ in this case is equal to

$$\text{Inf}\ (f) = \frac{k I_1(k)}{I_0(k)} \qquad \text{(see A.10).}$$

Hence we get

(7.2) e (Wilcoxon|best test; v. Mises distribution) =

$$\frac{3}{\pi^2} \frac{(I_0(k)e^{-k} - I_0(2k))^2}{k I_0(k)^3\, I_1(k)}.$$

Part of the graph of this efficiency curve is shown on Figure 1. We are here particularly interested in the two limiting cases $k \to 0$ and $k \to \infty$.

If k is close to zero we use the power-series expansions (A.11), (A.12), and a straightforward computation yields

$$(7.3) \qquad \lim_{k \to 0} e(\text{Wilcoxon}|\text{best test; v. Mises distribution}) = \frac{6}{\pi^2}.$$

If k is very large we may use the approximations (A.13), (A.14) and we obtain

$$(7.4) \qquad \lim_{k \to \infty} e(\text{Wilcoxon}|\text{best test; v. Mises distribution}) = \frac{3}{\pi}.$$

<u>Remark</u>: In the limiting case $k \to \infty$ the efficiency of the Wilcoxon test approximates $\frac{3}{\pi}$, which is equal to the efficiency of the Wilcoxon (one- or two-sample) test for shifts in case of a normal distribution. This is not too surprising if we note that for $k \to \infty$ the "shape" of the von Mises distribution, properly normalized, approaches that of a $N(0,1)$ distribution. This is a special case of a very general result obtained by Buehler (1965), page 1880.

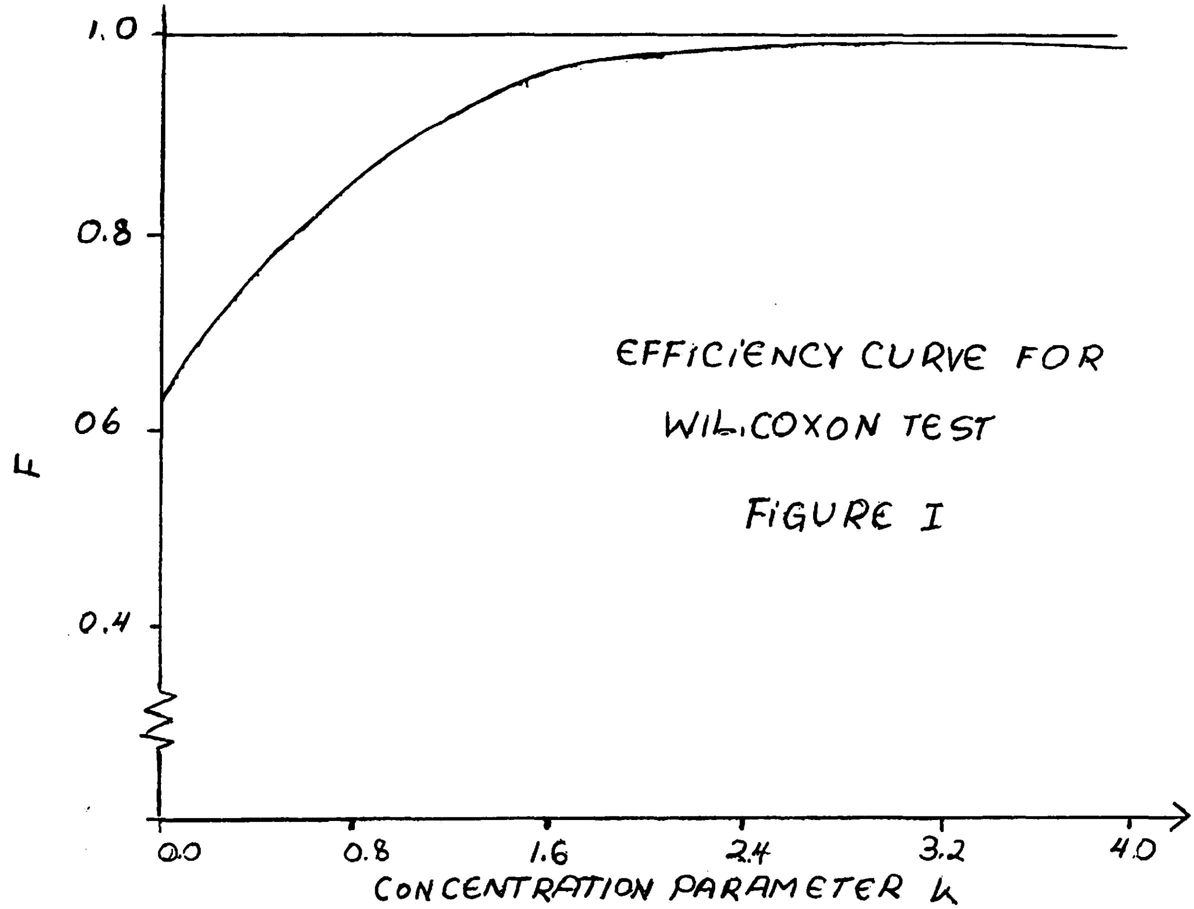
<u>Sign test</u>: The sign test rejects the hypothesis $\theta = 0$ against $K: \theta > 0$, if the number T_n of observations on the upper half-circle is too large. This test corresponds to $J(u) \equiv 1$, but here $J(\cdot)$ is a constant and our theory is not applicable. However, the Pitman-ARE can easily be computed directly.

It is well-known that

$$\text{under H:} \quad T_n \quad N\left(\frac{n}{2}, \frac{n}{4}\right),$$
$$\text{under K:} \quad T_n \quad N\left(np^{(\theta)}, np^{(\theta)}(1-p^{(\theta)})\right),$$

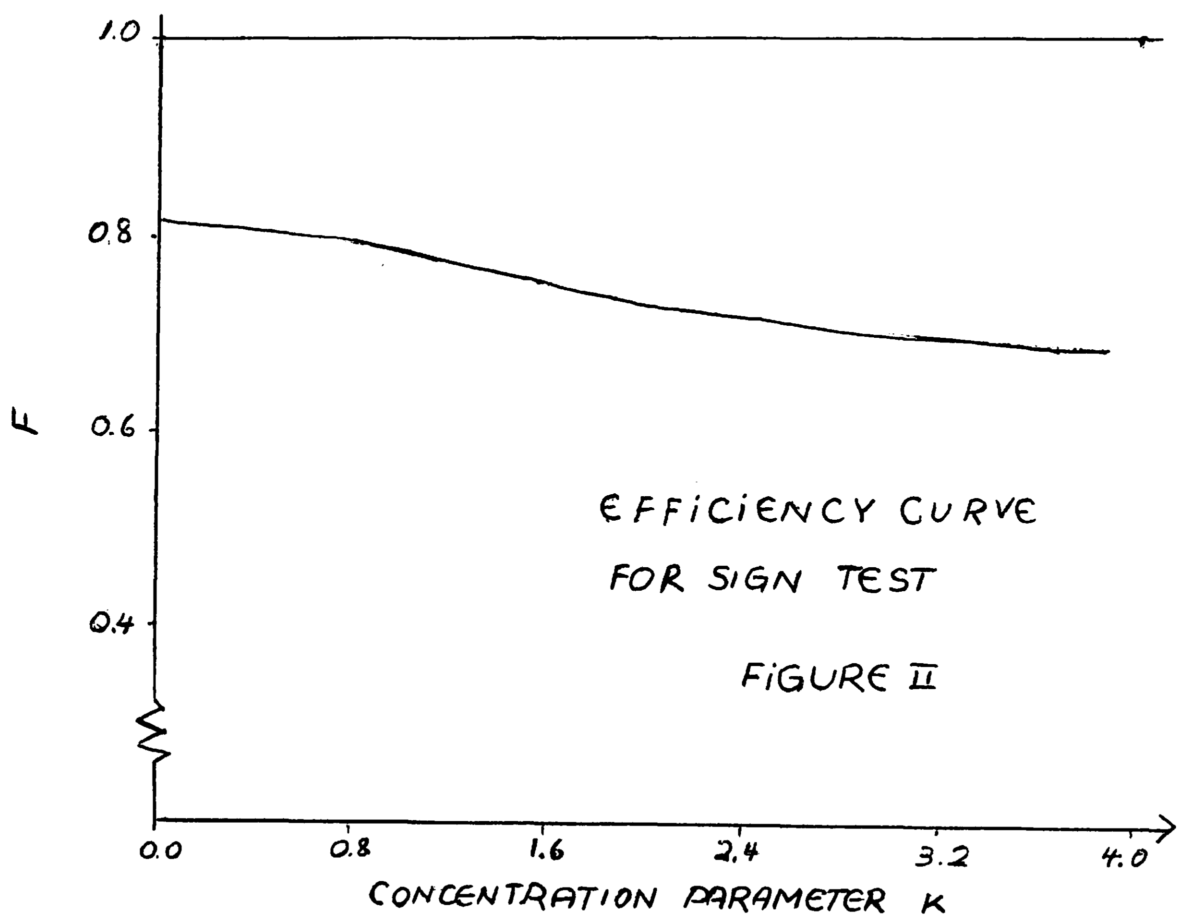
where $p^{(\theta)}$ is defined by

$$(7.5) \qquad p^{(\theta)} = \int_0^{\pi} C(k)\, e^{k \cos (x-\theta)} dx = \int_0^{\pi} f(x-\theta)\, dx.$$



EFFICIENCY CURVE FOR WILCOXON TEST

FIGURE I



EFFICIENCY CURVE FOR SIGN TEST

FIGURE II

The efficacy is

$$(7.6) \qquad \text{eff(sign test; f)} = \frac{n\left(\frac{\partial}{\partial\theta} p^{(\theta)}\Big|_{\theta=0}\right)^2}{\text{Var}_0\, T_n} = \frac{(f(0) - f(\pi))^2}{1/4},$$

since $\frac{\partial}{\partial\theta} p^{(\theta)}\Big|_{\theta=0} = [\frac{\partial}{\partial\theta} \int_0^{\pi} f(x-\theta)dx]_{\theta=0} = \frac{\partial}{\partial\theta} [\int_{-\theta}^{\pi-\theta} f(u)du]_{\theta=0} = f(0) - f(\pi).$

In the case of the v. Mises distribution we get $f(0) - f(\pi) = C(k)(e^k - e^{-k})$ $= \frac{1}{2\pi I_0(k)} (e^k - e^{-k})$, and hence we arrive at

$$(7.7) \qquad e(\text{sign test}|\text{best test; v. Mises distribution}) =$$

$$\frac{4\{\frac{1}{2\pi I_0(k)} (e^k - e^{-k})\}^2}{\frac{kI_1(k)}{I_0(k)}} = \frac{(e^k - e^{-k})^2}{\pi^2 I_0(k)\, kI_1(k)} .$$

This efficiency curve is plotted on Figure 2. We again investigate the cases $k \to 0$ and $k \to \infty$. Using the approximations (A.11) - (A.14) we obtain

$$(7.8) \qquad \lim_{k \to 0} e(\text{sign test}|\text{best test; v. Mises distribution}) = \frac{8}{\pi^2}.$$

$$(7.9) \qquad \lim_{k \to \infty} e(\text{sign test}|\text{best test; v. Mises distribution}) = \frac{2}{\pi}.$$

As one would expect, the last result again coincides with the efficiency of the sign test for detecting shifts for a normal distribution.

III. TWO-SAMPLE CASE

Let $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ be two samples from circular distributions with c.d.f. F and G, respectively. Our goal is now to find tests for the hypothesis H: the two samples have the same underlying distribution vs. K: the underlying distributions are different. In particular we will be concerned with the subset of K which consists of shift alternatives.

8. Reduction by Invariance.

As in the one-sample case the class of tests for H vs. K may be reduced considerably by the principle of invariance, if a suitable group of transformations is used. In the linear case one usually uses the order relation on R to define one-sided alternatives, either for the particular case where G is obtained by shifting F by a fixed amount $\theta > 0$, or for the more general case that G is stochastically larger than F. Since the points on a circle cannot be ordered in a "natural" way, it is impossible to define one-sided alternatives in a satisfactory way. We will therefore have to take a group of transformations which rules out one-sided tests.

Before we define a group of transformations of the sample space we redefine the term "c.d.f." in a somewhat unconventional, but for our purposes more convenient way. A monotone increasing distribution function defines a probability mass distribution μ by $\mu((a,b]) = F(b) - F(a)$. But we could just as well define a monotone decreasing c.d.f. corresponding to μ by $\mu([a,b)) = F(a) - F(b)$. We will allow both versions for representing the probability measure and hence we arrive at the

<u>Definition</u>: F is a c.d.f. for the circle if

(a) $F: R \to R$, monotone.

(b) $F(x + 2n\pi) = F(x) \pm n$ for every $x \in R$, every integer n.

It is obvious that such a function F defines a unique probability distribution on the interval $[0,2\pi)$ (and hence on the circle), and that any probability measure on the circle defines such a c.d.f. up to a constant sum or difference.

Throughout this chapter we assume that the two samples are from continuous distributions and that the distribution functions are strictly increasing. I.e., we assume that F and G satisfy part (a) of the above definition in the strict sense and that they are continuous. Under these conditions the states of nature Θ for our testing problem is given by the set of pairs

$$(8.1) \qquad \Theta = \{(F,G): \ F, G \text{ are continuous, strictly monotone c.d.f.'s}\}.$$

H is the subset defined by

$$(8.2) \qquad H = \{(F,G): \ F - G = \text{const.} \ \text{ or } \ F + G = \text{const.}\}.$$

Transformation group: As our class T of transformations we take the set of all homeomorphisms of the circle C onto itself, i.e., all bi-continuous, one-to-one mappings of C onto C. It follows from the definition of a homeomorphism that T has a group structure if the composition of two such mappings it taken as group operation. Because of the one-to-one correspondence between C and $J = [0,2\pi)$ any one-to-one mapping t of C onto C defines a corresponding mapping t' of J onto J, and vice versa. Continuity is preserved except possibly at the point $z = 0$, where t' may and in general will have a jump of size 2π. By defining a suitable mapping t' from J into $[0,4\pi]$ we can find a continuous version of t' which still corresponds to t in the sense that the homeomorphism t of C onto C can be written in the form

$$t: \ e^{ix} \to e^{it'(x)}.$$

m and n). Take an arbitrary cut-off point and an arbitrary direction on one of the circles and determine the ranks $R_1^{(1)},\ldots,R_m^{(1)}$ of the X-observations in the combined sample of the first pair.

Definition: The two pairs of samples have the same arrangement if it is possible to find a cut-off point and a direction on the second circle in such a way that for the resulting ranks $R_i^{(2)}$ of the X-observations we get $R_i^{(2)} = R_i^{(1)}$ $(i = 1,\ldots,m)$.

"Same arrangement" is an equivalence relation. The term "arrangement" is used for an equivalence class defined by this relation or for a particular sequence of ranks which is a member of this equivalence class.

Lemma 8.1: The arrangement of a combined sample is a maximal invariant under T.

Proof: Since a homeomorphism on C can be represented on J by a strictly monotone increasing or a strictly monotone decreasing function, it follows easily that the arrangement of a pair of samples remains invariant under any t. (The directions have to be the same iff t' is monotone increasing.) Conversely, if two sample pairs have the same arrangement we obtain a suitable $t \in T$ by first making a rotation which shifts the second cut-off point onto the first. If the directions are different we use a reflection, which leaves the cut-off points unchanged, and which converts the direction corresponding to the second sample pair to that of the first. Finally, spaces between successive observations can easily be adjusted by means of a suitable homeomorphism. The combination of these mappings is a member of T and it maps the second sample pair onto the first; this shows that "arrangement" is in fact a maximal invariant.

Distribution-free tests: By a well-known result by Lehmann the distribution of an invariant test statistic depends only on the orbit of (defined by the induces group $\overline{T}$) to which (F,G) belongs. We use this simple fact to show that all invariant test statistics are non-parametric, i.e., they have the same distribution for each pair $(F,G) \in H$.

<u>Lemma 8.2</u>: Let (F,G) and (F',G') be elements of H. Then there exists a $t \in T$ such that $\bar{t}(F,G) = (F',G')$.

<u>Proof</u>: Since every c.d.f. can be replaced by its negative without changing the distribution we may w.l.o.g. assume that F, G, F', G' are all increasing. We may even go one step further and assume that $F \equiv G$ and $F' \equiv G'$, since a shift by a constant has no influence on the distribution. The mapping $t' = (F')^{-1} \circ F$ from R onto R is continuous, strictly increasing and satisfies

$$t'(x + 2k\pi) = (F')^{-1} \circ F(x + 2k\pi) = (F')^{-1}(F(x) + k)$$

$$= (F')^{-1} \circ F(x) + 2\pi k = t'(x) + 2\pi k.$$

Hence $e^{ix} \to e^{it'(x)}$ is a homeomorphism t and

$$\bar{t}(F,G) = (F \circ (t')^{-1}, \ G \circ (t')^{-1}) = (F \circ F^{-1} \circ F', \ G \circ F^{-1} \circ F')$$

$$= (F', G') \qquad \text{since } F' = G'.$$

This completes our proof.

<u>Corollary</u>: Test statistics which are invariant under T are distribution free.

<u>Proof</u>: This follows directly from the theorem by Lehmann mentioned above. See, e.g., E. L. Lehmann (1959), p. 220, Theorem 3.

9. Efficacy of the Best Parametric Test for Detecting Shift Alternatives.

In Section 3 we discussed the concept of the relative efficiency of two test sequences as the limiting ratio $\frac{n'}{n}$ of the sample sizes, which make the two tests equally powerful for some given alternative. It turned out, that the so-called Bahadur-efficiency is an approximate measure of this ratio which is particularly suitable for our purpose, since it does not require assumptions on the distribution of the test statistics which are as strong as those needed for Pitman's concept of efficiency. For the same reason we will use Bahadur's notion of efficiency for the two-sample problem.

If for a given sequence of test statistics T_N either Pitman's conditions or Assumptions 1-4 of Section 3 are satisfied we can define and compute the efficacy eff (T_N) of this sequence in such a way that the ARE is simply the ratio of the efficacies.

In order to have a standard for comparison for any nonparametric test sequence, we now derive the maximum efficacy that any test for detecting shift alternatives might achieve, provided that certain regularity conditions are satisfied. Since we are not primarily interested in parametric tests, we will not state our results in the form of theorems with all the regularity conditions. We rather assume that the required derivatives exist and that integration and differentiation can be interchanged. The derivations below follow very closely the corresponding arguments for the one-sample case in Section 3.

For the class of parametric tests we are going to consider, we impose a (rather weak) invariance condition: We will only consider tests for which the critical region (and hence the acceptance region) is invariant under a rotation of the X- and Y-observations by the same angle φ (invariance under the circle group).

Let $f(\cdot) > 0$ be a density for a circular distribution, i.e.,

$$(9.1) \qquad f(x + 2k\pi) = f(x), \quad k = \pm 1, \pm 2 \ldots \text{ and } \int_0^{2\pi} f(x)dx = 1.$$

Let $X_1, \ldots, X_m$ be a sample from a distribution with density $f(x - \varphi)$ and let $Y_1, \ldots, Y_n$ be a sample from the distribution with density $f(x - \theta)$ $(0 \le x \le 2\pi)$. Assume that $\frac{m}{N} \to \lambda$, as $N \to \infty$, where $0 < \lambda < 1$. Set $\Delta = \theta - \varphi$. We are concerned with the problem of testing $H: \Delta = 0$ vs. $K: \Delta > 0$.

Upper bound for the efficacy: In order to derive an upper bound for the efficacy we may assume the particular situation $\varphi = -(1 - \lambda)\Delta$, $\theta = \lambda\Delta$. By the Neyman-Pearson lemma the best test for a specific $\Delta > 0$ rejects H if

$$(9.2) \qquad T'_{m,n,\Delta} = \sum_{i=1}^{m} \log \frac{f(x_i + (1-\lambda)\Delta)}{f(x_i)} + \sum_{j=1}^{n} \log \frac{f(y_j - \lambda\Delta)}{f(y_j)} > C_\Delta$$

or, equivalently,

$$(9.3) \qquad T_{m,n,\Delta} = \frac{1}{\Delta} \sum_{i=1}^{m} \log \frac{f(x_i + (1-\lambda)\Delta)}{f(x_i)} + \frac{1}{\Delta} \sum_{j=1}^{n} \log \frac{f(y_j - \lambda\Delta)}{f(y_j)} > C_\Delta.$$

By the Central Limit Theorem $T_{m,n,\Delta}$ is asymptotically normally distributed. Since tests based on $\{T_{m,n,\Delta}\}$ are tailor-made for the particular Δ under consideration, we obtain an upper bound for the efficacy of any test of $H: \Delta = 0$ vs. $K: \Delta > 0$ by formally computing the efficacy of the sequence $\{T_{m,n,\Delta}\}$ just as in the one-sample case:

$$\frac{E_\Delta T_{m,n,\Delta} - E_0 T_{m,n,0}}{\Delta} =$$

$$m \int_0^{2\pi} \frac{\log f(x + (1-\lambda)\Delta) - \log f(x)}{\Delta} \, \frac{f(x + (1-\lambda)\Delta) - f(x)}{\Delta} \, dx$$

$$+ \, n \int_0^{2\pi} \frac{\log f(x - \lambda\Delta) - \log f(x)}{\Delta} \, \frac{f(x - \lambda\Delta) - f(x)}{\Delta} \, dx$$

$$\xrightarrow[\Delta\to 0]{} m(1-\lambda)^2 \int_0^{2\pi} \left.\frac{\partial \log f(x+\Delta)}{\partial\Delta}\right|_{\Delta=0} \left.\frac{\partial f(x+\Delta)}{\partial\Delta}\right|_{\Delta=0} dx$$

$$+ n\lambda^2 \int_0^{2\pi} \left.\frac{\partial \log f(x+\Delta)}{\partial\Delta}\right|_{\Delta=0} \left.\frac{\partial f(x+\Delta)}{\partial\Delta}\right|_{\Delta=0} dx$$

$$= [m(1-\lambda)^2 + n\lambda^2] \int_0^{2\pi} \frac{f'(x)^2}{f(x)} dx = [m \frac{n^2}{N^2} + n \frac{m^2}{N^2}] \text{ Inf } (f)$$

$$= \frac{mn}{N} \text{ Inf } (f) = N\,\lambda(1-\lambda) \text{ Inf } (f).$$

In a similar way it can be shown that $\text{var}_0\, T_{m,n,\Delta}$ converges to the same quantity as $\Delta \to 0$. Thus an upper bound for the efficacy of any test sequence $\{T_N\}$ is given by

(9.4) $$\text{eff}_u\,(T_N) = \lambda(1-\lambda) \text{ Inf } (f).$$

Example of an efficient test: Next we claim that there exists an invariant (under the circle group) test which achieves this upper bound, provided that some regularity conditions are satisfied. This has been shown by Chernoff and Savage (1958), page 983. It is a direct consequence of the large sample theory of the maximum likelihood estimator: $\hat{\Delta} = \hat{\varphi} - \hat{\theta}$ ($\hat{\varphi}$ and $\hat{\theta}$ are MLE) is asymptotically normally distributed with mean Δ and variance $[N\,\lambda(1-\lambda) \text{ Inf } (f)]^{-1}$, hence the efficacy of a test based on $\hat{\Delta}$ is equal to $\lambda(1-\lambda)$ Inf (f), and thus it coincides with the upper bound given by (9.4).

For the special case of the v. Mises distribution we derive the likelihood ratio test, since it has an interesting geometric interpretation; we also show that it is invariant and efficient. Let $\hat{\varphi}$, $\hat{\theta}$ be the MLE of the first and second sample, respectively. Let $\hat{\hat{\varphi}}$ be the MLE of the combined sample. Then it is easy to see that

From likelihood ratio theory it is well-known that $-2 \log LR_N$ has a χ_1^2-distribution as $N \to \infty$; hence Assumption 1 of Section 3 is satisfied. It was shown by Bahadur (1960) that (3.2) is satisfied for the c.d.f. of a χ_1^2-variable with $a = 1$, $t = 1$. The correct normalizing sequence is $b(N) = N$, since under K:

$$(9.9) \qquad \frac{-2 \log LR_N}{N} = 2k \Bigg\{ [\lambda^2 (\frac{\Sigma \sin x_i}{m})^2 + \lambda^2 (\frac{\Sigma \cos x_i}{m})^2]^{\frac{1}{2}}$$

$$+ [(1-\lambda)^2 (\frac{\Sigma \sin y_i}{n})^2 + (1-\lambda)^2 (\frac{\Sigma \cos y_i}{n})^2]^{\frac{1}{2}}$$

$$- \Bigg[\lambda^2 [(\frac{\Sigma \sin x_i}{m})^2 + (\frac{\Sigma \cos x_i}{m})^2]$$

$$+ (1-\lambda)^2 [(\frac{\Sigma \sin y_i}{n})^2 + (\frac{\Sigma \cos y_i}{n})^2]$$

$$+ 2\lambda(1-\lambda) [\frac{\Sigma \Sigma (\sin x_i \sin y_j + \cos x_i \cos y_j)}{n\, m}] \Bigg]^{\frac{1}{2}} \Bigg\}$$

where $i = 1,\ldots,m$ and $j = 1,\ldots,n$.

According to formulae (A.6), (A.7) of the Appendix we get, by the strong law of large numbers,

$$(9.10) \qquad \frac{-2 \log LR_N}{N} \xrightarrow{a.s.} 2k \Big\{ \lambda E_0 \cos X[\sin^2 \theta + \cos^2 \theta]^{\frac{1}{2}}$$

$$+ (1-\lambda) E_0 \cos X[\sin^2 \varphi + \cos^2 \varphi]^{\frac{1}{2}}$$

$$- [\lambda^2 (E_0 \cos X)^2 (\sin^2 \theta + \cos^2 \theta)$$

$$+ (1-\lambda)^2 (E_0 \cos X)^2 (\sin^2 \varphi + \cos^2 \varphi)$$

$$+ 2\lambda(1-\lambda)(E_0 \cos X)^2 (\sin \theta \sin \varphi + \cos \theta \cos \varphi)]$$

$$= 2k E_0 \cos X \{1 - \sqrt{\lambda^2 + (1-\lambda)^2 + 2\lambda(1-\lambda) \cos (\theta - \varphi)} \}$$

$$= 2k E_0 \cos X \{1 - \sqrt{1 - 2\lambda(1-\lambda)(1 - \cos \Delta)} \}$$

$$= C(\Delta).$$

Thus (3.3) is satisfied. The efficacy of the test sequence is now given by

$$(9.11) \qquad \text{eff}\ (LR_N) = \frac{1}{2}\frac{d^2C(\Delta)}{d\Delta^2}\bigg|_{\Delta=0} = \lambda(1-\lambda)k\ E_0 \cos X = \lambda(1-\lambda)\ \text{Inf}\ (f)$$

by Appendix (A.5) and (A.10). This result shows that the likelihood ratio test is an efficient test.

10. Locally Most Powerful Invariant Tests for Shift Alternatives.

In Section 8 we showed that in the two-sample problem with continuous c.d.f.'s F and G, an invariant (under T) test statistic for testing $H: F \equiv G$ vs. $K: F \not\equiv G$ should depend only on the "arrangement" of the x's and the y's. Since it is easier to work with order statistics instead of with "arrangements" we take an arbitrary cut-off point on the circle and specify a direction (e.g., counterclockwise). Then we order the combined sample, starting at the cut-off point and going in the direction specified. Thus to each sample outcome there corresponds a rank-sequence $(r_1,\ldots,r_m)$ of the y's. A test can now be defined by choosing a critical region (CR) in the sample space of such rank-sequences. We are, however, not completely free in choosing such a CR if the test is to be invariant under the transformation group T. In fact, if $(r_1,\ldots,r_m)$ is to be in the CR, then, in order to make the test independent of the cut-off point, we have to put into the CR all rank sequences of the form $([r_1 + k],\ldots,[r_m + k])$ for $0 \le k \le N-1$, where

$$[\ell] = 1 + (\ell-1) \bmod N. \tag{10.1}$$

(This somewhat uncommon definition of the symbol $[\cdot]$ is convenient, since we get $[N] = N$, whereas $N \bmod N = 0$.) We call rank sequences of the form $([r_1 + k],\ldots,[r_m + k])$ $(0 \le k \le N-1)$ rotations of $(r_1,\ldots,r_m)$. In addition, if the test is to be independent of the direction, for every point $(r_1,\ldots,r_m) \in CR$ we have to have $(r_1',\ldots,r_m') = (N+1-r_m,\ldots,N+1-r_1) \in CR$. We call $(r_1',\ldots,r_m')$ the inversion of $(r_1,\ldots,r_m)$.

<u>Theorem 10.1</u>: Let $f(\cdot) > 0$ be the density of a circular distribution (i.e., $f(\cdot)$ is periodic with period 2π and $\int_0^{2\pi} f(x)\,dx = 1$). Assume that $f''(\cdot)$ exists and is continuous. If X has density $f(x)$ whereas Y has density $f(x-\Delta)$, then a locally most powerful T-invariant test for $H: \Delta = 0$ against $K: \Delta \ne 0$ exists. Its rejection region is

$$(10.2) \qquad \sum_{k=0}^{N-1} \sum_{j=1}^{m} \sum_{i=1}^{m} E\left[\frac{f'(V^{[r_i+k]})f'(V^{[r_j+k]})}{f(V^{[r_i+k]})\,f(V^{[r_j+k]})}\right] > C,$$

where. $V^{(1)}, V^{(2)},\ldots,V^{(N)}$ is the order statistic of a sample of size N from the X-distribution.[1]

<u>Proof</u>: If we take a specific cut-off point and direction, then Hoeffding's result, mentioned in Section 4, yields immediately

$$(10.3) \qquad P_\Delta(R_1 = r_1,\ldots,R_m = r_m) = \frac{1}{\binom{N}{m}} E\left[\frac{f(V^{(r_1)}-\Delta)\cdots f(V^{(r_m)}-\Delta)}{f(V^{(r_1)})\cdots f(V^{(r_m)})}\right].$$

If the test is to be invariant under T, then with each rank sequence $(r_1,\ldots,r_m) \varepsilon$ CR we have to take all its rotations into the CR.

(a) Let us first assume that for each rank order $(r_1,\ldots,r_m)$ all rotations are different, i.e., for no $0 \le k \le N-1$ is $([r_1+k],\ldots,[r_m+k])$ a permutation of $(r_1,\ldots,r_m)$. In this case the power we gain by putting $(r_1,\ldots,r_m)$ (and also the corresponding rotations and inversions) into the CR is equal to Q_Δ, where Q_Δ is defined by

$$(10.4) \qquad Q_\Delta(r_1,\ldots,r_m) = \frac{\delta(r_1,\ldots,r_m)}{\binom{N}{m}} \sum_{k=0}^{N-1} E\left[\frac{f(V^{[r_1+k]}-\Delta)\cdots f(V^{[r_m+k]}-\Delta)}{f(V^{[r_1+k]})\cdots f(V^{[r_m+k]})}\right];$$

here

$$(10.5) \qquad \delta\,(r_1,\ldots,r_m) = \begin{cases} 1 & \text{if the inversion of } (r_1,\ldots,r_m) \text{ is equal to some rotation;} \\ 2 & \text{otherwise.} \end{cases}$$

Under H the arrangement $(r_1,\ldots,r_m)$ together with its inversion and its rotations has the probability

$$(10.6) \qquad Q_0(r_1,\ldots,r_m) = \frac{\delta(r_1,\ldots,r_m)}{\binom{N}{m}} N.$$

[1]We write $V^{[r_j]}$ instead of the correct expression $V^{([r_j])}$.

Hence it follows that in order to maximize the power in a neighborhood of $\Delta = 0$ we have to put into CR those rank sequences $(r_1,\ldots,r_m)$ for which $\frac{1}{\delta(r_1,\ldots,r_m)} \frac{\partial}{\partial\Delta} Q_\Delta(r_1,\ldots,r_m)\Big|_{\Delta=0}$ is largest.

We now evaluate this expression. The Lebesgue dominated convergence theorem allows us to interchange expectation and differentiation, hence we get

$$\frac{\partial Q_\Delta(r_1,\ldots,r_m)}{\partial\Delta} = -\frac{\delta(r_1,\ldots,r_m)}{\binom{N}{m}} \sum_{k=0}^{N-1} \sum_{i=1}^{m} E\left[\frac{f(V^{[r_1+k]}-\Delta)\cdots f'(V^{[r_i+k]}-\Delta)\cdots f(V^{[r_m+k]}-\Delta)}{f(V^{[r_1+k]})\cdots f(V^{[r_i+k]})\cdots f(V^{[r_m+k]})}\right] \tag{10.7}$$

and thus

$$\frac{1}{\delta(r_1,\ldots,r_m)} \frac{\partial Q_\Delta(r_1,\ldots,r_m)}{\partial\Delta}\Big|_{\Delta=0} = -\frac{1}{\binom{N}{m}} \sum_{i=1}^{m} \sum_{k=0}^{N-1} E\left[\frac{f'(V^{[r_i+k]})}{f(V^{[r_i+k]})}\right] \tag{10.8}$$

$$= \text{const.},$$

i.e., the expression we want to maximize does not depend on $(r_1,\ldots,r_m)$. Thus it follows that we have to take those arrangements into the CR for which $\frac{1}{\delta(r_1,\ldots,r_m)} \frac{\partial^2}{\partial\Delta^2} Q_\Delta(r_1,\ldots,r_m)\Big|_{\Delta=0}$ is largest.

By applying the dominated convergence theorem again we obtain

$$\frac{1}{\delta(r_1,\ldots,r_m)} \frac{\partial^2 Q_\Delta(r_1,\ldots,r_m)}{\partial\Delta^2}\Big|_{\Delta=0} = \tag{10.9}$$

$$\frac{1}{\binom{N}{m}} \sum_{k=1}^{N-1} \sum_{i=1}^{m} \sum_{\substack{j=1 \\ i\neq j}}^{m} E\left[\frac{f'(V^{[r_i+k]})}{f(V^{[r_i+k]})} \frac{f'(V^{[r_j+k]})}{f(V^{[r_j+k]})}\right]$$

$$+ \sum_{k=0}^{N-1} \sum_{i=1}^{m} E\left[\frac{f''(V^{[r_i+k]})}{f(V^{[r_i+k]})}\right] .$$

The second term is again independent of the particular arrangement. In maximizing the second derivative of Q_Δ at $\Delta = 0$ we may and will omit this term. For convenience we add another constant corresponding to the "diagonal" terms where $i = j$:

$$\frac{1}{\binom{N}{m}} \sum_{k=0}^{N-1} \sum_{i=1}^{m} E\left[\frac{f'(V^{[r_i+k]})}{f(V^{[r_i+k]})}\right]^2 .$$

It thus follows that under the assumptions of this part of the proof the most powerful test, which is based on the rank statistic $(R_1,\ldots,R_m)$ and independent of the particular cut-off point, has critical region

$$(10.10) \qquad \sum_{k=0}^{N-1} \sum_{j=1}^{m} \sum_{i=1}^{m} E\left[\frac{f'(V^{[r_i+k]})}{f(V^{[r_i+k]})} \frac{f'(V^{[r_j+k]})}{f(V^{[r_j+k]})}\right] > C.$$

It is easy to see that

$$(10.11) \qquad \sum_{k=0}^{N-1} E\left[\frac{f'(V^{[r_i+k]})}{f(V^{[r_i+k]})} \frac{f'(V^{[r_j+k]})}{f(V^{[r_j+k]})}\right]$$

is a function of $r_i - r_j$ only. We denote it by $h_N'(\cdot)$, i.e.,

$$(10.12) \qquad h_N'(i - j) = \sum_{k=0}^{N-1} E\left[\frac{f'(V^{[i+k]})}{f(V^{[i+k]})} \frac{f'(V^{[j+k]})}{f(V^{[j+k]})}\right]; \qquad i,j = 0, \pm 1, \pm 2,\ldots$$

It is also quite obvious that $h_N'(i-j) = h_N'(j-i)$, and this shows that the test (10.10) is invariant under an inversion. Since the test (10.10) is locally most powerful among all tests invariant under rotations, and is also invariant under inversions, it is a locally most powerful invariant (under T) test for testing $H: \Delta = 0$ vs. $K: \Delta \neq 0$.

(b) Assume now that there exists a rank sequence $(r_1,\ldots,r_m)$ and a $\bar{k}$, $0 < \bar{k} \leq N-1$ such that $([r_1 + \bar{k}],\ldots,[r_m + \bar{k}])$ is a permutation of $(r_1,\ldots,r_m)$. Let $\bar{k}$ be the smallest number with this property. Then obviously $\bar{k}$ divides N. Let ℓ be defined by $N = \ell\bar{k}$. Then under H this particular arrangement has only probability

$\frac{N}{\ell} \frac{\delta(r_1,\ldots,r_m)}{\binom{N}{m}}$ instead of $N \frac{\delta(r_1,\ldots,r_m)}{\binom{N}{m}}$. But it is easy to see that in the definition (10.4) of the power function $Q_\Delta(r_1,\ldots,r_m)$ k should only take the values $0, 1,\ldots,\bar{k}-1$, i.e., $Q_\Delta(r_1,\ldots,r_m)$ should be divided by a factor $\ell = N/\bar{k}$ under K also. But then it follows immediately from the Neymann-Pearson lemma that such a division of the power function by the same factor under H and under K does not change the optimality of a test. Thus the test derived under (a) above, under more specific conditions, is also a locally most powerful invariant test in the general case

q.e.d.

11. Definition and Representations of a Class of Invariant Two-Sample Statistics.

In the previous section we derived a locally most powerful invariant test for detecting shift alternatives. Using (10.12) the test can be written in the form

$$(11.1) \qquad T_N' = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{m} h_N'(R_i - R_j) > C.$$

In this section we will define a class of test statistics with an arbitrary function h_N'. More specifically: Let $h_N(\cdot)$ be a function defined for all values of the form $\frac{i}{N}$, $i = 0, \pm 1, \ldots, \pm N$. Assume that $h_N(\cdot)$ is symmetric with respect to 0 and with respect to $\frac{1}{2}$. (It is easy to see that $h_N(\cdot)$ is then periodic with period 1.) We will consider tests based on statistics of the form

$$(11.2) \qquad T_N = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{m} h_N \left(\frac{R_j - R_i}{N}\right).$$

It is obvious that the LMP invariant test has the form $T_N > C$ if we define $h_N(\frac{i}{N}) = h_N'(i)$. It is also easy to see that statistics of the form (11.2) are invariant under the group T: They are based on ranks, are independent of the cut-off point (only differences of ranks enter into T_N, and h_N is periodic), and are invariant under inversions (since $h_N(\cdot)$ is symmetric with respect to 0). This also shows that T_N is non-parametric.

<u>Alternative representations of T_N</u>: If we define a vector $z = (z_1, \ldots, z_N)'$ by

$$(11.3) \qquad z_i = \begin{cases} 1 & \text{if } i^{th} \text{ observation in combined sample is from the X-sample} \\ 0 & \text{otherwise,} \end{cases}$$

where we assume that arbitrary, but fixed, cut-off point and direction have been chosen. Using these "indicators" we may obviously write

$$(11.4) \qquad T_N = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} h_N \left(\frac{i-j}{N}\right) z_i z_j .$$

This representation of T_N as a quadratic form in the z_i's will be of interest, when we derive the limiting distribution of T_N in the next section.

If we define $g_i = \sum_{k=1}^{N} z_k z_{[k+i]}$, $i = 1,\ldots,N$, then g_i represents the number of times a pair of X-values is separated by i "gaps," that is by $i-1$ X- or Y-observations. g_i depends on the direction, but $\bar{g}_i = g_i + g_{N-i}$, $i = 1,\ldots,N$, does not, as can be seen easily. We call the symmetric vector $\bar{g} = (\bar{g}_1,\ldots,\bar{g}_N)$ the "gap structure" of the sample.[1] In terms of this gap structure we obtain

$$(11.5) \qquad T_N = \frac{1}{N} \sum_{i=1}^{N} h_N \left(\frac{i}{N}\right) \bar{g}_i .$$

A final representation of T_N will prove useful when we derive the limiting distribution of T_N under shift alternatives in Section 13. Let $F_m(\cdot)$, $G_n(\cdot)$ be the empirical c.d.f. (on $[0,2\pi]$) of the X- and Y-observations, respectively. Set $\lambda_N = \frac{m}{N}$ and

$$(11.6) \qquad H_N(x) = \lambda_N F_m(x) + (1-\lambda_N) G_n(x),$$

then $H_N(\cdot)$ is the empirical c.d.f. of the combined sample. From (11.4) it follows immediately that

$$(11.7) \qquad T_N = \frac{m^2}{N} \int_0^{2\pi} \int_0^{2\pi} h_N(H_N(x) - H_N(y))\, dF_m(x)\, dF_m(y).$$

We now derive, under certain assumptions, the asymptotic distribution of T_N under H and under K.

---

[1] It is obvious that the gap structure of a sample is invariant under T. It has been conjectured that it is in fact a maximal invariant. We have not been able to either prove or disprove this conjecture.

12. The Asymptotic Distribution of T_N under H.

α. Limit Functions h and Equivalence Classes of Sequences $\{h_N\}$

Before any statement about the limiting behavior of a sequence $\{T_N, \; N = 1, 2,\dots\}$ of the form (11.2) can be made, we have to make some assumptions about the sequence $\{h_N(\cdot), \; N = 1, 2,\dots\}$. Actually $h_N(\cdot)$ has to be defined only at the points $\frac{K}{N}$, where $K = 0, \pm 1, \pm 2,\dots,\pm N$, in order to make the expression (11.2) meaningful, but we assume that each $h_N(\cdot)$ is defined on $[-1,1]$ in order to facilitate the presentation of proofs and results. This is no real restriction on the class of statistics T_N. By $L_2[0,1]$ or L_2 we denote the space of square integrable functions on $[0,1]$. L_2 is a Hilbert space if the usual definition of inner product is used, and we denote by $||h||_{L_2}$, or simply $||h||$, the norm of an element of this Hilbert space.

Approximating sequences and equivalence classes of sequences: Throughout this section we make the following two assumptions on a sequence written in the form $\{h_N(\cdot)\}$ or $\{h_N\}$:

(i) $h_N(\cdot)$ is defined on $[-1,1]$, symmetric with respect to 0 and periodic with period 1.

(ii) $h_N(\cdot)$ is a step function; it is constant on $(\frac{2K-1}{2N}, \frac{2K+1}{2N})$, $K = 0, \pm 1, \pm 2, \dots$.[1]

Definition: We say $\{h_N\} \Rightarrow h$ ("$\{h_N\}$ converges to h" in this particular sense) if the following two conditions are satisfied:

(a) h is continuous, $\int_0^1 h(x)dx = 0$ and

$$||h - h_N||_{L_2} \to 0 \text{ as } N \to \infty.^{2}$$

[1] Note that the elements $h_N(\cdot)$ form a linear subspace of $L_2[-1,1]$.

[2] This obviously implies that $h(\cdot)$ is symmetric and periodic, since a subsequence of $h_N(\cdot)$ converges pointwise a.e. and since continuity implies uniqueness of the limit function.

(b) $h_N(0) \to h(0)$ as $N \to \infty$.

We also assume that $\lambda_N = \frac{m}{n} = \frac{m_N}{N} \to \lambda$ as $N \to \infty$ $(0 < \lambda < 1)$.

On the basis of these assumptions we show that the limiting distribution of $\{T_N\}$, if it exists, is a function of $h(\cdot)$ only.

If we consider several sequences $\{h_N\}$, $\{g_N\}$, we sometimes denote the corresponding statistics by T_{h_N}, T_{g_N}, respectively.

Theorem 12.1: (a) Let $\{h_N\}$ be a converging sequence $(h_N \Rightarrow h)$. Then

$$ET_N = h(0)\,\lambda(1-\lambda) + o_N(1) \quad \text{and} \tag{12.1}$$

$$\text{var}\, T_N = 2\lambda^2 (1-\lambda)^2 \,||h||^2 + o_N(1). \tag{12.2}$$

(b) If $\{g_N\}$, $\{h_N\}$ are two sequences satisfying (i) and (ii) and if $g_N(0) - h_N(0) \to 0$, $||g_N - h_N||_{L_2} \to 0$, i.e., if $\{g_N - h_N\} \Rightarrow 0$, then $E(T_{g_N} - T_{h_N})^2 \to 0$ as $N \to \infty$.

Proof of (a): We first give the proof under the assumption that for each N, $\sum_{i=1}^{N} h_N(\frac{i}{N}) = 0$. Using representation 11.4 we first compute

$$ET_N^2 = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{i'=1}^{N} \sum_{j'=1}^{N} h_N\left(\frac{i'-j'}{N}\right) h_N\left(\frac{i'-j'}{N}\right) z_i z_j z_{i'} z_{j'}.$$

Obviously

$$Ez_i z_j z_{i'} z_{j'} = \begin{cases} \frac{m}{n} & \text{if all indices are equal,} \\ \frac{m}{N}\frac{m-1}{N-1} & \text{if any three indices are equal and the fourth is different or if any two pairs of indices are equal,} \\ \frac{m}{N}\frac{m-1}{N-1}\frac{m-2}{N-2} & \text{if there is exactly one pair of equal indices,} \\ \frac{m}{N}\frac{m-1}{N-1}\frac{m-2}{N-2}\frac{m-3}{N-3} & \text{if all four indices are different,} \end{cases}$$

for N large enough to make $m_N \geq 3$.

For convenience we use the symbol λ_{-k} to denote $\frac{m-k}{N-k}$, where we suppress the N on which λ_{-k} actually depends. Obviously $\lambda_{-k} \to \lambda$ as $N \to \infty$, k fixed. To compute ET_N^2 we partition the set of quadruples of indices into seven subsets P_k:

$P_1 = \{(i,j,i',j') : i = j = i' = j'\}$,

$P_2 = \{(i,j,i',j') :$ three indices are equal, one is different$\}$,

$P_3 = \{(i,j,i',j') : i = j, \ i' = j', \ i \neq i'\}$,

$P_4 = \{(i,j,i',j') : i = i', \ j = j', \ i \neq j$ or $i = j', \ j = i', \ i \neq j\}$,

$P_5 = \{(i,j,i',j') : i = j$ or $i' = j'$, and all other indices are different

$P_6 = \{(i,j,i',j') : i = i'$ or $i = j'$ or $j = j'$, and all other indices are different$\}$,

$P_7 = \{(i,j,i',j') :$ all indices are different$\}$.

On each of these partitions sets $Ez_i z_j z_{i'} z_{j'}$ is constant, hence it essentially suffices to compute

$$\sum_{P_k} h_N\left(\frac{i-j}{N}\right) h_N\left(\frac{i'-j'}{N}\right)$$

for each k. For convenience we define $c_h^{(N)} = \frac{1}{N} \sum_{i=1}^{N} h_N\left(\frac{i}{N}\right)^2$.

$$\sum_{P_1} h_N\left(\frac{i-j}{N}\right) h_N\left(\frac{i'-j'}{N}\right) = N\, h_N(0)^2.$$

$$\sum_{P_2} h_N\left(\frac{i-j}{N}\right) h_N\left(\frac{i'-j'}{N}\right) = 4 \sum_{i=1}^{N} \sum_{j\neq i}^{N} h_N\left(\frac{i-j}{N}\right) h_N(0)$$

$$= 4 \sum_{i=1}^{N} (-h_N(0)\, h_N(0)) = -4N\, h_N(0)^2,$$

(making use of the symmetry property and of the fact that $\sum_{1}^{N} h_N\left(\frac{i}{N}\right) = 0$).

$$\sum_{P_3} h_N\left(\frac{i-j}{N}\right) h_N\left(\frac{i'-j'}{N}\right) = N(N-1)\, h_N(0)^2.$$

$$\sum_{P_4} h_N\left(\frac{i-j}{N}\right) h_N\left(\frac{i'-j'}{N}\right) = 2\sum_{i=1}^{N} \sum_{j\neq i} h_N\left(\frac{i-j}{N}\right)^2 = 2\sum_{i=1}^{N} (Nc_h^{(N)} - h_N(0)^2)$$

$$= 2N(Nc_h^{(N)} - h_N(0)^2).$$

$$\sum_{P_5} h_N\left(\frac{i-j}{N}\right) h_N\left(\frac{i'-j'}{N}\right) = 2\sum_{i=1}^{N} \sum_{j\neq i} \sum_{\substack{i'\neq i \\ i'\neq j}} h_N\left(\frac{i-j}{N}\right) h_N(0) = -2(N-2)N\, h_N(0)^2.$$

$$\sum_{P_6} h_N\left(\frac{i-j}{N}\right) h_N\left(\frac{i'-j'}{N}\right) = 4\sum_{i=1}^{N} \sum_{j\neq i} \sum_{\substack{j'\neq j \\ j'\neq i}} h_N\left(\frac{i-j}{N}\right) h_N\left(\frac{i'-j'}{N}\right)$$

$$= 4\sum_{i=1}^{N} \sum_{j\neq i} \left(-h_N(0) - h_N\left(\frac{i-j}{N}\right)\right) h_N\left(\frac{i-j}{N}\right)$$

$$= 4N\, h_N(0)^2 + 4N\, h_N(0)^2 - 4N^2\, c_h^{(N)}$$

$$= 8N\, h_N(0)^2 - 4N^2\, c_h^{(N)}.$$

$$\sum_{P_7} h_N\left(\frac{i-j}{N}\right) h_N\left(\frac{i'-j'}{N}\right) = \sum_{i=1}^{N} \sum_{j\neq i} \sum_{\substack{i'\neq j \\ i'\neq i}} \sum_{\substack{j'\neq i' \\ j'\neq j \\ j'\neq i}} h_N\left(\frac{i-j}{N}\right) h_N\left(\frac{i'-j'}{N}\right)$$

$$= \sum_{i=1}^{N} \sum_{j\neq i} \sum_{\substack{i'\neq j \\ i'\neq i}} h_N\left(\frac{i-j}{N}\right)\left(-h_N(0) - h_N\left(\frac{i'-j}{N}\right) - h_N\left(\frac{i'-i}{N}\right)\right)$$

$$= \sum_{i=1}^{N} \sum_{j\neq i} h_N\left(\frac{i-j}{N}\right)\left(-(N-2)\, h_N(0) + 2h_N(0) + 2h_N\left(\frac{i-j}{N}\right)\right)$$

$$= +\, (N-4)\, N\, h_N(0)^2 + 2N(Nc_h^{(N)} - h_N(0)^2)$$

$$= 2N^2\, c_h^{(N)} + N(N-6)\, h_N(0)^2.$$

Collecting terms and multiplying by $Ez_i z_j z_{i'} z_{j'}$ we get

$$N^2 ET_N^2 = N\,h_N(0)^2\lambda - 4N\,h_N(0)^2\lambda\lambda_{-1} + N(N-1)\,h_N(0)^2\lambda\lambda_{-1}$$
$$+ 2N(Nc_h^{(N)} - h_N(0)^2)\lambda\lambda_{-1} - 2(N-2)N\,h_N(0)^2\lambda\lambda_{-1}\lambda_{-2}$$
$$+ (8N\,h_N(0)^2 - 4N^2 c_h^{(N)})\lambda\lambda_{-1}\lambda_{-2}$$
$$+ (2N^2 c_h^{(N)} + N(N-6)\,h_N(0)^2)\lambda\lambda_{-1}\lambda_{-2}\lambda_{-3}.$$

Since $h_N(\cdot)$ is a step function which satisfies (i) and (ii) above we obtain the relation

$$c_h^{(N)} = \frac{1}{N}\sum_{i=1}^{N} h_N\left(\frac{i}{N}\right)^2 = ||h_N||_{L_2}^2.$$

Upon dividing both sides of the above equation by N^2 and taking into account that $h_N(0) \to h(0)$, we obtain

(12.3) $$ET_N^2 = h_N(0)^2(\lambda^2-2\lambda^3+\lambda^4) + ||h_N||^2\,(2\lambda^2-4\lambda^3+2\lambda^4) + o_N(1)$$
$$= [h(0)\lambda(1-\lambda)]^2 + 2[||h||_{L_2}\lambda(1-\lambda)]^2 + o_N(1).$$

Also

(12.4) $$ET_N = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} h_N\left(\frac{i-j}{N}\right) Ez_i z_j = \frac{1}{N}\sum_{i=1}^{N} h_N(0)\,\frac{m}{N} - \frac{N}{N}\,h_N(0)\,\frac{m}{N}\,\frac{m-1}{N-1}$$
$$= h_N(0)\,\frac{m}{N}\left(1 - \frac{m-1}{N-1}\right) = h(0)\lambda(1-\lambda) + o_N(1).$$

Hence

(12.5) $$\operatorname{var} T_N = ET_N^2 - (ET_N)^2 = 2||h||_{L_2}^2\,\lambda^2(1-\lambda)^2 + o_N(1).$$

We now extend this result to the general case. Let $\bar{h}_N = \frac{1}{N}\sum_{i=1}^{N} h_N\left(\frac{i}{N}\right)$, then the sequence $h'_N(\cdot) = h_N(\cdot) - \bar{h}_N$ satisfies the assymptions of the particular case treated so far.

But $|\bar{h}_N| = |\frac{1}{N}\sum_{i=1}^{N} h_N(\frac{i}{N}) - \int_0^1 h(x)\,dx| = |\int_0^1 (h_N(x) - h(x))\,dx|$

$$\leq \int_0^1 |h_N(x) - h(x)|dx \to 0,$$

since on a finite measure space convergence in L_2 implies convergence in L_1. Hence, applying (12.3) we get for the general case

$$(12.6)\qquad ET_N^2 = [(h_N(0) - \bar{h}_N)\,\lambda(1-\lambda)]^2 + 2[\lambda(1-\lambda)||h_N - \bar{h}_N||^2 + o_N(1)$$

$$= [h_N(0)\,\lambda(1-\lambda)]^2 + o_N(1) + 2\lambda^2(1-\lambda)^2(||h_N||^2_{L_2} - \bar{h}_N^2) + o_N(1)$$

$$= [h(0)\,\lambda(1-\lambda)]^2 + 2[\lambda(1-\lambda)||h||^2_{L_2}]^2 + o_N(1).$$

Similarly we have, making use of (12.4),

$$(12.7)\qquad ET_N = (h_N(0) - \bar{h}_N)\,\frac{m}{N}\,(1 - \frac{m-1}{N-1}) = h(0)\,\lambda(1-\lambda) + o_N(1),$$

and thus

$$(12.8)\qquad \operatorname{var} T_N = 2[\lambda(1-\lambda)||h||_{L_2}]^2 + o_N(1).$$

<u>Proof of (b)</u>: Replacing h_N by $g_N - h_N$ and T_N by $T_{g_N} - T_{h_N}$ we may apply the result derived under (a). Hence from (12.6)

$$(12.9)\qquad E(T_{g_N} - T_{h_N})^2 = [(g_N(0) - h_N(0))\,\lambda(1-\lambda)]^2 + 2[\lambda(1-\lambda)||g_N - h_N||_{L_2}]^2 + o_N(1) = o_N(1).$$

This completes our proof.

If we define two sequences $\{g_N(\cdot)\}$ and $\{h_N(\cdot)\}$ satisfying $\{g_N - h_N\} \Rightarrow 0$ to be equivalent (the requirements of an equivalence relation are obviously satisfied), then equivalent sequences have identical limiting distributions, if a limiting distribution exists at all. We state this fact as a

Corollary: If $\{h_N\}$ and $\{h_N'\}$ are sequences converging to the same h, then T_{h_N} and $T_{h_N'}$ either both converge in distribution to the same limit or neither of them converges in distribution.

Proof: $\{h_N\} \Rightarrow h$ and $\{h_N'\} \Rightarrow h$ implies $\{h_N - h_N'\} \Rightarrow 0$ and hence by Theorem 12.1 (b), $T_{h_N} - T_{h_N'} \xrightarrow{pr.} 0$ which implies the desired result.

β. Asymptotic Distribution of T_N for h With Finite Fourier Expansion

Throughout this section we assume that $h(\cdot)$ has the Fourier expansion

$$(12.10) \qquad h(x) = \sum_{k=-K}^{K} d_k \, e^{2\pi ikx},$$

where K is arbitrary but finite. From our assumption $\int_0^1 h(x)\,dx = 0$ we get $d_0 = 0$.

Since $h(\cdot)$ is real we have $d_k = \bar{d}_{-k}$ for each k; but $h(\cdot)$ is also symmetric with respect to 0, hence

$$\sum_{k=-K}^{K} d_k \, e^{2\pi ikx} = h(x) = h(-x) = \sum_{k=-K}^{K} d_k \, e^{-2\pi ikx},$$

so that $d_k = d_{-k}$ (by uniqueness of expansion). It follows that d_k is real for all k. Combining the results we get

$$(12.11) \qquad d_0 = 0, \quad d_k = \bar{d}_k = d_{-k} \qquad k = \pm 1, \pm 2, \ldots, \pm K.$$

Matrix form of representation: By (11.4) T_N can be written in the form

$$(12.12) \qquad T_N = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} h_N\left(\frac{i-j}{N}\right) z_i z_j = \frac{1}{N} z_N' H_N z_N,$$

where $z_N = (z_1^{(N)}, z_2^{(N)}, \ldots, z_N^{(N)})'$ is the vector of indentically distributed, dependent random variables which are indicators of the X-sample. H_N is defined by

$$[H_N]_{r,s} = h_N\left(\frac{r-s}{N}\right) \qquad \begin{array}{l} r = 1, 2, \ldots, N. \\ s = 1, 2, \ldots, N. \end{array}$$

Diagonalization of H_N: H_N is a symmetric matrix (since $h_N(\cdot)$ is symmetric with respect to zero) which has the additional property that $[H_N]_{r,s}$ depends only on $(r - s)$ mod N. Matrices with this latter property are called "circulant" matrices.

G. Wahba (1967) has shown that a unitary matrix W_N, which diagonalizes circulant matrices of order N, is given by the symmetric matrix W_N defined by

$$(12.13) \qquad [W_N]_{r,s} = \frac{1}{\sqrt{N}} e^{2\pi irs/N}.$$

By W_N^* we denote the adjoint of W_N, which is also equal to $\overline{W}_N$. Hence we get the relation

$$(12.14) \qquad \frac{1}{N} H_N = W_N D_N W_N^*,$$

where D_N is a diagonal matrix. Since H_N is symmetric, the elements of D_N are real.

If we set

$$(12.15) \qquad \eta_N = W_N^* z_N = (\eta_1^{(N)}, \eta_2^{(N)}, \ldots, \eta_N^{(N)})'$$

we get

$$(12.16) \qquad T_N = \frac{1}{N} z_N' H_N z_N = z_N' W_N D_N W_N^* z_N = \overline{\eta}_N' D_N \eta_N = \sum_{\ell=1}^{N} d_\ell^{(N)} |\eta_\ell^{(N)}|^2$$

where $d_\ell^{(N)}$ are the diagonal elements of D_N $(\ell = 1, 2, \ldots, N)$. Another way of writing this is

$$(12.17) \qquad T_N = \sum_{\ell=1}^{N} d_\ell^{(N)} [\mathrm{Re}(\eta_\ell^{(N)})^2 + \mathrm{Im}(\eta_\ell^{(N)})^2].$$

We now determine $d_\ell^{(N)}$:

Since $W_N^{-1} = W_N^*$ we get the relation (from (12.14))

(12.18) $$D_N = \frac{1}{N} W_N^* H_N W_N,$$

$$[H_N W_N]_{r,s} = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} h_N(\tfrac{r-j}{N})\, e^{2\pi ijs/N} = \frac{1}{N} \sum_{j=1}^{N} h_N(\tfrac{j}{N})\, e^{2\pi i(r-j)s/N},$$

by periodicity and symmetry of $h_N(\cdot)$. Hence

$$\frac{1}{N}[W_N^* H_N W_N]_{t,s} = \frac{1}{N^2} \sum_{r=1}^{N} e^{-2\pi itr/N} e^{2\pi irs/N} \sum_{j=1}^{N} h_N(\tfrac{j}{N})\, e^{2\pi ijs/N}$$

(by symmetry of h_N)

$$= \begin{cases} \frac{1}{N} \sum_{j=1}^{N} h_N(\frac{j}{N})\, e^{2\pi is\frac{j}{N}} & \text{if } t = s, \\ 0 & \text{if } t \neq s, \end{cases}$$

using the orthogonality property of W_N.

It is easily recognized that

(12.19) $$d_\ell^{(N)} = \ell^{th} \text{ Fourier coefficient of } h_N(\cdot), \qquad \ell \leq N.$$

<u>Lemma 12.1</u>: If $\{h_N\} \Rightarrow h$, then $d_\ell^{(N)} \to d_\ell$ as $N \to \infty$, $\ell = 1, 2, \ldots$

<u>Proof</u>: If $(\cdot,\cdot)$ denotes the inner product in the Hilbert space $L_2[0,1]$, then

$$d_\ell = (h(x), e^{2\pi i\ell x}) \quad \text{and} \quad d_\ell^{(N)} = (h_N(x), e^{2\pi i\ell x}).$$

By the Schwarz inequality

$$|d_\ell - d_\ell^{(N)}|^2 = |(h(x) - h_N(x), e^{2\pi i\ell x})|^2 \leq ||h - h_N||_{L_2}^2 ||e^{2\pi i\ell x}||$$

$$\to 0 \quad \text{as } N \to \infty.$$

Equations (12.16) and (12.17) give a first indication as to what the limiting distribution of T_N might look like. The $\eta_\ell^{(N)}$'s are linear functions of the z_N's and can thus be expected to be asymptotically normal under quite general conditions. (We will come to this question later.

in this section.) The $d_\ell^{(N)}$ converge to known constants. This convergence, however, does not allow us to pass to the limit immediately, since it is not uniform in N. We shall see that in fact $d_\ell^{(N)} = d_{N-\ell}^{(N)}$. The next theorem will show how this obstacle can be overcome by exploiting more thoroughly the structure of W_N and by choosing a particularly suitable sequence $\{h_N\} \Rightarrow h$.

<u>Theorem 12.2</u>: Let $\{h_N\}$ be the sequence of step functions satisfying (i) and (ii), defined by $h_N(\frac{i}{N}) = h(\frac{i}{N})$, then $\{h_N\} \Rightarrow h$, and for $N > 2K$ we may write

$$T_{h_N} = 2 \sum_{\ell=1}^{K} d_\ell |\eta_\ell^{(N)}|^2 = 2\bar{\eta}_N D_N' \eta_N \tag{12.20}$$

where $\eta_N = W_N^* z_N$, $\eta_\ell^{(N)} = \ell^{th}$ component of η_N, $[D_N']_{k,\ell} = \delta_{k\ell} \cdot d_\ell$.

<u>Proof</u>: Since $h(\cdot)$ is uniformly continuous on $[0,1]$ it is obvious that $\{h_N\} \Rightarrow h$. By straightforward computation we obtain for the particular H_N with $[H_N]_{r,s} = h(\frac{r-s}{N})$:

$$[W_N^* H_N]_{r,s} = \frac{1}{\sqrt{N}} \sum_{t=1}^{N} e^{-2\pi irt/N} h(\frac{t-s}{N})$$

$$= \frac{1}{\sqrt{N}} \sum_{t=1}^{N} e^{-2\pi irt/N} \sum_{k=1}^{N} d_k \left(e^{2\pi ik(t-s)/N} + e^{-2\pi ik(t-s)/N}\right).$$

$$[W_N^* H_N W_N]_{r,u} = \frac{1}{N} \sum_{s=1}^{N} e^{2\pi ius/N} \sum_{t=1}^{N} e^{-2\pi irt/N} \sum_{k=1}^{N} d_k \left(e^{2\pi ik(t-s)/N} + e^{-2\pi ik(t-s)/N}\right) \tag{12.21}$$

$$= \frac{1}{N} \sum_{k=1}^{N} d_k \Big[\sum_{s=1}^{N} e^{2\pi i(u-k)s/N} \sum_{t=1}^{N} e^{-2\pi i(r-k)t/N} + \sum_{s=1}^{N} e^{2\pi i(u+k)s/N} \sum_{t=1}^{N} e^{-2\pi i(r+k)t/N} \Big]$$

$$= \frac{1}{N} \sum_{k=1}^{N} d_k(\delta(u,r,k) + \delta'(u,r,k)), \quad \text{say.}$$

Now

$$\delta(u,r,k) = \begin{cases} N^2 & \text{if } k = u \text{ and } k = r, \\ 0 & \text{otherwise.} \end{cases}$$

$$\delta'(u,r,k) = \begin{cases} N^2 & \text{if } k = N-u \text{ and } k = N-r, \\ 0 & \text{otherwise.} \end{cases}$$

Hence

$$\frac{1}{N} W_N^* H_N W_N = D_N,$$

where D_N is a diagonal matrix with diagonal elements $(d_1, d_2, \dots, d_k, 0, \dots, 0, d_k, \dots, d_2, d_1, 0)$.

From this result we get the representation

$$T_N = \frac{1}{N} z_N' H_N z_N = z_N' W_N D_N W_N^* z_N = \sum_{\ell=1}^{K} d_\ell(|\eta_\ell^{(N)}|^2 + |\eta_{N-\ell}^{(N)}|^2),$$

where $\eta_\ell^{(N)} = \ell^{th}$ component of η_N, $\eta_N = W_N^* z_N$.

Now $|\eta_\ell^{(N)}|^2 + |\eta_{N-\ell}^{(N)}|^2 = (\mathrm{Re}\eta_\ell^{(N)})^2 + (\mathrm{Im}\eta_\ell^{(N)})^2 + (\mathrm{Re}\eta_{N-\ell}^{(N)})^2 + (\mathrm{Im}\eta_{N-\ell}^{(N)})^2$

It is easy to see, that $\eta_\ell^{(N)}$ and $\eta_{N-\ell}^{(N)}$ are conjugate complex, and hence $|\eta_\ell^{(N)}|^2 + |\eta_{N-\ell}^{(N)}|^2 = 2|\eta_\ell^{(N)}|^2$. Hence we finally get

$$T_N = 2 \sum_{\ell=1}^{K} d_\ell |\eta_\ell^{(N)}|^2.$$

<u>Distribution of $\eta_\ell^{(N)}$</u>: From now on we assume that $N > 2K$, which is no real restriction, since we are interested in the limiting distribution of $\eta_\ell^{(N)}$ only.

<u>Theorem 12.3</u>: (a) For each $\ell < N/2$ we have

(12.22) $$E\eta_\ell^{(N)} = 0.$$

(12.23) $$\mathrm{var}\ \mathrm{Re}(\eta_\ell^{(N)}) = \mathrm{var}\ \mathrm{Im}(\eta_\ell^{(N)}) = \frac{1}{2}\frac{m}{N}\left(1 - \frac{m-1}{N-1}\right) \to \frac{1}{2}\lambda(1-\lambda), \text{ as } N \to \infty.$$

(12.24) $$\mathrm{cov}\ (\mathrm{Re}(\eta_\ell^{(N)}), \mathrm{Im}(\eta_\ell^{(N)})) = 0.$$

(b) If $0 < \lambda < 1$, then $\mathrm{Re}(\eta_\ell^{(N)})$ and $\mathrm{Im}(\eta_\ell^{(N)})$ are asymptotically normally distributed with means and variances given by (12.22), (12.23).

<u>Proof</u>: (a)

$$E\eta_\ell^{(N)} = \sum_{r=1}^{N} \frac{1}{\sqrt{N}} e^{-2\pi i \ell r/N} Ez_r = \frac{m}{N} \frac{1}{N} \sum_{r=1}^{N} e^{-2\pi i \ell r/N} = 0,$$

since the ℓ^{th} row of W_N^* is orthogonal to the N^{th}, which has the form $\frac{1}{\sqrt{N}}(1, 1, 1, \ldots, 1)$.

$$\mathrm{var}\ \mathrm{Re}(\eta_\ell^{(N)}) = E\ \mathrm{Re}(\eta_\ell^{(N)})^2$$

$$= \frac{1}{4N} E[\sum_{r=1}^{N} e^{2\pi i r\ell/N} z_r + \sum_{s=1}^{N} e^{-2\pi i s\ell/N} z_s]^2$$

$$= \frac{1}{4N} \sum_{r=1}^{N} \sum_{s \neq r} (e^{2\pi i r\ell/N} + e^{-2\pi i r\ell/N})(e^{2\pi i s\ell/N} + e^{-2\pi i s\ell/N})\lambda\lambda_{-1}$$

$$+ \frac{1}{4N} \sum_{r=1}^{N} (e^{2\pi i r\ell/N} + e^{-2\pi i r\ell/N})^2 \lambda$$

$$= \frac{1}{4N} \sum_{r=1}^{N} (e^{2\pi i r\ell/N} + e^{-2\pi i r\ell/N})^2 \lambda(1-\lambda_{-1})$$

$$= \frac{\lambda(1-\lambda_{-1})}{4N} \begin{cases} 0 + 2N + 0 & \text{if } 2\ell \neq N, \\ N + 2N + N & \text{if } 2\ell = N. \end{cases}$$

$$E|\eta_\ell^{(N)}|^2 = \frac{1}{N} E[\sum_{r=1}^{N} \sum_{s=1}^{N} e^{-2\pi i \ell r/N} e^{2\pi i \ell s/N} z_r z_s]$$

$$= \frac{1}{N} \sum_{r=1}^{N} \sum_{s \neq r} e^{-2\pi i \ell r/N} e^{2\pi i \ell s/N} \lambda\lambda_{-1} + \frac{1}{N} \sum_{r=1}^{N} e^0 \lambda$$

$$= \frac{1}{N}(-\lambda\lambda_{-1}N + \lambda N) = \lambda(1-\lambda_{-1}),$$

hence

$$\mathrm{var}\ \mathrm{Im}(\eta_\ell^{(N)}) = E|\eta_\ell^{(N)}|^2 - \mathrm{var}\ \mathrm{Re}(\eta_\ell^{(N)}) = \tfrac{1}{2}\lambda(1-\lambda_{-1}).$$

A similar straightforward computation shows that $\mathrm{Re}(\eta_\ell^{(N)})$ and $\mathrm{Im}(\eta_\ell^{(N)})$ are uncorrelated.

(b) We prove the result for $\alpha_N = \mathrm{Re}(\eta_\ell^{(N)})$, ℓ fixed. The proof for the imaginary part can be given in the same way.

Asymptotic normality follows easily from a very general result by Hájek (1961).

From his Theorems 4.1 and 4.2 we compile the following result:

Let $\{a_{Ni},\ i = 1, 2,\dots,N\}$ and $\{b_{Ni},\ i = 1, 2,\dots,N\}$ be double sequences of real numbers. Let $(R_{N1}, R_{N2},\dots,R_{NN})$ be a random vector which assumes the $N!$ permutations of $(1, 2,\dots,N)$ with equal probabilities. Set $\bar{a}_N = \frac{1}{N}\sum_{i=1}^{N} a_{Ni}$, $\bar{b}_N = \frac{1}{N}\sum_{i=1}^{N} b_{Ni}$.

Assume that

$$\lim_{N\to\infty} \frac{\max_{1\le i\le N} (a_{Ni} - \bar{a}_N)^2}{\sum_{i=1}^{N} (a_{Ni} - \bar{a}_N)^2} = 0, \tag{12.25}$$

$$\lim_{N\to\infty} \frac{\max_{1\le i\le N} (b_{Ni} - \bar{b}_N)^2}{\sum_{i=1}^{N} (b_{Ni} - \bar{b}_N)^2} = 0, \tag{12.26}$$

(12.27)[1]
$$\left[\lim_{N\to\infty} \frac{k_N}{N} = 0\right] \Longrightarrow \lim_{N\to\infty} \frac{\max_{1\le i_1 <\dots< i_{k_N}\le N} \sum_{\alpha=1}^{k_N} (a_{Ni_\alpha} - \bar{a}_N)^2}{\sum_{i=1}^{N} (a_{Ni} - \bar{a}_N)^2} = 0.$$

Under these conditions $\alpha_N = \sum_{i=1}^{N} b_{Ni} a_{NR_i}$ is asymptotically normally distributed with mean $E\alpha_N$ and variance $\sigma^2(\alpha_N)$.

In our case we have

$$\alpha_N = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \cos\,(2\pi j\ell/N)\, z_j. \tag{12.28}$$

---

[1] Here, of course, "$\Longrightarrow$" stands for logical implications.

If we set

$$(12.29) \qquad a_{Nj} = \begin{cases} 1 & \text{if } 1 \le j \le m_N, \\ 0 & \text{if } m_N + 1 \le j \le N, \end{cases}$$

and

$$(12.30) \qquad b_{Nj} = \frac{1}{\sqrt{N}} \cos (2\pi j \ell / N),$$

then it is easy to see that α_N defined by (12.28) has the form $\alpha_N = \sum_{j=1}^{N} b_{Nj} a_{NR_j}$.

Now we have to check conditions (12.25), (12.26), and (12.27).

$\max_j (a_{Nj} - \bar{a}_N)^2 \le 1$, $\sum_{j=1}^{N} (a_{Nj} - \bar{a}_N)^2 = m_N(1-\lambda_N)^2 + (N-m_N)\lambda^2_N = N\, \lambda_N(1-\lambda_N)$

$\doteq N\, \lambda(1-\lambda) \to \infty$, hence (12.25) is satisfied. $b^2_{Nj} \le \frac{1}{N}$ for all j, $(\bar{b}_N = 0)$

$\sum_{j=1}^{N} b^2_{Nj} = \frac{1}{N} \frac{1}{4} \sum_{j=1}^{N} (e^{2\pi i j \ell/N} + e^{-2\pi i j \ell/N})^2 = \frac{1}{4N} (0 + 2N + 0) = \frac{1}{2}$, this implies (12.26).

To check (12.27) let $\frac{k_N}{N} < \delta$ for $N \ge N_\delta$. Then $\sum_{\alpha=1}^{k_N} (a_{Nj_\alpha} - \bar{a}_N)^2 \le k_N < \delta N$ for $N \ge N_\delta$ and all indices $j_1 < j_2 < \ldots < j_{k_N}$. Since $\sum_{j=1}^{N} (a_{Nj} - \bar{a}_N)^2 \doteq N\, \lambda(1-\lambda)$ we get $\frac{\sum_{\alpha=1}^{k_N} (a_{Nj_\alpha} - \bar{a}_N)^2}{\sum_{j=1}^{N} (a_{Nj} - \bar{a}_N)^2} < \frac{\delta}{\lambda(1-\lambda)}$ for N large enough. This shows that (12.27) is satisfied and completes our proof.

We now extend the above result to any finite number of components $\eta_\ell^{(N)}$ $(\ell = 1, 2, \ldots, L)$, but before we state our theorem we prove a useful

<u>Lemma 12.2</u>: Let $w_\ell^{(N)} = \alpha_\ell^{(N)} + i\beta_\ell^{(N)}$ (α's and β's real) be the first L row vectors of W_N^*, (L fixed). Then any linear combination double sequence $\{b_{Nj}\}$ of the form $b_{Nj} = \sum_{\ell=1}^{L} c_\ell \alpha_{\ell j}^{(N)} + \sum_{\ell=1}^{L} d_\ell \beta_{\ell j}^{(N)}$ satisfies (12.26).

<u>Proof</u>: First note that for $N > 2L$ we have $\bar{\alpha}_\ell^{(N)} = \frac{1}{N} \sum_{j=1}^{N} \alpha_{\ell j}^{(N)} = 0$,

$\bar{\beta}_\ell^{(N)} = \frac{1}{N} \sum_{j=1}^{N} \beta_{\ell j}^{(N)} = 0$, $\sum_{j=1}^{N} \alpha_{\ell j}^{(N)} \alpha_{\ell' j}^{(N)} = \sum_{j=1}^{N} \beta_{\ell j}^{(N)} \beta_{\ell' j}^{(N)} = 0$

for $1 \le \ell < \ell' \le L$ and $\sum_{j=1}^{N} \alpha_{\ell j}^{(N)} \beta_{\ell' j}^{(N)} = 0$ for $1 \le \ell \le \ell' \le L$

(i.e., all the real and complex components of W_N^* are orthogonal if $L < N/2$). We give the proof for a weighted sum of two double sequences $c_1\alpha_{1j}^{(N)} + c_2\alpha_{2j}^{(N)}$. The general case can be proved in the same way.

Consider $c_1\alpha_{1j}^{(N)} + c_2\alpha_{2j}^{(N)}$, $c_1^2 + c_2^2 > 0$.

$$\sum_{j=1}^{N} (c_1\alpha_{1j}^{(N)} + c_2\alpha_{2j}^{(N)})^2 = c_1^2 \sum_{j=1}^{N} \alpha_{1j}^{(N)2} + c_2^2 \sum_{j=1}^{N} \alpha_{2j}^{(N)2} = \frac{c_1^2}{2} + \frac{c_2^2}{2} > 0$$

$$\max_j (c_1\alpha_{1j}^{(N)} + c_2\alpha_{2j}^{(N)})^2 \le \max_j c_1^2\alpha_{1j}^{(N)2} + \max_j c_2^2\alpha_{2j}^{(N)2} + 2|c_1c_2|\max_j \alpha_{1j}^{(N)}\alpha_{2j}^{(N)}$$

$\le \frac{1}{N} (|c_1| + |c_2|)^2 \to 0$ as $N \to \infty$. Hence (12.26) follows.

<u>Joint distribution of $\eta_\ell^{(N)}$, $\ell = 1, 2, \ldots, L$</u>:

<u>Theorem 12.4</u>: Let $0 < \lambda < 1$. For any constant L the joint distribution of $(\mathrm{Re}\eta_1^{(N)}, \mathrm{Im}\eta_1^{(N)}, \mathrm{Re}\eta_2^{(N)}, \mathrm{Im}\eta_2^{(N)}, \ldots, \mathrm{Re}\eta_L^{(N)}, \mathrm{Im}\eta_L^{(N)})$ is asymptotically normal with mean vector 0 and covariance matrix $\Sigma = \frac{1}{2} \lambda(1-\lambda)I_{2L}$, where I_{2L} is the identity matrix of order $2L$.

<u>Proof</u>: According to a well-known theorem by H. Cramér it suffices to prove asymptotic normality for any linear combination $\zeta = \sum_{\ell=1}^{L} c_\ell \mathrm{Re}\eta_\ell^{(N)} + \sum_{\ell=1}^{L} c_\ell' \mathrm{Im}\eta_\ell^{(N)}$. But because of Lemma 12.2 this follows directly from another application of Hájek's results. Computation of mean vector and covariance matrix is straightforward.

Asymptotic distribution of T_N:

Theorem 12.5: Let $0 < \lambda < 1$ and $\{h_N\} \to h$. Under the assumptions of the present section, i.e., if $h(x) = \sum_{k=-K}^{K} d_k e^{2\pi ikx}$, the asymptotic distribution of $T_N = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{m} h_N(\frac{R_i - R_j}{N})$ is the same as the distribution of $\lambda(1-\lambda) \sum_{k=1}^{K} d_k u_k$, where u_k are independent χ^2 random variables with 2 d.f.

Proof: It is easy to see that if $F_N(x_1,\ldots,x_{2L}) \to F(x_1,\ldots,x_{2L})$ in distribution, where F is absolutely continuous, then the distribution of any quadratic form $x'Ax$ converges to the distribution of the quadratic form of the limit (the sets $x'Ax \le t$ are ellipses, and the F-measure of its boundaries are zero). Hence the result follows immediately from Theorem 12.2, (12.23), Theorem 12.4, and the Corollary to Theorem 12.1.

γ. Extension of Results to a Class of Functions With Infinite Fourier Expansion

We now extend the results of the previous section to a class of functions h with infinitely many Fourier coefficients different from zero.

For a function $h(x) = \sum_{k=-K}^{K} d_k e^{2\pi ikx}$ we found the asymptotic distribution of T_N by first studying 2K-dimensional linear functions of the z_N's, passing to the limit $(N \to \infty)$, and then deriving the distribution of the quadratic form.

In the present situation it seems to be natural that a similar path could be followed if we study "infinite-dimensional" linear functions of the z_N's taking values in some space and then try to pass to the limit.

A separable, infinite dimensional, Hilbert space will suit our purpose, as will be seen in the following theorems. In Appendix β we present a number of definitions and results on measures and convergence of measures on Hilbert spaces. A capital letter A refers to theorems and definitions stated in the appendix.

Under our assumptions on h (real, symmetric with respect to zero) we still have $d_k = d_{-k} = \bar{d}_k$ in the expansion

$$h(x) = \sum_{-\infty}^{\infty} d_k e^{2\pi ikx}. \tag{12.31}$$

Throughout this section we make the following additional

<u>Assumption</u>:

$$\sum_{-\infty}^{\infty} |d_k| = 2 \sum_{1}^{\infty} |d_k| < \infty. \tag{12.32}$$

This assumption implies that $\sum_{-N}^{N} d_k \to h(0)$ as $N \to \infty$.

In this section it is convenient to use as approximating sequence the sequence $\{h_N\}$ satisfying (i) and (ii) and

$$h_N\left(\frac{\ell}{N}\right) = \sum_{k=-N}^{N} d_k e^{2\pi ik\ell/N}, \tag{12.33}$$

i.e., on the intervals of constancy h_N takes the values at $\frac{\ell}{N}$ of the N^{th} partial sum of the Fourier series.

<u>Lemma 3</u>: For the sequence $\{h_N\}$ satisfying (i) and (ii) defined by (12.33) we have $||h_N - h||_{L_2} \to 0$ and $h_N(0) \to h(0)$ as $N \to \infty$; i.e., $\{h_N\}$ is a legitimate approximation in the sense that $\{h_N\} \Rightarrow h$.

<u>Proof</u>: Since $\sum_{k=-N}^{N} d_k e^{2\pi ikx} \to h(x)$ uniformly in x ($h(\cdot)$ is continuous) we have $|h_N(x) - h^{(N)}(x)| < \epsilon$, all x, for $N \geq N_\epsilon$, where $h^{(N)}(x)$ is the approximating step function satisfying (i), (ii) and $h^{(N)}(\frac{i}{N}) = h(\frac{i}{N})$. Hence $||h_N - h^{(N)}||_{L_2} \to 0$ as $N \to \infty$. By uniform continuity of h we have $||h^{(N)} - h||_{L_2} \to 0$, and hence $||h_N - h||_{L_2} \to 0$ as $N \to \infty$.

We have already remarked that $h_N(0) \to h(0)$ because of Assumption (12.32).

We now define a sequence of measures on the real Hilbert space H of real sequences with finite sum of squares.

As always in probability theory we assume that there exists an underlying probability space $(\Omega, \mathcal{A}, P)$ where the sequences of i.i.d. random variables $X_1, X_2, \dots$ ad inf. and $Y_1, Y_2, \dots$ ad inf. (with continuous distribution function) are defined. We also assume that to each N there is defined an integer $m_N \leq N$, such that $\lambda_N = \frac{m_N}{N} \to \lambda$ as $N \to \infty$, and $0 < \lambda < 1$.

Since we are only interested in certain rank order statistics, all our random variables are functions of the vectors z_N defined by (11.3). But it should be kept in mind that they are actually measurable functions on $(\Omega, \mathcal{A}, P)$.

Any measurable transformation from the space of the z_N's to H induces a probability measure on H. We now define a sequence of measurable transformations S_N of the z_N's and hence a sequence of measures on H:

Let V_N be the $N \times 2N$ matrix defined by

$$(12.34) \qquad [V_N]_{r,s} = \begin{cases} \operatorname{Re} [W_N]_{r,\frac{s+1}{2}} & \text{if } s \text{ is odd,} \\ \operatorname{Im} [W_N]_{r,\frac{s}{2}} & \text{if } s \text{ is even.} \end{cases}$$

$$\text{Set } c_k^{(2N)} = \begin{cases} |d_\ell| + |d_{N-\ell}| & \text{where } \ell = [\frac{k+1}{2}], \quad 1 \leq k \leq 2[\frac{N}{2}], \\ 0 & \text{for } 2[\frac{N}{2}] < k \leq 2N-2, \\ |d_N| & \text{for } k = 2N-1 \text{ and for } k = 2N. \end{cases}$$

Define $\overline{C}_{2N}$ to be the diagonal matrix with elements $c_k^{(2N)}$ $(k = 1, 2, \dots, 2N)$. E.g., for N even $\overline{C}_{2N}$ has diagonal elements $(|d_1| + |d_{N-1}|, |d_1| + |d_{N-1}|, |d_2| + |d_{N-2}|, \dots, |d_{N/2}| + |d_{N/2}|, 0, \dots, 0, |d_N|, |d_N|)$.

If $x = (x_1, x_2, \dots)$ is a generic element of the Hilbert space H, we define the sequence $\{S_N\}$ of mappings from Ω to H by the relation

(12.36) $\qquad \omega \to z_N(\omega) \to x(\omega), \qquad \text{with} \begin{cases}(x_1, x_2, \ldots, x_{2N})' = \overline{C}_N V_N' z_N \\ x_i = 0 \quad \text{for} \quad i > 2N.\end{cases}$

Let μ_N be the probability measure on H induced by S_N.

<u>Theorem 12.6</u>: The sequence $\{\mu_N\}$ of probability measures is compact.

<u>Proof</u>: We use Theorem A.2 for the proof.

Condition 1: This is obviously satisfied.

Condition 2: As a basis for H we take the vectors $(1,0,0,\ldots)$, $(0,1,0,0,\ldots),\ldots$. Now let $\epsilon > 0$ be given. Take M_ϵ, even, such that $\sum_{k=\frac{1}{2}M_\epsilon}^{\infty} |d_k| < \frac{\epsilon}{4}$, which is possible by (12.32), and let $M > M_\epsilon$ be arbitrary. Then for $N < M$ we have

$$\int_H r_M^2(x)\, d\mu_N(x) \le \begin{cases} 0 & \text{if } 2N < M, \\ 2|d_N| \frac{1}{2} \lambda(1-\lambda_{-1}) < \epsilon & \text{for } M \le 2N < 2M, \text{ by (12.23)} \end{cases}$$

For $N \ge M$ we set $\xi_{2N} = V_N' z_N$ (this is the vector of real and imaginary parts of η_N) and use the fact that all the components $\xi_\ell^{(2N)}$ have the same expected square $\frac{1}{2}\lambda(1-\lambda_{-1})$ (by (12.23)). Hence

$$\int_H r_M^2(x)\, d\mu_N(x) = \sum_{k=M}^{2N} (c_k^{(2N)})^2 E(\xi_k^{(2N)})^2$$

$$= \frac{1}{2}\lambda(1-\lambda_{-1}) \sum_{k=M}^{2N} (c_k^{(2N)})^2$$

$$\le \frac{1}{2}\lambda(1-\lambda_{-1}) \Big[2 \sum_{\ell=[\frac{M+1}{2}]}^{[\frac{N+1}{2}]} (|d_\ell| + |d_{N-\ell}|) + 2|d_N|\Big]$$

$$\le \lambda(1-\lambda) \Big[\sum_{\ell=[\frac{M}{2}]}^{\infty} |d_\ell| + \sum_{\ell=[\frac{M}{2}]}^{\infty} |d_\ell| + |d_N|\Big] < \epsilon.$$

Hence for $M > M_\epsilon$ we have $\sup_N \int_H r_M^2(x)\, d\mu_N(x) \le \epsilon$, which completes our proof.

Lemma 12.4: Let $\{\mu_N\}$ be a compact sequence, let M be the closure (in the topology of weak convergence) of $\{\mu_N, \; N = 1, 2, \ldots \})$. Then the set of characteristic functionals $\{\chi(f,\mu):\mu \epsilon M\}$ is a uniformly equi-continuous set of functions of f (on H). (For the definition of $\chi(f,\mu)$ see Appendix (A16).)

Proof: Let $\epsilon > 0$ be given. By Theorem A.1 there exists a compact K_ϵ such that $\mu_N(K_\epsilon) \geq 1 - \frac{\epsilon}{4}$, all N. By Lemma A.1 any $\mu \epsilon M$ also satisfies this relation, since it is the limit of a suitably chosen subsequence μ_{N_i}. Now K_ϵ is bounded by some constant $K > 0$. Let $||f - g|| < \frac{\epsilon}{2K}$. Then for any $\mu \epsilon M$ $|\chi(f,\mu) - \chi(g,\mu)| = |\int_H (e^{i(f,x)} - e^{i(g,x)})\, d\mu(x)| \leq \int_{K_\epsilon} |e^{i(f,x)}| \; |1-e^{i(g-f,x)}| \, d\mu(x) + 2\int_{H-K_\epsilon} d\mu(x)$ $\leq \int_{K_\epsilon} |(g-f,x)| \, d\mu(x) + \frac{\epsilon}{2} \leq \int_{K_\epsilon} ||f-g|| \; ||x|| \, d\mu(x) + \frac{\epsilon}{2} \leq K\frac{\epsilon}{2K} + \frac{\epsilon}{2} = \epsilon.$

Theorem 12.7: $\mu_N \to$ some probability measure μ as $N \to \infty$.

Proof: If $\mu_N \to \mu$ in the sense of weak convergence, then $\mu_N(R) = 1 = \int 1 d\mu_N \to \int 1 d\mu = \mu(R)$, thus μ has to be a probability measure. We show weak convergence by using Theorem A.3.

Let $f^{(k)} = (f_1, f_2, \ldots, f_k, 0, 0, \ldots)$; then by a slight extension of Theorem 12.4 we get

$$\chi(f^{(k)},\mu_N) \to \exp \{-\frac{1}{2}\frac{\lambda(1-\lambda)}{2} \sum_{i=1}^{k} |d_{[\frac{i+1}{2}]}| f_i^2\} \quad \text{as } N \to \infty. \tag{12.37}$$

Let μ be any limit measure of a suitably chosen subsequence. Then by the definition of weak convergence

$$\chi(f^{(k)},\mu) = \exp \{-\frac{1}{2}\frac{\lambda(1-\lambda)}{2} \sum_{i=1}^{k} |d_{[\frac{i+1}{2}]}| f_i^2\}. \tag{12.38}$$

Since the $f^{(k)}$'s are dense in H, and since the left hand side of (12.38) is continuous in its first argument we must have:

$$\chi(f,\mu) = \exp\left\{-\frac{1}{2}\frac{\lambda(1-\lambda)}{2}\sum_{i=1}^{\infty}\left|d_{[\frac{i+1}{2}]}\right| f_i^2\right\} \quad \text{all } f \in H. \tag{12.39}$$

(I.e., any limiting measure is normal with S-operator of the form

$$[S]_{r,s} = \frac{\lambda(1-\lambda)}{2}\,\delta_{r,s}\,\left|d_{[\frac{r+1}{2}]}\right|.)$$

Now let $f \in H$ and $\epsilon > 0$ be given. By the preceding Lemma 12.4 we know that $|\chi(f,\bar{\mu}) - \chi(f^{(k)},\mu_N)| < \frac{\epsilon}{3}$ uniformly in $\bar{\mu} \in M$, if only $||f - f^{(k)}|| \le \delta_\epsilon$ for suitably chosen δ_ϵ. Fix such an $f^{(k)}$ and let N_ϵ be such that $|\chi(f^{(k)},\mu_N) - \chi(f^{(k)},\mu)| < \frac{\epsilon}{3}$ for $N \ge N_\epsilon$, and for the particular μ appearing in (12.38) and (12.39). Then for $N \ge N_\epsilon$ we get

$$\begin{aligned}
&|\chi(f,\mu_N) - \chi(f,\mu)| \\
&\quad \le |\chi(f,\mu_N) - \chi(f^{(k)},\mu_N)| + |\chi(f^{(k)},\mu_N) - \chi(f^{(k)},\mu)| \\
&\qquad\qquad + |\chi(f^{(k)},\mu) - \chi(f,\mu)| \\
&\quad \le \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.
\end{aligned}$$

By Theorem A.3 we get the desired result.

Theorem 12.8: Under the assumptions of this section T_N converges in distribution to the distribution with characteristic function

$$\varphi_T(t) = \prod_{k=1}^{\infty}(1 - 2i\lambda(1-\lambda)\,d_k t)^{-1}. \tag{12.40}$$

Proof:

$$\text{Let } \delta_k = \begin{cases} +1, & \text{if } d_k \ge 0, \\ -1, & \text{if } d_k < 0. \end{cases}$$

If $\eta_N = W_N^* z_N$ as previously, then for the components $\eta_\ell^{(N)}$ we have $\eta_{N-\ell}^{(N)} = \bar{\eta}_\ell^{(N)}$. With this notation a straightforward computation shows that

$$(12.41)\qquad T_N = \frac{1}{N} z_N' H_N z_N = z_N' W_N D_N W_N^* z_N = \bar{\eta}_N' D_N \eta_N$$

$$= \sum_{k=1}^{N} d_k^{(N)} |(\eta_N)|^2 = \sum_{k=1}^{N} d_k^{(N)} [Re\eta_k^{(N)})^2 + (Im\eta_k^{(N)})^2]$$

$$= \sum_{k=1}^{N} \delta_k (c_{2k}^{(2N)})^2 [(\xi_{2k-1}^{(2N)})^2 + (\xi_{2k}^{(2N)})^2]$$

where $d_k^{(N)}$ are the characteristic values of H_N obtained by diagonalizing H_N by the matrix W_N.

Combining (12.36) and (12.41) it is easy to see that $T_N(\omega)$ is a function of $S_N(\omega)$ and that on H T_N has the simple structure

$$(12.42)\qquad T_N = \sum_{k=1}^{N} \delta_{[\frac{k+1}{2}]} x_k^2 + \delta_N(x_{2N-1}^2 + x_{2N}^2) = \sum_{k=1}^{\infty} \delta_{[\frac{k+1}{2}]} x_k^2 \quad \text{a.e. } (\mu_N),$$

since $x_{N+1} = \ldots = x_{2N-2} = x_{2N+1} = \ldots = 0$ a.e. μ_N.

Now we know from weak convergence that

$$(12.43)\qquad \varphi_{T_N}(t) = E_{\mu_N} \exp \{it \sum_{k=1}^{N} \delta_{[\frac{k+1}{2}]} x_k^2\} = E_{\mu_N} \exp \{it \sum_{k=1}^{\infty} \delta_{[\frac{k+1}{2}]} x_k^2\}$$

$$\rightarrow E_\mu \exp \{it \sum_{1}^{\infty} \delta_{[\frac{k+1}{2}]} x_k^2\} = \varphi_T(t) \quad \text{as } N \rightarrow \infty.$$

This limit is continuous and hence T_N converges in distribution. We now evaluate $E_\mu \exp \{it \sum_{k=1}^{\infty} \delta_{[\frac{k+1}{2}]} x_k^2\}$. From Theorem 12.7 we conclude that for each K

$$(12.44)\qquad E_{\mu_N} \exp \{it \sum_{k=1}^{2K} \delta_{[\frac{k+1}{2}]} x_k^2\} \rightarrow \prod_{k=1}^{K} (1 - 2i\lambda(1-\lambda) d_k t)^{-1} = E_\mu \exp\{it \sum_{k=1}^{2K} \delta_{[\frac{k+1}{2}]} x_k^2\}.$$

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

$$\varphi_T(t) = E_\mu \exp \{it \sum_{k=1}^{\infty} \delta_{[\frac{k+1}{2}]} x_k^2\} = \prod_{k=1}^{\infty} (1 - 2i\lambda(1-\lambda)\, d_k t)^{-1}.$$

<u>Theorem 12.9</u>: Let h have a continuous derivative. Then T_N converges in distribution to a probability measure with characteristic function

$$\varphi(t) = \prod_{k=1}^{\infty} (1 - 2i\lambda(1-\lambda)\, d_k t)^{-1},$$

where $d_k = \int_0^1 e^{2\pi ikx} h(x)\, dx, \qquad k = 1, 2, \ldots$

<u>Proof</u>: It is a well-known fact in Fourier-analysis that under the above conditions $|d_k| \le \frac{M}{k^2}$ for some constant M. Hence the sequence of Fourier coefficients converges absolutely and the result follows from Theorem 12.8.

13. The Asymptotic Distribution of T_N under Alternatives.

So far we have only obtained the asymptotic distribution of T_N defined by (11.2) in the case $F \equiv G$. We now derive, under somewhat more restticive conditions, the limiting distribution of T_N in cases where $F \not\equiv G$. The ideas of the proof have been developed by Chernoff and Savage (1958) and have since been applied to several problems of a similar nature.

It will be convenient to define a few terms which will appear repeatedly in the proof of Theorem 13.1. As before $h(\cdot)$ is a function defined on $[-1,1]$, even, periodic with period 1, and $\int_0^1 h(x)\,dx = 0$.

Definitions:

(13.1) $$H(x) = \lambda F(x) + (1-\lambda) G(x) \qquad 0 \le x \le 2\pi.$$

(13.2) $$B(x) = \int_0^{2\pi} h(H(x) - H(y))\, dF(y) \qquad 0 \le x \le 2\pi.$$

(13.3) $$B^*(x) = -\int_0^x \int_0^{2\pi} h'(H(u) - H(y))\, dF(y)\, dF(u) \qquad 0 \le x \le 2\pi.$$

(13.4) $$B^{**}(x) = B(x) + \lambda B^*(x).$$

Let $X_1,\ldots,X_m$ and $Y_1,\ldots,Y_n$ be independent samples of circular random variables with continuous densities $f(\cdot) > 0$, $g(\cdot) > 0$, respectively. In this section it is more convenient to use a different normalizing constant for the test statistic T_N and we thus define

(13.5) $$T_N = \frac{1}{N^2} \sum_{i=1}^{m} \sum_{j=1}^{m} h\left(\frac{R_i - R_j}{N}\right).$$

Let μ_N and σ_N^2 be defined by

(13.6) $$\mu_N = \lambda^2 EB(X)$$

and

(13.7) $$N\sigma_N^2 = 4\lambda^2 \left[\operatorname{var} B^{**}(X) + \frac{\lambda}{\lambda-1} \operatorname{var} B^*(Y)\right].$$

With these definitions we get the following

Theorem 13.1: If $\sigma_N \neq 0$, if $h''(\cdot)$ exists and is continuous, and if $\lambda_N = \frac{m}{N}$ satisfies $\lambda_0 \leq \lambda_N \leq 1-\lambda_0$ for some $\lambda_0 > 0$, and $\lambda_N \to \lambda$ as $N \to \infty$, then

$$(13.8) \qquad \lim_{N \to \infty} P\left(\frac{T_N - \mu_N}{\sigma_N} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt.$$

Proof: Let again $F_m(\cdot)$, $G_n(\cdot)$ be the empirical c.d.f. of the X- and the Y-sample, respectively. Set $H_N = \lambda F_m + (1-\lambda)G_n$ and $H = \lambda F + (1-\lambda)G$. Then using the representation (11.7) with the new normalizing constant we obtain

$$(13.9) \qquad T_N = \left(\frac{m}{N}\right)^2 \int_0^{2\pi} \int_0^{2\pi} h(H_N(x) - H_N(y))\, dF_m(x)\, dF_m(y).$$

The idea of the proof is to show that T_N is the sum of a constant term, an asymptotically normal term, and a term $o_p(N^{-\frac{1}{2}})$.

Let

$$(13.10) \qquad B_N = \left(\frac{N}{m}\right)^2 T_N = \int_0^{2\pi} \int_0^{2\pi} h(H_N(x) - H_N(y))\, dF_m(x)\, dF_m(y).$$

By Taylor's formula we have for any x and any Δ

$$(13.11) \qquad h(x + \Delta) = h(x) + \Delta h'(x) + \tfrac{1}{2}\Delta^2 h''(x + \theta\Delta), \quad \text{where } 0 \leq \theta \leq 1.$$

If we apply this expansion to the B_N-term with

$$(13.12) \qquad \Delta = H_N(x) - H(x) - (H_N(y) - H(y))$$

we obtain

$$(13.13) \qquad B_N = C_N + D_N + E_N, \quad \text{where}$$

$$(13.14) \qquad C_N = \int_0^{2\pi} \int_0^{2\pi} h(H(x) - H(y))\, dF_m(x)\, dF_m(y);$$

$$(13.15) \qquad D_N = \int_0^{2\pi} \int_0^{2\pi} h'(H(x) - H(y))\, (H_N(x) - H(x) - H_N(y) + H(y))\, dF_m(x)\, dF_m(y)$$

$$(13.16) \qquad E_N = \tfrac{1}{2} \int_0^{2\pi} \int_0^{2\pi} h''(\theta(H_N(x) - H_N(y)) + (1-\theta)(H(x) - H(y)) \\ (H_N(x) - H(x) - H_N(y) + H(y))^2\, dF_m(x)\, dF_m(y).$$

From the linearity of the Rieman-Stieltjes integral as a function of the measure (or rather: signed measure) we obtain formally

$$(13.17)\qquad dF_m(x)\, dF_m(y) = d(F_m(x) - F(x))\, d(F_m(y) - F(y)) + dF(x)\, d(F_m(y) - F(y)) + d(F_m(x) - F(x))\, dF(y) + dF(x)\, dF(y).$$

We now investigate the C_N-term.

Using (13.17) we obtain

$$(13.18)\qquad C_N = C_{N1} + C_{N2} + C_{N3} + C_{N4}$$

where

$$(13.19)\qquad C_{N1} = \int_0^{2\pi}\int_0^{2\pi} h(H(x) - H(y))\, dF(x)\, dF(y) = EB(X), \text{ a non-random term;}$$

$$(13.20)\qquad C_{N2} = \int_0^{2\pi} [\int_0^{2\pi} h(H(x) - H(y))\, dF(x)]\, d(F_m(y) - F(y))$$

$$= \frac{1}{m}\sum_{i=1}^{m} [B(X_i) - EB(X)], \quad \text{(note that } h(u) = h(-u));$$

$$(13.21)\qquad C_{N3} = \int_0^{2\pi}\int_0^{2\pi} h(H(x) - H(y))\, d(F_m(x) - F(x))\, dF(y)$$

$$= \int_0^{2\pi} [\int_0^{2\pi} h(H(x) - H(y))\, dF(y)]\, d(F_m(x) - F(x)) \quad \text{(Fubini's theorem)}$$

$$= \frac{1}{m}\sum_{i=1}^{m} [B(X_i) - EB(X)];$$

$$(13.22)\qquad C_{N4} = \int_0^{2\pi}\int_0^{2\pi} h(H(x) - H(y))\, d(F_m(x) - F(x))\, d(F_m(y) - F(y))$$

$$= o_p(N^{-\frac{1}{2}}) \qquad \text{by Lemma A.4 of the appendix,}$$

since $\frac{\partial^2 h(x,y)}{\partial y \partial x} = h''(H(x) - H(y))\, (\lambda f(x) + (1-\lambda)g(x))\, (\lambda f(y) + (1-\lambda)g(y))$ in this case.

Now we analyze the D_N-term similarly. Using (13.17) again we get the decomposition

$$(13.23)\qquad D_N = D_{N1} + D_{N2} + D_{N3} + D_{N4}$$

where D_{Ni} is obtained from C_{Ni} by replacing $h(H(x) - H(y))$ by $h'(H(x) - H(y))\, (H_N(x) - H(x) - H_N(y) + H(y))$ $(i = 1, 2, 3, 4)$.

(13.24) $$D_{N1} = D_{N11} + D_{N12} + D_{N13} + D_{N14}$$

where

(13.25) $$D_{N11} = \lambda_N \int_0^{2\pi}\int_0^{2\pi} h'(H(x) - H(y))\,(F_m(x) - F(x))\,dF(x)\,dF(y)$$
$$= \frac{\lambda_N}{m} \sum_{i=1}^{m} (B^*(X_i) - EB^*(X)),$$

because of

$$\frac{1}{m}\sum_{i=1}^{m}(B^*(X_i) - E^*(X)) = -\int_0^{2\pi}\int_0^{x}\int_0^{2\pi} h'(H(u) - H(y))\,dF(y)\,dF(u)\,d(F_m(x) - F(x))$$

$$= -\,[(F_m(x) - F(x))\int_0^{x}\int_0^{2\pi} h'(H(u) - H(y))\,dF(y)\,dF(u)]_{x=0}^{x=2\pi}$$

$$+\int_0^{2\pi}(F_m(x) - F(x))\int_0^{2\pi} h'(H(x) - H(y))\,dF(y)\,dF(x)$$

by partial integration (F has a density)

$$= 0 + \int_0^{2\pi}[\int_0^{2\pi} h'(H(x) - H(y))\,dF(y)]\,(F_m(x) - F(x))\,dF(x)$$

$$= \lambda_N^{-1} D_{N11}$$ by Fubini's theorem,

and where

(13.26) $$D_{N12} = (1-\lambda_N)\int_0^{2\pi}\int_0^{2\pi} h'(H(x) - H(y))\,(G_n(x) - G(x))\,dF(x)\,dF(y)$$
$$= \frac{(1-\lambda_N)}{n}\sum_{j=1}^{n}(B^*(Y_j) - EB^*(Y)),$$ by the same argument,

(13.27) $$D_{N13} = -\,\lambda_N\int_0^{2\pi}\int_0^{2\pi} h'(H(x) - H(y))\,(F_m(y) - F(y))\,dF(x)\,dF(y)$$
$$= D_{N11}$$ by skew-symmetry of h',

(13.28) $$D_{N14} = -\,(1-\lambda_N)\int_0^{2\pi}\int_0^{2\pi} h'(H(x) - H(y))\,(G_n(y) - G(y))\,dF(x)\,dF(y)$$
$$= D_{N12}$$ by the same argument.

(13.29) $$D_{N2} = \int_0^{2\pi}\int_0^{2\pi} h'(H(x) - H(y))\,(H_N(x) - H(x) - H_N(y) + H(y))\,dF(x)\,d(F_m(y) - F(y))$$

$$= o_p(N^{-\frac{1}{2}}) \text{ by Lemma A.3, parts a, b, c, d.}$$

(13.30) $$D_{N3} = \int_0^{2\pi}\int_0^{2\pi} h'(H(x) - H(y))\,(H_N(x) - H(x) - H_N(y) + H(y))\,dF(y)\,d(F_m(x) - F(x))$$

$$= o_p(N^{-\frac{1}{2}}) \text{ by the same argument.}$$

(13.31) $$D_{N4} = \int_0^{2\pi}\int_0^{2\pi} h'(H(x) - H(y))\,(H_N(x) - H(x) - H_N(y) + H(y))\,d(F_m(x) - F(x))\,d(F_m(y) - F(y))$$

$$= o_p(N^{-\frac{1}{2}}) \text{ by a straightforward extension of Lemma A.3.}$$

(13.32) $E_N = o_p(N^{-\frac{1}{2}})$; this follows easily from Lemma A.2 by multiplying out the square and using (13.17).

Collecting terms we get

(13.33) $$B_N = EB(X) + \frac{2}{m}\sum_{i=1}^{m}[B(X_i) - EB(X)] + \frac{2\lambda_N}{m}\sum_{i=1}^{m}[B^*(X_i) - EB^*(X_i)] + \frac{2(1-\lambda_N)}{n}\sum_{j=1}^{n}[B^*(Y_j) - EB^*(Y)] + o_p(N^{-\frac{1}{2}}),$$

and hence by (13.10, (13.2), (13.3), and (13.4)

(13.34) $$T_N = \lambda_N^2\,[EB(X) + \frac{2}{\lambda_N N}\frac{1}{m}\sum_{i=1}^{m}[B^{**}(X_i) - EB^{**}(X)] + \frac{2}{(1-\lambda_N)N}\frac{1}{n}\sum_{j=1}^{n}[B^*(Y_j) - EB^*(Y)] + o_p(N^{-\frac{1}{2}})].$$

Thus, because of the Central Limit Theorem, $N^{\frac{1}{2}}(T_N - \lambda^2 EB(X))$ has asymptotically a normal distribution $N(\mu,\sigma^2)$ with $\mu = 0$ and $\sigma^2 = 4\lambda^4(\frac{1}{\lambda}\,\sigma^2(B^{**}(X)) + \frac{1}{1-\lambda}\,\sigma^2 B^*(Y))$, since $B^*(Y)$ and $B^{**}(X)$ are bounded random variables (and hence they have finite variances). This completes our proof.

It is interesting to see why this theorem does not give us any information about the asymptotic distribution of T_N in the null case $(F(\cdot) = G(\cdot))$. In this case we get

$$(13.35) \qquad B(x) = \int_0^{2\pi} h(H(x) - H(y))\, dF(y) = \int_0^{2\pi} h(F(x) - F(y))\, dF(y)$$

$$= \int_0^1 h(F(x) - y)\, dy = \text{const.} \quad \text{(by periodicity of } h\text{)}.$$

$$(13.36) \qquad B^*(x) = - \int_0^x \int_0^{2\pi} h'(H(u) - H(y))\, dF(y)\, dF(u)$$

$$= \int_0^x [h(F(u) - y)]_{y=0}^{y=1}\, dF(u) = 0 \quad \text{(again by periodicity of } h\text{)}.$$

Hence it follows directly from the proof of Theorem 13.1 that T_N converges in distribution to the constant

$$(13.37) \qquad EB(X) = \int_0^{2\pi} \int_0^{2\pi} h(F(x) - F(y))\, dF(y)\, dF(x)$$

$$= \int_0^1 \int_0^1 h(x - y)\, dx\, dy = \int_0^1 h(x)\, dx = 0.$$

14. Efficiency and Consistency of Tests based on T_N.

In this section we derive the asymptotic relative efficiency of tests based on sequences $\{T_N\}$ for testing shift alternatives. We then show that under quite general conditions there exist tests in this class whose efficiency is arbitrarily close to 1. This result is somewhat surprising since we are only considering tests which are invariant under a relatively large group of transformations.

Again let $h(\cdot)$ be symmetric with respect to 0, periodic with period 1 and such that $\int_0^1 h(x)\,dx = 0$. Furthermore we assume that $h''(\cdot)$ and $f'(\cdot)$ exist and are continuous. In this section we restrict our attention to test statistics of the form

$$(14.1) \qquad T_N = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} h\left(\frac{i-j}{N}\right) z_i z_j ,$$

since more general test statistics of the form (11.2) have the same limiting distribution by the results of Section 12, α.

We use Bahadur's concept of efficiency again and thus we have to check Assumptions 1-4 of Section 3. Since $h''(\cdot)$ exists we know from Section 12 that (3.1) is satisfied. It follows from Theorem 13.1 that (3.3) is satisfied with $b_1(N) = N$ and

$$(14.2) \qquad c_1(\theta) = \lambda^2 \int_0^{2\pi} \int_0^{2\pi} h(H_\theta(x) - H_\theta(y))\,dF(y)\,dF(x),$$

where

$$H_\theta(x) = \lambda F(x) + (1-\lambda)G(x) \quad \text{and} \quad G(x) = \int_0^x f(t-\theta)dt = F(x-\theta) - F(-\theta).$$

In the following lemma we will show that Assumption 2 is satisfied with $t_1 = 1$. We have seen in Section 9 that the test with highest efficacy has an asymptotically normal distribution, so that it satisfies (3.1) with $a_2 = 1$, $t_2 = 2$; it also satisfies (3.3) with $b_2(N) = N^{\frac{1}{2}}$.

Hence the relation (3.4) holds, and all we have to show is that (3.2) holds for the distribution corresponding to (12.40).

<u>Lemma 14.1</u>: If $f(\cdot) > 0$ has a continuous derivative the distribution function corresponding to the characteristic function (12.40) satisfies (3.2) with $t = 1$ and $a = \frac{1}{\lambda(1-\lambda) \max d_k}$, provided that $d_k \geq 0$ for all k.

<u>Proof</u>: This result follows immediately from a theorem obtained by Box (1954). In his Theorem 2.4 he shows that for a finite weighted sum of independent χ_2^2-variables the distribution function $G(x)$ can be expressed in the form

$$(14.3) \qquad 1 - G(x) = P\left(\sum_{k=1}^{K} d_k \chi_2^2 > x\right) = \sum_{k=1}^{K} \alpha_k^{(K)} \exp\left(-\frac{x}{2d_k}\right),$$

where $\alpha_k^{(K)}$ are given explicitely. From the explicit representation it follows easily that the $\alpha_k^{(K)}$ converge as $K \to \infty$, provided that $\sum_{k=1}^{\infty} |d_k| < \infty$. From (14.3) it is obvious that the tail of $G(x)$ goes to 1 at a rate which corresponds to the "slowest" part on the r.h.s., i.e., $-2 \log (1 - G(x)) = \frac{x}{\max d_k}(1 + o(1))$. Since $d_k \to 0$ as $k \to \infty$ we have $\max_{1 \leq k \leq K} d_k = \max_{1 \leq k < \infty} d_k$ for K large enough. Since in (12.40) the weights are $\lambda(1-\lambda) d_k$, we obtain

$$(14.4) \qquad -2 \log (1 - F(x)) = \frac{x}{\lambda(1-\lambda) \max d_k}(1 + o(1)),$$

i.e., (3.2) is satisfied with $t = 1$ and $a = \frac{1}{\lambda(1-\lambda) \max d_k}$.

In order to compute the efficacy of any test based on a sequence $\{T_N\}$ we first have to find $\frac{d^2}{d\theta^2} C(\theta)\Big|_{\theta=0}$, where $C(\theta)$ is given by (14.2). First we obtain

$$\frac{1}{\lambda^2} \frac{dC(\theta)}{d\theta} = \int_0^{2\pi} \int_0^{2\pi} h'(H_\theta(x) - H_\theta(y)) (1-\lambda)(-f(x-\theta) + f(-\theta) + f(y-\theta) - f(-\theta)) f(x) f(y) \, dx \, dy,$$

and hence in particular

$$\frac{1}{\lambda^2}\frac{dC(\theta)}{d\theta}\bigg|_{\theta=0} = (1-\lambda)\int_0^{2\pi}\int_0^{2\pi} h'(F(x)-F(y))\,(f(y)-f(x))\,f(x)\,f(y)\,dx\,dy$$

$$= 0.$$

$$\frac{1}{\lambda^2}\frac{d^2C(\theta)}{d\theta^2}\bigg|_{\theta=0} = (1-\lambda)^2\int_0^{2\pi}\int_0^{2\pi} h''(F(x)-F(y))\,(f(y)-f(x))^2\,f(x)\,f(y)\,dx\,dy$$

$$+\ 2(1-\lambda)\int_0^{2\pi}\int_0^{2\pi} h'(F(x)-F(y))(f'(x)-f'(y))f(x)f(y)\,dx\,dy$$

$$= I + II, \text{ say.}$$

It is easy to see that $II = 0$ by integrating with respect to y and using the periodicity of f and h. Upon substituting $u = F(x)$ and integrating by parts we obtain

$$I = (1-\lambda)^2\int_0^{2\pi}[h'(u-F(y))\,(f(F^{-1}(u))-f(y))^2]_{u=0}^{u=1}\,dF(y)$$

$$-\ 2(1-\lambda)^2\int_0^{2\pi}\int_0^1 h'(u-F(y))\,(f(F^{-1}(u))-f(y))\,\frac{f'(F^{-1}(u))}{f\,(F^{-1}(u))}du\,dF(y)$$

$$= III + IV, \text{ say.}$$

Using periodicity of h' and of f we see that $III = 0$. If we now define

(14.5) $\quad p(u) = \frac{f'}{f}\circ F^{-1}(u),$ we obtain easily

$$IV = 2(1-\lambda)^2\int_0^1[\int_0^1 h'(u-v)\,f(F^{-1}(v))\,dv]\,p(u)\,du$$

$$= 2(1-\lambda)^2\int_0^1[-(h(u-v)\,f(F^{-1}(v)))\Big|_0^1 + \int_0^1 h(u-v)\,p(v)]\,p(u)\,du$$

$$= 2(1-\lambda)^2\int_0^1\int_0^1 h(u-v)\,p(u)\,p(v)\,du\,dv.$$

Hence we get finally

(14.6) $$\frac{d^2C(\theta)}{d\theta^2}\bigg|_{\theta=0} = 2(1-\lambda)^2\,\lambda^2\int_0^1\int_0^1 h(u-v)\,p(u)\,p(v)\,du\,dv.$$

Bahadur efficiency of sequences $\{T_N\}$: The value of $\left.\frac{d^2c(\theta)}{d\theta^2}\right|_{\theta=0}$ shows how fast the probability goes to infinity as $N \to \infty$, for alternatives close to zero. If we use Fourier-series expansions for $h(\cdot)$ and for $p(\cdot)$, i.e.,

$$(14.7) \qquad h(x) = \sum_{-\infty}^{\infty} d_k \, e^{2\pi ikx},$$

$$(14.8) \qquad p(u) = \sum_{-\infty}^{\infty} p_k \, e^{2\pi ikx},$$

then the convolution $g(v) = \int_0^1 h(u-v)\, p(u)\, du$ has Fourier coefficients $g_k = d_k p_k$. Hence the integral in (14.6) (which in $L_2[0,1]$ is the inner product of $g(\cdot)$ and $p(\cdot)$) is equal to $\sum_{-\infty}^{\infty} d_k \, |p_k|^2$.

Since we want to have this term as large as possible (in order to obtain high efficiency), we will from now on assume that $d_k \geq 0$ for all k.

The "slope" $s(\theta)$ of the sequence $\{T_N\}$ is defined by $s(\theta) = a(C(\theta))^t$. Since in our present case $t = 1$ and $a = \frac{1}{\lambda(1-\lambda)\max d_k}$ by Lemma 14.1, we obtain

$$(14.9) \qquad s(\theta) = \frac{1}{\lambda(1-\lambda)\max d_k} \int_0^{2\pi}\int_0^{2\pi} h(H_\theta(x) - H_\theta(y))\, dF(y)\, dF(x).$$

Using (14.6) we compute the efficacy of $\{T_N\}$ as

$$(14.10) \qquad \text{eff}\,(T_N) = \left.\tfrac{1}{2}\frac{d^2}{d\theta^2} s(\theta)\right|_{\theta=0} = \frac{\lambda^2(1-\lambda)^2}{\max d_k \; \lambda(1-\lambda)} \int_0^1\int_0^1 h(u-v)\, p(u)\, p(v)\, du\, dv.$$

Since the efficacy of the "best" parametric test is equal to

$$(14.11) \qquad \lambda(1-\lambda)\, \text{Inf}\,(f)$$

by Section 9, we obtain the efficiency

$$(14.12) \qquad e\,(T_N, \text{best test}\,|\,f) = \frac{\int_0^1\int_0^1 h(u-v)\, p(u)\, p(v)\, du\, dv}{\max d_k \;\; \text{Inf}\,(f)}.$$

Consistency of sequences $\{T_N\}$: Using Lemma 3.1 we immediately get the following result:

Lemma 14.2: Let $d_k \geq 0$ and p_k be defined by (14.7) and (14.8) respectively. If $\sum_{k=-\infty}^{\infty} d_k |p_k|^2 > 0$, there exists a neighborhood U of 0 such that for $\theta \in U$, $\theta \neq 0$, the sequence of tests $\{T_N \geq C_{N,\alpha}\}$ is consistent.

Proof: We have seen above that $\left.\frac{dC(\theta)}{d\theta}\right|_{\theta=0} = 0$ and that

(14.13) $$\left.\frac{d^2C(\theta)}{d\theta^2}\right|_{\theta=0} = 2\lambda^2(1-\lambda)^2 \sum_{k=-\infty}^{\infty} d_k |p_k|^2 > 0.$$

Hence there exists a neighborhood U of 0 such that $\theta \in U$, $\theta \neq 0$ implies $C(\theta) > 0$. Hence for this neighborhood U, the slope is > 0 and by Lemma 3.1 we obtain the desired result.

Hilbert space interpretation of efficiency term: As in the one-sample case it is convenient to interpret the efficiency (14.12) as a relationship between elements of $L_2[0,1]$, the space of square integrable complex functions on $[0,1]$. First we note that

(14.14) $$\text{Inf}\ (f) = \int_0^{2\pi} \left[\frac{f'(x)}{f(x)}\right]^2 dF(x) = \int_0^1 \left[\frac{f'(F^{-1}(v))}{f(F^{-1}(v))}\right]^2 dv = ||p(\cdot)||^2_{L_2}.$$

Furthermore it is well-known that the relation $r(\cdot) \to s(\cdot)$ defined by

(14.15) $$s(x) = \int_0^1 h(x-y)\ r(y)\ dy \qquad 0 \leq x \leq 1$$

is a compact linear operator from L_2 into L_2. We denote it by H.

The sequence $\{e^{2\pi ikx},\ k = 0, \pm 1, \pm 2, \dots\}$ is a complete orthonormal system in $L_2[0,1]$. We now show that it is a set of characteristic functions (or eigenvectors) of H. Let

(14.16) $$h(x) = \sum_{-\infty}^{\infty} d_k\ e^{2\pi ikx},$$

then

$$(14.17) \qquad s(x) = \int_0^1 h(x-y)\, e^{2\pi imy} dy = \int_0^1 \sum_{k=-\infty}^{\infty} d_k\, e^{2\pi ik(x-y)}\, e^{2\pi imy}\, dy$$

$$= \sum_{k=-\infty}^{\infty} d_k \int_0^1 e^{2\pi ikx}\, e^{2\pi i(m-k)y}\, dy = d_m\, e^{2\pi imx},$$

(the interchangeability of integration and summation is a well-known fact in Fourier analysis, it also follows from general Hilbert-space theory). This result shows that the functions $e^{2\pi imx}$ are eigenvectors of the operator H and that the Fourier coefficients d_m are the corresponding eigenvalues. For compact operators in Hilbert spaces we have the relation

$$(14.18) \qquad ||H|| = \max_k |d_k|,$$

where d_k are the eigenvalues. Since we assume $d_k \geq 0$, real, we obtain

$$(14.19) \qquad ||H|| = \max_k d_k.$$

Finally the integral $\int_0^1\int_0^1 h(u-v)\, p(u)\, p(v)\, du\, dv$ can now be written in the form

$$(14.20) \qquad \int_0^1 \left(\int_0^1 h(u-v)\, p(u)\, du \right) p(v)\, dv = (Hp, p)_{L_2}.$$

Hence we obtain the following expression for the efficiency from (14.12), (14.14), (14.19) and (14.20):

$$(14.21) \qquad e\,(T_N, \text{ best test}|f) = \frac{(Hp,p)}{||H||\;||p||^2}.$$

<u>Maximization of the efficiency</u>: It follows from our definition of efficiency and also from general Hilbert space theory that $(Hp, p) \leq ||H||\;||p||^2$ We want to find out how close we can get to equality by a suitable choice of H (for given p). W.l.o.g. we may assume that $||H|| = 1$. It is immediate that equality is achieved if and only if $Hp = p$. If we express this condition by means of Fourier coefficients, we obtain

(14.22) $\qquad d_k p_k = p_k, \qquad k = \pm 1, \pm 2, \ldots, \max d_k = 1.$

(For $k = 0$ the relation is always satisfied.) From this relation we get the following

Theorem 14.1: If $p(\cdot)$ has a finite Fourier expansion, then there exists an efficient test sequence among the class (14.1). It is given by $h(x) = \sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} d_k e^{2\pi ikx}$, $d_k = 1$ if $p_k p_{-k} \neq 0$, 0 otherwise. If $p(\cdot)$ has an infinite Fourier-expansion, an efficient test sequence of the form (14.1) does not exist, but for every $\epsilon > 0$ there is a test sequence with efficiency $> 1-\epsilon$.

Proof: The first part of the Theorem follows directly from (14.22). For the second part we would have to take an infinite number of d_k's equal to 1 in order to satisfy (14.22). But then the Fourier series (14.7) cannot converge. However if we take $d_0 = 0$, $d_k = 1$ for $k = \pm 1, \pm 2, \ldots, \pm K$, then by (14.21) the efficiency of the test sequence based on the corresponding h is

(14.23) $$e\,(T_N | \text{best test}; f) = \sum_{k=-K}^{K} |p_k|^2 \Big/ \sum_{k=-\infty}^{\infty} |p_k|^2,$$

and this term approches 1 as $K \to \infty$.

Remark: It follows from Theorem 12.8 that under the hypothesis the test statistic $T_N/[2\lambda(1-\lambda)]$ corresponding to $h(\cdot)$ of Theorem 14.1 has an asymptotic χ_K^2-distribution (with K d.f.), where K is the number of elements d_k for which $d_k = 1$.

15. An Example Studied by Wheeler and Watson.

We give one illustrative example for a test statistic of the form (11.1): S. Wheeler and G. S. Watson (1964) proposed a non-parametric test for equality of two circular distributions which can be described as follows: Place the two samples on the circle, change the angles between "successive" observations in such a way that all are equal $(= \frac{2\pi}{N})$, compute the length R of the vector resultant of the "adjusted" X-observations and reject the Null hypothesis of equality if R is too large.

If points on the circle are considered as unit vectors on the complex plane, then the statistic R^2 is defined as follows:

$$R^2 = \sum_{j=1}^{m} \exp\{2\pi i R_j/N\} \left(\overline{\sum_{k=1}^{m} \exp\{2\pi i R_k/N\}} \right) = \sum_{j=1}^{m} \sum_{k=1}^{m} \exp\{2\pi i (R_j - R_k)/N\}$$

$$= \sum_{j=1}^{m} \sum_{k=1}^{m} \cos 2\pi \left(\frac{R_j - R_k}{N}\right) \qquad \text{(since } R^2 \text{ is real)},$$

where the ranks R_j $(j = 1, 2, \ldots, m)$ are defined by choosing an arbitrary cut-off point and an arbitrary direction on the circle. Hence R^2/N is of the form (11.2) with $h_N(x) = \cos 2\pi x$, $N = 1, 2, \ldots$.

For the calculation of the efficiency of this test in the case of the v. Mises distribution we use (14.12). Here

$$d_1 = d_{-1} = \tfrac{1}{2}, \qquad d_k = 0 \text{ for } k \neq \pm 1;$$

$$\text{Inf}(f) = k\, I_1(k)/I_0(k) \qquad \text{from (A.10)};$$

$$p(u) = \frac{f'}{f} \circ F^{-1}(u) = -k \sin F^{-1}(u);$$

$$h(u-v) = \cos 2\pi(u-v) = \cos 2\pi u \cos 2\pi v + \sin 2\pi u \sin 2\pi v.$$

Hence

$$\int_0^1 \int_0^1 h(u-v)\, p(u)\, p(v)\, du\, dv$$

$$= \left[-k \int_0^1 \cos 2\pi u \sin F^{-1}(u)\, du\right]^2 + \left[-k \int_0^1 \sin 2\pi u \sin F^{-1}(u)\, du\right]^2$$

$$= k^2 \left[\int_0^1 \sin 2\pi u \sin F^{-1}(u)\, du\right]^2, \qquad \text{by skew-symmetry of } \sin F^{-1}(u).$$

We thus get the efficiency expression

(15.1) $$e(\text{Wheeler-Watson} \mid \text{best test; v. Mises distribution}) = \frac{2k \left[\int_0^1 \sin 2\pi u \sin F^{-1}(u)\, du \right]^2}{I_1(k)/I_0(k)} .$$

In the limiting case $k \to 0$ we get $F^{-1}(u) \to 2\pi u$, and hence by the dominated convergence theorem

$$\int_0^1 \sin 2\pi u \sin F^{-1}(u)\, du \to \int_0^1 \sin^2 2\pi u\, du = \tfrac{1}{2}.$$

By (A.12) and (A.11)

$$I_1(k) \doteq \frac{k}{2}, \qquad I_0(k) \doteq 1. \qquad \text{as } k \to 0.$$

As $k \to 0$ we thus obtain

(15.2) $$\lim_{k \to 0} e(\text{W-W} \mid \text{best test; v. Mises distr.}) = \frac{2k \cdot 1/4}{k/2} = 1.$$

This shows that for values of k which are close to 0 the Wheeler-Watson statistic will have a high efficiency in the case of a v. Mises distribution.

IV. APPENDIX

α. Some Properties of the v. Mises Distribution.

The v. Mises distribution is defined by the density

$$\text{(A.1)} \qquad f_{k,\theta}(x) = C(k)\, e^{k \cos (x-\theta)} \qquad \text{for } 0 \le x < 2\pi \quad (k \ge 0, \quad 0 \le \theta < 2\pi)$$

where $C(k)$ is a suitable normalizing constant. We first evaluate $C(k)$. It is a well-known fact [see e.g. F. Bowman (1958), p. 89] that the function $e^{ik \sin x}$ (as a function of x) has a Fourier expansion with the Bessel functions as "constants." More specifically,

$$\text{(A.2)} \qquad e^{ik \sin x} = \sum_{n=-\infty}^{\infty} J_n(k)\, e^{inx} \qquad \text{for real } x,$$

where k is an arbitrary complex number and $J_n(k)$ is the Bessel function of order n. The convergence is uniform with respect to x. Using this fact we obtain

$$\text{(A.3)} \qquad C(k)^{-1} = \int_0^{2\pi} e^{k \cos x}\, dx = \int_0^{2\pi} e^{k \sin (x + \pi/2)}\, dx$$

$$= \int_0^{2\pi} \sum_{n=-\infty}^{\infty} J_n(\tfrac{k}{i})\, e^{in(x+\pi/2)}\, dx$$

$$= \sum_{n=-\infty}^{\infty} \int_0^{2\pi} J_n(\tfrac{k}{i})\, e^{in(x+\pi/2)}\, dx$$

$$= J_0(\tfrac{k}{i})\, 2\pi$$

$$= 2\pi\, I_0(k),$$

where in general we define

$$I_n(k) = i^{-n} J_n(ik).$$

$I_n(k)$ is real for real k. It is called Bessel function of purely imaginary argument.

At various places we need the expectations of $\cos X$, $\sin X$, $\cos^2 X$, $\sin^2 X$, where X has the v. Mises distribution. We now derive these expectations. We indicate the value of θ by writing E_θ for expectation.

(A.4) $E_0 \sin X = 0,$ by symmetry.

(A.5)
$$E_0 \cos X = C(k) \int_0^{2\pi} \cos x \, e^{k \cos x} \, dx = C(k) \int_0^{2\pi} \frac{\partial}{\partial k} e^{k \cos x} \, dx$$
$$= C(k) \frac{\partial}{\partial k} \int_0^{2\pi} e^{k \cos x} \, dx = \frac{2\pi I_0'(k)}{2\pi I_0(k)} = \frac{I_1(k)}{I_0(k)} .$$

The last step follows from the relation $J_0'(k) = -J_1(k)$ [F. Bowman (1958), p. 93], hence $I_0'(k) = \frac{d}{dk} J_0(ik) = - i J_1(ik) = I_1(k)$.

Using this result we get more generally

(A.6)
$$E_\theta \sin X = C(k) \int_0^{2\pi} \sin x \, e^{k \cos (x-\theta)} \, dx$$
$$= C(k) \int_0^{2\pi} (\sin x \cos \theta + \cos x \sin \theta) \, e^{k \cos x} \, dx$$
$$= \cos \theta \, E_0 \sin X + \sin \theta \, E_0 \cos X$$
$$= \sin \theta \, E_0 \cos X.$$

By a similar argument we obtain

(A.7)
$$E_\theta \cos X = \cos \theta \, E_0 \cos X.$$

(A.8)
$$E_0 \sin^2 X = 1 - E_0 \cos^2 x = 1 - C(k) \frac{\partial^2}{\partial k^2} \int_0^{2\pi} e^{k \cos x} \, dx$$
$$= 1 - C(k) \frac{d^2}{dk^2} (2\pi \, I_0(k)) = 1 - 2\pi \, C(k) \frac{dI_1(k)}{dk}$$
$$= 1 + 2\pi i \, C(k) \frac{dJ_1(ik)}{dk} = 1 - \frac{1}{I_0(k)} J_1'(ik)$$
$$= 1 - \frac{1}{I_0(k)} \left(J_0(ik) - \frac{J_1(ik)}{ik}\right) = \frac{I_1(k)}{k \, I_0(k)} .$$

Here we used the general relation

$$J_n'(k) = J_{n-1}(k) - \frac{nJ_n(k)}{k} ,$$

[F. Bowman (1958), p. 93].

Since, in the case of the v. Mises distribution, we have

(A.9) $$\text{Inf}\ (f) = E_0\, k^2 \sin^2 X$$

we obtain immediately

(A.10) $$\text{Inf}\ (f) = \frac{kI_1(k)}{I_0(k)}\ .$$

Finally we state a few results about $I_0(k)$, $I(k)$ for k close to zero and k close to $+\infty$.

(A.11) $$I_0(k) = 1 + \frac{k^2}{4} + o(k^4) \qquad \text{for } k \to 0.$$

(A.12) $$I_1(k) = \frac{k}{2} + o(k^3) \qquad \text{for } k \to 0.$$

(A.13) $$I_0(k) \doteq \frac{e^k}{2\pi k} \qquad \text{as } k \to +\infty.$$

$$I_1(k) \doteq \frac{e^k}{2\pi k} \qquad \text{as } k \to +\infty.$$

β. <u>Definitions and Results about Measures and Weak Convergence of Measures on Separable Hilbert Spaces</u>.

In this section we present a few definitions and results on convergence of measures on Hilbert spaces. Most of the results follow directly from the definitions or can be found in Y. V. Prokhorov (1956).

We always assume that the Hilbert space H is real and separable.

Let be the σ-field of subsets of H generated by the class of continuous linear functionals on H (i.e., the minimal σ-field with respect to which all elements of the dual space are measurable), then $(H, \ )$ is a measurable space. Any countably additive nonnegative set function $m(\cdot)$ defined on is called a measure on H.

We assume that $m(H) < \infty$.

<u>Definition</u>: A sequence of measures $\{\mu_N\}$ converges weakly to a measure μ (in symbols: $\mu_N \overset{\text{weakly}}{\ } \mu$), if

(A.14) $$\int_H f(x)\, d\mu_N(x) \rightarrow \int_H f(x)\, d\mu(x)$$

for all bounded continuous functions f on H.

This type of convergence is usually called "convergence in distribution" if H is finite-dimensional.

Convergence in distribution on finite dimensional spaces can be metrized. The so-called Levy-metric has the property that convergence of measures in this metric is equivalent to convergence in distribution.

A similar metric has been defined by Prokhorov for the class of finite measures on (H,). Convergence in this metric is the same as weak convergence. With this metric the class $M(H)$ of finite measures on (H,) becomes a complete separable metric space.

Lemma A.1: $\mu_N \xrightarrow{\text{weakly}} \mu$ implies $\overline{\lim_N}\, \mu_N(F) \le \mu(F)$ for any closed set F H.

Of great interest is the characterization of the compact subsets of $M(H)$, i.e., sets which can be covered by a finite ϵ-net for every $\epsilon > 0$.

Theorem A.1: A set M' $M(H)$ is compact if and only if

(1) $\sup_{\mu \in M'} \mu(H) < \infty$ and

(2) for any $\epsilon > 0$ there exists a compact K_ϵ H such that $\mu(H - K_\epsilon) \le \epsilon$ for every $\mu \in M'$.

A useful sufficient condition for compact subsets M' is given by the following theorem. Since H is separable there exists a countable complete orthonormal set $\{e_i\}$ of vectors. We assume that such a system has been chosen and define

(A.15) $$r_N^2(x) = \sum_{i=N}^{\infty} (x, e_i)^2.$$

Then we can state this

Theorem A.2: A set of measures $M' \subset M(H)$ for which

(1) $\sup_{\mu \in M'} \mu(H) < \infty$ and

(2) $\lim_{N \to \infty} \sup_{\mu \in M'} \int_H r_N^2(x)\, d\mu(x) = 0,$

is compact.

A powerful tool for analyzing convergence in distribution of measures on finite dimensional spaces is the characteristic function. A corresponding transformation can be defined on Hilbert spaces.

Definition: The function $\chi(\cdot,\mu)$ defined for any bounded linear functional f on H by the equation

$$\chi(f,\mu) = \int_H e^{i(f,x)}\, d\mu(x) \tag{A.16}$$

is called characteristic functional of the measure μ. Here $(\cdot,\cdot)$ is the inner product on H.

Since μ is finite it is obvious that the integral exists. The name characteristic functional is justified by the fact that a measure is uniquely determined by its characteristic functional.

A characteristic functional is continuous in f.

By the definition of weak convergence we have, of course,

$$\mu_N \xrightarrow{\text{weakly}} \mu \quad \chi(f,\mu_N) \to \chi(f,\mu), \qquad \text{every } f \in H.$$

The converse $[\chi(f,\mu_N) \to \chi(f), \text{ continuous}] \Rightarrow \mu_N \xrightarrow{\text{weakly}} \mu$ is unfortunately not true for infinite-dimensional Hilbert spaces (and the usual topology on H). We have however the weaker

Theorem A.3: If $\{\mu_N,\ N = 1, 2, \ldots\}$ is compact and $\chi(f,\mu_N) \to \chi(f)$ for every $f \in H$, then for some $\mu \in M(H)$

$$\mu_N \xrightarrow{\text{weakly}} \mu \quad \text{and} \quad \chi(f) = \chi(f,\mu).$$

Y. Higher Order Random Terms.

For the lemmas of this subsection we make the following assumptions: X, Y are bounded random variables with continuous c.d.f. $F(\cdot)$, $G(\cdot)$, respectively. For convenience we assume $F(0) = G(0) = 0$, $F(2\pi) = G(2\pi) = 1$. $F_m(\cdot)$, $G_m(\cdot)$ are the empirical c.d.f.'s of independent samples of X's and Y's of sizes m and n, respectively. As $N = m + n \to \infty$, $\lambda_N = \frac{m}{N} \to \lambda$ with $0 < \lambda < 1$.

Lemma A.2: Let $h(x,y)$ be a bounded measurable function defined on the square $0 \le x \le 2\pi$, $0 \le y \le 2\pi$, then

$$\text{(a)} \quad \int_0^{2\pi}\int_0^{2\pi} h(x,y)\,(F_m(x) - F(x))^2\, dF(x)\, dF(y) = o_p(N^{-\frac{1}{2}}).$$

$$\text{(b)} \quad \int_0^{2\pi}\int_0^{2\pi} h(x,y)\,(F_m(x) - F(x))\,(G_n(x) - G(x))\, dF(x)\, dF(y) = o_p(N^{-\frac{1}{2}}).$$

$$\text{(c)} \quad \int_0^{2\pi}\int_0^{2\pi} h(x,y)\,(F_m(x) - F(x))\,(G_n(y) - G(y))\, dF(x)\, dF(y) = o_p(N^{-\frac{1}{2}}).$$

$$\text{(d)} \quad \int_0^{2\pi}\int_0^{2\pi} h(x,y)\,(F_m(x) - F(x))\,(F_m(y) - F(y))\, dF(x)\, dF(y) = o_p(N^{-\frac{1}{2}}).$$

Proof: It is obvious that all the integrals exist. We prove part (a). The proofs (b), (c), and (d) follow the same pattern.

By a theorem of Kolmogorov [see e.g. M. Fisz, p. 394] $N^{\frac{1}{2}} \sup_x |F_m(x) - F(x)|$ converges in distribution to some (finite) limit as $N \to \infty$. Hence $N^{\frac{1}{4}} \sup_x |F_m(x) - F(x)| \to 0$ in distribution and thus in probability.

Let $\epsilon > 0$ be given and let $|h(x,y)| \le K$. There exists N_ϵ such that for the sets

$$M_{\epsilon,N} = \{\omega : \sup_x N^{\frac{1}{4}} |F_m(x) - F(x)| > \frac{\epsilon}{K}\}$$

we have

$$P(M_{\epsilon,N}) < \epsilon \quad \text{for all} \quad N \ge N_\epsilon.$$

On the complement $M^c_{\varepsilon,N}$ we have

$$|N^{\frac{1}{2}} \int_0^{2\pi} \int_0^{2\pi} h(x,y)\ (F_m(x) - F(x))^2\ dF(x)\ dF(y)|$$

$$\leq \int_0^{2\pi} \int_0^{2\pi} K\ \frac{\varepsilon}{K}\ dF(x)\ dF(y) = \varepsilon.$$

Hence $P\{|N^{\frac{1}{2}} \int_0^{2\pi} \int_0^{2\pi} h(x,y)\ (F_m(x) - F(x))^2\ dF(x)\ dF(y)| > \varepsilon\} < \varepsilon$ for $N \geq N_\varepsilon$.

This completes our proof.

Lemma A.3: Let $h(x,y)$ be a bounded continuous function on the square $0 \leq x \leq 2\pi$, $0 \leq y \leq 2\pi$, with a continuous partial derivative $\frac{\partial h(x,y)}{\partial x} = h_1(x,y)$, then

(a) $$\int_0^{2\pi} \int_0^{2\pi} h(x,y)\ (G_n(x) - G(x))\ d(F_m(x) - F(x))\ dF(y) = o_p(N^{-\frac{1}{2}}).$$

(b) $$\int_0^{2\pi} \int_0^{2\pi} h(x,y)\ (F_m(x) - F(x))\ d(F_m(x) - F(x))\ dF(y) = o_p(N^{-\frac{1}{2}}).$$

(c) $$\int_0^{2\pi} \int_0^{2\pi} h(x,y)\ (G_n(y) - G(y))\ d(F_m(x) - F(x))\ dF(y) = o_p(N^{-\frac{1}{2}}).$$

(d) $$\int_0^{2\pi} \int_0^{2\pi} h(x,y)\ (F_m(y) - F(y))\ d(F_m(x) - F(x))\ dF(y) = o_p(N^{-\frac{1}{2}}).$$

Proof: It is again obvious that all the integrals exist. Parts (a) and (b) have been proven by Chernoff and Savage (1958, p. 986-989) under considerably weaker assumptions (their C_{2N}-term and C_{1N}-term, respectively). We prove part (c): Integrating the inner integral by parts we obtain

$$\int_0^{2\pi} \int_0^{2\pi} h(x,y)\ (G_n(y) - G(y))\ d(F_m(x) - F(x))\ dF(y) =$$

$$\int_0^{2\pi} (G_n(y) - G(y))\ [h(x,y)\ (F_m(x) - F(x))]_0^{2\pi} dF(y)$$

$$- \int_0^{2\pi} \int_0^{2\pi} (G_n(y) - G(y))\ (F_m(x) - F(x))\ h_1(x,y)\ dx\ dF(y).$$

Since the first term of the last line is 0, the result is obtained by the argument that was used in proving Lemma A.2, part (c).

Part (d) can be obtained by a similar kind of reasoning.

Lemma A.4: Let $h(x,y)$, $h_1(x,y) = \frac{\partial h(x,y)}{\partial x}$, $h_{12}(x,y) = \frac{\partial^2 h(x,y)}{\partial y \partial x}$ exist and be continuous, then

$$\int_0^{2\pi} \int_0^{2\pi} h(x,y)\, d(F_m(x) - F(x))\, d(F_m(y) - F(y)) = o_p(N^{-\frac{1}{2}}).$$

Proof: By carrying out one partial integration with respect to x and one partial integration with respect to y, the result can be obtained by the reasoning used to prove Lemma A.2, part (d).

REFERENCES

Bahadur, R. R. (1960). "Stochastic comparison of tests" Ann. Math. Stat. 31, pp. 276-295.

Batschelet, E. (1965). Statistical Methods for the Analysis of Problems in Animal Orientation and Certain Biological Rhythms, The American Institute of Biological Sciences.

Bowman, F. (1958). Introduction to Bessel Functions, Dover Publications, Inc., New York.

Box, G. E. P. (1954). "Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification" Ann. Math. Stat. 25, pp. 290-302.

Buehler, R. J. (1965). "The limit of the n^{th} power of a density" Ann. Math. Stat. 36, pp. 1878-82.

Chernoff, H. and Savage, I. R. (1958). "Asymptotic normality and efficiency of certain nonparametric test statistics" Ann. Math. Stat. 29, pp. 972-994.

Fisz, M. (1963). Probability Theory and Mathematical Statistics, Wiley, New York.

Gleser, L. J. (1964). "On a measure of test efficiency proposed by R. R. Bahadur" Ann. Math. Stat. 35, pp. 1537-44.

Govindarajulu, Z. (1960). Central limit theorems and asymptotic efficiency for one-sample non-parametric procedures, Technical Report No. 11, University of Minnesota, Minneapolis, Minnesota.

Grenander, U. (1963). Probabilities on Algebraic Structures, Wiley, New York.

Gumbel, E. J., Greenwood, J. A., and Durand, D. (1953). "The circular normal distribution" J. Amer. Stat. Ass. 49, pp. 267-297.

Hajek, J. (1961). "Some extensions of the Wald - Wolfowitz - Noether theorem" Ann. Math. Stat. 32, pp. 506-523.

Lehmann, E. L. (1959). Testing Statistical Hypothesis, Wiley, New York.

Pitman, E. J. G. (1948). Lectures in Non-parametric Statistical Inference, Columbia University, unpublished.

Prokhorov, Y. V. (1956). "Convergence of random processes and limit theorems in probability theory" Theory of Probability and its Applications 1, pp. 158-214.

Wahba, G. (1967). On the joint distribution of circulant forms in jointly stationary time series, Technical Report No. 17, Stanford University, Stanford, California.

Wheeler, S. and Watson, G. S. (1964). "A distribution-free two-sample test on a circle" Biometrika 51, pp. 256-258.