CONTRIBUTIONS TO THE THEORY OF TWO-SAMPLE RANK TESTS

William John Rowland Eplett

A Thesis Submitted to the Faculty of Science
University of the Witwatersrand, Johannesburg for the
Degree of Doctor of Philosophy

Johannesburg, 1977 .

ABSTRACT

The first part of this thesis concerns tests with test statistics which are nonlinear functions of the rank vector. The discussion is within the basic framework of testing two samples of circular data in order to ascertain whether they are drawn from the same population. The relevance of these tests to the standard situation is also stressed since they can be used for testing against an unspecified alternative hypothesis.

The principal test statistics considered arise from the Mann-Whitney test statistic adapted to conform with the invariance requirements of a test for circular data. An important relationship exists between these statistics and the Smirnov test statistics. This relationship, for instance, enables us to determine the range of possible values of the statistics and simplifies their computation.

The null distribution for these statistics can be obtained in terms of the distribution for the Mann-Whitney test statistic. The reasoning used in obtaining this result also enables us to derive some well-known results about the Kolmogorov-Smirnov statistics in a new fashion.

Other aspects of these Mann-Whitney type tests are also discussed, including the limiting distributions of the test statistics, a lower bound for the power, their Bahadur efficiency against Kuiper's (two-sample) test and a confidence interval for a certain measure of the distance between two (cumulative) distributions. Tabulation of exact probabilities for small sample sizes is also included.

A general class of tests is derived from the linear rank statistics for regression by using the method of union-intersection. This class includes the Mann-Whitney type tests as a particular case. A useful feature of these tests is that in the two-sample case they provide a class of tests consistent against an unspecified alternative.

The limiting distribution of these test statistics under the null hypothesis of randomness is obtained as the distribution of a functional of a Gaussian process. A pleasing feature of this work is that it also enables us to obtain the limiting distribution of another class of statistics which are quadratic forms in the ranks. The most significant aspects of previous work in connection with these quadratic forms are incorporated and extended within this discussion.

An expression is derived for the exact Bahadur efficiencies of the two-sample tests within the general class. The main result is that a test in the general class is never less efficient than the linear rank statistic from which it was derived, *irrespective of the alternative being tested.* Conditions are given under which the nonlinear test is actually strictly more efficient than the linear test.

The first part of the thesis concerns tests which may be used against on unspecified alternative and in this sense the tests may be regarded as robust. The second part of the thesis examines the performance of linear rank statistics with a view to ascertaining their robustness to deviations from the theoretical model. One technique involves interpreting the Haar expansion of the weighting function of a two-sample test in terms of asymptotic relative efficiency in order to give a detailed picture of the (asymptotic) behaviour of the test.

The thesis is concluded with comments on some aspects of robustness and proposed directions for development. These include the idea of an influence curve (à la Hampel) for the limiting power of a test and also approximations to optimal tests which give robust tests having test statistics easy to compute in practice.

In conclusion, it should be stated that this thesis - and particularly Part I - is more concerned with theoretical ideas than with immediate application. In this regard a firm theoretical footing, even if it has not been brought to a completely satisfactory conclusion as far as application goes, seems of more worth than finding heuristic solutions with certified practical efficiency only for particular examples.

ACKNOWLEDGEMENTS

The useful advice of my supervisor Dr. G.V. Kass is acknowledged. Conversations with Dr. J.N. Ridley and with Dr. M.S. Sears have also been very useful. The computer programming required for the tables included in the thesis was mostly done by Miss L. Ogilvy.

The actual typing of this thesis was done with considerable efficiency by Mrs. M.E. van der Westhuizen.

CONTENTS PAGE

PART I

PART II

LIST OF TABLES

PART 1

CHAPTER 1. Introduction to Hypothesis Testing for Circular Data.

1.1 A general outline of the topics examined in Part I.

The idea of investigating a random sample of observations scattered on the circumference of a circle is well established. The applicability of such investigations to situations arising in practice is illustrated by numerous examples in Mardia (1972, pp.12-17). Among these examples of situations which give rise to data in the form of points on a circle are the following: wind directions, hydrologic cycles, the flight orientation of birds in homing and migration, observations on experiments in a bubble chamber (made through a circular window) and psychological experiments in determining directions without visual aid.

The question often arises naturally as to whether two random samples of independently drawn observations are drawn from the same population or whether they are drawn from different populations. If both samples are drawn from the same population, then the combined sample is randomly distributed over the circumference of the circle and hence this hypothesis is that of randomness. Another possible question which may arise in the one-sample situation is whether the sample is drawn from a particular, given population.

Seemingly the most important difference between dealing with observations on a circle and the usual case where the observations lie on the real line is concerned with a certain arbitrariness of measurement in the circular case. This is because we have to choose some point on the circumference of the circle and then measure the angular displacements of the observations relative to this fixed point, also choosing the direction in which we measure (clockwise or anticlockwise) arbitrarily. A basic requirement of any test for either the one-sample or two-sample situation is that it does not depend on the choice of the point on the circumference from which angular displacement is measured. This requirement is a standard one in hypothesis testing (Lehmann,1959, p.213).

This invariance requirement means that the standard two-sample rank tests are not suitable for testing whether two samples of circular observations are drawn from the same population. New rank tests have to be developed which satisfy the invariance criterion. A variety of such tests have already been suggested and these will be discussed further in section 1.4.

Part I of this thesis, which forms the contents of chapters 2 and 3, is concerned with the development of a new class of tests for testing whether a sample of circular observations is randomly distributed (which, as mentioned earlier, includes the two-sample testing problem). These tests arise as an application of the method of union-intersection, described by Roy (1953), to the rank statistics used for testing the randomness of a sample of real-valued (non-directional) data.

The test derived using this technique applied to the well-known Mann-Whitney test (Mann and Whitney, 1947) possesses special features which encourage a separate analysis. We decided to present the discussion of this statistic in chapter 2 before developing the theory of the general class of tests to which it belongs. This may appear contrary to conventional practice, but there do seem to be compelling reasons for doing it. These reasons are: this procedure reflects the order in which the research was actually carried out; there is little overlap between the results of the two sections and those results in chapter 2 which are particular instances of results in chapter 3 can be derived more easily by direct arguments; putting the discussion of the Mann-Whitney type test first might be expected to draw the reader's attention to what seems to be the most promising of the tests in the general class and provides grounds for the generalization in chapter 3.

With regard to the general class of tests, two major results are obtained. The asymptotic distributions of the test statistics under the hypothesis of randomness are derived using the techniques of Billingsley (1968) for studying the convergence of probability measures. As mentioned earlier, each test in our class of tests is derived from a rank test for data on the real line. It is shown that the real line test is at best no more efficient- in the sense of Bahadur efficiency- than the test derived from it, irrespective of the alternative hypothesis being considered.

The results obtained for the asymptotic distributions of the test statistics do not in general permit an actual evaluation of limiting probabilities. In particular this is true for the Mann-Whitney type test. Using an expression relating the test statistic for this test to those of the (two-sample) Kolmogorov-Smirnov tests enables us to obtain a generating function for the probabilities of the null distribution for finite sample sizes. This generating function is expressed in terms of generating functions for the standard Mann-Whitney test. The technique employed in obtaining this result is an application of the principle of inclusion-exclusion (in the form used by Bender and Goldman (1975)) to extend results obtained by Steck (1969, 1971) for the Kolmogorov-Smirnov statistics.

A recurrence formula for generating the probabilities of the null distribution of the Mann-Whitney type test for finite sample sizes is also derived. The exact tail probabilities for selected sample sizes are tabulated.

1.2 Notation and preliminaries for rank tests

Let us begin by describing the two-sample testing problem for real-valued data. Suppose we have a random sample of $N(=m + n)$ independently drawn observations $X_1,\ldots,X_m,Y_1,\ldots,Y_n$, the first sample drawn from a population with (cumulative) distribution function $F(x)$ while the second comes from a population with distribution function $G(y)$, where both distribution functions are defined over the real numbers. The null hypothesis is that the combined sample of N observations is randomly distributed. This hypothesis is denoted by H and for this situation it may also be described by $H : F(x) = G(x)$ for all x.

The letter K will be used as a generic notation for the alternative hypothesis against which H is being tested. Under the assumption that F and G are continuous functions, the general alternative is expressed by $K : F(x) \neq G(x)$ for some x. Throughout Part I of this thesis, H is usually being tested against the general alternative, unless otherwise indicated.

The idea of ranking the observations in the combined sample is fundamental to the entire discussion and $R = (R_1, \ldots, R_N)$ will be used to denote the vector of ranks for the combined sample. Thus the rank of X_i in the combined sample is R_i, while the rank of Y_j is R_{m+j}. In order to effectively exclude the possibility that R is not well-defined, it is supposed that F and G are determined by densities. The assumption is tacitly made throughout this thesis that ties in the sample of observations occur with zero probability.

We shall use the indicator statistics Z_{Ni} which are defined by

$Z_{Ni} = 1$ if the ith observation in the ordered combined sample is an X,

$= 0$ otherwise $(i = 1, \ldots, N)$.

The rank statistics employed in chapter 3 arise from a more general problem than the two-sample testing problem. In the more general situation the sample $X_1,\dots,X_N$ is being tested for randomness using the vector of ranks $R = (R_1,\dots,R_N)$ where R_i is the rank of X_i in the ordered sample. The (simple) linear rank statistics for this problem are defined by

$$(1.2.1) \qquad S_N = \sum_{i=1}^{N} c_i a_N(R_i),$$

where $c = (c_1,\dots,c_N)$ and $a_N = (a_N(1),\dots,a_N(N))$ are vectors of real numbers. The terminology used is that of Hájek and Šidák (1967, p.61).

When attention is restricted to the two-sample case, linear rank statistics are obtained from (1.2.1) by putting

$$\begin{aligned} c_i &= 1 \quad i = 1,\dots,m \\ &= 0 \quad i = m+1,\dots,N. \end{aligned}$$

These statistics can also be expressed in terms of indicator statistics by

$$(1.2.2) \qquad S_N = \sum_{i=1}^{N} a_N(i) Z_{Ni}.$$

Let us now turn our attention to the problem where F and G are circular distributions. Although the absence of a precise formulation of the concepts involved in testing circular data does not seem to have hampered some researchers, it seems worthwhile to spend a little time placing the investigation on a sure footing. In doing this we follow the presentation used by Schach (1969b).

Th ... le being dealt with is regarded as being the set C ... plex numbers of modulus one. The mapping which takes x into e^{ix} is an isomorphism between $[0,2\pi)$ and C. This enables us to adopt the usual treatment for distribution functions and densities on C which may now be regarded as being represented by distribution functions and densities on $[0,2\pi)$. The definition of the distribution function $F(x)$ can be extended from $[0,2\pi)$ to the whole real line by defining $F(x+2k\pi) = F(x) \quad (k = \pm 1, \pm 2, \ldots)$.

As mentioned previously, the position of a point on the circle is measured in terms of its angular displacement relative to some fixed point. If the circle is now regarded as being the set C, then we take as the fixed point from which angular displacement is measured the point where C intersects the positive real axis. Angular displacement is then measured anticlockwise from this point. This amounts to saying that the image of a point on C under the inverse of the isomorphism gives the angular displacement of the point. The distribution $F(x)$ is then the distribution of the angular displacement as we have defined it.

Together with the circular distribution function $F(x)$ it is convenient to introduce the following two functions:

$$(1.2.3) \qquad \begin{aligned} F_\gamma(x) &= F(\gamma+x) - F(\gamma) \qquad x \in [0,2\pi-\gamma)) \\ &= 1 - F(\gamma) + F(x+\gamma-2\pi)\,(x \in [2\pi-\gamma,2\pi)) \end{aligned}$$

and

$$F^{-}_{\gamma}(x) = F(\gamma) - F(\gamma - x) \qquad (x \in [0,\gamma])$$
$$= 1 + F(\gamma) - F(2\pi+\gamma-x) \qquad (x \in (\gamma,2\pi)),$$

where $\gamma \in [0,2\pi)$. These functions express the effect upon the distribution function of changing the direction and fixed point in the measurement of the angular displacement. For instance if the cut-off point is changed to γ, then $F_{\gamma}(x)$ is the distribution function of the corresponding angular displacement. *Alternatively* $F_{\gamma}(x)$ may *be thought of as* the distribution function obtained through the transformation $e^{ix} \rightarrow e^{i(x-\gamma)}$. If not only the cut-off point, but also the direction of measurement is changed, then F^{-}_{γ} is obtained as the distribution of the angular displacement.

1.3 <u>*The requirement of invariance for circular tests*</u>

In order to implement the requirement of invariance mentioned in section 1.1, it is necessary to introduce a group of transformations of the sample space. The group chosen for C is, rather naturally roup of homeomorphisms of C onto itself and is det. $\mathcal{T}$. *An element of* $\mathcal{T}$ *is a mapping of the form* $\exp(ix) \rightarrow \exp\{i(c+t(x))\}$, *where* $c \in [0,2\pi)$ and $t(x)$ is a continuous, strictly monotone mapping of $[0,2\pi)$ onto itself. The group operation is composition of mappings.

The set of hypotheses for the two-sample testing problem will be indexed by the set Γ of pairs (F,G) of continuous, strictly increasing distribution functions such that $F(0) = G(0) = 0$ and $F(2\pi) = G(2\pi) = 1$. Then the problem of testing H against $K : F(x) \neq G(x)$ for some x, remains invariant under transformations by elements of T (Schach 1969b).

By defining the appropriate circular ranking, a maximal invariant under T can be obtained. Let $R = \{(r_1,\ldots,r_N) : (r_1,\ldots,r_N)$ is a permutation of the integers $1,\ldots,N\}$ and define the transformations g_r and g_i of R onto itself by

$$g_r : (r_1,\ldots,r_N) \to (r_1-1,\ldots,r_N-1)$$

and

$$g_i : (r_1,\ldots,r_N) \to (N+1-r_1,\ldots,N+1-r_N),$$

where modulo N arithmetic is used for the components of the transformed vectors where necessary.

Let G be the group of transformations of R onto itself generated by g_r and g_i. The orbits of G define equivalence classes over R and the orbit containing $r \in R$ is denoted by $a(r)$. If $R = (R_1,\ldots,R_N)$ is the ranking of the combined sample with respect to an arbitrary cut-off point and direction of measurement, then $a(R)$ does not depend on these two parameters. Standard arguments (Lehman (1959)) show that $a(r)$ is a maximal invariant under T (Schach, 1969b). Any test invariant under T must be distribution-free.

In general, the requirement of invariance for the circular testing problem means that the appropriate rank tests are those which are invariant under the transformations of G.

1.4 A survey of related research

Individual rank tests which have been proposed for testing whether two samples are drawn from the same circular distribution include those suggested by Kuiper (1960), Watson (1962) and Wheeler and Watson (1964). Another well researched test for this problem is the number of runs test (Mardia, p.203). The tests associated with Watson and with Wheeler and Watson can be included within the general class of tests proposed by Schach (1970). The test statistics of these tests have the general form

$$(1.4.1) \qquad T_N = N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} h_N\left(\frac{i-j}{N}\right) z_{Ni} z_{Nj}$$

where $\{h_N(x)\}$ is a sequence of functions defined for $x \in [-1,1]$, each of which is symmetric about 0 and periodic with period 1.

Schach investigated the limiting null distributions of the statistics T_N and also the limiting null distributions of a more general class of statistics obtained by replacing the sequence of scores $\{h_N(x)\}$ by a sequence $\{h_N(x,y)\}$ defined on the unit square (Schach, 1969a). Furthermore Schach showed that a locally most powerful invariant test exists for testing H against rotation alternatives and that its test statistic is of the form (1.4.1), provided certain standard smoothness assumptions hold for the underlying distribution (Schach 1969b).

Beran (1969) has considered a particular subclass of the statistics of the form (1.4.1) for which the scores $h_N(x)$ are independent of N (the sample size). By observing that such statistics can be derived in a natural way from statistics used for testing the one-sample hypothesis of uniformity, he has obtained the limiting null distributions of these statistics in an easier way than Schach did (although Schach's results are, of course, much more general).

Beran (1975) has investigated the asymptotic distributions and local asymptotic efficiencies of a class of quadratic rank tests for trend-in-location alternatives, which, although not directly applicable to circular data - because the test statistics are not necessarily invariant under G - nevertheless does contain some of the circular tests (for example it contains Watson's test) and the methods of investigation employed are in many respects similar to those for circular tests.

The idea behind the class of tests introduced in chapter 3 is most clearly foreshadowed in a note written by Barr and Shudde (1973) indicating that Kuiper's (one-sample) test could be derived from the Kolmogorov-Smirnov (one-sample) test using the method of union-intersection. Another well-known test which can be written in a union-intersection form is Ajne's test (Ajne, 1968, pp.343-344). A test proposed by Killeen (described in Killeen and

Hettmansperger (1972)) for a bivariate location problem can also be written in union-intersection form. The reader may be surprised at this last reference. However it is common knowledge that there is often a close relationship between tests for bivariate data and tests for circular data (see, for instance the remark on p.198 of Mardia's book and the paper by Bhattacharyya and Johnson (1969)). Furthermore this bivariate test is derived using the one-sample Wilcoxon statistic written in a way which suggested the form of the Mann-Whitney statistic used in chapter 2 and also a result derived by Killeen and Hettmansperger (1972) for obtaining the large deviation of the sequence of test statistics is central to one of the results of chapter 3.

The techniques used in chapters 2 and 3 in order to establish the convergence of the relevant probability measures are based on the ideas of Prohorov (1956). Wichura (1971) has demonstrated - with a piece of remarkable insight - that the general weak convergence theorems we use can be derived in a straightforward manner bypassing the technical development required by Prohorov in order to reach the same results. As Wichura himself points out though (p.1772), the real difficulty arises in applying the general results to specific situations. In the approach used in chapter 3, the methods are suggested by the derivation of the asymptotic distributions of Kolmogorov-Smirnov type statistics in Hájek and Šidák (1967), pp.184-189) and use numerous results from the book by Billingsley (1968).

The evaluation of the relative performances of tests by means of their exact slopes is proposed and developed by Bahadur (1960, 1967 and references in the 1967 survey paper). Considerable further research has been carried out in connection with this method of test evaluation. For our purposes we call the reader's attention to the following authors: Abrahamson (1967), Woodworth (1970), Hájek (1974) and the references therein. A clear distinction should be made between the exact slopes we shall be concerned with and the so-called approximate slopes often computed (as, for example, in Schach (1969b)). A discussion of this distinction may be found on pp.311-314 of Bahadur (1967).

The finite sample results obtained in section 2 of chapter II are combinational in nature and were inspired by the work of Steck (1971). Different approaches to deriving essentially similar results to those of Steck's are to be found in Epanechnikov (1968), Steck (1969) and Mohanty (1971). Appreciative comments about these results are made by Vincze (1972, pp.462-464 passim).

It is perhaps rather trite to observe that the development of the theory of rank tests has received considerable attention during the last twenty-five years. It does explain, however, why we have only mentioned that research directly related to the major themes of chapters 2 and 3. A comprehensive survey of the background material in every area touched upon would be too extensive. For this kind of information the reader's attention is drawn to books by the following authors: Hájek and Šidak (1967), Hájek

(1969), Puri (1970; editor of collection of papers), Puri and Sen (1971), Mardia (1972) and Lehmann (1975).

Certain basic ideas of functional analysis and group theory are required for the thesis. This material appears in standard texts on the relevant subjects and the reader might consult Hewitt and Stromberg (1969) for material related to functional analysis and Hall (1970) for backgroun to group theory.

CHAPTER 2. A Mann-Whitney Type Test for Circular Data.

2.1 Description of the test.

The test statistic for this test constitutes a fairly obvious adaptation of the usual Mann-Whitney test statistic defined by

$$\tilde{W}_N = \sum_{i=1}^{m} R_i ,$$

the sum of the ranks f the X's in the combined sample. The test statistic f. the circular Mann-Whitney test is defined as the maximum value of $\tilde{W}_N$ obtained from the different rankings of the combined sample arising from all possible choices ot cut-off point and direction of measurement. This test statistic, which we shall call $\tilde{\xi}_N$, is more usefully defined in terms of the group G, introduced in section 1.3, as

$$(2.1.1) \qquad \tilde{\xi}_N = \max_{g\in G} \{\tilde{W}_N(g(R))\}.$$

Most of the discussion of chapters 2 and 3 is phrased in terms of tests for circular data. Clearly, though, the natural homeomorphism between the real line and $(0,2\pi)$, defined by $\theta(x) = 2 \arctan (x) + \pi$, can be used to transform problems of hypothesis testing on the real line to tests on C. This mapping is strictly increasing and so preserves the ranking of the data. This suggests that rank tests introduced for testing circular data may also serve satisfactorily for testing data on the real line against broad alternative hypotheses. Results introduced in sections 3.4 and 3.5 serve to reinforce this contention.

Bearing in mind our remarks about the applicability of $\tilde{\xi}_N$ to testing data on the real line, another test is suggested which is invariant under the transformations of a subgroup of G. Suppose that G_1 is the subgroup of G generated by g_r, then $\tilde{\xi}_{N1}$ is defined by

$$(2.1.2) \qquad \tilde{\xi}_{N1} = \max_{g \in G_1} \{\tilde{W}_N(g(R))\}.$$

The group G_1 is induced by the subgroup T_1 of T consisting of mappings of the form: $\exp(ix) \to \exp\{i(c+t(x))\}$, where $c \in [0,2\pi)$ and $t(x)$ is a continuous, strictly increasing mapping of $[0,2\pi)$ onto itself. Discussion of the implications of the requirement that a test be invariant under T_1 proceeds along the lines of section 1.3. It must be conceded that such a requirement appears somewhat arbitrary. By way of justification, though a test statistic like $\tilde{\xi}_{N1}$ may be regarded as the one-sided version of $\tilde{\xi}_N$ and will be useful for testing the randomness of data on the real line. Further mention of this will be made in section 3.1.

The tests corresponding to $\tilde{\xi}_N$ and $\tilde{\xi}_{N1}$ both reject the null hypothesis of randomness for large values of the respective test statistics.

The scope of the analysis of $\tilde{\xi}_N$ (and $\tilde{\xi}_{N1}$) is greatly extended by an important relationship with the (two-sample)

Kolmogorov-Smirnov statistics (hereafter referred to as the Smirnov test statistics). If $F_m(z)$ and $G_n(z)$ are used to denote the empirical distribution functions of the X's and Y's respectively , then these statistics are defined by

(2.1.3) $$D^+_{m,n} = \sup_z\{F_m(z) - G_n(z)\},$$

$$D^-_{m,n} = \sup_z\{G_n(z) - F_m(z)\},$$

$$D_{m,n} = \max\{D^+_{m,n}, D^-_{m,n}\}.$$

The derivation of this relationship seems to become more transparent if instead of using the usual definition of the Mann-Whitney statistic as given by $\tilde{W}_N$, we instead define it as

(2.1.4) $$W_N = \sum_{i=1}^{m} R_i - \sum_{i=m+1}^{N} R_i,$$

and consequently work with the statistics

(2.1.5) $$\xi_N = \max_{g \epsilon G}\{W_N(g(R))\}$$

and $$\xi_N = \max_{g \epsilon G}\{W_N(g(R))\}$$

(2.1.6) $$\xi_{N1} = \max_{g \epsilon G_1}\{W_N(g(R))\}.$$

Relationships between ξ_N, ξ_{N1} and $\tilde{\xi}_N$, $\tilde{\xi}_{N1}$ are easily derived since $W_N = 2\tilde{W}_N - \{N(N+1)\}/2$, a linear transformation of $\tilde{W}_N$. Hence,

$$(2.1.7)\quad \xi_N = 2\tilde{\xi}_N - \{N(N+1)\}/2, \xi_{N1} = 2\tilde{\xi}_{N1} - \{N(N+1)\}/2.$$

Analogous to Z_{Ni} we can define

$$\begin{aligned} A_{Ni} &= 1 \quad \text{if} \quad Z_{Ni} = 1, \\ &= -1 \quad \text{if} \quad Z_{Ni} = 0, \end{aligned}$$

so that $W_N = \sum_{i=1}^{N} i\, A_{Ni}$. Then,

$$(2.1.8)\quad W_N((g_r)^{i+1}(R)) = W_N((g_r)^i(R)) + (n-m) + NA_{N(i+1)} \qquad (i = 0,\ldots,N-1).$$

To establish (2.1.8) notice that, using modulo N arithmetic for the coefficients of A_{Nk} where necessary so that they fall between 1 and N, we may write

$$\begin{aligned} W_N((g_r)^{i+1}(R)) &= \sum_{k=1}^{N} (k-i-1)A_{Nk} \\ &= \sum_{\substack{k=1 \\ k\neq i+1}}^{N} (k-i)A_{Nk} - \sum_{\substack{k=1 \\ k\neq i+1}}^{N} A_{Nk} + NA_{N(i+1)} \\ &= \sum_{k=1}^{N} (k-i)A_{Nk} - \sum_{k=1}^{N} A_{Nk} + NA_{N(i+1)} \\ &= W_N((g_r)^i(R)) + (n-m) + NA_{N(i+1)}. \end{aligned}$$

Applying the reduction in (2.1.8) i times and writing $S_i = A_{N1} + \ldots + A_{Ni}$, we get

$$W_N((g_r)^i(R)) = W_N(R) + i(n-m) + NS_i \qquad (i = 1,\ldots,N), \tag{2.1.9}$$

and hence

$$W_N((g_r)^i(R)) = W_N(R) + 2mn\{F_m(X^{(i)}) - G_n(X^{(i)})\} \tag{2.1.10}$$

where $X^{(i)}$ is the ith observation in the ordered combined sample. As an immediate consequence of (2.1.10), it follows that

$$\xi_{N1} = W_N(R) + 2mnD^{+}_{m,n}. \tag{2.1.11}$$

Since

$$\begin{aligned} W_N(g_1(R)) &= \sum_{i=1}^{m} (N+1 - R_i) - \sum_{i=m+1}^{N} (N+1-R_i) \\ &= m(N+1)-n(N+1) - \{\sum_{i=1}^{m} R_i - \sum_{i=m+1}^{N} R_i\} \\ &= (m-n)(N+1) - W_N(R), \end{aligned}$$

it then follows that

$$\begin{aligned} \xi_N &= \max_{g \in G_1} \{\max[W_N(g(R)),(m-n)(N+1) - W_N(g(R))]\} \\ &= \max\{W_N(R) + 2mn\ D^{+}_{m,n}, (m-n)(N+1) - W_N(R) + 2mn\ D^{-}_{m,n}\}. \end{aligned} \tag{2.1.12}$$

The appropriate expression for $\tilde{\xi}_N$ is obtained by substituting for $W_N(R)$ in terms of $\tilde{W}_N(R)$ in (2.1.12), obtaining

$$\begin{aligned}\xi_N &= \max\{2\tilde{W}_N(R) - (N(N+1))/2 + 2mn\, D^+_{m,n},\\ &\qquad 2m(N+1) - 2\tilde{W}_N(R) + 2mn\, D^-_{m,n} - (N(N+1))/2\}\\ &= 2\max\{\tilde{W}_N(R) + mn\, D^+_{m,n}, m(N+1) - \tilde{W}_N(R)\\ &\qquad + mn\, D^-_{m,n}\} - (N(N+1))/2.\end{aligned}$$

In consequence,

$$(2.1.13)\quad \tilde{\xi}_N = \max\{\tilde{W}_N(R) + mn\, D^+_{m,n}, m(N+1) - \tilde{W}_N(R) + mn\, D^-_{m,n}\}.$$

One immediate benefit of the expressions (2.1.11) - (2.1.13) is that the test statistics involved are easy to calculate in practice. Suppose that, for instance, we require to calculate $\tilde{\xi}_N$ from a sample of data. It is a straightforward procedure to calculate $D^+_{m,n}$ and $D^-_{m,n}$ using a table or some form of empirical probability plot. Then by (2.1.13) the value of $\tilde{\xi}_N$ equals whichever value is the greater of $\tilde{W}_N(R) + mn\, D^+_{m,n}$ and $m(N+1) - \tilde{W}_N(R) + mn\, D^-_{m,n}$.

Another application of (2.1.13) lies in finding the range of values the statistic $\tilde{\xi}_N$ can assume. In view of its definition in (2.1.1), it is clear that the largest value $\tilde{\xi}_N$ can assume is the same as the largest value which $\tilde{W}_N$ may assume. This value is $\frac{1}{2}m(N+n+1)$ and is attained when the ranks of the X's are as large as possible, that is $n+1,\ldots,N$. The members of R for which $\tilde{\xi}_N$ attains its maximum are then the members of the orbits of G containing an $r \in R$ which corresponds to ranks of $n+1,\ldots,N$ for the X's.

The least value which $\tilde{\xi}_N$ may assume seems less obvious and requires an argument using elementary number theory. For this result and some of the material in section 2.2, it is convenient to change the meaning of some of the notation in order to take advantage of the two-sample nature of the test statistic. Thus we now regard the two samples as already ordered, so that $X_1<\ldots<X_m$ and $Y_1<\ldots<Y_n$. Then R_i will denote the rank of X_i in the ordered combined sample. This change has the effect of enabling us to use a new set

$$\bar{R}_m = \{(r_1,\ldots,r_m) : r_i \text{ integers, } 1 \le r_1<\ldots<r_m \le N\}$$

with the group $\bar{G}$ defined as the group generated by $\bar{g}_r$ and $\bar{g}_i$, where these two transformations map $\bar{R}_m$ onto itself and are defined by

$$\bar{g}_i : (r_1,\ldots,r_m) \to (N+1-r_m,\ldots,N+1-r_1)$$

and

$$\bar{g}_r : (r_1,\ldots,r_m) \to (r_1-1,\ldots,r_m-1) \quad (\text{if } r_1 > 1)$$

$$(r_1,\ldots,r_m) \to (r_2-1,\ldots,r_m-1,N) \,(\text{if } r_1 = 1).$$

This change in the meaning of R_i should not cause confusion as it only arises in discussion of the Mann-Whitney statistic and it will always be clearly indicated that this meaning of R_i is being used (as opposed to the meaning given R_i in section 1.2).

The result which gives us the least value $\tilde{\xi}_N$ may assume is contained in Proposition 2-1-2. Before stating that result, however, a lemma is required concerning the Smirnov test statistics. It was mainly in order to employ this lemma that the meaning of R_i has been changed at this stage. The lemma is theorem 2.3 (p.1452) of Steck (1969).

Lemma 2-1-1.

$$P\{mn\ D^{+}_{m,n} \le r,\ mn\ D^{-}_{m,n} \le s\}$$

$$= P\{(iN-r) \le mR_i \le (iN-n+s),\quad i = 1,\dots,m\}.$$

Proposition 2-1-2.

$$(2.1.14) \qquad \min_{R \in \tilde{R}_m} \max_{g \in \tilde{G}} \{\tilde{W}_N(g(R))\} = (N(m+1) + m-d)/2,$$

where $d = (m,n)$, the greatest common divisor (g.c.d.) of m and n.

Proof Without loss of generality we may reduce the problem to finding the least possible value of $\tilde{\xi}_N$ for those $R \in \tilde{R}_m$ for which $\tilde{W}_N(R) \ge \tilde{W}_N(g(R))$ for all $g \in \tilde{G}$. From (2.1.13), this last condition is equivalent to

$$(2.1.15) \qquad mn\ D^{+}_{m,n} \le 0 \text{ and } mn\ D^{-}_{m,n} \le 2 \sum_{i=1}^{m} R_i - m(N+1).$$

Using Lemma 2-1-1 to express mn $D^+_{m,n}$ and mn $D^-_{m,n}$ in terms of the R_i, (2.1.15) becomes

$$(2.1.16) \quad Ni \le mR_i \le 2\sum_{i=1}^{m} R_i - N(m+1) + Ni \qquad (i = 1,\ldots,m).$$

This suggests considering $R_i = \langle Ni/m\rangle, i = 1,\ldots,m$, where $\langle x\rangle$ denotes the smallest integer greater than or equal to x (if $[x]$ is used to denote the greatest integer less than or equal to x, then $\langle x\rangle = -[-x]$). *Then* $\sum_{i=1}^{m} R_i = N + \sum_{i=1}^{m-1} \langle Ni/m\rangle$. The problem is to now obtain an expression for the last term on the right hand side. This is achieved using an argument gleaned from the study of the Dedekind sum in number theory.

Put $((x)) = \langle x\rangle - x$, so that $((x))$ is periodic with period 1, and consider $\sum_{i=1}^{m-1} ((Ni/m))$. Then, since $(m,n) = d$ *implies that $(m,N) = d$, it follows that in the set of re*sidue classes $\{Ni : i = 1,\ldots,m-1\}$ each non-zero multiple of d appears d times and 0 appears d-1 times. This is a basic result of elementary number theory - for further details consult Le Veque (1965, p.32). In consequence,

$$\begin{aligned}
\sum_{i=1}^{m-1} ((Ni/m)) &= d \sum_{\substack{1\le k\le m-1\\ d|k}} ((k/m)) \\
&= d\Big\{\sum_{\substack{1\le k\le m-1\\ d|k}} \langle k/m\rangle - \sum_{\substack{1\le k\le m-1\\ d|k}} k/m\Big\} \\
&= d\Big\{(m/d - 1) - \frac{1}{m}\sum_{k=1}^{m/d-1} dk\Big\} \\
&= d\Big\{(m/d - 1) - \frac{d}{2m}\big(m/d\,(m/d - 1)\big)\Big\} \\
&= (m-d)/2.
\end{aligned}$$

Now the sum in which we are interested can be calculated as

$$\sum_{i=1}^{m-1} \langle Ni/m \rangle = (N(m-1))/2 + (m-d)/2$$

and hence

$$\sum_{i=1}^{m} R_i = \{N(m+1) + (m-d)\}/2.$$

The upper bound on mR_i from (2.1.16) then becomes $(m-d) + Ni$ for $i = 1,\ldots,m$ and since $R_i = \langle Ni/m \rangle$ does satisfy this upper bound, the proposition is proved.

(The assertion that $mR_i \le (m-d) + Ni$, $i = 1,\ldots,m$ is perhaps not obvious. It is clearly true when $m|(Ni)$. Therefore we need only consider $1 \le d \le m-1$ (necessarily $m > 1$) and $Ni = mq + r$, where q and r are positive integers with $1 \le r \le m-1$. Then $mR_i - Ni = m-r$. But since $d|m$ and $d|N$, $d|r$ and hence since $r \ge 1, d \le r$. Therefore $mR_i - Ni \le m-d$, as required).

Clearly it follows from Proposition 2-1-2 that the smallest possible value of $\tilde{\xi}_N$ is $(N(m+1) + m-d)/2$ obtained when $R_i = \langle Ni/m \rangle$ and that, up to orbits of $\tilde{G}$, these values of R_i are unique. This result is intuitively acceptable, since the smallest value of $\tilde{\xi}_N$ occurs when the ranks of the X's in the ordered combined sample are *evenly spaced*.

2.2 <u>The null distribution of the test statistic: asymptotic and finite sample results.</u>

An expression for the limiting distributions of ζ_N and ξ_{N1} (and hence $\tilde{\xi}_N$ and $\tilde{\xi}_{N1}$) can be easily derived. Full details of the derivation are delayed until section 3.3 in order to make use of the background material introduced in section 3.2. For the present merely the result is stated.

In the asymptotic theory of the Mann-Whitney test, it is assumed that $m,n \to \infty$ in such a way that $m/_N \to \lambda$, λ a constant such that $0 < \lambda < 1$.

Proposition 2-2-1. Suppose that $\xi(t)$ is the continuous, stationary, Gaussian process on $[0,1]$ with

$$E\{\xi(t)\} = 0 \qquad (t \in [0,1]),$$

and

$$R(s,t) = E\{\xi(s)\xi(t)\} = 1/_3 - 2(s-t)(1-(s-t))$$
$$(0 \le t \le s \le 1).$$

Then

(i) $(mnN)^{-\frac{1}{2}}\{\xi_N - \frac{1}{2}(m-n)(N+1)\}$ converges in distribution to $f(\xi(t))$;

(ii) $(mnN)^{-\frac{1}{2}}\{\xi_{N1} - \frac{1}{2}(m-n)(N+1)\}$ converges in distribution to $f_1(\xi(t))$,

where f, f_1 are continuous functionals on $C([0,1])$ (the space of continuous functions on $[0,1]$ with the metric $\rho(y,z) = \max_{0\le t\le 1}\{|y(t) - z(t)|\}$) defined by

$$f(z(t)) = \max_{0\le t\le 1}\{|z(t)|\}, f_1(z(t)) = \max_{0\le t\le 1}\{z(t)\}.$$

As far as the author is aware, the question of suitable expressions for the distributions of $f(\xi(t))$ and $f_1(\xi(t))$ - suitable for application in practice - does not appear to be adequately resolved. Further mention of this is made in section 3.3. It is, however, with these remarks in mind that we turn to examine the possibility of obtaining finite sample results.

Following the pattern of section 2.1, the discussion of finite sample results is made in terms of $\tilde{\xi}_N$ rather than ξ_N. Throughout this section the revised definition of $R = (R_1,\ldots,R_N)$, introduced first in section 2.1, is used. From (2.1.13) it follows that

$$(2.2.1) \qquad P(\tilde{\xi}_N < e) = \sum_t P(\tilde{W}_N(R) = t, mn\,D^+_{m,n} < e - t,$$
$$mn\,D^-_{m,n} < e + t - m(N+1)).$$

The range of values of summation for t can be usefully reduced by noting that from (2.1.1), $\tilde{W}_N(R) \le \tilde{\xi}_N$ and $\tilde{W}_N(g_1(R)) \le \tilde{\xi}_N$, so that only values of t such that $m(N+1) - e < t < e$ can give a positive contribution to the right hand side of (2.2.1). Using Lemma 2-1-1, (2.2.1) may be written as

$$(2.2.2) \qquad P(\tilde{\xi}_N < e) = \sum P(\tilde{W}_N(R) = t;$$
$$(iN-(e-t))/m < R_i < (e+t-(m+1-i)N)/m,$$
$$i = 1,\ldots,m).$$

This last expression suggests determining $W(b,c,t)$ *defined as* equal to the number of ways the event

$$\{\sum_{i=1}^{m} R_i = t;\ b_i < R_i < c_i,\ \ i = 1,\ldots,m\}$$

can occur in the ordered combined sample, where $b = (b_1,\ldots,b_m)$ and $c = (c_1,\ldots,c_m)$ are two increasing sequences of integers such that $i - 1 \le b_i < c_i \le n + i + 1$. The numbers $W(b,c,t)$ count the number of partitions of t into m unequal parts of size less than or equal to N where each of the parts is also subject to a further restriction defined by b and c. Any results obtained about $W(b,c,t)$ will thus also be of some combinatorial interest.

The major result obtained concerning $W(b,c,t)$ *is* suggested by the results of Steck given as Theorem A1-1 and A1-2 (of Appendix 1). In the appendix these results are actually derived as a consequence of Lemma 2-2-4.

Some unnecessary complication may be removed from the discussion by restating the problem of finding $W(b,c,t)$ in a different form. If $T_i = R_i - i$, $u_i = b_i - i + 1$, $v_i = c_i - i - 1$, $i = 1,\ldots,m$ *and* $w = t - \frac{1}{2}m(m+1)$, then

$$(2.2.3) \qquad \{\sum_{i=1}^{m} R_i = t;\ \ b_i < R_i < c_i,\ i = 1,\ldots,m\}$$

if and only if

$$\{\sum_{i=1}^{m} T_i = w;\ \ u_i \le T_i \le v_i,\ \ i = 1,\ldots,m\}$$

and $0 \le T_1 \le \ldots \le T_m \le n$. The number of vectors of integers which satisfy the second of these conditions will for convenience be denoted by $N_m(u,v,w)$ (where $u = (u_1,\ldots,u_m)$ and $v = (v_1,\ldots,v_m)$).

The main result is stated in terms of the following generating functions:

$$N_m^*(u,v,z) = \sum_{w=0}^{\infty} N_m(u,v,w)z^w$$

and

$$\pi^*(m,n,z) = \sum_{w=0}^{\infty} \pi(m,n,w)z^w,$$

where $\pi(m,n,w)$ equals the number of ways in which the event $\{\sum_{i=1}^{m} R_i = w,\ 1 \le R_1<\ldots<R_m \le N\}$ can occur and is just $\binom{N}{m}P(\tilde{W}_N(R) = w)$. Both N_m^* and π^* are in fact polynomials, since only a finite number of the terms in the summation are non-zero.

Theorem 2-2-2. If the components of the $m \times m$ matrix $(d(i,j))$ are given by

$$d(i,j) = z^{(b_j-j)(j-i+1)}\pi^*(j-i+1,c_i-b_j-2,z), \tag{2.2.4}$$

then

$$N_m^*(u,v,z) = \det\{(d(i,j))\}, \tag{2.2.5}$$

the determinant of the matrix $(d(i,j))$.

The generating function for the numbers $W(b,c,t)$ is then obtained from

$$(2.2.6) \qquad W^*(b,c,z) = \sum_t W(b,c,t)z^t = z^{(m(m+1))/2}N_m^*(u,v,z).$$

The significance of Theorem 2-2-2 then lies . .he fact that using (2.2.6) together with (2.2.2) we ·ave an expression for the probabilities of $\tilde{\xi}_N$ in terms of the probabilities of $\tilde{W}_N$, the test statistic from which $\tilde{\xi}_N$ was derived. The actual expression does not seem to gain from being explicitly formulated, but the theoretical importance of the result appears undeniable.

Let us make a few preliminary remarks about the form of the matrix defined by (2.2.4). Clearly if $i - j > 1$ or $c_i - b_j \le 1$, then $d(i,j) = 0$. If $i = j + 1$, then $d(i,j) = 1$ since we take $\pi^*(0,n,z) = 1$ for $n \ge 0$. Thus the matrix $(d(i,j))$ is of the same form as those obtained by Steck in connection with the distributions of the Kolmogorov-Smirnov statistics (Appendix 1), that is it has ones on the first subdiagonal and zeroes below the first subdiagonal. Since the $d(i,j)$ are members of the polynomial ring over the integers rather than members of a field, caution has to be exercised in applying results about the determinants of matrices whose components are elements of a field.

The principle of inclusion-exclusion will be used to derive Lemma 2-2-4 from which Theorem 2-2-2 may then by deduced. The form of inclusion-exclusion used here is derived from theorem 2 of Bender and Goldman (1975). The use of this less familiar form serves two purposes. The statement of the theorem given by Bender and Goldman is more general than the usual statement of the result (Riordan (1958, p.51)) and incorporates the kind of functions we are interested in. Secondly, in this case it seems simpler to deal with functions defined on the subsets of the set $\{1,\ldots,m-1\}$ than to introduce a list of $m-1$ properties in order to force this particular problem into a form suitable for the more usual statement of the result.

Lemma 2-2-3 gives the inclusion-exclusion result required and is an application of theorem 2 of Bender and Goldman (p.792) along the lines of example 2 of the same article (p.793).

Lemma 2-2-3. Suppose that $\mathcal{B}$ is the collection of 2^k subsets of $\{1,\ldots,k\}$ and that $N_=(x)$ is defined for all $x \in \mathcal{B}$. If

$$(2.2.7) \qquad N_{\leq}(x) = \sum_{y:y\subset x} N_=(y),$$

then

$$(2.2.8) \qquad N_=(x) = \sum_{y:y\subset x} (-1)^{|y|-|x|} N_{\leq}(y)$$

where $|x|$ denotes the number of elements in $x \subset \{1,\ldots,k\}$.

In accordance with convention, $\mathbb{Z}$ is used to denote the set of integers. Suppose that X^* is a function on the finite subsets of $\mathbb{Z}^m$ such that if S_1, S_2 are two such disjoint subsets, $X^*(S_1 \cup S_2) = X^*(S_1) + X^*(S_2)$. The number of elements in a finite subset S is denoted by $|S|$. The function X^* is said to be multiplicative if there exist functions $(X^*)_1, \ldots (X^*)_m$ such that if $\ell_1, \ldots, \ell_k > 0$, $\ell_1 + \ldots + \ell_k = m$ and $Z(i) \subset Z^{\ell_i}$, then $X^*(Z(1) \times \ldots \times Z(k)) = \prod_{i=1}^{k} (X^*)_{\ell_i}(Z(i))$ and then $(X^*)_{\ell_i}(Z(i))$ is written $X^*(Z(i))$.

Let $\chi_m(u,v) = \{(x_1, \ldots, x_m) : x_i \text{ integers}, 0 \le x_1 \le \ldots \le x_m \le n, u_i \le x_i \le v_i, i = 1, \ldots, m\}$. The sets $\mathcal{L}(i,j)$ are defined for $1 \le i \le j \le m$ by $\mathcal{L}(i,j) = \{(x_i, \ldots, x_j) : x_k \text{ integers}, v_i \ge x_i > \ldots > x_j \ge u_j\}$ (which will be empty if $v_i - u_j \le j-i-1$). By convention take $X^*(\mathcal{L}(i,i-1)) = 1$ and otherwise $X^*(\mathcal{L}(i,j)) = 0$ for $1 \le j + 1 < i \le m$.

<u>Lemma 2-2-4.</u> If the components of the $m \times m$ matrix $(d(i,j))$ are given by

$$d(i,j) = X^*(\mathcal{L}(i,j)), \tag{2.2.9}$$

then, if X^* is multiplicative,

$$X^*(\chi_m(u,v)) = \det((d(i,j))) \tag{2.2.10}$$

<u>Proof.</u> The matrix $(d(i,j))$ has the same special form as that of *Theorem 2-2-2. The first step in the proof consists* of showing that in the definition of the determinant of $(d(i,j))$ as a summation over elements of the symmetric group on m letters, S_m, only if $\phi \in S_m$ can be written as

$$\phi = (m_1(m_1-1)\ldots 1)((m_2+m_1)\ldots(m_1+1))\ldots((m_1+\ldots+m_k)\ldots(m_1+\ldots+m_{k-1} + 1))$$

for $m_1, \ldots, m_k > 0$ and $m_1 + \ldots + m_k = m$, can it possibly give a non-zero contribution to $\det\{(d(i,j))\}$.

To show this, suppose that $\phi \in S_m$ gives a non-zero contribution to $\det\{(d(i,j))\}$ and consider any cycle of its decomposition into disjoint cycles. Suppose this cycle is written $(i_1 .. i_\ell)$ where i_1 is the largest integer appearing in the cycle. If it is possible, choose the least subscript r $(2 < r \leq \ell)$ for which $i_r > i_{r-1}$ and call this subscript s. Clearly $i_{s-1} < i_s < i_1$. But in this case, $i_2 = i_1 - 1$ since $i_2 < i_1$ and ϕ gives a non-zero contribution to the determinant. Similarly $i_3 = i_2 - 1 = i_1 - 2$ and so on, so that each integer between i_1 and i_{s-1} has already appeared among the first s-1 numbers of the cycle. This would mean that i_s had already appeared in the cycle, giving two members of the cycle with the same value which is a contradiction. If it is not possible to find a subscript s, then $(i_1 .. i_\ell)$ must be of the required form for ϕ to give a non-zero contribution to the determinant.

The relationship between partitions of m and permutations which can make a non-zero contribution to $\det\{(d(i,j))\}$ enables us to deduce that

$$\det\{(d(i,j))\} = \sum_{\omega(m)} (-1)^{m-k(\omega)} \times \left(\prod_{\substack{i \neq [\omega(m)]_j + 1,\, j=0,\ldots,k-1}} d(i,i-1)\right)\left(\prod_{i=1}^{k(\omega)} d([\omega(m)]_{i-1} + 1, [\omega(m)]_i)\right),$$

where the following notation is used:

$\sum_{\omega(m)}$ denotes summation over all partitions $(m_1, \ldots, m_k)$ of m where the m_i are positive integers such that $m_1 + \ldots + m_k = m$ and there are $k(\omega)$ integers in the partition (the order of the partition is relevant); $[\omega(m)]_0 = 0$,

$[\omega(m)]_i = m_1 + \ldots + m_i$, $i = 1,\ldots,k(\omega)$. Since $d(i,i-1) = 1$, $i = 2,\ldots,m$, from (2.2.9) and (2.2.10) it follows that

$$(2.2.11)\quad X^*(\chi_m(u,v)) = \sum_{\omega(m)} (-1)^{m-k(\omega)} \prod_{i=1}^{k(\omega)} X^*(\mathcal{L}([\omega(m)]_{i-1} + 1, [\omega(m)]_i)),$$

and it is this statement of the lemma which is most easily proved.

We introduce the subsets $S(i(1),\ldots,i(k))$ for $1 \le i(1) <\ldots< i(k) \le m - 1$ defined by $S(i(1),\ldots,i(k)) = \{(x_1,\ldots,x_m) \in Z^m : u_i \le x_i \le v_i \quad i = 1,\ldots,m;$ $x_{i(\ell)} \le x_{i(\ell)+1}$, $\ell = 1,\ldots,k$ and otherwise $x_i > x_{i+1}\}$ and the subsets $T(i(1),\ldots,i(k-1))$ for $1 \le i(1)<\ldots< i(k-1) \le m-1$ defined by $T(i(1),\ldots,i(k-1)) = \{(x_1,\ldots,x_m) \in Z^m,$ $(u_i \le x_i \le v_i,\ i = 1,\ldots,m) : v_{i(\ell)+1} \ge x_{i(\ell)+1} >\ldots> x_{i(\ell+1)}$ $\ge u_{i(\ell+1)}$, $\ell = 0,\ldots, k-1$, $i(0) = 0$, $i(k) = m\}$. The bracketed condition, $u_i \le x_i \le v_i$, is automatically satisfied in view of the rest of the definition of $T(i(1),\ldots,i(k-1))$. Notice that different sets $S(i(1),\ldots,i(k))$ are disjoint.

Let p,q denote subsets of $r = \{1,\ldots,m-1\}$ where $p = \{p(1),\ldots,p(\ell)\}$, $p(1) <\ldots< p(\ell)$ and similarly for q. Then, applying Lemma 2-2-3, we get

$$X^*(\chi_m(u,v)) = X^*(S(1,\ldots,m-1))$$

$$= \sum_{p:p\subset r} (-1)^{m-1-|p|} \sum_{q:q\subset p} X^*(S(q(1),\ldots,q(\ell))$$

$$(2.2.12) \qquad = \sum_{k=1}^{m} \sum_{p:|p|=k-1} (-1)^{m-k} X^*(T(p(1),\ldots,p(k-1)).$$

It is because $T(p(1),\ldots,p(k-1)) = \bigcup_{q\subset p} S(q(1),\ldots,q(\ell))$ that (2.2.12) follows. To establish this fact, first suppose that $(x_1,\ldots,x_m) \in T(p(1),\ldots,p(k-1))$, then certainly it cannot satisfy $x_i \le x_{i+1}$ for $i = 1,\ldots,p(1)-1,p(1)+1,\ldots, p(k-1)-1, p(k-1)+1,\ldots,m-1$, so that $(x_1,\ldots,x_m) \in S(q(1),\ldots, q(\ell))$ where $\{q(1),\ldots,q(\ell)\} \subset \{p(1),\ldots,p(k-1)\}$. Hence $T(p(1),\ldots,p(k-1)) \subset \bigcup_{q\subset p} S(q(1),\ldots,q(\ell))$. To prove that $\bigcup_{q\subset p} S(q(1),\ldots,q(\ell)) \subset T(p(1),\ldots,p(k-1))$, suppose that $(x_1,\ldots,x_m) \in S(q(1),\ldots,q(\ell))$, where $q \subset p$. Then $x_{q(i)} \le x_{q(i)+1}$ for $i = 1,\ldots,\ell$, but otherwise $x_i > x_{i+1}$. Certainly therefore $x_i > x_{i+1}$ for $i \ne p(1),\ldots,p(k-1)$ since $q \subset p$. This implies that

$$v_{p(i)+1} \ge x_{p(i)+1} > \ldots > x_{p(i+1)} \ge u_{p(i+1)}, \; i = 0,\ldots,k-1$$

since $(x_1,\ldots,x_m)$ satisfies $u_j \le x_j \le v_j$, $j = 1,\ldots,m$. Consequently $(x_1,\ldots,x_m) \in T(p(1),\ldots,p(k-1))$ and the required result is established.

There exists an obvious one-one correspondence between the vectors $(p(1),\ldots,p(k-1))$ and $(m_1,\ldots,m_k)$, the partitions of m into k integers, obtained by putting $[\omega(m)]_i = p(i)$,

$i = 1,\ldots,k-1$ and $[\omega(m)]_k = m$. Thus (2.2.12) may be written as

$$(2.2.13) \quad X^*(\chi_m(u,v)) = \sum_{\omega(m)} (-1)^{m-k(\omega)} X^*(T([\omega(m)]_1,\ldots,[\omega(m)]_{k-1})).$$

This last expression is valid regardless of the properties of X^*. Since $T([\omega(m)]_1,\ldots,[\omega(m)]_{k-1}) = \mathcal{L}(1,[\omega(m)]_1) \times\ldots\times \mathcal{L}([\omega(m)]_{k-1} + 1,m))$, if X^* is multiplicative, we obtain (2.2.11) and the lemma is proved.

It is only in the last step of the proof of Lemma 2-2-4 that X^* *is required to be multiplicative.* This is a rather vague requirement and in the present situation probably the most interesting cases occur when X^* is the generating function for the values obtained from some function f, defined on Z^m, when it acts on a finite subset of Z^m. Suppose that $f(x_1,\ldots,x_m) = f_1(x_1) + \ldots + f_1(x_m)$, where f_1 is a function mapping Z into itself. Then X^* defined as the generating function (over finite subsets) for f is multiplicative and Lemma 2-2-4 applies.

In order to obtain Theorem 2-2-2, *the lemma* is applied to the case where X^* is the generating function for $f(x_1,\ldots,x_m) = x_1 + \ldots + x_m$. Notice that for integers x_k and $i \le j$,

$$\sum_{k=i}^{j} x_k = Q, \quad u_j \le x_j < \ldots < x_i \le v_i$$

if and only if

$$\sum_{k=i}^{j} z_k = Q - (j-i+1)(u_j-1),\ 1 \le z_j < \dots < z_i \le v_i - u_j + 1,$$

where $z_k = x_k - u_j + 1$, $k = i,\dots,j$. Consequently,

$$(2.2.14)\quad X^*(\mathcal{L}(i,j)) = \sum_k \pi(j-i+1, v_i-u_j-j+i, k - (j-i+1)(u_j-1))z^k.$$

But putting $\ell = k - (j-i+1)(u_j-1)$, it follows from (2.2.14) that

$$X^*(\mathcal{L}(i,j)) = z^{(j-i+1)(u_j-1)} \sum_{\ell=-(j-i+1)(u_j-1)} \pi(j-i+1, v_i-u_j-j+i, \ell)z^{\ell}$$

$$= z^{(j-i+1)(u_j-1)} \sum_{\ell=0}^{\infty} \pi(j-i+1, v_i-u_j-j+i, \ell)z^{\ell}$$

$$(2.2.15)\qquad = z^{(j-i+1)(u_j-1)}\ \pi^*(j-i+1, v_i-u_j-j+i, z),$$

since $(j-i+1)(1-u_j) \le j - i + 1$ $(u_j \ge 0)$ and $\pi(j-i+1,n,\ell) = 0$ for $\ell < j - i + 1$. If we substitute for u_i and v_i in (2.2.15) in terms of b_i and c_i we obtain Theorem 2-2-2.

Using (2.2.5) and (2.2.11), N_m^* can be expressed as

$$(2.2.16)\quad N_m^*(u,v,z) = \sum_{\omega(m)} (-1)^{m-k(\omega)} z^{\left(\sum_{i=1}^{k(\omega)} m_i b_{[\omega(m)]_i} - \sum_{i \le j} m_i m_j\right)} \times$$

$$\prod_{i=1}^{k(\omega)} \pi^*(m_i, c_{[\omega(m)]_{i-1}} + 1 - b_{[\omega(m)]_i} - 2, z).$$

Now (2.2.6) can be used to obtain an explicit expression for the numbers $W(b,c,t)$ in terms of the π's. From (2.2.6) it follows that $W(b,c,t)$ is the coefficient of $z^{(t-\frac{1}{2}m(m+1))}$

in the right hand side of (2.2.16). Thus,

$$(2.2.17)\quad W(b,c,t) = \sum_{\omega(m)} \sum_{\omega_k(a)} (-1)^{m-k(\omega(m))} \prod_{j=1}^{k(\omega(m))} \pi(m_j,$$

$$c_{[\omega(m)]_{j-1}} + 1 - b_{[\omega(m)]_j} - 2, i_j),$$

where $\omega_k(a)$ is the number of partitions $(i_1,\ldots,i_{k(\omega(m))})$ of $a = t - \frac{1}{2}m(m+1) + \sum_{i\leq j} m_i m_j - \sum_{i=1}^{k} m_i b_{[\omega(m)]_i}$. Noting that $m^2 = 2\sum_{i<j} m_i m_j + \sum_{i=1}^{k} m_i^2$, we can write $a = t - {}^{m}/_{2} + (\sum_{i=1}^{k} m_i^2)/2 - \sum_{i=1}^{k} m_i b_{[\omega(m)]_i}$.

An interesting recurrence formula for N_m^* can be obtained by expanding the determinant in Theorem 2-2-2 by the m^{th} column. In view of the special form of $(d(i,j))$, the following result is obtained:

$$(2.2.18)\quad N_m^*(u,v,z) = \sum_{k=1}^{m} (-1)^{k+1} N_{m-k}^*(u,v,z) \times$$

$$\pi^*(k, v_{m-k+1} - u_m - k + 1, z) z^{k(u_m-1)},$$

or equivalently,

$$(2.2.19)\quad N_m(u,v,w) = \sum_{k=1}^{m} (-1)^{k+1} \sum_{i=1}^{w} N_{m-k}(u,v,w-i) \times$$

$$\pi(k, v_{m-k+1} - u_m - k + 1, i - k(u_m-1)),$$

where $N_o(u,v,w) = 0$ if $w > 0$, $N_o(u,v,w) = 1$ if $w = 0$, so that $N_o^*(u,v,z) = 1$.

In order to apply (2.2.2) to obtaining the distribution of $\tilde{\xi}_N$, the results about $W(b,c,t)$ have to be expressed in terms of probabilities. This is achieved by dividing by $\binom{N}{m}$. The values of b_i and c_i depend upon t and e and are given by

$$b_i = \max([(iN-e+t)/m], i-1)$$
$$c_i = \min(<(e+t - (m+1-i)N)/m>, n+i+1), i = 1,\ldots,m.$$

Discussion of the null distribution of $\tilde{\xi}_{N1}$ follows much the same lines as that for $\tilde{\xi}_N$ and in particular the numbers $W(b,c,t)$ can be used by setting $c_i = n + i + 1$, $i = 1,\ldots,m$.

This section is concluded by mention of a recurrence relation for $W(b,c,t)$. This relationship together with (2.2.2) can be used to obtain the value of $P(\tilde{\xi}_N > e)$. Exact tail probabilities are tabulated in Appendix 2.

Suppose that $W(b,c,t)$ is now written in more detail as $W(m,n,b,c,t)$. Then if the first term R_1 of the partition is fixed ($b_1+1 \le R_1 \le c_1-1$), the number of partitions of t obeying the required conditions may be found as the number of vectors $(S_2,\ldots,S_m)$ for which

(a) S_i are integers, $1 \le S_2 < \ldots < S_m \le N - R_1$,

(b) $b_i - R_1 < S_i < c_i - R_1 \quad (i = 2,\ldots,m)$,

(c) $\sum_{i=2}^{m} S_i = t - mR_1$,

by putting $S_i = R_i - R_1$. This number is just $W(m-1, n-R_1+1, b'(R_1), c'(R_1), t-mR_1)$ where $b'(R_1) = (b_2-R_1,\ldots,b_m-R_1)$ and $c'(R_1) = (c_2-R_1,\ldots,c_m-R_1)$. Then it follows that

$$(2.2.20)\quad W(m,n,b,c,t) = \sum_{R_1=b_1+1}^{c_1-1} W(m-1,n-R_1+1,b'(R_1),c'(R_1),\\ t-mR_1),$$

which enables us to generate the numbers $W(m,n,b,c,t)$ from the values of $W(1,n,b,c,t)$, which are obvious.

2.3 Aspects of the performance of the test

This section will in many aspects be a survey of results which will be derived in chapter 3. They are presented here, though, since they form part of the evidence for regarding the Mann-Whitney type tests in a favourable light.

(i) The consistency of the tests with critical regions $\{r : \xi_N(r) > t_\alpha\}$ and $\{r : \xi_{N1}(r) > t'_\alpha\}$ against the general alternative $K : F \neq G$ follows immediately from its derivation as a union-intersection test. The detailed formulation of this is given in section 3.1, but the description of the test statistics at the beginning of the chapter indicate this aspect of the test fairly clearly. A result of Nandi (1965) ensures the consistency of a union-intersection test from the consistency of its component tests.

(ii) The details required for the computation of Bahadur efficiency are discussed in section 3.4. For the present it suffices to know that $e(T_N^{(1)}, T_N^{(2)})$ is used to denote the Bahadur efficiency of two sequences of tests $\{T_N^{(1)}\},\{T_N^{(2)}\}$ - satisfying certain conditions - against some specified alternative hypothesis. Bahadur efficiency is regarded as a suitable method of asymptotic comparison between two tests.

In choosing a set of alternative hypotheses, we thought it would be interesting to take circular analogs of hypotheses for which the usual Mann-Whitney test performs well. Lehmann(1953, pp.34-35) has shown that for the set of alternative hypotheses that $F = pG + qG^2$, $0 \le p < 1$, $p + q = 1$, the locally most powerful rank test for H against K is W_N (that is W_N is most powerful for H against members of K for which $0 < q \le \varepsilon$ for some $\varepsilon > 0$).

Consider the two sets of alternatives

$$(2.3.1) \qquad K_{c,p} : F_c(x) = pG_c(x) + q(G_c(x))^2$$

and

$$K^-_{c,p} : F^-_c(x) = pG^-_c(x) + q(G^-_c(x))^2,$$

$0 \le c < 2\pi$, $0 \le p < 1$, $p + q = 1$. These alternatives then provide circular analogs to the alternative that $F = pG + qG^2$, obtained by changing the cut-off point and the direction of measurement of angular displacement. The invariance of ξ_N under these different choices makes it suitable for testing H against any of the members of $K_{c,p}$ or $K^-_{c,p}$.

The performance of ξ_N for $K_{c,p}$ and $K^-_{c,p}$ is compared with that of another test suitable for circular data (satisfying the invariance criterion), defined by

(2.3.2) $$V_{m,n} = D^+_{m,n} + D^-_{m,n},$$

so called Kuiper's test. As indicated in section 3.1, this test is a circular analog of the Smirnov test $D_{m,n}$ (or also of $D^+_{m,n}$). Then it is shown that

(2.3.3) $$\lim_{q\to 0} e(\xi_N, V_{m,n}) = {}^4/_3$$

where the limiting operation *means* that a subset of the alternatives $K_{c,p}$ or $K^-_{c,p}$ for fixed c and variable q is considered and the limiting value of the efficiencies as q *tends to* 0 - so that the alternative hypothesis approaches the null hypothesis - is ${}^4/_3$ (independently of c). It also follows that $\lim_{q\to 0} e(\xi_{N1}, V_{m,n}) = {}^4/_3$ for the alternatives $K(c,p)$.

(iii) Suppose that a lower bound is required for the power of ξ_{N1} against some fixed alternative *hypothesis*. Define $W_N(c)$ as the value of W_N calculated with c as the cut-off point $(0 \le c < 2\pi)$. Then clearly

$$P(\xi_{N1} > r|K) \ge P(W_N(c) > r|K).$$

In order to make the best use of this lower bound, an idea used by Massey (1950) in order to obtain a lower bound for the Kolmogorov-Smirnov statistics is adapted to the present situation. Suppose that $\gamma \in [0,2\pi)$ is a value for which $\int_S G_\gamma(s)dF_\gamma(s)$ attains its maximum, then the general lower bound will be used *for the case* $c = \gamma$.

The limiting distribution of $W_N(\gamma)$ is normal with mean $\mu_N(\gamma)$ and variance $\sigma_N^2(\gamma)$, where the values of these two parameters can be obtained directly from van Dantzig (1951) as:

$$\mu_N(\gamma) = 2mn\,\alpha(\gamma) + \tfrac{1}{2}\{(m-n)(N+1) - 2mn\}$$

$$\sigma_N^2(\gamma) = 4mn\{(m-1)\sigma^2_{F(y)}(\gamma) + (n-1)\sigma^2_{G(x)}(\gamma) + \alpha(\gamma)(1-\alpha(\gamma))\},$$

where

$$\alpha(\gamma) = \int_S G_\gamma(s)\,dF_\gamma(s),\ \sigma^2_{F(y)}(\gamma) = \int\{F_\gamma(y) - 1 + \alpha(\gamma)\}^2 dG_\gamma(y)$$

and

$$\sigma^2_{G(x)}(\gamma) = \int\{G_\gamma(x) - \alpha(\gamma)\}^2 dF_\gamma(x).$$

This means that, provided m,n are large, an approximate lower bound for the power of the test based on ξ_{N1} is given by

$$\int_{(t_\alpha-\mu_N(\gamma))/\sigma_N(\gamma)} (2\pi)^{-\frac{1}{2}} \exp(-{}^{x^2}/{}_2)\,dx,$$

where $P(\xi_{N1} > t_\alpha | H) = \alpha$. For the limiting normality of $F_N(\gamma)$ it is also required to have $\int_S G_\gamma(s)\,dF_\gamma(s) < 1$.

The details of a lower bound for the power of a test based on ξ_N are similar except that the possibility of a different direction of measurement is also included.

2.4 A distribution-free confidence interval

Consider the following measure of distance between two distributions F and G:

$$(2.4.1) \qquad \tau(F,G) = \sup_{\gamma\in[0,2\pi)} \{\int_S G_\gamma(s)dF_\gamma(s)\}.$$

An identity rather like (2.1.11) then gives

$$(2.4.2) \quad \tau(F,G) = \int_0^{2\pi} G(s)dF(s) + \sup_{0\le x<2\pi} \{F(x) - G(x)\}.$$

If F and G are continuous, $\tau(F,G) \ge \frac{1}{2}$ and $\tau(F,G) = \frac{1}{2}$ if and only if $F = G$.

The possibilities in using τ as a measure of distance can be seen from the preceding paragraph. The probability that a random variable with distribution $G(y)$ is less than a random variable with distribution $F(x)$ when angular displacement is measured from γ is given by $\int_S G_\gamma(s)dF_\gamma(s)$. The maximum of such values over γ is given by τ. Then (2.4.2) shows that τ has a relationship with another measure of distance $\sup_x\{F(x) - G(x)\}$, but whereas this last measure is exclusively pointwise, τ also takes into account the 'global' relationship between F and G. These considerations lead us to construct a confidence interval for τ.

It is well-known (van Dantzig, 1951) that $U_n(\gamma) = \int_S G_{n\gamma}(s)dF_{m\gamma}(s)$, $0 \le \gamma < 2\pi$ gives an unbiased, consistent estimator of $\int_S G_\gamma(s)dF_\gamma(s)$. This suggests

$$\hat{\tau} = \sup_{\gamma\in[0,2\pi)} \{U_n(\gamma)\}$$

as a possible estimator for $\tau(F,G)$. Of course, $\hat{\tau}$ is just a rescaled version of ξ_{N1}. An idea used by Birnbaum (1955) to construct confidence intervals for $\int G(s)dF(s)$ is now adapted to give a one-sided confidence interval for $\tau(F,G)$.

Note that

$$\tau - \hat{\tau} = \sup_{\gamma\in[0,2\pi)} \left\{\int G_\gamma(s)dF_\gamma(s)\right\} - \sup_{\gamma\in[0,2\pi)} \left\{\int G_{n\gamma}(s)dF_{m\gamma}(s)\right\}$$

$$\le \sup_\gamma\left\{\int G_\gamma(s)dF_\gamma(s) - \int G_{n\gamma}(s)dF_{m\gamma}(s)\right\}$$

$$= \sup_\gamma\left\{\int (G_\gamma(s) - G_{n\gamma}(s))dF_\gamma(s) + \int G_{n\gamma}(s)d(F_\gamma(s) - F_{m\gamma}(s))\right\}$$

$$\le \sup_\gamma \sup_s\{F_\gamma(s) - F_{m\gamma}(s)\} + \sup_\gamma \sup_s\{G_\gamma(s) - G_{n\gamma}(s)\}$$

$$= V_{1m} + V_{2n},$$

where

$$V_{1m} = \sup_\gamma \sup_s\{F_\gamma(s) - F_{m\gamma}(s)\}$$

$$V_{2n} = \sup_\gamma \sup_s\{G_\gamma(s) - G_{n\gamma}(s)\}.$$

Although here V_{1m} and V_{2n} are not written in the usual form, they are in fact test statistics for the one-sample Kuiper test distributed under the null hypothesis. Further discussion of this is given in section 3.1.

Discussion of the confidence interval now rests upon the convolution of the independent statistics V_{1m} and V_{2n}, since

$$P(\tau \leq \hat{\tau} + \varepsilon) \geq P(V_{1m} + V_{2n} \leq \varepsilon).$$

There are problems with this convolution and some form of approximation is required (possibly along the lines of Birnbaum (1955)).

In fairness, it should be said that this confidence interval is more of theoretical interest than suited for practical applications. Criticisms applying to Birnbaum's technique apply here. The inequalities employed in the derivation of the result are crude. It does have the advantage, though, of being independent of assumptions on F and G. This is probably more important here than in Birnbaum's case since any alternative approach would rely upon the limiting distribution of ξ_{N1} under general alternatives, which does not seem easily obtainable and will require some restrictions on F and G. There is also fairly extensive tabulation of the null distribution of the Kuiper statistics (although for small sample sizes the resulting confidence interval might be rather uninformative).

A two-sided confidence interval for τ can also be derived, again using the Kuiper statistic.

CHAPTER 3. A General Class of Circular Rank Tests

3.1 Introduction to the tests as an application of the method of union-intersection.

The method of union-intersection as described in Roy (1953) has been applied to deriving tests for composite hypotheses in many different situations, particularly in multivariate analysis. The resulting tests often possess desirable properties - again this appears to be particularly true in multivariate analysis - although there does not appear to have been much progress in deriving general properties for these tests. The only properties which hold in general for these tests seem to be either obvious or to require restrictive independence assumptions for the component test statistics (for example Nandi (1965)).

The successful performance of union-intersection tests suggests their application to nonparametric problems. Possibly the only specific example of such a procedure is by Behnen (1975); unfortunately, in that case the resulting test has little to recommend it. In this section we shall introduce a class of tests which can be motivated by a union-intersection argument. It should be stressed that no beneficial consequences result immediately from doing this (beyond consistency properties for the tests). Rather the formulation suggests that further analysis of these tests might be worthwhile.

In order to construct a Type I test using Roy's method, the subset of Γ(Γ defined in section 1.3) determined by the alternative hypothesis K is expressed as the union of subsets of Γ each of which defines an alternative hypothesis of a suitable type. Basically the component alternative hypotheses are required to be such that we already have adequate tests for testing H against them. In order to achieve such a decomposition of the general alternative K, the idea of a stochastic ordering for random variables on C is introduced.

We shall say that the random variable X (distribution F) is stochastically larger than the random variable Y (distribution G) on C if there exists $\gamma \in [0,2\pi)$ such that

$$(3.1.1) \qquad F(x) - F(\gamma) \le G(x) - G(\gamma) \quad \text{for } x \in [0,2\pi)$$

with strict inequality for at least one x (F,G continuous). Condition (3.1.1) expresses the condition that $p_\gamma(X)$ is stochastically larger than $p_\gamma(Y)$ in the usual sense. If K denotes the hyp that X is stochastically larger than Y on C, the

$$(3.1.2) \qquad K = \bigcup_{\gamma\in[0,2\pi)} K(\gamma),$$

where $K(\gamma)$ denotes the hypothesis that (3.1.1) holds for γ fixed.

The most interesting part of our construction is the *simple observation* that the hypothesis that X is stochastically larger than Y on C is equivalent to $K : F \neq G$. Identifying subsets of Γ with the hypotheses they index, if $(F,G) \in K$, then $E(x) = F(x) - G(x) \neq 0$ for some x and must attain its maximum on the closed interval $[0,2\pi]$ But since $E(0) = E(2\pi) = 0$, there exists $\gamma \in [0,2\pi)$ such that $E(x) \leq E(\gamma)$ for $x \in [0,2\pi)$ as required for (3.1.1). Thus K implies X stochastically larger than Y on C, while the reverse is obviously true. The decomposition in (3.1.2) is therefore true for $K : F \neq G$ and enables us to introduce fairly natural tests for H against K.

Suppose that there is a 'good' rank test for the hypothesis of *randomness* against the alternative that the observations of the first sample are stochastically larger (in the usual sense) than the observations of the second sample (a suitable example of such a test is the Mann-Whitney test). This test has as its test statistic $T_N(R)$. In order to test for H against $K(\gamma)$, for any $\gamma \in [0,2\pi)$, the test statistic T_N may be used, where γ is the cut-off point (affecting the ranking of the combined sample). The critical region of size β for this test will be denoted by $\omega(H,K(\gamma),\beta)$.

The Roy Type I test for H against K is obtained by taking as critical region the union of the critical regions of the component tests for the component hypotheses. In this case the critical region for the Type I test for H against K is

$$\omega(H,K) = \bigcup_{\gamma\in[0,2\pi)} \omega(H,K(\gamma),\beta),$$

with the size of the test obtained from $P(R \in \omega(H,K)|H) = \alpha(H,\beta) > \beta$. If, as is usual, $\omega(H,K(0),\beta) = \{r \in R : T_N(r) > t_\beta\}$, then the test with critical region $\omega(H,K)$ is also a Type II test for H against K and in fact the critical region may be written

$$\omega(H,K) = \{r \in R : T_N(g(r)) > t_\beta \quad \text{for some } g \in G_1\},$$

so that the test is invariant under $\mathcal{T}_1$. Alternatively, we may write

$$(3.1.3) \qquad \omega(H,K) = \{r \in R : \max_{g \in G_1} \{T_N(g(r))\} > t'_\alpha\},$$

where $P(\max_{g \in G_1} \{T_N(g(R))\} > t'_\alpha | H) = \alpha(H,\beta)$. It is in the form (3.1.3) that the tests will be discussed.

The construction of a test invariant under G can be similarly rationalized. *Suppose that for a fixed* $\gamma \in [0,2\pi)$, the alternative hypothesis is that either $K(\gamma)$ holds or that

$$F(x) - F(\gamma) \geq G(x) - G(\gamma) \qquad (x \in [0,2\pi))$$

with strict inequality for at least one x; the two possibilities to be collectively denoted by $\bar{K}(\gamma)$. Then the general alternative can be written

$$K = \bigcup_{\gamma \in [0,2\pi)} \bar{K}(\gamma).$$

To test against the second possibility in $\bar{K}(\gamma)$, T_N can be used with the combined sample ranked clockwise from γ. The resulting test is described by replacing G_1 by G in the preceding paragraph to obtain a test invariant under T.

The right hand side of (3.1.3) suggests a natural way of constructing alternative hypotheses for circular data from standard alternative hypotheses for rank tests. Consider, for instance, the situation where U is a continuous, strictly increasing distribution function over $[0,1]$, then a possible alternative hypothesis is that $G(x) = U(F(x))$ for all x. Then as the circular analog we could take the following: there exists $\gamma \in [0,2\pi)$ such that either

$$(3.1.4) \quad G_\gamma(x) = U(F_\gamma(x)) \quad \text{or} \quad G_\gamma^-(x) = U(F_\gamma^-(x)) \quad (x \in [0,2\pi)),$$

so that the alternative hypothesis does not depend on the choice of cut-off point or direction of measurement. This procedure is quite natural, since it expresses the idea that for some frame of measurement a particular, standard alternative hypothesis holds. The alternative hypothesis defined by (2.3.1) is of this type.

The circular alternative hypotheses that have just been described lend themselves to union-intersection tests, since they can be expressed as a union of hypotheses along the lines of (3.1.2). This of course also applies to more general situations than the two-sample problem. The same kind of discussion as that pursued above shows that the tests with test statistics

$$(3.1.5) \qquad \nu_N(R) = \max_{g \in G} \{S_N(g(R))\},$$

where $S_N(R)$ are the linear rank statistics defined by (1.2.1), and rejecting H for large values of ν_N may be regarded as derived using the method of union-intersection. It is the limiting distributions of such statistics which will be one of our main concerns in this chapter.

As mentioned in section 2.1 testing problems for real-valued data can be projected onto *C*. Two points about this deserve comment. First, the set of alternatives we have called circular analogs is as a rule a much larger set of hypotheses than the original alternative hypotheses from which they were derived. *Secondly, the projection* may affect the form of the alternative hypothesis as is the case, for instance, when testing for location shift - that is the hypothesis in the two-sample problem that $G(x) = F(x+c)$, c a constant. This is due to the fact that θ^{-1} does not map $(0,2\pi)$ uniformly. Notice that hypotheses like $K : F \neq G$ and $K : G(x) = U(F(x))$ are not affected by *projection.*

This section is concluded with examples of the method of test construction we have just discussed.

Consider the following test for the two-sample problem. Using 0 as cut-off point, replace the observation with rank R_i by a point on *C* with angular displacement $2\pi(R_i-1)/N$. Then count the number of X's above the real axis excluding the point with angular displacement π. Rotate the axis through 2π and define the test statistic as the maximum number of X's on one side of the axis. Since only rotations through multiples of $2\pi/N$ need be considered, this test is the union-intersection test constructed from the median test which has its test statistic defined by

$$T_N = \sum_{i=1}^{[(N+1)/2]} Z_{Ni}$$

([x] here denotes the greatest integer less than or equal to x).

The description of the test statistic also draws attention to its relation with a test for the uniformity of a sample of circular data described by Ajne (1968) and it may be of interest to know that the test can also be derived from Ajne's test using the technique suggested by Beran (1969) (see also Mardia (1972), pp.204-205) for constructing two-sample tests from tests of uniformity. Incidently, Ajne's test may be regarded as a union-intersection test constructed from the sign test. (Ajne's test for uniformity: Suppose $X_1,\ldots,X_N$ are drawn independently from a circular distribution, then count the number of observations on each side of a line through the centre of the circle as it is rotated through π. The test statistic is the maximum such number.)

Clearly the Mann-Whitney type tests of section 2.1 are union-intersection tests.

The one-sample test proposed by Kuiper (1960) for testing whether the independently drawn observations $X_1,\ldots,X_N$ come from a population with continuous distribution F(x) is defined by $V_N = D_N^+ + D_N^-$, where $D_N^+ = \sup_x \{F_N(x) - F(x)\}$ and $D_N^- = \sup_x \{F(x) - F_N(x)\}$.

If $D_N^+(x)$ is used to denote the value obtained for D_N^+ with x as the cut-off point, then Barr and Shudde (1973) have shown that

$$(3.1.6) \qquad V_N = \max_{x\in[0,2\pi)} \{D_N^+(x)\},$$

and hence $V_N = \max_{x\in[0,2\pi)} \{D_N(x)\}$, where $D_N(x) = \max\{D_N^+(x), D_N^-(x)\}$. They point out that V_N can be regarded as obtained through the method of union-intersection applied to the Kolmogorov test. In fact (3.1.6) is also implicit in the proof of the invariance of V_N given by Mardia (1972, p.174).

A similar result holds for the two-sample test statistic $V_{m,n}$ defined by (2.3.2), which may be derived from $D_{m,n}^+$, the one-sided Smirnov statistic, or from $D_{m,n} = \max\{D_{m,m}^+, D_{m,n}^-\}$, the two-sided statistic.

3.2 Convergence of probability measures: definitions and background material

*This section is basically a collection of material re-*quired for reading section 3.3. The results of this section are well-known in that they are readily available in published form. Lemmas and theorems are given without proof. In each case before the statement of the result, a reference is given to where the reader may find a proof of the result. Most of the items in this section are drawn from the books by Billingsley (1968) and by Hájek and Šidák (1967). These two references will be abbreviated to B and HS respectively for this section.

Suppose that P_ν and P are probability measures on the class $\mathcal{S}$ of Borel sets in a metric space (S,ρ). If $\int_S f\,dP_\nu \to \int_S f\,dP$ for every bounded continuous real function f on S, then P_ν is said to converge weakly to P, written $P_\nu \Rightarrow P$. If X_ν is a measureable mapping from a probability space $(\Omega,\mathcal{B},Q)$ into S, then we say that X_ν converges in distribution to the random element X, written $X_\nu \to_{\mathcal{D}} X$, if the distributions P_ν of the X_ν converge weakly to the distribution P of X. Equivalently, $X_\nu \to_{\mathcal{D}} X$ if $E\{f(X_\nu)\} \to E\{f(X)\}$ for each bounded continuous real function f on S. The presentation adopted here is in terms of probability measures rather than random elements.

The following two general results prove to be extremely useful in many cases.

Theorem 3-2-1 (B p.25). Suppose that S is a separable metric space and that X_ν, Y_ν have common domain and range. If $X_\nu \to_{\mathcal{D}} X$ and $Q\{\rho(X_\nu,Y_\nu) \geq \varepsilon\} \to 0$ for each positive ε, then $Y_\nu \to_{\mathcal{D}} X$.

Theorem 3-2-2 (B pp.29-31). If h is a continuous mapping from S into another metric space S', then if

$$P_\nu \Rightarrow P,$$

it follows that

$$P_\nu h^{-1} \Rightarrow Ph^{-1}.$$

(Equivalently: if $X_\nu \to_{\mathcal{D}} X$, then $h(X_\nu) \to_{\mathcal{D}} h(X)$).

Two separable metric spaces are of particular concern to us. Let $C = C([0,1])$ denote the space of continuous functionals on $[0,1]$ with the metric defined by $\rho(x,y) = \sup_{0\le t\le 1} \{|x(t) - y(t)|\}$ for two elements x,y of C. The space C is unsuitable for a complete discussion of the random functions we are concerned with and thus the space $D = D([0,1])$ of functionals on $[0,1]$ that are right continuous with left-hand limits is introduced. Clearly D contains C. The metric for D is introduced next.

Let Λ denote the class of strictly increasing, continuous mappings of $[0,1]$ onto itself. If x,y are in D, then $\tau(x,y)$ is defined to be the infimum of those positive ε for which there exists a $\lambda \in \Lambda$ such that

$$\sup_{0\le t\le 1} \{|\lambda(t) - t|\} \le \varepsilon$$

and

$$\sup_{0\le t\le 1} \{|x(t) - y(\lambda(t))|\} \le \varepsilon.$$

Then (D,τ) is a metric space and the resulting topology is called the Skorohod topology.

The standard approach to determining the weak convergence of a sequence of probability measures on D involves two important ideas: finite-dimensional distributions and the tightness of a sequence of measures.

A sequence $\{P_\nu\}$ of probability measures on the metric space S is said to be tight, if for every positive ε there exists a compact set K such that $P_\nu(K) > 1 - \varepsilon$ for all P_ν. If X_ν are random elements of S, the sequence $\{X_\nu\}$ is said to be tight when $\{P_\nu\}$ is tight, where P_ν is the distribution of X_ν.

If $t(1),\ldots,t(k)$ are points in $[0,1]$, the natural projection from D to R^k is defined by

$$\pi_{t(1),\ldots,t(k)}(x) = (x(t(1)),\ldots,x(t(k)))$$

for $x \in D$. Suppose that P is a probability measure on D, using the Skorohod topology to define the Borel sets. Define T_P as those $t \in [0,1]$ for which π_t is continuous except at points forming a set of P-measure 0. Since π_0, π_1 are always continuous, $0,1 \in T_P$. If $0 < t < 1$, then $t \in T_P$ if and only if

$$P\{x : x(t-) \neq x(t)\} = 0,$$

where $x(t-) = \lim_{s \to t-} x(s)$, the left-hand limit at t.

The crucial result is the following:

<u>Theorem 3-2-3</u> (B pp.124-125). If $\{P_\nu\}$ is tight and if $P_\nu \pi^{-1}_{t(1),\ldots,t(k)} \Rightarrow P\pi^{-1}_{t(1),\ldots,t(k)}$ holds for $t(1),\ldots,t(k) \in T_P$, then $P_\nu \Rightarrow P$.

There are several different possible approaches to determining whether a sequence of probability measures is tight. To describe those which prove useful for section 3.3 requires the introduction of the idea of a modulus of continuity. Define

$$(3.2.1) \qquad w(x,\delta) = \sup_{|s-t|<\delta} \{|x(s) - x(t)|\}.$$

This modulus is particularly appropriate for C. Another useful measurement is provided by $w(x,T) = \sup_{s,t\in T} \{|x(s)-x(t)|\}$, where $T \subset [0,1]$. For use with elements of D, define

$$(3.2.2) \quad w'(x,\delta) = \sup \min\{|x(t) - x(t_1)|, |x(t_2) - x(t)|\},$$

where the supremum extends over t_1, t, t_2 satisfying $t_1 \le t \le t_2$ and $t_2 - t_1 \le \delta$.

If $\{P_\nu\}$ is a sequence of probability measures on C, then the following result is appropriate.

Theorem 3-2-4 (B p.55). The sequence $\{P_\nu\}$ is tight if and only if these two conditions hold:

(i) For each positive η, there exists an a such that

$$P_\nu\{x : |x(0)| > a\} \le \eta, \quad \nu \ge 1.$$

(ii) For each positive ε and η, there exists a δ, $0 < \delta < 1$, and an integer ν_0 such that

$$P_\nu\{x : w(x,\delta) \ge \varepsilon\} \le \eta \quad , \quad \nu \ge \nu_0.$$

The more general situation where $\{P_\nu\}$ is a sequence of probability measures on D is covered by the next result.

<u>Theorem 3-2-5</u> (B p.125) The sequence $\{P_\nu\}$ is tight if and only if these two conditions hold:

(i) For each positive η, there exists an a such that

$$P_\nu\{x : \sup_t |x(t)| > a\} \le \eta \;, \quad \nu \ge 1.$$

(ii) For each positive ε and η, there exist a δ, $0 < \delta < 1$, and an integer ν_0 such that

$$P_\nu\{x : w'(x,\delta) \ge \varepsilon\} \le \eta,\ P_\nu\{x : w(x,[0,\delta)) \ge \varepsilon\} \le \eta \ \text{ and}$$
$$P_\nu\{x : w(x,[1-\delta,1)) \ge \varepsilon\} \le \eta \ \text{ for } \ \nu \ge \nu_0.$$

The following two theorems supply sufficient conditions for tightness.

<u>Theorem 3-2-6</u> (B pp.126-127). Suppose that $P_\nu \pi^{-1}_{t(1),\ldots,t(k)} \Rightarrow P\pi^{-1}_{t(1),\ldots,t(k)}$ holds whenever $t(1),\ldots,$ $t(k)$ all lie in T_P. Suppose that $P\{x : x(1) \ne x(1-)\} = 0$. Suppose finally that, for each positive ε and η, there exist a δ, $0 < \delta < 1$, and an integer ν_0 such that

$$P_\nu\{x : w'(x,\delta) \ge \varepsilon\} \le \eta, \quad \nu \ge \nu_0.$$

Then

$$P_\nu \Rightarrow P.$$

<u>Theorem 3-2-7</u> (B pp.127-128). Suppose that, for each positive η, there exists an a such that

$$P_\nu\{x : |x(0)| > a\} \le \eta, \qquad \nu \ge 1.$$

Suppose further that, for each positive ε and η, there exist a δ, $0 < \delta < 1$, and an integer ν_0 *such that*

$$(3.2.3) \qquad P_\nu\{x : w(x,\delta) \ge \varepsilon\} \le \eta\ , \qquad \nu \ge \nu_0.$$

Then $\{P_\nu\}$ is tight, and if P is the weak limit of a subsequence $\{P_{\nu'}\}$ then $P(C) = 1$.

Theorem 3-2-7 is usually used by verifying (3.2.3) with the *aid of the following lemma*.

Lemma 3-2-8 (B p.56). If $0 = t(0) < t(1) < \dots < t(r) = 1$, and if $t(i) - t(i-1) \ge \delta$, $2 \le i \le r-1$, then for positive ε

$$P\{x : w(x,\delta) \ge 3\varepsilon\} \le \sum_{i=1}^{r} P\{x : \sup_{t(i-1)\le s\le t(i)} \left|x(s)-x(t(i-1))\right| \ge \varepsilon\}.$$

In order to utilize Lemma 3-2-8, the following general result often proves useful. Let $\xi_1,\dots,\xi_m$ be random variables and put $S_k = \xi_1 + \dots + \xi_k$ ($S_0 = 0$) and $M_m = \max_{0\le k\le m} \{|S_k|\}$.

Theorem 3-2-9 (B p.94). Suppose that there exist non-negative numbers $u_1,\dots,u_m$ such that

$$P\{|S_j - S_i| \ge \lambda\} \le \lambda^{-\gamma}\Big(\sum_{i<\ell\le j} u_\ell\Big)^{\alpha}, \quad 0 \le i < j \le m$$

for $\gamma \ge 0$, $\alpha > 1$ and all positive λ. Then, for all positive λ,

$$P\{M_m \geq \lambda\} \leq K_{\gamma,\alpha}\, \lambda^{-\gamma} (u_1 + \ldots + u_m)^{\alpha},$$

where $K_{\gamma,\alpha}$ depends only on γ and α.

The second aspect requiring discussion if Theorem 3-2-3 is to be used to study convergence in distribution is the convergence of the finite-dimensional distributions. An important result is the following:

Theorem 3-2-10 (HS p.168). Let $X_\nu = (X_\nu(1),\ldots,X_\nu(k))$, $\nu \geq 1$ and $X = (X(1),\ldots,X(k))$ be k-dimensional random vectors, then $X_\nu \to_{\mathcal{D}} X$ if and only if $\sum_{j=1}^{k} \lambda_j X_\nu(j) \to_{\mathcal{D}} \sum_{j=1}^{k} \lambda_j X(j)$ for every real vector $(\lambda_1,\ldots,\lambda_k)$.

A brief word of explanation is appropriate at this juncture as to precisely what is meant when speaking of the limiting distribution of $\nu_N(R)$ (as defined by (3.1.5)). In view of the relationship between ν_N and S_N (defined by (1.2.1)), it is sufficient to discuss the latter. This will be done along the lines of p.152 of HS.

The statistics S_N are in fact indexed by the real vectors $(c_1,\ldots,c_N)$ and convergence statements concern sequences of statistics $\{S_{c_\nu}\}$, $c_\nu = (c_{\nu 1},\ldots,c_{\nu N(\nu)})$, and hold under conditions on $\{c_\nu\}$. The statistics S_{c_ν} are functions of the rank vectors $(R_1,\ldots,R_{N(\nu)})$. The explicit formulation of the sequences is surpressed by dropping the index ν (as HS does) except where confusion might otherwise arise.

The following notation is convenient: $\bar{a}_N = \frac{1}{N} \sum_{i=1}^{N} a_N(i)$ and $\bar{c} = \frac{1}{N} \sum_{i=1}^{N} c_i$.

Suppose that the scores $a_N(i)$ (in the definition of

S_N) satisfy the following condition: there exists a square integrable function $\phi(u)$, $\int_0^1 \{\phi(u) - \bar{\phi}\}^2 du > 0$ - where $\bar{\phi} = \int_0^1 \phi(u)du$ - such that

$$(3.2.4) \qquad \int_0^1 \{a_N(1 + [uN]) - \phi(u)\}^2 du \to 0 \text{ as } N \to \infty.$$

Furthermore, the statistics S_N are indexed by vectors $(c_1,\ldots,c_N)$ for which

$$(3.2.5) \qquad \sum_{i=1}^{N} (c_i-\bar{c})^2 \Big/ \max_{1\le i\le N} \{(c_i-c)^2\} \to \infty.$$

Then the following theorem is basic to our study of finite-dimensional distributions.

Theorem 3-2-11 (HS pp.163-164). Under H, (3.2.4) and (3.2.5), the statistics S_N are asymptotically normal (μ_c, σ_c^2) with

$$\mu_c = N\,\bar{a}_N\,\bar{c},$$

$$\sigma_c^2 = \{\sum_{i=1}^{N} (c_i-\bar{c})^2\} \int_0^1 \{\phi(u) - \bar{\phi}\}^2 du.$$

The following straightforward result about the moments of $S_N(R)$ is required in section 3.3. To make the statement of the result more manageable - without losing any generality - it is supposed that $\bar{a}_N = 0$, $\bar{c} = 0$ and $\sum_{i=1}^{N} c_i^2 = 1$.

Theorem 3-2-12 (HS p.61 and p.82). Under H,

$$(3.2.6) \qquad var(S_N) = (N-1)^{-1}\{\sum_{i=1}^{N} (a_N(i))^2\}$$

and

$$(3.2.7) \qquad E\{(S_N - E(S_N))^4\} = A_N + B_N,$$

where

$$A_N = 3(N-1)(N+1)^{-1}\{(N-1)^{-1}\sum_{i=1}^{N}(a_N(i))^2\}^2$$

and

$$B_N = \{\sum_{i=1}^{N} c_i^4 - 3(N-1)(N(N+1))^{-1}\}((N-1)(N-2)(N-3))^{-1} \times$$
$$\{N(N+1)\sum_{i=1}^{N}(a_N(i))^4 - 3(N-1)(\sum_{i=1}^{N}(a_N(i))^2)^2\}.$$

The theory of the limiting distributions of the statistics S_N under contiguous alternatives is expounded in Chapter VI of HS. In order to outline the concepts involved, it is advisable to recall the precise formulation of the problem in terms of the limiting distribution of the sequence of statistics $\{S_{c_\nu}\}$. This is the test statistic, based on $X_1,\dots,X_{N(\nu)}$, for testing H_ν against K_ν. Eac. H_ν is just H, the hypothesis of randomness, but K_ν will for the rest of the discussion depend on the parameters $d_\nu(1),\dots,d_\nu(N(\nu))$.

The alternatives K) are obtained by supposing that $X_1,\dots,X_{N(\nu)}$ have joint nsity

$$q_{d_\nu} = \prod_{i=1}^{N(\nu)} f(x_i - d_\nu(i)), \tag{3.2.8}$$

where f has finite Fisher's information and $d_\nu = (d_\nu(1),\ldots, d_\nu(N(\nu)))$ defines a sequence of real vectors. Recall that if $f(x)$ is an absolutely continuous density and $\inf(f) = \int_{-\infty}^{\infty} (f'(x)/f(x))^2 f(x)dx < \infty$, then this is the Fisher's information.

Definition 3-2-13 (HS p.202). Consider the measures P_ν and Q_ν defined on the measure spaces $(\chi_\nu, A_\nu, \mu_\nu)$, $\nu \geq 1$, so that $dP_\nu = p_\nu d\mu_\nu$ and $dQ_\nu = q_\nu d\mu_\nu$. If for any sequence of events $\{A_\nu\}$, $A_\nu \in A_\nu$,

$$P_\nu(A_\nu) \to 0 \text{ implies that } Q_\nu(A_\nu) \to 0,$$

then we say that the densities q_ν are contiguous to the densities p_ν.

If H_ν is composite, q_ν is said to be contiguous to H_ν if for each ν the convex hull of H_ν contains a density p_ν such that the densities q_ν are contiguous to the densities p_ν.

The limiting distribution of $\{S_{c_\nu}\}$ is then established under the following conditions on d_ν:

$$\max_{1\leq i\leq N(\nu)} (d_\nu(i) - \bar{d}_\nu)^2 \to 0 \tag{3.2.9}$$

and

$$\inf(f)\sum_{i=1}^{N} (d_\nu(i) - \bar{d}_\nu)^2 \to b^2,\ 0 < b^2 < \infty. \tag{3.2.10}$$

The derivation of results about the limiting distribution of $\{S_{c_\nu}\}$ under $K(d_\nu)$ is made possible because the q_{d_ν} are contiguous to the densities p_{d_ν} defined by

$$p_{d_\nu} = \prod_{i=1}^{N(\nu)} f(x_i - \bar{d}_\nu)$$

(HS theorem VI 2.1, pp.210-213).

The result about the limiting distribution of $\{S_{c_\nu}\}$ under $K(d_\nu)$ is stated under the assumption that $\bar{c}_\nu = 0$ and the index ν is dropped.

Theorem 3-2-14 (HS pp.216-218). Let q_d be given by (3.2.8) and assume that (3.2.4), (3.2.5), (3.2.9) and (3.2.10) hold. Then S_N is asymptotically normal (μ_{dc}, σ_c^2) with

$$\mu_{dc} = \{\sum_{i=1}^{N} c_i(d_i - \bar{d})\} \int_0^1 \varphi(u)\psi(u)\,du,$$

where $\psi(u) = f'(F^{-1}(u))/f(F^{-1}(u))$, and

$$\sigma_c^2 = (\sum_{i=1}^{N} c_i^2) \int_0^1 \{\phi(u) - \bar{\phi}\}^2 du.$$

In order to verify the tightness of a sequence of probability measures, the following result about contiguous densities is useful. The notation follows that of Definition 3-2-13.

Lemma 3-2-15 (Jurečková (1969), pp.1894-1895). If the densities q_ν are contiguous to p_ν, then, corresponding to every positive ε, there exists a positive number δ such that $Q_\nu(A_\nu) < \varepsilon$ is satisfied for all ν sufficiently large for every sequence of sets $\{A_\nu\}$, $A_\nu \in \mathcal{A}_\nu$ $(\nu = 1,2,\ldots)$

satisfying $P_\nu(A_\nu) < \delta$ for ν sufficiently large.

In section 3.3. we also discuss the limiting distributions of a class of test statistics which are quadratic *forms in* $R_1,\ldots,R_N$. These involve the integral $\int_0^1 (S(t))^2 dt$, where $S(t)$ is a stationary Gaussian process which for the present discussion is assumed to be *normalized, so that*

$$E\{S(t)\} = 0, \quad E\{(S(t))^2\} = 1.$$

The covariance function is defined by

$$\rho(t-s) = E\{(S(s)S(t)\}, \quad 0 \le s,t \le 1.$$

The eigenvalues λ_r of a non-negative definite, symmetric kernel $K(s,t)$ on $[0,1] \times [0,1]$ are defined by

$$\lambda_r \Phi_r(s) = \int_0^1 K(s,t)\Phi_r(t)dt.$$

If $\int_0^1 \Phi_i(t)\overline{\Phi_j(t)} = \delta_{ij}$, then the $\Phi_r(t)$ are the orthonormal *eigenfunctions* of the kernel ($\overline{f(t)}$ denotes the complex conjugate of $f(t)$ and δ_{ij} is the Kronecker delta). The eigenvalues are real and non-negative. *In particular* let us suppose that the eigenvalues and orthonormal eigenfunctions of $\rho(t-s)$ are λ_r and $\Phi_r(t)$.

<u>Theorem 3-2-16</u> (Kac and Siegart (1947), pp.438-441). Suppose that $J_1,J_2,\ldots$ are independent, normally distributed random variables each having mean 0 and variance 1.

The expansion

$$(3.2.11) \qquad S(t) = \sum_k \lambda_k^{\frac{1}{2}} J_k \, \phi_k(t)$$

holds in the sense that, with probability one, the right hand side converges in the mean to the left hand side of (3.2.11). Furthermore

$$\int_0^1 (S(t))^2 dt = \sum_k \lambda_k J_k^2 ,$$

convergence being with probability one (almost everywhere convergence).

In the course of the rest of this chapter (and in Part II of the thesis), use is made of the terminology and properties of the $L_2([0,1])$ space. Often the use is simply as a kind of convenient shorthand in order to avoid unnecessary elaboration. The salient points required for our purposes are mentioned below; subsequent technical references (usually not essential to understanding the main theme of the work) are not elucidated as they are part of the standard theory of Hilbert space and can be located in any modern textbook on functional analysis (as, for instance, Hewitt and Stromberg (1969)).

The space $L_2([0,1])$ consists of those functions ϕ defined almost everywhere on $[0,1]$ for which $\int_0^1 (\phi(u))^2 du$ exists (all in the sense of Lebesgue measure and Lebesgue integration for $[0,1]$) together with an inner product $\langle , \rangle$ defined by

$$\langle\phi_1,\phi_2\rangle = \int_0^1 \phi_1(u)\phi_2(u)du$$

for $\phi_1,\phi_2 \in L_2([0,$ It follows that the norm on $L_2([0,1])$ is defi by $\|\phi\| = \{\int_0^1 (\phi(u))^2du\}^{\frac{1}{2}}$.

The norm of a bounded linear operator S mapping $L_2([0,1])$ into itself is defined by

$$\|S\| = \sup\{\|S(\phi)\| : \|\phi\| = 1\}.$$

Notice that a simple consequence of this is that $\|S(\phi)\| \le \|S\|\|\phi\|$ for all $\phi \in L_2([0,1])$.

3.3 Asymptotic distributions of the test statistics.

This section is concerned with the limiting distributions of the statistics

(3.3.1) $$\nu_N(R) = \max_{g\in G} \{S_N(g(r))\},$$

where $S_N(R)$ are defined by (1.2.1) as

(3.3.2) $$S_N(R) = \sum_{i=1}^{N} c_i\ a_N(R_i).$$

In fact, instead of studying ν_N, we shall begin by considering the limiting distribution of ν_{N1} defined by

$$\nu_{N1} = \max_{g\in G_1} \{S_N(g(R))\}.$$

The definition of $a_N(i)$ is extended from $i = 1,\ldots,N$ by using modulo N arithmetic on the argument, so that in particular $a_N(-i) = a_N(N-i)$ $(i = 1,\ldots,N)$ and $a_N(0) = a_N(N)$. Then a form of ν_{N1} more suitable for studying the limiting distribution is obtained by introducing the scores $a_{N1}(i,t)$ defined as follows:

$$(3.3.3) \qquad a_{N1}(i,t) = a_N(i - [tN]) \qquad (0 \le t \le 1)$$

where $[x]$ denotes the greatest integer less than or equal to x. The definition of the scores $a_{N1}(i,t)$ implies that for $k/N \le t < (k+1)/N$, $a_{N1}(i,t) = a_N(i-k)$ for $k = 0,\ldots,N-1$. Now define $S_{N1}(t) = \sum_{i=1}^{N} c_i \, a_{N1}(R_i,t)$ $(0 \le t \le 1)$, so that

$$(3.3.4) \qquad \nu_{N1} = \sup_{0\le t\le 1} \{S_{N1}(t)\}.$$

In addition to the scores defined by (3.3.3), the following scores will be used later:

$$(3.3.5) \quad a_{N2}(i,t) = a_N(N + 1 - i - [tN]) \qquad (-1 \le t \le 0).$$

If $-k/N \le t < -(k-1)/N$, then $a_{N2}(i,t) = a_N(N$ [illegible] $k)$ for $k = 1,\ldots,N$. Also define $S_{N2}(t) = \sum_{i=1}^{N} c_i$ [illegible] $,t)$.

The function $S_{N1}(t)$ is a right-continuous step-function and consequently the convergence in distribution of $S_{N1}(t)$ is studied in the space $D([0,1])$. By defining the scores $a_{N1}(i,t)$ so that they are linear over the intervals $k/N \le t \le (k+1)/N$, $k = 0,\ldots,N-1$, $S_{N1}(t)$ could be made to be continuous and its convergence could then be discussed within $C([0,1])$ without affecting (3.3.4). The main reason for not doing this is that the basic convergence theorem for $S_{N1}(t)$ as it is now defined can also be used to obtain the limiting distributions of another class of circular tests. Altering the definition of $a_{N1}(i,t)$ to render $S_{N1}(t)$ continuous would considerably complicate this other application of the convergence results.

The following assumptions are made:

(A1) *The Noether condition,*

$$\sum_{i=1}^{N} (c_i - \bar{c})^2 / (\max_{1 \leq i \leq N} (c_i - \bar{c})^2) \to \infty$$

as $N \to \infty$.

(A2) There exists a function $\phi(u)$, $0 < u < 1$, such that $\phi \in L_2([0,1])$, $\int (\phi(u) - \bar{\phi})^2 du > 0$ (where $\bar{\phi} = \int \phi(u)du$) and $a_N(1 + [uN]) \to \phi(u)$ with respect to the $L_2([0,1])$ norm. In future L_2 will be taken to mean $L_2([0,1])$ and when the range of integration is omitted, it is understood to be $[0,1]$. The definition of $\phi(u)$ is extended outside $(0,1)$ by periodicity, so that $\phi(u+k) = \phi(u)$, $k = \pm 1, \pm 2, \ldots$.

(A3) Define the translation operator on L_2 by $S_t(f(u)) = f(u-t)$ for $f \in L_2$ and $t \in [-1,1]$, using modulo 1 arithmetic to ensure that the argument of f lies between 0 and 1. Define the reflection operator by $R(f(u)) = f(1-u)$, $f \in L_2$. These operators are linear, continuous and unitary.

These properties of S_t and R are straightforward to obtain. For example, $\langle S_t f, S_t g\rangle = \int_0^t f(u+1-t)g(u+1-t)du + \int_t^1 f(u-t)g(u-t)du = \int_{1-t}^1 f(u)g(u)du + \int_0^{1-t} f(u)g(u)du = \langle f,g\rangle$, so that S_t is unitary .

From S_t and R construct an operator U_t defined by $U_t = R(S_{-t}(f))$ for $t \in [-1,0]$. For each t, U_t is also unitary. Consequently, $\|U_t\| = 1$ (and likewise $\|S_t\| = 1$).

Assume that there exists an $\alpha > \frac{1}{2}$ and L a nondecreasing, continuous function on [0,1] such that

$$(3.3.6) \qquad \|(S_s - S_t)\phi\|^2 \le |L(s) - L(t)|^\alpha$$

for $s,t \in [0,1]$.

(A4) There exists an integer N_o such that for $N \ge N_o$ there exists $M < \infty$ such that

$$|N^{-1} \sum_{i=1}^{N} (a_{N1}(i,s) - a_{N1}(i,t))^4| < M$$

for $s,t \in [0,1]$.

The required convergence result for $S_{N1}(t)$ is contained in the following theorem.

<u>Theorem 3-3-1</u> Under H, with the assumptions (A1)-(A4),

$$(3.3.7) \quad \{\sum_{i=1}^{N} (c_i-\bar{c})^2\}^{-\frac{1}{2}}(S_{N1}(t) - \bar{c}\sum_{i=1}^{N} a_N(i)) \to_{\mathcal{D}} S_1^o(t),$$

where $S_1^o(t)$ is the continuous Gaussian process on [0,1] with $E\{S_1^o(t)\} = 0$ for $t \in [0,1]$ and $R_1(s,t) = \mathrm{cov}(S_1^o(s), S_1^o(t)) = \int (S_s(\phi)) - \bar{\phi})\ (S_t(\phi(u)) - \bar{\phi})du$ for $s,t \in [0,1]$.

In order to prove Theorem 3-3-1 we make the following assumptions (without loss of generality):

$$\bar{a}_N = 0,\ \bar{c} = 0,\ \sum_{i=1}^{N} c_i^2 = 1.$$

From S_t and R construct an operator U_t defined by $U_t = R(S_{-t}(f))$ for $t \in [-1,0]$. For each t, U_t is also unitary. Consequently, $\|U_t\| = 1$ (and likewise $\|S_t\| = 1$).

Assume that there exists an $\alpha > \frac{1}{2}$ and L a nondecreasing, continuous function on [0,1] such that

$$(3.3.6) \qquad \|(S_s - S_t)\phi\|^2 \leq |L(s) - L(t)|^\alpha$$

for $s,t \in [0,1]$.

(A4) There exists an integer N_o such that for $N \geq N_o$ there exists $M < \infty$ such that

$$|N^{-1} \sum_{i=1}^{N} (a_{N1}(i,s) - a_{N1}(i,t))^4| < M$$

for $s,t \in [0,1]$.

The required convergence result for $S_{N1}(t)$ is contained in the following theorem.

<u>Theorem 3-3-1</u> Under H, with the assumptions (A1)-(A4),

$$(3.3.7) \quad \{\sum_{i=1}^{N}(c_i-\bar{c})^2\}^{-\frac{1}{2}}(S_{N1}(t) - \bar{c}\sum_{i=1}^{N} a_N(i)) \to_{\mathcal{D}} S_1^o(t),$$

where $S_1^o(t)$ is the continuous Gaussian process on [0,1]
where $S_1^o(t)$ is the continuous Gaussian process on [0,1]),
with $E\{S_1^o(t)\} = 0$ for $t \in [0,1]$ and $R_1(s,t) = \mathrm{cov}(S_1^o(s), S_1^o(t)) = \int (S_s(\phi)) - \bar{\phi})\ (S_t(\phi(u)) - \bar{\phi})du$ for $s,t \in [0,1]$.

In order to prove Theorem 3-3-1 we make the following assumptions (without loss of generality):

$$\bar{a}_N = 0, \ \bar{c} = 0, \ \sum_{i=1}^{N} c_i^2 = 1.$$

This implies that $\bar{\phi} = 0$ and the covariance kernel $R_1(s,t)$ can be neatly written as $R_1(s,t) = \langle S_s(\phi), S_t(\phi)\rangle$. In view of the properties of S_t, the process $s_1^o\ (t)$ is stationary.

In order to prove Theorem 3-3-1, a series of lemmas are required. The first of these is a result about the convergence of fourth moments. If it were not for the requirement of uniform convergence, the result would follow more easily from the asymptotic normality of $S_{Ni}(s) - S_{Ni}(t)$, $i = 1,2$ (see Theorem 3-2-11).

Lemma 3-3-2. Under H the following limits hold as $N \to \infty$:

$$(3.3.8)\quad E\{(S_{N1}(s) - S_{N1}(t))^4\} \to 3\{\int (S_s(\phi(u)) - S_t(\phi(u)))^2 du\}^2,$$

uniformly for $s,t \in [0,1]$;

$$(3.3.9)\quad E\{(S_{N2}(s) - S_{N2}(t))^4\} \to 3\{\int (U_s(\phi(u)) - U_t(\phi(u)))^2 du\}^2$$

uniformly for $s,t \in [-1,0]$.

Proof. From Theorem 3-2-12, it is clear that $E\{(S_{Ni}(s) - S_{Ni}(t))^4\}$ $i = 1,2$, can be written as the sum of two terms A_N and B_N whose convergence properties may be studied individually.

To study these convergence properties further, we first show that as far as convergence in L_2 is concerned, the scores $a_{N1}(1 + [uN],t)$ may be replaced by $S_t(a_N(1 + [uN]))$. Suppose that $0 < t < 1$, then

$$\| a_{N1}(1+[uN],t) - S_t(a_N(1+[uN]))\| = \| a_N(1+[uN]-[tN]) - a_N(1+[(u-t)N])\|.$$

Since for real numbers x and y, $[x-y]$ equals either $[x] - [y]$ or $[x] - [y] - 1$, then $a_N(1 + [(u-t)N])$ equals $a_N(1 + [uN] - [tN])$ or $a_N([uN] - [tN])$. Therefore it is certainly true that

$$(3.3.10) \quad \| a_{N1}(1 + [uN], t) - S_t(a_N(1 + [uN]))\| \leq$$

$$\| a_N(1 + [uN] - [tN]) - a_N([uN] - [tN])\| .$$

In order to simpli.. 'he right hand side of (3.3.10), notice that

$$\| a_N(1 + [uN] - [tN]) - a_N([uN] - [tN])\|^2$$

$$= N^{-1} \sum_{i=1}^{N} (a_N(1+i - [tN]) - a_N(i - [tN]))^2$$

(recall that $a_N(i)$ is periodic with period N)

$$= N^{-1} \sum_{1+[tN]}^{N} \{a_N(1 + (i - [tN])) - a_N(i-[tN])\}^2 +$$

$$N^{-1} \sum_{1}^{[tN]} \{a_N(1+(i+N-[tN]))-a_N(i+N-[tN])\}^2$$

$$= N^{-1} \sum_{j=1}^{N-[tN]} \{a_N(1+j)-a_N(j)\}^2 + N^{-1} \sum_{j=N-[tN]+1}^{N}$$

$$\{a_N(1+j) - a_N(j)\}^2$$

$$= N^{-1} \sum_{j=1}^{N} \{a_N(1+j) - a_N(j)\}^2$$

$$= \| a_N(1 + [uN]) - a_N([uN])\|^2 .$$

It now follows that (3.3.10) can be written as

$$\| a_{N1}(1 + [uN],t) - S_t(a_N(1 + [uN]))\| \le$$

$$\| a_N(1 + [uN]) - a_N(1 + [(u-N^{-1})N])\|$$

$$\le \| a_N(1 + [uN]) - \phi(u)\| + \|\phi(u) - \phi(u-N^{-1})\|$$

$$+ \| S_{N^{-1}}(a_N(1 + [uN]) - \phi(u))\|,$$

which, for N sufficiently large and $\varepsilon > 0$,

$$\le 2\varepsilon + \|\phi(u) - \phi(u - N^{-1})\|,$$

because $a_N(1 + [uN]) \to \phi(u)$ in L_2 (by (A2)). Provided that $\|\phi(u) - \phi(u - N^{-1})\| \to 0$ as $N \to \infty$, it follows that for $0 \le t \le 1$, $a_{N1}(1 + [uN],t) - S_t(a_N(1 + [uN])) \to 0$ in L_2, uniformly in t. Since $\phi(u)$ obeys (A3), $\|\phi(u)-\phi(u-N^{-1})\| \to 0$ as $N \to \infty$, but in fact the result is true for all $\phi \in L_2$ as may be seen by suitable modification of theorem 13.24 of Hewitt and Stromberg (1969, p.199) using theorem 13.21 on p.197 of the same book.

Thus

$$(3.3.11) \qquad a_{N1}(1 + [uN],t) - S_t(a_N(1 + [uN])) \to 0$$

in L_2, uniformly for $0 \le t \le 1$.

Now

$$|\{N^{-1}\sum_{i=1}^{N}(a_{N1}(i,s)-a_{N1}(i,t))^2\}^{\frac{1}{2}}-\{\int(S_s(\phi(u))-S_t(\phi(u)))^2\}^{\frac{1}{2}}|$$

$$= |\|a_{N1}(1+[uN],s)-a_{N1}(1+[uN],t)\|-\|S_s(\phi)-S_t(\phi)\||$$

$$\leq |\|S_s(a_N(1+[uN]))-S_t(a_N(1+[uN]))\|-\|S_s(\phi)-S_t(\phi)\|| + P_{s,t}$$

$$(3.3.12) \quad \leq \|S_s(a_N(1 + [uN])-S_s(\phi)\|+\|S_t(a_N(1+[uN]))$$
$$- S_t(\phi)\|+P_{s,t},$$

where

$$P_{s,t} = \|a_{N1}(1+[uN],s)-S_s(a_N(1+[uN]))\| +$$
$$\|a_{N1}(1+[uN],t)-S_t(a_N(1+[uN]))\|,$$

so that $P_{s,t} \to 0$ as $N \to \infty$ uniformly in s and t by (3.3.11).

From (3.3.12) it follows that

$$|\{N^{-1}\sum(a_{N1}(i,s)-a_{N1}(i,t))^2\}^{\frac{1}{2}}-\{\int(S_s(\phi(u))-S_t(\phi(u)))^2\}^{\frac{1}{2}}|$$

$$\leq \|S_s\|\|a_N(1+[uN])-\phi(u)\|+\|S_t\|\|a_N(1+[uN])-\phi(u)\|+P_{s,t}$$

$$= 2\|a_N(1+[uN])-\phi(u)\|+P_{s,t},$$

which by (A2) and the remarks above about $P_{s,t}$ tends to 0 as $N \to \infty$. Hence,

$$A_N \to 3\{\int (S_s(\phi(u)) - S_t(\phi(u)))^2\}^2$$

uniformly for $s,t \in [0,1]$.

To show that $B_N \to 0$ uniformly, the expression

$$(3.3.13)\ |((N-1)(N-2)(N-3))^{-1}\{N(N+1)\sum_i (a_{N1}(i,s)-a_{N1}(i,t))^4 \\ -3(N-1)(\sum_i (a_{N1}(i,s)-a_{N1}(i,t))^2)^2\}|$$

is bounded uniformly for $s,t \in [0,1]$ and N sufficiently large. This is achievedby using (A4) to bound the first term in (3.3.13). The second term in (3.3.13), that is for our purposes

$$3((N-2)(N-3))^{-1}(\sum_i (a_{N1}(i,s)-a_{N1}(i,t))^2)^2,$$

is uniformly bounded in view of the uniform convergence of A_N together with the fact that $(\| S_s(\phi) - S_t(\phi)\|^2)^2$ is uniformly bounded.

Having bounded (3.3.13) uniformly, the required uniform convergence of B_N follows from (A1) which implies $\sum c_i^4 \to 0$. The truth of this last statement seems to be usually regarded as self-evident. With this in mind, we apologize if the following quick outline proof is unnecessary.

Consider the following problem: the non-negative real numbers $r_1 \le \ldots \le r_N$ satisfy $\sum_{i=1}^{N} r_i = 1$; show that $\sum_{i=1}^{N} r_i^2 \le r_N$. This is obviously true for $N = 1$. Suppose that it is true for N-1, then apply the result to $(1-r_N)^{-1}r_i$, $i = 1,\ldots,N-1$, obtaining $\sum_{i=1}^{N-1} r_i^2 \le r_{N-1}(1-r_N)$. Hence $\sum_{i=1}^{N} r_i^2 \le r_{N-1} - r_{N-1}\, r_N + r_N^2 = r_{N-1} + r_N(r_N-r_{N-1}) \le r_N$ since $r_N \le 1$, and the result is true by induction. From the solution of this problem it follows that $\sum c_i^4 \le \max\{c_i^2\}$. This together with (A1) gives the required convergence result.

The proof of (3.3.9) is similar. First it is required to show that for $t \in [-1,0]$, $a_{N2}(1 + [uN],t) - RS_{-t}(a_N(1 + [uN])) \to 0$ in L_2 uniformly in t. *This presents* some tricky aspects which the previous part did not.

The required result is that $\| a_N(N + 1 - (1 + [uN]) - [tN]) - a_N(1 + [-uN - tN]))\| \to 0$ as $N \to \infty$. But,

$$\begin{aligned} &\| a_N(- [uN] - [tN]) - a_N(1 + [-uN - tN])\| \\ &\le \| a_N(- [uN] - [tN]) - a_N([-uN - tN])\| \\ &\quad + \| a_N([-uN - tN]) - a_N(1 + [-uN - tN])\| . \end{aligned}$$

The second term here is easily managed since it becomes

$$\begin{aligned} &\| a_N(1 + [-(u + t + N^{-1})N]) - a_N(1 + [-(u + t)N])\| \\ &= \| U_t S_{-N^{-1}}(a_N(1 + [uN])) - U_t(a_N(1 + [uN]))\| \\ &\le \| S_{N^{-1}}(a_N(1 + [uN])) - a_N(1 + [uN])\| , \end{aligned}$$

which tends to 0 as $N \to \infty$.

The first term in the expression above is tackled by using the observation that, for real numbers x and y:

$$- [x] - [y] \text{ equals } [-x] - [y] \text{ or } [-x] - [y] - 1 ,$$
$$[-x-y] \text{ equals } [-x] - [y] \text{ or } [-x] - [y] - 1 .$$

Therefore ,

$$\| a_N(- [uN] - [tN]) - a_N([- uN - tN])\|$$

$$\leq \| a_N([- uN] - [tN]) - a_N([- uN] - [tN] - 1)\| .$$

This last term tends to 0 uniformly as $N \to \infty$ using the same arguments as before. Consequently,

$$(3.3.14)\quad a_{N2}(1 + [uN],t) - \mathcal{U}_t(a_N(1 + [uN])) \to 0$$

in L_2, uniformly for $t \in [-1,0]$.

The rest of the proof of (3.3.9) is exactly like that of (3.3.8). To prove the convergence of the A_N term, simply replace a_{N1}, S_t by a_{N2} and $\mathcal{U}_t$ in the proof already given. As far as the convergence of the B_N term is concerned, all that need by noted is that (A4) implies that

$$|N^{-1} \sum_{i=1}^{N} (a_{N2}(i,s) - a_{N2}(i,t))^4| < M$$

for $s,t \in [-1,0]$ and $N \geq N_0$ from the way the scores $a_{N2}(i,t)$ were defined by (3.3.5).

The proof of the lemma is complete.

Examination of the proof of Lemma 3-3-2 shows that all that is required from (A4) is that

$$(\sum_{i=1}^{N} c_i^4)|N^{-1} \sum_{i=1}^{N} (a_N(i,s) - a_N(i,t))^4| \to 0,$$

uniformly for s,t. The assumption we have made seems to be an adequate working assumption.

From here on most of the rest of the work necessary to derive Theorem 3-3-1 amounts to taking limiting cases of the results in Billingsley (1968) cited in section 3.2. Accordingly repetition of material and mere rewriting of Billingsley's material is avoided wherever possible, so that Lemma 3-3-3 and 3-3-4 are of the nature of (detailed) outline proofs.

Lemma 3-3-3. Suppose that for positive integers N,m there corresponds a sequence of random variables $\xi_{N1}^{(m)},\ldots,\xi_{Nm}^{(m)}$ and define $S_{Nk}^{(m)} = \xi_{N1}^{(m)} + \ldots + \xi_{Nk}^{(m)}$, $k = 1,\ldots,m$ and $S_{NO}^{(m)} = 0$. If $\alpha > 1$ and $\gamma \geq 0$, α and γ constants, and there exist non-negative numbers $u_1^{(m)},\ldots,u_m^{(m)}$ such that as $N \to \infty$,

$$\limsup P\{|S_{Nj}^{(m)} - S_{Ni}^{(m)}| \geq \lambda\} < \lambda^{-\gamma}\Big(\sum_{i<\ell\leq j} u_\ell^{(m)}\Big)^\alpha,$$

$$0 \leq i < j \leq m,$$

holds for all $\lambda > 0$ uniformly for $m = 1,2,\ldots$, then there exists an integer N_o such that

$$P\{\max_{1\leq k\leq m} |S_{Nk}^{(m)}| \geq \lambda\} \leq K_{\gamma,\alpha}\lambda^{-\gamma}(u_1^{(m)} + \ldots + u_m^{(m)})^\alpha$$

for $N \geq N_o$ and $m = 1,2,\ldots$, where $K_{\gamma,\alpha}$ is a constant which depends only on α and γ.

Proof. Choose N_o so that for $N \geq N_o$

$$\sup_{r\geq N} P\{|S_{rj}^{(m)} - S_{ri}^{(m)}| \geq \lambda\} \leq \lambda^{-\gamma}\Big(\sum_{i<\ell\leq j} u_\ell^{(m)}\Big)^\alpha,$$

$0 \leq i < j \leq m$, $m = 1,2,\ldots$. The lemma then follows as an application of Theorem 3-2-9.

Lemma 3-3-4. The sequence $\{S_{N1}\}$ of random elements of D is tight.

Proof. The proof is modelled on the proof of theorem 12.3 of Billingsley (1968, pp.95-96). Using Lemma 3-3-2 and (A3) (after absorbing the 3 into L), we get

$$\lim_{N\to\infty} E\{(S_{N1}(s) - S_{N1}(t))^4\} \le |L(s) - L(t)|^{2\alpha} ,$$

uniformly for s,t. This moment condition implies that

$$(3.3.15) \quad \limsup_{N \to \infty} P\{|S_{N1}(s) - S_{N1}(t)| \ge \lambda\} \le \lambda^{-4}|L(s)-L(t)|^{2\alpha},$$

for positive λ, uniformly in s,t. The inequality in (3.3.15) may be replaced by strict inequality (by multiplying L by a positive constant greater than 1). If we choose $\Delta \in [0,1)$ and $\delta > 0$, $\Delta + \delta \le 1$, then the random variables

$$\xi_{Ni}^{(m)} = S_{N1}(\Delta + ({}^{i}/_{m})\delta) - S_{N1}(\Delta + ({}^{(i-1)}/_{m})\delta),$$

$$i = 1,\ldots,m,$$

satisfy Lemma 3-3-3 with

$$u_i^{(m)} = L(\Delta + i\delta m^{-1}) - L(\Delta + (i-1)\delta m^{-1}),\ i = 1,\ldots,m.$$

Then the lemma implies that for positive ε there exists N_0 such that

$$(3.3.16)\quad P\{\max_{0\le i\le m} |S_{N1}(\Delta + i\delta m^{-1}) - S_{N1}(\Delta)| \ge \varepsilon\}$$

$$\le K\,\varepsilon^{-4}\{L(\Delta + \delta) - L(\Delta)\}^{2\alpha}$$

for $N \ge N_o$, $m = 1,2,\ldots$ and all Δ.

If $\varepsilon' > 0$, then choose $\delta = n^{-1}$, where n is an integer such that L which is (uniformly) continuous on [0,1] satisfies $|L(x + \delta) - L(x)| < \varepsilon'$ for $0 \le x \le 1 - \delta$. Letting $m \to \infty$ in (3.3.16) implies that

$$\sum_{j<n} P\{\sup_{j\delta<s<(j+1)\delta} |S_{N1}(s) - S_{N1}(j\delta)| \ge \varepsilon\}$$

$$\le K\,\varepsilon^{-4}(L(1) - L(0))(\max_{j<n}\{L((j+1)\delta) - L(j\delta)\})^{2\alpha-1}$$

$$\le K\,\varepsilon^{-4}(L(1) - L(0)) \times (\varepsilon')^{2\alpha-1}$$

and since ε' can be assigned arbitrary positive values, this last term can be made arbitrarily small. Now Lemma 3-2-8 can be used to prove that $\{S_{N1}\}$ satisfies (3.2.3) of Theorem 3-2-7. The other condition of Theorem 3-2-7 - which requires that $\{S_{N1}(0)\}$ be tight - is a consequence of the asymptotic normality of $\{S_{N1}(0)\}$.

The proof of Theorem 3-3-1 is completed by showing that the finite-dimensional distributions of $S_{N1}(t)$ converge weakly to those of $S_1^o(t)$. This is done with the aid of Theorem 3-2-10. The proof of Lemma 3-3-2 essentially contains a proof that $a_{N1}(1 + [uN],t) \to S_t(\phi(u))$ in L_2 and hence, for $0 \le t(1) < \ldots < t(k) \le 1$ and an arbitrary vector of real numbers $(\lambda_1,\ldots,\lambda_k)$, $\sum_j \lambda_j\, a_{N1}(1 + [uN],t(j)) \to$

$\sum_j \lambda_j S_{t(j)}(\phi(u))$ in L_2. An application of Theorem 3-2-11 to $\sum_j \lambda_j \sum_i c_i a_{N1}(R_i,t(j)) = \sum_i c_i (\sum_j \lambda_j a_{N1}(R_i,t(j)))$ proves that this is asymptotically normal with mean 0 and variance $\int\{\sum \lambda_j S_{t(j)}(\phi(u))\}^2 du$, provided this last integral is non-zero. The other-degenerate - case is straightforward. Using Theorem 3-2-3, the proof of Theorem 3-3-1 is now complete.

Corollary 3-3-5. *The limiting distribution of* ν_{N1} *is given by*

$$(3.3.17)\quad \{\sum_{i=1}^{N}(c_i-\bar{c})^2\}^{-\frac{1}{2}}(\nu_{N1}-\bar{c}\sum_{i=1}^{N} a_N(i)) \to_{\mathcal{D}} \sup_{0\le t\le 1}\{S_1^0(t)\}.$$

Proof. The functional $z(t) \to \sup_{0\le t\le 1}\{z(t)\}$ is continuous on (D,τ) and so Theorem 3-2-2 can be applied to give (3.3.17). That the functional is in fact continuous is fairly obvious from the definition of τ.

Theorem 3-3-1 can be extended to obtain the limiting distribution of ν_N. Essentially this requires piecing the two random elements $S_{N1}(t)$ and $S_{N2}(t)$ together to give $S_N(t)$, a right-continuous step function on $[-1,1]$. Although the discussion in section 3.2 concerned $D([0,1])$, obviously it applies to $D([a,b])$ for $a<b$ - the space of right-continuous real functions on $[a,b]$ with left-hand limits - with the appropriate straightforward modifications.

The details are as follows. Define the scores $a_N(i,t)$ by

$$\begin{aligned} a_N(i,t) &= a_{N1}(i,t) && (0 \le t \le 1) \\ &= a_{N2}(i,t) && (-1 \le t < 0), \end{aligned}$$

together with the function $S_N(t) = \sum_{i=1}^{N} c_i\, a_N(R_i,t)$ for $t \in [-1,1]$. The operators S_t and U_t are replaced by V_t defined by

$$\begin{aligned} V_t &= S_t && (0 \le t \le 1) \\ &= U_t && (-1 \le t < 0). \end{aligned}$$

As with Corollary 3-3-5, Theorem 3-3-6 is stated in its full *generality*.

<u>Theorem 3-3-6</u>. Under H, with the assumptions (A1)-(A4),

$$(3.3.18)\quad \{\sum_{i=1}^{N} (c_i-\bar{c})^2\}^{-\frac{1}{2}}(\nu_N - \bar{c} \sum_{i=1}^{N} a_N(i)) \to_{\mathcal{D}} \sup_{-1\le t\le 1} \{S^o(t)\},$$

where $S^o(t)$ is the Gaussian process on $[-1,1]$ with $E\{S^o(t)\} = 0$ for $t \in [-1,1]$ and $R(s,t) = \mathrm{cov}(S^o(s), S^o(t))$ $= \int (V_s(\phi(u)) - \bar{\phi})(V_t(\phi(u)) - \bar{\phi})du.$

<u>Proof</u>. If it can be shown that $S_N(t) \to_{\mathcal{D}} S(t)$ in $D([-1,1])$, then (3.3.18) follows.

The proof of Lemma 3-3-4 consisted principally of showing that for each positive ε and η, there exists a δ, $0 < \delta < 1$, and an integer N_o such that, for $N \ge N_o$, $P\{w(S_{N1},\delta) \ge \varepsilon\} < \eta$. A similar result applies to S_{N2}, since it satisfies effectively the same conditions as S_{N1}. It satisfies (3.3.9) of Lemma 3-3-2 compared with (3.3.8) *for* S_{N1}. *Furthermore since* $\|(RS_{-s} - RS_{-t})\phi\| \le \|(S_{-s}-S_{-t})\phi\|$ $\le |L(-s) - L(-t)|^{\alpha}$, it follows that a condition analogous to (3.3.6) holds for U_t. Therefore the arguments of *Lemma* 3-3-4 apply equally well to S_{N2}.

Theorem 3-2-6 is used to prove that $S_N(t) \to_{\mathcal{D}} S(t)$. In particular it is required that for each positive ε and η, there exist a δ, $0 < \delta < 1$, and an integer N_o such that

$$(3.3.19) \qquad P\{w'(S_N,\delta) \geq \varepsilon\} \leq \eta \quad , \quad N \geq N_o \ .$$

This will be established by proving that

$$(3.3.20) \qquad w'(S_N,\delta) \leq \max\{w(S_{N1},\delta), w(S_{N2},\delta)\} \ ,$$

so that

$$P\{w'(S_N,\delta) \geq \varepsilon\} \leq P\{w(S_{N1},\delta) \geq \varepsilon\} + P\{w(S_{N2},\delta) \geq \varepsilon\}$$

and consequently (3.3.19) follows from the properties of $\{S_{N1}\}$ and $\{S_{N2}\}$.

In order to prove (3.3.20), there are three cases to consider. If $-1 \leq t_1 + \delta < 0$, then for all t, t_2 such that $t_1 \leq t \leq t_2$ and $t_2 - t_1 \leq \delta$,

$$\min\{|S_N(t) - S_N(t_1)|, |S_N(t_2) - S_N(t)|\} \leq w(S_{N2},\delta).$$

If $t_1 \geq 0$, then for t, t_2 such that $t_1 \leq t \leq t_2$ and $t_2 - t_1 \leq \delta$

$$\min\{|S_N(t) - S_N(t_1)|, |S_N(t_2) - S_N(t)|\} \leq w(S_{N1},\delta).$$

If $t_1 < 0 \leq t_1 + \delta$, two possibilities arise for t such that $t_1 \leq t \leq t_2$. If $t \geq 0$, then $|S_N(t_2) - S_N(t)| \leq w(S_{N1},\delta)$, while if $t < 0$, then $|S_N(t) - S_N(t_1)| \leq w(S_{N2},\delta)$. In either case for t, t_2 such that $t_1 \le t \le t_2$ and $t_2 - t_1 \le \delta$,

$$\min\{|S_N(t) - S_N(t_1)|, |S_N(t_2) - S_N(t)|\}$$

$$\le \max\{w(S_{N1}, \delta), w(S_{N2}, \delta)\}.$$

The required result, (3.3.20), now follows from the way w' was defined in (3.2.2).

A further requirement of Theorem 3-2-6 is that $P(S^o(1) \ne S^o(1-)) = 0$. This though is immediate since $P(S_1^o \in C) = 1$.

Convergence of the finite-dimensional distributions is treated in exactly the same way as for Theorem 3-3-1. In view of Theorem 3-2-3, the proof of the theorem is complete.

Theorem 3-3-6 is the main result about the limiting distribution of ν_N contained in this thesis. As it stands, it is clearly not suitable for practical application. A more convenient expression is required to enable an approximation to be made for t_α such that $P\{\sup_t S^o(t) > t_\alpha\} = \alpha$. As far as the author is aware, this question has not been adequately resolved.

Although the question of adequate approximations to t_α seems to remain open, there has been development in this direction (and considerable research into apparently closely related questions). The basic concepts of such work are outlined (in a more general situation) by Belyayev (1972, in particular pp.15-16). Unfortunately much of the relevant research is being conducted in Russian universities and disseminated in technical reports. Besides these obstacles, it was felt that any detailed approach to the problem was bound to remove the thesis from the realm of rank tests to that of stochastic processes. Hopefully the author will be able to make some future contribution to this question under more favourable conditions.

The following result is of technical interest and may also be useful for obtaining the kind of approximation that we have just mentioned. In stating and proving the result, the simplifying assumptions that $\bar{a}_N = 0$, $\bar{c} = 0$ and $\sum c_i^2 = 1$ are made.

Proposition 3-3-7. Suppose that $\phi(u)$ defined by (A2) satisfies $\phi(u) + \phi(1 - u) = 0$ for $0 < u < 1$, is bounded over $(0,1)$ and satisfies (A3). Then, under H,

$$\text{(3.3.21)} \qquad \nu_N \to_{\mathcal{D}} \sup_{0 \le t \le 1} \{|S_1^0(t)|\}.$$

Proof. Consider the class of statistics of the form given by (3.3.2) with $a_N(N + 1 - i) + a_N(i) = \text{constant}$ for $i = 1,\ldots,N$. Such statistics are sometimes called odd-translation invariant. Under the assumption that $\bar{a}_N = 0$, the constant is 0. Then for S_N odd-translation invariant,

$$S_N((g_1(g_r)^k)(R)) = -S_N((g_r)^k(R)), k = 1,\ldots,N, \text{ and so}$$

$$\nu_N = \max_{1 \le k \le N} \{|S_N((g_r)^k(R))|\}$$

$$= \max_{g \in G_1} \{|S_N(g(R))|\}.$$

Now put $a_N^\phi(i) = N \int_{(i-1)/N}^{i/N} \phi(u)du$ so that $S_{N1}^\phi = \sum c_i\ a_N^\phi(R_i)$ is an odd-translation invariant statistic. Then assumptions (A1)-(A4) hold for $a_N^\phi(i)$ and $\phi(u)$. Therefore by Theorem 3-3-1, $S_{N1}^\phi(t) \to_{\mathcal{D}} S_1^o(t)$.

Scrutiny of Lemma 3-3-2 shows that it contains a proof that $a_{N1}(1 + [uN],t) \to S_t(\phi(u))$ *in* L_2 *uniformly for* $t \in [0,1]$. Together with Theorem 3-2-12 this implies that $E\{(S_{N1}(t) - S_{N1}^\phi(t))^2\} \to 0$ uniformly for $t \in [0,1]$, which in turn implies that $\sup_{0\le t\le 1} \{|S_{N1}(t) - S_{N1}^\phi(t)|\} \to 0$ in probability. This last statement implies that $\tau(S_{N1}(t),S_{N1}^\phi(t)) \to 0$ in probability (τ being the metric on $D([0,1])$). Applying Theorem 3-2-1, it follows that $S_{N1}(t)$ and $S_{N1}^\phi(t)$ have the same asymptotic distribution. Since we know the asymptotic distribution of $S_{N1}^\phi(t)$ - it arises from an odd-translation invariant statistic - it follows that

$$\nu_N \to_{\mathcal{D}} \sup_{0\le t\le 1} \{|S_1^o(t)|\},$$

as required.

Clearly in the condition $\phi(u) + \phi(1 - u) = 0$, the 0 may be replaced by a constant if the assumption $\bar{a}_N = 0$ is not made.

<u>Remark</u>. *A similar result holds for score generating functions satisfying* $\phi(u) = \phi(1 - u)$, $0 < u < 1$, when

$$\nu_N \to_{\mathcal{D}} \sup_{0\le t\le 1} \{S_1^o(t)\}.$$

The statistics S_N associated with ϕ by (A2) in this case are a generalization of the even-translation statistics which satisfy $a_N(i) = a_N(N + 1 - i)$ for $i = 1,\ldots,N$.

Using a result due to Fernique (1964), an approximation may be obtained for $P\{\sup_{0\le t\le 1} |S_1^0(t)| > x\}$, under certain conditions. The result is stated here as Lemma 3-3-8 and was apparently first proved by Marcus (1970) as the corollary to a more general result (which under the same conditions yields a stronger statement *than Fernique's* lemma). *The reader* is referred to pp.305-308 of the paper by Marcus for the proof and the particular statement of the lemma used here.

Lemma 3-3-8 Let $X(t)$ be a real-valued, separable Gaussian process on $[0,1]$ with $E\{(X(t))^2\}^{\frac{1}{2}} \le \Gamma$ and $E\{(X(t) - X(s))^2\} \le \psi(|t-s|)$, where ψ is assumed to be continuous and non-decreasing on $[0,1]$ *and satisfies* the following *conditions:*

(A) $\int_1^\infty \psi(e^{-x^2})dx < \infty$

(B) $\psi^2(h) \log {}^1/_h$ decreases monotonically as h decreases to zero from the right. *Then for* ℓ a fixed integer,

$$P\{\sup_{0\le t\le 1} |X(t)| \ge x(\Gamma + \frac{2^{\frac{1}{2}}}{2^{\frac{1}{2}}-1} \int_1^\infty \psi(\ell^{-u^2})du)\}$$
$$\le C\ell^2 \int_x^\infty e^{-u^2/2}\, du,$$

where $x \ge (4 \log \ell)^{\frac{1}{2}}$ and $C = 1 + {}^1/_3(4 \log \ell)/(4 \log \ell - 1) \times (2^{\frac{1}{2}} - 1)^{-1}$.

Quite how precise the approximation in Lemma 3-3-8 is remains unclear, but it does *indicate the kind of approximation* result that can reasonably be expected. It also indicates that the type of result obtained in Theorem 3-3-6 is not merely esoteric and the statistics ν_N can be put to use.

Of the four assumptions (A1)-(A4),(A3) is the only one likely to cause difficulty. An alternative characterization of the *functions ϕ which satisfy* (3.3.6) does not seem easy. The following result shows however that there are many functions which do satisfy that requirement.

Lemma 3-3-9. If ϕ is absolutely continuous on $[0,1]$, then it satisfies (A3).

Proof. The proof is outlined. Without loss of generality, it is assumed that $\phi(0) = 0$ so that by a well-known result (Hewitt and Stromberg, p.286, theorem 18.17), $\phi(x) = \int_0^x f(v)dv$, where $f \in L_1[0,1]$. Suppose that $0 \le s \le t \le 1$, so that is is required to find L continuous and nondecreasing such that, putting $\alpha = 1$,

$$(3.3.22) \qquad \|\phi(u - s) - \phi(u - t)\|^2 \le L(t) - L(s).$$

The left hand side of (3.3.22) can be written

$$\|\phi(u - s) - \phi(u - t)\|^2 = \int_0^s \{\phi(u + 1 - s) - \phi(u + 1 - t)\}^2 du +$$

$$\int_s^t \{\phi(u - s) - \phi(1 + u - t)\}^2 du + \int_t^1 \{\phi(u - s) - \phi(u - t)\}^2 du.$$

Consider the last of these terms:

$$\int_t^1 \{\phi(u - s) - \phi(u - t)\}^2 du = \int_t^1 \{\int_0^{u-s} f - \int_0^{u-t} f\}^2 du$$

$$= \int_t^1 \{\int_{u-t}^{u-s} f\}^2 du$$

$$\leq \int_t^1 (\int_{u-t}^{u-s} f^2) du \quad \text{(by the Schwarz inequality)}$$

$$\leq \int_0^1 (\int_{u-t}^{u-s} f^2) du \quad \text{(extending } f \text{ by periodicity)}$$

$$= \int_0^1 \int_{-1}^{u-s} f^2 - \int_0^1 \int_{-1}^{u-t} f^2.$$

Therefore taking $L_3(s) = - \int_0^1 (\int_{-1}^{u-s} f^2 \, dv) du$, we obtain

$$\int_t^1 (\phi(u - s) - \phi(u - t))^2 du \leq L_3(t) - L_3(s).$$

A similar argument applies to the first term. Since ϕ is bounded, the second term is less than or equal to $K(t-s)$ for some $K > 0$. If L_1, L_2 are the functions corresponding to these terms, then $L = L_1 + L_2 + L_3$ satisfies (3.3.22).

Example. As a particular example of the class of statistics defined by (3.3.1), consider the Mann-Whitney type statistic $\tilde{\xi}_N$. Clearly the scores of $N^{-1}\tilde{W}_N$ satisfy (A2) with $\phi(u) = u$, $u \in (0,1)$. Since $\phi(u)$ is absolutely continuous and the Mann-Whitney statistic is odd-translation invariant, Proposition 3-3-7 then applies for obtaining the limiting distribution of $\tilde{\xi}_N$.

The parameters for the limiting distribution of $\tilde{\xi}_N$ may then be obtained from Theorem 3-3-1, noting that

$$\int_0^1 \phi(u)\phi(u - s)du = \int_0^s u(1 + u - s)du + \int_s^1 u(u - s)du$$
$$= {}^1/_3 - {}^1/_2(s(1 - s)),$$

so that $(mnN)^{-\frac{1}{2}}\{\tilde{W}_{N1}(t) - \frac{1}{2}m(N + 1)\}$ converges in distribution to a Gaussian process with mean 0 and covariance kernel ${}^1/_{12} - {}^1/_2(s - t)(1 - (s - t))$ for $0 \le t \le s \le 1$. The relationship between W_N and $\tilde{W}_N$ implies that $(mnN)^{-\frac{1}{2}}\{W_{N1}(t) - \frac{1}{2}(m - n)(N + 1)\} \to_{\mathcal{D}} \xi(t)$ where $\xi(t)$ is defined in Proposition 2-2-1.

Obtaining the limiting distributions of ξ_N and ξ_{N1} in this way, while perfectly adequate, does not exploit the possibilities of a direct approach based on the expressions (2.1.11)-(2.1.13) and using results for the Kolmogorov-Smirnov test statistics. This alternative approach is outlined below.

In studying the limiting distributions of Kolmogorov-Smirnov type statistics, Hajék and Šidák (19[illegible], p.186) introduce the scores $\tilde{a}_N(i,t)$ defined as follows: for every $t \in [0,1]$ and $N \ge 1$,

$$\begin{aligned}\tilde{a}_N(i,t) &= 0, & i \le tN,\\ &= i - tN, & tN \le i < tN + 1,\\ &= 1, & i \ge tN + 1.\end{aligned}$$

Then $\tilde{a}_N(i,t)$ is continuous in t and consequently the processes

$$T_N(t) = (mnN)^{-\frac{1}{2}}\{n \sum_{i=1}^{m} \tilde{a}_N(R_i,t) - m \sum_{i=m+1}^{N} \tilde{a}_N(R_i,t)\}$$

determine probability distributions on $C([0,1])$.

Let us also introduce

$$W_N(t) = (mnN)^{-\frac{1}{2}}\{W_N(R) - \tfrac{1}{2}(m - n)(N + 1)\} \ (t \in [0,1]),$$

which defines a sequence of constant functions (as functions of t) and the continuous processes

(3.3.23) $$\xi_N(t) = W_N(t) - 2T_N(t).$$

Then, in view of (2.1.11) and (2.1.12),

(3.3.24) $$(mnN)^{-\frac{1}{2}}\{\xi_N - \tfrac{1}{2}(m - n)(N + 1)\} = f(\xi_N(t))$$

and

(3.3.25) $$(mnN)^{-\frac{1}{2}}\{\xi_{N1} - \tfrac{1}{2}(m - n)(N + 1)\} = f_1(\xi_N(t)),$$

where f, f_1 are defined by Proposition 2-2-1.

All that is required now is to show that $\xi_N(t) \to_{\mathcal{D}} \xi(t)$. First we verify that $\{\xi_N(t)\}$ is tight. The proof of Theorem 3-3-1 was principally a proof that $\{S_{N1}(t)\}$ is tight, and so it might be expected that showing $\{\xi_N(t)\}$ is tight will require extensive labour. In fact it does not, since in view of (3.3.23) and Theorem 3-2-4 all that is required is that $\{T_N(t)\}$ is tight. This is proved by Hajék and Šidák (1967, pp.187-189) in establishing the limiting distributions of the Kolmogorov-Smirnov test statistics.

To show that for $0 \le t_1 < \ldots < t_p \le 1$, $(\xi_N(t_1),\ldots,\xi_N(t_p)) \to_{\mathcal{D}} (\xi(t_1),\ldots,\xi(t_p))$ is merely a particular instance of the argument used to prove the corresponding result in Theorem 3-3-1 (and is straightforward). Proposition 2-2-1 then follows by Theorem 3-2-3 together with (3.3.24) and (3.3.25).

Notice that Lemma 3-3-8 applies to $\xi(t)$ since $E\{(\xi(t) - \xi(s))^2\} \le 4|t - s|$, and $\psi(h) \equiv 4h$ satisfies conditions A and B (for $h < e^{-1}$) of the lemma.

The motive for investigating the limiting distribution of $S_{N1}(t)$ is further strengthened by noticing that the limiting distributions of another important class of circular rank tests can be derived using Theorem 3-3-1. Again suppose that the convenient assumption that $\bar{a}_N = 0$, $\bar{c} = 0$ holds. Then consider the test statistic defined by

$$\eta_N(R) = N^{-1} \sum_{g\in G_1} \{S_N(g(r))\}^2 \tag{3.3.26}$$

and critical region $\eta_N(R) > k_\eta$. Obviously $\eta_N(R)$ does not depend on the cut-off point but it is not clear that it does not depend on the direction of measurement of angular displacement. This is proved by showing that $\tilde{\eta}_N(R) = 2\eta_N(R)$, where

$$\tilde{\eta}_N(R) = N^{-1} \sum_{g \in G} \{S_N(g(R))\}^2.$$

To show this - and at the same time obtain a useful expression for η_N - the anti-ranks $D(1),\ldots,D(N)$ are introduced. These are defined by $D(k) = j$ if and only if the k^{th} order statistic of the sample is X_j. Then $S_N(R) = \sum_{i=1}^{N} a_N(i)c_{D(i)}$, and $\tilde{\eta}_N(R)$ can be written as

$$N\tilde{\eta}_N = \sum_{k=1}^{N} \{(\sum_{i=1}^{N} a_N(i - k)c_{D(i)})^2 + (\sum_{i=1}^{N} a_N(k - i)c_{D(i)})^2\},$$

so that when this last expression is expanded, the coefficient of $c_{D(i)}c_{D(j)}$ in the first term on the right hand side is $\sum_{k=1}^{N} a_N(i + k)a_N(j + k)$ and in the second term it is $\sum_{k=1}^{N} a_N(-i + k)a_N(-j + k)$. That these two sums are equal follows from the periodic nature of a_N over period N. Hence $\tilde{\eta}_N(R) = 2\eta_N(R)$. It also follows that,

$$(3.3.27)\quad \eta_N = N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \{\sum_{k=1}^{N} a_N(i + k)a_N(j + k)\}c_{D(i)}c_{D(j)}.$$

There is an overlap between the class of tests defined by (3.3.26) and the class of two-sample tests introduced by Schach and mentioned in the introduction as having their test statistics defined by (1.4.1). In fact we shall be able to derive and extend the most interesting of Schach's results about circular tests without too much difficulty. Considering the test statistics written in the form (3.3.26), this class of tests may also be compared to the class of Cramér-von Mises goodness-of-fit tests introduced by Hájek and Šidák (1967, p.103). Finally, note that

$$(3.3.28) \qquad \eta_N = \int_0^1 (J_{N1}(t))^2 dt.$$

In order to apply Theorem 3-3-1 to obtaining the limiting distribution of η_N, a technical result is required.

<u>Lemma 3-3-10</u>. The functional $z \to \int_0^1 z^2$ is continuous for $z \in D([0,1])$.

<u>Proof</u>. Two facts from Billingsley (1968) need to be recalled. Firstly, if $z \in D$, then it is bounded, so that if $z_n \to z$ in (D,τ), then the z_n are uniformly bounded. Secondly if $z_n \to z$ in (D,τ), then $z_n(t)$ converges to $z(t)$ for continuity points of $z(t)$ and hence $z_n(t)$ converges to $z(t)$ almost everywhere with respect to Lebesgue measure on $[0,1]$. Clearly therefore if $z_n \to z$ in (D,τ), then $\{z_n^2\}$ are uniformly bounded and $(z_n(t))^2$ converges to $(z(t))^2$ almost everywhere. Lebesgue's dominated convergence theorem then implies that if $z_n \to z$, it follows that $\int_0^1 (z_n(t))^2 dt \to \int_0^1 (z(t))^2 dt$, as required for continuity.

Applying Lemma 3-3-10, Theorem 3-2-2 and Theorem 3-3-1, assuming that $\sum_{i=1}^{N} c_i^2 = 1$, it follows that

$$(3.3.29) \qquad \eta_N \to_{\mathcal{D}} \int_0^1 (S_1^0(t))^2 dt.$$

Note that the *scaling assumptions that have been made* are strictly for convenience; the general result is

$$\{N\sum(c_i - \bar{c})^2\}^{-1} \sum_{g\epsilon G_1} (S_{N1}(g(r)) - N\bar{a}_N \bar{c})^2$$

$$\rightarrow_{\mathcal{D}} \int_0^1 (S_1^o(t))^2 dt.$$

The expression (3.3.29) for the asymptotic distribution of η_N can be improved by using the expansion for a Gaussian process mentioned as Theorem 3-2-16. Since $S_1^o(t)$ is stationary, the covariance kernel may be written as $R_1(t-s)$ for $-1 \le t - s \le 1$ and $R_1(u)$ - using the new definition - can be extended to the real line by defining $R_1(u + 2k) = R_1(u)$, $k = \pm 1, \pm 2, \ldots$ Clearly $R_1(u)$ is an even function and since $R_1(1-u) = \langle\phi, S_{1-u}(\phi)\rangle$, $R_1(u)$ has period 1. If $R_1(u)$ is expanded in terms of $\exp\{2\pi i n u\}$, $n = 0, \pm 1, \pm 2, \ldots$, the Fourier coefficient of $\exp\{-2\pi i n u\}$ is denoted by b_n (equals b_{-n}).

<u>Theorem 3-3-11</u>. Under H, the asymptotic distribution of η_N is equal to the distribution of $\sum_{k=1}^{\infty} b_k \chi^2_{2,k}$, where $\{\chi^2_{2,k}; k=1,2,\ldots\}$ is a sequence of independent χ^2_2 random variables.

<u>Proof</u>. In order to apply Theorem 3-2-16, the eigenvalues and orthonormal eigenfunctions of $R_1(s,t)$ are required. Suppose that $\phi_r(t) = \exp(2\pi i r t)$, then

$$b_r = \int_0^1 R_1(u)\exp(2\pi i r u)du$$

$$= \int_0^1 R_1(t - s)\exp(2\pi i r(t - s))dt \qquad (0 \le s \le 1),$$

so that

$$b_r \exp(2\pi irs) = \int_0^1 R_1(t - s)\exp(2\pi irt)dt$$

$$= \int_0^1 R_1(s,t)\exp(2\pi irt)dt$$

and hence $\phi_r(t)$ is an eigenfunction corresponding to the eigenvalue b_r.

In fact it is preferable to work with the real orthonormal functions given by $2^{\frac{1}{2}} \cos 2\pi rt$, $2^{\frac{1}{2}} \sin 2\pi rt$ $(r = 1,2,..)$ together with $I_{[0,1]}(t) = 1$ for $t \in [0,1]$, $I_{[0,1]}(t) = 0$ otherwise. This is possible since, for instance, $2^{\frac{1}{2}} \cos 2\pi rt = 2^{-\frac{1}{2}}\{\exp(2\pi irt) + \exp(-2\pi irt)\}$ and $\exp(2\pi irt), \exp(-2\pi irt)$ are conjugate eigenfunctions corresponding to the same eigenvalue b_r. Similarly for $2^{\frac{1}{2}} \sin 2\pi rt$. Another observation to be made is that $b_o = \int_0^1 R_1(u)du = \int_0^1 \int_0^1 \phi(x)\phi(x + u)dx\, du = \{\int_0^1 \phi(x)dx\}^2 = 0$ (because $\bar{\phi} = 0$).

In view of the preceding remarks, Theorem 3-2-16 implies that

$$(3.3.30) \qquad \int_0^1 (S_1^o(t))^2 dt = \sum_{k=-\infty}^{\infty} b_k J_k^2,$$

almost everywhere, where the J_k are mutually independent, normally distributed random variables with mean 0 and variance 1. Grouping together terms for $k = \pm 1, \pm 2, \ldots$ together with (3.3.29) yields the required result.

Some comment on the relationship between Theorem 3-3-11 and the results obtained by Schach (1970) concerning the asymptotic distribution of statistics of the form (1.4.1) should be made. Theorem 3-3-11 is implied by Schach's results when two-sample tests are being considered (and Schach's result is more general). The more general case where the c_i are subject only to (A1) is not covered by Schach's results, but the changes necessary to his work to cover that case do not appear to be difficult. What does seem rather convenient about our derivation of the result through Theorem 3-3-1 is that the asymptotic distribution of η_N has been derived using standard results about the asymptotic distribution of S_N, results which allow an extension of Theorem 3-3-11 to cover contiguous alternatives. This will be done later.

At first glance Theorem 3-3-11 appears so much weaker than Schach's results for two-sample circular tests that it will not cover many test statistics of interest. In fact this does not seem to be true. This is indicated by the next piece of work showing that the asymptotic null distributions of the test statistics of the locally most powerful invariant (under G) tests for H against rotation alternatives are covered by Theorem 3-3-11.

Consider the two-sample situation where K (the alternative hypothesis) consists of $F(x) = \int_0^x f(t)dt$ satisfying the conditions listed below, together with $G_\phi(y) = \int_0^y f(t-\phi)dt$ for $0 < |\phi| \leq \phi_0$ and $0 < \phi_0 < \pi$. The conditions on $f(x)$ are those required by Schach (1969b, p.1791, (1.1)-(1.3)) for the existence of the test statistics. These conditions are:

(3.3.31) $f(x) > 0$ almost everywhere (on $[0,2\pi)$), and not a constant;

(3.3.32) $f'(x)$ exists and is continuous for all x;

$$(3.3.33) \qquad \int_0^{2\pi} \{f'(x)/f(x)\}^2 f(x)dx = \inf(f) < \infty,$$

where $\inf(f)$ is the Fisher's information. Then Schach (1969b) has shown that the locally most powerful rank test invariant under G is given by

$$(3.3.34) \quad \mu_N = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} E\{\psi(U^{(i+k)})\psi(U^{(j+k)})\}Z_{Ni}Z_{Nj},$$

where

$$(3.3.35) \quad \psi(x) = f'(F^{-1}(x))/f(F^{-1}(x)), \quad 0 \le x \le 1,$$

and $U^{(i)}$ is the ith order statistic from a sample of N independent random variables uniformly distributed over $[0,1]$ (modulo N arithmetic is applied to the superscripts in (3.3.34)).

To find the limiting distribution of μ_N, we compare it with η_N obtained from (3.3.26) when the underlying linear rank statistic is

$$(3.3.36) \qquad S_N = \left(\frac{N}{mn}\right)^{\frac{1}{2}} \sum_{k=1}^{N} E\{\psi(U^{(k)})\}Z_{Nk}.$$

These are actually the test statistics for the locally most powerful rank tests for H against a (one-sided) shift in location for the two-sample problem on the real line (see Hájek and Šidák (1967), theorem II.4.4a on pp.67-68).

Comparison of μ_N and η_N is made using corollary 3.4 of Schach (1970, p.50). In order to use this result though, the statistics μ_N and η_N have to be written in the form (1.4.1), that is

$$T_N = N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} h_N(\tfrac{i-j}{N}) Z_{Ni} Z_{Nj},$$

where $\{h_N(x)\}$ satisfy the following requirements:

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

(ii) h_N is step function constant on the intervals $(\frac{2k-1}{N}, \frac{2k+1}{N})$, $k = 0, \pm 1, \pm 2, \ldots, \pm N$,

(iii) $\sum_{k=1}^{N} h_N(k/N) = 0$ for all N.

For this reason the statistics are written as follows:

(3.3.37) $$N^{-2}\mu_N - N^{-2}m^2 \inf(f) = N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} h_N(\tfrac{i-j}{N}) Z_{Ni} Z_{Nj}$$

and

(3.3.38) $$(mnN^{-2})\eta_N = N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{h}_N(\tfrac{i-j}{N}) Z_{Ni} Z_{Nj},$$

where

$$h_N(x) = N^{-1} \sum_{k=1}^{N} E\{\psi(U^{(i+k)})\psi(U^{(j+k)})\} - N^{-1} \inf(f)$$

and

$$\bar{h}_N(x) = N^{-1} \sum_{k=1}^{N} E\{\psi(U^{(i+k)})\}E\{\psi(U^{(j+k)})\}$$

for

$$(2(i - j) - 1)/2N < x \le (2(i - j) + 1)/2N.$$

Schach (1969b, pp.1795-1796) has shown that the right hand side of (3.3.37) satisfies conditions (i)-(iii). The only condition on $\bar{h}_N$ requiring verification is (iii). But since by theorem II.4.3 of Hájek and Sidák (1967, pp.66), $\sum_{i=1}^{N} E\{\psi(U^{(i)})\} = 0$, therefore $\sum_{k=1}^{N} \bar{h}_N(k/N) = N^{-1} \sum_{k=1}^{N} \sum_{j=1}^{N} E\{\psi(U^{(j)})\}E\{\psi(U^{(j+k)})\} = 0$.

Schach's result may now be stated. The proof is found on pp.49-50 of Schach (1970).

Lemma 3.3.12. *If $\{h_N\}$ and $\{\bar{h}_N\}$ are two sequences which satisfy conditions (i)-(iii) and furthermore $\|h_N - \bar{h}_N\| \to 0$, using the L_2 norm, and $h_N(0) - \bar{h}_N(0) \to 0$, then $E\{(T_N - \bar{T}_N)^2\} \to 0$ as $N \to \infty$, where $T_N, \bar{T}_N$ are the statistics corresponding to h_N and $\bar{h}_N$ through (1.4.1).*

The limit of $h_N(0)$ is inf(f) as shown by Schach (1969, p.1795). Since $\bar{h}_N(0) = N^{-1} \sum_{k=1}^{N} E\{(\psi(U^{(k)}))^2\}$, $\bar{h}_N(0) \to \inf(f)$ in view of the asymptotic normality of the linear rank statistics defined by (3.3.36) ($\bar{h}_N(0)$ is a limiting variance).

Suppose that $\psi(x)$ is continuous on [0,1]. Schach (1969b) shows that if $h(x) = \int \psi(t)\psi(t + x)dt$, then $h_N \to h$ in the L_2 norm. The same kind of argument shows that $\bar{h}_N \to h$ in the L_2 norm and is briefly outlined here.

Define $\eta_N(x,y) = E\{\psi(U^{(i)})\}E\{\psi(U^{(j)})\}$, for $(2i - 1)/2N < x \leq (2i + 1)/2N$ and $(2j - 1)/2N < y \leq (2j + 1)/2N$. Now if $f_N(u,i)$ is the density of $U^{(i)}$ (from a sample of size N), then the distributions corresponding to the densities $f_N(u, < Nx - \frac{1}{2}>)$ converge weakly to a distribution having all its mass at x - this is an application of lemma 3 of Hoeffding (1953). Then by the Helly-Bray theorem, $\int_0^1 \psi(u)f_N(u,<Nx - \frac{1}{2}>)du \to \psi(x)$, so that $\eta_N(x,y) \to \psi(x)\psi(y)$. Then since $\eta_N(x,y)$ is uniformly bounded, $\int_0^1 \eta_N(t,t + x)dt \to \int_0^1 \psi(t)\psi(t + x)dt$ in L_2 by Lebesgue's dominated convergence theorem and hence $\bar{h}_N \to h$ in L_2.

Lemma 3-3-12 may now be applied to the statistics defined by the left hand sides of (3.3.37) and (3.3.38). Since the asymptotic distribution of $(mnN^{-2})\eta_N$ is already known from Theorem 3-3-11, the asymptotic distribution of $N^{-2}\mu_N - N^{-2}m^2\inf(f)$ is the same (by the lemma) and from this the asymptotic distribution of $N^{-2}\mu_N$ may be derived. The result as we have derived it is given as Corollary 3-3-13 (a corollary to Theorem 3-3-11).

Corollary 3-3-13. If ψ is absolutely continuous and $^m/_N \to \lambda$, $0 < \lambda < 1$, then $N^{-2}\mu_N$ has under H an asymptotic distribution equal to the distribution of $\lambda^2 \inf(f) + \lambda(1 - \lambda)\sum_{k=1}^{\infty} b_k \chi^2_{2,k}$ where $\{\chi^2_{2,k} : k = 1,2,\ldots\}$ is a sequence of independent χ^2_2 variables and $h(x) = \sum_{k=-\infty}^{\infty} b_k \exp(2\pi ikx)$.

The statement of this result in theorem 4.2 of Schach (1969b) is slightly stronger in that the requirement that ψ be absolutely continuous (required to ensure that (A3) is satisfied) is replaced by the requirement of continuity. It seems safe to say that any reasonable locally most powerful invariant rank test for which ψ is continuous will also have ψ absolutely continuous.

Let us also take the opportunity to make quite clear that an analysis of η_N for its own sake is more naturally conducted in ℓ_2 (as Schach did) and such an analysis may be expected to yield stronger statements about the limiting distribution of η_N. It does seem a convenient piece of unification, though, to derive a basic result (Theorem 3-3-1) and then use that result to establish the limiting distributions of a wide range of interesting test statistics. Essentially this approach is shorter as well, since only technical material about D([0,1]) is used.

The following example complements the previous one in this section (after Lemma 3-3-9) since it discusses the statistic η_N obtained when the underlying linear rank statistic is $\tilde{W}_N$.

Example. Consider the statistic η_N obtained from (3.3.26) by taking $S_N = (mnN)^{-\frac{1}{2}} \sum_{i=1}^{N} (i - (N+1)/2) Z_{Ni}$. The two-sample Cramér-von Mises test statistic is defined by

$$M_N^2 = mnN^{-1} \int_{-\infty}^{\infty} \{F_m(x) - G_n(x)\}^2 d\{N^{-1}(mF_m(x) + nG_n(x))\}.$$

It will now be shown that $\eta_N = M_N^2 - S_N^2$. From (2.1.10) it follows that

$$\text{(3.3.39)} \quad \tilde{W}_N((g_r)^i(R)) = \tilde{W}_N(R) + \text{[illegible]}\{F_m(X^{(i)}) - G_n(X^{(i)})\} \qquad (i = 1,\ldots,N).$$

Notice that $\sum_{i=1}^{N} \tilde{W}_N((g_r)^i(R)) = \sum_{i=1}^{N} (1 + \ldots + N)Z_{Ni} = \frac{1}{2}mN(N + 1)$, so that, summing both sides of (3.3.39),

$$(3.3.40)\quad \tfrac{1}{2}mN(N + 1) = N\tilde{W}_N(R) + mn \sum_{i=1}^{N} \{F_m(X^{(i)}) - G_n(X^{(i)})\}.$$

Now, by definition,

$$\begin{aligned}
N\eta_N &= \sum_{i=1}^{N} \{S_N((g_r)^i(R))\}^2 \\
&= \sum_{i=1}^{N} \{S_N(R) + (mnN^{-1})^{\frac{1}{2}}(F_m(X^{(i)}) - G_n(X^{(i)}))\}^2 \\
&= NS_N^2 + 2(mnN^{-1})^{\frac{1}{2}}S_N \sum_{i=1}^{N}\{F_m(X^{(i)}) - G_n(X^{(i)})\} \\
&\qquad + mnN^{-1} \sum_{i=1}^{N} \{F_m(X^{(i)}) - G_n(X^{(i)})\}^2 \\
&= NS_N^2 + 2S_N\{(mnN)^{-\frac{1}{2}}mn \sum_{i=1}^{N} (F_m(X^{(i)}) - G_n(X^{(i)}))\} \\
&\qquad + mn \int_{-\infty}^{\infty} \{F_m(x) - G_n(x)\}^2 d\{N^{-1}(mF_m(x) \\
&\qquad\qquad + nG_n(x))\} \\
&= NS_N^2 - 2S_N\, N(mnN)^{-\frac{1}{2}}(\tilde{W}_N - \tfrac{1}{2}m(N + 1)) + NM_N^2,
\end{aligned}$$

by (3.3.40),

$$= NS_N^2 - 2NS_N^2 + NM_N^2 = N(M_N^2 - S_N^2),$$

and the required result follows.

While in view of the results in section 2.1, some relationship between η_N and M_N^2 was to be expected, the actual form of that relationship is a little surprising.

The final contribution to this section is to examine the limiting distributions of ν_N and η_N for the location alternatives $K(d_\nu)$ defined in section 3.2 (see (3.3.8)). For the purpose of these introductory remarks, we recall from section 3.2 that we are dealing with sequences of hypotheses and test statistics indexed by ν (which has nothing to do with the test statistic ν_N).

It is assumed that d_ν in the definition of $K(d_\nu)$ satisfy (3.2.9), (3.2.10) together with the following assumption:

$$(3.3.41)\quad \{\sum_i (c_{\nu i} - \bar{c}_\nu)^2\}^{-\frac{1}{2}}\{\sum_i (c_{\nu i} - \bar{c}_\nu)(d_\nu(i) - \bar{d}_\nu)\} \to b_{cd} > 0.$$

The index ν will be omitted for the rest of the discussion. The following assumptions apply for the rest of this section: $\bar{a}_N = 0$, $\bar{c} = 0$ and $\sum c_i^2 = 1$ for all N.

Theorem 3-3-14. Suppose that (A1)-(A4) are satisfied. Then under $K(d)$, $S_{N1}(t)$ converges in distribution to the Gaussian process $S_1^{oo}(t)$ defined on $[0,1]$ with

$$(3.3.42)\quad E\{S_1^{oo}(t)\} = b_{cd}\int S_t(\phi(u))\psi(u)du$$

($\psi(u)$ defined by $f'(F^{-1}(u))/f(F^{-1}(u))$ and

$$(3.3.43)\quad \operatorname{cov}\{S_1^{oo}(s),\ S_1^{oo}(t)\} = \int S_s(\phi(u))S_t(\phi(u))du.$$

Proof. Because the $K(d)$ are contiguous alternatives, Lemma 3-2-15 together with Theorem 3-2-5 shows that $\{S_{N1}\}$ is tight. The convergence of the finite-dimensional distributions follows from Theorem 3-2-14 in similar fashion to the proof of Theorem 3-3-1.

Now the asymptotic distribution of η_N can be obtained using Theorem 3-3-14.

<u>Corollary 3-3-15</u>. Suppose that (A1)-(A4) are satisfied and that $E\{S_1^{OO}(t)\} = \sum_{k=-\infty}^{\infty} r_k \exp(2\pi ikt)$. Then

$$\eta_N \to_D \sum_{k=1}^{\infty} b_k \, \chi^2_{2,k;\delta(k)}, \tag{3.3.44}$$

where $\{\chi^2_{2,k;\delta(k)} : k = 1,2,\dots\}$ is a sequence of independent non-central χ^2_2 random variables with non-centrality parameters $\delta(k) = |r_k|^2/(2b_k^2)$.

<u>Proof</u>. First observe that $b_k = 0$ implies $r_k = 0$, since

$$b_k = \int_0^1 R_1(x)\exp(-2\pi ikx)\,dx$$

$$= \int_0^1 \int_0^1 \phi(t)\phi(t+x)\exp(-2\pi ikx)\,dt\,dx$$

$$= \left|\int_0^1 \phi(t)\exp(-2\pi ikt)\,dt\right|^2$$ (and ϕ is a real function), while

$$r_k = b_{cd} \int_0^1 \int_0^1 \phi(u-t)\psi(u)\exp(-2\pi ikt)\,du\,dt$$

$$= b_{cd} \int_0^1 \psi(u)\exp(-2\pi iku)\left(\int_0^1 \phi(u-t)\exp(2\pi ik(u-t))\,dt\right)du$$

$$= b_{cd}\left(\int_0^1 \psi(u)\exp(-2\pi iku)\right)\overline{\left(\int_0^1 \phi(t)\exp(-2\pi ikt)\,dt\right)}.$$

Theorem 3-2-16 can be applied to the process $S_1^{OO}(t) - E\{S_1^{OO}(t)\}$ to obtain an expansion for that process in terms of the orthonormal system $I_{[0,1]}(t)$, $2^{\frac{1}{2}}\cos 2\pi kt$, $2^{\frac{1}{2}}\sin 2\pi kt$. Thus

$$(3.3.45)\quad S_1^{OO}(t) - E\{S_1^{OO}(t)\} = \sum_{k=1}^{\infty} b_k^{\frac{1}{2}}\, J_{1k}(2^{\frac{1}{2}}\cos 2\pi kt)$$

$$+ \sum_{k=1}^{\infty} b_k^{\frac{1}{2}}\, J_{2k}(2^{\frac{1}{2}}\sin 2\pi kt),$$

where J_{1k}, J_{2k} are mutually independent, normally distributed random variables with mean 0 and variance 1.

Now, $E\{S_1^{OO}(t)\} = \sum_{k=-\infty}^{\infty} r_k \exp(2\pi ikt)$

$$= \sum_{k=1}^{\infty}\{f_k(2^{\frac{1}{2}}\cos 2\pi kt) + g_k(2^{\frac{1}{2}}\sin 2\pi kt)\},$$

where $f_k = 2^{-\frac{1}{2}}(r_k + r_k)$ and $g_k = i2^{-\frac{1}{2}}(r_k - r_{-k})$.

Hence from (3.3.45) it follows that

$$S_1^{OO}(t) = \sum_{k=1}^{\infty}\{b_k^{\frac{1}{2}}\, J_{1k}(2 \qquad + b_k^{\frac{1}{2}}\, J_{2k}(2^{\frac{1}{2}}\sin 2\pi kt)\}$$

$$+ \sum_{k=1}^{\infty}\{f_k(2^{\frac{1}{2}}\cos 2\pi kt) + g_k(2^{\frac{1}{2}}\sin 2\pi kt)\}$$

$$= \sum_{k=1}^{\infty} b_k^{\frac{1}{2}}(J_{1k} + f_k b_k^{-\frac{1}{2}})(2^{\frac{1}{2}}\cos 2\pi kt) +$$

$$\sum_{k=1}^{\infty} b_k^{\frac{1}{2}}(J_{2k} + g_k b_k^{-\frac{1}{2}})(2^{\frac{1}{2}}\sin 2\pi kt).$$

Putting $\eta_{1k} = J_{1k} + f_k b_k^{-\frac{1}{2}}$ and $\eta_{2k} = J_{2k} + g_k b_k^{-\frac{1}{2}}$, it then follows that

$$(3.3.46)\quad S_1^{oo}(t) = \sum_{k=1}^{\infty} b_k^{\frac{1}{2}}\{\eta_{1k}(2^{\frac{1}{2}}\cos 2\pi kt) + \eta_{2k}(2^{\frac{1}{2}}\sin 2\pi kt)\},$$

where the η_{ik} are independent, normally distributed with variance 1 and $E(\eta_{1k}) = (r_k + r_{-k})/(2b_k)^{\frac{1}{2}}$, $E(\eta_{2k}) = i(r_k - r_{-k})/(2b_k)^{\frac{1}{2}}$.

The series expansion in (3.3.46) converges in the mean to $S_1^{oo}(t)$ with probability one. Parseval's identity for the complete orthonormal system $\{\exp(2\pi ikt)\}$ then implies that

$$\int_0^1 (S_1^{oo}(t))^2 dt = \sum_{k=1}^{\infty} b_k\ \eta_{1k}^2 + \sum_{k=1}^{\infty} b_k\ \eta_{2k}^2,$$

with probability one. Grouping together terms - together with Theorems 3-2-2 and 3-3-14 - gives the required result.

Remarks. (i) Expressions for the limiting distributions of ν_{N1} and ν_N analogous to (3.3.17) and (3.3.18) are immediate consequences of Theorem 3-3-14.

(ii) The limiting distribution of $\mu_{N(\nu)}$ under the sequence of alternatives $K(d_\nu)$ can be obtained from Corollary 3-3-15. As was the case with H, the asymptotic distributions of $(mnN^{-2})\eta_N$ and $N^{-2}\mu_N - N^{-2}m^2\inf(f)$ are the same under the alternatives K(d) since $|(mnN^{-2})\eta_N - N^{-2}\mu_N + N^{-2}m^2\inf(f)| \to 0$ in probability under K(d). This is a consequence of contiguity together with Lemma 3-2-15.

(iii) There is a certain generality latent in the results derived from Theorem 3-3-14 which it might be worth indicating. The density f is a density on C, identified with $[0,2\pi)$, and as such depends implicitly on the cut-off point and direction of measurement of angular displacement. The statistics ν_N and η_N are invariant under these arbitrary choices and hence their limiting distributions under the sequence of alternatives $K(d_\nu)$ will be the same in all the resulting cases. This implies that the results hold for a family of distributions determined by f (which is larger than for the case of an arbitrary rank test).

3.4 Exact Bahadur efficiencies.

The main result of this section is not particularly difficult to derive, and yet is rather interesting. Stated concisely, the result is that if S_N is a two-sample linear rank statistic, then ν_N (and for that matter ν_{N1}) can never be less efficient than S_N as measured by their Bahadur efficiency.

First some preliminaries about Bahadur efficiency are required. A full discussion of the motivation behind calculating the exact slope of a sequence of test statistics is contained in Bahadur (1967). The actual computation of the exact slope is in general performed as follows. Suppose that $X_1,\ldots,X_N$ is a random sample whose joint distribution is indexed by $\theta \in \Omega$, Ω an arbitrary set, and it is required to test $H : \theta = \theta_o$ against $K : \theta \in \Omega-\{\theta_o\}$. If T_N is a statistic based on this sample for which H is rejected when $T_N > k_N$, then it is assumed that:

(a) There exists a function $b(\theta)$, $0 < b(\theta) < \infty$

for $\theta \in \Omega-\{\theta_o\}$ such that $N^{-1}T_N \to b(\theta)$ with probability one.

(b) There is a continuous function $I(x)$ such that for any sequence $\{x_N\}$ of constants converging to x

$$\lim_{N\to\infty} \{-N^{-1} \log P_{\theta_o}(T_N \geq Nx_N)\} = I(x).$$

Then the exact slope of the sequence of test statistics $\{T_N\}$ evaluated at $\theta^* \in \Omega-\{\theta_o\}$ is $2I(b(\theta^*))$. The function $I(x)$ is referred to as the large deviation of $\{T_N\}$.

If the two sequences of test statistics $\{T_N^{(1)}\}$ and $\{T_N^{(2)}\}$ satisfy (a) and (b), then the exact Bahadur efficiency of $\{T_N^{(1)}\}$ relative to $\{T_N^{(2)}\}$, evaluated at θ^*, is defined as

$$e(T_N^{(1)},T_N^{(2)}) = I_1(b_1(\theta^*))/I_2(b_2(\theta^*)).$$

For the present discussion a further piece of notation is required. The empirical distribution function of the combined sample in the two-sample situation is denoted by $H_N(x)$, so that $H_N(x) = (^m/_N)F_m(x) + (^n/_N)G_n(x)$. Then consider the random function on $[-1,1]$ defined by

$$(3.4.1)\quad \hat{S}_N(t) = {}^m/_N\int_0^1 a_N(1+[(u-t)N])dF_m(H_N^{-1}(u)), t \in [0,1]$$

$$= {}^m/_N\int_0^1 a_N(1+[-(u+t)N])dF_m(H_N^{-1}(u)), t \in [-1,0).$$

It follows that

$$(3.4.2) \qquad N^{-1}\nu_N = \sup_{-1\le t\le 1} \{\hat{S}_N(t)\}.$$

Our approach to establishing the almost sure limit of $\hat{S}_N(t)$ will be to suitably modify the proof of theorem 1 of Hájek (1974, pp.76-77) which deals with the limit of $N^{-1}S_N$. It is assumed that F and G define densities f and g respectively (with respect to Lebesgue measure on the real line). As usual, suppose that $^m/_N \to \lambda$, $0 < \lambda < 1$. Put $H(x) = \lambda F(x) + (1 - \lambda)G(x)$ and then define $\bar{F}(x) = F(H^{-1}(x))$ and also $\bar{f}(x) = \frac{d\bar{F}(x)}{dx}$.

Proposition 3-4-1. Suppose that a_N satisfies (A2) of section 3.3 and in addition $a_N(1 + [uN])$ is of uniformly bounded variation for closed subintervals of (0,1). Then $\hat{S}_N(t) \to \hat{S}(t)$ uniformly for $t \in [-1,1]$ with probability one, where

$$(3.4.3) \qquad \hat{S}(t) = \lambda \int_0^1 \nu_t(\phi(u))\bar{f}(u)du$$

(ν_t is defined in section 3.3).

Proof. The first step is to reduce the problem to the case where $a_N(1 + [uN])$ is of uniformly bounded variation on [0,1]. For $\delta > 0$, it is possible to find $0 < \varepsilon_1 < \varepsilon_2 < 1$ so that the scores

$$\tilde{a}_N(1 + [uN]) = a_N(1 + [uN]) \qquad u \in [\varepsilon_1,\varepsilon_2]$$

$$= 0 \qquad \text{otherwise}$$

satisfy $N^{-1} \sum_{i=1}^{N} |a_N(i) - \bar{a}_N(i)| < \delta$ for all $N \geq N_o$ and also

$$\bar{\phi}(u) = \phi(u) \qquad u \in [\varepsilon_1, \varepsilon_2]$$
$$= 0 \qquad \text{otherwise}$$

satisfies $\int_0^1 |\phi(u) - \bar{\phi}(u)| du < \delta$.

This last assertion is justified as follows. Since $\phi \in L_1([0,1])$, therefore it follows that $\lim_{\varepsilon \to 0} \int_0^{\varepsilon} |\phi(u)| du = 0$ and $\lim_{\varepsilon \to 0} \int_{1-\varepsilon}^{1} |\phi(u)| du = 0$ and consequently we can choose $0 < \varepsilon_1 < \varepsilon_2 < 1$ so that $\int_0^{\varepsilon_1} |\phi(u)| du < \delta/_4$ and $\int_{\varepsilon_2}^{1} |\phi(u)| du < \delta/_4$. Choose an integer N_o such that for $N \geq N_o$, $\int_0^1 |a_N(1 + [uN]) - \phi(u)| du < \delta/_4$ (since (A2) is sufficient to imply that $a_N(1 + [uN]) \to \phi(u)$ in $L_1([0,1])$, that is $\int_0^1 |a_N(1 + [uN]) - \phi(u)| du \to 0$ as $N \to \infty$). Then,

$$\int_0^1 |\phi(u) - \bar{\phi}(u)| du = \int_0^{\varepsilon_1} |\phi(u)| du + \int_{\varepsilon_2}^{1} |\phi(u)| du < \delta/_2.$$

Furthermore, $\int_0^1 |a_N(1 + [uN]) - \bar{a}_N(1 + [uN])| du = \int_0^{\varepsilon_1} |a_N(1 + [uN])| du + \int_{\varepsilon_2}^{1} |a_N(1 + [uN])| du$. But $\int_0^{\varepsilon_1} |a_N(1 + [uN])| du \leq \int_0^{\varepsilon_1} |a_N(1 + [uN]) - \phi(u)| du + \int_0^{\varepsilon_1} |\phi(u)| du < \delta/_2$ for $N \geq N_o$ and similarly for $\int_{\varepsilon_2}^{1} |a_N(1 + [uN])| du$. Hence

$$N^{-1} \sum_{i=1}^{N} |a_N(i) - \bar{a}_N(i)| = \int_0^1 |a_N(1 + [uN]) - \bar{a}_N(1 + [uN])| du$$
$$< \delta \quad \text{for } N \geq N_o.$$

Now if $\bar{S}_N(R) = \sum_{i=1}^{N} \bar{a}_N(i) Z_{Ni}$, then $|\bar{S}_N(R) - S_N(R)| < \delta N$ for any $R = (R_1, \ldots, R_N)$ and $N \geq N_0$. Furthermore, $\int_0^1 |V_t(\bar{\phi}(u) - \phi(u))| du = \int_0^1 |\bar{\phi}(u) - \phi(u)| du$ and $\lambda \bar{f}(u) \leq 1$ so that $|\lambda \int_0^1 V_t(\bar{\phi}(u)) \bar{f}(u) du - \lambda \int_0^1 V_t(\phi(u)) \bar{f}(u) du| < \delta$ uniformly for $t \in [-1,1]$. It now follows that it is sufficient to establish the proposition for $a_N(1 + [uN])$ of uniformly bounded variation on $[0,1]$.

For $t \in [-1,1]$,

$$(3.4.4) \quad \hat{S}_N(t) = {}^m/_N \int_0^1 V_t(a_N(1 + [uN])) dF_m(H_N^{-1}(u))$$

$$= {}^m/_N \int_0^1 V_t(\phi) \bar{f} du + {}^m/_N \int_0^1 V_t(a_N - \phi) \bar{f} du +$$

$${}^m/_N \int_0^1 V_t(a_N) d(\bar{F}_m - \bar{F})$$

(where $\bar{F}_m = F_m(H_N^{-1})$). Now because

$$|{}^m/_N \int_0^1 V_t(a_N - \phi) \bar{f}\, du| \leq {}^m/_N \| a_N - \phi \| \| \bar{f} \|,$$

by the Schwarz inequality, therefore

$$|{}^m/_N \int_0^1 V_t(a_N - \phi) \bar{f}\, du| \to 0$$

uniformly for $t \in [-1,1]$ by (A2). Also, there exists an N_0 such that

$$|{}^{m}/_{N} \int_0^1 V_t(a_N)d(\bar{F}_m - \bar{F}) = |{}^{m}/_{N} \int(\bar{F}_m - \bar{F})d(V_t(a_N))|$$

$$\leq {}^{m}/_{N} \sup_{0<u<1} \{|\bar{F}_m(u) - \bar{F}(u)|\}M,$$

where for $N \geq N_0$, $V_0^1(a_N(1+[uN])) < M$. By the Glivenko-Cantelli lemma $\sup_{0<u<1}\{|\bar{F}_m(u) - \bar{F}(u)|\} \to 0$ with probability one. Since the last two terms on the right hand side of (3.4.4) tend to 0 uniformly for $t \in [-1,1]$, the proposition is proved.

Corollary 3-4-2. As an immediate consequence of Proposition 3-4-1, it follows that

$$(3.4.5) \quad N^{-1}\nu_N \to \sup_{-1\leq t\leq 1} \{\lambda \int_0^1 V_t(\phi(u))\bar{f}(u)du\}.$$

The crucial point in obtaining the exact slope of $\{N^{-1}\nu_N\}$ is that the large deviation of ν_N is just the large deviation of S_N. This is dealt with in Proposition 3-4-3. In order to distinguish between the large deviations of the *different* statistics, the large deviation of ν_N, for instance, is written $I(\nu,x)$.

Proposition 3-4-3 is due to Killeen and Hettmansperger (1972, p.1509). The actual statement of the proposition given below is more general than that given by Killeen and Hettmansperger and seems to reflect more clearly why the result is true. The proof - the same as that in Killeen and Hettmansperger (1972) - is repeated here because it is so short and yet has what appear to be significant consequences.

Proposition 3-4-3. Suppose that for each N, $X_{1,N},\dots,$ $X_{s(N),N}$ are identically distributed random variables, where $\lim_{N\to\infty} (s(N))^{1/N} = 1$. Then if $M_N = \max_{1\le i\le s(N)} \{X_{i,N}\}$ and

$$\lim_{N\to\infty} N^{-1} \log P(X_{1,N} \ge k_N) = c,$$

it follows that

$$\lim_{N\to\infty} N^{-1} \log P(M_N \ge k_N) = c.$$

Proof. Since it follows by Boole's inequality that

$$P(X_{1,N} \ge k_N) \le P(M_N \ge k_N) \le \sum_{i=1}^{s(N)} P(X_{i,N} \ge k_N),$$

then

$$(3.4.6)\quad N^{-1} \log P(X_{1,N} \ge k_N) \le N^{-1} \log P(M_N \ge k_N) \le N^{-1} \log s(N) + N^{-1} \log P(X_{1,N} \ge k_N)$$

because the $X_{i,N}$ are identically distributed. Taking limits in (3.4.6), the left hand side tends to c and so does the right hand side because $N^{-1} \log s(N) \to 0$. This proves the required result.

Corollary 3-4-4. For the two-sample rank statistics, $I(\nu,x) = I(S,x)$.

Proof. Proposition 3-4-3 can be applied, since under H the statistics $S_N(g(R))$ for the different $g \in G$ are identically distributed (see theorem 2A of Hájek (1969)).

The consequence of the previous results is now fairly obvious. It is stated, though, as a theorem.

Theorem 3-4-5. For ν_N derived from S_N, $e(\nu_N, S_N) \geq 1$, irrespective of the *alternative hypothesis*.

Proof. Corollary 3-4-2, Corollary 3-4-4 and the fact that $I(x)$ is a nondecreasing function.

Since explicit expressions for $I(S,x)$ are available, as for instance in Woodworth (1970), the exact slope of $\{\nu_N\}$ can now be written down. In general the expressions for $I(S,x)$ are rather complicated and the results only become *useful after numerical* tabulation.

In view of Theorem 3-4-5, optimal properties for ν_N may be derived from those of S_N. For instance Hájek has shown that it is possible to choose $\phi(u)$ (related to the scores $a_N(i)$ by (A2) of section 3.3) so that the best possible exact slope for the given alternative hypothesis is actually attained by the corresponding two-sample rank test. Clearly ν_N derived from that particular S_N shares this property. The invariance properties of ν_N also suggest that this optimality *holds for a wider range* of alternatives obtained by varying the cut-off point and direction of measurement in the definition of F and G regarded now as circular distributions.

Example. The efficiency properties attributed to ξ_N in section 2.3 (ii) can now be discussed in detail. In fact in order to use later a result obtained by Woodworth (1970), the statistic we actually work with is

$$\xi_N^o = \max_{g \in G} \{W_N^o(g(R))\},$$

where $W_N^o = (N+1)^{-1}\tilde{W}_N - {}^m/_2$. Clearly the value of the exact slope of $\{\xi_N^o\}$ is the same as that of $\{\xi_N\}$ and $\{\tilde{\xi}_N\}$.

The almost sure limit of $N^{-1}\tilde{\xi}_N$ can be obtained using Corollary 3-4-2. A direct approach using the Glivenko-Cantelli lemma and the limit of $N^{-1}W_N^o$ appears to be easier. As before, ${}^m/_N \to \lambda$, $0 < \lambda < 1$. By the Glivenko-Cantelli lemma,

$$(N(N+1))^{-1}\, mn\, D^+_{m,n} \to \lambda(1-\lambda) \sup_x (F(x) - G(x))$$

and

$$(N(N+1))^{-1}\, mn\, D^-_{m,n} \to \lambda(1-\lambda) \sup_x (G(x) - F(x)),$$

both with probability one. Furthermore

$$N^{-1}\, W_N^o \to w$$

with probability one, where $w = \lambda(1-\lambda)\alpha + \lambda^2/2 - \lambda/2$ with $\alpha = \int G(x)\,dF(x)$. Using (2 1.13) with the appropriate modification, it then follows that

$$(3.4.7) \qquad N^{-1}\xi_N^o \to \max\{w + \lambda(1-\lambda) \sup_x (F(x) - G(x)),\ -w + \lambda(1-\lambda) \sup_x (G(x) - F(x))\}$$

with probability one.

In the particular case when $F = pG + qG^2$, $\sup_x(F(x) - G(x)) = 0$, $\sup_x(G(x) - F(x)) = {}^q/_4$, $\alpha = \frac{1}{2} + {}^q/_6$ and hence

$$N^{-1}\xi_N^o \to (g\lambda(1-\lambda))/6 \tag{3.4.8}$$

with probability one.

Since $I(\xi^o,x) = I(W^o,x)$ by Corollary 3-4-4, an expression for $I(\xi^o,x)$ can be obtained from Woodworth (1970, p.260). On p.269 of Woodworth's paper, $I(W^o,x)$ is tabulated for $\lambda = {}^1/_2, {}^1/_4, {}^1/_8, {}^1/_{16}$ and selected values of x. This table can be used to find the exact slope of $\{\xi_N^o\}$ for the alternative hypothesis that $F = pG + qG^2$ $(p < 1)$. From (3.4.8) it follows that a value x for which $I(W^o,x)$ is tabulated corresponds to $q = 6(\lambda(1-\lambda))^{-1}x$.

A useful series expansion for $I(W^o,x)$ can be obtained from Woodworth (1970, p.262) and is used to derive (2.3.3). This expansion is

$$I(W^o,x) = 6(\lambda(1-\lambda))^{-1}x^2 + o(x^2) \tag{3.4.9}$$

The efficiency of ξ_N^o relative to $V_{m,n}$ (Kuiper's test) is required. Abrahamson (1967) has computed the exact slope of the sequence $\{V_{m,n}\}$ to be

$$4\lambda(1-\lambda)\{\sup_x(F(x) - G(x)) + \sup_x(G(x) - F(x))\}^2 + o(\{\sup_x(|F(x) - G(x)|)\}^2), \tag{3.4.10}$$

which, when $F = pG + qG^2$ $(p < 1)$, becomes

$$(q^2\lambda(1-\lambda))/4 + o(q^2). \tag{3.4.11}$$

A brief digression to discuss the derivation of (3.4.10) seems appropriate at this point. The large deviations of $V_{m,n}$ and V_N can be obtained directly from the large deviations of $D_{m,n}$ and D_N respectively, since it was observed in section 3.1 that the two statistics can be written in the form required for Proposition 3-4-3. This is more or less implicit in the proof of theorem 2 of Abrahamson's paper. The argument becomes much clearer though when the essential part is abstracted along the lines of Proposition 3-4-3.

From (3.4.9) and (3.4.11), it follows that

$$e(\xi_N^o, V_{m,n}) = ({}^4/_3 + o(1))/(1 + o(1)),$$

and hence

$$\lim_{q \to 0} e(\xi_N^o, V_{m,n}) = {}^4/_3.$$

The limiting efficiency that has just been calculated was for the alternative hypothesis that $F = pG + qG^2$ $(p < 1)$, but applies to any member of the class of alternatives defined by (2.3.1) because of the invariance of $\xi_N^o, V_{m,n}$. The efficiency we have just derived is also in fact that of W_N^o against $D_{m,n}$ for the alternative $F = pG + qG^2$.

This last result can be easily generalized to the case where the alternative hypothesis is that $F = pG + qG^k$, $k > 1$, $0 \leq p < 1$, $p + q = 1$. In this case

(3.4.12) $$\lim_{q\to 0} e(\xi_N^o, V_{m,n}) = 12\left(\frac{k-1}{k+1}\right)^2,$$

which increases quite rapidly with k; for instance when k = 4, it is 4.32. Again this result applies to the class of circular analogs of the hypothesis defined by (3.1.4).

It might be pointed out that the alternatives used here are very convenient for obtaining simple results for the limiting asymptotic efficiency. Other possible alternatives which might be considered appear less tractable.

It would be interesting to have some idea of the kind of alternative hypotheses for which the exact slope of ν_N is actually greater than that of S_N (so that $e(\nu_N, S_N) > 1$). For this purpose it suffices to consider the almost sure limits of $N^{-1}S_N$ and $N^{-1}\nu_N$ since $I(S,x)$ is strictly increasing. The following proposition is suggested by (3.4.7) which shows that ξ_N^o is more efficient than W_N^o if $F(x) > G(x)$ for some x.

Proposition 3-4-6. If ϕ is increasing, differentiable and the derivative is bounded away from 0, then for the corresponding S_N, ν_N whenever the alternative hypothesis is such that $F(x) > G(x)$ for some x it follows that $e(\nu_N, S_N) > 1$.

Proof. It is sufficient to show that

$$\int_0^1 \phi(u)d(F_a(H_a^{-1}(u)) - \int_0^1 \phi(u)d(F(H^{-1}(u)) > 0$$

for some $a \in (0,2\pi)$ where F_a, H_a are defined in terms of F,H by (1.2.3). Integrating by parts this becomes

$$\int_0^1 \phi'(u)\{F(H^{-1}(u)) - F_a(H_a^{-1}(u))\}du > 0$$

for some $a \in (0,2\pi)$. Now

$$\int_0^1 \phi'(u)\{F(H^{-1}(u))-F_a(H_a^{-1}(u))\}du \geq \varepsilon\left(\int_0^{2\pi} F(u)\,dH(u)-\int_0^{2\pi} F_a(u)\,dH_a(u)\right)$$

(where $\phi'(u) \geq \varepsilon > 0$, $u \in (0,1)$)

$$= \varepsilon(1-\lambda)\left(\int F dG - \int F_a dG_a\right)$$

$$= \varepsilon(1-\lambda)(F(a) - G(a))$$

$$> 0$$

for some $a \in (0,2\pi)$ by hypothesis.

The import of Proposition 3-4-6 is really quite satisfactory, since the only other alternative hypotheses possible are those when F,G satisfy $F(x) \leq G(x)$ for all x and for these alternatives S_N based on ϕ monotone increasing may be expected to perform well (if it is going to perform well at all). Otherwise ν_N is a definite improvement on S_N.

One other question might be raised if considering the derivation of the exact slopes of the test statistics η_N defined (for the two-sample case) by (3.3.26). Is there some method, along the lines of Proposition 3-4-3, by which the exact slope $\{\eta_N\}$ can be obtained from the exact slope of $\{S_N\}$? It appears not, although the exact slopes of these statistics can be obtained in a straightforward fashion by adapti..g the methods of Woodworth (1970). This work however doe· not really belong to the mainstream of the thesis and so *is deferred to Appendix 3*, where there is a brief outline of the results.

3.5 Conclusion

This section will be used to gather together a few observations about the tests we have introduced in Part I. These will be matters which have mostly not been touched on in the course of the discussion, but some of them might serve as useful guides to future research.

The two-sample tests A and A_1 defined as having the critical regions $\{r \in R : \nu_N(r) > t_\alpha\}$ and $\{r \in R : \nu_{N1}(r) > t'_\alpha\}$ respectively are essentially tests suited for use against a vaguely defined alternative hypothesis. In this sense, they might be compared with the Smirnov tests. It is as well therefore that by appropriately choosing the statistic S_N from which ν_N and ν_{N1} are derived, the basic requirement of consistency can be ensured for the tests A and A_1 aga ... al alternatives.

Proposition 3-5-1 Suppose that F,G are distributions determining densities f,g and that ϕ defined by (A2) of section 3.3 is a strictly mc ..tone, differentiable function on (0,1). Then the tests A anc $_1$ corresponding to ϕ (through ν_N and ν_{N1}) are consistent for the alternative $K : F(x) \neq G(x)$ for some x.

Proof. Let us consider A_1 and assume that ϕ is strictly increasing. In the light of section 3-1, it will be sufficient to show that the test based on S_N generated by ϕ is consistent for the alternative hypothesis that $F(x) \le G(x)$ with strict inequality at some point.

To prove this, we use the result of Hájek (1974) (consult Proposition 3-4-1 for the notation) which says that

$$N^{-1}S_N \to \lambda\int_0^1 \phi(u)\bar{f}(u)du$$

with probability one. If this limit can be shown to be strictly larger under the alternative hypothesis than it is under the null hypothesis, the result will be established. Using integration by parts, this would require that $\int_0^1 \phi'(u)(u - F(H^{-1}(u)))du > 0$.

Since $F(x) \le G(x)$ for all x, it follows that $u-F(H^{-1}(u)) \ge 0$ for $u \in [0,1]$. Furthermore F,G continuous and $F(x) < G(x)$ for some x imply that it is possible to find $0 < a < b < 1$ such that $H(x) - F(x) > 0$ for $x \in [a,b]$ and $H(a) < H(b)$. Then

$$\int_0^1 \phi'(u)(u - F(H^{-1}(u)))du \ge \int_{H(a)}^{H(b)} \phi'(u)(u - F(H^{-1}(u)))du$$

$$> r\int_a^b \phi'(H(u))dH(u)$$

(where $r = \inf\{H(x) - F(x) : x \in [a,b]\}$

$$= r\{\phi(H(b)) - \phi(H(a))\}$$

$$> 0$$

which gives the required result. The other cases in the proposition may be similarly handled.

The actual computation of the test statistic ν_N could be rather a lengthy process. There seems to be no general result which might help to make computation of ν_N less cumbersome. Each situation will call for its own simplifications. Otherwise it is necessary to settle for a few pretty obvious remarks. For instance, if S_N may be classified as either even - or odd-translation invariant, then instead of computing 2N values of S_N one need only compute the N values of $S_N(g(R))$ for $g \in G_1$. In the two-sample situation, if $S_N = \sum_{i=1}^{m} a_N(R_i)$ has scores which satisfy $a_N(1) \le \ldots \le a_N(N)$, then when calculating ν_{N1} only values of $S_N(g(R))$ corresponding to $g = (g_1)^{R_i}$, $i = 1,\ldots,m$ need be calculated since the maximum can only occur at one of these m values.

If we put $s(k) = S_N((g_r)^k(R))$ and $s(k') \ge s(k)$, then it follows that $s(k'+1) - 2s(k') + s(k'-1) < 0$. This may help in finding ν_{N1}, although it only seems to be at all useful when $a_N(R_i) = R_i$ or $R_i(R_i-1)$. Some breakthrough along these lines may be possible though. As we have indicated in section 2.2 the Mann-Whitney type statistics are no more difficult to compute than the test statistics for the Smirnov tests.

As far as Part I of this thesis goes, it just remains to point out some directions for future research. The main issue to be resolved must be obtaining suitable large sample size approximations for the distributions of the test statistics ν_N and ν_{N1}. This point has already been raised in section 3.3. Another consequence of such informati m might well be some insight - probably in the form of upper and lower bounds - concerning the asymptotic relative efficiencies of the various tests. This would be an important supplement to section 3.4.

Of some theoretical interest is the question of generalizations of the tests introduced during the course of chapter 3. Such a generalization might for instance be obtained by replacing G in the definition of ν_N by other sequences of subgroups of the s. metric group (the sequence of subgroups corresponding to different values of N). Another possibility lies in investigating other quotient spaces in the way we have done for the circle. In each case the identification of points is going to require new invariance properties.

Another important question is the robustness of performance of, for instance, the tests based on ν_{N1}. We believe that these tests may provide a class of systematically derived robust tests for general alternatives. Here we mean by robust that the performance of the test is not greatly affected by small deviations in the underlying distribution of the sample.

It should also be noted that the method of test construction suggested in section 3.1 can also be easily adapted to derive one-sample tests for the uniformity of a circular distribution from one-sample nonparametric tests. It seems that such a test derived from the Wilcoxon test does not share the pleasing possibilities of the test we discussed in chapter 2. An obvious question would then be how much of our analysis does carry over to this one-sample situation.

PART II

CHAPTER 4. On Determining the Effectiveness of Linear Rank Test Statistics.

4.1 An outline of the contents of Part II.

The aim of this part of the thesis is to gain further understanding about two-sample linear rank statistics, defined by (1.2.2) as

$$S_N = \sum_{i=1}^{N} a_N(i)\ z_{Ni}. \tag{4.1.1}$$

In this respect there is a clear division between the two parts of the thesis, with Part I concerning test statistics which were nonlinear functions of the rank vector and Part II concerning only linear rank statistics.

A deeper insight into the performance of these rank tests can be useful in several ways. Although the null distribution of a rank statistic under the hypothesis of randomness does not depend on the particular distribution of the members of the random sample, nevertheless the performance of the test is going to depend on the proposed alternative hypothesis. In many instances, there exist rank tests which are either locally most powerful or asymptotically most powerful for a given situation (LMPRT's and AMPRT's). It is usually not clear though how deviations from the proposed alternative hypothesis will effect the performance of these optimal tests. Some guide to the strengths and weaknesses of these optimal tests would be very useful.

Another p ' - with optimal tests is that they may not be readily appl ince their test statistics require too great an effor .or their computation. Some simple approximation to the test is required which yields a test statistic which is straightforward to compute and yet the corresponding test does not sacrifice too much of the optimality of the original test. An incidental benefit may also be that the test with the simpler form is more robust against deviations from the assumed model.

A very useful way of comparing the performance of two sequences of tests is provided by their asymptotic relative efficiency (ARE). In chapter 5 we obtain a more detailed picture of the relative performance of two tests by - in some sense - decomposing the ARE of the two tests into a number of other quantities each of which allows an interpretation in terms of the relative performance of the tests. This detailed analysis often confirms in precise terms what are otherwise fairly vague intuitive feelings about the various tests. It can also sometimes reveal details which were not so immediately apparent. It is essentially a diagnostic device for assessing why a test performs in the way it does.

Another benefit of the analysis conducted in chapter 5 is that the decomposition can be used to obtain a sequence of rank tests, the test statistics of which are relatively easv to compute and yet which can be made to have an efficiency relative to the AMPRT of as near to one as required.

The relevant terms and notation introduced in Part I of the thesis retain the same meanings for Part II.

4.2 Asymptotic relative efficiency

Let $\{A_N^{(1)}\}$ and $\{A_N^{(2)}\}$ be two sequences of tests for the null hypothesis H which can be expressed as $H : \theta = \theta_0$ against the alternative hypothesis that $\theta > \theta_0$ (in general θ may be taken to index the joint distribution of $X_1,\ldots,X_N$). Then the relative efficiency of $\{A_N^{(2)}\}$ relative to $\{A_N^{(1)}\}$ is the ratio $N^{(1)}/N^{(2)}$ where $N^{(1)}$, $N^{(2)}$ are the number of observations necessary to give $A^{(1)}_{N^{(1)}}$ and $A^{(2)}_{N^{(2)}}$ the same power for a given level of significance. Pitman suggested asymptotic relative efficiency as the limit of $N^{(1)}/N^{(2)}$ for a sequence of alternatives depending on the sample size and converging to H in such a way that the power of both tests converges to a limit less than one. A detailed discussion of these well-established ideas may be found in Noether (1955), who proves a result of which Theorem 4-2-1 is a special case.

Let A_N be a test for the hypothesis $H : \theta = \theta_0$ against the alternative hypothesis that $\theta > \theta_0$ base[illegible] observations, let T_N be the test statistic and let $\mu_N(\theta)$ [illegible]) and $\sigma_N^2(\theta) = \mathrm{Var}(T_N|\theta)$. Suppose that θ_N is a seq[illegible] of alternatives such that $\theta_N = \theta_0 + kN^{-\frac{1}{2}}$, where k is a positive constant independent of N.

Theorem 4-2-1. Let the following conditions be satisfied:

A. an $\varepsilon > 0$ such that for $\theta_0 \le \theta \le \theta_0 + \varepsilon$, $\mu_N'(\theta)$ exists,

B. $\lim_{N\to\infty} \mu_N'(\theta_N)/\mu_N'(\theta_0) = 1$,

C. $\lim_{N\to\infty} \sigma_N(\theta_N)/\sigma_N(\theta_0) = 1$,

D. $c = \lim_{N\to\infty} \mu_N'(\theta_N)/N^{\frac{1}{2}}\sigma_N(\theta_0)$ exists,

E. the distribution of $(T_N - \mu_N(\theta_N))/\sigma_N(\theta_N)$ tends to the Normal distribution as $N \to \infty$.

(Note: ' denotes a derivative).

Then,

(i) if A_N is a test satisfying A-E, the asymptotic power of A_N is given by

$$1 - \Phi(\lambda_\alpha - kc),$$

where λ_α is determined by the significance level α of the sequence of tests $\{A_N\}$ as satisfying $1 - \Phi(\lambda_\alpha) = \alpha$ and Φ is defined by $\Phi(x) = \int_{-\infty}^{x} (2\pi)^{-\frac{1}{2}} e^{-y^2/2} dy$;

(ii) if $c^{(1)}, c^{(2)} > 0$, then the ARE of $\{A_N^{(2)}\}$ relative to $\{A_N^{(1)}\}$ is given by

$$\text{ARE}(T^{(2)}, T^{(1)}) = (c^{(2)}/c^{(1)})^2$$

(identifying tests and test statistics), provided the sequences of tests $\{A_N^{(1)}\}$, $\{A_N^{(2)}\}$ satisfy A-E.

Our discussion will be exclusively in terms of the two-sample situation so that for asymptotic results it is assumed that $m/N \to \lambda$, $0 < \lambda < 1$.

4.3 A survey of related research

The work in chapter 5 rests principally on an expansion of the weighting functions of the test statistics in terms of a set of orthonormal functions. The use of orthonormal expansions of one sort or another is quite commonplace in statistics. Nevertheless, the possibility of using such expansions to study ARE does not seem to have been mentioned before.

Perhaps that area of research which lies closest to the investigations of chapter 5 is the recent interest shown in robust procedures. It is the concern for robust tests which suggests detailed investigation of the efficiency properties of tests. The main research into robust nonparametric procedures has been in connection with estimation. Very significant work has been done by Hampel (1968, 1971, 1974) in discussing measures of robustness and their use in obtaining 'good' estimators. As regards nonparametric tests, discussion of robustness properties appears in the work of Doksum (1966, 1969). Another significant contribution has been made by Gastwirth (1965, 1967, 1970) who suggested looking for quick tests - approximations to the AMPRT's.

The relationship between the performance of a rank test and the weight in the tails of the distributions of the sample is discussed in two of the papers already mentioned, those of Doksum (1969) and Gastwirth (1970). Another useful reference here is Hájek (1969, pp.150-151). We shall discuss this relationship in section 5.4.

CHAPTER 5 Orthonormal Expansions for Linear Rank Tests.

5.1 A Riesz representation theorem for ARE.

The essential features of the following result are well-known, although the actual statement lies somewhere between the results of Hájek (1962, and also in Hájek and Šidák (1967), pp.267-269) and a result obtained by van Eeden (1963). The idea is to use the familiar Riesz representation theorem for bounded linear functionals on the Hilbert space $L_2([0,1])$ to obtain a result about ARE.

Theorem 5-1-1 concerns the test statistics T_N^ϕ where $m\,T_N^\phi = \sum_{i=1}^{N} a_N(i)\; Z_{Ni}$ and $a_N(1 + [uN]) \to \phi(u)$ in L_2, $\int_0^1 \phi(u)du=0$.

Theorem 5-1-1. Assume that the test statistics T_N^ϕ are being used to test $H : \theta = 0$ against the alternative that $\theta = \theta_N$ where $\theta_N > 0$ and $\theta_N \to 0$ as $N \to \infty$. Suppose that the assumptions of Theorem 4-2-1 hold for the T_N^ϕ, $\phi \in L_2$ and $\int_0^1 \phi(u)du=0$. In addition, suppose that $\mu_N'(\theta_N) \to g(\phi)$, where $\{g(\phi) : \|\phi\| = 1\}$ is bounded and $g(\phi) > 0$ for at least one ϕ. Then there exists a unique $\tilde{\phi} \in L_2$ such that

$$(5.1.1) \qquad \mathrm{ARE}(T^\phi, T^{\tilde{\phi}}) = \left\{\frac{\langle\phi,\tilde{\phi}\rangle}{\|\phi\|\,\|\tilde{\phi}\|}\right\}^2$$

(the inner product and norm are for L_2 and were defined in section 3.2), where $\int_0^1 \tilde{\phi}(u)du = 0$.

The tests with test statistics generated by $\tilde{\phi}(u)$ are AMPRT's within the class of linear rank tests and the ARE between a sequence of such tests and any other sequence with test statistics T_N^ϕ is equal to the limiting correlation coefficient between the test statistics (under H).

Proof. Consider the efficacy of the sequence of tests based on T_N defined by $e(T) = \lim_{N\to\infty} \frac{N^{-\frac{1}{2}}\mu'_N(\theta_N)}{\sigma_N(0)}$. To find $\lim_{N\to\infty} N\sigma_N(0)^2$, recall that from p.61 of Hájek and Šidák (1967) it follows that

$$N\ \mathrm{Var}(T_N^{\phi}) = nm^{-1}\{(N-1)^{-1} \sum_{i=1}^{N} (a_N(i) - \bar{a}_N)^2\}$$

under H, so that taking limits, $N\sigma_N(0)^2 \to \frac{1-\lambda}{\lambda}\|\phi\|^2$. Hence, $e(T^{\phi}) = ((\frac{\lambda}{1-\lambda})^{\frac{1}{2}}\, g(\phi))/\|\phi\|$.

Now g is in fact a linear functional. Let us consider for instance, $g(\phi_1 + \phi_2)$. The value of g is independent of the particular sequence of test statistics T_N^{ϕ} provided the scores satisfy $a_N(1 + [uN]) \to \phi(u)$ in L_2. In particular therefore, given $\phi(u) \in L_2$, we can always consider T_N^{ϕ} defined by the scores $a_N(i) = N\int_{(i-1)/N}^{i/N} \phi(u)du$. Clearly with this particular choice of scores for $T_N^{\phi_1}$, $T_N^{\phi_2}$ and $T_N^{\phi_1+\phi_2}$, it follows that $T_N^{\phi_1+\phi_2} = T_N^{\phi_1} + T_N^{\phi_2}$, so that $E(T_N^{\phi_1+\phi_2}) = E(T_N^{\phi_1}) + E(T_N^{\phi_2})$. Then differentiating and taking limits we get $g(\phi_1 + \phi_2) = g(\phi_1) + g(\phi_2)$.

The Riesz representation theorem may now be applied to $h(\phi) = (\frac{\lambda}{1-\lambda})^{\frac{1}{2}}\, g(\phi)$ to obtain $h(\phi) = \langle\phi, \tilde{\phi}\rangle$ for some unique $\tilde{\phi} \in L_2$ such that $\int_0^1 \tilde{\phi}(u)du = 0$. Limiting ourselves to the cases where $e(T^{\phi}) > 0$ and applying part (ii) of Theorem 4-2-1, we obtain (5.1.1). That the right hand side of (5.1.1) is in fact the limiting correlation between T_N^{ϕ} and $T_N^{\tilde{\phi}}$ follows by taking the limit in the formula for the correlation of two linear rank statistics obtainable from p.62 of

Hájek and Šidák (1967).

Remarks. (i) The ARE is strictly speaking not defined when $e(T) \leq 0$. In our subsequent work this will be ignored. If $e(T)$ is zero, it still makes sense to speak of zero ARE relative to the AMPRT, while if $e(T) < 0$, it implies that the test is not suited for the alternative hypothesis. Instead of the scor[illegible] $a_N(i)$, the scores $-a_N(i)$ should be used.

(ii) The [illegible] important implication of Theorem 5-1-1 is the existence of asymptotically optimal rank tests within the restricted class of tests being discussed. To actually find the weighting function $\tilde{\phi}(u)$ of such tests the result on limiting normality obtained by Chernoff and Savage (1958) is often useful. From the parameters of that result it follows that we can take $\mu_N(\theta_N) = \int_{-\infty}^{\infty} \phi(H_{\theta_N}(x))dF(x)$, where $H_{\theta_N}(x) = {}^m/_N\, F(x) + {}^n/_N\, G_{\theta_N}(x)$.

Suppose that ϕ possesses a derivative and satisfies $\lim\limits_{u\to 0,1} \{\phi(u)\, \frac{\partial}{\partial\theta}(G_\theta(x))\Big|_{\theta=0,x=F^{-1}(u)}\} = 0$. Then provided the operations performed can be justified, integration by parts yields

$$\|\phi\|\, e(T^\phi) = -(\lambda(1-\lambda))^{\frac{1}{2}} \int_0^1 \phi(u)\left\{\frac{\frac{\partial^2}{\partial x\partial\theta}(G_\theta(x))\Big|_{\theta=0,x=F^{-1}(u)}}{f(F^{-1}(u))}\right\}du,$$

where f is the density corresponding to F. If this relationship can be verified for functions ϕ constituting a basis or a complete orthonormal set, then $\tilde{\phi}(u) = -\dfrac{\frac{\partial^2}{\partial x\partial\theta}(G_\theta(x))\Big|_{\theta=0,x=F^{-1}(u)}}{f(F^{-1}(u))}$ (compare with Capon (1961)).

The rest of this chapter will be concerned with obtaining a more detailed 'picture' of the relative performances of two-sample linear rank tests. This will be achieved by using the coefficients in the expansion of the weighting function $\phi(u)$ of the test statistic in terms of a specially chosen system of orth ormal functions. This system of functions is chosen so that the performances of the tests can be compared over subsets of the ranks.

One reason for obtaining such information is to further develop the readily acceptable intuitive argument that the lighter the tails of a density, the more information is contained in the extreme observations of a sample from that density. Thus it is expected that the AMPRT for the alternative hypothesis of location shift will place more weight on the extreme ranks when the underlying distribution has light tails. It is the intention of this chapter to systematically quantify the weight a test accords to various subsets of the ranks, using the notion of ARE as a measure of weight.

As the discussion will be exclusively in terms of asymptotic notions, it seems to be often convenient to identify the weighting function $\phi(u)$ with the rank tests it generates and speak of the test $\phi(u)$. This is permissible for the situations we shall be dealing with since the asymptotic properties of all the tests based on test statistics generated by $\phi(u)$ depend only on $\phi(u)$ (rather than on the actual scores $a_N(i)$).

The tests $\phi(u)$ are assumed to be normalized so that $\int_0^1 \phi(u)du = 0$ and $\int_0^1 (\phi(u))^2du = 1$ (this also serves to exclude degenerate tests). Theorem 5-1-1 then shows that the ARE of two tests based on ϕ_1, ϕ_2 (normalized) is given by ρ^2, where $\rho = \int_0^1 \phi_1(u)\phi_2(u)du$, provided that one of the two tests is the AMPRT for the sequence of alternatives. It will be easier to work with ρ, which we *shall call simply the efficiency*, always bearing in mind the interpretation in terms of ARE.

The idea is to use $\phi(u)$ to obtain a detailed 'profile' of how the test performs against the AMPRT as measured by efficiency. For this purpose an appropriate set of orthonormal functions is required in order to obtain an expansion for $\phi(u)$ open to useful interpretation. Such functions are introduced in section 5.2.

5.2 Mathematical preliminaries: the Haar expansion for functions in L_2.

Since $\phi(u) \in L_2$, if $\{\psi_k(u) : k = 1,2,\ldots\}$ is a complete orthonormal set in L_2, then $\phi(u)$ has an expansion

$$\phi(u) = \sum_{k=1}^{\infty} a_k \psi_k(u), \tag{5.2.1}$$

where $a_k = \langle\phi,\psi_k\rangle$ (the L_2 inner product) and convergence is with respect to the L_2 norm. Two immediate consequences of (5.2.1) are that

$$1 = \|\phi(u)\|^2 = \sum_{k=1}^{\infty} a_k^2 , \tag{5.2.2}$$

while if $\pi(u) \in L_2$ so that $\pi(u) = \sum_{k=1}^{\infty} b_k \psi_k(u)$, then

$$(5.2.3) \qquad \langle\phi(u),\pi(u)\rangle = \sum_{k=1}^{\infty} a_k\, b_k.$$

The orthonormal system used for the rest of this chapter is the Haar system, consisting of step functions defined on [0,1] as follows:

$$\chi_0^{(0)}(u) = 1 \qquad u \in [0,1];$$

$$\chi_0^{(1)}(u) = \begin{cases} 1 & u \in [0,\tfrac{1}{2}) \\ -1 & u \in [\tfrac{1}{2},1), \end{cases}$$

and for $m \geq 1$ and $1 \leq k \leq 2^m$;

$$\chi_m^{(k)}(u) = \begin{cases} 2^{m/2} & u \in (\frac{k-1}{2^m}, \frac{k-\frac{1}{2}}{2^m}) \\ -2^{m/2} & u \in (\frac{k-\frac{1}{2}}{2^m}, \frac{k}{2^m}) \\ 0 & u \in (\frac{\ell-1}{2^m}, \frac{\ell}{2^m}), \end{cases}$$

with $\ell \neq k$, $1 \leq \ell \leq 2^m$. The definition of these functions at points of discontinuity has not been included since these details are not required for our purposes. They can be found in Alexits (1961, pp.46-47).

5.3 Interpretation of Haar expansions

In future it will be assumed that $\phi(u)$ is the AMPRT so that inner products involving $\phi(u)$ can in fact be readily interupted in terms of ARE. For any orthonormal system, (5.2.3) offers a decomposition of the efficiency of $\pi(u)$ against $\phi(u)$. The coefficients a_k in the expansion of $\phi(u)$ may be regarded as potential efficiency since if $<\phi(u),\pi(u)>$ is to be near 1, then the coefficients a_k and b_k must mostly be of similar magnitude. The coefficient a_k is also itself an efficiency, of $\psi_k(u)$ relative to $\phi(u)$ and may be interpreted as the weight in terms of efficiency accorded to $\psi_k(u)$ by $\phi(u)$.

For a meaningful interpretation of these general remarks, Haar's system of functions appears more suitable than other, better known, orthonormal systems in L_2. The coefficient of $\chi_m^{(k)}(u)$ in the expansion of $\phi(u)$ is denoted by $c_{m,k}$. By normalization, $c_{0,0} = 0$ and so the first useful coefficient of the expansion is $c_{0,1}$. But $\chi_0^{(1)}$ *is just the weighting function* of the median test and hence $c_{0,1}$ is the efficiency of the median test relative to the test generated by $\phi(u)$. Examining $\chi_m^{(k)}(u)$, it becomes clear that this may also be considered as the weighting function of a kind of median test applied to a subset of the sample falling between two quantiles.

The weighting function indicates the limiting values of the scores assigned to the ranks and $\chi_m^{(k)}$ only assigns nonzero weight to the interval $(\frac{k-1}{2^m}, \frac{k}{2^m})$. In other words, the test generated by $\chi_m^{(k)}$ is only affected by X's falling between the $(\frac{k-1}{2^m})$-th and the $(\frac{k}{2^m})$-th quantile of the combined sample. The efficiency of $\chi_m^{(k)}$ relative to ϕ is $c_{m,k}$ and this also indicates the weight accorded by the test ϕ to this particular range of the sample. In particular, the weight attached to extreme observations may now be quantified in terms of the $c_{m,k}$'s for extreme intervals. Comparing the coefficients arising from the expansions of $\phi(u)$ and $\pi(u)$ allows us to evaluate the relative weights the two statistics attach to the ranks corresponding to a particular interval.

The detailed analysis of the test $\phi(u)$ obtained by computing the coefficients $c_{m,k}$ may be useful in two ways:

(a) The analysis is explanatory, since it breaks down into detail the relative performance of two tests as measured by their efficiency. This detail may be regarded as being of two kinds. Firstly there is the performance of the tests at different levels as measured by $c_{m,k}$ for different values of m. As a measure of the proportion of the total weight attributable to the mth level, one may define $w_{1m} = \sum_{j=1}^{2^m} c_{m,j}^2$. Then w_{1m} measures the ability of $\phi(u)$ to detect m-th level deviations from the hypothesis. The values of w_{1m} for increasing m indicate the efficiency attached to increasingly smaller subsets of the ranks and so may be regarded as an indication of the complexity of the test. A high degree of complexity would tend to suggest a certain lack of robustness when the alternative hypothesis is not one for which $\phi(u)$ is the AMPRT.

Secondly, the performance of tests over various subsets of the sample at a given level can be measured.

(b) The analysis can be exploratory, since analysis of the AMPRT may suggest a test whose weighting function has a simpler form - so that the test statistic is easier to compute - but which is still nearly as efficient as the AMPRT. If the rate of convergence in (5.2.2) warrants it, one such 'approximation' can be obtained by taking the terms in the expansion of $\phi(u)$ (the AMPRT) corresponding to the first few levels (values of m). This will be discussed in more detail in section 5.5.

Another possible benefit in such an approximation of $\phi(u)$ lies in obtaining a test which is more robust without sacrificing too much by way of efficiency relative to the AMPRT.

For a full appreciation of these interpretative possibilities, the discussion of some examples is required. This is done in section 5.4. Before proceeding to this, some simple preliminaries are required.

The alternative hypotheses under which the asymptotic performance of the rank statistics will be studied are (a) $G(x) = F(x+\theta)$, $\theta > 0$ (location shift) and (b) $G(x) = F(x\sigma)$, $\sigma > 1$ (change of scale). Assuming the appropriate conditions, the AMPRT against location shift for the distribution F with density f has as weighting function

$$\phi_F(u) = - I_f^{-\frac{1}{2}} \frac{f'(F^{-1}(u))}{f(F^{-1}(u))} ,$$

where $I_f = \inf(f) = \int_{-\infty}^{\infty} (f'(x)/f(x))^2 f(x)dx$. The AMPRT against the change of scale alternative has weighting function

$$\phi_{1F}(u) = - I_{1f}^{-\frac{1}{2}}\{1 + F^{-1}(u) \frac{f'(F^{-1}(u))}{f(F^{-1}(u))}\},$$

where $I_{1f} = \int_{-\infty}^{\infty} \{-1- \frac{xf'(x)}{f(x)}\}^2 f(x)dx.$

A definite advantage of using the Haar system of orthonormal functions is that the coefficients $c_{m,k}$ are easily obtained.

Lemma 5-3-1. (i) *The coefficient* $c_{m,k}$ of $\chi_m^{(k)}(u)$ in the expansion of $\phi_F(u)$ is given by

$$c_{m,k} = 2^{m/2}\ I_f^{-\frac{1}{2}}\{f(F^{-1}(\frac{k-1}{2^m})) + f(F^{-1}(\frac{k}{2^m})) - 2f(F^{-1}(\frac{k-\frac{1}{2}}{2^m}))\},$$

$k = 1,\ldots 2^m$, $m \geq 1$ and $m = 0$, $k = 1$.

(ii) The coefficient $c_{m,k}$ of $\chi_m^{(k)}(u)$ in the expansion of $\phi_{1F}(u)$ *is given by*

$$c_{m,k} = 2^{m/2}\ I_{1f}^{-\frac{1}{2}}\{F^{-1}(\frac{k-1}{2^m})f(F^{-1}(\frac{k-1}{2^m})) + F^{-1}(\frac{k}{2^m})f(F^{-1}(\frac{k}{2^m}))$$
$$- 2F^{-1}(\frac{k-\frac{1}{2}}{2^m})f(F^{-1}(\frac{k-\frac{1}{2}}{2^m}))\},$$

$k = 1,\ldots,2^m$, $m \geq 1$ and $m = 0$, $k = 1$.

Proof. The coefficient of $\chi_m^{(k)}$ is given by

$$c_{m,k} = \int_0^1 \phi_F(u)\ \chi_m^{(k)}(u)\,du$$

$$= 2^{m/2} \times \left\{\int_{\frac{k-1}{2^m}}^{\frac{k-\frac{1}{2}}{2^m}} \phi_F(u)\,du - \int_{\frac{k-\frac{1}{2}}{2^m}}^{\frac{k}{2^m}} \phi_F(u)\,du\right\}.$$

The expression given for $c_{m,k}$ then follows because $\int_a^b -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}du = f(F^{-1}(a))-f(F^{-1}(b))$. The proof of part (ii) follows similarly.

A further simplification is obtained as follows:

(a) if $\phi(u)$ is skew-symmetric about $\frac{1}{2}$ (so that $\phi(u) + \phi(1-u) = 0$, $0 < u < 1$), then $c_{m,2^{m-1}+r} = c_{m,2^{m-1}-r+1}$, $r = 1,\ldots,2^{m-1}$ $(m \geq 1)$;

(b) If $\phi(u)$ is symmetric about $\frac{1}{2}$ (so that $\phi(u) = \phi(1-u)$, $0 < u < 1$), then $c_{m,2^{m-1}+r} = -c_{m,2^{m-1}-r+1}$, $r = 1,\ldots,2^{m-1}$ $(m \geq 1)$ and $c_{0,1} = 0$.

Both (a) and (b) follow directly from the way $c_{m,k}$ is obtained In both cases it is thus sufficient to compute and display only $c_{m,1},\ldots,c_{m,2^{m-1}}$ and this is what will be done in section 5.4.

It might reasonably be supposed that the Haar expansion of $\phi_F(u)$ reflects the nature of the density f. In this respect the following result seems to tell us that the coefficients reflect in some way the exponential character (of degree $-x$) of f over a range determined by k and m.

<u>Theorem 5-3-2.</u> The density for which the test based on the weighting function $\chi_m^{(k)}(u)$, $m \geq 1$, is the AMPRT against the location shift alternative is

$$f(x) = \begin{cases} \frac{1}{2}e^{-\ell_2} & -\ell_1 \le x \le -\ell_2 \\ \frac{1}{2}e^{x} & -\ell_2 \le x \le 0 \\ \frac{1}{2}e^{-x} & 0 \le x \le \ell_2 \\ \frac{1}{2}e^{-\ell_2} & \ell_2 \le x \le \ell_3 \\ 0 & \text{elsewhere,} \end{cases}$$

where $\ell_2 = \log(1 + (2^m-1)^{-1})$ and ℓ_1, ℓ_3 are determined by $\frac{1}{2}(\ell_1-\ell_2)e^{-\ell_2} = \frac{k-1}{2^m}$ and $\frac{1}{2}(\ell_3-\ell_2)e^{-\ell_2} = 1 - \frac{k}{2^m}$.

Proof. Straightforward from the way the AMPRT is obtained. Of course, $\chi_0^{(1)}$ gives the *AMPRT for* $f(x) = \frac{1}{2}e^{-|x|}$, $-\infty < x < \infty$, since it is just the weighting function of the median test. The density in Theorem 5-3-2 is uniform except in the range between the $(\frac{k-1}{2^m})$-th and $(\frac{k}{2^m})$-th quantiles where it assumes a double exponential form.

5.4 Analysis of some AMPRT's

Two linear rank tests often used for testing against location shift are the Mann-Whitney test, which we regard as being generated by $\phi(u) = \sqrt{12}(\frac{1}{2}-u)$ and the Normal scores test generated by $\phi(u) = -\Phi^{-1}(u)$, where $\Phi(u) = \int_{-\infty}^{u} (2\pi)^{-\frac{1}{2}}e^{-x^2/2}dx$. Coefficients for the first four levels in the Haar expansion of the two weighting functions are presented in the first two columns of Table 5-1. Values of w_{1m} and $w_{2m} = \sum_{i=0}^{m} w_{1i}$ are also tabulated.

Values of the coefficients for the Mann-Whitney test can be obtained from the following formulae:

TABLE 5-1 <u>Coefficients in the Haar expansions of five weighting functions.</u>

$\phi(u)$	1 $\sqrt{12}(\frac{1}{2}-u)$ (Mann-Whitney)	2 $-\Phi(u)$ (Normal scores)	3 $-\sqrt{180}\{(u-\frac{1}{2})^2-\frac{1}{12}\}$ (Mood)	4 $\sqrt{48}\{\lvert\frac{1}{2}-u\rvert-\frac{1}{4}\}$ (Ansari-Bradley)	5 $\frac{1}{\sqrt{2}}\{(\Phi^{-1}(u))^2-1\}$ (Capon)
$c_{0,1}$	.866	.80	0	0	0
ψ_{10}	.75	.637	0	0	0
$c_{1,1}$	.3062	.33	.593	.6124	.43
ψ_{11}	.1875	.224	.703	.7516	.37
$c_{2,1}$	.1083	.19	.314	.2166	.37
$c_{2,2}$	.	.08	.105	.	.04
ψ_{12}	.0469	.085	.220	.1876	.27
$c_{3,1}$	.0383	.11	.130	.0766	.28
$c_{3,2}$	.	.04	.093	.	.05
$_{,3}$	.	.03	.056	.	.02
$c_{3,4}$	.	.03	.019	.	.01
ψ_{13}	.0117	.033	.058	.0468	.16
$c_{4,1}$	.0135	.07	.049	.027	.21
$c_{4,2}$	.	.02	.043	.	.04
$c_{4,3}$	.	.02	.036	.	.02
$c_{4,4}$	.	.01	.030	.	.01
$c_{4,5}$	.	.01	.023	.	.006
$c_{4,6}$	.	.01	.016	.	.004
$c_{4,7}$	.	.01	.010	.	.003
$c_{4,8}$	.	.01	.003	.	.002
ψ_{14}	.0029	.013	.015	.0116	.10
ψ_{24}	.999	.991	.995	.998	.895

(i) Coefficients with exactly the same value as the preceding one are indicated with a dot.

(ii) Values for the Normal scores tests are only meant to be rough approximations.

$$c_{m,k} = 3^{\frac{1}{2}}.2^{-3m/2-1} \;,\quad k = 1,\ldots,2^m;$$

$$w_{1m} = 3.2^{-2m-2} \;,\quad w_{2m} = 1 - 2^{-2m-2} \;.$$

The numbers in Table 5-1 clearly indicate the greater significance attached to extreme ranks by the Normal scores test. These differences are more pronounced for the higher levels. It is also noticeable that this difference in weights is principally located in the extreme intervals corresponding to $c_{m,1}$ and $c_{m,2^m}$. The values of $c_{m,k}$ for the Mann-Whitney test are constant for a given level, which seems a fair indication of the general robustness of the test. The greater emphasis the Normal scores test places on the extreme ranks is also reflected in the large values of w_{1m} for $m \geq 1$. While this suits the test for use with light-tailed distributions, it also contributes to a decline in efficiency which may be considerable when the distribution is heavy-tailed. Of course, there is very little new in all this, but these remarks can now be explicitly quantified by the numbers of Table 5-1.

Consider the test statistic generated by

$$\phi(u) = \begin{cases} \sqrt{96}\,(\frac{1}{4} - u) & 0 < u < \frac{1}{4} \\ 0 & \frac{1}{4} < u < \frac{3}{4} \\ \sqrt{96}\,(\frac{3}{4} - u) & \frac{3}{4} < u < 1. \end{cases}$$

This statistic is used by Randles and Hogg (1973) in their adaptive rank test to test the data for location shift after the underlying distribution has been classified as light-tailed. Clearly this statistic places more weight on the extreme ranks than does the Mann-Whitney test. If $m \geq 2$, the weights $c_{m,k}$ for this $\phi(u)$ are in fact simply the Mann-Whitney weights multiplied by a constant provided $k = 1,\ldots,2^{m-2}$, $3.2^{m-2},\ldots,2^m$ and otherwise $c_{m,k} = 0$. Hence although all the weight is concentrated on $(0,\frac{1}{4})$ and $(\frac{3}{4},1)$, it is still spread out evenly over subintervals of these intervals, in contrast to the emphasis laid on the extreme subintervals by the Normal scores test. It is, however, this emphasis on the extreme intervals which typifies the AMPRT's for light-tailed distributions. It might therefore be more appropriate to use $\phi(u)$ which involves some power of u, say u^4, in order to weight the extreme ranks more heavily.

The last example serves to indicate that the greater detail obtained from the Haar expansion of $\phi(u)$ can perhaps reveal information not fully apparent from an examination of $\phi(u)$ by itself.

A simple result relating the length of the tails of a distribution with the weight the AMPRT for location shift places upon the extreme ranks is possible. This result relates the coefficients $c_{m,k}$ to an ordering for densities defined in terms of the weighting functions of the corresponding AMPRT's. This ordering is due to Hájek (1969) and is defined as follows: suppose that for densities f,g, the weighting functions

$$\phi(u) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}, \qquad \psi(u) = -\frac{g'(G^{-1}(u))}{g(G^{-1}(u))}$$

are continuous, skew-symmetric, nondecreasing functions on $(0,1)$, then $f < g$ if $\phi(u) = b(u)\psi(u)$, $\frac{1}{2} < u < 1$, where $b(u)$ is nondecreasing and *nonconstant*.

Proposition 5-4-1. *If* $f < g$, *then for m sufficiently large,* $c_{m,2^m}$ in the expansion of $\psi(u)$ is larger than $c_{m,2^m}$ *in the* expansion of $\phi(u)$ (both coefficients are negative).

Proof. First notice that $\phi(u)$ and $\psi(u)$ can be assumed to be normalized so that $\|\phi\| = \|\psi\| = 1$, without affecting the fact that $\phi(u)/\psi(u)$ is nondecreasing for $\frac{1}{2} < u < 1$.

If $c_{m,k}(\phi)$ *denotes the coefficient* $c_{m,k}$ in the Haar expansion of $\phi(u)$, then

$$c_{m,k}(\phi) = 2^{m/2}\left\{\int_{\frac{k-1}{2^m}}^{\frac{k-\frac{1}{2}}{2^m}} \phi(u)du - \int_{\frac{k-\frac{1}{2}}{2^m}}^{\frac{k}{2^m}} \phi(u)du\right\}$$

$$= 2^{m/2}\left\{\int_{\frac{k-1}{2^m}}^{\frac{k-\frac{1}{2}}{2^m}} b(u)\psi(u)du - \int_{\frac{k-\frac{1}{2}}{2^m}}^{\frac{k}{2^m}} b(u)\psi(u)du\right\}$$

$$\leq 2^{m/2}\left\{b\left(\frac{k-\frac{1}{2}}{2^m}\right)\int_{\frac{k-1}{2^m}}^{\frac{k-\frac{1}{2}}{2^m}} \psi(u)du - b\left(\frac{k-\frac{1}{2}}{2^m}\right)\int_{\frac{k-\frac{1}{2}}{2^m}}^{\frac{k}{2^m}} \psi(u)du\right\}$$

$$= b\left(\frac{k-\frac{1}{2}}{2^m}\right)c_{m,k}(\psi).$$

The proposition will be established if it can be shown that there exists $v \in (\frac{1}{2},1)$ such that $b(v) > 1$. Suppose such a v does not exist, so that $b(u) \le 1$, $\frac{1}{2} < u < 1$. Then since $b(u)$ is nonconstant, there exists an x, $\frac{1}{2} < x < 1$, such that $b(x) < 1$ and hence, since ϕ, ψ, b are necessarily positive, $\int_{\frac{1}{2}}^{1}(\phi(u))^2du = \int_{\frac{1}{2}}^{1}(b(u))^2(\psi(u))^2du$ $< \int_{\frac{1}{2}}^{1}(\psi(u))^2du$. But $\int_{\frac{1}{2}}^{1}(\phi(u))^2du = \int_{\frac{1}{2}}^{1}(\psi(u))^2du = \frac{1}{2}$ by the skew-symmetry of ϕ and ψ. This contradiction implies the existence of the required v such that $b(v) > 1$.

Thus if $\frac{k-\frac{1}{2}}{2^m} \ge v$, then $c_{m,k}(\phi) < c_{m,k}(\psi)$.

A similar analysis using the Haar expansion can be performed for the AMPRT's used for detecting a change in scale. Three such test statistics are Mood's test statistic generated by $\phi(u) = (u - \frac{1}{2})^2$, the Ansari-Bradley statistic generated by $\phi(u) = |u - \frac{1}{2}|$ and Capon's (Normal scores) statistic generated by $\phi(u) = (\Phi^{-1}(u))^2$ (*the weighting functions* have not been normalized). The necessary information about the Haar expansions of these weighting functions is tabulated in the last three columns of Table 5-1.

For Mood's statistic the following formulae apply: $c_{m,k} = (45)^{\frac{1}{2}}.2^{-5m/2-1}(2^m + 1 - 2k)$, $k = 1,\ldots,2^m$; $\psi_{1m} = 15.2^{-4m-2}(2^{2m} - 1)$; $\psi_{2m} = 1 - 2^{-2m-2}(5 - 2^{-2m})$ (for $m \ge 1$). The *coefficients* $c_{m,k}$ for the Ansari-Bradley statistic are related to those of the Mann-Whitney statistic by $c_{m,k}$ (Ansari-Bradley) $= 2c_{m,k}$ (Mann-Whitney) for $m \ge 1$.

Two observations may be made about the Haar expansions of the change-in-scale test statistics. One is that the coefficients for the Ansari-Bradley statistic have the same general form as those of the *Mann-Whitney* test, *suggesting* a robust test. The second feature is the poor convergence- measured by w_{2m} - of the Haar expansion for $(\Phi^{-1}(u))^2$. As with the Normal scores test in column 2 of Table 5-1, the weighting function for Capon's test statistic emphasizes the extreme intervals at each level. But while for the other four statistics represented in Table 5-1 most of the *potential efficiency* is accounted for by the first four levels, this does not apply to Capon's test at all. This assigning of relatively large weights to extreme intervals for large values of m may be taken to indicate instability in the performance of the statistic in the face of deviation from Normality by the underlying distribution.

5.5 Quick rank tests.

The coefficients $c_{r,k}$ computed in order to analyse $\phi(u)$ can also be used to derive a rank test which approximates the test $\phi(u)$. This test will have as its weighting function

$$\hat{\phi}(u) = \sum_{r=1}^{m} \sum_{k=1}^{2^r} c_{r,k}\, X_r^{(k)}(u),$$

and the ARE of this test against $\phi(u)$ when $\phi(u)$ is the AMPRT is given by

$$w_{2m} = \sum_{r=1}^{m} \sum_{k=1}^{2^r} c_{r,k}^2.$$

By choosing m *sufficiently large*, $\hat{\phi}$ can be made as 'near' to ϕ in terms of efficiency as we please because Parseval's identity for $\|\phi\|^2$ holds for the Haar system.

In fact the test $\hat{\phi}(u)$ is a member of the class of grouped rank tests proposed by Gastwirth (1966) since $\hat{\phi}(u)$ is constant over the intervals $(\frac{k-1}{2^{m+1}}, \frac{k}{2^{m+1}})$, $k = 1,\ldots,2^{m+1}$. In general grouped rank tests are *generated* by step functions which jump at $0 < \lambda_1 < \ldots < \lambda_r < 1$. If $\phi(u)$ is the AMPRT, then $\hat{\phi}(u)$ is not usually the asymptotically most powerful grouped rank test (AMPGRT) for the fractiles $k/2^{m+1}$, $k = 1,\ldots,2^{m+1}$, and there will exist a grouped rank test based on these fractiles which is more efficient against $\phi(u)$ than $\hat{\phi}(u)$ is.

These remarks suggest another systematic procedure by which we can obtain grouped rank tests which approximate $\phi(u)$ arbitrarily closely in terms of efficiency. At the mth stage of the procedure, compute the AMPGRT for the fractiles $k/2^{m+1}$. When the alternative hypthothesis being tested is one of location shift, this can be done by using lemma 4.1 of Gastwirth (1966). The consequent *sequence of* AMPGRT's has *increasing efficiency* against $\phi(u)$ and this efficiency can be made to be as close to one as required by taking m sufficiently large. This last statement is true because of the corresponding property for the sequence of $\hat{\phi}(u)$'s obtained for increasing m.

The advantage in using a grouped rank test is that the test statistic can be easily computed since it is just

$$S_N = \sum_{j=1}^{2^{m+1}} d_j \left(\sum_{i=P_{j-1}}^{P_j - 1} Z_{Ni} \right),$$

where $P_k = [Nk2^{-m-1}] + 1$, $k = 1,\ldots,2^{m+1}$ and $P_0 = 1$. Nevertheless the corresponding test can be nearly as efficient as the AMPRT even for small m, while the AMPRT may have a test statistic which turns out to be difficult to compute in practice.

Example. Consider the Normal scores test for location shift. To an accuracy of two decimal places, the grouped rank test *obtained from* the Haar expansion of $\Phi^{-1}(u)$ up to level two and the AMPGRT for the fractiles k/8, $k = 1,\ldots,8$ have the same weighting function. The weights on the intervals are as follows:

Interval	$(0,\frac{1}{8})$	$(\frac{1}{8},\frac{1}{4})$	$(\frac{1}{4},\frac{3}{8})$	$(\frac{3}{8},\frac{1}{2})$
Weight	-1.65	-.90	-.49	-.16

(the remaining weights follow by symmetry). For both tests, the ARE against the Normal scores test is .946 aginst the alternative hypothesis of location shift in the Normal distribution.

5.6 An influence curve for rank tests

In his doctoral dissertation, Hampel (1968) introduced the idea of an influence curve to measure the effect 'infinitesimal' changes in the underlying distribution F had upon the estimator T(F) defined as a functional of F. Essentially the influence curve is a derivative. Hampel's idea has come to be regarded as yielding a tool of fundamental importance for assessing the robustness of an estimator. In this section we propose to introduce an influence curve for two-sample rank tests which we hope will yield some of the interpretative possibilities of Hampel's influence curve.

While the main concern of this section is the influence curve, this section has also been used for gathering together some observations about the robustness of two-sample rank tests and some comments on directions for future research. It is quite apparent that robust has been used in many different senses in the literature. In the present discussion it means roughly that the performance of a robust test as measured by its power remains adequate under small deviations from the theoretical formulation of the testing problem. Obviously the precise nature of these deviations has to be decided before any investigation is feasible.

When discussing the influence curve, technical niceties will for the most part be omitted. Various assumptions are required if all the operations performed are to be justified. Justification for introducing the influence curve must rest principally upon the possibility of there being an acceptable interpretation for it. Rather than spend time justifying the validity of the derivation, let us concentrate instead upon finding out whether the result is meaningful. One assumption though is implicit in our derivations and deserves to be stat 1 here: all distributions are assumed to have absolutely continuous densities (with respect to Lebesgue measure on the real line).

The subsequent discussion will only concern the two-sample problem of location shift where the underlying distribution is F (density f). In this case the limiting power of a rank test with weighting function $\phi(u)$ $(0 < u < 1)$ is effectively determined by

$$(5.6.1)\quad \beta(\phi,F) = -\int_{-\infty}^{\infty} \phi(F(u))f'du/\{\int_0^1 (\phi(u) - \bar{\phi})^2\}^{\frac{1}{2}}$$

(this seems the most convenient form for the present purposes). This follows either from Hájek and Šidák (1967, p.227) or from the results of Chernoff and Savage (1958) (or subsequent versions under weaker conditions) together with Theorem 4-2-1 (i). It is the local behaviour of $\beta(\phi,F)$ in terms of F that will be studied.

There is another useful way of viewing (5.6.1) suggested by the study of the approximate slopes of standard sequences of test statistics which may also be helpful. In this case, let us consider the test statistic S_N generated by $\phi(u)$ and scaled so that I of Bahadur (1960, p.276) is true (for the Normal distribution). For our purposes this amounts to requiring that the stochastic limit of $N^{-\frac{1}{2}} S_N$ is 0 when H holds (that is when the location shift θ equals 0) and the limiting variance of S_N is 1. Then the stochastic limit of $N^{-\frac{1}{2}}S_N$ under K can be expressed as $(\lambda(1-\lambda))^{\frac{1}{2}}\theta\beta(\theta,F) + o(\theta)$. Studying $\beta(\phi,F)$ *therefore describes the limiting behaviour of* the actual test statistic when $\theta > 0$ is small (of course for the interpretation as a limiting power it is assumed $\theta \to 0$). It also effectively describes the behaviour of the approximate slope of the standard sequence $\{S_N\}$ (Bahadur (1960), p.293). Both these alternatives are useful to bear in mind - especially since the *stochastic limit* exists under *very weak conditions* - although our discussion will revolve around $\beta(\phi,F)$ as an expression of the limiting power.

Denote by Δ_x the probability measure determined by the point mass 1 at a given point x. Then the influence curve $IC(\phi,F,x)$ for the limiting power that will be *discussed in* this section is *defined pointwise by*

$$IC(\phi,F,x) = \lim_{\varepsilon\to 0+} \{\beta(\phi,(1-\varepsilon)F + \varepsilon\Delta_x) - \beta(\phi,F)\}/\varepsilon,$$

if this limit is defined for all x. This should be compared with the definition given by Hampel (1974, p.383) which will be referred to as *Hampel's influence curve.*

Much of the basic discussion of Hampel's IC applies equally well in the present context since it only assumes a functional of F. In particular, the relationship between the IC and the von Mises (Volterra) and Fréchet derivatives discussed on pp.38-39 of Hampel (1968) still apply. When the various derivatives exist they coincide with the IC.

In interpreting the IC the definition of the von Mises derivative gives us that

$$(5.6.2) \qquad \lim_{\varepsilon \to 0+} \{\beta(\phi,(1-\varepsilon)F + \varepsilon G) - \beta(\phi,F)\}/\varepsilon$$

$$= \int IC(\phi,F,x)\,dG(x),$$

so that if the true underlying distribution is not F but instead $(1-\varepsilon)F + \varepsilon G$, then for small ε, $\varepsilon\int IC(\phi,F,x)dG(x)$ measures the change in the performance of ϕ (interpretable in terms of limiting power). This interpretation not only makes the role of the IC clear, but also in fact facilitates discussion of the result in terms of limiting power when assumptions on the nature of the distributions are required to ensure asymptotic normality.

One difference between the interpretation of the IC for $\beta(\phi,F)$ and that for Hampel's IC deserves attention. While changes in absolute value concern us when considering the stability of an estimator, when considering $\beta(\phi,F)$ on the other hand positive values of the IC are of less concern, especially in a minimax sense. It is only a decrease in power which really concerns us. At the same time one should be wary of ignoring large positive values of the IC which reflect instability in the test statistic. It is probably best to avoid gross instability, even if superficially it seems to be beneficial.

In general the non-normalized IC can be obtained as follows:

$$(5.6.3) \qquad IC(\phi,F,x) = \int \phi'(F)\,id(\Delta_x - F) - \int \phi'(F)f'(\Delta_x - F)du$$

(' denotes derivative). To normalize the IC divide by $\{\int_0^1 (\phi(u) - \bar{\phi})^2 du\}^{\frac{1}{2}}$. The IC's of several tests have been obtained using (5.6.3) and will now be discussed in some detail.

For the Mann-Whitney test, the normalized IC is given by

$$(5.6.4) \qquad IC(u,F,x) = 4\sqrt{3}(f(x) - \int (f(u))^2 du).$$

Consequently the effects of contamination are felt most strongly around the *mode* of the theoretical distribution and in the tails. Contamination in the tails is especially harmful since $IC(x) \to -4\sqrt{3} \int (f(u))^2 du$. As an example, if $f(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2}$ (the standard Normal), then $IC(0) = (\frac{12}{\pi})^{\frac{1}{2}} (2^{\frac{1}{2}}-1) = .81$ and $IC(x) \to -(\frac{12}{\pi})^{\frac{1}{2}} = -1.95$ as $x \to \pm\infty$.

This implies a certain reliability in using the Mann-Whitney test, since the power is never going to be too drastically affected by a small amount of contamination. This is because for all 'nice' distributions the IC is bounded.

The most interesting distribution at which to evaluate the IC of the Normal scores test for location shift is naturally the standard Normal distribution. This is also the most amenable procedure. *Then we find that the normalized*

IC is obtained as

$$(5.6.5) \qquad IC(\Phi^{-1}(u),\Phi(u),x) = \tfrac{1}{2}(1-x^2),$$

so that $IC(x) \to -\infty$ as $x \to \pm\infty$. Interpretation of this is easy - contamination located in the tails of the Normal distribution has a profound adverse effect upon the limiting power of the *Normal scores test* (since the IC is unbounded there). The IC and our interpretation should be compared to Hampel's IC for the arithmetic mean and his interpretation thereof (Hampel (1968, 1974)).

The IC in (5.6.5) confirms the *impressions of section* 5.4 where the large weight attached to small subsets of the ranks was noted. Since the Normal scores test is not robust against contamination in a region where such contamination is very likely, the question then arises *of how to obtain* a test suited for when the theoretical underlying distribution is Normal, but also reasonably robust against this sort of contamination.

This leads to the consideration of asymptotically *minimax tests*. *Tests are* ordered according to the infimum of their limiting power fo[r l]ocation shift over a set of underlying distributions of the form $F = (1-\varepsilon)G + \varepsilon H$, where $0 \le \varepsilon < 1$ is fixed, G is a fixed symmetric, strongly unimodal *distribution, H is* a variable symmetric distribution and all distributions have densities. This set of distributions is called the gross-contamination model.

Under these circumstances $\beta(\phi,F)$ (and hence the limiting power) has a saddlepoint. That is there exists an $F_0 = (1-\varepsilon)G + \varepsilon H_0$ and a ϕ_0 such that

$$(5.6.6) \qquad \inf_F \beta(\phi_0,F) = \beta(\phi_0,F_0) = \sup_\phi \beta(\phi,F_0).$$

Then ϕ_0 generates a robust rank test in a well-defined minimax sense.

In fact (5.6.6) follows immediately from theorem 3 of Jaeckel (1971), in particular his extension of the result of Huber (1964) to cover R-estimators of location. The asymptotic variance of these estimators (suitably rescaled) equals the inverse of $\beta(\phi,F)$ for the corresponding tests and so the saddlepoint property for the asymptotic variances of R-estimators implies (5.6.6).

The least favourable distribution F_0 ensured by (5.6.6) is defined on p.1026 of Jaeckel (1971). In the case where $G = \Phi$, the least favourable distribution has density

$$(5.6.7) \qquad f_0(x) = \frac{(1-\varepsilon)}{\sqrt{2\pi}} \exp\{kx + \tfrac{1}{2}k^2\} \quad x \le -k$$

$$= \frac{(1-\varepsilon)}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}x^2) \qquad -k < x < k$$

$$= \frac{(1-\varepsilon)}{\sqrt{2\pi}} \exp\{-kx + \tfrac{1}{2}k^2\} \quad x \ge k,$$

where k is defined by

$$(1-\varepsilon)^{-1} = \int_{-k}^{k} (2\pi)^{-\frac{1}{2}}\, e^{-\frac{1}{2}x^2}\, dx + \frac{1}{k}\left(\frac{2}{\pi}\right)^{\frac{1}{2}} e^{-\frac{1}{2}k^2}$$

(p.81 of Huber (1964)). The minimax test is then the AMPRT for f_0 and is generated by

$$(5.6.8) \qquad \phi_0(u) = -k \qquad 0 \le u \le (1-\varepsilon)\Phi(-k) + \varepsilon/2$$

$$= \Phi^{-1}\left(\frac{u-\varepsilon/2}{1-\varepsilon}\right) \qquad (1-\varepsilon)\Phi(-k) + \varepsilon/2 < u < (1-\varepsilon)\Phi(k) + \varepsilon/2$$

$$= k \qquad (1-\varepsilon)\Phi(k) + \varepsilon/2 \le u \le 1.$$

In the light of (5.6.8), a fairly natural test to use with the contaminated Normal is the one generated by

$$(5.6.9) \qquad \phi(u) = -k \qquad 0 \le u \le \Phi(-k)$$

$$= \Phi^{-1}(u) \qquad \Phi(-k) < u < \Phi(k)$$

$$= k \qquad \Phi(k) \le u \le 1$$

for some $k > 0$. In this case the non-normalized IC computed at the standard Normal gives

$$(5.6.10) \qquad IC(\phi,\Phi,x) = -\tfrac{1}{2}(k^2+1)(\Phi(k) - \Phi(-k)) - kf(k) \qquad |x| > k$$

$$= \tfrac{1}{2}(k^2-x^2) + 1 - \tfrac{1}{2}(k^2+1)(\Phi(k) - \Phi(-k)) - kf(k) \qquad |x| \le k,$$

where $f(k) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}k^2}$. In this case the IC is bounded and becomes constant as $x \to \pm \infty$. Notice the jump at $|x| = k$ *essentially because ϕ does not have a continuous derivative.* Whether this has any particular significance is not too clear, but it seems doubtful that it should. A distinction should be made here with Hampel's IC since in that case continuity of the IC is desirable to insure against the effects of round-off and grouping on the estimate. This consideration does not apply to the limiting power. It may however be useful to have the IC continuous to insure against the effects this kind of inaccuracy may have upon the test statistic (as indicated by the stochastic limit of the test statistic).

To normalize $IC(\phi,\Phi,x)$, divide by $(2k^2-1)\Phi(-k)+1-2kf(k)$.

The IC for the median test reveals great sensitivity to deviations at the median of the theoretical distribution. Using δ_x and I_x to denote the Dirac-delta and Heaviside functions respectively (I_x is the indicator function for $[x,\infty)$), the normalized IC when $\phi(u)$ generates the median test is given by

$$(5.6.11) \qquad IC(\phi,F,x) = 2\{\delta_{F^{-1}(\frac{1}{2})}(x) - f(F^{-1}(\tfrac{1}{2})) - f'(F^{-1}(\tfrac{1}{2}))/f(F^{-1}(\tfrac{1}{2}))(\tfrac{1}{2} - I_{F^{-1}(\frac{1}{2})}(x))\},$$

so that

$$IC(x) = -2f(F^{-1}(\tfrac{1}{2})) - f'(F^{-1}(\tfrac{1}{2}))/f(F^{-1}(\tfrac{1}{2})) \qquad x < F^{-1}(\tfrac{1}{2})$$

$$= -2f(F^{-1}(\tfrac{1}{2})) + f'(F^{-1}(\tfrac{1}{2}))/f(F^{-1}(\tfrac{1}{2})), \qquad x > F^{-1}(\tfrac{1}{2}).$$

If $f(x)$ is symmetric and unimodal, then $IC(x) = -2f(0)$ for $x \neq F^{-1}(\tfrac{1}{2})$. At $x = F^{-1}(\tfrac{1}{2})$, there is an 'infinite spire' and in addition there is also a jump of size $\frac{-2f'(F^{-1}(\frac{1}{2}))}{f(F^{-1}(\frac{1}{2}))}$ (which will be zero if f is symmetric and unimodal). This infinite spire occurs as the limiting case of a jump in the central scores (compare this with the IC of the median on p.385 of Hampel (1974)).

The grouped rank tests introduced in section 5.5 have IC's similar in many respects to that of the median test. In particular there will be infinite spires corresponding to jumps in the weighting function. Again these are the limiting cases of sensitivity to a jump in the scores. The IC of a grouped rank test is however bounded, apart from these infinite spires.

In order to consider the implications of the delta function further, recall that

$$\lim_{\varepsilon \to 0+} \{\beta(\phi,(1-\varepsilon)F + \varepsilon G) - \beta(\phi,F)\}/\varepsilon = 2(g(0) - f(0)),$$

provided f is symmetric and unimodal. In the limit the entire effect of a small amount of contamination is concentrated at the median. This concentration at a single point is all that the delta function implies and it should not be construed as indicating great instability in the median test.

The basic motivation for introducing grouped rank [illegible] - that the test statistics are easy to compute and [illegible] the test can be made to have a good ARE relative to the AMPRT - seems to be a worthwhile objective. The sensitivity of grouped rank tests at jumps in the weighting functions indicated by the IC is slightly disturbing. Perhaps a class of tests with test statistics which are still easy to compute and which are related to the Mann-Whitney test might be a better proposition. Certainly all the indications are that the Mann-Whitney test will provide a good guide.

For these reasons we would like to suggest for further investigation a class of two-sample rank tests generated by continuous, piecewise linear functions. These functions are defined by taking points $0 = \lambda_0 < \ldots < \lambda_r = 1$ and defining the weighting function ϕ at these points as being

$$(5.6.12) \quad \phi(\lambda_0) = 0; \quad \phi(\lambda_i) = \alpha_i(\lambda_i - \lambda_{i-1}) + \phi(\lambda_{i-1}), i=1,\ldots,r$$

for real numbers $\alpha_1,\ldots,\alpha_r$. The definition of ϕ is completed by taking it to be linear over the intervals $[\lambda_{i-1},\lambda_i]$, $i = 1,\ldots,r$.

For actual use, the test statistic might be expressed as

$$S_N = \sum_{j=1}^{r} \alpha_j \left(\sum_{i=p_{j-1}}^{p_j - 1} i \, Z_{Ni} \right)$$

where $P_j = [N\lambda_j] + 1$, $j = 0,\ldots,r$. The Mann-Whitney test corresponds to $r = 1$, $\alpha_1 = 1$.

The non-normalized IC is given by

$$(5.6.13) \quad IC(\phi,F,x) = 2\alpha_\ell f(x) - 2\sum_{i=1}^{r} \alpha_i \int_{J_i} f^2 - \alpha_\ell f(F^{-1}(\lambda_\ell))$$

$$+ \sum_{i=1}^{r} \alpha_i \{\lambda_i f(F^{-1}(\lambda_i)) - \lambda_{i-1} f(F^{-1}(\lambda_{i-1}))\}$$

$$- \sum_{i=\ell+1}^{r} \alpha_i \{f(F^{-1}(\lambda_i)) - f(F^{-1}(\lambda_{i-1}))\},$$

where $J_i = [F^{-1}(\lambda_{i-1}), F^{-1}(\lambda_i))$, $i = 1,\ldots,r$ and $x \in J_\ell$.

The expression (5.6.13) can be simplified by considering the case where f is symmetric about 0. Then it is quite reasonable to require $\lambda_0 < \ldots < \lambda_{2p}$ satisfying $\lambda_k + \lambda_{2p-k} = 1$, $k = 0,\ldots,p$ and also $\alpha_i = \alpha_{2p+1-i}$, $i = 1,\ldots,p$. Since it is principally contamination in the tails which concerns us, it is interesting to consider the limit of $IC(x)$ as $x \to \pm\infty$. Under our assumptions this is

$$(5.6.14) \quad \lim_{x\to\pm\infty} IC(x) = \sum_{i=1}^{2p} \alpha_i \{\lambda_i f(F^{-1}(\lambda_i)) - \lambda_{i-1} f(F^{-1}(\lambda_{i-1})) - 2\int_{J_i} f^2\}.$$

The limit here should be interpreted as a requirement that $|x|$ should be sufficiently large so that $x \in J_1$ (or J_{2p}) and $\alpha_1 f(x)$ is negligible.

Some interesting possibilities arise from (5.6.14). One reasonable course would be to choose $\alpha_1,\ldots,\alpha_{2p}$ so as to maximize the right hand side of (5.6.14) subject to the constraints that $\alpha_i \geq 0$ and $\int_0^1 (\phi(u) - \bar{\phi})^2 du = 1$. This is a routine variational problem and would yield the locally most robust test to contamination in the tails (among the tests generated by continuous, piecewise linear functions).

This point will not be discussed further here as it belongs to the detailed study of the special class of tests generated by continuous, piecewise linear functions rather than to a general discussion about influence curves. While the subject is before us however, another aspect of these *special tests might be mentioned.* Can asymptotically most powerful piecewise linear rank tests be obtained in the same way that lemma 4.1 of Gastwirth(1966) gives the AMPGRT's for location shift *with a specified underlying distribution?* The coefficients $\alpha_1,\ldots,\alpha_r$ of such an optimal test for given $\lambda_0,\ldots,\lambda_r$ can be expressed explicitly and the problem is probably well known in approximation theory, although we have not yet found a specific reference to it.

Our method for solving this problem will be briefly outlined although the details of this and other aspects of these tests are to be reported elsewhere than in this thesis. The problem reduces to the best approximation in the L_2 norm of the weighting function of the AMPRT by a continuous, piecewise linear function. A basis of r-1 roof functions on the intervals $[\lambda_{i-1},\lambda_{i+1}]$ (with peak at λ_i), $i = 1,\ldots,r-1$ then enables us to set up the normal equations for the minimum norm approximation. To solve these equations requires finding the inverse of a symmetric tridiagonal (Jacobi) matrix which can be determined as a Green's matrix. The coefficients α_i can then be obtained explicitly.

The remarks of the last two paragraphs provide us with two different ways of choosing piecewise linear rank tests for a given symmetric density. Here Hampel's suggestion (Hampel (1968, p.36)) might prove useful and we confine our attention to those tests which cannot be simultaneously bettered in both respects by any other piecewise linear test. Preliminary investigations into continuous, piecewise linear tests suggest that they do indeed offer a promising class of tests.

Some conclusions based on this chapter would seem appropriate. In the light of the JC and the Haar expansion, the Normal scores test for location shift must be construed as not robust against the possibility of gross-contamination. It is really rather like the mean considered as an estimate of location for the Normal distribution. This is in spite of the result obtained by Doksum (1966) indicating a favourable minimax property for the Normal scores test. In that case though a variance requirement on the distributions under consideration eliminated contamination by heavy-tailed distributions.

Remedies for this lack of robustness on the part of the Normal scores test have been suggested by way of the weighting functions (5.6.8) and (5.6.9). One wonders if perhaps these tests are not overpessimistic in the sense that they arise from the full neighbourhood implied by the gross-contamination model. This allows the heavy tails of the double exponential type distribution to determine the weighting for the extreme ranks. It might be an idea to replace the horizontal line at either extreme of the weighting function by a line with gradient chosen so that the resulting curve is smooth (differentiable). This seems a less radical option than simply grouping the extreme observations and is suggested by analogy with the Mann-Whitney test. This is not unreasonable in the light of the minimax result, since it means replacing the tails of the least favourable distribution by the tails of a logistic distribution, which is often regarded as a 'smoothed out' version of the double exponential.

The Mann-Whitney test handles satisfactorily the contamination of any theoretical distribution which does not have very heavy tails. The Haar expansion of the weighting function also confirms this. Certainly as long as the Mann-Whitney test is reasonably efficient for the theoretical distribution, this test can be used without fear of a sharp loss of limiting power through contamination. The result obtained by Doksum (1966) for different alternatives is another point in favour of the Mann-Whitney test.

The median test is essentially robust, although very sensitive to distortion at the median. Its low efficiency for testing location shift with light-tailed distributions rather militates against it though.

Naturally the influence curve discussed here can be applied to any alternative hypothesis - and not just location shift - provided the expressions for the limiting power are sufficiently regular. The influence curve and the Haar expansion can be used together, the one supplying information about the limiting power and the other supplying a detailed picture of the test's efficiency properties. The influence curve in particular seems to be a very useful heuristic tool.

LIST OF REFERENCES

If an article is in a language other than English, then the appropriate reference in '*Mathematical Reviews*' is included wherever this is possible. If the author is aware of an English translation of such an article, this is also cited.

ABRAHAMSON, I.G. (1967). Exact Bahadur efficiencies for the Kolmogorov-Smirnov and Kuiper one- and two-sample statistics. Ann. Math. Statist. 38, 1475-1490.

AJNE, B. (1968). A simple test for uniformity *of a circular* distribution. Biometrika 55, 343-354.

ALEXITS, G. (1961). Convergence Problems of Orthogonal Series. *Budapest: Pergamon Press.*

BAHADUR, R.R. (1960). Stochastic comparison of tests. Ann. Math. Stat[illegible] 31, 276-2[illegible]5.

BAHADUR, R.R. (1[illegible] [illegible]es of convergence of estimates and test statist[illegible]. Ann. Math. Statist. 38, 303-324.

BARR, D.R. and SHUDDE, R.H. (1973). A note on Kuiper's V_n statistic. Biometrika 60, 663-664.

BEHNEN, K. (1975). The Randles-Hogg test and an alternative proposal. Comm. *Statist.* 4, 203-238.

BELYAYEV, Yu. K. (1972). Point processes and first passage problems. Proceedings of the Sixth Berkely Symposium *on Mathematical* Statistics and Probability, Vol III. (ed. Le Cam, L.M., Neyman, J. and Scott, E.L.), 1-17. Berkely: University of California Press.

BENDER, E.A. and GOLDMAN, J.R. (1975). On the applications of Möbius inversion in combinatorial analysis. Amer. Math. Monthly 82, 789-803.

BERAN, R.J. (1969). The derivation of nonparametric two-sample tests from tests for uniformity of a circular distribution. Biometrika 56, 561-570.

BERAN, R.J. (1975). Local asymptotic power of quadratic rank tests for trend. Ann. Statist. 3, 401-412.

BERGE, C. (1971). Principles of Combinatorics. New York: Academic Press.

BHATTACHARYYA, G.K. and JOHNSON, R.A. (1969). On Hodge's bivariate sign test and a test for uniformity of a circular distribution. Biometrika 56, 446-449.

BILLINGSLEY, P. (1968). Convergence of Probability Measures. New York: John Wiley.

BIRNBAUM, Z.W. (1955). On a use of the Mann-Whitney statistic. Proceedings of the Third Berkely Symposium on Mathematical Statistics and Probability, Vol. I (ed. Neyman, J.), 13-17. Berkely: University of California Press.

CAPON, J. (1961). Asymptotic efficiency of certain locally most powerful rank tests. Ann. Math. Statist. 32, 88-100.

CHERNOFF, H. and SAVAGE, I.R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. Ann. Math. Statist. 29, 972-994.

DOKSUM, K. (1966). Asymptotically minimax distribution-free procedures. Ann. Math. Statist. 37, 619-628.

DOKSUM, K. (1969). Starshaped transformations and the power of rank tests. Ann. Math. Statist. 40, 1167-1176.

EPANECHNIKOV, V.A. (1968). The significance level and power of the two-sided Kolmogorov test in the case of small samples (in Russian). Teor. Verojatnost. i Primenen. 13, 725-730. [MR. 39, # 3664]. English translation: Theor. Probability Appl. 13 (1968), 686-690.

FERNIQUE, X. (1964). Continuité des processus Gaussiens. Comptes Rendus Acad. Sci. Paris 258, 6058-6060. [MR. 29, # 1662].

GASTWIRTH, J.L. (1965). Percentile modifications of two-sample rank tests. J. Amer. Stat. Ass. 60, 1127-1141.

GASTWIRTH, J.L. (1966). On robust procedures. J. Amer. Stat. Ass. 61, 929-948.

GASTWIRTH, J.L. (1970). On robust rank tests. Nonparametric Techniques in Statistical Inference, 89-109. Cambridge: University Press.

HÁJEK, J. (1962). Asymptotically most powerful rank-order tests. Ann. Math. Statist. 33, 1124-1147.

HÁJEK, J. and ŠIDÁK, Z. (1967). Theory of Rank Tests. New York: Academic Press.

HÁJEK, J. (1969). Nonparametric Statistics. San Francisco: Holden-Day Inc.

HÁJEK, J. (1974). Asymptotic sufficiency of the vector of ranks in the Bahadur sense. Ann. Statist. 2, 75-83.

HALL, M. (1970). The Theory of Groups. New York: Holt, Rinehart and Winston.

HAMPEL, F.R. (1968). Contributions to the Theory of Robust Estimation. Unpublished Ph.D. thesis, University of California, Berkely.

HAMPEL, F.R. (1971). A general qualitative definition of robustness. Ann. Math. Statist. 42, 1887-1896.

HAMPEL, F.R. (1974). The influence curve and its role in robust estimation. J. Amer. Stat. Ass. 69, 383-393.

HEWITT, E. and STROMBERG, K. (1969). Real and Abstract Analysis. New York: Springer-Verlag.

HOEFFDING, W. (1953). On the distribution of the expected values of the order statistics. Ann. Math. Statist. 24, 93-100.

HUBER, P.J. (1964). Robust estimation of a location parameter. Ann. Math. Statist. 35, 73-101.

JAECKEL, L.A. (1971). Robust estimates of location: symmetry and asymmetric contamination. Ann. Math. Statist. 42, 1020-1034.

JUREČKOVÁ, J. (1969). Asymptotic linearity of a rank statistic in regression parameter. Ann. Math. Statist. 40, 1889-1900.

KAC, M. and SIEGERT, A.J.F. (1947). An explicit representation of a stationary Gaussian process. Ann. Math. Statist. 18, 438-442.

KILLEEN, T.J. and HETTMANSPERGER, T.P. (1972). Bivariate tests for location and their Bahadur efficiencies. Ann. Math. Statist. 43, 1507-1516.

KREWERAS, G. (1965). Sur une classe de problèmes de dénombremont liés au treillis des partitions des entiers. Cahiers du Bur. Univ. de Recherche Opér. (ISUP) 6, 9-103.

KUIPER, N.H. (1960). Test concerning random points on a circle. Proc. Kon. Ned. Akad. Wet. A63, 38-47.

LEHMANN, E.L. (1953). The power of rank tests. Ann. Math. Statist. 24, 23-43.

LEHMANN, E.L. (1959). Testing Statistical Hypotheses. New York: John Wiley.

LEHMANN, E.L. (1975). Nonparametrics: Statistical Methods Based on Ranks. San Francisco: Holden-Day Inc.

LE VEQUE, W.J. (1965). Topics in Number Theory, Vol. I. Reading, Massachusetts: Addison-Wesley Publishing Company Inc.

LOÈVE, M. (1960). Probability Theory. New York: van Nostrand.

MANN, H.B. and WHITNEY, D.R. (1947). On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Statist. 18, 50-60.

MARCUS, M.B. (1970). A bound for the distribution of the maximum of continuous Gaussian processes. Ann. Math. Statist. 41, 305-309.

MARDIA, K.V. (1972). Statistics of Directional Data. London: Academic Press.

MASSEY, F.J. (1950). A note on the power of a non-parametric test. Ann. Math. Statist. 21, 440-443.

MOHANTY, S.G. (1971). A short proof of Steck's result on two-sample Smirnov statistics. Ann. Math. Statist. 42, 413-414.

MOHANTY, S.G. (1977). Acknowledgement of Priority to "A short proof of Steck's result on two-sample Smirrov statistics." Ann. Statist. 5, 429.

NANDI, H.K. (1963). On some properties of Roy's union-intersection tests. Cal. Stat. Ass. Bull. 14, 9-13.

NOETHER, G.E. (1955). On a theorem of Pitman. Ann. Math. Statist. 26, 64-68.

PROHOROV, Yu. V. (1956). Convergence of random processes and limit theorems in probability theory (in Russian). Teor. Verojatnost. i Primenen. 1, 177-238 [MR. 18, # 943]. English translation: Theor Probability Appl. 1, 157-214.

PURI, M.L. (editor) (1970). Nonparametric Techniques in Statistical Inference. Cambridge: University Press.

PURI, M.L. and SEN, P.K. (1971). Nonparametric Methods in Multivariate Analysis. New York: John Wiley.

RANDLES, R.H. and HOGG, R.V. (1971). Certain uncorrelated and independent rank statistics. J. Amer. Stat. Ass. 66, 569-574.

RANDLES, R.H. and HOGG, R.V. (1973). Adaptive distribution-free tests. Comm. Statist. 2, 337-356.

RIORDAN, J. (1958). An Introduction to Combinatorial Analysis. New York: John Wiley.

ROY, S.N. (1953). On a heuristic method of test construction and its use in multivariate analysis. Ann. Math. Statist. 24, 220-238.

SCHACH, S. (1969a). The asymptotic distribution of some non-linear functions of the two-sample rank vector. Ann. Math. Statist. 40, 1011-1020.

SCHACH, S. (1969b). On a class of nonparametric two-sample tests for circular distributions. Ann. Math. Statist. 40, 1791-1800.

SCHACH, S. (1970). The asymptotic distribution of a class of nonparametric test statistics. Metrika 15, 48-58.

STECK, G.P. (1969). The Smirnov two-sample tests as rank tests. Ann. Math. Statist. 40, 1449-1466.

STECK, G.P. (1971). Rectangular probabilities for uniform order statistics and the probability that the empirical distribution function lies between two distribution functions. Ann. Math. Statist. 42, 1-11.

VAN DANTZIG, D. (1951). On the consistency and power of Wilcoxon's two sample test. Proc. Kon. Ned. Akad. Wet. A 54, 1-8.

VAN EEDEN, C. (1963). The relation between Pitman's asymptotic relative efficiency of two tests and the correlation coefficient between their test statistics. Ann. Math. Statist. 34, 1442-1451.

VINCZE, I. (1972). On some results and problems in connection with statistics of the Kolmogorov-Smirnov type. Proceedings of the Sixth Berkely Symposium on Mathematical Statistics and Probability, Vol. I (ed. Le Cam, L.M., Neyman, J. and Scott, E.L.), 459-470.

WATSON, G.S. (1962). Goodness-of-fit tests on a circle, II. Biometrika 49, 57-63.

WHEELER , S. and WATSON, G.S. (1964). A distribution-free two-sample test on a circle. Biometrika 51, 256-257.

WICHURA, M.J. (1971). A note on the weak convergence of stochastic processes. Ann. Math. Statist. 42, 1769-1772.

WOODWORTH, G.G. (1970). Large deviations and Bahadur efficiency of linear rank statistics. Ann. Math. Statist. 41, 251-283.

APPENDIX 1

Some Comments on the Theorems of Steck

As mentioned in section 2.2, the main theorem of that section (Theorem 2-2-2) is suggested by results obtained by Steck in connection with the Kolmogorov-Smirnov statistics. These results are described below together with the appropriate references and an outline of how each result may be derived using the ideas of Lemma 2-2-4. The notation of section 2.2 is used wherever possible.

Theorem A1-1. Define $\binom{t}{s}_+$ as follows:

$$\begin{aligned}\binom{t}{s}_+ &= \binom{t}{s} && \text{if } 1 \le s \le t \\ &= 0 && \text{if } t < s.\end{aligned}$$

Then

$$|\chi_m(u,v)| = \det\left(\binom{c_i - b_j + j - i - 1}{j - i + 1}_+\right) \tag{A1.1}$$

Proof. This follows directly from Lemma 2-2-4 by taking $X^*(S) = |S|$ for finite subsets S of Z^m. With this definition of X^* it then follows that

$$\begin{aligned}X^*(\ell(i,j)) &= |\ell(i,j)| \\ &= \binom{v_i - u_j + 1}{j - i + 1}_+.\end{aligned}$$

Substituting for u_i and v_i in terms of b_i and c_i, (A1.1) follows from (2.2.10).

Theorem A1-1 corresponds to theorem 4.1 of Steck (1961) after bearing in mind that, from the way u_i and v_i were defined,

$$b_i < R_i < c_i \iff u_i \le T_i \le v_i \qquad (i = 1,\ldots,m).$$

The proof of Theorem A1-1 given by Steck (1969) rests upon expressing the determinant in (A1.1) as a recurrence relationship and then (by way of fairly lengthy calculations) verifying that this recurrence formula is indeed true. The proof of Theorem A1-1 given by Mohanty (1971) rests upon a *neat induction argument* and is much shorter than Steck's proof. The merit of a proof of the theorem using Lemma 2-2-4 would seem to be that it brings the result *into line* with a host of other applications of inclusion-exclusion (or more generally, generalized Möbius inversion). It also seems to the author that from Lemma 2-2-4 the form of the result as a determinant becomes more transparent. For instance Mohanty certainly proves the result, but - as so often with ingenious induction arguments - at the end of the proof one still does *not really understand why it worked*.

Theorem A1-2 is a result applicable to the (one-sample) Kolmogorov test and concerns order statistics rather than ranks. It does not, therefore, fall within the main theme of this thesis. Its importance, though, is that Steck (1971) proved the result using an inclusion-exclusion argument. It might consequently be expected that the proof of Lemma 2-2-4 would be similar to the proof of theorem 2 of Steck (1971). This however is not the case as will be explained in more detail.

In order to state Theorem A1-2 some preliminary notation is required. Suppose that $U^{(1)} \leq \ldots \leq U^{(m)}$ are the order statistics from a sample of m independent random variables uniformly distributed over [0,1]. Also take numbers $u_1,\ldots,u_m$, $v_1,\ldots,v_m$ such that $0 \leq u_1 \leq \ldots \leq u_m \leq 1$ and $0 \leq v_1 \leq \ldots \leq v_m \leq 1$ and also $u_i < v_i$, $i = 1,\ldots,m$.

Theorem A1-2. If $(x)_+$ is defined by $(x)_+ = \max(x,0)$, then

$$\text{(A1.2)}\quad P(u_i \leq U^{(i)} \leq v_i, i = 1,\ldots,m) = m!\det\{((v_i-u_j)_+^{j-i+1}/(j-i+1)!)\}.$$

Outline Proof. The basic ideas of Lemma 2-2-4 can be used. In chapter 2 there are certain restrictions included in order to simplify the discussion. These may be removed without affecting the validity of the proof of Lemma 2-2-4. Specifically X^* may be supposed to be a finitely additive function defined on certain subsets of E^m (m-dimensional Euclidean space). All the sets required for Lemma 2-2-4 must be suitably redefined for $x_1,\ldots,x_m$ real numbers and X^* must be defined on these sets. It is also required that

$$X^*(\mathcal{L}(1,[\omega(m)]_1) \times\ldots\times \mathcal{L}([\omega(m)]_{k-1} + 1,m)) =$$

$$X^*(\mathcal{L}(1,[\omega(m)]_1)) \times\ldots\times X^*(\mathcal{L}([\omega(m)]_{k-1} + 1,m)).$$

If X^* is the distribution of $(U_1, \ldots, U_m)$, where the U_i are independent random variables uniformly distributed over $[0,1]$, the necessary conditions are satisfied. Then

$$X^*(\mathcal{L}(i,j)) = P\{v_i \geq U_i > \ldots > U_j \geq u_j\}$$
$$= \{(v_i - u_j)_+^{j-i+1}\}/(j - i + 1)!$$

and (A1.2) follows from the revised form of Lemma 2-2-4.

In addition to the proof offered in Steck (1971), Steck also proved Theorem A1-2 by taking the limiting case of his two-sample result (Theorem A1-1). Mohanty (1971) also indicates a proof of Theorem A1-2 using an induction argument and Epanechnikov (1968) proves an equivalent result to Theorem A1-2 by verifying that the appropriate recurrence formula holds. Both the proofs of Epanechnikov and Mohanty involve multiple integrals of some degree of complexity. A proof of the result which uses the principle of inclusion-exclusion may consequently be deemed of some interest.

The rest of this appendix will be an attempt by the author to convince the reader that Steck's proof using the principle of inclusion-exclusion is not correct. In fact originally the author proposed a proof of Theorem 2-2-2 which closely followed Steck's argument. It was the detection of a flaw in that proof which led to the following discussion.

There seems to be little point in simply transposing Steck's proof on pp.2-4 of his paper (Steck (1971)) when the interested reader can consult the relevant material himself.

Some of the notation necessary for the discussion is introduced here even though all the notation is the same as in Steck's article.

Put $A = \{(x_1,\dots,x_m) : 0 \le x_i \le 1, \text{ all } i\}$ and $\Omega = \{(x_1,\dots,x_m) : 0 \le x_1 \le \dots \le x_m \le 1\}$. In addition let $F_k = \{(x_1,\dots,x_k) : u_i \le x_i \le v_i, i \le k \text{ and } 0 \le x_1 \le \dots \le x_k \le 1\}$ and then define $B_k = \{(x_1,\dots,x_m) : (x_1,\dots,x_{m-k}) \in F_{m-k}$ and $u_m \le x_{m-k+1} \le \dots \le x_m \le v_{m-k+1}\}$. A lot of the confusion then seems to arise from the assertion that if $\omega \in \bigcup_{k=1}^{m} B_k$, then there corresponds a unique largest integer $k_o(\omega)$ such that $\omega \in B_k$ for $k \le k_o$ (lines below equation (2.2) on p.3 of Steck's paper). Clearly this is true if $\omega \in \Omega$. Saying that $(x_1,\dots,x_m) \in B_k$ implies that $u_i \le x_i \le v_i$, $i = 1,\dots,m-k$ and $u_m \le x_i \le v_{m-k+1}$, $i = m-k+1,\dots,m$ so that if $u_m < v_{m-k}$, we can suppose that $x_{m-k+1} < x_{m-k}$, in which case $(x_1,\dots,x_m) \notin B_{k-1}$ since $(x_1,\dots,x_{m-k+1}) \notin F_{m-k+1}$. One can easily verify that the set of values for which this happens has non-zero probability (remember that each co-ordinate of the m-dimensional vector is assumed to be independently, uniformly distributed over [0,1]).

The remarks of the last paragraph may explain why Steck's proof is difficult to follow from then on. It seems that the proof rests upon the following statement: if $A_r = \{\omega : k_o(\omega) = r\}$, then

$$P(B_k) = \sum_{r=k}^{m} \binom{r}{k} P(F_m \cap A_r), \quad k = 1,\dots,m. \tag{A1.3}$$

Steck has certainly shown that

$$P(B_k) \geq \sum_{r=k}^{m} \binom{r}{k} P(F_m \cap A_r),$$

but does not seem to have established the full implication of (A1.3). *In more detail, it seems that* Steck shows that for instance $P(B_1 \cap A_r) \geq rP(F_m \cap A_r)$, but claims $P(B_1 \cap A_r) = rP(F_m \cap A_r)$.

Perhaps the most *serious observation* to be made about Steck's proof is that (A1.3) *does not seem* to be necessarily true. To show this, consider the *special* case of B_1. Let us suppose that $u_m < v_{m-2}$; $v_{m-3} < v_{m-2} < v_{m-1} < v_m$. Then in this special case if $(x_1,\ldots,x_m) \in F_m \cap A_r$, *then* $(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_m,x_i) \in B_1$, $i = m-r+1,\ldots,m$, *and if* the (disjoint) subsets of B_1 defined thus are denoted by $A_r(i)$, $i = m - r + 1,\ldots,m$; $r = 1,\ldots,m$, then

$$\text{(A1.4)} \qquad \bigcup_{r=1}^{m} \bigcup_{i=m-r+1}^{m} A_r(i) \subset B_1.$$

In addition $P(A_r(i)) = P(F_m \cap A_r)$. These last two sentences reflect the essential ideas of Steck's argument.

Consider the subset E of B_1 containing elements $x = (x_1,\ldots,x_m)$ which satisfy

$$\max\{v_{m-3},u_m\} \leq x_m < x_{m-2}; \quad v_{m-2} < x_{m-1} < v_{m-1}.$$

Then such $(x_1,\ldots,x_m)$ satisfy the following

(a) $\bar{x} = (x_1,\ldots,x_{m-3},x_m,x_{m-2},x_{m-1}) \in \Omega$

(b) since $x_{m-1} \notin [u_m,v_{m-2}]$, $k_o(\bar{x}) = 2$.

But then x - which does not have final component x_{m-2} or x_{m-1} - must satisfy $x \notin A_2(m+1-i)$, $i = 1,2$ and hence $x \notin \cup\cup A_r(i)$. Furthermore (obviously assuming $P(F_{m-3}) > 0$),

$$P(E) = \{(v_{m-2} - \max(v_{m-3},u_m))^2/2!\}(v_{m-1} - v_{m-2})P(F_{m-3}) > 0,$$

so that in view of (A1.4),

$$P(B_1) > \sum_{r=1}^{m} rP(F_m \cap A_r).$$

<u>Note</u>. After much of this thesis had been compiled, the author's attention was drawn to a question of priority over the discovery of Theorem A1-1. As a note by Mohanty (1977) indicates, the result is a special case of a theorem due to Kreweras (1965). This theorem is stated on pp.69-70 of Berge (1971), who indicates that the proof is by induction.

APPENDIX 2

The Null Distribution of the Mann-Whitney Type Tests

As indicated in Part I of the thesis, the question of approximations for obtaining the critical values of either of the Mann-Whitney type tests is not completely resolved. This appendix contains tables of the exact tail probabilities for the values of $\tilde{\xi}_{N1}$ and $\tilde{\xi}_N$ under the hypothesis of randomness for the sample sizes m = n = 4,5,6,7,8,9,10.

The tables were generated by computer and then the cases m = n = 4 and m = n = 5 were checked by hand.

Clearly far more extensive tabulation is possible and there is no need to consider only equal sample sizes. The only limitation to this method of obtaining the probabilities is the increasing number of calculations necessary as N increases. In fact in this regard the exploitation of various symmetries in the definitions of the test statistics can cut down the number of calculations required in order to obtain the probabilities. The relationships of section 2.1 are also useful in this respect. Nevertheless it was felt that g[illegible] pages and pages of tables did not really resolve the problem and could easily be done at any time should the need arise.

The recurrence relation given by (2.2.20) is unfortunately of little use as far as generating tables of probabilities is concerned since it introduces the vectors b and c into consideration as well. It might be used, however, to obtain a single probability provided m or n is small. This would give us $P(\tilde{\xi}_N \geq c)$ for a particular value of c which is all that is required to perform the test.

The approximation offered by Lemma 3-3-8 was tried out for the case m = n = 10 and found to be hopelessly inadequate. This was in spite of a certain amount of 'streamlining' of the approximation for this particular case. At the same time this was not really much of a surprise because of there being approximation in two directions, namely the asymptotic convergence and then the approximation to get the required probability from the limiting result. It is the accuracy of the second of these two approximations which is of major concern and requires further study, especially since Lemma 3-3-8 was only a general result developed for the study of continuity conditions for Gaussian processes.

Table A2-1 and A2-2 list the tail probabilities for $\tilde{\xi}_{N1}$ and $\tilde{\xi}_N$ respectively. For interest's sake we have also included the complete set of probabilities for $\tilde{\xi}_{N1}$ and $\tilde{\xi}_N$ when m = n = 10. These are Tables A2-3 and A2-4 respectively. They may possibly be of use in assessing approximate distributions for the two test statistics.

TABLE A2-1 Tail probabilities for $\tilde{\xi}_{N1}$.

$p = \text{Prob}(\tilde{\xi}_{N1} \geq e)$.

	e	p
m = n = 4	26	.114
	25	.229
	24	.457
	23	.686

	e	p
m = n = 5	40	.040
	39	.079
	38	.159
	37	.278
	36	.437
m = n = 6	57	.013
	56	.026
	55	.052
	54	.091
	53	.159
	52	.234
	51	.338
m = n = 7	77	.004
	76	.008
	75	.016
	74	.029
	73	.049
	72	.078
	71	.118
	70	.167
	69	.233
m = n = 8	100	.001
	99	.003
	98	.005
	97	.009

	e	p
m = n = 8	96	.015
	95	.024
	94	.037
	93	.055
	92	.078
	91	.108
	90	.147
m = n = 9	126	.0004
	125	.0007
	124	.002
	123	.003
	122	.004
	121	.007
	120	.011
	119	.017
	118	.024
	117	.034
	116	.048
	115	.064
	114	.086
	113	.112
	112	.143
m = n =10	155	.0001
	154	.0002
	153	.0004
	152	.0008

	$\underline{e}$	$\underline{p}$
$\underline{m = n = 10}$	151	.001
	150	.002
	149	.003
	148	.005
	147	.007
	146	.010
	145	.015
	144	.020
	143	.027
	142	.036
	141	.047
	140	.061
	139	.078
	138	.098
	137	.122

TABLE A2-2 Tail probabilities for $\tilde{\xi}_N$.

$p = \text{Prob}(\tilde{\xi}_N \geq e)$.

	$\underline{e}$	$\underline{p}$
$\underline{m = n = 4}$	26	.114
	25	.343
	24	.572
$\underline{m = n = 5}$	40	.040
	39	.119
	38	.278
	37	.357
	36	.595

	e	p
m = n = 6	57	.013
	56	.039
	55	.091
	54	.143
	53	.221
m = n = 7	77	.004
	76	.012
	75	.029
	74	.053
	73	.078
	72	.118
m = n = 8	100	.001
	99	.004
	98	.009
	97	.016
	96	.026
	95	.039
	94	.061
	93	.093
	92	.127
m = n = 9	126	.0004
	125	.001
	124	.003
	123	.005
	122	.009

	e	p
m = n = 9	121	.012
	120	.019
	119	.029
	118	.043
	117	.059
	116	.077
	115	.103
m = n = 10	155	.0001
	154	.0003
	153	.0008
	152	.001
	151	.003
	150	.004
	149	.006
	148	.009
	147	.013
	146	.019
	145	.026
	144	.034
	143	.046
	142	.061
	141	.082
	140	.102

TABLE A2-3 Probabilities for $\tilde{\xi}_{N1}$ when m = n = 10.

VALUE OF STATISTIC	PROBABILITY
110	0.0000
111	0.0001
112	0.0005
113	0.0014
114	0.0031
115	0.0059
116	0.0096
117	0.0143
118	0.0198
119	0.0259
120	0.0324
121	0.0399
122	0.0451
123	0.0507
124	0.0550
125	0.0580
126	0.0596
127	0.0598
128	0.0595
129	0.0563
130	0.0530
131	0.0494
132	0.0450
133	0.0407
134	0.0362
135	0.0319
136	0.0275
137	0.0237
138	0.0199
139	0.0168
140	0.0137
141	0.0113
142	0.0089
143	0.0071
144	0.0054
145	0.0042
146	0.0031
147	0.0024
148	0.0016
149	0.0012
150	0.0008
151	0.0005
152	0.0003
153	0.0002
154	0.0001
155	0.0001

TABLE A2-4 Probabilities for $\tilde{\xi}_N$ when m = n = 10.

VALUE OF STATISTIC	PROBABILITY
110	0.0000
111	0.0
112	0.0
113	0.0
114	0.0
115	0.0027
116	0.0045
117	0.0026
118	0.0010
119	0.0002
120	0.0210
121	0.0398
122	0.0337
123	0.0245
124	0.0144
125	0.0490
126	0.0825
127	0.0691
128	0.0529
129	0.0766
130	0.0572
131	0.0796
132	0.0639
133	0.0494
134	0.0359
135	0.0415
136	0.0481
137	0.0383
138	0.0290
139	0.0212
140	0.0204
141	0.0208
142	0.0155
143	0.0117
144	0.008[illegible]
145	0.0058
146	0.0061
147	0.0043
148	0.0028
149	0.0015
150	0.0013
151	0.0011
152	0.0006
153	0.0004
154	0.0002
155	.0001

One final remark needs to be made about $\tilde{\xi}_N$ to the effect that Batschelet has mentioned the question of its null distribution in his book 'Statistical Methods for the *Analysis of Problems* in Animal Orientation and Certain Biological Rhythms' (see Mardia (1972 p.204)).

APPENDIX 3

The Exact Slopes of Quadratic Rank Statistics

Define a quadratic rank statistic as one of the form

$$(A3.1) \qquad T_N = \sum_{i=1}^{N} \sum_{j=1}^{N} a_N(R_i/N+1,\ i/N+1,\ R_j/N+1,\ j/N+1),$$

where $a_N(u(1),\ldots u(4))$ is a function defined for $0 \le u(i) \le 1$, $i = 1,\ldots,4$ and R_i is the rank of X_i when the sample $X_1,\ldots,X_N$ is ordered. A result about the large deviations of such statistics can be obtained through a straightforward adaptation of the techniques used by Woodworth (1970). This result will now be outlined.

If $a = a(u(1),\ldots,u(4))$ and $b = b(u(1),\ldots,u(4))$ are functions defined for $0 \le u(i) \le 1$, $i = 1,\ldots,4$ and $H = \{h(x,y) : h(x,y) \ge 0, \int_0^1 h(x,y)dx = 1 = \int_0^1 h(x,y)dy\}$, then define the pseudometric d by

$$d(a,b) = \sup_{h \in H}\{\left|\iiiint (a(u(1),\ldots,u(4))-b(u(1),\ldots,u(4))h(u(1),u(2)) h(u(3),u(4))du(1)\ldots du(4)\right|\},$$

where the range of integration is understood to be (0,1) in each case.

Suppose that a_N satisfy the following two requirements:

(a) for each N, a_N is constant over the sets $\{i(k) - 1 \le Nu(k) < i(k),\ 1 \le i(k) \le N,\ k = 1,\ldots,4\}$;

(b) there exists a function $a(u(1),\ldots,u(4))$ such that $d(a_N,a) \to 0$ as $N \to \infty$.

<u>Proposition A3-1</u>. Under conditions (a) and (b),

$$I(T,x) = \lim_{N\to\infty}(N^{-1}\log P(T_N \geq N^2 x_N)),$$

where $\{x_N\}$ approaches x,

$$= \inf\{\iint h \log h : \int\ldots\int ah(u(1),u(2))h(u(3),u(4))\, du(1)\ldots du(4) \geq x,\ h \in H\}$$

integration being over the range $(0,1)$.

The proof of Proposition A3-1 simply involves making the appropriate straightforward adjustments to the contents of pp.251-257 of Woodworth's article. For example, we begin by considering the case where a_N is a step function over a fixed grid $C_{ij} \times C_{k\ell}$ where C_{ij} is the rectangle: $C_{ij} = \{(u,v) : u_{i-1} \leq u < u_i,\ v_{j-1} \leq v \leq v_j\}$ and $0 = u_o < u_1 < \ldots < u_\ell = 1$, $0 = v_o < v_1 < \ldots < v_k = 1$ are constants. Then $a_N(u(1),\ldots u(4))$ is defined by $a_N(u(1),\ldots,u(4)) = a_{ijk\ell}$ for $(u(1),\ldots,u(4)) \in C_{ij} \times C_{k\ell}$, for all N. If $X_{ij}^{(N)}$ is defined as the number of integers α such that $(R_\alpha/N+1,\alpha/N+1) \in C_{ij}$ $(i = 1,\ldots,\ell,$ $j = 1,\ldots,k)$, then

$$T_N = \sum\sum\sum\sum a_{ijk\ell}\, X_{ij}^{(N)}\, X_{k\ell}^{(N)}. \qquad \text{(A3.2)}$$

With T_N written in this form the arguments on pp.252-254 only require alteration in regard to what Woodworth calls the constraint.

Further details are omitted. I have not checked every detail of the adaptation of Woodworth's arguments rigorously, but I cannot see that any unmanageable problems should arise here. Proposition A3-1 is sufficient to cover all the two-sample quadratic forms in the rank vector introduced by Schach (see section 1.4).

The almost sure limits of $N^{-2}T_N$ for the two-sample quadratic statistics are easy to obtain (see, for instance, the arguments in Schach (1969b)). In this connection notice that from Proposition 3-4-1 it follows that

$$(A3.3) \qquad N^{-2}\eta_N \to \lambda^2 \int_0^1 \{\int_0^1 \phi(u-t)\bar{f}(u)du\}^2 dt$$

with probability one since $N^{-2}\eta_N = \int_0^1 (\hat{S}_N(t))^2 dt$ (η_N defined as the two-sample version of (3.3.26)). Using Fubini's theorem this becomes

$$N^{-2}\eta_N \to \lambda^2 \int_0^{2\pi} \int_0^{2\pi} \int_0^1 \phi(u-v+t)\phi(t)dt\ d\bar{F}(u)d\bar{F}(v)$$

with probability one.

Further discussion of these points lies beyond the scope of this thesis. The purpose of this brief account was merely to draw the reader's attention to the fact that in principle there are no difficulties to an evaluation of the exact slopes of quadratic forms in the two-sample rank vector. Schach (1969b) has discussed the approximate slopes of the locally most powerful invariant rank tests defined by (3.3.34).

APPENDIX 4

A NOTE ON TERMINOLOGY

The terms limiting distribution and asymptotic distribution have both been used frequently during the course of the thesis. When the asymptotic distribution of T_N is referred to, then we mean that T_N converges in distribution to this asymptotic distribution, while if referring to the limiting distribution of T_N some *rescaling is required* and then the rescaled version of T_N converges in distribution to the limiting distribution.