Sankhyā : The Indian Journal of Statistics
1981, Volume 43, Series A, Pt. 1, pp. 52–66.

AN APPLICATION OF LINEARIZATION IN NONPARAMETRIC MULTIVARIATE ANALYSIS

By J. M. BUHRMAN and F. H. RUYMGAART

Mathematisch Centrum

and

Katholieke Universiteit Nijmegen, The Netherlands

SUMMARY. Attention is restricted to two-dimensional random vectors. The underlying bivariate random structure can be investigated by means of statistics based on the linearized sample elements; the linear compounds are parametrized by the angle that the compounding unit vector makes with the positive first coordinate axis. The collection of statistics forms a stochastic process, certain functionals of which will be relevant statistics for the problem in its original bivariate setting. In this paper we focus on an application to the nonparametric bivariate regression problem, which is very similar to the union-intersection method employed in Roy (1953) and Roy and Bose (1953) in a parametric setting. In the present case the statistics are simple linear rank statistics based on the ranks of the linearized sample elements. The asymptotic distribution theory is developed under the null hypothesis and is to a large extent almost immediate from results in Hájek and Šidák (1967).

INTRODUCTION

For each $N \in \mathcal{N}$ we are given independent two-dimensional random vectors $X_1 = X_{1N} = (\xi_{1N}, \eta_{1N})$, $X_2 = X_{2N} = (\xi_{2N}, \eta_{2N})$, ..., $X_N = X_{NN} = (\xi_{NN}, \eta_{NN})$ with bivariate distribution functions (d.f.'s) $F_1 = F_{1N}$, $F_2 = F_{2N}, \ldots, F_N = F_{NN}$. All random elements to be mentioned in this paper are supposed to be defined on one single probability space $(\Omega, \mathcal{A}, P)$. The σ-field $\mathcal{A}$ is assumed to be complete with respect to the measure P. For each $t \in [0, \pi)$ let $e_t \in \mathcal{R}^2$ be the unit vector making the angle t with the positive ξ-axis ($e_0 = (1, 0)$ and $e_{\frac{1}{2}\pi} = (0, 1)$). Given elements a, $b \in \mathcal{R}^2$, the inner product is denoted by $< a, b >$ and the norm by $\|a\|$.

For each $t \in [0, \pi)$ let us introduce the N independent (univariate) random variables (r.v.'s)

$$X_{n,t} = < X_n, e_t >, \; n = 1, 2, \ldots, N. \qquad \ldots \; (1.1)$$

It is the purpose of the present paper to illustrate the possibility of investigating the bivariate random structure by means of suitably chosen univariate statistics $S_N(t)$, say, based on what will be called the *linearized sample* elements in (1.1). These statistics form a stochastic process

$$S_N = \{S_N(t), \; t \in [0, \pi)\}, \qquad \ldots \; (1.2)$$

certain functionals of which can be used to reach decisions concerning the original bivariate sample. This approach, leading to the so called union-intersection principle for tests of multivariate hypotheses, has been exploited in Roy (1953) and Roy and Bose (1953). Virtually the same idea is used in ordinary principal component analysis, where $S_N(t)$ is the sample variance of the projections corresponding to t. A certain linearization is e.g. also considered in Pyke (1975, Section 4) for the study of bivariate empirical processes. Here we have chosen (0, 0) as the centre of our bundle of lines. For certain applications it is possible that a centre different from (0, 0), may be even a random centre like e.g. $\bar{X} = N^{-1} \sum_{n=1}^{N} X_n$, is appropriate.

The special problem that we have in mind is that of testing the bivariate hypothesis of randomness

$$F_1 = F_2 = \ldots = F_N = F \quad \forall N \in \mathcal{N}, \qquad \ldots (1.3)$$

where F is a *continuous* bivariate d.f. that will occasionally also have to satisfy some further technical conditions (see Sections 3 and 4). Various (asymptotically) distribution free tests for this hypothesis can be found in Puri and Sen (1971); in addition to the extensive list of references in this monograph we may draw attention to the more recent papers by Mardia (1969, 1970), Bhattacharyya and Johnson (1970), and Friedman and Rafsky (1979). To the knowledge of the authors, however, none of the tests to be found in the literature is based on a process of the type (1.2). In fact the process approach leads to a class of statistics containing some of the statistics in Puri and Sen (1971) as a special case. For a further discussion see Section 3.

In order to proceed with a precise description of the process (1.2) to be used for the present problem we need to introduce some more notation and assumptions. The rank $R_{n,t}$ of $X_{n,t}$ among the r.v.'s in (1.1) is defined in the usual way by

$$R_{n,t} = \#\{m : X_{m,t} \leqslant X_{n,t}\}. \qquad \ldots (1.4)$$

Let $J : (0, 1) \to \mathcal{R}$ be square integrable with respect to Lebesgue measure and suppose that the scores $a_N(n)$ $(n = 1, 2, \ldots, N)$ satisfy

$$\int_0^1 (a_N(1+[uN]) - J(u))^2 du \to 0, \quad as \ N \to \infty. \qquad \ldots (1.5)$$

where (for $a \in \mathcal{R}$) $[a]$ denotes the greatest integer not exceeding a. Without loss of generality we shall assume that

$$\int_0^1 J(u)du = 0, \quad \int_0^1 J^2(u)du = \sigma^2 \in (0, \infty). \quad \ldots \quad (1.6)$$

The regression constants $c_n = c_{nN}$ $(n = 1, 2, \ldots, N)$ satisfy

$$\sum_{n=1}^{N} c_n = 0, \quad \sum_{n=1}^{N} c_n^2 = 1, \quad \textit{and} \quad \max_{n=1,2,\cdots N} c_n^2 \to 0, \textit{ as } N \to \infty. \quad \ldots \quad (1.7)$$

For each $t \in [0, \pi)$ we define $S_N(t)$ to be the rank statistic

$$S_N(t) = \sum_{n=1}^{N} c_n a_N(R_{n,t}), \quad \ldots \quad (1.8)$$

which is well known for testing against regression in univariate samples.

To describe the sample paths of the process S_N, let us draw a line through (0,0) perpendicular to the vector $X_m - X_n$ for each pair of indices (m, n) with $m < n$. Each of those lines, $\binom{N}{2}$ in number, makes an angle in $[0, \pi)$ with the positive ξ-axis. Let us denote the ordered angles by

$$0 = \tau_{(0)} \leqslant \tau_{(1)} \leqslant \tau_{(2)} \leqslant \ldots \leqslant \tau_{(M)} < \pi, \ M = \binom{N}{2}. \quad \ldots \quad (1.9)$$

The sample paths of S_N are obviously step functions that are constant on the intervals $[0, \tau_{(1)}), (\tau_{(1)}, \tau_{(2)}), \ldots, (\tau_{(M)}, \pi)$. Suppose that the pair of indices $(m(i), n(i))$ corresponds to the angle $\tau_{(i)}$. This entails that $X_{m(i),t}$ and $X_{n(i),t}$ are tied for $t = \tau_{(i)}$. We also have $|R_{m(i),t} - R_{n(i),t}| = 1$ for t approaching $\tau_{(i)}$ from below as well as from above. When t passes $\tau_{(i)}$ the only change in the vector of ranks is the transposition of $R_{m(i),t}$ and $R_{n(i),t}$, which in general will result in a jump of S_N at $t = \tau_{(i)}$. The value of S_N at this particular point is irrelevant. *It will be convenient to replace $S_N(\tau_{(i)})$ by the right hand limit $S_N(\tau_{(i)}+)$, so that the sample paths become continuous from the right.*

Let μ be a bounded measure on $[0, \pi)$ $(0 \leqslant \mu([0, \pi)) < \infty)$. Then we can write $\mu = \mu_1 + \mu_2$, where μ_1 is absolutely continuous and μ_2 singular with respect to Lebesgue measure. *We shall assume that the singular part is a discrete measure concentrating its total mass on a countable subset* $\{t_1, t_2, \ldots\} \subset [0, \pi)$. We obtain important special cases when we take for μ the Lebesgue measure on $[0, \pi)$ or a discrete measure concentrating its total mass on the finite number of points $t_1 < t_2 < \dots < t_m$ in $[0, \pi)$ such that $\mu(\{t_i\}) = 1$, $i = 1, 2, \dots, m$. The latter measure will be called the counting measure on $\{t_1, t_2, \dots, t_m\}$. Let $L_2([0, \pi), \mu)$ be the separable Hilbert space of Borel-measurable functions that are square integrable with respect to the measure μ. The inner product of $\phi, \psi \in L_2([0, \pi), \mu)$ is defined in the usual way by

$$< \phi, \psi >_\mu = \int_{[0, \pi)} \phi(t)\psi(t)d\mu(t), \qquad \dots \quad (1.10)$$

and the derived norm by $\|\phi\|_\mu$. If μ is the counting measure on $\{t_1, t_2, \dots, t_m\}$ each $\phi \in L_2([0, \pi), \mu)$ can of course be identified with the vector

$$(\phi(t_1), \phi(t_2), \dots, \phi(t_m)) \in \mathcal{R}^m.$$

It is clear that we may consider the process S_N as a random element in $L_2([0, \pi), \mu)$. This has the advantage that weak convergence is relatively easy to establish but a drawback is that the class of continuous functionals is rather restricted; in particular the supremum cannot be considered. For weak convergence of the S_N see Section 2. Due to the continuity of $\|\cdot\|_\mu^2$ we can consider the statistics

$$\|S_N\|_\mu^2 = \int_{[0, \pi)} S_N^2(t)d\mu(t), \qquad \dots \quad (1.11)$$

which seem rather natural for testing the hypothesis (1.3) that is to be rejected for large values. It turns out, however, that the statistics in (1.11) are not even asymptotically distribution free. This deficiency can be remedied if we take as our statistic the square of the norm of a suitable random transformation of the process S_N. For details and certain invariance properties of the statistics we refer to Section 3. *The asymptotic theory is restricted to the null hypothesis.*

The random transformation alluded to in the preceding paragraph is derived from a consistent estimator of a certain covariance function. In Section 4 we prove the existence of such consistent estimators. Section 5 is devoted to a discussion of a possible extension to the case of contiguous alternatives. Also a variation of the results, related to the study of the processes in the complete separable metric space $D([0, \pi))$ (of bounded right continuous functions having only discontinuities of the first kind) endowed with the Skorokhod metric is briefly considered.

2. Weak convergence of the processes

Under the null hypothesis the r.v.'s $X_{n,t}$ have a common continuous univariate d.f. that will be denoted by F_t. Let us observe that because of (1.6) we have

$$E(J(F_t(X_{n,t})) = 0 \quad \forall\, t \in [0, \pi), \qquad \ldots \quad (2.1)$$

$$\Sigma_F(s, t) = \operatorname{cov}(J(F_s(X_{n,s})),\ J(F_t(X_{n,t}))$$
$$= E(J(F_s(X_{n,s}))J(F_t(X_{n,t}))) \;\forall\, s, t \in [0, \pi). \qquad \ldots \quad (2.2)$$

These quantities are indeed well defined due to the square integrability of J and independent of n. The mean value function in (2.1) is apparently independent of F; the covariance function in (2.2) does in general depend on F, although its "diagonal elements" $\Sigma_F(t, t)$ are again independent of F. The function Σ_F is continuous and bounded by σ^2 on $[0, \pi)^2$.

The covariance function Σ_F will be identified with the covariance operator on $L_2([0, \pi), \mu)$, denoted by the same symbol and defined in the usual way by

$$\Sigma_F(\phi) = \int_{[0,\pi)} \Sigma_F(\cdot, t)\phi(t)d\mu(t),\ \phi \in L_2([0, \pi), \mu). \qquad \ldots \quad (2.3)$$

The operator is symmetric, semi-definite positive and nuclear. It follows that the eigenvalues can be written as a sequence

$$\lambda_{F,1} \geqslant \lambda_{F,2} \geqslant \ldots \downarrow 0,\ \textit{with} \sum_{k=1}^{\infty} \lambda_{F,k} = \text{trace } (\Sigma_F)$$
$$= \int_{[0,\pi)} \Sigma_F(t, t)d\mu(t) = \sigma^2\mu([0, \pi)) \in (0, \infty), \qquad \ldots \quad (2.4)$$

see (1.6). *By convention the number of times an eigenvalue is repeated in this sequence is equal to its multiplicity which is finite for each non zero eigenvalue.* By $\phi_{F,1}, \phi_{F,2}, \ldots,$ we denote a corresponding complete orthonormal sequence of eigenfunctions. Let $Z_1, Z_2, \ldots$ be i.i.d. standard normal r.v.'s. By G_F we shall understand the Gaussian random element in $L_2([0, \pi), \mu)$ defined by

$$G_F = \sum_{k=1}^{\infty} \lambda_{F,k}^{1/2} Z_k \phi_{F,k}. \qquad \ldots \quad (2.5)$$

This Gaussian process has mean value function 0 and covariance function Σ_F. For the material in this paragraph see e.g. Grenander (1963, Chapter 6). The law of a random element will be denoted by $\mathcal{L}(\cdot)$ and weak convergence by $\underset{w}{\rightarrow}$.

In the special case where μ is the counting measure on $\{t_1, t_2, \ldots, t_m\}$ the operator Σ_F in (2.3) can be obviously identified with the $m \times m$-matrix $\{\Sigma_F(t_i, t_j)\}$. In this case $\mathcal{L}(G_F)$ virtually reduces to an m-variate normal distribution with 0 mean value vector and covariance matrix $\{\Sigma_F(t_i, t_j)\}$.

Theorem 2.1 : *Let* (1.3) *and* (1.5)–(1.7) *be satisfied. Then we have*

$$\mathcal{L}(S_N) \underset{w}{\rightarrow} \mathcal{L}(G_F), \text{ as } N \to \infty, \text{ on } L_2([0, \pi), \mu), \qquad \ldots \quad (2.6)$$

where S_N *is given by* (1.2) *and* (1.8), *and* G_F *is the Gaussian random element in* (2.5).

Proof : Throughout this proof we shall liberally borrow from Hájek and Šidák (1967, Sections V.1.1–V.1.6). Let us write

$$S_N(t) = \widetilde{S}_N(t) + \rho_N(t), \; t \in [0, \pi), \qquad \ldots \quad (2.7)$$

where

$$S_N(t) = \sum_{n=1}^{N} c_n J(F_t(X_{n,t})), \quad \rho_N(t) = S_N(t) - \widetilde{S}_N(t). \qquad \ldots \quad (2.8)$$

It is clear that the process $\widetilde{S}_N$ is also a random element in $L_2([0, \pi), \mu)$. The assertion (2.6) is immediate from

$$\mathcal{L}(\widetilde{S}_N) \underset{w}{\rightarrow} \mathcal{L}(G_F), \quad \text{as } N \to \infty, \qquad \ldots \quad (2.9)$$

$$E(\|\rho\|_\mu^2) \to 0, \qquad \text{as } N \to \infty. \qquad \ldots \quad (2.10)$$

To prove (2.9) we shall first show that, for arbirtary $\phi \in L_2([0, \pi), \mu)$,

$$\mathcal{L}(< \widetilde{S}_N, \phi >_\mu) \underset{w}{\rightarrow} \mathcal{L}(< G_F, \phi >_\mu), \text{ as } N \to \infty. \qquad \ldots \quad (2.11)$$

Let us note that

$$< \widetilde{S}_N, \phi >_\mu = \sum_{n=1}^{N} c_n Y_n, \text{ with } Y_n = \int_{[0, \pi)} J(F_t(X_{n,t}))\phi(t) d\mu(t).$$

It is immediate that the Y_n are i.i.d. and, from (1.6) and Fubini's theorem, that

$$E(Y_n) = 0, \; \text{var}(Y_n) = < \Sigma_F(\phi), \phi >_\mu.$$

Since the c_n satisfy the Noether-condition (see (1.7)) it follows that $< \widetilde{S}_N, \phi >_\mu$ converges in distribution to a normal law with mean 0 and variance

$$\left(\sum_{n=1}^{N} c_n^2\right) < \mathfrak{T}_F(\phi), \phi >_\mu = < \mathfrak{T}_F(\phi), \phi >_\mu.$$

Because $< G_F, \phi >_\mu$ has the same normal distribution, assertion (2.11) follows.

It is a matter of straightforward computation to see that for any $\phi, \psi \in L_2([0, \pi), \mu)$, we have

$$\begin{aligned} \operatorname{cov}(< \widetilde{S}_N, \phi >_\mu, < \widetilde{S}_N, \psi >_\mu) &= E(< \widetilde{S}_N, \phi >_\mu < \widetilde{S}_N, \psi >_\mu) \\ &= < \mathfrak{T}_F(\phi), \psi >_\mu, \end{aligned}$$

with $\mathfrak{T}_F$ as in (2.2). Using the complete orthonormal sequence $\phi_{F,1}, \phi_{F,2}, \ldots$ defined below (2.4) we have

$$E(\| \sum_{i=k}^{\infty} < \widetilde{S}_N, \phi_{F,i} >_\mu \phi_{F,i} \|^2) = \sum_{i=k}^{\infty} \lambda_{F,i} \downarrow 0, \ as \ k \to \infty, \qquad \ldots \quad (2.12)$$

because of (2.4). This settles the tightness of the sequence $\{\widetilde{S}_N\}$, see Grenander (1963, Section 6.2). Combining (2.11) and (2.12) we arrive at (2.9).

For the proof of (2.10) let us observe that

$$E(\rho_N^2(t)) = \epsilon_N, \ \textit{independent of } t \in [0, \pi), \textit{ and } \epsilon_N \to 0, \ as \ N \to \infty. \qquad \ldots \quad (2.13)$$

It follows at once from (2.13) and Fubini's theorem that

$$E(\|\rho_N\|_\mu^2) = \mu([0, \pi))\epsilon_N \to 0, \ as \ N \to \infty,$$

which proves (2.10).

3. Weak convergence of the statistics

As an immediate corollary to Theorem 2.1 we obtain the limiting distribution of the statistics in (1.11).

Theorem 3.1 : *Under the assumptions of Theorem 2.1 we have*

$$\mathcal{L}(\|S_N\|_\mu^2) \underset{w}{\to} \mathcal{L}\left(\sum_{k=1}^{\infty} \lambda_{F,k} Z_k^2\right), \ as \ N \to \infty. \qquad \ldots \quad (3.1)$$

Proof: Theorem 2.1 entails that

$$\mathcal{L}(\|S_N\|_\mu^2) \underset{w}{\rightarrow} \mathcal{L}(\|G_F\|_\mu^2)$$

$$= \mathcal{L}\left(\left\| \sum_{k=1}^{\infty} \lambda_{F,k}^{1/2} Z_k \, \phi_{F,k} \right\|_\mu^2 \right) = \mathcal{L}\left(\sum_{k=1}^{\infty} \lambda_{F,k} Z_k^2 \right), \; as \; N \to \infty,$$

by the orthonormality of the $\phi_{F,k}$.

We see that the limiting distribution of $\|S_N\|_\mu^2$ depends on the eigenvalues $\lambda_{F,1}, \lambda_{F,2}, \dots$ that in turn depend on F, so that the statistics are not even asymptotically distribution free. By estimating a reasonable number of eigenvalues we can get an insight into this limiting distribution. Let us note in particular that

$$\lim_{N \to \infty} E(\|S_N\|_\mu^2) = \sum_{k=1}^{\infty} \lambda_{F,k} = \sigma^2 \mu([0, \pi)), \qquad \dots \quad (3.2)$$

which is independent of the underlying d.f. F, see (2.4). Further examination can be patterned on investigations of the limiting d.f.'s of Cramér-von Mises type statistics, which are similar to the one appearing on the right in (3.1); for a survey see e.g. Durbin (1971).

If we also estimate some eigenfunctions of Σ_F we may arrive at a sequence of statistics, based on random transforms of S_N, that is asymptotically distribution free. In order to facilitate the discussion we shall impose the technical condition on F that

> *the eigenvalues of* Σ_F *satisfy* $\lambda_{F,1} > \lambda_{F,2} > \dots > \lambda_{F,r} > 0$, *for some* $r \in \mathcal{N}$, *implying that the multiplicities of these eigenvalues equal* 1, $\dots$ (3.3)

according to the convention below (2.4). Let us also assume that there exists a sequence of random covariance functions $\{\hat{\Sigma}_N\}$ on $[0, \pi)^2$ such that

$$P\left(\left\{ \lim_{N \to \infty} \iint_{[0,\pi)^2} |\hat{\Sigma}_N(s,t) - \Sigma_F(s,t)|^2 \, d\mu(s) \, d\mu(t) = 0 \right\}\right) = 1. \quad \dots \quad (3.4)$$

Let $\hat{\lambda}_{N,1} \geqslant \hat{\lambda}_{N,2} \geqslant \dots \downarrow 0$ be the eigenvalues of $\hat{\Sigma}_N$, $N \in \mathcal{N}$. Conditions (3.3) and (3.4) entail that

$$P\left(\left\{ \lim_{N \to \infty} |\hat{\lambda}_{N,k} - \lambda_{F,k}| = 0 \right\}\right) = 1, \; i = 1, 2, \dots, r, \qquad \dots \quad (3.5)$$

and that for each N a complete sequence of orthonormal eigenfunctions $\hat{\phi}_{N,1}, \hat{\phi}_{N,2}, \ldots$ of $\hat{\Sigma}_N$ can be selected in such a way that

$$P\left(\left\{\lim_{N\to\infty} \|\hat{\phi}_{N,k}-\phi_{F,k}\|_\mu = 0\right\}\right) = 1,\ i = 1, 2, \ldots, r. \qquad \ldots \quad (3.6)$$

By χ^2_r we denote the chi-squared distribution with r degrees of freedom.

Theorem 3.2 : *Let condition (3.3) and (3.4) be satisfied in addition to the assumptions of Theorem 2.1 Then we have*

$$\mathcal{L}\left(\sum_{k=1}^{r} \hat{\lambda}_{N,k}^{-1} < S_N, \hat{\phi}_{N,k} >^2_\mu\right) \underset{w}{\rightarrow} \chi^2_r,\ as\ N \rightarrow \infty. \qquad \ldots \quad (3.7)$$

Proof : It follows from (3.5) and (3.6) that for N sufficiently large

$$\left|\sum_{k=1}^{r} \hat{\lambda}_{N,k}^{-1} < S_N, \hat{\phi}_{N,k} >^2_\mu - \sum_{k=1}^{r} \lambda_{F,k}^{-1} < S_N, \phi_{F,k} >^2_\mu\right|$$

$$\leqslant \sum_{k=1}^{r} |\hat{\lambda}_{N,k}^{-1} - \lambda_{F,k}^{-1}| < S_N, \hat{\phi}_{N,k} >^2_\mu$$

$$+ \sum_{k=1}^{r} \lambda_{F,k}^{-1} | < S_N, \hat{\phi}_{N,k} >^2_\mu - < S_N, \phi_{F,k} >^2_\mu|$$

$$\leqslant \|S_N\|^2_\mu \Big[\sum_{k=1}^{r} |\hat{\lambda}_{N,k}^{-1} - \lambda_{F,k}^{-1}|$$

$$+ 2 \sum_{k=1}^{r} \|\hat{\phi}_{N,k} - \phi_{F,k}\|_\mu \Big] = O_P(1),\ as\ N \rightarrow \infty. \qquad \ldots \quad (3.8)$$

because $\|S_N\|^2_\mu = O_P(1)$, as $N \rightarrow \infty$.

Furthermore we have by Theorem 2.1 (as $N \rightarrow \infty$)

$$\mathcal{L}\left(\sum_{k=1}^{r} \lambda_{F,k}^{-1} < S_N, \phi_{F,k} >^2\right) \underset{w}{\rightarrow} \mathcal{L}\left(\sum_{k=1}^{r} \lambda_{F,k}^{-1} < G_F, \phi_{F,k} >^2\right)$$

$$= \mathcal{L}\left(\sum_{k=1}^{r} Z_k^2\right) = \chi^2_r. \qquad \ldots \quad (3.9)$$

The theorem follows at once from (3.8) and (3.9).

By way of an example let us first consider the special case where μ is the counting measuring on $t_1 < t_2 < \ldots < t_m$. Condition (3.4) is now equivalent to

$$P\left(\left\{ \lim_{N \to \infty} \hat{\Sigma}_N(t_i, t_j) = \Sigma_F(t_i, t_j) \right\}\right) = 1 \;\; \forall\; i, j \in \{1, 2, \ldots, m\}, \quad \ldots \quad (3.10)$$

and the assertion of Theorem 3.2 reduces to

$$\mathcal{L}\left(\sum_{k=1}^{r} \lambda_{N;k}^{-1} \left[\sum_{i=1}^{m} S_N(t_i)\hat{\phi}_{N,k}(t_i) \right]^2 \right) \underset{w}{\to} \chi_r^2, \; as \; N \to \infty. \quad \ldots \quad (3.11)$$

The statistics are invariant under all translations but not in general under rotations. If the t are equidistant (in the sense, that $t_2 - t_1 = t_3 - t_2 = \ldots = t_m - t_{m-1} = \pi - t_m + t_1$), however, the tests are invariant under the finite group of rotations over angles that are multiples of π/m. In the present bivariate case some of the tests in Puri and Sen (1971) are based on statistics as in (3.11) with $t_1 = 0$ and $t_2 = \frac{1}{2}\pi$.

Let us next choose μ to be equal to Lebesgue measeure on $[0, \pi)$. Then condition (3.4) becomes

$$P\left(\left\{ \lim_{N \to \infty} \int_0^\pi \int_0^\pi | \hat{\Sigma}_N(s, t) - \Sigma_F(s, t)|^2 \, ds dt = 0 \right\}\right) = 1. \quad \ldots \quad (3.12)$$

The assertion of the theorem can now be written as

$$\mathcal{L}\left(\sum_{k=1}^{r} \lambda_{N;k}^{-1} \left[\int_0^\pi S_N(t)\hat{\phi}_{N,k}(t) dt \right]^2 \right) \underset{w}{\to} \chi_r^2, \; as \; N \to \infty. \quad \ldots \quad (3.13)$$

These statistics are invariant under translations as well as rotations; see, however, Section 5.

4. Estimation of the covariance function

For each $t \in [0, \pi)$ let $\hat{F}_{N,t}$ be the univariate empirical d.f. of the $X_{n,t}$ in (1.1), and for each $s, t \in [0, \pi)$ let $\hat{F}_{N,s,t}$ be the bivariate empirical d.f. of the $(X_{n,s}, X_{n,t})$. The bivariate empirical d.f. of the original random vectors X_n is denoted by $\hat{F}_N$. Throughout this section we let

$$\hat{\sigma}_N(s, t) = \iint_{\mathcal{R}^2} J\left(\frac{N}{N+1} \hat{F}_{N,s}(u)\right) J\left(\frac{N}{N+1} \hat{F}_{N,t}(v)\right) d\hat{F}_{N,s,t}(u, v), \quad \ldots \quad (4.1)$$

for $(s, t) \in [0, \pi)^2$. Observe that $\hat{F}_{N,s,t}(u, v)$ is equal to the fraction of vectors X_n in the closed convex set bounded by the line through ue_s perpendicular to e_s and the line through ve_t perpendicular to e_t.

Under the assumption that the underlying d.f.

$$F \textit{ has a density with respect to Lebesgue measure on } \mathcal{R}^2, \qquad \ldots \quad (4.2)$$

and that the score function

$$J \textit{ is continuous on } (0, 1), \textit{ with } |J(s)| \leqslant c[s(1-s)]^{-\alpha},$$

$$s \in (0, 1), \textit{ for some } c \in (0, \infty) \textit{ and } \alpha \in (0, \tfrac{1}{2}), \qquad \ldots \quad (4.3)$$

it fouows from Groeneboom, LePage and Ruymgaart (1976) that in the present notation

$$P\left(\left\{ \lim_{N \to \infty} \hat{\sigma}_N(s, t) = \Sigma_F(s, t) \right\}\right) = 1 \;\; \forall\, (s, t) \in [0, \pi)^2. \qquad \ldots \quad (4.4)$$

This result is the starting point for the construction of a sequence of estimators $\hat{\Sigma}_N$ that satisfy (3.4).

For this construction let $k(N) \in \mathcal{N}$ be such that $k(N) \uparrow \infty$, as $N \to \infty$, and let $\mathcal{P}_N = \{0 = t_{N,1}, t_{N,2}, \ldots, t_{N,k(N)} = \pi\}$, $t_{N,i} < t_{N,i+1}$, be partitions of $[0, \pi)$ with $\mathcal{P}_N \subset \mathcal{P}_{N+1}$ and $\bigcup \mathcal{P}_N = \mathcal{P}$ dense in $[0, \pi)$. It is convenient to write

$$R(N, i, j) = [t_{N,i}, t_{N,i+1}) \times [t_{N,j}, t_{N,j+1}).$$

We shall define

$$\hat{\Sigma}_N(s, t) = \sum_{i=1}^{k(N)-1} \sum_{j=1}^{k(N)-1} 1_{R(N,i,j)}(s, t) \hat{\sigma}_N(t_{N,i}, t_{N,j}), \quad \textit{for } (s, t) \in [0, \pi)^2. \qquad \ldots \quad (4.5)$$

Theorem 4.1 : *Let conditions* (4.2) *and* (4.3) *be satisfied. Then the* $\hat{\Sigma}_N$ *in* (4.5) *have the property*

$$P\left(\left\{ \lim_{N \to \infty} \hat{\Sigma}_N = \Sigma_F \textit{ on } [0, \pi)^2 \right\}\right) = 1; \qquad \ldots \quad (4.6)$$

they satisfy, moreover, condition (3.4).

Proof: Let us first introduce the sets

$$\Omega_{s,t} = \{\textit{the samples } \{X_{n,s}\} \textit{ and } \{X_{n,t}\} \textit{ are both united}$$

$$\textit{for all } N, \textit{ and } \hat{\Sigma}_N(s, t) \to \Sigma_F(s, t), \textit{ as } N \to \infty\}.$$

Then we have

$$P(\Omega_0) = 1, \; where \; \Omega_0 = \bigcap_{s,t\in \mathcal{P}} \Omega_{s,t}, \qquad \dots \quad (3.7)$$

in particular because of (4.4) and the countability of $\mathcal{P}$.

To prove (4.6) we shall first compare $\hat{\Sigma}_N$ with

$$\Sigma_{F,N}(s,t) = \sum_{i=1}^{k(N)-1} \sum_{j=1}^{k(N)-1} 1_{R(N,i,j)}(s,t)\, \Sigma_F(t_{N,i}, t_{N,j}), \; for \; (s,t) \in [0,\pi)^2. \quad \dots \quad (4.8)$$

Because Σ_F is continuous and bounded on $[0,\pi)^2$ it is uniformly continuous, so that

$$\sup_{s,t\in[0,\pi)} | \; \Sigma_{F,N}(s,t) - \Sigma_F(s,t) | \to 0, \; as \; N \to \infty. \qquad \dots \quad (4.9)$$

Hence it suffices to prove that

$$\lim_{N\to\infty} | \; \hat{\Sigma}_N(s,t) - \Sigma_{F,N}(s,t) | = 0 \; \forall \; (s,t) \in [0,\pi)^2, \quad \dots \quad (4.10)$$

$$and \; for \; any \; \omega \in \Omega_0,$$

in view of (4.7). This is, however, immediate because of the properties of the set Ω_0 and since (4.10) need only be satisfied for $s, t \in \mathcal{P}$, due to the way in which $\Sigma_{F,N}$ has been defined.

To prove that (3.4) is satisfied let us note that by the Schwarz inequality

$$\sup_{s,t\in[0,\pi)} | \; \hat{\Sigma}_N(s,t) | = \sup_{s,t\in \mathcal{P}} | \; \hat{\Sigma}_N(s,t) |$$

$$\leqslant \sup_{s\in \mathcal{P}} | \; \hat{\Sigma}_N(s,s) | = N^{-1} \sum_{n=1}^{N} J^2(n/(N+1)) \; on \; \Omega_0. \qquad \dots \quad (4.11)$$

It is in fact only in the last equality that the restriction to Ω_0 plays a role; the equality holds because $\hat{\Sigma}_N(s,s)$ is independent of $s \in \mathcal{P}$ for such $\omega \in \Omega_0$. It follows that on Ω_0 all the functions $|\hat{\Sigma}_N|$ are bounded by a finite constant (J is square integrable, see (4.3)), and (3.4) follows from (4.6) and an application of the Lebesgue dominated convergence theorem.

5. Discussion

Throughout the first part of this section let us focus on the choice $\mu = \lambda$, where λ denotes Lebesgue measure on $[0, \pi)$. Let us also restrict ourselves to a discussion of the statistics $\|S_N\|_\lambda^2$.

Although we have seen in Section 3 that these statistics have certain invariance properties, they are easily seen to be not invariant under unequal scale changes of the coordinates. Moreover, if a large value of S_N^2 is assumed on an interval $(\tau_{(i-1)}, \tau_{(i)})$ of relatively short length, such large values may not have much effect on the statistics $\|S_N\|_\lambda^2$. This phenomenon might reduce the power of the derived tests under certain alternatives.

These deficiencies can be remedied if we consider the processes T_N, defined by

$$T_N(s) = S_N \circ \phi_N(s), \quad s \in [0, \pi), \qquad \dots \quad (5.1)$$

where $\phi_N : [0, \pi) \to [0, \pi)$ is the random time change, given by

$$\phi_N(s) = \tau_{(i-1)}, \; for \;\; s \in [(i-1)\pi/M, \; i\pi/M), \qquad \dots \quad (5.2)$$

and $i = 1, 2, \dots, M$. These processes are easily seen to be also invariant under unequal scale changes of the coordinates, and the various values are assumed on intervals of equal length. Hence in some cases more powerful tests could be probably derived from the statistics

$$\|T_N\|_\lambda^2 = \int_0^\pi T_N^2(s)ds, \qquad \dots \quad (5.3)$$

and some of its modifications.

Although it is very likely that the ϕ_N will converge in the supremum norm to a continuous function ϕ_F (directly related to the inverse of the d.f. of the random variable $\frac{1}{2}\pi + \arctan(\eta_1 - \eta_2)/(\xi_1 - \xi_2))$), this is not sufficient to derive weak convergence of T_N on $L_2([0, \pi), \lambda)$ from that of S_N (see Billingsley (1968, Section 17). The reason is that the topology in $L_2([0, \pi), \lambda)$ is too coarse.

Due to the properties of S_N (see also the convention below (1.9)) we may as well consider S_N as a random element in the complete separable metric space $D([0, \pi))$ of bounded right continuous functions having discontinuities of the first kind only, endowed with the Skorokhod metric. We conjecture that weak convergence of S_N in $D([0, \pi))$ would entail that of T_N in $D([0, \pi))$. Weak convergence of the finite dimensional distributions is contained in (2.11) by choosing for μ the appropriate counting measure. The tightness, however, constitutes a problem since the usual condition on the moments of $|S_N(s)-S_N(t)|$ does not seem to work. An indication that tightness may not even be satisfied can be found in the circumstance that the number of jumps of the sample paths is of order N^2, whereas the height of these jumps is in general of the usual order $N^{-\frac{1}{2}}$.

It is not the purpose of this paper to go into any efficiency considerations. Nevertheless, the feeling seems to be justified that by taking into account a larger number of directions (and not only e.g. 0 and $\frac{1}{2}\pi$) the power will be improved for a rather considerable class of alternatives. Asymptotic normality of $\|S_N\|_\mu^2$ for arbitrary F_n (including fixed as well as local alternatives) follows, under more stringent conditions on J, from results in Ruymgaart and van Zuijlen (1978) in the case where μ is an arbitrary counting measure.

ACKNOWLEDGEMENTS

The authors wish to thank Professor E. V. Khmaladze for some useful comments leading to the inclusion of the asymptotically distribution free statistics in (3.7). They are also indebted to Professor P. K. Sen for pointing out the references Roy (1953) and Roy and Bose (1953).

REFERENCES

BILLINGSLEY, P. (1968): *Weak Convergence of Probability Measures*, Wiley, New York.

BHATTACHARYYA, G. K. and JOHNSON, R. A. (1970): A layer rank test for ordered bivariate alternatives. *Ann. Math. Statist.*, **41**, 1296-1310.

DURBIN, J. (1973): Distribution theory for tests based on the sample distribution function. *Reg. Conf. Ser. Appl. Math.*, 9, SIAM, Philadelphia.

FRIEDMAN, J. H. and RAFSKY, L. C. (1979): Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. *Ann. Statist.*, **7**, 697-717.

GRENANDER, U. (1963): *Probabilities on Algebraic Structures*, Almqvist & Wiksell-Wiley, Stockholm-New York.

GROENEBOOM, P., LEPAGE, Y. and RUYMGAART, F. H. (1976): Rank tests for independence with best strong Bahadur slope, *Z. Wahrscheinlich-keitstheorie Verw. Gebiete*, **36**, 119-127.

HÁJEK, J. and ŠIDÁK, Z. (1967): *Theory of Rank Tests*, Acad. Press, New York.

MARDIA, K. V. (1967): A non-parametric tests for the bivariate two-sample location problem, *J. R. Statist. Soc., B*, **29**, 320-342.

PURI, M. L. and SEN, P. K. (1971): *Nonparametric Methods in Multivariate Analysis*, Wiley, New York.

PYKE, R. (1975): Multidimensional empirical processes: some comments. *Statist. Inf. Rel. Topics* (M. L. Puri editor) *Vol.* 2, 45-58, Ac. Press, New York.

ROY, S. N. (1953): On a heuristic method of test construction and its use in multivariate analysis, *Ann. Math. Statist.*, **24**, 220-238.

ROY, S. N. and BOSE, R. C. (1953): Simultaneous confidence interval estimation. *Ann. Math. Statist.*, **24**, 513-536.

RUYMGAART, F. H. and ZUIJLEN, M. C. A. VAN (1978): Asymptotic normality of multivariate linear rank statistics in the non-i.i.d. case. *Ann. Statist.*, **6**, 588-602.

Paper received: December, 1979.