JOURNAL OF MULTIVARIATE ANALYSIS 11, 485–497 (1981)

A Robust Principal Component Analysis

F. H. RUYMGAART

Katholieke Universiteit, Toernooiveld, The Netherlands

Communicated by L. K. Schmetterer

A robust principal component analysis for samples from a bivariate distribution function is described. The method is based on robust estimators for dispersion in the univariate case along with a certain linearization of the bivariate structure. Besides the continuity of the functional defining the direction of the suitably modified principal axis, we prove consistency of the corresponding sequence of estimators. Asymptotic normality is established under some additional conditions.

1. INTRODUCTION AND FIRST RESULTS

Classical principal component analysis is based on the sample covariance matrix as an estimator of the covariance matrix of the unknown multivariate normal distribution from which the sample is taken. It is obvious that this method is not robust: not only will the estimated direction of the principal axis be rather sensitive to outliers but the very definition of the principal axis in terms of the population covariance matrix will not even make sense if the (mixed) moments of second order do not exist. The purpose of this paper is to describe a robust approach to principal component analysis. The main tools are robust estimators for dispersion in the univariate case along with a certain linearization of the multivariate structure. As a by-product we arrive at a class of stochastic processes that might be of independent interest. Attention will be restricted to the bivariate case and the emphasis is on asymptotic properties such as (strong) consistency and asymptotic normality of the estimated direction of the—properly modified—principal axis.

More specifically we consider a sequence $X_1 = (X_{11}, X_{12})$, $X_2 = (X_{21}, X_{22})$,... of independent and identically distributed (i.i.d.) bivariate random vectors, defined on a certain probability space $(\Omega, \mathscr{A}, P)$. For $N \in \mathbb{N}$ the first N of these vectors constitute a random sample of size N from their

Received June 20, 1980.

AMS 1970 subject classifications: Primary 62H25; Secondary 62G35.

Key words and phrases: Principal component, robustness, robust estimator for dispersion, linearization.

0047-259X/81/040485-13$02.00/0

common bivariate distribution function (d.f.) F on $\mathbb{R}^2$. The bivariate empirical d.f. based on this sample is, as usual, defined by

$$\hat{F}_N(x_1, x_2) = N^{-1} \sum_{n=1}^{N} 1_{(-\infty, x_1] \times (-\infty, x_2]}(X_{n1}, X_{n2}), \qquad (x_1, x_2) \in \mathbb{R}^2. \tag{1.1}$$

Let us introduce

$$\mathscr{F}_i = \{G\colon G \text{ is an } i\text{-variate d.f. on } \mathbb{R}^i\}, \qquad i \in \{1, 2\}. \tag{1.2}$$

For any $G \in \mathscr{F}_i$ and Borel set B in $\mathbb{R}^i$ we write

$$G\{B\} = \int_B dG. \tag{1.3}$$

For $t \in [0, 2\pi)$ let $e_t \in \mathbb{R}^2$ be the unit vector making an angle t with the x_1^+-axis, so that, e.g., $e_0 = (1, 0)$ and $e_{(1/2)\pi} = (0, 1)$. The inner product of $a, b \in \mathbb{R}^2$ is written as $\langle a, b\rangle$. With any bivariate d.f. $G \in \mathscr{F}_2$ we associate the collection of linearized d.f.'s $\{G_{(t)}, t \in [0, 2\pi)\}$, where

$$G_{(t)}(x) = G\{y \in \mathbb{R}^2\colon \langle y, e_t\rangle \leqslant x\}, \qquad x \in \mathbb{R}. \tag{1.4}$$

In particular we introduce, for any $t \in [0, 2\pi)$, the univariate random variables (r.v.'s)

$$X_{n,t} = \langle X_n, e_t\rangle, \qquad n \in \{1, 2,..., N\}. \tag{1.5}$$

These r.v.'s are a random sample from the univariate d.f. $F_{(t)}$, see (1.4), since it is obvious that

$$F_{(t)}(x) = P(\langle X_n, e_t\rangle \leqslant x), \qquad x \in \mathbb{R}. \tag{1.6}$$

This random sample will be referred to as the linearized sample (corresponding to the angle t). It is also obvious that the empirical d.f. of this linearized sample coincides with $\hat{F}_{N,(t)}$, i.e., that

$$\hat{F}_{N,(t)}(x) = N^{-1} \sum_{n=1}^{N} 1_{(-\infty, x]}(X_{n,t}), \qquad x \in \mathbb{R}. \tag{1.7}$$

Because of the relation $X_{n,t+\pi} = -X_{n,t}$, $t \in [0, \pi)$ one can essentially restrict attention to $t \in [0, \pi)$; it will be convenient, however, to use the closed interval $[0, \pi]$.

In order to establish robustness of certain functionals we first need to introduce a metric in $\mathscr{F}_i$. For $i \in \{1, 2\}$ we shall use the metric

$$\rho_i(G_1, G_2) = \sup_{\mathbb{H}} |G_1\{\mathbb{H}\} - G_2\{\mathbb{H}\}|, \qquad G_j \in \mathscr{F}_i, \tag{1.8}$$

for $j \in \{1, 2\}$, where the supremum is taken over all closed halfspaces $\mathbb{H} \subset \mathbb{R}^{i}$. It should be noted that we have

$$\rho_1(G_1, G_2) = \sup_{x \in \mathbb{R}} |G_1(x) - G_2(x)|, \qquad G_j \in \mathscr{F}_1, \tag{1.9}$$

and the relation

$$\rho_2(G_1, G_2) = \sup_{t \in [0, \pi)} \rho_1(G_{1,(t)}, G_{2,(t)}), \qquad G_j \in \mathscr{F}_2. \tag{1.10}$$

Classical principal component analysis is based on the non-robust functional

$$\int \left(x - \int x \, dF_{(t)}(x)\right)^2 dF_{(t)}(x), \tag{1.11}$$

where F is bivariate normal. The direction of the principal axis is the angle p, where this last expression assumes its maximum, if unique. A straightforward robustification is obtained by replacing the functional in (1.11) by a robust functional for dispersion. We therefore introduce a functional

$$\theta\colon \mathscr{F}_1 \to \mathbb{R}, \ \theta \textit{ robust, i.e., continuous w.r.t. the metric } \rho_1; \tag{1.12}$$

see Hampel [7] and Huber [8]. Application of this functional to the elements (in a subset) of the linearized d.f.'s associated with the underlying bivariate d.f. F yields the function

$$t \mapsto \theta(F_{(t)}), \qquad t \in [0, \pi]. \tag{1.13}$$

DEFINITION 1.1. The *direction of the modified principal axis* is the functional $\tau\colon \mathscr{F}_2 \to [0, \pi)$, defined by

$$\tau(G) = \inf\{s \in [0, \pi]\colon \sup_{t \in [0, s]} \theta(G_{(t)}) = \sup_{t \in [0, \pi]} \theta(G_{(t)})\}. \tag{1.14}$$

For any pair $G_1, G_2 \in \mathscr{F}_2$ we define, moreover,

$$\rho(\tau(G_1), \tau(G_2)) = |\tau(G_1) - \tau(G_2)| \wedge (\pi - |\tau(G_1) - \tau(G_2)|). \tag{1.15}$$

We shall assume that

$$\exists p \in [0, \pi)\colon \theta(F_{(p)}) > \sup_{t \notin I(p;\varepsilon)} \theta(F_{(t)}) \qquad \forall \varepsilon > 0, \tag{1.16}$$

where

$$\begin{aligned} I(p;\varepsilon) &= (p - \varepsilon, p + \varepsilon) \cap [0, \pi], \qquad & p \in (0, \pi), \\ &= [0, \varepsilon) \cup (\pi - \varepsilon, \pi], \qquad & p = 0. \end{aligned} \tag{1.17}$$

It is obvious that under this condition $\tau(F) = p$, where F is the underlying bivariate d.f.

Let $\{G_k\} \subset \mathscr{F}_2$ be an arbitrary sequence such that $\rho_2(G_k, F) \to 0$, as $k \to \infty$, where F is the underlying bivariate d.f. The robustness of θ (see (1.12)) and relation (1.10) entails that

$$\theta(G_{k,(t)}) \to \theta(F_{(t)}) \quad \forall t \in [0, \pi], \qquad \textit{as } k \to \infty. \tag{1.18}$$

For the present problem, however, we shall use the stronger property that

$$\sup_{t \in [0,\pi]} |\theta(G_{k,(t)}) - \theta(F_{(t)})| \to 0, \qquad \textit{as } k \to \infty. \tag{1.19}$$

THEOREM 1.1. *Suppose that the underlying bivariate d.f. F and the functional θ are such that* (1.16) *is satisfied and that* (1.19) *holds true for any $\{G_k\} \subset \mathscr{F}_2$ with $\rho_2(G_k, F) \to 0$, as $k \to \infty$. Then the functional τ in* (1.14) *is robust at F (w.r.t. the metrics ρ_2 in* (1.8) *and ρ in* (1.15)) *and*

$$P(\lim_{N \to \infty} \tau(\hat{F}_N) = \tau(F)) = 1, \tag{1.20}$$

i.e., $\{\tau(\hat{F}_N)\}$ is a strongly consistent sequence of estimators for the direction of the modified principal axis.

Proof. Choose an arbitrary $\varepsilon > 0$. Let us introduce the number

$$\delta_1(\varepsilon) = \theta(F_{(p)}) - \sup_{t \notin I(p;\varepsilon)} \theta(F_{(t)}). \tag{1.21}$$

In view of condition (1.19) there exists $\delta_2(\varepsilon) > 0$ such that

$$\sup_{t \in [0,\pi]} |\theta(G_{(t)}) - \theta(F_{(t)})| < \tfrac{1}{3}\delta_1(\varepsilon), \qquad \textit{as } \rho_2(G, F) < \delta_2(\varepsilon). \tag{1.22}$$

We shall prove that

$$\rho(\tau(G), \tau(F)) < \varepsilon, \qquad \textit{as } \rho_2(G, F) < \delta_2(\varepsilon), \tag{1.23}$$

which entails the robustness of τ at F.

It follows from (1.21) and (1.22) that

$$\sup_{t \notin I(p;\varepsilon)} \theta(G_{(t)}) < \theta(F_{(p)}) - \tfrac{2}{3}\delta_1(\varepsilon). \tag{1.24}$$

On the other hand, (1.22) entails

$$\theta(G_{(p)}) > \theta(F_{(p)}) - \tfrac{1}{3}\delta_1(\varepsilon). \tag{1.25}$$

Combination of (1.24), (1.25) and the definition of τ in (1.14) yields (1.23).

According to the version of the Glivenko–Cantelli theorem due to Wolfowitz [13, 14] we have

$$P(\lim_{N\to\infty} \rho_2(\hat{F}_N, F) = 0) = 1; \tag{1.26}$$

see also (1.8)–(1.10). The strong consistency in (1.20) is immediate from (1.26) and the robustness. ∎

Let us briefly comment on the assumptions (1.16) and (1.19). It will turn out (see Section 2) that assumption (1.19) is automatically fulfilled for an important class of functionals. Moreover, if $\theta(F_{(\cdot)})$ is continuous on $[0, \pi]$ with a unique maximum, (1.16) is fulfilled. It should be noted, however, that the function $\theta(F_{(\cdot)})$ need not be smooth. In particular if F concentrates mass 1 on a finite set of points, this function need not even be continuous. This lack of smoothness presents serious difficulties if one wishes to show the asymptotic normality of $\tau(\hat{F}_N)$. Therefore the remainder of this paper is devoted to the study of smoothened versions $\tilde{\theta}(F_{(\cdot)})$ and $\tilde{\theta}(\hat{F}_{N,(\cdot)})$ of $\theta(F_{(\cdot)})$ respectively $\theta(\hat{F}_{N,(\cdot)})$. Specifications and basic properties of the stochastic processes

$$\{N^{1/2}(\tilde{\theta}(\hat{F}_{N,(t)}) - \tilde{\theta}(F_{(t)})), t \in [0, \pi]\} \tag{1.27}$$

are given in Section 2. Processes of this type may have some independent interest. Robustness, strong consistency and asymptotic normality of the relevant quantities based on $\tilde{\tau}$, the corresponding modification of τ, are established in Section 3.

We conclude this section with some references. Linearization has been used in Roy [10] and Roy and Bose [11], where it leads to the so called union-intersection principle for testing certain multivariate hypotheses. A certain linearization is, e.g., also considered in Pyke [9, Sect. 4] in the study of bivariate empirical processes. An application in nonparametric multivariate analysis is given in Buhrman and Ruymgaart [4]. Finally classical principal component analysis itself is based on linearization as has already been observed. As an aside it should be observed that linearization plays a natural role in quantum mechanics. Given a pair of non-commuting (and consequently not simultaneously measurable) observables like, e.g., position and momentum one may consider the set of all standardized linear combinations of these observables. Each such linear combination has a univariate d.f. (depending on the state of the system); thus we are given a family $\{F_{(t)}, t \in [0, 2\pi)\}$ of univariate d.f.'s. According to a theorem due to Nelson (see Theorem 2.1 in Gudder [6]) there does not exist a bivariate d.f. the collection of linearized d.f.'s of which coincides with the aforementioned family.

2. A Smooth Version

First we shall specify the functional θ in (1.12). Let J: $(0, 1) \to \mathbb{R}$ satisfy

$$J(s) = 0 \quad \forall s \notin [\gamma, 1-\gamma] \qquad \textit{for some} \quad \gamma \in (0, \tfrac{1}{2}). \tag{2.1}$$

To obtain a suitable estimator for dispersion (see, e.g., David [5] and Bickel and Lehmann [1]) it furthermore is assumed that

$$J(\tfrac{1}{2} - s) = -J(\tfrac{1}{2} + s) < 0 \qquad \forall s \in (0, \tfrac{1}{2}). \tag{2.2}$$

To guarantee robustness of θ in a strong sense (see Lemma 2.1) we impose the condition that J has a derivative $J^{(1)}$ such that

$$J^{(1)} \textit{ is continuous and } \exists c \in (0, \infty): |J^{(1)}| \leqslant c \textit{ on } (0, 1). \tag{2.3}$$

We define as usual

$$G^{-1}(s) = \inf\{x: G(x) \geqslant s\}, \qquad s \in (0, 1) \textit{ for } G \in \mathscr{F}_1. \tag{2.4}$$

It is well known that the functional

$$\theta(G) = \int_0^1 G^{-1}(s)\, J(s)\, ds, \qquad G \in \mathscr{F}_1, \tag{2.5}$$

is robust on $\mathscr{F}_1$ (see, e.g., Huber [8] and Boos [3]). In Boos [3] it was demonstrated that this functional even has a Fréchet derivative on $\mathscr{F}_1$; in the course of the proof the relation

$$\begin{aligned}\theta(G_2) - \theta(G_1) &- \int_{\mathbb{R}} (G_2(x) - G_1(x))\, J(G_1(x))\, dx \\ &= -\int_{\mathbb{R}} [\bar{J}(G_2(x)) - \bar{J}(G_1(x)) - (G_2(x) - G_1(x))\, J(G_1(x))]\, dx \end{aligned} \tag{2.6}$$

was established, where

$$\bar{J}(u) = \int_0^u J(s)\, ds, \qquad s \in (0, 1).$$

Under the present conditions on J that are stronger than those in Boos [3], one can even show that the functional satisfies a uniform Lipschitz condition.

LEMMA 2.1. *Under the assumptions* (2.1) *and* (2.3) *we have*

$$\left|\theta(G_2)-\theta(G_1)-\int_{\mathbb{R}}(G_1(x)-G_2(x))\,J(G_1(x))\,dx\right|\leqslant\frac{1}{2}c[\rho_1(G_1,G_2)]^2,\quad (2.7)$$

for any pair $G_1, G_2 \in \mathscr{F}_1$. *The number* c *appearing on the right in* (2.7) *is the same as the number in* (2.3) *and is consequently independent of* G_1 *and* G_2.

Proof. The lemma is immediate from (2.6) by observing that

$$\begin{aligned}\tilde{J}(G_2(x))-\tilde{J}(G_1(x))&=(G_2(x)-G_1(x))\,J(G_1(x))\\&\quad+\tfrac{1}{2}(G_2(x)-G_1(x))^2\,J^{(1)}(G'(x)),\qquad x\in\mathbb{R},\end{aligned}$$

where $G'(x)$ is a point strictly between $G_1(x)$ and $G_2(x)$. ∎

LEMMA 2.2. *Let* γ *be as in* (2.1). *For each* $G\in\mathscr{F}_2$ *there exists a constant* $r(G)\in(0,\infty)$ *such that*

$$-r(G)\leqslant G_{(t)}^{-1}(\tfrac{1}{2}\gamma)\leqslant G_{(t)}^{-1}(1-\tfrac{1}{2}\gamma)\leqslant r(G)\qquad\forall t\in[0,\pi].\qquad(2.8)$$

In particular, under assumptions (2.1) *and* (2.3) *the function* $\theta(G_{(\cdot)})$ *is bounded on* $[0,\pi]$.

Proof. Given $G\in\mathscr{F}_2$ there exists an open sphere B with center 0 and radius $r(G)\in(0,\infty)$ such that

$$G\{B\}=G\{\bar{B}\}>1-\tfrac{1}{2}\gamma,\qquad(2.9)$$

where $\bar{B}$ is the closure of B. Since $G_{(t)}(-r(G))\leqslant G\{B^c\}$ and $G_{(t)}(r(G))\geqslant G\{\bar{B}\}$ relation (2.8) is immediate from (2.9). The last assertion of the lemma is obvious from (2.8) and (2.5). ∎

Let D be a domain in $\mathbb{R}^i$. The collection of the continuous functions on D for which all the (mixed partial) derivatives of order j are continuous on the interior of D is denoted by

$$\mathscr{C}^{(j)}(D);\ \textit{in particular we write}\ \mathscr{C}^{(0)}(D)=\mathscr{C}(D).\qquad(2.10)$$

Let $(\alpha,\beta)\supset[0,\pi]$ and consider a kernel

$$K\colon(\alpha,\beta)\times(\alpha,\beta)\to[0,\infty),\qquad K\in\mathscr{C}^{(3)}((\alpha,\beta)\times(\alpha,\beta)).\qquad(2.11)$$

For any bounded measurable function $\phi: [0, \pi] \to \mathbb{R}$ we define the smoothened version $\tilde{\phi}$ by

$$\tilde{\phi}(t) = \int_0^\pi K(t, s)\,\phi(s)\,ds, \qquad t \in [0, \pi]. \tag{2.12}$$

For brevity let us write

$$\partial^j K(t, s)/(\partial t^j) = K^{(j)}(t, s), \qquad (t, s) \in (\alpha, \beta) \times (\alpha, \beta). \tag{2.13}$$

Let us note that

$$\tilde{\phi} \in \mathscr{C}^{(3)}([0, \pi]); \qquad \tilde{\phi}^{(j)}(\,\cdot\,) = \int_0^\pi K^{(j)}(\cdot, s)\,\phi(s)\,ds, \tag{2.14}$$

for $j \in \{0, 1, 2, 3\}$. In order that $\tilde{\phi}$ remains (in some sense) close to ϕ one typically chooses K zero outside a narow strip along the main diagonal in $[0, \pi] \times [0, \pi]$. On a slightly smaller strip one might choose K constant.

DEFINITION 2.1. Let θ be of the type (2.5) with J satisfying (2.1) and (2.3). Because of Lemma 2.2, for any $G \in \mathscr{F}_2$ we may define a function $t \mapsto \tilde{\theta}(G_{(t)})$ according to

$$\tilde{\theta}(G_{(t)}) = \int_0^\pi K(t, s)\,\theta(G_{(s)})\,ds, \qquad t \in [0, \pi], \tag{2.15}$$

where K satisfies (2.11); see also (2.12). By the *direction of the smoothly modified principal axis* we shall understand the functional $\tilde{\tau}: \mathscr{F}_2 \to [0, \pi]$, obtained from τ in (1.14) by replacing $\theta(G_{(t)})$ by $\tilde{\theta}(G_{(t)})$ as in (2.15).

THEOREM 2.1. *Suppose that $\tilde{\theta}(F_{(\cdot)})$ has a unique maximum on $[0, \pi)$, where F is the underlying bivariate d.f. Then the functional $\tilde{\tau}$ in Definition* 2.1 *is robust at F* (*w.r.t. the metrics ρ_2 in* (1.8) *and ρ in* (1.15)) *and*

$$P(\lim_{N \to \infty} \tilde{\tau}(\hat{F}_N) = \tilde{\tau}(F)) = 1. \tag{2.16}$$

Proof. It follows from the way in which Theorem 1.1 was proved that we need only verify that conditions (1.16) and (1.19) are satisfied with θ replaced by $\tilde{\theta}$. Since by (2.14) the function $\tilde{\theta}(F_{(\cdot)})$ is continuous on $[0, \pi]$ the uniqueness of the maximum in $[0, \pi)$ trivially entails (1.16). As far as (1.19) is concerned it is immediate from the properties of K that (1.19) in its original form (i.e., with θ instead of $\tilde{\theta}$) entails its validity for $\tilde{\theta}$. Therefore we shall restrict ourselves to the proof of (1.19) for θ.

Let us choose an arbitrary sequence $\{G_k\} \subset \mathscr{F}_2$ such that $\rho_2(G_k, F) \to 0$, as $k \to \infty$. We see from (2.7) that

$$\sup_{t \in [0,\pi]} |\theta(G_{k,(t)}) - \theta(F_{(t)})|$$

$$\leqslant \rho_2(G_k, F) \sup_{t \in [0,\pi]} \int_{\mathbb{R}} |J(F_{(t)}(x))|\, dx + \frac{1}{2} c[\rho_2(G_k, F)]^2. \quad (2.17)$$

We used relation (1.10) to obtain this upper bound. It follows from (2.8) (applied with $G = F$) that

$$\sup_{t \in [0,\pi]} \int_{\mathbb{R}} |J(F_{(t)}(x))|\, dx \leqslant 2r(F) \max_{s \in (0,1)} |J(s)| < \infty. \quad (2.18)$$

The assertion (1.19) follows at once from (2.17) and (2.18). ∎

3. Asymptotic Normality

Motivated by (2.6) let us introduce the decomposition

$$N^{1/2}(\theta(\hat{F}_{N,(t)}) - \theta(F_{(t)})) = N^{-1/2} \sum_{n=1}^{N} \Delta_n(t) + R_N(t), \qquad t \in [0, \pi], \quad (3.1)$$

where

$$\Delta_n(t) = \int_{\mathbb{R}} [F_{(t)}(x) - 1_{(-\infty,x]}(X_{n,t})]\, J(F_{(t)}(x))\, dx, \qquad t \in [0, \pi], \quad (3.2)$$

and R_N is the remainder.

It is clear that the Δ_n are i.i.d. random elements in some function space, so that one might expect the process $N^{-1/2} \sum_{n=1}^{N} \Delta_n(t)$, $t \in [0, \pi]$, to converge weakly (note that the expectation function is identically equal to zero). It is remarkable, however, to find that some of the well-known moment conditions for tightness seem hard to verify, even in the case where F has a (smooth) density with respect to Lebesgue measure in $\mathbb{R}^2$.

For this as well as for other reasons we return to the smooth versions. It is convenient to write

$$\theta(\hat{F}_{N,(t)}) = T_N(t), \qquad \theta(F_{(t)}) = \mu(t), \qquad t \in [0, \pi]. \quad (3.3)$$

In the notation of (2.12) the smooth versions can be written as $\tilde{T}_N$ and $\tilde{\mu}$. According to (2.14) these (random) functions are elements of $\mathscr{C}^{(3)}([0, \pi])$, so that the processes $N^{1/2}(\tilde{T}_N^{(j)}(t) - \tilde{\mu}^{(j)}(t))$, $t \in [0, \pi]$, are random elements in $\mathscr{C}([0, \pi])$ for $j = 0, 1, 2$ (and even $j = 3$ but we do not need this). Following (3.1) these processes may be decomposed as

$$\begin{aligned} &N^{1/2}(\tilde{T}_N^{(j)}(t) - \tilde{\mu}^{(j)}(t)) \\ &\quad = N^{-1/2} \sum_{n=1}^{N} \tilde{\Delta}_n^{(j)}(t) + \tilde{R}_N^{(j)}(t), \qquad t \in [0, \pi], \end{aligned} \tag{3.4}$$

for $i = 0, 1, 2$, where the $\tilde{\Delta}_n^{(j)}$ and $\tilde{R}_N^{(j)}$ are random elements in $\mathscr{C}([0, \pi])$ as well.

Any of the random elements $\tilde{\Delta}_n^{(j)}$ has a well defined covariance function $\sum_j$ on $[0, \pi] \times [0, \pi]$. To be more explicit let us introduce the joint d.f.

$$F_{(s,t)}(x, y) = P(X_{n,s} \leqslant x, X_{n,t} \leqslant y), \qquad (x, y) \in \mathbb{R}^2, \tag{3.5}$$

and $(s, t) \in [0, \pi] \times [0, \pi]$. We have

$$E\tilde{\Delta}_n^{(j)}(t) = 0 \qquad \forall t \in [0, \pi], \quad j \in \{0, 1, 2\}, \tag{3.6}$$

$$\begin{aligned} \sum_j &= E\tilde{\Delta}_n^{(j)}(s)\, \tilde{\Delta}_n^{(j)}(t) \\ &= \int_0^{\pi} \int_0^{\pi} K^{(j)}(s, u)\, K^{(j)}(t, v) \left[\int_{\mathbb{R}} \int_{\mathbb{R}} (F_{(u,v)}(x, y) \right. \\ &\qquad \left. - F_{(u)}(x)\, F_{(v)}(y))\, J(F_{(u)}(x))\, J(F_{(v)}(y))\, dx\, dy \right] du\, dv \\ &\qquad\qquad \forall (s, t) \in [0, \pi] \times [0, \pi], \quad j \in \{0, 1, 2\}. \end{aligned} \tag{3.7}$$

For $j = 0, 1, 2$ we introduce Gaussian processes $\mathscr{G}_j$ such that

$$E\mathscr{G}_j(t) = 0, \qquad E\mathscr{G}_j(s)\, \mathscr{G}_j(t) = \sum_j (s, t), \tag{3.8}$$

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

A condition on the underlying d.f. F, sufficient for our purpose, is that

$$\textit{F has a bounded density w.r.t. Lebesgue measure on } \mathbb{R}^2. \tag{3.9}$$

With some additional effort (3.9) can be weakened to the assumption that F

is continuous on $\mathbb{R}^2$. We conjecture that we can do even without any condition on F. This is the case if one can show that

$$\exists \delta \in (0, \tfrac{1}{2}) \colon N^{1/2-\delta} \rho_2(\hat{F}_N, F) \xrightarrow[P]{} 0 \qquad (N \to \infty), \tag{3.10}$$

for any d.f. $F \in \mathscr{F}_2$. It is obvious from Stute [12, Theorem 1.1] that (3.9) implies (3.10).

LEMMA 3.1. *Let conditions* (2.1), (2.3), (2.11) *and either* (3.9) *or* (3.10) *be satisfied. Then we have the convergence in distribution*

$$N^{1/2}(\tilde{T}_N^{(j)} - \tilde{\mu}^{(j)}) \xrightarrow[d]{} \mathscr{G}_j \qquad (N \to \infty), \quad j \in \{0, 1, 2\}, \tag{3.11}$$

on the complete separable metric space $\mathscr{C}([0, \pi])$ *endowed with the supremum metric.*

Proof. It will be convenient to introduce the notation

$$c_j = \max_{(s,t) \in [0,\pi] \times [0,\pi]} |K^{(j)}(s, t)|, \qquad j \in \{0, 1, 2, 3\},$$

where $K^{(0)} = K$.

First we shall prove that

$$\max_{t \in [0,\pi]} |\tilde{R}_N^{(j)}(t)| \xrightarrow[P]{} 0 \qquad (N \to \infty), \quad j \in \{0, 1, 2\}. \tag{3.12}$$

It follows from (2.7), (2.12) and (3.1) that

$$\begin{aligned} \max_{t \in [0,\pi]} |\tilde{R}_N^{(j)}(t)| &\leqslant \tfrac{1}{2}\pi c_j c N^{1/2} [\sup_{t \in [0,\pi]} \rho_1(\hat{F}_{N,(t)}, F_{(t)})]^2 \\ &= \tfrac{1}{2}\pi c_j c N^{1/2} [\rho_2(\hat{F}_N, F)]^2, \end{aligned}$$

so that (3.12) follows at once from (3.10).

Relation (3.12) entails that we need only show the convergence in distribution

$$N^{-1/2} \sum_{n=1}^{N} \tilde{\Delta}_n^{(j)} \xrightarrow[d]{} \mathscr{G}_j \qquad (N \to \infty). \tag{3.13}$$

Since the $\tilde{\Delta}_n^{(j)}$ are i.i.d. random elements in $\mathscr{C}([0, \pi])$ the convergence of the finite dimensional distributions is trivially fulfilled. To prove tightness we use Billingsley [2, Theorem 12.3]. The random elements being centered and independent, we find

$$
\begin{aligned}
E\left(N^{-1/2}\sum_{n=1}^{N}(\bar{\Delta}_n^{(j)}(s)-\bar{\Delta}_n^{(j)}(t))\right)^2
&= N^{-1}\sum_{n=1}^{N}E(\bar{\Delta}_n^{(j)}(s)-\bar{\Delta}_n^{(j)}(t))^2 \\
&= E(\bar{\Delta}_1^{(j)}(s)-\bar{\Delta}_1^{(j)}(t))^2 \\
&\leqslant (s-t)^2\, c_{j+1}^2\, E\left(\int_0^{\pi}|\Delta_1(u)|\,du\right)^2 \\
&\leqslant (s-t)^2\, c_{j+1}^2\, \pi 2cr(F), \qquad (3.14)
\end{aligned}
$$

because of (2.8). This settles the tightness. ∎

THEOREM 3.1. *Suppose that $\tilde{\mu}$ has a unique maximum in* $[0,\pi)$; *according to Definition* 2.1 *this maximum is a fortiori assumed at the point* $\tilde{\tau}(F)$. *Let conditions* (2.1), (2.3), (2.11) *and either* (3.9) *or* (3.10) *be satisfied. Then we have*

$$
N^{1/2}(\tilde{\tau}(\hat{F}_N)-\tilde{\tau}(F)) \xrightarrow[d]{} \mathcal{N}(0,\sigma^2) \qquad (N\to\infty), \qquad (3.15)
$$

where $\sigma^2=[\tilde{\mu}^{(2)}(\tilde{\tau}(F))]^{-2}\sum_1(\tilde{\tau}(F),\tilde{\tau}(F))$; *see* (3.7).

Proof. Since $\tilde{T}_N$ is continuous on $[0,\pi]$ it assumes an absolute maximum in $[0,\pi]$; Definition 2.1 entails that $\tilde{\tau}(\hat{F}_N)$ is a point where this maximum is assumed, so that

$$
\begin{aligned}
N^{1/2}\tilde{T}_N^{(1)}(\tilde{\tau}(\hat{F}_N)) = 0 &= N^{1/2}\tilde{T}_N^{(1)}(\tilde{\tau}(F)) \\
&\quad + N^{1/2}(\tilde{\tau}(\hat{F}_N)-\tilde{\tau}(F))\,\tilde{T}_N^{(2)}(t_N), \qquad (3.16)
\end{aligned}
$$

where t_N is a random point strictly between $\tilde{\tau}(F)$ and $\tilde{\tau}(\hat{F}_N)$. Moreover, Lemma 3.1 for $j=2$ entails

$$
\sup_{t\in[0,\pi]}|\tilde{T}_N^{(2)}(t)-\tilde{\mu}^{(2)}(t)| \xrightarrow[P]{} 0 \qquad (N\to\infty). \qquad (3.17)
$$

Because $\tilde{\mu}$ has a maximum at $\tilde{\tau}(F)$ it follows that $\tilde{\mu}^{(2)}(\tilde{\tau}(F))<0$; by continuity of $\tilde{\mu}^{(2)}$ there exists an $\varepsilon>0$ such that $\tilde{\mu}^{(2)}<0$ on the interval $I(p;\varepsilon)$, defined as in (1.16) with $p=\tilde{\tau}(F)$. Let

$$
\Omega_N=\{\omega\in\Omega:\ \inf_{t\in I(p;\varepsilon)}|\tilde{T}_N^{(2)}(t)|>0\}; \qquad (3.18)
$$

it is clear from (3.17) that

$$
P(\Omega_N)\to 1 \qquad (N\to\infty). \qquad (3.19)
$$

Combination of (3.16) and (3.18) yields

$$N^{1/2}(\tilde{\tau}(\hat{F}_N) - \tilde{\tau}(F)) = -\frac{N^{1/2}(\tilde{T}_N^{(1)}(\tilde{\tau}(F)) - \tilde{\mu}^{(1)}(\tilde{\tau}(F)))}{\tilde{T}_N^{(2)}(t_N)}, \qquad on\ \Omega_N. \tag{3.20}$$

The (strong) consistency in Theorem 2.1 and (3.17) imply

$$\tilde{T}_N^{(2)}(t_N) \xrightarrow[P]{} \tilde{\mu}^{(2)}(\tilde{\tau}(F)) \qquad (N \to \infty), \tag{3.21}$$

and the theorem follows by straightforward combination of (3.19)–(3.21) and Lemma 3.1 for $i = 1$. ∎

References

[1] Bickel, P. J. and Lehmann, E. L. (1976). Descriptive statistics for nonparametric models. III. Dispersion. *Ann. Statist.* **4**, 1139–1158.

[2] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[3] Boos, D. D. (1979). A differential for L-statistics. *Ann. Statist.* **7**, 955–959.

[4] Buhrman, J. M. and Ruymgaart, F. H. (1979). An application of linearization in non-parametric multivariate analysis. *Sankhyā Ser. A.*, in press.

[5] David, H. A. (1970). *Order Statistics.* Wiley, New York.

[6] Gudder, S. P. (1979). *Stochastic Methods in Quantum Mechanics.* North-Holland, New York.

[7] Hampel, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42**, 1887–1896.

[8] Huber, P. J. (1972). Robust statistics: A review. *Ann. Math. Statist.* **43**, 1041–1067.

[9] Pyke, R. (1975). Multidimensional empirical processes: Some comments. In *Statistical Inference and Related Topics* (M. L. Puri, Ed.), Vol. 2, pp. 45–58. Academic Press, New York.

[10] Roy, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Ann. Math. Statist.* **24**, 220–238.

[11] Roy, S. N. and Bose, R. C. (1953). Simultaneous confidence interval estimation. *Ann. Math. Statist.* **24**, 513–536.

[12] Stute, W. (1977). Convergence rates for the isotrope discrepancy. *Ann. Probab.* **5**, 707–723.

[13] Wolfowitz, J. (1954). Generalizations of the theorem of Glivenko–Cantelli. *Ann. Math. Statist.* **25**, 131–138.

[14] Wolfowitz, J. (1960). Convergence of the empiric distribution function on half-spaces. In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, Stanford Univ. Press, Stanford.