The Annals of Statistics
1997, Vol. 25, No. 6, 2388–2409

TRANSFORMED EMPIRICAL PROCESSES AND MODIFIED KOLMOGOROV–SMIRNOV TESTS FOR MULTIVARIATE DISTRIBUTIONS[1]

BY A. CABAÑA AND E. M. CABAÑA

Instituto Venezolano de Investigaciones Científicas and Universidad de la República, Uruguay

A general way of constructing classes of goodness-of-fit tests for multivariate samples is presented. These tests are based on a random signed measure that plays the same role as the empirical process in the construction of the classical Kolmogorov–Smirnov tests. The resulting tests are consistent against any fixed alternative, and, for each sequence of contiguous alternatives, a test in each class can be chosen so as to optimize the discrimination of those alternatives.

1. Introduction. The design of tests specially adapted to detect some specific alternatives is a common procedure in nonparametric statistics. All three well-known tests for the two-sample problem, Wilcoxon, Fisher and Yates or median are all three distribution free, but each one is better than the others in detecting changes in position for specific sample distributions: the logistic, the normal or the double exponential, respectively [see Hájek and Sidák (1967), for instance].

As far as we know, the design of goodness-of-fit tests, consistent against all fixed alternatives and specially fitted to detect a specific family of them, has not been as well developed. That problem was considered and solved by one of the authors [Cabaña (1996)], by using a martingale transformation of the empirical process originally used in statistical inference by Khmaladze (1981, 1993), who introduces a goodness-of-fit test but does not focus his attention on the improvement of the behavior under specific alternatives.

In the present article, we pose again the problem of designing consistent goodness-of-fit tests, fitted to an arbitrarily given family of alternatives (such as changes in position for normal samples, changes in dispersion for double exponential ones, etc.). Our results extend the ones in Cabaña (1996) in two senses: (1) goodness-of-fit tests for multivariate samples are discussed, and (2) the basic transformations to be applied to the empirical process for the construction of the tests are obtained by means of L^2 arguments that disregard the martingale property characteristic of the particular transformation introduced in Khmaladze (1981) and used in Cabaña (1996).

Received September 1996; revised June 1997.

[1]Joint work supported by Project 94/007, CONICYT (National Council for Scientific and Technological Research), Montevideo, Uruguay.

AMS 1991 *subject classifications*. 62G10, 62G20, 62G30, 60G15.

Key words and phrases. Goodness-of-fit, power improvement, TEPs.

In fact, that martingale transformation is also an isometry in the L^2 space related to the probability to be fitted, and we call it below a *Laguerre isometry* or *L-isometry* because it is obtained by changing variables in the *Laguerre shift* that carries each Laguerre polynomial into the next one, with degree one unit larger. Our procedure is carried on starting from any arbitrarily given isometry, not necessarily the Laguerre one, although that one is the simplest we can choose, and all the resulting tests have the same distribution free behavior under the null hypothesis of fit and also under the alternatives for which the test is designed. This shows in particular that the martingale property of Laguerre isometry is not at all essential for testing purposes.

2. Preliminaries. The empirical process associated with the sample $\{X\}$ of size 1 with respect to the probability distribution F_0 on $\mathscr{E} = \mathbf{R}^d$, $d \geq 1$, is the signed measure b_x: $A \mapsto b_x(A) = \mathbf{1}_{\{x \in A\}} - F_0(A)$, evaluated at the random point $x = X$.

It is well known that when X has distribution F_0, b_X satisfies the properties: $\mathbf{E} b_X(A) = 0$ for every measurable set A, and $\mathbf{E} b_X(A) b_X(B) = F_0(A \cap B) - F_0(A) F_0(B)$ for every measurable A, B, which means that b_X has the same first- and second-order moments as an F_0-Brownian bridge.

As n goes to infinity, the empirical process

$$b_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} b_{X_i} \tag{1}$$

of the sample $\{X_1, \dots, X_n\}$ of i.i.d. variables with distribution F converges in distribution to an F_0-Brownian bridge b^{F_0} when $F = F_0$. When $F \neq F_0$ instead, b_n behaves asymptotically as the sum of an F-bridge b^F and the deterministic term $\sqrt{n}(F - F_0)$ that tends to infinity in the supremum norm as $n \to \infty$. This gives a well-known justification for the classical Kolmogorov–Smirnov test of $\mathscr{H}_0$: $F = F_0$ [see Donsker (1952)].

Let us introduce a family of alternatives $F^{(\tau)}$ (τ in a neighborhood of 0) contiguous to $F^{(0)} = F_0$, with density $f^{(\tau)}$ with respect to F_0, and such that there exists an $L^2(\mathscr{E}, dF_0)$ function k satisfying

$$\left\| \frac{1}{\tau}\left(\sqrt{f^{(\tau)}} - 1\right) - \frac{k}{2} \right\|_{L^2} \to 0 \quad \text{as } \tau \to 0^+, \ \tau \neq 0. \tag{2}$$

When (2) holds, the function k necessarily satisfies

$$\int k(x)\, dF_0(x) = 0, \tag{3}$$

and can be obtained as the $L^1(\mathscr{E}, dF_0)$-limit of $(f^{(\tau)} - 1)/\tau$ as $\tau \to 0$.

We shall assume that we are especially interested in detecting the sequence of alternatives $\mathscr{H}_n$: $F = F_{(n)} = F^{(\delta/\sqrt{n})}$, where n is the sample size, and δ is a fixed parameter introduced for further convenience of notation.

Under $\mathscr{H}_n$, $b_n(A)$ converges in distribution to the sum of b^{F_0} plus the deterministic drift

$$\lim_{n\to\infty}\sqrt{n}\,\mathbf{E}b_{X_{\delta/\sqrt{n}}}(A)=\delta\left.\frac{\partial\mathbf{E}b_{X_\tau}(A)}{\partial\tau}\right|_{\tau=0}=\delta\int_A k(x)\,dF_0(x)$$

[c.f. Oosterhoff and van Zwet (1979)]. That drift measures the sensitivity of the test based on b_n with respect to $\mathscr{H}_n$.

The aim of this paper is the construction of goodness-of-fit tests based on signed measures $\hat{w}_x$ $(x\in\mathscr{E})$ such that the following hold.

(a) When x is replaced by a random variable X, the resulting measure $\hat{w}_X(A)$ evaluated on any measurable set A is a random variable.

(b) The measure $\hat{w}_n=n^{-1/2}\sum_{i=1}^n\hat{w}_{X_i}$ associated with the sample has some normalized limit distribution under $\mathscr{H}_0$.

(c) The asymptotic distributions of $\hat{w}_n$ under $\mathscr{H}_0$ and under $\mathscr{H}_n$ differ as much as possible.

These random measures will play, in the construction of our tests, the same role as the empirical processes in classical Kolmogorov–Smirnov tests. Each of them will be called a *transformed empirical process* (in short: TEP).

For technical convenience, we shall require that the second-order moments of $\hat{w}_X$ coincide with those of a Wiener process, normalized by the requirement that its total variance be one. Thus, under a central limit property to be established, the limit distribution of $\hat{w}_n=n^{-1/2}\sum_{i=1}^n w_{X_i}$ will be a normalized Gaussian process with independent increments.

As a consequence, we pose ourselves the problem of finding signed measures $\hat{w}_x$ satisfying the following:

(A) $\mathbf{E}\hat{w}_X(A)=0$, for all measurable sets A and the random variable X distributed as F_0;

(B) $\mathbf{E}\hat{w}_X(A)\hat{w}_X(B)=V(A\cap B)$, for all measurable sets A, B, the random variable X distributed as F_0, and some probability V on $\mathscr{E}$;

(C) $\partial\mathbf{E}\hat{w}_{X_\tau}(\mathscr{E})/\partial\tau|_{\tau=0}\geq\partial\mathbf{E}\hat{w}'_{X_\tau}(\mathscr{E})/\partial\tau|_{\tau=0}$ for any $\hat{w}'_x$ satisfying the analogue of (1) and (2), and the random variable X_τ distributed as $F^{(\tau)}$.

Conditions (A) and (B) do not require further explanation. As for condition (C), notice that it requires the drift of our TEP to be as large as possible on the set $\mathscr{E}$ of maximum variance. The adequacy of such heuristic requirement will be verified a posteriori, from the properties of the resulting tests, since they will have optimum efficiency [see Section 4 and Cabaña (1996)].

Let us point out that the problem of finding processes $\hat{w}_X$ satisfying (A), (B), (C) is implicitly solved in Cabaña (1993, 1996), where goodness-of-fit tests asymptotically efficient are obtained for univariate samples. In this article, we solve it explicitly, extend the solution to a multivariate context, and, for each family of alternatives as in (2), provide consistent multivariate tests, efficient for that family.

In the sequel, $\mathscr{E}$ is always set equal to $\mathbf{R}^d$, and the examples in Section 9 are developed for $d = 1$ and $d = 2$, but most results apply to more general spaces, particularly the extension of isometries described in Section 8.

3. A formal L^2 construction of the TEP.

3.1. *Isometries and associated TEPs.* Let us assume for simplicity that the probability V appearing in condition (B) is absolutely continuous with respect to F_0, and call a^2 its density; that is,

$$V(A) = \int_A a^2(x)\,dF_0(x), \quad \int a^2(x)\,dF_0(x) = 1.$$

We shall assume further that a is F_0-a.e. different from zero; that is, F_0 and V are absolutely continuous with respect to each other.

We shall denote by $\langle \bullet, \bullet \rangle$ and $\langle \bullet, \bullet \rangle_V$ the inner products in $L^2(\mathscr{E}, dF_0)$ and $L^2(\mathscr{E}, dV)$, respectively. From an orthonormal basis $\Psi = (\psi_0 = 1, \psi_1, \psi_2, \dots)$ of $L^2(\mathscr{E}, dF_0)$, we may construct the new sequence of functions

$$\Psi^V = (\psi_i^V)_{i=0,1,\dots}, \qquad \psi_i^V(x) = \psi_i(x)/a(x),$$

which is an orthonormal basis for $L^2(\mathscr{E}, dV)$.

Conditions (A) and (B) can be replaced by the requirement that the Fourier coefficients

$$c_i(x) = \int \psi_i^V(y)\,d\hat{w}_x(y)$$

of $d\hat{w}_x/dV$ with respect to Ψ^V satisfy

$$\mathbf{E}c_i(X) = \langle c_i, 1\rangle = \mathbf{E}\int \psi_i^V(x)\,d\hat{w}_X(x) = 0$$

and

$$\begin{aligned}\mathbf{E}c_i(X)c_j(X) = \langle c_i, c_j\rangle &= \mathbf{E}\int \psi_i^V(y)\,d\hat{w}_X(y)\int \psi_j^V(z)\,d\hat{w}_X(z)\\ &= \int \psi_i^V(y)\psi_j^V(y)\,dV(y) = \langle \psi_i^V, \psi_j^V\rangle_V = \delta_{i,j}.\end{aligned}$$

Consequently, $C = (c_i)_{i=0,1,\dots}$ is required to be an orthonormal system in $L^2(\mathscr{E}, dF_0)$ with all its elements orthogonal to the constant 1, in order that (B) and (C) hold.

Expand now $\hat{w}_{X_\tau}(\mathscr{E}) = \sum_i c_i(X_\tau)\langle \psi_i^V, 1\rangle$, and compute

$$\begin{aligned}\left.\frac{\partial \mathbf{E}\hat{w}_{X_\tau}(\mathscr{E})}{\partial \tau}\right|_{\tau=0} &= \lim_{\tau\to 0}\sum_i \frac{\mathbf{E}(c_i(X_\tau) - c_i(X))}{\tau}\langle \psi_i^V, 1\rangle_V\\ &= \sum_i \int c_i(x)k(x)\,dF_0(x)\langle \psi_i, a\rangle \qquad (4)\\ &= \sum_i \langle c_i, k\rangle\langle \psi_i, a\rangle = \sum_i k_i^C a_i,\end{aligned}$$

where k_i^C, a_i are the ith Fourier coefficients of k and a with respect to the orthonormal systems C and Ψ, respectively.

Let us introduce now the isometry $\mathscr{T}$ that maps Ψ onto C; that is,

$$\mathscr{T}\colon L^2(\mathscr{E}, dF_0) \to L^2(\mathscr{E}, dF_0), \qquad \psi_i \mapsto \mathscr{T}\psi_i = c_i,\ i = 1, 2, \ldots. \tag{5}$$

The last term in (4) equals $\langle k, \mathscr{T}a\rangle$, and therefore condition (C) holds when, for a given $\mathscr{T}$, one chooses a to minimize the angle between $\mathscr{T}a$ and k. This is accomplished by selecting $\mathscr{T}a = k/\|k\|$, provided the span of C contains k. This will be ensured by imposing that the span of C, that is, the range of $\mathscr{T}$, be the orthogonal complement $1^\perp$ of 1 in $L^2(\mathscr{E}, dF_0)$.

The conclusions obtained so far can be summarized in the following statement.

PROPOSITION 1. *The measure $\hat{w}_x$ with formal Fourier expansion*

$$\hat{w}_x(A) = \sum_{i=0}^{\infty} c_i(x)\langle \hat{a}\mathbf{1}_A, \psi_i\rangle \tag{6}$$

satisfies (A), (B), (C), *when $\Psi = (\psi_i)_{i=1,2,\ldots}$ is an orthonormal basis of $L^2(\mathscr{E}, dF_0)$, $\mathscr{T}$ is an isometry on $L^2(\mathscr{E}, dF_0)$ with range $1^\perp$, $C = (c_i)_{i=1,2,\ldots}$ is the image of Ψ by $\mathscr{T}$ and the function $\hat{a}$ has the property $\mathscr{T}\hat{a} = k/\|k\|$.*

In that case, the objective of our optimization has the value

$$\left.\frac{\partial \mathbf{E}\hat{w}_{X_\tau}(\mathscr{E})}{\partial\tau}\right|_{\tau=0} = \langle k, \mathscr{T}\hat{a}\rangle = \|k\|. \tag{7}$$

REMARK 1. The preceding proposition gives a formal solution to our problem for each orthonormal basis of $L^2(\mathscr{E}, dF_0)$ and each isometry $\mathscr{T}$ with range $1^\perp$.

REMARK 2. By replacing $c_i(x)$ by $\mathscr{T}\psi_i(x)$ in (6), and using the linearity of $\mathscr{T}$ we obtain

$$\hat{w}_\bullet(A) = \mathscr{T}\sum_{i=0}^{\infty}\langle \hat{a}\mathbf{1}_A, \psi_i\rangle\psi_i = \mathscr{T}(\hat{a}\mathbf{1}_A), \tag{8}$$

and this implies in particular that $\hat{w}_x$ depends only on the isometry $\mathscr{T}$, but not on the orthonormal basis Ψ.

Let us finally introduce the notation

$$w_\bullet^{(a,\mathscr{T})}(A) = \mathscr{T}(a\mathbf{1}_A) \tag{9}$$

for the measure associated with the isometry $\mathscr{T}$ and the *score function* $a \in L^2(\mathscr{E}, dF_0)$.

After Remark 2, we may reformulate our previous statement in the following.

PROPOSITION 2. *The measure* (9) *satisfies conditions* (A), (B), (C) *and equation* (7), *when* $\mathscr{T}$ *is an isometry on* $L^2(\mathscr{E}, dF_0)$ *with range* $1^\perp$ *containing* k, *and the score function is chosen as* $\hat{a} = \mathscr{T}^{-1}k/\|k\|$.

3.2. *Constructing the TEP as a stochastic integral with respect to the empirical process.* From (9) and the orthogonality of the range of $\mathscr{T}$ with respect to 1, the expression

$$\begin{aligned} w_X^{(a,\mathscr{T})}(A) &= \mathscr{T}(a\mathbf{1}_A)(X) = \int \mathscr{T}(a\mathbf{1}_A)\,d\mathbf{1}_{\{\bullet\le X\}} \\ &= \int \mathscr{T}(a\mathbf{1}_A)\,d(\mathbf{1}_{\bullet\le X} - F_0) = \int \mathscr{T}(a\mathbf{1}_A)\,db_X \end{aligned} \tag{10}$$

follows.

From (10) we derive an expression for

$$w_n^{(a,\mathscr{T})} = \frac{1}{\sqrt{n}}\sum_{i=1}^n w_{X_i}^{(a,\mathscr{T})} \tag{11}$$

in terms of b_n, namely

$$w_n^{(a,\mathscr{T})}(x) = \int \mathscr{T}(a\mathbf{1}_x)\,db_n. \tag{12}$$

3.3. *Transformed empirical process in* $\mathscr{E} = \mathbf{R}^d$. We adopt the expressions (10) and (12) as the definition of the transformed empirical process.

DEFINITION 1. The transformed empirical process of the sample $\{X_1, X_2, \ldots, X_n\}$ of $\mathscr{E}$-valued random variables, associated with the probability distribution F_0 on $\mathscr{E}$, the isometry $\mathscr{T}$ on $L_2 = L_2(\mathscr{E}, dF_0)$ with range equal to the orthogonal complement $1^\perp$ of the constant function 1 and the L_2-function a with $\|a\| = 1$ is

$$w_n^{(a,\mathscr{T})}(A) = \int \mathscr{T}(a\mathbf{1}_A)\,db_n. \tag{13}$$

No attention has been paid to the convergence of the Fourier expansions involved in the arguments that motivated Definition 1, but straightforward computation of moments shows that properties (A), (B) hold when $\hat{w}$ is replaced by $w_X^{(a,\mathscr{T})}$ or $w_n^{(a,\mathscr{T})}$. Moreover

$$\left.\frac{\partial \mathbf{E} w_{X_\tau}^{(a,\mathscr{T})}(\mathscr{E})}{\partial\tau}\right|_{\tau=0} = \int k\mathscr{T}a\,dF_0,$$

so that (C) holds with $w_X^{(\hat{a},\mathscr{T})}$ substituted for $\hat{w}_X$, where $\hat{a} = \mathscr{T}^{-1}k/\|k\|$.

We show in Section 4 that, under suitable assumptions on $\mathscr{T}$, a, a central limit theorem holds for the TEP so defined.

4. Asymptotic properties of the TEP. Let us consider the TEPs $(w_n^{(a,\mathscr{T})})_{n=1,2,\ldots}$ with respect to F_0 constructed over triangular arrays of i.i.d.

variables $X_{n,1}, X_{n,2}, \dots, X_{n,n}$ with distribution function $F^{(n)}$. We describe separately the limit properties of the TEPs for $n \to \infty$ in two cases: when $F^{(n)} = F \neq F_0$ is the same fixed distribution for all samples and when $F^{(n)}$ is a sequence contiguous to F_0.

4.1. *Unboundedness under fixed alternatives.* The expectation of $w_X^{(a,\mathscr{T})}(A)$ when X has distribution F is

$$\int \mathscr{T}(a\mathbf{1}_A)(x)(dF(x) - F_0(x)) = \int \mathscr{T}(a\mathbf{1}_A)(x)dF(x) = \langle \mathscr{T}(a\mathbf{1}_A), dF/dF_0\rangle.$$

Let us call $\mathscr{T}g$ the projection of dF/dF_0 on the range of $\mathscr{T}$. Then $\mathbf{E}w_X^{(a,\mathscr{T})}(A) = \langle a\mathbf{1}_A, g\rangle = \int_A agdF_0$. From this expression we are led to the following.

THEOREM 1. *If $\int_A agdF_0 \neq 0$ for some A in a family of sets $\mathscr{J}$, then*

$$\lim_{n\to\infty} \sup_{A\in\mathscr{J}} |w_n(A)| = +\infty \quad a.s. \tag{14}$$

COROLLARY 1. *When $\mathscr{J} = \{(b, c]: b, c \in \mathbf{R}^d\}$ or $\mathscr{J} = \{(-\infty, c]: c \in \mathbf{R}^d\}$, then* (14) *holds.*

The conclusion follows from the assumptions that $\mathscr{T}$ has range $1^\perp$ and a is a.e. nonvanishing.

4.2. *Gaussian limit under the null hypothesis and contiguous alternatives.* Replacing the empirical process b_n in (13) by the F_0-Brownian bridge b, one obtains a Wiener process $w^{(a,\mathscr{T})}(A) = \int \mathscr{T}(a\mathbf{1}_A)\,db$, indexed on A (A in a given family of sets $\mathscr{J}$). Following Ossiander [Ossiander (1987)], the convergence in distribution of the transformed empirical process $w_n^{(a,\mathscr{T})}$ to $w^{(a,\mathscr{T})}$ is guaranteed, under the null hypothesis $\mathscr{H}_0$, by the assumptions

$$\mathscr{T}(a\mathbf{1}_A) \leq G \quad \text{for some } G \in L_2(\mathscr{E}, dF_0) \quad \text{and all } A \in \mathscr{J} \tag{15}$$

of uniform boundedness of the family

$$\mathscr{G} = \{\mathscr{T}(a\mathbf{1}_A): A \in \mathscr{J}\}, \tag{16}$$

and

$$\int_0^1 \sqrt{\log N_{[\,]}^{(2)}(\varepsilon, \mathscr{G}, F_0)}\,d\varepsilon < \infty \tag{17}$$

about the boundedness of the $L_2(F_0)$-*metric entropy with bracketing* $N_{[\,]}^{(2)}$, which is defined by

$$N_{[\,]}^{(2)}(\varepsilon, \mathscr{G}, F_0)] = \min\{\kappa: \text{there exist sets } \mathscr{U} \text{ and } \mathscr{L} \text{ with cardinal } \kappa \text{ of } L_2(F_0)\text{-functions such that for each } f \in \mathscr{G} \text{ there exist } u \in \mathscr{U} \text{ and } l \in \mathscr{L} \text{ such that } l < f < u \text{ and } \|u - l\|^2 \leq \varepsilon^2\}. \tag{18}$$

As for the asymptotic distribution of $w_n^{(a,\mathscr{T})}(A)$ under $\mathscr{H}_n$, it follows from the Le Cam third lemma and assumption (2) [see Le Cam and Yang (1990), Oosterhoff and van Zwet (1979)] that it is the same as under $\mathscr{H}_0$ plus the bias $\delta \int k\mathscr{T}(a\mathbf{1}_A)\, dF_0$.

The following statement summarizes the asymptotic behavior indicated above. The assumptions in (i) must be verified for each particular isometry, as will be done below for the examples in Section 9.

THEOREM 2. *Let $\{X_1, \dots, X_n\}$ be a sample of $\mathscr{E}$-valued i.i.d. random variables with distribution F and $w_n^{(a,\mathscr{T})}(A)$ the transformed empirical process of that sample associated to the probability distribution F_0 on $\mathscr{E}$, the isometry $\mathscr{T}$ on $L_2 = L_2(\mathscr{E}, dF_0)$ with range orthogonal to the constant function* 1 *and the L_2-function a ($\|a\| = 1$), as introduced in Definition* 1.

(i) *When assumptions* (15) *and* (17) *hold, $\{w_n^{(a,\mathscr{T})}(A){:}\ A \in \mathscr{J}\}$ converges in distribution to the Wiener process $\{w^{(a,\mathscr{T})}(A){:}\ A \in \mathscr{J}\}$ with covariance function*

$$\mathbf{E}w^{(a,\mathscr{T})}(A)w^{(a,\mathscr{T})}(B) = \int_{A\cap B} a^2\, dF_0,$$

under the null hypothesis $\mathscr{H}_0$: "$F = F_0$."

(ii) *When, in addition, the family of probabilities $F^{(\tau)}$ on $\mathscr{E}$ with density $f^{(\tau)}$ with respect to F_0 satisfies* (2), *then $\{w_n^{(a,\mathscr{T})}(A){:}\ A \in \mathscr{J}\}$ converges in distribution to $\{w^{(a,\mathscr{T})}(A) + \delta \int k\mathscr{T}(\mathbf{1}_A a)\, dF_0{:}\ A \in \mathscr{J}\}$, under the sequence of alternatives $\mathscr{H}_n$: "$F = F^{(\delta/\sqrt{n})}$."*

These results justify the test procedure described in next section.

5. The goodness-of-fit tests. We propose to test $\mathscr{H}_0$: "$F = F_0$" by means of the critical region

$$\sup_{A\in\mathscr{J}} \left| w_n^{(\hat{a},\mathscr{T})}(A) \right| > c(\alpha)$$

with $\mathscr{T}\hat{a} = k/\|k\|$ and $c(\alpha)$ such that

$$\mathbf{P}\left\{\sup_{A\in\mathscr{J}} \left| w^{(\hat{a}\mathscr{T})}(A) \right| > c(\alpha)\right\} \le \alpha,$$

where α is an upper bound of the asymptotic level desired for the test.

The family $\mathscr{J}$ of measurable subsets of $\mathscr{E}$ is chosen rich enough to generate the σ-field of measurable sets in $\mathscr{E}$, but not so large that $\sup_{A\in\mathscr{J}} \left| w^{(\hat{a}\mathscr{T})}(A) \right|$ be unbounded. It is assumed that $\mathscr{J}$ contains $\mathscr{E}$. Moreover, the compromise in choosing $\mathscr{J}$ is that a CLT

$$\lim_{n\to\infty} \sup_{A\in\mathscr{J}} |\hat{w}_n(A) - w^V(A)| = 0 \tag{19}$$

holds for copies $\hat{w}_n$ of $w_n^{(\hat{a},\mathscr{T})}$ and a V-Wiener process w^V, and also

$$\text{for } F \ne F_0, \text{ there is } A \in \mathscr{J} \text{ such that } F(A) \ne F_0(A). \tag{20}$$

We shall choose for $\mathscr{J}$ either the family of all generalized intervals

$$\mathscr{J}_1 = \Big\{ \bigtimes_{i=1}^{d} (p_i, q_i]\colon -\infty \le p_i < q_i \le \infty,\ i = 1, 2, \dots, d \Big\}$$

or the family of translations of the negative orthant

$$\mathscr{J}_0 = \Big\{ \bigtimes_{i=1}^{d} (-\infty, q_i]\colon q_i \le \infty,\ i = 1, 2, \dots, d \Big\}.$$

Since $\mathscr{E} \in \mathscr{J}$, the asymptotic power of the test for $\mathscr{H}_n$ is greater than

$$\mathbf{P}\big\{ w^{(\hat{a}, \mathscr{T})}(\mathscr{E}) + \delta \|k\| > c(\alpha) \big\} = \Phi(-c(\alpha) + \delta \|k\|).$$

An argument like the one used in Cabaña (1996) or Cabaña and Cabaña (1996) shows that when the level α and the probability β of type II error are sufficiently small, the asymptotic relative efficiency of our test with respect to the maximum likelihood ratio test can be chosen as close to one as desired. In this sense, the test has optimum efficiency.

For other values of α and β there is a reduction in the efficiency, due to the use as the test variable of the supremum of $w_n^{(\hat{a}, \mathscr{T})}$ over a family of sets instead of the value of the same process on $\mathscr{E}$. This is the price to be paid for the consistency against all alternatives.

Notice that the TEP could be defined on families of functions different from indicators of sets in $\mathscr{J}$, but we have chosen this particular index set because the reflection principle for Brownian motion provides the exact distribution of the supremum of the process over the family.

6. On the construction of isometries and their associated transformed empirical processes. Let us notice that any isometry $\mathscr{T}$ on $L^2(\mathscr{E}, dF_0)$ with range contained in $1^\perp = \{f \in L^2(\mathscr{E}, dF_0)\colon \langle f, 1\rangle = 0\}$ induces an orthonormal system $\Psi(\mathscr{T}) = (1, \mathscr{T}1, \mathscr{T}^2 1, \dots, \mathscr{T}^i 1, \dots)$.

Conversely, given any orthonormal system $\Psi = (1, \psi_1, \psi_2, \dots)$, the linear transformation $\mathscr{T}(\Psi)$ on $L^2(\mathscr{E}, dF_0)$ that maps each ψ_i onto ψ_{i+1}, $i = 0, 1, \dots,$ is an isometry with range in $1^\perp$. In addition, when Ψ is a basis, the range is $1^\perp$ and therefore it contains any function k with the property (3).

These observations imply that the isometries needed for the construction of TEPs may possibly be obtained from known orthonormal systems in $L^2(\mathscr{E}, dF_0)$.

In Section 6.1 we indicate the analytical form of the shift $\mathscr{T}_L$ induced by the Laguerre polynomials on $L^2(\mathbf{R}^+, e^{-x}\, dx)$. A simple analytical property of $\mathscr{T}_L$ reflected in (23) gives an alternative way to obtain this isometry.

For a given orthonormal system Ψ, it is not easy in general to find the explicit analytical expression for the shift operator that maps each ψ_n onto ψ_{n+1}. We describe it for a particular example in Section 6.2: the orthonormal system of Tchebyshev polynomials on $L_T^2 = L^2([-1, 1], dx/(\pi\sqrt{1 - x^2}))$.

The normalized Hermite polynomials $h_n(x) = H_n(x)/\sqrt{n!}$, $n = 0, 1, \dots,$ where $e^{tx - t^2/2} = \sum_{n=0}^{\infty} H_n(x)(t^n/n!)$ [Sansone (1959)], constitute an orthonormal basis of $L_H^2 = L^2(\mathbf{R}, e^{-x^2/2}\, dx/\sqrt{2\pi})$. An explicit writing of the shift that maps h_n onto h_{n+1} in terms of integrals of the Poisson kernel [see Morán and Urbina (1996) for the form of the inverse shift], leads to cumbersome computations so that the alternative approach, namely, the generalization of (23), is adopted in Section 6.3 to construct an isometry $\mathscr{T}_H$ on that space.

By means of changes of variables, the isometries in Sections 6.1, 6.2 and 6.3 induce others on different L^2 spaces on $\mathbf{R}$, as described in Section 7, and the latter ones can be used in the construction of new isometries on L^2 spaces on $\mathbf{R}^d$ as shown in Section 8.

The *Laguerre shift* (29) has been used many times in probability theory and statistical inference: for instance, let us mention that Brownian bridge is constructed from Wiener process by means of its inverse in Karatzas and Shreve (1991), that Efron and Johnstone (1990) and Ritov and Wellner (1988) use (29) and its inverse in connection with hazard rates [some properties of both isometries are described in Groeneboom and Wellner (1992)], and that Khmaladze (1981) introduced (29) in statistical inference, specially emphasizing some associated martingale properties. We use it as our main example because of its very simple analytical expression, but do not apply any martingale approach.

6.1. *The Laguerre shift.* The well-known Laguerre polynomials $\{L_n: n = 0, 1, \dots\}$ are an orthonormal basis of $L^2(\mathbf{R}, dF_0)$, with $F_0(x) = F_0([0, x)) = 1 - e^{-x}$, and they are obtained by means of the iterated application of the mapping

$$h(x) = \mathscr{T}_L g(x) = g(x) - \int_0^x g(t)\,dt \tag{21}$$

to the first element of the basis, the polynomial of degree zero $L_0 = 1$:

$$L_n(x) = (-1)^n \mathscr{T}_L^n L_0(x) \tag{22}$$

[see, for instance, Sansone (1959), as a general reference].

This implies that $\mathscr{T}_L$ is an isometry and, for each $g \in L^2(\mathbf{R}, dF_0)$, $\mathscr{T}_L g$ is orthogonal to the constant 1.

Both properties of $\mathscr{T}_L$ follow immediately after plain calculations, and, conversely, imply (22). The clue to show that $\mathscr{T}_L$ is an isometry is the equation

$$\mathscr{T}_L g(x) \mathscr{T}_L h(x) f_0(x) = g(x)h(x)f_0(x) - \frac{d}{dx}\left(f_0(x) \int_0^x g(t)\,dt \int_0^x h(t)\,dt \right) \tag{23}$$

and the fact that $\lim_{x\to\infty} f_0(x) \int_0^x g(t)\,dt \int_0^x h(t)\,dt = 0$.

6.2. *The Tchebycheff shift.* The Tchebycheff polynomials $T_0(x) = 1$, $T_n(x) = \sqrt{2}\cos(n \arccos x)$, $n = 1, 2, \dots$ are an orthonormal basis of $L_T^2 = L^2([-1, 1], dx/(\pi\sqrt{1 - x^2}))$. The isometry $\mathscr{T}_T$ that maps each T_n onto T_{n+1} can be described as follows.

Given $u(x) \in L_T^2$, let us assume first that $\langle u, 1\rangle = \int_{-1}^1 u(x)\,dx/(\pi\sqrt{1 - x^2}) = 0$. Perform the change of variable $\tilde{u}(|t|) = u(\cos t)$, $-\pi < t \le \pi$, obtain an analytic function $h^{(u)}$ on $D = \{\zeta = re^{it}: |\zeta| < 1\}$ such that $\tilde{u}_r(t) =$

$\Re\, h^{(u)}(re^{it})$ converges in $L^2((-\pi,\pi], dt)$-norm to $\tilde{u}(|\, t\,|)$ as $r \to 1$ and take the $L^2((-\pi,\pi], dt)$-limit

$$v(\cos t) = \lim_{r=|\zeta|\to 1} \Re\, \zeta h^{(u)}(\zeta).$$

This limit is the image $v = \mathscr{T}_T u$ of u.

The function $h^{(u)}$ is determined up to an imaginary additive constant, and the one satisfying $\Im h^{(u)}(0) = 0$ is obtained by integrating the Poisson kernel:

$$\begin{aligned} h^{(u)}(\zeta) &= \frac{1}{2\pi i}\int_C u(\Re\, z)\frac{z+\zeta}{z-\zeta}\frac{dz}{z} \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} u(\cos s)\left(\frac{1-r^2+2ir\sin(s-t)}{1-2r\cos(s-t)+r^2}\right) ds \\ &= u(\cos t) + \frac{1}{2\pi}\int_{-\pi}^{\pi}(u(\cos s)-u(\cos t))\left(\frac{1-r^2+2ir\sin(s-t)}{1-2r\cos(s-t)+r^2}\right) ds. \end{aligned}$$

Taking the limit in

$$\begin{aligned} \Re\zeta h^{(u)}(\zeta) = r\cos t\, u(\cos t) &+ \frac{1}{2\pi}\int_{-\pi}^{\pi}(u(\cos s)-u(\cos t)) \\ &\times \left(\frac{r(1-r^2)\cos t - 2r^2\sin(s-t)\sin t}{1-2r\cos(s-t)+r^2}\right) ds, \end{aligned}$$

one obtains

$$\begin{aligned} \mathscr{T}_T u(\cos t) &= \lim_{r\to 1} \Re\, \zeta h^{(u)}(\zeta) \\ &\quad + \frac{1}{2\pi}\int_{-\pi}^{\pi}(u(\cos s)-u(\cos t))\left(\frac{-\sin(s-t)\sin t}{1-\cos(s-t)}\right) ds \\ &= \cos t\, u(\cos t) - \frac{\sin(t)}{2\pi}\int_{-\pi}^{\pi}\frac{u(\cos s)-u(\cos t)}{\tan((s-t)/2)}\, ds. \end{aligned}$$

Any $u(x) \in L_T^2$ can be written as the sum of the constant $\langle u, 1\rangle$ and the function $u(x) - \langle u, 1\rangle$, orthogonal to 1. The image of $\langle u, 1\rangle$ by $\mathscr{T}_T$ is $\sqrt{2}\langle u, 1\rangle x$, and the image of $u(x) - \langle u, 1\rangle$ is obtained as indicated above.

It may be noticed, in particular, that the images by $\mathscr{T}_T$ of functions such as the product $a(\cdot)\mathbf{1}_x(\cdot)$ of a continuous nonvanishing function a times the indicator function of a half line are not bounded. This makes this isometry useless for our present purposes.

6.3. *An isometry associated with the normal distribution.* Let $\varphi(x) = e^{-x^2/2}/\sqrt{2\pi}$ denote the normal density. The analogue to (23),

$$\begin{aligned} &\mathscr{S} g(x)\mathscr{S} h(x)\varphi(x) \\ &\quad = g(x)h(x)\varphi(x) - \frac{d}{dx}\left(\sqrt{|x|}\int_0^x \sqrt{|t|}\, g(t)\, dt\sqrt{|x|}\int_0^x \sqrt{|t|}\, h(t)\, dt\varphi(x)\right), \end{aligned} \tag{24}$$

where

$$\mathscr{S}\colon f(x) \mapsto f(x) - \operatorname{sgn} x\sqrt{|x|}\int_0^x \sqrt{|t|}\, f(t)\, dt \tag{25}$$

and

$$\lim_{x\to\infty} \sqrt{|x|}\int_0^x \sqrt{|t|}\, g(t)\, dt\sqrt{|x|}\int_0^x \sqrt{|t|}\, h(t)\, dt\varphi(x) = 0$$

imply that (25) defines an isometry $\mathscr{S}$ on $L^2(\mathbf{R}, \varphi(x)\, dx)$.

Isometry $\mathscr{S}$ maps even functions onto even functions orthogonal to $\sqrt{|x|}$ and odd functions onto odd functions orthogonal to $\operatorname{sgn}(x)\sqrt{|x|}$. As a consequence, the range of $\mathscr{S}$ is orthogonal to $\operatorname{sgn}(x)\sqrt{|x|}$ and $\sqrt{|x|}$, but not to the constant 1 as required. The new isometry $\mathscr{T}_H$ given by

$$\mathscr{T}_H f = \begin{cases} f, & f \text{ odd},\\ \mathscr{S}f + \dfrac{1}{1-\cos\gamma}\langle \mathscr{S}f, 1\rangle\, (u-1), & f \text{ even}, \end{cases} \tag{26}$$

where $u(x) = \sqrt{|x|}/\|\sqrt{|\cdot|}\| = \sqrt[4]{\pi/2}\sqrt{|x|}$, $\cos\gamma = \langle 1, u\rangle = \sqrt[4]{1/\pi}\ \Gamma(3/4)$, has range orthogonal to 1.

Since any function f is the sum of an *odd part* $f^o(x) = (f(x) - f(-x))/2$ and an *even part* $f^e(x) = (f(x) + f(-x))/2$, (26) completely defines $\mathscr{T}_H$, which can also be written as

$$\mathscr{T}_H f(x) = f(x) - \sqrt{|x|}\int_0^{|x|} \sqrt{s} f^e(s)\, ds + \frac{1}{1-\cos\gamma}\langle \mathscr{S}f^e, 1\rangle(u(x)-1). \tag{27}$$

The inverse of $\mathscr{S}$ is given by

$$(\mathscr{S}^{-1}h)(x) = h(x) - \frac{\sqrt{|x|}}{\varphi(x)}\int_x^\infty \operatorname{sgn} t\sqrt{|t|}h(t)\varphi(t)\, dt$$

and hence

$$\begin{aligned}(\mathscr{T}_H^{-1}h)(x) &= h^o(x) + \mathscr{S}^{-1}\Bigl(h^e - \frac{1}{1-\cos\gamma}\langle h^e, u\rangle(u-1)\Bigr)\\ &= h(x) - \frac{1}{1-\cos\gamma}\langle h^e, u\rangle(u-1)\\ &\quad - \frac{\sqrt{|x|}}{\varphi(x)}\int_{|x|}^\infty \sqrt{t}\Bigl[h^e(t) - \frac{1}{1-\cos\gamma}\langle h, u\rangle(u(t)-1)\Bigr]\varphi(t)\, dt.\end{aligned} \tag{28}$$

In particular, (28) reduces to h when h is odd.

7. Construction of isometries on $L^2(\mathbf{R}, dF_0)$ for an arbitrary probability F_0, by means of a change of variables. Let us assume that we are given an isometry $\mathscr{T}_F$ on $L_2(\mathscr{E}, dF)$ and wish to construct a new isometry $\mathscr{T}_{F_0}$ on $L^2(\mathscr{E}, dF_0)$. The following lemma gives a simple and general procedure for such construction:

LEMMA 1. *If $\mathscr{T}_F$ is an isometry on $L_2(\mathscr{E}, dF)$, then the mapping $g_0 \mapsto \mathscr{T}_{F_0} g_0$ defined by*

$$(\mathscr{T}_{F_0} g_0)(x) = (\mathscr{T}_F g_0 \circ F_0^{-1} \circ F)(F^{-1}(F_0(x)))$$

is an isometry on $L^2(\mathscr{E}, dF_0)$.

PROOF. Let us assume $X_0 \sim F_0$, and $\mathbf{E} g_0^2(X_0), \mathbf{E} h_0^2(X_0) < \infty$.

The change of variables $X = F^{-1}(F_0(X_0))$ maps X_0 onto $X \sim F$, and hence $\mathbf{E} g_0^2(X_0) < \infty$, $\mathbf{E} h_0^2(X_0) < \infty$ are equivalent to $\mathbf{E} g^2(X) < \infty$, $\mathbf{E} h^2(X) < \infty$, respectively, for $g = g_0 \circ F_0^{-1} \circ F$, and $h = h_0 \circ F_0^{-1} \circ F$.

We may then compute

$$\begin{aligned}\int \mathscr{T}_{F_0} g_0 \mathscr{T}_{F_0} h_0 dF_0 &= \mathbf{E} \mathscr{T}_{F_0} g_0(X_0) \mathscr{T}_{F_0} h_0(X_0) \\ &= \mathbf{E}(\mathscr{T}_F g_0 \circ F_0^{-1} \circ F)(F^{-1}(F_0(X_0))) \\ &\quad \times (\mathscr{T}_F h_0 \circ F_0^{-1} \circ F)(F^{-1}(F_0(X_0))) \\ &= \mathbf{E}(\mathscr{T}_F g)(X)(\mathscr{T}_F h)(X).\end{aligned}$$

Since $\mathscr{T}_F$ is an isometry on $L_2(\mathscr{E}, dF)$, the right-hand term equals

$$\mathbf{E} g(X) h(X) = \mathbf{E} g_0(X_0) h_0(X_0) = \int g_0 h_0 \, dF_0,$$

and this proves Lemma 1. $\square$

7.1. *Example* 1. *L-isometries.* From Lemma 1 applied to the isometry in Section 6.1, we get, for each F_0, a new isometry

$$(\mathscr{T}_{L, F_0} g)(x) = g(x) - \int_{-\infty}^{x} \frac{g(t)}{1 - F_0(t)} dF_0(t), \tag{29}$$

on $L_2(\mathbf{R}, dF_0)$. Its inverse

$$(\mathscr{T}_{L, F_0}^{-1} h)(x) = h(x) + \frac{1}{1 - F_0(x)} \int_{-\infty}^{x} h(t) \, dF_0(t) \tag{30}$$

is obtained by solving (29) for g. In what follows, any isometry in the class defined by (29) will be called an *L-isometry*. Under suitable assumptions, $\mathscr{T}_{L, F_0}$ satisfies (15) and (17), and hence Theorem 2 applies.

LEMMA 2. *When $|a|/(1 - F_0)^\alpha$ belongs to $L^2(\mathbf{R}, dF_0)$ for some positive α, then $\mathscr{T}_{L, F_0}$ satisfies* (15) *and* (17) *with $\mathscr{J} = \{(-\infty, x] \colon x \in \mathbf{R}\}$.*

PROOF. The function $|\mathscr{T}_{L, F_0}(a \mathbf{1}_{(-\infty, y]})|$ is bounded by

$$G = |a| + \int_{-\infty}^{\cdot} \frac{|a(s)|}{1 - F_0(s)} \, dF_0(s)$$

uniformly in y. Let us assume with no loss of generality that $\alpha < 1/2$. The inequalities

$$\begin{aligned}
&\int_{-\infty}^{\infty}\left(\int_{-\infty}^{x}\frac{|a(s)|}{1-F_0(s)}\,dF_0(s)\right)^2 dF_0(x)\\
&\quad\le \int_{-\infty}^{\infty}\left(\int_{-\infty}^{x}\frac{a(s)^2}{(1-F_0(s))^{2\alpha}}\,dF_0(s)\int_{-\infty}^{x}\frac{dF_0(s)}{(1-F_0(s))^{2-2\alpha}}\right)dF_0(x)\\
&\quad\le \left\|\frac{|a|}{(1-F_0)^{\alpha}}\right\|^2\frac{1}{1-2\alpha}\int_{-\infty}^{\infty}\left(\frac{1}{(1-F_0(x))^{1-2\alpha}}-1\right)dF_0(x)\\
&\quad\le \left\|\frac{|a|}{(1-F_0)^{\alpha}}\right\|^2\frac{1}{2\alpha(1-2\alpha)}
\end{aligned}$$

and the assumptions on a imply that G is in $L^2(\mathbf{R}, dF_0)$, so that (15) holds.

Given a positive ε, let us construct a (finite) partition ($x_0 = -\infty$, $x_1, x_2, \dots, x_\nu = \infty$) of $\mathbf{R}$ such that for each $i = 1, 2, \dots, \nu$,

$$\int_{x_{i-1}}^{x_i} a^2(s)\,dF_0(s) \le \varepsilon^2/8 \tag{31}$$

and also

$$\int_{x_{i-1}}^{x_i}\frac{a^2(s)}{(1-F_0(s))^{2\alpha}}\,dF_0(s) \le \frac{\alpha(1-2\alpha)\varepsilon^2}{4},$$

so that

$$\int_{x_{i-1}}^{x_i}\left(\int_{x_{i-1}}^{x}\frac{|a|}{1-F_0}\,dF_0\right)^2 dF_0(x) \le \frac{\alpha(1-2\alpha)\varepsilon^2}{4}\,\frac{1}{2\alpha(1-2\alpha)} = \varepsilon^2/8. \tag{32}$$

Notice that a partition satisfying (31) can be constructed with $\nu \le 1 + 8/\varepsilon^2$ intervals, and another one satisfying (32) requires at most $1 + ((4\|a/(1-F_0)^{\alpha}\|^2)/\alpha(1-2\alpha)\varepsilon^2)$. The partition obtained by joining the points in both of the above partitions satisfies our requirements and can therefore be achieved with $\nu \le C/\varepsilon^2$, where C may depend on a (and α) but not on ε.

For all $y \in [x_{i-1}, x_i]$, $\mathscr{T}(a\mathbf{1}_{(-\infty, y]})$ is bounded from above by

$$u_i = a\mathbf{1}_{(-\infty, x_{i-1})} + |a|\mathbf{1}_{[x_{i-1}, x_i]} - \int_{-\infty}^{x_{i-1}\wedge\cdot}\frac{a}{1-F_0}\,dF_0 + \int_{x_{i-1}\wedge\cdot}^{x_i\wedge\cdot}\frac{|a|}{1-F_0}\,dF_0$$

and from below by

$$l_i = a\mathbf{1}_{(-\infty, x_{i-1})} - |a|\mathbf{1}_{[x_{i-1}, x_i]} - \int_{-\infty}^{x_{i-1}\wedge\cdot}\frac{a}{1-F_0}\,dF_0 - \int_{x_{i-1}\wedge\cdot}^{x_i\wedge\cdot}\frac{|a|}{1-F_0}\,dF_0.$$

These bounds satisfy

$$\|u_i - l_i\|^2 \le \left\|2\left(|a|\mathbf{1}_{[x_{i-1}, x_i]} + \int_{x_{i-1}\wedge\cdot}^{x_i\wedge\cdot}\frac{|a|}{1-F_0}\,dF_0\right)\right\|^2 \le \varepsilon^2$$

as a consequence of (31) and (32), hence $N_{[\ ]}^{(2)} \le C/\varepsilon^2$, and condition (17) holds. This ends the proof of Lemma 2. □

7.2. *Example* 2. *N-isometries.* When F_0 is a symmetric probability distribution function ($F_0(x)+F_0(-x)=1$), then $(g\circ F_0^{-1}\circ\Phi)^e = g^e\circ F_0^{-1}\circ\Phi$ and $(g\circ F_0^{-1}\circ\Phi)^o = g^o\circ F_0^{-1}\circ\Phi$, and hence, from the isometry in Section 6.3, we get the new isometry

$$
\begin{aligned}
(\mathscr{T}_{H,F_0}g)(x) &= g(x) - \sqrt{|\Phi^{-1}(F_0(x))|}\int_0^{|x|}\frac{\sqrt{|\Phi^{-1}(F_0(s))|}\,g^e(s)}{\varphi(\Phi^{-1}(F_0(s)))}\,dF_0(s)\\
&\qquad + \frac{1}{1-\cos\gamma}\int \mathscr{S}(g^e\circ F_0^{-1}\circ\Phi)(z)\varphi(z)\,dz\\
&\qquad\qquad \times\big(u(\Phi^{-1}(F_0(x)))-1\big)
\end{aligned}\tag{33}
$$

on $L_2(\mathbf{R}, dF_0)$. A general version of (33) for nonsymmetric F_0 is equally easy to obtain, but its expression is even more complicated. The mappings given by (33) are called *N-isometries* in the following.

LEMMA 3. *When $|a|^e/(\varphi(\Phi^{-1}(F_0(\cdot)))^\alpha$ belongs to $L^2(\mathbf{R}, dF_0)$ for some positive α, then T_{H,F_0} satisfies* (15) *and* (17) *with $\mathscr{J}=\{(-\infty,x]: x\in\mathbf{R}\}$.*

PROOF. Proceed as in the proof of Lemma 2: now $|\mathscr{T}_{H,F_0}(a\mathbf{1}_{(-\infty,y]})|$ is uniformly bounded by

$$
\begin{aligned}
G = |a| &+ \sqrt{|\Phi^{-1}(F_0(x))|}\int_0^{|x|}\frac{\sqrt{|\Phi^{-1}(F_0(s))|}\,|a|^e(s)}{\varphi(\Phi^{-1}(F_0(s)))}\,dF_0(s)\\
&+\frac{1}{1-\cos\gamma}\|a\|\big(1\vee u(\Phi^{-1}(F_0(s)))\big).
\end{aligned}
$$

The first and last term in G are square integrable, because $|a(\cdot)|$ and $u(\Phi^{-1}(F_0(\cdot)))$ are in $L^2(\mathbf{R}, dF_0)$. In order to verify that the middle term has the same property, we assume again $\alpha<1/2$ and derive the inequalities

$$
\begin{aligned}
&\int_{-\infty}^{\infty}|\Phi^{-1}(F_0(x))|\left(\int_0^x\frac{\sqrt{|\Phi^{-1}(F_0(s))|}\,|a|^e(s)}{\varphi(\Phi^{-1}(F_0(s)))}\,dF_0(s)\right)^2 dF_0(x)\\
&\quad\le\int_{-\infty}^{\infty}|\Phi^{-1}(F_0(x))|\left|\int_0^x\frac{|\Phi^{-1}(F_0(s))|\,dF_0(s)}{\varphi^{2-2\alpha}(\Phi^{-1}(F_0(s)))}\int_0^x\frac{(|a|^e(s))^2\,dF_0(s)}{\varphi^{2\alpha}(\Phi^{-1}(F_0(s)))}\right|dF_0(x)\\
&\quad\le M_a\int_{-\infty}^{\infty}|z|\left|\int_0^z\frac{t\varphi(t)\,dt}{\varphi^{2-2\alpha}(t)}\right|\varphi(z)\,dz\\
&\quad=\frac{M_a}{1-2\alpha}\int_{-\infty}^{\infty}|z|\varphi(z)\left(\frac{1}{\varphi^{1-2\alpha}(z)}-\frac{1}{\varphi^{1-2\alpha}(0)}\right)dz<\infty,
\end{aligned}
$$

thus providing (15).

The estimation of the $L_2(F_0)$-metric entropy with bracketing is made as in the proof of Lemma 2, and it leads to the conclusion that condition (17) also holds. □

8. Constructing isometries on $L_2(\mathscr{E}, dF_0)$, $\mathscr{E}$ space of dimension greater than one. When $\mathscr{E}$ is the Cartesian product $\mathscr{E}=\mathbf{R}\times\mathscr{E}_1$, the measure F_0 is written in terms of the *marginal measure* $\tilde F_0(J)=F_0(J\times\mathscr{E}_1)$

and the *conditional measures* $dF_0^c(y, \bullet)$ defined on $\mathscr{E}_1$ by $F_0(J \times B) = \int_{y \in J} d\tilde{F}_0(y) \int_{z \in B} dF_0^c(y, z)$.

In other words, if $X = (Y, Z) \sim F_0$, then $Y \sim \tilde{F}_0$ and the conditional distribution of Z given $Y = y$ is $\mathbf{P}\{Z \in B \mid Y = y\} = \int_B dF_0^c(y, \bullet)$.

For each g in $L_2(\mathscr{E}, dF_0)$, we denote

$$\tilde{g}(y) = \mathbf{E}(g(Y, Z) \mid Y = y). \tag{34}$$

This new function is in $L_2(\mathbf{R}, d\tilde{F}_0)$ and the mapping $g \mapsto \tilde{g}$ preserves the norm.

LEMMA 4. *If $g \in L_2(\mathscr{E}, dF_0)$ and $\tilde{\mathscr{T}}$ is an isometry on $(\mathbf{R}, d\tilde{F}_0)$, then*

$$g \mapsto \mathscr{T} g(y, z) = g(y, z) - \tilde{g}(y) + \tilde{\mathscr{T}} \tilde{g}(y) \tag{35}$$

is an isometry on $L_2(\mathscr{E}, dF_0)$.

PROOF. Given $g, h \in L_2(\mathscr{E}, dF_0)$, let us compute $\langle \mathscr{T} g, \mathscr{T} h\rangle = \mathbf{E}(g(Y, Z) - \tilde{g}(Y) + \tilde{\mathscr{T}}\tilde{g}(Y))(h(Y, Z) - \tilde{h}(Y) + \tilde{\mathscr{T}}\tilde{h}(Y))$. The equalities $\mathbf{E}g(Y, Z)\tilde{h}(Y) = \mathbf{E}\tilde{h}(Y)\mathbf{E}(g(Y, Z) \mid Y) = \mathbf{E}\tilde{g}(Y)\tilde{h}(Y)$, $\mathbf{E}(g(Y, Z) - \tilde{g}(Y))\tilde{\mathscr{T}}\tilde{h}(Y) = 0$, and similar ones obtained by interchanging g and h lead us to write $\langle \mathscr{T} g, \mathscr{T} h\rangle = \mathbf{E}g(Y, Z)h(Y, Z) - \mathbf{E}\tilde{g}(Y)\tilde{h}(Y) + \mathbf{E}\tilde{\mathscr{T}}\tilde{g}(Y)\tilde{\mathscr{T}}\tilde{h}(Y) = \langle g, h\rangle - \langle \tilde{g}, \tilde{h}\rangle + \langle \tilde{\mathscr{T}}\tilde{g}, \tilde{\mathscr{T}}\tilde{h}\rangle = \langle g, h\rangle$, and this ends the proof of the lemma. □

9. One application: consistent goodness-of-fit to the standard normal distribution. As an illustration of the general procedure for the design of consistent and efficient tests contained in the preceding sections, we describe the tests associated to the isometries in Sections 6.1 and 6.3 for goodness-of-fit to the standard normal distribution in $\mathbf{R}$ and $\mathbf{R}^2$. Two cases are considered: in Case 1, the tests are designed to have optimum sensitivity against shifts of the mean, while in Case 2, the alternative to be detected is a change of dispersion.

9.1. *Case* 1. *Tests designed for detection of shifts in the mean.*

9.1.1. *The one-dimensional test.* Let $\mathscr{E} = \mathbf{R}$, $F_0(x) = \Phi(x) = \int_{-\infty}^x \varphi(t)\,dt$, $\varphi(t) = (1/\sqrt{2\pi})e^{-t^2/2}$ and $F^{(\tau)}(x) = \Phi(x - \tau)$. The ratio of the densities is $((f^{(\tau)}(x))/f_0(x)) = 1 + x\tau + o(\tau)$, so that $k(x) = x$, $\|k\|^2 = \int_{-\infty}^{\infty} x^2\varphi(x)\,dx = 1$.

EXAMPLE 3 (TEP associated to the L-isometry). The score function is $\hat{a}(x) = \mathscr{T}_{L,\Phi}^{-1}(x) = x + (1/(1 - \Phi(x)))\int_{-\infty}^x t\varphi(t)\,dt = x - (\varphi(x)/(1 - \Phi(x)))$ [see (30)], and hence the TEP for the sample $\{X\}$ of size 1 is

$$\begin{aligned}
w_X(x) = w_X((-\infty, x]) &= \hat{a}(X)\mathbf{1}_{\{X \le x\}} - \int_{-\infty}^X \frac{\hat{a}(t)\mathbf{1}_{\{t \le x\}}\varphi(t)\,dt}{1 - \Phi(t)} \\
&= \left(X - \frac{\varphi(X)}{1 - \Phi(X)}\right)\mathbf{1}_{\{X \le x\}} + \int_{-\infty}^{X \wedge x} d\left(\frac{\varphi(t)}{1 - \Phi(t)}\right) \\
&= X\mathbf{1}_{\{X \le x\}} + \frac{\varphi(x)}{1 - \Phi(x)}\mathbf{1}_{\{x < X\}}.
\end{aligned}$$

In order to describe the general shape of the TEPs and their response to changes in the position or in the dispersion of the samples, we introduce the fictitious sample of size 9: $(\Phi^{-1}(i/10))_{i=1,2,\dots,9}$, which will be referred as *the special sample* in the following. Then we compute the TEPs associated with the optimum score function $\hat{a}$ for the special sample, and for the *shifted* and *dispersed special samples* $(\Phi^{-1}(i/10)+1)_{i=1,2,\dots,9}$ and $(2\Phi^{-1}(i/10))_{i=1,2,\dots,9}$. The corresponding three diagrams are presented in the left-hand side of Figure 1.

Since k is odd, then $\hat{a}(x) = \mathscr{T}_{N,\Phi}^{-1}k(x) = k(x) = x$, and

$$\begin{aligned}
w_X(x) &= w_X((-\infty, x]) = \mathscr{T}_{N,\Phi}(\cdot\mathbf{1}_{(-\infty,x]}(\cdot))(X) \\
&= X\mathbf{1}_{\{X\le x\}} - \sqrt{|X|}\int_0^{|X|}\sqrt{s}\left(\frac{-s}{2}\right)\mathbf{1}_{\{|x|\le s\}}\,ds \\
&\quad + \frac{u(X)-1}{1-\cos\gamma}\int\left(\frac{-s}{2}\mathbf{1}_{\{|x|\le|s|\}} - \sqrt{|s|}\int_0^{|s|}\sqrt{t}\left(\frac{-t}{2}\right)\mathbf{1}_{\{|x|<t\}}\,dt\right)\varphi(s)\,ds \\
&= X\mathbf{1}_{\{X\le x\}} + \frac{\sqrt{|X|}}{5}(|X|^{5/2} - |x|^{5/2})^+ \\
&\quad + \frac{u(X)-1}{1-\cos\gamma}\left[-\varphi(|x|) + \frac{2}{5}\int_{|x|}^{\infty}\varphi(s)\sqrt{s}(s^{5/2}-|x|^{5/2})\,ds\right].
\end{aligned}$$

The right-hand side of Figure 1 shows the shapes of the new TEPs.

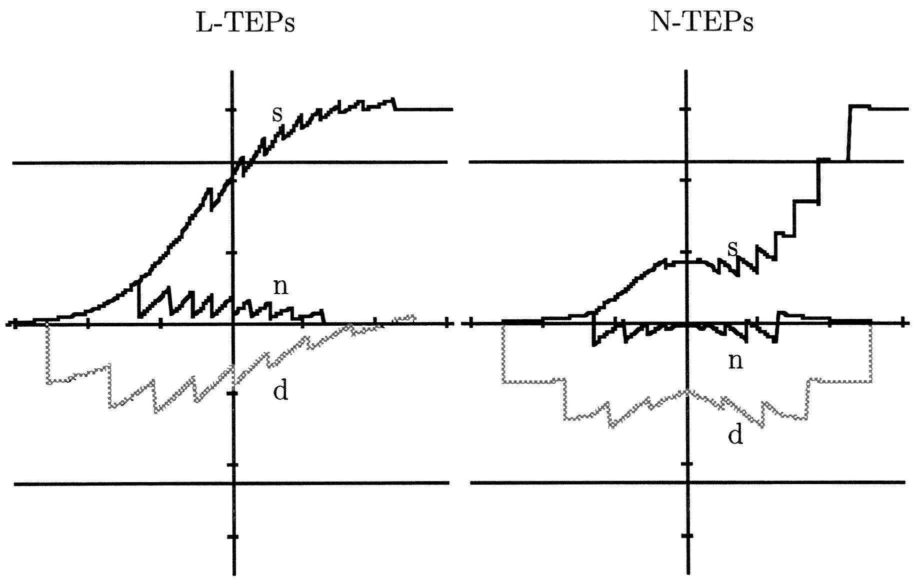

FIG. 1. *Responses to normal* (*n*), *shifted* (*s*) *and dispersed* (*d*) *samples*, *of the TEPs associated with L-isometry* (*L-TEPs*) *and with N-isometry* (*N-TEPs*), *optimized to detect changes in position.*

9.1.2. *The test for $d = 2$.* Let $F^{(\tau)}$ be the normal distribution with mean $\binom{\tau}{0}$ and variance matrix equal to the identity, and let $F_0 = F^{(0)}$. The density under $\mathscr{H}_0$ is $f_0(y, z) = \varphi(y)\varphi(z)$, and under the alternative, $f^{(\tau)}(y, z) = \varphi(y-\tau)\varphi(z)$. The densities ratio is $f^{(\tau)}(y, z)/f_0(y, z) = \varphi(y-\tau)/\varphi(y) = 1 + x\tau + o(\tau)$, and hence the drift is $k(y, z) = y$.

We choose now to project the plane measure on the direction of the shift, namely, the y axis, on, and obtain the corresponding marginal distribution $\tilde{F}_0 = \Phi$. Lemma 4 is now applied to construct two isometries on $L^2(\mathbf{R}^2, dF_0)$, by substituting $\mathscr{T}_{L,\tilde{F}_0}$ and $\mathscr{T}_{H,\tilde{F}_0}$ for $\tilde{\mathscr{T}}$ in (35):

$$\mathscr{T}_{L,F_0} g(y, z) = g(y, z) - \mathbf{E}g(y, Z) + \mathscr{T}_{L,\Phi}\mathbf{E}g(\cdot, Z)(y),$$
$$\mathscr{T}_{H,F_0} g(y, z) = g(y, z) - \mathbf{E}g(y, Z) + \mathscr{T}_{H,\Phi}\mathbf{E}g(\cdot, Z)(y),$$

Z standard normal.

When $g(y, z) = \tilde{g}(y)$ does not depend on z, $\mathscr{T} g = \tilde{\mathscr{T}}\tilde{g}$. In particular, since the drift k depends only on y, this implies that the score functions corresponding to $\mathscr{T}_{L,F_0}$ and $\mathscr{T}_{H,F_0}$ also depend only on y and each one is given by the formula obtained for $d = 1$, respectively

$$\hat{a}_L(y, z) = y - \frac{\varphi(y)}{1-\Phi(y)} \quad \text{and} \quad \hat{a}_H(y, z) = y.$$

Consequently, the TEP for a single observation is, in the former case:

$$\begin{aligned} w^L_{Y,Z}(A) &= w^{(\mathscr{T}_L,\hat{a})}_{Y,Z}(A) \\ &= \hat{a}_L(Y, Z)\mathbf{1}_{\{(Y,Z)\in A\}} - \iint_A \hat{a}_L(y, z)\varphi(y)\varphi(z)\,dy\,dz \\ &\quad + \iint_A \frac{\hat{a}_L(y, z)}{1-\Phi(y)}\big(\mathbf{1}_{\{Y\le y\}} - \Phi(y)\big)\varphi(y)\varphi(z)\,dy\,dz, \end{aligned}$$

and, for $A = (-\infty, y] \times (-\infty, z]$, it reduces to

$$w^L_{Y,Z}(y, z) = \left(Y - \frac{\varphi(Y)}{1-\Phi(Y)}\right)\mathbf{1}_{\{Y\le y,\, Z\le z\}} + \Phi(z)\frac{\varphi(Y\wedge y)}{1-\Phi(Y\wedge y)}.$$

As for the latter case,

$$\begin{aligned} w^H_{Y,Z}(y, z) &= w^{(\mathscr{T}_H,\hat{a})}_{Y,Z}((-\infty, y]\times(-\infty, z]) \\ &= Y\mathbf{1}_{\{Y\le y,\, Z\le z\}} + \frac{\sqrt{|Y|}\Phi(z)}{5}\big(|Y|^{5/2} - |y|^{5/2}\big)^+ \\ &\quad - \frac{u(Y)-1}{1-\cos\gamma}\Phi(z)\left[\varphi(y) - 2\int_{|y|}^{\infty}\varphi(s)\sqrt{s}\frac{s^{5/2}-|y|^{5/2}}{5}\,ds\right]. \end{aligned}$$

The goodness-of-fit is tested using the critical region

$$\max_{y,\,z\in\mathbf{R}} |w_n(y, z)| > c(\alpha)$$

with $w_n(y,z) = (1/\sqrt{n})\sum_{i=1}^n w^L_{Y_i,Z_i}(y,z)$ or $w_n(y,z) = (1/\sqrt{n})\sum_{i=1}^n w^H_{Y_i,Z_i}(y,z)$, respectively.

A conservative region of size α is obtained with $c(\alpha) = -\Phi^{-1}(\alpha/8)$, because of the well-known estimate

$$\mathbf{P}\Big\{\max_{y,\, z\in\mathbf{R}} w^V((-\infty, y], (-\infty, z]) \geq c\Big\} \leq 4\Phi(-c)$$

that holds for a V-Wiener process w associated to any probability measure V.

The general shape of the TEPs and how they are affected by changes in the samples is sketched in Figure 2, for the TEPs associated with the L-isometry.

The upper part of Figure 2 shows the graph $(y, z, w(y,z))$, $-3 < y, z < 3$ of the TEP for the *special sample* $\{(\Phi^{-1}(i/6), \Phi^{-1}(j/6)\}_{i,j=1,\dots,5}$, corresponding to the L-isometry. As in the one-dimensional pictures, we have introduced an arbitrary *special* (*two-dimensional*) *sample* with an empirical distribution abnormally close to its theoretical distribution, assumed to be the standard Gaussian.

The other two graphs in Figure 2 show the TEPs for the *shifted sample* $\{(\Phi^{-1}(i/6)+1, \Phi^{-1}(j/6)+1)\}_{i,j=1,\dots,5}$ and the *dispersed sample* $\{(2\Phi^{-1}(i/6), 2\Phi^{-1}(j/6))\}_{i,j=1,\dots,5}$, in the same domain.

The small diagrams in the left-hand side show the same graphs with the direction of vision changed to horizontal, in order to show the position of maxima and minima. The critical planes at $c(5\%)$ and $-c(5\%)$ are also shown.

A graph with points over $c(5\%)$ or under $-c(5\%)$ leads to rejecting the null hypothesis of goodness-of-fit, at a level smaller than 5%.

9.2. *Design of a test specially sensitive to changes in dispersion.* We choose now $F^{(\tau)}(x) = \Phi((1-\tau)x)$, and so

$$\frac{f^{(\tau)}(x)}{f_0(x)} = (1-\tau)\frac{\varphi(x-\tau x)}{\varphi(x)} = (1-\tau)(1+\tau x^2 + o(\tau)),$$

and $k(x) = x^2 - 1$, $\|k\|^2 = 2$.

EXAMPLE 4 (TEP associated to the L-isometry). The score function is

$$\begin{aligned}\hat{a}(x) &= \frac{1}{\sqrt{2}}\bigg[x^2 - 1 + \frac{1}{1-\Phi(x)}\int_{-\infty}^x (t^2-1)\varphi(t)\,dt\bigg]\\ &= \frac{1}{\sqrt{2}}\bigg[x^2 - 1 - \frac{x\varphi(x)}{1-\Phi(x)}\bigg]\end{aligned}$$

[see again (30)], and the TEP for the sample $\{X\}$ of size 1 is, consequently,

$$\begin{aligned}&\frac{1}{\sqrt{2}}\bigg[\bigg(X^2 - 1 - \frac{X\varphi(X)}{1-\Phi(X)}\bigg)\mathbf{1}_{\{X\leq x\}} - \int_{-\infty}^{X\wedge x}\frac{t^2-1-(t\varphi(t)/(1-\Phi(t))}{1-\Phi(t)}\varphi(t)\,dt\bigg]\\ &\quad= \frac{1}{\sqrt{2}}\bigg[(X^2-1)\mathbf{1}_{\{X\leq x\}} + \frac{x\varphi(x)}{1-\Phi(x)}\mathbf{1}_{\{x<X\}}\bigg].\end{aligned}$$

The shape of the TEPs for the same special samples used in previous one-dimensional diagrams is shown in Figure 3.

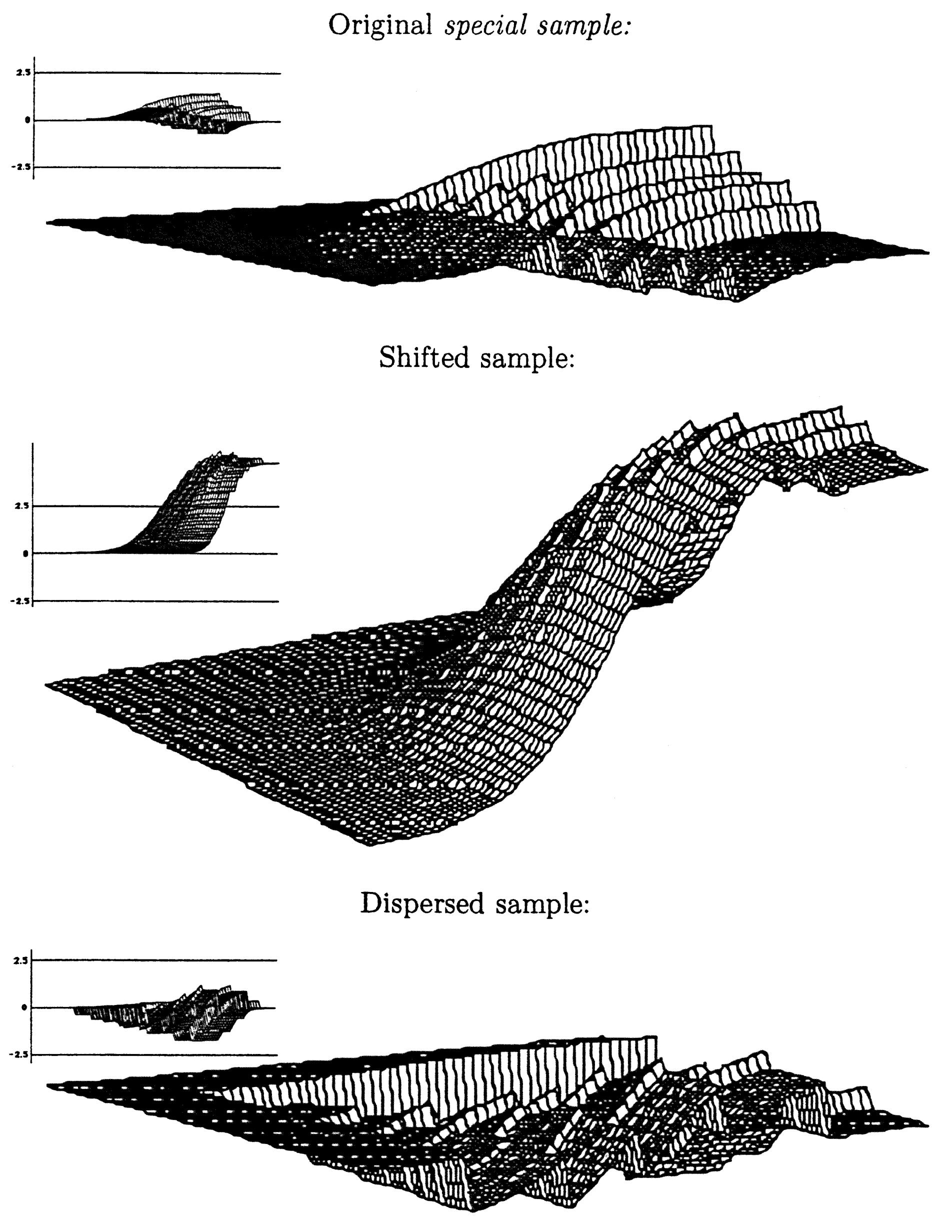

FIG. 2. *Graph of the TEPs for the two-dimensional special sample, corresponding to the L-isometry, and for the same sample after changes in position or dispersion.*

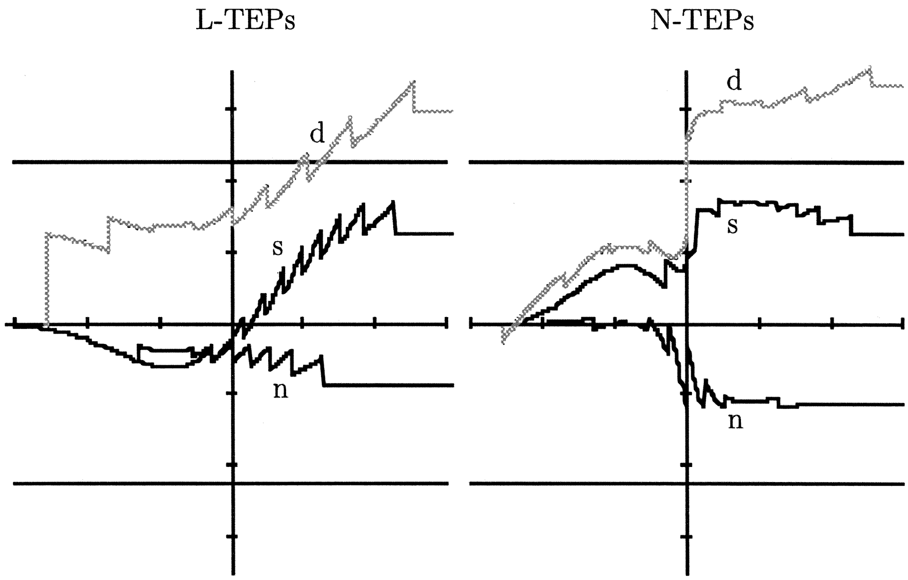

FIG. 3. *Responses to normal* (n), *shifted* (s) *and dispersed* (d) *samples, of the TEPs associated with L-isometry* (*L-TEPs*) *and with* N*-isometry* (*N-TEPs*), *optimized to detect changes in dispersion.*

EXAMPLE 5 (TEP associated to the N-isometry). From (28) and $k(x) = (x^2 - 1)/\sqrt{2}$, we evaluate the score function $\hat{a}$ numerically, and that evaluation is then used to compute the TEP $w_X(x) = \mathscr{T}_H(\hat{a}\mathbf{1}_{\{\cdot\leq x\}})(X)$ by numerical integration in (27).

The shape of the TEPs for the special samples is shown in the right-hand part of Figure 3.

The extension to $d = 2$ is similar to the one in Section 9.1.2.

Acknowledgments. Parts of this paper were written during visits of the authors to Departament d'Estadística, Universitat de Barcelona and Departamento de Estadística e Investigación Operativa, Universidad de Valladolid, Spain, respectively.

REFERENCES

CABAÑA, A. (1993). Transformations du pont brownien empirique et tests de type Kolmogorov–Smirnov. *C.R. Acad. Sci. Paris Sér. I. Math.* **317** 315–318.

CABAÑA, A. (1996). Transformations of the empirical process and Kolmogorov–Smirnov tests. *Ann. Statist.* **24** 2020–2035.

CABAÑA, A. and CABAÑA, E. M. (1996). Bridge-to-bridge transformations and Kolmogorov–Smirnov tests. *Comm. Statist. Theory Methods* **25** 227–234.

DONSKER, M. D. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Statist.* **23** 277–281.

EFRON, B. and JOHNSTON, I. (1990). Fisher's information in terms of hazard rate. *Ann. Statist.* **18** 38–62.

GROENEBOOM, P. and WELLNER, J. A. (1992). Information bounds and nonparametric maximum likelihood estimation. *DMV Sem.* **19**. Birkhäuser, Boston.

HÁJEK, J. and SIDÁK, Z. (1967). *Theory of Rank Tests.* Academic Press, New York.

KARATZAS, I. and SHREVE, S. (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. Springer, New York.

KHMALADZE, E. V. (1981). Martingale approach in the theory of goodness-of-fit. *Theory Probab. Appl.* **26** 240–257.

KHMALADZE, E. V. (1993). Goodness-of-fit problem and scanning innovation martingales. *Ann. Statist.* **21** 798–829.

LE CAM L. and YANG, G. L. (1990). *Asymptotics in Statistics. Some Basic Concepts*. Springer, New York.

MORÁN, M. and URBINA, W. (1996). Some properties of the Gaussian–Hilbert transform. Unpublished manuscript.

OOSTERHOFF, J. and VAN ZWET, W. R. (1979). A note on contiguity and Hellinger distance. In *Contributions to Statistics. Jaroslav Hájek Memorial Volume* (J. Jorečková, ed.) 157–166. Reidel, Dordrecht.

OSSIANDER, M. (1987). A central limit theorem under metric entropy with L_2 bracketing. *Ann. Probab.* **15** 897–919.

RITOV, J. and WELLNER, J. A. (1988). Censoring, martingales and the Cox model. *Contemp. Math.* **80** 191–220.

SANSONE, G. (1959). *Orthogonal Functions* (R. Courant, L. Bers and J. J. Stoker, eds.) *Pure and Applied Mathematics* **9**. Interscience, New York.

DEPARTAMENTO DE MATEMÁTICAS
INSTITUTO VENEZOLANO DE
INVESTIGACIONES CIENTÍFICAS
APARTADO 21827
CARACAS 1020-A
VENEZUELA
E-MAIL: acabana@ivic.ivic.ve

CENTRO DE MATEMÁTICA
FACULTAD DE CIENCIAS
UNIVERSIDAD DE LA REPÚBLICA
EDUARDO ACEVEDO 1139
11200 MONTEVIDEO
URUGUAY
E-MAIL: ecabana@cmat.edu.uy