The random projection method in goodness of fit for functional data

J.A. Cuesta-Albertos [a,*,1] E. del Barrio [b,1] R. Fraiman [c] C. Matrán [b,1]

[a] *Departamento de Matemáticas, Estadística y Computación, Universidad de Cantabria, Spain.*

[b] *Departamento de Estadística e Investigación Operativa. Universidad de Valladolid. Spain.*

[c] *Departamento de Matemáticas, Universidad de San Andrés, Argentina and Centro de Matemática, Universidad de la República, Uruguay*

Abstract

The possibility of considering random projections to identify probability distributions belonging to parametric families is explored. The results are based on considerations involving invariance properties of the family of distributions as well as on the random way of choosing the projections. In particular, it is shown that if a one-dimensional (suitably) randomly chosen projection is Gaussian, then the distribution is Gaussian.

In order to show the applicability of the methodology some goodness-of-fit tests based on these ideas are designed. These tests are computationally feasible through the bootstrap setup, even in the functional framework. The paper includes some simulations providing power comparisons of these projections-based tests with other available tests of normality, as well as to test the Black-Scholes model for a stochastic process.

Key words: Random projections, goodness of fit tests, families of distributions, gaussian distributions, Black-Scholes, stochastic processes.

---

[1] Research partially supported by the Spanish Ministerio de Educación y Ciencia, grant MTM2005-08519-C02-01 and 02 and by the Consejería de Educación y Cultura de la Junta de Castilla y León, grant PAPIJCL VA074/03.

* Corresponding author. Postal address: Departamento de Matemáticas, Estadística y Computación, Universidad de Cantabria, Avda. los Castros s.n., 39005 Santander, Spain. *Tel.*: (+34) 942201428. *Fax*: (+34) 942201402. Email address: `cuestaj@unican.es`

1 Introduction

The use of projections of very-high dimensional objects on randomly chosen subspaces is getting an increasing interest as a powerful tool in several applications of Mathematics. For instance, this idea is employed to obtain approximate algorithms in problems of high computational complexity (see, e.g., [14]) or, even, randomly chosen projections are starting to be employed as a tool to detect copyright violations on images posted in the Internet (see [9]).

Although in such those applications Statistics is at the basis of the procedure, a genuinely statistical analysis of the possibilities of the idea (including the design of methods based on it and a theoretical exploration of their power) has not been carried out yet. We are aware of some results in which random projections have been used to estimate mixtures of distributions (see [7] and [15]), but even these papers have not been written from a purely statistical point of view but rather from learning theory.

A first look at the problem with statistical motivation was [3]. To analyze to what extent random projections characterize a probability distribution, the closed cone

$$\mathbb{E}(P,Q) := \{h \in \mathbb{H} : P_h = Q_h\} \tag{1}$$

associated to two Borel distributions, P and Q, on a Hilbert space $\mathbb{H}$ is considered there. In (1), P_h denotes the marginal distribution of P along the h direction, i.e., the law of $\langle X, h\rangle$ if X is an $\mathbb{H}$-valued random element with law P and $\langle \cdot, \cdot \rangle$ is the scalar product in $\mathbb{H}$. In the sequel we will often refer to this distribution as the projection of P on h. With a similar meaning we will denote by P_V the projection of P on a linear subspace V.

A slight modification of one of the main results in [3] gives the following theorem (see Theorem 4.1 in [3]). It involves the property of a probability measure, P, being determined by its moments (in the sequel $P \in DM(\mathbb{H})$). A discussion of this property (including sufficient conditions like the so-called *Carleman condition*) can be found in [10] (see also Subsection 8.4 in [2]). A straightforward result is that an affine transformation of a probability measure determined by its moments is also determined by its moments.

Theorem 1.1 (Cuesta-Fraiman-Ransford) *Let $P \in DM(\mathbb{H})$ and Q be Borel probability measures on a separable Hilbert space, $\mathbb{H}$. Let μ be a non-degenerate Gaussian distribution on $\mathbb{H}$. Then $P = Q$ if and only if $\mu[\mathbb{E}(P,Q)] > 0$.*

This result was employed in [4] to obtain some consistent goodness-of-fit test to a fixed distribution.

We emphasize that Theorem 1.1 goes, somehow, in a direction opposite to that of older results in [5] where (p. 793) it is claimed that *"For many data sets, we show that most projections are nearly the same and approximately Gaussian"*. Below we include further comments on this point.

Our goal, in this paper, is to generalize Theorem 1.1 in some senses, to employ these results to obtain goodness-of-fit tests for families of distributions and to make a preliminary study to explore the possibilities of the technique. In particular, taking into account the above quotation from [5], we are particularly interested in seeing how this procedure works when applied to the design of goodness-of-fit tests for Gaussianity.

The proposed generalizations can be summarized as follows. First, in Section 2, we present a class of probability measures which can replace the Gaussian measure, μ, in Theorem 1.1. The possibility of choosing suitable (non-Gaussian) measures to increase the power against particular alternatives is still the object of current research.

In Section 3, the distribution P in Theorem 1.1 is replaced by a family of distributions. We consider two different cases. In the simplest situation one single random projection suffices to determine whether a distribution belongs to the family. This is the case for *invariant families* in the sense given in Definition 3.1. In particular, Theorem 3.6 states that if a randomly chosen projection of a distribution is Gaussian, then the distribution is Gaussian.

We want to stress the interest of this result. Let us assume that we are interested in knowing whether a given (multivariate) distribution, P, is Gaussian. Projection Pursuit techniques to reject this hypothesis are based on the fact that, if P is not Gaussian, then not every one-dimensional projection is Gaussian. However, since most projections of P are approximately Gaussian, an extensive search is required to find out one of the scarce directions in which the projection of P is clearly not Gaussian. On the other hand, according to Theorem 3.6, this search is not required because, if P is not Gaussian almost every projection of P is not Gaussian. Some simulations which are reported in Subsection 5.1 support this last point of view.

Subsection 3.2 focuses on non-invariant families. In this case, one projection is not sufficient to determine inclusion in the family. We show, though, that for a location-scale model with k-dimensional parameter, $(k+1)$ projections suffice. Notice that we are not considering here a projection on a $(k+1)$-dimensional subspace, but $(k+1)$ one-dimensional projections. Of course, it would suffice to handle the $(k+1)$-dimensional projection, but we want to remark that applying, for instance, a $(k+1)$-dimensional Kolmogorov-Smirnov test is, by far, more time consuming than taking the maximum of $(k+1)$ one-dimensional Kolmogorov-Smirnov tests.

Our results are applied in Section 4 to obtain goodness-of-fit tests to families of distributions. Unfortunately, when more than one projection is considered, the goodness-of-fit test is often not distribution free; this is the case for the Kolmogorov-Smirnov statistic that we consider in Section 5. We propose to apply bootstrap to estimate the distribution of the test under the null hypothesis. The analysis to prove that the bootstrap works in this setting is of a technical character and can be skipped by the readers interested only in the applications of the methodology. Therefore it has been included as an appendix

Finally, in Section 5, we present some simulations to give a general idea about the power of the proposed procedure under several conditions.

We will incorporate new notation as it becomes necessary. However, in addition to that already presented we use throughout the paper the following. $\mathbb{H}$ will be a separable Hilbert space with norm $\|\cdot\|$ and scalar product $\langle\cdot,\cdot\rangle$. All the random elements (r.e.'s) will be assumed to be defined on the same rich enough probability space $(\Omega, \mathcal{A}, \nu)$. For such a r.e., X, $\mathcal{L}(X)$ will denote its law and $X =_d Y$ will mean $\mathcal{L}(X) = \mathcal{L}(Y)$. ℓ_n will denote Lebesgue measure on $\mathbb{R}^n$.

If $F \subset \mathbb{H}$, then $\text{span}(F)$ will be the closed linear subspace spanned by F. $\{v_n\}_{n=1}^{\infty}$ will denote a generic orthonormal basis of $\mathbb{H}$ and $V_n = \text{span}(\{v_1, \dots, v_n\})$. Given any subspace, V, of $\mathbb{H}$ we will write $V^{\perp}$ for its orthogonal complement. If X is an $\mathbb{H}$-valued r.e., then X_V will denote the projection of X on the subspace V. $\mathcal{M}_d$ will denote the set of $d \times d$ positive definite matrices.

We will employ de Finetti's notation and, then $P(f)$ will denote the expectation of f with respect to the probability P. In agreement with this notation, when the probability belongs to a family $\{P(\cdot,\theta), \theta \in \Theta\}$, we will use $P(f,\theta)$ to denote the expectation of f with respect to the probability $P(\cdot,\theta)$.

2 Determination of a distribution

We begin this section by defining a key concept in this paper.

Definition 2.1 *We will say that a Borel probability measure on $\mathbb{H}$, μ, is probability determining (in short, $\mu \in PD(\mathbb{H})$) if for every $P \in DM(\mathbb{H})$ and Q Borel probability measures on $\mathbb{H}$ we have that $\mu[\mathbb{E}(P,Q)] > 0$ implies $Q = P$.*

For further reference we include the following easy property of probability determining distributions.

Lemma 2.2 *If $\mu \in PD(\mathbb{H})$, then $\mu(K) = 0$ for every hyperplane $K \subset \mathbb{H}$.*

If $\mathbb{H}$ is finite-dimensional, the following consequence of Corollary 3.2 in [3] gives a sufficient condition for the $PD(\mathbb{H})$ property.

Proposition 2.3 *If $\mathbb{H} = \mathbb{R}^d$ and μ is absolutely continuous with respect to the Lebesgue measure, then $\mu \in PD(\mathbb{H})$.*

For infinite dimensional spaces, Theorem 4.1 in [3] shows that every non-degenerate Gaussian measure on a Hilbert space is probability determining. A generalization of this result is given in Theorem 2.5 below. Its proof relies on the use of the well-known Lemma 2.4, reproduced from [2] (Corollary 2, p. 231) and the argument in the proof of Theorem 4.1 in [3]. It is included here for the sake of completeness.

Lemma 2.4 *Let X, Y be two random variables taking values on the measurable spaces $(\mathcal{X}, \sigma_{\mathcal{X}})$, $(\mathcal{Y}, \sigma_{\mathcal{Y}})$ respectively, where $(\mathcal{X}, \sigma_{\mathcal{X}})$ is a Borel space. Let $\phi(x, y)$ be a measurable map on the product space $(\mathcal{X}, \sigma_{\mathcal{X}}) \times (\mathcal{Y}, \sigma_{\mathcal{Y}})$ such that $E|\phi(X, Y)| < \infty$.*

If $P(\cdot/Y = y)$ is a regular conditional distribution for X given $Y = y$, then

$$E[\phi(X, Y)/Y = y] = \int \phi(x, y) P(dx/Y = y), \quad a.s.$$

Theorem 2.5 *Let U be an $\mathbb{H}$-valued r.e. satisfying that for some orthonormal basis, $\{v_n\}_n$, $\nu\left\{U_{V_n^\perp} = 0\right\} = 0$, for every $n \geq 2$, and that the conditional distribution of U_{V_n} given $U_{V_n^\perp}$ is absolutely continuous with respect to the Lebesgue measure, ℓ_n. Then $\mathcal{L}(U) \in PD(\mathbb{H})$.*

PROOF.- Let $\mu = \mathcal{L}(U)$ and let $P \in DM(\mathbb{H})$, Q be two Borel probabilities on $\mathbb{H}$ such that $\mu[\mathbb{E}(P, Q)] > 0$. Fix $n \geq 2$. From Lemma 2.4, we have that

$$\begin{aligned} 0 < \mu[\mathbb{E}(P, Q)] &= \int_{V_n^\perp} \nu\left[U \in \mathbb{E}(P, Q) \big/ U_{V_n^\perp} = z\right] d\mu_{V_n^\perp}(z) \\ &= \int_{V_n^\perp} \nu\left[(U_{V_n}, z) \in \mathbb{E}(P, Q) \big/ U_{V_n^\perp} = z\right] d\mu_{V_n^\perp}(z) \\ &= \int_{V_n^\perp} \nu\left[U_{V_n} \in \mathbb{E}^z \big/ U_{V_n^\perp} = z\right] d\mu_{V_n^\perp}(z), \end{aligned} \tag{2}$$

where $\mathbb{E}^z := \{h_{V_n} : h \in \mathbb{E}(P, Q) \text{ and } h_{V_n^\perp} = z\}$ is the z-section of $\mathbb{E}(P, Q)$ and the last equality comes from the fact that the sets $\{(U_{V_n}, z) \in \mathbb{E}(P, Q)\}$ and $\{U_{V_n} \in \mathbb{E}^z\}$ coincide.

From (2), taking into account the hypothesis on the distribution of $U_n^\perp$ and that the conditional distribution of U_{V_n} given $U_{V_n^\perp}$ is absolutely continuous with respect to ℓ_n, there exists $z_n \in V_n^\perp - \{0\}$ such that $\ell_n(\mathbb{E}^{z_n}) > 0$.

Since $\mathbb{E}(P,Q)$ is a cone, it also follows that $\ell_n(\mathbb{E}^{tz_n}) > 0$ for each $t \in \mathbb{R} - \{0\}$. Therefore, $\ell_{n+1}(\mathbb{E}(P,Q) \cap \tilde{V}_n) > 0$, where $\tilde{V}_n = \text{span}(V_n \cup \{z_n\})$.

By Proposition 2.3, we deduce that $P_{\tilde{V}_n} = Q_{\tilde{V}_n}$. In particular, since $V_n \subset \tilde{V}_n$, we obtain $P_{V_n} = Q_{V_n}$. Taking into account that the finite dimensional distributions determine the joint distribution, we have that $P = Q$ and so $\mu \in PD(\mathbb{H})$. $\square$

In this paper the term *dissipative* will make reference to the fact that a random element, U, (or indistinctly its law $\mathcal{L}(U)$) satisfies the hypotheses in Theorem 2.5.

Notice that in Theorem 2.5, the orthogonal basis may depend on U and, in consequence, this result includes Theorem 1.1 as a particular case. However, it also covers non-Gaussian distributions like the one we propose in the following example.

Example 2.6 Let $\{U_n\}$ be a sequence of real random variables with the following joint distribution.

- The distribution of U_1 is uniform on $[0, 1]$.
- For $m \geq 1$, the distribution of U_{n+1} given $\{U_1 = u_1, ..., U_n = u_n\}$ is uniform on $\left[0, (1 - u_1^2 - ... - u_n^2)^{1/2}\right]$.

Elementary computations show that $\sum_{n\geq 1} U_n^2 = 1,\ \nu$-a.s. Thus, if $\{v_n\}_{n=1}^{\infty}$ is an orthonormal basis of the Hilbert space $\mathbb{H}$, and we define

$$U := \sum_{n\geq 1} U_n v_n,$$

we have that U belongs to the unit sphere in $\mathbb{H}$, ν-a.s. However, it is obvious that the distribution of U is dissipative. $\square$

3 Families of distributions

In this section we analyze the problem of determining families of distributions through randomly chosen projections. We will present cases in which only one projection still suffices and some others in which it does not.

A useful criterion in relation with our goal is invariance that has received considerable attention mainly in connection to Decision Theory (see e.g. [8] for a general treatment regarding point-wise estimation). It involves a group of transformations, G, acting on the sample space, $\mathbb{H}$, that induces a new group,

$\tilde{G}$, acting on the family of probabilities. For $\tilde{g} \in \tilde{G}$ and P, a Borel probability on $\mathbb{H}$, $\tilde{g}P$ will denote the action of $\tilde{g}$ on P.

Definition 3.1 *Let $\tilde{G}$ be a group of transformations on the Borel probabilities on $\mathbb{H}$ and let $\mathcal{P}$ be a family of Borel probabilities on $\mathbb{H}$. We will say that $\mathcal{P}$ is $\tilde{G}$-invariant if for every $\tilde{g} \in \tilde{G}$ and every $P \in \mathcal{P}$, then $\tilde{g}P \in \mathcal{P}$.*

Some usual groups of transformations in Statistics include that of changes in location, $\tilde{G}_1 = \{\tilde{h} : h \in \mathbb{H}\}$, where $\tilde{h}P = \mathcal{L}(X + h)$ for any X such that $\mathcal{L}(X) = P$, or that of changes in scatter, $\tilde{G}_2 = \{\tilde{\Sigma} : \Sigma \in \mathcal{M}_d\}$, where $\tilde{\Sigma}P = \mathcal{L}(\Sigma X)$ for any X such that $\mathcal{L}(X) = P$. A $\tilde{G}_1$-invariant family will be called *l-invariant*, whereas a $\tilde{G}_2$-invariant family will be called *s-invariant*.

We begin by showing that invariant families are determined by just one randomly chosen projection. In particular, the results in this section show that randomly chosen one-dimensional projections determine Gaussian distributions. In Subsection 3.2 we will show that if we are in a non-invariant location-scale family and the parameter space has dimension k then $(k + 1)$ one-dimensional projections suffice to determine the family.

3.1 Invariant families

First we focus on the finite dimensional case.

Proposition 3.2 *Let $\mathbb{H} = \mathbb{R}^d$ and set $\mathcal{P} = \{\mathcal{L}(\Sigma X + h) : h \in \mathbb{H},\ \text{and}\ \Sigma \in \mathcal{M}_d\}$, where $\mathcal{L}(X) \in DM(\mathbb{H})$. Let Y be an $\mathbb{H}$-valued r.e. and define*

$$A = \{h \in \mathbb{H} :\ \exists g_h \in \mathbb{H}, M_h \in \mathcal{M}_d\ s.t.\ \langle Y, h\rangle =_d \langle M_h X + g_h, h\rangle\}.$$

Let $\mu \in PD(\mathbb{H})$. If $\mu(A) > 0$, then $\mathcal{L}(Y) \in \mathcal{P}$.

PROOF.- Without loss of generality we can assume that X is centered with identity covariance matrix. By hypothesis, for every $h \in A$ there exist $m_h \in \mathbb{H}$ and $M_h \in \mathcal{M}_d$ such that

$$\langle Y, h\rangle =_d \langle M_h X, h\rangle + \langle m_h, h\rangle. \tag{3}$$

From Lemma 2.2 it is obvious that A contains k independent vectors, and, then we have that $E\|Y\|^2 < \infty$. Call Σ the covariance matrix of Y and m its mean.

Obviously, we can take $M_h = \Sigma$ and $m_h = m$ in (3) and, in consequence, we have that $A = \mathbb{E}(\mathcal{L}(\Sigma X + m), \mathcal{L}(Y))$, which implies that

$$\mu\left[\mathbb{E}(\mathcal{L}(\Sigma X + m), \mathcal{L}(Y))\right] > 0.$$

From this, taking into account that $\mathcal{L}(\Sigma X + m) \in DM(\mathbb{H})$ and that $\mu \in PD(\mathbb{H})$, we have that $\mathcal{L}(Y) = \mathcal{L}(\Sigma X + m)$, and in consequence $\mathcal{L}(Y) \in \mathcal{P}$. $\square$

Some corollaries are easily obtained from Proposition 3.2. The first two state that the result in this proposition holds for l- or s-invariant families.

Corollary 3.3 *Let $\mathbb{H} = \mathbb{R}^d$ and set $\mathcal{P} = \{\mathcal{L}(X + h) : h \in \mathbb{H}\}$, where $\mathcal{L}(X) \in DM(\mathbb{H})$. Let Y be an $\mathbb{H}$-valued r.e. and define*

$$A = \{h \in \mathbb{H} :\, \exists g_h \in \mathbb{H} \text{ s.t. } \langle Y, h\rangle =_d \langle X + g_h, h\rangle\}.$$

Let $\mu \in PD(\mathbb{H})$. If $\mu(A) > 0$, then $\mathcal{L}(Y) \in \mathcal{P}$.

Corollary 3.4 *Let $\mathbb{H} = \mathbb{R}^d$ and set $\mathcal{P} = \{\mathcal{L}(\Sigma X) : \Sigma \in \mathcal{M}_d\}$, where $\mathcal{L}(X) \in DM(\mathbb{H})$. Let Y be an $\mathbb{H}$-valued r.e. and define*

$$A = \{h \in \mathbb{H} :\, \exists M_h \in \mathcal{M}_d \text{ s.t. } \langle Y, h\rangle =_d \langle M_h X, h\rangle\}.$$

Let $\mu \in PD(\mathbb{H})$. If $\mu(A) > 0$, then $\mathcal{L}(Y) \in \mathcal{P}$.

Our third corollary is related to elliptical families of distributions. Recall that an elliptical family on $\mathbb{R}^d$ consists of all the probability distributions with density functions

$$f_{m,\Sigma}(x) = C(\Sigma) f_0\left((x - m)'\Sigma^{-1}(x - m)\right),\ x \in \mathbb{H},$$

for $m \in \mathbb{H}$, $\Sigma \in \mathcal{M}_d$, where $C(\Sigma)$ is a normalizing constant and f_0 a suitable function. Note that an elliptical family is both l- and s-invariant.

Corollary 3.5 *Let $\mathbb{H} = \mathbb{R}^d$ and let $\mathcal{P}$ be an elliptical family on $\mathbb{H}$, such that some (hence all) its distributions are in $PD(\mathbb{H})$. For an $\mathbb{H}$-valued r.e. Y define*

$$A = \{h \in \mathbb{H} :\, \mathcal{L}(\langle Y, h\rangle) = P_h,\ \text{for some } P \in \mathcal{P}\}.$$

Let $\mu \in PD(\mathbb{H})$. If $\mu(A) > 0$, then $\mathcal{L}(Y) \in \mathcal{P}$.

Since the family of Gaussian distributions is elliptical, for this particular case Corollary 3.5 means that if the set $A = \{h \in \mathbb{H} : \mathcal{L}(\langle Y, h\rangle)$ is Gaussian$\}$ satisfies $\mu(A) > 0$, then Y is Gaussian.

The extension of these results to the infinite dimensional case will focus on the Gaussian case.

Theorem 3.6 *Let μ be a dissipative measure on $\mathbb{H}$. If Y is an $\mathbb{H}$-valued r.e. and the set*

$$A = \{h \in \mathbb{H} :\, \mathcal{L}(\langle Y, h\rangle) \text{ is Gaussian}\}$$

has positive μ-measure, then Y is Gaussian.

PROOF.- Let $\{v_n\}_n$ be an orthonormal basis as in Theorem 2.5. Fix $n \in \mathbb{N}$. For $z \in V_n^\perp$ consider the set $A^z = \{h_{V_n} : h \in A \text{ and } h_{V_n^\perp} = z\}$, the z-section of A.

By the same argument as in the proof of Theorem 2.5 we obtain that there exists $z_n \in V_n^\perp$ such that $\ell_{n+1}(A \cap \tilde{V}_n)$ is strictly positive, where $\tilde{V}_n = \text{span}(V_n \cup \{z\})$.

Taking into account that all Gaussian distributions on $\tilde{V}_n$ are projections on this subspace of the Gaussian distributions on $\mathbb{H}$, and Corollary 3.5 we obtain that the projection of Y on $\tilde{V}_n$ is Gaussian. Thus, we have that, for every $n \in \mathbb{N}$, the distribution of Y_{V_n} is Gaussian.

Moreover, since Y is the limit of the sequence $\{Y_{V_n}\}_n$ then $\mathcal{L}(Y)$ is also Gaussian. $\square$

The following corollaries are consequences of similar arguments and their proofs are hence omitted.

Corollary 3.7 *Let $m \in \mathbb{H}$. Consider the set $\mathcal{P}$ of all Gaussian distributions on $\mathbb{H}$ with mean m and let μ be a dissipative probability distribution on $\mathbb{H}$. If Y is an $\mathbb{H}$-valued r.e. such that the set*

$$A = \{h \in \mathbb{H} : \ \mathcal{L}(\langle Y, h\rangle) = P_h \ \ \textit{for some } P \in \mathcal{P}\}$$

is of positive μ-measure, then $\mathcal{L}(Y) \in \mathcal{P}$.

Corollary 3.8 *Let $\mathcal{P}$ be the set of all Gaussian distributions on $\mathbb{H}$ with a given covariance operator Σ. Let μ be a dissipative probability distribution on $\mathbb{H}$. If Y is an $\mathbb{H}$-valued r.e. such that the set*

$$A = \{h \in \mathbb{H} : \ \mathcal{L}(\langle Y, h\rangle) = P_h \ \ \textit{for some } P \in \mathcal{P}\}$$

is of positive μ-measure, then $\mathcal{L}(Y) \in \mathcal{P}$.

To end this subsection we give the generalization of Corollary 3.3 to l-invariant families on infinite-dimensional spaces.

Proposition 3.9 *Let $\mathcal{P} = \{\mathcal{L}(X+g) : g \in \mathbb{H}\}$, where X is an $\mathbb{H}$-valued r.e. such that $\mathcal{L}(X) \in DM(\mathbb{H})$. Let μ be a dissipative probability distribution on $\mathbb{H}$. If Y is an $\mathbb{H}$-valued r.e. and the set*

$$A = \{h \in \mathbb{H} : \ \langle Y, h\rangle =_d \langle X + g_h, h\rangle, \ \textit{for some } g_h \in \mathbb{H}\}$$

is of positive μ-measure, then $\mathcal{L}(Y) \in \mathcal{P}$.

PROOF.- The same proof as in Theorem 3.6 allows to obtain that, for every $n \in \mathbb{N}$, there exists $h_n \in V_n$ such that $Y_{V_n} =_d (X_{V_n} + h_n)$ and, in consequence, there exists a (real) random variable X_n^*, with the same distribution as $\langle X_{V_n}, e_n \rangle$, such that

$$\langle Y, e_n \rangle = X_n^* + \langle h_n, e_n \rangle, \ \nu - \text{almost surely.} \tag{4}$$

Since $\langle X_{V_n}, e_n \rangle$ and X_n^* have the same distribution, we have that the (real) random variable $Z = \sum_{n=1}^{\infty} |X_n^*|^2$ has finite expectation; in consequence, it is almost surely finite and $X^* = \sum_{n=1}^{\infty} X_n^* e_n$ is an $\mathbb{H}$-valued r.e.

Moreover, by Parseval's identity, $Y = \sum_{n=1}^{\infty} \langle Y, e_n \rangle e_n \in \mathbb{H}$. Thus, from (4), we have that $h_Y := \sum_{n=1}^{\infty} \langle h_n, e_n \rangle e_n \in \mathbb{H}$ and, in consequence, we can take $h_n = (h_Y)_{V_n}$ for every $n \in \mathbb{N}$.

The proof ends by taking into account that $\{X_{V_n}\}$, $\{Y_{V_n}\}$ and $(h_Y)_{V_n}$ converge to X, Y and h_Y in norm and, in consequence, Y has the same distribution as $X + h_Y$. $\square$

3.2 Non-invariant families

First we include two examples and a proposition in which a family of distributions is not determined by one-dimensional (or k-dimensional, with $k \geq 1$ fixed) projections. We build these examples by resorting to non-invariant families.

Example 3.10 Let v, w be two (non-random) not collinear vectors in $\mathbb{H}$, not necessarily infinite-dimensional, and for a given r.e., X, on $\mathbb{H}$, let us consider the parametric family of distributions, $\mathcal{P} = \{\mathcal{L}(X + \theta v) : \theta \in \mathbb{R}\}$. Set $Y := X + w$. Obviously $\mathcal{L}(Y) \notin \mathcal{P}$. However, if $h \in \mathbb{H}$ satisfies that $\langle h, v \rangle \neq 0$, it is obvious that

$$\langle Y, h \rangle = \langle X, h \rangle + \ \theta \langle v, h \rangle,$$

where $\theta = \frac{\langle w,h \rangle}{\langle v,h \rangle}$ and, therefore, the projection of the distribution of Y on the linear subspace generated by h belongs to the family of projections of $\mathcal{P}$ on this subspace. Thus one random projection is not enough to determine $\mathcal{P}$ as long as the distribution used to produce the random direction, h, satisfies that

$$\mu\{h : \ \langle v, h \rangle \neq 0\} > 0$$

(which is satisfied by every distribution in $PD(\mathbb{H})$ because of Lemma 2.2). $\square$

We can generalize Example 3.10 as follows.

Proposition 3.11 *Let $\mu \in PD(\mathbb{H})$. Consider a fixed value $n \in \mathbb{N}$ and let V_n be an n-dimensional subspace of $\mathbb{H}$. Let X be an $\mathbb{H}$-valued r.e. and let*

$$\mathcal{P} := \{\mathcal{L}(X+g) : g \in V_n\}.$$

Let $w \in \mathbb{H} - V_n$ and define the r.e. $Y := X + w$. If $k \leq n$, for μ^k-a.e. $(h_1, ..., h_k) \in \mathbb{H}^k$, there exists $g_0 \in V_n$ such that the distribution of the projection of Y on the subspace generated by $h_1, ..., h_k$ coincides with that of the projection of $X + g_0$.

PROOF.- Let $v_1, ..., v_n$ be an orthogonal basis of V_n. Because of Lemma 2.2 we have that

$$\mu^k\{(h_1, ..., h_k) : \langle h_i, v_i\rangle \neq 0,\ i = 1, ..., k\} = 1.$$

Then, for μ^k-almost every $(h_1, ..., h_k)$, the linear system

$$\langle w, h_i\rangle = \sum_{j=1}^{i} \lambda_j \langle v_j, h_i\rangle,\ i = 1, ..., k,$$

admits an unique solution $\lambda_1^0, ..., \lambda_k^0$ and the vector $g_0 = \sum_{i=1}^{k} \lambda_i^0 v_i$ is an element in V_n which satisfies the desired property. □

Example 3.12 Let $\{g_\theta : \theta \in \Theta\} \subset \mathbb{H}$. Notice that the argument in Example 3.10 holds under the following assumptions:

(1) $\mathcal{P} = \{\mathcal{L}(X + g_\theta) : \theta \in \Theta\}$, where X is an $\mathbb{H}$-valued r.e.
(2) $\{g_\theta : \theta \in \Theta\} \neq \mathbb{H}$
(3) For μ-almost every $h \in \mathbb{H}$, $\mathbb{R}$ coincides with the image set $\{\langle g_\theta, h\rangle : \theta \in \Theta\}$ □

Thus, non-invariant location families are not determined by just one random projection. However, we will show that, in some cases, if the parameter which defines the family of distributions has finite dimension n, then the family is determined by $k = (n + 1)$ one-dimensional projections.

We analyze first location families. The proof goes in two steps. We begin by dealing with the case when the candidate distribution is also a shift of the parent distribution. Then we apply Proposition 3.9 to extend this result to general distributions.

Proposition 3.13 *Let $\mu \in PD(\mathbb{H})$ and consider V_n, a proper subspace of $\mathbb{H}$,* *with $dim(V_n) = n \in \mathbb{N}$. Let X be an $\mathbb{H}$-valued r.e. and let*

$$\mathcal{P} := \{\mathcal{L}(X + g) : g \in V_n\}.$$

Let $w \in \mathbb{H} - V_n$ and let us consider the random element $Y := X + w$. Then, for μ^{n+1}-a.e. $(h_1, ..., h_{n+1}) \in \mathbb{H}^{n+1}$, there exists no $g_0 \in V_n$ such that the distributions of the projections of Y on the $(n+1)$ one-dimensional subspaces generated by $h_1, ..., h_{n+1}$ coincide with those of the projections of $X + g_0$.

PROOF.- Let $v_1, ..., v_n$ be an orthogonal basis of V_n. If $h_1, ..., h_{n+1} \in \mathbb{H}$ satisfy that there exists $g_0 = \sum_{j=1}^n \lambda_j^0 v_j \in V_n$ such that

$$\langle Y, h_i \rangle =_d \langle X, h_i \rangle + \langle g_0, h_i \rangle, i = 1, ..., n + 1,$$

then it is also satisfied that $\lambda_1^0, ..., \lambda_n^0$ is a solution of the linear system

$$\langle w, h_i \rangle = \sum_{j=1}^{n} \lambda_j \langle v_j, h_i \rangle, \ \ i = 1, ..., n + 1. \tag{5}$$

However, if $h_1, ..., h_{n+1}$ is a random sample taken from μ, Lemma 2.2 implies that there is probability zero for the event that h_{n+1} belongs to the subspace generated by $h_1, ..., h_n$, thus the linear system (5) has no solution with μ^{n+1}-probability one. $\square$

Proposition 3.14 *Let μ be a dissipative distribution on $\mathbb{H}$ and consider V_n, a proper n-dimensional subspace of $\mathbb{H}$. Let X be an $\mathbb{H}$-valued r.e. such that $\mathcal{L}(X) \in DM(\mathbb{H})$, and let*

$$\mathcal{P} := \{\mathcal{L}(X + g) : g \in V_n\}.$$

Let Y be an $\mathbb{H}$-valued r.e. such that $\mathcal{L}(Y) \notin \mathcal{P}$. Then, for μ^{n+1}-a.e. $(h_1, ..., h_{n+1}) \in \mathbb{H}^{n+1}$, there exists no $g_0 \in V_n$ such that the distributions of the projections of Y on the $(n+1)$ one-dimensional subspaces generated by $h_1, ..., h_{n+1}$ coincide with those of the projections of $X + g_0$.

PROOF.- Let us assume that the result is false. In particular we have that

$$\mu\{h \in \mathbb{H} : \langle Y, h \rangle =_d \langle X, h \rangle + \langle v, h \rangle, \text{ for some } v \in \mathbb{H}\} > 0.$$

From Proposition 3.9 there exists $w \in \mathbb{H}$ such that $Y + w =_d X$. Then the result follows from Proposition 3.13. $\square$

Finally, we analyze location-scale families. Thus, we will assume that X is an $\mathbb{H}$-valued r.e., such that $\mathcal{L}(X) \in DM(\mathbb{H})$ and that $\mathcal{P} := \{\mathcal{L}(sX + g) : g \in V \text{ and } s \in \mathbb{R}\}$, where V is a subspace of $\mathbb{H}$ which is not necessarily nonempty and not necessarily different from $\mathbb{H}$.

The next example shows that, even in the case $V = \emptyset$, one projection is not enough to identify this model.

Example 3.15 Let $\mathbb{H} = \mathbb{R}^2$. Let $X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$ centered Gaussian r.e.'s with independent marginals, and $\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = \mathrm{Var}(Y_1) = 1$ and $\mathrm{Var}(Y_2) = 2$.

It is easy to check that if μ is a distribution absolutely continuous with respect to the Lebesgue measure, then for μ-almost every $h \in \mathbb{R}^2$ there exists $s \in \mathbb{R}$ such that $s\langle X, h\rangle$ and $\langle Y, h\rangle$ are identically distributed. $\Box$

Now we show that, if dim $(V) = n < \infty$ then $(n+2)$ independent projections are enough to determine the family $\mathcal{P}$.

Proposition 3.16 *Let μ be a dissipative distribution on $\mathbb{H}$ and let V_n be an n-dimensional subspace of $\mathbb{H}$, where the dimension of $\mathbb{H}$ is greater than or equal to $(n+2)$. Let X, Y be $\mathbb{H}$-valued r.e.'s such that $\mathcal{L}(X) \in DM(\mathbb{H})$, and let*

$$\mathcal{P} := \{\mathcal{L}(sX + g) : s \in \mathbb{R} \text{ and } g \in V_n\}.$$

Then, if $\mathcal{L}(Y) \notin \mathcal{P}$, for μ^{n+2}-a.e. $(h_1, ..., h_{n+2}) \in \mathbb{H}^{n+2}$, there exists no $g_0 \in V_n$ and $s \in \mathbb{R}$ such that the distributions of the projections of Y on the $(n+2)$ one-dimensional subspaces generated by $h_1, ..., h_{n+2}$ coincide with those of the projections of $(sX + g_0)$.

PROOF.- To avoid trivialities let us assume that X is not ν-a.s. constant. In this case, Lemma 2.2 implies that, ν-a.s., the random variable $\langle X, h_1\rangle$ is not a.s. constant and s is determined by the variances of the random variables $\langle X, h_1\rangle$ and $\langle Y, h_1\rangle$. In other words, there exists at most one value for s which satisfies the equation

$$\langle Y, h_1\rangle =_d s\langle X, h_1\rangle + \langle g, h_1\rangle.$$

Let us denote this solution by s_{h_1} and set $X_1 = s_{h_1}X$. Obviously $\mathcal{L}(X_1) \in DM(\mathbb{H})$ and, by Proposition 3.14, there exists no $g_0 \in V_n$ such that the distributions of the projections of Y on the $(n+1)$ one-dimensional subspaces generated by $h_2, ..., h_{n+2}$ coincide with those generated by the projections of $(X_1 + g_0)$ which concludes the proof. $\Box$

From Proposition 3.9, using a similar argument to that in Proposition 3.16 it is possible to prove the following result.

Proposition 3.17 *Let $\mathcal{P} = \{\mathcal{L}(sX + g) : s \in \mathbb{R},\ g \in \mathbb{H}\}$, where X is an $\mathbb{H}$-valued r.e. such that $\mathcal{L}(X) \in DM(\mathbb{H})$. Assume that μ is a dissipative dis-*

tribution on $\mathbb{H}$. Let Y be an $\mathbb{H}$-valued r.e. and consider the set

$$A = \{(h_1, h_2) \in \mathbb{H}^2 : \exists\, g \in \mathbb{H} \text{ and } s \in \mathbb{R} \text{ s.t. } \langle Y, h_i\rangle =_d \langle sX, h_i\rangle + \langle g, h_i\rangle,\ i = 1, 2\}.$$

If $\mu^2(A) > 0$, then $\mathcal{L}(Y) \in \mathcal{P}$.

4 Goodness-of-fit tests

This section is devoted to the application of the above results to the goodness-of-fit setup. Rather than being exhaustive we have decided to focus on some simple examples to show how the random projection method could be used for testing fit to a parametric family of distributions. This could be easily modified to cover more general situations.

We will assume that, given i.i.d. $\mathbb{H}$-valued data, $X_1, \ldots, X_n$, with $\mathcal{L}(X_i) = P$, we are interested in testing the null hypothesis

$$H_0\colon P \in \mathcal{P} := \{P(\cdot, \theta) : \theta \in \Theta\}$$

against the alternative $P \notin \mathcal{P}$, where $\mathcal{P}$ is either an invariant family as in Subsection 3.1 or an non-invariant family like the one considered in Proposition 3.16. In any case the family will be determined by k μ-generated random projections, $h_1, \ldots, h_k$, meaning that μ^k-a.e.

$$(P_{h_1}, \ldots, P_{h_k}) \in \mathcal{P}_{h_1,\ldots,h_k} := \{(P_{h_1}(\cdot, \theta), \ldots, P_{h_k}(\cdot, \theta)) : \theta \in \Theta\} \tag{6}$$

iff $P \in \mathcal{P}$ (k can be taken to be 1 for invariant families, whereas for non-invariant families as in Proposition 3.16 it should be no less than $n + 2$).

We can try to use univariate Kolmogorov-Smirnov metrics to measure the deviation of a probability measure P with respect to $\mathcal{P}$. To be precise, for probability measures on the real line Q_1, Q_2, call $d(Q_1, Q_2) := \sup_x |Q_1(-\infty, x] - Q_2(-\infty, x]|$. With this notation (6) means that $P = P(\cdot, \theta)$ iff μ^k-a.e.

$$\max_{i=1,\ldots,k} d(P_{h_i}, P_{h_i}(\cdot, \theta)) = 0. \tag{7}$$

This suggest that we base our test on

$$D_n := \max_{i=1,\ldots,k} \sqrt{n}\, d((\mathbb{P}_n)_{h_i}, P_{h_i}(\cdot, \theta)),$$

the maximum of k univariate Kolmogorov-Smirnov statistics, where $\mathbb{P}_n$ denotes the empirical measure based on $X_1, \ldots, X_n$. From (7) we get that $D_n \to$

∞ if $P \notin \mathcal{P}$, while if $P = P(\cdot, \theta) \in \mathcal{P}$ then the multivariate version of Donsker's Theorem gives that $D_n \to_w D$, D being a finite random variable.

Obviously, in order to make the test useful, we need to know the distribution of D_n under the assumption $P = P(\cdot, \theta)$. If $k = 1$ and P has continuous marginals this is trivial because, in this case, the Kolmogorov-Smirnov test is distribution free. Regrettably, this is no longer true if $k > 1$. An added difficulty arises from the fact that θ must be estimated and we should replace D_n by

$$\hat{D}_n = \max_{i=1,\ldots,k} \sqrt{n} d((\mathbb{P}_n)_{h_i}, P_{h_i}(\cdot, \hat{\theta}_n)), \tag{8}$$

for some suitable estimator $\hat{\theta}_n = \theta(X_1, \ldots, X_n)$.

We propose to apply the bootstrap to approximate the null distribution of $\hat{D}_n$ as follows. Generate $X_1^*, \ldots, X_n^*$ i.i.d. data from distribution $P(\cdot, \hat{\theta}_n)$, the bootstrap sample. Call $\mathbb{P}_n^*$ the empirical distribution based on $X_1^*, \ldots, X_n^*$ and $\hat{\theta}_n^* = \theta(X_1^*, \ldots, X_n^*)$. Define

$$\hat{D}_n^* = \max_{i=1,\ldots,k} \sqrt{n} d((\mathbb{P}_n^*)_{h_i}, P_{h_i}(\cdot, \hat{\theta}_n^*)). \tag{9}$$

In the Appendix it is proved that the bootstrap works for this problem under suitable assumptions, that is, the conditional distribution of $\hat{D}_n^*$ given the data mimics that of $\hat{D}_n$ and critical values for the $\hat{D}_n$-test can be approximated by quantiles of $\hat{D}_n^*$ obtained by simulation. This Appendix is of technical character and can be skipped by the reader interested only in applications.

A consequence of the results in the Appendix is that the distribution of D under the null hypothesis depends indeed on the values of $h_1, ..., h_k$. Thus, the test is a randomized test.

5 Simulations

In this section we make some simulations in order to show how the proposed procedure behaves in practice. We have chosen two examples. We consider first the design of a test of multivariate normality. Then we propose a goodness of fit test for the Black-Scholes model.

Our tests will be based on the statistic $\hat{D}_n$ in (8). We use $\hat{D}_n$ because of its computational convenience and the good asymptotic properties shown in Subsection 6. Note, nevertheless, that the univariate K.S.-test is not optimal against the alternatives considered here. Thus, the power of the resulting test should increase if we replace K.S. by a better univariate test against these alternatives. We do not pursue this line here because we focus on providing a general idea on how the random projection method works in goodness-of-fit testing and, in particular, on how this kind of tests are affected when the dimension of the data set increases.

Computations have been carried out with MatLab. Programs are available from the authors upon request.

5.1 Gaussian goodness-of-fit test

The null hypothesis in this section is that the sample comes from a d-dimensional Gaussian distribution. We have considered several values of d. The alternatives we use are the same as in [12], which we describe succinctly: If we denote by $\mathbf{3}$ (resp. $\mathbf{0}$) the d-dimensional vector with all components equal to 3 (resp. 0) and by S (resp. I_d) the covariance matrix with all elements in the diagonal equal to 1 and the elements off-diagonal equal to .9 (resp. the identity), then the considered alternatives are the mixtures

$$pN_d(0, Id) + (1-p)N_d(m, \Sigma), \tag{10}$$

where m is $\mathbf{3}$ or $\mathbf{0}$ and Σ is S or Id, for several values of d and $p = .5, .79, .9$.

We consider the projection of the sample on k randomly chosen directions and the statistic $\hat{D}_n$ defined in (8). According to the results in Subsection 6, the rejection areas have been computed using the bootstrap as follows.

If $k < d$, we have started by selecting the k vectors, computing the projection of the sample on the k-dimensional subspace generated by those vectors and then we have taken $\hat{\theta}_n = (\hat{\mu}_n, \hat{\Sigma}_n)$ to be the sample mean and covariance matrix of the projections on this k-dimensional subspace. At this stage, we have generated 200 k-dimensional random samples from a Gaussian distribution with parameters $(\hat{\mu}_n, \hat{\Sigma}_n)$ and we have computed bootstrap replicas $\hat{D}^*_{n,r}$, $r = 1, \ldots, 200$. We have sorted the values and, if we denote by R the average of the values in positions 190 and 191 in the sorted list, then the rejection region is $[R, \infty)$.

If $k \geq d$, to reduce the computational effort, we calculate first the sample mean and covariance matrix. Then, we have generated 200 random samples from a d-dimensional Gaussian distribution with this mean and this covariance, we have projected each of these samples on the k randomly chosen vectors and we have worked as in the previous case.

In Tables 5.1 and 5.2, we show proportions of rejections along the 2000 repetitions we have done for each number of randomly chosen projections and alternatives. This number is the same as in [12] to make the results comparable. In both tables, k denotes the number of randomly chosen projections, d

the dimension of the space under consideration, p, m and Σ the corresponding value of the parameters in (10).

We compare first our results to those obtained with the Projection Pursuit method applied to the K.S-test as carried out in [12]. We consider this comparison particularly illuminating because of the reasons stated in the introduction. In [12], the authors have fixed $d = 2$ and, then, they have taken 15000 directions uniformly scattered on the unit sphere. Results are shown in Table 5.1.

Table 5.1
Comparison with the Projection Pursuit applied to the K-S-test. Dimension of the data is $d = 2$.

Sample size	p	m	Σ	projection pursuit	Projected tests $k=1$	$k=2$	$k=5$	$k=10$	$k=40$
$n=25$	.5	**3**	I_d	.22	.248	.283	.294	.317	.283
		0	S	.15	.083	.093	.109	.115	.144
	.79	**3**	I_d	.48	.346	.414	.481	.526	.520
	.1	**3**	Id	.36	.250	.307	.376	.398	.384
		0	S	.14	.070	.086	.093	.124	.148
$n=50$	.5	**3**	I_d	.61	.410	.566	.665	.706	.688
		0	S	.26	.126	.142	.185	.193	.254
	.79	**3**	I_d	.87	.507	.671	.832	.891	.897
	.1	**3**	Id	.67	.382	.533	.637	.670	.716
		0	S	.25	.089	.116	.136	.161	.217
$n=100$	.5	**3**	I_d	.97	.542	.759	.940	.977	.984
		0	S	.54	.159	.184	.291	.384	.491
	.79	**3**	I_d	1	.633	.823	.976	.999	1
	.1	**3**	Id	.94	.574	.735	.908	.933	.950
		0	S	.53	.102	.151	.201	.252	.365
Geometric means:				.4421	.2334	.2968	.3613	.4005	.4427

As a way to summarize the results, we have computed the geometric mean of the values provided by those authors and, also, the geometric mean of our results in the same cases.

We consider that the results in Table 5.1 are very promising. The obtained powers with just 40 randomly selected directions are very similar to those obtained with Projection Pursuit with 15000 projections. Moreover, the application of Projection Pursuit procedures in higher dimensions is quite involved, while the application of the random projection method is straightforward and the growth in computational time when the dimension increases is very small.

An important issue is how the method is affected by increasing dimensions. This is analyzed in Table 5.2. For the sake of comparison, when available, we provide the proportion of rejections obtained by Szekely and Rizzo (see [12]) when applying their (S.R.-)test. The comparison between our results and those obtained with the S.R.-test is now more difficult because both procedures are based on different univariate tests. In particular, it seems that increasing the dimension affects very negatively the S.R.-test when $m = \mathbf{3}$ and $\Sigma = I_d$, while the effect is clearly positive if $m = \mathbf{0}$ and $\Sigma = S$.

With respect to our test, it seems that it is not so affected as the S.R.-test. Roughly speaking, the effect appears to go in the same direction: when $m = \mathbf{3}$ and $\Sigma = I_d$ the power (slightly) decreases with the dimension, while if $m = \mathbf{0}$ and $\Sigma = S$, the power (slightly) increases with the dimension. It is worth noticing that if $k = 40$, then the power increases in both cases. Thus, it seems that a good protection against the curse of the dimensionality in this kind of problem is to take a number of projections around 40.

5.2 Goodness-of-fit test to the Black-Scholes model

The celebrated Black-Scholes model for the evolution of stock prices, $S(t)$, is described by the differential equation $dS(t) = \mu S(t)dt + \sigma S(t)dW(t)$, where $\mu \in \mathbb{R}$, $\sigma > 0$ and W is a standard Brownian motion (in $[0, 1]$), see, e.g., [11]. If a stochastic process $Y(t)$ follows the Black-Scholes model then

$$Z(t) = \log(Y(t)) = sW(t) + at \tag{11}$$

for some $s > 0$ and $a \in \mathbb{R}$. We will consider Z as an $L^2[0, 1]$-valued random element with the usual scalar product and norm.

After the logarithmic transform, testing fit to the Black-Scholes model fits in the setup of Proposition 3.14 where V_n is the one-dimensional subspace generated by the identity and the r.e. X in this proposition is a standard Brownian Motion. Hence, according to this proposition, three one-dimensional projections suffice to construct a goodness-of-fit test to the Black-Scholes model.

We analyze the behavior of our procedure against alternatives of type

$$Z(t) = (1 + s_2 f(t))W(t) + (a_1 t + a_2 g(t)),\ t \in [0, 1], \tag{12}$$

Table 5.2
Comparison with Szekely and Rizzo's test. Sample size is 50.

Dimension	p	m	Σ	Szekely and Rizzo	Projected tests $k=1$	$k=2$	$k=5$	$k=10$	$k=40$
$d=3$	.5	**3**	I_d	.58	.380	.540	.748	.865	.926
		0	S	.71	.127	.180	.214	.258	.344
	.79	**3**	I_d	.98	.451	.635	.854	.952	.985
	.1	**3**	Id	.91	.353	.521	.708	.801	.859
		0	S	.65	.114	.129	.167	.197	.266
$d=5$	.5	**3**	I_d	.20	.332	.518	.742	.875	.989
		0	S	.99	.139	.175	.230	.274	.410
	.79	**3**	I_d	.79	.411	.601	.842	.955	.998
	.1	**3**	Id	.93	.342	.508	.733	.847	.932
		0	S	.89	.100	.127	.180	.231	.326
$d=10$	.5	**3**	I_d	.05	.332	.501	.696	.863	.997
		0	S	1	.155	.201	.254	.304	.462
	.79	**3**	I_d	.27	.383	.550	.811	.946	1
	.1	**3**	Id	.66	.351	.459	.714	.840	.966
		0	S	.97	.114	.140	.190	.236	.392
$d=50$	.5	**3**	I_d	—	.310	.462	.687	.864	.997
		0	S	—	.154	.210	.269	.334	.470
	.79	**3**	I_d	—	.353	.540	.806	.933	1
	.1	**3**	Id	—	.321	.442	.668	.852	.967
		0	S	—	.110	.138	.203	.256	.416
$d=500$	.5	**3**	I_d	—	.326	.463	.687	.867	.995
		0	S	—	.170	.213	.281	.344	.487
	.79	**3**	I_d	—	.368	.540	.771	.937	1
	.1	**3**	Id	—	.308	.445	.694	.847	.973
		0	S	—	.113	.150	.204	.267	.446

where we have taken f, g equal to t^2 or $\sin(2\pi t)$ and $s_2, a_1, a_2 \in \{0, 1\}$. Thus, if $s_2 = a_2 = 0$ we are in the null hypothesis, otherwise the alternative holds.

According to Proposition 3.14, any dissipative distribution μ is valid for generating the random directions to be employed to project. Here we have taken μ to be Wiener measure on $L^2[0,1]$.

To generate replicas of W we have taken $W(0)=0$ and, then

$$W(t) = W\left(\frac{j-1}{N}\right) + w_j,\ j=1,...,N,\ t \in \left(\frac{j-1}{N}, \frac{j}{N}\right],$$

where $w_j,\ j=1,...,N$ are i.i.d. real r.v.'s with Gaussian distribution centered at 0 and variance N^{-1}. In our simulations, we have fixed $N=100$ and we have taken $n=50$ as sample size.

Concerning the estimates of a and s in (11), observe that, for the model (11) $E(Z(1)) = a$ and $\langle Z\rangle = s^2$, where $\langle Z\rangle$ denotes the quadratic variation in $[0,1]$ of Z (see, e.g., [11] for general background on this and other concepts in stochastic calculus). Thus, if we have the random sample $Z_1,...,Z_m$, this motivates that we consider the estimators

$$\hat{a}_1 = \frac{1}{n}\sum_{i=1}^{n} Z_i(1)$$

and

$$\hat{s}^2 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N}\left[Z_i\left(\frac{j}{N}\right) - Z_i\left(\frac{j-1}{N}\right)\right]^2.$$

It can be shown that these estimators satisfy the hypotheses in Corollary 6.3.

We have tried our test with $k=3,5,10$ and 40 one-dimensional projections. The rejection region has been obtained as in Subsection 5.1 taking $B=200$ bootstrap simulations. In Table 5.3, we show the proportion of rejections along 1000 repetitions.

Results in this table are very promising. We get quite high powers, even for the minimum value of $k=3$, except in the case where $s_2=1, a_2=0$ and $f(t)=t^2$, in which we obtain proportion of rejections around .150. The conclusion is that the procedure works well for data with sinusoidal volatility or with a non-linear drift. The only case in which the procedure provides not-so-good results is when the volatility is $f(t)=t^2$ and there is linear drift.

6 Appendix: Bootstrapping the estimated empirical process.

As noted in Section 4, projection on a finite number of directions reduces our data to i.i.d. observations in $\mathbb{R}^k$ and we want to test fit of this multivariate data

Table 5.3
Goodness-of-fit test to the Black-Scholes model. Sample size is 50. Samples have been taken from $Z(t) = (1 + s_2 f(t))W(t) + (a_1 t + a_2 g(t)),\ t \in [0,1]$, where W is a standard Brownian motion. Sample size is $n = 50$.

		s_2	a_1	a_2	Projected tests			
					$k=3$	$k=5$	$k=10$	$k=40$
$f(t) = t^2$	$g(t) = t^2$	0	0	0	.056	.054	.050	.057
		0	0	1	.713	.758	.809	.847
		0	1	0	.059	.053	.055	.049
		0	1	1	.712	.741	.772	.850
		1	0	0	.154	.154	.166	.144
		1	0	1	.526	.569	.650	.720
		1	1	0	.154	.147	.147	.123
		1	1	1	.513	.551	.614	.717
	$g(t) = \sin(2\pi t)$	0	0	1	.981	.998	1	1
		0	1	1	.976	.998	1	1
		1	0	1	.939	.981	1	1
		1	1	1	.942	.977	1	1
$f(t) = \sin(2\pi t)$	$g(t) = t^2$	1	0	0	.968	.992	1	1
		1	0	1	.994	1	1	1
		1	1	0	.960	.994	.999	1
		1	1	1	.995	1	1	1
	$g(t) = \sin(2\pi t)$	1	0	1	1	1	1	1
		1	1	1	.999	1	1	1

to some parametric family. Hence we consider the following setup. I$\{X_n\}_n$ are i.i.d. $\mathbb{R}^k$-valued random vectors with common distribution $P(\cdot, \theta_0)$ (the X_n's here correspond to the k-dimensional projections $[\langle X_n, h_1\rangle, \dots, \langle X_n, h_k\rangle]^T$ of the X_n in Section 4 and the $P(\cdot, \theta)$ here to the projection of the $P(\cdot, \theta)$ there on span$\{h_1, \dots, h_k\}$; to ease notation we avoid any reference to $h_1, \dots, h_k$ in this Appendix).

We assume this common distribution to be a member of the family $\mathcal{P} := \{P(\cdot, \theta) : \theta \in \Theta\}$, with $\Theta \subset \mathbb{R}^d$ an open set. We take $\mathcal{F}$ to be a class of real valued functions on $\mathbb{R}^k$ such that $\sup_{f\in\mathcal{F}} |f(x) - P(f, \theta)| < \infty$ for every x and θ. The *true* value of the parameter, θ_0, will be, in general, unknown and we cannot, therefore, base our inferences on the usual empirical process indexed by $\mathcal{F}$, namely, $\{\mathbb{G}_n(f)\}_{f\in\mathcal{F}}$ with $\mathbb{G}_n(f) = \sqrt{n}(\mathbb{P}_n(f) - P(f,\theta_0))$ ($\mathbb{G}_n(f)$ is considered as an $l^\infty(\mathcal{F})$-valued process). We can, instead, consider the *estimated empirical process*:

$$\left\{\hat{\mathbb{G}}_n(f) = \sqrt{n}(\mathbb{P}_n(f) - P(f,\hat{\theta}_n))\right\}_{f\in\mathcal{F}}, \tag{13}$$

where $\hat{\theta}_n = \theta(X_1,\dots,X_n)$ is some suitable estimator of θ_0. This estimated empirical process was first considered in [6] for $k=1$ and $\mathcal{F} = \{\mathbf{1}_{(-\infty,x]} : x \in \mathbb{R}\}$, giving weak convergence (in the Skorohod topology) to a certain Gaussian process under suitable regularity assumptions on $\mathcal{P}$ and $\hat{\theta}_n$. Later, [1] generalized these results to the class of lower rectangles in $\mathbb{R}^k$, $\mathcal{F} = \{\mathbf{1}_{(-\infty,x]} : x \in \mathbb{R}^k\}$, where $\mathbf{1}_{(-\infty,x]} = \mathbf{1}_{(-\infty,x_1]} \times \cdots \times \mathbf{1}_{(-\infty,x_k]}$ for $x = (x_1,\dots,x_k) \in \mathbb{R}^k$. We will give here a simple extension to general classes of functions. Further, in order to make our results usable in practice, we will provide sufficient conditions under which the (parametrically) bootstrapped estimated empirical process, namely,

$$\left\{\hat{\mathbb{G}}^*_n(f) = \sqrt{n}(\mathbb{P}^*_n(f) - P(f,\hat{\theta}^*_n))\right\}_{f\in\mathcal{F}}, \tag{14}$$

mimics the distribution of the estimated empirical process. In (14), $\mathbb{P}^*_n$ denotes the empirical measure based on the bootstrap sample $X^*_1,\dots,X^*_n$ which, conditionally given $X_1,\dots,X_n$, are i.i.d. random vectors with common law $P(\cdot,\hat{\theta}_n)$ and $\hat{\theta}^*_n = \theta(X^*_1,\dots,X^*_n)$.

As in [6] or [1] we will assume $\hat{\theta}_n$ to be *efficient*, meaning that, under $P(\cdot,\theta)$,

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^n l(X_i,\theta) + o_{P(\cdot,\theta)}(1). \tag{15}$$

We also make the following assumptions on function l in (15):

$$P(l(\cdot,\theta),\theta) = 0, \qquad P(l(\cdot,\theta)^T l(\cdot,\theta),\theta) < \infty. \tag{16}$$

If $P(f,\theta)$ is smooth enough and we denote $\nabla_\theta P(f,\theta)$ the vector of partial derivatives of $P(f,\theta)$ with respect to θ then we have

$$\begin{aligned}\hat{\mathbb{G}}_n(f) &= \mathbb{G}_n(f) - \sqrt{n}\left(P(f,\hat{\theta}_n) - P(f,\theta_0)\right)\\ &= \mathbb{G}_n(f) - \sqrt{n}\nabla_\theta P(f,\theta_0)^T(\hat{\theta}_n - \theta_0) + o_{P_{\theta_0}}(1)\\ &= \mathbb{G}_n(f) - \nabla_\theta P(f,\theta_0)^T\mathbb{G}_n(l(\cdot,\theta_0)) + o_{P_{\theta_0}}(1).\end{aligned}$$

Sufficient conditions to ensure that the $o_{P_{\theta_0}}(1)$ in the above display can be taken to be uniform in f (that is, to ensure that the $l^\infty(\mathcal{F})$-valued r.e.

$$\hat{\mathbb{G}}_n - \left(\mathbb{G}_n - \nabla_\theta P(\cdot,\theta_0)^T\mathbb{G}_n(l(\cdot,\theta_0))\right)$$

tends to 0 in outer probability) are given by

$$\nabla_\theta P(f,\theta) \text{ is uniformly bounded in } f \text{ for fixed } \theta. \tag{17}$$

and

$$\nabla_\theta P(f,\theta) \text{ is uniformly continuous in } f \text{ and } \theta \in \Lambda, \tag{18}$$

where Λ is some neighborhood of θ_0. Here, uniform continuity in f is with respect to the semi-metric

$$\rho_{\theta_0}(f,g) := \left(P((f-g)^2,\theta_0) - (P((f-g),\theta_0))^2\right)^{1/2}.$$

Observe now that, for a fixed element $T \in l^\infty(\mathcal{F})$, $\tau_T(z) := z(\cdot) - T(\cdot)z(l)$ is a continuous map from $l^\infty(\mathcal{F} \cup \{l\})$ into $l^\infty(\mathcal{F})$ (this follows from the inequality $\|\tau_T(z_1) - \tau_T(z_2)\|_{\mathcal{F}} \le (1 + \|T\|_{\mathcal{F}})\|z_1 - z_2\|_{\mathcal{F}\cup\{l\}}$). Thus, if

$$\mathcal{F} \text{ is } P(\cdot,\theta_0)\text{-Donsker}, \tag{19}$$

then so is $\mathcal{F} \cup \{l\}$ and this essentially proves (only straightforward changes that we omit are needed) the following result.

Theorem 6.1 *Under (15)-(19) $\hat{\mathbb{G}}_n$ converges weakly in $l^\infty(\mathcal{F})$ to the Gaussian process*

$$\hat{\mathbb{G}}(f) = \mathbb{G}(f) - \nabla_\theta P(f,\theta_0)^T \mathbb{G}(l(\cdot,\theta_0)),$$

where $\mathbb{G}$ is a $P(\cdot,\theta_0)$-Brownian bridge.

Let us turn now to the bootstrapped estimated empirical process defined in (14). We recall that weak convergence of a sequence of random elements $\{Z_n\}_n$ to a (tight, Borel measurable) limit Z in $l^\infty(\mathcal{F})$ is equivalent to

$$\sup_{h\in BL_1} |E^*h(Z_n) - Eh(Z)| \to 0$$

as $n \to \infty$, where BL_1 is the set of all functions $h : l^\infty(\mathcal{F}) \to \mathbb{R}$ bounded by 1 and such that $|h(x) - h(y)| \le \|x - y\|_{\mathcal{F}}$, see, e.g., [13], p 73. We will say that $\hat{\mathbb{G}}_n^*$ *converges weakly to* $\hat{\mathbb{G}}$ *in outer probability* if

$$\sup_{h\in BL_1} \left|\tilde{E}h(\hat{\mathbb{G}}_n^*) - Eh(\hat{\mathbb{G}})\right| \to 0$$

in outer probability. Here $\tilde{E}$ denotes outer expectation with respect to the probability $P(\cdot,\hat{\theta}_n)$, conditionally given $\hat{\theta}_n$. It is easy to see that if $\hat{\mathbb{G}}_n^*$ converges weakly to $\hat{\mathbb{G}}$ in outer probability, $H : l^\infty(\mathcal{F}) \to \mathbb{R}$ is a continuous function such that $H(\hat{\mathbb{G}}_n^*)$ is measurable and d metrizes weak convergence of probability measures on the line, then

$$d(\tilde{\mathcal{L}}(H(\hat{\mathbb{G}}_n^*)), \mathcal{L}(H(\hat{\mathbb{G}}))) \to 0$$

in probability, where $\tilde{\mathcal{L}}(H(\hat{\mathbb{G}}_n^*))$ denotes the conditional law of $H(\hat{\mathbb{G}}_n^*)$ given $\hat{\theta}_n$.

To show that $\hat{\mathbb{G}}_n^*$ converges weakly to $\hat{\mathbb{G}}$ in outer probability we will use the notation $\mathbb{G}_{n,\theta}$ for the empirical process when the underlying distribution is $P(\cdot,\theta)$. We can write now

$$\begin{aligned}\mathbb{G}_n^*(f) &= \sqrt{n}\left(\mathbb{P}_n^*(f) - P(f,\hat{\theta}_n^*)\right)\\ &= \mathbb{G}_{n,\hat{\theta}_n}(f) - \sqrt{n}\left(P(f,\hat{\theta}_n^*) - P(f,\hat{\theta}_n)\right).\end{aligned}$$

If we strengthen (15) to

$$L_\varepsilon(\theta_n) \to 0 \tag{20}$$

as $\theta_n \to \theta_0$, for every $\varepsilon > 0$, where $L_\varepsilon(\theta) = P\left(\left\|\sqrt{n}(\hat{\theta}_n - \theta) - \mathbb{G}_{n,\theta}(l)\right\| > \epsilon,\, \theta\right)$ (observe that, with this notation, (15) means $L_\varepsilon(\theta_0) \to 0$), then we can argue as before to get

$$\begin{aligned}\hat{\mathbb{G}}_n^*(f) &\simeq \mathbb{G}_{n,\hat{\theta}_n}(f) - \nabla_\theta P(f,\hat{\theta}_n)^T \mathbb{G}_{n,\hat{\theta}_n}(l(\cdot,\hat{\theta}_n))\\ &\simeq \mathbb{G}_{n,\hat{\theta}_n}(f) - \nabla_\theta P(f,\theta_0)^T \mathbb{G}_{n,\hat{\theta}_n}(l(\cdot,\hat{\theta}_n)),\end{aligned}$$

with $\simeq$ meaning that the difference between the corresponding $l^\infty(\mathcal{F})$-valued r.e.'s tends to 0 in outer probability. If we also assume

$$P\left((l(\cdot,\theta_n) - l(\cdot,\theta_0))^2, \theta_n\right) \to 0, \tag{21}$$

as $\theta_n \to \theta_0$, then we obtain

$$\hat{\mathbb{G}}_n^*(f) \simeq \mathbb{G}_{n,\hat{\theta}_n}(f) - \nabla_\theta P(f,\theta_0)^T \mathbb{G}_{n,\hat{\theta}_n}(l(\cdot,\theta_0)).$$

Thus, we see that the study of the asymptotic behavior of $\hat{\mathbb{G}}_n^*$ can be reduced to the consideration of a central limit theorem for a empirical process based on a triangular array, namely, $\{\mathbb{G}_{n,\hat{\theta}_n}(f)\}_{f\in\mathcal{F}\cup\{l(\cdot,\theta_0)\}}$. This, in turn, can be obtained as an application of the uniform central limit theorem, see Subsection 2.8.3 in [13]. Sufficient conditions for the desired conclusion are

$$\sup_{f,g\in\mathcal{F}\cup\{l(\cdot,\theta_0)\}} |\rho_{\hat{\theta}_n}(f,g) - \rho_{\theta_0}(f,g)| \to 0 \tag{22}$$

in probability,

$$\mathcal{F}\cup\{l(\cdot,\theta_0)\} \text{ is Donsker and pre-Gaussian uniformly in } \{P(\cdot,\theta)\}_{\theta\in\Theta} \tag{23}$$

and

$$\mathcal{F} \cup \{l(\cdot, \theta_0)\} \text{ has a measurable envelope } F \text{ such that } P(F^2 I_{F > \varepsilon\sqrt{n}}, \hat{\theta}_n) \to 0 \tag{24}$$

in probability, for every $\varepsilon > 0$. We summarize this argument in the following results.

Theorem 6.2 *If conditions (15)-(18) and (20)-(24), hold then $\hat{\mathbb{G}}_n^*$ converges weakly to $\hat{\mathbb{G}}$ in outer probability, where $\hat{\mathbb{G}}$ is the Gaussian process defined in Theorem 6.1*

Corollary 6.3 *(Consistency of the bootstrap for general Kolmogorov-Smirnov statistics with estimated parameters). Under the hypotheses of Theorem 6.2, if $\|\hat{\mathbb{G}}_n^*\|_{\mathcal{F}}$ is measurable and d metrizes weak convergence of probability measures on the line, then*

$$d(\tilde{\mathcal{L}}(\|\hat{\mathbb{G}}_n^*\|_{\mathcal{F}}), \mathcal{L}(\|\hat{\mathbb{G}}\|_{\mathcal{F}})) \to 0$$

in probability.

Example 6.4 Let us consider the problem of testing fit to the family of multivariate normal distributions $N(\mu, \Sigma)$ on $\mathbb{R}^k$ using the Kolmogorov-Smirnov statistic with estimated parameters, namely,

$$\hat{D}_n = \sqrt{n} \sup_{x \in \mathbb{R}^k} |F_n(x) - F(x; \hat{\mu}_n, \hat{\Sigma}_n)|,$$

where F_n denotes the k-variate empirical d.f., $F(x; \mu, \Sigma)$ the distribution function of the $N(\mu, \Sigma)$ law and $\hat{\mu}_n, \hat{\Sigma}_n$ are the usual estimators based on a sample of size n, namely, the sample mean and covariance matrix, respectively. Observe that $\hat{D}_n = \|\hat{\mathbb{G}}_n\|_{\mathcal{F}}$, with $\mathcal{F}$ the class of indicators of lower rectangles in $\mathbb{R}^k$, $\mathcal{F} = \{\mathbf{1}_{(-\infty,x]} : x \in \mathbb{R}^k\}$ (we denote $\mathbf{1}_{(-\infty,x]} = \mathbf{1}_{(-\infty,x_1]} \times \cdots \times \mathbf{1}_{(-\infty,x_k]}$ for $x = (x_1, \ldots, x_k) \in \mathbb{R}^k$). Here $\hat{\theta}_n = (\hat{\mu}_n, \hat{\Sigma}_n)$ satisfies (15) with $l(x, \theta) = (x - \mu, (x - \mu)(x - \mu)^T - \Sigma)$ and remainder $(0, -\sqrt{n}(\bar{X}_n - \mu)(\bar{X}_n - \mu)^T)$. We also have (16), whereas (17) and (18) follow from straightforward (but cumbersome) computations that we omit. It is well known that $\mathcal{F}$ is universally Donsker (see, e.g., [13], p 129), hence we also have (19). Applying Theorem 6.1 we can conclude

$$\hat{D}_n \underset{w}{\to} \|\hat{\mathbb{G}}\|_{\mathcal{F}}.$$

The distribution of $\|\hat{\mathbb{G}}\|_{\mathcal{F}}$ is hard to evaluate even in dimension one.

We can try to use a bootstrap approach based on

$$\hat{D}_n^* = \sqrt{n} \sup_{x \in \mathbb{R}^k} |F_n^*(x) - F(x; \hat{\mu}_n^*, \hat{\Sigma}_n^*)|$$

where F_n^* is the empirical d.f. associated to $X_1^*, \ldots, X_n^*$, i.i.d. $N(\hat{\mu}_n, \hat{\Sigma}_n)$ r.v.'s and $\hat{\mu}_n^*$ and $\hat{\Sigma}_n^*$ are the sample mean and sample covariance matrix, respectively, computed from $X_1^*, \dots, X_n^*$. Now, (20) reduces to check

$$P\left(\|\sqrt{n}(\bar{X}_n - \mu_n)(\bar{X}_n - \mu_n)^T\| > \varepsilon; \mu_n, \Sigma_n\right) \to 0,$$

when $(\mu_n, \Sigma_n) \to (\mu, \Sigma)$, which is certainly true. (21) follows from the fact that $l(x,\theta_n) - l(x,\theta_0) = (\theta_0 - \theta_n, (x-\mu_n)(x-\mu_n)^T - (x-\mu_0)(x-\mu_0)^T - (\Sigma_0 - \Sigma_n))$ and we can similarly show that (22) holds in this case. Condition (24) is an easy consequence of the fact that $\mathcal{F}$ is a uniformly bounded class of functions. Finally, (23) is a consequence of $\mathcal{F}$ being a uniformly bounded VC-class (plus some measurability properties). Summarizing, we obtain that

$$d(\tilde{\mathcal{L}}(\hat{D}_n^*), \mathcal{L}(\hat{D}_n)) \to 0$$

in probability and, therefore, that we can use the bootstrap to approximate the distribution of $\hat{D}_n$.

Example 6.5 Consider now the statistics $\hat{D}_n$ and $\hat{D}_n^*$ of (8) and (9). They can be written as $\|\hat{\mathbb{G}}_n\|_{\mathcal{F}}$ and $\|\hat{\mathbb{G}}_n^*\|_{\mathcal{F}}$, respectively, with $\mathcal{F} = \mathcal{F}_1 \cup \cdots \cup \mathcal{F}_k$ and $\mathcal{F}_j = \{\mathbb{R} \times \cdots \times (-\infty, x_j] \times \cdots \times \mathbb{R}\}_{x_j \in \mathbb{R}}$. We have again that $\mathcal{F}$ is a uniformly bounded VC-class and, as in Example 6.4 we can show that the assumptions of Theorem 6.2 are satisfied for suitable families of distributions (including the family of multivariate normal distributions) and, in those cases

$$d(\tilde{\mathcal{L}}(\hat{D}_n^*), \mathcal{L}(\hat{D}_n)) \to 0$$

in probability.

Acknowledgments. This research has been carried out partially during a stay of E. del Barrio and J.A. Cuesta-Albertos at Université Paul Sabatier, Toulouse, France. These authors want to acknowledge with thanks the warm hospitality they received there.

The authors also want to thank to an anonymous referee for its very careful reading of the manuscript.

References

[1] CSRGŐ, M.,CSRGŐ, S. and RVSZ, P. (1978). Approximations of the empirical process when parameters are estimated. *Ann. Probab.* **5** 790–810.

[2] CHOW, Y.S. and TEICHER, H. (1997). *Probability Theory. Independence, Interchangeability, Martingales. 3rd. edit.* Springer Verlag, New York.

[3] CUESTA-ALBERTOS, J.A., FRAIMAN, R. and RANSFORD, T. (2006). A sharp form of the Cramér-Wold theorem. To appear at *J. Theoret. Probab.*

[4] CUESTA-ALBERTOS, J.A., FRAIMAN, R. and RANSFORD, T. (2006). Random projections and goodness-of-fit tests in infinite-dimensional spaces. To appear in *Bull. Brazilian Math. Soc.*

[5] DIACONIS, P. and FREEDMAN, D. (1984). Asymptotics of graphical projection pursuit. *Ann. Statis.* **12** 793–815.

[6] DURBIN, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statis.* **1** 279–290.

[7] GLDÓFALVI, G. Random projection for non-Gaussian mixture models. Downloaded from http://www-cse.ucsd.edu/ elkan/254spring02/gidofalvireport.pdf

[8] LEHMANN, E. L., CASELLA, G. (1998). *Theory of Point Estimation.* Springer. New York.

[9] LEJSEK, H., ÁSMUNDSSON, F.H., JÓNSON, B.P. and AMSALEG, L. (2005). Efficient and effective image copyright enforcement. *Publication Interne* n. 1699. IRISA, Rennes.

[10] SHOHAT, J. A. and TAMARKIN, J. D. (1943) *The Problem of Moments*, Amer. Math. Soc., Providence, RI.

[11] SHREVE, S.E. (2004). *Stochastic Calculus for Finance II. Continuous-Time models.* Springer.

[12] SZEKELY, G.J. and RIZZO, M.L. (2005). A new test for multivariate normality. *J. Multivariate Anal.* **93** 58-80.

[13] van der VAART, A. W. and WELLNER, J.A. (1996). *Weak Convergence and Empirical Processes.* Springer Verlag. New York.

[14] VEMPALA, S. S. (2004). *The Random Projection Method.* American Mathematical Society, Rhode Island.

[15] VEMPALA, S. S. and WANG, G. (2004). A spectral algorithm for learning mixture models. *J. Computer and System Sciences*, **68** , Issue 4, 841 - 860.