JOURNAL OF MULTIVARIATE ANALYSIS 52, 1–14 (1995)

A Continuous Metric Scaling Solution for a Random Variable*

C. M. CUADRAS[†] AND J. FORTIANA[‡]

Universitat de Barcelona, Barcelona, Spain

As a generalization of the classical metric scaling solution for a finite set of points, a countable set of uncorrelated random variables is obtained from an arbitary continuous random variable X. The properties of these variables allow us to regard them as principal axes for X with respect to the distance function $d(u, v) = \sqrt{|u-v|}$. Explicit results are obtained for uniform and negative exponential random variables.

1. INTRODUCTION

Metric scaling or principal coordinate analysis, introduced by Torgerson [14] and especially Gower [9], is a method of ordination aiming to provide a graphical representation of a *finite* set of n elements. The method obtains an $n \times m$ matrix X from an $n \times n$ Euclidean distance matrix $\Delta = (\delta_{ij})$. The set of n rows of X, considered as points in $\mathbf{R}^m$, has inter-distances which reproduce those in Δ [11, p. 397]. Columns of X can be regarded as "variables" (*principal axes*), and each row as the set of values of these variables (*principal coordinates*) for the corresponding element of the original set.

This principal coordinate representation can be singled out between all possible Euclidean representations of the same set by duality with principal components. This property can be stated as follows: For any $n \times \tilde{m}$ matrix $\tilde{X}$ giving such a representation, the principal components for its "variables" (i.e., columns) are the principal coordinate axes.

Cuadras and Arenas (see [4, 5, 7]) take advantage of the good properties of these "variables" to define and study a distance-based model for prediction with mixed variables.

Received September 16, 1993; revised March 7, 1994.

AMS subject classification: 62H25.

Key words and phrases: principal components of a stochastic process, principal coordinate analysis.

* Work supported in part by CGYCIT Grant PS88-0032.

[†] E-mail: carlesm@porthos.bio.ub.es.

[‡] E-mail: josepf@porthos.bio.ub.es.

0047-259X/95 $6.00

Cuadras and Fortiana [8] proposed a continuous extension of this distance-based model, taking a uniform (0, 1) random variable U as predictor. A countable set of random variables were interpreted as principal axes of U with respect to a suitable distance. In this paper the construction of this set of principal coordinates is generalized to any continuous random variable.

2. Continuous Euclidean Configuration

Let X be a random variables on a probability space $(\Omega, \mathscr{A}, P)$, with values on a (possibly unbounded) interval $I \equiv (a, b) \subset \mathbf{R}^* = \mathbf{R} \cup \{-\infty\} \cup \{\infty\}$. Denote its c.d.f. by F, and let $\delta: I \times I \to \mathbf{R}_+$ be a distance function.

Definition 1. A *continuous Euclidean configuration* representing X with respect to δ, is a stochastic process $\mathbf{X} = \{X_t\}_{t \in I}$ such that for all $\omega_1, \omega_2 \in \Omega$ the *Euclidean distance between trajectories* $X_t(\omega_1)$, $X_t(\omega_2)$, defined as

$$D_E(\omega_1, \omega_2) \equiv \left\{ \int_I (X_t(\omega_1) - X_t(\omega_2))^2 \, dt \right\}^{1/2}$$

equals $\delta(X(\omega_1), X(\omega_2))$.

That is, $\mathbf{X}$ is defined as a process such that distances between its "rows" reproduce the interdistances of the original (continuous, one-dimensional) set of points. When δ is the Euclidean distance, a trivial representation of X is the degenerate process with $X_t = X$ for $t \in [0, 1]$ and $X_t = 0$ for $t \notin [0, 1]$.

Throughout the paper we take as distance δ the function d defined by

$$d(x, y) = \sqrt{|x - y|}, \qquad x, y \in (a, b). \tag{1}$$

One reason for this choice is that it has manifested good properties in the finite case (see [5, 7, 8]). In addition, a continuous Euclidean representation for this distance can easily be obtained, using the following.

Construction 1. Let X be a random variable as defined above. Consider the function $u: I \times I \to [0, 1]$ defined by

$$u(t, x) = \begin{cases} 1 & \text{if} \quad t < x, \\ 0 & \text{if} \quad t \geqslant x, \end{cases}$$

and let $\mathbf{X} = \{X_t\}_{t \in I}$ be defined as $X_t = u(t, X)$, for $t \in I$. That is, for each $t \in I$, X_t is the indicator of $[X > t] \in \mathscr{A}$, a Bernoulli random variable with $p = 1 - F(t)$.

PROPOSITION 1. *The process* $\mathbf{X} = \{X_t\}_{t \in I}$, *obtained from* X *using construction* 1, *is a continuous Euclidean representation of* X *with respect to distance* (1).

Proof. Given $\omega_1, \omega_2 \in \Omega$, let $x_i = X(\omega_i)$ and $u_i(t) = X_t(\omega_i) = u(t, x_i)$, for $i = 1, 2$. Assume, for instance, that $x_1 < x_2$. Then

$$D_E^2(\omega_1, \omega_2) = \int_I (u_1(t) - u_2(t))^2 \, dt = \int_{x_1}^{x_2} 1^2 \, dt = x_2 - x_1 = d^2(x_1, x_2). \quad \blacksquare$$

In the following, given a random variable X, $\mathbf{X}$ will denote the process obtained from X using Construction 1. Proposition 2 gives an additional relation between $\mathbf{X}$ and X, which in the continuous case allows us to write X as a sort of "continuous sum of indicators."

PROPOSITION 2.

$$\int_I X_t \, dF(t) = F \circ X. \tag{2}$$

In particular, when F *is continuous, this integral gives a uniform* (0, 1) *random variable, and* X *can be expressed as*

$$X = F^{-1}\left(\int_I X_t \, dF(t)\right).$$

Proof. Given $\omega \in \Omega$, let $x = X(\omega) \in (a, b)$. As $X_t(\omega) = 1$ for $t \in (a, x)$ and $= 0$ otherwise, we have $\int_I X_t(\omega) \, dF(t) = \int_a^x 1 \, dF(t) = F(x)$, and (2) holds. $\blacksquare$

PROPOSITION 3. *The covariance function of* $\mathbf{X}$ *is given by*

$$K(s, t) = \min\{F(s), F(t)\} - F(s) F(t), \qquad s, t \in I. \tag{3}$$

Proof. As $X_s X_t = X_{\max\{s, t\}}$, we have $K(s, t) = \mathrm{E}(X_s X_t) - \mathrm{E}(X_s) \mathrm{E}(X_t) = 1 - F(\max\{s, t\}) - (1 - F(s))(1 - F(t)) = F(s) + F(t) - F(\max\{s, t\}) - F(s) F(t)$. $\blacksquare$

K is a symmetric, positive semidefinite kernel, and when X is continuous, K also has this property. In any case, $0 \leqslant K(s, t) \leqslant 1$, for all $(s, t) \in I \times I$, and K tends to 0 on the boundary of its domain. It is worth noting that K is the difference between two bivariate distribution functions having F as both marginals, namely, the upper Fréchet bound $F^+(s, t) = \min\{F(s), F(t)\}$, and the product $F(s) F(t)$.

When X is a uniform (0, 1) random variable, (3) is the ubiquitous kernel $\min\{s, t\} - st$, which appears in probability theory as the covariance function of the Brownian bridge, in statistics, in the study by Anderson and Darling [1, 2] of empirical processes, in mechanics, as the Green function for the vibrating string, etc.

Whereas the continuous Euclidean representation (Construction 1) can be written in principle for any random variable, we will be interested in properties which require that kernel K verifies the following finitude conditions:

1. $\operatorname{tr}(K) \equiv \int_I K(s, s)\, ds < +\infty$.
2. $K \in \mathscr{L}^2(I \times I)$.

From Cauchy–Schwarz inequality, $|K(s, t)|^2 \leqslant K(s, s)\, K(t, t)$, we see that square integrability of K is implied by the finitude of $\operatorname{tr}(K)$. This condition can be translated in terms of geometrical properties of X with respect to distance (1) with the help of the following.

DEFINITION 2. The *geometric variability* of a (real valued) random variable X with respect to a distance function δ is the quantity

$$V_\delta(X) = \tfrac{1}{2} \int_{\mathbf{R}^2} \delta^2(s, t)\, dF(s)\, dF(t), \tag{4}$$

where F is the distribution function of X, provided that this integral exists.

$2V_\delta(X)$ is the expected value of the distance function $\delta^2(\cdot, \cdot)$, evaluated on two random variables, independent and identically distributed as X. When δ is the Euclidean distance, $V_\delta(X)$ coincides with $\operatorname{Var}(X)$. When δ is the distance (1), $V_d(X)$ is another measure of dispersion of X. Straightforward computations provide some examples of V_d as compared to $\sigma \equiv \sqrt{\operatorname{Var}(X)}$.

- For $X \sim N(\mu, \sigma^2)$, $V_d(X) = 2\sigma/\sqrt{\pi}$.
- For a uniform random variable, $V_d(X) = \sigma/\sqrt{3}$.
- For a negative exponential random variable, $V_d(X) = \sigma/2$.
- For a logistic random variable, $V_d(X) = \sqrt{3}\sigma/\pi$.

The geometric variability of a random vector is similarly defined. It was used by Cuadras [4–6], to perform a Discriminant Analysis based on distances between observations, and by Rao [12], to define dissimilarity coefficients between populations.

PROPOSITION 4. *Let X be a random variable such that* $\lim_{s \to -\infty} sF(s) = 0$, *and let d be the distance function* (1). *Then, when any of the quantities* $V_d(X)$, $\operatorname{tr}(K)$ *is finite, the other is also finite, and the equality* $V_d(X) = \operatorname{tr}(K)$ *holds.*

Proof. The distances we are considering are symmetric nonnegative functions, vanishing on the diagonal $[y=x]\subset \mathbf{R}^2$. Thus, for any δ, the geometric variability (4) is equivalent to an integral on a half-plane, and it can be computed through iterated one-dimensional integrals

$$V_\delta(X)=\int_{[y<x]} \delta^2(x, y)\, dF(x)\, dF(y)=\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{x} \delta^2(x, y)\, dF(y)\right) dF(x).$$

Specializing to $d^2(x, y)=|x-y|$, using Riemann–Stieljes integration by parts and taking into account the hypothesis $\lim_{y\to -\infty} yF(y)=0$, we see that for any $x\in \mathbf{R}$ the integrand equals

$$\int_{-\infty}^{x} (x-y)\, dF(y)=[(x-y)\, F(y)]_{-\infty}^{x}+\int_{-\infty}^{x} F(y)\, dy=\int_{-\infty}^{x} F(y)\, dy.$$

Since the integrand is non-negative, Fubini's theorem allows us to interchange the order of integration, giving

$$\begin{aligned} V_d(X)&=\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{x} F(y)\, dy\right) dF(x)=\int_{-\infty}^{+\infty}\left(\int_{y}^{+\infty} dF(x)\right) F(y)\, dy\\ &=\int_{-\infty}^{+\infty} (1-F(y))\, F(y)\, dy=\operatorname{tr}(K). \quad \blacksquare \end{aligned}$$

3. Principal Coordinates

In this section, X will denote an absolutely continuous random variable, with c.d.f F and probability density f (with respect to the Lebesgue measure), such that the finitude conditions $\mu=\mathrm{E}(|X|)<+\infty$, $\lim_{s\to -\infty} sF(s)=0$, and $V_d(X)<+\infty$ are verified. From Mercer's theorem (see, for example, Courant and Hilbert [3, Vol. I, Chap. 3]),

$$K(s, t)=\sum_{j=1}^{\infty} \lambda_j \psi_j(s)\, \psi_j(t)$$

is absolutely and uniformly convergent (in both s and t) on $I\times I$, where $\{\psi_j\}_{j\in \mathbf{N}}$ is a complete orthonormal (in $\mathscr{L}^2(I)$) set of solutions of

$$\int_I \psi_j(s)\, K(s, t)\, ds=\lambda_j \psi_j(t). \tag{5}$$

From the theorem of Kac and Siegert [10] (see Shorack and Wellner [13]), we obtain a countable decomposition,

$$\mathbf{X} = \sum_{j=1}^{\infty} Z_j \psi_j(t), \tag{6}$$

where $\{Z_j\}_{j \in \mathbf{N}}$ is an orthogonal (i.e., uncorrelated) set of square integrable random variables defined by

$$Z_j = \int_I X_t \psi_j(t)\, dt, \qquad j \in \mathbf{N}, \tag{7}$$

and verifying that $\mathrm{Var}(Z_j) = \lambda_j$.

By analogy with the finite case, each Z_j is called a *principal component* of $\mathbf{X}$. In the following theorem we show that principal components are obtained as continuous functions h of X, and we compute an explicit differential equation for h.

THEOREM 1. *Let ψ be an eigenfunction of K, with associated eigenvalue λ $(< \infty)$, and consider the function*

$$h(s) = \int_a^s \psi(t)\, dt, \qquad s \in (a, b).$$

Then

1. *The Principal Component Z corresponding to λ is given by $Z = h(X)$.*
2. $m = \mathrm{E}(Z) = \int_a^b [1 - F(t)]\, \psi(t)\, dt.$
3. *h is a solution of*

$$\lambda h'' + (h - m) f = 0, \qquad h(a) = 0, \qquad h'(a) = 0. \tag{8}$$

Proof. 1. Given $\omega \in \Omega$, $Z(\omega) = \int_a^b X_t(\omega)\, \psi(t)\, dt = \int_a^{X(\omega)} \psi(t)\, dt = h(X(\omega))$.

2. As $Z = h(X)$ is a square integrable random variable, in particular, the integral $m = \mathrm{E}(Z) = \int_a^b h(t)\, f(t)\, dt$ is finite. Integrating by parts, we obtain

$$m = \lim_{x \to b} \left(h(t)\, F(t) \Big|_{t=a}^{t=x} - \int_a^x F(t)\, \psi(t)\, dt \right) = \int_a^b [1 - F(t)]\, \psi(t)\, dt. \tag{9}$$

3. Using kernel (3),

$$\int_a^b K(s,t)\,\psi(t)\,dt = \int_a^s [F(t) - F(s)\,F(t)]\,\psi(t)\,dt + F(s)\int_s^b (1 - F(t))\,\psi(t)\,dt. \tag{10}$$

From 2, the integral in the second summand is $m - \int_a^s (1 - F(t))\,\psi(t)\,dt$. Substituting into (10) and simplifying, we obtain

$$\lambda\psi(s) = \int_a^s F(t)\,\psi(t)\,dt + F(s)(m - h(s)).$$

Since $\psi(s) \equiv h'(s)$, differentiation of this equation yields (8) (using the fundamental theorem of calculus). Initial conditions are immediate from the boundary condition for K and from the definition of h. ∎

Besides (8), the equation for $g(t) = h(t) - m$, i.e.,

$$\lambda g'' + fg = 0, \qquad g(a) = -m, \qquad g'(a) = 0, \tag{11}$$

will be also useful.

The following theorem summarizes properties of the set $\{Z_j\}_{j\in\mathbb{N}}$. Comparison with the classical (finite) Metric Scaling solution as described by Mardia *et al.* [11, p. 399], suggests that, once standardized, each Principal Component Z should be a "principal coordinate axis" for X with respect to distance (1).

We denote by C_j and C_j^* the results of standardizing Z_j to mean 0 and to mean 0 and variance 1, respectively.

THEOREM 2. 1. *The C_j's are uncorrelated absolutely continuous random variables. The sequence of variances $\{\mathrm{Var}(C_j)\}_{j\in\mathbb{N}}$ is decreasing and summable, the sum being equal to the trace of K.*

2. *Given $\omega_1, \omega_2 \in \Omega$, the Euclidean distance between the sequences $\{C_j(\omega_1)\}_{j\in\mathbb{N}}$ and $\{C_j(\omega_2)\}_{j\in\mathbb{N}}$, equals $d(X(\omega_1), X(\omega_2))$.*

Proof. 1. The first statement is a standard property of the principal components of a stochastic process.

2. Given $\omega_1, \omega_2 \in \Omega$, let $x_i = X(\omega_i)$ and $u_i(t) = X_t(\omega_i) = u(t, x_i)$, for $i = 1, 2$. Expanding $u_1(t) - u_2(t)$ with respect to the complete orthonormal set $\{\psi_j\}$ on I, we obtain the Fourier series $\sum_{j=1}^{\infty} a_j\psi_j(t)$, where $a_j = \int_I (u_1(t) - u_2(t))\;\psi_j(t)\,dt = h_j(x_1) - h_j(x_2)$, where h_j is defined in Theorem 1. Then Parseval equality gives

$$|x_1 - x_2| = \int_I (u_1(t) - u_2(t))^2\,dt = \sum_{j=1}^{\infty} a_j^2 = \sum_{j=1}^{\infty} (C_j(\omega_1) - C_j(\omega_2))^2. \quad ∎$$

Thus, each C_j can be called a *principal coordinate axis* for X and, given $\omega \in \Omega$, the sequence $\{C_j(\omega)\}_{j \in \mathbf{N}}$ can accordingly be called a *continuous metric scaling* representation for $X(\omega)$ with respect to distance (1).

4. Solution for a Uniform Distribution

In this section, explicit results are obtained when X is a (0, 1) uniform random variable. In this case, (11) becomes the familiar differential equation of trigonometric functions. An appropriate solution is

$$g_j(t) = -\frac{\sqrt{2}}{j\pi}\cos(j\pi t), \qquad \lambda_j = \frac{1}{(j\pi)^2}, \qquad 0 \leqslant t \leqslant 1,\ j \in \mathbf{N}.$$

The corresponding standardized variables, $C_j^* = -\sqrt{2}\cos(j\pi X)$ have the following remarkable property.

Proposition 5. *The standardized C_j^*'s are identically distributed, with probability density function*

$$g(x) = \begin{cases} \pi^{-1}(2-x^2)^{-1/2}, & \text{if } -\sqrt{2} < x < \sqrt{2}, \\ 0, & \text{otherwise}, \end{cases} \tag{12}$$

with respect to the Lebesgue measure.

Proof. The characteristic function of the symmetric r.v. C_j^* is given by

$$\varphi(t) = \int_0^1 \exp[it(-\sqrt{2}\cos(j\pi x))]\,dx = \int_0^1 \cos[t\sqrt{2}\cos(j\pi x)]\,dx.$$

Applying the change of variable $y = j\pi x$, we obtain $\varphi(t) = (1/j\pi) \int_0^{j\pi} \cos[t\sqrt{2}\cos y]\,dy$, which can be written as a sum of j integrals on the intervals $[(k-1)\pi, k\pi]$, $(k = 1, ..., j)$. As the integrand is an even, periodic (with period $= 2\pi$) function of y, all these summands coincide, thus

$$\varphi(t) = \frac{1}{\pi}\int_0^{\pi} \cos[t\sqrt{2}\cos y]\,dy,$$

independently of j. This integral equals $J_0(\sqrt{2}\,t)$, where J_0 is the zeroth order Bessel function (see, e.g., [15, Chap. III, Sect. 3.3]). The inverse Fourier transform

$$g(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \varphi(t)\exp(-itx)\,dt = \frac{1}{\pi}\int_0^{\infty} J_0(\sqrt{2}\,t)\cos(tx)\,dt$$

yields the probability density function (12), by using a standard property of Bessel functions (see, e.g., [15, Chap. XIII, Sect. 13.42]). ∎

Remark. An alternative, more elementary proof follows from the fact that given $j>0$ and $y \in (-\sqrt{2}, \sqrt{2})$, the equation $y = -\sqrt{2}\cos(j\pi x)$ has exactly j real roots $x_1, ..., x_j$, hence the value $g_j(y)$ of the density of C_j^* is given by

$$g_j(y) = \sum_{i=1}^{j} f_X(x_i) \Big/ \left|\frac{dy}{dx}\right|_{x=x_i}, \tag{13}$$

where $f_X(x_i) = 1$ (the density of the (0, 1) uniform r.v. X), and

$$\left|\frac{dy}{dx}\right|_{x=x_i} = j\pi\sqrt{2}\sin(j\pi x_i) = j\pi\sqrt{2-y^2}$$

Adding the j (equal) terms, we obtain (12).

The equality

$$\sum_{j=1}^{\infty} (C_j(x) - C_j(y))^2 = |x-y| \tag{14}$$

can be proved in this case by direct computation (see Appendix A). Finally, it is worth noting the formal analogy with the finite metric scaling solution [8]

$$C_j^* = -\sqrt{2}\, T_j(V),$$

where $V = \cos(\pi X)$ and T_j is the jth Chebyshev polynomial of the first kind.

5. Solution for an Exponential Distribution

Let X be a negative exponential random variable with c.d.f. $F(x) = 1 - \exp(-\alpha x)$, $x \in (0, \infty)$, $\alpha > 0$. Equation (11) is now

$$\lambda \frac{d^2 g(x)}{dx^2} + \alpha \exp(-\alpha x)\, g(x) = 0. \tag{15}$$

Integrating this differential equation we obtain

Proposition 6. *The jth standardized principal coordinate axis for X is given by*

$$C_j^* = \frac{1}{J_0(\xi_j)} J_0(\xi_j \exp(-\alpha X/2)), \qquad (j \in \mathbf{N}), \tag{16}$$

where J_0 is the zeroth order Bessel function (of the first kind), and ξ_k is the kth positive root of the first order Bessel function J_1.

The variance of the corresponding centered principal coordinate axis is given by

$$\mathrm{Var}(C_j)=\frac{4}{\alpha\xi_j^2}.$$

Proof. Applying to (15) the change of variable

$$t=2(\alpha\lambda)^{-1/2}\exp(-\alpha x/2),$$

and denoting $y(t)=g(x(t))$, we have

$$\frac{dg}{dx}=\frac{dy}{dt}\frac{dt}{dx}=-(\alpha/\lambda)^{1/2}\exp(-\alpha x/2)\frac{dy}{dt},$$

and

$$\frac{d^2g}{dx^2}=\frac{d}{dx}\left(\frac{dg}{dx}\right)=\frac{\alpha^{3/2}}{2\lambda^{1/2}}\exp(-\alpha x/2)\frac{dy}{dt}+\frac{\alpha}{\lambda}\exp(-\alpha x)\frac{d^2y}{dt^2},$$

which, substituting in (15), gives the Bessel differential equation

$$t\frac{d^2y}{dt^2}+\frac{dy}{dt}+ty=0, \tag{17}$$

where $t\in(0, 2(\alpha\lambda)^{-1/2})$.

As the solution of (17) must be finite for $t=0$, it will be of the form

$$y(t)=AJ_0(t),$$

where A is a constant. The contour condition $y'(2(\alpha\lambda)^{-1/2})=0$ imposes that $\xi=2(\alpha\lambda)^{-1/2}$ is a root of $y'=-AJ_1(t)$. Thus, we obtain a countable set of solutions of (15), substituting for ξ each of the positive roots ξ_j of J_1, $(j\in\mathbf{N})$, in

$$g(x)=AJ_0(\xi\exp(-\alpha x/2)). \tag{18}$$

Given ξ, the constant A is determined by imposing the condition that $\mathrm{Var}(g(X))=\lambda$, or equivalently, in the notation of Theorem 1, that the eigenfunction $\psi=g'$ is normalized so that $\int\psi^2=1$. In general, the kth moment $\mu_k=\mathrm{E}(g(X))$ is obtained by evaluating

$$\mu_k=A^k\int_0^\infty [J_0(\xi\exp(-\alpha x/2))]^k\,\alpha\exp(-\alpha x)\,dx.$$

The change of variable $t = \xi \exp(-\alpha x/2)$ gives

$$\mu_k = \frac{2A^k}{\xi^2} \int_0^\xi [J_0(t))]^k \, t \, dt.$$

For $k = 1$ and $k = 2$, this integral can be computed in closed form, (see, e.g., [15, Chap. V, Sect. 5.1]). The first moment is

$$\mu_1 = \frac{2A}{\xi^2} [tJ_1(t)]_0^\xi = 0,$$

as ξ is a root of J_1. The second moment is

$$\mu_2 = \frac{2A^2}{\xi^2} \left[\frac{t^2}{2} (J_0^2(t) + J_1^2(t)) \right]_0^\xi = A^2 J_0^2(\xi).$$

From this expression, the equality

$$\mu_2 = \lambda = \frac{4}{\alpha \xi^2}$$

gives

$$A = \frac{2}{\sqrt{\alpha}\, \xi J_0(\xi)}.$$

Substituting into (18), the centered principal coordinate axis corresponding to ξ is

$$C = \frac{2}{\sqrt{\alpha}\, \xi J_0(\xi)} J_0(\xi \exp(-\alpha X/2)),$$

and the corresponding standardized variable is

$$C^* = \frac{1}{J_0(\xi)} J_0(\xi \exp(-\alpha X/2)). \quad \blacksquare$$

The probability density function of the principal coordinate axes of the negative exponential r.v. cannot be written as a closed formula. However, from the equality

$$C_j^* = \frac{1}{J_0(\xi_j)} J_0(\xi_j U^{1/2}),$$

where $U=\exp(-\alpha X)$ is a uniform (0, 1) random variable, we can obtain a formal expression for the density g_j which is suitable for numerical computation (see the remark after Proposition 5).

An immediate observation is that in the exponential case, the C_j^*'s are not equally distributed. In particular, the range of values of C_j^* is the interval

$$I_j=\begin{cases}(1/\beta_j, \beta_1/\beta_j), & \text{if } j \text{ is odd,}\\ (\beta_1/\beta_j, 1/\beta_j), & \text{if } j \text{ is even,}\end{cases}$$

where $\beta_j=J_0(\xi_j)$, $j\geqslant 1$. This statement follows from the fact that the maximum value of J_0 is 1, the minimum (i.e., the negative value with maximum absolute value) is $\beta_1=J_0(\xi_1)=-0.402759$, and the sign of β_j is given by the parity of j.

Given $y\in I_j$, we can write $g_j(y)$ as a sum of terms, one for each of the roots of the equation

$$y=\frac{1}{\beta_j} J_0(\xi_j x^{1/2}).$$

In contrast with (13), the number of roots depends now on y. Let us denote this number by $k_j(y)$, and given one of these roots x_i, $1\leqslant i\leqslant k_j(y)$,

$$\left|\frac{dy}{dx}\right|_{x=x_i}=\frac{1}{|\beta_j|}\,|J_1(\zeta_{ij}(y))|\,\frac{\xi_j^2}{2\zeta_{ij}(y)},$$

where $\zeta_{ij}(y)=\xi_j x_i^{1/2}$. Hence

$$g_j(y)=\frac{2\,|\beta_j|}{\xi_j^2}\sum_{i=1}^{k_j(y)}\frac{\zeta_{ij}(y)}{|J_1(\zeta_{ij}(y))|}. \quad \blacksquare$$

Appendix A

The proof of equality (14) amounts to checking that

$$|x-y|=\frac{2}{\pi^2}\sum_{j=1}^{\infty}\frac{(\cos(j\pi x)-\cos(j\pi y))^2}{j^2},$$

which is obtained from the Fourier double series expansion of $|x-y|$ on $[0,1]\times[0,1]$

$$|x-y|=\frac{1}{4}A_{00}+\frac{1}{2}\sum_{m=1}^{\infty}A_{m0}\cos m\pi x+\frac{1}{2}\sum_{n=1}^{\infty}A_{0n}\cos n\pi y$$
$$+\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}A_{mn}\cos m\pi x\cos n\pi y,$$

where $A_{mn} = 4 \int_0^1 \int_0^1 |x-y| \cos m\pi x \cos n\pi y \, dx \, dy$ $(m, n \geqslant 0)$. Further computations give

$$\begin{aligned} A_{00} &= 4/3, & \\ A_{m0} &= 4(m\pi)^{-2} [1 + (-1)^m], & m > 0, \\ A_{0n} &= 4(n\pi)^{-2} [1 + (-1)^n], & n > 0, \\ A_{mn} &= -4(n\pi)^{-2} \delta_{mn}, & m, n > 0. \end{aligned}$$

Deleting null terms, using $\cos 2a = 2\cos^2 a - 1$ and $\sum_{n=1}^{\infty} 1/n^2 = \pi^2/6$, we obtain

$$\begin{aligned} |x-y| &= \frac{1}{3} + 2\,\frac{1}{2}\left(\frac{4}{\pi^2}\sum_{n=1}^{\infty}\frac{\cos^2 n\pi x}{n^2} - \frac{1}{3}\right) - \frac{4}{\pi^2}\sum_{n=1}^{\infty}\frac{\cos n\pi x \cos n\pi y}{n^2} \\ &= \frac{2}{\pi^2}\sum_{n=1}^{\infty}\frac{(\cos n\pi x - \cos n\pi y)^2}{n^2}. \end{aligned}$$

ACKNOWLEDGMENTS

The authors thank J. A. Cuesta-Albertos for useful comments, and J. Burbea for providing the elegant alternative proof of Proposition 4 which replaces our own.

REFERENCES

[1] ANDERSON, T. W., AND DARLING, D. A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Ann. Math. Statist.* **23** 193–212.

[2] ANDERSON, T. W., AND DARLING, D. A. A test of goodness of fit. *J. Amer. Statist. Assoc.* **49** 765–769.

[3] COURANT, R., AND HILBERT, D. (1953). *Methods of Mathematical Physics.* Wiley–Interscience, New York.

[4] CUADRAS, C. M. (1989). Distance analysis in discrimination and classification using both continuous and categorical variables. In *Statistical Data Analysis and Inference* (Y. Dodge, Ed.), pp. 459–473. North-Holland, Amsterdam.

[5] CUADRAS, C. M. (1991). A distance-based approach to discriminant analysis and its properties. *Univ. Barcelona Math. Preprint Ser.* **90**.

[6] CUADRAS, C. M. (1992). Some examples of distance-based discrimination. *Biometrical Lett.* **29**(1) 3–20.

[7] CUADRAS, C. M., AND ARENAS, C. (1990). A distance-based regression model for prediction with mixed data. *Comm. Statist. Theory Methods* **19** 2261–2279.

[8] CUADRAS, C. M., AND FORTIANA, J. (1993). Continuous metric scaling and prediction. In *Multivariate Analysis: Future Directions 2* (C. M. Cuadras and C. R. Rao, Eds.), pp. 47–66. Elsevier, Amsterdam.

[9] GOWER, J. C. (1966). Some distance properties of latent root and vector methods in multivariate analysis. *Biometrika* **53** 315–328.

[10] KAC, M., AND SIEGERT, A. J. F. (1947). An explicit representation of a stationary Gaussian process. *Ann. Math. Statist.* **18** 438–442.

[11] MARDIA, K. V., KENT, J. T., AND BIBBY, J. M. (1979). *Multivariate Analysis.* Academic Press, New York.

[12] RAO, C. R. (1982). Diversity and dissimilarity coefficients: A unified approach. *Theor. Popul. Biol.* **21** 24–43.

[13] SHORACK, G. R., AND WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

[14] TORGERSON, W. S. (1958). *Theory and Methods of Scaling.* Wiley, New York.

[15] WATSON, G. N. (1944). *A Treatise on the Theory of Bessel Functions,* 2nd ed. Cambridge Univ. Press, Cambridge.