Pergamon

Computers Math. Applic. Vol. 31, No. 8, pp. 107–116, 1996

0898-1221/96 \$15.00 + 0.00

S0898-1221(96)00035-1

Some Properties of the Hellinger Transform and Its Application in Classification Problems

I. FAZEKAS
Institute of Mathematics and Informatics, Kossuth University
P.O. Box 12, 4010 Debrecen, Hungary

F. LIESE
Department of Mathematics, University of Rostock
Universitätplatz 1, D-18055, Rostock, Germany

(Received and accepted July 1995)

Abstract—The Hellinger transform for stochastic processes is studied. The Hellinger transform is calculated for multidimensional Gaussian autoregressive processes and for Ornstein-Uhlenbeck processes.

Keywords—Hellinger transform, Radon-Nikodym derivative, Multidimensional Gaussian autoregressive process, Ornstein-Uhlenbeck process.

1. INTRODUCTION AND NOTATION

In this paper, the following general classification problem is studied. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\theta : \Omega \to \{1, \dots, N\}$ a random variable (parameter). Let

$$\pi_i = P(\theta = i) > 0, \qquad i = 1, \dots, N, \tag{1}$$

denote the *a priori* probabilities of classes.

$$P_i(\cdot) = P(\cdot \mid \theta = i), \qquad i = 1, \dots, N, \tag{2}$$

denotes the probability measure in the case of the i^{th} class.

Let μ be a dominating (probability) measure for $P_1, \dots, P_N$ and let X_i denote the density of P_i with respect to μ. The Hellinger transform of $P_1, \dots, P_N$ is defined as follows:

$$\mathbb{H}_{\underline{\alpha}}(P_1, \dots, P_N) = E_\mu \left\{ X_1^{\alpha_1} \cdots X_N^{\alpha_N} \right\}, \tag{3}$$

where $\underline{\alpha} = (\alpha_1, \dots, \alpha_N)$ is a fixed vector with $\alpha_1 \geq 0, \dots, \alpha_N \geq 0$, $\alpha_1 + \dots + \alpha_N = 1$ and E_μ denotes the expectation with respect to μ (see [1]). The Hellinger transform is one of the measures of affinity among several distributions used in the literature (see, e.g., [2,3]). It is easily seen that

$$0 \leq \mathbb{H}_{\underline{\alpha}}(P_1, \dots, P_N) \leq 1 \tag{4}$$

This research was supported by the Hungarian Foundation for Scientific Researches under Grant Number OTKA-1650/1991 and Grant Number OTKA-T-4047/1992.
The authors wish to thank M. Arató for helpful discussions.

Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX

and $\mathbb{H}_{\underline{\alpha}} = 1$ holds iff measures $P_1, \dots, P_N$ coincide. Furthermore, $\mathbb{H}_{\underline{\alpha}}$ is close to zero iff the 'distance' of $P_1, \dots, P_N$ is large in some sense. Therefore $\mathbb{H}_{\underline{\alpha}}$ can be used to measure the possibility of discrimination among several distributions. We mention that many deep investigations are devoted to various 'distances' between two probability measures and their application in hypothesis testing (see, e.g., [4–7]), but it seems that the case of several distributions is not completely known. In [8], the limit (when the number of observations tends to infinity) of the Hellinger transform of Gaussian autoregressive processes is given in terms of spectral densities. We remark that Michálek [9] calculated the asymptotic Rényi's rate for Gaussian processes.

In this paper, upper and lower bounds are given for $\mathbb{H}_{\underline{\alpha}}$ in the case of a filtered probability space (Propositions 2.8 and 2.9). An explicit formula for $\mathbb{H}_{\underline{\alpha}}$ is calculated in the case of multidimensional Gaussian AR processes in Section 3. In Section 4, the Hellinger transform of one-dimensional Ornstein-Uhlenbeck processes is given. In Section 5, the connection between $\mathbb{H}_{\underline{\alpha}}$ and the probability of misclassification is considered.

2. BASIC PROPERTIES OF HELLINGER TRANSFORMS

Let us introduce the notation

$$\underline{X}^{\underline{\alpha}} = X_1^{\alpha_1} \cdots X_N^{\alpha_N}, \tag{5}$$

where $\underline{X} = (X_1, \dots, X_N)$ is a vector with nonnegative coordinates and $\underline{\alpha} = (\alpha_1, \dots, \alpha_N)$ is an arbitrary vector.

DEFINITION 2.1. *(See [1].) Let $\underline{\alpha} = (\alpha_1, \dots, \alpha_N)$, where $\alpha_i \geq 0$ for every i and $\sum_{i=1}^N \alpha_i = 1$. The Hellinger transform of the probability measures $P_1, \dots, P_N$ is defined by the formula*

$$\mathbb{H}_{\underline{\alpha}}(P_1, \dots, P_N) = E_\mu \{X_1^{\alpha_1} \cdots X_N^{\alpha_N}\} = E_\mu \{\underline{X}^{\underline{\alpha}}\}, \tag{6}$$

where μ is a dominating measure for $P_1, \dots, P_N$ and $X_i = \frac{dP_i}{d\mu}$.

For $\underline{\alpha}_0 = (1/N, \dots, 1/N)$, $\mathbb{H}_{\underline{\alpha}_0}(P_1, \dots, P_N)$ is the Matusita affinity or Bhattacharyya coefficient (see [2,3]). For $N = 2$, it is the well-known Hellinger integral. The Hellinger transform is a special case of the f-affinity for $f(s_1, \dots, s_N) = \prod_{i=1}^N s_i^{\alpha_i}$.

DEFINITION 2.2. *Let $f(\underline{s}) = f(s_1, \dots, s_N)$, $s_i \geq 0$, for $i = 1, \dots, N$, be a continuous, concave, homogeneous function. The f-affinity of $P_1, \dots, P_N$ is defined by*

$$A_f(P_1, \dots, P_N) = \int f(X_1, \dots, X_N)\, d\mu. \tag{7}$$

As f is homogeneous, $A_f(P_1, \dots, P_N)$ does not depend on the choice of the dominating measure μ. The f-affinity is nothing else but the negative of the f-dissimilarity of Győrfi and Nemetz [3].

PROPOSITION 2.3. *(See [3].) Let $\mathcal{F}_s$ be a σ-subalgebra of $\mathcal{F}$ and let $P_{1,s}, \dots, P_{N,s}$ denote the restriction of $P_1, \dots, P_N$ on $\mathcal{F}_s$. Then*

$$A_f(P_1, \dots, P_N) \leq A_f(P_{1,s}, \dots, P_{N,s}). \tag{8}$$

For the proof, let $\mu = (P_1 + \cdots + P_N)/N$ and use the Jensen inequality.

COROLLARY 2.4.

$$\mathbb{H}_{\underline{\alpha}}(P_1, \dots, P_N) \leq \mathbb{H}_{\underline{\alpha}}(P_{1,s}, \dots, P_{N,s}). \tag{9}$$

Throughout this section, $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots$ denotes a nondecreasing sequence of σ-subalgebras of $\mathcal{F}$ for which $\sigma(\cup_n \mathcal{F}_n) = \mathcal{F}$. Let $P_{i,n}$, (respectively, μ_n) be the restriction of P_i, (respectively, μ) to $\mathcal{F}_n$ and $X_{i,n} = \frac{dP_{i,n}}{d\mu_n}$, $\underline{X}_n = (X_{1,n}, \dots, X_{N,n})$ $(i = 1, \dots, N,\ n = 0, 1, 2, \dots)$. T will denote a stopping time with respect to $\{\mathcal{F}_n, n = 0, 1, \dots\}$. The following generalization of the approximation theorem of Liese and Vajda [6] holds for Hellinger transforms.

PROPOSITION 2.5. *With the above notation,*

$$\lim_{n\to\infty} A_f(P_{1,n}, \dots, P_{N,n}) = A_f(P_1, \dots, P_N). \tag{10}$$

For the proof, let μ be the same as in the previous proof and use the supermartingale convergence theorem for $f(\underline{X}_n)$.

COROLLARY 2.6. *If $n \to \infty$, then*

$$\mathbb{H}_{\underline{\alpha}}(P_{1,n}, \dots, P_{N,n}) \downarrow \mathbb{H}_{\underline{\alpha}}(P_1, \dots, P_N). \tag{11}$$

We shall use the following multiplicative decomposition of a nonnegative supermartingale.

LEMMA 2.7. *Let $(S_n, \mathcal{F}_n, n = 0, 1, \dots)$ be a nonnegative supermartingale (with $\mathcal{F}_0 = \{\emptyset, \Omega\}$). Then there exist a martingale $(M_n, \mathcal{F}_n)$ and a nonnegative, nonincreasing, predictable sequence $(G_n, \mathcal{F}_n)$ such that*

$$S_n = M_n G_n, \qquad n = 0, 1, \dots. \tag{12}$$

PROOF. Let $M_0 = 1$, $G_0 = S_0$,

$$G_n = \frac{E\{S_n \mid \mathcal{F}_{n-1}\}}{S_{n-1}} G_{n-1}, \qquad n \geq 1 \tag{13}$$

(here we used the convention $1/0 = 0$). Let $M_n = S_n/G_n$ if $G_n \neq 0$ and $M_n = M_{n-1}$ if $G_n = 0$. Then M_n is a martingale. Since $G_n(\omega) = 0$ implies that $S_n(\omega) = 0$, then (12) is satisfied.

In the sequel, we use the notation

$$S_n(\underline{\alpha}) = \underline{X}_n^{\underline{\alpha}}. \tag{14}$$

Since $S_n(\underline{\alpha})$ is a nonnegative supermartingale, it admits a multiplicative decomposition

$$S_n(\underline{\alpha}) = M_n(\underline{\alpha}) G_n(\underline{\alpha}), \tag{15}$$

where $G_n(\underline{\alpha})$ is defined analogously to (13).

We derive inequalities between $\mathbb{H}_{\underline{\alpha}}(P_1, \dots, P_N)$ and $G_n(\underline{\alpha}')$ for an appropriate $\underline{\alpha}'$. Let $\underline{\alpha} = (\alpha_1, \dots, \alpha_N)$ be fixed. For any i $(1 \leq i \leq N)$ and γ, we associate with $\underline{\alpha}$ a new vector

$$[\underline{\alpha}, i, \gamma] = (\gamma\alpha_1, \dots, \gamma\alpha_{i-1}, 1 - \gamma(1 - \alpha_i), \gamma\alpha_{i+1}, \dots, \gamma\alpha_N). \tag{16}$$

PROPOSITION 2.8. *Let us assume that*

$$1 < \gamma_i < \min\left\{\frac{1}{\alpha_i}, \frac{1}{1 - \alpha_i}\right\} \tag{17}$$

is satisfied for γ_i for every $i = 1, \dots, N$. Then for every stopping time T,

$$\mathbb{H}_{\underline{\alpha}}(P_{1,T}, \dots, P_{N,T}) \leq \min_{1\leq i\leq N} \left(E_{P_i}\left\{G_T^{1/(\gamma_i-1)}([\underline{\alpha}, i, \gamma_i])\right\}\right)^{(\gamma_i-1)/\gamma_i}. \tag{18}$$

PROOF. We prove (18) for $i = 1$. Let $\gamma = \gamma_1$ satisfy (17). By the multiplicative decomposition, we can write

$$\left(X_{1,T}^{\alpha_1} \cdots X_{N,T}^{\alpha_N}\right)^{\gamma} = X_{1,T}^{\gamma-1} M_T([\alpha, 1, \gamma]) G_T([\underline{\alpha}, 1, \gamma]). \tag{19}$$

Hence

$$E_\mu \left\{X_{1,T}^{\alpha_1} \cdots X_{N,T}^{\alpha_N}\right\} = E_\mu \left\{X_{1T}^{(\gamma-1)/\gamma} G_T^{1/\gamma}([\underline{\alpha},1,\gamma]) M_T^{1/\gamma}([\underline{\alpha},1,\gamma])\right\}. \tag{20}$$

Applying Hölders inequality with $p = \gamma/(\gamma-1)$, $q = \gamma$, we get

$$E_\mu \left\{X_{1,T}^{\alpha_1} \cdots X_{N,T}^{\alpha_N}\right\} \le \left[E_\mu \left\{M_T([\underline{\alpha},1,\gamma])\right\}\right]^{1/\gamma} \left[E_\mu \left\{X_{1,T} G^{1/(\gamma-1)}([\underline{\alpha},1,\gamma])\right\}\right]^{(\gamma-1)/\gamma}. \tag{21}$$

One can easily derive that $E_\mu\{X_{1,T}\xi\} = E_{P_1}\{\xi\}$ for every $\mathcal{F}_T$-measurable ξ. By the proof of Lemma 2.7, we have $E_\mu\{M_n\} = 1$ for the martingale part of the multiplicative decomposition, so $E_\mu\{M_T\} \le 1$. Therefore, (21) implies that

$$\mathbb{H}_{\underline{\alpha}}(P_{1,T},\dots,P_{N,T}) \le \left[E_{P_1}\left\{G_T^{1/(\gamma-1)}([\underline{\alpha},1,\gamma])\right\}\right]^{(\gamma-1)/\gamma}. \tag{22}$$

To prove the following inequality, we need the inverse Hölder inequality: if ξ and η are positive random variables, $0 < p < 1$, $1/q + 1/p = 1$, then

$$E\{\xi\eta\} \ge \left(E\{\xi^p\}\right)^{1/p} \left(E\{\eta^q\}\right)^{1/q}. \tag{23}$$

PROPOSITION 2.9. *Suppose that the probability measures $P_1,\dots,P_N$ are equivalent and let μ be chosen such that μ and P_i are equivalent. Then for every stopping time T*

$$\mathbb{H}_{\underline{\alpha}}(P_{1,T},\dots,P_{N,T}) \ge \max_{1\le i\le N} \left[E_{P_i}\left\{G_T^{1/(\gamma_i-1)}([\underline{\alpha},i,\gamma_i])\right\}\right]^{(\gamma_i-1)/\gamma_i}, \tag{24}$$

where $0 < \gamma_i < 1$ for every $i = 1,\dots,N$.

PROOF. First we remark that equivalence of μ and P_i implies that $X_{i,n} > 0$. Therefore $G_n(\cdot) > 0$ and $M_n(\cdot) > 0$. Now, let T be a bounded stopping time. Let $\gamma = \gamma_1$ be fixed. Easy calculation and the inverse Hölder inequality (with $p = \gamma$) show that

$$\mathbb{H}_{\underline{\alpha}}(P_{1,T},\dots,P_{N,T}) = E_\mu \left\{(X_{1,T})^{(\gamma-1)/\gamma} G_T^{1/\gamma}([\underline{\alpha},1,\gamma]) M_T^{1/\gamma}([\underline{\alpha},1,\gamma])\right\} \tag{25}$$

$$\ge \left[E_\mu \left\{X_{1,T} G_T^{1/(\gamma-1)}([\underline{\alpha},1,\gamma])\right\}\right]^{(\gamma-1)/\gamma} \left[E_\mu\left\{M_T([\underline{\alpha},1,\gamma])\right\}\right]^{1/\gamma} \tag{26}$$

$$= \left[E_{P_1}\left\{G_T^{1/(\gamma-1)}([\underline{\alpha},1,\gamma])\right\}\right]^{(\gamma-1)/\gamma}. \tag{27}$$

This proves (24) for bounded stopping times. If T is an arbitrary stopping time, then approximate it by $T \wedge n$ and use Corollary 2.6.

REMARK 2.10. $G_n(\underline{\alpha})$ can easily be calculated for Markov processes. Let $y_1, y_2, \dots$ be a homogeneous Markov process with transition density function $p_i(\cdot,\cdot)$ and initial density function $p_i(\cdot)$ in the case $\theta = i$. Then

$$G_n(\underline{\alpha}) = \prod_{l=2}^{n} \left[\int_{-\infty}^{+\infty} \prod_{i=1}^{N} p_i^{\alpha_i}(y_{l-1}, v_l)\, dv_l\right] \cdot \left(\prod_{i=1}^{N} p_i^{\alpha_i}(y_1)\right) \cdot \int_{-\infty}^{+\infty} \prod_{i=1}^{N} p_i^{\alpha_i}(v_1)\, dv_1. \tag{28}$$

If, moreover, one of the measures P_i, $i = 1,\dots,N$, say P_1, is such that $y_1, y_2, \dots$ are independent, then inequality (18) is quite simple because if we take expectation of G_n with respect to P_1 and use (28) then the expectation sign E_{P_1} can be interchanged with the product sign with respect to l.

3. THE HELLINGER TRANSFORM FOR GAUSSIAN AR PROCESSES

Considerations in this section have been inspired by those of Pap, Bokor and Gáspár [10]. Let $z_{(k)} = (z_0^\top, \dots, z_k^\top)^\top$, where $z_0, \dots, z_k \in \mathbb{R}^m$ and $^\top$ denotes the transpose of a vector. Let $\mathcal{N}_m(\mu, C)$ denote the m-dimensional normal distribution with mean vector μ and covariance matrix C.

Let us suppose that the measure P_i is generated by an m-dimensional Gaussian AR(1) process. More precisely, we assume that the observed process belongs to the model

$$y_i(l+1) = A_i y_i(l) + w_i(l+1), \qquad l = 0, 1, 2, \dots, \tag{29}$$

if $\theta = i$, $i = 1, \dots, N$. Here $\{w_i(l),\ l = 1, 2, \dots\}$ is a Gaussian white noise with covariance Q_i; i.e., $w_i(l) \sim \mathcal{N}_m(0, Q_i)$ and $w_i(1), w_i(2), \dots$ are i.i.d. Suppose that the matrices A_i and Q_i are known and Q_i is invertible for every $i = 1, \dots, N$. For the i^{th} model, let the initial distribution be Gaussian with mean zero and with invertible covariance matrix D_i; i.e., $y_i(0) \sim \mathcal{N}_m(0, D_i)$ $(i = 1, \dots, N)$. Suppose that the process is observed until time k. P_i is the distribution of $y_i(0), \dots, y_i(k)$ (in the case of the i^{th} model, i.e., when $\theta = i$).

Then the Hellinger transform is

$$\mathbb{H}_{\underline{\alpha}}(P_1, \dots, P_N) = \mathbb{H}_k = \int \prod_{i=1}^N p\left(z_{(k)} \mid H_i\right)^{\alpha_i} \, dz_{(k)}, \tag{30}$$

where

$$p\left(z_{(k)} \mid H_i\right) = (2\pi)^{-((k+1)m)/2} \left(\det C_i\right)^{-1/2} \exp\left\{-\frac{1}{2} z_{(k)}^\top C_i^{-1} z_{(k)}\right\}, \tag{31}$$

for $z_{(k)} \in \mathbb{R}^{(k+1)m}$. (To emphasize that $\mathbb{H}_{\underline{\alpha}}(P_1, \dots, P_N)$ depends on the number of observations $y_i(0), \dots, y_i(k)$, we shall write $\mathbb{H}_k$ instead of $\mathbb{H}_{\underline{\alpha}}(P_1, \dots, P_N)$. Obviously, P_i and C_i depend on k, too.)

Using the fact that the integral of a density function is equal to 1 (for the density of $\mathcal{N}_{(k+1)m}$ $(0, (\sum_{i=1}^N \alpha_i C_i^{-1})^{-1})$), we obtain

$$\begin{aligned} \mathbb{H}_k &= \left(\det\left(\sum_{i=1}^N \alpha_i C_i^{-1}\right)^{-1}\right)^{1/2} \prod_{i=1}^N \left(\det C_i\right)^{-\alpha_i/2} \\ &= \left(\det\left(\sum_{i=1}^N \alpha_i C_i^{-1}\right)\right)^{-1/2} \prod_{i=1}^N \left(\det C_i^{-1}\right)^{\alpha_i/2}. \end{aligned} \tag{32}$$

Now, C_i is the following matrix

$$\begin{pmatrix} B_i^{(0)} & B_i^{(0)} A_i^\top & B_i^{(0)} A_i^{\top 2} & \cdots & B_i^{(0)} A_i^{\top k} \\ A_i B_i^{(0)} & B_i^{(1)} & B_i^{(1)} A_i^\top & \cdots & B_i^{(1)} A_i^{\top(k-1)} \\ A_i^2 B_i^{(0)} & A_i B_i^{(1)} & B_i^{(2)} & \cdots & B_i^{(2)} A_i^{\top(k-2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_i^k B_i^{(0)} & A_i^{k-1} B_i^{(1)} & A_i^{k-2} B_i^{(2)} & \cdots & B_i^{(k)} \end{pmatrix}, \tag{33}$$

where $B_i^{(j)} = A_i^j D_i A_i^{\top j} + \sum_{l=0}^{j-1} A_i^l Q_i A_i^{\top l}$. The inverse of C_i is

$$C_i^{-1} = \begin{pmatrix} D_i^{-1} + A_i^\top Q_i^{-1} A_i & -A_i^\top Q_i^{-1} & 0 & \cdots & 0 \\ -Q_i^{-1} A_i & Q_i^{-1} + A_i^\top Q_i^{-1} A_i & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & Q_i^{-1} + A_i^\top Q_i^{-1} A_i & -A_i^\top Q_i^{-1} \\ 0 & \cdots & 0 & -Q_i^{-1} A_i & Q_i^{-1} \end{pmatrix}, \tag{34}$$

and $C_i^{-1} = F_i F_i^\top$, where

$$F_i = \begin{pmatrix} D_i^{-1/2} & -A_i^\top Q_i^{-1/2} & 0 & \cdots & 0 \\ 0 & Q_i^{-1/2} & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -A_i^\top Q_i^{-1/2} \\ 0 & \cdots & 0 & 0 & Q_i^{-1/2} \end{pmatrix}. \tag{35}$$

We remark that the above matrices depend on k. We obtain

$$\mathbb{H}_k = \left(\det\left(\sum_{i=1}^N \alpha_i C_i^{-1}\right)\right)^{-1/2} \prod_{i=1}^N \left((\det Q_i)^k \det D_i\right)^{-\alpha_i/2}. \tag{36}$$

Here $\sum_{i=1}^N \alpha_i C_i^{-1}$ has the form

$$\sum_{i=1}^N \alpha_i C_i^{-1} = \begin{pmatrix} D+d_2 & -d_1 & 0 & \cdots & 0 \\ -d_1^\top & d_0+d_2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & d_0+d_2 & -d_1 \\ 0 & \cdots & 0 & -d_1^\top & d_0 \end{pmatrix}, \tag{37}$$

where

$$d_0 = \sum \alpha_i Q_i^{-1}, \quad d_1 = \sum \alpha_i Q_i^{-1} A_i, \quad d_2 = \sum \alpha_i A_i^\top Q_i^{-1} A_i, \quad D = \sum \alpha_i D_i^{-1}. \tag{38}$$

Eliminating the elements over the diagonal blocks, we get a matrix with diagonal blocks $M(1), \dots, M(k+1)$, where

$$M(1) = d_0, \tag{39}$$

$$M(l+1) = d_0 + d_2 - d_1^\top M(l)^{-1} d_1, \qquad l = 1, \dots, k-1, \tag{40}$$

$$M(k+1) = D + d_2 - d_1^\top M(k)^{-1} d_1. \tag{41}$$

Therefore (36) implies that

$$\mathbb{H}_k = \prod_{i=1}^N (\det D_i)^{-\alpha_i/2} \left[\prod_{i=1}^N (\det Q_i)^{-\alpha_i/2}\right]^k \left(\prod_{l=1}^{k+1} \det M(l)\right)^{-1/2}. \tag{42}$$

Here $M(1) = d_0$ is positive definite. Furthermore, one can prove that $M(2) \geq M(1)$ with equality iff $A_1 = \dots = A_N$. To prove this, we need the following lemmas. (Lemmas 3.1 and 3.2 were communicated to the authors by Losonczi and Páles. For the operator version of these lemmas, see [11].)

LEMMA 3.1. *Let U be a symmetric positive definite matrix. Then*

$$x^\top U^{-1} x + z^\top U z \geq x^\top z + z^\top x, \tag{43}$$

for each pair of vectors x, z. Equality holds in (43) if and only if $x = Uz$.

PROOF. (43) can be obtained from the inequality $\|U^{-1/2}x - U^{1/2}z\| \geq 0$, where $\|\cdot\|$ denotes the Euclidean norm.

LEMMA 3.2. *Let $V_1, V_2, \dots, V_N$ be symmetric positive definite matrices and let $V = V_1 + \cdots + V_N$. Then*

$$\sum_{i=1}^{N} x_i^\top V_i^{-1} x_i \geq \left(\sum_{i=1}^{N} x_i\right)^\top V^{-1} \left(\sum_{i=1}^{N} x_i\right), \tag{44}$$

for arbitrary vectors $x_1, \dots, x_N$. Equality holds in (43) if and only if $V_1^{-1}x_1 = V_2^{-1}x_2 = \cdots = V_N^{-1}x_N$.

PROOF. For $i = 1, \dots, N$, put $U = V_i$, $x = x_i$, and $z = V^{-1}(\sum_{j=1}^{N} x_j)$ in (43) and sum the inequalities obtained. We get

$$\sum_{i=1}^{N} x_i^\top V_i^{-1} x_i + z^\top \left(\sum_{j=1}^{N} x_j\right) \geq \sum_{i=1}^{N} x_i^\top V^{-1} \sum_{j=1}^{N} x_j + \sum_{j=1}^{N} z^\top x_j, \tag{45}$$

which is equivalent to (44). The condition for equality is $x_i = V_i z$; that is, $V^{-1}x_i$ is independent of i.

Now, we can return to the proof of $M(2) \geq M(1)$. We have

$$M(2) - M(1) = \sum_{i=1}^{N} \alpha_i A_i^\top Q_i^{-1} A_i - \left(\sum_{i=1}^{N} \alpha_i Q_i^{-1} A_i\right)^\top \left(\sum_{i=1}^{N} \alpha_i Q_i^{-1}\right)^{-1} \left(\sum_{i=1}^{N} \alpha_i Q_i^{-1} A_i\right).$$

Therefore, with notation $V_i = \alpha_i Q_i^{-1}$, we have

$$M(2) - M(1) = \sum_{i=1}^{N} \left(A_i^\top V_i\right) V_i^{-1} \left(V_i A_i\right) - \left(\sum_{i=1}^{N} V_i A_i\right)^\top \left(\sum_{i=1}^{N} V_i\right)^{-1} \left(\sum_{i=1}^{N} V_i A_i\right) \geq 0.$$

In the last step, we used (44) with $x_i = V_i A_i a$, where a is an arbitrary vector. Equality holds if $V_i^{-1} x_i = A_i a$ does not depend on i for every a. As

$$M(l+1) - M(l) = d_1^\top \left(M(l-1)^{-1} - M(l)^{-1}\right) d_1, \qquad l = 2, \dots, k-1, \tag{46}$$

it follows that

$$0 < d_0 = M(1) \leq M(2) \leq \cdots \leq M(k) \leq d_0 + d_2 \tag{47}$$

with equality iff $A_1 = \cdots = A_N$.

Substituting $M(l)$, $l = 2, \dots, k$, by $M(1)$ in equality (42), we get the estimation

$$\mathbb{H}_k \leq \text{const} \cdot c^{k/2}, \tag{48}$$

where

$$c = \prod_{i=1}^{N} \frac{\left(\det Q_i^{-1}\right)^{\alpha_i}}{\det\left(\sum_{i=1}^{N} \alpha_i Q_i^{-1}\right)}. \tag{49}$$

$c \leq 1$ with equality iff $Q_1 = \cdots = Q_N$. This estimation cannot be improved if $A_1 = \cdots = A_N$. Otherwise, using $M(l)$, $2 \leq l \leq k$, instead of $M(1)$, one can get a better estimation than (48),(49). In the special case when $Q_1 = \cdots = Q_N = Q$, one can use, e.g., $M(2)$ (for which $M(2) \geq M(1) \geq Q$ with equality iff $A_1 = \cdots = A_N$) to obtain the estimation

$$\mathbb{H}_k \leq \text{const} \cdot \left[\frac{\det Q}{\det M(2)}\right]^{k/2}. \tag{50}$$

We mention two possible choices of the initial distribution. For the first, if D_i is the solution of the equation $D_i = A_i D_i A_i^\top + Q_i$, $i = 1, \dots, N$, then we have stationary AR(1)-processes. For the second, if all the D_i's are equal to the dispersion of the stationary solution of model $\theta = 1$, then one can study a possible change at time 0 from the first model for one of the other models.

4. THE HELLINGER TRANSFORM OF CONTINUOUS TIME PROCESSES

First we consider one-dimensional continuous time AR(1) or Ornstein-Uhlenbeck processes. Let $y_i(t)$, $0 \le t \le T$, be the solution of the stochastic differential equation

$$dy_i(t) = -\lambda_i y_i(t)\,dt + dw(t), \qquad 0 \le t \le T, \tag{51}$$

where $w(t)$ is a standard Wiener process and $\lambda_i > 0$, $i = 0, 1, \dots, N$. Suppose that $y_i(t)$ is stationary ($y_i(0) \sim \mathcal{N}(0, 1/2\lambda_i)$). Then $Ey_i(t) = 0$ and $\operatorname{cov}(y_i(t), y_i(t+s)) = \exp[-\lambda_i s]/2\lambda_i$.

It follows from formula (2.3.41) of [12] that the Radon-Nikodym derivative with respect to P_{λ_0} has the form

$$\frac{dP_{\lambda_i}}{dP_{\lambda_0}}(y(t)) = \sqrt{\frac{\lambda_i}{\lambda_0}} \exp\left\{ \frac{\lambda_0^2 - \lambda_i^2}{2} \int_0^T y^2(t)\,dt + \frac{\lambda_i - \lambda_0}{2} T - \frac{\lambda_i - \lambda_0}{2} \left[y^2(T) + y^2(0)\right] \right\}, \tag{52}$$

where P_{λ_i} denotes the measure generated by the process given by equation (51). Then the Hellinger transform of measures $P_{\lambda_1}, \dots, P_{\lambda_N}$ is

$$\begin{aligned} \mathbb{H}_{\underline{\alpha}}\left(P_{\lambda_1}, \dots, P_{\lambda_N}\right) &= E_0 \left\{ \prod_{i=1}^N \left(\frac{dP_{\lambda_i}}{dP_{\lambda_0}} \right)^{\alpha_i} \right\} \\ &= E_0 \left\{ \sqrt{\frac{\prod_{i=1}^N \lambda_i^{\alpha_i}}{\lambda_0}} \exp\left[\frac{\lambda_0^2 - \sum_{i=1}^N \alpha_i \lambda_i^2}{2} \int_0^T y^2(t)\,dt \right.\right. \\ &\qquad \left.\left. + \frac{\sum_{i=1}^N \alpha_i \lambda_i - \lambda_0}{2} \left[T - \left(y^2(T) + y^2(0)\right)\right] \right] \right\}, \end{aligned} \tag{53}$$

where E_0 denotes that the expectation is taken with respect to the measure defined by the process determined by equation (51) with parameter λ_0. Using Novikov's method (see, e.g., [12–14]), put $\lambda_0^2 = \sum_{i=1}^N \alpha_i \lambda_i^2$. Then we obtain

$$\begin{aligned} &\mathbb{H}_{\underline{\alpha}}\left(P_{\lambda_1}, \dots, P_{\lambda_N}\right) \\ &= \sqrt{\frac{\prod_{i=1}^N \lambda_i^{\alpha_i}}{\lambda_0}} \exp\left[\frac{\sum_{i=1}^N \alpha_i \lambda_i - \lambda_0}{2} T \right] E_0 \left\{ \exp\left[-\frac{\sum_{i=1}^N \alpha_i \lambda_i - \lambda_0}{2} \left(y^2(T) + y^2(0)\right) \right] \right\}. \end{aligned} \tag{54}$$

Now we need the following fact. If the common distribution of ξ_1 and ξ_2 is normal with zero mean and $D^2\xi_1 = \sigma_1^2$, $D^2\xi_2 = \sigma_2^2$, $\operatorname{cov}(\xi_1, \xi_2) = \varrho$, then (the Laplace transform)

$$E\left\{ \exp\left[-\frac{t\left(\xi_1^2 + \xi_2^2\right)}{2} \right] \right\} = \left[\left(t\sigma_1^2 + 1\right)\left(t\sigma_2^2 + 1\right) - t^2\varrho^2 \right]^{-1/2}. \tag{55}$$

Therefore equation (54) implies the following proposition.

PROPOSITION 4.1. *The Hellinger transform of Ornstein-Uhlenbeck processes defined by (51) is*

$$\begin{aligned} \mathbb{H}_{\underline{\alpha}}\left(P_{\lambda_1}, \dots, P_{\lambda_N}\right) &= 2\sqrt{\lambda_0 \prod_{i=1}^N \lambda_i^{\alpha_i} \exp\left[\left(\sum_{i=1}^N \alpha_i \lambda_i - \lambda_0 \right) \frac{T}{2} \right]} \\ &\times \left[\left(\sum_{i=1}^N \alpha_i \lambda_i + \lambda_0 \right)^2 - \left(\sum_{i=1}^N \alpha_i \lambda_i - \lambda_0 \right)^2 \exp(-2\lambda_0 T) \right]^{-1/2}, \end{aligned} \tag{56}$$

where $\lambda_0^2 = \sum_{i=1}^N \alpha_i \lambda_i^2$.

Now we consider one-dimensional shifted Wiener processes. Let $y_i(t)$, $0 \le t \le T$, be defined by the equation

$$y_i(t) = m_i(t) + w(t), \qquad 0 \le t \le T, \tag{57}$$

$i = 0, 1, \dots, N$, where $w(t)$ is a standard Wiener process. Suppose that $m_i(0) = 0$ and $m_i(t)$ is absolute continuous with derivative $m_i'(t)$ for which $\int_0^T (m_i'(t))^2\, dt < \infty$. Then it follows from [12, formula (2.3.17)] that the Radon-Nikodym derivative with respect to P_{m_0} has the form

$$\begin{aligned}\frac{dP_{m_i}}{dP_{m_0}}(y(t)) = \exp\Bigg\{ &-\frac{1}{2}\int_0^T (m_i'(t))^2\, dt + \int_0^T m_i'(t)\, dy(t) \\ &+\frac{1}{2}\int_0^T (m_0'(t))^2\, dt - \int_0^T m_0'(t)\, dy(t)\Bigg\},\end{aligned} \tag{58}$$

where P_{m_i} denotes the measure generated by the process given by equation (57). Then the Hellinger transform of measures $P_{m_1}, \dots, P_{m_N}$ is

$$\begin{aligned}\mathbb{H}_{\underline{\alpha}}(P_{m_1}, \dots, P_{m_N}) = E_0 \Bigg\{ \exp\Bigg[&-\frac{1}{2}\int_0^T \sum_{i=1}^N \alpha_i (m_i'(t))^2\, dt \\ &+ \int_0^T \sum_{i=1}^N \alpha_i m_i'(t)\, dy(t) + \frac{1}{2}\int_0^T (m_0'(t))^2\, dt - \int_0^T m_0'(t)\, dy(t) \Bigg]\Bigg\},\end{aligned} \tag{59}$$

where E_0 denotes that the expectation is taken with respect to the measure P_{m_0}. Using Novikov's method, put $m_0'(t) = \sum_{i=1}^N \alpha_i m_i'(t)$. Then we obtain

$$\mathbb{H}_{\underline{\alpha}}(P_{m_1}, \dots, P_{m_N}) = \exp\left\{ -\frac{1}{2}\int_0^T \sum_{i=1}^N \alpha_i (m_i'(t))^2\, dt + \frac{1}{2}\int_0^T \left(\sum_{i=1}^N \alpha_i m_i'(t) \right)^2 dt \right\}. \tag{60}$$

Therefore we have the following proposition.

PROPOSITION 4.2. *If $m_i(t)$, $i = 1, \dots, N$, satisfy conditions listed above then the Hellinger transform of the processes defined by (57) is given by (60).*

5. APPLICATIONS TO THE PROBABILITY OF MISCLASSIFICATION

Consider the classification problem introduced in Section 1 by (1),(2). A classification rule is a measurable mapping

$$d : \Omega \to \{1, \dots, N\}. \tag{61}$$

$\theta = i$ is accepted on the event $\{d = i\}$. The Bayesian risk connected with the decision rule d is given by

$$R(d) = P\{\theta \ne d\} = \sum_{i=1}^N \pi_i P_i\{d \ne i\}. \tag{62}$$

It is well known that a decision rule d_0 with

$$\{d_0 \ne i\} \supseteq \{\pi_i X_i < \max(\pi_1 X_1, \dots, \pi_N X_N)\} \tag{63}$$

for $i = 1, \dots, N$ minimizes the risk (i.e., the probability of error). Furthermore, the resulting probability of error is

$$P_e = R(d_0) = 1 - \int \max_{1 \le i \le n} (\pi_i X_i)\, d\mu. \tag{64}$$

One can easily give an upper bound on P_e in terms of Hellinger integrals:

$$P_e \le \sum_{i<j} \pi_i^{\alpha_1^{(ij)}} \pi_j^{\alpha_2^{(ij)}} \mathbb{H}_{\underline{\alpha}^{(ij)}}(P_i, P_j) \tag{65}$$

(here $\underline{\alpha}^{(ij)} = (\alpha_1^{(ij)}, \alpha_2^{(ij)})$ is a vector with nonnegative coordinates and $\alpha_1^{(ij)} + \alpha_2^{(ij)} = 1$). The following inequality can be proved by the same way as the theorem of Bhattacharya and Toussaint [2]:

$$P_e \le \frac{N-2}{N-1} + \frac{1}{N-1}\left(\prod_{i=1}^{N} \pi_i^{\alpha_i}\right) \mathbb{H}_{\underline{\alpha}}(P_1, \dots, P_N). \tag{66}$$

REFERENCES

1. J. Jacod, Filtered statistical models and Hellinger processes, *Stoch. Proc. Appl.* **32**, 3–45 (1989).
2. B.K. Bhattacharya and G.T. Toussaint, An upper bound on the probability of misclassification in terms of Matusita's measure of affinity, *Ann. Inst. Statist. Math.* **34**, 161–165 (1982).
3. L. Győrfi and T. Nemetz, f-dissimilarity: A general class of separation measures of several probability measures, In *Topics of Information Theory*, (Edited by I. Csiszár and P. Elias), pp. 309–321, North-Holland, (1977).
4. I. Csiszár, Information type measures of difference of probability distributions and indirect observations, *Studia Sci. Math. Hungar.* **2**, 299–318 (1967).
5. J. Jacod and A.N. Shiryaev, *Limit Theorems for Stochastic Processes*, Springer, Berlin, (1987).
6. F. Liese and I. Vajda, *Convex Statistical Distances*, Volume 95, Teubuer-Texte zur Mathematik, Leipzig, (1987).
7. H. Chernoff, A measure of asymptotic efficiency for tests of a hyphothesis based on the sum of observations, *Ann. Math. Statist.* **23**, 497–507 (1952).
8. I. Fazekas, Hellinger transform of Gaussian autoregressive processes, *Computers Math. Applic.* **27** (7), 15–21 (1994).
9. J. Michálek, Asymptotic Rényi's rate of Gaussian processes, *Problems of Control and Information Theory* **19** (3), 209–227 (1990).
10. G. Pap, J. Bokor and P. Gáspár, Analysis of multiple model method for change detection of AR processes, *Computers Math. Applic.* **19** (1), 115–121 (1990).
11. L. Losonczi and Zs. Páles, A simple proof for a quadratic inequality, *International Series of Numerical Mathematics* **103**, 445–447 (1992).
12. M. Arató, *Linear Stochastic Systems with Constant Coefficients. A Statistical Approach* (in Russian), Nauka, Moscow, (1989); (English edition: *Lecture Notes in Control and Inf.*, Volume 45, Springer-Verlag, Berlin, (1982).
13. A. Novikov, On the estimation of parameters of diffusion processes, *Studia Sci. Math. Hungar.* **7**, 201–209 (1972).
14. G. Pap, The distribution of estimates of parameters of multidimensional stationary AR processes, *Computers Math. Applic.* **27** (2), 1–8 (1994).
15. B. Grigelionis, Hellinger processes and Hellinger integrals for solutions of martingale problems, In *Proc. 5th Vilnius Conference on Probability Theory and Mathematical Statistics*, Volume 1, (1990).
16. B. Grigelionis, On Hellinger transforms for solutions of martingale problems, In *Stochastic Processes; a Festschrift in Honour of Gopinath Kallianpur*, (Edited by C. Stomatis *et al.*), pp. 107–116, Springer-Verlag, Berlin, (1993).