JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 138, 280–292 (1989)

Measures of Distance between Probability Distributions*

J. K. CHUNG, PL. KANNAPPAN, C. T. NG, AND P. K. SAHOO

Centre for Information Theory, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

Submitted by David Aldous

Received July 6, 1987

In statistical estimation problems measures between probability distributions play significant roles. Hellinger coefficient, Jeffreys distance, Chernoff coefficient, directed divergence, and its symmetrization J-divergence are examples of such measures. Here these and like measures are characterized through a composition law and the sum form they possess. The functional equations

$$f(pr, qs) + f(ps, qr) = (r+s) f(p, q) + (p+q) f(r, s)$$

and

$$f(pr, qs) + f(ps, qr) = f(p, q) f(r, s)$$

are instrumental in their deduction.

1. INTRODUCTION

Let $\Gamma_n^0 = \{P = (p_1, p_2, ..., p_n) \mid 0 < p_k < 1, \sum_{k=1}^n p_k = 1\}$ denote the set of all discrete n-ary complete positive probability distributions. Let $I_0 =]0, 1[$, $\mathbf{R}_+ =]0, \infty[$, $\mathbf{R}$ be the set of real numbers and $\mathbf{C}$ be the set of complex numbers. The following are some known measures between two probability distributions P and Q in Γ_n^0:

(a) Hellinger coefficient [4, 9]

$$H_n(P, Q) = \sum_{k=1}^n \sqrt{p_k q_k}, \tag{1.1}$$

(b) Chernoff coefficient [5, 11]

$$C_{n,\alpha}(P \| Q) = \sum_{k=1}^n p_k^\alpha q_k^{1-\alpha}, \qquad \alpha \in I_0, \tag{1.2}$$

* This work is supported by NSERC of Canada grants.

0022-247X/89 $3.00

(c) Jeffreys distance [10]

$$K_n(P, Q) = \sum_{k=1}^{n} (\sqrt{p_k} - \sqrt{q_k})^2. \tag{1.3}$$

These measures have many applications in statistics, pattern recognition, and numerical taxonomy. In information theory, for P and Q in Γ_n^0, the directed divergence [15] was defined as

$$I_n(P \| Q) = \sum_{k=1}^{n} p_k \log_2 \frac{p_k}{q_k}, \tag{1.4}$$

and was characterized in [12, 18]. It serves as a separability measure and is frequently used in statistics [15] and pattern recognition [22, 24]. However, the directed divergence is not symmetric. In [15], *symmetric divergence* was introduced as

$$J_n(P, Q) = I_n(P \| Q) + I_n(Q \| P) \tag{1.5}$$

to restore the symmetry. It has the explicit form

$$J_n(P, Q) = \sum_{k=1}^{n} (p_k - q_k) \log_2 \frac{p_k}{q_k}, \tag{1.6}$$

and is also called the J-divergence in honor of Jeffreys [10] who first used this measure in some estimation problems. It is non-negative and attains minimum when $P = Q$. Moreover, it satisfies the *composition law*

$$J_{nm}(P * R, Q * S) + J_{nm}(P * S, Q * R) = 2J_n(P, Q) + 2J_m(R, S) \tag{1.7}$$

for all $P, Q \in \Gamma_n^0$ and $R, S \in \Gamma_m^0$ where

$$P * R = (p_1 r_1, p_1 r_2, ..., p_1 r_m, p_2 r_1, ..., p_2 r_m, ..., p_n r_m).$$

Evidently J_n has the sum form [12, 16, 18]. It was characterized in [14] through recursivity, symmetry, differentiability, and some normalization conditions. Applications of J_n to statistics may be found in [10, 11, 15, 19, 21], to pattern recognition in [22, 24], and to questionnaire analysis in [8]. The directed divergence of degree α defined as

$$I_{n,\alpha}(P \| Q) = (2^{\alpha-1} - 1)^{-1} \left(\sum_{k=1}^{n} p_k^\alpha q_k^{1-\alpha} - 1 \right), \tag{1.8}$$

where $\alpha \in \mathbf{R} - \{1\}$, was characterized in [17]. Like directed divergence, this measure is not symmetric either. In [21], thus the symmetric divergence (J-divergence) of degree α was defined as

$$J_{n,\alpha}(P, Q) = I_{n,\alpha}(P \| Q) + I_{n,\alpha}(Q \| P)$$

which indeed by (1.8) equals

$$J_{n,\alpha}(P, Q) = \frac{\sum_{k=1}^{n} (p_k^\alpha q_k^{1-\alpha} + p_k^{1-\alpha} q_k^\alpha) - 2}{2^{\alpha-1} - 1}. \tag{1.9}$$

It is a one parameter generalization of (1.6) since (1.9) tends to (1.6) as $\alpha \to 1$. Some of its properties can be found in [21]. In addition to having the sum form it also satisfies the following composition law

$$\begin{aligned} J_{nm,\alpha}(P * R, Q * S) &+ J_{nm,\alpha}(P * S, Q * R) \\ &= 2J_{n,\alpha}(P, Q) + 2J_{m,\alpha}(R, S) + \lambda J_{n,\alpha}(P, Q)\, J_{m,\alpha}(R, S), \end{aligned} \tag{1.10}$$

where $\lambda = 2^{\alpha-1} - 1$. This property will be called the *symmetric compositivity* of $J_{n,\alpha}$. If $\alpha \to 1$, then $\lambda \to 0$ and (1.10) yields (1.7).

Let $\mu_n: \Gamma_n^0 \times \Gamma_n^0 \to \mathbf{R} (n \geqslant 2)$ be *symmetric*, that is, $\mu_n(P, Q) = \mu_n(Q, P)$.

DEFINITION 1. A sequence of symmetric measures $\{\mu_n\}$ is said to be *symmetrically compositive* if for some $\lambda \in \mathbf{R}$,

$$\begin{aligned} \mu_{nm}(P * R, Q * S) &+ \mu_{nm}(P * S, Q * R) \\ &= 2\mu_n(P, Q) + 2\mu_m(R, S) + \lambda \mu_n(P, Q)\, \mu_m(R, S) \end{aligned} \tag{1.11}$$

for all $P, Q \in \Gamma_n^0$, $S, R \in \Gamma_m^0$. If $\lambda = 0$, then $\{\mu_n\}$ is said to be *symmetrically additive.*

DEFINITION 2. A sequence of symmetric measures $\mu_n: \Gamma_n^0 \times \Gamma_n^0 \to \mathbf{R}$ $(n \geqslant 2)$ is said to have the *sum form* [16, 18] if there exists a symmetric function $f: I_0^2 \to \mathbf{R}$ such that

$$\mu_n(P, Q) = \sum_{k=1}^{n} f(p_k, q_k) \tag{1.12}$$

for all $P, Q \in \Gamma_n^0$ (symmetry of f means $f(p, q) = f(q, p)$).

The function f in (1.12) is called a *generating function* of the sequence $\{\mu_n\}$. For details regarding the existence of f, interested readers should refer to [16].

The sum form and symmetric compositivity lead to the study of the functional equation

$$\sum_{i=1}^{n}\sum_{j=1}^{m} f(p_i r_j, q_i s_j) + \sum_{i=1}^{n}\sum_{j=1}^{m} f(p_i s_j, q_i r_j)$$
$$= 2\sum_{i=1}^{n} f(p_i, q_i) + 2\sum_{j=1}^{m} f(r_j, s_j) + \lambda \sum_{i=1}^{n} f(p_i, q_i) \sum_{j=1}^{m} f(r_j, s_j), \qquad (1.13)$$

where P, $Q \in \Gamma_n^0$, R, $Q \in \Gamma_m^0$, and $\lambda \in \mathbf{R}$.

A function A from an interval to $\mathbf{R}$ is said to be *additive* if $A(x+y) = A(x) + A(y)$. An additive $A: I_0 \to \mathbf{R}$ has a unique extension to additive $\bar{A}: \mathbf{R} \to \mathbf{R}$ [7, Theorem 4.3]. A mapping $M: I_0 \to \mathbf{R}$ is *multiplicative* if $M(xy) = M(x)\,M(y)$. A multiplicative $M: I_0 \to \mathbf{R}$ has a unique extension to multiplicative $\bar{M}: \mathbf{R}_+ \to \mathbf{R}$[2]. A function $L: I_0 \to \mathbf{R}$ is said to be *logarithmic* provided $L(xy) = L(x) + L(y)$. A logarithmic $L: I_0 \to \mathbf{R}$ has a unique extension $\bar{L}: \mathbf{R}_+ \to \mathbf{R}$[2].

The aim of this paper is to find all the symmetrically compositive sequences $\mu_n: \Gamma_n^0 \times \Gamma_n^0 \to \mathbf{R}$ having the sum form with a measurable symmetric generating function $f: I_0^2 \to \mathbf{R}$. The measures we obtain include such measures as Hellinger coefficient [9], Chernoff coefficient [5], Jeffreys distance [10], J-divergence [15], J-divergence of degree α [21], and more. The concept of measuring similarity between two distributions is dual to that of distance and the measures obtained here cover both.

This paper is organized as follows: In Section 2, we present the Lebesgue measurable solution of the functional equation (1.13). In Section 3, we display all the regular symmetric measures which have the sum form and the compositivity. Section 4 contains discussion about a new measure there contained and its relationship to Hellinger coefficient and Jeffreys distance.

2. Solutions of the Functional Equation (1.13)

We make use of the following two auxiliary results to obtain the solutions of the functional equations (1.13).

LEMMA 1. *A function $f: I_0^2 \to \mathbf{R}$ satisfies the functional equation*

$$f(pr, qs) + f(ps, qr) = (r+s) f(p, q) + (p+q) f(r, s) \qquad (2.1)$$

for all p, q, r, $s \in I_0$ if, and only if,

$$f(p, q) = p[L_1(q) - L_2(p)] + q[L_1(p) - L_2(q)] \qquad (2.2)$$

where L_1, L_2 are logarithmic on $\mathbf{R}_+$.

Proof. It is easy to verify that f given by (2.2) satisfies (2.1). Obviously, $f=0$ is a solution of (2.1) and is of the form (2.2). We now suppose $f\neq 0$. First, we will show that f satisfying (2.1) for all $p, q, r, s\in I_0$ can be extended (uniquely) from I_0^2 to $\mathbf{R}_+^2$. Setting $r=s=\lambda$ in (2.1), we get

$$f(\lambda p, \lambda q)=\lambda f(p, q)+\lambda(p+q)\, l(\lambda), \tag{2.3}$$

where $l(\lambda):=(2\lambda)^{-1} f(\lambda, \lambda)$ is logarithmic (replace λ by $\lambda_1\lambda_2$ in (2.3) and use (2.3) three times). Now, we extend f to $\bar{f}$, as follows. For any $p, q\in\mathbf{R}_+$, choose positive λ sufficiently small such that λ, λp, $\lambda q\in I_0$. Define

$$\bar{f}(p, q)=\frac{1}{\lambda} f(\lambda p, \lambda q)-(p+q)\, l(\lambda). \tag{2.4}$$

From $f(\lambda\mu p, \lambda\mu q)=f(\lambda.\mu p, \lambda.\mu q)=f(\mu.\lambda p, \mu.\lambda q)$, using (2.3), it follows that the right side of (2.4) is independent of λ and thus $\bar{f}$ is well defined. For p, q, r, $s\in\mathbf{R}_+$, choose $\lambda\in I_0$ such that λp, λq, λr, $\lambda s\in I_0$. Then

$$\begin{aligned}
&\bar{f}(pr, qs)+\bar{f}(ps, qr)\\
&\quad=\frac{1}{\lambda^2}\{f(\lambda^2 pr, \lambda^2 qs)+f(\lambda^2 ps, \lambda^2 qr)\}-(pr+qs)\, l(\lambda^2)-(ps+qr)\, l(\lambda^2)\\
&\quad=\frac{1}{\lambda^2}\{f(\lambda p\,\lambda r, \lambda q\,\lambda s)+f(\lambda p\,\lambda s, \lambda q\,\lambda r)\}-(p+q)(r+s)\, l(\lambda^2)\\
&\quad=\frac{1}{\lambda}\{(r+s) f(\lambda p, \lambda q)+(p+q) f(\lambda r, \lambda s)\}\\
&\qquad-2(p+q)(r+s)\, l(\lambda) \qquad \text{(by using (2.1) and } l \text{ logarithmic)}\\
&\quad=(r+s)\bar{f}(p, q)+(p+q)\bar{f}(r, s).
\end{aligned}$$

Thus $\bar{f}$ satisfies (2.1) on $\mathbf{R}_+$. From here on, let us simply assume that f satisfies (2.1) for $p, q, r, s\in\mathbf{R}_+$. Set $p=q$, $r=s$ in (2.1) to get

$$f(pr, pr)=r f(p, p)+p f(r, r).$$

From this it follows that

$$L(p):=\frac{1}{p} f(p, p) \tag{2.5}$$

is logarithmic on $\mathbf{R}_+$. Setting $q=s=1$ in (2.1), we get

$$f(p, r)=(1+r)\, g(p)+(1+p)\, g(r)-g(pr), \tag{2.6}$$

where $g(p) := f(1, p)$. Note that (2.1) implies $f(1, 1) = 0$ and f is symmetric. Now (2.5) and (2.6) give

$$g(p^2) = 2(1+p)\, g(p) - p\, L(p), \qquad p \in \mathbf{R}_+. \tag{2.7}$$

With $p = r$, $q = s$ (2.1) yields

$$f(p^2, q^2) + f(pq, pq) = 2(p+q) f(p, q). \tag{2.8}$$

Putting (2.5), (2.6), and (2.7) into (2.8), we have

$$\begin{aligned}(1-p)(1-q)[2g(pq) - L(pq)] &= (1-p)(1-pq)[2g(q) - L(q)] \\ &\quad + (1-q)(1-pq)[2g(p) - L(p)].\end{aligned}$$

Defining

$$2L_2(x) = \begin{cases} (1-x)^{-1}(2g(x) - L(x)), & x \neq 1 \\ 0, & x = 1, \end{cases}$$

we get from this definition and $g(1) = 0$ that

$$g(p) = \frac{1}{2} L(p) + (1-p)\, L_2(p), \qquad p \in \mathbf{R}_+,$$

and from the above equation $L_2(pq) = L_2(p) + L_2(q)$ follows whenever $p \neq 1$, $q \neq 1$, and $pq \neq 1$. L_2 evidently satisfies $L_2(pq) = L_2(p) + L_2(q)$ when $p = 1$ or $q = 1$. To check that this equation is also true for the case $p \neq 1$, $q \neq 1$ but $pq = 1$ we have to show that $L_2(p^{-1}) = -L_2(p)$ for $p \neq 1$. The latter is equivalent to

$$\left(1 + \frac{1}{p}\right) f(p, p) + 2p\, f\left(1, \frac{1}{p}\right) = 2f(1, p),$$

which can be obtained by putting $q = p$, $r = 1$, and $s = 1/p$ in (2.1). Thus, L_2 is logarithmic on $\mathbf{R}_+$. Now using (2.6), we get (2.2), where $L_1 = \frac{1}{2}L + L_2$. This proves Lemma 1.

LEMMA 2. *Suppose $f: I_0^2 \to \mathbf{R}$ satisfies the functional equation*

$$f(pr, qs) + f(ps, qr) = f(p, q)\, f(r, s) \tag{2.9}$$

for all p, q, r, $s \in I_0$. Then

$$f(p, q) = M_1(p)\, M_2(q) + M_1(q)\, M_2(p), \tag{2.10}$$

where M_1, $M_2 : I_0 \to \mathbf{C}$ are multiplicative. Further, either M_1 and M_2 are both real or M_2 is the complex conjugate of M_1. The converse is also true.

Proof. Without loss of generality, let us assume that $f \neq 0$. Observe first that with $r = s = \lambda$ in (2.9) we get the M-homogeneity law

$$f(\lambda p, \lambda q) = \frac{1}{2} f(\lambda, \lambda) f(p, q) = M(\lambda) f(p, q), \tag{2.11}$$

where $M: I_0 \to \mathbf{R}$, $M(\lambda) := \frac{1}{2} f(\lambda, \lambda)$. Since $f \neq 0$, M is multiplicative and can be uniquely extended to a multiplicative $\bar{M}(\neq 0)$ on $\mathbf{R}_+$ [2]. As in Lemma 1, it is easy to show that

$$\bar{f}(p, q) = \bar{M}\left(\frac{1}{\lambda}\right) f(\lambda p, \lambda q)$$

for $p, q \in \mathbf{R}_+$ (λ small such that $\lambda p, \lambda q \in I_0$) extends f and Eq. (2.9) to $\mathbf{R}_+^2$. We assume from here on that f has this extended meaning.

We then fix $s = q = 1$ in (2.9) and define $g(p) = f(p, 1)$ to get

$$f(p, q) = g(p)\, g(q) - g(pq). \tag{2.12}$$

The assumption $f \neq 0$ implies that g is not multiplicative, and $f(1, 1) = 2$. Setting $s = 1$ in (2.9) and using (2.12), we obtain

$$g(pr)\, g(q) - 2g(pqr) = g(p)\, g(q)\, g(r) - g(p)\, g(qr) - g(r)\, g(pq). \tag{2.13}$$

We fix q in (2.13) to get

$$F(pr) = F(p) \frac{g(r)}{2} + F(r) \frac{g(p)}{2}, \tag{2.14}$$

where $F(p) := g(p)\, g(q) - 2g(pq)$.

If F is identically 0, then $M_1 := 2^{-1} g$ is multiplicative and $f(p, q) = 2M_1(pq)$, and f thus is of the form (2.10) with $M_2 = M_1$. If $F \neq 0$, then from (2.14) [6, Lemma 5, Remark 4] and since g is not multiplicative we have $g = M_1 + M_2$ where $M_1, M_2 : \mathbf{R}_+ \to \mathbf{C}$ are multiplicative; that is, f is of the form (2.10) where M_1, M_2 are complex valued. Suppose $M_k(p) = a_k(p) + i\, b_k(p)$ with real $a_k(p)$, $b_k(p) (k = 1, 2)$. Since f is real valued, letting $q = 1$ in (2.12), we get $b_1(p) = -b_2(p) = b(p)$ say. Again using f real, that is, the imaginary part of $f = 0$, we get

$$[a_2(p) - a_1(p)]\, b(q) + [a_2(q) - a_1(q)]\, b(p) = 0. \tag{2.15}$$

Suppose there is a p_0 such that $a_1(p_0) \neq a_2(p_0)$. Then from (2.15), we have $b(q) = c(a_2(q) - a_1(q))$, that is,

$$c(a_2(p) - a_1(p))(a_2(q) - a_1(q)) = 0,$$

for all p, q. Then $c=0$, so $b(q)=0$ and M_1 and M_2 are real. Otherwise $a_1=a_2$ and M_2 equals the conjugate of M_1 as desired. The converse part is straightforward. This completes the proof of Lemma 2.

Remark 1. The general solutions of functional Eqs. 2.1) and (2.9) are obtained in Lemmas 1 and 2 without assuming any regularity conditions on f.

Now we proceed to find the Lebesgue measurable and symmetric solution of the functional equation (1.13). We first treat the case $\lambda=0$.

THEOREM 1. *Let $f\colon I_0^2 \to \mathbf{R}$ be symmetric and measurable in each variable. Then f satisfies the functional equation* (1.13) *with $\lambda=0$ for all pairs of positive integers $m, n(\geqslant 2)$ and for all $P, Q \in \Gamma_n^0$, $R, S \in \Gamma_m^0$ if, and only if, f is given by*

$$f(p,q)=a(p\log_2 p+q\log_2 q)+b(p\log_2 q+q\log_2 p), \tag{2.16}$$

where a, b are arbitrary constants.

Proof. For fixed $P, Q \in \Gamma_n^0$, define

$$\begin{aligned}F(r,s)=\sum_{i=1}^{n} [&f(p_i r, q_i s)+f(p_i s, q_i r)\\ &-(r+s)f(p_i,q_i)-(p_i+q_i)f(r,s)].\end{aligned} \tag{2.17}$$

First, (2.17) in (1.13) with $\lambda=0$ gives $\sum_{j=1}^{m} F(r_j, s_j)=0$ and then the measurability of f implies (refer to [3]) $F(r,s)=a_1 r+a_2 s+a_3$, where a_1, a_2, a_3 are functions of P and Q with $a_1+a_2+ma_3=0$. Using the symmetry of f and noting the fact that (1.13) holds for all m (enough for two values of m) we get $a_1=a_2=a_3=0$, that is, the right hand side of (2.17) is zero. Keeping r and s constant, defining

$$G(p,q)=f(pr,qs)+f(ps,qr)-(r+s)f(p,q)-(p+q)f(r,s),$$

and applying once again the above procedure to G in place of F, we get (2.1). Since f is measurable in each variable, so are L_1 and L_2 derived in Lemma 1, giving $L_1(p)=a\log_2 p$ and $L_2(p)=b\log_2 p$, where a and b are real constants (see [1]). Thus from (2.2) follows (2.16). The converse part is easy to verify. This proves the theorem.

Now we consider the case $\lambda \neq 0$ and prove the following theorem.

THEOREM 2. *Let $f\colon I_0^2 \to \mathbf{R}$ be symmetric and measurable in each variable. Then f satisfies the functional equation* (1.13) *for all pairs of positive integers*

$m, n (\geqslant 2)$ and for all $P, Q \in \Gamma_n^0$, $R, S \in \Gamma_m^0$ with $\lambda \neq 0$ if, and only if, f has the form

$$f(p, q) = -\frac{1}{\lambda}(p+q), \tag{2.18}$$

or

$$f(p, q) = \frac{1}{\lambda}[p^\alpha q^\beta + p^\beta q^\alpha - p - q], \tag{2.19}$$

or

$$f(p, q) = \frac{1}{\lambda}\left[2p^\beta q^\beta \cos\left(\alpha \log_2 \frac{p}{q}\right) - p - q\right], \tag{2.20}$$

where α, β are arbitrary real constants.

Proof. First, define $h: I_0^2 \to \mathbf{R}$ by

$$h(p, q) = p + q + \lambda f(p, q), \tag{2.21}$$

and rewrite (1.13) as

$$\sum_{i=1}^{n} \sum_{j=1}^{m} [h(p_i r_j, q_i s_j) + h(p_i s_j, q_i r_j)] = \sum_{i=1}^{n} h(p_i, q_i) \sum_{j=1}^{m} h(r_j, s_j). \tag{2.22}$$

Note that h in (2.21) is symmetric and measurable in each variable. If $h = 0$, then f has the form (2.18). From here on, let $h \neq 0$. For fixed $P, Q \in \Gamma_n^0$, defining $F: I_0^2 \to \mathbf{R}$ by

$$F(r, s) = \sum_{i=1}^{n} [h(p_i r, q_i s) + h(p_i s, q_i r) - h(r, s)\, h(p_i, q_i)]$$

as in Theorem 1, we first obtain $F(r, s) = 0$. Applying once again the same procedure, we obtain the functional equation (2.9) satisfied by h. It can be shown that the measurability of h implies that of M_1 and M_2 in (2.10). Thus for real M_1, M_2, we get $M_1(p) = p^\alpha$, $M_2(q) = q^\beta$ [1], and from (2.10) we obtain (2.19). In the case where M_1, M_2 are complex, we get $M_1(p) = p^{\beta + i\alpha} = p^\beta e^{i\alpha \log_e p}$ and $M_2(q) = q^\beta e^{-i\alpha \log_e q}$ [13], resulting in (2.20). The converse is easy to verify. This completes the proof of the theorem.

3. Characterization of Measures

In this section we display all regular symmetric measures having the sum form and the symmetric compositivity.

THEOREM 3. *Suppose that a sequence of measures* $\{\mu_n\}$ *has the sum form* (1.12) *with a measurable symmetric generating function* $f: I_0^2 \to \mathbf{R}$, *and is symmetrically compositive for all pairs of positive integers* m, n ($\leqslant 2$). *Then*

$$\mu_n(P, Q) = \sum_{k=1}^{n} \left[p_k(a \log_2 p_k + b \log_2 q_k) + q_k(a \log_2 q_k + b \log_2 p_k)\right] \tag{3.1}$$

if $\lambda = 0$, *or else*

$$\mu_n(P, Q) = \frac{1}{\lambda}\left[\sum_{k=1}^{n} (p_k^\alpha q_k^\beta + p_k^\beta q_k^\alpha) - 2\right], \tag{3.2}$$

or

$$\mu_n(P, Q) = -\frac{2}{\lambda}\left[1 - \sum_{k=1}^{n} (p_k q_k)^\beta \cos\left(\alpha \log_2 \frac{p_k}{q_k}\right)\right], \tag{3.3}$$

or

$$\mu_n(P, Q) = -\frac{2}{\lambda}, \tag{3.4}$$

where a, b, α, *and* β *are arbitrary real constants.*

Proof. Since μ_n satisfies the sum form and (1.11), we obtain the functional equation (1.13) for all pairs of positive integers m, n. Thus, using Theorem 1 (when $\lambda = 0$) we obtain (3.1); using Theorem 2 (when $\lambda \neq 0$) we obtain (3.2), (3.3), or (3.4). This completes the proof of the theorem.

If the sequence $\{\mu_n\}$ is measuring "distance" between P and Q, it is pleasant to assume that

$$\mu_n(P, P) = 0.$$

COROLLARY 1. *A measure of distance* $\{\mu_n\}$ *satisfies the hypotheses in Theorem* 3 *and* $\mu_n(P, P) = 0$ *if, and only if,*

$$\mu_n(P, Q) = a\, J_n(P, Q), \tag{3.5}$$

or

$$\mu_n(P, Q) = \frac{(2^{\alpha - 1} - 1)}{\lambda} J_{n,\alpha}(P, Q), \tag{3.6}$$

or

$$\mu_n(P, Q) = -\frac{2}{\lambda}\left[1 - \sum_{k=1}^{n} (p_k q_k)^{1/2} \cos\left(\alpha \log_2 \frac{p_k}{q_k}\right)\right]. \tag{3.7}$$

Proof. Since $\mu_n(P, P) = 0$, (3.1) yields $a = -b$ and thus for $\lambda = 0$ we get (3.5). Now using $\mu_n(P, P) = 0$ in (3.2) we obtain $\beta = 1 - \alpha$ and hence (3.6). Again $\mu_n(P, P) = 0$ in (3.3) yields $\beta = \frac{1}{2}$. Thus, one obtains (3.7). Since (3.4) does not satisfy the condition $\mu_n(P, P) = 0$, μ_n corresponding to (3.4) does not mature. This completes the proof of the corollary.

4. Cosine Divergence

While μ_n in (3.5) and (3.6) are constant multiples of known measures such as J-divergence and J-divergence of degree α, the measures μ_n in (3.7) seem to be new. Their standardization

$$D_{n,\alpha}(P, Q) = \frac{1}{2}\left[1 - \sum_{k=1}^{n} (p_k q_k)^{1/2} \cos\left(\alpha \log_2 \frac{p_k}{q_k}\right)\right], \tag{4.1}$$

which we call the *cosine divergence*, has the following nice and desirable properties:

(a) Positive definite; $D_{n,\alpha}(P, Q) \geqslant 0$ and it is zero if, and only if, $P = Q$.

(b) Symmetric; $D_{n,\alpha}(P, Q) = D_{n,\alpha}(Q, P)$.

(c) Bounded from above; $D_{n,\alpha}(P, Q) \leqslant 1$.

(d) Subadditive; $D_{nm,\alpha}(P * R, Q * S) + D_{nm,\alpha}(P * S, Q * R) \leqslant 2D_{n,\alpha}(P, Q) + 2D_{m,\alpha}(R, S)$.

(e) $D_{n,\alpha}$ can be extended continuously to the square of the space of complete distributions $\Gamma_n = \{P = (p_1, p_2, ..., p_n) \mid p_k \geqslant 0, \sum_{k=1}^{n} p_k = 1\}$.

Properties (a) and (c) follow from Holder inequality and (d) follows from (1.11), (3.3), and (4.1). For fixed $Q = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, we draw the level curves of $D_{3,\alpha}(P, Q)$ in the variables p_1, p_2 (Fig. 1).

Notice that if $\alpha = 0$, $D_{n,\alpha}(P, Q)$ reduces to

$$D_{n,0}(P, Q) = \frac{1}{4} \sum_{k=1}^{n} (\sqrt{p_k} - \sqrt{q_k})^2,$$

which is a multiple of the formerly known Jeffreys distance.

Measuring similarity between P and Q can be conceived as a dual to that of measuring distance. If a distance measure μ_n between probability distribution is bounded from above by a constant b then $b - \mu_n$ is naturally a similarity measure. Applying this principle to $D_{n,\alpha}$, the similarity measure

$$S_{n,\alpha}(P, Q) = \frac{1}{2}\left[1 + \sum_{k=1}^{n} (p_k q_k)^{1/2} \cos\left(\alpha \log_2 \frac{p_k}{q_k}\right)\right] \tag{4.2}$$

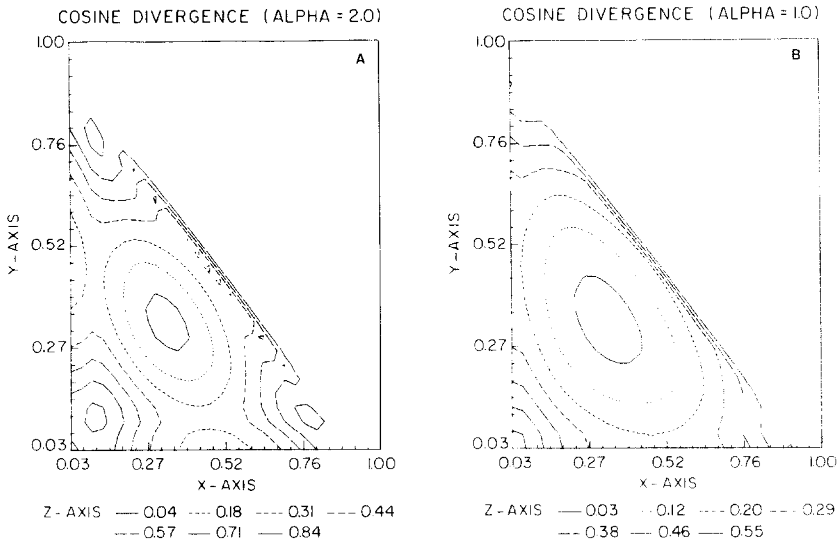

FIGURE 1

is obtained. This similarity measure has the following properties: (a) $S_{n,\alpha}(P, Q) \leqslant 1$ and equality holds if, and only if $P = Q$; (b) $S_{n,\alpha}(P, Q) = S_{n,\alpha}(Q, P)$; (c) $S_{n,\alpha}(P, Q) \geqslant 0$; (d) $S_{n,\alpha}$ is superadditive. This measure also satisfies most of the properties of a similarity measure listed in [23]. If $\alpha = 0$, then

$$S_{n,0}(P, Q) = \frac{1}{2}\left[1 + H_n(P, Q)\right],$$

where $H_n(P, Q)$ is the Hellinger (or Bhattacharyya) coefficient (refer to [4, 9]).

REFERENCES

1. J. ACZÉL, "Lectures on Functional Equations and Their Applications," Academic Press, New York/London, 1966.
2. J. ACZÉL, J. A. BAKER, D. Z. DJOKOVIĆ, PL. KANNAPPAN, AND F. RADÓ, Extensions of certain homomorphisms of subsemigroups to homomorphisms of groups, *Aequationes Math.* **6** (1971), 263–271.
3. J. ACZÉL, C. T. NG, AND C. WAGNER, Aggregation theorems for allocation problems, *SIAM J. Algebraic Discrete Methods* **5** (1984), 1–8.
4. A. BHATTACHARYYA, On a measure of divergence between two statistical populations defined by their probability distributions, *Calcutta Math. Soc. Bull.* **35** (1943), 99–109.
5. H. CHERNOFF, A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations, *Ann. Math. Stat.* **23** (1952), 493–507.

6. J. K. Chung, Pl. Kannappan, and C. T. Ng, A generalization of the cosine–sine functional equation on groups, *Linear Algebra Appl.* **66** (1985), 259–277.
7. J. Dhombres, "Some Aspects of Functional Equations," Chulalongkorn Univ. Press, Bangkok, 1979.
8. P. Halland, P. L. Brockett, and A. Levine, A characterization of divergence with application to questionnaire information, submitted for publication.
9. E. Hellinger, Die Orthogonalvarianten quadratischer Formen von unendlich vielen Variablelen, dissertation, Gottingen, 1907.
10. H. Jeffreys, An invariant form for the prior probability in estimation problems, *Proc. Roy. Soc London Ser. A* **186** (1946), 453–461.
11. T. Kailath, The divergence and Bhattacharyya distance measures in signal selection, *IEEE Trans. Comm. Tech.* **COM-15** (1967), 52–60.
12. Pl. Kannappan, On directed divergence and inaccuracy, *Z. Wahrsch. Verw. Gebiete* **25** (1972), 49–55.
13. Pl. Kannappan, "Theory of Functional Equations," Matscience Report 48, Madras, 1969.
14. Pl. Kannappan and P. N. Rathie, An axiomatic characterization of J-divergence, Submitted for publication.
15. S. Kullback, "Information Theory and Statistics," Peter Smith, Gloucester, MA. 1978.
16. C. T. Ng, Representation for measures of information with the branching property, *Inform. and Control* **25** (1974), 45–56.
17. P. N. Rathie and Pl. Kannappan, A directed divergence function of type β, *Inform. and Control* **20** (1972), 38–45.
18. P. Sahoo, On some functional equation connected to sum form information measures on open domains, *Utilitas Math.* **23** (1983), 161–175.
19. K. R. Sawyer, Testing separate families of hypotheses: An information criterion, *J. Roy. Statist. Soc. B* **45** (1983), 89–99.
20. K. R. Sawyer, A multiple divergence criterion for testing between separate hypotheses, *Statist. Probab. Lett.* **1** (1982), 26–30.
21. L. T. Sheng and P. N. Rathie, The J-divergence of order α, *J. Combin. Inform. System Sci.* **6** (1981), 197–205.
22. J. T. Tou and R. C. Gonzales, "Pattern Recognition Principle," Addison–Wesley, Reading, MA, 1974.
23. J. Veglius, S. Janson, and F. Johansson, Measures of similarity between distributions, *Quality and Quantity* **20** (1986), 437–441.
24. T. Y. Young and T. W. Calvert, "Classification, Estimation and Pattern Recognition," Amer. Elsevier, New York, 1974.

Printed in Belgium