PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 3, Fasc. 2 (1984), p. 241-258

DIFFERENTIAL METRICS IN PROBABILITY SPACES

BY

JACOB BURBEA AND C. RADHAKRISHNA RAO
(PITTSBURGH, PENNSYLVANIA)

Abstract. In this paper * we discuss the construction of differential metrics in probability spaces through entropy functionals and examine their relations with the information metric introduced by Rao using the Fisher information matrix in the statistical problem of classification and discrimination, and the classical Bergman metric. It is suggested that the scalar and Ricci curvatures associated with the Bergman information metric may yield results in statistical inference analogous to those of Efron using the Gaussian curvature.

1. Introduction. Distance measures between probability distributions play an important role in the discussion of problems of inference (see, e.g., [19], [20], p. 6-23, and [24], p. 317-332). A wider class of measures, which may not satisfy all the requirements of a distance function, called *dissimilarity* or *divergence measures*, are used extensively in problems of taxonomical classification in biology. Some discussion on the choice of these measures and their application to live data can be found, for instance, in [15], [18], [23], p. 19-34, and [27]. A unified approach to the construction of distance and dissimilarity measures is given in recent papers by Burbea and Rao [10] and Rao [25].

One method of specifying the difference between two probability distributions is through the geodesic distance induced by a suitably chosen quadratic differential metric in th space of probability distributions. This was done in earlier papers by Rao [21], [22] where the *Fisher information matrix* is used to

* The work of the second-named author is sponsored by the Air Force Office of Scientific Research under Contract F49620-79-C-0161. Reproduction in whole or in part is permitted for any purpose of the United States Government.

construct the differential metric for a parametric family of probability distributions, which will be referred to as the *information metric*. The choice of the information matrix arose in a natural way through the concept of statistical discrimination.

In this paper we discuss the construction of differential metrics through entropy functionals and examine their relations with the information metric. In particular, we study the connection between the classical *Bergman metric* and the information metric. We suggest the possibility of using the scalar and Ricci curvatures associated with the *Bergman information metric* in statistical inference.

2. Information differential metric. In this paper μ stands for a σ-finite measure on a σ-algebra of the subsets of a sample space $\mathscr{X}$. By

$$(f, g)_\mu = \int_{\mathscr{X}} f(t)\overline{g(t)}\, d\mu(t) \quad \text{and} \quad \|f\|_\mu = \sqrt{(f, f)_\mu}$$

we denote the inner product and norm, respectively, of the (separable) Hilbert space $L_2(\mathscr{X}; \mu)$ of complex-valued functions which are square-integrable on $\mathscr{X}$ with respect to the measure μ. We shall also write

$$S_\mu = \{f \in L_2(\mathscr{X}; \mu)\colon \|f\|_\mu = 1\}$$

and

$$P_\mu = \{|f| \in L_1(\mathscr{X}; \mu)\colon |f|^{1/2} \in S_\mu\}$$

for the unit sphere of $L_2(\mathscr{X}; \mu)$ and the set of probability densities, respectively.

Let D be a parameter space in $\mathbf{R}^k$ consisting of k-tuples $x = (x_1, \ldots, x_k)$ of real continuous parameters. Usually, D is a smooth manifold imbedded in $\mathbf{R}^k$; however, when $k = 2n$, it is convenient to consider D as a manifold imbedded in $\mathbf{C}^n \cong \mathbf{R}^{2n}$, consisting of n-tuples $z = (z_1, \ldots, z_n)$ of complex continuous parameters $z_j = x_j + iy_j$, $j = 1, \ldots, n$. In this case we shall use the formalism of complex differentiation

$$\partial_{z_j} = 2^{-1}(\partial_{x_j} - i\partial_{y_j}), \qquad \partial_{\bar{z}_j} = 2^{-1}(\partial_{x_j} + i\partial_{y_j}) \quad (j = 1, \ldots, n),$$

and thus, for a C^1-function f around $z \in D$,

$$df = (\partial + \bar{\partial}) f, \qquad df = df(z), \ z \in D,$$

with

$$\partial f = \sum_{j=1}^{n} \partial_{z_j} f\, dz_j, \qquad \bar{\partial} f = \sum_{j=1}^{n} \partial_{\bar{z}_j} f\, d\bar{z}_j.$$

In this notation, $\partial_{\bar{z}_j}$ can be written as $\bar{\partial}_{z_j}$, and if all the y_j are zero, i.e., if D is a manifold imbedded in $\mathbf{R}^n$, then the complex formalism reduces to the former real setting.

In this paper, in view of the above remarks, we shall assume that D is indeed a manifold imbedded in C^n and we shall freely use the complex formalism. A family $\mathscr{F} = \mathscr{F}(\mathscr{X}|D)$ of probability density functions $p = p(\cdot|\cdot)$ defined on $\mathscr{X} \times D$ with the property that, for any $z \in D$, $p(\cdot|z) \in P_\mu$ is specified by the following regularity conditions:

(i) for μ-almost all $t \in \mathscr{X}$, $p(t|\cdot)$ is differentiable in D;

(ii) $\partial_{z_j} \int_{\mathscr{X}} p(t|z)\, d\mu(t) = \int_{\mathscr{X}} \partial_{z_j} p(t|z)\, d\mu(t)$ and, therefore, the latter is equal to zero.

In certain instances of this paper it may be required to assume additional regularity properties on $p(\cdot|\cdot) \in \mathscr{F}(\mathscr{X}|D)$. Sometimes, however, these additional requirements will not be mentioned explicitly in order to avoid lengthy discussion.

The *Fisher information matrix* of $p = p(\cdot|\cdot) \in \mathscr{F}(\mathscr{X}|D)$ is the $(n \times n)$-hermitian matrix $F = [g_{k\bar{m}}]$ whose entries are

$$g_{k\bar{m}}(z) = \int_{\mathscr{X}} p^{-1}(\partial_{z_k} p)(\bar{\partial}_{z_m} p)\, d\mu(t).$$

This can also be written as

$$g_{k\bar{m}}(z) = \int_{\mathscr{X}} p(\partial_{z_k} \log p)(\bar{\partial}_{z_m} \log p)\, d\mu(t).$$

Since condition (ii) implies also that $\mathrm{E}\{\partial_{z_k} \log p(\cdot|z)\} = 0$, $k = 1, \ldots, n$, the information matrix at $z \in D$ is the (hermitian) variance-covariance matrix of $\{\partial_{z_k} \log p(\cdot|z)\colon k = 1, \ldots, n\}$.

The *information differential metric* with respect to $p(\cdot|\cdot) \in \mathscr{F}(\mathscr{X}|D)$ is the hermitian quadratic form

$$ds^2(z) = \sum_{k,m=1}^{n} g_{k\bar{m}}\, dz_k d\bar{z}_m, \quad g_{k\bar{m}} = g_{k\bar{m}}(z), \; z \in D.$$

In terms of the norm of $L_2(\mathscr{X}; \mu)$ this form admits the expression

$$ds^2(z) = \|p^{1/2} \partial \log p\|_\mu^2, \quad p = p(\cdot|z),$$

and therefore $ds^2(z)$ is positive definite. It follows that $ds^2(z)$ is (locally) invariant under holomorphic transformations of z. Indeed, if $\varphi\colon D^* \to D$ is a bijective holomorphic (or, in short, biholomorphic) mapping of D^* onto D, then $p^* = p^*(\cdot|\cdot)$ defined by

$$p^*(\cdot|w) = p(\cdot|\varphi(w)), \quad w \in D^*,$$

is in $\mathscr{F}(\mathscr{X}|D^*)$, and

$$g^*_{k\bar{m}}(w) = \int_{\mathscr{X}} p^*(\partial_{w_k} p^*)(\bar{\partial}_{w_m} p^*)\, d\mu(t)$$

are the entries of its information matrix while

$$ds_*^2(w) = \sum_{k,m=1}^{n} g_{k\bar{m}}^* dw_k d\bar{w}_m, \qquad g_{k\bar{m}}^* = g_{k\bar{m}}^*(w), \ w \in D^*,$$

is the corresponding information metric. Consequently, using the chain rule, we have

$$g_{k\bar{m}}^*(w) = \sum_{i,j=1}^{n} g_{i\bar{j}}(z) \frac{\partial z_i}{\partial w_k} \frac{\partial \bar{z}_j}{\partial \bar{w}_m},$$

which means that the entries of the information matrix are transformed by a biholomorphic transformation as components of a covariant tensor of the second order. Moreover, this also shows that $ds_*^2(w) = ds^2(z)$, as asserted.

The induced geodesic distance of ds^2, with respect to $p(\cdot|\cdot) \in \mathscr{F}(\mathscr{X}|D)$, defines a pseudo-distance S_D on $D \times D$, referred to as the *information pseudo-distance* with respect to $p = p(\cdot|\cdot)$ (see [2], [10], and [21]). By a *pseudo-distance* on $D \times D$ we mean any function δ on $D \times D$ that satisfies all axioms of a distance function except that $\delta(z, \zeta)$ can be zero even if $z \neq \zeta$. Under some additional regularity requirements on $p = p(\cdot|\cdot)$, which guarantee that ds^2 is strictly positive definite, the information pseudo-distance S_D becomes a proper distance on $D \times D$, known as the *information distance* with respect to $p = p(\cdot|\cdot)$. In the latter case, D becomes a metric space (D, S_D). This metric space is said to be *complete* if for each point $\zeta \in D$ and every $r > 0$ the closed ball $\{z \in D: \ S_D(z, \zeta) \leq r\}$ is a compact subset of D. Since S_D is induced from the Riemannian metric ds^2, completeness in the above sense is equivalent to completeness in the ordinary sense (see, e.g., [16], p. 53).

An alternative and simpler expression of $g_{k\bar{m}}(z)$ in the form of

$$g_{k\bar{m}}(z) = -\mathrm{E}\{\partial_{z_k}\bar{\partial}_{z_m} \log p(\cdot|z)\} \tag{2.1}$$

is also available, provided that $p(\cdot|\cdot) \in \mathscr{F}(\mathscr{X}|D)$ satisfies the additional regularity condition:

(iii) for μ-almost all $t \in \mathscr{X}$ and for every $z \in D$, $\partial_{z_k}\bar{\partial}_{z_m} \log p(t|z)$ exists.

If this condition is fulfilled, then $ds^2(z)$ takes the form

$$ds^2(z) = -\mathrm{E}\{\partial\bar{\partial} \log p(\cdot|z)\}.$$

An attractive feature of the information metric ds^2 is based on the following considerations: Let $z \in D$ and let f be μ-measurable on $\mathscr{X}$. Since $1 = (p(\cdot|z), 1)_\mu$, we deduce that $0 = (\partial p, 1)_\mu$, where $p = p(\cdot|z)$ and $p(\cdot|\cdot) \in \mathscr{F}(\mathscr{X}|D)$. Therefore

$$(\partial p, f)_\mu = (\partial p, f - \alpha \cdot 1)_\mu$$

for any scalar α and, by the Cauchy inequality, we have

$$|(\partial p, f)_\mu|^2 = |(\partial p, f - \alpha\cdot 1)_\mu|^2 = \left|\left(p^{1/2}\,\partial\log p,\ p^{1/2}(f-\alpha\cdot 1)\right)_\mu\right|^2$$
$$\leqslant \|p^{1/2}\,\partial\log p\|_\mu^2\,\|p^{1/2}(f-\alpha\cdot 1)\|_\mu^2$$
$$= ds^2(z)\,\|p^{1/2}(f-\alpha\cdot 1)\|_\mu^2.$$

Equality holds if and only if there exists a scalar $\beta = \beta(z)$ such that

$$f(t) = \alpha + \beta\,\partial\log p(t|\,z)$$

for μ-almost all $t\in\mathscr{X}$. Now, the minimum of $\|p^{1/2}(f-\alpha\cdot 1)\|_\mu^2$ is attained by $\alpha = \hat{f}(z)$, where $\hat{f}(z) = \left(f,\ p(\cdot\,|\,z)\right)_\mu$, and therefore

$$|(\partial p, f)_\mu|^2 \leqslant ds^2(z)\left\|p^{1/2}\left(f-\hat{f}(z)\cdot 1\right)\right\|_\mu^2.$$

This may also be written as

$$|\partial\hat{f}(z)|^2 \leqslant ds^2(z)\left\{\|p(\cdot\,|\,z)^{1/2} f\|_\mu^2 - |\hat{f}(z)|^2\right\}.$$

The expression in the last curly brackets is non-negative and it vanishes if and only if f is μ-almost everywhere a constant on $\mathscr{X}$.

The usefulness of the above lower-bound estimate is by now well known (see [24], p. 317-332), and therefore it will not be discussed here. For further details on the metric $ds^2(z)$, we refer the reader to [2], [10], and [21].

3. Entropies and divergence measures. For convenience we regard the family $\mathscr{F} = \mathscr{F}(\mathscr{X}|\,D)$ as a subset of an open set U, where U itself is an open subset of some Fréchet space $\bar{\mathscr{F}}$ of functions $f = f(\cdot|\cdot)$ defined on $\mathscr{X}\times D$, that includes the tangent space of U. The tangent of $f(t|\,z)$ at $z\in D$ in the direction of $(u, v)\in C^n\times C^n$ may be expressed as

$$d_{(u,v)} f(\cdot\,|\,z) = \partial_u f(\cdot\,|\,z) + \bar{\partial}_v f(\cdot\,|\,z),$$

where

$$\text{(3.1)}\qquad \partial_u f(\cdot\,|\,z) = \sum_{k=1}^{n} \partial_{z_k} f(\cdot\,|\,z)\,u_k,\qquad \bar{\partial}_v f(\cdot\,|\,z) = \sum_{k=1}^{n} \bar{\partial}_{z_k} f(\cdot\,|\,z)\,v_k.$$

The tangent, therefore, is composed of the holomorphic direction $\partial_u f(\cdot|\,z)$ and the anti-holomorphic direction $\bar{\partial}_v f(\cdot\,|\,z)$.

Let φ be a concave C^2-function on the interval $R_+ = (0, \infty)$ and consider the *φ-entropy functional*

$$\{H_\varphi(p)\}(z) = \int_{\mathscr{X}} \varphi\,[p(t|\,z)]\,d\mu(t),\qquad p = p(\cdot\,|\,\cdot)\in\mathscr{F}(\mathscr{X}|\,D).$$

We shall suppress the dependence on $z\in D$ and write

$$\text{(3.2)}\qquad H_\varphi(p) = \int_{\mathscr{X}} \varphi\,[p(t)]\,d\mu(t),\qquad p(t) = p(t|\,z)\in\mathscr{F}(\mathscr{X}|\,D).$$

The derivative of H_φ at $p \in U$ in the direction of $f \in \tilde{\mathscr{F}}$ is given by

$$dH_\varphi(p;f) = \frac{d}{ds} H_\varphi(p+sf)\Big|_{s=0},$$

and so

$$dH_\varphi(p;f) = \int_{\mathscr{X}} \varphi'[p(t)]\, f(t)\, d\mu(t).$$

The derivative of this at $p \in U$ in the direction of $g \in \tilde{\mathscr{F}}$ is

$$d^2 H_\varphi(p;f,g) = \int_{\mathscr{X}} \varphi''[p(t)]\, f(t)\, g(t)\, d\mu(t)$$

while the (complex) Hessian at $p \in U$ in the direction of $f \in \tilde{\mathscr{F}}$ is defined by

$$\Delta_f H_\varphi(p) = 4d^2 H_\varphi(p;f,\bar{f}),$$

and so

$$\Delta_f H_\varphi(p) = 4 \int_{\mathscr{X}} \varphi''[p(t)]\, |f(t)|^2\, d\mu(t).$$

Since φ is concave, we obtain

(3.3) $$-\Delta_f H_\varphi(p) \geqslant 0, \quad f \in \tilde{\mathscr{F}}.$$

In particular, when f is chosen to be $\partial_u p$, $u \in C^n$, we have

$$\Delta_{\partial_u p}\{H_\varphi(p)\}(z) = 4 \int_{\mathscr{X}} \varphi''[p(t|z)]\, |\partial_u p(t|z)|^2\, d\mu(t),$$

where $\partial_u p(\cdot|z)$ is as given in (3.1). When $u = (dz_1, \ldots, dz_n) \in C^n$, this may also be written as

(3.4) $$\Delta_{\partial p}\{H_\varphi(p)\}(z) = 4 \int_{\mathscr{X}} \varphi''[p(t|z)]\, |\partial p(t|z)|^2\, d\mu(t).$$

We specialize the above concave function φ to the α-*order entropy function* φ_α, $\alpha > 0$, defined by

(3.5) $$\varphi_\alpha(s) = \begin{cases} (\alpha-1)^{-1}(s-s^\alpha), & \alpha \neq 1, \\ -s \log s, & \alpha = 1. \end{cases}$$

This function is defined for $s \in R_+$ and can be extended to $s = 0$ by using the convention $0 \log 0 = 0$. With this choice of $\varphi = \varphi_\alpha$ we call $H_\alpha \equiv H_\varphi$ the α-*order entropy* (see [14]). It follows that

$$\{H_\alpha(p)\}(z) = \begin{cases} (\alpha-1)^{-1}\left[1 - \int_{\mathscr{X}} p^\alpha\, d\mu(t)\right], & \alpha \neq 1, \\ -\int_{\mathscr{X}} p \log p\, d\mu(t), & \alpha = 1, \ p = p(t|z) \in \mathscr{F}(\mathscr{X}|D). \end{cases}$$

From (3.3)-(3.5) we see that with H_α there is associated a (hermitian) differential metric

$$ds_\alpha^2(z) = -\frac{1}{4\alpha}\Delta_{\partial p}\{H_\alpha(p)\}(z)$$

which is positive definite and, therefore, a metric of a Riemannian geometry. This metric can be also expressed as

$$ds_\alpha^2(z) = \sum_{k,m=1}^{n} g_{km}^{(\alpha)}\, dz_k d\bar{z}_m,$$

where the metric coefficients are given by

$$g_{km}^{(\alpha)}(z) = \int_{\mathscr{X}} p^\alpha(\partial_{z_k}\log p)(\bar{\partial}_{z_m}\log p)\,d\mu(t).$$

The (hermitian) matrix $[g_{km}^{(\alpha)}]$ and the metric ds_α^2 will be called the α*-order entropy matrix* and the α*-order entropy metric,* respectively.

Of great importance as far as applications are concerned is the special case of $\alpha = 1$ and for this reason the index $\alpha = 1$ will be deleted from the above quantities. In this special case, $H = H_1$ is the familiar *Shannon's entropy,* and $ds^2 = ds_1^2$ and $[g_{k\bar{m}}] = [g_{k\bar{m}}^{(1)}]$ are the previously defined information differential metric and information matrix, respectively.

Another natural way for the derivation of the α-order entropy metric is via the notion of divergence. We consider any C^2-function $F(\cdot, \cdot)$ on $\boldsymbol{R}_+ \times \boldsymbol{R}_+$ so that $F(s, t) \geqslant 0$ and $F(s, s) = 0$ for $s, t \in \boldsymbol{R}$. For $p, q \in P_\mu$ we define the *divergence* of p and q with respect to F as

$$D_F(p, q) = \int_{\mathscr{X}} F[p(t), q(t)]\,d\mu(t), \qquad p, q \in P_\mu.$$

Fixing $p \in P_\mu$ and letting q vary we find that

$$D_F(p, p) = dD_F(p, q)|_{q=p} = 0, \qquad d^2 D_F(p, q)|_{q=p} \geqslant 0.$$

In particular, when $p = p(t|z)$ and $q = p(t|\zeta)$ with $z, \zeta \in D$ and $p(\cdot|\cdot) \in \mathscr{F}(\mathscr{X}|D)$, we have for the (complex) Hessian, in analogy with (3.4),

$$\Delta_{\partial p}\{D_F(p, p)\}(z) = 4\int_{\mathscr{X}} \partial_q^2 F(p, q)|_{q=p}\,|\partial p|^2\,d\mu(t).$$

This, of course, is also a (hermitian) positive definite differential quadratic form.

A very simple example is furnished by the J*-divergence* (see [11])

$$J_\varphi^{(\lambda)}(p, q) = H_\varphi(\lambda p + (1-\lambda)q) - \lambda H_\varphi(p) - (1-\lambda)H_\varphi(q), \qquad 0 < \lambda < 1,$$

induced by the φ-entropy functional H_φ. According to (3.2) we have

$$J_\varphi^{(\lambda)}(p, q) = \int_{\mathscr{X}} \{\varphi(\lambda p + (1-\lambda)q) - \lambda\varphi(p) - (1-\lambda)\varphi(q)\}\,d\mu(t),$$

and therefore

$$\Delta_{\partial p}\{J_{\varphi}^{(\lambda)}(p,p)\}(z) = -4\lambda(1-\lambda)\int_{\mathscr{X}} \varphi''(p)|\partial p|^2\,d\mu(t).$$

The concavity of φ implies that this is a positive definite form. In particular, when φ is the α-order entropy function φ_α of (3.5), the α-order entropy metric emerges again.

A more interesting example is furnished by the *K-divergence* (see [11]) which is defined as follows: Let ψ be a C^2-function in $\boldsymbol{R}_+$ with the property that $s\psi(s^{-1})+\psi(s)\geq 0$ and with the normalization $\psi''(1)>0$. The K-divergence is

$$K_\psi(p,q) = \frac{1}{8\psi''(1)}\int_{\mathscr{X}}\left\{p\psi\left(\frac{q}{p}\right)+q\psi\left(\frac{p}{q}\right)\right\}d\mu(t).$$

Its Hessian is therefore

$$\Delta_{\partial p}\{K_\psi(p,p)\}(z) = \int_{\mathscr{X}} p|\partial\log p|^2\,d\mu(t),$$

which is the information metric. If $\psi = -\varphi_\alpha$, where φ_α is the α-order entropy function of (3.5), the resulting K-divergence K_α is called the *α-order K-divergence* ($\alpha>0$). Note that in this case

$$-[s\varphi_\alpha(s^{-1})+\varphi_\alpha(s)] = \begin{cases}(\alpha-1)^{-1}(1-s^\alpha)(s^{1-\alpha}-1), & \alpha\neq 1,\\ (1-s)\log s^{-1}, & \alpha=1,\end{cases}$$

is indeed non-negative for $s\in\boldsymbol{R}_+$ and that $-\varphi_\alpha''(1)=\alpha>0$. Moreover,

$$8\alpha K_\alpha(p,q) = \begin{cases}(\alpha-1)^{-1}\left[\int_{\mathscr{X}}(p^{1-\alpha}q^\alpha+q^{1-\alpha}p^\alpha)\,d\mu(t)-2\right], & \alpha\neq 1,\\ \int_{\mathscr{X}}[q\log p^{-1}q+p\log q^{-1}p]\,d\mu(t), & \alpha=1.\end{cases}$$

In particular, $8K_1(p,q)$ is the familiar *Kullback-Leibler divergence* (see [17]).

Another example of a divergence measure is provided by the *Hellinger divergence*

$$H_\psi(p,q) = \int_{\mathscr{X}}[\psi(p)-\psi(q)]^2\,d\mu(t),$$

where ψ is any C^2-function on $\boldsymbol{R}_+$. Obviously, $\{H_\psi(p,q)\}^{1/2}$ defines a pseudo-distance on $P_\mu\times P_\mu$. In particular, if $p=p(t|z)$ and $q=p(t|\zeta)$ with $z,\zeta\in D$ and $p(\cdot|\cdot)\in\mathscr{F}(\mathscr{X}|D)$, we have the function

$$\varrho_\psi(z,\zeta) = \{H_\psi(p(\cdot|z),p(\cdot\zeta))\}^{1/2}.$$

which is the *Hellinger ψ-pseudo-distance* on $D\times D$,

$$\varrho_\psi(z,\zeta) = \left\{\int_{\mathscr{X}}[\psi(p(t|z))-\psi(p(t|\zeta))]^2\,d\mu(t)\right\}^{1/2}.$$

The Hessian, in analogy with (3.4), is given by

$$\Delta_{\partial p}\{H_\psi(p,p)\}(z) = 8 \int_{\mathscr{X}} [\psi'(p)]^2 |\partial p|^2 \, d\mu(t).$$

The choice

$$\psi(s) = \frac{1}{\alpha\sqrt{2}} s^{\alpha/2} \quad (s \in R_+, \alpha > 0\}$$

gives the *α-order Hellinger divergence*

$$H_\alpha(p,q) = \frac{1}{2\alpha^2} \int_{\mathscr{X}} [p^{\alpha/2} - q^{\alpha/2}]^2 \, d\mu(t)$$

and the *α-order Hellinger pseudo-distance*

$$\varrho_\alpha(z,\zeta) = \frac{1}{\alpha\sqrt{2}} \left\{ \int_{\mathscr{X}} ([p(t|z)]^{\alpha/2} - [p(t|\zeta)]^{\alpha/2})^2 \, d\mu(t) \right\}^{1/2}. \tag{3.6}$$

It follows that

$$\Delta_{\partial p}\{H_\alpha(p,p)\}(z) = 16 \int_{\mathscr{X}} p^\alpha |\partial \log p|^2 \, d\mu(t),$$

which is (modulo the factor of 16) the α-order entropy metric.

4. The projective pseudo-distance. The 1-order pseudo-distance ϱ_1 defined in (3.6) is known as the *Hellinger pseudo-distance* on D and is also denoted by ϱ. Evidently,

$$\varrho(z,\zeta) = \{1 - \int_{\mathscr{X}} [p(t|z)\,p(t|\zeta)]^{1/2} \, d\mu(t)\}^{1/2}, \qquad z,\zeta \in D,$$

with $p(\cdot|\cdot) \in \mathscr{F}(\mathscr{X}|D)$. It is also evident that

$$\varrho^2(z,z) = d\varrho^2(z,\zeta)|_{\zeta=z} = 0$$

and

$$d^2\varrho^2(z,\zeta)|_{\zeta=z} = \frac{1}{8} \int_{\mathscr{X}} p^{-1} |dp| \, d\mu(t).$$

In particular, $ds^2(z) = 2\Delta_{\partial p}\varrho^2(z,\zeta)|_{\zeta=z}$ is the information metric.

In spite of the importance of the Hellinger pseudo-distance ϱ in the theory of statistical inference (see [20], p. 6-23), it is found more convenient for our purpose to study an alternative pseudo-distance λ defined below. To do so we shall follow the convention of statistical quantum mechanics and consider $p(\cdot|\cdot) \in \mathscr{F}(\mathscr{X}|D)$ as a square of the modulus of a normalized *wave function* $\psi(\cdot|\cdot)$. Thus,

$$p(t|z) = |\psi(t|z)|^2, \qquad (t,z) \in \mathscr{X} \times D, \tag{4.1}$$

where $\psi(\cdot|z)$ is a complex-valued function in S_μ, the unit sphere of $L_2(\mathscr{X};\mu)$, i.e.,

(4.2) $$\|\psi(\cdot|z)\|_\mu^2 = 1, \qquad z \in D.$$

In this way we obtain

(4.3) $$\varrho(z,\zeta) = \{1 - \int_{\mathscr{X}} |\psi(t|z)\psi(t|\zeta)|\, d\mu(t)\}^{1/2}.$$

We define

$$\lambda(z,\zeta) = \frac{1}{\sqrt{2}} \min \|\exp\{i\theta_1\}\psi(\cdot|z) - \exp\{i\theta_2\}\psi(\cdot|\zeta)\|_\mu, \qquad z, \zeta \in D,$$

where the minimum is taken over all $\theta_1, \theta_2 \in [0, 2\pi]$. Since also

$$\lambda(z,\zeta) = \frac{1}{\sqrt{2}} \min_{0 \leqslant \theta \leqslant 2\pi} \|\psi(\cdot|z) - e^{i\theta}\psi(\cdot|\zeta)\|_\mu,$$

it is seen that λ is indeed a pseudo-distance on D. An alternative and useful expression for this pseudo-distance in the form of

(4.4) $$\lambda(z,\zeta) = \{1 - |\int_{\mathscr{X}} \psi(t|z)\psi(t|\zeta)\, d\mu(t)|\}^{1/2}$$

is also available. Comparing this expression with that of (4.3) we see that

$$\varrho(z,\zeta) \leqslant \lambda(z,\zeta), \qquad z, \zeta \in D.$$

A routine calculation based on

$$\lambda^2(z,\zeta) = 1 - |(\psi(\cdot|z), \psi(\cdot|\zeta))_\mu|$$

and $\lambda^2(z,z) = d\lambda^2(z,\zeta)|_{\zeta=z} = 0$ shows that

(4.5) $$d^2\lambda^2(z,\zeta) = \|d\psi\|_\mu^2 - |(\psi, d\psi)_\mu|^2, \qquad \psi = \psi(\cdot|z),$$

which is clearly non-negative.

We call λ the (μ,ψ)-*projective pseudo-distance* of D. The reasons for this name lie in the following more general considerations. Let H be an abstract Hilbert space and consider any two non-zero elements h_1 and h_2 of H. We say that h_1 is *equivalent* to h_2, in short $h_1 \sim h_2$, if there exists a non-zero scalar c such that $h_1 = ch_2$. The set of all equivalence classes $[h]$, $h \in H - \{0\}$, forms the projective space $P(H)$ which, in general, is of infinite dimension. This is a complete metric space with respect to the distance.

$$d([h_1],[h_2]) = \frac{1}{\sqrt{2}} \operatorname{dist}([h_1] \cap S, [h_2] \cap S),$$

where S is the unit sphere of H. It follows that

$$d^2([h_1],[h_2]) = \tfrac{1}{2} \inf_{\theta_1,\theta_2} \left\| \exp\{i\theta_1\} \frac{h_1}{\|h_1\|} - \exp\{i\theta_2\} \frac{h_2}{\|h_2\|} \right\|^2$$

$$= \min_{\theta_1,\theta_2} \left\{ 1 - \mathrm{Re}\left[\exp\{i(\theta_1-\theta_2)\} \frac{(h_1, h_2)}{\|h_1\|\,\|h_2\|} \right] \right\},$$

and thus

$$d([h_1],[h_2]) = \left\{ 1 - \frac{|(h_1, h_2)|}{\|h_1\|\,\|h_2\|} \right\}^{1/2}. \tag{4.6}$$

It is also clear that any linear isometry $T\colon H \to \tilde{H}$ between two Hilbert spaces induces an isometry $P_T\colon P(H) \to P(\tilde{H})$ between the corresponding projective spaces, given by $P_T([h]) = [Th]$.

In our case, the Hilbert space H is $L_2(\mathscr{X};\mu)$ and the unit sphere of H is S_μ. In view of (4.2) for any $z \in D$, the wave function $\psi(\cdot|z)$ belongs to S_μ. We define the mapping $\mathscr{T}\colon D \to P(L_2(\mathscr{X};\mu))$ by $\mathscr{T}(z) = [\psi(\cdot|z)]$, $z \in D$, and note that this mapping is injective if and only if for any two distinct points $z, \zeta \in D$ the wave functions $\psi(\cdot|z)$ and $\psi(\cdot|\zeta)$ are linearly independent. We also define

$$\lambda(z,\zeta) = d(\mathscr{T}(z), \mathscr{T}(\zeta)) = d([\psi(\cdot|z)], [\psi(\cdot|\zeta)]),$$

which in view of (4.6) is identical with (4.4).

The difference between the Hellinger and the projective pseudo-distance is now more apparent. Both are based on the normalized wave function $\psi(\cdot|\cdot)$ of S_μ, but while the Hellinger pseudo-distance admits the expression

$$\varrho^2(z,\zeta) = 1 - |(|\psi(\cdot|z)|, |\psi(\cdot|\zeta)|)_\mu|,$$

the projective pseudo-distance admits the more analytic expression of (4.4). Consequently, $0 \leqslant \varrho(z,\zeta) \leqslant \lambda(z,\zeta) \leqslant 1$ for every $z, \zeta \in D$. Moreover, while ϱ is a distance of D if and only if for any two distinct points $z, \zeta \in D$ the wave-function amplitudes $|\psi(\cdot|z)|$ and $|\psi(\cdot|\zeta)|$ are different on $\mathscr{X}$ except for a subset of $\mathscr{X}$ of zero μ-measure, λ is a distance on D if and only if for any two distinct points $z, \zeta \in D$ the wave functions $\psi(\cdot|z)$ and $\psi(\cdot|\zeta)$ are linearly independent.

5. Sesqui holomorphic kernels. In most cases a given wave function $g(\cdot|z) \in L_2(\mathscr{X};\mu)$, $z \in D$, is not necessarily normalized as in (4.2), in which case the quantities under considerations will involve certain norming constants detailed below. We write

$$g(\cdot|z) = \sqrt{K(z,\bar{z})}\,\psi(\cdot|z), \qquad \psi(\cdot|z) \in S_\mu,\ z \in D, \tag{5.1}$$

where the *norming constant*

$$K(z,\bar{z})=(g(\cdot|z),g(\cdot|z))_\mu \tag{5.2}$$

is a positive smooth function of $z\in D$. The reason for displaying $K(z,\bar{z})$ instead of, say, $K(z)$ is that the right-hand side of (5.2) is in effect a function of z as well as $\bar{z}$ $(z\in D)$, a fact which will become more apparent once we further require that $g(t|\cdot)$ be holomorphic in D for μ-almost all $t\in\mathscr{X}$. With the notation of (5.1) and (5.2), formula (4.5) becomes

$$d^2\lambda^2(z,\zeta)|_{\zeta=z}=K^{-2}[K\|dg\|_\mu^2-|(g,dg)_\mu|^2], \tag{5.3}$$

where $K=K(z,\bar{z})$ and $g=g(\cdot|z)$, $z\in D$.

A particularly important case occurs when the non-normalized wave function $g(\cdot|\cdot)$ has the property that $g(t|\cdot)$ is holomorphic in D for μ-almost all $t\in\mathscr{X}$. In this case, in view of (5.2) and Hartogs' theorem, $K(z,\bar{z})$ is holomorphic in $(z,\bar{z})$, $z\in D$, and therefore, by polarization,

$$K(z,\bar{\zeta})=(g(\cdot|z),g(\cdot|\zeta))_\mu,\quad z,\zeta\in D. \tag{5.4}$$

It follows that, for any fixed $\zeta\in D$, $K(\cdot,\bar{\zeta})$ is holomorphic in D, and because the kernel is hermitian, i.e., $K(z,\bar{\zeta})=\overline{K(\zeta,\bar{z})}$ for any $z,\zeta\in D$, it also follows that $K(\zeta,\cdot)$ is anti-holomorphic in D. Another application of Hartogs' theorem then shows that $K(z,\bar{\zeta})$ is holomorphic in $(z,\bar{\zeta})$ for $(z,\zeta)\in D\times D$. Such kernels are said to be *sesqui holomorphic* on $D\times D$. Any sesqui holomorphic kernel $K(z,\bar{\zeta})$ which is also positive definite on $D\times D$ is called a *Bergman kernel* on $D\times D$ (see, e.g., [12], p. 88-93). This means that, for any finite system of points $z_1,\ldots,z_N$ of D and any corresponding scalars $\alpha_1,\ldots,\alpha_N$ of C,

$$\sum_{k,m=1}^{N}K(z_k,\bar{z}_m)\alpha_k\bar{\alpha}_m\geqslant 0.$$

Now, it follows immediately from (5.4) that $K(z,\bar{\zeta})$ is indeed a Bergman kernel on $D\times D$. In fact,

$$\sum_{k,m=1}^{N}K(z_k,\bar{z}_m)\alpha_k\bar{\alpha}_m=\Big\|\sum_{k=1}^{N}\alpha_k g(\cdot|z_k)\Big\|_\mu^2\geqslant 0.$$

Applying the classical theory of reproducing kernels (see [1] and [12], p. 88-93) we deduce the existence of a unique Hilbert space $H(D)$ of functions which are holomorphic in D such that $k_\zeta(z)=K(z,\bar{\zeta})$, $z,\zeta\in D$, is its reproducing kernel. Let $(\cdot,\cdot)$ be the inner product of $H(D)$; then for any $\zeta\in D$ we have $k_\zeta\in H(D)$ and

$$f(\zeta)=(f,k_\zeta),\quad f\in H(D).$$

In particular,

$$K(z,\bar{\zeta})=k_\zeta(z)=(k_\zeta,k_z),\quad z,\zeta\in D.$$

In the other direction, assume that we are given a reproducing kernel space $H(D)$, with the inner product $(\cdot, \cdot)$, of holomorphic functions in D. Let $k_\zeta = K(\cdot, \bar{\zeta})$, $\zeta \in D$, be the reproducing kernel of $H(D)$. This kernel, as before, is a Bergman kernel on $D \times D$. We now consider any isometry of $H(D)$ onto $L_2(\mathscr{X}; \mu)$ and put

$$g(\cdot \mid z) = \overline{Tk_z}, \quad z \in D. \tag{5.5}$$

Therefore, it follows that, for $z, \zeta \in D$,

$$K(z, \bar{\zeta}) = (k_\zeta, k_z) = (Tk_\zeta, Tk_z)_\mu = \left(\overline{g(\cdot \mid \zeta)}, \overline{g(\cdot \mid z)}\right)_\mu$$

or, expressed in another form,

$$K(z, \bar{\zeta}) = \left(g(\cdot \mid z), g(\cdot \mid \zeta)\right)_\mu = \int_{\mathscr{X}} g(t \mid z)\overline{g(t \mid \zeta)}\, d\mu(t),$$

which is precisely that of (5.4). We see therefore that there is a canonical correspondence between holomorphic wave-functions $g(\cdot \mid z) \in L_2(\mathscr{X}; \mu)$ and the Bergman kernels k_z, $z \in D$, of Hilbert spaces $H(D)$ of holomorphic functions in D. This correspondence is determined by (5.5) and for this reason $g(\cdot \mid \cdot)$, which is defined on $\mathscr{X} \times D$, is called a *generating function* of $L_2(\mathscr{X}; \mu)$ for $H(D)$ (see [5]).

Under the above circumstances, the second differential in (5.3) takes the form

$$d^2\lambda^2(z, \zeta)|_{\zeta = z} = K^{-2}[K\partial\bar{\partial}K - |\partial K|^2]$$

or

$$d^2\lambda^2(z, \zeta)|_{\zeta = z} = \partial\bar{\partial}\log K, \quad K = K(z, \bar{z}),\ z \in D.$$

This is the *Bergman metric*

$$db^2(z) = \partial\bar{\partial}\log K = \sum_{k,m=1}^{n} T_{k\bar{m}}\, dz_k d\bar{z}_m, \tag{5.6}$$

where

$$T_{k\bar{m}}(z) = \partial_{z_k}\bar{\partial}_{z_m} \log K(z, \bar{z}), \quad z \in D,\ k, m = 1, .., n. \tag{5.7}$$

A special instance of this metric was first introduced and studied by Bergman in 1933 (see also [4], p. 182-186, and [6] and [8], for additional details). An alternative geometrical derivation of this metric is as follows: We define

$$S(D) = \{f \in H(D) \colon \|f\| \leqslant 1\} \quad \text{and} \quad S_\zeta(D) = \{f \in S(D) \colon f(\zeta) = 0\}$$

for a fixed $\zeta \in D$. For a direction $v \in C^n$, we define

$$b(\zeta; v) = \sqrt{K(\zeta, \bar{\zeta})}\, \max\{|\partial_v f(\zeta)| \colon f \in S_\zeta(D)\}, \tag{5.8}$$

where

$$\partial_v f(z) = \sum_{k=1}^{n} v_k \partial_{z_k} f(z), \qquad v = (v_1, \ldots, v_n) \in C^n.$$

Standard Hilbert space arguments show that the extremal problem (5.8) admits a solution which is unique modulo a rotation and that

(5.9) $$4b^2(\zeta; v) = 4\partial_v \bar{\partial}_v \log K = \Delta_v \log K, \qquad K = K(\zeta, \bar{\zeta}).$$

In particular, when $v = dz = (dz_1, \ldots, dz_n)$, we have

$$b^2(z; dz) = db^2(z)$$

and we once again obtain the Bergman metric. This metric is also Kähler, i.e., the two-form

$$\sum_{k,m=1}^{n} T_{k\bar{m}} dz_k \wedge d\bar{z}_m$$

is closed. For more details on such metrics we refer the reader to the monograph of Weil [28].

The (μ, ψ)-projective pseudo-distance λ in (4.4) assumes here a simpler expression. Indeed, using (4.4), (5.1), (5.2), and (5.4), we readily obtain

(5.10) $$\lambda(z, \zeta) = \{1 - [|K(z, \bar{\zeta})|^2 K(z, \bar{z})^{-1} K(\zeta, \bar{\zeta})^{-1}]^{1/2}\}^{1/2}, \qquad z, \zeta \in D.$$

In this form and when $K(z, \bar{\zeta})$ is the classical Bergman kernel (i.e., the reproducing kernel of $H(D) \equiv H_2(D)$, where $H_2(D)$ stands for the space of holomorphic functions in the bounded domain D in C^n which are also square-integrable with respect to the Lebesgue measure on D), this pseudo-distance becomes a proper distance which was first studied by Skwarczyński [26]. For this reason, λ in (5.10) will be also called the *Skwarczyński pseudo-distance* of D (see [9] for further details). A sufficient condition for λ to be a distance is provided by the following theorem:

THEOREM 1. *Suppose that* $1, f_1, \ldots, f_n \in H(D)$, *where* $f_j(z) = z_j$, $1 \leqslant j \leqslant n$, *and* $z = (z_1, \ldots, z_n) \in D$. *Then* λ *is a distance on* D.

Proof. Assume that $\lambda(z, \zeta) = 0$ for $z, \zeta \in D$. Therefore, by (5.10), $|K(z, \bar{\zeta})|^2 = K(z, \bar{z}) K(\zeta, \bar{\zeta})$, which means that $k_\zeta = \alpha k_z$ for some $\alpha \in C$. It follows that $f(\zeta) = \bar{\alpha} f(z)$ for every $f \in H(D)$. Putting $f = 1$ and $f = f_j$, $j = 1, \ldots, n$, we obtain $\zeta = z$, which completes the proof.

In the present setting, a relation of considerable importance, as far as applications of statistical inference are concerned, is the fact that the information metric ds^2 is nothing else but the Bergman metric db^2. Indeed, from (4.1), (4.2), and (5.1) we have

$$\log p(t|z) = \log g(t|z) + \log \overline{g(t|z)} - \log K(z, \bar{z}).$$

By virtue of the Cauchy-Riemann equations we deduce that

$$\partial_{z_k}\bar{\partial}_{z_m}\log p(t|z) = -\partial_{z_k}\bar{\partial}_{z_m}\log K(z, \bar{z}),$$

which is valid for μ-almost all $t \in \mathscr{X}$ and all $z \in D$. This, by (2.1) and (5.7), shows that $g_{k\bar{m}}(z) = T_{k\bar{m}}(z)$, $z \in D$, and therefore $ds^2(z) = db^2(z)$.

Summarizing the above results we have the following theorem:

THEOREM 2. *Assume that the probability density function* $p(\cdot|\cdot) \in \mathscr{F}(\mathscr{X}|D)$ *is given by*

$$p(t|z) = |g(t|z)|^2/\sqrt{K(z, \bar{z})}, \quad K(z, \bar{z}) > 0, \; z \in D,$$

where $g(t|\cdot)$ *is holomorphic in* D *for* μ*-almost all* $t \in \mathscr{X}$. *Then*

(i) $K(z, \bar{\zeta})$ *is the Bergman kernel of a Hilbert space* $H(D)$ *of holomorphic functions in* D;

(ii) $g(\cdot|\cdot)$, *which is defined on* $\mathscr{X} \times D$, *is the generating function of* $L_2(\mathscr{X}; \mu)$ *for* $H(D)$;

(iii) *the projective pseudo-distance* λ *in* (4.4) *is the Skwarczyński pseudo-distance in* (5.10);

(iv) *for any* $z, \zeta \in D$, $0 \leqslant \varrho(z, \zeta) \leqslant \lambda(z, \zeta) \leqslant 1$, *where* ϱ *is the Hellinger pseudo-distance in* (4.3);

(v) *the information metric* ds^2 *is the Bergman metric* db^2 *which may be expressed by* (5.6)-(5.9);

(vi) $\lambda^2(z, z) = d\lambda^2(z, \zeta)|_{\zeta=z} = 0$ *and* $d^2\lambda^2(z, \zeta)|_{\zeta=z} = ds^2(z) = db^2(z) \geqslant 0$.

Associated with the Bergman information metric ds^2 are the standard invariants as the scalar and the Ricci curvatures which perhaps deserve a further study in the present context. Such a study will probably yield results analogous to those of Efron (see [13] and the literature cited therein). For example, an analogue of the Gaussian curvature of ds^2 may be given (see [7]). For a direction $v \in C^n$ and $z \in D$, the *directional curvature* of $ds^2(z)$ is, in view of (5.9), given by

$$\varkappa(z: v) = -\frac{2}{\Delta_v \log K}\Delta_v \log(\Delta_v \log K), \quad K = K(z, \bar{z}).$$

For $n = 1$ this curvature reduces to the usual Gaussian curvature of the one-dimensional Kähler metric $ds^2(z) = (\partial_z\bar{\partial}_z \log K)\,dz d\bar{z}$.

The classical Bergman metric and the original Skwarczyński distance (see [26]) are also globally invariant under biholomorphic mappings of D. This property, in general, does not hold for our ds^2 and λ. It holds, however, when some additional assumptions are imposed. For this and related results we refer to [6].

In the next section we provide a simple example illustrating the theory. More examples can be generated along silimar lines (see also [5]).

6. An example. Let φ be a holomorphic mapping of the manifold D (imbedded in C^n) into the right half-plane $R = \{z \in C\colon \operatorname{Re} z > 0\}$, thus $\operatorname{Re} \varphi(z) > 0$ for every $z \in D$. Let $\alpha > 0$ be a real number and consider the generating function

$$g_\alpha(t|z) = t^{(\alpha-1)/2} e^{-t\varphi(z)}, \quad t \in R_+,\ z \in D.$$

This generates the Bergman kernel

$$K_\alpha(z, \bar\zeta) = \int_0^\infty g_\alpha(t|z)\overline{g_\alpha(t|\zeta)}\,dt, \quad z, \zeta \in D.$$

It follows that

$$K_\alpha(z, \bar\zeta) = \int_0^\infty e^{-t[\varphi(z)+\overline{\varphi(\zeta)}]} t^{\alpha-1}\,dt$$

or

$$K_\alpha(z, \bar\zeta) = \Gamma(\alpha)[\varphi(z)+\overline{\varphi(\zeta)}]^{-\alpha}, \quad \alpha > 0.$$

In the present situation the sample space $\mathscr{X}$ is R_+ and the measure μ is the ordinary Lebesgue measure. The probability density function $p(\cdot|\cdot) \in \mathscr{F}(\mathscr{X}|D)$ is therefore

$$p(t|z) = \frac{2^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-2t\operatorname{Re}\varphi(z)} [\operatorname{Re} \varphi(z)]^\alpha, \quad \operatorname{Re} \varphi(z) > 0,\ z \in D.$$

The Fisher information matrix has the entries

$$g_{km}^{(\alpha)}(z) = \partial_{z_k} \bar\partial_{z_m} \log K_\alpha(z, \bar z), \quad k, m = 1, \ldots, n,$$

and therefore

$$g_{km}^{(\alpha)}(z) = \alpha[2\operatorname{Re} \varphi(z)]^{-2} (\partial_{z_k}\varphi)\overline{(\partial_{z_m}\varphi)}, \quad \varphi = \varphi(z).$$

It follows that the rank of this matrix is 1. The Bergman information metric is then

$$ds_\alpha^2(z) = \alpha[2\operatorname{Re} \varphi(z)]^{-2} |d\varphi(z)|^2. \tag{6.1}$$

Here, of course, $ds_\alpha^2(z) = db_\alpha^2(z)$, and thus, in view of (5.9), for a direction $v \in C^n$ we have

$$b_\alpha^2(z\colon v) = \alpha[2\operatorname{Re} \varphi(z)]^{-2} |d\varphi(z\colon v)|^2.$$

It follows that

$$\Delta_v \log b_\alpha^2(z\colon v) = 8[2\operatorname{Re} \varphi(z)]^{-2} |d\varphi(z\colon v)|^2,$$

and hence the directional curvature is $\varkappa_\alpha(z, v) = -4\alpha^{-1}$, which is a negative constant and independent of the direction v.

We recognize the metric in (6.1) as the Poincaré hyperbolic metric of the right half-plane R. The geodesic pseudo-distance S_α induced by ds_α^2 is also the information pseudo-distance and is given by

$$S_\alpha(z, \zeta) = \frac{1}{2\sqrt{\alpha}} \log \frac{1+\delta}{1-\delta}, \quad z, \zeta \in D,$$

where

$$\delta \equiv \delta(z, \zeta) = \left| \frac{\varphi(z) - \varphi(\zeta)}{\varphi(z) + \overline{\varphi(\zeta)}} \right|. \tag{6.2}$$

The Skwarczyński pseudo-distance is now

$$\lambda_\alpha(z, \zeta) = \{1-(1-\delta^2)^{\alpha/2}\}^{1/2},$$

where δ is defined by (6.2). To write this in another form we put $w = \varphi(z)$ and $\tau = \varphi(\zeta)$, $z, \zeta \in D$, and thus $u = \operatorname{Re} w > 0$ and $v = \operatorname{Re} \tau > 0$. Then (see also [9])

$$\lambda_\alpha(z, \zeta) = \sqrt{1 - [\sqrt{uv} | (w+\bar{\tau})|/2]^\alpha}.$$

On the other hand, it is easily seen that the Hellinger pseudo-distance admits the expression

$$\varrho_\alpha(z, \zeta) = \sqrt{1 - [\sqrt{uv}\,(u+v)/2]^\alpha},$$

and thus, once again, $0 \leqslant \varrho_\alpha(z, \zeta) \leqslant \lambda_\alpha(z, \zeta) \leqslant 1$ for every $z, \zeta \in D$.

References

[1] N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc. 68 (1950), p. 337-404.

[2] C. Atkinson and A. F. S. Mitchell, *Rao's distance measure*, Sankhyā 43 (1981), p. 345-365.

[3] S. Bergman, *Über eine in der Theorie der Funktionen von zwei komplexen veränderlichen auftretende unitare Geometrie*, Akad. Wetensch. Amsterdam Proc. 36 (1933), p. 307-313.

[4] – *The kernel function and conformal mapping*, Math. Survey 5, Amer. Math. Soc., Providence, R. I., 1970.

[5] J. Burbea, *Total positivity of certain reproducing kernels*, Pacific J. Math. 67 (1976), p. 101-130.

[6] – *The Carathéodory metric and its majorant metrics*, Canad. J. Math. 29 (1977), p. 771-780.

[7] – *On the hessian of the Carathéodory metric*, Rocky Mountain J. Math. 8 (1978), p. 555-559.

[8] – *A generalization of Pick's theorem and its applications to intrinsic metrics*, Ann. Polon. Math. 39 (1981), p. 49-61.

[9] — *On metrics and distortion theorems, recent developments in several complex variables*, Ann. Math. Studies 100 (1981), p. 65-92.
[10] — and C. R. Rao, *Entropy differential metric, distance and divergence measures in probability spaces — a unified approach*, T. Multivariate Anal. 12 (1982), p. 575-596.
[11] — *On the convexity of some divergence measures based on entropy functions*, IEEE Trans. Information Theory 28 (1982), p. 489-495.
[12] W. F. Donoghue, Jr., *Monotone matrix functions and analytic continuation*, Springer-Verlag, Berlin 1974.
[13] B. Efron, *Defining the curvature of a statistical problem (with applications to second order efficiency)*, Ann. Statist. 3 (1975), p. 1189-1242.
[14] M. E. Havrda and F. Charvát, *Quantification method of classification processes: concept of structural α-entropy*, Kybernetica 3 (1967), p. 30-35.
[15] N. Jardine and R. Sibson, *Mathematical taxonomy*, J. Wiley, New York 1971.
[16] S. Kobayashi, *Hyperbolic manifolds and holomorphic mappings*, M. Dekker, New York 1970.
[17] S. Kullback and R. A. Leibler, *On information and sufficiency*, Ann. Math. Statist. 22 (1951), p. 79-86.
[18] P. C. Mahalanobis, D. N. Majumdar and C. R. Rao, *Anthropometric survey of the United Provinces — a statistical study*, Sankhyā 9 (1949), p. 90-324.
[19] K. Matusita, *Decision rules based on distance for problem of fit, two samples and estimation*, Ann. Math. Statist. 26 (1955), p. 631-640.
[20] E. J. G. Pitman, *Some basic theory for statistical inference*, Halsted Press, London 1979.
[21] C. R. Rao, *Information and accuracy attainable in the estimation of statistical parameters*, Bull. Calcutta Math. Soc. 37 (1945), p. 81-91.
[22] — *On the distance between two populations*, Sankhyā 9 (1949), p. 246-248.
[23] — *Taxonomy in anthropology, in mathematics, in the archaeological and historical sciences*, Edin. Univ. Press, Edinburgh 1971.
[24] — *Linear statistical inference and its applications*, J. Wiley, New York 1973.
[25] — *Diversity and dissimilarity coefficients: a unified approach*, J. Theoretical Population Biology. 21 (1982), p. 24-43.
[26] M. Skwarczyński, *The invariant distance in the theory of pseudo-conformal transformations and the Lu Qi Keng conjecture*, Proc. Amer. Math. Soc. 22 (1969), p. 305-310.
[27] R. R. Sokal and P. H. A. Sneath, *Principles of numerical taxonomy*, W. H. Freeman, San Francisco 1963.
[28] A. Weil, *Introduction à l'étude de variétés kählériennes*, Hermann, Paris 1958.

Department of Mathematics and Statistics
University of Pittsburgh
Pittsburgh, PA 15260, U.S.A.

Received on 10. 9. 1981