

AD-A093 875

PITTSBURGH UNIV PA INST FOR STATISTICS AND APPLICATIONS F/G 12/1
ENTROPY DIFFERENTIAL METRIC, DISTANCE AND DIVERGENCE MEASURES I--ETC(U)
OCT 80 J BURBEA, C R RAO F49620-79-C-0161

UNCLASSIFIED

TR-80-18

AFOSR-TR-80-1374

NL

40
1 of 1
20-11876

END
DATE
FILMED
2-84
DTIC

AFOSR-Tk- 80 - 1374

ENTROPY DIFFERENTIAL METRIC, DISTANCE AND
DIVERGENCE MEASURES IN PROBABILITY SPACES -
A UNIFIED APPROACH

J. Burbea
and
C. Radhakrishna Rao*

LEVEL

October 1980

Technical Report No. 80-18

✓ Institute for Statistics and Applications
Department of Mathematics and Statistics
University of Pittsburgh
Pittsburgh, PA 15260

DTIC
SELECTED
JAN 16 1981

*The work of this author is sponsored by the Air Force ✓
of Scientific Research under Contract [REDACTED] F49620-77-C-0161
Reproduction in whole or in part is permitted for any
purpose of the United States Government.

Approved for public release;
distribution unlimited.

81 1 16 023

AD A093875

DDC FILE COPY

19 REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM	
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER	
18 AFOSR-TR-80-1374	AD-A093875		
4. TITLE (and Subtitle)		5. TYPE OF REPORT & PERIOD COVERED	
ENTROPY DIFFERENTIAL METRIC, DISTANCE AND DIVERGENCE MEASURES IN PROBABILITY SPACES-A UNIFIED APPROACH.		(9) INTERIM rept.	
7. AUTHOR(s)		6. PERFORMING ORG. REPORT NUMBER	
10 J. Burbea & C. Radhakrishna/Rao			
9. PERFORMING ORGANIZATION NAME AND ADDRESS		8. CONTRACT OR GRANT NUMBER(s)	
Institute for Statistics & Applications Department of Mathematics and Statistics University of Pittsburgh, Pittsburgh, PA 15260		(15) F49624-79-C-0161	
11. CONTROLLING OFFICE NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS	
AFOSR/NM Bolling AFB, DC 20332		(16) 61102F 2304/A5 (17) A5	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE	
(12) T38		(11) Oct 1980	
		13. NUMBER OF PAGES	
		34	
		15. SECURITY CLASS. (of this report)	
		UNCLASSIFIED	
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE	
16. DISTRIBUTION STATEMENT (of this Report)			
(14) TR-80-1374 Approved for public release; distribution unlimited.			
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)			
18. SUPPLEMENTARY NOTES			
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)			
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)			
The paper is devoted to metrization of probability spaces through the introduction of a quadratic differential metric in the parameter space of the probability distributions. For this purpose, a ϕ -entropy functional is defined on the probability space and its Hessian along a direction of the tangent space of the parameter space is taken as the metric. The distance between two probability distributions is computed as the geodesic distance induced by the metric. The paper also deals with three measures of divergence between probability distributions and their inter-relationships.			

ENTROPY DIFFERENTIAL METRIC, DISTANCE
AND DIVERGENCE MEASURES IN PROBABILITY
SPACES - A UNIFIED APPROACH

J. Burbea and C. Radhakrishna Rao
University of Pittsburgh, Pittsburgh

Abstract. The paper is devoted to metrization of probability spaces through the introduction of a quadratic differential metric in the parameter space of the probability distributions. For this purpose, a ϕ -entropy functional is defined on the probability space and its Hessian along a direction of the tangent space of the parameter space is taken as the metric. The distance between two probability distributions is computed as the geodesic distance induced by the metric. The paper also deals with three measures of divergence between probability distributions and their inter-relationships.

AMS (MOS) Subject Classification: 62H30, 94A17, 53A35.

Key Words and Phrases: Divergence measures, Entropy, Geodesic distance, Information metric.

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is
approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

1. INTRODUCTION

In an early paper, one of the authors [13] introduced a quadratic differential metric over the parameter space of a parametric family of probability distributions and proposed the geodesic distance induced by the metric as a measure of dissimilarity between two probability distributions. This metric was derived from heuristic considerations and it was expressed in terms of the Fisher information matrix (Fisher [6], see Rao [16, pp. 329-332] for details). Such a choice of the matrix for the quadratic differential metric was shown to have attractive properties through the concepts of discrimination and divergence measures between probability distributions ([9], [14, 15] and [16, p.332]). Quite recently, Atkinson and Mitchell [1] obtained the geodesic distances induced by the metric introduced in [13], which will be referred to in this paper as the information metric, for a number of parametric family of probability distributions.

In this paper, we consider a general function space and study a metric based on the Hessian of the ϕ -entropy functional, which was also introduced in an earlier paper by the authors [5]. A special choice of ϕ leads to the α -order entropy of Havrda and Charvát [7], and this gives rise to a class of metrics, which are called α -order entropy metrics. The above mentioned information metric is a

Accession For	NTIS GRA&I	DTIC TAB	Unannounced	Justification	By	Distribution	Availability	Dist
								A

limiting member of this class as $\alpha \rightarrow 1$, which corresponds to the Shannon entropy [18].

The geodesic distances induced by the α -order entropy metric are obtained for the multinomial and normal distributions. Their relation to other distance measures due to Möbius, Poincaré, Hellinger and Carathéodory is examined. The relationship of the information metric to the Bergman metric will be discussed elsewhere.

We also extend the concepts of the J,K,L-divergence measures between multinomial populations considered in the earlier paper [5] to more general distributions, and study their inter-relationships and convexity properties.

Dissimilarity measures between probability distributions play an important role in the discussion of problems of statistical inference and in practical applications to study affinities among a given set of populations. (See for instance, Matusita [10, 11], Pitman [12, pp. 6-23], Rao [16, p. 352], [17]). This paper provides a unified approach for measuring dissimilarity between probability distributions through distance and divergence measures having some desirable properties.

2. ϕ -ORDER ENTROPY METRIC

Throughout this paper, F denotes a linear space of

functions, $p = p(x)$, $x \in X$, measurable with respect to a σ -finite measure μ on a σ -algebra of the subsets of X . The convex subset of probability density functions in F is denoted by F_1

$$F_1 = \{p \in F: \int_X p(x) d\mu(x) = 1, p(x) \geq 0 \text{ for } \mu\text{-almost all } x \in X\}. \quad (2.1)$$

Let U be an open convex subset of F and let ϕ be a C^2 -function on an interval I containing

$$\nu\{p(x) \in \mathbb{R}: p \in U, x \in X\}.$$

For $p \in U$, we define the ϕ -entropy functional

$$H_\phi(p) = \int_X \phi[p(x)] d\mu(x). \quad (2.2)$$

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

$$dH_\phi(p:f) = \left. \frac{d}{dt} H_\phi(p+tf) \right|_{t=0}, \quad t \in \mathbb{R},$$

and thus, by virtue of (2.2),

$$dH_\phi(p:f) = \int_X \phi'[p(x)] f(x) d\mu(x).$$

The second derivative at $p \in U$ along $g \in F$ is

$$d^2 H_\phi(p:f,g) = \int_X \phi''[p(x)] f(x) g(x) d\mu(x) \quad (2.3)$$

and, in particular, the Hessian is

$$\begin{aligned}\Delta_f H_\phi(p) &= d^2 H_\phi(p; f, f) \\ &= \int_X \phi''[p(x)] \{f(x)\}^2 d\mu(x). \end{aligned} \quad (2.4)$$

We note that $\Delta_f H_\phi(p) \leq 0$ for every $f \in F$ and $p \in U$, if and only if ϕ is concave on I . This is equivalent to the requirement that H_ϕ be a concave functional on U .

We shall also consider a parametric family of probability density functions $p=p(x|\theta)$ with $x \in X$ and $\theta=(\theta_1, \dots, \theta_n) \in \Omega$, a manifold in $\mathbb{R}^n$. We assume that the subfamily of F_1 in (2.1)

$$F_\Omega = \{p(\cdot|\theta) \in F_1 : \theta \in \Omega\} \quad (2.5)$$

is sufficiently smooth in $\theta \in \Omega$ and satisfies the usual regularity properties, not explicitly stated to avoid lengthy discussion. Accordingly, we shall write

$$dp = dp(\theta) = \sum_{j=1}^n [\partial \theta_j p(\cdot|\theta)] d\theta_j; \quad \theta \in \Omega, p(\cdot|\theta) \in F_\Omega. \quad (2.6)$$

Then the Hessian in (2.4) along a direction of the tangent space of the parameter space Ω is obtained by replacing f by dp in (2.6). Thus

$$\begin{aligned}\Delta_\theta H_\phi(p) &= d^2 \{H_\phi(p)\}(\theta) \\ &= \int_X \phi''(p) [dp]^2 d\mu(x); \quad p=p(x|\theta). \end{aligned} \quad (2.7)$$

In particular, when ϕ is concave in $\mathbb{R}_+ \equiv (0, \infty)$

$$ds_\phi^2(\theta) = -\Delta_\theta H_\phi(p) \quad (2.8)$$

is a positive definite form on the tangent space, which may be regarded as a differential metric of a Riemannian geometry. This can also be written as

$$ds_{\phi}^2(\theta) = \sum_{k,m=1}^n g_{km}^{(\phi)} d\theta_k d\theta_m \quad (2.9)$$

where

$$g_{km}^{(\phi)} = g_{km}^{(\phi)}(\theta) = - \int_{\mathcal{X}} \phi''(p) (\partial_{\theta_k} p) (\partial_{\theta_m} p) d(x), \quad p = p(x|\theta) \in F_{\Omega}. \quad (2.10)$$

The metric in (2.9) and the matrix $[g_{km}]$ in (2.10) will be called the ϕ -entropy metric and the ϕ -entropy matrix respectively. The distance between probability density functions in F_{Ω} is defined as the geodesic distance between their parameter values determined by the metric (2.9).

We shall now consider some special choices of ϕ . For this purpose, we define for $\alpha \in \mathbb{R}$, two families $\{\phi_{\alpha}\}$ and $\{\psi_{\alpha}\}$ of smooth functions on $\mathbb{R}_+$:

$$\phi_{\alpha}(x) = \begin{cases} (\alpha-1)^{-1} (x-x^{\alpha}) & , \alpha \neq 1 \\ -x \log x & , \alpha = 1 \end{cases} \quad (2.11)$$

and

$$\psi_{\alpha}(x) = \begin{cases} [\alpha(\alpha-1)]^{-1} (1-\alpha+\alpha x-x^{\alpha}), & \alpha \neq 0, 1 \\ 1-x+\log x & , \alpha = 0 \\ -x \log x + x - 1 & , \alpha = 1 \end{cases} \quad (2.12)$$

When the smooth function ϕ in (2.2) is chosen to be ϕ_{α} of (2.11), we shall write $H_{\alpha} = H_{\phi_{\alpha}}$ and $H = H_1$. In this way, for $p \in F_1$

$$H_{\alpha}(p) = \begin{cases} (\alpha-1)^{-1} [1 - \int_{\mathcal{X}} p^{\alpha} d\mu(x)], & \alpha \neq 1 \\ - \int_{\mathcal{X}} p \log p d\mu(x), & \alpha = 1 \end{cases} \quad (2.13)$$

is the α -order entropy [7], while the 1-order entropy $H = H_1$, is the Shannon entropy [18]. The metric (2.9) with $\phi = \phi_{\alpha}$, $\alpha \neq 0$ is denoted by $ds_{\alpha}^2(\theta)$. In order that the value $\alpha = 0$ be also included, we modify ϕ_{α} to ψ_{α} as in (2.12). In this way

$$ds_{\alpha}^2(\theta) = \sum_{k,m=1}^n g_{km}^{(\alpha)} d\theta_k d\theta_m, \quad \alpha \in \mathbb{R} \quad (2.14)$$

with

$$g_{km}^{(\alpha)} = g_{km}^{(\alpha)}(\theta) = \int_{\mathcal{X}} p^{\alpha} (\partial_{\theta_k} \log p) (\partial_{\theta_m} \log p) d\mu(x), \quad p = p(x|\theta) \in F_{\Omega}. \quad (2.15)$$

We call (2.14), the α -order entropy metric and the matrix $[g_{km}^{(\alpha)}]$ in (2.15), the α -order entropy matrix. The geodesic pseudo distance induced by $ds_{\alpha}^2(\theta)$ is denoted by S_{α} and is called the α -order entropy pseudo distance.

In the special case of $\alpha=1$, corresponding to the Shannon entropy which is widely used in applied research, we have (dropping the suffix $\alpha=1$)

$$ds^2(\theta) = \sum_{k,m=1}^n g_{km} d\theta_k d\theta_m \quad (2.16)$$

and $[g_{km}]$ with

$$g_{km} = g_{km}(\theta) = \int_{\mathcal{X}} p (\partial_{\theta_k} \log p) (\partial_{\theta_m} \log p) d\mu(x), \quad p = p(x|\theta) \in F_{\Omega}. \quad (2.17)$$

The expression (2.16) is the information metric [13] mentioned in the introduction while $[g_{km}]$ is the Fisher information matrix. The geodesic pseudo-distance S induced by $ds^2(\theta)$ will be called the information pseudo-distance (a pseudo-distance satisfies all the postulates of distance except that it may vanish for elements which are distinct).

3. THE J, K, L-DIVERGENCE MEASURES

3.1 Definitions and inter-relationships

We consider the convex subset U of F , the function ϕ on the interval I and the ϕ -entropy functional H_ϕ as defined in (2.2). For $p, q \in U$, the J-divergence (with respect to H_ϕ) is defined to be the Jensen difference

$$J_\phi(p, q) = 2 H_\phi\left(\frac{p+q}{2}\right) - H_\phi(p) - H_\phi(q)$$

which can be written in the explicit form

$$J_\phi(p, q) = \int_X \{2\phi\left(\frac{p+q}{2}\right) - \phi(p) - \phi(q)\} d\mu(x), \quad p=p(x), q=q(x) \in U. \quad (3.1)$$

We also consider other measures of divergence, special forms of which have received numerous practical applications: The K-divergence

$$K_\phi(p, q) = \int_X (p-q)[p^{-1}\phi(p) - q^{-1}\phi(q)] d\mu(x) \quad (3.2)$$

and the L-divergence

$$L_{\phi}(p, q) = \int_X \{p \phi(\frac{q}{p}) + q \phi(\frac{p}{q})\} d\mu(x). \quad (3.3)$$

The following theorem gives some results concerning the J, K, L-divergences and their inter-relationships.

Theorem 1. The following hold:

- (i) If ϕ is concave on $\mathbb{R}_+$, then $J_{\phi}(p, q) \geq 0$ for $p, q \in F_1$.
- (ii) If $F(x) \equiv x \phi(x^{-1}) + \phi(x)$ is non-positive on $\mathbb{R}_+$, then $L_{\phi}(p, q) \geq 0$ for $p, q \in F_1$.
- (iii) If $\psi(x) = \phi(x)/x$ is decreasing on $\mathbb{R}_+$, then $K_{\phi}(p, q) \geq 0$ for $p, q \in F_1$.
- (iv) If ψ is decreasing and convex on $\mathbb{R}_+$, then $K_{\phi}(p, q) \geq J_{\phi}(p, q)$ for $p, q \in F_1$.
- (v) If ϕ is concave and ψ is convex on $\mathbb{R}_+$, then $K_{\phi}(p, q) \geq J_{\phi}(p, q) \geq 0$ for $p, q \in F_1$.

Proof. Items (i) and (iii) are trivial. As for item (ii), we have

$$L_{\phi}(p, q) = - \int_X q F(\frac{p}{q}) d\mu(x); \quad p, q \in F_1,$$

and item (ii) follows. We now prove item (iv). From (3.1) and (3.2), we have

$$K_{\phi}(p, q) - J_{\phi}(p, q) = \int_X G(p, q) d\mu(x)$$

where

$$G(x, y) \equiv \frac{y}{x} \phi(x) + \frac{x}{y} \phi(y) - 2 \phi[(x+y)/2]; \quad x, y \in \mathbb{R}_+.$$

This may be written as

$$\frac{G(x, y)}{x+y} = \frac{y}{x+y} \psi(x) + \frac{x}{x+y} \psi(y) - \psi(\frac{x+y}{2})$$

$$\geq \psi\left(\frac{yx}{x+y} + \frac{xy}{x+y}\right) - \psi\left(\frac{x+y}{2}\right)$$

$$\geq \psi\left(\frac{2xy}{x+y}\right) - \psi\left(\frac{x+y}{2}\right) \geq 0.$$

The first inequality follows from the convexity of ψ and the second inequality results from the assumption that ψ is decreasing on $\mathbb{R}_+$. This proves (iv). Item (v) follows from items (i) and (iv), if we show that the assumptions of (v) imply that ψ is decreasing on $\mathbb{R}_+$. Indeed, from $\psi(x) = \phi(x)/x$ we have

$$\psi'(x) = -\frac{1}{x}[\psi(x) - \phi'(x)]$$

and thus

$$\psi''(x) = -\frac{1}{x}[2\psi'(x) - \phi''(x)].$$

Therefore,

$$2\psi'(x) = -x\psi''(x) + \phi''(x) < 0, \quad x \in \mathbb{R}_+.$$

This concludes the proof.

When the function ϕ is replaced by ϕ_α of (2.11) the resulting divergences J_{ϕ_α} , K_{ϕ_α} and L_{ϕ_α} will be called the " α -order J, K and L divergences" and they will be denoted by J_α , K_α and L_α respectively. As in the case of the α -order entropy H_α , the index $\alpha=1$ will be dropped from these divergences and, thus, $J=J_1$, $K=K_1$ and $L=L_1$. For $p, q \in F_1$, the explicit expressions of J_α , K_α and L_α are as follows:

$$J_{\alpha}(p,q) = \begin{cases} (\alpha-1)^{-1} \int_X [p^{\alpha} + q^{\alpha} - 2^{1-\alpha} (p+q)^{\alpha}] d\mu(x) & , \alpha \neq 1 \\ \int_X [p \log p + q \log q - (p+q) \log \{(p+q)/2\}] d\mu(x) & , \alpha = 1, \end{cases} \quad (3.4)$$

$$K_{\alpha}(p,q) = \begin{cases} (\alpha-1)^{-1} \int_X (p-q)(p^{\alpha-1} - q^{\alpha-1}) d\mu(x) & , \alpha \neq 1 \\ \int_X (p-q)(\log p - \log q) d\mu(x) & , \alpha = 1, \end{cases} \quad (3.5)$$

$$L_{\alpha}(p,q) = \begin{cases} (\alpha-1)^{-1} \left[\int_X (p^{1-\alpha} q^{\alpha} + q^{1-\alpha} p^{\alpha}) d\mu(x) - 2 \right] & , \alpha \neq 1 \\ \int_X [q \log(p^{-1} q) + p \log(q^{-1} p)] d\mu(x) & , \alpha = 1. \end{cases} \quad (3.6)$$

We note that for $\alpha = 1$,

$$K(p,q) = L(p,q) = \int_X (p-q)(\log p - \log q) d\mu(x) ; p, q \in F_1,$$

which is the familiar Jeffreys-Kullback-Leibler divergence.

In this connection, we also mention the α -order Hellinger psuedo distance

$$M(p,q) = 2|\alpha|^{-1} \left[\int_X (p^{\alpha/2} - q^{\alpha/2})^2 d\mu(x) \right]^{1/2} \quad (3.7)$$

The special case of (3.7) when $\alpha = 1$, $M(p,q) = M_1(p,q)$, has been extensively studied by Matusita [10,11] and recently discussed by Pitman [12, pp. 6-23] from the point of view of statistical inference.

The following corollary is a consequence of Theorem 1:

Corollary 1. Let $\alpha \geq 0$. Then, for $p, q \in F_1$:

(i) $J_{\alpha}(p,q) \geq 0$;

(ii) $K_{\alpha}(p,q) \geq 0$;

- (iii) $L_\alpha(p, q) \geq 0$;
- (iv) $K_\alpha(p, q) \geq J_\alpha(p, q) \geq 0$, provided $0 \leq \alpha \leq 2$;
- (v) $K(p, q) = L(p, q) \geq J(p, q)$.

Proof. We consider ϕ_α of (2.11) and define

$$F_\alpha(x) \equiv x\phi_\alpha(x^{-1}) + \phi_\alpha(x) , \quad \psi_\alpha(x) = x^{-1}\phi_\alpha(x) ; \quad x \in \mathbb{R}_+ .$$

Since the cases with $\alpha = 1$ are limiting cases of $\alpha \neq 1$ as $\alpha \rightarrow 1$, we may assume that $\alpha \neq 1$, $\alpha > 0$. In this case

$$\phi_\alpha(x) = (\alpha-1)^{-1}(x-x^\alpha), \quad \psi_\alpha(x) = (\alpha-1)^{-1}(1-x^{\alpha-1}),$$

$$F_\alpha(x) = (\alpha-1)^{-1}x^{1-\alpha}(1-x^\alpha)(x^{\alpha-1}-1).$$

Since

$$\phi_\alpha''(x) = -\alpha x^{\alpha-2} \leq 0, \quad F_\alpha(x) \leq 0, \quad \psi_\alpha'(x) = -x^{\alpha-2} < 0,$$

items (i)-(iii) follow from items (i)-(iii) of Theorem 1.

Also, since $\psi_\alpha''(x) = (2-\alpha)x^{\alpha-3}$, item (iv) follows from item (v) of Theorem 1. Finally, (v) follows from (iv) and (3.5)-(3.6).

It is worth pointing out that the divergence measures (3.1)-(3.3) based on the general ϕ and (3.4)-(3.7) based on the α -order entropy can be used to generate a metric in the parameter space defining the probability distributions by considering two contiguous distributions. This is easily done by considering the Hessian along the tangent space of F_Ω , namely when $p = p(\cdot | \theta)$ and $q \rightarrow p$. The

precise results are as follows:

$$\begin{aligned} \text{(i)} \quad d^2\{J_\phi(p,p)\}(\theta) &= -\frac{1}{2} \int_X \phi''(p) [dp(\theta)]^2 d\mu(x) \\ &= \frac{1}{2} ds_\phi^2(\theta) \end{aligned}$$

which is the ϕ -entropy metric defined in (2.9).

$$\text{(ii)} \quad d^2\{K_\phi(p,p)\}(\theta) = -2 \int_X [p^{-1}\phi(p)]' [dp(\theta)]^2 d\mu(x),$$

$$\begin{aligned} \text{(iii)} \quad d^2\{L_\phi(p,p)\}(\theta) &= -2 \phi''(1) \int_X p^{-1} [dp(\theta)]^2 d\mu(x) \\ &= -2 \phi''(1) ds^2(\theta) \end{aligned}$$

where $ds^2(\theta)$ is the information metric as in (2.16), so that when $\phi''(1) < 0$, this metric is essentially the information metric.

Further when $\phi = \phi_\alpha$ as in (2.11), we have

$$\text{(iv)} \quad d^2\{J_\alpha(p,p)\}(\theta) = \frac{\alpha}{4} ds_\alpha^2(\theta),$$

$$\text{(v)} \quad d^2\{K_\alpha(p,p)\}(\theta) = 2 ds_\alpha^2(\theta),$$

$$\text{(vi)} \quad d^2\{L_\alpha(p,p)\}(\theta) = \alpha ds_\alpha^2(\theta),$$

$$\text{(vii)} \quad d^2\{M_\alpha^2(p,p)\}(\theta) = \frac{\alpha^2}{4} ds_\alpha^2(\theta),$$

where $ds_\alpha^2(\theta)$ is the α -order entropy metric. The relations

(i)-(vii) reflect the local properties of the J , K , L , M -divergence measures. We shall now consider their global properties in terms of their convexity as functions on

$$F_1 \times F_1.$$

3.2 The J-divergence

We compute the Hessian of J_ϕ at $(p,q) \in U \times U$ along $(f,g) \in F \times F$;

$$\Delta_{(f,g)} J_\phi(p,q) = d^2 J_\phi[(p,q):(f,g),(f,g)].$$

By virtue of (2.4), we deduce that

$$\Delta_{(f,g)} J_\phi(p,q) = - \int_X \{a(p,q)f^2 + b(p,q)fg + a(q,p)g^2\} d\mu(x) \quad (3.8)$$

where

$$b(p,q) \equiv - \frac{1}{2} \phi''[\frac{1}{2}(p+q)] \quad (3.9)$$

and

$$a(p,q) \equiv \phi''(p) + b(p,q). \quad (3.10)$$

We therefore conclude that J_ϕ is convex (concave) on $U \times U$ if and only if $a(p,q) \leq 0$ ($a(p,q) \geq 0$) and

$$d(p,q) \equiv a(p,q)a(q,p) - [b(p,q)]^2 \geq 0. \quad (3.11)$$

From (3.9)-(3.11), we find that

$$a(p,q) = -2\phi''(p)\phi''[\frac{1}{2}(p+q)] \left\{ \frac{1}{\phi''(p)} - 2 \frac{1}{\phi''[\frac{1}{2}(p+q)]} \right\}$$

and

$$d(p,q) = -2\phi''(p)\phi''(q)\phi''[\frac{1}{2}(p+q)] \left\{ \frac{1}{\phi''(p)} + \frac{1}{\phi''(q)} - 2 \frac{1}{\phi''[\frac{1}{2}(p+q)]} \right\}$$

Since the expression in the last curly bracket is the Jensen-difference (or the J-divergence) of $(\phi'')^{-1}$ we conclude (see also [5]):

Theorem 2. $J_\phi(p,q)$ is convex (concave) on $U \times U$ (with respect to $F \times F \supset U \times U$) if and only if ϕ is concave (convex) and $(\phi'')^{-1}$ is convex (concave) on I .

As a corollary of this theorem we obtain the following result on $J_\alpha(p,q)$ of (3.4):

Corollary 2. $J_\alpha(p,q)$ is never concave on $F_1 \times F_1$. It is convex on $F_1 \times F_1$ if and only if $\alpha \in [1,2]$.

Proof. The case of $\alpha = 0$ is degenerate for $J_0(p,q) \equiv 0$. We therefore assume that $\alpha \neq 0$. Also, since the case $\alpha = 1$ is a limiting case of $\alpha \rightarrow 1$, we may also assume that $\alpha \neq 1$. From (2.11) we deduce that $\phi_\alpha''(x) = -\alpha x^{\alpha-2}$ for $x \in \mathbb{R}_+$, while for $f_\alpha(x) \equiv [\phi_\alpha''(x)]^{-1}$ we have

$$f_\alpha''(x) = \alpha^{-1}(\alpha-1)(2-\alpha)x^{-\alpha}, \quad x \in \mathbb{R}_+.$$

The result follows at once.

For the proof of the following corollary we refer the reader to [5]:

Corollary 3. Assume that U is an open convex subset of F , such that

$$\bigcup \{p(x) \in \mathbb{R} : p \in U, x \in X\} = I \equiv (0,1).$$

Let $f_\alpha(x) = \phi_\alpha(x) + \phi_\alpha(1-x)$, $x \in I$ where ϕ_α is given by (2.11). Then $J_{f_\alpha}(p,q)$ is never concave on $U \times U$. It is convex on $U \times U$ if and only if $\alpha \in [1,2]$ or $\alpha \in [3, 11/3]$.

When ϕ is of class C^4 on the interval I , the condition of Theorem 2 may be summarized as one single condition, namely that the matrix

$$M_{\phi}(x) = \begin{bmatrix} \phi^{(2)}(x) & \sqrt{2} \phi^{(3)}(x) \\ \sqrt{2} \phi^{(3)}(x) & \phi^{(4)}(x) \end{bmatrix} .$$

be negative (or positive) definite for all $x \in I$. This means that $\phi^{(2)}(x) \leq 0$ (or $\phi^{(2)}(x) \geq 0$) and $A_{\phi}(x) \equiv \det\{M_{\phi}(x)\} \geq 0$, for all $x \in I$. This condition may serve to single out $\phi_1(x)$ and $\phi_2(x)$ of (2.11) and therefore, the entropies $H_1(p)$ and $H_2(p)$ of (2.13). Indeed, the following hold:

Theorem 3. The general solution of

$$A_{\phi}(x) = \det\{M_{\phi}(x)\} \equiv 0, \quad \phi^{(2)}(x) > 0; \quad x \in \mathbb{R}_+, \quad (3.12)$$

is one of the following two forms:

$$\phi(x) = \frac{1}{c^2} [(cx+b)\log(cx+b) - cx] + dx + e$$

where c, b, d and e are constants with $c > 0$ and $b \geq 0$, or

$$\phi(x) = ax^2 + kx + r$$

where a, k and r are constants with $a > 0$. In particular

$\phi(x) = -\phi_1(x) = x \log x$ is the only solution of (3.12)

subject to the conditions:

$$\phi(1) = 0, \quad \phi^{(1)}(1) = 1, \quad \phi^{(2)}(1) = 1, \quad \phi^{(3)}(1) = -1.$$

Similarly, $\phi(x) = -\phi_2(x) = x^2 - x$ is the only solution of (3.12) subject to the conditions:

$$\phi(1)=0, \phi^{(1)}(1)=1, \phi^{(2)}(1)=2.$$

Proof. When $\phi^{(2)}(x) \equiv a = \text{const.}$, the second form is obtained. When $\phi^{(2)}(x) \neq \text{const.}$, we let $f(x) = [\phi^{(2)}(x)]^{-1}$.

Then

$$f^{(2)}(x) = - \frac{1}{[\phi^{(2)}(x)]^3} A_\phi(x)$$

and so $f^{(2)}(x) = 0$, which means $(1/\phi^{(2)})^{(2)} = 0$. The result follows now at once.

3.3 The K-divergence

As for the Hessian K_ϕ at $(p, q) \in F_1 \times F_1$ along $(f, g) \in F \times F$ we have by virtue of (3.2)

$$\Delta_{(f,g)} K(p, q) = - \int_X \{a(p, q)f^2 + 2b(p, q)fg + a(q, p)g^2\} d\mu(x) \quad (3.13)$$

where, for $x, y \in \mathbb{R}_+$,

$$a(x, y) = \phi''(x) - y\psi''(x), \quad \psi(x) \equiv \phi(x)/x, \quad (3.14)$$

and

$$b(x, y) = -[\psi'(x) + \psi'(y)].$$

It therefore follows that $K_\phi(p, q)$ is convex on $F_1 \times F_1$ if and only if $a(x, y) \leq 0$ and

$$d(x, y) \equiv a(x, y)a(y, x) - [b(x, y)]^2 \geq 0; \quad x, y \in \mathbb{R}_+. \quad (3.15)$$

However, from (3.14) it is seen that $a(x,y) \leq 0$ whenever ϕ is concave and ψ is convex on $\mathbb{R}_+$, a situation identical with that of Theorem 1(v). We clearly have:

Theorem 4. $K_\phi(p,q)$ is convex on $F_1 \times F_1$ (with respect to $F \times F$) if ϕ is concave, ψ is convex on $\mathbb{R}_+$ and (3.15) holds.

As a corollary we obtain the following result on $K_\alpha(p,q)$ of (3.5), the proof of which is to be found in [5]:

Corollary 4. $K_\alpha(p,q)$ is convex on $F_1 \times F_1$ for all $\alpha \in [1,2]$.

3.4 The L-divergence

The Hessian of L_ϕ at $(p,q) \in F_1 \times F_1$ along $(f,g) \in F \times F$ is, in view of (3.3),

$$\Delta_{(f,g)} L_\phi(p,q) = - \int_X \{a(p,q)f^2 + 2b(p,q)fg + a(q,p)g^2\} d\mu(x) \quad (3.16)$$

where

$$a(p,q) = \frac{1}{q} \phi''\left(\frac{p}{q}\right) + \frac{q^2}{p^3} \phi''\left(\frac{q}{p}\right) \quad (3.17)$$

and

$$b(p,q) = -\frac{p}{q^2} \phi''\left(\frac{p}{q}\right) - \frac{q}{p^2} \phi''\left(\frac{q}{p}\right).$$

In this case, the discriminant

$$d(x,y) = a(x,y)a(y,x) - [b(x,y)]^2$$

is identically zero on $\mathbb{R}_+ \times \mathbb{R}_+$. This leads to the following result (see also [5]):

Theorem 5. $L_\phi(p, q)$ is convex (concave) on $F_1 \times F_1$ (with respect to $F \times t$) if and only if the function $F(x) \equiv x\phi(x^{-1}) + \phi(x)$ is concave (convex) on $\mathbb{R}_+$.

Proof. Since $d(x, y) \equiv 0$ on $\mathbb{R}_+ \times \mathbb{R}_+$, $L_\phi(p, q)$ is convex on $F_1 \times F_1$ if and only if

$$a(x, y) = \frac{y^2}{x^3} \left\{ \frac{x^3}{y^3} \phi''\left(\frac{x}{y}\right) + \phi''\left(\frac{y}{x}\right) \right\} \leq 0 \quad ; \quad (x, y) \in \mathbb{R}_+ \times \mathbb{R}_+.$$

Putting $t = y/x$, this condition becomes

$$t^{-3} \phi''(t^{-1}) + \phi''(t) \leq 0 \quad ; \quad t \in \mathbb{R}_+.$$

This means that $F''(t) \leq 0$ and the result follows.

In particular, for $L_\alpha(p, q)$ of (3.6) we have:

Corollary 5. $L_\alpha(p, q)$ is convex and (concave) on $F_1 \times F_1$ for all $\alpha \geq 0$ (or all $\alpha \leq 0$).

Proof. The case $\alpha = 0$ is degenerate for then $K_0(p, q) \equiv 0$. Also, by continuity we may assume that $\alpha \neq 1$. From (2.11) and $F_\alpha(x) \equiv x\phi_\alpha(x^{-1}) + \phi_\alpha(x)$, we deduce that

$$F''_\alpha(x) = -\alpha(x^{\alpha-2} + x^{-\alpha-1}) \quad ; \quad x \in \mathbb{R}_+$$

and the result follows.

4. GEODESIC DISTANCES

We return to the α -order entropy metric in (2.14)-

(2.15). The emphasis of the subsequent analysis will be in finding the α -order entropy pseudo-distance S_α for known multiparametric families of probability distributions F_Ω . When $\alpha=1$, such an analysis was carried out by Rao [13] and more recently by Atkinson and Mitchell [1], where the distance S is explicitly evaluated for certain multiparametric families F_Ω . We shall not repeat the examples of [1,13] for their extensions to the case of $\alpha \neq 1$ is not particularly difficult. An exception will be made for families of normal distributions, where it seems that the present analysis is slightly more general and, perhaps, simpler than that found in [1,13].

Being the geodesic pseudo-distance induced by $ds_\alpha^2(\theta)$ of (2.14)-(2.15), S_α may be evaluated with the aid of the Euler-Lagrange equations which involve the Christoffel symbols based on the α -order entropy matrix $[g_{km}^{(\alpha)}(\theta)]$ of (2.15). In general, such an undertaking may prove difficult as far as an explicit closed expression for S_α is sought.

4.1 Multinomial Distributions

Consider a multinomial discrete distribution $p(x|\theta) = p(x|\theta_1, \dots, \theta_n)$ where the sample space X is the set of integers $X = X_n = \{1, 2, \dots, n\}$ and $p(k|\theta) = \theta_k$ for $k \in X_n$. In this case, $g_{km}^{(\alpha)}$ of (2.15) is

$$g_{km}^{(\alpha)}(\theta) = \int_{X_n} [p(k|\theta)]^{\alpha-2} \delta_{km} d\mu(k) = \delta_{km} \theta_k^{\alpha-2}; \quad k, m = 1, \dots, n.$$

we may use the identification

$$p_k \equiv p(k|\theta) = \theta_k \quad ; \quad k=1, \dots, n.$$

We shall assume first that

$$p \equiv (p_1, \dots, p_n) \in \mathbb{R}_+^n \quad ;$$

and then make the restriction of $p \in \Omega_n$, where

$$\Omega_n \equiv \{p \in \mathbb{R}_+^n : \sum_{k=1}^n p_k = 1, 0 < p_k < 1, k=1, \dots, n\}.$$

With these considerations the metric of (2.14) may be expressed as

$$ds_\alpha^2(p) = \sum_{k=1}^n p_k^{\alpha-2} (dp_k)^2, \quad p \in \mathbb{R}_+^n.$$

The fundamental tensor of the metric $g_{km}^{(\alpha)} = \delta_{km} p_k^{\alpha-2}$ is of rank n and, therefore, S_α is indeed a distance. The evaluation of this geodesic distance is immediate, and, for $p, q \in \mathbb{R}_+^n$, we have

$$S_\alpha(p, q) = \begin{cases} 2|\alpha|^{-1} \left\{ \sum_{k=1}^n [p_k^{\alpha/2} - q_k^{\alpha/2}]^2 \right\}^{1/2}, & \alpha \neq 0 \\ \left\{ \sum_{k=1}^n [\log p_k - \log q_k]^2 \right\}^{1/2}, & \alpha = 0 \end{cases}$$

which is (modulo a factor of $\sqrt{2}$) the α -order Hellinger distance $M_\alpha(p, q)$ as in (3.7). The same results hold with the restriction of $p, q \in \Omega_n$.

4.2 Normal Distributions

We first consider a two-parameter family of normal

distributions $p(\cdot | \mu, \sigma) \equiv N(\mu, \sigma^2)$ with mean μ and variance σ^2 ($-\infty < \mu < \infty; \sigma > 0$). Here, for reasons of convergence we must assume that $\alpha > 0$. Fixing $\alpha > 0$, it will be found convenient to introduce new variables x and y ($-\infty < x < \infty; y > 0$) via

$$y = \sigma, \quad x = \{A(\alpha)\}^{-\frac{1}{2}} \mu; \quad A(\alpha) \equiv (\alpha^{\frac{1}{2}} - \alpha^{-\frac{1}{2}})^2 + 2\alpha^{-1}, \quad \alpha > 0. \quad (4.1)$$

We may consider the complex parameter

$$z = x + iy \in W \equiv U = \{z \in \mathbb{C} : \text{Im} z > 0\} \quad (4.2)$$

with U being the upper half-plane. In this way $p(\cdot | \mu, \sigma)$ is replaced by $p(\cdot | z) \equiv N(\mu, \sigma^2)$ with $z \in U$ as in (4.2)-(4.3). Now, a routine calculation, omitted here, shows that the metric (2.14)-(2.15) admits the form

$$ds_{\alpha}^2(z) = B(\alpha) y^{-(\alpha+1)} |dz|^2 \quad (4.3)$$

where

$$B(\alpha) \equiv \alpha^{-3/2} (2\pi)^{\frac{1-\alpha}{2}} A(\alpha), \quad \alpha > 0. \quad (4.4)$$

The metric in (4.3) constitutes a Kähler metric on the upper half-plane U and when $\alpha=1$, it reduces to the familiar Poincaré metric. The Gaussian curvature of (4.3) is

$$\kappa_{\alpha}(z) = -(\alpha+1) \{2B(\alpha)\}^{-1} y^{\alpha-1}; \quad y = \text{Im} z > 0, \quad \alpha > 0,$$

and, is always negative. In particular, $\kappa_1(z) \equiv -2^{-1}$.
 In this case, S_α is indeed a distance on U and $S \equiv S_1$
 is the familiar hyperbolic distance of U .

We now treat this distance $S_\alpha (\alpha > 0)$:

1. The case of $\alpha=1$: In this case, by (4.1)-(4.4),

$$ds^2(z) = 2y^{-2} |dz|^2. \quad (4.5)$$

Elementary arguments based on the invariance properties
 of this metric of Poincaré lead to the following geodesic
 distance (or "Poincaré distance"):

$$S(z, \zeta) = \sqrt{2} \log \frac{1 + \delta(z, \zeta)}{1 - \delta(z, \zeta)} ; \quad z, \zeta \in U, \quad (4.6)$$

where

$$\delta(z, \zeta) = \left| \frac{z - \bar{\zeta}}{z - \zeta} \right| ; \quad z, \zeta \in U. \quad (4.7)$$

It should be noted that $\delta \equiv \delta(z, \zeta)$ is also a distance on
 U and is called the "Möbius distance" (see also [3,4]
 for further generalizations of these distances). Also,
 the geodesics of (4.5) (see for example, (4.16)) are
 given by the "semi-circles"

$$z = a + re^{i\theta} ; \quad r > 0, \quad 0 < \theta < \pi, \quad (4.8)$$

where a is a real fixed constant.

Expressed in terms of the original parameters μ
 and σ , the distance in (4.6), by virtue of (4.1) and
 (4.7), may be written as

$$S(\mu_1, \sigma_1; \mu_2, \sigma_2) = \sqrt{2} \log \frac{1 + \delta(\mu_1, \sigma_1; \mu_2, \sigma_2)}{1 - \delta(\mu_1, \sigma_1; \mu_2, \sigma_2)} \quad (4.9)$$

where

$$\delta(\mu_1, \sigma_1; \mu_2, \sigma_2) = \left[\frac{(\mu_1 - \mu_2)^2 + 2(\sigma_1 - \sigma_2)^2}{(\mu_1 - \mu_2)^2 + 2(\sigma_1 + \sigma_2)^2} \right]^{\frac{1}{2}} \quad (4.10)$$

is the Möbius distance (4.7) in terms of μ and σ . This is the required distance between $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$. It agrees with a rather more involved expression obtained by Atkinson and Mitchell [1]. The expression in [1] can be obtained from (4.9) by using (4.1), (4.8) and (4.10) note that always

$$0 \leq \delta(\mu_1, \sigma_1; \mu_2, \sigma_2) < 1.$$

On the other hand, the Poincaré distance $S(\mu_1, \sigma_1; \mu_2, \sigma_2)$ clearly satisfies

$$S(\mu_1, \sigma_1; \mu_2, \sigma_2) \geq 2\sqrt{2} \cdot \delta(\mu_1, \sigma_1; \mu_2, \sigma_2).$$

The Hellinger pseudo-distance (3.7) between $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ is, in this case, a proper distance with the following form:

$$M(\mu_1, \sigma_1; \mu_2, \sigma_2) = 2 \left[1 - \left(\frac{2\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2} \right)^{\frac{1}{2}} e^{-(\mu_1 - \mu_2)^2 / 4(\sigma_1^2 + \sigma_2^2)} \right]^{\frac{1}{2}} \quad (4.11)$$

2. The case of $\alpha \neq 1$ In this case, the geodesic distance S_α of the metric (4.3) is not easily explicated as in the former case. We shall first find all the geodesics of this metric. This may be, of course, done with the aid of the Christoffel symbols of the metric (4.3). We shall, however, proceed directly, for reasons of economy and clarity. Writing

$$\beta = (\alpha + 1)/2 \quad ; \quad \beta > 1/2, \quad (4.12)$$

finding the geodesics of (4.3) amounts to solving the following extremal problem of calculus of variations (the factor $B(\alpha) > 0$ is irrelevant here!):

$$\min \int_a^b y^{-\beta} \sqrt{1 + (y')^2} dx, \quad y > 0,$$

where the minimum is taken over all C^2 -paths $y = f(x)$, joining the points $(a, f(a))$ and $(b, f(b))$. A routine calculation based on the Lagrangian of $y^{-\beta} \sqrt{1 + (y')^2}$ shows that the Euler-Lagrange equations of this problem admit the simple form

$$yy'' = -\beta[1 + (y')^2]. \quad (4.13)$$

In order to solve (4.13) we proceed with standard methods, letting

$$p = \frac{dy}{dx}, \quad y'' = p \frac{dp}{dy}$$

to obtain

$$-\beta d(\log y) = 2 d[\log(1+p^2)] .$$

This shows that

$$y^{-\beta} = r^{-1}(1+p^2)^{\frac{1}{2}} ; \quad r > 0 ,$$

with $r > 0$ being a constant. Consequently,

$$x = \pm \int \frac{y^{\beta}}{(r^2 - y^{2\beta})^{\frac{1}{2}}} dy + a ; \quad r > y^{\beta} > 0, \quad (4.14)$$

where a is an arbitrary constant of integration. We may use the substitution $y = r^{1/\beta} \sin^{1/\beta} \theta$, $0 < \theta < \pi$ and upon introducing the one parameter family of functions

$$F_{\gamma}(\theta) \equiv -\gamma \int_{\pi/2}^{\theta} \sin^{\gamma} t dt ; \quad \gamma \in \mathbb{R}, \quad 0 < \theta < \pi, \quad (4.15)$$

the solution (4.14) may be written in the parametric form:

$$x = a \pm r^{1/\beta} F_{1/\beta}(\theta) , \quad y = r^{1/\beta} \sin^{1/\beta} \theta ; \quad r > 0 , \quad 0 < \theta < \pi. \quad (4.16)$$

When $\beta=1$, or, by (4.12), when $\alpha=1$, (4.16) reduces to (4.8). Equation (4.16) gives all the geodesics of the problem. We also note that the geodesics in (4.16) include the lines $x=\text{const.}$ as a limiting case, corresponding to $r \rightarrow \infty$.

An expression for $S_{\alpha}(z, \zeta)$, $z, \zeta \in U$, may now be given by using (4.3) and (4.16). We have

$$S_{\alpha}(z, \zeta) = \sqrt{B(\alpha)} \frac{r^{\frac{1}{\beta}-1}}{2\beta-1} \left| F_{\frac{1}{\beta}-2}(\theta_2) - F_{\frac{1}{\beta}-2}(\theta_1) \right| \quad (4.17)$$

where, after choosing, without loss, the (+) sign in (4.16),

$$\begin{aligned} z=x+iy, \quad x=a+r^{1/\beta} F_{1/\beta}(\theta_1), \quad y=r^{1/\beta} \sin^{1/\beta} \theta_1, \\ \zeta=\xi+i\eta, \quad \xi=a+r^{1/\beta} F_{1/\beta}(\theta_2), \quad \eta=r^{1/\beta} \sin^{1/\beta} \theta_2. \end{aligned} \quad (4.18)$$

Using (4.1), (4.4), (4.12) and (4.15) one deduces immediately that (4.17) reduces to (4.6) when $\alpha=1$. In general the quantities θ_1, θ_2 and r are determined by the given $z=x+iy, \zeta=\xi+i\eta \in U$ via (4.18). However, except for special values of $\alpha > 0$ where integrals of type (4.15) can be further explicated, finding a closed form formula for $S_\alpha(z, \zeta)$ in terms of z and ζ may prove difficult.

One may use an alternative expression for $S_\alpha(z, \zeta)$ which, sometimes, is simpler than that of (4.18). It is based on the recursive formula

$$F_{\gamma-2}(\theta) = \frac{\gamma-2}{\gamma-1} [F_\gamma(\theta) - \cos \theta \sin^{\gamma-1} \theta],$$

valid for all real γ and easily derived from (4.15).

Using this formula, together with (4.12) and (4.18), (4.17) becomes

$$S_\alpha(z) = \frac{2\sqrt{B(\alpha)}}{|1-\alpha|} \left| \frac{x-\xi}{r} + y^{\frac{1}{2}(1-\alpha)} (1-r^{-2}y^{\alpha+1})^{\frac{1}{2}} - \eta^{\frac{1}{2}(1-\alpha)} (1-r^{-2}\eta^{\alpha+1})^{\frac{1}{2}} \right| \quad (4.23)$$

Letting $r \rightarrow \infty$ in (4.19), corresponds to the geodesic $x=\text{const.}$, and, accordingly

$$S_{\alpha}(z) = \frac{2\sqrt{B(\alpha)}}{|1-\alpha|} \left| y^{\frac{1}{2}(1-\alpha)} - \eta^{\frac{1}{2}(1-\alpha)} \right|; z, \zeta \in U, \operatorname{Re} z = \operatorname{Re} \zeta.$$

This, of course, agrees with (4.6) as $\alpha \rightarrow 1$ and $\operatorname{Re} z = \operatorname{Re} \zeta$.

The α -order distance $S_{\alpha}(\mu_1, \sigma_1; \mu_2, \sigma_2)$ between $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ can be derived from (4.19) by using (4.1), (4.4) and (4.18). In particular,

$$S_{\alpha}(\mu, \sigma_1; \mu, \sigma_2) = \frac{2\sqrt{B(\alpha)}}{|1-\alpha|} \left| \sigma_1^{\frac{1}{2}(1-\alpha)} - \sigma_2^{\frac{1}{2}(1-\alpha)} \right|$$

which agrees with (4.9) as $\alpha \rightarrow 1$ and $\mu = \mu_1 = \mu_2$.

The α -order Hellinger distance between $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ is now

$$M_{\alpha}(\mu_1, \sigma_1; \mu_2, \sigma_2) = \frac{\sqrt{2}}{\alpha^{3/4}} (2\pi)^{\frac{1-\alpha}{4}} \left\{ (\sigma_1^{\frac{1-\alpha}{2}} - \sigma_2^{\frac{1-\alpha}{2}})^2 + 2(\sigma_1 \sigma_2)^{\frac{1-\alpha}{2}} E_{\alpha}(\mu_1, \sigma_1; \mu_2, \sigma_2) \right\}^{\frac{1}{2}}$$

where

$$E_{\alpha}(\mu_1, \sigma_1; \mu_2, \sigma_2) = 1 - \left(\frac{2\sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2} \right)^{\frac{1}{2}} e^{-\alpha(\mu_1 - \mu_2)^2 / 4(\sigma_1^2 + \sigma_2^2)}.$$

When $\alpha=1$, this formula reduces to (4.11).

4.3 Products of Normal Distributions

The previous methods can be extended to products of normal distributions

$$p(x|\theta) = \prod_{k=1}^n N(x_k; \mu_k, \sigma_k^2), \quad (4.20)$$

where $x=(x_1, \dots, x_n) \in X \subseteq \mathbb{R}^n$ and $\theta=(\mu_1, \sigma_1, \dots, \mu_n, \sigma_n) \in \mathbb{R}^{2n}$ with means μ_k and variances σ_k^2 ($-\infty < \mu_k < \infty, \sigma_k > 0; k=1, \dots, n$).

As in (4.1)-(4.2), we find it convenient to introduce new variables. Accordingly, we replace $x=(x_1, \dots, x_n)$ by $t=(t_1, \dots, t_n) \in X = \mathbb{R}^n$ and write for the parameters

$$y_k = \sigma_k, \quad x_k = \{A(\alpha)\}^{-\frac{1}{2}} \mu_k; \quad A(\alpha) \equiv (\alpha^{\frac{1}{2}} - \alpha^{-\frac{1}{2}})^2 + 2\alpha^{-1}, \quad \alpha > 0, \quad (4.21)$$

and

$$z=(z_1, \dots, z_n), \quad z_k = x_k + iy_k, \quad (-\infty < x_k < \infty; y_k > 0); k=1, \dots, n. \quad (4.22)$$

Plainly, we view the distribution in (4.20) as $p(t|z)$ with t in the sample space X and $z \in U^n$, n copies of the upper half-plane U .

As in (4.3) the metric (2.14)-(2.15) admits here the form

$$ds_\alpha^2(z) = B_n(\alpha) \left\{ \prod_{j=1}^n y_j^{1-\alpha} \right\} \sum_{k=1}^n y_k^{-2} |dz_k|^2 \quad (4.23)$$

where

$$B_n(\alpha) = \alpha^{-(n+2)/2} (2\pi)^{\frac{1}{2}n(1-\alpha)} A(\alpha), \quad \alpha > 0. \quad (4.24)$$

When $n=1$, (4.23)-(4.24) reduce, of course, to (4.4)-(4.5). The case of $\alpha=1$ is, as before, rather involved and since we cannot expect a closed form formula for the geodesic distance S_α , we shall only deal with the case of $\alpha=1$. In this case, by (4.21)-(4.24),

$$ds^2(z) = 2 \sum_{k=1}^n y_k^{-2} |dz_k|^2 \quad (4.25)$$

which is, as in (4.5), the Poincaré metric on U^n .

In order to find the geodesic distance S we exploit the fact that the metric (4.25) is (globally) invariant under biholomorphic mappings. Accordingly, we use the mapping

$$\omega_k = \frac{z_k - i}{z_k + i} \quad ; \quad z_k \in U, \quad 1 \leq k \leq n \quad (4.26)$$

which maps U^n biholomorphically onto the polydisk $D^n = \{\omega = (\omega_1, \dots, \omega_n) \in C^n : |\omega_k| < 1, k=1, \dots, n\}$. With this mapping the metric in (4.25) becomes

$$ds^2(\omega) = 8 \sum_{k=1}^n (1 - |\omega_k|^2)^{-2} |d\omega_k|^2 \quad (4.27)$$

which is the Poincaré metric on the polydisk D^n . We first find the geodesic distance $S(\omega, \tau)$ of this metric when $\omega, \tau \in D^n$. In order to do so we assume that $\tau = 0 = (0, \dots, 0)$ and evaluate $S(\omega, 0), \omega \in D^n$. We write

$$r = (r_1, \dots, r_n), \quad r_k = |\omega_k|, \quad 0 \leq r_k < 1; k=1, \dots, n$$

and note that due to the invariance of (4.27), $S(\omega, 0) = S(r, 0)$. In this way, we have

$$ds^2(r) = 8 \sum_{k=1}^n \left(\frac{dr_k}{1-r_k^2} \right)^2 = 2 \sum_{k=1}^n \left(d \log \frac{1+r_k}{1-r_k} \right)^2$$

and consequently

$$S(\omega, 0) = \sqrt{2} \left[\sum_{k=1}^n \log^2 \frac{1+|\omega_k|}{1-|\omega_k|} \right]^{\frac{1}{2}}.$$

This, as is well known, is sufficient for the determination

of the distance between any two points of D^n . Indeed, given two points $\omega, \tau \in D^n$ there exists a holomorphic automorphism ϕ of D^n on D^n so that $\phi(\omega) = v, \phi(\tau) = 0 \in D^n$. Again, by invariance, $S(\omega, \tau) = S(\phi(\omega), \phi(\tau)) = S(v, 0)$. Here, up to a rotation

$$k = \frac{\omega_k - \tau_k}{1 - \bar{\tau}_k \omega_k}, \quad k=1, \dots, n.$$

It therefore follows that

$$S(\omega, \tau) = \sqrt{2} \left[\sum_{k=1}^n \log 2 \frac{1 + \delta(\omega_k, \tau_k)}{1 - \delta(\omega_k, \tau_k)} \right]^{\frac{1}{2}}; \quad \omega, \tau \in D^n, \quad (4.28)$$

where

$$\delta(\omega_k, \tau_k) = \frac{\omega_k - \tau_k}{1 - \bar{\tau}_k \omega_k}, \quad k=1, \dots, n. \quad (4.29)$$

Returning to the metric in (4.25), its geodesic distance $S(z, \zeta)$ between two points $z, \zeta \in U^n$ is obtained from (4.28)-(4.29) and the mapping in (4.26). This gives

$$S(z, \zeta) = \sqrt{2} \left\{ \sum_{k=1}^n \log 2 \frac{1 + \delta(z_k, \zeta_k)}{1 - \delta(z_k, \zeta_k)} \right\}^{\frac{1}{2}}; \quad z, \zeta \in U^n, \quad (4.30)$$

with

$$\delta(z_k, \zeta_k) = \left| \frac{z_k - \zeta_k}{z_k - \bar{\zeta}_k} \right|, \quad k=1, \dots, n. \quad (4.31)$$

This generalizes (4.7)-(4.8). Finally, from (4.30)-(4.31) the information distance $S_n(\mu, \sigma; \nu, \rho)$ between a $\prod_{k=1}^n N(t_k; \mu_k, \sigma_k^2)$ distribution and a $\prod_{k=1}^n N(t_k; \nu_k, \rho_k^2)$ distribution is given by

$$S_n(\mu, \sigma; \nu, \rho) = \sqrt{2} \left[\sum_{k=1}^n \log^2 \frac{1 + \delta(\mu_k, \sigma_k; \nu_k, \rho_k)}{1 - \delta(\mu_k, \sigma_k; \nu_k, \rho_k)} \right]^{\frac{1}{2}} \quad (4.31)$$

with

$$\delta(\mu_k, \sigma_k; \nu_k, \rho_k) = \frac{[(\mu_k - \nu_k)^2 + 2(\sigma_k - \rho_k)^2]^{\frac{1}{2}}}{[(\mu_k - \nu_k)^2 + 2(\sigma_k + \rho_k)^2]^{\frac{1}{2}}}, \quad k=1, \dots, n \quad (4.32)$$

and, where

$$\mu = (\mu_1, \dots, \mu_n), \sigma = (\sigma_1, \dots, \sigma_n); \nu = (\nu_1, \dots, \nu_n), \rho = (\rho_1, \dots, \rho_n).$$

In view of (4.31)-(4.32) and (4.9) - (4.10) we may conclude the following desirable property of the information distance:

$$S_n^2(\mu, \sigma; \nu, \rho) = \sum_{k=1}^n S^2(\mu_k, \sigma_k; \nu_k, \rho_k). \quad (4.33)$$

5. THE CARATHÉODORY PSEUDO-DISTANCE

The information distance $S(\mu, \sigma; \nu, \rho)$ between $N(\mu, \sigma^2)$ and $N(\nu, \rho^2)$ given as in (4.9) suggests an introduction of a pseudo-distance on a theme of Carathéodory (see [4] for a further generalization). We briefly discuss this possibility and refer the reader to Burbea [2,3,4] and the book of Kobayashi [8, pp. 49-53] for further details.

We assume that the family of multiparametric probability distributions F_Ω is such that Ω is a complex manifold in $\mathbb{C}^n$. Thus $p(\cdot|z) \in F_\Omega$ with $z=(z_1, \dots, z_n) \in \Omega$ being an n -tuple of complex parameters $z_j = x_j + iy_j$, $1 \leq j \leq n$. We consider the Möbius and Poincaré distances δ and S on the upper half-plane U , as given in (4.6)-(4.7). Let $H(\Omega; U)$ denote the family of holomorphic functions from Ω into U . We define

$$\delta_\Omega(z, \zeta) = \sup \{ \delta(f(z), f(\zeta)) : f \in H(\Omega; U) \} ; z, \zeta \in \Omega.$$

A normal family argument shows that the supremum is attained. It is also clear that δ_Ω satisfies all axioms of a pseudo-distance on Ω . It is called the Möbius pseudo-distance of Ω . The Carathéodory pseudo-distance of Ω is defined by

$$S_\Omega(z, \zeta) = \sup \{ S(f(z), f(\zeta)) : f \in H(\Omega; U) \} ; z, \zeta \in \Omega.$$

Again, the supremum is attained and by (4.6)-(4.7)

$$S_\Omega(z, \zeta) = \sqrt{2} \log \frac{1 + \delta_\Omega(z, \zeta)}{1 - \delta_\Omega(z, \zeta)}.$$

Both pseudo-distances become distances on Ω when Ω is biholomorphically equivalent to a bounded domain in $\mathbb{C}^n$. It is also clear that

$$S_\Omega(z, \zeta) \geq 2\sqrt{2} \delta_\Omega(z, \zeta), \quad 0 \leq \delta_\Omega(z, \zeta) < 1 ; z, \zeta \in \Omega.$$

Let $\phi: \Omega \rightarrow \Omega^*$ be a holomorphic mapping of a complex manifold Ω of $\mathbb{C}^n$ into another complex manifold Ω^* of $\mathbb{C}^m$.

Then, for $z, \zeta \in \Omega$,

$$\delta_{\Omega}^*(\phi(z), \phi(\zeta)) \leq \delta_{\Omega}(z, \zeta)$$

and

$$S_{\Omega}^*(\phi(z), \phi(\zeta)) \leq S_{\Omega}(z, \zeta).$$

In particular, δ_{Ω} and S_{Ω} are biholomorphically invariants.

Also, in the case that Ω is the upper half-plane U , we have

$$\delta_U = \delta, \quad S_U = S$$

and, therefore, δ_{Ω} and S_{Ω} constitute a natural generalization of δ and S in (4.6)-(4.7)

when $\Omega = U^n$ we have, contrary to (4.33).

$$\delta_{U^n}(z, \zeta) = \max\{\delta(z_1, \zeta_1), \dots, \delta(z_n, \zeta_n)\}$$

and, therefore,

$$S_{U^n}(z, \zeta) = \max\{S(z_1, \zeta_1), \dots, S(z_n, \zeta_n)\}$$

where $z = (z_1, \dots, z_n)$, $\zeta = (\zeta_1, \dots, \zeta_n) \in U^n$.

References

- [1] Atkinson, C. and Mitchell, A. F. S. (1980). Rao's distance measure. Sankhyā, (in press).
- [2] Burbea, J. (1977). The Carathéodory metric and its majorant metrics. Canad. J. Math., 29, 771-780.
- [3] Burbea, J. (1980). A generalization of Pick's theorem and its applications to intrinsic metrics. Annal. Polon. Math., 39, (in press).
- [4] Burbea, J. (1980). On metrics and distortion theorems. Annals Math. Studies, 120 (in press).
- [5] Burbea, J. and Rao, C. R. (1980). On the convexity of some divergence measures based on entropy functions. Tech. Rep. No. 80-13, University of Pittsburgh.
- [6] Fisher, R. A. (1925). Theory of statistical estimation. Proc. Camb. Phil. Soc., 22, 700-725.
- [7] Havrda, M. E. and Charvát, F. (1967). Quantification method of classification processes: Concept of structural α -entropy. Kybernetika, 3, 30-35.
- [8] Kobayashi, S. (1970). Hyperbolic Manifolds and Holomorphic Mappings. Marcel Dekker, New York.
- [9] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. Ann. Math. Statist., 22, 79-86.
- [10] Matusita, K. (1955). Decision rules based on the distance, for problem of fit, two samples and estimation. Ann. Math. Statist., 26, 631-640.
- [11] Matusita, K. (1957). Decision rule based on the distance for the classification problem. Ann. Inst. Statist. Math., 8, 67-77.
- [12] Pitman, E. J. G. (1979). Some Basic Theory for Statistical Inference. Halsted Press.
- [13] Rao, C. R. (1945). Information and accuracy attainable in the estimation of statistical parameters. Bull. Calcutta Math. Soc., 37, 81-91.
- [14] Rao, C. R. (1949). On the distance between two populations. Sankhyā, 9, 246-248.

References

- [15] Rao, C. R. (1962). Efficient estimates and optimum inference procedures in large samples (with discussion). J. Roy. Statist. Soc. B., 24, 46-72.
- [16] Rao, C. R. (1973). Linear Statistical Inference and its Applications. John Wiley, New York.
- [17] Rao, C. R. (1980). Diversity and dissimilarity coefficients: a unified approach. Tech. Rep. 80-10, University of Pittsburgh.
- [18] Shannon, C. E. (1948). A mathematical theory of communications. Bell System Tech. J., 27, 379-423,

