PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 98, Number 1, September 1986

AN EXTENDED ČENCOV CHARACTERIZATION OF THE INFORMATION METRIC

L. L. CAMPBELL[1]

ABSTRACT. Čencov has shown that Riemannian metrics which are derived from the Fisher information matrix are the only metrics which preserve inner products under certain probabilistically important mappings. In Čencov's theorem, the underlying differentiable manifold is the probability simplex $\Sigma_1^n x_i = 1$, $x_i > 0$. For some purposes of using geometry to obtain insights about probability, it is more convenient to regard the simplex as a hypersurface in the positive cone. In the present paper Čencov's result is extended to the positive cone. The proof uses standard techniques of differential geometry but does not use the language of category theory.

1. Introduction. There has been a good deal of interest in the use of differential geometry to interpret certain operations on probability distributions in statistics [**2, 4, 5**], biomathematics [**9**], thermodynamics [**7**], and information theory [**3**]. Further references are to be found in those cited above. Much of this literature begins by introducing a Riemannian metric which is generated by the Fisher information matrix. One reason for singling out this particular metric is to be found in a theorem of Čencov [**4**, Theorem 11.1 or Lemma 11.3], which characterizes this information metric on the probability simplex as the only metric having a certain invariance property under some probabilistically natural mappings.

In this paper, we develop a characterization theorem which is closely related to Čencov's. The principal difference between the two theorems is that we characterize Riemannian metrics on the positive cones $\mathbf{R}_n^+ = \{x = (x_1, \dots, x_n)\colon\ x_i > 0\}$, while Čencov characterizes them on the probability simplexes $S_{n-1} = \{x \in \mathbf{R}_n^+ \colon \Sigma x_i = 1\}$. As will be seen later, the connection between geometry and probability is enhanced if S_{n-1} is regarded as a surface in the differentiable manifold $\mathbf{R}_n^+$. In addition, some of Shahshahani's development [**9**] requires a metric on $\mathbf{R}_n^+$. A second difference from Čencov's theorem is that neither the statement of our result nor the proof use the language of category theory.

2. Markov mappings. Let m and n be integers satisfying $2 \leqslant m \leqslant n$ and let $\{A_1, \dots, A_m\}$ be a partition of the set $\{1, 2, \dots, n\}$ into disjoint subsets. To each A_i,

Received by the editors February 21, 1985 and, in revised form, August 26, 1985.
1980 *Mathematics Subject Classification*. Primary 62E10; Secondary 53B20, 94A17.
Key words and phrases. Information metric, Markov mapping.
[1]Research supported by the Natural Sciences and Engineering Research Council of Canada, Grant No. A2151.

associate a probability vector $Q^{(i)} = (q_{i1}, \dots, q_{in})$ concentrated on A_i. That is,

$$q_{ij} = 0 \quad \text{if } j \notin A_i, \qquad q_{ij} > 0 \quad \text{if } j \in A_i, \qquad \sum_{j=1}^{n} q_{ij} = 1.$$

Define a mapping $f\colon \mathbf{R}_m^+ \to \mathbf{R}_n^+$ by

$$y_j = \sum_{i=1}^{m} x_i q_{ij}, \qquad j \in \{1, 2, \dots, n\}. \tag{1}$$

Note that f maps S_{m-1} into S_{n-1} (and, more generally, the simplex $\sum x_i = a$ into the simplex $\sum y_i = a$). Following Čencov [**4**, p. 56 and Lemma 9.5, p. 136], we shall refer to a mapping of this type as a congruent embedding of $\mathbf{R}_m^+$ in $\mathbf{R}_n^+$ by a Markov mapping.

Associated with the map f is another map $g\colon \mathbf{R}_n^+ \to \mathbf{R}_m^+$ defined by

$$x_i = \sum_{j \in A_i} y_j, \qquad i \in \{1, 2, \dots, m\}. \tag{2}$$

Note that f is one-to-one while g is many-to-one and that the composition $g \circ f$ is the identity map on $\mathbf{R}_m^+$.

The mappings (1) and (2), when restricted to the simplexes S_{m-1} and S_{n-1} respectively, have clear probabilistic interpretations. In fact, because of the special form of $Q^{(i)}$, there is exactly one positive term in each of the sums (1) because there is exactly one set A_i which contains j as an element. Hence, once the vectors $Q^{(i)}$ are chosen, the components y_j of $y = f(x)$ are each of the form $q_{ij} x_i$. Thus, in a quite elementary way, the image $f(S_{m-1})$ in S_{n-1} is like S_{m-1}. If we wish to change x in S_{m-1}, we can equally well look at the effect of changing y in $f(S_{m-1})$. This led Čencov to seek our Riemannian metrics on S_{m-1} and S_{n-1} which are invariant under congruent embeddings by Markov mappings. The notion of invariance used is that inner products should be unchanged when tangent vectors are mapped by the Jacobian map f_*. We develop a similar result on the manifolds $\mathbf{R}_m^+$ and $\mathbf{R}_n^+$.

Before going on, we remark that when $m = n$ the mapping f is just a permutation of the components of $(x_1, \dots, x_n)$ and g is the inverse permutation. When $m = n - 1$, we can, up to permutations, take $A_i = \{i\}$ for $i = 1, 2, \dots, n-2$ and $A_{n-1} = \{n-1, n\}$. For this special case,

$$g(y_1, \dots, y_n) = (y_1, y_2, \dots, y_{n-2}, y_{n-1} + y_n). \tag{3}$$

Permutations and mappings of the type (3) plan an essential role in the characterization of information measures [**1**]. More general mappings of the form (2) are easily seen to be compositions of mappings of the type (3).

3. Isometries. Our terminology and notation for differential geometry follow Hicks [**6**]. To each $x \in \mathbf{R}_m^+$ there is associated a tangent space M_x, the vector space of derivations evaluated at x. The set $\{X_i = \partial/\partial x_i, i = 1, 2, \dots, m\}$ is a basis for M_x and $\{Y_j = \partial/\partial y_j, i = 1, 2, \dots, n\}$ is similarly a basis for M_y, $y \in \mathbf{R}_n^+$. If $y = f(x)$, the Jacobian map f_* maps M_x to M_y so that $Y = f_* X$ means that $Y\phi = X\phi \circ f$, for every C^∞ function ϕ with domain a neighborhood of y. When f is defined by (1),

$$f_* X_i = \sum_{j=1}^{n} q_{ij} Y_j. \tag{4}$$

Now suppose that Riemannian metrics, i.e. C^∞ inner products on tangent spaces, are defined on $\mathbf{R}_m^+$ and $\mathbf{R}_n^+$. For vectors U and V in M_x, their inner product will be denoted by $\langle U, V\rangle_m(x)$; a similar notation is employed in $\mathbf{R}_n^+$. The mapping f is called an isometry if

$$\langle U,V\rangle_m(x) = \langle f_*U, f_*V\rangle_n(y) \tag{5}$$

for all $x \in \mathbf{R}_m^+$ and all U and V in M_x, where $y = f(x)$.

4. Characterization theorem. The principal result of this paper is the following variant of a result of Čencov [4, Theorem 11.1].

THEOREM. *Let $\langle\ ,\ \rangle_m$ be a Riemannian metric on $\mathbf{R}_m^+$ for $m \in \{2,3,\dots\}$. Let this sequence of metrics have the property that every congruent embedding by a Markov mapping is an isometry. Then*

$$\langle X_i, X_j\rangle_m(x) = A(|x|) + \delta_{ij}|x|B(|x|)/x_i, \tag{6}$$

where $|x| = \sum_i^m x_i$, δ_{ij} is the Kronecker delta, and A and B are C^∞ functions on $\mathbf{R}^+$ satisfying $B(\alpha) > 0$ and $A(\alpha) + B(\alpha) > 0$ for all $\alpha > 0$. Conversely, if A and B are C^∞ functions on $\mathbf{R}^+$ satisfying $B(\alpha) > 0$, $A(\alpha) + B(\alpha) > 0$, then (6) *defines a sequence of Riemannian metrics under which every congruent embedding by a Markov mapping is an isometry.*

PROOF. Let g_{ij} be defined by

$$g_{ij}^{(m)}(x) = \langle X_i, X_j\rangle_m(x), \quad g_{ij}^{(n)}(y) = \langle Y_i, Y_j\rangle_n(y), \tag{7}$$

where $y = f(x)$. Then $g_{ij} = g_{ji}$. Consider first the case $m = n$ and the Markov mapping f_{rs} which interchanges x_r and x_s while leaving the other coordinates unchanged. Then it is easily seen that f_{rs*} interchanges X_r and X_s; that is

$$f_{rs*}X_r = Y_s, \qquad f_{rs*}X_s = Y_r, \qquad f_{rs*}X_i = Y_i \quad \text{otherwise.}$$

Thus, by (5) and (7) and the symmetry of inner products,

$$\begin{aligned} g_{rj}^{(m)}(x) &= g_{sj}^{(m)}(y) \qquad \text{for } j \notin \{r,s\},\\ g_{rr}^{(m)}(x) &= g_{ss}^{(m)}(y),\\ g_{ij}^{(m)}(x) &= g_{ij}^{(m)}(y) \qquad \text{otherwise.}\end{aligned} \tag{8}$$

Next, consider (8) on the center line, $x_1 = x_2 = \cdots = x_m$. If we put $\bar{x} = (\alpha/m, \alpha/m, \dots, \alpha/m)$, where $\alpha = |\bar{x}|$, then $f_{rs}(\bar{x}) = \bar{x}$ and thus

$$g_{rr}^{(m)}(\bar{x}) = F_m(\alpha) \quad \text{for all } r \in \{1,2,\dots,m\} \tag{9}$$

and

$$g_{rs}^{(m)}(\bar{x}) = G_m(\alpha) \quad \text{for } r \neq s,\ r,\ s \in \{1,2,\dots,m\}, \tag{10}$$

where F_m and G_m are some C^∞ functions on $\mathbf{R}^+$.

Next, let $n = km$ where k is an integer larger than one and consider the Markov mapping $y = f_k(x)$ defined by

$$y = \left(\frac{x_1}{k}, \dots, \frac{x_1}{k}, \frac{x_2}{k}, \dots, \frac{x_2}{k}, \dots, \frac{x_m}{k}, \dots, \frac{x_m}{k}\right),$$

each component being repeated k times. From (4),

$$f_{k*}X_i = (1/k)\left(Y_{(i-1)k+1} + \cdots + Y_{ik}\right)$$

and, by (5) and (7),

$$g_{ij}^{(m)}(x) = \frac{1}{k^2}\sum g_{rs}^{(n)}(y), \tag{11}$$

where the summation is for $(i-1)k+1 \leqslant r \leqslant ik$, $(j-1)k+1 \leqslant s \leqslant jk$. Now if $x = \bar{x} = (\alpha/m, \ldots, \alpha/m)$, then $y = \bar{y} = (\alpha/n, \ldots, \alpha/n)$. Thus from (9)–(11),

$$F_m(\alpha) = \frac{1}{k}F_n(\alpha) + \frac{k-1}{k}G_n(\alpha) \tag{12}$$

and

$$G_m(\alpha) = G_n(\alpha), \tag{13}$$

where $n = km$.

Since m and k can be chosen arbitrarily, (13) implies that $G_m(\alpha) = A(\alpha)$, where A is some C^∞ function. Moreover, (12) can be written

$$\left(F_m(\alpha) - A(\alpha)\right)/m = \left(F_{mk}(\alpha) - A(\alpha)\right)/mk,$$

which implies that the left side is equal to some C^∞ function $B(\alpha)$, independent of m. Consequently, at points $\bar{x}$ on the center line,

$$g_{ij}^{(m)}(\bar{x}) = A(\alpha) + mB(\alpha)\delta_{ij}, \tag{14}$$

where $\alpha = |\bar{x}|$.

Next, let x be a point in $\mathbf{R}_m^+$ of the form

$$x = \left(\frac{\alpha k_1}{n}, \frac{\alpha k_2}{n}, \ldots, \frac{\alpha k_m}{n}\right) \tag{15}$$

where $\sum k_i = n$, all k_i are positive integers, and $\alpha = |x| > 0$. Consider the partition of $\{1, 2, \ldots, n\}$ into the sets

$$A_i = \{1, 2, \ldots, k_1\}, \qquad A_2 = \{k_1 + 1, \ldots, k_2\}, \quad \text{etc.}$$

with the probability vectors $Q^{(i)} = (q_{i1}, \ldots, q_{in})$ defined by

$$q_{ij} = \begin{cases} \dfrac{1}{k_i} & \text{if } \sum_{i=1}^{i-1} k_l < j \leqslant \sum_{l=1}^{i} k_l, \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

The corresponding Markov mapping f maps x to $\bar{y} = (\alpha/n, \alpha/n, \ldots, \alpha/n)$. Moreover, by (4),

$$f_* X_i = \frac{1}{k_i}\left(Y_{k_1 + \cdots + k_{i-1}+1} + \cdots + Y_{k_1 + \cdots + k_i}\right). \tag{17}$$

Consequently, by (5), (7), (14), and (17)

$$g_{ij}^{(m)}(x) = A(\alpha) \quad \text{for } i \neq j$$

and

$$g_{ii}^{(m)}(x) = \frac{1}{k_i^2}\left[k_i^2 A(\alpha) + k_i n B(\alpha)\right].$$

Hence, for any x of the form (15),

$$g_{ij}^{(m)}(x) = A(\alpha) + \delta_{ij}\alpha B(\alpha)/x_i, \tag{18}$$

since $x_i = \alpha k_i/n$.

Finally, any $x \in \mathbf{R}_m^+$ can be approximated arbitrarily well by an x of the form (15) with $\alpha = |x|$. Since $g_{ij}^{(m)}$ is a C^∞ function, (18) must hold for all x. Thus (6) holds.

The sign conditions on A and B follow from the positive-definiteness of inner products. Let X be any vector in the tangent space M_x and let $X = \sum a_i X_i$. Then by (6) or (18),

$$\langle X, X\rangle_m = A(|x|)\left(\sum a_i\right)^2 + |x|B(|x|)\sum\left(\frac{a_i^2}{x_i}\right).$$

If $\sum a_i = 0$ we see that $B(|x|) > 0$ is a necessary condition for positive definiteness. Also, if $a_i = x_i$, then

$$\langle X, X\rangle_m = |x|^2\left[A(|x|) + B(|x|)\right],$$

from which it follows that $A(|x|) + B(|x|) > 0$ is a necessary condition.

To prove the converse, observe first that by Cauchy's inequality,

$$|x|\sum\left(\frac{a_i^2}{x_i}\right) = \left(\sum x_i\right)\left(\sum\frac{a_i^2}{x_i}\right) \geqslant \left(\sum a_i\right)^2.$$

Thus if $B(|x|) > 0$ and $A(|x|) + B(|x|) > 0$

$$A(|x|)\left(\sum a_i\right)^2 + |x|B(|x|)\sum\left(\frac{a_i^2}{x_i}\right) \geqslant A(|x|)\left(\sum a_i\right)^2 + B(|x|)\left(\sum a_i\right)^2 \geqslant 0.$$

It is straightforward to show that this expression equals zero if and only if all $a_i = 0$. Thus (6) defines an inner product.

To check the isometry condition use (4) and (6) to get

$$\langle f_*X_i, f_*X_j\rangle_n(y) = \sum_k\sum_l q_{ik}q_{jl}\left[A(|y|) + \frac{\delta_{kl}|y|B(|y|)}{y_k}\right].$$

For Markov mappings $y = f(x)$ we have $|y| = |x|$ and $\sum_k q_{ik} = 1$. Thus

$$\langle f_*X_i, f_*X_j\rangle_n(y) = A(|x|) + |x|B(|x|)\sum_k\frac{q_{ik}q_{jk}}{y_k}.$$

However by the definition of Markov mappings $q_{ik}q_{jk} = 0$ if $i \neq j$. Also

$$q_{ik} = \begin{cases} y_k/x_i & \text{if } k \in A_i, \\ 0 & \text{if } k \notin A_i, \end{cases}$$

since, as pointed out earlier each sum (1) has only one nonzero term. Hence

$$\sum_k\frac{q_{ik}q_{jk}}{y_k} = \delta_{ij}\sum_{k\in A_i}\frac{q_{ik}}{x_i} = \frac{\delta_{ij}}{x_i},$$

since $\sum_k q_{ik} = 1$. Thus

$$\langle f_*X_i, f_*X_j\rangle_n(y) = \langle X_i, X_j\rangle_m(x),$$

as claimed.

It is worth noting that the method of proof used here, and particularly the use of x defined by (15) and the mapping by (16), is reminiscent of Khinchin's [**8**] characterization theorem for entropy.

5. Concluding remarks. If we take $A(|x|) = 0$ and $B(|x|) = 1$ in (6), we obtain the Riemannian metric of Čencov [**4**] and Shahshahani [**9**]. For vectors $X = \sum a_i X_i$ and $W = \sum b_i X_i$ in M_x,

$$\langle X, W\rangle_m(x) = A(|x|)\left(\sum a_i\right)\left(\sum b_i\right) + |x| B(|x|) \sum \left(\frac{a_i b_i}{x_i}\right).$$

If X is tangent to a simplex $|x| =$ constant, then $\sum a_i = 0$ and the choice of A is immaterial. Since Čencov [**4**] dealt with vectors in the tangent space to the manifold S_{m-1} the term involving $A(|x|)$ is absent from his work.

Even in $\mathbf{R}_m^+$, the choice $A(|x|) = 0$, $B(|x|) = 1$ leads to interesting geometric interpretations. This choice is employed for the remainder of this paper. Some questions related to information theory have been investigated in a separate paper [**3**]; here we indicate some connections with probability and statistics.

For $x \in S_{m-1}$, let $N = \sum x_i X_i$ $U = \sum u_i x_i X_i$, $V = \sum v_i x_i X_i$. Then, in this inner product, N is a unit vector which is normal to the simplex S_{m-1} in each tangent space M_x, $x \in S_{m-1}$. In addition,

$$\langle N, U\rangle_m(x) = \sum u_i x_i, \qquad \langle N, V\rangle_m(x) = \sum v_i x_i.$$

Thus, if $x_1, \ldots, x_m$ are interpreted as probabilities, the normal component of U appears as the expected value of a random variable which takes the values $u_1, \ldots, u_m$. Moreover, if U' is the projection of U on S_{m-1},

$$U' = U - \langle N, U\rangle_m(x) N$$

and V' is similarly the projection of V, then

$$\langle U', V'\rangle_m(x) = \sum u_i v_i x_i - \left(\sum u_i x_i\right)\left(\sum v_i x_i\right).$$

This is the covariance of the associated random variables. Thus there is a natural association between random variables and the special class of vector fields of the form $\sum u_i x_i X_i$ in which elementary geometric and probabilistic quantities coincide. Note that we need to have S_{m-1} embedded in a larger space to get the notion of normal component, which is connected to expectation.

REFERENCES

1. J. Aczél and Z. Daróczy, *On measures of information and their characterizations*, Academic Press, New York, 1975.

2. S.-I. Amari, *Differential geometry of curved exponential families–curvatures and information loss*, Ann. Statist. **10** (1982), 357–385.

3. L. L. Campbell, *The relation between information theory and the differential geometry approach to statistics*, Inform. Sci. **35** (1985), 199–210.

4. N. N. Čencov, *Statistical decision rules and optimal inference*, Transl. Math. Monographs, vol. 53, Amer. Math. Soc., Providence, R. I., 1981.

5. D. C. Hamilton, D. G. Watts, and D. M. Bates, *Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions*, Ann. Statist. **10** (1982), 386–393.

6. N. J. Hicks, *Differential geometry*, Van Nostrand, Princeton, N.J., 1965.

7. R. S. Ingarden, Y. Sato, K. Sugawa, and M. Kawaguchi, *Information thermodynamics and differential geometry*, Tensor (N. S.) **33** (1979), 347–353.

8. A. I. Khinchin, *Mathematical foundations of information theory*, Dover, New York, 1957.

9. S. Shahshahani, *A new mathematical framework for the study of linkage and selection*, Mem. Amer. Math. Soc. **17** No. 211 (1979).

DEPARTMENT OF MATHEMATICS AND STATISTICS, QUEEN'S UNIVERSITY, KINGSTON, ONTARIO K7L 3N6, CANADA