A Modification on Rao's Method to Evaluate the Distance for Geometric Distributions

Udie Sabrie
Basic Education college
University of Babylon

Jinan Hamza Farhood
College of Education
University of Babylon

Abstract

The concepts of distance between two distributions are fundamental in problems of statistical inference and in practical applications to study affinities among a given set of populations. (Rao, 1945) proposed a method, based on Fisher's information matrix, for measuring distance between distributions of a parametric family satisfying certain regularity conditions. He defined a metric tensor field via the Fisher information metric over the parameter space, which gives rise to a Riemannian metric. Then the geodesic distance induced by the metric is a measure of dissimilarity between two probability distributions, known as the Rao distance. In this paper, Rao's method is applied to obtain the distance between two geometric distributions. Some other properties are discussed too, see(Reverter and Oller, 2003).

الخلاصة

إن مفاهيم المسافة بين توزيعين هو أساسي في مشاكلِ الإستدلالِ الإحصائيِ وفي التطبيقاتِ العملية لدراسة الصلات بين مجموعة معطاة من المجتمعات.

اقترح راو طريقة معتمده على مصفوفة معلومات فيشر، لقياس المسافة بين توزيعات العائلة المعلمية التي تحقق شروط قياسية معينة وعرف الحقل المتري من خلال معلومات فيشر على الفضاء المعلمي. إن المسافة الجيودسية الناتجة من الفضاء المتري هي مقياس الاختلاف بين توزيعين احتماليين تعرف بمسافة راو.

في هذا البحث، طريقة راو طبقت لاستخراج المسافة بين توزيعين هندسيينِ. كذلك نوقشت بعض الخواص الأخرى.

Introduction

We introduce some basic ideas and important concepts of differential geometry such as tensor, first fundamental form and geodesics and introduce some interesting connections between statistics and differential geometry which includes Rao distance based on Riemannian metric whose elements are the entries in the Fisher information matrix. The set of distributions forms a manifold. The distance between two distributions is the distance along the geodesic that connects the two distributions.

The question of introducing a distance between different statistical populations has been considered by various authors. If we assume that all the information for constructing such a distance is contained in the probability density function of a random vector, supposedly existing and restricted to each population, it will not be generally satisfactory to characterize each population by their mean value of the random vector X, since the latter does not determine uniquely the probability density function associated with each population.

A reasonable alternative would be to allow that the probability density function of the random vector X, in any of the populations studied, to belong to a certain parameteric family, $p(./\theta)$.

It is also reasonable to require that the proposed distance on the parametric space Ω, possesses the property of being invariant under any admissible transformation of the parameters, since the latter does not affect the probability density function $p(./\theta)$ of the random vector X . In addition, the distance has to be invariant for admissible transformations of the random vector X, since it must be independent of the method by which the measurements are attained.

The method proposed by (Rao, 1945) and studied later by (Atkinson and Mitchell, 1981) and (Burbea and Rao, 1982), allows us to define a distance on the parametric space Ω with the above mentioned characteristics. If the parametric family satisfies certain regularity conditions, the fisher information matrix defines a covariant symmetric tensor field of the second order on the parametric space Ω .

$$g_{ij} = E\left(\frac{\partial \ln \mathrm{P}(X/\theta)}{\partial \theta^i}\frac{\partial \ln \mathrm{P}(X/\theta)}{\partial \theta^j}\right) \qquad (i, j = 1,...n) \qquad (1)$$

And it may also be taken as a metric tensor field on Ω , thereby rendering Ω as a Riemannian manifold. The Rao distance between two points θ_A and θ_B of Ω is then defined as their geodesic distance with respect to the metric induced by the equation (1), for more detail see (Gamero, Pichardo and Munoz, 2002) and (Amari, 2001).

While (Oller and Cuadras, 1985) applied Rao's Method to obtain a distance between two negative multinomial distributions and (Rios; Villarroya and Oller, 1992) applied Rao's Method to obtain a distance between multivariate linear models and their applications to the classification of response curves, Computational Statistic and Data Analysis. The Statistical Tests for the Inverse Gaussian Distribution based on Rao Distance is introduced by (Villarroya and Oller, 1993).

In this paper, Rao's method is applied to obtain the distance between two geometric distributions (Rao's Distance for Geometric Distributions).We also discuss some other differential geometric properties of these distributions.

1 Tensor (Jasim, 2006 and Lipschutz, 1969)

An *n*th- rank tensor in *m*- dimensional space is a mathematical object that has *n* indices and m^n components and obeys certain transformation rules. Each index of a tensor ranges over the number of dimensions of space. However, the dimension of the space is largely irrelevant in most tensor equations (with the notable exception of the contracted kronecker delta). Tensors are generalizations of scalars (that have no indices), vectors (that have exactly one index), and matrices (that have exactly two indices) to an arbitrary number of indices.

The notation for a tensor is similar to that of a matrix (i.e., $A = (a_{ij})$), except that a tensor $a_{ijk...}, a^{ijk...}, a_i^{jk}$, etc., may have an arbitrary number of indices where *i*, *j*, *k*...=1, 2, ..., *m*. In addition, a tensor with rank *r*+s may be of mixed type (*r*, s), consisting of *r* so-called "contravariant" (upper) indices and s "covariant" (lower) indices. Note that the positions of the slots in which contravariant and covariant indices are placed are significant so, for example, $a_{\mu\nu}^{\lambda}$ is distinct from $a_{\mu}^{\nu\lambda}$.

While the distinction between covariant and contravariant indices must be made for general tensor, the two are equivalent for tensors in three-dimensional Euclidean space, and such tensors are known as cartesian tensors.

The metric tensor is a tensor of rank 2 that is used to measure distance between any two points in a given space.

2 The First Fundamental Form (Gardner, 2000 and Jasim, 2006)

Suppose *M* is a surface determined by $\vec{X}(u,v) \subset E^3$ and suppose $\vec{\alpha}(t)$ is a curve on *M*, the variable *t* is called the parameter of the curve, $t \in [a,b]$ for $a,b \in R$. Then we can write $\vec{\alpha}(t) = \vec{X}(u(t), v(t))$ ((*u* (*t*), *v* (*t*)) is a curve in R^2 whose image under $\vec{X}$ is $\vec{\alpha}$). Then

$$\vec{\alpha}'(t)=\frac{\partial\vec{X}}{\partial u}\frac{du}{dt}+\frac{\partial\vec{X}}{\partial v}\frac{dv}{dt}=u'\vec{X}_1+v'\vec{X}_2 \tag{2}$$

If $s(t)$ represents the arc length along $\vec{\alpha}$ (with $s(a)=0$) then

$$s(t)=\int_a^t\left\|\vec{\alpha}'(r)\right\|dr \tag{3}$$

Where r is a real variable.
And

$$\frac{ds}{dt}=\left\|\vec{\alpha}'(t)\right\| \tag{4}$$

so

$$\left(\frac{ds}{dt}\right)^2=\left\|\vec{\alpha}'(t)\right\|^2=\vec{\alpha}'.\vec{\alpha}'=\left(u'\vec{X}_1+v'\vec{X}_2\right).\left(u'\vec{X}_1+v'\vec{X}_2\right)$$
$$=u'^2\left(\vec{X}_1.\vec{X}_1\right)+2u'v'\left(\vec{X}_1.\vec{X}_2\right)+v'^2\left(\vec{X}_2.\vec{X}_2\right). \tag{5}$$

Following Gauss' notation (briefly) we denote

$$E=\vec{X}_1.\vec{X}_1,\qquad F=\vec{X}_1.\vec{X}_2,\qquad G=\vec{X}_2.\vec{X}_2 \tag{6}$$

and have

$$\left(\frac{ds}{dt}\right)^2=E\left(\frac{du}{dt}\right)^2+2F\left(\frac{du}{dt}\frac{dv}{dt}\right)+G\left(\frac{dv}{dt}\right)^2 \tag{7}$$

or in differential notation

$$ds^2=E(du)^2+2F(dudv)+G(dv)^2 \tag{8}$$

Definition (Gardner, 2000; Jasim, 2006 and Lipschutz, 1969)

The matrix of the first fundamental form of a surface M determined by $\vec{X}(u,v)$ is

$$\begin{pmatrix}E & F\\ F & G\end{pmatrix}\equiv\begin{pmatrix}g_{11} & g_{12}\\ g_{21} & g_{22}\end{pmatrix}$$

Where E, F, G are as defined in formula (6). This matrix determines dot products of tangent vectors.

If $\vec{v}=a\vec{X}_1+b\vec{X}_2$ and $\vec{\omega}=c\vec{X}_1+d\vec{X}_2$ are vectors tangent to a surface M at a given point, then

$$\vec{v}.\vec{\omega}=\left(a\vec{X}_1+b\vec{X}_2\right).\left(c\vec{X}_1+d\vec{X}_2\right)$$
$$=Eac+F(ad+bc)+Gbd$$
$$=\begin{pmatrix}a & b\end{pmatrix}\begin{pmatrix}E & F\\ F & G\end{pmatrix}\begin{pmatrix}c\\ d\end{pmatrix}$$

Notation (Dodson and Scharcanski, 2003; Gardner, 2000 and Jasim, 2006)

We now replace the parameters u and v with u^1 and u^2 in formula (8) we then have

$$ds^2=g_{11}\left(du^1\right)^2+2g_{12}du^1du^2+g_{22}\left(du^2\right)^2=\sum_{i,j}g_{ij}du^idu^j. \tag{9}$$

Where the summation is taken over the set {1,2}. If $\vec{v}$ is a vector tangent to M

at a point $\vec{p}$ and $\vec{v} = (v^1, v^2)$ in the basis $\{\vec{X}_1, \vec{X}_2\}$ for the tangent plane at $\vec{p}$, then we have $\vec{v} = \sum_i v^i \vec{X}_i$.

3 Connections Between Statistics and Differential Geometry

3.1 Riemannian Metrics Based on Fisher Information (Chen, 1999 and Jasim, 2006)

We define the coefficients of the expected Fisher information matrix as equal to the coefficients of the first fundamental form (Riemannian metric) on the space of probabilities, known as Fisher information metric, which it is a metric tensor for a statistical differential manifold. It can be used to calculate the informational difference between measurement. It takes the following form where $f(x,\theta)$ be a class of probability densities.

$$g_{ij} = \int \frac{\partial \log f(x,\theta)}{\partial \theta_i} \frac{\partial \log f(x,\theta)}{\partial \theta_j} f(x,\theta) dx \tag{10}$$

which can be thought of intuitively as:

“The distance between two points on a statistical differential manifold is the amount of information between them, i.e. the informational difference between them”.

An equivalent form of the above equation is:

$$g_{ij} = -\int \frac{\partial^2 \log f(x,\theta)}{\partial \theta_i \, \partial \theta_j} f(x,\theta) dx = -E\left[\frac{\partial^2 \log f(x,\theta)}{\partial \theta_i \, \partial \theta_j}\right] \tag{11}$$

We define the coefficients of the first fundamental form as:

$$g_{11} = E = -E\left[\frac{\partial^2 \log f(x,\theta)}{\partial \theta_1^2}\right]$$

$$g_{12} = g_{21} = F = -E\left[\frac{\partial^2 \log f(x,\theta)}{\partial \theta_1 \, \partial \theta_2}\right]$$

$$g_{22} = G = -E\left[\frac{\partial^2 \log f(x,\theta)}{\partial \theta_2^2}\right]$$

where (θ_1, θ_2) are two parameters of the probability density functions. It is clear that *E, F* and *G* are functions of the parameters θ_1 and θ_2. The expectations apply to the whole sample space where the random variables are defined.

3.2 Riemannian Metrics and Geodesics (Gruber, 2003a and Gruber, 2003b).

We know that $f(x,\theta)$ be a class of probability densities, e.g. normal, binomial. Suppose that Θ be the set of all of the values of the parameter θ (Θ is a subset of R^n).

Assume that $\theta = (\theta_1, \theta_2, \ldots, \theta_n) \in \Theta$. Consider a quadratic differential metric in the form

$$ds^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij}(\theta) d\theta_i d\theta_j \tag{12}$$

Let $\theta = \theta(t)$, $t_1 \le t \le t_2$ represent a curve that joins the points in Θ $p = \theta(t_1)$, $Q = \theta(t_2)$. Let C represent the set of all such curves. The geodesic distance between p and Q will be defined by

$$d(p,Q)=\min_{c}\left|\int_{t_1}^{t_2}\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}g_{ij}(\theta)\theta_i\theta_j\right\}^{1/2}dt\right| \tag{13}$$

A curve for which this minimum is assumed will be called a geodesic curve. This curve may be found using the calculus of variations as a solution to the Euler Lagrange equations. For this problem they take the form of a system of ordinary differential equations

$$\sum_{i=1}^{n}g_{ij}\theta_i+\sum_{i=1}^{n}\sum_{j=1}^{n}\Gamma_{ijk}\theta_i\theta_j=0\,,\quad k=1,...,n \tag{14}$$

With boundary conditions of the form $\theta(t_1)=\alpha$, $\theta(t_2)=\beta$ where

$$\Gamma_{ijk}=\frac{1}{2}\left[\frac{\partial g_{jk}}{\partial\theta_i}+\frac{\partial g_{ki}}{\partial\theta_j}-\frac{\partial g_{ij}}{\partial\theta_k}\right] \tag{15}$$

3.3 Fisher Information Rao Distances between Probability Distributions (Gruber, 2003a and Gruber, 2003b)

We defined the elements of the Fisher information matrix as in form (11).
By substitution of the probability density into (11) the elements of the Fisher information (11) matrix and the Christoffel symbols (15) may be obtained. The system of equations (14) may be solved to get the geodesic curve and the geodesic distance may be obtained using (13). The geodesic distance may now be computed between two pdfs.

3.4 The Rao Distance for the Normal Distribution (Gruber, 2003b)

Three cases of distances to be considered for both the univariable cases include:
1- Distributions with the same standard deviations but different means.
2- Distributions with the same means but different standard deviations.
3- Distances between normal distributions where both the mean and standard deviation are different.

In the univariable distribution case we can see the following:
Consider two normal distributions with means μ_1 and μ_2 and common standard deviation σ .
The Rao distance is

$$d=\frac{|\mu_1-\mu_2|}{\sigma} \tag{16}$$

For two normal distributions with common mean μ and with standard deviations σ_1 and σ_2 . The Rao distance for this case is

$$d=\left|\sqrt{2}\log\left(\frac{\sigma_1}{\sigma_2}\right)\right|=\left|\frac{1}{\sqrt{2}}\log\left(\frac{\sigma_1^2}{\sigma_2^2}\right)\right| \tag{17}$$

For two normal distributions with respective means μ_1 and μ_2 and standard deviations σ_1 and σ_2

$$d=\sqrt{2}\left|\log\frac{1+\delta(1,2)}{1-\delta(1,2)}\right| \tag{18}$$

Observe that in (19) $\delta<1$
Where

$$\delta(1,2)=\sqrt{\frac{(\mu_1-\mu_2)^2+2(\sigma_1-\sigma_2)^2}{(\mu_1-\mu_2)^2+2(\sigma_1+\sigma_2)^2}} \tag{19}$$

Furthermore, the derivative of *d* with respect to *δ* is

$$d'=\sqrt{2}\left[\frac{2}{1-\delta^2}\right]>0 \ \text{for} \ |\delta|<1 \tag{20}$$

Thus, the Rao distance is an increasing function of *δ*. As the difference between the standard deviations remain constant *δ* and hence the Rao distance increases. For a constant difference between μ_1 and μ_2 the Rao distance becomes larger for increasing standard deviations when

$$|\mu_1-\mu_2|\le\sqrt{\frac{2(\sigma_1-\sigma_2)(\sigma_2+(1-\sigma_2)(\sigma_1-\sigma_2))}{\sigma_2}} \tag{21}$$

For the quantity under the square root sign in (21) to be positive either $\sigma_2>1$ and $\sigma_1>\sigma_2$ or $\sigma_2<1$ and $\sigma_1<\sigma_2$.

4. Rao's Distance for Geometric Distributions

We obtain the family of geometric distribution $\{P_r(\alpha/\theta)\}$

$P_r(\alpha/\theta)=\dfrac{\Gamma(|\alpha|+r)}{\alpha!\Gamma(r)}\theta^\alpha(1-\Theta)^r$. And if $r=1$, then

$$P_1(\alpha/\theta)=\frac{\Gamma(|\alpha|+1)}{\alpha!\Gamma(1)}\theta^\alpha(1-\Theta)^1 \tag{22}$$

In this case

$$f(z)=\frac{F'(z)}{F(z)}=(\log F(z))'=(1-z)^{-1} \qquad \frac{f'(z)}{f(z)}=(\log f(z))'=(1-z)^{-1} \tag{23}$$

$$g_{ij}(\theta)=\frac{r}{\theta^2}\left(\frac{1}{\theta^i}\delta_{ij}+\frac{1}{\theta^2}\right) \quad (i,j,r=1)$$

Therefore,

$$g_{11}(\theta)=\frac{1}{\theta^2}\left(\frac{1}{\theta^1}\delta_{11}+\frac{1}{\theta^2}\right) \tag{24}$$

Where θ^2 =1-Θ and δ_{11} is a Kronecker delta.

We proceed to calculate the Riemann-Christoffel tensor of the first kind (covariant curvature tensor) R_{1111} of the metric (24). This gives

$$R_{hijk}=\frac{r}{4\theta^i\theta^h\theta^4}\left(\delta_{ij}\delta_{kh}-\delta_{ik}\delta_{jh}+\frac{\theta^i(\delta_{kh}-\delta_{jh})+\theta^h(\delta_{ij}-\delta_{ik})}{\theta^2}\right).$$

And if $h,i,j,k,r=1$, then

$$R_{1111}=\frac{1}{4\theta^1\theta^1\theta^4}\left(\delta_{11}\delta_{11}-\delta_{11}\delta_{11}+\frac{\theta^1(\delta_{11}-\delta_{11})+\theta^1(\delta_{11}-\delta_{11})}{\theta^2}\right) \tag{25}$$

$R_{1111}=0$

The Riemannian curvature K is therefore

K=0 (26)

The inverse of the tensor-metric g_{11} is given by

$$g^{11}=\theta^2(\delta^{11}\theta^1-\theta^1\theta^1) \tag{27}$$

And defining $\delta_{111}=\delta_{11}\delta_{11}$, the Christoffel symbols of the second kind are

$$\Gamma^1_{11} = -\frac{\delta_{111}}{2\theta^1} + \frac{\delta_{11}+\delta_{11}}{2\theta^2} \tag{28}$$

Therefore, the differential equations of the geodesics may be written as

$$\frac{d^2\theta^1}{ds^2} - \frac{1}{2\theta^1}\left(\frac{d\theta^1}{ds}\right)^2 + \frac{1}{\theta^2}\frac{d\theta^1}{ds}\frac{d\theta^1}{ds} = 0 \tag{29}$$

$$\frac{d^2\theta^2}{ds^2} + \frac{1}{2\theta^1}\left(\frac{d\theta^1}{ds}\right)^2 - \frac{1}{\theta^2}\left(\frac{d\theta^2}{ds}\right) = 0 \tag{30}$$

$$\frac{1}{\theta^2}\left(\frac{1}{\theta^1}\left(\frac{d\theta^1}{ds}\right)^2 + \frac{1}{\theta^2}\left(\frac{d\theta^2}{ds}\right)^2\right) = 1, \tag{31}$$

we obtain from equations (30) and (31)that

$$\frac{1}{\theta^2}\frac{d^2\theta^2}{ds^2} - \frac{3}{2}\left(\frac{1}{\theta^2}\frac{d\theta^2}{ds}\right)^2 + \frac{1}{2} = 0\ . \tag{32}$$

The latter equation may be resolved via the transformation

$$u = \frac{1}{\theta^2}\frac{d\theta^2}{ds}\quad , \tag{33}$$

Yielding

$$u = -\tanh\left(\frac{s}{2} + C\right) \tag{34}$$

Where C is an integration constant.
From equations (29) and (34) we obtain

$$\frac{1}{\theta^1}\frac{d^2\theta^1}{ds^2} - \frac{1}{2}\left(\frac{1}{\theta^1}\frac{d\theta^1}{ds}\right)^2 + \frac{1}{\theta^1}\frac{d\theta^1}{ds}\tanh\left(\frac{s}{2} + C\right) = 0 \tag{35}$$

Equation (35) has the particular solution

$$\theta^1 = A_1 = \text{ const.} \tag{36}$$

More general solution can be obtained through the transformation

$$z^1 = \theta^1\left(\frac{d\theta^1}{ds}\right)^{-1} \tag{37}$$

Which reduces equation (35) to an independent set of linear differential equations of the form.

$$\frac{dz^1}{ds} - \tanh\left(\frac{s}{2} + C\right)z^1 = \frac{1}{2}\quad . \tag{38}$$

The general solution of equation (38) is

$$z^1 = \cosh^2\left(\frac{s}{2} + C\right)\left(\tanh\left(\frac{s}{2} + C\right) + B_1\right)\quad , \tag{39}$$

where the B_1 is constant of integration.
It follows, since

$$\ln\theta^1 = \int\frac{1}{z_1}ds\quad , \tag{40}$$

that

$$\theta^1 = A_1\left(B_1 + \tanh\left(\frac{s}{2} + C\right)\right)^2 \quad . \tag{41}$$

Equations (36) and (41) describe the geodesic lines, with the A_1 as integration constant.

$$d = 2\cosh^{-1}\left(\frac{1-\sqrt{a^1 b^1}}{\sqrt{a^2 b^2}}\right) \tag{42}$$

$$= 2\cosh^{-1}\left(\frac{1-\sqrt{a^1 b^1}}{\sqrt{(1-a^1)(1-b^1)}}\right)$$

5. Conclusions

(A) From equation (42), we conclude that the distance between two geometric distributions is not bounded. However, if we fix a^2 and b^2, easy to prove that

$$d \le 2\cosh^{-1}\left(\frac{1}{\sqrt{a^2 b^2}}\right) < \ln\left(\frac{4}{a^2 b^2}\right) \tag{43}$$

(B) It is also possible to establish a relationship between equation (42) and Bhattacharyya's distance (1946). This last distance can be obtained applying (Rao's, 1945) method to Geometric distribution .

If we consider the geometric distribution's family, with known index N and parameters θ^1, θ^2, where $\sum_{j=1}^{2}\theta^j = 1$,the Bhattacharyya's distance d_B between two points of the parametric space, $a^1 b^1$ is given by

$$d_B = 2\sqrt{N}\,\cos^{-1}\left(\sum_{j=1}^{2}\sqrt{a^j b^j}\right) \tag{44}$$

Hence, if we fix a^2 and b^2, both distances in equations (42) and (44), are decreasing functions of $\sqrt{a^1 b^1}$, and following (Shepard , 1962), we conclude that the preorders associated with equations (42) and (44) are equal.

(C) When d is very small, we have

$$\cosh\left(\frac{d}{2}\right) \cong 1 + \frac{d^2}{8}, \tag{45}$$

and thus

$$d \cong 2\sqrt{2\left(\frac{1-\sum_{j=1}^{2}\sqrt{a^j b^j}}{\sqrt{a^2 b^2}}\right)} \tag{46}$$

It follows from equations (44) and (46) that

$$d \cong 4\frac{\sin\left(\frac{d_B}{4\sqrt{N}}\right)}{\sqrt[4]{a^2 b^2}} \cong \sqrt{\frac{1}{N}}\,\frac{d_B}{\sqrt[4]{a^2 b^2}} \tag{47}$$

Thus, When a^2 and b^2 are fixed, d and d_B are nearly proportional, for small values of d .

(D) Let E sample space and consider two finite partitions of E, $\mathrm{P}_1 = \{A_1, A_2\}$ and $\mathrm{P}_2 = \{B_1, B_2\}$. Let P_2 be a refinement of P_1, and thus $B_2 \subset A_2$, and consider the parametric family of the geometric distributions in equation (1), with parameters $\theta^1 = \mathrm{P}_1(A_1)$ and $\theta^1 = \mathrm{P}_1(B_1)$ respectively. Let A and B two statistical populations with

$$\mathrm{P}_1(A_1) = a^1 \qquad \mathrm{P}_1(B_1) = p^1 \quad \text{(population A)}$$
$$\mathrm{P}_1(A_1) = b^1 \qquad \mathrm{P}_1(B_1) = q^1 \quad \text{(population B),} \tag{48}$$

then

$$d_1 = 2\cosh^{-1}\left(\frac{1-\sqrt{a^1 b^1}}{\sqrt{a^2 b^2}}\right) \le d_2 = 2\cosh^{-1}\left(\frac{1-\sqrt{p^1 q^1}}{\sqrt{p^2 q^2}}\right) \tag{49}$$

Actually it is sufficient to consider first the case

$$A_i = B_{i_1} \cup B_{i_{s_i}} \qquad A_i \in \mathrm{P}_1,\ B_{i_j} \in \mathrm{P}_2 \quad (i = 1,2) \tag{50}$$

and

$$a^i = p^{i_1} + p^{i_{s_i}} \qquad b^i = q^{i_1} + q^{i_{s_i}} \quad (i = 1,2) \tag{51}$$

Since P_2 is a refinement of P_1

$$\sqrt{a^i b^i} \ge \sum_{j=1}^{s_i} \sqrt{p^{i_j} q^{i_j}} \quad (i = 1,2) \tag{52}$$

because

$$\left(\sum_{j=1}^{s_i} \sqrt{p^{i_j} q^{i_j}}\right)^2 = a^i b^i - \sum_{j<1}\left(p^{i_j} q^{i_1} + p^{i_1} q^{i_j}\right) + 2\sum_{j<1}\sqrt{p^{i_j} q^{i_j} p^{i_1} q^{i_1}} \le a^i b^i \ (i = 1,2) \tag{53}$$

On the other hand, if

$$A_2 = B_{j_1} \cup B_2 \qquad \mathrm{A}_2 \in \mathrm{P}_1, B_s \in \mathrm{P}_2 \quad \text{then} \tag{54}$$

$$\sqrt{a^2 b^2} \ge \sqrt{p^{j1} q^{j1}} + \sqrt{p^2 q^2} \quad ,$$

but since

$$1-\sum_{j=1}^{2}\sqrt{a^j b^j} \ge 0 \tag{55}$$

We find that

$$\frac{1-\sqrt{a^1 b^1}}{\sqrt{a^2 b^2}} \le \frac{1-\sqrt{a^1 b^1} - \sqrt{p^{j1} q^{j1}}}{\sqrt{p^2 q^2}} \quad . \tag{56}$$

However, since

$$\sqrt{a^{11} b^{11}} + \sqrt{p^{11} q^{11}} \ge \sqrt{p^1 q^1} \tag{57}$$

Equation (49) follows at once, that is, $d_1 \le d_2$: refining the partition associated with geometric distribution, the distance in equation (32) increases .

References

Amari, S.I. (2001), Information Geometry on Hierarchy of Probability Distributions, Ieee Transactions on Information Theory, Volume 47, No. 5, pp. 1701-1711.

Atkinson, C. and Mitchell, A.F.S. (1981), Rao's Distance Measure, Sankhya, The Indian Journal of Statistics, A 43, pp. 345-365.

Bhattacharyya, A. (1946), On A measure of Divergence Between two Multinomial Populations, Sankhya, 7, pp. 401-406.

Burbea, J. and Rao, C.R. (1982), Entropy Differential Metric, Distance and Divergence Measure in Probability Spaces: A unified Approach, Journal of Multivariate Analysis, 12, pp. 575-596.

Chen, W.W.S. (1999), On Computing Gaussian Curvature of Some Well Known Distributions. http://www.irs.gov/pub/irs-soi/gauscurv.pdf

Dodson, C.T.J. and Scharcanski, J. (2003), Information Geometric Similarity Measurement for Near-Random Stochastic Processes, Ieee Transactions on Systems, Man, and Cybernetics-Part A : Systems and Humans, Volume 33, No. 4, pp. 435-440.

Gamero, M.D.J., Pichardo, J.M.M. and Munoz, J. (2002), Rao Distance as A measure of Influence in the Multivariate Linear Model, Journal of Applied Statistics, Volume 29, No, 6, pp. 841-854.

Gardner, B. (2000), Differential Geometry (and Relativity) Classnotes. http://www.etsu.edu/math/gardner/5310/notes.htm-10k-

Gruber, M.H.J. (2003**a**), Some Interesting Connections between Statistics and Differential Geometry. http://www.rit.edu/~mjgsma/maatalk/geometrystatistics.html

Gruber, M.H.J. (2003**b**), Using Differential Geometry to Study Robustness in Bayesian Statistics for Shrinkage Estimators. http://www.rit.edu/~mjgsma/slides.pdf

Jasim, W.A. (2006), Application of Central Limit Theorem on Probability Distributions by Using Differential Geometry, Master Thesis, Babylon University, Ministry of Higher Education and scientific Research, Iraq.

Lipschutz, M.M. (1969), Schaum's Outline Series: Theory and Problems of Differential Geometry, Mc Graw-Hill Book Company, New York.

Oller, J.M. and Cuadras, C.M. (1985), Rao's Distance for Negative Multinomial Distribution, Sankhya, The Indian Journal of Statistics, Series A, Volume 47, Part 1, pp. 75-83.

Rao, C.R. (1945), Information and the Accuracy Attainable in the Estimation of Statistical Parameters, Bull. Calcutta Math. Soc., 27, pp. 81-91.

Reverter, F. and Oller, J.M. (2003), Computing the Rao Distance for Gamma Distributions, Volume 157, Part 1, pp. 155-167.

Rios, M., Villarroya, A. and Oller, J.M.(1992), Rao Distance Between Multivariate Linear Models and their Applications to the classification of Response Curves, Computational Statistic and Data Analysis, 13, pp. 431-441.

Shepard, R.W. (1962), The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function. Psychometrika, 27, pp. 219-146.

Villarroya, A. and Oller, J. (1993), Statistical Tests for the Inverse Gaussian Distribution Based on Rao Distance, Sankhya, The Indian Journal of Statistics, Series A, Volume 55, Part 1, pp. 80-103.