ELSEVIER Journal of Geometry and Physics 17 (1995) 185–189

JOURNAL OF GEOMETRY AND PHYSICS

Brief communication

On the parameter space derived from the joint probability density functions and the property of its scalar curvature

Masashi Yamada [a], Manabu Miyata [a], Tomoaki Kawaguchi [b,*]

[a] *Department of Information Sciences, Saitama College, Kazo, 347 Japan*

[b] *Institute of Information Sciences and Electronics, University of Tsukuba, Tsukuba, 305 Japan*

Received 22 March 1994; revised 7 October 1994

Abstract

Extended parameter spaces were introduced by Kawaguchi et al. (1992) to define the geometrical distance between two probability distributions having different function forms. A statistical interpretation of the extended parameter spaces was introduced. In this paper, the property of the scalar curvature of extended parameter spaces is given.

Keywords: Probability density functions; Parameter space; Curvature

1991 MSC: 53B99, 53B50, 94A75, 62A99

1. Introduction

Traditionally, the parameter space in statistics has not been given a geometrical structure. It only means the set of parameters that specifies probability distribution. However, in 1945 Rao [1] introduced Riemannian structures into the statistical parameter space. Since then, many researchers have made geometrical approaches to the parameter space. In general, it is difficult to solve geodesic equations given by nonlinear differential equations derived from parameter spaces. Atkinson and Mitchel [2] solved geodesic equations for some distribution. Amari [3] has researched some relations between geometrical structures and statistical properties of the spaces. Burbea and Rao [4,5] have studied some relations of statistical divergences and statistical parameter spaces.

* Corresponding author.

SSDI 0393-0440(94)00045-X

In these works, statistical parameter spaces are defined for the probability distributions belonging to the same family of probability distribution. When probability density functions have different function forms from one another, it is impossible to treat them in the same space. This defect has restricted wider possibilities of applications of statistical parameter spaces. Extended parameter spaces were introduced by Kawaguchi et al. [6], to define the geometrical distance between two probability distributions having different function forms. We introduced a metric in the parameter space in a way that enables us to retain the form of Fisher's information matrix calculated from each marginal probability density function. In this paper we clarify the statistical significance of these geometrical concepts.

2. Extended parameter spaces

Let $\boldsymbol{x} = \left(x^1, x^2, \dots, x^m\right)$ be random variables, $\boldsymbol{\theta} = \left(\theta^1, \theta^2, \dots, \theta^n\right)$ real continuous parameters and $h(\boldsymbol{x}; \boldsymbol{\theta})$ a joint probability density function of random variables $\boldsymbol{x}$. Each probability density function of random variable x^i may be different from one another. Fisher's information matrix of $h(\boldsymbol{x}; \boldsymbol{\theta})$ is as follows:

$$I(\boldsymbol{\theta}) = -\begin{pmatrix} \int_D h(\boldsymbol{x}; \boldsymbol{\theta}) \dfrac{\partial^2 \log h(\boldsymbol{x}; \boldsymbol{\theta})}{\partial\left(\theta^1\right)^2} \mathrm{d}\boldsymbol{x} & \dots & \int_D h(\boldsymbol{x}; \boldsymbol{\theta}) \dfrac{\partial^2 \log h(\boldsymbol{x}; \boldsymbol{\theta})}{\partial\theta^1 \partial\theta^n} \mathrm{d}\boldsymbol{x} \\ \vdots & \ddots & \vdots \\ \int_D \dfrac{h(\boldsymbol{x}; \boldsymbol{\theta}) \partial^2 \log h(\boldsymbol{x}; \boldsymbol{\theta})}{\partial\theta^n \partial\theta^1} \mathrm{d}\boldsymbol{x} & \dots & \int_D h(\boldsymbol{x}; \boldsymbol{\theta}) \dfrac{\partial^2 \log h(\boldsymbol{x}; \boldsymbol{\theta})}{\partial(\theta^n)^2} \mathrm{d}\boldsymbol{x} \end{pmatrix}, \tag{2.1}$$

where D means the domain of integration of $f(\boldsymbol{x}; \boldsymbol{\theta})$ with respect to $\boldsymbol{x}$. It is well known that the quantity (2.1) is a Riemannian metric tensor [1]. Therefore, the statistical parameter space of a joint density function constitutes a Riemannian space.

Suppose that random variables of $h(\boldsymbol{x}; \boldsymbol{\theta})$ are independent of one another. Then this density function is denoted by

$$h(\boldsymbol{x}; \boldsymbol{\theta}) = f_{(1)}\left(x^1; \boldsymbol{\theta}_{(1)}\right) f_{(2)}\left(x^2; \boldsymbol{\theta}_{(2)}\right) \cdots f_{(m)}\left(x^m; \boldsymbol{\theta}_{(m)}\right), \tag{2.2}$$

where $f_{(i)}\left(x^i; \boldsymbol{\theta}_{(i)}\right)$ $(i = 1, 2, \dots, m)$ are marginal density functions of random variables x^i $(i = 1, 2, \dots, m)$ and $\boldsymbol{\theta}$ is the set made of $\boldsymbol{\theta}_{(1)}, \boldsymbol{\theta}_{(2)}, \dots, \boldsymbol{\theta}_{(m)}$. Thus the metric tensor (2.1) reduces to

$$g_{ij} = \begin{pmatrix} g_{(1)}\left(\boldsymbol{\theta}_{(1)}\right) & 0 & \dots & 0 \\ 0 & g_{(2)}\left(\boldsymbol{\theta}_{(2)}\right) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & g_{(m)}\left(\boldsymbol{\theta}_{(m)}\right) \end{pmatrix}, \tag{2.3}$$

where $g_{(i)}(\boldsymbol{\theta}_{(i)})$, $(i = 1, 2, \dots, m)$ is Fisher's information matrix derived from each marginal density function $f_{(i)}(x^i; \boldsymbol{\theta}_{(i)})$ $(i = 1, 2, \dots, m)$, namely the metric tensor in each marginal density functions, respectively.

The forms of these functions $f_{(i)}(x^i; \boldsymbol{\theta}_{(i)})$ $(i = 1, 2, \dots, m)$ and the number of parameters are not always the same. The metric tensor (2.3) enables an investigation of a statistical parameter space of probability density functions whose function forms are different from one another.

In the above discussion, it has been shown that the form of a metric tensor reduces to (2.3) when the random variables of a joint density function are independent of one another. Such a joint density function can be decomposed as the product of marginal density functions. Making use of this fact conversely, it is possible to make such a joint density from the product of some probability density functions. Accordingly, we can obtain the metric tensor derived from the product of probability density functions whose function forms are different from one another. Extended parameter spaces are interpreted as Riemannian space with a metric tensor made by this way.

Each specified point in the extended parameter spaces can correspond to a probability distribution. We can consider the geodesic distance between these points as the distance between two probability distributions whose forms of density function are different from each other.

3. Scalar curvature of extended parameter spaces

There is an important geometrical property between the scalar curvature of the extended parameter space and that of the parameter space derived from each marginal density function.

Theorem. *Let $f_{(i)}(x^i; \boldsymbol{\theta}_{(i)})$ $(i = 1, 2, \dots, m)$ be probability density functions of each independent random variable x^i $(i = 1, 2, \dots, m)$ and $h(\boldsymbol{x}; \boldsymbol{\theta})$ the product of these functions, where $\boldsymbol{\theta}_{(i)}(i = 1, 2, \dots, m)$ belongs to an open subset of $\mathbb{R}^{n_i}$, $\boldsymbol{x}$ is $(x^1, x^2, \dots, x^m)$ and $\boldsymbol{\theta}$ is $(\theta^1, \theta^2, \dots, \theta^n)$. If $h(\boldsymbol{x}; \boldsymbol{\theta})$ has Fisher's information matrix, then the scalar curvature of the extended parameter space of $h(\boldsymbol{x}; \boldsymbol{\theta})$ is the summation of the scalar curvature of the parameter spaces, the subspaces of extended parameter space, derived from each probability density function.*

Proof. The metric tensor of the extended parameter space made of the product of m probability density functions $f_{(1)}(x^1; \boldsymbol{\theta}_{(1)})\, f_{(2)}(x^2; \boldsymbol{\theta}_{(2)}) \cdots f_{(m)}(x^m; \boldsymbol{\theta}_{(m)})$ is

$$g_{ij} = \begin{pmatrix} g_{(1)}(\boldsymbol{\theta}_{(1)}) & 0 & \dots & 0 \\ 0 & g_{(2)}(\boldsymbol{\theta}_{(2)}) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & g_{(m)}(\boldsymbol{\theta}_{(m)}) \end{pmatrix} \quad (m \le n),$$

where $g_{(i)}(\boldsymbol{\theta}_{(i)})$, $(i = 1, 2, \ldots, m)$ are Fisher's information matrix derived from each marginal density function $f_{(i)}(x^i; \boldsymbol{\theta}_{(i)})$ $(i = 1, 2, \ldots, m)$.

Arranging the vectors of parameters $\boldsymbol{\theta}_{(i)}$ by the order of index i, we obtain

$$\begin{aligned}&(\boldsymbol{\theta}_{(1)}, \boldsymbol{\theta}_{(2)}, \ldots, \boldsymbol{\theta}_{(m)})\\ &= \left(\theta^1_{(1)}, \theta^2_{(1)}, \ldots, \theta^{n_1}_{(1)}, \theta^1_{(2)}, \theta^2_{(2)}, \ldots, \theta^{n_2}_{(2)}, \ldots, \theta^1_{(m)}, \theta^2_{(m)}, \ldots, \theta^{n_m}_{(m)}\right).\end{aligned} \tag{3.1}$$

Now we use the following convention of indices.

The convention of index: $\alpha, \beta, \ldots, \omega$ and $\alpha_{(i)}, \beta_{(i)}, \ldots, \omega_{(i)}$ run from 1 to n $(= \sum_{j=1}^{m} n_j)$ and from $\sum_{j=1}^{i-1} n_j + 1$ to $\sum_{j=1}^{i-1} n_j + n_i$, respectively. The subindex i runs from 1 to m. The values of both indices α and $\alpha_{(i)}$ indicate the order of parameters in the right-hand side of (3.1).

By making use of the above convention of indices, the scalar curvature K is defined as

$$K = g^{\mu\lambda}\left(\frac{\partial \Gamma_{\mu}{}^{\nu}{}_{\lambda}}{\partial \theta^{\nu}} - \frac{\partial \Gamma_{\nu}{}^{\nu}{}_{\lambda}}{\partial \theta^{\mu}} + \Gamma_{\nu}{}^{\nu}{}_{\alpha}\Gamma_{\mu}{}^{\alpha}{}_{\lambda} - \Gamma_{\mu}{}^{\nu}{}_{\alpha}\Gamma_{\nu}{}^{\alpha}{}_{\lambda}\right), \tag{3.2}$$

where $\Gamma_{\nu}{}^{\kappa}{}_{\alpha}$ is Christoffel symbols of the second kind. $\Gamma_{\nu}{}^{\kappa}{}_{\alpha}$ can be rewritten by the other index system as

$$\Gamma_{\alpha_{(i)}}{}^{\beta_{(j)}}{}_{\gamma_{(k)}} = \frac{1}{2} g^{\beta_{(j)}\delta}\left(\frac{\partial g_{\gamma_{(k)}\delta}}{\partial \theta^{\alpha_{(i)}}} + \frac{\partial g_{\alpha_{(i)}\delta}}{\partial \theta^{\gamma_{(k)}}} - \frac{\partial g_{\alpha_{(i)}\gamma_{(k)}}}{\partial \theta^{\delta}}\right),$$

where $\alpha_{(i)}, \beta_{(j)}, \gamma_{(k)}$ run from 1 to n, because i, j and k are not fixed. If two of the three indices i, j and k are not equal, then $g^{\alpha_{(i)}\beta_{(k)}}$, $g_{\alpha_{(i)}\beta_{(k)}}$ and $\partial g_{\alpha_{(i)}\delta}/\partial\theta^{\beta_{(k)}}$ vanish. Therefore, only when all of i, j and k are the same, $\Gamma_{\alpha_{(h)}}{}^{\beta_{(i)}}{}_{\gamma_{(k)}}$ does not vanish.

Thus (3.2) reduces to (3.3)

$$\begin{aligned}g^{\mu\lambda}K_{\nu\mu\lambda}{}^{\nu} &= g^{\mu_{(1)}\lambda_{(1)}}K_{\nu_{(1)}\mu_{(1)}\lambda_{(1)}}{}^{\nu_{(1)}} + g^{\mu_{(2)}\lambda_{(2)}}K_{\nu_{(2)}\mu_{(2)}\lambda_{(2)}}{}^{\nu_{(2)}}\\ &\quad + \cdots + g^{\mu_{(m)}\lambda_{(m)}}K_{\nu_{(m)}\mu_{(m)}\lambda_{(m)}}{}^{\nu_{(m)}}.\end{aligned} \tag{3.3}$$

Each term of the right-hand side in (3.3) is a scalar curvature of the parameter spaces derived from probability density functions $f_{(i)}(x^i; \boldsymbol{\theta}_{(i)})$ $(i = 1, 2, \ldots, m)$. Hence the scalar curvature of extended parameter space is the sum of the scalar curvatures of each of the subspaces. □

4. Conclusions

We have given the interpretation of extended parameter spaces, i.e., the extended parameter spaces are the statistical parameter spaces derived from a joint density function which consists of a product of density functions independent of one another.

In the conventional theory of statistical parameter spaces based on Fisher's information matrix, it is impossible to calculate the distance between two probability distributions whose density function forms are different from each other. The extended parameter spaces enable us to treat the distance between two probability density functions having different function forms.

As is well known, a scalar curvature is an important geometrical property. It is proved that the scalar curvature in the extended parameter space is the summation of the curvatures of each one of its subspaces. This shows an interesting property of extended parameter spaces.

References

[1] C.R. Rao, Information and accuracy attainable in the estimation of statistical parameters, Bull. Calcutta Math. Soc. 37 (1945) 81–91.

[2] C. Atkinson and A.F.S. Mitchell, Rao's distance measure, Sankhya 43, Series A, Part 3 (1981) 345–365.

[3] S. Amari, *Differential-Geometrical Methods in Statistics* (Springer, Berlin, 1985).

[4] J. Burbea and C.R. Rao, Entropy differential metric, distance and divergence measures in probability spaces: A unified approach, J. Multivariate Anal. 12 (1982) 575–596.

[5] J. Burbea and C.R. Rao, On the convexity of same divergence measures based on entropy functions, IEEE Trans. Inform. Theory 28 (1987) 489–495.

[6] T. Kawaguchi, M. Yamada and M. Miyata, An extended parameter space and its applications to some probability density functions, Tensor, N. S. 51 (1992) 301–308.