



# Hommage to Chentsov's theorem

Akio Fujiwara<sup>1</sup>

Received: 26 September 2022 / Revised: 6 November 2022 / Accepted: 12 November 2022 /  
Published online: 21 November 2022  
© The Author(s) 2022

## Abstract

Chentsov's theorem, which characterises Markov invariant Riemannian metric and affine connections of manifolds of probability distributions on finite sample spaces, is undoubtedly a cornerstone of information geometry. This article aims at providing a comprehensible survey of Chentsov's theorem as well as its modest extensions to generic tensor fields and to parametric models comprising continuous probability densities on  $\mathbb{R}^k$ .

**Keywords** Chentsov's theorem · Markov invariance · Fisher metric ·  $\alpha$ -connections · Amari–Chentsov tensor · Information geometry

## 1 Introduction

For each natural number  $n$  satisfying  $n \geq 2$ , let

$$\mathcal{S}_{n-1} := \left\{ p : \Omega_n \rightarrow \mathbb{R}_{++} \mid \sum_{\omega \in \Omega_n} p(\omega) = 1 \right\}$$

be the manifold of probability distributions on a finite sample space  $\Omega_n = \{1, 2, \dots, n\}$ , where  $\mathbb{R}_{++}$  denotes the set of strictly positive real numbers. The manifold  $\mathcal{S}_{n-1}$  is sometimes called the  $(n - 1)$ -dimensional *probability simplex*. In what follows, we identify each point  $p \in \mathcal{S}_{n-1}$  with the numerical vector  $(p(1), p(2), \dots, p(n)) \in \mathbb{R}_{++}^n$ .

---

Communicated by Hiroshi Matsuzoe.

---

✉ Akio Fujiwara  
fujiwara@math.sci.osaka-u.ac.jp

<sup>1</sup> Department of Mathematics, Osaka University, Toyonaka, Osaka 560-0043, Japan

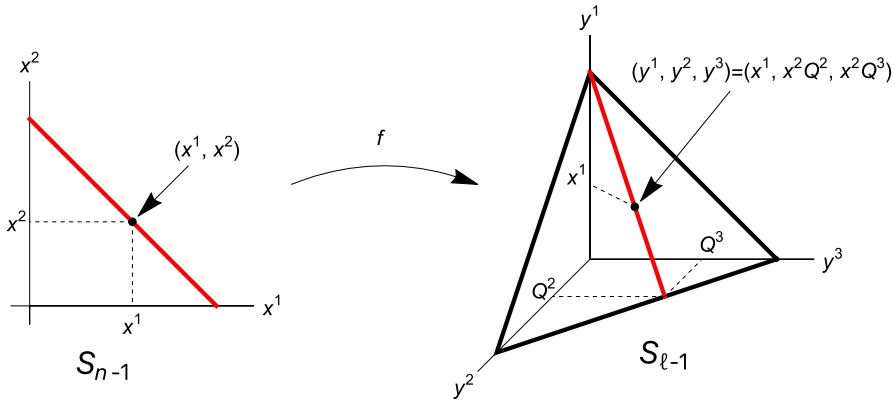


Fig. 1 A Markov embedding  $f : S_{n-1} \rightarrow S_{l-1}$  for  $n = 2$  and  $l = 3$ , where  $C_{(1)} = \{1\}$ ,  $C_{(2)} = \{2, 3\}$ , and  $Q_{(1)} = (1, 0, 0)$ ,  $Q_{(2)} = (0, Q^2, Q^3)$

In his seminal book [6], Chentsov characterised Riemannian metric  $g$  and affine connections  $\nabla$  on  $S_{n-1}$  that fulfil certain invariance property, now usually referred to as the Markov invariance. Given natural numbers  $n$  and  $l$  satisfying  $2 \leq n \leq l$ , let

$$\Omega_\ell = \bigsqcup_{i=1}^n C_{(i)} \tag{1}$$

be a direct sum decomposition of the index set  $\Omega_\ell = \{1, \dots, \ell\}$  into  $n$  mutually disjoint nonempty subsets  $C_{(1)}, \dots, C_{(n)}$ . A map

$$f : S_{n-1} \longrightarrow S_{l-1} : (x^1, \dots, x^n) \longmapsto (y^1, \dots, y^\ell)$$

is called a *Markov embedding*<sup>1</sup> associated with the partition (1) if it takes the form

$$y^j := \sum_{i=1}^n x^i Q_{(i)}^j \quad (j = 1, \dots, \ell), \tag{2}$$

where  $Q_{(i)} = (Q_{(i)}^1, Q_{(i)}^2, \dots, Q_{(i)}^\ell)$  is a probability distribution on  $\Omega_\ell$  whose support is  $C_{(i)}$  for each  $i$  ( $1 \leq i \leq n$ ). In other words, the image  $f(S_{n-1})$  is (the interior of) the convex hull of  $n$  extreme points  $\{Q_{(i)}\}_{1 \leq i \leq n}$  in  $S_{l-1}$ . A simple example of a Markov embedding is illustrated in Fig. 1, where  $n = 2$  and  $l = 3$ .

Since the image  $f(S_{n-1})$  of a Markov embedding  $f : S_{n-1} \rightarrow S_{l-1}$  is statistically isomorphic to the preimage  $S_{n-1}$  (due to the existence of a sufficient statistic), Chentsov claimed that the geometry of the submanifold  $f(S_{n-1})$  of  $S_{l-1}$  must be equivalent to the geometry of  $S_{n-1}$ .

Based on these observations, Chentsov introduced the notion of invariance/equivariance, now usually referred to as the *Markov invariance*, as follows. A series

<sup>1</sup> Chentsov called such an embedding a *congruent embedding* [6, Lemma 9.5].

$\{g^{[n]}\}_{n \in \mathbb{N}}$  of Riemannian metrics, each on  $\mathcal{S}_{n-1}$ , is said to be *invariant* [6, p. 157] if

$$g_p^{[n]}(X, Y) = g_{f(p)}^{[\ell]}(f_*X, f_*Y) \tag{3}$$

holds for all  $n, \ell \in \mathbb{N}$  satisfying  $2 \leq n \leq \ell$ , Markov embeddings  $f : \mathcal{S}_{n-1} \rightarrow \mathcal{S}_{\ell-1}$ , points  $p \in \mathcal{S}_{n-1}$ , and vector fields  $X, Y \in \Gamma(T\mathcal{S}_{n-1})$ , where  $f_*$  denotes the differential of  $f$ . Chentsov proved that, up to a constant factor, the only invariant metric satisfying (3) is the Fisher metric [6, Theorem 11.1]. For an accessible proof, see [5] (cf., [8]).

On the other hand, a series  $\{\nabla^{[n]}\}_{n \in \mathbb{N}}$  of affine connections, each on  $\mathcal{S}_{n-1}$ , is said to be *equivariant* [6, p. 62] if

$$f_* \left( \nabla_X^{[n]} Y \right)_p = \left( \nabla_{f_*X}^{[\ell]} f_*Y \right)_{f(p)} \tag{4}$$

holds for all  $n, \ell \in \mathbb{N}$  satisfying  $2 \leq n \leq \ell$ , Markov embeddings  $f : \mathcal{S}_{n-1} \rightarrow \mathcal{S}_{\ell-1}$ , points  $p \in \mathcal{S}_{n-1}$ , and vector fields  $X, Y \in \Gamma(T\mathcal{S}_{n-1})$ . Chentsov proved that the only equivariant affine connections satisfying (4) are the  $\alpha$ -connections [6, Theorem 12.2]. For the reader's convenience, we will give a proof, following the pattern of Chentsov's original argument, in Appendix. (An alternative proof based on a weaker condition (5) below is found in [8].)

Chentsov's theorem characterises all Markov invariant geometrical structures of the probability simplex  $\mathcal{S}_{n-1}$ , and thus is regarded as a cornerstone of information geometry. Nevertheless, it is natural to seek characterisations of Markov invariant tensor fields of generic-type and/or geometrical structures of probability spaces on infinite sample spaces, both of which being beyond the scope of Chentsov's theorem.

Motivated by these considerations, this article aims to give two variations of Chentsov's theorem. In Sect. 2, we extend the notion of Markov invariance to generic  $(r, s)$ -type tensor fields, and characterise all Markov invariant tensor fields on  $\mathcal{S}_{n-1}$ . This section also serves as a brief overview of the paper [7]. In Sect. 3, following Amari and Nagaoka's idea sketched out in [2], we demonstrate that the Fisher metric and the  $\alpha$ -connections are the only natural Markov invariant geometrical structures of parametric models comprising continuous probability densities on infinite sample spaces  $\mathbb{R}^k$ . Here, we employ only elementary calculus.

## 2 Markov invariant tensor fields of generic-types

Since the image  $f(\mathcal{S}_{n-1})$  of a Markov embedding  $f : \mathcal{S}_{n-1} \rightarrow \mathcal{S}_{\ell-1}$  is a submanifold of  $\mathcal{S}_{\ell-1}$ , its geometrical structure is canonically induced from the geometrical structure  $(g^{[\ell]}, \nabla^{[\ell]})$  of the ambient manifold  $\mathcal{S}_{\ell-1}$ . Specifically, the metric of  $f(\mathcal{S}_{n-1})$  is induced from the metric  $g^{[\ell]}$  by restricting it to the subspace  $T_{f(p)}f(\mathcal{S}_{n-1}) \subset T_{f(p)}\mathcal{S}_{\ell-1}$ , and the connection of  $f(\mathcal{S}_{n-1})$  is induced by projecting  $(\nabla_{f_*X}^{[\ell]} f_*Y)_{f(p)}$  onto the subspace  $T_{f(p)}f(\mathcal{S}_{n-1})$  with respect to  $g^{[\ell]}$ . Therefore, at first sight, the equivariance requirement (4) for a connection seems too strong, and one may instead define the Markov invariance for a connection in a weaker form as follows:

$$g_p^{[n]}(\nabla_X^{[n]}Y, Z) = g_{f(p)}^{[\ell]}(\nabla_{f_*X}^{[\ell]}f_*Y, f_*Z). \tag{5}$$

Nevertheless, it turns out that any sequence  $\{\nabla^{[n]}\}_n$  of affine connections satisfying (5) enjoys the property  $\nabla_{f_*X}^{[\ell]}f_*Y \in T_{f(p)}f(\mathcal{S}_{n-1})$ , and thus the requirements (4) and (5) are actually equivalent.

Since the sequence  $\bar{\nabla}^{[n]}$  of the Levi-Civita connections with respect to the Markov invariant Fisher metrics  $g^{[n]}$  automatically fulfils the requirement (5), the problem of characterising the Markov invariant connections is reduced to characterising the Markov invariant (0, 3)-type tensor fields

$$(X, Y, Z) \mapsto g((\nabla_X Y - \bar{\nabla}_X Y), Z).$$

Now, in a quite similar way to the derivation of the Fisher metric:

$$g_p(X, Y) := E_p[(X \log p)(Y \log p)],$$

where  $E_p[\cdot]$  denotes the expectation with respect to  $p$ , one can prove that, up to a constant factor, the only Markov invariant (0, 3)-type tensor field is given by

$$S_p(X, Y, Z) := E_p[(X \log p)(Y \log p)(Z \log p)], \tag{6}$$

which is usually referred to as the *Amari-Chentsov tensor*. In this way, the  $\alpha$ -connection  $\nabla^{(\alpha)}$  can also be defined by the formula

$$g(\nabla_X^{(\alpha)}Y, Z) := g(\bar{\nabla}_X Y, Z) - \frac{\alpha}{2}S(X, Y, Z) \quad (\alpha \in \mathbb{R}).$$

Note that the above argument naturally leads to a characterisation of Markov invariant (1, 2)-type tensor field  $F(X, Y)$  through the relation  $g(F(X, Y), Z) = S(X, Y, Z)$ . One may naturally generalise this idea to characterising Markov invariant (1,  $s$ )-type tensor fields in terms of Markov invariant (0,  $s + 1$ )-type tensor fields. However, one cannot simply extend this argument to generic ( $r, s$ )-type tensor fields. Now, a question naturally arises; how can one characterise Markov invariant ( $r, s$ )-type tensor fields? The purpose of this section is to answer this question.

Associated with each Markov embedding  $f : \mathcal{S}_{n-1} \rightarrow \mathcal{S}_{\ell-1}$  is a unique affine map

$$\varphi_f : \mathcal{S}_{\ell-1} \longrightarrow \mathcal{S}_{n-1} : (y^1, \dots, y^\ell) \mapsto (x^1, \dots, x^n)$$

that satisfies

$$\varphi_f \circ f = \text{id}.$$

In fact, it is explicitly given by the following relations

$$x^i = \sum_{j \in C(i)} y^j \quad (i = 1, \dots, n)$$

that allocate each event  $C_{(i)} (\subset \Omega_\ell)$  to the singleton  $\{i\} (\subset \Omega_n)$ . (For a proof, see [7].) We shall call the map  $\varphi_f$  the *coarse-graining* associated with a Markov embedding  $f$ . Note that the coarse-graining  $\varphi_f$  is determined only by the partition (1), and is independent of the internal ratios  $\{Q_{(i)}^j\}_{i,j}$  that specifies  $f$  as (2).

For example, let us consider a Markov embedding

$$f : \mathcal{S}_1 \longrightarrow \mathcal{S}_3 : (p_1, p_2) \longmapsto (\lambda p_1, (1 - \lambda)p_1, \mu p_2, (1 - \mu)p_2), \quad (0 < \lambda, \mu < 1)$$

associated with the partition  $\Omega_4 = C_{(1)} \sqcup C_{(2)}$ , where

$$C_{(1)} = \{1, 2\}, \quad C_{(2)} = \{3, 4\}.$$

Then, the coarse-graining  $\varphi_f : \mathcal{S}_3 \rightarrow \mathcal{S}_1$  associated with  $f$  is given by

$$\varphi_f : (q_1, q_2, q_3, q_4) \longmapsto (q_1 + q_2, q_3 + q_4).$$

Now we introduce a generalised Markov invariance. A series  $\{F^{[n]}\}_{n \in \mathbb{N}}$  of  $(r, s)$ -type tensor fields, each on  $\mathcal{S}_{n-1}$ , is said to be *Markov invariant* if

$$F_p^{[n]}(\omega^1, \dots, \omega^r, X_1, \dots, X_s) = F_{f(p)}^{[\ell]}(\varphi_f^* \omega^1, \dots, \varphi_f^* \omega^r, f_* X_1, \dots, f_* X_s) \quad (7)$$

holds for all  $n, \ell \in \mathbb{N}$  satisfying  $2 \leq n \leq \ell$ , Markov embeddings  $f : \mathcal{S}_{n-1} \rightarrow \mathcal{S}_{\ell-1}$ , points  $p \in \mathcal{S}_{n-1}$ , cotangent vectors  $\omega^1, \dots, \omega^r \in T_p^* \mathcal{S}_{n-1}$ , and tangent vectors  $X_1, \dots, X_s \in T_p \mathcal{S}_{n-1}$ . When no confusion arises, we simply use an abridged notation  $F$  for  $F^{[n]}$ .

The main result of this section is the following.

**Theorem 1** *Markov invariant tensor fields are closed under the operations of raising and lowering indices with respect to the Fisher metric  $g$ .*

In order to prove Theorem 1, we need some preliminary considerations. Suppose we want to know whether the  $(1, 2)$ -type tensor field  $F^i{}_{jk} := g^{im} S_{mjk}$  is Markov invariant in the sense of (7), where  $S$  is the Markov invariant  $(0, 3)$ -type tensor field defined by (6). Put differently, we want to investigate if, for some (then any) local coordinate system  $(\xi^a)$  of  $\mathcal{S}_{n-1}$ , the  $(1, 2)$ -type tensor field  $F$  defined by  $F \left( d\xi^a, \frac{\partial}{\partial \xi^b}, \frac{\partial}{\partial \xi^c} \right) := g^{ae} S_{ebc}$  exhibits

$$F_p \left( d\xi^a, \frac{\partial}{\partial \xi^b}, \frac{\partial}{\partial \xi^c} \right) = F_{f(p)} \left( \varphi_f^* d\xi^a, f_* \frac{\partial}{\partial \xi^b}, f_* \frac{\partial}{\partial \xi^c} \right). \quad (8)$$

In order to handle such a relation, it is useful to identify the Fisher metric  $g$  on the manifold  $\mathcal{S}_{n-1}$  and its inverse  $g^{-1}$  with the following linear maps:

$$g : T\mathcal{S}_{n-1} \longrightarrow T^*\mathcal{S}_{n-1} : \frac{\partial}{\partial \xi^a} \longmapsto g_{ab} d\xi^b,$$

$$g^{-1} : T^* \mathcal{S}_{n-1} \longrightarrow T \mathcal{S}_{n-1} : d\xi^a \longmapsto g^{ab} \frac{\partial}{\partial \xi^b}.$$

Note that these maps do not depend on the choice of a local coordinate system  $(\xi^a)$  of  $\mathcal{S}_{n-1}$ .

Now, observe that

$$\begin{aligned} \text{the left-hand side of (8)} &= S_p \circ (g_p^{-1} \otimes I \otimes I) \left( d\xi^a, \frac{\partial}{\partial \xi^b}, \frac{\partial}{\partial \xi^c} \right) \\ &= S_p \left( g_p^{ae} \frac{\partial}{\partial \xi^e}, \frac{\partial}{\partial \xi^b}, \frac{\partial}{\partial \xi^c} \right) \end{aligned}$$

and

$$\begin{aligned} \text{the right-hand side of (8)} &= S_{f(p)} \circ (g_{f(p)}^{-1} \otimes I \otimes I) \left( \varphi_f^* d\xi^a, f_* \frac{\partial}{\partial \xi^b}, f_* \frac{\partial}{\partial \xi^c} \right) \\ &= S_{f(p)} \left( g_{f(p)}^{-1} (\varphi_f^* d\xi^a), f_* \frac{\partial}{\partial \xi^b}, f_* \frac{\partial}{\partial \xi^c} \right). \end{aligned}$$

Since the  $(0, 3)$ -type tensor field  $S$  is Markov invariant, the following Lemma establishes (8).

**Lemma 2** *For any Markov embedding  $f : \mathcal{S}_{n-1} \rightarrow \mathcal{S}_{\ell-1}$ , it holds that*

$$f_* \left( g_p^{ae} \frac{\partial}{\partial \xi^e} \right) = g_{f(p)}^{-1} (\varphi_f^* d\xi^a). \tag{9}$$

*In other words, the diagram*

$$\begin{array}{ccc} T_p^* \mathcal{S}_{n-1} & \xrightarrow{\varphi_f^*} & T_{f(p)}^* \mathcal{S}_{\ell-1} \\ g^{-1} \downarrow & & g^{-1} \downarrow \\ T_p \mathcal{S}_{n-1} & \xrightarrow{f_*} & T_{f(p)} \mathcal{S}_{\ell-1} \end{array}$$

*is commutative.*

For the proof of Lemma 2, consult the original paper [7]. Lemma 2 has the following implication: raising indices with respect to the Fisher metric preserves Markov invariance. Note that this result is consistent with the observation in the opening paragraphs of this section where the Markov invariance of a  $(1, 2)$ -type tensor field  $F$  was connected with the Markov invariance of the  $(0, 3)$ -type tensor field  $S$ .

Let us proceed to the issue of lowering indices. Suppose that, given a Markov invariant  $(3, 0)$ -type tensor field  $T$ , we want to know whether the  $(2, 1)$ -type tensor field  $F$  defined by

$$F \left( \frac{\partial}{\partial \xi^a}, d\xi^b, d\xi^c \right) := g_{ae} T^{ebc}$$

satisfies Markov invariance:

$$F_p \left( \frac{\partial}{\partial \xi^a}, d\xi^b, d\xi^c \right) = F_{f(p)} \left( f_* \frac{\partial}{\partial \xi^a}, \varphi_f^* d\xi^b, \varphi_f^* d\xi^c \right)$$

or equivalently

$$T_p \left( (g_p)_{ae} d\xi^e, d\xi^b, d\xi^c \right) = T_{f(p)} \left( g_{f(p)} \left( f_* \frac{\partial}{\partial \xi^a} \right), \varphi_f^* d\xi^b, \varphi_f^* d\xi^c \right).$$

This question is resolved affirmatively by the following

**Lemma 3** *For any Markov embedding  $f : S_{n-1} \rightarrow S_{\ell-1}$ , it holds that*

$$\varphi_f^* ((g_p)_{ae} d\xi^e) = g_{f(p)} \left( f_* \frac{\partial}{\partial \xi^a} \right). \tag{10}$$

*In other words, the diagram*

$$\begin{array}{ccc} T_p^* S_{n-1} & \xrightarrow{\varphi_f^*} & T_{f(p)}^* S_{\ell-1} \\ g \uparrow & & g \uparrow \\ T_p S_{n-1} & \xrightarrow{f_*} & T_{f(p)} S_{\ell-1} \end{array}$$

*is commutative.*

**Proof** Since  $g$  is an isomorphism, the identity (10) is an immediate consequence of Lemma 2. In fact, it follows from (9) that

$$f_* \left( \frac{\partial}{\partial \xi^e} \right) = g_{f(p)}^{-1} \left( \varphi_f^* ((g_p)_{ea} d\xi^a) \right),$$

and thus

$$g_{f(p)} \left( f_* \frac{\partial}{\partial \xi^e} \right) = \varphi_f^* ((g_p)_{ea} d\xi^a),$$

proving the claim. □

Lemma 3 has the following implication: lowering indices with respect to the Fisher metric preserves Markov invariance. Theorem 1 is now an immediate consequence of Lemmas 2 and 3, as well as the line of arguments that precede those lemmas.

Theorem 1 has a remarkable consequence: every  $(r, s)$ -type Markov invariant tensor field can be obtained by raising indices of some  $(0, r + s)$ -type Markov invariant tensor field. For example,  $g^{ij}$  is, up to scaling, the only  $(2, 0)$ -type Markov invariant tensor field.

It may be worthwhile to mention that not every operation that is standard in tensor calculus preserves Markov invariance. The following example is due to Amari [1].

**Example 4** With a  $\nabla^{(+1)}$ -affine coordinate system  $\theta = (\theta^1, \dots, \theta^{n-1})$  of  $\mathcal{S}_{n-1}$  defined by

$$\log p(\omega) = \sum_{i=1}^{n-1} \theta^i \delta_i(\omega) - \log \left( 1 + \sum_{k=1}^{n-1} \exp \theta^k \right) \quad (\omega \in \Omega_n),$$

the Amari-Chentsov tensor field (6) has the following components:

$$S_{ijk} = \begin{cases} \eta_i(1 - \eta_i)(1 - 2\eta_i) & (i = j = k) \\ -\eta_i(1 - 2\eta_i)\eta_k & (i = j \neq k) \\ -\eta_j(1 - 2\eta_j)\eta_i & (j = k \neq i) \\ -\eta_k(1 - 2\eta_k)\eta_j & (k = i \neq j) \\ 2\eta_i\eta_j\eta_k & (i \neq j \neq k \neq i) \end{cases}.$$

Here,  $\eta = (\eta_1, \dots, \eta_{n-1})$  is a  $\nabla^{(-1)}$ -affine coordinate system of  $\mathcal{S}_{n-1}$  that is dual to  $\theta$ ; more succinctly,  $\eta_i = p(i)$ . By using the formula

$$g^{ij} = \frac{1}{\eta_n} + \frac{\delta^{ij}}{\eta_i} \quad \left( \eta_n := 1 - \sum_{i=1}^{n-1} \eta_i \right),$$

the (1, 2)-type tensor field  $T^i_{jk} := g^{im} S_{mjk}$  is readily calculated as <sup>2</sup>

$$T^i_{jk} = \begin{cases} 1 - 2\eta_i & (i = j = k) \\ -\eta_k & (i = j \neq k) \\ -\eta_j & (i = k \neq j) \\ 0 & (i \neq j, i \neq k) \end{cases}. \tag{11}$$

We know that  $T$  is Markov invariant (either from Theorem 1 or from the discussion in the opening paragraphs of this section). However, the following contracted (0, 1)-type tensor field

$$\tilde{F}_k := T^i_{ik} = 1 - n\eta_k$$

is non-zero, and hence is not Markov invariant.<sup>3</sup> This demonstrates that the contraction, which is a standard operation in tensor calculus, does not always preserve Markov invariance.

<sup>2</sup> Incidentally, the coefficients of  $\alpha$ -connections in the coordinate system  $\theta$  are related to  $S_{ijk}$  and  $T^i_{jk}$  as follows:

$$\Gamma^{(\alpha)}_{ij,k} = \frac{1 - \alpha}{2} S_{ijk} \quad \text{and} \quad \Gamma^{(\alpha)k}_{ij} = \frac{1 - \alpha}{2} T^k_{ij}.$$

<sup>3</sup> Recall that the only Markov invariant (0, 1)-type tensor field on  $\mathcal{S}_{n-1}$  is zero.



Chentsov’s idea of imposing the invariance of geometrical structures under Markov embeddings  $f : \mathcal{S}_{n-1} \rightarrow \mathcal{S}_{\ell-1}$  is based on the fact that  $\mathcal{S}_{n-1}$  is statistically isomorphic to  $f(\mathcal{S}_{n-1})$ . Put differently, the Markov invariance only involves direct comparison between  $\mathcal{S}_{n-1}$  and its image  $f(\mathcal{S}_{n-1})$ , and is nothing to do with the complement of  $f(\mathcal{S}_{n-1})$  in the ambient space  $\mathcal{S}_{\ell-1}$ . On the other hand, the partial trace operation  $T^i_{jk} \mapsto T^i_{ik}$  on  $\mathcal{S}_{\ell-1}$  (more precisely, on  $T_{f(p)}\mathcal{S}_{\ell-1} \otimes T^*_{f(p)}\mathcal{S}_{\ell-1}$ ) makes the output  $T^i_{ik}$  ‘contaminated’ with information from outside the submanifold  $f(\mathcal{S}_{n-1})$ . It is thus no wonder such an influx of extra information manifests itself as the non-preservation of Markov invariance. In this respect, a distinctive characteristic of Lemmas 2 and 3 lies in the fact that raising and lowering indices preserve Markov invariance although they are represented in the forms of contraction such as  $g^{i\ell}S_{mjk} \mapsto g^{im}S_{mjk}$  or  $g_{i\ell}T^{mjk} \mapsto g_{im}T^{mjk}$ .

**Remark 5** A related intriguing instance arises in curvature tensors: the Ricci curvature of the  $\alpha$ -connection  $\nabla^{(\alpha)}$  of  $\mathcal{S}_{n-1}$  is calculated to be

$$\text{Ric}^{\nabla^{(\alpha)}} = (n - 2) \frac{1 - \alpha^2}{4} g.$$

Specifically, the manifold  $\mathcal{S}_{n-1}$  is Einstein for all  $\alpha \in \mathbb{R}$ . Moreover, for  $n \geq 3$ , the Ricci curvature divided by  $n - 2$  is Markov invariant for all  $\alpha \in \mathbb{R}$ . Note that when  $n = 2$ , the manifold  $\mathcal{S}_{n-1}$  is one-dimensional and thus is flat for all  $\alpha$ .

### 3 Geometry of manifolds of continuous probability densities

In their celebrated book, Amari and Nagaoka stated that it is not so easy to extend Chentsov’s theorem to the case when the underlying set  $\mathcal{X}$  of outcomes is infinite [2, p. 38]. There have been several attempts to deal with infinite outcome spaces and/or general measure spaces such as [3, 4, 9], but they are all technically demanding. Amari and Nagaoka also suggested a completely different approach to comprehend the Fisher metric and the  $\alpha$ -connections of a parametric model  $M = \{p_\theta(x) : \theta \in \Theta \subset \mathbb{R}^d, x \in \mathcal{X}\}$  from the viewpoint of Chentsov’s theorem as follows <sup>4</sup> [2, p. 39]:

*First, let us finitely partition  $\mathcal{X}$  into the regions  $\Delta_1, \Delta_2, \dots, \Delta_n$ . In other words, each  $\Delta_i$  is a subset of  $\mathcal{X}$ ,  $\Delta_i \cap \Delta_j = \emptyset$  ( $i \neq j$ ), and  $\bigcup_{i=1}^n \Delta_i = \mathcal{X}$ . Now fix a particular partition  $\Delta = \{\Delta_1, \Delta_2, \dots, \Delta_n\}$  and let*

$$P_\theta^\Delta(i) := \int_{\Delta_i} p_\theta(x) dx.$$

*Then  $M^\Delta := \{P_\theta^\Delta(i)\}$  forms a model on  $\Delta$ . Since  $\Delta$  is a finite set, from Chentsov’s theorem we know that the Fisher metric and the  $\alpha$ -connections are introduced on  $M^\Delta$  by the invariance requirement. Now we may consider  $M$  to be the limit of  $M^\Delta$  as  $\Delta$  becomes finer and finer. Hence, if we require that the desired metrics and connections*

<sup>4</sup> Notations are slightly changed according to the context of the present article.

on models should be “continuous” with respect to such a limit, it is concluded that the metric and the connections on  $M$  should be given by the limit of the Fisher metric and the  $\alpha$ -connections on  $M^\Delta$ , and under some regularity condition they coincide with the Fisher metric and the  $\alpha$ -connections on  $M$ .

It is crucial to notice that the coarse-graining  $p_\theta \mapsto P_\theta^\Delta$  does not in general have a sufficient statistic.<sup>5</sup> This is in a striking contrast to the situation of the previous sections where we treated Markov embeddings that warranted the existence of sufficient statistics. In order to realise the above programme, therefore, it is important to scrutinise the limiting procedure. However, the meaning of “the limit of  $M^\Delta$  as  $\Delta$  becomes finer and finer” is mathematically unclear, and to the best of the author’s knowledge, this limiting procedure has not been treated explicitly in the literature. The purpose of this section is to demonstrate Amari and Nagaoka’s programme when the underlying sample space is  $\mathcal{X} = \mathbb{R}^k$  and the density function  $p_\theta(x)$  is continuous in  $x \in \mathbb{R}^k$ , a simple yet typical situation in statistics.

Let  $M = \{p_\theta(x) : \theta \in \Theta \subset \mathbb{R}^d, x \in \mathbb{R}^k\}$  be a  $d$ -dimensional parametric family of probability density functions on  $\mathbb{R}^k$ . We assume the following regularity conditions:

- (i) the support of  $p_\theta$  does not depend on  $\theta$ .
- (ii)  $p_\theta(x)$  is differentiable in  $\theta$ , and both  $p_\theta(x)$  and its derivative  $Xp_\theta(x)$  are continuous in  $x$  for all  $\theta \in \Theta$  and  $X \in T_{p_\theta}M$ .
- (iii) for all Jordan measurable<sup>6</sup> domains  $A \subset \mathbb{R}^k, \theta \in \Theta$ , and  $X \in T_{p_\theta}M$ ,

$$X \int_A p_\theta(x) dx = \int_A Xp_\theta(x) dx.$$

- (iv) for all  $\theta \in \Theta$  and  $X \in T_{p_\theta}M$ , the Amari-Chentsov tensor

$$S_\theta(X, X, X) = \int_{\mathbb{R}^k} p_\theta(x) \left( \frac{Xp_\theta(x)}{p_\theta(x)} \right)^3 dx$$

is absolutely convergent.

In condition (i), the support of  $p_\theta$  can be arbitrary; however, we assume in what follows that the support is  $\mathbb{R}^k$  for concreteness. Let  $\Delta = \{\Delta_1, \Delta_2, \dots, \Delta_n\}$  be a Jordan measurable finite partition of  $\mathbb{R}^k$  such that the interior of each  $\Delta_i$  is open and connected. We denote the totality of such finite partitions of  $\mathbb{R}^k$  by  $\mathcal{I}$ . Note that  $\mathcal{I}$  is

<sup>5</sup> Amari and Nagaoka mentioned this fact in the original Japanese edition of [2].

<sup>6</sup> A bounded set  $A \subset \mathbb{R}^k$  is called *Jordan measurable* if the inner Jordan measure of  $A$  (the supremum of volumes of nonoverlapping left-closed rectangles that belong to  $A$ ) equals the outer Jordan measure of  $A$  (the infimum of volumes of nonoverlapping left-closed rectangles that cover  $A$ ). In the present article, we further make an extended use of this terminology: a set  $A \subset \mathbb{R}^k$ , which can be unbounded, is called Jordan measurable if  $A \cap \{x \in \mathbb{R}^k : |x| \leq R\}$  is Jordan measurable for all  $R > 0$ .

a directed system <sup>7</sup> endowed with partial ordering  $\Delta < \Delta'$  having the interpretation that  $\Delta'$  is finer than  $\Delta$ .

Associated with a finite partition  $\Delta = \{\Delta_1, \Delta_2, \dots, \Delta_n\} \in \mathcal{I}$  is a parametric model  $M^\Delta = \{P_\theta^\Delta\}_\theta$  on the finite set  $\Omega_n = \{1, 2, \dots, n\}$  defined by

$$P_\theta^\Delta(i) := \int_{\Delta_i} p_\theta(x) dx \quad (i \in \Omega_n).$$

We are interested in the relationship between the original model  $M = \{p_\theta\}$  and the induced model  $M^\Delta = \{P_\theta^\Delta\}$ . Specifically, we want to know if the nets of the Fisher metrics  $\{g_\theta^\Delta\}_\Delta$  and the Amari-Chentsov tensors  $\{S_\theta^\Delta\}_\Delta$  on  $M^\Delta$  converge to the Fisher metric  $g_\theta$  and the Amari-Chentsov tensor  $S_\theta$  on  $M$ , respectively. The next theorem gives an affirmative answer to this question.

**Theorem 6** *Under regularity conditions (i)–(iv),*

$$\lim_{\Delta \in \mathcal{I}} g_\theta^\Delta(X, Y) = g_\theta(X, Y) \quad \text{and} \quad \lim_{\Delta \in \mathcal{I}} S_\theta^\Delta(X, Y, Z) = S_\theta(X, Y, Z)$$

hold for all  $X, Y, Z \in T_{p_\theta} M$ .

Theorem 6 could be paraphrased by saying that the Fisher metric and the  $\alpha$ -connections are the only natural Markov invariant geometrical structures of parametric models comprising continuous probability densities on  $\mathbb{R}^k$ .

Before proceeding to the proof, we introduce some notations that are used throughout the proof. For  $R > 0$ , let  $B^R$  denote the closed ball of radius  $R$  in  $\mathbb{R}^k$  centred at the origin, i.e.,

$$B^R := \{x \in \mathbb{R}^k : |x| \leq R\}.$$

Given a finite partition  $\Delta \in \mathcal{I}$ , let

$$\Delta^R = \{\Delta_j^R\}_{j=1}^{n_1+n_2}$$

denote a refinement of  $\Delta$  in  $\mathcal{I}$  such that  $\{\Delta_j^R\}_{j=1}^{n_1}$  and  $\{\Delta_j^R\}_{j=n_1+1}^{n_1+n_2}$  are partitions of  $B^R$  and its complement  $\mathbb{R}^k \setminus B^R$ , respectively. Note that  $n_1$  and  $n_2$  may depend both on  $\Delta$  and  $R$ .

Now we proceed to the proof of Theorem 6. By virtue of the standard polarisation argument using the identity

$$g(X, Y) = \frac{1}{2} \{g(X + Y, X + Y) - g(X, X) - g(Y, Y)\}$$

<sup>7</sup> A directed system is an index set  $\mathcal{I}$  together with a partial ordering  $<$  which satisfies the condition that if  $\alpha, \beta \in \mathcal{I}$ , then there exists  $\gamma \in \mathcal{I}$  so that  $\alpha < \gamma$  and  $\beta < \gamma$ . A net  $\{x_\alpha\}_{\alpha \in \mathcal{I}}$  in a topological space  $X$  (i.e., a mapping from a directed system  $\mathcal{I}$  to  $X$ ) is said to converge to a point  $x \in X$ , written  $\lim_{\alpha \in \mathcal{I}} x_\alpha = x$ , if for any neighbourhood  $\mathcal{N}_x$  of  $x$ , there exists a  $\beta \in \mathcal{I}$  so that  $\beta < \alpha$  implies  $x_\alpha \in \mathcal{N}_x$ . See [10] for more information.

and its analogue

$$\begin{aligned}
 S(X, Y, Z) = \frac{1}{6} \{ & S(X + Y + Z, X + Y + Z, X + Y + Z) \\
 & - S(X + Y, X + Y, X + Y) - S(X + Z, X + Z, X + Z) \\
 & - S(Y + Z, Y + Z, Y + Z) + S(X, X, X) + S(Y, Y, Y) \\
 & + S(Z, Z, Z) \},
 \end{aligned}$$

which are valid for symmetric tensors  $g$  and  $S$ , we see that Theorem 6 is proved simply by showing that

$$\lim_{\Delta \in \mathcal{I}} g_{\theta}^{\Delta}(X, X) = g_{\theta}(X, X) \tag{12}$$

and

$$\lim_{\Delta \in \mathcal{I}} S_{\theta}^{\Delta}(X, X, X) = S_{\theta}(X, X, X) \tag{13}$$

for all  $X \in T_{p_{\theta}}M$ . Since the proof of (12) is almost the same as that of (13), we shall present only the latter here.

The Amari–Chentsov tensor  $S_{\theta}^{\Delta R}(X, X, X)$  of the induced model  $M^{\Delta R} = \{P_{\theta}^{\Delta R}\}$  is decomposed into two parts:

$$S_{\theta}^{\Delta R}(X, X, X) = \sum_{i=1}^{n_1} P_{\theta}^{\Delta R}(i) \left( \frac{X P_{\theta}^{\Delta R}(i)}{P_{\theta}^{\Delta R}(i)} \right)^3 + \sum_{i=n_1+1}^{n_1+n_2} P_{\theta}^{\Delta R}(i) \left( \frac{X P_{\theta}^{\Delta R}(i)}{P_{\theta}^{\Delta R}(i)} \right)^3. \tag{14}$$

Firstly, let us evaluate the first term of the right-hand side of (14).

**Lemma 7**

$$\lim_{\Delta \in \mathcal{I}} \sum_{i=1}^{n_1} P_{\theta}^{\Delta R}(i) \left( \frac{X P_{\theta}^{\Delta R}(i)}{P_{\theta}^{\Delta R}(i)} \right)^3 = \int_{B^R} p_{\theta}(x) \left( \frac{X p_{\theta}(x)}{p_{\theta}(x)} \right)^3 dx.$$

**Proof** Due to the mean-value theorem, for each  $i = 1, \dots, n_1$ , there is an  $x_i \in \Delta_i^R$  such that

$$P_{\theta}^{\Delta R}(i) = \int_{\Delta_i^R} p_{\theta}(x) dx = p_{\theta}(x_i) \mu(\Delta_i^R),$$

where  $\mu(\Delta_i^R)$  is the Jordan measure of the region  $\Delta_i^R$ . Similarly, for each  $i = 1, \dots, n_1$ , there is a  $\xi_i \in \Delta_i^R$  such that

$$X P_{\theta}^{\Delta R}(i) = \int_{\Delta_i^R} X p_{\theta}(x) dx = X p_{\theta}(\xi_i) \mu(\Delta_i^R).$$

Thus,

$$\sum_{i=1}^{n_1} P_{\theta}^{\Delta^R}(i) \left( \frac{X P_{\theta}^{\Delta^R}(i)}{P_{\theta}^{\Delta^R}(i)} \right)^3 = \sum_{i=1}^{n_1} p_{\theta}(x_i) \left( \frac{X p_{\theta}(\xi_i)}{p_{\theta}(x_i)} \right)^3 \mu(\Delta_i^R),$$

and the next Lemma 8 proves the claim. □

**Lemma 8** *Let  $f$  and  $g$  be continuous functions on a Jordan measurable bounded closed domain  $D \subset \mathbb{R}^k$ . Given a Jordan measurable finite partition  $\Delta = \{\Delta_1, \dots, \Delta_n\}$  of  $D$ , take arbitrary points  $x_i$  and  $\xi_i$  in  $\Delta_i$  for each  $i = 1, \dots, n$ . Then*

$$\lim_{\Delta} \sum_{i=1}^n f(x_i)g(\xi_i)\mu(\Delta_i) = \int_D f(x)g(x)dx,$$

where the limit is taken over all Jordan measurable finite partitions  $\Delta$  of  $D$ .

**Proof** Since

$$\sum_{i=1}^n f(x_i)g(\xi_i)\mu(\Delta_i) = \sum_{i=1}^n f(x_i)g(x_i)\mu(\Delta_i) + \sum_{i=1}^n f(x_i)\{g(\xi_i) - g(x_i)\}\mu(\Delta_i),$$

it suffices to prove that

$$\lim_{\Delta} \sum_{i=1}^n f(x_i)\{g(\xi_i) - g(x_i)\}\mu(\Delta_i) = 0.$$

We see from the Cauchy-Schwarz inequality that

$$\begin{aligned} & \left( \sum_{i=1}^n f(x_i)\{g(\xi_i) - g(x_i)\}\mu(\Delta_i) \right)^2 \\ & \leq \left( \sum_{i=1}^n f(x_i)^2\mu(\Delta_i) \right) \left( \sum_{i=1}^n \{g(\xi_i) - g(x_i)\}^2\mu(\Delta_i) \right). \end{aligned}$$

For  $\Delta = \{\Delta_1, \dots, \Delta_n\}$ , let  $|\Delta| := \max_{1 \leq i \leq n} |\Delta_i|$ , where  $|\Delta_i|$  is the diameter of  $\Delta_i$ . Since  $g$  is uniformly continuous on  $D$ , for any  $\varepsilon > 0$ , there exists  $\delta > 0$  so that  $|\Delta| < \delta$  implies  $|g(\xi_i) - g(x_i)| < \varepsilon$  for all  $i = 1, \dots, n$ . As a consequence,

$$\sum_{i=1}^n \{g(\xi_i) - g(x_i)\}^2\mu(\Delta_i) < \varepsilon^2\mu(D).$$

Since

$$\lim_{\Delta} \sum_{i=1}^n f(x_i)^2\mu(\Delta_i) = \int_D f(x)^2dx < \infty,$$

the claim is verified. □

We next evaluate the second term of the right-hand side of (14).

**Lemma 9**

$$\lim_{R \rightarrow \infty} \sum_{i=n_1+1}^{n_1+n_2} P_{\theta}^{\Delta^R}(i) \left| \frac{X P_{\theta}^{\Delta^R}(i)}{P_{\theta}^{\Delta^R}(i)} \right|^3 = 0.$$

**Proof** Since  $p_{\theta}(x)/P_{\theta}^{\Delta^R}(i)$  is a probability density on the region  $\Delta_i^R$ , we apply Jensen’s inequality to the convex function  $t \mapsto |t|^3$ , to obtain

$$\begin{aligned} \frac{1}{P_{\theta}^{\Delta^R}(i)} \int_{\Delta_i^R} p_{\theta}(x) \left| \frac{X p_{\theta}(x)}{p_{\theta}(x)} \right|^3 dx &\geq \left| \frac{1}{P_{\theta}^{\Delta^R}(i)} \int_{\Delta_i^R} p_{\theta}(x) \frac{X p_{\theta}(x)}{p_{\theta}(x)} dx \right|^3 \\ &= \left| \frac{1}{P_{\theta}^{\Delta^R}(i)} \int_{\Delta_i^R} X p_{\theta}(x) dx \right|^3 \\ &= \left| \frac{X P_{\theta}^{\Delta^R}(i)}{P_{\theta}^{\Delta^R}(i)} \right|^3. \end{aligned}$$

Consequently,

$$\int_{\Delta_i^R} p_{\theta}(x) \left| \frac{X p_{\theta}(x)}{p_{\theta}(x)} \right|^3 dx \geq P_{\theta}^{\Delta^R}(i) \left| \frac{X P_{\theta}^{\Delta^R}(i)}{P_{\theta}^{\Delta^R}(i)} \right|^3.$$

Taking the sum over  $i = n_1 + 1, \dots, n_1 + n_2$ , we have

$$\int_{\mathbb{R}^k \setminus B^R} p_{\theta}(x) \left| \frac{X p_{\theta}(x)}{p_{\theta}(x)} \right|^3 dx \geq \sum_{i=n_1+1}^{n_1+n_2} P_{\theta}^{\Delta^R}(i) \left| \frac{X P_{\theta}^{\Delta^R}(i)}{P_{\theta}^{\Delta^R}(i)} \right|^3.$$

Since regularity condition (iv) implies

$$\lim_{R \rightarrow \infty} \int_{\mathbb{R}^k \setminus B^R} p_{\theta}(x) \left| \frac{X p_{\theta}(x)}{p_{\theta}(x)} \right|^3 dx = 0,$$

we have the claim. □

Applying Lemmas 7 and 9 to (14), we conclude that for any  $\varepsilon > 0$ , there exist  $\Delta \in \mathcal{I}$  and  $R > 0$  such that  $\Delta^R \prec \Delta'$  implies

$$\left| S_{\theta}^{\Delta'}(X, X, X) - S_{\theta}(X, X, X) \right| < \varepsilon.$$

This completes the proof of (13).

**Remark 10** The continuity of the density  $p_\theta(x)$  and its derivative  $Xp_\theta(x)$  in regularity condition (ii) is introduced solely for the sake of simplicity, and can be loosened depending on the situation. For example, Theorem 6 is still valid even if  $p_\theta(x)$  and  $Xp_\theta(x)$  have finitely many discontinuity points.

## Appendix

### A Chentsov's argument characterising affine connections

Introduce a coordinate system  $\theta = (\theta^i)_{1 \leq i \leq n-1}$  of  $\mathcal{S}_{n-1}$  as in Example 4, that is,

$$\log p(\omega) = \sum_{i=1}^{n-1} \theta^i \delta_i(\omega) - \log \left( 1 + \sum_{k=1}^{n-1} \exp \theta^k \right) \quad (\omega \in \Omega_n).$$

Then, Chentsov's theorem [6, Theorem 12.2] is restated as follows.

**Theorem 11** *The affine connections  $\nabla$  on  $\mathcal{S}_{n-1}$  satisfying the equivariance condition*

$$f_* \left( \nabla_X^{[n]} Y \right)_p = \left( \nabla_{f_* X}^{[\ell]} f_* Y \right)_{f(p)} \tag{15}$$

are all described by formulas

$$\nabla_{\partial_i} \partial_i = \gamma (1 - 2\eta_i) \partial_i \tag{16}$$

$$\nabla_{\partial_i} \partial_j = -\gamma (\eta_i \partial_j + \eta_j \partial_i) \quad (i \neq j) \tag{17}$$

where  $\partial_i := \partial/\partial\theta^i$ ,  $\eta_i := p(i)$ , and  $\gamma$  is a real parameter.

Note that (16) and (17) are rewritten as

$$\Gamma_{ij}^k = \gamma T^k_{ij}$$

where  $T^k_{ij}$  are defined by (11). Thus, by setting  $\gamma = (1 - \alpha)/2$ , we restore the  $\alpha$ -connections as demonstrated in the footnote of Example 4.

**Proof of Theorem 11** We divide the proof into four steps.

*Step 1.* For  $i = 1, \dots, n - 1$ , let  $X_i := \partial/\partial\theta^i$  be the vector fields associated with the coordinate system  $\theta = (\theta^i)_{1 \leq i \leq n-1}$ . In order to comprehend these vector fields in terms of elementary geometry, let us represent the tangent vector  $(X_i)_p$  at  $p \in \mathcal{S}_{n-1}$  by a numerical vector  $(\vec{X}_i)_p \in \mathbb{R}^n$  whose  $\omega$ th entry ( $1 \leq \omega \leq n$ ) is given by

$$\frac{\partial}{\partial\theta^i} p(\omega) = p(\omega) \frac{\partial}{\partial\theta^i} \log p(\omega) = p(\omega) \{ \delta_i(\omega) - p(i) \} = p(i) \{ \delta_i(\omega) - p(\omega) \}.$$

Note that the numerical vector  $\{ \delta_i(\omega) - p(\omega) \}_{1 \leq \omega \leq n}$  corresponds to the arrow connecting the initial point  $p$  with the terminal point  $e_i$ , the  $i$ th vertex of the probability

simplex  $\mathcal{S}_{n-1}$ . In this way, the tangent vector  $(X_i)_p$  is interpreted as the geometrical vector  $\vec{p}e_i = \vec{e}_i - \vec{p}$  multiplied by  $p(i)$ . Further, following Chentsov, we introduce another vector field  $X_n$  that has a geometrical vector interpretation  $\vec{p}e_n = \vec{e}_n - \vec{p}$  multiplied by  $p(n)$ , i.e., whose numerical vector representation  $(\vec{X}_n)_p$  has the form

$$p(n) \{ \delta_n(\omega) - p(\omega) \} \quad (1 \leq \omega \leq n).$$

In what follows, these representations are used interchangeably for tangent vectors  $(X_i)_p$  with  $i = 1, \dots, n$ . Note that the vector fields  $X_1, \dots, X_n$  satisfy the identity:

$$\sum_{i=1}^n X_i = 0. \tag{18}$$

Similarly, we introduce a set of vector fields  $Y_1, \dots, Y_\ell$  on  $\mathcal{S}_{\ell-1}$ . A crucial observation is that for a Markov embedding  $f : \mathcal{S}_{n-1} \rightarrow \mathcal{S}_{\ell-1}$  associated with the partition (1), we have

$$f_*((X_i)_p) = \sum_{k \in C(i)} (Y_k)_{f(p)} \quad (1 \leq i \leq n). \tag{19}$$

In fact, due to (2),

$$f(\vec{e}_i) = \sum_{k \in C(i)} Q_{(i)}^k \vec{\epsilon}_k,$$

where  $\epsilon_k$  is the  $k$ th vertex of  $\mathcal{S}_{\ell-1}$ . As a consequence,

$$\vec{q} := f(\vec{p}) = \sum_{\omega \in \Omega_n} p(\omega) f(\vec{e}_\omega) = \sum_{\omega \in \Omega_n} \sum_{k \in C(\omega)} p(\omega) Q_{(\omega)}^k \vec{\epsilon}_k = \sum_{k \in \Omega_\ell} q(k) \vec{\epsilon}_k,$$

where  $q(k) := p(\omega) Q_{(\omega)}^k$  with the index  $\omega$  being the one satisfying  $k \in C(\omega)$ . By using this equality, we have

$$\begin{aligned} f_*((X_i)_p) &= p(i) f_*(\vec{e}_i - \vec{p}) = p(i) \left( \sum_{k \in C(i)} Q_{(i)}^k \vec{\epsilon}_k - \vec{q} \right) \\ &= p(i) \sum_{k \in C(i)} Q_{(i)}^k (\vec{\epsilon}_k - \vec{q}) \\ &= \sum_{k \in C(i)} q(k) (\vec{\epsilon}_k - \vec{q}) = \sum_{k \in C(i)} (Y_k)_q, \end{aligned}$$

proving (19).



*Step 2.* Let us deal with the case when  $\ell = n$ . In this case, a Markov embedding  $f$  is reduced to a permutation of indices of events. Consider the barycentre

$$p_0 = \left( \frac{1}{n}, \dots, \frac{1}{n} \right)$$

of  $\mathcal{S}_{n-1}$ . Due to the symmetry of  $\mathcal{S}_{n-1}$  over permutations of indices at the barycentre, the tangent vector  $(\nabla_{X_i}^{[n]} X_i)_{p_0}$  must be parallel to  $(X_i)_{p_0}$ , so that there is a constant  $\lambda^{[n]}$  such that

$$\left( \nabla_{X_i}^{[n]} X_i \right)_{p_0} = \lambda^{[n]} (X_i)_{p_0}.$$

Similarly, there is a constant  $\mu^{[n]}$  such that for any distinct pair  $(i, j)$  of indices,

$$\left( \nabla_{X_i}^{[n]} X_j \right)_{p_0} = \mu^{[n]} (X_i + X_j)_{p_0}.$$

Now, using (18), we have

$$\left( \nabla_{X_i}^{[n]} X_i \right)_{p_0} = \left( - \sum_{j \neq i} \nabla_{X_j}^{[n]} X_i \right)_{p_0} = -\mu^{[n]} \sum_{j \neq i} (X_j + X_i)_{p_0} = -\mu^{[n]} (n - 2) (X_i)_{p_0},$$

which leads to

$$\mu^{[n]} = -\frac{\lambda^{[n]}}{n - 2}.$$

*Step 3.* Suppose that  $\ell = Nn$  for some  $N \in \mathbb{N}$ , and consider the Markov embedding

$$f(x^1, \dots, x^n) = \left( \underbrace{\frac{x^1}{N}, \dots, \frac{x^1}{N}}_N, \dots, \underbrace{\frac{x^n}{N}, \dots, \frac{x^n}{N}}_N \right).$$

This map corresponds to the partition

$$C_{(i)} := \{N(i - 1) + 1, \dots, N(i - 1) + N\} \quad (1 \leq i \leq n).$$

Since  $f$  maps the barycentre  $p_0$  of  $\mathcal{S}_{n-1}$  to the barycentre of  $\mathcal{S}_{\ell-1}$ , it follows from the equivariance condition (15) as well as (19) that

$$\begin{aligned} f_* \left( \nabla_{X_i}^{[n]} X_i \right)_{p_0} &= \left( \nabla_{f_* X_i}^{[\ell]} f_* X_i \right)_{f(p_0)} = \sum_{v \in C_{(i)}} \sum_{w \in C_{(i)}} \left( \nabla_{Y_v}^{[\ell]} Y_w \right)_{f(p_0)} \end{aligned}$$

$$\begin{aligned}
 &= \sum_{v \in C(i)} \left( \nabla_{Y_v}^{[\ell]} Y_v \right)_{f(p_0)} + \sum_{v, w \in C(i): v \neq w} \left( \nabla_{Y_v}^{[\ell]} Y_w \right)_{f(p_0)} \\
 &= \sum_{v \in C(i)} \lambda^{[\ell]} (Y_v)_{f(p_0)} + \sum_{v, w \in C(i): v \neq w} \mu^{[\ell]} (Y_v + Y_w)_{f(p_0)} \\
 &= \lambda^{[\ell]} \sum_{v \in C(i)} (Y_v)_{f(p_0)} - \frac{\lambda^{[\ell]}}{\ell - 2} \left( (N - 1) \sum_{v \in C(i)} Y_v + (N - 1) \sum_{w \in C(i)} Y_w \right)_{f(p_0)} \\
 &= \lambda^{[\ell]} \left( 1 - \frac{2(N - 1)}{\ell - 2} \right) f_*(X_i)_{p_0}.
 \end{aligned}$$

Since  $f_*$  is injective, this leads to

$$\lambda^{[n]} = \lambda^{[\ell]} \left( 1 - \frac{2(N - 1)}{\ell - 2} \right).$$

Since  $\ell = Nn$ , this is further equivalent to

$$\frac{n}{n - 2} \lambda^{[n]} = \frac{\ell}{\ell - 2} \lambda^{[\ell]}.$$

Consequently, there exists a constant  $\gamma$ , independent of  $n$ , such that

$$\frac{n}{n - 2} \lambda^{[n]} = \gamma.$$

*Step 4.* Take a rational point  $p$  in  $\mathcal{S}_{n-1}$ , and represent it by a common denominator as

$$p = \left( \frac{m_1}{\ell}, \dots, \frac{m_n}{\ell} \right), \quad (\ell, m_1, \dots, m_n \in \mathbb{N}).$$

Further, let us consider the Markov embedding

$$f(x^1, \dots, x^n) = \left( \underbrace{\frac{x^1}{m_1}, \dots, \frac{x^1}{m_1}}_{m_1}, \dots, \underbrace{\frac{x^n}{m_n}, \dots, \frac{x^n}{m_n}}_{m_n} \right).$$

This map corresponds to the partition

$$C(i) := \{M_i + 1, \dots, M_i + m_i\} \quad (1 \leq i \leq n),$$

where  $M_1 = 0$  and  $M_{j+1} = M_j + m_j$  for  $1 \leq j \leq n - 1$ .

Since the image  $f(p)$  is the barycentre of  $\mathcal{S}_{\ell-1}$ , it follows from the equivariance condition (15) as well as (19) that for  $i \neq j$ ,

$$\begin{aligned} f_* \left( \nabla_{X_i}^{[n]} X_j \right)_p &= \left( \nabla_{f_* X_i}^{[\ell]} f_* X_j \right)_{f(p)} = \sum_{v \in C(i)} \sum_{w \in C(j)} \left( \nabla_{Y_v}^{[\ell]} Y_w \right)_{f(p)} \\ &= \sum_{v \in C(i)} \sum_{w \in C(j)} \left( -\frac{\lambda^{[\ell]}}{\ell - 2} \right) (Y_v + Y_w)_{f(p)} \\ &= -\frac{\gamma}{\ell} \left( m_j \sum_{v \in C(i)} Y_v + m_i \sum_{w \in C(j)} Y_w \right)_{f(p)} \\ &= -\gamma \left( p(j) \sum_{v \in C(i)} Y_v + p(i) \sum_{w \in C(j)} Y_w \right)_{f(p)} \\ &= -\gamma \left( p(j) f_* (X_i)_p + p(i) f_* (X_j)_p \right). \end{aligned}$$

Since  $f_*$  is injective, this leads to

$$\left( \nabla_{X_i}^{[n]} X_j \right)_p = -\gamma \left( p(j) (X_i)_p + p(i) (X_j)_p \right). \tag{20}$$

Further, by using (18),

$$\begin{aligned} \left( \nabla_{X_i}^{[n]} X_i \right)_p &= - \sum_{j \neq i} \left( \nabla_{X_j}^{[n]} X_i \right)_p = \gamma \sum_{j \neq i} \left( p(i) (X_j)_p + p(j) (X_i)_p \right) \\ &= \gamma \left( p(i) \sum_{j \neq i} (X_j)_p + (1 - p(i)) (X_i)_p \right) \\ &= \gamma (1 - 2p(i)) (X_i)_p. \end{aligned} \tag{21}$$

Finally, the relations (20) and (21), which are valid for all rational points  $p \in \mathcal{S}_{n-1}$ , are uniquely extended to all  $p \in \mathcal{S}_{n-1}$  by continuity. This completes the proof.  $\square$

**Acknowledgements** The author would like to express his sincere gratitude to Professor Shun-ichi Amari for all his encouragement and inspiring discussions. He is also grateful to Professor Hiroshi Nagaoka for many insightful comments. The present study was supported by JSPS KAKENHI Grant no. JP17H02861.

**Data availability** Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

**Declarations**

**Conflict of interest** The author states that there is no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give

appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Amari, S.-I.: Private communication (2015)
2. Amari, S.-I., Nagaoka, H.: *Methods of Information Geometry*, Translations of Mathematical Monographs 191 (AMS and Oxford, Providence, 2000); Originally Published in Japanese. Iwanami Shoten, Tokyo (1993)
3. Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: Information geometry and sufficient statistics. *Probab. Theory Relat. Fields* **162**, 327–364 (2015)
4. Bauer, M., Bruveris, M., Michor, P.W.: Uniqueness of the Fisher–Rao metric on the space of smooth densities. *Bull. Lond. Math. Soc.* **48**, 499–506 (2016)
5. Campbell, L.L.: An extended Čencov characterization of the information metric. *Proc. Am. Math. Soc.* **98**, 135–141 (1986)
6. Čencov, N.N.: *Statistical Decision Rules and Optimal Inference*, Translations of Mathematical Monographs 53 (AMS, Providence, 1982); Originally Published in Russian. Nauka, Moscow (1972)
7. Fujiwara, A.: Complementing Chentsov's characterization. In: Ay, N. (ed.) *Information Geometry and Its Applications*, Springer Proceedings in Mathematics & Statistics, vol. **252**. Springer, pp. 335–347 (2018)
8. Fujiwara, A.: *Foundations of Information Geometry*. Kyoritsu Shuppan, Tokyo (2021).. **(in Japanese)**
9. Pistone, G., Sempì, C.: An infinite-dimensional structure on the space of all the probability measures equivalent to a given one. *Ann. Stat.* **5**, 1543–1561 (1995)
10. Reed, M., Simon, B.: *Methods of Modern Mathematical Physics I: Functional Analysis*. Academic Press, San Diego (1980)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.