RESEARCH PAPER

The Fisher metric as a metric on the cotangent bundle

Hiroshi Nagaoka[1]

Abstract

The Fisher metric on a manifold of probability distributions is usually treated as a metric on the tangent bundle. In this paper, we focus on the metric on the cotangent bundle induced from the Fisher metric with calling it the Fisher co-metric. We show that the Fisher co-metric can be defined directly without going through the Fisher metric by establishing a natural correspondence between cotangent vectors and random variables. This definition clarifies a close relation between the Fisher co-metric and the variance/covariance of random variables, whereby the Cramér-Rao inequality is trivialized. We also discuss the monotonicity and the invariance of the Fisher co-metric with respect to Markov maps, and present a theorem characterizing the co-metric by the invariance, which can be regarded as a cotangent version of Čencov's characterization theorem for the Fisher metric. The obtained theorem can also be viewed as giving a characterization of the variance/covariance.

1 Introduction

The Fisher metric on a statistical manifold (a manifold consisting of probability distributions) is one of the most important notions in information geometry [1]. It is usually treated as a Riemannian metric which is a metric on the tangent bundle. The subject of the present paper is the metric on the cotangent bundle corresponding to the Fisher metric, which we call the *Fisher co-metric*. The Fisher metric and the Fisher co-metric are essentially a single geometric object so that one is induced from another. Never-

---

✉ Hiroshi Nagaoka
  nagaoka@is.uec.ac.jp

1   The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu, Tokyo 182-8585, Japan

theless, studying the Fisher co-metric has several implications as mentioned below, which are what the present paper intends to show.

Firstly, as will be seen in Sect. 2, the Fisher co-metric is defined via the variance/covariance of random variables based on a natural correspondence between cotangent vectors and random variables. This definition is very natural and does not seem arbitrary. There is no room for questions such as why log p appears in the definition of the Fisher metric.

Secondly, the above relationship between cotangent vectors and random variables directly links the variance/covariance of an unbiased estimator and the Fisher co-metric, which trivializes the Cramér-Rao inequality. Recognizing this fact, the Fisher metric appears to be a detour for the Cramér-Rao inequality, at least conceptually.

Thirdly, once we focus on the Fisher co-metric, we are motivated to reconsider a known result for the Fisher metric as a source of similar problems for the Fisher co-metric and the variance/covariance, which may lead to a new insight. As an example, co-metric and variance/covariance versions of Čencov's theorem on characterization of the Fisher metric are investigated in this paper.

The paper is organized as follows. In Sect. 2, we introduce the Fisher co-metric on the manifold $\mathcal{P}(\Omega)$, which is the totality of positive probability distributions on a finite set Ω, via the variance/covariance of random variables on Ω. In Sect. 3, the Fisher co-metric is shown to be equivalent to the Fisher metric by a natural correspondence. In Sect. 4, the Fisher metric and co-metric on an arbitrary submanifold of $\mathcal{P}$ are discussed, where we see that the Cramér-Rao inequality is trivialized by considering the co-metric. Section 5 treats the e- and m-connections on $\mathcal{P}(\Omega)$, where it is clarified that, in application to estimation theory, the role of the m-connection as a connection on the cotangent bundle and its relation to the Fisher co-metric are crucial. Sections 2–5 can be considered to constitute a first half of the paper, which is aimed at showing the naturalness and the usefulness of considering the Fisher co-metric.

The second half of the paper focuses on the monotonicity and the invariance of the Fisher metric and co-metric with respect to Markov maps. In Sect. 6, we investigate the monotonicity. We show there that the monotonicity of the Fisher metric, which is well known as a characteristic property of the metric, is equivalently translated into the monotonicity of the Fisher co-metric and that of the variance. In Sect. 7, after reviewing the invariance of the Fisher metric and Čencov's theorem, we consider their co-metric versions. It is shown that, being different from the monotonicity, the invariance of the metric and that of the co-metric are not logically equivalent. We present a theorem on characterization of the Fisher co-metric in terms of the invariance, which corresponds to Čencov's theorem but does not follow from it. The obtained theorem can also be expressed as a theorem on characterization of the variance/covariance. In Sect. 8, we investigate a stronger version of the invariance, which can be regarded as the joint condition that combines the invariance of the metric and that of the co-metric. The formulation used for expressing this condition is applied to affine connections in Sect. 9, whereby a kind of invariance condition for affine connection is obtained. The condition is shown to be equivalent to a known version of invariance condition which is seemingly weaker than the original condition used by Čencov to characterize the α-connections, but actually characterizes the α-connection as well. Section 10 is devoted to concluding remarks.

Remark 1.1 Throughout this paper, we denote the tangent space and the cotangent space of a manifold M at a point $p \in M$ by $T_p(M)$ and $T_p^*(M)$, respectively. We also denote the totality of smooth vector fields and that of smooth differential 1-forms on M by $\mathfrak{X}(M)$ and $\mathfrak{D}(M)$, respectively. We generally use capital letters $X, Y, \ldots$ for vector fields in $\mathfrak{X}(M)$, which are maps assigning tangent vectors $X_p, Y_p, \ldots$ in $T_p(M)$ to each point $p \in M$. To save the symbols, we also denote general tangent vectors in $T_p(M)$ by $X_p, Y_p, \ldots$, not only when they are the values of vector fields. Similarly, We use Greek letters $\alpha, \beta, \ldots$ for 1-forms in $\mathfrak{D}(M)$, which are maps assigning cotangent vectors $\alpha_p, \beta_p, \ldots$ in $T_p^*(M)$ to each point $p \in M$, and also denote general cotangent vectors by $\alpha_p, \beta_p, \ldots$, not only when they are the values of 1-forms. The pairing of $X_p \in T_p(M)$ and $\alpha_p \in T_p^*(M)$ is expressed as $\alpha_p(X_p)$, considering a cotangent vector as a function on the tangent space. We keep the first capital letters $A, B, \ldots$ for random variables ($\mathbb{R}$-valued functions on sample spaces).

2 The Fisher co-metric

We introduce the Fisher co-metric in this section, while its equivalence to the Fisher metric will be shown in the next section.

Let Ω be a finite set with cardinality $|\Omega| \geq 2$, and let $\mathcal{P}(\Omega)$ be the totality of strictly positive probability distributions on Ω:

$$\mathcal{P} = \mathcal{P}(\Omega) := \left\{ p \,\middle|\, p : \Omega \to (0, 1), \; \sum_{\omega \in \Omega} p(\omega) = 1 \right\}, \tag{2.1}$$

which is regarded as a manifold with $\dim \mathcal{P}(\Omega) = |\Omega| - 1$. Let the totality of $\mathbb{R}$-valued functions on Ω be denoted by $\mathbb{R}^\Omega$, and define $\left(\mathbb{R}^\Omega \right)_c := \{A \in \mathbb{R}^\Omega \mid \sum_{\omega \in \Omega} A(\omega) = c\}$ for a constant $c \in \mathbb{R}$. Since $\mathcal{P}$ is an open subset of the affine space $\left(\mathbb{R}^\Omega \right)_1$, its tangent space can be identified with the linear space $\left(\mathbb{R}^\Omega \right)_0$. Following the terminology of [1], we denote this identification $T_p(\mathcal{P}) \to \left(\mathbb{R}^\Omega \right)_0$ by $X_p \mapsto X_p^{(\mathrm{m})}$, and call $X_p^{(\mathrm{m})}$ the *m-representation* of X_p.

For an arbitrary submanifold M of $\mathcal{P}$ (including the case when $M = \mathcal{P}$) we define $T_p^{(\mathrm{m})}(M) := \{X_p^{(\mathrm{m})} \mid X_p \in T_p(M)\}$, which is a linear subspace of $T_p^{(\mathrm{m})}(\mathcal{P}) = \left(\mathbb{R}^\Omega \right)_0$. When the elements of M are parametrized as p_ξ by a coordinate system $\xi = (\xi^i)$ of M, the m-representation of $(\partial_i)_p \in T_p(M)$, where $\partial_i := \frac{\partial}{\partial \xi^i}$, with $p = p_\xi$ is represented as

$$(\partial_i)_p^{(\mathrm{m})} = \partial_i p_\xi, \tag{2.2}$$

and $\{(\partial_i)_p^{(\mathrm{m})}\}_{i=1}^n$ ($n = \dim M$) constitute a basis of $T_p^{(\mathrm{m})}(M)$.

We denote the expectation of a random variable $A \in \mathbb{R}^\Omega$ w.r.t. a distribution $p \in \mathcal{P}$ by

$$\langle A \rangle_p := \sum_{\omega \in \Omega} p(\omega) A(\omega), \tag{2.3}$$

and define the function

$$\langle A \rangle : \mathcal{P} \to \mathbb{R}, \quad p \mapsto \langle A \rangle_p. \tag{2.4}$$

Since $\langle A \rangle$ is a smooth function on the manifold $\mathcal{P}$, its differential $(d\langle A \rangle)_p \in T_p^*(\mathcal{P})$ at each point $p \in \mathcal{P}$ is defined. We introduce the following map:

$$\delta_p : \mathbb{R}^\Omega \to T_p^*(\mathcal{P}), \quad A \mapsto \delta_p(A) := (d\langle A \rangle)_p, \tag{2.5}$$

for which we have

$$\forall A \in \mathbb{R}^\Omega, \forall X_p \in T_p(\mathcal{P}), \quad \delta_p(A)(X_p) = X_p \langle A \rangle = \sum_{\omega \in \Omega} X_p^{(\mathrm{m})}(\omega) A(\omega). \tag{2.6}$$

Proposition 2.1 *For every $p \in \mathcal{P}$, the linear map $\delta_p : \mathbb{R}^\Omega \to T_p^*(\mathcal{P})$ is surjective with $\mathrm{Ker}\, \delta_p = \mathbb{R}$, where $\mathbb{R}$ is regarded as a subspace of $\mathbb{R}^\Omega$ by identifying a constant $c \in \mathbb{R}$ with the constant function $\omega \mapsto c$. Hence, δ_p induces a linear isomorphism $\mathbb{R}^\Omega / \mathbb{R} \to T_p^*(\mathcal{P})$.*

Proof Every cotangent vector $\alpha_p \in T_p^*(\mathcal{P})$ is a linear functional on $T_p(\mathcal{P})$, which is represented as $\alpha_p : X_p \mapsto \sum_\omega X_p^{(\mathrm{m})}(\omega) A(\omega)$ by some $A \in \mathbb{R}^\Omega$. This means that $\alpha_p = \delta_p(A)$ due to (2.6). Hence, δ_p is surjective. For any $A \in \mathbb{R}^\Omega$, we have

$$A \in \mathrm{Ker}\, \delta_p \iff \forall X_p \in T_p(\mathcal{P}), \quad \delta_p(A)(X_p) = \sum_\omega X_p^{(\mathrm{m})}(\omega) A(\omega) = 0$$

$$\iff \forall B \in \left(\mathbb{R}^\Omega \right)_0, \quad \sum_\omega A(\omega) B(\omega) = 0$$

$$\iff A \in \mathbb{R},$$

which proves $\mathrm{Ker}\, \delta_p = \mathbb{R}$. $\square$

For each $p \in \mathcal{P}$, denote the L^2 inner product and the covariance of random variables $A, B \in \mathbb{R}^\Omega$ by

$$\langle A, B \rangle_p := \langle AB \rangle_p \quad \text{and} \tag{2.7}$$
$$\mathrm{Cov}_p(A, B) := \langle A - \langle A \rangle_p, B - \langle B \rangle_p \rangle_p. \tag{2.8}$$

Then $\mathrm{Cov}_p : (A, B) \mapsto \mathrm{Cov}_p(A, B)$ is a degenerate nonnegative bilinear form on $\mathbb{R}^\Omega$ with kernel $\mathbb{R}$, and defines an inner product on $\mathbb{R}^\Omega / \mathbb{R}$. Therefore, Proposition 2.1 implies that an inner product on $T_p^*(\mathcal{P})$, which we denote by g_p, can be defined by

$$g_p(\delta_p(A), \delta_p(B)) = \mathrm{Cov}_p(A, B). \tag{2.9}$$

Denoting the norm for g_p by $\|\cdot\|_p$, we have

$$\|\delta_p(A)\|_p^2 = V_p(A), \tag{2.10}$$

where the RHS is the variance $V_p(A) := \langle (A - \langle A \rangle_p)^2 \rangle_p$.

We have thus defined the map g which maps each point $p \in \mathcal{P}$ to the inner product g_p on $T_p^*(\mathcal{P})$. We generally call such a map (a metric on the cotangent bundle) a *co-metric*. Although a co-metric is essentially equivalent to a usual (Riemannian) metric (a metric on the tangent bundle) by the correspondence explained in the next section, it is often useful to distinguish them conceptually. The co-metric defined by (2.9) is called the *Fisher co-metric*, since it corresponds to the Fisher metric as will be shown later.

Remark 2.2 Eq. (2.10) is found in Theorem 2.7 of the book [1], where the norm and the inner product on the cotangent space were considered to be induced from the Fisher metric.

3 The correspondence between a metric and a co-metric

By a standard argument of linear algebra, an inner product $\langle \cdot, \cdot \rangle$ on an $\mathbb{R}$-linear space V establishes a natural linear isomorphism between V and its dual space V^*, which we denote by $\xleftrightarrow{\langle \cdot, \cdot \rangle}$. This gives a one-to-one correspondence between a metric on a manifold M and a co-metric on M as follows. Given a metric g on M, a tangent vector $X_p \in T_p(M)$ and a cotangent vector $\alpha_p \in T_p^*(M)$ at a point $p \in M$ correspond each other by

$$X_p \xleftrightarrow{g_p} \alpha_p \iff \forall Y_p \in T_p(M), \; \alpha_p(Y_p) = g_p(X_p, Y_p). \tag{3.1}$$

The correspondence is extended to the correspondence between a vector field $X \in \mathfrak{X}(M)$ and a 1-form $\alpha \in \mathfrak{D}(M)$ by

$$X \xleftrightarrow{g} \alpha \iff \forall p \in M, \; X_p \xleftrightarrow{g_p} \alpha_p. \tag{3.2}$$

(Note: some literature refers to this correspondence as the musical isomorphism with notation $\alpha = X^\flat$ and $X = \alpha^\sharp$, while we will use the symbol $\sharp$ for a different meaning later.) This correspondence determines a co-metric on M, which is denoted by the same symbol g, such that for every $p \in M$

$$X_p \xleftrightarrow{g_p} \alpha_p \text{ and } Y_p \xleftrightarrow{g_p} \beta_p \implies g_p(\alpha_p, \beta_p) = g_p(X_p, Y_p). \tag{3.3}$$

Conversely, given a co-metric g on M, the correspondence $\xleftrightarrow{g_p}$ is defined by

$$X_p \xleftrightarrow{g_p} \alpha_p \iff \forall \beta_p \in T_p^*(M), \; \beta_p(X_p) = g_p(\alpha_p, \beta_p), \tag{3.4}$$

and a metric on M is defined by the same relation as (3.3). It should be noted that when a metric and a co-metric correspond in this way, the relations (3.1) and (3.4) are equivalent, so that there arises no confusion even if we use the same symbol g for the corresponding metric and co-metric in $\xleftrightarrow{g_p}$ and $\xleftrightarrow{g}$.

Note that for an arbitrary coordinate system (ξ^i) of M, $g_{ij} := g(\frac{\partial}{\partial \xi^i}, \frac{\partial}{\partial \xi^j})$ and $g^{ij} := g(d\xi^i, d\xi^j)$ form the inverse matrices of each other at every point of M. Note also that the norms for $(T_p(M), g_p)$ and $(T_p^*(M), g_p)$ are linked by

$$\|X_p\|_p = \max_{\alpha_p \in T_p^*(M) \setminus \{0\}} \frac{|\alpha_p(X_p)|}{\|\alpha_p\|_p} \tag{3.5}$$

$$\text{and} \quad \|\alpha_p\|_p = \max_{X_p \in T_p(M) \setminus \{0\}} \frac{|\alpha_p(X_p)|}{\|X_p\|_p}, \tag{3.6}$$

where the max's in these equations are achieved by those X_p and α_p which correspond to each other by $\xleftrightarrow{g_p}$ up to a constant factor.

For a tangent vector $X_p \in T_p(\mathcal{P})$, define

$$L_{X_p} := X_p^{(\mathrm{m})}/p \in \{A \in \mathbb{R}^\Omega \mid \langle A \rangle_p = 0\}, \tag{3.7}$$

which is the derivative of the map $\mathcal{P} \to \mathbb{R}^\Omega$, $p \mapsto \log p$ w.r.t. X_p. (In [1], L_{X_p} is called the *e-representation* of X_p and is denoted by $X_p^{(\mathrm{e})}$.) Note that L_{X_p} is characterized by (cf. (2.6))

$$\forall A \in \mathbb{R}^\Omega, \ \delta_p(A)(X_p) = X_p \langle A \rangle = \langle L_{X_p}, A \rangle_p. \tag{3.8}$$

The following proposition shows that the metric induced from the Fisher co-metric g by the correspondence $\xleftrightarrow{g}$ is the Fisher metric.

Proposition 3.1 *For each point $p \in \mathcal{P}$, we have:*

1. *$\forall A \in \mathbb{R}^\Omega, \ \forall X_p \in T_p(\mathcal{P}), \ X_p \xleftrightarrow{g_p} \delta_p(A) \Leftrightarrow L_{X_p} = A - \langle A \rangle_p.$*
2. *$\forall X_p, Y_p \in T_p(\mathcal{P}), \ g_p(X_p, Y_p) = \langle L_{X_p}, L_{Y_p} \rangle_p.$*

Proof $\underline{1}$: According to (3.4), the condition $X_p \xleftrightarrow{g_p} \delta_p(A)$ is equivalent to

$$\forall B \in \mathbb{R}^\Omega, \ g_p(\delta_p(A), \delta_p(B)) = \delta_p(B)(X_p). \tag{3.9}$$

Here the LHS is equal to

$$\langle A - \langle A \rangle_p, B - \langle B \rangle_p \rangle_p = \langle A - \langle A \rangle_p, B \rangle_p,$$

while the RHS is equal to $\langle L_{X_p}, B \rangle_p$ by (3.8). Hence, (3.9) is equivalent to $L_{X_p} = A - \langle A \rangle_p$.

$\underline{2}$: Obvious from item 1 and (3.3). $\qquad\square$

4 The Fisher co-metric on a submanifold and the Cramér–Rao inequality

Let M be an arbitrary submanifold of $\mathcal{P}$. Then a metric on M is induced as the restriction of the Fisher metric g, which we denote by $g_M : p \mapsto g_{M,p} = g_p|_{T_p(M)^2}$. When a coordinate system $\xi = (\xi^i)$ is given on M, corresponding to (2.2) it holds that

$$L_{(\partial_i)_p} = \partial_i \log p_\xi \quad \text{at} \quad p = p_\xi. \tag{4.1}$$

We have

$$g_{M,ij}(p) := g_{M,p}((\partial_i)_p, (\partial_j)_p) = \langle \partial_i \log p_\xi, \partial_j \log p_\xi \rangle_p, \tag{4.2}$$

which defines the Fisher information matrix $G_M(p) = [g_{M,ij}(p)]$. The metric g_M induces a co-metric on M, which is denoted by the same symbol g_M. Letting

$$g_M^{ij}(p) := g_{M,p}((d\xi^i)_p, (d\xi^j)_p), \tag{4.3}$$

we have $G_M(p)^{-1} = [g_M^{ij}(p)]$.

Suppose that a cotangent vector $\alpha_p \in T_p^*(M)$ on M is the restriction of a cotangent vector $\tilde{\alpha}_p \in T_p^*(\mathcal{P})$ on $\mathcal{P}$; i.e., $\alpha_p = \tilde{\alpha}_p|_{T_p(M)}$. Then, it follows from (3.6) that

$$\begin{aligned}
\|\alpha_p\|_{M,p} &= \max_{X_p \in T_p(M)\setminus\{0\}} \frac{|\alpha_p(X_p)|}{\|X_p\|_{M,p}} \\
&= \max_{X_p \in T_p(M)\setminus\{0\}} \frac{|\tilde{\alpha}_p(X_p)|}{\|X_p\|_p} \\
&\leq \max_{X_p \in T_p(\mathcal{P})\setminus\{0\}} \frac{|\tilde{\alpha}_p(X_p)|}{\|X_p\|_p} = \|\tilde{\alpha}_p\|_p.
\end{aligned} \tag{4.4}$$

(Note that $\|X_p\|_p = \|X_p\|_{M,p}$ since the metric on M is the restriction of the metric on $\mathcal{P}$.) Furthermore, for an arbitrary $\alpha_p \in T_p^*(M)$, there always exists $\tilde{\alpha}_p \in T_p^*(\mathcal{P})$ satisfying $\alpha_p = \tilde{\alpha}_p|_{T_p(M)}$ and $\|\alpha_p\|_{M,p} = \|\tilde{\alpha}_p\|_p$. Indeed, letting $X_p \in T_p(M)$ be defined by $X_p \xleftrightarrow{g_{M,p}} \alpha_p$, such an $\tilde{\alpha}_p$ is obtained by $X_p \xleftrightarrow{g_p} \tilde{\alpha}_p$.

The above observations lead to the following proposition.

Proposition 4.1 *1. For any $\alpha_p \in T_p^*(M)$, we have*

$$\begin{aligned}
\|\alpha_p\|_{M,p} &= \min\{\|\tilde{\alpha}_p\|_p \mid \tilde{\alpha}_p \in T_p^*(\mathcal{P}) \ \text{and} \ \alpha_p = \tilde{\alpha}_p|_{T_p(M)}\} \\
&= \|(\alpha_p)^\sharp\|_p,
\end{aligned} \tag{4.5}$$

where $(\alpha_p)^\sharp := \arg\min_{\tilde{\alpha}_p}\{\|\tilde{\alpha}_p\|_p \mid \cdots\}$.

2. *For any $\alpha_p, \beta_p \in T_p^*(M)$, we have*

$$g_{M,p}(\alpha_p, \beta_p) = g_p((\alpha_p)^\sharp, (\beta_p)^\sharp). \tag{4.6}$$

The above proposition shows that the Fisher co-metric on M can be defined from the Fisher co-metric on $\mathcal{P}$ directly by (4.5) and (4.6), not by way of the Fisher metric.

Corollary 4.2 *(The Cramér-Rao inequality) Suppose that an n-tuple of random variables $\vec{A} = (A^1, \ldots, A^n) \in (\mathbb{R}^\Omega)^n$ satisfies*

$$\forall i \in \{1, \ldots, n\}, \quad (d\xi^i)_p = \delta_p(A^i)|_{T_p(M)} \tag{4.7}$$

for a coordinate system $\xi = (\xi^i)$ of an n-dimensional submanifold M of $\mathcal{P}$ and for a point $p \in M$. Letting $V_p(\vec{A}) = [v^{ij}] \in \mathbb{R}^{n \times n}$ be the variance-covariance matrix of $\vec{A}$ defined by

$$v^{ij} := \mathrm{Cov}_p(A^i, A^j) \tag{4.8}$$

and letting $G_M(p)$ be the Fisher information matrix, we have

$$V_p(\vec{A}) \geq G_M(p)^{-1}. \tag{4.9}$$

Proof For an arbitrary column vector $c = (c_i) \in \mathbb{R}^n$, let

$$\tilde{\alpha}_p := \sum_i c_i \delta_p(A^i) \in T_p^*(\mathcal{P}),$$

$$\alpha_p := \sum_i c_i (d\xi^i)_p \in T_p^*(M).$$

Since (4.7) implies that $\tilde{\alpha}_p|_{T_p(M)} = \alpha_p$, it follows from Proposition 4.1 that $\|\tilde{\alpha}_p\|_p \geq \|\alpha_p\|_{M,p}$. Noting that $\|\tilde{\alpha}_p\|_p^2 = {}^t c \, V_p(\vec{A}) \, c$ and $\|\alpha_p\|_{M,p}^2 = {}^t c \, G_M(p)^{-1} c$, where ^{t} denotes the transpose, we obtain (4.9). $\qquad\square$

5 On the e, m-connections

An affine connection is usually treated as a connection on the tangent bundle, while it corresponds to a connection on the cotangent bundle by the relation

$$\forall X, Y \in \mathfrak{X}(M), \ \forall \alpha \in \mathfrak{D}(M), \quad X\alpha(Y) = \alpha(\nabla_X Y) + (\nabla_X \alpha)(Y). \tag{5.1}$$

This correspondence is one-to-one, so that we can define an affine connection by specifying a connection on the cotangent bundle. Therefore, the α-connection in information geometry can also be introduced in this way. Although affine connections are out of the main subject of this paper, we will briefly discuss the significance of defining the

m-connection (i.e. ($\alpha = -1$)-connection) in this way, since it is closely related to the role of the Fisher co-metric in the Cramér-Rao inequality.

We start by introducing the m-connection $\nabla^{(m)}$ on $\mathcal{P} = \mathcal{P}(\Omega)$ as a flat connection on the cotangent bundle for which the 1-form $d\langle A \rangle$ is parallel for any $A \in \mathbb{R}^{\Omega}$; i.e.,

$$\forall X \in \mathfrak{X}(\mathcal{P}), \forall A \in \mathbb{R}^{\Omega}, \ \nabla_X^{(m)} d\langle A \rangle = 0. \tag{5.2}$$

Since $\dim\{d\langle A \rangle \mid A \in \mathbb{R}^{\Omega}\} = |\Omega| - 1 = \dim \mathcal{P}$, (5.2) implies that every parallel 1-from is represented as $d\langle A \rangle$ by some $A \in \mathbb{R}^{\Omega}$. Then the correspondence (5.1) determines a connection on the tangent bundle, which is denoted by the same symbol $\nabla^{(m)}$. Letting $\alpha = d\langle A \rangle$ in (5.1) and applying (5.2), we have

$$\forall X, Y \in \mathfrak{X}(\mathcal{P}), \forall A \in \mathbb{R}^{\Omega}, \ XY\langle A \rangle = (\nabla_X^{(m)} Y)\langle A \rangle. \tag{5.3}$$

This implies that, for any $Y \in \mathfrak{X}(\mathcal{P})$,

Y is m-parallel

$\Leftrightarrow \ \forall A \in \mathbb{R}^{\Omega}, \forall X \in \mathfrak{X}(\mathcal{P}), \ XY\langle A \rangle = 0$

$\Leftrightarrow \ \forall A \in \mathbb{R}^{\Omega}, \ Y_p\langle A \rangle = \sum_{\omega} Y_p^{(m)}(\omega) A(\omega)$ does not depend on $p \in \mathcal{P}$

$\Leftrightarrow \ Y_p^{(m)}$ does not depend on $p \in \mathcal{P}$, $\tag{5.4}$

where "m-parallel" means "parallel w.r.t. $\nabla^{(m)}$". Since this property characterizes the m-connection on $\mathcal{P}$ (e.g. Equation (2.39) of [1]), our definition of the m-connection is equivalent to the usual definition in information geometry.

Next, we define the e-connection $\nabla^{(e)}$ as the dual connection of $\nabla^{(m)}$ w.r.t. the Fisher metric g ([1], [2]), which means that

$$\forall X, Y, Z \in \mathfrak{X}(\mathcal{P}), \ Zg(X, Y) = g(\nabla_Z^{(e)} X, Y) + g(X, \nabla_Z^{(m)} Y). \tag{5.5}$$

Using (5.1), we can rewrite (5.5) into

$$\forall X, Y, Z \in \mathfrak{X}(\mathcal{P}), \forall \alpha \in \mathfrak{D}(\mathcal{P}),$$
$$X \xleftrightarrow{g} \alpha \ \Rightarrow \ (\nabla_Z^{(m)} \alpha)(Y) = g(\nabla_Z^{(e)} X, Y). \tag{5.6}$$

This implies that, for any $X \in \mathfrak{X}(\mathcal{P})$ and $\alpha \in \mathfrak{D}(\mathcal{P})$,

$$X \xleftrightarrow{g} \alpha \ \Rightarrow \ \Big[X \text{ is e-parallel} \ \Leftrightarrow \ \alpha \text{ is m-parallel} \Big]. \tag{5.7}$$

Now, let us recall the situation of Corollary 4.2. An estimator $\vec{A} = (A^1, \ldots, A^n)$ is said to be *efficient* for the statistical model (M, ξ) when it is unbiased (i.e. $\forall i$, $\xi^i = \langle A^i \rangle|_M$) and achieves the equality in the Cramér-Rao inequality (4.9) for every $p \in M$. Noting that the achievability at each $p \in M$ is represented by the condition

$\forall i, \delta_p(A^i) = (d\langle A^i\rangle)_p = ((d\xi^i)_p)^\sharp$ and recalling (5.2), we can see that the condition for (M, ξ) to have an efficient estimator is expressed as

$$\forall i, \ \exists \tilde{\alpha}^i \in \mathfrak{D}(\mathcal{P}), \ \ \tilde{\alpha}^i \text{ is m-parallel and } \ \forall p \in M, \ \tilde{\alpha}^i_p = ((d\xi^i)_p)^\sharp. \tag{5.8}$$

On the other hand, it is well known that the existence of an efficient estimator is equivalent to the condition that M is an exponential family and that ξ is an expectation coordinate system, which can be rephrased as (see Theorem 3.12 of [1])

$$M \text{ is an e-autoparallel submanifold of } \mathcal{P},$$
$$\text{and } \xi \text{ is an m-affine coordinate system.} \tag{5.9}$$

Therefore, the two conditions (5.8) and (5.9) are necessarily equivalent. These are both purely geometrical conditions for a submanifold of the dually flat space $(\mathcal{P}, g, \nabla^{(e)}, \nabla^{(m)})$, and we can prove their equivalence within this geometrical framework, forgetting its statistical background. Indeed, the equivalence can be proved for a more general situation where M is a submanifold of a manifold S equipped with a Riemannian metric g and a pair of dual affine connections ∇, ∇^* on the assumption that ∇^* is flat. Note that this assumption is weaker than the dually-flatness of (S, g, ∇, ∇^*) in that ∇ is allowed to have non-vanishing torsion, which is essential in application to quantum estimation theory. See section 7 of [4] for details.

6 Monotonicity

The monotonicity with respect to a Markov map is known to be an important and characteristic property of the Fisher metric. In this section we discuss the monotonicity of the Fisher co-metric and its relation to the variance of random variables.

Let Ω_1 and Ω_2 be arbitrary finite sets, and let $\mathcal{P}_i := \mathcal{P}(\Omega_i)$ for $i = 1, 2$. A map $\Phi : \mathcal{P}_1 \to \mathcal{P}_2$ is called a *Markov map* when it is affine in the sense that $\forall p, q \in \mathcal{P}_1$, $0 \le \forall a \le 1, \Phi(ap + (1-a)q) = a\,\Phi(p) + (1-a)\,\Phi(q)$. Every Markov map Φ is represented as

$$\forall p \in \mathcal{P}_1, \ \ \Phi(p) = \sum_{x \in \Omega_1} W(\cdot \,|\, x) p(x), \tag{6.1}$$

where W is a surjective channel from Ω_1 to Ω_2; i.e.,

$$\forall (x, y) \in \Omega_1 \times \Omega_2, \ W(y\,|\,x) \ge 0, \ \ \forall x \in \Omega_1, \ \sum_{y \in \Omega_2} W(y\,|\,x) = 1, \tag{6.2}$$

$$\text{and} \ \ \forall y \in \Omega_2, \ \exists x \in \Omega_1, \ W(y\,|\,x) > 0. \tag{6.3}$$

When Φ is represented as (6.1), we write $\Phi = \Phi_W$.

More generally, for a submanifold M of $\mathcal{P}_1$ and a submanifold N of $\mathcal{P}_2$, a map $\varphi : M \to N$ is called a *Markov map* when there exists a Markov map $\Phi : \mathcal{P}_1 \to \mathcal{P}_2$

such that $\varphi = \Phi|_M$. Since a Markov map φ is smooth, it induces at each $p \in M$ the differential

$$\varphi_* = \varphi_{*,p} = (d\varphi)_p : T_p(M) \to T_{\varphi(p)}(N) \tag{6.4}$$

and its dual

$$\varphi^* = \varphi_p^* = {}^{\mathrm{t}}(d\varphi)_p : T_{\varphi(p)}^*(N) \to T_p^*(M), \tag{6.5}$$

where $^{\mathrm{t}}$ denotes the transpose of a linear map. See Remark 6.2 below for the notation $\varphi^* = \varphi_p^*$.

As is well known, the Fisher metric satisfies the following monotonicity property for its norm:

$$\forall p \in M, \forall X_p \in T_p(M), \quad \|\varphi_*(X_p)\|_{\varphi(p)} \le \|X_p\|_p. \tag{6.6}$$

The cotangent version of the monotonicity is given below.

Proposition 6.1 *We have*

$$\forall p \in M, \forall \alpha_{\varphi(p)} \in T_{\varphi(p)}^*(N), \quad \|\varphi^*(\alpha_{\varphi(p)})\|_{M,p} \le \|\alpha_{\varphi(p)}\|_{N,\varphi(p)}, \tag{6.7}$$

where $\|\cdot\|_{M,p}$ and $\|\cdot\|_{N,\varphi(p)}$ denote the norms w.r.t. the Fisher co-metrics g_M and g_N, respectively.

Proof Since the inequality is trivial when $\|\varphi^*(\alpha_{\varphi(p)})\|_{M,p} = 0$, we assume $\|\varphi^*(\alpha_{\varphi(p)})\|_{M,p} > 0$. Then, invoking (3.6), we have

$$
\begin{aligned}
\|\varphi^*(\alpha_{\varphi(p)})\|_{M,p} &= \max_{X_p \in T_p(M) \setminus \{0\}} \frac{|\varphi^*(\alpha_{\varphi(p)})(X_p)|}{\|X_p\|_{M,p}} \\
&= \max_{X_p \in T_p(M) \setminus \{0\}} \frac{|\alpha_{\varphi(p)}(\varphi_*(X_p))|}{\|X_p\|_p} \\
&= \max_{X_p \in T_p(M) : \varphi_*(X_p) \neq 0} \frac{|\alpha_{\varphi(p)}(\varphi_*(X_p))|}{\|X_p\|_p} \\
&\le \sup_{X_p \in T_p(M) : \varphi_*(X_p) \neq 0} \frac{|\alpha_{\varphi(p)}(\varphi_*(X_p))|}{\|\varphi_*(X_p)\|_{\varphi(p)}} \\
&\le \max_{Y_{\varphi(p)} \in T_{\varphi(p)}(N) \setminus \{0\}} \frac{|\alpha_{\varphi(p)}(Y_{\varphi(p)})|}{\|Y_{\varphi(p)}\|_{\varphi(p)}} = \|\alpha_{\varphi(p)}\|_{N,\varphi(p)},
\end{aligned}
$$

where the third equality follows since X_p achieving $\max_{X_p \in T_p(M) \setminus \{0\}}$ should satisfy $\varphi_*(X_p) \neq 0$ due to the assumption $\|\varphi^*(\alpha_{\varphi(p)})\|_{M,p} > 0$, and the first $\le$ follows from (6.6). $\quad\square$

Remark 6.2 We have written $\varphi^*(\alpha_{\varphi(p)})$ for $\varphi_p^*(\alpha_{\varphi(p)})$ above (and will use similar notations throughout the paper), considering that omitting p from φ_p^* is harmless in

the context and is better for the readability of expressions. Note that the notation $\varphi^*(\alpha_{\varphi(p)})$, if it appears alone, is mathematically ambiguous unless $\varphi^{-1}(\varphi(p))$ is the singleton $\{p\}$ in that $\varphi_{p'}^*(\alpha_{\varphi(p)}) \in T_{p'}^*(M)$ depends on a choice of $p' \in \varphi^{-1}(\varphi(p))$. On the other hand, the notation $\varphi_*(X_p)$ has no such ambiguity, since we know that the argument X_p belongs to $T_p(M)$ and hence φ_* must be $\varphi_{*,p}$.

Let us consider the case when $M = \mathcal{P}_1$ and $N = \mathcal{P}_2$, and let $\Phi = \Phi_W : \mathcal{P}_1 \to \mathcal{P}_2$ be an arbitrary Markov map represented by a surjective channel W. Recalling (6.1) and the definition of m-representation of tangent vectors, we have

$$Y_{\Phi(p)} = \Phi_*(X_p) \Leftrightarrow Y_{\Phi(p)}^{(m)} = \sum_{x \in \Omega_1} W(\cdot \mid x) X_p^{(m)}(x) \tag{6.8}$$

for $X_p \in T_p(\mathcal{P}_1)$ and $Y_{\Phi(p)} \in T_{\Phi(p)}(\mathcal{P}_2)$. We claim that

$$\forall A \in \mathbb{R}^{\Omega_2}, \quad \Phi^*(\delta_{\Phi(p)}(A)) = \delta_p(E_W(A \mid \cdot)), \tag{6.9}$$

where $\Phi^* = \Phi_p^*$, and $E_W(A \mid \cdot) \in \mathbb{R}^{\Omega_1}$ denotes the conditional expectation of A defined by

$$\forall x \in \Omega_1, \quad E_W(A \mid x) = \sum_{y \in \Omega_2} W(y \mid x) A(y). \tag{6.10}$$

Eq. (6.9) is verified as follows; for every $\beta_p = \delta_p(B) \in T_p^*(\mathcal{P}_1)$, where $B \in \mathbb{R}^{\Omega_2}$, we have

$$\begin{aligned}
&\beta_p = \Phi^*(\delta_{\Phi(p)}(A)) \\
&\Leftrightarrow \forall X_p \in T_p(\mathcal{P}_1), \quad \beta_p(X_p) = \delta_{\Phi(p)}(A)(\Phi_*(X_p)) \\
&\Leftrightarrow \forall X_p \in T_p(\mathcal{P}_1), \quad \sum_{x \in \Omega_1} X_p^{(m)}(x) B(x) = \sum_{(x,y) \in \Omega_1 \times \Omega_2} W(y \mid x) X_p^{(m)}(x) A(y) \\
&\Leftrightarrow B - E_W(A \mid \cdot) \in \mathbb{R} \\
&\Leftrightarrow \beta_p = \delta_p(E_W(A \mid \cdot)),
\end{aligned} \tag{6.11}$$

where the second $\Leftrightarrow$ follows from (2.6) and (6.8), the third $\Leftrightarrow$ follows from $T_p^{(m)}(\mathcal{P}) = (\mathbb{R}^\Omega)_0$, and $\mathbb{R}$ is identified with the set of constant functions on Ω_1.

Invoking (2.10) and (6.9), we see that the monotonicity (6.7) is equivalent to the following well-known inequality for the variance:

$$\forall A \in \mathbb{R}^{\Omega_2}, \quad V_p(E_W(A \mid \cdot)) \leq V_{\Phi(p)}(A), \tag{6.12}$$

which we refer to as the *monotonicity of the variance*.

In the above proof of Proposition 6.1, we derived (6.7) from (6.6). Conversely, we can derive (6.6) from (6.7) by the use of (3.5) as follows; for any $X_p \in T_p(M)$,

$$
\begin{aligned}
\|\varphi_*(X_p)\|_{\varphi(p)} &= \max_{\alpha_{\varphi(p)} \in T^*_{\varphi(p)}(N) \setminus \{0\}} \frac{|\alpha_{\varphi(p)}(\varphi_*(X_p))|}{\|\alpha_{\varphi(p)}\|_{N,\varphi(p)}} \\
&= \max_{\alpha_{\varphi(p)} \in T^*_{\varphi(p)}(N) \setminus \{0\}} \frac{|\varphi^*(\alpha_{\varphi(p)})(X_p)|}{\|\alpha_{\varphi(p)}\|_{N,\varphi(p)}} \\
&= \max_{\alpha_{\varphi(p)} \in T^*_{\varphi(p)}(N) : \varphi^*(\alpha_{\varphi(p)}) \neq 0} \frac{|\varphi^*(\alpha_{\varphi(p)})(X_p)|}{\|\alpha_{\varphi(p)}\|_{N,\varphi(p)}} \\
&\leq \sup_{\alpha_{\varphi(p)} \in T^*_{\varphi(p)}(N) : \varphi^*(\alpha_{\varphi(p)}) \neq 0} \frac{|\varphi^*(\alpha_{\varphi(p)})(X_p)|}{\|\varphi^*(\alpha_{\varphi(p)})\|_{M,p}} \\
&\leq \max_{\beta_p \in T^*_p(M) \setminus \{0\}} \frac{|\beta_p(X_p)|}{\|\beta_p\|_{M,p}} = \|X_p\|_p, \quad (6.13)
\end{aligned}
$$

where we have assumed $\|\varphi_*(X_p)\|_{\varphi(p)} > 0$ with no loss of generality, which yields the third equality (cf. the proof of Proposition 6.1), and the first $\leq$ follows from (6.7). Thus, (6.6) and (6.7) are equivalent. Note that this equivalence is derived solely from a general argument on metrics and co-metrics, and does not rely on the special characteristics of the Fisher metric/co-metric. In this sense, we say that (6.6) and (6.7) are *logically equivalent*.

Recalling that the Fisher metric is characterized as the unique monotone metric up to a constant factor, we obtain the following propositions from the logical equivalence mentioned above.

Proposition 6.3 *The monotonicity (6.7) characterizes the Fisher co-metric up to a constant factor.*

Proposition 6.4 *The variance is characterized up to a constant factor as the positive quadratic form for random variables satisfying the monotonicity (6.12).*

Remark 6.5 We have described the above propositions in a rough form for the sake of readability. For the exact statement, we need a formulation similar to Theorems 7.1, 7.2 and 7.3 in the next section. See also Remark 8.4.

Remark 6.6 Since the monotonicity of the Fisher metric (6.6), that of the Fisher co-metric (6.7), and that of the variance (6.12) are all logically equivalent, we can derive (6.6) from the more popular (6.12).

7 Invariance

Čencov showed in [3] that the Fisher metric is characterized up to a constant factor as a covariant tensor field of degree 2 satisfying the invariance for Markov embeddings. Note that the invariance is weaker than the monotonicity and that the tensor field is not assumed to be positive nor symmetric. In this section we review Čencov's theorem

and then investigate its co-metric version, which will be shown to be equivalent to a theorem characterizing the variance/covariance of random variables.

We begin by reviewing the invariance property of the Fisher metric. Suppose that M and N are arbitrary submanifolds of $\mathcal{P}_1 = \mathcal{P}(\Omega_1)$ and $\mathcal{P}_2 = \mathcal{P}(\Omega_2)$, respectively, and that a pair of Markov maps

$$\varphi : M \to N \quad \text{and} \quad \psi : N \to M \tag{7.1}$$

satisfis

$$\psi \circ \varphi = id_M, \tag{7.2}$$

where $\circ$ denotes the composition of maps. Note that φ is injective while ψ is surjective. Given a pair of points $(p, q) \in M \times N$ satisfying

$$q = \varphi(p) \quad \text{and} \quad p = \psi(q), \tag{7.3}$$

we have

$$\psi_{*,q} \circ \varphi_{*,p} = id_{T_p(M)}. \tag{7.4}$$

It then follows from the monotonicity (6.6) that

$$\|X_p\|_{M,p} \geq \|\varphi_*(X_p)\|_{N,q} \geq \|\psi_*(\varphi_*(X_p))\|_{M,p} = \|X_p\|_{M,p}, \tag{7.5}$$

so that we have the invariance of the Fisher metric

$$\forall X_p \in T_p(M), \quad \|X_p\|_{M,p} = \|\varphi_*(X_p)\|_{N,q}, \tag{7.6}$$

which is equivalent to

$$\forall X_p, \forall Y_p \in T_p(M), \quad g_{M,p}(X_p, Y_p) = g_{N,q}(\varphi_*(X_p), \varphi_*(Y_p)). \tag{7.7}$$

This means that $\varphi_{*,p} : T_p(M) \to T_q(N)$ is isometry, which is represented as

$$(\varphi_{*,p})^{\dagger} \circ \varphi_{*,p} = id_{T_p(M)}, \tag{7.8}$$

where $(\varphi_{*,p})^{\dagger} : T_q(N) \to T_p(M)$ denotes the adjoint (Hermitian conjugate) of $\varphi_{*,p}$ w.r.t. the inner products $g_{M,p}$ and $g_{N,q}$.

A Markov map $\Phi : \mathcal{P}_1 \to \mathcal{P}_2$ is called an *Markov embedding* when there exists a Markov map $\Psi : \mathcal{P}_2 \to \mathcal{P}_1$ such that

$$\Psi \circ \Phi = id_{\mathcal{P}_1}. \tag{7.9}$$

Note that $|\Omega_1| \le |\Omega_2|$ necessarily holds in this case. As a special case of the invariance (7.7), we have

$$\forall p \in \mathcal{P}_1, \forall X_p, \forall Y_p \in T_p(\mathcal{P}_1), \quad g_p(X_p, Y_p) = g_{\Phi(p)}(\Phi_*(X_p), \Phi_*(Y_p)). \tag{7.10}$$

According to Čencov, this property characterizes the Fisher metric up to a constant factor. The exact statement is presented below.

Theorem 7.1 *(Čencov [3]) For $n = 2, 3, \ldots$, let $\Omega_n := \{1, 2, \ldots, n\}$ and $\mathcal{P}_n := \mathcal{P}(\Omega_n)$, and let g_n be the Fisher metric on $\mathcal{P}_n$. Suppose that we are given a sequence $\{h_n\}_{n=2}^{\infty}$, where h_n is a covariant tensor field of degree 2 on $\mathcal{P}_n$ which continuously maps each point $p \in \mathcal{P}_n$ to a bilinear form $h_{n,p} : T_p(\mathcal{P}_n)^2 \to \mathbb{R}$. Then the following two conditions are equivalent.*

(i) $\exists c \in \mathbb{R}, \forall n, \quad h_n = c g_n$.
(ii) For any $m \le n$ and any Markov embedding $\Phi : \mathcal{P}_m \to \mathcal{P}_n$, it holds that

$$\forall p \in \mathcal{P}_m, \forall X_p, \forall Y_p \in T_p(\mathcal{P}_m),$$
$$h_{m,p}(X_p, Y_p) = h_{n,\Phi(p)}(\Phi_*(X_p), \Phi_*(Y_p)). \tag{7.11}$$

We now proceed to the invariance property of co-metrics. Let us consider the same situation as (7.1)-(7.3), which implies that

$$\varphi_p^* \circ \psi_q^* = id_{T_p^*(M)}. \tag{7.12}$$

Then it follows from the monotonicity (6.7) that, for any $\alpha_p \in T_p^*(M)$,

$$\|\alpha_p\|_{M,p} \ge \|\psi^*(\alpha_p)\|_{N,q} \ge \|\varphi^*(\psi^*(\alpha_p))\|_{M,p} = \|\alpha_p\|_{M,p}, \tag{7.13}$$

so that we have the invariance of the Fisher co-metric

$$\forall \alpha_p \in T_p^*(M), \quad \|\alpha_p\|_{M,p} = \|\psi^*(\alpha_p)\|_{N,q}, \tag{7.14}$$

which is equivalent to

$$\forall \alpha_p, \beta_p \in T_p^*(M), \quad g_{M,p}(\alpha_p, \beta_p) = g_{N,q}(\psi^*(\alpha_p), \psi^*(\beta_p)) \tag{7.15}$$

This means that $\psi_q^* : T_p^*(M) \to T_q^*(N)$ is isometry, which is represented as

$$(\psi_q^*)^\dagger \circ \psi_q^* = id_{T_p^*(M)}, \tag{7.16}$$

where $(\psi_q^*)^\dagger : T_q^*(N) \to T_p^*(M)$ denotes the adjoint (Hermitian conjugate) of ψ_q^* w.r.t. the inner products $g_{M,p}$ and $g_{N,q}$ on the cotangent spaces. Due to $\psi_q^* = {}^t(\psi_{*,q})$, (7.16) can be rewritten as

$$\psi_{*,q} \circ (\psi_{*,q})^\dagger = id_{T_p(M)}. \tag{7.17}$$

We observed in Sect. 6 that the monotonicity of metrics (6.6) and that of co-metrics (6.7) are logically equivalent. On the other hand, such an equivalence does not hold for the invariance properties (7.6) and (7.14). Indeed, a linear-algebraic consideration shows that the implications $(7.4) \wedge (7.8) \Rightarrow (7.17)$ and $(7.4) \wedge (7.17) \Rightarrow (7.8)$ do not hold unless $\varphi_{*,p}$ is a linear isomorphism (cf. Lemma 8.2). This means that we cannot expect Čencov's theorem to yield a corollary which states that the Fisher co-metric is characterized by the invariance (7.14). Nevertheless, the statement itself is true as explained below.

Let us return to the situation of (7.9). We call a Markov map $\Psi : \mathcal{P}_2 \to \mathcal{P}_1$ a *Markov co-embedding* when there exists a Markov embedding $\Phi : \mathcal{P}_1 \to \mathcal{P}_2$ satisfying (7.9). As an example of (7.14), (7.9) implies the invariance

$$\forall p \in \mathcal{P}_1, \forall \alpha_p \in T_p^*(\mathcal{P}_1), \quad \|\alpha_p\|_p = \|\Psi^*(\alpha_p)\|_{\Phi(p)}, \qquad (7.18)$$

which can be rewritten as

$$\forall q \in \Phi(\mathcal{P}_1), \forall \alpha_{\Psi(q)} \in T_{\Psi(q)}^*(\mathcal{P}_1), \quad \|\alpha_{\Psi(q)}\|_{\Psi(q)} = \|\Psi^*(\alpha_{\Psi(q)})\|_q. \quad (7.19)$$

Actually, the range $\forall q \in \Phi(\mathcal{P}_1)$ in the above equation can be extended to $\forall q \in \mathcal{P}_2$ for the reason described below.

It is known (e.g. Lemma 9.5 of [3]) that every pair (Φ, Ψ) of Markov embedding and co-embedding satisfying (7.9) is represented in the following form:

$$\forall q \in \mathcal{P}_2, \quad \Psi(q) = q^F \quad \text{with} \quad q^F(x) := \sum_{y \in F^{-1}(x)} q(y), \qquad (7.20)$$

and

$$\forall p \in \mathcal{P}_1, \quad \Phi(p) = \sum_{x \in \Omega_1} p(x) \, r_x, \qquad (7.21)$$

where F is a surjection $\Omega_2 \to \Omega_1$ which yields the partition $\Omega_2 = \bigsqcup_{x \in \Omega_1} F^{-1}(x)$, and $\{r_x\}_{x \in \Omega_1}$ is a family of probability distributions on Ω_2 such that the support of r_x is $F^{-1}(x)$ for every $x \in \Omega_1$. We note that a Markov co-embedding Ψ is determined by F alone, while a Markov embedding Φ is determined by F and $\{r_x\}_{x \in \Omega_1}$ together. Consequently, Ψ is uniquely determined from Φ, while Φ for a given Ψ has the degree of freedom corresponding to $\{r_x\}_{x \in \Omega_1}$. According to this fact, when a Markov co-embedding Ψ and a distribution $q \in \mathcal{P}_2$ are arbitrarily given, we can always choose a Markov embedding Φ satisfying (7.9) and $q \in \Phi(\mathcal{P}_1)$; indeed, defining r_x by

$$r_x(y) := \begin{cases} q(y)/q^F(x) & \text{if } F(y) = x \\ 0 & \text{otherwise,} \end{cases} \qquad (7.22)$$

the resulting Φ satisfies $q = \Phi(q^F) \in \Phi(\mathcal{P}_1)$. This is the reason why $\forall q \in \Phi(\mathcal{P}_1)$ in (7.19) can be replaced with $\forall q \in \mathcal{P}_2$. We thus have

$$\forall q \in \mathcal{P}_2, \forall \alpha_{\Psi(q)} \in T^*_{\Psi(q)}(\mathcal{P}_1), \quad \|\alpha_{\Psi(q)}\|_{\Psi(q)} = \|\Psi^*(\alpha_{\Psi(q)})\|_q, \qquad (7.23)$$

or equivalently,

$$\forall q \in \mathcal{P}_2, \forall \alpha_{\Psi(q)}, \forall \beta_{\Psi(q)} \in T^*_{\Psi(q)}(\mathcal{P}_1),$$
$$g_{\Psi(q)}(\alpha_{\Psi(q)}, \beta_{\Psi(q)}) = g_q(\Psi^*(\alpha_{\Psi(q)}), \Psi^*(\beta_{\Psi(q)})) \qquad (7.24)$$

for every Markov co-embedding Ψ.

The invariance (7.23) characterizes the Fisher co-metric up to a constant factor. Namely, we have the following theorem.

Theorem 7.2 *For $n = 2, 3, \ldots$, let $\Omega_n := \{1, 2, \ldots, n\}$ and $\mathcal{P}_n := \mathcal{P}(\Omega_n)$, and let g_n be the Fisher co-metric on $\mathcal{P}_n$. Suppose that we are given a sequence $\{h_n\}_{n=2}^{\infty}$, where h_n is a contravariant tensor field of degree 2 on $\mathcal{P}_n$ which continuously maps each point $p \in \mathcal{P}_n$ to a bilinear form $h_{n,p} : T^*_p(\mathcal{P}_n)^2 \to \mathbb{R}$. Then the following two conditions are equivalent.*

(i) $\exists c \in \mathbb{R}, \forall n, \quad h_n = cg_n$.
(ii) For any $m \le n$ and any Markov co-embedding $\Psi : \mathcal{P}_n \to \mathcal{P}_m$, it holds that

$$\forall q \in \mathcal{P}_n, \forall \alpha_{\Psi(q)}, \forall \beta_{\Psi(q)} \in T^*_{\Psi(q)}(\mathcal{P}_m),$$
$$h_{m,\Psi(q)}(\alpha_{\Psi(q)}, \beta_{\Psi(q)}) = h_{n,q}(\Psi^*(\alpha_{\Psi(q)}), \Psi^*(\beta_{\Psi(q)})). \qquad (7.25)$$

The proof will be given by rewriting the statement in terms of variance/covariance for random variables. Suppose that a Markov co-embedding $\Psi : \mathcal{P}_1 \to \mathcal{P}_2$ is represented as (7.20) by a surjection $F : \Omega_2 \to \Omega_1$. Then Ψ is represented as $\Psi = \Phi_W$ by the channel W from Ω_2 to Ω_1 defined by

$$W(x \mid y) = \begin{cases} 1 & \text{if } x = F(y), \\ 0 & \text{otherwise.} \end{cases} \qquad (7.26)$$

For an arbitrary $A \in \mathbb{R}^{\Omega_1}$, its conditional expectation w.r.t. W is represented as

$$E_W(A \mid y) = \sum_{x \in \Omega_1} W(x \mid y)A(x) = A(F(y)), \qquad (7.27)$$

so that it follows from (6.9) that

$$\Psi^*(\delta_{q^F}(A)) = \delta_q(A \circ F), \qquad (7.28)$$

where we have invoked $\Psi(q) = q^F$ from (7.20). Hence, (7.23) and (7.24) are rewritten as

$$V_{q^F}(A) = V_q(A \circ F) \tag{7.29}$$

and

$$\mathrm{Cov}_{q^F}(A, B) = \mathrm{Cov}_q(A \circ F, B \circ F). \tag{7.30}$$

These identities themselves are obvious, but what is important is that they character-
ize the variance/covariance up to a constant factor. Namely, we have the following
theorem.

Theorem 7.3 *In the same situation as Theorem 7.2, suppose that we are given a
sequence $\{\gamma_n\}_{n=2}^\infty$, where γ_n is a map which continuously maps each point $p \in \mathcal{P}_n$ to
a bilinear form $\gamma_{n,p}$ on $\mathbb{R}^{\Omega_n}$. Then the following conditions (i) and (ii) are equivalent.*

(i) $\exists c \in \mathbb{R}, \forall n, \forall p \in \mathcal{P}_n, \quad \gamma_{n,p} = c\,\mathrm{Cov}_p$.
(ii) : (ii-1) $\wedge$ (ii-2)

(ii-1) $\forall n, \forall p \in \mathcal{P}_n, \forall A \in \mathbb{R}^{\Omega_n}, \quad \gamma_{n,p}(A, 1) = 0$.
(ii-2) For any $m \le n$ and any surjection $F : \Omega_n \to \Omega_m$, it holds that

$$\forall p \in \mathcal{P}_n, \forall A, B \in \mathbb{R}^{\Omega_m}, \quad \gamma_{m,p^F}(A, B) = \gamma_{n,p}(A \circ F, B \circ F). \tag{7.31}$$

Note that, if we assume that $\{\gamma_n\}_n$ are all symmetric tensors, then (7.31) can be
replaced with

$$\forall p \in \mathcal{P}_n, \forall A \in \mathbb{R}^{\Omega_m}, \quad \gamma_{m,p^F}(A, A) = \gamma_{n,p}(A \circ F, A \circ F), \tag{7.32}$$

which corresponds to (7.29).

See A1 in Appendix for the proof, where we use an argument similar to Čencov's
proof of Theorem 7.1. It is obvious that Theorem 7.2 immediately follows from this
theorem.

Remark 7.4 If we delete (ii-1) from (ii) in Theorem 7.3, then we have (i)$'$ $\Leftrightarrow$ (ii-2) by
replacing (i) with

(i)$'$ $\exists c_1, \exists c_2 \in \mathbb{R}, \forall n, \forall p \in \mathcal{P}_n,$
  $\forall A, B \in \mathbb{R}^{\Omega_n}, \quad \gamma_{n,p}(A, B) = c_1 \langle A, B \rangle_p + c_2 \langle A \rangle_p \langle B \rangle_p$.

We give a proof for (i)$'$ $\Leftrightarrow$ (ii-2) in A1, from which Theorem 7.3 is straightforwad.

8 Strong invariance

In the preceding two sections, we have observed the following facts.

- The monotonicity of metrics and that of co-metrics are logically equivalent.

- The monotonicity logically implies the invariance of metrics and that of co-metrics.
- The invariance of metrics and that of co-metrics are not logically equivalent.

In this section we introduce a new notion of invariance called the strong invariance, and show that:

- The strong invariance of metrics and that of co-metrics are logically equivalent.
- The monotonicity of metrics/co-metrics logically implies the strong invariance of metrics/co-metrics.
- The strong invariance of metrics/co-metrics logically implies the invariance of metrics and that of co-metrics.

Recall the situation of (7.1), (7.2) and (7.3), where we are given submanifolds $M \subset \mathcal{P}_1$ and $N \subset \mathcal{P}_2$, Markov mappings $\varphi : M \to N$ and $\psi : N \to M$ satisfying $\psi \circ \varphi = id_M$, and points $p \in M$ and $q \in N$ satisfying $q = \varphi(p)$ and $p = \psi(q)$. Then we have the following proposition.

Proposition 8.1 *The Fisher metrics g_M and g_N on M and N satisfy*

$$\forall X_p \in T_p(M), \forall Y_q \in T_q(N), \quad g_{M,p}(X_p, \psi_*(Y_q)) = g_{N,q}(\varphi_*(X_p), Y_q), \quad (8.1)$$

or equivalently,

$$\psi_{*,q} = (\varphi_{*,p})^\dagger. \tag{8.2}$$

In addition, $(\psi_{*,q})^\dagger \circ \psi_{*,q} = \varphi_{*,p} \circ \psi_{*,q}$ is the orthogonal projector from $T_q(N)$ onto $\varphi_{*,p}(T_p(M)) = T_q(\varphi(M))$.

The property (8.1) is called the *strong invariance* of the Fisher metric. The proposition will be proved by using the following lemma.

Lemma 8.2 *Let U and V be finite-dimensional metric linear spaces, and let $A : U \to V$ and $B : V \to U$ be linear maps satisfying $BA = I$. Then the following two conditions are equivalent.*

(i) $A^\dagger A = BB^\dagger = I$.
(ii) $B = A^\dagger$.

When these conditions hold, $B^\dagger B = AB$ is the orthogonal projector from V onto the image Im A of A.

Proof It is obvious that (ii) $\Rightarrow$ (i) under the assumption $BA = I$. Conversely, if we assume (i) with $BA = I$, then we have

$$(B - A^\dagger)(B - A^\dagger)^\dagger = BB^\dagger - BA - A^\dagger B^\dagger + A^\dagger A$$
$$= I - I - I + I = 0,$$

from which (ii) follows.

Assume (i) and (ii). Then $(B^\dagger B)^2 = B^\dagger B$ due to $BB^\dagger = I$, which implies that $B^\dagger B$ is the orthogonal projector onto Im $B^\dagger =$ Im A. □

Proof of Prop 8.1.

Letting $U := T_p(M)$, $V := T_q(N)$, $A := \varphi_{*,p}$ and $B := \psi_{*,q}$ in the previous lemma, we see from (7.4), (7.8) and (7.17) that the assumption $BA = I$ and the condition (i) are satisfied, so that we have (ii), which means (8.2). ☐

As can be seen from Lemma 8.2, the strong invariance (8.1) is equivalent to the condition that both the invariance for the metric (7.6)–(7.8) and the invariance for the co-metric (7.14)–(7.16) hold. Taking the transpose of both sides of (8.2), the strong invariance is also expressed

$$\psi_q^* = (\varphi_p^*)^\dagger, \tag{8.3}$$

which means that the Fisher co-metric satisfies

$$\forall \alpha_p \in T_p^*(M), \forall \beta_q \in T_q^*(N), \quad g_{M,p}(\alpha_p, \varphi^*(\beta_q)) = g_{N,q}(\psi^*(\alpha_p), \beta_q). \tag{8.4}$$

Since the strong invariance (8.4) logically implies the invariance (7.7) for the Fisher metric via (8.1), we see that the following proposition, which is stated in a rough form similar to Proposition 6.3, is obtained as a corollary of Čencov's theorem (Theorem 7.1).

Proposition 8.3 *The strong invariance (8.4) characterizes the Fisher co-metric up to a constant factor.*

An exact formulation of this proposition will be given in A2 of Appendix with a proof based on Čencov's theorem.

Remark 8.4 The above proposition is stronger than Proposition 6.3 and weaker than Theorem 7.2. To formulate Proposition 6.3 and Proposition 8.3 in exact forms similar to Theorem 7.2, it matters what assumptions should be imposed on bilinear forms $\{h_{n,p}\}$ on the cotangent spaces prior to the monotonicity or the strong invariance. Here we should keep in mind that the significance of these propositions, which are weaker than Theorem 7.2, lies in the fact that they follow from Čencov's theorem while Theorem 7.2 does not. For Proposition 6.3, we need to assume that $\{h_{n,p}\}$ are inner products (positive symmetric forms) to ensure that the monotonicity of them is translated into the monotonicity of the corresponding inner products on the tangent spaces. For Proposition 8.3, on the other hand, we only need to assume $\{h_{n,p}\}$ to be non-degenerate (nonsingular) and symmetric. See A2 for details.

Let us consider the strong invariance (8.4) for the case when (ϕ, ψ) is a Markov embedding/co-embedding pair (Φ, Ψ) and rewrite it into an identity for the covariance of random variables. Let Ω_1 and Ω_2 be arbitrary finite sets satisfying $|\Omega_1| \leq |\Omega_2|$ and let $\mathcal{P}_i := \mathcal{P}(\Omega_i)$, $i = 1, 2$. Given a surjection $F : \Omega_2 \to \Omega_1$ and a distribution $q \in \mathcal{P}_2$, let (Φ, Ψ) be defined by (7.20) and (7.21) with (7.22). For arbitrary $A \in \mathbb{R}^{\Omega_1}$ and $B \in \mathbb{R}^{\Omega_2}$, let $\alpha_{q^F} := \delta_{q^F}(A) \in T_{q^F}^*(\mathcal{P}_1)$ and $\beta_q := \delta_q(B) \in T_q^*(\mathcal{P}_2)$, for which the strong invariance (8.4) is represented as

$$g_{q^F}(\alpha_{q^F}, \Phi^*(\beta_q)) = g_q(\Psi^*(\alpha_{q^F}), \beta_q). \tag{8.5}$$

Recalling (7.28), we have

$$\Psi^*(\alpha_{q^F}) = \delta_q(A \circ F). \tag{8.6}$$

On the other hand, Φ is represented as $\Phi = \Phi_V$ by the channel V defined by

$$V(y \,|\, x) = \begin{cases} r_x(y) = \frac{q(y)}{q^F(x)} & \text{if } x = F(y) \\ 0 & \text{otherwise,} \end{cases} \tag{8.7}$$

so that (6.9) yields

$$\Phi^*(\beta_q) = \delta_{q^F}(E_V(B \,|\, \cdot)). \tag{8.8}$$

Hence, the strong invariance (8.5) is rewritten as

$$\mathrm{Cov}_{q^F}(A, E_V(B \,|\, \cdot)) = \mathrm{Cov}_q(A \circ F, B). \tag{8.9}$$

This identity can be verified directly as follows. Letting $a := \langle A \rangle_{q^F} = \langle A \circ F \rangle_q$ and $b := \langle B \rangle_q = \langle E_V(B \,|\, \cdot) \rangle_{q^F}$, we have

$$\begin{aligned} \mathrm{RHS} &= \sum_y q(y)(A(F(y)) - a)(B(y) - b) \\ &= \sum_x \sum_{y \in F^{-1}(x)} q(y)(A(F(y)) - a)(B(y) - b) \\ &= \sum_x (A(x) - a) \sum_{y \in F^{-1}(x)} q(y)(B(y) - b) \\ &= \sum_x q^F(x)(A(x) - a)(E_V(B \,|\, x) - b) \\ &= \mathrm{LHS}, \end{aligned} \tag{8.10}$$

where the fourth equality follows from

$$E_V(B \,|\, x) = \sum_y V(y \,|\, x)B(y) = \frac{1}{q^F(x)} \sum_{y \in F^{-1}(x)} q(y)B(y). \tag{8.11}$$

Note that (7.30) is obtained from (8.9) by substituting $B \circ F$ for B.

9 Weak invariance for affine connections

In addition to characterizing the Fisher metric by the invariance with respect to Markov embeddings, Čencov also gave a characterization of the α-connections by the invariance condition. In this section we show that a similar notion to the strong invariance

of metrics, which is described in terms of Markov embedding/co-embedding pairs, can be considered for affine connections.

Let Ω_1, Ω_2 be arbitrary finite sets satisfying $2 \leq |\Omega_1| \leq |\Omega_2|$, and let $\Phi : \mathcal{P}_1 \rightarrow \mathcal{P}_2$ be a Markov embedding, where $\mathcal{P}_i := \mathcal{P}(\Omega_i)$, $i = 1, 2$. Suppose that affine connections ∇ and ∇' are given on $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively. When these connections are the α-connection on $\mathcal{P}_1$ and that on $\mathcal{P}_2$ for some common $\alpha \in \mathbb{R}$, they satisfy

$$\forall X, Y \in \mathfrak{X}(\mathcal{P}_1), \quad \Phi_*(\nabla_X Y) = \nabla'_{\Phi_*(X)} \Phi_*(Y). \tag{9.1}$$

Some remarks on the meaning of the above equation are in order. First, we define $\Phi_*(X)$ for an arbitrary vector field X on $\mathcal{S}_1$ as a vector field on $K := \Phi(\mathcal{P}_1)$ that maps each point $q = \Phi(p) \in K$, where $p \in \mathcal{P}_1$, to

$$(\Phi_*(X))_q := \Phi_{*,p}(X_p) \in T_q(K). \tag{9.2}$$

Since K is a submanifold of $\mathcal{S}_2$ on which the connection ∇' is given, $\nabla'_{\Phi_*(X)} \Phi_*(Y)$ in (9.1) is defined as a map which maps each point $q \in K$ to a tangent vector in $T_q(\mathcal{P}_2)$, although $\nabla'_{\Phi_*(X)} \Phi_*(Y)$ does not belong to $\mathfrak{X}(K)$ in general. The condition (9.1) means that K is autoparallel in $\mathcal{P}_2$ with respect to ∇' and that the restricted connection of ∇' induced on the autoparallel K is obtained from ∇ by the diffeomorphism $\Phi : \mathcal{P}_1 \rightarrow K$. Čencov [3] showed that the invariance condition characterizes the family $\{\alpha\text{-connection}\}_{\alpha \in \mathbb{R}}$ by a formulation similar to Theorem 7.1.

Remark 9.1 As is mentioned above, the fact that $\{\alpha\text{-connection}\}_{\alpha \in \mathbb{R}}$ satisfy the invariance (9.1) implies that $K = \Phi(\mathcal{P}_1)$ is autoparallel in $\mathcal{P}_2$ w.r.t. the α-connection for every $\alpha \in \mathbb{R}$. A kind of converse result is found in [5], which states that if a submanifold K of $\mathcal{P}_2 = \mathcal{P}(\Omega_2)$ is autoparallel w.r.t. the α-connection for every $\alpha \in \mathbb{R}$ (or, for some two different values of α), then K is represented as $\Phi(\mathcal{S}_1)$ by some Markov embedding Φ from some $\mathcal{S}_1 = \mathcal{S}(\Omega_1)$ into $\mathcal{S}_2$.

Let g and g' be the Fisher metrics on $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively. Noting that Φ_* is an isometry with respect to these metrics due to the invariance of the Fisher metric, (9.1) implies that

$$\forall X, Y, Z \in \mathfrak{X}(\mathcal{S}_1), \quad g(\nabla_X Y, Z) = g'(\nabla'_{\Phi_*(X)} \Phi_*(Y), \Phi_*(Z)) \circ \Phi, \tag{9.3}$$

where the RHS denotes the function on $\mathcal{P}_1$ such that

$$p \mapsto g'_{\Phi(p)}\big((\nabla'_{\Phi_*(X)} \Phi_*(Y))_{\Phi(p)}, \Phi_{*,p}(Z_p)\big). \tag{9.4}$$

Since (9.3) is apparently weaker than (9.1), we call this property the *weak invariance* of the connections ∇, ∇'. In an actual fact, however, as is mentioned in [6] and is proved in [7], the weak invariance (9.3) characterizes the family $\{\alpha\text{-connection}\}_{\alpha \in \mathbb{R}}$ as well as the stronger condition (9.1).

Now, recalling the strong invariance (8.1) of the Fisher metric, we see that the weak invariance (9.3) is equivalent to

$$\forall X, Y \in \mathfrak{X}(\mathcal{S}_1), \quad \nabla_X Y = \Psi_*\big(\nabla'_{\Phi_*(X)} \Phi_*(Y)\big), \tag{9.5}$$

where the RHS denotes the map

$$\mathcal{P}_1 \ni p \mapsto \Psi_{*,\Phi(p)}\big(\big(\nabla'_{\Phi_*(X)} \Phi_*(Y)\big)_{\Phi(p)}\big). \tag{9.6}$$

The fact that the weak invariance is a condition on connections is more clearly expressed in (9.5) than (9.3) in that (9.5) does not include the Fisher metric.

10 Concluding remarks

In this paper we have focused on the face of the Fisher metric as a metric on the cotangent bundle, calling it the Fisher co-metric to distinguish it from the original Fisher metric on the tangent bundle. What we have shown are listed below.

1. Based on a correspondence between cotangent vectors and random variables, the Fisher co-metric is defined via the variance/covariance in a natural way (Sect. 2).
2. The Cramér-Rao inequality is trivialized by considering the Fisher co-metric (Sect. 4).
3. The role of the m-connection as a connection on the cotangent bundle is important in considering the achievability condition for the Cramér-Rao inequality (Sect. 5).
4. The monotonicity of the Fisher metric is equivalently translated into the monotonicity of the Fisher co-metric and that of the variance (Sect. 6).
5. The invariance of the Fisher metric and that of the Fisher co-metric are not logically equivalent, and a new Čencov-type theorem for characterizing the Fisher co-metric by the invariance is established, which can also be regarded as a theorem for characterizing the variance/covariance (Sect. 7).
6. The notion of strong invariance is introduced, which combines the invariance of the Fisher metric and that of the Fisher co-metric (Sect. 8).
7. The weak invariance of the α-connections is expressed in a formulation similar to the strong invariance of the Fisher metric/co-metric (Sect. 9).

It should be noted that, although this paper emphasizes the importance of the Fisher co-metric, this does not diminish the importance of the Fisher metric at all. Apart from the importance as a metric on the tangent bundle itself, which is essential for geometry of statistical manifolds, we should not forget that the Fisher information matrix (i.e. the components of the Fisher metric) is of primary importance as a practical tool to compute the Fisher co-metric. Even if we know that $g_M^{ij}(p) = g_{M,p}((d\xi^i)_p, (d\xi^j)_p)$ in (4.3) can be defined by (4.5) and (4.6) and that understanding $g_M^{ij}(p)$ in this way is important for conceptual understanding of the Cramér-Rao inequality, this does not tell us a method to compute $g_M^{ij}(p)$ for a given statistical model (M, ξ) better than computing the inverse of the Fisher information matrix $G_M(p) := [g_{M,ij}(p)]$ in general.

Finally, we note that some of the results obtained here can be extended to the quantum case in several directions, which will be discussed in a forthcoming paper.

Data Availability Data sharing is not applicable to this article as no data sets were generated or analyzed during the current study.

Declarations

Conflict of interest The author is a current member of the Editorial Board of Information Geometry. The author states that there is no other conflict of interest.

Appendix

A1 Proof of Theorem 7.3

As mentioned in Remark 7.4, we show the equivalence (i)$'$ $\Leftrightarrow$ (ii-2), with expressing (i)$'$ as

(i)$'$ $\exists c_1, \exists c_2 \in \mathbb{R}, \ \forall n, \forall p \in \mathcal{P}_n, \ \gamma_{n,p} = c_1 \langle \cdot, \cdot \rangle_p + c_2 \langle \cdot \rangle_p \langle \cdot \rangle_p$.

The implication (i)$' \Rightarrow$ (ii-2) is obvious. To show the converse, we assume (ii-2) and will derive (i)$'$ by several steps. The uniform distribution on Ω_n is denoted by u_n in the sequel.

(a) $\forall n, \ \exists c_{1,n}, \exists c_{2,n} \in \mathbb{R}, \ \gamma_{n,u_n} = c_{1,n} \langle \cdot, \cdot \rangle_{u_n} + c_{2,n} \langle \cdot \rangle_{u_n} \langle \cdot \rangle_{u_n}$.

Proof Fix $n \geq 2$ arbitrarily, and let $m = n$, and F be a permutation on Ω_n. Define $e_{n,i} \in \mathbb{R}^{\Omega_n}$ for $i \in \Omega_n$ by $e_{n,i}(j) = \delta_{i,j}$. Then $u_n{}^F = u_n$, and condition (ii-2) claims that

$$\forall i, \forall j \in \Omega_n, \ \gamma_{n,u_n}(e_{n,i}, e_{n,j}) = \gamma_{n,u_n}(e_{n,i} \circ F, e_{n,j} \circ F)$$
$$= \gamma_{n,u_n}(e_{n,F^{-1}(i)}, e_{n,F^{-1}(j)}).$$

Since this holds for any permutation F, there exist $a_n, b_n \in \mathbb{R}$ such that

$$\forall i, \forall j \in \Omega_n, \ \gamma_{n,u_n}(e_{n,i}, e_{n,j}) = a_n \delta_{i,j} + b_n.$$

Comparing this with

$$\langle e_{n,i}, e_{n,j} \rangle_{u_n} = \frac{1}{n} \delta_{i,j} \quad \text{and} \quad \langle e_{n,i} \rangle_{u_n} \langle e_{n,j} \rangle_{u_n} = \frac{1}{n^2},$$

we have the assertion of (a) with letting $c_{1,n} := na_n$ and $c_{2,n} := n^2 b_n$. □

(b) $\exists c_1, \exists c_2 \in \mathbb{R}, \ \forall n, \ \gamma_{n,u_n} = c_1 \langle \cdot, \cdot \rangle_{u_n} + c_2 \langle \cdot \rangle_{u_n} \langle \cdot \rangle_{u_n}$.

Proof It suffices to show that the constants $\{c_{1,n}\}$ and $\{c_{2,n}\}$ in (a) satisfy $c_{1,n} = c_{1,m}$ and $c_{2,n} = c_{2,m}$ for every $n, m \geq 2$. Let $F : \Omega_{mn} \to \Omega_n$ be defined by

$$F(i) = k \quad \text{if} \quad (k-1)m + 1 \leq i \leq km, \quad k \in \mathbb{N}.$$

Then we have $u_{mn}^F = u_n$, and (ii-2) yields

$$\forall A, \forall B \in \mathbb{R}^{\Omega_n}, \quad \gamma_{n,u_n}(A, B) = \gamma_{mn,u_{mn}}(A \circ F, B \circ F).$$

Due to (a), the LHS of the above equation is represented as

$$\gamma_{n,u_n}(A, B) = c_{1,n}\langle A, B \rangle_{u_n} + c_{2,n}\langle A \rangle_{u_n} \langle B \rangle_{u_n},$$

while the RHS is represented as

$$\gamma_{mn,u_{mn}}(A \circ F, B \circ F) = c_{1,mn}\langle A \circ F, B \circ F \rangle_{u_{mn}} + c_{2,mn}\langle A \circ F \rangle_{u_{mn}} \langle B \circ F \rangle_{u_{mn}}$$
$$= c_{1,mn}\langle A, B \rangle_{u_n} + c_{2,mn}\langle A \rangle_{u_n} \langle B \rangle_{u_n}.$$

Thus we have $c_{1,n} = c_{1,mn}$ and $c_{2,n} = c_{2,mn}$. Similarly, we can show that $c_{1,m} = c_{1,mn}$ and $c_{2,m} = c_{2,mn}$. Hence we obtain $c_{1,n} = c_{1,m}$ and $c_{2,n} = c_{2,m}$. □

(c) For any $n \geq 2$ and any $p \in \mathcal{P}_n$ whose components $\{p(i)\}_{i \in \Omega_n}$ are all rational numbers, we have $\gamma_{n,p} = c_1 \langle \cdot, \cdot \rangle_p + c_2 \langle \cdot \rangle_p \langle \cdot \rangle_p$, where c_1 and c_2 are the constants in (b).

Proof Suppose that the components of p are represented as $p(i) = k_i/m$ by natural numbers $m, k_1, \ldots, k_n$ satisfying $\sum_{i=1}^n k_i = m$. Let $\Omega_m = \bigsqcup_{i=1}^n K_i$ be a partition of Ω_m such that $|K_i| = k_i$, and define $F : \Omega_m \to \Omega_n$ by

$$F(j) = i \quad \text{if} \quad j \in K_i.$$

Then we have $u_m^F = p$. Letting $\gamma'_{n,p} := c_1 \langle \cdot, \cdot \rangle_p + c_2 \langle \cdot \rangle_p \langle \cdot \rangle_p$ and noting that both $\{\gamma_{n,p}\}$ and $\{\gamma'_{n,p}\}$ satisfy condition (ii-2), we obtain

$$\forall A, B \in \mathbb{R}^{\Omega_n}, \quad \gamma_{n,p}(A, B) = \gamma_{m,u_m}(A \circ F, B \circ F)$$
$$= \gamma'_{m,u_m}(A \circ F, B \circ F)$$
$$= \gamma'_{n,p}(A, B),$$

where the second equality follows from (b). □

(d) (i)′ is derived from (c) and the continuity of $p \mapsto \gamma_{n,p}$.

□

A2 An exact form of Proposition 8.3

We present a proposition of Čencov-type that can be regarded as an exact version of Proposition 8.3, and show that the proposition follows from Theorem 7.1 (Čencov's theorem).

Proposition A.1 *For $n = 2, 3, \ldots$, let $\Omega_n := \{1, 2, \ldots, n\}$ and $\mathcal{P}_n := \mathcal{P}(\Omega_n)$, and let g_n be the Fisher co-metric on $\mathcal{P}_n$. Suppose that we are given a sequence $\{h_n\}_{n=2}^{\infty}$, where h_n is a contravariant tensor field of degree 2 on $\mathcal{P}_n$ which continuously maps each point $p \in \mathcal{P}_n$ to a non-degenerate and symmetric bilinear form $h_{n,p} : T_p^*(\mathcal{P}_n)^2 \to \mathbb{R}$. Then the following two conditions are equivalent.*

(i) $\exists c \in \mathbb{R} \setminus \{0\}$, $\forall n$, $h_n = cg_n$.
(ii) For any $m \leq n$ and any pair (Φ, Ψ) of Markov embedding $\Phi : \mathcal{P}_m \to \mathcal{P}_n$ and Markov co-embedding $\Psi : \mathcal{P}_n \to \mathcal{P}_m$ satisfying $\Psi \circ \Phi = id_{\mathcal{P}_m}$, it holds that

$$\forall p \in \mathcal{P}_m, \forall \alpha_p \in T_p^*(\mathcal{P}_m), \forall \beta_{\Phi(p)} \in T_{\Phi(p)}^*(\mathcal{P}_n),$$
$$h_{m,p}(\alpha_p, \Phi^*(\beta_{\Phi(p)})) = h_{n,\Phi(p)}(\Psi^*(\alpha_p), \beta_{\Phi(p)}). \tag{A.1}$$

Proof Since the Fisher co-metric satisfies (8.4), we have (i) $\Rightarrow$ (ii). To prove the converse, assume that $\{h_n\}_{n=2}^{\infty}$ satisfies (ii). Due to the assumption that $h_{n,p}$ is non-degenerate for every n and $p \in \mathcal{P}_n$, $h_{n,p}$ induces the correspondence $\overset{h_{n,p}}{\longleftrightarrow}$ between $T_p(\mathcal{P}_n)$ and $T_p^*(\mathcal{P}_n)$ together with the bilinear form $h_{n,p} : T_p(\mathcal{P}_n)^2 \to \mathbb{R}$ as in the case of inner products by

$$X_p \overset{h_{n,p}}{\longleftrightarrow} \alpha_p \Leftrightarrow \forall \beta_p \in T_p^*(\mathcal{P}_n), \ \beta_p(X_p) = h_{n,p}(\alpha_p, \beta_p)$$

and

$$X_p \overset{h_{n,p}}{\longleftrightarrow} \alpha_p \text{ and } Y_p \overset{h_{n,p}}{\longleftrightarrow} \beta_p \Rightarrow h_{n,p}(X_p, Y_p) = h_{n,p}(\alpha_p, \beta_p).$$

Then, similar to the equivalence between (8.1) and (8.4), we see that (A.1) is equivalent to

$$\forall p \in \mathcal{P}_m, \forall X_p \in T_p(\mathcal{P}_m), \forall Y_{\Phi(p)} \in T_{\Phi(p)}(\mathcal{P}_n),$$
$$h_{m,p}(X_p, \Psi_*(Y_{\Phi(p)})) = h_{n,\Phi(p)}(\Phi_*(X_p), Y_{\Phi(p)}). \tag{A.2}$$

Indeed, if

$$X_p \overset{h_{m,p}}{\longleftrightarrow} \alpha_p \text{ and } Y_q \overset{h_{n,q}}{\longleftrightarrow} \beta_q, \quad \text{where} \quad q := \Phi(p),$$

then

$$h_{m,p}(\alpha_p, \Phi^*(\beta_q)) = \Phi^*(\beta_q)(X_p) = \beta_q(\Phi_*(X_p)) = h_{n,q}(\Phi_*(X_p), Y_q)$$

and

$$h_{n,q}(\Psi^*(\alpha_p), \beta_q) = \Psi^*(\alpha_p)(Y_q) = \alpha_p(\Psi_*(Y_q)) = h_{m,p}(X_p, \Psi_*(Y_q)),$$

which shows the equivalence of (A.1) and (A.2). Due to $\Psi_* \circ \Phi_* = id_{T_p(\mathcal{P}_m)}$, Eq. (A.2) implies (7.11) in (ii) of Theorem 7.1. Thus we have the following train of implications:

$$(\mathrm{ii}) \implies (\mathrm{ii}) \text{ of Theorem } 7.1 \implies (\mathrm{i}) \text{ of Theorem } 7.1 \implies (\mathrm{i}).$$

$\square$

References

1. Amari, S., Nagaoka, H.: Methods of Information Geometry. American Mathematical Society, Providence, RI; Oxford University Press, Oxford (2000)
2. Nagaoka, H., Amari, S.: Differential geometry of smooth families of probability distributions. METR 82-7, Dept Math Eng and Instr Phys, Univ. of Tokyo (1982). https://www.keisu.t.u-tokyo.ac.jp/research/techrep/y1982/. https://bsi-ni.brain.riken.jp/database/item/104
3. Čencov, N.N.: Statistical decision rules and optimal inference. American Mathematical Society, Providence, RI (1982) (original Russian edition: Nauka, Moscow (1972))
4. Nagaoka, H., Fujiwara, A.: Autoparallelity of quantum statistical manifolds in the light of quantum estimation theory. arXiv:2307.03431
5. Nagaoka, H.: Information-geometrical characterization of statistical models which are statistically equivalent to probability simplexes. Proc. 2017 IEEE ISIT, pp. 1346–1350 (2017)
6. Fujiwara, A.: Hommage to Chentsov's theorem. Inf. Geo. **7**(Suppl 1), 79–98 (2024)
7. Fujiwara, A.: Foundations of Information Geometry. Kyoritsu Shuppan, Tokyo (2021). (**in Japanese**)