

The uniqueness of the Fisher metric as information metric

Hông Vân Lê¹

Received: 12 July 2013 / Revised: 4 April 2016 / Published online: 14 May 2016
© The Institute of Statistical Mathematics, Tokyo 2016

Abstract We define a mixed topology on the fiber space $\cup_{\mu} \oplus^n L^n(\mu)$ over the space $\mathcal{M}(\Omega)$ of all finite non-negative measures μ on a separable metric space Ω provided with Borel σ -algebra. We define a notion of strong continuity of a covariant n -tensor field on $\mathcal{M}(\Omega)$. Under the assumption of strong continuity of an information metric, we prove the uniqueness of the Fisher metric as information metric on statistical models associated with Ω . Our proof realizes a suggestion due to Amari and Nagaoka to derive the uniqueness of the Fisher metric from the special case proved by Chentsov by using a special kind of limiting procedure. The obtained result extends the monotonicity characterization of the Fisher metric on statistical models associated with finite sample spaces and complement the uniqueness theorem by Ay–Jost–Lê–Schwachhöfer that characterizes the Fisher metric by its invariance under sufficient statistics.

Keywords Monotonicity of the Fisher metric · Chentsov’s theorem · Mixed topology

1 Introduction

Recent successful applications of information geometry, see e.g. [Amari \(1987\)](#), [Amari and Nagaoka \(2000\)](#), [Ay et al. \(2011\)](#), [Shahshahani \(1979\)](#), where the Fisher metric plays a fundamental role, motivate us to find an answer to the following important question. Is there another metric on statistical models with natural properties, which we could name information metric?

H.V.L. is partially supported by RVO: 67985840.

✉ Hông Vân Lê
hvle@math.cas.cz

¹ Institute of Mathematics of ASCR, Zitna 25, 11567 Prague 1, Czech Republic

Intuitively, information metric should reflect the amount of non-negative information of a statistical model, moreover

- it should measure “information loss” associated with a data processing and this information loss is a non-negative quantity (Chentsov 1978, Axiom A);
- it must be invariant under sufficient statistics, that is, mappings between sample spaces that preserve all information about the parameter x .

In statistical decision theory, a data processing is a statistical decision rule, which can be deterministic or randomized. A deterministic decision rule is a measurable map, which is also called a statistic. An indeterministic decision rule is a Markov transition distribution (Chentsov 1982). Recently, Ay–Jost–Lê–Schwachhöfer showed that a transformation between statistical models which is associated with a Markov transition distribution is a composition of the inverse of a transformation, which is associated with a sufficient statistic, and a transformation which is associated with a statistic (Ay et al. 2015, Theorem 4.10). Hence, assuming the condition of invariance under sufficient statistics, the “information loss” condition is reduced to the case where data processing is associated with a statistic.

Using the concept of a continuous local statistical covariant tensor field on statistical models (Ay et al. 2015, Definition 2.8), see also Definition 10 below, and utilizing the above discussion, we propose the following.

Definition 1 Given a class $\{\Omega\}$ of measure spaces, an *information metric* on statistical models (Definition 6), or more generally, on parametrized measure models (M, Ω, μ, p) where $\Omega \in \{\Omega\}$ is a continuous local statistical non-negative definite quadratic form $F_{(M, \Omega, \mu, p)}$ (Definitions 8, 9, 10) that satisfies the following two conditions:

1. the “information loss” $F_{(M, \Omega, \mu, p)} - F_{(M, \Omega_1, \kappa_*(\mu), \kappa_*(p))}$ is a non-negative definite quadratic form for any statistic $\kappa : \Omega \rightarrow \Omega_1$;
2. the “information loss” $F_{(M, \Omega, \mu, p)} - F_{(M, \Omega_1, \kappa_*(\mu), \kappa_*(p))}$ is zero (quadratic form) if κ is sufficient with respect to the parameter $x \in M$.

Each of the conditions (1) and (2) in Definition 1 is natural and has its own appeal. The condition (2) has been considered in Ay et al. (2015) as a criterion for a natural metric on parametrized measure models. The condition (1) is simpler formulated than the condition (2), since it does not depend on the notion of a sufficient statistic, that depends on a statistical model under consideration and depends on the notion of information implicitly. (For a modern definition of a sufficient statistic we refer the reader to Ay et al. (in preparation), where Ay–Jost–Lê–Schwachhöfer propose a geometric definition of a sufficient statistic associated with a (signed) parametrized measure model in terms of Banach manifolds in consideration, which agrees with the old concept of sufficient statistics that uses the Fisher–Neyman characterization.)

In 1972 Chentsov proved that on statistical models (M, Ω, μ, p) associated with finite sample spaces Ω the Fisher metric g^F (Example 11) is a unique metric, up to a multiplicative constant, that satisfies (2) (Chentsov 1982). In Ay et al. (2015, Corollary 4.11), for finite sample spaces Ω , we derived the uniqueness (up to a multiplicative constant) of a metric that satisfies the condition (1) on statistical models associated

with Ω from the uniqueness of a metric on statistical models that satisfies the condition (2) on Ω , see Proposition 25 and the Appendix at the end of this note for a discussion on the Chentsov theorem. The converse statement, every metric that satisfies the condition (2) also satisfies the condition (1), follows from the monotonicity theorem for the Fisher metric on statistical models associated with finite sample spaces.

In 2012 Ay–Jost–Lê–Schwachhöfer proved that the Fisher metric is a unique metric, up to a multiplicative constant, on statistical models that satisfies (2) (Ay et al. 2015, Remark 3.23). [On parametrized measure models there are many information metrics that satisfy the condition (2) (Ay et al. 2015, Theorem 2.10). This fact has been observed earlier for parametrized measure models associated with finite sample spaces by Campbell (1986)]. Further, Theorem 3.11 in Ay et al. (2015) states that, the Fisher metric satisfies (1) if Ω, Ω_1 are smooth manifolds and μ is dominated by a Lebesgue measure.

In our paper, we extend the aforementioned results as follows. Our first observation is the following:

Theorem 2 (The monotonicity of the Fisher metric) *Let Ω_1, Ω_2 be topological spaces with Borel σ -algebra, $\kappa : \Omega_1 \rightarrow \Omega_2$ a statistic. Assume that $(M, \Omega_1, \mu_1, p_1)$ and $(M, \Omega_2, \kappa_*(\mu_1), \kappa_*(p_1))$ are 2-integrable parametrized measure models. Then for all $x \in M$ and $V \in T_x M$ we have $g_{(M, \Omega_1, \mu_1, p_1)}^F(V, V) \geq g_{(M, \Omega_2, \kappa_*(\mu_1), \kappa_*(p_1))}^F(V, V)$.*

Theorem 2 is possibly known to experts in the field, but we include it here as well as its short proof since we have not seen a precise statement with a proof of it in an available source and we wish to discuss its consequence. We obtain immediately from the Ay–Jost–Lê–Schwachhöfer theorem (Ay et al. 2015, Remark 3.23) and Theorem 2 the following

Corollary 3 *Let $\{\Omega\}$ be the class of topological spaces provided with Borel σ -algebra. Any continuous local statistical non-negative definite quadratic form F on statistical models associated with $\{\Omega\}$ that satisfies the condition (2) in Definition 1 also satisfies the condition (1) in Definition 1. In other words, the condition (2) is stronger than the condition (1) for those F .*

To prove the uniqueness result for an information metric that satisfies the weaker monotonicity condition (1) in Definition 1 we pose a topological condition on such an information metric. This condition is formulated in terms of the strong continuity, the notion we introduce in Definition 18.

For a measurable space (Ω, Σ) let us denote by $\mathcal{M}(\Omega)$ the subset of all finite non-negative measures on Ω .

Theorem 4 (The uniqueness of the Fisher metric) *Let $\{\Omega\}$ be the class of separable metrizable topological spaces provided with Borel σ -algebra. Assume that F is a continuous local statistical non-negative definite quadratic form defined on all 2-integrable statistical models (M, Ω, μ, p) (Definitions 6, 10) where $\Omega \in \{\Omega\}$. If F satisfies the monotonicity condition (1) in Definition 1.1 and the associated quadratic form \tilde{F} on $\mathcal{M}(\Omega)$ (Definition 10) is strongly continuous for all Ω , then F is the Fisher quadratic form up to a multiplicative constant.*

Corollary 5 *Let $\{\Omega\}$ be the class of separable metrizable topological spaces provided with Borel σ -algebra. Any continuous local statistical non-negative definite quadratic form F on statistical models associated with $\{\Omega\}$ that satisfies the condition (1) in Definition 1 also satisfies the condition (2) in Definition 1, if the associated form \bar{F} on $\mathcal{M}(\Omega)$ satisfies the strong continuity condition for all Ω . In other words, the combination of the condition (1) and the strong continuity condition is stronger than the condition (2) for those F .*

In Remark 26 below we argue how we consider Theorem 4 as a generalization of the characterization the Fisher metric by its monotonicity in the case of finite sample spaces, which is equivalent to the Chentsov theorem. Since there are many measure classes which are invariant under statistics, see e.g. Bogachev (2007, vol. II, Chapter 9) for discussion, we conjecture that without the strong continuity assumption there exists a local statistical continuous metric that satisfies (1) but does not satisfy (2).

The remainder of our paper is organized as follows. In Sect. 2, we recall the notion of a k -integrable parametrized measure model and the notion of a local statistical continuous covariant tensor field that have been introduced by Ay–Jost–Lê–Schwachhöfer in Ay et al. (2015). In Sect. 3, we prove Theorem 2. In Sect. 4, we assume that Ω is a separable metrizable topological space provided with Borel σ -algebra. We introduce a mixed topology on the space $\mathcal{L}_n^n(\Omega) := \cup_{\mu \in \mathcal{M}(\Omega)} \oplus^n L^n(\Omega, \mu)$, which enjoys nice properties (Proposition 17). Using this topology we introduce the notion of strongly continuous covariant n -tensors on $\mathcal{M}(\Omega)$ (Definition 18). In Sect. 5, we prove Theorem 4 by deriving it from the special case associated with finite sample spaces. Finally, we include an appendix containing a note on the Chentsov uniqueness theorem.

The idea to derive the uniqueness of the Fisher metric from its special case proved by Chentsov for finite sample spaces has been proposed by Amari and Nagaoka (2000, p. 39) as follows “Here we shall only observe that Chentsov’s theorem leads to the Fisher metric and the α -connections if a kind of limiting procedure is permitted”, see also Remark 26(3) on a similar idea due to Morozova–Chentsov. In this note we have found such limiting procedure in terms of strong continuity associated with the mixed topology.

2 k -integrable parametrized measure models and local statistical continuous tensor fields

For $\mu_0 \in \mathcal{M}(\Omega)$ denote by

$$\begin{aligned}\mathcal{M}_+(\Omega, \mu_0) &:= \{\mu = \phi\mu_0 \mid \phi \in L^1(\Omega, \mu_0), \phi > 0, \mu_0\text{-a.e.}\}, \\ \mathcal{P}_+(\Omega, \mu_0) &= \{\mu \in \mathcal{M}_+(\Omega, \mu_0) : \mu(\Omega) = 1\}.\end{aligned}$$

Definition 6 (Ay et al. 2015, Definition 2.4; cf. Amari 1987, §2, p. 25; Amari and Nagaoka 2000, §2.1) Let $k \geq 1$. A k -integrable parametrized measure model is a quadruple (M, Ω, μ, p) consisting of a smooth (finite or infinite dimensional) Banach manifold M and a continuous map $p : M \rightarrow \mathcal{M}_+(\Omega, \mu)$ provided with the L^1 -topology such that there exists a density potential $\bar{p} = \frac{dp}{d\mu} : M \times \Omega \rightarrow \mathbb{R}$ satisfying $p(x) = \bar{p}(x, \omega)d\mu(\omega)$ and the following conditions:

1. the function $x \mapsto \ln \bar{p}(x, \omega) = \ln \frac{dp(x)}{d\mu}(\omega) : M \rightarrow \mathbb{R}$ is defined and continuously Gâteaux-differentiable for μ -almost all $\omega \in \Omega$,
2. for all continuous vector field V on M the function $\omega \mapsto \partial_V \ln \bar{p}(x, \omega)$ belongs to $L^k(\Omega, p(x))$; moreover, the function $x \mapsto \|\partial_V \ln \bar{p}(x, \omega)\|_{L^k(\Omega, p(x))}$ is continuous on M .

We call M the *parameter space* of (M, Ω, μ, p) . We call (M, Ω, μ, p) a *statistical model* if $p(M) \subset \mathcal{P}_+(\Omega, \mu)$.

In Definition 6 the continuous Gâteaux-differentiability of $\ln \bar{p}(x, \omega)$ in $x \in M$ means the continuity of the Gateaux-differential $D \ln \bar{p}(x, \omega)$ as a function on TM (Hamilton 1982, chapter I.3).

Remark 7 In Definition 6 we represent a tangent vector $V \in T_x M$ by the function $\partial_V \ln \bar{p}(x, \omega) \in L^1(\Omega, p(x))$. This representation is independent of the choice of a reference measure in $\mathcal{M}_+(\Omega, \mu)$, it depends only on the map $p : M \rightarrow \mathcal{M}_+(\Omega, \mu)$.

Definition 8 (Ay et al. 2015, Definition 2.2) A section τ of the bundle $T^*M \otimes_n$ times $\otimes T^*M$ is called a *weakly continuous covariant n -tensor*, if the value $\tau(V_n)$ is a continuous function for any continuous n -vector field V_n on M .

Definition 9 (Ay et al. 2015, Definition 2.1) A *covariant n -tensor field* on $\mathcal{M}(\Omega)$ assigns to each $\mu \in \mathcal{M}(\Omega)$ a multilinear map $\tau_\mu : \bigoplus^n L^n(\Omega, \mu) \rightarrow \mathbb{R}$ that is continuous w.r.t. the product topology on $\bigoplus^n L^n(\Omega, \mu)$.

Definition 10 (*Locality and continuity condition*) (Ay et al. 2015, Definition 2.8) Given a class $\{\Omega\}$ of measure spaces, a *statistical covariant continuous n -tensor field* A assigns to each parametrized measure model (M, Ω, μ, p) where $\Omega \in \{\Omega\}$ a *weakly continuous* (in the sense of Definition 8) covariant n -tensor field $A|_{(M, \Omega, \mu, p)}$ on M (cf. Definition 9). A statistical covariant continuous n -tensor field A is called *local* if there is a covariant n -tensor field \tilde{A} on $\mathcal{M}(\Omega)$ with the following property for any parametrized measure model (M, Ω, μ, p) and any $V_i \in T_x M$

$$A|_{(M, \Omega, \mu, p)}(V_1, \dots, V_n) = \tilde{A}_{p(x)}(\partial_{V_1} \ln \bar{p}(x), \dots, \partial_{V_n} \ln \bar{p}(x)). \tag{1}$$

From now on, if a weakly continuous covariant tensor A on a k -integrable statistical model (M, Ω, μ, p) satisfies (1) for $A|_{(M, \Omega, \mu, p)} = A$, we shall write $A = p^*(\tilde{A})$.

Example 11 (cf. Remark 22). In Ay et al. (2015) Ay–Jost–Lê–Schwachhöfer showed that the Fisher quadratic form

$$g^F(V, W)_x := \int_{\Omega} \partial_V \ln \bar{p}(x, \omega) \partial_W \ln \bar{p}(x, \omega) dp(x) \tag{2}$$

and the Amari–Chentsov 3-symmetric tensor

$$T^{AC}(V, W, X)_x := \int_{\Omega} \partial_V \ln \bar{p}(x, \omega) \partial_W \ln \bar{p}(x, \omega) \partial_X \ln \bar{p}(x, \omega) dp(x) \tag{3}$$

are local statistical continuous covariant tensor fields.

3 The monotonicity of the Fisher metric

In this section, we consider topological spaces Ω provided with Borel σ -algebra. We prove Theorem 2, and discuss some related problems (Remark 14).

Recall that a statistic $\kappa : \Omega_1 \rightarrow \Omega_2$ induces the linear operator $\kappa_* : L^1(\Omega_1, \mu_1) \rightarrow L^1(\Omega_2, \kappa_*(\mu_1))$ defined by Ay et al. (2015, (3.2))

$$\kappa_* f(y) := \frac{d\kappa_*(f \cdot \mu_1)}{d\kappa_*(\mu_1)}(y) \tag{4}$$

for $f \in L^1(\Omega_1, \mu_1)$ and $y \in \Omega_2$.

Remark 12 The operator κ_* is well defined, since by the Radon–Nikodym theorem, $f \in L^1(\Omega_1, \mu_1)$ if and only if $f \cdot \mu_1$ is a measure dominated by μ_1 , i.e., the null set of μ_1 is also a null set of $f \cdot \mu_1$. Now assume that $Z \subset \Omega_2$ is a null set of $\kappa_*(\mu_1)$. Then $\kappa^{-1}(Z)$ is also a null set of μ and hence of $f \cdot \mu_1$. It follows that Z is a null set of $\kappa_*(f \cdot \mu_1)$, and by the Radon–Nikodym theorem $\kappa_*(f \cdot \mu_1)$ is dominated by $\kappa_*(\mu_1)$.

Some time we will write $\kappa_*^{\mu_1}(f)$, if f may belong to $L^p(\Omega_1, \mu_1)$ for different μ_1 .

The following Lemma 13 is an expression of the well-known fact that condition expectation reduces the L^p -norm, see e.g. Neveu (1965, §4.3).

Lemma 13 *For all $p \geq 2$ we have $\kappa_*(L^p(\Omega_1, \mu_1)) \subset L^p(\Omega_2, \kappa_*(\mu_1))$. The linear map κ_* contracts L^p -norm:*

$$\|\kappa_*(f)\|_{L^p(\Omega_2, \kappa_*(\mu_1))} \leq \|f\|_{L^p(\Omega_1, \mu_1)}$$

for all $f \in L^p(\Omega_1, \mu_1)$.

Proof Let $f \in L^p(\Omega_1, \mu_1)$ and $y \in \Omega_2$. For a sequence of open sets $\Omega_2 = A_0 \supset \dots \supset A_n \supset \dots \ni y$ and a statistic $\kappa : \Omega_1 \rightarrow \Omega_2$ we set

$$|f|_n(y) := \frac{\int_{\kappa^{-1}(A_n)} |f(x)| d\mu_1}{\mu_1(\kappa^{-1}(A_n))}.$$

By the Hölder inequality, we have

$$(|f|_n(y))^p \leq \frac{\int_{\kappa^{-1}(A_n)} |f(x)|^p d\mu_1}{\mu_1(\kappa^{-1}(A_n))}.$$

Since $\lim_{n \rightarrow \infty} |f|_n(y) = \kappa_*(|f|)(y)$, we deduce from the above inequality

$$(\kappa_*(|f|)(y))^p \leq \kappa_*(|f|^p)(y). \tag{5}$$

Using (5), we obtain

$$\begin{aligned} \|\kappa_*(f)\|_{L^p(\Omega_2, \kappa_*(\mu_1))}^p &= \int_{\Omega_2} |\kappa_*(f)|^p \, d\kappa_*(\mu_1) \\ &\leq \int_{\Omega_2} (\kappa_*(|f|))^p \, d\kappa_*(\mu_1) \leq \int_{\Omega_2} \kappa_*(|f|^p) \, d\kappa_*(\mu_1) = (\|f\|_{L^p(\Omega_1, \mu_1)})^p, \end{aligned}$$

which implies immediately Lemma 13. □

Proof of Theorem 2 By Remark 7 the geometry of a parametrized measure model $(M, \Omega_1, \mu_1, p_1)$ does not depend on the choice of a reference measure μ_1 . Thus, to prove Theorem 2 at a point $x \in M$, we can assume that $p_1(x) = \mu_1$ and hence $\bar{p}_1(x, \omega) = 1$. Abusing the notation, for a function $\bar{p} : M \times \Omega \rightarrow \mathbb{R}$ and for $x \in M$, we denote by $\bar{p}(x)$ the function $\Omega \rightarrow \mathbb{R}$ such that $\bar{p}(x)(\omega) = \bar{p}(x, \omega)$. Then we have $\kappa_*^{\mu_1}(\bar{p}_1(x))(\kappa(\omega)) = 1$ for all ω . Now let $V \in T_x M$. Then we have

$$\partial_V (\ln \kappa_*^{\mu_1}(\bar{p}_1(x))) = \partial_V \kappa_*^{\mu_1}(\bar{p}_1(x)). \tag{6}$$

Next, we shall prove the following equality

$$\partial_V \kappa_*^{\mu_1}(\bar{p}_1(x)) = \kappa_*^{\mu_1}(\partial_V \bar{p}_1(x)). \tag{7}$$

To prove (7) it suffices to show that the following equality holds

$$\partial_V \kappa_*^{\mu_1}(p_1(x)) = \kappa_*^{\mu_1}(\partial_V p_1(x)), \tag{8}$$

where the RHS and LHS of (8) are understood as signed measures.

The condition (2) in Definition 6 implies that $\partial_V \bar{p}_1 \in L^2(\Omega_1, \mu_1) \subset L^1(\Omega_1, \mu_1)$, since $(M, \Omega_1, \mu_1, p_1)$ is a 2-integrable parametrized measure model. Hence, for any measurable subset A in Ω_2 , we can apply the differentiation under integral [see e.g. Jost (2005, Theorem 16.11, p. 213)] to obtain the following [cf. Ay et al. (2015, (2.6)) for a slightly different proof]

$$\partial_V \int_{\kappa^{-1}(A)} \bar{p}_1 \, d\mu_1 = \int_{\kappa^{-1}(A)} \partial_V \bar{p}_1 \, d\mu_1.$$

This equality implies (8) immediately. Hence (7) holds.

Let us continue the proof of Theorem 2. Using (7), and recalling that $\bar{p}_1(x, \omega) = 1$, we obtain from (6)

$$\partial_V (\ln \kappa_*^{\mu_1}(\bar{p}_1(x))) = \kappa_*^{\mu_1}(\partial_V \bar{p}_1(x)) = \kappa_*^{\mu_1}(\partial_V \ln \bar{p}_1(x)).$$

Since $\partial_V \ln \bar{p}_1(x) \in L^2(\Omega_1, p_1(x))$ for all $x \in M$, by Lemma 13, we obtain

$$\|\kappa_*^{\mu_1}(\partial_V \ln \bar{p}_1(x))\|_{L^2(\Omega_2, \kappa_*^{\mu_1}(p_1(x)))} \leq \|\partial_V \ln \bar{p}_1(x)\|_{L^2(\Omega_1, p_1(x))}. \tag{9}$$

Noting that the RHS of (9) is equal to $g_{(M, \Omega_1, \mu_1, p_1)}^F$ and the LHS of (9) is equal to $g_{(M, \Omega_2, \kappa_*(\mu_1), \kappa_*(p_1))}^F$, we deduce Theorem 2 from (9). \square

Remark 14 1. It is not hard to see that if Ω_1, Ω_2 are metric topological spaces, κ and f are continuous, then the inequality (9) becomes an equality if and only if $f(\omega) = \kappa_*(f)(\kappa(\omega))$ for all ω .

2. Lemma 13 implies that the absolute value \hat{T}^{AC} of the Amari–Chentsov tensor defined by $\hat{T}^{AC}(V) := |T^{AC}(V, V, V)|$ for $V \in TM$ also satisfies the version of Definition 1 on statistical fields which measure “information loss”.

4 Mixed topology and strongly continuous covariant tensor fields

In this section, we assume that Ω is a separable metric topological space provided with Borel σ -algebra. Let $\mathbb{R}_{\geq 0}^n := [0, \infty)^n$. We introduce a mixed topology on the spaces

$$\mathcal{L}_n^n(\Omega) := \cup_{\mu \in \mathcal{M}(\Omega)} \oplus^n L^n(\Omega, \mu) \quad \text{and} \quad \mathcal{L}_1^n(\Omega) := \cup_{\mu \in \mathcal{M}(\Omega)} L^n(\Omega, \mu)$$

which has good properties (Proposition 17). Using the mixed topology, we introduce the notion of strongly continuous covariant n -tensor fields on $\mathcal{M}(\Omega)$ (Definition 18), whose examples are the Fisher quadratic form (Remark 22) and all continuous functions on $\mathcal{L}_k^k(\Omega_n)$ (Example 19), where Ω_n is a finite sample space consisting of n elementary events.

4.1 Mixed topology on $\mathcal{L}_n^n(\Omega)$

It is known that $\mathcal{M}(\Omega)$ possesses many different important topologies, e.g., the total variation topology, the strong topology and the weak topology. The total variation is used in Definition 6. Now we recall the notion of weak topology on $\mathcal{M}(\Omega)$, which plays prominent role in measure theory and especially in probability theory (Bogachev 2007; Billingsley 1999). Denote by $C_b(\Omega)$ the space of all bounded continuous real functions on Ω .

Definition 15 (cf. Bogachev 2007, vol. II, Definition 8.1.1) A sequence of Borel measures μ_α on Ω is called *weakly convergent* to a Borel measure μ (writing as $\mu_\alpha \implies \mu$) if for every $f \in C_b(\Omega)$ one has

$$\lim_\alpha \int_\Omega f \, d\mu_\alpha = \int_\Omega f \, d\mu.$$

It is known that the weak topology on $\mathcal{M}(\Omega)$ is generated by fundamental neighborhoods of $\mu, \mu \in \mathcal{M}(\Omega)$, defined as follows (Bogachev 2007, Definition 8.1.2)

$$U_{f_1, \dots, f_k, \varepsilon}(\mu) := \left\{ \nu : \left| \int_\Omega f_i \, d\mu - \int_\Omega f_i \, d\nu \right| < \varepsilon \text{ for } i \in [1, k] \right\}, \quad (10)$$

where $f_i \in C_b(\Omega), k \in \mathbb{N}$ and $\varepsilon > 0$.

- Remark 16* 1. The weak topology on $\mathcal{M}(\Omega)$ is weaker than the total variation topology, hence for any k -integrable parametrized measure model (M, Ω, μ, p) the embedding $p : M \rightarrow \mathcal{M}_+(\Omega, \mu) \rightarrow \mathcal{M}(\Omega)$ is also continuous with respect to the weak topology on $\mathcal{M}(\Omega)$.
2. Since Ω is a separable metric topological space, for each $\mu \in \mathcal{M}(\Omega)$ the subspace $C_b(\Omega)$ is a dense subset in $L^n(\Omega, \mu)$ with respect to the $L^n(\Omega, \mu)$ -topology (Adams and Fournier 2006; Jost 2005; Bogachev 2007).

Let us denote by $\mathcal{L}_n^n(\Omega)$ the fibration over $\mathcal{M}(\Omega)$ whose fiber over $\mu \in \mathcal{M}(\Omega)$ is the space $\oplus^n L^n(\Omega, \mu)$. Note that the product topology on $\oplus^n L^n(\Omega, \mu)$ is generated by the product norm defined as follows. For $\vec{f} = (f_1, \dots, f_n) \in \oplus^n L^n(\Omega, \mu)$ let

$$\|\vec{f}\|_{L_n^n(\mu)} := \sum_{i=1}^n \|f_i\|_{L^n(\Omega, \mu)}.$$

Denote by π the projection $\mathcal{L}_n^n(\Omega) \rightarrow \mathcal{M}(\Omega)$.

We are going to define a topology on $\mathcal{L}_n^n(\Omega)$ by specifying its base. For an n -tuple of functions $\vec{f} \in \oplus^n C_b(\Omega) = (C_b(\Omega))^n$, an open set $U \subset \mathcal{M}(\Omega)$ in the weak topology and $\varepsilon > 0$ we set

$$O(\vec{f}, U, \varepsilon) := \{[\vec{g}, \mu] : \mu \in U, \vec{g} \in \oplus^n L^n(\Omega, \mu) \text{ and } \|\vec{g} - \vec{f}\|_{L_n^n(\mu)} < \varepsilon\}, \tag{11}$$

where $[\vec{g}, \mu]$ means a pair.

Note that

$$O(\vec{f}, \cup_i U_i, \varepsilon) = \cup_i O(\vec{f}, U_i, \varepsilon) \quad \text{and} \quad O(\vec{f}, \cap_i U_i, \varepsilon) = \cap_i O(\vec{f}, U_i, \varepsilon). \tag{12}$$

Proposition 17 *The collection B of all subsets $O(\vec{f}, U, \varepsilon)$ where $\vec{f} \in (C_b(\Omega))^n$, U is open set in $\mathcal{M}(\Omega)$ and $\varepsilon > 0$ generates a unique topology on $\mathcal{L}_n^n(\Omega)$, which we shall call the mixed topology. Furthermore, the restriction of this topology to each fiber $\oplus^n L^n(\Omega, \mu)$ is equal to the $L^n(\Omega, \mu)$ -topology on the fiber. Consequently, the space $(C_b(\Omega))^n \times \mathcal{M}(\Omega)$ is a dense subset in the mixed topology. The projection $\pi : \mathcal{L}_n^n(\Omega) \rightarrow \mathcal{M}(\Omega)$ is continuous with respect to the mixed topology on the domain and the weak topology on the target space.*

Proof To prove the first assertion of Proposition 17 it suffices to show that the following conditions hold.

1. The (base) elements in B cover $\mathcal{L}_n^n(\Omega)$.
2. Let $O(\vec{f}_1, U_1, \varepsilon_1)$ and $O(\vec{f}_2, U_2, \varepsilon_2)$ be base elements. If their intersection I is non-empty, then for each $[\vec{f}, \mu] \in I$, there is a base element $O(\vec{f}_3, U_3, \varepsilon_3)$ such that $[\vec{f}, \mu] \in O(\vec{f}_3, U_3, \varepsilon_3) \subset I$.

The first condition (1) holds by Remark 16.2.

Now let us prove that (2) holds. For $\vec{f} \in \oplus^n L^n(\Omega, \mu)$ we set

$$B(\vec{f}, \varepsilon, \mu) := \{\vec{f}' \in \oplus^n L^n(\Omega, \mu) \text{ and } \|\vec{f}' - \vec{f}\|_{L_n^n(\Omega, \mu)} < \varepsilon\}.$$

Note that $I \cap \pi^{-1}(\mu)$ is an open subset of $\pi^{-1}(\mu)$ in $L^n(\Omega, \mu)$ -topology, since it is the intersection of two open balls $B(\vec{f}_1, \varepsilon_1, \mu)$ and $B(\vec{f}_2, \varepsilon_2, \mu)$. Using (12), we can assume w.l.o.g.

$$\begin{aligned} U_1 &= U_{\tilde{f}_1, \dots, \tilde{f}_k, \varepsilon_1}(\mu_1), \\ U_2 &= U_{\tilde{g}_1, \dots, \tilde{g}_m, \varepsilon_2}(\mu_2). \end{aligned}$$

Let δ_1 be a number such that

$$U_{\tilde{f}_1, \dots, \tilde{f}_k, \tilde{g}_1, \dots, \tilde{g}_m, \delta_1}(\mu) \subset U_1 \cap U_2, \tag{13}$$

and moreover $\delta_1 \leq \min\{1, \varepsilon_1, \varepsilon_2\}$. Next we choose a positive number $\delta_2 \leq \delta_1$ such that

$$\|\vec{f} - \vec{f}_1\|_{L^n_\mu(\mu)} < \varepsilon_1 - \delta_2 \quad \text{and} \quad \|\vec{f} - \vec{f}_2\|_{L^n_\mu(\mu)} < \varepsilon_2 - \delta_2. \tag{14}$$

Then we choose $[\vec{f}_3, \mu] \in I \cap \pi^{-1}(\mu)$ with the following properties

$$\vec{f}_3 \in (C_b(\Omega))^n \quad \text{and} \quad \|\vec{f}_3 - \vec{f}\|_{L^n_\mu(\mu)} < \frac{1}{4}\delta_2. \tag{15}$$

We obtain from (14) and (15)

$$\|\vec{f}_3 - \vec{f}_i\|_{L^n_\mu(\mu)} < \varepsilon_i - \frac{3}{4}\delta_2 \quad \text{for } i = 1, 2. \tag{16}$$

We write $\vec{f}_3 = (f_3^1, \dots, f_3^n)$. Note that $|f_3^i - f_1^i|^n$ and $|f_3^i - f_2^i|^n$ are continuous bounded functions on Ω for all $i \in [1, n]$. Now we set

$$U_3 := U_{\tilde{f}_1, \dots, \tilde{f}_k, \tilde{g}_1, \dots, \tilde{g}_m, |f_3^i - f_1^i|^n, |f_3^i - f_2^i|^n, i \in [1, n], (\frac{1}{8}\delta_2)^n}(\mu). \tag{17}$$

Since $\delta_2 \leq \delta_1$ we obtain from (17) and (13)

$$U_3 \subset U_{\tilde{f}_1, \dots, \tilde{f}_k, \tilde{g}_1, \dots, \tilde{g}_m, \delta_1}(\mu) \subset U_1 \cap U_2.$$

Clearly, (15) implies that $[\vec{f}, \mu] \in O(\vec{f}_3, U_3, \frac{1}{4}\delta_2)$. Hence, setting $\varepsilon_3 := \frac{1}{4}\delta_2$, to complete the proof of the first assertion of Proposition 17, it suffices to show that

$$O\left(\vec{f}_3, U_3, \frac{1}{4}\delta_2\right) \subset I. \tag{18}$$

Let $[\vec{h}, \mu'] \in O(\vec{f}_3, U_3, \frac{1}{4}\delta_2)$. To prove (18) we need to show that $[\vec{h}, \mu'] \in I$, or equivalently

$$[\vec{h}, \mu'] \in O(\vec{f}_i, U_i, \varepsilon_i) \quad \text{for } i = 1, 2. \tag{19}$$

Since $\mu' \in U_3 \subset U_i$ for $i = 1, 2$, (19) is equivalent to

$$\|\vec{h} - \vec{f}_i\|_{L^n_\mu(\mu')} < \varepsilon_i \quad \text{for } i = 1, 2. \tag{20}$$

Taking into account $[\vec{h}, \mu'] \in O(\vec{f}_3, U_3, \frac{1}{4}\delta_2)$, we obtain

$$\|\vec{h} - \vec{f}_3\|_{L^n_\mu(\mu')} < \frac{1}{4}\delta_2. \tag{21}$$

Since $\mu' \in U_3$, we derive from (16) and (17)

$$\|\vec{f}_3 - \vec{f}_1\|_{L^n_\mu(\mu')} < \|\vec{f}_3 - \vec{f}_1\|_{L^n_\mu(\mu)} + \frac{1}{8}\delta_2 < \varepsilon_1 - \frac{5}{8}\delta_2. \tag{22}$$

In the same way we obtain

$$\|\vec{f}_3 - \vec{f}_2\|_{L^n_\mu(\mu')} < \varepsilon_2 - \frac{5}{8}\delta_2. \tag{23}$$

Clearly, (21), (22), and (23) imply (20). This proves the first assertion of Proposition 17.

The second assertion of Proposition 17 follows from Remark 16.2, observing that a ball $B(\vec{f}, \varepsilon, \mu)$ is the intersection of the open set $O(\vec{f}, U(\mu), \varepsilon)$ with the fiber $\oplus^n L^n(\Omega, \mu)$.

Finally, the last assertion is obvious, since the preimage $\pi^{-1}(U)$ of an open set $U \subset \mathcal{M}(\Omega)$ is the union of all open sets of the form $O(f, U, \varepsilon)$, $f \in (C_b(\Omega))^n$ and $\varepsilon > 0$. This completes the proof of Proposition 17. □

4.2 Strongly continuous covariant n -tensor on $\mathcal{M}(\Omega)$

Definition 18 A covariant n -tensor field on $\mathcal{M}(\Omega)$ is called *strongly continuous*, if it is a continuous function on $\mathcal{L}^n_n(\Omega)$ with respect to the mixed topology.

Example 19 Let $\Omega_n := \{\omega_1, \dots, \omega_n\}$ be a finite sample space of n elementary events. Let δ_{ω_i} denote the Dirac measure concentrated at ω_i . Let $\mu_l = \sum_{i=1}^l c_i \delta_{\omega_i} \in \mathcal{M}(\Omega_n)$, where $l \leq n$ and $c_i > 0$. Then, for all $k \geq 1$, $L^k(\Omega_n, \mu_l)$ is homeomorphic to $C_b(\Omega_l) = \mathbb{R}^l$, which is provided with the usual (vector space) topology. Furthermore, the weak topology on $\mathcal{M}(\Omega_n) = \mathbb{R}^n_{\geq 0}$ coincides with the usual topology on $\mathbb{R}^n_{\geq 0} \subset \mathbb{R}^n$. Hence the subset $\mathcal{M}_+(\Omega_n)$ consisting of positive measures on Ω_n is dense in $\mathcal{M}(\Omega_n)$. We observe that $\pi : \mathcal{L}^k_k(\Omega_n) \rightarrow \mathcal{M}(\Omega_n)$ is a fiber bundle whose fiber over μ_l is homeomorphic to $(\mathbb{R}^l)^k$. A covariant k -tensor field \tilde{F} on $\mathcal{M}(\Omega_n) = \mathbb{R}^n_{\geq 0}$ is a continuous function on $\mathcal{L}^k_k(\Omega_n)$. Since $\pi^{-1}(\mathcal{M}_+(\Omega_n))$ is open and dense in $\mathcal{L}^k_k(\Omega_n)$, the function \tilde{F} is defined uniquely by its restriction to $\pi^{-1}(\mathcal{M}_+(\Omega_n))$. In particular, the Fisher metric defined on $\mathcal{M}_+(\Omega_n)$ is associated with the quadratic form $\tilde{g}^F : \mathcal{L}^2_2(\Omega_n) \rightarrow \mathbb{R}$ defined by $\tilde{g}^F([f_1, f_2, \mu]) = \int_{\Omega_n} f_1 \cdot f_2 d\mu$, see also Remark 22.

Proposition 20 *Let $g \in C_b(\Omega)$ and $c : \mathcal{M}(\Omega) \rightarrow \mathbb{R}$ be a continuous function with respect to the weak topology. We define a covariant n -tensor field $T_{g,c}$ on $\mathcal{M}(\Omega)$ by setting*

$$T_{g,c}([f_1, \dots, f_n, \mu]) := c(\mu) \cdot \int_{\Omega} g \cdot f_1 \cdots f_n \, d\mu.$$

Then $T_{g,c}$ is a strongly continuous covariant n -tensor field on $\mathcal{M}(\Omega)$.

Proof By Proposition 17, $\pi : \mathcal{L}_n^n(\Omega) \rightarrow \mathcal{M}(\Omega)$ is a continuous function, hence $c(\mu)$ is a continuous function on $\mathcal{L}_n^n(\Omega)$. Thus to prove Proposition 20 it suffices to assume that $c(\mu) = 1$, i.e., it suffices to show that $T_{g,1}$ descends to a continuous function on $\mathcal{L}_n^n(\Omega)$ provided with the mixed topology. Equivalently, we need to show that the set

$$O(a, b) := \{[\vec{f}, \mu] \in \mathcal{L}_n^n(\Omega) \mid a < T_{g,1}(\vec{f}, \mu) < b\}$$

is an open set in the mixed topology for any $-\infty < a < b < \infty$.

Let $[\vec{f}, \mu] \in O(a + \varepsilon, b - \varepsilon)$, where $\varepsilon < \frac{1}{4}(b - a)$. We will show that there is an open set $O(\vec{f}_1, U_1, \delta) \ni [\vec{f}, \mu]$ such that

$$T_{g,1}(O(\vec{f}_1, U_1, \delta)) \subset (a, b). \tag{24}$$

Lemma 21 *The restriction of $T_{g,1}$ to each fiber $\oplus^n L^n(\Omega, \mu)$ is continuous in the product $L^n(\Omega, \mu)$ -topology. Moreover, if $\|\vec{h} - \vec{f}\|_{L^n(\mu)} \leq 1$ then*

$$\begin{aligned} |T_{g,1}([\vec{f}, \mu]) - T_{g,1}([\vec{h}, \mu])| &\leq \sup_{\Omega} g(\omega) \cdot 2^n \cdot \|\vec{h} - \vec{f}\|_{L^n(\mu)} \\ &\cdot \left(1 + \sum_{i=1}^n \sum_{k=1}^n \|f_i\|_{L^n(\Omega, \mu)}^k \right). \end{aligned}$$

Proof Write $\vec{f} - \vec{h} = \vec{a} = (a_1, \dots, a_n)$. Expanding $h_1 \cdots h_n = \prod_{i=1}^n (f_i - a_i)$ and using Holder’s inequality, we obtain

$$\begin{aligned} &|T_{g,1}([\vec{f}, \mu]) - T_{g,1}([\vec{h}, \mu])| \\ &\leq \sup_{\Omega} g(\omega) \cdot \sum_{[1,n]=\{i_1, \dots, i_k\} \cup \{j_1, \dots, j_{n-k}\}} \int_{\Omega} |a_{i_1} \cdots a_{i_k} f_{j_1} \cdots f_{j_{n-k}}| \, d\mu \\ &\leq \sup_{\Omega} g(\omega) \cdot 2^n \cdot \max_{[1,n]=\{i_1, \dots, i_k\} \cup \{j_1, \dots, j_{n-k}\}} \|a_{i_1}\|_{L^n(\Omega, \mu)} \cdots \|f_{j_{n-k}}\|_{L^n(\Omega, \mu)}. \end{aligned} \tag{25}$$

Note that in (25) the set $\{j_1, \dots, j_{n-k}\}$ may be empty but the set $\{i_1, \dots, i_k\}$ is always non-empty. Since $\sum_{i=1}^n \|a_i\|_{L^n(\Omega, \mu)} \leq 1$, we have

$$\begin{aligned} & \max_{[1,n]=\{i_1,\dots,i_k\}\cup\{j_1,\dots,j_{n-k}\}} \|a_{i_1}\|_{L^n(\Omega,\mu)} \cdots \|f_{j_{n-k}}\|_{L^n(\Omega,\mu)} \\ & \leq \sum_{i=1}^n \|a_i\|_{L^n(\Omega,\mu)} \left(1 + \sum_{i=1}^n \sum_{k=1}^n \|f_i\|_{L^n(\Omega,\mu)}^k \right). \end{aligned} \tag{26}$$

Clearly Lemma 21 follows from (25) and (26). □

Continuation of the proof of Proposition 20. We define a function $G : \mathcal{L}_n^n(\Omega) \rightarrow \mathbb{R}$ by setting

$$G([\vec{f}, \mu]) := \sup_{\Omega} g(\omega) \cdot 2^n \left(1 + \sum_{i=1}^n \sum_{k=1}^n \|f_i\|_{L^n(\Omega,\mu)}^k \right).$$

Let us pick an element $\vec{f}_1 = ((f_1)_1, \dots, (f_1)_n) \in (C_b(\Omega))^n \cap B(\vec{f}, \delta, \mu)$ where δ is so small such that the following equalities hold:

$$\delta < \min \left\{ \frac{1}{2}, \frac{\varepsilon}{16G([\vec{f}, \mu])} \right\}, \tag{27}$$

$$|T_{g,1}(\vec{f}_1, \mu) - T_{g,1}(\vec{f}, \mu)| < \frac{\varepsilon}{16}, \tag{28}$$

and

$$|G([\vec{h}, \mu]) - G([\vec{f}_1, \mu])| \leq \frac{\varepsilon}{8} \tag{29}$$

for all $h \in B(\vec{f}_1, \delta, \mu)$. The existence of δ follows from the positivity of G , from Lemma 21 and from the continuity of the restriction of G to each fiber $\oplus^n L^n(\Omega, \mu)$.

We define a neighborhood $U_1 = U_1([\vec{f}_1, \mu])$ containing μ as follows:

$$U_1 := U_{(g \cdot f_1 \cdots f_n), |(f_1)_1|^n, \dots, |(f_1)_n|^n, \lambda}(\mu),$$

where λ depends on g, \vec{f}_1, μ and is so small such that

$$\lambda < \frac{\varepsilon}{8} \tag{30}$$

and for $\mu' \in U_1$ we have

$$|G([\vec{f}_1, \mu']) - G([\vec{f}_1, \mu])| \leq \frac{\varepsilon}{8}. \tag{31}$$

The existence of λ satisfying (31) is ensured by the continuity of the function $G([\vec{f}_1, \mu])$ in variable μ .

Now we shall show that $O(\vec{f}_1, U_1, \delta) \ni [\vec{f}, \mu]$ satisfies (24). Assume that $[\vec{h}, \mu'] \in O(\vec{f}_1, U_1, \delta)$. Then

$$|T_{g,1}([\vec{h}, \mu']) - T_{g,1}([\vec{f}, \mu])| \leq |T_{g,1}([\vec{h}, \mu']) - T_{g,1}([\vec{f}_1, \mu'])| + |T_{g,1}([\vec{f}_1, \mu']) - T_{g,1}([\vec{f}_1, \mu])| + |T_{g,1}([\vec{f}_1, \mu]) - T_{g,1}([\vec{f}, \mu])|. \tag{32}$$

Let us estimate the first term in the RHS of (32). By Lemma 21 we have

$$|T_{g,1}([\vec{h}, \mu']) - T_{g,1}([\vec{f}_1, \mu'])| \leq \|\vec{h} - \vec{f}_1\|_{L^n(\mu')} \cdot G([\vec{f}_1, \mu']). \tag{33}$$

Taking into account (31), (29), and the choice of δ in (27), we obtain from (33), noting that $\vec{f}_1 \in B(\vec{f}, \delta, \mu) \implies \vec{f} \in B(\vec{f}_1, \delta, \mu)$:

$$|T_{g,1}([\vec{h}, \mu']) - T_{g,1}([\vec{f}_1, \mu'])| \leq \delta \cdot G([\vec{f}_1, \mu']) \leq \delta \left(\frac{\varepsilon}{8} + G([\vec{f}_1, \mu]) \right) \leq \delta \left(\frac{\varepsilon}{8} + \frac{\varepsilon}{8} + G([\vec{f}, \mu]) \right) < \frac{3\varepsilon}{16}. \tag{34}$$

We estimate the second term in the RHS of (32) as follows, using the fact $\mu' \in U_1 = U_1([\vec{f}_1, \mu])$ with λ satisfying (30):

$$|T_{g,1}([\vec{f}_1, \mu']) - T_{g,1}([\vec{f}_1, \mu])| < \lambda < \frac{\varepsilon}{8}. \tag{35}$$

Using (34), (35) and estimating the last term in the RHS of (32) by (28), we obtain from (32)

$$|T_{g,1}([\vec{h}, \mu']) - T_{g,1}([\vec{f}, \mu])| \leq \frac{3\varepsilon}{16} + \frac{\varepsilon}{8} + \frac{\varepsilon}{16} = \frac{3\varepsilon}{8}. \tag{36}$$

Equation (36) implies that $T_{g,1}([\vec{h}, \mu']) \in (a, b)$. Hence (24) holds. The proof of Proposition 20 is completed. \square

Remark 22 Let $1_\Omega : \Omega \rightarrow \mathbb{R}$ denote the constant function taking the value 1. Then $1_\Omega \in C_b(\Omega)$. Let (M, Ω, μ, p) be a 2-integrable parametrized measure model. By (1) the 2-tensor field $T_{1_\Omega,1}$ induces the following local statistical 2-tensor g on (M, Ω, μ, p) :

$$\begin{aligned} g_x(V, W) &= (T_{1_\Omega,1})_{p(x)}(\partial_V \ln \bar{p}(x), \partial_W \ln \bar{p}(x)) \\ &= \int_{\Omega} \partial_V \ln \bar{p}(x) \cdot \partial_W \ln \bar{p}(x) \, dp(x). \end{aligned} \tag{37}$$

The RHS of (37) is the Fisher metric g^F . Thus, the Fisher metric is induced from the strongly continuous covariant 2-tensor field $T_{1_\Omega,1}$ on $\mathcal{M}(\Omega)$. In the same way, the Amari–Chentsov tensor T^{AC} is induced from the strongly continuous covariant 3-tensor field $T_{1_\Omega,1}$ on $\mathcal{M}(\Omega)$.

5 The uniqueness of the Fisher metric

Recall that δ_ω denotes the Dirac measure concentrated at $\omega \in \Omega$.

Lemma 23 (cf. Bogachev 2007, Example 8.1.6) *The set of all measures of the form $\sum_{i=1}^N c_i \delta_{\omega_i}$, $c_i > 0$, is dense in $\mathcal{M}(\Omega)$ in the weak topology. The convex hull of the set of Dirac measures is dense in the space $\mathcal{P}(\Omega)$.*

Proof 1. A version of Lemma 23 for finite Baire measures is proved in Bogachev (2007, Example 8.1.6). We apply Bogachev’s argument for the proof of Lemma 23. Suppose we are given a neighborhood $U \ni \mu$ of the form (10). We may assume that the total variation norm $\|\mu\| \leq 1$. There are simple (step) functions g_i such that $\sup_{\omega \in \Omega} |f_i(\omega) - g_i(\omega)| < \varepsilon/4$ for all $i \in [1, k]$. To prove Lemma 23 it suffices to show that U contains a measure $\nu = \sum_{i=1}^N c_i \delta_{\omega_i}$ such that for all $i \in [1, k]$ we have

$$\int_{\Omega} g_i \, d\mu = \int_{\Omega} g_i \, d\nu. \tag{38}$$

Let $\Omega = \cup_{j=1}^{n_i} A_i^j$ be a finite partition into disjoint measurable sets corresponding to g_i , i.e., $g_i = \sum a_i^j \chi_{A_i^j}$. Then

$$\Omega = \cup_{l_1, \dots, l_k} A_1^{l_1} \cap A_2^{l_2} \cap \dots \cap A_k^{l_k}$$

is a finite partition corresponding to g_i for all $i \in [1, k]$. Set $c_{l_1 \dots l_k} := \mu(A_1^{l_1} \cap A_2^{l_2} \cap \dots \cap A_k^{l_k})$ and let $\omega_{l_1 \dots l_k}$ be a point in $A_1^{l_1} \cap A_2^{l_2} \cap \dots \cap A_k^{l_k}$. Then (38) holds for $\nu = \sum_{l_1, \dots, l_k} c_{l_1 \dots l_k} \delta_{\omega_{l_1 \dots l_k}}$. Since μ is a non-negative measure, we have $c_{l_1 \dots l_k} \geq 0$. This completes the proof of the first assertion of Lemma 23.

2. The second assertion follows immediately, since by the above construction of $\sum_{i=1}^N c_i \delta_{\omega_i}$ we have $\sum c_i = \mu(\Omega)$. □

Proof of Theorem 4 Assume that F is a metric defined on all 2-integrable statistical models (M, Ω, μ, p) that satisfies the condition of Theorem 4 and \tilde{F}_Ω denotes the associated strongly continuous quadratic form on $\mathcal{M}(\Omega)$. Denote by \tilde{g}_Ω^F the quadratic form on $\mathcal{M}(\Omega)$ that is associated with the Fisher metric g^F . We shall show that $\tilde{F}_\Omega = c \cdot \tilde{g}_\Omega^F$ for some constant c .

By Proposition 25 it suffices to consider the case Ω is non-discrete. Let $\kappa_n : \Omega \rightarrow \Omega_n$ be a statistic such that $\kappa_n(\Omega) = \Omega_n$. Let us choose points $\{\omega_1, \dots, \omega_n\} \in \Omega$ such that $\kappa_n(\omega_i)$ are distinct points in Ω_n . Let us consider the following map

$$\Omega_n \xrightarrow{i_n} \Omega \xrightarrow{\kappa_n} \Omega_n,$$

where i_n identifies $\kappa_n(\omega_i)$ with ω_i for all $i \in [1, n]$. Note that i_n is also a statistic and $\kappa_n \circ i_n = \text{Id}$. Let $\mu_n^+ \in \mathcal{P}_+(\Omega_n)$. Observe that $(\mathcal{P}_+(\Omega_n), \Omega_n, \mu_n^+, \text{Id})$ is a 2-integrable statistical model. By the monotonicity assumption of F , and using $\kappa_n \circ$

$i_n = \text{Id}$, we conclude that the metric F defined on the 2-integrable statistical model $(\mathcal{P}_+(\Omega_n), \Omega, (i_n)_*(\mu_n^+), (i_n)_*(\text{Id}))$ is defined uniquely by the metric F defined on the 2-integrable statistical model $(\mathcal{P}_+(\Omega_n), \Omega_n, \mu_n^+, \text{Id})$. By Proposition 25 the metric F defined on the 2-integrable statistical model $(\mathcal{P}_+(\Omega_n), \Omega_n, \mu_n^+, \text{Id})$ coincides with the Fisher metric up to a multiplicative constant c . Hence, the restriction of \tilde{F}_Ω to the subspace of $\mathcal{L}_2^2(\Omega)$

$$\mathcal{L}_2^2(\omega_1, \dots, \omega_n) : \left\{ [f_1, f_2, \mu_n] \in \mathcal{L}_2^2(\Omega) \mid \mu_n = \sum_{i=1}^n c_i \delta_{\omega_i}, c_i > 0 \right\}$$

coincides with the restriction of \tilde{g}_Ω^F up to the multiplicative constant c , since \tilde{F}_Ω is strongly continuous.

Now we shall show that the constant c does not depend on the choice of a collection $\{\omega_1, \dots, \omega_n\}$. Let $\{\omega'_1, \dots, \omega'_m\}$ be another collection of distinct m points on Ω . Let $\Omega_N := \{\omega''_1, \dots, \omega''_N\}$ be the union of $\{\omega_1, \dots, \omega_n\}$ and $\{\omega'_1, \dots, \omega'_m\}$. We consider the following sequence of statistics

$$\Omega_n \xrightarrow{i_{n,N}} \Omega_N \xrightarrow{i_N} \Omega \xrightarrow{\kappa_N} \Omega_N \xrightarrow{\kappa_{N,n}} \Omega_n,$$

where $i_{n,N}$ and i_N are the natural embeddings and κ_N and $\kappa_{N,n}$ are sufficient statistics such that $\kappa_N \circ i_N = \text{Id}$ and $\kappa_{N,n} \circ i_{n,N} = \text{Id}$. By Proposition 25, the constant c that depends on $\{\omega_1, \dots, \omega_n\}$ equals the constant c'' that depends on $\{\omega''_1, \dots, \omega''_N\}$. In the same way we prove that the constant c' that depends on $\{\omega'_1, \dots, \omega'_m\}$ equals the constant c'' that depends on $\{\omega''_1, \dots, \omega''_N\}$. Hence the constant c does not depend on the choice of $\{\omega_1, \dots, \omega_n\}$.

We denote by $\mathcal{D}^+(\Omega)$ the set of all measures $\mu_n = \sum_{i=1}^n c_i \delta_{\omega_i}, c_i > 0$, where $\omega_i \in \Omega$. By Lemma 23 the subset

$$\mathcal{L}_2^2(\Omega, \mathcal{D}^+) := \{[f_1, f_2, \mu] \in \mathcal{L}_2^2(\Omega) \mid \mu \in \mathcal{D}^+(\Omega)\}$$

is dense in $\mathcal{L}_2^2(\Omega)$ in the mixed topology. Since the restriction of \tilde{F}_Ω to $\mathcal{L}_2^2(\Omega, \mathcal{D}^+)$ coincides with the restriction of \tilde{g}_Ω^F up to the multiplicative constant c , taking into account the strong continuity of \tilde{F}_Ω , this completes the proof of Theorem 4. \square

Acknowledgements The author thanks Shun-ichi Amari, Nihat Ay, Lorenz Schwachhöfer and Alesha Tuzhilin for valuable conversations. She is grateful to Vladimir Bogachev and Jürgen Jost for their helpful comments and suggestions. The final version of this manuscript is greatly improved thanks to critical helpful suggestions of the anonymous referees. She acknowledges the VNU for Sciences in Hanoi for excellent working conditions and financial support during her visit when a part of this note has been done.

Appendix: The Chentsov uniqueness theorem

In this appendix we recall a reformulation of the Chentsov theorem (Chentsov 1978, Theorem 11.1, p. 159) on the uniqueness of the Fisher metric in the language of information geometry by Amari and Nagaoka (Proposition 24), which is simpler than the original formulation by Chentsov in the category language. In Proposition 25 we

formulate a result in Ay et al. (2015) that characterizes the Fisher metric on finite sample spaces via the monotonicity. Then we discuss in Remark 26 some problems in generalizing the Chentsov theorem to a larger class of measure spaces that contains also non-discrete measure spaces.

Let us denote by $\mathcal{P}_+(\Omega_n)$ the subset of $\mathcal{P}(\Omega_n)$ that consists of positive measures.

Proposition 24 (Amari and Nagaoka 2000, Theorem 2.6, p. 38) *Assume that $\{(g_n)\}_{n=1}^\infty$ is a sequence of Riemannian metrics on $\mathcal{P}_+(\Omega_n)$ for each n that are invariant with respect to sufficient statistics; i.e., for all n, m , $S \subset \mathcal{P}_+(\Omega_n)$, and $F : \Omega_n \rightarrow \Omega_m$ such that F is a sufficient statistic for S , the induced metrics on S and $F_*(S)$ are assumed to be invariant. Then there exists a positive real number c such that, for all n , g_n coincides with the Fisher metric on $\mathcal{P}_+(\Omega_n)$ scaled by a factor of c .*

Amari and Nagaoka did not supply their proof of Proposition 24. We recommend the reader to Campbell (1986) for a slight generalization of the Chentsov theorem, whose proof is close to the original Chentsov's proof. For the reader convenience we recall the following monotonicity characterization of the Fisher metric on finite sample spaces.

Proposition 25 (Ay et al. 2015, Corollary 4.11) *Let F be a continuous local statistical quadratic 2-form defined on statistical models associated with finite sample spaces $\{\Omega_n\}$ such that F is monotone under statistics. Then F coincides with the Fisher metric up to a multiplicative constant.*

Remark 26 1. Chentsov defined the Fisher metric only on the positive sector $\mathcal{P}_+(\Omega_n)$ of the space of all probability measures, because the expression for the Fisher metric in (2) is well defined only on $\mathcal{P}_+(\Omega_n)$. In this paper we follow the approach in Ay et al. (2015) by requiring that an information metric F is obtained by (1) from the associated 2-form \tilde{F} , which is not only defined on $\mathcal{P}_+(\Omega_n)$ but also defined on $\mathcal{M}(\Omega_n)$ (in general case, on $\mathcal{M}(\Omega)$) and hence on $\mathcal{P}(\Omega_n)$ (resp. on $\mathcal{P}(\Omega)$). This small difference is important, since for a non-discrete space Ω we do not know how to define a notion of a positive measure without using a reference measure μ_0 . Since the Fisher metric g^F satisfies the mentioned requirement, see Example 19, Proposition 25 is equivalent to the Chentsov uniqueness theorem. Clearly, Theorem 4 generalizes Proposition 25.

2. As we mentioned above, the original Chentsov theorem can be equivalently reformulated in terms of the associated form \tilde{F} . Note that the space $\mathcal{P}(\Omega_n)$ (resp. $\mathcal{M}(\Omega_n)$) is not a manifold, or a manifold with boundary, but a stratified space which admits different embeddings into Euclidean spaces. In Ay et al. (2016) and in the present paper we do not consider smooth tensor fields on $\mathcal{P}(\Omega_n)$ (resp. on $\mathcal{M}(\Omega_n)$) but (strongly or point-wise) continuous tensor fields on $\mathcal{M}(\Omega)$ which do not require the notion of a smooth structure on $\mathcal{M}(\Omega)$.
3. In Morozova et al. (1991, §5) Morozova–Chentsov also suggested a method to extend the Chentsov uniqueness theorem to the case of non-discrete measure spaces Ω . Their idea is similar to the Amari–Nagaoka idea, namely they wanted to consider a Riemannian metric on infinite measure spaces as limit of Riemannian metrics on finite measure spaces. They did not discuss a condition under which such a limit

exists. In fact, they did not give a definition of limit of such metrics. If the limit exists they called it *finitely generated*. They stated that the Fisher metric is the unique finitely generated metric that is invariant under sufficient statistics (resp. that is monotone). One may speculate that since such a Riemannian metric depends on base measures μ and tangent vectors at μ Morozova–Chentsov’s approach requires a definition of topology on the space $\mathcal{L}_2^2(\Omega)$.

References

- Adams, R.A., Fournier, J.J.F. (2006). *Sobolev spaces*. Amsterdam: Elsevier/Academic Press.
- Amari, S. (1987). Differential geometrical theory of statistics. In: *Differential geometry in statistical inference. Lecture note-monograph series*, (Vol. 10). California: Institute of Mathematical Statistics.
- Amari, S., Nagaoka, H. (2000). *Methods of information geometry. Translations of mathematical monographs* (Vol. 191). Providence/Oxford: American Mathematical Society/Oxford University Press.
- Ay, N., Jost, J., Lê, H. V., Schwachhöfer, L. (2015). Information geometry and sufficient statistics. *Probability Theory and related Fields*, 162, 327–364. [arXiv:1207.6736](https://arxiv.org/abs/1207.6736).
- Ay, N., Jost, J., Lê, H. V. and Schwachhöfer, L., *Information geometry* (book in preparation).
- Ay, N., Olbrich, E., Bertschinger, N., Jost, J. (2011). A geometric approach to complexity. *Chaos*, 21, 37–103.
- Billingsley, P. (1999). *Convergence of probability measures*. New York: Wiley.
- Bogachev, V.I. (2007). *Measure Theory* (Vol. I, II). Berlin: Springer.
- Campbell, L. L. (1986). An extended Chentsov characterization of a Riemannian metric. *Proceedings of the American Mathematical Society*, 98, 135–141.
- Chentsov, N. (1978). Algebraic foundation of mathematical statistics. *Mathematische Operationsforschung und Statistik Serie Statistics*, 9, 267–276.
- Chentsov, N. (1982). *Statistical decision rules and optimal inference. Translation of mathematical monographs* (Vol. 53). Providence: American Mathematical Society.
- Hamilton, R. (1982). The inverse function theorem of Nash and Moser. *Bulletin of the American Mathematical Society*, 7, 65–222.
- Jost, J. (2005). *Postmodern analysis*. Berlin: Springer.
- Morozova, E., Chentsov, N. (1991). *Natural geometry of families of probability laws*, Itogi Nauki i Techniki, Current problems of mathematics, Fundamental directions 83 (pp. 133–265). Moscow.
- Neveu, J. (1965). *Mathematical foundations of the calculus of probability*. San Francisco: Holden-Day Inc.
- Shahshahani, S. (1979). A new mathematical framework for the study of linkage and selection. *Memoirs of the American Mathematical Society*, volume 17, Nr. 211.