Information Geometry and Sufficient Statistics

Nihat Ay
Jürgen Jost
Hong Van Le
Lorenz Schwachhoefer

SFI WORKING PAPER: 2012-11-020

INFORMATION GEOMETRY AND SUFFICIENT STATISTICS

NIHAT AY[1,5], JÜRGEN JOST[1,4,5], HÔNG VÂN LÊ[2] AND LORENZ SCHWACHHÖFER[3]

Abstract. Information geometry provides a geometric approach to families of statistical models. The key geometric structures are the Fisher quadratic form and the Amari-Chentsov tensor. In statistics, the notion of sufficient statistic expresses the criterion for passing from one model to another without loss of information. This leads to the question how the geometric structures behave under such sufficient statistics. While this is well studied in the finite sample size case, in the infinite case, we encounter technical problems concerning the appropriate topologies. Here, we introduce notions of parametrized measure models and tensor fields on them that exhibit the right behavior under statistical transformations. Within this framework, we can then handle the topological issues and show that the Fisher metric and the Amari-Chentsov tensor on statistical models in the class of symmetric 2-tensor fields and 3-tensor fields can be uniquely (up to a constant) characterized by their invariance under sufficient statistics, thereby achieving a full generalization of the original result of Chentsov to infinite sample sizes. More generally, we decompose Markov morphisms between parametrized measure models in terms of statistics. In particular, the Cramér-Rao inequality, a monotonicity result for the Fisher information, naturally follows.

Keywords: Fisher quadratic form, Amari-Chentsov tensor, sufficient statistic, Chentsov theorem

Contents

1. Introduction

Let us begin with a short synopsis of our work and its context. Parametrized statistics deals with families of probability measures on some sample space Ω parametrized by a parameter x from some space M which we shall take to be a finite dimensional manifold. This parameter is to be estimated, and for that purpose, one wishes to quantify the dependence of the model on that parameter. That is achieved by the Fisher metric as first suggested by Rao [29], and this line was then systematically developed by Chentsov and Morozova [12], [13], [26]. Moreover, there exists a natural affine structure on spaces of probability measures as discovered by Amari [1], [2] and Chentsov [14]. Such structures should be invariant under reparametrizations, and this leads us into the realm of differential geometry, the field of mathematics that systematically

Date: November 3, 2012.
J.J. is partially supported by ERC Advanced Grant FP7-267087; H.V.L. is partially supported by RVO: 67985840.

investigates geometric invariances. Statistics, however, requires more. There is the concept of a sufficient statistic, that is, a mapping between sample spaces that preserves all information about the parameter x. Therefore, it is natural to also require the invariance of the geometric structures under sufficient statistics. It is relatively easy to see that the Fisher metric and the Amari-Chentsov tensor are invariant, but whether they are the only such invariant structures is more subtle. This is the question we are addressing and solving in the present paper. For finite sample spaces, this has been achieved by Chentsov long ago. The case of infinite sample spaces, however, is more difficult. The space of probability measures on an infinite sample space is infinite dimensional, and therefore, standard constructions from finite dimensional differential geometry may fail. Let us first try to develop some intuition about this difficulty as it reveals itself in our setting. A natural way to change a measure μ consists in multiplying it by nonnegative functions ϕ. Such functions obviously need to satisfy some technical properties. In order to apply techniques from differential geometry, we need to define tangent spaces of such families. In order to define a (Hilbert type) metric on such a tangent space, it should ideally consist of L^2-functions f w.r.t. the measure μ under consideration. In order to obtain a variation of the base measure, however, we should then consider $e^f\mu$. This leads to the obvious problem that for $f \in L^2$, e^f need not be in L^1, and the construction might therefore be ill-defined. Thus, one needs to consider restricted classes of functions, and one might try to construct a more refined Banach space structure. This has been achieved by Pistone with Sempi [28] and other coworkers [10], [16]. We shall describe their construction in detail in the last section of our paper, but let us point out some fact that might give some useful intuition. When $e^f \in L^1$, then $e^{tf} \in L^{1/t}$, and one can then play with t, as it suffices to exponentiate e^{sf} for sufficiently small s in order to pass from an infinitesimal variation to a local family. One may then use techniques from Orlicz spaces. However, there are technical difficulties, caused for instance by the fact that the topology on the considered Banach manifolds is so strong that the space of bounded random variables is not dense in that topology [10, Lemma 2].
Our approach is different. Our essential idea is that while the space of all probability measures $\mathcal{M}(\Omega)$ on the sample space Ω in general will not carry the required geometric structures, it can still induce such structures on all finite dimensional models, that is, on statistical families with a finite dimensional manifold of parameters. For that purpose, however, those families need to be embedded into the space of all measures, and including the embedding p as part of our notion of a statistical model allows us to treat the elements of M as measures on Ω. We can then pull back tensors from $\mathcal{M}(\Omega)$ to M and then require the needed regularity properties not on $\mathcal{M}(\Omega)$, where we might not be able to define them, but on M, where they can be naturally defined. This leads us to a notion of a *statistical model* or *statistical manifold* [22, 23] as a manifold M equipped with a (Fisher) metric g and an (Amari-Chentsov) 3-tensor which are induced by an embedding p into $\mathcal{M}(\Omega)$.
Our approach combines concepts from measure theory, information theory, and statistics. It thus is situated in information geometry, a new mathematical field that recently emerged, where geometric ideas and methods are exploited as principal tools to study mathematical statistics and related problems in information theory, neural networks, system theory [4]. Information geometry has also been identified as a natural formalism for complexity theory [5, 6]. In particular, complex networks can be analyzed with tools from information geometry [27].

The structure of our paper is as follows. In Section 2 we introduce the notion of a k-integrable parametrized measure model, which encompasses all known examples in statistics. We compare our concept with the concept of a geometrically regular statistical model proposed by Amari. At the end of this section, we state our Main Theorem 2.9. In Section 3 we introduce the notion of sufficient statistics based on the Fisher-Neyman characterization (Definition 3.1, Lemma 3.3). We give a simple proof that the Amari-Chentsov structure is invariant under sufficient statistics (Theorem 3.5). At the end of the section we discuss Chentsov's results on the uniqueness of the

Fisher metric and the Amari-Chentsov tensor (Proposition 3.17, Lemma 3.18). At the end of that Section, we prove our Main Theorem. In Section 4 we introduce the notion of a Markov morphism. A novel aspect of our concept of Markov morphisms between parametrized measure models is the consideration of smooth maps between the parameter spaces (Definition 4.4, Example 4.5). Thus, the geometry of parametrized measure models is intrinsic. We decompose a Markov morphism as a composition of a right inverse of a sufficient statistic and a statistic (Theorem 4.10). As a consequence we give a geometric proof of the monotonicity theory for Markov morphisms (Corollary 4.11). Finally, in Section 5, we study the relations between k-integrable parametrized measure models and statistical models in the Pistone-Sempi theory.

2. Parametrized measure models

In this section we describe the geometry of spaces of measures and of parametrized families of measures. In technical terms, we introduce the notion of a k-integrable parametrized measure model (Definition 2.3) and the notion of tensor fields on them, following the locality and continuity condition (Definition 2.1, Remark 2.4). We show that our notion of generalized statistical models encompasses all statistical models considered by Chentsov, Amari, Pistone-Sempi (Remark 2.4, Example 2.5), and we compare our concept with that by Amari (Remark 2.6).

Let (Ω, Σ) be a measurable space. Later on, Ω will also have to carry a differentiable structure. We consider the Banach space of all signed finite measures on Ω with the total variation $\|\cdot\|_{TV}$ as Banach norm. More precisely, the total variation of such a measure μ is defined as

$$\|\mu\|_{TV} \;:=\; \sup \sum_{i=1}^{n} |\mu(A_i)|$$

where the supremum is taken over all finite partitions $\Omega = A_1 \dot\cup \ldots \dot\cup A_n$ with disjoint sets $A_i \in \Sigma$. We consider the subset $\mathcal{M}(\Omega)$ of all finite non-negative measures on Ω, and, with a σ-finite non-negative measure μ_0, we also consider the subspace

$$\mathcal{S}(\Omega, \mu_0) \quad := \quad \{\mu = \phi\, \mu_0 \;:\; \phi \in L^1(\Omega, \mu_0)\}.$$

of signed measures dominated by μ_0. This space can be identified in terms of the canonical map $i_{can} : \mathcal{S}(\Omega, \mu_0) \to L^1(\Omega, \mu_0)$, $\mu \mapsto \frac{d\mu}{d\mu_0}$. Note that

$$\|\mu\|_{TV} \;=\; \left\| \frac{d\mu}{d\mu_0} \right\|_{L^1(\Omega, \mu_0)},$$

which implies that i_{can} is a Banach space isomorphism. Therefore, we refer to the topology of $\mathcal{S}(\Omega, \mu_0)$ also as *L^1-topology*. This is independent of the particular choice of the reference measure μ_0, because if $\phi \in L^1(\Omega, \mu_0)$ and $\psi \in L^1(\Omega, \phi\mu_0)$, then $\psi\phi \in L^1(\Omega, \mu_0)$. Throughout the paper, we consider the following hierarchy of subsets of $\mathcal{S}(\Omega, \mu_0)$:

$$\begin{aligned}
\mathcal{M}(\Omega, \mu_0) &= \{\mu = \phi\, \mu_0 \;:\; \phi \in L^1(\Omega, \mu_0),\;\; \phi \geq 0\} \\
\mathcal{M}_+(\Omega, \mu_0) &= \{\mu = \phi\, \mu_0 \;:\; \phi \in L^1(\Omega, \mu_0),\;\; \phi > 0\} \\
\mathcal{M}^a(\Omega, \mu_0) &= \{\mu = \phi\, \mu_0 \;:\; \phi \in L^1(\Omega, \mu_0),\;\; \phi \geq 0,\; \mu(\Omega) = \|\mu\|_{TV} = a\} \\
\mathcal{P}(\Omega, \mu_0) &= \{\mu \in \mathcal{M}(\Omega, \mu_0) \;:\; \mu(\Omega) = \|\mu\|_{TV} = 1\} \\
\mathcal{P}_+(\Omega, \mu_0) &= \{\mu \in \mathcal{M}_+(\Omega, \mu_0) \;:\; \mu(\Omega) = \|\mu\|_{TV} = 1\}
\end{aligned}$$

In particular, for $\mu = \phi\mu_0 \in \mathcal{M}_+(\Omega, \mu_0)$, i.e., $\phi > 0$, μ_0 and μ have the same null sets and are equivalent, that is, $\mu_0 = \phi^{-1}\mu \in \mathcal{M}_+(\Omega, \mu)$. Thus, we have some kind of multiplicative structure on $\mathcal{M}_+(\Omega, \mu_0)$, and one might hope to generate this via an exponential map from the linear structure on $L^1(\Omega, \mu_0)$. The problem, however, is that if $f \in L^1(\Omega, \mu_0)$, then we do not necessarily have $e^f \in L^1(\Omega, \mu_0)$. When it is, then $e^f\mu_0 \in \mathcal{M}_+(\Omega, \mu_0)$, but when it is not, the measure $e^f\mu_0$ is not well defined. Thus, certain infinitesimal deformations are obstructed, that is, cannot be integrate into local ones. Of course, this does not happen when Ω is finite, the case treated by Chentsov, and this is the technical reason why we need to work harder for our main result. (Pistone and Sempi have analyzed the underlying topological structure, and we shall describe their construction from our perspective in Section 5. The essential point for an intuitive understanding of this topology is that if $e^f \in L^1$, then for $0 < t < 1$, $e^{tf} \in L^p$ for $p = 1/t > 1$.)

In order to avoid this issue and in order to make contact with the basic construction of parametric statistics, we shall consider parametrized families of measures, that is, differentiable maps $M \to \mathcal{M}(\Omega)$ of smooth Banach manifolds M into the "universal measure set" $\mathcal{M}(\Omega)$ and attempt to pull geometric structures from $\mathcal{M}(\Omega)$ back to M by such maps. Since, however, we may not be able to fully define these objects on $\mathcal{M}(\Omega)$, we shall have to push forward tensors from M instead, and integrate them w.r.t. the measures $p(x)$ defined by a parametrized family. (In fact, the smoothness requirement for M that we really need is continuous differentiability, that is, a C^1-structure, but since in statistics, the precise smoothness requirement is usually not an important issue, we do not elaborate upon this point.) We shall now introduce the technical conditions needed to realize universal objects on $\mathcal{M}(\Omega)$ on such parametrized families.

If we have a differentiable map $p : M \to \mathcal{M}(\Omega)$ that assigns to each $x \in M$ a measure $p(x, .) \in \mathcal{M}_+(\Omega, \mu_0)$, then we can push forward vector fields (or other contravariant tensor fields) on M, and we can pull back covariant tensors from $\mathcal{M}_+(\Omega, \mu_0)$. We need to impose various conditions on the tensors and on the maps which we are now going to develop.

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

In this definition, continuity refers to the continuity of the linear maps τ_μ for fixed μ. (This is different from requiring that τ_μ be continuous as a function of μ.)
Such objects then will be pulled back to M under a map $p : M \to \mathcal{M}(\Omega)$, and they then operate on n vector fields on M. When these vector fields are continuous, their evaluation under the pulled back covariant tensor field should also be continuous. This requirement is formalized in

Definition 2.2. A covariant n-tensor field τ on a smooth Banach manifold M is called (*weakly*) *continuous* if for any continuous contravariant n-tensor field A on M the function $\tau(A)$ is a continuous function on M.

In contrast to the preceding definition which was only concerned with the continuity of a linear operator at each point, this definition requires that the objects be continuous as functions of the point $x \in M$.

For a map $p : M \to \mathcal{M}(\Omega, \mu)$ the composition $\bar{p} := p \circ i_{can} : M \to L^1(\Omega, \mu)$, $x \mapsto \bar{p}(x) := \frac{dp(x)}{d\mu}$, will play a central role. Thus, $\bar{p}$ is a map from M to $L^1(\Omega, \mu)$ which we can then evaluate at some $\omega \in \Omega$. We can therefore consider $\bar{p}$ also as a map $M \times \Omega \to \mathbb{R}$, $(x, \omega) \mapsto \bar{p}(\omega, x) = \frac{dp(x)}{d\mu}(\omega)$, which we refer to as *the density potential*. However, this notation is slightly misleading, and the infinitesimal tangent vector of the family rather corresponds to $\ln \bar{p}(\omega, x)$ (recall our discussion above of the exponentiation of $f \in L^1(\Omega, \mu)$, and taking the logarithm of course is the inverse of exponentiation.) In particular, the pushforward of a tangent vector $V \in T_xM$ is $\partial_V \ln \bar{p}(x, \omega)$, and we often simply identify V with its pushforward when the map p is fixed in a given context.
Our parametrized families of measures will need to satisfy some further important technical requirements that we shall now list and that will lead us to our technical concept of a parametrized measure model.

(1) The parameter space M is a smooth manifold (of class at least C^1, to be precise).
(2) There is a continuous mapping $p : M \to \mathcal{M}_+(\Omega, \mu)$ provided with the L^1-topology.

(3) The composition $\bar{p} = i_{can} \circ p$ is differentiable as a map from the manifold M to the Banach space $L^1(\Omega, \mu)$ (Gâteaux-differentiability).

(4) The 1-form

$$A(V)_x := \int_\Omega \partial_V \ln \bar{p}(x,\omega)\, dp(x), \tag{2.1}$$

the *Fisher quadratic form*

$$g^F(V,W)_x := \int_\Omega \partial_V \ln \bar{p}(x,\omega) \partial_W \ln \bar{p}(x,\omega)\, dp(x) \tag{2.2}$$

and the *Amari-Chentsov 3-symmetric tensor*

$$T^{AC}(V,W,X)_x := \int_\Omega \partial_V \ln \bar{p}(x,\omega) \partial_W \ln \bar{p}(x,\omega) \partial_X \ln \bar{p}(x,\omega)\ dp(x) \tag{2.3}$$

are well-defined and continuous in the sense of Definition 2.2.

We can now state our general definition of a parametrized measure model.

Definition 2.3. (cf. [3, §2 , p. 25], [4, §2.1]) Let $k \geq 1$. A *k-integrable parametrized measure model* is a quadruple (M, Ω, μ, p) consisting of a smooth (finite dimensional or infinite dimensional) Banach manifold M and a continuous map $p : M \to \mathcal{M}_+(\Omega, \mu)$ provided with the L^1-topology such that

(1) the function $x \mapsto \ln \bar{p}(x,\omega) := \ln \frac{dp(x)}{d\mu}(\omega) : M \to \mathbb{R}$ is defined and continuously Gâteaux-differentiable for μ-almost all $\omega \in \Omega$,

(2) for all $1 \leq p \leq k$ and for all continuous vector fields V on M the function $\omega \mapsto \partial_V \ln \bar{p}(x,\omega)$ belongs to $L^p(\Omega, p(x))$; moreover, the function $x \mapsto ||\partial_V \ln \bar{p}(x,\omega)||_{L^p(\Omega,p(x))}$ 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 $p(M) \subset \mathcal{P}_+(\Omega, \mu)$. A k-integrable parametrized measure model (M, Ω, μ, p) is called *immersed* if $d_x \ln \bar{p} : T_x M \to L^k(\Omega, p(x))$ is injective for all $x \in M$.

Remark 2.4. 1. Note that, as explained above, the choice of a reference measure in $\mathcal{M}_+(\Omega, \mu)$ is immaterial for a k-integrable parametrized measure model (M, Ω, μ, p).

2. For a *statistical* model, (2.1) vanishes identically. Recalling the identification of the tangent vector V on M with its pushforward $\partial_V \ln \bar{p}$, this simply means

$$\int_\Omega V d\mu = 0. \tag{2.4}$$

To obtain (2.4) we argue as follows. For a curve $x(t)$, $t \in (-\varepsilon, \varepsilon)$, on M with $\partial_t := \dot{x}(t) = V((x(t))$ the condition (2) in Definition 2.3 implies that

$$f(t) := \int_\Omega \partial_t \ln \bar{p}(x(t),\omega)\, dp(x(t))$$

is continuous and hence integrable over $(-\varepsilon, \varepsilon)$. In particular, $A(V)_x$ is continuous in x. Apply the Fubini theorem and the condition (1) in Definition 2.3 we have

$$\int_{-\varepsilon}^{\varepsilon} \int_\Omega (\partial_t \ln \bar{p}(x(t),\omega)) p(x(t)) d\mu\, dt = \int_\Omega \int_{-\varepsilon}^{\varepsilon} (\partial_t \ln \bar{p}(t,\omega)) p(t,\omega) dt d\mu$$

$$= \int_\Omega (p(\varepsilon,\omega) - p(-\varepsilon,\omega)) d\mu = 0.$$

Observe that the above formula for general k-integrable parametrized measure models implies

$$\partial_V \int_\Omega dp(x) = \int_\Omega \partial_V \ln p(x,\omega)\, dp(x) \tag{2.5}$$

for all $x \in M$ and for all tangent vector $V \in T_x M$.

3. For any k-integrable parametrized measure model (M, Ω, μ, p) the composition $i_{can} \circ p : M \to L^1(\Omega, \mu)$ is Gâteaux-differentiable by (2.5) and taking into account

$$\int_\Omega |\partial_V e^{\ln \bar{p}(x)}| d\mu = \int_\Omega |\bar{p}(x) \partial_V \ln \bar{p}(x)| d\mu = \int_\Omega |\partial_V \ln \bar{p}(x)| dp(x) < \infty.$$

4. Any 3-integrable parametrized measure model carries the Fisher quadratic form and the Amari-Chentsov tensor, which are continuous in the sense of Definition 2.2. On a k-integrable parametrized measure model (M, Ω, μ, p) the covariant symmetric n-tensor field $T^n(V, \cdots, V) := (\partial_V \ln \bar{p}(x, \omega))^n$ satisfies the locality and continuity conditions required in the introduction.

5. In [14] Chentsov considered only statistical models (M, Ω_n, μ_n, p) where M is a submanifold in $\mathcal{P}_+(\Omega_n, \mu_n)$ and p is the canonical embedding, see also Example 2.5. Amari and all authors before Pistone and Sempi considered only statistical models (M, Ω, μ, p) where M is finite dimensional and $p(M) \subset \mathcal{P}_+(\Omega, \mu)$ [4]. Their examples satisfy the conditions in Definition 2.3.

Example 2.5. 1. Let Ω_n be a finite set of n elements and μ_n a measure of maximal support on Ω_n. It is evident that $\mathcal{M}_+(\Omega_n, \mu_n)$ is diffeomorphic to $\mathbb{R}^n$. Let S be a C^1-submanifold in $\mathcal{P}_+(\Omega_n, \mu_n)$ and $i_S : S \to \mathcal{P}_+(\Omega_n, \mu_n)$ the canonical embedding. Then $(S, \Omega_n, \mu_n, i_S)$ is an immersed k-integrable statistical model for all $k \geq 1$. In particular, $(\mathcal{P}_+(\Omega_n, \mu_n), \Omega_n, \mu_n, Id)$ is a k-integrable statistical model. Conversely, for any immersed 1-integrable statistical model (M, Ω_n, μ_n, p) the map $p : M \to \mathcal{P}_+(\Omega_n, \mu_n)$ defines an immersion $M \to \mathcal{P}_+(\Omega_n, \mu_n)$ between differentiable manifolds.

2. If $s : N \to M$ is a smooth map and (M, Ω, μ, p) is a k-integrable parametrized measure model, then $(N, \Omega, \mu, p \circ s)$ is a k-integrable parametrized measure model.

On a 3-integrable parametrized measure model (M, Ω, μ, p) the pair of the Fisher quadratic form and the Amari-Chentsov tensor is also called the *Amari-Chentsov structure*.

Remark 2.6. We would like to compare our concept of a k-integrable parametrized measure model with the concept of a geometrical regular statistical model proposed by Amari, for instance in [3, §2]. Amari listed 6 properties a geometrically regular statistical model $\{p(x, \omega)\}$ must satisfy [3, A_1-A_6, p. 25]. The condition A_1 says that the domain of parameter x is homeomorphic to $\mathbb{R}^n$. The conditions A_2 and A_3 are equivalent to our condition (2) listed just before Definition 2.3. The condition A_4 requires that $\bar{p}(x, \omega)$ is smooth in x uniformly in ω, and moreover the relation (2.5) holds. The condition A_5 requires that the statistical model is 3-integrable, moreover the function $x \mapsto ||\partial_V \ln \bar{p}(x, \omega)||_{L^p(\Omega, p(x))}$ exists and smooth for $1 \leq p \leq 3$. The last condition A_6 requires that the Fisher quadratic form is positive definite. Amari's conditions are slightly stronger than ours, but in general our concept agrees with his concept. Note that similar regularity conditions have been posed by Cramer [11, p.500-501], see also [20, Chapter 2, §6].

As mentioned above, we consider tensor fields on parametrized measure models (M, Ω, μ, p) that are inherited from a corresponding field on the "universal measure set" $\mathcal{M}(\Omega)$ in terms of the parametrization p.

Note that we do not impose any strong regularity conditions on tensor fields on $\mathcal{M}(\Omega)$. Instead, we assume the required regularity and continuity conditions to be satisfied on the pull-back of the field with respect to a parametrization $p : M \to \mathcal{M}(\Omega)$. In addition to these conditions, the existence of a global tensor on $\mathcal{M}(\Omega)$ sets some compatibility constraints on the associated fields on the class of parametrized measure models (M, Ω, μ, p). In the following definition we summarize necessary regularity and compatibility conditions for tensor fields, which are, in particular, satisfied in the case of the Fisher quadratic form and the Amari-Chentsov tensor.

Definition 2.7 (Locality and continuity condition)**.** A *statistical* covariant *(continuous)* n-tensor field A assigns to each parametrized measure model (M, Ω, μ, p) a *continuous (in the sense of*

Definition 2.2) covariant n-tensor field $A|_{(M,\Omega,\mu,p)}$ on M (cf. Definition 2.1). A statistical covariant n-tensor field A is called *local* if for any parametrized measure model (M, Ω, μ, p) and any $V_i \in T_x M$ the value $A|_{(M,\Omega,\mu,p)}(V_1, \ldots, V_n)$ depends only on $p(x)$ and the values $\partial_{V_1} \ln \bar{p}(x), \ldots, \partial_{V_n} \ln \bar{p}(x) \in L^n(\Omega, p(x))$ (but not on $p(M)$).

In particular, this means that the value depends only on $p(x)$, but not on the manifold M defining the parametrized family of which $p(x)$ is a member.

Remark 2.8. 1. Assume that A is a local statistical covariant continuous n-tensor field. Using Lemma 3.20 and condition (1) in Definition 2.3 we note that A defines a point-wise continuous n-tensor field $\tilde{A}$ on $\mathcal{M}(\Omega)$ (Definition 2.1) by setting

$$\tilde{A}_{p(x)}(\partial_{V_1} \ln \bar{p}(x), \cdots, \partial_{V_n} \ln \bar{p}(x)) = A|_{(M,\Omega,\mu_0,p)}(V_1(x), \cdots, V_n(x)). \tag{2.6}$$

Thus, in order to define A it suffices to determine the associated point-wise continuous n-tensor field $\tilde{A}$ on $\mathcal{M}(\Omega)$ and then verify if the original statistical field A is continuous.

Condition (2.6) holds for the Fisher quadratic form field and the Amari-Chentsov tensor field. The choice of $\partial_V \ln \bar{p}(x)$ is also related to the Gâteaux-differentiability of p (Remark 2.4.3). We choose $L^n(\Omega, p(x))$ as a natural condition for the value $\partial_{V_n} \ln \bar{p}(x)$ since it is a natural extension of the condition for the existence of the Fisher quadratic form and the Amari-Chentsov tensor on a parametrized measure model. In fact, it is also possible to replace the value space $L^n(\Omega, p(x))$ by another function space depending on the measure $p(x)$ (Remark 3.21).

2. The locality and continuity condition holds obviously for tensor fields on statistical models associated with finite sample spaces as in the Chentsov work [14].

3. In [22] and [23], Lê proved the following variant of the locality condition, which has been asked by Lauritzen [21] and Amari-Nagaoka [4]. For any statistical model (M, g, T) there exist a finite sample space Ω_n provided with a dominant measure μ_n and an immersion $p : M \to \mathcal{M}(\Omega_n, \mu_n) = \mathcal{M}(\Omega_n)$ such that the statistical structure (g, T) is induced from the Amari-Chentsov structure on $(\mathcal{M}(\Omega_n, \mu_n), \Omega_n, \mu, Id)$ via p.

Our main theorem uses the notion of a sufficient statistic and the associated invariance property. As already stated in the introduction, sufficient statistics are important transformations between parametrized measure models, since they preserve the information of the underlying models. Although we introduce the corresponding definitions later in the paper, we present our main theorem already here so that its main structure guides the arguments and motivates further results of the paper.

Theorem 2.9 (Main Theorem)**.** *(1) Assume that A is a local statistical continuous 1-form field. If A is invariant under sufficient statistics then there is a continuous function $c : \mathbb{R} \to \mathbb{R}$ such that for all finite measures μ on Ω and for all $V \in L^1(\Omega, \mu)$ we have*

$$\tilde{A}_\mu(V) = c(\int_\Omega d\mu) \cdot \int_\Omega V \, d\mu.$$

In particular, recalling (2.4), there is no weakly continuous 1-form field on statistical models that is invariant under sufficient statistics. On a parametrized measure model (M, Ω, μ, p) the field A is expressed as follows

$$A(V)_x = c(\int_\Omega dp(x)) \cdot \partial_V(\int_\Omega dp(x)). \tag{2.7}$$

(2) Assume that F is a local statistical continuous quadratic form field. If F is invariant under sufficient statistics then there is a continuous function $f : \mathbb{R} \to \mathbb{R}$ such that $F(x) = f(\int_\Omega p(x)) g^F(x) + A(x)^2$, where A is the field in (1) and g^F is the Fisher quadratic form. In

particular, the Fisher quadratic form is the unique up to a constant weakly continuous quadratic form field on statistical models that is invariant under sufficient statistics.

(3) Assume that T is a local statistical continuous covariant symmetric 3-tensor field. If T is invariant under sufficient statistics then there is a continuous function $t : \mathbb{R} \to \mathbb{R}$ such that $T(x) = t(\int_\Omega p(x))T^{AC}(x) + A_1(x)^3 + A_2(x) \cdot g^F(x)$, where A_1, A_2 are the fields described in (1), g^F and T^{AC} are the Fisher quadratic form and the Amari-Chentsov tensor respectively. In particular, the Amari-Chentsov tensor is the unique up to a constant weakly continuous 3-symmetric tensors field on statistical models that is invariant under sufficient statistics.

Campbell noticed that the Fisher metric on parametrized measure models associated with a finite sample space Ω_n coincides with the Shahshahani metric [9], which is important in mathematical biology and game theory [30]. It is interesting to find applications in this direction of the Fisher metric and other natural metrics on generalized statistical models described in the Main Theorem.

3. Sufficient statistics and the Amari-Chentsov structure

A statistic κ is a measurable map between a measure space (Ω_1, μ_1) and a measurable space Ω_2. One of the most important properties of the Fisher quadratic form and the Amari-Chentsov tensor is the invariance of these structures under statistics $\kappa : \Omega_1 \to \Omega_2$ that are sufficient (a notion introduced by Fisher in 1922) for the parameter $x \in (M, \Omega_1, \mu, p)$ (Definition 3.1, Theorem 3.5). In other words, the Fisher quadratic form and the Amari-Chentsov tensor on (M, Ω_1, μ, p) and $(M, \Omega_2, \kappa_*(\mu), \kappa_*(p))$ coincide, if κ is sufficient. Sufficient statistics represent important transformations between parametrized measure models, since they preserve the information of the underlying models. Thus one wishes to know whether there are other quadratic forms and 3-symmetric tensors on parametrized measure models which are invariant under sufficient statistics. This question has been solved by Chentsov in the negative for statistical models associated with finite sample spaces [14], see Proposition 3.17. However, one naturally wishes to consider *infinite* sample spaces Ω, and in this case the space of measures becomes *infinite dimensional*, and the topological aspects then become more subtle. More precisely, the main difficulty for an extension of the Chentsov theorem to all parametrized measure models is caused by two facts. Firstly, a statistical model associated with a finite sample space can be regarded locally as a submanifold in a universal statistical model $(\mathcal{P}_+(\Omega_n, \mu_n), \Omega_n, \mu_n)$, which is a finite-dimensional open simplex (Example 2.5). In this case, it suffices to consider the Fisher metric, the Amari-Chentsov tensor and other tensor fields on this open simplex. Secondly, the structure of sufficient statistics associated with the considered statistical models can be described in terms of Markov congruent embeddings [14], see also our discussion at the end of Section 4. It is not easy to generalize these facts to statistical models associated with infinite sample space, since, in particular, there is no canonical smooth structure on the set $\mathcal{M}_+(\Omega, \mu)$ of all measures equivalent to μ, or on the set $\mathcal{M}(\Omega, \mu)$ of all measures dominated by μ.

In this section, we first give a simple proof that the Amari-Chentsov structure is invariant under sufficient statistics (Theorem 3.5). We also give a geometric proof of the Fisher-Neyman factorization theorem which characterizes a sufficient statistic $\kappa : (\Omega_1, \mu_1) \to \Omega_2$ under the assumption that $\kappa : \Omega_1 \to \Omega_2$ is a smooth map (Theorem 3.10). Using Theorem 3.10 we present a proof of the monotonicity theorem, also called the Cramér-Rao inequality (Theorem 3.11). We also consider examples of sufficient statistics, which are associated with Markov congruent embeddings from $\mathcal{M}_+(\Omega_n, \mu_n)$ to $\mathcal{M}_+(\Omega_m, \mu_m)$ (Example 3.14). Using them we discuss Chentsov's results on geometric structures which are invariant under sufficient statistics between finite sample spaces (Proposition 3.17, Lemma 3.18). We shall then be in a position to prove our Main Theorem.

For a measurable map $\kappa : (\Omega_1, \mu_1) \to \Omega_2$ let us denote by $\kappa_*(\mu_1)$ the push-forward measure on Ω_2.

Definition 3.1. (cf. [4, (2.17)], [8, Theorem 1, p. 117]) Assume that $(M, \Omega_1, \mu_1, p_1)$ is a k-integrable parametrized measure model and Ω_2 is a measurable space. A statistic $\kappa : (\Omega_1, \mu_1) \to \Omega_2$ is said to be *sufficient for the parameter* $x \in M$ if there exist a function $s : M \times \Omega_2 \to \mathbb{R}$ and a function $t \in L^1(\Omega_1, \mu_1)$ such that for all $x \in M$ we have $s(x, \omega_2) \in L^1(\Omega_2, \kappa_*(\mu_1))$ and

$$\bar{p}_1(x, \omega_1) = s(x, \kappa(\omega_1))t(\omega_1) \qquad \mu_1 - a.e.. \tag{3.1}$$

Remark 3.2. Definition 3.1 is a version of the Fisher-Neyman characterization theorem, which states that a statistic is sufficient for the parameter $x \in M$ if and only if (3.1) holds. The Fisher-Neyman characterization theorem is simpler to formulate than the corresponding definition in textbooks on mathematical statistics, e.g. in [8, Definition 1, p.116], which involves the notion of conditional distribution.

A measurable map $\kappa : (\Omega_1, \mu_1) \to \Omega_2$ transforms a parametrized measure model $(M, \Omega_1, \mu_1, p_1)$ into the parametrized measure model $(M_1, \Omega_2, \kappa_*(\mu_1), \kappa_*(p_1))$ whose density potential $\kappa_*(\bar{p}_1)$ is defined by

$$\kappa_*(\bar{p}_1) := \frac{d\kappa_*(p_1)}{d\kappa_*(\mu_1)}. \tag{3.2}$$

Lemma 3.3. *A statistic $\kappa : (\Omega_1, \mu_1) \to \Omega_2$ is sufficient for the parameter $x \in M$ if and only if the function*

$$r(x, \omega_1) := \frac{\bar{p}_1(x, \omega_1)}{\kappa_*(\bar{p}_1)(x, \kappa(\omega_1))}$$

does not depend on x for almost all $\omega_1 \in (\Omega_1, \mu_1)$.

Proof. The "if" part of Lemma 3.3 is obvious. Now we assume that (3.1) holds, i.e. $\bar{p}_1(x, \omega_1) = s(x, \kappa(\omega_1)) \cdot t(\omega_1)$ for all $x \in M$ and almost everywhere on (Ω_1, μ_1). Then for all $x \in M$ and almost all $\omega_1 \in (\Omega_1, \mu_1)$ we have

$$\kappa_*(\bar{p}_1)(x, \kappa(\omega_1)) = \kappa_*(t)(\kappa(\omega_1)) \cdot s(x, \kappa(\omega_1)). \tag{3.3}$$

From (3.3) we obtain for all $x \in M$

$$r(x, \omega_1) = \frac{t(\omega_1)s(x, \kappa(\omega_1))}{\kappa_*(t)(\kappa(\omega_1)) \cdot s(x, \kappa(\omega_1))} = \frac{t(\omega_1)}{\kappa_*(t)\kappa(\omega_1)} \qquad \mu_1 - a.e.. \tag{3.4}$$

This completes the proof of Lemma 3.3. □

We get immediately

Corollary 3.4. *Assume that $\kappa : \Omega_1 \to \Omega_2$ is a sufficient statistic for the parameter $x \in M$ where $(M, \Omega_1, \mu_1, p_1)$ is a k-integrable parametrized measure model. Then $(M, \Omega_2, \kappa_*(\mu_1), \kappa_*(p_1))$ is also a k-integrable parametrized measure model.*

Let $\kappa : (\Omega_1, \mu_1) \to (\Omega_2, \mu_2)$ be a statistic and $(M, \Omega_1, \mu_1, p_1)$ a k-integrable parametrized measure model. The Fisher quadratic form $\tilde{g}^F$ on the transformed parametrized measure model $(M, \Omega_2, \kappa_*(\mu_1), \kappa_*(p_1))$ is defined by

$$\tilde{g}^F(V, V)_x = \int_{\kappa(\Omega_1)} (\partial_V \ln(\kappa_*(\bar{p}_1)(x, \omega_2)))^2 d\kappa_*(p_1(x)). \tag{3.5}$$

Theorem 3.5. *If a statistic κ is sufficient for the parameter $x \in M$, then the Amari-Chentsov structure transformed by κ is equal to the original structure.*

Proof. Assume that a statistic κ is sufficient for the parameter $x \in M$. By Lemma 3.3 we have for all $x \in M$

$$p_1(x, \omega_1) = r(\omega_1)\kappa_*(p_1(x))(\kappa(\omega_1)) \qquad \mu_1 - a.e.. \tag{3.6}$$

Hence for all $x \in M$ and all $V \in T_xM$

$$\partial_V \ln p_1(x, \omega_1) = \partial_V \ln \kappa_*(p_1(x))(\kappa(\omega_1)) \qquad \mu_1 - a.e.. \tag{3.7}$$

It follows for all $x \in M$ and all $V \in T_xM$

$$\begin{aligned} g^F(V,V)_x &= \int_{\Omega_1} (\partial_V \ln \kappa_*(p_1(x))(\kappa(\omega_2)))^2 r(\omega_1)\kappa_*(p_1(x))(\kappa(\omega_1)) d\mu_1 \\ &= \tilde{g}^F(V,V)_x. \end{aligned}$$

This proves the invariance of the Fisher metric under sufficient statistics. The invariance of the Amari-Chentsov tensor under sufficient statistics is proved in the same way. □

Corollary 3.6. *Assume that Ω is a differentiable manifold provided with the Borel σ-algebra. The Amari-Chentsov structure on any k-integrable parametrized measure model (M, Ω, μ, p) is invariant under the action of the diffeomorphism group of Ω.*

Remark 3.7. The first known variant of Theorem 3.5 is the second part of the Cramér-Rao inequality (Theorem 3.11) [11], [29]. The invariance of the Amari-Chentsov structure on statistical models associated with finite sample spaces under sufficient statistics has been discovered first by Chentsov [14].

In what follows we interpret the function $r(x, \omega_1)$ assuming that Ω_1 and Ω_2 are smooth manifolds supplied with the Borel σ-algebra and κ is smooth. Furthermore, we assume that μ_1 is dominated by a Lebesgue measure on Ω_1, i.e. a measure that is locally equivalent to the Lebesgue measure on $\mathbb{R}^n$. Then the set Ω_2^{sing} of singular values of κ is a null set in $(\Omega_2, \kappa_*(\mu_1))$. Let ω_2 be a regular value of κ. Then $\kappa^{-1}(\omega_2)$ is a smooth submanifold of Ω_1. Furthermore, any sufficiently small open neighborhood $U_\varepsilon(\omega_2) \subset \Omega_2$ of ω_2 consists only of regular values of κ. Without loss of generality we assume that the preimage $\kappa^{-1}(U_\varepsilon(\omega_2))$ is a direct product $U_\varepsilon(\omega_2) \times \kappa^{-1}(\omega_2)$, which is the case if $U_\varepsilon(\omega_2)$ is diffeomorphic to a ball. The measure μ_1 (respectively, $p_1(x)$) on the source space and the induced measure $\kappa_*(\mu_1)$ (respectively, $\kappa_*(p_1(x))$) on the target space define a "vertical" measure $\mu^\perp_{\omega_2}$, which depends on μ_1, on each fiber $\kappa^{-1}(\omega_2)$ by the following formula:

$$d\mu^\perp_{\omega_2}(\mu_1, y) := \frac{d\mu_1(\omega_2, y)}{d\kappa_*(\mu_1)(\omega_2)} \tag{3.8}$$

for all $y \in \kappa^{-1}(\omega_2)$. (Respectively, we replace μ_1 by $p_1(x)$ in the LHS and RHS of (3.8)). Here we identify a point $(\omega_2, y \in \kappa^{-1}(\omega_2))$ with the image of y in Ω_1 via the inclusion $f^{-1}(\omega_2) \to \Omega_1$. Note that $d\mu^\perp_{\omega_2}(\mu_1, y)$ is well-defined only if $\omega_2 \in \kappa(\Omega_1)$.

Lemma 3.8. *Assume that the value ω_2 of a statistic κ is regular. Then $\mu^\perp_{\omega_2}(\mu_1)$ is a probability measure on $\kappa^{-1}(\omega_2)$ for any finite measure μ_1 on Ω_1.*

Proof. We need to show that

$$\int_{\kappa^{-1}(\omega_2)} d\mu^\perp_{\omega_2}(\mu_1, y) = 1. \tag{3.9}$$

Let g be a Riemannian metric on Ω_2. Denote by $D_\varepsilon(\omega_2)$ the disk with center at ω_2 and of radius ε. Using (3.8) and Fubini's formula we obtain

$$\int_{D_\varepsilon(\omega_2)} d\kappa_*(\mu_1) \int_{\kappa^{-1}(\omega_2)} d\mu^\perp_{\omega_2}(\mu_1, y) = \int_{\kappa^{-1}(D_\varepsilon(\omega_2))} d\mu_1. \tag{3.10}$$

Taking into account

$$\int_{\kappa^{-1}(D_\varepsilon(\omega_2))} d\mu_1 = \int_{D_\varepsilon(\omega_2)} d\kappa_*(\mu_1), \tag{3.11}$$

we derive from (3.10)

$$\int_{\kappa^{-1}(\omega_2)} d\mu^\perp_{\omega_2}(\mu_1, y) = \lim_{\varepsilon\to 0} \frac{\int_{D_\varepsilon(\omega_2)} d\kappa_*(\mu_1)}{\int_{\kappa^{-1}(D_\varepsilon(\omega_2))} d\mu_1} = 1. \tag{3.12}$$

This proves (3.9) and Lemma 3.8. □

Remark 3.9. The measure $\mu^\perp_{\omega_2}$ is the conditional distribution $(d\omega_1|\omega_2)$ of the variable (elementary event) ω_1 subject to the condition $\kappa = \omega_2$. In general, a conditional distribution $(d\omega_1|\omega_2)$ of the variable ω_1 subject to condition $\kappa = \omega_2$ can be defined for measurable mappings, which need not be smooth. We refer to [18, p. 81], [8, p. 106] for a definition of a conditional distribution in a general case.

Theorem 3.10. *Assume that Ω_1 and Ω_2 are smooth manifolds supplied with Borel σ-algebras and μ_1 is a measure on Ω_1 dominated by a Lebesgue measure. Let $(M, \Omega_1, \mu_1, p_1)$ be a k-integrable parametrized measure model. A smooth statistic $\kappa : (\Omega_1, \mu_1) \to \Omega_2$ is sufficient for the parameter $x \in M$ if and only if the conditional distribution $\mu^\perp_{\omega_2}(p_1(x))$ defined on the set of regular values ω_2 of κ is independent of $x \in M$.*

Proof. Representing a point ω_1 by the pair $(\kappa(\omega_1), y)$, $y \in \kappa^{-1}(\kappa(\omega_1))$, we write

$$d\mu^\perp_{\kappa(\omega_1)}(p_1(x), y) = \tilde\mu^\perp_{\kappa(\omega_1)}(x, y) d\mu^\perp_{\kappa(\omega_1)}(\mu_1, y), \tag{3.13}$$

where

$$\tilde\mu^\perp_{\kappa(\omega_1)}(x, y) = \frac{p(x)(\kappa(\omega_1), y)}{\kappa_*(p(x))(\kappa(\omega_2))}. \tag{3.14}$$

Observe that (3.13) is equivalent to the following

$$p_1(x, (\kappa(\omega_1), y)) = \tilde\mu^\perp_{\kappa(\omega_1)}(x, y)\kappa_*(\bar p_1)(x, \kappa(\omega_1)). \tag{3.15}$$

(3.15) implies that $\tilde\mu^\perp_{\kappa(\omega_1)}(x, y)$ coincides with $r(x, \omega_1)$. Now we obtain Theorem 3.10 from Lemma 3.3 immediately. □

Using Lemma 3.3 and Theorem 3.10 we will present a proof of the monotonicity theorem (Theorem 3.11), also called the Cramér-Rao inequality, which characterizes sufficient statistics in terms of the Fisher information metric.

Theorem 3.11. *(Cramér-Rao inequality, cf.* [4, Theorem 2.1]*). Assume that Ω_1 and Ω_2 are smooth manifolds provided with Borel σ-algebra and μ_1 is a measure on Ω_1 dominated by a Lebesgue measure. Let $(M, \Omega_1, \mu_1, p_1)$ be a k-integrable parametrized measure model and $\kappa : \Omega_1 \to \Omega_2$ a statistic. Denote by $\tilde g^F$ the Fisher metric on the transformed parametrized measure model $(M, \Omega_2, \kappa_*(\mu_1), \kappa_*(p_1))$. For each $x \in M$ and each $V \in T_xM$ we have*

$$\tilde g^F(V, V)_x \le g^F(V, V)_x, \tag{3.16}$$

Inequality (3.16) becomes an equality for all $x \in M$ and for all $V \in T_xM$ if and only if the statistic κ is sufficient for the parameter $x \in M$.

Proof. Since the space of smooth maps is dense in the space of measurable maps with respect to the L^p topology, for any $p \ge 1$, and taking into account $\partial_V\kappa_*(\bar p) = \kappa_*(\partial_V \bar p)$, for a proof of Theorem

3.11, we can assume that κ is smooth. Denote by Ω_2^{reg} the set of regular values of κ. Using (3.8), we obtain

$$g^F(V,V)_x = \int_{\Omega_2^{reg}} d\kappa_*(p_1(x)) \int_{\kappa^{-1}(\omega_2)} (\partial_V \ln \bar{p}_1(x,y))^2 \, d\mu^{\perp}_{\kappa(\omega_1)}(p_1(x), y). \tag{3.17}$$

Recall that

$$\tilde{g}^F(V,V)_x = \int_{\Omega_2^{reg}} (\partial_V \ln \kappa_*(\bar{p}_1)(x,\omega_2))^2 d\kappa_*(p_1(x)). \tag{3.18}$$

To prove Theorem 3.11, comparing (3.17) with (3.18), it suffices to show that for each $x \in M$ and for each $\omega_2 \in \Omega_2^{reg}$ the following inequality holds

$$\int_{\kappa^{-1}(\omega_2)} (\partial_V \ln \bar{p}_1(x,y))^2 \, \mu^{\perp}_{\omega_2}(p_1(x), y) \ge (\partial_V \ln \kappa_*(\bar{p}_1)(x,\omega_2))^2, \tag{3.19}$$

and the equality holds for all $x \in M$ and all regular values ω_2 if and only if κ is sufficient for the parameter $x \in M$.

Taking into account (3.15) and Lemma 3.8, we note that (3.19) is equivalent to the following inequality

$$\int_{\kappa^{-1}(\omega_2)} (\partial_V \ln \kappa_*(\bar{p}_1)(x,\omega_2) + \partial_V \ln \tilde{\mu}^{\perp}_{\omega_2}(x,y))^2 \, d\mu^{\perp}_{\omega_2}(p_1(x), y) \ge$$

$$\int_{\kappa^{-1}(\omega_2)} (\partial_V \ln \kappa_*(\bar{p}_1)(x,\omega_2))^2 \, d\mu^{\perp}_{\omega_2}(p_1(x), y). \tag{3.20}$$

Lemma 3.12. *For all $x \in M$ we have*

$$\int_{\{y \in \kappa(\omega_1)\}} \partial_V \ln \tilde{\mu}^{\perp}_{\kappa(\omega_1)}(x,y) d\mu^{\perp}_{\kappa(\omega_1)}(p_1(x), y) = 0. \tag{3.21}$$

Proof. Writing $\mu^{\perp}_{\kappa(\omega_1)}(p_1(x)) = \tilde{\mu}^{\perp}_{\kappa(\omega_1)}(x,y)\mu^{\perp}_{\kappa(\omega_1)}(\mu_1)$, we observe that (3.21) is a consequence of the following identity for all $x \in M$:

$$\int_{\{y \in \kappa(\omega_1)\}} \tilde{\mu}^{\perp}_{\kappa(\omega_1)}(x,y) d\mu^{\perp}_{\kappa(\omega_1)}(\mu_1, y) = 1,$$

whose validity follows from Lemma 3.8. □

Clearly (3.20) follows from Lemma 3.12, since $\partial_V \ln \kappa_*(\bar{p}_1)(x,\omega_2)$ does not depend on y. Note that (3.20), and hence (3.19), becomes an equality if and only if $\mu^{\perp}_{\kappa(\omega_1)}(p_1(x))$ is independent of x. By Theorem 3.10 the last condition is equivalent to the sufficiency of the statistic κ for the parameter $x \in M$. This proves Theorem 3.11. □

Remark 3.13. 1. Assume that a statistic κ is smooth. Denote by $\hat{g}^F_{\omega_2}$ the Fisher quadratic form on the statistical model $\mu^{\perp}_{\omega_2}(p_1(x), y)$ with respect to the reference measure $\mu^{\perp}_{\kappa(\omega_1)}(\mu_1, y)$ as in (3.14). Taking into account (3.17), (3.18) and (3.21) we obtain immediately the following equality for all $x \in M$ and all $V \in T_x M$ (cf. [4, Theorem 2.1])

$$g^F(V,V) = \tilde{g}^F(V,V) + \int_{\Omega_2} \hat{g}^F_{\omega_2}(V,V) d\kappa_*(p_1(x)). \tag{3.22}$$

The integral in the RHS of (3.22) is called the information loss [4, p.30].

Example 3.14. Let Ω_n be a finite set of n elements $E_1, \cdots, E_n$. Let μ_n be the probability distribution on Ω_n such that $\mu_n(E_i) = 1/n$ for $i \in [1,n]$. Clearly, the space $\mathcal{P}(\Omega_n, \mu_n)$ consists of all probability distributions p on Ω_n which can be represented as

$$p(E_i) = f(E_i)\mu_n \text{ for } i \in [1,n] \tag{3.23}$$

for some non-negative function $f : \Omega_n \to \mathbb{R}$ such that $\sum_{i=1}^n f(E_i) = n$. Denote by E_i^* the Dirac measure on Ω_n concentrated at E_i. The space $\mathcal{M}_+(\Omega_n, \mu_n)$ of measures equivalent to μ_n consists of all measures $p = \sum_{i=1}^n p_i E_i^*, p_i > 0$, so it is the positive cone $\mathbb{R}^n_+$. Let $n \le m < \infty$. Let $\{\hat{F}_1, \cdots, \hat{F}_n\}$ be a partition of the set $\Omega_m := \{F_1, \cdots, F_m\}$ into disjoint subsets. Denote this partition by $\bar{\kappa}$. We associate $\bar{\kappa}$ with a map $\kappa : \Omega_m \to \Omega_n$ by setting

$$\kappa(x) := E_i \text{ if } x \in \hat{F}_i.$$

We identify $\mathcal{M}_+(\Omega_m, \mu_m)$ with $\mathbb{R}^m_+$ which is generated by the Dirac measures $F_j^*, j \in [1, m]$. Recall that a linear mapping $\Pi : \mathbb{R}^n \to \mathbb{R}^m$, $\Pi(E_k^*) := \sum_{j=1}^m \Pi_{kj} F_j^*$, is called *a Markov mapping*, if $\Pi_{ij} \ge 0$ and $\sum_{j=1}^m \Pi_{kj} = 1$ (cf. Example 4.6). Following Chentsov [14, p. 56 and Lemma 9.5, p. 136], we call Π *a Markov congruent embedding subjected to a partition* $\bar{\kappa}$ if

- $F_j \notin \kappa^{-1}(E_i) \implies \Pi(E_i^*)(F_j) = 0$,
- $\Pi(E_i^*) \ne 0$ for all $i \in [1, n]$.

Note that $\Pi(\mathcal{M}(\Omega_n, \mu_n)) \subset \mathcal{M}(\Omega_m, \mu_m)$. The restriction of Π to $\mathbb{R}_{\ge 0} = \mathcal{M}(\Omega_n, \mu_n)$ is also denoted by Π.

Proposition 3.15. *Let $\Pi : \mathcal{M}(\Omega_n, \mu_n) \to \mathcal{M}(\Omega_m, \mu_n)$ be the restriction of a Markov mapping such that the image $(\Pi(\mathcal{M}_+(\Omega_n, \mu_n)), \Omega_n, \mu_n, i)$ is an immersed statistical model of dimension n, where i is the canonical embedding. A statistic $\kappa : \Omega_m \to \Omega_n$ is sufficient for the parameter $x \in \Pi(\mathcal{M}_+(\Omega_n, \mu_n))$, if Π is a Markov congruent embedding subjected to κ.*

Proof. Assume that Π is a Markov congruent embedding subjected to a statistic $\kappa : \Omega_m \to \Omega_n$. Then $\kappa_* \circ \Pi = Id$. By the monotonicity of the Markov morphism, see Corollary 4.11 below, κ must be sufficient for the parameter $x \in \Pi(\mathcal{M}_+(\Omega_n, \mu_n))$. □

Since $\kappa_* \circ \Pi = Id$ for Markov congruent embeddings Π, using Theorem 3.5 we obtain immediately

Corollary 3.16. *Let $\Pi : \mathcal{M}(\Omega_n, \mu_n) \to \mathcal{M}(\Omega_m, \mu_m)$ be a Markov congruent embedding. Then the Amari-Chentsov structure on $\mathcal{M}_+(\Omega_n, \mu_n)$ coincides with the Amari-Chentsov structure on $(\mathcal{M}_+(\Omega_n, \mu_n), \Omega_m, \mu_m, \Pi)$.*

A variant of Proposition 3.15 has been proved by Chentsov [14, Lemma 6.1, p.77 and Lemma 9.5, p.136], see also Proposition 4.7 below. It plays a decisive role in the Chentsov's work [14] on geometric structures on statistical models (M, Ω_n, μ_n, p) that are invariant under sufficient statistics. It implies that such geometric structures are preserved under Markov congruent embeddings, which are easier to understand. (Chentsov's arguments for finding invariant geometric structures have been re-exposed by Campbell in [9].) Since $\mathcal{M}(\Omega_n, \mu_n)$ is canonically isomorphic to $\mathbb{R}^n_{\ge 0}$, any map $p : M \to \mathcal{M}(\Omega_n, \mu_n)$ can be written as $p := (p_1, \cdots, p_n)$. We resume some results in [14], which are important for our paper, in the following

Proposition 3.17. *(1) (cf.* [14, Lemma 11.1 p. 157]*) Assume that C is a continuous function on statistical models (M, Ω_n, μ_n, p) associated with finite sample spaces $\{\Omega_n\}$ such that C is invariant under sufficient statistics. Then C is a constant.*

(2) (cf. [14, Lemma 11.2, p. 158]*) Assume that A is a continuous 1-form field on parametrized measure models $(M, \Omega_n, \mu_n, p = (p_1, \cdots, p_n))$ associated with finite sample spaces $\{\Omega_n\}$ such that A is invariant under sufficient statistics. Then there is a continuous function $c : \mathbb{R} \to \mathbb{R}$ such that for all $x \in M$ and all $V \in T_x M \subset \mathbb{R}^n$, $A_x(V) = c(\sum_{i=1}^n p_i(x)) \sum_{i=1}^n p_i \partial_V \ln p_i(x)$. In particular, there is no continuous 1-form field on statistical models associated with finite sample spaces that is invariant under sufficient statistics.*

(3) (cf. [14, Theorem 11.1, p. 159]*) Assume that F is a continuous quadratic form field on parametrized measure models (M, Ω_n, μ_n, p) associated with finite sample spaces $\{\Omega_n\}$ such that F*

is invariant under sufficient statistics. Then there is a continuous function $f : \mathbb{R} \to \mathbb{R}$ *such that* $F = f(\sum_{i=1}^n p_i(x)) \cdot g^F + A^2$ *where* A *is the 1-form field described in (2) and* g^F *is the Fisher metric. In particular, the Fisher metric is the unique up to a constant quadratic form on statistical models that is invariant under sufficient statistics.*

(4) (cf. [14, Theorem 12.2, p.175]*) Assume that* T *is a continuous covariant symmetric 3-tensor field on parametrized measure models* (M, Ω_n, μ_n, p) *associated with finite sample spaces* $\{\Omega_n\}$ *such that* F *is invariant under sufficient statistics. Then there is a continuous function* $t : \mathbb{R} \to R$ *such that* $T = t(\sum_{i=1}^n p_i(x)) \cdot T^{AC} + g^F \cdot A_2 + A_1^3$ *where* g^F *and* T^{AC} *are the Fisher metric and the Amari-Chentsov tensor respectively, and* A_1, A_2 *are the fields described in (2).*

The argument of Chentsov for proving (2) rests on the permutation invariance, because a map from Ω_n to itself that permutes the points of Ω is clearly a sufficient statistic. And from this permutation invariance, one easily obtains that A has to be of the form given in (2), that is, constant, and that this constant has to vanish in the statistical case. (3) and (4) can then be deduced from some general arguments about tensors. Note that in [14] Chentsov only gave a proof of Proposition 3.17 for statistical models (M, Ω_n, μ_n, p). The extension of Proposition 3.17 to parametrized measure models associated with Ω_n can be obtained easily using the following lemma.

Lemma 3.18. *Assume* (M, Ω, μ, p) *is a parametrized measure model and* $\kappa : \Omega \to \Omega'$ *is sufficient for the parameter* $x \in (M, \Omega, \mu, p)$*. Then* κ *is also sufficient for the parameter* $x \in (M \times (0,1), \Omega, \mu, \hat{p}(x,t) := tp(x))$*.*

Proof. Since $\kappa_*(t\mu) = t\kappa_*(\mu)$ for any finite measure μ on Ω and $t \in \mathbb{R}^+$, we get

$$\frac{d(tp(x))}{d\kappa_*(tp(x))} = \frac{dp(x)}{d\kappa_*(p(x))}.$$

Taking into account Lemma 3.3, this proves Lemma 3.18. □

Using Lemma 3.18 we obtain the second assertion (2) of Proposition 3.17 from its particular case for statistical models and the first assertion (1), since each 1-form $A \in T_p^*\mathcal{M}_+(\Omega_n, \mu_n)$ is a sum of two linear independent 1-forms A_0 and $A^\perp$, where A_0 annihilates the tangent hyperplane $T_p\mathcal{M}_+^{p_1+\cdots+p_n}(\Omega_n, \mu_n) \subset T_p\mathcal{M}_+(\Omega_n, \mu_n)$, and $A^\perp = A - A_0$.

Using the same argument we obtain the third assertion of Proposition 3.17 from its particular case for statistical models and the second assertion.

The last assertion of Proposition 3.17 is obtained from its particular case for statistical models, the second and third assertion.

We also note that in [9] Campbell gave a detailed proof of the third assertion of Proposition 3.17 using Chentsov's argument in [14]. Our proof of the Main Theorem (Theorem 2.9) is based on the following main observation. For each step function τ on (Ω, μ) subject to a statistic $\kappa : (\Omega, \mu) \to \Omega_n := \{E_1, \cdots, E_n\}$ (Definition 3.19) there exists a parametrized measure model (M, Ω, μ, p) and a vector $V \in T_xM$ such that $p(x) = \mu$ and $\partial_V \ln p = \tau$, moreover, κ is sufficient with respect to the parameter $x \in M$ (Lemma 3.20). Thus, the computation of any pointwise continuous covariant k-tensor field on $\mathcal{M}(\Omega)$, whose induced k-tensor field on parametrized measure models is invariant under sufficient statistics, is reduced to the case $\Omega = \Omega_n$, which has been considered by Chentsov for $k = 1, 2, 3$.

Definition 3.19. (cf. Example 3.14) Let (Ω, μ) be a finite measure space and let $\bar{\kappa}$ be a decomposition $\Omega = D_1 \dot{\cup} \ldots \dot{\cup} D_n$ where D_i is measurable. Denote by κ the associated statistic $\Omega \to \Omega_n$, $\kappa(D_i) := E_i$. A function $\tau : \Omega \to \mathbb{R}$ is called *a step function subject to* κ, if $\tau(\omega) = \tau_i \cdot \chi_{D_i}(\omega)$, where $\tau_i \in \mathbb{R}$ and χ_{D_i} is the characteristic function of D_i.

Lemma 3.20. *Let $M = (0,1)$ and Ω be a smooth manifold. Given a finite measure $\mu \in \mathcal{M}(\Omega)$, a point $x_0 \in M$, and a step function $\tau := \sum_i \tau_i \chi_{D_i}$ on Ω subject to a statistic $\kappa : (\Omega, \mu) \to \Omega_n$, there exist a k-integrable parametrized measure model (M, Ω, μ, p) and $V \in T_{x_0} M$ such that*

(1) *κ is sufficient for the parameter in M,*
(2) *$p(x_0) = \mu$,*
(3) *$\partial_V \ln \bar{p} = \sum_i \tau_i \chi_{D_i}$.*

Proof. Note that κ is a sufficient statistic for a k-integrable parametrized measure model (M, Ω, μ, p) iff p is given as in Definition 3.1, i.e.

$$\ln \bar{p}(x, \omega) = \ln \bar{p}(x, \kappa(\omega)) + \ln t(\omega) = \sum_{i=1}^n s_i(x) \chi_{D_i}(\omega) + \ln t(\omega)$$

for smooth functions $s_i : M \to \mathbb{R}$ and $t \in L^1(\Omega)$. For such $\ln \bar{p}(x, \omega)$ the conditions (2) and (3) are equivalent to the following

- $\sum_{i=1}^n s_i(x_0) \chi_{D_i}(\omega) + \ln t(\omega) = 0$,
- $\sum_{i=1}^n \partial_V s_i(x_0) \chi_{D_i}(\omega) = \sum_i \tau_i \chi_{D_i}(\omega)$.

Set $t(\omega) = 1$. The existence of functions $s_i(s)$ satisfying the listed conditions is obvious: it suffices to choose smooth s_i such that $s_i(x_0) = 0$ and $\partial_V s_i(x) = \tau_i$. In fact, we can simply take $V = \partial_x$ and $s_i(x) = (x - x_0)\tau_i$. Finally, one verifies that the defined parametrized measure model is k-integrable, since the s_i are smooth. □

Proof of the Main Theorem. 1. Let A be a pointwise continuous 1-tensor field on $\mathcal{M}(\Omega)$ satisfying the condition (1) in the Main Theorem. To prove the first assertion of the Main Theorem, it suffices to assume that V is a step function $\sum_i \tau_i \chi_{D_i}$ (using again the identification between the tangent vector V and $\partial_V \ln \bar{p}$) subject to a statistic $\kappa : \Omega \to \Omega_n$. By Lemma 3.20 there exists a 1-generalized statistical model (M, Ω, μ, p) such that

(1) $p(x, \omega) = e^{s_i(x) \chi_{D_i}}$, where $s_i \in C^\infty(M)$, hence κ is sufficient for the parameter $x \in M$,
(2) $p(x_0) = \mu$,
(3) $\partial_V \ln \bar{p}(x, \omega) = \sum_i \tau_i \chi_{D_i}$.

Set

$$d_i := \int_{D_i} d\mu.$$

Then $\kappa_*(\mu) = d_i E_i^*$, where E_i^* is the Dirac measure concentrated at E_i. Since A is associated with a statistical field which is invariant under κ_* we have

$$A_\mu(\tau) = (A_{\kappa_*(\mu)}(\partial_V(\ln \kappa_*(\bar{p}))) = A_{(d_1, \cdots, d_n)}(\tau_1, \cdots, \tau_n) = c(\sum_{i=1}^n d_i) \sum_{i=1}^n d_i \tau_i, \tag{3.24}$$

where c is the function defined in Proposition 3.17.2. Note that

$$\sum_i d_i = \int_\Omega d\mu,$$

$$\sum_i d_i \tau_i = \sum_i (\int_{D_i} \tau_i d\mu) = \int_\Omega \tau d\mu.$$

This proves the first assertion in the Main theorem. The next assertions of the Main theorem concerning specification of the covariant 1-tensor field A follows immediately.

2. Now assume that F is a pointwise continuous quadratic form on $\mathcal{M}(\Omega)$ and μ is a finite measure. To prove the second assertion of the Main Theorem we follow the same line of arguments as above. It suffices to prove the validity of the second assertion for a step function τ on Ω, since F

is a quadratic form (otherwise we have to consider step functions subjected to different statistics). We deduce the second assertion of the Main Theorem from Proposition 3.17.2 using the observation that the Fisher metric on $\mathcal{M}(\Omega,\mu)$ applied to τ

$$g^F_\mu(\tau) = \int_\Omega \tau^2 d\mu = \sum_{i=1}^n d_i \tau_i^2$$

is equal to the Fisher metric applied to $\kappa_*(\tau) = (\tau_1, \cdots, \tau_n)$

$$g^F_{(d_1,\cdots,d_n)}([\tau_1,\cdots,\tau_n]) = \sum_{i=1}^n d_i \tau_i^2.$$

3. The last assertion of the Main Theorem is proven in the same way. It follows from Proposition 3.17.2 using the observation that the Amari-Chentsov 3-symmetric tensor on $\mathcal{M}(\Omega,\mu)$ applied to τ

$$T^{AC}_\mu(\tau) = \int_\Omega \tau^3 d\mu = \sum_{i=1}^n d_i \tau_i^3$$

is equal to the Amari-Chentsov tensor applied to $\kappa_*(\tau) = (\tau_1, \cdots, \tau_n)$

$$T^{AC}_{(d_1,\cdots,d_n)}([\tau_1,\cdots,\tau_n]) = \sum_{i=1}^n d_i \tau_i^3.$$

To complete the proof of the Main Theorem we need to show that

(1) all the tensor fields described in the Main Theorem are weakly continuous on n-integrable parametrized measure models,
(2) the tensor field A is invariant under sufficient statistics.

Note that (1) holds since the associated k-tensor fields τ on $\mathcal{M}(\Omega)$ are weakly bounded, i.e., for any $\mu \in \mathcal{M}(\Omega)$ there is a continuous function $c : \mathcal{M}(\Omega,\mu) \to \mathbb{R}$ such that $|\tau_\mu(V)| \le c(\mu)||V||_{L^n(\Omega,\mu)}$. The proof of (2) is similar to the proof of Theorem 3.5, observing that

$$\partial_V p(x) = \partial_V \ln \bar{p}(x)\bar{p}(x)\mu$$

for $p(x) = \bar{p}(x)\mu$ (cf. Remark 2.4), and hence omitted. □

Remark 3.21. Our proof of the Main Theorem is based on the fact that the step functions are dense in the value space $L^n(\Omega,\mu)$. The same proof applies if we replace the value space $L^n(\Omega,\mu)$ by another function space depending on the finite measure μ where the step functions are dense.

4. Markov morphisms and sufficient statistics

In this section we introduce the notions of a Markov morphism, a μ-representable Markov morphism, and a restricted Markov morphism (Definitions 4.1, 4.2, 4.4) extending the Chentsov notion of a Markov morphism [12], and the notion of a statistical morphism introduced independently by Morse and Sacksteder in [24]. These notions are needed for comparing two statistical models; they stem from the Blackwell concept of "comparison of experiments" in [7]. A novel aspect is our consideration of a parametrization of the parameter space M of a parametrized measure model (M,Ω,μ,p) as a restricted Markov morphism (Definition 4.4, Example 4.5). Thus, the geometry of parametrized measure models is intrinsic (Example 4.5). We decompose a Markov morphism associated with a (positive) Markov transitition kernel as a composition of a right inverse of a sufficient statistic and a statistic (Theorem 4.10). As a consequence we give a geometric proof of the monotonicity theory for Markov morphisms (Corollary 4.11).

Positivity assumption. In this section, for the simplicity of the exposition of the theory, we enlarge the class of parametrized measure models to include also (M,Ω,μ,p), where $p : M \to$

$\mathcal{M}(\Omega, \mu)$. This assumption is caused solely by the fact that if the Markov transition kernel $\Pi(\omega, \omega')$ is not everywhere positive, then the transformed density $p(x, \omega') := \int_\Omega \Pi(\omega, \omega') d\mu$ need not be everywhere positive on $(\Omega', \Pi_*(\mu))$. Alternatively, when considering Markov transition kernels we restrict ourselves to positive ones.

Definition 4.1. ([12, p. 194], [24, p. 205]) *A Markov transition from a measurable space* (Ω, Σ) *to a measurable space* (Ω', Σ') is a map $T : \Omega \to \mathcal{P}(\Omega', \Sigma')$ such that for each $S \in \Sigma'$ the function $\int_S d(T(x))$ is a Σ-measurable function. A Markov transition $T : \Omega \to \mathcal{P}(\Omega', \Sigma')$ defines *a Markov morphism* $T_* : \mathcal{M}(\Omega, \Sigma) \to \mathcal{M}(\Omega', \Sigma')$ by

$$T_*(\nu)(S) := \int_\Omega \int_S d(T(\omega)) d\nu \tag{4.1}$$

for $S \in \Sigma'$.

Since $T(\Omega) \subset \mathcal{P}(\Omega', \Sigma')$, substituting $S := \Omega'$ in (4.1), we obtain

$$T_*(\mathcal{M}^a(\Omega, \Sigma)) \subset \mathcal{M}^a(\Omega', \Sigma')$$

for all $a \in \mathbb{R}^+$.

Next, we assume that $T(\omega)$ is dominated by a probability measure $\mu' \in \mathcal{P}(\Omega', \Sigma')$. Then there exists a measurable function $\Pi_\omega : \Omega' \to \mathbb{R}$ such that for all $S \in \Sigma'$ we have

$$T(\omega)(S) = \int_S \Pi_\omega(\omega') d\mu'. \tag{4.2}$$

If $T(\Omega) \subset \mathcal{P}(\Omega', \mu')$, by (4.2), there exists *a Markov transition kernel* $\Pi : \Omega \times \Omega' \to \mathbb{R}$ *from* Ω *to* $\mathcal{M}(\Omega', \mu')$ such that

$$T(\omega)(S) = \int_S \Pi(\omega, \omega') d\mu'. \tag{4.3}$$

Definition 4.2. If (4.3) holds, $T(\Pi) := T$ is called *a μ'-representable Markov transition*, and $T(\Pi)_*$ is called *a μ'-representable Markov morphism.*

Note that any Markov transition kernel $\Pi : \Omega \times \Omega' \to \mathbb{R}$ from Ω to $\mathcal{P}(\Omega', \mu')$ satisfies

$$\Pi(\omega, \omega') \geq 0 \text{ for all } (\omega, \omega') \in \Omega \times \Omega', \tag{4.4}$$

$$\int_{\Omega'} \Pi(\omega, \omega') d\mu' = 1 \text{ for all } \omega \in \Omega. \tag{4.5}$$

Abbreviate $T(\Pi)_*$ as Π_*. For any measure $\nu \in \mathcal{M}(\Omega)$ and $S \in \Sigma'$ we have

$$\Pi_*(\nu)(S) = \int_\Omega \int_S \Pi(\omega, \omega') d\mu' \, d\nu. \tag{4.6}$$

It follows

$$\frac{d\Pi_*(\nu)}{d\mu'}(\omega') = \int_\Omega \Pi(\omega, \omega') d\nu. \tag{4.7}$$

If Ω, Ω' are finite sets, then any Markov morphism $T : \mathcal{M}(\Omega, \Sigma) \to \mathcal{M}(\Omega', \Sigma')$ is μ-representable for any dominant measure μ on Ω', see also Example 4.6. This is not true, if Ω, Ω' are open domains in $\mathbb{R}^n$, $n \geq 1$.

Example 4.3. 1. (cf. [12, p. 511]) Let (Ω, Σ) be a measurable space. We define a Markov transition T^{Id} on (Ω, Σ) by setting

$$T^{Id}(\omega)(A) := \chi_A(\omega) \text{ for } \omega \in \Omega,$$

where χ_A is the indicator function of $A \in \Sigma$. Clearly T_*^{Id} defines a Markov morphism which is the identity transformation of $\mathcal{P}(\Omega, \Sigma)$. Note that T_*^{Id} is not a μ-representable Markov morphism for any measure $\mu \in \mathcal{M}(\Omega, \Sigma)$, if Ω is an open domain in $\mathbb{R}^n$ with Borel σ-algebra Σ, and $n \geq 1$. To see this, we note that if μ dominates all the measures $T^{Id}(\omega), \omega \in \Omega$, then μ has no null set, in particular $\mu(\{\omega\}) > 0$ for all $\omega \in \Omega$. It is easy to see that this is impossible, since $\dim \Omega \geq 1$.

2. Assume that $\kappa : \Omega_1 \to \Omega_2$ is a statistic. Then κ defines a Markov transition T^κ from (Ω_1, Σ_1) to (Ω_2, Σ_2) by setting

$$T^\kappa(\omega_1)(A) := \chi_A(\kappa(\omega_1)) \text{ for } \omega_1 \in \Omega_1 \tag{4.8}$$

and $A \in \Sigma_2$. For $\nu \in \mathcal{M}(\Omega_1)$ and $S \in \Sigma_2$, using (4.1), we get

$$T_*^\kappa(\nu)(S) = \int_{\Omega_1} \int_S d\chi_A(\kappa(\omega_1)) d\nu = \int_{\kappa^{-1}(S)} d\nu.$$

Hence $T_*^\kappa = \kappa_*$. Then T_*^κ is not a μ_2-representable Markov morphism for any $\mu_2 \in \mathcal{M}(\Omega_2)$, if for instance $\kappa(\Omega_1)$ and Ω_2 are open domains in $\mathbb{R}^n$, $n \geq 1$, since there exists $\nu \in \mathcal{M}(\Omega_1)$ such that $\kappa_*(\nu)$ is not dominated by μ_2.

Denote by $C^1(M_1, M_2)$ the space of all differentiable maps from a differentiable manifold M_1 to a differentiable manifold M_2. Let (Ω_1, Σ_1) and (Ω_2, Σ_2) be measurable spaces. Denote by $\mathfrak{M}(\Omega_1, \Omega_2)$ the set of all Markov morphisms from $\mathcal{M}(\Omega_1)$ to $\mathcal{M}(\Omega_2)$.

Definition 4.4. Assume that $(M_1, \Omega_1, \mu_1, p_1)$ and $(M_2, \Omega_2, \mu_2, p_2)$ are parametrized measure models. A pair $(f \in C^1(M_1, M_2), T \in \mathfrak{M}(\Omega_1, \Omega_2))$ is called *a restricted Markov morphism*, if for all $x \in M$

$$p_2(f(x)) = T_*(p_1(x)). \tag{4.9}$$

Example 4.5. 1. Assume that $(M, \Omega_1, \mu_1, p_1)$ is a parametrized measure model and $\kappa : \Omega_1 \to \Omega_2$ is a statistic. Then $(M, \Omega_2, \kappa_*(\mu_1), \kappa_*(p_1))$ is a parametrized measure model. By Example 4.3.2 the pair (Id, κ_*) is a Markov morphism. We also call (Id, κ_*) a statistic if no misunderstanding occurs.

2. Assume that $(M_2, \Omega_2, \mu_2, p_2)$ is a parametrized measure model and $f : M_1 \to M_2$ is a smooth map. Then $(M_1, \Omega_2, \mu_2, p_1 := p_2 \circ f)$ is a parametrized measure model and the pair (f, Id) is a Markov morphism. Such a Markov morphism is called *generated by a smooth map* f. It is easy to see that, if f is a diffeomorphism, then the Amari-Chentsov structure on M_1 is obtained from the Amari-Chentsov structure on M_2 via the pull-back map f^*.

Example 4.6. Let (Ω_n, μ_n) and (Ω_m, μ_m) be the measure spaces in Example 3.14. Let $\Pi : \Omega_n \times \Omega_m \to \mathbb{R}$ be a mapping such that $\Pi_{i,j} := \Pi(E_i, F_j)$ satisfies the following conditions

$$\Pi_{i,j} \geq 0 \text{ for all } 1 \leq i \leq n,\ 1 \leq j \leq m,$$

$$\sum_{j=1}^m \Pi_{i,j} = 1 \text{ for all } 1 \leq i \leq n. \tag{4.10}$$

Clearly, Π is a Markov transition kernel from Ω_n to $\mathcal{M}(\Omega_m, \mu_m)$. By (4.6) Π induces a map

$$\Pi_* : \mathbb{R}^n_{\geq 0} = \mathcal{M}(\Omega_n, \mu_n) \to \mathcal{M}(\Omega_m, \mu_m) = \mathbb{R}^m_{\geq 0},$$

$$\Pi_*(E_k^*)(F_j) := \sum_{i=1}^n \Pi_{i,j} E_k^*(E_i) = \Pi_{kj}. \tag{4.11}$$

Hence

$$\Pi_*(E_k^*) = \sum_{j=1}^m \Pi_{kj} F_j^*. \tag{4.12}$$

Let

$$(M_1 := \mathcal{P}_+(\Omega_n, \mu_n), \Omega_n, \mu_n, p_1(x) := x),$$
$$(M_2 := \mathcal{P}_+(\Omega_m, \mu_m), \Omega_m, \mu_m, p_1(y) := y)$$

be statistical models. By (4.9), a pair $(f \in \mathrm{Diff}(M_1, M_2), \Pi \in \mathfrak{M}(\Omega_n, \Omega_m))$ is a Markov morphism, if and only if for all $x \in M_1$

$$f(x)(F_j) = \Pi_*(x)(F_j) \text{ for all } 1 \le j \le m. \tag{4.13}$$

Thus for $\Pi \in \mathfrak{M}(\Omega_n, \Omega_m)$ the pair (f, Π) is a Markov morphism if and only if $f = \Pi_*|_{M_1}$. We also abbreviate $(\Pi_*|_{M_1}, \Pi)$ as Π if no misunderstanding occurs.

Next we drop the assumption that $n \le m$. Note that there is a canonical map

$$\chi_n : \Omega_n \to \mathcal{M}(\Omega_n, \mu_n), E_i \mapsto E_i^*.$$

Let $\kappa : \Omega_n \to \Omega_m$ be a statistic. The composition $\chi_m \circ \kappa : \Omega_n \to \mathcal{M}(\Omega_m, \mu_m)$ defines the following map $\Pi^\kappa : \Omega_n \times \Omega_m \to \mathbb{R}$

$$\Pi^\kappa(E_i, F_j) := \langle \chi_m \circ \kappa(E_i), F_j \rangle \tag{4.14}$$

Clearly $\sum_{j=1}^m \Pi^\kappa(E_i, F_j) = 1$ for all i. Hence Π^κ is a Markov transition kernel. Note that $\Pi^\kappa_* : \mathcal{M}(\Omega_n, \mu_n) \to \mathcal{M}(\Omega_m, \mu_m)$ coincides with the push-forward map $\kappa_* : \mathcal{M}(\Omega_n, \mu_n) \to \mathcal{M}(\Omega_m, \mu_m)$.

Proposition 4.7. *A linear mapping $\Pi : \mathbb{R}^n \to \mathbb{R}^m$ is a Markov congruent embedding subjected to a statistic κ, if and only if $\Pi^\kappa_* \circ \Pi(x) = x$ for all $x \in \mathbb{R}^n_{\ge 0}$. A Markov mapping $\Pi : \mathbb{R}^n \to \mathbb{R}^m$ has a left inverse if and only if it is a Markov congruent embedding.*

The first assertion of Proposition 4.7 is obvious. The second assertion of Proposition 4.7 is a reformulation of [14, Lemma 6.1, p. 77 and Lemma 9.5, p.136].

Let $(M, \Omega_1, \mu_1, p_1)$ be a parametrized measure model, (Ω_2, μ_2) a measure space and $\Pi : \Omega_1 \times \Omega_2 \to \mathbb{R}$ a Markov transition kernel from Ω_1 to $\mathcal{P}(\Omega_2, \mu_2)$. Set for each $x \in M$

$$\Pi^{[p]}(x, \omega_1, \omega_2) := \Pi(\omega_1, \omega_2)\bar{p}(x, \omega_1) \tag{4.15}$$

By (4.5) we get

$$\int_{\Omega_1 \times \Omega_2} \Pi^{[p]}(x, \omega_1, \omega_2)\mu_1\mu_2 = \int_{\Omega_1} \bar{p}_1(x, \omega_1) d\mu_1. \tag{4.16}$$

Lemma 4.8. *Then $(M, \Omega_1 \times \Omega_2, \mu_1\mu_2, \Pi^{[p]}(x, \omega_1, \omega_2))$ is a (generalized) statistical model. Moreover, the Amari-Chentsov structure on $(M, \Omega_1 \times \Omega_2, \mu_1\mu_2, \Pi^{[p]})$ coincides with the Amari-Chentsov structure on (M, Ω_1, μ_1, p).*

Proof. The first assertion of Lemma 4.8 follows from Lemma 3.3 and Corollary 4.9.

We present two proofs of the second assertion of Lemma 4.8.

First proof. Let us compute the Fisher quadratic form on $(M, \Omega_1 \times \Omega_2, \mu_1\mu_2, \Pi^{[p]}\mu_1\mu_2)$ using (4.15) and (4.5)

$$\begin{aligned} g^F(V, W)_x = \int_{\Omega_1 \times \Omega_2} (\partial_V \ln \Pi^{[p]}(x, \omega_1, \omega_2))(\partial_W \ln \Pi^{[p]}(x, \omega_1, \omega_2))\Pi^{[p]}(x, \omega_1, \omega_2) d\mu_2 d\mu_1 = \\ = \int_{\Omega_1} \int_{\Omega_2} (\partial_V \ln \bar{p}(x, \omega_1))(\partial_W \ln \bar{p}(x, \omega_1))\bar{p}(x, \omega_1)\Pi(\omega_1, \omega_2) d\mu_2 d\mu_1 = \\ = \int_{\Omega_1} (\partial_V \ln \bar{p}(x, \omega_1))(\partial_W \ln \bar{p}(x, \omega_1))\bar{p}(x, \omega_1) d\mu_1. \end{aligned} \tag{4.17}$$

In the same way, taking into account

$$\partial_V \ln \Pi^{[p]}(x,\omega_1,\omega_2) = \partial_V \ln \bar{p}(x,\omega_1),$$
$$\int_{\Omega_2} \Pi(\omega_1,\omega_2)\, d\mu_2 = 1,$$

we conclude that the Amari-Chentsov tensor T^{AC} on $(M, \Omega_1 \times \Omega_2, \mu_1\mu_2, \Pi^{[p]}_{112})$ coincides with the Amari-Chentsov tensor on (M, Ω_1, μ_1, p). This completes the proof of Lemma 4.8.

Second proof. Comparing (3.1) with (4.15) we observe that $\pi_1 : \Omega_1 \times \Omega_2 \to \Omega_1$ is a sufficient statistic with respect to the parameter $x \in (M, \Omega_1 \times \Omega_2, \mu_1\mu_2, \Pi^{[p]}(x,\omega_1,\omega_2))$. Thus Lemma 4.8 is a consequence of Theorem 3.11.2. □

Combining the second proof of Lemma 4.8 and Example 4.5 we obtain

Corollary 4.9. *Let $(M, \Omega_1, d\mu_1, p)$ be a parametrized measure model. The projection $\pi_1 : \Omega_1 \times \Omega_2 \to \Omega_1$ is a sufficient statistic for the parametrized measure model $(M, \Omega_1 \times \Omega_2, \mu_1\mu_2, \Pi^{[p]})$.*

Theorem 4.10. *Let $(Id, \Pi_*) : (M_1, \Omega_1, \mu_1, p_1) \to (M_1, \Omega_2, \mu_2, p_2)$ be a restricted Markov morphism, where Π_* is μ_2-representable by a positive Markov kernel. Then (Id, Π_*) is a composition of a right inverse of a sufficient statistic and a statistic.*

Proof. Let us denote by $\pi_2 : \Omega_1 \times \Omega_2 \to \Omega_2$ the projection onto the second factor. We observe that (Id, Π_*) is a composition of two maps $(Id, \Pi_{1,12}) : (M_1, \Omega_1, \mu_1, p_1) \to (M_1, \Omega_1 \times \Omega_2, \bar{p}_{12}(x,\omega_1,\omega_2) := \bar{p}(x,\omega_1) \cdot \Pi(\omega_1,\omega_2))$ and the push forward map $(\pi_2)_*$. The map $(Id, \Pi_{1,12})$ is the inverse of the sufficient statistic $(Id, (\pi_1)_*)$ by Corollary 4.9. This completes the proof of Theorem 4.10. □

Let $M_1 = M_2$. A restricted Markov morphism of form (f, T_*) is called *representable* if f is a diffeomorphism, and T_* is μ-representable.

Corollary 4.11. *(cf.* [4, p. 31]*) Representable restricted Markov morphisms decrease the Fisher metric on statistical models.*

2. The Fisher metric is the unique up to a constant weakly continuous quadratic 2-form field on statistical models associated with finite sample spaces $\{\Omega_n\}$ that is monotone under representable restricted Markov morphisms.

Proof. The first assertion of Corollary 4.11 is an immediate consequence of Theorem 4.10 and Theorem 3.11.2.

The second assertion of Corollary 4.11 is a consequence of Theorem 4.10 and Proposition 3.17, taking into account the following fact. A congruent Markov embedding $\Pi : \mathcal{P}(\Omega_n, \mu_n) \to \mathcal{P}(\Omega_m, \mu_m)$ subjected to a statistic κ satisfies $\Pi^\kappa_* \Pi(x) = Id$ by Proposition 4.7. Since any quadratic form field on $(\mathcal{M}_+(\Omega_n, \mu_n), \mu_n, p(x) := x)$ that is monotone under Markov morphisms is monotone under Markov congruent embeddings, it follows that such a quadratic form is invariant under sufficient statistics κ_* and also invariant under Markov congruent embeddings. Chentsov's result implies that such a quadratic form is the Fisher metric up to a constant. □

The second assertion of Corollary 4.11 is also valid for statistical models associated with infinite sample spaces. A proof of this assertion will be given in a forthcoming paper.

5. The Pistone-Sempi structure

In this section we study the relations between k-integrable parametrized measure models and statistical models in the Pistone-Sempi theory. First, we show that the Pistone-Sempi manifold is a k-integrable parametrized measure model for any k (Proposition 5.11). We also construct an example of a k-integrable parametrized measure model which does not admit a continuous map into the space $\mathcal{M}_+(\Omega, \mu_0)$ with the topology of Pistone and Sempi (Example 5.12).

In Section 2, we considered the L^1-topology of $\mathcal{M}_+(\Omega, \mu_0)$. However, this set carries also a stronger natural topology, discovered by Pistone and Sempi, which is referred to as *the exponential topology* (also *e-topology*) [28, §2.1]. In fact, Pistone and Sempi considered only the space $\mathcal{P}_+(\Omega, \mu)$ but their theory works also for $\mathcal{M}_+(\Omega, \mu) = \mathcal{P}_+(\Omega, \mu) \times \mathbb{R}^+$. Let us briefly recall the notion of the e-topology, which is defined using the notion of convergence of sequences.

Definition 5.1. [28, Definition 1.1] The sequence $(\mu_n)_{n\in\mathbb{N}}$ in $\mathcal{M}_+(\Omega, \mu)$ is e-convergent (exponentially convergent) to μ if $(\mu_n)_{n\in\mathbb{N}}$ tends to μ in the L^1-topology as $n \to \infty$, and, moreover, the sequences $(d\mu_n/d\mu)_{n\in\mathbb{N}}$ and $(d\mu/d\mu_n)_{n\in\mathbb{N}}$ are eventually bounded in each $L^p(\Omega, \mu)$, $p > 1$, that is, $d\mu_n/d\mu$ and $d\mu/d\mu_n$ converge to 1 with respect to all p-seminorms $L^p(\Omega, \mu)$, $p > 1$.

While $\mathcal{M}_+(\Omega, \mu_0)$ is connected with respect to the L^1-topology, its set of connected components with respect to the e-topology is more interesting. In what follows we briefly describe these components and their structure. Although the stated facts are known from the work of Pistone an Sempi, our presentation is slightly different and illuminates more abstract aspects.

5.1. **Orlicz spaces.** In this section, we briefly recall the theory of Orlicz spaces which is needed in section 5.2 for the description of the geometric structure on $\mathcal{M}(\Omega)$. Most of the results can be found e.g. in [19].

A function $\phi : \mathbb{R} \to \mathbb{R}$ is called a *Young function* if $\phi(0) = 0$, ϕ is even, convex, strictly increasing on $[0, \infty)$ and $\lim_{t\to\infty} t^{-1}\phi(t) = \infty$. Given a finite measure space (Ω, μ) and a Young function ϕ, we define the *Orlicz space*

$$L^\phi(\mu) := \left\{ f : \Omega \to \mathbb{R} \mid \int_\Omega \phi\left(\frac{f}{a}\right) d\mu < \infty \text{ for some } a > 0 \right\},$$

and on $L^\phi(\mu)$ we define the *Orlicz norm*

$$||f||_{\phi,\mu} := \inf \left\{ a > 0 \mid \int_\Omega \phi\left(\frac{f}{a}\right) d\mu \leq 1 \right\}.$$

For any Young function, $(L^\phi(\mu), ||\cdot||_{\phi,\mu})$ is a Banach space. Moreover, a sequence $(f_n)_{n\in\mathbb{N}} \in L^\phi(\mu)$ converges to 0 if and only if

$$\lim_{n\to\infty} \int_\Omega \phi(pf_n)\ d\mu = 0 \qquad \text{for all } p > 0.$$

Proposition 5.2. *Let (Ω, μ) be a finite measure space, and let $\phi_1, \phi_2 : \mathbb{R} \to \mathbb{R}$ be two Young functions. If*

$$\limsup_{t\to\infty} \frac{\phi_1(t)}{\phi_2(t)} < \infty,$$

then $L^{\phi_2}(\mu) \subset L^{\phi_1}(\mu)$, and the inclusion is continuous, i.e., $||f||_{\phi_1,\mu} \leq c\ ||f||_{\phi_2,\mu}$ for some $c > 0$ and all $f \in L^{\phi_2}(\mu)$. In particular, if

$$0 < \liminf_{t\to\infty} \frac{\phi_1(t)}{\phi_2(t)} \leq \limsup_{t\to\infty} \frac{\phi_1(t)}{\phi_2(t)} < \infty,$$

then $L^{\phi_1}(\mu) = L^{\phi_2}(\mu)$, and the Orlicz norms $||\cdot||_{\phi_1,\mu}$ and $||\cdot||_{\phi_2,\mu}$ are equivalent.

Proof. By our hypothesis, $\phi_1(t) \leq K\phi_2(t)$ for some $K \geq 1$ and all $t \geq t_0$. Let $f \in L^{\phi_2}(\mu)$ and $a > ||f||_{\phi_2,\mu}$. Moreover, decompose

$$\Omega := \Omega_1 \dot\cup \Omega_2 \qquad \text{with} \qquad \Omega_1 := \{\omega \in \Omega \mid |f(\omega)| \geq at_0\}.$$

Then

$$\begin{aligned}
K &\geq K\int_\Omega \phi_2\left(\frac{|f|}{a}\right)d\mu \geq \int_{\Omega_1} K\phi_2\left(\frac{|f|}{a}\right)d\mu \\
&\geq \int_{\Omega_1}\phi_1\left(\frac{|f|}{a}\right)d\mu \qquad\qquad \text{as } \frac{|f|}{a}\geq t_0 \text{ on } \Omega_1 \\
&= \int_\Omega \phi_1\left(\frac{|f|}{a}\right)d\mu - \int_{\Omega_2}\phi_1\left(\frac{|f|}{a}\right)d\mu \\
&\geq \int_\Omega \phi_1\left(\frac{|f|}{a}\right)d\mu - \int_{\Omega_2}\phi_1(t_0)d\mu \qquad \text{as } \frac{|f|}{a}< t_0 \text{ on } \Omega_2 \\
&\geq \int_\Omega \phi_1\left(\frac{|f|}{a}\right)d\mu - \phi_1(t_0)\mu(\Omega).
\end{aligned}$$

Thus, $\int_\Omega \phi_1\left(\frac{|f|}{a}\right) \leq K + \phi_1(t_0)\mu(\Omega) =: c$, hence $f \in L^{\phi_1}(\mu)$. Convexity and $\phi_1(0) = 0$ implies that $\phi_1(c^{-1}t) \leq c^{-1}\phi_1(t)$, as $c > 1$ and hence,

$$\int_\Omega \phi_1\left(\frac{|f|}{ac}\right)d\mu \leq c^{-1}\int_\Omega \phi_1\left(\frac{|f|}{a}\right)d\mu \leq 1,$$

so that $ac \geq ||f||_{\phi_1,\mu}$ whenever $a > ||f||_{\phi_2,\mu}$, and this shows the claim. □

The following lemma is a straightforward consequence of the definitions and we omit the proof.

Lemma 5.3. *Let (Ω, μ) be a finite measure space, let $\phi : \mathbb{R} \to \mathbb{R}$ be a Young function, and let $\tilde{\phi}(t) := \phi(\lambda t)$ for some constant $\lambda > 0$.*

Then $\tilde{\phi}$ is also a Young function. Moreover, $L^\phi(\mu) = L^{\tilde{\phi}}(\mu)$ and $||\cdot||_{\tilde{\phi},\mu} = \lambda||\cdot||_{\phi,\mu}$, so that these norms are equivalent.

Furthermore, we investigate how the Orlicz spaces relate when changing the measure μ to an equivalent measure $\mu' \in \mathcal{M}(\Omega, \mu)$.

Proposition 5.4. *Let $0 \neq \mu' \in \mathcal{M}(\Omega, \mu)$ be a measure such that $d\mu'/d\mu \in L^p(\Omega, \mu)$ for some $p > 1$, and let $q > 1$ be the dual index, i.e., $p^{-1} + q^{-1} = 1$. Then for any Young function ϕ we have*

$$L^{\phi^q}(\mu) \subset L^\phi(\mu'),$$

and this embedding is continuous.

Proof. Let $h := d\mu'/d\mu \in L^p(\Omega, \mu)$ and $c := ||h||_p > 0$. If $f \in L^{\phi^q}(\mu)$ and $a > ||f||_{\phi^q,\mu}$, then by Hölder's inequality we have

$$\int_\Omega \phi\left(\frac{|f|}{a}\right)\, d\mu' = \int_\Omega \phi\left(\frac{|f|}{a}\right) h\, d\mu \leq c\left|\left|\phi\left(\frac{|f|}{a}\right)\right|\right|_q = c\underbrace{\left|\left|\phi^q\left(\frac{|f|}{a}\right)\right|\right|_1^{1/q}}_{\leq 1} \leq c.$$

Thus, $f \in L^\phi(\mu')$, and $a \geq ||f||_{c^{-1}\phi,\mu'}$ whenever $a > ||f||_{\phi^q,\mu}$, hence $||f||_{\phi^q,\mu} \geq ||f||_{c^{-1}\phi,\mu'}$. This shows the claim as $||\cdot||_{c^{-1}\phi,\mu'}$ and $||\cdot||_{\phi,\mu'}$ are equivalent norms on $L^\phi(\mu')$ by Proposition 5.2. □

5.2. **Exponential tangent spaces.** For an arbitrary $\mu \in \mathcal{M}_+(\Omega, \mu_0)$, we define the set

$$\hat{B}_\mu(\Omega) := \{f : \Omega \to [-\infty, +\infty] \;:\; e^f \in L^1(\Omega, \mu)\},$$

which by Hölder's inequality is a convex cone inside the space of measurable functions $\Omega \to [-\infty, +\infty]$. For μ_0, there is a bijection

$$\log_{\mu_0} : \mathcal{M}_+(\Omega, \mu_0) \to \hat{B}_{\mu_0}(\Omega), \qquad \phi\,\mu_0 \mapsto \log(\phi),$$

and for $\mu_0' \in \mathcal{M}_+(\Omega, \mu_0)$ we have $\log_{\mu_0'} = \log_{\mu_0} - u$ where $u := \log_{\mu_0'}(\mu_0')$. That is, $\log_{\mu_0}$ canonically identifies $\mathcal{M}_+(\Omega, \mu_0)$ with a convex set. Moreover, we let

$$\begin{aligned} B_\mu(\Omega) &:= \hat{B}_\mu(\Omega) \cap (-\hat{B}_\mu(\Omega)) \\ &= \{f : \Omega \to [-\infty, \infty] \mid e^{\pm f} \in L^1(\Omega, \mu)\} \\ &= \{f : \Omega \to [-\infty, \infty] \mid e^{|f|} \in L^1(\Omega, \mu)\} \end{aligned}$$

and

$$B_\mu^0(\Omega) := \{f \in B_\mu(\Omega) \mid (1+s)f \in B_\mu(\Omega) \text{ for some } s > 0\}.$$

The points of $B_\mu^0(\Omega)$ are called *inner points* of $B_\mu(\Omega)$.

Note that for $\mu \in \mathcal{M}_+(\Omega, \mu_0)$ we have $B_\mu(\Omega) \subset B_{\mu_0}(\Omega)$.

Definition 5.5. Let $\mu \in \mathcal{M}_+(\Omega, \mu_0)$. Then

$$T_\mu \mathcal{M}_+(\Omega, \mu_0) := \{f : \Omega \to [-\infty, \infty] \mid \ tf \in B_\mu(\Omega) \text{ for some } t \neq 0\}$$

is called the *exponential tangent space of $\mathcal{M}_+(\Omega, \mu_0)$ at μ.*

Evidently, this space coincides with the Orlicz space $T_\mu \mathcal{M}_+(\Omega, \mu_0) = L^{\cosh t - 1}(\mu)$ and hence has a Banach norm. Moreover, $B_\mu(\Omega) \subset T_\mu \mathcal{M}_+(\Omega, \mu_0)$ contains the unit ball w.r.t. the Orlicz norm and hence is a neighborhood of the origin. Furthermore, $\lim_{t\to\infty} t^p/(\cosh t - 1) = 0$ for all $p \geq 1$, so that Proposition 5.2 implies that

$$L^\infty(\Omega, \mu_0) \ \subset\ T_\mu \mathcal{M}_+(\Omega, \mu_0) \ \subset\ \bigcap_{p \geq 1} L^p(\Omega, \mu), \tag{5.1}$$

where all inclusions are continuous.

Remark 5.6. In [16, Definition 6], $T_\mu \mathcal{M}_+(\Omega, \mu_0)$ is called the *Cramer class* of μ. Moreover, in [16, Proposition 7] and [28, Definition 2.2], the subspace of *centered Cramer class* is defined as the functions $u \in T_\mu \mathcal{M}_+(\Omega, \mu_0)$ with $\int_\Omega u \, d\mu = 0$. Thus, the space of centered Cramer classes is a closed subspace of codimension one.

In order to understand the topological structure of $\mathcal{M}_+(\Omega, \mu_0)$ with respect to the e-topology, it is useful to introduce the following preorder on $\mathcal{M}_+(\Omega, \mu_0)$:

$$\mu' \preceq \mu \quad \text{if and only if} \quad \mu' = \phi\mu \text{ with } \phi \in L^p(\Omega, \mu) \text{ for some } p > 1. \tag{5.2}$$

In order to see that $\preceq$ is indeed a preorder, we have to show transitivity, as the reflexivity of $\preceq$ is obvious. Thus, let $\mu'' \preceq \mu'$ and $\mu' \preceq \mu$, so that $\mu' = \phi\mu$ and $\mu'' = \psi\mu'$ with $\phi \in L^p(\Omega, \mu)$ and $\psi \in L^{p'}(\Omega, \mu')$, then $\phi^p, \psi^{p'}\phi \in L^1(\Omega, \mu)$ for some $p, p' > 1$. Let $\lambda := (p'-1)/(p+p'-1) \in (0,1)$. Then by Hölder's inequality, we have:

$$L^1(\Omega, \mu_1) \ni (\psi^{p'}\phi)^{1-\lambda}(\phi^p)^\lambda = \psi^{p'(1-\lambda)}\phi^{1+\lambda(p-1)} = (\psi\phi)^{p''},$$

where $p'' = pp'/(p+p'-1) > 1$, so that $\psi\phi \in L^{p''}(\Omega, \mu)$, and hence, $\mu'' \preceq \mu$ as $\mu'' = \psi\phi\mu$.

From the preorder $\preceq$ we define the equivalence relation on $\mathcal{M}_+(\Omega, \mu_0)$ by

$$\mu' \sim \mu \qquad \text{if and only if } \mu' \preceq \mu \text{ and } \mu \preceq \mu', \tag{5.3}$$

in which case we call μ and μ' *similar*, and hence we obtain a partial ordering on the set of equivalence classes $\mathcal{M}_+(\Omega, \mu_0)/_\sim$

$$[\mu'] \preceq [\mu] \qquad \text{if and only if} \qquad \mu' \preceq \mu.$$

If $\mu' \preceq \mu$, then $T_\mu \mathcal{M}_+(\Omega, \mu_0) \subset T_{\mu'} \mathcal{M}_+(\Omega, \mu_0)$ is continuously embedded. Namely, $\lim_{t\to\infty}(\cosh t - 1)^q/(\cosh(qt) - 1) = 2^{1-q}$, and then we apply Propositions 5.2 and 5.4 as well as Lemma 5.3.

In particular, if $\mu \sim \mu'$ then $T_\mu\mathcal{M}_+(\Omega,\mu_0) = T_{\mu'}\mathcal{M}_+(\Omega,\mu_0)$, and this space we denote by $T_{[\mu]}\mathcal{M}_+(\Omega,\mu_0)$. This space is therefore equipped with a family of equivalent Banach norms, and we have continuous inclusions

$$T_{[\mu']}\mathcal{M}_+(\Omega,\mu_0) \supset T_{[\mu]}\mathcal{M}_+(\Omega,\mu_0) \qquad \text{if} \qquad [\mu'] \preceq [\mu]. \tag{5.4}$$

Remark 5.7. In general, the subspace in (5.4) will be neither closed nor dense. Indeed, it is not hard to show that $f \in T_{[\mu']}\mathcal{M}_+(\Omega,\mu_0)$ lies in the closure of $T_{[\mu]}\mathcal{M}_+(\Omega,\mu_0)$ if and only if

$$(|f| + \epsilon \log(d\mu'/d\mu))_+ \in T_{[\mu]}\mathcal{M}_+(\Omega,\mu_0) \qquad \text{for all } \epsilon > 0.$$

The following now is a reformulation of Propositions 3.4 and 3.5 in [28].

Proposition 5.8. *A sequence $(g_n)_{n\in\mathbb{N}} \in \mathcal{M}(\Omega,\mu_0)$ is e-convergent to $g \in \mathcal{M}(\Omega,\mu_0)$ if and only if $g_n\mu_0 \sim g\mu_0$ for large n, and $u_n := \log|g_n| \in T_{g\mu_0}\mathcal{M}_+(\Omega,\mu_0)$ converges to $u_0 := \log|g| \in T_{g\mu_0}\mathcal{M}_+(\Omega,\mu_0)$ in the Banach norm on $T_{g\mu_0}\mathcal{M}_+(\Omega,\mu_0)$ described above.*

By virtue of this proposition, we shall refer to the topology on $T_\mu\mathcal{M}_+(\Omega,\mu_0)$ obtained above as the *topology of e-convergence* or the *e-topology*. Our description allows us to describe in a different way the Banach manifold structure on $\mathcal{M}(\Omega,\mu_0)$ defined in [28].

Theorem 5.9. *Let $K \subset \mathcal{M}_+(\Omega,\mu_0)$ be an equivalence class w.r.t. $\sim$, and let $T := T_{[\mu]}\mathcal{M}_+(\Omega,\mu_0)$ for $\mu \in K$ be the common exponential tangent space, equipped with the e-topology. Then for all $\mu \in K$,*

$$A_\mu := \log_\mu(K) \subset T$$

is open convex. In particular, the identification $\log_\mu : A_\mu \to K$ allows us to canonically identify K with a open convex subset of the affine space associated to T.

Remark 5.10. This theorem shows that the equivalence classes w.r.t. $\sim$ are the connected components of the e-topology on $\mathcal{M}(\Omega,\mu_0)$, and since each such component is canonically identified as a subset of an affine space whose underlying vector space is equipped with a family of equivalent Banach norms, it follows that $\mathcal{M}(\Omega,\mu_0)$ is a Banach manifold. This is the affine Banach manifold structure on $\mathcal{M}(\Omega,\mu_0)$ described in [28], therefore we refer to it as the *Pistone-Sempi structure*.

Proof. (Theorem 5.9) If $f \in A_\mu$, then, by definition, $(1+s)f, -sf \in \hat{B}_\mu(\Omega)$ for some $s > 0$. In particular, $sf \in B_\mu(\Omega)$, so that $f \in T$ and hence, $A_\mu \subset T$. Moreover, if $f \in A_\mu$ then $\lambda f \in A_\mu$ for $\lambda \in [0,1]$.

Next, if $g \in A_\mu$, then $\mu' := e^g\mu \in K$. Therefore, $f \in A_{\mu'}$ if and only if $K \ni e^f\mu' = e^{f+g}\mu$ if and only if $f + g \in A_\mu$, so that $A_{\mu'} = g + A_\mu$ for a fixed $g \in T$. From this, the convexity of A follows.

Therefore, in order to show that $A_\mu \subset T$ is open, it suffices to show that $0 \in A_{\mu'}$ is an inner point for all $\mu' \in K$. For this, observe that for $f \in B^0_{\mu'}(\Omega)$ we have $(1+s)f \in B_{\mu'}(\Omega)$ and hence $e^{\pm(1+s)f} \in L^1(\Omega,\mu')$, so that $e^f \in L^{1+s}(\Omega,\mu')$ and $e^{-f} \in L^{1+s}(\Omega,\mu') \subset L^s(\Omega,\mu')$, whence $e^f\mu' \sim \mu' \sim \mu$, so that $e^f\mu' \in K$ and hence, $f \in A_{\mu'}$. Thus, $0 \in B^0_{\mu'}(\Omega) \subset A_{\mu'}$, and since $B^0_{\mu'}(\Omega)$ contains the unit ball of the Orlicz norm, the claim follows. □

In the terminology which we developped, we can formulate the significance of the Pistone-Sempi structure on $\mathcal{M}_+(\Omega,\mu_0)$ as follows.

Proposition 5.11. *The quadruple $(\mathcal{M}_+(\Omega,\mu),\Omega,\mu,i_{can})$ is a k-integrable statistical model for all $k \geq 1$.*

Proof. Note that for $x \in \mathcal{M}_+(\Omega,\mu)$ we have $\ln \bar{p}(x,\omega) = \ln x(\omega)$. Using this and the definition of the Pistone-Sempi manifold, we conclude that the first condition in Definition 2.3 holds for the Pistone-Sempi manifold. The second condition in Definition 2.3 also holds for the Pistone-Sempi manifolds, since by (5.1) the inclusion $T_\mu\mathcal{M}_+(\Omega,\mu) \to L^k(\Omega,\mu)$ is continuous for all $k \geq 1$. □

The following example shows that the notion of a k-integrable parametrized measure model is more general than the corresponding notion within the theory of Pistone and Sempi.

Example 5.12. Let $\Omega := (0,1)$, and consider the 1-parameter family of finite measures

$$p(x) \;:=\; \bar{p}(x,t)\,dt \;:=\; \exp\left(-\frac{x^2}{t^{\frac{1}{k}}}\right) dt \;\in\; \mathcal{M}_+((0,1),dt), \qquad x \in \mathbb{R}.$$

This family defines a $(k-1)$-integrable parametrized measure model: Consider the map

$$\ln \bar{p}(\cdot,t) :\; x \mapsto \; -\frac{x^2}{t^{\frac{1}{k}}}.$$

It is continuously differentiable for all $t \in (0,1)$ and therefore satisfies condition (1) of Definition 2.3. Now we come to condition (2): With a continuous vector field $V : \mathbb{R} \to \mathbb{R}$, we have

$$\partial_V \ln \bar{p}(x,t) \;=\; V(x)\,\frac{\partial}{\partial x}\,\ln \bar{p}(x,t) \;=\; V(x)\,\frac{\partial}{\partial x}\left(-\frac{x^2}{t^{\frac{1}{k}}}\right) \;=\; -V(x)\,\frac{2\,x}{t^{\frac{1}{k}}}.$$

We now show that the function $t \mapsto \partial_V \ln \bar{p}(x,t)$ belongs to $L^j((0,1),p(x))$ for all $j \leq k-1$:

$$\begin{aligned}
I^{(j)}(x) \;&:=\; \|\partial_V \ln \bar{p}(x,t)\|^j_{L^j((0,1),p(x))} \\
&=\; \int_0^1 \left(\frac{2\,|x\,V(x)|}{t^{\frac{1}{k}}}\right)^j \exp\left(-\frac{x^2}{t^{\frac{1}{k}}}\right) dt \\
&\leq\; \frac{\big(2\,|x\,V(x)|\big)^j}{e} \int_0^1 \frac{1}{t^{\frac{j}{k}}}\,dt \\
&=\; \frac{\big(2\,|x\,V(x)|\big)^j}{e}\,\frac{k}{k-j} \\
&<\; \infty.
\end{aligned}$$

Finally, we now have to show that the function $x \mapsto I^{(j)}(x)$ is continuous. In order to verify the continuity in a point $x_0 \in \mathbb{R}$ it is sufficient to consider the restriction of $I^{(j)}$ to the closed interval $[x_0 - \varepsilon, x_0 + \varepsilon]$ with some positive number ε. On this interval, the corresponding integrand is upper bounded by a function that only depends on t and is integrable:

$$\left(\frac{2\,|x\,V(x)|}{t^{\frac{1}{k}}}\right)^j \exp\left(-\frac{x^2}{t^{\frac{1}{k}}}\right) \;\leq\; \frac{c}{t^{\frac{j}{k}}}, \qquad c \geq 0.$$

Therefore, by the continuity lemma for integrals, $I^{(j)}$ is continuous, which completes the proof that our family is $(k-1)$-integrable parametrized measure model. However, it does not define a model in the sense of Pistone and Sempi. In order to see this we show that for all $x \neq 0$, $p(x)$ and $p(0)$ are not similar: Obviously,

$$dt \;=\; \exp\left(\frac{x^2}{t^{\frac{1}{k}}}\right) dp(x).$$

The similarity of $dp(x)$ and dt would imply that $\frac{dt}{dp(x)}$ is in $L^{1+s}((0,1),dp(x))$ for some $s > 0$ (see 5.2 and 5.3). However, for all $s > 0$, we have

$$\begin{aligned}
\int_0^1 \left(\exp\left(\frac{x^2}{t^{\frac{1}{k}}}\right)\right)^{1+s} dp(x) \;&=\; \int_0^1 \exp\left(\frac{s\,x^2}{t^{\frac{1}{k}}}\right) dt \\
&\geq\; \int_0^1 \frac{1}{k!}\left(\frac{s\,x^2}{t^{\frac{1}{k}}}\right)^k dt \\
&=\; \infty.
\end{aligned}$$

Thus, $p(x)$ and dt are in different e-connected components of $\mathcal{M}_+((0,1),dt)$ and, therefore, the map p cannot be continuous with respect to the e-topology. Hence, the parametrized measure model cannot be considered as a submanifold of $\mathcal{M}_+((0,1),dt)$ in the sense of Pistone and Sempi.

We end this section with the following result which illustrates how the ordering $\preceq$ provides a stratification of $\hat{B}_{\mu_0}(\Omega)$.

Proposition 5.13. *Let $\mu_0', \mu_1' \in \mathcal{M}_+(\Omega,\mu_0)$ with $f_i := \log_{\mu_0}(\mu_i') \in \hat{B}_{\mu_0}(\Omega)$, and let $\mu_\lambda' := \exp(f_0 + \lambda(f_1 - f_0))\mu_0$ for $\lambda \in [0,1]$ be the segment joining μ_0' and μ_1'. Then the following hold.*

(1) *The measures μ_λ' are similar for $\lambda \in (0,1)$.*
(2) *$\mu_\lambda' \preceq \mu_0'$ and $\mu_\lambda' \preceq \mu_1'$ for $\lambda \in (0,1)$.*
(3) *$T_{\mu_\lambda'}\mathcal{M}_+(\Omega,\mu_0) = T_{\mu_0'}\mathcal{M}_+(\Omega,\mu_0) + T_{\mu_1'}\mathcal{M}_+(\Omega,\mu_0)$ for $\lambda \in (0,1)$.*

Proof. Let $\delta := f_1 - f_0$ and $\phi := \exp(\delta)$. Then for all $\lambda_1, \lambda_2 \in [0,1]$, we have

$$\mu_{\lambda_1}' = \phi^{\lambda_1 - \lambda_2}\mu_{\lambda_2}'. \tag{5.5}$$

For $\lambda_1 \in (0,1)$ and $\lambda_2 \in [0,1]$, we pick $p > 1$ such that $\lambda_2 + p(\lambda_1 - \lambda_2) \in (0,1)$. Then by (5.5) we have

$$\phi^{p(\lambda_1-\lambda_2)}\mu_{\lambda_2}' = \mu_{\lambda_2+p(\lambda_1-\lambda_2)}' \in \mathcal{M}_+(\Omega,\mu_0),$$

so that $\phi^{p(\lambda_1-\lambda_2)} \in L^1(\Omega,\mu_{\lambda_2}')$ or $\phi^{\lambda_1-\lambda_2} \in L^p(\Omega,\mu_{\lambda_2})$ for small $p-1>0$. Therefore, $\mu_{\lambda_1}' \preceq \mu_{\lambda_2}'$ for all $\lambda_1 \in (0,1)$ and $\lambda_2 \in [0,1]$, which implies the first and second statement.

This implies that $T_{\mu_i'}\mathcal{M}_+(\Omega,\mu_0) \subset T_{\mu_\lambda'}\mathcal{M}_+(\Omega,\mu_0) = T_{\mu_{1/2}'}\mathcal{M}_+(\Omega,\mu_0)$ for $i = 0,1$ and all $\lambda \in (0,1)$ which shows one inclusion in the third statement.

In order to complete the proof, observe that

$$T_{\mu_{1/2}'}\mathcal{M}_+(\Omega,\mu_0) = T_{\mu_{1/2}'}\mathcal{M}_+(\Omega_+,\mu_0) \oplus T_{\mu_{1/2}'}\mathcal{M}_+(\Omega_-,\mu_0),$$

where $\Omega_+ := \{\omega \in \Omega \mid \delta(\omega) > 0\}$ and $\Omega_- := \{\omega \in \Omega \mid \delta(\omega) \le 0\}$. If $g \in T_{\mu_{1/2}'}\mathcal{M}_+(\Omega_+,\mu_0)$, then for some $t \neq 0$

$$\begin{aligned}\int_\Omega \exp(|tg|)d\mu_0' &\le \int_{\Omega_+} \exp(|tg| + \frac{1}{2}\delta)d\mu_0' + \int_{\Omega_-} d\mu_0' \\ &= \int_{\Omega_+} \exp(|tg|)d\mu_{1/2}' + \int_{\Omega_-} d\mu_0' < \infty,\end{aligned}$$

so that $g \in T_{\mu_0'}(\Omega,\mu_0)$ and hence, $T_{\mu_{1/2}'}\mathcal{M}_+(\Omega_+,\mu_0) \subset T_{\mu_0'}(\Omega,\mu_0)$. Analogously, one shows that $T_{\mu_{1/2}'}\mathcal{M}_+(\Omega_-,\mu_0) \subset T_{\mu_0'}(\Omega,\mu_1)$ which completes the proof. □

Acknowledgements

H.V.L. would like to thank Shun-ichi Amari for many fruitful discussions, and Giovanni Pistone for providing the articles [10, 16]. We thank Holger Bernigau for his critical helpful comments on an early version of this paper. This work has been supported by the Max Planck Institute for Mathematics in the Sciences in Leipzig, the BSI at RIKEN in Tokyo, the ASSMS, GCU in Lahore-Pakistan, the VNU for Sciences in Hanoi, the Mathematical Institute of the Academy of Sciences of the Czech Republic in Prague, and the Santa Fe Institute. We are grateful for excellent working conditions and financial support of these institutions during extended visits of some of us.

References

[1] S. Amari, Theory of information spaces a geometrical foundation of statistics. POST RAAG Report 106, 1980.

[2] S. Amari, Differential geometry of curved exponential families curvature and information loss. The Annals of Statistics, 10(1982),357-385.

[3] S. Amari, Differential Geometrical Theory of Statistics, in: Differential geometry in statistical inference, Institute of Mathematical Statistics, Lecture Note-Monograph Series, Volume 10, California (1987).

[4] S. Amari, H. Nagaoka, Methods of information geometry, Translations of mathematical monographs; v. 191, American Mathematical Society, 2000.

[5] N. Ay, An Information-geometric approach to a theory of pragmatic structuring, Ann. of Prob., v. 30, N. 1 (2002), 416-436.

[6] N. Ay, E. Olbrich, N. Bertschinger, J. Jost, A geometric approach to complexity, Chaos 21, (2011), 037103.

[7] D. Blackwell, Equivalent comparisons of experiments, Ann. Math. Stat. 24(1953), 265-272.

[8] A. A. Borovkov, Mathematical statistics, Gordon and Breach Science Publishers, 1998.

[9] L. L. Campbell, An extended Chentsov characterization of a Riemannian metric, Proceedings of the AMS, 98, (1986), 135-141.

[10] A. Cena and G. Pistone, Exponential statistical model, AISM 59 (2007), 27-56.

[11] H. Cramer, Mathematical Methods of Statistics, Princeton Univ. Press, Princeton, 1946.

[12] N. Chentsov, Category of mathematical statistics, Dokl. Acad. Nauk USSR 164 (1965), 511-514.

[13] N. Chentsov, Algebraic foundation of mathematical statistics, Math. Operationsforsch. statist. Serie Statistics. v.9 (1978), 267-276.

[14] N. Chentsov, Statistical decision rules and optimal inference, Translation of Math. Monograph 53, AMS, Providence, 1982.

[15] R. A. Fisher, On the mathematical foundations of theoretical statistics, Philosophical Transactions of the Royal Society of London. Series A 222(1922), 309-368.

[16] P. Gibilisco and G. Pistone, Connections on non-parametric statistical models by Orlicz space geometry, Infinite Dimensional Analysis, Quantum Probability and Related Topics, 1(2) (1998), 325-347.

[17] J. Jost, Information geometry, preprint, 2010.

[18] O. Kallenberg, Foundations of Modern Probability, 2nd edition. Springer-Verlag, New York, Berlin, Heidelberg, 2001.

[19] M.A. Krasnoselskii and Ya.B. Rutickii, Convex functions and Orlicz spaces. Fizmatgiz, Moskava (1958). [In Russian; English translation: P. Noordfoff Ltd., Groningen (1961)]

[20] S. Kullback, Information Theory and Statistics, Dover Publications, 1968.

[21] S. Lauritzen, Statistical models, in: Differential geometry in statistical inference, Institute of Mathematical Statistics, Lecture Note-Monograph Series, Volume 10, California (1987).

[22] H.V. Lê, Statistical manifolds are statistical models, J. Geometry 84 (2005), 83-93.

[23] H.V. Lê, Monotone invariants and embedding of statistical models, Advances in Deterministic and Stochastic Analysis, World Scientific 2007, 231-254, arXiv:math/0506163.

[24] N. Morse and R. Sacksteder, Statistical isomorphism, Annals of Math. Statistics, 37 (1966), 203-214.

[25] E. Morozova and N. Chentsov, Markov invariant geometry on manifolds of states, Itogi Nauki i Techniki, Current problems of mathematics, Newest achievements 6 (1990), Moscow, 69-102.

[26] E. Morozova and N. Chentsov, Natural geometry on families of probability laws, Itogi Nauki i Techniki, Current problems of mathematics, Fundamental directions 83 (1991), Moscow, 133-265.

[27] E. Olbrich, T. Kahle, N. Bertschinger, N. Ay, J. Jost, Quantifying structure in networks, The European Physical Journal B - Condensed Matter and Complex Systems, 77 (2010) 239-247.

[28] G. Pistone and C. Sempi, An infinite-dimensional structure on the space of all the probability measures equivalent to a given one,The Annals of Statistics (5)(1995), 1543-1561.

[29] C. R. Rao, Information and the accuracy attainable in the estimation of statistical parameters, Bulletin of the Calcutta Mathematical Society 37(1945), 81-89.

[30] S. Shahshahani, A new mathematical framework for the study of linkage and selection, Mem. Amer. Math. Soc. 17 No. 211 (1979).

[1]Max-Planck-Institut für Mathematik in den Naturwissenschaften, Inselstrasse 22, 04103 Leipzig, Germany

[2]Institute of Mathematics of ASCR, Zitna 25, 11567 Praha 1, Czech Republic

[3]Fakultät für Mathematik, Technische Universität Dortmund, Vogelpothsweg 87, 44221 Dortmund, Germany

[4]Mathematisches Institut, Universität Leipzig, 04081 Leipzig, Germany
[5]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA