Parametrized Measure Models

Nihat   Ay
Jürgen   Jost
Hông Vân   Lê
Lorenz   Schwachhöfer

SANTA FE INSTITUTE

Parametrized measure models

Nihat Ay, Jürgen Jost, Hông Vân Lê, Lorenz Schwachhöfer

October 25, 2015

Abstract

We develope a new and general notion of parametric measure models and statistical models on an arbitrary sample space Ω. This is given by a diffferentiable map from the parameter manifold M into the set of finite measures or probability measures on Ω, respectively, which is differentiable when regarded as a map into the Banach space of all signed measures on Ω. Furthermore, we also give a rigorous definition of roots of measures and give a natural definition of the Fisher metric and the Amari-Chentsov tensor as the pullback of tensors defined on the space of roots of measures. We show that many features such as the preservation of this tensor under sufficient statistics and the monotonicity formula hold even in this very general set-up.

1 Introduction

Information geometry is concerned with the use of differential geometric methods in probability theory. An important object of investigation are families of probability measures or, more generally, of finite measures on a given sample space Ω which depend differentiably on a finite number of parameters. Associated to such a family there are two symmetric tensors on the parameter space M. The first is a quadratic form (i.e., a Riemannian metric), called the *Fisher metric* $\mathfrak{g}^F$, and the second is a 3-tensor, called the *Amari-Chentsov tensor* $\mathbf{T}$. The Fisher metric was first suggested by Rao [19], followed by Jeffreys [12], Efron [11] and then systematically developed by Chentsov and Morozova [8], [9] and [16]; the Amari-Chentsov tensor and its significance was discovered by Amari [1], [2] and Chentsov [10].

These tensors are of interest from the differential geometric point of view as they do not depend on the particular choice of parametrization of the family, but they are also natural objects from the point of view of statistics, as they are unchanged under sufficient statistics and are in fact characterized by this property; this was shown in the case of finite sample spaces by Chentsov in [9] and more recently for general sample spaces in [5]. In fact, Chentsov not only showed the invariance of these tensors under sufficient statistics, but also under what he called *congruent embeddings* of probability measures. These are Markov kernels between finite sample spaces which are right inverses of a statistic. We use this property to give a precise definition of congruent embeddings between arbitrary sample spaces (cf. Definition 3.1). As it turns out, every Markov kernel induces a congruent embedding in this sense, but there are congruent embeddings which are *not* induced by Markov kernels, cf. Theorem 3.1.

The main conceptual difficulty in the investigation of families of probability measures is the lack of a canonical manifold structure on the spaces $\mathcal{M}(\Omega)$ and $\mathcal{P}(\Omega)$ of finite measures and probability measures on Ω. If Ω is finite, then this issue is not a problem, as in this case a measure is given by finitely many parameters, allowing to identify $\mathcal{M}(\Omega)$ with the positive orthant in $\mathbb{R}^{|\Omega|}$ and $\mathcal{P}(\Omega)$ with the intersection of this orthant with an affine hyperplane in $\mathbb{R}^{|\Omega|}$, so that both are (finite dimensional) manifolds in a canonical way. But this is no longer true if Ω is infinite.

Attempts have been made to provide $\mathcal{P}(\Omega)$ and $\mathcal{M}(\Omega)$ with a Banach manifold structure. For instance, Pistone and Sempi [18] equipped these spaces with a topology, the so-called *e-topology*. With this, $\mathcal{P}(\Omega)$ and $\mathcal{M}(\Omega)$ become Banach manifolds and have many remarkable features, see e.g. [7], [17]. On the other hand, the e-topology is very strong in the sense that many families of measures on Ω fail to be continuous w.r.t. the e-topology, so it cannot be applied as widely as one would wish.

Another approach was recently pursued by Bauer, Bruveris and Michor [6] under the assumption that Ω is a manifold. In this case, the space of smooth densities also carries a natural topology, and they were able to show that the invariance under diffeomorphisms already suffices to characterize the Fisher metric of a family of such densities.

In [5], the authors of the present article proposed to define parametrized measure models as a family given as

$$p(\xi) = p(\omega; \xi)\mu. \tag{1.1}$$

for some reference measure μ and a positive function p on $\Omega \times M$ which is differentiable in $\xi \in M$, an idea which closely follows the notion of Amari [1]. While this embraces many interesting families of measures, it is still restricted as it heavily depends on the choice of the reference measure μ which is a priori not naturally given.

It is the aim of the present article to provide a yet more general definition of parametrized measure models which embraces all of the aforementioned definitions, but is more general and more natural than these at the same time. Namely, in this article we define *parametrized measure models* and *statistical models*, respectively, as families $(p(\xi))_{\xi \in M}$ which are given by a map p from M to $\mathcal{M}(\Omega)$ and $\mathcal{P}(\Omega)$, respectively, which is differentiable when regarded as a map between the (finite or infinite dimensional) manifold M and the Banach space $\mathcal{S}(\Omega)$ of finite signed measures on Ω, since evidently $\mathcal{P}(\Omega)$ and $\mathcal{M}(\Omega)$ are subsets of $\mathcal{S}(\Omega)$. That is, the geometric structure on $\mathcal{M}(\Omega)$ and $\mathcal{P}(\Omega)$ is given by the inclusions $\mathcal{P}(\Omega) \hookrightarrow \mathcal{M}(\Omega) \hookrightarrow \mathcal{S}(\Omega)$.

If the model is given as in the definition from [5] by (1.1), then we say that it is given by a *regular density function*. We shall show that most of the statements shown in [5] for parametrized measure models or statistical models with a regular density function also hold without this assumption.

As it turns out, neither the Fisher metric nor the Amari-Chentsov tensor may be regarded as pull-backs of a tensor on $\mathcal{S}(\Omega)$ via the map p. In order to overcome this problem, let us for the moment assume that the family is given by a regular density function as in (1.1). In this case, the definitions of $\mathfrak{g}^F$ and $\mathbf{T}^{AC}$ can be rewritten as

$$
\begin{aligned}
\mathfrak{g}^F(V, W) &= \int_\Omega \partial_V \log p(\omega; \xi)\, \partial_W \log p(\omega; \xi)\, dp(\xi), \\
&= 4 \int_\Omega \partial_V \sqrt{p(\omega; \xi)}\, \partial_W \sqrt{p(\omega; \xi)}\, d\mu \\
&= 4 \int_\Omega d\left(\partial_V \sqrt{p(\xi)}\, \partial_W \sqrt{p(\xi)} \right)
\end{aligned}
\tag{1.2}
$$

and

$$
\begin{aligned}
\mathbf{T}^{AC}(V, W, U) &= \int_\Omega \partial_V \log p(\omega; \xi)\, \partial_W \log p(\omega; \xi)\, \partial_U \log p(\omega; \xi)\, dp(\xi) \\
&= 27 \int_\Omega \partial_V \sqrt[3]{p(\omega; \xi)}\, \partial_W \sqrt[3]{p(\omega; \xi)}\, \partial_U \sqrt[3]{p(\omega; \xi)}\, d\mu \qquad (1.3) \\
&= 27 \int_\Omega d\left(\partial_V \sqrt[3]{p(\xi)}\, \partial_W \sqrt[3]{p(\xi)}\, \partial_U \sqrt[3]{p(\xi)} \right).
\end{aligned}
$$

Of course, the last lines in (1.2) and (1.3) do not make sense a priori, as it is not clear what a root of a measure should be, but this is precisely the way we shall make use of this. Namely, for $r \in (0, 1]$ we define the space $\mathcal{S}^r(\Omega)$ of *r-th powers of a measure*, which has $\mathcal{M}^r(\Omega)$ and $\mathcal{P}^r(\Omega)$ as subsets. Again, $\mathcal{S}^r(\Omega)$ is a Banach space and $\mathcal{S}^1(\Omega) = \mathcal{S}(\Omega)$. Furthermore, there is a bounded bilinear multiplication map

$$
\cdot : \mathcal{S}^r(\Omega) \times \mathcal{S}^s(\Omega) \longrightarrow \mathcal{S}^{r+s}(\Omega),
$$

for $r, s, r + s \in (0, 1]$ as well as differentiable maps taking the (signed) k-th power,

$$
\pi^k, \hat{\pi}^k : \mathcal{S}^r(\Omega) \longrightarrow \mathcal{S}^{kr}(\Omega)
$$

for $r, kr \in (0, 1]$. That is, one can work with these objects in a very suggestive way.
We call a parametrized measure model $p : M \to \mathcal{M}(\Omega)$ *k-integrable*, if the map $p^{1/k} : M \to \mathcal{M}^{1/k}(\Omega) \subset \mathcal{S}^{1/k}(\Omega)$ is (formally) differentiable. This notion of k-integrability turns out to be equivalent to that given in [5] for models with a regular density function. On $\mathcal{S}^{1/n}(\Omega)$ we define the *canonical n-tensor* as

$$
L_\Omega^n(\nu_1, \ldots, \nu_n) := n^n \int_\Omega d(\nu_1 \cdots \nu_n).
$$

Then by (1.2) and (1.3) the Fisher metric and the Amari-Chentsov tensor are given as the pull-backs

$$
\mathfrak{g}^F = (p^{1/2})^*(L_\Omega^2) \qquad \text{and} \qquad \mathbf{T}^{AC} = (p^{1/3})^*(L_\Omega^3),
$$

if the model is k-integrable for $k \geq 2$ or $k \geq 3$, respectively. That is, these tensors may be defined as pullbacks of objects which are naturally defined in terms of Ω only.
We also discuss the behaviour of the Fisher metric under statistics. For this, let $\kappa : \Omega \to \Omega'$ be measurable and $\kappa_* : \mathcal{S}(\Omega) \to \mathcal{S}(\Omega')$ be the push-forward of (signed) measures. This is a bounded linear map which maps $\mathcal{M}(\Omega)$ to $\mathcal{M}(\Omega')$ and $\mathcal{P}(\Omega)$ to $\mathcal{P}(\Omega')$, respectively. In particular, given a parametrized measure model $p : M \to \mathcal{M}(\Omega)$, it induces a map $p' := \kappa_* p$. We then show that this process preserves k-integrability, i.e., if p is k-integrable, then so is p' (cf. Theorem 5.1). Moreover, we show that in this generality the monotonicity formula holds (Theorem 5.2):

$$
\mathfrak{g}^F(V, V) \geq \mathfrak{g}'^F(V, V), \qquad (1.4)
$$

where $\mathfrak{g}^F$ and $\mathfrak{g}'^F$ denote the Fisher metrics of p and p', respectively, and where $V \in TM$. Moreover, if $p(\xi)$ has a regular density function p, then equality holds for all V iff κ is a sufficient statistic for the model. In fact, we show (1.4) even in the case where κ is replaced by any Markov kernel $K : \Omega \to \mathcal{P}(\Omega')$.
While the monotonicity formula has been known in special cases, e.g. if one of Ω, Ω' is finite, or if both are manifolds and κ is differentiable, our approach shows that (1.4) holds without any further

assumption on the model whatsoever. In particular, it is also valid for rather peculiar statistics as e.g. Example 3.2.

This paper is structured as follows. In Section 2 we give the formal definition of the spaces of powers of measures, setting up the formalism needed later on. In Section 3 we provide a precise definition of congruent embeddings for arbitrary sample spaces Ω and discuss their relations with Markov kernels and the existence of transverse measures. In the following Section 4 we establish the notion of k-integrability, which is applied in the final Section 5 to the discussion of sufficient statistics and the proof of the monotonicity formula.

2 The spaces of measures and their powers

2.1 The space of (signed) finite measures

Let (Ω, Σ) be a measurable space, that is an arbitrary set Ω together with a sigma algebra Σ of subsets of Ω. Regarding the sigma algebra Σ on Ω as fixed, we let

$$
\begin{aligned}
\mathcal{P}(\Omega) &:= \{\mu \ : \ \mu \text{ a probability measure on } \Omega\} \\
\mathcal{M}(\Omega) &:= \{\mu \ : \ \mu \text{ a finite measure on } \Omega\} \\
\mathcal{S}(\Omega) &:= \{\mu \ : \ \mu \text{ a signed finite measure on } \Omega\} \\
\mathcal{S}_0(\Omega) &:= \{\mu \in \mathcal{S}(\Omega) \ : \ \int_\Omega d\mu = 0\}.
\end{aligned}
$$

Clearly, $\mathcal{P}(\Omega) \subset \mathcal{M}(\Omega) \subset \mathcal{S}(\Omega)$, and $\mathcal{S}_0(\Omega), \mathcal{S}(\Omega)$ are real vector spaces. In fact, both $\mathcal{S}_0(\Omega)$ and $\mathcal{S}(\Omega)$ are Banach spaces whose norm is given by the total variation of a signed measure, 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 $A_i \in \Sigma$. By the *Jordan decomposition theorem*, each measure $\mu \in \mathcal{S}(\Omega)$ can be decomposed uniquely as

$$
\mu = \mu_+ - \mu_- \qquad \text{with } \mu_\pm \in \mathcal{M}(\Omega), \ \mu_+ \perp \mu_-. \tag{2.1}
$$

Thus, if we define

$$
|\mu| := \mu_+ + \mu_- \in \mathcal{M}(\Omega),
$$

then (2.1) implies

$$
|\mu(A)| \le |\mu|(A) \qquad \text{for all } \mu \in \mathcal{S}(\Omega) \text{ and } A \in \Sigma, \tag{2.2}
$$

so that

$$
\|\mu\|_{TV} = \| \, |\mu| \, \|_{TV} = |\mu|(\Omega).
$$

In particular,
$$\mathcal{P}(\Omega) = \{\mu \in \mathcal{M}(\Omega) \ : \ \|\mu\|_{TV} = 1\}.$$
Moreover, fixing a measure $\mu_0 \in \mathcal{M}(\Omega)$, we let

$$
\begin{aligned}
\mathcal{P}(\Omega, \mu_0) &:= \{\mu \in \mathcal{P}(\Omega) \ : \ \mu \text{ is dominated by } \mu_0\} \\
\mathcal{M}(\Omega, \mu_0) &:= \{\mu \in \mathcal{M}(\Omega) \ : \ \mu \text{ is dominated by } \mu_0\} \\
\mathcal{S}(\Omega, \mu_0) &:= \{\mu \in \mathcal{S}(\Omega) \ : \ \mu \text{ is dominated by } \mu_0\} \\
\mathcal{S}_0(\Omega, \mu_0) &:= \mathcal{S}(\Omega, \mu_0) \cap \mathcal{S}_0(\Omega),
\end{aligned}
\tag{2.3}
$$

where we say that μ_0 dominates μ if every μ_0-null set is also a $|\mu|$-null set. We may canonically identify $\mathcal{S}(\Omega, \mu_0)$ with $L^1(\Omega, \mu_0)$ by the correspondence

$$\imath_{can} : L^1(\Omega, \mu_0) \longrightarrow \mathcal{S}(\Omega, \mu_0), \qquad \phi \longmapsto \phi \, \mu_0.$$

By the Radon-Nikodým theorem, this is an isomorphism whose inverse is given by the Radon-Nikodým derivative $\mu \mapsto \frac{d\mu}{d\mu_0}$. Observe that $\imath_{can}$ is an isomorphism of Banach spaces, since evidently

$$\|\phi\|_{L^1(\Omega, \mu_0)} = \int_\Omega |\phi| \, d\mu_0 = \|\phi \, \mu_0\|_{TV}.$$

2.2   Differential maps between Banach manifolds and tangent spaces

In this section, we shall recall some basic notions of maps between Banach manifolds. For simplicity, we shall restrict ourselves to maps between open subsets of Banach spaces, even though this notion can be generalized to general Banach manifolds, see e.g. [13].

Let V and W be Banach spaces and $U \subset V$ an open subset. A map $\phi : U \to W$ is called *differentiable at* $x \in U$, if there is a bounded linear operator $d_x\phi \in Lin(V, W)$ such that

$$\lim_{h \to 0} \frac{\|\phi(x+h) - \phi(x) - d_x\phi(h)\|_W}{\|h\|_V} = 0. \tag{2.4}$$

In this case, $d_x\phi$ is called the *(total) differential of ϕ at x*. Moreover, ϕ is called *continuously differentiable* or shortly a *C^1-map*, if it is differentiable at every $x \in U$, and the map $d\phi : U \to Lin(V, W)$, $x \mapsto d_x\phi$ is continuous. Furthermore, a differentiable map $c : (-\varepsilon, \varepsilon) \to W$ is called a *curve in W*.

Definition 2.1. Let $X \subset V$ be an arbitrary subset and let $x_0 \in X$. Then $v \in V$ is called a *tangent vector of X at x_0*, if there is a curve $c : (-\varepsilon, \varepsilon) \to X \subset V$ such that $c(0) = x_0$ and $\dot{c}(0) := d_0c(1) = v$. The set of all tangent vectors at x_0 is called the *tangent space of X at x_0* and is denoted by $T_{x_0}X$. We also let
$$TX := \dot{\bigcup}_{x_0 \in X} T_{x_0}X \subset X \times V \subset V \times V,$$
equipped with the induced topology.

For instance, if $X = U \subset V$ is an open set, then $T_{x_0}U = V$ for all x_0 and hence, $TU = U \times V$. Indeed, the curve $c(t) = x_0 + tv \in U$ for small $|t|$ satisfies the properties required in the definition. While reparametrization of c implies that $T_{x_0}X$ is invariant under scalar multiplication, this set fails to be a linear subspace in general; if $X \subset V$ is a submanifold, however, then $T_{x_0}X$ coincides with the standard notion of the tangent space of a (Banach) manifold, justifying our notation.

If $U \subset V$ is open and $\phi : U \to W$ is a C^1-map whose image is contained in $X \subset W$, then $d_{x_0}\phi(V) \subset T_{\phi(x_0)}X$, whence ϕ induces a continuous map

$$d\phi : TU = U \times V \longrightarrow TX, \qquad (u, v) \longmapsto d_u\phi(v).$$

Proposition 2.1. *Let $V = \mathcal{S}(\Omega)$ be the Banach space of signed measures on Ω. Then the tangent spaces of $\mathcal{M}(\Omega)$ and $\mathcal{P}(\Omega)$ are*

$$T\mathcal{M}(\Omega) := \dot{\bigcup}_{\mu \in \mathcal{M}(\Omega)} \mathcal{S}(\Omega, \mu) \subset \mathcal{M}(\Omega) \times \mathcal{S}(\Omega)$$

and

$$T\mathcal{P}(\Omega) := \dot{\bigcup}_{\mu \in \mathcal{P}(\Omega)} \mathcal{S}_0(\Omega, \mu) \subset \mathcal{P}(\Omega) \times \mathcal{S}(\Omega).$$

Remark 2.1. This result is not obvious, because, in the case of an infinite sample space, $\mathcal{P}(\Omega)$ and $\mathcal{M}(\Omega)$ are not Banach submanifolds of $\mathcal{S}(\Omega)$. Our proof will handle a more general situation, as needed in Proposition 2.2 below.

To prove Proposition 2.1, we first show the following simple

Lemma 2.1. *Let $\{\nu_n : n \in \mathbb{N}\} \subset \mathcal{S}(\Omega)$ be a countable family of measures. Then there is a measure $\mu_0 \in \mathcal{M}(\Omega)$ dominating ν_n for all n.*

Proof. We assume w.l.o.g. that $\nu_n \neq 0$ for all n and define

$$\mu_0 := \sum_{n=1}^{\infty} \frac{1}{2^n \|\nu_n\|_{TV}} |\nu_n|.$$

Since $\|\nu_n\|_{TV} = |\nu_n|(\Omega)$, it follows that this sum converges, so that $\mu_0 \in \mathcal{M}(\Omega)$ is well defined. Moreover, if $\mu_0(A) = 0$, then $|\nu_n|(A) = 0$ for all n, showing that μ_0 dominates all ν_n as claimed. $\square$

Proof of Proposition 2.1. Given $\nu \in T_{\mu_0}\mathcal{M}(\Omega) \subset \mathcal{S}(\Omega)$, let $(\mu_t)_{t \in (-\varepsilon, \varepsilon)}$ be a curve in $\mathcal{M}(\Omega)$ with $\dot{\mu}_0 = \nu$. Pick two sequences $t_n^{\pm} \to 0$, $t_n^+ > 0$, $t_n^- < 0$, and let $\hat{\mu}_0 \in \mathcal{M}(\Omega)$ be a measure which dominates the measures $\mu_{t_n^{\pm}}$, μ_0 and ν. This measure exists by Lemma 2.1. Thus, there are functions $\phi_0, \phi_0', \phi_n^{\pm} \in L^1(\Omega, \hat{\mu}_0)$ such that

$$\mu_{t_n^{\pm}} = \phi_n^{\pm}\hat{\mu}_0, \qquad \mu_0 = \phi_0\hat{\mu}_0, \qquad \nu = \phi_0'\hat{\mu}_0.$$

By hypothesis, $\phi_n^{\pm}, \phi_0 \geq 0$. Adapting the definition of a C^1-map in (2.4), the differential quotient $(t_n^{\pm})^{-1}(\mu_{t_n^{\pm}} - \mu_0) = (t_n^{\pm})^{-1}(\phi_n^{\pm} - \phi_0)\hat{\mu}_0$ converges in $\mathcal{S}(\Omega)$ to $\nu = \phi_0'\hat{\mu}_0$, whence

$$\frac{\phi_n^{\pm} - \phi_0}{t_n^{\pm}} \xrightarrow{n \to \infty} \phi_0'$$

in $L^1(\Omega, \hat{\mu}_0)$. After passing to subsequences this implies pointwise convergence $\hat{\mu}_0$-a.e.

If $\phi_0(\omega) = 0$, then $\frac{\phi_n^+(\omega) - \phi_0(\omega)}{t_n^+} \geq 0$ while $\frac{\phi_n^-(\omega) - \phi_0(\omega)}{t_n^-} \leq 0$. Thus, if both converge to $\phi_0'(\omega)$, then $\phi_0'(\omega) = 0$. That is: on $\phi_0^{-1}(0)$ we have $\phi_0' = 0$ $\hat{\mu}_0$-a.e., showing that $\nu = \phi_0'\hat{\mu}_0$ is dominated by $\mu_0 = \phi_0\hat{\mu}_0$ and hence $\nu \in \mathcal{S}(\Omega, \mu_0)$. Thus, $T_{\mu_0}\mathcal{M}(\Omega) \subset \mathcal{S}(\Omega, \mu_0)$.

Conversely, given $\nu = \phi\mu_0 \in \mathcal{S}(\Omega, \mu_0)$, define $\mu_t := (1 + t\phi)_+\mu_0 \in \mathcal{M}(\Omega, \mu_0)$. Then

$$
\begin{aligned}
\frac{\|\mu_t - \mu_0 - t\nu\|_{TV}}{|t|} &= \frac{\|((1 + t\phi)_+ - 1 - t\phi)\mu_0\|_{TV}}{|t|} \\
&= \frac{\|(1 + t\phi)_-\|_1}{|t|} \\
&\leq \|(|t|^{-1} - |\phi|)_-\|_1 \xrightarrow{t\to 0} 0,
\end{aligned}
$$

using the monotone convergence theorem in the last step. Thus, ν is the derivative of $(\mu_t)_{t\in(-\varepsilon,\varepsilon)}$, showing that $\nu \in T_{\mu_0}\mathcal{M}(\Omega)$ as claimed.

To show the statement for $\mathcal{P}(\Omega)$, let $(\mu_t)_{t\in(-\varepsilon,\varepsilon)}$ be a curve in $\mathcal{P}(\Omega)$ with $\dot\mu_0 = \nu$. Then as μ_t is a probability measure for all t, we conclude

$$
\left|\int_\Omega d\nu\right| = \left|\int_\Omega \frac{1}{t}d(\mu_t - \mu_0 - t\nu)\right| \leq \frac{\|\mu_t - \mu_0 - t\nu\|_{TV}}{|t|} \xrightarrow{t\to 0} 0,
$$

so that $\nu \in \mathcal{S}_0(\Omega)$. Since $\mathcal{P}(\Omega) \subset \mathcal{M}(\Omega)$, it follows that $T_{\mu_0}\mathcal{P}(\Omega) \subset T_{\mu_0}\mathcal{M}(\Omega) \cap \mathcal{S}_0(\Omega) = \mathcal{S}_0(\Omega, \mu_0)$ for all $\mu_0 \in \mathcal{P}(\Omega)$.

Conversely, given $\nu = \phi\mu_0 \in \mathcal{S}_0(\Omega, \mu_0)$, define the curve $\lambda_t := \mu_t\|\mu_t\|_{TV}^{-1} \in \mathcal{P}(\Omega)$ with $\mu_t = (1 + t\phi)_+\mu_0$ as before, so that $\lambda_0 = \mu_0$. We set

$$
c_t^\pm := \int_\Omega (1 + t\phi)_\pm \, d\mu_0 \geq 0,
$$

so that $c_t^+ = \|\mu_t\|_{TV}$. Moreover,

$$
c_t^+ - c_t^- = \int_\Omega (1 + t\phi) \, d\mu_0 = 1 \implies \|\mu_t\|_{TV} = 1 + c_t^- \geq 1,
$$

as $\mu_0 \in \mathcal{P}(\Omega)$ and $\phi\mu_0 \in \mathcal{S}_0(\Omega, \mu)$. Thus,

$$
\begin{aligned}
\|\lambda_t - \lambda_0 - t\nu\|_{TV} &= \int_\Omega \left|\frac{(1 + t\phi)_+}{\|\mu_t\|_{TV}} - 1 - t\phi\right| d\mu_0 \\
&= \int_\Omega \left|-\frac{c_t^-}{\|\mu_t\|_{TV}}(1 + t\phi)_+ + (1 + t\phi)_-\right| d\mu_0 \\
&= \frac{c_t^-}{\|\mu_t\|_{TV}} \underbrace{\int_\Omega (1 + t\phi)_+ d\mu_0}_{=c_t^+=\|\mu_t\|_{TV}} + \underbrace{\int_\Omega (1 + t\phi)_- \, d\mu_0}_{=c_t^-} = 2c_t^-.
\end{aligned}
$$

But as shown above, $\lim_{t\to 0} \frac{c_t^-}{|t|} = 0$, whence $\nu = \dot\lambda(0)$ and $\nu \in T_{\mu_0}\mathcal{P}(\Omega)$. $\qquad\square$

2.3 Powers of densities

Let us now give the formal definition of roots of measures. On the set $\mathcal{M}(\Omega)$ we define the preordering $\mu_1 \leq \mu_2$ if μ_2 dominates μ_1. Then $(\mathcal{M}(\Omega), \leq)$ is a directed set, meaning that for any pair $\mu_1, \mu_2 \in \mathcal{M}(\Omega)$ there is a $\mu_0 \in \mathcal{M}(\Omega)$ dominating both of them (use e.g. $\mu_0 := \mu_1 + \mu_2$, but observe that by Lemma 2.1, this even holds for countable families of measures).

For fixed $r \in (0, 1]$ and measures $\mu_1 \leq \mu_2$ on Ω we define the linear embedding

$$\imath_{\mu_2}^{\mu_1} : L^{1/r}(\Omega, \mu_1) \longrightarrow L^{1/r}(\Omega, \mu_2), \qquad \phi \longmapsto \phi \left(\frac{d\mu_1}{d\mu_2}\right)^r.$$

Observe that

$$\|\imath_{\mu_2}^{\mu_1}(\phi)\|_{1/r} = \left|\int_\Omega |\imath_{\mu_2}^{\mu_1}(\phi)|^{1/r} \, d\mu_2\right|^r = \left|\int_\Omega |\phi|^{1/r} \frac{d\mu_1}{d\mu_2} \, d\mu_2\right|^r \qquad (2.5)$$

$$= \left|\int_\Omega |\phi|^{1/r} \, d\mu_1\right|^r = \|\phi\|_{1/r},$$

so that $\imath_{\mu_2}^{\mu_1}$ is an isometry. Moreover, we have evidently $\imath_{\mu_2}^{\mu_1} \imath_{\mu_3}^{\mu_2} = \imath_{\mu_3}^{\mu_1}$ whenever $\mu_1 \leq \mu_2 \leq \mu_3$. Then we define the *space of r-th roots of measures on Ω* to be the directed limit over the directed set $(\mathcal{M}(\Omega), \leq)$

$$\mathcal{S}^r(\Omega) := \varinjlim L^{1/r}(\Omega, \mu). \qquad (2.6)$$

Let us give a more concrete definition of $\mathcal{S}^r(\Omega)$. On the disjoint union of the spaces $L^{1/r}(\Omega, \mu)$ for $\mu \in \mathcal{M}(\Omega)$ we define the equivalence relation

$$L^{1/r}(\Omega, \mu_1) \ni \phi \sim \psi \in L^{1/r}(\Omega, \mu_2) \qquad \Longleftrightarrow \qquad \imath_{\mu_0}^{\mu_1}(\phi) = \imath_{\mu_0}^{\mu_2}(\psi)$$

$$\Longleftrightarrow \qquad \phi\left(\frac{d\mu_1}{d\mu_0}\right)^r = \psi\left(\frac{d\mu_2}{d\mu_0}\right)^r$$

for some $\mu_0 \geq \mu_1, \mu_2$. Then $\mathcal{S}^r(\Omega)$ is the set of all equivalence classes of this relation.
If we denote the equivalence class of $\phi \in L^{1/r}(\Omega, \mu)$ by $\phi\mu^r$, then the equivalence relation yields

$$\mu_1^r = \left(\frac{d\mu_1}{d\mu_2}\right)^r \mu_2^r \qquad (2.7)$$

whenever $\mu_1 \leq \mu_2$, justifying this notation. In fact, from this description in the case $r = 1$ we see that

$$\mathcal{S}^1(\Omega) = \mathcal{S}(\Omega).$$

Observe that by (2.5) $\|\phi\|_{1/r}$ is constant on equivalence classes, whence there is a norm on $\mathcal{S}^r(\Omega)$, also denoted by $\|.\|_{1/r}$, for which the inclusions

$$L^{1/r}(\Omega, \mu) \longrightarrow \mathcal{S}^r(\Omega), \qquad \phi \longmapsto \phi\mu^r$$

are isometries. For $r = 1$, we have $\|.\|_1 = \|.\|_{TV}$. Thus,

$$\|\phi\mu^r\|_{1/r} = \|\phi\|_{1/r} = \left|\int_\Omega |\phi|^{1/r} \, d\mu\right|^r \quad \text{for } 0 < r \leq 1 \qquad (2.8)$$

Note that the equivalence relation also preserves nonnegativity of functions, whence we may define the subsets

$$\mathcal{M}^r(\Omega) := \{\phi\mu^r : \mu \in \mathcal{M}(\Omega), \phi \geq 0\}$$

$$\mathcal{P}^r(\Omega) := \{\phi\mu^r : \mu \in \mathcal{P}(\Omega), \phi \geq 0, \|\phi\|_{1/r} = 1\}. \qquad (2.9)$$

In analogy to (2.3) we define for a fixed measure $\mu_0 \in \mathcal{M}(\Omega)$ and $r \in (0,1]$ the spaces

$$
\begin{aligned}
\mathcal{S}^r(\Omega, \mu_0) &:= \{\phi\,\mu_0^r \ : \ \phi \in L^{1/r}(\Omega, \mu_0)\} \\
\mathcal{M}^r(\Omega, \mu_0) &:= \{\phi\,\mu_0^r \ : \ \phi \in L^{1/r}(\Omega, \mu_0), \phi \geq 0\} \\
\mathcal{P}^r(\Omega, \mu_0) &:= \{\phi\,\mu_0^r \ : \ \phi \in L^{1/r}(\Omega, \mu_0), \phi \geq 0, \|\phi\|_{1/r} = 1\} \\
\mathcal{S}_0^r(\Omega, \mu_0) &:= \left\{\phi\mu_0^r \ : \ \phi \in L^{1/r}(\Omega, \mu_0), \int_\Omega \phi\, d\mu = 0\right\}.
\end{aligned}
$$

The elements of $\mathcal{P}^r(\Omega, \mu_0), \mathcal{M}^r(\Omega, \mu_0), \mathcal{S}^r(\Omega, \mu_0)$ are said to be *dominated by* μ_0^r.
From Lemma 2.1, we can now conclude the following statement:

Any sequence in $\mathcal{S}^r(\Omega)$ is contained in $\mathcal{S}^r(\Omega, \mu_0)$ for some $\mu_0 \in \mathcal{M}(\Omega)$.

In particular, any Cauchy sequence in $\mathcal{S}^r(\Omega)$ is a Cauchy sequence in $\mathcal{S}^r(\Omega, \mu_0) \cong L^{1/r}(\Omega, \mu_0)$ for some μ_0 and hence convergent. Thus, $(\mathcal{S}^r(\Omega), \|.\|_{1/r})$ is a Banach space.
In analogy to Proposition 2.1, we can also determine the tangent spaces of the subsets $\mathcal{P}^r(\Omega) \subset \mathcal{M}^r(\Omega) \subset \mathcal{S}^r(\Omega)$. The proof of the statement is identical to that of Proposition 2.1 and whence is omitted.

Proposition 2.2. *For each $\mu \in \mathcal{M}(\Omega)$ ($\mu \in \mathcal{P}(\Omega)$, respectively), the tangent spaces of $\mathcal{P}^r(\Omega) \subset \mathcal{M}^r(\Omega) \subset \mathcal{S}^r(\Omega)$ at μ^r are given as*

$$
T_{\mu^r}\mathcal{M}^r(\Omega) \ := \mathcal{S}^r(\Omega, \mu) \subset \mathcal{S}^r(\Omega)
$$

and

$$
T_{\mu^r}\mathcal{P}^r(\Omega) \ := \mathcal{S}_0^r(\Omega, \mu) \subset \mathcal{S}^r(\Omega).
$$

The product of powers of measures can now be defined for all $r, s \in (0,1]$ with $r + s \leq 1$ and for measures $\phi\mu^r \in \mathcal{S}^r(\Omega, \mu)$ and $\psi\mu^s \in \mathcal{S}^s(\Omega, \mu)$:

$$
(\phi\mu^r) \cdot (\psi\mu^s) := \phi\psi\mu^{r+s}.
$$

By definition $\phi \in L^{1/r}(\Omega, \mu)$ and $\psi \in L^{1/s}(\Omega, \mu)$, whence Hölder's inequality implies that $\|\phi\psi\|_{1/(r+s)} \leq \|\phi\|_{1/r}\|\psi\|_{1/s} < \infty$, so that $\phi\psi \in L^{1/(r+s)}(\Omega, \mu)$ and hence, $\phi\psi\mu^{r+s} \in \mathcal{S}^{r+s}(\Omega, \mu)$. Since by (2.7) this definition of the product is independent of the choice of representative μ, it follows that it induces a bilinear product

$$
\cdot : \mathcal{S}^r(\Omega) \times \mathcal{S}^s(\Omega) \longrightarrow \mathcal{S}^{r+s}(\Omega), \qquad \text{where } r, s, r+s \in (0,1], \tag{2.10}
$$

satisfying the Hölder inequality

$$
\|\nu_r \cdot \nu_s\|_{1/(r+s)} \leq \|\nu_r\|_{1/r}\|\nu_s\|_{1/s},
$$

so that the product in (2.10) is a bounded bilinear map.

Definition 2.2. (Canonical pairing)
For $r \in (0,1]$ we define the pairing

$$
(.;.) : \mathcal{S}^r(\Omega) \times \mathcal{S}^{1-r}(\Omega) \longrightarrow \mathbb{R}, \qquad (\nu_1; \nu_2) := \int_\Omega d(\nu_1 \cdot \nu_2). \tag{2.11}
$$

It is straightforward to verify that this pairing is non-degenerate in the sense that

$$(\nu_r; .) = 0 \Longleftrightarrow \nu_r = 0. \tag{2.12}$$

Besides multiplication of roots of measures, we also wish to take their powers. Here, we have two possibilities to deal with signs. For $0 < k \leq r^{-1}$ and $\nu_r = \phi \mu^r \in \mathcal{S}^r(\Omega)$ we define

$$|\nu_r|^k := |\phi|^k \mu^{rk} \qquad \text{and} \qquad \tilde{\nu}_r^k := \text{sign}\,(\phi)|\phi|^k \mu^{rk}.$$

Since $\phi \in L^{1/r}(\Omega, \mu)$, it follows that $|\phi|^k \in L^{k/r}(\Omega, \mu)$, so that $|\nu_r|^k, \tilde{\nu}_r^k \in \mathcal{S}^{rk}(\Omega)$. By (2.7) these powers are well defined, independent of the choice of the measure μ, and, moreover,

$$\| \, |\nu_r|^k \, \|_{1/(rk)} = \|\tilde{\nu}_r^k\|_{1/(rk)} = \|\nu_r\|_{1/r}^k. \tag{2.13}$$

Proposition 2.3. *Let $r \in (0, 1]$ and $0 < k \leq 1/r$, and consider the maps*

$$\pi^k, \tilde{\pi}^k : \mathcal{S}^r(\Omega) \longrightarrow \mathcal{S}^{rk}(\Omega), \qquad \begin{matrix} \pi^k(\nu) := |\nu|^k \\[2mm] \tilde{\pi}^k(\nu) := \tilde{\nu}^k. \end{matrix}$$

Then $\pi^k, \tilde{\pi}^k$ are continuous maps. Moreover, for $1 < k \leq 1/r$ they are C^1-maps between Banach spaces, and their derivatives are given as

$$d_{\nu_r}\tilde{\pi}^k(\rho_r) = k\,|\nu_r|^{k-1} \cdot \rho_r \qquad \text{and} \qquad d_{\nu_r}\pi^k(\rho_r) = k\,\tilde{\nu}_r^{k-1} \cdot \rho_r. \tag{2.14}$$

Observe that for $k = 1$, $\pi^1(\nu_r) = |\nu_r|$ fails to be C^1, whereas $\tilde{\pi}^1(\nu_r) = \nu_r$, so that $\tilde{\pi}^1$ is the identity and hence a C^1-map.

Proof. Let us first assume that $0 < k \leq 1$. We assert that in this case, there are constants $C_k, \tilde{C}_k > 0$ such that for all $x, y \in \mathbb{R}$

$$\begin{matrix} \left| |x + y|^k - |x|^k \right| \leq C_k |y|^k \\[2mm] \text{and} \\[2mm] \left| \text{sign}\,(x + y)|x + y|^k - \text{sign}\,(x)|x|^k \right| \leq \tilde{C}_k |y|^k. \end{matrix} \tag{2.15}$$

Namely, by homogeneity it suffices to show this for $y = 1$, and since the functions

$$x \longmapsto |x + 1|^k - |x|^k \qquad \text{and} \qquad x \longmapsto \text{sign}\,(x + 1)|x + 1|^k - \text{sign}\,(x)|x|^k$$

are continuous and have finite limits for $x \to \pm\infty$, it follows that they are bounded, showing (2.15). Let $\nu_1, \nu_2 \in \mathcal{S}^r(\Omega)$, and choose $\mu_0 \in \mathcal{M}(\Omega)$ such that $\nu_1, \nu_2 \in \mathcal{S}^r(\Omega, \mu_0)$, i.e., $\nu_i = \phi_i \mu_0^r$ with $\phi_i \in L^{1/r}(\Omega, \mu_0)$. Then

$$\begin{aligned} \|\pi^k(\nu_1 + \nu_2) - \pi^k(\nu_1)\|_{1/(rk)} &= \| |\phi_1 + \phi_2|^k - |\phi_1|^k \|_{1/(rk)} \\ &\leq C_k \| \, |\phi_2|^k \|_{1/rk} \qquad \text{by (2.15)} \\ &= C_k \|\nu_2\|_{1/r}^k \qquad \text{by (2.13)}, \end{aligned}$$

so that $\lim_{\|\nu_2\|_{1/r} \to 0} \|\pi^k(\nu_1 + \nu_2) - \pi^k(\nu_1)\|_{1/(rk)} = 0$, showing the continuity of π^k for $0 < k \leq 1$. The continuity of $\tilde{\pi}^k$ follows analogously.

Now let us assume that $1 < k \leq 1/r$. In this case, the functions

$$x \longmapsto |x|^k \qquad \text{and} \qquad x \longmapsto \text{sign}\,(x)|x|^k$$

with $x \in \mathbb{R}$ are C^1-maps with respective derivatives

$$x \longmapsto k\,\text{sign}\,(x)|x|^{k-1} \qquad \text{and} \qquad x \longmapsto k|x|^{k-1}.$$

Thus, if we pick $\nu_i = \phi_i\mu_0^r$ as above, then by the mean value theorem we have

$$\begin{aligned}
\pi^k(\nu_1 + \nu_2) - \pi^k(\nu_1) &= (|\phi_1 + \phi_2|^k - |\phi_1|^k)\mu_0^{rk} \\
&= k\,\text{sign}\,(\phi_1 + \eta\phi_2)|\phi_1 + \eta\phi_2|^{k-1}\phi_2\mu_0^{rk} \\
&= k\,\text{sign}\,(\phi_1 + \eta\phi_2)|\phi_1 + \eta\phi_2|^{k-1}\mu_0^{r(k-1)} \cdot \nu_2
\end{aligned}$$

for some function $\eta : \Omega \to (0,1)$. If we let $\nu_\eta := \eta\phi_2\mu_0^r$, then $\|\nu_\eta\|_{1/r} \leq \|\nu_2\|_{1/r}$, and we get

$$\pi^k(\nu_1 + \nu_2) - \pi^k(\nu_1) = k\tilde{\pi}^{k-1}(\nu_1 + \nu_\eta) \cdot \nu_2.$$

With the definition of $d_{\nu_1}\tilde{\pi}^k$ from (2.14) we have

$$\begin{aligned}
\|\pi^k(\nu_1 + \nu_2) - \pi^k(\nu_1) &- d_{\nu_1}\pi^k(\nu_2)\|_{1/(rk)} \\
&= \|k(\tilde{\pi}^{k-1}(\nu_1 + \nu_\eta) - \tilde{\pi}^{k-1}(\nu_1)) \cdot \nu_2\|_{1/(rk)} \\
&\leq k\|\tilde{\pi}^{k-1}(\nu_1 + \nu_\eta) - \tilde{\pi}^{k-1}(\nu_1)\|_{1/(r(k-1))}\|\nu_2\|_{1/r}
\end{aligned}$$

and hence,

$$\frac{\|\pi^k(\nu_1 + \nu_2) - \pi^k(\nu_1) - d_{\nu_1}\pi^k(\nu_2)\|_{\frac{1}{rk}}}{\|\nu_2\|_{\frac{1}{r}}} \leq k\|\tilde{\pi}^{k-1}(\nu_1 + \nu_\eta) - \tilde{\pi}^{k-1}(\nu_1)\|_{\frac{1}{r(k-1)}}.$$

Thus, the differentiability of π^k will follow if

$$\|\tilde{\pi}^{k-1}(\nu_1 + \nu_\eta) - \tilde{\pi}^{k-1}(\nu_1)\|_{1/(r(k-1))} \xrightarrow{\|\nu_2\|_{1/r} \to 0} 0,$$

and because of $\|\nu_\eta\|_{1/r} \leq \|\nu_2\|_{1/r}$, this is the case if $\tilde{\pi}^{k-1}$ is continuous.

Analogously, one shows that $\tilde{\pi}^k$ is differentiable if π^{k-1} is continuous.

Since we already know continuity of π^k and $\tilde{\pi}^k$ for $0 < k \leq 1$, and since C^1-maps are continuous, the claim now follows by induction on $\lceil k \rceil$. $\qquad\square$

Thus, (2.14) implies that the differentials of π^k and $\tilde{\pi}^k$ (which coincide on $\mathcal{P}^r(\Omega)$ and $\mathcal{M}^r(\Omega)$) yield continuous maps

$$d\pi^k = d\tilde{\pi}^k : \begin{array}{ccc} T\mathcal{P}^r(\Omega) & \longrightarrow & T\mathcal{P}^{rk}(\Omega) \\ \\ T\mathcal{M}^r(\Omega) & \longrightarrow & T\mathcal{M}^{rk}(\Omega) \end{array}, \qquad (\mu, \rho) \longmapsto k\mu^{rk-r} \cdot \rho.$$

3 Congruent embeddings

3.1 Statistics and congruent embeddings

Given two measurable sets (Ω, Σ) and (Ω', Σ'), a measurable map

$$\kappa : \Omega \longrightarrow \Omega'$$

will be called a *statistic*. If μ is a (signed) measure on (Ω, Σ), it then induces a (signed) measure $\kappa_* \mu$ on (Ω', Σ'), via

$$\kappa_* \mu(A) := \mu(\kappa^{-1} A),$$

which is called the *push-forward of μ by κ*. Note that

$$\kappa_* : \mathcal{S}(\Omega) \longrightarrow \mathcal{S}(\Omega') \tag{3.1}$$

is a bounded linear map. Indeed, the norm on both spaces is given by the total variation, and we have for any partition $\Omega' = A'_1 \dot\cup \ldots \dot\cup A'_n$

$$\sum \left| \kappa_* \mu(A'_i) \right| = \sum \left| \mu(\kappa^{-1} A'_i) \right| \leq \|\mu\|_{TV},$$

whence $\|\kappa_* \mu\|_{TV} \leq \|\mu\|_{TV}$. Moreover, κ_* is *monotone*, i.e., it maps nonnegative measures to nonnegative measures, and for these, the total variation is preserved:

$$\|\kappa_* \mu\|_{TV} = \|\mu\|_{TV} \qquad \text{for all } \mu \in \mathcal{M}(\Omega).$$

In particular, κ_* maps probability measures to probability measures, i.e.,

$$\kappa_*(\mathcal{P}(\Omega)) \subset \mathcal{P}(\Omega').$$

We also define the pull-back of a measurable function $\phi' : \Omega' \to \mathbb{R}$ as

$$\kappa^*(\phi') := \phi \circ \kappa.$$

With this, for subsets $A', B' \subset \Omega'$ and $A := \kappa^{-1}(A'), B := \kappa^{-1}(B') \subset \Omega$ we have

$$
\begin{aligned}
\int_{B'} d(\kappa_*(\kappa^*(\chi_{A'})\,\mu)) &= \int_B \kappa^*(\chi_{A'})\, d\mu = \int_{A \cap B} d\mu \\
&= \int_{A' \cap B'} d\kappa_*(\mu) = \int_{B'} \chi_{A'}\, d\kappa_*(\mu),
\end{aligned}
$$

since $\kappa^*(\chi_{A'}) = \chi_A$. Thus, $\kappa_*(\kappa^*(\chi_{A'})\mu) = \chi_{A'}\kappa_*(\mu)$. By linearity, this equation holds when replacing $\chi_{A'}$ by a step function on Ω', whence by the density of step functions in $L^1(\Omega', \mu')$ we obtain

$$\kappa_*(\kappa^*(\phi')\,\mu) = \phi'\,\kappa_*(\mu) \qquad \text{for all } \phi' \in L^1(\Omega', \kappa_*(\mu)). \tag{3.2}$$

Recall that $\mathcal{M}(\Omega)$ and $\mathcal{S}(\Omega)$ denote the spaces of *all* (signed) measures on Ω, whereas $\mathcal{M}(\Omega, \mu)$ and $\mathcal{S}(\Omega, \mu)$ denote the subspaces of the (signed) measures on Ω which are dominated by μ.

Definition 3.1. (Congruent embedding)
Let $\kappa : \Omega \to \Omega'$ be a statistic and $\mu' \in \mathcal{M}(\Omega')$. A *$\kappa$-congruent embedding* is a bounded linear map $K_* : \mathcal{S}(\Omega', \mu') \to \mathcal{S}(\Omega)$ such that

12

1. K_* is monotone, i.e., it maps nonnegative measures to nonnegative measures, or shortly: $K_*(\mathcal{M}(\Omega', \mu')) \subset \mathcal{M}(\Omega)$.

2. $\kappa_*(K_*(\nu')) = \nu'$ for all $\nu' \in \mathcal{S}(\Omega', \mu')$.

Furthermore, the image of a κ-congruent embedding K_* in $\mathcal{S}(\Omega)$ is called a *κ-congruent subspace of $\mathcal{S}(\Omega)$*.

Example 3.1. Let $\kappa : \Omega \to \Omega'$ be a statistic, let $\mu \in \mathcal{M}(\Omega)$ and $\mu' := \kappa_*(\mu) \in \mathcal{M}(\Omega')$. Then the map

$$K_\mu : \mathcal{S}(\Omega', \mu') \longrightarrow \mathcal{S}(\Omega, \mu) \subset \mathcal{S}(\Omega), \qquad \phi'\mu' \longmapsto \kappa^*(\phi')\mu \tag{3.3}$$

for all $\phi' \in L^1(\Omega', \mu')$ is a κ-congruent embedding, since

$$\kappa_*(K_\mu(\phi'\mu')) = \kappa_*(\kappa^*(\phi')\mu) = \phi'\kappa_*(\mu) = \phi'\mu'$$

by (3.2).

We shall now see that the above example exhausts *all* possibilities of congruent embeddings.

Proposition 3.1. *Let $\kappa : \Omega \to \Omega'$ be a statistic, let $K_* : \mathcal{S}(\Omega', \mu') \to \mathcal{S}(\Omega)$ for some $\mu' \in \mathcal{M}(\Omega')$ be a κ-congruent embedding, and let $\mu := K_*(\mu') \in \mathcal{M}(\Omega)$.*
Then $K_* = K_\mu$ with the map K_μ given in (3.3).

Proof. We have to show that $K_*(\phi'\mu) = \kappa^*(\phi')\mu$ for all $\phi' \in L^1(\Omega', \mu')$. By continuity, it suffices to show this for step functions, as these are dense in $L^1(\Omega', \mu')$, whence by linearity, we have to show that for all $A' \subset \Omega'$, $A := \kappa^{-1}(A') \subset \Omega$

$$K_*(\chi_{A'}\mu') = \chi_A\mu. \tag{3.4}$$

Let $A_1' := A'$ and $A_2' = \Omega' \backslash A'$, and let $A_i := \kappa^{-1}(A_i')$. We define the measures $\mu_i' := \chi_{A_i'}\mu' \in \mathcal{M}(\Omega')$, and $\mu_i := K_*(\mu_i') \in \mathcal{M}(\Omega)$. Since $\mu_1' + \mu_2' = \mu'$, it follows that $\mu_1 + \mu_2 = \mu$ by the linearity of K_*. Taking indices mod 2, and using $\kappa_*(\mu_i) = \kappa_*(K_*(\mu_i')) = \mu_i'$ by the κ-congruency of K_*, note that

$$\mu_i(A_{i+1}) = \mu_i(\kappa^{-1}(A_{i+1}')) = \kappa_*(\mu_i)(A_{i+1}') = \mu_i'(A_{i+1}') = 0.$$

Thus, for any measurable $B \subset \Omega$ we have

$$
\begin{array}{llll}
\mu_1(B) &=& \mu_1(B \cap A_1) & \text{since } \mu_1(B \cap A_2) \leq \mu_1(A_2) = 0 \\
&=& \mu_1(B \cap A_1) + \mu_2(B \cap A_1) & \text{since } \mu_2(B \cap A_1) \leq \mu_2(A_1) = 0 \\
&=& \mu(B \cap A_1) & \text{since } \mu = \mu_1 + \mu_2 \\
&=& (\chi_A\mu)(B) & \text{since } A_1 = A.
\end{array}
$$

That is, $\chi_A\mu = \mu_1 = K_*(\mu_1') = K_*(\chi_{A'}\mu')$, so that (3.4) follows. $\qquad\square$

As a consequence, any κ-congruent subspace of $\mathcal{S}(\Omega)$ must be of the form

$$\mathcal{C}_{\kappa,\mu} := \{\kappa^*(\phi')\mu \ : \ \phi' \in L^1(\Omega', \kappa_*(\mu))\} \tag{3.5}$$

for some $\mu \in \mathcal{M}(\Omega)$.

3.2 Markov kernels and Markov morphisms

Definition 3.2. (Markov kernel and Markov morphism, cf. [8], [15])
A *Markov kernel* between two measurable spaces (Ω_1, Σ_1) and (Ω_2, Σ_2) is a map $K : \Omega_1 \to \mathcal{P}(\Omega_2)$, associating to each $\omega_1 \in \Omega_1$ a probability measure on Ω_2 such that for each fixed measurable $A_2 \subset \Omega_2$ the map

$$\Omega_1 \longrightarrow [0,1], \qquad \omega_1 \longmapsto K(\omega_1; A_2) := K(\omega_1)(A_2)$$

is measurable. The *Markov morphism induced by K* is the linear map

$$K_* : \mathcal{S}(\Omega_1) \longrightarrow \mathcal{S}(\Omega_2), \qquad K(\mu_1; A_2) = K_*(\mu_1)(A_2) := \int_{\Omega_1} K(\omega_1; A_2) \, d\mu_1(\omega_1). \qquad (3.6)$$

Since $K(\omega_1) \in \mathcal{P}(\Omega_2)$, it follows that $K(\omega_1; \Omega_2) = 1$ and hence (3.6) implies that $K_*(\mu_1)(\Omega_2) = \mu_1(\Omega_1)$. Thus,

$$\|K_*(\mu_1)\|_{TV} = \|\mu_1\|_{TV} \qquad \text{for all } \mu_1 \in \mathcal{M}(\Omega_1). \qquad (3.7)$$

In particular, a Markov morphism maps probability measures to probability measures. For a general measure $\mu_1 \in \mathcal{S}(\Omega_1)$, (2.2) implies that $|K_*(\mu_1; A_2)| \le K_*(|\mu_1|; A_2)$ for all $A_2 \in \Sigma_2$ and hence,

$$\|K_*(\mu_1)\|_{TV} \le \|K_*(|\mu_1|)\|_{TV} = \|\mu_1\|_{TV} \qquad \text{for all } \mu_1 \in \mathcal{S}(\Omega_1),$$

so that $K_* : \mathcal{S}(\Omega_1) \to \mathcal{S}(\Omega_2)$ is a bounded linear map.
Observe that we can recover the Markov kernel K from K_* using the relation

$$K(\omega_1) = K_*(\delta^{\omega_1}) \qquad \text{for all } \omega_1 \in \Omega_1,$$

where δ^{ω_1} denotes the Dirac measure supported at $\omega_1 \in \Omega_1$.

Remark 3.1. From (3.6) it is immediate that K_* preserves dominance of measures, i.e., if μ_1 dominates μ_1', then $K_*(\mu_1)$ dominates $K_*(\mu_1')$. Thus, for each $\mu_1 \in \mathcal{M}(\Omega_1)$ there is a restriction

$$K_* : \mathcal{S}(\Omega_1, \mu_1) \longrightarrow \mathcal{S}(\Omega_2, \mu_2),$$

where $\mu_2 := K_*(\mu_1)$.

Definition 3.3. (Composition of Markov kernels)
Let (Ω_i, Σ_i), $i = 1, 2, 3$ be measurable spaces, and let $K_i : \Omega_i \to \mathcal{P}(\Omega_{i+1})$ for $i = 1, 2$ be Markov kernels. The *composition of K_1 and K_2* is the Markov kernel

$$K_2 K_1 : \Omega_1 \longrightarrow \mathcal{P}(\Omega_3), \qquad \omega \longmapsto (K_2)_*(K_1(\omega)).$$

Since $\|(K_2)_*(K_1(\omega))\|_{TV} = \|K_1(\omega)\|_{TV} = 1$ by (3.7), $(K_2)_*(K_1(\omega))$ is a probability measure, hence this composition yields indeed a Markov kernel. Moreover, it is straightforward to verify that this composition is associative, and for the induced Markov morphism we have

$$(K_2 K_1)_* = (K_2)_*(K_1)_*. \qquad (3.8)$$

Markov kernels are generalizations of statistics. In fact, a statistic $\kappa : \Omega \to \Omega'$ induces a Markov kernel by

$$K^\kappa(\omega) := \delta^{\kappa(\omega)}, \qquad \text{so that} \qquad K^\kappa(\omega; A') := \chi_{\kappa^{-1}(A')}(\omega).$$

In this case, the Markov morphism induced by K^κ is the map $\kappa_* : \mathcal{S}(\Omega) \to \mathcal{S}(\Omega')$ from (3.1). We shall write the Markov kernel K^κ also as κ if there is no danger of confusion.

14

Definition 3.4. (Congruent Markov kernels)

A Markov kernel $K : \Omega' \to \mathcal{P}(\Omega)$ is called κ-*congruent* for a statistic $\kappa : \Omega \to \Omega'$ if

$$\kappa_*(K(\omega')) = \delta^{\omega'} \qquad \text{for all } \omega' \in \Omega', \tag{3.9}$$

or, equivalently,

$$(K^\kappa K)_* = Id_{\mathcal{S}(\Omega')} : \mathcal{S}(\Omega') \longrightarrow \mathcal{S}(\Omega').$$

In this case, we also call the induced Markov morphism $K_* : \mathcal{S}(\Omega') \to \mathcal{S}(\Omega)$ κ-congruent.

In order to relate the notions of κ-congruent Markov morphism and κ-congruent embeddings from Definition 3.1, we need the notion of κ-transverse measures.

Definition 3.5. (Transverse measures)

Let $\kappa : \Omega \to \Omega'$ be a statistic. A measure $\mu \in \mathcal{M}(\Omega)$ is said to admit κ-*transverse measures* if there are measures $\mu_{\omega'}^\perp$ on $\kappa^{-1}(\omega')$ such that for all $\phi \in L^1(\Omega, \mu)$

$$\int_\Omega \phi \, d\mu = \int_{\Omega'} \left(\int_{\kappa^{-1}(\omega')} \phi \, d\mu_{\omega'}^\perp \right) d\mu'(\omega'), \tag{3.10}$$

where $\mu' := \kappa_*(\mu)$. In particular, the function

$$\Omega' \longrightarrow \hat{\mathbb{R}}, \qquad \omega' \longmapsto \int_{\kappa^{-1}(\omega')} \phi \, d\mu_{\omega'}^\perp$$

is measurable for all $\phi \in L^1(\Omega, \mu)$.

Observe that the choice of κ-transverse measures $\mu_{\omega'}^\perp$ is not unique, but rather, one can change these measures for all ω' in a μ'-null set.

Proposition 3.2. *Let $\kappa : \Omega \to \Omega'$ be a statistic and $\mu \in \mathcal{M}(\Omega)$ a measure which admits κ-transverse measures $\{\mu_{\omega'}^\perp : \omega' \in \Omega'\}$. Then $\mu_{\omega'}^\perp$ is a probability measure for almost every $\omega' \in \Omega'$ and hence, we may assume w.l.o.g. that $\mu_{\omega'}^\perp \in \mathcal{P}(\kappa^{-1}(\omega'))$ for all $\omega' \in \Omega'$.*

Proof. Given $\varepsilon > 0$, define $A_\varepsilon' := \{\omega' \in \Omega' : \mu_{\omega'}^\perp(\kappa^{-1}(\omega')) \geq 1 + \varepsilon\}$. Then for $\phi := \chi_{\kappa^{-1}(A_\varepsilon')}$ the two sides of equation (3.10) read

$$\int_\Omega \chi_{\kappa^{-1}(A_\varepsilon')} \, d\mu = \mu(\kappa^{-1}(A_\varepsilon')) = \mu'(A_\varepsilon')$$

$$\int_{\Omega'} \left(\int_{\kappa^{-1}(\omega')} \chi_{\kappa^{-1}(A_\varepsilon)} \, d\mu_{\omega'}^\perp \right) d\mu'(\omega') = \int_{A_\varepsilon'} \left(\int_{\kappa^{-1}(\omega')} d\mu_{\omega'}^\perp \right) d\mu'(\omega')$$

$$= \int_{A_\varepsilon'} \mu_{\omega'}^\perp(\kappa^{-1}(\omega')) \, d\mu'(\omega')$$

$$\geq (1 + \varepsilon)\mu'(A_\varepsilon').$$

Thus, (3.10) implies

$$\mu'(A_\varepsilon') \geq (1 + \varepsilon)\mu'(A_\varepsilon'),$$

and hence, $\mu'(A'_\varepsilon) = 0$ for all $\varepsilon > 0$. Thus,

$$\mu'(\{\omega' \in \Omega' \, : \, \mu^\perp_{\omega'}(\kappa^{-1}(\omega')) > 1\}) = \mu'\left(\bigcup_{n=1}^\infty A'_{1/n}\right) \leq \sum_{n=1}^\infty \mu'(A'_{1/n}) = 0,$$

whence $\{\omega' \in \Omega' \, : \, \mu^\perp_{\omega'}(\kappa^{-1}(\omega')) > 1\}$ is a μ'-null set. Analogously, $\{\omega' \in \Omega' \, : \, \mu^\perp_{\omega'}(\kappa^{-1}(\omega')) < 1\}$ is a μ'-null set, that is, $\mu^\perp_{\omega'} \in \mathcal{P}(\kappa^{-1}(\omega'))$ and hence $\|\mu^\perp_{\omega'}\|_{TV} = 1$ for μ'-a.e. $\mu' \in \Omega'$.
Thus, if we replace $\mu^\perp_{\omega'}$ by $\tilde\mu^\perp_{\omega'} := \|\mu^\perp_{\omega'}\|^{-1}_{TV}\mu^\perp_{\omega'}$, then $\tilde\mu^\perp_{\omega'} \in \mathcal{P}(\kappa^{-1}(\omega'))$ for *all* $\omega' \in \Omega'$, and since $\tilde\mu^\perp_{\omega'} = \mu^\perp_{\omega'}$ for μ'-a.e. $\omega' \in \Omega'$, it follows that (3.10) holds when replacing $\mu^\perp_{\omega'}$ by $\tilde\mu^\perp_{\omega'}$. $\qquad\square$

We are now ready to relate the notions of κ-congruent embeddings and κ-congruent Markov kernels.

Theorem 3.1. *Let $\kappa : \Omega \to \Omega'$ be a statistic and $\mu' \in \mathcal{M}(\Omega')$ be a measure.*

1. *If $K : \Omega' \to \mathcal{P}(\Omega)$ is a κ-congruent Markov kernel, then the restriction of K_* to $\mathcal{S}(\Omega', \mu') \subset \mathcal{S}(\Omega')$ is a κ-congruent embedding and hence, for $\phi' \in L^1(\Omega', \mu')$ we have*

$$K_*(\phi'\mu') = \kappa^*(\phi')K_*(\mu').$$

2. *Conversely, if $K_* : \mathcal{S}(\Omega', \mu') \to \mathcal{S}(\Omega)$ is a κ-congruent embedding, then the following are equivalent.*

    (a) *K_* is the restriction of a κ-congruent Markov morphism to $\mathcal{S}(\Omega', \mu') \subset \mathcal{S}(\Omega')$.*

    (b) *$\mu := K_*(\mu') \in \mathcal{S}(\Omega)$ admits κ-transverse measures.*

Theorem 3.1 implies that the two notions of congruency are equivalent for large classes of statistics κ, since the existence of transversal measures is guaranteed under rather mild hypotheses, e.g. if one of Ω, Ω' is a finite set, or if Ω, Ω' are differentiable manifolds equipped with a Borel measure μ and κ is a differentiable map.
However, there *are* examples of statistics and measures which do *not* admit κ-transverse measures, cf. Example 3.2 below.

Proof. The first statement follows directly from $(K^\kappa K)_* = (K^\kappa)_* K_* = \kappa_* K_*$ by (3.8) and Proposition 3.1.
For the second, suppose that $K_* : \mathcal{S}(\Omega', \mu') \to \mathcal{S}(\Omega)$ is a κ-congruent embedding. Then $K_* = K_\mu$ given in (3.3) for the measure $\mu := K_*(\mu')$ by Proposition 3.1.
If we assume that K_* is the restriction of a κ-congruent Markov morphism induced by the κ-congruent Markov kernel $K : \Omega' \to \mathcal{P}(\Omega)$, then we define the measures

$$\mu^\perp_{\omega'} := K(\omega')|_{\kappa^{-1}(\omega)} \in \mathcal{M}(\kappa^{-1}(\omega')).$$

Note that for $\omega' \in \Omega'$

$$\begin{aligned}
K(\omega'; \Omega\backslash\kappa^{-1}(\omega')) &= \int_{\Omega\backslash\kappa^{-1}(\omega')} dK(\omega') = \int_{\Omega'\backslash\omega'} d\kappa_*(K(\omega')) \\
&\overset{(3.9)}{=} \int_{\Omega'\backslash\omega'} d\delta^{\omega'} = 0.
\end{aligned}$$

That is, $K(\omega')$ is supported on $\kappa^{-1}(\omega')$ and hence, for an arbitrary set $A \subset \Omega$ we have

$$K(\omega'; A) = K(\omega'; A \cap \kappa^{-1}(\omega')) = \mu_{\omega'}^{\perp}(A \cap \kappa^{-1}(\omega')) = \int_{\kappa^{-1}(\omega')} \chi_A \, d\mu_{\omega'}^{\perp}.$$

Substituting this into the definition of K_* we obtain for a subset $A \subset \Omega$

$$\begin{aligned}
\int_{\Omega} \chi_A \, d\mu \ &= \ \mu(A) = K_*(\mu'; A) \overset{(3.6)}{=} \int_{\Omega'} K(\omega'; A) \, d\mu'(\omega') \\
&= \ \int_{\Omega'} \left(\int_{\kappa^{-1}(\omega')} \chi_A \, d\mu_{\omega'}^{\perp} \right) d\mu'(\omega'),
\end{aligned}$$

showing that (3.10) holds for $\phi = \chi_A$. But then, by linearity (3.10) holds for any step function ϕ, and since these are dense in $L^1(\Omega, \mu)$, it follows that (3.10) holds for all ϕ, so that the measures $\mu_{\omega'}^{\perp}$ defined above yield indeed κ-transverse measures of μ.

Conversely, suppose that $\mu := K_*(\mu')$ admits κ-transverse measures $\mu_{\omega'}^{\perp}$, and by Proposition 3.2 we may assume w.l.o.g. that $\mu_{\omega'}^{\perp} \in \mathcal{P}(\kappa^{-1}(\omega'))$. Then we define the map

$$K : \Omega' \longrightarrow \mathcal{P}(\Omega), \qquad K(\omega'; A) := \mu_{\omega'}^{\perp}(A \cap \kappa^{-1}(\omega')) = \int_{\kappa^{-1}(\omega')} \chi_A \, d\mu_{\omega'}^{\perp}.$$

Since for fixed $A \subset \Omega$ the map $\omega' \mapsto \int_{\kappa^{-1}(\omega')} \chi_A \, d\mu_{\omega'}^{\perp}$ is measurable by the definition of transversal measures, K is indeed a Markov kernel. Moreover, for $A' \subset \Omega'$

$$\kappa_*(K(\omega'))(A') = K(\omega'; \kappa^{-1}(A')) = \mu_{\omega'}^{\perp}(\kappa^{-1}(A') \cap \kappa^{-1}(\omega')) = \chi_{A'}(\omega'),$$

so that $\kappa_* K(\omega') = \delta^{\omega'}$ for all $\omega' \in \Omega'$, whence K is κ-congruent. Moreover, for any $\phi' \in L^1(\Omega', \mu')$ and $A \subset \Omega$ we have

$$\begin{aligned}
K_\mu(\phi'\mu')(A) \ &\overset{(3.3)}{=} \ \kappa^*(\phi')\mu(A) = \int_{\Omega} \chi_A \kappa^*(\phi') \, d\mu \\
&\overset{(3.10)}{=} \ \int_{\Omega'} \left(\int_{\kappa^{-1}(\omega')} \chi_A \kappa^*(\phi') \, d\mu_{\omega'}^{\perp} \right) d\mu'(\omega') \\
&= \ \int_{\Omega'} \left(\int_{\kappa^{-1}(\omega')} \chi_A \, d\mu_{\omega'}^{\perp} \right) \phi'(\omega') \, d\mu'(\omega') \\
&= \ \int_{\Omega'} K(\omega'; A) \, d(\phi'\mu')(\omega') \overset{(3.6)}{=} K_*(\phi'\mu')(A).
\end{aligned}$$

Thus, $K_\mu(\phi'\mu') = K_*(\phi'\mu')$ for all $\phi' \in L^1(\Omega', \mu')$ and hence, $K_\mu(\nu) = K_*(\nu)$ for all $\nu \in \mathcal{S}(\Omega', \mu')$. That is, the given congruent embedding K_μ coincides with the Markov morphism K_* induced by K, and this completes the proof. $\qquad \square$

Now we give an example of a statistic which does not admit κ-transverse measures.

Example 3.2. Let $\Omega := S^1$ be the unit circle group in the complex plain with the 1-dimensional Borel algebra $\mathfrak{B}$. Let $\Gamma := \exp(2\pi\sqrt{-1}\mathbb{Q}) \subset S^1$ be the subgroup of rational rotations, and let $\Omega' := S^1/\Gamma$ be the quotient space with the canonical projection $\kappa : \Omega \to \Omega'$. Let $\mathfrak{B}' := \{A' \subset \Omega' :$

17

$\kappa^{-1}(A') \in \mathfrak{B}\}$, so that $\kappa : \Omega \to \Omega'$ is measurable. For $\gamma \in \Gamma$, we let $m_\gamma : S^1 \to S^1$ denote the multiplication by γ.

Let λ be the 1-dimensional Lebesgue measure on Ω and $\lambda' := \kappa_*(\lambda)$ be the induced measure on Ω'. Suppose that λ admits κ-transverse measures $(\lambda_{\omega'}^\perp)_{\omega' \in \Omega'}$. Then for each $A \in \mathfrak{B}$ we have

$$\lambda(A) = \int_{\Omega'} \left(\int_{A \cap \kappa^{-1}(\omega')} d\lambda_{\omega'}^\perp \right) d\lambda'(\omega'). \tag{3.11}$$

Since λ is invariant under rotations, we have on the other hand for $\gamma \in \Gamma$

$$\lambda(A) = \lambda(m_\gamma^{-1} A) = \int_{\Omega'} \left(\int_{(m_\gamma^{-1} A) \cap \kappa^{-1}(\omega')} d\lambda_{\omega'}^\perp \right) d\lambda'(\omega')$$

$$= \int_{\Omega'} \left(\int_{A \cap \kappa^{-1}(\omega')} d((m_\gamma)_* \lambda_{\omega'}^\perp) \right) d\lambda'(\omega'). \tag{3.12}$$

Comparing (3.11) and (3.12) implies that $((m_\gamma)_* \lambda_{\omega'}^\perp)_{\omega' \in \Omega'}$ is another familiy of κ-transverse measures of λ which implies that $(m_\gamma)_* \lambda_{\omega'}^\perp = \lambda_{\omega'}^\perp$ for λ'-a.e. $\omega' \in \Omega'$, and as Γ is countable, it follows that

$$(m_\gamma)_* \lambda_{\omega'}^\perp = \lambda_{\omega'}^\perp \qquad \text{for all } \gamma \in \Gamma \text{ and } \lambda'\text{-a.e. } \omega' \in \Omega'.$$

Thus, for a.e. $\omega' \in \Omega'$ we have $\lambda_{\omega'}^\perp(\{\gamma \cdot x\}) = \lambda_{\omega'}^\perp(\{x\})$, and since Γ acts transitively on $\kappa^{-1}(\omega')$, it follows that singleton subsets have equal measure, i.e., there is a constant $c_{\omega'}$ with

$$\lambda_{\omega'}^\perp(A') = c_{\omega'}|A'|$$

for all $A' \subset \kappa^{-1}(\omega')$. As $\kappa^{-1}(\omega')$ is countable and infinite, this implies that $\lambda_{\omega'}^\perp = 0$ if $c_{\omega'} = 0$, and $\lambda_{\omega'}^\perp(\kappa^{-1}(\omega')) = \infty$ if $c_{\omega'} > 0$. Thus, $\lambda_{\omega'}^\perp$ is not a probability measure for a.e. $\omega' \in \Omega'$, contradicting Proposition 3.2. This shows that λ does not admit κ-transverse measures.

We conclude this section by the following result (cf. [5, Theorem 4.10]).

Theorem 3.2. *Any Markov kernel $K = \Omega \to \mathcal{P}(\Omega')$ can be decomposed into a statistic and a congruent Markov kernel. That is, there is a Markov kernel $K^{cong} : \Omega \to \mathcal{P}(\hat{\Omega})$ which is congruent w.r.t. some statistic $\kappa_1 : \hat{\Omega} \to \Omega$, and a statistic $\kappa_2 : \hat{\Omega} \to \Omega'$ such that*

$$K = K^{\kappa_2} K^{cong}.$$

Proof. Let $\hat{\Omega} := \Omega \times \Omega'$ and let $\kappa_1 : \hat{\Omega} \to \Omega$ and $\kappa_2 : \hat{\Omega} \to \Omega'$ be the canonical projections. We define the Markov kernel

$$K^{cong} : \Omega \longrightarrow \mathcal{P}(\hat{\Omega}), \qquad K^{cong}(\omega_1; \hat{A}) := K(\omega_1; \kappa_2(\hat{A} \cap (\{\omega_1\} \times \Omega'))),$$

and we assert that it is κ_1-congruent. Namely,

$$
\begin{aligned}
(\kappa_1)_*(K^{cong}(\omega_1))(A_1) &= K^{cong}(\omega_1; \kappa_1^{-1}(A_1)) = K^{cong}(\omega_1; A_1 \times \Omega') \\
&= K(\omega_1; \kappa_2((A_1 \times \Omega') \cap (\{\omega_1\} \times \Omega'))) \\
&= \begin{cases} K(\omega_1; \Omega') = 1 & \text{if } \omega_1 \in A_1 \\ K(\omega_1; \emptyset) = 0 & \text{if } \omega_1 \notin A_1 \end{cases} \\
&= \chi_{A_1}(\omega_1),
\end{aligned}
$$

18

whence $(\kappa_1)_*(K^{cong}(\omega_1)) = \delta^{\omega_1}$ as claimed. Observe that

$$
\begin{aligned}
(\kappa_2)_*(K^{cong}(\omega_1))(A_2) &= K^{cong}(\omega_1; \kappa_2^{-1}(A_2)) = K^{cong}(\omega_1; \Omega_1 \times A_2) \\
&= K(\omega_1; \kappa_2((\Omega_1 \times A_2) \cap (\{\omega_1\} \times \Omega'))) \\
&= K(\omega_1; \kappa_2(\{\omega_1\} \times A_2)) \\
&= K(\omega_1; A_2).
\end{aligned}
$$

Therefore, $K = K^{\kappa_2} K^{cong}$, so the claim follows. $\qquad\square$

3.3 Powers of densities and congruent embeddings

Given a statistic $\kappa : \Omega \to \Omega'$, recall that by Proposition 3.1 any κ-congruent embedding is of the form $K_\mu : \mathcal{S}(\Omega', \mu') \to \mathcal{S}(\Omega, \mu)$ for some $\mu \in \mathcal{M}(\Omega)$ with $\mu' := \kappa_*(\mu)$, and the image of K_μ, denoted by $\mathcal{C}_{\kappa,\mu}$, is called a κ-congruent subspace.

We now wish to generalize the notion of congruent embeddings and congruent subspaces to powers of measures. Namely, for $r \in (0,1]$ we define the congruent embedding of r-th powers to be the map

$$
K_\mu^r : \mathcal{S}^r(\Omega', \mu') \longrightarrow \mathcal{S}^r(\Omega), \qquad \phi\mu'^r \longmapsto \kappa^*(\phi)\mu^r,
$$

and denote its image by

$$
\mathcal{C}_{\kappa,\mu}^r := \{\kappa^*(\phi')\mu^r \ : \ \phi' \in L^{1/r}(\Omega', \mu')\}. \tag{3.13}
$$

As before, we denote the spaces of the form $\mathcal{C}_{\kappa,\mu}^r$ for $\mu \in \mathcal{M}(\Omega)$ as κ-*congruent subspaces*. Observe that $K_\mu^1 = K_\mu$ and $\mathcal{C}_{\kappa,\mu}^1 = \mathcal{C}_{\kappa,\mu}$ with the definitions from (3.3) and (3.5).

Since the pull-back of functions preserves products and powers, we have immediately the following properties.

$$
\|K_\mu^r(\nu^r)\|_{1/r} = \|\nu^r\|_{1/r}
$$

$$
K_\mu^{r_1+r_2}(\nu'^{r_1} \cdot \nu'^{r_2}) = K_\mu^{r_2}(\nu'^{r_1}) \cdot K_\mu^{r_2}(\nu'^{r_2}) \tag{3.14}
$$

$$
K_\mu^{r\alpha}(\pi^\alpha(\nu'^r)) = \pi^\alpha(K_\mu^r(\nu'^r)) \quad \text{and} \quad K_\mu^{r\alpha}(\tilde{\pi}^\alpha(\nu'^r)) = \tilde{\pi}^\alpha(K_\mu^r(\nu'^r))
$$

for $r_1 + r_2 \leq 1$ and $0 < \alpha < 1/r$. We now show the following decomposition result.

Proposition 3.3. *Let $\kappa : \Omega \to \Omega'$, $\mu \in \mathcal{M}(\Omega)$ and $\mu' := \kappa_*(\mu) \in \mathcal{M}(\Omega')$ be as above. Then for the congruent subspaces $\mathcal{C}_{\kappa,\mu}^r$, we have the decomposition*

$$
\mathcal{S}^r(\Omega, \mu) = \mathcal{C}_{\kappa,\mu}^r \oplus \left((\mathcal{C}_{\kappa,\mu}^{1-r})^\perp \cap \mathcal{S}^r(\Omega, \mu) \right), \tag{3.15}
$$

where

$$
(\mathcal{C}_{\kappa,\mu}^{1-r})^\perp = \{\nu^r \in \mathcal{S}^r(\Omega) \ : \ (\nu^r; \rho^{1-r}) = 0 \quad \text{for all } \rho^{1-r} \in \mathcal{C}_{\kappa,\mu}^{1-r}\}
$$

with the pairing $(.;.)$ from (2.11). In particular, for $r = 1/2$ (3.15) is an orthogonal decomposition w.r.t. the Hilbert metric on $\mathcal{S}^{1/2}(\Omega, \mu)$. Moreover,

$$
(\mathcal{C}_{\kappa,\mu}^{1-r})^\perp \cap \mathcal{S}^r(\Omega, \mu) = \{\nu^r \in \mathcal{S}^r(\Omega, \mu) \ : \ \kappa_*(\nu^r \cdot \mu^{1-r}) = 0\}. \tag{3.16}
$$

Proof. We begin by showing (3.16). Let $\nu^r = \phi\mu^r \in \mathcal{S}^r(\Omega, \mu)$ with $\phi \in L^{1/r}(\Omega, \mu)$. Then $\nu^r \in (\mathcal{C}_{\kappa,\mu}^{1-r})^\perp$ if and only if for all $\psi' \in L^{1/(1-r)}(\Omega', \mu')$ we have

$$
\begin{aligned}
0 &= \int_\Omega d\left(\nu^r \cdot \left(\kappa^*(\psi')\mu^{1-r}\right)\right) = \int_\Omega \phi\kappa^*(\psi')\, d\mu = \int_{\Omega'} d\kappa_*(\phi\kappa^*(\psi')\mu) \\
&\stackrel{(3.2)}{=} \int_{\Omega'} \psi'\, d\kappa_*(\phi\mu).
\end{aligned}
$$

But $\int_{\Omega'} \psi'\, d\kappa_*(\phi\mu) = 0$ for all $\psi' \in L^{1/(1-r)}(\Omega', \mu')$ holds if and only if $\kappa_*(\phi\mu) = 0$, which shows (3.16) since $\phi\mu = \nu^r \cdot \mu^{1-r}$.

The spaces on the right hand side of (3.15) are obviously contained in $\mathcal{S}^r(\Omega, \mu)$. To see that $\mathcal{C}_{\kappa,\mu}^r \cap (\mathcal{C}_{\kappa,\mu}^{1-r})^\perp = 0$, observe that for $\nu^r = \kappa^*(\phi')\mu^r \in \mathcal{C}_{\kappa,\mu}^r$ we have

$$
\kappa_*(\nu^r \cdot \mu^{1-r}) = \kappa_*(\kappa^*(\phi')\mu) = \phi'\mu'
$$

by (3.2), and by (3.16) this is contained in $(\mathcal{C}_{\kappa,\mu}^{1-r})^\perp$ only if $\phi' = 0$.

Since $K_\mu^r : \mathcal{S}^r(\Omega', \mu') \to \mathcal{C}_{\kappa,\mu}^r$ is an isometry by (3.14) and $\mathcal{S}^r(\Omega', \mu')$ is complete, it follows that $\mathcal{C}_{\kappa,\mu}^r$ is complete and hence closed in $\mathcal{S}^r(\Omega, \mu)$. Also $(\mathcal{C}_{\kappa,\mu}^{1-r})^\perp \subset \mathcal{S}^r(\Omega)$ is a closed subspace being the orthogonal complement of the subspace $\mathcal{C}_{\kappa,\mu}^{1-r}$. Therefore, the right hand side of (3.15) is a closed subspace of the left, hence it suffices to show that the right hand side contains the dense subspace $\{\nu^r = \phi\mu^r \; : \; \phi \in L^\infty(\Omega, \mu)\}$ of $\mathcal{S}^r(\Omega, \mu)$.

Thus, for such a $\nu^r = \phi\mu^r \in \mathcal{S}^r(\Omega, \mu)$ define $\phi' \in L^1(\Omega', \mu')$ by

$$
\kappa_*(\nu^r \cdot \mu^{1-r}) = \kappa_*(\phi\mu) = \phi'\mu'.
$$

Then for $A' \subset \Omega'$ we have

$$
|(\kappa_*(\phi\mu))(A')| = \left|\int_{\kappa^{-1}(A')} \phi\, d\mu\right| \leq \|\phi\|_\infty \mu(\kappa^{-1}A') = \|\phi\|_\infty \mu'(A'),
$$

whence $\|\phi'\|_\infty \leq \|\phi\|_\infty$, so that ϕ' is bounded as well. In particular, $\phi'\mu'^r \in \mathcal{S}^r(\Omega', \mu')$, and we decompose

$$
\nu^r = \underbrace{\kappa^*(\phi')\mu^r}_{\in \mathcal{C}_{\kappa,\mu}^r} + (\nu^r - \kappa^*(\phi')\mu^r).
$$

Thus, it remains to show that $\nu_1^r := \nu^r - \kappa^*(\phi')\mu^r \in (\mathcal{C}_{\kappa,\mu}^{1-r})^\perp \cap \mathcal{S}^r(\Omega, \mu)$. Since evidently $\nu_1^r \in \mathcal{S}^r(\Omega, \mu)$, we may use (3.16) to verify this. Namely,

$$
\kappa_*(\nu_1^r \cdot \mu^{1-r}) = \kappa_*(\nu^r \cdot \mu^{1-r}) - \kappa_*(\kappa^*(\phi')\mu^r \cdot \mu^{1-r}) = \phi'\mu' - \phi'\kappa_*(\mu) = 0
$$

by (3.2) and the definition of ϕ'. This completes the proof. $\qquad\square$

4 Parametrized measure models and k-integrability

In this section, we shall now present our notion of a parametrized measure model.

Definition 4.1. (Parametrized measure model)
Let Ω be a measure space.

1. A *parametrized measure model* is a triple (M, Ω, p) where M is a (finite or infinite dimensional) Banach manifold and $p : M \to \mathcal{M}(\Omega) \subset \mathcal{S}(\Omega)$ is a C^1-map in the sense explained in Section 2.2.

2. The triple (M, Ω, p) is called a *statistical model* if it consists only of probability measures, i.e., such that the image of p is contained in $\mathcal{P}(\Omega)$.

3. We call such a model *dominated by* μ_0 if the image of p is contained in $\mathcal{S}(\Omega, \mu_0)$. In this case, we use the notation (M, Ω, μ_0, p) for this model.

Remark 4.1. Evidently, for the applications we have in mind, we are interested mainly in statistical models. However, we can take the point of view that $\mathcal{P}(\Omega)$ is the projectivisation of $\mathcal{P}(\Omega) = \mathbb{P}(\mathcal{M}(\Omega)\backslash 0)$ via rescaling. Thus, given a parametrized measure model (M, Ω, p), normalisation yields a statistical model (M, Ω, p_0) defined by

$$p_0(\xi) := \frac{p(\xi)}{\|p(\xi)\|_{TV}}.$$

which is again a C^1-map. Indeed, the map $\mu \mapsto \|\mu\|_{TV}$ on $\mathcal{M}(\Omega)$ is a C^1-map, being the restriction of the linear (and hence continuous) map $\mu \mapsto \int_\Omega d\mu$ on $\mathcal{S}(\Omega)$.
Observe that while $\mathcal{S}(\Omega)$ is a Banach space, the subsets $\mathcal{M}(\Omega)$ and $\mathcal{P}(\Omega)$ do not carry a canonical manifold structure.

If a parametrized measure model (M, Ω, μ_0, p) is dominated by μ_0, then there is a *density function* $p : \Omega \times M \to \mathbb{R}$ such that

$$p(\xi) = p(.; \xi)\mu_0. \tag{4.1}$$

From the context, i.e., from the number of arguments, it will be clear which map p is meant, whence we will denote both maps by the same symbol. Evidently, we must have $p(.; \xi) \in L^1(\Omega, \mu_0)$ for all ξ. In particular, for fixed ξ, $p(.; \xi)$ is defined only up to changes on a μ_0-null set.

Definition 4.2. (Regular density function)
Let (M, Ω, μ_0, p) be a parametrized measure model dominated by μ_0. We say that this model has a *regular density function* if the density function $p : \Omega \times M \to \mathbb{R}$ satisfying (4.1) can be chosen such that for all $V \in T_\xi M$ the partial derivative $\partial_V p(.; \xi)$ exists and lies in $L^1(\Omega, \mu_0)$.

Remark 4.2. The standard notion of a statistical model always assumes that it is dominated by some measure and has a regular density function (e.g. [3, §2 , p. 25], [4, §2.1], [19], [5, Definition 2.4]). In fact, the definition of a parametrized measure model or statistical model in [5, Definition 2.4] is equivalent to a parametrized measure model or statistical model with a regular density function in the sense of Definition 4.2.
Let us point out why the present notion is indeed more general. The formal definition of differentiability of p implies that for each C^1-path $\xi(t) \in M$ with $\xi(0) = \xi$, $\dot{\xi}(0) =: V \in T_\xi M$, the curve $t \mapsto p(.; \xi(t)) \in L^1(\Omega, \mu_0)$ is differentiable. This implies that there is a $d_\xi p(V) \in L^1(\Omega, \mu_0)$ such that

$$\left\| \frac{p(.; \xi(t)) - p(.; \xi)}{t} - d_\xi p(V)(.) \right\|_1 \xrightarrow{t \to 0} 0.$$

If this is a *pointwise* convergence, then $d_\xi p(V) = \partial_V p(.; \xi)$ is the partial derivative and whence, $\partial_V p(.; \xi)$ lies in $L^1(\Omega, \mu_0)$, so that the density function is regular.

However, in general convergence in $L^1(\Omega, \mu_0)$ does *not* imply pointwise convergence, whence there are parametrized measure models in the sense of Definition 4.1 without a regular density function, cf. Example 4.1.2below. Nevertheless, for simplicity we shall frequently use the notation $\partial_V p(\cdot; \xi)$ instead of $d_\xi p(V)(.)$, even if the density function is *not* regular.

By this convention, for a parametrized measure model (M, Ω, μ_0, p) we can describe its derivative in the direction of $V \in T_\xi M$ as

$$d_\xi p(V) = \partial_V p(.; \xi)\, \mu_0.$$

Example 4.1. 1. The family of normal distributions on $\mathbb{R}$

$$p(\mu, \sigma) := \frac{1}{\sqrt{2\pi\sigma}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})\, dx$$

is a statistical model with regular density function on the upper half plane $H = \{(\mu, \sigma) \ : \ \mu, \sigma \in \mathbb{R}, \sigma > 0\}$.

2. To see that there are parametrized measure models without a regular density function, consider the family of measures on $\Omega = (0, \pi)$

$$p(\xi) := \begin{cases} \left(1 + \xi\, (\sin^2(t - 1/\xi))^{1/\xi^2}\right)\, dt & \text{for } \xi \neq 0 \\[2ex] dt & \text{for } \xi = 0. \end{cases}$$

This model is dominated by the Lebesgue measure dt, with density function $p(t; \xi) = 1 + \xi\, (\sin^2(t - 1/\xi))^{1/\xi^2}$ for $\xi \neq 0$, $p(t; 0) = 1$. Thus, the partial derivative $\partial_\xi p$ does not exist at $\xi = 0$, whence the density function is not regular.

On the other hand, $p : \mathbb{R} \to \mathcal{M}(\Omega, dt)$ is differentiable in the above sense at $\xi = 0$ with $d_0 p(\partial_\xi) = 0$, so that (M, Ω, p) is a parametrized measure model in the sense of Definition 4.1. To see this, we calculate

$$\left\| \frac{p(\xi) - p(0)}{\xi} \right\|_1 = \left\| (\sin^2(t - 1/\xi))^{1/\xi^2}\, dt \right\|_1$$

$$= \int_0^\pi (\sin^2(t - 1/\xi))^{1/\xi^2}\, dt$$

$$= \int_0^\pi (\sin^2 t)^{1/\xi^2}\, dt \xrightarrow{\xi \to 0} 0.$$

which shows the claim. Here, we used the π-periodicity of the integrand for fixed ξ and dominated convergence in the last step.

Since for a parametrized measure model (M, Ω, p) the map p is C^1, it follows that its derivative yields a continuous map between the tangent spaces

$$dp : TM \longrightarrow T\mathcal{M}(\Omega) = \dot{\bigcup}_{\mu \in \mathcal{M}(\Omega)} \mathcal{S}(\Omega, \mu).$$

That is, for each tangent vector $V \in T_\xi M$, its differential $d_\xi p(V)$ is contained in $\mathcal{S}(\Omega, p(\xi))$ and hence dominated by $p(\xi)$.

Definition 4.3. Let (M, Ω, p) be a parametrized measure model. Then for each tangent vector $V \in T_\xi M$ of M, we define

$$\partial_V \log p(\xi) := \frac{d\{d_\xi p(V)\}}{dp(\xi)} \in L^1(\Omega, p(\xi)) \tag{4.2}$$

and call this the *logarithmic derivative of p at ξ in direction V.*

If such a model is dominated by μ_0 and has a regular density function p for which (4.1) holds, then we can calculate the Radon-Nikodým derivative as

$$\begin{aligned}
\frac{d\{d_\xi p(V)\}}{dp(\xi)} &= \frac{d\{d_\xi p(V)\}}{d\mu_0} \cdot \left(\frac{dp(\xi)}{d\mu_0}\right)^{-1} \\
&= \partial_V p(.; \xi)(p(.; \xi))^{-1} = \partial_V \log p(.; \xi),
\end{aligned}$$

where we use the convention $\log 0 = 0$. This justifies the notation in (4.2) even for models without a regular densitiy function.

For a parametrized measure model (M, Ω, p) and $k > 1$ we consider the map

$$p^{1/k} := \pi^{1/k} \circ p : M \longrightarrow \mathcal{S}^{1/k}(\Omega), \qquad \xi \longmapsto p(\xi)^{1/k}.$$

Since $\pi^{1/k}$ is continuous by Proposition 2.3, it follows that $p^{1/k}$ is continuous as well. Let us pretend for the moment that $p^{1/k}$ is a C^1-map, so that $d_\xi p^{1/k}(V) \in T_{p(\xi)^{1/k}} \mathcal{M}^{1/k}(\Omega) = \mathcal{S}^{1/k}(\Omega, p(\xi))$. In this case, because of $\pi^k \circ \pi^{1/k} = Id$, we have

$$p = \pi^k \circ p^{1/k},$$

whence by the chain rule and (2.14) we have for $\xi \in M$ and $V \in T_\xi M$

$$d_\xi p(V) = k \, p(\xi)^{1-1/k} \cdot \left(d_\xi p^{1/k}(V)\right).$$

Thus with (4.2) this implies

$$d_\xi p^{1/k}(V) = \frac{1}{k} \partial_V \log p(\xi) \, p^{1/k}(\xi) \in \mathcal{S}^{1/k}(\Omega, p(\xi)) \tag{4.3}$$

and hence, in particular, $\partial_V \log p(\xi) \in L^k(\Omega, p(\xi))$, and depends continuously on $V \in TM$. This motivates the following definition.

Definition 4.4. (k-integrable parametrized measure model)
A parametrized measure model (M, Ω, p) is called *k-integrable* for $k \geq 1$ if for all $\xi \in M$ and $V \in T_\xi M$ we have

$$\partial_V \log p(\xi) = \frac{d\{d_\xi p(V)\}}{dp(\xi)} \in L^k(\Omega, p(\xi)),$$

and moreover, the map

$$dp^{1/k} : TM \longrightarrow T\mathcal{S}^{1/k}(\Omega)$$

given in (4.3) is continuous. $dp^{1/k}$ is called the *formal derivative of $p^{1/k}$*. Furthermore, we call the model *∞-integrable* if it is k-integrable for all $k \geq 1$.

Since $p(\xi)$ is a finite measure, we have $L^k(\Omega, p(\xi)) \subset L^l(\Omega, p(\xi))$ for all $1 \leq l \leq k$. Thus, k-integrability implies l-integrability for all such l.

Remark 4.3.   1. By our previous discussion, a parametrized measure model (M, Ω, p) for which $p^{1/k}$ is a C^1-map is always k-integrable, and the derivative coincides with the formal derivative. However, it is not clear if there are k-integrable parametrized measure models for which $p^{1/k}$ is *not* a C^1-map.

    2. Observe that for parametrized measure models with a regular density function the notion of k-integrability coincides with that given in [5, Definition 2.4].

Definition 4.5. (Canonical n-tensor)
For $n \in \mathbb{N}$, the *canonical n-tensor* is the covariant n-tensor on $\mathcal{S}^{1/n}(\Omega)$, given by

$$L_\Omega^n(\nu_1, \ldots, \nu_n) = n^n \int_\Omega d(\nu_1 \cdots \nu_n), \qquad \text{where } \nu_i \in \mathcal{S}^{1/n}(\Omega). \tag{4.4}$$

For $n = 2$, the pairing $(.;.) : \mathcal{S}^{1/2}(\Omega) \times \mathcal{S}^{1/2}(\Omega) \to \mathbb{R}$ from (2.11) satisfies

$$(\nu_1; \nu_2) = \frac{1}{4} \, L_\Omega^2(\nu_1, \nu_2).$$

Since $(\nu; \nu) = \|\nu\|_2^2$ by (2.8), it follows:

$$\left(\mathcal{S}^{1/2}(\Omega), \frac{1}{4} \, L_\Omega^2\right) \text{ is a Hilbert space with norm } \|.\|_2.$$

The main purpose of defining the notion of k-integrability is that for a k-integrable model, there is a well defined pullback of the canonical n-tensor L_Ω^n via the map $p^{1/n}$ for all $n \leq k$. That is, we define for $V_1, \ldots, V_n \in T_\xi M$

$$\tau_{(M,\Omega,p)}^n(V_1, \ldots, V_n) \quad := \quad L_\Omega^n(d_\xi p^{1/n}(V_1), \ldots, d_\xi p^{1/n}(V_n))$$

$$= \quad \int_\Omega \partial_{V_1} \log p(\xi) \cdots \partial_{V_n} \log p(\xi) \, dp(\xi),$$

where the second line follows immediately from (4.3) and (4.4).

Example 4.2.   1. For $n = 1$, the canonical 1-form is given as

$$\tau_{(M,\Omega,p)}^1(V) := \int_\Omega \partial_V \log p(\xi) \, dp(\xi) = \partial_V \|p(\xi)\|.$$

Thus, it vanishes if and only if $\|p(\xi)\|$ is locally constant, e.g., if (M, Ω, p) is a *statistical model*.

2. For $n = 2$, $\tau_{(M,\Omega,p)}^2$ coincides with the *Fisher metric*

$$\mathfrak{g}^F(V, W)_\xi := \int_\Omega \partial_V \log p(\xi) \, \partial_W \log p(\xi) \, dp(\xi) \tag{4.5}$$

3. For $n = 3$, $\tau^3_{(M,\Omega,p)}$ coincides with the *Amari-Chentsov 3-symmetric tensor*

$$T^{AC}(V, W, X)_\xi := \int_\Omega \partial_V \, \log p(\xi) \, \partial_W \log p(\xi) \, \partial_X \log p(\xi) \, dp(\xi).$$

Remark 4.4. While the Fisher metric and the Amari-Chentsov tensor give an interpretation of $\tau^n_{(M,\Omega,p)}$ for $n = 2, 3$, we do not know of any statistical significance of $\tau^n_{(M,\Omega,p)}$ for $n \geq 4$. However, one might hope to get an interpretation of e.g. $\tau^4_{(M,\Omega,p)}$ as some kind of curvature tensor of a suitable connection. Moreover, in [14, p.212] the question is posed if there are other significant tensors on statistical manifolds, and the canonical n-tensors may be considered as natural candidates.

5 Parametrized measure models and sufficient statistics

Given a parametrized measure model (statistical model, respectively) (M, Ω, p) and a Markov kernel $K : \Omega \to \mathcal{P}(\Omega')$ which induces the Markov morphism $K_* : \mathcal{M}(\Omega) \to \mathcal{M}(\Omega')$ as in (3.6), we obtain another parametrized measure model (statistical model, respectively) (M, Ω', p') by defining $p'(\xi) := K_*(p(\xi))$. It is the purpose of this section to investigate the relation between these two models in more detail.

Theorem 5.1. *Let* (M, Ω, p), $K : \Omega \to \mathcal{P}(\Omega')$ *and* (M, Ω', p') *be as above, and suppose that* (M, Ω, p) *is k-integrable for some $k \geq 1$. Then the following hold.*

1. (M, Ω', p') *is also k-integrable.*

2. *If K is congruent w.r.t. some statistic $\kappa : \Omega' \to \Omega$, then*

$$\|\partial_V \log p(\xi)\|_k = \|\partial_V \log p'(\xi)\|_k \qquad \text{for all } V \in T_\xi M.$$

3. *If K is induced by a statistic $\kappa : \Omega \to \Omega'$, then*

$$\left(\partial_V \log p(\xi) - \kappa^*(\partial_V \log p'(\xi))\right) \cdot p(\xi)^{1/k} \in (\mathcal{C}^{1-1/k}_{\kappa,p(\xi)})^\perp.$$

Proof. Since K_* is the restriction of a bounded linear map, it is obvious that $p' : M \to \mathcal{M}(\Omega')$ is again differentiable, and in fact,
$$d_\xi p'(V) = K_*(d_\xi p(V)).$$
for all $V \in T_\xi M$, $\xi \in M$.

Let us now assume that K is κ-congruent w.r.t. a statistic $\kappa : \Omega' \to \Omega$. Applying Proposition 3.1 (with $\Omega_1 := \Omega'$ and $\Omega_2 := \Omega$) to $\mu' := p(\xi) \in \mathcal{M}(\Omega)$ and $\mu := K_*(\mu') = p'(\xi) \in \mathcal{M}(\Omega')$, it follows that

$$\begin{aligned}
d_\xi p'(V) &= K_*(d_\xi p(V)) = K_*(\partial_V \log p(\xi) \mu') \\
&= \kappa^*(\partial_V \log p(\xi)) \mu = \kappa^*(\partial_V \log p(\xi)) p'(\xi),
\end{aligned}$$

and hence, for any κ-congruent Markov morphism K, we have

$$\partial_V \log p'(\xi) = \frac{d\{d_\xi p'(V)\}}{dp'(\xi)} = \kappa^*(\partial_V \log p(\xi)),$$

25

Thus,

$$
\begin{aligned}
\|\partial_V \log p'(\xi)\|_k^k &= \int_{\Omega'} |\partial_V \log p'(\xi)|^k \, dp'(\xi) \\
&= \int_{\Omega'} \kappa^*(|\partial_V \log p(\xi)|^k) \, dp'(\xi) = \int_\Omega |\partial_V \log p(\xi)|^k \, dp(\xi) \\
&= \|\partial_V \log p(\xi)\|_k^k,
\end{aligned}
$$

using that $\kappa_*(p'(\xi)) = \kappa_*(K_*(p(\xi)) = p(\xi)$ by the κ-congruency of K, which shows the second statement.

Next, suppose that K is induced by a statistic $\kappa : \Omega \to \Omega'$, i.e., $p'(\xi) = \kappa_*(p(\xi))$ for all $\xi \in M$. If (M, Ω, p) is k-integrable, then by Proposition 3.3, we may decompose

$$
\mathcal{S}^{1/k}(\Omega, p(\xi)) \ni \partial_V \log p(\xi) \cdot p(\xi)^{1/k} = \kappa^*(\phi') \cdot p(\xi)^{1/k} + \nu^{1/k},
$$

where $\phi' \in L^k(\Omega', p'(\xi))$ and $\nu^{1/k} \in (\mathcal{C}_{\kappa,p(\xi)}^{1-1/k})^\perp$. Thus,

$$
\begin{aligned}
d_\xi p'(V) &= \kappa_*(\partial_V \log p(\xi) \cdot p(\xi)) = \kappa_*((\partial_V \log p(\xi) \cdot p(\xi)^{1/k}) \cdot p(\xi)^{1-1/k}) \\
&= \kappa_*(\kappa^*(\phi') \cdot (p(\xi))) + \underbrace{\kappa_*(\nu^{1/k} \cdot p(\xi)^{1-1/k})}_{=0} \\
&= \phi' p'(\xi),
\end{aligned}
$$

using (3.2) and (3.16). That is, $\phi' = \partial_V \log p'(\xi)$, and this shows the third statement.

Finally, to see the first statement, observe that by Theorem 3.2, every Markov kernel is the composition of a statistic and a congruent Markov kernel. Since from the second and the third statement it follows that both a statistic and a congruent Markov kernel preserve k-integrability, the general case follows. $\qquad\square$

Definition 5.1. (Sufficient statistic)
Let (M, Ω, p) be a parametrized measure model. Then $\kappa : \Omega \to \Omega'$ is called a *sufficient statistic for* p if there is a $\mu \in \mathcal{M}(\Omega)$ such that
$$
p(\xi) = \phi'(\kappa(\cdot); \xi)\mu
$$
for some $\phi'(\cdot; \xi) \in L^1(\Omega', \mu')$. In this case,

$$
p'(\xi) = \kappa_* p(\xi) = \phi'(\cdot; \xi)\mu',
$$

where $\mu' = \kappa_*(\mu)$.

Evidently, by (3.5) this is equivalent to saying that $p(\xi) \subset \mathcal{C}_{\kappa,\mu}$ for all $\xi \in M$ and some fixed measure $\mu \in \mathcal{M}(\Omega)$. If in addition (M, Ω, p) is k-integrable, then (3.13) implies that $p(\xi)^{1/k} \in \mathcal{C}_{\kappa,\mu}^{1/k}$. We now can show the following

Theorem 5.2. *(Monotonicity theorem) (cf. [5], [4])*
Let (M, Ω, p) be a 2-integrable parametrized measure model, let $K : \Omega \to \mathcal{P}(\Omega')$ be a Markov kernel, so that the induced parametrized measure model (M, Ω', p') with $p'(\xi) = K_*(p(\xi))$ is also

2-integrable by Theorem 5.1. Moreover, let $\mathfrak{g}$ and $\mathfrak{g}'$ denote the Fisher metric of (M, Ω, p) and (M, Ω', p'), respectively. Then

$$\mathfrak{g}(V, V) \geq \mathfrak{g}'(V, V) \qquad \text{for all } V \in T_\xi M \text{ and } \xi \in M. \tag{5.1}$$

Moreover, if K is induced by a statistic $\kappa : \Omega \to \Omega'$ then equality in (5.1) holds iff $\partial_V \log p(\xi) = \kappa^*(\partial_V \log p'(\xi))$. In particular, if $p(\xi) = p(\cdot; \xi)\mu_0$ with regular and positive density function $p : M \times \Omega \to (0, \infty)$, and M is connected, then equality in (5.1) holds for all V if and only if κ is a sufficient statistic for the model (M, Ω, p).

Remark 5.1. 1. The difference $\mathfrak{g}(V, V) - \mathfrak{g}'(V, V) \geq 0$ is called the *information loss of the model* (M, Ω, p) *under* κ. Thus, the interpretation of the monotonicity theorem is that every statistic produces some (non-negative) information loss which vanishes if and only if this statistic is sufficient.

2. Theorem 5.1, 2. already implies that a congruent Markov morphism preserves the Fisher metric. Thus, due to the decomposition of Markov kernels from Theorem 3.2, it suffices to show that statistics decrease the Fisher metric unless the statistic is sufficient.

3. Note that our approach allows to prove the monotonicity theorem with no further assumption on the model (M, Ω, p). In order for (5.1) to hold we can even work with arbitrary Markov kernels, not just statistics, and there is no condition on the statistic κ. In particular, we do not need to assume that Ω is a topological space with its Borel σ-algebra, nor do we need to assume the existence of transversal measures of the map κ which are assumptions made in the literatur (e.g. [4, Theorem 2.1] or [5, Theorem 3.11]). In this sense, our statement generalizes these versions of the monotonicity theorem, as it even covers a statistic as in Example 3.2.

Proof. As before, Theorem 3.2 allows us to reduce the proof of the monotonicity to the case of congruent Markov morphisms and to Markov morphisms induced by a statistic.
Observe that by the definition of the Fisher metric in (4.5) it follows that

$$\mathfrak{g}^F(V, V) = \|\partial_V \log p(\xi)\|_2^2, \qquad \mathfrak{g}'^F(V, V) = \|\partial_V \log p'(\xi)\|_2^2,$$

where these norms are taken in $L^2(\Omega, p(\xi))$ and $L^2(\Omega', p'(\xi))$, respectively. Thus, if K is congruent w.r.t. some statistic, then Theorem 5.1, 2. implies that $\mathfrak{g}^F(V, V) = \mathfrak{g}'^F(V, V)$ in this case.
If K is induced by a statistic $\kappa : \Omega \to \Omega'$, let us use the decomposition (3.15) for $r = 1/2$ to write

$$\mathcal{S}^{1/2}(\Omega, p(\xi)) \ni \partial_V \log p(\xi) = \kappa^*(\phi')p(\xi) + \nu \in \mathcal{C}_{\kappa, p(\xi)}^{1/2} \oplus (\mathcal{C}_{\kappa, p(\xi)}^{1/2})^\perp,$$

where the two summands are orthogonal in the Hilbert space $\mathcal{S}^{1/2}(\Omega, p(\xi))$. Thus,

$$\mathfrak{g}^F(V, V) = \|\partial_V \log p(\xi)\|_2^2 = \|\kappa^*(\phi')\|_2^2 + \|\nu\|_2^2 \geq \|\phi'\|_2^2,$$

where the norms are taken in the Hilbert spaces $L^2(\Omega, p(\xi))$, $L^2(\Omega', p'(\xi))$ and $\mathcal{S}^{1/2}(\Omega, p(\xi))$, respectively. By Theorem 5.1, 3. this implies that $\phi' = \partial_V \log p'(\xi)$, so that

$$\|\phi'\|_2^2 = \|\partial_V \log p'(\xi)\|_2^2 = \mathfrak{g}'^F(V; V),$$

so that the estimate $\mathfrak{g}^F(V, V) \geq \mathfrak{g}'^F(V; V)$ follows, with equality iff $\partial_V \log p(\xi) = \kappa^*(\partial_V \log p'(\xi))$.

If (M, Ω, p) is given by a regular positive density function $p(\xi) = p(\cdot; \xi)\mu_0$, then $\log p(\cdot; \xi)$ and $\log p'(\cdot; \xi)$ are indeed well defined differentiable functions on $M \times \Omega$ and $M \times \Omega'$, respectively. In particular, $\kappa^*(\partial_V \log p'(\xi)) = \partial_V(\log \kappa^* p'(\xi))$, so that equality in (5.1) holds iff

$$\partial_V \log \frac{p(\cdot; \xi)}{\kappa^* p'(\cdot; \xi)} = \partial_V(\log p(\cdot; \xi) - (\log \kappa^* p'(\cdot; \xi))) = 0.$$

If M is connected, then this is the case for *all* $V \in TM$ iff $h(\cdot) := \frac{p(\cdot; \xi)}{\kappa^* p'(\cdot; \xi)}$ is positive and does not depend on $\xi \in M$. Thus, setting $\tilde{\mu}_0 := h\mu_0$ this implies that

$$p(\xi) = \kappa^*(p'(\xi))\tilde{\mu}_0,$$

showing that this happens iff κ is a sufficient statistic for (M, Ω, p). $\qquad\square$

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, J.Jost, H.V.Lê, L.Schwachhöfer, Information geometry and sufficient statistics, Probability Theory and Related Fields 162 no. 1-2, (2015), 327-364.

[6] M. Bauer, M. Bruveris, P. Michor, Uniqueness of the Fisher-Rao metric on the space of smooth densities, arXiv:1411.5577 (2014)

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

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

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

[10] N. Chentsov, Statistical decision rules and optimal inference, Moscow, Nauka, 1972 (in Russian), English translation in: Translation of Math. Monograph 53, AMS, Providence, 1982.

[11] B. Efron, Defining the curvature of a statistical problem (with applications to second order efficiency), with a discussion by C. R. Rao, Don A. Pierce, D. R. Cox, D. V. Lindley, Lucien LeCam, J. K. Ghosh, J. Pfanzagl, Niels Keiding, A. P. Dawid, Jim Reeds and with a reply by the author, Ann. Statist. 3 (1975), 1189-1242.

[12] H. Jeffreys, An invariant form for the prior probability in estimation problems, Proc. Roy. Soc. London. Ser. A. 186, 453-461, 1946.

[13] S. Lang, Introduction to Differentiable Manifolds, 2nd ed., Universitext, Springer (2002)

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

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

[16] 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.

[17] G. Pistone, Nonparametric information geometry. Geometric science of information, Lecture Notes in Comput. Sci., 8085, Springer, Heidelberg (2013)

[18] 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.

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