Parametrized measure models and a generalization of Chentsov's Theorem

Lorenz Schwachhöfer

technische universität dortmund

on the occasion of the 80th birthday of Prof. Shun-ichi Amari

IGAIA IV, Liblice, Czech Republic
June 16, 2016

(based on joint work with Nihat Ay, Jürgen Jost, Hông Vân Lê)

Outline of the talk

1. Statistical models
    - Amari's Definition of statistical models
    - Canonical tensors and k-integrability
2. Geometrization of the space of measures on Ω
    - Construction of Pistone-Sempi
    - Tangent cones and tangent fibrations
    - Definition of parametrized measure models
3. Statistics and information loss
    - Sufficient statistics
4. Invariant tensors and Chentsov's theorem
    - Invariant tensors
    - Theorems of Chentsov and Campbell
    - Generalization of the Chentsov-Campbell Theorem

Statistical models

What is a statistical model or a parametrized measure model?

Heuristicly speaking, a statistical model is a family $p(\xi)_{\xi \in M}$ of probability measures on a fixed sample space Ω which vary "differentiably" with $\xi \in M$, where the parameter space M is a (finite dimensional) manifold. More precisely, we make the following definition:

Definition (Amari, 1980)

Let Ω be a measure space.
A **parametrized measure model with a regular density function and reference measure** μ_0 is a family of measures given by

$$p(\xi) = \bar{p}(\xi; \omega)\mu_0,$$

where $\bar{p} > 0$ is differentiable in the ξ-variable, and $\partial_v \log \bar{p}(\xi; \cdot)$ is integrable w.r.t. $p(\xi)$, i.e., $\partial_v \log \bar{p}(\xi; \cdot) \in L^1(\Omega, p(\xi))$.

We call such a model **statistical** if all $p(\xi)$ are probability measures, i.e., if $\|p(\xi)\| := p(\xi)(\Omega) = 1$.

Remark:
We can always get a statistical model from a parametrized measure model by normalization:
If $p(\xi)_{\xi \in M}$ is a parametrized measure model, then we obtain a statistical model by setting

$$p_0(\xi) := \frac{p(\xi)}{\|p(\xi)\|}, \qquad \text{where } \|p(\xi)\| := p(\xi)(\Omega)$$

(i.e. use *projectivization of a finite measure*). But sometimes it is more convenient to work without this normalization.

Canonical tensors and k-integrability

Now define the **Canonical Tensors** for a parametrized measure model $p(\xi)_{\xi \in M}$ on Ω to be the following symmetric tensor fields:

- $\tau_\xi^1(v) := \int_\Omega \partial_v \log \bar{p}(\xi;\cdot)\, dp(\xi) =$
  $\partial_v \int_\Omega \bar{p}(\xi;\cdot) d\mu_0 = \partial_v \|p(\xi)\|.$
  Thus, $\tau^1 \equiv 0$ for statistical models (i.e., if $\|p(\xi)\| \equiv 1$).

- The **Fisher metric** is a symmetric 2-tensor:

$$\mathfrak{g}_F(v,w) = \tau_\xi^2(v,w) := \int_\Omega \partial_v \log \bar{p}(\xi;\cdot) \partial_w \log \bar{p}(\xi;\cdot)\, dp(\xi)$$

- The **Amari-Chentsov tensor** is a symmetric 3-tensor:

$$\mathbf{T}^{AC}(v,w,u) = \tau_\xi^3(v,w,u)$$
$$:= \int_\Omega \partial_v \log \bar{p}(\xi;\cdot)\partial_w \log \bar{p}(\xi;\cdot)\partial_u \log \bar{p}(\xi;\cdot)\; dp(\xi)$$

- We can generalize this to arbitrary degrees. The **canonical** *n*-**tensor**:

$$\tau_\xi^n(v_1,\dots,v_n) := \int_\Omega \partial_{v_1} \log \bar{p}(\xi;\cdot)\cdots\partial_{v_n} \log \bar{p}(\xi;\cdot)\; dp(\xi)$$

Observe that τ^n is only defined if $\partial_v \log \bar{p}(\xi;\cdot) \in L^n(\Omega, p(\xi))$.

This motivates the following definition:

k-integrability (cf. AJLS 2015)

A parametrized measure model (M, p, Ω) is called k-**integrable** if $\partial_v \log \bar{p}(\xi; \cdot) \in L^k(\Omega, p(\xi))$ for all $v \in T_\xi M$ (plus some continuity condition).

Example

Let $\Omega := (0,1)$, dt the the Lebesgue measure and $\alpha \in (0,1)$ be fixed. For $\xi \in (0, \infty)$ define

$$p(\xi) = \bar{p}(t; \xi)dt := (2 + sin(\xi t^{-\alpha}))dt.$$

Then $\partial_\xi \log \bar{p} = O(t^{-\alpha})$, whence (M, p, Ω) **is k-integrable iff $k < \alpha^{-1}$.**

In particular, the Fisher metric and the AC-tensor of this model is defined iff $\alpha < 1/2$ and $\alpha < 1/3$, respectively.

Geometrization of the space of measures on Ω

Question: Can we describe the geometry of parametrized measure models $(p(\xi))_{\xi \in M}$ on Ω by describing some geometric structures on the set of measures on Ω?

We define the following sets of measures on Ω:

$$\begin{aligned} \mathcal{P}(\Omega) &:= \{\text{probability meas.}\} \\ \mathcal{M}(\Omega) &:= \{\text{finite measures}\} \\ \mathcal{S}(\Omega) &:= \{\text{signed finite meas.}\} \end{aligned}$$

$$\left.\begin{array}{r} \mathcal{P}_+(\Omega, \mu_0) \\ \mathcal{P}(\Omega, \mu_0) \\ \mathcal{M}_+(\Omega, \mu_0) \\ \mathcal{M}(\Omega, \mu_0) \\ \mathcal{S}(\Omega, \mu_0) \end{array}\right\} \text{meas. dominated by } \mu_0$$

Observe that $\mathcal{S}(\Omega)$ and $\mathcal{S}(\Omega, \mu_0)$ are **Banach spaces** w.r.t. to the total variation, $\|\mu\|_{TV} := |\mu|(\Omega)$.

Construction of Pistone-Sempi

Brief review of the idea of Pistone-Sempi (1995):

$$\mathcal{M}_+(\Omega, \mu_0) \overset{\cong}{\leftrightarrow} \mathfrak{L}(\Omega, \mu_0) := \{f : \Omega \to \mathbb{R} \mid \int_\Omega e^f \mu_0 < \infty\}.$$

Then $\mathfrak{L}(\Omega, \mu_0)$ can be given the structure of a **Banach manifold**, which decomposes into disjoint connected components, each of which is canonically embedded as an open convex set into some Banach space. Thus, there is a canonical affine structure, and much more....... cf. talk from last Monday.

Advantage: Very strong and beautiful geometric structure on $\mathcal{M}_+(\Omega, \mu_0) \cong \mathfrak{L}(\Omega, \mu_0)$.

Disadvantage: The condition that $(p(\xi))_{\xi \in M} \in \mathcal{M}_+(\Omega, \mu_0)$ is differentiable (or even continuous) w.r.t. this Banach manifold structure is very strong. ($\to$ *e*-continuity)

Example. *e*-continuity of the map $\xi \mapsto p(\xi) \in \mathcal{M}_+(\Omega, \mu_0)$ implies that the model is k-integrable for all k.
But not even this suffices: e.g. the family $(p(\xi))_{\xi \in \mathbb{R}} \in \mathcal{M}_+((0,1), dt)$ given by

$$p(\xi) := \exp(-\xi^2 (\log t)^2) \; dt$$

is k-integrable for all k, but not *e*-continuous at $\xi = 0$ and hence not differentiable as a map into $\mathcal{M}_+((0,1), dt)$.

Tangent cones and tangent fibrations

Definition

Let V be a topological vector space and $X \subset V$ an arbitrary subset, $x \in X$.
$v \in V$ is called a (both sided) *tangent vector of X in x*, if there is a differentiable curve $c : (-\varepsilon, \varepsilon) \to X$ with $c(0) = x$, $\dot{c}(0) = v$.
$T_x X := \{\text{tangent vectors in x}\} \subset X$ is called the *(double) tangent cone of X in x*.
Let $TX := \dot{\bigcup}_{x \in X} T_x X \subset X \times V$. Then the canonical projection $TX \to X$ onto the first component is called the *tangent fibration of X*.

We apply this to $V := \mathcal{S}(\Omega)$ and $X := \mathcal{P}(\Omega)$ or $X := \mathcal{M}(\Omega)$, respectively.

We get the following result:

Proposition (AJLS 2015)

For $\mu \in \mathcal{M}(\Omega) \subset \mathcal{S}(\Omega)$ we have

$$T_\mu \mathcal{M}(\Omega) = \mathcal{S}(\Omega, \mu) = \{\phi\mu \mid \phi \in L^1(\Omega, \mu)\}$$

For $\mu \in \mathcal{P}(\Omega) \subset \mathcal{S}(\Omega)$ we have

$$T_\mu \mathcal{P}(\Omega) = \mathcal{S}_0(\Omega, \mu) = \{\phi\mu \mid \phi \in L^1(\Omega, \mu), E_\mu(\phi) = 0\}.$$

Definition of parametrized measure models

Definition (AJLS, 2015)

Let Ω be a measure space.
A **parametrized measure model (statistical model)** is a map $p : M \to \mathcal{M}(\Omega)$ ($p : M \to \mathcal{P}(\Omega)$), which is differentiable when regarded as a map between Banach manifolds $M \to S(\Omega)$.

Thus, the differential is a bounded linear map

$$d_\xi p : T_\xi M \to T_{p(\xi)}\mathcal{M}(\Omega) = \mathcal{S}(\Omega, p(\xi))$$
$$(d_\xi p : T_\xi M \to T_{p(\xi)}\mathcal{P}(\Omega) = \mathcal{S}_0(\Omega, p(\xi)))$$

That is: The directional derivatives $\partial_v p$ are dominated by $p(\xi)$ for $v \in T_\xi M$.

Advantages:

- The previous definitions of statistical models of Amari and of Pistone-Sempi are parametrized measure models in this sense.
- If there is a reference measure μ_0 dominating all $p(\xi)$ (which is always the case if M is finite dimensional), so that $p(\xi) = \bar{p}(\xi; \cdot)\mu_0$ for some density function $\bar{p} : M \times \Omega \to \mathbb{R}$, **we do not assume that $\bar{p} > 0$ a.e.!!!**.
- We do not assume that the measures $p(\xi)$ have the same null sets, but inequivalent measures are allowed.

BUT: If $\log \bar{p}$ is not defined, can we define the canonical forms τ^n?

Suppose that $p(\xi) = \bar{p}(\xi;\cdot)\mu_0$ with $\bar{p} > 0$. Then for $v \in T_\xi M$,

$$\frac{d(\partial_v p(\xi))}{dp(\xi)} = \frac{\partial_v \bar{p}(\xi;\cdot)}{\bar{p}(\xi;\cdot)} = \partial_v \log \bar{p}(\xi;\cdot),$$

and as $\partial_v p(\xi)$ is dominated by $p(\xi)$, this expression makes always sense.

Thus, we may **define**

$$\partial_v \log \bar{p}(\xi,\cdot) := \frac{d(\partial_v p(\xi))}{dp(\xi)} \in L^1(\Omega, p(\xi))$$

for any parametrized measure model, and this makes sense even though $\log \bar{p}$ does not!!!

In particular, the canonical tensors

$$\tau_\xi^n(v_1, \ldots, v_n) := \int_\Omega \partial_{v_1} \log \bar{p}(\xi; \cdot) \cdots \partial_{v_n} \log \bar{p}(\xi; \cdot)\, dp(\xi)$$

are still well defined for this notion of parametrized measure model, if $\partial_v \log \bar{p}(\xi; \cdot) \in L^n(\Omega, p(\xi))$, and the condition of k-integrability is well defined.

Question: What is a good interpretation of the k-integrability condition $\partial_v \log \bar{p}(\xi; \cdot) \in L^k(\Omega, p(\xi)) \subset L^1(\Omega, p(\xi))$?

Consider the following rewriting of the canonical tensor τ^n in the presence of a regular density function: $p(\xi) = \bar{p}(\xi;\omega)\mu_0$.

$$
\begin{aligned}
\tau^n(v_1,\ldots,v_n) &= \int_\Omega \partial_{v_1}\log\bar{p}\cdots\partial_{v_n}\log\bar{p}\; dp(\xi) \\
&= \int_\Omega \frac{\partial_{v_1}\bar{p}}{\bar{p}}\cdots\frac{\partial_{v_n}\bar{p}}{\bar{p}}\;\bar{p}\;d\mu_0 \\
&= \int_\Omega \frac{\partial_{v_1}\bar{p}}{\bar{p}^{1-1/n}}\cdots\frac{\partial_{v_n}\bar{p}}{\bar{p}^{1-1/n}}\;d\mu_0 \\
&= n^n\int_\Omega \partial_{v_1}\sqrt[n]{\bar{p}}\cdots\partial_{v_n}\sqrt[n]{\bar{p}}\;d\mu_0 \\
&= n^n\int_\Omega d(\partial_{v_1}\sqrt[n]{\bar{p}\mu_0}\cdots\partial_{v_n}\sqrt[n]{\bar{p}\mu_0}) \\
&= n^n\int_\Omega d(\partial_{v_1}p(\xi)^{1/n}\cdots\partial_{v_n}p(\xi)^{1/n}).
\end{aligned}
$$

Can we make sense out of *n*-th roots of measures?

Answer: YES, WE CAN!

For $0 < r \leq 1$, we can define Banach spaces $(\mathcal{S}^r(\Omega), \| \cdot \|_r)$, whose elements may be interpreted as "r-th powers of a measure". They have the subsets $\mathcal{P}^r(\Omega) \subset \mathcal{M}^r(\Omega) \subset \mathcal{S}^r(\Omega)$ of r-th powers of (probability) measures.

We can work with these quite intuitively:

- There is a multiplication map $\cdot : \mathcal{S}^r(\Omega) \times \mathcal{S}^s(\Omega) \to \mathcal{S}^{r+s}(\Omega)$, which is bilinear and bounded, if $r, s, r+s \in (0,1]$.
- There is a power raising map $\pi^k : \mathcal{S}^r(\Omega) \to \mathcal{S}^{kr}(\Omega)$ for all $r, kr \in (0,1]$. This map is continuous for all $k > 0$ and differentiable for $k \geq 1$.

We now say that a (general) parametrized measure model $p : M \to \mathcal{M}(\Omega)$ is k-integrable, if $p^{1/k} : M \to \mathcal{M}^{1/k}(\Omega) \subset \mathcal{S}^{1/k}(\Omega)$ is (weakly) differentiable.

Then the equation

$$\tau^n(v_1, \ldots, v_n) = n^n \int_\Omega d(\partial_{v_1} p(\xi)^{1/n} \cdots \partial_{v_n} p(\xi)^{1/n})$$

means the following:
If we define the canonical n-form on $\mathcal{S}^{1/n}(\Omega)$ as

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

then

$$\tau^n = (p^{1/n})^*(L^n_\Omega).$$

That is: **the canonical forms are pull-backs of the natural forms L^n_Ω on $\mathcal{S}^{1/n}(\Omega)$ via $p^{1/n} : M \to \mathcal{M}^{1/n}(\Omega)$.**

Statistics and information loss

A *statistic* is a measurable map $\kappa : \Omega \to \Omega'$ between measure spaces Ω and Ω'. (We only consider deterministic maps here, but all results presented here are also true if we consider Markov kernels, i.e. "noisy statistics").

Such a statistic induces a bounded lineas map $\kappa_* : \mathcal{S}(\Omega) \to \mathcal{S}(\Omega')$ by

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

Then κ_* maps $\mathcal{M}(\Omega)$ to $\mathcal{M}(\Omega')$ and $\mathcal{P}(\Omega)$ to $\mathcal{P}(\Omega')$.

In particular, if $p : M \to \mathcal{M}(\Omega)$ is a parametrized measure model, then so is $p' : M \to \mathcal{M}(\Omega')$ with $p'(\xi) = \kappa_* p(\xi)$.

Sufficient statistics

Heuristic Definition by Fisher:

> *... the criterion of sufficiency [...] requires that the whole of the relevant information supplied by a sample shall be contained in the statistics calculated.* (Fisher, 1922)

The standard definition of sufficient statistic:

Definition Sufficient statistic (Fisher-Neyman)

Let $(p(\xi))_{\xi \in M}$ be a statistical model. A statistic $\kappa : \Omega \to \Omega'$ is called *sufficient for the model*, if there is a measure $\mu \in \mathcal{P}(\Omega)$ such that

$$p(\xi) = \bar{p}'(\xi; \kappa(\cdot))\mu, \qquad \bar{p}' : M \times \Omega' \to \mathbb{R}^+.$$

In this case, $p'(\xi) = \kappa_*(p(\xi)) = \bar{p}'(\xi; \cdot)\mu'$, where $\mu' := \kappa_*(\mu) \in \mathcal{P}(\Omega')$. Thus, p' yields the same information as p.

Proposition (AJLS 2015)

Let $(p(\xi))_{\xi \in M}$ be a k-integrable parametrized measure model on Ω and $\kappa : \Omega \to \Omega'$ a statistic.
Then the induced model $(p'(\xi))_{\xi \in M}$ on Ω' with $p'(\xi) := \kappa_*(p(\xi))$ is also k-integrable.

E.g. if $(p(\xi))_{\xi \in M}$ is 2-integrable, so that the Fisher metric $\mathfrak{g} = \tau^2$ is defined, then so is $(p'(\xi))_{\xi \in M}$, whence the Fisher metric $\mathfrak{g}' = {\tau'}^2$ of p' is also defined.

Monotonicity Theorem (AJLS (2015))

Let $(p(\xi))_{\xi \in M}$ be a k-integrable parametrized measure model (statistical model) on Ω for $k \geq 2$, let $\kappa : \Omega \to \Omega'$ be a statistic, and let $p' := \kappa_* p$ as before, so that p' is also k-integrable.
Then the Fisher metrics $\mathfrak{g}_F, \mathfrak{g}'_F$ of p, p' satisfy the **monotonicity condition**

$$\mathfrak{g}_F(v, v) - \mathfrak{g}'_F(v, v) \geq 0.$$

for all $v \in T_\xi M$.

In Amari's book (2000), this theorem is shown if Ω, Ω' are manifolds and κ is differentiable or at least admits transversal measures. We show this result without any assumptions on Ω or κ.
The quantity $\mathfrak{g}_F(v, v) - \mathfrak{g}'_F(v, v)$ is the information loss of the model under κ.

Observe that $\mathfrak{g}_F(v,v) = \|\partial_v \log p(\xi;\cdot)\|_2^2$, and $\mathfrak{g}'_F(v,v) = \|\partial_v \log p'(\xi;\cdot)\|_2^2$ where the norms are taken in $L^2(\Omega, p(\xi))$ and $L^2(\Omega', p'(\xi))$, respectively.
We have the following generalization:

Theorem (AJLS (2015))

Let $(p(\xi))_{\xi\in M}$ be a k-integrable parametrized measure model (statistical model) on Ω for $k \geq 2$, let $\kappa : \Omega \to \Omega'$ be a statistic, and let $p' := \kappa_* p$ as before, so that p' is also k-integrable.
Then for all $l \in (1,k]$ and $v \in T_\xi M$ we have

$$\|\partial_v \log p\|_l^l - \|\partial_v \log p'\|_l^l \geq 0.$$

Moreover, equality in this equation holds either for **no** $l \in (1,k]$ or for **all** $l \in (1,k]$.

As in the case $k = 2$ we can define the l-th order information loss of the model under κ to be the quantity

$$\|\partial_v \log p\|_l^l - \|\partial_v \log p'\|_l^l \geq 0$$

That is, the information loss either vanishes for *no* l or for *all* l.

Theorem (AJLS 2016)

Let $p : M \to \mathcal{M}(\Omega)$, $\kappa : \Omega \to \Omega'$ and $p' := \kappa_* p$ as above, and suppose that $p(\xi) = \bar{p}(\xi; \cdot)\mu_0$ with $\bar{p}(\xi; \cdot) > 0$ differentiable. Then the information loss of the model under κ vanishes **iff κ is sufficient for the model**.

Again, this was shown in Amari's book already in the case of manifolds Ω, Ω' and differentiable κ.

The condition $\bar{p}(\xi; \cdot) > 0$ is crucial in this Theorem:
Example: $\Omega := (-1,1) \times (0,1)$, $\Omega' := (-1,1)$, $\kappa : \Omega \to \Omega'$ canonical projection. Define $p(\xi)_{\xi \in \mathbb{R}} = \bar{p}(\xi; s, t)$ ds dt with

$$p(s,t;\xi) := \begin{cases} h(\xi) & \text{for } \xi \geq 0 \text{ and } s \geq 0 \\ 2h(\xi)t & \text{for } \xi < 0 \text{ and } s \geq 0 \\ 1 - h(\xi) & \text{for } s < 0 \end{cases},$$

where $h(\xi) := \exp(-|\xi|^{-1})$ for $\xi \neq 0$ and $h(0) := 0$. Then $p(\xi)$ is a probability measure, and

$$p'(\xi) := \kappa_* p(\xi) = \bar{p}'(s;\xi)\, ds, \bar{p}'(s;\xi) = \begin{cases} (1 - h(\xi)) & s < 0 \\ h(\xi) & s \geq 0 \end{cases}.$$

Then for all $k \geq 1$,

$$\|\partial_\xi \log \bar{p}(s,t;\xi)\|_k = \|\partial_\xi \log \bar{p}'(s;\xi)\|_k,$$

whence there is no information loss, **but κ is not sufficient**.

What is going on?

κ is a sufficient statistic when restricting to $\xi > 0$ or to $\xi < 0$:

$$p(\xi) = \begin{cases} \bar{p}'(s;\xi)\ dsdt & \text{for } \xi > 0 \\ \bar{p}'(s;\xi)(\chi_{(-1,0)}(s) + 2t\chi_{[0,1)}(s))\ dsdt, & \text{for } \xi < 0 \end{cases}$$

But at $\xi = 0$, we have $\bar{p}(s,t;0) = 0$ for $s > 0$, so that $\bar{p} > 0$ is violated.

Invariant tensors

Definition

Suppose for every measure space Ω, Θ_Ω is a covariant n-tensor on $\mathcal{M}^{1/k}(\Omega)$ for some fixed $k > 1$. Then the family (Θ_Ω) is called *invariant under sufficient statistics*, if the following holds:
For any k-integrable parametrized measure model $(p(\xi))_{\xi \in M}$ on Ω and any statistic $\kappa : \Omega \to \Omega'$ which is sufficient for the model we have

$$(p^{1/k})^*(\Theta_\Omega) = ({p'}^{1/k})^*(\Theta_{\Omega'}).$$

Examples: The canonical tensor fields L^n_Ω on $\mathcal{M}^{1/n}(\Omega)$ are invariant. Indeed, the pullback $(p^{1/n})^*(\Theta_\Omega)$ is the canonical form τ^n.
As $\tau^n(v, \ldots, v) = \|\partial_v \log p\|^n_n$, it follows that this is invariant under sufficient statistics by the monotonicity theorem.

Further examples:

We can obtain further invariant tensors by taking tensor products of the canonical tensors, taking linear combinations where the coefficients are continuous functions of $\|p(\xi)\| = p(\xi)(\Omega)$. For instance, we can consider the following 4-tensors:

$$\begin{aligned}\Theta_\mu(v_1, v_2, v_3, v_4) =& a(\mu(\Omega))\tau^2(v_1, v_3)\tau^2(v_2, v_4) \\ &+ b(\mu(\Omega))\tau^3(v_1, v_3, v_4)\tau^1(v_2) \\ &+ c(\mu(\Omega))\tau^4(v_1, v_2, v_3, v_4) \\ &+ \dots\end{aligned}$$

Tensors of this form are called *algebraically generated by the canonical tensors.*

For $n = 2$ and 3, there is the following classification by Chentsov and Campbell if Ω is finite.

Theorem (Chentsov (1976), Campbell (1986))

Let $\Omega = I$ be a **finite measure space**.
The only invariant 2-tensors on a parametrized measure model are of the form

$$\begin{aligned}\sigma(v, w) &= f\,\tau^2(v, w) + g\,\tau^1(v)\tau^1(w)\\ &= f\,\mathfrak{g}_F(v, w) + g\,\partial_v\|p(\xi)\|\partial_w\|p(\xi)\|,\end{aligned}$$

where f, g are continuous functions depending on $\|p(\xi)\|$.
In particular, for a **statistical model** (i.e. $\|p(\xi)\| \equiv 1$), the only such tensor is (up to a constant) the Fisher metric.

Theorem (Chentsov (1976), Campbell (1986))

Let $\Omega = I$ be a **finite measure space**.
The only invariant 3-tensors on a parametrized measure model are of the form

$$\begin{aligned}\sigma(v,w,u) \;=\; & f\,\tau^3(v,w,u) \\ & +g_1\,\tau^2(v,w)\tau^1(u) + g_2\,\tau^2(w,u)\tau^1(v) \\ & \quad +g_3\,\tau^2(u,v)\tau^1(w) \\ & +h\,\tau^1(v)\tau^1(w)\tau^1(u)\end{aligned}$$

for functions f, g_1, g_2, g_3, h depending on $\|p(\xi)\|$.
In particular, for a **statistical model** (i.e. $\|p(\xi)\| \equiv 1$), the only such tensor is (up to a constant) the Amari-Chentsov tensor τ^3.

What about invariant tensors for $n \geq 4$?

We can define other invariant tensors by

- Forming arbitrary tensor products of these.
  For instance, the following are invariant 4-tensor fields:

$$\begin{array}{rcll}
\sigma_0(v_1, v_2, v_3, v_4) & = & \tau^4(v_1, v_2, v_3, v_4), & \text{or} \\
\sigma_1(v_1, v_2, v_3, v_4) & = & \tau^2(v_1, v_3)\tau^2(v_2, v_4), & \text{or} \\
\sigma_2(v_1, v_2, v_3, v_4) & = & \tau^1(v_3)\tau^3(v_1, v_2, v_4), & \text{or} \\
\sigma_3(v_1, v_2, v_3, v_4) & = & \tau^1(v_1)\tau^1(v_4)\tau^2(v_2, v_3), & \text{or} \\
\vdots & \vdots & \vdots &
\end{array}$$

- Taking linear combiations of such tensors with functions depending on $\|p(\xi)\|$, for instance,

$$\|p(\xi)\|^2\sigma_1(v_1, v_2, v_3, v_4) + \sigma_3(v_1, v_2, v_3, v_4).$$

Generalization of the Chentsov-Campbell Theorem

We can generalize extend the result of Chentsov / Campbell to **arbitrary measure spaces** and **tensors of arbitrary degree**:

Theorem (Ay, Lost, Lê, S. (2014))

Any family (Θ_Ω^k) of k-tensors which is invariant under sufficient statistics is algebraically generated by $(\tau^n)_{n\in\mathbb{N}}$ in the sense specified above.

Remark:

- Our proof shows that it suffices that the family is invarant under sufficient statistics of the form $\kappa : \Omega \to I$, where I is a finite set.
- If Ω is s manifold, then any diffeomorphism $\kappa : \Omega \to \Omega$ may be regarded as a statistic. Thus, an invariant family of tensors must be invariant under the diffeomorphism group.
  According to Michor et al., this property is already enough to characterize the tensors algebraically generated by the canonical tensors, if one assumes that the family $(p(\xi))_{\xi \in M}$ only consists of densities on Ω. (cf. Michor's talk)

References:

This talk is based on the following references:

- N.Ay, J.Jost, H.V.Lê, L.S., Information geometry and sufficient statistics, Probability Theory and Related Fields 162 no. 1-2, (2015), 327-364.
- N.Ay, J.Jost, H.V.Lê, L.S., Parametrized measure models, arXiv:1510.07305
- N.Ay, J.Jost, H.V.Lê, L.S., Information Geometry, (Textbook, Springer, 2016?)

Thank you for your attention!