The Fisher metric and probabilistic mappings

Hông Vân Lê
Institute of Mathematics, Czech Academy of Sciences

Workshop "Information Geometry",
Toulouse, October 16, 2019

1. Motivations / Introduction

• $\mathcal{P}(\mathcal{X})$ - the space of all probability measures on $\mathcal{X}$, $\mathcal{S}(\mathcal{X})$ - the space of all finite signed measures on $\mathcal{X}$,and $i : \mathcal{P}(\mathcal{X}) \to \mathcal{S}(\mathcal{X})$.

• $(M, \mathcal{X}, \mathbf{p})$ is a parameterized statistical model if M is a Banach manifold, and $i \circ \mathbf{p} : M \stackrel{\mathbf{p}}{\to} \mathcal{P}(\mathcal{X}) \stackrel{i}{\to} \mathcal{S}(\mathcal{X})$ is a C^1-map.

• there does not exist $(M, N, \mathbf{p})$ s.t. $\mathbf{p}(M) = \{f \cdot \nu_0 | \, f \in C^\infty(N)\}$, N is a compact smooth manifold.

- Allow M to be a Fréchet manifold?

- My suggestion: Endow statistical models, i.e., subsets in $\mathcal{P}(\mathcal{X})$, with diffeology, since

(1) Every statistical model has a natural diffeology.

(2) We can extend many concepts and results in AJLS theory to diffeological statistical models conceptually.

(3) Working with diffeological statistical models in many cases is simpler than with working with Fréchet statistical models.

2. Almost 2-integrable diffeological statistical models

• Any statistical model $P_{\mathcal{X}} \subset \mathcal{P}(\mathcal{X})$ is endowed with a natural geometric structure induced from the Banach space $(\mathcal{S}(\mathcal{X}), ||, ||_{TV})$.

- $v \in \mathcal{S}(\mathcal{X})$ is called a tangent vector of $P_{\mathcal{X}}$ at ξ_0, if $\exists$ a C^1-map $c : \mathbb{R} \to P_{\mathcal{X}} \xrightarrow{i} \mathcal{S}(\mathcal{X})$ s.t. $c(0) = \xi_0$ and $\dot{c}(0) = v$.

- The tangent (double) cone $C_\xi P_{\mathcal{X}} :=$ $\{v \in \mathcal{S}(\mathcal{X}) | \, v$ is a tangent vector of $P_{\mathcal{X}}$ at $\xi\}$.

- The tangent space $T_\xi P_{\mathcal{X}}$:= the linear hull of $C_\xi P_{\mathcal{X}}$.

- The tangent cone fibration $CP_{\mathcal{X}}$:= $\cup_{\xi \in \mathcal{X}} C_x \mathcal{X} \subset \mathcal{S}(\mathcal{X}) \times \mathcal{S}(\mathcal{X})$

- The tangent fibration $TP_{\mathcal{X}}$:= $\cup_{\xi \in \mathcal{X}} T_\xi P_{\mathcal{X}} \subset \mathcal{S}(\mathcal{X}) \times \mathcal{S}(\mathcal{X})$.

• We have $v \ll \xi$ for all $v \in C_\xi P_{\mathcal{X}}$. Hence we define the logarithmic representation $\log v$ of v by $\log v := dv/d\xi \in L^1(\mathcal{X}, \xi)$.

• $\log(C_\xi P) := \{\log v | v \in C_\xi P_\mathcal{X}\}$ is a subset in $L^1(\mathcal{X}, \xi)$. We denote call it the logarithmic representation of $C_\xi P$.

• Let $P_\mathcal{X} := \{p_\eta(x)\mu_0 | \mu_0 \in \mathcal{P}(\mathcal{X})\}$ (M1)
$p_\eta(x) = \sum_{i=1}^3 g^i(x)\eta_i$ for $x \in \mathcal{X}$.
$g^i(x) \geq 0$ for all $x \in \mathcal{X}$ and $\mathbb{E}_\mu(g^i) = 1$,
$\eta = (\eta_1, \eta_2) \in D_b \subset [0,1] \times [0,1] \subset \mathbb{R}^2$,
$\eta_3 = 1 - e_1 - e_2$.
$C_{p_\eta} P_\mathcal{X}$ can be $\mathbb{R}^2$ or $\mathbb{R}$ or $\mathbb{R} \cup \mathbb{R}$.

• We want to put a Riemannian metric on $P_\mathcal{X}$ i.e., to put a positive quadratic form $\mathfrak{g}$ on each tangent space $T_\xi P_\mathcal{X}$.

Definition A statistical model $P_{\mathcal{X}}$ will be called almost 2-integrable, if

$$\log(C_\xi P) \subset L^2(\mathcal{X}, \xi) \, \forall \xi \in P_{\mathcal{X}}.$$

Then the Fisher metric is defined by

$$\mathfrak{g}_\xi(v, w) := \langle \log v, \log w \rangle_{L^2(\mathcal{X}, \xi)} \text{ for } v, w \in C_\xi P$$

and extend to $v, w \in T_\xi P_{\mathcal{X}}$.

(M1). $P_{\mathcal{X}} = \mathbf{p}(D_b) \subset \mathcal{S}(\mathcal{X})$,

$$\mathbf{p} : \mathbb{R}^2 \to \mathcal{S}(\mathcal{X}), \eta \mapsto p_\eta \cdot \mu_0,$$

is an affine map, defined by the same formula. Hence $\forall\ \tilde{v} \in T_\eta P_{\mathcal{X}}$ we have $\tilde{v} = d\mathbf{p}(v)$.

$$d\mathbf{p}(v_1, v_2) = [(g^1 - g^3)v_1 + (g^2 - g^3)v_2]\mu_0.$$

If $g^i(x) > 0\ \forall\ x \in \mathcal{X}$, $\forall i$,

$$\log d\mathbf{p}(v)_{|\mathbf{p}(\eta)} = \frac{(g^1 - g^3)v_1 + (g^2 - g^3)v_2}{p_\eta}.$$

Hence $P_{\mathcal{X}}$ is almost 2-integrable, if

$$\frac{g^1 - g^3}{\sqrt{p_\eta}}, \frac{g^2 - g^3}{\sqrt{p_\eta}} \in L^2(\mathcal{X}, \mu_0)\ \forall \eta \in D_b.$$

• A C^k-diffeology on a nonempty $\mathcal{X}$ is any set $\mathcal{D}$ of parametrizations of $\mathcal{X}$, s.t.:

D1. Covering. The set $\mathcal{D}$ contains the constant maps $\mathbf{x} : r \mapsto x$, defined on $\mathbb{R}^n$, for all $x \in \mathcal{X}$ and for all $n \in \mathbb{N}$.

D2. Locality. Let $P : U \to \mathcal{X}$ a map. If $\forall\, r \in U$ there exists an open neighborhood $V \ni r$ such that $P_{|V}$ belongs to $\mathcal{D}$ then P belongs to $\mathcal{D}$.

D3. Smooth compatibility. If $P : U \to \mathcal{X}$ belongs to $\mathcal{D}$, then $\forall$ domain V, for every $F \in C^k(V, U)$, $P \circ F$ belongs to $\mathcal{D}$.

• A C^k-diffeological space is a nonempty set equipped with a C^k-diffeology.

• A statistical model $\mathcal{P}_{\mathcal{X}}$ endowed with a diffeology $\mathcal{D}$ will be called a diffeological statistical model, if $P : U \to \mathcal{P}_{\mathcal{X}}$ belongs to $\mathcal{D}$ then $i \circ P : U \to \mathcal{S}(\mathcal{X})$ is a C^1-map.

• A diffeological statistical model $(\mathcal{P}_{\mathcal{X}}, \mathcal{D})$ will be called almost 2-integrable, if $\log(T_\xi(\mathcal{P}_{\mathcal{X}}, \mathcal{D})) \subset L^2(\mathcal{X}, \xi)$ for all $\xi \in P_{\mathcal{X}}$.

• An almost diffeological statistical model $(\mathcal{P}_{\mathcal{X}}, \mathcal{D})$ will be called 2-integrable, if for any map $\mathbf{p} : U \to \mathcal{P}_{\mathcal{X}}$ in $\mathcal{D}$ the function $v \mapsto |d\mathbf{p}(v)|_{\mathfrak{g}}$ is continuous on TU.

(PD) Let $(M, \mathcal{X}, \mathbf{p})$ be a parametrized statistical model. Then $(\mathbf{p}(M), \mathbf{p}_*(\mathcal{D}_M))$ is a diffeological statistical model, where D_M is the standard diffeology (C^1-structure) on M. In other words $\mathbf{p}_*(\mathcal{D}_M)$ consists of all C^1-map $q : \mathbb{R}^n \supset U \to \mathbf{p}(M)$ such that there exists a smooth map $q^M : U \to M$ and $q = \mathbf{p} \circ q^M$.

By Theorem 3.2 in AJLS 2017, a parametrized statistical model $(M, \mathcal{X}, \mathbf{p})$ is 2-integrable, iff and only if $(\mathbf{p}(M), \mathbf{p}_*(\mathcal{D}_M))$ is a 2-integrable diffeological statistical model.

(NPD) Let $\mathcal{X}$ be a smooth finite dimensional compact manifold with a volume form μ. Let $\mathcal{P}_{\mathcal{X}} := \{f \cdot \mu | \, f \in C^\infty(\mathcal{X}),\, f(x) > 0,\, \mathbb{E}_\mu f = 1\}$.

• $\mathcal{P}_{\mathcal{X}}$ cannot be C^1-parameterized by a Banach manifold M. Assume the opposite, i.e., $\exists$ a C^1-map $\mathbf{p} : M \to \mathcal{P}(\mathcal{X})$ s.t. $\mathbf{p}(M) = \mathcal{P}_{\mathcal{X}}$. Then $\forall m \in M$ we have $d\mathbf{p}(T_m(M)) = T_{\mathbf{p}(m)}\mathcal{P}_{\mathcal{X}} = \{f \in C^\infty(\mathcal{X}) | \, \mathbb{E}_\mu f = 0\}$. Hence $C^\infty(\mathcal{X})$ can be endowed with a compatible Banach norm. But it is not the case.

• $(\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})$ is a 2-integrable diffeological statistical model.

3. Probabilistic mappings

• $\mathcal{F}_s(\mathcal{X})$ - the linear space of simple functions on $\mathcal{X}$.

• $I : \mathcal{F}_s(\mathcal{X}) \to Hom(S(\mathcal{X}), \mathbb{R})$, $f \mapsto I_f$, $I_f(\mu) := \int_{\mathcal{X}} f d\mu$ for $\mu \in \mathcal{S}(\mathcal{X})$.

• Σ_w - the smallest σ-algebra on $\mathcal{S}(\mathcal{X})$ such that I_f is measurable for all $f \in \mathcal{F}_s(\mathcal{X})$. We also denote by Σ_w the restriction of Σ_w to $\mathcal{M}(\mathcal{X})$, $\mathcal{M}^*(\mathcal{X}) := \mathcal{M}(\mathcal{X}) \setminus \{0\}$, and $\mathcal{P}(\mathcal{X})$.

If $\mathcal{X}$ is topological we let $\Sigma_{\mathcal{X}} := \mathcal{B}(\mathcal{X})$.

- $C_b(\mathcal{X})$ - the space of bounded continuous functions on $\mathcal{X}$,
- τ_v - the smallest topology on $\mathcal{S}(\mathcal{X})$ such that for any $f \in C_b(\mathcal{X})$ the map $I_f : (\mathcal{S}(\mathcal{X}), \tau_v) \to \mathbb{R}$ is continuous.
- The restriction of τ_v to $\mathcal{M}(\mathcal{X})$, $\mathcal{P}(\mathcal{X})$ is the weak topology.
- If $\mathcal{X}$ is separable and metrizable then the Borel σ-algebra $\mathcal{B}(\tau_v)$ on $\mathcal{P}(\mathcal{X})$ coincides with Σ_w.

- A probabilistic mapping $T : \mathcal{X} \rightsquigarrow \mathcal{Y}$ is a measurable mapping $\overline{T} : \mathcal{X} \to (\mathcal{P}(\mathcal{Y}), \Sigma_w)$.

- For a measurable mapping $\mathbf{p} : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ we shall denote by $\underline{\mathbf{p}} : \mathcal{X} \rightsquigarrow \mathcal{Y}$ the generated probabilistic mapping.

- $Id_{\mathcal{P}} : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{X})$ generates $ev : (\mathcal{P}(\mathcal{X}), \mathcal{B}(\tau_v)) \rightsquigarrow (\mathcal{X}, \mathcal{B}(\mathcal{X}))$, i.e., $\overline{ev} = Id_{\mathcal{P}}$.

- Any measurable map $\kappa : \mathcal{X} \to \mathcal{Y}$ generates a measurable map $\overline{\kappa} : \mathcal{X} \stackrel{\delta \circ \kappa}{\to} \mathcal{P}(\mathcal{Y})$.

• For $T : \mathcal{X} \rightsquigarrow \mathcal{Y}$, the Markov morphism $S_*(T) : \mathcal{S}(\mathcal{X}) \to \mathcal{S}(\mathcal{Y})$ is defined by

$$S_*(T)(\mu)(B) := \int_{\mathcal{X}} \overline{T}(x)(B) d\mu(x)$$

for $\mu \in \mathcal{S}(\mathcal{X})$ and $B \in \Sigma_{\mathcal{Y}}$.

• (Chentsov1982) $T : \mathcal{X} \rightsquigarrow \mathcal{Y}$ induces a linear bounded map $S_*(T) : \mathcal{S}(\mathcal{X}) \to \mathcal{S}(\mathcal{Y})$.

$S_*(T)(\mathcal{M}(\mathcal{X})) \subset \mathcal{M}(\mathcal{Y}), S_*(T)(\mathcal{P}(\mathcal{X})) \subset \mathcal{P}(\mathcal{Y})$.

• (Giry1982, JLLT2019) Probabilistic mappings are morphisms in the category of measurable spaces, i.e., for any $T_1 : \mathcal{X} \rightsquigarrow \mathcal{Y}$, $T_2 : \mathcal{Y} \rightsquigarrow \mathcal{Z}$

$$M_*(T_2 \circ T_1) = M_*(T_2) \circ M_*(T_1),$$

$$P_*(T_2 \circ T_1) = P_*(T_2) \circ P_*(T_1).$$

• $M_* := S_*(T)_{|\mathcal{M}(\mathcal{X})}$ and $P_* := S_*(T)_{|\mathcal{P}(\mathcal{X})}$ are faithful functors, since

$$S_*(T)(\delta_x) = \overline{T}(x) \in \mathcal{P}(\mathcal{Y}).$$

• (Morse-Sacksteder1966) If $\nu \ll \mu \in \mathcal{M}^*(\mathcal{X})$ then $M_*(T)(\nu) \ll M_*(T)(\mu)$.

• We also denote by T_* the map $S_*(T)$ if no confusion can arise.

Theorem(L. 2019) Let $(\mathcal{P}_{\mathcal{X}}, \mathcal{D})$ is a diffeological statistical model and $T : \mathcal{X} \rightsquigarrow \mathcal{Y}$.

(1) Then $(T_*(P_{\mathcal{X}}), T_*(\mathcal{D}))$ is a diffeological statistical model.

(2) If $(P_{\mathcal{X}}, \mathcal{D})$ is an almost 2-integrable, then $(T_*(P_{\mathcal{X}}), T_*(\mathcal{D}))$ is also an almost 2-integrable.

(3) If $(P_{\mathcal{X}}, \mathcal{D})$ is a 2-integrable, then $(T_*(P_{\mathcal{X}}), T_*(\mathcal{D}))$ is also a 2-integrable.

• $L(\mathcal{Y})$: $=$ the set of all bounded measurable functions on $\mathcal{X}$.

• Given $T : \mathcal{X} \rightsquigarrow \mathcal{Y}$, let $T^* : L(\mathcal{Y}) \to L(\mathcal{X})$,

$$T^*(f)(x) := I_f(\overline{T}(x)) = \int_{\mathcal{Y}} f d\overline{T}(x)$$

• If $\kappa : \mathcal{X} \to \mathcal{Y}$ is a measurable mapping, then $\kappa^*(f)(x) = f(\kappa(x))$, since $\overline{\kappa} = \delta \circ \kappa$.

• A morphism $T : (\mathcal{X}, P_{\mathcal{X}}) \rightsquigarrow (\mathcal{Y}, P_{\mathcal{Y}})$ will be called sufficient if there exists a probabilistic mapping $\underline{\mathbf{p}} : \mathcal{Y} \rightsquigarrow \mathcal{X}$ such that for all $\mu \in P_{\mathcal{X}}$ and $h \in L(\mathcal{X})$ we have

$$T_*(h\mu) = \underline{\mathbf{p}}^*(h) T_*(\mu) \qquad (SP)$$

$$\Longleftrightarrow \underline{\mathbf{p}}^*(h) = \frac{dT_*(h\mu)}{dT_*(\mu)} \in L^1(\mathcal{Y}, T_*(\mu)).$$

In this case we shall call $T : \mathcal{X} \rightsquigarrow \mathcal{Y}$ a probabilistic mapping sufficient for $\mathcal{P}_{\mathcal{X}}$ and we shall call the measurable mapping $\mathbf{p} : \mathcal{Y} \to \mathcal{P}(\mathcal{X})$ in (SP) a conditional mapping for T.

• Assume that $\kappa : (\mathcal{X}, P_{\mathcal{X}}) \rightsquigarrow (\mathcal{Y}, P_{\mathcal{Y}})$ is a sufficient morphism, where $\kappa : \mathcal{X} \to \mathcal{Y}$ is a statistic. Let $\mathbf{p} : \mathcal{Y} \to \mathcal{P}(\mathcal{X}),\ y \mapsto \mathbf{p}_y$, be a conditional mapping for κ. Since $\underline{\mathbf{p}}^*(1_A)(y) = \mathbf{p}_y(A)$, for all $\mu \in \mathcal{P}(\mathcal{X})$ we have

$$\mathbf{p}_y(A) = \frac{d\kappa_*(1_A \mu)}{d\kappa_* \mu} \in L^1(\mathcal{Y}, \kappa_*(\mu)) \quad (CM)$$

The RHS of (CM) is the conditional measure of μ applied to $A \in \Sigma_{\mathcal{X}}$ w.r.t. the measurable mapping κ. The equality (CM) implies that this conditional measure is regular and independent of μ.

Thus the notion of sufficiency of κ for $P_{\mathcal{X}}$ coincides with the classical notion of sufficiency of κ for $P_{\mathcal{X}}$. We also note that the equality in (CM) is understood as equivalence class in $L^1(\mathcal{Y}, \kappa_*(\mu))$ and hence every statistic κ' that coincides with a sufficient statistic κ except on a zero μ-measure set, for all $\mu \in P_{\mathcal{X}}$, is also a sufficient statistic for $P_{\mathcal{X}}$.

- Assume that $\mu \in \mathcal{P}(\mathcal{X})$ has a regular conditional distribution w.r.t. to a statistic $\kappa : \mathcal{X} \to \mathcal{Y}$, i.e., there exists a measurable mapping $\mathbf{p} : \mathcal{Y} \to \mathcal{P}(\mathcal{X}), y \mapsto \mathbf{p}_y$, such that

$$\mathbb{E}_{\mu}^{\sigma(\kappa)}(\mathbf{1}_A | y) = \mathbf{p}_y(A)$$

for any $A \in \Sigma_{\mathcal{X}}$ and $y \in \mathcal{Y}$.

• Let $P := \{\nu_\theta \in \mathcal{P}(\mathcal{X}) | \, \theta \in \Theta, \, \nu_\theta \ll \mu\}$. If there exist a function $h : \mathcal{Y} \times \Theta \to \mathbb{R}$ such that for all $\theta \in \Theta$ and we have

$$\nu_\theta = h(\kappa(x))\mu \quad (FN)$$

then κ is sufficient for P, since for any $\theta \in \Theta$

$$\mathbf{p}^*(1_A) = \frac{d\kappa_*(1_A \nu_\theta)}{d\kappa_* \nu_\theta}$$

does not depend on θ. The condition (FN) is the Fisher-Neymann sufficiency condition for a family of dominated measures.

Theorem (L.2019) Let $P_{\mathcal{X}} \subset \mathcal{P}(\mathcal{X})$ be an almost 2-integrable statistical model. Then for any $T : \mathcal{X} \leadsto \mathcal{Y}$, any $\mu \in P_{\mathcal{X}}$ and any $v \in T_\mu P_{\mathcal{X}}$ we have $T_*(v) \in T_{T_*\mu} T_*(P_{\mathcal{X}})$ and

$$\mathfrak{g}_\mu(v, v) \geq \mathfrak{g}_{T_*\mu}(T_*v, T_*v)$$

with the equality if T is sufficient w.r.t. $P_{\mathcal{X}}$.

Thank you for your attention!