Diffeological statistical models and diffeological Hausdorff measures

Hông Vân Lê
Institute of Mathematics, CAS, Praha

Praha March 10, 2021

OUTLINE

1. Diffeological spaces and diffeological statistical models.

2. Diffeological Fisher metric, diffeological Hausdorff measure.

3. Probabilistic morphisms and naturalilty of the diffeological Fisher metric.

4. Diffeological Fisher metric and diffeological Hausdorff measure in statistical physics and machine learning.

1. Diffeological spaces and diffeological statistical models

• Till the later part of the last century, differential calculus, invented by Newton as part of his investigations in physics and geometry, has been developed only for "regular" geometric spaces (differentiable manifolds).

• Diffeology, invented by Souriau in 1980, gives a unified (and relatively simple) treatment of regular and new singular objects in physics.

1.1. Definition (Le 2020, cf. IZ2013) For $k \in \mathbf{N} \cup \infty$ and $X \neq \emptyset$, a C^k-diffeology of X is a set $\mathcal{D}$ of mappings $\mathbf{p} : \mathbf{R}^n \supset U$ (open) $\to X$, $n \in \mathbf{N}$, s.t. the following axioms hold.

D1. Covering. $\mathcal{D}$ contains the constant mappings $\mathbf{x} : r \mapsto x$, $\forall\ x \in X$, $\forall\ n \in \mathbf{N}$.

D2. Locality. Let $\mathbf{p} : U \to X$ be a mapping. If for every point $r \in U$ there exists an open neighborhood V of r such that $\mathbf{p}_{|V}$ belongs to $\mathcal{D}$ then the map $\mathbf{p}$ belongs to $\mathcal{D}$.

D3. Smooth compatibility. For every element $\mathbf{p}: U \to X$ of $\mathcal{D}$, for every real domain V, for every $\psi \in C^k(V, U)$, $\mathbf{p} \circ \psi$ belongs to $\mathcal{D}$.

A C^k-diffeological space X is a nonempty set X equipped with a C^k-diffeology $\mathcal{D}$. Elements $\mathbf{p}: U \to X$ of $\mathcal{D}$ will be called C^k-maps from U to X.

A map $f : (X, \mathcal{D}) \to (X', \mathcal{D}')$ between two C^k-diffeological spaces is called a C^k-map, if for any $\mathbf{p} \in \mathcal{D}$ we have $f \circ \mathbf{p} \in \mathcal{D}'$.

• Originally the concept of diffeological spaces covers only C^∞-maps, because the category of C^∞-diffeology is "convenient", in particular Cartesian closed.

1.2. Examples • Let V be a Banach space. Then V has the canonical C^k-diffeology $\mathcal{D}^k_{can}$ that consists of all C^k-differentiable mappings $\mathbf{p} : U \to V$, where U is an open domain in $\mathbf{R}^n$. The space V has also another C^k-diffeology $\mathcal{D}^k_w$ that consists of all weakly C^k-differentiable mappings $U \to V$, where U is an open domain in $\mathbf{R}^n$.

• Pullback diffeology: $(X', \mathcal{D}')$ - a C^k-diffeological space, $f : X \to X'$ is a map. Then

$$f^*(\mathcal{D}') := \{\mathbf{p} : U \to X |\, f \circ \mathbf{p} \in \mathcal{D}'\},$$

is the pullback C^k-diffeology on X.

• Pushforward diffeology $(X, \mathcal{D})$ - a C^k-diffeological space, $f : X \to X'$ a map. Then

$$f_*(\mathcal{D}) : \{\mathbf{p} : U \to X' |\, \exists \mathbf{q} \in \mathcal{D} s.t. \mathbf{p} = f \circ \mathbf{q}\},$$

is the pushforward diffeology on X'.

• The space $\mathcal{S}(\mathcal{X})$ of all finite signed measures on a measurable space $\mathcal{X}$ is a Banach space with the total variation norm $\|\cdot\|_{TV}$.

• A statistical model $\mathcal{P}_{\mathcal{X}}$ is a subset in $\mathcal{P}(\mathcal{X}) \subset \mathcal{S}(\mathcal{X})$. Write $\mathcal{P}_{\mathcal{X}} \xrightarrow{i} \mathcal{S}(\mathcal{X})$.

1.3. Definition (Le2020, LT2021) (1) A statistical model $\mathcal{P}_{\mathcal{X}}$ endowed with a C^k-diffeology $\mathcal{D}_{\mathcal{X}}$ is called a C^k-diffeological statistical model (resp. a weakly C^k-diffeological statistical model), if $i_*(\mathcal{D}_{\mathcal{X}}) \subset \mathcal{D}^k_{can}$ (resp. if $i_*(\mathcal{D}_{\mathcal{X}}) \subset \mathcal{D}^k_w$).

1.6. Examples of C^k-diffeological statistical models are the image $(\mathbf{p}(M), \mathbf{p}_*(\mathcal{D}^k_{can}))$ of parameterized statistical models $(M, \mathcal{X}, \mathbf{p})$, where M is a smooth Banach manifold and $i \circ \mathbf{p} : M \to \mathcal{S}(\mathcal{X})$ is a C^k-map.

1.7. Definition. Let $(\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})$ be a (weakly) C^k-diffeological statistical model and $c : (-\varepsilon, \varepsilon) \to (\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})$ a C^k-map. The tangent vector $\partial_t c(0)$ at $c(0)$ is the image of the map $dc_0(\partial t) \in \mathcal{S}(\mathcal{X})$, where dc_0 is the (weak) differential of c at 0.

- For $\xi \in \mathcal{P}_{\mathcal{X}}$, the tangent cone $C_\xi(\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})$ consists of all tangent vectors $\partial_t^w c(0)$ at $c(0) = \xi$, where $c : (0,1) \to (\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})$ be a C^k-map, and the tangent space $T_\xi(\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})$ is the linear hull of $C_\xi(\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})$.

- For any $v \in X_\xi(\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})$ we have $v \ll \xi$. Write $\log v := dv/d\xi \in L^1(\mathcal{X}, \xi)$

- Using the concept of a tangent cone we can distinguish different C^1-diffeologies on a given statistical model (LT2020).

2. Diffeological Fisher metric, diffeological Hausdorff measure

• A C^k-diffeological statistical model $(\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})$ will be called almost 2-integrable, if $\log\Big(C_\xi(\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})\Big) \subset L^2(\mathcal{X}, \xi)\ \forall\ \xi \in \mathcal{P}_{\mathcal{X}}$.
In this case we define the diffeological Fisher metric $\mathfrak{g}$ on $\mathcal{P}_{\mathcal{X}}$ as follows. For $v, w \in C_\xi(\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})$

$$\mathfrak{g}_\xi(v, w) := \langle \log v, \log w \rangle_{L^2(\mathcal{X},\xi)} = \int_{\mathcal{X}} \log v \cdot \log w \, d\xi.$$

The Fisher metric on $C_\xi(\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})$ extends naturally to a positive quadratic form on $T_\xi(\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})$, which is also called the Fisher metric.

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

2.3. Definition. Let $(\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})$ be a 2-integrable C^k-diffeological statistical model.

(1) A map $c : [a, b] \to (\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})$ will be called a C^k-curve, if $\exists\ \varepsilon > 0$ and a C^k-map: $c_\varepsilon : (a - \varepsilon, b + \varepsilon) \to (\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})$ s.t. the restriction of c_ε to $[a, b]$ is c.

(2) A continuous map $c : [0,1] \to (\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})$ will be called a piece-wise C^k-curve, if there exists a finite number of points $0 = a_0 < a_1 < a_2 \cdots < a_m = 1$ such that the restriction of c to $[a_{i-1}, a_i]$ is a C^k-curve for $i \in [1, m]$.

(3) Let $c : [0,1] \to (\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})$ be a C^k-curve connecting $q_1, q_2 \in \mathcal{P}_{\mathcal{X}}$ such that $c(0) = q_1$ and $c(1) = q_2$. We define the length of c by

$$L(c) = \int_0^1 |\dot{c}(t)|_{\mathfrak{g}} \, dt.$$

The length of a piece-wise C^k-curve will be defined as the sum of the lengths of its C^k-smooth sub-intervals.

(4) The diffeological Fisher distance $\rho_{\mathfrak{g}}(x, y)$ between two points $x, y \in \mathcal{P}_{\mathcal{X}}$ will be defined as the infimum of the length over the space of piece-wise C^k-curves connecting x, y. In particular, if there is no C^k-path connecting x, y then $\rho_{\mathfrak{g}}(x, y) = \infty$.

2.4. Theorem(LT2020) The distance function $\rho_{\mathfrak{g}}(x, y)$ is an extended metric.

- The Hausdorff measure on (E, d).

$$\text{diam}\,(S \subset E) := \sup\{d(x,y) \,|\, x, y \in S\}.$$

- $\alpha_k := \text{vol} B^k(0,1)$.
- the k-dimensional Hausdorff measure of S

is defined as follows:

$$\mathcal{H}^k(S) := \lim_{\delta \to 0} \alpha_k \inf \left\{ \sum_{j \in I} \left(\frac{\text{diam}\,(S_j)}{2} \right)^k \right.$$

$$\left. \Big| \text{diam}\,(S_j) \le \delta \,\&\, S \subset \bigcup_{j \in I} S_i \right\}.$$

It is known that $\mathcal{H}^k$ is a regular Borel measure.

The definition of the k-dimensional Hausdorff measure extends to any nonnegative real dimension k, by extending the definition of α_k with the following definition

$$\alpha_k := \frac{\pi^{k/2}}{\Gamma(1+k/2)} \text{ where } \Gamma(t) := \int_0^\infty x^{t-1} e^{-x}\, dx.$$

The Hausdorff dimension $\mathcal{H}\text{-}\dim(S)$ is

$$\mathcal{H}\text{-}\dim(A) := \inf\{m \geq 0 \,|\, \mathcal{H}^m(A) = 0\}.$$

It is known that if $k > \mathcal{H}\text{-}\dim(A)$ then $\mathcal{H}^k(B) = 0$ and if $k < \mathcal{H}\text{-}\dim(A)$ then $\mathcal{H}^k(A) = \infty$.

• (1) Let (M^m, g) be a Riemannian manifold, regarded as a metric space with the Riemannian distance d_g. Then the Hausdorff measure $\mathcal{H}^m$ on M^m coincides with the standard volume.

• (2) Let $\varphi : A \subset (M^k, g) \to (N^n, g')$ be a Lipschitz map from an open domain $A \subset (M^k, g)$ to (N^n, g'). Then $d\varphi$ and

$$\mathbf{J}\, d\varphi := \sqrt{\mathrm{det}\Big((d\varphi)^* \circ (d\varphi)\Big)}$$

are defined $\mathcal{H}^k$-a.e. on A. If $k = n$ then we have the following area formula

$$\mathcal{H}^n(\varphi(A)) = \int_A \mathbf{J} d\varphi d\mathcal{H}^n(x).$$

• Let $(\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})$ be a 2-integrable C^k-diffeological statistical model with the diffeological Fisher distance $d_{\mathfrak{g}}$ and $m \in \mathbf{R}$ the Hausdorff dimension of $(\mathcal{P}_{\mathcal{X}}, d_{\mathfrak{g}})$. Then the Hausdorff measure $\mathcal{H}^m_{\mathfrak{g}}$ on $(\mathcal{P}_{\mathcal{X}}, d_{\mathfrak{g}})$ will be called the diffeological Hausdorff–Jeffrey measure of $(\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})$.

3. Probabilistic morphisms and naturality of the diffeological Fisher metric.

• In 1962 Lawvere proposed a categorical approach to probability theory, where morphisms are Markov kernels, and supplied the space $\mathcal{P}(\mathcal{X})$ with a natural σ-algebra Σ_w, making many constructions in probability theory and mathematical statistics functorial.

• $\mathcal{F}_s(\mathcal{X})$ the space of simple functions on $\mathcal{X}$. $I : \mathcal{F}_s(\mathcal{X}) \to \mathrm{Hom}\Big(S(\mathcal{X}), \mathbf{R}\Big)$, $f \mapsto I_f$, $I_f(\mu) := \int_{\mathcal{X}} f d\mu$ for $f \in \mathcal{F}_s(\mathcal{X})$ and $\mu \in \mathcal{S}(\mathcal{X})$.

- Σ_w is the smallest σ-algebra on $\mathcal{S}(\mathcal{X})$ s.t. I_f is measurable for all $f \in \mathcal{F}_s(\mathcal{X})$. Let $\mathcal{M}(\mathcal{X})$ denote the space of all finite nonnegative measures on $\mathcal{X}$. 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 $\mathcal{P}(\mathcal{X})$.

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

- Giry (1982) $i : \mathcal{X} \to \mathcal{P}(\mathcal{X})$, $x \mapsto \delta_x$, is a measurable mapping.

Hence any measurable mapping $\kappa : \mathcal{X} \to \mathcal{Y}$ assigns $\underline{i \circ \kappa} : \mathcal{X} \rightsquigarrow \mathcal{Y}$.

• $T : \mathcal{X} \rightsquigarrow \mathcal{Y} \Longrightarrow S_*(T) : \mathcal{S}(\mathcal{X}) \to \mathcal{S}(\mathcal{Y})$

$$S_*(T)(\mu)(B) := \int_{\mathcal{X}} \overline{T}(x)(B) d\mu(x) \qquad (1)$$

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

Write T_* as $S_*(T)$. Then $T_*(\mathcal{P}(\mathcal{X})) \subset \mathcal{P}(\mathcal{Y})$.

• Given $T : \mathcal{X} \rightsquigarrow \mathcal{Y}$ and a C^k-diffeological space $(\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})$ let $T_*(\mathcal{D}_{\mathcal{X}}) := (T_*)_*(\mathcal{D}_{\mathcal{X}})$, where $T_* : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{Y})$ is a smooth map.

3.2 Proposition(Le2020)

- $\Big(T_*(\mathcal{P}_{\mathcal{X}}), T_*(\mathcal{D}_{\mathcal{X}})\Big)$ is a C^k-diffeological statistical model.
- If $(\mathcal{P}_{\mathcal{X}}, \mathcal{D}_{\mathcal{X}})$ is an (almost) 2-integrable, then $\Big(T_*(\mathcal{P}_{\mathcal{X}}), T_*(\mathcal{D}_{\mathcal{X}})\Big)$ is also (almost) 2-integrable.

- (LT2020) Proposition 3.2 also holds for weakly C^k-diffeological statistical models.

3.3. Proposition. Let $T : \mathcal{X} \rightsquigarrow \mathcal{Y}$ be a probabilistic morphism and $(\mathcal{P}_\mathcal{X}, \mathcal{D}_\mathcal{X})$ an almost 2-integrable C^k-diffeological statistical model. Then for any $\mu \in \mathcal{P}_\mathcal{X}$ and any $v \in T_\mu(\mathcal{P}_\mathcal{X}, \mathcal{D}_\mathcal{X})$ we have the followingmonotonicity

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

with the equality if T is sufficient for $\mathcal{P}_\mathcal{X}$.

• A 1-1 measurable mapping $\kappa : \mathcal{X} \to \mathcal{Y}$ induces a sufficient for any $\mathcal{P}_\mathcal{X}$ probabilistic morphism $\underline{\kappa} : \mathcal{X} \to \mathcal{Y}$.

3.4. Example Let $\mathcal{X}$ be a measurable space and λ be a σ-finite measure. Friedrich (1991) considered a family $P(\lambda) := \{\mu \in \mathcal{P}(\mathcal{X}) | \, \mu \ll \lambda\}$ that is endowed with the following diffeology $\mathcal{D}(\lambda)$. A curve $c : \mathbf{R} \to P(\lambda)$ is a C^1-curve, iff

$$\log \dot{c}(t) \in L^2(\mathcal{X}, c(t)).$$

Hence $(P(\lambda), \mathcal{D}(\lambda))$ is an almost 2-integrable C^1-diffeological statistical model (Le2020) but not 2-integrable(LT2020).

In his 1991 paper Friedrich considered the group $\mathcal{G}(\mathcal{X}, \Sigma_{\mathcal{X}}, \lambda)$ of all measurable 1-1 mappings $\Phi : \mathcal{X} \to \mathcal{X}$ such that $\Phi_*(\lambda) \ll \lambda$. Clearly $\Phi_*(P(\lambda)) \subset P(\lambda)$. Note that Φ is a sufficient statistic w.r.t. $P(\lambda)$. Hence Proposition 3.3 implies the following

Corollary (Friedirch 1991) The group $\mathcal{G}(\mathcal{X}, \Sigma_{\mathcal{X}}, \lambda)$ acts isometrically on $P(\lambda)$.

4. Applications.

• Let $P_{\mathcal{X}} \subset \mathcal{P}(\mathcal{X})$ be a statistical model. An estimator is a map $\hat{\sigma} : \mathcal{X} \to P_{\mathcal{X}}$. Let $\varphi : \mathcal{P}_{\mathcal{X}} \to V$ be a "coordinate" mapping.

$$MSE_{\xi}[\varphi \circ \hat{\sigma}](l,h) := \mathbb{E}_{\xi}[(\varphi^{l} \circ \hat{\sigma}(x) - \varphi^{l}(\xi))$$

$$\cdot(\varphi^{h} \circ \hat{\sigma}(x) - \varphi^{h}(\xi))].$$

$b^{\varphi}_{\hat{\sigma}} := \mathbb{E}_{\xi}(\varphi_{\hat{\sigma}}) - \varphi$ will be called the bias of the φ-estimator $\varphi \circ \hat{\sigma}$.

• Diffeological Crámer-Rao inequality (Le2020, LT2020) genaralizes the classical Crámer-Rao inequality for an unbiased estimator where φ is the inclusion coordinate map

$$MSE_{\xi}(\hat{\sigma}) \geq \mathfrak{g}^{-1}(\xi).$$

• Using the diffeological Fisher metric we can define the notion of (natural) gradient flow.

• Using the Hausdorff-Jeffrey measure we define a 'natural'' prior measure in Bayesian statistical model.

5. Open questions

- Investigate the existence of affine connection and Levi-Civita connections on tangent cone of diffeological statistical models.

- Theory of weakly diffeological statistical models is not yet well-understood

- The concept of C^k-diffeological spaces are not yet well understood.

- Many natural statistical models are not 2-integrable and we should consider Finsler metrics instead of Fisher-Rao metrics.

- Application of Metric Geometry and Geometry Measure Theory to statistical physics (e.g. theory of Gibbs measures) and ML.

- H. V. Lê, Diffeological statistical models, the Fisher metric and probabilistic mappings, arXiv:1912.02090, Mathematics 2020, 8(2), 167.

- H. V. Lê and A. Tuzhilin, Nonparametric estimations and the diffeological Fisher metric, arXiv:2011.13418

Thank you for your attention!