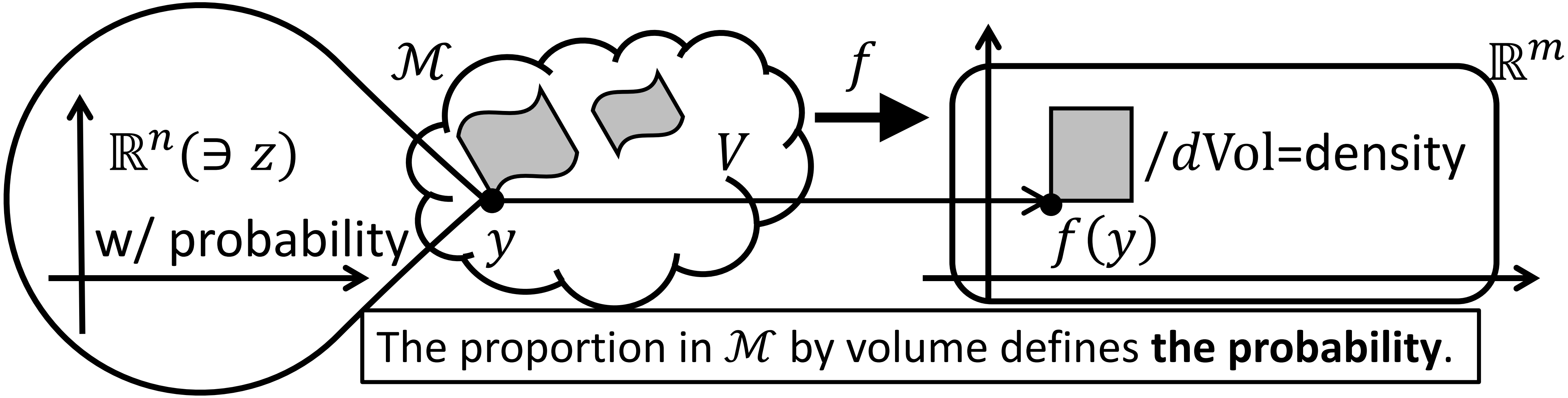
Symplectic/Contact Geometry Related to Bayesian Statistics

by

Atsuhide Mori (Osaka Dental University, Japan)

Geometric setting (1) The mind and statistics

- $\mathcal{M}$: manifold that stands for (a part of) the mind of an agent.
- Each point of $\mathcal{M}$ presents a probability density function on $\mathbb{R}^n$.
- Fix the state $V \in \mathcal{V} = \{$volume forms with finite total on $\mathcal{M}\}$ of $\mathcal{M}$. Then any statistic $f: \mathcal{M} \to \mathbb{R}^m$ obeys a probability distribution.



The proportion in $\mathcal{M}$ by volume defines **the probability**.

This probability distribution is the FULL estimation of the statistic f.
(A value and a confidence interval with error bar are not enough!)

Geometric setting (2) Bayesian updating

- Given a point $z \in \mathbb{R}^n$, the agent updates $V \in \mathcal{V}$ in the following way: Each point y of the mind $\mathcal{M}$ presents a probability density ρ_y of $\mathbb{R}^n$. Change it to the **likelihood** $\rho(z)(y) = \rho_y(z)$, and define the map
$$\varphi: \mathbb{R}^n \times \mathcal{V} \ni (z, V) \mapsto \rho(z)V \in \mathcal{V}$$: "**updating map**".

- Example.
$$\mathcal{M} = \{\mathrm{N}(\mu, \Sigma) | \, \mu \in \mathbb{R}^n\} \; (\Sigma \in \mathcal{P}_n), \qquad f = \mu \colon \mathcal{M} \to \mathbb{R}^n.$$

This is the estimation of the mean μ of a normal distribution with a fixed covariance Σ in the space $\mathcal{P}_n$ of positive definite real symmetric matrix. Then the updating map is $\varphi(z, V) = \mathrm{N}(z, \Sigma)V \in \mathcal{V}$, namely,
$$V \mapsto \mathrm{N}(z_1, \Sigma)V \mapsto \mathrm{N}(z_2, \Sigma)\mathrm{N}(z_1, \Sigma)V \mapsto \cdots.$$
(The symmetry $\mathrm{N}(\mu, \Sigma)(z) = \mathrm{N}(z, \Sigma)(\mu)$ implies $\rho(z) = \mathrm{N}(z, \Sigma)$.)

Geometric setting (3) Conjugate prior

- Our subject is the **updating map** $\varphi\colon \mathbb{R}^n \times \mathcal{V} \ni (z, V) \mapsto \rho(z)V \in \mathcal{V}$.
- **Conjugate prior** is a proper subset $\tilde{\mathcal{U}} \subset \mathcal{V}$ with $\varphi\left(\mathbb{R}^n \times \tilde{\mathcal{U}}\right) \subset \tilde{\mathcal{U}}$. Putting $\mathcal{U} = \left\{V \in \tilde{\mathcal{U}} \mid \int_{\mathcal{M}} V = 1\right\}$, we have $\tilde{\mathcal{U}} = \{kV \mid V \in \mathcal{U},\ k > 0\}$.
- Example. $\mathcal{M} = \{\mathrm{N}(\mu, \Sigma) | \, \mu \in \mathbb{R}^n\}$ $(\Sigma \in \mathcal{P}_n)$, $f = \mu\colon \mathcal{M} \to \mathbb{R}^n$. Put $\mathcal{U} = \{\mathrm{N}(m, A) d\mathrm{Vol} \mid m \in \mathbb{R}^n,\ A \in \mathcal{P}(n, \mathbb{R})\} \Rightarrow \tilde{\mathcal{U}}$: conjugate prior.
- Suppose that the conjugate prior $\tilde{\mathcal{U}}$ is a manifold. We fix a ``distance'' $\widetilde{D}\colon \tilde{\mathcal{U}} \times \tilde{\mathcal{U}} \to \mathbb{R}$, which satisfies **non** of the axioms of distance, as

$$\widetilde{D}(V_1, V_2) = \int_{\mathbb{R}^n} V_1 \ln \frac{V_2}{V_1} \text{ (the relative entropy)}$$

- The restriction $\widetilde{D}|_{\mathcal{U} \times \mathcal{U}} = D$ is non-negative (**KL-divergence**).

Geometric setting (4) Bayesian Information Geom.

- Each $y \in \mathcal{M}$ presents a **volume form** $\rho_y d\mathrm{Vol}$ on $\mathbb{R}^n (\ni z)$.
- Given points $z_1, z_2, \ldots \in \mathbb{R}^n$, one updates the prior $P \in \tilde{\mathcal{U}}$ as

$$P \ \mapsto \ \rho(z_1)P \ \mapsto \ \rho(z_2)\rho(z_1)P \ \mapsto \cdots \ \big(\rho(z)(y) = \rho_y(z)\big).$$

This corresponds to a point move on $\mathcal{U}$ by normalizing the density.

- **Generalized IG** = the Fisher metric g & α-connections $\nabla - \alpha g^* T$

  g: the quadratic term of $\widetilde{D}(P, P + dP) + \widetilde{D}(P + dP, P)$

  T: the cubic term of $3\widetilde{D}(P, P + dP) - 3\widetilde{D}(P + dP, P)$

The usual IG looks at the restrictions to the hypersurface $\mathcal{U}$.

Bayesian IG is the geometric study on the updating maps in $\tilde{\mathcal{U}}$ and $\mathcal{U}$.

Example of Bayesian IG (1) Two operations

- $\mu = y$ presents $\exp\left(-\frac{1}{2}(z-\mu)^{\mathrm{T}}\Sigma^{-1}(z-\mu)\right) d\mathrm{Vol}$ on $\mathbb{R}^n (\ni z)$.
- We have $\mathcal{U} = \{\mathrm{N}(m, A) d\mathrm{Vol} \mid m \in \mathbb{R}^n,\ A \in \mathcal{P}(n, \mathbb{R})\}$ and $\tilde{\mathcal{U}}$.
- If z repeats, the agent updates e.g. $\exp\left(-\frac{1}{2v}\|\mu\|^2\right) d\mathrm{Vol} \in \tilde{\mathcal{U}}$ into

$$\rho(z)^n V = \exp\left(-\frac{1}{2v}\|\mu\|^2 - \frac{n}{2}(\mu - y)^{\mathrm{T}}\,\Sigma^{-1}(\mu - y)\right) d\mathrm{Vol}.$$

- Two operations on $\mathcal{U} = \{\mathrm{N}(m, A) d\mathrm{Vol} \mid m \in \mathbb{R}^n,\ A \in \mathcal{P}(n, \mathbb{R})\}$:

  "$*$" from the **convolution** $\mathrm{N}(m, A) * \mathrm{N}(m', A')$ presenting $z + z'$,

  "$\cdot$" from the **normalized pointwise product** $k\mathrm{N}(m, A) \cdot \mathrm{N}(m', A')$.

The above updating roughly corresponds to the **iteration of "$\cdot$" on $\mathcal{U}$.**

Example of Bayesian IG (2) Symmetry of D

- Assume $n = 1$ (temporarily). Write $P = (m, s) \in \mathcal{U}$, where $A = s^2$.

$$\mathrm{N}(m, A) * \mathrm{N}(m', A') = \mathrm{N}(m + m', A + A') \Rightarrow P * P' = \left(m + m', \sqrt{s^2 + s'^2}\right)$$

$$\mathrm{N}(m, A) \cdot \mathrm{N}(m', A') = \mathrm{N}\left(\frac{mA' + Am'}{A + A'}, \frac{AA'}{A + A'}\right) \Rightarrow P \cdot P' = \left(\frac{ms'^2 + m's^2}{s^2 + s'^2}, \frac{ss'}{\sqrt{s^2 + s'^2}}\right)$$

- The correspondence $F = \left\{\left((m, s), (M, S)\right) \mid \frac{m}{s} + \frac{M}{S} = 0,\ sS = 1\right\}$ defines a diffeomorphism of $\mathcal{U}$ which **interchanges** "$*$" and "$\cdot$", i.e.,

$$(p, P), (p', P') \in F(\subset \mathcal{U} \times \mathcal{U}) \Rightarrow (p * p', P \cdot P'), (p \cdot p', P * P') \in F.$$

- Take the "stereogram" $f(p, P) = D(p, p')$ of D under $(p', P) \in F$. Then $f: \mathcal{U} \times \mathcal{U} \to \mathbb{R}$ is **preserved** under the transformations

$$\left((m, s), (M, S)\right) \mapsto \left((e^t m, e^t s), (e^{-t} M, e^{-t} S)\right)\ (t \in \mathbb{R})$$

Perhaps this is the first found symmetry of the KL-divergence D.

Example of Bayesian IG (3) Symplectic geometry

- The space $\mathcal{U} \times \mathcal{U}$ carries the **positive&negative symplectic structures** $d\lambda_\pm = \frac{dm \wedge ds}{s^2} \pm \frac{dM \wedge dS}{S^2}$ and their primitives $\lambda_\pm = \frac{dm}{s} \pm \frac{dM}{S}$.
- Restricting the primitives $\lambda_\pm$ to the hypersurface $N = \{sS = 1\}(\supset F)$, we obtain a **bi-contact structure**, i.e., a transverse pair of positive & negative contact structures. Then $\lambda_\pm$ are their natural extensions.
- In general, a contact form η & a function h on a manifold M determine the **contact Hamiltonian** vector field X via $\boldsymbol{\eta(X) = h}$ & $\boldsymbol{\eta \wedge \mathcal{L}_X \eta = 0}$. X is the push-forward of the Hamiltonian vector field of $e^t h$ on the product $\mathbb{R}(\ni \mathrm{t}) \times M$ with respect to the symplectic form $d(e^t \eta)$.
- $s = e^{-t-u}, S = e^{-t+u} \Rightarrow \lambda_\pm = e^t(e^u dm \pm e^{-u} dM),\ N = \{t = 0\}$.
- Unless $h = h(m, s)$, there is no **bi-contact Hamiltonian** vector field.

Example of Bayesian IG (4) The Bayesian flow

- The correspondence $F \subset \mathcal{U} \times \mathcal{U}$ is **Lagrangian** with respect to $d\lambda_{-}$.
- There is a bi-contact Hamiltonian flow **preserving the correspondence** $F \subset N(\subset \mathcal{U} \times \mathcal{U})$. It is the one for the function $h = k\frac{m}{s}$ $(k \in \mathbb{R})$.
- The restriction of the flow to the correspondence F can be presented by a flow on the second factor. Then the flow interpolates the iteration of "$\cdot$" product in a logarithmic time. Thus we call it the **Bayesian flow**.
- The diffeomorphism of $\mathcal{U}$ defined by $F \subset \mathcal{U} \times \mathcal{U}$ **sends any e-geodesic to an e-geodesic** (as a image). Particularly, the iteration of "$*$" product is a discretization of an e-geodesic, which the diffeomorphism sends to a flow line of the above Bayesian flow.
- This has an application concerning **the smoothness of a smoothing**.

Example of Bayesian IG (5) Multivariate Case

- Take **the extended Cholesky decomposition** of the covariance A.
- This defines the fiber-bundle projection (and therefore the foliation by fibers) of the space of normal distributions to the **unitriangular group**.
- Then the fibers (i.e., the leaves) have special properties:
  - They are **affine** (thus flat) with respect to the e-connection.
  - They are **closed** under the two operations "$*$" and "$\cdot$".
  - The product of any two leaves carries a pair of symplectic forms, the Lagrangian correspondence, the bi-contact hypersurface, and the **Bayesian bi-contact Hamiltoninan flow**.
- The Bayesian approach could explain **the extra dimensions** in physics.