FY2023/2023年度

Functional Analytic Learning Team

Ha Quang Minh

関数解析的学習チーム ハクァンミン

RIKEN

Center for Advanced Intelligence Project

Main Research Directions

1) Mathematical theory and structures coming from/related to **Infinite-dimensional Information Geometry** and **Optimal Transport**, in particular in the setting of **reproducing kernel Hilbert spaces** (RKHS), **infinite-dimensional Gaussian measures**, and **Gaussian processes**.

2) Theory of RKHS and related methodologies in machine learning and statistics

Focus: Information Geometry in Statistics and Machine Learning

- Our focus is on Information Geometry in infinite dimension, in particular in the reproducing kernel Hilbert space (RKHS) and stochastic process settings. In the setting of Gaussian measures on Hilbert space and Gaussian processes, many quantities admit explicit formulas.
- The regularized formulation possesses many favorable theoretical properties, such as dimension-independent convergence for empirical estimations.

Fisher-Rao metric on the set of Gaussian densities (with respect to Lebesgue measure) on $\mathbb{R}^n$. Let $\mathrm{Sym}^{++}(n)$ denote the set of $n \times n$ SPD (symmetric, positive definite) matrices. There is a one-to-one correspondence between $\mathrm{Sym}^{++}(n)$ and the set $\mathcal{S}$ of multivariate zero-mean Gaussian densities on $\mathbb{R}^n$, $\mathcal{S} = \left\{P(x;\theta) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma(\theta))}} \exp\left(-\frac{1}{2}x^T\Sigma(\theta)^{-1}x\right), \theta \in \Theta\right\}$, where $\Theta = \left\{\theta = [\theta^1, \dots, \theta^k], k = \frac{n(n+1)}{2} : \Sigma(\theta) \in \mathrm{Sym}^{++}(n)\right\}$. The Fisher information matrix is defined by $g_{ij}(\theta) = \int_{\mathbb{R}^n} \frac{\partial \ln P(x;\theta)}{\partial\theta^i}\frac{\partial \ln P(x;\theta)}{\partial\theta^j} P(x;\theta)dx$, $1 \leq i,j \leq k$. If g is strictly positive, it defines a Riemannian metric on $\mathcal{S}$, called *Fisher-Rao metric*, or *Fisher information metric*. Explicit expression for Fisher-Rao metric on $\mathcal{S}$: $g_{ij}(\theta) = \frac{1}{2}\mathrm{tr}\left[\Sigma^{-1}\left(\frac{\partial}{\partial\theta^i}\Sigma\right)\Sigma^{-1}\left(\frac{\partial}{\partial\theta^j}\Sigma\right)\right]$, $1 \leq i,j \leq k$. This corresponds to the affine-invariant Riemannian metric on $\mathrm{Sym}^{++}(n)$

$$\langle A, B\rangle_\Sigma = \frac{1}{2}\langle\Sigma^{-1/2}A\Sigma^{-1/2}, \Sigma^{-1/2}B\Sigma^{-1/2}\rangle_F = \frac{1}{2}\mathrm{tr}(\Sigma^{-1}A\Sigma^{-1}B), \quad A, B \in \mathrm{Sym}(n) \cong T_\Sigma(\mathrm{Sym}^{++}(n))$$

Information Geometry of infinite-dimensional Gaussian measures - sample results

Challenges in the infinite-dimensional setting: (i) no Lebesgue measure on infinite-dimensional Hilbert space; (ii) density functions do not exist; (iii) we **cannot** define the Fisher-Rao metric on the set $\mathrm{Gauss}(\mathcal{H})$ of all Gaussian measures on a Hilbert space $\mathcal{H}$.

Solution: Consider set of probability measures equivalent to a fixed measure μ_0, so that Radon-Nikodym densities with respect to μ_0 exist. Let $C_0 \in \mathrm{Sym}^+(\mathcal{H}) \cap \mathrm{Tr}(\mathcal{H})$ be fixed, with $\ker(C_0) = 0$. Let $\mu_0 = \mathcal{N}(0, C_0)$ be the corresponding Gaussian measure. Define the following set

$$\mathrm{SymHS}(\mathcal{H})_{<I} = \{S : S \in \mathrm{Sym}(\mathcal{H}) \cap \mathrm{HS}(\mathcal{H}), I - S > 0\} \subset \mathrm{HS}(\mathcal{H}) \text{ (set of Hilbert-Schmidt operators)}$$

which is a Hilbert manifold. Define the following subset of the set of positive, trace class operators

$$\mathrm{Tr}(\mathcal{H}, C_0) = \{C \in \mathrm{Sym}^+(\mathcal{H}) \cap \mathrm{Tr}(\mathcal{H}) : C = C_0^{1/2}(I-S)C_0^{1/2} \text{for some } S \in \mathrm{SymHS}(\mathcal{H})_{<I}\}.$$

The corresponding set of Gaussian measures equivalent to μ_0

$$\mathrm{Gauss}(\mathcal{H}, \mu_0) = \{\mu = \mathcal{N}(0, C), \;\; C \in \mathrm{Tr}(\mathcal{H}, C_0)\}$$

This is an infinite-dimensional statistical manifold parametrized by $S \in \mathrm{SymHS}(\mathcal{H})_{<I}$. For a fixed $S \in \mathrm{SymHS}(\mathcal{H})_{<I}$ and $\mu = \mathcal{N}(0, C_0^{1/2}(I-S)C_0^{1/2})$, the Fisher-Rao metric at S is defined to be, for $V_1, V_2 \in T_S(\mathrm{SymHS}(\mathcal{H})_{<I}) \cong \mathrm{Sym}(\mathcal{H}) \cap \mathrm{HS}(\mathcal{H})$, the tangent space of $\mathrm{SymHS}(\mathcal{H})_{<I}$ at S,

$$g_S(V_1, V_2) = \int_{\mathcal{H}} D\log\left\{\frac{d\mu}{d\mu_0}(x)\right\}(S)(V_1) D\log\left\{\frac{d\mu}{d\mu_0}(x)\right\}(S)(V_2)d\mu(x)$$

Theorem 1 (Riemannian metric). *Let $S \in \mathrm{SymHS}(\mathcal{H})_{<I}$ be fixed. Then*

$$g_S(V_1, V_2) = \frac{1}{2}\mathrm{tr}[(I-S)^{-1}V_1(I-S)^{-1}V_2], \quad V_1, V_2 \in \mathrm{Sym}(\mathcal{H}) \cap \mathrm{HS}(\mathcal{H}).$$

The corresponding Riemannian metric on $\mathrm{Tr}(\mathcal{H}, C_0)$ is given as follows. Let $\Sigma \in \mathrm{Tr}(\mathcal{H}, C_0)$ be fixed. For $A_1, A_2 \in T_\Sigma(\mathrm{Tr}(\mathcal{H}, C_0)) \cong \mathrm{SymHS}(\mathcal{H}, C_0) = \mathrm{SymHS}(\mathcal{H}, \Sigma) = \{V = C_0^{1/2}AC_0^{1/2}, A \in \mathrm{SymHS}(\mathcal{H})\}$,

$$\langle A_1, A_2\rangle_\Sigma = \frac{1}{2}\langle\Sigma^{-1/2}A_1\Sigma^{-1/2}, \Sigma^{-1/2}A_2\Sigma^{-1/2}\rangle_{\mathrm{HS}} = \frac{1}{2}\mathrm{tr}(\Sigma^{-1/2}A_1\Sigma^{-1}A_2\Sigma^{-1/2}).$$

There is a unique geodesic connecting any pair $A \in \mathrm{Tr}(\mathcal{H}, C_0)$, $B = A^{1/2}(I-S)A^{1/2} \in \mathrm{Tr}(\mathcal{H}, C_0)$

$$\gamma_{AB}(t) = A^{1/2}\exp[t\log(I-S)]A^{1/2}.$$

The length of this geodesic is the Riemannian distance between A and B

$$d_{\mathrm{FR}}(A, B) = \frac{1}{\sqrt{2}}||\log(A^{-1/2}BA^{-1/2})||_{\mathrm{HS}} = \frac{1}{\sqrt{2}}||\log(I-S)||_{\mathrm{HS}}.$$

Connection with the affine-invariant Riemannian geometry of positive definite Hilbert-Schmidt operators: *for* $d_{\mathrm{aiHS}}[(\gamma_1 I + A), (\gamma_2 I + B)] = ||\log[(A + \gamma_1 I)^{-1/2}(B + \gamma_2 I)(A + \gamma_1 I)^{-1/2}]||_{\mathrm{HS}_X}$,

$$\lim_{\gamma \to 0^+} d_{\mathrm{aiHS}}[(\gamma I + A), (\gamma I + B)] = ||\log(I-S)||_{\mathrm{HS}}, \;\; \text{for} \;\; B = A^{1/2}(I-S)A^{1/2} \in \mathrm{Tr}(\mathcal{H}, C_0)$$

Gaussian process setting

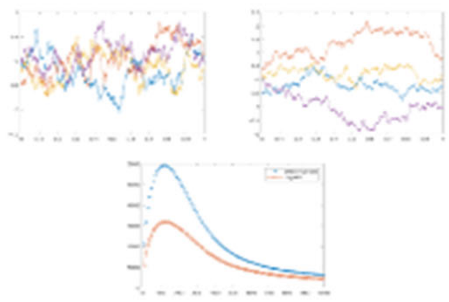
Let T = compact metric space, ν = non-degenerate Borel probability measure on T. Consider the Gaussian process $\xi = (\xi)_{t\in T} = (\xi(\omega, t))_{t\in T}$ on a probability space $(\Omega, \mathcal{F}, P)$ with mean function $m(t) = \mathbb{E}\xi(t)$ and covariance function $K(s,t) = \mathbb{E}[(\xi(s) - m(s))(\xi(t) - m(t))]$. Assume that $\int_T m^2(t)d\nu(t) < \infty$, $\int_T K(t,t)d\nu(t) < \infty$. There is a one-to-one correspondence between measurable Gaussian process $\mathrm{GP}(m, K) \iff \mathcal{N}(m, C_K)$ (Gaussian measure) on $\mathcal{H} = \mathcal{L}^2(T, \nu)$, with covariance operator $(C_K f)(s) = \int_T K(s,t)f(t)d\nu(t)$.

Distances/divergences between Gaussian processes $\xi^i = \mathrm{GP}(m_i, K^i)$, $i = 1, 2$

$$D_{\mathrm{GP}}(\xi^1||\xi^2) = D(\mathcal{N}(m_1, C_{K^1})||\mathcal{N}(m_2, C_{K^2}))$$

Theorem 2 (Dimension-independent sample complexity for regularized version). *Let $\gamma \in \mathbb{R}, \gamma > 0$ be fixed. Let $\mathbf{X} = (x_j)_{j=1}^m$ be independently sampled from (T, ν). For any $0 < \delta < 1$, with probability at least $1 - \delta$,*

$$\left|d_{\mathrm{aiHS}}\left[\left(\gamma I + \frac{1}{m}K^1[\mathbf{X}]\right), \left(\gamma I + \frac{1}{m}K^2[\mathbf{X}]\right)\right] - d_{\mathrm{aiHS}}[(\gamma I + C_{K^1}), (\gamma I + C_{K^2})]\right|$$
$$\leq \frac{1}{\gamma^2}\left(1 + \frac{\kappa_1^2}{\gamma}\right)^3\left[(\kappa_1 + \kappa_2)^2 + \frac{\kappa_1^2\kappa_2^2}{\gamma}\right] \times \left(\kappa_1 + \kappa_2 + \frac{\kappa_1\kappa_2}{\gamma}\right)^2\left[\frac{2\log\frac{6}{\delta}}{m} + \sqrt{\frac{2\log\frac{6}{\delta}}{m}}\right].$$



Samples of the centered Gaussian processes $\mathcal{N}(0, K^1)$, $\mathcal{N}(0, K^2)$ on $T = [0, 1]$ and approximations of squared distances between them. Left: $K^1(x,y) = \exp(-a||x-y||)$, $a = 1$. Right: $K^2(x,y) = \exp(-a||x-y||)$, $a = 1.2$. Here the number of sample paths is $N = 10, 20, \dots, 1000$, and $\gamma = 10^{-7}$

References

[1] H.Q.Minh. *Fisher-Rao Riemannian geometry of equivalent Gaussian measures on Hilbert space*, International Conference on Geometric Science of Information (GSI 2023), Saint Malo, France, August 2023.

[2] H.Q.Minh. *Fisher-Rao geometry of equivalent Gaussian measures on infinite-dimensional Hilbert spaces*, preprint, extended journal submission (59 pages), under review.

[3] H.Q.Minh. *Infinite-dimensional distances and divergences between positive definite operators, Gaussian measures, and Gaussian processes*, preprint under journal review.