FY2022/2022年度

Functional Analytic Learning Unit

Ha Quang Minh

関数解析的学習ユニット 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 & Optimal Transport in Machine Learning & Statistics

- Our current focus is on Information Geometry and Optimal Transport in the infinite-dimensional setting, in particular in the reproducing kernel Hilbert space (RKHS) and Gaussian process settings. For Gaussian measures on Hilbert space and Gaussian processes, many quantities of interest admit explicit formulas that can be efficiently computed.
- Regularization formulations of distances/divergences in both Information Geometry and Optimal Transport possess many favorable theoretical properties over exact formulations, such as dimension-independent sample complexities for empirical estimations.

Case study: Regularized Rényi and Kullback-Leibler (KL) divergences

For two Gaussian densities $\mathcal{N}(m_1, C_1), \mathcal{N}(m_2, C_2)$ on $\mathbb{R}^n$, the Rényi and KL divergences are

$$D_{\mathrm{R},r}(\mathcal{N}(m_1,C_1)||\mathcal{N}(m_2,C_2)) = \frac{1}{2}\langle m_2 - m_1, [(1-r)C_1 + rC_2]^{-1}(m_2-m_1)\rangle + \frac{1}{2}d^{2r-1}_{\mathrm{logdet}}(C_1,C_2),$$

$$\mathrm{KL}(\mathcal{N}(m_1,C_1)||\mathcal{N}(m_2,C_2)) = \frac{1}{2}\langle m_2 - m_1, C_2^{-1}(m_2-m_1)\rangle + \frac{1}{2}d^{1}_{\mathrm{logdet}}(C_1,C_2).$$

Here $d^{\alpha}_{\mathrm{logdet}}(A,B) = \frac{4}{1-\alpha^2}\log\frac{\det(\frac{1-\alpha}{2}A + \frac{1+\alpha}{2}B)}{\det(A)^{\frac{1-\alpha}{2}}\det(B)^{\frac{1+\alpha}{2}}}$, $-1 < \alpha < 1$, is the Alpha Log-Determinant divergences with limiting cases $d^1_{\mathrm{logdet}}(A,B) = \lim_{\alpha\to 1} d^{\alpha}_{\mathrm{logdet}}(A,B) = \mathrm{tr}(B^{-1}A - I) - \log\det(B^{-1}A)$ $d^{-1}_{\mathrm{logdet}}(A,B) = \lim_{\alpha\to -1} d^{\alpha}_{\mathrm{logdet}}(A,B) = \mathrm{tr}(A^{-1}B - I) - \log\det(A^{-1}B)$. These formulas do **not** generalize to the infinite-dimensional setting since tr and log det are not always well-defined.

Regularized Rényi and KL divergences between Gaussian measures on a Hilbert space $\mathcal{H}$. Let $m_1, m_2 \in \mathcal{H}$ and $C_1, C_2 \in \mathrm{Sym}^+(\mathcal{H}) \cap \mathrm{Tr}(\mathcal{H})$. Let $\gamma \in \mathbb{R}, \gamma > 0$ be fixed. The *regularized Rényi divergence* of order r, between the Gaussian measures $\mathcal{N}(m_1,C_1), \mathcal{N}(m_2,C_2)$, is defined to be

$$D^{\gamma}_{\mathrm{R},r}[\mathcal{N}(m_1,C_1)||\mathcal{N}(m_2,C_2)] = \frac{1}{2}\langle m_1 - m_2, [(1-r)(C_1+\gamma I) + r(C_2+\gamma I)]^{-1}(m_1-m_2)\rangle + \frac{1}{2}d^{2r-1}_{\mathrm{logdet}}[(C_1+\gamma I),(C_2+\gamma I)],\ 0 \le r \le 1.$$

Here for $A, B \in \mathrm{Sym}(\mathcal{H}) \cap \mathrm{Tr}(\mathcal{H})$, $A + \gamma I > 0, B + \gamma I > 0$,

$$d^{\alpha}_{\mathrm{logdet}}[(A+\gamma I),(B+\gamma I)] = \frac{4}{1-\alpha^2}\log\left[\frac{\mathrm{det_X}(\frac{1-\alpha}{2}(A+\gamma I) + \frac{1+\alpha}{2}(B+\gamma I))}{\mathrm{det_X}(A+\gamma I)^{\frac{1-\alpha}{2}}\mathrm{det_X}(B+\gamma I)^{\frac{1+\alpha}{2}}}\right],$$

$$d^{1}_{\mathrm{logdet}}[(A+\gamma I),(B+\gamma I)] = \mathrm{tr_X}[(B+\gamma I)^{-1}(A+\gamma I) - I] - \log\mathrm{det_X}[(B+\gamma I)^{-1}(A+\gamma I)],$$

$$d^{-1}_{\mathrm{logdet}}[(A+\gamma I),(B+\gamma I)] = \mathrm{tr_X}[(A+\gamma I)^{-1}(B+\gamma I) - I] - \log\mathrm{det_X}[(A+\gamma I)^{-1}(B+\gamma I)].$$

Here $\mathrm{tr_X}(A+\gamma I) = \mathrm{tr}(A) + \gamma$, $\mathrm{det_X}(A+\gamma I) = \gamma\det\left(I + \frac{A}{\gamma}\right)$ are the extended trace and extended Fredholm determinant. The regularized divergences are finite for all pairs of Gaussian measures.

Theorem 1. *Let $\nu_0 = \mathcal{N}(m, C_0)$, $\nu = \mathcal{N}(m, C)$, $m \in \mathcal{H}$, $C_0, C \in \mathrm{Sym}^{++}(\mathcal{H}) \cap \mathrm{Tr}(\mathcal{H})$, be two equivalent Gaussian measures, that is $m - m_0 \in \mathrm{range}(C_0^{1/2})$ and there exists $S \in \mathrm{Sym}(\mathcal{H}) \cap \mathrm{HS}(\mathcal{H})$ such that $C = C_0^{1/2}(I-S)C_0^{1/2}$. Then, with the Hilbert-Carleman determinant $\det_2(I-S) = \det[(I-S)\exp(S)]$,*

$$\lim_{\gamma\to 0^+} D^{\gamma}_{\mathrm{R},r}(\nu||\nu_0) = D_{\mathrm{R},r}(\nu||\nu_0) = \frac{1}{2r(1-r)}\log\det[(I-S)^{r-1}(I-(1-r)S)],$$

$$\lim_{\gamma\to 0^+} \mathrm{KL}^{\gamma}(\nu||\nu_0) = \mathrm{KL}(\nu||\nu_0) = -\frac{1}{2}\log\det_2(I-S).$$

Theorem 2 (Estimation of regularized Rényi and KL divergences between centered Gaussian processes from finite covariance matrices). *Let $\gamma \in \mathbb{R}, \gamma > 0$ be fixed. Let $0 < r < 1$ 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 $1 - \delta$,*

$$\left|D^{\gamma}_{\mathrm{R},r}\left[\mathcal{N}\left(0, \frac{1}{m}K^1[\mathbf{X}]\right)\Big|\Big|\mathcal{N}\left(0,\frac{1}{m}K^2[\mathbf{X}]\right)\right] - D^{\gamma}_{\mathrm{R},r}[\mathcal{N}(0,C_{K^1})||\mathcal{N}(0,C_{K^2})]\right|$$

$$\le \frac{1}{2\gamma^2}\left[\frac{\kappa_1^4}{r} + \frac{\kappa_2^4}{1-r} + \frac{[(1-r)\kappa_1^2 + r\kappa_2^2]^2}{r(1-r)}\right]\left(\frac{2\log\frac{6}{\delta}}{m} + \sqrt{\frac{2\log\frac{6}{\delta}}{m}}\right).$$

$$\left|\mathrm{KL}^{\gamma}\left[\mathcal{N}\left(0, \frac{1}{m}K^1[\mathbf{X}]\right)\Big|\Big|\mathcal{N}\left(0,\frac{1}{m}K^2[\mathbf{X}]\right)\right] - \mathrm{KL}^{\gamma}[\mathcal{N}(0,C_{K^1})||\mathcal{N}(0,C_{K^2})]\right|$$

$$\le \frac{1}{2\gamma^2}\left[\kappa_1^4 + \kappa_2^4 + \kappa_1^2\kappa_2^2\left(2 + \frac{\kappa_2^2}{\gamma}\right)\right]\left(\frac{2\log\frac{6}{\delta}}{m} + \sqrt{\frac{2\log\frac{6}{\delta}}{m}}\right).$$

It is **not** clear how to obtain sample complexities for the exact divergences.

Gaussian process setting

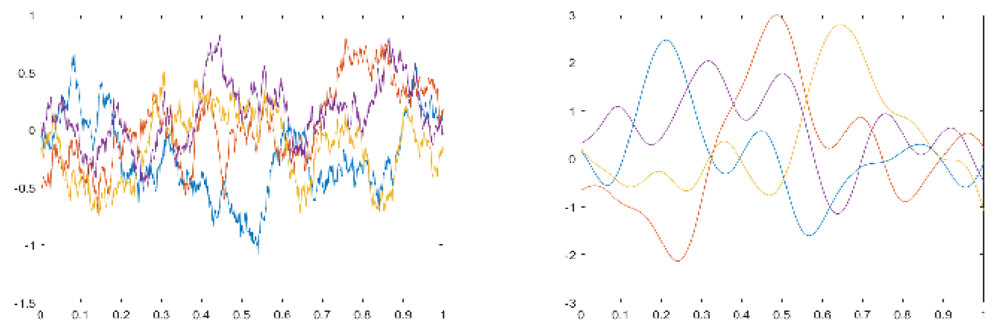
Let T = compact metric space (in general σ-compact metric space), ν = nondegenerate Borel probability measure on T. Consider the Gaussian process $\xi = (\xi)_{t\in T} = (\xi(\omega,t))_{t\in T}$ on a probability space $(\Omega, \mathcal{F}, P)$ with mean function $m(t) = \mathbb{E}\xi(t)$ and covariance 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 $\mathrm{GP}(m,K) \Longleftrightarrow \mathcal{N}(m, C_K)$ (Gaussian measure) on $\mathcal{H} = \mathcal{L}^2(T,\nu)$, with covariance operator

$$(C_K f)(s) = \int_T K(s,t)f(t)d\nu(t)$$

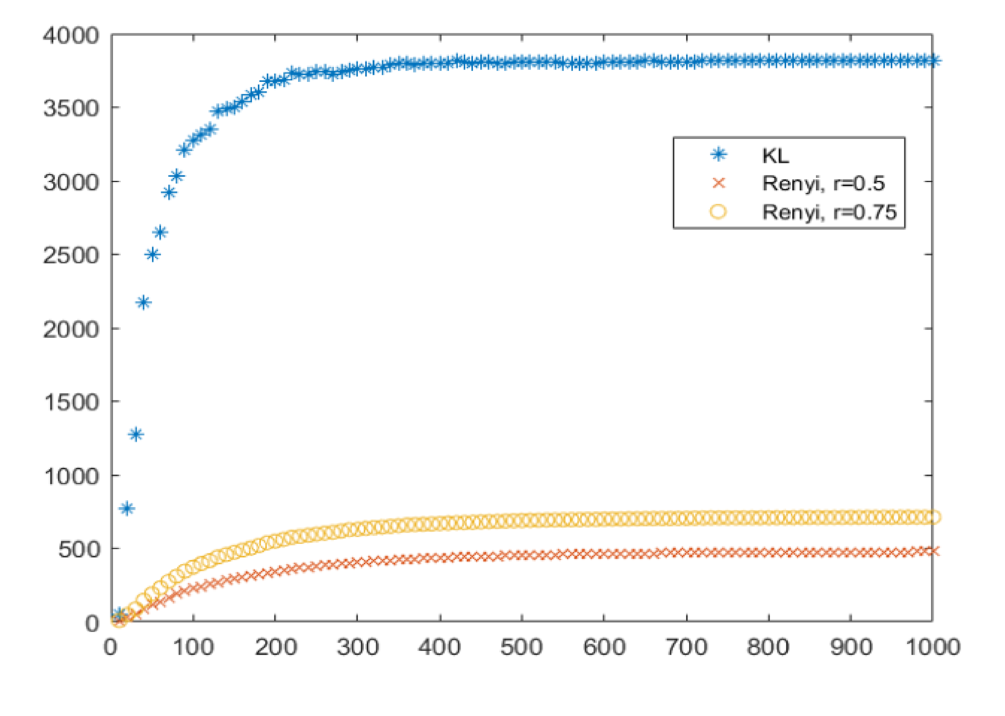
Distances/divergences between Gaussian processes $\xi^i = \mathrm{GP}(m_i, K^i)$, $i = 1,2$

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

Finite-dimensional approximations rely crucially on RKHS covariance and cross-covariance operators $R_{ij} = R_{K^i}R^*_{K^j} : \mathcal{H}_{K^j} \to \mathcal{H}_{K^i}$, $R_{ij} = \int_T (K^i_t \otimes K^j_t)d\nu(t)$, $R_{ij}f(x) = \int_T K^i(x,t)f(t)d\nu(t)$, $f \in \mathcal{H}_{K^j}$.



Samples of the centered Gaussian processes $\mathrm{GP}(0,K^1)$, $\mathrm{GP}(0,K^2)$ on $T = [0,1]$. Left: $K^1(x,y) = \exp(-a||x-y||)$, $a = 1$. Right: $K^2(x,y) = \exp(-||x-y||^2/\sigma^2)$, $\sigma = 0.1$



Approximate divergences between the above centered Gaussian processes from finite covariance matrices, with $m = 10, 20, \ldots, 1000$. The regularization parameter is $\gamma = 10^{-6}$.

Selected bibliography

[1] H.Q.Minh. Regularized Divergences Between Covariance Operators and Gaussian Measures on Hilbert Spaces. **Journal of Theoretical Probability**, 2021.

[2] H.Q.Minh. Convergence and finite sample approximations of entropic regularized Wasserstein distances in Gaussian and RKHS settings, **Analysis and Applications**, 2022.

[3] H.Q.Minh. Kullback-Leibler and Renyi divergences in reproducing kernel Hilbert space and Gaussian process settings, arXiv preprint, 2022.