Distributional Convergence of Empirical Entropic Optimal Transport and Applications

Javier Cárcamo [*] Antonio Cuevas [†]
Luis Alberto Rodríguez Ramírez [‡]

July 24, 2025

Contents

Abstract

Keywords
MSC 2020 subject classification Primary: 62G20, 62E20 Secondary: 62-08

1 Introduction

The goal of this manuscript is showing the asymptotic behavior of the kernel distance when the parameter of the family is estimated from the data (data-driven parameter). Recall first some notation: $\{k_\lambda : \lambda \in \Lambda\}$ is a family of kernels where Λ is a parameter space to be specified later. For each of the kernels k_λ we will denote by $\mathcal{H}_{k,\lambda}$ its associated RKHS and the unit ball of such space as $\mathcal{F}_{k,\lambda}$. For a given Borel's measure S, the mean embedding is defined as

$$\mu_{\mathrm{S}}(\cdot) = \int_{\mathcal{X}} k_\lambda(\cdot, y)\ \mathrm{dS}(y), \tag{1}$$

[*]EHU-UPV
[†]UAM
[‡]Institute for Mathematical Stochastics, University of Göttingen, Goldschmidtstraße 7, 37077 Göttingen

where the integral is understand in the Pettis' sense. The interest of mean embedding lays in the following definition property: for every $f \in \mathcal{H}_{k,\lambda}$, we have that $\mathrm{S}(f) = \langle f, \mu_{\mathrm{S}} \rangle_{\mathcal{H}_{k,\lambda}}$, where $\langle\ ,\ \rangle_{\mathcal{H}_{k,\lambda}}$ denotes the inner product. In terms of the Riesz's representation theorem for Hilbert's spaces, the mean embedding is the dual element of the integral functional induced by S in $\mathcal{H}_{k,\lambda}$ (provided integrability assumptions).

In Cárcamo et al. (2024) we used the following set of assumptions.

(Reg) *Regularity assumption.* $\mathcal{X}$ is a separable metric space and each kernel is continuous as a real function of one variable (with the other kept fixed).

(Dom) *Dominance assumption.* There exists a constant $c > 0$ such that $k_\lambda \ll c\, k$, for all $\lambda \in \Lambda$. Further, k is bounded on the diagonal, that is, $\sup_{x \in \mathcal{X}} (k(x,x)) < \infty$.

(Ide) *Identifiability assumption.* If $\mathrm{P} \neq \mathrm{Q}$, there exists $\lambda \in \Lambda$ such that $\mu_{\mathrm{P}}^{\lambda} \neq \mu_{\mathrm{Q}}^{\lambda}$.

(Par) *Continuous parametrization.* Λ is a compact subset of $\mathbb{R}^k$ (with $k \in \mathbb{N}$) and, for a fixed $(x,y) \in \mathcal{X} \times \mathcal{X}$, the function $\lambda \mapsto k_\lambda(x,y)$ is continuous from Λ to $\mathbb{R}$.

(Sam) *Sampling scheme.* The sampling scheme is balanced, that is, $\frac{n}{(n+m)} \to \theta$, with $\theta \in [0,1]$, as $n, m \to \infty$.

Under a small variation of them, we will exploit the following Prof. Cárcamo's idea: if

$$\begin{aligned} \sigma(\lambda, \mathrm{P}-\mathrm{Q}) &= \sup_{f \in \mathcal{F}_{k,\lambda}} \left((\mathrm{P}-\mathrm{Q})(f) \right) = \| \mu_{\mathrm{P}} - \mu_{\mathrm{Q}} \|_{\mathcal{H}_{k,\lambda}} \\ &= \left(\int_{\mathcal{X}} \int_{\mathcal{X}} k_\lambda(x,y)\ \mathrm{d}(\mathrm{P}-\mathrm{Q})(y)\ \mathrm{d}(\mathrm{P}-\mathrm{Q})(x) \right)^{1/2}, \end{aligned} \tag{2}$$

we can use the integral expression to compute the derivative explicitly.

Some questions around (2):

1. What is the appropriate domain for the new functional in order to compute the Hadamard directional derivative? As we can see, in (2), the argument of σ has been extended. Additionally, the integral expression is valid for every element of $\ell^\infty(\mathcal{F}_{k,\Lambda})$.

2. What is the new process? Obviously the empirical process is involved in the second argument. But for the first we should have to add assumptions on the parameter estimation (M-estimators, etc).

3. Empirical results, code (`C++`) and so: having the asymptotic distribution under the alternative, we can detect or explore examples where Gretton's heuristics is not working (interaction between the two terms of the limit, see below).

Extension of mean embedding to the space $\mathcal{C}_{\text{bpl}}\left(\mathcal{F}_{k,\Lambda},\rho\right)$

Draft of the proof of differentiability

Our object of desire is the asymptotic behavior of the increment quotient

$$\frac{\sup\limits_{f\in\mathcal{F}_{k,\lambda+t_j\,h_j^{\lambda}}}\left(\left(g^{*}+t_j\,h_j^{g^{*}}\right)(f)\right)-\sup\limits_{f\in\mathcal{F}_{k,\lambda}}\left(g^{*}\left(f\right)\right)}{t_j},\tag{3}$$

where $t_j\searrow 0$, $h_j^{\lambda}\longrightarrow h^{\lambda}$ in Λ and $h_j^{g^{*}}\longrightarrow h^{g^{*}}$ in $\mathcal{C}_{\text{bpl}}\left(\mathcal{F}_{k,\Lambda},\rho\right)$.

References

Cárcamo, J., Cuevas, A., & Rodríguez, L.-A. (2024). A uniform kernel trick for high and infinite-dimensional two-sample problems. *Journal of Multivariate Analysis*, *202*, 105317. https://doi.org/https://doi.org/10.1016/j.jmva.2024.105317

2 Supplementary material