Distributional Convergence of Empirical Entropic Optimal Transport and Applications

Javier Cárcamo * Antonio Cuevas †
Luis Alberto Rodríguez Ramírez ‡

July 29, 2025

Contents

Abstract

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

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

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

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

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, $\sup_{x \in \mathcal{X}} (k(x,x)) < \infty$ (CAMBIAR LA CONDICIÓN DE DOMINANCIA. ES EXCESIVA).

(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 $\mathbb{R}$ MODIFICAR LA REGULARIDAD PARA INCLUIR LO QUE NECESITAMOS SOBRE DIFERENCIABILIDAD.

(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

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

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

$$h^{+,\lambda} = \frac{\mu_{\mathrm{P}}^{\lambda} - \mu_{\mathrm{Q}}^{\lambda}}{\left\|\mu_{\mathrm{P}}^{\lambda} - \mu_{\mathrm{Q}}^{\lambda}\right\|_{\mathcal{H}_{k,\lambda}}}, \tag{3}$$

1.1 Preliminaries

- Reproducing Kernel Hilbert Spaces.
- Mean embedding.
- Empirical process.
- U-statistics.

Generalized mean embedding

Las aplicaciones continuas y prelineales satisfacen la mayoría de propiedades que les pedimos a las medidas de Borel $\mathcal{MB}_{\mathrm{p}}(\mathcal{X})$ para que haya mean embedding.

Lemma 1. *Given* $g \in \mathcal{C}_{\mathrm{bpl}}\left(\mathcal{F}_{k,\Lambda}, \rho\right)$ *and* $\lambda \in \Lambda$

- *For every* $\lambda \in \Lambda$, $\|g\|_{\ell^\infty\left(\mathcal{F}_{k,\lambda}\right)} = \|g\|_{\mathcal{H}_{k,\lambda}^*}$.
-

Proof. The expression $\|g\|_{\ell^\infty\left(\mathcal{F}_{k,\lambda}\right)} = \|g\|_{\mathcal{H}_{k,\lambda}^*}$ is a direct consequence of Dudley (2014, Lemma 2.30, p. 88) □

1.2 State of the art: different approaches to parametric families of kernels

- Median.
- Argmax.
- Rayleigh's quotient.

1.3 Our contribution

- Asymptotic result for data-driven kernel distance.
- Negative result on median heuristic for Gaussian kernel.
- Our proposal: data-driven corrected estimation (with the true asymptotic distribution).

2 Main results

2.1 Differentiability results

Previous lemma to make life easier later.

Lemma 2. *Resultado sobre la propiedad Lipschitz marginal (extensión del lema de Shapiro).*

Proof. $\square$

Differentiability under the null. Continuity of ψ

Theorem 3. *Let us assume that the family of kernels $\{k_\theta : \theta \in \Lambda\}$ satisfies (Dom) and (Par). The mapping ψ in (2) is Hadamard directionally differentiable at $(\lambda, 0)$ tangentially to $\Lambda \times \mathcal{C}_{\text{bpl}}(\mathcal{F}_{k,\Lambda}, \rho)$. In such a case, the (directional) derivative of ψ at $(\lambda, 0)$ is given by*

$$\psi'_{(\lambda,0)}(\zeta, g) = \|g\|_{\ell^\infty(\mathcal{F}_{k,\lambda})} = \|g\|_{\mathcal{H}^*_{k,\lambda}}, \tag{4}$$

with $g \in \mathcal{C}(\mathcal{F}_{k,\Lambda}, \rho)$.

Proof. By definition of Hadamard directional differentiability and Lemma 2, given $g \in \mathcal{C}_{\text{bpl}}(\mathcal{F}_{k,\Lambda}, \rho)$ and $\zeta \in \Lambda$ and sequences $(\zeta_j)_{j\in\mathbb{N}} \in \Lambda^{\mathbb{N}}$ and $(t_j)_{j\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ such that $d_\Lambda(\zeta_j, \zeta) \longrightarrow 0$ and $t_j \searrow 0$ when $j \longrightarrow \infty$ we have to show that

$$\lim_{j\longrightarrow\infty} \frac{\psi(\lambda + t_j\,\zeta_j, t_j\,g) - \psi(\lambda, 0) - t_j\,\psi'_{(\lambda,0)}(\zeta, g)}{t_j} = 0. \tag{5}$$

Under the hypothesis of the theorem, note that

$$\frac{\psi(\lambda + t_j\,\zeta_j, t_j\,g) - \psi(\lambda, 0) - t_j\,\psi'_{(\lambda,0)}(\zeta, g)}{t_j} = \|g\|_{\ell^\infty(\mathcal{F}_{k,\lambda+t_j\,\zeta_j})} - \|g\|_{\ell^\infty(\mathcal{F}_{k,\lambda})}. \tag{6}$$

Now observe that $\|g\|_{\ell^\infty(\mathcal{F}_{k,\lambda+t_j\,\zeta_j})} = \sqrt{g\left(g\left(k_{\lambda+t_j\,\zeta_j}(\cdot_1, \cdot_2)\right)\right)}$. By (Dom), (Par) and dominated convergence theorem, $k_{\lambda+t_j\,\zeta_j}(\cdot_1, y)$ converges to $k_\lambda(\cdot_1, y)$ on the metric ρ for every $y \in \mathcal{X}$. Hence, by continuity of g, $g\left(k_{\lambda+t_j\,\zeta_j}(\cdot_1, y)\right)$ to $g(k_\lambda(\cdot_1, y))$ pointwise in $y \in \mathcal{X}$.

At this point, it is worth to mention that by (Dom)

$$\left|g\left(k_{\lambda+t_j\,\zeta_j}(\cdot_1, y)\right) - g(k_\lambda(\cdot_1, y))\right| \le 2\,\|g\|_{\ell^\infty(\mathcal{F}_{k,\Lambda})} \sqrt{k(y, y)}, \tag{7}$$

so, by dominated convergence theorem the convergence of $g\left(k_{\lambda+t_j\,\zeta_j}(\cdot_1, \cdot_2)\right)$ to $g(k_\lambda(\cdot_1, \cdot_2))$ is also given in the metric ρ. Thanks to the continuity of g, the first part of the proof is ended. $\square$

Differentiability under the alternative

Theorem 4. *Let us assume that the family of kernels* $\{k_\theta : \theta \in \Lambda\}$ *satisfies (Dom), (Ide) and (Par). If* $\mathrm{P}, \mathrm{Q} \in \mathcal{MB}_{\mathrm{p}}(\mathcal{X})$ *such that* $\mathrm{P} \neq \mathrm{Q}$*, then the mapping* ψ *in* (2) *is Hadamard directionally differentiable at* $(\lambda, \mathrm{P}-\mathrm{Q})$ *tangentially to* $\Lambda \times \mathcal{C}_{\mathrm{bpl}}\left(\mathcal{F}_{k,\Lambda}, \rho\right)$*, the subset of* $\ell^\infty\left(\mathcal{F}_{k,\Lambda}\right)$ *constituted by bounded, prelinear and continuous functionals with respect to the distance* ρ *in* (??)*. In such a case, the (directional) derivative of* ψ *at* $(\lambda, \mathrm{P}-\mathrm{Q})$ *is given by*

$$\psi'_{(\lambda,\mathrm{P}-\mathrm{Q})}(\zeta, g) = g\left(h^{+,\lambda}\right) + \frac{1}{2\left\|\mu^\lambda_{\mathrm{P}-\mathrm{Q}}\right\|_{\mathcal{H}_{k,\lambda}}} \int_{\mathcal{X}}\int_{\mathcal{X}} \partial_\lambda k_\lambda(x,y)(\zeta)\, \mathrm{d}(\mathrm{P}-\mathrm{Q})(y)\, \mathrm{d}(\mathrm{P}-\mathrm{Q})(x), \tag{8}$$

with $g \in \mathcal{C}\left(\mathcal{F}_{k,\Lambda}, \rho\right)$ *and* $\zeta \in \Lambda$*; where the functions* $h^{+,\lambda}$ *are defined in* (3).

Proof. By definition of Hadamard directional differentiability and Lemma 2, given $g \in \mathcal{C}_{\mathrm{bpl}}\left(\mathcal{F}_{k,\Lambda}, \rho\right)$ and $\zeta \in \Lambda$ and sequences $(\zeta_j)_{j\in\mathbb{N}} \in \Lambda^{\mathbb{N}}$ and $(t_j)_{j\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ such that $d_\Lambda\left(\zeta_j, \zeta\right) \longrightarrow 0$ and $t_j \searrow 0$ when $j \longrightarrow \infty$ we have to show that

$$\lim_{j\longrightarrow\infty} \frac{\psi\left(\lambda + t_j\,\zeta_j, \mathrm{P}-\mathrm{Q} + t_j\, g\right) - \psi(\lambda, \mathrm{P}-\mathrm{Q}) - t_j\, \psi'_{(\lambda,\mathrm{P}-\mathrm{Q})}(\zeta, g)}{t_j} = 0. \tag{9}$$

To begin with, note that

$$\frac{\psi\left(\lambda + t_j\,\zeta_j, \mathrm{P}-\mathrm{Q} + t_j\, g\right) - \psi(\lambda, \mathrm{P}-\mathrm{Q}) - t_j\, \psi'_{(\lambda,\mathrm{P}-\mathrm{Q})}(\zeta, g)}{t_j} = L_1 + L_2 + L_3, \tag{10}$$

where

$$\begin{aligned}
L_1 &= \frac{\sup_{f\in\mathcal{F}_{k,\lambda}} \left(\left(\mathrm{P}-\mathrm{Q} + t_j\, g\right)(f)\right) - \sup_{f\in\mathcal{F}_{k,\lambda}} \left(\left(\mathrm{P}-\mathrm{Q}\right)(f)\right)}{t_j},\\
L_2 &= \frac{1}{L_4}\int_{\mathcal{X}}\int_{\mathcal{X}} \frac{k_{\lambda+t_j\,\zeta_j}(x,y) - k_\lambda(x,y)}{t_j}\, \mathrm{d}(\mathrm{P}-\mathrm{Q})(y)\, \mathrm{d}(\mathrm{P}-\mathrm{Q})(x),\\
L_3 &= \frac{1}{L_4}\left(g\left(\mu^{\lambda+t_j\,\zeta_j}_{\mathrm{P}-\mathrm{Q}} - \mu^\lambda_{\mathrm{P}-\mathrm{Q}}\right) + \int_{\mathcal{X}} g\left(k_{\lambda+t_j\,\zeta_j}(\cdot,y) - k_\lambda(\cdot,y)\right) \mathrm{d}(\mathrm{P}-\mathrm{Q})(y)\right.\\
&\quad \left. + t_j\left(\|g\|^2_{\ell^\infty\left(\mathcal{F}_{k,\lambda+t_j\,\zeta_j}\right)} - \|g\|^2_{\ell^\infty\left(\mathcal{F}_{k,\lambda}\right)}\right)\right),\\
L_4 &= \left(\left\|\mu^{\lambda+t_j\,\zeta_j}_{\mathrm{P}-\mathrm{Q}+t_j\,g}\right\|_{\mathcal{H}_{k,\lambda+t_j\,\zeta_j}} + \left\|\mu^{\lambda}_{\mathrm{P}-\mathrm{Q}+t_j\,g}\right\|_{\mathcal{H}_{k,\lambda}}\right)^{1/2}.
\end{aligned} \tag{11}$$

The convergence of L_1 was proved in Cárcamo et al. (2024, Lemma 4) and the limit is $g\left(h^{+,\lambda}\right)$, where $h^{+,\lambda}$ was defined in (3). Provided the convergence of L_4, by (Par), L_2 converges to

$$\frac{1}{2\left\|\mu_{\mathrm{P}-\mathrm{Q}}^{\lambda}\right\|_{\mathcal{H}_{k,\lambda}}}\int_{\mathcal{X}}\int_{\mathcal{X}}\partial_\lambda k_\lambda(x,y)(\zeta)\ \mathrm{d}(\mathrm{P}-\mathrm{Q})(y)\ \mathrm{d}(\mathrm{P}-\mathrm{Q})(x). \tag{12}$$

Now we continue with L_4. By definition

$$\begin{aligned}\left\|\mu_{\mathrm{P}-\mathrm{Q}+t_j\,g}^{\lambda+t_j\,\zeta_j}\right\|_{\mathcal{H}_{k,\lambda+t_j\,\zeta_j}} = \bigg(&\int_{\mathcal{X}}\int_{\mathcal{X}} k_{\lambda+t_j\,\zeta_j}(x,y)\ \mathrm{d}(\mathrm{P}-\mathrm{Q})(y)\ \mathrm{d}(\mathrm{P}-\mathrm{Q})(x)\\ &+t_j\,g\left(\mu_{\mathrm{P}-\mathrm{Q}}^{\lambda+t_j\,\zeta_j}\right)\\ &+t_j\int_{\mathcal{X}} g\left(k_{\lambda+t_j\,\zeta_j}(\cdot,y)\right)\ \mathrm{d}(\mathrm{P}-\mathrm{Q})(y)\\ &+t_j^2\,\|g\|_{\ell^\infty\left(\mathcal{F}_{k,\lambda+t_j\,\zeta_j}\right)}\bigg)^{1/2}.\end{aligned} \tag{13}$$

From top to bottom in (13):

1. $\int_{\mathcal{X}}\int_{\mathcal{X}} k_{\lambda+t_j\,\zeta_j}(x,y)\ \mathrm{d}(\mathrm{P}-\mathrm{Q})(y)\ \mathrm{d}(\mathrm{P}-\mathrm{Q})(x)$ converges to $\int_{\mathcal{X}}\int_{\mathcal{X}} k_\lambda(x,y)\ \mathrm{d}(\mathrm{P}-\mathrm{Q})(y)\ \mathrm{d}(\mathrm{P}-\mathrm{Q})(x)=\left\|\mu_{\mathrm{P}-\mathrm{Q}}^{\lambda}\right\|_{\mathcal{H}_{k,\lambda}}^2$ by (Dom) and (Par) by Cauchy-Schwarz's inequality and dominated convergence theorem.

2. By $g\in\mathcal{C}_{\mathrm{bpl}}\left(\mathcal{F}_{k,\Lambda},\rho\right)$

$$\left|g\left(\mu_{\mathrm{P}-\mathrm{Q}}^{\lambda+t_j\,\zeta_j}\right)\right|\leq\|g\|_{\ell^\infty\left(\mathcal{F}_{k,\lambda+t_j\,\zeta_j}\right)}\left\|\mu_{\mathrm{P}-\mathrm{Q}}^{\lambda}\right\|_{\mathcal{H}_{k,\lambda}}, \tag{14}$$

   so by (Dom),

$$\left|g\left(\mu_{\mathrm{P}-\mathrm{Q}}^{\lambda+t_j\,\zeta_j}\right)\right|\leq\|g\|_{\ell^\infty\left(\mathcal{F}_{k,\Lambda}\right)}\int_{\mathcal{X}}\sqrt{k(x,x)}\ \mathrm{d}(\mathrm{P}+\mathrm{Q})(y), \tag{15}$$

   and $t_j\,g\left(\mu_{\mathrm{P}-\mathrm{Q}}^{\lambda+t_j\,\zeta_j}\right)$ goes to 0.

3. Analogously, $\left|g\left(k_{\lambda+t_j\,\zeta_j}\right)\right|\leq\|g\|_{\ell^\infty\left(\mathcal{F}_{k,\Lambda}\right)}\sqrt{k(y,y)}$, so we can conclude that $t_j\int_{\mathcal{X}} g\left(k_{\lambda+t_j\,\zeta_j}(\cdot,y)\right)\ \mathrm{d}(\mathrm{P}-\mathrm{Q})(y)$ is also tending to 0.

4. Finally, since $\|g\|_{\ell^\infty\left(\mathcal{F}_{k,\lambda+t_j\,\zeta_j}\right)}\leq\|g\|_{\ell^\infty\left(\mathcal{F}_{k,\Lambda}\right)}$, then $t_j^2\,\|g\|_{\ell^\infty\left(\mathcal{F}_{k,\lambda+t_j\,\zeta_j}\right)}$ converges to 0.

In conclusion, $L_4 \longrightarrow 2 \left\| \mu^{\lambda}_{\mathrm{P}-\mathrm{Q}} \right\|_{\mathcal{H}_{k,\lambda}}$ when $j \longrightarrow \infty$. To end up this proof, we see the convergence of L_3. Firstly, we tackle $\int_{\mathcal{X}} g\left(k_{\lambda+t_j\,\zeta_j}(\cdot,y) - k_\lambda(\cdot,y)\right)\,\mathrm{d}(\mathrm{P}-\mathrm{Q})(y)$. Pointwise convergence of $k_{\lambda+t_j\,\zeta_j}(\cdot,y)$ to $k_\lambda(\cdot,y)$ is given by (Par) and by (Dom) and Folland (1999, Theorem 2.27), convergence in metric ρ is also given. By continuity of g, $g\left(k_{\lambda+t_j\,\zeta_j}(\cdot,y) - k_\lambda(\cdot,y)\right)$ is converging to 0 pointwise when $j \longrightarrow \infty$. Since this expression is also bounded by the third item of the previous enumeration, by dominated convergence theorem we have the desired limit.

Secondly, convergence $g\left(\mu^{\lambda+t_j\,\zeta_j}_{\mathrm{P}-\mathrm{Q}} - \mu^{\lambda}_{\mathrm{P}-\mathrm{Q}}\right)$ to 0 is to be proved. By (Dom), (Par), and definition of meam embedding, $\mu^{\lambda+t_j\,\zeta_j}_{\mathrm{P}-\mathrm{Q}}$ converges to $\mu^{\lambda}_{\mathrm{P}-\mathrm{Q}}$ pointwise. Now, recall that

$$\left|\mu^{\lambda+t_j\,\zeta_j}_{\mathrm{P}-\mathrm{Q}}(x) - \mu^{\lambda}_{\mathrm{P}-\mathrm{Q}}(x)\right| \leq 2\sqrt{k(x,x)} \int_{\mathcal{X}} \sqrt{k(y,y)}\,\mathrm{d}(\mathrm{P}+\mathrm{Q})(y), \tag{16}$$

by (Dom). So, by virtue of dominated convergence theorem, the convergence is also given in the metric ρ. By continuity of g, this term is also done.

Finally, for the last term it is enough to observe that $\left| \|g\|^2_{\ell^\infty(\mathcal{F}_{k,\lambda+t_j\,\zeta_j})} - \|g\|^2_{\ell^\infty(\mathcal{F}_{k,\lambda})} \right| \leq 2\,\|g\|^2_{\ell^\infty(\mathcal{F}_{k,\Lambda})}$. Hence, L_3 is converging to 0 and the proof is ended. □

2.2 Statistic results

Our asymptotic results

Delta method y palante.

Negative result on the median heuristic for the Gaussian kernel

Con el Teorema 3 y la teoría de Gretton, sale que bajo la nula va a 0.

2.3 Empirical results

3 Notas

1. What happens when the estimated parameter goes to 0 or ∞ in the Gaussian kernel? The limit of the estimated parameter should belong to the parameter space (see Theorem 3). Hablando en plata, bajo la nula y usando la heurística de la mediana, el parámetro va a infinito (la mediana va a 0). Se va a la mierda.

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).

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

Dudley, R. M. (2014). *Uniform central limit theorems* (Vol. 142). Cambridge university press.

Folland, G. B. (1999). *Real analysis: Modern techniques and their applications* (Vol. 40). John Wiley & Sons.

4 Supplementary material