

---

# A data-driven calibration for a non-asymptotic kernel two-sample test

Perrine Lacroix\*<sup>1</sup>

<sup>1</sup>université de Nantes – Université de Nantes - UMR 6629 – France

## Résumé

We observe two populations of multivariate data described by  $p$  variables, where  $p$  is significantly larger than the population sizes. A two-sample test has to be performed to decide between the null hypothesis (the distributions of both populations are equal) and the alternative hypothesis (distributions are different). To take into account the complex structure of variables and overcome the curse of dimensionality problem, data are embedded in a well-chosen Reproducing Kernel Hilbert Space (RKHS). In our work, we study a test statistic inspired by Harchaoui et al. (2008) generalizing the student t-test in a RKHS, and propose a non-asymptotic and implementable method to calibrate the test. First, through a spectral analysis, a theoretical upper bound of the test quantile is proposed. Second, a data-driven algorithm is implemented satisfying a control of the type I error and including the calibration of the unknown regularization hyperparameter. Joint work with Bertrand Michel (Univ. Nantes), Franck Picard (ENS de Lyon) and Vincent Rivoirard (Univ. Paris-Dauphine).

---

\*Intervenant