Beyond Mean Embedding: The Power of Cumulants in RKHSs*

Zoltán Szabó†

Abstract

Maximum mean discrepancy (MMD) and Hilbert-Schmidt independence criterion (HSIC) rely on the mean embedding of probability distributions and are among the most successful approaches in machine learning to quantify the difference and the independence of random variables, respectively. We present higher-order variants of MMD and HSIC by extending the notion of cumulants to reproducing kernel Hilbert spaces. The resulting kernelized cumulants have various benefits: (i) they are able to characterize the equality of distributions and independence under very mild conditions, (ii) they are easy to estimate with minimal computational overhead compared to their degree one (MMD and HSIC) counterparts, (iii) they achieve improved power when applied in two-sample and independence testing for environmental and traffic data analysis.

Paper: arXiv.

*Advanced Course on Data Science and Machine Learning (ACDL), Tuscany, Italy (June 10-14, 2023); abstract.

†This is joint work with Patric Bonnier and Harald Oberhauser.