The Power of Cumulants in Reproducing Kernel Hilbert Spaces[*]

Zoltán Szabó[†]

Abstract

Maximum mean discrepancy (MMD, also called energy distance) and Hilbert-Schmidt independence criterion (HSIC, a.k.a. distance covariance) rely on the mean embedding of probability distributions and are among the most successful approaches in machine learning and statistics to quantify the difference and the independence of random variables, respectively. In this talk, I am going to 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 in two-sample and independence testing as illustrated in the context of environmental and traffic data analysis. [This is joint work with Patric Bonnier and Harald Oberhauser, and appeared at NeurIPS-2023.]

Paper: arXiv.