# Data-Efficient Independence Testing with Analytic Kernel Embeddings*

**Zoltán Szabó**† (CMAP, École Polytechnique)

## Abstract

A central task in the analysis of complex, heterogeneous data is the efficient measurement and testing of statistical dependencies. In my talk, I will present a new easy-to-compute dependency measure and an associated independence test. The independence measure captures the difference of the joint distribution and the product of the marginals via analytic kernel embeddings of distributions, evaluated at finite many locations (features). These features are optimized in a task-specific way, to maximize a proxy on the test power. The result is a linear-time nonparametric independence testing technique which also delivers indicating regions if the independence hypothesis is violated. We illustrate the data-efficiency of the proposed approach in two real-world benchmarks (Million Song Data, VideoStory46K dataset): it achieves comparably performance to the state-of-the-art quadratic-time HSIC test, and outperforms competing $\mathcal{O}(n)$ and $\mathcal{O}(n \log n)$ tests.

- Preprint: `https://arxiv.org/abs/1610.04782`
- Code: `https://github.com/wittawatj/fsic-test`

*PASADENA Seminar, Télécom ParisTech, Paris, France, 17 May 2017; abstract.

†Joint work with Wittawat Jitkrittum and Arthur Gretton.