A Fast Goodness-of-Fit Test with Analytic Kernel Embeddings*

Wittawat Jitkrittum†
†Gatsby Unit, University College London

Zoltán Szabó
CMAP, École Polytechnique

Kenji Fukumizu
Dept. of Mathematical Analysis and Statistical Inference,
The Institute of Statistical Mathematics

Arthur Gretton†

Abstract

We propose a fast, interpretable, nonparametric goodness-of-fit test for testing the null hypothesis that a given sample is distributed according to a specified model (probability density). The test requires only the gradient of the log density of the model; the normalizer of the density is not assumed to be known. Our test extends the recently developed kernelized Stein operator of Liu et al., 2016, Chwialkowski et al., 2016 to produce an analytic witness function observing the differences between the empirical distribution and the model. The test statistic is given by the evaluations of this witness function at finitely many test locations, and can be computed in linear-time (w.r.t. sample size). We show that these test locations can be automatically determined in such a way as to maximize the test power. A unique advantage of the new approach is that the optimized locations also indicate where the empirical distribution differs from the specified model, resulting in a goodness-of-fit test which gives a clue on where the model does not fit the data well. Experiments on restricted Boltzmann machines and Gaussian-vs-Laplace problems verify the efficiency of the new test both in terms of runtime and its test power.

*Greek Stochastics Workshop - Model Determination, Milos, Greece, 14-17 July 2017; abstract.