Minimax Nonparametric Testing in Wasserstein Distance

Tudor Manole
(Carnegie Mellon University)

June 12, 2024

Nonparametric hypothesis testing is a prominent motivation for the development of inferential results for empirical Wasserstein distances. Indeed, Wasserstein distances are increasingly used for tasks such as multivariate two-sample testing and independence testing, and are also implicitly used as test statistics in generative modeling algorithms for machine learning applications. Despite these methodological advances, theory has lagged behind on deriving optimal procedures for testing under the Wasserstein distance, and this talk will aim to provide some new steps in this direction.

We will focus on the goodness-of-fit and two-sample testing problems, where the composite alternative hypotheses are separated from the null in Wasserstein distance. We adopt the minimax perspective, and seek to find the critical testing radii for these problems under various assumptions on the set of alternatives. First, absent any smoothness assumptions, we show that the goodness-of-fit critical radius decays polynomially faster than the corresponding two-sample critical radius. This suggests that the Wasserstein two-sample testing problem is statistically harder than its one-sample counterpart, contrary to the related problem of estimating the Wasserstein distance, for which the one- and two-sample minimax rates coincide. Second, we show how these two critical radii improve when smoothness assumptions are placed on the alternatives. Third, we present some surprises regarding the (sub)optimality of various commonly-used test statistics.