Summary of Progress

Alden Green

August 15, 2019

A brief summary of the progress we've made so far.

1 Progress to date.

We observe samples $x_1, \ldots, x_n \sim P$ and $y_1, \ldots, y_m \sim Q$, (for convenience, assume P and Q are supported on $[0,1]^d$ absolutely continuous with respective densities p and q which are lower bounded away from 0, and pper bounded away from ∞, over their support.) We started with the following IPM-based test statistic; construct a neighborhood graph G with incidence matrix D, and let

$$T_C := \sup_{\theta : \|D\theta\|_2 \leq C} \langle \theta, z \rangle,$$

where $z = 1_X - 1_Y$ is a contrast vector and C is a scaling parameter which controls the size of the function class we compute the IPM over. We made the following alterations to the above setup and test statistic:

- We modified the setup of the problem in several ways to ease tractability of analysis. First, we assumed we observed values $z_i = f(x_i) + \varepsilon_i$, where $x_i \sim (p+q)/2$, and $f = (p-q)/(p+q)$, and ε_i represented the error of a Bernoulli random variable minus its expectation. Then, we let ε_i be normally distributed.
- More fundamentally, we changed the norm dictating the function class over which we take supremum. Let C be such that if $f = (p-q) \in W^{2,1}(L)$, then $\|Df\|_2 \lesssim C$. This choice of C results in the IPM being computed over too rich of a function class, so must restrict the function class further. To do so, we modified the IPM as follows: letting $D^T D = L = V \Lambda V^T$ be the eigendecomposition of the Laplacian matrix L, let $g_C(\lambda) = \infty \cdot \mathbf{1}(\lambda > C)$, and let $g_C(L) = V g_C(\Lambda) V^T$. Then, consider the following choice of test statistic:

  $$T_C := \sup_{\theta : \theta^T g_C(L) \theta \leq C} \langle \theta, z \rangle$$

  or equivalently

  $$T_C := \sum_{k : \lambda_k \leq C^2} \langle v_k, z \rangle$$

For an appropriately formed neighborhood graph, and an appropriate choice of C, does T_C achieve non-trivial power uniformly over a nonparametric function class , that is assume that

$$\|f\|_{W^{2,1}} := \sum_{|\alpha| \leq 1} \|\mathcal{D} f\|_{L^2}^2 \leq L$$