Graph Testing: Notes for the Week of 12/11 - 12/18

Alden Green

December 14, 2018

1 Setup

Data model. We are given two distributions, P and Q, defined over compact set $\mathcal{X} \subset \mathbb{R}^d$, with the ability to sample from either one. Our goal is to test the hypothesis $H_0 : P = Q$ vs. the alternative $H_a : P \neq Q$.

Under the **binomial data model**, our sampling procedure is to draw i.i.d Rademacher labels $L_i \in \{1, -1\}$ for $i \in \{1, \dots, N\}$, and then sample $Z_i \sim P$ if $L_i = 1$ and $Z_i \sim Q$ otherwise. Define 1_X to be the length-N indicator vector for $L_i = 1$

$$1_X[i] = \begin{cases} 1, L_i = 1 \\ 0 \text{ otherwise} \end{cases}$$

and similarly for 1_Y

$$1_Y[j] = \begin{cases} 1, L_i = -1 \\ 0 \text{ otherwise} \end{cases}$$

and define $a = \frac{1_X}{N/2} - \frac{1_Y}{N/2}$.

Under the **fixed label data model** we use the same data generating process as above, except fix $\mathcal{L}_X = \{1, \dots, N/2\}$ and $\mathcal{L}_Y = \{N/2, \dots, N\}$. Say that $L_i = 1$ for $i \in \mathcal{L}_X$ and $L_i = -1$ for $i \in \mathcal{L}_Y$, and call $\{X_1, \dots, X_{|\mathcal{L}_X|}\} = \{Z_i : i \in \mathcal{L}_X\}$ and likewise for Y.

Graph. Form an $N \times N$ Gram matrix A, where $A_{ij} = K(Z_i, Z_j)$ for **kernel function** K. Let $G = (V, E)$ with $V = \{Z_1, \dots, Z_n\}$ and $E = \{A_{ij} : 1 \leq i < j \leq n\}$. Take $L = D - A$ to be the (unnormalized) **Laplacian matrix** of A (where D is the diagonal degree matrix with $D_{ii} = \sum_{j \in [n+m]} A_{ij}$). Let M be the number of non-zero entries of A. Denote by B the $M \times N$ **incidence matrix** of A, where we denote the ith row of B as B_i and set B_i to have entry A_{ij} in position i, $-A_{ij}$ in position j, and 0 everywhere else.

Test statistics. We define our **laplacian smooth** test statistic.

$$T_2 = \left(\max_{\theta : \|B\theta\|_2 \le 1} a^T \theta \right)^2 = a^T L^\dagger a.$$

Distances between probability measures. An **integral probability metric** (IPM) with respect to a function class $\mathcal{F}$ is defined

$$\sup_{f \in \mathcal{F}} \mathbb{E}\left[f(X)\right] - \mathbb{E}\left[f(Y)\right]$$

for $X \sim P$, $Y \sim Q$.

Hereafter, we will assume P and Q are absolutely continuous with respect to Lebesgue measure, with density functions p and q, respectively. Denote the **mixture density** by $\mu = \frac{p+q}{2}$.

Denote the **gradient** of a function f by ∇_x. Then we can define the **Sobolev semi-norm** and **dot product**, $\|f\|_{W_0^{1,2}(\mathcal{X},\mu^2)}$ and $\langle f, g \rangle_{W_0^{1,2}(\mathcal{X},\mu^2)}$, by

$$\langle f, g \rangle_{W_0^{1,2}(\mathcal{X},\mu)} = \int_{\mathcal{X}} \langle \nabla_x f(x), \nabla_x g(x) \rangle_{\mathbb{R}^d} \mu^2(x), \quad \|f\|_{W_0^{1,2}(\mathcal{X},\mu)} = \sqrt{\int_{\mathcal{X}} \|\nabla_x f(x)\|^2 \mu^2(x) dx}$$

Let the **Sobolev space**, $W^{1,2}(\mathcal{X}, \mu^2)$, be

$$W^{1,2}(\mathcal{X}, \mu^2) = \left\{ f : \mathcal{X} \to \mathbb{R}, \int_{\mathcal{X}} \|\nabla_x f(x)\|^2 \mu^2(x) dx < \infty \right\}.$$

and denote by $W_0^{1,2}(\mathcal{X}, \mu^2)$ the restriction of $W^{1,2}(\mathcal{X}, \mu^2)$ to functions which vanish at the boundary of $\mathcal{X}$. Note that $\|f\|_{W_0^{1,2}(\mathcal{X},\mu^2)}$ defines a semi-norm over $W_0^{1,2}(\mathcal{X}, \mu^2)$. Finally, let $B_W(\mathcal{X}, \mu^2)$ be the **unit ball** of $W_0^{1,2}(\mathcal{X}, \mu^2)$, meaning

$$B_W(\mathcal{X}, \mu^2) = \left\{ f \in W_0^{1,2}(\mathcal{X}, \mu^2) : \|f\|_{W_0^{1,2}(\mathcal{X},\mu^2)} \le 1 \right\}$$

Now we can define the **Sobolev IPM**, $\mathcal{S}_{\mu^2}(P, Q)$ It is simply an IPM where the function class is the Sobolev unit ball with respect to μ^2.

$$\mathcal{S}_{\mu^2}(P, Q) \stackrel{\text{def}}{=} \sup_{f \in B_W} \left\{ \mathbb{E}\left[f(X)\right] - \mathbb{E}\left[f(Y)\right] \right\}$$

We will show that the Laplacian constraint $\|B\theta\|_2 \le 1$ is very similar to the constraint $f_\theta \in B_W(X, \mu^2)$ for the right choice of K, over all Holder functions.

Holder functions For mapping $f : \mathbb{R}^d \to \mathbb{R}$ and β a positive integer, we say f is a β-**Holder function** if there exists $C > 0$ such that for all $x, y \in \mathcal{X}$

$$\left| f^{(\beta-1)}(x) - f^{(\beta-1)}(y) \right| \leq K \left\| x - y \right\|$$

Roughly speaking, this means the functions have bounded β partial derivatives.

2 DESIRED RESULTS

Theorem 1. For bandwidth parameter $h > 0$ and decreasing function $k(\cdot, \cdot)$, write

$$K(Z_i, Z_j) = \frac{1}{h^m} k(\left\| Z_i - Z_j \right\|^2 / h^2).$$

For Sobolev IPM $\mathcal{S}_{\mu^2}(P, Q)$ as defined above,

$$\sqrt{T_2} \xrightarrow{p} \mathcal{S}_{\mu^2}(P, Q)$$

Proof attempt of Proposition 1. Recall that, for incidence matrix B,

$$\sqrt{T_2} = \left(\max_{\theta : \left\| B\theta \right\|_2 \leq 1} a^T \theta \right).$$

We expand $\left| \sqrt{T_2} - \mathcal{S}_{\mu^2}(P, Q) \right|$,

$$\begin{aligned} \left| \sqrt{T_2} - \mathcal{S}_{\mu^2}(P, Q) \right| \leq & \left| \max_{\theta : \left\| B\theta \right\|_2 \leq 1} \left\{ a^T \theta \right\} - \sup_{f \in B_W(\mathcal{X}, \mu^2)} \left\{ \mathbb{P}_n(f) - \mathbb{Q}_n(f) \right\} \right| \\ & + \left| \sup_{f \in B_W(\mathcal{X}, \mu^2)} \left\{ \mathbb{P}_n(f) - \mathbb{Q}_n(f) \right\} - \sup_{f \in B_W(\mathcal{X}, \mu^2)} \left\{ \mathbb{P}(f) - \mathbb{Q}(f) \right\} \right| \end{aligned} \tag{1}$$

(The following statement would hold only if Proposition 1 held over $B_W(\mathcal{X}, \mu^2)$, rather than over $B_W([0, 1], \lambda)$ for λ Lebesgue measure.)

By Proposition 1, the second term in the summand on the right hand side of (1) is $o_P(1)$.

(The following statement would hold only if Proposition 2 were uniform over $B_W(\mathcal{X}, \mu^2)$ rather than over the class of α-Holder functions $\mathcal{F}_\alpha$)

Then, Proposition 2 implies that for any $\epsilon > 0$, there exists N such that for $n \geq N$,

$$\sup_{f \in B_W(\mathcal{X}, \mu^2)} \left\{ \mathbb{P}_n(f) - \mathbb{Q}_n(f) \right\} - \max_{\theta : \left\| B\theta \right\|_2 \leq 1} \left\{ a^T \theta \right\} \leq \epsilon$$

with high probability.

To complete the proof, we will have to show that for any $\epsilon > 0$, there exists N such that for $n \geq N$,

$$\max_{\theta : \|B\theta\|_2 \leq 1} \left\{a^T \theta\right\} - \sup_{f \in B_W(\mathcal{X}, \mu^2)} \{\mathbb{P}_n(f) - \mathbb{Q}_n(f)\} \leq \epsilon$$

with high probability.

□

3 SUPPLEMENTAL RESULTS

Empirical process over Sobolev classes. The following theorem is a stand-in; it handles only functions with domain on the unit interval, and is stated specifically with respect to Lebesgue measure.

Proposition 1. Let $\mathcal{F}$ be the set of all absolutely continuous functions $f : [0, 1] \to \mathbb{R}$ such that $\|f\|_\infty \leq 1$ such that $\int (f'(x))^2 dx \leq 1$. Then, there exists a constant K such that for every $\epsilon > 0$,

$$\log N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|_\infty) \leq K \left(\frac{1}{\epsilon}\right).$$

Thus, the class $\mathcal{F}$ is P-Donsker (and P-Glivenko-Cantelli) for all P.

Regularization functional.

Proposition 2. Let $\mathcal{F}_\alpha$ be a unit ball in the space of α-Holder functions, and define $k(\cdot, \cdot)$ as in Theorem 1. For function $f \in \mathcal{F}_\alpha$, denote f evaluated on the data, $\mathbf{f} = (f(Z_1), \ldots, f(Z_N))$. Then, there exists a constant c depending only on k such that for $\alpha \geq 3$ and a sequence $(h_n) \to 0$ such that

$$\sup_{f \in \mathcal{F}_\alpha} \left| \|B\mathbf{f}_2\| - \|f\|_{W_0^{1,2}(\mathcal{X}, \mu^2)} \right| \xrightarrow{p} 0$$

4 PROOFS