Notes for Week 7/10/19 - 7/16/19

Alden Green

August 1, 2019

1 Testing over graphs.

Suppose we observe $G = (V, E)$, an undirected graph over $V = [n]$. Let D be the $m \times n$ incidence matrix of G, with singular value decomposition $D = U\Lambda^{1/2}V^T$, so that the Laplacian matrix $L = V\Lambda V^T$. For $i \in [n]$, we observe

$$z_i = \beta_i + \varepsilon_i, \quad \varepsilon_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1)$$

Our statistical goal is hypothesis testing. We wish to distinguish

$$\mathbf{H}_0 : \|\beta\|_2 = 0 \quad \text{vs.} \quad \mathbf{H}_a : \|\beta\|_2 > 0.$$

We will evaluate our performance using the notion of *worst-case error*. For a given "function" class $\mathcal{H}$, test function $\phi : \mathbb{R}^n \to \{0,1\}$, and $\epsilon > 0$, let

$$\mathcal{R}_\epsilon(\phi; \mathcal{H}) = \mathbb{E}_0(\phi) + \sup_{\beta \in \mathcal{H}: \|\beta\|_2 > \epsilon} \mathbb{E}_\beta(1 - \phi)$$

An example function class we will consider will be unit balls in discrete Sobolev norms, with known smoothness. For s, d positive integers, and radius $C_n > 0$, let

$$\mathcal{S}_d^s(C_n) = \Big\{\beta : \|D^{(s)}\beta\|_2 \leq C_n\Big\}$$

where

$$D^{(s)} = \begin{cases} L^{s/2}, & s \text{ even} \\ DL^{(s-1)/2}, & s \text{ odd.} \end{cases}$$

We note that the constraint is equivalent to $\beta^T V \Lambda^s V \beta \leq C_n^2$.

1.1 Test statistic

Let $0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{n-1}$ denote the ordered eigenvalues of L, and let v_k denote the eigenvector corresponding to λ_k. For a constant $C > 0$ to be specified later, let

$$T_C = \sum_{k: \lambda_k^s \leq C^2} y_k^2$$

where

$$y_k = \frac{1}{\sqrt{n}} \langle z, v_k \rangle.$$

We first compute the expectation $\mathbb{E}(T_C)$. Let $\theta \in \mathbb{R}^n$ represent the expectation of (y_k), meaning

$$\theta_k = \frac{1}{\sqrt{n}} \langle \beta, v_k \rangle$$

and let $\Pi_C\theta$ have entries $(\Pi_C\theta)_k = \mathbb{I}(\lambda_k^s \le C^2)\theta_k$. Finally, let $N(C) = \sharp\left\{k : \lambda_k^s \le C^2\right\}$.

Lemma 1. *For any $\beta \in \mathbb{R}^n$,*

$$\mathbb{E}(T_C) = \frac{N(C)}{n} + \|\Pi_C\theta\|_2^2 \tag{1}$$

If additionally $\beta \in \mathcal{H}$, the following lower bound holds:

$$\mathbb{E}(T_C) \ge \frac{N(C)}{n} + \frac{\|\beta\|^2}{n} - \frac{C_n^2}{nC^2} \tag{2}$$

Proof. We can write

$$\begin{aligned}
\mathbb{E}(T_C) &= \sum_{k:\lambda_k^s \le C} \mathbb{E}(y_k^2) \\
&= \frac{1}{n}\sum_{k:\lambda_k^s \le C} \mathbb{E}\big(\langle\beta, v_k\rangle^2 + \langle\varepsilon, v_k\rangle^2 + 2\langle\varepsilon, v_k\rangle\langle\beta, v_k\rangle\big) \\
&= \sum_{k:\lambda_k^s \le C} \theta_k^2 + \frac{1}{n} \\
&= \|\Pi_C\theta\|^2 + \frac{N(C)}{n},
\end{aligned}$$

showing (1). Now, assuming, $\|D^{(s)}\beta\|_2 \le C_n$, we can further obtain

$$\begin{aligned}
\|\Pi_C\theta\|^2 &= \|\theta\|^2 - \sum_{k:\lambda_k^s > C^2} \theta_k^2 \\
&\ge \|\theta\|^2 - \frac{1}{C^2}\sum_{k:\lambda_k^s > C^2} \theta_k^2\lambda_k^s \\
&\ge \|\theta\|^2 - \frac{1}{nC^2}\beta^T V\Lambda^s V^T\beta \\
&\ge \|\theta\|^2 - \frac{C_n^2}{nC^2}
\end{aligned}$$

and (2) is shown. □

We now turn to computing the variance $\mathrm{Var}(T_C)$.

Lemma 2.

$$\mathrm{Var}(T_C) = \frac{2N(C)}{n^2} + \frac{4\|\Pi_C\theta\|_2^2}{n}$$

Proof. To begin, we rewrite

$$\begin{aligned}
T_C &= \sum_{k:\lambda_k^s \le C^2} y_k^2 \\
&= \frac{1}{n}\sum_{k:\lambda_k^s \le C^2} \langle z, v_k\rangle^2 \\
&= \frac{1}{n}\sum_{k:\lambda_k^s \le C^2} z^T v_k v_k^T z \\
&=: \frac{1}{n} z^T P_C z.
\end{aligned}$$

where $P_C := \sum_{k:\lambda_k^s \le C^2} v_k v_k^T$. Therefore $\mathrm{Var}(T_C) = \mathrm{Var}(z^T P_C z)/n^2$. We expand $z = \beta + \varepsilon$ to obtain

$$
\begin{aligned}
\mathrm{Var}(z^T P_C z) &= \mathrm{Var}((\beta+\epsilon)^T P_C(\beta+\epsilon)) \\
&= \mathrm{Var}(\beta^T P_C \beta + 2\epsilon P_C \beta + \epsilon^T P_C \epsilon) \\
&= 4\beta^T P_C I P_C \beta + \mathrm{Var}(\epsilon^T P_C \epsilon) + 4\mathrm{Cov}(\epsilon P_C \beta, \epsilon^T P_C \epsilon) \\
&= 4n\|\Pi_C \theta\|_2^2 + \mathrm{Var}(\epsilon^T P_C \epsilon) \qquad (3)
\end{aligned}
$$

where the last equality follows from the Gaussianity of ϵ, as

$$
\mathbb{E}\big((\epsilon P_C \beta)(\epsilon^T P_C \epsilon)\big) = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n (P_C\beta)_k (P_C)_{ij} \mathbb{E}(\epsilon_i \epsilon_j \epsilon_k) = 0.
$$

Also by the Gaussianity of ϵ, $\epsilon^T P_C \epsilon \sim \chi^2_{N(C)}$, and therefore $\mathrm{Var}(\epsilon^T P_C \epsilon) = 2N(C)$. Plugging back into (3), we obtain

$$
\mathrm{Var}(z^T P_C z) = 4n\|\Pi_C\theta\|_2^2 + 2N(C)
$$

and therefore the desired result is proved. $\square$

We will consider now the test $\phi_C = \mathbf{1}\{T(C) \ge N(C)/n + \tau(b)\}$, where for $b \ge 1$, $\tau(b) = b\sqrt{2N(C)/n^2}$. We first upper bound the type I error.

Lemma 3. *Under the null hypothesis $\beta = 0$, and for any $C > 0$,*

$$
\mathbb{E}_{\beta=0}(\phi_C) \le \frac{1}{b^2}.
$$

Proof. The desired result follows from Chebyshev's inequality,

$$
\begin{aligned}
\mathbb{E}_{\beta=0}(\phi) &= \mathbb{P}_{\beta=0}\big(T(C) \ge N(C)/n + \tau(b)\big) \\
&= \mathbb{P}_{\beta=0}\Big(T(C) - \frac{N(C)}{n} \ge \tau(b)\Big) \\
&\le \mathbb{P}_{\beta=0}\Big(\Big|T(C) - \frac{N(C)}{n}\Big| \ge \tau(b)\Big) \\
&\le \frac{\mathrm{Var}_{\beta=0}(T_C)}{\tau(b)^2} = \frac{1}{b^2}.
\end{aligned}
$$

$\square$

The calculation for the type II error will be slightly more involved.

Lemma 4. *Let $b \ge 1$ be fixed. For every $\beta \in \mathcal{S}_d^s$ such that*

$$
\frac{\|\beta\|^2}{n} \ge 2b\sqrt{2\frac{N(C)}{n^2}} + \frac{C_n^2}{nC^2} \qquad (4)
$$

we have that

$$
\mathbb{E}_\beta(1-\phi) \le \frac{2}{b^2} + \frac{2}{b\sqrt{2N(C)}}.
$$

Proof. Let $\Delta = \mathbb{E}_\beta(T_C) - N(C)/n = \|\Pi_C\theta\|^2$, and observe that by Lemma 1 and (4),

$$
\Delta \ge \frac{\|\beta\|_2^2}{n} - \frac{C_n^2}{nC^2} \ge 2\tau(b).
$$

An application of Chebyshev's inequality yields

$$
\begin{aligned}
\mathbb{E}_\beta\big(1-\phi\big) &= \mathbb{P}_\beta\big(T_C \le N(C)/n + \tau(b)\big) \\
&= \mathbb{P}_\beta\big(T_C - \mathbb{E}_\beta(T_C) \le \tau(b) - \Delta\big) \\
&\le \mathbb{P}_\beta\big(|T_C - \mathbb{E}_\beta(T_C)| \le \Delta - \tau(b)\big) && \text{(since } \Delta \ge \tau(b)\text{)} \\
&\le \frac{\mathrm{Var}_\beta(T_C)}{(\Delta - \tau(b))^2} \\
&\le 4\frac{\mathrm{Var}_\beta(T_C)}{\Delta^2} && \text{(since } \Delta \ge 2\tau(b)\text{)} \\
&\le 4\frac{2N(C)/n^2 + \|\Pi_C\theta\|_2^2/n}{\Delta^2}.
\end{aligned}
$$

We now handle each summand separately. For the first term, since $\Delta \ge 2\tau(b)$, we have

$$
\frac{2N(C)}{n^2\Delta^2} \le \frac{1}{2b^2}.
$$

For the second term, since $\Delta = \|\Pi_C\theta\|^2$, we have

$$
\begin{aligned}
\frac{\|\Pi_C\theta\|_2^2/n}{\Delta^2} &\le \frac{1}{n\Delta^2} \\
&\le \frac{1}{2n\tau(b)} \\
&= \frac{1}{2b\sqrt{2N(C)}}.
\end{aligned}
$$

□

To more explicitly specify the critical radius $\epsilon : \|\beta\|_2 \ge \epsilon$, we will need to make an assumption on the relation between $N(C)$ and C. In particular, let $C = C^*$, where

$$
C^* = \frac{(C_n n^{s/d})^{4s/(4s+d)}}{n^{s/d}}.
$$

We will assume the following bounds on $N(C^*)$:

(A1) Tail decay:

$$
N(C^*) \le (C^*)^{d/s} n
$$

(A2) Asymptotic consistency:

$$
\lim_{n\to\infty} N(C^*) = \infty
$$

Corollary 1. *Under (A2) and (A1), letting*

$$
\epsilon^2 = (2\sqrt{2}b + 1)\left(\frac{(C_n n^{s/d})^{2d/(4s+d)}}{n}\right)
$$

we have that

$$
\mathcal{R}_\epsilon(\phi_{C^*}; \mathcal{S}_d^s) \le \frac{2}{b^2} + o(1)
$$

Proof. Recall that

$$\mathcal{R}_\epsilon(\phi_{C^*}; \mathcal{S}^s_d) = \mathbb{E}_{\beta=0}(\phi_{C^*}) + \sup_{\beta \in \mathcal{H}: \|\beta\|_2/n > \epsilon} \mathbb{E}_\beta(1 - \phi_{C^*})$$

By Lemma 3, we have that

$$\mathbb{E}_{\beta=0}(\phi_{C^*}) \leq \frac{1}{b^2}.$$

Now, we verify that $\epsilon^2 \geq 2b\sqrt{2N(C^*)/n^2} + C_n^2/n(C^*)^2$. By Assumption (A1) and the choice of $C^\star$, we have

$$N(C^*) \leq (C^*)^{d/s} n = (C_n n^{s/d})^{4d/(4s+d)}$$

and therefore

$$2b\sqrt{\frac{N(C^*)}{n^2}} \leq \frac{2b}{n}(C_n n^{s/d})^{2d/(4s+d)}$$

Moving on to the second term, we have

$$\begin{aligned}
\frac{C_n^2}{n(C^\star)^2} &= \frac{C_n^2 n^{2s/d}}{n(C_n n^{s/d})^{8s/(4s+d)}} \\
&= \frac{C_n^{2d/(4s+d)} n^{(s/d)2d/(4s+d)}}{n} \\
&= \frac{(C_n n^{s/d})^{2d/(4s+d)}}{n}
\end{aligned}$$

and therefore $\epsilon \geq 2b\sqrt{2N(C^*)/n^2} + C_n^2/n(C^*)^2$. As a result, by Lemma 4 for any $\beta \in \mathcal{S}^s_d$ such that $\|\beta\|/n \geq \epsilon$,

$$\mathbb{E}_\beta(1 - \phi_{C^\star}) \geq \frac{2}{b^2} + \frac{2}{b\sqrt{2N(C)}}$$

and by Assumption (A2), the latter term tends to infinity with n. □

2 Applications

2.1 Grid

Let G be the d-dimensional grid graph, and let $C_n = n^{1/2 - s/d}$, so that

$$C^\star = \frac{n^{2s/(4s+d)}}{n^{s/d}}$$

The following bound holds for eigenvalues of the Laplacian L of the grid graph:

$$\lambda_k^s \geq 4\sin^{2s}(\pi k^{1/d}/(2n^{1/d})) \geq \frac{\pi^{2s} k^{2s/d}}{4^s n^{2s/d}}$$

Therefore, for any $C > 0$, if $\lambda_k^s < C^2$, then

$$\begin{aligned}
\frac{\pi^{2s} k^{2s/d}}{4^s n^{2s/d}} < C^2 \implies \\
k < \frac{4^s C^{d/s} n}{\pi^{2s}}
\end{aligned}$$

Since this holds in particular with respect to $C = C^\star$, assumption (A1) is satisfied. One can similarly shown (A2) holds as well. By Corollary 1, we therefore have that when

$$\epsilon^2 = (2b+1)\left(\frac{n^{d/(4s+d)}}{n}\right) = (2b+1)n^{-4s/(4s+d)}$$

we have

$$\mathcal{R}_\epsilon(\phi_{C^*}; \mathcal{S}^s_d) \leq \frac{2}{b^2} + o(1).$$

3 Additional Theory

We consider now a naive test statistic,

$$T = \frac{\|z\|^2}{n}$$

with associated test $\Phi_b(T) = \mathbf{1}(T \leq 1 + 2b/\sqrt{n})$. and prove an upper bound showing that when $\epsilon \geq n^{-1/4}$, the worst-case error is bounded for all $\beta \in L^2(V)$.

Lemma 5. *Let $\epsilon = n^{-1/4}$. Then*

$$\mathcal{R}_\epsilon(\Phi_b; L^2(V)) \leq \frac{2}{b^2} + \frac{1}{n}$$