Notes for Week 5/15/19 - 5/17/19

Alden Green

May 30, 2019

Consider absolutely continuous distributions $\mathbb{P}$ and $\mathbb{Q}$ with density functions f and g, respectively. For fixed $n \geq 0$, let $Z = (z_1, \ldots, z_n)$, where for $i = 1, \ldots, n$, $z_i \sim \frac{\mathbb{P}+\mathbb{Q}}{2}$ are independent. Given Z, for $i = 1, ..., n$ let

$$\ell_i = \begin{cases} 1 \text{ with probability } \frac{f(z_i)}{f(z_i)+g(z_i)} \\ -1 \text{ with probability } \frac{g(z_i)}{f(z_i)+g(z_i)} \end{cases}$$

be conditionally independent labels, and write

$$1_X = \begin{cases} 1, & l_i = 1 \\ 0, & \text{otherwise} \end{cases} \qquad 1_Y = \begin{cases} 1, & l_i = -1 \\ 0, & \text{otherwise.} \end{cases}$$

We will write $X = \{x_1, \ldots, x_{N_X}\} := \{z_i : \ell_i = 1\}$ and similarly $Y = \{y_1, \ldots, y_{N_Y}\} := \{z_i : \ell_i = -1\}$, where N_X and N_Y are of course random but $N_X + N_Y = n$.

Our statistical goal is hypothesis testing: that is, we wish to construct a test function ϕ which differentiates between

$$\mathbb{H}_0 : f = g \text{ and } \mathbb{H}_1 : f \neq g.$$

For a given function class $\mathcal{H}$, some $\epsilon > 0$, and test function ϕ a Borel measurable function of the data with range $\{0, 1\}$, we evaluate the quality of the test using *worst-case risk*

$$R_\epsilon^{(t)}(\phi; \mathcal{H}) = \sup_{f \in \mathcal{H}} \mathbb{E}_{f,f}^{(t)}(\phi) + \sup_{\substack{f,g \in \mathcal{H} \\ \delta(f,g) \geq \epsilon}} \mathbb{E}_{f,g}^{(t)}(1 - \phi)$$

where

$$\delta^2(f, g) = \int_{\mathcal{D}} (f - g)^2 dx.$$

Test statistic. For $r \geq 0$, define the *r-graph* $G_r = (V, E_r)$ to have vertex set $V = \{1, \ldots, t\}$ and edge set E_r which contains the pair (i, j) if and only if $\|z_i - z_j\|_2 \leq r$. Let D_r denote the incidence matrix of G_r.

Define the *Laplacian Smooth* test statistic over the neighborhood graph to be

$$T_{LS} = \sup_{\theta : \|D_r \theta\|_2 \leq C(n,r)} \langle \theta, \frac{1_X}{N_X} - \frac{1_Y}{N_Y} \rangle$$

where we note that the test statistic is implicitly a function of r and $C(n, r)$.

General assumptions. We will need to make the following assumptions in order to obtain meaningful theoretical results.

Assumption 1 (General assumptions.)**.** *For each $f \in \mathcal{H}$, assume* $\operatorname{supp}(f)$ *is an open connected, bounded subset of $\mathbb{R}^d$ with l-Lipschitz boundary. Further assume $\mathcal{H}$ is uniformly bounded above and below: that is, there exists $f_{\min} < f_{\max}$ such that for each $f \in \mathcal{H}$,*

$$0 < f_{\min} < f(x) < f_{\max} < \infty, \quad \textit{for all } x \in \operatorname{supp}(f)$$

We motivate these assumptions through the following theorem relating the r-neighborhood graph to the d-dimensional grid, a useful result as the spectrum of grid graphs is well known.

Formally, for $\kappa = n^{1/d}$, consider the *grid graph*

$$G_{grid} = (V_{grid}, E_{grid}), \quad V_{grid} = \left\{k : k \in [\kappa]^d\right\}, \quad E_{grid} = \{(k, k') : k, k' \in V_{grid}, \|k - k'\|_1 = 1\}$$

with associated incidence matrix D_{grid}.

Lemma 1. *Let $z_1, \ldots, z_n \sim p$ for some density function $p \in \mathcal{H}$ which satisfies Assumption 1.*

Fix $a > 2$. There exists $c_1 > 0$ (potentially depending on $d, l, f_{\min}, f_{\max}$ and a) such that for any $r \geq c_1 \left(\frac{\log n}{n}\right)^{1/d}$, the following statement holds with probability at least $1 - n^{-a}$: there exists a bijection $T : [n] \to [\kappa]^d$ such that for all $\theta = (\theta_1, \ldots, \theta_n) \in \mathbb{R}^n$, letting

$$(\theta_T)_k = \theta_{T^{-1}(k)}, \textit{for all } k \in [\kappa]^d$$

we have

$$\|D_{\text{grid}}\theta_T\|_2 \leq \|D_r\theta\|_2 .$$

Prove this.

We use this to control the empirical process T_{LS} under the null hypothesis.

Lemma 2. *Assume $\mathcal{H}$ satisfies Assumption 1, and fix $a \geq 2$. Then, there exists $c_1 > 0$ (potentially depending on $d, l, f_{\min}$, $f_{\max}$, and a) and c_2 (potentially depending on d) such that for any $r \geq c_1 \left(\frac{\log n}{n}\right)^{1/d}$, and any $C_{n,r} > 0$,*

$$\mathbb{P}_{\mathbb{H}_0}\left(T_{LS} \leq \frac{c_2 C(n,r)\sqrt{\log n}}{a\sqrt{n}}\right) \leq 1 - 2\exp(2a) - a$$

Proof. We write

$$\begin{aligned}
T_{LS} &\leq \left| T_{LS} - \sup_{\substack{\theta : \|D_r\theta\|_2 \leq C(n,r), \\ \theta^T \mathbf{1} = 0}} \frac{1}{n/2}\langle \theta, 1_X - 1_Y\rangle \right| + \sup_{\substack{\theta : \|D_r\theta\|_2 \leq C(n,r), \\ \theta^T \mathbf{1} = 0}} \frac{1}{n/2}\langle \theta, 1_X - 1_Y\rangle \\
&\leq \left| T_{LS} - \sup_{\substack{\theta : \|D_r\theta\|_2 \leq C(n,r), \\ \theta^T \mathbf{1} = 0}} \frac{1}{n/2}\langle \theta, 1_X - 1_Y\rangle \right| + \sup_{\substack{\theta : \|D_{\text{grid}}\theta\|_2 \leq C(n,r), \\ \theta^T \mathbf{1} = 0}} \frac{1}{n/2}\langle \theta, 1_X - 1_Y\rangle \qquad \text{(Lemma 3)} \\
&\leq \left| T_{LS} - \sup_{\substack{\theta : \|D_r\theta\|_2 \leq C(n,r), \\ \theta^T \mathbf{1} = 0}} \frac{1}{n/2}\langle \theta, 1_X - 1_Y\rangle \right| + C(n,r)\frac{1}{n/2}\sqrt{\sigma^T L_{\text{grid}}^{\dagger}\sigma}
\end{aligned}$$

whence the statement follows by Lemmas 4 and 5. ∎

1 Additional Theory

The Laplacian smooth test statistic has a closed form solution. Let $L = D^T D$ be the Laplacian of G, and write $L^\dagger$ for the pseudoinverse of L.

Lemma 3. *For any unweighted, undirected, connected graph $G = (V, E)$ with incidence matrix D, number $C > 0$, and any vector $v \in \mathbb{R}^n$ with $\sum_{i=1}^n v_i = 0$,*

$$\sup_{\theta : \|D\theta\|_2 \le C} \langle \theta, v \rangle = C\sqrt{v^T L^\dagger v}$$

Additionally, for any vector $v \in \mathbb{R}^n$ (not necessarily $\sum_{i=1}^n v_i = 0$), under the additional constraint $\theta^T \mathbf{1} = 0$, the same statement holds. That is,

$$\sup_{\substack{\theta : \|D\theta\|_2 \le C, \\ \theta^T \mathbf{1} = 0}} \langle \theta, v \rangle = C\sqrt{v^T L^\dagger v}$$

Proof. Note that the condition $\|D\theta\|_2 \le C$ is equivalent to $\theta^T L \theta \le C$. The solution then follows from the KKT conditions. $\square$

Lemma 4. *Under the general assumptions,*

$$\left| T_{LS} - \sup_{\substack{\theta : \|D_r\theta\|_2 \le C(n,r), \\ \theta^T \mathbf{1} = 0}} \frac{1}{n/2} \langle \theta, 1_X - 1_Y \rangle \right| \le \frac{8C(n,r)}{an^{1-1/d}\pi}$$

with probability at least $1 - 2\exp(2a)$.

Proof. Note that as $(1_X/N_X - 1_Y/X_Y)^T \mathbf{1} = 0$, we may write

$$T_{LS} = \sup_{\substack{\theta : \|D_r\theta\|_2 \le 1, \\ \theta^T \mathbf{1} = 0}} \langle \theta, \frac{1_X}{N_X} - \frac{1_Y}{N_Y} \rangle$$

and therefore by Cauchy-Schwarz,

$$\begin{aligned}
\left| T_{LS} - \sup_{\substack{\theta : \|D_r\theta\|_2 \le C, \\ \theta^T \mathbf{1} = 0}} \frac{1}{n/2} \langle \theta, 1_X - 1_Y \rangle \right| &\le \sup_{\substack{\theta : \|D_r\theta\|_2 \le C, \\ \theta^T \mathbf{1} = 0}} \{\|\theta\|\} \cdot \|1_X(2/n - 1/N_X) - 1_Y(1/N_Y - 2/n)\| \\
&\le \sup_{\substack{\theta : \|D_r\theta\|_2 \le C, \\ \theta^T \mathbf{1} = 0}} \{\|\theta\|\} \cdot \|1_X(2/n - 1/N_X) - 1_Y(1/N_Y - 2/n)\| \\
&\le \sup_{\substack{\theta : \|D_r\theta\|_2 \le C, \\ \theta^T \mathbf{1} = 0}} \{\|\theta\|\} \cdot \left(\|1_X\| \frac{|n/2 - N_X|}{n/2 N_X} + \|1_Y\| \frac{|n/2 - N_Y|}{n/2 N_Y} \right) \\
&\le \sup_{\substack{\theta : \|D_r\theta\|_2 \le C, \\ \theta^T \mathbf{1} = 0}} \{\|\theta\|\} \cdot \left(\frac{2\,|n/2 - N_X|}{\sqrt{N_X} n} + \frac{2\,|n/2 - N_Y|}{\sqrt{N_Y} n} \right).
\end{aligned}$$

The reasoning in the proof of Lemma 5 leads to the bound

$$\sup_{\substack{\theta : \|D_r\theta\|_2 \le C, \\ \theta^T \mathbf{1} = 0}} \{\|\theta\|\} \le \frac{Cn^{1/d}}{\pi},$$

and a standard application of Hoeffding's inequality to the quantities N_X and N_Y yields, for any $a > 0$

$$|N_X - n/2|\,, |N_Y - n/2| \le a\sqrt{n}$$

with probability at least $1 - 2\exp(-2a)$. As a result,

$$\sup_{\substack{\theta:\|D_r\theta\|_2 \le 1,\\ \theta^T\mathbf{1}=0}} \{\|\theta\|\} \cdot \left(\frac{2\,|n/2 - N_X|}{\sqrt{N_X} n} + \frac{2\,|n/2 - N_Y|}{\sqrt{N_Y} n} \right) \le \frac{8aCn^{1/d}}{n\pi}$$

with probability at least $1 - 2\exp(-2a)$, which proves the claim. □

1.1 Type I error.

Denote the eigenvalues and eigenvectors of $L_{\text{grid}} = D_{\text{grid}}^T D_{\text{grid}}$ as $\{\lambda_k : k \in [\kappa]^d\}$ and $\{u_k : k \in [\kappa]^d\}$, respectively.

Lemma 5. *Let $\sigma_1, \ldots, \sigma_n \sim \text{Rademacher}(1/2)$ be independent random variables, and write $\sigma = (\sigma_1, \ldots, \sigma_n)$. Then,*

$$\sigma^T L_{\text{grid}}^{\dagger} \sigma \le \frac{(2d)^d \kappa^2}{a\pi^2} \left(8 + 2\pi \log(\sqrt{2}\kappa) \left(\sum_{p=2}^{d} \kappa^p \right) \right)$$

with probability at least $1 - a$.

Proof. Taking the eigendecomposition $L_{\text{grid}} = U\Lambda U^T$, we can then write $L_{\text{grid}}^{\dagger} = U\Lambda^{\dagger}U^T$. Therefore

$$\mathbb{E}\big(\sigma^T L_{\text{grid}}^{\dagger} \sigma\big) = \sum_{k\in[\kappa]^d} \frac{\mathbb{E}\big(u_k^T \sigma\big)^2}{\lambda_k}$$

and as

$$\mathbb{E}\big(u_k^T\sigma\big)^2 = \sum_{i,j=1}^{n} \mathbb{E}(u_{k_i} u_{k_j} \sigma_i \sigma_j) = \sum_{i=1}^{n} u_{k_i}^2 = 1$$

we are left with

$$\mathbb{E}\big(\sigma^T L_{\text{grid}}^{\dagger} \sigma\big) = \sum_{\substack{k\in[\kappa]^d,\\ k\neq \mathbf{0}}} \frac{1}{\lambda_k}. \tag{1}$$

It is well known HR, CB, that the d-dimensional grid can be written as the Kronecker product of d path graphs, and exploiting this fact as in SWT we obtain

$$\lambda_k = 4\sin^2\left(\frac{\pi(k_1 - 1)}{2\kappa}\right) + \cdots + 4\sin^2\left(\frac{\pi(k_d - 1)}{2\kappa}\right).$$

The inequality $\sin(x) \ge x/2$ holds for $x \in [0, \pi/2]$, and so we have

$$\begin{aligned}
\sum_{\substack{k\in[\kappa]^d,\\ k\neq \mathbf{0}}} \frac{1}{\lambda_k} &\le \frac{\kappa^2}{\pi^2} \sum_{\substack{k\in[\kappa]^d,\\ k\neq \mathbf{0}}} \left(\sum_{j=1}^{d} (k_j - 1)^2 \right)^{-1} \\
&\le \frac{(2d)^d \kappa^2}{\pi^2} \left(8 + 2\pi \log(\sqrt{2}\kappa) \left(\sum_{p=2}^{d} \kappa^{p-2} \right) \right) \\
&= \frac{(2d)^d \kappa^2}{\pi^2} \left(8 + 2\pi \log(\sqrt{2}\kappa) \left(\sum_{p=2}^{d} \kappa^{p} \right) \right) && \text{(Lemma 6)}
\end{aligned}$$

whence the claim follows by Markov's inequality. □

Lemma 6. *There exists a universal constant $c_1 > 0$ such that for any $\kappa \geq 1$ and for all integers $d \geq 3$,*

$$\sum_{\substack{k\in[\kappa]^d,\\ k\neq \mathbf{0}}} \left(\sum_{j=1}^{d}(k_j-1)^2\right)^{-1} \leq (2d)^d\left(8+2\pi\log(\sqrt{2}\kappa)\left(\sum_{p=2}^{d}\kappa^{p-2}\right)\right)$$

Proof. We will prove by induction on d. As a base case, consider $d = 2$. Rewrite

$$\sum_{\substack{k\in[\kappa]^d,\\ k\neq \mathbf{0}}} \left(\sum_{j=1}^{d}(k_j-1)^2\right)^{-1} \leq 4\sum_{k_1=1}^{\kappa-1}\frac{1}{k_1^2}+\sum_{k_1=2}^{\kappa-1}\sum_{k_2=2}^{\kappa-1}(k_1^2+k_2^2)^{-1}.$$

Then, upper bounding sums by integrals, we obtain

$$\sum_{k_1=1}^{\kappa-1}\frac{1}{k_1^2}\leq 1+\int_{x=1}^{\kappa}\frac{1}{x^2}\,\mathrm{d}x\leq 2,$$

and

$$\begin{aligned}\sum_{k_1=2}^{\kappa-1}\sum_{k_2=2}^{\kappa-1}(k_1^2+k_2^2)^{-1} &\leq \int_{x=1}^{\kappa}\int_{y=1}^{\kappa}\frac{1}{x^2+y^2}\,\mathrm{d}x\,\mathrm{d}y\\ &\leq 2\pi\log(\sqrt{2}\kappa),\end{aligned}$$

so that

$$\sum_{\substack{k\in[\kappa]^d,\\ k\neq \mathbf{0}}} \left(\sum_{j=1}^{d}(k_j-1)^2\right)^{-1} \leq 8+2\pi\log(\sqrt{2}\kappa).$$

and the base case is shown.

We proceed by induction. For $d \geq 3$ we write

$$\begin{aligned}\sum_{\substack{k\in[\kappa]^d,\\ k\neq \mathbf{0}}} \left(\sum_{j=1}^{d}(k_j-1)^2\right)^{-1} &\leq 2d\sum_{\substack{k\in[\kappa]^{d-1},\\ k\neq \mathbf{0}}} \left(\sum_{j=1}^{d-1}(k_j-1)^2\right)^{-1}+\sum_{k_1=2}^{\kappa-1}\cdots\sum_{k_d=2}^{\kappa-1}\left(\sum_{j=1}^{d}k_j^2\right)^{-1}\\ &\leq 2d\left((2d)^{d-1}\left(8+2\pi\log\sqrt{2}\kappa\left(\sum_{p=2}^{d-1}\kappa^{p-2}\right)\right)\right)+\sum_{k_1=2}^{\kappa-1}\cdots\sum_{k_d=2}^{\kappa-1}\left(\sum_{j=1}^{d}k_j^2\right)^{-1}\\ &=\left((2d)^{d}\left(8+2\pi\log\sqrt{2}\kappa\left(\sum_{p=2}^{d-1}\kappa^{p-2}\right)\right)\right)+\sum_{k_1=2}^{\kappa-1}\cdots\sum_{k_d=2}^{\kappa-1}\left(\sum_{j=1}^{d}k_j^2\right)^{-1}\end{aligned}$$

and again upper bounding sums by integrals, we obtain

$$\begin{aligned}\sum_{k_1=2}^{\kappa-1}\cdots\sum_{k_d=2}^{\kappa-1}\left(\sum_{j=1}^{d}(k_j)^2\right)^{-1} &\leq \int_1^{\kappa}\cdots\int_1^{\kappa}\frac{1}{x_1^2+\cdots+x_d^2}\,\mathrm{d}x_d\ldots\mathrm{d}x_1\\ &\leq \kappa^{d-2}\int_1^{\kappa}\int_1^{\kappa}\frac{1}{x^2+y^2}\,\mathrm{d}x\,\mathrm{d}y\\ &\leq \kappa^{d-2}2\pi\log(\sqrt{2}\kappa)\end{aligned}$$

and the proof is complete. □

Type II error. Lemma 7 will be useful in lower bounding the approximation error of the T_{LS} test statistic. Write $P_1^\perp$ for the projection operator onto the subspace of $\mathbb{R}^n$ orthogonal to the constant vector $\mathbf{1}$.

Lemma 7. *For any vector $v \in \mathbb{R}^n$ and Laplacian matrix L of a connected graph G,*

$$v^T L^\dagger v \geq \frac{(v^T P_1^\perp v)^2}{v^T L v}$$

Proof. We can expand $v^T P_1^\perp v = v^T (L)^{1/2} (L^\dagger)^{1/2} v$, for any matrix square root of L and $L^\dagger$. The statement follows by Cauchy-Schwarz. □