# Notes for Week 5/29/19 - 5/31/19

Alden Green

June 4, 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

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

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 $\phi$ 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 $\phi$ 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

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

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

# 1 Additional Theory

Let $\theta^\star = (\theta_i^\star)_{i=1}^n$, with $\theta_i^\star := f(z_i) - g(z_i)$. Write $L = D^T D$ for the Laplacian matrix of the $r$-neighborhood graph, $\mathcal{D} f$ for the gradient of a function $f$, and $\mathcal{D}^2 f$ for the Hessian of a function $f$. We write $C^2(L)$ for the set of functions $f$ twice continuously differentiable over $\mathbb{R}^d$, with bounded Hessian $\left\|\mathcal{D}^2 f\right\|_\infty \leq L$.

**Lemma 1.** *For all density functions $f, g \in C^2(L)$ with $\int (f-g)^2 \geq \epsilon^2$, if $r \geq c_1(\log n/n)^{1/d}$ and $\epsilon \geq c_2 n r^{d+1}$ then*

$$\sup_{\|D\theta\|_2 \leq 1} \langle \theta, \theta^\star \rangle \geq \frac{c_2}{2} (\log n)^{1+2/d} n^{-2/d}$$

*with probability at least $1-\delta$.*

*Proof.* By Lemmas 4 and 5, we have

$$\begin{aligned} \sup_{\|D\theta\|_2 \leq 1} \langle \theta, \theta^\star \rangle &= \frac{1}{n} \sqrt{(\theta^\star)^T L^\dagger \theta^\star} \\ &\geq \frac{1}{\lambda_k} \left( \langle \theta^\star, \theta^\star \rangle - \langle P_k^\perp \theta^\star, P_k^\perp \theta^\star \rangle \right)^2 \end{aligned} \tag{1}$$

where the latter inequality holds for any $k = 1, \ldots, n-1$. We upper bound $\langle P_k^\perp \theta^\star, P_k^\perp \theta^\star \rangle$ using the following relations

$$(\theta^\star)^T L \theta^\star \geq (P_k^\perp \theta^\star)^T L^\dagger (P_k^\perp \theta^\star) \geq \lambda_k \langle P_k^\perp \theta^\star, P_k^\perp \theta^\star \rangle$$

and obtain

$$\begin{aligned} \frac{1}{\lambda_k} \left( \langle \theta^\star, \theta^\star \rangle - \langle P_k^\perp \theta^\star, P_k^\perp \theta^\star \rangle \right)^2 &\geq \frac{1}{\lambda_k} \left( \langle \theta^\star, \theta^\star \rangle - \frac{(\theta^\star)^T L \theta^\star}{\lambda_k} \right)^2 \\ &\geq \frac{1}{\lambda_k} \left( c_1 n \epsilon^2 - \frac{c_2 n^2 r^{d+2}}{\lambda_k} \right)^2 \qquad \text{(Lemmas 2 and 3)} \end{aligned}$$

where the latter inequality occurs with probability at least $1 - 2\delta$. Choose $k = n-1$. As $r \geq c_1(\log n/n)^{1/d}$, we have that $G \preceq \text{Grid}$, and as a result $4 \geq \lambda_{n-1} \geq 1$. Therefore,

$$\begin{aligned} \frac{1}{\lambda_k} \left( c_1 n \epsilon^2 - \frac{c_2 n^2 r^{d+2}}{\lambda_k} \right)^2 &\geq \frac{1}{4} \left( c_1 n \epsilon^2 - c_2 n^2 r^{d+2} \right)^2 \\ &\geq \frac{1}{4} \left( c_2 n^2 r^{d+2} \right)^2 \\ &\geq \frac{1}{4} \left( c_2 n^{1-2/d} \log(n)^{1+2/d} \right)^2 = \frac{c_2^2}{4} n^{2-4/d} \log(n)^{2+4/d}, \end{aligned}$$

and the proof is complete. ∎

**Lemma 2.** *For any $f, g \in L^2(\mathbb{R}^d)$ satisfying the regularity conditions, such that $\int_{\mathbb{R}^d} (f-g)^2 dx \geq \epsilon^2$ there exists constant $c_1$ such that*

$$\langle \theta^\star, \theta^\star \rangle \geq c_1 n \epsilon^2$$

*with probability at least $1-\delta$.*

**Lemma 3.** *For any $f, g \in C^2(L)$ satisfying the regularity conditions, there exists a constant $c_2$ such that*

$$(\theta^\star)^T L \theta^\star \leq c_2 n^2 r^{d+2}$$

*with probability at least $1-\delta$.*

## 2 Linear Algebra

**Lemma 4.** *For any unweighted, undirected, connected graph $G = (V, E)$ with incidence matrix $D$, 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. □

Let $L$ be the Laplacian matrix of a connected graph, and $L^\dagger$ be the pseudo-inverse $L$. Write the eigendecomposition of $L^\dagger = U\Lambda^\dagger U^T$, where $\Lambda$ is an $n \times n$ diagonal matrix with entries $0 = \lambda_0 < \lambda_1 \le \lambda_2 \le \ldots \le \lambda_{n-1}$ and $U$ is an orthogonal matrix with columns $U = (u_0 \ldots u_{n-1})$. For each $k = 0, \ldots, n-1$, write $U_k = (u_0 \ldots u_{k-1})$ for the first $k$ columns of $U$, $P_k$ as the projection operator onto the span of $U_k$, and $P_k^\perp$ for the projection operator onto the subspace orthogonal to the span of $U_k$.

**Lemma 5.** *For any $k = 1, \ldots, n$,*

$$(\theta^\star)^T L^\dagger \theta^\star \ge \frac{1}{\lambda_k}\left(\langle \theta^\star, \theta^\star \rangle - \langle P_k^\perp \theta^\star, P_k^\perp \theta^\star \rangle\right)^2$$

*Proof.* Note that $\Lambda^\dagger$ is a diagonal matrix, with entries $\rho_{1,1} = 0$ and $\rho_{k,k} = \frac{1}{\lambda_k}$ for $k = 1, \ldots, n$. Therefore,

$$(\theta^\star)^T L^\dagger \theta^\star = \sum_{k=1}^n \frac{\langle \theta^\star, u_k \rangle^2}{\lambda_k}.$$

Clearly, for any $k = 1, \ldots, n$,

$$\sum_{k=1}^n \frac{\langle \theta^\star, u_k \rangle^2}{\lambda_k} \ge \frac{\langle P_k \theta^\star, P_k \theta^\star \rangle^2}{\lambda_k}$$

and as $\langle P_k \theta^\star, P_k \theta^\star \rangle + \langle P_k^\perp \theta^\star, P_k^\perp \theta^\star \rangle = \langle \theta^\star, \theta^\star \rangle$, we obtain

$$\frac{\langle P_k \theta^\star, P_k \theta^\star \rangle^2}{\lambda_k} = \frac{1}{\lambda_k}\left(\langle \theta^\star, \theta^\star \rangle - \langle P_k^\perp \theta^\star, P_k^\perp \theta^\star \rangle\right)^2.$$

□