Notes for Week 6/9/19 - 6/11/19

Alden Green

June 12, 2019

Let f, g be L^2 density functions supported on the d-dimensional unit cube $[0,1]^d$, meaning

$$\int_{[0,1]^d} f(x)^2 dx, \int_{[0,1]^d} g(x)^2 dx < \infty, \quad \text{and} \quad \int_{[0,1]^d} f(x) dx, \int_{[0,1]^d} g(x) dx = 1.$$

with f,g bounded away from 0, $f(x), g(x) > p_{\min}$ for all $x \in [0,1]^d$.

We observe data (X, ℓ), a design matrix and associated labels, specified as follows. We let $x_1, \ldots, x_n$ be the rows of X, each sampled independently from $\mu = (f+g)/2$. For each $i = 1, \ldots, n$, we then sample ℓ_i according to

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

and let $\ell = (\ell_i)$.

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

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

We will consider those density functions which belong to $\mathcal{H} := \mathcal{H}_1^d(L)$, the d-dimensional Lipschitz ball with norm L. Formally, a function $f \in \mathcal{H}$ if for all $x, y \in [0,1]^d$, $|f(x) - f(y)| \leq L \|x - y\|$.[1]

1 Sieve-based test.

For a radius $r > 0$ to be determined later, the r-neighborhood graph $G = (V, E)$ consists of vertices $V = [n]$ corresponding to the n data points, and edges $E = \{(u, v) : \|x_u - x_v\| \leq r\}$. Write D for the incidence matrix of this graph, and $L = D^T D$ for the graph Laplacian. Denote the singular value decomposition of D by $D = U \Lambda^{1/2} V^T$, where U and V are orthonormal matrices and Λ is a diagonal matrix with entries

[1] Note that the Lipschitz requirement, along with the restriction that f be a density function over a bounded domain, imply that f has finite L^2 norm.

$0 = \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \ldots \leq \lambda_n$, so that $L = V \Lambda V^T$. For $k = 1, \ldots, n$, write $V_k = (v_1 \ldots v_k)$ for the $n \times k$ matrix containing the first k columns of V–that is, the first k eigenvectors of L. Our first test statistic will be as follows:

$$T_{LE}^{(k)} := \sup_{\substack{\theta \in \mathrm{col}(V_k), \\ \|\theta\|=1}} \langle \theta, \ell \rangle^2 = \ell^T V_k V_k^T \ell. \tag{1}$$

1.1 Assumptions

The following are statements I wish to hold in order for the later theory to follow, but I cannot yet prove:

1. There exists a constant $c_1 > 0$ – which may depend on dimension d but not on n – such that for all $h \in 1, \ldots, k$ and all $i = 1, \ldots, n$,
$$|v_{hi}| \leq \frac{c}{\sqrt{n}} \tag{2}$$

If v_k were eigenvectors of the grid–rather than of the neighborhood graph–the first item would hold with $c = 2^{d/2}$.

1.2 Main result.

Let $d = 3, 4$, choose $r = c(\log(n)/n)^{1/d}$ to satisfy the condition of Lemma 6, and let $k = n^{2d/(d+4)}$. Let

$$\phi_{LE} = \begin{cases} 1, & \text{if } T_{LE}^{(k)} \geq k + a\sqrt{2k + c^4\left(k + \frac{k^2}{n}\right)} \\ 0, & \text{otherwise.} \end{cases}$$

The following theorem holds.

Theorem 1. *There exists some number $c > 0$ constant in sample size n such that for any $\epsilon \geq cn^{-2/(4+d)}$, we have*

$$R_{\epsilon}^{(t)}(\phi_{LE}; \mathcal{H}_1^d(L)) \leq \frac{1}{a^2}.$$

when $d = 3$ or 4.

Proof. □

1.3 Relating the eigenvectors to a grid.

For $\kappa = n^{1/d}$, let $\widetilde{X} = \{\widetilde{x} \in [\kappa]^d/\kappa\}$ be a grid over the unit cube. Consider the lattice graph $\mathrm{Grid} = (\widetilde{V}, \widetilde{E})$, with vertices $\widetilde{V} = \{v : v \in [\kappa]^d\}$, and edges $\widetilde{E} = \left\{(u, v) : \widetilde{u}, \widetilde{v} \in \widetilde{V}, \|\widetilde{u} - \widetilde{v}\|_1 = 1\right\}$, and write $\widetilde{L} = \widetilde{V} \Lambda \widetilde{V}^T$ for the eigendecomposition of the associated Laplacian matrix.

We wish to exhibit a mapping $P : \mathbb{R}^V \to \mathbb{R}^{\widetilde{V}}$ such that the difference between the statistics $T_{LE}^{(k)} = \ell^T V_k V_k^T \ell$ and $\widetilde{T}_{LE}^{(k)} = (P\ell)^T \widetilde{V}_k \widetilde{V}_k^T (P\ell)$ is small, which we accomplish through the use of optimal transport theory.

Lemma 1. *Fix $0 < \delta < 1$. For some number $c_1 > 0$, dependent on $f+g$, d, and δ, but not n, with probability at least $1-\delta$ there exists a mapping $T : X \to \widetilde{X}$ such that*

$$\|X - TX\|_\infty \leq c_1 \left(\frac{\log(n)}{n}\right)^{1/d} \tag{3}$$

If $r = c_1(\log n/n)^{1/d}$ and 3 holds, then letting $P(v)$ be the index of $T(x_v)$ on the grid graph, we have for every $\theta \in \mathbb{R}^V$,

$$(P\theta)^T \widetilde{L} P\theta \leq \theta^T L\theta \leq (\log n)^{2+1/d}(P\theta)^T \widetilde{L} P\theta. \tag{4}$$

1.4 Type I error

Under the null hypothesis, ℓ consists of n independent and identically distributed Rademacher random variables. Using this fact, we can bound the first two moments of T_{LE} under the null hypothesis.

Lemma 2. *If $f = g$, then*

$$\mathbb{E}(T_{LE}^{(k)}) = k \tag{5}$$

and

$$\mathrm{Var}(T_{LE}^{(k)}) \leq 2k + c_1^4\left(k + \frac{2k^2}{n}\right) \tag{6}$$

Proof. Taking the expectation first, we have

$$\begin{aligned}
\mathbb{E}(T_{LE}^{(k)}) &= \sum_{j=1}^{k} \mathbb{E}((\ell^T v_j)^2) \\
&= \sum_{j=1}^{k}\sum_{i=1}^{n}\sum_{i'=1}^{n} \mathbb{E}(\ell_i \ell_{i'}) v_{ji} v_{ji'} \\
&= \sum_{j=1}^{k}\sum_{i=1}^{n} v_{ji}^2 = k.
\end{aligned}$$

We defer the proof of (6) until the following subsection, where we show it in greater generality. □

1.5 Type II error.

Write $\ell =: \theta^\star + w$, where $\theta^\star = (\theta_i^\star)$ is defined by

$$\theta_i^\star = \frac{f(x_i) - g(x_i)}{f(x_i) + g(x_i)}$$

and $w = (w_i)$ therefore consists of n independent (although not identically distributed) zero-mean noise terms. We expand

$$\ell^T V_k V_k^T \ell = (\theta^\star)^T V_k V_k^T \theta^\star + w^T V_k V_k w + 2w^T V_k V_k^T \theta^\star, \tag{7}$$

and focus at first our attention on the second term in (7).

Lemma 3. *Assuming (2) holds,*

$$\mathbb{E}\left(w^T V_k V_k w\right) \geq k - \frac{k c_1^2 \delta^2(f,g)}{p_{\min}^2} \tag{8}$$

and

$$\mathrm{Var}(T_{LE}^{(k)}) \leq 2k + c_1^4\left(k + \frac{2k^2}{n}\right) \tag{9}$$

Proof. We note that, conditional on X,

$$w_i = \begin{cases} 2\frac{g(x_i)}{f(x_i)+g(x_i)}, & \text{with probability } \frac{f(x_i)}{f(x_i)+g(x_i)} \\ -2\frac{f(x_i)}{f(x_i)+g(x_i)}, & \text{with probability } \frac{g(x_i)}{f(x_i)+g(x_i)} \end{cases}$$

We first show (8). Expanding the quadratic form as a double sum, and using the law of iterated expectation, we have

$$\begin{aligned} \mathbb{E}\left(w^T V_k V_k w\right) &= \sum_{h=1}^{k}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\big(w_i w_j v_{hi} v_{hj}\big) \\ &= \sum_{h=1}^{k}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\Big(\mathbb{E}\Big(w_i w_j \Big| x_i, x_j\Big) v_{hi} v_{hj}\Big) \\ &= \sum_{h=1}^{k}\sum_{i=1}^{n}\mathbb{E}\Big(\mathbb{E}\Big(w_i^2 \Big| x_i\Big) v_{hi}^2\Big). \end{aligned} \tag{10}$$

The conditional expectation $\mathbb{E}\Big(w_i^2 \Big| x_i\Big)$ can be directly computed,

$$\begin{aligned} \mathbb{E}\Big(w_i^2 \Big| x_i\Big) &= 4\left(\frac{g(x_i)}{f(x_i)+g(x_i)}\right)^2 \frac{f(x_i)}{f(x_i)+g(x_i)} + 4\left(\frac{f(x_i)}{f(x_i)+g(x_i)}\right)^2 \frac{g(x_i)}{f(x_i)+g(x_i)} \\ &= 4\frac{f(x_i)g(x_i)(f(x_i)+g(x_i))}{(f(x_i)+g(x_i))^3} \\ &= 4\frac{f(x_i)g(x_i)}{(f(x_i)+g(x_i))^2}. \end{aligned}$$

and plugging this back into (10), we obtain

$$\begin{aligned} \mathbb{E}\left(w^T V_k V_k w\right) &= \sum_{h=1}^{k}\sum_{i=1}^{n}\mathbb{E}\left(4\frac{f(x_i)g(x_i)}{(f(x_i)+g(x_i))^2} v_{hi}^2\right) \\ &= \sum_{h=1}^{k}\sum_{i=1}^{n}\mathbb{E}\left(v_{hi}^2\left(1-\left(\frac{f(x_i)-g(x_i)}{f(x_i)+g(x_i)}\right)^2\right)\right) \\ &\geq k - \frac{c^2}{n}\sum_{h=1}^{k}\sum_{i=1}^{n}\mathbb{E}\left(\frac{f(x_i)-g(x_i)}{f(x_i)+g(x_i)}\right)^2 \\ &\geq k - \frac{k c_1^2 \delta^2(f,g)}{p_{\min}^2} \end{aligned}$$

and (8) is shown.

We now show (9). It will be helpful to note that for all $i, i', j, j' \in [n]$,

$$\mathrm{Cov}(w_i w_j, w_{i'} w_{j'} | X) = \begin{cases} 0, & \text{if } \{i, i', j, j'\} \text{ has four distinct elements} \\ 0, & \text{if } \{i, i', j, j'\} \text{ has three distinct elements} \\ 0, & \text{if } \{i, i', j, j'\} \text{ has two distinct elements and } i = j \\ \mathbb{E}(w_i^2|X)\mathbb{E}(w_j^2|X), & \text{if } \{i, i', j, j'\} \text{ has two distinct elements and } i \neq j \\ \mathrm{Var}(w_i^4|X), & \text{if } \{i, i', j, j'\} \text{ has one distinct element.} \end{cases}$$

and additionally

$$\mathbb{E}(w_i w_j|X) = \begin{cases} 0, & \text{if } i \neq j \\ E(w_i^2|X), & \text{if } i = j. \end{cases}$$

Then, by the law of total covariance,

$$\begin{aligned}
\mathrm{Var}(w^T V_k V_k^T w) &= \sum_{h,h'=1}^{k} \sum_{i,i'=1}^{n} \sum_{j,j'=1}^{n} \mathrm{Cov}\big(w_i w_j v_{hi} v_{hj}, w_{i'} w_{j'} v_{h'i'} v_{h'j'}\big) \\
&= \sum_{h,h'=1}^{k} \sum_{i,i'=1}^{n} \sum_{j,j'=1}^{n} \mathbb{E}\Big(v_{hi} v_{hj} v_{h'i'} v_{h'j'} \mathrm{Cov}\big(w_i w_j, w_{i'} w_{j'}|X\big)\Big) + \\
&\quad \mathrm{Cov}\Big(v_{hi} v_{hj} \mathbb{E}(w_i w_j|X), v_{h'i'} v_{h'j'} \mathbb{E}(w_i' w_j'|X)\Big) \\
&= 2 \sum_{h,h'=1}^{k} \sum_{i \neq j}^{n} \mathbb{E}\Big(v_{hi} v_{hj} v_{h'i} v_{h'j} \mathbb{E}(w_i^2|X) \mathbb{E}(w_j^2|X)\Big) + \\
&\quad \sum_{h,h'=1}^{k} \sum_{i=1}^{n} \mathbb{E}\left(v_{hi}^2 v_{h'i}^2 \mathrm{Var}(w_i^4|X)\right) + \\
&\quad \sum_{h,h'=1}^{k} \sum_{i,i'=1}^{n} \mathrm{Cov}\big(v_{hi}^2 \mathbb{E}(w_i^2|X), v_{h'i'}^2 \mathbb{E}(w_{i'}^2|X)\big)
\end{aligned}$$

Then, using the upper bound $\mathbb{E}(w_i^2|X), \mathbb{E}(w_j^2|X) \leq 1$, we can bound the first term

$$\begin{aligned}
\sum_{h,h'=1}^{k} \sum_{i \neq j}^{n} \mathbb{E}\Big(v_{hi} v_{hj} v_{h'i} v_{h'j} \mathbb{E}(w_i^2|X) \mathbb{E}(w_j^2|X)\Big) &\leq \sum_{h,h'=1}^{k} \sum_{i \neq j}^{n} \mathbb{E}\Big(v_{hi} v_{hj} v_{h'i} v_{h'j}\Big) \\
&\leq \sum_{h,h'=1}^{k} \mathbb{E}\Big(\sum_{i=1}^{n} v_{h'i} v_{hi} \sum_{j=1}^{n} v_{h'j} v_{hj}\Big) \\
&= \sum_{h=1}^{k} 1 = k.
\end{aligned}$$

To bound the second term, we note that $\mathrm{Var}(w_i^4|X) \leq 1$, and along with (2) this yields

$$\sum_{h,h'=1}^{k} \sum_{i=1}^{n} \mathbb{E}\left(v_{hi}^2 v_{h'i}^2 \mathrm{Var}(w_i^4|X)\right) \leq \sum_{h,h'=1}^{n} \sum_{i=1}^{n} \frac{c_1^4}{n^2} = \frac{c_1^4 k^2}{n}.$$

Finally, to bound the third term we note that if $h \neq h'$ and $i \neq i'$, $\mathrm{Cov}\big(v_{hi}^2 \mathbb{E}(w_i^2|X), v_{h'i'}^2 \mathbb{E}(w_{i'}^2|X)\big) = 0$.

Therefore,

$$\begin{aligned}\sum_{h,h'=1}^{k}\sum_{i,i'=1}^{n}\mathrm{Cov}\big(v_{hi}^2\mathbb{E}(w_i^2|X),v_{h'i'}^2\mathbb{E}(w_{i'}^2|X)\big) &= \sum_{h=1}^{k}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}\big(v_{hi}^2\mathbb{E}(w_i^2|X),v_{hi'}^2\mathbb{E}(w_{i'}^2|X)\big)+\\ &\quad\sum_{h\neq h'}^{k}\sum_{i=1}^{n}\mathrm{Cov}\big(v_{hi}^2\mathbb{E}(w_i^2|X),v_{h'i}^2\mathbb{E}(w_i^2|X)\big)\\ &\leq \sum_{h=1}^{k}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{c_1^4}{n^2}+\sum_{h\neq h'}^{k}\sum_{i=1}^{n}\frac{c_1^4}{n^2}\\ &= c_1^4k+c_1^4\frac{k^2}{n},\end{aligned}$$

and putting the pieces together yields (9). $\square$

Now, we turn our attention to the third term of (7).

Lemma 4.

$$\mathbb{E}(w^TV_kV_k^T\theta^\star)=0 \tag{11}$$

and assuming (2) *holds,*

$$\mathrm{Var}(w^TV_kV_k^T\theta^\star)\leq k\frac{c_1^2\delta^2(f,g)}{p_{\min}^2}$$

Finally, we examine the first term in (7). We will need the following result.

Lemma 5. *For any* $0<\delta<1$*, there exists a number* $c_2>0$ *depending only on* d *and* δ *such that, if* $r\geq c_2(\log n/n)^{1/d}$*, then*

$$\|D\theta^\star\|^2\leq\frac{16L^4}{p_{\min}^2}n^2r^{d+2}$$

and

$$\lambda_k\geq\frac{4k^{2/d}}{\pi^2n^{2/d}}$$

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

Lemma 6. *For any* $\delta>0$*, there exists a number* $c_2>0$ *constant in sample size such that if* $r\geq c_2(\log n/n)^{1/d}$*,*

$$(\theta^\star)^TV_kV_k^T\theta^\star\geq n(\delta(f,g))^2-\frac{n^2r^{d+2}n^{2/d}}{k^{2/d}}$$

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

Proof. Letting $z=D\theta^\star$, we have

$$\begin{aligned}(\theta^\star)^TV_kV_k^T\theta^\star &= \|\theta^\star\|^2-(\theta^\star)^T\left(I-V_kV_k^T\right)\theta^\star\\ &= \|\theta^\star\|^2-z^T(D^\dagger)^T(I-V_kV_k^T)D^\dagger z\\ &\geq \|\theta^\star\|^2-z^Tz\lambda_{\max}\Big((D^\dagger)^T(I-V_kV_k^T)D^\dagger\Big)\\ &\geq \|\theta^\star\|^2-\frac{z^Tz}{\lambda_k}\end{aligned}$$

Examining $\|\theta^\star\|^2$, we have that

$$\mathbb{E}(\|\theta^\star\|^2)=n\delta^2(f,g)\quad\text{and}\quad\mathrm{Var}(\|\theta^\star\|^2)\leq n\delta^2(f,g)$$

Turning to the second term, we have given our choice of r, along with the fact $f - g \in \mathcal{H}_1^d(2L)$,

$$z^T z = \|D\theta^\star\|^2 \leq \frac{16L^4}{p_{\min}^2} n^2 r^{d+2}$$

and additionally

$$\lambda_k \geq \frac{4k^{2/d}}{\pi^2 n^{2/d}},$$

each with probability $1 - \delta$. $\square$