

Notes for Week 6/1/19 - 6/5/19

Alden Green

June 7, 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 $f(x), g(x) > 0$ for all $x \in [0, 1]^d$.

We observe data (X, ℓ) , a design matrix and associated labels, specified as follows. We consider a fixed design setting, and more precisely we fix X to be a grid over the unit cube. Formally, for a given sample size $n \in \mathbb{N}$ and $\kappa = n^{1/d}$, let $X = \{x_k : k \in [\kappa]^d\}$ with $x_k = k/\kappa$. Our labels are then $\ell = \{\ell_k : k \in [\kappa]^d\}$, with ℓ_k independent random variables sampled according to

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

where dividing by $f(x) + g(x)$ is permissible as $f(x), g(x) > 0$ for all $x \in [0, 1]^d$.

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 Discretization error in fixed design setting.

Our first result is a lower bound. We exhibit density functions $f, g \in \mathcal{H}_1^d(\pi/2)$ such that $\delta^2(f, g) = n^{-2/d}$ for which, assuming fixed design X as we have done, no test can distinguish $f - g$ from the 0 function. For simplicity, assume κ is an odd integer.

¹Note 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.

Proposition 1. For $x \in [0, 1]^d$, let $f, g \in \mathcal{H}_1^d(\pi/2)$ be density functions given by

$$f(x) = 1 + \frac{1}{2\kappa} \prod_{i=1}^d \sin(\kappa\pi x_i), \quad g(x) = 1 - \frac{1}{2\kappa} \prod_{i=1}^d \sin(\kappa\pi x_i)$$

Then, $\delta^2(f, g) = n^{-2/d}$, and $f(x_k) = g(x_k)$ for every $k \in [\kappa]^d$.

As a result, $(f - g)(x_k) = 0$ for every $x_k \in X$, and so the labels ℓ have the same distribution as they would under the null hypothesis $f = g$.

Proof of Proposition 1. We begin by showing that f and g are density functions with Lipschitz norm bounded by 1. We compute the integral of f over its support,

$$\begin{aligned} \int_{[0,1]^d} f(x) dx &= 1 + \frac{1}{2\kappa} \int_{[0,1]^d} \prod_{i=1}^d \sin(\kappa\pi x_i) dx_i \\ &= 1 + \frac{1}{2\kappa} \left(\int_0^1 \sin(\kappa\pi x) dx \right)^d \\ &= 1 + \frac{1}{2\kappa} \left(-\cos(\kappa\pi x) \Big|_0^1 \right) = 1. \end{aligned}$$

where the last equality follows as κ is an odd integer. Similar calculations follow for g , and therefore f and g are density functions.

Next, we upper bound the Lipschitz norm of f (again, similar calculations will hold for g). Taking the partial derivative of f along any coordinate x_i , we obtain

$$\begin{aligned} \left| \frac{\partial f}{\partial x_i} \right| &= \left| \frac{\kappa\pi}{2\kappa} \cos(\kappa\pi x_i) \prod_{j \neq i}^d \sin(\kappa\pi x_j) \right| \\ &\leq \frac{\pi}{2}. \end{aligned}$$

Since this holds for each $i = 1, \dots, d$, f has Lipschitz norm upper bounded by $\pi/2$.

Now, we show $\delta^2(f - g) \geq 2^{-(d+2)} n^{-2/d}$,

$$\begin{aligned} \int_{[0,1]^d} (f(x) - g(x))^2 dx &= \frac{1}{4\kappa^2} \int_{[0,1]^d} \left(\prod_{i=1}^d \sin(\kappa\pi x_i) \right)^2 dx_i \\ &= \frac{1}{4\kappa^2} \left(\int_0^1 \sin^2(\kappa\pi x) dx \right)^d \\ &= \frac{1}{4\kappa^2} \left(\int_0^1 \frac{1 - \cos(2\kappa\pi x)}{2} dx \right)^d \\ &= \frac{1}{4\kappa^2} \left(\frac{1 - \frac{\sin(2\kappa\pi x)}{2\kappa\pi}}{2} \Big|_0^1 \right)^d \\ &= \frac{1}{42^d \kappa^2} \end{aligned}$$

with the last equality following from the fact κ is an integer.

Finally, for any $k \in [\kappa]^d$ and $x_k = k/\kappa$, we have

$$\sin(\kappa\pi(x_k)_i) = \sin(\pi k_i) = 0$$

and so $f(x_k) = 1 = g(x_k)$. \square

Remark 1. We note that by [1], the *minimax rate* for hypothesis testing over the Holder ball $\mathcal{H}_\beta^d(L)$ is of order $n^{-4\beta/(4\beta+d)}$. In particular, the aforementioned Lipschitz ball is simply $\mathcal{H}_1^d(L)$, and the minimax rate becomes $n^{-4/(4+d)}$. We note that this is faster than the lower bound $n^{-2/d}$ provided by Proposition 1. The correct interpretation is that the discretization error due to considering fixed grid design (and choosing negative examples with knowledge of that design) dominates other forms of error (i.e. squared bias and variance). Interestingly, this appears not to be true for estimation in the analogous case. Here, the minimax rate of estimation over the Lipschitz ball with L^2 risk is $n^{-2/(2+d)}$, which dominates the exhibited discretization error $n^{-2/d}$.

2 Sub-optimality of the discrete Sobolev IPM for fixed design.

For a function $h : [0, 1]^d \rightarrow \mathbb{R}$, write $\theta^h := (h(x_k) : x_k \in X)$ for the evaluation of h at the data. In the previous section, we exhibited a function $h := f - g$ such that $\|h\|_{L^2} \sim n^{-1/d}$, but $\|\theta^h\|_2 = 0$. To avoid this issue, let us shift slightly our requirements on f, g , to be placed directly on discrete norms rather than continuous norms. In particular, let D be the incidence matrix of a lattice graph $G = (V, E)$ over the design points X , where $V = [\kappa]^d$ and $(k, k') \in E$ if $\|k - k'\|_1 = 1$. Now, we will consider those density functions $f : [0, 1]^d \rightarrow \mathbb{R}$ such that the discrete Sobolev norm is bounded $\|D\theta^f\|_2 \leq n^{1/2-1/d}$. In particular, letting

$$\overline{\mathcal{W}}_1^d(L) = \left\{ f : \|D\theta^f\|_2 \leq n^{1/2-1/d} \right\}$$

for $f, g \in \overline{\mathcal{W}}_1^d(L)$ we now wish to test whether $\|\theta^f - \theta^g\|_2 = 0$.

A potential test statistic in this context is the discrete Sobolev integral probability metric (IPM), formally defined as

$$T_S := \sup_{\theta : \|D\theta\|_2 \leq C, \theta^T \mathbf{1} = 0} \frac{1}{n} \langle \theta, \ell \rangle$$

Letting $L = D^T D$ be the discrete Laplacian over the lattice, and L^\dagger be the pseudoinverse of L , a quick calculation yields $T_S = \frac{C}{n} \sqrt{\ell^T L^\dagger \ell}$. Is T_S an optimal test over vectors θ^f, θ^g in the discrete Sobolev ball, with respect to the empirical norm $\|\theta^f - \theta^g\|_2$? Propositions 2, 3 and 4 show that it is not. Hereafter, we let $\sqrt{\kappa}$ be an integer.

Proposition 2. Let $f, g \in \overline{\mathcal{W}}_1^d(\pi/2)$ be functions given by

$$f(x) = 1 + \frac{1}{2\sqrt{\kappa}} \prod_{i=1}^d \cos\left(x_i \sqrt{\kappa} \pi - \frac{\pi}{2\sqrt{\kappa}}\right), \quad g(x) = 1 - \frac{1}{2\sqrt{\kappa}} \prod_{i=1}^d \cos\left(x_i \sqrt{\kappa} \pi - \frac{\pi}{2\sqrt{\kappa}}\right)$$

Then the following statements hold:

- $\|\theta^f - \theta^g\|_2 = 2^{-(d+2)/2} n^{1/2-1/2d}$.
- $\|D\theta^f\|_2 \leq (2^{-(d+2)/2} \sqrt{d} \pi) n^{1/2-1/d}$.
- $\sup_{\theta : \|D\theta\|_2 \leq C, \theta^T \mathbf{1} = 0} \frac{1}{n} \langle \theta, \theta^f - \theta^g \rangle \leq C(d^{-1/2} 2^{-d/2}) n^{-1/2}$

Proof. Write $\theta^h := \theta^f - \theta^g$. In the one-dimensional case (i.e G being the path graph), it is well known that for any $k \in 1, \dots, \kappa$, the vector

$$v = (v_k)_{j=1}^\kappa, v_k = \cos\left(\frac{(j-1/2)k\pi}{\kappa}\right)$$

is a (scaled) eigenvector of L with eigenvalue $4 \sin^2(k\pi/2\kappa)$; precisely

$$Lv = (4 \sin^2(j\pi/2\kappa))v, \text{ and } \|v\| = (\kappa/2)^{1/2}$$

As $\sqrt{\kappa}$ is an integer, one can use standard facts about Cartesian product graphs to derive that θ^h is itself a (scaled) eigenvector of L ; precisely

$$L\theta^h = (4d \sin^2(\pi\sqrt{\kappa}/2\kappa))\theta^h, \text{ and } \|\theta^h\|_2 = \frac{1}{2\sqrt{\kappa}} \left(\frac{\kappa}{2}\right)^{d/2}.$$

The second equality proves the first claim. Both equalities together yield

$$\begin{aligned} \|D\theta^h\|_2 &= 2\sqrt{d} \sin(\pi\sqrt{\kappa}/2\kappa) \frac{1}{2\sqrt{\kappa}} \left(\frac{\kappa}{2}\right)^{d/2} \\ &\leq \frac{1}{2\kappa} \sqrt{d}\pi \left(\frac{\kappa}{2}\right)^{d/2} \end{aligned}$$

which proves the second claim. Finally, as we have noted $\sup_{\theta: \|D\theta\|_2 \leq C, \theta^T \mathbf{1} = 0} \frac{1}{n} \langle \theta, \theta^f - \theta^g \rangle = \frac{C}{n} \sqrt{(\theta^h)^T L^\dagger \theta^h}$, and since

$$\begin{aligned} \sqrt{(\theta^h)^T L^\dagger \theta^h} &= \frac{\|\theta^h\|_2}{2\sqrt{d} \sin(\pi\sqrt{\kappa}/2\kappa)} \\ &\leq \frac{1}{\sqrt{d}} \left(\frac{\kappa}{2}\right)^{d/2} \end{aligned}$$

the third claim follows. \square

Proposition 3 demonstrates that under the null hypothesis, T_S scales at approximately the same rate, $n^{-1/2}$, as it does for the choice of f and g in Proposition 2.

Proposition 3. Fix $0 < \delta < 1$. If $\theta^f - \theta^g = 0$, there exists a constant $C(d, \delta)$ based only on d and δ , such that with probability at least $1 - \delta$,

$$T_S \geq C(d, \delta) \frac{C}{\sqrt{n}}.$$

To prove Proposition 3, it will be helpful to introduce multiindex notation to refer to the eigenvalues and eigenvectors of L . Specifically, for $k, j \in [\kappa]^d, k = (k_1, \dots, k_d), j = (j_1, \dots, j_d)$, it is well known that the eigenvector v_k has coordinates $(v_k)_j$ and corresponding eigenvalue λ_k given by

$$\begin{aligned} (v_k)_j &= \left(\frac{2}{\kappa}\right)^{d/2} \prod_{i=1}^d \cos\left(\frac{(j_i - 1/2)k_i\pi}{\kappa}\right) \\ \lambda_k &= 4 \sum_{i=1}^d \sin^2\left(\frac{\pi(k_i - 1)}{2\kappa}\right). \end{aligned}$$

Proof. Throughout, let c be a constant which depends only on d and δ , which may change from line to line. We will show that under the null hypothesis $\theta^f - \theta^g = 0$, $\mathbb{E}(\ell^T L^\dagger \ell) = cn$, in which case the result follows from (e.g.) an application of Markov's inequality.

Under the null hypothesis $\theta^f - \theta^g = 0$, we note that ℓ consists of n i.i.d Rademacher random variables. Hence

$$\begin{aligned}
\mathbb{E}(\ell^T L^\dagger \ell) &= \sum_{k \in [\kappa]^d} \frac{(\ell^T v_k)^2}{\lambda_k} \\
&= \sum_{k \in [\kappa]^d} \frac{1}{\lambda_k} \\
&= \frac{1}{4} \sum_{k \in [\kappa]^d} \frac{1}{\sum_{i=1}^d \sin^2 \left(\frac{\pi(k_i-1)}{2\kappa} \right)} \\
&\geq \frac{\kappa^2}{\pi^2} \sum_{k \in [\kappa]^d} \frac{1}{\sum_{i=1}^d (k_i - 1)^2} \\
&\geq \frac{\kappa^2}{\pi^2} \int_{x \in [1, \kappa]^d} \frac{1}{\sum_{i=1}^d x_i^2} dx \\
&= c\kappa^2 \int_0^\kappa \frac{1}{r^2} r^{d-1} dr \\
&= c\kappa^2 \kappa^{d-2} = c\kappa^d.
\end{aligned}$$

and the claim follows since $\kappa^d = n$. \square

Together, Propositions 2 and 3 show that there exists functions $f, g \in \overline{\mathcal{W}}_1^d$ with empirical norm $\|\theta^f - \theta^g\|_2 \gtrsim n^{1/2-1/2d}$ such that T_S cannot distinguish $\theta^f - \theta^g$ from the zero vector.

Now, we turn our attention to a different test. Denote by $V_m = [v_1 \dots v_m]$ the $n \times m$ matrix containing the first m eigenvectors of L (that is, the eigenvectors corresponding to the smallest m eigenvalues of L , excluding the eigenvalue $\lambda = 0$). The *Laplacian eigenmaps* estimate of $\mathbb{E}(\ell) = \frac{\theta^f - \theta^g}{\theta^f + \theta^g}$ is the projection of ℓ onto the first m eigenvectors of L ,

$$\widehat{\theta}^{LE} = (V_m V_m^T) y.$$

Proposition 4 shows that a test based on $\|\widehat{\theta}^{LE}\|_2$ can distinguish any two $f, g \in \overline{\mathcal{W}}_1^d$ such that $\|\theta^f - \theta^g\|_2$ is at least roughly $n^{1/2-1/(d+2)}$. It is a consequence of the estimation results derived in [2].

Proposition 4. Fix $0 < \delta < 1$, and write $\widetilde{\theta}^h := (\theta^f - \theta^g)/(\theta^f + \theta^g)$. Assume $f, g \in \overline{\mathcal{W}}_1^d(L)$ satisfy some lower bound $f(x), g(x) \geq p_{\min}$ for all $x \in [0, 1]^d$. Then, there exists a constant $C(L, d, \delta, p_{\min})$ such that

- $\|\widehat{\theta}^{LE}\|_2 \geq \|\widetilde{\theta}^h\|_2 - C(L, d, \delta, p_{\min}) n^{1/2-1/(d+2)}$
- $\|\widehat{\theta}^{LE}\|_2 \leq \|\widetilde{\theta}^h\|_2 + C(L, d, \delta, p_{\min}) n^{1/2-1/(d+2)}$

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

Proof. We begin by noting that

$$\|D\widetilde{\theta}^h\|_2 \leq \frac{L}{p_{\min}}$$

In [2], it is shown that for $y = \widetilde{\theta}^h + w$, where w is i.i.d Normal noise with unit variance,

$$\|\widehat{\theta}^{LE} - \widetilde{\theta}^h\|_2^2 \leq cn \frac{1}{n^{\frac{2}{d+2}}}$$

where c is a constant depending on d , L , and p_{\min} . We take for granted that this result can be extended to the case $\ell = \tilde{\theta}^h + w$, where w consists of independent but not necessarily identically distributed noise with mean zero and variance no greater than 1. Both claims then follow from the triangle inequality. \square

REFERENCES

- [1] Ery Arias-Castro, Bruno Pelletier, and Venkatesh Saligrama. Remember the curse of dimensionality: the case of goodness-of-fit testing in arbitrary dimension. *Journal of Nonparametric Statistics*, 30(2):448–471, 2018.
- [2] Veeranjaneyulu Sadhanala, Yu-Xiang Wang, and Ryan J Tibshirani. Total variation classes beyond 1d: Minimax rates, and the limitations of linear smoothers. In *Advances in Neural Information Processing Systems*, pages 3513–3521, 2016.