

# Notes for Week 6/13/19 - 6/18/19

Alden Green

June 22, 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 above and below,  $0 < p_{\min} < f(x), g(x) < p_{\max} < \infty$  for all  $x \in [0, 1]^d$ .

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

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

We will consider the function class  $\mathcal{H} := \mathcal{H}_1^d(L)$ , the  $d$ -dimensional Lipschitz ball with norm  $L$ . Formally, a function  $f \in \mathcal{H}$  if for all  $x, y \in [0, 1]^d$ ,  $|f(x) - f(y)| \leq L \|x - y\|$ .<sup>1</sup>

## 1 Test Statistic

Let  $r > 0$  be a bandwidth parameter, let  $\eta_r : \mathcal{X} \times \mathcal{X} \rightarrow [0, \infty)$  be a Mercer kernel given by

$$\eta_r(x, y) = \begin{cases} 1, & \|x - y\| \leq r \\ 0, & \|x - y\| > r, \end{cases}$$

---

<sup>1</sup>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.

and let  $\tilde{A}$  be an  $n \times n$  matrix with entries  $\tilde{A}_{ij} = n^{-1}\eta(x_i, x_j)$ . We obtain  $A$  from  $\tilde{A}$  by setting all diagonal entries of  $\tilde{A}$  to be zero. Our test statistic is then

$$T_\eta := \ell^T A \ell.$$

## 2 Theory

We show that a simple test based on the statistic  $T_\eta$ , with proper choice of the bandwidth parameter  $r$ , achieves minimax optimality over the class of Lipschitz functions  $\mathcal{H}_1^d(L)$ .

**Theorem 1.** Fix  $r = n^{-2/(4+d)}$ , and consider the test

$$\phi_{\eta, \tau} = \mathbb{I}\{T_\eta \geq \tau\}, \quad \tau = an^{-d/(4+d)}$$

with  $a \geq 1$ . There are constants  $c_1$  depending only on  $p_{\max}$ , and  $c_2$  depending only on  $(L, p_{\max}, p_{\min})$ , such that

$$R_\epsilon^n(\phi_{\eta, \tau}; \mathcal{H}_1^d(L)) \leq c_1/a^2, \quad \text{if } \epsilon \geq c_2 an^{-2/(4+d)}$$

*Proof.* Our first objective is to show that, for an appropriate choice of  $c_1$ ,  $\mathbb{E}_{f, f}^n(\phi_{\eta, \tau}) \leq \frac{c_1}{2a^2}$  for all  $f \in \mathcal{H}_1^d(L)$ . This is equivalent to proving that

$$\mathbb{P}(T_\eta \geq \tau) \leq c_1/2a^2 \tag{1}$$

which we now proceed to show.

From Lemma 2, we have that  $\mathbb{E}_{f, f}(T_\eta) = 0$ . Therefore, by Chebyshev's inequality,

$$\mathbb{P}(T_\eta \geq \tau) \leq \text{Var}(T_\eta)/\tau^2 \tag{2}$$

In Lemma 3, we have upper bounded  $\text{Var}(T_\eta)$  as function of  $r$  and  $\delta^2(f, g)$ . Under the null hypothesis  $f = g$ ,  $\delta^2(f, g) = 0$ , and along with our choice of  $r = n^{-2/(d+4)}$ , the upper bound on  $\text{Var}(T_\eta)$  reduces to

$$\text{Var}(T_\eta) \leq \frac{p_{\max}^2}{2} n^{-2d/(d+4)}$$

By our choice of  $\tau = an^{-d/(4+d)}$ , we therefore have

$$\text{Var}(T_\eta)/\tau^2 \leq \frac{p_{\max}^2}{2a^2}$$

and therefore (1) holds with  $c_1 := p_{\max}^2$ .

We turn now to bound the type II error,  $\mathbb{E}_{f, g}^{(n)}(\phi_{\eta, \tau})$ , for all  $f, g \in \mathcal{H}_1^d(L)$  such that  $\delta^2(f, g) \geq \epsilon = c_2 n^{-2/(4+d)}$ . Our goal now will be to show

$$\mathbb{P}(T_\eta \leq \tau) \leq c_1/2a^2.$$

By Chebyshev's inequality we obtain

$$\mathbb{P}(T_\eta \leq \tau) \leq \frac{\text{Var}(T_\eta)}{(\mathbb{E}_{f, g}(T_\eta) - \tau)^2} \tag{3}$$

We obtain the following lower bound on  $\text{Var}(T_\eta)$  in Lemma 3 in terms of  $r$  and  $\delta(f, g)$ :

$$\text{Var}(T_\eta) \leq 2n \frac{p_{\max}^2}{p_{\min}^2} r^d (r^d \delta^2(f, g) + L r^{d+1} \delta(f, g)) + \frac{p_{\max}^2}{2} r^d$$

We will split our analysis into cases, based on which term on the right hand side of the preceding equation dominates. **TODO: Complete this proof.**

**Case 1, 1st term dominates:** We assume

$$2n \frac{p_{\max}^2}{p_{\min}^2} r^d (r^d \delta^2(f, g) + Lr^{d+1} \delta(f, g)) > \frac{p_{\max}^2}{2} r^d,$$

so that the variance is upper bounded

$$\text{Var}(T_\eta) \leq 4n \frac{p_{\max}^2}{p_{\min}^2} r^d (r^d \delta^2(f, g) + Lr^{d+1} \delta(f, g)).$$

Additionally, in the following case, we will show that for a sufficiently large choice of  $c_2$ , depending on  $p_{\max}, p_{\min}, L$  and  $d$ , we obtain

$$\mathbb{E}_{f,g} T_\eta \geq 2\tau$$

so that

$$(\mathbb{E}_{f,g} T_\eta - \tau)^2 \geq (\mathbb{E}_{f,g} T_\eta)^2 / 4.$$

In Lemma 2, moreover, we prove the following lower bound on  $\mathbb{E}_{f,g} T_\eta$  as a function of  $r$  and  $\delta(f, g)$ :

$$\mathbb{E}_{f,g} T_\eta \geq \frac{(n-1)p_{\min}^2}{p_{\max}^2} \left( \delta^2(f, g) \frac{r^d}{2^d} - L\delta(f, g) r^{d+1} \right).$$

We can therefore upper bound  $\text{Var}(T_\eta) / (\mathbb{E}_{f,g} T_\eta - \tau)^2$ ,

$$\text{Var}(T_\eta) / (\mathbb{E}_{f,g} T_\eta - \tau)^2 \leq \frac{p_{\max}^6}{p_{\min}^6} \frac{16nr^{2d}\delta(f, g)(\delta(f, g) + Lr)}{(n-1)^2 r^{2d} \delta^2(f, g)(\delta(f, g) 2^{-d} - Lr)^2}$$

**Case 2, 2nd term dominates:** We first lower bound the denominator, and then upper bound the numerator. In Lemma 2, we prove the following lower bound on  $\mathbb{E} T_\eta$  when  $f, g \in \mathcal{H}_1^d(L)$ , stated as a function of  $r$  and  $\delta^2(f, g)$ :

$$\mathbb{E}_{f,g} T_\eta \geq \frac{(n-1)p_{\min}^2}{p_{\max}^2} \left( \delta^2(f, g) \frac{r^d}{2^d} - L\delta(f, g) r^{d+1} \right). \quad (4)$$

Given the choice of bandwidth parameter  $r = n^{-2/(d+4)}$  and the lower bound  $\delta(f, g) \geq c_2 a n^{-2/(d+4)}$ , we may further simplify the expression on the right hand side of (4),

$$\begin{aligned} \frac{(n-1)p_{\min}^2}{p_{\max}^2} \left( \delta^2(f, g) \frac{r^d}{2^d} - L\delta(f, g) r^{d+1} \right) &= \frac{(n-1)p_{\min}^2}{p_{\max}^2} n^{-2d/(d+4)} \delta(f, g) \left( \frac{\delta(f, g)}{2^d} - L n^{-2/(d+4)} \right) \\ &\geq \frac{(n-1)p_{\min}^2}{p_{\max}^2} c_2 a n^{-(2d+2)/(d+4)} \left( \frac{c_2 a n^{-2/(d+4)}}{2^d} - L n^{-2/(d+4)} \right) \\ &\geq 2a^2 n^{1-(2d+2)/(d+4)-2/(d+4)} = 2a^2 n^{-d/(d+4)}. \end{aligned}$$

where the last inequality follows when  $c_2$  is chosen to be sufficiently large relative to  $p_{\min}, p_{\max}, L$  and  $d$ . Given our choice  $\tau = a n^{-d/(d+4)}$ , the denominator of (3) can therefore be lower bounded

$$(\mathbb{E}_{f,g}(T_\eta) - \tau)^2 \geq (2a^2 - a)^2 n^{-2d/(d+4)} \geq a^4 n^{-2d/(d+4)}$$

where the last inequality follows from the fact  $2a^2 - a \geq a^2$  for all  $a \geq 1$ .

We turn now to upper bounding the numerator of (3). In Lemma 3 we give the following upper bound on  $\text{Var}_{f,g} T_\eta$  as a function of  $r$  and  $\delta(f, g)$ ,

$$\text{Var}_{f,g} T_\eta \leq 2n \frac{p_{\max}^2}{p_{\min}^2} r^d (r^d \delta^2(f, g) + Lr^{d+1} \delta(f, g)) + \frac{p_{\max}^2}{2} r^d$$

Similar to before, we apply our choice of  $r$  and lower bound on  $\delta(f, g)$  and obtain

$$2n \frac{p_{\max}^2}{p_{\min}^2} r^d (r^d \delta^2(f, g) + Lr^{d+1} \delta(f, g)) + \frac{p_{\max}^2}{2} r^d \leq$$

□

### 3 Supporting Theory

We begin by recalling Chebyshev's inequality, which we will use to bound deviations of  $T_\eta$  from its expectation.

**Lemma 1** (Chebyshev). *Let  $X$  be a random variable with  $\mathbb{E}X < \infty$  and  $0 < \text{Var}(X) < \infty$ . Then,*

$$\mathbb{P}(|X - \mathbb{E}X| \geq a) \leq \frac{\text{Var}(X)}{a^2}.$$

In order to employ Lemma 1, we must bound the expectation and variance of  $T_\eta$  under the null and alternative hypotheses. We begin with the expectation  $\mathbb{E}T_\eta$ .

**Lemma 2.** *Under the null hypothesis  $h = g$ , for any  $h, g \in \mathcal{H}_1^d(L)$ ,*

$$\mathbb{E}T_\eta = 0.$$

*Under the alternative hypothesis  $h \neq g$ , for any  $h, g \in \mathcal{H}_1^d(L)$ ,*

$$\mathbb{E}T_\eta \geq \frac{(n-1)p_{\min}^2}{p_{\max}^2} \left( \delta^2(f, g) \frac{r^d}{2^d} - L\delta(f, g)r^{d+1} \right)$$

*Proof.* Let  $\theta^*$  be a vector in  $\mathbb{R}^n$ , with entries  $\theta_i^* = \mathbb{E}(\ell_i | x_i)$ . Letting  $w = \ell - \theta^*$ , we note that conditional on  $X$ , the noise component  $w$  consists of  $n$  independent and zero mean terms.

We expand the test statistic

$$T_\eta = (\theta^*)^T A \theta^* + w^T A w + 2w^T A \theta^*$$

and note that by the law of iterated expectation,

$$\mathbb{E}(T_\eta) = \mathbb{E}(T_\eta | X) = \mathbb{E}((\theta^*)^T A \theta^*)$$

Now, let  $X$  and  $Y$  be independent random variables with distribution  $\mu$ , and let  $\Delta(x) := f(x) - g(x)$ . By the linearity of expectation,

$$\begin{aligned} \mathbb{E}(\theta^*)^T A \theta^* &= \sum_{i \neq j=1}^n \mathbb{E}(\theta_i^* \theta_j^* A_{ij}) \\ &= (n-1) \mathbb{E}(\theta_1^* \theta_2^* \eta_r(x_1, x_2)) \\ &\geq \frac{4(n-1)}{p_{\max}^2} \mathbb{E}(\Delta(x) \Delta(y) \eta_r(X, Y)) \end{aligned}$$

We can lower bound  $\mathbb{E}(\Delta(x) \Delta(y) \eta_r(X, Y))$  by  $\delta^2(f, g)$ , minus a remainder term, as follows

$$\begin{aligned} \mathbb{E}(\Delta(x) \Delta(y) \eta_r(X, Y)) &= \iint \Delta(x) \Delta(y) \eta_r(x, y) d\mu(x) d\mu(y) \\ &\geq \frac{p_{\min}^2}{4} \iint \Delta(x) \Delta(y) \eta_r(x, y) dx dy \\ &\geq \frac{p_{\min}^2}{4} \left( \iint \Delta^2(x) \eta_r(x, y) dx dy - \iint |\Delta(x)| |\Delta(x) - \Delta(y)| \eta_r(x, y) dx dy \right) \end{aligned}$$

To lower bound the first term, observe that for any  $x \in [0, 1]^d$ ,

$$\int_{[0, 1]^d} \eta_r(x, y) dx \geq \frac{r^d}{2^d}$$

and therefore

$$\iint \Delta^2(x) \eta_r(x, y) dx dy \geq \frac{r^d}{2^d} \int_{[0,1]^d} \Delta^2(x) dx \geq \frac{r^d}{2^d} \delta^2(f, g).$$

To upper bound the remainder term, observe that  $\Delta \in \mathcal{H}_1^d(2L)$ . Using this fact, we obtain the upper bound

$$\begin{aligned} \iint |\Delta(x)| |\Delta(x) - \Delta(y)| \eta(x, y) dx dy &\leq: \iint |\Delta(x)| \|x - y\| \eta(x, y) dx dy \\ &\leq L r^{d+1} \int_{[0,1]^d} |\Delta(x)| dx \\ &\leq L r^{d+1} \delta(f, g). \end{aligned}$$

where the last inequality follows from Jensen's inequality.  $\square$

**Lemma 3.** For any  $h, g \in \mathcal{H}_1^d(L)$ ,

$$\text{Var}(T_\eta) \leq 2n \frac{p_{\max}^2}{p_{\min}^2} r^d (r^d \delta^2(f, g) + L r^{d+1} \delta(f, g)) + \frac{p_{\max}^2}{2} r^d$$

*Proof.* We begin by computing the second moment  $\mathbb{E}T_\eta^2$ . Expanding the quadratic form, we obtain

$$\mathbb{E}T_\eta^2 = \sum_{i, i'=1}^n \sum_{j, j'=1}^n \mathbb{E}(\ell_i \ell_{i'} \ell_j \ell_{j'} A_{ii'} A_{jj'}) \quad (5)$$

We next compute this expectation, based on the number of distinct indices  $i, i', j, j'$ . Let  $X, X', Y, Y'$  be four independent random variables, each with distribution  $\mu$ , and write  $\tilde{\Delta} := 2\Delta/\mu$ . Then, applying the law of conditional expectation, we obtain

$$n^2 \mathbb{E}(\ell_i \ell_{i'} \ell_j \ell_{j'} A_{ii'} A_{jj'}) = \begin{cases} \mathbb{E}(\tilde{\Delta}(X) \tilde{\Delta}(X') \eta_r(X, X'))^2, & \text{if } \{i, i', j, j'\} \text{ contains four distinct elements,} \\ \mathbb{E}(\tilde{\Delta}(X') \tilde{\Delta}(Y') \eta_r(X, X') \eta_r(X, Y')), & \text{if } \{i, i', j, j'\} \text{ contains three distinct elements and } i = j, \\ \mathbb{E}(\eta_r(X, Y)^2), & \text{if } \{i, i', j, j'\} \text{ contains two distinct elements and } i = j. \\ 0, & \text{if } i = i' \text{ or } j = j'. \end{cases}$$

and as a result we obtain

$$\begin{aligned} n^2 \mathbb{E}T_\eta^2 &= n(n-1)(n-2)(n-3) \mathbb{E}(\tilde{\Delta}(X) \tilde{\Delta}(X') \eta_r(X, X'))^2 + \\ &\quad 2n(n-1)(n-2) \mathbb{E}(\tilde{\Delta}(X') \tilde{\Delta}(Y') \eta_r(X, X') \eta_r(X, Y')) + \\ &\quad 2n(n-1) \mathbb{E}(\eta_r(X, Y)^2). \end{aligned} \quad (6)$$

The first term – covering the case when  $\{i, i', j, j'\}$  contains four distinct elements – will be canceled out by the mean, as we will see, and we do not simplify it further.

The second term we upper bound by methods similar to those used to obtain the upper bound in Lemma 2. We will need the following upper bound,

$$\begin{aligned} \int \eta(x, x') \eta(x, y') dx &\leq r^d \mathbb{I}(\|x' - y'\| \leq 2r) \\ &= r^d \eta_{2r}(x', y'). \end{aligned}$$

We proceed,

$$\begin{aligned}
\mathbb{E}\left(\tilde{\Delta}(X')\tilde{\Delta}(Y')\eta_r(X, X')\eta_r(X, Y')\right) &\leq \frac{p_{\max}^2}{p_{\min}^2} \iiint \Delta(x')\Delta(y')\eta(x, x')\eta(x, y') dx dy' dx' \\
&\leq \frac{p_{\max}^2}{p_{\min}^2} r^d \iint \Delta(x')\Delta(y')\eta_{2r}(x', y') dy' dx' \\
&\leq \frac{p_{\max}^2}{p_{\min}^2} r^d \left(r^d \delta^2(f, g) + Lr^{d+1} \delta(f, g)\right). \tag{7}
\end{aligned}$$

The following upper bound on the third term is a straightforward consequence of the fact  $\eta_r^2 = \eta_r$ , and therefore

$$\mathbb{E}(\eta_r(X, Y)^2) = \mathbb{E}(\eta_r(X, Y)) \leq \frac{p_{\max}^2}{4} r^d. \tag{8}$$

Finally, we have that

$$n^2 \mathbb{E}(T_\eta^2) = \left(n(n-1) \mathbb{E}(\tilde{\Delta}(X)\tilde{\Delta}(X')\eta_r)\right)^2. \tag{9}$$

Combining (6)-(9), we obtain

$$\text{Var}(T_\eta) \leq 2n \frac{p_{\max}^2}{p_{\min}^2} r^d \left(r^d \delta^2(f, g) + Lr^{d+1} \delta(f, g)\right) + \frac{p_{\max}^2}{2} r^d.$$

□