Notes for Week 7/4/19 - 7/9/19

Alden Green

August 1, 2019

Let $s \geq 0$. Suppose $\theta \in \ell^2(\mathbb{Z}^d)$ is a sequence such that

$$\sum_{k \in \mathbb{Z}^d} \theta_k^2 c_k^2 \leq 1$$

where for $k = (k_1, \ldots, k_d) \in \mathbb{Z}^d$

$$c_k = \left(\sum_{i=1}^{d} (2\pi k_i)^2 \right)^{s/2}.$$

Our goal is hypothesis testing; precisely, we would like to determine whether

$$\mathbf{H_0} : \|\theta\| = 0, \quad \text{or } \mathbf{H_a} : \|\theta\| \neq 0.$$

1 Testing in the Gaussian Sequence Model

We observe

$$y_k = \theta_k + \frac{1}{\sqrt{n}} \epsilon_k, \quad k \in \mathbb{Z}^d$$

where $\epsilon_k \sim \mathcal{N}(0,1)$ are independent and identically distributed for all $k \in \mathbb{Z}^d$. To test whether θ belongs to $\mathbf{H}_0$ or $\mathbf{H}_a$, we consider the test statistic

$$T_C = \sum_{k : c_k \leq C} y_k^2.$$

1.1 Supplemental Theory

Let $N(C) := \sharp \left\{ k \in \mathbb{Z}^d : c_k \leq C \right\}$. Recall that $\ell^2(\mathbb{Z}^d)$ is a normed space, where for $\theta \in \ell^2$

$$\|\theta\|^2 = \sum_{k \in \mathbb{Z}^d} \theta_k^2$$

Lemma 1. *Under $\mathbf{H}_0$, $\mathbb{E}(T_C) = N_C/n$. Under $\mathbf{H}_a$,*

$$\mathbb{E}(T_C) \geq \frac{N_C}{n} + \|\theta\|^2 - (2\pi C)^{-2}.$$

Proof. We have that

$$\begin{aligned} \mathbb{E}(T_C) &= \sum_{k : c_k \leq C} \mathbb{E}(y_k^2) \\ &= \frac{N(C)}{n} + \sum_{k : c_k \leq C} \theta_k^2 \end{aligned}$$

whence the expectation under the null hypothesis is obvious. Under $\mathbf{H}_a$, we have

$$\begin{aligned}\sum_{k:c_k\leq C}\theta_k^2 &\geq \|\theta\|^2 - \sum_{k:c_k>C}\theta_k^2\\ &\geq \|\theta\|^2 - (2\pi C)^{-2}\sum_{k:c_k>C}\theta_k^2 c_k^2\\ &\geq \|\theta\|^2 - (2\pi C)^{-2}.\end{aligned}$$

□

Lemma 2. *Under either $\mathbf{H}_0$ or $\mathbf{H}_a$, we have*

$$\mathrm{Var}(T_C) \leq 2\frac{N(C)}{n^2} + 4\frac{\mathbb{E}(T_C)}{n}$$

Proof.

$$\begin{aligned}\mathrm{Var}(T_C) &= \mathrm{Var}\Big(\sum_{k:c_k\leq C} y_k^2\Big)\\ &= \sum_{k:c_k\leq C}\mathrm{Var}(y_k^2)\\ &= \sum_{k:c_k\leq C}\frac{\mathrm{Var}(\epsilon_k^2)}{n^2} + \frac{4\theta_k^2\mathrm{Var}(\epsilon_k)}{n}\end{aligned}$$

and the statement follows from properties of the standard normal distribution. □

For $b \geq 1$, let $\tau(b) := b\sqrt{6N(C)/n^2}$. From here forward, let $C = n^{2s/(4s+d)}$. Our test will be

$$\Gamma = \mathbb{I}\left\{T_C \geq N(C)/n + \tau(b)\right\}$$

The following bound on the Type I error of our test follows immediately from Chebyshev's inequality.

Lemma 3. *Under the null hypothesis $\mathbf{H}_0$*

$$\mathbb{P}_0(T_C \geq N(C)/n + \tau(b)) \leq \frac{1}{b^2}$$

More technical work will be required to show the desired bound on type II error, which holds when $\|\theta\|^2$ is sufficiently large.

Lemma 4. *Suppose*

$$\|\theta\|^2 \geq (2\pi C)^{-2} + 2\tau(b)$$

Then

$$\mathbb{P}_a(T_C \leq N(C)/n + \tau(b)) \leq 2\left(\frac{1}{6b^2N(C)} + 2\frac{1}{b\sqrt{6N(C)}} + \frac{1}{4b^2}\right)$$

Before proving Lemma 4, we note that $N(C) \leq C^{d/s}$. (In fact, tighter bounds exist, but we will not need them.) Therefore

$$(2\pi C)^{-2} + 2\tau(b) \leq \frac{1}{4\pi^2}n^{-4s/(4s+d)} + \sqrt{24}bn^{-4s/(4s+d)}$$

and so the critical radius $\|\theta\|^2 \geq (2\pi C)^{-2} + 2\tau(b)$ is minimax optimal. We turn now to the proof of Lemma 4.

Proof of Lemma 4. We note that by hypothesis, we have that

$$\mathbb{E}_a(T_C) \geq \frac{N(C)}{n} + 2\tau(b). \tag{1}$$

By Chebyshev's inequality, we therefore have

$$\begin{aligned}
\mathbb{P}_a(T_C \leq N(C)/n + \tau(b)) &= \mathbb{P}_a(T_C - \mathbb{E}_a(T) \leq N(C)/n + \tau(b) - E_a(T)) \\
&\leq \frac{\mathrm{Var}_a(T_c)}{(\mathbb{E}_a(T) - \tau(b) - N(C)/n)^2} \\
&\leq 4\frac{\mathrm{Var}_a(T_c)}{(\mathbb{E}_a(T) - N(C)/n)^2} && (1) \\
&\leq 8\frac{N(C)/n^2 + \mathbb{E}_a(T_C)/n}{(\mathbb{E}_a(T) - N(C)/n)^2} && (\text{Lemma } 2)
\end{aligned}$$

Letting $\Delta = \mathbb{E}_a(T) - N(C)/n$, and noting that $\Delta \geq 2\tau(b)$, we obtain

$$\begin{aligned}
8\frac{N(C)/n^2 + \mathbb{E}_a(T_C)/n}{(\mathbb{E}_a(T) - N(C)/n)^2} &= 8\frac{N(C)/n^2 + \Delta/n + N(C)/n}{\Delta^2} \\
&\geq 2\left(\frac{N(C)}{\tau^2(b)n^2} + 2\frac{1}{\tau(b)n} + \frac{N(C)}{n\tau^2(b)}\right) \\
&= 2\left(\frac{1}{6b^2 N(C)} + 2\frac{1}{b\sqrt{6N(C)}} + \frac{1}{4b^2}\right)
\end{aligned}$$

□

2 Testing in the Nonparametric Regression Model

Let $\mathcal{D} = [0,1]^d$ and consider the Sobolev unit ball $W_d^{s,2}(1)$ of functions supported over $\mathcal{D}$. Let $\{\phi_k : k \in \mathbb{Z}^d\}$ be the tensor product Fourier basis of $W_d^{s,2}$. Note that for any $f \in W_d^{s,2}(1)$, letting

$$f(x) := \sum_{k \in \mathbb{Z}^d} \theta_k \phi_k(x)$$

we can show that

$$\|f\|_{L^2} = \|\theta\|_2, \ \|f\|_{W^{2,s}}^2 = \sum_{k \in \mathbb{Z}^d} \theta_k^2 c_k^2.$$

Therefore, testing whether $\|f\|_{L^2} = 0$ or $\|f\|_{L^2} > 0$ is exactly the same as testing whether $\|\theta\|_2 = 0$.

Now, however, for $i \in [n]$, assume we observe

$$z_i = f(x_i) + \varepsilon_i, \quad x_i \overset{\text{i.i.d}}{\sim} \mathrm{Unif}\big([0,1]^d\big), \quad \varepsilon_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1)$$

To test whether $\|\theta\|_2 = 0$, we consider the test statistic

$$T_C = \sum_{k : c_k \leq C} \widetilde{y}_k^2$$

where for $k \in \mathbb{Z}^d$,

$$\widetilde{y}_k := \frac{1}{n}\sum_{i=1}^n z_i \phi_k(x_i).$$

2.1 Supplemental Theory

We will need to make a pair of additional regularity assumptions beyond $f \in W_d^{k,2}(1)$.

(A1) For every $x \in [0,1]^d$ and for all $C > 0$,

$$\sum_{k: c_k \le C} \phi_k^2(x) = N(C)$$

(A2) The regression function $f \in L^4$.

It is not hard to check that the tensor product Fourier basis satisfies (A1).

Lemma 5. *Under* $\mathbf{H}_0$, $\mathbb{E}(T_C) = N(C)/n$. *Under* $\mathbf{H}_a$,

$$\mathbb{E}(T_C) \ge \frac{N(C)}{n} + \frac{\|\theta\|^2 N(C)}{n} + \frac{(n-1)}{n}\left(\|\theta\|^2 - (2\pi C)^{-2}\right)$$

Proof. We write

$$\widetilde{y}_k = \frac{1}{n}\sum_{i=1}^n f(x_i)\phi_k(x_i) + \frac{1}{n}\sum_{i=1}^n \varepsilon_i \phi_k(x_i) =: \widetilde{\theta}_k + \widetilde{\epsilon}_k.$$

To compute the expectation of $\widetilde{y}_k^2$, we therefore must compute the expectation of each of $\widetilde{\theta}_k^2$ and $\widetilde{\epsilon}_k^2$. (It is not hard to see that $\mathbb{E}(\widetilde{\theta}_k \widetilde{\epsilon}_k) = 0$). We have

$$\begin{aligned}
\mathbb{E}(\widetilde{\theta}_k^2) &= \mathbb{E}\left(\frac{1}{n^2}\sum_{i,j=1}^n f(x_i)f(x_j)\phi_k(x_i)\phi_k(x_j)\right) \\
&= \frac{1}{n}\mathbb{E}(f^2(x_1)\phi_k^2(x_1)) + \frac{(n-1)}{n}\mathbb{E}\big(f(x_1)\phi_k(x_1)\big)^2 \\
&= \frac{1}{n}\mathbb{E}(f^2(x_1)\phi_k^2(x_1)) + \frac{(n-1)}{n}\theta_k^2.
\end{aligned}$$

In addition,

$$\mathbb{E}(\widetilde{\epsilon}_k^2) = \frac{1}{n}\mathbb{E}(\varepsilon_1^2 \phi_k^2(x_1)) = \frac{1}{n}.$$

Therefore,

$$\mathbb{E}(T_C) = \sum_{k: c_k \le C} \frac{1}{n}\mathbb{E}(f^2(x_1)\phi_k^2(x_1)) + \frac{(n-1)}{n}\theta_k^2 + \frac{1}{n}$$

Under the null hypothesis $\|f\| = 0$, the first two terms are zero, and we are left with

$$\mathbb{E}(T_C) = \frac{N(C)}{n}$$

Under the alternative, calculations similar to those used in the proof of Lemma 1 along with assumption (A1) lead to the desired conclusion. □

For any $C > 0$, let the projection operator $P_C : \ell_2(\mathbb{Z}^d) \to \ell_2(\mathbb{Z}^d)$ be given by:

$$(P_C\theta)_k = \theta_k \mathbf{1}\{c_k \le C\}.$$

Lemma 6. *Under either* $\mathbf{H}_0$ *or* $\mathbf{H}_a$, *we have*

$$\mathrm{Var}(T_C) \le \frac{1}{n}\left(\|P_C\theta\|^2 + N(C)\|\theta\|^2\|P_C\theta\|^2\right) + \frac{1}{n^2}\left(N(C) + N(C)^2\|\theta\|^4 + 2N(C)\|\theta\|^2\right) + \frac{\mu_4 N(C)^2}{n^3} \tag{2}$$

Proof. We seek to upper bound $\mathrm{Cov}(\widetilde{y}_k, \widetilde{y}_{k'})$. We begin with the following decomposition:

$$\mathrm{Cov}(\widetilde{y}_k, \widetilde{y}_{k'}) = \frac{1}{n^4} \sum_{i,i',j,j'=1}^{n} \mathrm{Cov}\big(z_i z_j \phi_k(x_i)\phi_k(x_j), z_{i'} z_{j'} \phi_{k'}(x_{i'})\phi_{k'}(x_{j'})\big)$$

We split the summand into six cases based on the number of distinct elements $\{i, j, i', j'\}$, and analyze each case separately. Let

$$\chi_k^2 := \mathbb{E}(f^2(x_1)\phi_k(x_1)^2), \quad \mu_4 := \mathbb{E}(f^4(x_1))$$

and note that by assumption (A1), $\sum_{k:c_k\le C} \chi_k^2 = \|\theta\|^2 N(C)$ and by assumption (A2), $\mu_4 < \infty$.

Case 1: $\{i, j, i', j'\}$ has 4 distinct elements

$$\mathrm{Cov}\big(z_i z_j \phi_k(x_i)\phi_k(x_j), z_{i'} z_{j'} \phi_{k'}(x_{i'})\phi_{k'}(x_{j'})\big) = 0$$

.

Case 2: $\{i, j, i', j'\}$ has 3 distinct elements, $i = j$

$$\mathrm{Cov}\big(z_i z_j \phi_k(x_i)\phi_k(x_j), z_{i'} z_{j'} \phi_{k'}(x_{i'})\phi_{k'}(x_{j'})\big) = 0$$

.

Case 3: $\{i, j, i', j'\}$ has 3 distinct elements, $i = i'$ or $i = j'$

$$\begin{aligned}
\mathrm{Cov}\big(z_i z_j \phi_k(x_i)\phi_k(x_j), z_{i'} z_{j'} \phi_{k'}(x_{i'})\phi_{k'}(x_{j'})\big) &= \mathrm{Cov}\big(z_1 z_2 \phi_k(x_1)\phi_k(x_2), z_1 z_3 \phi_{k'}(x_1)\phi_{k'}(x_3)\big) \\
&= \mathbb{E}\big(z_1^2 z_2 z_3 \phi_k(x_1)\phi_{k'}(x_1)\phi_k(x_2)\phi_{k'}(x_1)\big) - \theta_k^2\theta_{k'}^2 \\
&= \mathbb{E}\Big(\big(\varepsilon_1^2 + f^2(x_1)\big)\phi_k(x_1)\phi_{k'}(x_1)\Big)\theta_k\theta_{k'} - \theta_k^2\theta_{k'}^2 \\
&\qquad \text{(independence properties)} \\
&= \Big(\mathbf{1}\{k = k'\} + \mathbb{E}\big(f^2(x_1)\phi_k(x_1)\phi_{k'}(x_1)\big)\Big)\theta_k\theta_{k'} - \theta_k^2\theta_{k'}^2 \\
&\le \Big(\mathbf{1}\{k = k'\} + \sqrt{\mathbb{E}\big(f^2(x_1)\phi_k^2(x_1)\big)\mathbb{E}\big(f^2(x_1)\phi_{k'}^2(x_1)\big)}\Big)\theta_k\theta_{k'} - \theta_k^2\theta_{k'}^2 \\
&= \Big(\mathbf{1}\{k = k'\} + \chi_k\chi_{k'}\Big)\theta_k\theta_{k'} - \theta_k^2\theta_{k'}^2
\end{aligned}$$

Case 4: $\{i, j, i', j'\}$ has 2 distinct elements, $i = j$

$$\mathrm{Cov}\big(z_i z_j \phi_k(x_i)\phi_k(x_j), z_{i'} z_{j'} \phi_{k'}(x_{i'})\phi_{k'}(x_{j'})\big) = 0$$

Case 5: $\{i, j, i', j'\}$ has 2 distinct elements, $i = i'$ or $i = j'$.

$$\begin{aligned}
\mathrm{Cov}\big(z_i z_j \phi_k(x_i)\phi_k(x_j), z_{i'} z_{j'} \phi_{k'}(x_{i'})\phi_{k'}(x_{j'})\big) &= \mathrm{Cov}\big(z_1 z_2 \phi_k(x_1)\phi_k(x_2), z_1 z_2 \phi_{k'}(x_1)\phi_{k'}(x_2)\big) \\
&= \mathbb{E}(z_1^2 z_2^2 \phi_k(x_1)\phi_k(x_2)\phi_{k'}(x_1)\phi_{k'}(x_2)) - \theta_k^2\theta_{k'}^2 \\
&= \mathbb{E}(z_1^2 \phi_k(x_1)\phi_{k'}(x_1))^2 - \theta_k^2\theta_{k'}^2 \\
&\le (\mathbf{1}\{k = k'\} + \chi_k\chi_{k'})^2 - \theta_k^2\theta_{k'}^2
\end{aligned}$$

Case 6: $\{i, j, i', j'\}$ has 1 distinct element.

$$\mathrm{Cov}\big(z_i z_j \phi_k(x_i)\phi_k(x_j), z_{i'} z_{j'} \phi_{k'}(x_{i'})\phi_{k'}(x_{j'})\big) \le \mathbb{E}\big(z_1^4 \phi_k^2(x_1)\phi_{k'}^2(x_1)\big)$$

Putting the cases together, we obtain the following upper bound on $\mathrm{Cov}(\widetilde{y}_k, \widetilde{y}_{k'})$:

$$\mathrm{Cov}(\widetilde{y}_k, \widetilde{y}_{k'}) \le \frac{1}{n}\left[(\mathbf{1}\{k = k'\} + \chi_k\chi_{k'})\theta_k\theta_{k'}\right] + \frac{1}{n^2}(\mathbf{1}\{k = k'\} + \chi_k\chi_{k'})^2 + \frac{1}{n^3}\mathbb{E}(z_1^4\phi_k^2(x_1)\phi_{k'}^2(x_1))$$

We compute the sum over k, k' of each term in the summand on the right hand side separately.

1st term. Observe that

$$\sum_{k,k':c_k,c_k'\leq C} \chi_k \chi_{k'} \theta_k \theta_{k'} = \left(\sum_{k:c_k\leq C} \chi_k \theta_k\right)^2$$
$$\leq \left(\sum_{k:c_k\leq C} \chi_k^2\right)\left(\sum_{k:c_k\leq C} \theta_k^2\right)$$
$$= N(C)\|\theta\|^2\|P_C\theta\|^2;$$

therefore

$$\sum_{k,k':c_k,c_k'\leq C} \frac{1}{n}\left[\left(\mathbf{1}\left\{k=k'\right\}+\chi_k\chi_{k'}\right)\theta_k\theta_{k'}\right] \leq \frac{1}{n}\left(\|P_C\theta\|^2 + N(C)\|\theta\|^2\|P_C\theta\|^2\right).$$

2nd term.

$$\sum_{k,k':c_k,c_k'\leq C} \frac{1}{n^2}\left(\mathbf{1}\left\{k=k'\right\}+\chi_k\chi_{k'}\right)^2 = \frac{1}{n^2}\left(\sum_{k:c_k\leq C} 1 + \sum_{k,k':c_k,c_k'\leq C} \chi_k^2\chi_{k'}^2 + 2\sum_{k:c_k\leq C}\chi_k^2\right)$$
$$\leq \frac{1}{n^2}\left(N(C) + N(C)^2\|\theta\|^4 + 2N(C)\|\theta\|^2\right)$$

3rd term.

$$\sum_{k,k':c_k,c_{k'}\leq C} \frac{1}{n^3}\mathbb{E}(z_1^4\phi_k^2(x_1)\phi_{k'}^2(x_1)) = \mathbb{E}\left(z_1^4 \sum_{k,k':c_k,c_{k'}\leq C} \phi_k^2(x_1)\phi_{k'}^2(x_1)\right)$$
$$= \mathbb{E}\left(z_1^4\left\{\sum_{k:c_k\leq C}\phi_k^2(x_1)\right\}\right)$$
$$\leq \mu_4 N(C)^2$$

We can therefore write

$$\sum_{k,k':c_k,c_{k'}\leq C} \mathrm{Cov}(\widetilde{y}_k,\widetilde{y}_{k'}) \leq \frac{1}{n}\left(\|P_C\theta\|^2 + N(C)\|\theta\|^2\|P_C\theta\|^2\right) + \frac{1}{n^2}\left(N(C) + N(C)^2\|\theta\|^4 + 2N(C)\|\theta\|^2\right)$$
$$+ \frac{\mu_4 N(C)^2}{n^3}$$

which is the desired result. □

From now on, we will fix $C = n^{2s/(4s+d)}$, and recall that $N(C) \leq C^{d/s}$. Let $\tau(b) = b\sqrt{N(C)/n}$. We will consider the following test:

$$\Gamma(T_C) = \mathbf{1}\left\{T_C \geq N(C)/n + \tau(b)\right\}$$

The following bound on Type I error follows immediately from Chebyshev's inequality.

Lemma 7. *For any $b \geq 1$,*

$$\mathbb{P}_0(T_C \geq N(C)/n + \tau(b)) \leq \frac{1}{b^2}.$$

Similar as before, a bound on Type II error is more subtle, and will require that $\|f\|_{L^2} = \|\theta\|_2$ be sufficiently far from zero.

Lemma 8. *Suppose that for some $b \geq 1$,*

$$\|\theta\|^2 \geq 2(2\pi C)^{-2} + 2\tau(b) \tag{3}$$

Then, there exist universal constants c_1, c_2 and $c_3 > 0$ such that

$$\mathbb{P}_a\left(T_C \leq \frac{N(C)}{n} + \tau(b)\right) \leq \frac{1}{4b^2} + c_1 n^{-d/(4s+d)} + c_2 \min\left\{\frac{N(C)}{n-1}, \frac{n}{N(C)}\right\} + c_3 \frac{N(C)}{n^2} \tag{4}$$

Before proving Lemma 8, we remark that i): the critical radius is the same as in the Gaussian white noise setting, and ii): except for $1/(4b^2)$, each summand on the right hand side of (4) is neglible for sufficiently large n, assuming $4s \neq d$.

Proof of Lemma 8. We derive two important facts from (3). The first is that by Lemma 1,

$$\Delta := \mathbb{E}_a(T_C) - \frac{N(C)}{N} \geq 2\tau(b)$$

and therefore $(\Delta - \tau(b))^2 \geq \frac{\Delta^2}{4}$. The second is that

$$\|P_C\theta\|^2 \geq \|\theta\|^2 - (2\pi C)^{-2} \geq \frac{\|\theta\|^2}{2}.$$

We now proceed to use Chebyshev's inequality, obtaining

$$\begin{aligned}
\mathbb{P}_a\left(T_C \leq \frac{N(C)}{n} + \tau(b)\right) &= \mathbb{P}\left(T_C - E_a(T_C) \leq \frac{N(C)}{n} + \tau(b) - E_a(T_C)\right) \\
&\leq \frac{\mathrm{Var}_a(T_C)}{(E_a(T_C) - N(C)/n - \tau(b))^2} \\
&\leq 4\frac{\mathrm{Var}_a(T_C)}{\Delta^2}
\end{aligned}$$

There are six terms in the summand on the right hand side of (2), which jointly upper bound $\mathrm{Var}_a(T_C)$ We bound the ratio of each over Δ^2 in turn.

Term 1: Note that $\Delta^2 \geq (n-1)/n\|P_C\theta\|^2$. Therefore,

$$\begin{aligned}
\frac{\|P_C\theta\|^2}{n\Delta^2} &\leq \frac{1}{(n-1)\|P_C\theta\|^2} \\
&\leq \frac{2}{(n-1)\|\theta\|^2} \\
&\leq 4\pi^2\frac{C^2}{n} \\
&\leq 4\pi^2 n^{-d/(4s+d)}.
\end{aligned}$$

Term 2: Note that $\Delta^2 \geq \max\left\{N(C)\|\theta\|^2/n, (n-1)/n\|P_C\theta\|^2\right\}^2$. Therefore,

$$\begin{aligned}
\frac{N(C)\|\theta\|^2\|P_C\theta\|^2}{n\Delta^2} &\leq \min\left\{\frac{N(C)\|\theta\|^2}{(n-1)\|P_C\theta\|^2}, \frac{n\|P_C\theta\|^2}{N(C)\|\theta\|^2}\right\} \\
&\leq \min\left\{4\frac{N(C)}{(n-1)}, \frac{n}{N(C)}\right\}
\end{aligned}$$

Term 3:

$$\frac{N(C)}{n^2\Delta^2} \leq \frac{1}{4b^2}$$

Term 4: By similar analysis to term 2, we obtain

$$\frac{N(C)^2\|\theta\|^4}{n^2\Delta^2} \leq \min\left\{4\frac{N(C)}{(n-1)}, \frac{n}{N(C)}\right\}^2$$

Term 5: As $\|\Delta\| \geq \|P_C\theta\|^2 \geq \|\theta\|^2/2$, we obtain

$$\frac{N(C)\|\theta\|^2}{n^2\Delta^2} \leq 2\frac{N(C)}{n^2}$$

Term 6: By similar analysis to term 1, we obtain

$$\frac{\mu_4 N(C)^2}{n^3\Delta^2} \leq \mu_4 n^{-d/(4s+d)}.$$

Combining terms, we have that

$$\mathbb{P}_a\left(T_C \leq \frac{N(C)}{n} + \tau(b)\right) \leq 4\pi^2 n^{-d/(4s+d)} + \min\left\{4\frac{N(C)}{(n-1)}, \frac{n}{N(C)}\right\} + \frac{1}{4b^2} + \min\left\{4\frac{N(C)}{(n-1)}, \frac{n}{N(C)}\right\}^2 + 2\frac{N(C)}{n^2} + \mu_4 n^{-d/(4s+d)}.$$

□

3 Testing with Eigenvectors in the Nonparametric Regression Model

Consider the same setup as in the previous section. We now use a modified test statistic. Let $\eta : \mathcal{D}\times\mathcal{D} \to \mathbb{R}$ be a Mercel kernel with the expansion

$$\eta(x, y) := \sum_{k\in\mathcal{Z}^d} c_k^2 \phi_k(x)\phi_k(y)$$

with associated operators $T = T_{\eta,q} : L^2(q) \to L^2(q)$ (for q a distribution over $\mathcal{D}$)

$$Tf(x) := \int_{\mathcal{D}} \eta(x, y)f(y)dq(y)$$

Denote $T_n := T_{\eta,P_n}$, and let $\{(\widehat{\phi}_j, \lambda_j)\}_{j=1}^n$ be the eigenvector/eigenvalue pairs of T_n, so that

$$T_n\widehat{\phi}_j = \lambda_j\widehat{\phi}_j, \quad \|\widehat{\phi}_j\|_{L^2(P_n)} = 1.$$

Our test statistic will be

$$T_C = \sum_{j:\sqrt{\lambda_k}\leq C} \widehat{y}_k^2$$

where $\widehat{y}_k = \langle z, \widehat{\phi}_k\rangle_{L^2(P_n)}$.

3.1 Supplemental Theory

Let $\widehat{N}(C) = \sharp j : \sqrt{\lambda_j} \le C$, and $\widehat{\theta} \in L^2(P_n)$ be the sequence with elements $\widehat{\theta}_j = \langle f, \widehat{\phi}_j \rangle_{L^2(P_n)}$.

Lemma 9. *Under* $\mathbf{H}_0$, $\mathbb{E}_0(T_C) = \mathbb{E}(\widehat{N}(C))/n$. *Under* $\mathbf{H}_a$,

$$\mathbb{E}_a(T_C) = \frac{\mathbb{E}(\widehat{N}(C))}{n} + \mathbb{E}\left(||P_C\widehat{\theta}||_2^2\right) \tag{5}$$

$$\ge \frac{\mathbb{E}(\widehat{N}(C))}{n} + \|\theta\|_2^2 - \frac{1}{C^2}\left(\frac{n-1}{n} + \frac{\mathbb{E}(f^2(x)\eta(x,x))}{n}\right) \tag{6}$$

Proof. By linearity,

$$\mathbb{E}(T_C) = \sum_{j:\sqrt{\lambda_j}<C} \mathbb{E}(\widehat{y}_k^2),$$

so that it is sufficient to compute $\mathbb{E}(\widehat{y}_k^2)$. We have

$$\mathbb{E}(\widehat{y}_j^2) = \mathbb{E}\left(\widehat{\theta}_j^2\right) + \mathbb{E}\left(\langle \epsilon, \widehat{\phi}_j\rangle_{L^2(P_n)}^2\right) + 2\mathbb{E}\left(\langle \epsilon, \widehat{\phi}_j\rangle_{L^2(P_n)}\widehat{\theta}_j^2\right).$$

By the law of iterated expectation, the third summand is 0. The second term can be computed as

$$\begin{aligned}
\mathbb{E}\left(\langle \epsilon, \widehat{\phi}_j\rangle_{L^2(P_n)}^2\right) &= \frac{1}{n^2}\sum_{i=1,j=1}^{n} \mathbb{E}\big(\varepsilon_i\varepsilon_j\widehat{\phi}_j\widehat{\phi}_i\big) \\
&\ge \frac{1}{n^2}\mathbb{E}\left(\sum_{i=1}^{n}\widehat{\phi}_j^2\right) \\
&= \frac{1}{n},
\end{aligned}$$

and summing over $\left\{j : \sqrt{\lambda_j} \le C\right\}$, we obtain the representation (5). We can then expand

$$\begin{aligned}
\mathbb{E}\Bigg(\sum_{j:\sqrt{\lambda_j}>C}\widehat{\theta}_j^2\Bigg) &= \mathbb{E}\Big(\big\|\widehat{\theta}\big\|_2\Big) - \mathbb{E}\Bigg(\sum_{j:\sqrt{\lambda_j}<C}\widehat{\theta}_j^2\Bigg) \\
&= \|\theta\|_2 - \mathbb{E}\Bigg(\sum_{j:\sqrt{\lambda_j}<C}\widehat{\theta}_j^2\Bigg) \\
&\ge \|\theta\|_2 - \frac{1}{C^2}\mathbb{E}\Bigg(\sum_{j:\sqrt{\lambda_j}<C}\widehat{\theta}_j^2\lambda_j\Bigg) \\
&\ge \|\theta\|_2 - \frac{\mathbb{E}\left(\langle T_n f, f\rangle_{L^2(P_n)}\right)}{C^2}
\end{aligned}$$

and further examining $\mathbb{E}\left(\langle T_n f, f\rangle_{L^2(P_n)}\right)$, we obtain

$$\begin{aligned}
\mathbb{E}\left(\langle T_n f, f\rangle_{L^2(P_n)}\right) &= \frac{1}{n}\mathbb{E}(f^2(x)\eta(x,x)) + \frac{(n-1)}{n}\int_{\mathcal{D}}\int_{\mathcal{D}} f(x)f(y)\eta(x,y)dP(x)dP(y) \\
&= \frac{1}{n}\mathbb{E}(f^2(x)\eta(x,x)) + \frac{(n-1)}{n}\sum_{k\in\mathbb{Z}^d} c_k^2\theta_k^2 \\
&\le \frac{1}{n}\mathbb{E}(f^2(x)\eta(x,x)) + \frac{(n-1)}{n},
\end{aligned}$$

and therefore (6). $\square$