Notes for Week 3/6/20 - 3/12/20

Alden Green

March 12, 2020

Suppose we observe independent design points $X = x_1, \ldots, x_n \sim P$ i.i.d – where we assume P has density p which is supported on $[0,1]^d$ – and responses

$$y_i = f(x_i) + \varepsilon_i, \varepsilon_i \overset{i.i.d}{\sim} N(0,1). \tag{1}$$

Our goal is to test

$$\mathbf{H}_0 : f = 0, \quad \text{vs.} \quad \mathbf{H}_a : f \neq 0.$$

We will use as our test statistic the empirical norm of a Laplacian smoothing estimator. Let $G = (X, E)$ be a graph formed over the design points X, the Laplacian smoothing estimator $\widehat{\theta}_{\mathrm{LS}}(G) \in \mathbb{R}^n$ is defined as

$$\widehat{\theta}_{\mathrm{LS}}(G) = \operatorname*{argmin}_{\theta \in \mathbb{R}^n} \sum_{i=1}^{n} (y_i - \theta_i)^2 + \rho \theta^T L_G \theta \tag{2}$$

i.e. $\widehat{\theta}_{\mathrm{LS}}(G) = (I + \rho L_G)^{-1} y$; here $\rho > 0$ is a tuning parameter controlling how much shrinkage the estimator performs. Then our Laplacian smoothing test statistic will simply be

$$T_{\mathrm{LS}}(G) := \|\widehat{\theta}_{\mathrm{LS}}(G)\|_n^2. \tag{3}$$

with the corresponding test $\phi_{\mathrm{LS}}(G) := \mathbf{1}\{T_{\mathrm{LS}}(G) \geq \tau(b)\}$, where $b > 1$ is a user specified hyperparameter which controls the level of Type I and Type II error the user is willing to tolerate, and τ is a function of b (and also, implicitly, of G) to be specified later.

Let $G_{n,r}$ be the random geometric graph of radius r, i.e $G_{n,r} = (X, E_{n,r})$ where $E_{n,r} \subseteq X \times X$ contains the edge $e(x_i, x_j) \in E_{n,r}$ if and only if $\mathbf{1}(\|x_i - x_j\|_2 \leq r)$. When $f \in H^1(\mathcal{X}; L)$ for $d = 1, 2$ or 3, and the density p satisfies typical regularity conditions, the test $\phi_{\mathrm{LS}}(G_{n,r})$ achieves minimax optimal testing rates.

Theorem 1. *Suppose we observe samples $(x_i, y_i)_{i=1}^n$ according to the model (1). Let $L > 0$ and $b \geq 1$ be fixed constants, and $d = 1, 2$ or 3. Suppose that P is an absolutely continuous probability measure over $\mathcal{X} = [0,1]^d$ with density function $p(x)$ bounded away from zero and infinity,*

$$0 < p_{\min} < p(x) < p_{\max} < \infty, \quad \textit{for all } x \in \mathcal{X}.$$

and the test $\phi_{\mathrm{LS}}(G_{n,r})$ is performed with parameter choices

$$c\frac{(\log n)^{p_d}}{n^{1/d}} \leq r(n) \leq n^{-4/((4+d)(2+d))}, \quad \rho = \lambda_\kappa^{-1}, \quad \kappa = n^{2d/(4+d)}$$

$$\tau(b) = \frac{1}{n}\sum_{k=1}^{n}\Big(\frac{1}{1+\rho\lambda_k}\Big)^2 + \frac{4b}{n}\sqrt{\sum_{k=1}^{n}\Big(\frac{1}{1+\rho\lambda_k}\Big)^4}$$

for c a constant that depends only on $\mathcal{X}, p_{\min}$ and $p_{\max}$. Then the following statements holds for every n sufficiently large: there exists constants c_1, c_2 which do not depend on n, b or R such that for every $\epsilon \geq 0$ satisfying

$$\epsilon^2 \geq c_1^2 \cdot b^2 \cdot L^2 \cdot n^{-4/(4+d)} \tag{4}$$

the worst-case risk is upper bounded

$$\mathcal{R}_\epsilon(\phi_{\mathrm{spec}}(G_{n,r}); H^1(\mathcal{X}; L)) \leq \frac{c_2}{b}. \tag{5}$$

1 Fixed Graph Analysis

As usual, our analysis will proceed by showing that for any function f, there exists a set E_f satisfying $\mathbb{P}(E_f) \geq 1 - \|S_\rho(\beta)\|_n^2$, such that conditional on $X = x$ for any $x \in E_f$ our test has non-trivial power. This latter step amounts to analyzing the behavior of our test in the fixed graph setting. Formally, suppose we observe fixed design points $x_1, \ldots, x_n$, and random responses

$$y_i = \beta_i + \epsilon_i, \quad \epsilon_i \overset{\text{i.i.d}}{\sim} N(0,1) \tag{6}$$

In the following Lemma, we bound the Type I and Type II error of $\phi_{\mathrm{LS}}(G)$. For convenience, we denote $S_\rho = (I + \rho L)^{-1}$.

Lemma 1. *Fix $\rho > 0$. Suppose we observe data according to model (6), and perform the test $\phi_{\mathrm{LS}}(G)$ with threshold*

$$\tau(b) = \frac{1}{n}\sum_{k=1}^{n}\Bigl(\frac{1}{1+\rho\lambda_k}\Bigr)^2 + \frac{b}{n}\sqrt{\sum_{k=1}^{n}\Bigl(\frac{1}{1+\rho\lambda_k}\Bigr)^4}.$$

1. ***Type I error:*** *Under the null hypothesis $\beta = \beta_0 = 0$, the Type I error of $\phi_{\mathrm{LS}}(G)$ is upper bounded,*

$$\mathbb{E}_{\beta_0}\Bigl[\phi_{\mathrm{LS}}(G)\Bigr] \leq \frac{1}{b^2}$$

2. ***Type II error:*** *For any $b \geq 1$ and β such that*

$$\|S_\rho(\beta)\|_n^2 \geq \frac{2b}{n}\sqrt{\sum_{k=1}^{n}\Bigl(\frac{1}{1+\rho\lambda_k}\Bigr)^4} \tag{7}$$

the Type II error of $\phi_{\mathrm{LS}}(G)$ is upper bounded,

$$\mathbb{E}_\beta\Bigl[1 - \phi_{\mathrm{LS}}(G)\Bigr] \leq \frac{8}{b\sqrt{n}}\Biggl(\sum_{k=1}^{n}\Bigl(\frac{1}{1+\rho\lambda_k}\Bigr)^4\Biggr)^{-1/2} + \frac{2}{b^2}. \tag{8}$$

In particular, there exist universal constants c_1 and c_2 such that if

$$\frac{c_1}{n}\sum_{i=1}^{n}\beta_i^2 \geq c_2\frac{\rho\beta^T L\beta}{n} + \frac{4b}{n}\sqrt{\sum_{k=1}^{n}\Bigl(\frac{1}{1+\rho\lambda_k}\Bigr)^4}, \tag{9}$$

then (7) and thus (8) follow.

2 Analysis

2.1 Fixed Graph Testing

Decomposing $y = \beta + \varepsilon$, the Laplacian smoothing test statistic may be written as

$$T_{\mathrm{LS}}(G) = \frac{1}{n}(\beta+\varepsilon)^T S_\rho^2 (\beta+\varepsilon) = \frac{1}{n}\Big(\beta^T S_\rho^2 \beta + 2\beta^T S_\rho^2 \varepsilon + \varepsilon^T S_\rho^2 \varepsilon\Big)$$

Writing the spectral decomposition of the Laplacian $L = V\Lambda V^T$ – where V is orthonormal and Λ diagonal– and invoking the rotational invariance of the Gaussian distribution, we conclude that

$$\varepsilon^T S_\rho^2 \varepsilon \overset{d}{=} \sum_{k=1}^{n} \Big(\frac{1}{1+\rho\lambda_k}\Big)^2 \varepsilon_k^2 \tag{10}$$

This equality (in distribution) will be useful for computing both the mean and variance of $T_{\mathrm{LS}}(G)$.

Mean of $T_{\mathrm{LS}}(G)$. Noting that ε is mean-zero, we have

$$\begin{aligned} \mathbb{E}\big[T_{\mathrm{LS}}(G)\big] &= \frac{1}{n}\Big(\beta^T S_\rho^2 \beta + \mathbb{E}\big[\varepsilon^T S_\rho^2 \varepsilon\big]\Big) \\ &= \|S_\rho(\beta)\|_n^2 + \sum_{k=1}^{n} \frac{1}{(1+\rho\lambda_k)^2} \end{aligned} \tag{11}$$

Variance of $T_{\mathrm{LS}}(G)$. Note that since $\beta^T S_\rho^2 \varepsilon$ is symmetric about zero,

$$\mathrm{Cov}\big[\beta^T S_\rho^2 \varepsilon, \varepsilon^T S_\rho^2 \varepsilon\big] = 0,$$

and therefore

$$\begin{aligned} \mathrm{Var}\big[T_{\mathrm{LS}}(G)\big] &= \frac{1}{n^2}\Big(4\mathrm{Var}\big[\beta^T S_\rho^2 \varepsilon\big] + \mathrm{Var}\big[\varepsilon^T S_\rho^2 \varepsilon\big]\Big) \\ &= \frac{1}{n^2}\Big(4\beta^T S_\rho^4 \beta + \mathrm{Var}\big[\varepsilon^T S_\rho^2 \varepsilon\big]\Big) \\ &\leq \frac{1}{n^2}\Big(4\beta^T S_\rho^2 \beta + \mathrm{Var}\big[\varepsilon^T S_\rho^2 \varepsilon\big]\Big) \\ &= \frac{1}{n^2}\bigg(4\beta^T S_\rho^2 \beta + \sum_{k=1}^{n}\Big(\frac{1}{1+\rho\lambda_k}\Big)^4 \mathrm{Var}\big[\varepsilon_k^2\big]\bigg) \\ &= \frac{1}{n^2}\bigg(4\beta^T S_\rho^2 \beta + 2\sum_{k=1}^{n}\Big(\frac{1}{1+\rho\lambda_k}\Big)^4\bigg). \end{aligned}$$

where the inequality in previous display follows from $\lambda_{\min}(S_\rho) \geq 1$.

For convenience, we introduce the notation

$$t(b) := \frac{b}{n}\sqrt{\sum_{k=1}^{n}\Big(\frac{1}{1+\rho\lambda_k}\Big)^4}.$$

Type I error. Using Chebyshev's inequality, we obtain

$$\begin{aligned}
\mathbb{P}_{\beta=0}\Big(T_{\text{LS}}(G) \geq \tau(b)\Big) &= \mathbb{P}_{\beta=0}\Big(T_{\text{LS}}(G) - \mathbb{E}\big[T_{\text{LS}}(G)\big] \geq t(b)\Big) \\
&\leq \mathbb{P}_{\beta=0}\Big(\big|T_{\text{LS}}(G) - \mathbb{E}\big[T_{\text{LS}}(G)\big]\big| \geq t(b)\Big) \\
&\leq \frac{\text{Var}_{\beta=0}\big[T_{\text{LS}}(G)\big]}{\big[t(b)\big]^2} \\
&\leq \frac{2}{b^2}.
\end{aligned}$$

Type II error. We note that (11) along with assumption (7) implies that

$$\mathbb{E}[T_{\text{LS}}(G)] - \tau(b) = \|S_\rho(\beta)\|_n^2 - t(b) \geq t(b).$$

Again applying Chebyshev's inequality, we find

$$\begin{aligned}
\mathbb{P}_{\beta}\Big(T_{\text{LS}}(G) < \tau(b)\Big) &= \mathbb{P}_{\beta}\Big(T_{\text{LS}}(G) - \mathbb{E}_\beta\big[T_{\text{LS}}(G)\big] < t(b) - \|S_\rho(\beta)\|_n^2\Big) \\
&\leq \mathbb{P}_{\beta}\Big(\big|T_{\text{LS}}(G) - \mathbb{E}_\beta\big[T_{\text{LS}}(G)\big]\big| > \frac{1}{2}\|S_\rho(\beta)\|_n^2\Big) \\
&\leq 4\frac{\text{Var}_\beta(T_{\text{LS}}(G))}{\|S_\rho(\beta)\|_n^4} \\
&\leq \frac{16\|S_\rho(\beta)\|_n^2/n + 8\sum_{k=1}^n \Big(\frac{1}{1+\rho\lambda_k}\Big)^4/n^2}{\|S_\rho(\beta)\|_n^4}
\end{aligned}$$

Since $\|S_\rho(\beta)\|_n^2 \geq 2t(b)$,

$$\frac{1}{n\|S_\rho(\beta)\|_n^2} \leq \frac{1}{2nt(b)}, \quad \frac{1}{n^2\|S_\rho(\beta)\|_n^4}\sum_{k=1}^n\Big(\frac{1}{1+\rho\lambda_k}\Big)^4 \leq \frac{1}{4b^2}.$$

and (8) follows.

(9) implies (7). Note that S_ρ lets constant signals pass through unfiltered, i.e. decomposing $\beta = a_1\widetilde{1} + a_2\beta_\perp$ where $\widetilde{1} = n^{-1/2}(1,\ldots,1) \in \mathbb{R}^n$, we have

$$\beta^T S_\rho^2 \beta = a_1^2 + a_2^2 \beta_\perp^T S_\rho^2 \beta_\perp \tag{12}$$

We use the following crude but sufficient lower bound on the quadratic form,

$$\beta_\perp^T S_\rho^2 \beta_\perp = (\beta_\perp + (S_\rho - I)\beta_\perp)^T(\beta_\perp + (S_\rho - I)\beta_\perp) \geq \Big(1 - \frac{1}{\sqrt{2}}\Big)\beta_\perp^T\beta_\perp - (4\sqrt{2}-1)\Big(\beta^T(S_\rho - I)^T(S_\rho - I)\beta\Big).$$

Then the derivations in (Sadhanala) show

$$\beta^T(S_\rho - I)^T(S_\rho - I)\beta \leq \frac{\rho}{4}\beta^T L\beta$$

and plugging back in to (12), we conclude that

$$\beta^T S_\rho^2 \beta \geq c_1\beta^T\beta - c_2\rho\beta^T L\beta$$

for $c_1 = 1 - 1/\sqrt{2}$ and $c_2 = \sqrt{2} - 1/4$, which suffices to prove the claim.