Lecture 23: April 16

Lecturer: Siva Balakrishnan *Scribe: Tian Tong*

23.1 Recap of non-parametric least squares

1. **Wainwright's book Theorem 13.5:** Suppose that the shifted function class $\mathcal{F}^* = \{f - f^* : f \in \mathcal{F}\}$ is star-shaped, and δ_n satisfies the critical inequality

$$\frac{\mathcal{G}_n(\delta_n; \mathcal{F}^*)}{\delta_n} \leq \frac{\delta_n}{2\sigma},$$

then for any $t \geq \delta_n$, the nonparametric least-squares estimate $\widehat{f}$ satisfies the bound

$$\mathbb{P}\left(\|\widehat{f} - f^*\|_n^2 \geq 16t\delta_n\right) \leq e^{-\frac{n\delta_n}{2\sigma^2}}.$$

The mean squared error is upper bounded by

$$\mathbb{E}\|\widehat{f} - f^*\|_n^2 \lesssim \delta_n^2.$$

2. Adaptivity: the rates can vary as a function of f^*, since the local Gaussian width $\mathcal{G}_n(\delta_n; \mathcal{F}^*)$ can vary as the true function f^* varies. Intuitively, if f^* is "simple", $\mathcal{G}_n(\delta_n; \mathcal{F}^*)$ can be small, and one would get a faster rate automatically.

3. **Oracle inequality, Wainwright's book Theorem 13.13:** Suppose that $\partial\mathcal{F} = \{f_1 - f_2 : f_1, f_2 \in \mathcal{F}\}$ is star-shaped, and δ_n satisfies the critical inequality

$$\frac{\mathcal{G}_n(\delta_n; \partial\mathcal{F})}{\delta_n} \leq \frac{\delta_n}{2\sigma},$$

then for any $t \geq \delta_n$, the the nonparametric least-squares estimate $\widehat{f}$ satisfies the bound

$$\|\widehat{f} - f^*\|_n^2 \leq \inf_{\gamma \in (0,1)} \left\{\frac{1+\gamma}{1-\gamma}\|f - f^*\|_n^2 + \frac{c_0}{\gamma(1-\gamma)} t\delta_n\right\} \quad \text{for all } f \in \mathcal{F}$$

with probability greater than $1 - c_1 e^{-c_2 \frac{n\delta_n}{\sigma^2}}$.

23.2 Minimax hypothesis testing

Given samples $X_1, \ldots, X_n \sim P$, consider the hypothesis testing:

$$\begin{aligned} \text{null } H_0 &: P \in \mathcal{P}_0; \\ \text{alternative } H_1 &: P \in \mathcal{P}_1(\epsilon) = \mathcal{P}_1 \cap \{P : \rho(P, \mathcal{P}_0) \geq \epsilon\}. \end{aligned}$$

Define the test function as $\phi_T(X_1, \ldots, X_n) \in \{0, 1\}$, where $\phi_T = 1$ means to reject the null and $\phi_T = 0$ means to retain the null. Define the risk as

$$R_\epsilon(T) = \underbrace{\sup_{P \in \mathcal{P}_0} \mathbb{E}[\phi_T]}_{\text{Type I error}} + \underbrace{\sup_{P \in \mathcal{P}_1(\epsilon)} \mathbb{E}[1 - \phi_T]}_{\text{Type II error}}.$$

The minimax test is $\widehat{T} = \arg\min_T R_\epsilon(T)$.

We will define the critical radius to be some ϵ^* such that

$$\begin{aligned} \inf_T R_{C_1 \epsilon^*}(T) &\leq 1/6, \\ \inf_T R_{C_2 \epsilon^*}(T) &\geq 1/3. \end{aligned}$$

23.3 Gaussian mean testing

Consider $y = \theta^* + \eta$, where $\eta \sim N(0, \sigma^2 I_d)$. The two hypotheses are

$$\begin{aligned} &H_0 : \theta^* = 0, \quad \mathcal{P}_0 = \{N(0, \sigma^2 I_d)\}; \\ &H_1 : \|\theta^*\|_2 \geq \epsilon, \quad \mathcal{P}_1(\epsilon) = \{N(\theta^*, \sigma^2 I_d) : \|\theta^*\|_2 \geq \epsilon\}. \end{aligned}$$

In the general case, we observe n samples $y_1, \ldots, y_n$ and the critical radius scales as

$$\epsilon^* \asymp \sigma \frac{d^{1/4}}{\sqrt{n}}.$$

As a comparison, the minimax estimation rate is

$$\epsilon^* \asymp \sigma \sqrt{\frac{d}{n}}.$$

In general, it is often the case that minimax testing is a statistically easier problem than its estimation counterpart.

This problem is also related to two problems that you might come across in the minimax hypothesis testing literature.

- Our Gaussian mean testing problem is a special case of the problem of *signal detection*. Broadly, this is the problem of detecting the presence of a signal (generally, testing $\theta^* = 0$ against some class of alternatives).
- One can also view this as a multiple testing problem. Concretely, for each hypothesis $\{1, \ldots, d\}$ we can convert the y_i to a p-value (just use $\Phi(|y_i|)$, where $\Phi(\cdot)$ is the standard Gaussian CDF). Now, our mean testing problem is equivalent to the problem of *global null testing*. In that literature, one might study the minimax power against various types of alternatives (sparse, dense, ...) and our setup roughly corresponds to testing against a possibly dense alternative.

In the remainder, we suppose that the sample size $n = 1$. One can replace σ by $\sigma/\sqrt{n}$ in case of n samples.

23.3.1 Upper bound

Design the test statistics $T = \sum_{i=1}^d y_i^2$, and the test function $\phi_T = 1\{T \geq \mathbb{E}_0 T + C\sqrt{\mathrm{Var}_0(T)}\}$.

Lemma 23.1 *If $\mathbb{E}_1 T - \mathbb{E}_0 T \geq C(\sqrt{Var_0(T)} + \sqrt{Var_1(T)})$, then $R(T) \leq 2/C^2$.*

Proof: Type I error is bounded by Chebshev's inequality as

$$\mathbb{P}_0(T \geq \mathbb{E}_0 T + C\sqrt{\mathrm{Var}_0(T)}) \leq 1/C^2.$$

Type II error is bounded similarly as

$$\begin{aligned}\mathbb{P}_1(T \leq \mathbb{E}_0 T + C\sqrt{\mathrm{Var}_0(T)}) &= \mathbb{P}_1(T - \mathbb{E}_1 T \leq \mathbb{E}_0 T - \mathbb{E}_1 T + C\sqrt{\mathrm{Var}_0(T)}) \\ &\leq \mathbb{P}_1(T - \mathbb{E}_1 T \leq -C\sqrt{\mathrm{Var}_1(T)}) \leq 1/C^2.\end{aligned}$$

Combine them to obtain

$$R(T) = \sup_{P \in \mathcal{P}_0} \mathbb{E}[\phi_T] + \sup_{P \in \mathcal{P}_1(\epsilon)} \mathbb{E}[1 - \phi_T] \leq 2/C^2.$$

■

The means and variances are calculated as following:

- $\mathbb{E}_0 T = \sigma^2 d$.
- $\mathbb{E}_1 T = \|\theta^*\|_2^2 + \sigma^2 d$.
- $\mathrm{Var}_0(T) = \sum_{i=1}^d \mathrm{Var}_0(y_i^2) = 2\sigma^4 d$, where

$$\mathrm{Var}_0(y_i^2) = \mathbb{E}[\eta_i^4] - (\mathbb{E}[\eta_i^2])^2 = 3\sigma^4 - \sigma^4 = 2\sigma^4.$$

- $\mathrm{Var}_1(T) = \sum_{i=1}^d \mathrm{Var}_1(y_i^2) = 4\sigma^2\|\theta^*\|_2^2 + 2\sigma^4 d$, where

$$\begin{aligned}\mathrm{Var}_1(y_i^2) &= \mathrm{Var}(\theta_i^{*2} + 2\theta_i^*\eta_i + \eta_i^2) = \mathrm{Var}(2\theta_i^*\eta_i + \eta_i^2)\\ &= \mathbb{E}[(2\theta_i^*\eta_i + \eta_i^2)^2] - (\mathbb{E}[2\theta_i^*\eta_i + \eta_i^2])^2\\ &= \mathbb{E}[4\theta_i^{*2}\eta_i^2 + \eta_i^4] - \sigma^4\\ &= 4\sigma^2\theta_i^{*2} + 2\sigma^4.\end{aligned}$$

The condition $\mathbb{E}_1 T - \mathbb{E}_0 T \geq C(\sqrt{\mathrm{Var}_0(T)} + \sqrt{\mathrm{Var}_1(T)})$ is $\|\theta^*\|_2^2 \geq C(\sqrt{2\sigma^4 d} + \sqrt{2\sigma^4 d + 4\sigma^2\|\theta^*\|_2^2})$. Therefore as long as $\|\theta^*\|_2 \gtrsim \sigma d^{1/4}$, the test will have a low risk $R(T) \leq 1/6$. This in turn proves our upper bound on the critical radius.

23.3.2 Lower bound

If we are testing a simple null against a simple alternative we know the optimal test, and we understand its Type I and Type II errors (by the Neyman-Pearson Lemma). To use this however one needs to transform the hypotheses into simple null v.s. simple alternative. Design the Bayesian counterpart as

$$\begin{aligned}H_0 &: \theta^* = 0;\\ H_1 &: \theta^* \sim \epsilon\mathcal{S}_{d-1},\end{aligned}$$

where $\mathcal{S}_{d-1}$ denotes the d-dimensional unit sphere, and the parameter $\epsilon \asymp \sigma d^{1/4}$.

Remark: It is crucial to introduce the spherical uniform sampling as the alternative hypothesis. It does not work if one uses a one-point alternative hypothesis

$$\begin{aligned}H_0 &: \theta^* = 0;\\ H_1 &: \theta^* = \epsilon(1/\sqrt{d}, \ldots, 1/\sqrt{d}).\end{aligned}$$

In this case, the minimax test function is $\phi_T = 1\{\sum_{i=1}^d y_i \geq \epsilon\sqrt{d}/2\}$, and the minimax risk is

$$R(T) = 2\mathbb{P}_0\left(\sum_{i=1}^d y_i \geq \frac{\epsilon}{2}\sqrt{d}\right) = 2\Phi\left(\frac{\epsilon}{2\sigma}\right),$$

where $\Phi(\cdot)$ is the tail distribution function of the standard Gaussian distribution. In order to have a risk $R(T) \asymp 1$, the parameter $\epsilon \asymp \sigma$. The dimension d disappears, and the problem degenerates into a one-dimensional hypothesis testing problem.

To be continued...