MATH888: High-dimensional probability and statistics Fall 2021

Lecture 30 – 15th November, 2021

Sebastien Roch, UW-Madison *Scribe: Zhenmei Shi, Sebastien Roch*

1 Overview

In the last lecture, we finished the proof of Proposition 15.1 in Wainwright's book [1]. In this lecture, we will give a second example in low dimension.

2 From Estimation to Testing: A Second Example

2.1 Reminder

Recall that the minimax risk of the estimation problem is:

$$\mathfrak{M}(\theta(\mathcal{P}); \Phi \circ \rho) = \inf_{\hat{\theta}} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\Phi(\rho(\hat{\theta}, \theta(\mathbb{P}))) \right]. \tag{1}$$

Hypothesis testing problem setting:

1. In the space $\theta(\mathcal{P})$, $\{\theta^1, ..., \theta^M\}$ is a 2δ-separated set.
2. For each θ^j, choose distribution $\mathbb{P}_{\theta^j} \in \mathcal{P}$ such that $\theta(\mathbb{P}_{\theta^j}) = \theta^j$.
3. Generate Z by the following procedure:

   Pick J uniformly at random in the index set $[M] := \{1, ..., M\}$.

   Given $J = j$, sample $Z \sim \mathbb{P}_{\theta^j}$.

Let $\mathbb{Q}$ be the joint distribution of (J, Z).

Assuming this setting, the main result is:

Theorem 1 (From estimation to a testing problem)**.** *For any increasing function Φ and choice of 2δ-separated set, the minimax risk of the estimation problem is lower bounded as*

$$\mathfrak{M}(\theta(\mathcal{P}); \Phi \circ \rho) \geq \Phi(\delta) \inf_{\psi} \mathbb{Q}[\psi(Z) \neq J], \tag{2}$$

where the infinimum ranges over all testing functions from the range of Z to $[M]$.

2.2 Corollary for M=2

Corollary 2 (Special case when M=2)**.**

$$\mathfrak{M} \geq \frac{\Phi(\delta)}{2}(1 - \|\mathbb{P}_{\theta^1} - \mathbb{P}_{\theta^2}\|_{TV}). \tag{3}$$

Proof. Recall that

$$\|\mathbb{P}_{\theta^1} - \mathbb{P}_{\theta^2}\|_{TV} = \sup_A |\mathbb{P}_{\theta^1}(A) - \mathbb{P}_{\theta^2}(A)|.$$

Note that

$$\inf_{\psi} \mathbb{Q}[\psi(Z) \neq J] = 1 - \sup_{\psi} \mathbb{Q}[\psi(Z) = J].$$

Also, by the law of total probability, we have

$$\sup_{\psi} \mathbb{Q}[\psi(Z) = J] = \sup_{\psi} \left\{ \frac{1}{2}\mathbb{P}_{\theta^1}[\psi(Z) = 1] + \frac{1}{2}\mathbb{P}_{\theta^2}[\psi(Z) = 2] \right\}. \tag{4}$$

Defining $A = \{\psi(Z) = 1\}$ we see that

$$\begin{aligned}(4) &= \frac{1}{2}\left(\sup_{\psi}\{\mathbb{P}_{\theta^1}[\psi(Z) = 1] - \mathbb{P}_{\theta^2}[\psi(Z) = 1]\} \right) + \frac{1}{2} \\ &= \frac{1}{2} \sup_A |\mathbb{P}_{\theta^1}(A) - \mathbb{P}_{\theta^2}(A)| + \frac{1}{2}.\end{aligned}$$

Replacing above gives the claim. □

2.3 Example: Uniform Location Family

Here we consider the uniform location family (Example 15.5 in Wainwright's book [1]). The setup is the following:

1. $\mathbb{U}_\theta$ is the uniform distribution over $[\theta, \theta + 1]$.
2. $\mathbb{U}_\theta^n$ is the distribution of n i.i.d. samples from $\mathbb{U}_\theta$.
3. Goal is to lower bound the minimax risk (in the MSE case)

   $$\inf_{\hat{\theta}} \sup_{\theta \in \mathbb{R}} \mathbb{E}_\theta[(\hat{\theta} - \theta)^2]),$$

   where $\hat{\theta}$ is a function of $(X_1, \ldots, X_n) \sim \mathbb{U}_\theta^n$.
4. First we try an estimator based on the sample mean:

   $$\bar{X}_n = \frac{1}{n}\Sigma_{i=1}^n X_i,$$

   Defining

   $$\hat{\theta}_0 = \bar{X}_n - \frac{1}{2},$$

we have

$$\mathbb{E}\bar{X}_n = \mathbb{E}X_1 = \theta + \frac{1}{2},$$

and

$$\begin{aligned}\mathbb{E}_\theta[(\hat{\theta}_0 - \theta)^2] &= \mathbb{E}_\theta[(\bar{X}_n - \frac{1}{2} - \theta)^2] = \mathbb{E}_\theta[(\bar{X}_n - (\frac{1}{2} + \theta))^2] \\ &= \mathrm{Var}_\theta(\bar{X}_n) = \frac{1}{n^2}\, n\, \mathrm{Var}_\theta(X_1) = \frac{1}{n}\,\mathrm{Var}_0(X_1),\end{aligned}$$

where we used translation invariance in the last equality. Thus, this estimator based on the sample mean estimator risk rate of the order of $\frac{1}{n}$.

5. By Corollary 2,

$$\mathfrak{M} \geq \frac{\Phi(\delta)}{2}(1 - \|\mathbb{U}^n_{\theta^1} - \mathbb{U}^n_{\theta^2}\|_{TV}),$$

where $|\theta^1 - \theta^2| = 2\delta$.

6. Using inequalities we previously discussed,

$$\begin{aligned}\|\mathbb{U}^n_{\theta^1} - \mathbb{U}^n_{\theta^2}\|_{TV} &\leq \mathbb{H}^2(\mathbb{U}^n_{\theta^1}, \mathbb{U}^n_{\theta^2}) &(5)\\ &\leq n\mathbb{H}^2(\mathbb{U}_{\theta^1}, \mathbb{U}_{\theta^2}) \\ &= n\int_{-\infty}^{\infty}\left(\sqrt{f_{\theta^1}(x)} - \sqrt{f_{\theta^1}(x)}\right)^2 dx \\ &= \min(2n, 2n|\theta^1 - \theta^2|), &(6)\end{aligned}$$

where $\mathbb{H}$ is Hellinger distance (equation 15.9 in [1]) and $f_{\theta^i}(x)$ is density function under $\mathbb{U}_{\theta^i}$. Eq. (5) is based on the conjunction with Lemma 15.3 in [1]. Taking $\delta = \frac{1}{8n}$,

$$\begin{aligned}(6) &= 2n \times 2\frac{1}{8n} = \frac{1}{2} \\ \Rightarrow & 1 - \|\mathbb{U}^n_{\theta^1} - \mathbb{U}^n_{\theta^2}\|_{TV} \geq 1 - \frac{1}{\sqrt{2}} \\ \Rightarrow & \mathfrak{M} \geq \frac{1 - \frac{1}{\sqrt{2}}}{128}\frac{1}{n^2}, &(7)\end{aligned}$$

where (7) is based on Corollary 2.

2.4 A Better Estimator

Now consider estimator $R_n := \min\{X_1, \ldots, X_n\}$. We will show that R_n achieves better risk. It can be proved that (the second claim follows from the argument below in fact)

$$\begin{aligned}&\mathbb{E}_\theta R_n > \theta, \\ &R_n \underset{p}{\to} \theta \text{ as } n \to \infty.\end{aligned}$$

W.l.o.g. set $\theta = 0$, then

$$\mathbb{E}[R_n^2] = \int_0^1 \mathbb{P}(R_n^2 \geq t)dt, ,$$

where

$$\begin{aligned}\mathbb{P}(R_n^2 \geq t) &= \mathbb{P}(R_n \geq \sqrt{t}) = \mathbb{P}(X_1, \ldots, X_n \geq \sqrt{t}) \\ &= \Pi_{i=1}^n P(X_i \geq \sqrt{t}) = P(X_1 \geq \sqrt{t})^n = (1-\sqrt{t})^n.\end{aligned}$$

Define $t_j = \frac{j^2}{n^2}, j = 0, \ldots, n$. On the interval $[t_j, t_{j+1})$,

$$(1-\sqrt{t})^n \leq e^{-\sqrt{t}n} \leq e^{-\sqrt{t_j}n} = e^{-j}. \tag{8}$$

Thus,

$$\begin{aligned}(8) &= \int_0^1 (1-\sqrt{t})^n dt \qquad (9)\\ &\leq \sum_{j=0}^{\infty} e^{-j} \left(\frac{(j+1)^2}{n^2} - \frac{j^2}{n^2} \right) \qquad (10)\\ &= \frac{1}{n^2} \sum_{j=0}^{\infty} e^{-j}(2j+1) \lesssim \frac{1}{n^2}. \qquad (11)\end{aligned}$$

We can see that R_n achieves the better risk rate of order $\frac{1}{n^2}$ than the estimator based on the sample mean estimator whose risk rate is of order $\frac{1}{n}$. By (7), we know that R_n is an optimal estimator up to a constant factor.

References

[1] Wainwright, Martin J., *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Cambridge Series in Statistical and Probabilistic Mathematics, 2019.