MATH888: High-dimensional probability and statistics	Fall 2021
Lecture 29 — November 12, 2021	
Sebastien Roch, UW-Madison	*Scribe: Qiaomei Li, Sebastien Roch*

1 Overview

In the last lecture, we explained a connection between estimation and testing that is useful to derive lower bounds on the minimiax risk. In this lecture we will give a formal proof of this technique, which follows Proposition 15.1 of Wainwright's book [1].

2 From estimation to testing: proof of the lower bound

Recall that the minimax risk of the estimation problem is :

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

In order to provide a lower bound on the minimax risk, we show it can be obtained through a related hypothesis testing problem.

2.1 Reminder: setting and statement

1. In the space $\theta(\mathbb{P})$, $\{\theta^1, ..., \theta^M\}$ 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 infimum ranges over all testing functions from the range of Z to $[M]$.

Proof. **Step I:** By Markov's inequality, we have

$$\mathbb{E}_{\mathbb{P}}[\Phi(\rho(\hat{\theta}, \theta(\mathbb{P})))] \geq \Phi(\delta)\mathbb{P}[\Phi(\rho(\hat{\theta}, \theta)) \geq \Phi(\delta)]. \tag{3}$$

Since Φ is increasing, we have

$$\Phi(\delta)\mathbb{P}[\Phi(\rho(\hat{\theta}, \theta)) \geq \Phi(\delta)] \geq \Phi(\delta)\mathbb{P}[\rho(\hat{\theta}, \theta) \geq \Phi(\delta)] \geq \Phi(\delta)] \geq \Phi(\delta)\mathbb{P}[\rho(\hat{\theta}, \theta) \geq \delta]. \tag{4}$$

Hence we arrive at the bound

$$\mathbb{E}_{\mathbb{P}}[\Phi(\rho(\hat{\theta}, \theta(\mathbb{P})))] \geq \Phi(\delta)\mathbb{P}[\rho(\hat{\theta}, \theta) \geq \delta]. \tag{5}$$

Step II: Recall $\mathbb{Q}$ is the joint distribution of (J, Z), and note that we can bound the supremum as follows:

$$\sup_{\mathbb{P}} \mathbb{P}[\rho(\hat{\theta}, \theta(\mathbb{P})) \geq \delta] \geq \frac{1}{M} \sum_{j=1}^{M} \mathbb{P}_{\theta^j}[\rho(\hat{\theta}, \theta^j) \geq \delta] = \mathbb{Q}[\rho(\hat{\theta}, \theta^J) \geq \delta]. \tag{6}$$

Step III: For any choice of estimator $\hat{\theta}$, define the testing function $\psi(Z) := \operatorname{argmin}_{j \in [M]} \rho(\theta^j, \hat{\theta})$. We claim that $\{\rho(\hat{\theta}, \theta^J) < \delta\} \subseteq \{\psi(Z) = J\}$, which easily follows from the triangle inequality. See Figure 15.1 in Wainwright's book for a proof by picture [1].

Step IV: Since $\{\rho(\hat{\theta}, \theta^j) < \delta\}$ implies $\{\psi(Z) = j\}$ by the previous claim (for all j), we have

$$\mathbb{P}_{\theta^j}[\rho(\hat{\theta}, \theta^j) \geq \delta] \geq \mathbb{P}_{\theta^j}[\psi(Z) \neq j]. \tag{7}$$

By taking averages over j, we get

$$\mathbb{Q}[\rho(\hat{\theta}, \theta^J) \geq \delta] = \frac{1}{M} \sum_{j=1}^{M} \mathbb{P}_{\theta^j}[\rho(\hat{\theta}, \theta^j) \geq \delta] \geq \mathbb{Q}[\psi(Z) \neq J]. \tag{8}$$

Combining the result with the lower bound of the supremum we found in Step II,

$$\sup_{\mathbb{P}} \mathbb{E}_{\mathbb{P}}[\Phi(\rho(\hat{\theta}, \theta(\mathbb{P})))] \geq \Phi(\delta)\, \mathbb{Q}[\psi(Z) \neq J]. \tag{9}$$

Step V: Take an infimum over all estimators $\hat{\theta}$ on the l.h.s. and then take an infimum over testing function on the r.h.s. That completes the proof. ☐

2.2 Example: uniform location family

Consider n i.i.d samples from a uniform distribution over a unit length interval starting at an unknown parameter θ. That is, $\mathbb{P}_\theta = \mathcal{U}(\theta, \theta + 1)$, $\theta \in \mathbb{R}$, the positive increasing function is $\Phi(\delta) = \delta^2$, and the semi-metric is $\rho(\hat{\theta}, \theta) = |\hat{\theta} - \theta|$.

Consider the sample mean, which we denote here by $\hat{\theta}_m$, minus $\frac{1}{2}$. It is easy to see that $\hat{\theta}_m - \frac{1}{2}$ is an unbiased estimator of θ. By previous calculations, its mean sqaured error is $\frac{\sigma^2}{n}$. We will show next time that a better risk can be achieved by a different estimator.

References

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