Minimax: Lower Bounds Based on Two Hypotheses

Hao Yan

June 1, 2022

1 Introduction

Given a semi-distance d and an monotone increasing function w such that

$$w : [0, \infty) \to [0, \infty), \quad w(0) = 0,$$

we would like to obtain lower bounds of the form

$$\liminf_{n\to\infty} \inf_{\hat{\theta}_n} \sup_{\theta\in\Theta} \mathbb{E}_\theta \left[w(\psi_n^{-1} d(\hat{\theta}_n, \theta)) \right] \geq c > 0.$$

A general reduction scheme is to consider

$$\begin{aligned}\mathbb{E}_\theta \left[w(\psi_n^{-1} d(\hat{\theta}_n, \theta)) \right] &\geq w(A) P_\theta(\psi_n^{-1} d(\hat{\theta}_n, \theta) \geq A) \\ &= w(A) P_\theta(d(\hat{\theta}_n, \theta) \geq s),\end{aligned}$$

with $s = s_n = A\psi_n$.

Consider $2s$-separated hypotheses, $\{\theta_0, \dots, \theta_M\}$. We call a test any measurable function $\boldsymbol{\psi} : \mathcal{X} \to \{0, \dots, M\}$. Then

$$\inf_{\hat{\theta}_n} \sup_{\theta\in\Theta} P_\theta(d(\hat{\theta}_n, \theta) \geq s) \geq p_{e,M},$$

where

$$p_{e,M} := \inf_{\boldsymbol{\psi}} \max_{0\leq j\leq M} P_j(\boldsymbol{\psi} \neq j).$$

If $p_{e,M} \geq c'$, the together, we have

$$\liminf_{n\to\infty} \inf_{\hat{\theta}_n} \sup_{\theta\in\Theta} \mathbb{E}_\theta \left[w(\psi_n^{-1} d(\hat{\theta}_n, \theta)) \right] \geq w(A) c'.$$

Here, we first consider taking $M = 1$.

2 Lower bounds based on two hypotheses

One approach is to use the following proposition.

Proposition 2.1.

$$p_{e,1} \geq \sup_{\tau>0} \left\{ \frac{\tau}{1-\tau} P_1 \left(\frac{dP_0^a}{dP_1} \geq \tau \right) \right\}.$$

The other more common approach is to use distances and divergences between probabilities measures. See their definitions in the appendix. The following theorem is our key tool.

Theorem 2.2. *Let P_0 and P_1 be two probability measures on $(\mathcal{X}, \mathcal{A})$*

(i) If $V(P_1, P_0) \leq \alpha < 1$, then

$$p_{e,1} \geq \frac{1-\alpha}{2}$$

(total variation version)*.*

(ii) If $H^2(P_1, P_0) \leq \alpha < 2$, then

$$p_{e,1} \geq \frac{1}{2}(1 - \sqrt{\alpha(1-\alpha/4)})$$

(Hellinger version)*.*

(iii) If $K(P_1, P_0) \leq \alpha < \infty$ (or $\chi^2(P_1, P_0) \leq \alpha < \infty$), then

$$p_{e,1} \geq \max\left(\frac{1}{4}\exp(-\alpha), \frac{1-\sqrt{\alpha/2}}{2}\right)$$

(Kullback $/\chi^2$ version)*.*

We will illustrate their usage with the following examples.

Example 2.1 (Gaussian location family)**.** Consider $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} \mathcal{N}(\theta, \sigma^2)$, we would like to estimate $\theta \in \mathbb{R}$. Since we know that the sample mean converges at rate $O(\frac{\sigma}{\sqrt{n}})$, consider $\theta_{0n} = 0$ and $\theta_{1n} = c_0 \frac{\sigma}{\sqrt{n}}$. We have

$$K(P_1^{(n)}, P_0^{(n)}) = n\frac{\theta_{1n}^2}{2\sigma^2} = \frac{c_0^2}{2}.$$

Thus,

$$p_{e,1} \geq \max\left(\frac{1}{4}\exp\left(-\frac{c_0^2}{2}\right), \frac{1-c_0/2}{2}\right).$$

Since $\psi_n = \frac{\sigma}{\sqrt{n}}$, $A = \frac{c_0}{2}$, we have

$$\liminf_{n\to\infty} \inf_{\hat{\theta}_n} \sup_{\theta\in\Theta} \mathbb{E}_\theta\left[w\left(\psi_n^{-1} d\left(\hat{\theta}_n, \theta\right)\right)\right] \geq w\left(\frac{c_0}{2}\right) \max\left(\frac{1}{4}\exp\left(-\frac{c_0^2}{2}\right), \frac{1-c_0/2}{2}\right).$$

Take $d(\hat{\theta}_n, \theta) = |\hat{\theta}_n - \theta|$, we have

$$\liminf_{n\to\infty} \inf_{\hat{\theta}_n} \sup_{\theta\in\Theta} \mathbb{E}_\theta\left[\left|\hat{\theta}_n - \theta\right|\right] \geq \frac{\sigma}{\sqrt{n}} \cdot \frac{c_0}{2} \max\left(\frac{1}{4}\exp\left(-\frac{c_0^2}{2}\right), \frac{1-c_0/2}{2}\right)$$

and

$$\liminf_{n\to\infty} \inf_{\hat{\theta}_n} \sup_{\theta\in\Theta} \mathbb{E}_\theta\left[\left(\hat{\theta}_n - \theta\right)^2\right] \geq \frac{\sigma^2}{n} \cdot \frac{c_0^2}{4} \max\left(\frac{1}{4}\exp\left(-\frac{c_0^2}{2}\right), \frac{1-c_0/2}{2}\right).$$

For the first one, we take $c_0 = 1$ and obtain

$$\liminf_{n\to\infty} \inf_{\hat{\theta}_n} \sup_{\theta\in\Theta} \mathbb{E}_\theta\left[\left|\hat{\theta}_n - \theta\right|\right] \geq \frac{\sigma}{8\sqrt{n}}.$$

For the second one, we take $c_0 = 4/3$ and get

$$\liminf_{n\to\infty} \inf_{\hat{\theta}_n} \sup_{\theta\in\Theta} \mathbb{E}_\theta\left[\left(\hat{\theta}_n - \theta\right)^2\right] \geq \frac{2\sigma^2}{27n}.$$

In fact, for the second case, we can actually show that the sample mean is a minimax estimator using the classical Bayesian-type minimaxity result. See Jun Shao's book for it.

Example 2.2 (Uniform location family)**.** Consider now $Y_1, \dots, Y_n \overset{i.i.d.}{\sim} U(\theta, \theta+1)$. The KL-divergence approach won't work here since the support for different θ has non-overlapping region. In this case, we turn to the Hellinger distance. Since $\hat{\theta}_n = Y_{(1)}$ converges at the rate $O(n^{-1})$, we consider $\theta_{0n} = 0$ and $\theta_{1n} = \frac{2c_0}{n}$. For this case, $\psi_n = \frac{1}{n}$ and $A = c_0$. For n suitably large, we have

$$H^2(P_0, P_1) = \int_0^{\theta_{1n}} dt + \int_1^{\theta_{1n}+1} dt = \frac{4c_0}{n}$$

so

$$H^2(P_0^{(n)}, P_1^{(n)}) = 2\left(1 - \left(1 - \frac{2c_0}{n}\right)^n\right) \leq 4c_0.$$

Thus, we have

$$\liminf_{n\to\infty} \inf_{\hat{\theta}_n} \sup_{\theta\in\Theta} \mathbb{E}_\theta \left[\left|\hat{\theta}_n - \theta\right|\right] \geq \frac{c_0}{2n}(1 - 2\sqrt{c_0(1-c_0)})$$

and

$$\liminf_{n\to\infty} \inf_{\hat{\theta}_n} \sup_{\theta\in\Theta} \mathbb{E}_\theta \left[\left(\hat{\theta}_n - \theta\right)^2\right] \geq \frac{c_0^2}{2n^2}(1 - 2\sqrt{c_0(1-c_0)}).$$

Take c_0 to be a suitable constant and we complete the bound.

The previous two examples are somewhat too naive, as they can be completely solved using classical Bayesian theory. We conclude with a more involved example.

Example 2.3 (Nonparametric Regression)**.** Consider

$$Y_i = f(X_i) + \epsilon_i, \ i = 1, \dots n,$$

where $f : [0,1] \to \mathbb{R}$, $X_i = \frac{i}{n}$ and $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. For a given x_0, we are interested in estimating $f(x_0)$, and consider the lower bound for

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{f\in\mathcal{F}} \mathbb{E}_f \left[|T_n(x_0) - f(x_0)|\right].$$

Here the semi-distance is taken to be

$$d(f,g) := |f(x_0) - g(x_0)|.$$

Consider the Hölder class $\mathcal{F} = \Sigma(\beta, L)$, which satisfies

$$\left|f^{(l)}(x) - f^{(l)}(x')\right| \leq L\left|x - x'\right|^{\beta - l}, \ \forall x, x' \in [0,1]$$

for $l \leq \beta$. The rate we are aim to obtain is

$$\psi_n = n^{-\frac{\beta}{2\beta+1}}.$$

First, we consider $f_{0n} = 0$. To find a suitable f_{1n}, we need to find a function which is very close to 0, yet as far away from 0 as possible at x_0. Thus, we can first consider a function symmetric around 0. It decays as fast as possible when it gets away from $x = 0$, while satisfying the smoothness requirement. We can thus make use of smooth kernels used in nonparametric regression, for example

$$K_0(u) = \exp\left(-\frac{1}{1-u^2}\right) I(|u| \leq 1),$$

and set

$$f_{1n}(x) = Lh_n^\beta K\left(\frac{x - x_0}{h_n}\right), \quad x \in [0,1],$$

where

$$h_n = c_0 n^{-\frac{1}{2\beta+1}}, \quad c_0 > 0,$$

and where the function $K : \mathbf{R} \to [0, +\infty)$ satisfies

$$K \in \Sigma(\beta, 1/2) \cap C^\infty(\mathbf{R}) \quad \text{and} \quad K(u) > 0 \Longleftrightarrow u \in (-1/2, 1/2).$$

It is straightforward to check that $f_{1n} \in \mathcal{F}$. We have

$$d(f_{1n}, f_{0n}) = L c_0^\beta K(0) n^{-\frac{\beta}{2\beta+1}},$$

so $s = \frac{1}{2} L c_0^\beta K(0) n^{-\frac{\beta}{2\beta+1}}$ and $A = \frac{1}{2} L c_0^\beta K(0)$. The KL-divergence is

$$\begin{aligned}
K(P_1^{(n)}, P_0^{(n)}) &= \sum_{i=1}^n \frac{f_{1n}(X_i)^2}{2\sigma^2} \\
&= \frac{L^2 h_n^{2\beta}}{2\sigma^2} \sum_{i=1}^n K^2\left(\frac{X_i - x_0}{h_n}\right) \\
&\leq \frac{L^2 h_n^{2\beta}}{2\sigma^2} K^2(0) \sum_{i=1}^n I\left(\left|\frac{X_i - x_0}{h_n}\right| \leq \frac{1}{2}\right) \\
&= \frac{L^2 h_n^{2\beta}}{2\sigma^2} K^2(0) \sum_{i=1}^n I\left(x_0 - \frac{h_n}{2} \leq X_i \leq x_0 + \frac{h_n}{2}\right) \\
&\leq \frac{a_0}{\sigma^2} L^2 K^2(0) h_n^{2\beta} \max(nh_n, 1) \\
&\leq \frac{a_0}{\sigma^2} L^2 K^2(0) n h_n^{2\beta+1} \\
&= \frac{a_0}{\sigma^2} L^2 K^2(0) c_0^{2\beta+1}.
\end{aligned}$$

The rest part is to pick suitable constants. Since it does not affect the rate, we skip it.

A Definitions of Distances and Divergences

Definition A.1 (**Hellinger distance**).

$$H(P,Q) = \left(\int (\sqrt{p} - \sqrt{q})^2 d\nu \right)^{1/2}.$$

If P and Q are product measures, $P = \otimes_{i=1}^n P_i, \quad Q = \otimes_{i=1}^n Q_i$, then

$$H^2(P,Q) = 2\left(1 - \prod_{i=1}^n \left(1 - \frac{H^2(P_i, Q_i)}{2}\right)\right).$$

Definition A.2 (**Total variation distance**).

$$V(P,Q) = \sup_A |P(A) - Q(A)| = \frac{1}{2}\int |p-q| d\nu = 1 - \int \min(p,q) d\nu.$$

Definition A.3 (**Kullback divergence**).

$$K(P,Q) = \begin{cases} \int \log \frac{dP}{dQ} dP, & \text{if } P \ll Q \\ +\infty, & \text{otherwise} \end{cases}.$$

We will focus on minimax results with normal-distributed noise. For two normal distribution $P_0 = \mathcal{N}(\mu_0, \sigma^2)$ and $P_1 = \mathcal{N}(\mu_1, \sigma^2)$, we have

$$K(P_0, P_1) = \frac{(\mu_1 - \mu_0)^2}{2\sigma^2}.$$

If P and Q are product measures, $P = \otimes_{i=1}^n P_i, \quad Q = \otimes_{i=1}^n Q_i$, then

$$K(P,Q) = \sum_{i=1}^n K(P_i, Q_i).$$

This is a very handy property when dealing with independent observations, which is our primary focus.

Definition A.4 (**χ^2 divergence**).

$$\chi^2(P,Q) = \begin{cases} \int \left(\frac{dP}{dQ} - 1\right)^2 dQ, & \text{if } P \ll Q \\ +\infty, & \text{otherwise} \end{cases}.$$