Minimax Lower Bound

JB

Let $\{\mathbb{P}_\theta\}_{\theta\in\Theta}$ be the collection of probability measures on $(\mathcal{V}, \mathcal{B}(\mathcal{V}))$, $\mathcal{V}$ and Θ are metric spaces. Let $\hat{\theta} : \mathbb{M} \to \Theta$ be measurable, which is a estimator of θ. Let d be the metric of Θ.

Definition 1. We say $\hat{\theta}_0$ is minimax optimal over Θ if $\exists$ constant $C \geq c$ and $\phi \geq 0$ which relates to model s.t.

$$\text{(Upper bound) } \sup_{\theta\in\Theta} \mathbb{E}_\theta d(\hat{\theta}_0, \theta) < C\phi,$$
$$\text{(Lower bound) } \inf_{\hat{\theta}} \sup_{\theta\in\Theta} \mathbb{E}_\theta d(\hat{\theta}, \theta) \geq c\phi.$$

Among ϕ is called the minimax optimal rate of the estimation θ over Θ.

Remark 1. General reduction scheme for lower bound is to find ϕ (a constant relates to model) s.t. $\inf_{\hat{\theta}} \sup_{\theta\in\Theta} \mathbb{P}_\theta\{d(\hat{\theta}, \theta) \geq \phi\} \geq c$ since $\mathbb{E}_\theta \frac{d(\hat{\theta},\theta)}{\phi} \geq \mathbb{P}_\theta\{d(\hat{\theta}, \theta) \geq \phi\}$ by Markov Inequality.

The second step is to find finite $\{\theta_j\}_{j\leq m}$ satisfied $d(\theta_j, \theta_k) > 2\phi$, then

$$\inf_{\hat{\theta}} \sup_{\theta\in\Theta} \mathbb{P}_\theta\{d(\hat{\theta}, \theta) \geq \phi\} \geq \inf_{\hat{\theta}} \sup_{j} \mathbb{P}_{\theta_j}\{d(\hat{\theta}, \theta_j) \geq \phi\}$$

Let $\varphi(\hat{\theta}) = arg\min_j d(\hat{\theta}, \theta_j)$. Claim: $\inf_{\hat{\theta}} \sup_j \mathbb{P}_{\theta_j}\{d(\hat{\theta}, \theta_j) \geq \phi\} \geq \inf_{\hat{\theta}} \sup_j \mathbb{P}_{\theta_j}\{\varphi(\hat{\theta}) \neq j\}$ since if $\varphi(\hat{\theta}) \neq j$, then $\exists k \neq j$ s.t. $d(\hat{\theta}, \theta_k) < d(\hat{\theta}, \theta_j) \Rightarrow d(\hat{\theta}, \theta_j) \geq d(\theta_k, \theta_j) - d(\hat{\theta}, \theta_k) \geq d(\theta_k, \theta_j) - d(\hat{\theta}, \theta_j) \Rightarrow d(\hat{\theta}, \theta_j) \geq \frac{1}{2} d(\theta_j, \theta_k) \geq \phi$. So

$$\inf_{\hat{\theta}} \sup_{\theta\in\Theta} \mathbb{P}_\theta\{d(\hat{\theta}, \theta) \geq \phi\} \geq \inf_{\hat{\theta}} \sup_{j} \mathbb{P}_{\theta_j}\{\varphi(\hat{\theta}) \neq j\}$$

The last step is to transform minimax problem into a hypothesis testing problem. Let $\hat{\varphi} \in \{1, ..., m\}$. Note that $\varphi(\hat{\theta})$ is also a testing statistics valued in $\{1, ..., m\}$, then

$$\inf_{\hat{\theta}} \sup_{j} \mathbb{P}_{\theta_j}\{\varphi(\hat{\theta}) \neq j\} \geq \inf_{\hat{\varphi}} \sup_{j} \mathbb{P}_{\theta_j}\{\hat{\varphi} \neq j\}$$

Summary: For lower bound, we need to specify ϕ, find $\{\theta_j\}_{j\leq m}$ s.t. $d(\theta_j, \theta_k) > 2\phi$ and $c > 0$ s.t. $\inf_{\hat{\varphi}} \sup_j \mathbb{P}_{\theta_j}\{\hat{\varphi} \neq j\} \geq c$. To show ϕ is optimal, we need to find $\hat{\theta}_0$ s.t. $\sup_{\theta\in\Theta} \mathbb{E}_\theta d(\hat{\theta}_0, \theta) < C\phi$, which means that $\inf_{\hat{\theta}} \sup_{\theta\in\Theta} \mathbb{E}_\theta d(\hat{\theta}, \theta) \asymp \phi$ and $\sup_{\theta\in\Theta} \mathbb{E}_\theta d(\hat{\theta}_0, \theta) \asymp \phi$.

Definition 2. Let $\mathcal{V}$ be $\mathbb{R}^n$ and probability measure $\mathbb{P}$ on $\mathbb{R}^n$. Define the total variation of $\mathbb{P}_i$:

$$||\mathbb{P}_1 - \mathbb{P}_2||_{TV} = \sup_{A \in \mathcal{B}(\mathbb{R}^n)} |\mathbb{P}_1(A) - \mathbb{P}_2(A)|$$

If $\mathbb{P}_i << \lambda$, let f_i be the Radon derivative of $\mathbb{P}_i$. Then $||\mathbb{P}_1 - \mathbb{P}_2||_{TV} = \frac{1}{2}\int |f_1 - f_2| = 1 - \int \min(f_1, f_2)$.

Lemma 1. (Neyman-Pearson) $\mathbb{P}_{\theta_2}\{\hat{\varphi} = 1\} + \mathbb{P}_{\theta_1}\{\hat{\varphi} = 2\} \geq 1 - ||\mathbb{P}_1 - \mathbb{P}_2||_{TV}$.

Lemma 2. (Le Cam) $\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_\theta d(\hat{\theta}, \theta) \geq \frac{\phi}{2}(1 - ||\mathbb{P}_{\theta_1} - \mathbb{P}_{\theta_2}||_{TV})$, if $d(\theta_1, \theta_2) > 2\phi$.

Proof.

It's sufficient to show $\inf_{\hat{\varphi}} \sup_{j=1,2} \mathbb{P}_{\theta_j}\{\hat{\varphi} \neq j\} \geq \frac{1 - ||\mathbb{P}_{\theta_1} - \mathbb{P}_{\theta_2})||_{TV}}{2}$, which is true since that $\sup_{j=1,2} \mathbb{P}_{\theta_j}\{\hat{\varphi} \neq j\} \geq \frac{\mathbb{P}_{\theta_2}\{\hat{\varphi}=1\} + \mathbb{P}_{\theta_1}\{\hat{\varphi}=2\}}{2}$.

□

Remark 2. $\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_\theta d(\hat{\theta}, \theta) \geq \frac{d(\theta_1, \theta_2)}{4}(1 - ||\mathbb{P}_{\theta_1} - \mathbb{P}_{\theta_2}||_{TV})$

Example 1. (Minimax rate for Gaussian mean) Let $\mathbb{P}_\theta \sim N(\theta, \sigma^2)$.

Upper bound: We apply the MLE $\bar{X}$ for θ, then $\sup_{\theta \in \Theta} \mathbb{E}|\bar{X} - \theta|^2 = \frac{\sigma^2}{n}$.

Lower bound: To bound $||\mathbb{P}_1 - \mathbb{P}_2||_{TV}$ when $\mathbb{P}$ is a Gaussian measure, a more convenient divergence is K-L divergence: $D(\mathbb{P}_1 || \mathbb{P}_2) = \int f_1 \log \frac{f_1}{f_2}$.

(Pinsker) $||\mathbb{P}_1 - \mathbb{P}_2||_{TV} \leq \sqrt{D(\mathbb{P}_1||\mathbb{P}_2)}$. Then we have

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_\theta d(\hat{\theta}, \theta) \geq \frac{d(\theta_1, \theta_2)}{4}(1 - \sqrt{D(\mathbb{P}_1||\mathbb{P}_2)})$$

Note that $D(\mathbb{P}_{\theta_1}||\mathbb{P}_{\theta_2}) = \frac{n||\theta_1 - \theta_2||_2^2}{2\sigma^2}$. Choose θ_1 and θ_2 s.t. $||\theta_1 - \theta_2||^2 = \frac{\alpha}{n}$, then we have $||\mathbb{P}_{\theta_1} - \mathbb{P}_{\theta_2}||_{TV} \leq \sqrt{D(\mathbb{P}_{\theta_1}||\mathbb{P}_{\theta_2})} = \sqrt{\frac{n||\theta_1 - \theta_2||_2^2}{2\sigma^2}} = \sqrt{\frac{\alpha}{2\sigma^2}}$. Choose a α s.t. $1 - ||\mathbb{P}_{\theta_1} - \mathbb{P}_{\theta_2}||_{TV} > 0$, then we know $\frac{1}{n}$ is optimal.

Theorem 1. (Fano) $\inf_{\hat\varphi}\sup_{j\le m}\mathbb{P}_{\theta_j}\{\hat\varphi\ne j\}\ge 1-\frac{\frac{1}{m^2}\sum_{j,k}D(\mathbb{P}_{\theta_j}||\mathbb{P}_{\theta_k})+\log 2}{\log m}$.

Proof.

Let $P_j=\mathbb{P}_{\theta_j}\{\hat\varphi=j\}$, $Q_j=\frac{1}{m}\sum_k\mathbb{P}_{\theta_k}\{\hat\varphi=j\}$, then $\bar Q=\frac{1}{m}$. Let $K(p_1,p_2)=D(\mathbb{P}_1||\mathbb{P}_2)$, $\mathbb{P}_i\sim B(1,p_i)$, then $K(p,q)=p\log\frac{p}{q}+(1-p)\log\frac{1-p}{1-q}$ and K is convex. So

$$K(\bar P,\bar Q)=\bar P\log\bar P+(1-\bar P)\log(1-\bar P)-\bar P\log\bar Q-(1-\bar P)\log(1-\bar Q)\ge-\log 2+\bar P\log m$$

$$\Rightarrow\bar P\le\frac{K(\bar P,\bar Q)+\log 2}{\log m}\le\frac{\frac{1}{m}\sum_j K(P_j,Q_j)+\log 2}{\log m}\le\frac{\frac{1}{m^2}\sum_{j,k}K(\mathbb{P}_{\theta_j}\{\hat\varphi=j\},\mathbb{P}_{\theta_k}\{\hat\varphi=j\})+\log 2}{\log m}$$

$$\begin{aligned}\Rightarrow\sup_{j\le m}\mathbb{P}_{\theta_j}\{\hat\varphi\ne j\}\ge\frac{1}{m}\sum_j\mathbb{P}_{\theta_j}\{\hat\varphi\ne j\}\ge 1-\frac{1}{m}\sum_j(1-\mathbb{P}_{\theta_j}\{\hat\varphi\ne j\})=1-\bar P\\ \ge 1-\frac{\frac{1}{m^2}\sum_{j,k}K(\mathbb{P}_{\theta_j}\{\hat\varphi=j\},\mathbb{P}_{\theta_k}\{\hat\varphi=j\})+\log 2}{\log m}\end{aligned}$$

What we need to do is to show $K(\mathbb{P}_{\theta_j}\{\hat\varphi=j\},\mathbb{P}_{\theta_k}\{\hat\varphi=j\})\le D(\mathbb{P}_{\theta_j}||\mathbb{P}_{\theta_k})$.

$$D(\mathbb{P}_{\theta_j}||\mathbb{P}_{\theta_k})=\int f_i\log\frac{f_i}{f_k}=\int_{\hat\varphi=j}f_j\log\frac{f_j}{f_k}+\int_{\hat\varphi\ne j}f_j\log\frac{f_j}{f_k}$$

Let $f_k^j(\cdot)=f_k(\cdot|\hat\varphi=j)=\frac{f_k(\cdot)\mathbb{I}(\hat\varphi(\cdot)=j)}{\mathbb{P}_{\theta_k}\{\hat\varphi=j\}}$,

$$\begin{aligned}\int_{\hat\varphi=j}f_j\log\frac{f_j}{f_k}=\mathbb{P}_{\theta_j}\{\hat\varphi=j\}(\int f_j^j\log\frac{\mathbb{P}_{\theta_j}\{\hat\varphi=j\}}{\mathbb{P}_{\theta_k}\{\hat\varphi=k\}}+D(\mathbb{P}_{\theta_j}^{\hat\varphi=j}||\mathbb{P}_{\theta_k}^{\hat\varphi=j}))\\ \ge\mathbb{P}_{\theta_j}\{\hat\varphi=j\}\log\frac{\mathbb{P}_{\theta_j}\{\hat\varphi=j\}}{\mathbb{P}_{\theta_k}\{\hat\varphi=k\}}\end{aligned}$$

Similarly, $\int_{\hat\varphi\ne j}f_j\log\frac{f_j}{f_k}\ge\mathbb{P}_{\theta_j}\{\hat\varphi\ne j\}\log\frac{\mathbb{P}_{\theta_j}\{\hat\varphi\ne j\}}{\mathbb{P}_{\theta_k}\{\hat\varphi\ne k\}}$.

□

Remark 3. $\inf_{\hat\theta}\sup_{\theta\in\Theta}\mathbb{E}_\theta d(\hat\theta,\theta)\ge(\min_{\theta_i\ne\theta_j}\frac{d(\theta_i,\theta_j)}{2})\left(1-\frac{\frac{1}{m^2}\sum_{j,k}D(\mathbb{P}_{\theta_j}||\mathbb{P}_{\theta_k})+\log 2}{\log m}\right)$

Example 2. (Minimax rate for sparse vector) $\mathbb{P}_\theta \sim N(\theta, \sigma^2 I_p)$, $||\theta||_0 \leq s$.

Upper bound: Let $\hat{\theta}_0 = arg\min_{||a|| \leq s} ||x - a||_2^2$, then $\sup_{\theta \in \Theta} \mathbb{E}_\theta ||\theta - \hat{\theta}_0||^2 = O(\frac{s \log(1+\frac{p}{2s})}{n})$.

Lower bound: If $\exists\, \beta > 0$ and $\{\theta_j\}_{j \leq m}$ s.t. $\frac{s\log(1+\frac{p}{2s})}{n}\beta \leq ||\theta_j - \theta_k||^2$ and $1 - \frac{\frac{1}{m^2}\sum_{j,k} D(\mathbb{P}_{\theta_j}||\mathbb{P}_{\theta_k}) + \log 2}{\log m} > 0$. Then ϕ is optimal.

(Gilbert-Varshamov) If $1 \leq s \leq \frac{p}{8}$, then there exist $\{\omega_j\}_{j \leq m} \subset \{0,1\}^p$ s.t. $||\omega||_0 = s$, $\log m \geq \frac{1}{16} s \log(1 + \frac{p}{2s})$ and $\rho(\omega_i, \omega_j) := \sum_{k \leq q} I(\omega_i^{(k)} \neq \omega_j^{(k)}) \geq \frac{s}{2}$.

Let $\theta_j = \omega_j \sqrt{\frac{\log(1+\frac{p}{2s})}{n}\beta}$, then

$$\frac{1}{2}\frac{s\log(1+\frac{p}{2s})}{n}\beta \leq ||\theta_j - \theta_k||^2 \leq 2\frac{s\log(1+\frac{p}{2s})}{n}\beta \leq \frac{32\beta}{n}\log m$$

$$\Rightarrow 1 - \frac{\frac{1}{m^2}\sum_{j,k} D(\mathbb{P}_{\theta_j}||\mathbb{P}_{\theta_k}) + \log 2}{\log m} \geq 1 - \frac{\frac{32\beta}{2\sigma^2}\log m + \log 2}{\log m}$$

Let β be small enough s.t. $1 - \frac{\frac{16\beta}{\sigma^2}\log m + \log 2}{\log m} > 0$.

Lemma 3. (Assouad) Let $\Theta = \{0, 1\}^p$ and $\hat{T}$ is an estimator of $\psi(\theta)$, then

$$\inf_{\hat{T}} \max_{\theta\in\Theta} \mathbb{E}_\theta 2^s d^s(\hat{T}, \psi(\theta)) \geq (\min_{\theta\neq\theta'} \frac{d^s(\psi(\theta), \psi(\theta'))}{\rho(\theta, \theta')}) \frac{p}{2} \min_{\rho(\theta,\theta')=1} (1 - ||\mathbb{P}_\theta - \mathbb{P}_{\theta'}||_{TV}).$$

Proof.

$$\max_{\theta\in\Theta} \mathbb{E}_\theta (2d(\hat{T}, \psi(\theta)))^s \geq \frac{1}{2^p} \sum_{\theta\in\Theta} \mathbb{E}_\theta (2d(\hat{T}, \psi(\theta)))^s$$

Let $\hat{\theta} = arg \min_{\theta\in\Theta} d(\hat{T}, \psi(\theta))$, then

$$\frac{1}{2^p} \sum_{\theta\in\Theta} \mathbb{E}_\theta (2d(\hat{T}, \psi(\theta)))^s \geq \frac{1}{2^p} \sum_{\theta\in\Theta} \mathbb{E}_\theta (d(\hat{T}, \psi(\theta) + d(\hat{T}, \psi(\hat{\theta})))^s \geq$$

$$\frac{1}{2^p} \sum_{\theta\in\Theta} \mathbb{E}_\theta (d(\psi(\hat{\theta}), \psi(\theta)))^s \geq \frac{1}{2^p} \sum_{\theta\in\Theta} \mathbb{E}_\theta \frac{d^s(\psi(\hat{\theta}), \psi(\theta))}{\max\{\rho(\theta, \hat{\theta}), 1\}} \rho(\theta, \hat{\theta})$$

$$\geq \min_{\theta\neq\theta'} \frac{d^s(\psi(\theta), \psi(\theta'))}{\rho(\theta, \theta')} \frac{1}{2^p} \sum_{\theta\in\Theta} \mathbb{E}_\theta \rho(\theta, \hat{\theta})$$

$$\sum_{\theta\in\Theta} \frac{1}{2^p} \mathbb{E}_\theta \rho(\theta, \hat{\theta}) = \frac{1}{2^p} \sum_{\theta\in\Theta} \sum_{j\leq p} \mathbb{P}_\theta\{\hat{\theta}_j \neq \theta_j\} = \frac{1}{2^p} \sum_{j\leq p} \sum_{\theta_{-j}} \sum_{\theta_j=1,0} \mathbb{P}_\theta\{\hat{\theta}_j \neq \theta_j\}$$

$$\geq \frac{1}{2^p} \sum_{j\leq p} \sum_{\theta_{-j}} \min_{\rho(\theta^*,\theta')=1} (1 - ||\mathbb{P}_{\theta^*} - \mathbb{P}_{\theta'}||_{TV}) = \frac{p}{2} \min_{\rho(\theta^*,\theta')=1} (1 - ||\mathbb{P}_{\theta^*} - \mathbb{P}_{\theta'}||_{TV}).$$

□

Example 3. (Minimax rate for functional data) $Y_{ij} = \mathcal{X}_i(t_{ij}) + \varepsilon_{ij}$, $i = 1, ..., n$, $j = 1, ..., k$. ε_{ij} are mutually independent and zero mean s.t. $\mathbb{E}\varepsilon^2 = \sigma^2 < \infty$.

$\mathcal{X}_i$ is iid random function valued in $\mathbb{W}_q[0, 1]$, which is independent to ε_{ij} and $\mathbb{E}\mathcal{X}_i = m$.

t_{ij} is the observed location which is fixed and $\mathbb{E}||\mathcal{X}^{(q)}||_2^2 \leq M_0$. Let $\mathcal{P}(q, M_0)$ be all the probability measure of $\mathcal{X}$ s.t. $\mathbb{E}||\mathcal{X}^{(q)}||_2^2 \leq M_0$.

Upper bound: One can show that if we use smoothing spline $\hat{m}_\lambda$ for estimation of m, then $\sup_{\mathbb{P}\in\mathcal{P}(q,M_0)} \mathbb{E}_\mathbb{P}||\hat{m}_\lambda - m||_2^2 = O(k^{-2q} + n^{-1})$, if the tuning parameter satisfies some conditions.

Lower bound: Let $\hat{m}$ be the estimator, let $\mathcal{P}_1 = \{\mathbb{P} \in \mathcal{P}(q, M_0); \mathbb{P}\{\mathcal{X} \textit{ is a constant function}\} = 1\}$, then

$$\inf_{\hat{m}} \sup_{\mathbb{P}\in\mathcal{P}(q,M_0)} \mathbb{E}_\mathbb{P}||\hat{m} - m||_2^2 \geq \inf_{\hat{m}} \sup_{\mathbb{P}\in\mathcal{P}_1} \mathbb{E}||\hat{m} - m||_2^2 \gtrsim n^{-1}$$

Let $\phi_j \in \mathbb{W}_q[0,1]$ with distinct support and same norm $\sqrt{Ck^{-(2q+1)}}$, $j = 1, ..., 2k$. Define $\mathcal{P}_2$: the collection of all the probability measure valued in $\{\sum_j \theta_j \phi_j; \theta \in \{0,1\}^{2k}\}$. It's sufficient to show $\min_{\theta \neq \theta'} \frac{||(\psi(\theta) - \psi(\theta^{'})||_2^2}{\rho(\theta, \theta^{'})} = Ck^{-(2q+1)}$, where $\psi(\theta) = \sum_j \theta_j \phi_j$. This is true since $\min_{\theta \neq \theta'} \frac{||(\psi(\theta) - \psi(\theta^{'})||_2^2}{\rho(\theta, \theta^{'})} = ||\phi_j||_2^2 = Ck^{-(2q+1)}$.