Minimax lower bounds I

Kyoung Hee Kim

Sungshin University

Setting

- Family of probability measures $\{\mathbb{P}_\theta : \theta \in \Theta\}$ on a sigma field $\mathcal{A}$
- Estimator $\hat{\theta}$: measurable map from Ω to Θ
- $L(\hat{\theta}, \theta)$: loss function
- Maximum risk $R(\Theta, \hat{\theta}) := \sup_{\theta \in \Theta} \mathbb{E}_\theta L(\hat{\theta}, \theta)$
- Minimax risk with a minimax estimator $\hat{\theta}_{mm}$,

$$R(\Theta) = \inf_{\tilde{\theta}} \sup_{\theta \in \Theta} E_\theta L(\tilde{\theta}, \theta) = \sup_{\theta \in \Theta} E_\theta L(\hat{\theta}_{mm}, \theta).$$

Why minimax?

- Uniformity excludes super-efficient estimators
- Optimal rates reveal the difficulty in the model of interest
- Guidance on the choice of estimators

Various distances

P, Q: two probability measures with densities p, q w.r.t ν.

- **Total variation** distance

$$V(P,Q) = \sup_{A \in \mathcal{A}} |P(A) - Q(A)| = \sup_{A \in \mathcal{A}} \left| \int_A (p-q) d\nu \right|$$

- Squared **Hellinger** distance

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

- **Kullback–Leibler** (KL) divergence

$$KL(P,Q) = \int p \log \frac{p}{q} d\nu.$$

- **Chi-squared** χ^2 distance

$$\chi^2(P,Q) = \int \frac{p^2}{q} d\nu - 1.$$

Relation between distances

P, Q: two probability measures with densities p, q w.r.t ν.

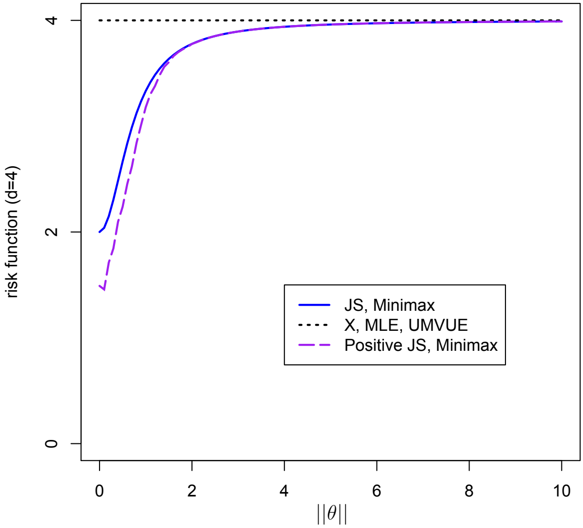
1. $\frac{1}{2}h^2(P,Q) \leq V(P,Q) \leq h(P,Q)\sqrt{1 - \frac{h^2(P,Q)}{4}}$
2. $V(P,Q) \leq h(P,Q) \leq \sqrt{KL(P,Q)} \leq \sqrt{\chi^2(P,Q)}$
3. $h^2(P^n, Q^n) \leq nh^2(P,Q)$
4. (Pinsker) $V(P,Q) \leq \sqrt{\frac{KL(P,Q)}{2}}$ and $V(P,Q) \leq 1 - \frac{1}{2}\exp(-KL(P,Q))$.

Relation to Bayes estimator

- Let $\hat{\theta}_\Pi$ be Bayes estimator using a prior Π.
- *Least favourable* prior gives a maximal Bayes risk (BR) for all prior distributions
- If Bayes risk of $\hat{\theta}_\Pi$ is equal to a maximal risk of $\hat{\theta}_\Pi$, then $\hat{\theta}_\Pi$ is minimax, and Π is least favourable.
- If maximal risk of $\hat{\theta}$ is equal to $\lim_{n\to\infty}$ BR of $\hat{\theta}_{\Pi_n}$, converging limit of Bayes risks with a sequence of priors (which is at least any Bayes risk for any prior), then $\hat{\theta}$ is minimax.

Examples of minimax estimator

1. Sample mean
   - Let X_i be i.i.d. $N(\theta, 1)$, $i = 1, \ldots, n$ and let $L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$.
   - $\bar{X}$ is minimax with a sequence of prior $\Pi_b = N(\mu, b)$ since $\hat{\theta}_{\Pi_b} = (n\bar{X} + \mu/b)/(n + 1/b)$ and the posterior variance $1/(n + 1/b)$ converges to $1/n$ as $b \to \infty$, which is equal to $Var(\bar{X})$.
2. James–Stein estimator
   - Let $X \sim N_d(\theta, I_d)$ where $d \geq 3$, and let $L(\hat{\theta}, \theta) = \sum_{i=1}^{d} (\hat{\theta}_i - \theta_i)^2$.
   - Minimax estimator $\hat{\theta}_{mm}$ (dominating X) is found as $\hat{\theta}_{mm} = \left(1 - \frac{d-2}{\sum_{i=1}^{d} X_i^2}\right) X$
   - But $\hat{\theta}_{mm}$ is not unique minimax (and not admissible). For instance, positive part estimator $\left(1 - \frac{d-2}{\sum_{i=1}^{d} X_i^2}\right)_+ X$ dominates $\hat{\theta}_{mm}$.


4
2
0
risk function (d=4)
0
2
4
6
8
10
$||\theta||$
JS, Minimax
X, MLE, UMVUE
Positive JS, Minimax

Minimax optimality

- Usually difficult to find minimax risk and minimax estimator.
- Typically satisfied if we find a 'good' lower bound $\ell(n)$ on $R(\Theta)$ and if we find a 'good' upper bound $u(n)$ by using a specific estimator $\tilde{\theta}$, where

$$\ell(n) \leq R(\Theta) \leq \sup_{\theta\in\Theta} \mathbb{E}_\theta L(\tilde{\theta}, \theta) \leq u(n) \tag{1}$$

$$\lim_{n\to\infty} \frac{u(n)}{\ell(n)} \leq C.$$

- If $\tilde{\theta}$ satisfies (1), then $\tilde{\theta}$ is called a **minimax optimal** estimator.

Possible questions

1. In a nonparametric regression problem where the regression function θ satisfies some smooth conditions (e.g. twice conti. differentiable), with the squared error loss function at one point, what is a minimax lower bound? That is,

$$\sup_{\theta\in\Theta} \mathbb{E}_\theta(\tilde{\theta}(x_0) - \theta(x_0))^2 \geq ...?$$

2. In a density estimation problem where a density f on $\mathbb{R}$ is assumed to be s times differentiable, with the integrated squared error loss function, what is a mnimax lower bound? That is,

$$\sup_{f\in\mathcal{F}_s} \mathbb{E}_f \int (\tilde{f}(x) - f(x))^2 dx \geq ...?$$

Lower bounds via TV distance

- d is a metric on Θ
- $d(\theta_0, \theta_1) \geq 2\delta$ where $\theta_0, \theta_1 \in \Theta$
- Denote

$$A_0 := \{\omega : d(\theta_0, \hat{\theta}(\omega)) < \delta\}.$$

(1) Risk bound implies TV bound.

- Suppose

$$\mathbb{P}_\theta\{d(\theta, \hat{\theta}) \geq \delta\} \leq \epsilon, \quad \forall\ \theta \in \Theta$$

- Then $\mathbb{P}_{\theta_0} A_0 \geq 1 - \epsilon$ and $\mathbb{P}_{\theta_1} A_0 \leq \epsilon$, which shows

$$V(\mathbb{P}_{\theta_0}, \mathbb{P}_{\theta_1}) \geq \mathbb{P}_{\theta_0} A_0 - \mathbb{P}_{\theta_1} A_0 \geq 1 - 2\epsilon$$

Lower bounds via TV distance

- d is a metric on Θ
- $d(\theta_0, \theta_1) \geq 2\delta$ where $\theta_0, \theta_1 \in \Theta$
- Denote

$$A_0 := \{\omega : d(\theta_0, \hat{\theta}(\omega)) < \delta\}.$$

(2) TV bound implies risk bound.

- Suppose

$$V(\mathbb{P}_{\theta_0}, \mathbb{P}_{\theta_1}) < 1 - 2\epsilon.$$

- Then, $\sup_{\theta\in\Theta} \mathbb{P}_\theta \left(d(\theta, \hat{\theta}) \geq \delta\right) = R(\Theta, \hat{\theta}) > \epsilon$ since

$$\begin{aligned} 2R(\Theta, \hat{\theta}) &\geq \mathbb{P}_{\theta_0}\left(d(\theta_0, \hat{\theta}) \geq \delta\right) + \mathbb{P}_{\theta_1}\left(d(\theta_1, \hat{\theta}) \geq \delta\right) \\ &\geq \mathbb{P}_{\theta_0} A_0^c + \mathbb{P}_{\theta_1} A_0 \\ &\geq 1 - \sup_A |\mathbb{P}_{\theta_0} A - \mathbb{P}_{\theta_1} A| > 2\epsilon. \end{aligned}$$

Lower bounds via testing

Above argument uses

- $L(\xi, \theta) = \mathbb{1}_{\{d(\xi,\theta) \geq \delta\}}$
- $L(\xi, \theta_0) + L(\xi, \theta_1) \geq 1$ for all ξ if $d(\theta_0, \theta_1) \geq 2\delta$.

For a general loss function,

- Suppose $\inf_{\xi \in \Theta} (L(\xi, \theta_0) + L(\xi, \theta_1)) \geq d(\theta_0, \theta_1)$ for a pair θ_0 and θ_1 in Θ.
- Then

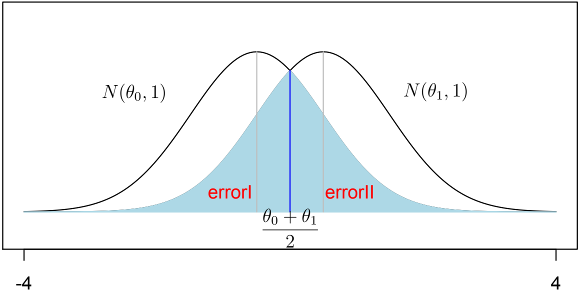
$$\inf_{\xi \in \Theta} \left(\frac{L(\xi, \theta_0)}{d(\theta_0, \theta_1)} + \frac{L(\xi, \theta_1)}{d(\theta_0, \theta_1)} \right) \geq 1.$$

Lower bounds via testing

$$
\begin{aligned}
2R(\Theta,\hat{\theta}) &\geq \mathbb{E}_{\theta_0} L(\hat{\theta},\theta_0) + \mathbb{E}_{\theta_1} L(\hat{\theta},\theta_1) \\
&= d(\theta_0,\theta_1)\left[\mathbb{E}_{\theta_0}\frac{L(\hat{\theta},\theta_0)}{d(\theta_0,\theta_1)} + \mathbb{E}_{\theta_1}\frac{L(\hat{\theta},\theta_1)}{d(\theta_0,\theta_1)}\right] \\
&\geq d(\theta_0,\theta_1) \inf_{f_0+f_1\geq 1, f_0, f_1 \geq 0} \left[\mathbb{E}_{\theta_0} f_0 + \mathbb{E}_{\theta_1} f_1\right]
\end{aligned}
$$

- Note that

$$
\begin{aligned}
\inf_{f_0+f_1\geq 1, f_0, f_1 \geq 0} \mathbb{E}_{\theta_0} f_0 + \mathbb{E}_{\theta_1} f_1 &= \int (p_{\theta_0} \wedge p_{\theta_1})\, d\nu \\
&= 1 - \frac{1}{2}\int |p_{\theta_0} - p_{\theta_1}| d\nu \\
&= 1 - V(\mathbb{P}_{\theta_0}, \mathbb{P}_{\theta_1})
\end{aligned}
$$


$N(\theta_0, 1)$
$N(\theta_1, 1)$
errorI
errorII
$\frac{\theta_0 + \theta_1}{2}$
-4
4

Testing affinity (TA)

- Testing affinity: $\int (p_{\theta_0} \wedge p_{\theta_1})\, d\nu := ||\mathbb{P}_{\theta_0} \wedge \mathbb{P}_{\theta_1}||_1$
  - $TA = \mathbb{E}_{\theta_0} \mathbb{1}_{\{p_{\theta_0} \leq p_{\theta_1}\}} + \mathbb{E}_{\theta_1} \mathbb{1}_{\{p_{\theta_1} \leq p_{\theta_0}\}} =$ Sum of two errors
  - If $\mathbb{P}_{\theta_0} = \mathbb{P}_{\theta_1}$, then $TA = 1$ meaning testing impossible
  - If $\text{supp}(\mathbb{P}_{\theta_0}) = [0, 1]$, $\text{supp}(\mathbb{P}_{\theta_1}) = [2, 3]$, then $TA = 0$ meaning perfect test
- Calculation of TA for product measures $\mathbb{P}_{\theta_0} = P_0^n, \mathbb{P}_{\theta_1} = P_1^n$
  - $(1 - TA)^2 \leq nh^2(P_0, P_1)$
  - $h^2(P_0, P_1) \leq KL(P_0, P_1) \leq \chi^2(P_0, P_1)$ for $P_0 \ll P_1$

Reduction strategy

- Restrict attention to a finite subset

$$\Theta_F := \{\theta_\alpha : \alpha \in A\} \subseteq \Theta$$

  where A is a finite index set.
- With an estimator $\hat{\theta}$, define

$$\hat{\alpha} := \arg\min_{\alpha \in A} L(\hat{\theta}, \theta_\alpha).$$

- Assume with a metric d, for some constant $\delta > 0$,

$$\inf_{\xi \in \Theta} \left(L(\xi, \theta_\alpha) + L(\xi, \theta_\beta)\right) \geq \delta d(\theta_\alpha, \theta_\beta) > 0. \ \ \forall\ \alpha, \beta \in A \quad (2)$$

Reduction strategy

- Reduce maximum risk by the average risk over the finite sub parameter space

$$
\begin{aligned}
R(\Theta, \hat{\theta}) &= \frac{1}{2} \sup_{\theta \in \Theta} E_\theta 2L(\hat{\theta}, \theta) \\
&\geq \frac{1}{2} \max_{\alpha \in A} E_{\theta_\alpha} 2L(\hat{\theta}, \theta_\alpha) \\
&\geq \frac{1}{2} \max_{\alpha \in A} E_{\theta_\alpha} \left(L(\hat{\theta}, \theta_\alpha) + L(\hat{\theta}, \theta_{\hat{\alpha}}) \right) \quad \text{by def. of } \hat{\alpha} \\
&\geq \frac{\delta}{2} \max_{\alpha A} E_{\theta_\alpha} d(\theta_{\hat{\alpha}}, \theta_\alpha) \quad \text{by condition (2)} \\
&\geq \frac{\delta}{2|A|} \sum_{\alpha \in A} E_{\theta_\alpha} d(\theta_{\hat{\alpha}}, \theta_\alpha). \quad \text{bound max by average}
\end{aligned}
$$

Two-point tests

- Let $A = \{0, 1\}$ and $d(\theta_\alpha, \theta_\beta) = \mathbb{1}_{\{\theta_\alpha \neq \theta_\beta\}}$ where $\alpha, \beta \in A$.
- By the above calculation,

$$\begin{aligned} R(\Theta, \hat{\theta}) &\geq \frac{\delta}{4}\left(E_{\theta_0} d(\theta_{\hat{\alpha}}, \theta_0) + E_{\theta_1} d(\theta_{\hat{\alpha}}, \theta_1)\right) \\ &\geq \frac{\delta}{4}\left(E_{\theta_0}(\hat{\alpha} = 1) + E_{\theta_1}(\hat{\alpha} = 0)\right) \\ &\geq \frac{\delta}{4} \inf\left\{E_{\theta_0} f_0 + E_{\theta_1} f_1 : 0 \leq f_0, f_1 \leq 1, f_0 + f_1 = 1\right\} \\ &\geq \frac{\delta}{4} \|\mathbb{P}_{\theta_0} \wedge \mathbb{P}_{\theta_1}\|_1 . \end{aligned}$$

- Thus

$$R(\Theta) \geq \frac{\delta}{4} \|\mathbb{P}_{\theta_0} \wedge \mathbb{P}_{\theta_1}\|_1 \geq \frac{\delta}{4}\left(1 - h(\mathbb{P}_{\theta_0}, \mathbb{P}_{\theta_1})\right). \tag{3}$$

Le Cam's Lemma

Construct

$$A := \{\theta_0, \theta_1\} \subseteq \Theta$$

such that

1. $\inf_{\xi \in \Theta} \left(L(\xi, \theta_0) + L(\xi, \theta_1) \right) \geq \delta$,
2. $||\mathbb{P}_{\theta_0} \wedge \mathbb{P}_{\theta_1}||_1 \geq c > 0$.

Then, for every estimator $\hat{\theta}$,

$$\sup_{\theta \in \Theta} \mathbb{E}_\theta L(\hat{\theta}, \theta) \geq \frac{c\delta}{4}.$$

Examples

(1) Normal location models

Let $Y_1, \dots, Y_n \sim N(\theta, 1)$, where $\theta \in \Theta \in \mathbb{R}$. Then for any estimator $\tilde{\theta}$,

$$\sup_{\theta \in \Theta} \mathbb{E}_\theta \left(\tilde{\theta} - \theta\right)^2 \geq Cn^{-1}.$$

- $\mathbb{P}_\theta = P_\theta^n$ (i.i.d. sample) and $P_\theta = N(\theta, 1)$ with a loss function $L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$.
- (loss cond.) $L(\xi, \theta_0) + L(\xi, \theta_1) \geq \frac{1}{2} L(\theta_0, \theta_1) \geq_{(?)} \delta$
- (testing cond.) $V^2(\mathbb{P}_{\theta_0}, \mathbb{P}_{\theta_1}) \leq h^2(\mathbb{P}_{\theta_0}, \mathbb{P}_{\theta_1}) \leq nh^2(P_{\theta_0}, P_{\theta_1}) \leq nKL(P_{\theta_0}, P_{\theta_1}) = n\frac{(\theta_1 - \theta_0)^2}{2} \leq_{(?)} 1/2$
- Fix $\theta_0 \in \mathbb{R}$ and take $\theta_1 = \theta_0 + 1/\sqrt{n}$. Then (2) is satisfied with $\delta = 1/(2n)$, and $h^2(\mathbb{P}_{\theta_0}, \mathbb{P}_{\theta_1}) \leq 1/2$.
- By (3), we have $R(\Theta) \geq \frac{1}{8n}(1 - \sqrt{1/2})$.

(2) Nonparametric regression

Let $Y_i = \theta(x_i) + \epsilon_i$ for $i = 1, \ldots, n$, where $\epsilon_i \sim N(0, 1)$, $x_i = i/n$, $\theta \in \Theta$ with Θ the set of all twice continuously differentiable functions on $[0, 1]$, $\theta''(x) < M$. Then for any estimator $\tilde{\theta}$ and any $x_0 \in [0, 1]$,

$$\sup_{\theta \in \Theta} \mathbb{E}_\theta \left(\tilde{\theta}(x_0) - \theta(x_0) \right)^2 \geq C n^{-4/5}.$$

- $\mathbb{P}_\theta = \prod_{i=1}^n P_{\theta,i}$ and $P_{\theta,i} = N(\theta(i/n), 1)$ with a loss function $L(\hat{\theta}, \theta) = (\hat{\theta}(x_0) - \theta(x_0))^2$.
- (loss cond.) $L(\xi, \theta_0) + L(\xi, \theta_1) \geq \frac{1}{2} L(\theta_0, \theta_1) \geq_{(?)} \delta$
- (testing cond.) $h^2(\mathbb{P}_{\theta_0}, \mathbb{P}_{\theta_1}) \leq KL(\mathbb{P}_{\theta_0}, \mathbb{P}_{\theta_1}) \leq_{(?)} 1/2$

- Let $\theta_0 = 0$ on $[0,1]$ and $\theta_1(x) = Mh_n^2 K\left(\frac{x-x_0}{h_n}\right)$ where $K(t) = a\exp\left(-\frac{1}{1-(2t)^2}\right)\mathbb{1}_{\{|2t|\leq 1\}}$ where a is a normalizing constant so $K(t)$ is a kernel, and let $h = \tilde{c}n^{-1/5}$.
- Check $\theta_0, \theta_1 \in \Theta$.
- Check $L(\xi,\theta_0) + L(\xi,\theta_1) \geq \frac{1}{2}L(\theta_0,\theta_1) = \frac{M^2}{2}h_n^4 K^2(0) = \frac{M^2}{2}h_n^4 a^2\exp(-2)$
- Check

$$KL(\mathbb{P}_{\theta_0},\mathbb{P}_{\theta_1}) = \int\cdots\int\prod_{i=1}^n \phi(u_i)\log\frac{\prod_{i=1}^n\phi(u_i)}{\prod_{i=1}^n\phi(u_i-\theta_1(x_i))}du_1\ldots du_n$$

$$= \int\cdots\int\prod_{i=1}^n\phi(u_i)\sum_{i=1}^n\left(-u_i\theta_1(x_i) + \frac{1}{2}\theta_1(x_i)^2\right)du_1\ldots du_n$$

$$
\begin{aligned}
KL(\mathbb{P}_{\theta_0},\mathbb{P}_{\theta_1}) &= \frac{1}{2}\sum_{i=1}^{n}\theta_1(x_i)^2 \\
&= \frac{M^2}{2}\sum_{i=1}^{n} h_n^4 a^2 \exp\left(-\frac{2}{1-4\left(\frac{x_i-x_0}{h_n}\right)^2}\right)\mathbb{1}_{\{|x_i-x_0|\le\frac{h_n}{2}\}} \\
&\le \frac{M^2}{2}h_n^4 a^2 \sum_{i=1}^{n}\mathbb{1}_{\{x_0-\frac{h_n}{2}\le\frac{i}{n}\le x_0+\frac{h_n}{2}\}} \\
&\le \frac{M^2}{2}h_n^4 a^2 n h_n = \frac{1}{2}a^2 n h_n^5
\end{aligned}
$$

Since $h \sim n^{-1/5}$, by choosing a sufficiently small, we have the lower bound of order $h_n^4 \sim n^{-4/5}$.

Multiple tests

- Let $A = \{0,1\}^m$ and $d(\theta_\alpha, \theta_\beta) = H(\alpha, \beta) = \sum_{k=1}^m \mathbb{1}_{\{\alpha_k \neq \beta_k\}}$ where $\alpha, \beta \in A$.
- e.g. $\alpha = (0, 0, \ldots, 1, 0, 1, \ldots, 1)$
- Start from

$$R(\Theta, \hat{\theta}) \geq \frac{\delta}{2|A|} \sum_{\alpha \in A} \mathbb{E}_{\theta_\alpha} d(\theta_{\hat{\alpha}}, \theta_\alpha)$$

$$
\begin{aligned}
R(\Theta, \hat{\theta}) &\geq \frac{\delta}{2|A|} \sum_{\alpha \in A} \mathbb{E}_{\theta_\alpha} d(\theta_{\hat{\alpha}}, \theta_\alpha) \\
&= \frac{\delta}{2 \cdot 2^m} \sum_{\alpha \in A} \sum_{k=1}^{m} \mathbb{E}_{\theta_\alpha} \mathbb{1}\{\theta_{\hat{\alpha}_k} \neq \theta_{\alpha_k}\} \\
&= \frac{\delta}{4} \sum_{k=1}^{m} \left[\frac{1}{2^{m-1}} \sum_{\alpha \in A_{0,k}} \mathbb{E}_{\theta_\alpha} \mathbb{1}_{\{\hat{\alpha}_k \neq 0\}} + \frac{1}{2^{m-1}} \sum_{\alpha \in A_{1,k}} \mathbb{E}_{\theta_\alpha} \mathbb{1}_{\{\hat{\alpha}_k \neq 1\}} \right] \\
&= \frac{\delta}{4} \sum_{k=1}^{m} \left(\bar{\mathbb{P}}_{0,k} \mathbb{1}_{\{\hat{\alpha}_k \neq 0\}} + \bar{\mathbb{P}}_{1,k} \mathbb{1}_{\{\hat{\alpha}_k \neq 1\}} \right) \\
&\geq \frac{\delta}{4} \sum_{k=1}^{m} ||\bar{\mathbb{P}}_{0,k} \wedge \bar{\mathbb{P}}_{1,k}||_1 \geq \frac{\delta}{4} m \min_{H(\alpha,\beta)=1} ||\mathbb{P}_{\theta_\alpha} \wedge \mathbb{P}_{\theta_\beta}||_1
\end{aligned}
$$

where $\bar{\mathbb{P}}_{i,k} = \frac{1}{2^{m-1}} \sum_{\alpha \in A_{i,k}} \mathbb{P}_{\theta_\alpha}$, and $A_{i,k} = \{\alpha \in A : \alpha_k = 0\}$ where $i = 0$ or 1.

Assouad's Lemma

Construct

$$A := \{\theta_\alpha,\ \alpha \in \{0,1\}^m\} \subseteq \Theta$$

such that

1. $\inf_{\xi\in\Theta} \left(L(\xi,\theta_\alpha) + L(\xi,\theta_\beta)\right) \geq \delta H(\alpha,\beta)\ \forall\ \alpha,\beta \in \{0,1\}^m$,
2. $||\mathbb{P}_{\theta_\alpha} \wedge \mathbb{P}_{\theta_\beta}||_1 \geq c > 0$ if $H(\alpha,\beta) = 1$.

Then, for every estimator $\hat{\theta}$,

$$\sup_{\theta\in\Theta} \mathbb{E}_\theta L(\theta,\hat{\theta}) \geq \frac{c\delta}{4} m.$$

Notation: $H(\alpha,\beta) = ||\alpha-\beta||_0 = \sum_{k=1}^m \mathbb{1}_{\{\alpha_k\neq\beta_k\}}$.

Examples

(3) Gaussian sequence models

Let $Y_i = \theta_i + \sigma\epsilon_i$ where $\epsilon_i \sim N(0,1)$ and $\sigma \to 0$ with $\theta = (\theta_1, \theta_2, ...) \in \Theta = \left\{\theta : \sum_k k^{2s}\theta_k^2 \leq M\right\}$ where M is a constant. Then for any estimator $\tilde{\theta}$,

$$\sup_{\theta \in \Theta} \mathbb{E}_\theta \|\theta - \tilde{\theta}\|_2^2 \geq C\sigma^{\frac{4s}{1+2s}}.$$

- $P_{\theta_k} = N(\theta_k, \sigma^2)$ for $k = 1, 2, ...$ where M is a constant.
- $\mathbb{P}_\theta = \prod_{k \in \mathbb{N}} P_{\theta_k}$
- $L(\theta, t) = \sum_k (\theta_k - t_k)^2$ satisfying

$$L(\theta_\alpha, t) + L(\theta_\beta, t) \geq \frac{1}{2} L(\theta_\alpha, \theta_\beta).$$

- Need to construct $\theta_\alpha \in \Theta$ with $\alpha \in \{0,1\}^m$ s.t.
  1. $\frac{1}{2}\sum_k(\theta_{\alpha,k} - \theta_{\beta,k})^2 \geq \delta||\alpha - \beta||_0$ for all $\alpha, \beta \in \{0,1\}^m$.
  2. $||\mathbb{P}_{\theta_\alpha} \wedge \mathbb{P}_{\theta_\beta}||_1 \geq c > 0$ for $||\alpha - \beta||_0 = 1$.

- Construct a finite subset by setting

$$\begin{aligned}\theta_\alpha &= (\theta_{\alpha_1}, \theta_{\alpha_2}, ..., \theta_{\alpha_m}, 0, 0, ...)\\ &= \epsilon(\alpha_1, \alpha_2, ..., \alpha_m, 0, 0, ...).\end{aligned}$$

Then

$$\frac{1}{2}\sum_k(\theta_{\alpha_k} - \theta_{\beta_k})^2 = \frac{1}{2}\epsilon^2 H(\alpha, \beta)$$

with $\delta = \frac{1}{2}\epsilon^2$.

- TA calculation: w.l.o.g. $\alpha_1 \neq \beta_1$, then

$$\begin{aligned} ||\mathbb{P}_{\theta_\alpha} \wedge \mathbb{P}_{\theta_\beta}||_1 &= \left\| \left(P_0 \times \mathcal{Q} \times \prod_{i>m} P_0 \right) \wedge \left(P_\epsilon \times \mathcal{Q} \times \prod_{i>m} P_0 \right) \right\|_1 \\ &= ||P_0 \wedge P_\epsilon||_1. \end{aligned}$$

- With normal distribution,

$$\begin{aligned} ||P_0 \wedge P_\epsilon||_1 &= 2 \int_{-\infty}^{\epsilon/2} \phi_{\sigma^2}(x - \epsilon) dx \\ &= 2 \left(1 - \Phi(\frac{\epsilon}{2\sigma}) \right) \\ &> 0 \end{aligned}$$

if $\epsilon = c_0 \sigma$ for $\sigma \to 0$.

- Check $\theta_\alpha \in \Theta$ by letting

$$m = \lfloor (M\epsilon^{-2})^{1/(1+2s)} \rfloor$$

since $\sum_k k^{2s}\theta_k^2 \le \sum_{k=1}^m k^{2s}\epsilon^2 \le m^{1+2s}\epsilon^2 \le M$.

- Lower bound is obtained as

$$m\epsilon^2 \sim \epsilon^{4s/(1+2s)} \sim \sigma^{4s/(1+2s)}$$

which gives $n^{-2s/(1+2s)}$ with the usual choice $\sigma = n^{-1/2}$.

(4) Smooth density estimation

Let $Y_1, \ldots, Y_n$ be an i.i.d. sample from a density $f \in \mathcal{F}_s$ where $\mathcal{F}_s$ is a class of s times differentiable densities with Hölder type condition $\int \left(f^{(s-1)}(x+h) - f^{(s-1)}(x)\right)^2 dx \leq (M|h|)^2$. Then

$$\sup_{f \in \mathcal{F}_s} \mathbb{E}_f \int (\tilde{f}(x) - f(x))^2 dx \geq C n^{-\frac{2s}{1+2s}}.$$

- $\mathbb{P}_f = P_f^n$ where P_f has a density f.
- Loss $L(f, g) = \int (f(x) - g(x))^2 dx$ satisfying

$$L(\xi, f_\alpha) + L(\xi, f_\beta) \geq \frac{1}{2} L(f_\alpha, f_\beta)$$

- Typical construction: small perturbations on a base function

$$f_\alpha(x) = f_0(x) + \epsilon \sum_{k=1}^{m} \alpha_k \varphi_k(x)$$

- For the loss separation condition,

$$\frac{1}{2}\int (f_\alpha(x) - f_\beta(x))^2 = \frac{1}{2}\epsilon^2 \int \left(\sum_{k=1}^{m} (\alpha_k - \beta_k)\varphi_k(x)\right)^2$$

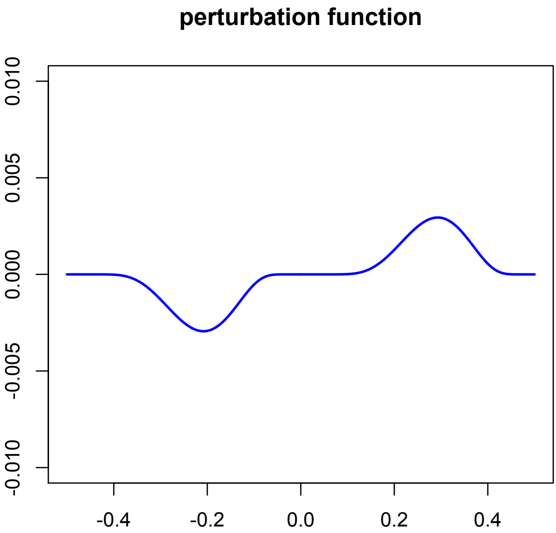
$$\geq_{(?)} \delta \sum_{k=1}^{m} (\alpha_k - \beta_k)^2$$

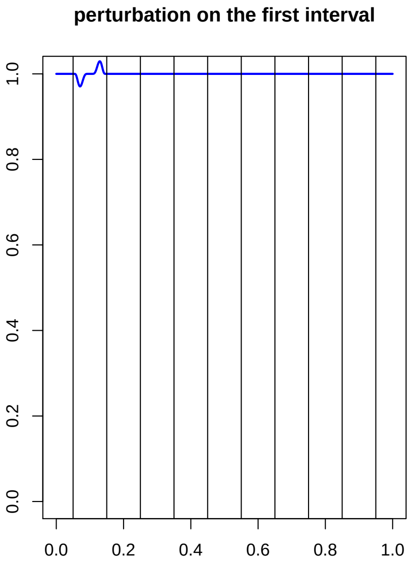
- Convenient if $\int \varphi_k \varphi_{k'} = V \mathbb{1}_{\{k=k'\}}$ s.t. $\delta = V\epsilon^2/2$.

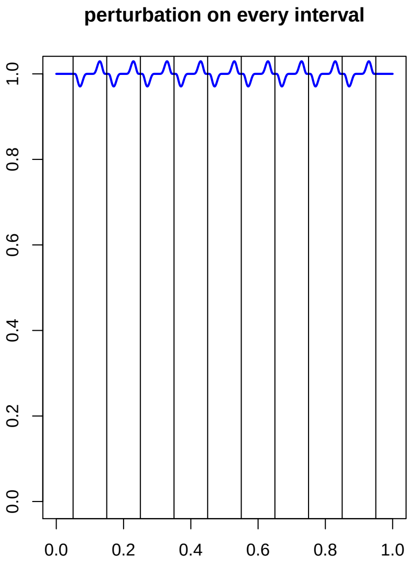
- Ibragimov and Has'minskii(1983)
  - Construct perturbed uniform densities on $[0, 1]$
  - $f_0(x) = 1$
  - $\varphi_k(x) = \varphi(mx - k)$ for $k = 1, ..., m - 1$ where

$$\varphi(x) = c_0 \exp\left(-\frac{1}{x} - \frac{1}{1-2x}\right) \mathbb{1}_{\{0 \le x \le \frac{1}{2}\}}$$

    with $\varphi(-x) = -\varphi(x)$, where c_0 is small enough to satisfy $\varphi \in \mathcal{F}$.
  - Perturbation function $\varphi(x)$: infinitely differentiable, $\int_{-1/2}^{1/2} \varphi(x)dx = 0$.
  - $\int_0^1 f_\alpha(x)dx = \int_0^1 \left[1 + \epsilon \sum_{k=1}^m \alpha_k \varphi_k(x)\right] dx = 1$ with $f_\alpha(x) \ge c_1 > 0$


perturbation function
0.010
0.005
0.000
-0.005
-0.010
-0.4
-0.2
0.0
0.2
0.4


perturbation on the first interval
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0


perturbation on every interval
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

- $\int \varphi_k \varphi_{k'} = \mathbb{1}\{k = k'\}\frac{C}{m}$, then $\delta = C\epsilon^2/(2m)$
- Then for the testing condition with $||\alpha - \beta||_0 = 1$,

$$\begin{aligned}\chi^2(P_{f_\alpha}, P_{f_\beta}) &:= \int \frac{(f_\alpha - f_\beta)^2}{f_\alpha} dx \\ &\leq \frac{1}{c_1} \int (f_\alpha - f_\beta)^2 \\ &= \frac{1}{c_1}\delta ||\alpha - \beta||_0 = \frac{\delta}{c_1},\end{aligned}$$

we need

$$\frac{1}{n} \sim \delta = \frac{C\epsilon^2}{2m}$$

with the lower bound $m\delta \sim \epsilon^2 \sim m/n$ up to a constant.

- [*] $\varphi_k^{(s-1)}(x) = m^{s-1}\varphi^{(s-1)}(mx - k + 1/2)$
- Check $f_\alpha \in \mathcal{F}$.

$$\int (f_\alpha^{(s-1)}(x+h) - f_\alpha^{(s-1)}(x))^2 dx$$
$$= \epsilon^2 \int \left(\sum_k \alpha_k (\varphi_k^{(s-1)}(x+h) - \varphi_k^{(s-1)}(x)) \right)^2 dx$$
$$\leq_{[*]} \epsilon^2 m^{2s-2} \int \left(\varphi^{(s-1)}(x+mh) - \varphi^{(s-1)}(x) \right)^2 dx$$
$$\leq \epsilon^2 m^{2s-2} (M|mh|)^2 \quad \text{since } \varphi \in \mathcal{F}$$
$$= M^2 \epsilon^2 m^{2s} |h|^2 \leq_{(?)} (M|h|)^2$$

if $m \sim \epsilon^{-1/s}$.

- From testing condition, $1/n \sim \epsilon^2/m \sim \epsilon^{2+(1/s)}$. That is, $\epsilon \sim n^{-s/(1+2s)}$. Then the lower bound is $\epsilon^2 \sim n^{-2s/(1+2s)}$.

Lemma

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

Proof.

Let $A_0 = \{x \in \mathcal{X} : p(x) \geq q(x)\}$.

1. Then $V \geq P(A_0) - Q(A_0) = \frac{1}{2}\int |p-q|d\nu = 1 - \int (p \wedge q)$.
2. For all $A \in \mathcal{A}$, $|\int_A (p-q)d\nu| = |\int_{A\cap A_0}(p-q) + \int_{A \cap A_0^c}(p-q)| \leq \max\left\{\int_{A_0}(p-q), \int_{A_0^c}(q-p)\right\} = \frac{1}{2}\int |p-q|$.

By the above two, we have proved the claims. □

Lemma

$\frac{1}{2}h^2(P,Q) \le V(P,Q) \le h(P,Q)\sqrt{1-\frac{h^2(P,Q)}{4}}$.

Proof.

1. $\frac{1}{2}h^2(P,Q) = 1 - \int\sqrt{pq} = 1 - \left(\int_{p>q}\sqrt{pq} + \int_{q>p}\sqrt{pq}\right) \le 1 - \left(\int_{p>q} q + \int_{q>p} p\right) = 1 - \int(p\wedge q) = V(P,Q)$.
2. $(1-\frac{1}{2}h^2(P,Q))^2 = \left(\int\sqrt{(p\wedge q)(p\vee q)}\right)^2 \le \int(p\wedge q)\int(p\vee q) = \int(p\wedge q)\left(2-\int(p\wedge q)\right) = (1-V)(1+V) = 1-V^2$. Thus $V(P,Q)^2 \le 1-(1-\frac{1}{2}h^2(P,Q))^2 = h(P,Q)^2(1-\frac{1}{4}h^2(P,Q))$.

□

Lemma

$h^2(P,Q) \leq KL(P,Q) \leq \chi^2(P,Q)$.

Proof.

1. $KL(P,Q) = \int_{pq>0} p \log \frac{p}{q} = 2\int_{pq>0} p \log \sqrt{\frac{p}{q}} \geq -2\int_{pq>0} p \log \left(\sqrt{\frac{q}{p}} - 1 + 1\right) \geq -2\int_{pq>0} p\left(\sqrt{\frac{q}{p}} - 1\right) = -2\int \sqrt{pq} + 2 = h^2(P,Q)$.
2. $KL(P,Q) = \int_{pq>0} p \log \frac{p}{q} = \int_{pq>0} p \log \left(\frac{p}{q} - 1 + 1\right) \leq \int \frac{p^2}{q} - 1 = \chi^2(P,Q)$.

□

Lemma

$$h^2(P^n, Q^n) = 2\left(1 - \left(1 - \frac{h^2(P,Q)}{2}\right)^n\right) \leq nh^2(P,Q).$$

Proof.

1. $1 - \frac{1}{2}h^2(P^n, Q^n) = \int \sqrt{p^n q^n} = \left(\int \sqrt{pq}\right)^n = \left(1 - \frac{h^2(P,Q)}{2}\right)^n.$
2. Use $(1-a)^n \geq 1 - an$ for $0 \leq a < 1$.

□

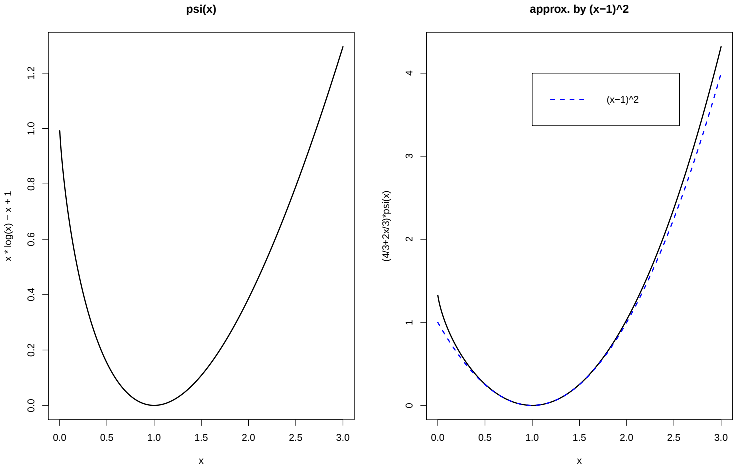
Lemma (Pinsker)

$$V(P,Q) \leq \sqrt{\frac{KL(P,Q)}{2}}.$$

Proof.

1. Let $\psi(x) = x\log x - x + 1$ for $x \geq 0$ where $0\log 0 := 0$. Note that $\psi(0) = 1$, $\psi(1) = 0$, $\psi'(1) = 0$, $\psi''(x) = 1/x \geq 0$ and $\psi(x) \geq 0$ for all $x \geq 0$.
2. Use $\left(\frac{4}{3} + \frac{2}{3}x\right)\psi(x) \geq (x-1)^2$ where $x \geq 0$. If $P \ll Q$, $V(P,Q) = \frac{1}{2}\int |p-q| = \frac{1}{2}\int_{q>0} |\frac{p}{q} - 1|q \leq$ $\frac{1}{2}\int_{q>0} q\sqrt{\left(\frac{4}{3} + \frac{2p}{3q}\right)\psi\left(\frac{p}{q}\right)} \leq \frac{1}{2}\sqrt{\int\left(\frac{4q}{3} + \frac{2p}{3}\right)}\sqrt{\int_{q>0} q\psi\left(\frac{p}{q}\right)} =$ $\sqrt{\frac{1}{2}\int_{pq>0} p\log\frac{p}{q}} = \sqrt{\frac{KL(P,Q)}{2}}$.

□


psi(x)
x * log(x) − x + 1
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0.0 0.5 1.0 1.5 2.0 2.5 3.0
x
approx. by (x−1)^2
(4/3+2x/3)*psi(x)
0
1
2
3
4
(x−1)^2
0.0 0.5 1.0 1.5 2.0 2.5 3.0
x

Lemma

$V(P,Q) \leq 1 - \frac{1}{2}\exp(-KL(P,Q))$.

Proof.

1. W.l.o.g., $KL(P,Q) < \infty$ (i.e. $P \ll Q$).
2. $(\int \sqrt{pq})^2 = \exp\left(2\log \int_{pq>0} \sqrt{pq}\right) = \exp\left(2\log \int_{pq>0} p\sqrt{\frac{q}{p}}\right) \geq \exp\left(2\int_{pq>0} p \log \sqrt{\frac{q}{p}}\right) = \exp\Big(-KL(P,Q)\Big)$.
3. Use $2\int(p\wedge q) \geq \left(2 - \int(p\wedge q)\right)\int(p\wedge q) \geq \left(\int\sqrt{(p\wedge q)(p\vee q)}\right)^2 = \left(\int\sqrt{pq}\right)^2 \geq \exp\Big(-KL(P,Q)\Big)$ and $2\int(p\wedge q) = 2(1 - V(P,Q))$.

□

Theorem

Suppose Π *is a distribution on* Θ *such that*

$$\int R(\theta, \delta_\Pi) d\Pi(\theta) = \sup_\theta R(\theta, \delta_\Pi).$$

Then

1. δ_Π *is minimax.*
2. Π *is least favourable.*

Proof.

(i) Let δ be any other procedure. Then $\sup_\theta R(\theta, \delta) \geq \int R(\theta, \delta) d\Pi(\theta) \geq \int R(\theta, \delta_\Pi) d\Pi(\theta) = \sup_\theta R(\theta, \delta_\Pi)$.
(ii) Let Π' be some other distribution of θ. Then $\int R(\theta, \delta_{\Pi'}) d\Pi'(\theta) \leq \int R(\theta, \delta_\Pi) d\Pi'(\theta) \leq \sup_\theta R(\theta, \delta_\Pi) = \int R(\theta, \delta_\Pi) d\Pi(\theta)$. □

Theorem

Suppose $\{\Pi_n\}$ *is a sequence of prior distributions with Bayes risks* $\int R(\theta, \delta_n) d\Pi_n(\theta) =: r_{\Pi_n}$*, and that* δ *is an estimator for which* $\sup_\theta R(\theta, \delta) = \lim_{n\to\infty} r_{\Pi_n}$*. Then*

1. δ *is minimax.*
2. *the sequence* $\{\Pi_n\}$ *is least favourable.*

Proof.

(i) Suppose δ' is any other estimator. Then $\sup_\theta R(\theta, \delta') \geq \int R(\theta, \delta') d\Pi_n(\theta) \geq r_{\Pi_n}$, and this holds for every n. Hence $\sup_\theta R(\theta, \delta') \geq \sup_\theta R(\theta, \delta)$. Thus δ is minimax.
(ii) If Π is any distribution, then
$\int R(\theta, \delta_\Pi) d\Pi(\theta) \leq \int R(\theta, \delta) d\Pi(\theta) \leq \sup_\theta R(\theta, \delta) = \lim_{n\to\infty} r_{\Pi_n}$.
The claim holds by definition. □