Minimax bounds III

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 $\hat{\theta}_{mm}$,

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

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 $\hat{\theta}$,

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

Assouad's Lemma

Construct

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

such that

1. $\inf_{\xi\in\Theta}\left(L(\xi,\theta_\alpha) + L(\xi,\theta_\beta)\right) \geq \delta\|\alpha-\beta\|_0\ \forall\ \alpha,\beta \in \{0,1\}^m$,
2. $||\mathbb{P}_{\theta_\alpha} \wedge \mathbb{P}_{\theta_\beta}||_1 \geq c > 0$ if $\|\alpha-\beta\|_0 = 1$.

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

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

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

Fano's Lemma

Construct

$$F := \{\theta_j, j \in J\} \subseteq \Theta$$

where $|F| = M$ satisfying the following: suppose $\forall\ \theta_j, \theta_{j'} \in F$,

1. (loss cond.) $L(\theta_j, \theta_{j'}) \geq \delta$
2. (testing cond.) $KL(\mathbb{P}_{\theta_j}, \mathbb{P}_{\theta_{j'}}) \leq \epsilon$.

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

$$\sup_{\theta \in \Theta} \mathbb{E}_\theta L(\theta, \hat{\theta}) \geq \frac{\delta}{2}\left(1 - \frac{\epsilon + \log 2}{\log M}\right).$$

Yang and Barron (1999)

Construct

$$F := \{\theta_j, j \in J\} \subseteq \Theta$$

where $|F| = M$ satisfying the following: suppose $\forall\ \theta_j, \theta_{j'} \in F$,

1. $L(\theta_j, \theta_{j'}) \geq \delta$.

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

$$\sup_{\theta \in \Theta} \mathbb{E}_\theta L(\theta, \hat{\theta}) \geq \frac{\delta}{2}\left(1 - \frac{\log N_c(\epsilon^2, \mathcal{P}_\Theta, KL) + \epsilon^2 + \log 2}{\log M}\right).$$

Possible questions

- For Le Cam, one vs. one $\Rightarrow$ one vs. multiple? Or, multiple vs. multiple?
- Assouad + Le Cam?
- For Fano, KL divergence $\Rightarrow$ more general distance?
- Non-finite sub-parameter space?

Let $co(\mathcal{P})$ be the convex hull of $\mathcal{P}$, and let $\mathcal{P}_i = \{\mathbb{P}_\theta : \theta \in A_i\}$.

Le Cam's Lemma II

Construct

$$A_0 \subseteq \Theta, \quad A_1 \subseteq \Theta$$

such that for $\theta_0 \in A_0, \theta_1 \in A_1$,

1. $\inf_{\xi \in \Theta} L(\xi, \theta_0) + L(\xi, \theta_1) \geq \delta$
2. $\sup_{\mathbb{Q}_i \in co(\mathcal{P}_i)} \|\mathbb{Q}_0 \wedge \mathbb{Q}_1\| > c > 0$

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

$$\sup_{\theta \in \Theta} \mathbb{E}_\theta L(\hat{\theta}, \theta) \geq \frac{c\delta}{2}.$$

Remarks

- When $|A_0| = 1$, $|A_1| = 1$: Le Cam I (one versus one testing)
- When $|A_0| = 1$ and $|A_1| = m$: One versus multiple testing
- When $|A_0| = m_1$ and $|A_1| = m_2$: Multiple versus multiple testing

Proof: Let $w_i(\theta) \geq 0$ and $\sum_{\theta \in A_i} w_i(\theta) = 1$ for $i = 0, 1$.

$$
\begin{aligned}
2R(\Theta, \hat{\theta}) &= \sup_{\theta \in \Theta} \mathbb{E}_\theta 2L(\hat{\theta}, \theta) \\
&\geq \max_{\theta_0 \in A_0} \mathbb{E}_{\theta_0} L(\hat{\theta}, \theta_0) + \max_{\theta_1 \in A_1} \mathbb{E}_{\theta_1} L(\hat{\theta}, \theta_1) \\
&\geq \sum_{\theta_0 \in A_0} w_0(\theta_0) \mathbb{E}_{\theta_0} L(\hat{\theta}, \theta_0) + \sum_{\theta_1 \in A_1} w_1(\theta_1) \mathbb{E}_{\theta_1} L(\hat{\theta}, \theta_1) \\
&= \delta \left(\sum_{\theta_0 \in A_0} w_0(\theta_0) \mathbb{E}_{\theta_0} \left(\frac{L(\hat{\theta}, \theta_0)}{\delta} \right) + \sum_{\theta_1 \in A_1} w_1(\theta_1) \mathbb{E}_{\theta_1} \left(\frac{L(\hat{\theta}, \theta_1)}{\delta} \right) \right) \\
&\geq \delta \inf_{\{f_0, f_1 \geq 0, f_0 + f_1 \geq 1\}} \left\{ \left(\sum_{\theta_0 \in A_0} w_0(\theta_0) \mathbb{E}_{\theta_0} f_0 + \sum_{\theta_1 \in A_1} w_1(\theta_1) \mathbb{E}_{\theta_1} f_1 \right) \right\} \\
&\geq \delta \|\bar{\mathbb{P}}_0 \wedge \bar{\mathbb{P}}_1\|_1,
\end{aligned}
$$

where $\bar{\mathbb{P}}_i := \sum_{\theta_i \in A_i} w_i(\theta_i) \mathbb{P}_{\theta_i}$ for $i = 0, 1$. Since the above inequality is satisfied for any prob. measure w_i on A_i, we can take the supremum over all convex hull.

Examples

(1) Gaussian sequence model

Let $Y_j = \mu_j + \xi_j/\sqrt{n}$ for $j = 1, \dots, n$ where $\xi_j \sim N(0,1)$. Suppose $\Theta = \{\theta = (\mu_1, \dots, \mu_n) : \|\theta\|_0 \le 1, \|\theta\|_2 \le C\}$. Then

$$\sup_{\theta \in \Theta} \mathbb{E}_\theta \|\hat{\theta} - \theta\|_2^2 \ge c \frac{\log n}{n}.$$

Remarks. 1) In Minimax I, when we assume $\sum_k k^{2s} \mu_k^2 \le M$, we had the lower bound $n^{-2s/(1+2s)}$ using Assouad I.
2) What happens if we assume $\|\theta\|_0 \le k$?

Using Le Cam I

Use one to one test.

- $L(\xi, \theta_0) = \|\xi - \theta_0\|_2^2$
- $\theta_0 = (0, \ldots, 0)$ and $\theta_1 = (\tilde{\delta}, 0, \ldots, 0)$.
- (loss cond.) $L(\xi, \theta_0) + L(\xi, \theta_1) \geq \frac{1}{2} L(\theta_0, \theta_1) = \frac{\tilde{\delta}^2}{2} =: \delta$.
- $\mathbb{P}_{\theta_0} = \otimes_{i=1}^n N(0, 1/n)$, $\mathbb{P}_{\theta_1} = N(\tilde{\delta}, 1/n) \otimes_{i=2}^n N(0, 1/n)$.
- (testing cond.) $\|\mathbb{P}_{\theta_0} \wedge \mathbb{P}_{\theta_1}\|_1 = 1 - TV(\mathbb{P}_{\theta_0}, \mathbb{P}_{\theta_1}) \geq 1 - \sqrt{\chi^2(\mathbb{P}_{\theta_0}, \mathbb{P}_{\theta_1})} = 1 - \sqrt{\int \frac{\phi_{1/n}(x_1 - \tilde{\delta})^2}{\phi_{1/n}(x_1)} - 1} = 1 - \sqrt{\exp(n\tilde{\delta}^2) - 1} \geq c$ if $\tilde{\delta}^2 \sim 1/n$.
- This gives n^{-1} rate, which is not enough!

Using Le Cam II

Use one versus multiple test.

- Let

$$A_0 = \{\theta_0\} = \{(0, \ldots, 0)\}$$
$$A_1 = \{\delta(1, 0, \ldots 0), \delta(0, 1, 0, \ldots, 0), \ldots, \delta(0, \ldots, 0, 1)\}$$

- For any $\theta_1 \in A_1$, $\|\theta_0 - \theta_1\|_2^2 = \delta^2$.
- $\mathbb{Q}_0 = \otimes_{i=1}^n N(0, \frac{1}{n})$
- $\mathbb{Q}_1 = \frac{1}{n} \sum_{j=1}^n \left(\left(\prod_{i \neq j} N(0, \frac{1}{n}) \right) \otimes N(\delta, \frac{1}{n}) \right)$.
- Suffices to show $\chi^2(\mathbb{Q}_0, \mathbb{Q}_1) \leq c$.

- Note that

$$\begin{aligned}
\chi^2(\mathbb{Q}_0, \mathbb{Q}_1) &= \int \frac{\left(\frac{1}{n}\sum_{j=1}^n \prod_{i\neq j} \phi_{1/n}(x_i)\phi_{1/n}(x_j - \delta)\right)^2}{\prod_{i=1}^n \phi_{1/n}(x_i)} - 1 \\
&= \int \frac{\frac{1}{n^2}\sum_{j=1}^n \prod_{i\neq j} \phi_{1/n}(x_i)\phi_{1/n}(x_j - \delta)^2}{\phi_{1/n}(x_j)} - \frac{1}{n} \\
&= \frac{1}{n}\Big(\int \frac{\phi_{1/n}(x_j - \delta)^2}{\phi_{1/n}(x_j)} - 1\Big) \\
&= \frac{1}{n}\Big(\exp(n\delta^2) - 1\Big).
\end{aligned}$$

- The above can be arbitrarily small by choosing $\delta = \sqrt{c \log n / n}$.
- This gives $\log n/n$ rate.

(2) Estimation of a functional (Bickel & Ritov, 1988)

Suppose we have i.i.d. sample $X_1, \ldots, X_n$ from a density $\theta \in \Theta$ where

$$\Theta := \{\theta \text{ on } [0,1],\ 0 < c_0 \leq \theta(x) \leq c_1 < \infty, |\theta^{(2)}(x)| \leq c_2 < \infty\}.$$

and let $T(\theta) = \int_0^1 (\theta'(x))^2 dx$ for $\theta \in \Theta$. Then

$$\sup_{\theta \in \Theta} \mathbb{E}_\theta |\hat{T} - T(\theta)| \geq c n^{-4/9}.$$

cf. Use Le Cam II where the first condition is changed by using $T(\theta)$ instead of θ.

- $\theta_0(x) = \mathbb{1}_{\{x\in[0,1]\}}$, $T(\theta_0) = 0$.
- g: twice differentiable function on $[0,1]$ s.t. $\int_0^1 g(x)dx = 0$, $\int_0^1 g^2(x)dx = a > 0$, $\int (g'(x))^2 dx = b > 0$.
- Divide $[0,1]$ into m disjoint intervals, and for $j = 1, \ldots, m$, let

$$g_j(x) = cm^{-2} g\Big(m\big(x - \frac{j-1}{m}\big)\Big)$$

  where c is small enough so that $|g_j| < 1$.
- Construct

$$\Theta_1 = \{\theta_\alpha = \theta_0 + \sum_{j=1}^m \alpha_j g_j : \alpha \in \{-1,1\}^m\} \subseteq \Theta,$$

  so that for $\theta_\alpha \in \Theta_1$, $T(\theta_\alpha) = \sum_{j=1}^m \int_0^1 \left(g_j'\right)^2 = c^2 b m^{-2}$.

- $\delta \sim \inf_\alpha |T(\theta_0) - T(\theta_\alpha)| \sim m^{-2}$.
- $\mathbb{Q}_0 = \theta_0^n$, $\mathbb{Q}_1 = \frac{1}{2^m} \sum_\alpha \theta_\alpha^n$
- It suffices to show that $\chi^2(\mathbb{Q}_0, \mathbb{Q}_1) \leq c$.

Crude bound: $\int \theta_\alpha \theta_\beta = 1 + \sum_{j=1}^m \alpha_j \beta_j \int g_j^2 \leq 1 + c^2 m^{-4} a$. Then

$$\begin{aligned}\chi^2(\mathbb{Q}_0, \mathbb{Q}_1) &= \int_0^1 \frac{\left(\frac{1}{2^m} \sum_\alpha \prod_{i=1}^n \theta_\alpha(x_i)\right)^2}{\prod_{i=1}^n \theta_0(x_i)} - 1 \\ &= \frac{1}{2^{2m}} \int_0^1 \Big(\sum_\alpha \prod_{i=1}^n \theta_\alpha(x_i)\Big)\Big(\sum_\beta \prod_{i=1}^n \theta_\beta(x_i)\Big) - 1 \\ &\leq \left(1 + c^2 m^{-4} a\right)^n - 1 \leq \exp(c^2 a n m^{-4}) - 1 \leq c,\end{aligned}$$

for $m = \tilde{c} n^{1/4}$, where we use $(1 + x/n)^n \leq e^x$ for $|x| \leq n$.

- But then we would have $n^{-1/2}$ rate, which is not enough!

- Better bound by careful calculation.
- Treat $P(\alpha_j = 1) = P(\alpha_j = -1) = 1/2$. Then $\alpha_j \beta_j =: \gamma_j$ has the same distribution. Define $\Gamma := \sum_j \gamma_j$, then $\mathbb{E}\Gamma^k = 0$ for all odd k, and $\mathbb{E}\Gamma^2 = m$ and by denoting $\int g_j^2 = c^2 a m^{-5} =: u$,

$$\chi^2(\mathbb{Q}_0, \mathbb{Q}_1) = \mathbb{E}\left(1 + u \sum_j \gamma_j\right)^n$$
$$= \sum_{k=0}^{n} \binom{n}{k} u^k \mathbb{E}\left(\Gamma^k\right)$$
$$\lesssim \binom{n}{2} u^2 \mathbb{E}\Gamma^2 = \binom{n}{2} c^4 a^2 m^{-10} m \sim n^2 m^{-9}.$$

- Take $m \sim n^{2/9}$, then the lower bound is obtained as $m^{-2} \sim n^{-4/9}$.

Remarks

- Difficult if we use Assouad since we take squares with $\alpha \in \{-1, 1\}^m$!
- If we take $\alpha \in \{0, 1\}^m$, essentially we would calculate the same quantity in our *crude bound* so that we get $n^{-1/2}$ rate via Assouad (check).

Multiple multiple testing example

(3) Key idea in manifold lower bound (Kim & Zhou, 2015)

Let $\theta \in \Theta = \{0, 1/m^2\}^{2m} \subseteq \mathbb{R}^{2m}$ and let $X_i \in \mathbb{R}^2$ and for $i = 1, \ldots, n$, we let $P(X_i = x_j) = 1/(2m)$ where $x_j = (j/(2m), \theta_j)$ where $j = 1, \ldots, 2m$. Using the maximum loss $L(\theta, \hat{\theta}) = \max_j |\theta_j - \hat{\theta}_j|$,

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_\theta L(\theta, \hat{\theta}) \gtrsim \left(\frac{\log n}{n}\right)^2.$$

cf. In estimating the support of $U[0, \theta]$ where $\theta \in (0, 1]$, the minimax rate is of order $1/n$ under the L_1 loss.

Two points argument

- Let $\theta = (0, \ldots, 0)$ and $\theta' = (1/m^2, 0, \ldots, 0)$.
- $\delta \sim 1/m^2$.
- Check $P_\theta \wedge P_{\theta'}$ puts mass $1/(2m)$ on $(2m-1)$ points $(2/(2m), 0), (3/(2m), 0), \ldots, (1, 0)$. Thus $\|P_\theta \wedge P_{\theta'}\|_1 = 1 - 1/(2m)$. We need to take $m \sim n$ to show $\|P_\theta^n \wedge P_{\theta'}^n\|_1 \geq (\|P_\theta \wedge P_{\theta'}\|_1)^n > 0$.
- This gives the lower bound $1/n^2$ (not enough).

Using Le Cam II

- Let A_0 be the set of θ's s.t. m locations are zeros and m locations are $1/m^2$ so that $|A_0| = \binom{2m}{m}$.
- Let A_1 be the set of θ's s.t. $m+1$(or $m-1$) locations are zeros and $m-1$(or $m+1$) locations are $1/m^2$.
- $\delta \sim \inf_{\theta \in A_0, \theta' \in A_1} L(\theta, \theta') = 1/m^2$.
- Suffices to prove

$$\left\| \frac{1}{|A_0|} \sum_{\theta} \mathbb{P}_\theta \wedge \frac{1}{|A_1|} \sum_{\theta'} \mathbb{P}_{\theta'} \right\|_1 \geq 0.$$

  Quite involved, but provable where $m \sim n/\log n$.
- This gives the lower bound $\left(\frac{\log n}{n}\right)^2$.

Idea

- Assouad I in one direction & Le Cam in another direction
- Parameter space: for a given positive integer m and a finite set $B \subseteq \mathbb{R}^p \backslash \{\mathbf{0}_{1\times p}\}$, let $\Gamma = \{0,1\}^m$ and $\Lambda \subseteq B^m$. Define

$$A = \{\alpha = (\gamma, \lambda) : \gamma \in \Gamma, \lambda \in \Lambda\}.$$

- Projection: For $\alpha = (\gamma, \lambda) \in A$, denote the projection of α to Γ by $\gamma(\alpha) = \gamma$, to Λ by $\lambda(\alpha) = \lambda$.
- $\gamma_i(\alpha)$: ith coordinate of the first component of α.

Notation

- Average: For a given $a \in \{0, 1\}$ and $1 \leq i \leq m$, denote

$$A_{i,a} = \{\alpha \in A : \gamma_i(\alpha) = a\}.$$

Define mixture distribution $\bar{\mathbb{P}}_{i,a}$ by

$$\bar{\mathbb{P}}_{i,a} = \frac{1}{|A_{i,a}|} \sum_{\alpha \in A_{i,a}} \mathbb{P}_{\theta_\alpha}.$$

Assouad's Lemma II (Cai & Zhou, 2012)

Consider $A = \{\alpha = (\gamma, \lambda) : \gamma \in \Gamma, \lambda \in \Lambda\}$. Suppose

1. $L(\theta_\alpha, \theta_\beta) \geq \delta \|\gamma(\alpha) - \gamma(\beta)\|_0 \ \forall \alpha, \beta \in A$,
2. $\min_{1 \leq k \leq m} \|\bar{\mathbb{P}}_{k,0} \wedge \bar{\mathbb{P}}_{k,1}\|_1 \geq c$, where $\bar{\mathbb{P}}_{i,a} = \frac{1}{|A_{i,a}|} \sum_{\alpha \in A_{i,a}} \mathbb{P}_{\theta_\alpha}$, $A_{i,a} = \{\alpha \in A : \gamma_i(\alpha) = a\}$.

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

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

Remarks

- Effectively treat "two-directional" problems.
- Simultaneous application of Le Cam's in one direction and Assouad's in another.
- Special cases
  - Assouad's Lemma I considers Γ only.
  - Le Cam's Lemma I considers Λ with $m = 1$.

Proof of Assouad II:

$$R(\Theta,\hat\theta) \ge \frac{1}{2}\max_{\alpha\in A}\mathbb{E}_{\theta_\alpha}\Big(L(\hat\theta,\theta_\alpha)+L(\hat\theta,\theta_{\hat\alpha})\Big)$$

$$\ge \frac{\delta}{2|A|}\sum_\alpha \mathbb{E}_{\theta_\alpha} L(\theta_\alpha,\theta_{\hat\alpha})$$

$$\ge \frac{\delta}{2\cdot 2^m|\Lambda|}\sum_\alpha\sum_{k=1}^m \mathbb{E}_{\theta_\alpha}\mathbb{1}_{\{\gamma(\hat\alpha)_k\neq\gamma(\alpha)_k\}} \text{ by loss cond.}$$

$$= \frac{\delta}{4}\sum_{k=1}^m\frac{1}{2^{m-1}|\Lambda|}\left(\sum_{\{\alpha:\gamma(\alpha)_k=0\}}\mathbb{E}_\alpha\{\gamma(\hat\alpha)_k\neq 0\}+\sum_{\{\alpha:\gamma(\alpha)_k=1\}}\mathbb{E}_\alpha\{\gamma(\hat\alpha)_k\neq 1\}\right)$$

$$\ge \frac{\delta}{4} m \min_k \|\bar{\mathbb{P}}_{k,0}\wedge\bar{\mathbb{P}}_{k,1}\|_1,$$

where $\bar{\mathbb{P}}_{k,a}=\frac{1}{|A_{i,a}|}\sum_{\alpha\in A_{i,a}}\mathbb{P}_{\theta_\alpha}$ for $a=0,1$ with $A_{i,a}=\{\alpha\in A:\gamma_i(\alpha)=a\}$. By testing condition, the proof is complete.

Example: covariance matrix estimation

- Let $Z = (z_{ij})$ be $m \times n$ matrix.
- Matrix ℓ_q norm: for $1 \leq q \leq \infty$

$$|||Z|||_q = \max_{\|z\|_q = 1} \|Az\|_q.$$

- Special cases:
  - Spectral norm $|||Z|||_2 =: |||Z|||$: square root of the largest singular value of Z^*Z
  - Matrix ℓ_1 norm $|||Z|||_1 = \max_j \sum_i |z_{i,j}|$: maximum absolute column sum
- $\mathcal{G}_0(k)$ = matrices with at most $k+1$ nonzero elements on each row/column.

(4) Sparse covariance matrix estimation (Cai & Zhou, 2012)

Suppose we have an i.i.d. sample X_i from p variate Gaussian with a covariance matrix Σ. Assuming a sparse Σ (at most $k(\leq Mn^{1/2}(\log p)^{-3/2})+1$ nonzero elements on each row and column),

$$\sup_{\Sigma \in \mathcal{G}_0(k)} \mathbb{E}_\Sigma \left|\left|\left| \hat{\Sigma} - \Sigma \right|\right|\right|^2 \geq C \left(k^2 \frac{\log p}{n} \right).$$

Using Le Cam I

- Construct $\Sigma_\alpha = I_{p\times p} + \epsilon Z_\alpha$ for $\alpha \in \{0,1\}$, where

$$Z_\alpha = \begin{pmatrix} 0 & \alpha & \cdots & \alpha & 0 & \cdots \\ \alpha & 0 & \cdots & 0 & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \alpha & 0 & \vdots & 0 & 0 & \cdots \\ 0 & 0 & \vdots & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

where the number of α in the first row (and column) is k.

Using Le Cam I

- (loss cond.) $\|\|\Sigma_1 - \Sigma_0\|\|^2 = \epsilon^2 \|\|Z_1\|\|^2 \sim k\epsilon^2$.
- Useful facts: Let g_i be the density function of $N_p(0, \Sigma_i)$ for $i = 0, 1, 2$. Then

$$\int \frac{g_1 g_2}{g_0} = \left[\det\left(I - \Sigma_0^{-2}(\Sigma_1 - \Sigma_0)(\Sigma_2 - \Sigma_0)\right)\right]^{-1/2}.$$

- (testing cond.) $\chi^2(\mathbb{P}_0, \mathbb{P}_1) = \left[\det\left(I - \epsilon^2 Z_1^2\right)\right]^{-n/2} - 1 \sim (1 - k\epsilon^2)^{-n/2} - 1$ gives $k\epsilon^2 \sim 1/n$ (not enough!).

Using Assouad I

- Construct $\Sigma_\alpha = I_{p\times p} + \epsilon Z_\alpha$ for $\alpha \in \{0,1\}^m$, where

$$Z_\alpha = \begin{pmatrix} 0 & \alpha_1 & \cdots & \alpha_1 & 0 & \cdots \\ \alpha_1 & 0 & \alpha_2 & \cdots & \alpha_2 & \cdots \\ \vdots & \alpha_2 & \ddots & \vdots & \vdots & \vdots \\ \alpha_1 & \vdots & \vdots & 0 & 0 & \cdots \\ 0 & \alpha_2 & \vdots & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

where the number of α_1 in the first k row (and column) is $k/2$. Let $m \sim k$.

Using Assouad I

- (loss cond.) $\|\|\Sigma_\alpha - \Sigma_\beta\|\|^2 = \epsilon^2 \|\|Z_\alpha - Z_\beta\|\|^2 \sim k\epsilon^2 \|\alpha - \beta\|_0$
- (testing cond.) For α, β s.t. $\|\alpha - \beta\|_0 = 1$,

$$\begin{aligned} TV^2(\mathbb{Q}_\alpha, \mathbb{Q}_\beta) &\le 2KL(\mathbb{Q}_\alpha, \mathbb{Q}_\beta) \quad \text{Pinsker ineq.} \\ &= n\left[\operatorname{tr}(\Sigma_\alpha \Sigma_\beta^{-1}) - \log\det(\Sigma_\alpha \Sigma_\beta^{-1}) - p\right]. \end{aligned}$$

Write $\Sigma_\alpha - \Sigma_\beta =: D_1$, then

$$TV(\mathbb{Q}_\alpha, \mathbb{Q}_\beta) \le n\left[\operatorname{tr}(D_1 \Sigma_\beta^{-1}) - \log\det\left(D_1 \Sigma_\beta^{-1} + I_{p\times p}\right)\right] \sim nk\epsilon^2.$$

- Thus, the lower bound is k/n (not enough!).

Using Assouad II

- For convenience, let p be even and let $m = p/2$.
- Consider the index $\alpha = (\gamma, \lambda) = \begin{pmatrix} \gamma_1 \cdot \lambda_1 \\ \gamma_2 \cdot \lambda_2 \\ \vdots \\ \gamma_m \cdot \lambda_m \end{pmatrix}$ where $\gamma_i \in \{0, 1\}$ and $\lambda_i \in \{0, 1\}^m$ where each λ_i has k nonzero elements.
- Let Λ be the set of all elements λ so that the each column sum of λ is less than or equal to k.
- Define $Z_\alpha = \alpha$ where $\alpha = (\gamma, \lambda)$ with $\gamma \in (0, 1)^m$ and $\lambda \in \Lambda$.

Using Assouad II

- Construct $\Sigma_\alpha = I_p + \epsilon \begin{pmatrix} 0_{p/2 \times p/2} & Z_\alpha \\ Z_\alpha^T & 0_{p/2 \times p/2} \end{pmatrix}_{p \times p}$.
- One possible Z_α would be like this:

$$\begin{pmatrix} 0 & \gamma_1 & \cdots & 0 & \gamma_1 & \cdots \\ \gamma_2 & 0 & \cdots & \gamma_2 & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \gamma_j & \vdots & 0 & \vdots & \cdots \\ \gamma_{j+1} & \gamma_{j+1} & \vdots & 0 & \gamma_{j+1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Using Assouad II

- (loss cond.)
  $|||\Sigma_\alpha - \Sigma_\beta|||^2 = \epsilon^2 |||Z_\alpha - Z_\beta|||^2 \sim \frac{k^2\epsilon^2}{p} H(\gamma(\alpha), \gamma(\beta))$.
  Indeed, let $\nu = \mathbb{1}_{\{p/2+1 \le i \le p\}} \in \{0,1\}^p$, then
  $\|(Z_\alpha - Z_\beta)\nu\|_2^2 \ge H(\gamma(\alpha), \gamma(\beta))k^2$. Since $\|\nu\|_2^2 = m = p/2$,

$$|||\Sigma_\alpha - \Sigma_\beta|||^2 \ge \epsilon^2 \frac{\|(Z_\alpha - Z_\beta)\nu\|_2^2}{\|\nu\|_2^2} \ge \frac{H(\gamma(\alpha), \gamma(\beta))(k\epsilon)^2}{p/2}.$$

- (testing cond.) By taking $\epsilon^2 \sim \log p/n$, we can bound the affinity bounded away from zero. This gives the lower bound $\frac{k^2 \log p}{n}$. For more details, see the proof of Lemma 6 of Cai and Zhou(2012).

Preliminaries

- f-divergence: for a convex function $f : [0, \infty) \to \mathbb{R}$ satisfying $f(1) = 0$,

$$D_f(P\|Q) := \int f\Big(\frac{dP}{dQ}\Big)dQ.$$

  - $f(x) = x \log x$ gives KL
  - $f(x) = x^2 - 1$ gives Chi-squared
  - $f(x) = 1 - \sqrt{x}$ gives Hellinger (times half)
  - $f(x) = |x - 1|/2$ gives TV.
  - $f(x) = x^l - 1$ ($l > 1$, $l \neq 2$).
- Let $R(\Theta, \hat{\theta}) = \sup_{\theta\in\Theta} \mathbb{E}L(\theta, \hat{\theta}) = \sup_{\theta\in\Theta} \mathbb{E}\ell(\rho(\theta, \hat{\theta}))$, where $\ell : [0, \infty) \to [0, \infty)$ is a nondereasing function and ρ is a metric on Θ.

Theorem: Fano's Lemma II (Guntuboyina, 2011)

Let

- $N_p(\eta)$: a lower bound on the η-packing number of the metric space (Θ, ρ)
- $N_{c,f}(\epsilon, \mathcal{P}_\Theta)$: an upper bound on the ϵ covering number of the space $\{\mathcal{P}_\Theta\}$ (using f-divergence distance).

Then

$$R(\Theta, \hat{\theta}) \geq \sup_{\eta>0, \epsilon>0} \ell(\eta/2)\,(1 - *)\,,$$

where $*$ is one of the followings:

1. $\frac{\log 2 + \log N_{c,KL}(\epsilon, \mathcal{P}_\Theta) + \epsilon^2}{\log N_p(\eta)}$
2. $\frac{1}{N_p(\eta)} + \sqrt{\frac{(1+\epsilon^2) N_{c,C}(\epsilon, \mathcal{P}_\Theta)}{N_p(\eta)}}$
3. for $l > 1, l \neq 2$, $\left(\frac{1}{N_p(\eta)^{l-1}} + \frac{(1+\epsilon^2) N_{c,l}(\epsilon, \mathcal{P}_\Theta)^{l-1}}{N_p(\eta)^{l-1}}\right)^{1/l}$.

Remarks

- Yang and Barron(1999) gives suboptimal lower bound on the minimax risk in many parametric problems.
- Finite dimensional situations can be dealt with the current approaches.
- This implies that global packing and covering characteristics are enough to obtain optimal minimax lower bounds.

(5) Example: location normal (revisited)

Let $X_1, \ldots, X_n \sim N(\theta, 1)$ where $\Theta = [a, b] \in \mathbb{R}$ where $-\infty < a < b < \infty$. Prove the minimax lower bound under the squared error loss by YB method (i.e. Fano's II (1)) and Fano's II (2).

- Note that $KL(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) = \frac{n}{2}(\theta - \theta')^2$ and $\chi^2(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) = \exp(n(\theta - \theta')^2) - 1$. Thus

$$N_{c,KL}(\epsilon, \mathcal{P}_\Theta) = \frac{c\sqrt{n}}{\epsilon}, \; N_{c,C}(\epsilon, \mathcal{P}_\Theta) = \frac{c\sqrt{n}}{\sqrt{\log(1+\epsilon^2)}}$$

- Also since we use the Euclidean ρ,

$$N_p(\eta, \Theta) = \frac{c'}{\eta}$$

(5) Example: location normal (revisited)

- (1) gives the lower bound

$$\sup_{\eta \leq \eta_0, \epsilon \leq \epsilon_0} \frac{\eta^2}{4} \left(1 - \frac{\log 2 + \log(c\sqrt{n}/\epsilon) + \epsilon^2}{\log(c'/\eta)} \right).$$

Optimize with respect to ϵ by taking $\epsilon = \epsilon_1 \in [0, \epsilon_0]$, then we need to take $\eta \sim 1/\sqrt{n}$, where the second term is of order $1/(\log n)$.

- (2) gives

$$\sup_{\eta \leq \eta_0, \epsilon \leq \epsilon_0} \frac{\eta^2}{4} \left(1 - \frac{\eta}{c'} - \sqrt{\eta\sqrt{n}} \sqrt{\frac{c(1+\epsilon^2)}{c'\sqrt{\log(1+\epsilon^2)}}} \right).$$

Optimize again by taking $\epsilon = \epsilon_0$, then we need to take $\eta = c_1/\sqrt{n}$, where the second term is bounded away from zero, which gives the lower bound $\sim 1/n$ by taking c_1 sufficiently small.

Proof of Theorem: Start with the same definition of $Z = \arg\min_{j \in J} L(\hat{\theta}, \theta_j) =: \tilde{j}$, and use (as in the proof of Fano I)

$$R(\Theta, \hat{\theta}) \geq \max_{j \in J} \mathbb{E}_{\theta_j} L(\theta_j, \hat{\theta}) \geq \ell(\frac{\eta}{2}) \max_{j \in J} \mathbb{P}_j(Z \neq j) \geq \ell(\frac{\eta}{2}) \sum_{j \in J} w_j \mathbb{P}_j(Z \neq j),$$

by bounding the max by some average, where $w_j = w\{j\}$ is the prior on J. For any $\tilde{j} \in J$, we have

$$\begin{aligned}
\sum_{j \in J} w_j \mathbb{P}_j(\tilde{j} \neq j) &= 1 - \sum_{j \in J} w_j \mathbb{P}_j(\tilde{j} = j) \\
&= 1 - \sum_{j \in J} w_j \int_{\mathcal{X}} \{\tilde{j}(x) = j\} p_j(x) d\mu(x) \\
&= 1 - \int_{\mathcal{X}} w\{\tilde{j}(x)\} p_{\tilde{j}(x)}(x) d\mu(x) \\
&\geq 1 - \int_{\mathcal{X}} \max_{j \in J} \Big\{ w_j p_j(x) \Big\} d\mu(x) =: \bar{r}_w.
\end{aligned}$$

Let $\bar{r}$ be $\bar{r}_w$ when w is the uniform on $\{1, \ldots, N\}$ where $|J| =: N$.

Lemma: Bounding testing risk (Guntuboyina, 2011)

Suppose $f : [0, \infty)$ is a differentiable convex function and let

$$g(a) := f(N(1-a)) + (N-1)f\Big(\frac{Na}{N-1}\Big).$$

Then,

1. for every probability measure Q on χ,

$$\sum_{j \in J} D_f(P_j \| Q) \geq g(\bar{r}).$$

2. for every $a \in [0, 1 - 1/N]$, we have

$$\bar{r}_w \geq \bar{r} \geq a + \frac{\inf_Q \sum_{j \in J} D_f(P_j \| Q) - g(a)}{g'(a)}.$$

Proof of (2) in Fano's Lemma II.

1. Let $f(x) = x^2 - 1$, then $D_f(P\|Q) = \int p^2/q - 1$. Note that $g(a) = \frac{N^3}{N-1}\left(1 - \frac{1}{N} - a\right)^2$.
2. Using Lemma (1), $\sum_{j\in J} \chi^2(P_j\|Q) \geq \frac{N^3}{N-1}\left(1 - \frac{1}{N} - \bar{r}_w\right)^2$. Since $r \leq 1 - 1/N$[1] and $N^3/(N-1) \geq N^2$,

$$\bar{r} \geq 1 - \frac{1}{N} - \frac{1}{\sqrt{N}}\sqrt{\frac{\inf_Q \sum_{j\in J} \chi^2(P_j\|Q)}{N}}.$$

[1] Indeed, $1 - \frac{1}{N}\int \max_{j\in J} p_j d\mu(x) \leq 1 - 1/N$.

Proof of (2) in Fano's Lemma II.

3. Fix $\epsilon > 0$, and use the definition of $N_{c,C}(\epsilon, \Theta)$ to get a finite set G and probability measures Q_k's s.t. $\sup_\theta \min_{k \in G} \chi^2(P_\theta, Q_k) \leq \epsilon$.
4. Check [2]

$$\frac{\inf_Q \sum_{j \in J} \chi^2(P_j, Q)}{N} \leq N_{c,C}\left(1 + \max_{j \in J} \min_{k \in G} \chi^2(P_{\theta_j}, Q_k)\right) - 1$$
$$\leq N_{c,C}\left(1 + \epsilon^2\right) - 1$$

where $N_{c,C} = |G|$. Then the proof of (2) is done.

cf) YB uses

$$\frac{\inf_Q \sum_{j \in J} KL(P_j, Q)}{N} \leq \log N_{c,KL} + \max_{j \in J} \min_{k \in G} KL(P_j, Q_k).$$

[2] See Theorem 3.1 & Example III.3 in Guntuboyina(2011).

Proof of Lemma: Fix a probability measure Q. First, assuming (1), we prove (2). That is, $\sum_{j \in J} D_f(P_j \| Q) \geq g(\bar{r})$. Since f is convex, g is also convex. Hence, for every $a \in [0, 1 - 1/N]$,

$$g(\bar{r}) \geq g(a) + g'(a)(\bar{r} - a).$$

Also

$$\frac{g'(a)}{N} = f'\Big(\frac{Na}{N-1}\Big) - f'(N(1-a)).$$

Convexity of g gives $g'(a) \leq g'(1 - 1/N) = 0$ for $a \leq 1 - 1/N$. Rearranging the above, we have proved the claim.

Proof of Lemma (1): Start with a simple inequality for nonnegative numbers $a_j, j \in J$ with $\tau := \arg\max_{j \in J}\{w_j a_j\}$.

$$\sum_{j \in J} w_j f(a_j) = w_\tau f(a_\tau) + (1 - w_\tau) \sum_{j \neq \tau} \frac{w_j}{1 - w_\tau} f(a_j).$$

By convexity,

$$\sum_{j \neq \tau} \frac{w_j}{1 - w_\tau} f(a_j) \geq f\left(\sum_{j \neq \tau} \frac{w_j}{1 - w_\tau} a_j\right) = f\left(\frac{\sum_{j \in J} w_j a_j - w_\tau a_\tau}{1 - w_\tau}\right).$$

Apply the inequality to $a_j := p_j(x)/q(x)$ for $x \in \chi$ s.t. $q(x) > 0$. Let $T(x) = \arg\max_{j \in J} w_j p_j(x)$ so that $\tau = T(x)$ and $\bar{r}_w = 1 - \int_\chi w_{T(x)} p_{T(x)} d\mu(x)$. Then

$$\sum_{j \in J} w_j f\left(\frac{p_j(x)}{q(x)}\right) \geq w_{T(x)} f\left(\frac{p_{T(x)}(x)}{q(x)}\right)$$

$$+ \left(1 - w_{T(x)}\right) f\left(\frac{\sum_{j \in J} w_j \frac{p_j(x)}{q(x)} - w_{T(x)} \frac{p_{T(x)}(x)}{q(x)}}{1 - w_{T(x)}}\right) =: (1) + (2)$$

Continue: We integrate the above with respect to Q. [3]

- LHS is $\sum_{j\in J} w_j D_f(P_j \| Q)$.
- $(1) = \int_\chi w_{T(x)} f\left(\frac{p_{T(x)}(x)}{q(x)}\right) q(x) d\mu(x)$. Let $\tilde{Q}$ be the prob measure on χ having the density $\tilde{q}(x) := w_{T(x)} q(x) / \int_\chi w_{T(x)} dQ(x) =: w_{T(x)} q(x)/W$. Then

$$
\begin{aligned}
(1) &= W \int_\chi \tilde{q}(x) f\left(\frac{p_{T(x)}(x)}{q(x)}\right) d\mu(x) \\
&\geq W f\left(\int \frac{w_{T(x)} p_{T(x)}}{W} d\mu(x)\right) \\
&= W f\left(\frac{1-\bar{r}_w}{W}\right).
\end{aligned}
$$

[3] Recall $1 - \bar{r}_w = \int_\chi w_{T(x)} p_{T(x)} d\mu(x)$.

Continue: We integrate the above with respect to Q. [4]

- LHS is $\sum_{j\in J} w_j D_f(P_j\|Q)$.
- Similarly, (2) $\geq (1-W)f\left(\frac{\bar{r}_w}{1-W}\right)$.
- When w is the uniform probability on J, then $W = 1/N$, which gives

$$\frac{1}{N}\sum_{j\in J} D_f(P_j\|Q) \geq \frac{1}{N} f(N(1-\bar{r})) + (1-\frac{1}{N}) f(\frac{\bar{r}}{1-1/N}).$$

hence,

$$\sum_{j\in J} D_f(P_j\|Q) \geq f(N(1-\bar{r})) + (N-1) f\left(\frac{N\bar{r}}{N-1}\right) = g(\bar{r}).$$

[4] Recall $1-\bar{r}_w = \int_{\chi} w_{T(x)} p_{T(x)} d\mu(x)$.

(6) Example: d-dimensional normal mean

- Let $X \sim N_d(\theta, \sigma^2 I_d)$ where $\theta \in \Theta$ with Θ being the ball of radius Γ centered at the origin. Use squared error loss and let ρ be the Euclidean distance on $\mathbb{R}^d$.
- $\chi^2(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) = \exp(|\theta - \theta'|^2/\sigma^2) - 1$.
- Volume argument gives

$$N_p(\eta) = \left(\frac{\Gamma}{\eta}\right)^d, \quad N_{c,C}(\epsilon) = \left(\frac{3\Gamma}{\sigma\sqrt{\log(1+\epsilon^2)}}\right)^d$$

  where $\sigma\sqrt{\log(1+\epsilon^2)} \leq \Gamma$.
- Calculations in Fano's II (2) give the lower bound $d\sigma^2$.

Volume argument:

1. Bound $N_p(\eta)$ below.
   - $N_p(\eta) \geq N_c(\eta)$.
   - For any η covering set, Θ is contained in the union of the balls of radius η with centers in the covering set.
   - Volume of Θ must be smaller than the sum of the volumes of these balls.
   - Hence $N_p(\eta) \geq (\Gamma/\eta)^d$.
2. Bound $N_{c,C}(\epsilon, \mathcal{P}_\Theta)$ above.
   - Note $\chi^2(P_\theta, P_{\theta'}) = \exp(\rho^2(\theta, \theta')/\sigma^2) - 1$
   - $\chi^2(P_\theta, P_{\theta'}) \leq \epsilon^2$ iff $\rho(\theta, \theta') \leq \sigma\sqrt{\log(1+\epsilon^2)} =: \epsilon'$
   - This implies $N_{c,C}(\epsilon, \mathcal{P}_\Theta) \leq N_{c,C}(\epsilon', \Theta)$
   - $N_{c,C}(\epsilon', \Theta) \leq N_p(\epsilon')$. For any ϵ' packing set, the balls of radius $\epsilon'/2$ with centers in the packing set are all disjoint and their union is contained in the ball of radius $\Gamma + \epsilon'/2$ centered at the origin. Thus $N_p(\epsilon') \leq (1 + (2\Gamma/\epsilon'))^d \leq (3\Gamma/\epsilon')^d$.

Two fuzzy hypotheses I (Tsybakov §2.7.4)

Construct $A_0 \subseteq \Theta, A_1 \subseteq \Theta$ and consider two priors π_0 and π_1 supported on A_0 and A_1. Let $\mathbb{Q}_j(S) = \int \mathbb{P}_\theta(S) d\pi_j$ for $j = 0, 1$. Suppose we estimate $T(\theta)$ and

1. $\exists\; c \in \mathbb{R},\ s > 0, 0 \leq \beta_0, \beta_1 < 1$ s.t.

$$\pi_0(T(\theta) \leq c) \geq 1 - \beta_0, \quad \pi_1(T(\theta) \geq c + 2\delta) \geq 1 - \beta_1.$$

2. $$TV(\mathbb{Q}_0, \mathbb{Q}_1) \leq \alpha < 1.$$

Then for any estimator $\hat{T}$,

$$\sup_{\theta \in \Theta} \mathbb{E}_\theta |\hat{T} - T(\theta)| \geq \delta \left(\frac{1 - \alpha - \beta_0 - \beta_1}{2} \right).$$

Ideas in the proof:

$$\begin{aligned} 2R(\Theta, \hat{T}) &= \sup_{\theta \in \Theta} \mathbb{E}_\theta 2|\hat{T} - T(\theta)| \\ &\geq \int \mathbb{E}_\theta |\hat{T} - T(\theta)| d\pi_0(\theta) + \int \mathbb{E}_\theta |\hat{T} - T(\theta)| d\pi_1(\theta) = (1) + (2), \end{aligned}$$

where

$$\begin{aligned} (1) := \int \mathbb{E}_\theta |\hat{T} - T(\theta)| d\pi_0(\theta) &\geq \int \mathbb{E}_\theta |\hat{T} - T(\theta)| \mathbb{1}_{\{\hat{T} \geq c+\delta, T(\theta) < c\}} d\pi_0(\theta) \\ &\geq \delta \left(\int \mathbb{P}_\theta(\hat{T} \geq c + \delta) d\pi_0(\theta) - \beta_0 \right) \\ (2) := \int \mathbb{E}_\theta |\hat{T} - T(\theta)| d\pi_1(\theta) &\geq \int \mathbb{E}_\theta |\hat{T} - T(\theta)| \mathbb{1}_{\{\hat{T} \leq c+\delta, T(\theta) > c+2\delta\}} d\pi_1(\theta) \\ &\geq \delta \left(\int \mathbb{P}_\theta(\hat{T} \leq c + \delta) d\pi_1(\theta) - \beta_1 \right). \end{aligned}$$

Suffices to bound $\mathbb{Q}_0 f_0 + \mathbb{Q}_1(1 - f_0)$ from below.

Two fuzzy hypotheses II (Cai & Low, 2011)

Construct $A_0 \subseteq \Theta, A_1 \subseteq \Theta$ and consider two priors π_0 and π_1 supported on A_0 and A_1. Let $\mathbb{Q}_i(S) = \int \mathbb{P}_\theta(S) d\pi_i(\theta)$ for $i = 0, 1$. Suppose we estimate $T(\theta)$ and let $m_i = \int T(\theta) d\pi_i(\theta)$ and $\nu_i^2 = \int (T(\theta) - m_i)^2 d\pi_i(\theta)$.

1. Suppose $|m_1 - m_0| > \nu_0 \chi^2(\mathbb{Q}_0, \mathbb{Q}_1)$.

Then

$$\sup_{\theta \in \Theta} \mathbb{E}_\theta (\hat{T} - T(\theta))^2 \geq \frac{(|m_1 - m_0| - \nu_0 \chi^2(\mathbb{Q}_0, \mathbb{Q}_1))^2}{(\chi^2(\mathbb{Q}_0, \mathbb{Q}_1) + 2)^2}.$$

Proof ideas:

- Supremum risk is bounded by the average:

$$\sup_{\theta\in\Theta} \mathbb{E}_\theta(\hat{T} - T(\theta))^2 (*) \geq \int \mathbb{E}_\theta(\hat{T} - T(\theta))^2 (\lambda d\pi_0 + (1-\lambda) d\pi_1)$$

- If $\int \mathbb{E}_\theta(\hat{T} - T(\theta)) d\mu_0 \leq \epsilon^2$, then

$$\int \mathbb{E}_\theta(\hat{T} - T(\theta))^2 d\pi_1 \geq \int (\mathbb{E}_\theta \hat{T} - T(\theta))^2 d\pi_1$$
$$\geq \max(m_1 - m_0 - \nu_0\chi^2 - (\chi^2+1)\epsilon, 0)^2 =: (i).$$

- Then[5],

$$(*) \geq \lambda\epsilon^2 + (1-\lambda) \times (i)$$

is minimized at ϵ^*, and by plugging in this ϵ^* and optimizing w.r.t. $\lambda = \frac{\chi^2+1}{\chi^2+2}$, we get the lower bound.

[5]For the proof of (i), see the proof of Theorem 2 of Cai and Low (2011)

Example

Non-smooth functional (Cai & Low, 2011)

Let $Y_j \sim N(\theta_i, 1)$ and our interest is to estimate $T(\theta) = \frac{1}{n}\sum_{i=1}^{n} |\theta_i|$ assuming $|\theta_i| \leq 1$. Let $\Theta = \{\theta = (\theta_1, \ldots \theta_n), |\theta_i| \leq 1\}$. Then

$$\inf_{\hat{T}} \sup_{\theta \in \Theta} \mathbb{E}(\hat{T} - T(\theta))^2 \gtrsim \beta_*^2 \left(\frac{\log \log n}{\log n} \right)^2 .$$

- Use the existence of μ_0 and μ_1 on $[-1,1]$, for any $k \in \mathbb{N}$
  - μ_0 and μ_1 are symmetric around 0.
  - $\int t^l \mu_1(dt) = \int t^l \mu_0(dt)$, for $l = 0, \dots, k$.
  - $\int |t|\mu_1(dt) - \int |t|\mu_0(dt) = 2\delta_k$ where $\delta_k \sim k^{-1}(1 + o(1))$.
  - δ_k: distance (uniform norm) on $[-1,1]$ from $f(x) = |x|$ to the space of polynomials of no more than degree k.
- Let $\pi_0 = \mu_0^n, \pi_1 = \mu_1^n$.
- Note $m_1 - m_0 = \mathbb{E}_{\pi_1} T(\theta) - \mathbb{E}_{\pi_0} T(\theta) = \mathbb{E}_{\mu_1}|\theta_1| - \mathbb{E}_{\mu_0}|\theta_0| = 2\delta_k$, and

$$\nu_i^2 = \mathbb{E}_{\pi_0}(T(\theta) - m_0)^2 = \frac{1}{n}\mathbb{E}_{\mu_0}(|\theta_1| - \mathbb{E}_{\mu_0}|\theta_1|)^2 \leq \frac{1}{n}$$

- By choosing $k \sim \log n / \log\log n$, we can bound $\chi^2(\mathbb{Q}_0, \mathbb{Q}_1)$ as small as possible.