Minimax Lower Bounds

Jiuzhou Miao

April 23, 2025

Proof of Proposition 1. By using Markov's inequality, one has that

$$\mathbb{E}_{\mathbb{P}}\Big\{\phi\Big(\varrho\big(\hat{\theta},\theta(\mathbb{P})\big)\Big)\Big\} \geq \phi(\delta)\mathbb{P}\Big\{\phi\Big(\varrho\big(\hat{\theta},\theta(\mathbb{P})\big)\Big) \geq \phi(\delta)\Big\} \geq \phi(\delta)\mathbb{P}\Big\{\varrho\big(\hat{\theta},\theta(\mathbb{P})\big) \geq \delta\Big\}.$$

Thus, it suffices to lower bound the quantity

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\Big\{\varrho\big(\hat{\theta},\theta(\mathbb{P})\big) \geq \delta\Big\}.$$

Note that

$$\begin{aligned}
\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\Big\{\varrho\big(\hat{\theta},\theta(\mathbb{P})\big) \geq \delta\Big\} &\geq \max_{1\leq j\leq M} \mathbb{P}_{\theta_j}\big\{\varrho(\hat{\theta},\theta_j) \geq \delta\big\} \\
&\geq M^{-1}\sum_{j=1}^{M} \mathbb{P}_{\theta_j}\big\{\varrho(\hat{\theta},\theta_j) \geq \delta\big\} \\
&= \sum_{j=1}^{M} \mathbb{Q}(J=j)\mathbb{Q}\big\{\varrho(\hat{\theta},\theta_j) \geq \delta \big| J=j\big\} \\
&= \mathbb{Q}\big\{\varrho(\hat{\theta},\theta_J) \geq \delta\big\},
\end{aligned}$$

so we have reduced the problem to lower bounding the quantity $\mathbb{Q}\big\{\varrho(\hat{\theta},\theta_J) \geq \delta\big\}$.

Note that any estimator $\hat{\theta}$ can be used to define a test via

$$\psi(Z) = \arg\min_{1\leq l\leq M} \varrho(\theta_l,\hat{\theta}).$$

Suppose that the true parameter is θ_j, then the event $\big\{\varrho(\hat{\theta},\theta_j) < \delta\big\}$ ensures that $\psi(Z) = j$ because for any other $1 \leq k \leq M$ and $k \neq j$, one has that

$$\varrho(\theta_k,\hat{\theta}) \geq \varrho(\theta_k,\theta_j) - \varrho(\theta_j,\hat{\theta}) > 2\delta - \delta = \delta.$$

Therefore, one has that

$$\mathbb{Q}\big\{\varrho(\hat{\theta},\theta_J) \geq \delta\big\} = M^{-1}\sum_{j=1}^{M} \mathbb{P}_{\theta_j}\big\{\varrho(\hat{\theta},\theta_j) \geq \delta\big\} \geq \mathbb{Q}\big\{\psi(Z) \neq J\big\}.$$

□

Proof of Proposition 2. We only prove thee first equality. Let $A_0 = \{q \geq p\}$. Note that

$$\begin{aligned}
\int_{A_0^c} (p-q)\mathrm{d}\nu &= \int_{A_0^c} p\mathrm{d}\nu - \int_{A_0^c} q\mathrm{d}\nu \\
&= 1 - \int_{A_0} p\mathrm{d}\nu - 1 + \int_{A_0} q\mathrm{d}\nu \\
&= \int_{A_0} (q-p)\mathrm{d}\nu,
\end{aligned}$$

one has that

$$\int_{\mathcal{X}} |p-q|\mathrm{d}\nu = \int_{A_0^c} (p-q)\mathrm{d}\nu + \int_{A_0} (q-p)\mathrm{d}\nu = 2\int_{A_0} (q-p)\mathrm{d}\nu.$$

Then

$$\begin{aligned}
\mathrm{TV}(\mathbb{P}\|\mathbb{Q}) &= \sup_{A\in\mathcal{A}} \Big| \int_A (p-q)\mathrm{d}\nu \Big| \\
&\geq \int_{A_0} (q-p)\mathrm{d}\nu \\
&= 2^{-1} \int_{\mathcal{X}} |p-q|\mathrm{d}\nu.
\end{aligned}$$

On the other hand, for any $A \in \mathcal{A}$,

$$\begin{aligned}
\mathrm{TV}(\mathbb{P}\|\mathbb{Q}) &= \sup_{A\in\mathcal{A}} \Big| \int_A (p-q)\mathrm{d}\nu \Big| \\
&= \sup_{A\in\mathcal{A}} \Big| \int_A (p-q)\mathbb{I}_{A_0}\mathrm{d}\nu + \int_A (p-q)\mathbb{I}_{A_0^c}\mathrm{d}\nu \Big| \\
&= \max\Big\{ \sup_{A\in\mathcal{A}} \int_A (q-p)\mathbb{I}_{A_0}\mathrm{d}\nu, \sup_{A\in\mathcal{A}} \int_A (p-q)\mathbb{I}_{A_0^c}\mathrm{d}\nu \Big\} \\
&\leq \max\Big\{ \int_{A_0} (q-p)\mathrm{d}\nu, \int_{A_0^c} (p-q)\mathrm{d}\nu \Big\} \\
&= 2^{-1} \int_{\mathcal{X}} |p-q|\mathrm{d}\nu.
\end{aligned}$$

□

Proof of Proposition 3. Let $h(x) = x\log x - x + 1$ for $x \geq 0$, where $0\log 0 := 0$. We claim that

$$3^{-1}(4+2x)h(x) \geq (x-1)^2$$

for all $x \geq 0$. Then one has that

$$\begin{aligned}
\mathrm{TV}(\mathbb{P}\|\mathbb{Q}) &= 2^{-1}\int_{\mathcal{X}} |p-q|\mathrm{d}\nu \\
&= 2^{-1}\int_{\mathcal{X}} |p/q-1|q\mathrm{d}\nu \\
&\leq 2^{-1}\int_{\mathcal{X}} q\sqrt{3^{-1}(4+2p/q)h(p/q)}\mathrm{d}\nu \\
&= 2^{-1}\int_{\mathcal{X}} \sqrt{3^{-1}(4q+2p)qh(p/q)}\mathrm{d}\nu \\
&\leq 2^{-1}\sqrt{\int_{\mathcal{X}} 3^{-1}(4q+2p)\mathrm{d}\nu}\sqrt{\int_{\mathcal{X}} qh(p/q)\mathrm{d}\nu} \\
&= \sqrt{2^{-1}\int_{\mathcal{X}} qh(p/q)\mathrm{d}\nu} \\
&= \sqrt{2^{-1}\mathrm{KL}(\mathbb{P}\|\mathbb{Q})}.
\end{aligned}$$

□

Proof of Proposition 4. Note that

$$\begin{aligned}
p \vee q &= 2^{-1}\big(p+q+|p-q|\big), \\
p \wedge q &= 2^{-1}\big(p+q-|p-q|\big),
\end{aligned}$$

one has that

$$\int_{\mathcal{X}}(p\vee q)\mathrm{d}\nu + \int_{\mathcal{X}}(p\wedge q)\mathrm{d}\nu = 2.$$

Then

$$\begin{aligned}
\big\{1-2^{-1}\mathrm{H}^2(\mathbb{P}\|\mathbb{Q})\big\}^2 &= \Big(\int_{\mathcal{X}}\sqrt{pq}\mathrm{d}\nu\Big)^2 \\
&= \Big\{\int_{\mathcal{X}}\sqrt{(p\vee q)(p\wedge q)}\mathrm{d}\nu\Big\}^2 \\
&\leq \int_{\mathcal{X}}(p\vee q)\mathrm{d}\nu\int_{\mathcal{X}}(p\wedge q)\mathrm{d}\nu \\
&= \int_{\mathcal{X}}(p\wedge q)\mathrm{d}\nu\Big\{2-\int_{\mathcal{X}}(p\wedge q)\mathrm{d}\nu\Big\} \\
&= \big\{1-\mathrm{TV}(\mathbb{P}\|\mathbb{Q})\big\}\big\{1+\mathrm{TV}(\mathbb{P}\|\mathbb{Q})\big\} \\
&= 1-\mathrm{TV}^2(\mathbb{P}\|\mathbb{Q})
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{TV}^2(\mathbb{P}\|\mathbb{Q}) &\leq 1-\big\{1-2^{-1}\mathrm{H}^2(\mathbb{P}\|\mathbb{Q})\big\}^2 \\
&= 2^{-1}\mathrm{H}^2(\mathbb{P}\|\mathbb{Q})\big\{2-2^{-1}\mathrm{H}^2(\mathbb{P}\|\mathbb{Q})\big\} \\
&= \mathrm{H}^2(\mathbb{P}\|\mathbb{Q})\big\{1-4^{-1}\mathrm{H}^2(\mathbb{P}\|\mathbb{Q})\big\}.
\end{aligned}$$

□

Proof of Theorem 5. For any estimator $\hat{\theta}$, define

$$V_j(\hat{\theta}) = 2^{-1}\delta^{-1} \inf_{\mathbb{P}_j \in \mathcal{P}_j} \varrho\big(\hat{\theta}, \theta(\mathbb{P}_j)\big)$$

for $j = 0, 1$. One has that

$$\begin{aligned}\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\Big\{\varrho\big(\hat{\theta}, \theta(\mathbb{P})\big)\Big\} &\geq 2^{-1}\mathbb{E}_{\mathbb{P}_0}\Big\{\varrho\big(\hat{\theta}, \theta(\mathbb{P}_0)\big)\Big\} + 2^{-1}\mathbb{E}_{\mathbb{P}_1}\Big\{\varrho\big(\hat{\theta}, \theta(\mathbb{P}_1)\big)\Big\} \\ &\geq \delta\mathbb{E}_{\mathbb{P}_0}\big\{V_0(\hat{\theta})\big\} + \delta\mathbb{E}_{\mathbb{P}_1}\big\{V_1(\hat{\theta})\big\}.\end{aligned}$$

Since the right-hand side is linear in $\mathbb{P}_0$ and $\mathbb{P}_1$, we can take suprema over the convex hulls and thus obtain the lower bound

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\Big\{\varrho\big(\hat{\theta}, \theta(\mathbb{P})\big)\Big\} \geq \delta \sup_{\mathbb{P}_j\in\mathrm{conv}(\mathcal{P}_j), j=0,1} \Big[\mathbb{E}_{\mathbb{P}_0}\big\{V_0(\hat{\theta})\big\} + \mathbb{E}_{\mathbb{P}_1}\big\{V_1(\hat{\theta})\big\}\Big].$$

By the triangle inequality, one has that

$$\varrho\big(\hat{\theta}, \theta(\mathbb{P}_0)\big) + \varrho\big(\hat{\theta}, \theta(\mathbb{P}_1)\big) \geq \varrho\big(\theta(\mathbb{P}_0), \theta(\mathbb{P}_1)\big) \geq 2\delta.$$

Then infima over $\mathbb{P}_j \in \mathcal{P}_j$ for each $j = 0, 1$,

$$\inf_{\mathbb{P}_0\in\mathcal{P}_0} \varrho\big(\hat{\theta}, \theta(\mathbb{P}_0)\big) + \inf_{\mathbb{P}_1\in\mathcal{P}_1} \varrho\big(\hat{\theta}, \theta(\mathbb{P}_1)\big) \geq 2\delta,$$

which is equivalent to $V_0(\hat{\theta}) + V_1(\hat{\theta}) \geq 1$. Since $V_j(\hat{\theta}) \geq 0$ for $j = 0, 1$, the variational representation of the TV distance implies that

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_0}\big\{V_0(\hat{\theta})\big\} + \mathbb{E}_{\mathbb{P}_1}\big\{V_1(\hat{\theta})\big\} &\geq 1 + \mathbb{E}_{\mathbb{P}_0}\big\{V_0(\hat{\theta})\big\} - \mathbb{E}_{\mathbb{P}_1}\big\{V_0(\hat{\theta})\big\} \\ &\geq 1 - \mathrm{TV}(\mathbb{P}_0\|\mathbb{P}_1).\end{aligned}$$

□

Lemma 1. *Let $g(x) = h(x) + x\log\{(M-1)\}$ with $h(x) = -x\log x - (1-x)\log(1-x)$. For $\{p_j\}_{j=1}^M$ satisfy $p_j \geq 0$ for all $1 \leq j \leq M$ and $\sum_{j=1}^M p_j = 1$, one has that*

$$g\Big(\sum_{j\neq i} p_j\Big) \geq -\sum_{j=1}^M p_j \log p_j.$$

Proof.

$$\begin{aligned}\sum_{j=1}^M p_j\log p_j &= p_i\log p_i + \sum_{j\neq i} p_j\log p_j \\ &= p_i\log p_i + \Big(\sum_{j\neq i} p_j\Big)\log\Big(\sum_{j\neq i} p_j\Big) + \sum_{j\neq i} p_j \log\frac{p_j}{\sum_{j\neq i} p_j} \\ &= -h\Big(\sum_{j\neq i} p_j\Big) - \sum_{j\neq i}\Big(1 - \sum_{j\neq i} p_j\Big)\log\frac{\sum_{j\neq i} p_j}{1 - \sum_{j\neq i} p_j} \\ &\geq -h\Big(\sum_{j\neq i} p_j\Big) - \log(M-1) \\ &\geq -g\Big(\sum_{j\neq i} p_j\Big).\end{aligned}$$

□

Proof of Theorem 6. Let $h(x) = -x\log x - (1-x)\log(1-x)$ and denote $q_e = \mathbb{Q}(\psi(Z) \neq J)$. We will show that

$$g(q_e) = h(q_e) + q_e \log(M-1) \geq \log M - \mathcal{I}(Z, J).$$

Then by using that $h(q_e) \leq \log 2$, one has that

$$\log 2 + q_e \log(M-1) \geq g(q_e) \geq \log M - \mathcal{I}(Z, J)$$

and

$$q_e \geq \frac{\log M - \log 2 - \mathcal{I}(Z, J)}{\log(M-1)} \geq 1 - \frac{\log 2 + \mathcal{I}(Z, J)}{\log M}.$$

Note that

$$\begin{aligned}
q_e &= M^{-1}\sum_{j=1}^{M} \mathbb{P}_{\theta_j}\{\psi(Z) \neq j\} \\
&= M^{-1}\sum_{j=1}^{M} \int \mathbb{I}\{\psi(Z) \neq j\} \mathrm{d}\mathbb{P}_{\theta_j} \\
&= \int \sum_{j=1}^{M} \mathbb{I}\{\psi(Z) \neq j\} \frac{\mathrm{d}\mathbb{P}_{\theta_j}}{M\mathrm{d}\bar{\mathbb{Q}}} \mathrm{d}\bar{\mathbb{Q}} \\
&= \mathbb{E}_{\bar{\mathbb{Q}}}\Big(\sum_{j=1}^{M} \mathbb{I}\{\psi(Z) \neq j\} \frac{\mathrm{d}\mathbb{P}_{\theta_j}}{M\mathrm{d}\bar{\mathbb{Q}}}\Big) \\
&= \mathbb{E}_{\bar{\mathbb{Q}}}\Big(\sum_{j \neq \psi(Z)} \frac{\mathrm{d}\mathbb{P}_{\theta_j}}{M\mathrm{d}\bar{\mathbb{Q}}}\Big).
\end{aligned}$$

By using Lemma 1, one has that

$$\begin{aligned}
g(q_e) &= g\bigg(\mathbb{E}_{\bar{\mathbb{Q}}}\Big(\sum_{j \neq \psi(Z)} \frac{\mathrm{d}\mathbb{P}_{\theta_j}}{M\mathrm{d}\bar{\mathbb{Q}}}\Big)\bigg) \\
&\geq \mathbb{E}_{\bar{\mathbb{Q}}}\bigg\{g\Big(\sum_{j \neq \psi(Z)} \frac{\mathrm{d}\mathbb{P}_{\theta_j}}{M\mathrm{d}\bar{\mathbb{Q}}}\Big)\bigg\} \\
&\geq -\mathbb{E}_{\bar{\mathbb{Q}}}\Big(\sum_{j=1}^{M} \frac{\mathrm{d}\mathbb{P}_{\theta_j}}{M\mathrm{d}\bar{\mathbb{Q}}} \log \frac{\mathrm{d}\mathbb{P}_{\theta_j}}{M\mathrm{d}\bar{\mathbb{Q}}}\Big) \\
&= \log M - \mathcal{I}(Z, J).
\end{aligned}$$

Proof of Proposition 7. For any other distribution $\tilde{\mathbb{Q}}$, one has that

$$\begin{aligned}
&M^{-1}\sum_{j=1}^{M} \mathrm{KL}(\mathbb{P}_{\theta_j}\|\bar{\mathbb{Q}}) - M^{-1}\sum_{j=1}^{M} \mathrm{KL}(\mathbb{P}_{\theta_j}\|\tilde{\mathbb{Q}}) \\
&= M^{-1}\sum_{j=1}^{M} \int \log \frac{\mathrm{d}\tilde{\mathbb{Q}}}{\mathrm{d}\bar{\mathbb{Q}}} \mathrm{d}\mathbb{P}_{\theta_j} \\
&= \int \log \frac{\mathrm{d}\tilde{\mathbb{Q}}}{\mathrm{d}\bar{\mathbb{Q}}} \mathrm{d}\bar{\mathbb{Q}} \\
&\leq 0,
\end{aligned}$$

then one has that

$$\mathcal{I}(Z,J) = M^{-1}\sum_{j=1}^{M}\mathrm{KL}(\mathbb{P}_{\theta_j}\|\bar{\mathbb{Q}}) \leq M^{-1}\sum_{j=1}^{M}\mathrm{KL}(\mathbb{P}_{\theta_j}\|\tilde{\mathbb{Q}}) \leq \max_{1\leq j\leq M}\mathrm{KL}(\mathbb{P}_{\theta_j}\|\tilde{\mathbb{Q}}).$$

Let $\{\gamma_1, \gamma_N\}$ be an ϵ-covering of parameter space Ω in the square-root KL divergence and set $\tilde{\mathbb{Q}} = N^{-1}\sum_{k=1}^{N}\mathbb{P}_{\gamma_k}$. For each θ_j, $1 \leq j \leq M$, one can find some γ_k such that $\mathrm{KL}(\mathbb{P}_{\theta_j}\|\mathbb{P}_{\gamma_k}) \leq \epsilon^2$. Then

$$\begin{aligned}
\mathrm{KL}(\mathbb{P}_{\theta_j}\|\tilde{\mathbb{Q}}) &= \int \log \frac{\mathrm{d}\mathbb{P}_{\theta_j}}{N^{-1}\sum_{k=1}^{N}\mathrm{d}\mathbb{P}_{\gamma_k}}\mathrm{d}\mathbb{P}_{\theta_j} \\
&\leq \int \log \frac{\mathrm{d}\mathbb{P}_{\theta_j}}{N^{-1}\mathrm{d}\mathbb{P}_{\gamma_k}}\mathrm{d}\mathbb{P}_{\theta_j} \\
&= \mathrm{KL}(\mathbb{P}_{\theta_j}\|\mathbb{P}_{\gamma_k}) + \log N \\
&\leq \epsilon^2 + \log N.
\end{aligned}$$

□

□