Minimax Lower Bounds

Jiuzhou Miao

School of Statistics and Mathematics, Zhejiang Gongshang University

April 23, 2025

Minimax risks

- Given a family of distributions $\mathcal{P}$, let $\theta : \mathcal{P} \to \Omega$ be a functional.

Minimax risks

- Given a family of distributions $\mathcal{P}$, let $\theta : \mathcal{P} \to \Omega$ be a functional.
- Let X be a random variable taking values on $\mathcal{X}$ drawn according to a distribution $\mathbb{P} \in \mathcal{P}$ for which $\theta(\mathbb{P}) = \theta^*$.

Minimax risks

- Given a family of distributions $\mathcal{P}$, let $\theta : \mathcal{P} \to \Omega$ be a functional.
- Let X be a random variable taking values on $\mathcal{X}$ drawn according to a distribution $\mathbb{P} \in \mathcal{P}$ for which $\theta(\mathbb{P}) = \theta^*$.
- An estimator $\hat{\theta}$ can be viewed as a measurable function from the domain $\mathcal{X}$ to the parameter space Ω.

Minimax risks

- Given a family of distributions $\mathcal{P}$, let $\theta : \mathcal{P} \to \Omega$ be a functional.
- Let X be a random variable taking values on $\mathcal{X}$ drawn according to a distribution $\mathbb{P} \in \mathcal{P}$ for which $\theta(\mathbb{P}) = \theta^*$.
- An estimator $\hat{\theta}$ can be viewed as a measurable function from the domain $\mathcal{X}$ to the parameter space Ω.
- In order to assess the quality of any estimator, let $\varrho : \Omega \times \Omega \to \mathbb{R}_+$ be a semi-metric.

Minimax risks

- Given a family of distributions $\mathcal{P}$, let $\theta : \mathcal{P} \to \Omega$ be a functional.
- Let X be a random variable taking values on $\mathcal{X}$ drawn according to a distribution $\mathbb{P} \in \mathcal{P}$ for which $\theta(\mathbb{P}) = \theta^*$.
- An estimator $\hat{\theta}$ can be viewed as a measurable function from the domain $\mathcal{X}$ to the parameter space Ω.
- In order to assess the quality of any estimator, let $\varrho : \Omega \times \Omega \to \mathbb{R}_+$ be a semi-metric.
- The quantity

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left\{ \varrho(\hat{\theta}, \theta^*) \right\}$$

reflects the worst-case performance of the estimators.

Minimax risks

- The minimax risk is the smallest worst-case risk over all possible estimators:

$$\mathfrak{M}\big(\theta(\mathcal{P}), \varrho\big) = \inf_{\hat{\theta}} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\Big\{\varrho\big(\hat{\theta}, \theta(\mathbb{P})\big)\Big\}.$$

Minimax risks

- The minimax risk is the smallest worst-case risk over all possible estimators:

$$\mathfrak{M}\big(\theta(\mathcal{P}), \varrho\big) = \inf_{\hat{\theta}} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\Big\{\varrho\big(\hat{\theta}, \theta(\mathbb{P})\big)\Big\}.$$

- Let $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ be an increasing function, one can define a slight generalization of the ϱ-minimax risk

$$\mathfrak{M}\big(\theta(\mathcal{P}), \phi\circ\varrho\big) = \inf_{\hat{\theta}} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\Bigg\{\phi\Big(\varrho\big(\hat{\theta}, \theta(\mathbb{P})\big)\Big)\Bigg\}.$$

Minimax risks

- The minimax risk is the smallest worst-case risk over all possible estimators:

$$\mathfrak{M}\big(\theta(\mathcal{P}),\varrho\big)=\inf_{\hat{\theta}}\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}\Big\{\varrho\big(\hat{\theta},\theta(\mathbb{P})\big)\Big\}.$$

- Let $\phi:\mathbb{R}_+\to\mathbb{R}_+$ be an increasing function, one can define a slight generalization of the ϱ-minimax risk

$$\mathfrak{M}\big(\theta(\mathcal{P}),\phi\circ\varrho\big)=\inf_{\hat{\theta}}\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}\Big\{\phi\Big(\varrho\big(\hat{\theta},\theta(\mathbb{P})\big)\Big)\Big\}.$$

- A particularly common choice is $\phi(t)=t^2$, which can be used to obtain minimax risks for the MSE associated with ϱ.

From estimation to testing

- Let $\{\theta_1, \ldots, \theta_M\}$ be a 2δ-separated set contained in the space $\theta(\mathcal{P})$, meaning that $\varrho(\theta_j, \theta_k) \geq 2\delta$ for all $j \neq k$.

From estimation to testing

- Let $\{\theta_1, \ldots, \theta_M\}$ be a 2δ-separated set contained in the space $\theta(\mathcal{P})$, meaning that $\varrho(\theta_j, \theta_k) \geq 2\delta$ for all $j \neq k$.
- For each θ_j, let us choose some representative distribution $\mathbb{P}_{\theta_j}$ such that $\theta(\mathbb{P}_{\theta_j}) = \theta_j$.

From estimation to testing

- Let $\{\theta_1, \ldots, \theta_M\}$ be a 2δ-separated set contained in the space $\theta(\mathcal{P})$, meaning that $\varrho(\theta_j, \theta_k) \geq 2\delta$ for all $j \neq k$.
- For each θ_j, let us choose some representative distribution $\mathbb{P}_{\theta_j}$ such that $\theta(\mathbb{P}_{\theta_j}) = \theta_j$.
- We generate a random variable Z by the following procedure:

From estimation to testing

- Let $\{\theta_1, \ldots, \theta_M\}$ be a 2δ-separated set contained in the space $\theta(\mathcal{P})$, meaning that $\varrho(\theta_j, \theta_k) \geq 2\delta$ for all $j \neq k$.
- For each θ_j, let us choose some representative distribution $\mathbb{P}_{\theta_j}$ such that $\theta(\mathbb{P}_{\theta_j}) = \theta_j$.
- We generate a random variable Z by the following procedure:
  (1) Sample a random integer J from the uniform distribution over $\{1, \ldots, M\}$.

From estimation to testing

- Let $\{\theta_1, \ldots, \theta_M\}$ be a 2δ-separated set contained in the space $\theta(\mathcal{P})$, meaning that $\varrho(\theta_j, \theta_k) \geq 2\delta$ for all $j \neq k$.
- For each θ_j, let us choose some representative distribution $\mathbb{P}_{\theta_j}$ such that $\theta(\mathbb{P}_{\theta_j}) = \theta_j$.
- We generate a random variable Z by the following procedure:
  (1) Sample a random integer J from the uniform distribution over $\{1, \ldots, M\}$.
  (2) Given $J = j$, sample $Z|J = j \sim \mathbb{P}_{\theta_j}$.

From estimation to testing

- Denote $\mathbb{Q}$ as the joint distribution of (Z, J) and $\bar{\mathbb{Q}} = M^{-1}\sum_{j=1}^{M}\mathbb{P}_{\theta_j}$ as the marginal distribution of Z.

From estimation to testing

- Denote $\mathbb{Q}$ as the joint distribution of (Z, J) and $\bar{\mathbb{Q}} = M^{-1} \sum_{j=1}^{M} \mathbb{P}_{\theta_j}$ as the marginal distribution of Z.
- Consider the M-ary hypothesis testing problem of determining the randomly chosen index J.

From estimation to testing

- Denote $\mathbb{Q}$ as the joint distribution of (Z, J) and $\bar{\mathbb{Q}} = M^{-1} \sum_{j=1}^{M} \mathbb{P}_{\theta_j}$ as the marginal distribution of Z.
- Consider the M-ary hypothesis testing problem of determining the randomly chosen index J.
- Let $\psi : \mathcal{Z} \to \{1, \ldots, M\}$ be the testing function for this problem, and the associated probability of error is given by $\mathbb{Q}\{\psi(Z) \neq J\}$.

From estimation to testing

Proposition 1 (From estimation to testing)

For any increasing function ϕ and and choice of 2δ-separated set the minimax risk is lower bounded as

$$\mathfrak{M}\big(\theta(\mathcal{P}), \phi \circ \varrho\big) \geq \phi(\delta) \inf_{\psi} \mathbb{Q}\big\{\psi(Z) \neq J\big\}.$$

- By choosing δ such that $\mathbb{Q}\big\{\psi(Z) \neq J\big\} \geq 1/2$, then

$$\mathfrak{M}\big(\theta(\mathcal{P}), \phi \circ \varrho\big) \geq 2^{-1}\phi(\delta).$$

Some divergence measures

- Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures and ν be a σ-finite measure on a measurable space $(\mathcal{X}, \mathcal{A})$.

Some divergence measures

- Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures and ν be a σ-finite measure on a measurable space $(\mathcal{X}, \mathcal{A})$.
- Suppose that $\mathbb{P} \ll \nu$ and $\mathbb{Q} \ll \nu$ and define $p = \mathrm{d}\mathbb{P}/\mathrm{d}\nu$ and $q = \mathrm{d}\mathbb{Q}/\mathrm{d}\nu$.

Some divergence measures

- Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures and ν be a σ-finite measure on a measurable space $(\mathcal{X}, \mathcal{A})$.
- Suppose that $\mathbb{P} \ll \nu$ and $\mathbb{Q} \ll \nu$ and define $p = \mathrm{d}\mathbb{P}/\mathrm{d}\nu$ and $q = \mathrm{d}\mathbb{Q}/\mathrm{d}\nu$.
- Total variation (TV) distance:

$$\mathrm{TV}(\mathbb{P}\|\mathbb{Q}) = \sup_{A\in\mathcal{A}} \big|\mathbb{P}(A) - \mathbb{Q}(A)\big| = \sup_{A\in\mathcal{A}} \bigg|\int_A (p-q)\mathrm{d}\nu\bigg|.$$

Some divergence measures

- Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures and ν be a σ-finite measure on a measurable space $(\mathcal{X}, \mathcal{A})$.
- Suppose that $\mathbb{P} \ll \nu$ and $\mathbb{Q} \ll \nu$ and define $p = \mathrm{d}\mathbb{P}/\mathrm{d}\nu$ and $q = \mathrm{d}\mathbb{Q}/\mathrm{d}\nu$.
- Total variation (TV) distance:

$$\mathrm{TV}(\mathbb{P}\|\mathbb{Q}) = \sup_{A\in\mathcal{A}} \big|\mathbb{P}(A) - \mathbb{Q}(A)\big| = \sup_{A\in\mathcal{A}} \left|\int_A (p-q)\mathrm{d}\nu\right|.$$

- Kullback-Leibler (KL) divergence: Assume that $\mathbb{P} \ll \mathbb{Q}$,

$$\mathrm{KL}(\mathbb{P}\|\mathbb{Q}) = \int \log\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\mathrm{d}\mathbb{P} = \int_{\mathcal{X}} p\log\frac{p}{q}\mathrm{d}\nu.$$

Some divergence measures

- Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures and ν be a σ-finite measure on a measurable space $(\mathcal{X}, \mathcal{A})$.
- Suppose that $\mathbb{P} \ll \nu$ and $\mathbb{Q} \ll \nu$ and define $p = \mathrm{d}\mathbb{P}/\mathrm{d}\nu$ and $q = \mathrm{d}\mathbb{Q}/\mathrm{d}\nu$.
- Total variation (TV) distance:

$$\mathrm{TV}(\mathbb{P}\|\mathbb{Q}) = \sup_{A\in\mathcal{A}} \left|\mathbb{P}(A) - \mathbb{Q}(A)\right| = \sup_{A\in\mathcal{A}} \left|\int_A (p-q)\mathrm{d}\nu\right|.$$

- Kullback-Leibler (KL) divergence: Assume that $\mathbb{P} \ll \mathbb{Q}$,

$$\mathrm{KL}(\mathbb{P}\|\mathbb{Q}) = \int \log \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} \mathrm{d}\mathbb{P} = \int_{\mathcal{X}} p \log \frac{p}{q} \mathrm{d}\nu.$$

- Squared Hellinger distance:

$$\mathrm{H}^2(\mathbb{P}\|\mathbb{Q}) = \int \left(\sqrt{\mathrm{d}\mathbb{P}} - \sqrt{\mathrm{d}\mathbb{Q}}\right)^2 = \int_{\mathcal{X}} \left(\sqrt{p} - \sqrt{q}\right)^2 \mathrm{d}\nu.$$

Properties of TV distance

- $\mathrm{TV}(\mathbb{P}\|\mathbb{Q})$ satisfies the axioms of distance.

Proposition 2 (Scheffé)

$$\mathrm{TV}(\mathbb{P}\|\mathbb{Q}) = 2^{-1}\int_{\mathcal{X}} |p-q|\mathrm{d}\nu = 1 - \int_{\mathcal{X}} (p \wedge q)\mathrm{d}\nu.$$

Properties of TV distance

- $\mathrm{TV}(\mathbb{P}\|\mathbb{Q})$ satisfies the axioms of distance.
- $0 \leq \mathrm{TV}(\mathbb{P}\|\mathbb{Q}) \leq 1.$

Proposition 2 (Scheffé)

$$\mathrm{TV}(\mathbb{P}\|\mathbb{Q}) = 2^{-1} \int_{\mathcal{X}} |p - q| \mathrm{d}\nu = 1 - \int_{\mathcal{X}} (p \wedge q) \mathrm{d}\nu.$$

Properties of KL divergence

- $\mathrm{KL}(\mathbb{P}\|\mathbb{Q})$ is not a distance.

Properties of KL divergence

- $\mathrm{KL}(\mathbb{P}\|\mathbb{Q})$ is not a distance.
- $\mathrm{KL}(\mathbb{P}\|\mathbb{Q}) \geq 0.$

Properties of KL divergence

- $\mathrm{KL}(\mathbb{P}\|\mathbb{Q})$ is not a distance.
- $\mathrm{KL}(\mathbb{P}\|\mathbb{Q}) \geq 0$.
- Let $\mathbb{P}^{1:n} = \otimes_{i=1}^{n}\mathbb{P}_i$ and $\mathbb{Q}^{1:n} = \otimes_{i=1}^{n}\mathbb{Q}_i$. Then

$$\mathrm{KL}(\mathbb{P}^{1:n}\|\mathbb{Q}^{1:n}) = \sum_{i=1}^{n}\mathrm{KL}(\mathbb{P}_i\|\mathbb{Q}_i).$$

Properties of Squared Hellinger distance

- $\mathrm{H}(\mathbb{P}\|\mathbb{Q}) = \sqrt{\mathrm{H}^2(\mathbb{P}\|\mathbb{Q})}$ satisfies the axioms of distance.

Properties of Squared Hellinger distance

- $\mathrm{H}(\mathbb{P}\|\mathbb{Q}) = \sqrt{\mathrm{H}^2(\mathbb{P}\|\mathbb{Q})}$ satisfies the axioms of distance.
- $0 \leq \mathrm{H}^2(\mathbb{P}\|\mathbb{Q}) \leq 2.$

Properties of Squared Hellinger distance

- $\mathrm{H}(\mathbb{P}\|\mathbb{Q}) = \sqrt{\mathrm{H}^2(\mathbb{P}\|\mathbb{Q})}$ satisfies the axioms of distance.
- $0 \leq \mathrm{H}^2(\mathbb{P}\|\mathbb{Q}) \leq 2$.
- $$\mathrm{H}^2(\mathbb{P}\|\mathbb{Q}) = 2\Big(1 - \int \sqrt{\mathrm{d}\mathbb{P}\mathrm{d}\mathbb{Q}}\Big) = 2\Big(1 - \int \sqrt{pq}\,\mathrm{d}\nu\Big).$$

Properties of Squared Hellinger distance

- $\mathrm{H}(\mathbb{P}\|\mathbb{Q}) = \sqrt{\mathrm{H}^2(\mathbb{P}\|\mathbb{Q})}$ satisfies the axioms of distance.
- $0 \le \mathrm{H}^2(\mathbb{P}\|\mathbb{Q}) \le 2$.
-
$$\mathrm{H}^2(\mathbb{P}\|\mathbb{Q}) = 2\Big(1 - \int \sqrt{\mathrm{d}\mathbb{P}\mathrm{d}\mathbb{Q}}\Big) = 2\Big(1 - \int \sqrt{pq}\,\mathrm{d}\nu\Big).$$
- Let $\mathbb{P}^{1:n} = \otimes_{i=1}^n \mathbb{P}_i$ and $\mathbb{Q}^{1:n} = \otimes_{i=1}^n \mathbb{Q}_i$. Then
$$\mathrm{H}^2(\mathbb{P}^{1:n}\|\mathbb{Q}^{1:n}) = 2\Big[1 - \prod_{i=1}^n \{1 - 2^{-1}\mathrm{H}^2(\mathbb{P}_i\|\mathbb{Q}_i)\}\Big].$$

Two inequalities

Proposition 3 (Pinsker-Csiszár-Kullback inequality)

$$\mathrm{TV}(\mathbb{P}\|\mathbb{Q}) \leq \sqrt{2^{-1}\mathrm{KL}(\mathbb{P}\|\mathbb{Q})}.$$

Proposition 4 (Le Cam's inequality)

$$\mathrm{TV}(\mathbb{P}\|\mathbb{Q}) \leq \mathrm{H}(\mathbb{P}\|\mathbb{Q})\sqrt{1 - 4^{-1}\mathrm{H}^2(\mathbb{P}\|\mathbb{Q})}.$$

1. Basic framework
2. Binary testing and LeCam's method
3. Fano's method

Binary testing and TV distance

- In a binary testing problem with equally weighted hypotheses, assume that Z drawn according to the distribution $\bar{\mathbb{Q}} = 2^{-1}(\mathbb{P}_0 + \mathbb{P}_1)$.

Binary testing and TV distance

- In a binary testing problem with equally weighted hypotheses, assume that Z drawn according to the distribution $\bar{\mathbb{Q}} = 2^{-1}(\mathbb{P}_0 + \mathbb{P}_1)$.
- For a given decision rule $\psi : \mathcal{Z} \to \{0, 1\}$, the associated probability of error is given by

$$\mathbb{Q}\{\psi(Z) \neq J\} = 2^{-1}\mathbb{P}_0\{\psi(Z) \neq 0\} + 2^{-1}\mathbb{P}_1\{\psi(Z) \neq 1\}.$$

Binary testing and TV distance

- In a binary testing problem with equally weighted hypotheses, assume that Z drawn according to the distribution $\bar{\mathbb{Q}} = 2^{-1}(\mathbb{P}_0 + \mathbb{P}_1)$.
- For a given decision rule $\psi : \mathcal{Z} \to \{0, 1\}$, the associated probability of error is given by

$$\mathbb{Q}\{\psi(Z) \neq J\} = 2^{-1}\mathbb{P}_0\{\psi(Z) \neq 0\} + 2^{-1}\mathbb{P}_1\{\psi(Z) \neq 1\}.$$

- Let $A = \{x \in \mathcal{X} : \psi(x) = 1\}$, one has that

$$\begin{aligned}\sup_{\psi} \mathbb{Q}\{\psi(Z) = J\} &= \sup_{A \subseteq \mathcal{X}} \{2^{-1}\mathbb{P}_0(A^c) + 2^{-1}\mathbb{P}_1(A)\} \\ &= 2^{-1}\{\mathrm{TV}(\mathbb{P}_1 \| \mathbb{P}_0) + 1\}.\end{aligned}$$

Binary testing and TV distance

- Then

$$\inf_{\psi} \mathbb{Q}\{\psi(Z) \neq J\} = 1 - \sup_{\psi} \mathbb{Q}\{\psi(Z) = J\}$$
$$= 2^{-1}\{1 - \mathrm{TV}(\mathbb{P}_1 \| \mathbb{P}_0)\}.$$

Binary testing and TV distance

- Then

$$\inf_{\psi} \mathbb{Q}\{\psi(Z) \neq J\} = 1 - \sup_{\psi} \mathbb{Q}\{\psi(Z) = J\}$$
$$= 2^{-1}\{1 - \mathrm{TV}(\mathbb{P}_1 \| \mathbb{P}_0)\}.$$

- By using Proposition 1, for any pair of distributions $\mathbb{P}_0, \mathbb{P}_1 \in \mathcal{P}$ such that $\varrho\big(\theta(\mathbb{P}_0), \theta(\mathbb{P}_1)\big) \geq 2\delta$,

$$\mathfrak{M}\big(\theta(\mathcal{P}), \phi \circ \varrho\big) \geq 2^{-1}\phi(\delta)\{1 - \mathrm{TV}(\mathbb{P}_1 \| \mathbb{P}_0)\}.$$

Le Cam's convex hull method

Theorem 5 (Le Cam's convex hull method)

For any 2δ-separated classes of distributions $\mathcal{P}_0$ and $\mathcal{P}_1$ contained within $\mathcal{P}$ in the sense that $\varrho\big(\theta(\mathbb{P}_0), \theta(\mathbb{P}_1)\big) \geq 2\delta$ for all $\mathbb{P}_0 \in \mathcal{P}_0$ and $\mathbb{P}_1 \in \mathcal{P}_1$, one has that

$$\inf_{\hat{\theta}} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\Big\{\varrho\big(\hat{\theta}, \theta(\mathbb{P})\big)\Big\} \geq \delta \sup_{\mathbb{P}_j\in\mathrm{conv}(\mathcal{P}_j), j=0,1} \big\{1 - \mathrm{TV}(\mathbb{P}_1\|\mathbb{P}_0)\big\}.$$

Fano's lower bound

Theorem 6 (Fano)

Let $\{\theta_1, \ldots, \theta_M\}$ be a 2δ-separated set in the ϱ semi-metric on $\theta(\mathcal{P})$ and suppose that J is uniformly distributed over $\{1, \ldots, M\}$, and $Z|J = j \sim \mathbb{P}_{\theta_j}$. Then for any increasing function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$, the minimax risk is lower bounded by

$$\mathfrak{M}\big(\theta(\mathcal{P}), \phi \circ \varrho\big) \geq \phi(\delta)\Big\{1 - \frac{\mathcal{I}(Z, J) + \log 2}{\log M}\Big\},$$

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

Bounds based on local packings

- By using Jensen's inequality, one has that

$$\mathcal{I}(Z, J) \leq M^{-2} \sum_{j=1}^{M} \sum_{k=1}^{M} \mathrm{KL}(\mathbb{P}_{\theta_j} \| \mathbb{P}_{\theta_k}).$$

Bounds based on local packings

- By using Jensen's inequality, one has that

$$\mathcal{I}(Z, J) \le M^{-2} \sum_{j=1}^{M} \sum_{k=1}^{M} \mathrm{KL}(\mathbb{P}_{\theta_j} \| \mathbb{P}_{\theta_k}).$$

- If one can construct a 2δ-separated set of parameter space Ω such that for some $c > 0$, $\mathrm{KL}(\mathbb{P}_{\theta_j} \| \mathbb{P}_{\theta_k}) \le c^2\delta^2$ for all $j \neq k$, then one has that $\mathcal{I}(Z, J) \le c^2\delta^2$.

Bounds based on local packings

- By using Jensen's inequality, one has that

$$\mathcal{I}(Z, J) \leq M^{-2} \sum_{j=1}^{M} \sum_{k=1}^{M} \mathrm{KL}(\mathbb{P}_{\theta_j} \| \mathbb{P}_{\theta_k}).$$

- If one can construct a 2δ-separated set of parameter space Ω such that for some $c > 0$, $\mathrm{KL}(\mathbb{P}_{\theta_j} \| \mathbb{P}_{\theta_k}) \leq c^2\delta^2$ for all $j \neq k$, then one has that $\mathcal{I}(Z, J) \leq c^2\delta^2$.
- By taking $M = M(2\delta)$ such that

$$2^{-1} \log M(2\delta) \geq \log 2 + c^2\delta^2,$$

then

$$\mathfrak{M}(\theta(\mathcal{P}), \phi \circ \varrho) \geq 2^{-1}\phi(\delta).$$

Yang-Barron version of Fano's method

Proposition 7 (Yang-Barron)

Let $\mathcal{N}_{\mathrm{KL}}(\epsilon, \mathcal{P})$ be the ϵ-covering number of $\mathcal{P}$ in the square-root KL divergence. Then

$$\mathcal{I}(Z, J) \leq \inf_{\epsilon > 0} \left\{ \epsilon^2 + \log \mathcal{N}_{\mathrm{KL}}(\epsilon, \mathcal{P}) \right\}.$$

Bounds based on local packings

- By choosing ϵ such that $\epsilon^2 \geq \log \mathcal{N}_{\mathrm{KL}}(\epsilon, \mathcal{P})$, then

$$\frac{\epsilon^2 + \log \mathcal{N}_{\mathrm{KL}}(\epsilon, \mathcal{P}) + \log 2}{\log M} \leq \frac{2\epsilon^2 + \log 2}{\log M}.$$

Bounds based on local packings

- By choosing ϵ such that $\epsilon^2 \geq \log \mathcal{N}_{\mathrm{KL}}(\epsilon, \mathcal{P})$, then

$$\frac{\epsilon^2 + \log \mathcal{N}_{\mathrm{KL}}(\epsilon, \mathcal{P}) + \log 2}{\log M} \leq \frac{2\epsilon^2 + \log 2}{\log M}.$$

- By taking $M = M(2\delta)$ such that

$$2^{-1} \log M(2\delta) \geq 4\epsilon^2 + 2\log 2,$$

then

$$\mathfrak{M}(\theta(\mathcal{P}), \phi \circ \varrho) \geq 2^{-1}\phi(\delta).$$

Thank You