Minimax lowerbounds

Eric Ziebell

May 6, 2025

General comments

1. Warning: This document is dedicated to my personal learning process, and might contain errors.
2. If you find any typos or errors, feel free to make a pull request or contact me directly.
3. If you would simply like to contribute, please reach out as well.

1 The essentials

Definition 1.1. Let $(\mathcal{X}_n, \mathcal{F}_n, (\mathbb{P}_{\vartheta,n})_{\vartheta\in\Theta})$ be a statistical model and (Θ, d) a metric space[1]. Suppose that $(v_n)_{n\in\mathbb{N}}$ is a null sequence. Then, $(v_n)_{n\in\mathbb{N}}$ is called optimal (minimax) convergence rate over Θ if

1. There exists an estimator $\hat{\vartheta}_n^*$ such that
$$\limsup_{n\to\infty} v_n^{-2} \sup_{\vartheta\in\Theta} \mathbb{E}_{\vartheta,n}[d(\hat{\vartheta}_n^*, \vartheta)^2] < \infty.$$
2. We have the uniform lowerbound
$$\liminf_{n\to\infty} v_n^{-2} \inf_{\hat{\vartheta}_n} \sup_{\vartheta\in\Theta} \mathbb{E}_{\vartheta,n}[d(\hat{\vartheta}_n, \vartheta)^2] > 0,$$
where the infimum is taken over all measurable functions (estimators) in model n.

The following reduction scheme can, for instance, be found in Tsybakov [2, Chapter 2].

Proposition 1.2 (Reduction scheme to a testing problem)**.** *Let $(\mathcal{X}_n, \mathcal{F}_n, (\mathbb{P}_{\vartheta,n})_{\vartheta\in\Theta})$ be a statistical model and (Θ, d) a metric space. Suppose that $(v_n)_{n\in\mathbb{N}}$ is a null sequence satisfying an upper bound (1) and*
$$\liminf_{n\to\infty} \inf_{\psi_n} \max_{j=1,\dots,M} \mathbb{P}_{\vartheta_j,n}(\psi_n \neq j) > 0,$$
for parameters $\{\vartheta_1, \dots, \vartheta_M\} \subset \Theta$ separated according to $(v_n)_{n\in\mathbb{N}}$, see (A.1). Then, v_n is the optimal minimax convergence rate over Θ.

Let us first consider the situation where we consider only two hypotheses for the reduction from Proposition 1.2.

Lemma 1.3. *Let $(\mathcal{X}_n, \mathcal{F}_n, (\mathbb{P}_{\vartheta,n})_{\vartheta\in\Theta})$ be a statistical model and (Θ, d) be a metric space. Suppose that $(v_n)_{n\in\mathbb{N}}$ is a null sequence satisfying an upper bound (1) and*
$$\liminf_{n\to\infty} \frac{1}{2}(1 - \|\mathbb{P}_{\vartheta_1,n} - \mathbb{P}_{\vartheta_1+2\alpha v_n,n}\|_{\mathrm{TV}}) > 0.$$
Then, v_n is the optimal minimax convergence rate over Θ.

[1] In general $d : \Theta \times \Theta \to [0, \infty)$ is also allowed to be a semi-distance.

Definition 1.4.

1. *f-divergence*: For a convex function $f : [0, \infty) \to \mathbb{R}$ satisfying $f(1) = 0$ the f-divergence between two probability measures P and Q is given by
$$D_f(P|Q) := \int f\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right) \mathrm{d}Q,$$
provided that P is absolutely continuous with respect to Q and ∞ otherwise.

2. *f-spread*: Let $(\mathcal{X}_n, \mathcal{F}_n, (\mathbb{P}_{\vartheta,n})_{\vartheta\in\Theta})$ be a statistical model. Suppose that f induces a f-divergence. Then, we call
$$\mathcal{J}_f(\Theta, n) := \inf_Q \sum_{\vartheta\in\Theta} D_f(P_\vartheta|Q), \tag{1.1}$$
the f-spread over Θ.[2]

Now it is important to note that usually it does not matter which f-divergence we take here. To see this, we require a slight generalisation of Proposition 1.2. (The original result is recovered by setting $\tilde{\Theta} = \{\vartheta_1, \ldots, \vartheta_M\}$ and π as the discrete uniform. Note also that many ideas from the proof of Proposition 1.2 and Lemma 1.3 are also visible in this result.) The following result provides a lowerbound for the minimax risk, based on an f-divergence and is proven in [1, II.2].

Proposition 1.5. *Let $(\mathcal{X}_n, \mathcal{F}_n, (\mathbb{P}_{\vartheta,n})_{\vartheta\in\Theta})$ be a statistical model and (Θ, d) a metric space. Suppose $\tilde{\Theta} \subset \Theta$ and $((\tilde{\Theta}, d), \Pi, \pi)$ is a polish space. Then, we have*

1. *Then, the minimax risk is lower bounded by the Bayes risk according to π:*
$$\begin{aligned}\liminf_{n\to\infty} v_n^{-2} \inf_{\hat{\vartheta}_n} \sup_{\vartheta\in\Theta} \mathbb{E}_{\vartheta,n}[d(\hat{\vartheta}_n, \vartheta)^2] &\geq \liminf_{n\to\infty} v_n^{-2} \inf_{\psi_n} \int_{\tilde{\Theta}} \mathbb{P}_{\vartheta,n}(\psi_n \neq \vartheta)\mathrm{d}\pi(\vartheta) \\ &= \liminf_{n\to\infty}\left(1 - \int_{\mathcal{X}} \max_{\vartheta\in\tilde{\Theta}}(\pi(\vartheta)p_{\vartheta,n}(x))\mathrm{d}\mu(x)\right).\end{aligned}$$

2. *Suppose that $|\tilde{\Theta}| < \infty$ and $f : [0, \infty) \to \mathbb{R}$ induces and f-divergence according to Definition 1.4. If f is differentiable, we have for every $a \in [0, 1 - 1/|\tilde{\Theta}|]$ a lowerbound for the minimax risk, based on the f-spread within the (reduced) statistical model $(\mathcal{X}_n, \mathcal{F}_n, (\mathbb{P}_{\vartheta,n})_{\vartheta\in\tilde{\Theta}})$:*

   what happens if we allow $|\tilde{\Theta}| = \infty$?
$$\liminf_{n\to\infty}\left(1 - \int_{\mathcal{X}} \max_{\vartheta\in\tilde{\Theta}}(\pi(\vartheta)p_{\vartheta,n}(x))\mathrm{d}\mu(x)\right) \geq \liminf_{n\to\infty}\left(a + \frac{\mathcal{J}_f(\tilde{\Theta}, n) - g_{\tilde{\Theta}}(a)}{g'_{\tilde{\Theta}}(a)}\right),$$
with
$$g(a) = f(|\tilde{\Theta}|(1-a)) + (|\tilde{\Theta}| - 1) f\left(\frac{\tilde{\Theta} a}{|\tilde{\Theta} - 1|}\right),$$
and $\mathcal{J}_f(\tilde{\Theta}, n)$ defined according to (1.1).

Remark 1.6.

1. Usually, the f-spread $\mathcal{J}_f$ cannot be computed exactly and also needs to be bounded.
2. If $\tilde{\Theta} = \{1, 2\}$, we have $\mathcal{J}_f(\{1, 2\}, n) \geq f(1 + \|P_{1,n} - P_{2,n}\|_{\mathrm{TV}}) + f(1 - \|P_{1,n} - P_{2,n}\|_{\mathrm{TV}})$.

References

[1] Adityanand Guntuboyina. Lower bounds for the minimax risk using \$f\$-divergences and applications, 2011. *Preprint*, arXiv: 1002.0042.

[2] Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, 2009.

[2] Note that in this definition it might also be sensible to additionally normalize by $|\Theta|$.

A Proofs for Section 1 (The essentials)

Proposition 1.2 (Reduction scheme to a testing problem)**.** *Let $(\mathcal{X}_n, \mathcal{F}_n, (\mathbb{P}_{\vartheta,n})_{\vartheta\in\Theta})$ be a statistical model and (Θ, d) a metric space. Suppose that $(v_n)_{n\in\mathbb{N}}$ is a null sequence satisfying an upper bound (1) and*

$$\liminf_{n\to\infty} \inf_{\psi_n} \max_{j=1,\dots,M} \mathbb{P}_{\vartheta_j,n}(\psi_n \neq j) > 0,$$

for parameters $\{\vartheta_1, \dots, \vartheta_M\} \subset \Theta$ separated according to $(v_n)_{n\in\mathbb{N}}$, see (A.1). Then, v_n is the optimal minimax convergence rate over Θ.

Proof of Proposition 1.2.

1. With Markov's inequality, we observe that forall $\alpha > 0$, we have

$$v_n^{-2}\mathbb{E}_{\vartheta,n}[d(\hat{\vartheta}_n, \vartheta)^2] \geq \alpha^2 \mathbb{P}_{\vartheta,n}(d(\hat{\vartheta}_n, \vartheta) \geq \alpha v_n),$$

   such that

$$\liminf_{n\to\infty} v_n^{-2} \inf_{\hat{\vartheta}_n} \sup_{\vartheta\in\Theta} \mathbb{E}_{\vartheta,n}[d(\hat{\vartheta}_n, \vartheta)^2] \geq \liminf_{n\to\infty} \inf_{\hat{\vartheta}_n} \sup_{\vartheta\in\Theta} \alpha^2 \mathbb{P}_{\vartheta,n}(d(\hat{\vartheta}_n, \vartheta) \geq \alpha v_n).$$

2. For any subset $\{\vartheta_1, \dots, \vartheta_M\} \subset \Theta$, we obtain the lowerbound

$$\liminf_{n\to\infty} \inf_{\hat{\vartheta}_n} \sup_{\vartheta\in\Theta} \alpha^2 \mathbb{P}_{\vartheta,n}(d(\hat{\vartheta}_n, \vartheta) \geq \alpha v_n) \geq \liminf_{n\to\infty} \inf_{\hat{\vartheta}_n} \max_{j=1,\dots,M} \alpha^2 \mathbb{P}_{\vartheta_j,n}(d(\hat{\vartheta}_n, \vartheta_j) \geq \alpha v_n).$$

3. Suppose that $\{\vartheta_1, \dots, \vartheta_M\} \subset \Theta$ are seperated according to

$$d(\vartheta_i, \vartheta_j) > 2\alpha v_n = \gamma, \quad i \neq j, \quad i, j = 1, \dots, M. \tag{A.1}$$

   Let us now consider the minimum distance test $\psi^* : \mathcal{X}_n \to \{1, \dots, M\}$:

$$\psi^* = \mathrm{argmin}_{k=1,\dots,M} d(\hat{\vartheta}_n, \vartheta_k).$$

   Since the hypothesis are seperated with radius γ, we observe that $\{\hat{\vartheta}_n \in B_\gamma(\vartheta_j)\} \subset \{\psi^* = j\}$. Thus, we clearly have the inclusion $\{\psi^* \neq j\} \subset \{\hat{\vartheta}_n \in B_\gamma(\vartheta_j)\}^c = \{\hat{\vartheta}_n \in B_\gamma(\vartheta_j)^c\}$. However, it is still possible that $\hat{\vartheta}_n$ is the closest to ϑ_j but not inside of a γ ball, so the inclusion might be strict. All in all, we obtain the bound

$$\mathbb{P}_{\vartheta_j,n}(d(\hat{\vartheta}_n, \vartheta_j) \geq \gamma) \geq \mathbb{P}_{\vartheta_j,n}(\psi^* \neq j), \quad j = 1, \dots, M.$$

4. Since for any $\hat{\vartheta}_n$, we can construct such a test, we may replace the infimum over all estimators with an infimum over all tests in model n, yielding the inequality

$$\liminf_{n\to\infty} \inf_{\hat{\vartheta}_n} \max_{j=1,\dots,M} \alpha^2 \mathbb{P}_{\vartheta_j,n}(d(\hat{\vartheta}_n, \vartheta_j) \geq \alpha v_n) \geq \alpha^2 \liminf_{n\to\infty} \inf_{\psi_n} \max_{j=1,\dots,M} \mathbb{P}_{\vartheta_j,n}(\psi_n \neq j).$$

   Thus, if the latter term is positive, we obtain the desired lower bound. □

Lemma 1.3. *Let $(\mathcal{X}_n, \mathcal{F}_n, (\mathbb{P}_{\vartheta,n})_{\vartheta\in\Theta})$ be a statistical model and (Θ, d) be a metric space. Suppose that $(v_n)_{n\in\mathbb{N}}$ is a null sequence satisfying an upper bound (1) and*

$$\liminf_{n\to\infty} \frac{1}{2}(1 - \|\mathbb{P}_{\vartheta_1,n} - \mathbb{P}_{\vartheta_1+2\alpha v_n,n}\|_{\mathrm{TV}}) > 0.$$

Then, v_n is the optimal minimax convergence rate over Θ.

Proof. The result follows from Proposition 1.2 for $M = 2$, by setting $\vartheta_2 = \vartheta_1 + 2\alpha v_n$ and observing

$$\liminf_{n\to\infty} \inf_{\psi_n} \max_{j=1,2} \mathbb{P}_{n,\vartheta_j}(\psi_n \neq j) \geq \liminf_{n\to\infty} \inf_{\psi_n} \frac{1}{2}\left(\mathbb{P}_{\vartheta_1,n}(\psi_n = 2) + \mathbb{P}_{\vartheta_2,n}(\psi_n = 1)\right),$$

where we have used that $\max(a,b) \geq \frac{1}{2}(a+b)$. By going over to the complementary event, we obtain

$$\begin{aligned}
&\frac{1}{2}\left(\mathbb{P}_{\vartheta_1,n}(\psi_n = 2) + \mathbb{P}_{\vartheta_2,n}(\psi_n = 1)\right) \\
&= \frac{1}{2}\left(1 - \mathbb{P}_{\vartheta_1,n}(\psi_n = 1) + 1 - \mathbb{P}_{\vartheta_2,n}(\psi_n = 2)\right) \\
&= 1 - \frac{1}{2}\left(\mathbb{P}_{\vartheta_1,n}(\psi_n = 1) + \mathbb{P}_{\vartheta_2,n}(\psi_n = 2)\right) \\
&= 1 - \int_{\mathcal{X}_n} \frac{1}{2}(\mathbb{1}_{\{\psi_n(x)=1\}} p_1(x) + \mathbb{1}_{\{\psi_n(x)=2\}} p_2(x)) \mathrm{d}\mu(x) \\
&\geq 1 - \frac{1}{2}\int_{\mathcal{X}_n} \max(p_0(x), p_1(x)) \mathrm{d}\mu(x). \\
&= \frac{1}{2}(1 - \|\mathbb{P}_{\vartheta_1,n} - \mathbb{P}_{\vartheta_1+2\alpha v_n,n}\|_{\mathrm{TV}}),
\end{aligned}$$

concluding the proof. $\square$