Statistical Decision Theory

Muchang Bahng

Spring 2025

Contents

1 Minimax Risk

The foundational work was done by Wald in 1949 [Wal49]. In practice, minimax estimation isn't something that you actually write out for a model and then solve for an estimator that achieves this minimax risk. It's not really the point of minimax estimation, nor is it really practical. It is a more abstract diagnosis of determining how difficult a certain problem is, similar to classify problems in P or NP in computer science.

Definition 1.1 (Minimax Risk)

Given a family of probability distributions $\mathscr{P}$ and an estimator θ, the **minimax risk** is defined

$$R = \inf_{\hat{\theta}} \sup_{P \in \mathscr{P}} \mathbb{E}_P \left[L\big(\hat{\theta}, \theta(P)\big) \right] \tag{1}$$

In essence, we want to do best (inf) in the worst case (sup) scenario. This gives you more insight and quantifies this difficulty. Ideally, this risk might go to 0 under some asymptotic conditions, and we would like to know the rate at which it does. We will see that for nonparameteric problems, the rate goes something like

$$\left(\frac{1}{n} \right)^{\frac{2\beta}{2\beta + d}} \tag{2}$$

where β is some measure of smoothness and d is the dimension.

Example 1.1 (Gaussian)

Let $\mathscr{P} = \{N(\theta, 1), \theta \in \mathbb{R}\}$, a family of Gaussians. Then, the sample mean $\hat{\theta} = \bar{X}$ minimizes

$$\inf_{\hat{\theta}} \sup_{\theta} \mathbb{E}[(\hat{\theta} - \theta)^2] \tag{3}$$

Definition 1.2 (Kullback-Leibler Divergence)

Recall that the KL divergence is

$$\mathrm{KL}(P, Q) = \int p \log \frac{p}{q} \tag{4}$$

Definition 1.3 (Total Variation Distance)

The total variation distance

$$\mathrm{TV}(P, Q) = \sup_A |P(A) - Q(A)| \tag{5}$$

which can be rewritten as

$$\frac{1}{2} \int |P(x) - Q(x)| \, dx \tag{6}$$

if we take the measurable set $A = \{x \mid p(x) > q(x)\}$.

References

[Wal49] Abraham Wald. Statistical decision functions. *The Annals of Mathematical Statistics*, 20(2):165–205, 1949.