A short notes on minimax estimator

Tong Zhou*

October 10, 2020

This notes is mainly based on professor Will Fithian's notes on the course STATS 210 at UC Berkeley and section 12.4 of the book titled *All of Statistics* by Larry Wasserman.

The main idea for choosing the minimax estimator for a risk function is to minimize the *worst-case* risk, as opposed to the Bayes estimator which aims to minimize the *average* risk. More specifically, the minimax estimator should satisfy:

$$\delta^* := \arg\min_{\delta} \left(\sup_{\theta} R\left(\theta; \delta\right) \right)$$

We first introduce notations that will be used in the sequel:

1. <u>the minimax risk of the estimation problem:</u> $r^* := \inf_{\delta} \sup_{\theta} R\left(\theta, \delta\right)$.

2. <u>minimax estimator δ^*:</u> if $r^* = \sup_{\theta} R\left(\theta; \delta^*\right)$.

The notation of the minimax risk implicitly reveals a game theory interpretation. We can imagine the agent and the Nature are two players. In this game, the Nature chooses parameter θ to maximize the agent's risk, while the agent chooses estimator δ to minimize the risk. It is not difficult to see the minimax estimator is a Nash-equilibrium in this game.

Compared to Bayes, where the Nature chooses prior distributions as its strategy, in the minimax estimation the Nature chooses parameter. Another contrast between them is from a key observation: the average-case risk from the computation of Bayes risk must be no greater than the worst-case risk implied by the minimax risk. More specifically, given a proper prior Λ, the <u>Bayes risk</u> smaller than the <u>minimax risk</u> can be illustrated by

$$\begin{aligned} r_{\Lambda} &= \inf_{\delta} \int R\left(\theta; \delta\right)\, \mathrm{d}\Lambda(\theta) \\ &\leqslant \inf_{\delta} \sup_{\theta} R(\theta; \delta) = r^*. \end{aligned} \tag{1}$$

*Johns Hopkins University, tzhou11@jhu.edu

If δ_Λ denotes the Bayes estimator, i.e. $r_\Lambda = \int R(\theta;\delta_\Lambda)\,\mathrm{d}\Lambda(\theta)$, then Equation (1) further implies the Bayes risk of any Bayes estimator are lower bounds for r^*.

Another notation will be used is the *least favorable prior* Λ^* which dictates

$$r_{\Lambda^*} \coloneqq \sup_{\Lambda} r_\Lambda$$

Then it immediately follows the relation:

$$\sup_{\theta} R\left(\theta;\delta\right) \geqslant r^* \geqslant r_{\Lambda^*} \geqslant r_\Lambda \tag{2}$$

Note that the δ and Λ are given. Why the second inequality is true? Because $r_\Lambda \leqslant r_*$ for ANY Λ. Taking sup on the LHS gives the second one.

The most important theorem is

Theorem 1. *If $r_\Lambda = \sup_\theta R(\theta;\delta_\Lambda)$ with Bayes estimator δ_Λ then:*

- *δ_Λ is minimax.*
- *If δ_Λ is a unique Bayes (up to a.s.) for Λ, it is the unique minimax.*
- *Λ is the least favorable prior.*

The condition in Theorem 1 is just $\int R(\theta;r_\Lambda)\,\mathrm{d}\Lambda(\theta) = \sup_\theta R(\theta;\delta_\Lambda)$, i.e. the average risk = sup risk. This theorem also delivers arguably the only take-away message from this section: Bayes estimators with a constant risk function are minimax.