1 Minimax Framework

Recall that the Bayes estimator minimizes the average risk

\[ r(\Lambda, \delta) = \int R(\theta, \delta) \, d\Lambda(\theta), \]

where \( \Lambda \) is the prior. One complaint about the Bayes estimator is that the choice of \( \Lambda \) is often arbitrary and hard to defend.

An alternative approach is to assume that nature is malicious, and will always pick the worst value of \( \theta \) in response to the statistician’s choice of \( \delta \). In this case, we instead seek to minimize the maximum risk (or worst-case risk):

\[ \min_{\delta} \sup_{\theta} \int L(\theta, \delta) \, dP_{X \mid \theta}. \]

Claim 1. If the loss function \( L(\theta, \delta) \) is convex in \( \delta \), then finding the minimax rule can be reduced to solving a convex optimization problem.

Why is this claim true? We note that the average of convex functions is convex, hence \( \int L(\theta, \delta) \, dP_{X \mid \theta} \) is convex in \( \delta \). Also, the supremum of convex functions is easily checked to be convex, hence \( \sup_{\theta} \int L(\theta, \delta) \, dP_{X \mid \theta} \) is convex in \( \delta \). As a result, the problem

\[ \min_{\delta} \sup_{\theta} \int L(\theta, \delta) \, dP_{X \mid \theta} \]

aims to minimize a convex function, and is thus a convex optimization problem.

Even though there is substantial literature on solving convex optimization problems, this particular problem is difficult because it’s hard to evaluate the supremum over \( \theta \). Indeed, if we are given a decision rule \( \delta \), we cannot evaluate the gradient because computing \( \sup_{\theta} \) is computationally intractable.

2 Minimax Theorem

The first minimax theorem was proved by von Neumann, in the setting of zero-sum games. It states that every two-person finite-strategy zero-sum game has a mixed-strategy Nash equilibrium. For the purposes of this class, however, we introduce the generalization by Sion and Kakutani:

**Theorem 2** (Sion-Kakutani Minimax Theorem). Let \( X \) and \( \Lambda \) be compact convex sets in linear spaces (i.e., topological vector spaces). Let \( H(\lambda, x) : \Lambda \times X \to \mathbb{R} \) be a continuous function such that

- \( H(\lambda, \cdot) \) is convex for every fixed \( \lambda \in \Lambda \),
- \( H(\cdot, x) \) is concave for any fixed \( x \in X \).

Then

1. The strong duality holds:

\[ \max_{\lambda} \min_{x} H(\lambda, x) = \min_{x} \max_{\lambda} H(\lambda, x). \] (1)
2. There exists a "saddle point" \((\lambda^*, x^*)\), for which

\[ H(\lambda, x^*) \leq H(\lambda^*, x) \leq H(\lambda^*, x^*) \]

for all \(\lambda \in \Lambda, x \in X\). \hspace{1cm} (2)

We omit the proof, but strongly recommend the following exercises:

- Show that (2) implies (1).
- Show that for an arbitrary function \(F(\lambda, x)\), we have

\[
\max_{\lambda} \min_{x} F(\lambda, x) \leq \min_{x} \max_{\lambda} F(\lambda, x).
\]

This is called the weak duality.

There is a game-theoretic way to view weak duality. Two people, named Max and Min, are playing a zero-sum game. Max chooses \(\lambda \in \Lambda\) and Min chooses \(x \in X\), then \(F(\lambda, x)\) is the reward function for Max (and \(-F(\lambda, x)\) is the reward function for Min). Weak duality states that going first is advantageous for Max, no matter the choice of the function \(F\).

3 Applications of the Minimax Problem

3.1 A failed approach

Since we are interested in finding \(\min_\delta \sup_\theta \int L(\theta, \delta) \, dP_x|\theta\), a first approach is to lower bound this quantity using weak duality. The following is slightly sloppy mathematically:

\[
\min_\delta \sup_\theta \int L(\theta, \delta) \, dP_x|\theta \geq \sup_\theta \min_\delta \int L(\theta, \delta) \, dP_x|\theta.
\]

The problem is that in the expression on the RHS, the statistician is allowed to choose \(\delta\) after nature has chosen \(\theta\). If \(L(\theta, g(\theta)) = 0\), the statistician may pick the constant predictor \(\delta = g(\theta)\), so the right hand side is just 0. We have given up too much by applying weak duality.

3.2 Wald’s approach

What if, instead of operating directly on \(\theta\), we instead maximize over priors on \(\theta\)? Recall that the risk function \(R(\theta, \delta)\) is defined to be \(\int L(\theta, \delta) \, dP_x|\theta\).

Claim 3. For any \(\delta\), we have

\[
\sup_\theta R(\theta, \delta) = \sup_{\Lambda(\theta)} \int R(\theta, \delta) \, d\Lambda(\theta).
\]

Proof Since the right hand side is an average of \(R(\theta, \delta)\) over some values of \(\theta\) and the left hand side is the supremum of \(R(\theta, \delta)\), we have LHS \(\geq\) RHS. For the reverse inequality, given any \(\epsilon > 0\), there exists some \(\theta_0\) such that \(R(\theta_0, \delta) + \epsilon > \sup_\theta R(\theta, \delta)\). Taking \(\Lambda_0\) to be the delta distribution with point mass at \(\theta = \theta_0\), we get

\[
\sup_{\Lambda} \int R(\theta, \delta) \, d\Lambda(\theta) \geq \int R(\theta, \delta) \, d\Lambda_0(\theta) = R(\theta_0, \delta) > \sup_\theta R(\theta, \delta) - \epsilon.
\]

Taking \(\epsilon \to 0^+\), we see that RHS \(\geq\) LHS, so we are done. \qed
Thus, maximizing over $\theta$ is the same as maximizing over $\Lambda(\theta)$. We use the minimax theorem:

$$
\min_{\delta} \sup_{\theta} \int L(\theta, \delta) \, dP_{X|\theta} = \min_{\delta} \sup_{\theta} R(\theta, \delta) = \min_{\Lambda} \sup_{\delta} \int R(\theta, \delta) \, d\Lambda(\theta) = \sup_{\Lambda} \min_{\delta} \int L(\theta, \delta) \, dP_{X|\theta} \, d\Lambda(\theta).
$$

We have not verified that the hypotheses of the minimax theorem are satisfied. The main idea is that, when $\Lambda$ is fixed, then this function is convex in $\delta$ (as discussed in the first section when $L(\theta, \cdot)$ is convex), and when $\delta$ is fixed, this function is linear in $\Lambda$. We gloss over the details regarding compactness and convexity, the norms on $X$ and $\Lambda$, continuity, etc.

The above result may be succinctly summarized:

$$
\min_{\delta} \sup_{\theta} R(\theta, \delta) = \sup_{\Lambda} \min_{\delta} r_{\Lambda}(\Lambda, \delta).
$$

Define the Bayes risk

$$
r_{\Lambda} = \min_{\delta} r(\Lambda, \delta) = \min_{\delta} E_{\Lambda}[R(\theta, \delta)]
$$

to be the risk of the Bayes estimator under the prior $\theta \sim \Lambda(\theta)$. Then the above result is $\min_{\Lambda} \sup_{\theta} R(\theta, \delta) = \sup_{\Lambda} r_{\Lambda}$.

The left-hand side’s expression may be interpreted as the statistician choosing $\delta$ first, then nature picking $\theta$, which seems very bad for the statistician. However, this is not as pessimistic as it initially seems: in the right-hand side’s expression, nature acts first by choosing a prior $\Lambda$, then the statistician picks $\delta$.

An objection to Wald’s minimax formulation was that nature is not malicious, but this way of viewing the problem reveals that it is not so pessimistic.

### 3.3 Difficulties

Similar to the fact that the first formulation of the minimax problem is hard to solve, this second form also has its own difficulties. While it is easy to evaluate $r_{\Lambda}$ for each choice of $\Lambda$, there are too many possible choices for $\Lambda$ to evaluate $\sup_{\Lambda} r_{\Lambda}$. Ultimately, we have traded the problem of being unable to evaluate the objective $\sup_{\theta} R(\theta, \delta)$ with the problem of searching over another possibly infinite-dimensional space.

### 4 Minimax in Action

Unfortunately, the minimax problem is difficult to solve exactly, even when it has been reformulated. Very few results of exact optimality are known. Fortunately, statisticians have two ways of dealing with this problem:

- Instead of searching for optimal solutions, we can instead seek potential solutions, which are asymptotically optimal. More on this next week.

- We may upper bound and lower bound the minimax risk by constants. Indeed, in the expression

$$
\min_{\delta} \sup_{\theta} R(\theta, \delta) = \sup_{\Lambda(\theta)} r_{\Lambda},
$$

the left hand side is a minimum over $\delta$, so for any choice $\delta_0$, the quantity $\sup_{\theta} R(\theta, \delta_0)$ is an upper bound on minimax risk. Similarly, the right hand side is a supremum over $\Lambda$, so for any choice of $\Lambda_0$, the
the quantity \( r_{\Lambda_0} \) is a lower bound on the minimax risk. If we get upper and lower bounds which are close, then we are quite satisfied with the result. This is still a difficult task, however.

**Definition 4.** A prior \( \Lambda \) is called least favorable if, for any prior \( \Lambda' \), we have \( r_{\Lambda} \geq r_{\Lambda'} \).

This definition is completely motivated by the minimax equation. The notion is very formal in the sense that it’s usually not possible to check whether \( \Lambda \) is least favorable.

**Theorem 5.** Suppose \( \Lambda(\theta) \) is a prior on \( \theta \), and \( \delta_{\Lambda} \) is the Bayes estimator under \( \Lambda \). Suppose also that \( r(\Lambda, \delta_{\Lambda}) = r_{\Lambda} = \sup_\theta R(\theta, \delta_{\Lambda}) \). Then

1. \( \delta_{\Lambda} \) is the minimax estimator;
2. If \( \delta_{\Lambda} \) is the unique Bayes estimator with respect to \( \Lambda \), then it is the unique minimax estimator;
3. \( \Lambda \) is least favorable.

This theorem is almost tautological. Nevertheless, we prove the first assertion, and leave the rest as exercises.

**Proof**  Let \( \delta \) be any decision rule. Then

\[
\sup_\theta R(\theta, \delta) \geq \int R(\theta, \delta) \, d\Lambda(\theta) \\
\geq \int R(\theta, \delta_{\Lambda}) \, d\Lambda(\theta) \\
= r_{\Lambda} \\
= \sup_\theta R(\theta, \delta_{\Lambda}).
\]

Hence \( \delta_{\Lambda} \) minimizes the minimax risk.

5  Reference

B. Levit (2010). “Minimax Revisited: I, II”. This paper is a survey of several techniques people have used to bound the minimax risk of the following problem: suppose you have a single observation \( X \sim N(\mu, \sigma^2) \), where \( \sigma^2 \) is known, and you know that \( |\mu| \leq 1 \). You want to estimate \( \mu \). The minimax estimator is not known! You can have a feeling of the difficulty in solving the exact minimax estimator!