Chapter 5. Minimax Analysis

- **The Minimax Rule**

Let

$$\mathcal{D}^* = \{\text{all randomized decision rules } \delta^* : \\ R(\theta, \delta^*) < \infty \text{ for all } \theta \in \Theta\}.$$

A decision rule δ_0 is a **minimax decision rule** if it minimizes $\sup_\theta R(\theta, \delta)$ among all decision rules in $\mathcal{D}^*$, i.e., if

$$\sup_{\theta\in\Theta} R(\theta, \delta_0) = \inf_{\delta\in\mathcal{D}^*} \sup_{\theta\in\Theta} R(\theta, \delta^*).$$

The quantity, $\sup_{\theta\in\Theta} R(\theta, \delta_0)$, is called the **minimax value**.

• Admissibility of the Minimax Rule

Theorem 1: *If δ_0 is a unique minimax rule, then δ_0 is admissible.*

Proof: Suppose δ_0 is not admissible. Then, there exists δ_1 such that

$$R(\theta, \delta_1) \leq R(\theta, \delta_0) \ \text{ for all } \theta \in \Theta$$

and

$$R(\theta, \delta_1) < R(\theta, \delta_0)$$

for some θ. It follows that

$$\sup_{\theta \in \Theta} R(\theta, \delta_1) \leq \sup_{\theta \in \Theta} R(\theta, \delta_0),$$

which implies that δ_1 is also minimax. This contradicts the uniqueness of the minimax rule. $\square$

• Determination of Minimax Rules

Theorem 2: *If δ_0 is admissible and has a constant risk function over Θ, then δ_0 is minimax.*

Proof: If δ_0 is not minimax, then there exists δ_1 such that

$$\sup_{\theta \in \Theta} R(\theta, \delta_1) < \sup_{\theta \in \Theta} R(\theta, \delta_0).$$

Since $R(\theta, \delta_0) = c$ for some constant c, then

$$R(\theta, \delta_1) < c = R(\theta, \delta_0)$$

for all $\theta \in \Theta$. This contradicts the admissibility of δ_0. $\square$

Let δ^π be the Bayes rule with respect to π. Then the Bayes risk is

$$r(\pi) = r(\pi, \delta^\pi) = \int_\Theta R(\theta, \delta^\pi) dF^\pi(\theta).$$

Definition: A prior π is said to be *least favorable* if for all prior distribution π',

$$r(\pi) = r(\pi, \delta^\pi) \geq r(\pi') = r(\pi', \delta^{\pi'}).$$

Theorem 3: *Suppose π is a distribution on Θ such that*

$$r(\pi, \delta^\pi) = \int_\Theta R(\theta, \delta^\pi) dF^\pi(\theta) = \sup_{\theta \in \Theta} R(\theta, \delta^\pi).$$

Then

(1) δ^π is minimax.

(2) If δ^π is the unique Bayes rule with respect to π, then it is the unique minimax rule.

(3) π is least favorable.

Note: The condition

$$r(\pi, \delta^\pi) = \int_\Theta R(\theta, \delta^\pi) dF^\pi(\theta) = \sup_{\theta \in \Theta} R(\theta, \delta^\pi)$$

says that the average of $R(\theta, \delta^\pi)$ equals to its maximum. This will happen when the risk function is constant or when π assigns probability 1 to the set on which the risk function taken on its maximum.

Proof: (1) Let δ be any other decision rule. Then

$$\begin{aligned}
&\sup_{\theta\in\Theta} R(\theta,\delta) \geq \int_\Theta R(\theta,\delta)dF^\pi(\theta) \\
&\geq \int_\Theta R(\theta,\delta^\pi)dF^\pi(\theta) = r(\pi,\delta^\pi) \\
&= \sup_{\theta\in\Theta} R(\theta,\delta^\pi).
\end{aligned}$$

Thus, δ^π is minimax.

The proof of (2) is analogous to that of (1) with the replacement of "$\geq$" by "$>$".

(3) Let π' be any other prior distribution on Θ. Then

$$\begin{aligned}
&r(\pi',\delta^{\pi'}) = \int_\Theta R(\theta,\delta^{\pi'})dF^{\pi'}(\theta) \\
&\leq \int_\Theta R(\theta,\delta^\pi)dF^{\pi'}(\theta) \quad (\delta^{\pi'} \text{ is Bayes w.r.t. } \pi') \\
&\leq \sup_{\theta\in\Theta} R(\theta,\delta^\pi) = r(\pi,\delta^\pi).
\end{aligned}$$

This completes the proof. $\square$

Corollary 1 (Theorem 17 of the textbook): *If $\delta_0^* \in \mathcal{D}^*$ is Bayes with respect to $\pi_0 \in \Theta^*$, where Θ^* denotes the set of all π for which $L(\pi, a) < \infty$ for all $a \in \mathcal{A}$, and*

$$R(\theta, \delta_0^*) \leq r(\pi_0, \delta_0^*)$$

for all $\theta \in \Theta$, then δ^* is minimax and π_0 is least favorable.

Proof: If

$$R(\theta, \delta_0^*) \leq r(\pi_0, \delta_0^*),$$

then

$$r(\pi_0, \delta_0^*) = \sup_{\theta \in \Theta} R(\theta, \delta_0^*).$$

Thus, Corollary 1 directly follows from Theorem 3. $\square$

Corollary 2 (Equalizer Rules): *If a Bayes rule δ^π has constant risk, then it is minimax.*

Corollary 3 (Sub-Equalizer Rules): *Let*

$$W_\pi = \{\theta : \ R(\theta, \delta^\pi) = \sup_{\theta' \in \Theta} R(\theta, \delta^\pi)\}.$$

If $P^\pi(W_\pi) = 1$, then δ^π is minimax.

The proofs of the above two corollaries are straightforward.

Let π_n be a sequence of prior distributions and also let δ_n is the Bayes rule with respect to π_n. Write

$$r_n = r(\pi_n, \delta_n) = \int_\Theta R(\theta, \delta_n) dF^{\pi_n}(\theta).$$

Assume

$$r = \lim_{n \to \infty} r_n < \infty.$$

Definition: A sequence π_n is said to be least favorable if for every π,

$$r(\pi, \delta_\pi) \leq r.$$

Theorem 4: *Suppose π_n is a sequence of prior distributions with Bayes risk $r_n \to r$, and δ is a decision rule such that*

$$\sup_{\theta\in\Theta} R(\theta,\delta) \le r.$$

Then

(1) δ is minimax.

(2) the sequence π_n is least favorable if the equality holds.

Proof: (1) Suppose δ' is any other decision rule. Then

$$\sup_{\theta\in\Theta} R(\theta,\delta') \ge \int_\Theta R(\theta,\delta')dF^{\pi_n}(\theta) \ge r_n$$

for any n. Thus,

$$\sup_{\theta\in\Theta} R(\theta,\delta') \ge \lim_{n\to\infty} r_n = r \ge \sup_{\theta\in\Theta} R(\theta,\delta),$$

which implies that δ is minimax.

(2) Let π denote any prior distribution. Then

$$r_\pi = r(\pi,\delta_\pi) = \int_\Theta R(\theta,\delta^\pi)dF^\pi(\theta)$$

$$\le \int_\Theta R(\theta,\delta)dF^\pi(\theta) \le \sup_{\theta\in\Theta} R(\theta,\delta) = r.$$

Hence, $\{\pi_n\}$ is least favorable. $\square$

Example 1: Suppose that $X \sim N(\theta, 1)$ and that it is desired to estimate θ under squared-error loss. We seek to prove that the usual estimator, $\delta_0(x) = x$, is minimax.

Solution:

Method 1: By the Bylth's theorem, we have shown that δ_0 is admissible. Since

$$R(\theta, \delta_0) = E_\theta[(\theta - \delta_0)^2] = 1,$$

which is constant. Thus, Theorem 2 gives that δ_0 is admissible.

Method 2: Suppose that we cannot apply the Bylth's theorem. Let $\pi(\theta) = 1$.Then δ_0 is the generalized Bayes rule. Consider a sequence of proper prior $\pi_n = N(0, n)$. Then,

$$\delta_n = \left(\frac{n}{n+1}\right) x,$$

and

$$\begin{aligned} r_n =& r(\pi_n, \delta_n) = \int_{-\infty}^{\infty} R(\theta, \delta_n)\pi_n(\theta)d\theta \\ =& \int_{-\infty}^{\infty} \frac{1}{(n+1)^2}\left(n^2 + \theta^2\right)\pi_n(\theta)d\theta = \frac{n}{n+1}. \end{aligned}$$

Thus,

$$\lim_{n\to\infty} r_n = 1.$$

Since $R(\theta, \delta_0) = 1$, $\sup_\theta R(\theta, \delta_0) = 1$. Theorem 4 leads to that $\delta_0 = x$ is minimax and the sequence $\{N(0, n)\}$ is least favorable. $\square$

Example 2: Assume $X \sim \mathcal{B}(n, \theta)$ is observed, and that it is desired to estimate θ under squared-error loss. Then, $\delta = \frac{x}{n}$ is admissible and it is also UMVUE. Is δ minimax?

Solution: How to show that $\delta = \frac{x}{n}$ is admissible?

To examine whether $\delta = \frac{x}{n}$ is minimax, we find an equalizer rule of the form $\delta(x) = ax + b$. Clearly,

$$\begin{aligned} R(\theta, \delta) &= E_\theta[(aX + b - \theta)^2] \\ &= E_\theta[\{a(X - n\theta) + b + (an - 1)\theta\}^2] \\ &= a^2 n\theta(1-\theta) + [b + (an-1)\theta]^2 \\ &= \theta^2[-a^2 n + (an-1)^2] + \theta[a^2 n + 2b(an-1)] + b^2. \end{aligned}$$

For the risk to be constant in θ, we must have

$$-a^2 n + (an-1)^2 = 0$$

and

$$a^2 n + 2b(an-1) = 0.$$

Solving these equations for a and b gives

$$a = (n + \sqrt{n})^{-1} \text{ and } \ b = \sqrt{n}/[2(n + \sqrt{n})].$$

Thus,

$$\delta_0(x) = ax + b = \frac{x + \sqrt{n}/2}{n + \sqrt{n}}$$

is an equalizer rule.

To complete the argument, we must show that δ_0 is Bayes. It is easy to see that if $\theta \sim \mathcal{B}e(\alpha, \beta)$, then the Bayes estimator is

$$\frac{x + \alpha}{n + \alpha + \beta}.$$

Hence, the equalizer rule $\delta 0$ is clearly of this form with

$$\alpha = \beta = \sqrt{n}/2.$$

Hence δ_0 is Bayes, and Corollary 2 gives that δ_0 is minimax and the least favorable prior is $\mathcal{B}e(\sqrt{n}/2, \sqrt{n}/2)$. Since δ_0 is δ, then $\delta = x/n$ cannot be minimax.

Straightforward calculation yields

$$R(\theta, \delta_0) = \frac{1}{[4(1 + \sqrt{n})^2]}$$

and

$$R(\theta, \delta = x/n) = \frac{\theta(1-\theta)}{n}.$$

From Figure 5.9 on page 375 of the textbook, we can see that the minimax rule compares very favorably with $\delta(x) = x/n$ for small n but rather unfavorably for large n.