Lecture 3 Minimax and More on Admissibility

3.1 Minimax rules

The minimax principle presents an alternative and conservative approach to summarize risk functions. It compares decision rules based on "the worst case scenario". $d \in \mathcal{D}$ is called a *minimax rule* if

$$\sup_{\theta \in \Theta} R(\theta, d) = \min_{d' \in \mathcal{D}} \sup_{\theta \in \Theta} R(\theta, d').$$

Note: In some sense, the minimax principle and admissibility complement each other. An admissible rule d usually has a low risk at a certain value of θ, which is hard to beat, but d can also yield a high risk at some other value of θ; a minimax rule d precisely prevents such a high risk from happening even though the more rosy scenario of d need not look very attractive. Therefore, a minimax and admissible rule d appears desirable. Although the Bayesian paradigm is drastically different from the minimax principle, it can help us find minimax rules.

Theorem 1 (i) *If d^π is a Bayes rule that satisfies $R(\theta, d^\pi) \leq B(\pi, d^\pi)\ \forall\ \theta \in \Theta$, then d^π is a minimax rule.*

(ii) *Suppose for every $k = 1, 2, \ldots,$ $d_k = d^{\pi_k}$ is a Bayes rule with respect to prior π_k. If d satisfies*

$$R(\theta, d) \leq c = \lim_{k \to \infty} B(\pi_k, d_k) \quad \forall\ \theta \in \Theta,$$

then d is a minimax rule.

Proof: It suffices to show (ii). By contradiction, suppose d is not a minimax rule, i.e. there exists d' such that

$$\sup_{\theta \in \Theta} R(\theta, d') < \sup_{\theta \in \Theta} R(\theta, d).$$

Hence there exist $\epsilon = \epsilon(d, d') > 0$ and $k = k(\epsilon)$ such that

$$B(\pi_k, d') \leq \sup_{\theta \in \Theta} R(\theta, d') < \sup_{\theta \in \Theta} R(\theta, d) - \epsilon \leq c - \epsilon \leq B(\pi_k, d_k),$$

contradicting to "d_k is a Bayes rule with respect to π_k". *QED.*

Note that the condition in case (i) appears too restrictive because the inequality in an opposite direction, $B(\pi, d^\pi) \leq \sup_{\theta\in\Theta} R(\theta, d^\pi)$, always holds. Two definitions are needed before introducing the next result.

- d is called an *equalizer rule* if $R(\theta, d)$ is constant (does not depend on θ).
- d is called an *extended Bayes rule* if there exists a sequence of priors π_k, $k = 1, 2, ...$ such that

$$\lim_{k\to\infty} B(\pi_k, d) = \lim_{k\to\infty} B(\pi_k, d_k).$$

Corollary 1 *If d is an equalizer and an extended Bayes, then d is minimax.*

A recipe for finding a minimax rule d:

Step 1: Make a guess on an equalizer d with $R(\theta, d) \equiv c$, and propose a sequence of priors $\{\pi_k\}$;

Step 2: Find a Bayes rule $d_k = d^{\pi_k}$ for each π_k;

Step 3: Check if $\lim_{k\to\infty} B(\pi_k, d_k) = c$.

Example 3.1 Let $X_1, ..., X_n$ be iid $N(\theta, \sigma^2)$ random variables where $\sigma > 0$ is known. For estimation of θ with the squared error loss, show that $\bar{X}$ is minimax.

Proof: First, $\bar{X}$ is an equalizer: $E_\theta(\bar{X} - \theta)^2 = \sigma^2/n$. Next, we show that $\bar{X}$ is an extended Bayes. Consider conjugate prior $\pi_k(\theta) \sim N(0, k)$, $k = 1, 2, ...$ which implies

$$\begin{aligned}
B(\pi_k, d_k) &= \int_{\mathcal{X}} r(x^n, d_k(x^n))\ m(x^n)\ dx^n \\
&= \int_{\mathcal{X}} \underbrace{\int_\Theta (d_k(x^n) - \theta)^2\ \pi_k(\theta|x^n)\ d\theta}_{\text{posterior variance}}\ m(x^n)\ dx^n \\
&= \int_{\mathcal{X}} \frac{k\sigma^2/n}{k + \sigma^2/n}\ m(x^n)\ dx^n \\
&= \frac{k\sigma^2/n}{k + \sigma^2/n} \stackrel{k\to\infty}{\longrightarrow} \sigma^2/n.
\end{aligned}$$

QED.

Note:

- An example of a non-equalizer minimax rule: Let $X \sim N(\theta, 1)$ with $\theta \in [-b, b]$ where $b \in (0, 1)$ is known. Estimate θ with the squared error loss. Define

$$d(x) = b\ tanh(bx) = b\ \frac{e^{bx} - e^{-bx}}{e^{bx} + e^{-bx}}, \quad x \in I\!R.$$

Verify the following claims:

- ∘ d is not an equalizer.
- ∘ For the "two-point" prior $\pi(b) = \pi(-b) = 1/2$, $d(x) = E(\theta|x)$, thus d is a Bayes estimator under π.
- ∘ $R(\theta, d) \leq B(\pi, d)$, $\forall\ \theta \in [-b, b]$.

Hence d is minimax by Theorem 1 (i).

- Although (exact) minimax rules based on finite samples may not be appealing these days, the study of asymptotic minimax procedures is an important area in nonparametric statistics.

3.2 Finding admissible rules via Bayes rules

One way to find admissible rules is to use Bayes rules and approximations.

Theorem 2 *Let Θ be a subset of Euclidean space $\mathbb{R}^k$. Assume that for every decision rule d of interest, the risk function $R(\theta, d)$ is continuous in θ. Suppose the prior $\pi(\theta)$ on Θ is a continuous distribution which satisfies $\pi(U_\theta(\epsilon)) > 0\ \forall \epsilon > 0, \theta \in \Theta$, where $U_\theta(\epsilon)$ is the ϵ-neighborhood of θ in Θ. Then a Bayes rule d^π is admissible.*

Proof: By contradiction. Suppose $\exists\ d'$ that dominates d^π. Then $R(\theta_0, d^\pi) - R(\theta_0, d') \triangleq b > 0$ for some θ_0. The continuity of $R(\theta, d)$ in θ implies that $R(\theta, d^\pi) - R(\theta, d') \geq b/2$ $\forall \theta \in U_{\theta_0}(\epsilon)$ and for some $\epsilon > 0$. Therefore,

$$B(\pi, d^\pi) - B(\pi, d') \geq \int_{U_{\theta_0}(\epsilon)} [R(\theta, d^\pi) - R(\theta, d')]\ \pi(\theta)\ d\theta \geq \frac{b}{2}\ \pi(U_{\theta_0}(\epsilon)) > 0,$$

which contradicts that d_π is a Bayes rule. *QED.*

Remarks:

(a) The case of discrete priors over Θ can be treated similarly.

(b) Theorem 2 can be generalized by considering a sequence of Bayes rules and their limit. Instead of stating such a general result, we illustrate it by the following example.

Example 3.2 Let $X_1, ..., X_n$ be iid $N(\theta, \sigma^2)$ random variables, where $\sigma > 0$ is a known constant while θ is an unknown parameter. Estimate θ with the squared error loss. We will show that $\bar{X}$ is admissible. In fact, we can reduce the problem to the simple case $n = 1$ by using sufficient statistics (a topic next week). Hence we assume $X \sim N(\theta, \sigma^2)$ and claim that X is admissible.

For each $k = 1, 2, ...$, introduce the conjugate prior $\pi_k \sim N(0, k)$. Note that as $k \to \infty$, the Bayes estimator

$$d_k(x) \stackrel{\triangle}{=} d^{\pi_k}(x) = [1 + \sigma^2/k]^{-1} x \longrightarrow x.$$

Suppose $d(X) = X$ is inadmissible, then $\exists\ d'$ better than d.

(i) For some $\theta_0 \in I\!R$, $\epsilon > 0$ and $b > 0$,

$$B(\pi_k, d) - B(\pi_k, d') \geq \int_{\theta_0-\epsilon}^{\theta_0+\epsilon} [R(\theta, d) - R(\theta, d')]\ \pi_k(\theta)\ d\theta \geq b\ \int_{\theta_0-\epsilon}^{\theta_0+\epsilon} \pi_k(\theta)\ d\theta > 0,$$

but we need a finer result because the lower bound tends to 0 with a rate $\frac{1}{\sqrt{k}}$ as $k \to \infty$.

(ii) The Bayes risk

$$B(\pi_k, d_k) = \int_{\mathcal{X}} \underbrace{\int_{\Theta} [d_k(x) - \theta]^2 \pi_k(\theta|x)\ d\theta}_{\text{posterior variance}}\ m(x)\ dx = \frac{k\sigma^2}{k + \sigma^2},$$

thus

$$B(\pi_k, d_k) - B(\pi_k, d) = \frac{k\sigma^2}{k + \sigma^2} - \sigma^2 = \frac{-\sigma^4}{k + \sigma^2}.$$

Combining (i) and (ii), we have

$$\sqrt{k}\ [B(\pi_k, d_k) - B(\pi_k, d')] \geq \frac{-\sqrt{k}\ \sigma^4}{k + \sigma^2} + b\sqrt{k}\ \int_{\frac{\theta_0-\epsilon}{\sqrt{k}}}^{\frac{\theta_0+\epsilon}{\sqrt{k}}} \frac{e^{-z^2/2}}{\sqrt{2\pi}}\ dz \stackrel{k\to\infty}{\longrightarrow} \sqrt{\frac{2}{\pi}}\ b\ \epsilon > 0,$$

which contradicts that d_k is a Bayes rule with respect to π_k. $\quad QED.$