Minimax Procedures

MIT 18.655

Dr. Kempthorne

Spring 2016

Outline

Minimax Procedures

Decision Problem: Basic Components

- $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$: parametric model.
- $\Theta = \{\theta\}$: Parameter space.
- $\mathcal{A}\{a\}$: Action space.
- $L(\theta, a)$: Loss function.
- $R(\theta, \delta) = E_{X|\theta}[L(\theta, \delta(X))]$

Minimax Criterion

- Two decision procedures δ_1 and δ_2 in $\mathcal{D}$.
  δ_1 is preferred to δ_2 if
  $$sup_{\theta\in\Theta} R(\theta, \delta_1) < \sup_{\theta\in\Theta} R(\theta, \delta_2)$$
- δ^* is **minimax** if
  $$sup_{\theta\in\Theta} R(\theta, \delta^*) \leq \sup_{\theta\in\Theta} R(\theta, \delta) \text{ for all } \delta \in \mathcal{D}$$
  i.e.,
  $$sup_{\theta\in\Theta} R(\theta, \delta^*) = \inf_{\delta\in\mathcal{D}} [\sup R(\theta, \delta)]$$

Outline

Minimax Procedures: Game Theory

Two-Person Zero-Sum Game

- *Nature*(N) : chooses $\theta \in \Theta$, using distribution $\pi(\cdot)$.
- *Statistician*(S) : chooses $\delta \in \mathcal{D}$.
- Outcome of Game: Payoff paid by S to N

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

Partial-Information Case I

- S chooses δ first
- N specifies $\pi(\cdot)$ knowing δ
- S knows that N will have knowledge of δ.

Optimal strategies:

- Given δ, N will specify π as π_δ:

$$r(\pi_\delta, \delta) = \sup_\pi r(\pi, \delta)), \quad \pi_\delta \text{ is } \textbf{least favorable} \text{ to } \delta.$$

- Given knowlege of N's strategy, S will choose δ^*:

$$r(\pi_{\delta^*}, \delta^*) = \sup r(\pi, \delta^*) = \inf_\delta \sup r(\pi, \delta)$$

Claim: δ^* is **minimax**

- For any prior π and decision procedure δ
$$r(\pi, \delta) = \int_\Theta R(\theta, \delta)\pi(d\theta) \le \sup_\theta R(\theta, \delta)$$
- Given any δ, the least-favorable prior to δ (π_δ) gives positive weight only to those θ_*:
$$R(\theta_*, \delta) = \sup_\theta R(\theta, \delta).$$
It follows that
$$\sup_\pi r(\pi, \delta) = r(\pi_\delta, \delta) = \sup_\theta R(\theta, \delta).$$
- Player S will choose δ^* such that
$$\begin{aligned} r(\pi_{\delta^*}, \delta^*) &= \sup_\pi r(\pi, \delta^*) = \sup_\theta R(\theta, \delta^*) \\ &= \inf_\delta \sup_\pi r(\pi, \delta) = \inf_\delta \sup_\theta R(\theta, \delta) \end{aligned}$$
a **minimax** procedure.

Partial-Information Case II

- N chooses π first.
- S specifies δ knowing π.
- N knows that S will have knowledge of π.

Optimal strategies:

- Given π, S will choose the Bayes procedure δ_π.
$$r(\pi, \delta_\pi) = \inf_\delta r(\pi, \delta)$$
- Given knowlege of S's strategy, N will choose π^*:
$$r(\pi^*, \delta_{\pi^*}) = \sup_\pi r(\pi, \delta_\pi) = \sup_\pi \inf_\delta r(\pi, \delta).$$
π^* is the **Least Favorable Prior Distribution**.

Minimax Procedures: Game Theory

Theorem 3.3.1 (von Neumann). For the Two-Person Zero-Sum Game define:

- The *Lower Value* of the Game is
  $$\underline{v} \equiv \sup_{\pi} \inf_{\delta} r(\pi, \delta)$$
- The *Upper Value* of the Game is
  $$\overline{v} \equiv \inf_{\delta} \sup_{\pi} r(\pi, \delta)$$

If Θ and $\mathcal{D}$ are finite, then

- the *least favorable* π_* and *minimax* δ^* exist and
  $$\underline{v} = r(\pi^*, \delta^*) = \overline{v}$$
- $\delta^* = \delta_{\pi^*}$, the Bayes procedure for prior π^*
- $\pi* = \pi_{\delta*}$, the least-favorable prior against $\delta*$

Outline

Minimax Theorems

Theorem 3.3.2 Suppose δ^* is such that

- $sup_\theta R(\theta, \delta^*) = r < \infty$.
- There exists a prior π^* such that:

  δ^* is Bayes for π^*

  $\pi^*\{\theta : R(\theta, \delta^*) = r\} = 1$

Then

δ^* is minimax.

Proof. This theorem follows from the following two propositions.

Proposition 3.3.2 π_δ is least favorable against δ if and only if

$$\pi_\delta\{\theta : R(\theta, \delta) = \sup_{\theta'} R(\theta', \delta)\} = 1.$$

Only maximal-risk points θ get positive weight from π_δ

Proof: π_δ is least favorable against δ iff

$$r(\pi_\delta, \delta) = \sup_\pi r(\pi, \delta) = \sup_\theta R(\theta, \delta),$$

which is true iff the condition of the proposition is satisfied.

Proposition 3.3.1 Suppose the procedure δ^{**} and the prior π^{**} can be found such that

$$\delta^{**} \text{ is Bayes for } \pi^{**}$$
$$\pi^{**} \text{ is least-favorable against } \delta^{**}.$$

Then

$$\underline{v} = \overline{v} = r(\pi^{**}, \delta^{**})$$

and

π^{**} is a **least favorable prior**

δ^{**} is a **minimax** procedure.

Proof:

First we show that $\underline{v} \leq \overline{v}$

1. $\inf_\delta r(\pi, \delta) \leq r(\pi, \delta')$, for all π, δ'.
2. Take sup_π of both sides:
   $\underline{v} \equiv \sup_\pi \inf_\delta r(\pi, \delta) \leq \sup_\pi r(\pi, \delta')$, for all δ'
3. Take $inf_{\delta'}$ of both sides:
   $\underline{v} \equiv \sup_\pi \inf_\delta r(\pi, \delta) \leq \inf_{\delta'} \sup_\pi r(\pi, \delta') \equiv \overline{v}$

Proof (continued):
Second we show that $\underline{v} \geq \overline{v}$

1. By definition:
$$\underline{v} \equiv \sup_\pi \inf_\delta r(\pi, \delta) \geq \inf_\delta r(\pi', \delta), \text{ for all } \pi'$$
2. For $\pi' = \pi^{**}$ this gives
$$\underline{v} \geq \inf_\delta r(\pi^{**}, \delta) = r(\pi^{**}, \delta^{**}).$$
3. By definition of π^{**}
$$r(\pi^{**}, \delta^{**}) = \sup_\pi r(\pi, \delta^{**})$$
But
$$\sup_\pi r(\pi, \delta^{**}) \geq \inf_\delta \sup_\pi r(\pi, \delta) \equiv \overline{v}$$
4. Putting 2 and 3 together we have:
$$\underline{v} \geq \overline{v}.$$

Minimax Example

Minimax Estimation of Binomial Probability (Case I)

- For a sample $(X_1, \ldots, X_n)$ iid $Bernoulli(\theta)$, $S = \sum_1^n X_i$ is sufficient and $S \sim Binomial(n, \theta)$.
- Relative squared-error loss:
$$L(\theta, a) = \frac{(\theta - a)^2}{\theta(1-\theta)}, \ 0 < \theta < 1.$$
- $\delta(S) = S/n = \overline{X}$ has constant risk
$$R(\theta, \overline{X}) = \frac{1}{\theta(1-\theta)} E[(\overline{X} - \theta)^2] = \frac{1}{\theta(1-\theta)} \times [\frac{\theta(1-\theta)}{n}] = \frac{1}{n}$$
- Since $\overline{X}$ is Bayes for the prior $\theta \sim \pi = \beta(1,1)$, it must be that
  $\overline{X}$ is minimax, and
  $\pi = Unif(0,1)$ is least favorable.

Minimax Example

Minimax Estimation of Binomial Probability (Case II)

- For a sample $(X_1, \ldots, X_n)$ iid $Bernoulli(\theta)$, $S = \sum_1^n X_i$ is sufficient and $S \sim Binomial(n, \theta)$.
- Ordinary squared-error loss:
$$L(\theta, a) = (\theta - a)^2 \ , \ 0 < \theta < 1.$$
- $\delta^*(S) = \frac{S + \frac{1}{2}\sqrt{n}}{n + \sqrt{n}} = \frac{\sqrt{n}}{\sqrt{n}+1}\overline{X} + \frac{1}{\sqrt{n}+1}\frac{1}{2}$
has constant risk
- $\delta^*(S)$ is Bayes for $\pi = \beta(\sqrt{n}/2, \sqrt{n}/2)$
- From the theorem it follows that
$\delta^*(S)$ is minimax, and
$\pi = \beta(\sqrt{n}/2, \sqrt{n}/2)$ is least favorable.

Note: as $\lim_{n\to\infty} \frac{R(\theta, \delta^{**})}{R(\theta, \overline{X})} > 1$, for all $\theta \neq 1/2$, and
$$\lim_{n\to\infty} \frac{R(\theta, \delta^{**})}{R(\theta, \overline{X})} = 1, \text{ for l } \theta = 1/2.$$

Theorem 3.3.3 Let δ^* be a rule such that

$$\sup_\theta R(\theta, \delta^*) = r < \infty.$$

Let $\{\pi_k\}$ be a sequence of prior distributions with Bayes risks

$$r_k = \inf_\delta r(\pi_k, \delta).$$

If

$$\lim_{k \to \infty} r_k = r,$$

then

δ^* is minimax.

Proof:

Consider any other procedure δ. It must be that

$$\sup_\theta R(\theta, \delta) \geq E_{\pi_k}[R(\theta, \delta)] \geq r_k$$

Taking the limit as $k \to \infty$, it follows that

$$\sup_\theta R(\theta, \delta) \geq \lim_{k \to \infty} r_k = r = \sup_\theta R(\theta, \delta^*).$$

Thus, δ^* is minimax.

Example 3.3.3 $\overline{X}$ is minimax for estimating a $Normal(\theta, \sigma^2)$ mean under squared error loss.

- $R(\theta, \overline{X}) = \sigma^2/n$.
- $\pi_k = N(\eta_0, \tau^2 = k)$.
- Bayes risk $r_k = [\frac{n}{\sigma^2} + \frac{1}{k}]^{-1}$.
- $\lim_{k \to \infty} r_k = \sigma^2/n$

It follows that $\overline{X}$ is minimax.

Example 3.3.4 Minimaxity of $\overline{X}$ in Nonparametric Model.

- $X_1, \ldots, X_n$ iid $P \in \mathcal{P}$.
- $\mathcal{P} = \{P : Var_P(X_i) \leq M\}$
- Decision problem: estimate $\theta(P) = E_P(X_i)$ with squared-error loss.

Apply Theorem 3.3.3: define a sequence of prior distributions $\{\pi_k\}$ such that

$$r_k = \inf_{\delta} r(\pi_k, \delta) \to r,$$

where

$$r = \max_P R(P, \overline{X}).$$

Define π_k :

- π_k gives positive weight only to $P : Var_P(X_i) = M$.
  $$\pi_k(\{P : Var_P(X_i) < M\}) = 0.$$
- π_k gives positive weight only to $P : P = N(\mu, M)$ for some μ.
  $$\pi_k(\{P : P \neq N(\mu, M), \text{ for some } \mu\}) = 0.$$
- π_k is a Gaussian mixture of Gaussian distributions

- π_k is a Gaussian mixture of Gaussian distributions
  Let $\mu(P) = E[X_i \mid P]$ be the mixing parameter:
  $$\mu \sim N(0, k) \text{ and}$$
  $$P \mid \mu = N(\mu, M)$$
  Note: the prior-predictive distribution of all X_i is $N(0, M + k)$.

The problem is identical to Example 3.3.3 with $\sigma^2 = M$, and $\eta_0 = 0$. It follows that

$$r_k = [\tfrac{n}{M} + \tfrac{1}{k}]^{-1} \to M/n = R(P, \overline{X})$$

18.655 Mathematical Statistics
Spring 2016