Minimax Estimation

Outline

1) Minimax risk, estimator
2) Least favorable priors
3) Examples

Minimax risk

Last idea for choosing an estimator: worst-case risk

$$\underset{\delta}{\text{minimize}} \quad \sup_{\theta} \quad R(\theta;\delta)$$

The minimum achievable sup-risk is called the <u>minimax risk</u> of the estimation problem

$$r^* = \inf_{\delta} \sup_{\theta} R(\theta;\delta)$$

An estimator δ^* is called <u>minimax</u> if it achieves the minimax risk, i.e.

$$\sup_{\theta} R(\theta;\delta^*) = r^*$$

Game theory interpretation: (Minimax game)

1) Analyst chooses estimator δ
2) Nature chooses parameter θ to max. risk

<u>NB</u>: Nature chooses θ adversarially, not X

Compare to Bayes ("Maximin" game)

1) Nature chooses prior Λ (mixed strategy)
2) Analyst chooses estimator to min. (avg) risk

We will look for Nature's Nash-equil. strategy in this problem

Least Favorable Priors

Key observation: average-case risk $\leq$ worst-case risk

For proper prior Λ, the Bayes risk is

$$r_\Lambda = \inf_\delta \int R(\theta;\delta)\, d\Lambda(\theta)$$
$$\leq \inf_\delta \sup_\theta R(\theta;\delta) = r^*$$

If δ_Λ Bayes then $r_\Lambda = \int R(\theta;\delta_\Lambda)\, d\Lambda(\theta)$

$\Rightarrow$ Bayes risk of any Bayes estimator lower bounds r^*

Least favorable prior Λ^* gives best lower bound: $r_{\Lambda^*} = \sup_\Lambda r_\Lambda$

Sup-risk of any estimator upper bounds r^*

$$\underbrace{r_\Lambda}_{\text{(any } \Lambda)} \leq r_{\Lambda^*} \leq r^* \leq \underbrace{\sup_\theta R(\theta;\delta)}_{\text{(any } \delta)}$$

Idea: try to match upper & lower bounds

<u>Theorem</u> If $r_\Lambda = \sup_\theta R(\theta; \delta_\Lambda)$ with Bayes estimator δ_Λ then:

(a) δ_Λ is minimax

(b) If δ_Λ is unique Bayes (up to $\overset{a.s.}{=}$) for Λ, it is unique minimax

(c) Λ is least fav.

<u>Proof</u> a) Any other δ:

$$
\begin{aligned}
\sup_\theta R(\theta; \delta) &\geq \int R(\theta; \delta)\, d\Lambda(\theta) \\
&\geq \int R(\theta; \delta_\Lambda)\, d\Lambda(\theta) \qquad (*) \\
&= r_\Lambda \\
&= \sup_\theta R(\theta; \delta_\Lambda) \quad \text{by assumption}
\end{aligned}
$$

$\Rightarrow$ r_Λ is minimax risk, δ_Λ is minimax,

b) Replace "$\geq$" with "$>$" in 2^{nd} ineq. $(*)$

c) Any other prior $\tilde{\Lambda}$:

$$
\begin{aligned}
r_{\tilde{\Lambda}} &= \inf_\delta \int R(\theta; \delta)\, d\tilde{\Lambda}(\theta) \\
&\leq \int R(\theta; \delta_\Lambda)\, d\tilde{\Lambda}(\theta) \\
&\leq \sup_\theta R(\theta; \delta_\Lambda) = r_\Lambda \qquad \boxtimes
\end{aligned}
$$

The above theorem gives a checkable condition:

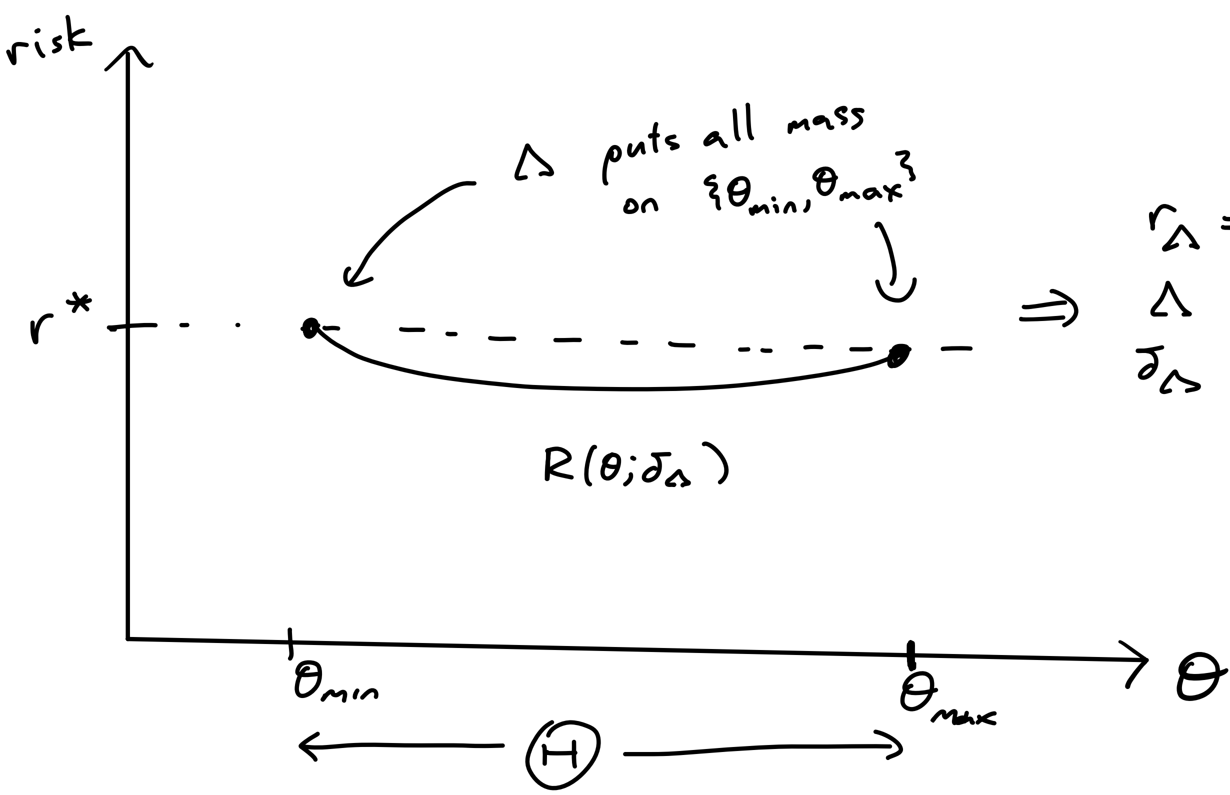
does avg risk = sup risk ?

True if :

mistake on final: saying r_Λ is const. doesn't prove anything

1) $R(\theta; \delta_\Lambda)$ is constant

2) $\Lambda\left(\{\theta : R(\theta; \delta_\Lambda) = \max_{\xi} R(\xi; \delta_\Lambda)\}\right) = 1$



$\Rightarrow$ $r_\Lambda = r^*$

Λ least fav.

δ_Λ minimax

Example (Binomial)

$X \sim \mathrm{Binom}(n, \theta)$, estimate θ, sq. err.

Try $\mathrm{Beta}(\alpha, \beta)$, hope to get one with const. risk

$$\delta_{\alpha,\beta}(X) = \frac{\alpha + X}{\alpha+\beta+n}.$$ Try $\alpha = \beta$ (symmetric)

$$\begin{aligned}
\mathrm{MSE}(\theta; \delta_{\alpha,\alpha}) &= \left(\frac{\alpha+\theta n}{2\alpha+n} - \theta\right)^2 + (2\alpha+n)^{-2}\,\mathrm{Var}_\theta(X) \\
&= (2\alpha+n)^{-2}\left[\alpha^2(1-2\theta)^2 + n\theta(1-\theta)\right] \\
&= (2\alpha+n)^{-2}\left[(4\alpha^2-n)\theta^2 - (4\alpha^2-n)\theta + \alpha^2\right]
\end{aligned}$$

(set $\alpha^* = \sqrt{n}/2$)

$$= \frac{n/4}{(n+\sqrt{n})^2}$$

$$\Rightarrow \quad \delta^*(x) = \frac{\sqrt{n}/2 + X}{\sqrt{n} + n}$$

$$r^* = n/4(n+\sqrt{n})^2$$

$$\Lambda^* = \mathrm{Beta}(\sqrt{n}/2, \sqrt{n}/2)$$

$$\propto_\theta \theta^{\sqrt{n}/2 - 1}(1-\theta)^{\sqrt{n}/2 - 1}$$

Q: why concentrated at $\theta = 1/2$?

Least Favorable Sequence

Sometimes there is no least favorable prior,

$X \sim N(\theta, 1)$: LF prior should spread mass everywhere, but that is not a proper prior.

Def: A sequence $\Lambda_1, \Lambda_2, \ldots$ is LF if $r_{\Lambda_n} \longrightarrow \sup_{\Lambda} r_{\Lambda}$

Thm: Suppose $\Lambda_1, \Lambda_2, \ldots$ is a prior sequence and δ satisfies $\sup_{\theta} R(\theta; \delta) = \lim_n r_{\Lambda_n}$

Then a) δ is minimax

b) $\Lambda_1, \Lambda_2, \ldots$ is LF

Proof a) Other est. $\tilde{\delta}$. Then $\forall n$,

$$\sup_{\theta} R(\theta; \tilde{\delta}) \geq \int R(\theta; \tilde{\delta}) \, d\Lambda_n(\theta)$$

$$\geq r_{\Lambda_n}$$

$$\Rightarrow \sup_{\theta} R(\theta; \tilde{\delta}) \geq \sup_n r_{\Lambda_n}$$

$$\geq \lim_n r_{\Lambda_n}$$

$$= \sup_{\theta} R(\theta; \delta)$$

b) Prior Λ

$$r_\Lambda = \int R(\theta;\delta_\Lambda)\, d\Lambda(\theta)$$
$$\leq \int R(\theta;\delta)\, d\Lambda(\theta)$$
$$\leq \sup_\theta R(\theta;\delta)$$
$$= \lim_n r_{\Lambda_n}$$

$\boxtimes$

<u>Ex</u> $X \sim N_d(\theta, I_d)$. Est. θ, use MSE

$\delta_0(X) = X$ unbiased, $MSE = d$

$\zeta = \frac{1}{1+n}$

Try prior Λ_n: $\theta_i \overset{iid}{\sim} N(0,n) \Rightarrow \delta_{\Lambda_n}(x) = (1-\zeta)X$

$$MSE(\theta;\delta_{\Lambda_n}) = \zeta^2\|\theta\|^2 + (1-\zeta)^2 d$$

$$r_{\Lambda_n} = \mathbb{E}\, MSE(\theta)$$
$$= \zeta^2 nd + (1-\zeta)^2 d$$
$$= d\left(\frac{n}{(1+n)^2} + \frac{n^2}{(1+n)^2}\right)$$
$$\to d$$

So $\Lambda_1, \Lambda_2, \ldots$ is LF

$\delta_0(X)$ is minimax (but inadmissible?)

$r^* = d$

Bounding minimax risk

Our theorem gives an idea of how to bound r^* for a problem:

Upper bound: If δ is any estimator then

$$r^* \le \sup_\theta R(\theta;\delta) \qquad (= \text{ if } \delta \text{ minimax})$$

Lower bound: If Λ is any prior then

$$r^* \ge \int R(\theta;\delta_\Lambda)\, d\Lambda(\theta) \qquad (= \text{ if } \Lambda \text{ LF})$$

Minimax estimators are very hard to find but minimax bounds are often used in stat theory to characterize hardness (esp. lower)

Ex: Propose practical estimator δ, find Λ for which r_Λ close to $\sup_\theta R(\theta;\delta)$
(or same rate, or cvgs asymptotically)

$\Rightarrow$ Conclude δ can't be improved "much" (*)

Ex: Quantify hardness of a problem by its minimax rate in some asy. régime.

Caveat: A problem might be easy throughout most of par. space but very hard in some bizarre corner you never encounter in practice!