Minimax Estimation

Outline

1) Minimax risk, estimator

2) Least favorable priors

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 minimax risk of the estimation problem

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

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

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

Game theory interpretation:

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

NB: Nature chooses θ adversarially, not X

Compare to Bayes, where Nature chooses prior from a known distribution

$\Rightarrow$ Nature plays a specific mixed strategy

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

Least Favorable Priors

Minimax closely related to Bayes

<u>Key observation:</u> 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 <u>any</u> Bayes estimator lower bounds r^*

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

Sup-risk of any estimator upper bounds r^*

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

Theorem 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.

Proof a) Any other δ:

$$\sup_\theta R(\theta;\delta) \geq \int R(\theta;\delta)\, d\Lambda(\theta)$$

$$\geq \int R(\theta;\delta_\Lambda)\, d\Lambda(\theta) \qquad (*)$$

$$= r_\Lambda$$

$$= \sup_\theta R(\theta;\delta_\Lambda) \quad \text{by assumption}$$

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

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

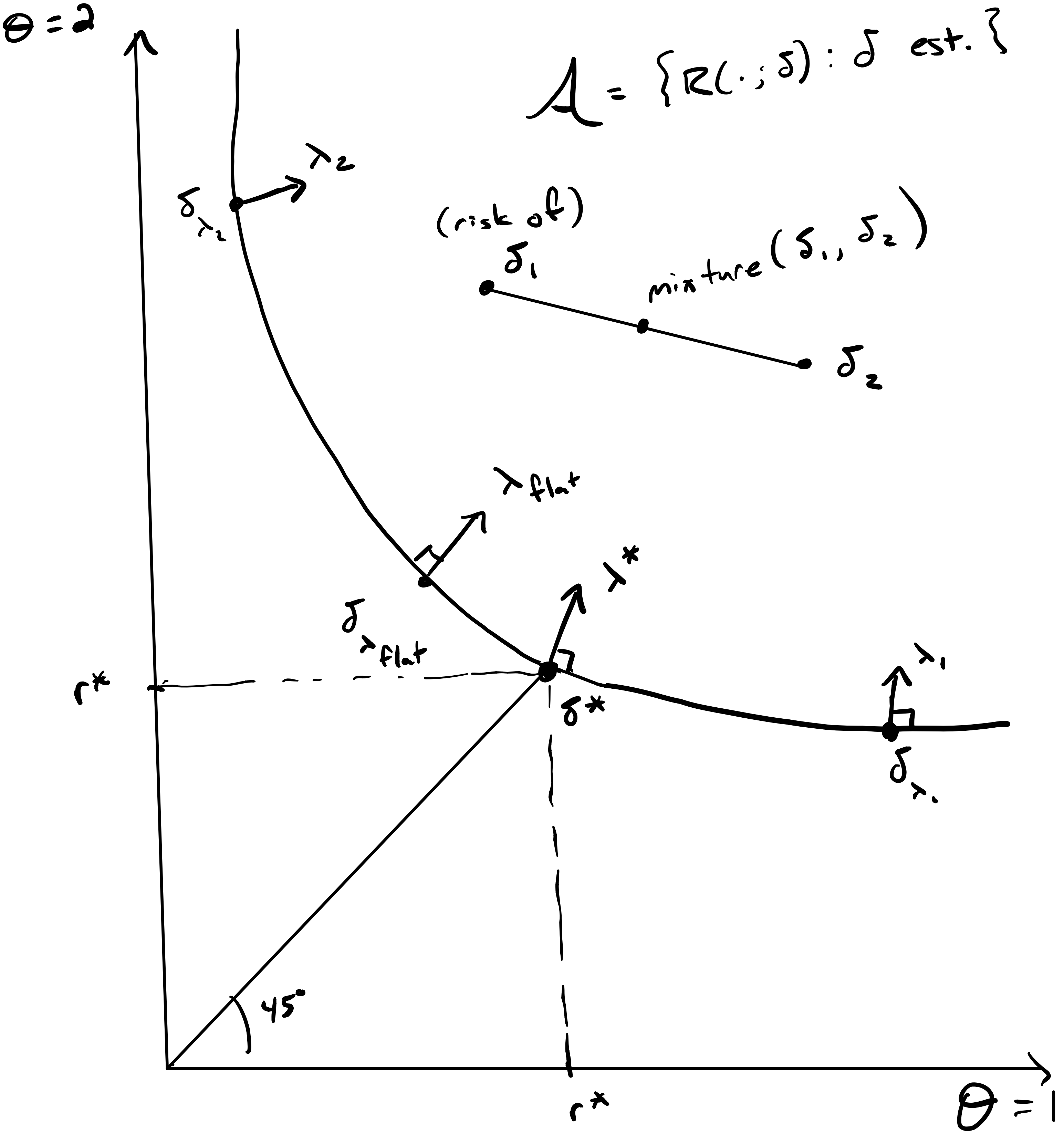
c) Any other prior $\tilde{\Lambda}$:

$$r_{\tilde{\Lambda}} = \inf_\delta \int R(\theta;\delta)\, d\tilde{\Lambda}(\theta)$$

$$\leq \int R(\theta;\delta_\Lambda)\, d\tilde{\Lambda}(\theta)$$

$$\leq \sup_\theta R(\theta;\delta_\Lambda) = r_\Lambda \qquad \boxtimes$$

Picture $\Theta = \{1, 2\}$



The above theorem gives a checkable condition:

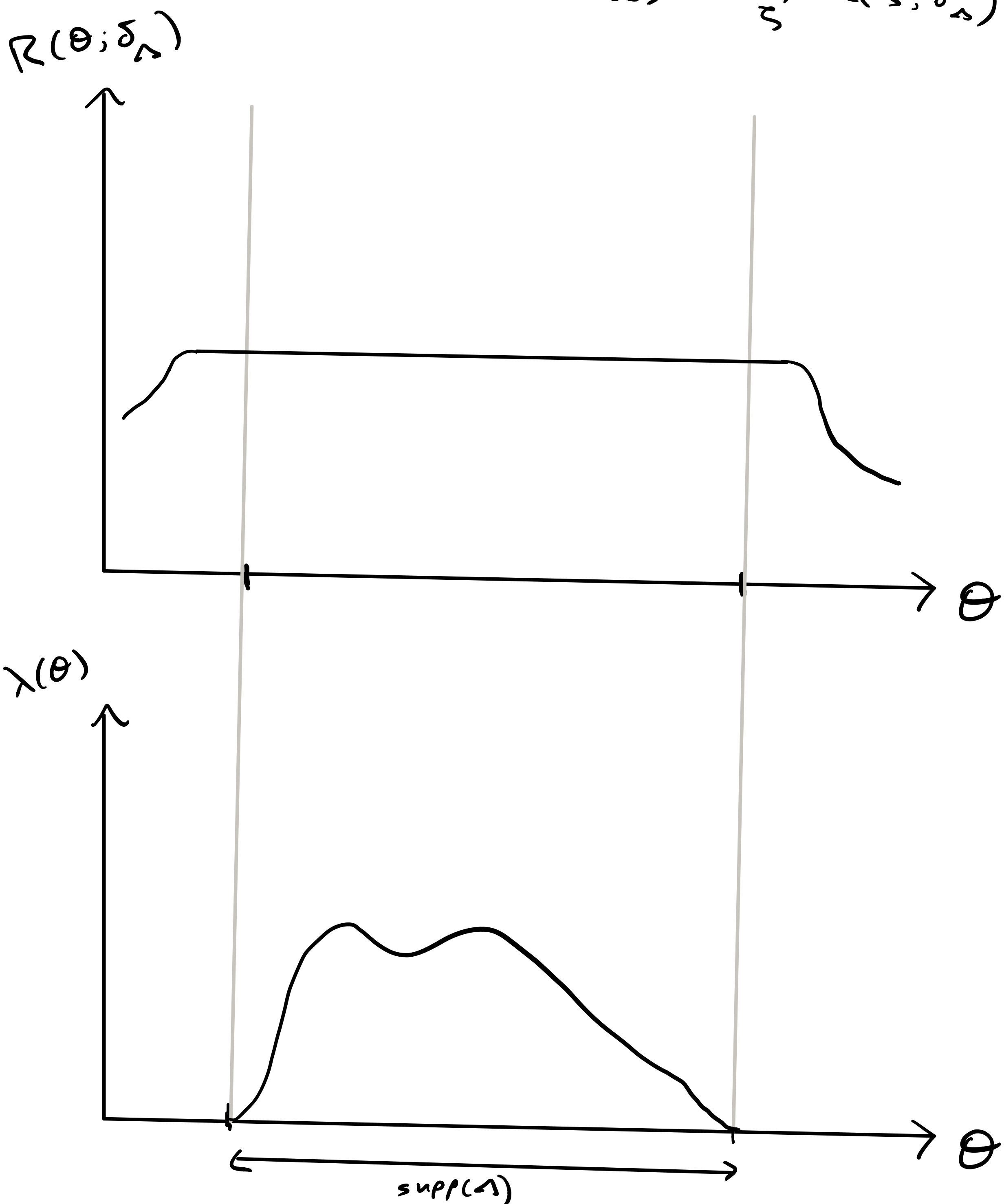
does avg risk = sup risk ?

True if :

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

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

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



Example (Binomial)

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

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

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

$$R(\theta; \delta_{\alpha,\beta}(X)) = \mathbb{E}_\theta\left[\left(\frac{\alpha+X}{\alpha+\beta+n} - \theta\right)^2\right]$$

$$= \text{Var}_\theta\left(\frac{X}{\alpha+\beta+n}\right) + \left(\frac{\alpha+\theta n}{\alpha+\beta+n} - \theta\right)^2$$

$$= (\alpha+\beta+n)^{-2} \cdot \left[n\theta(1-\theta) + (\alpha - (\alpha+\beta)\theta)^2\right]$$

$$\propto_\theta \underbrace{\left[(\alpha+\beta)^2 - n\right]}_{\text{Set } = 0}\theta^2 + \underbrace{\left[n - 2\alpha(\alpha+\beta)\right]}_{\text{Set } = 0}\theta + \alpha^2$$

Set $(\alpha+\beta)^2 = n$, $2\alpha(\alpha+\beta) = n$

$\Rightarrow \alpha + \beta = \sqrt{n} \Rightarrow 2\alpha\sqrt{n} = n$

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

$\Longrightarrow \text{Beta}\left(\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2}\right)$ is LF

$\dfrac{X + \sqrt{n}/2}{n + \sqrt{n}}$ is minimax ✓

We got lucky!

Bounding minimax risk

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

<u>Upper bound</u>: If δ is any estimator then

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

<u>Lower bound</u>: 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)

<u>Ex</u>: 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" (*)

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

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

Least Favorable Sequence

Sometimes there is no least favorable prior, e.g. if par. space isn't compact.

$X \sim N(\theta, 1)$: 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>Basic Picture:</u>

$\sup_\theta R(\theta; \delta)$ generic δ

$\geq \inf_\delta \sup_\theta R(\theta; \delta)$ $\Big(= \sup_\theta R(\theta; \delta^*)$ if minimax est. exists$\Big)$

$\geq \sup_\Lambda r_\Lambda$ $\Big(= r_{\Lambda^*}$ if LF prior exists$\Big)$

$\geq r_\Lambda$ generic Λ