Chap 1. Wald decision theory

Instructor: Fang Han

$$X \in \mathcal{X}$$

data $\quad$ data space

P_θ $\leftarrow$ data generating dist.
non-random

If θ is finite dimensional $\Rightarrow$ parametric model

If θ is infinite dimensional $\Rightarrow$ non-parametric model

θ is usually unknown

Decision theory:

How to make the "best" decision based on (X, P_θ)

Notation

$x \longleftarrow$ realization of Data

$a \longleftarrow$ action

$\mathcal{A} \longleftarrow$ action space

$D(\cdot | x) \longleftarrow$ the action to be made based on the particular data realization x

Remark:

this action can be $\boxed{\text{random}}$

$T(x) \longleftarrow$ this corresponds to the case for the $\boxed{\text{Non-random}}$ decision.

Goal: To make decision based on what we observed:

"decision is some action a based on x"

① The decision based on $X = x$ is denoted by $D(\cdot | x)$ which is a conditional probability dist. of the action given $X = x$.

Sometimes we use

$$\hat{T}_x : x \longmapsto \mathcal{J}$$

Emphasize it is a random function

to represent the random decision sampled from $D(\cdot | x)$

② In some cases, the decision can be a deterministic function, i.e. $D(\cdot | x)$ is the probability of a point mass. Then

$$T : x \longmapsto \mathcal{J}$$

without hat

③ The "decision function" space:

$\mathcal{D}$: the class of all possible random decisions to be considered.

$\mathcal{J}$: Class of all deterministic decisions to be considered.

Example (Point estimation)

(X, P_θ) : to "estimate" a certain transformation $\psi(\theta)$

with $\boxed{\psi}$ known

θ unknown

based on X.

Normal mean estimation

$P_\theta = N(\mu, \sigma^2)^{\otimes n}$ → product measure

$\theta = (\mu, \sigma^2)$

$\psi((\mu, \sigma^2)) = \mu$.

Then an action is a certain real value.

the action space is $\mathcal{A} = \mathbb{R}$

Q: How to compare different decisions?

To set up a loss function. (Inevitable)

Loss function :

$L(a, \theta)$ which describes the quality of the decision a at $\theta \in \Theta$.

$$L : \mathcal{A} \times \Theta \longmapsto \mathbb{R}_+$$

We expect a smaller loss leading to a better decision.

e.g. In point estimation problem,

$$L(a, \theta) = [a - \psi(\theta)]^2$$

With the loss function, we can introduce risk, about certain decision $D(\cdot | x)$ in the decision space $\mathcal{D}$.

$$R(D, \theta) := \boxed{\int_{\mathcal{X}} \boxed{\int_{\mathcal{A}} L(a, \theta) D(da | x)} \, dP_\theta}$$

expected loss w.r.t. "randomness" of random decision

expected loss w.r.t. the data randomness.

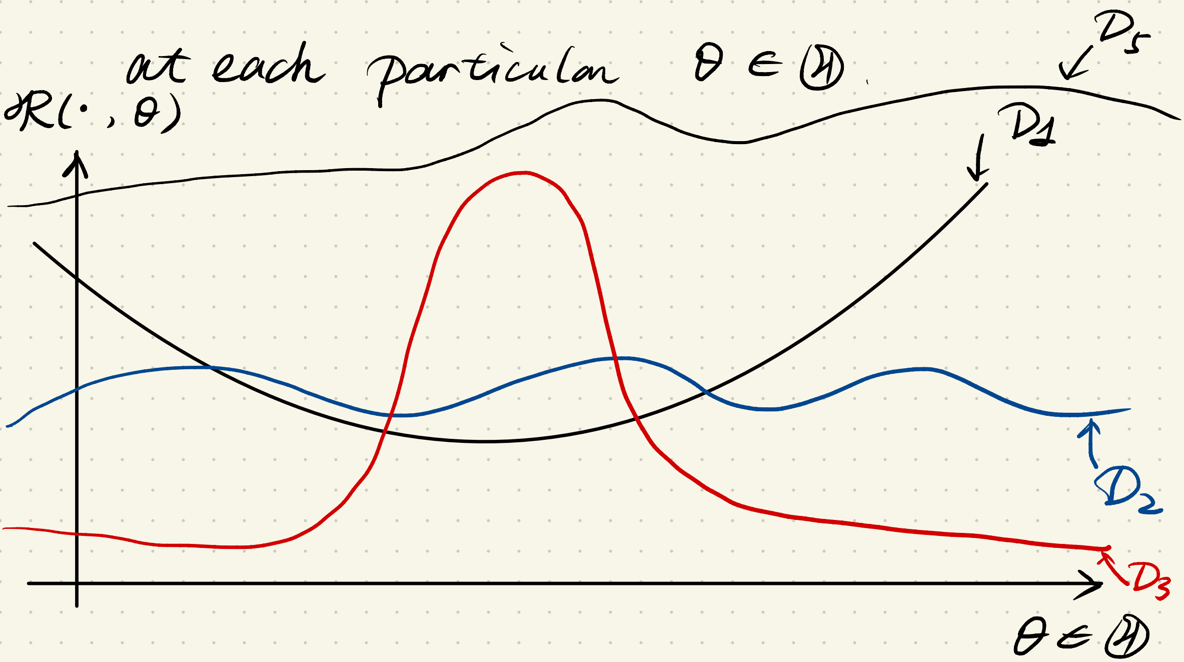
$$= \mathbb{E}\left[\mathbb{E}\left[L(\hat{T}_x, \theta) \mid X = x\right]\right]$$

$$= \mathbb{E}\left[L(\hat{T}, \theta)\right]$$

Tower Law

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X]$$

Since $\mathcal{R}(D, \theta)$ [← deterministic] is non-random, we can compare the performance of different D's at each particular $\theta \in \Theta$.



1st criterion

"admissibility"

D_5 is inadmissible and should be bad.

$$
\begin{tikzcd}
\hat{T} & \text{randomness from } X \arrow[l] & \text{random decision} \arrow[ll, dashed, bend left] \\
X \arrow[u] & \text{random data} \arrow[l] & \\
P_\theta \arrow[u] & \text{DGD} \arrow[l] & \\
\mathcal{P} \arrow[u] & \text{DFamily} \arrow[l] &
\end{tikzcd}
$$

$$
\mathcal{R}(\overset{\hat{T}}{D}, \theta) = \left[E\{ L(\hat{T}, \theta) \} \right]
$$

To compare different $\hat{T}$'s (D's)

① Admissiblity

A decision D is admissible if it is NOT dominated by another decision

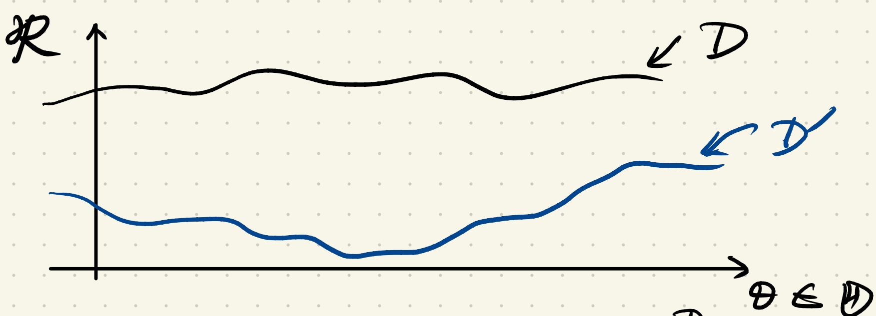
i.e.,

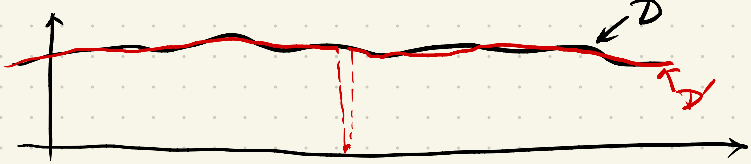
there does NOT exist another decision $D' \in \mathcal{D}$ s.t.

(1) $\mathcal{R}(D, \theta) \geq \mathcal{R}(D', \theta)$ $\forall \theta \in \Theta$

(2) $\exists\, \theta_0 \in \Theta$ s.t.

$$\mathcal{R}(D, \theta) > \mathcal{R}(D', \theta)$$



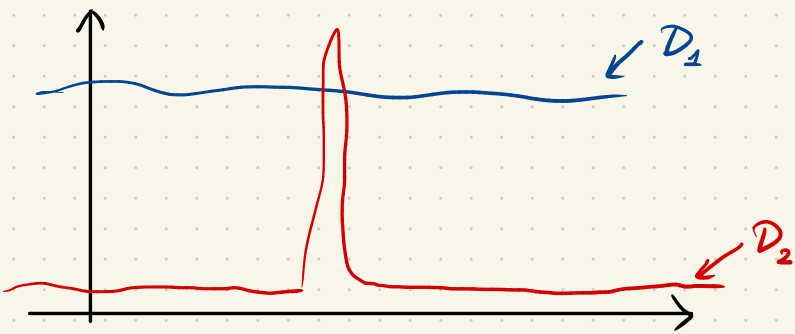


② $\underbrace{\text{Mini}}_{\text{best}}\underbrace{\text{max}}_{\text{worst case}}$ [Pathethic conservative]

Philosophy:

Prefer D_1 over D_2 in a minimax framework if

$$\sup_{\theta \in \Theta} \mathcal{R}(D_1, \theta) < \sup_{\theta \in \Theta} \mathcal{R}(D_2, \theta)$$

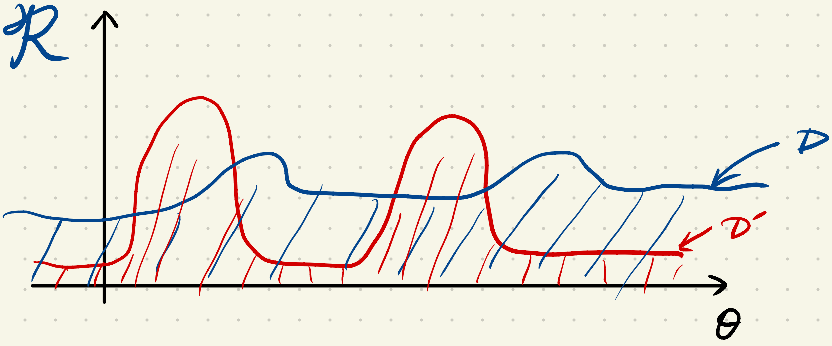


Then, a minimax rule is a decision $D^* \in \mathcal{D}$

s.t. $$\sup_{\theta \in \Theta} \mathcal{R}(D^*, \theta) = \underbrace{\inf_{D \in \mathcal{D}}}_{\text{mini}} \underbrace{\sup_{\theta \in \Theta}}_{\text{max}} \mathcal{R}(D, \theta).$$

③ Think about what θ is coming from.

Bayesian decision theory



Bayesian : compare the "average" performance w.r.t.

$$P(\theta) \sim \pi \quad (\leftarrow \text{prior})$$

over Θ.

$$\boxed{\begin{aligned} & r(D,\pi) \\ & := \int \mathcal{R}(D,\theta)\,\pi(d\theta) \end{aligned}}$$

In Bayesian, we can think θ is a realization of $\Theta \sim \pi$

$$r(D,\pi) = \mathbb{E}\left[L(\hat{T}, \Theta)\right]$$

$$\mathbb{E}\left[\begin{array}{c} \hat{T} \\ \uparrow \\ X \\ \uparrow \\ P_\theta \\ \uparrow \\ \pi \end{array}\right]$$

Then, the Bayes rule then is the decision D_π s.t.

$$r(D_\pi, \pi) = \inf_{D \in \mathcal{D}} r(D, \pi)$$

Result 1.

Suppose $\Theta \sim \Pi$

$$X \mid \Theta = \theta \sim P_\theta$$

$$\hat{T} \sim D(\cdot \mid x)$$

$$L(\cdot \mid \cdot) \geq 0$$

Then if one decision $D_\pi \in \mathcal{D}$ satisfies it minimizes

$$E[L(\hat{T}_{D_\pi}, \Theta) \mid X = x]$$

for all $x \in \mathcal{X}$

Claim: D_π is a Bayes rule.

Result 2 Assume

① $a \mapsto L(a,\theta)$ is convex for all $\theta \in \Theta$;

② $\mathcal{D}$ is unrestricted

③ $\mathcal{A}$ is a convex set

④ $\exists$ at least one Bayes rule

Assume Jensen can be applied

Claim: then $\exists$ a non-random Bayes rule.

Proof: Using Jensen

$$[f \text{ convex} \Rightarrow \mathbb{E}f(X) \geq f(\mathbb{E}X)]$$

a Bayes rule

$$r(D_\pi, \pi)$$

$$= \mathbb{E}[L(\hat{T}_{D_\pi}, \pi)]$$

$$= \mathbb{E}[\mathbb{E}(L(\hat{T}_{D_\pi}, \theta) \mid \Theta = \theta, X = x)]$$

$$\geq \mathbb{E}\left[L\left(\mathbb{E}(\hat{T}_{D\pi} \mid \Theta = \theta, X = x), \theta\right)\right]$$

① is a non-random rule

② If $\hat{T}$ is Bayes rule, is also a Bayes rule.

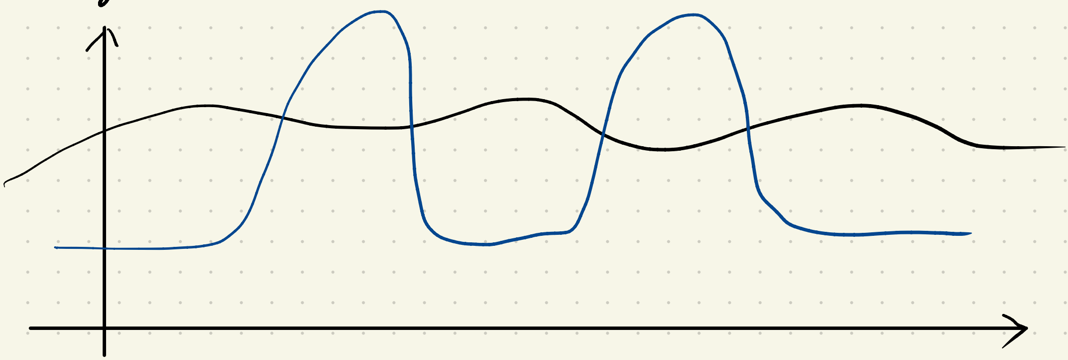
Result 3 Built result 1 & 2,

it suffices to consider a T that is minimizing

$$\mathbb{E}[L(T,\theta) \mid X = x]$$

$$= \underbrace{\int L(T,\theta)\, \underbrace{p(\theta \mid x)}_{\text{posterior dist.}}\, d\theta}_{\text{posterior risk}}$$

Bayes risk



averaging the pointwise risk

w.r.t. your preference π

prior dist.

$$\boxed{\begin{array}{l} P_\theta \sim X \longleftarrow D \text{ or } \hat{T} \text{ or } T \\ \quad \wr \\ \quad \pi \end{array}}$$

$$r(D, \pi) := \mathbb{E}[L(D, \theta)]$$

$\uparrow$

Bayes risk for D w.r.t. π

<u>Claim 1</u>. minimizing $\mathbb{E}[L(D, \theta) \mid X = x]$ is sufficient to minimize $r(D, \pi)$

<u>Claim 2</u>. As long as Jensen, there exists a non-random Bayes rule.

<u>Claim 3</u>. To find a non-random Bayes rule, it suffices to find the T that minimizes $\mathbb{E}[L(T, \theta) \mid X = x]$

i.e. find the T that minimizes $\mathbb{E}[\cdot \mid \cdot]$ w.r.t. $p(\theta \mid x)$.

Example [point estimation, square error loss]

$$\Theta \sim \Pi$$

$$X \mid \Theta = \theta \sim P_\theta$$

$$L(a, \theta) = (a - \underbrace{\psi(\theta)}_{\text{known}})^2$$

Q: What is the Bayes rule ???

A: It suffices to consider

$$T: \mathcal{X} \mapsto \mathcal{A}$$

Then we only have to find the one

T that minimizes

$$\mathbb{E}[L(T, \Theta) \mid X = x]$$

$$= \mathbb{E}[(\underbrace{T}_{a = T(x)} - \psi(\Theta))^2 \mid X = x]$$

In other words, we only have to find an a s.t.

$$\mathbb{E}[(a-\psi(\Theta))^2 \mid X=x]$$

is minimized

Claim: $a = \underbrace{\mathbb{E}[\psi(\Theta) \mid X=x]}_{\text{if it exists.}}$

Proof:

$$\frac{d}{da}\mathbb{E}[(a-\psi(\Theta))^2 \mid X=x]$$

$$\overset{\uparrow}{=} \mathbb{E}\left[\frac{d}{da}(a-\psi(\Theta))^2 \mid X=x\right]$$

measure theory

$$= \mathbb{E}[2(a-\psi(\Theta)) \mid X=x]$$

$$\text{LHS}=0 \Leftarrow a = \mathbb{E}[\psi(\Theta) \mid X=x]$$

Q: How to calculate the posterior expectation?

A: It is highly non-trivial.

By Bayes,

$$\theta \sim \pi \quad (\leftarrow \pi(\theta))$$

$$X \mid \theta = \theta \sim P_\theta(x)$$

$$\|$$

Q: $\theta \mid X = x$?

A: $P(X, \theta) = P(X \mid \theta)\,\pi(\theta)$

$$\Rightarrow P(\theta \mid x) = \frac{P(X \mid \theta)\,\pi(\theta)}{P(x)}$$

$$= \frac{P(X \mid \theta)\,\pi(\theta)}{\int P(X \mid \theta)\,\pi(\theta)\,d\theta} \leftarrow \text{usually no closed form.}$$

Example of conjugate prior.

Example (Poisson – Gamma) L^2, estimate θ

By prev. discussion, the Bayes rule is

$$E[\theta \mid X = x]$$

w. $\theta \mid X = x \sim \mathrm{Gamma}(\alpha + x, \beta + 1)$

$\Rightarrow$ Bayes rule

$$T_{\pi} : x \longmapsto \frac{\alpha + x}{\beta + 1}$$

observe

prior

$$= \frac{1}{\beta+1} x + \frac{\beta}{\beta+1}\left(\frac{\alpha}{\beta}\right)$$

Example (Point, L1 loss)

$$L(a, \psi(\theta)) = |a - \psi(\theta)|$$

$\Rightarrow$ the Bayes rule will be posterior median

Remarks ① Bayes rule may not, usually not unique.

② Bayes rule may not exist.

③ Bayes rule is not necessarily admissible

Bayes procedure

Peter Hoff

Subjective (Bayesian philosophy)

Pragmatic (Decision Theory)

Minimax rule

D^* is one of the rules that minimises the maximal risk:

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

Q: How to find them?

Answer 1: Information-theoretic:

Guess which one is minimax

+ establishing the minimax risk

Answer 2: building connection to Bayes rule.

(STAT 581)

---

minimax rule $\Longleftrightarrow$ "least favorable" Bayes rule

Def. [LF prior]

A prior π^* is said to be a least favorable prior (LFP) if

$$r(D_{\pi^*}, \pi^*)$$

$$= \sup_{\pi} r(D_\pi, \pi)$$

Thm. If π satisfies

$$r(D_\pi, \pi) = \sup_{\theta \in \Theta} \mathcal{R}(D_\pi, \theta)$$

then

(i) D_π is a minimax rule ;

(ii) If D_π is the unique Bayes rule, then it is the unique minimax rule ;

(iii) π is a LFP.

Proof:

(i) Consider any $D \in \mathcal{D}$. Then

$$\sup_{\theta \in \Theta} \mathcal{R}(D, \theta)$$

$$\geq \int_{\Theta} \mathcal{R}(D, \theta) \underline{\pi(d\theta)}$$

as long as a prob. measure

$$\overset{(*)}{\geq} \int_{\Theta} \mathcal{R}(D_\pi, \theta) \pi(d\theta)$$

By def. of Bayes rule

$$= \sup_{\theta \in \Theta} \mathcal{R}(D_\pi, \theta)$$

Thm. condition

Because D is arbitrary, D_π is a minimax rule.

(ii) Uniqueness

If D_π is unique, for any other D, $(*)$ is a strict " $>$ ".

Then,

$$\sup_{\theta \in \Theta} R(D, \theta) > \sup_{\theta \in \Theta} R(D_\pi, \theta)$$

$\therefore$ minimax rule is also unique

(iii) To prove π is LFP, for $\forall \pi'$

$$\begin{aligned} & r(D_{\pi'}, \pi') \\ \le\ & r(D_\pi, \pi') \\ \le\ & \sup_{\theta \in \Theta} R(D_\pi, \theta) \\ =\ & r(D_\pi, \pi) \end{aligned}$$

Since π' is arbitrary, π is the LFP.

Message:

Minimax is really related to the LFP

Corollary. If Π is a prior s.t.

$\mathcal{R}(D_\pi, \cdot)$ is a constant valued function,

then D_π is minimax.

Remark: It is very rare that we can use such result; one example in HW2. However, we can generalise the above result a little bit.

<u>Def</u> [Generalized LFP ; LF sequence of priors]

Let $\{\Pi_K ; K = 1, 2, \dots\}$

be a sequence of prior dist. on Θ, and

define

$$r_0 := \liminf_{K \to \infty} r(D_{\pi_K}, \pi_K)$$

Then, this seq. is said to be least-favorable if $\forall \pi$,

$$r(D_\pi, \pi) \le r_0$$

Thm. [Generalization of previous Thm.]

If ① $\{\pi_K\}$ is a prior seq.

② $r_0 := \liminf_{K\to\infty} r(D_{\pi_K}, \pi_K)$

③ If some $D \in \mathcal{D}$ satisfies

$$\sup_{\theta\in\Theta} R(D, \theta) = r_0$$

then ① D is minimax

② π_K is LF seq.

Example [Normal mean problem]

$$X_1, \dots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$$

σ^2 → known

$\theta \in \Theta = \mathbb{R}$

$L(\cdot, \cdot)$ = squared error

$$(a, \theta) \longrightarrow (a-\theta)^2$$

Claim: $\overline{X}_n \left(:= \frac{1}{n}\sum x_i\right)$ is minimax

(Use def. can be very difficult ...)

Step 1. $T: \overset{\in \mathbb{R}^n}{x} \longmapsto \frac{1}{n}\sum x_i$

$\sup R(T, \theta)$

$\overset{\sup}{=} \mathbb{E}[(\bar{X}_n - \theta)^2]$

$= \frac{\theta^2}{n}$ (indep. of θ)

this should be r_0

It remains to find $\{\pi_k\}$ s.t.

$$\liminf_{n\to\infty} r(D_{\pi_k}, \pi_k) = \theta^2/n$$

Step 2. By Guessing,

$$\pi_k \sim N(0, k)$$

HW2:

$$r(D_{\pi_k}, \pi_k) \xrightarrow{k\to\infty} \theta^2/n$$

Step 3. Just use the above Thm. $\square$

Proof. Consider any decision $D' \in \mathcal{D}$

Then $\forall k = 1, 2, 3, \dots$

$$\sup_{\theta \in \Theta} R(D', \theta)$$

$$\geq \int_{\Theta} R(D', \theta)\, \pi_k(d\theta)$$

$$\geq r(D_{\pi_k}, \pi_k)$$

The above is true $\forall k$,

$\Downarrow$

$$\sup_{\theta \in \Theta} R(D', \theta) \geq \liminf_{k \to \infty} r(D_{\pi_k}, \pi_k)$$

$$= r_0$$

$$= \sup_{\theta \in \Theta} R(D, \theta)$$

$\therefore$ D is a minimax rule

(ii) Consider any π, then

$$r(D_\pi, \pi)$$

$$\leq r(D, \pi)$$

$$\leq \sup_{\theta \in \Theta} R(D, \pi) = r_0$$

$\therefore$ $\{\pi_k\}$ is LF seq.

We have shown

$\overline{X}_n$ is minimax optimal

(i.e. a minimax rule)

w.r.t.

known

$$\left\{ X_1, \dots, X_n \overset{iid}{\sim} N(\theta, \boxed{\sigma^2}) \right\}$$

Q: Is the sample mean still minimax optimal

when ① Normal

② $Var(X_1)$ known

is violated ?

A: Yes !!!

Previously,

$$\inf_{D \in \mathcal{D}} \boxed{\sup_{\underbrace{\theta \in \Theta}_{P \in \mathcal{P}}}} \mathcal{R}(D, \theta)$$

DGP   the family/model of DGPs

To show $\overline{X}_n$ is minimax w.r.t. a sufficiently large $\mathcal{P}$, we need the theorem.

Thm. Consider $\mathcal{P}_1$, $\mathcal{P}_2$ are two families of DGPs $\mathcal{P}_1 \subseteq \mathcal{P}_2$. If

(i) D_1 is minimax w.r.t. $\mathcal{P}_1$

i.e. $\sup\limits_{P \in \mathcal{P}_1} \mathcal{R}(D_1, P) = \inf\limits_{D} \sup\limits_{P \in \mathcal{P}_1} \mathcal{R}(D, P)$

(ii) $\sup\limits_{P \in \mathcal{P}_1} \mathcal{R}(D_1, P)$

$= \sup\limits_{P \in \mathcal{P}_2} \mathcal{R}(D_1, P)$

Then, D_1 is minimax w.r.t. $\mathcal{P}_2$.

Corollary 1. Consider

$$\mathcal{P}_2 = \{ P = Q^{\otimes n}, \quad \mathrm{supp}(Q) \subset \mathbb{R}, \quad \mathrm{Var}_Q(X) = \sigma^2 \}$$

Then: $\bar{X}_n$ is minimax w.r.t. $\mathcal{P}_2$.

Proof. (i) We know

$$T: X \mapsto \bar{X}_n$$

has the risk

$$\sup_{P \in \mathcal{P}_2} \mathcal{R}(T, P) = \sup \mathbb{E}_{X_1, \dots, X_n \overset{iid}{\sim} P} (\bar{X}_n - \mathbb{E}_P X_1)^2$$

$$= \sigma^2/n = \sup_{P \in \mathcal{P}_1} \mathcal{R}(T, P)$$

(ii) $\mathcal{P}_1 \subseteq \mathcal{P}_2$

$\Longrightarrow$ $\bar{X}_n$ is minimax in $\mathcal{P}_2$

$\square$

Corollary 2

$$\mathcal{P}_3 = \{ P \in Q^{\otimes n},\ supp(Q) \subset \mathbb{R},\ Var_Q(X) \leq \sigma^2 < \infty \}$$

Then, $\bar{X}_n$ is minimax in $\mathcal{P}_3$.

Proof. ① $\mathcal{P}_1 \subseteq \mathcal{P}_3$

② $\sup_{P \in \mathcal{P}_3} \mathcal{R}(T, P)$

$$= \sup_{P \in \mathcal{P}_3} \frac{Var_Q(X)}{n}$$

$$= \sigma^2/n$$

$$= \sup_{P \in \mathcal{P}_1} \mathcal{R}(T, P)$$

<u>Proof.</u> $\mathcal{P}_1 \subseteq \mathcal{P}_2$

$$\sup_{P\in\mathcal{P}_1} \mathcal{R}(T, P) \underset{=}{\uparrow}\!\!\le \sup_{P\in\mathcal{P}_2} \mathcal{R}(T, P)$$

To formalize it, proof by contradiction.

Suppose

D_1 is NOT minimax over $\mathcal{P}_2$.

Then $\exists D_2 \in \mathcal{D}$ s.t. it has smaller worst risk in $\mathcal{P}_2$.

Then
$$\sup_{P\in\mathcal{P}_1} \mathcal{R}(D_2, P)$$
$$\le \sup_{P\in\mathcal{P}_2} \mathcal{R}(D_2, P)$$
$$< \sup_{P\in\mathcal{P}_2} \mathcal{R}(D_1, P)$$
$$= \sup_{P\in\mathcal{P}_1} \mathcal{R}(D_1, P)$$

$\mathcal{P}_1$ is also not minimax in $\mathcal{P}_1$.

Contradiction!

Chapter 1.5. Admissiblity

Q: Is the $\bar{X}_n$ admissible??

HW1: $X \sim Bin(\theta, n)$

Prove: $T: x \mapsto 0.5$ is admissible

Q: $T: x \mapsto 1$ admissible?

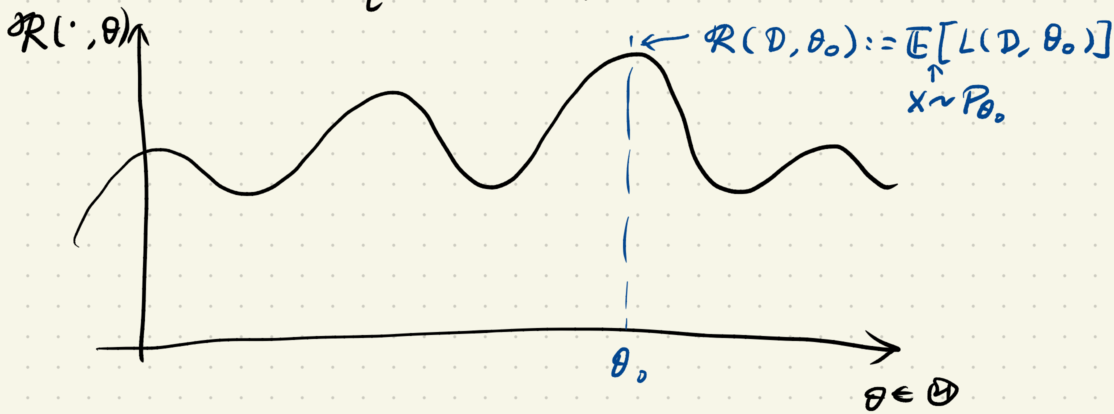
A: $\frac{X}{n}$ is dominating this T.

Def. [Uniquessness of a Bayes rule]

For a prior π, a rule D_π is said to be unique

if $\forall \theta \in \Theta$, a rule is Bayes iff

it is equal to D_π a.e. w.r.t. P_θ



In other words, to above def. of uniqueness is the best possible.

Def. [Uniqueness of minimax rule]

A rule D^* is said to be unique minimax if $\forall \theta \in \Theta$, a rule is minimax iff it is identical to D^* [a.e. w.r.t. P_θ]

P_θ- a.e.

Thm. [Admissiblity of Bayes / Minimax rule]

As long as the Bayes ~~minimax~~ rule is unique, it is admissible.

Proof HW2

Thm. [Uniqueness of the Bayes rule]

π : prior

D_π : Bayes rule

Q : the marginal dist. of x

$$\left[\begin{array}{l} \Theta \sim \pi \ (P(\theta)) \\ X \mid \Theta = \theta \sim P_\theta \ (P(X \mid \theta)) \\ X \sim Q \ (P(x)) \end{array}\right]$$

defined as, $\forall$ meas-set A,

$$Q(A) = \int P_\theta (x \in A) \, \pi(d\theta)$$

Then, as long as the following are true.

(i) the loss func.

$$L : (a, \theta) \longmapsto L(a, \theta)$$

is either squared error or strictly convex w.r.t. a, $\forall \theta \in \Theta$

(ii) $r(D_\pi, \pi) < \infty$

absolutely continuous

* (iii) For any $\theta \in \Theta$, $P_\theta \boxed{\ll} Q$,

i.e. $\forall$ meas. set A,

$$Q(A) = 0 \implies P_\theta(A) = 0$$

Or

$$P_\theta(A) > 0 \implies Q(A) > 0.$$

then, D_π is unique Bayes.

Lemma (Sufficient cond. of Thm (iii))

Cond (iii) is true as long as

$\exists$ measure ν on $(\mathcal{X}, \mathcal{A})$ s.t.

$\forall \theta \in \Theta$,

$$P_\theta \ll \nu \quad \text{and} \quad \nu \ll P_\theta$$

Proof: Fix $\theta_0 \in \Theta$. The goal is to show

as long as $P_{\theta_0}(A) > 0 \Rightarrow Q(A) > 0$

Suppose $P_{\theta_0}(A) > 0$. Because $P_{\theta_0} \ll \nu$,

$\Rightarrow \nu(A) > 0$

$\Longrightarrow P_\theta(A) > 0, \ \forall \theta \in \Theta$

because $\forall \theta \in \Theta$,
$\nu \ll P_\theta$

$$\Longrightarrow Q(A) = \int_\Theta \underbrace{P_\theta(A)}_{>0} \pi(d\theta)$$

> 0

$\square$

Example [Bayesian Normal Mean is admissible]

$$X_1, \dots, X_n \mid \Theta = \theta \overset{iid}{\sim} N(\theta, \sigma^2)$$

$$\Theta \sim N(\mu, \tau^2)$$

$$\Rightarrow T_\pi : x \longmapsto (1 - P_n)\bar{x}_n + P_n \mu$$

with $P_n := \dfrac{1/\tau^2}{1/\tau^2 + n/\sigma^2} \in (0, 1)$

is admissible

Because

Approach 1 : Using the Lemma

$P_\theta \ll \lambda$ and $\lambda \ll P_\theta$

$\mathcal{N}(\theta, \sigma^2)$ Lebesgue measure

Approach 2 : Q is normal

$\Rightarrow P_\theta \ll Q$

Q: $(1-P_n)\bar{X}_n + P_n \mu$ is admissible over $P_n \in (0,1)$. What will happen if

$P_n \notin (0,1)$???

A: ① If $P_n = 1$, then

$T: x \mapsto \mu$

is admissible.

② If $P_n = 0$, then

$T: x \mapsto \bar{X}_n$

is minimax and admissible

very difficult

Friday :)

③ If $P_n > 1$ or $P_n < 0$, then

it is inadmissible

Proof of ③: By construction,

If $P_n < 0$, will be dominated by $T_{0,\mu}$

if $P_n > 1$, will be dominated by $T_{1,\mu}$

How to show normal mean is admissible?

$$X_1, X_2, \dots, X_n \mid \Theta = \theta \overset{iid}{\sim} N(\theta, \sigma^2)$$

$$\Theta \sim N(\mu, \tau^2)$$

$\Longrightarrow$ Bayes rule/estimator

Case 1:

$$(1-P_n)\bar{X}_n + P_n \cdot \mu$$

with $P_n = \dfrac{1/\tau^2}{1/\tau^2 + n/\sigma^2} \in (0, 1)$

is provably admissible.

Case 2 : $P_n \notin [0,1]$

$T_{P_n,\mu}$

Claim : $(1-P_n)\bar{X}_n + P_n\mu$ is inadmissible.

Proof.

If $P_n < 0$, then

$$\begin{aligned}
&\mathcal{R}(T_{P_n,\mu}, \theta) \\
&= \text{bias}^2(T_{P_n,\mu}, \theta) + \text{Var}(T_{P_n,\mu}) \\
&\geq \text{Var}(T_{P_n,\mu}) \\
&= \frac{(1-P_n)^2\sigma^2}{n} > \theta^2/n \\
&= \mathcal{R}(\bar{X}_n, \theta)
\end{aligned}$$

algebra

$T_{0,\mu}$

If $P_n > 1$,

$$\begin{aligned}
&\mathcal{R}(T_{P_n,\mu}, \theta) \\
&= \text{bias}^2(T_{P_n,\mu}) + \text{Var}(T_{P_n,\mu}) \\
&\geq \text{bias}^2(T_{P_n,\mu})
\end{aligned}$$

$$= \{\underbrace{(1-P_n)\theta + P_n\mu}_{\mathbb{E}_\theta T_{P_n,\mu}} - \theta\}^2$$

$$= P_n^2(\mu-\theta)^2 > (\mu-\theta)^2$$

$$= R(T_{1,\mu}, \theta)$$

$T_{1,\mu} : x \longmapsto \mu.$

Case 3: $P_n = 0$ or 1

If $P_n = 1$, $T_{1,\mu}$ is admissible.

Claim: If $P_n = 0$, $T : x \longmapsto \bar{X}_n$.

is admissible

in the model $X_1, \dots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$, $\theta \in \Theta \subset \mathbb{R}$, σ^2 known

under MSE

Proof. WLOG, assume $\sigma^2 = 1$.

Step 1. Let's construct a Bayes estimator

$$\theta \sim \mathcal{N}(0, \tau^2) \leftarrow \Pi_\tau$$

$$\Downarrow$$

$$T_\tau \; (= T_{\pi_\tau})$$

$$\Rightarrow r(T_\tau, \Pi_\tau) = \frac{\tau^2}{1+n\tau^2}$$

In addition, using $T (= T_{0,\mu})$ to represent $\bar{X}_n$, $\forall \theta$, $R(T, \theta) = \frac{1}{n}$.

Step 2. The diff between the risks is super small when τ is large

$$r(T_\tau, \Pi_\tau) - R(T, \theta)$$

$$= \frac{\tau^2}{1+n\tau^2} - \frac{1}{n} = -\frac{1}{n(1+n\tau^2)}$$

Step 3. Let's consider an arbitrary estimator T_1.

$$-\frac{1}{n(1+n\tau^2)} = r(T_\tau, \Pi_\tau) - R(T, \theta)$$

$$\leq r(T_1, \Pi_\tau) - \frac{1}{n}$$

$x^+ = x\mathbb{1}\{x \geq 0\}$

$x^- = -x\mathbb{1}\{x \leq 0\}$

$\Downarrow$

$x = x^+ + x^-$

$$= \int \left[R(T_1,\theta) - \frac{1}{n} \right] \pi_\tau(d\theta)$$

$$= \int \left[R(T_1,\theta) - \frac{1}{n} \right]^+ \pi_\tau(d\theta) - \int \left[R(T_1,\theta) - \frac{1}{n} \right]^- \pi_\tau(d\theta)$$

Step 4. We separately consider two cases for π

(i) $\forall \theta \in \Theta$, $R(T_1, \theta) \geq R(T, \theta)$ ✓

(ii) $\exists\ \theta \in \Theta$ s.t.

$$R(T_1, \theta) < R(T, \theta)$$

Goal: In the (ii) case, it is always true that

$$\int \left[R(T_1, \theta) - \frac{1}{n} \right]^+ \pi_\tau(d\theta) > 0$$

for some τ.

Step 5. Step 3 gives

$$\int \left[R(T_1,\theta) - \frac{1}{n} \right]^+ \pi_\tau(d\theta) \geq -\frac{1}{n(1+n\tau^2)} + \int \left[R(T_1,\theta) - \frac{1}{n} \right]^- \pi_\tau(d\theta)$$

$-\frac{1}{n(1+n\tau^2)}$

Claim: as long as π_τ $\exists(\theta_1-\delta, \theta_1+\delta)$ s.t. $\mathcal{R}(T_1,\theta) < \frac{1}{n}$

$$\Downarrow$$

$\exists(\theta_1-\delta, \theta_1+\delta)$, s.t. $\forall \theta \in (\theta_1-\delta, \theta_1+\delta)$

$$\mathcal{R}(T_1,\theta) \le \frac{1}{n} - \varepsilon$$

Proof. It is true because

$\mathcal{R}(T_1, \cdot)$ is cont. w.r.t. $\theta \in \mathbb{R}$

$$\mathcal{R}(T_1,\cdot) = \mathbb{E}[T_1-\theta]^2 \to \text{smooth}.$$

<u>Step 6.</u>

$$\int [\mathcal{R}(\cdot,\cdot) - \tfrac{1}{n}]^- \pi_\tau(d\theta)$$

$$\ge \int_{\theta_1-\delta}^{\theta_1+\delta} [\ \cdot\ ]^- \pi_\tau(d\theta)$$

$$\ge \int_{\theta_1-\delta}^{\theta_1+\delta} \varepsilon\, \pi_\tau(d\theta) \qquad \mathbb{P}(\mathcal{N}(0,\tau^2) \in (\theta_1-\delta, \theta_1+\delta))$$

$$= \varepsilon \cdot \pi_\tau((\theta_1-\delta, \theta_1+\delta))$$

$$\Downarrow$$

$$\int [\ ]^+ \pi_\tau(d\theta) \ge -\frac{1}{n(1+n\tau^2)} + \varepsilon\, \pi_\tau((\theta_1-\delta, \theta_1+\delta))$$

$$\int [\ ]^{+} \pi_\tau(d\theta) \geq \underbrace{-\frac{1}{n(1+n\tau^2)}}_{①} + \underbrace{\Sigma\, \pi_\tau((\theta_1-\delta, \theta_1+\delta))}_{②}$$

Notice :

First, ① $\to 0$ at rate $\frac{1}{\tau^2}$ as $\tau \to \infty$.

Second, ② $\to 0$ at rate $\frac{1}{\tau}$ as $\tau \to \infty$

$\Downarrow$

$\exists\ \tau_0$ large enough s.t.

① + ② > 0

$\square$