Module 4

Bayes and Minimax Estimation

In most practical situations, it is impossible to find a decision rule δ_* that has smallest risk $R_\delta(F)$ for every $F \in \mathcal{P}$ among all decision rules. Finding optimal rule may be possible in some situations if we consider some characteristics of risk function $R_\delta(F)$ that does not depend on $F \in \mathcal{P}$ and then try to optimize that characteristics among all decision rules.

Suppose that some prior information is available about $F \in \mathcal{P}$ in the form of a probability distribution on $\mathcal{P}$ (or Θ where $\mathcal{P} = \{F_{\underline{\theta}} : \underline{\theta} \in \Theta\}$ and functional form of F_θ is known). Consider an estimation problem with

$\mathcal{D}$: class of all (randomized) d.r.s / estimators for an estimation problem

Π: known probability distribution on $\mathcal{P}$. (Called prior distribution)

Define

$$r_\delta(\Pi) = \int_{\mathcal{P}} R_\delta(F)\, d\Pi(F)$$

= Bayes risk of d.r. $\delta \in \mathcal{D}$ with respect to Π (does not depend on F)

Definition: (a) A decision rule / estimator δ_Π is said to be a Bayes rule w.r.t. Π if

$$r_{\delta_\Pi}(\Pi) = \inf_{\delta \in \mathcal{D}} R_\delta(\Pi)$$

(b) An estimator/ d.r. δ^* is said to be a minimax estimator if

$$\sup_{F \in \mathcal{P}} R_{\delta^*}(F) = \inf_{\delta \in \mathcal{D}} \sup_{F \in \mathcal{P}} R_\delta(F).$$

Note: (a) Bayes d.r. minimizes average risk $R_\delta(F)$ when average is taken with respect to prior π w.r.t. π.

(b) Minimax d.r. minimizes the maximum risk.

Consider

- $\underline{X} = (X_1, \ldots, X_n) \overset{d.f.}{\sim} F \in \mathcal{P} = \{F_{\underline{\theta}} : \underline{\theta} \in \Theta\}$, where functional form of $F_{\underline{\theta}}$ is known but $\underline{\theta} \in \Theta$ is unknown.
- $\underline{\theta} \in \Theta$ is seen as realization of a r.v. S whose (prior) distribution is π
- $\pi(\cdot)$ is based on past experience, past data or Statistician's belief.
- $\underline{X} = (X_1 \ldots X_n)$: sample from $F_{\underline{\theta}} \equiv F_{\underline{X}|S=\underline{\theta}}$ (Conditional distribution of $\underline{X}$ given $S = \underline{\theta}$)
- $f_{\underline{\theta}}(\cdot)$, $\underline{\theta} \in \Theta$: Conditional p.d.f. of $\underline{X}$ given $S = \underline{\theta}$
- The observed sample $\underline{X} = \underline{x}$ is used to obtain an updated prior distribution, which is called the posterior distribution.

Posterior distribution of $S \equiv$ Distribution of S given $\underline{X} = \underline{x}$

$$\equiv \frac{\text{Jt. dist of } (\underline{X}\ S)}{\text{marginal dist. of } \underline{X}}$$

- $f_{S|\underline{X}}(\theta|\underline{x})$ = posterior p.d.f. of S given $\underline{X}=\underline{x}$

$$= \frac{f_{\underline{X},S}(\underline{x},\theta)}{f_{\underline{X}}(\underline{x})}$$

$$= \frac{f_{\underline{X}|S}(\underline{x}|\theta)\, f_S(\theta)}{\int_{\Theta} f_{\underline{X},S}(\underline{x},\theta)\, d\theta} = \frac{f_\theta(\underline{x})\, f_S(\theta)}{m(\underline{x})}, \text{ say,}$$

where

$$m(\underline{x}) = \int_{\Theta} f_{\underline{X},S}(\underline{x},\theta)\, d\theta, \quad \underline{x} \in \mathbb{R}^n$$

$\equiv$ marginal dist. of $\underline{X}$.

- Posterior distribution (distribution of S given $\underline{X}=\underline{x}$) contains all the available information about unknown $\underline{\theta}$.

- Throughout we assume that $L(\underline{\theta}, a) \geq 0$, $\forall\, \underline{\theta} \in \Theta$, $a \in \mathcal{A}$.

<u>Theorem</u> Let $\delta_0(\cdot)$ be a non-randomized d.r. such that for any $\underline{x} \in \mathcal{X}$

$$E_{S|\underline{X}=\underline{x}}\left(L(S, \delta_0(\underline{x}))\right) = \min_{a \in \mathcal{A}} E_{S|\underline{X}=\underline{x}}\left(L(S, a)\right)$$

where expectation is w.r.t. the posterior distribution of S given $\underline{X}=\underline{x}$. Then δ_0 is Bayes w.r.t. π.

Proof. Let $\delta(\cdot|\cdot)$ be any d.r.. For given $\underline{x}$, let $Y_{\underline{x}}$ be the r.v. corresponding to $Y_{\underline{x}}$. Then

$$L^*(\theta, \delta, \underline{x}) = E_{Y_{\underline{x}}}\left(L(\theta, Y_{\underline{x}})\right)$$

$$R_\delta(\theta) = E_{\underline{X}|S=\theta}\left[L^*(\theta, \delta, \underline{X})\right] = E_{\underline{X}|S=\theta}\left[E_{Y_{\underline{X}}} L(\theta, Y_{\underline{X}})\right]$$

$$r_\pi(\delta) = E_S\left(R_\delta(S)\right) = E_S\, E_{\underline{X}|S}\, E_{Y_{\underline{X}}}\left(L(S, Y_{\underline{X}})\right)$$

$\boxed{3/4}$

$$= E_{\underline{X}} E_{S|\underline{X}} E_{\gamma_{\underline{x}}} (L(S, \gamma_{\underline{x}}))$$

$$= E_{\underline{X}} E_{\gamma_{\underline{x}}} E_{S|\underline{X}} (L(S, \gamma_{\underline{x}}))$$

$$\geq E_{\underline{X}} E_{\gamma_{\underline{x}}} E_{S|\underline{X}} (L(S, \delta_0(\underline{X})))$$

$$= E_{\underline{X}} E_{S|\underline{X}} (L(S, \delta_0(\underline{X})))$$

$$= r_\pi(\delta_0).$$

Hence the result follows.

Remark: (i) For finding a Bayes rule with respect to a given prior distribution it is enough to consider only non-randomized decision rules.

(ii) The non-randomized d.r. δ_0 s.t. $E_{S|\underline{X}=\underline{x}}(L(S, \delta_0(\underline{x}))) = \min_{a \in \mathcal{A}} E_{S|\underline{X}=\underline{x}}(L(S, a))$ is called the Bayes action.

Example: $g(\cdot)$: a real valued function

$\underline{X} \sim F \in \mathcal{P} = \{F_{\underline{\theta}} : \underline{\theta} \in \Theta\}$

$\pi(\cdot)$: a probability distribution on Θ with

$$E_S(g^2(S)) < \infty$$

$\mathcal{A}$ = range of $g(\underline{\theta})$, $\underline{\theta} \in \Theta$

Goal: To estimate $g(\underline{\theta})$ under the loss functions

$$L_1(\underline{\theta}, a) = (a - g(\underline{\theta}))^2, \quad L_2(\underline{\theta}, a) = \left(\frac{a}{g(\underline{\theta})} - 1\right)^2,$$

$$\underline{\theta} \in \Theta, \ a \in \mathcal{A}.$$

Let

$$\psi_1(a) = E_{S|\underline{X}=\underline{x}}((a - g(S))^2) = a^2 - 2a E_{S|\underline{X}=\underline{x}}(g(S)) + E_{S|\underline{X}=\underline{x}}(g^2(S))$$

$$\psi_2(a) = a^2 E_{S|\underline{X}=\underline{x}}\left(\frac{1}{g^2(S)}\right) - 2a E_{S|\underline{X}=\underline{x}}\left(\frac{1}{g(S)}\right) + 1$$

Clearly, for given $\underset{\sim}{x}$,

ψ_1 is minimized at $a \equiv a_1^*(\underset{\sim}{x}) = E_{S|\underset{\sim}{X}=\underset{\sim}{x}}(g(S))$

$\rightarrow$ posterior mean of $g(S)$

and ψ_2 is minimized $a = a_2^*(\underset{\sim}{x}) = \dfrac{E_{S|\underset{\sim}{X}=\underset{\sim}{x}}\left(\frac{1}{g(S)}\right)}{E_{S|\underset{\sim}{X}=\underset{\sim}{x}}\left(\frac{1}{g^2(S)}\right)}$.

<u>Example</u> Let $X_1, \ldots, X_n$ be i.i.d. $P(\theta)$ r.v.s, where $\theta \in \Theta = (0, \infty)$ is unknown. Consider estimation of $g(\theta) = \theta^m$, for a given positive integer m.

Then

$$f_{\underset{\sim}{\theta}}(\underset{\sim}{x}) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{\lfloor x_i} I_{\{0, 1, 2, \ldots\}}(x_i)$$

$$= \frac{e^{-n\theta}\,\theta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} \lfloor x_i}, \quad \underset{\sim}{x} = (x_1, \ldots, x_n) \in \{0, 1, \ldots\}^n.$$

Consider the prior distribution $\Pi(\cdot)$ as

$$f_S(\theta) = \frac{\theta^{\alpha-1} e^{-\frac{\theta}{\nu}}}{\Gamma\alpha\, \nu^{\alpha}}, \quad \theta > 0, \rightarrow \text{Gamma}(\alpha, \nu)$$

where $\alpha > 0$ and $\nu > 0$ are known constants.

Then

$$f_{S|\underset{\sim}{X}}(\theta|\underset{\sim}{x}) \propto \text{Jt p.d.f. of } (S, \underset{\sim}{X}) \text{ at point } (\underset{\sim}{\theta}, \underset{\sim}{x})$$

$$= f_{S, \underset{\sim}{X}}(\underset{\sim}{\theta}, \underset{\sim}{x})$$

$$= f_{\theta}(\underset{\sim}{x})\, f_S(\theta)$$

$$\propto e^{-(n+\frac{1}{\nu})\theta}\, \theta^{\sum_{i=1}^{n} x_i + \alpha - 1} \rightarrow \text{Gamma}\left(\sum_{i=1}^{n} x_i + \alpha, \frac{\nu}{n\nu+1}\right)$$

If $L(\theta, a) = (g(\theta) - a)^2$, $\theta \in \Theta$, $a \in \mathcal{A} = (0, \infty)$. Then the Bayes estimator is

$$\delta_\pi(\underline{x}) = E_{S|\underline{X}=\underline{x}}(g(S))$$
$$= E_{S|\underline{X}=\underline{x}}(S^m)$$
$$= \frac{\Gamma\left(\sum_{i=1}^n x_i+\alpha+m\right)\left(\frac{\vartheta}{n\vartheta+1}\right)^{\sum_{i=1}^n x_i+\alpha+m}}{\Gamma\left(\sum_{i=1}^n x_i+\alpha\right)\left(\frac{\vartheta}{n\vartheta+1}\right)^{\sum_{i=1}^n x_i+\alpha}}$$

$$= (T+\alpha+m-1)\cdots(T+\alpha)\left(\frac{\vartheta}{n\vartheta+1}\right)^m$$

Note: In the above example, the prior distribution and the posterior distribution are in the same parametric family of distributions

Definition: If the prior distribution is such that the prior and the posterior distribution are in the same parametric family, then such a prior is called a Conjugate prior.

Definition: (i) A measure $\pi(\cdot)$ is called an improper prior if $\pi(\Theta)=\infty$

(ii) A measure π with $\pi(\Theta)=1$ is called a proper prior;

(iii) An estimator $\delta(\underline{x})$ that minimizes

$$\psi(a) = \int_\Theta L(\theta,a)\, f_\theta(\underline{x})\, f_S(\theta)\, d\theta$$

with an improper prior $\pi(\cdot)$ is called a generalized Bayes action

Note: Let π be a prior such that $\pi(\Theta)<\infty$. One can make π to be proper prior by considering

$$\pi^*(A) = \frac{\pi(A)}{\pi(\Theta)}, \quad A \subseteq \mathbb{R}, \quad (\because \text{through multiplication of a constant } \pi(\Theta).)$$

Clearly $\pi^*(\Theta)=1$.

Need for Generalized Bayesian Approach

- Suppose that one has no prior information and it has to be chosen subjectively.
- In such cases one would like to select a non-informative prior that treats all values in Θ equitably, in some sense.
- A non-informative prior is often improper (see Berger (1985) for details)

Example $X_1, \ldots, X_n$ are i.i.d. $N(\theta, \sigma^2)$, where $\theta \in \mathbb{R} = \Theta$ is unknown and $\sigma > 0$ is known.

Goal: To estimate θ under the squared error loss (SEL) function :

$$L(\theta, a) = (a-\theta)^2, \quad a \in \mathcal{A} = \mathbb{R}, \ \theta \in \Theta.$$

If $\Theta = [a, b]$, then a non-informative prior is $U(a, b)$.
If $\Theta = \mathbb{R}$, the corresponding uniform distribution is the Lebesgue measure on $\mathbb{R}$ (weight to an interval just depends on its length and not on location) which is an improper prior on $\mathbb{R}$. The generalized Bayes estimator $\hat{\delta}_\pi$ corresponding to the Lebesgue measure $\hat{\pi}$ on $\mathbb{R}$ minimizes,

$$\psi(a) = (2\pi\sigma^2)^{-\frac{n}{2}} \int_{-\infty}^{\infty} (a-\theta)^2 e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2} d\theta$$

$$\approx k(\underline{x}) \int_{-\infty}^{\infty} (a-\theta)^2 e^{-\frac{1}{2\sigma^2}(\theta-\bar{x})^2} d\theta$$

The minimizing a for given $\underline{x}$, is

$$a \equiv a_*(\underline{x}) = \bar{x}$$

$$\Rightarrow \quad \delta_\pi(\underline{X}) = \bar{X} \text{ is the generalized Bayes estimator}$$

under SEL function.

Definition A decision rule/estimator δ is said to be inadmissible if there exists a decision rule δ_0 such that

$$R_{\delta_0}(\theta) \le R_{\delta}(\theta), \quad \forall \theta \in \Theta,$$

with strict inequality for some $\theta \in \Omega$.

Definition A decision rule δ is said to be admissible if it is not inadmissible, i.e., if for any d.r. δ_0 with

$$R_{\delta_0}(\theta) \le R_{\delta}(\theta), \quad \forall \theta \in \Omega$$

we have $R_{\delta_0}(\theta) = R_{\delta}(\theta), \quad \forall \theta \in \Omega.$

Example: Let $X \sim N(\theta, 1)$, where $\theta \in \mathbb{R} = \Theta$ is unknown.
Consider estimation of θ under the SEL

$$L(\theta, a) = (a-\theta)^2, \quad a, \theta \in \mathbb{R}.$$

Let $c \in \mathbb{R}$ be fixed
Consider estimator

$$\delta_c(\underline{x}) = c \qquad \text{(no-data rule).}$$

Then

$$R_{\delta_c}(\theta) = (\theta - c)^2$$

Claim: δ_c is admissible

On contrary suppose δ_c is inadmissible. Then there exists a d.r. δ^* s.t

$$R_{\delta^*}(\theta) \le R_{\delta_c}(\theta) \quad \forall \theta \in \mathbb{R} \quad \dots \quad (*)$$

with strict inequality for some θ

$$\Rightarrow \quad E_c\left((\delta^*(x) - c)^2 \right) \le 0 \qquad (\theta = c)$$

$$\Rightarrow \quad \delta^*(x) = c \quad \text{a.s. under } N(c, 1)$$

$$\Rightarrow \quad \delta_c(x) = \delta(x) \quad \text{a.s.} \quad \forall \theta \in \Theta$$

$$\Rightarrow \quad \longrightarrow \text{Contradiction } (*)$$

Thus an admissible estimator/rule may not be a good estimator/rule.
However any good rule $\boxed{8/4}$ should be admissible.

Definition A decision rule / estimator δ_π is called unique Bayes with respect to prior π if for any other decision rule / estimator δ with

$$r_\delta(\pi) \le r_{\delta_\pi}(\pi)$$

we have $\delta = \delta_\pi$ with probability one (under distribution of $\underline{X} \mid \Theta = \theta$), $\forall\, \theta \in \Theta$.

Theorem Let δ_π be a Bayes rule with respect to a prior π.

(a) If δ_π is unique Bayes rule, then δ_π is admissible;

(b) If Θ is countable and $\pi(\{\theta\}) > 0$ $\forall\, \theta \in \Theta$, then δ_π is admissible;

(c) If, for any d.r. δ with $R_\delta(\theta) < \infty$ $\forall\, \theta \in \Theta$ $R_\delta(\theta)$ is a continuous function of $\theta \in \Theta$ and $\pi(B) > 0$, $\forall\, B \subseteq \Theta$ open set s.t. $B \neq \phi$, then δ_π is admissible.

Proof. (a) On contrary suppose δ_π is inadmissible. Then there exists a d.r. δ^* s.t.

$$R_{\delta^*}(\theta) \le R_{\delta_\pi}(\theta), \quad \forall\, \theta \in \Theta$$

with strict inequality for some $\theta \in \Theta$ ---- $(*)$

$$\Rightarrow \quad E_S(R_{\delta^*}(S)) \le E_S(R_{\delta_\pi}(S))$$

$$\Rightarrow \quad r_{\delta^*}(\pi) \le r_\delta(\pi)$$

$$\Rightarrow \quad \delta^* = \delta_\pi \quad \text{a.s.} \quad \underline{X} \mid \Theta = \theta, \quad \forall\, \theta \in \Theta$$

$$\Rightarrow \quad R_{\delta^*}(\theta) = R_{\delta_\pi}(\theta), \quad \forall\, \theta \in \Theta,$$

contradicting $(*)$.

(b) Let $\Theta = \{\theta_1, \theta_2, \ldots\}$ or $\Theta = \{\theta_1, \ldots, \theta_k\}$ for some $k \in \mathbb{N}$. Let $\pi(\{\theta_i\}) = p_i$, $i = 1, 2, \ldots$, so that $p_i > 0$, $\forall\, i = 1, 2, \ldots$ and $\sum_{i:\, \theta_i \in \Theta} p_i = 1$.

On contrary suppose that δ_π is inadmissible. Then there exists a d.r. δ^* s.t.

$$R_{\delta^*}(\theta_i) \leq R_{\delta_\pi}(\theta_i), \quad \forall i$$

and $$R_{\delta^*}(\theta_j) < R_{\delta_\pi}(\theta_j), \text{ for some } j$$

$$\Rightarrow \sum_{i:\, \theta_i \in \Theta} R_{\delta^*}(\theta_i) p_i < \sum_{i:\, \theta_i \in \Theta} R_{\delta_\pi}(\theta_i) p_i$$

$$\Rightarrow r_{\delta^*}(\pi) < r_{\delta_\pi}(\pi),$$

contradicting the fact that δ_π is Bayes w.r.t. π.

(c) On contrary suppose that δ_π is inadmissible. Then there exists a d.r. δ^* s.t.

$$R_{\delta^*}(\theta) \leq R_{\delta_\pi}(\theta), \quad \forall \theta \in \Theta$$

with $$R_{\delta^*}(\theta_0) < R_{\delta_\pi}(\theta_0), \quad \text{for some } \theta_0 \in \Theta$$

Let $$R_{\delta_\pi}(\theta_0) - R_{\delta^*}(\theta_0) = \varepsilon,$$

so that $\varepsilon > 0$. Define

$$A = \{\theta \in \Theta : R_{\delta_\pi}(\theta) - R_{\delta^*}(\theta) > \frac{\varepsilon}{2}\}.$$

Then $A \neq \emptyset$ $(\theta_0 \in A)$. Also, since $R_{\delta_\pi}(\theta)$ and $R_{\delta^*}(\theta)$ are continuous functions of $\theta \in \Theta$, $R_{\delta_\pi}(\theta) - R_{\delta^*}(\theta)$ is a continuous function of $\theta \in \Theta$.

$\Rightarrow$ A is non-empty open set (for a continuous function inverse image of an open set is open)

Therefore

$$r_{\delta_\pi}(\pi) - r_{\delta^*}(\pi) = \int_\Theta [R_{\delta_\pi}(\theta) - R_{\delta^*}(\theta)]\, d\pi(\theta)$$

$$= \int_A [R_{\delta_\pi}(\theta) - R_{\delta^*}(\theta)]\, d\pi(\theta) + \int_{A^c} [R_{\delta_\pi}(\theta) - R_{\delta^*}(\theta)]\, d\pi(\theta)$$

$$\geq \int_A [R_{\delta_\pi}(\theta) - R_{\delta^*}(\theta)]\, d\pi(\theta)$$

$$\geq \frac{\varepsilon}{2} \int_A d\pi(\theta)$$

$$= \frac{\varepsilon}{2} \pi(A) > 0,$$

Contradicting the hypothesis that δ_π is Bayes w.r.t. π.

Remark: (i) Even if one does not support Bayesian approach, this approach can be used to generate admissible rules.

(ii) Generalized Bayes rules are not necessarily admissible. Many generalized Bayes rules are limits of Bayes rules and are often admissible.

Theorem: Suppose that $\Theta \subseteq \mathbb{R}^k$ is open and for any d.r./estimator with $R_\delta(\underline{\theta}) < \infty$, $\forall\, \underline{\theta} \in \Theta$, $R_\delta(\underline{\theta})$ is a continuous function of $\underline{\theta} \in \Theta$. Further suppose there exists a sequence $\{\pi_m\}_{m \geq 1}$ of priors (possibly improper) and a d.r./estimator δ_0 s.t.

(a) $\pi_m(B) > 0$, $\forall\, m \geq 1$ and for every non-empty open set $B \subseteq \cancel{\mathbb{R}^k}\ \Theta$;

(b) the Bayes risks $r_{\delta_0}(\pi_m)$, $m = 1, 2, \ldots$ are finite;

(c)
$$\liminf_{m \to \infty} \frac{[r_{\delta_0}(\pi_m) - r_{\delta_{\pi_m}}(\pi_m)]}{\pi_m(B)} = 0,$$

for every non-empty open set $B \subseteq \cancel{\mathbb{R}^k}\ \Theta$; here δ_{π_m} is a Bayes (or generalized Bayes) rule w.r.t. prior π_m.

Then δ_0 is admissible.

Proof. On contrary suppose δ_0 is inadmissible, i.e. there exists a d.r. δ^* s.t.

$$R_{\delta^*}(\underline{\theta}) \leq R_{\delta_0}(\underline{\theta}), \quad \forall\ \underline{\theta} \in \Theta$$

$$\text{with } R_{\delta^*}(\underline{\theta}_0) < R_{\delta_0}(\underline{\theta}_0), \text{ for some } \underline{\theta}_0 \in \Theta.$$

Let

$$R_{\delta_0}(\theta_0) - R_{\delta^*}(\theta_0) = \varepsilon,$$

so that $\varepsilon > 0$.

Define

$$B = \{\underline{\theta} \in \Theta : R_{\delta_0}(\underline{\theta}) - R_{\delta^*}(\underline{\theta}) > \tfrac{\varepsilon}{2}\}.$$

so that $B \neq \phi$ ($\theta_0 \in B$) and B is open (since $R_{\delta_0}(\underline{\theta}) - R_{\delta^*}(\underline{\theta})$ is continuous in $\underline{\theta} \in \Theta$). Thus

$$\begin{aligned}
r_{\delta_0}(\Pi_m) - r_{\delta_{\Pi_m}}(\Pi_m) &\geq r_{\delta_0}(\Pi_m) - r_{\delta^*}(\Pi_m) \\
&= \int_{\Theta} [R_{\delta_0}(\theta) - R_{\delta^*}(\theta)]\, d\Pi_m(\underline{\theta}) \\
&\geq \int_{B} [R_{\delta_0}(\underline{\theta}) - R_{\delta^*}(\underline{\theta})]\, d\Pi_m(\underline{\theta}) \\
&\geq \frac{\varepsilon}{2}\, \Pi_m(B)
\end{aligned}$$

$$\Rightarrow \quad \frac{r_{\delta_0}(\Pi_m) - r_{\delta_{\Pi_m}}(\Pi_m)}{\Pi_m(B)} \geq \frac{\varepsilon}{2}$$

$$\Rightarrow \quad \liminf_{m \to \infty} \frac{r_{\delta_0}(\Pi_m) - r_{\delta_{\Pi_m}}(\Pi_m)}{\Pi_m(B)} \geq \frac{\varepsilon}{2},$$

Contradicting hypothesis (c).

Hence the result follows.

Remark. Suppose that $\Theta \subseteq \mathbb{R}^k$ is open and for any d.r. δ with $R_\delta(\underline{\theta}) < \infty$, $\forall\ \underline{\theta} \in \Theta$, $R_\delta(\underline{\theta})$ is a continuous fn of $\underline{\theta} \in \Theta$. Let Π be a prior (possibly improper) s.t. $\Pi(B) > 0$, $\forall$ non-empty open set $B \subseteq \Theta$. If δ_0 is Bayes (generalised Bayes) with $r_{\delta_0}(\Pi) < \infty$, then δ_0 is admissible.

$\boxed{12/4}$

Lemma: Let $\{f_{\underline{\theta}} : \underline{\theta} \in \Theta\}$ be the natural exponential family of p.d.f.s dominated by some measure λ on $(\mathbb{R}^k, \mathcal{B}^k)$,

i.e.,

$$f_{\underline{\theta}}(\underline{x}) = e^{\underline{\theta}' T(\underline{x}) - \beta(\underline{\theta})} h(\underline{x}), \quad \underline{x} \in \mathcal{X} \subseteq \mathbb{R}^k,$$

where

$$\beta(\underline{\theta}) = \ln \left(\int_{\mathcal{X}} e^{\underline{\theta}' T(\underline{x})} h(\underline{x})\, d\lambda \right), \quad \underline{\theta} \in \Theta$$

If $\underline{\theta}_0$ is an interior point of natural parameter space and $\psi(\cdot)$ is a Borel function satisfying $E_{\underline{\theta}_0}(|\psi(\underline{X})|) < \infty$, then

$$E_{\underline{\theta}}(\psi(\underline{X})) = \int_{\mathcal{X}} \psi(\underline{x}) e^{\underline{\theta}' T(\underline{x}) - \beta(\underline{\theta})} h(\underline{x})\, d\lambda$$

is infinitely often differentiable in a neighborhood of $\underline{\theta}_0$ and the derivatives can be evaluated by differentiation under the integral sign.

Theorem: Let $\delta_0(\underline{X})$ be a Bayes estimator of $g(\underline{\theta})$ w.r.t. a prior Π, under the squared error loss function. If $\delta_0(\underline{X})$ is unbiased then $r_{\delta_0}(\Pi) = 0$ (i.e., if $r_{\delta_0}(\Pi) > 0$ and δ_0 is unbiased then δ_0 can not be Bayes)

Proof

Under SEL, the Bayes estimator is

$$\delta_0(\underline{X}) = E_{S|\underline{X}}(g(S)) \quad \cdots\cdots \quad (1)$$

Also

$$E_{\underline{X}|S}(\delta_0(\underline{X})) = g(S) \quad \cdots\cdots \quad (2)$$

(δ_0 is unbiased for $g(\Theta)$)

Thus

$$r_{\delta_0}(\Pi) = E_{S,\underline{X}}\left((\delta_0(\underline{X}) - g(S))^2 \right)$$

$$= E_{\underline{X}}(\delta_0^2(\underline{X})) - 2E_{S,\underline{X}}(\delta_0(\underline{X}) g(S)) + E_S(g^2(S))$$

But

$$E_{\underset{\sim}{S}\,\underset{\sim}{X}}(\delta_0(\underset{\sim}{X})\, g(S)) = E_{\underset{\sim}{X}}(E_{S|\underset{\sim}{X}}(\delta_0(\underset{\sim}{X})\, g(S)))$$

$$= E_{\underset{\sim}{X}}(\delta_0^2(\underset{\sim}{X})) \qquad \text{(using (i))}$$

and

$$E_{\underset{\sim}{S}\,\underset{\sim}{X}}(\delta_0(\underset{\sim}{X})\, g(S)) = E_S(E_{\underset{\sim}{X}|S}[\delta_0(\underset{\sim}{X})\, g(S)])$$

$$= E_S(g^2(S)) \qquad \text{(using (ii))}$$

Thus

$$r_{\delta_0}(\pi) = 0$$

Remark: Since $r_{\delta_0}(\pi) = 0$, under SEL, occurs usually in some trivial cases, a Bayes estimator is typically not unbiased. However a generalized Bayes estimator can be unbiased (see the example on page 7/4)

Example Let $X_1, \dots, X_n$ be i.i.d. $N(\mu, \sigma^2)$ r.v.s where $\theta = (\mu) \in \mathbb{R} = \Theta$ is unknown, and $\sigma > 0$ is known. Show that $\bar{X}$ can not be Bayes w.r.t. any prior π

Solution Since $\bar{X}$ is unbiased, for it to be Bayes w.r.t. π we must have

$$r_{\bar{X}}(\pi) = 0$$

$$\Rightarrow E^S(E^{X|S}((\bar{X}-\mu)^2)) = 0$$

$$\Rightarrow E^S\left(\frac{\sigma^2}{n}\right) = 0 \Rightarrow \sigma^2 = 0 \longrightarrow \text{Contradiction}$$

Hence $\bar{X}$ can not be Bayes w.r.t. proper prior.

Example Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ distribution, where $\mu \in \Theta = \mathbb{R}$ is unknown and $\sigma > 0$ is known. Under the SEL

$$L(\mu, a) = (a-\mu)^2, \quad a, \mu \in \mathbb{R},$$

show that $\bar{X}$ is admissible for estimating μ.

Solution Since the jt pdf of $(X_1, \ldots, X_n)$ belongs to exponential family, the risk functions of all d.r.s with finite risk are continuous. Let

$$\Pi_m \sim N(0, \sigma_m^2), \quad m = 1, 2, \ldots$$

Then

$$\Pi_m((a,b)) > 0, \quad \forall \ -\infty < a < b < \infty$$

$$\Rightarrow \quad \Pi_m(D) > 0, \quad \forall \text{ non-empty open set } D \subseteq \mathbb{R}^1$$

(open sets in $\mathbb{R}$ are countable union of open intervals)

$$f_{S|X}(\theta|x) \propto f_\theta(x)\, f_S(\theta)$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \times \frac{1}{\sigma_m\sqrt{2\pi}} e^{-\frac{\mu^2}{2\sigma_m^2}}$$

$$= c_1(x)\, e^{-\frac{1}{2}\left(\mu^2\left(\frac{1}{\sigma_m^2}+\frac{n}{\sigma^2}\right) - \frac{2\mu n\bar{x}}{\sigma^2}\right)}$$

$$\sim N\left(\frac{n\bar{x}/\sigma^2}{\frac{1}{\sigma_m^2}+\frac{n}{\sigma^2}}, \left(\frac{1}{\sigma_m^2}+\frac{n}{\sigma^2}\right)^{-1}\right)$$

$$\Rightarrow \quad \delta_{\Pi_m}(X) = \frac{\bar{X}}{\frac{\sigma^2}{n\sigma_m^2}+1}$$

$$r_{\pi_m}(\delta_{\pi_m}) = E_{\pi_m}\left[E_\mu \left(\frac{\bar{X}}{\frac{\sigma^2}{n\sigma_m^2}+1} - \mu \right)^2 \right]$$

$$= E_{\pi_m}\left[\frac{\sigma^2/n}{\left(\frac{\sigma^2}{n\sigma_m^2}+1\right)^2} + \left(\frac{\mu}{\frac{\sigma^2}{n\sigma_m^2}+1} - \mu \right)^2 \right]$$

$$= \frac{\sigma^2/n}{\left(\frac{\sigma^2}{n\sigma_m^2}+1\right)^2} + \left(\frac{\sigma^2/n\sigma_m^2}{\frac{\sigma^2}{n\sigma_m^2}+1} \right)^2 \sigma_m^2$$

$$= \frac{\sigma^2/n}{\left(\frac{\sigma^2}{n\sigma_m^2}+1\right)^2} \left(1 + \frac{\sigma^2}{n\sigma_m^2}\right) = \frac{\sigma^2/n}{\frac{\sigma^2}{n\sigma_m^2}+1}$$

Let $\delta_0(\underline{X}) = \bar{X}$. Then

$$\frac{r_{\pi_m}(\delta_0) - r_{\pi_m}(\delta_{\pi_m})}{\pi_m((a,b))} = \frac{\frac{\sigma^2}{n} - \frac{\sigma^2/n}{\frac{\sigma^2}{n\sigma_m^2}+1}}{\Phi\left(\frac{b}{\sigma_m}\right) - \Phi\left(\frac{a}{\sigma_m}\right)}$$

$$= \frac{1}{\left(\frac{\sigma^2}{n} + \sigma_m^2\right)\left[\Phi\left(\frac{b}{\sigma_m}\right) - \Phi\left(\frac{a}{\sigma_m}\right)\right]}$$

Let us choose σ_m such that $\sigma_m \to \infty$, as $m \to \infty$.

Then

$$\Phi\left(\frac{b}{\sigma_m}\right) - \Phi\left(\frac{a}{\sigma_m}\right) = \frac{b-a}{\sigma_m}\phi(c^*), \quad \frac{a}{\sigma_m} < c^* < \frac{b}{\sigma_m}$$

and

$$\frac{r_{\pi_m}(\delta_0) - r_{\pi_m}(\delta_{\pi_m})}{\pi_m((a,b))} \to 0, \quad \text{as } m \to \infty$$

$\Rightarrow \delta_0(\underline{X}) = \bar{X}$ is admissible for estimating μ.

Assignment Problems (1) Let $X_1, \dots, X_n$ be a random sample from $E(0, \frac{1}{\theta})$ distribution ($E(X_i) = \frac{1}{\theta}$), where $\mu \in \Theta = \mathbb{R}$ is unknown. Consider estimation of $\psi(\theta) = \frac{1}{\theta}$ under the SEL and the gamma prior pdf for $Y\theta$ as $\hat{h}(\mu) = \frac{\mu_0^{\alpha_0} e^{-\mu_0 \mu} \mu^{\alpha_0 - 1}}{\Gamma \alpha_0}, \mu > 0,$

where $\mu_0, \alpha_0 \in \mathbb{R}_{++}$ are fixed constants. Find the Bayes estimator, and show that it is not unbiased.

Show that $\delta_0(\underline{X}) = \bar{X}$ can not be Bayes wrt any proper prior.

(2) Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$ RVs, where $\underline{\theta} = (\mu, \sigma)$ $\in \Theta = \mathbb{R} \times \mathbb{R}_{++}$ is unknown. Consider the prior pdf for $(\mu, \tau) = (\mu, \frac{1}{2\sigma^2})$ as

$$h(\mu, \tau) = h_1(\mu|\tau)\, h_2(\tau), \quad \mu \in \mathbb{R}, \tau > 0,$$

where $h_1(\mu|\tau)$ is the pdf of $N(\theta_0, \frac{\sigma_0^2}{\tau})$ and $h_2(\tau)$ is the pdf of $G(\alpha_0, \mu_0)$, where $\theta_0 \in \mathbb{R}$ and $\sigma_0, \mu_0, \alpha_0 \in \mathbb{R}_{++}$ are fixed constants. Find the Bayes estimators of μ and σ^2 under the SEL.

Minimaxity and Admissibility

A minimax estimator is the one which minimizes $\sup_{\underline{\theta} \in \Theta} R_\delta(\underline{\theta})$, $\forall\, \delta \in \mathcal{D}$.

Definition An estimator δ is said to be unique minimax estimator if for any other estimator δ_0 with

$$\sup_{\theta \in \Theta} R_{\delta_0}(\theta) = \sup_{\theta \in \Theta} R_\delta(\theta),$$

we have $\delta = \delta_0$ a.s. $\underline{X} \,|\, \underline{\theta} = \underline{\theta}$, $\forall\, \underline{\theta}$.

Theorem Let δ_π be a Bayes estimator w.r.t. a prior Π. If δ_π has a constant risk $R_{\delta_\pi}(\theta)$ (i.e., $R_{\delta_\pi}(\theta) = c$, $\forall\, \underline{\theta} \in \Theta$ where c is a constant), then δ_π is minimax. Furthermore if δ_π is unique Bayes w.r.t. Π then it is the unique ~~Bayes~~ minimax estimator.

Proof. Let δ_0 be any other estimator of $g(\theta)$. Then

$$\sup_{\theta\in\Theta} R_{\delta_0}(\theta) \geq \int_{\Theta} R_{\delta_0}(\theta)\, d\pi(\theta) \quad (\sup \geq \text{average})$$

$$\geq \int_{\Theta} R_{\delta_\pi}(\theta)\, d\pi(\theta) \quad (\delta_\pi \text{ is Bayes w.r.t. } \pi)$$

$$= C = \sup_{\theta\in\Theta} R_{\delta_\pi}(\theta) \quad \ldots\ldots \quad (\mathrm{I})$$

Now suppose δ_π is unique Bayes but not unique minimax. Then $P_\theta(\delta_1 \neq \delta_\pi) > 0$ and there exists an estimator δ_1 such that

$$\sup_{\theta\in\Theta} R_{\delta_1}(\theta) = \sup_{\theta\in\Theta} R_{\delta_\pi}(\theta)$$

$$\Rightarrow \int_{\Theta} R_{\delta_1}(\theta)\, d\pi(\theta) = \int_{\Theta} R_{\delta_\pi}(\theta)\, d\pi(\theta) \quad (\text{in view of (I)})$$

$$\Rightarrow P_\theta(\delta_1 = \delta_\pi) = 1 \quad \forall\, \theta\in\Theta$$

$$\Rightarrow P_\theta(\delta_1 \neq \delta_\pi) = 0 \quad \forall\, \theta\in\Theta \longrightarrow \text{Contradiction.}$$

Hence δ_π is unique minimax if it is unique Bayes with constant risk

Example Let $X_1, \ldots, X_n$ be a random sample from $\mathrm{Bin}(m, \theta)$ distribution, where $\theta\in\Theta = (0,1)$ is unknown and m is a fixed positive integer. Consider estimation of θ under the SEL $L(\theta, a) = (a-\theta)^2$, $a, \theta\in\Theta$. Find the minimax estimator and show that $\delta_0(\underline{X}) = \overline{X}$ is not minimax.

Solution Consider $Be(\alpha, \beta)$ prior for θ, where $\alpha > 0$ and $\beta > 0$ are fixed positive constants. Then posterior distribution of θ given $\underline{X} = \underline{x}$ has p.d.f.

$$\pi(\theta|\underline{x}) \propto \left\{\prod_{i=1}^{n} \binom{m}{x_i} \theta^{x_i} (1-\theta)^{m-x_i}\right\} \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

$$\Rightarrow \pi(\theta|\underline{x}) = \frac{\theta^{t+\alpha-1}(1-\theta)^{mn-t+\beta-1}}{B(t+\alpha, mn-t+\beta)}, \quad 0<\theta<1, \quad \text{where } t = \sum_{i=1}^{n} x_i$$

The unique Bayes estimator of θ is

$$\delta_\pi(\underset{\sim}{X}) = E^{S|\underset{\sim}{X}}(S) = \frac{T+\alpha}{mn+\alpha+\beta},$$

where $T = \sum_{i=1}^{n} X_i \sim Bin(mn, \theta)$

If we can choose $\alpha>0$ and $\beta>0$ s.t. $\delta_\pi(\underset{\sim}{X})$ has the constant risk then the corresponding Bayes estimator will be the unique minimax estimator. We have, for $\theta \in \Theta$,

$$R_{\delta_\pi}(\theta) = E_\theta\left((\delta_\pi(\underset{\sim}{X}) - \theta)^2 \right)$$

$$= E_\theta\left(\left(\frac{T+\alpha}{mn+\alpha+\beta} - \theta\right)^2 \right)$$

$$= \frac{Var_\theta(T)}{(mn+\alpha+\beta)^2} + \left(\frac{mn\theta+\alpha}{mn+\alpha+\beta} - \theta\right)^2$$

$$= \frac{mn\,\theta(1-\theta)}{(mn+\alpha+\beta)^2} + \frac{(\alpha-(\alpha+\beta)\theta)^2}{(mn+\alpha+\beta)^2}$$

$$= \frac{((\alpha+\beta)^2 - mn)\theta^2 - \theta(2\alpha(\alpha+\beta) - mn) + \alpha^2}{(mn+\alpha+\beta)^2}$$

which does not depend on θ iff $(\alpha+\beta)^2 = mn$ and $2\alpha(\alpha+\beta) = mn$

$$\Leftrightarrow \alpha+\beta = \sqrt{mn} \qquad \alpha = \frac{\sqrt{mn}}{2}$$

$$\Leftrightarrow \alpha = \beta = \frac{\sqrt{mn}}{2}$$

Thus, for the choice $\alpha = \beta = \frac{\sqrt{mn}}{2}$, the Bayes estimator

$$\delta_\pi^*(\underset{\sim}{X}) = \frac{T + \frac{\sqrt{mn}}{2}}{mn + \sqrt{mn}}$$

is the unique minimax estimator.

$$R_{\delta_{\Pi}^*}(\theta) = \left[\frac{\alpha^2}{(mn+\alpha+\beta)^2}\right] \quad \alpha=\beta=\frac{\sqrt{mn}}{2}$$

$$= \frac{1}{4(\sqrt{mn}+1)^2}, \quad \forall \theta$$

$$R_{\bar{X}}(\theta) = \frac{mn\theta(1-\theta)}{m^2n^2} = \frac{\theta(1-\theta)}{mn}$$

$$\sup_{\theta\in\Theta} R_{\bar{X}}(\theta) = \frac{1}{4mn} > \frac{1}{4(\sqrt{mn}+1)^2} = \sup_{\theta\in\Theta} R_{\delta_{\Pi}^*}(\theta)$$

$\Rightarrow \bar{X}$ is not minimax.

Remark: (i) The minimax estimator heavily depends on the loss function. For example in the above case if we take

$$L(\theta, a) = \frac{(a-\theta)^2}{\theta(1-\theta)}, \quad 0<\theta<1, \quad 0\le a\le 1,$$

then $\bar{X}$ is the constant risk unique Bayes (and hence unique minimax) w.r.t $U(0,1)$ prior for θ. ↳ estimator

In this case the estimator

$$\delta_{\Pi}'(\underline{x}) = \frac{n\bar{x} + \frac{\sqrt{mn}}{2}}{mn+\sqrt{mn}} \qquad \text{(Exercise)}$$

has the unbounded risk.

(ii) In many situations a constant risk estimator is not minimax (e.g. unbiased estimator under SEL) but a limit of Bayes estimators w.r.t. a sequence of priors. For such situations we have the following result.

Theorem Let $\{\Pi_j\}_{j\ge1}$ be a sequence of prior and let $\{r_j\}_{j\ge1}$ be the sequence of Bayes risks of corresponding Bayes estimators of $\eta = g(\theta)$. Let δ_0 be a constant risk estimator of $\eta = g(\theta)$ with constant risk R_0. If $\lim_{j\to\infty} r_j = R_0$ then δ_0 is minimax

Proof. Fix $\varepsilon > 0$. Then $\exists\ j_0 \equiv j_0(\varepsilon)$ s.t.

$$R_0 < r_j + \varepsilon, \quad \forall\ j \geq j_0$$

Let δ^* be any estimator, ~~Then~~ and let δ_{π_j} be the Bayes estimator corresponding to prior π_j, $j = 1, 2, \ldots$. Then

$$\begin{aligned}
\sup_{\theta \in \Theta} R_{\delta_0}(\theta) &= R_0 \\
&< r_j + \varepsilon, \quad \forall\ j \geq j_0 \\
&= \int_{\Theta} R_{\delta_{\pi_j}}(\theta)\, d\pi_j(\theta) + \varepsilon, \quad \forall\ j \geq j_0, \\
&\leq \int_{\Theta} R_{\delta^*}(\theta)\, d\pi_j(\theta) + \varepsilon, \quad \forall\ j \geq j_0 \\
&\leq \sup_{\theta \in \Theta} R_{\delta^*}(\theta) + \varepsilon, \quad \forall\ j \geq j_0.
\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, we have

$$\sup_{\theta \in \Theta} R_{\delta_0}(\theta) \leq \sup_{\theta \in \Theta} R_{\delta^*}(\theta)$$

$\Rightarrow$ δ_0 is minimax.

Theorem Let $\Theta_0 \subseteq \Theta$ and let δ_0 be a minimax estimator of $\eta = g(\theta)$ when Θ_0 is the parameter space. Then δ_0 is minimax under parameter space Θ if

$$\sup_{\theta \in \Theta} R_{\delta_0}(\theta) = \sup_{\theta \in \Theta_0} R_{\delta_0}(\theta) \quad \ldots\ldots (I)$$

Proof. Suppose that (I) holds. On contrary suppose that δ_0 is not minimax on Θ. Then $\exists\ \delta^*$ such that

$$\sup_{\theta \in \Theta} R_{\delta^*}(\theta) < \sup_{\theta \in \Theta} R_{\delta_0}(\theta)$$

$$\Rightarrow \sup_{\theta \in \Theta_0} R_{\delta^*}(\theta) \leq \sup_{\theta \in \Theta} R_{\delta^*}(\theta) < \sup_{\theta \in \Theta} R_{\delta_0}(\theta) = \sup_{\theta \in \Theta_0} R_{\delta_0}(\theta)$$

$\Rightarrow$ δ_0 is not minimax under Θ_0, which leads to contradiction.

Example: Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$ rvs, where $\underline{\theta} = (\mu, \sigma) \in \Theta = \mathbb{R} \times (0, c]$, where c is a ~~unknown~~ known positive constant. Show that $\bar{X}$ is minimax for estimating μ under the SEL function.

Solution Let $\Theta_0 = \mathbb{R} \times \{c\}$. Under Θ_0 consider prior for μ as

$$\Pi_m \sim N(0, \sigma_m^2).$$

Then the Π_m-Bayes estimator is

$$\delta_{\Pi_m}(\underline{X}) = \frac{\bar{X}}{\frac{c^2}{n\sigma_m^2} + 1}$$

$$r_m = r_{\delta_{\Pi_m}}(\Pi_m) = \frac{c^2/n}{\frac{\sigma^2}{n\sigma_m^2} + 1} \rightarrow \frac{c^2}{n} = \text{Constant risk of } \bar{X} = \delta_0, \text{ as } \sigma_m \to \infty$$

$\Rightarrow$ $\bar{X}$ is minimax under Θ_0

$$\sup_{\underline{\theta} \in \Theta} R_{\delta_0}(\underline{\theta}) = \sup_{\underline{\theta}} \left[\frac{\sigma^2}{n}\right] = \frac{c^2}{n} = \sup_{\underline{\theta} \in \Theta_0} R_{\delta_0}(\underline{\theta})$$

$\Rightarrow$ $\bar{X}$ is also minimax under the parameter space Θ.

Assignment: Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$ r.v.s, where $\underline{\theta} = (\mu, \sigma) \in \Theta = \mathbb{R} \times (0, \infty)$ is unknown. Consider estimation of μ under the SEL function. Show that all estimators are minimax.

Soln. For fixed σ^2 since $\bar{X}$ is minimax, we have

$$\sup_{\mu \in \mathbb{R}} R_\delta(\underline{\theta}) \geq \frac{\sigma^2}{n}, \quad \forall \text{ estimator } \delta$$

$$\Rightarrow \sup_{\underline{\theta} \in \Theta} R_\delta(\underline{\theta}) = \infty, \quad \forall \text{ estimator } \delta$$

$\Rightarrow$ All estimators are minimax

Theorem An ~~an~~ admissible estimator with constant risk is strictly minimax. If the loss function $L(\underline{\theta}, a)$ is strictly convex in $a \in \mathcal{A}$, $\forall\, \underline{\theta} \in \Theta$, then an admissible estimator with constant risk is unique minimax estimator.

Proof. Suppose that δ_0 is admissible and

$$R_{\delta_0}(\underline{\theta}) = c, \qquad \forall\ \underline{\theta} \in \Theta,$$

where c is a constant.

On contrary suppose that δ_0 is not minimax. Then $\exists\ \delta^* \in \mathcal{D}$ s.t.

$$\sup_{\underline{\theta} \in \Theta} R_{\delta^*}(\underline{\theta}) < \sup_{\underline{\theta} \in \Theta} R_{\delta_0}(\theta) = c$$

$$\Rightarrow \quad R_{\delta^*}(\theta) < c = R_{\delta_0}(\theta), \qquad \forall\ \theta \in \Theta$$

$$\Rightarrow \quad \delta_0 \text{ is inadmissible} \quad (\text{a Contradiction})$$

Thus δ_0 is minimax.

Now suppose that the loss function $L(\underline{\theta}, a)$ is strictly convex in $a \in \mathcal{A}$, $\forall\ \underline{\theta} \in \Theta$. Let δ^* be any other estimator with

$$\sup_{\underline{\theta} \in \Theta} R_{\delta^*}(\underline{\theta}) = \sup_{\underline{\theta} \in \Theta} R_{\delta_0}(\theta) = c$$

and $\delta^* = \delta_0$ does not hold a.s., $\forall\ \underline{\theta} \in \Theta$. Consider

$$\delta^{**} = \frac{\delta^* + \delta_0}{2}.$$

Then

$$R_{\delta^{**}}(\theta) = E_\theta\left(L\left(\theta, \frac{\delta^* + \delta_0}{2}\right)\right)$$

$$< E_\theta\left(\frac{1}{2} L(\underline{\theta}, \delta^*) + \frac{1}{2} L(\underline{\theta}, \delta_0)\right)$$

$$= \frac{1}{2} R_{\delta^*}(\theta) + \frac{1}{2} R_{\delta_0}(\theta)$$

$$\leq \frac{c}{2} + \frac{c}{2} = c = R_{\delta_0}(\theta), \qquad \forall\ \theta \in \Theta$$

$$\Rightarrow \quad \delta_0 \text{ is inadmissible} \quad (\text{a Contradiction})$$

Hence δ_0 is unique minimax.

Assignment. (1) Let $X \sim \text{Poisson}(\theta)$, $\theta \in \Theta = (0, \infty)$. Consider estimation of θ under the SEL $L(\underline{\theta}, a) = (a-\theta)^2$, $a \in \mathcal{A} = \Theta$, $\theta \in \Theta$. Find Bayes estimator of θ under the prior Π_n having pdf $\pi_n(\theta) = \frac{1}{n^\alpha \Gamma\alpha} \theta^{\alpha-1} e^{-\frac{\theta}{n}}$, $\theta > 0$. Show that $\delta_0(X) = X$ is an admissible estimator of θ and that $\sup_{\theta \in \Theta} R_\delta(\theta) = \infty$ for any estimator δ. Can you find a minimax estimator under the loss function $L(\underline{\theta}, a) = \frac{(a-\theta)^2}{\theta}$, $a \in \mathcal{A}$, $\theta \in \Theta$.

Formulation of Testing Problem as a Decision Problem

Let $X_1, \ldots, X_n$ be i.i.d. with p.d.f. $f \in \{f_0, f_1\}$ w.r.t. a dominating measure λ; here f_0 and f_1 are known pdfs.

Consider testing

$$H_0 : f \equiv f_0 \quad \text{(null hypothesis)}$$

$$\text{against } H_1 : f \equiv f_1 \quad \text{(alternate hypothesis)}$$

MP(α) test is

$$\text{Accept } H_0 \text{ if } \frac{\prod_{i=1}^{n} f_0(x_i)}{\prod_{i=1}^{n} f_1(x_i)} > k$$

$$\text{Accept } H_1 \text{ if } \quad '' \quad < k$$

$$\text{Accept } H_0 \text{ w.p. } \gamma \ (\equiv \gamma(\underline{x})) \quad '' \quad = k$$

The randomization may be necessary to achieve the exact level α if the distribution of statistics involved is discrete.

Decision Theoretic Formulation: $\Omega = \{f_0, f_1\} \equiv \{0, 1\}$, $\mathcal{A} = \{a_0, a_1\}$, where

$$a_i : \text{accept } H_i, \quad i = 0, 1.$$

Suppose penalties for wrong actions are

True State of Nature	Actions: a_0	a_1
f_0	0	b
f_1	c	0

$b > 0$, $c > 0$

→ 0-1 loss function $L(\theta, a)$

__Decision Rule:__ $\delta : \mathcal{X} \times \mathcal{A} \to [0, 1]$

$\delta(a_0 | \underline{x}) = P(\text{taking action } a_0 | \underline{X} = \underline{x})$

$\delta(a_1 | \underline{x}) = P(\text{taking action } a_1 | \underline{X} = \underline{x}) = 1 - \delta(a_0 | \underline{x})$,

$\underline{x} \in \mathcal{X}$.

We may denote, for $\underline{x} \in \mathcal{X}$,

$$\delta(a_1 | \underline{x}) \equiv \delta(\underline{x})$$

and $\delta(a_0 | \underline{x}) \equiv 1 - \delta(\underline{x})$.

Thus a decision rule $\delta : \mathcal{X} \to [0, 1]$, where

$\delta(\underline{x}) = P(\text{taking action } a_1 | \underline{X} = \underline{x})$, $\underline{x} \in \mathcal{X}$

$1 - \delta(\underline{x}) = P(\text{taking action } a_0 | \underline{X} = \underline{x})$, $\underline{x} \in \mathcal{X}$.

$E_{f_0}(\delta(\underline{X}))$ = (unconditional) probability of taking action a_1 when sampling from f_0
(probability of type-I error)

$E_{f_1}(\delta(\underline{X}))$ = (unconditional) probability of taking action a_1 when sampling from f_1
(power of test)

$$L^*(\theta, \delta, \underline{x}) = E^{\delta(\cdot|\underline{x})}(L(\theta, a))$$

$$= (1 - \delta(\underline{x})) L(\theta, a_0) + \delta(x) L(\theta, a_1).$$

$\theta \in \Omega = \{0, 1\}$

where 0 corresponds to f_0 and 1 corresponds to f_1.

$$R_\delta(\theta) = E_\theta(L^*(\theta, \delta, \underline{X}))$$

$$= \int_{\mathcal{X}} [(1 - \delta(\underline{x})) L(\theta, a_0) + \delta(x) L(\theta, a_1)] \prod_{j=1}^{n} f_i(x_j) \, dx$$

$i \in \{0, 1\}$

$$R_\delta(0) = b \int_{\mathcal{X}} \delta(\underline{x}) \prod_{i=1}^{n} f_0(x_i) \, dx = b \, P(\text{type I error} | \delta)$$

$$R_\delta(1) = c \int_{\mathcal{X}} (1 - \delta(\underline{x})) \prod_{i=1}^{n} f_1(x_i) \, dx = c \, [1 - \text{power of test}]$$

Classical Approach: Fix $R_\delta(0)$ and then minimize $R_\delta(\phi)$. (which is the same as minimizing the power).

Bayes Rule: Let π be a prior distribution on $\Omega = \{0, 1\}$ with p.m.f.

$$\pi(0) = \alpha \quad \text{and} \quad \pi(1) = 1-\alpha,$$

where $\alpha \in [0, 1]$

Then

$$r_\delta(\pi) = \alpha R_\delta(0) + (1-\alpha) R_\delta(1)$$

$$\cancel{=}$$

$$= E^{\underline{X}}\left(E^{\Theta|\underline{X}}\left(L^*(\Theta, \delta, \underline{x})\right)\right).$$

We aim to minimize

$$E^{\Theta|\underline{X}=\underline{x}}\left(L^*(\Theta, \delta, \underline{x})\right), \quad \forall \underline{x} \in A.$$

$$P(\theta = 0 | \underline{X} = \underline{x}) = \frac{\alpha \prod_{i=1}^{n} f_0(x_i)}{\alpha \prod_{i=1}^{n} f_0(x_i) + (1-\alpha) \prod_{i=1}^{n} f_1(x_i)}, \quad \underline{x} \in A,$$

$$P(\theta = 1 | \underline{X} = \underline{x}) = \frac{(1-\alpha) \prod_{i=1}^{n} f_1(x_i)}{\alpha \prod_{i=1}^{n} f_0(x_i) + (1-\alpha) \prod_{i=1}^{n} f_1(x_i)}$$

$$E^{\Theta|\underline{X}=\underline{x}}\left(L^*(\Theta, \delta, \underline{x})\right) = (1-\delta(\underline{x}))\, E^{\Theta|\underline{X}=\underline{x}}\left(L(\Theta, a_0)\right) + \delta(x)\, E^{\Theta|\underline{X}=\underline{x}}\left(L(\Theta, a_1)\right)$$

$$= (1-\delta(\underline{x}))\, L(1, a_0)\, P(\Theta = 1 | \underline{X} = \underline{x}) + \delta(\underline{x})\, L(0, a)\, P(\Theta = 0 | \underline{X} = \underline{x})$$

$$= \frac{b\, \delta(\underline{x})\, \alpha \prod_{i=1}^{n} f_0(x_i)}{\alpha \prod_{i=1}^{n} f_0(x_i) + (1-\alpha) \prod_{i=1}^{n} f_1(x_i)} + \frac{c(1-\delta(\underline{x}))(1-\alpha) \prod_{i=1}^{n} f_1(x_i)}{\alpha \prod_{i=1}^{n} f_0(x_i) + (1-\alpha) \prod_{i=1}^{n} f_1(x_i)}$$

$$= k_1 \delta(\underline{x}) + k_2 (1-\delta(\underline{x})) \quad \cdots\cdots \quad (I),$$

Where

$$k_1(\underline{x}) = \frac{b\alpha \prod_{i=1}^{n} f_0(x_i)}{\alpha \prod_{i=1}^{n} f_0(x_i) + (1-\alpha) \prod_{i=1}^{n} f_1(x_i)}, \quad \underline{x} \in \mathcal{X}$$

and

$$k_2(\underline{x}) = \frac{c(1-\alpha) \prod_{i=1}^{n} f_1(x_i)}{\alpha \prod_{i=1}^{n} f_0(x_i) + (1-\alpha) \prod_{i=1}^{n} f_1(x_i)}, \quad \underline{x} \in \mathcal{X}.$$

Clearly (I) is minimized for $\delta(\underline{x}) \equiv \delta_\pi(\underline{x})$, where

$$\delta_\pi(\underline{x}) = \begin{cases} 0, & \text{if } k_1 > k_2 \\ \partial, & \text{if } k_1 = k_2 \\ 1, & \text{if } k_1 < k_2 \end{cases}$$

We have

$$k_1 < k_2 \iff \frac{\prod_{i=1}^{n} f_0(x_i)}{\prod_{i=1}^{n} f_1(x_i)} < \frac{c(1-\alpha)}{b\alpha}$$

Thus the Bayes rule is

$$\delta_\pi(\underline{x}) = \begin{cases} 1, & \text{if } \frac{\prod_{i=1}^{n} f_0(x_i)}{\prod_{i=1}^{n} f_1(x_i)} < \frac{c(1-\alpha)}{b\alpha} \\ \partial, & \text{if } \quad '' \quad = \quad '' \\ 0, & \text{if } \quad '' \quad > \quad '' \end{cases}$$

$\rightarrow$ form of the test is MP test

It follows that all Bayes rules are MP tests for some $\alpha \in [0,1]$.

$\boxed{30/3}$

Geometric Interpretation for Testing or Two Actions Problem.

$\Omega = \{0, 1\}$ $\qquad$ $\mathcal{A} = \{0, 1\}$

Decision rule $\delta : \mathcal{X} \times \mathcal{A} \to [0, 1]$ s.t

$\delta(x) \equiv P(\text{taking action } 1 \mid \underline{X} = \underline{x})$

$1 - \delta(\underline{x}) = P(\text{taking action } 0 \mid \underline{X} = \underline{x})$

$L(\theta, a)$: loss function, $(\theta, a) \in \Omega \times \mathcal{A}$.

$L^*(\theta, \delta, \underline{x}) = L(\theta, 0)(1 - \delta(\underline{x})) + L(\theta, 1)\delta(\underline{x})$

For a d.r. δ define

$$V(\delta) = \begin{pmatrix} R_\delta(0) \\ R_\delta(1) \end{pmatrix} \in \mathbb{R}^2$$

$\hookrightarrow$ risk vector of δ.

Let

$$\mathcal{S} = \{V(\delta) : \delta \in \mathcal{D}\} \quad (\text{risk set}).$$

Result: The risk set $\mathcal{S}$ is a convex set.

Proof: Let $\underline{x}, \underline{y} \in \mathcal{S}$ and let $\beta \in [0, 1]$. Then $\exists$ d.r.s $\delta_1 \in \mathcal{D}$ and $\delta_2 \in \mathcal{D}$ s.t.

$$\underline{x} = \begin{pmatrix} R_{\delta_1}(0) \\ R_{\delta_1}(1) \end{pmatrix} \quad \text{and} \quad \underline{y} = \begin{pmatrix} R_{\delta_2}(0) \\ R_{\delta_2}(1) \end{pmatrix}$$

Consider

$$\delta_\beta = \beta\delta_1 + (1 - \beta)\delta_2$$

Clearly $\delta_\beta \in \mathcal{S}$ (i.e., δ_β is a proper d.r.).

$$L^*(\theta, \delta_\beta, \underline{x}) = (1-\delta_\beta(\underline{x}))\, L(\theta, a_0) + \delta_\beta(\underline{x})\, L(\theta, a_1)$$

$$= [1 - \beta\delta_1(\underline{x}) - (1-\beta)\delta_2(\underline{x})]\, L(\theta, a_0) + (\beta\delta_1(\underline{x}) + (1-\beta)\delta_2(x))\, L(\theta, a_1)$$

$$= [1-\beta + \beta(1-\delta_1(x)) - (1-\beta)\delta_2(x)]\, L(\theta, a_0) + (\beta\delta_1(x) + (1-\beta)\delta_2(x))\, L(\theta, a_1)$$

$$= [\beta(1-\delta_1(\underline{x})) + (1-\beta)(1-\delta_2(x))]\, L(\theta, a_0) + [\beta\delta_1(x) + (1-\beta)\delta_2(x)]\, L(\theta, a_1)$$

$$= \beta[(1-\delta_1(x))L(\theta, a_0) + \delta_1(x)\, L(\theta, a_1)] + (1-\beta)[(1-\delta_2(x))\, L(\theta, a_0) + \delta_2(x)\, L(\theta, a_1)]$$

$$= \beta\, L^*(\theta, \delta_1, x) + (1-\beta)\, L^*(\theta, \delta_2, x)$$

$$\Rightarrow R_{\delta_\beta}(\theta) = \beta R_{\delta_1}(\theta) + (1-\beta) R_{\delta_2}(\theta) \qquad \theta \in \{0, 1\}$$

$$\Rightarrow \beta\underline{x} + (1-\beta)\underline{y} = \begin{pmatrix} R_{\delta_\beta}(0) \\ R_{\delta_\beta}(1) \end{pmatrix} \in S.$$

<u>Remark:</u> It can be shown that for any convex set $S \subseteq \mathbb{R}^2$ there exists a decision problem with S as the risk set.

Bayes Rule

Suppose $\pi: P(\theta=0) = p = 1 - P(\theta=1)$, $0 < p < 1$.

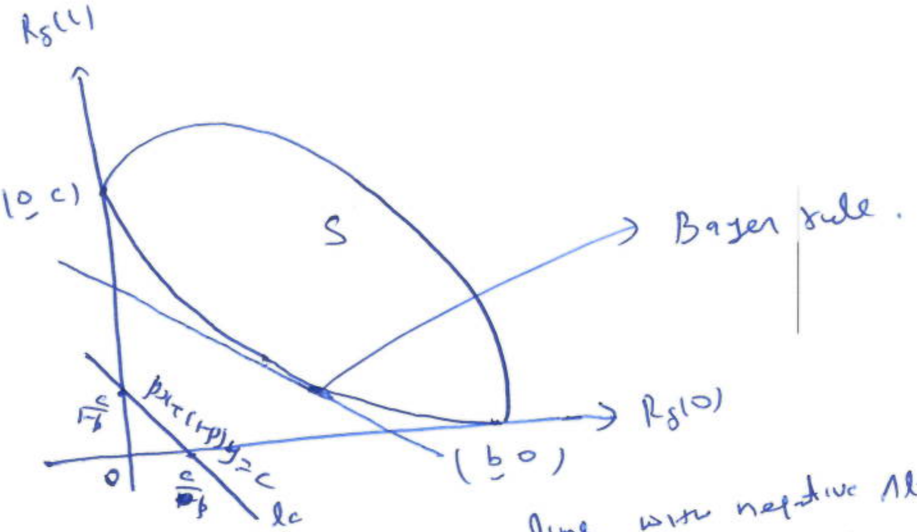
$$r_\delta(\pi) = p R_\delta(0) + (1-p) R_\delta(1) = px + (1-p)y$$

We need to minimize $px + vy$.

Consider lines $px + (1-p)y = c$

Loss matrix

		Actions 0	1
State of Nature	0	0	b
	1	c	0

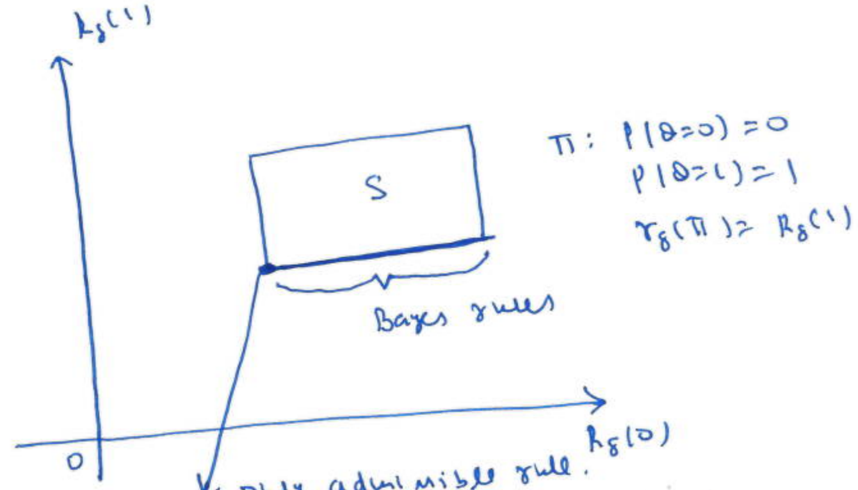


Since $l_c = \{(x,y) : px + vy = c\}$ is a line with negative slope, as $c \uparrow$ the line l_c moves up. For minimizing $r_\delta(\pi)$, p fixed, we want to pick the smallest c s.t. $l_c \cap S \neq \emptyset$, then if $(x_0, y_0) \in l_c \cap S$, $(x_0, y_0) = V(\delta_0)$, then δ_0 is Bayes w.r.t. π.

If the above picture truly represents the risk set then the admissible rules correspond to the lower boundary of S. But we have just shown that the Bayes rules are on the lower boundary. Thus we may conjecture that

set of admissible rules = set of Bayes rules.

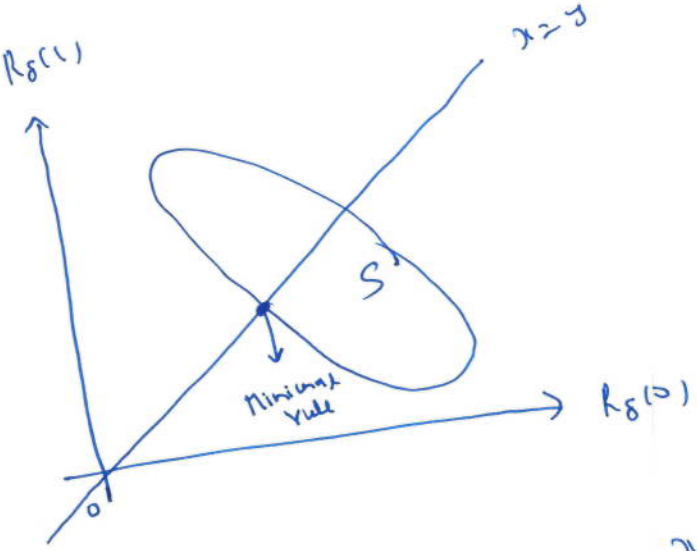
However it is false as shown below.

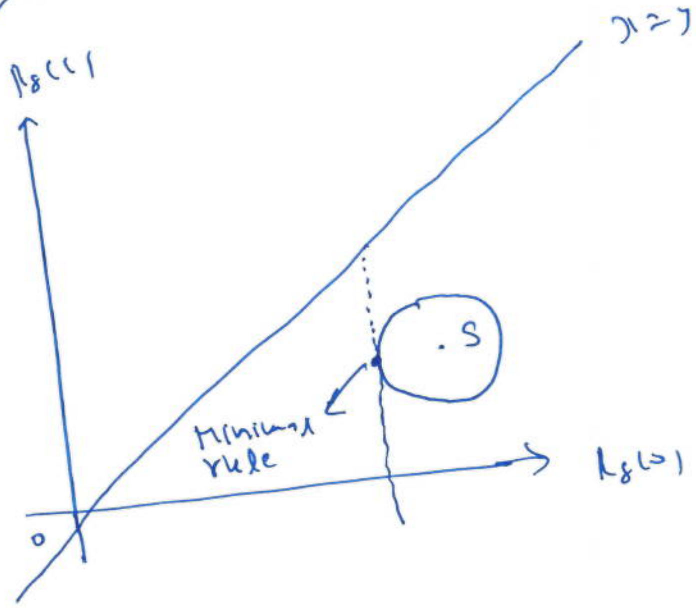


Except for left most vertex all the points on the line which are Bayes solutions are corresponding to inadmissible rules. Also in this example we have only one minimax solution.

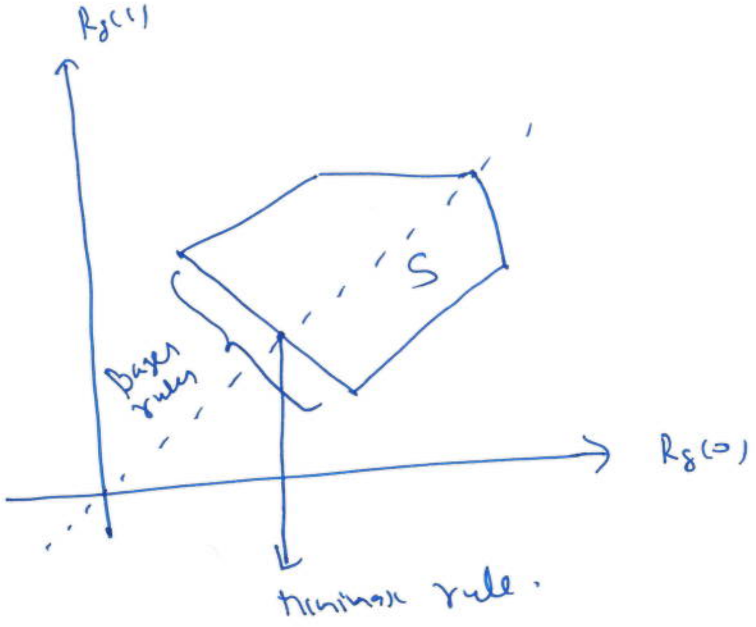
Remark: Most of the above discussion were heuristic, without formal theory. For example we must prove that the risk set is closed before we can pick off lower boundary points. It can be shown that this is true in many situations but we will not provide the details here.

Minimax Rule





Bayes Rule may not be minimax



Minimax Rule may not be Unique