5 Minimaxity

- $\delta \in \mathcal{D}$ is <u>minimax</u> if $\sup_{\theta \in \Theta} R(\theta, \delta) = \inf_{\delta' \in \mathcal{D}} \sup_{\theta \in \Theta} R(\theta, \delta')$

$= \overline{V} =$ upper (minimax) value of the decision problem

- generally a minimax rule exists but it may not be unique or might even be absurd.

Lemma 1 $\sup_{\theta \in \Theta} R(\theta, \delta) = \sup_{\pi} E_{\pi}(R(\theta, \delta))$

Proof: Let $\varepsilon > 0$ and $\Theta_\varepsilon = \{\theta : R(\theta, \delta) \geq L_\varepsilon\}$ where $L_\varepsilon \uparrow \sup_\theta R(\theta, \delta)$ as $\varepsilon \downarrow 0$. Then if $\pi_\varepsilon(L_\varepsilon) = 1$, $L_\varepsilon \leq E_{\pi_\varepsilon}(R(\theta, \delta)) \leq \sup_{\theta \in \Theta} R(\theta, \delta)$.

Corollary 2 $\overline{V} = \inf_{\delta} \sup_{\pi} E_{\pi}(R(\theta, \delta))$

eg <u>estimating mean of $N(\mu, \sigma_0^2)$</u>

- $\mu \in \mathbb{R}$, $\Psi(\mu) = \mu$

$$L(\mu, a) = (\mu - a)^2$$

- problem is convex so restrict to nonrandomized rules based on MSS $\bar{x}$

- consider prior $\pi = N(0, \tau_0^2)$ then posterior is

$$\mu \mid \bar{x} \sim N\left(\left(\frac{n}{\sigma_0^2}+\frac{1}{\tau_0^2}\right)^{-1}\left(\frac{n}{\sigma_0^2}\bar{x}+\frac{1}{\tau_0^2}0\right), \left(\frac{n}{\sigma_0^2}+\frac{1}{\tau_0^2}\right)^{-1}\right)$$

so $d(\bar{x}) = \left(\frac{n}{\sigma_0^2}+\frac{1}{\tau_0^2}\right)^{-1}\frac{n}{\sigma_0^2}\bar{x} = a\bar{x}$ is the Bayes rule

and the Bayes risk is $E_\pi(R(\mu,d))$

where $R(\mu,d) = E_{N(\mu,\frac{\sigma_0^2}{n})}((\mu - a\bar{x})^2)$

$$= (\mu - a\mu)^2 + a^2\frac{\sigma_0^2}{n} = (1-a)^2\mu^2 + a^2\frac{\sigma_0^2}{n}$$

$$\therefore E_\pi(R(\mu,d)) = (1-a)^2\tau_0^2 + a^2\frac{\sigma_0^2}{n}$$

$$\le E_\pi(R(\mu,d^*)) \le \sup_\mu R(\mu,d^*)$$

$\forall d^* \in D$ and $\forall a = \frac{n/\sigma_0^2}{n/\sigma_0^2 + 1/\tau_0^2} \in (0,1)$

$$\therefore \lim_{\tau_0^2\to\infty} (1-a)^2\tau_0^2 + a^2\frac{\sigma_0^2}{n} = \frac{\sigma_0^2}{n} \le \sup_\mu R(\mu,d^*)$$

$\forall d^*$ and since $R(\mu,\bar{x}) \equiv \sigma_0^2/n$

this proves $\bar{x}$ is minimax.

- this argument can be generalized to the problem $X = (x_1,\ldots,x_n) \overset{iid}{\sim} N_k(\mu, \sigma_0^2 I)$ with $L(\mu, t) = \|\mu - t\|^2$ (Exercise)

- also there are many minimax rules here.

eg - $x = (x_1, \ldots, x_n) \sim N(\mu, \sigma^2)$ $\quad (\mu, \sigma^2) \in R^1 \times (0, \infty)$

- $\Psi(\mu, \sigma^2) = \mu$.

- $L((\mu, \sigma^2), a) = (\mu - a)^2$

- $L((\mu, \sigma^2), \cdot)$ is convex so restrict to nonrandomized rules

- put $d_0(x) = \bar{x}$

- so $R((\mu, \sigma^2), \bar{x}) = E_{(\mu, \sigma^2)}((\mu - \bar{x})^2)$

$= Var_{(\mu, \sigma^2)}(\bar{x}) = \frac{\sigma^2}{n}$

- for fixed σ^2, d_0 is minimax

- so $\sup_{\mu} R((\mu, \sigma^2), d) \geq \frac{\sigma^2}{n}$ for any d

- $\therefore \sup_{(\mu, \sigma^2)} R((\mu, \sigma^2), d) \geq \sup_{(\mu, \sigma^2)} R((\mu, \sigma^2), d_0) = \infty$

$\therefore$ any d is minimax

- so problems with minimaxity as a general principle

- solution (?) : restrict to subclass $\mathcal{D}_0$ and look for constrained minimax (local minimax) procedure (strategy (?))

Hypothesis testing

$$L(\theta,a) = \begin{cases} c_0 & \theta \in H_0,\ a=1 \\ c_1 & \theta \in H_a,\ a=0 \\ 0 & \text{otherwise} \end{cases}$$

$$R(\theta,\phi) = \begin{cases} c_0 E_\theta[\phi] & \theta \in H_0 \\ c_1(1-E_\theta[\phi]) & \theta \in H_a \end{cases}$$

$$\sup_\theta R(\theta,\phi) = \max\left\{ c_0 \sup_{\theta\in H_0} E_\theta[\phi],\ c_1\left(1-\inf_{\theta \in H_a} E_\theta[\phi]\right)\right\}$$

- suppose $c_0 = c_1 = 1$ and we restrict to size α.
- $\phi_\alpha(x) \equiv \alpha$ has size α and power α
- if $\alpha < \frac{1}{2}$ then $\alpha < 1-\alpha$ and so

$$\sup_\theta R(\theta,\phi_\alpha) = 1-\alpha = 1-\inf_{\theta\in H_a} E_\theta[\phi_\alpha]$$

$\therefore$ when $\alpha < \frac{1}{2}$ need only look for test maxing $\inf_{\theta \in H_a} E_\theta[\phi_\alpha]$ called a size α maximin ~~test~~

- locally minimax tests

- $\underline{V} = \sup_{\pi} \inf_{\delta} r_\pi(\delta)$ = lower (maximin) value of the dec problem

- if a prior exists that gives Bayes risk equal to $\underline{V}$ it is called a least favorable prior

Lemma 3 $\underline{V} \le \overline{V}$ (the problem has a value if $\underline{V} = \overline{V}$)

Proof: For any δ^*, π $\inf_{\delta} r_\pi(\delta) \le r_\pi(\delta^*)$

so $\underline{V} = \sup_{\pi} \inf_{\delta} r_\pi(\delta) \le \sup_{\pi} r_\pi(\delta^*) \overset{L1}{=} \sup_{\theta} R(\theta, \delta^*)$

and so $\underline{V} \le \inf_{\delta} \sup_{\pi} r_\pi(\delta) = \overline{V}$.

Theorem 4 If δ_0 is a Bayes rule wrt π and $R(\theta, \delta_0) \le r_\pi(\delta_0)$ $\forall \theta$ then δ_0 is minimax, problem has value $r_\pi(\delta_0)$ and π is least favorable.

Proof: $\overline{V} \le \sup_{\theta} R(\theta, \delta_0) \le r_\pi(\delta_0) = \inf_{\delta} r_\pi(\delta) \le \underline{V}$

and so $\underline{V} = \overline{V}$ which implies the result.

eg – if δ_0 is Bayes wrt $\bar{\pi}$ and has constant risk then it is minimax & $\bar{\pi}$ is least favorable and $\underline{V} = \bar{V}$

eg – $x = (x_1, \ldots, x_n) \overset{iid}{\sim}$ Bernoulli(θ), $\theta \in (0,1)$

– $\Psi(\theta) = \theta \in (0,1)$

– $L(\theta, \psi) = \dfrac{(\theta - \psi)^2}{\theta(1-\theta)}$ (absolute errors more important when $\theta \approx 0$ or $\theta \approx 1$)

– $\bar{\pi} = U(0,1) = \text{beta}(1,1)$

$\therefore \theta | x \overset{Ex}{\sim} \text{beta}(n\bar{x}+1, n(1-\bar{x})+1)$

– Bayes rule is $\bar{x}$ (Ex)

– $R(\theta, \bar{x}) = \frac{1}{n}$ $\left(\text{Var}_\theta(\bar{x}) = \frac{\theta(1-\theta)}{n} \right)$

$\therefore \bar{x}$ is minimax and $\bar{\pi}$ is least favorable

Def δ_0 is extended Bayes if $\forall \varepsilon > 0$ $\exists$ a prior π_ε s.t. $r_{\pi_\varepsilon}(\delta_0) \le \inf_\delta r_{\pi_\varepsilon}(\delta) + \varepsilon$

(so δ_0 is ε-Bayes wrt π_ε $\forall \varepsilon$)

Theorem 5 If δ_0 is extended Bayes with constant risk then δ_0 is minimax.

Proof: Suppose δ_0 is not minimax and $c = R(\theta, \delta_0)$.

Then $\exists$ δ_1 st $\sup_\theta R(\theta, \delta_1) = \sup_\theta R(\theta, \delta_0) - \varepsilon = c - \varepsilon$

for some $\varepsilon > 0$. But for $\pi_{\varepsilon/2}$, $r_{\pi_{\varepsilon/2}}(\delta_0)$

$\forall\theta$
$= R(\theta, \delta_0) \overset{=c}{\le} \inf_\delta r_{\pi_{\varepsilon/2}}(\delta) + \frac{\varepsilon}{2} \overset{\forall\theta}{\le} c - \varepsilon + \frac{\varepsilon}{2}$

$= c - \frac{\varepsilon}{2}$ (X)

eg $x \sim$ Poisson (θ)

- $\Theta = (0, \infty)$, $\Psi(\theta) = \theta$

- $L(\theta, a) = \dfrac{(\theta - a)^2}{\theta}$ (convex)

- $d_0(x) = x$, $R(\theta, d_0) = 1$ $(\mathrm{Var}_\theta(x) = \theta)$

- $\pi_\varepsilon = \text{gamma}_{\text{rate}}(1, \frac{\varepsilon}{1-\varepsilon})$

- so likelihood $\times$ prior $\dfrac{\theta^x}{x!} e^{-\theta} \left(\dfrac{\varepsilon}{1-\varepsilon}\right) \exp\left(-\left(\dfrac{\varepsilon}{1-\varepsilon}\right)\theta\right)$

$$\propto \theta^x \exp\left\{-\left(1+\frac{\epsilon}{1-\epsilon}\right)\theta\right\}$$

$$= \theta^x \exp\left\{-\frac{1}{1-\epsilon}\,\theta\right\}$$

$$\therefore\ \theta \mid x \sim \text{gamma}_{\text{rate}}\left(x+1, \frac{1}{1-\epsilon}\right)$$

and Bayes rule minimizes

$$E\left[\frac{(\theta-d(x))^2}{\theta} \,\Big|\, x\right]$$

$$= \int_0^\infty \frac{(\theta-d(x))^2}{\theta}\,\frac{1}{\Gamma(x+1)}\left(\frac{\theta}{1-\epsilon}\right)^x e^{-\theta/(1-\epsilon)}\,\frac{1}{1-\epsilon}\,d\theta$$

$$= \frac{1}{x}\,\frac{1}{1-\epsilon}\int_0^\infty (\theta-d(x))^2\,\frac{1}{\Gamma(x)}\left(\frac{\theta}{1-\epsilon}\right)^{x-1} e^{-\theta/(1-\epsilon)}\,\frac{1}{1-\epsilon}\,d\theta$$

$$\sim \text{gamma}_{\text{rate}}\left(x, \frac{1}{1-\epsilon}\right)$$

$\therefore$ Bayes rule is $x(1-\epsilon)$ with post. risk

$$\frac{1}{x}\,\frac{1}{1-\epsilon}\,x(1-\epsilon)^2 = 1-\epsilon$$ which is the Bayes risk

- now $r_{\pi_\epsilon}(d_0) = 1$ and $\therefore$ d_0 is extended Bayes and $\therefore$ minimax

Theorem 6 (Minimax Theorem)
If $\Theta = \{\theta_1, \dots, \theta_n\}$ then $\underline{V} = \overline{V}$ and there exists a least-favorable prior π_0. If R is closed from below then $\exists$ an admissible minimax s_0 and s_0 is Bayes wrt π_0.

Proof: Put $V = \sup\{\alpha : Q_{\alpha \underset{\sim}{1}} \cap \overline{R} = \phi\}$. Then for each n $\exists s_n$ s.t. $\sup_\theta R(\theta, s_n) \leq V + \frac{1}{n}$, so $r_\pi(s_n) \leq V + \frac{1}{n}$ for any prior π and so $\sup_\pi r_\pi(s_n) \leq V + \frac{1}{n}$. This implies $\overline{V} \leq V$.

Now $Q^0_{V\underset{\sim}{1}}$ and $\overline{R}$ are disjoint convex sets. Thus by the Separating Hyperplane Thm $\exists \underset{\sim}{p}, c$ s.t. $\underset{\sim}{p}'\underset{\sim}{x} \geq c \ \forall \underset{\sim}{x} \in \overline{R}$ and $\underset{\sim}{p}'\underset{\sim}{x} \leq c \ \forall \underset{\sim}{x} \in Q^0_{V\underset{\sim}{1}}$ and as before each $p_i \geq 0$, $\sum_{i=1}^n p_i = 1$. Let $\pi_0(\{\theta_i\}) = p_i$. Now $\underset{\sim}{p}' (V-\varepsilon)\underset{\sim}{1} = V - \varepsilon \leq c \ \forall \varepsilon > 0$ and so $V \leq c$. Also $\forall s$, $r_{\pi_0}(s) = \sum_{i=1}^n p_i R(\theta_i, s) \geq c$ and so $\underline{V} = \sup_\pi \inf_s r_\pi(s) \geq V$ which implies $V \leq \underline{V} \leq \overline{V} \leq V$ which proves $\underline{V} = \overline{V}$

and π_0 is least favourable (ie. $V \le c \le \inf_s r_{\pi_0}(s) \le \sup_\pi \inf_s r_\pi(s) \le V$)

Consider the sequence of points $(R(\Theta_1, s_N), \ldots, R(\Theta_k, s_N))$ $\le (V + \frac{1}{n})\underset{\sim}{1}$. Since the sequence is in $[0, V+\frac{1}{n}]^k$ $\exists$ a limit point $\underset{\sim}{r}_0 \in \bar{R}$ and since $Q_{\underset{\sim}{r}_0} \wedge \bar{R} \neq \phi$, the proof of Thm 4.4.(i) implies $\lambda(Q_{\underset{\sim}{r}_0} \wedge \bar{R}) \neq \phi$. For $\underset{\sim}{s}_0 \in \lambda(Q_{\underset{\sim}{r}_0} \wedge \bar{R})$ we have $\{\underset{\sim}{s}_0\} = Q_{\underset{\sim}{s}_0} \wedge Q_{\underset{\sim}{r}_0} \wedge \bar{R} = Q_{\underset{\sim}{s}} \wedge \bar{R}$ and so $\underset{\sim}{s}_0 \in \lambda(R)$. Thus $s_0 \in R$ st. $R(\Theta_i, s_0) = s_{0i}$ for $i = 1, \ldots, k$ is admissible (Thm 4.3) and $r_{\pi_0}(s_0) = \sum_{i=1}^k p_i s_{0i} \le \sum_{i=1}^k r_{0i} = V$. But we already showed $r_{\pi_0}(s) \geq c \geq V$ $\forall s$ and so $r_{\pi_0}(s_0) = V$ so s_0 is Bayes wrt π_0.