II.2 Strategies for "Solving" Decision Problems

- 3 basic strategies.

(i) Restrict the class $\mathcal{D}$

- use $\mathcal{D}_0 \subseteq \mathcal{D}$ and look for optimal $\delta_0 \in \mathcal{D}_0$

eg unbiasedness, invariance later

- note - if for each $\delta \in \mathcal{D}$ $\exists\ \delta_0 \in \mathcal{D}_0$ s.t. $R(\theta, \delta_0) \le R(\theta, \delta)$ $\forall \theta$ then there is no loss in this.

Lemma ① If Ψ is convex, $L(\theta, \cdot)$ is convex for all $\theta \in \Theta$, $d_\delta(x) = \int_\Psi t\, \delta(x, dt) \in \Psi$ $\forall x$ (so $d_\delta(x) \in \Psi$ since Ψ is convex) then $R(\theta, \delta) \ge R(\theta, d_\delta)$ $\forall \theta$.

Proof: $R(\theta, \delta) = \int_X \int_\Psi L(\theta, t)\, \delta(x, dt)\, P_\theta(dx)$

Jensen
$\ge \int_X L(\theta, d_\delta(x))\, P_\theta(dx) = R(\theta, d_\delta)$

- note - this inequality is strict unless $\delta(x, \cdot)$ is degenerate at a.e. $\Psi(\theta)$ a.e. P_θ

eg $L(\theta, t) = (\Psi(\theta) - t)' A (\Psi(\theta) - t)$

- if $\int_\Psi t\, \delta(x, dt)$ doesn't exist then $\int L(\theta, a) \delta(x, da) = \infty$

- define δ^* by $\delta^*(x, \cdot) = \delta(x, \cdot)$ when $d_\delta(x)$ exists and by $\delta^*(x, \cdot) =$ degenerate at a_0 otherwise, then $R(\theta, \delta^*) \le R(\theta, \delta)$

- so in this problem we can replace $\mathcal{D}$ by $\mathcal{D}_0 = \{\delta : d_\delta(x) \text{ exists } \forall x \in \mathcal{X}\}$ and, then replace $\mathcal{D}_0$ by $D_0 = \{d_\delta : \delta \in \mathcal{D}_0\}$ by Lemma 5

Lemma 2 Suppose T is sufficient and $\delta \in \mathcal{D}$. Then define $\delta_T(x, \cdot)$ by $z \sim P(\cdot \mid T)(T(x))$ and $a \sim \delta(z, \cdot)$. Then $\delta_T(x_1, \cdot) = \delta_T(x_2, \cdot)$ when $T(x_1) = T(x_2)$ and $R(\theta, \delta) = R(\theta, \delta_T)$ $\forall \theta$.

Proof: For $A \subseteq \mathcal{A}$ we have

$\delta_T(x_1, A) = \int_{\mathcal{X}} \delta(z, A) P(dz \mid T)(T(x_1))$

$= \delta_T(x_2, A)$. Also $R(\theta, \delta)$

$= \int_{\mathcal{X}} \int_{\mathcal{A}} L(\theta, a) \delta(x, da) P_\theta(dx)$

$= \int_{\mathcal{T}} \int_{\mathcal{X}} \int_{\mathcal{A}} L(\theta, a) \delta(z, da) P(dz \mid T)(t) P_{\theta T}(dt)$

$= \int_{\mathcal{X}} \int_{\mathcal{A}} L(\theta, a) \int_{\mathcal{X}} \delta(z, da) P(dz \mid T)(T(x)) P_\theta(dx)$

$= \int_{\mathcal{X}} \int_{\mathcal{A}} L(\theta, a)\, \delta_T(x, da)\, P_\theta(dx)$ (9)

$= R(\theta, \delta_T)$,

- so we can always replace $\mathcal{D}$ by $\mathcal{D}_T = \{\delta : \delta$ depends on x only thru MSS $T\}$

- makes sense to replace $\mathcal{X}$ by $\mathcal{T}$ and consider only δ s.t. $\delta(t, \cdot)$ is a prob. meas on $\mathcal{A}$ when T is a MSS

- note if δ is nonrandomized then δ_T will in general be random (hypothesis testing)

Lemma (3) (Rao-Blackwell)

If $\mathcal{A}$ is convex, $L(\theta, \cdot)$ is convex $\forall \theta$, T is suff., $d : \mathcal{X} \to \mathcal{A}$ is s.t. $d_T(x) = E_{P(\cdot|T)(T(x))}(d)$ exists then $d_T(x_1) = d_T(x_2)$ when $T(x_1) = T(x_2)$ and $R(\theta, d_T) \leq R(\theta, d)$.

Proof: $R(\theta, d) = \int_{\mathcal{X}} L(\theta, d(x))\, P_\theta(dx)$

$= \int_{\mathcal{T}} \int_{\mathcal{X}} L(\theta, d(x))\, P(dx \mid T)(T(x))\, P_{\theta T}(dt)$

Jensen $\geq \int_{\mathcal{T}} L(\theta, E_{P(\cdot|T)(t)}(d))\, P_{\theta T}(dt)$

$= \int_{\mathcal{X}} L(\theta, d_T(x))\, P_\theta(dx) = R(\theta, d_T)$

- Rao-Blackwellize an estimator = average estimator wrt cond. given MSS

- in quadratic loss estimation problems we can restrict eg Hypothesis Testing

- $R(\theta, \phi) = c_0 E_\theta(\phi) I_{H_0}(\theta) + c_a (1 - E_\theta(\phi)) I_{H_a}(\theta)$

- can't find ϕ simultaneously minimizing $E_\theta(\phi)$ when $\theta \in H_0$ and maximizing $E_\theta(\phi)$ when $\theta \in H_a$.

- restrict attention to size α test fns, namely, $E_\theta(\phi) \le \alpha \;\; \forall \theta \in H_0$, say $\phi \in \mathcal{D}_\alpha$

- then try to find the size α test fn that maximizes $E_\theta(\phi)$ for $\theta \in H_a$.

+ "sometimes" such a solution exists in $\mathcal{D}_\alpha$

- there are other possibilities, such as find ϕ minimizing
$$\alpha \sup_{\theta \in H_0} E_\theta(\phi) + (1-\alpha) \sup_{\theta \in H_a} (1 - E_\theta(\phi))$$

(ii) <u>Report the class of admissible rules</u>

- δ is better than δ' if $R(\theta,\delta) \le R(\theta,\delta')$ $\forall\theta$ and $\exists\, \theta_0$ s.t. $R(\theta_0,\delta) < R(\theta_0,\delta')$

- δ is as good as δ' if $R(\theta,\delta) \le R(\theta,\delta')$ $\forall\theta$ and they are equivalent if $R(\theta,\delta) = R(\theta,\delta')$ $\forall\theta$

- **Def** δ is admissible if $\nexists\, \delta'$ s.t. δ' is better than δ and $A \subseteq D$ be the class of all admissible rules

- **note** - admissibility is defined relative to L

- so one solution to the decision problem is to simply state A

Lemma 4 If $\exists\, \theta_0$ s.t. $P_\theta \ll P_{\theta_0}$ $\forall\theta$ then δ_{θ_0} is admissible.

Proof: Suppose $\exists\, \delta$ as good as δ_{θ_0}. Then

$0 = R(\theta_0, \delta_{\theta_0}) = R(\theta_0, \delta)$ which

implies δ is also degenerate at $\Psi(\theta_0)$ a.e. P_{θ_0}

which implies $\delta = \delta_{\theta_0}$ a.e. P_θ $\forall\theta$

- so admissibility is necessary (for decision theory) but not sufficient.

<u>Def</u> $\mathcal{C}$ is a (essentially) complete class of decision functions if for any $\delta \notin \mathcal{C}$ $\exists$ $\delta_0 \in \mathcal{C}$ st. δ_0 is (as good as) better than δ.

<u>Lemma 5</u> If $\mathcal{C}$ is a complete class then $A \subseteq \mathcal{C}$.

Proof: If $\delta \in A \setminus \mathcal{C}$ then $\exists$ $\delta_0 \in \mathcal{C}$ that is better than δ ($\Rightarrow \Leftarrow$)

<u>Lemma 6</u> If $\mathcal{C}$ is essentially complete and $\delta \in A \setminus \mathcal{C}$ then $\exists$ $\delta' \in \mathcal{C}$ st. δ' is eq. to δ.

Proof: By hypothesis $\exists$ $\delta' \in \mathcal{C}$ st $R(\theta,\delta') \le R(\theta,\delta)$ $\forall \theta$ and so $R(\theta,\delta') = R(\theta,\delta)$ $\forall \theta$ since δ is admissible.

<u>Def</u> $\mathcal{C}$ is minimal (essentially) complete if $\mathcal{C}$ is (essentially) complete and no proper subclass is (essentially) complete.

- so a solution to the decision problem is to reveal a minimal complete or essentially complete class. (max reduction)

Theorem 7 If a minimal complete class $\mathcal{C}$ exists then $\mathcal{C} = \mathcal{A}$.

Proof: We have $\mathcal{A} \subseteq \mathcal{C}$. So suppose $\delta_0 \in \mathcal{C} \setminus \mathcal{A}$. Since $\delta_0 \notin \mathcal{A}$ $\exists$ δ_{00} s.t. δ_{00} is better than δ_0.

Consider $\mathcal{C}_1 = \mathcal{C} \setminus \{\delta_0\}$. If $\delta_{00} \notin \mathcal{C}$ then $\exists$ $\delta \in \mathcal{C}$ s.t. δ is better than δ_{00} and so δ is better than δ_0 and $\delta \in \mathcal{C}_1$. Also δ is better than any δ' that δ_0 is better than. If $\delta_{00} \in \mathcal{C}$ then $\delta_{00} \in \mathcal{C}_1$ and better than any δ' that δ_0 is better than. In either case this shows $\mathcal{C}_1$ is complete ⊗. Therefore $\mathcal{C} \subseteq \mathcal{A}$.

eg sufficiency

- we should class of δ dependent on data only thru the value of a sufficient statistic is essentially complete.

eg nonrandomized rules

- $L(\theta, \cdot)$ convex $\forall_\theta$ on convex $\mathcal{A}$
- if $\int_{\mathcal{A}} L(\theta, a) \delta(x, da) < \infty$ implies $\int_{\mathcal{A}} a\, \delta(x, da) \in \mathcal{A}$ then the class of nonrandomized rules is essentially complete

- given that $\mathcal{A}$ contains many rules that are absurd this is not really a solution.

- also, there are inadmissible rules that seem much more natural than the rules that are better

(iii) Apply a Total ordering to $\mathcal{D}$

- assign a single number to each δ that orders the δs
- leads to minimax (worst case) and Bayes rules (average case)
- these satisfy: if δ_1 is as good as δ_2 then it will be under both minimax and Bayes criteria.
- still dependent on loss L
- Bayes rules require the addition of another ingredient: the prior π on Θ
- typically minimax or Bayes produce a solution although not necessarily unique.